

Persistent and Tangential Homology for Defect Classification in Time of Flight Diffraction

Dissertation zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik
und Naturwissenschaften

Fachbereich Mathematik
der Universität Hamburg

vorgelegt von
Jose Fernando Cuenca Jimenez

Hamburg, 2016

Als Dissertation angenommen vom Fachbereich Mathematik der
Universität Hamburg

Auf Grund der Gutachten von:

Prof. Dr. Armin Iske (Universität Hamburg)
und Prof. Dr. Gerlind Plonka-Hoch (Universität Göttingen)

Hamburg, den 09.11.2016

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, den 29. August 2016

Contents

Introduction	vii
1 Ultrasonic Non-Destructive Testing	1
1.1 First steps	1
1.1.1 Why ultrasonic NDT	1
1.1.2 Basic principles of ultrasonic NDT	1
1.1.3 Advantages and disadvantages of ultrasonic NDT	3
1.2 Physics of ultrasound	4
1.2.1 Wave propagation	4
1.2.2 Huygens principle	5
1.2.3 Diffraction	6
1.2.4 Reflection	6
1.2.5 Refraction	7
1.3 Equipment and transducers: the piezoelectric effect	9
1.4 Ultrasonic NDT methods	10
1.4.1 The Time of Flight Diffraction Method (TOFD)	12
1.4.2 Wall thickness measurement	13
1.5 Welding defects	15
1.5.1 Lack of deposition	16
1.5.2 Lack of penetration	17
1.5.3 Over deposition	17
1.5.4 Undercut	18
1.5.5 Spatter	18
1.5.6 Surface cracks	19
1.5.7 Internal cracks	19
1.5.8 Slag inclusions	20
1.5.9 Porosity	21
1.5.10 Wormholes	21
1.5.11 Lamellar tear	22
2 Automatic classification of TOFD defects: state of the art	25
2.1 Image pre-processing stage	25
2.1.1 Sparse deconvolution methods for NDT	26
2.2 Image segmentation stage	29
2.2.1 Density-based spatial clustering of applications with noise (DB-SCAN)	31

2.3	Image classification stage	32
2.3.1	Statistical classification, linear classifiers and artificial neural networks	33
3	An introduction to topological data analysis	39
3.1	Simplices and simplicial complexes	39
3.2	Filtrations. The Vietoris-Rips complex	40
3.3	Homology groups and Betti numbers	41
3.4	Barcodes and persistent diagrams	44
3.5	Persistent homology with different birth times	45
3.6	Other filtrations	46
3.6.1	Čech complex	46
3.6.2	Alpha complex	47
3.6.3	Witness complex	49
3.6.3.1	Definition of the witness complex $W(D)$	50
3.6.3.2	The weak witnesses theorem	51
3.6.3.3	How to choose the landmarks	53
3.6.3.4	Building a witness complex filtration	54
3.6.3.5	Example: Lazy witness filtration for points on S^2	59
4	Discrete Morse Theory	61
4.1	Definitions	61
4.2	Basic elements of discrete Morse theory	62
4.3	Main theorem of discrete Morse theory	64
4.4	Morse inequalities	69
4.5	Gradient vector fields	70
4.6	Computing persistent homology	75
4.7	Stability of persistent diagrams	79
5	Classification methods	83
5.1	Classification method 1: persistent homology and TOFD defects	83
5.2	Classification method 2: tangential homology and TOFD defects	87
5.2.1	Part 1: creation of a catalogue of TOFD defects	88
5.2.1.1	Tangent complex $T(X)$	90
5.2.1.2	Approximated $T(X)$	91
5.2.1.3	Approximated $T(X)$ filtered by curvature	92
5.2.2	Part 2: classification of a new defect	94
6	Nearest neighbour search	97
6.1	Introduction and motivation of kd -trees	97
6.2	Construction of $2d$ -trees	97
6.3	Some operations in kd trees	98
6.3.1	Range search in a $2d$ -tree	99
6.3.2	Nearest neighbour search in a $2d$ -tree	100

7	Numerical Results	105
7.1	Results: persistent homology and TOFD defects	105
7.1.1	Test Case 1: simulated TOFD Crack	105
7.1.2	Test Case 2: real TOFD data with lack of fusion	105
7.1.3	Test Case 3: simulated TOFD data with two pore defects	108
7.2	Results: tangential homology and TOFD defects	108
8	Summary and conclusions	113
	CURRICULUM VITAE - September 2016	123

Acknowledgments

First of all, I would like to say that completing my PhD would not have been possible without the support and guidance that I received from many people.

I would like to thank my supervisor Prof. Dr. Armin Iske for giving me the opportunity to do a PhD at the University of Hamburg. He has always helped me contributing with new ideas, supervision and encouragement. I have never worked before in a research group with such a wonderful atmosphere, and with so many motivated and friendly people.

Further, I would like to thank Mannesmann Salzgitter Forschung GmbH, and especially Oliver Nemitz, who have provided me with real data since the beginning of my PhD. Moreover, a big thanks to all the members of the ZeMat Project and especially to Dr. Florian Boßmann for all the interesting and productive discussions during these three years.

Adeleke, Benedikt, Claus, Libor, Marcel, Matthias, Mijail and Sara, thank you for being so nice to me since I arrived to Germany. You rock!

Naturally, I could have never finished this work without the continuous support of my family and my friends. They have been there for me since I came to this world and especially during these three years that I have been living abroad.

Last, a huge thanks to my girlfriend. She has made my life better in infinite ways since I know her. Every weekend has been an adventure in the last years. I will always remember our salsa dances in the street!

Introduction

Non-destructive testing (NDT) techniques are widely used in industry to evaluate properties of a component, material or system without causing any damage. One of the main advantages of these methods is that they do not permanently alter the material being inspected, and therefore time and money can be saved. The employment of ultrasonic waves meant a big achievement for NDT since until that point, x-rays and radioactive isotopes were the best techniques and the equipment was big and expensive as well as the testing process pretty slow.

Among the NDT techniques based on ultrasonic waves for defect detection, the Time of Flight Diffraction (TOFD) is one of the most sophisticated methods, with high significance in many applications of steel industry, where welding quality is essential to avoid productivity losses. TOFD has different advantages that make it desirable to other techniques. For instance, defect detection does not depend on the defect orientation. Moreover, it provides sizing accuracy and a high probability of detection, reducing the risk of false indications.

Notwithstanding, the interpretation of TOFD images and classification of defects is still done manually and therefore frequently questioned since it mostly depends on the experience of the human operator, and when the data volume increases, this can lead to errors. In consequence, the recent trend is to fully automate the classification of defects. To this end, different approaches have been used in the past, including artificial neural networks [27, 69], mathematical morphology [64], Hough transform [18, 97] to propose match filtering for parabola detection [79], etc.

In this thesis we propose two innovative methods for automated detection and classification of defects. Both are based on the principles of persistent homology, moreover the second method goes one step further using tangential homology, which is persistent homology applied to some suitable tangential constructions [29, 25].

Persistent homology is a recently and powerful developed algebraic method for detecting topological features of data, especially big and noisy data sets [39, 40, 104, 28]. The fields of applications of this tool has increased a lot in the last years, and some important examples can be found in areas like biological systems [73, 21], sensor networks [36], Alzheimer [77], computer graphics [54, 44], chemistry [100] and computer vision [24]. The key idea of persistent homology is to build a filtration from a Point Cloud Data (PCD), i.e., a nested sequence of simplicial subcomplexes. During this process, the “birth” and “death” of different topological features such as connected components and n -dimensional holes, can be tracked down. There exist different ways to create a filtration using different kind of simplicial complexes such as Vietoris-Rips complexes [103], Alpha complexes [38], Lazy-Witness complexes [35], or Čech complexes [34].

Tangential homology is a concept introduced by [29, 25]. It has proved to be a pow-

Introduction

erful tool for shape classification of curve PCD. The idea of tangential homology is to compute persistent homology not to the space itself, but to spaces that are enriched with geometrical information about the shape.

Software packages are used in this work to compute the persistent homology of a PCD including Perseus [72] (based on [66]), and Phom [91].(based on [104]).

This thesis is organized as follows. Chapter 1 shortly introduces ultrasonic non-destructive testing, where the physics behind ultrasounds are also well presented. This chapter also touches upon the current ultrasonic NDT methods and the different welding defects that normally appear. Chapter 2 provides the state of the art of the existing techniques for automatic classification of TOFD defects, where image pre-processing, image segmentation and image classification stages are introduced. In Chapter 3, an introduction to topological data analysis is given. It provides the necessary theoretical background to understand our new approach to the problem. Chapter 4 focuses on Discrete Morse theory, which allows a fast pre-processing step before the computation of persistent homology. Chapter 5 describes our new approach to the problem, where two automated classification methods are presented. Chapter 6 deals with neighbouring search, a constant in many steps of our methods. Chapter 7 will present numerical results where our techniques are tested with real and simulated data. Finally, Chapter 8 will conclude this work.

1 Ultrasonic Non-Destructive Testing

In this chapter, we introduce non-destructive testing of materials. Section 1.1. includes the reasons that explain why people started to use these type of methods, the basic principles of ultrasonic Non-Destructive Testing (NDT) techniques and their advantages and disadvantages. In section 1.2., the physics of ultrasound theory are presented, describing the most significant physical processes involved in ultrasonic NDT. An explanation of how to create and measure an ultrasound is included in section 1.3. Section 1.4. introduces the main methods for ultrasonic testing. To conclude this chapter, in section 1.5. we show and explain the typical welding defects that may occur. For a more comprehensive account on the topic, we refer to [55, 16, 59].

1.1 First steps

Ultrasonic testing usually employs high frequency waves, i.e, frequencies between 0.5 and 10 Mhz [55], in order to carry measurements in a very wide range of fields, such as engineering (flaw detection [4], material characterization [56], etc) or medical applications (sonography, cancer detection [15], etc.).

Ultrasonic NDT methods can be classified according to the type of waves that are used for gathering information: *i) pulse-echo* techniques, which capture information given by reflected waves and *ii) through transmission* techniques, which use transmitted waves. Normally pulse echo systems are more frequently used since they only require one side access to the inspected object or material.

1.1.1 Why ultrasonic NDT

Ultrasonic testing methods are relatively new. The idea was proposed by Sokolov in 1935 [90]. Before that, the employed techniques for detection of internal defects were just x-ray or radioactive isotopes, and by this time, x-ray equipment was big and expensive and the testing process was dangerous and slow. Moreover, only thin material could be tested. After the Second World War, equipment was developed to perform ultrasonic testing measurements. The key of ultrasonic testing is that sound waves are not only reflected or diffracted by the interfaces, but also by the internal defects.

1.1.2 Basic principles of ultrasonic NDT

An ultrasonic measurement is arranged as follows (Figure 1.1):

- (i) Transducers acting as emitter or receivers have to be placed into the inspected area. Depending on the method, one can place two devices acting as emitter and

receiver or just one device acting as both. These transducers are normally piezo-elements, which means that they are able to convert an excited voltage into an ultrasound and viceversa. The piezoelectric effect will be addressed later in this chapter.

- (ii) The ultrasound travels through the material.
- (iii) When the wave finds a flaw, some wave energy will be reflected to the receiver.
- (iv) The receiver converts that energy into electricity again and that electrical signal is displayed on a screen.
- (v) The position, size, orientation and other properties of the defect can be found out from the measured signals.

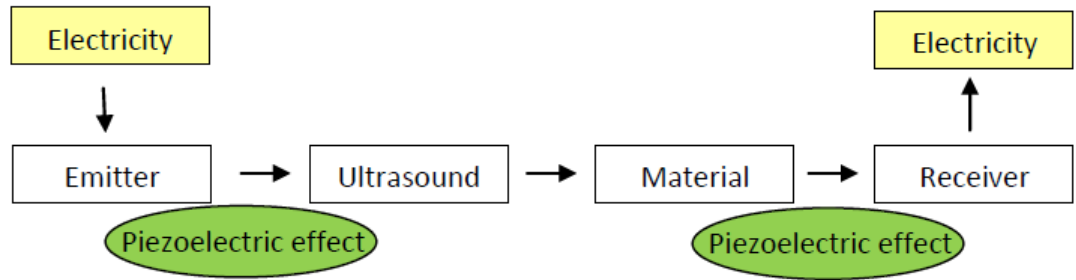


Figure 1.1: typical arrangement for ultrasonic testing.

The most interesting properties of an acoustic wave propagated in isotropic material are the following (Figure 1.2):

- Amplitude A : it is the height from the equilibrium point to the highest point of a crest. It measures how "big" the wave is.
- Frequency f : it gives the number of oscillations of a particle per second.
- Wavelength λ : it is the distance between two points in the wave with the same phase.
- Speed of the wave v : it depends on the medium.

Another important remark is that the smaller the wavelength is, the higher is the interaction between sound waves and the material.

Let λ be the wavelength in mm , v the sound velocity in km/s and f the frequency of the wave in Mhz . They are related as follows:

$$\lambda = \frac{v}{f}.$$

To get small wavelengths, one needs to use high frequencies. If the frequency is not high enough, the defect detection is dubious.

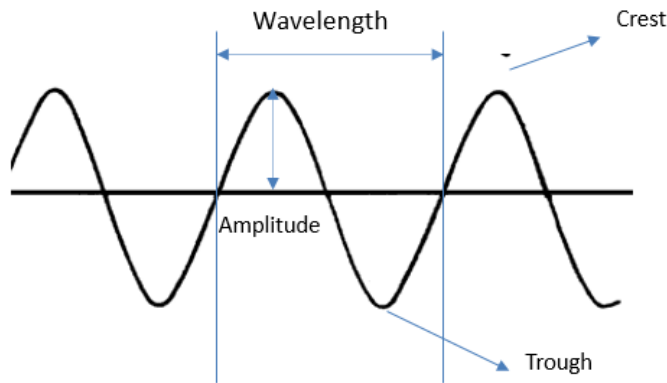


Figure 1.2: Basic elements of a wave.

1.1.3 Advantages and disadvantages of ultrasonic NDT

Here we list advantages and disadvantages of ultrasonic NDT methods in comparison with other NDT methods. We summarize them as follows:

Advantages:

- We usually only need access to one side of the material inspected in order to place the transducers.
- The results of the measurements are displayed instantaneously on a screen.
- The equipment is small, cheap and can be automated, i.e., complete system to move it along the inspected material.
- It is not dangerous and operators are safe in any case.
- It is very sensitive to surface or inner defects. Therefore volumetric and crack like defects can be detected, and the detection does not depend on their orientation.

Disadvantages:

- Operators need training especially for the interpretation of ultrasonic data. This is precisely the motivation of our work. There are not so many current automatic interpretation techniques for ultrasonic data.
- The obtained data is quite noisy.
- Inhomogeneous materials are difficult to inspect.
- Spurious indications, and misreading of signals, can result in unnecessary repairs.

1.2 Physics of ultrasound

To understand how ultrasonic testing works, we need to go deeper and comprise all physical processes that are involved in this method. Here we explain the most important ones. For a more detailed overview on these topics the reader may consult [101, 67].

1.2.1 Wave propagation

As we noted before, ultrasonic testing uses ultrasonic waves to gather important information about the existing flaws or indications inside a material. Therefore, it is very important to understand how sound waves propagate in a solid material.

Every body is made of atoms, which can combine and create bigger particles. Due to many reasons, these particles can start vibrating. We are interested in studying the vibration that induces a mechanical wave. A body can be subjected to a force field and considering the case where its elastic limit has not been reached, the body particles start oscillating. Whenever a particle is displaced from its equilibrium position, some internal forces arise in order to restore that equilibrium. At the end, these forces added to the inertia of the particles, lead to oscillatory motions of the medium. The relation between the particles oscillation and propagation of the sound leads us to different ways of wave propagation. Here we mention the most used in ultrasonic testing:

- Longitudinal waves: It is a wave motion in which the particles of the medium oscillate in the same direction of the wave propagation (Figure 1.3). They are caused by compression and expansion forces, creating compression and rarefaction areas. They can be generated in gases, liquids or solids. They are also referred as *compression* waves.

As illustrated in Figure 1.3, in compression areas particles are closer to each other and density of the medium increases. There are many examples in nature where longitudinal waves are generated, i.e., tsumamis, ultrasounds, internal water waves, primary or *P* waves, etc.

- Transverse waves: it is a wave motion in which the particles of the medium oscillate perpendicular to the direction of the wave propagation (Figure 1.4). It is important to note that particles do not move with the wave. They just move up and down from their original equilibrium point as the wave passes. One example of these waves are the secondary or *S* waves.

The velocity of these types of waves in different materials is well known [86]. In particular we are interested in steel, since in this project we are working with steel pipelines. In this medium, longitudinal waves have approximately a speed of $5.94 \frac{mm}{\mu s}$ while the speed of transversal waves is around $3.29 \frac{mm}{\mu s}$. Therefore, longitudinal waves

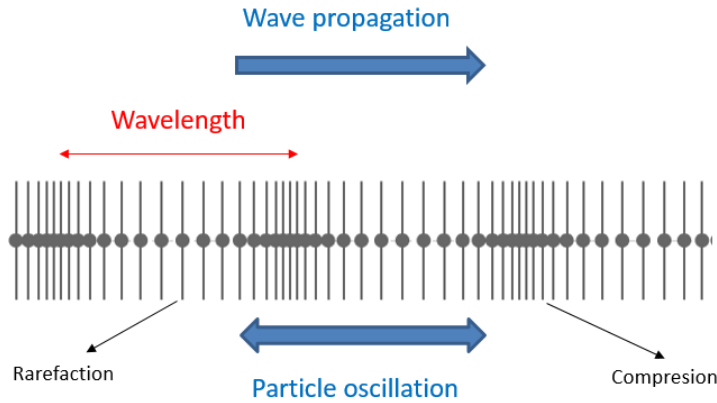


Figure 1.3: Longitudinal wave: wave propagation in oscillation direction.

arrive generally much earlier than transversal waves to the receiver, and in most of the cases, these waves will be the only ones considered and measured.

1.2.2 Huygens principle

Several effects like reflection, refraction or diffraction occur to waves. In order to be able to fully understand them, it is very important to explain the principle that governs them, and this is the *Huygens principle*. It was published by Huygens in 1690 in [50], a work about light waves:

"There is the further consideration in the emanation of these waves, that each particle of matter in which a wave spreads, ought not to communicate its motion only to the next particle which is in the straight line drawn from the luminous point, but that it also imparts some of it necessarily to all the others which touch it and which oppose themselves to its movement. So it arises that around each particle there is made a wave of which that particle is the centre. ..."

Let us try to fully understand what this paragraph is saying. First of all, it is necessary to give some definitions. A *wavefront* shows all the points of a wave that are in phase with one another [59, 5]. This is illustrated in Figure 1.5, where we have a 3D view of a wave and a wavefront is represented by a line.

We can represent all the points with the same wave crest as a straight line (Figure 1.6). Huygens principle basically says that it is possible to predict the shape and position of the next wavefront. We can think on every point on a wavefront as a source of secondary wavelets. Let us apply Huygens principle to the wave of Figure 1.6. Remember that a small vibrating source produces circular waves that radiate outwards. Therefore, if we apply Huygens principle to all the points of a wavefront, we can draw little circles around each point to represent the wavelets formed by each point. The radii of all the circles are the same, showing that all the wavelets move at the same speed, so they move the

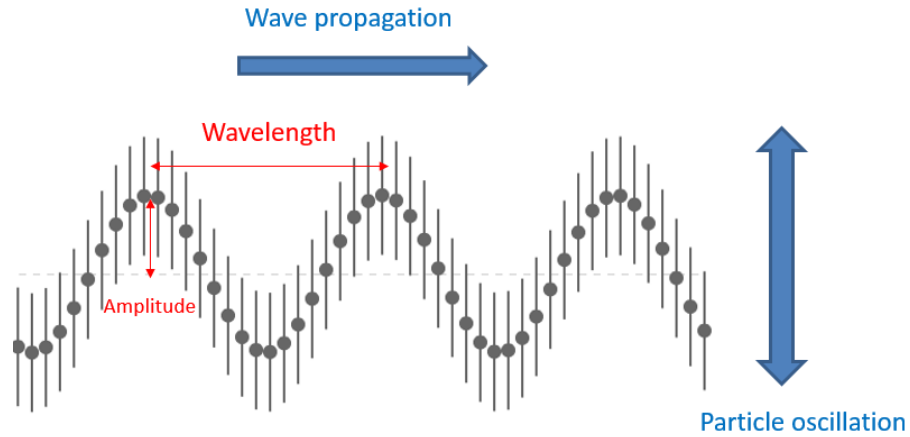


Figure 1.4: Transverse wave: propagation perpendicular to the oscillation.

same distance after some time. Notice that one tangent line can be drawn to all of these circles. The tangent shows the boundary of the region that contains all the little wavelets. Huygens says that the tangent line represents the new wavefront (Figure 1.7).

1.2.3 Diffraction

As we explained, ultrasonic waves are sent through the material. These waves can find obstacles in their way to the receiver such as defects. Due to the Huygens principle, diffraction occurs at the tips of the defect and the wave can travel round it.

Let us use Huygens principle to predict how a wavefront would look like for a wave that finds a barrier in its way (Figure 1.8). Some of the points in the wavefront are blocked by the barrier. This causes the shape of the new wavefront to be curved at the point where the barrier has cut off some of the point sources. The curvature of the wavefront as it passes the barrier is called *diffraction*.

What happens if we add another barrier? (Figure 1.9 left). The wave now passes through a gap. The wavefront is now curved on both sides. The narrower we make the gap (Figure 1.9 right), the more the wave spreads out behind the gap. The more the wave is diffracted, as it passes through the gap. It is important to note that the diffraction pattern formed by the narrow gap is very similar to the pattern formed by a single spherical vibrating source.

1.2.4 Reflection

This phenomenon occurs at the boundaries. Consider two materials M_1 and M_2 one over the other. The wave is reflected at the boundary if both materials have different

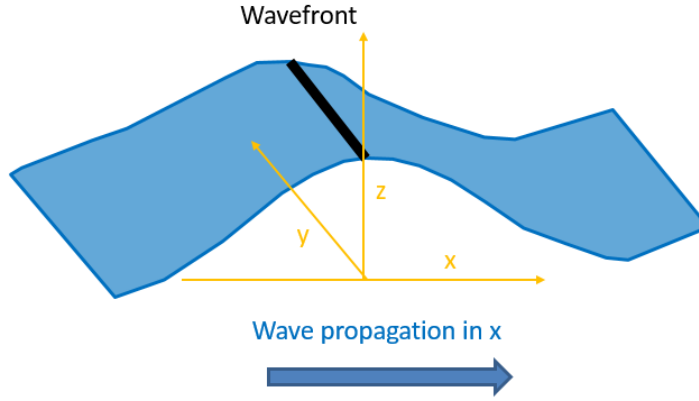


Figure 1.5: Representation of a wave and one wavefront in 3D.

acoustic impedances [22]. The greater this difference is, the greater the percentage of the wave energy that will be reflected. The acoustic impedance Z of a material is defined as follows:

$$Z = \rho v,$$

where ρ is the density of the material and v is the sound velocity. Ultrasonic testing is widely used for welding testing, and in most of cases steel and air are the two materials at each side of the boundary. The proportion of wave energy that is reflected is measured by the so-called *reflection coefficient* R and can be computed as follows [12]:

$$R = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2,$$

where Z_1 and Z_2 are the impedances of the two materials at the interface or boundary. Note that in this whole chapter we are assuming that there is no loss of energy by attenuation, absorption or dispersion. Therefore, when the wave arrives to the interface, the total energy is split into reflected plus transmitted energy, and the *transmission coefficient* T is calculated by

$$T = 1 - R.$$

Due to the Huygens principle, the wave at the boundary builds a new wave front that will be propagated as a reflected wave. In reflection, the ultrasound speed remains the same, and therefore the angle of incidence is equal to the angle of reflection. Reflection of waves is illustrated in Figure 1.10.

1.2.5 Refraction

When a wave arrives to an interface, besides reflection, another phenomenon called refraction takes place. Consider two materials M_1 and M_2 at each side of the interface. Refraction takes place at the interface when the velocity of sound in M_1 is different than

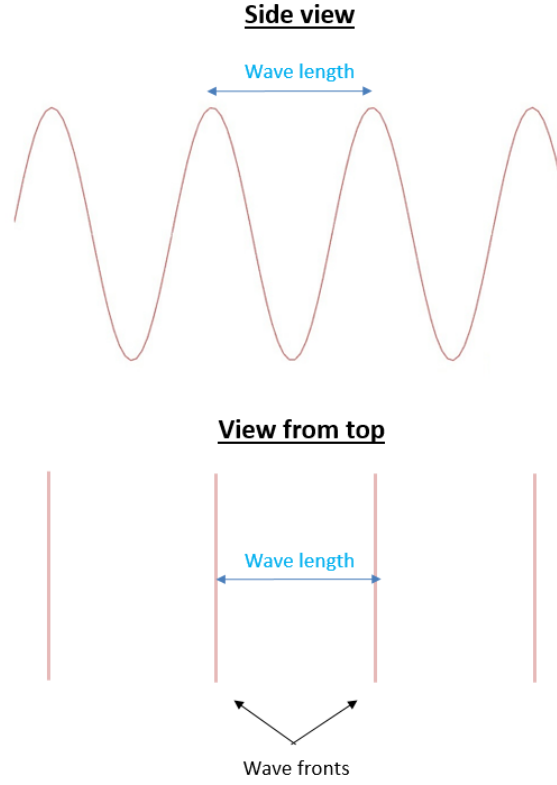


Figure 1.6: Representation of a wavefront from top view of the wave.

in M_2 . Sound velocity is inherent of the material, since it depends only on material properties such that the elastic modulus and density. When the wave "enters" in the second material, there is a change in the direction of the travelling wave.

Refraction is governed by *Snell's law* [17], which describes the relation between angles and velocities of waves. It basically says that the proportion between material velocities v_1 and v_2 is equal to the proportion between sinus of incident angle α_1 and refracted angle α_2 (Figure 1.11):

$$\frac{\sin(\alpha_1)}{v_1} = \frac{\sin(\alpha_2)}{v_2}.$$

Note that v_1 and v_2 are the velocities that correspond to the longitudinal waves in those materials.

When the wave passes from a material with some velocity to another material where its velocity is higher, there is a certain angle of incidence that causes a refraction angle of 90° . This is known as critical angle. To find it, one needs to substitute $\alpha_2 = \pi/2$ and calculate α_1 . When this happens, most of the acoustic energy is in the form of an inhomogeneous compression wave, and this wave is known as "creep wave". It decays pretty fast along the interface, and because of this, they are barely used in NDT.

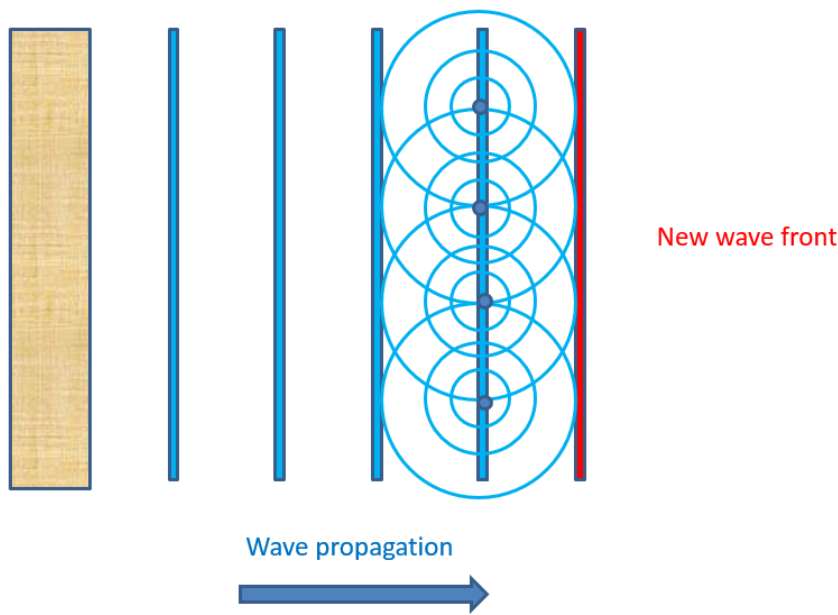


Figure 1.7: It is possible to predict the position and shape of the new wavefront according to the Huygens principle.

1.3 Equipment and transducers: the piezoelectric effect

Before getting to know what kind ultrasonic testing methods exist, it is important to describe how to create an ultrasound. This would not be possible without the so-called *piezoelectric effect*. Here we give a brief introduction to it. For a more comprehensive account on the topic, we refer to [55, 59, 51].

Ultrasonic testing is based on the conversion of electricity into mechanical vibrations and the reverse process. There are some materials called piezo-elements, which can make possible this process converting electrical energy into acoustic energy and viceversa. The role of the piezoelectric effect in NDT is the following:

- An electric pulse or voltage is created.
- This electricity passes through a piezo-element, which is placed in the heart of the emitter. Due to the piezoelectric effect, this device converts the electrical energy into an ultrasound that travels through the material.
- When the wave arrives to the receiver, another piezo-element converts back the energy into a voltage that can be measured and related to wave properties.

A piezo-element is normally a piece of a polarized material (which means that some parts of the molecule have a positive charge, while other have negative charge) with electrodes attached to two of its opposite faces. If we apply an electric field across the material, the polarized molecules will start to align with the electric field and therefore

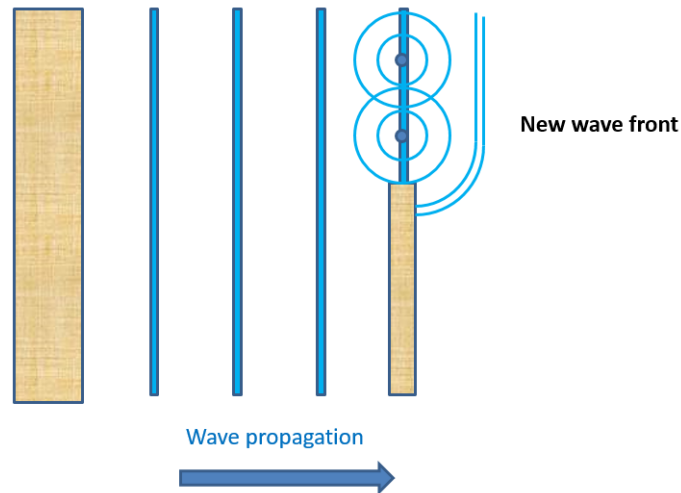


Figure 1.8: Diffraction of waves. It is possible to predict the new wave front using the Huygens principle.

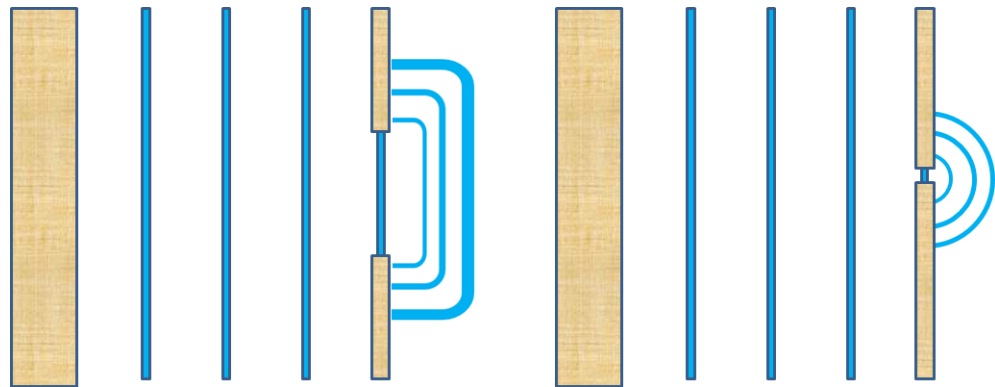


Figure 1.9: Diffraction of a wave through a gap.

dipoles are induced within the molecular or crystal structure of the material. As a consequence, the dimensions of the material change and a mechanical wave is created. This phenomenon is known as the *reverse piezoelectric effect* [46].

Some materials like barium titanate create electricity when the material changes in size due to an imposed mechanical force. This is the *direct piezoelectric effect*. A list with different materials that are piezoelectric can be found in [52] and the phenomenon is illustrated in Figure 1.12.

1.4 Ultrasonic NDT methods

In this section, we introduce different available methods. Each of those use the explained effects of reflection, refraction and diffraction to gather interesting information of the

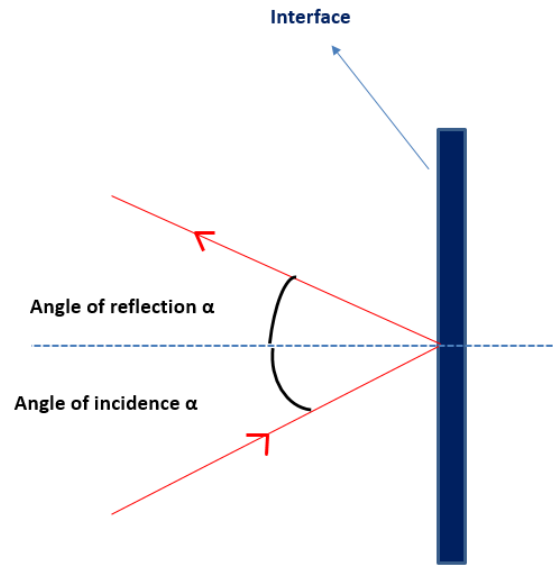


Figure 1.10: Reflection of waves.

tested materials. It is possible to classify these methods precisely according to the incoming ultrasound [46]:

- (i) Impulse-echo techniques: the measured waves will be reflections or diffractions of the inner defects. Therefore, one needs to place the transducers in positions such that this will happen.

In Figure 1.13 we show one example where one transducer is acting as emitter and receiver. In case a), the reflected waves at the defect do not reach the receiver, but in case b), after moving a bit the transducer until it is directly above of the inner defect, some information can be measured at the receiver. There are many different ways of placing the probes. The Time of Flight Diffraction method (TOFD), that will be presented later, place the probes in a way to measure diagonal reflection and diffraction waves.

- (ii) Transmission techniques: now the receiver is situated opposite to the emitter at the other side of the material (Figure 1.14). When there is no defect in between, the ultrasonic wave will always reach the receiver (even if it is in an attenuated version). When there is a defect in between, some alterations will be shown in the measured data.

Impulse-echo methods are normally more employed than transmission techniques. For obvious reasons, the last need access to both sides of the inspected material, and hence this is more restrictive. A second reason is that in impulse-echo techniques the received

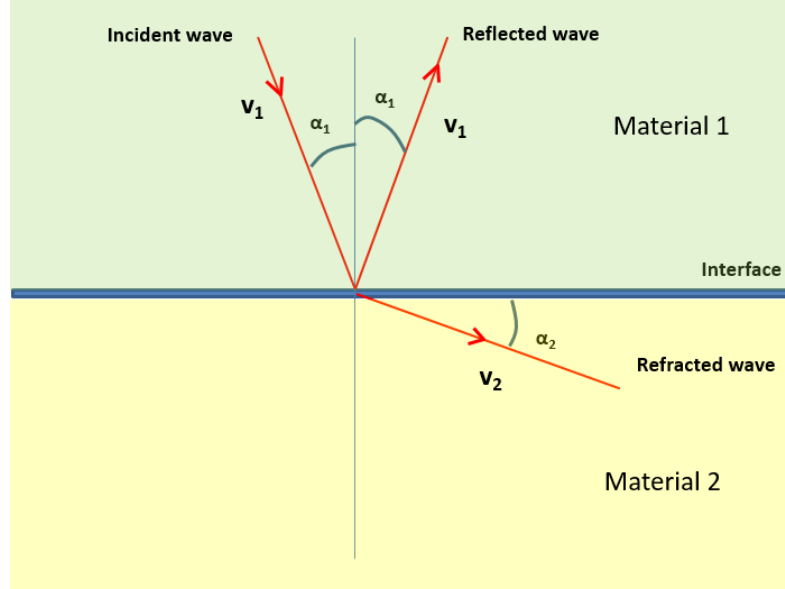


Figure 1.11: Schematic view of refraction and Snell's law. The incident wave is refracted and reflected at the interface.

waves contain boundary defect information, and moreover one can easily find out the distance defect-probes because it is directly related to the ultrasound velocity and the time of flight that the wave spends from emitter to receiver. In this research project we focus on this techniques and in the last part of this chapter we present two important methods: the Time of Flight Diffraction (TOFD) [68, 93, 80], which is worldwide employed for weld defect detection and the Wall Thickness measurement [55].

1.4.1 The Time of Flight Diffraction Method (TOFD)

The TOFD was first introduced in 1977 [89], and since then, it has been one of the most reliable methods in non invasive condition assessment. The main difference of TOFD with respect to all other ultrasonic NDT techniques is that it relies on the detection of diffracted rather than reflected signals. A typical set-up for the TOFD method is shown in Figure 1.15. Two angle beam probes are used in a transmitter-receiver arrangement, where the distance between them is normally calculated according to the wall thickness. Ultrasonic waves are sent through the inspected material. The received information at the receiver is recorded in an *a-scan* (Figure 1.16), which is a representation of the amplitudes of the waves versus time. An *a-scan* normally monitors:

- Diffracted signals at the edges of the defects.
- Lateral wave, which is the fastest wave that travels directly from emitter to receiver.
- Back wall echo, which is the reflection of the back wall and the latest wave arriving to the receiver.

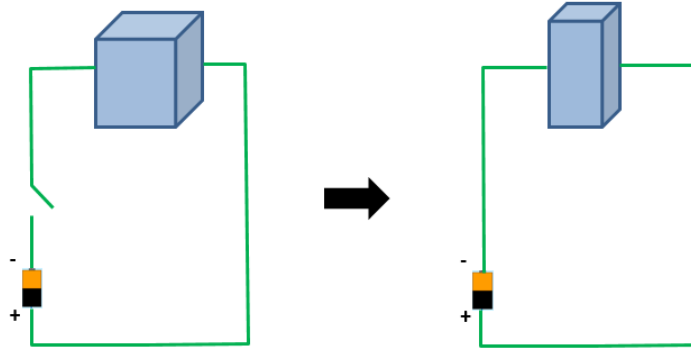


Figure 1.12: Piezoelectric effect. When a voltage is applied across a piezo-element, it changes in size and viceversa.

- Diffractions from inhomogeneities of the material (noise).

For acquiring more data, probes are moved in parallel to the weld. Whenever they are fixed in a new position, another a-scan is taken. All the combinations of a-scans create a *b-scan* (Figure 1.16), which is a representation of the position of probes versus time. The echo amplitude is normally displayed as gray scale, with light gray for zero amplitude and black and white for negative and positive maximum amplitude respectively.

The two classification methods proposed in this paper are based on the detection and recognition of the shapes that TOFD indications may present in b-scans. We focus our work in the following indications:

- *Volumetric defects*: pore defects or gas inclusions, which usually exhibit a parabolic shape in b-scans (Figure 1.16).
- *Planar defects*: lack of fusion and crack propagations, which usually produce linear shapes in b-scans (Figure 7.3).

TOFD b-scans are usually quite noisy. Before applying any classification method, a preprocessing step is applied for denoising. One typical deconvolved b-scan is shown in Figure 1.16.

1.4.2 Wall thickness measurement

Usually surface defects are really easy to detect even with the human eye, and most of the times there is no need of a non-destructive testing evaluation. But there can be some situations where the material surface is not visible. For instance one can consider a pipeline that is half buried. Then one needs an NDT evaluation for checking the state of the buried surface.

This is the purpose of wall thickness measurements. A probe, acting as emitter and receiver at the same time, is placed on the accessible surface (Figure 1.17a), and ultra-sounds are sent to the interior into the direction of the inner surface defect. Once the

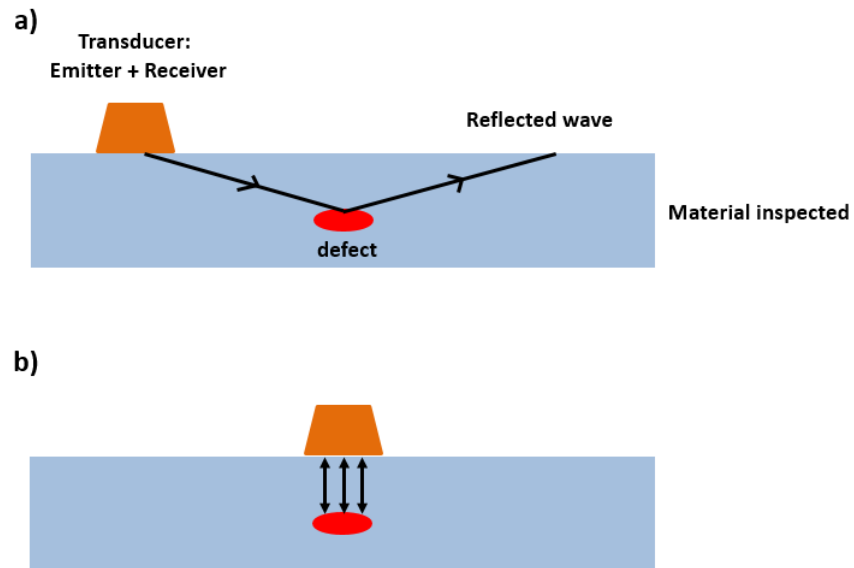


Figure 1.13: a) The reflected waves from the defect do not reach the receiver; b) In this position of the transducer the reflected waves can be measured at the receiver.

wave arrives to the defect itself, it is reflected back reaching the transducer. As in every pulse-echo method, there is a correlation between the time of flight and the distance or wall thickness. Accordingly, it is possible to reconstruct the shape of the inner surface. Note that in Figure 1.17a we just show a 2D view, but of course the defect is 3D and one needs to move the probe across the whole outer surface to reconstruct the inner surface. This leads to a three-dimensional b-scan (Figure 1.17b), where each column corresponds to a probe position on a line on the surface.

Let us analyse why a wall thickness b-scan looks like that. To that end, we follow the journey of the waves in Figure 1.17a. The trip starts at the transducer. From there, the wave or parts of it can take several paths.

- Part of it does not even enter the material and it is reflected back to the transducer (point *D*). That is why one sees a region in the b-scan that corresponds to this reflection of the surface.
- Part of the ultrasound reaches the defect and it is reflected to the transducer back.
- There are many backwall echo reflections. Consider for example the ultrasound travelling from *A* to *B*. Then it is reflected back and when it arrives again to *A*, some part may continue to the transducer and other is again reflected and starts travelling in direction to the inner surface again. This is why in the b-scan one can see several back wall echoes with a decreasing amplitude.

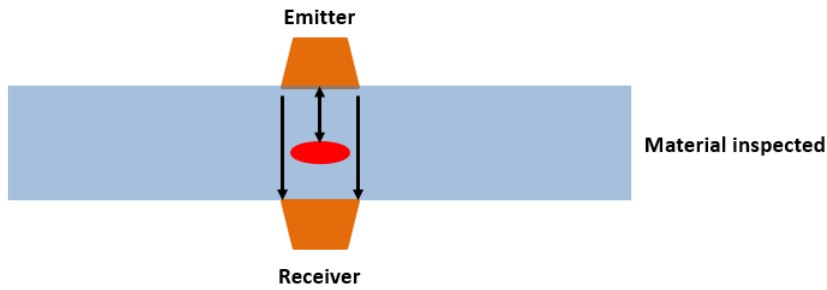


Figure 1.14: Transmission technique.

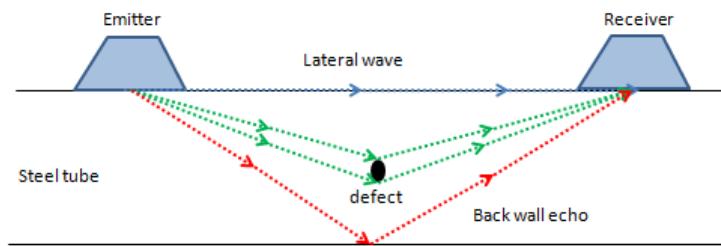


Figure 1.15: Time of Flight Diffraction method (TOFD).

1.5 Welding defects

In this section we describe the most common welding defects and the principal mechanisms that cause them. The coordinate system used in the text is chosen as illustrated in Figure 1.18, where a pipeline is welded along the z -axis.

Welding imperfections are classified according to ISO 6520 [3] while their acceptable limits are specified in ISO 5817 [2] and ISO 10042 [1]. The most observed defects can be classified as follows:

- 2.5.1. Surface defects:
 - 2.5.1.1. Lack of deposition
 - 2.5.1.2. Lack of penetration
 - 2.5.1.3. Over deposition
 - 2.5.1.4. Undercut
 - 2.5.1.5. Spatter
 - 2.5.1.6. Surface cracks
- 2.5.2. Internal defects:
 - 2.5.2.1. Internal cracks

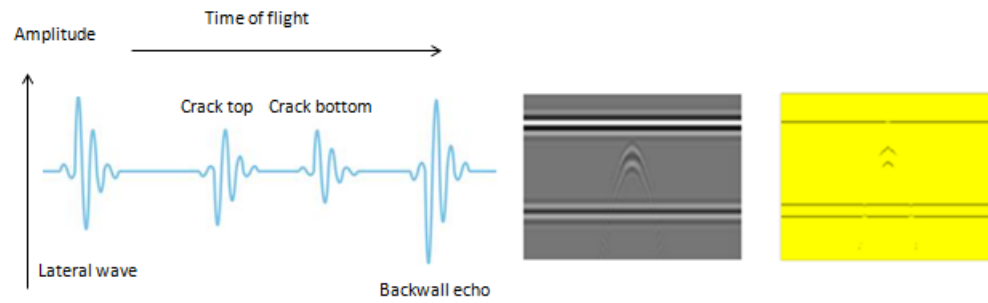


Figure 1.16: Left to right: a-scan, b-scan, deconvolved b-scan (preprocessed data).

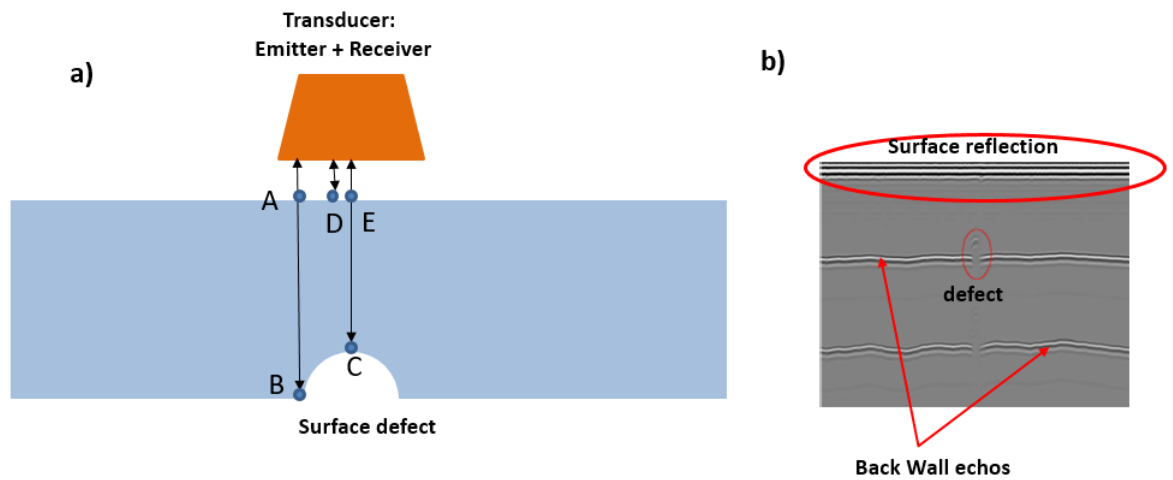


Figure 1.17: a) Wall thickness measurement; b) Typical corresponding b-scan.

2.5.2.2. Slag inclusion

2.5.2.3. Porosity

2.5.2.4. Wormholes

2.5.2.5. Lamellar tear

The following subsections include a brief description of every welding defect.

1.5.1 Lack of deposition

This kind of defect is generally not acceptable. If we consider a cross section xy of the weld, this should have a top reinforcement as shown in Figure 1.19a top. When this is not the case, the weld presents lack of deposition (Figure 1.19a bottom and b).

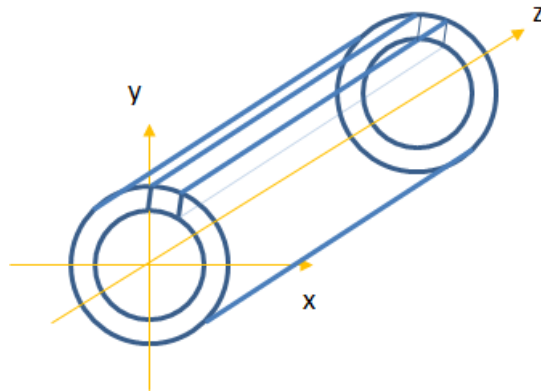


Figure 1.18: Representation of a pipeline and the chosen coordinate system where the tube is welded along the z axis.

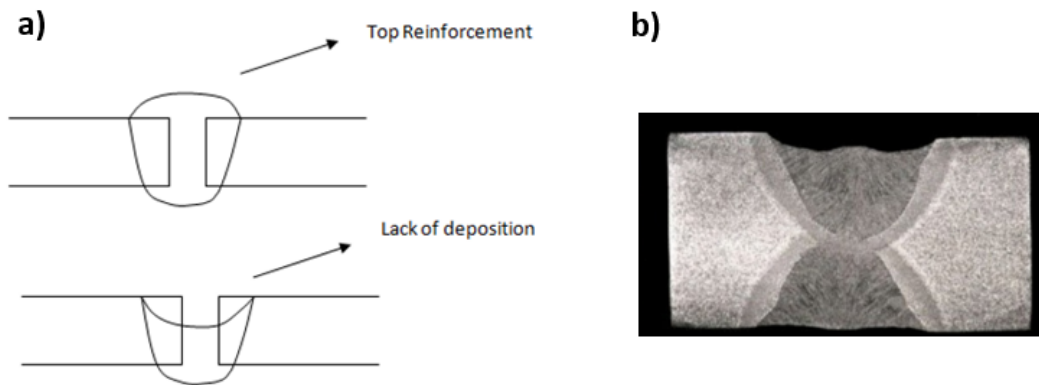


Figure 1.19: a) Ideal cross section xy representing top reinforcement and lack of deposition; b) Real image showing lack of deposition [75].

1.5.2 Lack of penetration

It is illustrated in Figure 1.20 and it can be found in two ways: i) when the weld bead does not penetrate the entire thickness of the base plate or ii) when two opposing weld beads do not interpenetrate. It is normally caused due to a low welding current and can be fixed easily by simply increasing the amperage.

1.5.3 Over deposition

It happens when too much heat is added to the system (Figure 1.21a). This increases the heat affected zone (HAF), which is the area of base material that is not melted and has had its microstructure and properties altered by welding or heat intensive cutting operations. Therefore, this is an indirect defect that might affect the pipeline in the long

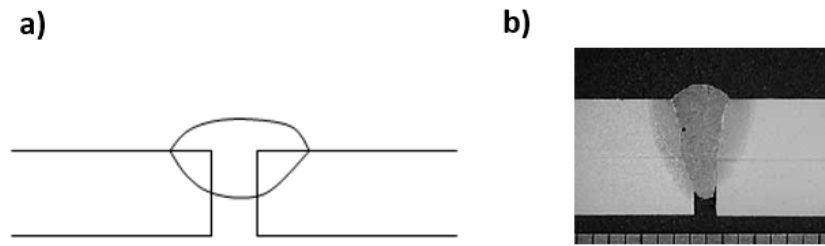


Figure 1.20: a) Cross section xy showing lack of penetration; b) Real image showing lack of penetration [87].

term, increasing the corrosion rate. This also creates local stress concentration.

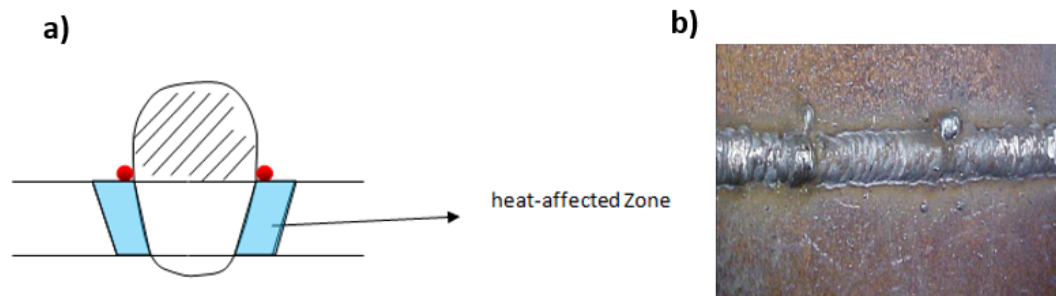


Figure 1.21: a) Cross section xy showing over deposition. The HAZ is represented by the blue area while the local stress concentration by the red circles; b) Real image showing over deposition [84].

1.5.4 Undercut

If the welding speed is too high, the metal will get melted and solidified before it can flow back and fill the whole place (Figure 1.22). Undercut also takes place when a poor technique is employed and not enough filler metal is deposited along the edges of the weld. Other causes are a dampened electrode or excessive arc length [82]. To prevent undercut, using a multi-run technique is one of the best solutions. It helps to produce fine welds with improved mechanical properties.

1.5.5 Spatter

Spatter are zones where there are small droplets of metal (Figure 1.23a and b). It is not a defect itself, but it indicates that might be a flaw in its location. An explosion of metal has taken place for some reason. It is possible that gas has got entrapped in the flaw.

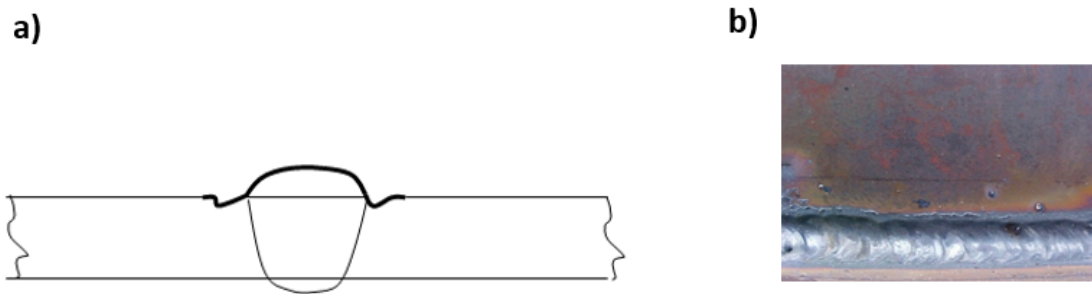


Figure 1.22: a) Cross section xy showing undercut; b) Real image showing undercut [92].

Spatter indicates a possibility of porosity inside and therefore needs to be removed. It is possible to do so by grinding, and it is compulsory, otherwise corrosion will start soon when painting.

1.5.6 Surface cracks

They appear as linear openings at the metal surface (Figure 1.24a and b). They can be critical and have led to frequent failures. They are generally known as "hot cracks", since they take place during the welding process itself when the weld is still hot. Cracking normally results from poor welding practice, inadequate joint preparation, improper electrodes, inadequate preheat, and an excessive cooling rate. All this leads to unfavorable and residual thermal stresses, which causes the crack. The remedy is cutting out and rewelding.

By residual stresses, we mean stresses that remain in the material after having removed the original cause of the stresses. These can cause a premature failure of a structure and in most cases are highly undesirable. There are several physic mechanisms that can cause residual stresses [85]: inelastic (plastic) deformations, temperature gradients (during thermal cycle) and structural changes (phase transformation). For instance, when the welding process is finished, after some time the weld starts cooling, but this cooling process can be different between different areas, and then residual stresses are created.

1.5.7 Internal cracks

They are generally known as cold or delayed cracks. You may find them even even few days after the welding.

At the beginning, they are generated internally and then they can propagate outside. Internal cracks generally happen due to a phenomenon called "*hydrogen embrittlement*". Basically, the material loses ductility becoming brittle, and under thermal stresses cracks [61, 102]. When one takes a look into the microstructure of the material (Figure 1.25b), it can be appreciated that there exist grains and voids.

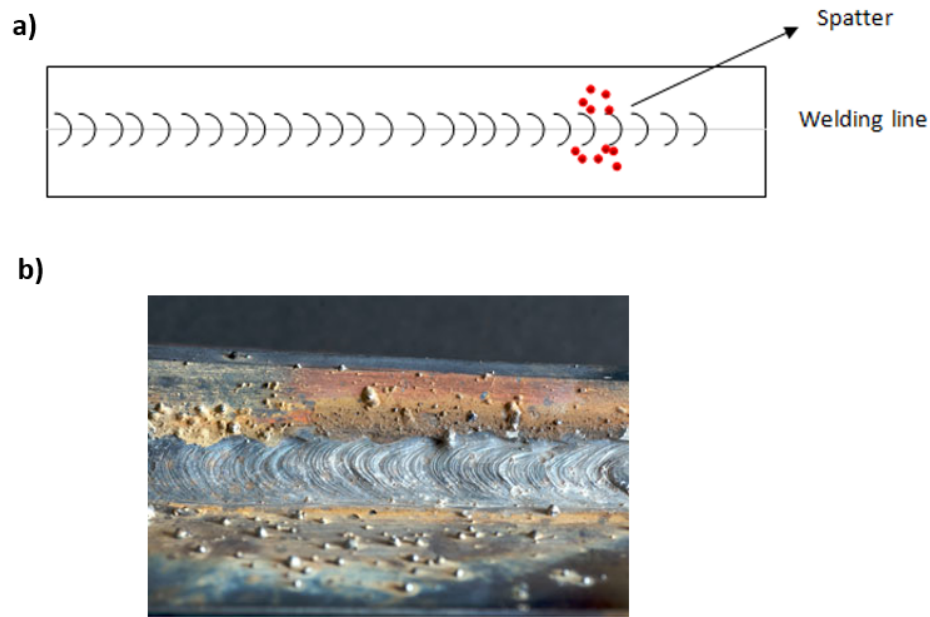


Figure 1.23: a) Representation of spatter from a top view. The red dots indicate small droplets of metal; b) Real image showing spatter [7]

During the welding process there exits hydrogen pressure and if this process is not done correctly, the hydrogen gets dissociated and its atomic form occupies the voids (Figure 1.25b left). When the welding is over and the whole material gets cooled down, these atoms combine and form some molecules which size is much bigger than the sum of the size of the individual atoms. Therefore, the lack of space for the generated molecules, causes an enormous force in every direction and the material cracks (Figure 1.25b right).

In order to avoid hydrogen embrittlement, several techniques can be employed and all of them are focused on minimizing contact between the metal and hydrogen, particularly during fabrication. One solution is to apply sulfides to the metal [95].

1.5.8 Slag inclusions

This defect usually occurs in welds that need multiple passes and there is a poor overlap between them (Figure 1.26a and b). As a consequence the slag from the previous weld cannot melt out and rises to the top of the new weld bead. Slag inclusions can have any arbitrary shape and with all probability, will have many sharp peaks.

Because of this, it will behave like crack tips, meaning chances of stress concentration and a potential location for a crack initiator. The slag that appears between passes or runs must be cleaned via grinding, wire brushing, or chipping.

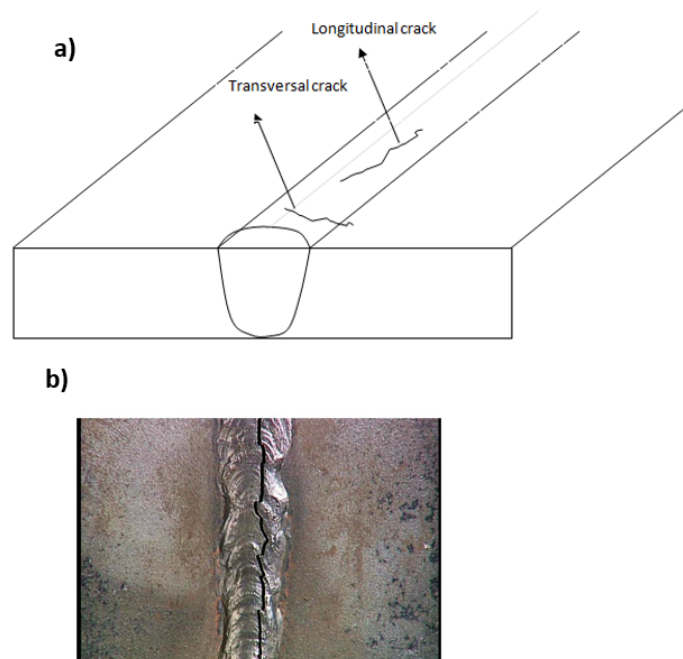


Figure 1.24: a) Representation of the two types of surface cracks (transversal and longitudinal) along the weld; b) Image of a real surface crack [14].

1.5.9 Porosity

Pore defects are gas inclusions inside the welded area (Figure 1.27). During the welding process, some gases are injected. If the metal solidifies too fast, the gas bubbles get entrapped before they can float out. Porosity can be caused by high welding speed, presence of impurities in the welding side, high sulphur content in the electrode or wrong welding current or polarity [81]. Porosity is an intolerable defect and therefore has to be removed. One solution is to use electrode formulation [60].

1.5.10 Wormholes

This volume errors only occur within the seam and have a tube-like geometry. This defect is also not tolerated and has to be removed. They emerge at one point and then grow along the solidification direction upward (chimney effect). They can even reach the surface and that is why they are called wormholes (Figure 1.28). The diameter of the tube is usually kept constant.

To prevent wormholes, one should eliminate the gas and cavities. To this end, one needs: to clean the workpiece surfaces around the weld, to remove any residue such as oil or grease after every NDT operation and to eliminate the surface coatings from the joint area to expose bright material.

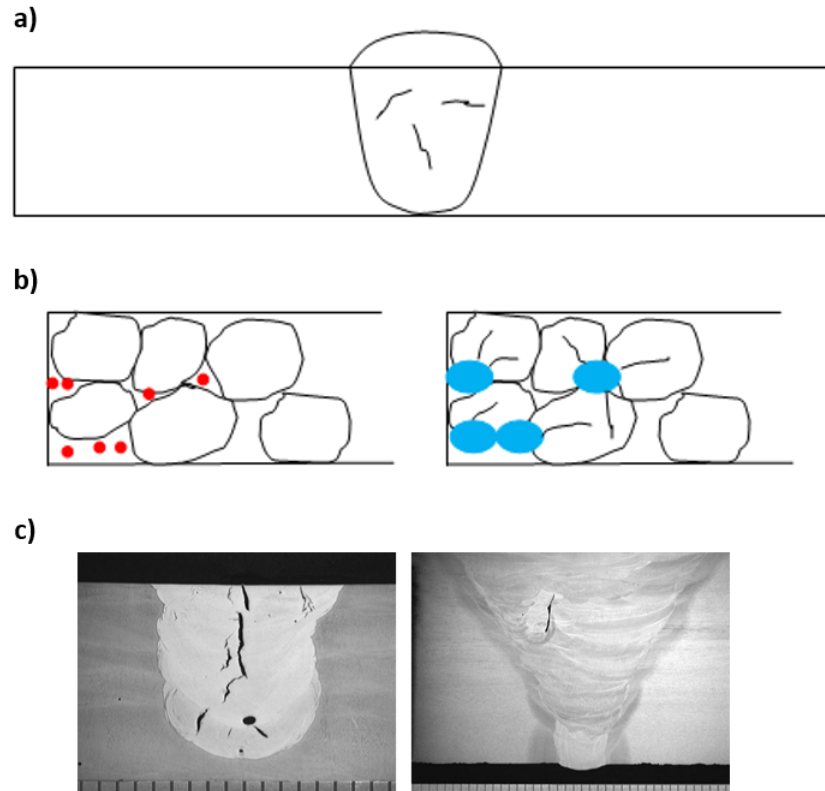


Figure 1.25: a) Representation of internal cracks in a weld; b) Schematic view of hydrogen embrittlement process; c) Real image showing internal cracks [30].

1.5.11 Lamellar tear

This defect appears in T-joint and fillet welds. The crack is always related to points of high stress concentration and they appear parallel to the plate surface (Figure 1.29). Several conditions must be satisfied in order to observe a lamellar tear:

- There exists a shrinkage force acting through the plate thickness.
- Cracks are perpendicular to the direction of the shrinkage force.
- The fusion boundary is almost parallel to the plane of the defects.
- The plate normally has poor ductility in the through-thickness direction.

Changing the construction process to use casted or forged parts can solve this problem, since these defects only arise in welded parts.



Figure 1.26: a) Cross section xy representing slag inclusions in red; b) Real image showing a slag inclusion [75].

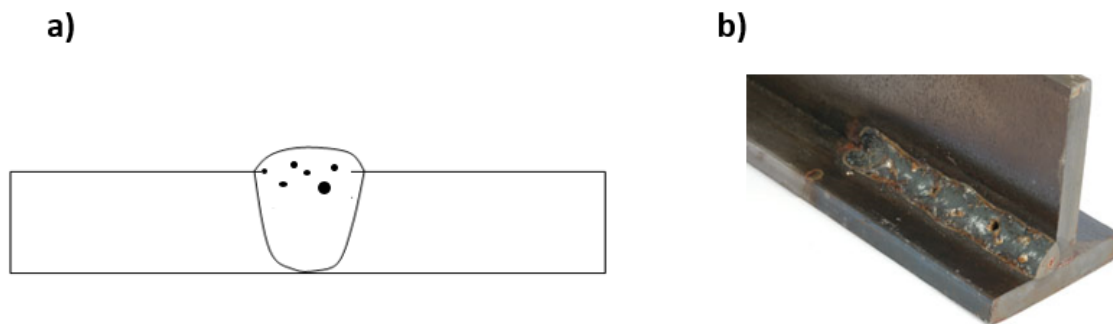


Figure 1.27: a) Cross section xy representing porosity. b) Real image: round holes in the weld bead are a sign of metal porosity [43].

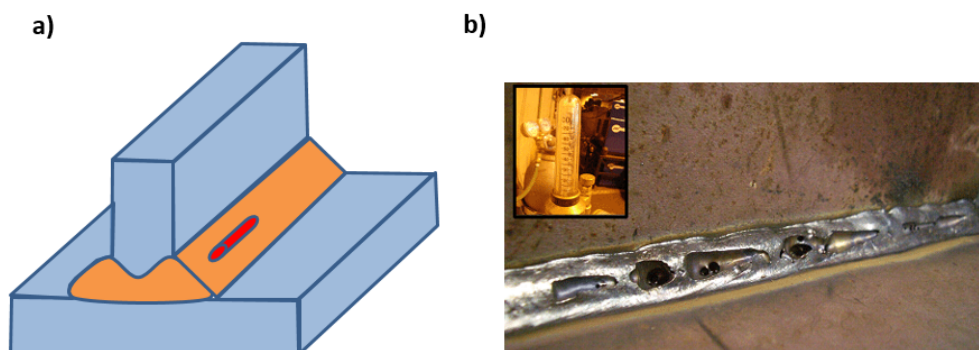


Figure 1.28: a) Representation of a wormhole in a T-joint weld (shown in red); b) Real image showing wormholes [31].

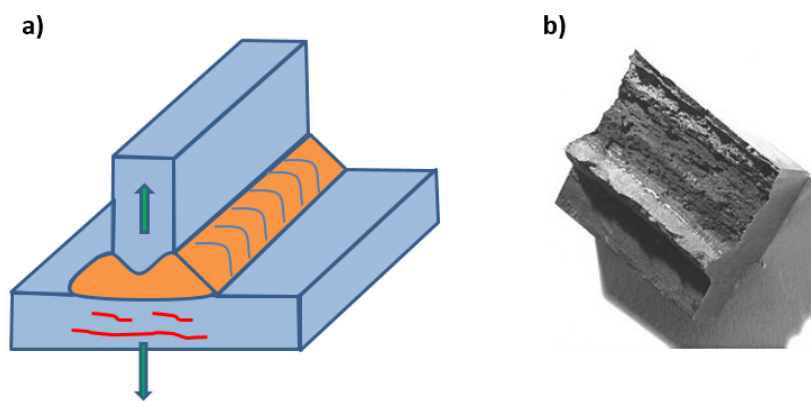


Figure 1.29: a) Representation of a lamellar tear (red) in a T-joint weld; b) Real image of an appearance of fracture face of lamellar tear [47].

2 Automatic classification of TOFD defects: state of the art

TOFD is one of the most popular ultrasonic NDT methods. However, it also has disadvantages, as is the lack of techniques to automate the interpretation of TOFD images and the classification of defects, which are still done manually.

Before we present our classification methods, we want to give an overview of the existing current techniques that have been used to overcome this last disadvantage. Even though this research field is very open and there is still a lot to be done, there have been some remarkable attempts and it is worthy to mention them.

The goal is to interpret TOFD images without the help of a human operator. The general procedure for automatic interpretation of TOFD images is normally divided into four main stages, as shown in Figure 2.1.

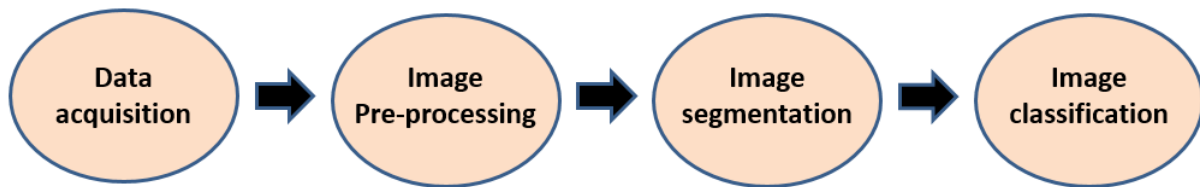


Figure 2.1: Procedure's scheme for the automatic interpretation and classification of TOFD defects.

After the data is obtained using TOFD equipment, the next steps are preprocessing, image segmentation and image classification. In the rest of this chapter each of this steps is presented and some of the current techniques are shown to the reader. Please note that here we only give a brief introduction to them.

2.1 Image pre-processing stage

After a TOFD measurement has been taken (as explained before in section 1.4.1), a raw image is obtained. This is represented by a b-scan (Figure 1.16 center). To perform an image segmentation in a later step, an image pre-processing has to be implemented. This stage comprises two processes:

- Removing everything from the raw image that does not belong to the region of interest (ROI).
- Denoising or background removal.

In section 1.4.1, we noted that the first wave that arrives to the receiver is the lateral wave and the latest is the backwall echo. The ROI is the region that exists between these two signals (Figure 2.2), and therefore, the rest of the image is not further processed. It is also important to remove these two strong signals as well (lateral wave and back wall echo), in order to reduce the amount of data analyzed in the next steps.

TOFD b-scans are usually quite noisy and denoising is needed in the remaining image of the ROI. One can use linear processing methods [99] for this purpose. In the present research project, a sparse deconvolution method [19, 20] is used, where an Orthogonal Matching pursuit (OMP) algorithm adjusted to NDT is applied. This method eliminates the low amplitudes in every a-scan, removing an important amount of noise and it is explained in detail in the following subsection. After applying this method, we obtain a deconvolved b-scan (Figure 1.16 right), where the amount of data has been reduced substantially while retaining the fundamental image, in this case, the defect shape. Processes involved in the preprocessing stage are illustrated in Figure 2.2.

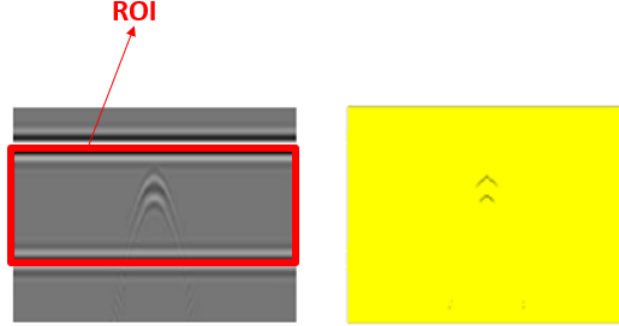


Figure 2.2: Left: Selection of the region of interest; right: deconvolved b-scans after using a sparse deconvolution method. Some noise still remains.

Finally, the basics of the sparse deconvolution method used in our work for preprocessing are explained in the next subsection:

2.1.1 Sparse deconvolution methods for NDT

In the TOFD method, the receiver measures the amplitudes of the arriving waves. These signals come either from defects or noise and the measured amplitude is a function of the form [20]:

$$m(t, p) = \left(g(t) * \left(\sum_{k=1}^K A_k(t, p) + N(t, p) \right) \right)(t),$$

where:

- t is time and p is a certain position of the probes.

- $m(t, p)$ is the measured amplitude and it is represented in a b-scan.
- K is the total number of defects.
- g is an ultrasound impulse function (i.e., Gabor impulse, etc).
- A is the amplitude function, which takes into account several physical waves phenomena such as divergence, absorption or dispersion.
- N is the noise that can come from reflections of small inhomogeneities of the material.

Nowadays, due to the high quality manufacturing processes in steel industry, it is reasonable to assume that only few defects are present. Accordingly, $\sum_{k=1}^K A(t, p)$ is a sparse signal, while the noise $N(t, p)$ is not (many values with low amplitude). As already explained, high amplitude values come from strong signals and the Noise $N(t, p)$ will influence with a minor role the function m . Due to the sparsity shown of the main signals, sparse deconvolution methods can be used. These try to find a sparse solution X of the problem:

$$m = g * X.$$

The solution X will be a good approximation of the solution m and at the same time a large amount of noise will be filtered out. We assume that an upper bound $L \in \mathbb{R}$ of the non zero elements of X is given and the impulse function g to be known. In many applications, this is not the case. For a more comprehensive account, we refer the reader to [20], where techniques are discussed in order to solve a blind deconvolution problem where g is not known.

What kind of sparse deconvolution methods are suitable for our problem? In ultrasonic NDT, there are two main requirements that must be satisfied [19]:

- (i) Stability respect to noise.
- (ii) Computational efficiency.

The so-called *Greedy methods* satisfy both conditions. These techniques try to find a global solution of a problem by solving several local problems and then combining them. One of the handicaps of these methods is that the solution obtained is approximate and not exact, although it is very fast and stable. In our work, we have used the adjustments of a Matching Pursuit (MP) and an Orthogonal Matching Pursuit (OMP) to NDT. In the following paragraphs, the basics of these Greedy methods considered in this work are explained. For a more detailed explanation the reader is referred to [20].

Let f be a signal in a Hilbert space H and $D = \{g_1, \dots, g_P\}$ a dictionary. The idea of the MP method is to represent this signal as a weighted sum of functions g_{γ_n} taken from D where $\gamma_n \neq \gamma_k \forall n \neq k$:

$$f(t) = \sum_{n=0}^L x_n g_{\gamma_n}(t),$$

2 Automatic classification of TOFD defects: state of the art

where the functions g are normally called atoms, n is the number of atoms chosen and x_n refers to the amplitudes of each atom.

Having chosen a dictionary D , MP finds first the atom x_n that has the largest inner product with the signal $f(t)$. Next, the residual r_n is calculated subtracting the contribution due to that atom to the signal. Finally, the process is repeated until the residual is small enough and the signal has been decomposed adequately. We can schematise the MP algorithm as follows:

Given a signal $f(t)$ and a dictionary D :

- (i) Set $r_1 = f(t)$ and $n=1$. Repeat 2 and 3 until some stopping criterion is satisfied, i.e., $\|r_n\| < \text{threshold}$.
- (ii) Find $g_{\gamma_n} \in D$ with the biggest inner product $|\langle r_n, g_{\gamma_n} \rangle|$ i.e., the best match with the residual.
- (iii) $x_n = \langle r_n, g_{\gamma_n} \rangle$, $r_{n+1} = r_n - x_n g_{\gamma_n}$, $n = n + 1$.

OMP gives better results but it is computationally less efficient. The difference with MP, is that OMP takes into account the correlations between active dictionary elements [94].

It has been explained how MP works in general, but how is it applied to ultrasonic NDT? The signal $f(t)$ is an a-scan in our case. Therefore, in order to denoise or deconvolve a whole b-scan one can repeat the process for every a-scan. We assume that our ultrasound data $m = g * X$ is sampled on a time grid $1 \leq t_1 < \dots < t_N < \infty$ and g is known. Therefore $(m(t_k))_{k=1}^N$ is given. We assume m is of the form:

$$m = \sum_{k=1}^L x_k f(\cdot - \tau_k),$$

with L the number of ultrasonic impulses and shift parameters τ_1, \dots, τ_L and $x_k \in \mathbb{R}$ for $k = 1, \dots, L$. The problem is that the exact shifts are unknown so we need to choose a dictionary $D = \{f(\cdot - \tau_k)\}_{k=1}^M$ that at least contains good approximations. The sampling points $1 \leq \mathbf{t}_1 < \dots < \mathbf{t}_M < \infty$ are chosen such that $|\mathbf{t}_{j_k} - \tau_k|$ is small enough, so we can guarantee that there exists t_{j_k} such that $|t_{j_k}|$ is small and

$$m = \sum_{k=1}^L x_k f(\cdot - \tau_k) \approx \sum_{k=1}^L \tilde{x}_k g(\cdot - \mathbf{t}_{j_k}).$$

Now it is possible to discretize the equation $m = g * X$:

$$m := \begin{bmatrix} m(t_1) \\ \vdots \\ m(t_N) \end{bmatrix} \quad G := \begin{bmatrix} g(t_1 - \mathbf{t}_1) & \dots & g(t_1 - \mathbf{t}_M) \\ \vdots & \ddots & \vdots \\ g(t_N - \mathbf{t}_1) & \dots & g(t_N - \mathbf{t}_M) \end{bmatrix} \quad X := \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix} \quad (2.1)$$

To perform one MP iteration of the data m with dictionary D we need to calculate:

$$\arg \max_{j=1,\dots,M} |\langle m, g(\cdot - \mathbf{t}_j) \rangle| = \arg \max_{j=1,\dots,M} \left| \sum_{k=1}^N m(t_k) g(t_k - \mathbf{t}_j) \right| = \arg \max |G^T m|.$$

We have assumed that the number of ultrasonic impulses L is known for every a-scan. This assumption will never be realistic since we do not know the exact number of defects. Instead, it is reasonable to fix an upper bound for L . Considering a TOFD measurement, we will always get two strong signals (lateral wave + backwall echo) plus the signals coming from the different defects:

$$L \approx 2 + Y,$$

being Y our guess for the number of defects inside the tube. In many a-scans there will be just two signals caused by the lateral wave and backwall echo, so the MP algorithm will reconstruct at least Y signals caused by noise. That is why a pure MP algorithm will never denoise completely our data.

2.2 Image segmentation stage

A typical ROI of a deconvolved b-scan (after using a sparse deconvolution method) could look like Figure 2.3, where two parabolic shapes and one linear shape could correspond to pore and crack defects respectively. It is clear that defect areas are just a very small part of the whole ROI. That is why several techniques are employed to subdivide an image into its constituent regions or objects. In this way, a more exhaustive image processing can be applied in the next steps. The image segmentation in different regions or objects can be done using several techniques, including morphological processing techniques [37], clustering algorithms [83, 58], etc. The present research work uses clustering algorithms to segment the image into different regions. These divide a given Point Cloud of Data (PCD) into different clusters. For example, in Figure 2.3, a clustering algorithm would allow us to cluster the three observed shapes and separate them from the remaining noise.

By clusters, we mean a set of points that have the following properties:

- Elements that belong to the same cluster should be related to each other / to the cluster.
- Elements that belong to two different clusters should be unrelated to each other / to the cluster.

In this case, the relation is defined as the distance between points. For example, given a set of points $S = \{p_1, p_2, \dots, p_m\}$ we could say that two elements p_i and p_j are related if $\|p_i - p_j\|_2$ is smaller than a certain threshold. So basically, a clustering algorithm tries to create clusters according to certain relation measure between the set elements.

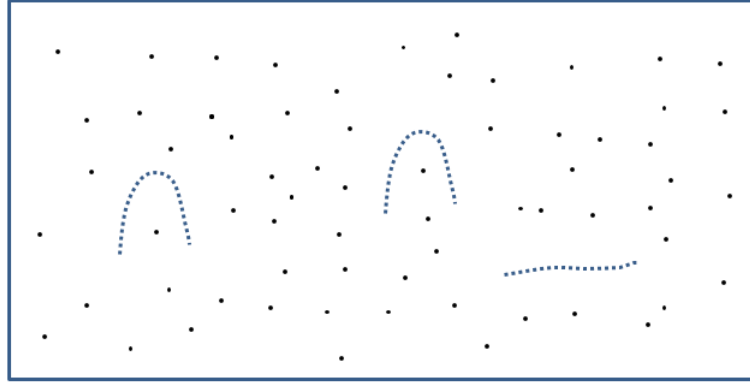


Figure 2.3: Deconvolved b-scan after pre-processing step. Three clusters and some remaining noise represented.

There are many clustering algorithms available in literature, such as the Complete and Single-Linkage algorithm [53], k-means algorithm [41], fuzzy clustering algorithms [10], etc.

One can limit the number of suitable algorithms for each application. In order to do that, one needs to consider the following questions:

- Which is going to be the function used for relation between the data elements?
- Do we allow overlapping clusters?
- Do we know the number of clusters a priori?

One algorithm that is suitable for TOFD data is the so-called Density-based spatial clustering of applications with noise (DBSCAN) [42] due to the following reasons:

- It does not need to know the number of clusters in advance. This is our case in NDT data since we do not know the number of defects inside the material.
- It does not require one element to be related with all other elements of the cluster. This is our case in TOFD data since there might be a large distance between the first and the last peak caused by a defect.
- It can handle noisy data and will serve as an extra denoising process, since noise points are not clustered and therefore eliminated.

One disadvantage is that it cannot properly handle overlapping clusters (intersection of defects in b-scans). This will be addressed in section 5.2.2. In the following subsection, a short introduction to DBSCAN is included. For a more detailed overview, the reader is referred to [42].

2.2.1 Density-based spatial clustering of applications with noise (DBSCAN)

DBSCAN is a data clustering algorithm that models clusters as high-density clumps of points. When looking at Figure 2.3, one can confirm with no doubt that there are “clusters” of points and noise points which do not belong to those clusters.

By clusters, we mean areas with a density of points appreciably higher than outside those areas. Indeed, the density in the noise areas is smaller than in any cluster area.

Before we explain how the algorithm works itself, some definitions are needed. We consider the 2D case. Note that all the next definitions can be easily extend to higher dimensions. Let S be a PCD. Then the definitions are as follows:

Definition 2.1. The ϵ -neighbourhood N_ϵ of a point $p \in S$ is defined as follows:

$$N_\epsilon(p) = \{q \in S : d(p, q) \leq \epsilon\}.$$

Basically, a ball with radius ϵ can be drawn around p , and then the points q contained in that ball are considered to be the neighbourhood of p (Figure 2.4a). With this definition in mind, now we can classify all points from the PCD S :

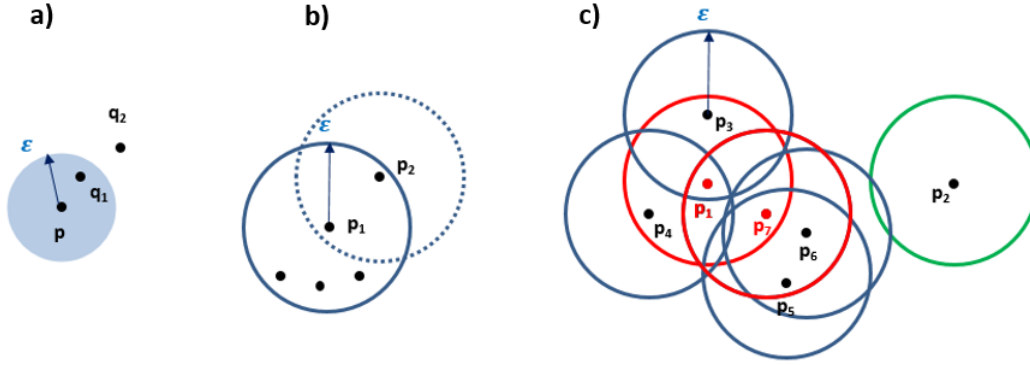


Figure 2.4: a) Taking ϵ as radius, we can draw an ϵ -ball. q_1 is a neighbour of p whereas q_2 is not; b) Given ϵ and $P_{min} = 3$, p_1 is a core point while p_2 is not; c) Representation of core, density reachable and noise points.

Definition 2.2. Let $P_{min} \in \mathbb{N}$. We say that $p \in S$ is a core point if the number of neighbours $|N_\epsilon(p)| \geq P_{min}$.

In Figure 2.4b we further explain this concept.

Definition 2.3. Given ϵ and P_{min} , we say that a point q is directly density-reachable from p if $q \in N_\epsilon(p)$ and $|N_\epsilon(p)| \geq P_{min}$.

Looking at the example of Figure 2.4c, for the given ϵ and considering $P_{min} = 3$, we can see that p_3 is directly density-reachable from p_1 , since p_3 belongs to the neighbourhood of p_1 and $|N_\epsilon(p_1)| = 3 \geq P_{min}$, so both conditions are satisfied.

On the other hand, p_6 is not directly density-reachable from p_1 , since it does not belong to its neighbourhood.

Definition 2.4. We say that a point q is density reachable from p w.r.t. ϵ and P_{min} if there exists a path p_1, \dots, p_n being $p_1 = p$ and $p_n = q$ such that p_{i+1} is directly density-reachable from p_i .

In Figure 2.4c, red dots represent core points. We can see that p_6 is density reachable from p_1 since one can find the path p_1, p_7, p_6 with the properties defined above. A core point forms a cluster with all points that are reachable from it. Points that are not reachable from any others are *noise* (see p_2 in Figure 2.4c).

Once these definitions are clear, one can schematise the DBSCAN algorithm as follows:

- (i) The only two parameters of the algorithm are set: ϵ and P_{min} .
 - (ii) It starts with a random point p . If $|N_\epsilon(p)| \geq P_{min}$ then a cluster is started. If this is not the case this point is considered as noise, but take into account that later it could join a cluster if it is density reachable or for another value of ϵ . The algorithm retrieves all points density-reachable from p w.r.t. ϵ and P_{min} .
 - (iii) If p is a border point, no points are density-reachable from p and DBSCAN looks for the next point of the point cloud, in order to look for new clusters or noise points.
- An example of clustering using DBSCAN algorithm can be found in Figure 2.5.

The DBSCAN clustering algorithm is a powerful method to divide the image generated by the ultrasonic NDT techniques into different clusters. However, some drawbacks has to be mentioned:

- It is not able to properly handle data sets with different densities, since in that situation it is very complicated to find a suitable value for ϵ .
- In some cases, representation of defects in b-scans could intersect. For instance, in Figure 2.6 we see that the two parabolas intersect each other. DBSCAN could fail in these cases, since you may start clustering from a parabola branch, but when getting to the cross intersection, you may continue with the wrong path. For such cases, we have elaborated in section 5.2.2 an algorithm based on central differences which is able to cluster intersecting defects without problem.

2.3 Image classification stage

After we have segmented the image and extracted the main features from denoised b-scans, it is time to classify the different types of defects. The main part of our work precisely focuses on new techniques at this step to automatically classify TOFD defects.

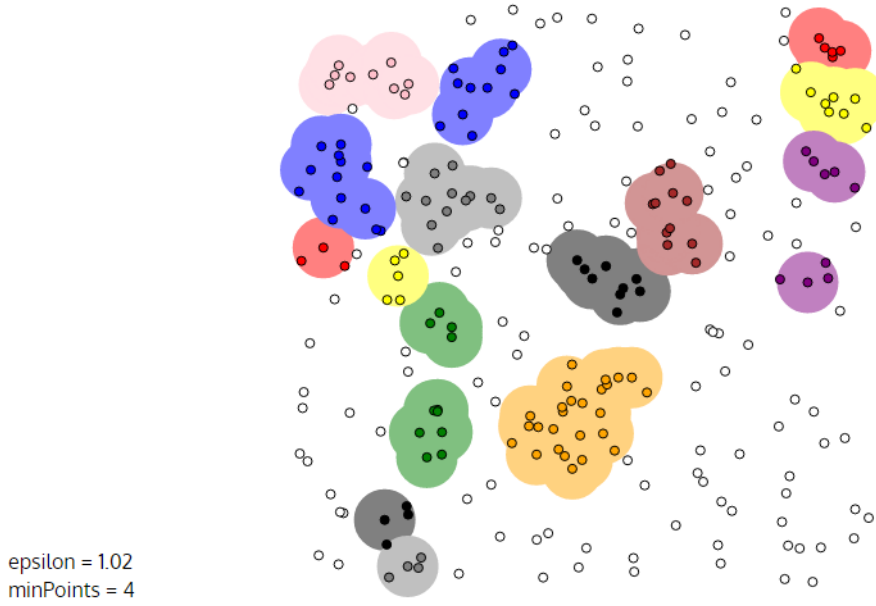


Figure 2.5: Clustering after applying DBSCAN to the PCD shown. Different colours represent different clusters and white dots are noise points. Source: [71].

Different techniques are available in the literature for extracting different features of the defects and classifying them. Among them we can name wavelet features [74], texture features [96, 88], first order statistical descriptors [48], geometric descriptors [74], etc.

Given first statistical descriptors, in [27, 26] they use Artificial Neural Networks (ANN) and linear classifiers to perform a classification based on those descriptors. In the rest of this section, we explain deeper some of these ideas.

2.3.1 Statistical classification, linear classifiers and artificial neural networks

Let us suppose that a set of categories is given. In statistics, classification can be seen as the process of identifying to which of these categories a new observation belongs. These are created based on different data observations. For example, rocks can be classified according to the hardness, rugosity, colour, geomorphologic origin, etc. Each of these characteristics of the rocks could be weighted, i.e., some of them have greater importance than others for classification. Basically, classification is a problem of pattern recognition. When a new observation is taken, there is a function who maps it into one of these categories. These kind of algorithms are known as classifiers, in which a mathematical function is built to assign one category to the input data.

In the NDT field, and especially in TOFD data measurements, the categories would be the different types of defects. After image pre-processing and segmentation, one is able to extract or cluster the different PCD that correspond to the different defects (parabolic

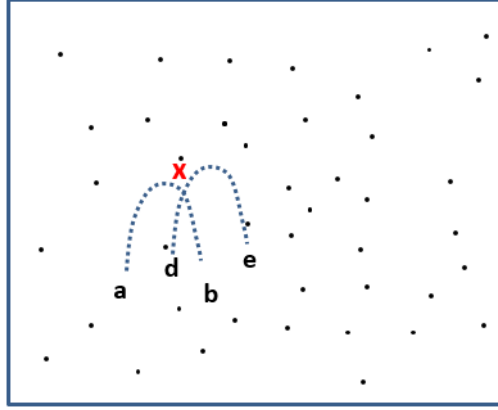


Figure 2.6: b-scan with two parabolic defects that intersect. Ideally, one would like to find two clusters: the parabola ab and the parabola db . The problem with DBSCAN is that you may start clustering from a but when getting to x , the algorithm may continue through the path xe .

PCD, linear PCD). One could describe those features statistically, i.e., computing the mean, standard deviation, energy, skewness, etc. The task of a classification algorithm is to identify a new defect with the created categories that are based on the experience and observations of data.

Recently, Artificial Neural Networks (ANN) have gained quite some popularity as automatic systems for interpretation and classification of defects [27, 69, 63, 62]. ANN structures are similar to that of the human brain, and are able to process a huge amount of data in short periods of time. Their structure is normally based on a system of interconnected "neurons" which can interact between each other (Figure 2.7). These interactions can be based on experience, and therefore ANN are able to learn, which make ANN a really interesting tool. We can reformulate the problem in a mathematical way as follows:

Let G be a class of functions. By learning we mean applying different observations to find a function $f \in G$ such that the problem is solved optimally, i.e., for a cost function $C : G \rightarrow \mathbb{R}$, the target is to find a function f such that f is the solution with least cost:

$$C(f) \leq C(g) \quad \forall g \in G$$

In many cases, the solution will depend on the data, and clearly the cost must be a function of the observations in order to model according to the data. Let us consider one of the easiest examples for a better understanding. Let (x, y) be pairs of some distribution D . Imagine we want to find the function f that minimizes

$$G = E[(f(x) - y)^2]$$

Probably, we only have N samples from the whole PCD and therefore the cost is

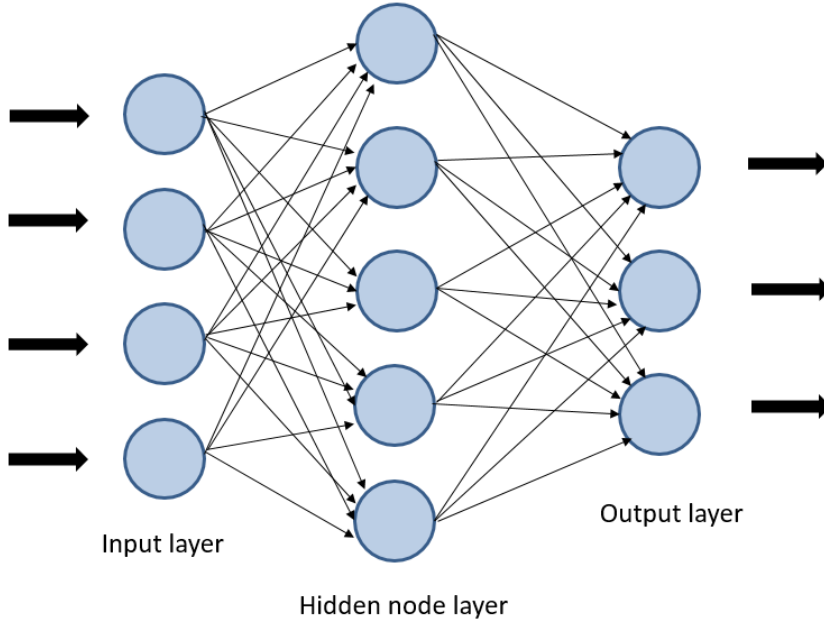


Figure 2.7: Representation of an Artificial Neural Network. In TOFD classification of defects, as input we would have a TOFD image or measurement. The output is the type of defect group to which it belongs.

minimized only over a sample:

$$\hat{G} = \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2$$

ANN can use linear classifiers to make classifications based on values of a linear combination of the characteristics or features values. These values are normally represented in a input vector. The mathematical formulation of the problem is the following:

Let \bar{x} be the input feature vector to the classifier. Then the output is

$$y = f(\bar{w} \cdot \bar{x}) = f\left(\sum_j w_j x_j\right),$$

being \bar{w} the vector of weights. This means there will be variables “more important” than others for the classification. We consider another illustrative example.

Imagine that given a sentence, we would like to classify it as *positive* or *negative*. To this end, a linear classifier can be used. The structure of the problem is illustrated in Figure 2.8.

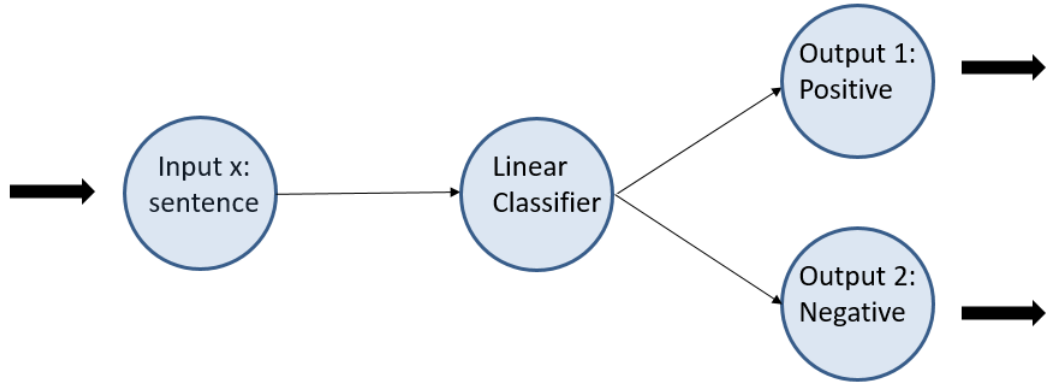


Figure 2.8: Structure of the proposed example for understanding of a linear classifier.

How does this classifier work? One can imagine a very simple threshold classifier. This would be just counting the number of positive and negative words in a sentence. If $\text{Number of positive words} > \text{number of negative words} \implies y = \text{positive}$. Otherwise $y = \text{negative}$.

Given an input sentence like “We went to the cinema. The movie was great and the chairs good. On the other hand, the company was disgusting“, we find two positive words and one negative. Accordingly $y > 0$. This threshold lists have some limitations:

- How is created the positive-negative word list?
- Words have different degree of sentiment (superb > great). How do we weight different words or variables in general?
- Single words may not be enough, i.e., “not good” is negative.

The first two problems in the list are normally addressed by learning a classifier and the last problem needs a more elaborate solution. Therefore a linear classifier uses training data to give a weight for each word. In the table below an example of weighted words is given. Then for the given input sentence “We went to the cinema. The movie was great and the chairs good. On the other hand, the company was disgusting“, now we compute the score taking the weights into account:

$$y = 1 + 1.5 - 2 = 0.5 > 0 \quad \text{positive}$$

This method receives the name of linear classifier because the output is a weighted sum of the input. Following ideas of this mini example, linear classifiers have been used in content based-recommendation systems [78].

Words	Weights
horrible	-3.1
bad	-1
disgusting	-2
fantastic	2.3
awesome	3
superb	2
great	1.5
good	1
the, food, cinema, chair, movie, etc	0

Table 2.1: Weights defined for each word.

3 An introduction to topological data analysis

As we said in the Introduction, our classification methods are innovative in the NDT field and are based on persistent and tangential homology ideas. In this chapter, an introduction to this topic is given. This is necessary in order to understand the foundations of our methods. For a more complete analysis on topological data analysis, the reader is invited to consult [73, 23, 70].

3.1 Simplices and simplicial complexes

Simplices are generalizations of triangles or tetrahedrons to arbitrary dimensions, and are the simplest cells used in persistent homology. Let $U = \{u_0, \dots, u_n\} \subset \mathbb{R}^n$ be $n + 1$ affinely independent points in \mathbb{R}^n , which means $u_1 - u_0, \dots, u_n - u_0$ are linearly independent. This is needed so that we don't end up with collinear vertices. Then, an n -simplex $\Delta_n \equiv \Delta_n(U)$ is defined as a n -dimensional polytope, given by the convex hull of its $n + 1$ vertices:

$$\Delta_n = \left\{ U = \sum_{j=0}^n \lambda_j u_j : 0 \leq \lambda_j \leq 1 \text{ and } \sum_{j=0}^n \lambda_j = 1 \right\} \subset \mathbb{R}^n.$$

The most regularly used simplices in \mathbb{R}^3 are shown in Figure 3.1.

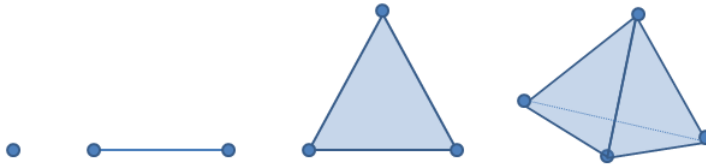


Figure 3.1: 0-, 1-, 2-, and 3- simplices: vertex, edge, triangle, tetrahedron.

Taking simplices as building units, it is possible to combine them. A *simplicial complex* K is a finite collection of simplices that satisfy the two following requirements:

- (i) Every face of a simplex from K is also in K .

- (ii) The intersection of any two simplices τ_1 and τ_2 in K is either the empty set or a face that belongs to both τ_1 and τ_2 .

A d -face σ_d of a simplex τ of dimension p is the d -dimensional subset of $d+1$ vertices, where $0 \leq d \leq p$. We write K_d to indicate the set of d -dimensional simplices in K . In Figure 3.2 we represent a valid simplicial complex.

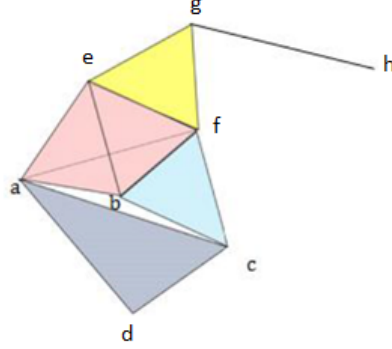


Figure 3.2: A simplicial complex K consisting on vertices, edges, triangles and one tetrahedron, where $K_0 = \{a, b, c, d, e, f, g, h\}$; $K_1 = \{ab, bc, cd, da, ae, eb, bf, ef, eg, fg, gh\}$; $K_2 = \{acd, afb, aeb, bef, efg\}$; $K_3 = \{abfe\}$ and $k_d = 0 \quad \forall \quad d > 3$.

3.2 Filtrations. The Vietoris-Rips complex

We denote by L a *simplicial subcomplex*, which is a subcollection of simplices from K which forms a simplicial complex by itself. Let $N \in \mathbb{N}$ be a natural number. We define a *filtration* F of K as a nested collection of simplicial complexes $F_n K \rightarrow K$, for $n \in \{0, \dots, N\}$, which goes from the empty set until K in the following way:

$$\emptyset = F_0 K \rightarrow F_1 K \rightarrow \dots \rightarrow F_{N-1} K \rightarrow F_N K \rightarrow K,$$

where N denotes the length of the filtration. Some interesting topological features (e.g., connected components and holes) appear and disappear when building a simplicial complex K from subcomplexes through a filtration F . We show this process in Figure 3.3.

In this research work, we use *Vietoris-Rips complexes (VR) with same birth times* for building a filtration F . To understand how this approach works, consider the example from Figure 3.4. Let $X = \{x_1, \dots, x_N\}$ be a set of N vertices forming a point cloud. This PCD is the first step of F (Figure 3.4a). Then N ϵ -balls are centered at each vertex in X , being ϵ the radii of the balls. Further steps of the filtration are created increasing gradually the radius of the balls.

Now the VR complex at each step of the filtration is built as follows:

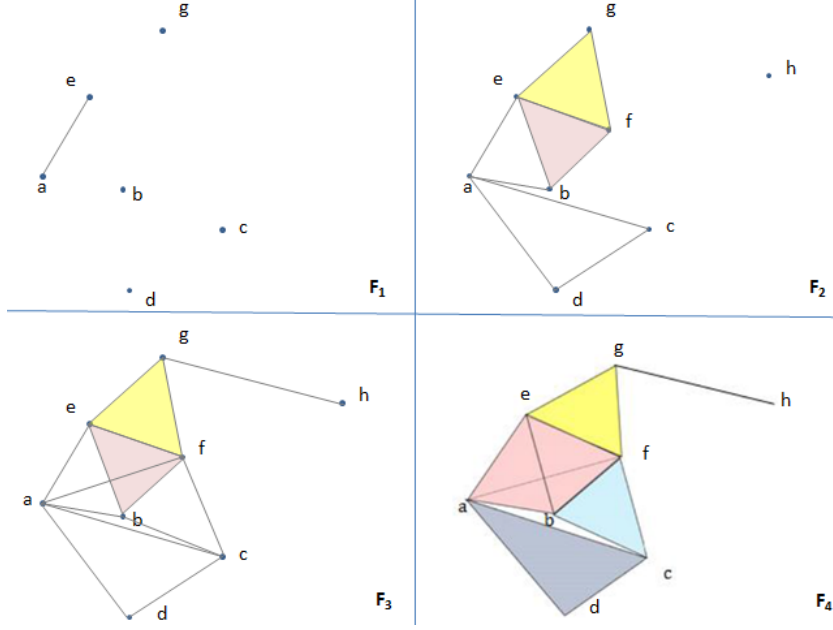


Figure 3.3: Filtration of length 4 of the simplicial complex K of Figure 3.2. Note that the intermediate subcomplexes are very different from the final state. Connected components and holes appear and disappear across F .

- When two balls intersect, an edge (1-simplex) is added between the two centers (Figure 3.4b).
- When there is a closed loop between three 1-simplices, i.e., pairwise intersection between three balls, one 2-simplex (triangle) is added (Figure 3.4b,c and d). The same applies for n -simplices of higher dimensions.

3.3 Homology groups and Betti numbers

Every simplicial complex has different topological invariants such as connected components and holes. These can be detected by homology, a well known topological tool. The *homology groups* of a simplicial complex K , denoted by $H_n(K)$, give the number of connected components for $n = 0$ and the number of n -dimensional holes $\forall n \geq 0$.

In order to understand how $H_n(K)$ are computed, we refer to the simplicial complex in Figure 3.5, a triangle with its interior.

To compute the homology groups we first need to describe the *n-dimensional chain groups* $C_n(K)$ of K , which by definition are formal linear combinations of n -simplices. For instance, in Figure 3.5 C_0 is isomorphic to \mathbb{Z}^3 with basis v_1, v_2 and v_3 . C_1 is isomorphic to \mathbb{Z}^3 with a basis given by the 1-simplices e_1, e_2 and e_3 . C_2 is isomorphic to \mathbb{Z} with basis the 2-simplex T . Finally, $C_n = 0 \forall n > 2$.

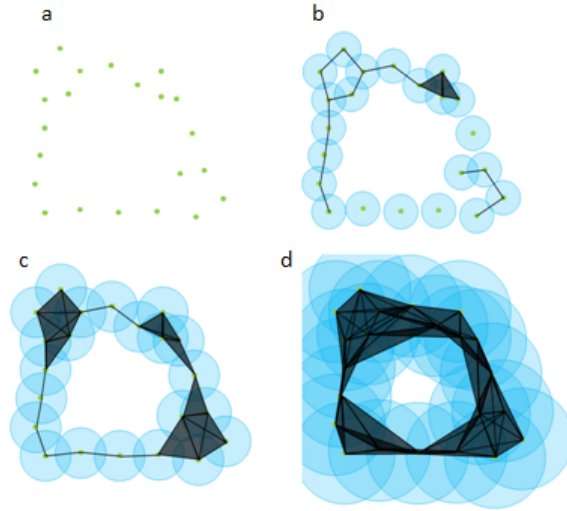


Figure 3.4: Four steps of a filtration from random PCD where in each step a Vietoris-Rips complex is built. Number of connected components and holes are different at each step of the filtration. For instance in *b* there are six connected components and one 1-dimensional hole.

An oriented simplicial complex is a simplicial complex K equipped with a partial ordering on its vertices that restricts to a linear ordering on each simplex (Figure 3.5). Let $\sigma = (\sigma_0, \dots, \sigma_n)$ be an oriented n -simplex. The boundary operator $\partial_n : C_n \rightarrow C_{n-1}$ of σ is the homomorphism:

$$\partial_n(\sigma) = \sum_{i=0}^n (-1)^i (\sigma_0, \dots, \sigma_{i-1}, \dots, \sigma_{i+1}, \dots, \sigma_n),$$

where $(\sigma_0, \dots, \sigma_n)$ are the faces of σ . In Figure 3.5, we see that

$$\begin{aligned} \partial(v_1) &= \partial(v_2) = \partial(v_3) = 0, \\ \partial(e_1) &= v_2 - v_1, \quad \partial(e_2) = v_3 - v_2, \quad \partial(e_3) = v_3 - v_1, \quad \text{and} \\ \partial(T) &= e_1 + e_2 + e_3. \end{aligned}$$

Geometrically, a *cycle* is a closed loop (e.g. $e_1 + e_2 + e_3$ in Figure 3.5). From the algebraic viewpoint, the *cycle group* $Z_n(K)$ is the subspace generated by the kernel of the boundary operator ∂_n :

$$Z_n = \ker(\partial_n).$$

The *boundary group* $B_n(K)$ is the subspace generated by the image of the boundary operator ∂_{n+1} .

$$B_n = \text{im}(\partial_{n+1}).$$

We compute all the cycle and boundary groups for the example of Figure 3.5:

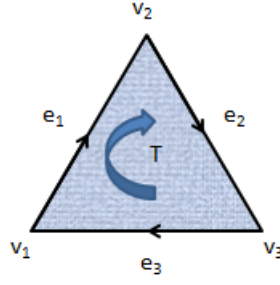


Figure 3.5: An oriented simplicial complex K with 0-simplices v_1, v_2, v_3 , 1-simplices e_1, e_2, e_3 and 2-simplex T .

- Since the boundary of a 0-simplex or vertex is 0, then Z_0 is the space generated by the 3 vertices,

$$Z_0 = \ker(\partial_0) = C_0 = \langle v_1, v_2, v_3 \rangle.$$

- Computing $Z_1 = \ker(\partial_1)$ means solving the problem $\partial(\alpha e_1 + \beta e_2 + \gamma e_3) = 0$, i.e.,

$$\alpha(v_2 - v_1) + \beta(v_3 - v_2) + \gamma(v_1 - v_3) = 0, \quad \forall \alpha, \beta, \gamma \in \mathbb{R}.$$

This leads to $\alpha = \beta = \gamma$ and therefore $Z_1 = \ker(\partial_1) = \langle e_1 + e_2 + e_3 \rangle = \mathbb{Z}$.

- For the rest of cycles groups

$$Z_n = 0 \quad \forall n \geq 2.$$

- Then boundary groups are computed:

$$B_0 = \text{im}(\partial_1) = \langle v_2 - v_1, v_3 - v_2, v_1 - v_3 \rangle.$$

$$B_1 = \text{im}(\partial_2) = \langle e_1 + e_2 + e_3 \rangle,$$

$$B_n = 0 \quad \forall n \geq 2.$$

Note that $(v_2 - v_1) + (v_1 - v_3) = -(v_3 - v_2)$, hence $B_0 = \langle v_2 - v_1, v_3 - v_2 \rangle = \mathbb{Z}^2$.

The n -dimensional homology group is defined as the quotient of the n -cycle group and the n -boundary group,

$$H_n(K) = Z_n(K)/B_n(K).$$

For the example of Figure 3.5, the homology groups are

$$H_0 = Z_0/B_0 = \langle v_1, v_2, v_3 \rangle / \langle v_2 - v_1, v_3 - v_2, v_1 - v_3 \rangle = \mathbb{Z},$$

$$H_1 = Z_1/B_1 = \mathbb{Z}/\mathbb{Z} = 0.$$

We define the *Betti numbers* β_n as the rank of each homology group. β_0 measures the number of connected components in a simplicial complex and β_n the number of

n -dimensional holes $\forall n \geq 0$. In our example from Figure 3.5, $\beta_0 = 1$ and $\beta_1 = 0$. To summarize, the reader should keep in mind that homology is a tool that allows us to calculate the number of connected components and holes of a simplicial complex.

We remark that Betti numbers are sensitive with respect to perturbations in the simplicial complex, e.g., Betti numbers may change when adding only a single vertex (Figure 3.6). To avoid such instabilities, persistent homology considers a filtration F of a simplicial complex K . Betti numbers are essentially computed in each step of the filtration, keeping track of topological changes (e.g. for holes and connected components).

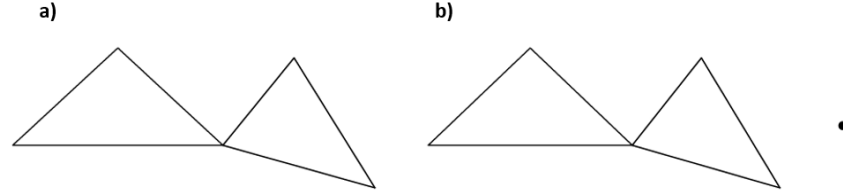


Figure 3.6: By adding a vertex, the Betti Number is incremented by one from a) to b).

3.4 Barcodes and persistent diagrams

As we noted before:

- *Persistent homology* keeps track of the "birth" and "death" of topological features across a filtration F .
- *Betti numbers* are computed in every step of F .

To monitor the change of homology classes a *barcode* is normally used (Figure 3.7). The beginning and end of every bar represents the birth and death (in terms of balls radii) of a topological attribute (1-dimensional holes in case of Figure 3.7). Long bars represent *stable* holes across F while short bars correspond to unstable or noisy features. Note that a similar barcode can be generated in every dimension.

Another equivalent representation to the barcode is the so-called *persistent diagram* (Figure 3.7), which is a 2-dimensional graph where the points represent the topological features. The x and y coordinates give the birth and death of them. Points far away from the diagonal indicate the existence of holes with a long lifetime across the filtration.

Currently, there is no mechanism in most of homology softwares to extract the points involved in a given homology class or hole, but one should be careful about asking which holes in the complex correspond to which generator. In particular, if the complex comes from points sampled on the annulus $1 < ||x|| < 2$, then any collection of points which goes all the way around the hole is a legitimate homology generator, and any two such generators are equivalent in the sense that they generate the same class.

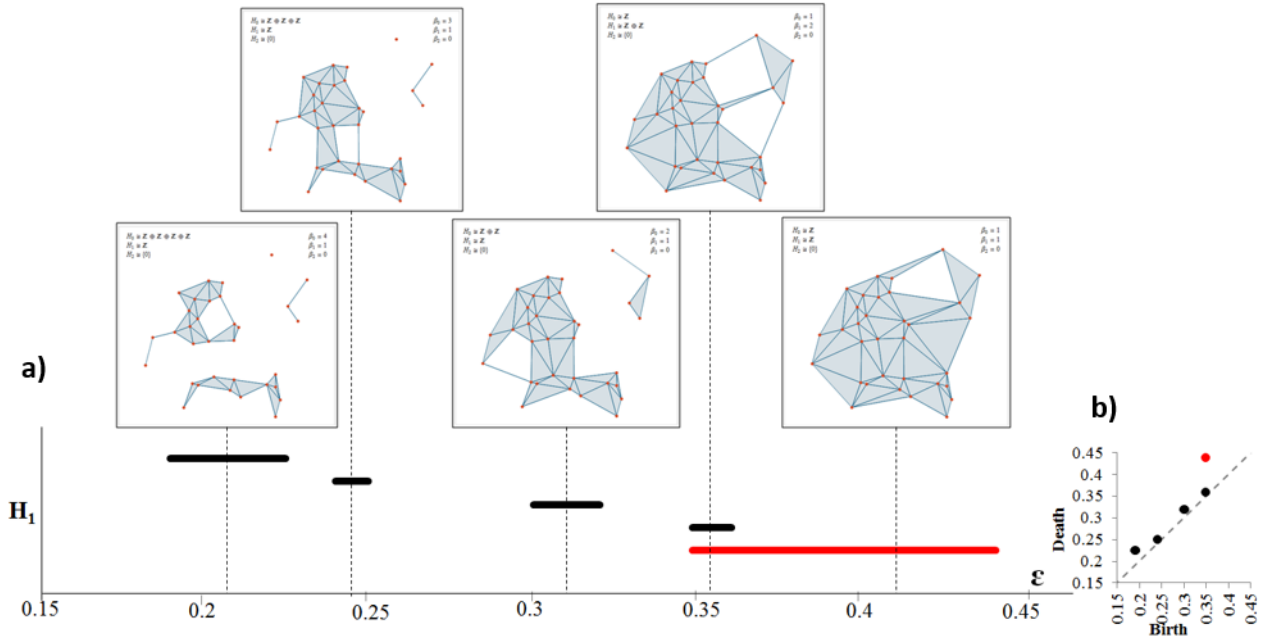


Figure 3.7: a) Barcode representation of $H_1(K)$. Five steps of the filtration are shown. Stable holes correspond to long bars (red). The Betti numbers are equal to the number of bars intersecting the dashed line; b) Equivalent persistent diagram generated with [49].

3.5 Persistent homology with different birth times

We would also like the reader to keep in mind another way of building a filtration, the so-called “*persistent homology with different birth times*”, which will be used in our second classification method exposed in section 5.2.

This type of Vietoris Rips filtration is created with vertex-radius-birth triples [72]. A radius for every vertex ball is fixed and it will not change. The idea is that every vertex gets its own birth time in the filtration. The complex is built in the same way i.e., two vertices get an edge between them if their corresponding balls intersect. This edge appears only when the 2 corresponding vertices have emerged in the filtration.

In Figure 3.8 this procedure is explained. We consider nine points in 2D. Green balls are drawn around every vertex with a certain value for the radius. In the top-right of the figure, the coloured numbers indicate when the vertices appear in the filtration F (bottom-left). In each step of F the VR-complex is built as we explained in the previous subsection. Similarly to Figure 3.7, but now looking at the 0 betti numbers, the barcode of Figure 3.8 keeps track of the birth and death of the number of connected components across F (two connected components in steps 1 and 2 of F , four in step 3, two in step 4 and 1 from step 5 to n).

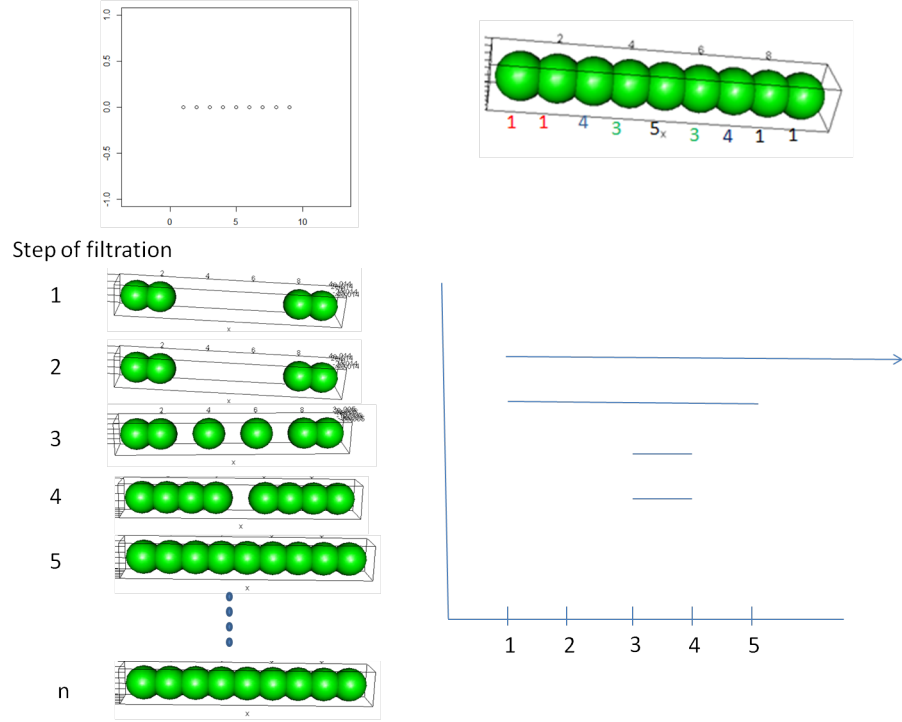


Figure 3.8: VR complex with different birth times.

3.6 Other filtrations

In this section, we present other interesting complexes used to create a filtration. The motivation for their use is presented, as well as the advantages and disadvantages of them.

3.6.1 Čech complex

Let $X = \{x_1, \dots, x_N\}$ be again a set of N vertices forming a PCD. Then N ϵ -balls are centered at each vertex in X , being ϵ the radii of the balls. We define the Čech complex at each scale $\epsilon > 0$ in the following way:

Definition 3.1. Let $\tau = (p_1, \dots, p_m)$ be collection of points in X . We say that τ is an $m - 1$ dimensional simplex of a Čech complex C_ϵ if

$$\bigcap_{j=1}^m B_\epsilon(p_i) \neq \emptyset,$$

or equivalently, if the corresponding ϵ balls have a nonempty intersection.

For instance, if we consider the PCD a, b, c given in Figure 3.9 and the ϵ -balls shown, one can extract the following conclusions:

- ab belongs to the Čech complex C_ϵ because the intersection of the corresponding balls is nonempty. The same applies for ac and bc .
- The simplex abc does not belong to C_ϵ , since the triple intersection is empty (small white area of Figure 3.9 left). Instead the associated Rips complex does contain abc as explained in the previous section.

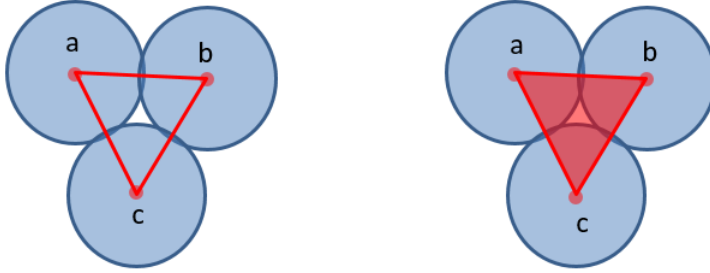


Figure 3.9: Left: Čech complex for the given PCD and ϵ ; right: the Rips complex does contain the 2-dimensional simplex abc .

From the definition of a Čech complex, it is clear that a simplex τ will appear in the step of the filtration where one finds the smallest radius ϵ such that the intersection between all balls of points belonging to τ is nonempty. This problem can be reformulated in another way [73]:

*Given a collection τ of points in \mathbb{R}^d , what is the smallest radius r so that there exists some point x in \mathbb{R}^d whose distance from each point in τ is less than r ? This is known as the *smallest enclosing ball problem* and for large PCD it is pretty complicated.*

Therefore, from the computational point of view, building a Vietoris-Rips filtration is cheaper than a Čech for the same PCD since for the first of them just simple pairwise distance calculations are needed. Once we know how the Čech complex is built, it is easy to define how to create the filtration. It is done in the same way that a Vietoris-Rips Filtration by changing the value of the parameter ϵ .

We already mentioned the main disadvantage of the Čech filtration, which is related to the computational complexity of the smallest enclosing ball problem. On the other hand, this filtration “beats” the Rips in the sense that it is faithful to the topology of the union of balls. Taking a look at Figure 3.9, the 2-dimensional simplex will not appear until the triple intersection is nonempty. Algorithms to construct a Čech filtration can be found in [34].

3.6.2 Alpha complex

Vietoris-Rips or Čech complexes may contain many simplices when the value ϵ -balls are big. There exists a variant of the Čech complex called alpha complex, which considers the intersection of ϵ -balls with the Voronoi cells for the given PCD.

3 An introduction to topological data analysis

Let S be a finite set of points in \mathbb{R}^d and ϵ a positive real number. For each $p \in S$ let $B_\epsilon(p)$ a ball of radius ϵ at each point p , and let

$$U(\epsilon) = \bigcup_{p \in S} B_\epsilon(p)$$

be the union of these balls.

Definition 3.2. [57]. Considering the 2-dimensional case, let $S = \{p_1, p_2, \dots, p_n\}$ be a finite set of n points in \mathbb{R}^2 . Let $d(p_i, p_j)$ the euclidean distance between two points p_i and p_j . The locus of points which are closer to p_i than to p_j , denoted by $L(p_i, p_j)$, is given by one of the half planes that results from computing the bisector between p_i and p_j . The locus of points closer to p_i than to any other point in S , denoted by $V(i)$ is given by

$$V_i = \bigcap_{i \neq j} L(p_i, p_j),$$

the intersection of all the half planes associated with p_i . Repeating the process for every point in S , this will result in a set of N regions bounded or unbounded, which are known as the Voronoi diagram $V(S)$ of the set S .

This definition can be generalized to higher dimensions. An example of a Voronoi diagram is illustrated in Figure 3.10.

Let us consider the intersection between each ϵ ball and their corresponding Voronoi cell:

$$R_p(\epsilon) = B_p(\epsilon) \cap V_p.$$

This is illustrated in Figure 3.11. At each scale $\epsilon > 0$, the alpha complex $\alpha(\epsilon)$ can be defined as follows:

Definition 3.3. A simplex $\tau = (p_1, p_2, \dots, p_q)$ belongs to $\alpha(\epsilon)$ if

$$\bigcap_{p \in \tau} R_p(\epsilon) \neq \emptyset.$$

To further explain this, we refer to the example of Figure 3.11. The simplex $\{127\}$ exists because there is triple intersection between $R_1(\epsilon)$, $R_2(\epsilon)$ and $R_7(\epsilon)$. On the other hand, the edge $\{26\}$ does not belong to the complex since the intersection between $R_2(\epsilon)$ and $R_6(\epsilon)$ is empty. From the definition of alpha complex $\alpha(\epsilon)$ the following holds:

- $R_p(\epsilon) \subseteq V_p$.
- $R_p(\epsilon) \subseteq B_p(\epsilon)$ and therefore $\alpha(\epsilon) \subseteq C_\epsilon$.

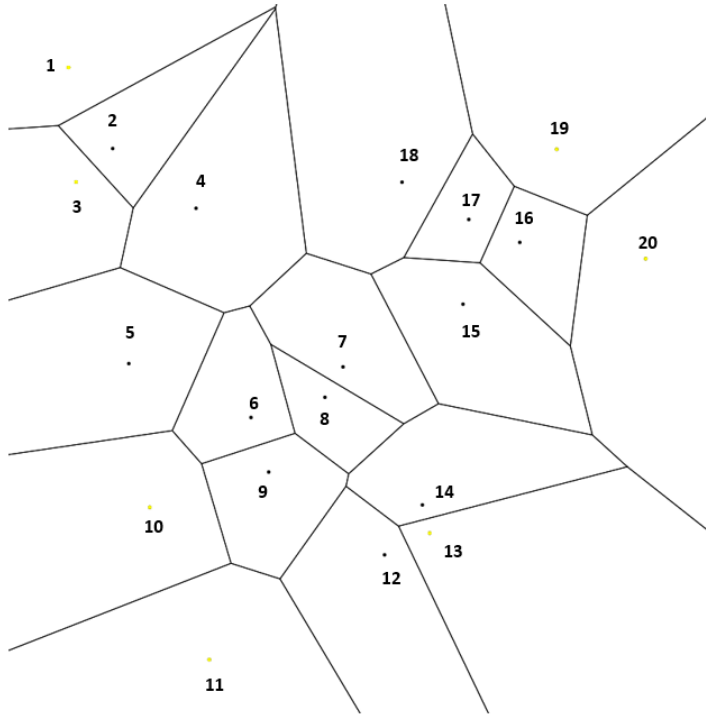


Figure 3.10: The Voronoi diagram of a set of 20 points in 2D.

One of the advantages of alpha complexes respect to Čech complexes is that they use less simplices. This is something that can be clearly seen in the example of Figure 3.11, where for example the edge $\{74\}$ will never exist even for a large value of ϵ since the limit of Voronoi regions makes it impossible. On the contrary, this edge will appear for some ϵ big enough in the Čech or Vietoris-Rips filtration.

As in the previous sections, once we have defined how the complex is built, the last step is the creation of the filtration. Given our collection of points S in \mathbb{R}^d , we can increase now the value of ϵ in order to get a nested sequence of alpha complexes, where we denote by K^i the i -alpha complex in the filtration

$$\emptyset = K^0 \subset K^1 \subset \dots \subset K^m.$$

3.6.3 Witness complex

It has been seen that simplicial complexes such as Čech, Vietoris-Rips or Alpha use all points from the PCD in order to build the filtration. For huge complexes, this can be very expensive from the computational point of view. There are some cases in which a smaller vertex set can be representative of the topology of the studied object. In this case, it would be very interesting to create simplicial complexes based only on certain

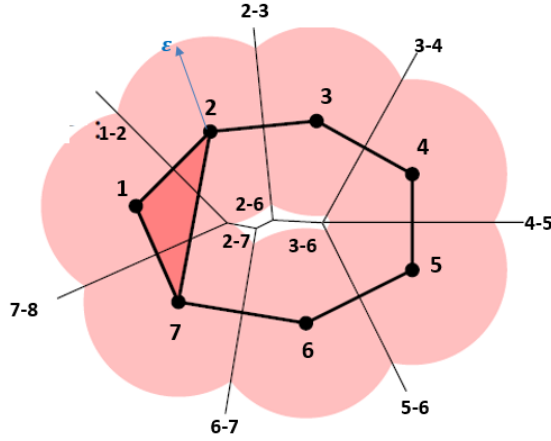


Figure 3.11: Intersection between ϵ -balls and the Voronoi regions. The associated alpha-complex is superimposed.

points of the PCD. These structures are called witness complexes and were introduced in [35], where the reader can find a more comprehensive account on the topic.

Smaller complexes are obtained by choosing a set of landmarks from the PCD, and then building a witness complex on this set. This means that the remaining points that are not used as landmarks, will be employed as witnesses to the existence of higher dimensional simplices in the complex (edges, triangles, etc.).

3.6.3.1 Definition of the witness complex $W(D)$

Let P be the total number of points of a PCD. We consider N data points and a set of landmark points n such that $N + n = P$. Let D be an $n \times N$ matrix with entries greater or equal than 0:

$$D := \begin{bmatrix} d_{11} & \dots & d_{1N} \\ \vdots & \ddots & \vdots \\ d_{n1} & \dots & d_{nN} \end{bmatrix}, \quad (3.1)$$

where d_{ij} represents the distance between a landmark point i and a point j . The *strict* witness complex $W_\infty(D)$ with vertex set $\{1, 2, \dots, n\}$ is defined as follows:

- (i) An edge $\tau = [ab]$ belongs to $W_\infty(D)$ if there exists a point $1 \leq i \leq N$ s.t. $d(a, i)$ and $d(b, i)$ are the two smallest entries in the i -th column of D .
- (ii) By induction in p : imagine all the faces of a p -dimensional simplex $\tau = [a_0, a_1, \dots, a_p]$ belong to $W_\infty(D)$. Then τ also belongs to $W_\infty(D)$ iff one can find a point $1 \leq i \leq N$ s.t. $d(a_0, i), d(a_1, i), \dots, d(a_p, i)$ are the smallest $p + 1$ entries in the i -th column of D .

In both cases, we say that i is a “witness” to the existence of τ . To further explain this, we refer to the example of Figure 3.12 left, where we consider a PCD of ten points. We consider the case where the landmarks are a, b, c, d . Therefore the complex is going to be built only around those vertices and the rest of points will determine the existence of edges and triangles in the $W_\infty(D)$ complex. First of all, one needs to construct the matrix D of distances between landmarks and the rest of points:

$$D := \begin{matrix} & \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \end{matrix} \\ \begin{pmatrix} 4.1 & 3.4 & 1.2 & 4.5 & 5 & 0.8 \\ 4.2 & 3.3 & 0.9 & 4.2 & 4.5 & 1 \\ 4.9 & 4 & 1.8 & 4.7 & 5.1 & 0.5 \\ 1.3 & 1.6 & 2 & 0.8 & 0.7 & 5 \end{pmatrix} & \begin{matrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{matrix} \end{matrix}$$

From the definition on $W_\infty(D)$ and taking a look on D , one can extract the following conclusions:

- 3 is a witness of the existence of the edge ab , since $d(a, 3)$ and $d(b, 3)$ are the two smallest entries in the third column.
- 6 is a witness of the existence of the edge ac .
- The edge bc has no witnesses and therefore does not occur in the complex $W_\infty(D)$.
- 1 is a witness of the edge ad .
- 2, 4 and 5 are witnesses of the edge bd .
- dc has not witnesses and does not occur in $W_\infty(D)$.
- The 2-simplex abd appears in $W_\infty(D)$ and there are three witnesses 2, 4, 5.

In Figure 3.12 right we show the complex $W_\infty(D)$.

In practice, it is costly to compute $W_\infty(D)$ and normally a “lazy version” $W_1(D)$ of the witness complex is used. $W_1(D)$ has the same skeleton than $W_\infty(D)$ and it is defined as the maximal simplicial complex containing $W_\infty(D)$ (Figure 3.13). For instance, in 2D this means that whenever there are three 1-dimensional simplices forming an empty triangle, the 2-simplex will exist in $W_1(D)$. We will write $W(D)$ to mean $W_1(D)$.

3.6.3.2 The weak witnesses theorem

$W_\infty(D)$ can be motivated by comparing it with the Delaunay triangulation in Euclidean space. Here we will present the necessary theorem to address this motivation and we also explain what is a Delaunay complex.

Definition 3.4. Delaunay triangulation and complex. Let $L \subset \mathbb{R}^D$ be a finite set of points. A p -simplex $\tau = [l_0 l_1 \dots l_p]$ is contained in a Delaunay triangulation $\text{Del}(L)$ when $\exists x \in \mathbb{R}^D$ s.t. x is equidistant from the points $l_0 l_1 \dots l_p$ and has no nearer neighbour in L . In such a case x is a strong witness to the existence of τ .

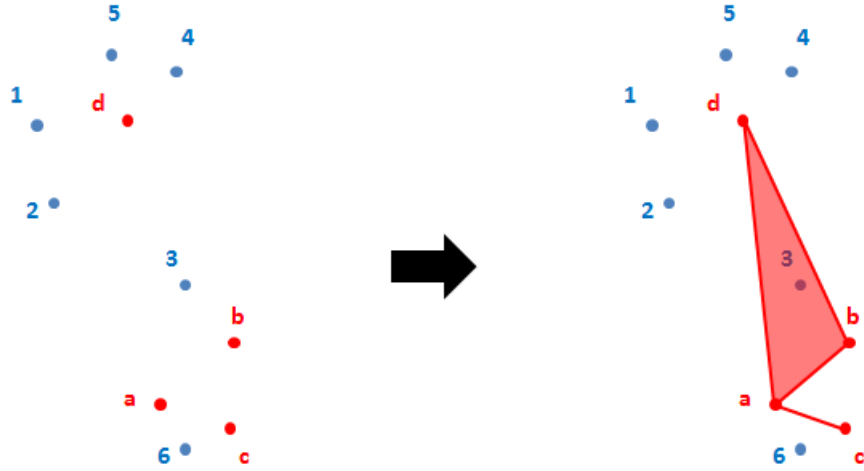


Figure 3.12: Left: PCD consisting on 4 landmark points a, b, c, d and 6 vertices 1, 2, 3, 4, 5, 6. Right: representation of the associated $W_\infty(D)$: 0-simplices a, b, c, d , 1-simplices ab, ad, db, ac and 2-simplex adb .

In Figure 3.14 this definition is explained graphically. We consider the set L of five points l_0, l_1, l_2, l_3, l_4 where x is equidistant to l_0, l_1 and l_2 and there is no nearer neighbour in L (l_3 and l_4 are outside the circumcenter). Therefore x is a strong witness and the 2D simplex $l_0l_1l_2$ exists in a Delaunay triangulation. If the set of witnesses is discrete, it is clear that the probability of finding a strong witness is 0. This is the motivation of the following definition.

Definition 3.5. [35] We say that $x \in \mathbb{R}^D$ is a weak witness for $\tau = [l_0l_1\dots l_p]$ with respect to L if the $p+1$ nearest neighbours of x in L are l_0, l_1, \dots, l_p (tolerating equality).

So, weak witnesses are not necessarily equidistant from l_0, l_1, \dots, l_p . As a consequence we can redefine the notion of $W_\infty(D)$ saying that τ is a p -simplex of $W_\infty(D)$ iff τ and its faces have weak witnesses.

Theorem 3.6. [35] Let $L \in \mathbb{R}^D$ be a finite set of points and $l_0, l_1, \dots, l_p \in L$. Then $\tau = [l_0l_1\dots l_p]$ has a strong witness with respect to L iff τ and its faces have weak witnesses with respect to L .

The consequences of this theorem are pretty important, since now one does not look for a strong witness but instead look for a certain number of weak witnesses.

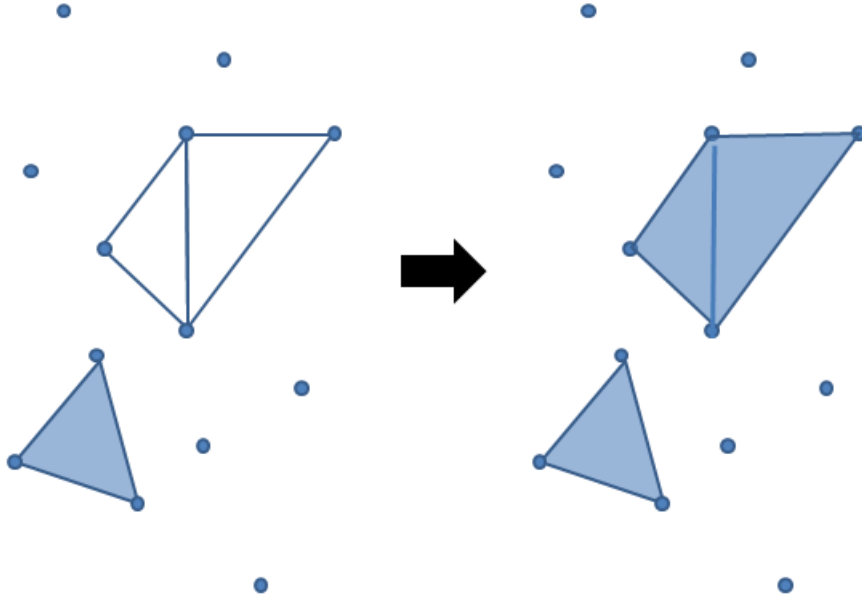


Figure 3.13: Suppose the left figure is a witness complex $W_\infty(D)$. Then the lazy version $W_1(D)$ is shown in the right.

3.6.3.3 How to choose the landmarks

Witness complexes are built just on a certain number of “landmark” points. The first and most important question that arises is how to choose those landmark points? It is clear that they should be representative of the topology of the samples. Different methods have been used to this end [35]:

- Randomly.
- By maxmin algorithm.

Consider a PCD $Z \in \mathbb{R}^D$. The aim is to select i landmark points. Maxmin algorithm works as follows:

- (i) The first landmark $l_1 \in Z$ is chosen randomly.
- (ii) Inductively l_1, l_2, \dots, l_{i-1} have been chosen.
- (iii) Then l_i is the point that maximizes the function

$$z \mapsto \min\{d(z, l_1), d(z, l_2), \dots, d(z, l_{i-1})\},$$

where d is the distance between the potential new landmark candidate and the existing landmarks.

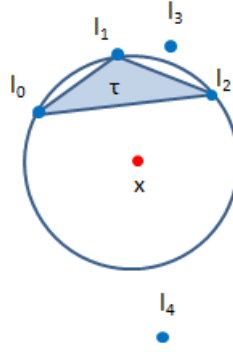


Figure 3.14: x strong witness to the existence of the 2D simplex $\tau = l_0 l_1 l_2$ with respect to L .

(iv) The process is repeated until the number of desired landmarks is reached.

To further explain how maxmin algorithm works, we look at the example of Figure 3.15, where we consider six points $Z = a, b, c, d, l_1, l_2$ in 2D. Imagine that two landmarks l_1 and l_2 have been chosen already and let us assume that the target is to choose a third landmark for this PCD. How is l_3 chosen?

For every candidate (in this case a, c, d, c), we calculate the distances to each landmark and take the minimum between the two of them. The loser distances are the dashed lines of Figure 3.15 and the winners for every case are: $\overline{al_1}$, $\overline{bl_1}$, $\overline{cl_2}$ and $\overline{dl_2}$.

The next step of the maxmin algorithm is to take the maximum of these distances, and hence, the new landmark is found. In the example of Figure 3.15, the winner is $\overline{dl_2}$ so the new landmark is d .

Therefore, the maxmin algorithm gives even spaced landmarks, but one of the disadvantages is that it takes extremal points as well. Both random and maxmin algorithms give good results according to [35].

Another important issue is the number of landmarks that one should use in order to build the complex. In [35] $N/n \geq 20$ gave good results.

3.6.3.4 Building a witness complex filtration

Let D be the matrix $n \times N$ defined previously, where n is a set of landmarks and N are the remaining points in the PCD. Let $R \in [0, \infty)$ and $\nu \in \mathbb{N}$. The cases where $\nu = 0, 1, 2$ are of special importance. The nested family of simplicial complexes $W(D; R, \nu)$ is created with the following idea:

- (i) A value for ν is fixed.
- (ii) R will be the changing parameter that will be used to create the different steps of the filtration. It is equivalent to the radius parameter of the Vietoris-Rips filtration.

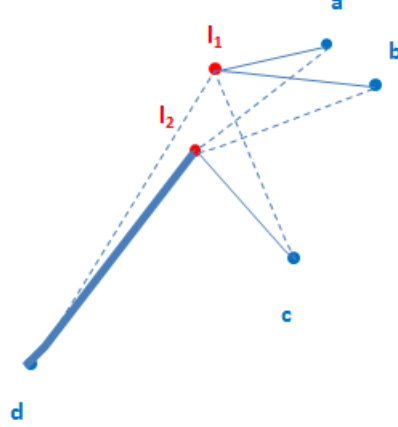


Figure 3.15: Illustration of maxmin algorithm.

The vertex set of $W(D; R, \nu)$ is a set of landmarks $\{1, 2, \dots, n\}$ and the next steps of the filtration are built as follows:

- If $\nu = 0$, then $m_i = 0 \quad \forall \quad i = 1, 2, \dots, N$.
- If $\nu > 0$, then for $i = 1, 2, \dots, N$, let m_i be the ν -smallest entry of the i -column of D . For instance, if $\nu = 2$, then m_3 is the second smallest entry of the third column of D . Intuitively, m_i is the distance between a potential witness i and the ν -closest landmark.

- One edge $\tau = [a_0 a_1 \dots a_p]$ belongs to $W(D; R, \nu)$ iff there exists a witness $i \in \{1, 2, \dots, n\}$ s.t.

$$\max(D(a, i), D(b, i)) \leq R + m_i.$$

- We consider the lazy version of the complex, meaning this that a simplex $\tau = [a_0 a_1 \dots a_p]$ belongs to $W(D; R, \nu)$ if all its faces belong to $W(D; R, \nu)$.

Note that for $R = 0$ and $\nu = 2$, $W(D; R, \nu)$ is equivalent to $W(D)$. To further explain this with an example, we consider again the PCD given in Figure 3.12 left. We are going to build the filtrations ($\nu = 0$, and $\nu = 1$) and the changing parameter R will take the values 0, 1.5 and 3.

First we try to build the filtration considering $\nu = 0$. As explained before in the definition of $W(D; R, \nu)$, if $\nu = 0$, $m_i = 0 \quad \forall i = 1, 2, \dots, N$ and therefore in our example $m_1 = \dots m_6 = 0$. Taking a look on Figure 3.16 left, we see that we need to check when the six edges ab, bc, ac, ad, db and dc appear in the filtration. To check if the edge ab exists

$$\max(D(a, i), D(b, i)) \leq R + m_i$$

needs to be computed. To check if such witness i exists:

3 An introduction to topological data analysis

- Compute

$$D := \begin{matrix} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ \begin{pmatrix} 4.1 & 3.4 & 1.2 & 4.5 & 5 & 0.8 \\ 4.2 & 3.3 & 0.9 & 4.2 & 4.5 & 1 \\ 4.9 & 4 & 1.8 & 4.7 & 5.1 & 0.5 \\ 1.3 & 1.6 & 2 & 0.8 & 0.7 & 5 \end{pmatrix} & \mathbf{a} \\ & & & & & & \mathbf{b} \\ & & & & & & \mathbf{c} \\ & & & & & & \mathbf{d} \end{matrix}$$

- Compare the the rows corresponding to a and b and take the maximum for every column the resulting vector max is:

$$max_{ab} := \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ (4.2 & 3.4 & 1.2 & 4.5 & 5 & 1) \end{matrix} \text{ max between row a and row b}$$

The same vector has to be computed for the rest of edges:

$$max_{ac} := \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ (4.9 & 4 & 1.8 & 4.7 & 5.1 & 0.8) \end{matrix} \text{ max between row a and row c}$$

$$max_{bc} := \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ (4.9 & 4 & 1.8 & 4.7 & 5.1 & 1) \end{matrix} \text{ max between row b and row c}$$

$$max_{ad} := \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ (4.1 & 3.4 & 2 & 4.5 & 5 & 5) \end{matrix} \text{ max between row a and row d}$$

$$max_{db} := \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ (4.2 & 3.3 & 2 & 4.2 & 4.5 & 5) \end{matrix} \text{ max between row d and row b}$$

$$max_{dc} := \begin{matrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ (4.9 & 4 & 2 & 4.7 & 5.1 & 5) \end{matrix} \text{ max between row d and row c}$$

This values are the result of the left side of the above equation.

- Finally we can compare:

$$max(i) \leq R + m_i,$$

for the different cases:

- (i) For $R = 0$, since $m_i = 0$, there is not i such that $\max(i) \leq R + m_i = 0$. There are no edges, so $W(D; 0, 0)$ is given only by the four vertices a, b, c, d (Figure 3.16 left).
- (ii) For $R = 1.5$:
- edge ab : 3 and 6 are witnesses of its existence since $\max_{ab}(3) = 1.2 \leq 1.5$ and $\max_{ab}(6) = 1 \leq 1.5$.
 - edges ac and bc also exist since for example 6 is again a witness. The 2-dimensional simplex abc also exists since we are considering the lazy version of the complex. (Figure 3.16 middle).
- (iii) For $R = 3$: every edge exists. One can find witnesses for every case. For instance 3 is a witness to the existence of the six edges. The 2-dimensional associated simplices also appear (Figure 3.16 right).

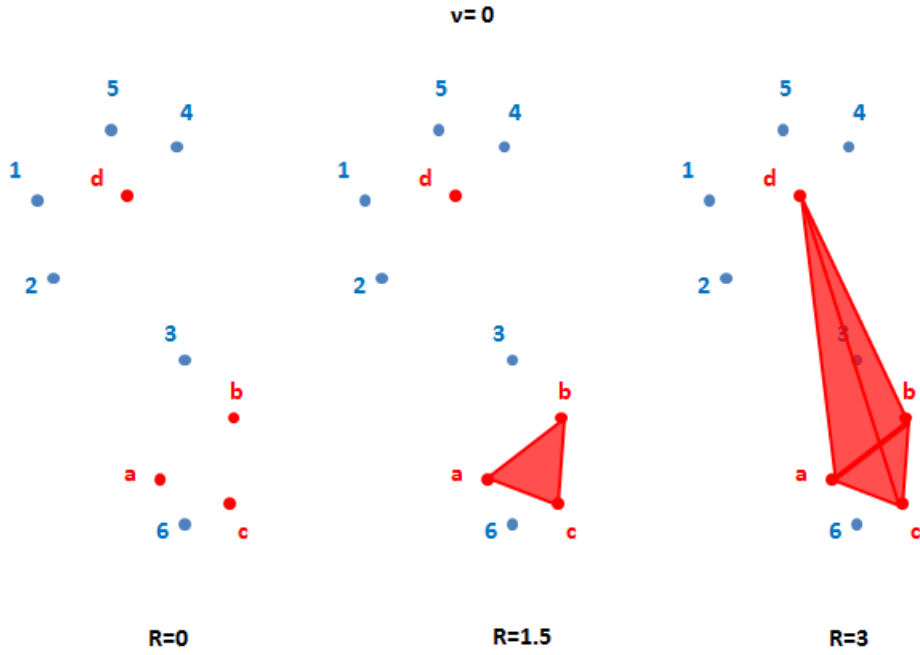


Figure 3.16: Three steps of the filtration for given PCD and landmarks a, b, c, d , where ν is considered to be 0; left: $R = 0$. The complex only consists on the four vertices a, b, c, d ; middle: $R = 1.5$ with complex formed by abc and d ; right: $R = 3$ with complex formed by abc , adc and acb .

3 An introduction to topological data analysis

We build the filtration considering $\nu = 1$. The only thing that changes with respect to the case where $\nu = 0$ is that the values m_i are now different from 0 and equal to the smallest entry of the i -column of D . Therefore $m_1 = 1.3$, $m_2 = 1.6$, $m_3 = 0.9$, $m_4 = 0.8$, $m_5 = 0.7$ and $m_6 = 0.8$.

(i) $R = 0 \rightarrow W(D; 0, 1)$ (Figure 3.17 left). i is a witness if

$$\max(D(a, i), D(b, i)) \leq m_i$$

holds. We analyze only the case of edges ab and ac .

-edge ab . $\max_{ab} = (4.2, 3.4, 1.2, 4.5, 5, 1)$ and the vector M that contains all values m_i is $M = (1.3, 1.6, 0.9, 0.8, 0.7, 0.8)$. As one can see $\max_{ab}(i) > M(i) \quad \forall i$. Therefore, ab does not exist in this step of the filtration.

-edge ac . The point 6 is a witness to the existence of ac since $\max_{ac}(6) = 0.8 \leq m_6 = 0.8$.

If one continues with the rest of edges, one concludes that they do not belong to $W(D; 0, 1)$.

(ii) $R = 1.5 \rightarrow W(D; 1.5, 1)$. One can check that all edges exist since for instance the vertex 3 is a witness for all of them. Therefore, all the 2D simplices as well as the tetrahedron also appear (Figure 3.17 middle).

(iii) $R = 3 \rightarrow W(D; 3, 1)$. The complex does not change (Figure 3.17 right).

It is also interesting to make some general observations about the three cases where $\nu = 0, 1, 2$.

$\nu = 0$: The filtration $W(D; R, 0)$ is very similar to the VR filtration noted by Rips($L; R$) and built over the same landmarks L . In particular, there are inclusions:

$$W(D; R, 0) \subseteq \text{Rips}(L; 2R) \subseteq W(D; 2R, 0)$$

$\nu = 1$: This family is very interesting since it can be understood as arising from a family of coverings of the space X by Voronoi-like regions surrounding each landmark point, which overlap increasingly as $R \rightarrow \infty$.

$\nu = 2$: there is a particular coincidence for $R = 0$ in this family:

$$W(D; 0, 2) = W(D)$$

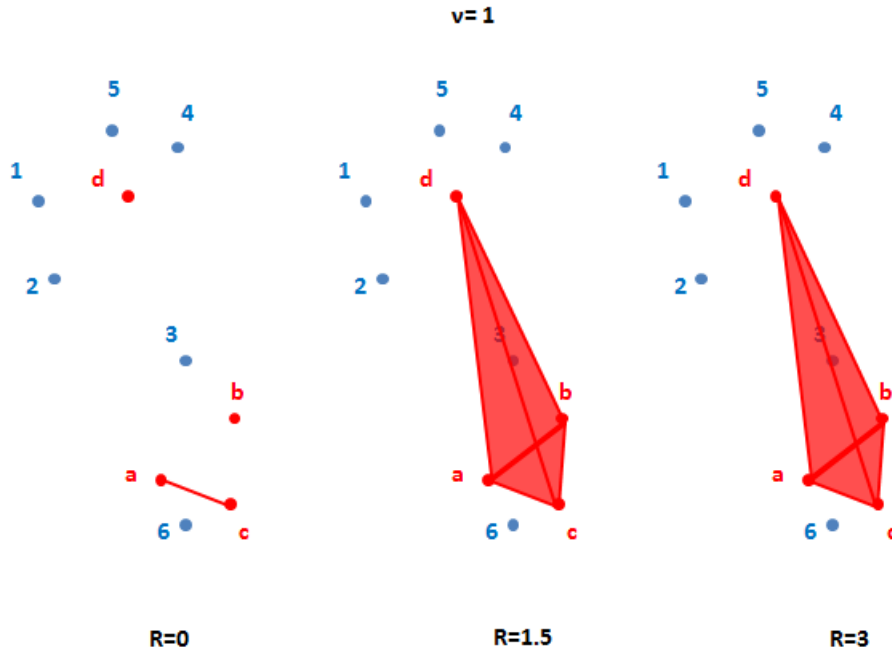


Figure 3.17: Analogous filtration to Figure 3.15 but considering $\nu = 1$

3.6.3.5 Example: Lazy witness filtration for points on S^2

We generated 1000 points randomly on S^2 as one can see in Figure 3.18. The aim is to recover the topology of the sphere. To this end, persistent homology is computed using a Lazy witness filtration. The software used is the Phom package for R [91]. The number of landmarks chosen is 40.

Results are shown in Figure 3.19, where one can see that there is one stable connected component (red dot in the upper left corner), no stable 1-dimensional holes (all green dots close to the diagonal) and one stable 2-dimensional hole (blue dot). This agrees with the betti numbers of S^2 which are $\beta_0=1$, $\beta_1=0$ and $\beta_2=1$.

Therefore, only with a sample of 40 landmark points out of 1000, one can construct a proper lazy witness filtration where the topology is well recovered. Indeed, the computational cost is much cheaper than using a Vietoris-Rips filtration that uses the 1000 points.

This computational advantage is not without a price: until now there are no explicit results about how faithfully a witness complex is to the topology of the underlying union of balls.

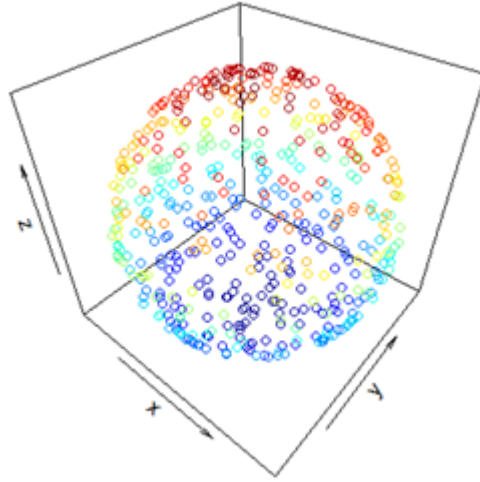


Figure 3.18: 1000 random points on S^2 .

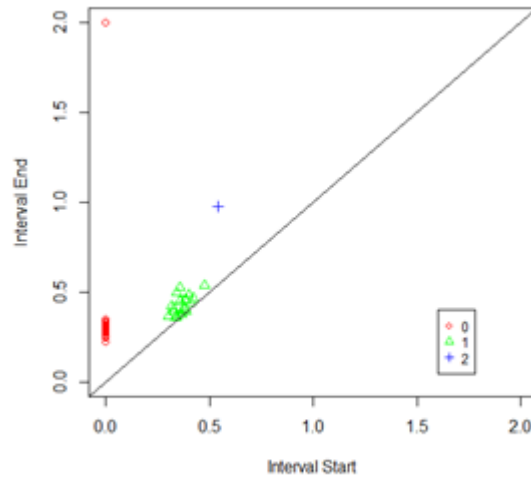


Figure 3.19: Persistent diagram corresponding to the persistent homology computation of 1000 points over S^2 . Red dots correspond to connected components. Green dots represent 1D holes and blue dots 2D.

4 Discrete Morse Theory

Almost every available software for computing persistent homology is based on discrete Morse theory. The computation of persistent homology is one of the key parts in topological data analysis since for huge complexes, computing homology is extremely expensive. This is precisely the problem that discrete Morse theory tries to overcome or at least reduce. It does so by reducing the size of the input complexes preserving the homotopy of them. Basically, it is a tool for determining equivalences between topological spaces arising from simplicial complexes. This way, a pre-processing step is performed before the homology computation. We will focus our attention in simplicial complexes, since this is the structure that we use.

4.1 Definitions

In this section we introduce some useful definitions that are used in the rest of the chapter. For all of these definitions, let K be a simplicial complex and τ , σ and γ simplices of K .

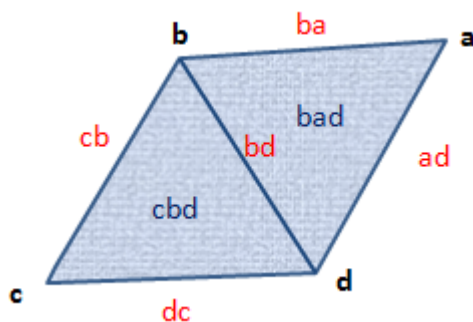


Figure 4.1: Simplicial complex

Definition 4.1. Immediate faces of a simplex τ are those faces σ such that $\dim(\sigma) = \dim(\tau) - 1$.

In the example of Figure 4.1, the simplicial complex K consists on four 0-dimensional simplices or vertices a, b, c, d , five 1D simplices or edges cb, ba, bd, ad, dc and two 2D simplices or triangles cbd, bad . If we consider the simplex $\tau = cbd$, then the faces are c, b, d, cb, bd and dc . The immediate faces are just cb, bd and dc .

Definition 4.2. We say that γ is a *coface* of a simplex τ if τ is a face of γ .

An immediate coface γ of τ is one that satisfies $\dim(\gamma) = \dim(\tau) + 1$. In the example of Figure 4.1, if we consider the simplex $\tau = a$, the cofaces are ba, ad and bad , but the immediate cofaces are just ba and ad .

Definition 4.3. We say a topological space X is contractible if the identity map on X is null-homotopic, which means that it is homotopic to a constant map. In a more informal way, X is contractible if it can be constantly shrunk to a point within the space (homotopic equivalent to a point).

Examples of contractible spaces are euclidean and complex spaces \mathbb{R}^n and \mathbb{C}^n . In Figure 4.2 more examples of contractible and non contractible spaces can be found.

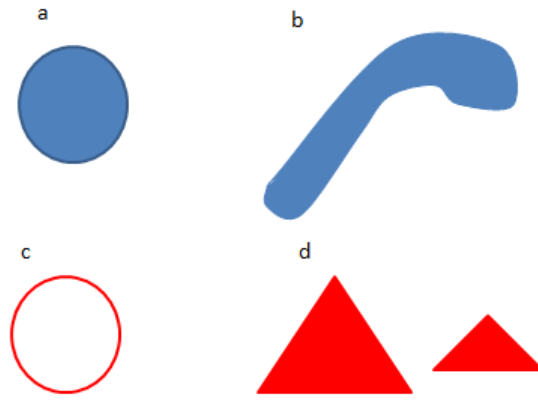


Figure 4.2: Contractible spaces (a and b) and non contractible spaces (c and d).

4.2 Basic elements of discrete Morse theory

There are two basic elements in discrete Morse theory from which everything is built. These are the *discrete Morse function* and the *critical simplices*.

Definition 4.4. Let K be a *simplicial complex*. Let $\tau^{(p)} \in K$ be a simplex of dimension p . Let $\gamma^{(p+1)}, \sigma^{(p-1)} \in K$ immediate cofaces and faces of τ respectively. Then we say that a function $f : K \rightarrow \mathbb{R}$ is a discrete Morse function if for every simplex $\tau \in K$, the two following conditions hold:

- (i) $\#\{\gamma > \tau | f(\gamma) \leq f(\tau)\} \leq 1,$
- (ii) $\#\{\sigma < \tau | f(\sigma) \geq f(\tau)\} \leq 1.$

(i) means that at most one immediate face σ of τ can have a value greater than or equal to that of τ and (ii) says that at most one immediate coface γ of τ can have a value less than or equal to that of τ . Basically, a discrete Morse function assigns values (real numbers) to every simplex existing in our simplicial complex. As stated in the previous definition, this function normally gives higher values to higher dimensional simplices.

To further explain this definition, we consider the example of Figure 4.3, where the shown simplicial complex is a triangle with its interior. In the left, a trivial Morse function is represented, where each simplex gets the value of its dimension, i.e., vertices or 0-dimensional simplices get the value 0, edges get 1 and the triangle abc gets 2. In the middle, another valid Morse function is shown. (i) and (ii) hold for every simplex. For example if we consider the 1D simplex $\tau = ab$, we see that $f(\tau) = 4$ and the values at the two 0-dimensional faces are $f(\sigma_1) = 1$ and $f(\sigma_2) = 6$. Therefore, (ii) holds. (i) also holds if we look at the coface abc .

On the other hand, in Figure 4.3 right, we show a non valid Morse function. The problem arises on the 0-dimensional simplex a , where $f(a) = 8$. This vertex, has two cofaces $\gamma_1 = ab$ and $\gamma_2 = ac$ where $f(\gamma_1) = 3$ and $f(\gamma_2) = 5$. Two cofaces cannot have smaller values due to (i) and therefore, this is not a discrete Morse function.

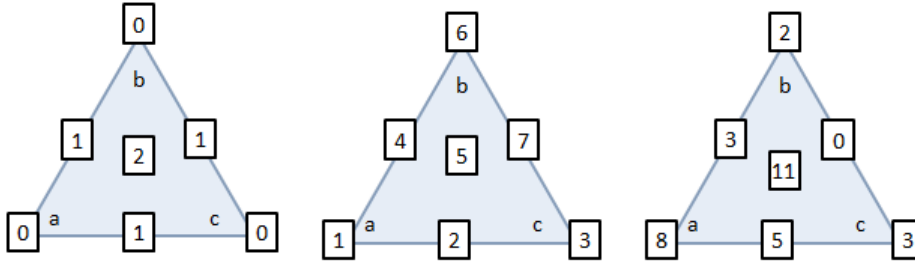


Figure 4.3: Left: trivial Morse function; middle: A valid Morse function; right: not a valid Morse function.

Definition 4.5. Let K be a *simplicial complex*. A simplex τ is critical if

$$(i) \quad \#\{\gamma > \tau | f(\gamma) \leq f(\tau)\} = 0$$

$$(ii) \quad \#\{\sigma < \tau | f(\sigma) \geq f(\tau)\} = 0$$

Basically, this means that if you consider a simplex τ , the function values must decrease when you move to the immediate faces and must increase when looking at the immediate cofaces. To further explain this, we refer to Figure 4.4. The simplices that are not critical are called *regular*.

Now we briefly introduce the notion of cell and CW complexes. For a more detailed explanation they reader may consult [65].

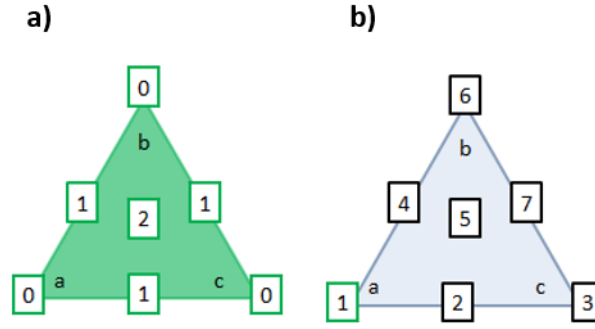


Figure 4.4: Critical simplices are drawn in green. a) All simplices are critical in a trivial Morse function; b) There is only one critical simplex.

Definition 4.6. A cell complex is a Hausdorff space X together with a specific cell decomposition of X .

Definition 4.7. A CW complex is a cell complex satisfying two extra conditions:

- (i) \forall n -dimensional open cell C in the partition of X , \exists a continuous map f from the n -dimensional closed ball to X s.t.
 - the restriction of f to the interior of the closed ball is a homeomorphism onto the cell C .
 - the image of the boundary of the closed ball is contained in the union of a finite number of elements of the partition, each having cell dimension less than n .
- (ii) A subset of X is closed if and only if it meets the closure of each cell in a closed set.

4.3 Main theorem of discrete Morse theory

Once we have defined what is a discrete Morse function and a critical simplex, we can state the main result of discrete Morse theory.

Theorem 4.8. [45] Suppose K is a simplicial complex with a discrete Morse function. Then K is homotopy equivalent to a CW complex with exactly one cell of dimension p for each critical simplex of dimension p .

The impact of this theorem is enormous since the size of a huge simplicial complex can be reduced substantially while keeping homotopy. The new size will be determined exclusively by the critical simplices of K . For example, the simplicial complex of Figure 4.4b

with the defined Morse function is homotopy equivalent to a vertex, since there is only one critical 0-dimensional simplex. We need some more definitions in order to be able to prove the previous theorem.

Definition 4.9. Let K be a simplicial complex. If one can find two faces α and β such that β is the only face that properly contains α , then α is a free face and $K_1 = K - \alpha - \beta$ is another simplicial complex with the same homotopy.

The removal of α and β is called an *elementary collapse* and K and K_1 are homotopy equivalent. The reverse process is called *expansion*. Therefore, we say that two simplicial complexes K and K_1 are *homotopy equivalent* if one can be reached from the other by a finite sequence of elementary collapses or expansions.

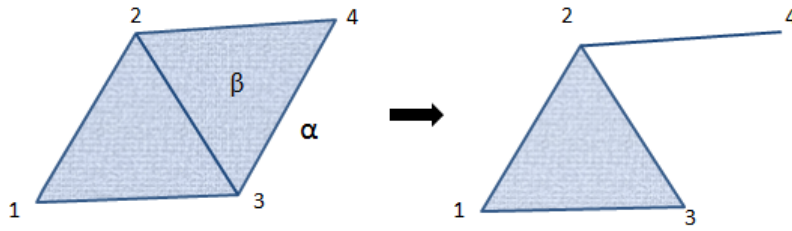


Figure 4.5: Schematic view of an elementary collapse in a simplicial complex K . K and K_1 are homotopy equivalent after the removal of the faces α and β .

Definition 4.10. A *collapsible* complex is a complex that can be reduced to a vertex by applying a finite sequence of elementary collapses.

An example is shown in Figure 4.5.

Lemma 4.11. Let K be a simplicial complex and f a discrete Morse function on K . Then there is always a minimum and it is a 0-face.

Proof. If a face F is a “minimum” and its dimension is ≥ 1 , then F at least contains two vertices x and y .

$$f(x) \leq f(F),$$

$$f(y) \leq f(F),$$

and this violates condition (ii) of definition 4 of a Morse function. \square

To illustrate this, we consider the example of Figure 4.6.

Lemma 4.12. Let K be a simplicial complex and f a discrete Morse function on it. Then there is always a maximum, either it is critical, in which case is a facet, or it consists of a pair (α, β) with α a free face of β .

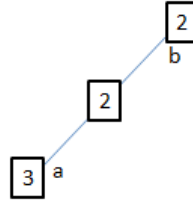


Figure 4.6: Simplicial complex where we assume that the simplex ab is the minimum. This cannot occur since it is not possible to obtain a valid Morse function.

Proof. The existence of a maximum can be easily proved since a map from a finite set attains its maximum.

If the maximum is critical (Figure 4.7ii), suppose by contradiction that it is not a facet but a face F (let G be a facet containing F). Then $f(F) \leq f(G)$ but then F is not critical.

If instead there are two faces α and β s.t. $f(\alpha) = f(\beta) = \max(f)$, then we are claiming that β is the only face that contains α . This can be seen in the example of Figure 4.7i where ab is the only edge that contains a . If we want to attach a new edge ac to a (Figure 4.7iii), this is actually not possible:

- if $f(ac) > 1$, then ab and a would not be the maximum anymore.
- if $f(ac) \leq 1$ then the function is not a valid Morse function anymore (violation of condition (ii) of definition 4 of Morse functions).

□

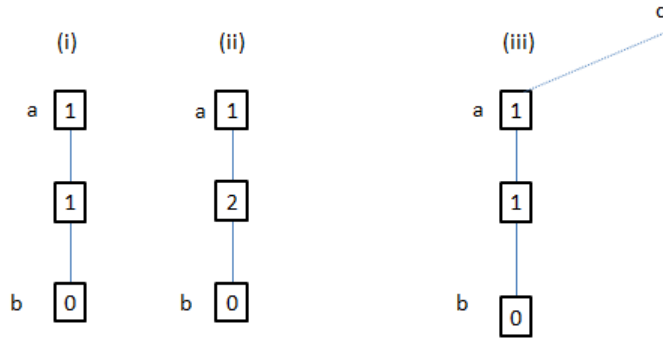


Figure 4.7: (i) and (ii): two different Morse functions applied on a simplicial complex; (iii): a is a free face of ab . A new edge (dashed line) cannot be added.

Definition 4.13. [45]. Let K be a simplicial complex with a discrete Morse function f on it. Let $c \in \mathbb{R}$ any real number. Then we define the level subcomplex $K(c)$ as the subcomplex consisting of all simplices τ of K s.t. $f(\alpha) \leq c$, and including all their faces.

To illustrate this, we refer to Figure 4.8. Finally, we state the last two lemmas that help us to prove the Theorem 4.8.

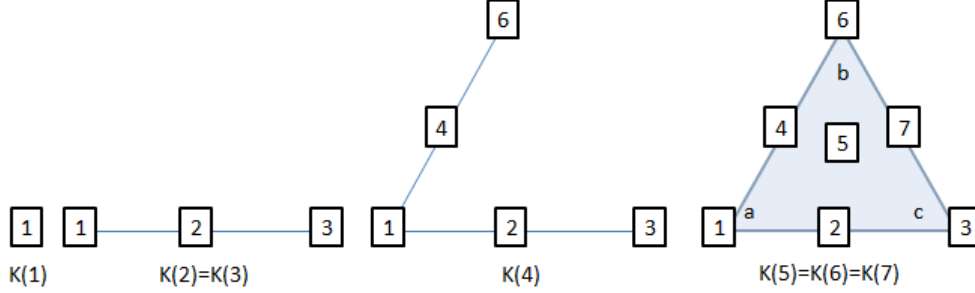


Figure 4.8: Different level subcomplexes for the shown discrete Morse function.

Lemma 4.14. [45]. Let K be a simplicial complex. If there are no critical simplices τ with $f(\tau) \in (a, b]$, then $K(b)$ is homotopy equivalent to $K(a)$.

Proof. Starting from the highest level, by Lemma 4.12 we know that if there is not a critical cell, we do have a pair of faces α and β s.t. α is a free face and therefore both faces can be removed while homotopy is maintained. As a consequence $K(b)$ collapses to $K(a)$ \square

Lemma 4.15. [45]. If there is a single critical simplex τ with $f(\tau) \in (a, b]$ then there is a map $F : S^{(d-1)} \rightarrow K(a)$, where d is the dimension of τ , such that $K(b)$ is homotopy equivalent to $K(a) \cup_F B^d$.

In fact, what this lemma is saying is that if you find a critical simplex, you can remove it but you have to keep track of it, attaching a d -ball at the end of your simplification process.

Here we just explain the ideas behind the proof. For the complete proof, the reader may consult [45]. If τ is a critical d -dimensional simplex, then by definition all its faces have smaller discrete Morse values. Because of this, they will appear in an earlier level subcomplex. As a consequence, when the critical simplex τ is added, it is being attached along the entire boundary, and this is equivalent to attach a d -cell.

Finally, we are ready to prove theorem 4.8.

Proof. By lemma 4.12, we know there exists a maximum that it is either critical or a pair of faces (α, β) with α a free face of β . Lemma 4.14 guarantees that homotopy is maintained when passing regular simplices and lemma 4.15 tells us that we must attach a cell of dimension p when a p -dimensional simplex is found. \square

In order to help the reader to understand the theorem 4.8 better, we consider two examples. In Figure 4.9i, we see that the only critical simplex is the vertex d . Therefore, theorem 4.8 states that K is homotopy equivalent to a vertex. We start looking for the maximum, that always exists according to lemma 4.12. We see that the maximum consists of a pair of faces ab and abd where ab is a free face of abd . Therefore by lemma 4.14., we can remove both faces and homotopy is maintained. The process is repeated in the next steps (Figure 4.9 ii, iii, iv and v) where free removable faces are marked in a red circle. Finally, after all elementary collapses, we see that the initial simplicial complex K is homotopy equivalent to a vertex (Figure 4.9vi).

The second example shown in Figure 4.10i corresponds to an empty polygon of five edges. It is a bit different since when we start looking for the maximum we see that corresponds to the 1D face bc and it is critical. Therefore according to Lemma 4.15, we can delete it and at the end of the simplification process attach a B_1 ball to the remaining simplicial complex. In Figure 4.10ii we see that b is a free face of ab and c is a free face of cd , so we can remove them according to lemma 4.14 and homotopy is maintained. In Figure 4.10iii the process is repeated (elementary collapses) and finally, in the last step, the B_1 ball is attached. The initial simplicial complex is homotopy equivalent to a circle and actually this is what theorem 4.8 was predicting.

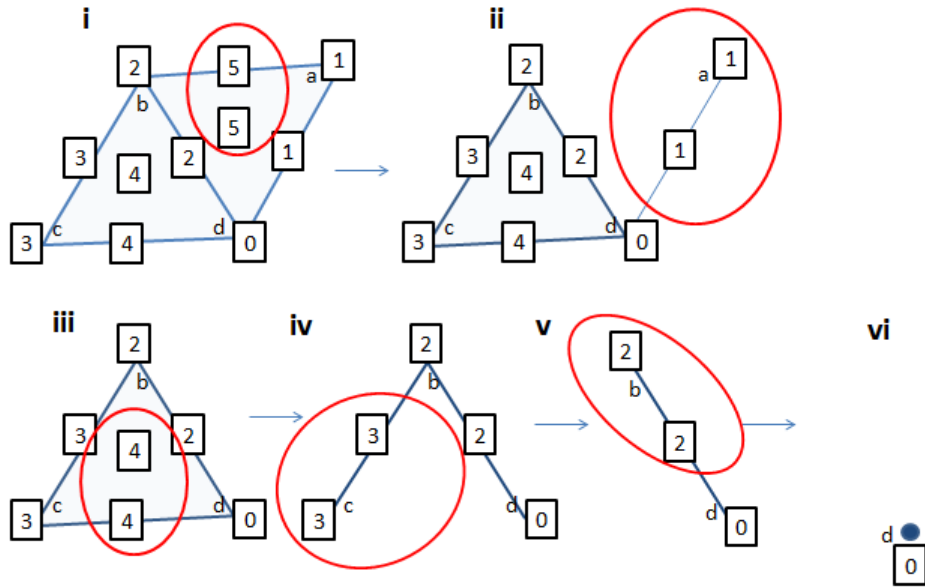


Figure 4.9: Reduction of the size of a simplicial complex maintaining homotopy. The initial simplicial complex is homotopy equivalent to a vertex.

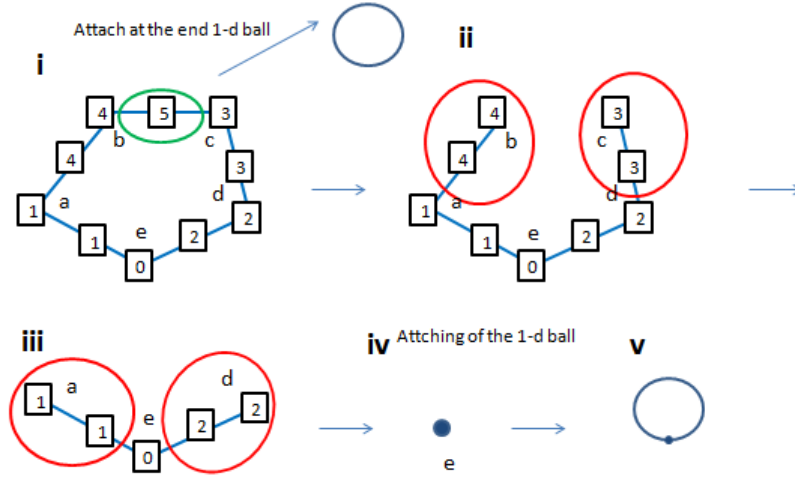


Figure 4.10: Reduction of the size of a simplicial complex maintaining homotopy. The initial simplicial complex is homotopy equivalent to a circle.

4.4 Morse inequalities

The Morse inequalities stated by Forman in [45] provide a bound for Betti numbers in a simplicial complex. One of the main results is:

Theorem 4.16. *For every $p = 0, 1, \dots, n$ being n the dimension of the simplicial complex K . Let m_p be the number of critical simplices in dimension p . Then*

$$m_p \geq b_p.$$

As explained in Chapter 3, the Betti numbers measure the number of connected components and dimensional holes in a simplicial complex. This theorem says that the Betti numbers for each dimension are bounded by the number of critical cells in that dimension. This underlines the importance of critical simplices and again it is clear that if we are able to construct a Morse function with less critical simplices, then the information about Betti numbers is maximized.

As it can be inferred from Figure 4.11, $m_0 = 3$ and $m_1 = 2$. Therefore, we know from the Morse function that the number of connected components is less or equal than three and that the number of 1D holes in K is less or equal than two:

$$\beta_0 \leq 3,$$

$$\beta_1 \leq 2.$$

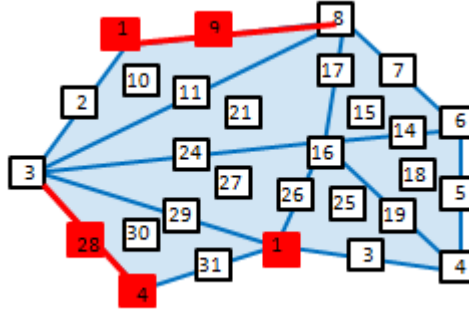


Figure 4.11: Simplicial complexes and a Morse function on it. Critical simplices are marked in red.

4.5 Gradient vector fields

Clearly trying to assign real numbers to all simplices in a simplicial complex and then trying to see if this corresponds to a Morse function, looks quite dubious from the efficiency point of view.

As we explained already, the Morse function is not the most important aspect of discrete Morse theory, but the associated critical simplices, since they provide a bound for the Betti numbers. Morse functions can be quite hard to compute, and in many cases, it will be enough to consider what it is known in the literature as the *gradient vector field of a Morse function*.

Definition 4.17. If $\tau^{(p)}$ is a regular simplex with $\tau^{(p)} < \gamma^{(p+1)}$ satisfying $f(\gamma) \leq f(\tau)$, we say that τ and γ are paired.

We can represent this paired simplices by an arrow from τ to γ . The diagram defined by all arrows is known as *gradient vector field*. The goal, rather than building a Morse function, is to create a gradient vector field that corresponds to a Morse function. How this is accomplished, will be explained in the rest of this section. To illustrate the concept of gradient vector field, we consider the simplicial complex of Figure 4.12. The following simplices are paired:

- b and ab since $b < ab$ and $f(b) = 6 > f(ab) = 4$.
- c and ac since $c < ac$ and $f(c) = 3 > f(ac) = 2$.
- bc and abc since $bc < abc$ and $f(bc) = 7 > f(abc) = 5$.

Basically, the gradient vector field indicates us where the function decreases while increasing the simplex dimensions. The corresponding gradient vector field of the Morse function shown in Figure 4.11 is displayed in Figure 4.13. Note that a simplex is critical if and only if, it is neither the tail nor the head of any arrow. It is possible to characterize a gradient vector field as a discrete vector field with some particularities.

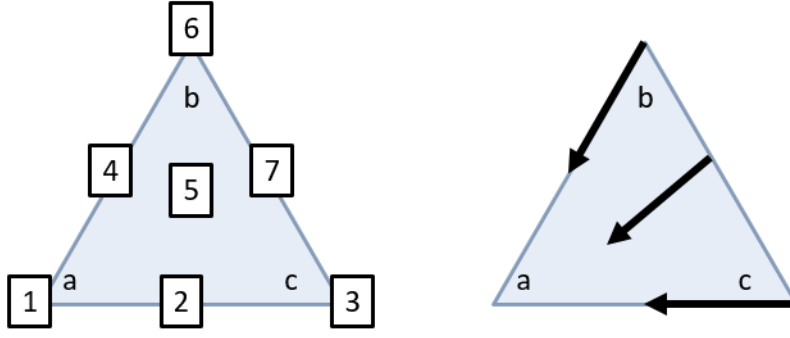


Figure 4.12: left: Morse function on the simplicial complex form by vertices a, b, c ; right: Associated gradient vector field.

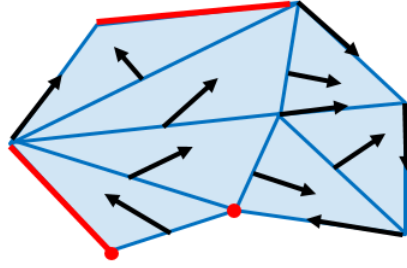


Figure 4.13: Gradient vector field that corresponds to the Morse function shown on Figure 4.11.

Definition 4.18. We say that a discrete vector field V on a simplicial complex K is a collection of pairs $\{\tau^{(p)} < \gamma^{(p+1)}\}$ of simplices such that each simplex is in at most one pair of V .

For example, in Figure 4.12 one discrete vector field V could be formed by the following pairs: $\{b, ab\}$, $\{a, ac\}$ and $\{bc, abc\}$. Note that in this definition we do not say anything about the Morse function values.

Definition 4.19. Let V be a discrete vector field on a simplicial complex K . Then a V -path is a sequence of simplices

$$\tau_0^{(p)}, \gamma_0^{(p+1)}, \tau_1^{(p)}, \gamma_1^{(p+1)}, \tau_2^{(p)}, \dots, \gamma_r^{(p+1)}, \tau_{r+1}^{(p)},$$

s.t. $\forall i = 0, \dots, r, \{\tau < \gamma\} \in V$ and $\gamma_i > \tau_{i+1} \neq \tau_i$. We say that the V -path is a *non-trivial closed path* if $r \geq 0$ and $\tau_0 = \tau_{r+1}$. If V is the gradient vector field of a discrete Morse function f , then this V -path is called a gradient path of f .

In Figure 4.14 left we show a simplicial complex given by the edges ab, bc, cd, da, de and ea . A V -path is shown in Figure 4.14 middle, with $\tau_0 = a, \gamma_0 = ab, \tau_1 = b, \gamma_1 = bc, \tau_2 = c, \gamma_2 = cd, \tau_3 = d$. All requirements of a V -path are satisfied.

In Figure 4.14 right, a non-trivial closed path is shown, where $\tau_4 = \tau_0$ and the requirements specified in the above definition are satisfied.

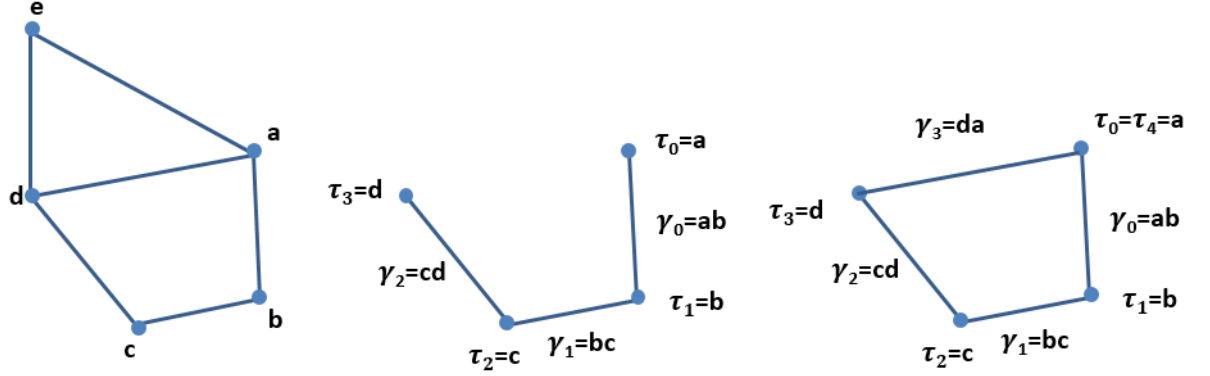


Figure 4.14: Left: simplicial complex; middle: V -path; right: non-trivial V -closed path

These definitions take us to the next result.

Theorem 4.20. [45]. *Let V be the gradient vector field of a discrete Morse function f . Then a sequence of simplices as those defined in definition 4.19, is a V -path if and only if $\tau_i < \gamma_i > \tau_{i+1}$ for*

$$f(\tau_0) \geq f(\gamma_0) > f(\tau_1) \geq f(\gamma_1) > \dots \geq f(\gamma_r) > f(\tau_{r+1}).$$

In Figure 4.15c and d we show two possible V -paths for the simplicial complex with associated Morse function shown in Figure 4.15a. Note that in Figure 4.15c, if we choose $\tau_0 = a$ as the first vertex of the path, then according to the previous theorem, $\gamma_0 = ab$ would not be a valid choice since $f(a) < f(ab)$. Instead the path $\{a, ac, c\}$ is valid. in Figure 4.15d, another valid path $\{ab, abc, ac\}$ is displayed where $f(ab) > f(abc) > f(ac)$.

Basically, the gradient path of f are sequences of simplices along which f is decreasing. What this theorem is also saying is that given V a gradient vector field, then there are no nontrivial closed V -paths. The main result in this field is actually that the converse is true:

Theorem 4.21. [45]. *A discrete vector field V is the gradient vector field of a discrete Morse function if and only if there are no non-trivial closed V -paths.*

Therefore, computational strategies are based in finding such gradient vector fields such that there are no non-trivial closed V -paths. This way, one makes sure that gradient

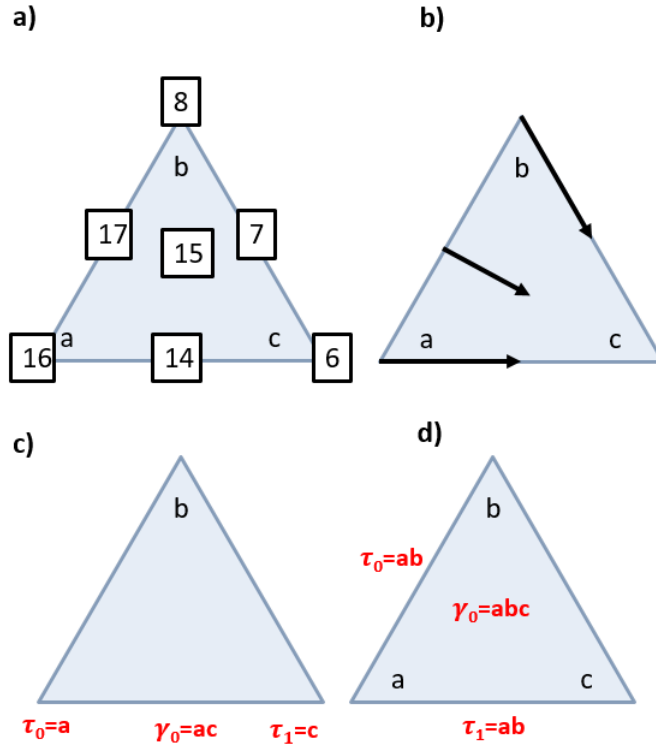


Figure 4.15: a) Simplicial complex with a valid Morse function on it; b) Associated gradient vector field; c) V -path; d) Another valid V -path.

vector field corresponds to a Morse function on the simplicial complex K . There is another possibility to characterize gradient vector fields and it is going to be discussed in the rest of this section.

It is also possible to describe a gradient vector field from a combinatorial point of view [45], and in these terms, discrete Morse can be reformulated.

Definition 4.22. Let K be a simplicial complex. The *Hasse Diagram* of K is a directed graph in which the simplices are ordered by the face relation. There is an edge between τ and γ if and only if $|\dim(\gamma) - \dim(\tau)| = 1$, i.e., if one is an immediate face of the other.

In Figure 4.16a, we illustrate this definition. Creating a Hasse diagram is like decomposing a human body in the different parts that it is made of. It is the same but working with simplicial complexes. We start from the highest level, i.e., the 2D simplex abc . This simplex can be decomposed in three dimensional faces ab, ac , and bc . Each of these faces is split into two vertices or 0-dimensional simplices. This is how the directed graph is created.

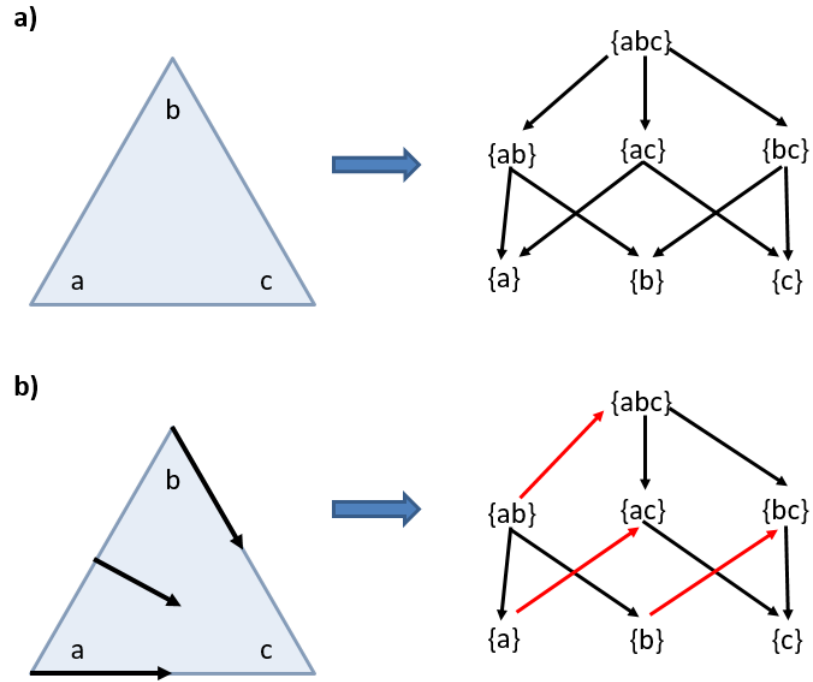


Figure 4.16: a) Triangle with its interior is the considered simplicial complex. In the right, the associated Hasse diagram; b) Modification of the Hasse diagram according to the shown gradient vector field.

Then the Hasse diagram is modified in the following way: the direction of one arrow is reversed where the gradient vector field indicates it. In the example of Figure 4.16, one can see that there are arrows in the gradient vector field $\{b, bc\}$, $\{a, ac\}$, and $\{ab, abc\}$, so the arrows are reversed in red. A V -path can be seen as a directed path in this modified graph. This leads us to the next result.

Theorem 4.23. [45]. *There are no nontrivial closed V -paths if and only if there are no nontrivial closed directed paths in the corresponding directed Hasse diagram.*

Therefore, we can see a discrete vector field as a partial matching of the Hasse diagram, and this discrete vector field is a gradient vector field corresponding to a Morse function if the modified Hasse Diagram contains no cycles. Finally, we can reformulate the main theorem of discrete Morse Theory in the following way:

Theorem 4.24. [45]. *Let V be an acyclic partial matching of the Hasse diagram of K (of the kind described above — assume that the empty set is not paired with another simplex). Let u_p denote the number of unpaired p -simplices. Then K is homotopy equivalent to a CW-complex with exactly u_p cells of dimension p , for each $p > 0$.*

4.6 Computing persistent homology

Several concepts have been presented in Chapters 2, 3 and 4. The reader should pay special attention to the followings:

- How to build different types of filtrations such as Čech, Rips, Alpha or Lazy witness.
- Discrete Morse theory, where ideas have been shown for pre-processing a simplicial complex in order to reduce its size and save computational time.
- How to compute homology given a simplicial complex.
- How to encode persistent homology information in barcodes or persistent diagrams.

But we have not explained yet how the algorithm for computing persistent homology works given a filtration. Once we have preprocessed the simplicial complex with ideas from discrete Morse theory, how do we compute the barcode representation given a filtration? This question is answered in the present section. We first give the algorithm [11] and later, we show an illustrative example.

We assume that we have a filtration that goes simplex by simplex or cell by cell. In discrete Morse theory this means that we assume that every cell is critical in f . This is a reasonable assumption since we can always reach that by a small perturbation ϵ , see Figure 4.17. The perturbation is performed to create some order in the filtration. Once this is done, one can forget about ϵ .

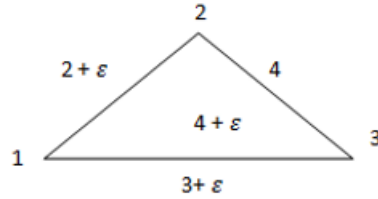


Figure 4.17: Morse function where each cell is critical.

Let D be the matrix of boundary maps $\partial_* : C_* \rightarrow C_{*-1}$. This matrix can be split according to the different dimensions. We define the pivot $pivot_R(j)$ of some matrix R as the largest row index in column j with $R_{ij} \neq 0$. The algorithm works as follows:

Algorithm 1 Persistent homology computation algorithm

$R=D$ **for** $j=1$ to n

while $\exists j_0 < j$ with $pivot(j_0) = pivot(j)$

 add column j_0 to column j ($1+1=0; 1+0=1; 0+0=0$)

4 Discrete Morse Theory

Let us state some properties of this reduced matrix:

- Column j in D is a cycle (the boundary of a cell σ_i in K_{j-1}).
- Adding “older” columns does not change homology class in K_{j-1} , because we only add columns that correspond to boundaries of earlier complexes.
- Reducing the matrix means minimizing $pivot(j)$ and this is just finding the “earliest” representative cycle.
- Column j in R non-zero $\rightarrow pivot_R(j) = i$, $\rightarrow z_i$ is not homologous in K_{j-1} to any earlier cycle ($[Z_i] \notin im(h_{i-1,i})$).
- $[z_i] = 0$ in j (boundary in K_j).

Another important remark is related with the gradient vector field of a Morse function. Remember that each of those arrows represent two pair of simplices “paired”. It turns out that if (σ_i, σ_j) is a pair in the discrete gradient of f , then $i = j - 1$ and it will be a persistence pair.

The explanation is that σ_i is the cell with highest index in $\partial\sigma_j$. Moreover, σ_j is the cell with highest index having σ_i in its boundary (there is no way to reduce the pivot) and column j in D is already reduced.

All these remarks and definitions can look a bit confusing. That is why we consider the following example in order to help the reader to understand the algorithm better. In Figure 4.18 a filtration of a tetrahedron is depicted.

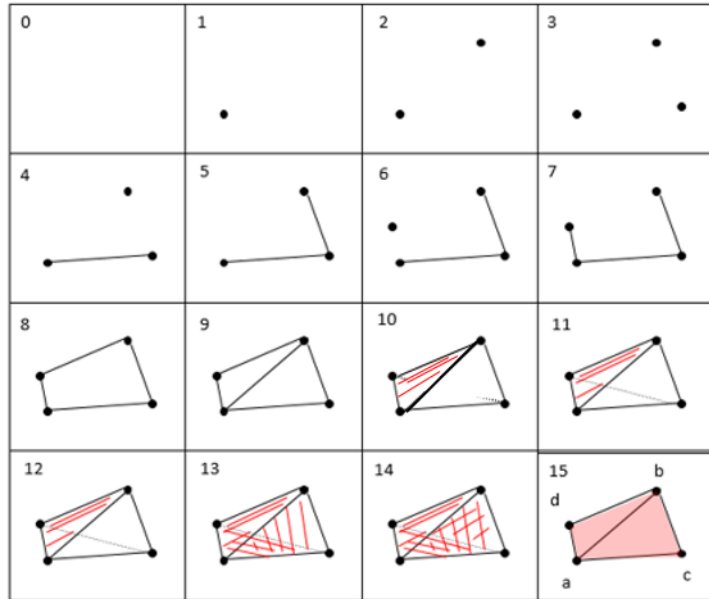


Figure 4.18: Filtration of the tetrahedron a, b, c, d .

The tetrahedron is filtered simplex by simplex. We proceed as follows:

- (i) Write the boundary matrices.
- (ii) Reduce them using the proposed algorithm. In this way, one is able to recover the persistence pairs for this filtration.

Across the filtration, we have 0, 1, 2 and 3D simplices. Therefore, we can write the boundary matrices for each of these dimensions and split the computation of persistence pairs separately. Let us start with dimension 0.

$$D_0 = \begin{matrix} & a & b & c & d \\ (0 & 0 & 0 & 0) \end{matrix}$$

On the top of the matrix D_0 we write the name of the 0-dimensional simplices. All of them have no boundary so the 0-dimension is not interesting. Now we compute the matrix of boundaries for the first dimension:

$$D_1 = \begin{matrix} & ac & bc & ad & bd & ab & cd \\ a & \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} & 1 \\ b & \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} & 2 \\ c & \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} & 3 \\ d & \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} & 6 \\ & 4 & 5 & 7 & 8 & 9 & 11 \end{matrix}$$

This matrix is built as follows: On top of the matrix we name the 1D simplices. These are: ac, bc, ad, bd, ab and cd . In every row we represent the 0-dimensional simplices a, b, c, d (vertices are the boundaries of edges). For every column, we write 1 if that 0-dimensional simplex is contained in the 1-dimensional simplex analyzed. For example in the first column, we write 1 in the first and third row because a and c are the boundaries of ac .

On the bottom of the matrix we show indices of the filtration where the edges appear (or equivalently the Morse function values). ac appears in step 4, and bc, ad, bd, ab, dc in steps 5, 7, 8, 9, 11 respectively. The same is displayed on the right, where indices of the filtration where the vertices appear are displayed (a, b, c, d in 1, 2, 3, 6 respectively).

We continue reducing this matrix according to the proposed algorithm. We go column by column trying to reduce the pivot index by adding a column with a lower index as long as one can do that.

$$D_1 = \begin{matrix} & ac & bc & ad & bd & ab & cd \\ a & \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} & 1 \\ b & \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} & 2 \\ c & \begin{pmatrix} \mathbf{1} & \mathbf{1} & 0 & 0 & 0 & 1 \end{pmatrix} & 3 \\ d & \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} & 6 \\ & 4 & 5 & 7 & 8 & 9 & 11 \end{matrix} \longrightarrow D_1^* = \begin{matrix} & ac & bc & ad & bd & ab & cd \\ a & \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} & 1 \\ b & \begin{pmatrix} 0 & \mathbf{1} & 0 & 1 & 0 & 0 \end{pmatrix} & 2 \\ c & \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 1 \end{pmatrix} & 3 \\ d & \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} & 6 \\ & 4 & 5 & 7 & 8 & 9 & 11 \end{matrix}$$

The first column can never be reduced. We look at the second column and see that the pivot is in the third row (marked in red, D_{23}). One can observe that the earlier column has the same pivot and therefore according to the algorithm, we need to add the first column to the second: (1 and 0 will result in a 1, 0 and 0 in 0, 1 and 1 in 0). D_1^* is the result from this operation. The new pivot is in position D_{22} and this column cannot longer be reduced since

$$\text{pivot}(D(,1)) \neq \text{pivot}(D(,2)).$$

We keep checking the rest of the matrix. The pivot of the third column is D_{43} and cannot be reduced. The pivot for the fourth column is D_{44} . Both pivots are marked by red in D_1^{**} . Since $\text{pivot}(\text{column}3) = \text{pivot}(\text{column}4)$, we add the third column to the fourth, resulting in D_1^{***} .

$$D_1^{**} = \begin{array}{c} \begin{array}{cccccc} & ac & bc & ad & bd & ab & cd \\ a & \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} & 1 \\ b & \begin{pmatrix} 0 & \color{red}{1} & 0 & 1 & 0 & 0 \end{pmatrix} & 2 \\ c & \begin{pmatrix} \color{red}{1} & 0 & 0 & 0 & 0 & 1 \end{pmatrix} & 3 \\ d & \begin{pmatrix} 0 & 0 & \color{red}{1} & \color{red}{1} & 1 & 1 \end{pmatrix} & 6 \\ & 4 & 5 & 7 & 8 & 9 & 11 \end{array} \end{array} \longrightarrow D_1^{***} = \begin{array}{c} \begin{array}{cccccc} & ac & bc & ad & bd & ab & cd \\ a & \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} & 1 \\ b & \begin{pmatrix} 0 & \color{red}{1} & 0 & \color{red}{1} & 0 & 0 \end{pmatrix} & 2 \\ c & \begin{pmatrix} \color{red}{1} & 0 & 0 & 0 & 0 & 1 \end{pmatrix} & 3 \\ d & \begin{pmatrix} 0 & 0 & \color{red}{1} & 0 & 1 & 1 \end{pmatrix} & 6 \\ & 4 & 5 & 7 & 8 & 9 & 11 \end{array} \end{array}$$

The new pivot of column 4 is D_{24} and can be still reduced. Adding the second column to the fourth we reach D_1^{****} . Repeating the procedure with the third and fifth columns, third and sixth and first and sixth, we finally obtain the reduced matrix R , which cannot be longer reduced.

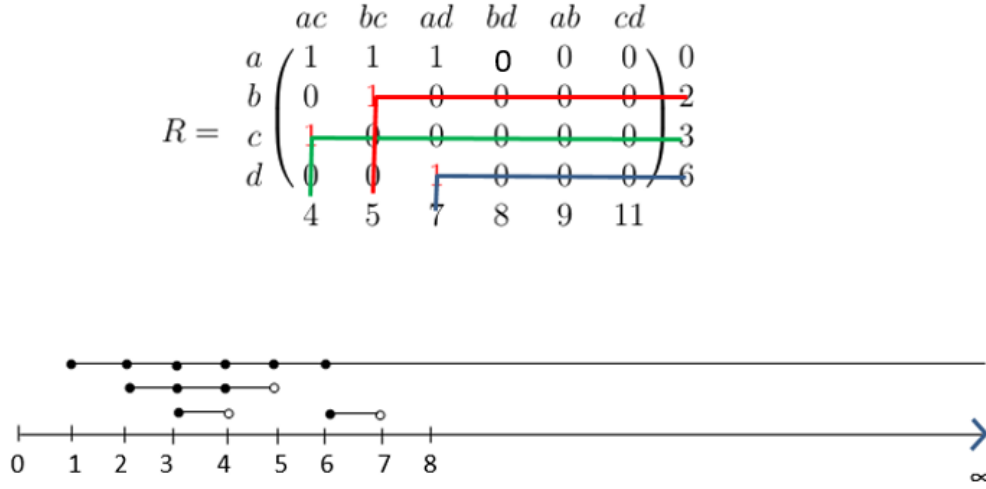
$$D_1^{****} = \begin{array}{c} \begin{array}{cccccc} & ac & bc & ad & bd & ab & cd \\ a & \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} & 1 \\ b & \begin{pmatrix} 0 & \color{red}{1} & 0 & 0 & 0 & 0 \end{pmatrix} & 2 \\ c & \begin{pmatrix} \color{red}{1} & 0 & 0 & 0 & 0 & 1 \end{pmatrix} & 3 \\ d & \begin{pmatrix} 0 & 0 & \color{red}{1} & 0 & \color{red}{1} & 1 \end{pmatrix} & 6 \\ & 4 & 5 & 7 & 8 & 9 & 11 \end{array} \end{array} \longrightarrow R = \begin{array}{c} \begin{array}{cccccc} & ac & bc & ad & bd & ab & cd \\ a & \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} & 1 \\ b & \begin{pmatrix} 0 & \color{red}{1} & 0 & 0 & 0 & 0 \end{pmatrix} & 2 \\ c & \begin{pmatrix} \color{red}{1} & 0 & 0 & 0 & 0 & 0 \end{pmatrix} & 3 \\ d & \begin{pmatrix} 0 & 0 & \color{red}{1} & 0 & 0 & 0 \end{pmatrix} & 6 \\ & 4 & 5 & 7 & 8 & 9 & 11 \end{array} \end{array}$$

The next step is to depict the barcode for H_0 , since we analysed the boundary matrix for dimension 1. From the reduced matrix R it is possible to read the persistence pairs. In this case, they are $\{2, 5\}$, $\{6, 7\}$ and $\{3, 4\}$. They are represented in the barcode (Figure 4.19).

One can compute the Betti numbers at each step of the filtration just counting the number of basis elements that intersect with a vertical line in the barcode. For instance

$$\beta_0(4) = 2.$$

This barcode keeps track of the birth and death of number of connected components in the tetrahedron filtration of Figure 4.18. In the next section, we consider a slight variation of the barcodes, the persistent diagrams and we also analyze their stability.


 Figure 4.19: Barcode representation in H_0 .

4.7 Stability of persistent diagrams

Let f be a discrete Morse function. We assume that every cell of f is critical. Consider the pairs (v_i, v_j) being $v_i = f(\sigma_i)$ and $v_j = f(\sigma_j)$ for all persistence pairs (i, j) . We consider σ_i an essential cell (it creates homology that it is never killed by another cell) and we allow ∞ as a possible value. Then one can plot the persistent diagram of the function f

$$Dgm(f) \subset \bar{\mathbb{R}}^2.$$

In Figure 4.20 we show one function f and a small perturbation g of the function f is considered such that

$$\|f - g\|_\infty \leq \delta.$$

Some persistence pairs AB and CD of f are represented in a persistent diagram in the right side. This section raises the following question: if given f and g , are the persistent diagrams related somehow? A metric needs to be defined for comparison between persistent diagrams.

Definition 4.25. Let $X, Y \subset \bar{\mathbb{R}}^2$. The bottleneck distance $d_B(X, Y)$ is defined as the infimum over all bijections $\gamma : X \rightarrow Y$ of the

$$\sup_{x \in X} \|x - \gamma(x)\|_\infty$$

We define some operations in this metric. Let $a, b \in \mathbb{R}$. Then

- $(a, \infty) - (b, \infty) = a - b$

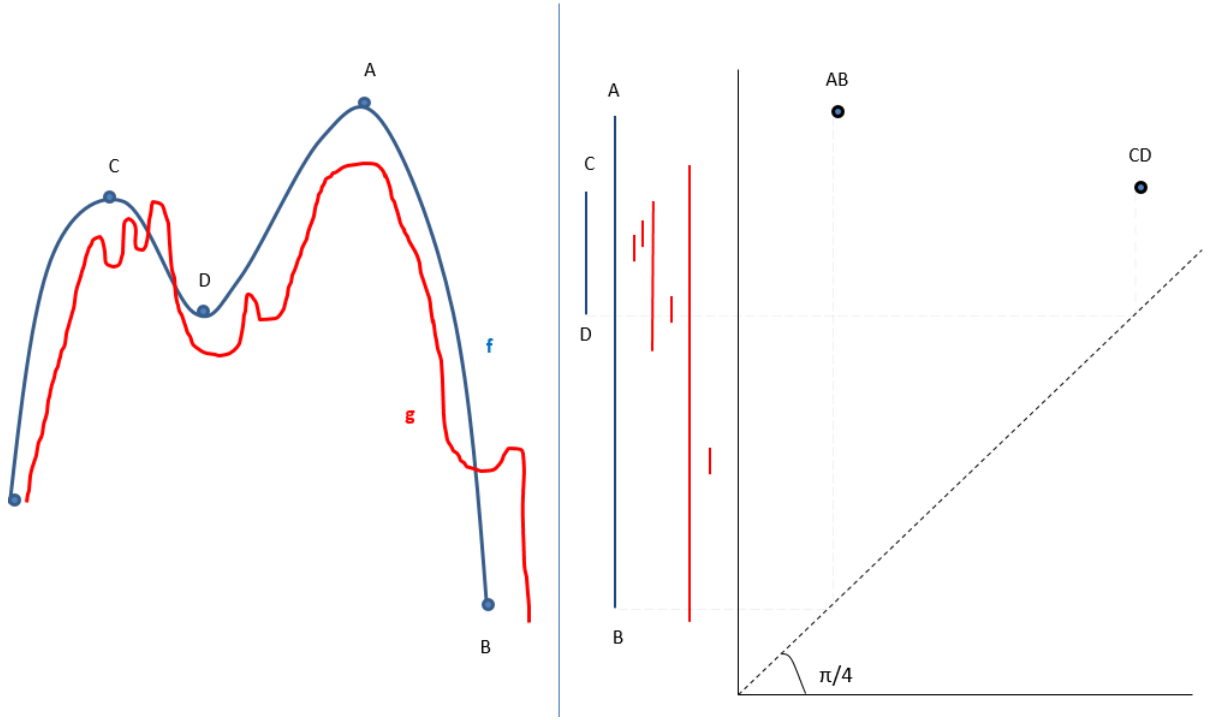


Figure 4.20: Morse function f and perturbation g . In the right side some persistent pairs represented in a persistent diagram. Given f and g , are the persistent diagrams of both related?.

- $(a, \infty) - (b, c) = \infty$

These definitions allow us to derive the main theorem about stability of persistent diagrams, which can be proved using discrete Morse theory.

Theorem 4.26. [28] *Let f, g be two discrete Morse functions. Then*

$$d_B(Dgm(f), Dgm(g)) \leq \|f - g\|_\infty$$

Proof. The idea is to interpolate linearly between f and g . Let $f_t = (1 - t)f + tg$, which is also a discrete Morse function for every t . The order of the function f_t changes at a finite number values $0 < t_1 \dots < t_k < 1$. This means within an interval $[t_i, t_{i+1}]$ we can assume a fixed order of cells induced by the functions. Let p_i be a set of persistence pairs for this family of functions. Then for each pair $r, s \in [t_i, t_{i+1}]$

$$\begin{aligned} d_B(Dgm(f_r), Dgm(f_s)) &\leq \max_{(\sigma, \tau) \in p_i} \|(f_r(\sigma), f_r(\tau)) - (f_s(\sigma), f_s(\tau))\|_\infty \leq \|f_r - f_s\|_\infty \\ &= |s - r| \|f - g\|_\infty. \end{aligned}$$

By the triangle inequality

$$\begin{aligned} d_B(Dgm(f), Dgm(g)) &\leq \sum_{i=0}^{k-1} d_B(Dgm(f_{t_i}), Dgm(f(t_{i+1}))) \\ &\leq \sum_{i=0}^{k-1} \|t_{i+1} - t_i\|_\infty = \|f - g\|_\infty. \end{aligned}$$

□

This theorem has interesting implications regarding the number of critical points of discrete Morse functions, because every persistence pair can only move as much as the perturbation of the function.

Corollary 4.27. *Let f be a discrete Morse function and $\delta \geq 0$. Then for every Morse function f_σ with $\|f - f_\sigma\| \leq \delta$, the number of critical points of f_σ is greater or equal than the number of critical points of f with persistence less or equal than 2δ .*

Proof. Persistence pairs in the discrete gradient of f V_f are pairs with persistence 0, and by stability theorem we know that this corresponds to pairs with persistence less or equal than 2δ . □

5 Classification methods

In Chapter 1, ultrasonic NDT, the TOFD method and the most common welding defects were presented. Chapter 2 showed the state of the art of automatic classification of TOFD defects. Then, in Chapters 3 and 4, we learned the basics of persistent homology and discrete Morse theory. Once these concepts have been discussed, we can present the methods for the automatic classification developed in the present research work. Therefore, in this chapter, two automated methods are presented for classification of TOFD defects:

- The first method is able to differentiate between defects from different classes, in particular volumetric and planar defects. This technique is based on ideas from persistent homology.
- The second method does the same job as method 1, but it is also able to distinguish between same type of defects that present slight different curves in the b-scan. This technique requires a complete noise removal of the b-scan and it is a shape classification method based on tangential homology, which is persistent homology applied to some suitable tangential constructions. Every defect is represented as a curve PCD in the b-scan and the method extracts a compact shape descriptor for every defect [29].

5.1 Classification method 1: persistent homology and TOFD defects

In this section, the first classification method is presented. We show briefly how ideas from persistent homology can be applied for detection and classification of TOFD indications, in particular volumetric (pores, wormholes, etc) and planar defects (cracks and lack of fusion).

Before getting into the method details, we show an schematic view (Figure 5.1) in order to understand the general procedure of this method. Imagine one company has manufactured one steel tube and then they want to know if there is a defect inside and if that is the case, find out the type. This method operates as follows:

- (i) A TOFD setup is installed and measurements are taken in the inspected weld.
- (ii) A b-scan is obtained, where in case the weld contains any defect, it will appear at the b-scan between the lateral wave and backwall echo.
- (iii) A sparse deconvolution method is used to reduce the amount of data and to denoise the image. Lateral wave and backwall echo are also filtered out.

Classification method 1: persistent homology and TOFD defects

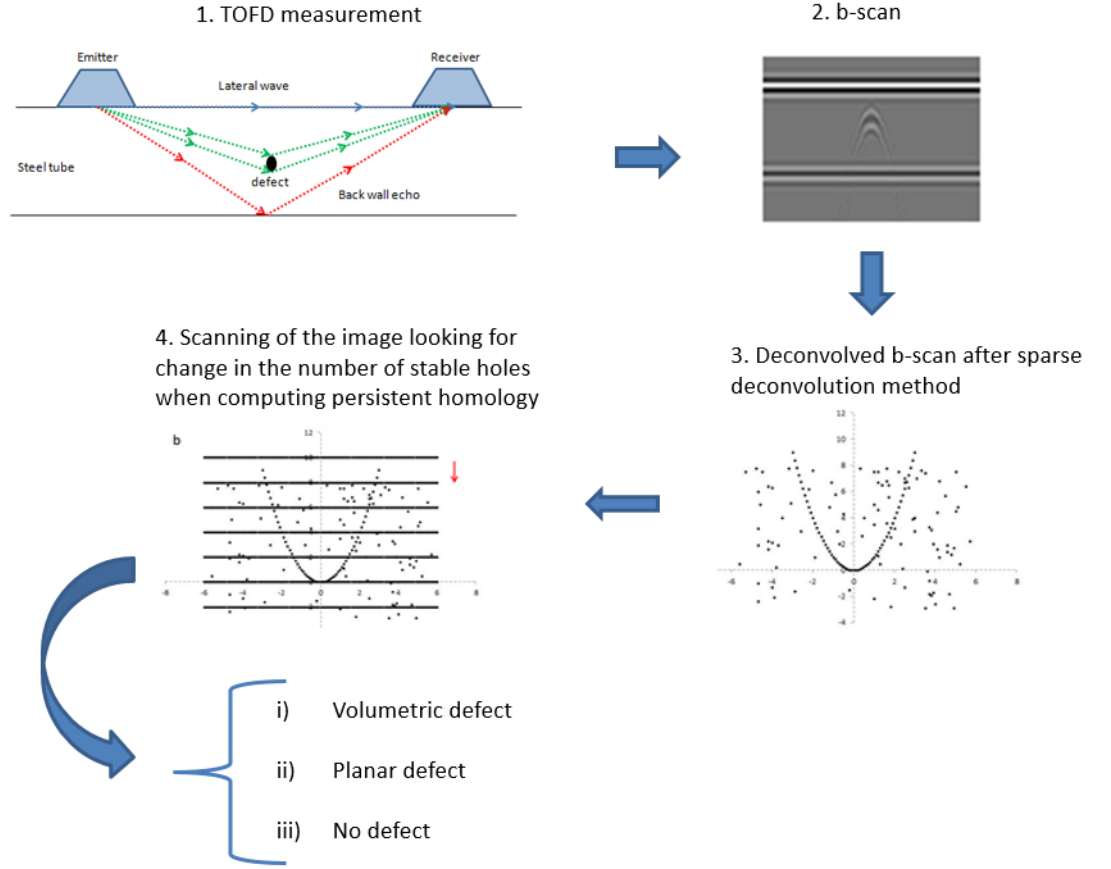


Figure 5.1: Scheme of classification method 1.

- (iv) A scan of the image is performed. Extra suitable data depending on the defect is added and persistent homology is computed. The defects will be related to the existence of stable holes in the PCD.
- (v) The defect type is found.

Pore defects present a parabolic shape in deconvolved b-scans (Figure 5.2a, where the lateral wave and backwall echo signals have been filtered out). The idea for detecting such parabolic features surrounded by noise is a scanning method based on persistent homology. The procedure works as follows:

- (i) An horizontal line of points is added (Figure 5.2b for $y = 10$) at the top of the image.
- (ii) Persistent homology is computed for the initial PCD plus the added line of points.

5.1 Classification method 1: persistent homology and TOFD defects

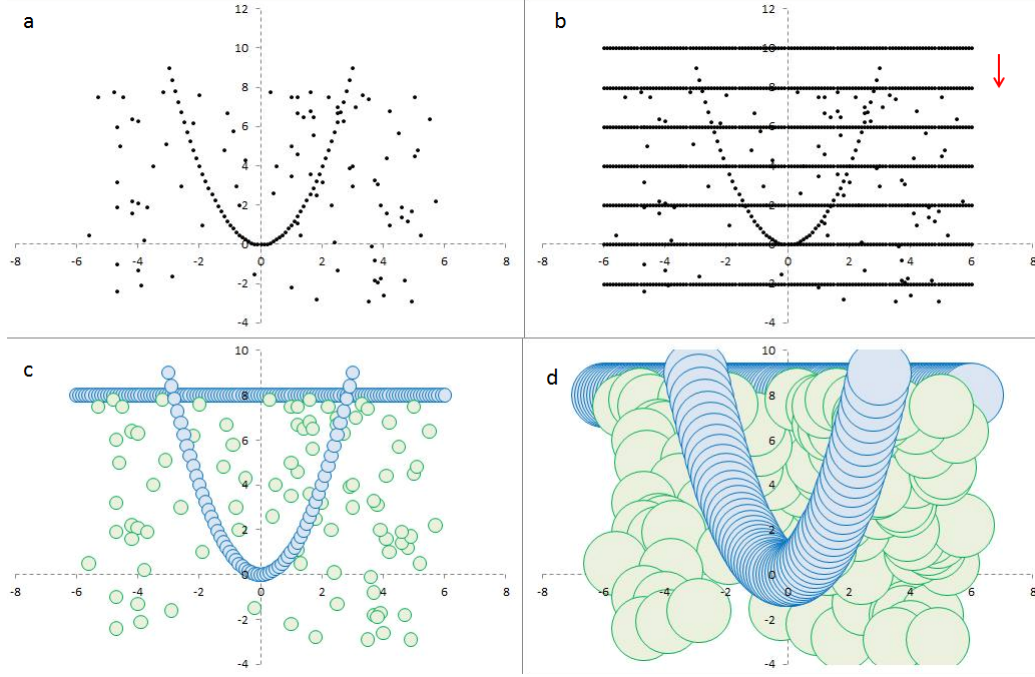


Figure 5.2: (a) Deconvolved b-scan including a parabolic defect; (b) a horizontal line is added and moved down; (c) small radii are needed to create a hole between the line and the parabola; (d) a large radius is needed to eliminate the hole [32].

- (iii) The line is moved down and the process is repeated until the whole image has been scanned.

When we start building the filtration by increasing the radii of the balls around every vertex, a *stable* 1D hole will be created very soon (Figure 5.2c,d) in cases where the parabola intersects with the line. This hole will be detected by persistent homology. As soon as the line surpasses the parabola (Figure 5.2b, for $y = -2$ or $y = 10$), no stable hole will be detected. Summarizing, one-dimensional stable holes are generated by pore defects, whereas unstable holes are due to noise.

Crack defects present a linear shape in deconvolved b-scans. Figure 5.3 shows one example, where the crack is marked by blue dots. To detect the crack, instead of adding lines, we add data points lying on a rectangle, see Figure 5.3 left. A scan of the image is performed from left to right. When the rectangle does not intersect with the crack (e.g. Figure 5.3, left), persistent homology detects only one big stable hole created in the rectangle. But as soon as the rectangle arrives at the crack indication (Figure 5.3 right), two stable holes will be detected by persistent homology. In conclusion, changes in the number of stable holes are indicators of the existence of pore or crack defects depending on the extra data added.

One important remark is that the rectangles of points shown in Figure 5.3 are asym-

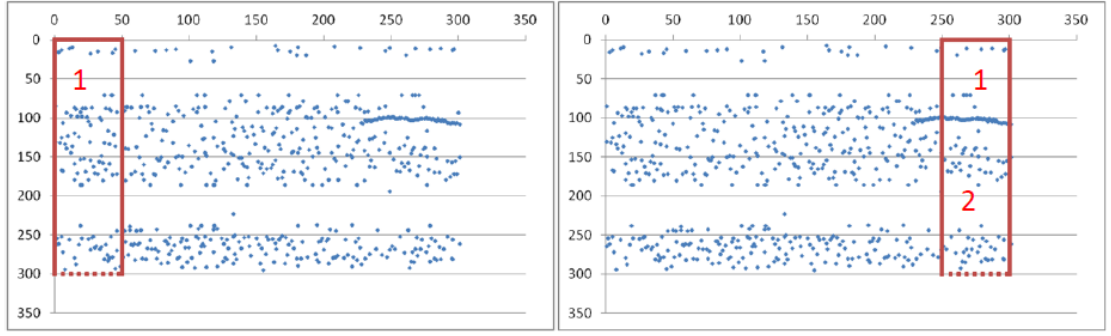


Figure 5.3: Two steps with adding points on rectangle. Left: one stable generator is detected; Right: two stable generators are detected by persistent homology [32].

metric (in our case points in the lower part of the rectangle are closer than in the upper part and in the laterals). This is done in order to visualize the results in a clearer way. If the rectangle is symmetric, then the 2-stables holes shown in Figure 5.3 right, would appear and disappear at the same time, which means that we would see just one point in the persistent diagram. This is illustrated in Figure 5.4.

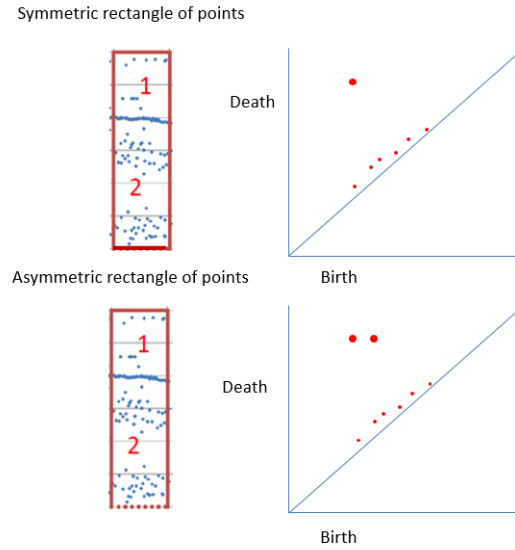


Figure 5.4: If the rectangle of points is symmetric (distance between every point is the same), the two created stable holes when computing persistent homology will be represented as just one point in the corresponding persistent diagram.

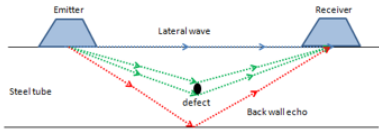
5.2 Classification method 2: tangential homology and TOFD defects

This method is divided into two parts. As in the previous section, we first summarize in Figure 5.5 and 5.6 the general procedure of the method before analyzing the details.

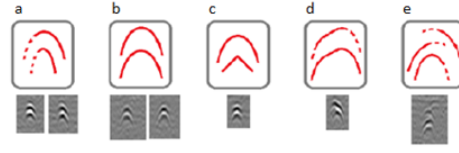
In the first part (Figure 5.5) we create a catalogue of the most common defects that appear in TOFD measurements. To this end, several hundreds of weld were tested and the different indications were grouped by the shapes presented in the b-scans [33]. For each of those shapes, a barcode shape descriptor was computed using ideas from tangential homology.

Classification method 2. i) creation of a catalogue of barcode shape descriptors

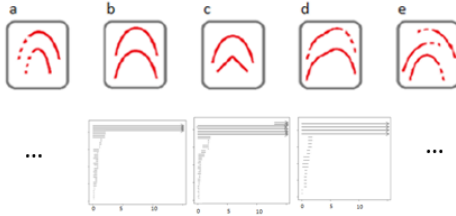
1. With a TOFD setup several hundreds of weld seams are tested and evaluated



2. Indications can be grouped by similar appearances of flaws in the TOFD Scan.



4. Each TOFD defect has an associated barcode which is representative of its shape.



3. For each of those shapes, tangent complex $T(X)$, approximated $T(X)$ and approximated $T(X)$ filtered by curvature are computed.

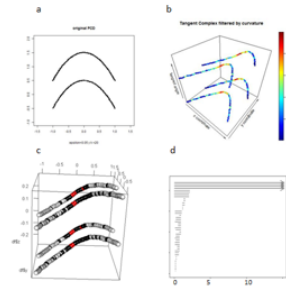


Figure 5.5: Scheme of classification method 2 part 1.

Given a new TOFD measurement and therefore a b-scan, the second part of the method describes how to find what defect type of the catalogue it belongs to (Figure 5.6). Deconvolution and total noise removal for extraction of the defect shape is performed. Next, the barcode shape descriptor is computed and a comparison between the computed barcode and the ones from the catalogue allow us to find out the defect type.

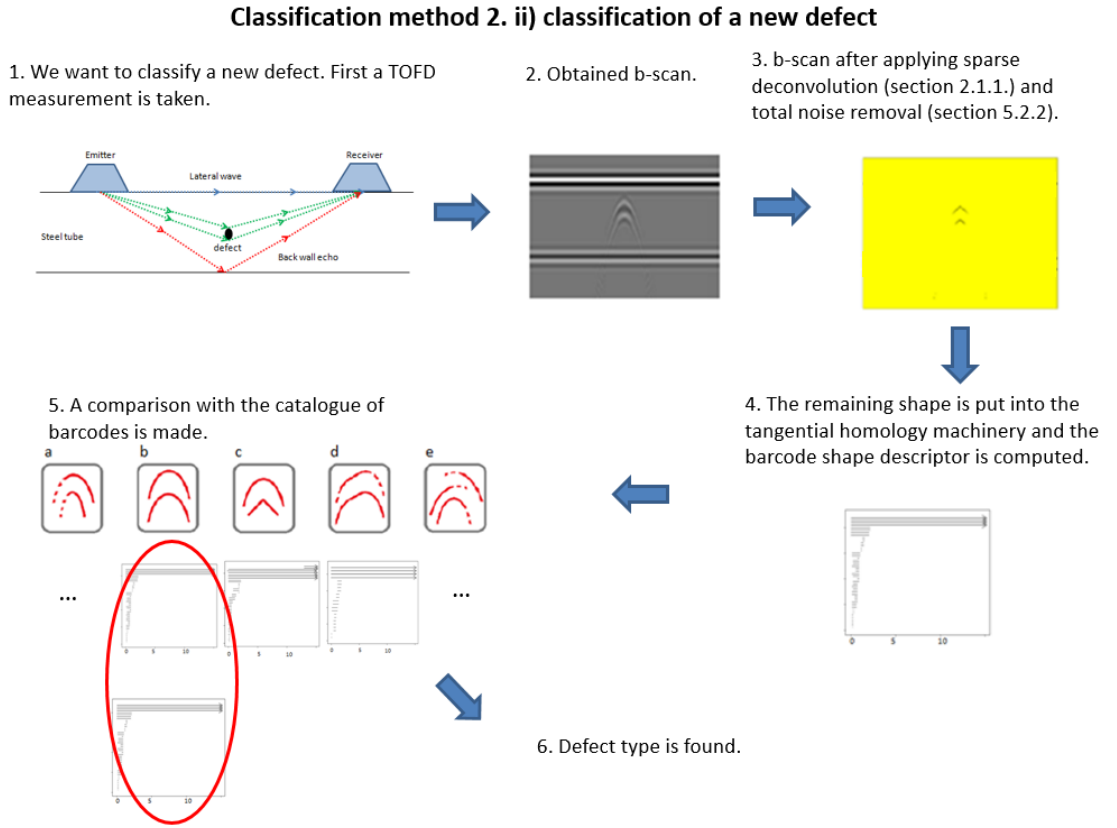


Figure 5.6: Scheme of classification method 1 part 2.

5.2.1 Part 1: creation of a catalogue of TOFD defects

Different TOFD measurements have been taken in the laboratory to establish a correspondence between the b-scan shapes and the different defects represented with them.

The department of non-destructive testing of "Salzgitter Mannesmann Forschung" GmbH, a partner in our research project, was commissioned by a global supplier of welded large-diameter steel pipes for the oil and gas industry to establish a TOFD setup in order to examine weld seams of longitudinal welded pipes inline. For this purpose a ToFD setup was designed, constructed and installed in the mill where measurements were mostly performed on pipes with an outer diameter of approximately 0.813 m (32") and a wall thickness of 0.039 m. The ultrasonic beam angle was set by a wedge to 60° and a centre frequency of 5 MHz (transducer Olympus V310-SM) was used according to the standard specification ISO EN DIN 10863. Furthermore, the divergence angle of the emitting transducer was high enough to probe almost the entire weld seam volume. The probe center separation was set to approximately 0.09 m, which resulted in strong analogue signal strength which was recorded and digitized by an Olympus OmniScan MX test unit and finally saved on a computer.

With this setup, several hundreds of weld seams were tested and evaluated. In doing

so, it became apparent that it is possible to classify indications by similar appearances of flaws in the TOFD scan [33]. This classification was done manually and contains along with the different appearances also the frequency of occurrence.

The found indications are shown in Figure 5.7. Even though most of them present a parabolic shape, there are still important geometric differences that have to be recognized by our automatic classification method.



Figure 5.7: Most common types of volumetric indications experienced in our TOFD real measurements.

The method we suggest for the classification is based on [29], but applied to TOFD defects. A compact shape descriptor is computed for given PCD. This shape descriptor is a barcode, that will be based on several techniques from differential geometry and algebraic topology.

Basically, persistent homology is applied not to the space itself, but to spaces that will be enriched with information about the shapes (tangential information and curvature). This technique allows to distinguish between topologically identical objects but with a slight different geometry. One of the big advantages of this method is that it can be used to any curve PCD without any previous knowledge about it.

5 Classification methods

In order to create a shape descriptor for the found indications, we perform the following steps, which are explained in details in the next subsections:

- (i) Building the tangent complex $T(X)$.
- (ii) Building the approximated $T(X)$.
- (iii) Filtering of the approximated $T(X)$ by curvature.
- (iv) Computation of the barcode shape descriptor (β_0 barcode).

5.2.1.1 Tangent complex $T(X)$

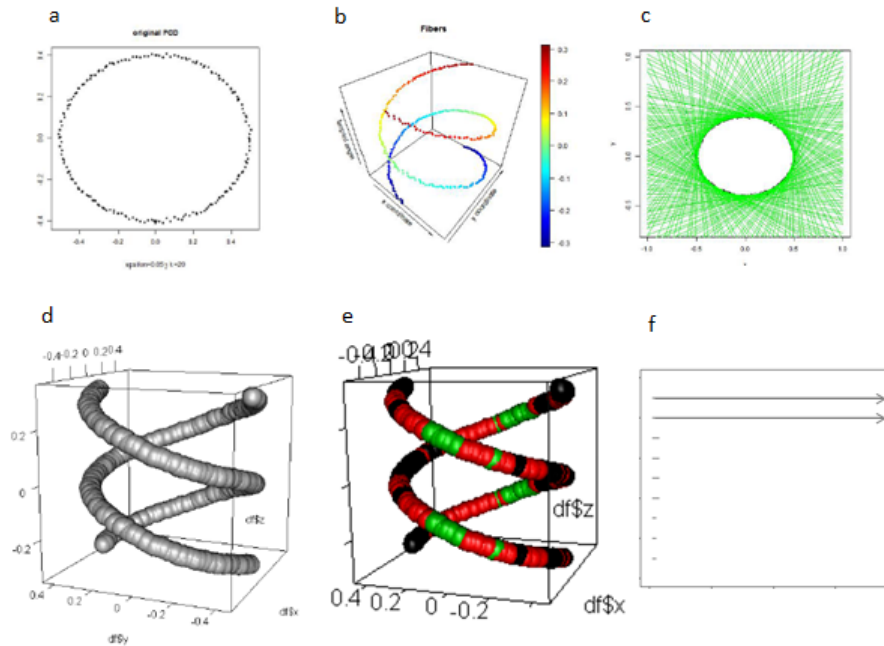


Figure 5.8: (a) Curve PCD (analogous to deconvolved and de-noised b-scan); (b) Tangent Complex $T(X)$; (c) Graphic control of the $T(X)$ depending on the neighbourhood S ; (d) Approximated $T(X)$; (e) Approximated $T(X)$ filtered by curvature where every colour represents a different value for the curvature; (f) β_0 barcode representation.

We focus our attention on curve PCD in 2D, the structure that a deconvolved b-scan shows. We compute the so-called Tangent complex to our PCD. Let $X \in \mathbb{R}^2$ be a curve.

First, we consider $T^0(X) \subseteq X \times \mathbb{S}^1$ to be the set of all possible tangents in every point $x \in X$, i.e.,

$$T^0(X) = \left\{ (x, \nu) \mid \lim_{t \rightarrow 0} \frac{d(x + t\nu, X)}{t} \right\},$$

where (x, ν) represents a point x with tangent vector in the direction ν . The *Tangent Complex* of X is defined as the closure of T^0 , meaning $T(X) = \overline{T^0(X)} \subseteq \mathbb{R}^2 \times \mathbb{S}^1$. As a direct consequence, the projection $\pi : T(X) \rightarrow X$ sends a point $(x, \nu) \in T(X)$ onto its basepoint and $\pi^{-1}(x) \subseteq T(X)$ is the fiber at x .

So how to compute the fibers in our PCD? We take as example the elliptic PCD P from Figure 5.8. It is not so trivial since we do not know the underlying space X from which P was sampled. $\forall p \in P$ we must estimate the two possible tangent directions. We proceed as follows:

- (i) We choose a neighbourhood $S = \{x_1, \dots, x_k\}$ (k nearest neighbors of p).
- (ii) We assume that the best tangent line passes through $x_0 = (1/k) \sum_{i=1}^k x_i$ i.e., the average coordinate of S .
- (iii) To find the orientation of the line, we minimize the sum of the squares of the perpendicular distances of the line to the point's nearest neighbors, solving a least squares problem.

Let M be the matrix whose i th row is $(x_i - x_0)^T$ representing the distances between every point in the neighbourhood x_i and x_0 . Then, if n is the unitary perpendicular vector to the tangent line, Mn is the vector of perpendicular distances. The quantity to be minimized is $|Mn|^2$. The solution to this problem, and therefore the fiber for $p \in P$ is the eigenvector corresponding to the larger eigenvalue of the covariance matrix $M^T M$, as well as the vector pointing in the reverse direction.

We measure the angle $\theta \in [-\pi, \pi]$ that exists between the tangent line and a reference, being the x axis our choice. This is how we come from Figure 5.8a to Figure 5.8b, where 3 coordinates are represented: x and y coordinates of the original curve, and z coordinate corresponding to θ angle. The top and the bottom of the volume are glued. Take into account that there are two curves because there are always two possible tangent directions. Therefore there is a map $\mathbb{R}^2 \rightarrow \mathbb{R}^3 : (x, y) \rightarrow (x, y, \theta)$ between Figure 5.8a and Figure 5.8b.

Another aspect that comes into consideration is how to choose the neighborhood S . It was determined empirically taking a certain number of neighbors that give us a suitable $T(x)$ (Figure 5.8 c). In the future, it could be interesting to try automatic neighbour estimation depending on the given data.

5.2.1.2 Approximated $T(X)$

The aim is to study the topology of the tangent complex $T(X)$. The problem is that this set consists just of points, and points have trivial topology. As we already explained

in section 3.5., it is possible to follow a *persistent homology with different birth times* approach, where we first center and fix an ϵ ball of radius ϵ at each point p .

$$B_\epsilon(p) = \{x | d(p, x) \leq \epsilon\}.$$

The approximated $T(X)$ shown in Figure 5.8d is the union of balls in the fiber points:

$$T(X) \approx \bigcup_{p \in \pi^{-1}(P)} B_\epsilon(p).$$

In the previous section it was shown that $T(X)$ lies on $\mathbb{R}^2 \times \mathbb{S}^1$, therefore a metric is necessary to define what an ϵ ball is. An Euclidean metric can be defined as follows:

$$ds^2 = dx^2 + \omega^2 d\nu^2,$$

where $x \in \mathbb{R}^2$ and $\nu \in \mathbb{S}^1$. ω is a scaling factor that is used to define the importance of a change in the tangent direction. A large value ω will spread the fibers out in $T(X)$ in angular directions.

Another issue is how to choose ϵ . It needs to be chosen large enough so that the basepoints are connected when $\omega = 0$. In our case, the basepoints will correspond to the defect PCD in a deconvolved b-scan. We consider the following rule of thumb proposed by [29]:

$$\epsilon \approx \frac{\sqrt{\Delta x^2 (1 + (\omega k)^2)}}{2},$$

where Δx is the average separation between points, k is the maximum value of curvature and ω is the scaling factor.

5.2.1.3 Approximated $T(X)$ filtered by curvature

The next step is to create a filtration. Following the “*Persistent homology with different birth times*” approach, each vertex will get a value indicating in which step of the filtration appears. This value is the curvature at each point (Figure 5.8e). The question is, how to compute it for every point in the original PCD shown in Figure 5.8a? The curvature k at a point $x \in X$ with tangent direction ν is

$$k(x, \nu) = \frac{1}{\rho(x, \nu)},$$

where ρ is the radius of the osculating circle. A good approximation for ρ is the radius of the osculating parabola, and it is cheaper to compute [29]. Again a neighbourhood is chosen as shown in Figure 5.9.

To find the best parabolas $\forall p$ in the PCD, the idea is to minimize the sum of the squares of the vertical distances of the parabola to the point’s nearest neighbors. To this end, the nearest neighbors of S are moved into a coordinate frame with horizontal axis the tangent direction, vertical axis the normal direction to the tangent and the origin to be x_0 (Figure 5.9 and 5.10). The resulting parabola will depend on the frame coordinate

5.2 Classification method 2: tangential homology and TOFD defects

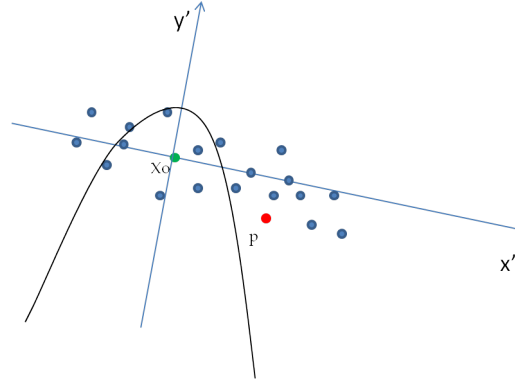


Figure 5.9: For a point p , a neighbourhood S is chosen (blue dots), being x_0 the average coordinate in S .

in which the points are expressed, and this will change $\forall p$, since the tangent directions are different.

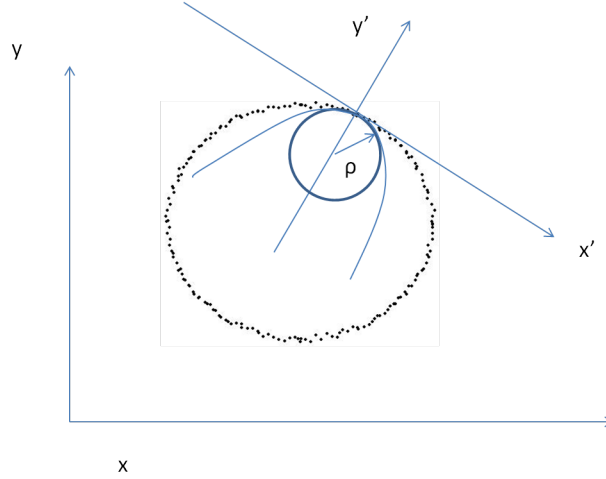


Figure 5.10: Rotation and translation of axes.

At this point we are ready to find a parabola $f(x') = c_0 + c_1x' + c_2x'^2$ in the following way:

- Let $S = \{(x'_1, y'_1), \dots, (x'_k, y'_k)\}$ be the considered neighbourhood.

- Let $A = \begin{bmatrix} 1 & x'_1 & x'^2_1 \\ 1 & x'_2 & x'^2_2 \\ \vdots & \vdots & \vdots \\ 1 & x'_k & x'^2_k \end{bmatrix}$

5 Classification methods

- Let $C = (c_0, c_1, c_2)^T$
- Let $Y = (y'_1, \dots, y'_k)^T$

Defined all the variables above, $\mu = AC - Y$ is the vector that quantifies the distance between $f(x')$ and S . Therefore, we want to find the vector C s.t. $|\mu|$ is minimized. The solution to this problem is:

$$C = (A^T A)^{-1} A^T Y,$$

being the curvature of $f(x')$ at the vertex $2c_2$, and this will be our estimation for the curvature at every point p . Finally, we are ready to build a Vietoris-Rips Filtration and therefore to compute the β_0 barcodes of the shapes (Figure 5.8f). If the barcodes are computed for the shapes of Figure 5.7, then the barcode is a shape descriptor and fingerprint of the studied defect, since it contains tangential and curvature information.

5.2.2 Part 2: classification of a new defect

Given a new TOFD measurement and b-scan, our method will operate as follows in order to find out to which defect from the catalogue belongs to:

- (i) Sparse deconvolution.
- (ii) Total noise removal.
- (iii) Building the tangent complex $T(X)$.
- (iv) Building the approximated $T(X)$.
- (v) Filtering of the approximated $T(X)$ by curvature.
- (vi) Computation of the barcode shape descriptor (β_0 barcode).
- (vii) Comparison with the catalogue of barcodes.
- (viii) The defect type is found.

In the rest of this section we explain how to completely denoise the b-scan in order to extract the defect shape (total noise removal). Sparse deconvolution methods are not 100% effective, as it has been discussed in section 2.1.1. For this reason, we propose a complementary algorithm to clean deconvolved b-scans.

A typical deconvolved b-scan with two pore defects is shown in Figure 5.11 after sparse deconvolution. A standard clustering algorithm such as DBSCAN could fail in the parabolic intersection as explained in section 2.2.1. Our algorithm works as follows:

- (i) *Filtering of lateral wave and backwall echo*: removal of points with high horizontal density, i.e., when certain amount of points are found in a thin horizontal rectangle, it means that they belong to the lateral wave or back wall echo. The algorithm eliminates them (Figure 5.11b).

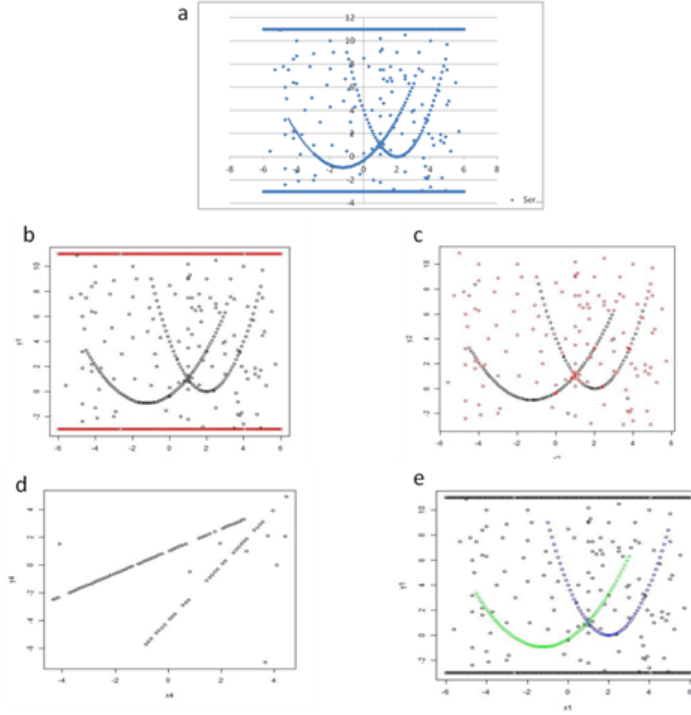


Figure 5.11: (a) TOFD deconvolved b-scan with two pore defects. (b) Filtering of back wall echo and lateral wave (red). (c) Filtered points after step ii) and iii) of the algorithm (red). (d) Transformation to the derivative domain. (e) Recovered parabolic shapes and complete noise removal.

- (ii) *Threshold condition:* Let $t \in \mathbb{R}$ be a threshold value, $X = \{p_1, \dots, p_n\}$ the whole PCD and $\{f_1, \dots, f_n\}$ the 1st nearest neighbour $\forall p \in X$. Since the defect points are very close to each other in comparison to noise points, then if

$$|p_i - f_i| > t,$$

we assume that it is a noise related point and the algorithm eliminates it.

- (iii) *Angle method:* Let $\{s_1, \dots, s_n\}$ be the 2nd nearest neighbour $\forall p \in X$. As shown in Figure 5.12, for defect points

$$\theta = \angle p_i f_i p_i s_i \simeq \pi.$$

On the other hand, for noise related points, θ will differ substantially from π in most of cases. So, when θ deviates a certain value from π , points are eliminated (Figure 5.11 c). There will be some cases where defect points will be eliminated, especially in the intersection. These will be recovered in the last step of this algorithm.

5 Classification methods

- (iv) *Transformation to the derivative domain*: this is necessary in cases where we have two or more intersecting defects, since the intersection point is problematic. The main idea is to compute the derivative domain where this problematic point will not exist. We use a first order central difference to approximate the derivative. Then, $\forall p \in X$ as follows:

$$f'(x_{p_i}) = (y_{f_i} - y_{s_i}) / (x_{f_i} - x_{s_i}).$$

As shown in Figure 5.11d, the 2 parabolic defects are now lines that are much easier to cluster. Going back to the original domain, we know which points correspond to which parabola.

- (v) *Least squares problem*: in order to recover some defect points that were eliminated unintentionally in steps ii) and iii), we solve a least squares problem. The recovered parabolas with the whole noise removal are shown in Figure 5.11e. This algorithm has been tested with several different samples being able to clean 97% of noise-related points.

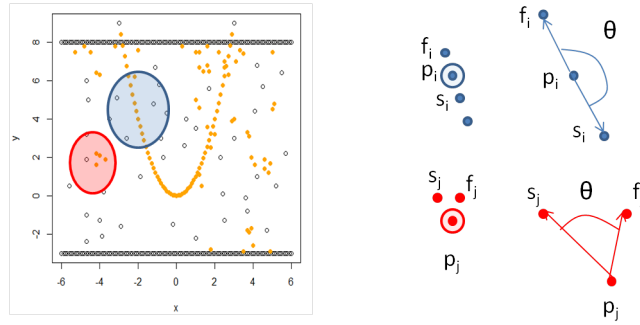


Figure 5.12: Schematic illustration of the “angle method”, step 3 of the algorithm. The blue zoomed area corresponds to defect points zone while the red zoomed area is a noisy zone.

After having extracted the defect shape (curve PCD), it is put into the tangential homology machinery, after which the barcode shape descriptor can be computed and compared to the catalogue barcodes, and the defect type can be found.

6 Nearest neighbour search

As explained in section 5.2.2 and 5.2.1.1, the algorithm of our classification methods looks quite often for nearest neighbours. In a nearest neighbour search problem (NNP) a set of data points P in a d -dimensional space is given. These points are pre-processed into a data structure, so that given any query point q , the nearest or generally k nearest points of P to q can be reported efficiently.

In order to do that, we are using the package RANN of the software *R*. RANN finds the k nearest neighbours for every point in a given dataset based on Arya and Mount's ANN library [9, 8]. This library can give approximate near neighbours to within a specified error bound. The library implements a number of different data structures, based on kd -trees. A kd -tree runs in $O(M \log M)$ time. In this Chapter we explain the basics about kd -trees, giving a motivation for its use, an explanation of how to construct them and an introduction to the typical operations that can be performed in a kd -tree.

6.1 Introduction and motivation of kd -trees

A kd -tree is a data structure that stores a finite set of points from a k -dimensional space. Since the data in our research is mainly two dimensional (deconvolved b-scans), we will focus on $2d$ trees. kd -trees were introduced by [13], and allow us to do efficient processing of points in space, being very flexible and useful in a lot of applications.

In geometric data, in many cases the data points are not evenly distributed. Therefore, one needs a data structure adapted to that distribution of the data. Consider the example of Figure 6.1, where the green dots represent locations of olive trees in Andalusia, Spain. Given a query point, if the target is to find the nearest olive tree to that point, it would be extremely inefficient to start looking at every cell, since the olive trees are concentrated in certain areas. This is the problem that precisely kd -trees try to overcome.

kd -trees have been applied in different fields such as ray tracing [98], N -body simulation [6] and nearest neighbour search [13].

6.2 Construction of $2d$ -trees

A $2d$ -tree is a data structure based in a set of points that facilitates efficient data processing of these points. The idea is to build a tree that corresponds to a recursive partitioning of the plane. We consider the example of Figure 6.2 to explain the procedure where the data consists on eight 2D points a, b, c, d, e, f, g, h . A $2d$ -tree is built as follows:

- We choose a point that will be the root of the tree. Our choice is a .

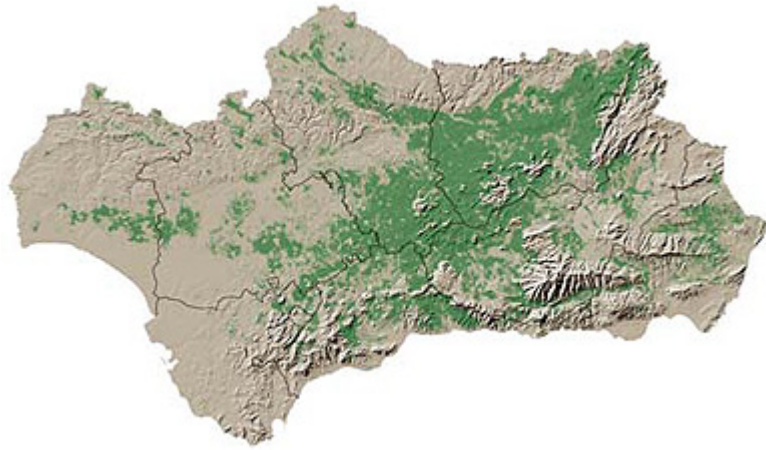


Figure 6.1: Map of olive trees in Andalusia (Spain). The green dots correspond to locations of existing olive trees in the region. Source: [76].

- We cut the space in half. In two dimensions the plane can be cut with an horizontal or vertical line. In this case we choose the horizontal one (green line). Points located in one half will be siblings belonging to the left branch of the tree and points located in the other half will be located in the right branch.
- Consider a second point b of the set. Now one needs to alternate with the other dimension of the space y . A vertical line is added. Whenever a new point is added to the tree, one needs to check to which branch of the tree corresponds and subdivide the space in a different dimension than the previous node did.

The generated tree is shown in Figure 6.3. Let us try to sketch how the whole tree is built. The node of the tree a divided the space in two halves. b belongs to the down part. The next point we consider is c , which belongs to the upper part, the right branch of the tree. Since its previous node in the tree a cuts the space horizontally, c does it now vertically. d is in the upper part of a (right branch of the tree) and in the right side of c (right branch of c) and cuts the space horizontally. The same procedure applies to the rest of nodes.

6.3 Some operations in kd trees

In this section we explain two important operations that can be performed in kd trees: range search and nearest neighbour search.

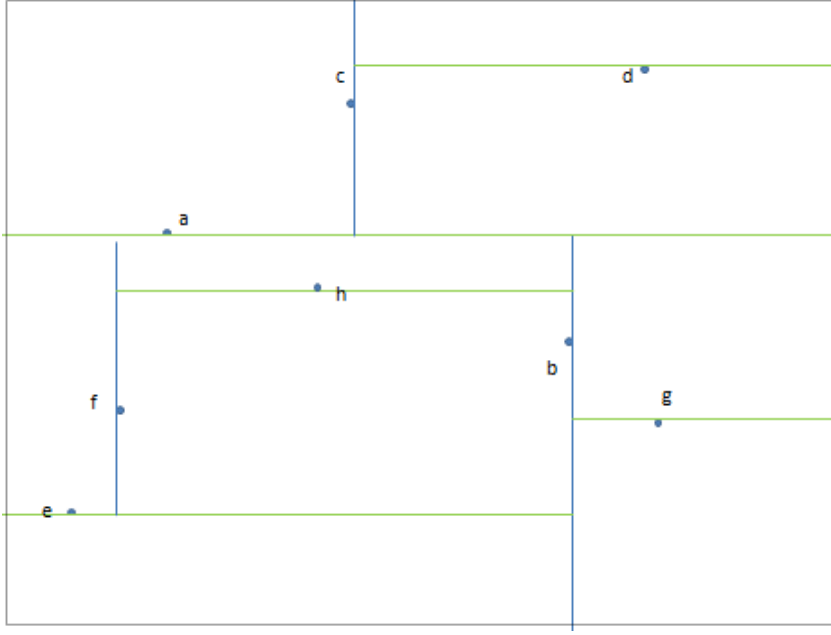


Figure 6.2: Graphic representation of the space partition given by a $2d$ -tree of the set of points from a to h .

6.3.1 Range search in a $2d$ -tree

In this operation the goal is to find all points in a query axis-aligned rectangle. In order to do that, we proceed as follows:

- Check if the root of the tree lies in a given rectangle.
- Recursively search left/bottom (if any could fall in rectangle).
- Recursively search right/top (if any could fall in rectangle).

To further explain this, imagine we wish to find which points in our dataset are contained by the yellow rectangle of Figure 6.4.

We proceed as it was explained and first we see that the root of the tree a does not lie on the rectangle. Therefore, we now check in which half divided by a the rectangle lies. It lies on the down part or left branch of the tree (Figure 6.3). Accordingly, we only examine this branch of the tree and here we already see one of the advantages of *kd* trees.

We reach point b which does not lie on the rectangle and we keep looking on the left side of the space divided by b . Point e is reached which does not match the yellow query but we can appreciate that the rectangle is now in both spaces divided by e . Therefore, an examination in both regions is needed. In one of the regions there are no points so it is rejected, and in the other we see that the next point f is actually in the yellow rectangle and as a consequence it is the champion.

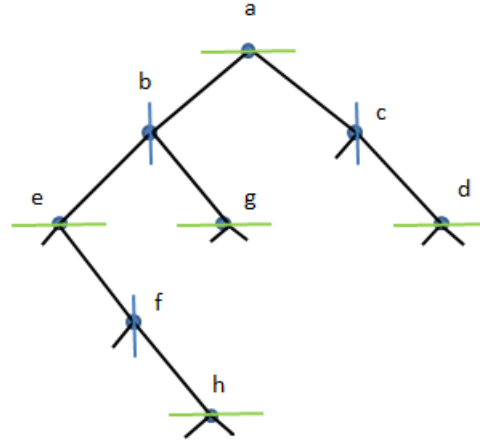


Figure 6.3: 2d-tree of the dataset $a-h$.

6.3.2 Nearest neighbour search in a 2d-tree

Aiming to find the closest point to a query point q , the next procedure is followed:

- Check distance from point in the root to query point.
- Recursively search left/bottom (if it could contain a closer point).
- Recursively search right/top (if it could contain a closer point).
- Organize the method so that it begins by searching for a query point.

We consider the example of Figure 6.4 where a given query point q is marked with red. We start calculating the distance from the root a to q . Then we start going down the tree. q is in the left branch of a (Figure 6.3) so we just check the points that belong to that branch. The next point in the tree in that direction is b and

$$d(q, b) < d(q, a),$$

so b is our new champion. Next we have to check the left branch of b and we see that

$$d(q, e) > d(q, b),$$

so b remains as the champion. Due to the location of q we just have to examine the right branch of e (Figure 6.3 and 6.4). The next point to check is f and

$$d(q, f) > d(q, b),$$

so b keeps being the champion. Finally, we just need to check the right branch of f and the last point h with

$$d(q, h) < d(q, b),$$

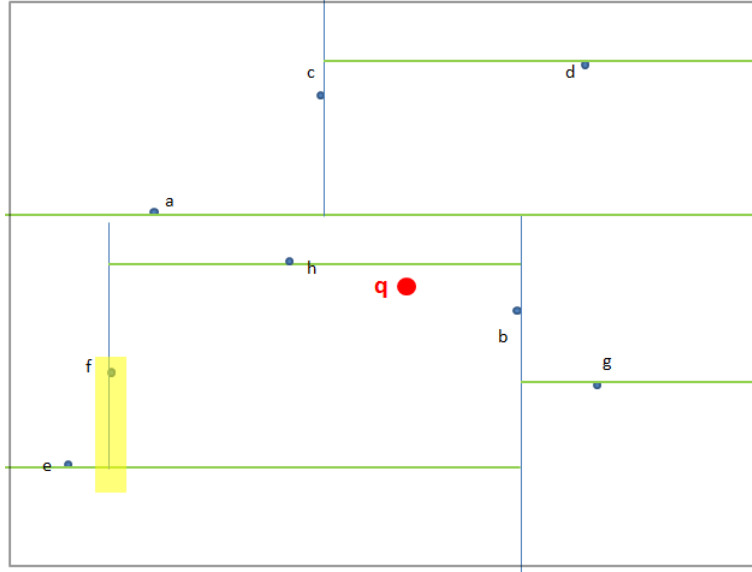


Figure 6.4: Graphic representation of the space partition given by a $2d$ -tree of the set of points from a to h , where a query yellow rectangle and point q are also represented.

and therefore h is the champion and the nearest neighbour of q .

One can also divide the space according to the median of x and y coordinates. Let us introduce a second example where we consider the following data set S in $2D$: $[(0,7), (3,4), (2,3), (2,8), (6,2), (9,9), (4,1), (8,6), (6,6), (10,8)]$. It is geometrically represented in Figure 6.5, with points in blue. The tree is build as follows:

- We pick an attribute i.e., x , and calculate the median. In our case the median is equal to 5.
- The data is split along that median (drawing a vertical line in $x = 5$).
- We examine the points which its x coordinate is less or equal than the median, i.e., $[(0,7), (3,4), (2,3), (2,8), (4,1)]$ and compute the median of the y coordinate that turns out to be 4. An horizontal line is drawn in $y = 4$.
- The process is repeated for the points lying at the right side of $x = 5$.

This way, we fracture the space in halves and this process could go on. At the end, points are located in different hypercubes. The $2d$ -tree is represented in Figure 6.6

If a query point q is added (Figure 6.5), one can walk down the $2d$ -tree looking in the hypercube where q is located. In this case, the algorithm would choose $(2,8)$ as the nearest neighbour of q . This is of course an approximate technique and in this example is clearly seen, since the real nearest neighbour of q is the point $(3,4)$.

In a typical case, the nearest neighbour search in kd -trees runs in $O(\log N)$ and in the worst case in $O(N)$. $2d$ -trees are extremely effective in quickly processing huge amount

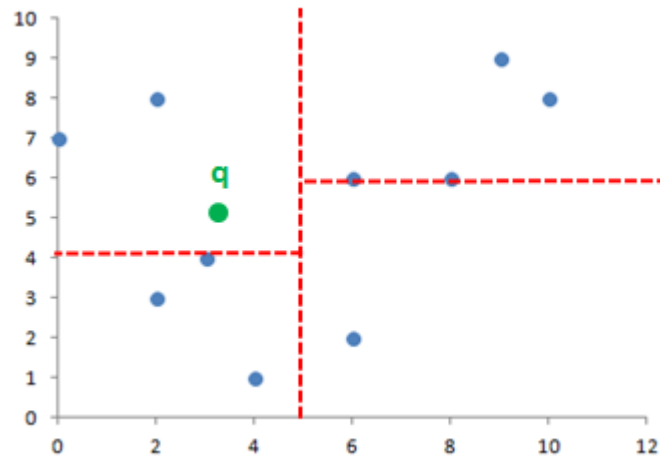


Figure 6.5: Graphic representation of the dataset S and the query point q .

of geometric data. One can easily extend a $2d$ tree to higher dimensions. Even if there are k dimensions, one will do a recursive partitioning one adimension at a time. For instance in $3D$, planes are used to partition the space.

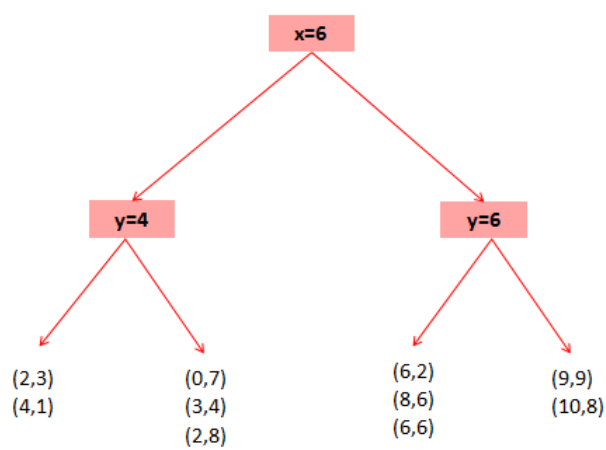


Figure 6.6: The 2d tree of the dataset S .

7 Numerical Results

Numerical results are presented and analyzed in the present chapter. The two automatic classification methods of TOFD defects using ultrasonic NDT techniques explained in Chapter 5, were tested with real and simulated data. Section 7.1. includes the numerical results for the classification method based on persistent homology and section 7.2., those results for the classification method based on tangential homology.

7.1 Results: persistent homology and TOFD defects

We have analyzed three different cases of TOFD defects. In the test cases 1 and 3 the data was simulated with the model proposed in [19]. Test 2 corresponds to real data. To compute the persistent homology of the PCD we have used the software packages Perseus [72] and the Phom [91].

7.1.1 Test Case 1: simulated TOFD Crack

A component of 20 mm thickness is simulated with the weld at $z = 0$. A crack is located at $(x = 40 - 75, y = -11, z = 0)$. Two probes, one emitter and one receiver, are moving along the weld (Figure 7.2b). The angle of aperture of the emitter is 12 degrees, other parameters of ultrasound are: bandwidth: 30 MHz, frequency: 5 MHz, shift = $\pi/4$. a-scans are taken every 0.5 mm, 300 in total. The velocity of ultrasound in steel tube is known and defined ($c_l = 5.92$ mm/s and $c_t = 3.23$ mm/s). Gaussian noise was also added to the geometrical model considering a variance of 0.035. The deconvolved b-scan is shown in Figure 7.1 (blue dots), where the backwall echo and lateral wave have also been filtered out. The crack is detected very well (Figure 7.1). The entire detection process takes less than 3 seconds CPU time only.

7.1.2 Test Case 2: real TOFD data with lack of fusion

The data (Figure 7.3a) was obtained from a sample of a pipe with outer diameter 1.166 m and wall thickness 23.3 mm. The weld seam was tested with a 10 MHz transducer, 6 mm diameter (Olympus C563-SM). The angle of incidence is 70 degrees. The data exhibits a lack of fusion at the end of the pipe. a-scans were taken every 0.5 mm. Again, 300 a-scans were performed. Barcodes for the first homology group are shown in Fig. 7.4. The defect is detected very well.

7 Numerical Results

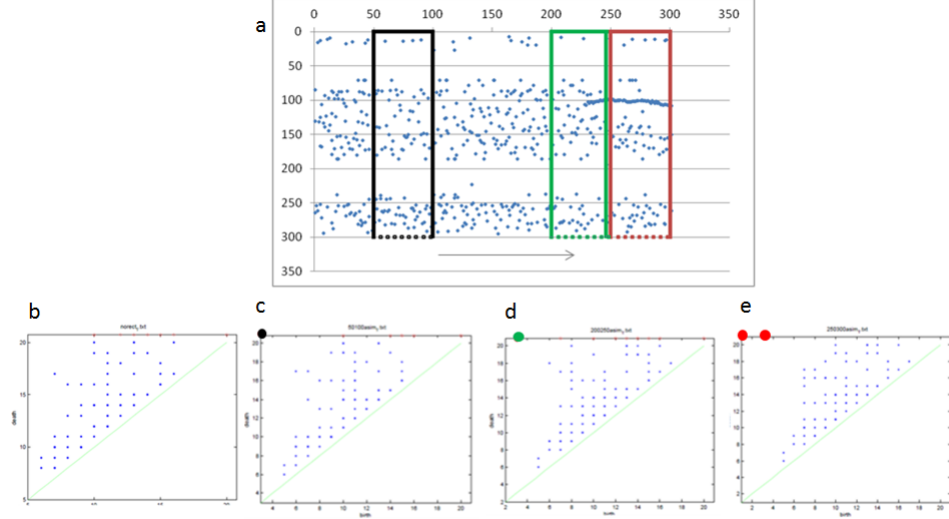


Figure 7.1: (a) Move asymmetric rectangle of points scanning the whole image. (b) persistent diagram for the PDC without adding any extra data; (c), (d), (e) persistent diagram for the PCD (deconvolved b-scan adding the black, green, red rectangle of points). Note the change from 1 to 2 stable 1D holes (dots in the upper left corner of diagrams) when the rectangle matches the crack.

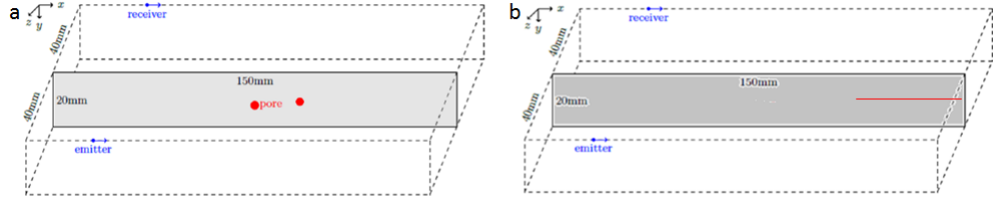


Figure 7.2: (a) Simulation for TOFD pore defects; (b) for crack.

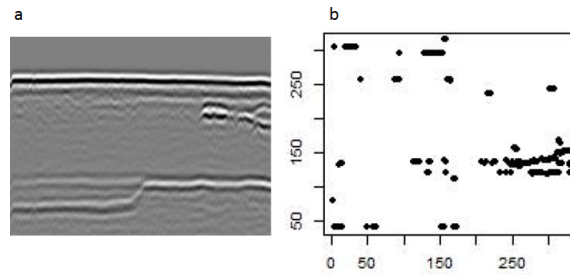


Figure 7.3: (a) b-scan Real TOFD data; (b) deconvolved and filtered b-scan.

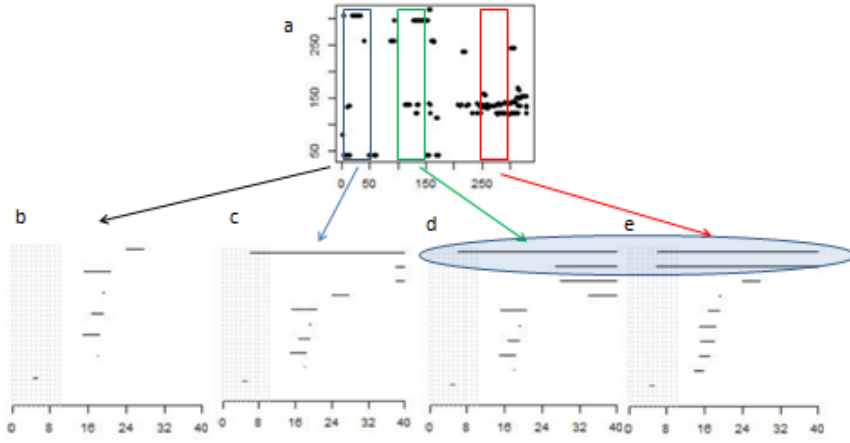


Figure 7.4: Barcode representation of persistent homology of PCD representing a lack of fusion. (b) no extra data is added, therefore no stable hole is detected; (c) blue rectangle of points added: one stable hole; (d) green rectangle of points added: one stable hole; (e) red rectangle of points added: two stable holes. This change in stable one-dimensional holes detects the defect.

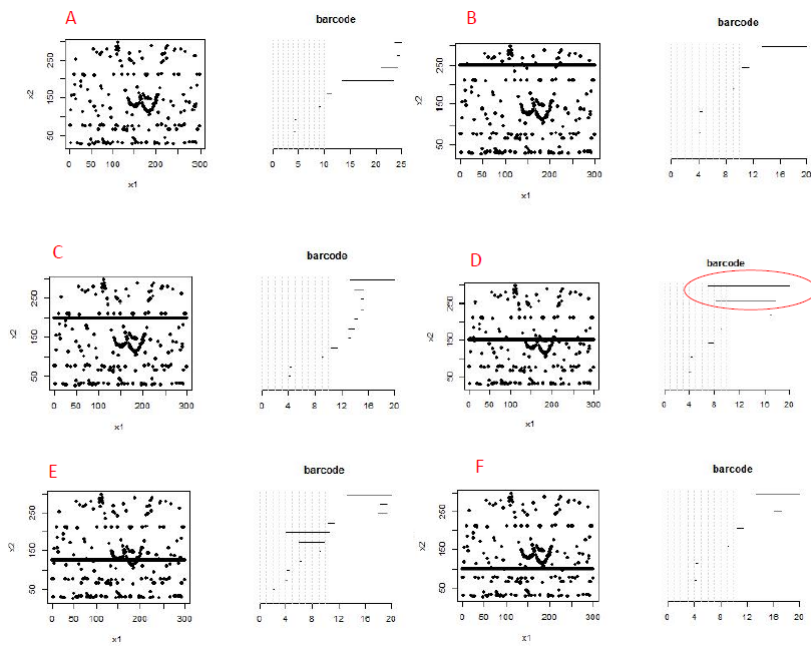


Figure 7.5: Barcode representation of persistent homology computed for the PCD representing two pore defects.

7.1.3 Test Case 3: simulated TOFD data with two pore defects

We used the same weld as in test case 1, but now with two pore defects (Figure 7.2a). We chose a bandwidth of 50 Mhz, and set the variance of the noise to 0.02. The pore defects are at $(0, -13.5, 0)$ and $(16, -12, 0)$. Our results are shown in Figure 7.5. The two pore defects are detected at the jump from 0 to 2 stable holes in the barcode representation. This happens when the horizontal line intersects with the parabolic defects. The total CPU time for all computations is less than 3 seconds only.

7.2 Results: tangential homology and TOFD defects

In this section the barcode shape descriptor is computed for shapes that simulate the different real indications shown in Figure 5.7. This way a catalogue of the most common defects has been created. We analyze defects b , c and j of Figure 5.7.

The data is generated manually as follows:

- Defect b Figure 5.7: we select 50 points per unit length along the x and y axis and project them into the parabolas of equations $f(x) = -x^2 + 0.5$ and $f(x) = -x^2 + 1.5$ with $x \in [-1, 1]$. Next Gaussian noise is added to each point in the PCD with mean 0 and standard deviation 0.01.
- Defect c Figure 5.7: same procedure than before projecting the points to the parabola $f(x) = -x^2 + 1.5$ and the lines $f(x) = 0.625x + 0.2$ e $f(x) = -0.625x + 0.2$.
- Defect j Figure 5.7: the points are projected over a semi-ellipse with semiaxes $a = 1$ and $b = 0.65$ and the line $f(x) = 0.14$.

We set the necessary parameters to input the three PCD into the tangential homology machinery:

- (i) ϵ is set to 0.05 and it is enough to connect properly all the defects points.
- (ii) the scaling factor ω is set to 0.1.
- (iii) the number of neighbours k is set to 20.

The results are summarized in Figures 7.6, 7.7 and 7.8 where the barcode shape descriptor is created for every shape.

Let us analyze the efficiency of this method for shape classification. It is obvious that pure topology is not able to distinguish between the defects in Figures 7.6a and Figure 7.7a, since both present two connected components and no holes. Nevertheless, we can use the proposed method for this goal. Defect from Figure 7.6a present two smooth parabolic shapes. On the contrary, defect in Figure 7.7a has a kink or singularity.

This kink will split the tangent complex in many pieces since the curvature varies extremely fast in the kink. Defect from Figure 7.6a presents 4 half-infinite intervals in the β_0 barcode that correspond to 4 separated elements in the approximated $T(X)$

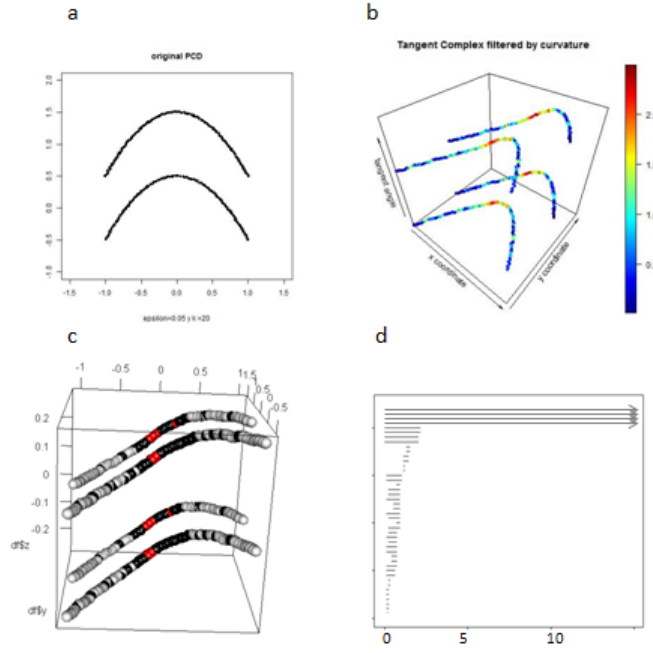


Figure 7.6: Barcode shape descriptor computed for pore defect type b. (a) defect PCD; (b) $T(X)$ filtered by curvature; (c) Approximated $T(X)$ filtered by curvature; (d) β_0 barcode representation.

(Figure 7.6c,d). On the other hand, defect from Figure 7.7a presents 8 (Figure 7.7 d). So, even though they have the same topology, they can be differentiated by the number of half-infinite intervals.

Even in the case when two defects present the same topology and smoothness (Figure 7.6a and Figure 7.8a), the curvature information can be used to distinguish them. In this case both present the same number of half-infinite intervals, but the finite intervals will differ due to the information provided by the curvature of the shapes (Figure 7.6d and Figure 7.8d). To compare different barcodes a metric has been proposed in [29], where one can measure “distance” between barcodes.

Summarizing, when a new TOFD measurement is taken in the laboratory in order to find out the type of defect d :

- The obtained deconvolved and denoised b-scan is put into the tangential homology machinery.
- A barcode is generated.
- This shape descriptor is compared with the catalogue of barcodes that we computed from real data experience (Figure 5.7).

7 Numerical Results

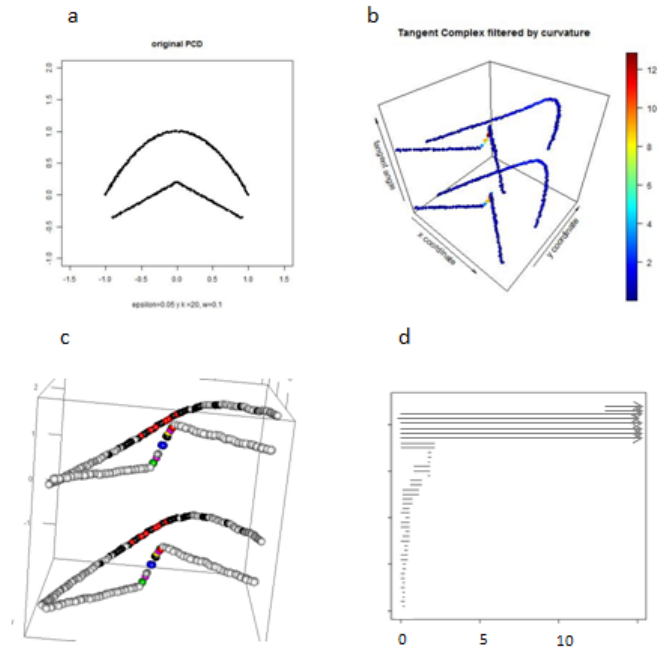


Figure 7.7: Barcode shape descriptor computed for pore defect type c. (a) defect PCD; (b) $T(X)$ filtered by curvature; (c) Approximated $T(X)$ filtered by curvature; (d) β_0 barcode representation.

- After this comparison, we find out what kind of indication d exists inside our pipeline.

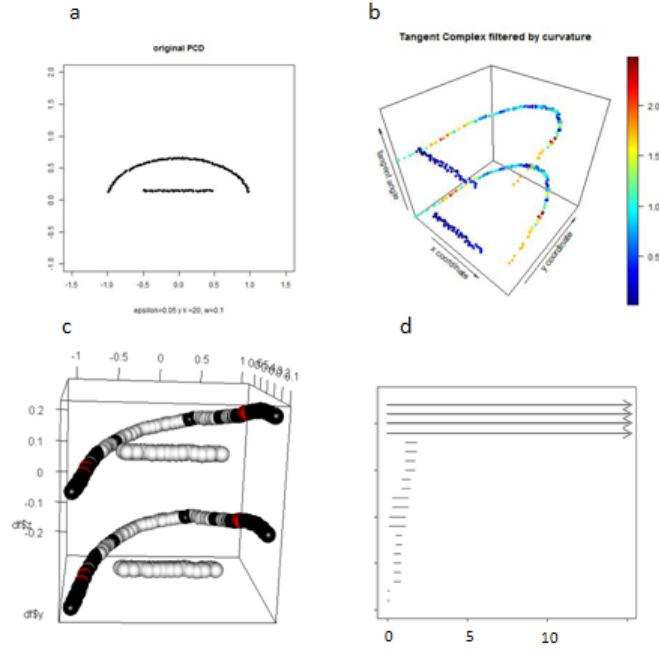


Figure 7.8: Barcode shape descriptor computed for pore defect type j. (a) defect PCD; (b) $T(X)$ filtered by curvature; (c) Approximated $T(X)$ filtered by curvature; (d) β_0 barcode representation.

8 Summary and conclusions

Finally we summarize the results of this thesis. As we explained in the first chapters, the Time of Flight Diffraction (TOFD) technique is one of the most advanced methods in ultrasonic NDT for automatic weld inspection, especially in the steel industry, where it is rather important to avoid productivity losses. Even though it has a low rate of false indications, and it is fast and precise in terms of size accuracy, it has a huge drawback, and this is the interpretation of the images that are generated by the ultrasounds used in TOFD. To be more precise, most of industrial companies working with TOFD perform a manual interpretation of b-scans, and this implies that every decision regarding interpretation of TOFD images depends on the experience and knowledge of one human operator. For large amount of scans and data, this is of course inefficient and cost time, money and moreover can lead to errors.

In this research project innovative methods for automatic classification of TOFD defects have been presented. We show a successful application of persistent homology based methods to the task of automating the decision making process involved in the interpretation of TOFD images.

Before applying our methods, image pre-processing is performed to the raw b-scan in order to immensely reduce the amount of data and save computational time in next steps. To this end, sparse deconvolution methods are used on the region of interest (ROI) of the TOFD b-scans. Two automated classification methods have been proposed at this point:

- In the first method a scanning technique based on persistent homology is used to detect and distinguish between volumetric defects (pores, wormholes, etc.) and planar defects (cracks and lack of fusion), where the change in the number of stable holes in the PCD when adding extra suitable data determines the type of defect. The method turns out to be fast and able to handle noisy datasets. Moreover, the classification is done without the image segmentation stage, which is always needed in others classification methods. One of its limitations is that it is only useful for linear and parabolic shape defects. Even though these are the most common in TOFD, the method has to be adjusted for other shapes.
- The second method performs a deeper classification being able to differentiate defects that even belong to the same class. In a first step, a catalogue of shape descriptors describing the most typical shapes that defects present in TOFD b-scans is created in the following way:
 - (i) Typical b-scan defect shapes are found out taking real TOFD measurements where several hundreds of weld seams were tested and evaluated.

8 Summary and conclusions

- (ii) The approximated tangent complex filtered by curvature is computed.
- (iii) Computation of the 0 homology group is performed to obtain the barcode shape descriptor.

Given a new TOFD measurement and its corresponding b-scan this method finds out the defect type as proceeds:

- (i) The b-scan is deconvolved using sparse deconvolution methods.
- (ii) Complete noise removal is performed for defect shape extraction.
- (iii) The remaining PCD is put into the tangential homology machinery to compute the approximated tangent complex filtered by curvature.
- (iv) The barcode shape descriptor is computed.
- (v) A comparison is made between the barcode and the ones from the catalogue to find out the defect type.

Summarizing, this technique studies the connected components of a complex constructed from a curve PCD using its tangential information. One of the main advantage of this approach is that it can be used for any curve PCD (defect shape) without any prior knowledge about it. One limitation is that the shape cannot contain any remaining noise, since it directly affects to the number of connected components. Therefore powerful pre-processing and denoising techniques are needed.

The proposed methods have been validated by our numerical examples with real TOFD data obtained in the laboratory through several measurements. Our methods are efficient and give a relevant contribution to the development of an automatic system of detection and classification of welding TOFD defects.

Following the work here presented, multiple possible routes could be considered in order to continue and enrich this research:

- Application of our methods to other ultrasonic NDT techniques such as SAFT, etc.
- Extension to higher dimensions, i.e., persistent homology methods for interpretation of 3D scans.
- Analysis of other defects that have not been studied in this thesis (undercut, lamellar tear, etc).
- Use of other complexes such as Witness in order to save computational time.
- Incorporation of our methods to the real practice of detection and classification of welding defects in different industries and materials.
- Application of our methods to other scientific fields where stable holes or topological features can be related to properties of the PCD, such as mask of satellite precipitation data.

Bibliography

- [1] BS EN ISO 10042. Welding. arc-welded joints in aluminium and its alloys. quality levels for imperfections. 2005.
- [2] BS EN ISO 5817. Welding-fusion-welded joints in steel, nickel, titanium and their alloys-quality levels for imperfections. 2007.
- [3] BS EN ISO 6520-1. Welding and allied processes-classification of geometric imperfections in metallic materials-part 1: Fusion welding. 2007.
- [4] Agostino Abbate, Jeff Koay, Julius Frankel, Stephan C Schroeder, and Pankaj Das. Signal detection and noise suppression using a wavelet transform signal processor: application to ultrasonic flaw detection. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 44(1):14–26, 1997.
- [5] Habib Ammari. *Mathematical Modeling in Biomedical Imaging II: Optical, Ultrasound, and Opto-Acoustic Tomographies*, volume 2035. Springer Science & Business Media, 2011.
- [6] Richard J Anderson. Tree data structures for n-body simulation. *SIAM Journal on Computing*, 28(6):1923–1940, 1999.
- [7] Welding answers S.A. Understanding weld discontinuities. Available at <http://weldinganswers.com/434/>, 2015.
- [8] Sunil Arya and David M Mount. Approximate nearest neighbor queries in fixed dimensions. In *SODA*, volume 93, pages 271–280, 1993.
- [9] Sunil Arya, David M Mount, Nathan S Netanyahu, Ruth Silverman, and Angela Y Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)*, 45(6):891–923, 1998.
- [10] Andrea Baraldi and Palma Blonda. A survey of fuzzy clustering algorithms for pattern recognition. i. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(6):778–785, 1999.
- [11] Ulrich Bauer. *Persistence in discrete Morse theory*. PhD thesis, Citeseer, 2011.
- [12] Petr Beckmann and Andre Spizzichino. The scattering of electromagnetic waves from rough surfaces. *Norwood, MA, Artech House, Inc., 1987, 511 p.*, 1987.
- [13] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

Bibliography

- [14] W Benton. Welding inspection and quality control. Available at <http://slideplayer.com/slide/4225338/>, n.d.
- [15] Wendie A Berg, Jeffrey D Blume, Jean B Cormack, Ellen B Mendelson, Daniel Lehrer, Marcela Böhm-Vélez, Etta D Pisano, Roberta A Jong, W Phil Evans, Marilyn J Morton, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *Jama*, 299(18):2151–2163, 2008.
- [16] Michael Berke. *Nondestructive material testing with ultrasonics: introduction to the basic principles*. Krautkammer GmbH & Company, 1996.
- [17] David T Blackstock. *Fundamentals of physical acoustics*. John Wiley & Sons, 2000.
- [18] P Bolland, LFC Lew Yan Voon, B Gremillet, L Pillet, A Diou, and P Gorria. The application of hough transform on ultrasonic images for the detection and characterization of defects in non-destructive inspection. In *Signal Processing, 1996., 3rd International Conference on*, volume 1, pages 393–396. IEEE, 1996.
- [19] Florian Boßmann. *Model Based Defect Reconstruction in Ultrasonic Non-Destructive Testing*. Der Andere Verlag, 2014.
- [20] Florian Boßmann, Gerlind Plonka, Thomas Peter, Oliver Nemitz, and Till Schmitte. Sparse deconvolution methods for ultrasonic ndt. *Journal of Nondestructive Evaluation*, 31(3):225–244, 2012.
- [21] Jacob Brown and Tomáš Gedeon. Structure of the afferent terminals in terminal ganglion of a cricket and persistent homology. *PloS one*, 7(5):e37278, 2012.
- [22] Michel Bruneau. *Fundamentals of acoustics*. John Wiley & Sons, 2013.
- [23] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
- [24] Gunnar Carlsson, Tigran Ishkhanov, Vin De Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *International journal of computer vision*, 76(1):1–12, 2008.
- [25] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas J Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(02):149–187, 2005.
- [26] AA Carvalho, JMA Rebello, LVS Sagrilo, CS Camerini, and IVJ Miranda. Mfl signals and artificial neural networks applied to detection and classification of pipe weld defects. *Ndt & E International*, 39(8):661–667, 2006.

- [27] CF Theresa Cenate, B Sheela Rani, B Venkatraman, and DN Sangeetha. Classification of defects in time of flight diffraction (tofd) images using artificial neural network.
- [28] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.
- [29] Anne Collins, Afra Zomorodian, Gunnar Carlsson, and Leonidas J Guibas. A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, 28(6):881–894, 2004.
- [30] M Consonni, Chen Fun Wee, and C Schneider. Manufacturing of welded joints with realistic defects. *Insight-Non-Destructive Testing and Condition Monitoring*, 54(2):76–81, 2012.
- [31] E Craig. Flux cored welds, part 2, n.d.
- [32] Jose F Cuenca and Armin Iske. Persistent homology for defect detection in non-destructive evaluation of materials. *The e-Journal of Nondestructive Testing*, 21(01), 2016.
- [33] Jose F. Cuenca, Armin Iske, Patrick A. Labud, and Oliver Nemitz. Tangential homology for defect detection in the time of flight diffraction method (tofd). *Proceedings in Applied Mathematics and Mechanics*, 16(1):619–620, 2016.
- [34] Stefan Dantchev and Ioannis Ivrissimtzis. Efficient construction of the čech complex. *Computers & Graphics*, 36(6):708–713, 2012.
- [35] Vin De Silva and Gunnar Carlsson. Topological estimation using witness complexes. *Proc. Sympos. Point-Based Graphics*, pages 157–166, 2004.
- [36] Vin De Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(1):339–358, 2007.
- [37] Edward R Dougherty, Roberto A Lotufo, and The International Society for Optical Engineering SPIE. *Hands-on morphological image processing*, volume 71. SPIE Optical Engineering Press Washington, 2003.
- [38] Herbert Edelsbrunner. The union of balls and its dual shape. In *Proceedings of the ninth annual symposium on Computational geometry*, pages 218–231. ACM, 1993.
- [39] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [40] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.

Bibliography

- [41] Charles Elkan. Using the triangle inequality to accelerate k-means. In *ICML*, volume 3, pages 147–153, 2003.
- [42] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [43] P Evans. 22 possible causes of weld metal porosity. Available at www.thefabricator.com, 2010.
- [44] X Feng and Y Tong. Dept. of comput. sci. & eng., michigan state univ., east lansing, mi, usa. visualization and computer graphics. *IEEE Transactions on*, 19(8):1298–1306, 2013.
- [45] Robin Forman. A user’s guide to discrete morse theory. *Sém. Lothar. Combin*, 48:35pp, 2002.
- [46] Takeo Furukawa and Naoya Seo. Electrostriction as the origin of piezoelectricity in ferroelectric polymers. *Japanese Journal of Applied Physics*, 29(4R):675, 1990.
- [47] The Welding Institute. TWI Group. Defects - lamellar tearing. job knowledge 47. Available at <http://www.twi-global.com/technical-knowledge/job-knowledge/defects-lamellar-tearing-047/>, 2000.
- [48] Simon Haykin and Richard Lippmann. Neural networks, a comprehensive foundation. *International journal of neural systems*, 5(4):363–364, 1994.
- [49] R Hennigan. Simplicial homology of the alpha complex, 2013.
- [50] Christiaan Huygens. *Treatise on light*. tredition, 2012.
- [51] Bernard Jaffe. *Piezoelectric ceramics*, volume 3. Elsevier, 2012.
- [52] H Jaffe and DA Berlincourt. Piezoelectric transducer materials. *Proceedings of the IEEE*, 53(10):1372–1386, 1965.
- [53] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.
- [54] Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek. *Computational homology*, volume 157. Springer Science & Business Media, 2006.
- [55] J Krautkramer and H Krautkramer. Ultrasonic testing of materials, 1990. *Cited on*, page 8, 1953.
- [56] Tribikram Kundu. *Ultrasonic nondestructive evaluation: engineering and biological material characterization*. CRC press, 2003.
- [57] Der-Tsai Lee et al. On k-nearest neighbor voronoi diagrams in the plane. *IEEE Trans. Computers*, 31(6):478–487, 1982.

- [58] Richard CT Lee. Clustering analysis and its applications. In *Advances in Information Systems Science*, pages 169–292. Springer, 1981.
- [59] Brian Michael Lempriere. *Ultrasound and elastic waves: frequently asked questions*. Academic Press, 2003.
- [60] S Liu, DL Olson, and S Ibarra. Electrode formulation to reduce weld metal hydrogen and porosity. Technical report, American Society of Mechanical Engineers, New York, NY (United States), 1994.
- [61] MR Louthan, GR Caskey, JA Donovan, and DE Rawl. Hydrogen embrittlement of metals. *Materials Science and Engineering*, 10:357–368, 1972.
- [62] FW Margrave, K Rigas, David A Bradley, and P Barrowcliffe. The use of neural networks in ultrasonic flaw detection. *Measurement*, 25(2):143–154, 1999.
- [63] A Masnata and M Sunseri. Neural network classification of flaws detected by ultrasonic means. *NDT & E International*, 29(2):87–93, 1996.
- [64] Thouraya Merazi-Meksen, Malika Boudraa, and Bachir Boudraa. Mathematical morphology for tofd image analysis and automatic crack detection. *Ultrasonics*, 54(6):1642–1648, 2014.
- [65] John Milnor. On spaces having the homotopy type of a cw-complex. *Transactions of the American Mathematical Society*, 90(2):272–280, 1959.
- [66] Konstantin Mischaikow and Vidit Nanda. Morse theory for filtrations and efficient computation of persistent homology. *Discrete & Computational Geometry*, 50(2):330–353, 2013.
- [67] Paul E Mix. *Introduction to nondestructive testing: a training guide*. John Wiley & Sons, 2005.
- [68] Shyamal Mondal and T Sattar. An overview tofd method and its mathematical model. *NDT. net*, 5(4), 2000.
- [69] Elineudo P Moura, Romeu R Silva, Marcio HS Siqueira, and Joao Marcos A Rebello. Pattern recognition of weld defects in preprocessed tofd signals using linear classifiers. *Journal of Nondestructive Evaluation*, 23(4):163–172, 2004.
- [70] James R Munkres. *Elements of algebraic topology*, volume 2. Addison-Wesley Menlo Park, 1984.
- [71] Harris Naftali. Visualizing dbscan clustering. Available at <http://www.naftaliharris.com/blog/visualizing-dbscan-clustering/>, 2015.
- [72] Vidit Nanda. Perseus: The persistent homology software (2012), 2012.

Bibliography

- [73] Vidit Nanda and Radmila Sazdanović. Simplicial models and topological inference in biological systems. In *Discrete and topological models in molecular biology*, pages 109–141. Springer, 2014.
- [74] NM Nandhitha, N Manoharan, B Sheela Rani, B Venkataraman, and Baldev Raj. A comparative study on the performance of the classical and wavelet based edge detection for image denoising on defective weld thermographs. In *12th Asia-Pacific Conference of Non-Destructive Testing*, pages 1–5. Citeseer, 2006.
- [75] V Novakovic. Defects and discontinuities. esc defects training qa/qc. Available at <http://www.slideshare.net/VlastimirNovakovic/welding-defects-45427484>, 2006.
- [76] Nelson Orobio Caballero. ¿por qué hay tantos olivos en andalucía? Available at <http://iberhistoria.es/edad-antigua/romanos/entradas-romanos/olivos-en-andalucia/>, 2014.
- [77] Deepti Pachauri, Chris Hinrichs, Moo K Chung, Sterling C Johnson, and Vikas Singh. Topology-based kernels with application to inference problems in alzheimer’s disease. *IEEE transactions on medical imaging*, 30(10):1760–1770, 2011.
- [78] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.
- [79] PA Petcher and S Dixon. Parabola detection using matched filtering for ultrasound b-scans. *Ultrasonics*, 52(1):138–144, 2012.
- [80] Koyickal Gopalan Prabhakaran, Brian Stephen Wong, and Yeo Yan Teng. Time of flight diffraction: An alternate non-destructive testing procedure to replace traditional methods. In *Microtechnologies for the New Millennium 2005*, pages 534–539. International Society for Optics and Photonics, 2005.
- [81] JE Ramirez, B Han, and S Liu. Effect of welding variables and solidification substructure on weld metal porosity. *Metallurgical and Materials Transactions A*, 25(10):2285–2294, 1994.
- [82] Hoobasarl Rampaul. *Pipe welding procedures*. Industrial Press, 2003.
- [83] NA Rathod and SA Ladhake. Review of data clustering algorithms. *IJCIS*, (2(3)):56–58, 2011.
- [84] Simeon Rony. Discontinuities and defects. welding-defects. Available at <http://www.slideshare.net/RonySimeon/9488085-weldingdefects>, 2012.
- [85] Gary S Schajer. *Practical residual stress measurement methods*. John Wiley & Sons, 2013.

- [86] Udo Schlengermann. *Taschenbuch Ultraschall-Werkstoffprüfung*. Vulkan-Verlag Classen, 1992.
- [87] Steven G Shi, Paul Hilton, Steve Mulligan, and Geert Verhaeghe. Hybrid nd: Yag laser-mag welding of thick section steel with adaptive control. In *Proceedings of the 23th International Congress on Applications of Lasers & Electro-Optics*, 2004.
- [88] C'Shekhar N Shitole, O Zahran, and W Al-Nuaimy. Combining fuzzy logic and neural networks in classification of weld defects using ultrasonic time-of-flight diffraction. *Insight-Non-Destructive Testing and Condition Monitoring*, 49(2):79–82, 2007.
- [89] MG Silk. The interpretation of tofd data in the light of asme xi and similar rules. *British Journal of Non-Destructive Testing*, 31(5):242–251, 1989.
- [90] SJ Sokolov. Ultrasonic oscillations and their applications. *Tech Phys USSR*, 2:522–534, 1935.
- [91] Andrew Tausz. phom: Persistent homology in r. *Software available at <http://cran.rproject.org/web/packages/phom>*, 2011.
- [92] ME Mechanical Team. Defects in welding joints: Internal and external, 2016.
- [93] Norbert Trimborn. The time-of-flight-diffraction-technique, 1997.
- [94] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12):4655–4666, 2007.
- [95] Alfred CC Tseung, Ho C Chan, Nani B Fouzder, and Thurairajah Sriskandara-jah. Prevention of hydrogen embrittlement of metals in corrosive environments, December 18 1984. US Patent 4,488,578.
- [96] Ioannis Valavanis and Dimitrios Kosmopoulos. Multiclass defect detection and classification in weld radiographic images using geometric and texture features. *Expert Systems with Applications*, 37(12):7606–7614, 2010.
- [97] Lew FLY Voon, Patrice Bolland, Olivier Laligant, Patrick Gorria, B Gremillet, and L Pillet. Gradient-based hough transform for the detection and characterization of defects during nondestructive inspection. In *Electronic Imaging'97*, pages 140–146. International Society for Optics and Photonics, 1997.
- [98] Ingo Wald and Vlastimil Havran. On building fast kd-trees for ray tracing, and on doing that in $O(n \log n)$. In *2006 IEEE Symposium on Interactive Ray Tracing*, pages 61–69. IEEE, 2006.
- [99] Zhenjian Wang, Changwen Qu, and Lei Cui. Denoising images using wiener filter in directionalet domain. In *2006 International Conference on Computational Intelligence for Modelling Control and Automation and International Conference on*

Bibliography

- Intelligent Agents Web Technologies and International Commerce (CIMCA'06)*, 2006.
- [100] Kelin Xia, Xin Feng, Yiyong Tong, and Guo Wei Wei. Persistent homology for the quantitative prediction of fullerene stability. *Journal of computational chemistry*, 36(6):408–422, 2015.
- [101] James A Zagzebski. *Essentials of ultrasound physics*. Mosby, 1996.
- [102] CA Zapffe and CE Sims. Hydrogen embrittlement, internal stress and defects in steel. *Trans. AIME*, 145(1941):225–271, 1941.
- [103] Afra Zomorodian. Fast construction of the vietoris-rips complex. *Computers & Graphics*, 34(3):263–271, 2010.
- [104] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.

CURRICULUM VITAE - September 2016

Personal information

- **Name:** José Fernando Cuenca Jiménez
- **Address:** University of Hamburg, Department of Mathematics, Bundesstrasse 55, D-20146 Hamburg, Germany
- **E-mail:** jose.cuenca@uni-hamburg.de
- **Nationality:** Spanish
- **Date of birth:** 17.03.1987

Education

- PhD. in applied mathematics at the University of Hamburg, Germany (2013-2016)
- Erasmus Mundus MSc MATHMODS. Mathematical Modelling in Engineering: Theory, Numeric, Applications (2011-2013). Universities of L'Aquila (Italy), Hamburg (Germany) and Gdansk (Poland)
- MSc in civil engineering at the University of Granada, Spain (2005-2010)

Publications

- (i) Cuenca, J. F., & Iske, A. (2016). Persistent homology for defect detection in non-destructive evaluation of materials. The e-Journal of Nondestructive Testing, 21(01). ISSN 1435-4934.
- (ii) Cuenca, J. F., Iske, A., Labud, P. A., & Nemitz, O. (2016). Tangential Homology for Defect Detection in the Time of Flight Diffraction Method (TOFD). Proceedings in Applied Mathematics and Mechanics. Submitted.
- (iii) Cuenca, J. F., & Iske, A. (2016-7). Persistent Homology for Defect Detection in Non-Destructive Evaluation of Materials. Hamburger Beiträge zur Angewandten Mathematik.

- (iv) Cuenca, J. F. (2013). Particle agglomeration in flow modelled with molecular dynamics coupled to a thermal Lattice Boltzmann code. *TASK Quarterly*, 17(3-4), 181-213.
- (v) J. Cuenca, J. Rodriguez, M. Martin-Morales, Z. Sanchez-Roldan, M. Zamorano (2013) "Effects of olive residue biomass fly ash as filler in self-compacting concrete". *Construction and Building Materials*, 40: 702-709.
- (vi) M.Martin-Morales, J.Cuenca, P.Lopez; "Characterization of the biomass ashes-Application in Mortars" in Spanish national congress of civil engineering: Society, Economy and Environment (2011).
- (vii) Martin-Morales, M., Cuenca, J., Lopez, P., Rodriguez, J., Zamorano, M., & Valverde, P. I. (2011, October). Self-compacting concrete with biomass fly ash: Preliminary results. In *Proceeding of the Thirteenth International Waste Management and Landfill Symposium*, Cagliari, Italy (pp. 3-7).

Conferences and Workshops

2016

- CSE Workshop "Recent trends and future developments in Computational Science & Engineering" (oral presentation), March 22-24 in Plön, Germany.
- Joint annual meeting of DMV and GAMM (oral presentation), March 7-11 in Braunschweig, Germany.
- 26. Rhein-Ruhr-Workshop (poster), January 29-30 in Bestwig, Germany.

2015

- Malaysia International NDT Conference and Exhibition 2015 MINDTCE-15 (oral presentation), November 22-24 in Kuala Lumpur, Malaysia.
- The 18th International Workshop for Young Mathematicians "Algebraic and Differential Topology" (poster), September 13-18 in Cracow, Poland.
- Lothar-Collatz-Seminar (oral presentation), June 9 in Hamburg, Germany.
- Young Researchers Meeting and CSE workshop "Recent trends and future developments in Computational Science & Engineering" (oral presentation), March 9-13 in Plön, Germany.
- 25. Rhein-Ruhr-Workshop (oral presentation), January 30-31 in Bestwig, Germany.

2014

- Mathematical modelling workshop Mamow14: Mathmods & EMA Alumni meeting (poster), July 14-19 in L'Aquila, Italy.
- Young Researchers Meeting and CSE workshop "Recent trends and future developments in Computational Science & Engineering" (poster), March 24-25 in Plön, Germany.
- Rhein-Ruhr-Workshop (poster), Jan-Feb 31-1 in Bestwig, Germany.

Work experience

- Researcher at University of Hamburg (2013-2016).
- Researcher at Institute for Computational Physics in Stuttgart, Germany (March-September 2013).
- Researcher at Project of Excellence – P08-RNM-03584 by the Autonomous Government of Andalusia (2010).
- Professional chess player (Chess Grandmaster GM currently ranked with 2541 ELO points).
- 4 Times World Championship participant and Spanish Champion under 20 in 2007.
- Chess author, commentator and coach.
- Player of German, Swedish, Italian, British, Danish and Spanish chess leagues in first division.