

Universität Koblenz-Landau,  
Campus Landau  
Institut für Mathematik

# Components of a Spatial-Toolbox for processing geocoded mapping information in the context of decision support

Dissertation zur Erlangung des akademischen Grades eines  
Doktors der Naturwissenschaften,  
Fachbereich 7: Natur- und Umweltwissenschaften,  
Universität Koblenz-Landau

vorgelegt von  
Ralf Wagner

Datum der Einreichung: 21. Mai 2010

Tag der mündlichen Prüfung: 23. August 2010

1. Berichterstatter: Prof. Dr. Engelbert Niehaus, Universität Koblenz-Landau, Campus Landau
2. Berichterstatterin: Prof. Dr. Dr. Marlien Herselman, Nelson Mandela Metropolitan University, Port Elizabeth and George, Südafrika

# List of abbreviations

- AI** Artificial Intelligence
- ANN** Artificial Neural Network
- ANNs** Artificial Neural Networks
- BNN** Biological Neural Network
- BNNs** Biological Neural Networks
- BPN** Backpropagation Network
- BPNs** Backpropagation Networks
- CAD** Computer Aided Design
- CKP** Construction Kit Principle
- DSS** Decision Support Systems
- EWARS** Early Warning And Response System
- GIS** Geographical Information System
- GUI** Graphical User Interface
- KN** Kohonen Network
- KNs** Kohonen Networks
- MAS** Multi-agent system
- MLFFN** Multi-layer feed-forward-network
- OO** Object Orientation
- OOA** Object Oriented Analysis
- OOP** Object Oriented Programming
- ReGLaN** Research Group Learning and Neurosciences



**SDSS** Spatial Decision Support System

**SPATTB** “Spatial-Toolbox”

**SOM** Self Organizing Map

**SOMs** Self Organising Maps

**UML** Unified Modeling Language

# Contents

<b>1</b>	<b>Preface</b>	<b>1</b>
1.1	Aim and methods used . . . . .	1
1.2	Structure of the dissertation . . . . .	3
<b>2</b>	<b>Contribution to the project ReGLaN-Health &amp; Logistics</b>	<b>8</b>
<b>3</b>	<b>Embedding in the context of current research</b>	<b>12</b>
3.1	Description and limitation of the context of current research . . . . .	12
3.1.1	General concepts of spatial analysis . . . . .	12
3.1.2	Usability of Artificial Neural Networks (ANNs) and Geographical Information System (GIS) in the context of spatial analysis . . . . .	18
3.1.3	General concepts of decision support systems . . . . .	20
3.1.4	Decision support in current research . . . . .	23
3.2	Integration of this thesis . . . . .	25
<b>4</b>	<b>Artificial Neural Networks</b>	<b>31</b>
4.1	Historical development . . . . .	31
4.2	Fundamental mathematical modeling of Artificial Neural Networks based on Biological Neural Networks . . . . .	34
4.3	Description of the relevant models for this thesis . . . . .	53
4.3.1	Backpropagation Networks . . . . .	55

4.3.2	Self Organising Maps and Kohonen Networks . . . . .	80
<b>5</b>	<b>Principles of object-oriented analysis</b>	<b>87</b>
<b>6</b>	<b>Proposal of a “Spatial Toolbox”</b>	<b>91</b>
6.1	Localisation of the “Spatial-Toolbox” (SPATTB) . . . . .	91
6.2	Application and customisation of OpenSource components . . . . .	92
6.3	Workflow for building up SPATTB components . . . . .	93
<b>7</b>	<b>Measure theory on topological and function spaces</b>	<b>96</b>
7.1	Basic definitions and prerequisites . . . . .	96
7.1.1	Topological space . . . . .	96
7.1.2	Continuity . . . . .	100
7.1.3	Product topology . . . . .	104
7.1.4	Classification of topological spaces . . . . .	106
7.1.5	Compactness . . . . .	109
7.1.6	Metric spaces . . . . .	109
7.1.7	Normed space and semi-normed space . . . . .	111
7.1.8	Locally compact spaces . . . . .	114
7.2	Convergence . . . . .	115
7.2.1	Generalised convergence on topological spaces . . . . .	115
7.3	Completeness . . . . .	122
7.4	Topological vector spaces . . . . .	125
7.4.1	Basic characteristics of topological vector spaces . . . . .	125
7.4.2	Locally convex spaces . . . . .	125
7.5	Measurement and integration . . . . .	128
7.5.1	Basic principles of measures . . . . .	128
7.5.2	Measurement on topological spaces and elements of integra- tion theory . . . . .	130

7.5.3	Representation theorem of F. Riesz . . . . .	152
<b>8</b>	<b>Coherences of approximation theory</b>	<b>154</b>
8.1	The theorems of Weierstraß and Stone . . . . .	154
8.2	The Bézier interpolation approach describing curves . . . . .	163
8.3	The Bézier interpolation approach describing surfaces . . . . .	179
8.3.1	The bilinear tensor product approach . . . . .	179
8.3.2	The approach based on triangular surfaces . . . . .	182
<b>9</b>	<b>Results of measure theory applied on the “Spatial-Toolbox”</b>	<b>192</b>
9.1	Fundamental considerations . . . . .	192
9.2	Processing an approximation of the functional coherences by the given data at the particular coordinates . . . . .	194
9.3	Interpolation using methods from computer-oriented geometry . . .	217
9.3.1	Structuring the given data within the regarded geographical area using <i>triangulation</i> . . . . .	218
9.3.2	Topological triangular Bézier surface interpolation . . . . .	222
9.4	Application of Kohonen Networks in the SPATTB . . . . .	242
<b>10</b>	<b>Conclusion</b>	<b>252</b>
	<b>List of Figures</b>	<b>254</b>
	<b>Index</b>	<b>258</b>
	<b>Bibliography</b>	<b>261</b>
	<b>Declaration</b>	<b>271</b>

# 1 Preface

## 1.1 Aim and methods used

This dissertation provides an interdisciplinary contribution to the project ReGLaN-Health & Logistics. ReGLaN-Health & Logistics, as described in chapter 2, is an international cooperation deriving benefits from the capabilities of scientists working on different fields. The aim of the project is the development of a so-called Spatial Decision Support System that supports decision makers working within health systems with a special focus on rural areas. In this dissertation, one important component for the development of the Decision Support Systems (DSS) named Early Warning And Response System is proposed and described in detail. This component called “**Spatial-Toolbox**” is developed with the intention of dealing with spatial data, i.e. data with additional geocoded information with regard to the special requirements of the Early Warning And Response System (EWARS).

An important component in the process of developing the EWARS is the concept of Geographical Information Systems. Classically, geocoded information with a vectorial character numerically describing spatial phenomena is managed and processed in a GIS. For the system GRASS that will be used within this project, these capabilities and possibilities are described in detail in [Net08] and for GIS in general, the book [Lan07] can be recommended. But for the development of the EWARS, the manageability of the type of data exemplarily given by  $(x, y, o)$  with

## 1 Preface

coordinates  $(x, y)$  and  $O_3$ -concentration  $o$  is not sufficient. As described in section 3.2 of chapter 3 and later on in chapter 9, the manageable data has to be extended to data of type  $(x, y, f)$ , where  $(x, y)$  are the geocoded information, but where  $f$  is not only a numerical value but a functional description of a certain phenomenon. An example for the existence and appearance of that type of data is the geocoded information about the variation of the  $O_3$ -concentration in time or depending on temperature. In section 3.1 of chapter 3, a knowledge-base as important subsystem of DSS containing expert knowledge is mentioned. This expert-knowledge can be made manageable when using methods from the field of fuzzy logic (more detailed information about them can be found in section 9.2). Thereby mappings, so-called fuzzy-sets, are generated. Within the EWARS, these mappings will be used with respect to additional geocoded data. The knowledge about the geocoded mapping information only at a finite set of locations  $(x_i, y_i)$  associated with mapping information  $f_i, i \in \{1, \dots, n\}$  is not sufficient in applications that need continuous statements in a certain geographical area. To provide a contribution towards solving this problem, in section 9.3 of chapter 9, methods from the field of computer geometry and Computer Aided Design, so-called Béziermethods, are used for interpolating this geocoded mapping information. Classically, these methods operates on vectors in the space  $\mathbb{R}^d, d \in N$ , but in terms of dealing with mapping information, there has to be an extension on topological vector spaces since mapping spaces can be defined as such spaces. This builds a new perspective and possibility in the application of these methods. Therefore, the according algorithms have to be extended; this work is presented in chapter 8.

The field of Artificial Neural Networks plays an important role for the processing and management of the data within the EWARS, where features of biological processes and structures are modeled and implemented as algorithms. Generally, the developed methods can be divided as usable in terms of interpolation

or approximation functional coherences and in such being applicable to classification problems. In this dissertation one method from each type is regarded in more detailed. Thereby, the classical algorithms of the so-called Backpropagation Networks for approximation and the Kohonen Networks for classification are described in chapter 4. In chapter 9, an extension of these algorithms is then proposed using coherences from mathematical measure-theory and approximation theory described in chapters 7 and 8. Thereby, the training data is of the type  $\{(f_1, c_1), \dots, (f_n, c_n)\}$  with mapping data  $f_i: X \rightarrow \mathbb{C}$  and associated vectorial data  $c_i, i \in \{1, \dots, n\}$ . Measure theory operating on mappings plays an important role if the data is of the following type:  $\{(K_1, c_1), \dots, (K_n, c_n)\}$ , where the  $K_i$  are compact subsets of the domain  $X$ , which is a topological vector space. The mentioned extension of these algorithms is based on a preprocessing of the mapping data using integration methods from measure theory.

## 1.2 Structure of the dissertation

In this section, the structure of this thesis is described to provide an overview for the reader.

**Chapter 2 Contribution to the project ReGLaN-Health & Logistics:** Here the aim and the objectives of the project are described in more detail and the contribution of this dissertation is specified. It describes which methods will be used, where the proposed SPATTB is located in the EWARS and which applications are possible in this context.

**Chapter 3 Embedding in the context of current research:** In the process of developing methods for handling spatial respectively geocoded data, the knowledge about the understanding of these terms in publications within the according scientific field is essential. Hence in this chapter, possibilities and

## 1 Preface

applications in the context of managing spatial data, especially referring to current publications, are described. Additionally, basic coherences about DSS are described with appropriate references to relevant parts of this thesis. Finally, the specific contribution of this dissertation to the scientific community is described highlighting its pertinence.

**Chapter 4 Artificial Neural Networks:** The field of Artificial Neural Networks is huge and growing rapidly. Within this chapter basic historical facts from the beginning through to actual applications and methods are described. In addition the biological background and a common model of an artificial neuron are depicted, followed by a possibility of classification. Subsequent to this, the relevant Artificial Neural Network (ANN)-models for this thesis, the Backpropagation Networks and the Kohonen Networks are characterised in more detail, describing mathematically the relevant algorithms on which the developments and extensions in chapter 9 are based.

**Chapter 5 Principles of object-oriented analysis:** Handling semi-structured or unstructured problems or dealing with complicated situations is a challenge in every scientific field. A way of dealing with this problem is the application of the so-called Object Oriented Analysis. This method provides a possibility of structuring complicated systems in a standardised way, that provides the possibility of simplifying the working relationships within a project. This method will also be used for the development of the EWARS. Important elements of Object Oriented Analysis (OOA) are depicted in this chapter. One of that is a so-called *object*, which exemplarily can be a representation of a concrete element within the regarded health system, for example a doctor treating a patient. For the process or even the viability of that treatment, several circumstances can play a role. E.g., the kind of disease can be mentioned, the distance between the two people, the available infras-



## 1 Preface

structure or the actual weather situation. If the doctor for example has the possibility of using telemedicine and the regarded disease allows this treatment, the availability of electric current and the risk of having heavy rainfall become essential information. So regarding them in terms of OOA, these circumstances have an important influence on the behaviour of the representing objects. If geocoded mapping information is the basis information for a relevant influence statement that has to be given in a continuous way, the methods for dealing with the kind of data proposed in chapter 9 play an important role for providing appropriate information to particular objects and influencing their behaviour.

In the process of describing the EWARS with methods from Object Oriented Analysis, the SPATTB should also be regarded in this way. The SPATTB itself represents an object-class with the ability of deriving several methods as sub-classes. The methods important to this thesis are the extensions of the Backpropagation Network (BPN)- and the Kohonen Network (KN)-algorithms for processing geocoded mapping information. These can therefore be regarded as subclasses of the SPATTB, which a user can call providing geocoded mapping information.

**Chapter 6 Proposal of a “Spatial Toolbox”:** Here the general structure of the SPATTB is proposed and a localisation within the whole concept of the EWARS is provided. Additionally, the importance of OpenSource software components within the whole system and particularly for the SPATTB is described. Finally, a workflow for the development of additional components for the toolbox is proposed and illustrated.

**Chapter 7 Measure theory on topological and function spaces:** Geocoded mapping information is considered as elements of a topological vector space of functions. When regarding data of the type  $\{(f_1, c_1), \dots, (f_n, c_n)\}$  already

## 1 Preface

mentioned above, the described preprocessing in terms of extending the algorithms from the field of ANN and the extended Bézier-interpolation require the theoretical background of measure theory. Characterizing differences in a mathematical way requires adequate methods of measuring. As mentioned above, there are components developed for managing geocoded mapping information, so if the differences should be described in this context, appropriate methodologies have to be used. In this chapter the basic facts about topological spaces and generalised concepts of convergence in terms of so-called *nets* and *filters* are given. Furthermore, the concept of topological vector spaces is characterised and finally the important methods from mathematical measure theory are described. Thereby, the approach *inter alia* mainly proposed by Bourbaki in [Bou04] based on linear functionals is used, but also shown is the coherence to the measuring approach based on systems of sets like  $\sigma$ -algebras. This coherence is described by the theorems of Riesz, whereof one formulation is described in this chapter.

**Chapter 8 Coherences of approximation theory:** As mentioned above, there is a proposed method for interpolating geocoded mapping information by approaches from the field of computer geometry and Computer Aided Design (CAD). In this chapter these approaches are described, but the according coherences are not depicted in the classical way as given in common literature, but in an extended way. Conventually, these methods are based on vectorial described elements out of a space  $\mathbb{R}^d, d \in \mathbb{N}$ . This concept is extended to gain the ability of elements out of an arbitrary topological vector space. This description is in chapter 9 for application of this on a special topological vector space, whose elements are of the type  $(x, y, f)$  with (geo-)coordinates  $(x, y)$  and a mapping of the type  $f : X \longrightarrow \mathbb{C}$ . The above mentioned fuzzy sets are of this type and the interpolation of them can then be realised

through this method.

**Chapter 9 Results of measure theory applied to the “Spatial-Toolbox”:** At this point, the described and partly extended methods from the fields of ANN, parameter-based Bézier-interpolation and measure theory are combined with respect to the operation of geocoded mapping information. Within this chapter, the already described approach of interpolating geocoded mapping information using Béziermethods is proposed. Furthermore, an algorithm is developed, which extends the classical BPN-algorithm described in chapter 4 for gaining the ability of training them but using mapping information. Finally, a similar extension is described for the KN-algorithm, where a programmed application as a component of the SPATTB based on the KN-algorithm and a proposed gearing of several OpenSource components (GRASS, maxima, R) is described. This application provides a possibility of gaining continuous statements about the density of the given discrete geocoded data.

## 2 Contribution to the project

# ReGLaN-Health & Logistics

The methods which are developed and described in this thesis should offer a contribution to the project ReGLaN-Health & Logistics. The description of the coherences in this context will be the topic of the current chapter.

The project ReGLaN-Health & Logistics (ReGLaN = Research Group Learning and Neurosciences) deals with the development and description of adaptive methods and algorithms in the context of health services and decision support.

The project was founded by the South African Gerhard Ackermann. The main aim is the development and provision of methods and applications for improving the health system especially of rural areas.

In this project there is an international collaboration of different institutions and people. The main persons leading this project are Dipl.-Ing Gerhard Ackermann, Prof. Dr. Engelbert Niehaus, Prof. Dr. Dr. Marlien Herselman, Dr. med. Ruth Niehaus and Dipl.-Ing David Niehaus.

The advantage drawn from the heterogeneity of those participating is the access to the expert knowledge and points of view in the particular scientific field.

One important part of our contribution to the optimisation of the described health systems is the development of a so-called *digital decision support system* (DDSS).

## 2 Contribution to the project ReGLaN-Health & Logistics

This should be used to provide decision support to the different decision makers within the regarded health system. In rural areas, particularly in non-industrial nations, the circumstances within which first-aid is applied are often not optimal. Therefore it is important to get the best results by using the resources available. The system will help to make a decision on the treatment of a patient by processing the existing resources and analysing the possible treatments.

The system that will be developed within this project is called EWARS. The focus here lies on the ability to early warn, especially in terms of epidemiology and in providing a logistically optimised response that optimises the usage of medical resources in rural areas. In [Nie09], Niehaus, Herselman and Babu write, “the objective of a spatial decision support provided by EWARS is to optimize the deployment of the existing limited resources in rural areas to improve health service and delivery according to risk”.

In this context, the EWARS is additionally under the umbrella of Action Team 6 of the United Nations, and has been presented to the Scientific and Technical Subcommittee of the Committee on the Peaceful Uses of Outer Space in Vienna ([Nie08]).

In this thesis, a so-called SPATTB is proposed with the aim of providing a contribution to the development of the EWARS. In figure 2.1, the processing within the EWARS is illustrated, where a so-called *expert knowledge layer* is visible. This expert knowledge has to be expressed in a mathematical way for further processing. One possibility to do this consists in using methods from fuzzy logic as described in section 9.2. With these mappings, so-called membership mappings, the grade validity of linguistic values as “warm” or “fast” can be described in a fuzzy way. Regarding geocoded data, such statements and mappings are located on certain geographical coordinates. This information should not only be used within the system on the given coordinates but also on the area “in between”. This thesis

## *2 Contribution to the project ReGLaN-Health & Logistics*

proposes a method for approximating the geocoded mapping information in section 9.3 for a whole area based on “discrete” mapping information only at a finite number of locations. This method is proposed as part of the SPATTB as described in chapter 6.

Coherences drawn from the expert knowledge in coherence with measured data are also of high importance for the processing within the system. An example is the relationship between expert knowledge about the amount of oxygen and temperature concerning a certain body of water and an associated number of insects acting as disease vectors, that develop in this body of water within a period of time. In section 9.2, a method of approximately describing such coherences as mappings that provide the ability to calculate the associated information for variations of the mappings is proposed.

The proposed SPATTB described in chapter 6 is thereby considered for interacting in different situations within the EWARS. Another field of application is described in section 9.4. The ability of providing suitable statements in the context of health and epidemics is obviously related to the available information; so regard to the original data and its relation is thus essential before drawing a conclusion on a certain application. In the mentioned section, a method based on an algorithm from the field of neuroempiricism and an according software implementation is proposed for describing the density of given measurement data.

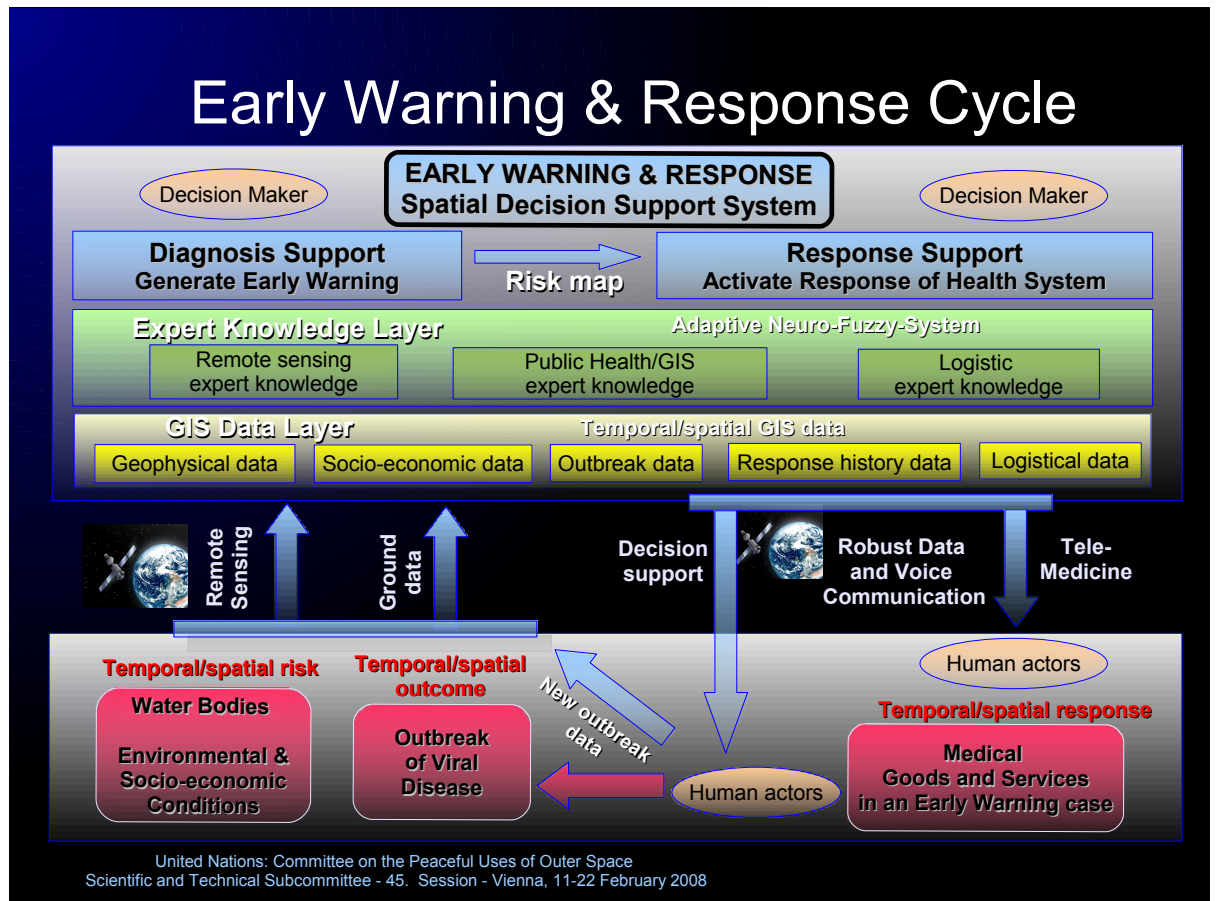


Figure 2.1: EWARS, [Nie08], p.14

# 3 Embedding in the context of current research

## 3.1 Description and limitation of the context of current research

### 3.1.1 General concepts of spatial analysis

As mentioned in chapter 2, the EWARS inter alia deals with *spatial data*. At this point, some general information about the processing of *spatial data* in the context of *spatial analysis* is given. Furthermore important characteristics related in terms of research and application in this field are pointed out and some explicit studies of current interest are mentioned.

In the introduction of [San07] it described that the terms *spatial analysis* and *spatial modeling* cannot be defined in a universal way. It is stated that within the modeling of spatial phenomena and processes, there is a need for scientific knowledge and skills from various fields. The special part of each discipline stands in close relation to the regarded application and/or modeling process. So on the one hand, different disciplines can provide methodologies and frameworks in the content of spatial analysis and spatial modeling. On the other hand, methods as



### 3 Embedding in the context of current research

well as models developed within different fields such as mathematics can be tested, applied and evaluated using spatial data.

In this context, the term *modeling* is important and thus its meaning and definition has to be outlined and defined. This term can be interpreted in different ways. At this point, the following general definition based on the one given in [Hag65] should be mentioned: “(...) *a model is a schematic representation of reality, developed with the goal of understanding and explaining it.*” ([San07], p.1). In section 4.2 the term *modeling* and in particular the term *mathematical modeling* is looked at in more detail.

In the context of dealing with spatial data and developing models, the relevant aim is of crucial importance, because the usage of methods and concepts is dependent on this. One example is the spatial dynamics of population growth, where the aim can only be the description and a clear presentation of the given data or it can be an attempted explanation or even a forecast. In addition to this, the type of data given is also important because for example methods for computing discrete data in general cannot deal with non discrete data directly.

In chapter 9 extensions of algorithms that provide the ability of not only dealing with vectorial data but also with geocoded mapping information are proposed.

Throughout modeling in terms of spatial analysis, the following steps are important (basing on [San07], p.2):

- Regard to the real situation, detection of the existing logical structure and making a choice about the relevant components.
- Formalizing the given facts and chosen components.
- Undertaking a calculation and/ or processing in order to gain concrete results from the model/ the models.
- A comparison of these results with empirical observations and evaluation of

### 3 Embedding in the context of current research

the developed model/ models.

- Highlighting the “gaps” between modeling and observations and trying to improve the model/ models.

The study of these gaps, which in some literature are called *residuals*, can suggest the usage of new models or lead to a modification of the developed model/ models e.g. by taking into account additional important parameters and factors.

These steps are very similar to the ones in the so-called modeling cycle for mathematical modeling mentioned in section 4.2.

Spatial models can be represented and illustrated by so-called *flow-diagrams* or *sagittal-diagrams*.

According to [San07], p.16 et seqq., the following criteria or characteristics for the classification of spatial models can be stated:

- The model is *aggregated*, i.e. a group of factors and parts are regarded as whole system, or the model is *disaggregated*, i.e. they are regarded individually.
- The model is either a static or a dynamic one.
- The model is deterministic and/ or probabilistic.

No model in spatial analysis can create a perfect representation of reality. Hence it makes sense in many applications to not only use one kind of modeling, but to take different models into account. This way a multi-perspective view on the regarded situation can be achieved. In doing this, the advantages and disadvantages of each used model and methodology must not be disregarded to get appropriate results.

Dealing with spatial data, there is often a need to regard the factor of *time* as well ([San07], p. 97) highlighting the process which evolve the regarded spatial structure. In the consideration of time, two main approaches can be mentioned.

### 3 *Embedding in the context of current research*

On the one hand, the evolution of the spatial structure is in the foreground and time is regarded for measuring the differences of certain periods. As examples models of urban growth or diffusion models in epidemics can be given. On the other hand, the time of the individuals is regarded, while space is regarded as a variable in their life path. This approach is applied in demographic models using biographic data.

In the following, some methodologies of spatial modeling are described. According to [San07], p. 160 et seqq., one possibility of gaining an understanding of a complex system consists of regarding the basic entities of this system. This idea of so-called *spatial microsimulation models* goes back to Orcutt, who attempted an employment forecast in the 1960s not by regarding a high macroeconomic model but looking at the household level. With the increasing development of computers, this kind of modeling became more and more relevant for multiple disciplines like population dynamics, epidemics and even for computer graphic technology (refer e.g. on [Nvi]).

The basic idea of microsimulation is the formalising of change at an individual level and in this way describing the whole system. This formalising can be realised in two different ways. The first approach uses *statistical procedures*. The models developed therewith are based on probabilistic rules for the impact of certain events on certain types of individuals in certain situations. An example from the field of population dynamics is the case of a women giving birth to a baby in a certain period of time. This depends inter alia on her age, her income and her education. Population models using this type of approach are the CORSIM (USA), the DYNACAN (Canada) and the SVERIGE (Sweden).

The other approach mentioned above is the *agent based* or *multi-agent systems*. Here, there are defined rules for the behaviour of the regarded individuals, but them is given a certain degree of autonomy. In this way, the used agents can

### 3 Embedding in the context of current research

interact with each other following the given rules. At this point, two definitions of multi-agents are given. Durfee defines the following in [Dur89]:

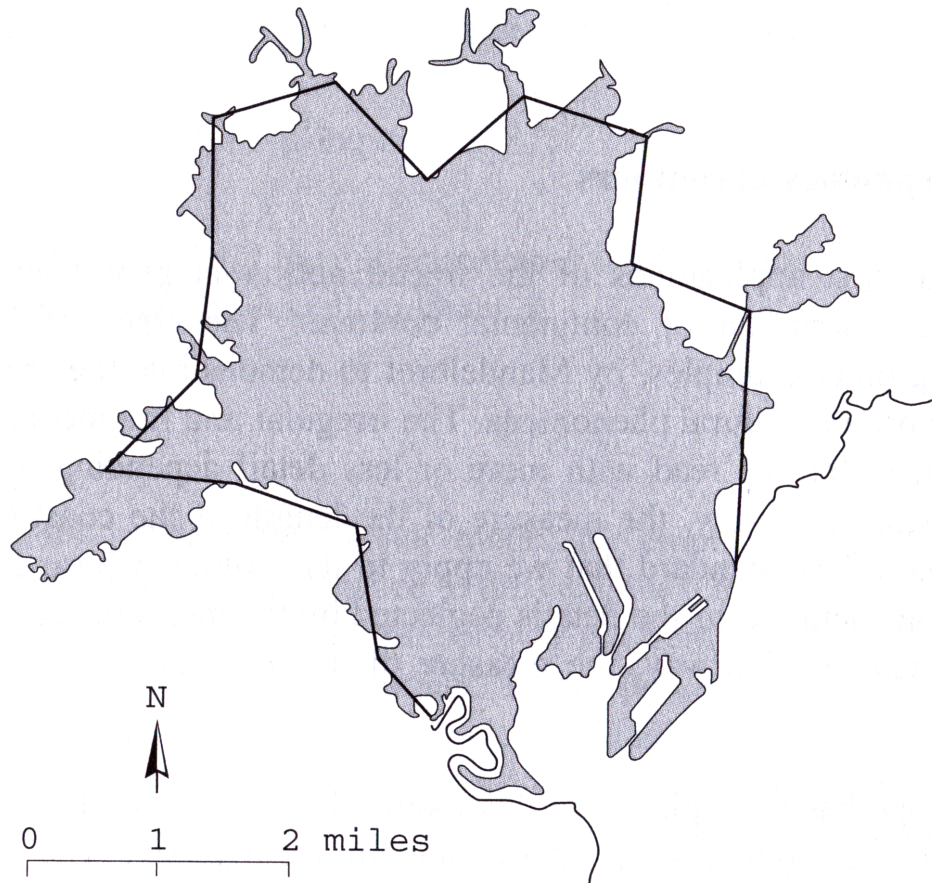
*A Multi-agent system (MAS) is a network loosely coupled with entities acting together to resolve problems that were beyond their individual capabilities.*

Another definition is given by Wooldridge in [Woo02]:

*An agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives.*

An example for the usage of MAS is located in the field of epidemiology. The spreading of a disease distributed by a certain virus can be described using the so-called *SIR approach* based on the work of Kermack and McKendrick (refer to [Krä03] and [Prü08]). In this approach the three stadiums *susceptible* (S), *infectious* (I) and *resistant* (R) are regarded. The passage from one stadium to another can be described using rules and therewith, the spreading of the disease can be described by multi-agents. There are several simulations in existence but exemplary is [Xjt] and should be mentioned at this point.

Furthermore, the field of *fractals* is regarded as a supply of methodologies, which is useful for certain problems in spatial analysis. According to [San07], p.281 et seqq., this geometric approach is used in certain applications dealing inter alia with contour problems. Mandelbrot gave the following definition in 1982 (see [Man82]): “(A fractal is) a rough or fragmented geometric shape that can be split into parts, each of which is (at least approximately) a reduced-size copy of the whole.” For more information about this field refer to [Man82] or [Pei04] for example. In figure 3.1, an application of fractals in terms of spatial data is illustrated.



**Figure 3.1:** Fractal analysis of the urban perimeter of Cardiff ([San07], p.296)

### **3.1.2 Usability of ANNs and GIS in the context of spatial analysis**

In actual research ANNs are used in many contexts. In section ??, there is given an overview of important applications and according references is given. In this section actual applications in the context of spatial data and GIS are discussed. In these studies, some special types of ANNs are used in the according situation with different results. The reason for mentioning these studies at this point is on the one hand to clarify the actuality of ANNs and the resulting necessity of conducting research in this field in a theoretical and practical way. On the other hand, I would like to point out that there are real situations for the application of the SPATTB as a generalised concept.

In 1996 Zhou and Civco published an article ([Zho96]) in which they described one of their studies. In this study they used sample data of their GIS IDRISI ([idr]) and an exercise out of the instruction manual concerning the optimal location of a light manufacturing plant. As the name of the article suggests, they used an ANN with a genetic learning algorithm and compared the results with those from conventional methods such as multi-criteria evaluation produced in this case. A detailed description of this method is not given at this point, but for further details the reader can refer to [Van77]. They pointed out, that there are certain problems using the conventionell methods. There is the heterougenity of spatial data, the required knowledge for the usage of these methods and the mostly complex user interfaces.

In this study, they used different network configurations for the calculation of a grade of suitability based on multi-dimensional information like slope or forest coverage. Comparing the results with those from the conventional methods produced,

### *3 Embedding in the context of current research*

they drew the conclusion that the ANNs are in this case more appropriate in dealing with the inaccuracy of the data. In addition they mentioned that in their opinion the user interface concerning the ANNs is less complex and less difficult to use, because the main interaction of the user is the input of training data. As a main result they pointed out that for them, the used ANNs are a real alternative to the conventional methods.

Lee, Ryu, Won and Lee studied landslides in a corean area called Boun. In this context they undertook a study where they used a special type of ANNs called Backpropagation Networks (BPNs). Their goal was to detect susceptible areas in which landslides can occur, using several information sources such as aerial photographs, data on the soil type, timber cover and land use. The results they obtained were published in 2003 in [Lee03]. With the available data, they trained the BPN and made a susceptibility map of the regarded area using the network data in combination with a GIS. To verify their results, they compared this map with actual landslide occurrences. They used three different networks, each basing on a different number of input factors such as slope, aspect and soil material, and drew the conclusion that the network using eight factors is more suitable to this case than the ones with fourteen or four factors. To determine which of the fourteen factors were available to use, they used a certain algorithm described in the article. As a main result they pointed out that the ANN they used can be applied for effectively producing a susceptibility map in the regarded area and for gaining qualitative results.

For the usage of wireless communication especially in Wireless Local Area Networks (WLAN), the electromagnetic coverage is essential. In this context, Şen, Gümüştay, Kavas and Bulucu completed a study in 2008 on indoor radio wave propagation using a GIS database and a BPN for interpolation. Therefore, they used the electromagnetic field values of 1085 observation points and stored the data

### 3 Embedding in the context of current research

in the GIS database. The applied BPN uses the three dimensional coordinates of the observation points and calculates a number that represents the intensity of the electromagnetic field. To verify the results they divided the data into 672 observation points for the training and 413 points for testing. In addition, they undertook a comparison of the results produced by the BPN and the results obtained through a statistical interpolation method named *kriging* (for more details refer e.g. to [Fah09], p. 331) and came to the conclusion, that both methods produced suitable predictions, but the results of the kriging were more accurate. However they drew the conclusion that the use of such a BPN is viable in this context because of the underlying learning algorithm, which allows an update of the existing interpolation using new data easily.

This study is a good example to show the importance of a generalised description of ANNs and has the ability to compare different models in a better way. A generalisation of the BPN-algorithm is proposed in chapter 9 for extending its capabilities.

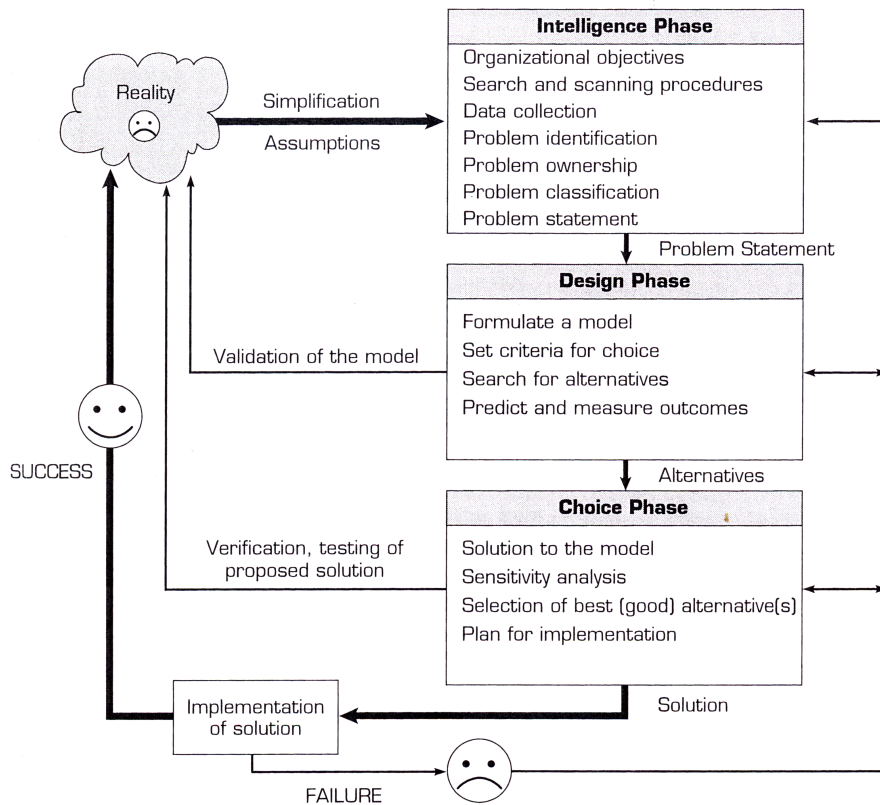
#### 3.1.3 General concepts of decision support systems

According to the introduction of [Tur00], DSS is nowadays becoming more and more important, especially due to the possibility of providing access to databases and web-based applications. In this subsection, basic information about DSS is provided based on [Tur00].

People have to make decisions every day. Thereby, in most cases there has to be a choice among several alternatives with the aim of achieving possibly the best results for the regarded problem. The term “best” in this context depends strongly on the problem, the available alternatives and the aims and abilities of the decision maker respective to the decision maker. Turban et al. mention ([Tur00], p.41):



### 3 Embedding in the context of current research



**Figure 3.2:** Phases of decision making ([Tur00], p.50)

“each discipline has its own set of assumptions about reality and methods. Each also contributes a unique, valid view of how people make decisions. Finally, there is a lot of variation in what constitutes a successful decision in practice.”

The process of decision making can be classified according to Simon ([Sim77]) in four phases: *intelligence*, *design*, *choice* and *implementation*. In figure 3.2 the decision making process is illustrated. The intelligent phase includes the identification of the problem and the formulation of the aim or aims that should be achieved. Furthermore, the availability and quality of data is analysed. In [Tur00], p.54, it is mentioned that the collection of data is one of the most difficult steps in analysis. In the design phase, models are generated with look at possible alternatives. Ad-

### 3 Embedding in the context of current research

ditionally here, the principle of choice has to be selected to define criteria for the evaluation of the model outcomes. At this point it should be decided what a model has to accomplish and what restrictions can be accepted in the regarded context. In the choice phase one or more of the found solutions are chosen. Thereby, a sensitivity analysis to determine the robustness of the choices has to be performed. Furthermore, a plan for implementation is made and realised in the next implementation phase. As illustrated in figure 3.2 the outcomes of implementation are used as logical reasons to adapt the decision and possibly reiterate some of the described phases to improve the results.

Turban et al. describe in [Tur00], p. 27 et seqq., that computerisation of the decision process can improve it. On p. 23 et seqq. and additionally on p. 72 et seqq., how the single phases can be supported by different technologies and computerized algorithms is described and methods from the field of Artificial Intelligence (AI) are especially mentioned. This field is illustrated in figure 3.3, where three aspects are highlighted: experts systems, fuzzy logic and Artificial Neural Networks among others relevant for the EWARS. In the course of this thesis, the terms fuzzy logic (section 9.2) and Artificial Neural Networks (chapter 4) will be regarded in more detail in the coherence of dealing with spatial data and particularly geocoded mapping data. As mentioned in chapter 2, expert systems play an important role in the EWARS.

Figure 3.4 illustrates important key characteristics and capabilities of Decision Support Systems. Dealing with semistructured and unstructured problems mentioned in this figure will be realised within the description of the EWARS using the advantages of Object Oriented Analysis described in chapter 5.

Basing on [Tur00], p.109 et seqq., the following subsystems of a DSS are important:

(DSS-S1) *Data-management subsystem* for storing and managing the relevant data.

### 3 Embedding in the context of current research

- (DSS-S2) *Model management subsystem* managing necessary models from relevant scientific fields.
- (DSS-S3) *User interface subsystem* for managing the interaction with possible users.
- (DSS-S4) *Knowledge-based management subsystem* supporting other subsystems as a repository of expert knowledge or acting as an independent component.

Coherences between these subsystems are illustrated in figure 3.5.

Regarding the conception of the EWARS, an important component of the (DSS-S1) is the Geographical Information System GRASS and the basing database, where geocoded information is stored and administrated. Thereby, the database-type is not restricted to the specific GRASS-database, but there is the possibility of using other interfaces e.g. to a MySQL- or PostgreSQL-database. The elements within the (DSS-S2) *inter alia* will be integrated as scripts (e.g. written in PHP or Perl) accessible by appropriate web-interfaces which belong to the (DSS-S3). In section 9.4 of chapter 9, a component of the (DSS-S2) as a collection of Perl-scripts is developed, which make use of the OpenSource softwares *maxima* and *R* interacting with GRASS using the according web-interface. The expert knowledge within the (DSS-S4) will be implemented using methods from fuzzy theory. Thereby, geocoded mapping information, which can be handled with the methods developed in chapter 9 of this thesis is generated.

#### 3.1.4 Decision support in current research

Decision support and the development of decision support systems plays an important role in current research concerning industry regarding the field of data warehousing and also the field of epidemics and medical informatics.

In this subsection, there are mentioned representative examples from both fields.

### *3 Embedding in the context of current research*

As it is described in [Nem02], DSS plays an important role in the context of data warehousing within firms. The authors mention, that information is conventionally stored and provided by data warehouses, but that there also is a huge amount of implicit knowledge that is not taken into account. So, they propose a so-called knowledge warehouse, that “enhances retrieval and sharing knowledge across the organisation.”

They deal with so-called knowledge management for retrieving explicit data basing on implicit one. In this article, the regarded data is text data and the process of knowledge management concerning this is described and proposed to be supported by methods from the fields of IT and AI (a good overview of that field is illustrated in figure 3.3).

The authors are regarding the following different phases of knowledge management: Sharing tacit (implicit) knowledge, converting tacit knowledge to explicit knowledge, converting explicit knowledge to new knowledge and learning new knowledge. Thereby, methods from AI, especially ANNs (also see chapter 4), are proposed to be used more or less within a certain phase. Exemplarily, in the phase of converting explicit knowledge into new knowledge, methods for data clustering are mentioned whereby the Kohonen algorithm (also see chapter 4) is one possibility. As written down in the article, the main goal is to provide and/or propose an “intelligent analysis platform” enhancing all phases of knowledge management. In this proposition the inclusion of AI methods should “amplify the cognitive capabilities of the decision maker”. This knowledge warehouse is proposed there as an extension of existing data warehouse systems like Warehouse 5.0 from SAP. Yang et al. published an interesting approach in [Yan07]. There, they propose a DSS concept basing on the GIS ArcGIS using the technologies COM or additionally DCOM, ASP and ActiveX for the prevention of epidemic diseases. It is especially

highlighted the fact of inadequate integration of epidemic models in Geographical Information System on which is mainly focussed in that article. In figure 3.6, there proposed system configuration is illstrated. There, the approach of implementing epidemic models as SIS/SIR models using the Component Object Model (COM) technology developed by Microsoft, is visible.

In a study for improving vector control programmes in Madagascar, Rakotomanana et alt. draw the conclusion, that “a geographical information system is a potentially valuable tool for decision-making and optimising interventions” (see [Rak07]).

## 3.2 Integration of this thesis

In this thesis, a SPATTB is proposed and described. This will be a part of the Spatial Decision Support System EWARS that is developed within the project Research Group Learning and Neurosciences (ReGLaN)-Health & Logistics (see chapter 2). An important tool in this context is the GIS GRASS, that will be used for the storage and the visual representation of the data. In common GIS software like GRASS ([GRA10]), raster- and vector-data are managed and stored. With vector-data, elements like boundaries, streets or cities can be represented, whereas raster-data is used for describing continuous data such as the amount of pollutants  $(x,y,p)$ , the altitude  $(x,y,a)$  or the pH-value at any location  $(x,y)$  in the regarded area (see [Lan07], p.45 et seqq.). But as described in chapter 2, the handling of  $(x,y,f)$  as geocoded mapping information  $f : X \longrightarrow \mathbb{C}$  at location  $(x,y)$  is essential within the concept of the EWARS. In chapter 9 solutions for this problem are developed and described. Therefore the approaches proposed and described are innovative in this field and extend the ability of dealing with geocoded data especially in the context of GIS-useage. In section 9.2, an algorithm from the field of Artificial Neural Networks, that is classically used for the approximation

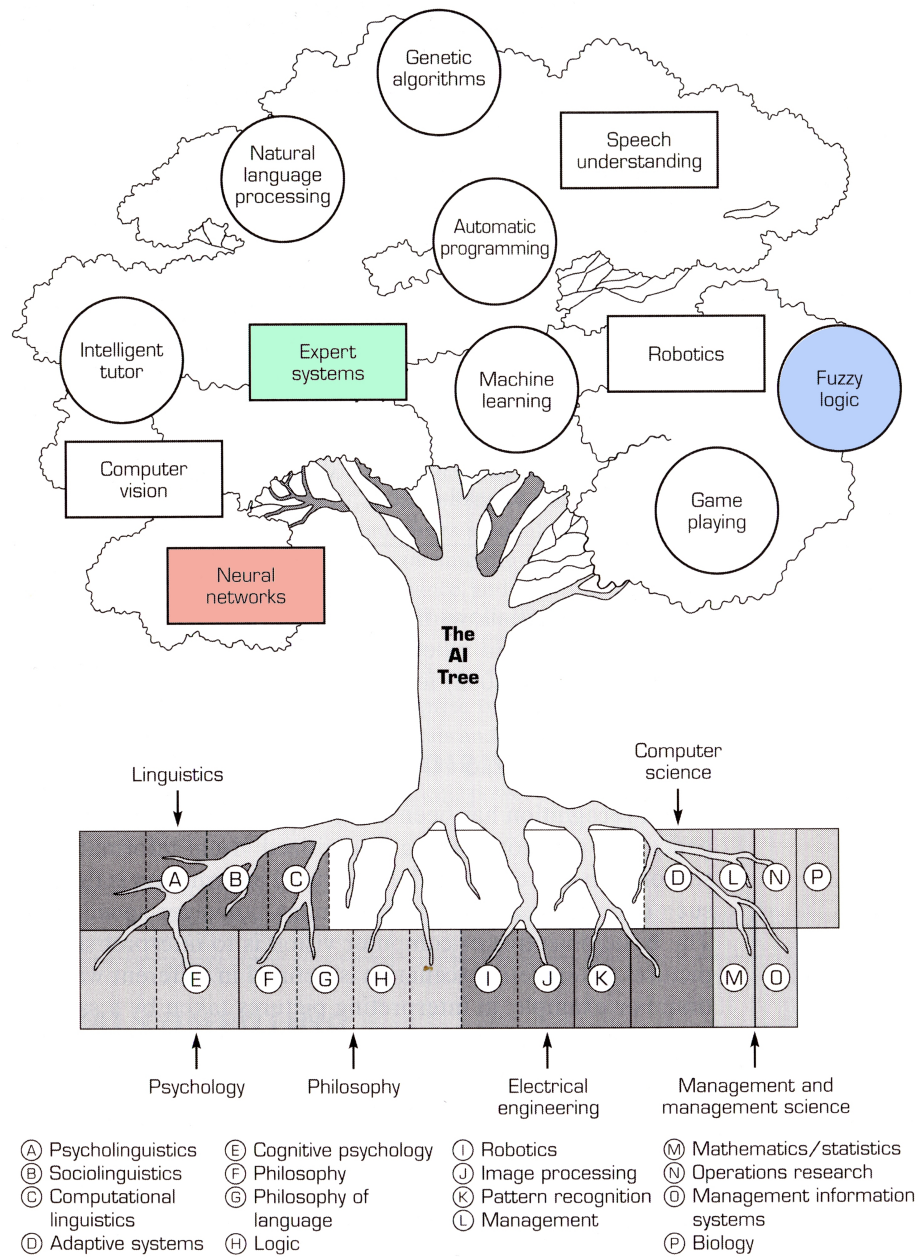
### *3 Embedding in the context of current research*

of functional relationships basing on discrete vector data, is extended to gain the ability of handling mapping information. A possibility of gaining continuous statements about geocoded mapping information within a certain geographical area is described in section 9.3. Thereby, geometrical methods from the field of Computer Aided Design described in section 8.3 of chapter 8 are used.

To gain the ability of classify mapping information in clusters, the classical Kohonen-algorithm (see section 4.3) is extended based on the propositions made in section 9.2.

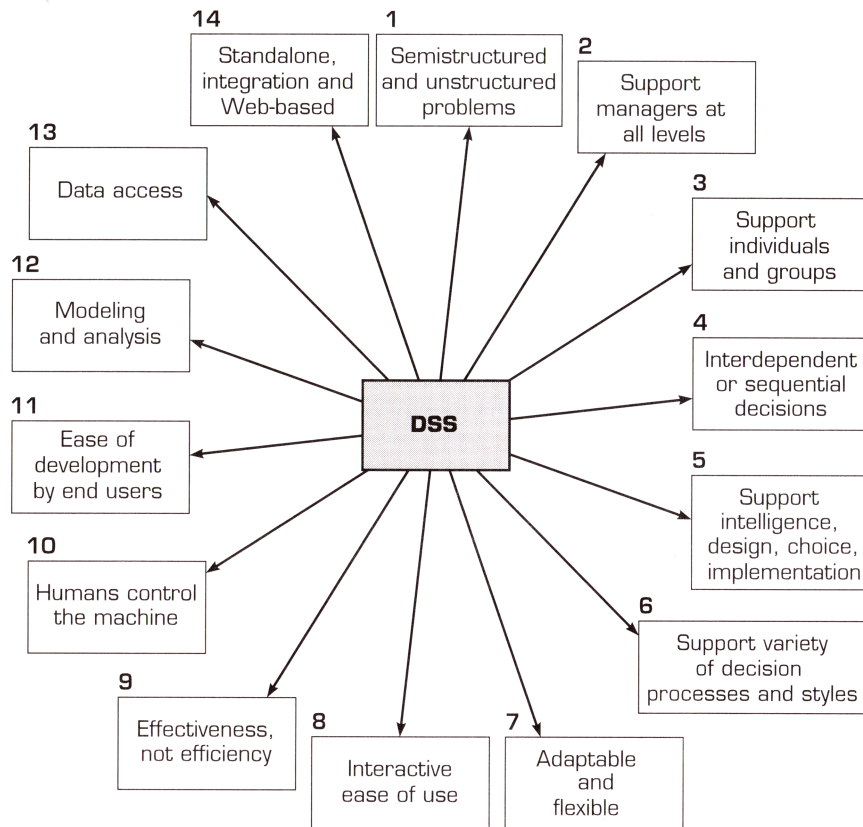
There, methods from measure theory operating on mappings are used in a similar way as when extending the approximation-algorithm as mentioned above.

### 3 Embedding in the context of current research



**Figure 3.3:** Overview of the scientific field of AI with possible applications ([Tur00], p.545, modified)

### 3 Embedding in the context of current research



**Figure 3.4:** Key characteristics and capabilities of DSS ([Tur00], p.107)



3 Embedding in the context of current research

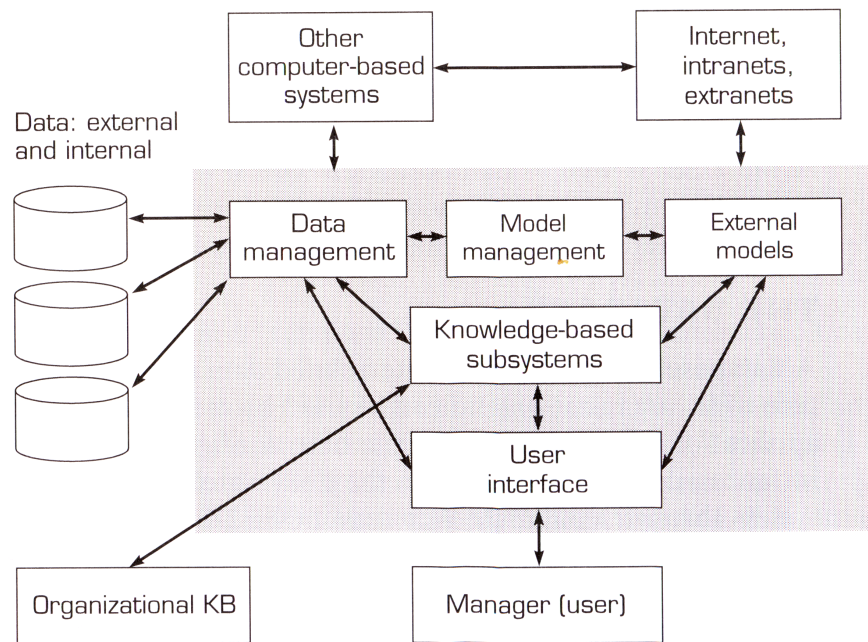


Figure 3.5: Schematic view on subsystems of a DSS ([Tur00], p.109)

3 Embedding in the context of current research

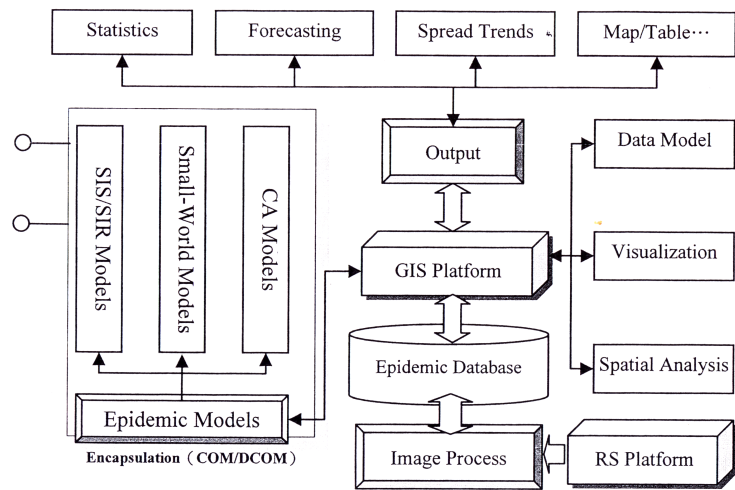


Figure 3.6: Proposed system configuration ([Yan07])

# 4 Artificial Neural Networks

## 4.1 Historical development

In this section a description of the historical development of the field of Artificial Neural Networks is given. According to [Zel96], p.23, artificial neural networks can be described as systems for processing information, which consist of a more or less huge number of atomic units called neurons.

In general, there are two main motivations to examine this field. Many models are based on biological models of the information processing of a mammal's brain. So some scientists tried to gain new knowledge on biological neural networks by regarding connected artificial neurons.

Another motivation is based on the fact that artificial neural networks can be regarded and treated as massively parallel systems which can be used as a special kind of algorithms in different fields of science such as physics, informatics, bioinformatics and mathematics. This connection of different scientific fields with the field of artificial neural networks is described in [Zel96], pp.23-24.

Zell mentions the ability of "learning" as a main feature of an ANN which is the main reason for the huge amount of possible applications. Some learning algorithms will be part of the next section.

The following summary of the historical development of ANNs is based on the

#### 4 Artificial Neural Networks

description in [Zel96], pp.28-33.

Neural networks are not only used from the perspective of mathematicians, many scientific areas can use them for research in their respective fields. Thus the models partially orientate themselves to the biological model, which is why biologists, psychologists and doctors can use particularly similar models to better understand biological neural networks.

Besides information scientists use neural networks as massively parallel algorithms for carrying out many arithmetic steps at the same time. Physicists use them to gain the ability to better describe complicated functional coherences of physical models. In this possibility where neural networks are considered as functions, mathematics is also interested. It is for this reason that it is even more exactly entered in this thesis on neural networks. In addition, certain neural networks are able to learn using training patterns and thus a special meaning comes into consideration of suitable learning algorithms (compare [Zel96], p. 23-24).

The development of neural networks since 1943 will now be described. In this year, the article “*A Logical Calculus of the Ideas Immanent in Nervous Activity*” was published by **McCulloch and Pitts** in the journal *Bulletin of Mathematical Biology* ([McC43]), which describes a model that is able to represent every limited boolean function ([Pat00], p. 22). Some years later (1949) Hebb published a theory where the connection of two nerve cells becomes stronger, if the synaptic activity becomes bigger. This is called “**Hebb’s learning rule**”.

To simulate the processing of sensory inputs, Rosenblatt developed a special form of neural networks, so-called **Perceptrons**. He published his knowledge in 1958 ([Ros58]). The basic construction exists of a layer which takes up photo-sensory inputs and passes them to a processing layer. Aside from this, different connections of the nerve cells of the input layer exist with those of the processing layer. Over

#### 4 Artificial Neural Networks

these connections, signals are passed on and compared by the respective nerve cells of the processing layer to a threshold value. If the signal is big enough, 1 is passed on to the so-called output layer, otherwise 0. Then, the nerve cells of the output layer calculate the output of the whole neural network ([Pat00], S.23).

Moreover, Rosenblatt developed a learning algorithm for the Perceptron, which allows the learning of certain outputs. The discoveries of this time enabled many scientists to suppose that thereby the bases of self-learning systems were already known.

Nevertheless, this acceptance was disproved in the book *Perceptrons* by Minsky and Papert ([Min69]). They carried out an exact mathematical analysis of this kind of neural networks and showed that they cannot solve many problems at all, especially the so-called “XOR-problem”. The interest in the field of research of the neural networks thereby decreased and many scientists turned to other areas. However, some did not cease their research and thus compiled the base for many models that were developed later. At this point, only some of them should be named. The **delta rule** developed by Widrow and Hoff is a specialised case of the Backpropagation Network-algorithm, which will be looked at in more detail further below.

The mathematician Teuvo Kohonen published works about the so-called self-organising maps which can adapt themselves by unsupervised learning independently to different input patterns (compares [Zel96], p. 31). The different possibilities of the learning with neural networks are explained in [Zel96], p.93 et seqq.

Rumelhart and Hinton made an important contribution to today’s interest in neural networks with the publication of the Backpropagation rule in 1986 ([RHW86]). This was already discovered some years before (1974) by Werbos in his thesis

([Wer74]), but was not perceived until later by a broader public. By this algorithm it was possible to also solve some of the problems which could not be solved by the Perceptrons ([Pat00], p. 25).

Newer research deals with the recognition of patterns, for example Fukushima and his colleagues developed the so-called **Neokognitron** which is able to process handwritten texts. Another example is the ability of robots to orientate themselves in the available space (compare [N.91], p. 6).

## 4.2 Fundamental mathematical modeling of Artificial Neural Networks based on Biological Neural Networks

In this section, the fundamental ideas and approaches of assigning the concepts of Biological Neural Networks (BNNs) to ANNs are presented. In this regard, the term *modeling* plays a decisive role. Hence, in the first part of this section, information about modeling in science is pointed out whereupon the application to ANNs is regarded.

According to [Imb03], p.4-9, the term *model* is in close connection to the term *system*. In ancient Greek, the word *systema* stands for *integrated whole*, which means a combination of different elements and processes, that are naturally given or human made. In this context, the *system boundary* is important but also are the *inner relations* of the regarded system. The following quote illustrates this in a significant way: “A system is more than its elements sum” ([Imb03], p.4).

Systems can be very simple but also very complex. An example in the context of this thesis is the health system in rural areas, which contains complicated inner

## 4 Artificial Neural Networks

relations. Important in this section are Biological Neural Networks as part of the human brain and the dealing with information within them.

To analyse systems, *models* can be used. In [Tur00], p.47, it is said that: “Models can represent systems or problems with various degrees of abstraction”. Imboden and Koch ([Imb03], p.8) illustrate the relationship between a system and a model with the association of glasses. A model is the glasses we use for regarding a system. Obviously, there is not only one possible model for a certain system. The construction of the model eachtime is related to the point of view or the perspective of the creators and their meanings. So when modeling a health system, a very rough point of view can consist in only considering the number of treated persons. An even finer model would take the different layers of this system and their relationship into account.

With regard to a Biological Neural Network (BNN), it is similar. A “simple” model would for example only regard the input of information and the reaction of the whole network. But if the inner computation of this network is taken into account, the model would be more complex.

If mathematics are used to develop a model, one could name it a *mathematical model*.

The stages of mathematical modeling are clearly arranged as illustrated by figure 4.1.

For more information about mathematical modeling regarding applications in ecologies and economics, refer to [Son01] and [Kra97].

At this point, basic information about the biological information processing in neurons and BNNs are described following a possibility of modeling some of these coherences in terms of Artificial Neural Networks.

**Biological neurons** are used for gathering information, processing and trans-

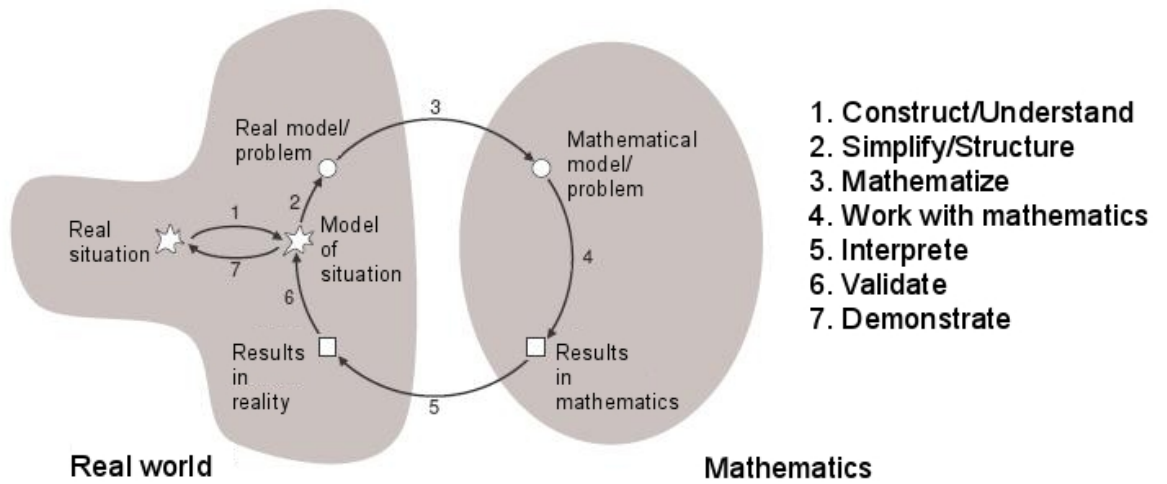


Figure 4.1: Modeling cycle ([Blu05])

porting it. An important part of each neuron is the so-called **soma** (cell body), which contains the nucleus and is responsible for the cellular metabolism. From the soma, the neurites starts. These are thin appendages and they are divided into the two main categories of **dendrites** and **axons**. Dendrites are branched out. Their task is it to get signals from other neurons and lead them to the soma. After processing, signals are transported from the soma by the axons. The bundled

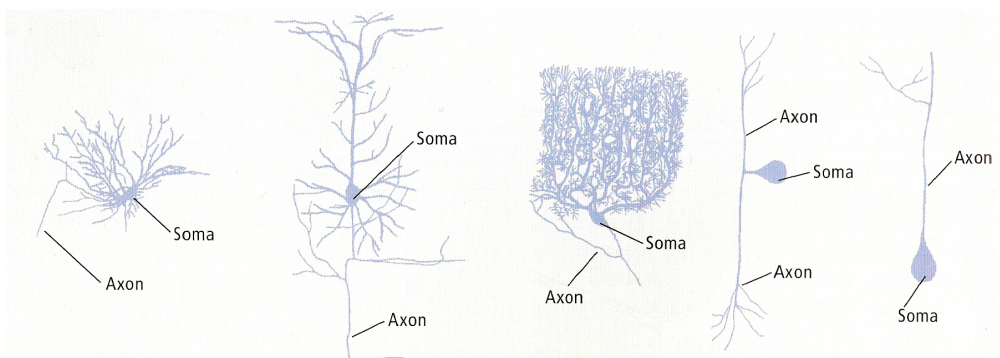
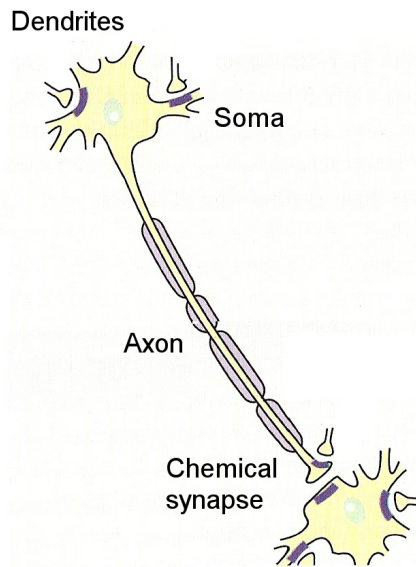


Figure 4.2: Different occurrences of nerve cells ([G.99], S. 48)





**Figure 4.3:** Important components of a nerve cell ([G.99], p. 92, modified)

up axons which leave the central nervous system and move through the body are called nerves (see [R.00], p. 133). Different types of neurons are shown in figure 4.2 and the most important parts of a single neuron can be found in figure 4.3.

The **equilibrium rest potential of the membrane** denotes a voltage (70-90 mV) between the inner and the surface of the cell. This is a reason for an asymmetrical ion distribution between both rooms. Ions are electrically loaded particles which originate from the atoms that take up electrons or deliver them. Moreover, the permeability of the membrane is different for different kinds of ions. Aside from this, the concentration of the  $K^+$ -ions and the protein-ions is greater in the inner than in the outer where the concentration of  $Na^+$ - and  $Cl^-$ -ions is greater. This different ion distribution originates from the fact that an equilibrium appears between the existing concentration slope and the counter-acting electric potential. The rest potential is formed by the equilibria of the involved ions and the respective concentration gradients are maintained by the ion pumps. Nerve cells and

#### 4 Artificial Neural Networks

other cells have the ability to bring on a so-called action potential ([R.00], p. 152). This consists of a short depolarisation up to possibly 30 mV and the following repolarisation. Thereby, a posthyperpolarisation appears, before the rest potential appears again. In addition, the passed on information is encoded by the frequency and the unloading pattern.

For a nerve cell to pass on a stimulus, the nerve cell must be depolarised strongly enough so that a threshold value is crossed and an action potential can be generated. With weak stimuli no action potential is brought on because the threshold value is not crossed. Numerous  $Na^+$  canals on the body and the axon hill are opened by the incoming action potential by which the conductivity of the membrane is increased to  $Na^+$ . Through this  $Na^+$  follows the concentration gradient by which another depolarisation takes place and therefore increasingly  $Na^+$  canals are opened. This is called the Hodgkin cycle. The whole process forms the behaviourdepolarisation phase of the action potential (see [H.96], p.132-160).

After the aperture of the  $Na^+$  canals, the  $K^+$  canals are activated with a short delay and at the same, time the  $Na^+$  canals are closed. This leads to the fact, that  $K^+$  flows out outwardly. Through this, the membrane potential becomes negative again. One calls this process the repolarisation-phase of the action potential. The activation of the  $K^+$ -and the deactivation of the  $Na^+$  canals continues for some milli-seconds, then the  $Na^+$  canals return in an activatable and closed state, the  $K^+$  canals shut but nevertheless do not become inactive. The control of the  $K^+$ -and the  $Na^+$  canals is dependent on the actual voltage.

Often it still comes to a posthyperpolarisation by which for a short time no other action potential can be brought on, because the  $Na^+$  canals are still inactive. This is called an **absolute refractory period**.

An action potential is passed on within an axon as follows (see figure 4.4). Within

#### 4 Artificial Neural Networks

the axon it comes to local currents beginning at the location where only just an action potential appears. These are then passed outwardly and are led back by the  $K^+$  canals. Through this a depolarisation appears in the neighbouring areas of the membrane by which an action potential is also brought on. By the period of absolute refractory it is prevented that an action potential can be passed on in the opposite direction (compare [R.00], pp.134-170).

In the course of the evolution, the original axons without a mark which own no boundary have developed themselves into myelinated fibres. These are surrounded by the so-called myelin sheath which are separated by the nodes of Ranvier. This development is important for the quick excitation forwarding in the vertebrates brain.

Within the group of vertebrates a quick further pipe of the action potential is possible with the myelinated axons. The myelin sheaths have a big ohmic resistor by which the electrotonic propagation within an axon is improved. Therefore, a depolarisation can only appear on the nodes of the Ranvier lying in between. Through this, the so-called **saltatoric conduction of impulses** occurs.

Now after considering of the forwarding of information by a nerve cell, it should be explained how this information is transferred between the nerve cells. It was identified, that no direct connection exists between the axon of a nerve cell and the dendrite or the cell membrane of another nerve cell, but that at this point a small space exists. This crossing of two nerve cells is called chemical synapse and consists of the membrane of the first nerve cell (presynaptic membrane), the space itself (synaptic gap) and the membrane of the second nerve cell (postsynaptic membrane). At this point, the electric signal passed on in the axon is converted into a chemical one. This separation causes the transport of information to always be directed. Based on [G.99], p.91 these synapses are additionally perceived to undertake the task of learning and minding, because it has been identified that

4 Artificial Neural Networks

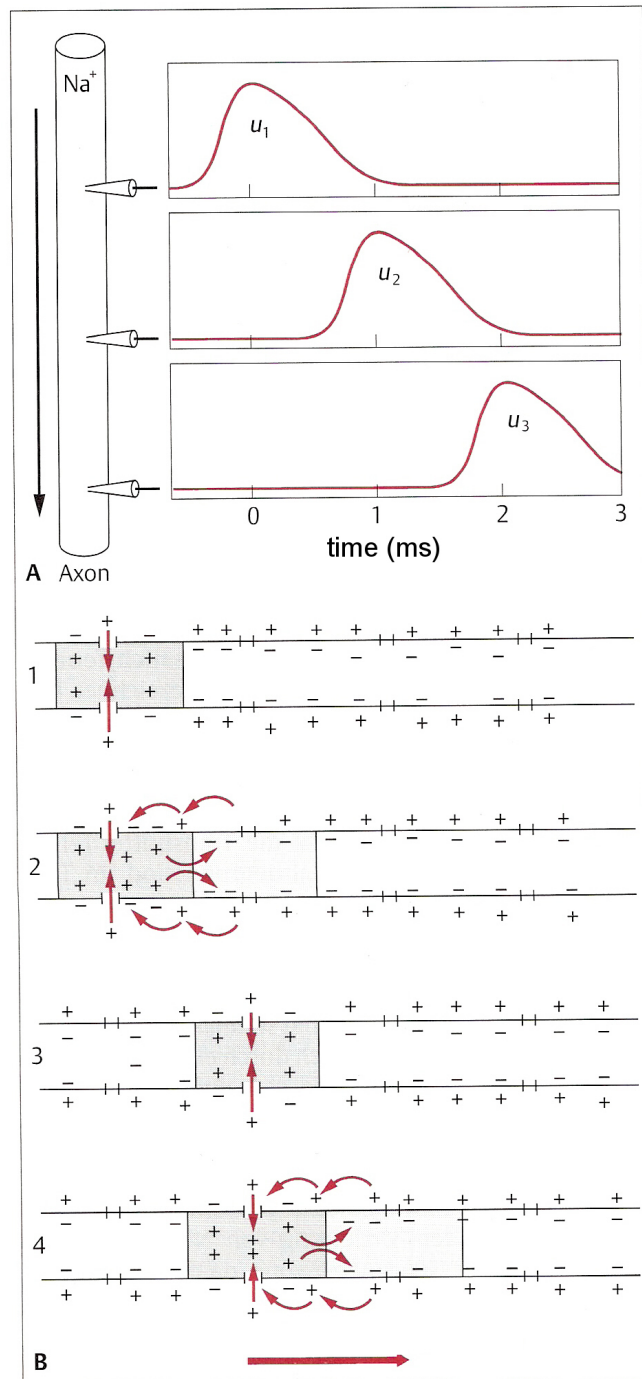


Figure 4.4: Action potential within an axon ([R.00], p. 168, modified)

#### 4 Artificial Neural Networks

through frequent use a lighter forwarding of information takes place.

At the end of the axons there exists a so-called synaptic end, which covers numerous vesicles with chemical carrier substances. If an action potential comes to this point, then the amount of  $Ca^{2+}$  increases in this area. This leads to the fact, that the vesicles do melt with the presynaptic membrane and deliver the carrier substances in the synaptic gap. These the reach the postsynaptic membrane, where receptors are present. These are connection places to which fitting molecules can attach themselves. If the carrier substances reach the receptors, this leads to a change of the second nerve cell's potential, and exciting and restraining synapses are distinguished. A depolarisation (excitatory postsynaptic potential, EPSP) occurs through exciting synapses by which an action potential can be released in the second neuron. By restraining synapses, a hyperpolarisation (inhibitory postsynaptic potential, IPSP) occurs, which complicates the release of an action potential (compare [G.99], p.94-95). The charging of the different synapses is shown in figure 4.5.

At the dendrites or the cell membrane of a nerve cell generally more than one synapse exists. By the cooperation of exciting and restraining synapses, it is decided whether an action potential is brought on, or not. Moreover, the size of the action potential is dependent on this process. Through these mechanisms it is possible that a complicated forwarding and processing of information can occur through the boarding of many nerve cells. An example of this is the retina of the eye which is shown in figure 4.6. At this point, information is taken up perpetually over sensors and is processed over a nerve cell network.

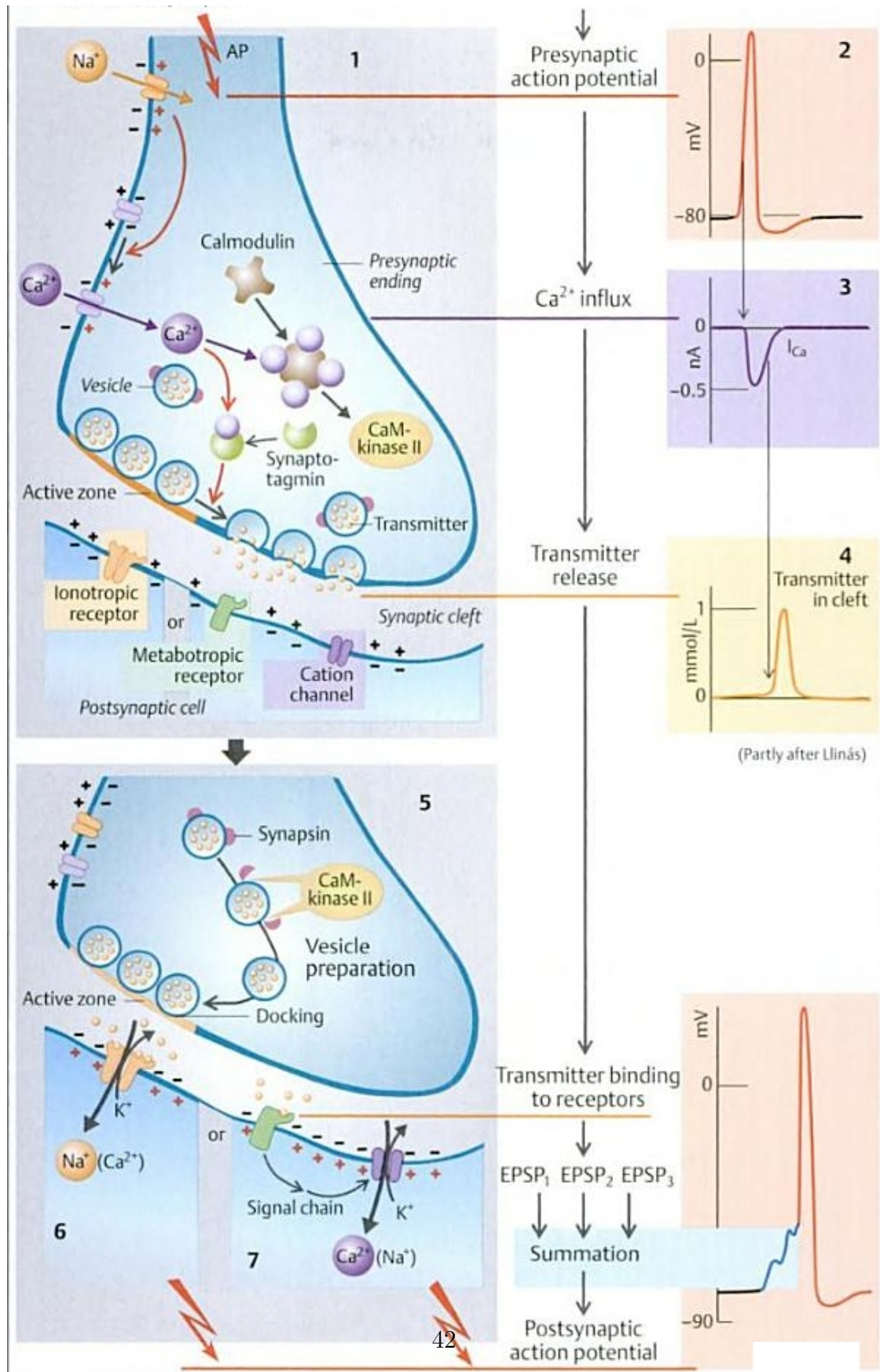


Figure 4.5: Chemical synapse ([Sil09], p. 51)

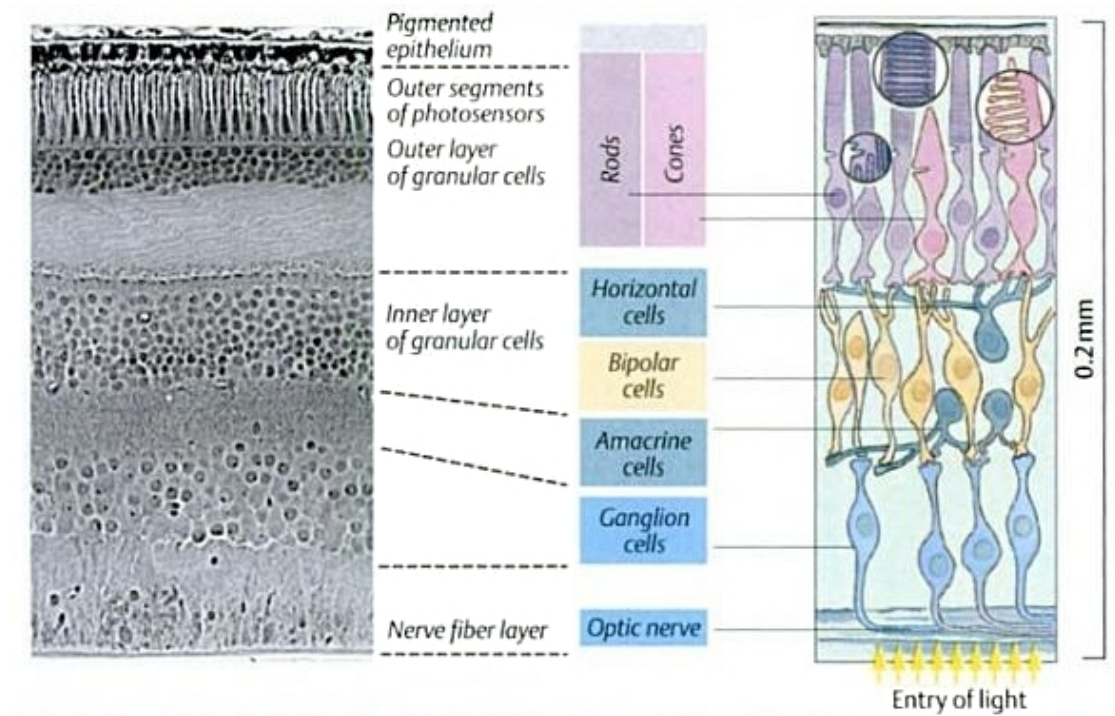


Figure 4.6: Nerve cell network of the human eye retina ([Sil09], p. 351)



#### 4 Artificial Neural Networks

A possibility of modeling ANNs with respect to the biological coherences as used by many authors is now described based on [Zel96].

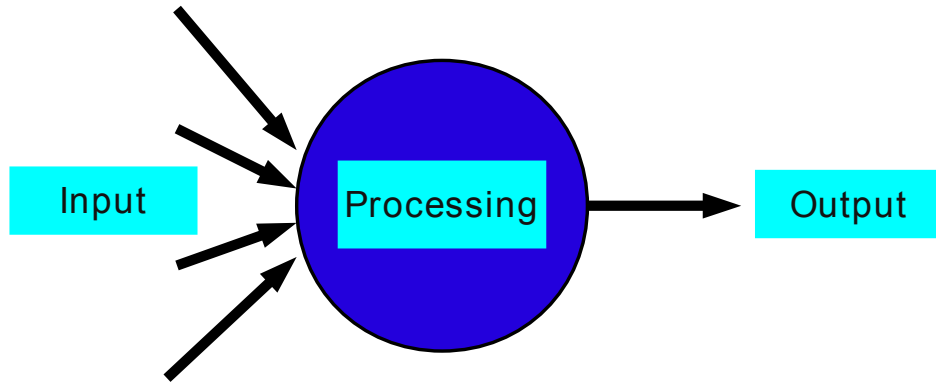
As with the biological model, the atomic units of ANNs should also be called **nerve cells** or **neurons**. Some authors use the terms **cells** or **processing element**.

These nerve cells have to take up information, process it and then pass it on. Because of this fact, a necessity for input elements which have the same function as the dendrites exists. These elements connect the nerve cell with other cells and take up information. It is important that these connections are directed and that the nerve cell can only take up information from the entrances . These entrances should be illustrated by lines.

The nerve cell calculates from the entrances value which is passed on by a directed connection to other nerve cells, it can be connected with a finite number of other nerve cells. The biological nerve cells are connected with each other by the synapses. With the forwarding of information single nerve cells can work by restraining or exciting others. This process which is steered in the biological model through the amount of chemical transmitters and the size of the respective synapse is realised in the mathematical model with weights assigned to the suitable connections.

Figure 4.7 illustrates the coherence of *input*, *processing* and *output*. This processing is looked at in more detail in the following. According to [Zel96], p.72 et seqq., three different mappings classically used, a *propagation-mapping*, an *activation-mapping* and an *output-mapping*. Often one of the last two mappings is the identic mapping, but for some applications the flexibility of having two different mappings is necessary. Regarding an nerve cell  $N_j$  with  $I$  inputs,  $I \in \mathbb{N}$ , the following definitions can be given.





**Figure 4.7:** Input, processing and output in general

$$\begin{aligned} net : \mathbb{R}^I \times \mathbb{R}^I &\longrightarrow \mathbb{R} \\ (O_j, W_j) &\mapsto net(O_j, W_j) \end{aligned}$$

with

$$net(O_j, W_j)(s) = \sum_{\tilde{i}=1}^I o_{\tilde{i}}(s) \cdot w_{\tilde{i}j}(s) \quad (4.1)$$

Thereby  $O_j$  represents a vector of input information with  $I$  components also-called the *input-vector*.  $W_j$  contains the particular weights of the connections and is called the *weight-vector*. The  $s \in \mathbb{N}_0$  represents the actual step of iteration, which is necessary to enable the distinguishment of the different steps. In the following this should be omitted to gain a clear arrangement. For a shorter notation of the value of the propagation-mapping, the following is used in the context of neuron  $N_j$ :

$$net_j(s) = \sum_{\tilde{i}=1}^I o_{\tilde{i}}(s) \cdot w_{\tilde{i}j}(s) \quad (4.2)$$

## 4 Artificial Neural Networks

The following definition of the activation-mapping is given:

$$\begin{aligned} f_{act} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto f_{act}(x) \end{aligned}$$

with

$$f_{act}(net_j(s)) = a_j(s) \tag{4.3}$$

Thereby, the codomain can often be restricted to a finite interval. The output-mapping is defined in a similar way:

$$\begin{aligned} f_{out} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto f_{out}(x) \end{aligned}$$

with

$$f_{out}(a_j(s)) = o_j(s) \tag{4.4}$$

According to [Kra91], pp.24 - 25, the generated state of activation can be classified in the following way:

1. Discrete state of activation
  - binary:  $\{0, 1\}$ ,  $\{-1, 1\}$
  - non-binary:  $\{-1, 0, 1\}$ ,  $\{-100, \dots, 100\}$
2. Continuous state of activation
  - infinite:  $\mathbb{R}$
  - finite intervals:  $[0, 1]$ ,  $[-1, 1]$

## 4 Artificial Neural Networks

The following four mappings in particular should be mentioned at this point:

$$\begin{aligned} f_{act_1} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\mapsto f_{act_1}(x) := \begin{cases} 0 & \text{for } x < \theta_j \\ 1 & \text{for } x \geq \theta_j \end{cases} \end{aligned}$$

$$\begin{aligned} f_{act_2} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\mapsto f_{act_2}(x) := \begin{cases} -1 & \text{for } x < -0,5 \cdot \pi \\ \sin(x) & \text{for } -0,5 \cdot \pi \leq x < 0,5 \cdot \pi \\ 1 & \text{for } x \geq 0,5 \cdot \pi \end{cases} \end{aligned}$$

$$\begin{aligned} f_{act_3} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\mapsto f_{act_3}(x) := \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned}$$

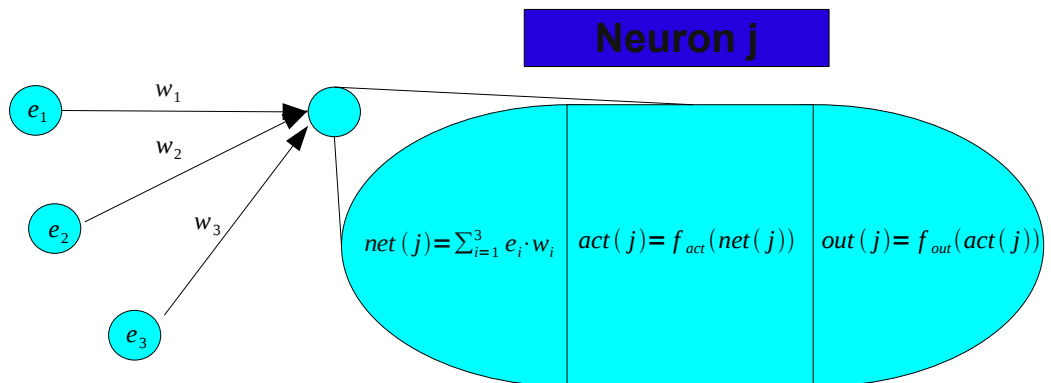
$$\begin{aligned} f_{act_4} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\mapsto f_{act_4}(x) := \frac{1}{1 + e^{-x}} \end{aligned}$$

The coherences of the different mappings within an artificial nerve cell are illustrated in figure 4.8.

Alltogether, the processing of information of a nerve cell consists of the network input, the state of activation and the output. Such a nerve cell is only one small elementary component of a neural network. By the description of ANNs, different topologies can be used according to the use. Here the term topology means the connectivity structure of the nerve cells.

ANNs can be described with respect to the connectivity structure and the relationships between the single neurons using terms and coherences from *graph theory*.

At this point, some basic coherences of graph theory are described and then these methods are applied to the Artificial Neural Networks.



**Figure 4.8:** Coherence of the propagation-, activation- and output-mapping

Based on [Bal00], pp. 2-3, a *graph* can be defined and characterised in the following way:

**Definition 4.2.1 (Graph)**

A **graph** is an ordered triple  $G = (V(G), E(G), I_G)$ , where  $V(G)$  is a non-empty set,  $E(G)$  is a set disjoint from  $V(G)$ , and  $I_G$  is an **incidence map** that associates with each element of  $E(G)$ , an unordered pair of elements (same or distinct) of  $V(G)$ . Elements of  $V(G)$  are called the **vertices** (or **nodes** or **points**) of  $G$ , and elements of  $E(G)$  are called the **edges** (or **lines**) of  $G$ . If, for the edge  $e$  of  $G$ ,  $I_G(e) = \{u, v\}$  is valid, the notation  $I_G(e) = uv$  is used.

**Definition 4.2.2 (Characterisation of a graph)**

In this definition, some characteristic terms describing a graph are listed.

- (1) **End or End vertex:** If  $I_G(e) = uv$ , then the vertices  $u$  and  $v$  are called **end vertices** of the edge  $e$ .
- (2) **Incidence:** If  $u$  and  $v$  are the **end vertices** of the edge  $e$ , they are denoted as **incident** with  $e$ .
- (3) **Parallelism:** If a set of two or more edges have the same ends, these edges are called **multiple** or **parallel** edges.
- (4) **Adjacency:**
  - A pair of vertices  $u$  and  $v$  is called **adjacent** in  $G$ , if, and only if, there is an edge in  $G$  with  $u$  and  $v$  as its ends.
  - A pair of distinct edges  $e$  and  $f$  is called **adjacent** in  $G$ , if, and only if, there exists a common end vertex.
- (5) **Loop:** An edge  $e$  is called **loop**, if  $I_G(e) = uu$ .
- (6) **Simple graph:** A graph is called **simple**, if there are no loops and multiple

*edges.*

According to [Bal00], p. 3, a graph can be represented by a *digram*, where, the vertices are drawn as *points* and the edges as *connections* joining the vertices. This representation allows a visualised overview of the described issue using a graph.

**Definition 4.2.3 (Finiteness, order and size of a graph)**

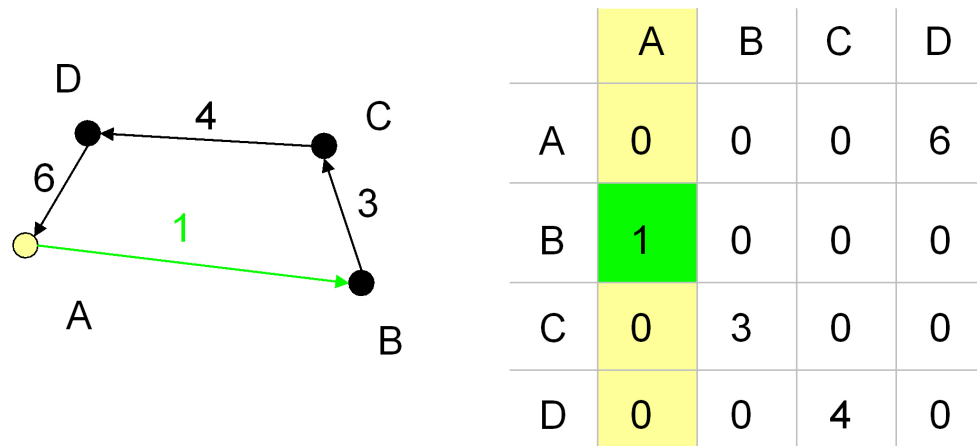
*If both,  $V(G)$  and  $E(G)$  are finite, a graph is called **finite graph**. A graph, that is not finite is called an **infinite graph**.*

*The number  $o(G)$  of vertices of a graph is called the **order** of this graph, the number  $s(G)$  of edges is called the **size** of this graph.*

A special form of graphs are the so-called digraphs (see [Nie05], p.24). where for every edge a direction is set. In addition, the possibility to assign a number, the so-called weight to every edge exists. Such graphs are then called directed weighted graphs. This kind of graphs play an important role for describing BPNs.

To maintain the overview about the connecting structure with regard to even more complicated graphs, the existing connections can be represented in a scheme called adjacency matrix (compare [Nie05], p.7). If no weights exist, a 1 is put down for every available connection, otherwise a 0. Here, every column represents a vertex and the connection to other vertices is represented by the particular rows. If the edges are weighted, the suitable weights are put down, instead. With directed graphs the connection is put down only for the vertex from which the edge starts. A graph and the suitable adjacency matrix are shown in figure 4.9.

As already described neural networks consisting of single neurons are characterised by a network input, an activation function and an output function. The whole data processing of a neural network can be adapted by different arrangements and different kinds of connections. Further, the nerve cells are summarised within most



**Figure 4.9:** Directed weighted graph and according adjacency matrix

models to so-called layers. The first of the layers is often called the **input layer** and the last one the **output layer**. All other layers which are localised between them are the so-called **hidden layers**. Common topologies are illustrated in figure 4.10.

Adaptivity of an Artificial Neural Network to a given situation can be realised by learning procedures to a certain grade. In this respect, three main models have been developed, the supervised learning, the encouraging learning and the unsupervised learning (compare [Pat00], p.42). Besides, the nerve cells of the input layer receive an input. The totality of these inputs which appear together is called the input pattern.

**Supervised learning:** With this form of learning a suitable expected output pattern exists for each of the used input patterns. One would like to reach the point where the neural network adapts itself during the learning process in such a way that one possibly receives the suitable output pattern or approximately this pattern. Here, a teacher who gives the input patterns and the expected output patterns to the neural network is necessary. Then from

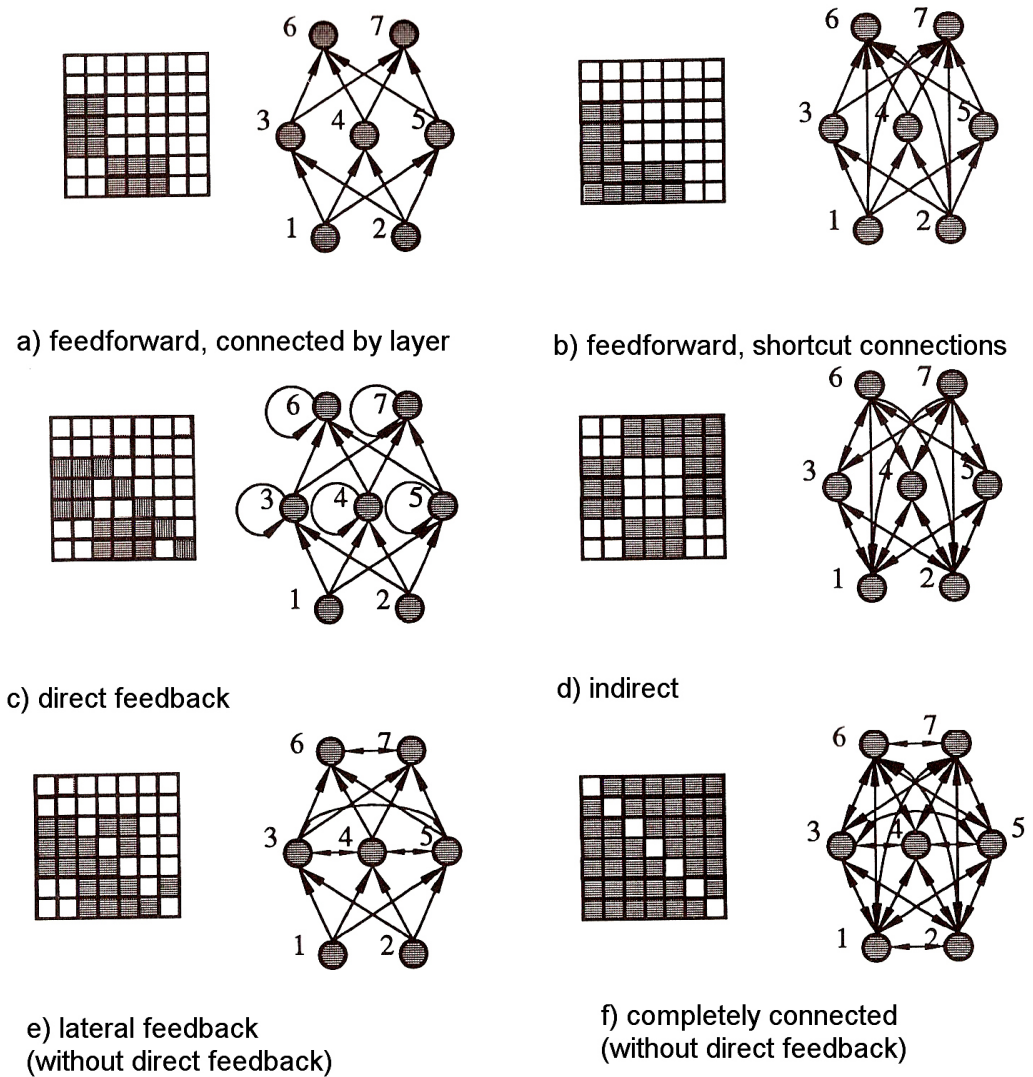


Figure 4.10: Network topologies ([Zel96], p. 79, modified)



the input patterns the generated outputs are compared to the expected outputs. This comparison can be realised by an error-function which calculates the difference between the expected ones and the preserved outputs. Such functions will be looked at further.

If the difference has been calculated, this is the used with the regarded learning procedure to adapt the components of the neural network.

A realisation of this learning-type is the so-called Backpropagation rule which is explained in more detail in the following section.

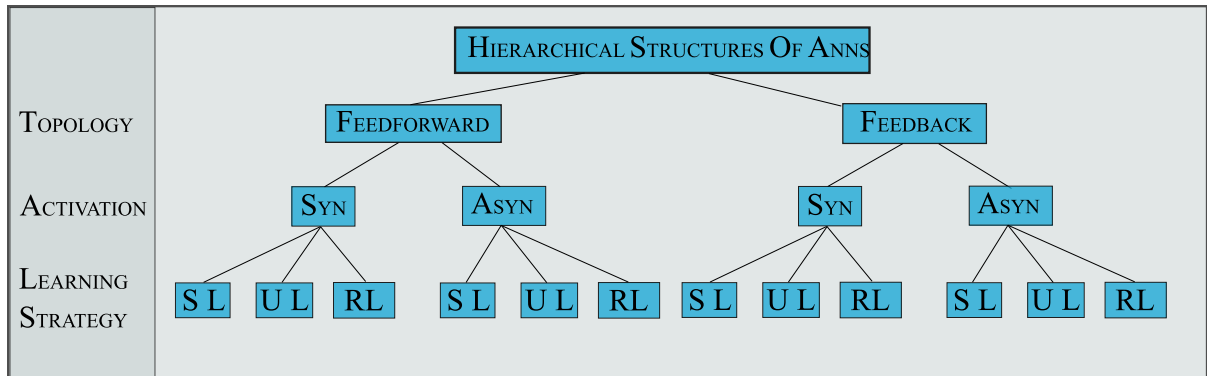
**Encouraging learning:** The encouraging learning is characterised by the fact that the neural network does not know the expected issue pattern. The teacher gives only instructions, while the weights of the components which have calculated a good issue become bigger, the others smaller.

**Unsupervised learning:** With this form of learning, the neural network does not know the expected issue pattern and there is also no teacher. The learning process now consists because of the fact that the neural network adapts its components to the input patterns and their structure. A learning procedure from this group is the so-called Kohonen algorithm (see further below). Further, with every step of the iteration the weights of the nerve cell which has the biggest state of activation in terms of the actual input pattern are adapted following a certain rule.

### 4.3 Description of the relevant models for this thesis

In literature, a huge amount of ANN-models are described and there are many possibilities for classification. At this point, a possible hierarchical structure concerning the basing topology and processing algorithm is discussed. A concise

overview at this structure is provided in figure 4.11. This structure is only a proposal and theoretically, the layers shown in figure 4.11 can be arranged in another way.



**Figure 4.11:** Hierarchical structures of ANNs. Abbreviations: **syn**: synchronous activation, **asyn**: asynchronous activation, **sl**: supervised learning, **ul**: unsupervised learning, **rl**: reinforced learning

Based on [Zel96], pp.71-96, there are pointed out three main hierarchical layers concerning ANNs are pointed out, the **network topology**, the **way of activation** and the used **learning strategy**.

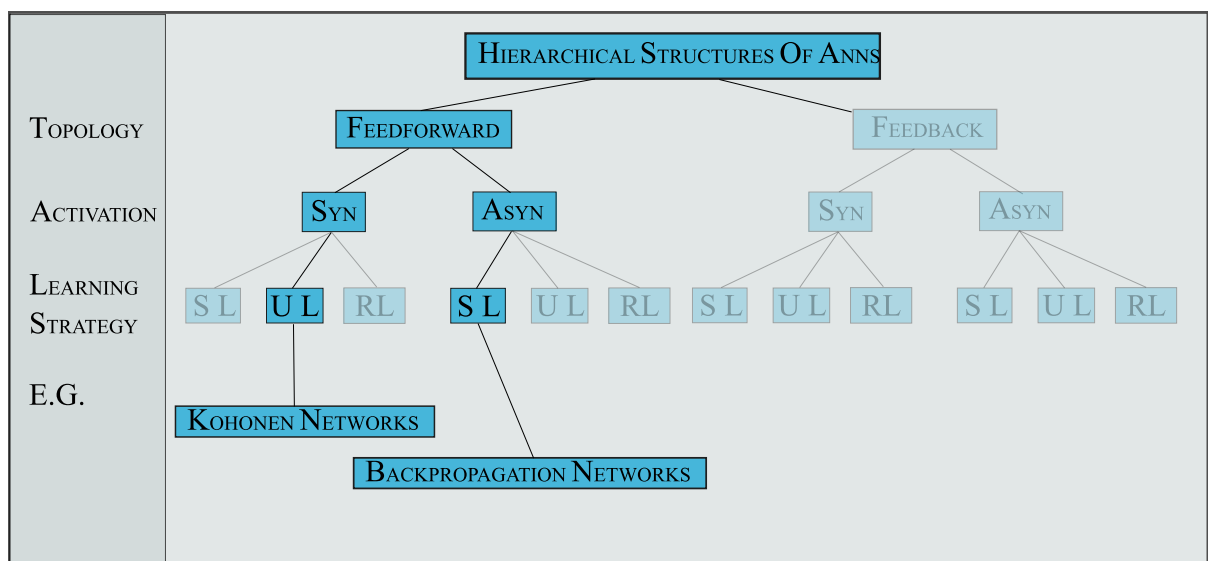
The network topology can be classified into **feed-forward-networks** and **feedback-networks**.

There are two main application fields of ANNs, **approximation** and **classification**. In this section the ANN-models that are directly associated with this thesis are looked at. The so-called Backpropagation Networks are representatives of approximating networks. These networks are regarded and extended in the context of mathematical measure-theory in chapter 9 to describe their characteristics in

the face of usage and applicability in the SPATTB.

The second model are the so-called Kohonen Networks (KNs), which can be used for classification. For this model an explicit application that has been programmed as a prototype by the author in the context of the EWARS is described and further addressed in chapter 9.

The classification of these two network types with respect to the hierarchical structure described in section 4.2 is shown in figure 4.12.



**Figure 4.12:** Classification of the BPNs and the KNs in the regarded hierarchical structure

### 4.3.1 Backpropagation Networks

Backpropagation Network are multi-layer feed-forward-networks connected by layer and trained with the Backpropagation-algorithm. This algorithm should be described in the following. The aim of the training is to get a network representing pairs of input-vectors and desired output-vectors using the according de facto

## 4 Artificial Neural Networks

output-vectors. Thereby, a decreasing of the generated error is aspired.

The idea of the Backpropagation Network-algorithm has been published in the dissertation of Paul Werbos in 1974 ([Wer74]) the first time, but it only attracted interest with the publication of Rumelhart et al. in the journal *Nature* ([RHW86]). For describing the processing within this kind of networks, the number of so-called *trainable layers* is called  $n \in \mathbb{N}$ . Thereby, the regarded network has  $n + 1$  layers, whose connections can be represented by  $n$  adjacency matrices, if the network is regarded as directed weighed graph. Such a network is illustrated in figure 4.13. The usage of adjacency matrices leads to a more clearer representation. Additionally, the calculation of the output of the whole network can be managed more easier. In figure 4.14, the first adjacency matrix of a Multi-layer feed-forward-network (MLFFN) is illustrated. There, the number of the layer is realised by superscription. The according adjacency matrices should then be denoted as *weight matrices*.

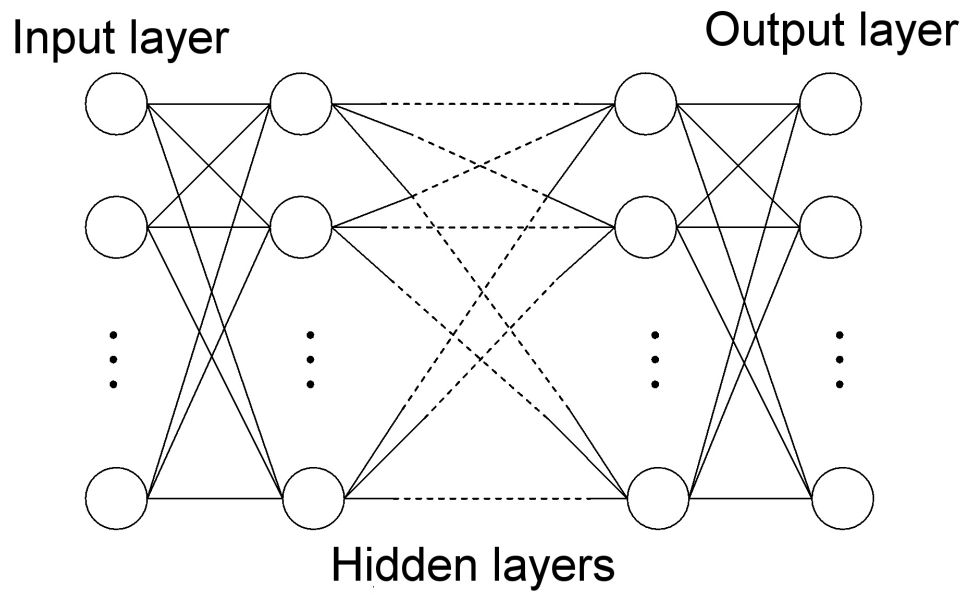
In general, weight matrices should be denoted in this thesis in the following way:

$$W_{\tilde{i},\tilde{i}+1}(s)$$

Thereby,  $\tilde{i}$  is the position of the first layer,  $\tilde{i} + 1$  th one of the second layer counting in terms of figure 4.13.

In the example shown in figure 4.14,  $r$  represents the number of neurons in the first layer,  $s$  the one in the second layer. The following  $s \times r$ -matrix is generated

$$W_{1,2} = \begin{pmatrix} w_{N_1^1, N_1^2} & w_{N_2^1, N_1^2} & \dots & w_{N_r^1, N_1^2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N_1^1, N_s^2} & w_{N_2^1, N_s^2} & \dots & w_{N_r^1, N_s^2} \end{pmatrix} \quad (4.5)$$



**Figure 4.13:** Multi-layer *feed-forward-network*

4 Artificial Neural Networks

	$N_1^1$	$N_2^1$	$\dots$	$N_r^1$	$N_1^2$	$N_2^2$	$\dots$	$N_s^2$
$N_1^1$								
$N_2^1$								
$\vdots$								
$N_r^1$								
$N_1^2$	$w_{N_1^1, N_1^2}$	$w_{N_2^1, N_1^2}$	$\dots$	$w_{N_r^1, N_1^2}$				
$N_2^2$	$w_{N_1^1, N_2^2}$	$w_{N_2^1, N_2^2}$	$\dots$	$w_{N_r^1, N_2^2}$				
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$				
$N_s^2$	$w_{N_1^1, N_s^2}$	$w_{N_2^1, N_s^2}$	$\dots$	$w_{N_r^1, N_s^2}$				

**Figure 4.14:** Adjacency scheme for generating the first weight matrix

#### 4 Artificial Neural Networks

Regarding the input-vector  $o_{L1}$ , the following value of propagation can be calculated:

$$\begin{aligned}
 W_{1,2}o_{L1} &= \begin{pmatrix} w_{N_1^1, N_1^2} & w_{N_2^1, N_1^2} & \dots & w_{N_r^1, N_1^2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N_1^1, N_s^2} & w_{N_2^1, N_s^2} & \dots & w_{N_r^1, N_s^2} \end{pmatrix} \begin{pmatrix} o_{N_1^1} \\ \vdots \\ o_{N_r^1} \end{pmatrix} \\
 &= \begin{pmatrix} o_{N_1^1}w_{N_1^1, N_1^2} + o_{N_2^1}w_{N_2^1, N_1^2} + \dots + o_{N_r^1}w_{N_r^1, N_1^2} \\ \vdots \\ o_{N_1^1}w_{N_1^1, N_s^2} + o_{N_2^1}w_{N_2^1, N_s^2} + \dots + o_{N_r^1}w_{N_r^1, N_s^2} \end{pmatrix} \\
 &= \begin{pmatrix} \sum_{i=1}^r o_{N_i^1}w_{N_i^1, N_1^2} \\ \vdots \\ \sum_{i=1}^r o_{N_i^1}w_{N_i^1, N_s^2} \end{pmatrix} = \\
 &\begin{pmatrix} net_{N_1^2} \\ \vdots \\ net_{N_s^2} \end{pmatrix}
 \end{aligned}$$

For calculating the output of the  $n + 1$  different layers, the following mapping  $F_{ges}^{m_{\tilde{i}+1}}$ ,  $\tilde{i} \in \{1, \dots, n\}$  is defined:

$$\begin{aligned}
 F_{act}^{m_{\tilde{i}+1}} \quad \mathbb{R}^{m_{\tilde{i}+1}} &\longrightarrow \mathbb{R}^{m_{\tilde{i}+1}} \\
 \begin{pmatrix} x_1 \\ \vdots \\ x_{m_{\tilde{i}+1}} \end{pmatrix} &\longrightarrow F_{act}^{\tilde{j}} \left( \begin{pmatrix} x_1 \\ \vdots \\ x_{m_{\tilde{i}+1}} \end{pmatrix} \right) = \begin{pmatrix} f_{act}(x_1) \\ \vdots \\ f_{act}(x_{m_{\tilde{i}+1}}) \end{pmatrix}
 \end{aligned}$$

and

#### 4 Artificial Neural Networks

$$F_{out}^{m_{\tilde{i}+1}} : \mathbb{R}^{m_{\tilde{i}+1}} \longrightarrow \mathbb{R}^{m_{\tilde{i}+1}}$$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_{m_{\tilde{i}+1}} \end{pmatrix} \longrightarrow F_{out}^{\tilde{j}} \left( \begin{pmatrix} x_1 \\ \vdots \\ x_{m_{\tilde{i}+1}} \end{pmatrix} \right) = \begin{pmatrix} f_{out}(x_1) \\ \vdots \\ f_{out}(x_{m_{\tilde{i}+1}}) \end{pmatrix}$$

Here,  $m_{\tilde{i}+1}$  represents the number of neurons in layer  $\tilde{i} + 1$ .

Finally,

$$F_{ges}^{m_{\tilde{i}+1}} := F_{out}^{m_{\tilde{i}+1}} \circ F_{act}^{m_{\tilde{i}+1}}$$

is defined with:

$$F_{ges}^{m_{\tilde{i}+1}} : \mathbb{R}^{m_{\tilde{i}+1}} \longrightarrow \mathbb{R}^{m_{\tilde{i}+1}}$$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_{m_{\tilde{i}+1}} \end{pmatrix} \longrightarrow F_{ges}^{\tilde{j}} \left( \begin{pmatrix} x_1 \\ \vdots \\ x_{m_{\tilde{i}+1}} \end{pmatrix} \right) = F_{out}^{\tilde{j}} \left( F_{act}^{m_{\tilde{i}+1}} \begin{pmatrix} x_1 \\ \vdots \\ x_{m_{\tilde{i}+1}} \end{pmatrix} \right)$$

After describing the calculation of the output of a MLFFN, there should be given a description of calculating output errors with regard to the BPN-algorithm.

Because these networks should be used for the approximation of more or less complicated unknown functions, the coherence between actual output and desired output has to be regarded. Besides, the input lies in the domain of the unknown mapping and the output in the codomain. Such a network exactly then has an optimum compound of the weights, when it calculates the desired output fitting to the according input. Nevertheless, this will not be generally the case. Even more it is necessary to use procedures which approximate this state as good as possible. Thereby, the network is enabled to approximately learn the function.



#### 4 Artificial Neural Networks

The possibilities which components can be changed with such a learning process were already described. Besides, with the learning procedure basing on the BPN-algorithm, the topology of the network will maintain. The learning consists in the fact that the weights of the connections change with every learning step and the neural network adapts in this way.

So that one can apply such a procedure, it is necessary to know whether and how the preserved outputs deviate from the expected ones. Then this difference between preserved and expected output is the mistake which the neural network makes with the image of the suitable input. How this mistake can be calculated, is further described in this segment.

If one liked to represent a functional coherence with a neural network, one knows only an finite amount of inputs and their images, the expected outputs. One can summarise this into a set of tuples with two components in each case. The first component is the input and the second one the suitable output. This set is often divided into a so-called training set and a testing-set. Then, the training is done with the training-set and after that, the adapted network is tested with the testing-set.

The following assumptions are made:

$\tilde{P}$  is the number of trainings-patterns.

$t^P$  is the desired output-vector of pattern  $P$ .

$o^P$  is the actual output-vector according to the input-vector  $o_1^P$ ,  $P \in \{1, \dots, \tilde{P}\}$ .

The training-set  $T$  is then the following:

$$T := \left\{ (o_1^1, t^1), \dots, (o_1^{\tilde{P}}, t^{\tilde{P}}) \right\} .$$

With this, the following error-mapping  $E$  is defined:

#### 4 Artificial Neural Networks

$$\begin{aligned}
 E : \quad \mathbb{R}^{m_1} \times \mathbb{R}^{m_{n+1}} &\longrightarrow \mathbb{R} \\
 (t^P, o^P) &\mapsto E(t^P, o^P) \\
 := \frac{1}{2} \sum_{\tilde{j}=1}^{m_{n+1}} (t_{\tilde{j}}^P - o_{\tilde{j}}^P)^2
 \end{aligned}$$

Abbreviative,  $E_P$  is used instead of  $E(t^P, o^P)$ .

If one liked to receive, nevertheless, an information about the whole amount in training patterns, it is necessary to summarise the respective errors with another mapping. An easy possibility consists in forming the sum from the calculated values.

$$\begin{aligned}
 E_g &:= \sum_{P=1}^{\tilde{P}} E_P \\
 &= \sum_{P=1}^{\tilde{P}} \frac{1}{2} \sum_{\tilde{j}=1}^{m_{n+1}} (t_{\tilde{j}}^P - o_{\tilde{j}}^P)^2
 \end{aligned} \tag{4.6}$$

The Backpropagation Network-algorithm is used to minimise the error in the context of *feed forward networks* with an set of training patterns. Besides, a gradient-descent procedure is applied.

Thereby, the function  $E$  is regarded as a function of the weights of the respective neural network. At last, these weights are the parameters in a neural network which form the basis for the calculation of the respective outputs.

To minimise the originating error between expected output and preserved output, the weights must be changed accordingly. Besides, one receives from the number of these weights the dimension of the domain. The dimension of the codomain is

#### 4 Artificial Neural Networks

one. The whole number of the weights should be denoted as  $n_{whole}$ . With this, the mapping  $E$  can be described in the following way:

$$E : \quad \mathbb{R}^{n_{whole}} \longrightarrow \mathbb{R}$$

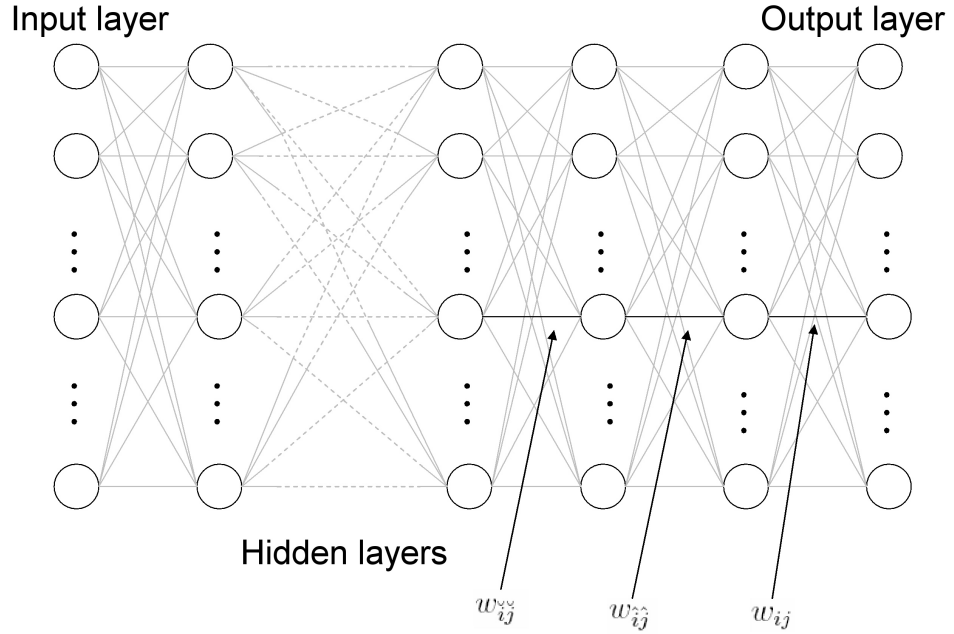
$$\begin{pmatrix} w_1 \\ \vdots \\ w_{n_{whole}} \end{pmatrix} \longrightarrow E \left( \begin{pmatrix} w_1 \\ \vdots \\ w_{n_{whole}} \end{pmatrix} \right)$$

Besides, the weight  $w_1$  describes the first weight of the neural network. The vector which summarizes all weights is called weight vector. The remaining components of the weight vector consists of the remaining weights which are arranged in order by the input layer up to the output layer.

With the calculation of  $E_P$ , the respective weights play a big role, because with other weights one becomes generally also another error for the tuple  $(t^P, o^P)$ .

Regarding MLFFN, differentiability of the mapping  $E$  is necessary because of the applied gradient-descent processing. This differentiability is given using differentiable component mappings. According to [Zel96]), therefore often the activating mapping is the identity and the output-mapping is a *sigmoid* mapping, exemplarily the logistic mapping defined in 4.2.

The error should be minimised by the fact that the gradient from the partial derivatives is calculated with respect to the weights of the neural network. With two-layer networks there is only one layer of trainable weights which are used for processing the network output directly. If more than one layer of weights exists, one receives for every hidden layer an output. Then from the output of the last of these layers and the last weight matrix, the output of the whole neural network is calculated. Hence, the partial derivatives can be calculated only for the weights in the last weight matrix directly. The other partial derivatives have to be calculated



**Figure 4.15:** Situation within a BPN

using the already calculated partial derivatives of the following layer.

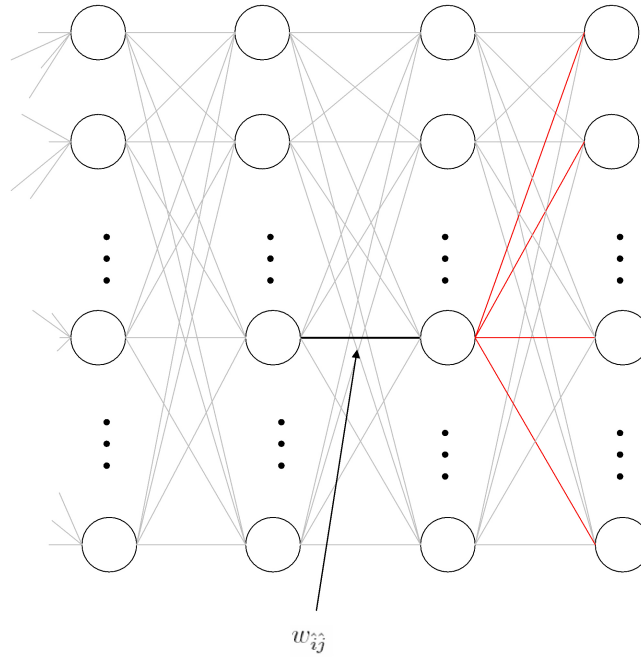
So, the calculation can be described in the following way:

$$\begin{aligned}
 \frac{\partial E_P}{\partial w_{ij}} &= \frac{\partial \frac{1}{2} \sum_{j=1}^{m_{n+1}} (t_j^P - o_j^P)^2}{\partial w_{ij}} \\
 &= \frac{\partial \frac{1}{2} \left( (t_1^P - o_1^P)^2 + \dots + (t_j^P - o_j^P)^2 + \dots + (t_{m_l}^P - o_{m_l}^P)^2 \right)}{\partial w_{ij}} \\
 &= \frac{\partial \frac{1}{2} (t_1^P - o_1^P)^2 + \dots + \frac{1}{2} (t_j^P - o_j^P)^2 + \dots + \frac{1}{2} (t_{m_l}^P - o_{m_l}^P)^2}{\partial w_{ij}} \\
 &= \left( \frac{\partial \frac{1}{2} (t_1^P - o_1^P)^2}{\partial w_{ij}} + \dots + \frac{\partial \frac{1}{2} (t_j^P - o_j^P)^2}{\partial w_{ij}} + \dots + \frac{\partial \frac{1}{2} (t_{m_l}^P - o_{m_l}^P)^2}{\partial w_{ij}} \right)
 \end{aligned}$$

#### 4 Artificial Neural Networks

$$\begin{aligned}
&= \left( 0 + \dots + \frac{\partial \frac{1}{2} (t_j^P - o_j^P)^2}{\partial w_{ij}} + \dots + 0 \right) \\
&= \frac{\partial \frac{1}{2} (t_j^P - o_j^P)^2}{\partial w_{ij}} \\
&\stackrel{(*)}{=} \frac{\partial \frac{1}{2} (y_j^P)^2}{\partial y_j^P} \frac{\partial y_j^P}{\partial w_{ij}} \\
&= \frac{\partial \frac{1}{2} (y_j^P)^2}{\partial y_j^P} \frac{\partial (t_j^P - o_j^P)}{\partial o_j^P} \frac{\partial o_j^P}{\partial w_{ij}} \\
&= y_j^P \left( \frac{\partial t_j^P}{\partial o_j^P} + \left( -\frac{\partial o_j^P}{\partial o_j^P} \right) \right) \frac{\partial o_j^P}{\partial w_{ij}} \\
&= -y_j^P \frac{\partial f_{out}(net_j^P)}{\partial net_j^P} \frac{\partial \left( \sum_{i=1}^{m_n} o_i^P w_{ij} \right)}{\partial w_{ij}} \\
&= -y_j^P f'_{out}(net_j^P) \frac{\partial (o_1^P w_{1j} + \dots + o_i^P w_{ij} + \dots + o_{m_n}^P w_{m_n j})}{\partial w_{ij}} \\
&= -y_j^P f'_{out}(net_j^P) \left( \frac{\partial o_1^P w_{1j}}{\partial w_{ij}} + \dots + \frac{\partial o_i^P w_{ij}}{\partial w_{ij}} + \dots + \frac{\partial o_{m_n}^P w_{m_n j}}{\partial w_{ij}} \right) \\
&= -y_j^P f'_{out}(net_j^P) \left( 0 + \dots + \frac{\partial o_i^P w_{ij}}{\partial w_{ij}} + \dots + 0 \right) \\
&= -y_j^P f'_{out}(net_j^P) \frac{\partial o_i^P w_{ij}}{\partial w_{ij}} \\
&= -y_j^P f'_{out}(net_j^P) o_i^P
\end{aligned}$$

#### 4 Artificial Neural Networks



**Figure 4.16:** Weight within the last but one trainable layer

$$\stackrel{(**)}{=} -\delta_{j_{oB}}^P o_i^P \quad (4.7)$$

$$(*) := y_j^P := t_j^P - o_j^P, j \in \{1, \dots, n+1\}$$

$$(**) := \delta_{j_{oB}}^P := (t_j^P - o_j^P) f'_{out}(net_j^P)$$

Using the variables  $\delta_{j_{oB}}^P$  the  $o$  marks, that the regarded weight  $w_{ij}$  belongs to the last weight matrix.  $B$  means that the Backpropagation rule is applied.

If one looks, nevertheless, at a weight  $w_{ij}$  with which the nerve cell  $i$  and the nerve cell  $j$  are in two hidden layers following on each other, then the partial derivatives cannot be calculated directly.

If one considers  $j$  as a nerve cell of the last hidden layer, this is connected with all nerve cells of the output layer. By a change of the weight  $w_{ij}$  the outputs of the nerve cells of the output layer change accordingly. The calculation of the particular partial derivatives can be done in the following way:

#### 4 Artificial Neural Networks

$$\begin{aligned}
\frac{\partial E_P}{\partial w_{ij}^{\hat{}}} &= \frac{\partial \frac{1}{2} \left( (t_1^P - o_1^P)^2 + \dots + (t_j^P - o_j^P)^2 + \dots + (t_{m_l}^P - o_{m_l}^P)^2 \right)}{\partial w_{ij}^{\hat{}}} \\
&= \frac{\partial \frac{1}{2} (t_1^P - o_1^P)^2 + \dots + \frac{1}{2} (t_j^P - o_j^P)^2 + \dots + \frac{1}{2} (t_{m_l}^P - o_{m_l}^P)^2}{\partial w_{ij}^{\hat{}}} \\
&= \left( \frac{\partial \frac{1}{2} (t_1^P - o_1^P)^2}{\partial w_{ij}^{\hat{}}} + \dots + \frac{\partial \frac{1}{2} (t_j^P - o_j^P)^2}{\partial w_{ij}^{\hat{}}} + \dots + \frac{\partial \frac{1}{2} (t_{m_l}^P - o_{m_l}^P)^2}{\partial w_{ij}^{\hat{}}} \right) \\
&\stackrel{(***)}{=} \left( \frac{\partial f_1(y_1^P)}{\partial w_{ij}^{\hat{}}} + \dots + \frac{\partial f_1(y_j^P)}{\partial w_{ij}^{\hat{}}} + \dots + \frac{\partial f_1(y_{m_l}^P)}{\partial w_{ij}^{\hat{}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( \frac{\partial f_1(y_k^P)}{\partial y_k^P} \frac{\partial y_k^P}{\partial w_{ij}^{\hat{}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( \frac{\partial \frac{1}{2} (y_k^P)^2}{\partial y_k^P} \frac{\partial (t_k^P - o_k^P)}{\partial w_{ij}^{\hat{}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( y_k^P \frac{\partial (t_k^P - o_k^P)}{\partial o_k^P} \frac{\partial o_k^P}{\partial w_{ij}^{\hat{}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( y_k^P (-1) \frac{\partial o_k^P}{\partial w_{ij}^{\hat{}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P \frac{\partial f_{out}(net_k^P)}{\partial w_{ij}^{\hat{}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P \frac{\partial f_{out}(net_k^P)^P}{\partial z_k} \frac{\partial net_k^P}{\partial w_{ij}^{\hat{}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P \frac{\partial f_{out}(net_k^P)^P}{\partial z_k} \frac{\partial \sum_{i=1}^{m_n} o_i^P w_{ik}}{\partial w_{ij}^{\hat{}}} \right)
\end{aligned}$$

#### 4 Artificial Neural Networks

$$\begin{aligned}
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) \frac{\partial o_1^P w_{1k} + \dots + o_j^P w_{jk} + \dots + o_{m_n}^P w_{m_n k}}{\partial w_{\hat{i}\hat{j}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) \left( \frac{\partial o_1^P w_{1k}}{\partial w_{\hat{i}\hat{j}}} + \dots + \frac{\partial o_j^P w_{jk}}{\partial w_{\hat{i}\hat{j}}} + \dots + \frac{\partial o_{m_n}^P w_{m_n k}}{\partial w_{\hat{i}\hat{j}}} \right) \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) \left( 0 + \dots + \frac{\partial o_j^P w_{jk}}{\partial w_{\hat{i}\hat{j}}} + \dots + 0 \right) \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) \frac{\partial o_j^P w_{jk}}{\partial w_{\hat{i}\hat{j}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) \frac{\partial o_j^P w_{jk}}{\partial o_j^P} \frac{\partial o_j^P}{\partial w_{\hat{i}\hat{j}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) w_{jk} \frac{\partial o_j^P}{\partial w_{\hat{i}\hat{j}}} \right) \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) w_{jk} \right) \frac{\partial o_j^P}{\partial w_{\hat{i}\hat{j}}} \\
&= \sum_{k=1}^{m_{n+1}} \left( -y_k^P f'_{out} (net_k^P) w_{jk} \right) f'_{out} (net_j^P) o_i^P \\
&= \sum_{k=1}^{m_{n+1}} \left( - (t_k^P - o_k^P) f'_{out} (net_k^P) w_{jk} \right) f'_{out} (net_j^P) o_i^P \\
&= \sum_{k=1}^{m_{n+1}} (\delta_{k_a B}^P w_{jk}) f'_{out} (net_j^P) o_i^P \\
&\stackrel{****}{=} \delta_{j_h B}^{P'} o_i^P \tag{4.8}
\end{aligned}$$

$$(***) := y_k^P := t_k^P - o_k^P, k \in \{1, \dots, n\}$$

$$(***) := \delta_{j_h B}^{P'} := \sum_{k=1}^{m_{n+1}} (\delta_{k_a B}^P w_{jk}) f'_{out} (net_j^P)$$



#### 4 Artificial Neural Networks

Within the variable  $\delta_{jhB}^{P'}$ ,  $h$  marks, that the weight  $w_{ij}$  belongs to a connection of two neurons lying in two neighbouring hidden layers.

The  $\delta$  - components of the weights between the hidden layers should be defined in general and not only for the layers  $n - 1$  and  $n$  by the following:

$$\delta_{jB}^P := \sum_k (\delta_{kB}^P w_{jk}) f'_{out}(net_j^P)$$

Using the Backpropagation Network-rule, the particular partial derivations of the error-mapping with respect to the single weights can be calculated. With them, the according gradient-vector can be constructed. In the following, some coherences of minimization methods and especially the algorithm of gradient-descent are described. The term extremum of a mapping with  $n$ -dimensional domain and one-dimensional codomain can be defined as follows (see [Neu93], p.163).

**Definition 4.3.1**

Let

$$F : U \longrightarrow \mathbb{R}$$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longrightarrow F \left( \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right)$$

an arbitrary mapping with  $U \subset \mathbb{R}^n$ . Furthermore, let  $x, x_0 \in U$  be arbitrarily chosen.

Then, the mapping has

1. a locale maximum in  $x_0$ , if the following is valid:

$$\begin{aligned} \exists_{\epsilon > 0} & : B_\epsilon(x_0) \subset U, \text{ with} \\ \forall_{x \in B_\epsilon(x_0)} & : F(x) \leq F(x_0) \end{aligned}$$

2. a locale minimum in  $x_0$ , if the following is valid:

$$\begin{aligned} \exists_{\epsilon > 0} & : B_\epsilon(x_0) \subset U, \text{ with} \\ \forall_{x \in B_\epsilon(x_0)} & : F(x) \geq F(x_0) \end{aligned}$$

If thereby  $<$  or  $>$  are valid, there exists a strict locale maximum or minimum in the regarded point.

Clearly regarded, every partial derivation is equal to zero at a maximum or minimum. This coherence is described in the following theorem.

**Theorem 4.3.1**

Let  $U \subset \mathbb{R}^n$  be an open subset and  $F : U \longrightarrow \mathbb{R}$  partial differentiable. If this mapping then has a local maximum or minimum in  $x_0 \in U$  the following is valid:

$$\frac{\partial F(x_0)}{\partial x_i} = 0 \text{ for } i \in \{1, \dots, n\}, \text{ also } \text{grad}F(x_0) = 0.$$

#### 4 Artificial Neural Networks

Thereby one receives a necessary condition for the existence of a local maximum or minimum. However, this is insufficient, because the gradient can become zero also at other places on which is entered in the other course.

At this point procedures should be explained by which it is possible to move bit by bit in the direction of the negative gradient and to reach thus to a local minimum.

The mappings used in this section should be differentiable at least one time at the regarded element of the domain.

Following [C.99], p. 1 one understands by an unrestringiated minimization problem the task to get a point  $x_M \in U$  for which is valid:

$$\forall x \in U : F(x_M) \leq F(x)$$

In general one can distinguish in numerical mathematics direct searching procedures or searching procedures which use locale information about the gradient. Besides, the second group shows a better convergence and reaches in many cases to a local minimum (compare [P.93], p. 91).

Methods from the group of the direct searching procedures are applied according to [A.00], p.316 on the not differentiable functions or on such whose differentiation is complicated. Moreover, it is possible to receive by this one start value for applying methods from the second group.

These consists in methods, where starting from a point  $x_0 \in U$  a direction of decreasing values of mapping is followed. Thereby, a sequence  $(x_k)_k$  is generated basing on the following iteration:

$$x_{k+1} = x_k + \eta_k d_k \tag{4.9}$$

$\eta_k$  describes the so-called step range,  $d_k$  the used direction. Thereby, at every step

## 4 Artificial Neural Networks

the size of the step is set starting in  $x_k$ .

With the use of this procedure with respect to the described minimization problem it is desirable, that the result converges to a value  $x_M \in U$ . Nevertheless, in many uses it is also enough if a local minimum is reached.

First the step width and the direction should be regarded closer.

So that a suitable direction can be identified, it is necessary to define this in general ([C.99], p. 25):

### Definition 4.3.2

*Let  $F : U \rightarrow \mathbb{R}$  and  $x_k \in U$ . Then the following is a necessary condition for having an adequate direction  $d_k$  starting from  $x_k$ :*

$$\exists \eta'_k \forall \eta_k \in (0, \eta'_k] : F(x_k + \eta_k d_k) < F(x_k)$$

The so-called Armijo-rule for calculating the actual step width is defined in the following (compare [C.99], p.36)

### Definition 4.3.3

*Let  $F : U \rightarrow \mathbb{R}$  be continuously differentiable. Further, let  $\sigma, \beta \in (0, 1)$  be fixed for any step of iteration. Then for  $x_k, d_k \in U$  with  $\text{grad}F(x_k) d_k < 0$  there has to be found a  $\eta_k := \max \{ \beta^l \mid l = 0, 1, 2, \dots \}$  to guarantee the validity of the following:*

$$F(x_k + \eta_k d_k) \leq F(x_k) + \sigma \eta_k \text{grad}F(x_k) d_k \quad (4.10)$$

*The Armijo-rule consists in checking this inequation successively for the  $\beta^l$  and in choosing the value with which it is valid for the first time. This value shows the maximum of all possible step widths valid for this inequation, because of  $\beta \in (0, 1)$ .*

#### 4 Artificial Neural Networks

To be able to apply the Armijo-rule, it is necessary, that it is well-defined. This is guaranteed by the following coherence:

**Theorem 4.3.2**

Let  $F : U \rightarrow \mathbb{R}$  be continuously differentiable and  $\sigma, \beta \in (0, 1)$  fixed. Then, for  $x_k, d_k \in U$  with  $\text{grad}F(x_k) d_k < 0$  there exist a finite  $l \in \mathbb{N}$  with

$$F(x_k + \beta_k^l d_k) \leq F(x_k) + \sigma \beta_k^l \text{grad}F(x_k) d_k$$

*Proof.* Suppositionally, for any  $l \in \mathbb{N}$  the following would be valid:

$$F(x_k + \beta_k^l d_k) > F(x_k) + \sigma \beta_k^l \text{grad}F(x_k) d_k$$

Then, this follows:

$$\frac{F(x_k + \beta_k^l d_k) - F(x_k)}{\beta_k^l} > \sigma \text{grad}F(x_k) d_k$$

With this, the following is valid:

$$\lim_{l \rightarrow \infty} \frac{F(x_k + \beta_k^l d_k) - F(x_k)}{\beta_k^l} > \lim_{l \rightarrow \infty} \sigma \text{grad}F(x_k) d_k$$

$$\Leftrightarrow \text{grad}F(x_k) d_k > \sigma \text{grad}F(x_k) d_k$$

#### 4 Artificial Neural Networks

This is true, because  $\beta \in (0, 1)$  and thereby  $\beta^l$  is converging to zero.

With  $\sigma \in (0, 1)$ , the following can be derived

$$\Rightarrow \text{grad}F(x_k) d_k > 0$$

This is a contradiction to the prerequisites and so, the Armijo-rule is well-defined.

□

After the concept of the direction was defined and a procedure was introduced to the regulation of a suitable step width, it should be looked at this point on the algorithm by which the respective steps of iteration can be calculated (compare to [C.99], p.26).

##### **Algorithm 4.3.1 (Directed minimization)**

- (DM1) A  $x_0 \in U$  is chosen and  $k := 0$  is set.
- (DM2) If  $x_k$  suffices to a suitable criteria of stopping, the algorithm stops.
- (DM3) A suitable direction  $d_k$  is chosen.
- (DM4) A suitable step width  $\eta_k$  is calculated.
- (DM5) The step of iteration  $x_{k+1} = x_k + \eta_k d_k$  is done,  $k := k + 1$  is set and it is gone to step 2.

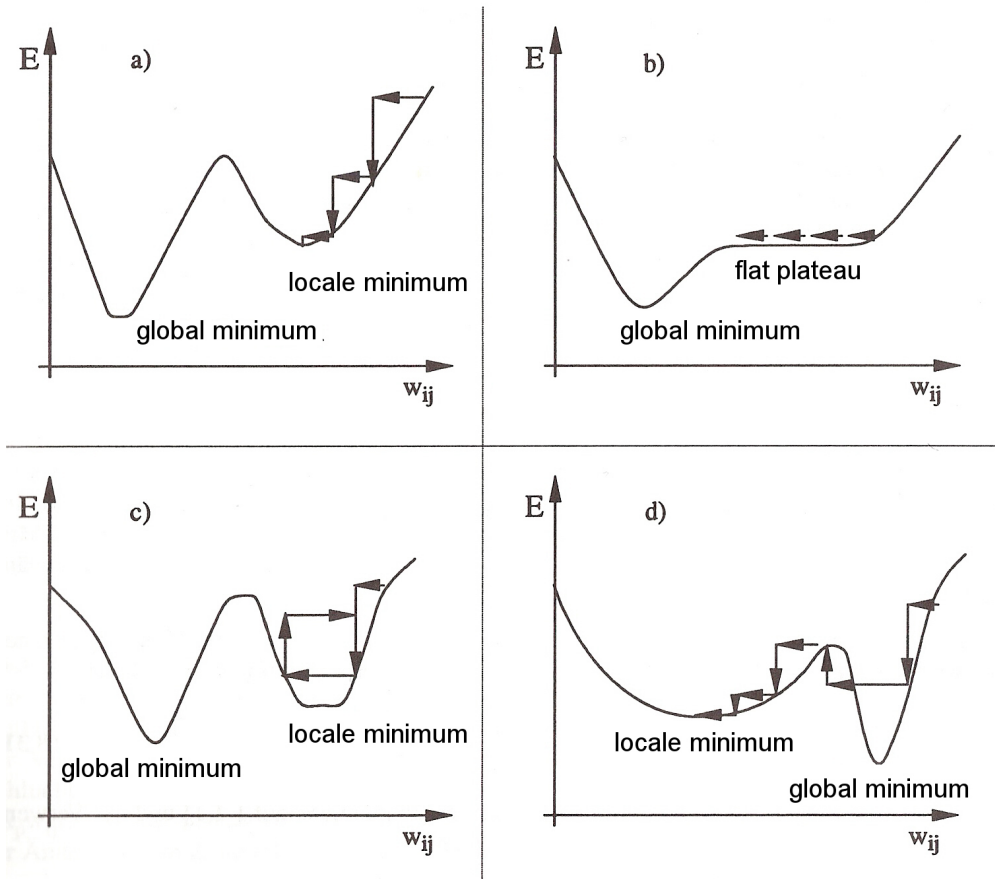
At this point it should be made clear, that the behaviour of this algorithm and the result generated with its substantially depend on how the direction and the step width have been chosen. Though by the choice of the direction it is guaranteed, that one moves to decreasing functional values there, but the whole consideration rests only on local information. If one looks, for example, at a direction which lies near with the gradient in the particular point or which is the gradient, then

it can be that one moves towards a local minimum. However, besides there are even other areas like saddle surfaces or maxima in which the gradient becomes smaller. With available saddle surfaces one can imagine that the tangents of single partial functions become horizontal in a whole area. Then the procedure will move further in the direction which is given by the course of the other partial functions. If one is in the area around a maximum, the algorithm “moves” in the direction of decreasing functional values from there. Nevertheless, an interesting aspect is if one uses by chance as a start value a point whose gradient is already equal to zero. Besides, one cannot suppose immediately that it concerns a local minimum in request. A possibility at this point in which the algorithm is finished to get on, is to consider itself the functional values in certain environments around the suitable point.

This can appear not only with the definition of the start value, but also otherwise. Besides, the choice of the step width plays an important role. With the Armijo-rule described on top it can appear that one reaches to himself with the calculated value to a point whose gradient is equal to zero which explains, however, for example, a local maximum. The step width is so calculated by this procedure that moves to a smaller functional value in the regarded direction and the biggest possible step width is chosen in a certain area. However, this does not mean that closer to the starting point a good minimum exist which is jumped over then. An other possibility consists in the fact that one reaches though to a local minimum, however, it would be possible with a bigger step width to find a better one. Some of the possible problems of this method are illustrate in figure 4.17.

Thereby it should be made clear that the described procedure can only use local information and, therefore, one can make only statements about a certain area.

This general procedure should be more exactly looked in the next segment by the definition of a certain direction given by the negative gradient vector in a certain



**Figure 4.17:** Problems of the directed minimization ([Zel96], p. 113, modified)



point.

**Algorithm 4.3.2 (Gradient descent algorithm)**

(GD1) A starting point  $x_0 \in U$  is chosen,  $\sigma, \beta \in (0, 1)$ ,  $\epsilon \geq 0$  and  $k := 0$  are set.

(GD2) If  $\|\text{grad}F(x_k)\| \leq \epsilon$ , then the algorithm stops.

(GD3)  $d_k = -\text{grad}F(x_k)$  is chosen.

(GD4) The step width  $\eta_k$  is calculated using the Armijo-rule.

(GD5) The step of iteration  $x_{k+1} = x_k + \eta_k d_k$  is done,  $k := k + 1$  set and gone to step 2.

A sequence  $(x_k)_k$  originates from the described algorithm. According to [C.99], p.69, to this procedure a convergence result is described which shows, that if there exist an accumulation point of this sequence, the gradient is of value zero at this point.

Applying on the partial derivations calculated by the Backpropagation Network-algorithm, the iterations for the particular weight matrices can be described in the following way:

$$\text{grad}E_P(W(k)) = \begin{pmatrix} \frac{\partial E_P(W(k))}{\partial w_1} \\ \vdots \\ \frac{\partial E_P(W(k))}{\partial w_{n_{whole}}} \end{pmatrix} \quad (4.11)$$

Using the gradient descent algorithm, it follows:

$$W(k+1) = W(k) + \eta_k \cdot (-\text{grad}E_P(W(k))) \quad (4.12)$$

#### 4 Artificial Neural Networks

With these nets two equations are distinguished, one for a weight in the last connecting layer and one for a weight in another layer. Besides, the consideration occurs in this order. One receives

- for a weight from the last connecting layer

$$\begin{aligned}
 w_{ij}(s+1) &= w_{ij}(s) + \eta_k \cdot \left( - \left( -\delta_{j \circ B}^P o_i^P \right) \right) \\
 &= w_{ij}(s) + \eta_k \cdot \left( - \left( - \left( t_j^P - o_j^P \right) f'_{out} \left( net_j^P \right) o_i^P \right) \right) \\
 &= w_{ij}(s) + \eta_k \cdot \left( t_j^P - o_j^P \right) f'_{out} \left( net_j^P \right) o_i^P
 \end{aligned}$$

- for a weight in another layer

$$\begin{aligned}
 w_{ij}(s+1) &= w_{ij}(s) + \eta_k \cdot \left( - \left( -\delta_{jB}^P o_i^P \right) \right) \\
 &= w_{ij}(s) + \eta_k \cdot \left( - \left( - \sum_k \left( \delta_{kB}^P w_{jk} \right) f'_{out} \left( net_j^P \right) \right) \right) \\
 &= w_{ij}(s) + \eta_k \cdot \sum_k \left( \delta_{kB}^P w_{jk} \right) f'_{out} \left( net_j^P \right)
 \end{aligned}$$

This calculation can be written down more clearly using matrices.

- Calculation for the last weight matrix

#### 4 Artificial Neural Networks

$$\begin{aligned}
 W_{l-1,l}(s+1) &= W_{l-1,l}(s) + \eta_k \cdot (- (G_{l-1,l}(s))) \\
 &= W_{l-1,l}(s) + \eta_k \cdot \left( - \begin{pmatrix} -\delta_{1aB}^P o_1 & \dots & \delta_{1aB}^P o_{m_n} \\ \vdots & \ddots & \vdots \\ -\delta_{m_l aB}^P o_1 & \dots & -\delta_{m_l aB}^P o_{m_n} \end{pmatrix} \right)
 \end{aligned}$$

- Calculation for the last weight matrix

$$\begin{aligned}
 W_{\tilde{i},\tilde{i}+1}(s+1) &= W_{\tilde{i},\tilde{i}+1}(s) + \eta_k \cdot (- (G_{\tilde{i},\tilde{i}+1}(s))) \\
 &= W_{\tilde{i},\tilde{i}+1}(s) + \eta_k \cdot \left( - \begin{pmatrix} \delta_{1B}^P o_1 & \dots & \delta_{1B}^P o_{m_{\tilde{i}}} \\ \vdots & \ddots & \vdots \\ \delta_{m_{\tilde{i}+1} B}^P o_1 & \dots & \delta_{m_{\tilde{i}+1} B}^P o_{m_{\tilde{i}}} \end{pmatrix} \right)
 \end{aligned}$$

In the historical development of analysing *feed-forward-networks* and especially BPNs, there was developed a statement of existence concerning the representation of mappings with certain properties by networks with certain characteristics. This result is related to the 13. problem of *Hilbert* postulated in [Hil00] and the solution given by *Kolmogorov* and *Lorentz* (see [Kol57] and [Lor86] for more details).

There, it was proofed, that an arbitrary multidimensional mapping  $F : K \longrightarrow \mathbb{R}^m$  with a compact domain  $K \subset \mathbb{R}^n$  can exactly be described by a *superposition* and *composition* of onedimensional mappings  $f : \mathbb{R} \longrightarrow \mathbb{R}$ .

Hecht-Nielsen used this in [Hec87] to proof, that there exists a *feed-forward-network* with *three layers*, which describes a continuous mapping  $F : K \longrightarrow \mathbb{R}^m$  with

a compact domain  $K \subset \mathbb{R}^n$  approximately with a certain error using onedimensional mappings of sigmoidal type. In [Hec92], he even showed, that this is valid for any  $L_2$ -mapping  $F : [0, 1]^n \rightarrow \mathbb{R}^m$ .

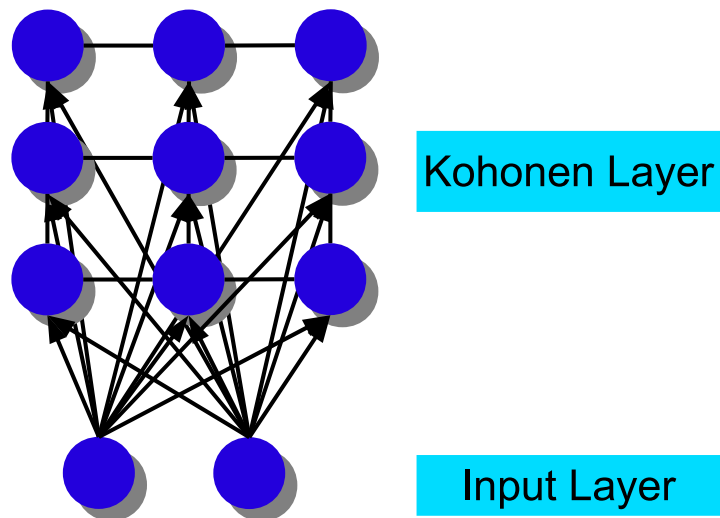
### 4.3.2 Self Organising Maps and Kohonen Networks

The concept of Self Organizing Maps or Kohonen Maps has been developed by Teuvo Kohonen in 1982 [Koh82]. The basic idea is to get a representation of a given input domain by using a neuron layer of a certain dimension, the so-called Kohonen layer. From the used input domain the network gets certain inputs and by using the Kohonen algorithm, the input domain is clustered into disjunct areas. Every neuron of the Kohonen layer is then responsible for a certain of this clustering areas.

The topology of the original Kohonen Network consists of two layers, an input layer and the mentioned Kohonen layer. In this Kohonen layer, the neurons are connected in a certain way, but these connections are only topological ones, there are no signals transported on them.

The dimension of the input layer depends on the dimension of the regarded input domain. The input neurons are connected to every neuron of the Kohonen layer with weighthened connections. Let the dimension of the input domain be  $n$ , then there are  $n$  input vectors and  $n$  weighthened connections lead to every neuron of the Kohonen layer (They are called Kohonen neurons from now on). These are summarized to the so-called codebook vector of each Kohonen neuron. In this case, an element of the input domain with  $n$  components can be presented to the ANN.

In figure 4.18, there is shown a SOM with an two dimensional input domain and a two dimensional square Kohonen layer of dimension  $3 \times 3$ .



**Figure 4.18:** Exemplary SOM with two dimensional input and a square Kohonen layer

While proceeding the Kohonen algorithm, stochastic input vectors are presented to the SOM and the codebook vectors are changed using similarities to these input vectors. Through this progress, the codebook vectors are moving in the input domain and so cluster it step by step.

The time, the ANN takes to reach an adequate level of representation depends on different parameters explained later on. According to them, there can also be some problems like a knotting of the codebook vectors.

The single steps of the Kohonen Algorithm developed by Teuvo Kohonen is well described in [Roj96], pp.391-399, on which the following description is related but written down in a precise mathematical way.

**Definition 4.3.4 (Characteristics of the Kohonen Networks)**

*The Kohonen algorithm uses the following parameters and functions:*

## 4 Artificial Neural Networks

- *Neighbourhood function  $\phi$*

$$\begin{aligned}\phi : \mathbb{R}_0^+ \times \mathbb{N}_0 &\longrightarrow [0, 1] \\ (n, t) &\mapsto \phi(n, t)\end{aligned}$$

where  $n \in \mathbb{R}_0^+$  represents the so-called topological distance and  $t$  the actual iteration step.

Alternatively,  $\phi$  can be described as a function depending on a certain neighbourhood radius  $r \in \mathbb{R}^+$  instead of  $t$ , what according to [Zel96], pp.182, can be better used for implementations.

- *Learning function  $\eta$*

$$\begin{aligned}\eta : \mathbb{N}_0 &\longrightarrow [0, 1] \\ (t) &\mapsto \eta(t)\end{aligned}$$

where  $t$  represents the actual iteration step.

Furthermore, there are an input layer with a dimension  $n$  depending on the regarded input domain and a Kohonen layer of a certain topology containing  $m$  neurons existing.

The neurons of the Kohonen layer are numbered and named as  $N_i$ ,  $i \in \{1, \dots, m\}$ , so that each one can be identified in a unique way. According to this, the  $n$ -dimensional codebook vector of each neuron is named as  $w_i \in \mathbb{R}^n$ ,  $i \in \{1, \dots, m\}$ .

### **Algorithm 4.3.3 (Kohonen algorithm)**

The following steps describe the general algorithm using a given SOM.

The start step would be a stochastic choose of the  $m$  weight or codebook vectors following a certain algorithm.

Then, the following steps are used within each iteration  $t$ :

## 4 Artificial Neural Networks

1. Stochastic choose of an input vector  $e_t$  following a certain algorithm depending on the iteration step  $t$ .
2. Computation of the codebook neuron  $w_k$  with minimal distance to the given input vector.
3. Changing of the weights using the following rule:

$$w_i(t+1) := w_i(t) + \eta(t) \cdot \phi(n_{i,k}, t) \cdot (e_t - w_i(t)), i \in \{1, \dots, m\} \quad (4.13)$$

where  $n_{i,k}$  is the topological distance of the codebook vectors  $w_i$  and  $w_k$ .

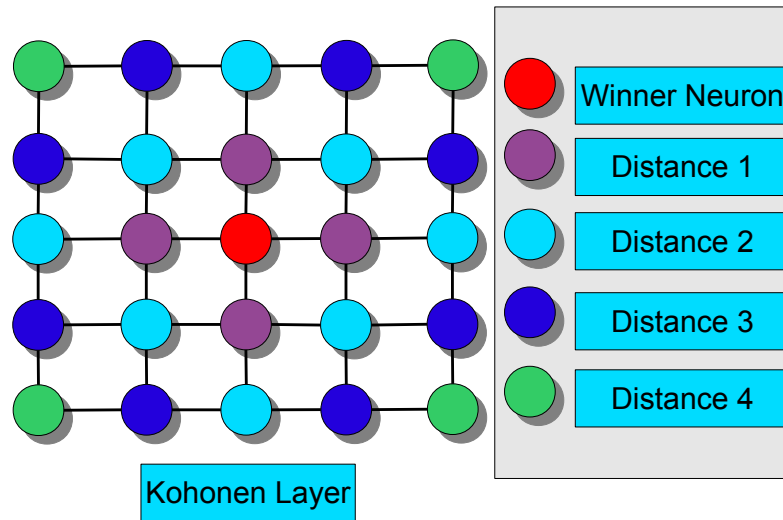
The minimal distance mentioned in step two can be mathematically described and computed by the term *metric* defined as a real-valued function  $d: \mathbb{R}_0^+ \rightarrow \mathbb{R}$  in chapter 7.1.

Using this term, the codebook neuron  $k$  with minimal distance to the given input vector  $e_t$  at iteration step  $t$  can be identified as:

$$d(w_k, e_t) = \min_{j \in \{1, \dots, m\}} d(w_j, e_t) \quad (4.14)$$

When this so-called “winner neuron” is known, the codebook vectors are updated following equation 4.13. By this, the single codebook vectors are moved a certain step in the direction of the regarded input vector depending on the topological relation to the winner neuron. Because of this, the distance to the input neuron is only important to get the winner neuron, the update of the codebook vectors depends on the topological structure of the Kohonen layer and so this topology is preserved while mapping the input domain. The topological neighbourhood can

be defined in different way, the so-called “block party distance” is often used and is calculated as shown in figure 4.19 (refer to [Hec90], p.140).



**Figure 4.19:** Kohonen Layer with block party distance

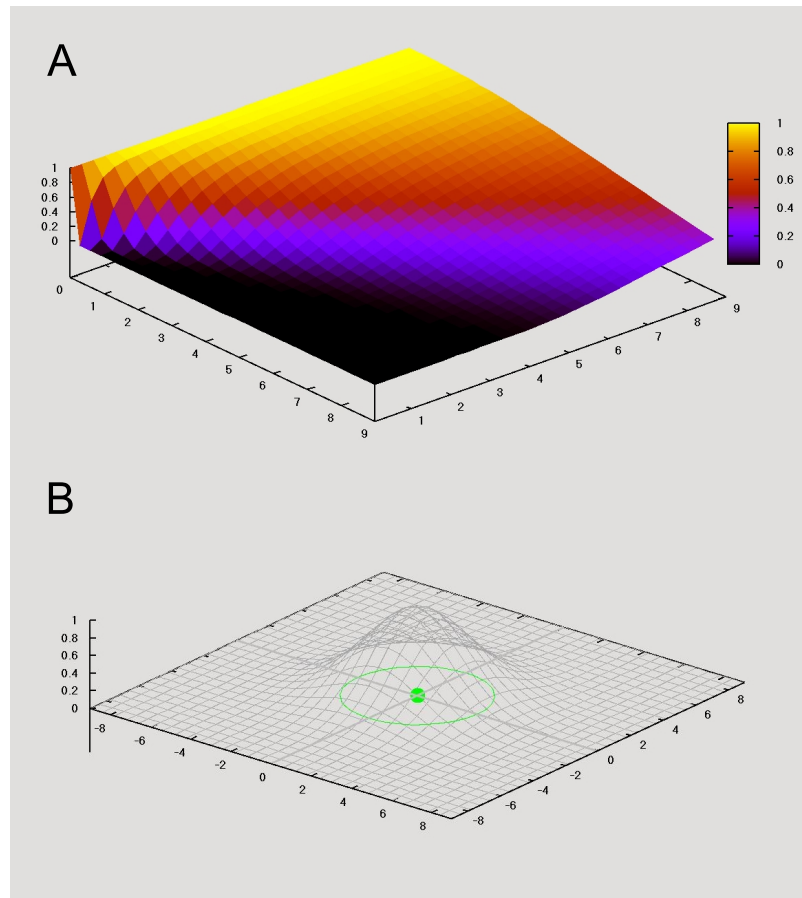
With all these parameters, the update rule 4.13 can be applied. The neighbourhood function  $\phi$  can also be very different, a popular choice is a so-called gauss function which can be defined as follows according to [Zel96], p. 182:

$$\begin{aligned} \phi_{gauss} : \mathbb{R}_0^+ \times \mathbb{R}^+ &\longrightarrow \mathbb{R}^+ \\ (n, d) &\longrightarrow \phi_{gauss}(n, d) := e^{-\left(\frac{n}{d}\right)^2} \end{aligned}$$

where  $n$  is the topological distance and  $d$  is the neighbourhood radius. In figure 4.20 the graph of this function is shown on the top. On the bottom, there is shown a graph, that illustrates the situation of neighbourhood with the neighbourhood radius 3 of a neuron at the point  $(0, 0)$ .

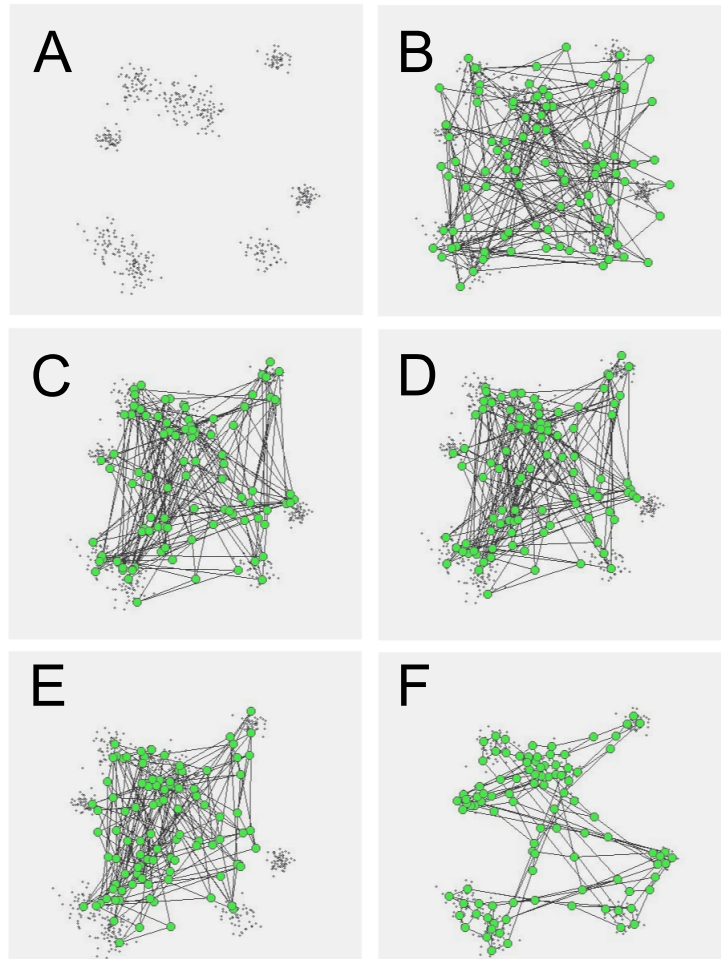
As an example, there is a given two dimensional discrete input domain. The used Kohonen layer is two dimensional with the dimension  $10 \times 10$ . The figure 4.21 shows the development of the codebook vectors after certain iteration steps using





**Figure 4.20:** A Graph of the function  $\phi_{gauss}$ , B Illustration of the neighbourhood situation with radius 3

the java applet DemoGNG on [Loo].



**Figure 4.21:** Representation process of a discrete input domain: A input data, B initial codebook vectors, C 10 iterations, D 30 iterations, E 100 iterations, F 30000 iterations

## 5 Principles of object-oriented analysis

When regarding a complex system like the **environment** or a certain **health system** in cases of modeling, it is necessary to concentrate on a reasonable collection of important system components using developed and/or existing methods. One possible perspective is the concept of Object Orientation (OO), which is applicable in many contexts. The advantage of using this method is a quality that can be described with the term *Construction Kit Principle (CKP)*. Which means that certain constructs of OO named *classes*, can be “plugged” together within a system under the condition of the availability of adequate “interfaces”. There are thus many advantages for describing a complex system with methods of OO.

In this section, the basic elements and background information concerning OO are provided.

The term and concept OO is mainly used in informatics as Object Oriented Programming (OOP). However, not only the process of programming is in the focus, but prior to this also the analysis of the relevant structures, which are described by the term OOA.

The beginning of OOP can be allocated to the programming language “*Smalltalk-80*”, which was developed in the period from 1970 until 1980 by Xerox in the

## 5 Principles of object-oriented analysis

USA. To date, different programming languages concerning OOP have been developed. In 1995, a generalised description language named *Unified Modeling Language (UML)* was published by Booch, Rumbaugh and Jacobson in [BRJ95] and in 1997, the *Object Management Group* accepted this as a standard notation. For a more detailed description of the historical coherence refer to [Bal99], p.3-4. Based on [Boo06], p.15, UML can be described as a graphical language for visualising, specifying and documenting of a software system. Using this language, it is possible to describe the structure, the data flow and the existing elements in a standardised way. The notation in this thesis follows the specifications of UML as they can be found for example in [Bal01].

A good software tool for working with OO is *ArgoUML*, which is available at [arg]. Because this is OpenSource software, it can be freely be used and the application of OO principles is simplified, particularly concerning the aim of a consequent use of OpenSource within the whole project.

The advantage of ArgoUML consists in the possibility of working whilst using the specifications of UML in a clear and structured way. The atomic elements are represented here. Furthermore, the user has the possibility of analysing and structuring a certain project using the main types of UML diagrams as they are described in [Bal01], p.19 et seqq. Additionally, even so-called *checklists* are implemented in here to provide a tool to using UML in an optimal and economic manner. The possibility of creating source code for the developed elements directly in diverse formats (SQL, C++, PHP4, PHP5, JAVA, CSHARP) is of great benefit.

The following atomic elements of UML are described more detailed at this point:

- Object
- Class

## 5 Principles of object-oriented analysis

- Attribute
- Operation
- Associations
- Interface
- Package

Generally spoken, a *class* is an abstract concept of an *object*. In this coherence, an object is associated with a certain class. Every object can be characterised by its *name*, its *attributes* and their *values*, and it reacts with certain *operations* on the surrounding. In addition to that, an object can be linked to other objects, one refers to the “knowledge” of them. These links are realised as so-called *associations*. An object is a dynamic instance of a class. The according class defines the attributes, the operations and the associations for a collection of objects.

The kind of data, that can be used by the objects of a certain class are defined using attributes. Every attribute is characterised by its name, its data type, its initial value and its features. According to [Bal01], p.7, the data type can be one of the types usually used in a programming language (boolean, string, integer), a kind of enumeration or even classes itself. A possible feature is for example specified by mentioning the term *frozen*, which inhibits the modification of the according value.

To define activities for the objects of a certain class, so-called *operations* are used. These operations can be applied to every attribute within the regarded class and the according name should represent the performed activity. *Associations* identify linkages between classes and as a consequence between derived objects. They can be named and labeled with a cardinality of association.

To facilitate classes the application of another classes methods to this class, so-called *interfaces* can be used. From a class, more specialised classes with additional

## 5 Principles of object-oriented analysis

features can be derived. In this coherence, one refers to transmission and the original class is then called the *basis class*.

For grouping and summarizing elements of UML, for example classes, so-called *packages* are defined. In doing this, a higher level of abstraction can be used for the description of the regarded system.

The way of describing a system using UML is dependent on the aim and the perspective of this description. So in a project group, some people are only interested in the abilities of certain components, while others are interested in the detailed processing of the same components. Even though they pursue the same main aim, there may be the necessity of regarding certain steps within this progress of development from different perspectives.

# 6 Proposal of a “Spatial Toolbox”

## 6.1 Localisation of the SPATTB

The concept of the SPATTB is incorporated into the concept of the EWARS, where the two main tasks are identified in the process of conception in this thesis. The first one acts like an *appropriate “tool”* for other components of the EWARS by providing calculation and processing methods. For doing this, *compatible interfaces* have to be defined. The second *use case* consists of directly dealing with queries of the system. The type of usage is dependent on the actual query and the intention of the decision maker or respectively the expert. In figure 6.1, this use cases are illustrated in a structured way.

As already mentioned above, the term *interface* is central to the conceptualisation and building up of the EWARS. For implementing the “Spatial-Toolbox” as part of the whole system, it has to be guaranteed that the system and respectively the using person can apply it. Thus special attention has to be paid to the manner though which input is gauned and output is provided.

The proposed “Spatial-Toolbox” requires the two abilities of *flexibility* and *usability*. This means that it should be possible to extend the provided methods in an appropriate way and as already described, define the interfaces accurately.

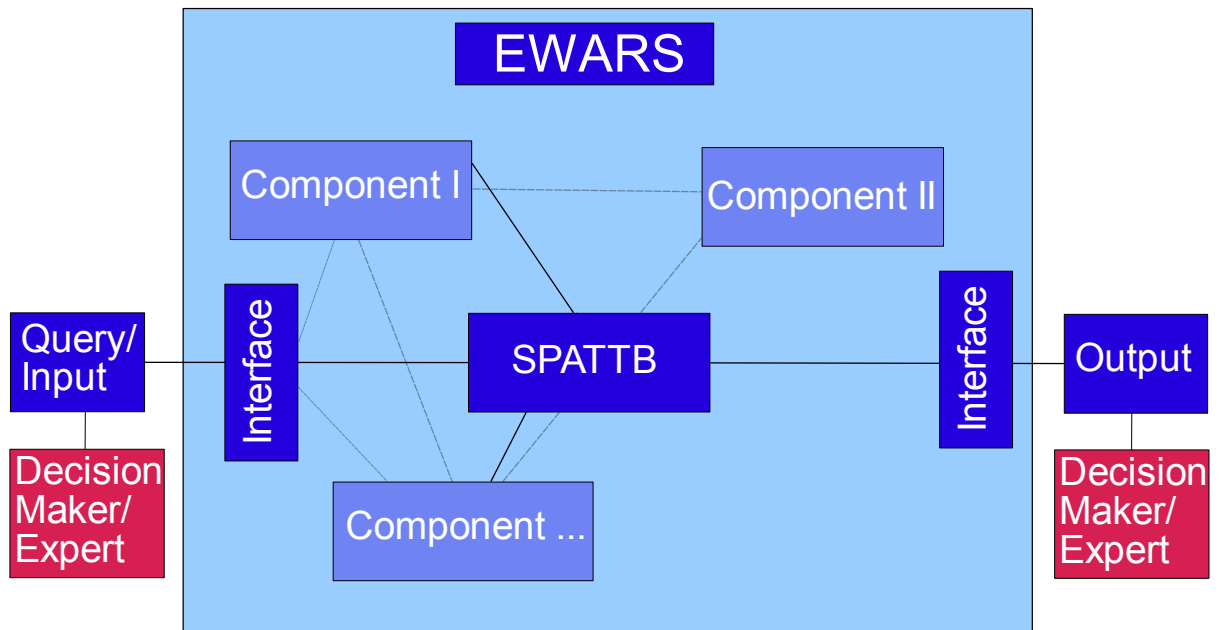


Figure 6.1: Use cases related to the SPATTB

## 6.2 Application and customisation of OpenSource components

As conceptualised for the whole EWARS, the “Spatial-Toolbox” itself is defined and described only using OpenSource components. The advantages of doing this are the availability of free components, the alignment and conformation with existing licenses and the benefits bained through the knowledge of the OpenSource community. The challenge that has to be mastered within this concept is the description of a certain “use case”, which means a special application in the context of the SPATTB, by interfacing the used components in an appropriate way. In the following section the conditions that are required in the authors opinion for choosing such a component are described.



## 6.3 Workflow for building up SPATTB components

To support the claim of flexibility in the meaning of extending and improving the SPATTB, an appropriate workflow is proposed in this section.

**WF I      Definition of the according “use case(s)”**

There is much importance to know, what is the aim of constructing that component and what benefits can be achieved using it. It has to be defined, if there should be one or more applications and what they are about. This knowledge features the basis for the following steps. In terms of modeling, it has to be taken into account, which way of describing a certain situation and application makes sense and which restrictions and features can be accepted.

**WF II      Description of the according input and output information**

The awareness of these facts is the prerequisite for providing accurate interfaces. There has to be particular information about the specifications of the input and the output. Based on this, it can be decided, which methods can be taken into account or what modifications are eventually necessary for using them.

**WF III      Accurate identification of sub-processes within the regarded component**

The single steps of each use case have to be defined in order to construct these as optimal as possible.

**WF IV      Mathematical analysis of the different sub-processes**

A significant benefit of mathematics is the ability of abstraction and getting an arbitrary point of view. Mathematics is necessary to guarantee the usability of the considered methods for a certain application.

**WF V      Finding of appropriate software for implementation**

One important condition of choosing a certain software is that it is Open-Source software. In regard to implementation in an appropriate way with the ability of creating usable webinterfaces, command line use is necessary. Furthermore, the definition and implementation of interfaces for describing the whole process of the regarded use case as a sequence of sub-processes has to be taken into account.

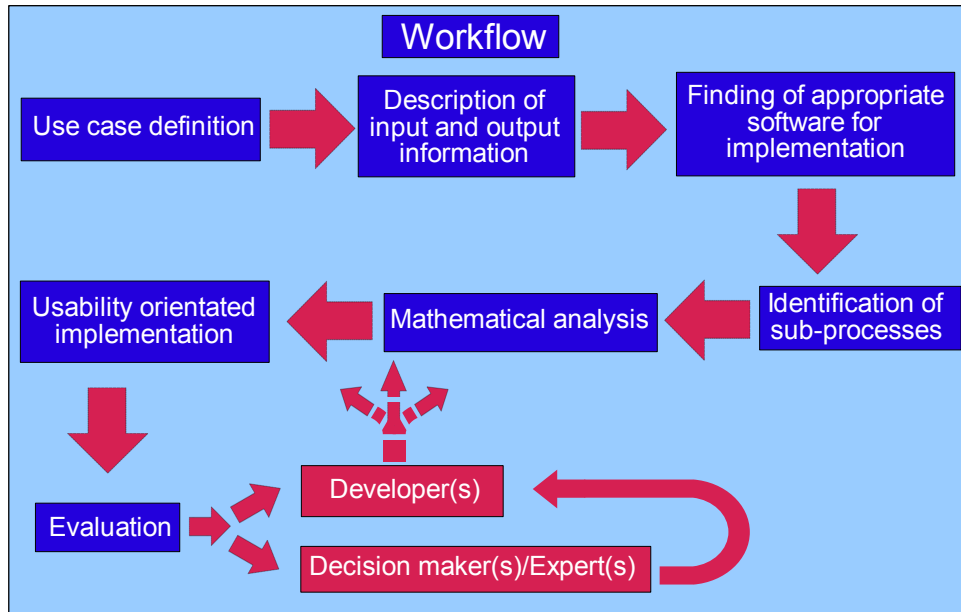
**WF VI Usability orientated implementation**

Implementing the regarded component, the certain user and its role has to be taken into account. Balzert also describes this in terms of describing software using the UML (refer to [Bal01], p.39 et seqq.). It is essential to regard the particular kind of expert knowledge in view of creating an appropriate structure for the according Graphical User Interface (GUI).

**WF VII Evaluation**

One important kind of evaluation is testing the usability of the created component. It has to be regarded, if the whole process is described. Thereby, not only “test” data has to be used, but also “real”, “authentic” data, because with that kind of data the system should work in everyday use. If there are more than one components, that are usable for a certain use case, it is an asset to provide information about differences of the provided output for the decision maker. In this coherence, it has to be taken into account, if that feature makes sense for the particular persons in terms of the ability of using that kind of information. So, there are two kinds of evaluation. The first one is important for the developer(s) of that component in regard of optimisation it. The second one is important for the users with respect to the described restrictions. Both kinds are important elements of this workflow. A fact that is also very important is the possibility for the users of giving feedback to the developers with regard to the optimisation.

This workflow is illustrated in figure 6.2.



**Figure 6.2:** Workflow for develop an additional component for the SPATTB

In the context of evaluation, the term *quality* is used sometimes in writing and/or in speaking. To obviate misunderstandings, the authors insight for this term will be stated at this point. In this coherence, it is important to separate the term *quality* from the term *difference*, because they usually cannot be used as synonyms.

# 7 Measure theory on topological and function spaces

## 7.1 Basic definitions and prerequisites

### 7.1.1 Topological space

Since in chapter 9, there are developed methods for implementing ANNs in the context of dealing with spatial data within the SPATTB, different definitions and prerequisites concerning the term *space* and especially *topological space* in mathematics are required. These different ideas and concepts are pointed out describing also the coherences and the basing *hierarchical structure*.

A very general concept of a mathematical space is the *topological space* based on the general idea of a so-called *neighbourhood*. This concept was mainly developed at the beginning of the 20th century by *Felix Hausdorff* in [Hau14], based on the concept of a *set* developed by *Georg Cantor* in [Can97].

I will refer to a special type of topological spaces in definition 7.1.16. This is the concept of a so-called *metric space*, which was described by *Maurice Fréchet* in [Fre06]. Here, the term *distance* plays an important role. For a more detailed

description of the historical development refer to [Jän97], pp. 3-6 and [Que01], pp. 327-334.

The basic idea of mathematical topology is the mentioned *topological space*, which can be defined in the following way according to [Jän97], p.5:

**Definition 7.1.1 (Topological Space)**

A **topological space** is a pair  $(X, \mathcal{O})$  consisting of a set  $X \neq \emptyset$  and a set  $\mathcal{O}$  of subsets of  $X$  with the following properties, while the elements  $U$  of  $\mathcal{O}$  are called **open sets**:

- *Axiom 1: Any union of open sets is open again.*

$$\Upsilon \subseteq \mathcal{O} \Rightarrow \bigcup_{U \in \Upsilon} U \in \mathcal{O}$$

- *Axiom 2: The intersection of two open sets is open.*

$$U, V \in \mathcal{O} \Rightarrow U \cap V \in \mathcal{O}$$

- *Axiom 3: The empty set  $\emptyset$  and the set  $X$  are open.*

$$\emptyset \in \mathcal{O}, \quad X \in \mathcal{O}$$

This set  $\mathcal{O}$  of subsets of  $X$  is then called a **topology** on the set  $X$ .

By this definition, the set  $X$  is divided into two types of subsets, the open ones and the closed ones. The **open subsets** are  $U \in \mathcal{O}$ , the **closed subsets** are the corresponding *complements*  $V := X \setminus U$ .

An important idea in topology is the concept of *neighbourhood*. This concept is essential for the definition and usage of the terms *convergence* and *continuity* in a topological space.

According to [Que01], p. 25, this can be defined in the following way:

**Definition 7.1.2 (Neighbourhood)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space and  $x_0$  an arbitrary Element of  $X$ . A

subset  $U \subset X$  is then named **neighbourhood** of the Element  $x_0$  when there exists an open set  $O \in \mathcal{O}$  with  $x_0 \in O \subset U$ .

The set  $\mathfrak{U}(x_0)$  which contains all existing neighbourhoods of an element  $x_0$  is named **neighbourhood system** of  $x_0$  or **neighbourhood filter** (the term filter is defined in the following section 7.2).

In this context, the term *neighbourhood basis* as a subset of the neighbourhood system is an important concept. In some cases it is possible, to assign statements or theorems according to neighbourhood filters on neighbourhood bases and reduce the number of regarded sets. One of this cases is lemma 7.2.1 in section 7.2.

**Definition 7.1.3 (Neighbourhood basis)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space,  $x_0$  an arbitrary Element of  $X$  and  $\mathfrak{U}(x_0)$  the according neighbourhood system. A subset  $\mathcal{B}(x_0) \subset \mathfrak{U}(x_0)$  is then named **neighbourhood basis** of the Element  $x_0$  when the following is true:

$$\forall_{U \in \mathfrak{U}(x_0)} \exists_{B \in \mathcal{B}(x_0)} : B \subset U$$

In topology, a certain set  $A \subset X$ , while  $(X, \mathcal{O})$  is an arbitrary topological space, can be characterised using the terms *border*, *interior* and *closure*. The definitions and interpretative descriptions of these terms are given in the following, basing on [Que01], p.28:

**Definition 7.1.4 (Topological characteristics of a set)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space,  $x_0$  an arbitrary Element of  $X$  and  $A \subset X$  an arbitrary subset of  $X$ . The following terms define characteristics of this set and its elements:

- (1) An element  $x_0 \in A$  is named **interior point** of  $A$ , when  $A$  is a neighbourhood of this element.

- (2) The set  $\mathring{A}$  with  $\mathring{A} := \{x_0 \in A \mid x_0 \text{ is an interior point}\}$  is named the **interior** of the set  $A$ .
- (3) The interior  $\mathring{A}$  is the biggest open set within the set  $A$  and can be constructed in the following way:  $\mathring{A} = \bigcup_{U \subset A, \text{ open}} U$
- (4) An element  $x_0 \in A$  is named **border point** of  $A$ , when each neighbourhood of  $x_0$  does intersect both,  $A$  and its complement set  $A^c$ . This means the following:  $\forall_{U \in \mathfrak{U}(x_0)} : U \cap A \neq \emptyset \neq U \cap A^c$
- (5) The set  $\partial A$  with  $\partial A := \{x_0 \in A \mid x_0 \text{ is an border point}\}$  is named the **border** of the set  $A$ .
- (6) The set  $\bar{A} := \left\{ x \in X \mid \forall_{U \in \mathfrak{U}(x_0)} : U \cap A \neq \emptyset \right\}$  is named the **closure** of the set  $A$ .
- (7) The closure  $\bar{A}$  is the smallest closed set which contains the set  $A$  and can be constructed in the following way:  $\bar{A} = \bigcap_{U \supset A, \text{ closed}} U$
- (8) The set  $A$  is named **dense** in  $X$ , when  $\bar{A} = X$ .
- (9) The border of the set  $A$  is the following:  $\partial A = \bar{A} \setminus \mathring{A}$

The topology  $\mathcal{O}$  of a topological space  $X$  can be characterised regarding a special system of sets in  $X$  named *basis of the topological space*. This system can be defined in the following way as it is done in [Que01], p.23:

**Definition 7.1.5 (Basis and subbasis of a topological space)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space.

- (1) A system  $\mathfrak{B} \subset \mathcal{O}$  is then named **basis of the topological space**, if each arbitrary open set in  $X$  can be denoted as a union of sets out of  $\mathfrak{B}$ .
- (2) A system  $\mathcal{S} \subset \mathcal{O}$  is named **subbasis of the topological space**, if the set of all finite intersections of sets out of  $\mathcal{S}$  is a basis of the topology.

Hence, a topological space can be described using its basis and additionally a pair of topological spaces is equal, if their basis is equal ([Que01], p.23).

Using the terms *neighbourhood basis* and *basis of a topological space*, a classification of topological spaces is possible in the following way ([Que01], p.27):

**Definition 7.1.6 (First and second axiom of countability)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space.

- (1) *The first axiom of countability is fulfilled, if each element of  $X$  features a countable neighbourhood basis.*
- (2) *The second axiom of countability is fulfilled, if the set  $\mathcal{O}$  of open sets features a countable basis of the topological space.*

In the following a topology according to an arbitrary subset of a given topological space  $(X, \mathcal{O})$  is regarded according to [Bar07], pp.108-109.

**Definition 7.1.7 (Relative topology, subspace)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space. Further, let  $A \subseteq X$  be an arbitrary subspace. Then, the **relative topology** of  $\mathcal{O}$  in  $A$  is defined by the following set:

$$\mathcal{O}|_A := \{O \cap A \mid O \in \mathcal{O}\}$$

The according topological space  $(A, \mathcal{O}|_A)$  is then called **subspace** of  $(X, \mathcal{O})$ . Instead of the term **relative topology**, the terms **subspace topology**, **induced topology** and **trace topology** can synonymously be used.

**7.1.2 Continuity**

To describe relationships between topological spaces, the term *continuity* can be used. Regarding the set  $\mathbb{C}$ , continuity of a mapping  $f : \mathbb{C} \rightarrow \mathbb{C}$  can be described in the following way: Evanescent changings in the *argument* of that



function leads to evanescent changings in the value of the mapping. In *metric spaces* (see 7.1.16), this changings can be described with the term *distance*, but in arbitrary topological spaces, the definition of *continuity* has to be done more abstract, because distance is not defined there in general.

According to [Que01], p.30, the concept of *continuity* in topological spaces can be defined in the following way:

**Definition 7.1.8 (Continuity)**

Let  $(X, \mathcal{O}_1)$  and  $(Y, \mathcal{O}_2)$  be arbitrary topological spaces. A mapping  $f : X \longrightarrow Y$  is named **continuous**, if the inverse images of open sets in  $(Y, \mathcal{O}_2)$  are open in  $(X, \mathcal{O}_1)$ , that means:

$$f : X \longrightarrow Y \text{ continuous} \Leftrightarrow \forall_{O \in \mathcal{O}_2} : f^{-1}(O) \in \mathcal{O}_1$$

This definition is in an equivalent way possible using closed sets.

This definition of continuity describes this term according to the mapping as a whole. Continuity at the element  $x_0 \in X$  can be defined in the following way:

**Definition 7.1.9 (Pointwise continuity)**

A mapping  $f : X \longrightarrow Y$  between two arbitrary topological spaces  $(X, \mathcal{O}_1)$  and  $(Y, \mathcal{O}_2)$  is named **pointwise continuous**, if for all neighbourhoods  $V \in \mathfrak{U}_2(f(x_0))$  of  $f(x_0)$  exists a neighbourhood  $U \in \mathfrak{U}_1(x_0)$  of  $x_0$  and  $f(U) \subset V$ , that means:

$$f : X \longrightarrow Y \text{ pointwise continuous at } x_0 \in X \Leftrightarrow \forall_{V \in \mathfrak{U}_2(f(x_0))} : f^{-1}(V) \in \mathfrak{U}_1(x_0).$$

The following theorem provides a facilitation for proving the continuity of a certain mapping (see [Que01], p.30).

In the proof, the following coherence is used:

$$\bigvee_{\substack{A_i \in \mathcal{O}_1 \\ i \in I}} f^{-1} \left( \bigcup_{i \in I} A_i \right) = \bigcup_{i \in I} f^{-1}(A_i)$$

and

$$\bigvee_{\substack{A_i \in \mathcal{O}_1 \\ i \in I}} f^{-1} \left( \bigcap_{i \in I} A_i \right) = \bigcap_{i \in I} f^{-1}(A_i)$$

**Theorem 7.1.1 (Subbase criterion of continuity)**

Let  $(X, \mathcal{O}_1)$  and  $(Y, \mathcal{O}_2)$  be arbitrary topological spaces and let  $\mathcal{S}(\mathcal{O}_1)$  and  $\mathcal{S}(\mathcal{O}_2)$  be arbitrary particular subbases.

A mapping  $f: X \rightarrow Y$  is continuous, if and only if for any set  $S \in \mathcal{S}(\mathcal{O}_2)$   $f^{-1}(S)$  is open in  $(X, \mathcal{O}_1)$ .

*Proof.* Let  $O_Y \in \mathcal{O}_2$  be an arbitrary open set in  $Y$  and let  $O_Y = \bigcup_{j \in J} \bigcap_{i=1}^n S_i$  with appropriate sets  $S_i \in \mathcal{S}(\mathcal{O}_2), n \in \mathbb{N} \cdot$ . This can be done for every open set in  $\mathcal{O}_2$  according to definition 7.1.5. Then, the following is valid:

$$\begin{aligned} f^{-1}(O_Y) &= f^{-1} \left( \bigcup_{j \in J} \bigcap_{i=1}^n S_i \right) \\ &= \bigcup_{j \in J} f^{-1} \left( \bigcap_{i=1}^n S_i \right) \\ &= \bigcup_{j \in J} \bigcap_{i=1}^n f^{-1}(S_i) \end{aligned} \tag{7.1}$$

Any open set in  $\mathcal{O}_2$  can be expressed in the way done above using a finite subset of the regarded subbase.

Hence, it is adequate to proof, that for any set  $S \in \mathcal{S}(\mathcal{O}_2)$   $f^{-1}(S)$  is open in  $(X, \mathcal{O}_1)$ , because the last expression in 7.1 is an element out of  $\mathcal{O}_1$  regarding definition 7.1.1.

□

So, the term continuity draws conclusions about the *preimages* of open sets, but none about the *images* of open sets. But if there exists a relation between open sets in the domain and open sets in the codomain in the following way, the mapping is called *open* ([Que01], p.32):

**Definition 7.1.10 (Open mapping)**

Let  $(X, \mathcal{O}_1)$  and  $(Y, \mathcal{O}_2)$  be arbitrary topological spaces. A mapping  $f : X \longrightarrow Y$  is named **open**, if the images of open sets in  $(X, \mathcal{O}_1)$  are open in  $(Y, \mathcal{O}_2)$ , that means:

$$f : X \longrightarrow Y \text{ open} \Leftrightarrow \forall_{O \in \mathcal{O}_1} : \{f(O) \mid O \in \mathcal{O}_1 \subset \mathcal{O}_2\}.$$

Mappings between topological spaces, which are continuous, bijective and open are named *topological homeomorphisms* and can be defined as described in [Que01], p.32 as follows:

**Definition 7.1.11 (Topological homeomorphism)**

A bijective mapping  $f : (X, \mathcal{O}_1) \longrightarrow (Y, \mathcal{O}_2)$  is named **topological mapping** or **topological homeomorphism**, if  $f$  and  $f^{-1}$  are both continuous mappings. Then, the topological spaces  $(X, \mathcal{O}_1)$  and  $(Y, \mathcal{O}_2)$  are called **homeomorph**.

According to [Jän97], p.17 and definition 7.1.8, the following equivalence can be formulated:

$$f : X \longrightarrow Y \text{ is an homeomorphism} \Leftrightarrow \left( \forall_{O \in \mathcal{O}_1} : O \text{ open} \Leftrightarrow f(O) \text{ open} \right)$$

In this thesis, homeomorph spaces are written as:

$$(X, \mathcal{O}_1) \cong (Y, \mathcal{O}_2).$$

### 7.1.3 Product topology

While regarding families of topological spaces  $(X_i, \mathcal{O}_i)$ , it is interesting to have a look at the product of the particular topologies. In the following definition, this fact is described more precisely basing on [Que01], p.40:

**Definition 7.1.12 (Product space, product topology)**

Let  $I$  be an indexing set and  $(X_i, \mathcal{O}_i)$ ,  $i \in I$  be a family of topological spaces. Furthermore,  $X := \prod_{i \in I} X_i$  describes the cartesian product of these spaces and  $p_i : X \longrightarrow X_i$  the particular projection. Then, the **product topology**  $(X, \mathcal{O})$  is defined using the following basis:

$$\mathcal{B} := \left\{ \bigcap_{k \in K} p_k^{-1}(O_k) \mid \forall_{k \in K} O_k \text{ is open, } K \text{ is a finite subset of } I \right\}$$

$(X, \mathcal{O})$  is then named **product space** or **topological product** of the spaces  $(X_i, \mathcal{O}_i)$ ,  $i \in I$ .

Regarding this product topology, the set  $\mathcal{S} := \left\{ p_i^{-1}(O_i) \mid \forall_{i \in I} O_i \in \mathcal{O}_i \right\}$  defines a subbasis.

In the following, an important theorem concerning continuity of mappings between product spaces is given according to [Que01], p.41.

**Theorem 7.1.2 (Product spaces and continuity)**

Let  $(Y, \mathcal{O})$  be an arbitrary topological space. Additionally, let  $I$  be an indexing set and  $(X_i, \mathcal{O}_i)$ ,  $i \in I$  be a family of topological spaces. Further,  $X := \prod_{i \in I} X_i$  defines the according product space as defined in 7.1.12.

## 7 Measure theory on topological and function spaces

A mapping  $g: Y \longrightarrow X$  is continuous, if and only if for each  $i \in I$  the mapping  $g_i := p_i \circ g$  is continuous.

*Proof.*  $\Rightarrow$ : It has to be proven:  $g$  is continuous  $\Rightarrow \forall_{j \in I} : g_j = p_j \circ g$  is continuous

It is valid, that a composition of mappings is continuous, if every participating mapping is continuous. As a prerequisite,  $g$  is continuous. Hence, it has to be shown, that  $\forall_{j \in I} : p_j$  is continuous. Since every  $p_j$  is defined as  $p_j : X \longrightarrow X_j$ , this means, that the preimage  $p_j^{-1}(O)$  of an arbitrary open set  $O \subseteq X_j$  has to be open in  $X$ . Regarding the basis  $\mathcal{B}$  of the product topology defined in definition 7.1.12, every preimage of that type is an element of  $\mathcal{B}$  and consequently an open set in  $X$ , what had to be shown.

$\Leftarrow$ : It has to be proven:  $\forall_{j \in I} : g_j = p_j \circ g$  is continuous  $\Rightarrow g$  is continuous

It is sufficient to show, that every preimage according to  $g$  of an element out of the subbasis  $\mathcal{S}$  defined in definition 7.1.12 is an open set in the space  $Y$ . The elements out of  $\mathcal{S}$  are of the type  $p_i^{-1}(O)$ ,  $O$  is open in  $X_i, i \in I$ . As prerequisite, the mappings  $g_i, i \in I$  are continuous. So let  $O \in X_i$  be an arbitrary open set out of the space  $X_i, i \in I$ . The according element out of the subbasis  $\mathcal{S}$  is  $p_i^{-1}(O)$ . The preimage of that element according to the mapping  $g$  is then  $g^{-1}(p_i^{-1}(O))$ . Regarding the definition of the mappings  $g_i, i \in I$ , the following equation is valid:

$$g^{-1}(p_i^{-1}(O)) = g_i^{-1}(O)$$

Because of the continuity of every  $g_i, i \in I$ , this is an open set in the space  $Y$ . Due to the arbitrary choice of  $O$ , the mapping  $g$  itself is continuous, what had to be shown.

□

### 7.1.4 Classification of topological spaces

In the historical development of topology, there has been established a number of various topological spaces with certain features and restrictions. During this section, some of these spaces are regarded in a closer way. A system that bases on the topological discrimination of elements and sets in topological spaces is named *separation axioms* or *Tychonoff separation axioms* after the russian mathematician Andrei Nikolajewitsch Tychonoff living in the 20th century. These axioms are normally denoted with the letter “T”, which is derived from the german word “Trennungsaxiom”. The notation of these separation axioms used in this thesis is the following (basing on [Que01], p.84):

**Definition 7.1.13 (Tychonoff separation axioms)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space. This space can be restricted by the following **separation axioms**:

$T_0$  (Kolmogoroff): For each pair of different elements  $x_1$  and  $x_2$  in  $X$  one of them has a neighbourhood, which does not contain the other one.

$T_1$  (Fréchet): For each pair of different elements  $x_1$  and  $x_2$  in  $X$  both of them have neighbourhoods, which does not contain the other one.

$T_2$  (Hausdorff): Each pair of different elements  $x_1$  and  $x_2$  in  $X$  have disjunct neighbourhoods.

$T_3$  (Vietoris): Each closed set  $A \subset X$  and each element  $x \in X \setminus A$  have disjunct neighbourhoods.

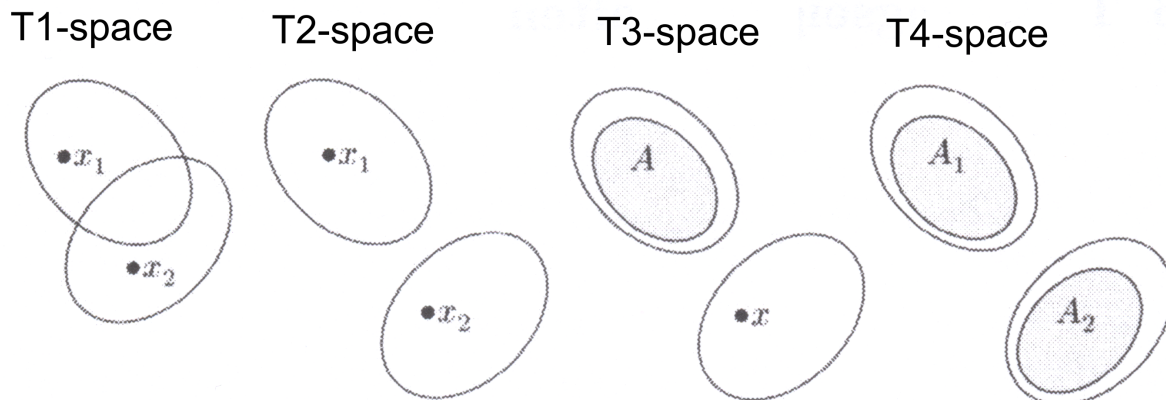
$T_{3_a}$  For each closed set  $A \subset X$  and each element  $x \in X \setminus A$  there exists a continuous mapping  $f : X \longrightarrow [0, 1]$ , with  $f(x) = 1$  and  $f(A) \subset \{0\}$ .

$T_4$  (Tietze): For each disjunct pair of closed subsets  $A_1$  and  $A_2$  of  $X$  there does exist disjunct neighbourhoods.

The topological space  $(X, \mathcal{O})$  is named  $T_i$ -space for  $i \in \{1, \dots, 4\}$  or rather  $i = 3_a$ ,

if it fullfills the according axiom.

These axioms are illustrated in figure 7.1.



**Figure 7.1:** Illustration of the defined separation axioms ([Que01], p.84)

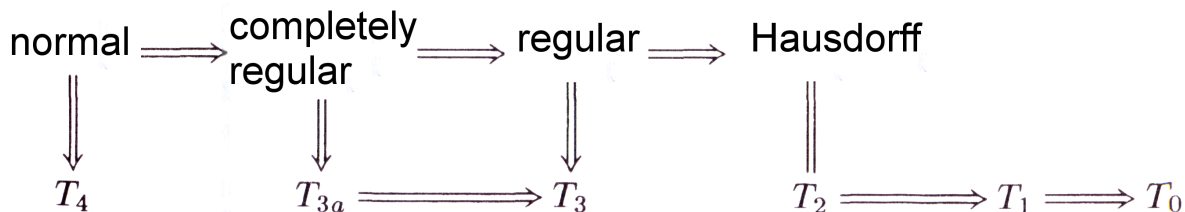
Thus, the  $T_1$ -spaces can be characterised with the fact, that each set consisting of only one element out of  $X$  is closed. This is stated by the following theorem of equivalence (see [Que01], p.84-85):

**Theorem 7.1.3 (Characteristics of a  $T_1$ -space)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space. Then, the following equivalence is true:

- (1)  $X$  is a  $T_1$ -space.
- (2) Each set consisting of only one element out of  $X$  is closed.
- (3) Each subset  $A \subset X$  is the intersection of all its neighbourhoods.

*Proof.* (a)  $\Rightarrow$  (b) : Let  $x \in X$  arbitrarily but consistently. Then for all  $y \neq x$  exists an open neighbourhood  $U_y$ , witch does not contain  $x$ . Thus,  $\{x\} = X \setminus \bigcup \{U_y \mid y \in X \setminus \{x\}\}$  is closed.



**Figure 7.2:** Overview of the relations between the "T"-spaces ([Que01], p.87)

(b)  $\Rightarrow$  (c) : For each  $x \notin A$ ,  $X \setminus \{x\}$  is an open neighbourhood of  $A$  and  $A = \bigcap \{X \setminus \{x\} \mid x \notin A\}$ .

(c)  $\Rightarrow$  (a) : Hence  $\{x\}$  is the intersection of all its neighbourhoods and for each  $y \neq x$  exists a neighbourhood of  $x$ , which does not contain  $y$ .

□

Such topological spaces which can be classified as  $T_1$ -space are named in a certain way, if other characteristics out of definition 7.1.13 are fulfilled as well:

**Definition 7.1.14 (Extended nomenclature of  $T_1$ -spaces)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space. It is named as

- (1) regular, if it is a  $T_1$ - and  $T_3$ -space.
- (2) completely regular, if it is a  $T_1$ - and  $T_{3a}$ -space.
- (3) normal, if it is a  $T_1$ - and  $T_4$ -space.

An overview about the relations of the different defined spaces is given in figure 7.2.



### 7.1.5 Compactness

An important role is played by so-called *compact* spaces.

Compact spaces provide the ability of gathering *global* features from *local* features of the regarded space. Jänich illustrates this in [Jän97], p.24, by the following description, which is analogously noted: If there is a certain feature for an open subset of a compact space, that is handed to unions of open subsets and each element has a neighbourhood providing this feature, then the space has that feature itself.

According to [Que01], p.105, compactness can be defined in the following way:

**Definition 7.1.15 (Compactness)**

*Let  $(X, \mathcal{O})$  be an arbitrary topological space. This space is named **compact**, if each open covering of  $X$  contains a finite subcovering.*

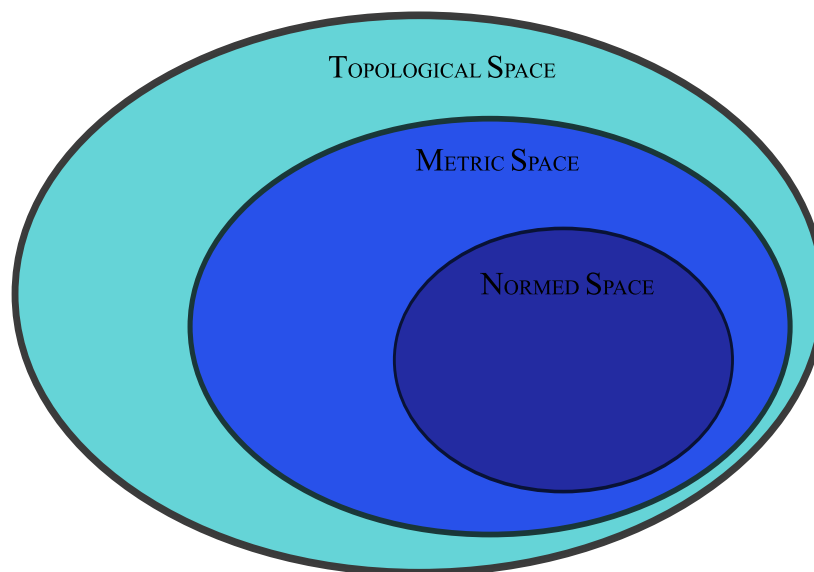
**Theorem 7.1.4 (Tychonoff's theorem)**

*Let  $(X_i)_{i \in I}$  be an arbitrary indexed set of compact topological spaces. Then, the product space  $X := \prod_{i \in I} X_i = \times_{i \in I} X_i$  is compact as well.*

The according proof can be found in [Que01], p.109.

### 7.1.6 Metric spaces

To have the ability of calculating differences between two elements of a certain set, it is necessary to define the term *distance* in a mathematical way. This is done by a so-called *metric* and the used concept is based on the everyday experience with the term *distance*. Hence, this mathematical distance is not negative, it is independent from the order of the regarded elements and using detours leads to a increased distance. But now, there is given an exact mathematical definition of this concept.



**Figure 7.3:** Hierarchical coherence of the regarded mathematical spaces

In [Jän97], p. 10-11, a *metric space* is defined in the following way:

**Definition 7.1.16 (Metric space)**

A **metric space** is a pair  $(X, d)$  consisting of a set  $X \neq \emptyset$  and a real-valued function  $d: X \rightarrow \mathbb{R}$  called **metric** with the following properties:

(1)  $\forall_{x,y \in X} : d(x,y) \geq 0$  and  $\forall_{x,y \in X} : d(x,y) = 0 \Leftrightarrow x = y$ .

(2) *symmetry:*

$$\forall_{x,y \in X} : d(x,y) = d(y,x).$$

(3) *triangle inequation:*

$$\forall_{x,y,z \in X} : d(x,z) \leq d(x,y) + d(y,z).$$

The metric spaces are topological spaces with additional characteristics as already defined. But how can the topology of a metric space be described? To handle this, the concept of a so-called *epsilon ball* or *open ball* is used. According to [Jän97], p.11, the topology of such a space can be defined in the following way:

**Definition 7.1.17 (Topology of a metric space)**

Let  $(X, d)$  be an arbitrary metric space. A subset  $Y \subset X$  is then named **open**, if the following is true:

$$\forall_{x_0 \in Y} \exists_{\epsilon > 0} : B_\epsilon(x_0) \subset Y, \text{ with } B_\epsilon(x_0) := \{x \in X \mid d(x_0, x) \leq \epsilon\}.$$

The set  $\mathcal{O}(d)$  containing all open subsets of  $X$  is then named the **topology of a metric space**.

**7.1.7 Normed space and semi-normed space**

The term *norm* can be defined as following (compare [Fis93], p. 189) regarding an arbitrary vectorspace  $V$  with null vector  $\mathfrak{o}_V$ . If the regarded vector space is clear, the null vector is denoted as  $\mathfrak{o}$ .

**Definition 7.1.18 (Norm)**

Let  $V$  be an arbitrary vectorspace above the field  $\mathbb{K}$ . A mapping

$$\begin{aligned} \|\cdot\| : V &\longrightarrow \mathbb{R}_+^0 \\ v &\longmapsto \|v\| \end{aligned}$$

is called *norm over  $V$* , if  $\forall v, w \in V$  and  $\forall \lambda \in \mathbb{K}$  the following is valid:

*N1 homogeneity:*

$$\|\lambda v\| = |\lambda| \cdot \|v\|.$$

*N2 triangle inequality:*

$$\|v + w\| \leq \|v\| + \|w\|.$$

*N3  $\|v\| = 0 \Rightarrow v = \mathfrak{o}$ .*

The real number  $\|v\|$  is called **norm** of the vector  $v$ .

The pair  $(V, \|\cdot\|)$  is then called **normed vectorspace**.

From  $N1$  can be concluded, that  $\|\mathbf{o}\| = 0$ , because of:

$$\|\mathbf{o}\| = \|0 \cdot \mathbf{o}\| = 0 \cdot \|\mathbf{o}\| = 0$$

By a norm, a topology is induced on the basing space. An example for the space  $\mathbb{R}^n$  is given in definition 9.2.3 in chapter 9.

The hierarchical structure of the spaces discussed until now is illustrated in figure 7.3.

In the case, that the property  $N3$  cannot be fulfilled, this definition can be reduced to that of a so-called *semi-norm* following [Els05], p.133.

**Definition 7.1.19 (Semi-norm)**

Let  $V$  be a vectorspace above the field  $\mathbb{K}$ . A mapping

$$\begin{aligned} \mathcal{S} : V &\longrightarrow \mathbb{R}_+^0 \\ v &\longmapsto \mathcal{S}(v) \end{aligned}$$

is called *semi-norm* over  $V$ , if  $\forall v, w \in V$  and  $\forall \lambda \in \mathbb{K}$  the following is valid:

1. *homogeneity*:

$$\mathcal{S}(\lambda v) = |\lambda| \cdot \mathcal{S}(v).$$

2. *triangle inequality*:

$$\mathcal{S}(v + w) \leq \mathcal{S}(v) + \mathcal{S}(w).$$

The real number  $\mathcal{S}(v)$  is called *semi-norm* of the vector  $v$ .

The pair  $(V, \mathcal{S})$  is then called *semi-normed vectorspace*.

**Remark 7.1.1 (Semi-norm)**

From the first attribute using the characteristics of the basing vector space, there can be derived the following:

$$v = \mathbf{o} \Rightarrow \mathcal{S}(v) = 0.$$

## 7 Measure theory on topological and function spaces

Let  $\mathbf{o}_V$  be the zero vector of the regarded vector space  $V$ , then the following is valid:

$$0 \cdot \mathbf{o}_V = (1 - 1) \cdot \mathbf{o}_V = 1 \cdot \mathbf{o}_V - 1 \cdot \mathbf{o}_V = \mathbf{o}_V - \mathbf{o}_V = \mathbf{o}_V$$

And so using attribute 1:

$$\mathcal{S}(\mathbf{o}_V) = \mathcal{S}(0 \cdot \mathbf{o}_V) = |0| \cdot \mathcal{S}(\mathbf{o}_V) = 0.$$

The concept of “length” described by the term *norm* can be regarded more generalised by using the term *Minkowski functional* defined in [Wer05], p.101:

### Definition 7.1.20 (Minkowski functional, gauge)

Let  $V$  be an arbitrary vectorspace above the field  $\mathbb{K}$ . A mapping

$$\begin{aligned} p_A : V &\longrightarrow \mathbb{R}_+^0 \\ v &\longmapsto p_A(v) := \inf \{ \lambda \in \mathbb{K} \mid \lambda > 0 \wedge v \in \lambda \cdot A \} \end{aligned}$$

is called *Minkowski functional* or *gauge* in  $V$ .

According to [Sch99], p.39, the terms *norm* and *semi-norm* can alternatively be defined using the term *gauge*. The coherence is disclosed in subsection 7.4.2.

### 7.1.8 Locally compact spaces

#### Definition 7.1.21 (Locally compact space)

Let  $(X, \mathcal{O})$  be an arbitrary topological space. This space is called **locally compact**, if it is a  $T_2$ -space (Hausdorff) and if any element  $x \in X$  has a compact neighbourhood.

Kowalsky shows in [Kow61], that the following conclusions can be drawn:

#### Corollary 7.1.1 (Characteristics of locally compact spaces)

Let  $(X, \mathcal{O})$  be an arbitrary locally compact topological space. Then the following is valid:

(1)  $(X, \mathcal{O})$  is a  $T_3$ - and a  $T_{3a}$ -space.

(2) Let  $K \subset U \subset X$ ,  $K$  compact and  $U$  open. Then, there exists a continuous mapping  $f : X \rightarrow [0, 1]$  where

$$f(K) = \left\{ y \in [0, 1] \mid \exists_{x \in K} f(x) = y \right\} = \{1\}, \text{ with the compact set } \overline{\{x \in X \mid f(x) \neq 0\}} \subset U.$$

Characteristic (2) will be important for *radon measures* in section 7.5.

## 7.2 Convergence

### 7.2.1 Generalised convergence on topological spaces

In this thesis, convergences is looked at in generally defined metric spaces and topological vector spaces. Hence, a more general meaning of convergence is necessary, because the description by *sequences* which can be used in metric spaces like  $\mathbb{R}^n$ , is not sufficient enough.

According to [Que01], p.75, the idea of sequences is sufficient in such topological spaces, which feature a countable neighbourhood base, but not for those, which does not fulfill the first axiom of countability in 7.1.6. For general topological spaces, there can be found examples, which show, that inter alia the concept of continuity cannot be described by sequences. One example is given in [Que01], p. 74.

In the historical development, there have been worked out two main ideas of generalizing the idea of sequences, so that there can be given a sufficient description of topological spaces without a countable neighbourhood base. A short presentation of this development is shown in [Que01], p. 331. The one main idea is given by the so-called **filters**, which have been first described by H. Cartan in 1937. This concept is used in many literature about *set topology* like [Que01] or about *measure theory* like [Els05]. The second concept is called **net** or according to the developers E.H. Moore and H.L. Smith **moore-smith sequence**. In this section, there will be shown the coherences of these two concepts.

A complex valued sequence can be defined as a mapping from the set  $\mathbb{N}$  to the set  $\mathbb{C}$  like it is done in [Neu96], p.89.

In general, a sequence is a mapping from the set  $\mathbb{N}$  to an arbitrary set  $X$ . The idea

of a mapping as a left-total and right-unique *relation* (for explanation see below) of the set  $A$  and the set  $B$  can be interpreted in two central ways. The first is on the one side the notation as a *function*:

$$\begin{aligned} f: A &\longrightarrow B \\ a &\mapsto f(a) \end{aligned} .$$

On the other side, the view is more focused on the indexing of the elements of the set  $B$  by the elements of the set  $A$ . In this case, the term *indexed family of elements* is used.  $A$  is then called the *index set* and the following notation is used:

$$(b_i)_{i \in A}, \quad \forall_{i \in A} : b_i \in B.$$

The elements  $b_i$  are named as the *members* of the *indexed family of elements* with *index*  $i \in A$ . Sequences are mostly noted in this way.

Basing on [Jän97], p.23, a converging sequence in a topological space can be defined as follows:

**Definition 7.2.1 (Converging sequence in a topological space)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space.  $(x_n)_{n \in \mathbb{N}}$  is then a sequence in this space. An element  $x \in X$  is named **limit** of this sequence, if for each neighbourhood  $U$  of  $x$  there exists a number  $n_0 \in \mathbb{N}$ , that  $x_n \in U$  for all  $n \geq n_0$ .

Filters can be defined in the following way as it is found in [Que01], p. 77:

**Definition 7.2.2 (Filter)**

Let  $X \neq \emptyset$  be an arbitrary set. A system  $\mathfrak{F}$  of subsets in  $X$  is called **filter** in  $X$ , when the following four characteristics are true:

- (1)  $\emptyset \notin \mathfrak{F}, X \in \mathfrak{F}$
- (2)  $F \in \mathfrak{F} \Rightarrow F \neq \emptyset$



$$(3) F, G \in \mathfrak{F} \Rightarrow F \cap G \in \mathfrak{F}$$

$$(4) F \in \mathfrak{F}, F \subset G \Rightarrow G \in \mathfrak{F}$$

As mentioned above, the convergence theory basing on filters can for example be found in [Que01], p.77 et seqq.

According to [Que01], p.78, convergence of filters in a topological space can be defined in the following way:

**Definition 7.2.3 (Convergence of filters in topological spaces)**

Let  $(X, \mathcal{O})$  be a topological space,  $\mathfrak{F}$  be a filter in this space and  $\mathfrak{U}(x_0)$  be the neighbourhood filter of an element  $x_0$  of  $X$ . This filter converges to  $x_0 \in X$ , when the following is true:

$\mathfrak{F} \supset \mathfrak{U}(x_0)$ . The element  $x_0$  is then called **limit point** of  $\mathfrak{F}$ .

At this point, exemplarily the so-called *Frechét filter* is defined according to [Oss09], p.45.

**Definition 7.2.4 (Frechét filter)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space and  $(x_n)_{n \in \mathbb{N}}$  a sequence of elements in  $X$  with  $\forall_{n \in \mathbb{N}} x_n \in X$ . A so-called **Frechét filter** can be defined by the following set:

$$\mathfrak{F}_F = \left\{ F \subset X \mid \exists_{n_0 \in \mathbb{N}} \forall_{n \geq n_0} : x_n \in F \right\}$$

Hence, such a Frechét filter consists of sets with tail ends of the regarded sequence as elements.

In the following, the generalization of the sequence idea using nets is described.

The index set of a sequence is  $\mathbb{N}$ . This is a *linearly ordered* set. In general, the elements of a set can be compared using the idea of a so-called *relation*. In math-

ematics, a relation  $R$  according to a certain set  $M$  is defined as a subset of the cartesian product of this set and itself:  $R \subseteq M \times M$ . Instead of  $(a, b) \in R$ , one can write  $aRb$  or use the explicit definition of the regarded relation. To characterize the set and the used relation, the following notation is used:  $(M, R)$ .

**Example 7.2.1 (Relation)**

Let the regarded set be  $\mathbb{N}$ . A possible relation is  $<$ . Hence, the pair  $(2, 3)$  is an element of this relation, because  $2 < 3$  is true.

Regarding sequences and nets, especially the *linearly ordered relation* and the *partially ordered relation* have to be regarded and defined as follows.

**Definition 7.2.5 (Linearly ordered relation)**

Let  $M$  be a set and  $R$  be a relation on this set. This relation is called **linearly ordered**, when the following is true:

- (1)  $\forall_{a \in M} : aRa$  (reflexivity)
- (2)  $\forall_{a, b \in M} : aRb \wedge bRa \Leftrightarrow a = b$
- (3)  $\forall_{a, b, c \in M} : aRb \wedge bRa \Leftrightarrow aRc$  (transitivity)
- (4)  $\forall_{a, b \in M} : aRb \vee bRa$

**Definition 7.2.6 (Partially ordered relation)**

Let  $M$  be a set and  $R$  be a relation on this set. This relation is called **partially ordered**, when the following is true:

- (1)  $\forall_{a \in M} : aRa$  (reflexivity)
- (2)  $\forall_{a, b \in M} : aRb \wedge bRa \Rightarrow a = b$  (antisymmetry)
- (3)  $\forall_{a, b, c \in M} : aRb \wedge bRa \Leftrightarrow aRc$  (transitivity)

For the definition of a net, it is necessary to define the term *directed set* using a partially ordered relation:

**Definition 7.2.7 (Directed set)**

Let  $(I, R)$  be a partially ordered set with the relation  $R$ . This set is called a **directed set**, when the following is true:

$$\forall_{a,b \in I} \exists_{c \in I} : aRc \wedge bRc.$$

Alternatively, the following notation is used:  $(I, \triangleleft)$

So, according to [Que01], p.75, a net can be defined in the following way:

**Definition 7.2.8 (Net)**

Let  $X \neq \emptyset$  be an arbitrary set and  $(I, \triangleleft)$  a directed set. Then, a **net** or a **moore-smith sequence** is a mapping  $\mathfrak{N} : I \longrightarrow X$ .

The following notation is used:  $(\mathfrak{N}_i)_{i \in I}$

Since I will regard the generalised *convergence* in arbitrary topological spaces, I have to define the convergence of a net in relation to such a topological space:

**Definition 7.2.9 (Convergence of nets in topological spaces)**

Let  $(X, \mathcal{O})$  be a topological space and  $(\mathfrak{N}_i)_{i \in I}$  be a net. Let in addition to that  $\mathfrak{U}(x_0)$  be the neighbourhood filter of the element  $x_0$  of  $X$ . The net  $(\mathfrak{N}_i)_{i \in I}$  is then converging to  $x_0$  in the given topological space, when the following is true:

$$\forall_{U \in \mathfrak{U}(x_0)} \exists_{i_0 \in I} \forall_{i \geq i_0} : x_i \in U,$$

when  $(I, \triangleleft)$  is the corresponding directed set of the net.

In this context, the following notation is also used:

$$x_i \xrightarrow{i \rightarrow \infty} x_0.$$

**Lemma 7.2.1 (Convergence of nets in topological spaces)**

Let  $(\mathfrak{N}_i)_{i \in I}$  be a converging net in an arbitrary topological space  $(X, \mathcal{O})$ . Let in addition to that  $\mathfrak{U}(x_0)$  be the neighbourhood filter of the element  $x_0$  of  $X$  and

## 7 Measure theory on topological and function spaces

$\mathcal{B}(x_0)$  be the neighbourhood base of  $x_0$ . Then the following equivalence can be written down:

$$\begin{aligned} x_i \xrightarrow{i \rightarrow \infty} x_0 &\Leftrightarrow \bigvee_{U \in \mathfrak{U}(x_0)} \bigexists_{i_0 \in I} \bigvee_{i \geq i_0} : x_i \in U \\ &\Leftrightarrow \bigvee_{B \in \mathcal{B}(x_0)} \bigexists_{i_0 \in I} \bigvee_{i \geq i_0} : x_i \in B \end{aligned}$$

*Proof.* The first equivalence is just the notation defined in definition 7.2.9, so there is nothing to proof there. So I will focus the proof on the second equivalence:

“ $\Rightarrow$ ” : Since the set  $\mathcal{B}(x_0)$  is defined as a subset of the set  $\mathfrak{U}(x_0)$  with certain characteristics given in 7.1.3 and general propositions are passed on to subsets, this implication is true.

“ $\Leftarrow$ ” : Let:  $\bigvee_{B \in \mathcal{B}(x_0)} \bigexists_{i_0 \in I} \bigvee_{i \geq i_0} : x_i \in B$  be.

Let at this point  $B_1 \in \mathcal{B}(x_0)$  be arbitrary. this leads to the following implication:

$$B_1 \in \mathcal{B}(x_0) \Rightarrow \bigexists_{i_1 \in I} \bigvee_{i \geq i_1} : x_i \in B_1.$$

Since by definition  $\bigvee_{U \in \mathfrak{U}(x_0)} \bigexists_{B \in \mathcal{B}(x_0)} B \subset U$  is true, let  $U_1 \in \mathfrak{U}(x_0)$  with  $B_1 \subset U_1$ .

This leads to the validity of the following implication:

$$U_1 \in \mathfrak{U}(x_0) \wedge B_1 \subset U_1 \Rightarrow \bigexists_{i_1 \in I} \bigvee_{i \geq i_1} : x_i \in U_1.$$

Since the set  $U_1$  was arbitrary, the following is valid:

$$\bigvee_{U \in \mathfrak{U}(x_0)} \bigexists_{i_0 \in I} \bigvee_{i \geq i_0} : x_i \in U.$$

□

This lemma shows, that convergence can also be proofed using the according neighbourhood base and it is not necessary to take all neighbourhoods into account.

## 7 Measure theory on topological and function spaces

A very close relationship between topological filters and nets bases on the fact, that converging filters can be used as the index set of a net. A converging filter  $\mathfrak{F}$  provided with the relation *subset*: " $\subset$ " is a directed set. Hence, it can be used as the index set of a topological net.

## 7.3 Completeness

The term *completeness* in a mathematical space is in strong coherence to the convergence in this space. In real and complex analysis this term can be defined using so-called *Cauchy-sequences*. Using the definition and the according theorem (for the proof refer to the given literature) given in [Kön04], pp.52-53, completeness in  $\mathbb{C}$  (and equivalent in  $\mathbb{R}$ ) can be described in the following way:

### Definition 7.3.1 (Cauchy-sequence)

Let  $(x_n)_{n \in \mathbb{N}}$  be an arbitrary sequence in  $\mathbb{C}$ .  $(x_n)_{n \in \mathbb{N}}$  is called **Cauchy-sequence** in  $\mathbb{C} : \Leftrightarrow \forall_{\epsilon > 0} \exists_{n_0 \in \mathbb{N}} \forall_{n, m > n_0} : |x_m - x_n| < \epsilon$

### Theorem 7.3.1 (Cauchy-sequence in $\mathbb{C}$ )

Let  $(x_n)_{n \in \mathbb{N}}$  be an arbitrary sequence in  $\mathbb{C}$ . Then, the following equivalence is valid:

$(x_n)_{n \in \mathbb{N}}$  converges in  $\mathbb{C} \Leftrightarrow (x_n)_{n \in \mathbb{N}}$  is a Cauchy-sequence in  $\mathbb{C}$ .

Using this, the demand of **convergence of every Cauchy-sequence in the regarded space** is postulated as axiom of completeness in  $\mathbb{R}$  and  $\mathbb{C}$  and in general in every archimedean field.

An example for a space that is not complete in this definition can be given regarding  $\mathbb{Q}$ . There, Cauchy-sequences can be defined, whose limit is  $\sqrt{2}$ , which is not an element of  $\mathbb{Q}$ .

Regarding arbitrary metric spaces, completeness can be described in a generalised way using the given metric (basing on [Rud05], p.60):

**Definition 7.3.2 (Completeness in metric spaces)**

Let  $(X, d)$  be an arbitrary metric space. A sequence  $(x_n)_{n \in \mathbb{N}}$  in this space is called

*Cauchy-sequence*  $:\Leftrightarrow \forall_{\epsilon > 0} \exists_{n_0 \in \mathbb{N}} \forall_{n, m > n_0} : d(x_m, x_n) < \epsilon$

If every Cauchy-sequence is converging in a metric space, this space is called **complete**.

Until now, the definition of *completeness* always makes use of the idea of “distance” with respect to a metric. But in arbitrary topological spaces, there is no assurance to have such a metric. Hence, the term *completeness* has to be regarded in a way using general characteristics of topological spaces. In topological spaces that have additionally a compatible vector space structure (see 7.4), is it possible to use algebraic compositions of elements, so *completeness* can there be described as follows (refer to [Jän97], p.35).

**Definition 7.3.3 (Completeness of a topological vector space)**

Let  $(X, \mathcal{O})$  be an arbitrary topological vector space, where  $(V, +, \cdot)$  is the basing vector space above the field  $\mathbb{K}$ .

A sequence  $(x_n)_{n \in \mathbb{N}}$  in this space is then called **Cauchy-sequence**

$:\Leftrightarrow \forall_{U \in \mathfrak{U}(0)} \exists_{n_0 \in \mathbb{N}} \forall_{n, m > n_0} : x_n + (x_m) \in U$

If every Cauchy-sequence converges in this space, it is called **complete**.

An even more generalised way is the use of *filters* for describing *completeness* in topological spaces that have a *uniform structure*.

In such spaces, the definitions given in [Que01], pp.154-155, can be used:

**Definition 7.3.4 (Completeness in an uniform space)**

Let  $(X, \mathcal{U})$  be an arbitrary uniform space.

A filter  $\mathfrak{F}$  in this space is called **Cauchy-filter**

$$:\Leftrightarrow \forall_{V \in \mathcal{U}} \exists_{F \in \mathfrak{F}} : F \times F \subset V.$$

This uniform space  $X$  is called **complete**, if any Cauchy-filter converges in  $X$ .



## 7.4 Topological vector spaces

### 7.4.1 Basic characteristics of topological vector spaces

Basing on the description in [Jän97], p.33 et seqq., the term *topological vector space* is defined.

**Definition 7.4.1 (Topological vector space)**

Let  $(X, +, \cdot)$  be an arbitrary vector space above the field  $\mathbb{K}$ . If  $(X, \mathcal{O})$  also is a topological space and if the topology  $\mathcal{O}$  as well as the linear structure are compatible in terms of continuity of the following mappings *SM* and *AD*,  $X$  is called **topological vector space**.

$$\begin{array}{l} \text{TVS1} \quad SM : \mathbb{K} \times V \longrightarrow V \quad (\text{scalar multiplication}) \\ \quad \quad \quad (\lambda, v) \mapsto \lambda \cdot v \end{array}$$

$$\begin{array}{l} \text{TVS2} \quad AD : V \times V \longrightarrow V \quad (\text{addition}) \\ \quad \quad \quad (v, v) \mapsto v + v \end{array}$$

### 7.4.2 Locally convex spaces

As defined in [Jän97], p.36, a *locally convex space* can be described by the following:

**Definition 7.4.2 (Locally convex topological vector space)**

Let  $(X, +, \cdot)$  be an arbitrary topological vector space above the field  $\mathbb{K}$

$(X, +, \cdot)$  is called **locally convex topological vector space**, if every neighbourhood of zero contains a convex neighbourhood of zero.

Equivalent to this formulation is the demand of a convex basis of neighbourhoods of zero, what obviously follows using definition 7.1.3.

Meise and Vogt showed in [Mei92], p.231, that this is equivalent to the demand of a convex basis of neighbourhoods for any element  $x \in X$ .

What is interesting now is how to construct such locally convex vector spaces.

Schaefer pointed out in [Sch99], p.48, that there exists two ways of constructing such spaces. The first one as a geometric one using a filter basis with certain properties. At this point, it is concentrated on the other way, an analytical one.

This approach is described in [Wer05], pp.389 - 392, using the term *seminorm* defined in definition 7.1.19:

It has to be verified, that using semi-norms, a basis of neighbourhoods of zero of an appropriate topology can be constructed. This is done in the following lemma after giving some definitions basing on [Wer05], p.101 and p.390 as well as additionally [Sch99], p.11.

**Definition 7.4.3 (Characterization of sets in a vector space)**

Let  $(X, +, \cdot)$  be an arbitrary vector space above the field  $\mathbb{K}$ ,  $A$  be an arbitrary subset of  $X$  and  $\lambda$  be an arbitrary element of  $\mathbb{K}$ .

*CI Circled:*  $A$  is called **circled**, if

$$\forall_{\substack{\lambda \in \mathbb{K} \\ |\lambda| \leq 1}} \lambda \cdot A \subset A$$

*CII Convex:*  $A$  is called **convex**, if

$$\forall_{\substack{u, v \in A \\ \lambda \in \mathbb{K} \\ 0 \leq \lambda \leq 1}} \lambda \cdot u + (1 - \lambda) \cdot v \in A$$

*AI Absolutely convex:*  $A$  is called **absolutely convex**, if  $A$  is circled and convex.

This is equivalent to:  $A$  is absolutely convex

$$\Leftrightarrow \left( \forall_{\substack{u, v \in A \\ \lambda, \mu \in \mathbb{K}}} : (|\lambda| + |\mu| \leq 1 \Rightarrow \lambda \cdot u + \mu \cdot v \in A) \right)$$

*AII Absorbing:*  $A$  is called **absorbing**, if

$$\forall_{v \in V} p_A(v) < \infty, \text{ whereby } p_A \text{ is the Minkowski functional defined in definition}$$

7.1.20.

**Lemma 7.4.1 (Topology of a locally convex topological vector space)**

Let  $(X, +, \cdot)$  be an arbitrary vector space above the field  $\mathbb{K}$  and let  $P$  be a set of semi-norms in  $X$ . Let the system  $\mathfrak{U}_0$  of sets be defined by the following:

$$\mathfrak{U}_0 = \{U_{F,\epsilon} \mid F \subset P \text{ is finite, } \epsilon > 0\}$$

$$U_{F,\epsilon} := \left\{ x \in X \mid \forall_{p \in F} p(x - \mathbf{o}) \leq \epsilon \right\}, \epsilon > 0, F \subset P \text{ is finite.}$$

Then, for the system  $\mathfrak{U}_0$ , the following properties are valid:

- (1)  $\forall_{U \in \mathfrak{U}_0} : \mathbf{o} \in U$
- (2)  $\forall_{U_1, U_2 \in \mathfrak{U}_0} \exists_{U \in \mathfrak{U}_0} : U \subset U_1 \cap U_2$
- (3)  $\forall_{U \in \mathfrak{U}_0} \exists_{V \in \mathfrak{U}_0} : V + V := \{v + v \mid v \in V\} \subset U$
- (4)  $\forall_{U \in \mathfrak{U}_0} : U \text{ is absorbing}$
- (5)  $\forall_{\substack{U \in \mathfrak{U}_0 \\ \lambda \in \mathbb{K}}} \exists_{V \in \mathfrak{U}_0} : \lambda \cdot V := \{\lambda \cdot v \mid \lambda \in \mathbb{K}, v \in V\} \subset U$
- (6)  $\forall_{U \in \mathfrak{U}_0} : U \text{ is circled}$
- (7)  $\forall_{U \in \mathfrak{U}_0} : U \text{ is absolutely convex}$
- (8) Using (1) – (6), for  $(X, +, \cdot)$  can be constructed a topology by:

$$O \subset X \text{ is open} \Leftrightarrow \forall_{x \in O} \exists U \in \mathfrak{U}_0 : x + U := \{x + u \mid u \in U\} \subset O,$$

such that  $(X, +, \cdot)$  is a topological vector space.

The proof can be found in [Wer05], p.389 et seqq.

## 7.5 Measurement and integration

### 7.5.1 Basic principles of measures

The term *measuring* is in close coherence to our everyday life. We often measure things characterising them by a length or a volume. The *mathematical measure theory* generalizes this measuring for applying it on certain sets and spaces. It follows the concepts of the everyday use of *measuring*, and so, the methods are characterised as follows:

- (a) A set, that does not contain any element is measured by 0.
- (b) The value, that is produced by using a measurement method is positive (a positive element of  $\mathbb{R}$  or  $\mathbb{C}$ ).
- (c) The measurement of disjoint subset is done by adding up the particular measure values.

In the historical development of *measure theory*, many approaches have been developed. Two main proceedings of constructing a theory of measurement can be found. The first one well described in [Els05] and [Bau01] uses *systems of set* like so-called  $\sigma$ -algebras, measure spaces and measurable functions as basing concepts and leads then to measurability in topological spaces. A team of french authors using the pseudonym Nicolas Bourbaki (see [Bou04]) uses *linear functionals* describing measurement on topological spaces as a starting point. In this subsection, the basic ideas of the measuring in measure spaces are described. After that, the approach using linear functionals is looked at more detailed, because it is used for this thesis regarding the space of geocoded mappings as trainings-data for BPNs in chapter 9 in the context of dealing with spatial data within the EWARS. The coherence between both approaches finally is shown regarding the so-called *Representation theorem of Riesz* in the last subsection of this section.

## 7 Measure theory on topological and function spaces

The definition of measurement basing on sets is described briefly at this point, for further information refer to [Els05] or [Bau01].

This way of measuring uses the mentioned system of sets called  $\sigma$ -algebra ([Els05], p.14):

### Definition 7.5.1 ( $\sigma$ -algebra)

Let  $X$  be an arbitrary set and  $\mathfrak{P}(X)$  the corresponding power set. The system  $\mathfrak{A} \subset \mathfrak{P}(X)$  is called  **$\sigma$ -algebra**, if the following is valid:

- (1)  $X \in \mathfrak{A}$
- (2)  $\forall_{A \in \mathfrak{P}X} : A \in \mathfrak{A} \Rightarrow X \setminus A \in \mathfrak{A}$
- (3) For any sequence  $(A_n)_{n \in \mathbb{N}}$  in  $\mathfrak{P}X$  is valid:  $\forall_{n \in \mathbb{N}} : A_n \in \mathfrak{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$

Basing on this, a *measure* can be defined in the following way:

### Definition 7.5.2 (Measure)

Let  $X$  be a set,  $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$  and  $\mathfrak{A} \subset \mathfrak{P}(X)$  a  $\sigma$ -algebra. Then, a mapping  $\mu : \mathfrak{A} \rightarrow \overline{\mathbb{R}}$  is called **measure** with the following propositions:

- (1)  $\mu(\emptyset) = 0$
- (2)  $\forall_{A \in \mathfrak{A}} \mu(A) \geq 0$
- (3) For any disjoint sequence  $(A_n)_{n \in \mathbb{N}}$  in  $\mathfrak{A}X$  is valid:  $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$

A very important  $\sigma$ -algebra is the  $\sigma$ -algebra of Borel sets ([Els05], p.18):

### Definition 7.5.3 ( $\sigma$ -algebra of Borel sets)

Let  $(X, \mathcal{O})$  be an arbitrary topological space. The  $\sigma$ -algebra  $\sigma(\mathcal{O})$  generated by  $\mathcal{O}$  is then called the  **$\sigma$ -algebra of the Borel sets in  $X$**  and denoted by  $\mathfrak{B}(X)$ .

W.H.Young used this definition of measurement to define integration and in this context,  $\int_X f d\mu$  is called the  $(\mu)$ -Integral of a measurable mapping  $f : X \longrightarrow \overline{\mathbb{R}}$ . It is not necessary to broaden this concept here more deeply. The coherences are well described in [Els05], p. 119 et seqq. and the historical development on p.136 et seqq.

Regarding the concept of measuring described in the following subsection, the representation theorem of F. Riesz formulates the coherence of these two concepts. This theorem is given in a separate subsection (7.5.3) after the following, because there are needed some definitions out of this. In that way, both concepts are described before giving the theorem of Riesz. building a bridge between them.

## 7.5.2 Measurement on topological spaces and elements of integration theory

Now, the focus lays on the measuring approach using linear functionals.

### Definition 7.5.4 (Linear operator and linear functional)

Let  $\mathcal{MAP}(X, Y)$  be the set of any mappings  $F : X \longrightarrow Y$  between the sets  $X$  and  $Y$  as in section 7.2. A so-called **operator** is an element out of this set, thus it is a synonym for the term **mapping**. Hence, a **linear operator** is a **linear mapping**.

In this context, a **linear functional** of a vectorspace  $(V, +, \cdot)$  above the field  $\mathbb{K}$  is an operator of the type  $F : V \longrightarrow \mathbb{K}$  with the following properties:

$$\forall_{\substack{\lambda, \mu \in \mathbb{K} \\ x_1, x_2 \in V}} : F(\lambda \cdot x_1 + \mu \cdot x_2) = \lambda \cdot F(x_1) + \mu \cdot F(x_2)$$

### Definition 7.5.5 (Support)

Let  $(X, \mathcal{O})$  be an arbitrary topological space and  $f \in \mathcal{MAP}(X, \mathbb{C})$  a continuous mapping. Then, the **support** of  $f$  is defined by the following bounded set:

$$\text{Supp}(f) := \overline{\{x \in X \mid f(x) \neq 0_{\mathbb{C}}\}}.$$

**Definition 7.5.6 (Continuous mappings with compact support)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space. The set

$\mathcal{C}_C(X) := \left\{ f \in \mathcal{MAP}(X, \mathbb{C}) \mid f: X \longrightarrow \mathbb{C}, \text{Supp}(f) \text{ is compact} \right\}$  is called the space of all **continuous mappings with compact support** of the type  $f: X \longrightarrow \mathbb{C}$ .

**Definition 7.5.7 (Supremum semi-norm on compact sets in  $\mathcal{C}_C(X)$ )**

Let  $f$  be an arbitrary element of  $\mathcal{C}_C(X)$  and let  $K \in \mathcal{K}(X)$  be an arbitrary compact subset of  $X$ . Furthermore, let  $I_K$  be the characteristic function of  $K$ . Then by

$$\|\cdot\|_K: \mathcal{C}_C(X) \longrightarrow \mathbb{R}_0^+ \text{ with}$$

$$\|f\|_K := \|f \circ I_K\|_{\infty} = \sup\{|(f \circ I_K)(x)| \mid x \in X\},$$

there is a semi-norm defined on  $\mathcal{C}_C(X)$ .

To summarize the regarded function spaces, the following remark is given:

**Remark 7.5.1 (Nomenclature of function spaces)**

In this thesis, the following nomenclature concerning function spaces is used. Let therefor  $X$  be an arbitrary topological space and  $\mathbb{C}$  provided with the euclidian topology:

- The set of **any mappings**  $F: X \longrightarrow Y$  between the sets  $X$  and  $Y$  is denoted as  $\mathcal{MAP}(X, Y)$
- The set of **any continuous mappings**  $F: X \longrightarrow Y$  between the sets  $X$  and  $Y$  is denoted as  $\mathcal{C}(X, Y)$
- The set of **any continuous mappings with compact support**  $F: X \longrightarrow Y$  between the sets  $X$  and  $Y$  is denoted as  $\mathcal{C}_C(X, Y)$
- The set of any **continuous mappings**  $F: X \longrightarrow \mathbb{C}$  between the sets  $X$  and  $\mathbb{C}$  is denoted as  $\mathcal{C}(X)$

- The set of any **continuous mappings with compact support**  $F : X \longrightarrow \mathbb{C}$  between the sets  $X$  and  $\mathbb{C}$  is denoted as  $\mathcal{C}_C(X)$
- A mapping  $f \in \mathcal{C}(X)$  respectively  $f \in \mathcal{C}_C(X)$  is called **positive**, if it is **real-valued** ( $f : X \longrightarrow \mathbb{R}$ ) and if  $\forall_{x \in X} f(x) \geq 0$  is valid. The subset of the **positive** mappings of one of the above sets  $\mathcal{C}(X)$  and  $\mathcal{C}_C(X)$  is denoted as  $\mathcal{C}^+(X)$  respectively  $\mathcal{C}_C^+(X)$ .
- $\mathcal{K}(X) := \{K \subset X \mid K \text{ compact}\}$  is the set of any **compact subset** of  $X$

Using this, the term *positive complex-valued linear functional of positive continuous mappings* is defined:

**Definition 7.5.8 (Positive linear functional)**

Let  $F : \mathcal{C}(X) \longrightarrow \mathbb{C}$  respectively  $F : \mathcal{C}_C(X) \longrightarrow \mathbb{C}$  be a linear functional. This linear functional is called **positive**, if the following is valid:

$$f \in \mathcal{C}^+(X) \Rightarrow F(f) \geq 0 \text{ real-valued}$$

respectively

$$f \in \mathcal{C}_C^+(X) \Rightarrow F(f) \geq 0 \text{ real-valued}$$

Following [Bou04], a complex radon measure is defined in the following way:

**Definition 7.5.9 (Complex radon measure)**

Let  $(X, \mathcal{O})$  be an arbitrary locally compact  $T_2$  topological space. Then a **complex radon measure** on  $X$  is a **linear functional**  $\mu : \mathcal{C}_C(X) \longrightarrow \mathbb{C}$  with the following properties:

$$\forall_{K \in \mathcal{K}(X)} \exists_{C > 0} : f \in \mathcal{C}_C(X), \text{Supp}(f) \subset K \Rightarrow |\mu(f)| \leq C \cdot \|f\|_\infty$$

This idea of measuring is used in this thesis as the main measuring concept.



The following theorem shows with the help of a lemma, that for any positive linear form  $\mu: \mathcal{C}_C(X) \rightarrow \mathbb{C}$  there exists an adequate constant  $C$  to fulfill the postulation of the above definition.

**Lemma 7.5.1 (Positive linear functionals)**

Let  $\mu: \mathcal{C}_C(X) \rightarrow \mathbb{C}$  an arbitrary **positive linear functional** on  $\mathcal{C}_C(X)$ .

Then, the following is valid:

$$\forall_{f \in \mathcal{C}_C(X)} : |\mu(f)| \leq \mu(|f|)$$

*Proof.* Let  $\alpha \in \mathbb{C}$  with  $|\alpha| = 1$  and  $\alpha \cdot \mu(f) = |\mu(f)|$ , where  $f$  is an arbitrary element of  $\mathcal{C}_C(X)$ .

In  $\mathcal{C}_C(X)$ , any real-valued mapping  $g$  can be written as  $g^+ - g^-$ , with  $g^+ := \frac{1}{2} \cdot (|g| + g)$  and  $g^- := \frac{1}{2} \cdot (|g| - g)$  and  $\forall_{x \in X} : |g|(x) := |g(x)|$  easily shown by insertion.

It is valid, that  $g^+$  and  $g^-$  both are elements of  $\mathcal{C}_C^+(X)$ , because let  $g(x_0) = -y \in \mathbb{R}$  for an arbitrary  $x_0 \in X$ , then  $\frac{1}{2} \cdot (|g(x_0)| + g(x_0)) = \frac{1}{2} \cdot (y + (-y)) = 0$  and equivalently shown for  $g^-$ .

The mapping  $\alpha \cdot f$  can be written as  $g_1 + i \cdot g_2$ ,  $g_1, g_2 \in \mathcal{C}_C(X)$  real-valued. So, the following is valid using the linearity of  $\mu$ :

$$(1) \quad \forall_{i \in \{1,2\}} : \mu(g_i) = \mu(g_i^+ - g_i^-) = \mu(g_i^+ + (-1) \cdot g_i^-) = \mu(g_i^+) + (-1) \cdot \mu(g_i^-) \in \mathbb{R}.$$

$$(2) \quad \forall_{x \in X} : g_1(x) \leq |g_1|(x) = |g_1(x)| = \sqrt{g_1(x)^2} \leq \sqrt{g_1(x)^2 + g_2(x)^2} = |g_1(x) + i \cdot g_2(x)| = |(g_1 + i \cdot g_2)(x)| = |\alpha \cdot f(x)| = |\alpha \cdot f|(x).$$

So, using (1) and (2) and in addition the linearity of  $\mu$ , the following chain of inequalities can be written down:

$$|\mu(f)| = \alpha \cdot \mu(f) = \mu(\alpha \cdot f) = \mu(g_1 + i \cdot g_2) = \mu(g_1) + \underbrace{i \cdot \mu(g_2)}_{=0} = \mu(g_1) \stackrel{(2)}{\leq} \mu(\alpha \cdot f) = \mu(|f|)$$

□

**Theorem 7.5.1 (Positive linear functionals as complex radon measures)**

Let  $\mu : \mathcal{C}_C(X) \longrightarrow \mathbb{C}$  an arbitrary **positive linear functional** on  $\mathcal{C}_C(X)$ , then there exists an adequate constant  $C$  to fullfill definition 7.5.9 for this  $\mu$ . Such functionals are called then **positive radon measures**.

*Proof.* Let  $\mu : \mathcal{C}_C(X) \longrightarrow \mathbb{C}$  an arbitrary **positive linear functional** on  $\mathcal{C}_C(X)$  and let  $K \in \mathcal{K}(X)$  be an arbitrary compact subset of  $X$ . Furthermore, let  $f$  be an arbitrary element of  $\mathcal{C}_C(X)$  with  $Supp(f) \subset K$ .

Using corollary 7.1.1, there exists a continuous mapping  $h : X \longrightarrow [0, 1]$  where

$$h(K) = \left\{ y \in [0, 1] \mid \exists_{x \in K} h(x) = y \right\} = \{1\}.$$

Obviously,  $h$  is an element of  $\mathcal{C}_C^+$ .

So,  $|f| \leq \|f\|_\infty \cdot h$  is valid and using this and the above lemma, also the following is valid:

$$|\mu(f)| \leq \mu(|f|) \leq \mu(\|f\|_\infty \cdot h) \leq \|f\|_\infty \cdot \mu(h) = \|f\|_K \cdot \mu(h) \stackrel{C := \mu(h)}{=} \|f\|_K \cdot C.$$

□

**Remark 7.5.2 (Set of positive linear functionals)**

Let  $(V, +, \cdot)$  be an arbitrary vector space above the field  $\mathbb{K}$ . The set of any **positive linear functional** of the type  $F : V \longrightarrow \mathbb{K}$  is then denoted as  $\mathcal{PLF}(V, \mathbb{K})$

Until now, measurement is only described and defined for mappings of the type  $\mu : \mathcal{C}_C(X) \longrightarrow \mathbb{C}$ , but this is not adequate for regarding arbitrary mappings. Hence, the described method of measurement has to be extended to gain the ability of handling a huger set of mappings. This process of enlargement is shown at this point. The main idea is to describe *integrability* of mappings by using

approximations of mappings  $f \in \mathcal{C}_C(X)$ . The starting point is the definition of a so-called *upper integral* or *upper measure*, but before doing this, some more definitions are necessary. The illustration of the coherences is based on [Edw94], p.182 et seqq.

Hence,  $(X, \mathcal{O})$  denotes an arbitrary locally compact  $T_2$  topological space and  $\mathcal{F}$  set is a set of mappings with the following definition:  $\mathcal{F} \subset \mathcal{MAP}(X, [-\infty, \infty])$

In the following, the set  $\overline{\mathbb{R}}$  is defined as  $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ . For this set, the following is defined:

$$\begin{aligned} \forall_{x \in \mathbb{R}} : -\infty < x < \infty \\ \forall_{\substack{x \in \overline{\mathbb{R}} \\ x \neq -\infty}} : x + \infty = \infty + x = \infty \\ \forall_{\substack{x \in \overline{\mathbb{R}} \\ x \neq \infty}} : x + (-\infty) = (-\infty) + x = -\infty \end{aligned}$$

The following definition describes the term *directed* or *filtered set* given in definition 7.2.7 in the context of function sets:

**Definition 7.5.10 (Filtered set of mappings)**

Let  $(X, \mathcal{O})$  be an arbitrary locally compact  $T_2$  topological space. A set  $\mathcal{F} \subset \mathcal{MAP}(X, [-\infty, \infty])$  is called:

- (a) **increasing filtered**, if  $\forall_{f, g \in \mathcal{F}} \exists_{h \in \mathcal{F}} : f \leq h \wedge g \leq h$
- (b) **decreasing filtered**, if  $\forall_{f, g \in \mathcal{F}} \exists_{h \in \mathcal{F}} : h \leq f \wedge h \leq g$

For describing the upper measure, the definition of semicontinuity is necessary:

**Definition 7.5.11 (Semicontinuity of mappings)**

Let  $f \in \mathcal{F}$ . This mapping is called

- (a) **upper semicontinuous**,  $:\Leftrightarrow \forall_{\substack{\alpha \in \mathbb{R} \\ \alpha > f(x)}} \exists_{U \in \mathfrak{U}(x)} : y \in U \Rightarrow \alpha > f(y)$
- (b) **lower semicontinuous**,  $:\Leftrightarrow \forall_{\substack{\alpha \in \mathbb{R} \\ \alpha < f(x)}} \exists_{U \in \mathfrak{U}(x)} : y \in U \Rightarrow \alpha < f(y)$

## 7 Measure theory on topological and function spaces

Using this, the sets  $\mathcal{USC}(X)$  and  $\mathcal{LSC}(X)$  are characterised in the following way:

$$\mathcal{USC}(X) := \{f \in \mathcal{F} \mid f \text{ is upper semicontinuous}\}$$

$$\mathcal{LSC}(X) := \{f \in \mathcal{F} \mid f \text{ is lower semicontinuous}\}$$

$\mathcal{USC}^+(X)$  and  $\mathcal{LSC}^+(X)$  describes the particular subsets of positive mappings (see remark 7.5.1).

Using the set  $\mathcal{LSC}^+(X)$ , there is defined an *upper integral* on this set:

### Definition 7.5.12 (Upper integral on $\mathcal{LSC}^+(X)$ )

Let  $\mu^*$  be an arbitrary **positive radon measure** and  $\mu$  be an arbitrary **radon measure**. Then the **upper integral** of a mapping  $\phi \in \mathcal{LSC}^+(X)$  is defined in the following way:

$$\mu^*(\phi) = \int^* \phi d\mu = \int^* \phi(x) d\mu(x) := \text{Sup} \{ \mu(h) \mid h \in \mathcal{C}_C^+(X), h \leq \phi \}$$

Hence, the mapping  $\mu^* : \mathcal{LSC}^+(X) \longrightarrow [0, \infty]$  is a continuation of the mapping  $\mu : \mathcal{C}_C^+(X) \longrightarrow [0, \infty[$  for arbitrary positive radon measure  $\mu^*$  and radon measure  $\mu$ .

Now, arbitrary positive mappings  $f : X \longrightarrow [0, \infty]$  are regarded to extend the given definition:

### Definition 7.5.13 (Upper integral for mappings $f : X \longrightarrow [0, \infty]$ )

Let  $f \in \mathcal{MAP}(X, [0, \infty])$  an arbitrary mapping on  $X$  taking values in  $[0, \infty] \subset \overline{\mathbb{R}}$ . Then, the **upper integral** of this mapping is defined by the following:

$$\mu^*(f) = \int^* f d\mu = \int^* f(x) d\mu(x) := \text{Inf} \{ \mu^*(\phi) \mid \phi \in \mathcal{LSC}^+(X), \phi \leq f \}$$

With this definition, the following convergence statement can be given:

**Theorem 7.5.2 (Convergence of upper integrals)**

Let  $(f_n)_{n \in \mathbb{N}}$  be a sequence of mappings with  $\forall_{n \in \mathbb{N}} : f : X \longrightarrow [0, \infty]$  and  $f_n \nearrow f$ , what means, that  $\forall_{n \in \mathbb{N}} f_n \leq f_{n+1}$ ,  $\forall_{x \in X} \lim_{n \rightarrow \infty} f_n(x) = f(x)$ . Then for an arbitrary upper integral the following is valid:

$$\mu^*(f_n) \nearrow \mu^*(f)$$

The according proof can be found in [Edw94], p.190.

Regarding sets and their subsets, there can be defined a so-called *characteristic function*  $\chi_A$  of the subset  $A$ .

**Definition 7.5.14 (Characteristic function)**

Let  $\Omega \neq \emptyset$  be an arbitrary nonempty set and  $A \subseteq \Omega$  an arbitrary subset of  $\Omega$ . Then, the following mapping  $\chi_A$  with:

$$\begin{aligned} \chi_A : \Omega &\longrightarrow \{0, 1\} \\ x &\mapsto \chi_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \in \Omega \setminus A \end{cases} \end{aligned}$$

is called **characteristic function** of the subset  $A$ .

Using this mapping, which obviously is of the type  $f : X \longrightarrow [0, \infty]$ , there can be defined the *exterior*  $(\mu)$ -measure of a subset  $A \subset X$ .

**Definition 7.5.15 (Exterior  $(\mu)$ -measure of a subset)**

Let  $A$  be an arbitrary subset of  $X$ . Then, the so-called **exterior**  $(\mu)$ -measure of this subset is defined in the following way:

$$\mu^*(A) := \mu^*(\chi_A).$$

According to this definition, the term **set of  $\mu$ -measure zero** is described by the validity of  $\mu^*(A) = 0$  for the subset  $A$ .

This term is then used to describe the following:

If  $\forall_{x \in X} : P(x)$  is a property, then this **property** is said to hold  $\mu$ -**almost everywhere**, if the set  $\{x \in X \mid P(x) \text{ is false}\}$  is a set of  $\mu$ -measure zero. In the following, the abbreviation  $\mu\mathcal{AE}$  is used instead of  $\mu$ -almost everywhere.

The next step of extending the idea of measuring consists of introducing a seminorm on a subspace of  $\mathcal{MAP}(X, \mathbb{C})$ . Basing on this spaces, the spaces  $\mathcal{L}^P(X, \mu)$  can be constructed. With this, the term *integrability* can be defined finally.

**Theorem 7.5.3**

Let  $f, g : X \rightarrow [0, \infty]$  be arbitrary mappings. Then, the following is valid:

1.  $f \leq g \mu\mathcal{AE} \Rightarrow \mu^*(f) \leq \mu^*(g)$ .
2.  $f = g \mu\mathcal{AE} \Rightarrow \mu^*(f) = \mu^*(g)$ .
3.  $\mu^*(f) = 0 \Leftrightarrow f = 0 \mu\mathcal{AE}$ .
4.  $\mu^*(f) < \infty \Rightarrow f$  is finite  $\mu\mathcal{AE}$ , i.e.  $\neq \infty$ .

*Proof.* 1. There exists a set of  $\mu$ -measure zero  $A \subset X$  with  $f \leq g + \infty\chi_A$ .

Hence,

$$\begin{aligned}
 \mu^*(f) &\leq \mu^*(g + \infty\chi_A) \\
 &\leq \mu^*(g) + \mu^*(\infty\chi_A) \\
 &= \mu^*(g) + \lim_n \mu^*(n\chi_A) \\
 &= \mu^*(g) + \lim_n n\mu^*(A) \\
 &= \mu^*(g)
 \end{aligned}$$

2. There exists a set of  $\mu$ -measure zero  $A \subset X$  with  $g = f + \infty\chi_A$ . Hence,

$$\begin{aligned}
 \mu^*(g) &\leq \mu^*(f + \infty\chi_A) \\
 &\leq \mu^*(f) + \mu^*(\infty\chi_A) \\
 &= \mu^*(f) + \lim_n \mu^*(n\chi_A) \\
 &= \mu^*(f) + \lim_n n\mu^*(A) \\
 &= \mu^*(f)
 \end{aligned}$$

Using 1., the proposition is valid.

3. Let  $A = \{x \in X \mid f(x) > 0\}$ . Then,  $\chi_A \leq \infty f$ , what leads to the fact, that

$$\begin{aligned}
 \mu^*(A) &\leq \mu^*(\infty f) \\
 &= \mu^*(\lim_n n f) \\
 &= \lim_n \mu^*(n f) \\
 &= 0
 \end{aligned}$$

4. Let  $B = \{x \in X \mid f(x) = \infty\}$ . Then,  $\infty\chi_B \leq f$  is valid. Hence,  
 $\lim_n n\mu^*(B) \leq \mu^*(f) < \infty, \mu^*(B) = 0.$

□

The semi-norm mentioned above is defined in the following way:

**Definition 7.5.16 ( $P$  semi-norm)**

Let  $P$  be an arbitrary element of the interval  $[1, \infty]$  and let  $f$  be an arbitrary mapping out of  $\mathcal{MAP}(X, \mathbb{C})$ . Then the mapping:

$$\begin{aligned}
 \|\cdot\|_P : \mathcal{MAP}(X, \mathbb{C}) &\longrightarrow [0, \infty] \\
 f &\mapsto \|f\|_P := (\mu^*|f|^P)^{\frac{1}{P}} = \sqrt[P]{\int^* |f|^P d\mu}
 \end{aligned}$$

defines the  $P$  semi-norm.

Using this definitions, the following conclusions can be drawn (the proof can be found in [Edw94], p.215 et seq.):

**Corollary 7.5.1 (Characteristics of the  $P$  semi-norm)**

- (1)  $\forall_{f \in \mathcal{MAP}(X, \mathbb{C})} (\|f\|_P \geq 0, \|f\|_P = 0 \Leftrightarrow f = 0 \mu\mathcal{AE})$
- (2)  $\forall_{f, g \in \mathcal{MAP}(X, \mathbb{C})} (f = g \mu\mathcal{AE} \Rightarrow \|f\|_P = \|g\|_P)$
- (3)  $\forall_{\substack{f \in \mathcal{MAP}(X, \mathbb{C}) \\ \alpha \in \mathbb{C}}} (\|\alpha \cdot f\|_P = |\alpha| \cdot \|f\|_P)$
- (4)  $\forall_{f, g \in \mathcal{MAP}(X, \mathbb{C})} (\|f + g\|_P \leq \|f\|_P + \|g\|_P)$  (*Minkowski's inequality*)

At this point, there has to be mentioned, in which space the mapping  $\|\cdot\|_P$  defines a semi-norm. This is valid for the vectorspace  $\mathcal{F}^P(X, \mu)$  above the field  $\mathbb{C}$  with:  
 $\mathcal{F}^P(X, \mu) := \{f \in \mathcal{MAP}(X, \mathbb{C}) \mid \|f\|_P < \infty\}$ .

This space contains the space  $\mathcal{C}_C(X)$  as a vector subspace. The closure of this subspace relative to the defined semi-norm now defines the vector subspace  $\mathcal{L}^P(X, \mu)$ .

This subspace in turn can be characterised in the following way:

$$\mathcal{L}^P(X, \mu) = \left\{ f \in \mathcal{MAP}(X, \mathbb{C}) \mid \forall_{\epsilon > 0} \exists_{g \in \mathcal{C}_C(X)} : \|f - g\|_P < \epsilon \right\}.$$

**Definition 7.5.17 ( $P^{\text{th}}$ -power integrable)**

Let  $f \in \mathcal{L}^P(X, \mu)$  with:

$$\mathcal{L}^P(X, \mu) = \left\{ f \in \mathcal{MAP}(X, \mathbb{C}) \mid \forall_{\epsilon > 0} \exists_{g \in \mathcal{C}_C(X)} : \|f - g\|_P < \epsilon \right\}.$$

Then,  $f$  is called  $P^{\text{th}}$ -power integrable. For  $P = 1$ , the term *integrable* is used.

**Remark 7.5.3**

(R I) The term *integrability* should also be used for mappings

$f : X \rightarrow [-\infty, \infty]$ , if the following proposition is valid

$$\forall \epsilon > 0 \exists g \in \mathcal{C}_C(X) : \|f - g\|_1 < \epsilon.$$



(R II) Let  $f, g$  be two mappings with  $f, g: X \rightarrow \mathbb{C}$  respectively

$f, g: X \rightarrow [-\infty, \infty]$  matching  $\mu\mathcal{AE}$ , so  $g \in \mathcal{L}^P(X, \mu)$  is valid, if  $f \in \mathcal{L}^P(X, \mu)$ .

That means, that, if  $f$  is integrable, this is valid for  $g$ .

(R III) An arbitrary mapping  $f: X \rightarrow \mathbb{C}$  is an element out of  $\mathcal{L}^P(X, \mu)$ , if and only if  $\text{Rea}(f)$  and  $\text{Ima}(f)$  are elements out of  $\mathcal{L}^P(X, \mu)$ .

(R IV) Due to the validity of  $\forall_{g \in \mathcal{C}_C(X)} : \| |f| - |g| \|_p \leq \| f - g \|_p$ , the following characteristics are valid:

- $f \in \mathcal{L}^P(X, \mu) \Rightarrow |f| \in \mathcal{L}^P(X, \mu)$
- $f$  integrable  $\Rightarrow |f|$  integrable,

(R V) Let  $K \subseteq X$  be a compact subset. Then, for any real-valued mapping  $f: K \rightarrow$  the following is valid:

$$f \in \mathcal{L}^P(X, \mu) \Leftrightarrow f^+, f^- \in \mathcal{L}^P(X, \mu).$$

(R VI) With  $f, g \in \mathcal{L}^P$  real-valued, also  $\max(f, g)$  und  $\min(f, g)$  are elements out of  $\mathcal{L}^P(X, \mu)$ , because they can be characterised in the following way:

$$\begin{aligned} \max(f, g) &= (f - g)^+ + g, \\ \min(f, g) &= -(f - g)^- + g. \end{aligned}$$

(R VII) Due to the fact, that an integrable mapping  $f: X \rightarrow [-\infty, \infty]$  is finite  $\mu\mathcal{AE}$  (see 7.5.3), this is also valid for them.

### Lemma 7.5.2

A mapping  $f: X \rightarrow [0, \infty]$  is integrable, if and only if

$$\forall_{\epsilon > 0} \exists_{g, h \in \mathcal{L}^{SC^+}(X)} : -g \leq f \leq h \wedge \mu^*(h) < \infty \wedge \mu^*(h + g) < \epsilon.$$

*Proof.*  $\Rightarrow$ : Let  $g \in \mathcal{C}_C^+(X)$  with  $\mu^*(|f - g|) < \varepsilon$ .

Take  $h \in \mathcal{LSC}^+(X)$  with  $|f - g| \leq h, \mu^*(h) < \varepsilon$ . then,  $h + g \in \mathcal{LSC}^+(X)$  and  $h - g \in \mathcal{LSC}^+(X)$ . Hence,  $-(h - g) \leq f \leq h + g$  and  $\mu^*(h + g) = \mu^*(h) + \mu^*(g) < \infty$ ,  $\mu^*(h + g + h - g) = 2\mu^*(h) < 2\varepsilon$ . is valid

$\Leftarrow$ : Let  $\varepsilon > 0$ . Take  $g, h \in \mathcal{LSC}^+(X)$  with  $-g \leq f \leq h, \mu^*(h) < \infty$  and  $\mu^*(g + h) < \varepsilon$ .

Take then  $k \in \mathcal{C}_C^+(X)$  with  $k \leq h, \mu(k) \geq \mu^*(h) - \varepsilon$ .

Then,  $f - k \leq h - k$ ,

$k - f \leq k + g \leq h + g$ , is valid, i.e.  $|f - k| \leq (h - k) + (h + g)$ ,

$\mu^*(|f - k|) \leq \mu^*(h - k) + \mu^*(h + g) < 2\varepsilon$  using  $\mu^*(h - k) + \mu^*(k) = \mu^*(h)$ .

□

### Corollary 7.5.2

Let  $f \in \mathcal{LSC}^+(X)$ . Then,  $f$  is integrable  $\Leftrightarrow \mu^*(f) < \infty$  is valid.

The according proof can be found in [Edw94], p.189.

### Remark 7.5.4

Let  $f$  be an integrable mapping on  $X$ . Regard a sequence  $(g_n)$  in  $\mathcal{C}_C(X)$  with  $\lim \|f - g_n\|_1 = 0$ .

Then,  $(g_n)$  is a cauchy sequence in  $L^1(X)$  according to the halfnorm  $\|\cdot\|_1$ , i.e.

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} \forall n, m \geq n_0 : \|g_n - g_m\|_1 < \varepsilon.$$

With  $\forall_{g \in \mathcal{C}_C(X)} |\mu(g)| \leq \mu(|g|) = \mu^*(|g|) = \|g\|_1$  also  $(\mu(g_n))$  is a cauchy sequence in  $\mathbb{C}$ . I.e.,  $\lim \mu(g_n)$  exists.

Replacement of  $(g_n)$  by  $(h_n)$  in  $\mathcal{C}_C(X)$  with the same characteristics provides, that these characteristics are also valid for the sequence  $g_1, h_1, g_2, h_2, \dots$ . Hence,  $\lim \mu(g_n)$  and  $\lim \mu(h_n)$  are both equal to the limit of the sequence  $\mu(g_1), \mu(h_1), \mu(g_2), \dots$

**Definition 7.5.18 (Integral of integrable mappings)**

Let  $f \in \mathcal{L}^1(X, \mu)$  be an arbitrary integrable mapping. Further, let  $g_n$  be a sequence in  $\mathcal{C}_C(X)$  with  $\lim \|f - g_n\|_1 = 0$ .

Then, the limit  $\lim \mu(g_n)$  without reference to the choice of  $(g_n)$  is called the **integral** of the mapping  $f$  with the following nomenclature:

$$\mu(f) = \int f d\mu = \int_X f(x) d\mu(x).$$

**Remark 7.5.5**

It is easily seen, that

$$\begin{aligned} \mu : \mathcal{L}^1(X, \mu) &\longrightarrow \mathbb{C} \\ f &\longmapsto \int f d\mu \end{aligned}$$

defines a linear mapping, which extends the linear mapping  $\mu : \mathcal{C}_C(X) \rightarrow \mathbb{C}$ .

The following is valid:

- $\forall_{f \in \mathcal{L}^1(X, \mu)} |\mu(f)| \leq \mu(|f|)$ .  
Regarding  $\lim \|f - g_n\|_1 = 0$  the validity of  $\lim \| |f| - |g_n| \|_1 = 0$  is given for any sequence  $(g_n)$  in  $\mathcal{C}_C(X)$ , i.e.  $|\mu(f)| = |\lim_n \mu(g_n)| = \lim_n |\mu(g_n)| \leq \lim_n \mu(|g_n|) = \mu(|f|)$ .
- If  $f \in \mathcal{L}^1(X, \mu)$  and  $f \geq 0$ , so  $\mu(f) = \mu^*(f)$ .  
This is valid, because if  $g_n \in \mathcal{C}_C^+(X)$  with  $\|g_n - f\|_1 = \mu^*(|g_n - f|) \rightarrow 0$ , so  $|\mu^*(g_n) - \mu^*(f)| \rightarrow 0$ . Hence  $\mu(f) = \lim_n \mu(g_n) = \lim_n \mu^*(g_n) = \mu^*(f)$

**Theorem 7.5.4 (Theorem of the monotonous convergence (B. Levi))**

Let  $f_n : X \rightarrow [-\infty, \infty] \in \mathcal{L}^1(X, \mu)$  and  $n \in, f_n \nearrow f \mu\mathcal{A}\mathcal{E}$  with a suitable  $f : X \rightarrow [-\infty, \infty]$ . The following is proposed:

$$f \in \mathcal{L}^1(X, \mu) \Leftrightarrow \sup \mu(f_n) < \infty,$$

If that is valid, the following is also valid:

$$\mu(f) = \lim_n \mu(f_n) = \sup \mu(f_n)$$

*Proof.* Modification of the mapping onto a set of  $\mu$ -measure zero leads to the fact, that  $f_n \nearrow f$  everywhere and that every  $f_n$  is finite.

Then,  $f_n + f_1^- \geq f_1 + f_1^- = f_1^+ \geq 0$  and  $f_n + f_1^- \nearrow f + f_1^-$ , are valid.

$\Rightarrow$ : and the equality  $\mu(f) = \lim_n \mu(f_n)$  can be derived from 7.5.2 by:

$$\mu(f) + \mu(f_1^-) = \mu(f + f_1^-) = \mu^*(f + f_1^-) = \lim \mu^*(f_n + f_1^-) = \lim_n \mu(f_n) + \mu(f_1^-).$$

$\Leftarrow$ : For  $\varepsilon > 0$  choose  $n$  with  $m \geq n \Rightarrow \mu(f_m) - \mu(f_n) < \varepsilon$ , i.e.  $\mu^*(f_m - f_n) < \varepsilon$ .

With  $f_m - f_n \nearrow f - f_n$  ( $m \rightarrow \infty$ ),  $\mu^*(f - f_n) \leq \varepsilon$ ,  $\|f - f_n\|_1 \leq \varepsilon$  is valid.

Hence,  $\|f - g\|_1 \leq \|f - f_n\|_1 + \|f_n - g\|_1 < 2\varepsilon$  with appropriate  $g \in \mathcal{C}_C(X)$ .

□

### Corollary 7.5.3

Let  $f_n : X \rightarrow [-\infty, \infty]$  be integrable,  $n \in \mathbb{N}$ . If a mapping  $g : X \rightarrow [0, \infty]$  with  $\mu^*(g) < \infty$  and  $f_n \leq g$   $\mu\mathcal{A}\mathcal{E}$  exists for any  $n \in \mathbb{N}$ , then  $\sup_n f_n$  is integrable.

*Proof.* The mappings  $g_n := \max(f_1, \dots, f_n)$  are integrable,  $g_n \nearrow \sup f_n$ , and  $\mu(g_n) \leq \mu^*(g) < \infty$ . is valid.

□

### Lemma 7.5.3 (Fatou)

Let  $f_n, g$  as described in 7.5.3. If  $c := \limsup_{n \rightarrow \infty} \mu(f_n) > -\infty$ , then  $f := \limsup_{n \rightarrow \infty} f_n$  is integrable and  $c \leq \mu(f)$ . is valid.

*Proof.* Using 7.5.3,  $g_n := \sup_{k \geq n} f_k$  is integrable and

$$\begin{aligned}
 \mu(g_n) &\geq \sup_{k \geq n} \mu(f_k) \\
 &\geq \inf_n \sup_{k \geq n} \mu(f_k) \\
 &= \limsup_{n \rightarrow \infty} \mu(f_n) \\
 &= c
 \end{aligned}$$

This is valid using  $g_n \searrow f$ .

I.e.  $-g_n \nearrow -f$  is derived from 7.5.4. Hence,  $-f$  is integrable. This leads to the fact, that  $f$  is integrable as well and  $\mu(f) = -\lim_n \mu(-g_n)$   
 $= \lim_n \mu(g_n)$   
 $\geq c$ .

□

**Theorem 7.5.5 (Lebesgue)**

Let  $(f_n)$  be a sequence of integrable mappings on  $X$  (with values in  $\mathbb{C}$  or in  $[-\infty, \infty]$ ), converging to  $f \mu\mathcal{A}\mathcal{E}$ .

Let  $g : X \rightarrow [0, \infty]$  be a mapping with  $\mu^*(g) < \infty$  and  $|f_n| \leq g \mu\mathcal{A}\mathcal{E}$  and for any  $n \in \mathbb{N}$ . Then,  $f$  is integrable and  $\mu(f) = \lim_n \mu(f_n)$  and  $\|f - f_n\|_1 \rightarrow 0$  are valid.

*Proof.* For proving the integrability of  $f$  and the validity of  $\mu(f) = \lim_n \mu(f_n)$ , any  $f_n$  can be regarded as real valued.

Then,  $-g \leq f_n \leq g$ , and so  $\limsup \mu(f_n) \geq -\mu^*(g) > -\infty$  are valid. sing 7.5.3,  $f$  is integrable with  $f = \limsup f_n \mu\mathcal{A}\mathcal{E}$  and  $\mu(f) \geq \limsup \mu(f_n)$  is valid.

Changing all the signs, also  $\mu(f) \leq \liminf \mu(f_n)$ , is valid.

Hence,  $\mu(f) = \lim \mu(f_n)$  is valid.

Further  $|f - f_n| \mu\mathcal{A}\mathcal{E}$  converges to 0 and  $\leq |f| + |f_n| \leq 2g$

It follows  $\|f - f_n\|_1 = \mu(|f - f_n|) \rightarrow \mu(0) = 0$

□

**Theorem 7.5.6**

Let  $\lambda, \mu$  two positive measures on  $X$ ,  $f : X \rightarrow [0, \infty]$  an arbitrary mapping and further  $A \subset X$ .

Then, the following is valid:

1.  $\lambda^x(f) + \mu^*(f) = (\lambda + \mu)^x(f)$ .
2. A set of  $\lambda + \mu$ -measure zero  $\Leftrightarrow$  A set of  $\lambda$ -measure zero and A set of  $\mu$ -measure zero.
3.  $\mathcal{L}^1(X, \lambda + \mu) = \mathcal{L}^1(X, \lambda) \cap \mathcal{L}^1(X, \mu)$ .

*Proof.* 1. For  $f \in \mathcal{C}_C^+(X)$ , the statement obviously is valid.

For  $f \in \mathcal{LSC}^+(X)$  and any  $g \in \mathcal{C}_C^+$  with  $g \leq f$  the following is valid:

$$\begin{aligned} (\lambda + \mu)(g) &= \lambda(g) + \mu(g) \\ &\leq \lambda^x(f) + \mu^*(f), \text{, hence } (\lambda + \mu)^x(f) \leq \lambda^x(f) + \mu^*(f) \end{aligned}$$

For  $g, h \in \mathcal{C}_C^+$  with  $g \leq f, h \leq f$ , the following is valid:  $\lambda(g) + \mu(h) \leq \lambda \max(g, h) + \mu(\max(g, h)) \leq (\lambda + \mu)^x(f)$ ,. Due to the variability of  $g$  and  $h$ ,  $\lambda(g) + \mu^*(f) \leq (\lambda + \mu)^x(f)$ ,  
 $\lambda^x(f) + \mu^*(f) \leq (\lambda + \mu)^x(f)$  follows.

In the same way, the step from  $f \in \mathcal{LSC}^+(X)$  to an arbitrary  $f$  follows.

2. A set of  $\lambda + \mu$ -measure zero  $\Leftrightarrow (\lambda + \mu)^*(A) = 0 = 0 + 0 = \lambda^*(A) + \mu^*(A) \Leftrightarrow$   
A set of  $\lambda$ -measure zero and A set of  $\mu$ -measure zero.
3. Using  $\lambda \leq \lambda + \mu, \mu \leq \lambda + \mu$ , “ $\subset$ ” is valid with the already proven facts.

“ $\supset$ ”: Let  $f \in \mathcal{L}^1(X, \lambda) \cap \mathcal{L}^1(X, \mu)$ .  $f \geq 0$  can be assumed. With 7.5.2 for any  $\epsilon > 0$  lower semicontinuous mappings  $g_i, h_i : X \rightarrow (-\infty, \infty]$  exist with  $-g_i \leq f \leq h_i, i = 1, 2$ , so that  $\lambda^x(h_1) < \infty$ ,

## 7 Measure theory on topological and function spaces

$$\mu^*(h_2) < \infty, \lambda^x(g_1 + h_1) < \epsilon \text{ and } \mu^*(g_2 + h_2) < \epsilon.$$

Setting  $g = \min(g_1, g_2), h = \min(h_1, h_2)$ , then  $-g \leq f \leq h$ ,  $(\lambda + \mu)^*(h) = \lambda^*(h) + \mu^*(h) \leq \lambda^*(h_1) + \mu^*(h_2) < \infty$  and  $(\lambda + \mu)^*(g + h) = \lambda^*(g + h) + \mu^*(g + h) \leq \lambda^*(g_1 + h_1) + \mu^*(g_2 + h_2) < 2\epsilon$

Using 7.5.2  $f \in L^1(X, \lambda + \mu)$  is valid.

□

### Remark 7.5.6

*It can be proven, that 7.5.6 (3.) is valid for any  $P \in [1, \infty)$ :*

$$\mathcal{L}^P(X, \lambda + \mu) = \mathcal{L}^P(X, \lambda) \cap \mathcal{L}^P(X, \mu).$$

*Using the already proven facts,  $A \subset X$  is  $(\lambda + \mu)$ -integrable, if and only if  $A$  is  $\lambda$ -integrable and  $\mu$ -integrable. The same is valid for the term measurable instead of integrable.*

*Is  $f \in \mathcal{L}^P(\lambda) \cap \mathcal{L}^P(\mu)$ , so  $f$  is  $(\lambda + \mu)$ -measurable*

$$(\lambda + \mu)^x(|f|^p) = \lambda^x(|f|^p) + \mu^*(|f|^p) < \infty \text{ leads then to } f \in \mathcal{L}^P(\lambda + \mu).$$

At this point, the restriction of a positive measure is suspended.

### Definition 7.5.19 ( $\mu$ -integrability)

*Let  $\mu \in M(X)$  be an arbitrary measure. Then, the following is defined:*

$$\mathcal{L}^P(X, \mu) := \mathcal{L}^P(X, |\mu|), p \in [1, \infty),$$

*A mapping  $f$  on  $X$  is named  $\mu$ -integrable, if it is  $|\mu|$ -integrable, the sets of  $\mu$ -measure zero are the sets of  $|\mu|$ -measure zero.*

$$\|f\|_{p,\mu} := (|\mu|^x |f|^p)^{1/p},$$

Now, the spaces  $L^P(X, \mu)$  are defined, but first, some definitions are required.

**Definition 7.5.20 (Banachspace)**

A space  $E$  is called **Banachspace**, if  $E$  is complete and normed by a norm  $\|\cdot\|$ .

**Definition 7.5.21 (Positiv semidefinite Hermitian form, inner product)**

A **positiv semidefinite Hermitian form** or an **inner product** on a complex vector space  $E$  is a mapping

$$\begin{aligned} (\cdot|\cdot) : E \times E &\longrightarrow \mathbb{C} \\ (u, v) &\longmapsto (u|v) \end{aligned}$$

Thereby, the mapping  $(\cdot|\cdot)$  has the following characteristics ( $u, v, w \in E, \alpha \in \mathbb{C}$ ):

(IP1)  $(u + v|w) = (u|w) + (v|w)$ ,

(IP2)  $(\alpha u|v) = \alpha(u|v)$ ,

(IP3)  $(u|v) = \overline{(v|u)}$ ,  $\overline{\cdot}$  means complex conjugation

(IP4)  $(u|u) \geq 0$ ,

With 1. 2. and 3. follows:

(IP5)  $(u|v + w) = (u|v) + (u|w)$ ,

(IP6)  $(u|\alpha v) = \bar{\alpha}(u|v)$

(IP7) If  $(u|u) = 0 \Rightarrow u = 0$  is valid,  $(\cdot|\cdot)$  is named **positive definite**.

**Remark 7.5.7**

Set  $\|u\| := \sqrt{(u|u)}$  for any  $u \in E$ .

Then,  $|(u|v)| \leq \|u\|\|v\|$ ,  $u, v \in E$  is valid. This can be derived from  $(u + \alpha v|u + \alpha v) \geq 0$  for every real valued multiple of  $(v|u)$ .

Thus estimate leads to the fact, that  $\|\cdot\|$  is a semi-norm on  $E$ .



Is  $(\cdot|\cdot)$  even positive definite, then  $\|\cdot\|$  is a norm.

From the semi-norm  $\|\cdot\|$ , the inner product can be reversely derived using the so-called **polarization identity** with  $u, v \in E$ :

$$4(u|v) = \sum_{\nu=0}^3 i^\nu \|u + i^\nu v\|^2$$

**Definition 7.5.22 (Hilbert space)**

A complex vector space  $E$ , which is a Banach space using the inner product  $(\cdot|\cdot)$  and the according derived semi-norm  $\|\cdot\|$ , is named **complex Hilbert space**.

**Remark 7.5.8**

If  $\|\cdot\|$  is a semi-norm on the vectorspace  $E$ , then  $N := \{u \in E \mid \|u\| = 0\}$  defines a vector subspace of  $E$ , on whose coset classes  $u + N, u \in E$  this semi-norm is constant.

This can be seen by  $v \in N \Rightarrow \|u + v\| \leq \|u\| + \|v\| = \|u\| \leq \|u + v\|$ .

On the vector space  $E/N$ , a semi-norm can be defined representively by:

$$\|u/N\| = \|u + N\| := \|u\|, u \in E$$

Thereby,  $u/N$  describes the class generated by  $u$ . Using this,  $E/N$  even is a normed space. Also  $E/N$  is complete.

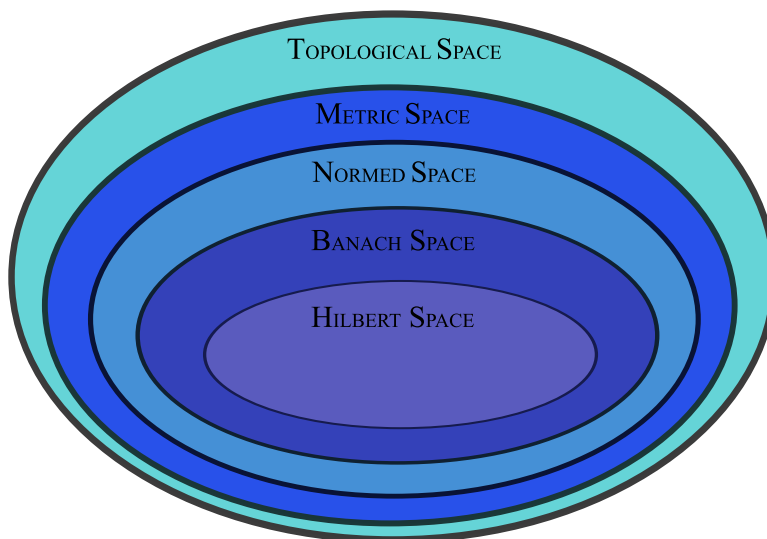
Similar coherences can be derived for an inner product  $(\cdot|\cdot)$  on  $E$ :

Using the semi-norm generated by  $(\cdot|\cdot)$ ,  $N$  is defined as done above.

Adding an arbitrary  $w \in N$  to arbitrary  $u, v \in E$ ,  $(u|v)$  remains constant (Regard  $|(u + v|v) - (u|v)| = |(w|v)| \leq \|w\|\|v\| = 0$ ).

In this way, an inner product can be defined representively on  $E/N$  by:

$$(u/N|v/N) := (u|v), u, v \in E,$$



**Figure 7.4:** Extended hierarchical structure of mathematical spaces

At this point, the hierarchical structure of spaces, that has been described until now, is illustrated in figure 7.4.

These considerations are now applied on the spaces  $\mathcal{L}^P(X, \mu)$ ,  $1 \leq p < \infty$ .

**Definition 7.5.23 ( $L^P(X, \mu)$  space)**

Let  $P$  be an arbitrary element of the interval  $(0, \infty)$  and  $\|\cdot\|_P$  the mapping defined above. Then, the **quotient space**  $L^P(X, \mu)$  is characterised by the following:

$$L^P(X, \mu) := \mathcal{L}^P(X, \mu) / \mathcal{N}^P,$$

with  $\mathcal{N}^P := \{f \in \mathcal{L}^P(X, \mu) \mid f = 0 \mu\mathcal{A}\mathcal{E}\}$ .

**Theorem 7.5.7 (Fischer-Riesz)**

For any  $p \in [1, \infty)$ ,  $\mathcal{L}^P(X, \mu)$  is complete. Hence,  $\mathcal{L}^P(X, \mu)$  is a Banach space.

*Proof.* Let  $(f_n)$  be a Cauchy-sequence in  $\mathcal{L}^P$ . It is appropriate to prove, that a partial sequence  $(f_n)$  converges with reference to the semi-norm  $\|\cdot\|_p$ . Hence,  $\|f_{n+1} - f_n\|_p < 2^{-n}$  for any  $n \in \mathbb{N}$  can be assumed.

$$\text{Set } g := \sum_{n=1}^{\infty} |f_{n+1} - f_n|$$

## 7 Measure theory on topological and function spaces

Regarding the partial sums of this sum to the power of  $P$ , they monotonously converge to  $g^p$ , the following is valid:

$$\begin{aligned}
 \mu^*(g^p) &= \lim_{n \rightarrow \infty} \mu^*\left(\sum_{k=1}^n |f_{k+1} - f_k|\right)^p \\
 &= \lim_{n \rightarrow \infty} \left\| \sum_{k=1}^n |f_{k+1} - f_k| \right\|_p^p \\
 &\leq \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \|f_{k+1} - f_k\|_p\right)^p \\
 &\leq 1.
 \end{aligned}$$

Hence,  $g$  is finite  $\mu\mathcal{AE}$ , the sequence  $(f_n)$  converges to a mapping  $f : X \rightarrow \mathbb{C}$   $\mu\mathcal{AE}$ .

For any  $m \in \mathbb{N}$ ,  $|f_n - f_m|^p \rightarrow |f - f_m|^p$  converges for  $n \rightarrow \infty$   $\mu\mathcal{AE}$ .

For  $n > m$ , the following is valid:

$$\begin{aligned}
 |f_n - f_m|^p &= \left| \sum_{k=m}^{n-1} (f_{k+1} - f_k) \right|^p \\
 &\leq \left( \sum_{k=m}^{n-1} |f_{k+1} - f_k| \right)^p \\
 &\leq g^p
 \end{aligned}$$

Using the theorem of Lebesgue (7.5.5), the following is valid for  $n \rightarrow \infty$ :  $\mu(|f_n - f_m|^p) \rightarrow \mu(|f - f_m|^p)$

This leads to the following:

$$\begin{aligned}
 \|f - f_m\|_p^p &= \lim_{n \rightarrow \infty} \|f_n - f_m\|_p^p \\
 &= \lim_{n \rightarrow \infty} \left\| \sum_{k=m}^{n-1} (f_{k+1} - f_k) \right\|_p^p \\
 &\leq \lim_{n \rightarrow \infty} \left( \sum_{k=m}^{n-1} \|f_{k+1} - f_k\|_p \right)^p \\
 &\leq 2^{-pm+p}
 \end{aligned}$$

Hence,  $\|f - f_m\| \leq 2^{-m+1}$  is valid and so  $f \in \mathcal{L}^P(X, \mu)$  and  $f_m \rightarrow f$  using the  $\|\cdot\|_p$ -Norm.

□

### 7.5.3 Representation theorem of F. Riesz

The representation theorem of Riesz describes the coherence between measurement using systems of sets and measurement using linear forms. In literature there is not only *one* formulation of this theorem, but different ones concerning different types of mathematical spaces. This is well described in [Els05], p.328 et seqq. In this context, a formulation concerning an arbitrary locally compact  $T_2$  topological space  $(X, \mathcal{O})$  is given, which describes the coherence to the complex radon measure in definition 7.5.9. In this definition, the term *radon measure* is used in the meaning of measuring using sets. This term is defined in [Els05], p. 313 as a special type of a so-called *borel measure* and has to be distinguished from the one defined in the last subsection.

#### Theorem 7.5.8 (Representation theorem of F. Riesz)

*Let  $(X, \mathcal{O})$  be an arbitrary locally compact  $T_2$  topological space. Furthermore, let  $I$  be a **positive linear functional**  $I : C_C(X) \rightarrow \mathbb{C}$ . Then, there exists one*

## 7 Measure theory on topological and function spaces

and only one radon measure (concerning measuring using sets)  $\mu : \mathfrak{B}(X) \longrightarrow [0, \infty]$  ,

with:

$$I(f) = \int_X f d\mu, f \in \mathcal{C}_C(X).$$

# 8 Coherences of approximation theory

## 8.1 The theorems of Weierstraß and Stone

In this section, the term *approximation* is broadly considered, whereas in the beginning it is restricted looking on the usage within this thesis.

Regarding [Col73], p.9, the term *approximation* can be used to describe a wide field of methods of numerical mathematics. Hence, it is defined, that this term here is understood in the meaning of *describing mappings using other mappings* taking into account the generated error.

Collatz and Krabs describe in [Col73], p.165 et seqq. a main problem of approximation theory:

*In which way is it possible to approximate elements out of the vector space  $\mathcal{C}(X, \mathbb{R})$  using a subspace of  $\mathcal{C}(X, \mathbb{R})$ , when  $X$  is a compact metric space?*

This question is answered in the theorems of *Weierstraß* and *Stone*, which is focused in this section.

### **Definition 8.1.1 (Maximum norm)**

*Let  $(X, d)$  be a compact metric space and  $\mathcal{C}(X, \mathbb{R})$  the vector space of the continuous real-valued mappings on  $X$ . Then, the mapping:*

$$\begin{aligned} \|\cdot\|_\infty : \mathcal{C}(X, \mathbb{R}) &\longrightarrow \mathbb{R}_+^0 \\ f &\mapsto \|f\|_\infty := \max_{x \in X} |f(x)| \end{aligned}$$

defines a **norm**, the so-called **maximum norm** or **Tschebychow norm**.

In general, a polynomial is a formal expression  $p(t) = \sum_{i=1}^n a_i \cdot t^i$ , where  $a_i \in \mathbb{K}$  for  $i \in \{1, \dots, n\}$  are elements of an arbitrary field  $\mathbb{K}$  and  $t$  is an arbitrary variable. The set  $\mathbb{K}[t]$  then contains any of this polynomials. If  $t$  only takes values out of  $\mathbb{K}$ , then such a polynomial can be identified as a mapping of the type  $p : \mathbb{K} \longrightarrow \mathbb{K}$  according to [Fis93], p.61.

For an exact argument with methods of elementary algebra, one can refer to [Kar08], p.145 et seqq.

### Definition 8.1.2 (Algebraic group)

A pair  $(G, *)$  of an arbitrary set  $G$  and an arbitrary binary operator

$*$  :  $G \times G \longrightarrow G$  is called **(algebraic) group**, if the following properties are valid:

- (A)  $\forall_{a,b,c \in G} : (a * b) * c = a * (b * c)$  (associativity)
- (N)  $\exists_{e \in G} \forall_{a \in G} : a * e = e * a = a$  (identity element)
- (I)  $\forall_{a \in G} \exists_{a^{-1} \in G} : a * a^{-1} = a^{-1} * a = e$  (inverse element)

If additionally the following property

- (C)  $\forall_{a,b \in G} : a * b = b * a$  (commutativity)

is valid, the **group** is called **commutative group** or **abelian group**.

### Definition 8.1.3 (Algebraic ring)

A triple  $(R, *, \circ)$  of an arbitrary set  $R$  and two arbitrary binary operators

$*$  :  $R \times R \longrightarrow R$ ,  $\circ$  :  $R \times R \longrightarrow R$  is called **(algebraic) ring**, if the following properties are valid:

- (AG)  $(R, *)$  is an abelian group

## 8 Coherences of approximation theory

(HG)  $(R, \circ)$  is a half group, that means  $\circ$  is an associative binary operator in  $R$ .

(DI)  $\forall_{a,b,c \in R} : a \circ (b * c) = (a \circ b) * (b \circ c)$  (distributivity I)

(DII)  $\forall_{a,b,c \in R} : (a * b) \circ c = (a \circ c) * (b \circ c)$  (distributivity II)

According to [Kar08],  $(\mathbb{K}[t], +, \cdot)$  is an algebraic ring. These coherences are used to define polynomials above the field  $\mathbb{R}$ .

### Definition 8.1.4 (Polynomial ring above $\mathbb{R}$ )

The set  $\mathbb{R}[x]$  of polynomials with real coefficients can be denoted in the following way:

$$\mathbb{R}[x] := \left\{ p \in \mathcal{MAP}(\mathbb{R}, \mathbb{R}) \mid \forall_{x \in \mathbb{R}} : p(x) = \sum_{k=1}^{\infty} p_k x^k \wedge \forall_{k \in \mathbb{N}_0} p_k \in \mathbb{R} \wedge \exists_{n \in \mathbb{N}_0} \forall_{k \geq n} p_k = 0 \right\}$$

**Addition and scalar multiplication** of this polynomials are defined by the following mappings:

$$+ : \mathbb{R}[x] \times \mathbb{R}[x] \longrightarrow \mathbb{R}[x], (p, q) \mapsto p + q \text{ with } (p + q)(x) := \sum_{k=0}^{\infty} (p_k + q_k) \cdot x^k$$

$$\cdot : \mathbb{R}[x] \times \mathbb{R}[x] \longrightarrow \mathbb{R}[x], (\lambda, p) \mapsto \lambda \cdot p \text{ with } (\lambda \cdot p)(x) := \sum_{k=0}^{\infty} (\lambda \cdot p_k) \cdot x^k.$$

**Equality** of polynomials means:

$$\forall_{p, q \in \mathbb{R}[x]} \left( p = q : \iff \forall_{k \in \mathbb{N}_0} p_k = q_k \right)$$

Hence, according to the note given above,  $(\mathbb{R}[t], +, \cdot)$  is an **algebraic ring**.

For the proof of the theorem of Weierstraß, the definition of so-called *Bernstein polynomials* is necessary (see [Sch97], p.162):



**Definition 8.1.5 (Bernstein polynomial)**

Regarding the interval  $[0, 1]$ , the *i*-th **Bernstein polynomial of grade  $n$**  is defined by the following mapping:

$$\begin{aligned} \mathcal{BS}_i^n : [0, 1] &\longrightarrow [0, 1] \\ \lambda &\mapsto \mathcal{BS}_i^n(\lambda) := \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \end{aligned}$$

for  $i \in \{0, 1, \dots, n\}$ .

**Theorem 8.1.1 (Approximation theorem of Weierstraß)**

Let  $I$  be the closed interval  $I := [a, b] \subset \mathbb{R}$  and  $f \in \mathcal{C}(I, \mathbb{R})$ . Then,  $f$  can be approximated uniformly  $[a, b]$  by using polynomials out of  $\mathbb{R}[x]$ , that means:

$$\forall \epsilon > 0 \quad \exists p \in \mathbb{R}[x] : \|f - p\|_\infty < \epsilon$$

The proof given at this point was developed by Sergei Natanowitsch Bernstein in [Ber12] and bases on the description in [Naa97], p.51 et seqq.

*Proof.* In the following, the interval  $[0, 1]$  is regarded. This is no restriction, because any  $x \in [a, b]$  of an arbitrary closed interval in  $\mathbb{R}$  can be mapped onto an element  $\lambda \in [0, 1]$  by using the following affine mapping:

$$\begin{aligned} \lambda : [a, b] &\longrightarrow [0, 1] \\ x &\mapsto \lambda(x) := \frac{x-a}{b-a} \end{aligned}$$

Using the binomial theorem, the following can be concluded:

$$1 = (\lambda + (1 - \lambda))^n = \sum_{i=0}^n \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} = \sum_{i=0}^n B_i^n(\lambda)$$

So, 1 is separated by  $(n + 1)$  Bernstein polynomials.

Due to the compactness of  $[0, 1]$ , the mapping  $f$  is uniformly continuous there. Hence, for any  $\epsilon > 0$  there exists a  $\delta > 0$ , such that for any  $x \in [0, 1]$  with  $|x - \frac{i}{n}| < \delta$  always  $|f(x) - f(\frac{i}{n})| < \frac{\epsilon}{2}$  is valid.

For the approximation of  $f$ , the following polynomials are defined:

## 8 Coherences of approximation theory

$$\begin{aligned} \mathcal{BS}_{n,f} : [0, 1] &\longrightarrow \mathbb{R} \\ \lambda &\longmapsto \mathcal{BS}_{n,f}(\lambda) := \sum_{i=0}^n \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot f\left(\frac{i}{n}\right) \end{aligned}$$

for  $n \in \mathbb{N}$

Using this polynomials,  $\forall_{\epsilon > 0} \exists_{p \in \mathbb{R}[x]} : \|f - p\|_{\infty} < \epsilon$  is equivalent to

$$\forall_{\epsilon > 0} \exists_{n_0(\epsilon) \in \mathbb{N}} \forall_{n \geq n_0} : \|\mathcal{BS}_{n,f} - f\|_{\infty} < \epsilon$$

So, regarding  $f$ , the following equation is valid:

$$\begin{aligned} \mathcal{BS}_{n,f}(\lambda) - f(\lambda) &= \left( \sum_{i=0}^n \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot f\left(\frac{i}{n}\right) \right) - f(\lambda) \\ &= \left( \sum_{i=0}^n \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot f\left(\frac{i}{n}\right) \right) \\ &\quad - \left( \sum_{i=0}^n \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot f(\lambda) \right) \\ &= \sum_{i=0}^n \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left( f\left(\frac{i}{n}\right) - f(\lambda) \right) \end{aligned}$$

.

This last sum is now separated into two parts:

$$\{0, \dots, n\} \supseteq K := \left\{ i \in \{0, \dots, n\} \mid \left| \frac{i}{n} - \lambda \right| < \delta \right\}$$

$$\{0, \dots, n\} \supseteq L := \left\{ i \in \{0, \dots, n\} \mid \left| \frac{i}{n} - \lambda \right| \geq \delta \right\}$$

$$\begin{aligned} \sum_{i=0}^n \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left( f\left(\frac{i}{n}\right) - f(\lambda) \right) &= \sum_{i \in K} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left( f\left(\frac{i}{n}\right) - f(\lambda) \right) + \\ &\quad \sum_{i \in L} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left( f\left(\frac{i}{n}\right) - f(\lambda) \right) \end{aligned}$$

Now, an estimation for each of these two parts is regarded:

$$\begin{aligned}
 |SK| &= \left| \sum_{i \in K} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left( f\left(\frac{i}{n}\right) - f(\lambda) \right) \right| \\
 &\leq \sum_{i \in K} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left| f\left(\frac{i}{n}\right) - f(\lambda) \right| \\
 &\leq \sum_{i \in K} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \frac{\epsilon}{2} \\
 &= \frac{\epsilon}{2} \cdot \sum_{i \in K} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \\
 &= \frac{\epsilon}{2}
 \end{aligned}$$

Here it can be seen, that this estimation is valid for any  $n \in \mathbb{N}$ .

For  $|SL|$ , the validity of the following is used:

- (i)  $|f\left(\frac{i}{n}\right) - f(\lambda)| \leq |f\left(\frac{i}{n}\right)| + |f(\lambda)| \leq 2 \cdot \max_{\lambda \in [0,1]} |f(\lambda)| := 2M$
- (ii)  $\left|\frac{i}{n} - \lambda\right| \geq \delta \Leftrightarrow \left(\frac{i}{n} - \lambda\right)^2 \geq \delta^2 \Leftrightarrow \frac{1}{\left(\frac{i}{n} - \lambda\right)^2} \leq \frac{1}{\delta^2}$
- (iii)  $\frac{\lambda(1-\lambda)}{n} = \sum_{i=0}^n \left(\frac{i}{n} - \lambda\right)^2 \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left(f\left(\frac{i}{n}\right) - f(\lambda)\right)$
- (iv)  $\forall_{\lambda \in [0,1]} : (-4\lambda^2 + 4\lambda - 1 = -(2\lambda - 1)^2 \leq 0 \Rightarrow \lambda(1 - \lambda) \leq \frac{1}{4})$

8 Coherences of approximation theory

$$\begin{aligned}
|SL| &= \left| \sum_{i \in L} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left( f\left(\frac{i}{n}\right) - f(\lambda) \right) \right| \\
&\leq \sum_{i \in K} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \left| f\left(\frac{i}{n}\right) - f(\lambda) \right| \\
&\stackrel{(i)}{\leq} 2M \cdot \sum_{i \in K} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \frac{\epsilon}{2} \\
&= 2M \cdot \sum_{i \in K} \frac{\left(\frac{i}{n} - \lambda\right)^2}{\left(\frac{i}{n} - \lambda\right)^2} \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \frac{\epsilon}{2} \\
&\stackrel{(ii)}{\leq} \frac{2M}{\delta^2} \cdot \sum_{i \in K} \left(\frac{i}{n} - \lambda\right)^2 \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} \cdot \frac{\epsilon}{2} \\
&\stackrel{(iii)}{\leq} \frac{2M}{\delta^2} \cdot \frac{\lambda(1 - \lambda)}{n} \\
&\stackrel{(iv)}{\leq} \frac{2M}{\delta^2} \cdot \frac{1}{4n} \\
&= \frac{M}{\delta^2 \cdot 2n}
\end{aligned}$$

Now, let  $n_0(\epsilon) := \frac{M}{\epsilon\delta^2}$  for a given  $\epsilon > 0$ . Then, the following inequality is valid:

$$\begin{aligned}
|\mathcal{BS}_{n,f}(\lambda) - f(\lambda)| &\leq |SK| + |SL| \\
&\leq \frac{\epsilon}{2} + \frac{M}{\delta^2 \cdot 2 \cdot \frac{M}{\epsilon\delta^2}} \\
&= \frac{\epsilon}{2} + \frac{M\epsilon\delta^2}{\delta^2 \cdot M \cdot 2} \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon
\end{aligned}$$

This estimation is valid for any  $\lambda \in [0, 1]$  and so is shown the following:

$$\|\mathcal{BS}_{n,f} - f\|_\infty < \epsilon,$$

and due to the generality of  $\epsilon$ , the statement of this theorem is proven. □

In literature, other proofs of this theorem can be found, one example is the use of convolutions of Landau levels in [Kön04], p.313-314.

The statement given in the theorem of Weierstraß is formulated in a more generalised way in the theorem of Stone-Weierstraß, which is formulated in [Naa97], p.57.:

**Theorem 8.1.2 (Approximation theorem of Stone-Weierstraß)**

*Let  $(X, \mathcal{O})$  be an arbitrary compact topological space. In addition to that, let  $\mathcal{R}$  be a subring of  $\mathcal{C}(X)$  with the following properties:*

(i) *Any constant mapping in  $X$  is an element of  $\mathcal{R}$ .*

(ii)  $\forall_{x,y \in X} \exists_{p \in \mathcal{R}} : p(x) \neq p(y)$

*Then, the following is valid:*

$$\overline{\mathcal{R}} = \mathcal{C}(X)$$

*That means, that  $\mathcal{R}$  is a dense subset of  $\mathcal{C}(X)$ , thus*

$$\forall_{\substack{f \in \mathcal{C}(X) \\ \epsilon > 0}} \exists_{p \in \mathcal{R}} : \|f - p\|_{\infty} < \epsilon$$

The proof given in [Naa97], p.56-61 uses the construction of a finite number of mappings  $p_{yi} \in \mathcal{R}$ ,  $i \in \{1, \dots, n\}$  for each  $y \in X$  in the context of a given mapping  $f \in \mathcal{C}(X)$ . For a given  $\epsilon > 0$ , the maximum  $g$  out of these mappings can be characterised by:

$$\|f - g\|_{\infty} \leq \frac{\epsilon}{2}.$$

Furthermore it is shown, that there can be found an appropriate mapping  $h \in \mathcal{R}$  with:

## 8 Coherences of approximation theory

$$\|g - h\|_{\infty} < \frac{\epsilon}{2},$$

so the following is valid

$$\|f - h\|_{\infty} \leq \|f - g\|_{\infty} + \|g - h\|_{\infty} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Due to the arbitrary of  $\epsilon$ , the statement  $\overline{\mathcal{R}} = \mathcal{C}(X)$  is valid.

In [Naa97], p.61-63 and in [Kön97], p.320, the generalised statement of this theorem is applied on a subset of polynomials in  $\mathcal{C}(\mathbb{R}^n)$ .

## 8.2 The Bézier interpolation approach describing curves

In this section, a so-called *Bézier-splines* interpolation approach is described.

Regarding his description is done in a more generalised way as it can normally be found in literature. There, this approach only describes interpolation in the space  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ , what is satisfying there, because as described in the next passage, this method is usually used in the field of computer graphics.

The coherences of the Bézier-approach are mainly based on the explanations in [Sch97], pp.161-182 and [Deu93], pp.203-219, but extended to a generalised description with exact mathematical formalization. Compared to other spline approaches, this one does not make use of complicated equation systems but it uses geometric characteristics in combination with parametrisations. This fact is one reason to take this method into account in this thesis, because computing cost can be reduced using it. As found in [Brü01], p.201, the Bézier interpolation bases on a set of *controlling points*, which define the according so-called *characteristic polygon*. This technique has been developed in the context of describing *curves* in CAD software in car industry and the mainly involved persons were the two applied mathematicians *Pierre Etienne Bézier (Renault)* and *Paul de Faget de Casteljau (Citroën)*. For more detailed historical notes refer to [Sal06], p.175 et seqq.

For the description and definition of the Bézier-interpolation, some characteristics of the Bernstein polynomials defined in section 8.1 are necessary.

### Theorem 8.2.1 (Characteristics of Bernstein polynomials)

Let  $\mathcal{BS}_i^n : [0, 1] \longrightarrow [0, 1]$  be the  $i$ -th Bernstein polynomial of grade  $n$ . Then, the following characteristics can be formulated:

$$(i) \quad \forall_{\substack{\lambda \in [0,1] \\ i \in \{0, \dots, n\}}} : \mathcal{BS}_i^n(\lambda) = \mathcal{BS}_{n-i}^n(1 - \lambda)$$

## 8 Coherences of approximation theory

$$(ii) \quad \bigvee_{\substack{\lambda \in [0,1] \\ i \in \{0, \dots, n\}}} : (1 - \lambda) \cdot \mathcal{BS}_0^n(\lambda) = \mathcal{BS}_0^{n+1}(\lambda)$$

$$(iii) \quad \bigvee_{\substack{\lambda \in [0,1] \\ i \in \{0, \dots, n\}}} : \lambda \cdot \mathcal{BS}_n^n(\lambda) = \mathcal{BS}_{n+1}^{n+1}(\lambda)$$

$$(iv) \quad \bigvee_{\substack{\lambda \in [0,1] \\ i \in \{0, \dots, n\}}} : \mathcal{BS}_i^n(\lambda) = \lambda \cdot \mathcal{BS}_{i-1}^{n-1}(\lambda) + (1 - \lambda) \cdot \mathcal{BS}_i^{n-1}(\lambda)$$

*Proof.* Let  $\lambda \in [0, 1]$  and  $i \in \{0, \dots, n\}$  be arbitrary.

At (i):  $\mathcal{BS}_i^n(\lambda) = \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i} = \binom{n}{n-i} (1 - \lambda)^{n-i} \cdot \lambda^{n-(n-i)} = \mathcal{BS}_{n-i}^n(1 - \lambda).$

At (ii):  $(1 - \lambda) \cdot \mathcal{BS}_0^n(\lambda) = (1 - \lambda) \cdot \binom{n}{0} \lambda^0 \cdot (1 - \lambda)^n = \binom{n+1}{0} (1 - \lambda)^{n+1} = \mathcal{BS}_0^{n+1}(\lambda).$

At (iii):  $\lambda \cdot \mathcal{BS}_n^n(\lambda) = \lambda \cdot \binom{n}{n} \lambda^n = \binom{n+1}{n+i} \cdot \lambda^{n+1} = \mathcal{BS}_{n+1}^{n+1}(\lambda)$

At (iv): The following chain of equalities for the binomial coefficient is valid:

$$\binom{n-1}{i-1} + \binom{n-1}{i} = \frac{(n-1)!}{(i-1)! \cdot (i-1)!} + \frac{(n-1)!}{i! \cdot (n-1-i)!} = \frac{n!}{i! \cdot (n-i)!} = \binom{n}{i}$$

And so, the following is valid:

$$\begin{aligned} \lambda \cdot \mathcal{BS}_{i-1}^{n-1}(\lambda) + (1 - \lambda) \cdot \mathcal{BS}_i^{n-1}(\lambda) &= \lambda \cdot \binom{n-1}{i-1} \cdot \lambda^{i-1} \cdot (1 - \lambda)^{n-i} + (1 - \lambda) \cdot \\ &\binom{n-1}{i} \cdot \lambda^i \cdot (1 - \lambda)^{n-1-i} = \binom{n-1}{i-1} \cdot \lambda^i \cdot (1 - \lambda)^{n-i} + \binom{n-1}{i} \cdot \lambda^i \cdot (1 - \lambda)^{n-i} = \\ &\mathcal{BS}_i^n(\lambda) \end{aligned}$$

□

In literature of numerical mathematics, it is shown, that these Bernstein polynomials define a basis of real valued polynomial vector spaces. Regarding the set  $\mathbb{R}[x]$  defined in 8.1.4 by

$$\mathbb{R}[x] := \left\{ p \in \mathcal{MAP}(\mathbb{R}, \mathbb{R}) \mid \bigvee_{x \in \mathbb{R}} : p(x) = \sum_{k=1}^{\infty} p_k x^k \wedge \bigvee_{k \in \mathbb{N}_0} p_k \in \mathbb{R} \wedge \bigvee_{n \in \mathbb{N}_0} \bigvee_{k \geq n} p_k = 0 \right\}$$

,

this is a vector space  $(\mathbb{R}[x], +, \cdot)$  above the field  $\mathbb{K}$  using the following mappings  $+$  and  $\cdot$ :



## 8 Coherences of approximation theory

$$+ : \mathbb{R}[x] \times \mathbb{R}[x] \longrightarrow \mathbb{R}[x], (p, q) \mapsto p + q \text{ with } (p + q)(x) := \sum_{k=0}^{\infty} (p_k + q_k) \cdot x^k$$

$$\cdot : \mathbb{K} \times \mathbb{R}[x] \longrightarrow \mathbb{R}[x], (\lambda, p) \mapsto \lambda \cdot p \text{ with } (\lambda \cdot p)(x) := \sum_{k=0}^{\infty} (\lambda \cdot p_k) \cdot x^k.$$

**Equality** of polynomials means:

$$\forall_{p, q \in \mathbb{R}[x]} \left( p = q \iff \forall_{k \in \mathbb{N}_0} p_k = q_k \right)$$

Hence, the space  $(\mathbb{R}_n[x], +, \cdot)$  together with this two mappings is a vector space as well as the set  $(\mathbb{R}_n[x], +, \cdot)$  of real valued polynomials with maximum grade  $n \in \mathbb{N}$  with:

$$\mathbb{R}_n[x] := \left\{ p \in \mathcal{MAP}(\mathbb{R}, \mathbb{R}) \mid \forall_{x \in \mathbb{R}} : p(x) = \sum_{k=0}^n p_k x^k \wedge \forall_{k \in \mathbb{N}_0} p_k \in \mathbb{R} \right\}$$

Regarding this vector spaces, the set  $\mathcal{BS} := \{\mathcal{BS}_0^n, \mathcal{BS}_1^n, \dots, \mathcal{BS}_n^n\}$  defines a *basis*.

For the according theorem and the proof refer to [Sch97], pp.164-165.

Conventionally, Bézier-curves are used in computer graphics as mappings of the type  $f : \mathbb{R} \longrightarrow \mathbb{R}^d$ ,  $d \in \mathbb{N}$  basing on given points out of the space  $\mathbb{R}^d$ .

According to [Sch97], every polynomial out of  $\mathbb{R}_n[x]$  but also polynomials basing on elements out of the space  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ , can be piecewise realised by so-called *Bézier curves*. Therefor, the space  $\mathbb{R}_n^d[x]$  is defined in the following way:

$$\mathbb{R}_n^d[x] := \left\{ p \in \mathcal{MAP}(\mathbb{R}, \mathbb{R}^d) \mid \forall_{x \in \mathbb{R}} : p(x) = \sum_{k=0}^n p_k x^k \wedge \forall_{k \in \mathbb{N}_0} p_k \in \mathbb{R}^d \right\}$$

The so-called *Bézier-realization* of a polynomial is then defined in the following way.

**Definition 8.2.1 (Bézier curves according to a given interval of parameters)**

Let  $CM := \{b_0, \dots, b_n\} \subseteq \mathbb{R}^d$  be a discrete subset. Then, the **Bézier-realization**  $C$  of  $CM$  using an arbitrary interval  $[a, b] \subset \mathbb{R}$  of parameters is the following:

$$C = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(\cdot; a, b), \text{ with } \forall_{i \in \{0, \dots, n\}} b_i \in \mathbb{R}^d \text{ and}$$

$$\forall_{t \in [a, b]} C(t) = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(t; a, b)$$

If the standard interval  $[0, 1]$  of parameters is used, the according **Bézier-realization**  $C^s$  is denoted in the following way:

$$C^s = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n, \text{ with } \forall_{i \in \{0, \dots, n\}} b_i \in \mathbb{R}^d \text{ and}$$

$$\forall_{t \in [0, 1]} C^s(t) = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(t)$$

The  $b_i \in \mathbb{R}^d, i \in \{0, \dots, n\}$

This Bézier-realization is a polynomial  $P \in \mathbb{R}[x]$  with maximum grade  $n$ .

are called **controlling-** or **Bézier-points**. The according polygonal chain that connects these points is called **Bézier polygon** or **characteristic polygon**.

The terms *graph* and *image* of a mapping can be described in the following way:

**Definition 8.2.2 (Graph and image of a mapping)**

Let  $F$  be an arbitrary mapping out of  $\mathcal{MAP}(\mathbb{D}, \mathbb{W})$ , where  $\mathbb{D}$  and  $\mathbb{W}$  are arbitrary sets. Then, the **graph** of this mapping is defined as the following subset of the cartesian product of  $\mathbb{D}$  and  $\mathbb{W}$ :

$$\text{Graph}(F) := \{(d, w) \in \mathbb{D} \times \mathbb{W} \mid F(d) = w\} \quad .$$

## 8 Coherences of approximation theory

The according **image** of this mapping can then be described by:

$$\text{Imag}(F) := F(\mathbb{D}) = \left\{ w \in \mathbb{W} \mid \exists_{d \in \mathbb{D}} : F(d) = w \right\} .$$

So, the following relationship between the polynomial  $P \in \mathbb{R}_n^d[x]$  and the Bézier-realization  $C$  respectively  $C^s$  can be noted:

$$\text{Imag}(C) \subset \text{Graph}(P) \text{ respectively } \text{Imag}(C^s) \subset \text{Graph}(P).$$

The approach that is proposed in connection with these geometric coherences consists in a kind of generalization. Instead of only regarding spaces of the type  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ , arbitrary **topological vector spaces** are used. One example, that is looked at more detailed further below, is the space of continuous mappings of the type  $f : X \longrightarrow \mathbb{R}$ , where  $X$  is a topological vector space itself.

So the following concepts and terms, that are classically used with the space  $\mathbb{R}^d$  in the context of Bézier-Interpolation are described in the mentioned generalised way in this thesis. As a matter of principle, these coherences are basing on the characteristics of the Bernstein polynomials, especially on the ability of building a partition of 1. For that reason, the set  $X$  is regarded as an arbitrary topological vector space in the following statements and definitions.

In the context of generalization, the term *topological Bézier-realization* of elements out of  $X$  is defined.

### **Definition 8.2.3 (Topological Bézier curves)**

Let  $\mathcal{BM} := \{b_0, \dots, b_n\} \subset X$  be a discrete subset of  $X$  with  $n + 1$  elements. Then, the **topological Bézier-realization**  $C$  of  $\mathcal{BM}$  using an arbitrary interval

## 8 Coherences of approximation theory

$[a, b] \subset \mathbb{R}$  of parameters is the following:

$$C = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(\cdot; a, b), \text{ with } \forall_{i \in \{0, \dots, n\}} b_i \in \mathcal{BM} \text{ and}$$

$$\forall_{t \in [a, b]} C(t) = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(t; a, b)$$

If the standard interval  $[0, 1]$  of parameters is used, the according **Bézier-realization**  $C^s$  is denoted in the following way:

$$C^s = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n, \text{ with } \forall_{i \in \{0, \dots, n\}} b_i \in \mathcal{BM} \text{ and}$$

$$\forall_{t \in [0, 1]} C^s(t) = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(t)$$

The  $b_i \in \mathcal{BM}$ ,  $i \in \{0, \dots, n\}$

are called **controlling-** or **Bézier-points**. The according polygonal chain that connects these points is called **Bézier polygon** or **characteristic polygon**.

This kind of **topological Bézier-realization**. should be synonymously used with the term **Topological Bézier curve**.

For the characterization of the Bézier-points, the terms *convex set* and *convex hull* are necessary.

### Definition 8.2.4 (Convex set, convex combination and and convex hull)

Let  $A \subset X$  be an arbitrary subset. This set  $A$  is called **convex**, if for every pair  $x, y \in A$  the **line segment**  $\overline{xy}$  also contains to  $A$ , that means, that the following proposition is true:

$$\forall_{x, y \in A} : \overline{xy} := \{\lambda \cdot x + (1 - \lambda) \cdot y \mid \lambda \in [0, 1]\} \subset A$$

A **linear combination** of the type:

## 8 Coherences of approximation theory

$x = \sum_{i=1}^k \lambda_i \cdot x_i, \quad \forall_{i \in \{1, \dots, k\}} : x_i \in X, \quad \forall_{i \in \{1, \dots, k\}} : \lambda_i \geq 0 \text{ and } \sum_{i=1}^k \lambda_i = 1$   
 is called **convex combination** of elements out of  $X$ .

The **convex hull**  $\text{conv}(A)$  of  $A$  is the smallest convex subset in  $X$ , which contains  $A$ . This is the intersection of all convex supsets of  $A$  or the set of all finite convex combinations with elements out of  $A$ :

$$\begin{aligned}
 \text{conv}(A) &:= \bigcap_{\substack{A \subset B \subset X \\ B \text{ convex}}} B \\
 &= \left\{ x = \sum_{i=1}^k \lambda_i \cdot x_i \mid k \in \mathbb{N} \wedge m < \infty, \right. \\
 &\quad \left. \forall_{i \in \{1, \dots, k\}} : x_i \in A, \quad \forall_{i \in \{1, \dots, k\}} : \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1 \right\}
 \end{aligned}$$

### Theorem 8.2.2 (Convex hull of Bézier curves)

The set  $M$  containing any point of a topological Bézier-realization according to the standard interval with

$$M := \left\{ \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(\lambda) \mid \lambda \in [0, 1] \right\}$$

is a subset of the convex hull of the  $(n + 1)$  Bézier-points  $b_0, b_1, \dots, b_n$ .

*Proof.* As shown above, the Bernstein polynomials can represent a non-negative finite representation of 1 by  $\forall_{\substack{n \in \mathbb{N} \\ \lambda \in [0, 1]}} : \sum_{i=0}^n \mathcal{BS}_i^n(\lambda) = 1$ . This leads to the fact, that

$$\forall_{\substack{n \in \mathbb{N} \\ \lambda \in [0, 1]}} : \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(\lambda) \text{ is a finite convex combination of } b_0, b_1, \dots, b_n.$$

So, any of these sums is an element of the convex hull of these the  $(n + 1)$  Bézier-points and the theorem is valid.

□

For calculating the coordinates of a topological Bézier-realization regarding a given element out of the domain  $[0, 1]$ , there exists a stable algorithm called *algorithm of de Casteljau*. This algorithm bases on the partition of the Bézier curve using so-called *partial curves* for approximation. The proposed generalised description is demonstrated in the following.

**Definition 8.2.5 (Partial curves of a Bézier curve)**

Let  $\mathcal{BM} := \{b_0, \dots, b_n\} \subset X$  be a discrete subset of  $X$  with  $n + 1$  elements and Bézier-realization  $C^s$  using the interval  $[0, 1]$  of parameters. Then, the **partial curves**  $b_i^k \in \mathcal{MAP}(\mathbb{R}, X)$  of  $C^s$  with  $i \in \{0, \dots, n - k\}$  and  $k \in \{0, \dots, n\}$  are defined by the following:

$$b_i^k := \sum_{j=0}^k b_{i+j} \cdot B_j^k = \sum_{j=i}^{i+k} b_j \cdot B_{j-i}^k$$

Hence, these partial curves  $b_i^k$  are Bézier curves of grade  $k$  using the Bézier-points  $b_i, \dots, b_{i+k}$ .

It is easily seen, that  $\forall_{\lambda \in [0,1]} b_0^n(\lambda) = \sum_{j=0}^n b_j \cdot \mathcal{BS}_j^n(\lambda) = C^s(\lambda)$  and  $b_i^0(\lambda) = \sum_{j=0}^0 b_{i+j} \cdot \mathcal{BS}_j^0(\lambda) = b_i \cdot \mathcal{BS}_0^0(\lambda) = b_i \cdot \lambda^0 \cdot (1 - \lambda)^0 = b_i$ .

Then, the following recursion can be done:

**Lemma 8.2.1 (Recursion of partial Bézier curves)**

Let  $\mathcal{BM} := \{b_0, \dots, b_n\} \subset X$  be a discrete subset of  $X$  with  $n + 1$  elements and Bézier-realization  $C^s$  using the interval  $[0, 1]$  of parameters. Then, for the **partial curves**  $b_i^k \in \mathcal{MAP}(\mathbb{R}, X)$  of  $C^s$  with  $i \in \{0, \dots, n - k\}$  and  $k \in \{0, \dots, n\}$ , the following **recursion** is valid:

$$\forall_{\substack{\lambda \in [0,1] \\ i \in \{0, \dots, n - k\} \\ k \in \{0, \dots, n\}}} b_i^k(\lambda) = (1 - \lambda) \cdot b_i^{k-1} + \lambda \cdot b_{i+1}^{k-1}$$

*Proof.* Let  $\mathcal{BM} := \{b_0, \dots, b_n\} \subset X$  be a discrete subset of  $X$  with  $n + 1$  elements with Bézier-realization  $C^s$  using the interval  $[0, 1]$  of parameters. Furthermore, let

8 Coherences of approximation theory

$i \in \{0, \dots, n - k\}$ ,  $k \in \{0, \dots, n\}$  and  $\lambda \in [0, 1]$  be arbitrary. Then, the following chain of equations can be build:

$$\begin{aligned}
 b_i^k(\lambda) &= b_i \mathcal{BS}_0^k(\lambda) + \sum_{j=1}^{k-1} b_{i+j} \mathcal{BS}_j^k(\lambda) + b_{i+k} \mathcal{BS}_k^k(\lambda) \\
 &= b_i \cdot (1 - \lambda) \cdot \mathcal{BS}_0^{k-1}(\lambda) + \\
 &\quad \sum_{j=1}^{k-1} b_{i+j} \cdot ((1 - \lambda) \mathcal{BS}_j^{k-1}(\lambda) + \lambda \cdot \mathcal{BS}_{j-1}^{k-1}(\lambda)) + b_{i+k} \mathcal{BS}_k^k(\lambda) \\
 &= \sum_{j=0}^{k-1} b_{i+j} \cdot (1 - \lambda) \mathcal{BS}_j^{k-1}(\lambda) + \sum_{j=1}^k b_{i+j} \lambda \cdot \mathcal{BS}_{j-1}^{k-1}(\lambda) \\
 &= (1 - \lambda) \cdot b_i^{k-1}(\lambda) + \lambda \cdot b_{i+1}^{k-1}(\lambda)
 \end{aligned}$$

Due to the arbitrary choice of  $i, k$  and  $\lambda$ , the above lemma is valid. □

Using the characteristic  $\forall_{\substack{i \in \{0, \dots, n\} \\ \lambda \in [0, 1]}} : b_i^0(\lambda) = b_i$  and lemma 8.2.1, a value  $C^s(\lambda) = b_0^n$  for an arbitrary parameter  $\lambda \in [0, 1]$  can be calculated by an iterative convex combination following the so-called *scheme of Casteljau* describing the process while using the *algorithm of de Casteljau*.

$$\begin{array}{ccccccc}
 b_n & = & b_n^0 & & & & \\
 & & & \searrow & & & \\
 b_{n-1} & = & b_{n-1}^0 & \rightarrow & b_{n-1}^1 & & \\
 & & \vdots & & \ddots & & \\
 b_1 & = & b_1^0 & \rightarrow & \dots & \rightarrow & b_1^{n-1} \\
 & & & \searrow & & & \searrow \\
 b_0 & = & b_0^0 & \rightarrow & \dots & \rightarrow & b_0^{n-1} & \rightarrow & b_0^n
 \end{array}$$

**Algorithm 8.2.1 (Algorithm of de Casteljau)**

Let  $\mathcal{BM} := \{b_0, \dots, b_n\} \subset X$  be a discrete subset of  $X$  with  $n + 1$  elements and Bézier-realization  $C^s$  using the interval  $[0, 1]$  of parameters. Then, a value  $C^s(\lambda) = b_0^n$  for an arbitrary parameter  $\lambda \in [0, 1]$  can be calculated by the following algorithm:

For  $i = 0, \dots, n$ :

$$b_i^0 := b_i;$$

For  $k = 1, \dots, n$

For  $i = 0, \dots, n - k$

$$b_i^k(\lambda) := (1 - \lambda) \cdot b_i^{k-1}(\lambda) + \lambda \cdot b_{i+1}^{k-1}(\lambda)$$

**Example 8.2.1 (Bézier curve on functional spaces)**

Let  $X$  be the space  $\mathcal{C}(\mathbb{R}, \mathbb{R})$  of continuous mappings with the set  $\mathbb{R}$  as domain and codomain. Furthermore, let the following two elements  $f_0$  and  $f_1$  out of  $\mathcal{C}(\mathbb{R}, \mathbb{R})$  be given with:

$$\forall_{x \in \mathbb{R}} : f_0(x) := \sin(x)$$

$$\forall_{x \in \mathbb{R}} : f_1(x) := \cos(x)$$

The according Bézier curve  $C_f$  of grade 2 is the following:

$$C_f = \sum_{i=0}^1 f_i \cdot \mathcal{BS}_i^1 = \sin(x) \cdot \mathcal{BS}_0^1 + \cos(x) \cdot \mathcal{BS}_1^1$$

with

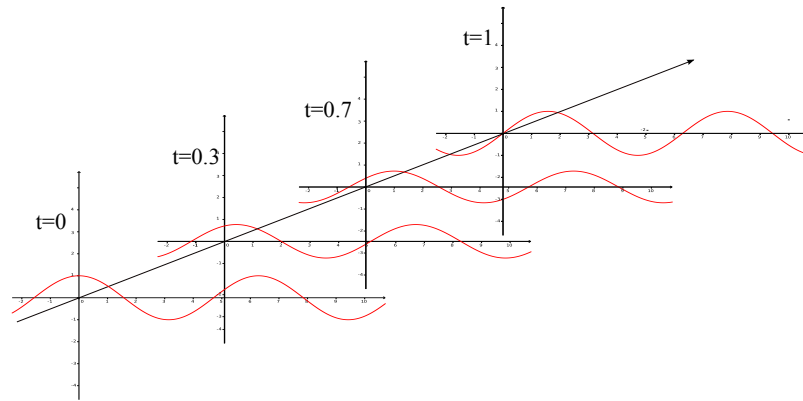
$$\forall_{t \in [0,1]} C_f(t) = \sin \cdot \mathcal{BS}_0^1(t) + \cos \cdot \mathcal{BS}_1^1(t) = \sin \cdot t^0 \cdot (1 - t)^1 + \cos \cdot t^1 \cdot (1 - t)^0$$

In figure 8.1, there are shown four cuttings of Graph ( $C^s$ ) using different parameters  $t \in [0, 1]$ .

In the following, the term  $\mathcal{C}^k$ -continuous,  $k \in \mathbb{N}_0$  as a characteristic of a mapping means continuity of this mapping right up to the  $k$ -th derivation.

As seen in definition 8.2.3, using  $n + 1$  Bézier-points, a Bézier curves of grade  $n$





**Figure 8.1:** Four cuttings of the graph of the curve  $C_f$

is generated. So, the more the number of Bézier-points increases, the more the grade of the resulting curve increases. But this growth of the according grade leads to an increasing effort of iteration. Hence, using arbitrary finite numbers of Bézier-points, a composition of different Bézier curves is more practical. To avoid a mixing-up of the terms, the “main” Bézier curve is called **Bézier-spline curve**, the participating ones are called **Bézier segments**. This composition is often done (see [Bun02], p.98) using cubic Bézier segments basing on four Bézier-points. One possibility of getting such a Bézier-spline curve is to regard the Bézier segments of every four neighboring Bézier-points. But in general, the resulting curve is only  $\mathcal{C}^0$ -continuous. To get an allover smooth curve, some relationships and characteristics between the single Bézier segments are regarded in the following.

To have the ability to guarantee the uniqueness of the first derivation or even of the curvilinearity of the Bézier-spline curve, the first and/ or respectively the second derivation of neighboring Bézier segments has to be equal at the linking Bézier-points.

The proofs of the following theorems and lemmas can be found in [Sch97], p.165 et seqq.

**Theorem 8.2.3 (Derivation of Bernstein polynomials)**

Let  $\mathcal{BS}_i^n : [0, 1] \longrightarrow [0, 1]$   
 $\lambda \mapsto \mathcal{BS}_i^n(\lambda) := \binom{n}{i} \lambda^i \cdot (1 - \lambda)^{n-i}$

for  $i \in \{0, 1, \dots, n\}$ .

be the *i*-th Bernstein polynomial of grade *n* above the interval  $[0, 1]$ .

Then, the first derivation is given by:

$$\frac{d}{d\lambda} \mathcal{BS}_i^n(\lambda) = \begin{cases} -n \cdot \mathcal{BS}_0^{n-1}(\lambda) & \text{for } i = 0 \\ n \cdot [\mathcal{BS}_{i-1}^{n-1}(\lambda) - \mathcal{BS}_i^{n-1}(\lambda)] & \text{for } i \in \{1, \dots, n-1\} \\ n \cdot \mathcal{BS}_{n-1}^{n-1}(\lambda) & \text{for } i = n \end{cases}$$

**Theorem 8.2.4 (Characteristics of border points)**

Let  $\mathcal{BM} := \{b_0, \dots, b_n\} \subset X$  be a discrete subset of *X* with *n* + 1 elements and Bézier-realization *C<sup>s</sup>* using the interval  $[0, 1]$  of parameters, while  $n > 2$ .

There, for *C<sup>s</sup>*, the following is applying:

$$C^s = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n, \text{ with } \forall_{i \in \{0, \dots, n\}} b_i \in X \text{ and}$$

$$\forall_{t \in [0, 1]} C^s(t) = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n(t)$$

Furthermore, let  $i \in \{0, \dots, n - k\}$ ,  $k \in \{0, \dots, n\}$  and  $\lambda \in [0, 1]$  be arbitrary.

Then, the following is valid:

- $C^s(0) = b_0$
- $C^s(1) = b_n$
- $\frac{d}{dt} C^s(0) = n \cdot (b_1 - b_0)$
- $\frac{d}{dt} C^s(1) = n \cdot (b_n - b_{n-1})$
- $\frac{d}{d^2t} C^s(0) = n \cdot (n - 1)(b_2 - 2 \cdot b_1 + b_0)$
- $\frac{d}{d^2t} C^s(1) = n \cdot (n - 1)(b_n - 2 \cdot b_{n-1} + b_{n-2})$

This leads to the fact, that the Bézier-point  $b_0$  defines the beginning and  $b_n$  defines

the end of the resulting Bézier curve. Furthermore it can be seen, that the first respectively the last segment of the regarded Bézier polygon is a *tangent* of this Bézier curve and that the second derivation on the beginning and the end of this curve can be described only using the two next neighboring Bézier-points.

At this point, a Bézier-spline curve is built by connecting a number of Bézier segments in an appropriate way. For the resulting Bézier-spline curve, an arbitrary parameter interval  $[a, b]$  is used and partitioned into  $r$  intervals  $[t_{j-1}, t_j], j \in \{1, \dots, r\}$  in accordance to  $a = t_0 < t_1 < \dots < t_r = b$ .

In this way,  $r$  Bézier segments of the type

$$C_j : [t_{j-1}, t_j] \longrightarrow \mathbb{R}$$

$$t \longmapsto C_j(t) = \sum_{i=0}^n b_{ij} \cdot \mathcal{BS}_i^n(t; t_{j-1}, t_j), j \in \{1, \dots, r\}$$

are constructed by the algorithm of de Casteljau.

In the following description,  $j \in \{1, \dots, r\}$  is valid until there is written anything else.

As described above, the  $j$ -th and the  $(j + 1)$ -th Bézier segments are  $\mathcal{C}^0$ -continuous at the point  $t_j \in [t_{j-1}, t_j]$  respectively  $t_j \in [t_j, t_{j+1}]$ , if the following condition is valid:

$$b_{n,j} = b_{0,j+1}.$$

According to that,  $\mathcal{C}^1$ -continuity is given at the regarded point  $t_j$ , if

$$\frac{d}{dt} C_j(t_j) = \frac{d}{dt} C_{j+1}(t_j) \text{ for } t_j \in [t_{j-1}, t_j] \text{ respectively } t_j \in [t_j, t_{j+1}].$$

This coherence is transformed into an expression of Bézier segments according to the standard interval  $[0, 1]$  by using the following mapping  $\lambda$  already mentioned in the proof of theorem 8.1.1:

$$\lambda : [t_{j-1}, t_j] \longrightarrow [0, 1]$$

$$t \longmapsto \lambda(t) := \frac{t - t_{j-1}}{t_j - t_{j-1}}$$

## 8 Coherences of approximation theory

In addition to that, the abbreviation  $\forall_{j \in \{1, \dots, r\}} \Delta_j := t_j - t_{j-1}$  is used.

$$\begin{aligned}
& \frac{d}{dt} C_j(t_j) = \frac{d}{dt} C_{j+1}(t_j) \\
\Leftrightarrow & \frac{d}{dt} \sum_{i=0}^n b_{ij} \cdot \mathcal{BS}_i^n(t_j; t_{j-1}, t_j) = \frac{d}{dt} \sum_{i=0}^n b_{ij+1} \cdot \mathcal{BS}_i^n(t_j, t_j, t_{j+1}) \\
\Leftrightarrow & \frac{d}{dt} \sum_{i=0}^n b_{ij} \cdot \mathcal{BS}_i^n\left(\frac{t_j - t_{j-1}}{t_j - t_{j-1}}\right) = \frac{d}{dt} \sum_{i=0}^n b_{ij+1} \cdot \mathcal{BS}_i^n\left(\frac{t_j - t_j}{t_{j+1} - t_j}\right) \\
& \Leftrightarrow \frac{d}{dt} \sum_{i=0}^n b_{ij} \cdot \mathcal{BS}_i^n(\lambda(t_j)) = \frac{d}{dt} \sum_{i=0}^n b_{ij+1} \cdot \mathcal{BS}_i^n(\lambda(t_j)) \\
\Leftrightarrow & \frac{d}{d\lambda} \sum_{i=0}^n b_{ij} \cdot \mathcal{BS}_i^n(\lambda(t_j)) \cdot \frac{d}{dt} \lambda(t_j) = \frac{d}{d\lambda} \sum_{i=0}^n b_{ij+1} \cdot \mathcal{BS}_i^n(\lambda(t_j)) \cdot \frac{d}{dt} \lambda(t_j) \\
\Leftrightarrow & \frac{d}{d\lambda} \sum_{i=0}^n b_{ij} \cdot \mathcal{BS}_i^n(\lambda(t_j)) \cdot \frac{1}{t_j - t_{j-1}} = \frac{d}{d\lambda} \sum_{i=0}^n b_{ij+1} \cdot \mathcal{BS}_i^n(\lambda(t_j)) \cdot \frac{1}{t_{j+1} - t_j} \\
& \Leftrightarrow n \cdot (b_{n,j} - b_{n-1,j}) \cdot \frac{1}{t_j - t_{j-1}} = n \cdot (b_{1,j+1} - b_{0,j+1}) \cdot \frac{1}{t_{j+1} - t_j}
\end{aligned}$$

This leads to the following coherence:

$$\begin{aligned}
b_{n,j} &= \frac{t_{j+1} - t_j}{t_j - t_{j-1} + t_{j+1} - t_j} \cdot b_{n-1,j} + \frac{t_j - t_{j-1}}{t_j - t_{j-1} + t_{j+1} - t_j} \cdot b_{1,j+1} \\
\Leftrightarrow b_{n,j} &= \frac{\Delta_{j+1}}{\Delta_j + \Delta_{j+1}} \cdot b_{n-1,j} + \frac{\Delta_j}{\Delta_j + \Delta_{j+1}} \cdot b_{1,j+1} \tag{8.1}
\end{aligned}$$

In a geometric interpretation (see [Sch97], p.171), this means, that the Bézier-point  $b_{n,j}$  has to divide the line segment between  $b_{n-1,j}$  and  $b_{1,j+1}$  in the ratio  $\Delta_j : \Delta_{j+1}$ . For getting  $\mathcal{C}^2$ -continuity at the regarded point  $t_j$ , the following coherences and transformations are necessary. Thereby, the abbreviation  $\eta_j := \frac{\Delta_{j+1}}{\Delta_j}$  is used.

## 8 Coherences of approximation theory

Using 8.1, the Bézier-point  $b_{1,j+1}$  is expressed by:

$$\begin{aligned}
 b_{n,j} &= \frac{\Delta_{j+1}}{\Delta_j + \Delta_{j+1}} \cdot b_{n-1,j} + \frac{\Delta_j}{\Delta_j + \Delta_{j+1}} \cdot b_{1,j+1} \\
 \Leftrightarrow b_{1,j+1} &= \left(1 + \frac{\Delta_{j+1}}{\Delta_j}\right) \cdot b_{n,j} - \frac{\Delta_{j+1}}{\Delta_j} \cdot b_{n-1,j} \\
 \Leftrightarrow b_{1,j+1} &= (1 + \eta_j) \cdot b_{n,j} - \eta_j \cdot b_{n-1,j} \tag{8.2}
 \end{aligned}$$

With the prerequisite of the necessary  $\mathcal{C}^0$ -continuity at the regarded point  $t_j$ , 8.2 and 8.2.4, the following is valid:

$$\begin{aligned}
 \frac{d}{d^2t} C_j(t_j) &= \frac{d}{d^2t} C_{j+1}(t_j) \\
 \Leftrightarrow \frac{n \cdot (n-1) (b_{n,j} - 2 \cdot b_{n-1,j} + b_{n-2,j})}{\Delta_j^2} &= \frac{n \cdot (n-1) (b_{2,j+1} - 2 \cdot b_{1,j+1} + b_{0,j+1})}{\Delta_{j+1}^2}
 \end{aligned}$$

This leads to the following:

$$\begin{aligned}
 b_{2,j+1} &= 2 \cdot b_{1,j+1} - b_{n,j} + \eta_j^2 (b_{n,j} - 2b_{n-1,j} + b_{n-2,j}) \\
 &= (1 + 2\eta_j + \eta_j^2) b_{n,j} - 2\eta_j (1 + \eta_j) b_{n-1,j} + \eta_j^2 b_{n-2,j} \\
 &= (1 + \eta_j) [(1 + \eta_j) b_{n,j} - \eta_j b_{n-1,j}] - \eta_j [(1 + \eta_j) b_{n-1,j} - \eta_j b_{n-2,j}] \\
 &\stackrel{8.2}{=} (1 + \eta_j) b_{1,j+1} - \eta_j H_j
 \end{aligned}$$

There, the aiding point  $H_j$  was defined as  $H_j := [(1 + \eta_j) b_{n-1,j} - \eta_j b_{n-2,j}]$ .

This means, that the Bézier-point  $b_{2,j+1}$  can be expressed by a so-called *barycentric combination* (see section 8.3) of the two points  $b_{1,j+1}$  and  $H_j$ .

This way of utilize geometric coherences leads to the ability of theoretically getting  $\mathcal{C}^k$ -continuous,  $k \in \mathbb{N}_0$  in a regarded Bézier-point.

## 8 Coherences of approximation theory

In the following section, this geometric approach of interpolation is extended regarding the parametric description of surfaces basing on rectangular and triangular sets of parameters. On the last case there will be special attention.

## 8.3 The Bézier interpolation approach describing surfaces

### 8.3.1 The bilinear tensor product approach

In the last section, the interpolation of discrete data by parameter Bézier-curves was looked at. Doing this, one gets a possibility of giving statements along the resulting curve. Hence, if the intention of dealing with the given data is not of that kind but of getting “areal statements”, the method of interpolation and processing the data has to be modified.

So in this section, two methods of areal interpolation are discussed. These are both basing on the geometric idea of Bézier-interpolation. For interpolating curves, a onedimensional interval of parameters is satisfying, but the interpolation of surfaces does require a parameter space of higher dimension.

The first approach shortly described in this subsection bases on a two-dimensional space of parameters. There will be more attention to the second approach using a three-dimensional space of parameters in the next subsection.

In the twodimensional case, the construction of so-called *tensorproduct surfaces* is used. So in the beginning, this term and before, the term *tensorproduct on functional vector spaces* are defined basing on [Aum93], p.477-479.

**Definition 8.3.1 (Tensorproduct on functional vector spaces)**

*Let  $V$  and  $W$  be two vectorspaces, which elements are functions of type*

*$f : \mathbb{R} \longrightarrow \mathbb{R}$ . Additionally, let the particular basis be  $B_V := \{f_0, \dots, f_m\}$  and  $B_W := \{f_0, \dots, f_n\}$ . The vectorspace  $V \otimes W$  of dimension  $(n + 1) \cdot (m + 1)$  with the following basis:*

$$B_{V \otimes W} := \{t_{00}, \dots, t_{nm}\},$$

$$t_{ij} : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

$$(u, v) \mapsto f_i(u) g_j(v)$$

is called **tensorproduct** of  $V$  and  $W$ .

**Definition 8.3.2 (Tensorproduct surface)**

Let  $V$  and  $W$  be two vectorspaces, which elements are functions of type

$f : \mathbb{R} \longrightarrow \mathbb{R}$ . Additionally, let the particular basis be  $B_V := \{f_0, \dots, f_m\}$  and  $B_W := \{g_0, \dots, g_n\}$  and the vectorspace  $V \otimes W$  of dimension  $(n + 1) \cdot (m + 1)$  be the **tensorproduct** of  $V$  and  $W$ .

Furthermore, let  $\forall \substack{i \in \{0, \dots, n\} \\ j \in \{0, \dots, m\}} : b_{ij} \in \mathbb{R}^d, d \in \mathbb{N}$

Then, the following mapping  $\Phi$  with

$$\Phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^d$$

$$(u, v) \mapsto \Phi((u, v)) := \sum_{i=0}^n \sum_{j=0}^m b_{ij} f_i(u) g_j(v)$$

is called **tensorproduct surface** of  $V, W$  and  $b_{ij} \in \mathbb{R}^d, d \in \mathbb{N}, i \in \{0, \dots, n\}, j \in \{0, \dots, m\}$ .

Regarding the vectorspace of polynomials with grade  $n_0 \leq n$  respectively  $m_0 \leq m$ , the Bernstein-polynomials  $\mathcal{BS}_i^n, i \in \{0, \dots, n\}$  respectively  $\mathcal{BS}_j^m, j \in \{0, \dots, m\}$  define a basis.

Using this, the interpolation basing on  $(n + 1) \cdot (m + 1)$  points  $b_{ij} \in \mathbb{R}^d, d \in \mathbb{N}, i \in \{0, \dots, n\}, j \in \{0, \dots, m\}$  by the mapping

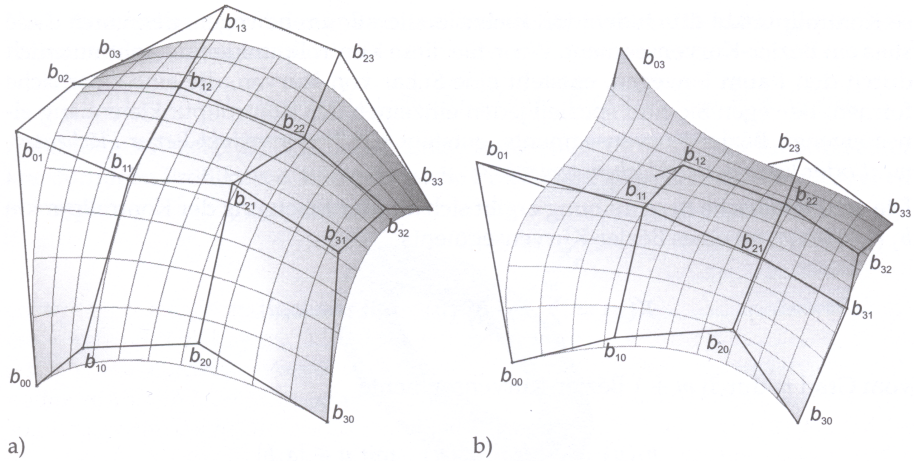
$$\Phi_{\mathcal{BS}} : [0, 1] \times [0, 1] \longrightarrow \mathbb{R}^d$$

$$(u, v) \mapsto \Phi_{\mathcal{BS}}((u, v)) := \sum_{i=0}^n \sum_{j=0}^m b_{ij} \mathcal{BS}_i^n(u) \mathcal{BS}_j^m(v)$$

defines a so-called **tensorproduct Bézier surface**. According to [Hos89], p.220 et seqq., this can be interpreted as a transformation of a Bézier-curve  $P = \sum_{i=0}^n c_i \cdot \mathcal{BS}_i^n$  described by another Bézier-curve

$$Q = \sum_{j=0}^m b_{ij} \cdot \mathcal{BS}_j^m$$





**Figure 8.2:** Two tensorproduct Bézier surfaces with according Bézier-grid ([Ben03], p.172)

The regarded coefficients  $b_{ij} \in \mathbb{R}^d$ ,  $d \in \mathbb{N}$ ,  $i \in \{0, \dots, n\}$ ,  $j \in \{0, \dots, m\}$  then define a so-called *Bézier-polyeder* or *Bézier-grid*. Two examples of such tensor-product Bézier surfaces are illustrated in figure 8.2.

Any sequence  $(b_{ij})_{i \in \{0, \dots, n\}}$  or  $(b_{ij})_{j \in \{0, \dots, m\}}$  with either  $i = \text{const.}$  or  $j = \text{const.}$  is called *Bézier-filament*.

Focussing the aim of proposing generalised statements for the interpolation within arbitrary topological vector spaces, the term *topological tensorproduct Bézier surface* is defined. This generalization makes sense, because as described after that definition, the according calculation algorithm bases on the coherences made in section 8.2.

**Definition 8.3.3 (Topological tensorproduct Bézier surface)**

Let  $X$  be an arbitrary topological vector space and  $\mathcal{TS} := \{b_{0,0}, \dots, b_{n,m}\} \subset X$  be a discrete subset of  $X$  with  $(n + 1) \cdot (m + 1)$  elements. Then, the **topological**

**tensorproduct Bézier surface**  $TBS$  of  $\mathcal{TS}$  using the parameter space  $[0, 1] \times [0, 1]$  is the following. Thereby, the Bernstein-polynomials  $\mathcal{BS}_i^n$ ,  $i \in \{0, \dots, n\}$  respectively  $\mathcal{BS}_j^m$ ,  $j \in \{0, \dots, m\}$  are used:

$$TBS : [0, 1] \times [0, 1] \longrightarrow \mathbb{R}^d$$

$$(u, v) \longmapsto \Phi_{BS}((u, v)) := \sum_{i=0}^n \sum_{j=0}^m b_{ij} \mathcal{BS}_i^n(u) \mathcal{BS}_j^m(v)$$

A value  $TBS((u_0, v_0))$  can be calculated using the algorithm of Casteljau in two ways. On the one hand side, the  $(n + 1)$  Bézier-points  $b_i = \sum_{j=0}^m b_{ij} \cdot \mathcal{BS}_j^m(v_0)$  and then the point  $TBS((u_0, v_0))$  on the surface is calculated by:

$$TBS((u_0, v_0)) = \sum_{i=0}^n b_i \cdot \mathcal{BS}_i^n.$$

On the other hand side, one can get the  $(m + 1)$  Bézier-points  $b_j = \sum_{i=0}^n b_{ij} \cdot \mathcal{BS}_i^n(u_0)$  and then the point  $TBS((u_0, v_0))$  by:

$$TBS((u_0, v_0)) = \sum_{j=0}^m b_j \cdot \mathcal{BS}_j^m.$$

Similar to Bézier-curves, there can be calculated compositions of tensorproduct Bézier surfaces as well. There, it is also possible to have  $\mathcal{C}^k$ -continuity for  $k \in \mathbb{N}_0$  in a regarded Bézier-point at the connection of two surfaces.

This coherences can be found in [Aum93], pp.481-490 or in [Hos89], pp.231-235. for coefficients  $b_{ij} \in \mathbb{R}^d$ ,  $d \in \mathbb{N}$ ,  $i \in \{0, \dots, n\}$ ,  $j \in \{0, \dots, m\}$ .

### 8.3.2 The approach based on triangular surfaces

In graphic software, in engineering regarding the *finite-element-method* (e.g. refer to [Kle07]), but especially also in Geographical Information System, complex structures are often described by subdividing them using *triangles*.

Regarding the characteristics of the used triangles and their influences on the whole system, it is useful to define a locale coordinate system. The following description of this coherence is based on [Far03], p.137 et seqq.

At this point, the geometric concept of a triangle is regarded in a more generalised way using the term *n-dimensional simplex*. The definition is based on [Dir96], p.157.

**Definition 8.3.4 (n-dimensional simplex)**

Let  $A$  be an affine space of dimension  $m$  and  $\{P_0, \dots, P_n\}$  be a set of  $n+1$  elements out of  $A$  with  $n \leq m$ . Then, a  $n$  dimensional simplex  $\boxplus$  according to this points is defined as the convex hull  $\text{conv}(\{P_0, \dots, P_n\})$ , if the  $n$  vectors  $\overrightarrow{P_0P_i}$ ,  $i \in \{1, \dots, n\}$  are linearly independent.

According to this definition, a triangle is a 2 dimensional Simplex.

**Definition 8.3.5 (Barycentric coordinates)**

Let  $A$  be an affine space of dimension  $m$  and  $\{P_0, \dots, P_n\}$  be a set of  $n+1$  elements out of  $A$  with  $n \leq m$ , which define a  $n$ -dimensional simplex. Then, a tuple  $(a_0, \dots, a_n) \in \mathbb{R}^n$  is called **barycentric coordinate realization** of a  $P \in A$ , if and only if the following equation is valid:

$$(a_0, \dots, a_n) \cdot P = a_0 \cdot P_0 + \dots + a_n \cdot P_n$$

This realization is not unique. To get uniqueness, these coordinates has to be normed. So, a tuple  $(\lambda_0, \dots, \lambda_n) \in [0, 1]^{n+1}$  is called **normed barycentric coordinate realization** of a  $P \in A$ , if and only if the following equation:

$$(\lambda_0, \dots, \lambda_n) \cdot P = \lambda_0 \cdot P_0 + \dots + \lambda_n \cdot P_n$$

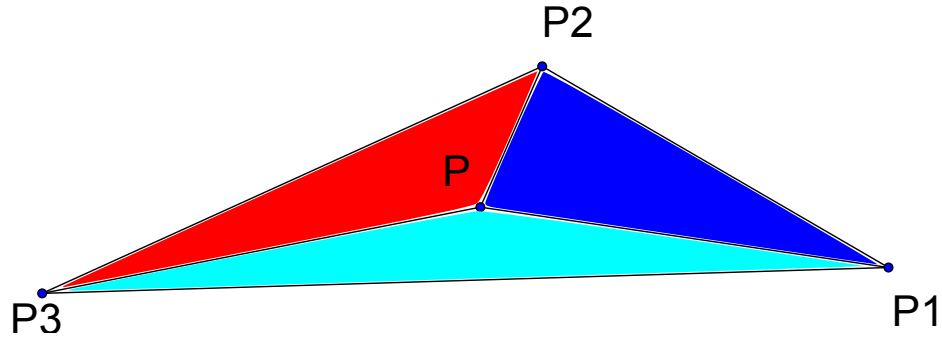
and the proposition

$$\sum_{i=0}^n \lambda_i = 1$$

are valid. The process of norming is given by the following:

$$\forall_{i \in \{0, \dots, n\}} \lambda_i = \frac{a_i}{\sum_{i=0}^n a_i}$$

Due to this definition, in definition 8.2.3,  $(\mathcal{BS}_0^n(t), \dots, \mathcal{BS}_n^n(t))$  is a normed barycentric coordinate realization of the according Bézier-curve.



**Figure 8.3:** Barycentric coordinates of a point  $P$

This idea of barycentric coordinates goes back to August F. Möbius, who published this 1827 in [Möb27]. Möbius used the term *barycentric* because of the following fact: If there is a weight of mass  $\lambda_i$  put on each vertex  $P_i$  of the regarded simplex, then the according point  $P$  is the *barycentre* (i.e. the centre of gravity) of this simplex.

In figure 8.3, an exemplary triangle with vertices  $P_1$ ,  $P_2$  and  $P_3$  is illustrated. The according barycentric coordinates of the shown point  $P$  can be calculated areal relations in the following way according to [Far03], p. 139:

$$\begin{aligned}\lambda_1 &= \frac{A(P, P_2, P_3)}{A(P_1, P_2, P_3)} \\ \lambda_2 &= \frac{A(P, P_3, P_1)}{A(P_1, P_2, P_3)} \\ \lambda_3 &= \frac{A(P, P_1, P_2)}{A(P_1, P_2, P_3)}\end{aligned}$$

The aim of this subsection is to describe surface interpolation using the Bézier-technique with a triangular space of parameters. So, Bernstein polynomials using barycentric coordinates have to be used and defined:

**Definition 8.3.6 (Barycentric Bernstein polynomials)**

Let  $i, j, k \in \mathbb{N}_{\leq n}$ ,  $i + j + k = n$  with  $n \in \mathbb{N}$ . Furthermore, let  $\mathbb{D}_{n+1}$  be a  $(n + 1)$ -dimensional set of normed barycentric coordinates with:

## 8 Coherences of approximation theory

$$\forall_{n \in \mathbb{N}} \mathbb{D}_{n+1} := \left\{ (\lambda_0, \dots, \lambda_n) \in [0, 1]^{n+1} \mid \sum_{k=0}^n \lambda_k = 1 \right\}$$

Then, the **barycentric Bernstein polynomials** of grade  $n$  with index  $(i, j, k)$  are defined by:

$$\begin{aligned} \mathcal{BS}_{(i,j,k)}^n : \quad \mathbb{D}_3 &\longrightarrow [0, 1] \\ (u, v, w) &\mapsto \mathcal{BS}_{i,j,k}^n((u, v, w)) := \frac{n!}{i!j!k!} u^i v^j w^k \end{aligned}$$

Since  $(u + [v + w])^n = \sum_{i=0}^n \binom{n}{i} u^i (v + w)^{n-i}$  and  $(v + w)^{n-i} = \sum_{j=0}^{n-i} \binom{n-i}{j} v^j w^{n-i-j}$  is valid, the following chain of equations shows, that those barycentric Bernstein polynomials define a partition of 1:

$$1 = (u + v + w)^n = \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} \frac{n!}{i!j!k!} u^i v^j w^k$$

### Remark 8.3.1 (Abbreviative notation)

For better clearness, the following abbreviations are used for the notation of barycentric Bernstein polynomials:

$$\forall_{\substack{n \in \mathbb{N} \\ i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} : I := (i, j, k)$$

$$\forall_{(u,v,w) \in \mathbb{D}_3} : U := (u, v, w)$$

Hence, with this, the value of a barycentric Bernstein polynomial of grade  $n$  at the point  $(u, v, w) \in \mathbb{D}_3$  can be noted in the following way:

$$\mathcal{BS}_I^n(U) = \frac{n!}{i!j!k!} u^i v^j w^k$$

Additionally,  $|I| := i + j + k$  and  $|U| := u + v + w$  are defined.

### Remark 8.3.2 (Coherence to Bernstein polynomials)

Using definition 8.3.6, the occurrence of “normal” Bernstein polynomials used in

the sections 8.1 and 8.2 at the “margins” of the regarded triangle can be seen. E.g., the following is valid:

$$\mathcal{BS}_{0,j,k}^n((0, v, w)) = \mathcal{BS}_j^n(v) = \mathcal{BS}_k^n(w)$$

In the following, the term *triangular Bézier surface* is defined and regarded further.

**Remark 8.3.3**

Let  $A := \{b_{(i,j,k)} \in \mathbb{R}^d \mid d \in \mathbb{N}, i, j, k \in \mathbb{N}_{\leq n}, i + j + k = n\}$  be an arbitrary discrete set in the space  $\mathbb{R}^d$ . According to [Hos89], p.245,  $|A| = \binom{n+2}{2} = \frac{1}{2}(n+1)(n+2)$  is valid.

**Definition 8.3.7 (Triangular Bézier surface)**

Let  $SM := \{b_{(i,j,k)} \in \mathbb{R}^d \mid d \in \mathbb{N}, i, j, k \in \mathbb{N}_{\leq n}, i + j + k = n\}$  be an arbitrary discrete set in the space  $\mathbb{R}^d$ .

Then, the so-called **triangular Bézier surface**  $CT$  according to  $SM$  using normed barycentric coordinates is defined in the following way:

$$CT = \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)} \cdot \mathcal{BS}_{(i,j,k)}^n, \text{ with } \forall_{(i,j,k) \in \mathbb{N}_{\leq n}^3} b_{(i,j,k)} \in \mathbb{R}^d \text{ and}$$

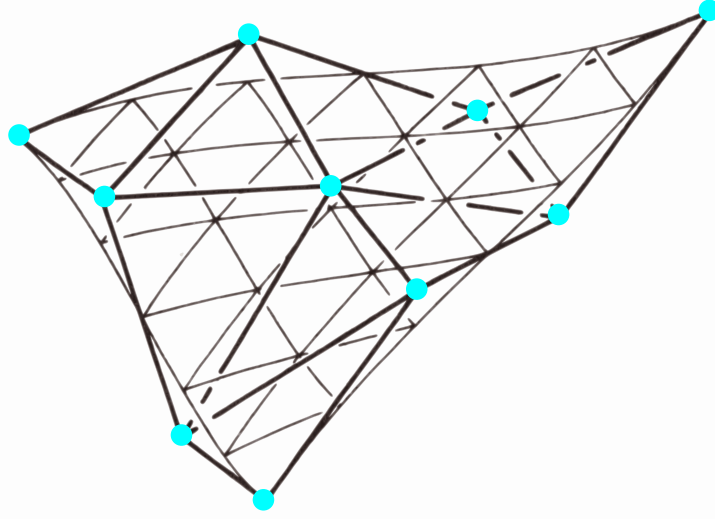
$$\forall_{(u,v,w) \in \mathbb{D}_3} CT((u, v, w)) = \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)} \cdot \mathcal{BS}_{(i,j,k)}^n((u, v, w))$$

The  $b_{(i,j,k)} \in \mathbb{R}^d$ ,  $(i, j, k) \in \mathbb{N}_{\leq n}^3$  are called **controlling- or Bézier-points**. They define a so-called *triangular Bézier-polyeder* or *triangular Bézier-grid*.

**Remark 8.3.4**

The above defined triangular Bézier surface  $CT$  represents the subset of the graph of a multidimensional polynomial  $P_{d-1}^n$  of maximum grade  $n \in \mathbb{N}$  in a certain triangle restricted by the set  $SM$ . In figure 8.4, there is exemplarily shown such a triangular Bézier surface of grade  $n = 3$ .

In terms of generalizing this coherences with regard to the application proposed in section 9.3 of chapter 9, an arbitrary topological vector space  $(X, \mathcal{O})$  is regarded



**Figure 8.4:** Triangular Bézier surface of grade  $n = 3$  and according triangular Bézier-grid, ([Hos89], p. 247, modified by the author)

instead of only  $\mathbb{R}^d$  as done in common literature. So, the term *topological triangular Bézier surface* is defined first:

**Definition 8.3.8 (Topological triangular Bézier surface)**

Let  $(X, \mathcal{O})$  an arbitrary topological vector space.

Further, let  $\mathcal{BTM} := \{b_{(i,j,k)} \in X \mid i, j, k \in \mathbb{N}_{\leq n}, i + j + k = n\}$  be an arbitrary discrete set in the space  $X$ .

Then, the so-called **topological triangular Bézier surface**  $ST$  according to  $\mathcal{BTM}$  using normed barycentric coordinates is defined in the following way:

$$ST = \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)} \cdot \mathcal{BS}_{(i,j,k)}^n, \text{ with } \forall_{(i,j,k) \in \mathbb{N}_{\leq n}^3} b_{(i,j,k)} \in X \text{ and}$$

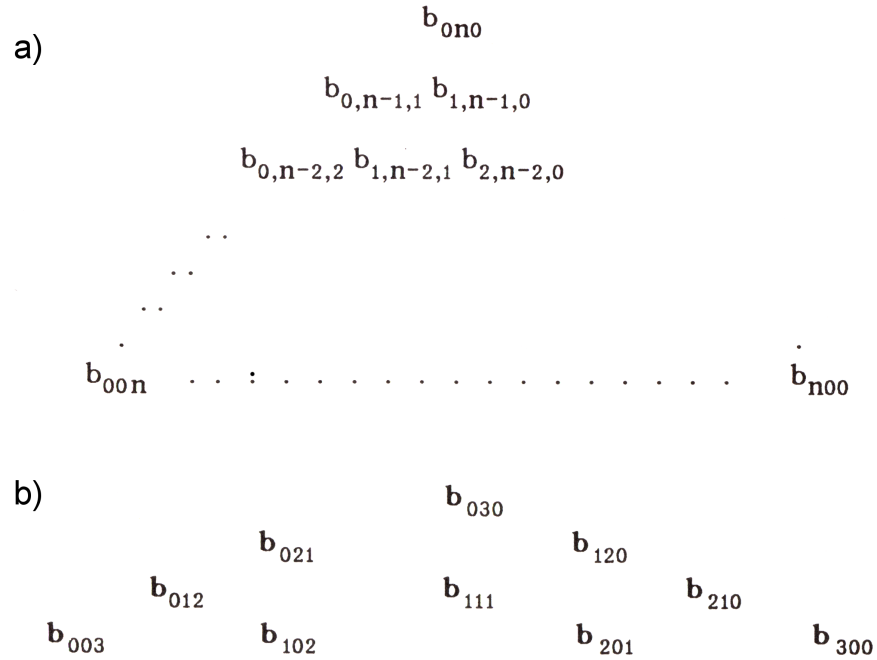
$$\forall_{(u,v,w) \in \mathbb{D}_3} ST((u,v,w)) = \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)} \cdot \mathcal{BS}_{(i,j,k)}^n((u,v,w))$$

The  $b_{(i,j,k)} \in X, (i,j,k) \in \mathbb{N}_{\leq n}^3$  are called **controlling- or Bézier-points**. They define a so-called *triangular Bézier-polyeder* or *triangular Bézier-grid*.

8 Coherences of approximation theory

In the further description, this definition is used as main concept.

In figure 8.5, the topological order of Bézier-points is illustrated and in figure 8.6, the according parameters are shown.



**Figure 8.5:** Topological order of Bézier-points: a)  $n \in \mathbb{N}$ , b)  $n = 3$ , ([Hos89], p. 246)

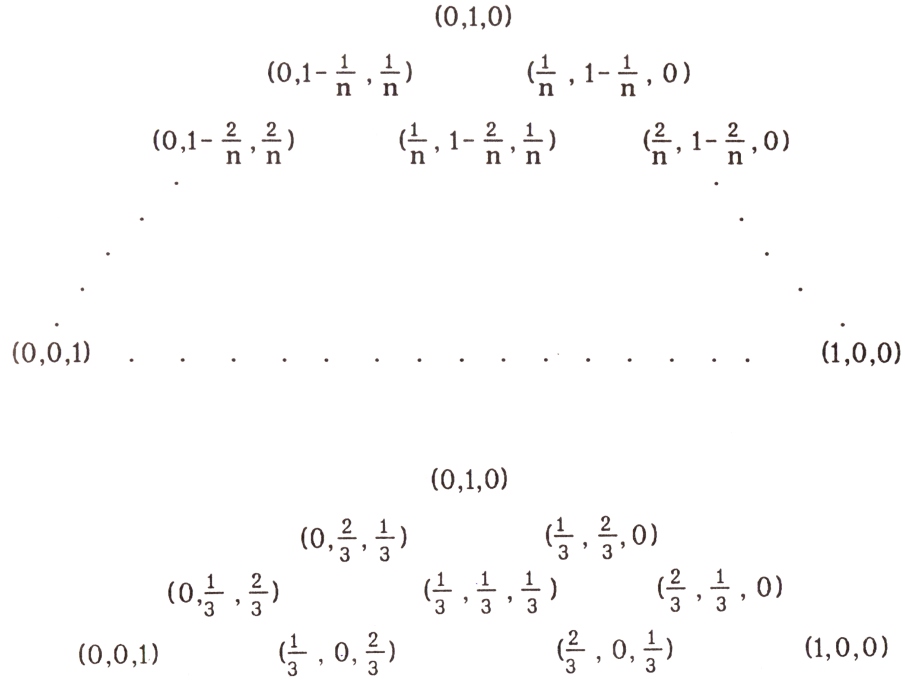
For the calculation of a certain value  $ST(U)$ , the following recursion is given basing on [Hos89], p.249. (there done for the space  $\mathbb{R}^d, d \in \mathbb{N}$ ).

**Remark 8.3.5 (Casteljau recursion concerning triangular Bézier surfaces)**

Let  $i, j, k \in \mathbb{N}_{\leq n}, i + j + k = n$  with  $n \in \mathbb{N}$  and  $U \in \mathbb{D}_3$ . Further, let  $b_I^0 := b_I$  and  $b_I^n(U) := ST(U)$ , while  $ST$  is a topological triangular Bézier surface. Then, the **Casteljau recursion concerning triangular Bézier surfaces** is given by the following:

$$b_I^{r+1} = u \cdot b_{I+e_1}^r + v \cdot b_{I+e_2}^r + w \cdot b_{I+e_3}^r,$$





**Figure 8.6:** Topological order of parameters: a)  $n \in \mathbb{N}$ , b)  $n = 3$ , ([Hos89], p. 247)

with  $|I| := n - r - 1$ ,  $r \in \mathbb{N}_{\leq n}$  and  $e_1 := (1, 0, 0)$ ,  $e_2 := (0, 1, 0)$ ,  $e_3 := (0, 0, 1)$ .

**Algorithm 8.3.1 (Triangular algorithm of de Casteljau)**

Let  $\mathcal{BTM} := \{b_{(i,j,k)} \in X \mid i, j, k \in \mathbb{N}_{\leq n}, i + j + k = n\}$  be an arbitrary discrete set in the space  $X$  with  $\binom{n+2}{2}$  elements and topological triangular Bézier surface  $\mathcal{ST}$  using normed barycentric coordinates. Then, a value  $\mathcal{ST}(U) = b_I^n(U)$  for an arbitrary element  $U \in \mathbb{D}_3$  can be calculated by the following algorithm:

For  $r = 0, \dots, n - 1, |I| := n - r - 1$ :

$$b_I^{r+1} = u \cdot b_{I+e_1}^r + v \cdot b_{I+e_2}^r + w \cdot b_{I+e_3}^r;$$

$$\mathcal{ST}(U) := b_I^n(U)$$

In the following chapter, it is necessary, to compute topological triangular Bézier surfaces of grade  $n = 3$  starting with **three** elements out of the regarded topological

vector space. According to 8.3.3, there are needed ten elements to describe such a surface. To compute the missing ones using convex combinations of the given ones, there exists an *algorithm of grade elevation* as described in [Hos89], p.248.

**Algorithm 8.3.2 (Triangular Bézier grade elevation)**

Let  $\mathcal{ST} = \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)} \cdot \mathcal{BS}_{(i,j,k)}^n$ , with  $\forall_{(i,j,k) \in \mathbb{N}_{\leq n}^3} b_{(i,j,k)} \in X$ . be a topological triangular Bézier surface of grade  $n \in \mathbb{N}$ . Then, the Bézier-points  $b_{(i,j,k)}^*$  of the derived topological triangular Bézier surface of grade  $n + 1$  can be calculated in the following way:

$$b_{(i,j,k)}^* = \frac{i}{n+1}b_{(i-1,j,k)} + \frac{j}{n+1}b_{(i,j-1,k)} + \frac{k}{n+1}b_{(i,j,k-1)}$$

Thereby,  $b_{(i,j,k)} := \mathfrak{D}_X$ , if  $i \notin [0, 1] \vee j \notin [0, 1] \vee k \notin [0, 1]$  by this iteration.

In figure 8.7, the processing of that algorithm is illustrated.

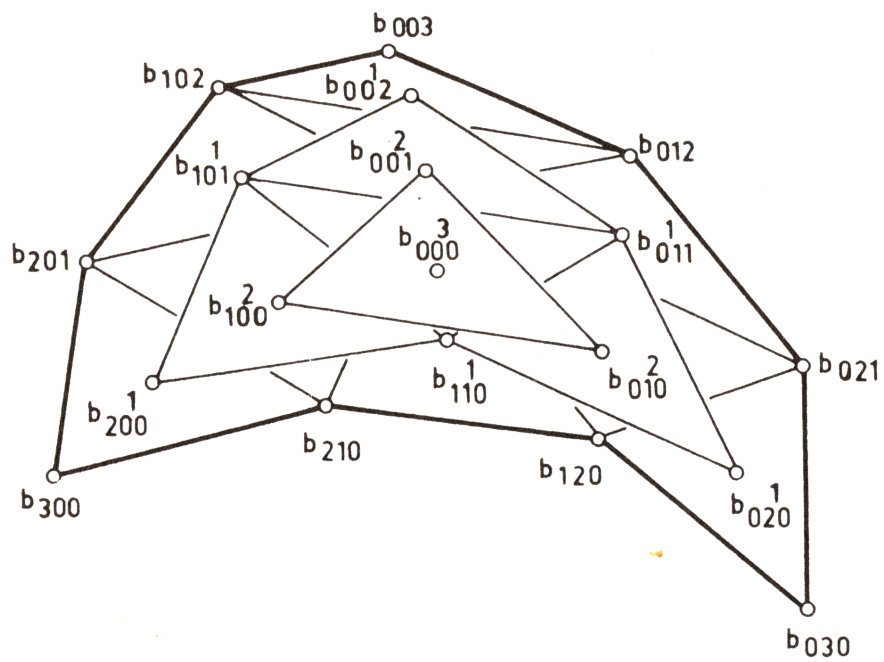


Figure 8.7: Triangular Casteljau algorithm ([Hos89], p.250)

# 9 Results of measure theory applied to the “Spatial-Toolbox”

## 9.1 Fundamental considerations

When dealing with spatial data, *discrete* and *non-discrete* information can be available and be important. So, it is important to be able to deal with this data in an adequate way. The SPATTB already mentioned in chapter 2 and in chapter 5, is supposed to make a contribution to that.

In this chapter, the focus is straightened on certain aspects of this SPATTB more detailed in a generalised way. Thereby, especially mathematization is looked at.

There should be regarded functionally associated geocoded data. The aim is to describe this data by appropriate mappings. The solution for this situation proposed in this thesis is a *surface interpolation of mappings with preprocessing of the given data*. Therefor, the algorithm of the Backpropagation Networks described in chapter 4, the Bézier-methods in section 8.3 from the field of computeroriented geometry and the concepts of arbitrary mathematical spaces and measurement out of chapter 7 are used.

Two main steps are suggested in this context:

- 1. Processing an approximation of the functional coherences by the given data at the particular coordinates**
- 2. Interpolation of these mappings by methods from computer-oriented geometry**

In the following, these two steps are discussed in separate sections of this chapter. In these sections, the particular problem is viewed in detail. In doing so, the applicability of the methods in the proposed concept and relevant relations to conventual methods are described.

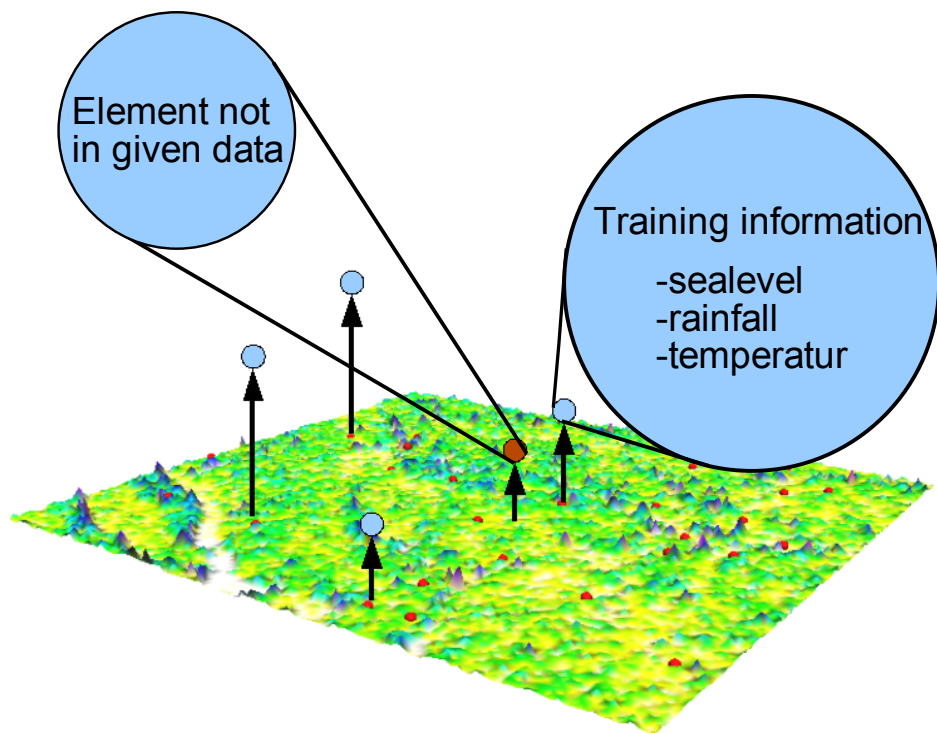
## 9.2 Processing an approximation of the functional coherences by the given data at the particular coordinates

As described in chapter 4, the classical Backpropagation-algorithm can be used for approximating functional relationships of the type  $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ ,  $n, m \in \mathbb{N}$ . The approximation using this algorithm bases on appropriate training- and test-data. Hence, to every iteration, there exists one mapping of the mentioned type representing the actual state of training. The characteristic *fitting* of test- and training-data means, that there must be input and nominal output data. Thereby, the input data has to consists of  $n$ -dimensional vectors, the nominal output data of  $m$ -dimensional ones.

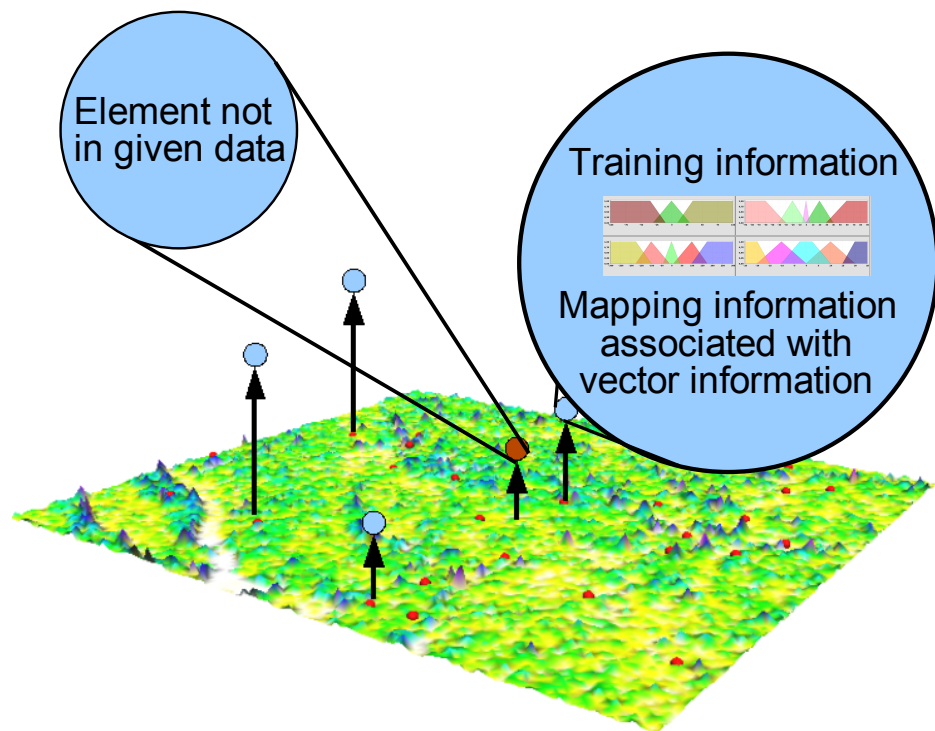
Dealing with spatial data, not only this kind of data is given in generally. This method matches exemplarily for measured temperature at certain coordinates. In this example, the input data is given by a two- or three-dimensional vector describing the according coordinates and the nominal output is given by the value of temperature. In this case, the classical algorithm can be considered for interpolating these measured data directly, because only discrete information are regarded. This situation is illustrated in figure 9.1.

Suppositionally, the data is not discrete, but for example, for certain areals of the regarded geographical area the amount of pollutants is known by measurement. Basing on this data, statements about other areals not contained in the given data should be produced. In this case, the classical algorithm cannot be used in a direct way.

Another possible application that is looked at more detailed in this thesis, is given by the coherence of mappings as input data itself and according vector data as



**Figure 9.1:** Discrete data for processing by BPN



**Figure 9.2:** Mapping data for processing by BPN

output. For example,  $k$  membership functions representing a particular fuzzy set of ecological factors are associated with a vector of dimension  $r$ . Thereby, the components of that vector represent an estimate number of certain insects using disease vectors. This coherence is illustrated in figure 9.2.

This examples illustrate, that there is a necessity of combining the generalised perspective given by using topological vector spaces and algorithm based methods like the BPN algorithm, that are not able to handle that kind of spatial problem in the classical way.

In this thesis, there is proposed a possible dealings with this problem as a solution that can be implemented as a part of the suggested “Spatial-Toolbox”.

So within this chapter, the mathematical analysis part of the implementation



process proposed in chapter 6 is regarded.

First of all, the structure of the regarded data is defined more precise. In general, the input information should be given by subsets or elements of a topological vector space. Depending on the processing method, a certain level of integrability is demanded for those elements. This is specified more precisely at a suitable point of the description.

In the following, the terms *set based dataset* and *function based dataset* are defined.

**Definition 9.2.1 (Set based and function based dataset)**

Let  $(X, \mathcal{O})$  be an arbitrary topological space. Furthermore, let  $S_i \subset X$  and  $M_i \in \mathbb{R}^d$  for  $i \in \{1, \dots, k\}$ ,  $k, d \in \mathbb{N}$ . Then, the following set is called **set based dataset**:

$$\mathfrak{D}_S := \{(S_1, M_1), \dots, (S_k, M_k)\}$$

Let  $(\mathfrak{F}, \mathcal{O})$  be an arbitrary topological space, whose elements are functions. Furthermore, let  $F_i \in X$  and  $Y_i \in \mathbb{R}^d$  for  $i \in \{1, \dots, k\}$ ,  $k, d \in \mathbb{N}$ . Then, the following set is called **function based dataset**:

$$\mathfrak{D}_F := \{(F_1, Y_1), \dots, (F_k, Y_k)\}$$

For the description of a coherence between a number of such mappings and a vector, the term **multidimensional function based dataset** is used:

Let  $(\mathfrak{F}, \mathcal{O})$  be an arbitrary topological space, whose elements are functions. Furthermore, let  $F_{ij} \in X$  and  $Y_i \in \mathbb{R}^d$  for  $i \in \{1, \dots, k\}$ ,  $j \in \{1, \dots, r\}$ ,  $k, d \in \mathbb{N}$ . Then, the following set is called **multidimensional function based dataset**:

$$\mathfrak{D}_{MF} := \left\{ \left( \begin{pmatrix} F_{11} \\ \vdots \\ F_{1r} \end{pmatrix}, Y_1 \right), \dots, \left( \begin{pmatrix} F_{k1} \\ \vdots \\ F_{kr} \end{pmatrix}, Y_k \right) \right\}$$

The proposal stated here can shortly be described in the following way.

*There should be done a kind of preprocessing of the given input data for getting discrete data, which can be used as input information for an appropriate Back-propagation Network. This is then trained using the classical algorithm as part of this new one.*

In this thesis, the function based dataset is regarded more closely.

The contribution of this thesis at this point is on the one hand side the proposal of new ways in dealing with spatial data. On the other hand side, a generalised mathematical description of the relevant parts is developed and described with regard on the implementation in the SPATTB.

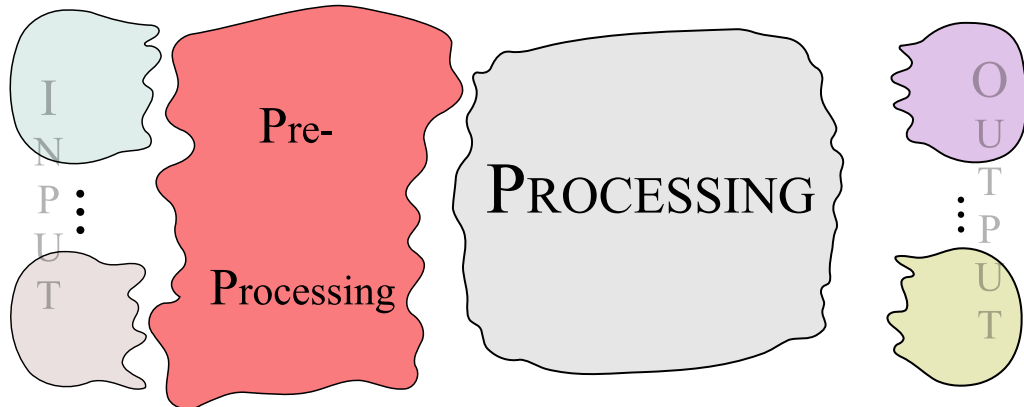
The effort, that should be achieved in this case, is a *discretization* of non-discrete input data. Then, it is possible to applicate the BPN-algorithm. The workflow is given in figure 9.3.

By this process, a mapping should be created, whose domain consists of a topological vector space and whose codomain consists of the space  $\mathbb{R}^n$ .

The preprocessing illustrated in figure 9.3 should be described by a mapping  $g^X$  defined by the following:

$$\begin{aligned} g^X : X &\longrightarrow \mathbb{R}^n \\ x &\longmapsto g^X(x) \end{aligned}$$

The method, that is used for approximation the coherence described above, is

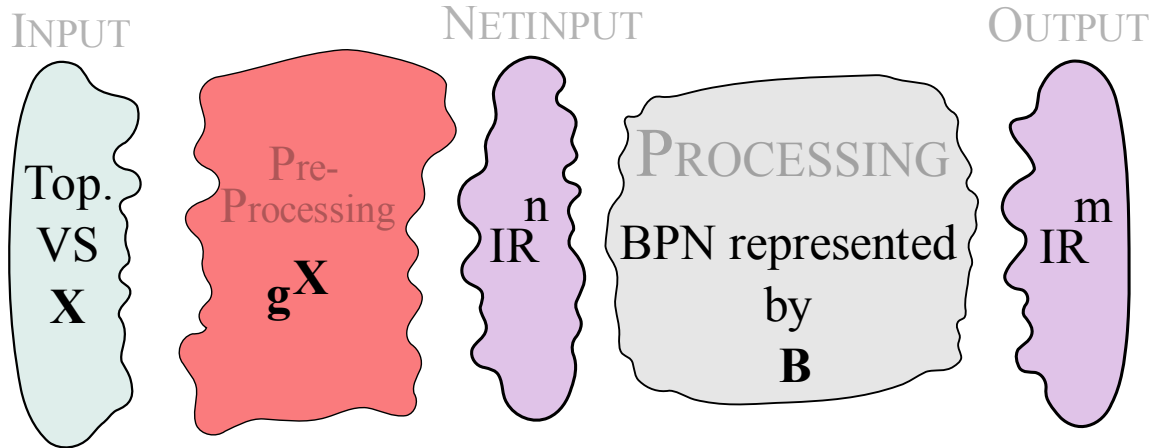


**Figure 9.3:** Overview of the proposed processing using a BPN in the context of discrete and non-discrete spatial data

the application of Backpropagation Networks as already mentioned. In chapter 4 is described, that the processing of such ANNs in the classical usage can be represented by different mappings. The definition of the particular used mapping depends on the actual point of view. On the one hand side, the ANN can be represented by a mapping, whose domain factors in the network mappings involved in the process of computing outputs basing on the given inputs. On the other hand side, such a BPN can be represented as a mapping  $B_t : R^n \longrightarrow \mathbb{R}^m$ , where these participating mappings only are used for defining the mapping process. The index  $t$  is used here to identify the actual state of the represented BPN.

At this point, this regarding is used. Hence, the processing illustrated in figure 9.3 can be described by such a mapping. The coherence of the mentioned spaces and mappings is shown in figure.

The aim of this raising to a higher level of abstraction is the provision of a more flexible and universal usability of the methods as it is given by the classical point of view. For the implementation in the developed concept “Spatial-Toolbox”, it is



**Figure 9.4:** Detailed overview of the processing

not only important to provide consistent ways of processing the available data, but also to describe possibilities of calculating information about differences between several methods, if they are applied to the same request. So, measurability of the described methods has to be guaranteed and has not to be neglected. Hence, in this chapter, measurability is regarded using the terms and methods described in chapter 7.

The mapping describing the whole processing should be denoted with  $F_t^X$  and is defined by the following composition:

$$F_t^X := B_t \circ g^X$$

This leads to the following description:

$$\begin{aligned} F_t^X : X &\longrightarrow \mathbb{R}^m \\ x &\longmapsto F_t^X(x) = B_t(g^X(x)) \end{aligned}$$

This mapping  $F_t^X$  consists of several component mappings, which can be described as follows:

9 Results of measure theory applied on the “Spatial-Toolbox”

$\forall_{i \in \{1, \dots, m\}}$  :  $F_{i,t}^X : X \longrightarrow \mathbb{R}$  with:

$$F_t^X = \begin{pmatrix} F_{1,t}^X \\ \vdots \\ F_{m,t}^X \end{pmatrix}$$

The mapping  $g^X$  as a part of the definition of the mapping  $F_t^X$  can be expressed using according component mappings as well:

$\forall_{i \in \{1, \dots, n\}}$  :  $g_i^X : X \longrightarrow \mathbb{R}$  with:

$$g^X = \begin{pmatrix} g_1^X \\ \vdots \\ g_n^X \end{pmatrix}$$

Last, but not least, the mapping  $B_t$  can be regarded more detailed on in the following way:

$\forall_{i \in \{1, \dots, m\}}$  :  $B_{i,t} : \mathbb{R}^n \longrightarrow \mathbb{R}$  with:

$$B_t = \begin{pmatrix} B_{1,t} \\ \vdots \\ B_{m,t} \end{pmatrix}$$

For the postulation of measurability according to the measuring approach using linear functionals defined in chapter 7, the following definition is required. In this, the space  $L^P(X, \mu, \mathbb{K}^n)$  is defined basing on the definition of a space  $L^P(X, \mu)$ .

**Definition 9.2.2 ( $L^P(X, \mu, \mathbb{K}^n)$ )**

Let  $f$  be a mapping of type  $f : X \longrightarrow \mathbb{K}^n$ , while  $X$  is an topological vec-

## 9 Results of measure theory applied on the “Spatial-Toolbox”

torspace and  $l \in \mathbb{N}$  and let  $\mathbb{K}$  be an arbitrary field.

The mapping  $f$  should be describable by component mappings  $f_i, i \in \{1, \dots, n\}$  and can be thereby denoted as:

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

This mapping is then called  $L^P(X, \mu, \mathbb{K}^n)$  - **integrable**,  $n \in \mathbb{N}$ , if and only if the following is valid:

$$\forall_{i \in \{1, \dots, n\}} \exists_{\mu_i \in \mathcal{MAP}(L^P(X, \mu_i), \mathbb{C})} : f_i \in L^P(X, \mu_i)$$

The mapping definition of this  $\mu$  is the following, while  $l, n \in \mathbb{N}$ :

$$\begin{aligned} \mu : L^P(X, \mu, \mathbb{K}^n) &\longrightarrow \mathbb{K}^n \\ f &\mapsto \mu(f) = \mu \left( \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \right) := \begin{pmatrix} \mu_1(f_1) \\ \vdots \\ \mu_n(f_n) \end{pmatrix} \end{aligned}$$

### Lemma 9.2.1

Let  $(X, \mathcal{O})$  be an arbitrary topological space. Furthermore, let  $g^X : X \longrightarrow \mathbb{R}^n$  with  $g^X \in \mathcal{C}(X, \mathbb{R}^n)$  be a mapping and let  $K \subseteq X$  be compact.

Then  $\overline{g^X(K)} := \overline{\left\{ x \in \mathbb{R}^n \mid \exists_{k \in K} : g^X(k) = x \right\}}$  is a compact set in the space  $\mathbb{R}^n$ .

*Proof.* Let  $g^X : X \longrightarrow \mathbb{R}^n$  with  $g^X \in \mathcal{C}(X, \mathbb{R}^n)$  be a mapping and let  $K \subseteq X$  be compact.

## 9 Results of measure theory applied on the “Spatial-Toolbox”

We consider an open cover of  $g^X(K)$  realised by a union of open sets. Let  $y \in g^X(K)$  an arbitrary element out of the image of the compact set  $K$  and let  $V_y := B_\epsilon(y) \subseteq \mathbb{R}^n$  be the open ball with radius  $\epsilon \in \mathbb{R}^+$  and center  $y \in g^X(K)$ . Then, the set  $M = \bigcup_{y \in g^X(K)} V_y$  is also an open set. Obviously,  $K \subseteq (g^X)^{-1}(M)$ .

Due to the continuity of the mapping  $g^X$ , the preimage of any open set is open and this is valid especially for the open set  $U_y := (g^X)^{-1}(V_y)$ , whereby  $(g^X)^{-1}(\{y\}) \subseteq K$ . This leads to the fact, that the set  $C := \bigcup_{y \in g^X(K)} U_y$  an open cover of  $K$  is.

Due to the compactness of the set  $K$ , there exists a finite subcover  $\bigcup_{i=1}^k U_i, k \in \mathbb{N}$ , while  $\forall_{i \in \{1, \dots, k\}} : U_i \in \{U_y \mid y \in g^X(K)\}$ .

This characteristic of covering is also valid for the image  $g^X(K)$  of  $K$ , i.e.  $g^X(K) \subseteq \bigcup_{i=1}^k U_i$ .

Regarding the construction of the sets  $U_i$ , it is valid, that  $\forall_{i \in \{1, \dots, k\}} \exists_{y_i \in g^X(K)} : g^X(U_i) \subseteq B_\epsilon(y_i) = V_{y_i}$ .

This leads to the fact, that the set  $g^X(K)$  can be covered by a finite set of  $\epsilon$ -balls, i.e.  $g^X(K) \subseteq \bigcup_{i=1}^k B_\epsilon(y_i)$ .

From this, it follows, that  $\forall_{y \in g^X(K)} \exists_{i \in \{1, \dots, k\}} : y \in V_{y_i} = B_\epsilon(y_i)$ .

From this reason, the following is valid:  $\forall_{y \in g^X(K)} \exists_{i \in \{1, \dots, k\}} : \|y\| \leq \|y_i\| + \epsilon$ .

Hence, it is is valid, that  $\forall_{y \in g^X(K)} \|y\| \leq \max_{i \in \{1, \dots, k\}} \|y_i\| + \epsilon$ .

From this, the set  $g^X(K)$  is bounded. Using the theorem of Heine-Borel (see [Rud05], p.42), a closed and bounded subset of the space  $\mathbb{R}^n, n \in \mathbb{N}$ , is compact. This can be applied on the set  $K_{g^X} := \overline{g^X(K)}$ , what consequently is a compact set.

□

### Theorem 9.2.1 (Integrability of $F_t^X$ )

*Let  $(X, \mathcal{O})$  be an arbitrary topological space. Furthermore, let the mapping*

9 Results of measure theory applied on the “Spatial-Toolbox”

$F_t^X : X \longrightarrow \mathbb{R}^m$  be defined as composition of the mappings  $g^X$  and  $B_t$  with  $F_t^X := B_t \circ g^X$ ,  $g^X : X \longrightarrow \mathbb{R}^n$ ,  $B_t : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ ,  $n, m \in \mathbb{N}$ . Thereby,  $g^X \in \mathcal{C}(X, \mathbb{R}^n)$  is postulated. Further, let  $B_t, t \in \mathbb{N}$  be a mapping representing a Backpropagation Network at iteration step  $t$ .

Then, there exist measures  $\mu|_K$ ,  $\lambda|_K$  and  $\eta$ , so that  $F_{t|K}^X \in L^P(K, \mu|_K, \mathbb{R}^m)$ ,  $g_{|K}^X \in L^P(K, \lambda|_K, \mathbb{R}^n)$  and  $B_{t|\overline{g^X(K)}} \in L^P(\overline{g^X(K)}, \eta, \mathbb{R}^m)$  for an arbitrary compact subset  $K \subseteq X$ .

*Proof.* In the theorem, the existence of a measure  $\mu|_K$  is postulated, where  $K$  is an arbitrary compact subset of  $X$ .

$F_t^X$  can be characterised by  $F_t^X : X \longrightarrow \mathbb{R}^m$ . The postulated  $\mu|_K$  is not defined for measuring this mapping  $F_t^X$ , but for the according mapping  $F_{t|K}^X$ , which can be characterised by :

$F_{t|K}^X : K \longrightarrow \mathbb{R}^m$ . In the process of proving, the according function space of which  $F_{t|K}^X$  is an element, will be specified more detailed.

The mapping definition of  $\mu|_K$  is the following:

$$\mu|_K : L^P(K, \mu|_K, \mathbb{R}^m) \longrightarrow \mathbb{R}^m .$$

Hence, the existence of such a measure  $\mu|_K$  operating on the space  $L^P(K, \mu|_K, \mathbb{R}^m)$  whose elements have an arbitrary compact subset of  $X$  as their domain has to be shown here. This fact is an essential information for the further proving.

In the following, first the integrability of  $B_{t|\overline{g^X(K)}}$  is regarded, then  $g_{|K}^X$  and finally  $F_{t|K}^X$ .

As mentioned above, the mapping  $B_t$  can be characterised by  $m$  component mappings  $B_{i,t}$  with  $\forall_{i \in \{1, \dots, m\}} : B_{i,t} : \mathbb{R}^n \longrightarrow \mathbb{R}$ . Each of these component mappings is defined as a composition of continuous mappings and so continuous itself.



## 9 Results of measure theory applied on the “Spatial-Toolbox”

The space  $\mathbb{R}^m$  as codomain of  $B_t$  is the cartesian product of  $m$  spaces  $\mathbb{R}$ . Hence, using definition 7.1.12, it represents a topological product with according product topology. The same is valid for the domain  $\mathbb{R}^n$ . Using this information, and the continuity of the  $B_{i,t}, i \in \{1, \dots, m\}$ , theorem 7.1.2 is applicable and  $B_t \in \mathcal{C}(\mathbb{R}^n, \mathbb{R}^m)$  results.

Regarding the mapping  $F_t^X|_K$ , the domain is a compact set and application of 9.2.1 provides the compactness of  $\overline{g^X(K)}$ .

Within the composition defined for  $F_t^X|_K$ , only elements out of  $g^X(K)$  are mapped by  $B_t$ . Hence, in the following a modified mapping with  $\overline{g^X(K)}$  as its domain is regarded. In this context, the following abbreviation is used:

$$\hat{K} := \overline{g^X(K)}$$

This mapping is then denoted as  $B_{t|\hat{K}} = \begin{pmatrix} B_{1,t|\hat{K}} \\ \vdots \\ B_{m,t|\hat{K}} \end{pmatrix}$ , with  $B_{t|\hat{K}} \in \mathcal{C}_C(\hat{K}, \mathbb{R}^m)$ .

So, for the particular component mapping, the following is valid:

$$\forall_{i \in \{1, \dots, m\}} : B_{i,t|\hat{K}} \in \mathcal{C}_C(\hat{K}, \mathbb{R})$$

From this characteristic follows, that there exist integrals  $\eta_i, i \in \{1, \dots, m\}$ , so that  $B_{i,t|\hat{K}} \in L^P(\hat{K}, \eta_i)$ .

Applying definition 9.2.2,  $B_{t|\hat{K}} \in L^P(\hat{K}, \eta, \mathbb{R}^m)$  is valid. Thereby,  $\eta := \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_m \end{pmatrix}$ .

In the next step, the mapping  $g_{|\hat{K}}^X$  is regarded.

According to the prerequisites, every component mapping of  $g^X$  is an element out of the space  $\mathcal{C}(X, \mathbb{R})$ .

9 Results of measure theory applied on the “Spatial-Toolbox”

So, for the particular component mappings  $g_{i|K}^X, i \in \{1, \dots, n\}$  of  $g|_K^X$ , the following is valid:

$$\forall_{i \in \{1, \dots, n\}} : g_{i|K}^X \in \mathcal{C}_C(K, \mathbb{R})$$

This leads to the fact, that there exist integrals  $\lambda_{i|K}, i \in \{1, \dots, n\}$ , so that  $g_{i|K}^X \in L^P(K, \lambda_{i|K})$ .

Applying definition 9.2.2,  $g|_K^X \in L^P(K, \lambda, \mathbb{R}^n)$  is valid. Thereby,  $\lambda := \begin{pmatrix} \lambda_{1|K} \\ \vdots \\ \lambda_{n|K} \end{pmatrix}$ .

Regarding now the mapping  $F_t^X$ , a similar explanatory statement can be given. Due to the definition  $F_t^X = B_t \circ g^X$ , the prerequisite  $g^X \in \mathcal{C}(X, \mathbb{R}^n)$  and the fact  $B_t \in \mathcal{C}(\mathbb{R}^n, \mathbb{R}^m)$  shown above, the characteristic  $F_{i|K}^X \in \mathcal{C}(K, \mathbb{R}^m)$  and even  $F_{i|K}^X \in \mathcal{C}_C(K, \mathbb{R}^m)$  follows.

So, for the particular component mappings  $F_{i,t|K}^X, i \in \{1, \dots, m\}$  of  $F_t^X$ , the following is valid:

$$\forall_{i \in \{1, \dots, m\}} : F_{i,t|K}^X \in \mathcal{C}_C(K, \mathbb{R})$$

This leads to the fact, that there exist integrals  $\mu_{i|K}, i \in \{1, \dots, m\}$ , so that  $F_{i,t|K}^X \in L^P(K, \mu_{i|K})$ .

Applying definition 9.2.2,  $F_t^X \in L^P(K, \mu, \mathbb{R}^m)$  is valid. Thereby,  $\mu := \begin{pmatrix} \mu_{1|K} \\ \vdots \\ \mu_{m|K} \end{pmatrix}$ .

## 9 Results of measure theory applied on the “Spatial-Toolbox”

These mappings  $\eta$ ,  $\lambda|_K$  and  $\mu|_K$  are the postulated integrals.

□

A this point, there is described a special topological vector space with regard on a possible application in the context of the SPATTB.

Therefor, the regarded spaces should be topologized in a certain way described in the following two definitions.

### Definition 9.2.3 (Euclidian topology on $\mathbb{R}^n$ )

In the space  $\mathbb{R}^n$ , the so-called **euclidian topology** is induced by the euclidian norm  $\|\cdot\|_e$  with the following basis  $\mathfrak{B}(\mathcal{O}_{\mathbb{R}^n})$ :

$$\begin{aligned} \mathfrak{B}(\mathcal{O}_{\mathbb{R}^n}) &:= \{B_\epsilon(x_0) \subset \mathbb{R}^n \mid \epsilon > 0 \wedge x_0 \in \mathbb{R}^n \\ &\quad \wedge B_\epsilon(x_0) := \{x \in \mathbb{R}^n \mid \|x - x_0\|_e < \epsilon\} \} \end{aligned}$$

### Definition 9.2.4 (Integral induced topology)

Let  $X_F$  be a topological vector space of mappings  $f: \tilde{X} \rightarrow \mathbb{R}$  with:

$$X_F := \{ f: \tilde{X} \rightarrow \mathbb{R} \mid \tilde{X} \text{ is a topological vector space} \}.$$

Additionally, let  $J$  be an appropriate indexing set.

Further, let the sets  $\mathcal{M}$  and  $X_{F\tilde{\mathcal{M}}}$  be defined as:

$$\mathcal{M} := \left\{ \tilde{\mu} \mid \tilde{\mu} \in \mathcal{P}\mathcal{L}\mathcal{F} \left( L^P \left( \tilde{X}, \tilde{\mu} \right), \mathbb{C} \right) \wedge P \in [1, \infty] \right\}$$

$$\bigvee_{\tilde{\mathcal{M}} \subseteq \mathcal{M}} X_{\tilde{\mathcal{M}}} := \left\{ f \in X_F \mid \exists_{P \in [1, \infty]} \exists_{\mu \in \tilde{\mathcal{M}}} f \in L^P \left( \tilde{X}, \mu \right) \right\} = \bigcap_{P \in [1, \infty]} L^P \left( \tilde{X}, \mu \right)$$

Then, using the following subbasis  $\mathcal{S}(\mathcal{O}_{X_{\tilde{\mathcal{M}}}, \tilde{\mathcal{M}}})$ , the open sets in  $\tilde{\mathcal{M}}$  are defined:

$$\mathcal{S}(\mathcal{O}_{X_{\tilde{\mathcal{M}}}, \tilde{\mathcal{M}}}) := \left\{ U_{\mu_j, B_\epsilon(y)} := \mu_j^{-1}(B_\epsilon(y)) \mid \mu_j \in \tilde{\mathcal{M}} \wedge \epsilon > 0 \wedge y \in \mu_j(X_{FI}) \right\}.$$

**Theorem 9.2.2**

Let  $X_F$  be a topological vector space of mappings  $f : \tilde{X} \longrightarrow \mathbb{R}$  with:

$$X_F := \{ f : \tilde{X} \longrightarrow \mathbb{R} \mid \tilde{X} \text{ is a topological vector space} \}$$

Further, let the set  $\mathcal{M}$  be defined as:

$$\mathcal{M} := \left\{ \tilde{\mu} \mid \tilde{\mu} \in \mathcal{P}\mathcal{L}\mathcal{F} \left( L^P \left( \tilde{X}, \tilde{\mu} \right), \mathbb{C} \right) \wedge P \in [1, \infty] \right\}$$

$$\text{Additionally, let } \mu := \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \in (\tilde{\mathcal{M}})^n, n \in \mathbb{N}$$

Then, there exists a mapping  $g^X \in \mathcal{C} \left( \tilde{X}_F, \mathbb{R}^n \right)$  with  $n \in \mathbb{N}$  and  $\tilde{X}_F \subseteq X_F$ .

*Proof.* Let  $X_F$  and  $\mathcal{M}$  be as defined above.

The mentioned subspace should be explicitly chosen as  $X_{\tilde{\mathcal{M}}} \subseteq X_F$  with:

$$X_{\tilde{\mathcal{M}}} := \left\{ f \in X_F \mid \exists_{P \in [1, \infty]} \exists_{\mu \in \tilde{\mathcal{M}}} f \in L^P \left( \tilde{X}, \mu \right) \right\} = \bigcap_{P \in [1, \infty]} L^P \left( \tilde{X}, \mu \right)$$

with  $\tilde{\mathcal{M}} \subseteq \mathcal{M}$ , as done in definition 9.2.4.

Then, with the subbase  $\mathcal{S} \left( \mathcal{O}_{X_{\tilde{\mathcal{M}}}, \tilde{\mathcal{M}}} \right)$  defined in 9.2.4,  $\left( X_{\tilde{\mathcal{M}}}, \mathcal{S} \left( \mathcal{O}_{X_{\tilde{\mathcal{M}}}, \tilde{\mathcal{M}}} \right) \right)$  is a topological space.

The base  $\mathfrak{B} \left( \mathcal{O}_{\mathbb{R}^n} \right)$  of the topological vector space  $(\mathbb{R}^n, \mathcal{O}_{\mathbb{R}^n})$  certainly as well is a subbase.

Let now the mapping  $g^X$  be defined in the following way:

## 9 Results of measure theory applied on the “Spatial-Toolbox”

$$g^X : X_{FI} \longrightarrow \mathbb{R}^n$$

$$f \longrightarrow g^X(f) := \mu(f) = \begin{pmatrix} \mu_1(f) \\ \vdots \\ \mu_n(f) \end{pmatrix}$$

This mapping can be described using the component mappings  $g_i^X, i \in \{1, \dots, n\}$ .

$$g_i^X : X_{FI} \longrightarrow \mathbb{R}$$

$$f \longrightarrow g_i^X(f) = \mu_i(f)$$

Using theorem 7.1.1 while regarding the both mentioned subbases,

$\forall_{i \in \{1, \dots, n\}} : g_i^X \in \mathcal{C}(X_{FI}, \mathbb{R})$  is valid:

Let therefor  $x_0 \in \mathbb{R}^n, \epsilon > 0$  and  $B_\epsilon(x_0) \in \mathfrak{B}(\mathcal{O}_{\mathbb{R}^n})$  be arbitrarily chosen. Then, for the preimage  $g_i^{X^{-1}}(B_\epsilon(x_0)) = \mu_i^{-1}(B_\epsilon(x_0)) \in \mathcal{S}(\mathcal{O}_{X_{\mathcal{M}}, \tilde{\mathcal{M}}})$  is valid due to the definition of  $\mathcal{S}(\mathcal{O}_{X_{FI}, UL})$ .

Using this continuity and applying theorem 7.1.2,  $g^X \in \mathcal{C}(X_{FI}, \mathbb{R}^n)$  is valid.

□

### Remark 9.2.1

*Fuzzy sets According to [Ber99], p.269, in most “real-world scenarios”, precise measurement is usually not given but a certain degree of uncertainty. One approach to handle this is the **fuzzy logic**, which can be traced back to L.A.Zadeh, who follows up this idea from 1965 on. His results and approaches can inter alia be found in [Zad65], [Zad73] and [Zad75].*

*At this point, elementary definitions concerning fuzzy theory are described basing on [Nau94], p.234 et seqq.*

### Definition 9.2.5 (Fuzzy set)

*Let  $X$  be a set and  $\mathcal{F}_\mu$  be a mapping defined in the following way:*

## 9 Results of measure theory applied on the “Spatial-Toolbox”

$$\begin{aligned} \mathcal{F}_\mu : X & \rightarrow [0, 1] \\ x & \mapsto \mathcal{F}_\mu(x) \end{aligned}$$

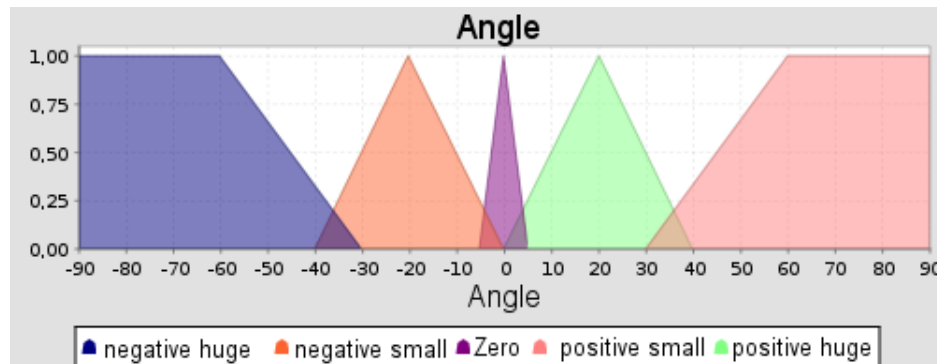
Then,  $\mathcal{F}_\mu$  is called a **fuzzy set**.

The set  $[0]^X$  should be the set of any fuzzy set according to the basing set  $X$ .

Instead of fuzzy set, in literature there is synonymously used the term membership function. Additionally, the use of these terms is nonuniform, some authors define the term fuzzy set as graph of the according membership function.

Fuzzy sets can be used to describe so-called linguistic variables using linguistic terms. According to the committee draft [iec], on which bases the international standard CEI IEC 61131-7 published in 2000, p. 6, Linguistic variables take values in the range of linguistic terms. Each linguistic term itself is described by an appropriate fuzzy set. Exemplarily, the linguistic value temperature is regarded. Possible linguistic terms are cold and warm.

In figure 9.5, fuzzy sets of the linguistic variable angle according to the deflection of a pendulum are illustrated.



**Figure 9.5:** Linguistic variable Angle (according to a pendulum's deflection)

For the handling in software systems, a horizontally representation of fuzzy sets using so-called  $\alpha$ -sections is more useful ([Nau94], p. 240).

**Definition 9.2.6 ( $\alpha$ -intersection)**

Let  $X$  be a set,  $\mathcal{F}_\mu \in [0]^X$  arbitrarily chosen and  $\alpha \in [0, 1]$ . Then, the  $\alpha$ -section of  $\mathcal{F}_\mu$  is defined by the following set :

$$[\mathcal{F}_{\mu\alpha}] := \{x \in X \mid \mathcal{F}_\mu(x) \geq \alpha\}$$

For implementation and further use of fuzzy sets, unions and intersections of them have to be defined. Hence, appropriate operators have to be used. Generally,  $t$ -norms define the characteristics, that are minimally necessary for an intersection operator and  $t$ -conorm describe that in terms of union operators.

**Definition 9.2.7**

$t$ -norm A mapping  $T : [0, 1]^2 \longrightarrow [0, 1]$  is called  **$t$ -norm**, if the following conditions are fulfilled:

$$(TN1) \quad \forall_{a \in [0,1]} : T(a, 1) = a$$

$$(TN2) \quad \forall_{a,b,c \in [0,1]} : a \leq b \Rightarrow T(a, c) \leq T(b, c)$$

$$(TN3) \quad \forall_{a,b \in [0,1]} : T(a, b) = T(b, a)$$

$$(TN4) \quad \forall_{a,b,c \in [0,1]} : T(a, T(b, c)) = T(T(a, b), c)$$

**Definition 9.2.8**

$t$ -conorm A mapping  $\perp : [0, 1]^2 \longrightarrow [0, 1]$  is called  **$t$ -norm**, if the following conditions are fulfilled:

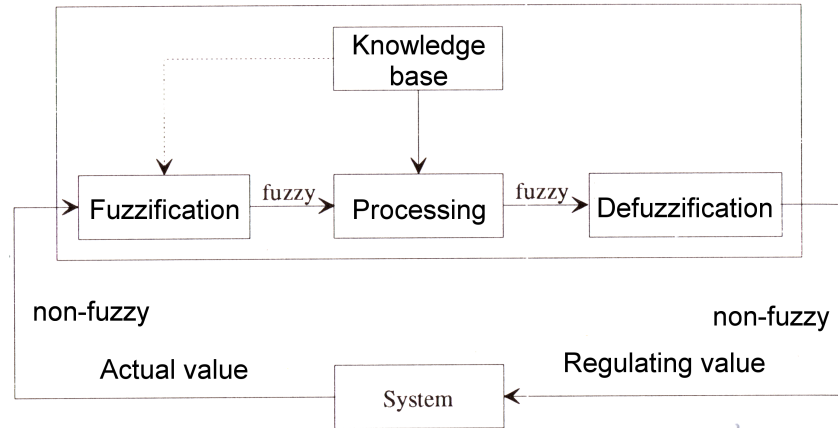
$$(TN1) \quad \forall_{a \in [0,1]} : \perp(a, 0) = a$$

$$(TN2) \quad \forall_{a,b,c \in [0,1]} : a \leq b \Rightarrow \perp(a, c) \leq \perp(b, c)$$

$$(TN3) \quad \forall_{a,b \in [0,1]} : \perp(a, b) = \perp(b, a)$$

$$(TN4) \quad \forall_{a,b,c \in [0,1]} : \perp(a, \perp(b, c)) = \perp(\perp(a, b), c)$$

At this point, there is mentioned one examples for a  $t$ -norm and a  $t$ -conorm:



**Figure 9.6:** General architecture of a fuzzy-controller ([Nau94], p.246, modified by the author)

$$T_{Luka} : [0, 1]^2 \longrightarrow [0, 1]$$

$$(a, b) \mapsto T_{Luka}(a, b) := \max\{0, a + b - 1\}$$

$$\perp_{Luka} : [0, 1]^2 \longrightarrow [0, 1]$$

$$(a, b) \mapsto T_{Luka}(a, b) := \min\{a + b, 1\}$$

*In implementations, there is often made use of so-called fuzzy-controller to control regulating values by processing the actual value of it. Detailed information can be found in [Nau94], [Ber99] and in [iec]. In figure such a fuzzy-controller is illustrated in a generalised way.*

In the following example, a pointwise evaluation of a mapping is used. First, it has to be shown, that this method defines an integral, i.e., that this defines a positive linear functional according to the definitions 7.5.4 and 7.5.8.

**Theorem 9.2.3**

*Let  $f, g$  be arbitrary mappings of type  $F : A \longrightarrow B$ . Further, let  $x_0 \in A$  be an arbitrary element and  $\mu_{x_0}$  be defined as:*



## 9 Results of measure theory applied on the “Spatial-Toolbox”

$$\begin{aligned} \mu_{x_0} : \mathcal{MAP}(A, B) &\longrightarrow \mathbb{R} \\ f &\longmapsto \mu_{x_0}(f) := f(x_0) \end{aligned}$$

Then,  $\mu_{x_0}$  is a positive linear functional.

*Proof.* The following has to be proven:

1.  $f \in \mathcal{MAP}(A, B)^+ \Rightarrow \mu_{x_0}(f) \geq 0$  real-valued

2.  $\forall_{\substack{\lambda, \eta \in \mathbb{R} \\ f, g \in \mathcal{MAP}(A, B)}} : \mu_{x_0}(\lambda \cdot f + \eta \cdot g) = \lambda \cdot \mu_{x_0}(f) + \eta \cdot \mu_{x_0}(g)$

• Let  $f \in \mathcal{MAP}(A, B)^+$  be arbitrarily chosen.

Then:  $\mu_{x_0}(f) \stackrel{\text{Def. } \mu_{x_0}(f)}{=} f(x_0) \stackrel{f \in \mathcal{MAP}(A, B)^+}{>} 0$

• Let  $f, g \in \mathcal{MAP}(A, B)$  and  $\lambda, \eta \in \mathbb{R}$  be arbitrarily chosen.

Then:  $\mu_{x_0}(\lambda \cdot f + \eta \cdot g) \stackrel{\text{Def. } \mu_{x_0}(f)}{=} (\lambda \cdot f + \eta \cdot g)(x_0) = (\lambda \cdot f)(x_0) + (\eta \cdot g)(x_0) = \lambda \cdot f(x_0) + \eta \cdot g(x_0) \stackrel{\text{Def. } \mu_{x_0}(f)}{=} \lambda \cdot \mu_{x_0}(f) + \eta \cdot \mu_{x_0}(g)$

□

### Example 9.2.1 (Approximation of mapping information)

The topological vector space, that should be regarded, is the vector space  $X_F$  of mappings  $f : \tilde{X} \longrightarrow \mathbb{R}$  :

$$X_F := \{ f : \tilde{X} \longrightarrow \mathbb{R} \mid \tilde{X} \text{ is a topological vector space} \}$$

regarded in theorem 9.2.2.

The situation in which that space plays an important role in the context of implementing the proposed processing in the SPATTB is described at this point. Suppositionally, there exists data consisting of geocoded membership functions associated with certain real numbers described by a vector out of the space  $\mathbb{R}^m$ . An example is the relationship between fuzzy information about the amount of oxygen and temperature concerning a certain water body and an associated number of mosquitos, that develop there in a period of time. In this case, there is only an association with

## 9 Results of measure theory applied on the “Spatial-Toolbox”

one real number described, but it also is conceivable to regard vectors describing different insects species.

According to remark 9.2.1, a fuzzy set has not to be continuous. Because of this, complex radon measures defined in definition 7.5.9 is not sufficient in any cases. Hence, the extended integration concerning  $\left(L^P\left(\tilde{X}, \mu\right)\right)$  is necessary in this context. So, the generalised perspective in theorem 9.2.2 is sensefull.

The basing data is described with the terms defined in 9.2.1 as follows:

Let  $F_i \in X_F$  and  $Y_i \in \mathbb{R}^m$  for  $i \in \{1, \dots, k\}$ ,  $k, m \in \mathbb{N}$ . Then, the following function based dataset is regarded:

$$\mathfrak{D}_F := \{(F_1, Y_1), \dots, (F_k, Y_k)\}$$

For applying the BPN-algorithm on the elements out of  $X_F$ , a mapping  $g^X$  has to be specified. It has to be ensured, that the regarded mapping components in  $\mathfrak{D}_F$  are integrable in terms of the definition of  $g^X$ .

According to the current application, the number  $n \in \mathbb{N}$  as dimension of the domain has to be set in an appropriate way. The dimension  $m \in \mathbb{N}$  is already set by the given desired outputs.

An essential information for describing the processing is the definition of  $g_{X_F}$ .

As proposed in theorem 9.2.2, this mapping consists of  $n$  onedimensional integrals and can be denoted as:

## 9 Results of measure theory applied on the “Spatial-Toolbox”

$$g^X : X_{FI} \longrightarrow \mathbb{R}^n$$

$$f \longrightarrow g^X(f) := \mu(f) = \begin{pmatrix} \mu_1(f) \\ \vdots \\ \mu_n(f) \end{pmatrix}$$

The definition is done in this general way to guarantee flexibility in choosing appropriate integrals. Restricted by the given mapping information, any appropriate integral out of the space  $\tilde{\mathcal{M}}$  (see definition 9.2.4) can be used within the discretization process. It is thinkable, to use the same integral  $\tilde{\mu}$  for every component of

$$g^X = \mu \text{ as } \mu = \begin{pmatrix} \tilde{\mu} \\ \vdots \\ \tilde{\mu} \end{pmatrix} \text{ or to use any combination of appropriate integrals.}$$

This would lead to the fact, that for the associated mapping, the function  $B_t$  gets a vector as input, whose components are all equal. Regarding the application of the Backpropagation Network-algorithm while the initial weights are generated by a randomization process, this type of input would lead to a leaping between different locations of minimal error. To improve these inputs, the following possibilities should be described.

The chosen integrals can describe a **pointwise evaluation** or an **integration concerning certain compact subsets** of  $\tilde{X}$ . So, discrete and non-discrete integration is considerable. Additionally, those two methods can be used separately, i.e. only pointwise evaluation for elements  $x_i \in \tilde{X}, i \in I$ , or an integration on compact subsets  $K_j \subset \tilde{X}, j \in J$  can be done or a combination of them.

So, let  $f \in X_F$  be such a mapping,  $\{x_0, x_1\}$  be a discrete subset of  $\tilde{X}$  and  $\{K_0, K_1, K_2\}$  be a set of compact subsets of this space.

Then, exemplarily, the following mapping  $\mu_1$  only using the pointwise evaluation can be defined:

9 Results of measure theory applied on the “Spatial-Toolbox”

$$\mu_1 = \begin{pmatrix} \mu_{x_0}(f) \\ \mu_{x_1}(f) \end{pmatrix} := \begin{pmatrix} f(x_0) \\ f(x_1) \end{pmatrix}$$

At the other hand side, the mapping  $\mu_2$  only using one compact subset of  $\tilde{X}$  can be defined:

$$\mu_2 = \begin{pmatrix} \mu_{21}(f) \\ \mu_{22}(f) \\ \mu_{23}(f) \end{pmatrix} := \begin{pmatrix} \int_{K_0} f d\mu_{21} \\ \int_{K_0} f d\mu_{22} \\ \int_{K_0} f d\mu_{23} \end{pmatrix}$$

The next level of modification is proposed as the regarding of different compact subsets at the particular components of the following mapping  $\mu_3$ :

$$\mu_3 = \begin{pmatrix} \mu_{31}(f) \\ \mu_{32}(f) \\ \mu_{33}(f) \end{pmatrix} := \begin{pmatrix} \int_{K_0} f d\mu_{31} \\ \int_{K_1} f d\mu_{32} \\ \int_{K_2} f d\mu_{32} \end{pmatrix}$$

Finally, a combination of the pointwise evaluation and the integration on compact subsets is proposed regarding the following mapping  $\mu_4$ :

$$\mu_4 = \begin{pmatrix} \mu_{x_0}(f) \\ \mu_{42}(f) \\ \mu_{x_1}(f) \\ \mu_{44}(f) \\ \mu_{45}(f) \end{pmatrix} := \begin{pmatrix} f(x_0) \\ \int_{K_0} f d\mu_{42} \\ f(x_1) \\ \int_{K_0} f d\mu_{44} \\ \int_{K_0} f d\mu_{45} \end{pmatrix}$$

This situation is illustrated in figure 9.7.

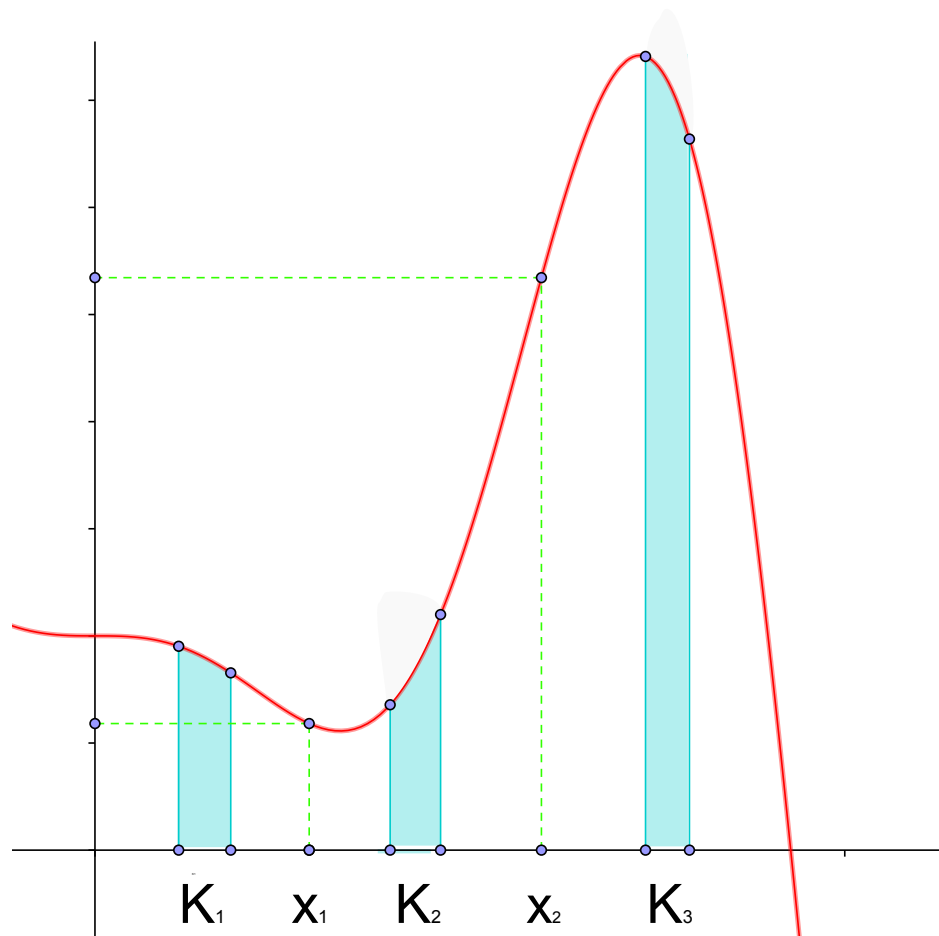


Figure 9.7: Exemplarily data for preprocessing

### 9.3 Interpolation using methods from computer-oriented geometry

In this section, the approach proposed in the last section is combined with the methods generalised in section 8.3.

The general situation, that should be regarded, is the following. Within a certain geographical region, there are given *discrete geocoded mapping information*. The aim is now, to interpolate this data to allow statements on geographical points, that

are not among the given data. In the following, there is proposed an interpolation using the generalised topological triangular Bézier surface from section 8.3.2 in chapter 8.

Therefore, the following processing steps are proposed:

- (PS1) Structuring the given data within the regarded geographical area using *triangulation*.
- (PS2) Topological triangular Bézier surface interpolation by taking into consideration  $\mathcal{C}^k$ -continuity at the connections of the single surfaces.
- (PS3) Description of the dealings with a certain geocoded request using this parameter based interpolation.

### 9.3.1 Structuring the given data within the regarded geographical area using triangulation

A triangulation is a separation of a given polygon or set of points in the space  $\mathbb{R}^l, l \in \mathbb{N}$  using triangles. For the description of triangulation, the term *point* should be used in the meaning of point in the space  $\mathbb{R}^l, l \in \mathbb{N}$ . According to [Far03], p. 143, a triangulation basing on a finite number of points can be characterised by the following:

- (TRI) The vertices of the triangles are in coincidence with the given points.
- (TRI) The inner of any pair of triangles is disjunct.
- (TRI) Two triangles have points in common if and only if there exist a common vertex or a common edge.
- (TRI) The union of the triangles is the convex hull of the given points.

For the description of a commonly used algorithm, the *algorithm of Delaunay triangulation*, the term *voronoi diagram* is necessary and defined basing on [Sch96],

p. 160 et seqq. and [Aum93], p.132 et seqq. There are only regarded the basing facts, because otherwise it would go beyond the scope of this thesis. In the following, the metric space  $(\mathbb{R}^l, d)$ ,  $l \in \mathbb{N}$  is used.

**Definition 9.3.1 (Voronoi-area)**

Let  $\mathcal{V} := \{P_1, \dots, P_n\} \subset \mathbb{R}^l, n \in \mathbb{N}$ . Then, the **Voronoi-areas**  $V(P_i), i \in \{1, \dots, n\}$  according to  $\mathcal{V}$  are defined in the following way:

$$\forall_{i \in \{1, \dots, n\}} : V(P_i) := \left\{ Q \in \mathbb{R}^l \mid \forall_{\substack{j \in \{1, \dots, n\} \\ j \neq i}} d(Q, P_i) \leq d(Q, P_j) \right\}$$

**Remark 9.3.1 (Half-planes and Voronoi-areas)**

If for  $A, B \in \mathbb{R}^l, A \neq B$ , the following set  $H_A$  defines the half-plane delimited by the perpendicular bisector of the side  $\overline{AB}$  containing  $A$ :

$$H_A := \{Q \in \mathbb{R}^l \mid d(A, Q) \leq d(B, Q)\}.$$

Then, the following is valid ([Aum93], p.132):

$$\forall_{i \in \{1, \dots, n\}} : V(P_i) = \bigcap_{\substack{j=1 \\ j \neq i}} H_{P_i}(P_j)$$

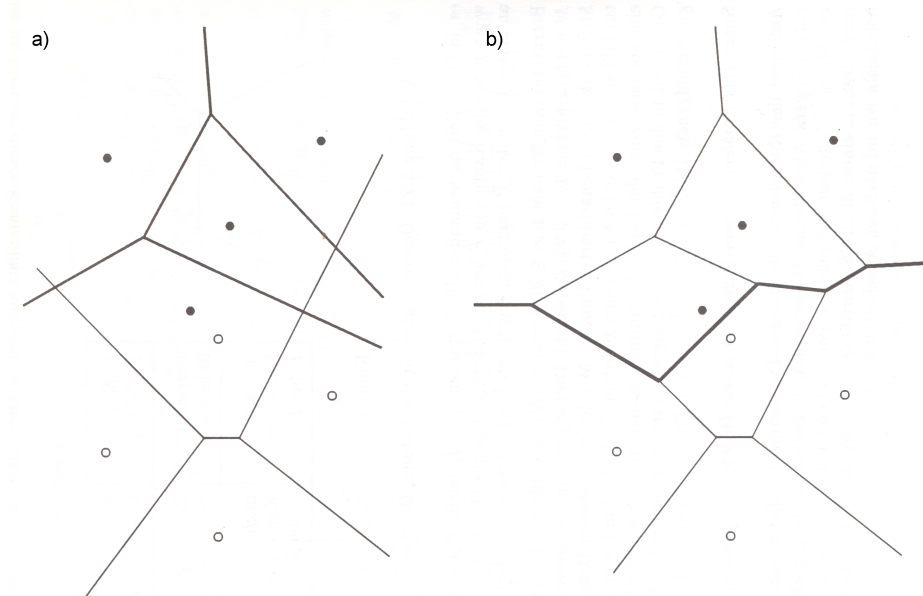
**Definition 9.3.2 (Voronoi-diagram)**

Let  $\mathcal{V} := \{P_1, \dots, P_n\} \subset \mathbb{R}^l, n \in \mathbb{N}$ . Then, the set  $\mathcal{VOR}$  containing any Voronoi-area of  $\mathcal{V}$  is named **Voronoi-diagram** or **Voronoi-tessellation**:

$$\mathcal{VOR} := \{V(P_i) \mid i \in \{1, \dots, n\} \wedge P_i \in \mathcal{V}\}$$

The following algorithm to calculate the according Voronoi-diagram for a given set of points is of the type *Sdivide et impera* deviding the given data and then doing iterative processings. In substance, the given set of points is divided in subsets containing three points, the according Voronoi-diagrams are calculated and stepwise merged. In figure 9.8, there is illustrated, that just merging make problems, because there can be generated areas containing more than one point

in the inner. To avoid this, an *algorithm of generating a separating edge chain* is used and is found in [Sch96], p.166-167.



**Figure 9.8:** Necessity of a separating edge chain: a) without separating edge chain, b) with separating edge chain ([Aum93], p.142)

**Algorithm 9.3.1 (Voronoi-diagram by divide et impera)**

(VA1) Start procedure  $Voronoi(\{P_1, \dots, P_n\})$

(VA2) Sort the given points to get a sorted list  $\{P_1^*, \dots, P_n^*\}$

(VA3) if  $n \leq 3$

- Calculate Voronoi-diagram
- Calculate convex hull
- End

(VA4) else

- Divide the sorted list by  $A_L := \{P_1^*, \dots, P_{\lfloor \frac{n}{2} \rfloor}^*\}$  and  $A_R := \{P_{\lfloor \frac{n}{2} \rfloor + 1}^*, \dots, P_n^*\}$



## 9 Results of measure theory applied on the “Spatial-Toolbox”

- $Voronoi(A_L)$
- $Voronoi(A_R)$
- Merge  $(A_L)$  and  $(A_R)$  as well as the according convex hulls
- Application of the algorithm of generating a separating edge chain
- End

(VA5) Return the final Voronoi-diagram

(VA6) End

Basing on this, the *Delaunay-triangulation* can be calculated using the following algorithm ([Sch96], p.184) regarding a set  $\mathcal{V} := \{P_1, \dots, P_n\} \subset \mathbb{R}^l, n \in \mathbb{N}$ .

### **Algorithm 9.3.2 (Delaunay-triangulation)**

(DA1) Calculate the according Voronoi-diagram

(DA2) Calculate the dualseparation of this Voronoi-diagram

Thereby, the dual separation is done by connecting any pair of points out of  $\mathcal{V}$ , that has an edge of the Voronoi-diagram in common. Delaunay proofed in 1934 ([Del34]), that by this method a triangulation fullfilling the conditions mentioned above is constructed.

Regarding a triangulation of a set  $\mathcal{V} := \{P_1, \dots, P_n\}$ , the ability of localizing a point  $P \in \text{conv}(V)$  within the constructed triangles is important for the following. A simple method of doing that localisation consists in a stepwise calculation of the barycentric coordinates of  $P$  for the single triangles. This is done, until a triangle is found according to what any barycentric coordinate is positive for  $P$ . The worst case would be the necessity of regarding any triangle. In [Far03], p.146-147, there is mentioned a more efficient algorithm of localisation.

**Algorithm 9.3.3 (Localization within a given triangulation)**

(LT1) Given: Triangulation of a set  $\mathcal{V} := \{P_1, \dots, P_n\}$

A point  $P \in \text{conv}(V)$

(LT2) Initializing a starting triangle using a certain algorithm

(LT3) Calculating the according barycentric coordinates

(LT4) if they all are positive, then the actual triangle is the output

(LT5) else, set the neighboring triangle in whose direction

the biggest negative barycentric coordinate was calculated as actual triangle and go to LT3

In figure 9.9, there is visualized the convex hull, the Voronoi-diagram and the Delaunay-triangulation of a set of 30 points. This has been done within the GIS GRASS using the sample dataset *spearfish 60* with mapset *PERMANENT* and raster map *aspect*. This OpenSource software and the sample data can be found on [GRA10]. The following commands were used:

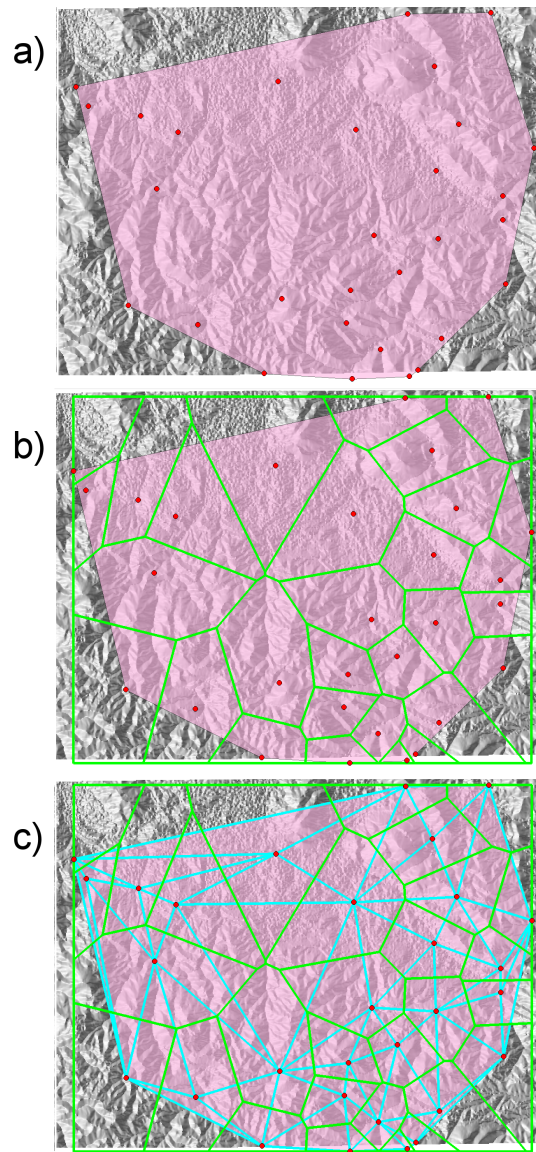
- Convex hull: `v.hull`
- Voronoi-diagram: `v.voronoi`
- Delaunay-triangulation: `v.delaunay`

Additionally, in figure 9.10, only the triangulation and as well the coherence between the Voronoi-diagram and the according delaunay-triangulation are illustrated. In figure 9.11, a threedimensional view is shown.

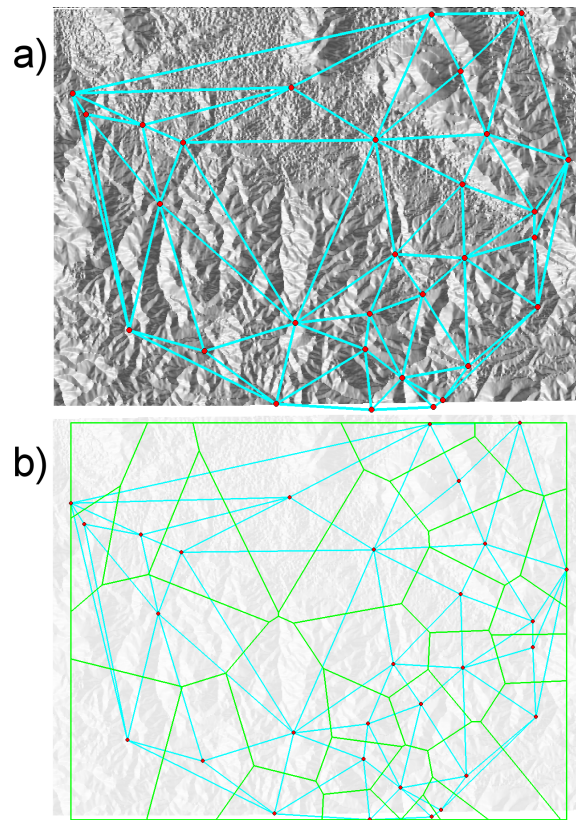
### 9.3.2 Topological triangular Bézier surface interpolation

For the following considerations, some prerequisites are necessary:

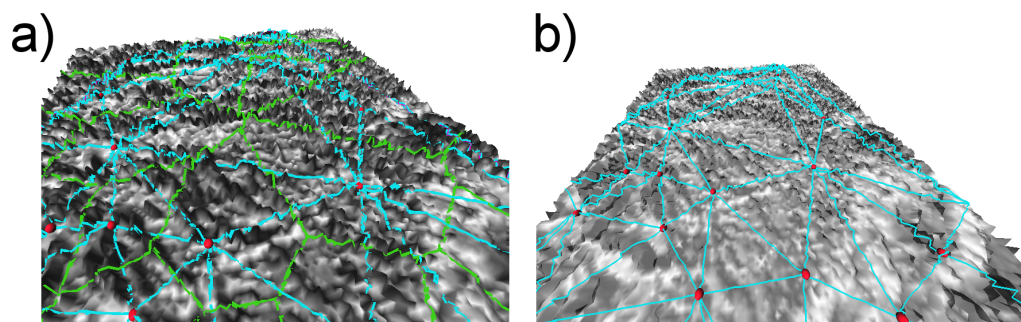
In general, the following cartesian product  $\mathcal{GF}_{P,n} := L^P(\tilde{X}, \mu) \times \mathbb{R}^n$  is regarded as basing space of *geocoded mapping information*, while  $\tilde{X}$  is a topological vector space



**Figure 9.9:** Convex hull, Voronoi-diagram and Delaunay-triangulation using GRASS: a) Points and according convex hull, b) Addition of the Voronoi-diagram, c) Addition of the Delaunay-triangulation



**Figure 9.10:** a) Given data and according Delaunay-triangulation, b) Coherence between Voronoi-diagram and Delaunay-triangulation



**Figure 9.11:** Threedimensional view: a) Voronoi-diagram and Delaunay-triangulation, b) Delaunay-triangulation

## 9 Results of measure theory applied on the “Spatial-Toolbox”

itself. In the following, the case  $\mathcal{GF}_{P,2}$  should be regarded. There, the elements are interpreted as geocoded  $L^P(\tilde{X}, \mu)$ -integrable mappings using twodimensional geo-coordinates. In terms of geocoded information, the cases  $n \geq 2$  are sensfull as well regarding e.g. additional information about altitude or time.

The set  $\mathcal{MI} \subset \mathcal{GF}_{P,2}$  should consist of the following entries:

$$\mathcal{MI} := \{(f_1, c_1), \dots, (f_k, c_k)\}, k \in \mathbb{N}.$$

This data should in the following be used as basis for the proposed interpolation processing. In section 8.3 of chapter 8, there has been proposed a generalization of surface interpolation regarding discrete subsets in a topological vector space using Bézier-methods. At this point, there are given reasons for the applicability of that method on the set  $\mathcal{MI}$ .

This method is applicable on  $\mathcal{MI}$ , because the space  $\mathcal{GF}_{1,2}$  is a topological vector space. The reason therefore is given by the fact, that the cartesian product of a family of topological vector spaces is a topological vector space as well, if the same scalar field  $\mathbb{K}$  and the according product topology are used ([Sch99], p.19). Therefor, in the space  $\mathcal{GF}_{2,1}$ , addition and scalar multiplication have to be defined in the following way:

$$\begin{aligned} + : \quad & \mathcal{GF}_{P,2} \times \mathcal{GF}_{P,2} \longrightarrow \mathcal{GF}_{P,2} \\ & ((f_1, x_1, y_1), (f_2, x_2, y_2)) \mapsto +(((f_1, x_1, y_1), (f_2, x_2, y_2))) \\ & \qquad \qquad \qquad \qquad \qquad \qquad := (f_1 + f_2, x_1 + x_2, y_1 + y_2) \\ \\ \cdot : \quad & \mathbb{K} \times \mathcal{GF}_{P,2} \longrightarrow \mathcal{GF}_{P,2} \\ & (\lambda, (f_2, x_2, y_2)) \mapsto \cdot((\lambda, (f_2, x_2, y_2))) := (\lambda \cdot f_1, \lambda \cdot x_1, \lambda \cdot y_1) \end{aligned}$$

The space  $\mathbb{R}$  is a topological vector space and so, the cartesian product  $\mathbb{R}^2$  with

the above postulations in mind is as well a topological vector space.

Before interpolation with the described triangular Bézier-methods is possible, the data has to be structured by triangulation. Doing this, only the components describing the geocoding of each datum are regarded. Using the interpolation algorithm, there is generated a set of parameterized surfaces, which can respectively be functionally described by:

$$\begin{aligned} \tilde{\mathcal{P}}\mathcal{S} : \quad \mathbb{D}_3 &\longrightarrow \mathcal{GF}_{P,2} \\ (i, j, k) &\mapsto \tilde{\mathcal{P}}\mathcal{S}((i, j, k)) \end{aligned}$$

Altogether, there should be described mapping from the convex hull of the given coordinates to the according elements out of  $L^P(\tilde{X}, \mu)$ . As described further, this mapping cannot be explicitly defined, but there is proposed an algorithm for calculating the according element out of  $L^P(\tilde{X}, \mu)$  for a given coordinate within the regarded convex hull.

For describing the mapping definition, the coordinate information has to be extracted for the particular elements of the set  $\mathcal{MI}$ . Such an extraction can be formulated in a generalised way for the space  $\mathcal{GF}_{P,n}$  by the following localisation mapping  $\mathcal{LOK}$ :

$$\begin{aligned} \mathcal{LOK} : \mathcal{GF}_{P,n} &\longrightarrow \mathbb{R}^n \\ (f, x) &\mapsto \mathcal{LOK}((f, x)) := x \end{aligned}$$

while  $x \in \mathbb{R}^n$  is assumed.

So let  $A \subseteq \mathcal{GF}_{P,n}$ . Then the extraction should be described by the following set:

$$A_L := \mathcal{LOK}(A) = \left\{ x \in \mathbb{R}^n \mid \exists_{g \in A} : x = \mathcal{LOK}(g) \right\}$$

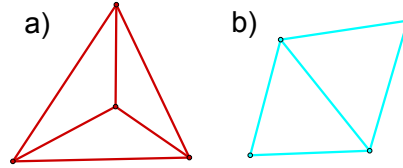
Applying this on the general space  $\mathcal{GF}_{P,n}$ , the following mapping definition results:

$$\begin{aligned} \mathcal{F}_{P,n,A} : \text{conv}(A_L) &\longrightarrow L^P(\tilde{X}, \mu) \\ a &\mapsto \mathcal{F}_{P,n,A}(a) \end{aligned}$$

With  $n := 2$  and  $\tilde{A} \subseteq \mathcal{GF}_{P,2}$ , the following description can be given:

$$\begin{aligned} \mathcal{F}_{P,2,\tilde{A}} : \text{conv}(\tilde{A}_L) &\longrightarrow L^P(\tilde{X}, \mu) \\ a &\mapsto \mathcal{F}_{P,2,A}(a) \end{aligned}$$

For the process of interpolating using Bézier-surfaces, the number of triangles generated by the according triangulation is an essential information. In figure 9.12, there are illustrated two triangulations of four points. Obviously, the number of triangles does not only depend on the number of points used for the triangulation, but also on the subset of these points that are elements of the according convex hull's border.



**Figure 9.12:** Triangulations of four points: a) Three points in the according convex hull, b) Four points in the according convex hull

At this point, there should be given a possibility of calculating the number of triangles generated by a triangulation using methods of graph theory. In the following, the term *planar graph* and *euler's polyeder theorem* are described basing on [Die00], p. 70 et seqq.

**Definition 9.3.3 (Planar graph)**

Let  $G = (V(G), E(G), I_G)$  be a graph, with  $|V(G)| < \infty$  and  $|E(G)| < \infty$ . If additionally the following prerequisites are valid,  $G$  is called **planar graph**:

(PG1)  $V \subseteq \mathbb{R}^2$

(PG2) Every edge is a union of line segments in  $\mathbb{R}^2$  between two vertices.

(PG3) Every pair of different edges have at most one end in common

(PG4) The interior of every edge contains no vertex and does not cross any other edge

$G$  is called **connected**, if every vertex is the end of an edge.

**Definition 9.3.4 (Faces of plane graphs)**

Let  $G$  be a plane graph. Then, the regions of the set  $\mathbb{R}^2 \setminus E(G)$  separated by the edges of  $G$  are called the **faces** of that graph.

**Theorem 9.3.1 (Euler’s polyeder theorem)**

Let  $G$  be a connected plane graph with  $n$  vertices,  $m$  edges and  $l$  faces. Then, the following is valid:

$$n - m + l = 2 \tag{9.1}$$

The proof can exemplarily be found in [Die00], p.74-75.

Using this, the number of triangles generated by a triangulation is derived in the following way:

**Remark 9.3.2 (Number of triangles)**

Let  $A \subset \mathbb{R}^2$  be a finite subset. Further, let  $T_A$  be an triangulation basing on the elements of  $A$ .  $T_A$  is then a plain graph with  $e_A$  edges,  $v_A$  vertices and  $f_A$  faces. Additionally, let  $c_A \leq v_A$  be the quantum of vertices, that are element of the border of  $\text{conv}(A)$ . Then, there are generated  $2v_A - e_A - c_A$ .

This can be seen by the following considerations:

Let  $tr_A$  be the number of generated triangles. Then, the number of faces  $f_A$  is  $f_A = tr_A + 1$  consisting of the number of triangles and the outer face. Regarding



## 9 Results of measure theory applied on the “Spatial-Toolbox”

the edges, each triangle has three edges and the outer face  $c_A$ . Further, for every edge, two neighboring faces can be found. So, the following is valid:

$$e_A = \frac{3tr_A + c_A}{2}$$

Now, Euler’s polyeder theorem can be applied in the following way:

$$\begin{aligned} 2 &= v_A - e_A + f_A \\ &= v_A - \frac{3tr_A + c_A}{2} + tr_A + 1 \\ \Leftrightarrow 4 &= 2v_A - 3tr_A - c_A + 2tr_A + 2 \\ \Leftrightarrow tr_A &= 2v_A - 2 - c_A \end{aligned}$$

So now, it is focused on the set  $\mathcal{MI}$  again. In the context of triangulation, the set  $\mathcal{MI}_L$  has to be regarded. With  $|\mathcal{MI}| = k$  also  $|\mathcal{MI}_L| = k$  is valid. With  $c_{\mathcal{MI}_L}$ , the number of elements is described, that are element of  $conv(\mathcal{MI}_L)$ . Using this and remark 9.3.2, the number of triangles  $tr_{\mathcal{MI}_L}$  generated by the according Delaunay-triangulation is the following:

$$tr_{\mathcal{MI}_L} = 2k - 2 - c_{\mathcal{MI}_L}$$

Thereby,  $tr_{\mathcal{MI}_L}$  also describes the the number of topological triangular Bézier surfaces, that have to be generated. These should be denoted as  $\mathcal{GFB}_i, i \in \{1, \dots, tr_{\mathcal{MI}_L}\}$ .

To be able to build a topological triangular Bézier surface consisting of a number of single topological triangular Bézier surface, the continuous combination of those surfaces has to be regarded. At this point, the coherences are described basing on [Hos89], pp.251-252. (done there for triangular Bézier surfaces).

### **Remark 9.3.3 (Neighboring topological triangular Bézier surfaces)**

*In the following, for identifying the Bézier-points of the surface with grade  $n = 1$ , the nomenclature **outer vertices** is used.*

## 9 Results of measure theory applied on the “Spatial-Toolbox”

At this point, there should be regarded  $\mathcal{C}^0$ - and  $\mathcal{C}^1$ -continuity for the descent of two neighboring topological triangular Bézier surface of grade  $n \in \mathbb{N}$ . Neighboring means coincidence of two outer vertices.

For getting the ability of describing  $\mathcal{C}^1$ -continuity, there has to be regarded a Bézier-surface of grade 3 and therefor, the missing Bézier-points have to be calculated using the algorithm of grade elevation 8.3.2.

Therefor, the participating elements out of  $\mathcal{MI}$  have to be identified for the single triangle respectively the single Bézier-surface. At this point, it is assumed, that the three points  $c_{k_1}$ ,  $c_{k_2}$  and  $c_{k_3}$  defines the triangle associated with the Bézier-surface  $\mathcal{GFB}_t, t \in \{1, \dots, tr_{\mathcal{MI}_L}\}$ . Additionally,  $k_1 < k_2 < k_3$  is assumed. The outer vertices are named as:

$$b_{(1,0,0)}^{(t)} := (f_{k_1}, c_{k_1})$$

$$b_{(0,1,0)}^{(t)} := (f_{k_2}, c_{k_2})$$

$$b_{(0,0,1)}^{(t)} := (f_{k_3}, c_{k_3})$$

For describing the according grade elevation, let  $\hat{i}, \hat{j}, \hat{k} \in \mathbb{N}_{\leq 1}, i + j + k = 1$ ,  $\tilde{i}, \tilde{j}, \tilde{k} \in \mathbb{N}_{\leq 2}, i + j + k = 2$  and  $i, j, k \in \mathbb{N}_{\leq 3}, i + j + k = 3$ . Then, using the mentioned algorithm, the missing Bézier-points are calculated in the following way:

$$b_{(\hat{i}, \hat{j}, \hat{k})}^{(t)} := \frac{\tilde{i}}{2} b_{(\tilde{i}-1, \tilde{j}, \tilde{k})}^{(t)} + \frac{\tilde{j}}{2} b_{(\tilde{i}, \tilde{j}-1, \tilde{k})}^{(t)} + \frac{\tilde{k}}{2} b_{(\tilde{i}, \tilde{j}, \tilde{k}-1)}^{(t)}$$

$$b_{(i, j, k)}^{(t)} := \frac{i}{3} b_{(i-1, j, k)}^{(t)} + \frac{j}{3} b_{(i, j-1, k)}^{(t)} + \frac{k}{3} b_{(i, j, k-1)}^{(t)}$$

Thereby,  $b_{(\hat{i}, \hat{j}, \hat{k})}^{(t)} := \mathfrak{D}_{\mathcal{GF}_{P,2}}$ , if  $\hat{i} \notin [0, 1] \vee \hat{j} \notin [0, 1] \vee \hat{k} \notin [0, 1]$  and  $b_{(\tilde{i}, \tilde{j}, \tilde{k})}^{(t)} := \mathfrak{D}_{\mathcal{GF}_{P,2}}$ , if  $\tilde{i} \notin [0, 1] \vee \tilde{j} \notin [0, 1] \vee \tilde{k} \notin [0, 1]$  by this iteration.

Using this, the topological triangular Bézier surface  $\mathcal{GFB}_t, t \in \{1, \dots, tr_{\mathcal{MI}_L}\}$  can be generated with:

## 9 Results of measure theory applied on the “Spatial-Toolbox”

$$\begin{aligned}
 \forall t \in \{1, \dots, tr_{\mathcal{M}\mathcal{I}_L}\} : \mathcal{GFB}_t : \quad \mathbb{D}_3 &\longrightarrow \mathcal{GF}_{P,2} \\
 (u, v, w) &\mapsto \mathcal{GFB}_t((u, v, w)) \\
 &:= \sum_{\substack{i, j, k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)}^{(t)} \cdot \mathcal{BS}_{(i,j,k)}^n((u, v, w))
 \end{aligned}$$

The considerations necessary for getting  $\mathcal{C}^0$ - and  $\mathcal{C}^1$ -continuity are described using two neighboring topological triangular Bézier surfaces  $\mathcal{GFB}_r, r \in \{1, \dots, tr_{\mathcal{M}\mathcal{I}_L}\}$  and  $\mathcal{GFB}_s, s \in \{1, \dots, tr_{\mathcal{M}\mathcal{I}_L}\}$  with  $i \neq j$ .

### $\mathcal{C}^0$ -continuity:

In this case, the according marginal curve for  $u = 0$  has to be equal:

$$\mathcal{GFB}_r((0, v, w)) = \mathcal{GFB}_s((0, v, w)) \quad (9.2)$$

Using the coherence described in remark 8.3.2, this leads to the following:

$$\begin{aligned}
 \sum_{\substack{j, k \in \mathbb{N}_{\leq n} \\ j+k=n}} b_{(0,j,k)}^{(r)} \cdot \mathcal{BS}_{(0,j,k)}^n((0, v, w)) &= \sum_{\substack{j, k \in \mathbb{N}_{\leq n} \\ j+k=n}} b_{(0,j,k)}^{(s)} \cdot \mathcal{BS}_{(0,j,k)}^n((0, v, w)) \\
 \Leftrightarrow \sum_{\substack{j, k \in \mathbb{N}_{\leq n} \\ j+k=n}} b_{(0,j,k)}^{(r)} \cdot \frac{n!}{j!k!} \cdot v^j \cdot (1-v)^k &= \sum_{\substack{j, k \in \mathbb{N}_{\leq n} \\ j+k=n}} b_{(0,j,k)}^{(s)} \cdot \frac{n!}{j!k!} \cdot v^j \cdot (1-v)^k
 \end{aligned}$$

### $\mathcal{C}^1$ -continuity:

For getting  $\mathcal{C}^1$ -continuity:, along the marginal curve for  $u = 0$ , the linear dependence of the three vectors  $\frac{d\mathcal{GFB}_r((0,v,w))}{dv}$ ,  $\frac{d\mathcal{GFB}_r((u,v,w))}{du} \Big|_{u=0}$  and  $\frac{d\mathcal{GFB}_s((u,v,w))}{du} \Big|_{u=0}$ .

9 Results of measure theory applied on the “Spatial-Toolbox”

$$\begin{aligned}
\frac{d\mathcal{GF}\mathcal{B}_r((0, v, w))}{dv} &= \sum_{\substack{j,k \in \mathbb{N}_{\leq n} \\ j+k=n}} b_{(0,j,k)}^{(r)} \cdot \frac{n!}{j!k!} \cdot v^{(j-1)} \cdot (1-v)^k \\
&\quad - b_{(0,j,k)}^{(r)} \cdot \frac{n!}{j!k!} \cdot v^j \cdot (1-v)^{(k-1)} \\
&= \sum_{\substack{j,k \in \mathbb{N}_{\leq n} \\ j+k=n}} b_{(0,j,k)}^{(r)} \cdot \mathcal{BS}_{(0,j,k)}^n((0, v, w)) (jv^{-1} - k(1-v)^{-1}) \\
&= n \cdot \sum_{\substack{j,k \in \mathbb{N}_{\leq n} \\ j+k=n}} \left( b_{(0,j+1,k)}^{(r)} - b_{(0,j,k+1)}^{(r)} \right) \cdot \mathcal{BS}_{(0,j,k)}^{n-1}((0, v, w))
\end{aligned}$$

$$\begin{aligned}
\frac{d\mathcal{GF}\mathcal{B}_r((u, v, w))}{du} \Big|_{u=0} &= \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)}^{(r)} \cdot \frac{n!}{i!j!k!} \cdot iu^{(i-1)}v^j \cdot (1-u-v)^k \\
&\quad - b_{(i,j,k)}^{(r)} \cdot \frac{n!}{i!j!k!} \cdot u^i \cdot v^j \cdot k(1-u-v)^{(k-1)} \Big|_{u=0} \\
&= \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i,j,k)}^{(r)} \cdot \mathcal{BS}_{(i,j,k)}^n((u, v, w)) (iu^{-1} - k(1-v)^{-1}) \Big|_{u=0} \\
&= n \cdot \sum_{\substack{i,j,k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} \left( b_{(i+1,j,k)}^{(r)} - b_{(i,j,k+1)}^{(r)} \right) \cdot \mathcal{BS}_{(i,j,k)}^{n-1}((u, v, w)) \Big|_{u=0}
\end{aligned}$$

9 Results of measure theory applied on the “Spatial-Toolbox”

$$\begin{aligned}
\frac{d\mathcal{GFB}_s((u, v, w))}{du} \Big|_{u=0} &= \sum_{\substack{i, j, k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i, j, k)}^{(s)} \cdot \frac{n!}{i!j!k!} \cdot iu^{(i-1)}v^j \cdot (1 - u - v)^k \\
&\quad - b_{(i, j, k)}^{(s)} \cdot \frac{n!}{i!j!k!} \cdot u^i \cdot v^j \cdot k(1 - u - v)^{(k-1)} \Big|_{u=0} \\
&= \sum_{\substack{i, j, k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} b_{(i, j, k)}^{(s)} \cdot \mathcal{BS}_{(i, j, k)}^n((u, v, w)) (iu^{-1} - k(1 - v)^{-1}) \Big|_{u=0} \\
&= n \cdot \sum_{\substack{i, j, k \in \mathbb{N}_{\leq n} \\ i+j+k=n}} \left( b_{(i+1, j, k)}^{(s)} - b_{(i, j, k+1)}^{(s)} \right) \cdot \mathcal{BS}_{(0, j, k)}^{n-1}((u, v, w)) \Big|_{u=0}
\end{aligned}$$

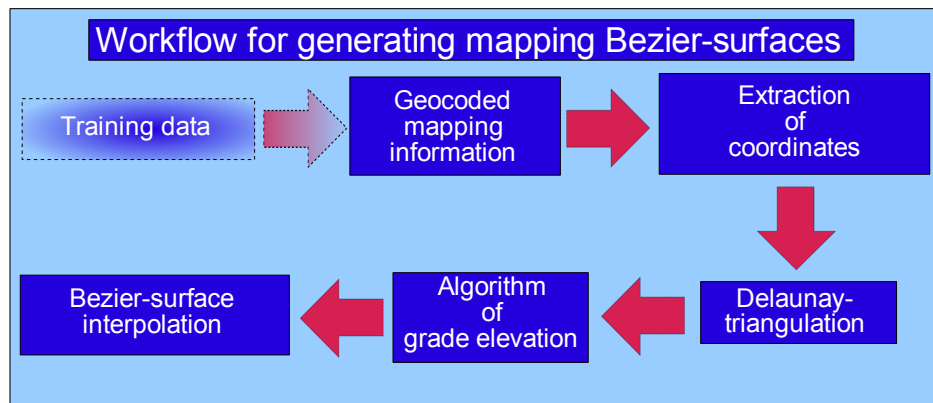
The prerequisite of linear dependence can be expressed by the following equation:

$$\forall_{\substack{\lambda_1, \lambda_2 \in \mathbb{R} \\ \lambda_1 + \lambda_2 > 0}} : \frac{d\mathcal{GFB}_r((0, v, w))}{dv} = \lambda_1 \cdot \frac{d\mathcal{GFB}_r((0, v, w))}{du} + \lambda_2 \cdot \frac{d\mathcal{GFB}_s((0, v, w))}{du} \quad (9.3)$$

At this point, the considerations made until now are joined to the proposition of an algorithm. With this algorithm, Bézier-surface interpolation of mapping information is possible in consideration of the necessary prerequisites. This algorithm describes the processing of mapping information of similar type as given by the set  $\mathcal{MI}$  in this section. If there is partly training data of similar type as used in section 9.2, that is additionally geocoded, the processing proposed there can be applicated to get appropriate geocoded mapping information. The proposed algorithm is illstrated in figure 9.13 with respect to the extension for eventually regarding training data.

**Algorithm 9.3.4 (Bézier-surfaces basing on mapping information)**

- (BSM1) *Input: Geocoded mapping information, that means a number of mappings out of the space  $L^P(\tilde{X}, \mu)$  with associated coordinate information*
- (BSM2) *Extraction of the particular coordinate information*
- (BSM3) *Application of Delaunay-triangulation on that coordinate information*
- (BSM4) *Application of the algorithm of grade elevation for getting the ability of generating cubic Bézier-surfaces*
- (BSM5) *Generating Bézier-surfaces for the particular triangles within the convex hull of the given coordinates*



**Figure 9.13:** Workflow of algorithm for generating Bézier-surfaces basing on mapping information

The workflow for processing a query with a coordinate in the regarded convex hull is summarized in the following algorithm and illustrated in figure.

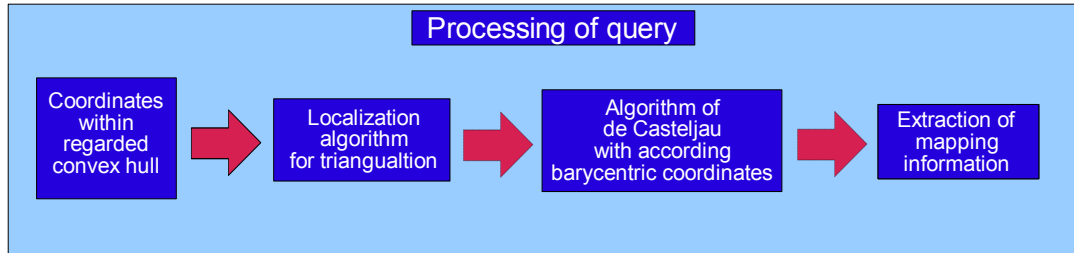
**Algorithm 9.3.5 (Processing a coordinate query)**

- (PCQ1) *Input: Coordinate information within the regarded convex hull*
- (PCQ2) *Localization using an appropriate algorithm, for example the one described in algorithm 9.3.3*

9 Results of measure theory applied on the “Spatial-Toolbox”

(PCQ3) Application of Casteljau algorithm on the barycentric coordinates of the given point in the localized triangle

(PCQ4) Extraction of the mapping information



**Figure 9.14:** Workflow for the processing of a coordinate query

In the last part of this section there should given an interpolation example using geocoded mapping information at four locations. Thereby, three geocoded fuzzy sets  $f_1, f_2, f_3 \in [0, 1]^{\mathbb{R}}$  are and a request for the resulting mapping at certain given coordinates are regarded. Because of the fact, that the convex combinations of mappings is done pointwise, the resulting mapping is an element of the space  $[0, 1]^{\mathbb{R}}$  as well. The following data should be given:

$$\mathcal{MI} := \{(f_1, c_1), (f_2, c_2), (f_3, c_3)\}$$

with:

$$\begin{aligned} f_1 : \mathbb{R} &\longrightarrow [0, 1] \\ x &\mapsto f_1(x) := (\sin(x) + 1) \cdot \frac{1}{2} \end{aligned}$$

$$\begin{aligned} f_2 : \mathbb{R} &\longrightarrow [0, 1] \\ x &\mapsto f_2(x) := \frac{1}{1+|x-3|} \end{aligned}$$

$$\begin{aligned} f_3 : \mathbb{R} &\longrightarrow [0, 1] \\ x &\mapsto f_3(x) := \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned}$$

## 9 Results of measure theory applied on the “Spatial-Toolbox”

and

$$c_1 := (3.49, 3.94), c_2 := (6, 3), c_3 := (5.8, 6.31)$$

To interpolate this data using the Bézier-methods, the following renaming is done:

$$b_{0,0,1}^{(1)} := (f_1, c_1), b_{1,0,0}^{(1)} := (f_2, c_3), b_{0,1,0}^{(1)} := (f_3, c_3)$$

The resulting triangle using the set  $\mathcal{MT}_L = \{c_1, c_2, c_3\}$  is illustrated as top view in figure 9.15 a). There, additionally the according mapping information is shown.

Using this information, the algorithm of grade elevation 8.3.2, what leads to the following:

$$\begin{aligned} \forall_{x \in \mathbb{R}} : b_{0,0,2}^{(1)} &:= b_{0,0,1}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{0,2,0}^{(1)} &:= b_{0,1,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{2,0,0}^{(1)} &:= b_{1,0,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{1,1,0}^{(1)} &:= \frac{1}{2}b_{0,1,0}^{(1)} + \frac{1}{2}b_{1,0,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{1,0,1}^{(1)} &:= \frac{1}{2}b_{0,0,1}^{(1)} + \frac{1}{2}b_{1,0,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{0,1,1}^{(1)} &:= \frac{1}{2}b_{0,0,1}^{(1)} + \frac{1}{2}b_{0,1,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{0,0,3}^{(1)} &:= b_{0,0,2}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{0,3,0}^{(1)} &:= b_{0,2,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{3,0,0}^{(1)} &:= b_{2,0,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{2,1,0}^{(1)} &:= \frac{2}{3}b_{1,1,0}^{(1)} + \frac{1}{3}b_{2,0,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{2,0,1}^{(1)} &:= \frac{2}{3}b_{1,0,1}^{(1)} + \frac{1}{3}b_{2,0,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{1,2,0}^{(1)} &:= \frac{1}{3}b_{0,2,0}^{(1)} + \frac{2}{3}b_{1,1,0}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{1,0,2}^{(1)} &:= \frac{1}{3}b_{0,0,2}^{(1)} + \frac{2}{3}b_{1,0,1}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{0,2,1}^{(1)} &:= \frac{2}{3}b_{0,1,1}^{(1)} + \frac{1}{3}b_{0,2,0}^{(1)} \end{aligned}$$



## 9 Results of measure theory applied on the “Spatial-Toolbox”

$$\begin{aligned} \forall_{x \in \mathbb{R}} : b_{0,1,2}^{(1)} &:= \frac{1}{3}b_{0,0,2}^{(1)} + \frac{2}{3}b_{0,1,1}^{(1)} \\ \forall_{x \in \mathbb{R}} : b_{1,1,1}^{(1)} &:= \frac{1}{3}b_{0,1,1}^{(1)} + \frac{1}{3}b_{1,0,1}^{(1)} + \frac{1}{3}b_{1,1,0}^{(1)} \end{aligned}$$

The resulting Bézier-points of grade 3 are illustrated in figure 9.16 in top view. It has to be pointed out, that only the components describing the coordinates are illustrated there.

Now, the point  $P$  with coordinates  $P := (4.36, 4.19)$  is regarded. It is assumed, that these coordinates within  $\text{conv}(\mathcal{M}\mathcal{I}_L)$  are given with the asking for the according mapping. This should be done terms of the Bézier-surface interpolation done with the calculated Bézier-points of grade 3.

The first step consists in calculating the normed barycentric coordinates of  $P$  with respect to  $c_1, c_2$  and  $c_3$ . This calculation can be done using areal relations as described in section 8.3 of chapter 8 or by solving a system of linear equations. At this point, the second way is gone. The according linear system of equations is solved using the CAS Maxima, which results in the following:

```
wxMaxima 0.8.3a http://wxmaxima.sourceforge.net
Maxima 5.19.2 http://maxima.sourceforge.net
Using Lisp CLISP 2.44.1 (2008-02-23)
Distributed under the GNU Public License. See the file COPYING.
Dedicated to the memory of William Schelter.
The function bug_report() provides bug reporting information.
```

```
Activating float mode
```

```
(%i1) if numer#false then numer:false else numer:true;
```

9 Results of measure theory applied on the “Spatial-Toolbox”

(%o1) true

Input data

(%i2) B001: [3.49, 3.94];

(%o2) [3.49, 3.94]

(%i3) B100: [6, 3];

(%o3) [6, 3]

(%i4) B010: [5.8, 6.31];

(%o4) [5.8, 6.31]

Query point

(%i5) P: [4.36, 4.19];

(%o5) [4.36, 4.19]

Solving the linear system of equations

(%i6) lgs:linsolve([lambda1\*B001[1]+lambda2\*B100[1]

+lambda3\*B010[1]=P[1], lambda1\*B001[2]+lambda2\*B100[2]

+lambda3\*B010[2]=P[2], lambda1+lambda2+lambda3=1],

[lambda1, lambda2, lambda3]);

9 Results of measure theory applied on the “Spatial-Toolbox”

(%o6) [lambda1 = .639204, lambda2 = .182806, lambda3 = .177990]

Checking the results

(%i7) rhs(lgs[1])\*B001+rhs(lgs[2])\*B100+rhs(lgs[3])\*B010;

(%o7) [4.36, 4.19]

For further calculating, the values  $\lambda_1 = 0.64$ ,  $\lambda_2 = 0.18$  and  $\lambda_3 = 0.18$  should be used.

The requested mapping at the given coordinates is now calculated using these normed barycentric coordinates in the algorithm of de Casteljau described in 8.3.1. The results are the following elements out of the space  $[0, 1]^{\mathbb{R}} \times \mathbb{R}^2$ :

$$\forall_{x \in \mathbb{R}} : b_{0,0,2}^{(1)1} := \lambda_1 b_{1,0,2}^{(1)} + \lambda_2 b_{0,1,2}^{(1)} + \lambda_3 b_{0,0,3}^{(1)}$$

$$\forall_{x \in \mathbb{R}} : b_{0,2,0}^{(1)1} := \lambda_1 b_{1,2,0}^{(1)} + \lambda_2 b_{0,3,0}^{(1)} + \lambda_3 b_{0,2,1}^{(1)}$$

$$\forall_{x \in \mathbb{R}} : b_{2,0,0}^{(1)1} := \lambda_1 b_{3,0,0}^{(1)} + \lambda_2 b_{2,1,0}^{(1)} + \lambda_3 b_{2,0,1}^{(1)}$$

$$\forall_{x \in \mathbb{R}} : b_{0,1,1}^{(1)1} := \lambda_1 b_{1,1,1}^{(1)} + \lambda_2 b_{0,2,1}^{(1)} + \lambda_3 b_{0,1,2}^{(1)}$$

$$\forall_{x \in \mathbb{R}} : b_{1,1,0}^{(1)1} := \lambda_1 b_{2,1,0}^{(1)} + \lambda_2 b_{1,2,0}^{(1)} + \lambda_3 b_{1,1,1}^{(1)}$$

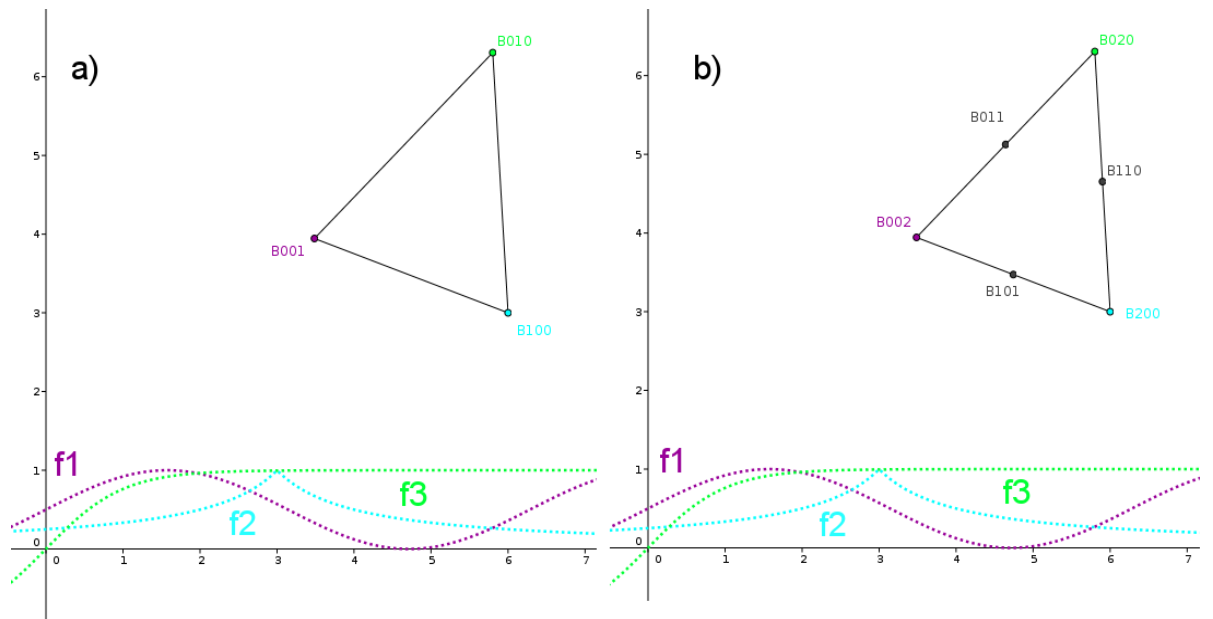
$$\forall_{x \in \mathbb{R}} : b_{1,0,1}^{(1)1} := \lambda_1 b_{2,0,1}^{(1)} + \lambda_2 b_{1,1,1}^{(1)} + \lambda_3 b_{1,0,2}^{(1)}$$

$$\forall_{x \in \mathbb{R}} : b_{0,0,1}^{(1)2} := \lambda_1 b_{1,0,1}^{(1)1} + \lambda_2 b_{0,1,1}^{(1)1} + \lambda_3 b_{0,0,2}^{(1)1}$$

$$\forall_{x \in \mathbb{R}} : b_{0,1,0}^{(1)2} := \lambda_1 b_{1,1,0}^{(1)1} + \lambda_2 b_{0,2,0}^{(1)1} + \lambda_3 b_{0,1,1}^{(1)1}$$

$$\forall_{x \in \mathbb{R}} : b_{1,0,0}^{(1)2} := \lambda_1 b_{2,0,0}^{(1)1} + \lambda_2 b_{1,1,0}^{(1)1} + \lambda_3 b_{1,0,1}^{(1)1}$$

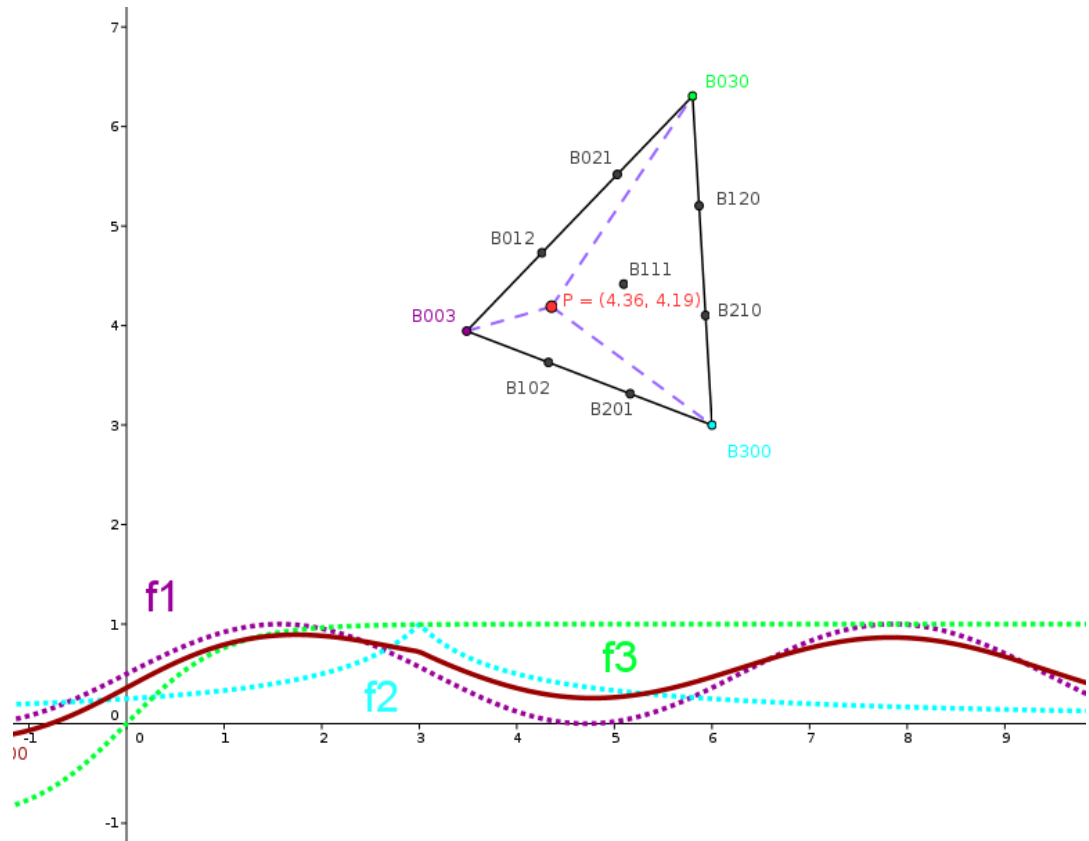
$$\forall_{x \in \mathbb{R}} : b_{0,0,0}^{(1)3} := \lambda_1 b_{1,0,0}^{(1)2} + \lambda_2 b_{0,1,0}^{(1)2} + \lambda_3 b_{0,0,1}^{(1)2}$$



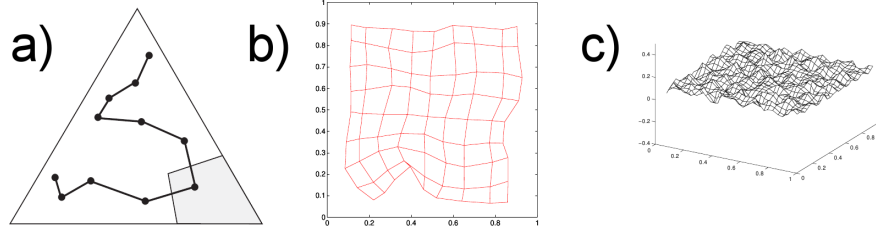
**Figure 9.15:** Bézier-points: a) grade 1, b) grade 2

The first component of  $b_{0,0,0}^{(1)3}$  represents the requested functional coherence at an arbitrary element  $x$  out of the domain  $\mathbb{R}$  of that mapping. An partial view on the graph of this mapping is illustrated in figure 9.16.

So, this example illustrated how a triangular Bézier-surface is generated using geocoded mapping information and how then a geocoded request is processed and the according mapping information is calculated in the proposed way.



**Figure 9.16:** Bézier-points of grade 3, point  $P$  with requested coordinates and calculated mapping



**Figure 9.17:** Different Kohonen topologies: a) one-dimensional Kohonen layer, triangular input domain, b) two-dimensional Kohonen layer, two-dimensional input domain, c) two-dimensional Kohonen layer, three-dimensional input domain

## 9.4 Application of Kohonen Networks in the SPATTB

In this subsection, there should be briefly proposed and described two possibilities of using Self Organizing Maps in the context of the SPATTB.

Regarding the Kohonen Maps described in section 4.3, a subspace  $U \subset \mathbb{R}^n, n \in \mathbb{N}$  is clustered by a Kohonen layer of a certain dimension and topology. Thereby, differences between input vectors and the codebook vectors containing the particular weights of the connections are regarded for modifying the codebook vectors. By this, a mapping of the space  $U$  is realised. In figure 9.17, there are shown codebook vectors according to a certain topology of a Kohonen layer clustering the given input domain.

At this point, an extension of the classical algorithm is proposed using the coherences described in section 9.2. There, a topological vector space of mappings is mapped by a function  $g^X$  to the space  $\mathbb{R}^n$ .

The topological vector space, that should be regarded, is the vector space  $X_F$  of mappings  $f: \tilde{X} \rightarrow \mathbb{R}$ :

## 9 Results of measure theory applied on the “Spatial-Toolbox”

$X_F := \{ f : \tilde{X} \longrightarrow \mathbb{R} \mid \tilde{X} \text{ is a topological vector space} \}$

The mapping  $g^X$  consists of  $n$  onedimensional integrals and can be denoted as it is done in the mentioned section:

$$g^X : X_{FI} \longrightarrow \mathbb{R}^n$$

$$f \longrightarrow g^X(f) := \mu(f) = \begin{pmatrix} \mu_1(f) \\ \vdots \\ \mu_n(f) \end{pmatrix}$$

In the context of Kohonen Maps, the combination of this mapping and the classical algorithm is proposed as an extension. By doing this, one gets the ability of clustering a subspace of the regarded  $X_F$  by preprocessing it. So, mappings can be used as inputs, they first are preprocessed in a certain way to  $n$  parameters, which now are the inputs for the classical algorithm.

In the second part of this subsection, an algorithm and software implementation for the calculation of so-called information densities are proposed. The context of this application should be described first, and then the algorithm is introduced.

The ability of giving suitable statements in the context of health, but also in other contexts, is obviously related to the available information. So, it is necessary to regard the original data and its relation before drawing a conclusion in a certain application. The regarded data should be geocoded data. That means data which is related to unique geographical coordinates. Some example for such data is the occurrence of malaria cases in certain areas. Regarding this kind of data, it is possible, to make statements like *“in area A, 70 percent of the population was infected by malaria”*, but it may be, that there is only information about a small number of persons living there. So, this statement is not very suitable. At this point, it is important to mention, that statements like this are part of a chain of

decisions, and so they could be a reason for a dissipation of valuable resources. So it is very important to be able to give a statement about the suitability of such conclusions using the amount of information in a certain area. This is called *density of data* in this thesis and the proposed calculation is described more precisely in the following. The calculation bases on the generating of so-called influence factors followed by a kind of interpolation for getting a continuous statement in the regarded area.

In general, interpolation methods are used to describe unknown functional coherence using given local information. In this field, different algorithms has been developed. Exemplarely, polynomial interpolation and various spline methods can be mentioned. For further information, the reader can refer to literature out of the field of numerical mathematics like [Spä95] or [Sch97].

The algorithm, that we used is based on gauss type mappings. The domain of the mapping that is interpolated consists of a subset of the space  $\mathbb{R}^2$  and represents the geo-coordinates of the given data. The codomain is the space  $\mathbb{R}$  and the particular value of that function at a given point  $(x, y) \in \mathbb{R}^2$  represents an influence factor on the neighbourhood. The computation of this value is described more detailed further down.

In the following, a set  $D$  of  $m$  threedimensional discrete datapoints is regarded. There, the first two components describe the according coordinates and the third one describes the influence factor.

It is  $D := \{(x_1, y_1, \mu_1), \dots, (x_m, y_m, \mu_m)\}$ .

Then, the resulting interpolating function  $I$  we used is defined as follows:

$$\begin{aligned} E : \quad \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto E(x, y) \end{aligned}$$

The following vectors should be regarded:



## 9 Results of measure theory applied on the “Spatial-Toolbox”

- Codebook vectors  $w_j, j \in \{1, \dots, m\}$
- Input vectors  $e_i, i \in \{1, \dots, n\}$

According to this, the distance between a certain codebook vector and an input vector can be written in the following way:

$$d(w_j, e_i)$$

From this distances, there is computed a new value. Our general idea of modeling the mentioned density value is, that every input vector has a certain influence on the density value of every codebook vector. The farther such an input vector is located from a certain codebook vector, that means the higher the distance is, the lower should the according influence be. This models the fact, that the information density is high in an area where many input vectors are located.

The according value for a certain input vector and a certain codebook vector is calculated as follows:

The first step is to calculate the value  $e_{ij}$ :

$$e_{ij} := \frac{1}{1 + d(w_j, e_i)}$$

Adding up all influence values for one certain input vector, there should be the sum of 1 and every single influence value represents the partial influence of the regarded input vector on a certain codebook vector. So, the value  $e_{ij}$  has to be divided by the sum of all values  $e_{ik}$  of the input vector  $e_i$ :

$$e_{is} := \sum_{k=1}^m e_{ik}$$

The influence value is the following:

$$\lambda_{ij} := \frac{e_{ij}}{e_{is}}$$

Regarding the sum of these, the required value 1 is calculated:

$$\sum_{j=1}^m \lambda_{ij} = \sum_{j=1}^m \frac{e_{ij}}{e_{is}} = \frac{e_{is}}{e_{is}} = 1$$

These  $\lambda_{ij}$  are the influence values for a certain input vector  $e_i$  and a certain codebook vector  $w_j$ . To describe the influence of all input vectors to a certain codebook vector, the following value  $\mu_j$  of the codebook vector  $w_j$  is calculated:

$$\mu_j := \frac{\sum_{i=1}^n \lambda_{ij}}{\sum_{i=1}^n e_{is}}$$

The whole system is built up using only OpenSource Software components. Doing this, it can be guaranteed, that there will not be any restriction by license for anyone who wants to use it.

One of the aims of the development of the EWARS is to make this application be usable for everyone, so a webinterface basing on the script language Perl and the UNIX bash script language was built up in this context. Because of this, it necessary, that all of the used applications can be used in a terminal mode.

The innovation consists in the computing of density mappings by the proposed software configuration. In the following, the role of each component is briefly described. Furthermore, there are shown our suggested possibilities of computing the mentioned data density needed for the interpolation.

The applications that are used, are the described in the following:

- Geographical Information System GRASS
- Statistical software environment and language R
- Computer Algebra System (CAS) Maxima
- Perl based Web interface

### **GIS GRASS**

The database of the GIS GRASS is used to store the given geocoded data and to represent it by maps. Doing this, we can use the advantage of this system, which is made for dealing with geographical data. In addition to that, it is possible to build maps which are based on different data sets and export them in bitmap (PNG, TIFF) or vector (SVG) graphics format.

### **Statistical software environment and language R**

The software environment R (available on [Ope09]) is made for statistical calculations. R is the OpenSource pendant to the commercial S or S Plus which was originally developed in the Bell Labs of ATT. There exists a huge community of developers similar to other OpenSource projects and a lot of useful extensions have been written. We decided to use R because it is a very powerful statistical application and there are special extensions concerning the field of artificial neural networks. So the *kohonen* extension makes it possible to analyze imported data using the Kohonen algorithm described in section 4.3.

One other advantage is the fact that using another extension called *spgrass6*, it is possible to directly import data stored in the database of the GIS GRASS. Hence, the combination of GRASS and R is a powerful tool for organizing geocoded data and statistically analyzing it in a very professional way.

### **Computer Algebra System Maxima**

With this CAS, the interpolation is calculated basing on the influence values generated in R. Furthermore, there are extracted sets of discrete values from this interpolation, which is exported to GRASS with the help of R for represent it then as geocoded information again.

The coherences between the different software components are shown in figure .

The according webinterface can be reached at [Wag].

In figure 9.19, the webinterface is illustrated in the actual version (May 2010) and in figures 9.20, 9.21 and 9.22, geocoded data from the used GRASS-database and the calculated and by the script produces density mappings in maxima and in GRASS are illustrated.

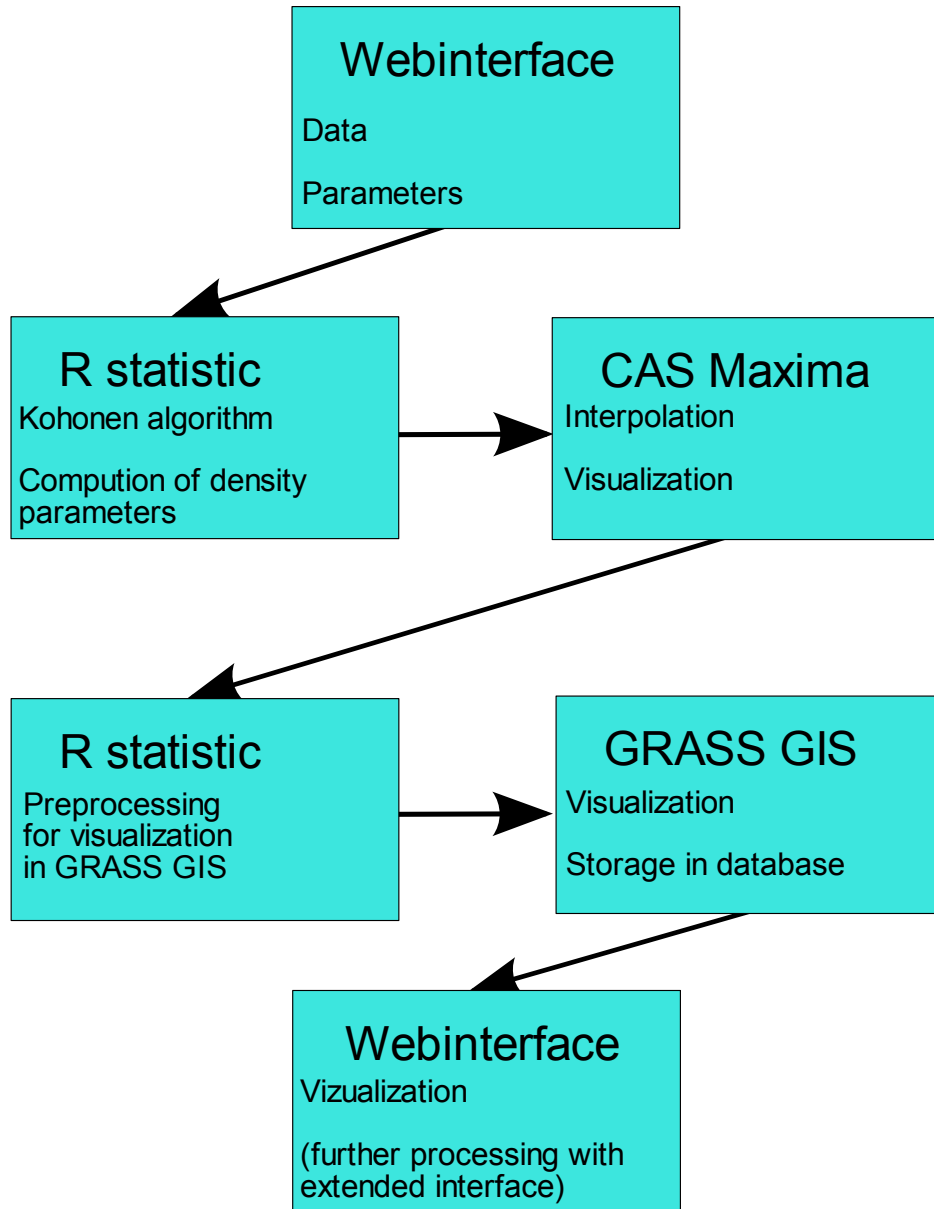
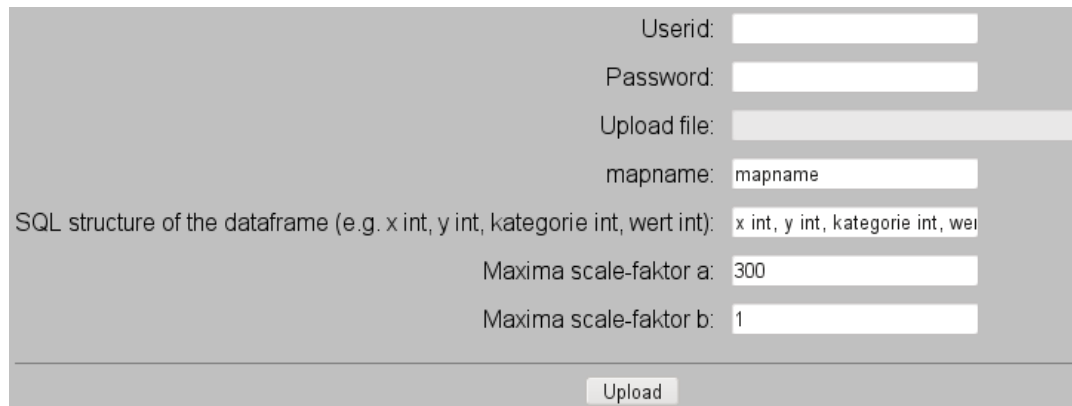


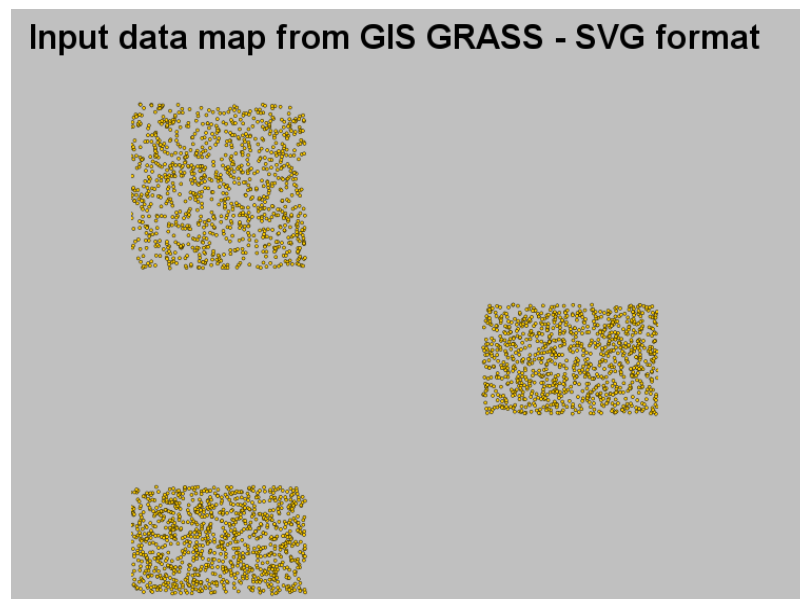
Figure 9.18: Coherences of used software configuration

9 Results of measure theory applied on the “Spatial-Toolbox”



The screenshot shows a web interface with a grey background. It contains several input fields and a button. At the top right, there are three input fields labeled 'Userid:', 'Password:', and 'Upload file:'. Below these, there is a 'mapname:' label followed by an input field containing the text 'mapname'. Further down, there is a label 'SQL structure of the dataframe (e.g. x int, y int, kategorie int, wert int):' followed by an input field containing 'x int, y int, kategorie int, wert int'. Below this, there are two more input fields: 'Maxima scale-faktor a:' with the value '300' and 'Maxima scale-faktor b:' with the value '1'. At the bottom center, there is a button labeled 'Upload'.

**Figure 9.19:** Developed webinterface



**Figure 9.20:** Basing geocoded data

9 Results of measure theory applied on the “Spatial-Toolbox”

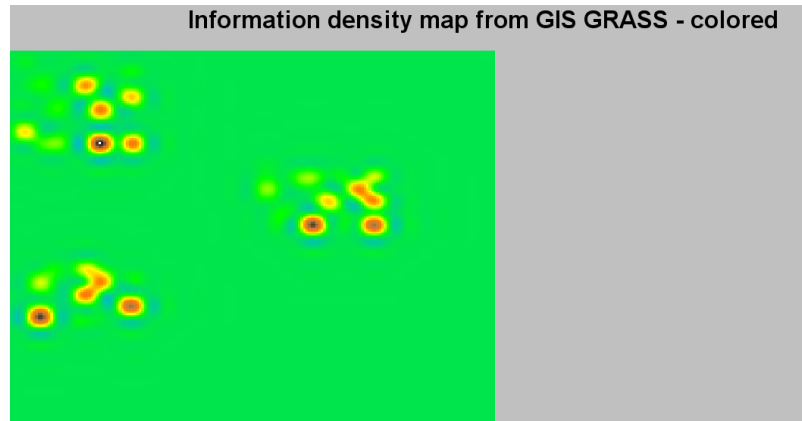


Figure 9.21: Information density mapping in maxima

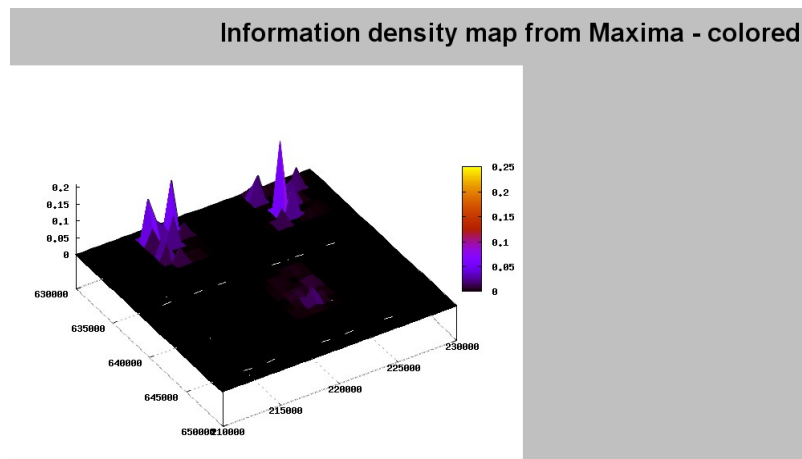


Figure 9.22: Information density mapping in GRASS

# 10 Conclusion

Within the course of this dissertation several approaches have been proposed and integrated within the concept of the EWARS as components of the SPATTB. This toolbox concept was described in chapter 6 as part of the EWARS.

The main contribution of this thesis to the scientific community is provided in chapters 8 and 9. In chapter 8, the conventional methodologies of parameter-based interpolating data has been extended from the case  $\mathbb{R}^d, d \in \mathbb{N}$  to the characterization of the coherences with respect to elements out of an arbitrary topological vector space and especially elements  $(x, y, f)$  out of a mapping space. This leads to the proposition of interpolating the geocoded mapping information made in chapter 9, what is not an obvious fact. An interesting result can be seen in the fact, that thereby a continuous interpolation of mappings is possible, that do not have to be differentiable itself. This generalisation has been done for Bézier-curves and for Bézier-surfaces especially for the case of triangular sets of parameters with respect to the possibility of structuring data by triangulation within a GIS.

In this context, in chapter 9, an algorithm of interpolating given geocoded mapping information is developed regarding the whole progress of structuring the data by Delaunay-triangulation, generating missing Bézier-points for interpolation using the algorithm of grade-elevation described in [Hos89] and finally processing queries basing on given coordinates with ask for the according mapping information using an generalised formulation of the algorithm of de Casteljaou.



## 10 Conclusion

An important achievement is the extension of the BPN- and the KN-algorithms, which can be scientifically located at the interface between mathematics and informatics. This is realised using methods from measure theory operating on mappings and the situation is exemplarily illustrated in figure 9.2 within chapter 9.2. The extension allows then the application of that algorithms on geocoded-mapping information.

With regard to the SPATTB with methods of Object Oriented Analysis described in chapter 5, these extended methods from the field of Artificial Neural Network represent subclasses of the main class SPATTB. Thereby, the mapping data provided to the system by a user is preprocessed by a mapping  $g^X$ , which works as a method in the context of these subclasses.

For future work, the implementation of any of these algorithms by adequate software solutions is planned for promoting the development of the EWARS. Thereby, the workflow proposed in chapter 6 is used. One implementation is proposed in section 9.4 basing only on OpenSource components. For this application, an improvement of the according webinterface and an integration of evaluation-methods with respect to real-data is intended to be achieved as following step of development.

With regard to the extended algorithms from the field of ANNs and the derived applications, an exact mathematical analysis of convergence regarding differences with respect to real data is thinkable and planned for further research.

Recapitulating these achievements, an important contribution to the development process of the EWARS is provided in this thesis, the further development of which will be worthwhile and prolific.

# List of Figures

2.1	EWARS, [Nie08], p.14 . . . . .	11
3.1	Fractal analysis of the urban perimeter of Cardiff ([San07], p.296 . .	17
3.2	Phases of decision making ([Tur00], p.50) . . . . .	21
3.3	Overview of the scientific field of AI with possible applications ([Tur00], p.545, modified) . . . . .	27
3.4	Key characteristics and capabilities of DSS ([Tur00], p.107) . . . . .	28
3.5	Schematic view on subsystems of a DSS ([Tur00], p.109) . . . . .	29
3.6	Proposed system configuration ([Yan07]) . . . . .	30
4.1	Modeling cycle ([Blu05]) . . . . .	36
4.2	Different occurrences of nerve cells ([G.99], S. 48) . . . . .	36
4.3	Important components of a nerve cell ([G.99], p. 92, modified) . . .	37
4.4	Action potential within an axon ([R.00], p. 168, modified) . . . . .	40
4.5	Chemical synapse ([Sil09], p. 51) . . . . .	42
4.6	Nerve cell network of the human eye retina ([Sil09], p. 351) . . . . .	43
4.7	Input, processing and output in general . . . . .	45
4.8	Coherence of the propagation-, activation- and output-mapping . .	48
4.9	Directed weigthened graph and according adjacency matrix . . . . .	51
4.10	Network topologies ([Zel96], p. 79, modified) . . . . .	52

*List of Figures*

4.11	Hierarchical structures of ANNs. Abbreviations: <b>syn:</b> synchronous activation, <b>asyn:</b> asynchronous activation, <b>sl:</b> supervised learning, <b>ul:</b> unsupervised learning, <b>rl:</b> reinforced learning . . . . .	54
4.12	Classification of the BPNs and the KNs in the regarded hierarchical structure . . . . .	55
4.13	Multi-layer <i>feed-forward-network</i> . . . . .	57
4.14	Adjacency scheme for generating the first weight matrix . . . . .	58
4.15	Situation within a BPN . . . . .	64
4.16	Weight within the last but one trainable layer . . . . .	66
4.17	Problems of the directed minimization ([Zel96], p. 113, modified) . . . . .	76
4.18	Exemplary SOM with two dimensional input ad a square Kohonen layer . . . . .	81
4.19	Kohonen Layer with block party distance . . . . .	84
4.20	A Graph of the function $\phi_{gauss}$ , B Illustration of the neighbourhood situation with radius 3 . . . . .	85
4.21	Representation process of a discrete input domain: A input data, B initial codebook vectors, C 10 iterations, D 30 iterations, E 100 iterations, F 30000 iterations . . . . .	86
6.1	Use cases related to the SPATTB . . . . .	92
6.2	Workflow for develop an additional component for the SPATTB . . . . .	95
7.1	Illustration of the defined separation axioms ([Que01], p.84) . . . . .	107
7.2	Overview of the relations between the "T"-spaces ([Que01], p.87) . . . . .	108
7.3	Hierarchical coherence of the regarded mathematical spaces . . . . .	110
7.4	Extended hierarchical structure of mathematical spaces . . . . .	150
8.1	Four cuttings of the graph of the curve $C_f$ . . . . .	173

*List of Figures*

8.2	Two tensorproduct Bézier surfaces with according Bézier-grid ([Ben03], p.172) . . . . .	181
8.3	Barycentric coordinates of a point $P$ . . . . .	184
8.4	Triangular Bézier surface of grade $n = 3$ and according triangular Bézier-grid, ([Hos89], p. 247, modified by the author) . . . . .	187
8.5	Topological order of Bézier-points: a) $n \in \mathbb{N}$ , b) $n = 3$ , ([Hos89], p. 246) . . . . .	188
8.6	Topological order of parameters: a) $n \in \mathbb{N}$ , b) $n = 3$ , ([Hos89], p. 247) . . . . .	189
8.7	Triangular Casteljaou algorithm ([Hos89], p.250 . . . . .	191
9.1	Discrete data for processing by BPN . . . . .	195
9.2	Mapping data for processing by BPN . . . . .	196
9.3	Overview of the proposed processing using a BPN in the context of discrete and non-discrete spatial data . . . . .	199
9.4	Detailed overview of the processing . . . . .	200
9.5	Linguistic variable Angle (according to a pendulums deflection) . . . . .	210
9.6	General architecture of a fuzzy-controller ([Nau94], p.246, modified by the author) . . . . .	212
9.7	Exemplarily data for preprocessing . . . . .	217
9.8	Necessarity of a separating edge chain: a) without separating edge chain, b) with separating edge chain ([Aum93], p.142) . . . . .	220
9.9	Convex hull, Voronoi-diagram and Delaunay-triangulation using GRASS: a) Points and according convex hull, b) Addition of the Voronoi-diagram, c) Addition of the Delaunay-triangulation . . . . .	223
9.10	a) Given data and according Delaunay-triangulation, b) Coherence between Voronoi-diagram and Delaunay-triangulation . . . . .	224

*List of Figures*

9.11	Threedimensional view: a) Voronoi-diagram and Delaunay-triangulation, b) Delaunay-triangulation . . . . .	224
9.12	Triangulations of four points: a) Three points in the according covex hull, b) Four points in the according covex hull . . . . .	227
9.13	Workflow of algorithm for generatining Bézier-surfaces basing on mapping information . . . . .	234
9.14	Workflow for the processing of a coordinate query . . . . .	235
9.15	Bézier-points: a) grade 1, b) grade 2 . . . . .	240
9.16	Bézier-points of grade 3, point $P$ with requested coordinates and calculated mapping . . . . .	241
9.17	Different Kohonen topologies: a) one-dimensional Kohonen layer, triangular input domain, b) two-dimensiona Kohonen layer, two- dimensional input domain, c) two-dimensional Kohonen layer, three- dimensional input domain . . . . .	242
9.18	Coherences of used software configuration . . . . .	249
9.19	Developed webinterface . . . . .	250
9.20	Basing geocoded data . . . . .	250
9.21	Information density mapping in maxima . . . . .	251
9.22	Information density mapping in GRASS . . . . .	251

# Index

## A

- algorithm of de Casteljau ..... 169
- algorithm of de Casteljau, Triangular  
187

## B

- Banachspace ..... 146
- Basis of a topological space ..... 97
- Border ..... 96
- Border point ..... 96

## C

- Characteristic function ..... 135
- Closure ..... 96
- Compactness ..... 107
- Complex radon measure ..... 130
- Continuity ..... 99
- Continuous mappings with compact  
support ..... 129

- Convergence of upper integrals .. 135

- Countability, first and second axiom  
of ..... 98

## D

- Dense ..... 96

## E

- Euler's polyeder theorem ..... 226
- Exterior ( $\mu$ )-measure ..... 135

## F

- Faces of a planar graph ..... 226
- Fatou, Lemma of ..... 142
- Filtered set of mappings ..... 133
- Fischer-Riesz, Theorem of ..... 148
- Fuzzy set ..... 207

## G

- Gauge ..... 111

*Index*

**H**

Hermitian form, positiv semidefinite  
146  
Hilbert space ..... 147

**I**

Inner product ..... 146  
integrability,  $\mu$ - ..... 145  
Integrable ..... 138  
Integral of integrable mappings .. 141  
Interior ..... 96  
Interior point ..... 96

**L**

Lebesgue, Theorem of ..... 143  
Linear functional ..... 128  
Linear operator ..... 128

**M**

Metric space ..... 108  
Minkowski functional ..... 111

**N**

Neighbourhood ..... 95  
Neighbourhood basis ..... 96  
Norm ..... 109

**O**

Open mapping ..... 101

**P**

P semi-norm ..... 137  
 $P^{th}$ -power integrable ..... 138  
Planar graph ..... 225  
Pointwise continuity ..... 99  
polarization identity ..... 147  
Positive linear functional ..... 130  
Positive linear functionals ..... 131  
Product space ..... 102  
Product topology ..... 102

**R**

Relative topology ..... 98

**S**

Semi-norm ..... 110  
Semicontinuity of mappings ..... 133  
Separation axioms ..... 104  
Subbase criterion of continuity ... 100  
Subbasis of a topological space .... 97  
Subspace ..... 98  
Support ..... 128  
Supremum semi-norm ..... 129

**T**

Theorem of the monotonous convergence (B. Levi) ..... 141  
Topological Bézier surface ..... 185  
Topological homeomorphism ..... 101  
Topological Space ..... 95  
Triangular Bézier surface ..... 184  
Tychonoff separation axioms ..... 104  
Tychonoff's theorem ..... 107

**U**

Upper integral ..... 134



# Bibliography

- [A.00] A., QUARTERONI: *Numerische Mathematik 1*. Springer, Berlin, 2000.
- [arg] *ArgoUML, OpenSource developing community*. Website. Available online at <http://argouml.tigris.org/>; visited on May 20th 2010.
- [Aum93] AUMANN, G. AND SPITZMÜLLER, K.: *Computerorientierte Geometrie*. BI Wissenschaftsverlag, Mannheim, 1993.
- [Bal99] BALZERT, H.: *Lehrbuch der Objektmodellierung*. Spektrum, 1999.
- [Bal00] BALAKRISHNAN, R. AND RANGANATHAN, K.: *A textbook of graph theory*. Springer, 2000.
- [Bal01] BALZERT, H.: *UML kompakt*. Spektrum, 2001.
- [Bar07] BARTSCH, R.: *Allgemeine Topologie I*. Oldenbourg Wissenschaftsverlag, München, 2007.
- [Bau01] BAUER, H. AND BURCKEL, R.B.: *Measure and integration theory*. Walter de Gruyter, Berlin, 2001.
- [Ben03] BENDER, M. AND BRILL, M.: *Computergrafik*. Hanser, München, 2003.
- [Ber12] BERNSTEIN, S.N.: *Démonstration du théoreme de Weierstrass fondée sur le calcul des probabilités*. Comm. Soc. Math. Kharkow, 13:1–2, 1912.
- [Ber99] BERTHOLD, M. AND HAND, D.J.: *Intelligent Data Analysis: An Introduction*. 1999.

## Bibliography

- [Blu05] BLUM, W. AND LEISS, D.: *Modellieren im Unterricht mit der "Tanken-Aufgabe"*. *mathematik lehren*, (128):18–21, 2005.
- [Boo06] BOOCH, G. AND RUMBAUGH, J. AND JACOBSON, I.: *Das UML-Benutzerhandbuch*. Pearson Education, 2006.
- [Bou04] BOURBAKI, N. AND BERBERIAN, S.K.: *Integration: Chapters 1-6*. Springer, Berlin, 2004.
- [BRJ95] BOOCH, G., J. RUMBAUGH I. JACOBSON: *Unified method for object-oriented development Version 0.8*. Rational Software Corporation, 1995.
- [Brü01] BRÜDERLIN, B. AND MEIER, A.: *Computergrafik und geometrisches Modellieren*. Vieweg+ Teubner, Stuttgart, 2001.
- [Bun02] BUNGARTZ, H.J. AND GRIEBEL, M. AND ZENGER, C.: *Einführung in die Computergraphik: Grundlagen, Geometrische Modellierung, Algorithmen*. Vieweg+Teubner, Braunschweig, 2002.
- [C.99] C., GEIGER: *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer, Berlin, 1999.
- [Can97] CANTOR, G.: *Beiträge zur Begründung der transfiniten Mengenlehre*. *Mathematische Annalen*, 49(2):207–246, 1897.
- [Col73] COLLATZ, L. AND KRABS, W.: *Approximationstheorie*. Teubner, Stuttgart, 1973.
- [Del34] DELAUNAY, B.: *Sur la sphere vide*. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7:793–800, 1934.
- [Deu93] DEUFLHARD, P. AND BORNEMANN, F.: *Numerische Mathematik. Eine algorithmisch orientierte Einfhrgung*. de Gruyter, Berlin, 1993.
- [Die00] DIESTEL, R.: *Graph theory , Graduate Texts in Mathematics 173*. Springer-Verlag, New York, 2000.

## Bibliography

- [Dir96] DIRSCHMID, H.J.: *Tensoren und Felder*. Springer, Vienna, 1996.
- [Dur89] DURFEE, E.H. AND LESSER, V.R. AND CORKILL, D.D.: *Trends in cooperative distributed problem solving*. IEEE Trans. Knowl. Data Eng., KOE-11, 1:63–83, 1989.
- [Edw94] EDWARDS, R. E.: *Functional analysis : theory and applications*. Dover Publications, New York, 1994.
- [Els05] ELSTRODT J.: *Maß- und Integrationstheorie*, 3. Springer, Berlin, 2005.
- [Fah09] FAHRMEIR, L. AND KNEIB, T. AND LANG, S.: *Regression: Modelle, Methoden und Anwendungen*. Springer, Berlin, 2009.
- [Far03] FARIN, G. AND HANSFORD, D.: *Lineare Algebra: Ein geometrischer Zugang*. Springer, Berlin, 2003.
- [Fis93] FISCHER G.: *Lineare Algebra*. Vieweg, Wiesbaden, 1993.
- [Fre06] FRECHÉT, M.M.: *Sur quelques points du calcul fonctionnel*. Rendiconti del Circolo Matematico di Palermo (1884-1940), 22(1):1–72, 1906.
- [G.99] G., THEWS: *Anatomie, Physiologie, Pathophysiologie des Menschen*. Wissenschaftliche Verlagsgesellschaft, Stuttgart, 1999.
- [GRA10] GRASS DEVELOPMENT TEAM: *Geographic Resources Analysis Support System (GRASS GIS) Software*. Open Source Geospatial Foundation, USA, 2010.
- [H.96] H., PENZLIN: *Lehrbuch der Tierphysiologie*. Fischer, Stuttgart, 1996.
- [Hag65] HAGGETT, P.: *Locational analysis in Modern Geography*. Arnold, London, 1965.
- [Hau14] HAUSDORFF, F.: *Grundzüge der Mengenlehre*. Verlag von Veit, Leipzig, 1914.
- [Hec87] HECHT-NIELSEN, R.: *Kolmogorovs mapping neural network existence*

## Bibliography

- theorem. Proceedings of the international conference on Neural Networks*, 3, 11–14. New York: IEEE Press, 1987.
- [Hec90] HECHT-NIELSEN, R.: *Neurocomputing: The Technology of Non-Algorithmic Information Processing*. Addison Wesley, Bonn, 1990.
- [Hec92] HECHT-NIELSEN, R.: *Theory of the backpropagation neural network*. Neural networks for perception (Vol. 2): computation, learning, architectures, 65–93, 1992.
- [Hil00] HILBERT, D.: *Mathematische probleme. vortrag, gehalten auf dem internationalen mathematiker-kongreß zu paris 1900*. Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, 253–297, 1900.
- [Hos89] HOSCHEK, J. AND LASSER, D.: *Grundlagen der geometrischen Datenverarbeitung*. Teubner, Stuttgart, 1989.
- [idr] *IDRISI GIS, Clark Labs, Clark University, Worcester, USA*. Website. Available online at <http://www.idrиси.com/>; visited on May 20th 2010.
- [iec] *INTERNATIONAL ELECTROTECHNICAL COMMISSION, Committee Draft CD 1.0 - IEC 1131 - PROGRAMMABLE CONTROLLERS, Part 7 - Fuzzy Control Programming*. Website. Available online at <http://www.fuzzytech.de/binaries/ieccd1.pdf>; visited on May 20th 2010.
- [Imb03] IMBODEN, D.M. AND KOCH, S.: *Systemanalyse: Einführung in die mathematische Modellierung natürlicher Systeme*. Springer, 2003.
- [Jän97] JÄNICH K.: *Topology*. Springer, Berlin, 1997.
- [Kar08] KARPFFINGER, C. AND MEYBERG, K.: *Algebra: Gruppen-Ringe-Körper*. Spektrum Akademischer Verlag, Heidelberg, 2008.
- [Kle07] KLEIN, B.: *FEM: Grundlagen und Anwendungen der Finite-element-*

## Bibliography

- methode im Maschinen-und Fahrzeugbau*. Vieweg+ Teubner, Stuttgart, 2007.
- [Koh82] KOHONEN, T. : *Self-Organized Formation of Topologically Correct Feature Maps*. Biological Cybernetics, 43:59–69, 1982.
- [Kol57] KOLMOGOROV, A.: *On the representation of continuous functions of many variables by superposition of continuous functions of one function and addition*. Dokl. Akad. Navk. USSR, 14:953–956, 1957.
- [Kow61] KOWALSKY, H.J.: *Topologische Räume*. Birkhäuser, 1961.
- [Kra91] KRATZER K.: *Neuronale Netze*. Hanser, München, 1991.
- [Kra97] KRABS, W.: *Mathematische Modellierung*. Teubner, Stuttgart, 1997.
- [Krä03] KRÄMER A. AND REINTJES R.: *Infektionsepidemiologie*. Springer, Berlin, 2003.
- [Kön97] KÖNIGSBERGER, K.: *Analysis 2*, 2. Springer, Berlin, 1997.
- [Kön04] KÖNIGSBERGER, K.: *Analysis 1*, 6. Springer, Berlin, 2004.
- [Lan07] LANG, S. AND BLASCHKE, T.: *Landschaftsanalyse mit GIS*. Ulmer, Stuttgart, 2007.
- [Lee03] LEE, S. AND RYU, J.H. AND LEE, M.J. AND WON, J.S.: *Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea*. Environmental Geology, 44(7):820–833, 2003.
- [Loo] LOOS H. AND FRITZKE B.: *DemoGNG - A java implementation of several competitive learning algorithms*. Website. Available online at <http://www.neuroinformatik.ruhr-uni-bochum.de/ini/VDM/research/gsn/DemoGNG/GNG.html>; visited on May 20th 2010.
- [Lor86] LORENTZ, G.: *Approximation of Functions*. Chelsea Publishing Company, Incorporated, Providence, 1986.

## Bibliography

- [Man82] MANDELBROT, B.B.: *The fractal geometry of nature*. W.H. Freeman and Company, New York, 1982.
- [McC43] MCCULLOCH, W.S. AND PITTS, W.: *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biology, 5(4):115–133, 1943.
- [Mei92] MEISE, R. AND VOGT, D.: *Einführung in die Funktionalanalysis*. Vieweg+Teubner, Braunschweig, 1992.
- [Min69] MINSKY, M. AND PAPERT, S.: *Perceptrons*. Cambridge, MA, 10:1–20,73, 1969.
- [Möb27] MÖBIUS, A.F.: *Der barycentrische Calcul: ein neues Hilfsmittel zur analytischen Behandlung der Geometrie: mit 4 Kupfertafeln*. JA Barth, 1827.
- [N.91] N., HOFFMANN: *Simulation Neuronaler Netze*. Vieweg, Braunschweig, 1991.
- [Naa97] NAAS, J. AND TUTSCHKE, W.: *Große Sätze und schöne Beweise der Mathematik: Identität des Schönen, Allgemeinen, Anwendbaren, 2*. Verlag Harri Deutsch, Thun, 1997.
- [Nau94] NAUCK, D. AND KLAWONN, F. AND KRUSE, R.: *Neuronale Netze und Fuzzy-Systeme*. Vieweg, Braunschweig, 1994.
- [Nem02] NEMATI, HAMID R. AND STEIGER, DAVID M. AND IYER, LAKSHMI S. AND HERSHEL, RICHARD T. : *Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing*. Decision Support Systems, 33(2):143–161, 2002.
- [Net08] NETELER, M. AND MITASOVA, H.: *Open source GIS: a GRASS GIS approach*. Springer, New York, 2008.

## Bibliography

- [Neu93] NEUNZERT H.: *Analysis 2*. Springer, Berlin, 1993.
- [Neu96] NEUNZERT H.: *Analysis 1*. Springer, Berlin, 1996.
- [Nie05] NIETZSCHE M.: *Graphen für Einsteiger*. Vieweg, Wiesbaden, 2005.
- [Nie08] NIEHAUS, E. (): *Early Warning: Tele-medicine and Logistical Response*, Vienna, February 2008. *Forty-fifth session of the Scientific and Technical Subcommittee*, Committee on the Peaceful Uses of Outer Space, United Nations. Available online at <http://www.oosa.unvienna.org/oosa/en/COPUOS/stsc/2008/presentations.html>; visited on May 20th 2010.
- [Nie09] NIEHAUS, E. AND HERSELMAN, M. AND BABU, A.N.: *Principles of Neuroempiricism and generalization of network topology for health service delivery*. Indian Journal of Medical Informatics, 4(1), 2009.
- [Nvi] NVIDIA: *PhysX*. Website. Available online at <http://developer.nvidia.com/object/physx.html>; visited on May 20th 2010.
- [Ope09] OPENSOURCE PROJECT: *R Statistic*. Website, 2009. Available online at <http://www.r-project.org/>; visited on May 20th 2010.
- [Oss09] OSSA E.: *Topologie: Eine anschauliche Einführung in die geometrischen und algebraischen Grundlagen, 2*. Vieweg+Teubner, Wiesbaden, 2009.
- [P.93] P., SPELLUCCI: *Numerische Verfahren der nichtlinearen Optimierung*. Birkhäuser, Basel, 1993.
- [Pat00] PATTERSON D.: *Künstliche neuronale Netze*. Prentice Hall, München, 2000.
- [Pei04] PEITGEN, H.O. AND JÜRGENS, H. AND SAUPE, D.: *Chaos and fractals: New frontiers of science*. Springer, New York, 2004.
- [Prü08] PRÜSS, J. AND SCHNAUBELT, R. AND ZACHER, R.: *Mathematische*

## Bibliography

*Modelle in der Biologie*. Birkhäuser Verlag, Basel, 2008.

- [Que01] QUERENBURG B.V.: *Mengentheoretische Topologie*, 3. Springer, Berlin, 2001.
- [R.00] R., ECKERT: *Tierphysiologie*. Thieme, Stuttgart, 2000.
- [Rak07] RAKTOTMANANA, FANJASO: *Determining areas that require indoor insecticide spraying using Multi-Criteria Evaluation , a decision support tool for malaria vector control programmes in the Central Highland of Madagascar*. International Journal of Health Geographics 2007, 6, January 2007.
- [RHW86] RUMELHART, D.E., G.E. HINTON R.J. WILLIAMS: *Learning representations by back-propagating errors*. Nature, 323:533–536, 1986.
- [Roj96] ROJAS, R. AND FELDMAN, J.: *Neural networks: a systematic introduction*. Springer, 1996.
- [Ros58] ROSENBLATT, F.: *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological review, 65(6):386–408, 1958.
- [Rud05] RUDIN, W.: *Analysis*, 3. Oldenburg Wissenschaftsverlag, München, 2005.
- [Sal06] SALOMON, D.: *Curves and surfaces for computer graphics*. Springer, New York, 2006.
- [San07] SANDERS LENA: *Geographical information systems series - Models in spatial analysis*. iSTE Ltd., London, 2007.
- [Sch96] SCHMITT, A. AND DEUSSEN, O. AND KREEB, M.: *Einführung in graphisch-geometrische Algorithmen*. Teubner, Stuttgart, 1996.
- [Sch97] SCHWARZ, H.R.: *Numerische Mathematik*, 4. Teubner Verlag,



## Bibliography

- Stuttgart, 1997.
- [Sch99] SCHAEFER, H.H. AND WOLFF, M.P.H.: *Topological vector spaces*, 2. Springer, New York, 1999.
- [Sil09] SILBERNAGL, S. AND DESPOPOULOS, A.: *Color atlas of physiology*. Thieme, New York, 2009.
- [Sim77] SIMON, H.A.: *The new science of management decision*. Prentice Hall, New Jersey, 1977.
- [Son01] SONAR, T.: *Angewandte Mathematik, Modellbildung und, Informatik: Eine Einführung für Lehramtsstudenten, Lehrer und Schüler. Mit Java-Übungen im Internet*. Vieweg+ Teubner Verlag, 2001.
- [Spä95] SPÄTH, H.: *Two dimensional spline interpolation algorithms*. A K Peters, Wellesley, Massachusetts, 1995.
- [Tur00] TURBAN, E. AND ARONSON, J.E.: *Decision Support Systems and Intelligent Systems*, 7. Prentice Hall, New Jersey, 2000.
- [Van77] VAN DELFT, A. AND NIJKAMP, P.: *Multi-criteria analysis and regional decision-making*. Springer, 1977.
- [Wag] WAGNER R.: *Calculation of information density by SOM*. Website. Available online at [http://mathematik.uni-landau.de/cgi-bin/informationdensity/upload\\_data\\_ascii.cgi/](http://mathematik.uni-landau.de/cgi-bin/informationdensity/upload_data_ascii.cgi/); visited on May 20th 2010.
- [Wer74] WERBOS, P.J.: *Beyond regression: New tools for prediction and analysis in the behavioral sciences, PhD-thesis*. , Harvard University, 1974.
- [Wer05] WERNER, D.: *Funktionalanalysis*. Springer, Berlin, 2005.
- [Woo02] WOOLDRIDGE M.: *Multitagent System*. John Wiley & Sons, London, 2002.

## Bibliography

- [Xjt] XJTEK: *AnyLogic*. Website. Available online at [http://www.xjtek.com/anylogic/demo\\_models/](http://www.xjtek.com/anylogic/demo_models/); visited on May 20th 2010.
- [Yan07] YANG, KUN AND PENG, S. AND XU, QUAN-LI AND CAO, YAN-BO: *A study on Spatial Decision Support Systems for Epidemic Disease Prevention Based on ArcGIS*, GIS for Health and Environment 1, 30–43. Lecture Notes in Geoinformation and Cartography. Springer, 2007.
- [Zad65] ZADEH, L.A.: *Fuzzy sets*. Information and control, 8(3):338–353, 1965.
- [Zad73] ZADEH, L.A.: *Outline of a new approach to the analysis of complex systems and decision processes*. IEEE transactions on Systems, Man, and Cybernetics, 3(1):28–44, 1973.
- [Zad75] ZADEH, L.A.: *The concept of a linguistic variable and its application to approximate reasoning–I 1*. Information sciences, 8(3):199–249, 1975.
- [Zel96] ZELL A.: *Simulation Neuronaler Netze*. Addison-Wesley, Bonn, 1996.
- [Zho96] ZHOU, J., CIVCO, D.L.: *Using genetic learning neural networks for spatial decision making in GIS*. Photogrammetric engineering and remote sensing, 62(11):1287–1295, 1996.

# Declaration

”I herewith declare that I have autonomously carried out the PhD-thesis entitled with “Components of a Spatial-Toolbox for processing geocoded mapping information in the context of decision support”. Any assistance used has been declared and the thoughts taken directly or indirectly from external sources are appropriately sourced. This or another thesis have never been previously submitted to another academic or non-academic institution for an exam, as a thesis or for evaluation in a similar context and has not yet been published.”

---

Place, Date

---

Signature