UiT The Arctic
University of Norway

Kristoffer Meyer Tangrand

Some new Contributions to Neural Networks and Wavelets with Applications

A dissertation for the degree of Philosophiae Doctor

Faculty of Engineering Science and Technology
Department of Computer Science and Computational Engineering

March 2023

Dedicated to my late grandparents

Abstract

In this Ph.D. thesis, we focus on some problems of general interest both in engineering sciences and applied mathematics. The close connection between some problems concerning neural networks, wavelets, structural health monitoring, and modern Fourier analysis is highlighted and applied in various ways. The main body of the Ph.D. thesis consists of six papers, A–F, which are put into a more general frame in the introduction.

In **Paper A** we present a case for how systematic use of energy flexibility can be an important instrument for managing peak loads and voltage problems in weak power grids. The FLEXNETT Simulator addresses production and energy dynamics down to every 10 minutes. A recurrent neural network was used to generate realistic values for the simulator.

In **Paper B** we made a case for using a combination of time series from non-intrusive ambient sensors and recurrent neural networks to predict room usage at a university campus. Training data was created by collecting measurements from ambient sensors measuring room CO_2, humidity, temperature, light, motion, and sound.

The findings in **papers A** and **B** led to inquiries concerning the learning ability of machine learning models.

In **Paper C** we propose a new approach to machine learning of geometric manifolds in R^n using single-layer or deep neural networks, Wavelet-Based Neural Networks (WBNN). Deep WBNNs provide a highly efficient computing architecture for the acceleration of the rate of convergence of the approximation process by using iterative algorithms.

The investigations in **paper C** inspired further research on actual engineering problems where, e.g., wavelets are of crucial importance.

In **Paper D** we investigate the impact of extreme arctic conditions on civil engineering infrastructures. Research and development of new methods are needed for damage detection in these structures. Advances in artificial intelligence could help solve the problem of structural damage detection, especially in arctic regions.

In **paper E**, a new example of the applications of operational modal analysis (OMA) techniques to a concrete railway arch bridge located over the Kalix river in Långforsen is presented. Results from the OMA techniques are used for finite element model (FEM) updating. Furthermore, artificial intelligence algorithms that can be useful for addressing the problem of missing data sets in structural health monitoring technologies are presented.

The questions discussed in **papers D** and **E** are not only related to neural networks and wavelets but also to modern Fourier analysis.

In **paper F** we prove some new inequalities and sharpness results concerning the Walsh-Fourier series. Moreover, a close connection between these series and wavelet theory is pointed out.

Preface

This Ph.D. thesis in Engineering Science is submitted in fulfillment of the requirements for the degree of Doctor of Philosophy at UiT The Arctic University of Norway. The research presented here was done under the supervision of Professor Bernt Bremdal and the co-supervision of Professor Børre Bang, Professor Lars-Erik Persson, Professor Lubomir Dechevsky, and Associate Professor Asbjørn Danielsen.

The main body of the Ph.D. thesis consists of six research articles, A – F, and a corresponding introduction. In the introduction, the papers are discussed and put into a more general frame. The introduction is also of independent interest since it contains a brief discussion on the important interplay between applied mathematics and engineering applications, illustrated by a comparison with some relevant international research presented in this light.

A brief description of the main content of the six papers can be found in the abstract above, and a more complete description is given at the end of the introduction.

List of Papers

Paper A: Kristoffer Tangrand and Bernt Bremdal, The FlexNett Simulator, *IOP Conference Series: Earth and Environmental Science (EES)*, Vol. 352, 2019, 8 pages.

Paper B: Kristoffer Tangrand and Bernt Bremdal, Using Deep Learning Methods to Monitor Non-Observable States in a Building, *Technical report, UiT The Arctic University of Norway*, 2020, 13 pages.

Remark: An abbreviated version was published in *Proceedings of the Northern Lights Deep Learning Workshop*, Vol. 1, 2020, 6 pages.

Paper C: Lubomir Dechevsky and Kristoffer Tangrand, Wavelet Neural Networks versus Wavelet-based Neural Networks, *Technical report, UiT The Arctic University of Norway, 2022*, 48 pages. *Submitted, available on Arxiv.org*

Paper D: Kristoffer Tangrand and Harpal Singh, Analysis of Civil Engineering Infrastructure in Norway With Solutions Based on Structural Health Monitoring and Artificial Intelligence, *To appear in Nonlinear Studies, 2023*, 15 pages.

Paper E: Kristoffer Tangrand, Harpal Singh, and Niklas Grip, A Comprehensive Study of Wavelets and Artificial Intelligence Algorithms for SHM and

its Application on a Concrete Railway Arch Bridge, *Technical report, UiT The Arctic University of Norway, 2022*, 17 pages. *Submitted to 14th International Workshop on Structural Health Monitoring September 12-14, 2023, Stanford University*, 17 pages

Paper F: David Baramidze, Lars-Erik Persson, Kristoffer Tangrand and George Tephnadze, $H_p - L_p$ Type Inequalities for Subsequences of Nörlund Means of Walsh-Fourier Series, *Accepted for publication by Journal of Inequalities and Applications, Springer Nature*, 16 pages

In addition to the papers above, the following paper is related to this Ph.D. thesis:

[*] Kristoffer Tangrand and Bernt Bremdal. "Using Ant Colony Optimization to determine influx of EVs and Charging Station capacities". Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON), 6 pages.

Acknowledgements

First of all, I express my deepest gratitude to my supervisors, Professor Bernt Bremdal, Professor Børre Bang, Professor Lars-Erik Persson, Professor Lubomir Dechevsky, and Associate Professor Asbjørn Danielsen, for introducing me to the topics covered in this Ph.D. thesis and for their invaluable support, advice, help, encouragement, and care during all of my work in this connection.

Secondly, I am indebted to the Faculty of Engineering Science and Technology at UiT The Arctic University of Norway, for providing me the economic opportunity to work with the questions studied and presented in this Ph.D. thesis.

Furthermore, I thank my co-authors for the fruitful collaborations. In particular, I want to thank Harpal Singh for his inspirational demeanor. Moreover, I want to thank my friends, who have always been there for me and provided me with help and support in any capacity they were able to.

Last but not least, my deepest thanks go to my parents, Olaf and Sylvi Tangrand, and my sisters and relatives for their love and support during this special time of my life.

Finally, I dedicate my Ph.D. thesis to my late grandparents, Knut and Anne-Lise Meyer. They passed away while I was working on my Ph.D.

Kristoffer Meyer Tangrand
Narvik, March 2023

Chapter 1

Introduction

1.1 Background

It is part of human nature to record and remember information. This is pure biology. Even the simplest life forms monitor and make sense of the world around them. This is performed using biological sensors, such as eyes, ears, or heat sensors. The information is then processed using a biological computer, commonly known as the brain. Monitoring their environment is crucial for all creatures.

Humanity has been collecting and recording information in various forms since the dawn of civilization – and probably even before that. In most cases, such information or data was used not only to remember the past but also to say something about the future. Throughout history, people have made and used things like written language, books, and, more recently, computers to record and store information. Today, enormous amounts of information are recorded, stored, and processed every day by computers.

In fact, many of the largest, wealthiest, and most influential companies in the world came into being due to their ability to collect, process, and structure information. To understand the data and extract meaningful insights, various mathematical and statistical techniques, known as algorithms, can be applied. These algorithms provide a set of precise steps for completing a task or making a decision based on a given situation.

1.2 Time Series

A common, universal procedure for making sense of information is to order it sequentially. For biological sensors and intelligence, this process is apparently done in an automated fashion that is not yet fully understood by neuroscience. However, for artificial sensors, it is common to collect the information in a timely fashion – simply leave the measurements in the order they were collected.

Such information or data is commonly referred to as *time series*. Since any chronologically ordered data can constitute a time series, they are ubiquitous. Data from fields such as statistics, signal processing, finance, and weather forecasting is usually represented as a time series, but most domains of applied science or engineering can make use of time series.

1.2.1 Time Series Analysis

Methods for extracting statistics and data characteristics are commonly called *time series analysis*. Methods for predicting future values of such chronological data are usually referred to as *time series forecasting*. Depending on the field, methods for analysis or forecasting are preferred. Commonly applied are methods from statistics for forecasting or signal processing methods for signal detection.

Some of the statistical properties of time series data include trend, seasonality, autocorrelation, and stationarity. Trend refers to the overall direction in which the data is moving, such as upwards or downwards. Seasonality refers to regular patterns that repeat over time, such as increased sales during the holiday season. Autocorrelation is the degree to which the data points are correlated with each other. Stationarity is the property of a time series in which the statistical properties do not change over time. These properties can be quantified and analyzed to better understand the underlying dynamics of the data.

These properties can be quantified and analyzed using statistical techniques such as time series decomposition, autocorrelation function plots, and stationarity tests. Time series decomposition involves breaking down the data into its trend, seasonality, and residual components. This can help identify the underlying patterns and trends in the data. Autocorrelation function plots can be used to visualize the degree of correlation between the data points. Stationarity tests can be used to determine whether a time series is stationary or not. These tests typically involve looking at the mean and variance of the data over time and checking for any significant changes.

Ergodicity is a property of a system that describes how its statistical properties evolve over time. A system is said to be ergodic if the long-term time average of a system's properties is equal to the average of the system's properties over a single time period. In other words, a system is ergodic if its statistical properties do not change over time.

For time series data, ergodicity is an important concept because it allows us to make predictions about the future behavior of the system based on its past behavior. If a time series is ergodic, we can make predictions about its future behavior by examining its past behavior and assuming that its statistical properties will remain the same over time. However, if a time series is not ergodic, we cannot make predictions about its future behavior based solely on its past behavior because its statistical properties may change over time.

However, approaches such as machine learning can be used for both purposes [1].

1.3 Machine Learning

The field of machine learning is devoted to building models that "learn." "Learning," in this context, means utilizing information in such a manner that the model, or algorithm, itself designs the proper set of rules or instructions for performing or improving upon some sort of task. Machine learning algorithms

build a model based on sample data, called training data, to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in many applications, such as medicine, speech recognition, and computer vision, where it is difficult or impossible to develop conventional algorithms to perform the necessary tasks.

1.3.1 Generalization and Learnability

The primary objective of a machine learning algorithm is to generalize from unknown information or experience. Generalization in this context means the ability of the algorithm to correctly classify or predict new and unseen data after having gone through a set of training data. The training examples come from a normally unknown probability distribution (which is considered representative of the occurrence space), and the learner must build a general model on this space that allows him to make accurate predictions in new cases. Because the training sets are finite and the future is uncertain, learning theory in general makes no guarantees about the performance of algorithms. If the hypothesis is less complex than the function, then the model is fully equipped with the data. This is further discussed and addressed in Paper C. Learning theorists investigate the time complexity and feasibility of learning in addition to performance limitations. In computational learning theory, a calculation is considered feasible if it can be performed in polynomial time.

1.3.2 Neural Networks

Some machine learning methods make use of data in ways that are similar to how biological brains work. The most common model is known as the neural network.

A neural network is based around a collection of connected units, or nodes, that receive or send signals between each other. These signals are real numbers, and all signals going to a node are (usually) summed upon arrival. The node then employs some sort of non-linear function to transform this sum. This transformed signal then becomes the output of the node. This output is then ready for transfer to connected nodes. The connections between nodes are often referred to as "weights" and are also represented by a real number. The weights change during evaluation to increase or decrease the signal strength between nodes. Usually, nodes and weights are organized into layers, but they do not have to be.

Since non-linearity is built into the neural network, it is able to reproduce and model non-linear processes. This leads to neural networks having almost endless application disciplines.

Since the output, or result, the neural network calculates is independent of its inner workings, it can be used for both regression and classification tasks. The concept of a neural network is also extendable to endless mathematical or structural changes.

1.4 Smart Buildings and Cities

Smart buildings and smart houses can be defined as a set of communication technologies enabling different sensors and functions within a building to communicate and interact with each other and also be managed, controlled, and automated in a remote way. Also, smart buildings and homes need to be sustainable on at least three fronts: the environment, the economy, and the quality of life for the people who live in them. Smart building projects are usually different from smart house projects because smart building projects are more business-to-business while smart house projects are more consumer-focused. Over the past years, overall EU energy expenditure has been reduced by 5-6% due to energy efficiency measures in buildings.

The main drivers are new and modern needs from users and owners, the automation of functions, and the search for better efficiency and productivity. There are also worries about using more renewable energy sources like photovoltaic (PV) panels, wind turbines, and generators made from bio-based materials. Smart buildings can be part of a larger ecosystem in which the building is linked to other parts of the smart city. The smart city is, in some ways, the macro-version of the smart building. Before, functions in a building were independent of each other. The smart building tries to link them all together so that they work better and use less energy. In the same way, the smart city uses digital solutions to connect different groups so they can work together, avoid duplication, and improve synergies. A major element of this emerging technology is artificial intelligence and machine learning. In a smart setting, energy efficiency and using non-fossil energy sources are the most important things. In May 2019, with effect from 2020, the EU introduced the "Clean Energy for All European Citizens" package. The idea of local energy markets and energy communities is part of this set of directives.

In paper A, the neural network architecture long short-term memory (LSTM) was investigated for the prediction of rooftop solar energy production in a neighborhood. The number of rooftop solar panels has been steadily increasing in recent years. Overproduction on sunny days can lead to households feeding electricity back to the grid. Most electricity grids are designed to feed electricity only one way, towards the customers. Thus, predicting production could lead to better control of production on local grids. A web application for simulating different neighborhoods was developed during this work but is no longer maintained.

The same neural network model (LSTM) was used in a different way in paper B, specifically to predict human occupancy numbers in university group rooms. These rooms are available for students to use throughout the day. As such, these rooms could see uneven usage patterns, such as students using them during overnight hours or for just a few hours at a time. Many modern buildings have sensors that measure the CO_2, humidity, temperature, sound level, and light level in the air. This monitoring is usually used to adjust ventilation and air conditioning as changes in sensor measurements are detected. Such measurements can also be collected and used with prediction algorithms. In

fact, the measurements are represented as a multivariate time series.

It is commonly known within the field of neural networks that collecting and curating training data can be challenging. One of these problems can be caused by how the network weights are set up. Usually, the weights are chosen at random. Depending on how the random numbers play out in the initial distribution, the training phase may go very smoothly or it may not converge at all. This is in relation to the distribution of the training dataset.

Another common caveat relates to deciding on network architecture. This is generally an unsolved problem, but approaches such as Neural Architecture Search exist. This approach, as can be assumed, is very computationally demanding and, in practical terms, not very feasible for large models. Some of these problems are discussed, and in some regards, addressed in paper C.

During the course of this paper, the problem of missing data, or in this case, irregular time series, came up. The sensors used to collect the indoor climate time series were not necessarily synchronized with the manual headcount used to establish the training data. As such, it was assumed that the machine learning models used for the prediction would learn biases because of the lack of synchronization. Paper E delves deeper into and describes this direction.

1.5 Structural Health Monitoring

The societal impact of solutions from science and technology is described in Section 1.4. Related to this, the creation and maintenance of public infrastructure represent an enormous cost for society, both in the public and private sectors. Structures sustain wear and tear during their operational lifetime due to a variety of environmental or human factors. Lack of maintenance and monitoring can result in the accumulation of damage over time, which can significantly reduce the performance of structures, cause changes in natural symmetry, or even lead to destruction. Civil engineering structures are typically designed to last 50 to 100 years. Structures are assumed to have the expected structural integrity during this lifetime. However, in general, structures are vulnerable to unforeseeable and unexpected damage caused by a variety of factors over the course of a structure's lifetime.

The deterioration of civil engineering infrastructures such as bridges, tunnels, and buildings causes numerous problems with significant consequences, both practically and economically. Governments and municipalities around the world must devote considerable resources to maintenance, repairs, or the construction of new structures to replace deteriorated or damaged ones in order to provide adequate service to citizens. Infrastructure maintenance costs for governments around the world are rising as many infrastructures near the end of their life cycles. Furthermore, the scarcity of expert labor to analyze such challenges exacerbates the problem. Analyzing these kinds of problems is very important, especially in places like northern Scandinavia, where harsh arctic conditions make them worse.

Large amounts of seafood cargo are exported along public roads in northern Norway. As of 2021, Norway's seafood industry exported 12 billion euros and contributed roughly 10% of Norwegian export earnings. Since the year 2000, the seafood industry has grown at a 7% annual rate, effectively doubling every ten years (see [2]) and [3]). The seafood industry is expected to continue growing at the same rate and has already seen 20 percent year-on-year growth in the first half of 2022 (see [4]). This is expected to increase the strain on an already – deteriorating infrastructure, particularly in sparsely populated northern Norway.

In Norway, the majority of civil engineering infrastructure operators and owners are municipalities or government-owned enterprises. Infrastructure asset management decisions are made for now based only on visual inspections. Non-destructive testing methods like acoustic, ultrasonic, or magnetic field testing could help with localized diagnosis. However, these testing methodologies have several limitations, including the inaccessibility of some parts of the structure, the inability to detect internal damage, the localization of the damage, and the difficulty of performing continuous monitoring with such techniques.

Structures vibrate as a result of natural or artificial excitations such as earthquakes, wind, or other vibrations. The output signals from these vibrations, such as accelerations, strains, or displacements, can be recorded as time series.

In a typical SHM system, sensors are placed all over the building and are used to calculate the structure's state. Damage is described as a deliberate or accidental alteration of a structure's material or geometric features, including alterations to boundary conditions or system connectivity that have an adverse effect on the structure's performance now or in the future (see [5]).

These signals have a non-stationary nature, which means that with time, their features change. But these signals can contain lots of information, which can be useful for stating the health of the structure. Windowing the signal can provide time localization with Fourier transforms, but it does not matter what the frequency component of the signal is.

1.6 Fourier Analysis

Since the first fundamental discoveries of Jean-Baptiste Joseph Fourier (1768-1830) in connection with his attempts to solve the heat equation, Fourier analysis has been greatly developed and applied.

The impact of Fourier analysis is heavily dependent on its many scientific applications, e.g. in physics, partial differential equations, number theory, combinatorics, signal processing, probability theory, statistics, forensics, option pricing, cryptography, numerical analysis, acoustics, oceanography, sonar, optics diffraction, geometry, protein structure analysis, etc. The standard variants of Fourier analysis are the Fourier transform, which is an integral transform, and the Fourier series. Other popular variants are the discrete-time Fourier transform and the fast Fourier transform. Fourier analysis has been very important

for the applications described in papers D and E. The most recent variant of Fourier analysis is that in the new book, [6]. Instead of having building blocks like sine and cosine, the basic elements here are sequences:

$x := (x_0, x_1, ..., x_j, ...)$ where $x_j = 0 \vee 1$

The study is done on a group G, which is the complete direct product of the group $z_2 := \{0, 1\}$. This new approach gives a theory that, in great parts, gives a similar theory as in the classical case, but there are many differences too (see [6]). This new approach seems to fit well with the problems of sequential data in papers A and B. It could also have implications for general computer science due to these being sequences of binary code. In paper F of this Ph.D. thesis, we have proven some inequalities and sharpness results that are new also vis-à-vis the new, fairly complete book [6].

1.7 Wavelet Theory

The Haar wavelet is a sequence of rescaled "square-shaped" functions that together form a wavelet family as its basis. Wavelet analysis is similar to Fourier analysis in that it allows a target function over an interval to be represented in terms of an orthonormal basis. These wavelets are named after the Hungarian mathematician Alfred Haar (1885–1933). The Haar system forms an orthonormal basis in the Lebesgue space L^2. For more information, see also the Appendix of paper D. After these first discoveries, it has been a fantastic development of the wavelet theory, and also in this case, the main reason is the great importance for applications, e.g. for most of those mentioned in Section 1.6. For several of these applications, wavelet theory has great advantages for various reasons. Also, in this case, there exists a continuous form (corresponding to the Fourier transform) and a discrete form (corresponding to the Fourier series). But as in the case of Fourier theory, there exist today a number of variants and generalizations in wavelet theory.

In papers D and E of this Ph.D. thesis, a number of these variants are discussed and applied, e.g. in connection to the serious problem of controlling and discovering in time, serious damage problems in bridges and buildings.

The most important theoretical contribution in this connection is found in Paper C. In particular, here we propose and describe a new machine learning approach combining neural networks with multiresolution wavelet analysis. It is described in very general and precise form, namely as the frame of a general Besov space (instead of L^2) and R^n, $n \in \mathbb{Z}$ (instead of $\mathbb{R}$).

1.8 A short description of the results in papers A - F

1.8.1 Paper A

The introduction of distributed, renewable energy sources has necessitated research into how production at the grid's terminal points should be best

catered for. The grid was never designed to handle a two-way flow of energy with production facilities at its terminal points.

In this paper, we describe a new analysis tool that we developed to study the impact of increasing rooftop solar installations in Norway. This simulation tool was created using geospatial data from the database of Norwegian buildings. A recurrent neural network (LSTM) was trained and applied to simulate the consequences of the high-density deployment of solar panels in different areas at Hvaler. The primary goal was to use the tool to investigate the impact of prosumers on the local distribution grid. By selecting a house or a group of houses linked to the same part of the grid, the tool generates the dynamics of loads for a single household as well as a neighborhood or a larger area.

We were able to provide evidence that the magnitude of peak loads on the local grid sections would not exceed the distribution limits. This is even with both large panels and a high number of local energy producers. We also showed that the local infrastructure for the grid areas studied at Hvaler (the target area) was robust enough to handle a high density of large rooftop panels. In this paper, we also determined the costs for single households using the power tariff introduced by the local distribution system operator, compared with the regular tariff. This analysis showed that the economic benefits of rooftop solar panels combined with peak hour consumption combined with power tariffs would be significant.

This paper cites the following sources: [7], [8], [9], [10], [11], [12], [13], [14], [15], and [16].

1.8.2 Paper B

With the advent of the Internet of Things (IoT), a multitude of monitoring and control opportunities have arisen. The development of smarter buildings, neighborhoods, and cities has already embraced this. Energy use and indoor climate control are central aspects related to the performance of buildings. Selective energy use can lead to more efficient buildings (see [17]). Monitoring the number of people in the specific rooms of a building can be used to achieve a more focused and efficient use of energy in a building. That in turn requires the ability to compare an estimate of space occupied with energy use. CO_2, illumination, and sound are known to be highly correlated with human occupancy (see [18]).

In this paper, we first present a correlation analysis between training features. Traditional algorithms like Pearson correlation and principal component analysis showed that the six different features were correlated to a mild degree. In particular, this fact implied that around 50% of variance could be retained in one component and that at least four components were needed to retain 90% variance. Because of this, we used all the data features for training the machine learning models.

Furthermore, we present results from a neural architecture search. This was done to investigate if certain architectures would perform better, given the

limited dataset. This search made it apparent that the activation function and loss function were the crucial parameters. Our search showed that rectified linear units and sigmoid units showed the same performance. For the loss function, mean squared error (L2 loss) was most effective.

We also present a search for feature selection. This was done by training on select combinations of features, i.e., only CO_2 and humidity. This search revealed that CO_2 was the feature contributing the most to the learning ability of the neural network.

Finally, we show that recurrent neural networks such as LSTM are more efficient than regular neural networks, support vector machines, and random forest regression at modeling the relationship for prediction for the sequential data in this paper.

This paper cites the following sources: [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37] and [38].

1.8.3 Paper C

In this paper, we present a new type of wavelet-based neural network (WBNN). We compare it to existing wavelet neural networks (WNNs) and demonstrate that WBNNs outperform WNNs. This superior performance is due in part to the advanced hierarchical tree structure of WBNNs, which is based on biorthonormal multiresolution analysis. Additionally, our new approach of incorporating wavelet tree depth into the neural width of the network allows for increased functionality and more efficient learning.

We show that WBNNs are able to efficiently learn not only regular distributions but also singular distributions like the Dirac delta and its derivatives. We also provide the general characteristics of the various activation operators that can be used in WBNNs and discuss the differences between non-threshold and threshold activation in learning fractal and piecewise smooth manifolds. We then introduce a new activation method based on the concept of decreasing rearrangement and provide proof of its consistency and optimality. Finally, we present four model examples and compare their results.

This paper cites the following sources:

[39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [12], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76] and [77].

1.8.4 Paper D

Damages to structures occur during their operational lifetime due to various environmental or human factors. Operators or owners of civil engineering infrastructure such as bridges, dams, and tunnels are mostly municipalities or government-owned enterprises in Norway. As for infrastructure assets, management decisions are based on visual inspections, which could be aided by lo-

calized diagnosis techniques such as the use of acoustic, ultrasonic, or magnetic field non-destructive testing methodologies. These testing methodologies have several limitations, such as inaccessibility to some parts of the structure and the inability to detect internal damage. For example, in a vibration-based SHM system, accelerometers are used to find the key parameters: mode shapes, mode frequencies, and mode damping. Once these parameters have been estimated, damage detection algorithms can be utilized. For example, in the case of a bridge, a label of critical damage, the need for inspection, or the need for maintenance can be assigned by comparing the bridge to data for a healthy bridge.

In this paper, we first describe and investigate this serious problem in northern Norway. As a basis we use a dataset from the Norwegian Public Road Administration. We conclude, from this dataset and according to internal classification by the said agency, that almost 1 in 10 bridges in Norway are in "critical" or "serious" states with regard to renovation. Moreover, almost half of the bridges are classified as "missing inspection", which is a mandatory regulation imposed by the agency itself. It is also apparent that around 1 in 3 bridges has had their planned renovation delayed. It is thus clear that the agency is not able to deliver on its own standards. Moreover, in this paper we put this struggling infrastructure into context in relation to the bustling Norwegian seafood industry. This seafood industry has quadrupled in economic magnitude in the last 20 years and is expected to continue to grow at the same rate.

In order to be able to handle this serious problem, we present an overview of the current state-of-the-art methods in structural health monitoring, such as finite element and operational modal analysis. We also give an overview of recently introduced techniques in machine learning in the context of structural health monitoring. We remark that very little work has been conducted in relation to vibration analysis, part of structural health monitoring using machine learning.

In this paper, we also present a pipeline for combining machine learning techniques with traditional monitoring techniques. Due to large amounts of data, machine learning techniques could assist in this aspect. This database can be accumulated over time and be used for training a mathematical framework or machine learning algorithms.

This paper cites the following sources: [2], [3], [4], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115], [116], [5], [117], [118], [119], [120], [121], [122], [123] and [124].

1.8.5 Paper E

As described in the previous article, bridge infrastructure in Norway is in a stressful situation due to various challenges. For example, in August 2022, the Tretten bridge catastrophically collapsed in Norway, where a truck and a

car became stuck. In another incident in May 2022, a bridge in Kvænangen municipality, Troms County, suffered serious damage, leading to the closure of a bridge over the important E6 European road. We conclude that more research is needed in this area so that precautionary measures can be taken to prevent such incidences in the future.

In this paper, we present a new example of the application of operational modal analysis techniques to a concrete railway arch bridge located over the Kalix river in Långforsen. Results from the operational modal analysis techniques are used for finite element model updating. During sensor data collection, wind speeds decreased drastically. This led to a greatly reduced signal-to-noise ratio, effectively crippling the data quality. The calibration of sensors also led to a decrease in data quality.

In this paper, we also present an overview of recent machine learning models inherently designed for handling missing data. It is revealed that such models are very recent and most have not been tested in application areas such as structural health monitoring. They do, however, show great promise for solving such missing data problems. We conclude that, however, such testing and bench-marking will require careful data collection and domain knowledge of the application area.

This paper cites the following sources: [39], [40], [109], [110], [111], [112], [117], [120], [121], [122], [123], [125], [126], [127], [128], [129], [130], [131], [132], [133], [134], [135], [136], [137], [138], [139], [140], [141], [142], [143], [144], [145], [146], [147], [148], [149], [150], [151], [152], [153], [154], [155], [156], [157], [158], [159], [160], [161], [162], [163], [164], [165] and [166].

1.8.6 Paper F

Fourier analysis is very important for various types of applications, see Section 1.6 and e.g. those in papers D and E of this Ph.D. thesis.

In this paper we make a new contribution in connection to the most modern form of Fourier analysis presented in the recent book [6] (with one of my supervisors as co-author). In particular, we investigate some new inequalities connected to the Walsh-Fourier series. The main result (see Theorem 1) shows that, in a special sense, the investigated inequalities are the sharpest possible. Moreover, we point out that there is a deep relation between the corresponding Walsh polynomials, Vilenkin groups, and wavelet frames (see [6] and especially [167] and [168]).

This paper cites the following sources: [169], [170], [171], [172], [173], [174], [175], [176], [177], [178], [179], [180], [181], [182], [183], [184], [185], [186], [187], [188], [189], [190], [191], [192], [193], [194], [195], [196],

Bibliography

[1] Bontempi, G., Ben Taieb, S., and Le Borgne, Y.-A. "Machine Learning Strategies for Time Series Forecasting". 2013, pp. 62–77.

[2] Fisk.no. *Sjømateksporten passerte 120 milliarder kroner i 2021*. https://fisk.no/fiskeri/7553-sjomateksporten-passerte-120-milliarder-kroner-i-2021, (accessed: 06.06.2022).

[3] Menon Economics. *EKSPORTMELDINGEN 2021*. https://www.menon.no/wp-content/uploads/2021-58-Eksportmeldingen-2021.pdf, (accessed: 06.06.2022).

[4] Riksrevisjonen. *Riksrevisjonens undersøkelse av overføring av godstransport fra vei til sjø og bane*. https://www.riksrevisjonen.no/globalassets/rapporter/no-2017-2018/godstransport.pdf, (accessed: 06.06.2022).

[5] Farrar, C. R. and Worden, K. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, 2012.

[6] Persson, L., Tephnadze, G., and Weisz, F. *Martingale Hardy Spaces and Summability of the One-Dimensional Vilenkin-Fourier Series*. Birkhäuser/Springer, 2022.

[7] *SINTEF Flexnett*. https://www.sintef.no/prosjekter/flexnett/, (accessed April 2019).

[8] Ożadowicz, A. "A New Concept of Active Demand Side Management for Energy Efficient Prosumer Microgrids with Smart Building Technologies". *Energies* vol. 10, no. 11 (2017), p. 22.

[9] Ottesen, S. et al. *Simplified battery operation and control algorithm*. Tech. rep. Deliverable D5.3 v.1, INVADE Project, (Available at https://h2020invade.eu/). 2017.

[10] Bremdal, B. A., Sæle, H., Mathisen, G., and Degefa, M. Z. "Flexibility offered to the distribution grid from households with a photovoltaic panel on their roof: Results and experiences from several pilots in a Norwegian research project". In: *2018 IEEE International Energy Conference (ENERGYCON)*.

[11] Bremdal, B. A. and Tangrand, K. *Investigations of prosumers' aggregated loads and energy flexibility at Hvaler(In Norwegian): Undersøkelse av aggregerte last- og fleksibilitetsbidrag fra plusskunder på Hvaler FLEXNETT rapport, Smart Innovation Norway*. Tech. rep. 2018.

[12] Schmidhuber, J. "Deep learning in neural networks: An overview". *Neural Networks* vol. 61 (2015), pp. 85–117.

[13] Feng, X. et al. “Topic-to-Essay Generation with Neural Networks”. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 2018, pp. 4078–4084.

[14] Martinsen, T. “A business model for an EV charging station with battery energy storage”. In: *CIRED Workshop 2016*.

[15] Nugroho, A., Rijanto, E., Wijaya, F. D., and Nugroho, P. “Battery state of charge estimation by using a combination of Coulomb Counting and dynamic model with adjusted gain”. In: *2015 International Conference on Sustainable Energy Engineering and Application (ICSEEA)*, pp. 54–58.

[16] Sæle, H. and Bremdal, B. “Economic evaluation of the grid tariff for households with solar power installed”. *CIRED - Open Access Proceedings Journal* vol. 2017, no. 1 (2017), pp. 2707–2710.

[17] Bremdal, B., Skjerve-Nielsen, C., and Nereng, G. “How the Prosumer Role in Smart Grids Redefines the Energy Efficiency Concept of Buildings”. In: World Sustainable Building Conference SB 2011, Helsinki.

[18] Ang, I. B. A., Dilys Salim, F., and Hamilton, M. “Human occupancy recognition with multivariate ambient sensors”. In: *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. 2016, p. 6.

[19] *Grand View Research: Bluetooth Beacons Market Analysis By Technology (iBeacon, Eddystone), By End-use (Retail, Travel & Tourism, Healthcare, Financial Institutions), By Region, And Segment Forecasts, 2018 – 2025*. 2017.

[20] Bockstael, N. and Jadin, A. “CO2 based room occupancy detection”. MA thesis. Ecole polytecnique de Louvain(EPL), 2018.

[21] Szczurek, A., Maciejewska, M., and Pietrucha, T. “Occupancy determination based on time series of CO2 concentration, temperature and relative humidity”. *Energy and Buildings* vol. 147 (2017), pp. 142–154.

[22] Danielsen, A. and Bremdal, B. “Predicting Bedside Falls using Current Context”. In: *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, p. 10.

[23] Kingma, D. and Ba, J. “Adam: A Method for Stochastic Optimization”. *International Conference on Learning Representations* (2014), p. 15.

[24] Hochreiter, S. and Schmidhuber, J. “Long short-term memory”. *Neural computation* vol. 9, no. 8 (1997), pp. 1735–1780.

[25] Cho, K. et al. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734.

[26] Autonomio Talos [Computer software]. *Talos*. Version 0.6.4. 2019. URL: http://github.com/autonomio/talos.

[27] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B., and Schmidhuber, J. "LSTM: A Search Space Odyssey". *IEEE Transactions on Neural Networks and Learning Systems* vol. 28 (2017), pp. 2222–2232.

[28] Chung, J., Gülçehre, C., Cho, K., and Bengio, K. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: *NIPS 2014 Workshop on Deep Learning*.

[29] Jiang, C., Masood, M. K., Soh, Y. C., and Li, H. "Indoor occupancy estimation from carbon dioxide concentration". *Energy and Buildings* vol. 131 (2016), pp. 132–141.

[30] Elsken, T., Metzen, J. H., and Hutter, F. "Neural Architecture Search: A Survey". *J. Mach. Learn. Res.* vol. 20, no. 1 (2019), pp. 1997–2017.

[31] Ang, I. B. A., Dilys Salim, F., and Hamilton, M. "Human occupancy recognition with multivariate ambient sensors". In: *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, p. 6.

[32] Singh, A. P. et al. "Machine Learning-Based Occupancy Estimation Using Multivariate Sensor Nodes". In: *2018 IEEE Globecom Workshops (GC Wkshps)*, p. 6.

[33] Ekwevugbe, T., Brown, N., Pakka, V., and Fan, D. "Real-time building occupancy sensing using neural-network based sensor network". In: *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*, pp. 114–119.

[34] Sultan, Z. et al. "Predicting occupancy counts using physical and statistical Co2-based modeling methodologies". *Building and Environment* vol. 123 (2017), pp. 517–528.

[35] Huang, Q. and Mao, C. "Occupancy Estimation in Smart Building using Hybrid CO2/Light Wireless Sensor Network". *Journal of Applied Sciences and Arts* vol. 1, no. 2 (2016), p. 13.

[36] Pearson, K. "LIII. On lines and planes of closest fit to systems of points in space". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* vol. 2, no. 11 (1901), pp. 559–572.

[37] Pearson, K. "Note on Regression and Inheritance in the Case of Two Parents". *Proceedings of the Royal Society of London Series I* vol. 58 (1895), pp. 240–242.

[38] Schmidhuber, J. "Deep learning in neural networks: An overview". *Neural Networks* vol. 61 (2015), pp. 85–117.

[39] Zhang, Q. and Benveniste, A. "Wavelet networks". *IEEE Trans Neural Networks* vol. 3, no. 6 (1992), pp. 889–898.

[40] Alexandridis, A. K. and Zapranis, A. D. "Wavelet neural networks: a practical guide". *Neural Networks* vol. 42 (2013), pp. 1–27.

[41] Chen, S., Cowan, C., and Grant, P. “Orthogonal least squares learning algorithm for radial basis function networks”. *IEEE Transactions on Neural Networks* vol. 2, no. 2 (1991), pp. 302–309.

[42] LeCun, Y., Bengio, Y., and Hinton, G. “Deep learning”. *Nature* vol. 521, no. 7553 (2015), pp. 436–444.

[43] Daubechies, I. *Ten Lectures on Wavelets*. USA: Society for Industrial and Applied Mathematics, 1992.

[44] Dahmen, W. “Wavelet and multiscale methods for operator equations”. *Acta Numerica* vol. 6 (1997), pp. 55–228.

[45] Bergh, J. and Löfström, J. *Interpolation spaces. An introduction*. Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York, 1976, pp. x+207.

[46] Dechevsky, L. and Penev, S. “On shape-preserving probabilistic wavelet approximators”. *Stochastic Analysis and Applications* vol. 15, no. 2 (1997), pp. 187–215.

[47] Cohen, A., Daubechies, I., and Vial, P. “Multiresolution analysis, wavelets and fast algorithms on an interval”. *C.R. Acad. Sci. Paris*. Ser. I Math. Vol. 316, no. 5 (1993), pp. 417–421.

[48] Cohen, A., Daubechies, I., and Vial, P. “Wavelets on the interval and fast wavelet transforms”. *Appl. Comp. Harmonic Anal.* vol. 1, no. 1 (1993), pp. 54–81.

[49] Cybenko, G. “Approximation by superpositions of a sigmoidal function”. *Math. Control Signal Systems* vol. 2.4 (1989), pp. 303–314.

[50] Cybenko, G. “Correction: "Approximation by superpositions of a sigmoidal function", [Math. Control Signal Systems, 2.4 (1989), pp. 303-314]”. *Math. Control Signal Systems* vol. 5.4 (1992), p. 455.

[51] Reed, M. and Simon, B. *Methods of Modern Mathematical Physics, Vol. 1: Functional Analysis. 2nd edition.* 2nd ed. Academic Press [Harcourt Brace Jovanovich, Publishers], New York, 1980.

[52] Lu, Z., Pu, H., Wang, F., Hu, Z., and Wang, L. “The expressive power of neural networks: A view from the width”. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[53] Dechevsky, L. and Penev, S. “On shape-preserving wavelet estimators of cumulative distribution functions and densities”. *Stochastic Analysis and Applications* vol. 16, no. 3 (1998), pp. 423–462.

[54] Triebel, H. *Theory of Function Spaces*. Vol. 78. Monographs in Mathematics. Birkhäuser Verlag, Basel, 1983.

[55] Samko, S. G., Kilbas, A. A., and Marichev, O. I. *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers, Yverdon, 1993.

[56] Frazier, M., Jawerth, B., and Weiss, G. *Littlewood-Paley Theory and the Study of Function Spaces*. Vol. 79. CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence, RI, 1991.

[57] Iba, H. *Evolutionary Approach to Machine Learning and Deep Neural Networks. Neuro-evolution and Gene Regulatory Networks*. Springer, Singapore, 2018.

[58] Iba, H. *Swarm Intelligence and Deep Evolution. Evolutionary Approach to Artificial Intelligence*. Taylor & Francis CRC Press, 2022.

[59] Hoyle, F. *The Black Cloud*. William Heinemann Ltd., 1957.

[60] Lem, S. *Niezwyciezony*. Wydawnictwo MON, 1964.

[61] Lem, S. *The Invincible*. The MIT Press, 2020.

[62] Leighton, F. T. *Introduction to Parallel Algorithms and Architectures. Arrays, Trees, Hypercubes*. Morgan Kaufman Publishers, Inc., San Mateo, CA, 1992.

[63] Dechevsky, L., Ramsay, J., and Penev, S. "Penalized wavelet estimation with Besov regularity constraints". *Math. Balkanica (N.S.)* vol. 13, no. 3-4 (1999), pp. 257–376.

[64] Dechevsky, L. "Atomic decomposition of function spaces and fractional integral and differential operators. TMSF, AUBG '99, Part A (Blagoevgrad)". *Fract. Calc. Appl. Anal.* vol. 2, no. 4 (1999), pp. 367–381.

[65] Okuta, R., Unno, Y., Nishino, D., Hido, S., and Loomis, C. "CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations". In: *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*. 2017. URL: http://learningsys.org/nips17/assets/papers/paper_16.pdf.

[66] Brenner, P., Thomée, V., and Wahlbin, L. B. *Besov Spaces and Applications to Difference Methods for Initial Value Problems*. Lecture Notes in Mathematics, No. 434. Springer-Verlag, Berlin-New York, 1975.

[67] Dechevsky, L. and Gulliksen, L. "A multirigid dynamical programming algorithm for discrete dynamical systems and its applications to numerical computation of global geodesics". *Int. J. Pure Appl. Math.* vol. 33, no. 2 (2006), pp. 257–286.

[68] Dechevsky, L. and Gulliksen, L. "Application of a multirigid dynamical programming algorithm to optimal parametrization , and a model solution of an industrial problem". *Int. J. Pure Appl. Math.* vol. 33, no. 3 (2006), pp. 381–406.

[69] Dechevsky, L., Person, L.-E., Singh, H., and Tangrand, K. *Learning non-parametric regression-functions and densities by univariate wavelet-based neural networks*. Tech. rep. 2022.

[70] Dechevsky, L., Person, L.-E., Singh, H., and Tangrand, K. *Learning of multidimensional geometric manifolds with wavelet-based neural networks*. Tech. rep. 2022.

[71] Jetten, K., Smale, S., and Zhou, D.-X. "Learning Theory and Approximation". In: *Mathematisches Forschungsintitut Oberwolfach Workshop, June 29th - July 5th. Oberfolfach Report 30/2008*. 2008, pp. 1655–1705.

[72] Binev, P., Cohen, A., Dahmen, W., and DeVore, R. "Classification algorithms using adaptive partitioning". *Ann. Statist.* vol. 42, no. 6 (2014), pp. 2141–2163.

[73] Petrushev, P. and Popov, V. *Rational Approximation of Real Functions*. Encyclopedia of Mathematics and its Applications, No. 28. Cambridge University Press, Cambridge, 1987.

[74] Dechevsky, L., Bratlie, J., Bang, B., Lakså, A., and Gundersen, J. "Wavelet-based lossless one- and two-dimensional representation of multidimensional geometric data". In: *AIP Conf. Proc.* Vol. 1410. Amer. Inst. Phys., Melville, NY, 2011, pp. 83–97.

[75] Dechevsky, L., Bratlie, J., and Gundersen, J. "Index mapping between tensor-product wavelet bases of different number of variables, and computing multivariate orthogonal discrete wavelet transforms on graphics processing units". *Lecture Notes in Comput. Sci.* vol. 7116 (2012), pp. 402–410.

[76] Binev, P., Cohen, A., Dahmen, W., and DeVore, R. "Universal algorithms for learning theory. II. Piecewise polynomial functions." *Constr. Approx.* vol. 26, no. 2 (2007), pp. 127–152.

[77] Hairer, M. and Labbé, C. "The reconstruction theorem in Besov spaces". *Journal of Functional Analysis* vol. 273, no. 8 (2017), pp. 2578–2618.

[78] Graps, A. "An Introduction to Wavelets". *IEEE Computational Science and Engineering* vol. 2, no. 2 (1995), pp. 50–61.

[79] Mallat, S. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. 3rd. USA: Academic Press, Inc., 2008.

[80] Hubbard, B. B. *The World According to Wavelets: The Story of a Mathematical Technique in the Making*. USA: A. K. Peters, Ltd., 1996.

[81] Fugal, D. L. *Conceptual Wavelets in Digital Signal Processing: An In-depth, Practical Approach for the Non-mathematician*. Space & Signals Technical Publications, 2009.

[82] Kaiser, G. *A Friendly Guide to Wavelets*. USA: Birkhäuser Boston Inc., 2011.

[83] Burrus, C. S., Gopinath, R. A., and Guo, H. *Introduction to Wavelets and Wavelet Transforms: A Primer*. Upper Saddle River, N.J: Prentice Hall, 1998.

[84] Meyer, Y. *Wavelets and Operators*. Cambridge University Press, 1992.

[85] Jaffard, S., Meyer, Y., and Ryan, R. *Wavelets: Tools for Science and Technology*. Society for Industrial and Applied Mathematics, 1987.

[86] Dechevsky, L., Grip, N., and Persson, L. "Sharp Error Estimates for Approximation by Wavelet Frames in Lebesgue Spaces". *Journal of Analysis and Applications* vol. 1, no. 1 (2003), pp. 11–31.

[87] Dechevsky, L. T. and Zanaty, P. "Triangular Beta-Function B-Spline Finite Elements: Evaluation and Graphical Comparisons". In: *Large-Scale Scientific Computing*. Springer Berlin Heidelberg, 2012.

[88] Goodfellow, I. et al. "Generative Adversarial Nets". In: *PAdvances in Neural Information Processing Systems (NIPS 2014)*. Vol. 27, p. 9.

[89] He, K., Zhang, X., Ren, S., and Sun, J. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

[90] Srivastava, R. K., Greff, K., and Schmidhuber, J. "Training Very Deep Networks". In: *Advances in Neural Information Processing Systems*. Vol. 28. 2015, p. 9.

[91] Rubanova, Y., Chen, R. T. Q., and Duvenaud, D. K. "Latent Ordinary Differential Equations for Irregularly-Sampled Time Series". In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019, p. 11.

[92] Shi, A. and Yu, X.-H. "Structural damage detection using artificial neural networks and wavelet transform". In: *2012 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*, pp. 7–11.

[93] Hochreiter, S. and Schmidhuber, J. "Long Short-Term Memory". *Neural Computation* vol. 9, no. 8 (1997), pp. 1735–1780.

[94] LeCun, Y., Bengio, Y., and Hinton, G. "Deep learning". *Nature* vol. 521 (2015), pp. 436–444.

[95] Silver, D. et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". *Science* vol. 362 (2018), pp. 1140–1144.

[96] Brown, T. et al. "Language Models are Few-Shot Learners". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 1877–1901.

[97] Dadman, S., Bremdal, B., and Tangrand, K. "The role of electric snowmobiles and rooftop energy production in the Arctic: The case of Longyearbyen". *Journal of Clean Energy Technologies* vol. 9, no. 4 (2021), pp. 46–53.

[98] Tangrand, K. and Bremdal, B. A. "Using ant colony optimization to determine influx of EVs and charging station capacities". In: *2016 IEEE International Energy Conference (ENERGYCON)*, p. 6.

[99] Tangrand, K. and Bremdal, B. "Using Deep Learning Methods to Monitor Non-Observable States in a Building". In: *Proceedings of the Northern Lights Deep Learning Workshop*. Vol. 1. 2020, p. 6.

[100] Tangrand, K. and Bremdal, B. "The FlexNett Simulator". *IOP Conference Series: Earth and Environmental Science* vol. 352 (2019), p. 8.

[101] Yaghoubi, V., Cheng, L., Paepegem, W. V., and Kersemans, M. "CNN-DST: Ensemble deep learning based on Dempster–Shafer theory for vibration-based fault recognition". *Structural Health Monitoring* vol. 21, no. 5 (2022), p. 20.

[102] Luleci, F., Catbas, F. N., and Avci, O. "Generative Adversarial Networks for Data Generation in Structural Health Monitoring". *Frontiers in Built Environment* vol. 8 (2022), p. 17.

[103] Kamada, S., Ichimura, T., and Iwasaki, T. "An Adaptive Structural Learning of Deep Belief Network for Image-based Crack Detection in Concrete Structures Using SDNET2018". In: *2020 International Conference on Image Processing and Robotics (ICIP)*, p. 6.

[104] Oring, A. "Fast and Robust Structural Damage Analysis of Civil Infrastructure Using UAV Imagery". In: *Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures*. Ed. by Pellegrino, C. et al. Cham: Springer International Publishing, 2022, pp. 1251–1260.

[105] Bukhsh, Z. A., Jansen, N., and Saeed, A. "Damage detection using in-domain and cross-domain transfer learning". *Neural Computing and Applications* vol. 33 (2021), pp. 16921–16936.

[106] Bai, Y., Zha, B., Sezen, H., and Yilmaz, A. "Engineering deep learning methods on automatic detection of damage in infrastructure due to extreme events". *Structural Health Monitoring (OnlineFirst)* (2022), p. 15.

[107] Soleimani-Babakamali, M. H., Soleimani-Babakamali, R., and Sarlo, R. "A general framework for supervised structural health monitoring and sensor output validation mitigating data imbalance with generative adversarial networks-generated high-dimensional features". *Structural Health Monitoring* vol. 21, no. 3 (2022), pp. 1167–1182.

[108] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. "Learning representations by back-propagating errors". *Nature* vol. 323 (1986), pp. 533–536.

[109] Singh, H. and Grip, N. "Recent trends in operation modal analysis techniques and its application on a steel truss bridge". *Nonlinear Studies* vol. 26, no. 4 (2019), pp. 911–927.

[110] Singh, H., Grip, N., and Nicklasson, P. J. "A comprehensive study of signal processing techniques of importance for operation modal analysis (OMA) and its application to a high-rise building". *Nonlinear Studies* vol. 28, no. 2 (2021), pp. 389–412.

[111] Singh, H. "The Hålogaland bridge - descriptions, challenges and related research under arctic condition". In: *9th International Operational Modal Analysis Conference, IOMAC 2022*, pp. 46–60.

[112] Singh, H. "Some new mathematical and engineering results connected to structural problems". PhD thesis. UiT The Arctic University of Norway, 2022.

[113] Verdens Gang. *The Neglected Bridges (in Norwegian)*. https://www.vg.no/spesial/2017/de-forsomte-broene/inspeksjoner/, (accessed: 05.09.2021).

[114] Verdens Gang. *The Neglected Bridges Map (in Norwegian)*. https://www.vg.no/spesial/2017/de-forsomte-broene/kart/inspeksjon, (accessed: 05.09.2021).

[115] Norsk Rikskringkasting (NRK). *Skader trugar bereevna på 1000 norske bruer (in Norwegian)*. https://www.nrk.no/vestland/1000-norske-bruer-har-skader-som-trugar-bereevna-1.15819652, (accessed: 03.06.2022).

[116] Amundsen, B. O. *Denne brua synger på absolutt siste verset: Nå må hus rives for å gi plass til ny (in Norwegian)*. https://www.veier24.no/artikler/denne-brua-synger-pa-absolutt-siste-verset-na-ma-hus-rives-for-a-gi-plass-til-ny/516133 (accessed: 16.06.2022).

[117] Grip, N., Sabourova, N., Tu, Y., and Elfgren, L. *Vibrationsanalys för tillståndsbedömning av byggkonstruktioner: Tillämpningsexempel:(Main results and summary in Swedish. Detailed results in English Appendices.)* Tech. rep. Luleå University of Technology , 2017.

[118] Grip, N., Sabourova, N., and Tu, Y. "Sensitivity-based model updating for structural damage identification using total variation regularization". *Mechanical systems and signal processing* vol. 84 (2017), pp. 365–383.

[119] Huang, Z. et al. "Modelling of damage and its use in assessment of a pre-stressed concrete bridge". In: *19th IABSE Congress, Stockholm 2016, September 21-23*, pp. 2093–2108.

[120] Akishev, G., Persson, L.-E., and Singh, H. "Inequalities for the Fourier coefficients in unbounded orthogonal systems in generalized Lorentz spaces". *Nonlinear Studies* vol. 27, no. 4 (2020), pp. 1137–1155.

[121] Akishev, G., Persson, L.-E., and Singh, H. "Some new Fourier inequalities and Jackson–Nikolskii type inequalities in unbounded orthonormal systems". *Constructive Mathematical Analysis* vol. 4, no. 3 (2021), pp. 291–304.

[122] Samko, N. and Singh, H. "A note on contributions concerning nonseparable spaces with respect to signal processing within Bayesian frameworks". *Mathematical Methods in the Applied Sciences* (2022), p. 7.

[123] Baramidze, D., Persson, L.-E., Singh, H., and Tephnadze, G. "Some new results and inequalities for subsequences of Nörlund logarithmic means of Walsh-Fourier series". *Journal of Inequalities and Applications* vol. 30 (2022), p. 13.

[124] Reynders, E. and De Roeck, G. "Reference-based combined deterministic–stochastic subspace identification for experimental and operational modal analysis". *Mechanical Systems and Signal Processing* vol. 22, no. 3 (2008), pp. 617–637.

[125] Moraud, E.M. *Wavelet Networks*. Tech. rep. "https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/martinmoraud.pdf".

[126] Lechner, M. and Hasani, R. *Learning Long-Term Dependencies in Irregularly-Sampled Time Series*. "https://arxiv.org/abs/2006.04418". 2020.

[127] Cho, K. et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Vol. abs/1406.1078. Doha, Qatar: Association for Computational Linguistics, 2014, pp. 1724–1734.

[128] Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. "Neural Ordinary Differential Equations". In: *Advances in Neural Information Processing Systems*. Vol. 31. 2018, p. 13.

[129] Brouwer, E. D., Simm, J., Arany, A., and Moreau, Y. "GRU-ODE-Bayes: Continuous Modeling of Sporadically-Observed Time Series". *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. 2019.

[130] Baytas, I. M. et al. "Patient Subtyping via Time-Aware LSTM Networks". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '17. 2017, pp. 65–74.

[131] Weerakody, P. B., Wong, K. W., Wang, G., and Ela, W. "A review of irregular time series data handling with gated recurrent neural networks". *Neurocomputing* vol. 441 (2021), pp. 161–178.

[132] Neil, D., Pfeiffer, M., and Liu, S.-C. "Phased LSTM: Accelerating Recurrent Network Training for Long or Event-based Sequences". In: *Advances in Neural Information Processing Systems*. Ed. by Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R. Vol. 29. 2016.

[133] Zhou, J. and Huang, Z. "Recover Missing Sensor Data with Iterative Imputing Network". In: *The Workshops of the The Thirty-Second AAAI Conference on Artificial Intelligence*. Vol. WS-18. AAAI Technical Report. AAAI Press, 2018, pp. 209–216.

[134] Vecoven, N., Ernst, D., and Drion, G. "A bio-inspired bistable recurrent cell allows for long-lasting memory". *PLoS ONE* vol. 16 (2021).

[135] Shukla, S. N. and Marlin, B. "Multi-Time Attention Networks for Irregularly Sampled Time Series". In: *International Conference on Learning Representations*. 2021.

[136] Horn, M., Moor, M., Bock, C., Rieck, B., and Borgwardt, K. "Set Functions for Time Series". In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 4353–4363.

[137] Zaheer, M. et al. "Deep Sets". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[138] Zhang, Y., Schlüter, A., and Waibel, C. *SolarGAN: Synthetic Annual Solar Irradiance Time Series on Urban Building Facades via Deep Generative Networks*. https://arxiv.org/abs/2206.00747. 2022.

[139] Shorten, C. and Khoshgoftaar, T. M. "A survey on Image Data Augmentation for Deep Learning". *Journal of Big Data* vol. 6, no. 1 (July 2019), p. 60.

[140] Yang, Z. et al. "SurfelGAN: Synthesizing Realistic Sensor Data for Autonomous Driving". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2020, pp. 11115–11124.

[141] Jeha, P. et al. "PSA-GAN: Progressive self attention GANs for synthetic time series". In: *ICML 2021 Time Series Workshop, ICLR 2022*. 2021.

[142] Nguyen, V. N., Jenssen, R., and Roverso, D. "Intelligent Monitoring and Inspection of Power Line Components Powered by UAVs and Deep Learning". *IEEE Power and Energy Technology Systems Journal* vol. 6, no. 1 (2019), pp. 11–21.

[143] Shukla, S. N. and Marlin, B. M. "Interpolation-Prediction Networks for Irregularly Sampled Time Series". In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9*.

[144] Graps, A. "An introduction to wavelets". *IEEE computational science and engineering* vol. 2, no. 2 (1995), pp. 50–61.

[145] Grip, N. "Wavelet and Gabor frames and bases: Approximation, sampling and applications". PhD thesis. Luleå University of Technology, 2002.

[146] Mallat, S. *A wavelet tour of signal processing*. Elsevier, 1999.

[147] Chen, H.-P. and Ni, Y.-Q. *Structural health monitoring of large civil engineering structures*. Wiley Online Library, 2018.

[148] Hou, Z., Noori, M., and Amand, R. S. "Wavelet-based approach for structural damage detection". *Journal of Engineering mechanics* vol. 126, no. 7 (2000), pp. 677–683.

[149] Kim, H. and Melhem, H. "Damage detection of structures by wavelet analysis". *Engineering structures* vol. 26, no. 3 (2004), pp. 347–362.

[150] Ruzzene, M., Fasana, A., Garibaldi, L., and Piombo, B. "Natural frequencies and dampings identification using wavelet transform: application to real data". *Mechanical systems and signal processing* vol. 11, no. 2 (1997), pp. 207–218.

[151] Rafiei, M. H. and Adeli, H. "A novel machine learning-based algorithm to detect damage in high-rise building structures". *The Structural Design of Tall and Special Buildings* vol. 26, no. 18 (2017), pp. 1–11.

[152] Daubechies, I., Lu, J., and Wu, H.-T. "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool". *Applied and computational harmonic analysis* vol. 30, no. 2 (2011), pp. 243–261.

[153] Mihalec, M., Slavič, J., and Boltežar, M. "Synchrosqueezed wavelet transform for damping identification". *Mechanical Systems and Signal Processing* vol. 80 (2016), pp. 324–334.

[154] Amezquita-Sanchez, J. P. and Adeli, H. "Synchrosqueezed wavelet transform-fractality model for locating, detecting, and quantifying damage in smart highrise building structures". *Smart Materials and Structures* vol. 24, no. 6 (2015), p. 15.

[155] Gilles, J. "Empirical wavelet transform". *IEEE transactions on signal processing* vol. 61, no. 16 (2013), pp. 3999–4010.

[156] Qu, H., Li, T., and Chen, G. "Adaptive wavelet transform: Definition, parameter optimization algorithms, and application for concrete delamination detection from impact echo responses". *Structural Health Monitoring* vol. 18, no. 4 (2019), pp. 1022–1039.

[157] Belsak, A. and Flasker, J. "Adaptive wavelet transform method to identify cracks in gears". *EURASIP Journal on Advances in Signal Processing* vol. 2010 (2010), p. 8.

[158] Grip, N. and Sabourova, N. "Simple non-iterative calibration for triaxial accelerometers". *Measurement Science and Technology* vol. 22, no. 12 (2011), p. 12.

[159] Forsberg, T., Grip, N., and Sabourova, N. "Non-iterative calibration for accelerometers with three non-orthogonal axes, reliable measurement setups and simple supplementary equipment". *Measurement Science and Technology* vol. 24, no. 3 (2013), p. 14.

[160] Teughels, A., Maeck, J., and De Roeck, G. "Damage assessment by FE model updating using damage functions". *Computers & structures* vol. 80, no. 25 (2002), pp. 1869–1879.

[161] Sabourova, N. et al. "The railway concrete arch bridge over Kalix river: dynamic properties and load carrying capacity". In: *FIB Symposium: Concrete Structures for Sustainable Community 11/06/2012-14/06/2012*. Swedish Concrete Association. 2012, pp. 609–612.

[162] Tangrand, K. and Singh, H. "Analysis of Civil Engineering Infrastructure in Norway With Solutions Based on Structural Health Monitoring and Artificial Intelligence". *Nonlinear Studies* (2022), p. 21.

[163] Rainieri, C. and Fabbrocino, G. *Operational modal analysis of civil engineering structures*. Springer, 2014.

[164] Brincker, R. and Ventura, C. *Introduction to operational modal analysis*. John Wiley & Sons, 2015.

[165] Wilhelms, H. and Grindstein, G. *Fare for at bro på E6 i Kvænangen kollapser: – Bare snakk om tid (in Norwegian)*. https://www.nrk.no, (accessed: 16.08.2022).

[166] Børresen, M. F., Trøen, M. I. N., Vogt, L. F., and Hågensen, K. *Vil fjerne brua i morgen - E6 fortsatt stengt (in Norwegian)*. https://www.nrk.no, (accessed: 16.08.2022).

[167] Farkov, Y. A., Lebedeva, E. A., and Skopina, M. "Wavelet frames on Vilenkin groups and their approximation properties". *Int. J. Wavelets Multiresolution Inf. Process.* vol. 13, no. 5 (2015), 1550036:1–1550036:19.

[168] Farkov, Y., Goginava, U., and Kopaliani, T. "Unconditional Convergence of Wavelet Expansion on the Cantor Dyadic Group". *Jaen Journal on Approximation* vol. 3, no. 1 (2011), pp. 117–133.

[169] Baramidze, D., Persson, L.-E., Tephnadze, G., and Wall, P. "Sharp $H_p - L_p$ type inequalities of weighted maximal operators of Vilenkin-Nörlund means and its applications". *Journal of Inequalities and Applications* vol. 242 (2016), p. 20.

[170] Baramidze, D., Persson, L.-E., Singh, H., and Tephnadze, G. "Some new results and inequalities for subsequences of Nörlund logarithmic means of Walsh-Fourier series". *Journal of Inequalities and Applications* vol. 30 (2022), p. 13.

[171] Blahota, I., Nagy, K., and Tephnadze, G. "Approximation by Θ-Means of Walsh-Fourier Series". *Anal. Math.* vol. 44, no. 1 (2018), pp. 57–71.

[172] Blahota, I., Nagy, K., and Tephnadze, G. "Approximation by Marcinkievicz theta-means of double Walsh-Fourier series". *Math. Inequal. Appl.* vol. 22, no. 3 (2019), pp. 837–853.

[173] Blahota, I., Persson, L.-E., and Tephnadze, G. "On the Nörlund means of Vilenkin-Fourier series". *Czechoslovak Math. J.* vol. 65, no. 4 (2015), pp. 983–1002.

[174] Fridli, S., Manchanda, P., and Siddiqi, A. "Approximation by Walsh-Nörlund means". *Acta Sci. Math. (Szeged)* vol. 74, no. 3-4 (2008), pp. 593–608.

[175] Gát, G. and Goginava, U. "Uniform and L-convergence of Logarithmic Means of Walsh-Fourier Series". *Acta Math. Sin., (Engl. Ser.)* vol. 22, no. 2 (2006), pp. 497–506.

[176] Goginava, U. "The maximal operator of the (C, α) means of the Walsh-Fourier series". *Ann. Univ. Sci. Budapest. Sect. Comput.* vol. 26 (2006), pp. 127–135.

[177] Goginava, U. "Almost everywhere convergence of subsequence of logarithmic means of Walsh-Fourier series". *Acta Math. Paed. Nyíreg.* vol. 21 (2005), pp. 169–175.

[178] Lukkassen, D., Persson, L., Tephnadze, G., and Tutberidze, G. "Some inequalities related to strong convergence of Riesz logarithmic means of Vilenkin-Fourier series". *Journal of Inequalities and Applications* (2020), p. 17.

[179] Móricz, F. and Siddiqi, A. "Approximation by Nörlund means of Walsh-Fourier series". *Journal of Approximation Theory* vol. 70, no. 3 (1992), pp. 375–389.

[180] Memić, N., Persson, L.-E., and Tephnadze, G. "A note on the maximal operators of Vilenkin-Nörlund means with non-increasing coefficients". *Studia Sci. Math. Hungar.* vol. 53, no. 4 (2016), pp. 545–556.

[181] Nagy, K. and Tephnadze, G. "Walsh-Marcinkiewicz means and Hardy spaces". *Cent. Eur. J. Math.* vol. 12, no. 8 (2014), pp. 1214–1228.

[182] Nagy, K. and Tephnadze, G. "The Walsh–Kaczmarz–Marcinkiewicz means and Hardy spaces". *Acta Math. Hungar.* vol. 149, no. 2 (2016), pp. 346–374.

[183] Nagy, K. and Tephnadze, G. "Strong convergence theorem for Walsh-Marcinkiewicz means". *Math. Inequal. Appl.* vol. 19, no. 1 (2016), pp. 185–195.

[184] Nagy, K. and Tephnadze, G. “Approximation by Walsh–Marcinkiewicz means on the Hardy space $H_{2/3}$”. *Kyoto J. Math.* vol. 54, no. 3 (2014), pp. 641–652.

[185] Persson, L.-E. and Tephnadze, G. “A Sharp Boundedness Result Concerning Some Maximal Operators of Vilenkin–Fejér Means”. *Mediterr. J. Math.* vol. 13, no. 4 (2016), pp. 1841–1853.

[186] Persson, L.-E., Tephnadze, G., and Wall, P. “Maximal operators of Vilenkin–Nörlund means”. *J. Fourier Anal. Appl.* vol. 21, no. 1 (2015), pp. 76–94.

[187] Persson, L.-E., Tephnadze, G., and Wall, P. “On the Nörlund logarithmic means with respect to Vilenkin system in the martingale Hardy space H_1”. *Acta Math. Hungar.* vol. 154, no. 2 (2018), pp. 289–301.

[188] Persson, L.-E., Tephnadze, G., and Tutberidze, G. “On the boundedness of subsequences of Vilenkin-Fejér means on the martingale Hardy spaces”. *Oper. Matrices* vol. 14, no. 1 (2020), pp. 283–294.

[189] Simon, P. “Strong convergence theorem for Vilenkin–Fourier series”. *J. Math. Anal. Appl.* vol. 245, no. 1 (2000), pp. 52–68.

[190] Simon, P. and Weisz, F. “Weak inequalities for Cesàro and Riesz summability of Walsh–Fourier series”. *J. Approx. Theory* vol. 151, no. 1 (2008), pp. 1–19.

[191] Tephnadze, G. “The maximal operators of logarithmic means of one-dimensional Vilenkin-Fourier series”. *Acta Math. Acad. Paedagog. Nyházi* vol. 27, no. 2 (2011), pp. 245–256.

[192] Tephnadze, G. “On the partial sums of Vilenkin-Fourier series”. *J. Contemp. Math. Anal.* vol. 49, no. 1 (2014), pp. 23–32.

[193] Tephnadze, G. and Tutberidze, G. “A note on The maximal operators of the Nörlund logaritmic means of Vilenkin-Fourier series”. *Trans. A. Razmadze Math. Inst.* vol. 174, no. 1 (2020), pp. 107–112.

[194] Weisz, F. *Martingale Hardy Spaces and their Applications in Fourier Analysis*. Springer, Berlin-Heidelberg-New York, 1994.

[195] Weisz, F. “Hardy spaces and Cesaro means of two-dimensional Fourier series”. *Bolyai Soc. Math. Studies* vol. 5 (1996), pp. 353–367.

[196] Weisz, F. “(C, α) summability of Walsh-Fourier series”. *Anal. Math.* vol. 27 (2001), pp. 141–156.

Papers

Paper A

The FlexNett Simulator

Kristoffer Tangrand, Bernt Bremdal

Published in *IOP Conference Series: Earth and Environmental Science (EES)*, Volume 352, 8 pages, 2019.

The FlexNett Simulator

Kristoffer Tangrand, Bernt Bremdal

PhD Research Fellow: IVT: UiT Narvik: Norway, Professor: UiT Narvik, Smart Innovation Norway: Norway

E-mail: kristoffer.tangrand@uit.no, bernt.bremdal@uit.no

Abstract. This paper documents research conducted in the Norwegian FLEXNETT project. It describes a new tool that was developed to study the future impact of prosumers with PV panels on the grid in Norway and the potential energy flexibility that lies with residential prosumers. Systematic use of energy flexibility can be an important instrument for managing peak loads and voltage problems in weak power grids. The influx of distributed energy resources can amplify this problem, but also help to resolve it. Self-balancing neighborhoods can be very attractive. This implies that loads related to energy demands can be curtailed and leveled out by different controllable devices or managed by using local energy production in the area to reduce the impact on the general distribution grid. The simulation tool is GIS based and can be applied to study the situation related to a single household, a neighborhood or in a specific transformer area. Unlike similar tools that address production yields over a period, the FLEXNETT Simulator addresses production and energy dynamics down to every 10 minutes. Due to the relatively low solar angle in Norway and rapidly changing weather these dynamics can be very prominent and induce local impact that is specific to a house or a neighborhood. The paper further describes how a recurrent neural network has been used as an engine to produce realistic values for the simulator.

1. Introduction

This work was conducted as part of the FLEXNETT project [1] lead by SINTEF Energi and with contribution of more than 20 industrial partners. The overall focus of the project was to investigate the present and future role of prosumers in the distribution grid. In particular, the role of private prosumers as instruments of energy flexibility for the local Distribution System Operator(DSO) was investigated. Energy flexibility relates to the possibility of reducing and leveling loads in order to avoid congestions in the electricity infrastructure and to maintain a high quality electricity supply. One of the central research questions specified was how prosumers with roof top based photovoltaic (PV) panels, with and without batteries, can help reduce peak loads in the grid. The grid was never designed to handle a two-way flow of energy with production facilities at its terminal points. The advent of distributed, renewable energy sources has called for research to understand how production at the grid's terminal points best should be catered for. As the use of PVs are still in its infancy in Norway it was considered important to determine new policies and methods early and be proactive to avoid situations that can cause problems for the operation of the grid. Experiences from southern Germany, where certain parts of the distribution grid have suffered uncontrolled and rapid peak accumulations of feeds, alternated by significant aggregated consumption loads triggered part of the research in FLEXNETT. High loads and rapid peak shifts can be a challenge. Congestions and voltage problems related to the

influx of distributed resources are not uncommon. Besides it has been shown that rapid peak shifts can cause life degradation on transformers. One important cause for these problems are related to the intermittency of solar and wind production. Another is that PV based production typically tops when consumption is low. This is quite common for residential areas where people spend time away from their house when the sun is at its highest and PV based production reaches its maximum.

A proactive investigation of PV based production and the role of prosumers in a Norwegian context was therefore called for. Some of the challenges experienced in southern Europe could be amplified further north due to geographical latitude and the low azimuth levels of the sun during larger parts of the year. The impact of local topography, vegetation and house architecture could cause greater local variations and together with cloud conditions cause even a higher frequency of alternating, aggregated loads in the distribution grid. At the same time opportunities were spotted. If local production could be organized so that local load balance could be maintained this could benefit the grid operations. Demand-response programs where loads are moved to periods of sunshine have been introduced and tested [2]. As the quality and price levels of batteries are improving they could also make a difference. Finally, pro-active planning could help to eliminate or alleviate the load issues, especially for new residential areas to be built where solar power is considered as an integral part of the construction project. But the same procedure could also be used to for existing residential areas in combination with a distributed set of batteries [3]. All of these issues were embraced by the FLEXNETT project. To answer part of this it was decided to create a novel simulation tool. Members of UiT tied up with Smart Innovation Norway to produce a solution that will be presented here. In the next paragraphs we will first explain some prerequisites for the work undertaken.

Then we will describe the tool created and how it was applied to produce relevant results. Finally, we will discuss the benefits and how we see the way forward.

2. Empirical Studies

Prior to the development of the FLEXNETT simulator empirical studies were conducted. The data used were harvested from the municipality of Hvaler in the county of Østfold in Norway. Hvaler was the first municipality in Norway where every residence was equipped with a smart meter connected to an AMS infrastructure deployed by Norgesnett, the local DSO. This happened in 2011. A substantial data history with a fair resolution was therefore available. This data allowed us to study the hourly load profiles of many different residences across the municipality. The double hump consumption profiles were quite uniform around the year with a distinct morning peak and a similar one in the late afternoon or early evening. However, there were certain differences with respect to when these peaks appeared. On the level of the principal substation these differences tended to even out, suggesting non-uniform behavior across areas and individual residences. However, within a single neighborhood the opposite seemed to be the case. Here, acute peaks were more common since the origin of the data received was not disclosed for the study we could only compare with consumption profiles of people we were able to get closer too. These were residential owners who took a special interest in the project. A statistical analysis showed, not surprisingly, that there was a relationship between demographic parameters and the type of consumption profile obtained. But we also found that many of the neighborhoods represented in the area were populated by people in the same age group and with matching daily routines. This could be related to the fact that many of these neighborhoods, mostly single residences and chained houses, were established at distinct points in time as the needs for housing emerged. Thus, older neighborhoods were typically inhabited by older and retired people, while recent property developments were dominated by families with smaller children. Hvaler was also the first municipality in Norway with a distinct population of residences with roof top panels. An active and facilitating policy introduced in

2015 created a set of incentives for residents to invest in panels. A small subset of these shared their production data with the project from the start, thus making it possible to create a history on local production, albeit not more than records for 20 units. To determine the conditions for production we studied the local topography and weather conditions for each one. For half of these we were allowed to inspect the property physically to specify features like height and angle of the roof and its orientation with respect to vegetation and other structures. The records harvested, and the observations made were used for a wider set of project purposes, some of which have been published elsewhere [4, 5]. One important conclusion from these studies was that production varied significantly for different neighbors in the same area and with the same type and size of roof top PV panels. It was apparent that local conditions causing shadows, both permanent, arbitrary and periodic made an impact. On an aggregated level this increased the intermittency, although much of this could be considered periodic.

One of the objectives of FLEXNETT was to carry out analyses to create the basis for the FLEXNETT simulation tool. The basic task was to examine empirical data and to use it to develop a tool to study the non-observed and the non-existent. In other words, we would use the tool to study what would happen if whole new areas were equipped with rooftop panels. What impact would a high-density deployment of PVs cause and how can it potentially be exploited for the benefit of the grid? A simulation tool was an obvious answer, if not the sole option. There was enough time series to power the generation of records for the simulations. But the set with production data was significantly leaner. These records represented far less households, despite significant histories for each of them. Standard statistical methods were considered but were rejected in favor of machine learning techniques based on a form of recurrent neural networks called the Long Short Term Method (LSTM).

3. Tool Making

One reason that standard statistical methods were rejected for the simulation was the inclusion of circumstantial data such as local weather and contextual elements representing causes for non-permanent shadows. All of this influence the rate of production and thus the degree of intermittency. Hence, the current state of production may, in the general case, not honor the Markov assumption. In other words, the current input does not hold all the information needed to compute the next event. LSTMs represent a type of deep learning techniques [6] that have proven useful for different applications, including time series and text analysis. Like most regression techniques these type of neural networks are typically used for prediction. And as such they can also be used for generating series of data based on the history [7]. The FLEXNETT problem was treated as a Partially Observable Markov Decision Process (POMDP). Each house in the municipality and their future production history could be considered the full state space where only part of it could be observed. Like other vanilla neural networks, a recurrent network can be seen as a function approximator for such a universe. But a recurrent network like LSTM produces a dense state specification as a function of the entire history. This is due to their ability to carry out operations over sequences of vectors. In addition, it combines the vector containing the input with a state vector that in addition to the output vector creates a new state vector. Hence neural networks like the LSTM can be effectively used as a world model generator. The LSTM in FLEXNETT was constructed accordingly. The output $y(t, R)$, representing the production at any given time of the year, $\mathbf{t}$, of any residence $\mathbf{R}$, equipped with a solar panel would then be approximated with the LSTM representing the function $\mathbf{f}(P_1, P_2, P_3, P_4, ..., P_n, R(PV, x_1, x_2, ..., x_n))$. Here, P_i, PV, x_i represent historic production at specific times, peak panel production and a vector of geographical and topographical parameters x_i, respectively. Different versions of the part vector $R(PV, x_1, x_2, ..., x_n))$ were attempted, but problems with convergence was experienced. The final version of R was limited to PV panel size and physical orientation in 3 dimensions. The LSTM was trained on the time series collected

and tested on a part of the collected data history separated from the training set.

Since PV production follows the sunlight, it has an inherent Sin Wave-like property. As such, variations in production would come as a consequence of physical objects existing between the sun rays and the PV panel. Such objects could come in the form of weather, such as clouds, and stationary or mobile ground objects. The LSTM should be able to account for ground objects simply from the recurring change in production. For the weather, including data from nearby PV panels should include information about changing weather conditions. A multi-dimensional training data consisting of production times series from multiple PVs was tested to see if it could provide better predictions than using only the time series from one single PV. For any given installation P_i, its training data consisted of time series $P_{i1}, P_{i2}, P_{i3}, ...$ as well as time series from other PVs P_j, P_k, P_l in the area. The training data was organized in sliding windows(batches) of 36 time steps (6 hours X 6 values per hour), where the first (35 x number of PVs) values were input. The sliding windows was shifted by 1. The output then, was a single floating point value that represents the predicted next production value.

Real consumption data was applied. The rich data set made available to the project permitted this. Time series from the complete data set was chosen at random. Bias was further reduced by introducing multiple replications for the simulations. One issue was the difference in time resolution. As the training data was sampled on a 10-minute interval this became the standard for the simulations too. Since the meters provide on hourly data for the consumption we used that. For the purpose at hand it sufficed to use average data across the hour. In the future higher resolution consumption data is desired. A graphical interface was developed to control the simulation. It included a Geographical Information System (GIS) section that also provided information to the simulation engine (see Figure 1). By selecting a house or a group of houses linked to the same part of the grid the tool would generate the dynamics of loads for a single household as well as a neighborhood or a larger area. Similarly, it would be possible to choose simulations for a particular period i.e. day, month, season or a full year and the results could be shown in real time or produced as a batch.

Figure 1. The FLEXNETT user panel showing time series for consumption and production on the left, the GIS panel in the middle where a selection of houses belonging to the same grid connection is investigated. Right a set of tabs can display aggregated power and cost data for the group and per building.

A selection of houses could also be equipped with batteries. For the economic calculations, battery degradation was taken into account, but the battery management was held very simple.

Battery would always have priority to surplus when fully or partly discharged. Degradation was based on the rainfall method [8]. Beyond this we applied a Coloumb Counting [9] method to monitor the state of charge. Economic calculations used the principal grid tariff at Hvaler introduced by Norgesnett as default. This includes a power part [kW]. The tariff for regular households is shown in Table 1 below. But the tool can be also used to analyze economic impacts due to tariffs with fixed and energy related fees only.

Table 1. Tariffs introduced by Norgesnett at Hvaler. Note the power related part

Fixed fee(NOK/year)	Period	Energy part NOK/100*kWh	Power part NOK pr. max hour per month [NOK/100*kW]
625	May-Oct	26.36	61.25
0	Nov-April	28.23	61.25

4. Results

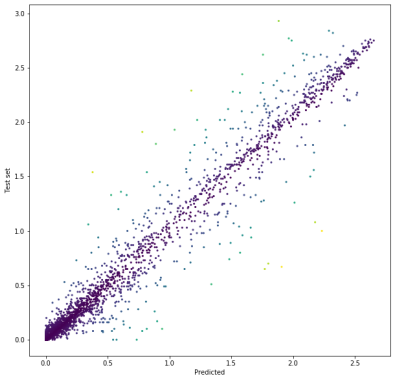


Figure 2. Scatter plot shows predicted values (x-axis) vs. actual values in test set (y-axis).

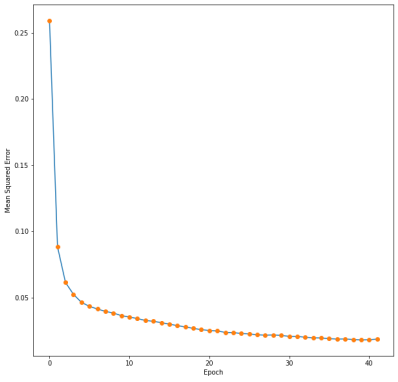


Figure 3. : Learning rate- showing the reduction of the Mean Square Error (y-axis) per training epoch (x-axis).

Figure 2 and Figure 3 shows the result of the training of the LSTM. The method showed good fit and acceptable error rates (MSE). It should be noted that the potential for overfit was present due to the limited sample size available. The network was trained until standard overfitting cutoff measures kicked in. Figure 4 shows the graphs produced for aggregated loads across time for a group of houses connected to the same point of coupling.

The tool was used to determine the impact of PV size on the grid and the potential role of the battery. The orientation of PV-panels on self-consumption was also investigated with the tool

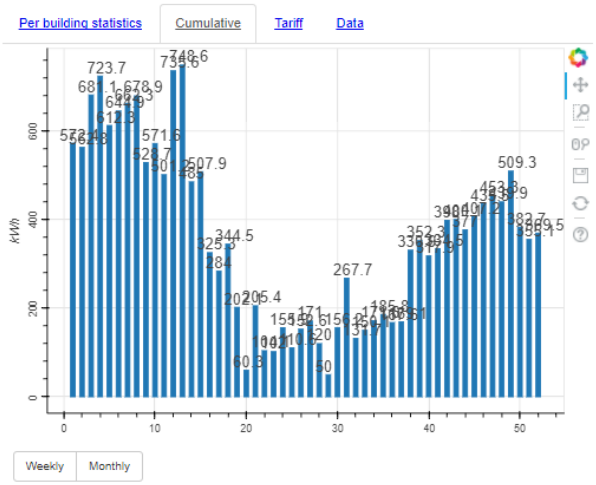


Figure 4. The cumulative net loads (kWh/h) for different hours for a period of app. 50 hours.

[3]. For the selection of households shown in Figure 2 the maximum loads on the neighboring transformer was calculated. Examples are given for three cases: No production facility, all houses equipped with 3,1kWp and 9,3kWp panels.

Table 2. Aggregated net loads for a selection of houses (kWh/h) (see Figure 1) equipped with roof top panels. Min specifies the minimum net load. Max the hour during the year with highest accumulated consumption.

Panel Size	0 kWp	3.1 kWp	9.3 kWp
Min	-9.94	31.58	86.61
Max	-118.93	119.43	-129.45

As can be observed from Table 2, the panel sizes need to be large in order to reach the magnitude of negative loads typically during the month of January with no production. During the warm and sunny part of the year the simulations also showed huge differences in max and min from minute to minute. The tool thus provided evidence that the local infrastructure for the grid areas studied at Hvaler were robust enough to handle a high density of large rooftop panels. Another issue studied was the impact of house orientation. Figure 5 shows the production profiles based on actual metering in the month of June for three different houses in the same neighborhood at Hvaler with PV panels oriented at 106, 182 and 200 degrees. In addition, and arbitrary consumption profile is shown for comparison in order to highlight self-consumption issues and the possibility of self-balancing by design. As can be observed, the easterly oriented panel meet the morning peak of the consumption better than the other two. But the westerly and southerly oriented profile meets the afternoon/evening peak better. The westerly oriented panel could have better absorbed the consumption if it was not cutoff as the curve in Figure 5 shows. This illustrates the susceptibility to local topography and vegetation that obstruct the sun at early and late hours during the summer. Using the tool to investigate the degree of

self-consumption related to PV orientation the easterly or westerly evening peaks did not match the consumption as well as expected. One explanation is that the neighborhoods investigated was inhabited by a several retired couples. (See Table 3). The dominant morning and evening peaks in the pilot areas at Hvaler came later than expected compared to the average records for the country and that generally show an earlier morning peak and a a more distinct afternoon peak. This fact favored the panels facing south. With the power tariffs introduced the economic benefit of the PV panels increased significantly for the prosumers. The tool further highlighted the need for diversity in orientation. An average neighborhood in Norway with a high density of panels facing south are more likely to accumulate high peaks at noon than areas with roof tops pointing slightly different ways. Moreover, with power tariffs it makes economic sense to adjust panels according to the households consumption profile. But it is important that production does curb the consumption at the hour with the highest energy use. For more on the economic part of FLEXNETT see [10].

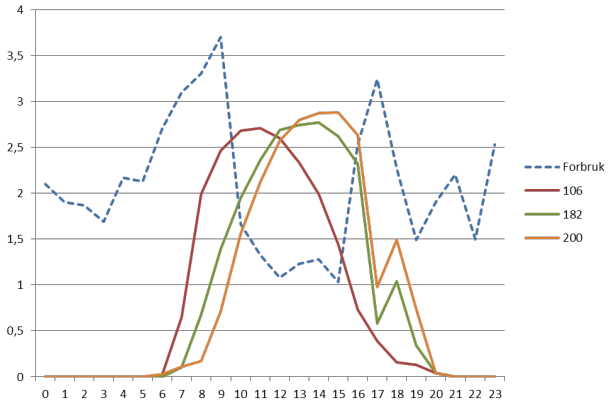


Figure 5. Shows production profiles for an easterly oriented rooftop PV (106), a panel pointing south (182) and a westerly oriented one (200). A consumption profile is shown for comparison.

Table 3. Max and min power for the average household equipped with a 3.1kWp panel. Max net consumption is obviously experienced during winter time. Max net consumption is positive during the summer.

Compass direction of PV	Max power consumption[kW]	Month	Min power (feed-in) [kW]	Month
South	-12.82	january	2.41	june
East	-12.82	january	2.06	june
Southeast	-12.82	january	1.99	june
West	-12.82	january	1.91	june

5. Conclusive Remarks

The FLEXNETT tool for analysis of load dynamics imposed by consumption and rooftop solar panels have been explained. UiT is currently maintaining the software. The simulation tool

created applies a machine learning approach for the purposes. A LSTM neural network was trained and applied to simulate the consequences of high-density deployment of solar panels in different areas at Hvaler. The primary idea was to use the tool to investigate the impact of prosumers on the local distribution grid. The tool proved to be useful to determine the magnitude of peaks caused by local feeds and consumption and how rapidly changes in loads can occur. Local conditions play an important role on the production profile and yield at all hours of the day. The tool accommodates this by means of the GIS. The tool developed made it possible to analyse and conclude that the grid sections investigated would be able to absorb a high number of prosumers. However, the tool also showed its strengths in exposing potential vulnerabilities where grid weaknesses can be a reality. The tool can help analyses where collective self-consumption is sought. This would be a kind of self-balancing "by design" where residential areas can be planned to minimize the impact on the grid and to avoid potential congestions. The results from the simulations combined with the empirical studies, suggest that local demography can impact the aggregated, net profiles for sections of the grid. When time of production does not match consumption well demographic homogenity may be an important reason for this. The tool was also useful to determine the cost for single households exposed to Norgesnetts power tariff. The benefit of rooftop PVs that meet hours with maximum consumption can be economically significant if power tariffs are introduced. The tool includes options where batteries can be deployed, and their effects studied. This part has not yet been fully developed and validated. But UiT is currently addressing the issue. Furthermore, the tool is considered as an engine for future studies of local energy markets. The plan is also to extend it to be able to address other areas in Norway. It is believed that the tool will be especially important for studies in the northern parts of the country where the average azimuth is even lower.

6. Acknowledgements

The FLEXNETT project was supported by the Norwegian Research Councils ENERGIX program. The research reported here was further supported by Smart Innovation Norway and GeoData (GIS system).

References

[1] https://www.sintef.no/prosjekter/flexnett/ (accessed April 2019)
[2] Andrzej Ozadowicz, A New Concept of Active Demand Side Management for Energy Efficient Prosumer Microgrids with Smart Building Technologies Energies (accessible at https://www.mdpi.com/)
[3] Stig Ødegaard Ottesen, Pol Olivella-Rosell, Pau Lloret, Ari Hentunen, Pedro Crespo del Granado, Sigurd Bjarghov, Venkatachalam Lakshmanan, Jamshid Aghaei, Magnus Korpås and Hossein Farahmand, Simplified Battery operation and control algorithm, Deliverable D5.3 v.1, INVADE Project, 20/12/2017 (Available at https://h2020invade.eu/)
[4] Bernt A Bremdal, Hanne Sæle , Geir Mathisen , Merkebu Zenebe Degefa; Flexibility offered to the distribution grid from households with a photovoltaic panel on their roof. IEEE: Energycon Conf, Limassol, June 2018.
[5] Bernt A. Bremdal and Kristoffer Tangrand: Investigations of prosumers aggregated loads and energy flexibility at Hvaler(In Norwegian): Undersøkelse av aggregerte last- og fleksibilitetsbidrag fra plusskunder på Hvaler, FLEXNETT rapport, Smart Innovation Norway. Published by SINTEF Energi, March 2018
[6] Jürgen Schmidhuber: Deep learning in neural networks: An overview in Neural Networks 61(2015) p. 85-117
[7] Xiaocheng Feng, Ming Liu, Jiahao Liu, Bing Qin , Yibo Sun, Ting Liu, Topic-to-Essay Generation with Neural Networks Proceed. of 21st International Joint Conference on Artificial Intelligence (IJCAI-18)
[8] Thomas Martinsen, A Business Model for an EV Charging Station with Battery Energy Storage, Paper 0374 CIRED Workshop, Helsinki, June 2016
[9] Nugroho, Asep. Battery State of Charge Estimation by Using a Combination of Coulomb Counting and Dynamic Model With Adjusted Gain. 10.1109/ICSEEA.2015.7380745.
[10] Hanne Sæle, Bernt A. Bremdal: Economic Evaluation of the Grid Tariff for Households with Solar Power Installed. Paper 0556, CIRED Conference, Glasgow, 12-15 June 2017

Paper B

Using Deep Learning Methods to Monitor Non-Observable States in a Building

B

Kristoffer Tangrand, Bernt Bremdal

Published in *Proceedings of the Northern Lights Deep Learning Workshop*, Volume 1, 6 pages, 2020.

Using Deep Learning Methods to Monitor Non-Observable States in a Building

Kristoffer Tangrand[1] and Bernt Bremdal[2]

[1,2]Department of Computer Science and Computational Engineering, UiT The Arctic University of Norway

Abstract

This paper presents results from ongoing research with a goal to use a combination of time series from non-intrusive ambient sensors and recurrent neural networks(RNNs) to predict room usage at a university campus. Training data was created by collecting measurements from ambient sensors measuring room CO_2, humidity, temperature, light, motion and sound, while the ground-truth counts was counted manually by human observers. Results include analyses of relationships between different sensor data sequences and recommendations for a prototype predictive model using recurrent neural networks.

Index terms— Indoor Air Quality, Occupancy Prediction, PCA, LSTM, GRU, Neural Architecture Search, Deep Learning, Internet of Things

1 Introduction

With the advent of Internet of Things (IoT) a multitude of monitoring and control opportunities arise. The development of smarter buildings, neighborhoods and cities have already embraced this. Energy use and indoor climate control are central aspects related to the performance of buildings. Today there is little knowledge on how a particular building is actually used. What part of the indoor area is populated at different hours? Selective energy use can lead to more efficient buildings [1]. Monitoring the number of people in the specific rooms of a building can be used to achieve a more focused and efficient use of energy in a building and more optimal space management. That in turn requires the ability to compare an estimate of space occupation and energy use.

Room occupancy has historically been monitored by means of cameras and smart phones(Wi-Fi, Bluetooth). But this raises privacy issues. However, ambient sensors that can monitor levels of CO2, temperature, humidity, illumination and noise are present in many office and educational buildings today.

The research contribution presented here involves applying recurrent neural networks as a predictive model for estimating the number of people in a room based on measured indoor climate variables, formulated as a regression problem.

2 Related work

CO_2 has proven a reliable indicator for occupancy detection [2]. Further, CO_2, illumination and sound are known to be highly correlated with human occupancy [3]. Machine Learning algorithms like Support Vector Machines(SVM) and Random Forest have shown promise on such sensor data [4]. Feed-forward Neural Networks(FNN) are used in [5] to predict occupancy numbers from CO_2, sound, temperature and motion. A combination of algorithms k-Nearest Neighbors(k-NN), FNN and Random Forest was used to achieve 85% accuracy in predicting the exact number of people in two rooms, it was however unable to predict accurately for unseen rooms [6]. In [7], k-NN is used to accurately determine the number of people according the classification criteria used by the authors. [8] found that utilizing a combination of averaged and first order differential CO_2 signal trained with SVM and FNN achieved an accuracy between 70-76% for occupation numbers of up to 200 people. In [9] a

combination of CO_2 and light signals was used to create a prototype smart HVAC(heating, ventilating, and air conditioning) system.

In recent years, RNN architecture Long Short-Term Memory(LSTM) has proven its effectiveness in a number of problem areas involving sequential data [10]. RNNs are known to outperform FNNs in problems involving sequential data. The main trick to achieving such great performance is an elaborate setup of information gates which lets each neuron in the RNN control which information to forget or remember depending on the patterns in the training sequence [11]. A simplified version called Gated Recurrent Units(GRU) was proposed more recently in [12]. Comparison between LSTM and GRU has been described in the literature, with performance being found to be roughly equal in [13].

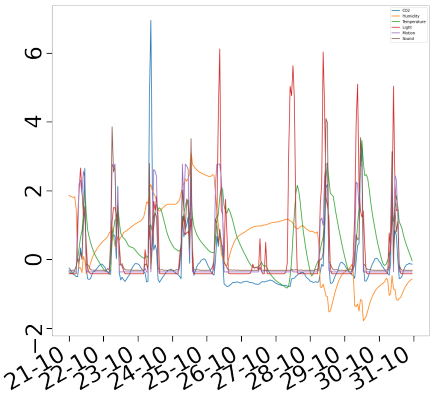


Figure 1: 10 days of data from a sensor placed in a study room. Note that the values are scaled and normalized for the plot.

3 Method

To explore the possibilities described above a device that combines different sensors and enable synchronization of time series from each sensor was used (see Figure 1). The sampling rate and number of sensors varied somewhat between devices. The two most common sampling rates were 10 and 40 samples per hour. However, the sensor data was pre-processed by a third-party that re-sampled the data to 4 samples each hour.

The data set was collected from ambient sensors placed in 10 relatively small study rooms that students usually use between classes or for studying alone and in groups. The data the sensors provided:

- CO_2, as measured in parts per million(ppm)
- Humidity, as measured by the amount of water vapour in the air
- Temperature, measured in Celsius
- Illumination, measured in lux
- Motion, which is a Passive infrared(PIR) sensor that returns a binary signal
- Sound level, measured in decibel

Human observers manually counted people to establish the ground-truth data for occupancy. The distribution of occupancy numbers in the collected dataset are shown in Table 1. It can be observed that samples with higher occupancy numbers are lacking.

The Method chapter consists of 3 separate subsections:

1. **Correlation Analysis** Here the statistical relationship between variables is described
2. **Time-dependency Inspection** This section describes a hands-on visual analysis of the variables and how they change through time
3. **Prediction Models** A walk-through of the process leading to the final LSTM model

3.1 Correlation Analysis

When using multivariate distributions as a basis for function approximation the statistical relationship between variables is of major importance. This is because many such function approximators learn better from uncorrelated data. First, the Pearson correlation coefficient [14] between features was investigated. Pearson correlation returns a value between -1 and 1 where -1 is the maximum negative correlation and 1 is the maximum positive. The closer the value gets to 0 the lower the correlation in any direction, and at 0 there is no correlation.

A common approach to pre-process training data and reduce over-fit before training neural networks

Number of people	0	1	2	3	4	5	6	7	8	9
Percentage of samples	60.6	14.7	8.4	4.9	4.8	3.0	1.3	0.5	1.6	0.3

Table 1: The distribution of ground truth occupancy numbers

is to apply Principal Component Analysis(PCA) [15] to reduce the number of correlated features. PCA can be used to determine the explained variance between the features. PCA helps to determine a variable's relative contribution to the overall. It is also used to determine statistical co-variance between variables.

Since the dataset is multivariate PCA can give indications on which variables correlate with each other, and more importantly, which do not. Since the sensor data is collected from multiple locations, it is also of importance to inspect the variation therein.

Before PCA, data should be standardized by removing the mean and dividing by the standard deviation. After standardization, each variable is centered on the origin of the principal components, which makes comparing the spatial relations possible. PCA does not handle missing values, so these were set to zero.

3.2 Time-dependency Inspection

As can be seen in Figure 1: Time series of this type often display a high degree of seasonality on a daily basis. Inspection of the data suggested that there were profound differences between rooms depending on their use, location in the building, heating and ventilation. The CO2 variable would dominate in one part while temperature and humidity gradients would be more pronounced in other parts of the building. Latency issues became evident too. It takes time for CO2 levels, humidity and temperature to increase when people meet in a room. Ventilation may flush stale air more or less effectively. Seasonal and daily changes in weather may affect conditions too. Spring or autumn with low solar altitudes may produce rays that effectively heat up a room in one part, while rooms on the shady side may require heating. These occurrences suggested a high degree of dynamics. States could be affected by situations or actions that happened several hours or even days before. Dry periods with a lot of sun would heat up concrete constructions and create more dust than rainy and chilly days. A shift could impact the observations profoundly. The conditions in a room caused by six people having a morning meeting could be logged quite differently from a similar situation in the afternoon as former meetings that day could create an "atmospheric legacy".

3.3 Prediction Models

The sequential nature of the recordings in this paper infers that the ordering of variable measurements can be taken advantage of. This is especially advantageous when training cyclic learners such as RNNs [16]. RNNs can be used for both classification and regression, depending on the properties of the output layer. Since a single scalar value such as an occupancy integer number can be easily scaled between 0 and 1, the learning process was formulated as a regression predicting a single floating number.

Designing architectures for neural networks has traditionally been manually done by humans. With the advent of great generalization tools and more powerful hardware, this is beginning to change, and Neural Architecture Search(NAS) is gaining traction [17]. NAS is a brute-force method that can be applied in various ways. Due to the many variations that are possible with(already heavy to train) neural networks like LSTM, narrowing down the search space is a trade-off that is almost unavoidable. As such, the selective focus of the NAS was to find the optimal activation functions for each of the layers, as well as finding the optimal optimizer rule. The NAS was also set up to give feedback on whether dropout or recurrent dropout increased precision. Finally, loss measurement and learning rate was also part of the search.

With this in mind, 3 investigations were performed:

1. **Neural Architecture Search** First, an architecture search was performed using the Keras wrapper Talos [18]. This package makes it simple to set up extensive architecture

	CO2	Humid	Temp	Light	Motion	Sound
CO2	1.000	0.128	0.433	0.217	0.325	0.221
Humidity	0.128	1.000	-0.004	-0.100	-0.075	-0.030
Temperature	0.433	-0.004	1.000	0.393	0.482	0.347
Light	0.217	-0.100	0.393	1.000	0.532	0.300
Motion	0.325	-0.075	0.482	0.532	1.000	0.570
Sound	0.221	-0.030	0.347	0.300	0.570	1.000

Figure 2: Pearson correlation between training features, as a mean between different rooms

searches to automate the process of optimizing hyperparameters.

2. **How different features influenced the learning process** A second analysis was performed to investigate the learning potential and influence each of the training features had on the model.
3. **Reproducibility and the influence of random initial parameters** Uncertainty is often experienced with regards to achieving reproducible results when training neural networks. The initial weights of a neural network are usually selected according to a random distribution.

Before training, the data was scaled between 0 and 1. If the feature had outliers, these were clipped to a value that realistically matched the rest of the distribution. The data was split into training(90%) and test(10%) sets, and then scaled with explicit range between 0 and 1. During the training process, 20% of the training data is used for validation.

4 Results

4.1 Correlation

First, the Pearson correlation between the features was examined. The result is shown in Figure 2. Observe that most signals correlate to varying degrees, except humidity.

Next, PCA was performed, shown in Figure 5,Figure 3 and Figure 4. The result is in agreement with the findings from the Pearson coefficients. Humidity is diverging from the other parameters, and there is not even correlation between humidity in different rooms. Of special interest, it can be noted that rooms with numbers 09 and 21 stand out with regards to both the CO_2 and humidity variables. In addition rooms 15, 18 and 19 are also displaying a different behavior compared to the rest of the rooms with regards to the same variables, but in opposite direction.

In Figure 5, observe that for most rooms, 40-50% information can be retained in only one component. From the same figure, it is clear that at least 4 principal components are needed to retain 90% of variance. It is also of importance that the variance is quite similar between rooms. But even the component with least variance contributes 5-7%, which can be quite substantial for a function approximator learning to distinguish between signals.

4.2 Prediction Model

4.2.1 Neural Architecture Search

The next step was finding a suitable RNN architecture for the data set. The NAS tested more than 2000 models varying hyperparameters and number of layers or nodes per layer. The architecture search showed that the dimensions of the neural network and training batches had little importance. The activation function and loss function were found to be the crucial parameters. In general, among the 20 best performing models all used Rectified Linear Unit(ReLU) or Exponential Linear Unit(ELU) as activation function between the first and middle layer with minuscule differences in performance. For the middle(hidden) layers, ReLU, Tanh and Sigmoid layers all achieved similar performance. For the activation of the last node, ReLU selected in 19 out of 20 cases. As for the loss function, Mean Squared Error(MSE) was common for all the models in the top 20. Adaptive Moment Estimation(ADAM)[19] was chosen as optimizing function in all the top 20 cases. Dropout was found by searches to lead to worse performance. However, recurrent dropout led to better performance.

The result in Figure 6 and Figure 7 shows the relation between prediction and truth predicted by the best-performing LSTM model(GRU performed similarly with equal parameters) on the test set. The input vector consist of a 16×6 where the former represents time steps and the latter training

features. 4 recordings per hour was used, such that each training sample contains 4 hours of previous data. The model has 3 hidden layers with shape $(64, 64, 64, 32)$ (including input layer) in which the last layer is densely connected to a single ReLU node. The network was trained with MSE as the loss function and ADAM as optimizer.

4.2.2 Feature Effectiveness

A feature effectiveness search was performed to investigate how well each feature and all combinations of features trained the network. The shape of the timestep vector remained the same. Here, it is observed that a model trained only on CO_2 had good learning potential, and that the other features alone had little or even no learning potential. A combination of all features except CO_2 did however have almost as good potential as CO_2 alone.

4.2.3 Reproducibility

At last, a number of tests with different initial parameters and random seeds were performed to test for reproducible results. For most seeds the network performed as expected(such as in Figure 6), but in rare cases with some seeds the neural network would not train at all. If in such case one is using a callback function to stop a network from over-training by monitoring loss, the network will not train and the data may appear useless.

4.2.4 Final Model

It is observed that the model is able to generalize patterns in the training set, even if the low amount of available training data probably leads to underperformance on accuracy. The prediction was inversely transformed and rounded to the closest integer. The mean absolute error from the model in Figure 6 prediction on the test set was at 0.47. In 68% of cases of the test set the model was able to precisely predict the number of persons, meaning the absolute error was lower than 0.5. In 85% of test samples, the error was lower than 1, and in 96% of samples the error was lower than 2 (Table 2). However, as seen in Table 2, a few outliers contributed heavily to the overall error. These outliers often correlate with higher occupancy numbers, as seen in Figure 7. Due to the low number of training samples for states with a high number of people(Table 1), it is likely that the network is undertrained on these occurrences of larger groups and as such scores particularly bad on these samples.

Error less than	0.5	1.0	2.0
Percentage of samples	68	85	96

Table 2: Prediction errors grouped by magnitude

Algorithm	SVR	RFR	FNN	LSTM
RMSE	0.151	0.110	0.102	0.097

Table 3: Comparison between machine learning models

Finally, LSTM was compared to some of the ML models from the literature on the same data set, namely SVR, Random Forest Regression and FNN. LSTM performs marginally better than its acyclic cousin, the FNN.

5 Discussion

The PCA analysis indicates that creating a general model for a certain type of room that has adequate precision might be difficult, due to the differences found in the correlation patterns between rooms. All of the rooms share the same HVAC system and are located at the same floor in the same building. Most rooms display similar behavior, but the analysis shows that a few rooms show different behavior from the rest. A floor plan drawing of the rooms was investigated to check if there were obvious architectural properties that could be influencing the data. The two rooms with numbers 09 and 21 were, unlike the other rooms, both situated next to the building's HVAC and/or plumbing system, which could explain the disparity these rooms displayed with regards to CO_2 and humidity. A solution to circumvent this problem could be to include a one-hot encoded feature vector which would give the learning model information about which room each sample was collected from. However, there is a chance that this could decrease the generality of such a model. Further investigations are needed to establish the degree of generality that can be achieved.

The data patterns found in the training samples are hard to analyse due to a large probability of

noisy interventions because of the dynamic and unpredictable states the sensors are subject to. While some of the signals clearly contain more seasonality due the nature of the data it monitors, especially CO_2, others are a lot more noisy and may only occasionally contain trends, such as sound and light. CO_2, humidity and temperature signals are influenced by pre-existing building monitor systems such as ventilation and thermostats.

The scaling techniques used to pre-process the training data might also "skew" certain features into influencing the training process more, especially since the distribution of the variables are not necessarily similar.

It is not surprising that the most of the information is contained in the CO_2 time series. This is consistent with previous research. Interestingly, during the feature effectiveness search a combination of the 5 other features is able to train a model that approaches the performance of a model trained only with CO_2 data on models trained from scratch. One could assume that the patterns contained in light and sound measurements would be noisy due to external influences from beyond the room itself. I.e. a solo person would probably not change the sound signal very much. Such is also the case for light, which is not exclusively influenced by indoor lighting, but also from sunlight. The motion sensor returns binary values and as such does not say anything about the actual amount of people, only if there are people or not. However, there was found little difference in training precision between networks that incorporate these features contrary to those that do not. If anything, the model using all training features are seemingly more robust since their performance is equal or better. It seems that recurrent neural network models like LSTM and GRU are able to filter the noisy parts of these signals and only use them in cases where they actually have predictive capability.

The same seems to also be true for the temperature feature. Temperature is controlled by thermostats and as such would balance out any human intervention in the heat signal. But this is also a signal that would contribute to the training of a recurrent neural network, since the temperature would first rise, and the thermostat would respond and adjust. These patterns could be present in the sliding time window training samples, depending on the time resolution.

Using human observers to gather ground-truth is costly and the yield is limited. However, attempts to introduce more automated means failed. The room booking system proved to be a very unreliable source for the same purpose. Any effort to let room users systematically share reliable measurements of room use and space occupancy proved very unreliable.

Employing NAS was useful in the sense that it eliminated inefficient hyperparameters. However, it seems that among the upper tiers of architectures for the given dataset and problem the robustness of trained models are very similar. This suggests that the quality of the dataset is more important in this case - a multitude of models manage to achieve similar performance. The initial distribution of parameters also influence the training outcome of certain models.

6 Conclusion

Earlier research on the correlation between indoor climate and occupancy numbers suggested that a number of machine learning models were able to model this relationship for prediction. This research set out to investigate how a recurrent neural network would fare in this regard, and more specifically LSTM.

The research shows that LSTM is within the same league or better than the machine learning models referred to in the literature for modelling this relationship, in the case of our dataset. This version of work focused on rather low-res time windows on 15 minutes. It can be assumed, by the notion that LSTM performance relative to other models increase as sequential data gets more detailed, that higher resolution data would lead to even better relative performance.

7 Further Work

Low-resolution thermal cameras with RNNs has proven promising as a non-intrusive method for monitoring the presence or actions of persons [20]. Such data could be used as training labels in the setting this paper describes, or act as a counting/monitoring device on its own.

References

[1] B. Bremdal, C. Skjerve-Nielsen, and G.' Nereng. "How the Prosumer Role in Smart Grids Redefines the Energy Efficiency Concept of Buildings". In: World Sustainable Building Conference SB 2011, Helsinki, Oct. 2011.

[2] Chaoyang Jiang et al. "Indoor occupancy estimation from carbon dioxide concentration". In: *ArXiv* abs/1607.05962 (2016).

[3] I. B. A. Ang, F. Dilys Salim, and M. Hamilton. "Human occupancy recognition with multivariate ambient sensors". In: *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. Mar. 2016, pp. 1–6. DOI: 10.1109/PERCOMW.2016.7457116.

[4] A. P. Singh et al. "Machine Learning-Based Occupancy Estimation Using Multivariate Sensor Nodes". In: *2018 IEEE Globecom Workshops (GC Wkshps)*. Dec. 2018, pp. 1–6. DOI: 10.1109/GLOCOMW.2018.8644432.

[5] T. Ekwevugbe et al. "Real-time building occupancy sensing using neural-network based sensor network". In: *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*. July 2013, pp. 114–119. DOI: 10.1109/DEST.2013.6611339.

[6] N. Bockstael and A. Jadin. *CO2 based room occupancy detection*. 2018.

[7] Andrzej Szczurek, Monika Maciejewska, and Tomasz Pietrucha. "Occupancy determination based on time series of CO2 concentration, temperature and relative humidity". In: *Energy and Buildings* 147 (2017), pp. 142–154. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2017.04.080. URL: http://www.sciencedirect.com/science/article/pii/S0378778817302396.

[8] M.S. Zuraimi et al. "Predicting occupancy counts using physical and statistical Co2-based modeling methodologies". In: *Building and Environment* 123 (2017), pp. 517–528. ISSN: 0360-1323. DOI: https://doi.org/10.1016/j.buildenv.2017.07.027. URL: http://www.sciencedirect.com/science/article/pii/S0360132317303268.

[9] Chen Mao and Qian Huang. "Occupancy Estimation in Smart Building using Hybrid CO2/Light Wireless Sensor Network". In: Oct. 2016.

[10] Klaus Greff et al. "LSTM: A Search Space Odyssey". In: *IEEE Transactions on Neural Networks and Learning Systems* 28 (2017), pp. 2222–2232.

[11] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[12] Kyunghyun Cho et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". In: *ArXiv* abs/1406.1078 (2014).

[13] Junyoung Chung et al. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: *ArXiv* abs/1412.3555 (2014).

[14] Karl Pearson. "Note on Regression and Inheritance in the Case of Two Parents". In: *Proceedings of the Royal Society of London Series I* 58 (Jan. 1895), pp. 240–242.

[15] Karl Pearson F.R.S. "LIII. On lines and planes of closest fit to systems of points in space". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), pp. 559–572. DOI: 10.1080/14786440109462720. eprint: https://doi.org/10.1080/14786440109462720. URL: https://doi.org/10.1080/14786440109462720.

[16] Jürgen Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural Networks* 61 (Jan. 2015), pp. 85–117. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2014.09.003. URL: http://dx.doi.org/10.1016/j.neunet.2014.09.003.

[17] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. "Neural Architecture Search: A Survey". In: *J. Mach. Learn. Res.* 20 (2018), 55:1–55:21.

[18] Autonomio Talos [Computer software]. *Talos*. Version 0.6.4. 2019. URL: http://github.com/autonomio/talos.

[19] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015. 2014. URL: http://arxiv.org/abs/1412.6980.

[20] Asbjørn Danielsen and Bernt Bremdal. "Predicting Bedside Falls using Current Context". In: Dec. 2017. DOI: 10.1109/SSCI.2017.8280988.

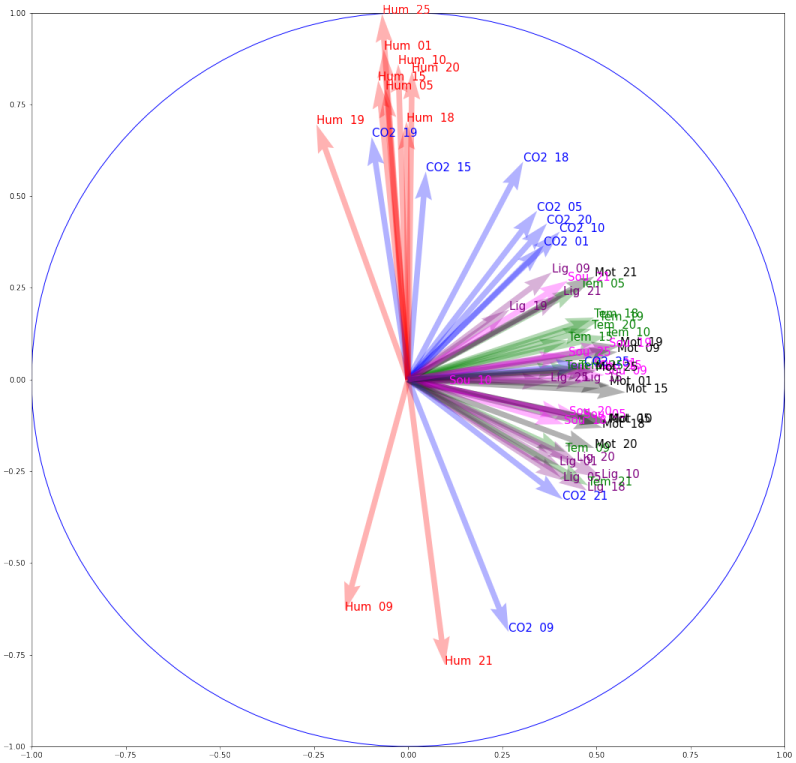


Figure 3: Correlation of principal components 1 and 2 with the original features. The same-colored arrows respond to the same components in 10 different rooms, while the number after each feature name simple refers to the rooms internal name

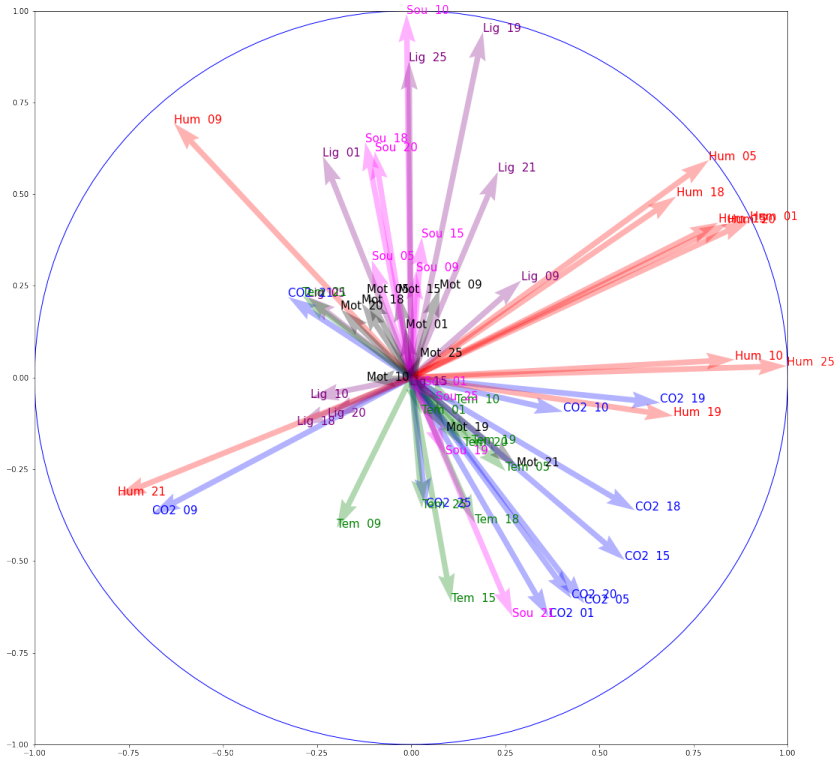


Figure 4: Correlation of principal components 2 and 3 with the original features. The same-colored arrows respond to the same components in 10 different rooms, while the number after each feature name simple refers to the rooms internal name

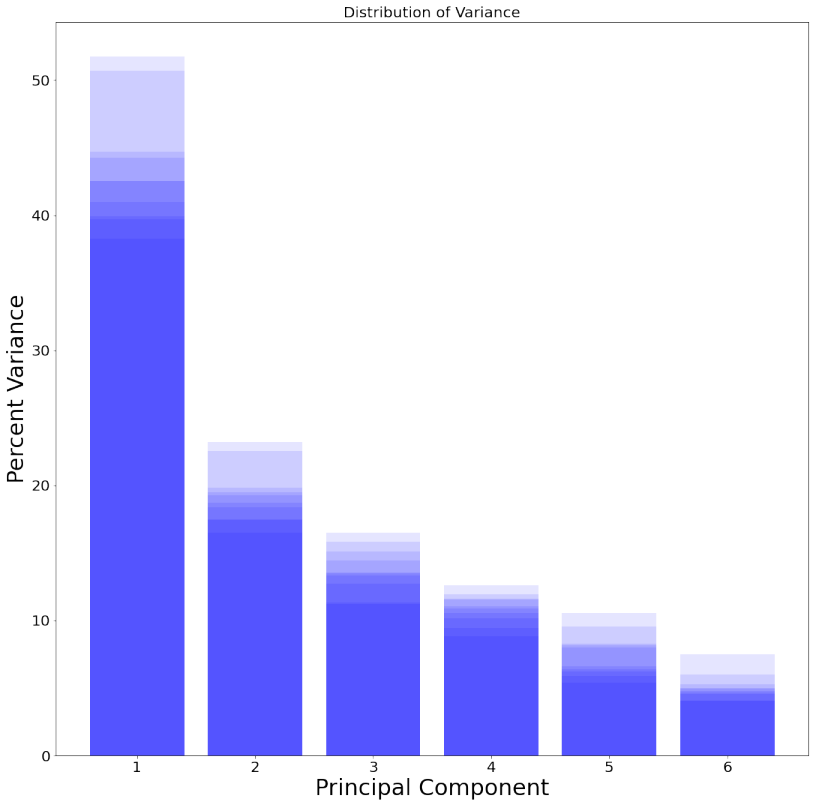


Figure 5: Variance retained per principal component. The color strength of each bar scales with how much variance is retained per component, per room. Here, result for 10 rooms are shown.

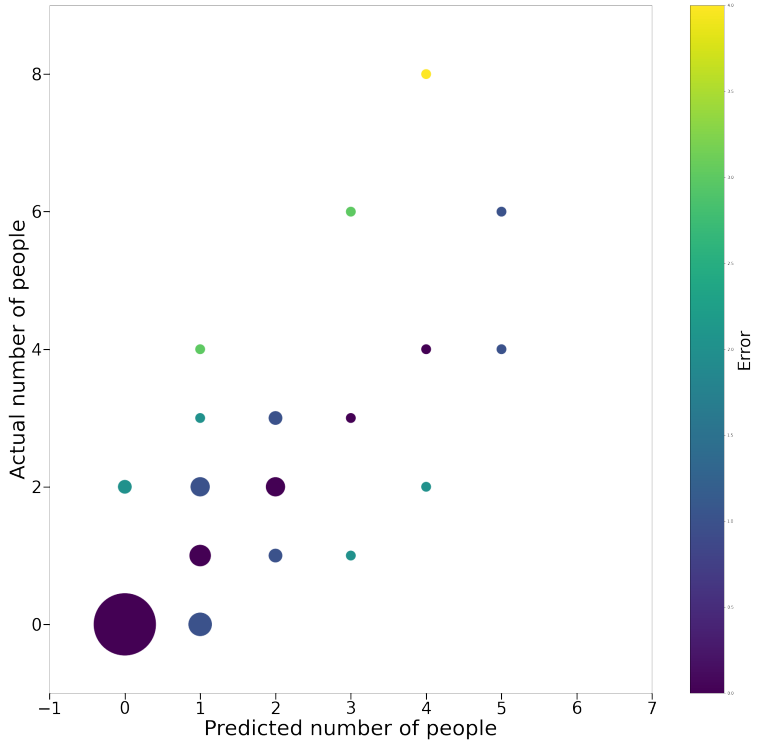


Figure 6: Prediction compared to the truth values. Here, the prediction is rounded to whichever is closest, 0 or any natural number. The size of the circles describes the amount of samples of this size, i.e. there are quite a few 0's. For a more precise interpretation of the same data, look at Figure 7

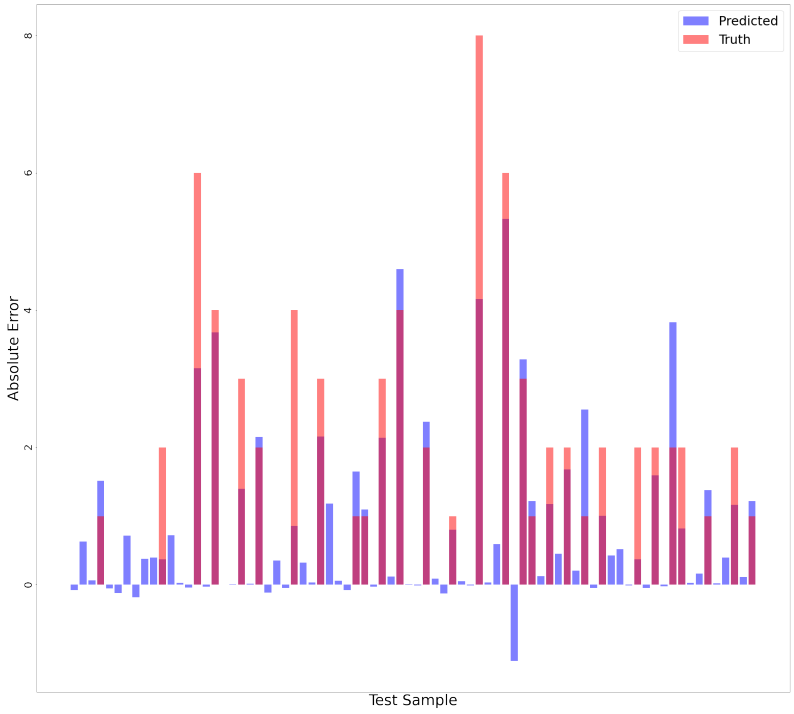


Figure 7: Prediction vs Truth after inversely scaling the prediction output. Note that the darker red are simply where the bars are drawn on top of each other, such that the colors are mixed. When there is no red, the truth value is 0

Paper C

Wavelet Neural Networks versus Wavelet-based Neural Networks

Lubomir Dechevsky, Kristoffer Tangrand

Technical report, UiT The Arctic University of Norway, 2022

C

WAVELET NEURAL NETWORKS VERSUS WAVELET-BASED NEURAL NETWORKS

A PREPRINT

Lubomir T. Dechevsky, Kristoffer M. Tangrand
lde009@uit.no, ktangrand@gmail.com
Faculty of Engineering Science and Technology
UiT – The Arctic University of Norway
P. O. Box 385, N-8505 Narvik, Norway

March 2, 2023

ABSTRACT

This is the first paper in a sequence of studies including also [DPS+–2022a] and [DPS+–2022b] in which we introduce a new type of *neural networks* (NNs) – *wavelet-based neural networks* (WBNNs) – and study their properties and potential for applications. We begin this study with a comparison to the currently existing type of *wavelet neural networks* (WNNs) and show that WBNNs vastly outperform WNNs. One reason for the vast superiority of WBNNs is their advanced hierarchical tree structure based on *biorthonormal multiresolution analysis* (MRA). Another reason for this is the implementation of our new idea to incorporate the wavelet tree depth into the neural width of the NN. The separation of the roles of wavelet depth and neural depth provides a conceptually and algorithmically simple but very highly efficient methodology for sharp increase in functionality of swarm and deep WBNNs and rapid acceleration of the machine learning process. In Theorem 1 (Section 2) we obtain a new result for the established WNNs: we propose a type of activation which is shown to lead to optimal performance of WNNs and show that even optimal performance of WNNs is vastly outperformed by WBNNs. In Section 3, in Theorems 2 and 3 we obtain new results about the

learnability via WNNs and WBNNs and in Corollary 1 we show that WBNNs can be used to learn efficiently not only any regular distribution in $L_{1,loc}$ but also singular distributions like the Dirac delta and its derivatives. In the same section we provide the general characteristics (i–iii) of the rich diversity of activation operators that can be used in machine learning via WBNNs of univariate and multivariate manifolds in two, three and higher-dimensional spaces. Here we establish the principal differences between non-threshold and threshold activation in learning fractal and piecewise smooth manifolds, respectively. In Section 4 we briefly address the importance of interconnection and interaction between *swarm AI* and *deep evolutionary AI* and the relevance to computational implementations using CPU and GPU parallelism. In Section 5 we introduce a new activation method based on the concept of *decreasing rearrangement* in functional analysis and function space theory. Theorem 4 is a uniform approximation theorem (UAT) providing qualitative proof of the consistency of the learning process when using the *decreasing rearrangement activation*. Theorem 5 provides an important quantitative upgrade of the UAT in Theorem 4 by showing that decreasing rearrangement activation of WBNNs results in machine learning process which is optimal in two key aspects: *fastest learning* and *maximal compression*. In Section 6 we consider four representative model examples which are then subjected to comprehensive graphical comparison the results of which have been systematized into a collection of conclusions and comments. Section 7 is comprised of the proofs. In the concluding Section 8 we discuss the connection of the present results with the studies in [DPS+–2022a], [DPS+–2022b], as well as some additional computational and research topics.

MSC2020: Primary 68T07; Secondary 42C40, 46E35, 65T60, 68Q32

Keywords and Phrases: artificial intelligence, swarm, deep evolutionary, neural networks, wavelet-based, machine learning, fastest learning, maximal compression, neural activation, threshold, decreasing rearrangement, Sobolev embedding

1 Introduction

The purpose of this paper is to propose a new approach to machine learning of geometric manifolds in $\mathbb{R}^n$, where $n = 1, 2, 3, 4, ...$ using single-layer or deep neural networks (NNs) based on Riesz unconditional bases of biorthonormal wavelets.

The first attempt to marry the theory of NNs with wavelet theory dates back to the early 1990s [ZB–1992]. This initial study gave rise to particularly constructed NNs which were named by the authors of [ZB–1992] as *wavelet NNs* (WNNs). In the course of the next twenty years, the theory and applications of WNNs were studied by numerous authors. The results of these studies have been summarized in [AZ–2013] which constitutes a comprehensive account of the current status of the study of WNNs. Before going into the mathematical details of the construction and functioning of WNNs, let us note that this type of networks was introduced relatively early, when wavelet theory was still quite new to the developers of applications in the field of Artificial Intelligence (AI). Due to this, WNNs make use only of a very small subset of the useful properties of wavelet bases. Thus, while the theory of WNNs relies on the basic property of wavelet basis functions that they are dilations and translations of one and the same function, this theory ignores the more advanced properties of wavelet bases related to *Multi-Resolution Analysis* (MRA). As a consequence of this, methods using WNNs are no more than a variant of meshless kernel estimation methods. The typical representative of these meshless methods are the ones using radial basis functions [CCG–1991]. The only essential difference with the variant of WNNs is that radial basis functions are replaced by tensor-product functions with sufficient number of vanishing moments. Not surprisingly, the mathematical apparatus used with WNNs is identical with the one for radial basis functions: iterative gradient or subgradient optimization methods. Unfortunately, these iterative methods can guarantee providing the global extrema only when the respective criterial functionals (objective functions) are convex. (Of course, in the particular case when the convex criterial functional is quadratic, possibly with linear constraints, in addition to iterative methods there exists also a broad variety of methods of computational linear algebra.) In practical applications, however, the realistic criterial functionals are most often non-convex, with multiple local extrema and saddlepoint singularities. In this general situation, the optimization methods used with WNNs and radial basis functions produce only local extrema which are close to the global extrema only if a very good initial starting point of the iterative algorithm is proposed. The usual defence of this type of results is to claim that all the local extremal values are close in value to the global extremal value. Here is a typical exposition of this type of argument [LBH–2015]: "..., In particular, it was commonly thought that simple gradient

descent would get trapped in poor local minima – weight configurations for which no small change would reduce the average error. In practice, poor local minima are rarely a problem with large networks. Regardless of the initial conditions, the system nearly always reaches solutions of very similar quality. Recent theoretical and empirical results strongly suggest that local minima are not a serious issue in general. Instead, the landscape is packed with a combinatorially large number of saddlepoints where the gradient is zero, and the surface curves up most dimensions and curves down in the remainder. The analysis seems to show that saddlepoints with only a few downward curving directions are present in very large numbers, but almost all of them have very similar values of the objective function. Hence, it does not much matter which of these saddlepoints the algorithm gets stuck at."

Some critical analysis of the above text in [LBH–2015] is due, as follows. While there is some rationale in the above claims for large and very large sample and network sizes, these claims cannot be accepted even as basic "rules of thumb"; expressions like "rarely", "nearly always", "very similar", "strongly suggest", "not a serious issue in general", "seems to show", "almost all", "does not much matter" are not good replacements for logical quantors. The argument about saddlepoints with only very few negative components in the signature (that is - to use fuzzy terminology in the spirit of [LBH–2015] - 'saddlepoints which are almost local minima'), is also unconvincing as a qualitative statement without any criteria or means for quantitative measurement. Even if a saddlepoint in a problem with a very large size has only one downward-curving dimension, the respective value of the criterial functional (objective function) can be much larger than the global minimum, if the downward curve is sufficiently steep. The one rigorous conclusion that can be drawn from the above excerpt of [LBH–2015] is that, once the criterial functional (objective function) ceases to be (globally) convex, iterative gradient/subgradient optimization is no longer a reliable approach to achieving quality learning results. The only way to achieve best (or, at least, sufficiently high) quality results is to start from a very good initial point of the iterative process, but the traditional way of achieving this is by human intervention, i.e., the use of natural, rather than artificial intelligence. In fact, the authors of [LBH–2015] acknowledge this in another excerpt of their text, as follows: "... The conventional option is to hand design good feature extractors, which requires a considerable amount of engineering skill and domain expertise. But this can all be avoided if good features can be learned automatically using a general-purpose learning procedure. This is the key advantage of deep learning ...". Our comment to this excerpt is that in the general case of non-convex criterial functionals, gradient search only plays the role of an auxiliary tool for improvement of already good

results. Achieving these good initial results using AI is thus claimed in [LBH–2015] to be the main aim of deep learning, and, in general, this aim cannot be attained by only using local optimization methods. The major weakness of WNNs proposed in [ZB–1992] is that the crucial problem of finding a good initial starting point for iterative local optimization has only been addressed by the vague recommendation that some 'explicit link between the network coefficients and the wavelet transform' should be provided. This weakness persists also in the later developments and upgrades of WNNs discussed in [AZ–2013]. Our present study shows that for sufficiently large samples this weakness can be overcome, at least partially, via tools from functional analysis. Namely, there is a rigorous mathematical general way to automatically improve the quality of the starting point of the iterative optimization process, valid for all cases when the criterial functional can be interpreted as distance between two mathematical objects. Since, to the best of the authors' knowledge, this systematic approach seems to be new in the context of NNs (and much more certainly so in the specific context of WNNs), we shall outline its main idea already in this early stage of our exposition, as follows. If the metric criterial functional of an optimization problem has an *equivalent metric which can be computed efficiently without the need of iterative optimization*, then this equivalent metric d_1:

$$0 < c_0 d_1(x, y) \leq d(x, y) \leq c_1 d_1(x, y) \tag{1}$$

can be used to generate a consistently good starting point for the optimization of the original metric d, provided that the equivalence constants $c_j,\ j = 0, 1$, with $0 < c_0 \leq 1 \leq c_1 < \infty$, do not depend on x, y and, in numerical problems, they are independent of the sample size of the numerical data (in the sequel of this exposition we shall use the notation $d \asymp d_1$). In most practical applications in numerical analysis the metrics d and d_1 are induced by respective equivalent norms or quasinorms [BL–1976] or their seminorm variants. This refers not only to deterministic quantities, but also in the indeterministic case, e.g. when considering equivalent risks in statistical estimation. A typical model example, with various applications in deterministic approximation and statistical risk estimation, is the Peetre $\mathbb{K}$-*functional* between Lebesgue space L_p and homogeneous Sobolev space $\dot{W}_p^k,\ 1 \leq p \leq \infty,\ k \in \mathbb{N}$, [BL–1976],

$$K(h^k,\ f;\ L_p, \dot{W}_p^k) = \inf_{\varphi \in \dot{W}_p^k} (||f - \varphi||_{L_p} + h^k ||\varphi||_{\dot{W}_p^k}) \tag{2}$$

where $f \in L_p + \dot{W}_p^k$ (the algebraic sum of the two spaces) and h is the step (in applications, related to the sample size). An equivalent seminorm of $K(h^k, f; L_p, \dot{W}_p^k)$ is $||f - f_{k,h}||_{L_p} + h^k||f_{k,h}||_{\dot{W}_p^k}$ where $f_{k,h}$ is the Steklov mean value of f with parameters k and h [DP–1997], the equivalence constants being independent on f and h. The numerical computation of $f_{k,h}$ is based on quadrature formulae and does not involve optimization. Thus, $\varphi_0 = f_{k,h}$ can be used as a starting point of iterative optimization. The quality of φ_0 as initial solution of the optimization problem depends on the size of the equivalence constants c_0 and c_1: if $c_0 = 1 = c_1$ (isometric equivalence) then φ_0 is the exact solution of the optimization problem. In the considered example, c_1 increases rapidly with increase of k and so for a fixed step $h > 0$ (fixed sample) the quality of φ_0 as initial solution deteriorates with the increase of k. Fix $k \in \mathbb{N}$ and let $h \to 0+$ (take sufficiently large sample): since c_j, $j = 0, 1$, do not depend on h (the sample size), for sufficiently large sample sizes φ_0 will be a consistently good starting point of the iterative optimization. Our first new result in Section 2 is to use the above idea for automatic generation of an initial starting point of iterative optimization in the case of WNNs. Although satisfactory from theoretical point of view, the practical usefulness of this generation would be rather limited, because the generated initial solution of the optimization problem would be consistently close to the global optimum only for sufficiently large sample sizes. Nothing is guaranteed for large samples of any a priori fixed size, let alone samples of medium or small size (in our numerical examples in the sequel of this exposition we shall consider sample sizes with $N \geq 2^{12}$ as very large, $2^{10} < N < 2^{12}$ as large, $2^9 \leq N \leq 2^{10}$ as medium-to-large, $2^8 \leq N < 2^9$ as medium, $2^7 \leq N < 2^8$ as medium-to-small, and $1 \leq N < 2^7$ as small). Our conclusion is that WNNs can be used efficiently for finding a consistently good local extremum only for very large sample sizes.

This weakness cannot be overcome within the conceptual construction of WNNs: a more advanced construction of relevant NNs is needed which we shall introduce in the present paper and call *Wavelet-Based Neural Network* (WBNN). The principal difference between WNNs and WBNNs is explained, as follows. Let φ be a scaling function (father wavelet) and ψ be the corresponding wavelet (mother wavelet) obtained by MRA [D–1992],[D–1997], so that for any $j_0 \in \mathbb{Z}$ the functions

$$\varphi_{j_0k_0}(x_1) = 2^{\frac{j_0}{2}}\varphi(2^{j_0}x_1 - k_0), \quad \psi_{jk}(x_1) = 2^{\frac{j}{2}}\psi(2^j x_1 - k) \tag{3}$$

$x_1 \in \mathbb{R}$, $j = j_0, j_0+1, ..., k_0 \in \mathbb{Z}$, $k \in \mathbb{Z}$, form an orthonormal basis of $L_2(\mathbb{R})$ and $\int\limits_{\text{supp }\psi} x_1^\lambda \psi(x_1)\,\mathrm{d}x = 0$ holds for all $\lambda = 0, 1, ...$ with $\lambda < r$ for some, henceforward fixed, $r > 0$, where $\text{supp }\psi$ is the support of ψ in $\mathbb{R}$.

Let φ be compactly supported in $\mathbb{R}$ (which implies the same for ψ). Let $B^s_{pq}(\mathbb{R}), -\infty < s < +\infty, 0 < p \le \infty, 0 < q \le \infty$ be the inhomogeneous Besov space with smoothness index s, metric power index p and metric logarithmic index q (a definition will be given below). Assume that $\varphi \in B^r_{\infty\infty}(\mathbb{R})$, $\psi \in B^r_{\infty\infty}(\mathbb{R})$. Then, since φ and ψ are compactly supported, $\varphi \in B^r_{p\infty}(\mathbb{R})$, $\psi \in B^r_{p\infty}(\mathbb{R})$ holds for every $p : 0 < p \le \infty$. For $x = (x_1, ..., x_n) \in \mathbb{R}^n$, $k = (k_1, ..., k_n) \in \mathbb{Z}^n$, consider [D–1999]

$$\begin{aligned}
\varphi^{[0]}_{0k}(x) &= \varphi_{0k_1}(x_1)\varphi_{0k_2}(x_2)\varphi_{0k_3}(x_3)...\varphi_{0k_n}(x_n)\\
\psi^{[1]}_{jk}(x) &= \psi_{jk_1}(x_1)\varphi_{jk_2}(x_2)\varphi_{jk_3}(x_3)...\varphi_{jk_n}(x_n)\\
\psi^{[2]}_{jk}(x) &= \varphi_{jk_1}(x_1)\psi_{jk_2}(x_2)\varphi_{jk_3}(x_3)...\varphi_{jk_n}(x_n)\\
&\dots\\
\psi^{[2^n-1]}_{jk}(x) &= \psi_{jk_1}(x_1)\psi_{jk_2}(x_2)\psi_{jk_3}(x_3)...\psi_{jk_n}(x_n)\ .
\end{aligned} \tag{4}$$

Denote $\varphi^{[0]} = \varphi^{[0]}_{00}$, $\psi^{[l]} = \psi^{[l]}_{00}$, $l = 1, 2, ..., 2^n - 1$. Then, $\varphi^{[0]} \in B^r_{p\infty}(\mathbb{R}^n)$, $\psi^{[0]} \in B^r_{p\infty}(\mathbb{R}^n$ for any $p : 0 < p \le \infty$, where $\psi^{[l]}$ is orthogonal to all polynomials of n variables of total degree less than r. Besides, $\{\varphi^{[0]}_{0k}, \psi^{[l]}_{jk}\}_{k\in\mathbb{Z}^n, j=0,1,...,2^n-1}$ is an orthonormal basis of $L_2(\mathbb{R}^n)$.

Moreover, for $f \in B^s_{pq}(\mathbb{R}^n), 0 < p \le \infty, 0 < q \le \infty, n(\frac{1}{p} - 1)_+ < s < r$,

$$f(x) = \sum_{k\in\mathbb{Z}^n} \alpha_{0k}\varphi^{[0]}_{0k}(x) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}^n}\sum_{l=1}^{2^n-1} \beta^{[l]}_{jk}\psi^{[l]}_{jk}(x) \tag{5}$$

for Lebesgue almost everywhere (Lebesgue – a.e.) $x \in \mathbb{R}^n$ holds, where $\alpha_{0k} =< \varphi^{[0]}_{0k}, f >= \int_{\mathbb{R}^n} \varphi^{[0]}_{0k}(x)f(x)dx$, $\beta^{[l]}_{jk} =< \psi^{[l]}_{jk}, f >$ and $a_+ = \max\{a, 0\}, a \in \mathbb{R}$. Convergence in (5) is in the quasinorm topology of the inhomogeneous Besov space $B^s_{pq}(\mathbb{R}^n)$ and, in view of the lower constraint about s, also in every Lebesgue point of f, i.e., Lebesgue a.e. on $\mathbb{R}^n$. Here, $B^s_{pq}(\mathbb{R}^n)$ admits the following quasinorm in terms of wavelet coefficients:

$$\|f\|_{B^s_{pq}(\mathbb{R}^n)} = \left\{ \Big(\sum_{k\in\mathbb{Z}^n} |\alpha_{0k}|^p \Big)^{\frac{q}{p}} + \sum_{j=0}^{\infty} \left[2^{j[s+n(\frac{1}{2}-\frac{1}{p})]} \Big(\sum_{k\in\mathbb{Z}^n} \sum_{l=1}^{2^n-1} |\beta^{[l]}_{jk}|^p \Big)^{\frac{1}{p}} \right]^q \right\}^{\frac{1}{q}}. \tag{6}$$

The construction introduced in (3-6) above generates an MRA with an orthonormal wavelet basis

$$\{\varphi_{0\mu}, \mu \in \mathbb{Z}\} \cup \{\psi^{(l)}_{j\nu}, j = 0, 1, ..., \nu \in \mathbb{Z}^n, l = 1, ..., 2^{n-1}\} \tag{7}$$

A typical example of such compactly supported wavelets are the Daubechies wavelets [D–1992] which will be the ones used in the remaining part of this paper. It is possible to generalize this construction to generate a broader class of MRAs based on bi-orthonormal wavelets [D–1997]. These are of considerable interest in the case of polynomial spline-wavelets which are the type preferred in image processing for $n = 2$ and surface processing for $n = 3$. In this case it is imperative to use bi-orthonormal and not orthonormal spline-wavelets, because only in the proper bi-orthonormal case can the spline-wavelet be compactly supported. (Moreover, an additional advantage in image processing is that there exist proper bi-orthonormal spline-wavelets which are compactly supported and whose graphs are symmetric.) There are no MRAs with compactly supported orthonormal polynomial spline-wavelet bases. We intend to consider the use of bi-orthonormal compactly supported spline-wavelets in a subsequent publication dedicated to deep image learning.

An important property of (bi)-orthonormal MRAs which follows from (3) is that $j = 0$ in (4) can be replaced by any $j_0 \in \mathbb{Z}$, such that (5) continues to hold true with $j = 0$ replaced by $j = j_0$. In this case, (6) defines an equivalent norm in $B^s_{pq}(\mathbb{R}^n)$ for $p \geq 1$, $q \geq 1$ (quasinorm for $0 < p < 1$ and/or $0 < q < 1$) with equivalence constants dependant on j_0. (The concept of equivalent metrics continues to hold true for quasinorms, because quasinormed abelian groups are metrizable [BL–1976, Section 3.10] – see also Section 3.

Consider now the above construction with $j_0 \in \mathbb{Z}$. Let $J \in \mathbb{Z}$ be such that $j_0 < J < \infty$ and consider the truncation $\sum_{j=j_0}^{J}$ of the series $\sum_{j=j_0}^{\infty}$ in (5) and (6). This defines a subspace $V_J \subset B^s_{pq}(\mathbb{R}^n)$ such that

$$V_J = \operatorname{span}\big(\{\varphi_{j_0\mu} : \mu \in \mathbb{Z}\} \cup \{\psi^{[l]}_{j\nu} : l = 1, ..., 2^n - 1, \nu \in \mathbb{Z}^n, j = j_0, ..., J\}\big). \tag{8}$$

Due to the properties of MRA, the following sequence of nested inclusions holds:

$$V_{j_0} \subset V_{j_0+1} \ \subset ... \subset V_J \subset V_{J+1} \subset ... \tag{9}$$

with

$$\overline{\bigcup_{j=j_0}^{\infty}} V_j = L_2(\mathbb{R}^n) \tag{10}$$

holds where $\bar{X}$ is the topological closure in Y of $X \subset Y$, where Y is a complete topological space. (In the case of MRA, the complete topological space Y is L_2 with respect to the topology induced by the inner product in L_2 or, equivalently, the norm in L_2.) Consider also the spaces $W_j = \text{span}\{\psi_{j\nu}^{[l]} : l = 1, ..., 2^{n-1}, \nu \in \mathbb{Z}^n\}$ where $j = j_0, ..., J$.

By the properties of MRA, $f \in V_J$ admits two equivalent representations, as follows:

$$\sum_{k_J \in \mathbb{Z}^n} \alpha_{Jk_J} \varphi_{Jk_J}(x) = f(x) = \sum_{k_{j_0} \in \mathbb{Z}^n} \alpha_{j_0 k_{j_0}} \varphi_{j_0 k_{j_0}}(x) + \sum_{j=j_0}^{J} \sum_{k_j \in \mathbb{Z}^n} \sum_{l_j=1}^{2^n-1} \beta_{jk_j}^{[l_j]} \psi_{jk_j}^{[l_j]}(x), x \in \mathbb{R}^n, \tag{11}$$

where the equalities in (11) are in the sense of Lebesgue – a.e. Invoking the introduced spaces W_j, (11) can be equivalently rewritten as

$$V_J = V_{j_0} \bigoplus W_{j_0} \bigoplus W_{j_0+1} \bigoplus ... \bigoplus W_J = V_{j_0} \bigoplus \bigoplus_{j=j_0}^{J} W_j. \tag{12}$$

In applications involving processing of numerical data with sample size N, J is chosen dependent on $N : J = J(N)$. Since $\int_{\mathbb{R}^n} \psi(x)dx = 0$, the usual selection of $J(N)$ is such, that the size of the support of $\psi_{Jk_J}^{[l_J]}$ is comparable to the average step h_N between adjacent data points:

$$\text{diam}(\text{supp}\, \psi_{Jk_J}^{[l_J]}) \asymp h_N, \text{where}\, h_N \asymp \frac{1}{N}, \tag{13}$$

with equivalence constants independent of N. In view of the definition of $\psi_{Jk_J}^{[l_J]}$, (13) implies

$$J(N) \asymp \log_2 N \tag{14}$$

with equivalence constants independent of N. With this selection, the father wavelet (scaling function) φ_{Jk_J} acts as a consistent approximation of the Dirac δ-function at the point x, as long as $x \in \text{supp}\, \varphi_{Jk_J}$, and the rate of this approximation improves with the number of consecutive vanishing moments of ψ additional to the condition $\int\limits_{\mathbb{R}} \psi(x)dx = 0$ needed for consistent approximation (these would be $\int\limits_{\mathbb{R}} x\psi(x)dx = 0$, $\int\limits_{\mathbb{R}} x^2\psi(x)dx = 0$ etc.). With this selection, the coefficient α_{Jk_J} is taken to be equal to the value of f at the point where φ_{Jk_J} is concentrated as a δ-function.

So far, we have been considering (3 - 12) for $B^s_{pg}(\mathbb{R}^n)$ of functions defined on the whole space $\mathbb{R}^n$. This means that the subspaces V_j and W_j, $j \in \mathbb{Z}$, are (countably) infinite-dimensional. To make the construction computationally feasible, in numerical applications we limit the consideration to only those $f \in B^s_{pq}(\mathbb{R}^n)$ which are *compactly supported* with $\text{diam}(\text{supp}\, f)$ comparable to the diameter of the convex hull of the numerical data set (this numerical data set is finite, therefore its convex hull is a bounded subset of $\mathbb{R}^n$, so its closure is compact $(n < \infty)$). In this case, the subspaces $V_j, W_j, j \in \mathbb{Z}$ are all finite-dimensional, with dimensions depending on j, the distribution of the data set and its sample size N.

Now, let us extend the consideration to include also $f \in B^s_{pq}(\Omega)$, where $\Omega \subset \mathbb{R}^n$ is a nonvoid compact hyper-rectangle, i.e.,

$$\Omega = \overset{n}{\underset{i=1}{\times}} [a_i, b_i] \tag{15}$$

where " $\times$ " indicates Cartesian product, and $-\infty < a_i < b_i < +\infty,\ i = 1, ..., n$. The construction (3-12) continues to hold also in this case which is very similar to the case of $f \in B^s_{pq}(\mathbb{R}^n)$ such that it is compactly supported, with $supp\ f$ contained in the closure of the convex hull of the numerical data set. In the present case, the finite dimensions of V_j, W_j depend on j, the sizes of $b_i - a_i,\ i = 1, ..., n$, and the sample size N. Moreover, $j_0 \in \mathbb{Z}$ is bounded from below by $diam(\Omega)$:

$$j_0 \asymp \log_2 \text{diam}(\Omega), \tag{16}$$

with constants of equivalence possibly dependent on n, but not on N. There is one notable technical modification: the orthonormal wavelet bases in V_j and W_j, $j = j_0, j_0 + 1, ..., J$, contain *boundary-corrected* wavelet basis functions [CDV–1993a], [CDV–1993b]. For the theory of deep learning of n-dimensional manifolds developed here, there is no principal difference between the case of compactly supported $f \in B^s_{pq}(\mathbb{R}^n)$ and the case of $f \in B^s_{pq}(\Omega)$ with Ω a compact hyper-rectangle

in $\mathbb{R}^n$. Therefore, in our application we shall focus on the former one of these two cases, to avoid the construction of boundary-corrected wavelets.

Now, we are in a position to provide a definition of WNNs which is equivalent to the original definition in [ZB–1992] and [AZ–2013], but is in a new form which allows an insightful comparison with the new type of WBNNs.

Consider the left-hand side (LHS) of the identity in (11). Define first a 1-layer WNN with its Jk_J-th node being a neuron processing the α_{Jk_J}-th coefficient in the expansion in the LHS of (11). The edges of the WNN's graph are only the ones connecting the input to the Jk_J-th neuron and the ones connecting the Jk_J-th neuron to the output neuron where received results are summed up. In contrast to this construction, repeat the 1-layer NN but with 1–1 correspondence to the α-coefficients in V_{j_0} and the β-coefficients in $W_j, j = j_0, ..., J$ (in the right-hand side (RHS) of (11)). The definition of the edges of the graph of the 1-layer WBNN is the same as with the previous 1-layer WNN construction. The widths of the so-defined 1-layer WNN and WBNN are, of course, the same. A crucial advantage of the WBNN layer is its telescopic ordering which incorporates *the wavelet depth* into the *neural width* of the WBNN layer. Adding *neural depth* to the 1-layer WNN and WBNN is done in one and the same way: the next layers are added as intermediate between the already defined 1st layer and the output neuron, and each intermediate layer has exactly the same structure and ordering as the 1st layer.

As we shall see in the next sections, the learnability conditions and universal approximation theorems for each of WNNs and WBNNs ensure that 1-layer NNs (with the widths specified via the LHS and RHS of (11) and the compactness of $\operatorname{supp} f$) are sufficient for learning every element of the range of the respective approximation theorem. From this point of view, deep WNNs and WBNNs are theoretically redundant, but as we shall see, they provide a highly efficient computing architecture for acceleration of the rate of convergence of the approximation process by using iterative algorithms. A maximally sparse structure of the edges between the l-th and the $(l+1)$-st layer should be used (a neuron on the $(l+1)$-st level is only connected with its corresponding neuron at level l by way of the 1–1 correspondence between levels l and $l+1$). In practice, in the context of WNNs, the intermediate levels of the NN are used for iterative local optimization starting with the initial approximation provided at level 1. Although the performance of the deep WNN is expected to be better than the one of the 1-layer WNN, this can be expected to be noticeable only for very large sample sizes and respective very large number of iterations (very deep WNN). For example, in the case of Fig. 4 of [ZB–1992] the number of iterations is 10000. In the case of

WBNNs, the quality of initial approximation is expected to be very high, due to the efficient use of the wavelet depth within the layer of neural depth 1. As a consequence, in comparison with the very large sample size needed for acceptable performance of WNN, it can be expected for the initial approximation of the 1-layer WBNN to be sufficiently good for large to medium samples sizes, while as the ultimate approximation achieved by a deep WBNN (with the 1-layer WBNN as its initial layer) to be sufficiently good already for medium to small samples.

Notice the distinction we make between *learnability conditions* and *universal approximation theorems* for a given type of single-layer NN computing $f : \mathbb{R}^n \to \mathbb{R}$ for a given $n \in \mathbb{N}$. (To study the general case of parametric manifolds on $\mathbb{R}^n$, i.e., $f : \mathbb{R}^n \to \mathbb{R}^m$, $m \in \mathbb{N}$, $n \in \mathbb{N}$, it is sufficient to study it coordinate by coordinate, i.e., for $m = 1$ only.)

A *universal approximation theorem* (UAT) for a single-layer NN of width N is a *qualitative consistency* result (i.e., refers to existence of convergence in a given topology without specifying *quantitative rates of convergence*) when $N \to +\infty$ under the assumption of an *activation function* of specific type being used in the neural computation. For example, in the case of sigmoid activation, the respective UAT is due to Cybenko [C–1989], [C–1992] [1], and refers to continuous functions, while in the case of *Rectified Linear Unit* (ReLU) activation [LPW+–2017], the respective UAT refers to the more general class of functions in L_1 (see [LPW+–2017, Theorem 1]. The *learnability set* (LS) for a given single-layer NN of width N is the largest set of $f : \mathbb{R}^n \to \mathbb{R}$ which can be approximated by neural computation via this NN when $N \to +\infty$ *without the invocation of a specific activation function*, i.e., when *activation is via the default identity function*.

As mentioned above, both LSs and UATs are qualitative consistency results. In the next sections we shall show that it is possible to obtain quantitative upgrades of LSs and UATs where the consistency results are strengthened to results about concrete rates of approximation.

2 A new result about WNNs

In this section we shall show how the idea of using equivalent metric (as discussed in Section 1) can be used to generate an initial solution in a single-layer WNN which is a good starting point for local optimization in a deep WNN (having as 1st layer the said single-layer WNN). The universal approximation theorem invoked in [ZB–1992] is Cybenko's classical result, [C–1989], [C–1992], valid for continuous functions f. As noted in [AZ–2013], after the publication of [ZB–1992]

[1] In [ZB–1992] Cybenko's work has been imprecisely and incompletely cited. Here we provide the relevant corrected and complete citation [C–1989], [C–1992]

more general learnability conditions and universal approximation theorems were derived about WNNs. Namely, results about UATs for ReLU NNs were formulated and proved in [LPW+–2017] in relevance to the larger space of Lebesgue-integrable functions $f \in L_1(\mathbb{R}^n)$. The ideas of the proofs of the results in [LPW+–2017] make it possible to identify the LS for sufficiently wide NNs, including the case of WNNs with (14).

To present this result, here we shall use terminology which will allow us to compare this result to the respective result for WBNNs (derived in section 3 below). Namely, in the case of wavelets on $\mathbb{R}^n$, learnable by sufficiently wide single-layer WNN are all *regular distributions* f in $\mathcal{S}'(\mathbb{R}^n)$ - the dual of the Laurent Schwartz space $\mathcal{S}(\mathbb{R}^n)$ [RS–1980]. In the case of boundary-corrected wavelets on Ω-compact hyper-rectangle in $\mathbb{R}^n$ (see Section 1), learnable by sufficiently wide single-layer WNN are all regular distributions $f \in D'(\Omega)$ - the dual of the space $D(\Omega)$ [RS–1980] [2]. In both cases, the regular distributions f are exactly the elements of $L_{1,loc}(\Delta)$, $\Delta = \mathbb{R}^n$ or $\Delta = \Omega \subset \mathbb{R}^n$ (for Ω see Section 1). Here, as usual, $L_{1,loc}$ consists of all g defined Lebesgue – a.e. on Δ and such that for every compact subset $C \subset \Delta$ the statement $f\chi_C \in L_1(\Delta)$ holds true, where χ_C is the characteristic function of C :

$\chi_C(x) = 1$ for $x \in C$ and $\chi_C(x) = 0$ for $x \in \Delta \setminus C$

In the case $\Delta = \Omega$, Ω considered in Section 1 is a compact in $\mathbb{R}^n$, therefore, $L_{1,loc}(\Omega) = L_1(\Omega)$. Notice that in the wavelet context the width of WNN is exponential (solving (14) for N yields the equivalent $N \asymp 2^J$) and sufficiently large for $L_{1,loc}$ to be learnable via single-layer WNN, according to [LPW+–2017]. Therefore, 'deepening' the WNN does not result in increasing the set of learnable functions. However, deep WNN may offer the following advantage: while the universal approximation theorem for sufficiently wide single-layer WNN provides only consistency of the approximation (i.e., convergence exists but may be arbitrarily slow), the use of a deep WNN with the said single-layer WNN forming its 1st layer may result in *increasing the speed (rate) of approximation*.

Now we shall formulate a model problem for which we shall be able to automatically generate a single-layer WNN that is asymptotically optimal with respect to the paradigm based on (1) in Section 1. According to this paradigm, local optimization with starting point at the automatically generated solution at the 1st layer of a deep WNN is expected to provide the global minimum for asymptotically large sample sizes $(N \to +\infty)$.

[2] In [RS–1980] a slightly different notation is used: D_Ω instead of $D(\Omega)$ and D'_Ω instead of $D'(\Omega)$.

Model problem. Let $n = 1$. From a random sample with size N, learn the density f of an absolutely continuous cumulative distribution function $F : \mathbb{R} \to [0, 1]$. In this case, $f = F'$, $f \in L_1(\mathbb{R})$, $f \geq 0$ on $\mathbb{R}$, $\int_{\mathbb{R}} |f(x)|dx = \int_{\mathbb{R}} f(x)dx = 1$.

This problem was addressed in [DP–1997] and [DP–1998], as follows. The density f was approximated by a (father) wavelet expansion, using the basis of a frame more general than a biorthonormal upgrade of LHS in (11). To consider the construction in [DP–1997] and [DP–1998] strictly in our present context, it is necessary to consider only those particular cases for which the frame is orthonormal, thus corresponding to the LHS in (11). In [DP–1998, Remark 2.3.2] were identified all cases when the frame used in [DP–1997], [DP–1998] is orthonormal: namely, these are exactly the cases in the LHS of (11) where the scaling function φ in (11) is of Haar type, i.e., the normalized characteristic function

$$\varphi(x) = \frac{1}{a} \cdot \chi_{[-\frac{a}{2}, \frac{a}{2}]}(x - x_0) \tag{17}$$

where $a > 0, x_0 \in [-\frac{a}{2}, \frac{a}{2}]$, and a is chosen so, as to match with the selection of j_0 in the RHS of (11).

With this choice of φ in the LHS of (11), the random estimator of f is obtained by plugging in the LHS of (11) the *empirical density*

$$\hat{f}_N(x) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i) \tag{18}$$

where $\delta(\cdot)$ is the delta-function and $\{x_i,\ i = 1, ..., N\}$ is the sample data set. Here, as earlier, the selection of the level J in (11) is such that (14) holds. In view of (18), the Jk_J-th neuron in the WNN associated with the basis function φ_{Jk_J} in the LHS of (11) computes the *empirical coefficient*

$$\hat{\alpha}_{Jk_J} = \frac{1}{N} \sum_{i=1}^{N} \varphi_{Jk_J}(X_i),\ k_J \in \mathbb{Z} \tag{19}$$

In [DP–1998] the selection of metric in which the risk is measured was relevant to the expectation of "the generalized Cramér-von Mises loss" [DP–1998, sections 2.2 and 2.3]. One very valuable feature of the estimates of this risk obtained in [DP–1998] was that they revealed the precise interconnection between the density's smoothness and the weight of its tails as $x \to \pm\infty$. In our present study we made the natural assumption of compactness of the density's support in correspondence with the

compactness of the convex hull of the sample data set. For densities with compact support with fixed diameter of the support, the risk estimates in [DP–1998] simplify and are only dependent on the smoothness of the density, and the following new result holds true.

Theorem 1. *Under the above assumption about the compact support of the density f, assume $f \in B^s_{p\infty}(\mathbb{R}),\ 0 < p \leq \infty,\ 0 < s < 2$, and (14) holds. Let the single-layer WNN associated with the Haar-type basis (17) used in (11) be defined as above, with (19) holding true. Let the risk of estimating f via the empirical density $\hat{f}_N$ in (18) be defined as in [DP–1998, Section 2.2]. Then,*

(i) The risk $R(f, \hat{f}_N)$ of learning f by the neural computations (19) in the LHS of (11) is:

$$R(f, \hat{f}_N) \asymp N^{-\frac{s}{1+2s}} \tag{20}$$

with constants of equivalence dependent on s and the fixed size of the support of f, but not on the choice of $f \in B^s_{p\infty}(\mathbb{R})$.

(ii) The rate in the RHS of (20) is asymptotically optimal *in the sense of [DP–1998, Theorem 2, 3.2].*

As a consequence of Theorem 1, under its assumptions, the single-layer WNN computing (19) generates automatically an element of V_J in (12) which is a good starting point for optimization search in V_J when the sample size N (and $J \asymp \log_2 N$) is asymptotically large ($N \to \infty$). For such N and J, using a deep WNN upgrade of the single-layer WNN (with the latter being the 1st layer of the deep WNN) it is possible to obtain an essential improvement of the learning of f within V_J and, possibly, even obtain globally optimal solution of the iterative optimization search in V_J performed by the deep WNN computing architecture. Thus, we have provided an instance when the equivalent-metric paradigm based on (1) in Section 1 offers an efficient AI alternative of the use of natural intellect in designing good starting point of deep WNN-compatible iterative optimization search, as discussed in [LBH–2015] – see Section 1. Although results of the type of Theorem 1 provide a rigorous mathematical justification of the use of AI based on deep WNNs, in practice, notable improvement can be generally expected only, or almost only, for very large samples sizes N.

Let us note two additional new features in Theorem 1 and its proof.

- Theorem 1 shows that the optimal estimation rate can be achieved when the activation function is the default identity (i.e., for the empirical density).

- The quantitative result involving rates is achieved in Theorem 1 under less restrictive assumptions for WNNs than the assumptions on generic NNs for the qualitative universal approximation theorem in [LPW+–2017, Theorem 1] in the sense that the latter NN must be *fully connected* (i.e., with *densely distributed edges* between the nodes of the NN) while the former WNN has very sparse edge distribution.

In the remaining part of this exposition, we shall show that the alternative of using WBNNs associated with the RHS of (11) and (12) provides highly efficient AI algorithms with quality practical results achieved already for medium to small sample sizes N.

3 WBNNs: learnability and universal approximation

To study learnability conditions and universal approximation theorems in the case of WBNNs it will be necessary to study some properties of the scale of Besov spaces $B^s_{pq}(\mathbb{R}^n)$ as defined via (5, 6). (The case of $B^s_{pq}(\Omega)$ using boundary-corrected wavelets in (5, 6) can be studied, mutatis mutandis, but our focus will continue to be on the case $\Omega = \mathbb{R}^n$.) To study the necessary aspects of the properties of the Besov-space scale $\{B^s_{pq}(\mathbb{R}^n),\ 0 < p, q \leq \infty, s \in \mathbb{R}\}, n \in \mathbb{N}$, we need some preparation, as follows.

- For the concept of *quasinorm* (or c-norm (c-quasinorm) with $c \geq 1$ being the constant in the quasi-triangle inequality), we refer to [BL–1976, Section 3.10]. See also there for the definition of *quasinormed abelian groups*.
- By [BL–1976, Lemma 3.10.1], a quasinormed abelian group A with c-quasinorm $||.||_A, : c \geq 1$, is *metrizable*, in the sense that the ρ-power A^ρ of A with 1-quasinorm $||.||^\rho_A$, $0 < \rho = \frac{1}{1+\log_2 c} \leq 1$, is a metric space with $d(a, b) = ||a - b||^\rho_A$.
- A necessary and sufficient condition for a normed space A to be *complete* (i.e. for A to be a *Banach space* is given in [BL–1976, Lemma 2.2.1].
- The previous item is being generalized in [BL–1976, Lemma 3.10.2] to a necessary and sufficient condition for a quasinormed abelian group A to be *complete*. (If A is not only abelian group, but also a vector space, then, endowed with the quasinorm $||.||_A$, it is called *quasinormed space* and, if it is also complete, *quasi-Banach space*.
- For example, consider the vector space l_2 of all sequences: $= (x_1, ..., x_n, ...), n \in \mathbb{N}, x_j \in \mathbb{R}$ or $x_j \in \mathbb{C}, j \in \mathbb{N}$, with quasinorm $||x||_{l_r} = (\sum_{j=1}^{\infty} |a_j|^r)^{\frac{1}{r}}$, $0 < r < \infty$, or $||x||_{l_\infty} = \max_{j \in \mathbb{N}} |a_j|$,

$r = \infty$. By the theory in [BL–1976, Section 3.10], l_r is a *Banach space* for $r : 1 \leq r \leq \infty$ and *quasi-Banach space* when $r : 0 < r < 1$. In the latter case, the constant c in the quasi-triangle inequality for $||.||_{l_r}$ is $c = 2^{\frac{1-r}{r}} > 1$; for the power ρ one gets $\rho = r \in (0, 1)$ and the 1-quasinormed abelian group $(l_r)^r$ with 1-quasinorm $||.||_{l_r}^r$ is a metric space with metric $d(a, b) = ||a - b||_{l_r}^r$.

- Using the properties of l_r from the previous item, it is possible to establish that B_{pq}^s, as defined via (5, 6), are Banach spaces for $1 \leq p \leq \infty$, $1 \leq p \leq \infty$, and quasi-Banach spaces when $0 < p < 1$ and/or $0 < q < \infty$ [T–1983].

- Another aspect of the theory of the Besov-space scale which proves to be relevant is *the lifting property* of *the Bessel potential* J^σ, $\sigma \in \mathbb{R}$, in the Besov-space scale. Following the exposition in [T–1983, Section 1.2.1], we define the Fourier transform $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$ first on $\mathcal{S}(\mathbb{R}^n)$, and then extend it to $\mathcal{S}'(\mathbb{R}^n)$, after which, following [T–1983, Section 2.3.8], we define the Bessel potential $J^\sigma : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R}^n)$, as follows:

$$J^\sigma f = \mathcal{F}^{-1}[(1 + |.|^2)^{-\frac{\sigma}{2}} \mathcal{F} f], f \in \mathcal{S}'(\mathbb{R}^n). \tag{21}$$

where $\sigma \in \mathbb{R}$. Now the *σ-lifting property* of the Bessel potential in the Besov-space scale can be formulated, as follows [T–1983, Section 2.3.8]. J^σ acts bijectively on $\mathcal{S}'(\mathbb{R}^n)$ and the restriction of J^σ on $\mathcal{S}(\mathbb{R}^n)$ acts bijectively on $\mathcal{S}(\mathbb{R}^n)$. Moreover, if s, p, q are as in (5, 6) and $f \in B_{pq}^{s_1}(\mathbb{R}^n)$, where $s_1 \in \big(-\infty, n(\frac{1}{p} - 1)_+ \big] \cup \big[r, \infty \big)$ with $\sigma : s = s_1 + \sigma$, then $J^\sigma f \in B_{pq}^s$ and formulae (5, 6) can be applied on $g = J^\sigma f$ and

$$||J^\sigma f||_{B_{pq}^s(\mathbb{R}^n)} \asymp ||f||_{B_{pq}^{s-\sigma}(\mathbb{R}^n)} \tag{22}$$

Moreover, using (22) when $s_1 \in (-\infty, n(\frac{1}{p} - 1)_+]$, i.e., for $\sigma > 0$ allows to extend the wavelet-based representation (5) and the quasinorm definition (6) for arbitrary $s \in \mathbb{R}$, i.e., including also singular distributions like the δ-function and its derivatives which are not in $L_{1,loc}$ [1]. Indeed, choose and fix any $s_1 \in \mathbb{R}$ and select and fix σ such that $n(\frac{1}{p} - 1)_+ < s - \sigma < r$. Then (5,6) will make sense for f replaced by $J^\sigma f$ and (22) can be used to define an equivalent quasinorm in $B_{pq}^{s_1}(\mathbb{R}^n)$.

Now we are ready to formulate and prove the following results about WBNNs.

[1] This fact is relatively easy to derive even in the n-dimensional case, using the theory of Fourier multipliers on $L_p(\mathbb{R}^n)$, $1 \leq p \leq \infty$, see [BL–1976, Section 6.1], [BTW–1975, Chapter 1] and [HL–2017, Introduction].

Theorem 2. *Using arbitrary samples with size N with $J(N)$ satisfying (14), $f \in \mathcal{S}'(\mathbb{R}^n)$, is learnable for $N \to \infty$ by WNNs if, and only if (iff) f is also learnable by WBNNs, i.e. the learnability sets by WNNs and by WBNNs coincide.*

Theorem 3. *Let N and $J(N)$ be as in Theorem 2, and let $f \in B^s_{pq}(\mathbb{R}^n)$, $0 < p \leq \infty$, $0 < q \leq \infty$, $s \in \mathbb{R}$. Then, for any $r : n(\frac{1}{p} - 1)_+ < r < \infty$ and orthonormal wavelet basis satisfying (3, 4) and for every $\sigma \in \mathbb{R}$ such that $s - \sigma \in (n(\frac{1}{p} - 1)_+, r)$ it holds true that $J^\sigma f$ is learnable by the WBNN generated by the said wavelet basis.*

Corollary 1. *The space $B^s_{pq}(\mathbb{R}^n)$ is contained in the learnability set of WBNN for every $s \in \mathbb{R}$, $0 < p \leq \infty$, $0 < q \leq \infty$.*

Corollary 1 implies that the learnability set of WBNNs contains not only all regular distributions in $\mathcal{S}'(\mathbb{R}^n)$, but also singular distributions, since Besov spaces with $s < 0$ do contain singular distributions.

Theorem 2 now suggests that Corollary 1 extends also to WNNs, but here lies one big difference between the *efficient* use of WBNNs and WNNs. Recovering f from $g = J^\sigma f$ requires approximate numerical computation of $J^{-\sigma} g$ which is very numerically sensitive to errors in the computation of g especially when f can be a singular distribution. Since for a given sample with size N the quality of learning g via WNNs is expected to be much worse than via WBNNs, the deterioration of the recovery of f from g when using WNNs would be much more exacerbated compared to the use of WBNNs so that the only case of σ for which the use of WNNs could be marginally acceptable is the trivial case $\sigma = 0$. (A detailed error analysis of the computations for $\sigma \neq 0$ would require the invocation of aspects of Paley-Wiener theory [FJW–1991, Sections 4–6], including sampling results of Shannon type [FJW–1991, Theorem 6.4] which goes beyond the study of AI aspects considered here.) Theorems 3 and 2 now imply that WNNs can be efficiently used (albeit only marginally for very large sample sizes N only) for learning $f \in B^s_{pq}(\mathbb{R}^n)$ only for the original range s, p, q for which (5) was formulated. Note that for the these values of s, p, q $B^s_{pq}(\mathbb{R}^n) \subset L_{1,loc}(\mathbb{R}^n)$.

The results obtained here for wavelet bases on $\mathbb{R}^n$ can in principle be reformulated for boundary-corrected wavelets on a compact hyperrectangle in $\mathbb{R}^n$, this modification is technically involved. For example, the lifting property of the Bessel potential has to be replaced by a respective property of fractional integro-differential operators of Riemann-Lionville, Grünwald-Letnikov, Caputo and other types under additional assumptions for each of these types [SKM–1993].

As far as UAT for WBNNs is concerned, it is much more rich and diversified than UAT for WNNs, due to the much more flexible telescopic structure of the RHS in (12). While in the case of WNNs the focus has been only on sigmoid and ReLU activation, in the case of WBNNs there is a great variety of meaningful activation methods, each of which is with its own UAT and own range of practical applications. In this section we study the common features of all these activation methods and provide a classification of these methods into two general subclasses, together with the general range of applications for each of these subclasses.

Any activation method can be defined as a (generally, *nonlinear*) operator Λ acting on the space sum in the RHS of (12) and being of *shrinkage* type, i.e. having the following properties.

(i) The restriction of Λ on V_{j_0} coincides with the identity on V_{j_0}, i.e.,

$$\Lambda(\alpha_{j_0k_0}\varphi_{j_0k_0}) = \alpha_{j_0k_0}\varphi_{j_0k_0} \tag{23}$$

for every $j_0, k_0, ...$

(ii) Using the Euler representation of $z \in \mathbb{C}$

$$z = |z|(\cos(\arg z) + i\sin(\arg z)), \arg z \in [0, 2\pi), \tag{24}$$

the action of Λ on the space $\bigoplus_{j=j_0}^{J} W_j$ in the RHS of (12) is defined such, that

$$\Lambda(\beta_{jk_j}^{[l_j]}\psi_{jk_j}^{[l_j]}) = \tilde{\beta}_{jk_j}^{[l_j]}\psi_{jk_j}^{[l_j]} \tag{25}$$

where

$$\arg\tilde{\beta}_{jk_j}^{[l_j]} = \arg\beta_{jk_j}^{[l_j]} \tag{26}$$

$$|\tilde{\beta}_{jk_j}^{[l_j]}| \leq |\beta_{jk_j}^{[l_j]}| \tag{27}$$

for every (j, k_j) participating in the formation of $\bigoplus_{j=j_0}^{J} W_j$.
Notice that when $\beta_{jk_j}^{[l_j]} \in \mathbb{R}$ (26) reduces to

$$\operatorname{sgn}\tilde{\beta}_{jk_j}^{[l_j]} = \operatorname{sgn}\beta_{jk_j}^{[l_j]} \tag{28}$$

where for $x \in \mathbb{R}\backslash\{0\}$

$$\operatorname{sgn} x = \begin{cases} +1, & x > 0; \\ -1, & x < 0; \\ \text{undefined}, & x = 0; \end{cases} \tag{29}$$

and for the case $z = x = 0$ it is convenient to define

$$\arg 0 = \operatorname{sgn} 0 = 0 \tag{30}$$

Clearly, Λ has the special property that it preserves the directrice and respective orientation of every basis function in V_{j_0} and $\bigoplus_{j=j_0}^{J} W_j$. It is also clear that, in general, Λ is nonlinear, since the shrinkage is individual for every basis function. Now, we divide all possible activation methods Λ with properties (i) and (ii) into two disjoint subclasses, as follows.

A. *Non-threshold-type* activation methods have the following additional property

(iii) for any selection of the coefficient vector $\{\alpha_{j_0 k_0}, \beta_{jk_j}^{[l_j]}\}$ in the RHS of (11), such that $\beta_{jk_j}^{[l_j]} \neq 0$ for some triple (j, k_j, l_j), it is fulfilled that $\tilde{\beta}_{jk_j}^{[l_j]} \neq 0$ holds true. (In other words, there is only reducing $|\beta_{jk_j}^{[l_j]}| > 0$ without ever "killing" the coefficient $\tilde{\beta}_{jk_j}^{[l_j]}$, i.e., having $|\tilde{\beta}_{jk_j}^{[l_j]}| = 0$.

B. We shall say that the activation method Λ is of *threshold-type*, if (iii) is not fulfilled for Λ.

In the second part of this study we shall study an important model example of activation of WBNNs resulting in *learning geometric manifolds with compression*. The analysis of concrete examples will show that there is an intrinsic separation of geometric manifolds into ones that are highly compressible and ones that are not. From a geometric point of view, the latter class of manifolds will be identified as *fractal-type* while the former class consists of manifolds of *piecewise-smooth type*. Our forthcoming study [DPS+–2022a] of diverse activation methods of both threshold and non-threshold type (type B and A) will demonstrate that activation of threshold type is performing well when learning piecewise-smooth manifolds, while activation of non-threshold type performs well when learning manifolds of fractal type.

4 Particle vs multiagent simulation and swarm vs. deep evolutionary AI

A very new research topic in AI research is establishing connection between *swarm* AI and deep neural networks by the invocation of *evolutionary algorithms* [I–2018], [I–2022]. Scientifically, this is a very new field, but conceptually it appeared in some of the most famous early sci-fi novels

[H–1957], [L–1964] (latest English translation [L–2020]) which were written in the first few years after the concept of AI emerged as a term at the workshop "Dartmouth Summer Research Project on Artificial Intelligence" at Darthmouth College, Hannover, NH, USA in 1956, to designate a specialized branch of cybernetics. Both of the novels of Fred Hoyle and Stanisław Lem successfully predicted the development of important modern scientific trends: the latter – nanotechnology and swarm AI; the former – deep learning by AI systems and connections with evolutionary algorithms.

Our present interest to the connection between swarm and deep evolutionary AI is only limited to its computational aspects. From this limited point of view, the above-said connection can be considered as a particular case of particle simulation and multiagent simulation (i.e., simulation of systems involving large numbers of relatively simple agents vs. simulation of systems involving small to moderate number of agents with relatively high level of individual intelligence features). Notice that the most efficient simulation of each of these two types of system is performed on different types of parallel computing architectures.

(a) Swarm of sufficiently broad single-layered NNs (including single-layered WBNNs with (14)) – CPU parallelism – (relatively expensive) large-size multi-CPU supercomputing architectures; e.g., hypercubic [L–1992].

(b) Deep (multi-layered, sufficiently broad) NNs (including deep WBNNs with (14)) – GPU parallelism – (relatively cheap) small-size multi-GPU computing architectures using GPGPU-programming – currently in CUDA, and more recently, Python [OUN+–2017].

Using connections between swarm intelligence and deep NNs [I–2018], [I–2022], it is possible to emulate the performance of the expensive computing architectures in item (a) by the cheap computing architectures in item (b), but at the inevitable price of some loss of efficiency. Ideally, hybrid multi-CPU multi-GPU computing architectures should be recommended.

5 Best activation of WBNNs for fastest learning and maximal compression

In [DRP–1999] was considered and systematized a rich diversity of threshold and non-threshold wavelet shrinkage methods for *nonparametric statistical estimation* of *densities* and *denoising of nonparametric regression functions*. In [DPS+–2022b] we shall show that each of these shrinkage methods gives rise to respective activation of WBNNs, generating highly efficient learning algorithms. Moreover, in some cases these learning algorithms can be shown to be *best possible* with respect to certain aspects which are important for applications. As a model example of the high

efficiency of learning with WBNNs, we shall study here the activation induced by only one of these wavelet shrinkage methods, namely, the one discussed in [DRP–1999, Appendix B10 b)].

For the sake of maximal clarity, we shall consider here only the simplest univariate case $n = 1$. This will be a very clear illustration of the optimal speed of learning and compression in model examples of curve learning in the next sections. The general case $n \in \mathbb{N}$ and some graphical visualization for the cases $n = 2$ and $n = 3$ will be considered in [DPS+–2022b].

Assume $f \in B^s_{pq}(\Omega)$ where $\Omega = \mathbb{R}$ or $\Omega = [a, b]$, $-\infty < a < b < \infty, 0 < p \leq \infty, 0 < q \leq \infty$ and $(\frac{1}{p} - 1)_+ < s < r$. Assume also that both the metric power index p_1, the metric logarithmic index q and the smoothness index s are *exact*, that is, $f \notin B^{s_1}_{p_1 q_1}(\Omega)$ for any $p_1 : 0 \leq p_1 < p, 0 < q_1 \leq q$ and any $s_1 : s_1 > s$.

As usual in our present study, when considering the domain $\mathbb{R}$, we shall be making the default assumption about compactness of $\mathrm{supp}\, f$ (in the case of boundary-corrected wavelets and $\Omega = [a, b]$ with $-\infty < a < b < +\infty$, we do not need this default assumption, i.e., it is possible that $f(a+) \neq 0$ and/or $f(b-) \neq 0$). For the index triple (p, q, s) consider now the *Sobolev embedding plane* passing through the point (p, q, s), i.e.,

$$\{(\rho, \eta, \sigma) : \sigma - \frac{1}{\rho} = \tau(p, s) = s - \frac{1}{p}, 0 < \rho \leq \infty, 0 < \eta \leq \infty\} \tag{31}$$

What is important about this selection is the well-known embedding

$$B^s_{pq}(\Omega) \hookrightarrow B^\sigma_{\rho\eta}(\Omega),\ \sigma - \frac{1}{\rho} = s - \frac{1}{p},\ 0 < p \leq \rho \leq \infty,\ 0 < q \leq \eta \leq \infty \tag{32}$$

where $\Omega = \mathbb{R}$ or $\Omega = [a, b]$.

(For two quasinormed spaces A and B, the notation $B \hookrightarrow A$ ("B is embedded/imbedded in A") means $B \subset A$ and $\exists c \in (0, \infty) : ||b||_B \leq c||b||_A$ for any $b \in B$.)

Due to the Sobolev embedding/imbedding (32), our assumption $f \in B^s_{pq}(\Omega)$ implies

$$f \in B^\sigma_{\rho\eta}(\Omega) \tag{33}$$

for any $(\rho, \eta, \sigma) : \sigma - \frac{1}{\rho} = s - \frac{1}{p}$, $0 < p \leq \rho \leq \infty$, $0 < q \leq \eta \leq \infty$. In [DRP–1999, Appendix B10 b)] it was explained that for the Besov spaces $B^\sigma_{\rho\eta}$ with (ρ, η, σ) lying on one and the same

Sobolev embedding plane, an important part of the function-space theory is related with the so-called *decreasing rearrangement* of f in all Besov spaces $B^{\sigma}_{\rho\eta}$ with (ρ, η, σ) on the same Sobolev embedding plane. A detailed consideration of the concept of decreasing rearrangement can be found in [BL–1976, Section 1.3] and the Peetre-Krée formula [BL–1976, Theorem 3.6.1] together with [BL–1976, 3.14.5.6 and Theorem 5.2.1 (2) for $q = p$ in local notation] (see also [DRP–1999, (B8)]. For our purposes in our present context it is sufficient to consider the *normalized decreasing rearrangement* of the β-coefficients in the series (5) and in $\bigoplus_{j=j_0}^{J} W_j$ in its truncation (11, 12), as follows (compare with [DRP–1999, Appendix B10 b), items b1 and b2]). Recall that here we consider only the case $n = 1$ in (5-16) - in particular, in (5-8, 11) this implies $l = 1$. Thus, in the sequel of the present definition of *decreasing rearrangement*, we shall be skipping the index l.

b1) Fix $j_0 \in \mathbb{Z}$ (with no loss of generality, it can be assumed that $j_0 = 0$). Consider all (j, k) in $(5, 6)$ such that $\operatorname{supp} \psi_{jk} \cap \operatorname{supp} f \neq \emptyset$. Denote the set of all such (j, k) by $I(f, \psi) = I(f, \psi, j_0)$. It is clear that for every fixed $j = j_0, j_0 + 1, ...$ in the generalization of (5, 6) involving j_0 the number M_j of elements of $I(f, \psi)$ from the j-th level does not exceed $c(f, \psi) \cdot 2^j$, for some $c(f, \psi) \in (0, \infty)$. Therefore, M_j is finite for any $j = j_0, j_0 + 1, ...$, but $M = \sum\limits_{j=j_0}^{\infty} M_j$ is, generally, *not finite*. On the other hand, for the truncation (11,12) the number $m(j_0, J) = \sum\limits_{j=j_0}^{J} M_j$ is finite, with $M = \lim_{J\to\infty} m(j_0, J)$. Denote by $i(f, \psi, j_0, J)$ the subset of $I(f, \psi, j_0)$ such that (j, k) participates in the formation of the truncation (11) and $\bigoplus_{j=j_0}^{J} W_j$ in (12).

The number of elements of $i(f, \psi, j_0, J)$ is

$$m(j_0, J) \leq c(f, \psi) \cdot 2^{j_0} \sum_{k=0}^{J-j0} 2^k = c(f, \psi) 2^{j_0} \cdot \frac{2^{J-j_0+1} - 1}{2 - 1} = c(f, \psi)(2^{J+1} - 2^{j_0}) \leq 2c(f, \psi) 2^J, \tag{34}$$

regardless of the choice of j_0.

b2) Recalling that $\tau = \tau(p, s) = s - \frac{1}{p}$ in (31), for every $(j, k) \in I(f, \psi)$ normalize $|\beta_{jk}|$ by multiplying with the factor $2^{j(\tau+\frac{1}{2})}$ and consider the decreasing rearrangement $\{b_\nu : \nu = 1, ..., M\}$ of the (possibly, infinite) set $\{2^{j(\tau+\frac{1}{2})}|\beta_{jk}| : (j, k) \in I(f, \psi)\}$.

In the case of the truncation (11, 12), we get the subset $\{2^{j(\tau+\frac{1}{2})}|\beta_{jk}| : (j, k) \in i(j_0, J)\}$ which is finite, with number of elements $O_{f,\psi}(2^J)$, according to (34).

For this model case, the activation operator $\Lambda = \Lambda_\delta$ is of threshold type, with threshold $\delta \in (0, \infty)$, defined in the following way. Let the decreasing rearrangement of $I(f, \psi)$ be $\{b_\nu, \nu = 1, ..., M\}$, with $(j, k) \in I(f, \psi)$ being ordered in a respective sequence $\{(j_\nu, k_\nu), \nu = 1, ..., M\}$ where (j_ν, k_ν) corresponds to b_ν for any $\nu = 1, ..., M$.

Then

$$\Lambda(\beta_{j_\nu k_\nu}\psi_{j_\nu k_\nu}) = \begin{cases} 0, & \text{if } 2^{j_\nu(\tau+\frac{1}{2})}|\beta_{j_\nu k_\nu}| \in (0, \delta) \\ \beta_{j_\nu k_\nu}\psi_{j_\nu k_\nu}, & \text{if } 2^{j_\nu(\tau+\frac{1}{2})}|\beta_{j_\nu k_\nu}| \geq \delta \end{cases} \tag{35}$$

The selection of the threshold δ depends on the concrete context of the learning process. We shall call every threshold activation method designed via the sequence of steps b1) and b2) *a decreasing rearrangement activation method*. For this type of activation method with threshold δ, the UAT corresponds to $\delta \to 0+$ and is given by the following theorem.

Theorem 4. *Let $\delta \to 0+$ in (35), and let f be as assumed above. Then, the summands in the series in the RHS of (5) can be commuted in such a way that (5) becomes*

$$\sum_{k\in\mathbb{Z}} \alpha_{j_0 k}\varphi_{j_0 k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x) = f(x) = \sum_{k\in\mathbb{Z}} \alpha_{j_0 k}\varphi_{j_0 k} + \sum_{\nu=1}^{M} \beta_{j_\nu k_\nu}\psi_{j_\nu k_\nu}(x) \tag{36}$$

where the RHS converges to $f(x)$ Lebesgue – a.e. in $\mathbb{R}$, as well as in the topology of $B^s_{pq}(\mathbb{R})$ and $B^\sigma_{\rho\eta}(\mathbb{R})$ for any $(\rho, \eta, \sigma) : 0 < p \leq \rho \leq \infty, 0 < q \leq \eta \leq \infty$, $\sigma - \frac{1}{\rho} = s - \frac{1}{p} = \tau$.

Theorem 4 continues to hold true for boundary–corrected wavelets and $\Omega = [a, b]$, with respective modifications in (5) and (36).

We shall now upgrade the qualitative result of Theorem 4 by formulating a quantitative result which proves to be best possible in a certain sense specified below. Among all Besov spaces with (quasi)norm (6), the ones which are Hilbert spaces are exactly

$$B^s_{pq}(\Omega) \text{ with } p = q = 2 \text{ and } 0 < s < r, \tag{37}$$

where, for $n = 1$ considered here, $\Omega = \mathbb{R}$ or $\Omega = [a, b], -\infty < a < b < +\infty$. Choose arbitrary triple (p, q, s) such that $0 < p \leq 2$, $0 < q \leq 2$, $(\frac{1}{p} - 1)_+ < s < r$, and consider the respective triple $\rho = \eta = 2$, $\sigma - \frac{1}{\rho} = s - \frac{1}{p}$. For (5, 6) to hold for this choice of (ρ, η, σ) it is necessary that

$$0 \leq \sigma < r \tag{38}$$

holds where $\sigma = 0$ corresponds to the case $B^0_{22}(\Omega) = L_2(\Omega)$. Therefore, (5, 6) hold simultaneously for the couples (p, q, s) and $(2, 2, \sigma)$, iff the following inequalities and equalities are simultaneously

$$(\frac{1}{p} - 1)_{+} < s < r, \quad 0 \leq \sigma < r, \quad \sigma = s - \frac{1}{p} + \frac{1}{2}, \quad p \leq 2, \quad q \leq 2. \tag{39}$$

Solving (39) for p, q and s, we obtain

$$\frac{1}{r + \frac{1}{2}} < p \leq 2, \quad 0 < q \leq 2, \quad \frac{1}{p} - \frac{1}{2} \leq s < r, \tag{40}$$

under which assumptions (38) holds. Consider the orthocomplement

$$O_{j_0\sigma} = O\big(V_{j_0}, B^\sigma_{22}(\Omega)\big) = V_{j_0}^{\perp(B^\sigma_{22}(\Omega))} \tag{41}$$

of V_{j_0} in $B^\sigma_{22}(\Omega)$, with respect to the inner product in $B^\sigma_{22}(\Omega)$, $0 \leq \sigma < r$. This orthocomplement is well defined because $\{\varphi_{j_0k_0}, \psi_{jk_j}\}$ is an unconditional Riesz basis in all Besov spaces where (5, 6) hold, and $B^\sigma_{22}(\Omega)$ is a Hilbert space, so that $V_{j_0}^\perp$ is well-defined with respect to the inner product in $B^\sigma_{22}(\Omega)$, $0 \leq \sigma < r$. For a given $f \in B^\sigma_{22}(\Omega)$, define $f_{j_0} \in V_{j_0}$ as follows

$$f_{j_0} = \sum_k \alpha_{j_0k}(f)\varphi_{j_0k} \tag{42}$$

Let $k \in \mathbb{N}$ and consider an arbitrary subspace S_k with $\dim S_k = k$, such that

$$S_k \subset O_{j_0\sigma} \subset B^\sigma_{22}(\Omega), \tag{43}$$

and define the best-approximation functional

$$E_k(f; B^\sigma_{22}(\Omega)) = \inf_{s \in S_k} \inf_{S_k \subset O_{j_0\sigma}} ||f - f_{j_0} - s||_{B^\sigma_{22}}. \tag{44}$$

Now, we are in the position to formulate the following remarkable result.

Theorem 5. *Assume that f is as in Theorem 4 with the additional assumption that (40) holds. Then,*

$$||f - \sum_{\nu=1}^{k} \beta_{j_\nu k_\nu} \psi_{j_\nu k_\nu}||_{B_{22}^{s-\frac{1}{p}+\frac{1}{2}}(\mathbb{R})} = E_k\big(f; B_{22}^{s-\frac{1}{p}+\frac{1}{2}}(\mathbb{R})\big), k = 1, 2, ... \tag{45}$$

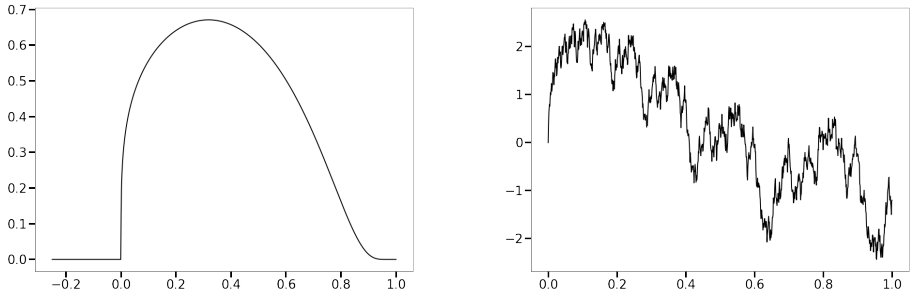
The result (45) holds, *mutatis mutandis*, also for the case of boundary-corrected wavelets and $\Omega = [a, b]$, $-\infty < a < b < +\infty$.

Theorem 5 shows that using a sufficiently broad (i.e., satisfying (14)) single-layered WBNN for learning curves with Besov regularity while using the current activation method results in a learning strategy which is *optimal* in the following two senses.

1. *Fastest learning* – using a fixed number of active neurons, the learned function is closest possible to the original, with the distance measuring the closeness being in terms of $||\cdot||_{B_{22}^{\sigma}}$, $0 \leq \sigma < r$, that is, taking into account not only position in space $(\sigma = 0)$ but also fractional derivatives up to order r.
2. *Maximal compression* – for a benchmark determined by a fixed distance between the target function and its learned version measured in terms of $||\cdot||_{B_{22}^{\sigma}}$, the benchmark result is achieved with the fewest possible activated neurons.

In the remaining part of the present study, we shall illustrate graphically aspects 1. and 2. of the optimality of the learning process with WBNNs when the current activation method is being used.

6 Representative model examples



(a) "λ-tear" (b) Weierstrass type fractal

Figure 1: Piecewise smooth type vs fractal type manifolds

We shall consider in detail graphical visualization related to two model examples which are representative in three important aspects.

1. The first one (the "λ-tear") is a typical manifold of *piecewise smooth type*, while the second one (Weierstrass-type curve) is a typical manifold of *fractal type*.
2. For both of them, their *exact* metric power, metric logarithmic and smoothness index of their Besov regularity is known.
3. The exact metric power, metric logarithmic and smoothness index of Besov regularity can be selected to be the *same* for both examples, which allows for insightful graphical comparisons.

Example 1 (see Fig. 1a). *The "λ-tear"*

$$f_1(x) = \begin{cases} x_+^{\lambda} \exp(-\frac{x^2}{1-x^2}) & x \in (0,1) \\ 0 & x \in [-1,0] \cup [1,2] \end{cases} \tag{46}$$

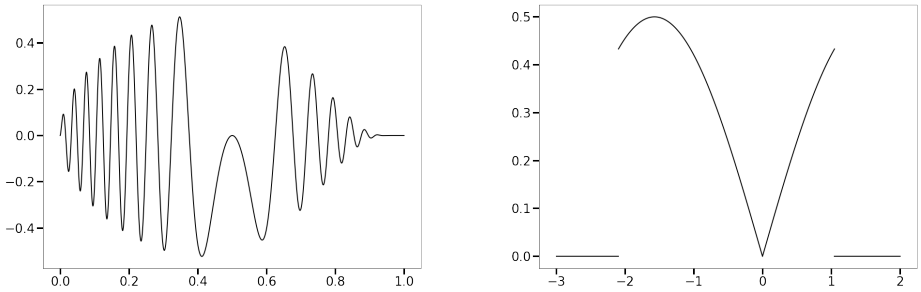
where $\lambda \in (0,1)$. This function is analytic for $x \in [-1,0) \cup (0,1) \cup (1,2]$; it is C^∞, but not analytic at $x = 1$; at $x = 0$ it is only C^0. Its exact Besov regularity for $p : 1 \le p \le \infty$ is $f \in B_{p\infty}^{\lambda+\frac{1}{p}}(\Omega)$, where $\Omega = \mathbb{R}$ or $\Omega = [-1,2]$ [BTW–1975, Proposition 2.4.2], see also [DRP–1999, Section 7, Example 1].

Example 2 (see Fig. 1b). *Weierstrass-type curve*

$$f_2(x) = \sum_{k=0}^{\infty} 1.5^{-\tau k} \sin(1.5^k \times 5x), \quad x \in \mathbb{R} \tag{47}$$

where $\tau \in (0,2)$. For the purpose of comparing with Example 1, we shall consider only the restriction of f_2 onto $\operatorname{supp} f_1$, i.e., for $x \in [0,1]$.

The graph of f_2 is a typical *self-similar monofractal* with constant local Hölder index τ and constant local fractal dimension $2 - \tau$ which is also its global fractal dimension on $[0,1]$. When considering $\Omega = \mathbb{R}$, for any compactly supported $\chi \in C^\infty(\mathbb{R})$ such that $[0,1] \subset \operatorname{supp} \chi$ and $\chi \equiv 1$ on $[0,1]$, $\chi \cdot f_2 \in B_{p\infty}^{\tau}(\mathbb{R})$ [BTW–1975, Proposition 2.4.1 for the imaginary part G_t in local notations, with additional rescaling], see also [DRP–1999, Section 7, Example 2]. For the restriction $\overline{f}_2 = f_2\Big|_{\operatorname{supp} f_1}$ in the case of boundary-corrected wavelets with $\Omega = [0,1]$, we have directly $\overline{f}_2 \in B_{p\infty}^{\tau}([0,1])$. For



(a) "Double chirp" (b) Sinusoidal density

Figure 2: Curves used to study Besov regularity

every $p : 1 \leq p \leq \infty$ this Besov regularity of $\overline{f}_2$ is *exact*. Clearly, when $\tau = \lambda + \frac{1}{p}$, f_1 and f_2 have exactly the same exact Besov regularity.

Besides the detailed comparative study of Examples 1 and 2, we shall study some additional geometric aspects of the learning process on two other model examples: "Double chirp" and "Sinusoidal density".

Example 3 (see Fig. 2a). *"Double chirp"*

$$f_3(x) = \sqrt[4]{x}\exp(-\frac{x^2}{1-x^2})\sin[64\pi x(1-x)], x \in [0,1], \tag{48}$$

Compare also [DRP–1999, Section 7, Example 3]. The graph of f_3 is very spatially inhomogeneous, containing at the endpoints 0 and 1 two chirps of a very different nature. f_3 in (48) is a product of "λ-tear" for $\lambda = \frac{1}{2}$ and C^∞-smooth function, so its Besov regularity is exactly the same as the Besov regularity of a "λ-tear" (Example 1) for $\lambda = \frac{1}{2}$. Of special interest is to compare how the optimal learning algorithm deals in the spatially different parts of the graph for large, moderate and small samples, or for small, medium or high compression percentage.

Example 4 (see Fig. 2b). *"Sinusoidal density"*

$$f_4(x) = \begin{cases} \frac{1}{2}|\sin x| & x \in [-\frac{2\pi}{3}, \frac{\pi}{3}] \\ 0 & x \in (-\infty, -\frac{2\pi}{3}) \cup (\frac{\pi}{3} + \infty) \end{cases} \tag{49}$$

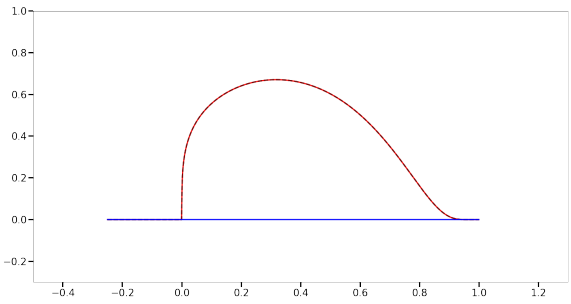
f_4 is analytic on $(-\infty, -\frac{2\pi}{3}) \cup (-\frac{2\pi}{3}, 0) \cup (0, \frac{\pi}{3}) \cup (\frac{\pi}{3}, +\infty)$; f_4 is C^0 at $x = 0$; f_4 is discontinuous at $x = -\frac{2\pi}{3}$ and $x = \frac{\pi}{3}$.

Because of the presence of the two jumps, $f_4 \in B^{\frac{1}{p}}_{p\infty}(\Omega)$, $1 \leq p \leq \infty$, $\Omega = \mathbb{R}$ or, for boundary corrected wavelets, $\Omega = [-\pi, \pi]$. This result about the Besov regularity of f_4 follows from the result about Besov regularity of the Heaviside step function which is present in implicit form in the embeddings in [DRP–1999, Appendix B12b, item (iv)]. The function exhibits considerable spatial inhomogeneity in the neighbourhoods of the three points of singularity ($x = -\frac{2\pi}{3}$, $x = 0$ and $x = \frac{\pi}{3}$). Of particular interest is to compare the performance of the optimal learning algorithm in a neighbourhood of each of the three singularities. The comparison of the performance between the two jump-singularities at $x = -\frac{2\pi}{3}$ and $x = \frac{\pi}{3}$ should include also comparative study of the local Gibbs phenomenon.

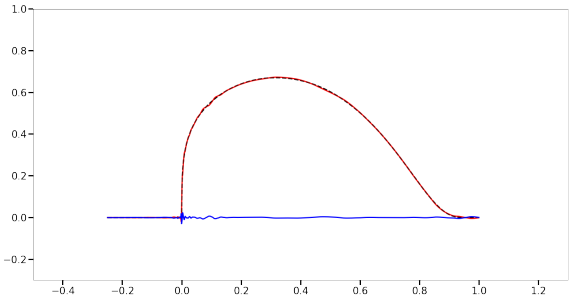
The sample sizes in Figures 1-13 are $N = 2^{10}$ or less.

First we focus on faster learning and maximal compression, according to items 1 and 2 in Section 5. The comparative graphical analysis of Figures 3–6 leads to the following conclusions.

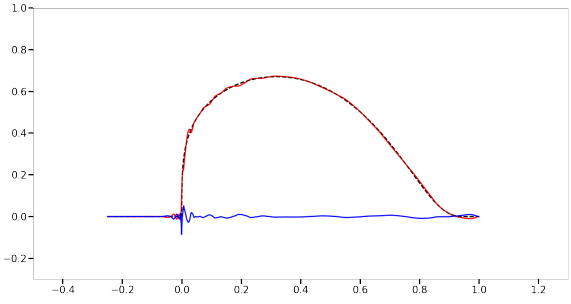
1. Examples 1, 3, 4 are of piecewise smooth type, while Example 2 is of the fractal type.
2. The decreasing rearrangement activation allows very fast learning combined with very high compression rate for the piecewise smooth curves: the quality of learning is superb at 85% compression. In comparison, retaining such high quality of learning for the fractal curve in Figure 2 is possible only at compression rate up to $3 - 4\%$. (See item (a) in each of Figures 3–6.)
3. Approximation of the target function by the learned one is very good even for superhigh levels of compression ($98 - 99\%$). This also indicates that if the large-to-medium sample size $N = 2^{10}$ be reduced to moderate or even small sizes (cf. Section 1), the rate of learning can be expected to be quite good, while the compression rates will decrease, but remain still quite good.
4. The effect of subjecting the fractal-type curve to high or superhigh compression rates is that the learned curve get smoothed out to a piecewise smooth (few isolated singularities, similar to Examples 1,3 and 4) or even smooth – no singularities at all. In [DPS+–2022a] and [DPS+–2022b] we shall show that if the fractality of the manifold is due to noise, then, learning the manifold with WBNNs where the decreasing arrangement activation is applied on the noisy (empirical) wavelet β–coefficients results in denoising and high-quality



(a) 85 % compression

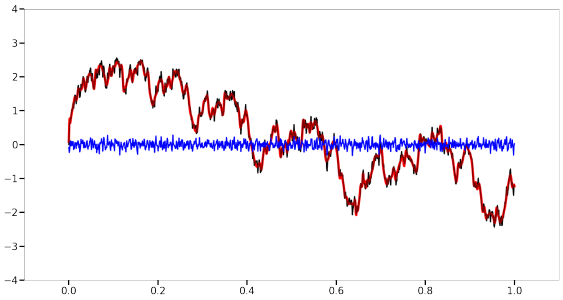


(b) 98 % compression

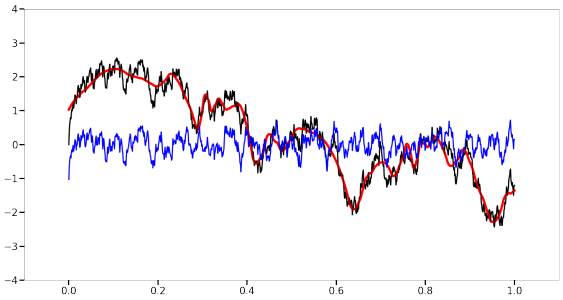


(c) 99 % compression

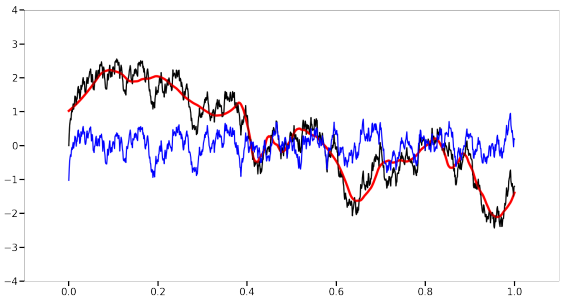
Figure 3: Target function (dashed black), learned function (red) and the error between the two (blue) for the "λ-tear" under high compression.



(a) 85 % compression

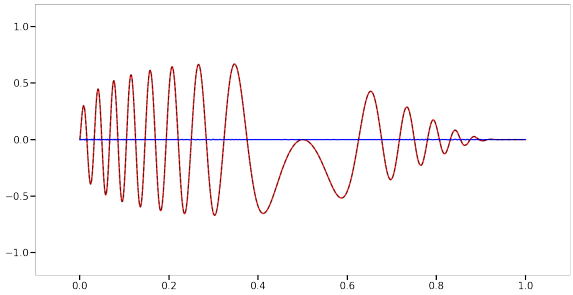


(b) 98 % compression

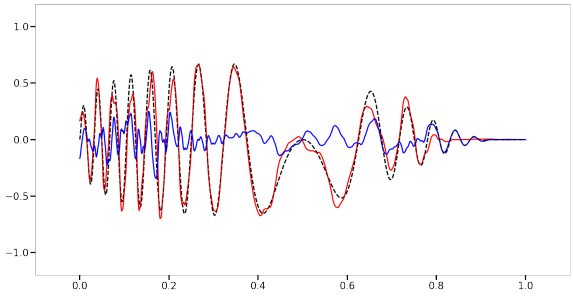


(c) 99 % compression

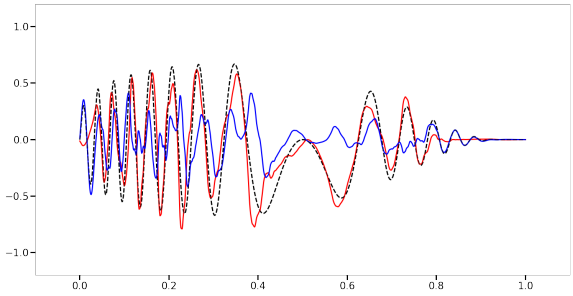
Figure 4: Target function (dashed black), learned function (red) and the error between the two (blue) for the "Weierstrass function" under high compression.



(a) 85 % compression

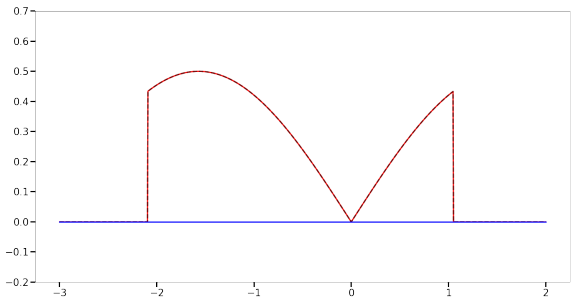


(b) 98 % compression

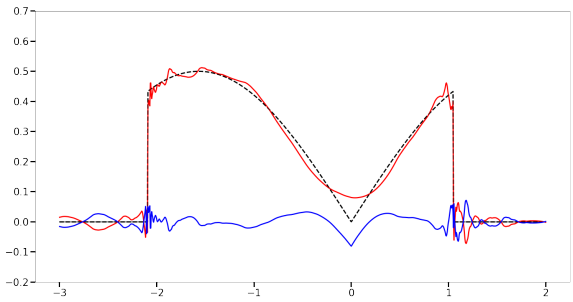


(c) 99 % compression

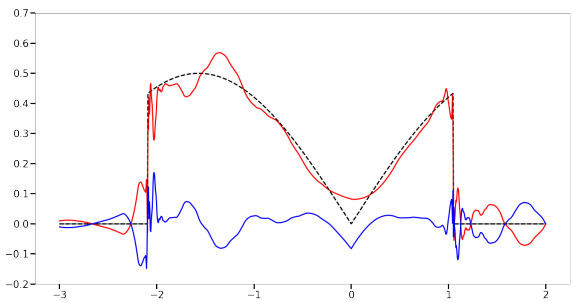
Figure 5: Target function (dashed black), learned function (red) and the error between the two (blue) for "the double chirp" under high compression.



(a) 85 % compression



(b) 98 % compression



(c) 99 % compression

Figure 6: Target function (dashed black), learned function (red) and the error between the two (blue) for "the sinusoidal density" under high compression.

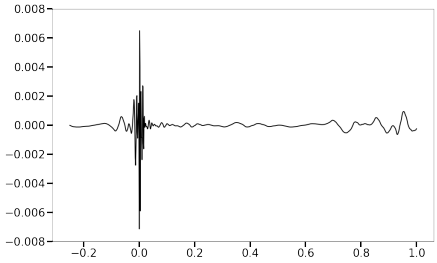
statistical estimation of the manifold. The level of compression resulting from this process can be useful in determining the chances for the manifold to have certain regularity.

The comparison of graphical data for the four examples on Figure 7 leads to the following observations.

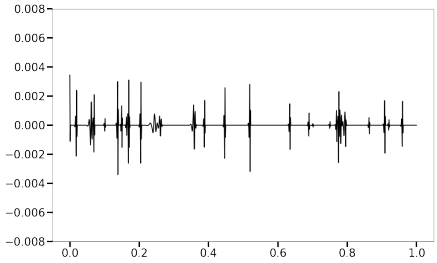
5. For piecewise smooth manifolds having isolated singularities of *the first kind* only (left and right onesided limits exist at the singularity) the distribution of the benchmark MISE error is narrowly concentrated in small neighbourhoods of the respective singularities. This also shows that the vector of β–coefficients of such manifolds tends to be sparse, i.e., it contains large in absolute value coefficients on all levels j only in small neighbourhoods of the singularities. Thus, manifolds of this type are highly compressible with threshold activation methods and are the fastest to learn with WBNNs. Typical examples are 1 and 4, see Figure 7 (a) and (d), resp.

6. Piecewise-smooth manifolds with isolated singularities of *the second kind* (at least one of the onesided limits does not exist at the singularity). Typical case of this type of singularity is the presence of a chirp on the side of the missing onesided limit (for example, functions $g(x) = x^a \sin \frac{1}{x^b}$, $x > 0, a > 0, b > 0$). Chirps can be very spatially inhomogeneous and even exhibit some fractal properties (for example, the function g in the above formula, with $a = 0$, has unbounded variation in a neighbourhood of $x = 0$ and the part of its graph in this neighbourhood is infinitely long. Thus, functions with chirps take a somewhat intermediate place between the functions in item 5 and the function in the next item 7. Typical example is 3 – see Figure 7 (c).

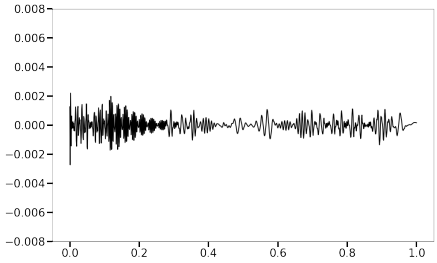
7. Fractal-type curves, especially ones with locally constant Hölder index, gather their Besov regularity from a dense vector of β–coefficients where all, or, at least, the vast majority of β–coefficients provide significant contribution which can only be ignored at the price of slowing the rate and decreasing the quality of the learning process. In [DPS+–2022a] and [DPS+–2022b] we shall show that manifolds of fractal type can be learned well only using non-threshold shrinkage activation. This type of activation produces 0% compression. Threshold activation methods, including the decreasing rearrangement activation, oversmooth the manifold, thereby altering its fractal type. This explains the low compression rate when achieving the benchmark MISE in the typical Example 2 – see Figure 7 (b).



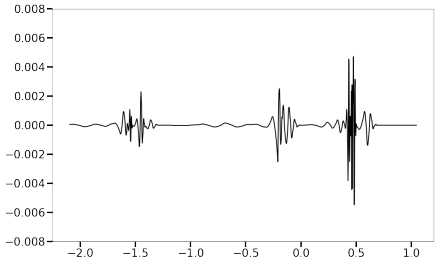
(a) Distribution of benchmark MISE for "λ-tear", at compression rate $\approx 99.610\%$. This is the benchmark MISE for all cases (a) – (d).



(b) Distribution of benchmark MISE for "Weierstrass function", attained at compression rate $\approx 3.418\%$.



(c) Distribution of benchmark MISE for "double chirp", attained at compression rate $\approx 83.984\%$.



(d) Distribution of benchmark MISE for "sinusoidal density", attained at compression rate $\approx 83.984\%$.

Figure 7: Benchmark error distribution for Examples 1-4. Benchmark error was MISE for Example 1 at compression $\approx 99.610\%$.

8. (Remark.) One of the main goals of research in [DPS+–2022a] and [DPS+–2022b] will be to design *hybrid activation strategies for adaptive learning* by composing a sequence of activation strategies of diverse – threshold and non–threshold – nature which will be achieved by the use of *deep WBNNs*, each single-layer WBNN in which will contribute with its own activation strategy.

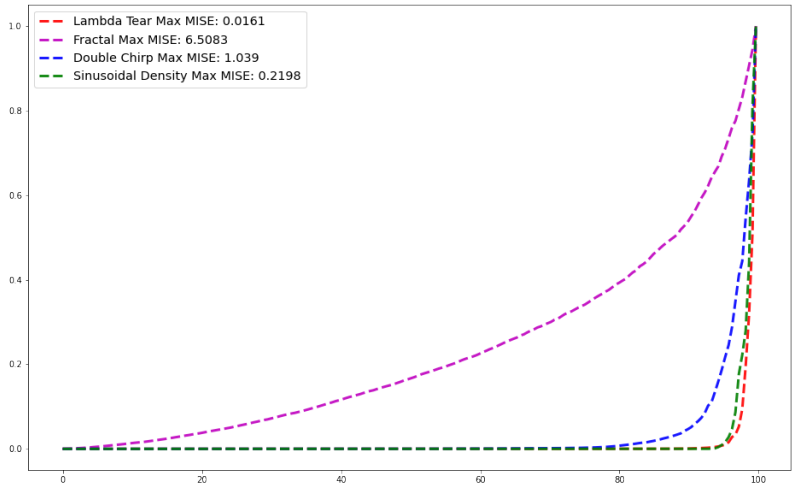
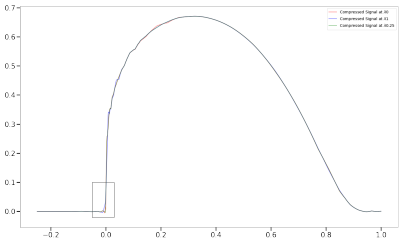


Figure 8: Compression percentage (x-axis) vs standardized relative MISE for Examples 1-4

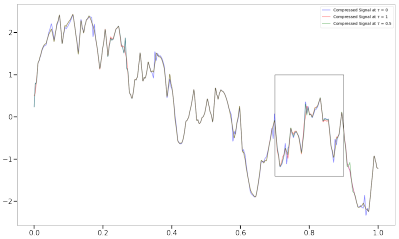
While Figure 7 provided detailed information about the local distribution of a benchmark MISE, Figure 8 provides an insightful comparison of the ratio between compression percentage and relative MISE for each of the four considered examples.

9. The aspect, in which Figure 8 is most insightful, is the comparative determination of the fractality type of the curves in each of Examples 1–4: "the Weierstrass curve" of Example 2 exhibits markedly fractal behaviour, followed by the "double chirp" of Example 3 exhibiting a somehow 'semi–fractal' behaviour, and with the curves in Example 1 and 4 being of markedly piecewise smooth type.

10. (Remark.) Tracing the behaviour of the "double chirp" of Example 3 in Figures 5, 7(c) and 8, some notable differences are observed with "the λ–tear" in Example 1, despite of the

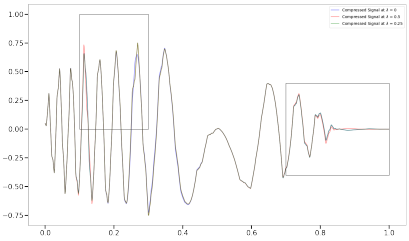


(a) Zoomed region for Example 1 (see Fig. 11a), compression rate 98%

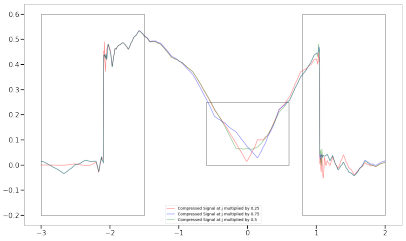


(b) Zoomed region for Example 2 (see Fig. 11b), compression rate 91%

Figure 9: Local zooming for Examples 1 and 2



(a) Zoomed regions for Example 3 (from left to right, see Fig. 12 12a) and 12b), resp.), compression rate 96%

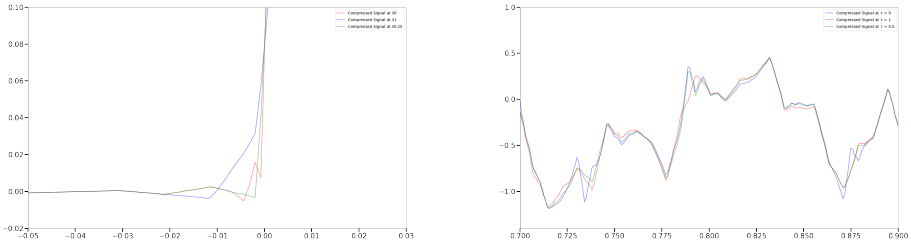


(b) Zoomed regions for Example 4 (from left to right, see (13a), 13b) and 13c), resp.), compression rate 98%

Figure 10: Local zooming for Examples 3 and 4

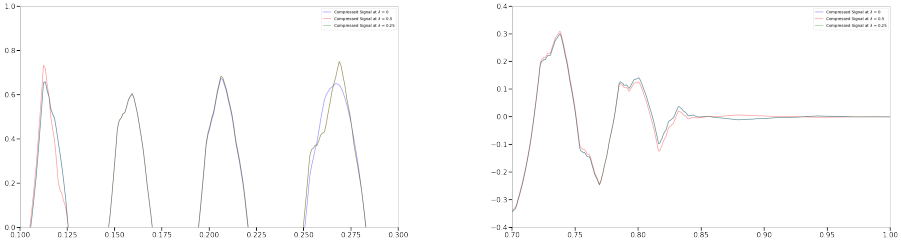
fact that they are both of the piecewise smooth type and, especially, despite of the fact that they have *exactly the same Besov regularity*. The explanation of this phenomenon is that although the Besov norms of f_1 and f_3 with the same exact parameters are both finite, the norm of f_3 is several orders of decimal magnitude larger than the norm of f_1, mainly due to the presence of the factor $64 = 2^6$ in the sine component of the formula (48) for f_3.

In Section 6 the exact Besov regularity of the example was known, and it was used in the construction of the activation operator. What if only approximate information is available about each of the parameters (p, q, s) of the Besov regularity? The algorithmically simplest way to overcome this ambiguity is to use a swarm of sufficiently broad single-layer WBNNs. As discussed in Section 4, using a deep WBNN is possible, but requires adjustments, with some loss of efficiency. This is why here the new research topic about relation between swarm and deep evolutionary AI ([I–2022]) is of great interest.



(a) (b)

Figure 11: Region zoom for (a) Example 1, (b) Example 2 (see Fig. 9a and 9b), resp.



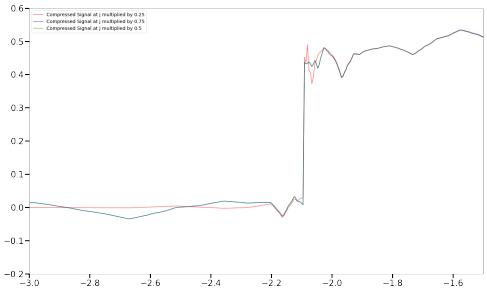
(a) (b)

Figure 12: Region zoom for Example 3: see Fig 10a

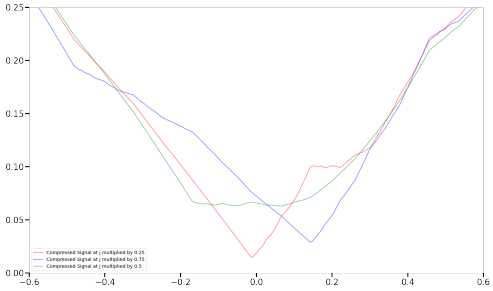
In Figures 9–13 we produce the graphical results of sequential emulation of the parallel learning process of a 'swarm' of 3 single-layered sufficiently broad (satisfying (14)) WBNNs, one of which is biased towards underestimating the Besov regularity (blue colour), the second one is using the exact Besov regularity information (green colour) and the last one is biased towards overestimating the Besov regularity. The graphical results for Examples 1–4 are presented in Figure 9(a) and 9(b), and Figure 10(a) and 10(b), resp. The rectangular regions on these figures, where the differences are most notable, are marked with window frames.

In Figures 11–13 are given zoomings of all windows in Figures 9 and 10, as follows.

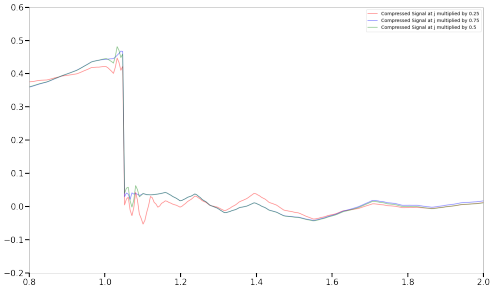
- The window in Figure 10(a): Figure 11(a)
- The window in Figure 10(b): Figure 11(b)
- The two windows in Figure 10(a): Figure 12(a) and (b)
- The three windows in Figure 10(b): Figure 13(a), (b) and (c)



(a)



(b)



(c)

Figure 13: Region zooms for Example 4: see Fig 10b

Some observations from the graphical comparison, as follows:

11. The learning process using WBNNs activated by decreasing rearrangement method proves to be *very robust with respect to errors of estimation of the Besov regularity information* when *manifolds without singularities* are being learned.
12. In the presence of singularities, the robustness in item 11 *decreases*: the more the singularities, the less the robustness, with the maximal deterioration of robustness being in the case of learning fractal-type manifolds. However, due to the relatively uniform distribution over all, or most of the β–coefficients of a fractal manifold (see item 7 above), for the low compression rate in Figure 7(b), the differences between the blue, green and red lines in Figure 9(b) would hardly be noticeable (see also item 13).
13. Due to the robustness properties in items 11 and 12 above, it makes sense to use *large-size swarms of single-layered WBNNs* only in the case of learning fractal-type manifolds, for a reason that will be explained in item 15.
14. Learning of piecewise smooth manifolds with WBNNs activated via the decreasing rearrangement method is robust with respect to small errors of underestimating or overestimating the manifold's Besov regularity, as long as compression rate is none or relatively small. This robustness *rapidly deteriorates* with the increase of the compression rate, but at the same time the quality of learning *deteriorates slowly* with the increase of the compression rate (which is an equivalent way of saying that piecewise smooth manifolds are being learnt fast). This is why in the case of learning a piecewise smooth manifold, small to medium sized swarms of sufficiently broad single-layered WBNNs are expected to be sufficient when the Besov regularity of the manifold is only approximately known, see Figure 11(a) and Figures 12 and 13.
15. In the case of fractal manifolds, the crucial difference is that, unlike the case of piecewise smooth manifolds, the quality of learning of the fractal *deteriorates rapidly together with the rapid deterioration of robustness* with respect to errors in the Besov regularity estimation, when the compression rates increases. This is why, contrary to the comparable compression rates in Figure 7(a) vs. Figure 9(a), Figure 7(c) vs. Figure 10(a) and Figure 7(d) vs. Figure 10(b), resp., there is a sharp difference in the compression rates in Figure 7(b) vs. Figure 9(b): less than 4% vs. more than 90%, resp. Because of this important difference, it can be expected that in the current context large-size swarms of sufficiently broad single-layered WBNNs would be needed to attain high quality of learning fractal manifolds. Notice that in

the context of item 4, when the fractality is due to noise, the denoised manifold may still be of fractal type, or it may be of piecewise smooth type. In the context of the present item, in this case determination of the size of the swarm requires special care (this topic will be addressed in [DPS+–2022a]).

16. Finally, let us turn our attention on the three singularities in Figure 13. Comparing the three singularities in Figure 13 with the singularity in Figure 11(a), it is observed that the latter singularity is somehow intermediate between the one in Figure 13(b) and the two in Figure 13(a) and (c). This is so, because, on the one hand, the singularities in Figure 13(b) and 11(a) are of the same type – discontinuity of the first derivative f_1', resp. f_4' and, on the other hand, $f_1'(0+) = +\infty$, which forces the graph of f_1 left of 0 to resemble that of the graph of f_4 at the left discontinuity point in Figure 13(a) and, modulo vertical axial symmetry, also the graph of f_4 in Figure 13(c). What is remarkable about the discontinuity singularities of the first kind in Figure 13(a) and (c) (and, to some less expressed extent, also about the singularity in Figure 11(a)) is the presence of *Gibbs phenomenon*. As it can be seen from Figure 3(a) and Figure 6(a), *even at high compression rates* there is *no Gibbs phenomenon at all*, which is due to the selection of *compactly supported* wavelet basis. If in this context a trigonometric basis is used there will be very significant Gibbs phenomenon even at compression rate 0%. Thus, we conclude that in the case of use of compactly supported wavelet basis, notable Gibbs phenomenon may eventually appear at jump points only at superhigh levels of compression, and it is due to uncompressed β-coefficients with low j ($j = j_0$ or j near to j_0). It is also at these superhigh levels of compression, and for the same reason, that errors in the estimation of Besov regularity can result in the decreasing rearrangement activation producing notable differences in regions with Gibbs phenomenon.

7 Proofs

Proof of Theorem 1. Let first $s : 0 < s < 1$. Using the the definitions and notations of [DP–1998, Section 2.2] for ρ, k^*, p and q, the upper bound

$$\exists\, c_1 < +\infty : R(f, \hat{f}_N) \leq c_1 N^{-\frac{s}{1+2s}} \tag{50}$$

where $c_1 = c_1(s, \mathrm{diam}(\mathrm{supp}(f)))$, follows after taking power $\frac{1}{\rho}$ from the two sides of [DP–1998, (2.2.3] where the choice $q = +\infty$ has been made. Under the assumptions of [DP–1998, Corollary

2.2.4] on f, taking in consideration the compactness of $\mathrm{supp} f$ implies the equivalence of the assumptions in [DP–1998, Corollary 2.2.4] and the current assumption $f \in B^s_{p\infty}(\mathbb{R})$

The lower bound

$$\exists\, c_0 > 0 : R(f, \hat{f}_N) \geq c_0 N^{-\frac{s}{1+2s}} \tag{51}$$

where $c_0 = c_0(s, \mathrm{diam}(\mathrm{supp} f))$, follows from [DP–1998, Theorem 2.3.2] for $q = \infty$. The selection of J in (14) assumes that in all cases considered in [DP–1998, Section 2.2], the optimal level k^* defined there, always satisfies

$$j_0 \leq k^* \leq J \tag{52}$$

which ensures the validity of all upper and lower bounds in [DP–1998, Corollaries 2.2.2-11 and Theorem 2.3.2]. Under the assumptions $q = \infty$ and compactness of $\mathrm{supp}\, f$, the norm $||.||_{p,s}$ defined in the formulation of [DP–1998, Theorem 2.3.2] can be replaced by the simpler $||.||_{B^s_{p\infty}(\mathbb{R})}$. This proves (i). The optimality of the rate $N^{-\frac{s}{1+2s}}$ in the context of the assumptions of Theorem 1 follows by the standard argument in risk estimation: with the increase of N, the bias term decreases and tends to 0 when $N \to \infty$, while the variance term increases and tends to $+\infty$ when $N \to \infty$. So, the optimal rate in N is achieved when the contributions of the bias and variance terms are equal. Under the assumptions of Theorem 1, it follows from the proof of [DP–1998, Theorem 2.2.1, under the assumptions of Corollary 2.2.4] that the rate for which the bias and variance terms are balanced is $N^{-\frac{s}{1+2s}}$. (ii) is proved.

Now let $s : 1 \leq s < 2$. In this case the proof is based on the same line of arguments, but with [DP–1998, Corollary 2.2.4] being replaced by by [DP–1998, Corollary 2.2.8] and noting that the expression $\frac{s-\frac{1}{r}+\frac{1}{q}}{1+2s-\frac{2}{r}+\frac{2}{q}}$ in [DP–1998, (2.2.4)] becomes $\frac{s}{1+2s}$ for $r = q = +\infty$. □

Proof of Theorem 2. Follows straight-forwardly from the chain of equalities in (11). □

Proof of Theorem 3. To prove the theorem for every quadruple $(p, q, s, \sigma) : 0 < p \leq \infty$, $0 \leq q \leq \infty$, $n(\frac{1}{p} - 1)_+ < s + \sigma < r$ we invoke [BL–1976, Lemma 3.10.2], as follows. Assume that $g = J^\sigma f$ and $g \in B^{s+\sigma}_{pq}(\mathbb{R}^n)$, that is, the RHS of (6) is finite when s is replaced by $s + \sigma$. Then, by [BL–1976, Lemma 3.10.2], the series (5) for g is convergent in the topology of $B^{s+\sigma}_{pq}(\mathbb{R}^n)$, therefore, $g = J^\sigma f$ is learnable by the WBNN generated by the specified wavelet basis.

In the particular case $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, the above proof can be simplified, by using [BL–1976, Lemma 2.2.1] instead of [BL–1976, Lemma 3.10.2]. □

Proof of Corollary 1. Follows from Theorem 3 by using the lifting property of J^{σ}. □

Proof of Theorem 4. The RHS of (36) is just a commuted version of its LHS. Since the basis $\{\varphi_{j_0k}, \psi_{jk}\}$ is a Riesz unconditional basis [D–1992] all that has to be shown is that the series in the LHS of (36) is absolutely convergent, because then [BL–1976, Lemma 3.10.2] (or, alternatively in the particular case of Theorem 4 when $p \geq 1$ and $q \geq 1$, [BL–1976, Lemma 2.2.1]) implies the statement of the theorem. The absolute convergence of the LHS in (36) in B^s_{pq} follows from $f \in B^s_{pq}$, implying the the finiteness of $||\cdot||_{B^s_{pq}}$ in (6). From here, the absolute convergence of the LHS of (36) in $B^{\sigma}_{\rho\eta}$ for every (ρ, η, σ) specified in the theorem follows from the Sobolev embedding $B^s_{pq} \hookrightarrow B^{\sigma}_{\rho\eta}$, see (32). □

Proof of Theorem 5. The space B^{σ}_{22} is a Hilbert space and, by Theorem 4, the RHS of (36) holds true. Because of the Hilbertian geometry of B^{σ}_{22}, removing the term involving any one $\psi_{j_{\nu_0}k_{\nu_0}}$ from $\sum\limits_{\nu=1}^{M} \beta_{j_\nu k_\nu}\psi_{j_\nu k_\nu}$ in (36) has the geometric meaning of orthogonal projection of

$\sum\limits_{\nu=1}^{M} \beta_{j_\nu k_\nu}\psi_{j_\nu k_\nu} \in \mathrm{span}\{\psi_{j_\nu k_\nu}\}_{\nu=1}^{M}$ onto

$$\left(\sum_{\nu=1}^{\nu_0-1} + \sum_{\nu=\nu_0+1}^{M}\right)\beta_{j_\nu k_\nu}\psi_{j_\nu k_\nu} \in \mathrm{span}\left\{\{\psi_{j_\nu k_\nu}\}_{\nu=1}^{\nu_0-1} \bigcup \{\psi_{j_\nu k_\nu}\}_{\nu=\nu_0+1}^{M}\right\}.$$

Since $|\beta_{j_\nu k_\nu}|$ are ordered as decreasing rearrangement with factor $2^{j_\nu(s-\frac{1}{p}+\frac{1}{2})}$ and $\sigma : \sigma - \frac{1}{2} = s - \frac{1}{p}$, (45) follows, which proves the theorem. Note the following remarkable geometric fact which remained implicit, but is crucial for the proof of the theorem: the basis $\{\varphi_{j_0k_0}, \psi_{jk}\}$ is orthonormal only in $L_2(\sigma = 0)$ but remains orthogonal in B^{σ}_{22} $(0 \leq \sigma < 2)$ which has the norm of a weighted l_2–sequence space with weight $2^{j\sigma}$. □

8 Concluding remarks

The present work is the first part of a sequence of studies dedicated to the new WBNNs. The next two parts of this series [DPS+–2022a] and [DPS+–2022b] are currently in preparation. The main focus in [DPS+–2022a] will be on a detailed study of the rich variety of threshold and non-threshold activation methods for learning curves in 2, 3 and higher dimensions. [DPS+–2022b]

will be dedicated to the diverse problems which arise when learning multivariate multidimensional geometric manifolds (surfaces, volume deformations and manifolds in dimensions higher than 3). One topic will be to reduce the dimensionality of high-dimensional WBNNs to 1- and 2-dimensional WBNNs with full preservation of their functional efficiency. One important application of this approach is to enable the use of GPGPU programming algorithms for learning parametric manifolds with arbitrary number of parameters, immersed in arbitrarily high-dimensional space [DBB+–2011], [DBG–2012]. Another topic in these two studies is to make progress in understanding the connection between learning and approximation [JSZ–2008] in the context of the new WBNNs. It should be noted that the essence of our new approach in the present paper – (a) separating the roles of wavelet depth and neural depth, (b) incorporating wavelet depth into the WBNN width to achieve consistency of learning, and (c) using the neural depth for accelerating the rate of consistent learning – can in principle be used also in the much more general context of arbitrary tree–based adaptive partitions [BCD+–2007], [BCD+–2014].

We conjecture that Theorem 5 can be generalized for a broader range than $(2, 2, \sigma)$ with $\sigma \in [0, r)$, namely, for the general assumptions on (p, q, s) and ρ, η, σ in (36) of Theorem 4. However, to investigate this conjecture, one needs to resort to a very different and much more technically involved and spacious research approach, beginning with the derivation of *direct inequalities* (Jackson-type, etc.) and *inverse inequalities* (Bernstein/Markov-type, etc.) and then, based on the derived inequalities establish a connection between appropriately selected best-approximation functionals and Peetre *K*-functionals. We refer to [PP–1987, Chapter 3] for an early, but sufficiently complete general exposition of this line of argument.

In conclusion, we note that by focusing on gradient/subgradient iterative optimization method in learning algorithms for NNs in the introduction, we left an important methodological gap which needs to be filled here [S–2015]. Numerical methods for optimal control based on Bellman's principle are very powerful in learning theory, both by swarm and deep evolutionary AI, including optimal control using feedback for supervised problems [S–2015]. Although, theoretically, Bellman's principle allows finding *global extrema* for a very general class of criterial functionals (including non-convex, non-smooth (including non-Lipschitzian) ones), and under complicated sets of constraints (including ones induced by technological standards in real-life engineering problems): computing the/a global extremum is often unfeasible due to the huge computational complexity. So, in many cases, a tradeoff is needed between affordable computational complexity and sufficiently high quality of a local extremum attained [DG–2006a], [DG–2006b]. So far, similar to gradient

methods, optimal tradeoffs in dynamical programming are also achieved via natural, rather than artificial intelligence. Nevertheless, if a dynamical programming algorithm is being applied in the context of machine learning using WBNNs, we may now have an acceptable automatic alternative.

References

[AZ–2013] A. K. Alexandridis and A. D. Zapranis. Wavelet neural networks: a practical guide. In: *Neural Networks* 42 (2013), pp. 1–27.

[BCD+–2007] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Universal algorithms for learning theory. II. Piecewise polynomial functions. In: *Constr. Approx.* 26.2 (2007), pp. 127–152.

[BCD+–2014] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Classification algorithms using adaptive partitioning. In: *Ann. Statist.* 42.6 (2014), pp. 2141–2163.

[BL–1976] J. Bergh and J. Löfström. *Interpolation spaces. An introduction*. Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York, 1976, pp. x+207.

[BTW–1975] P. Brenner, V. Thomée, and L. B. Wahlbin. *Besov Spaces and Applications to Difference Methods for Initial Value Problems*. Lecture Notes in Mathematics, No. 434. Springer-Verlag, Berlin-New York, 1975.

[C–1989] G. Cybenko. Approximation by superpositions of a sigmoidal function. In: *Math. Control Signal Systems* 2.4 (1989), pp. 303–314.

[C–1992] G. Cybenko. Correction: "Approximation by superpositions of a sigmoidal function", [Math. Control Signal Systems, 2.4 (1989), pp. 303-314]. In: *Math. Control Signal Systems* 5.4 (1992), p. 455.

[CCG–1991] S. Chen, C.F.N. Cowan, and P.M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. In: *IEEE Transactions on Neural Networks* 2.2 (1991), pp. 302–309.

[CDV–1993a] A. Cohen, I. Daubechies, and P. Vial. Multiresolution analysis, wavelets and fast algorithms on an interval. In: *C.R. Acad. Sci. Paris*. Ser. I Math. 316.5 (1993), pp. 417–421.

[CDV–1993b] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. In: *Appl. Comp. Harmonic Anal.* 1.1 (1993), pp. 54–81.

[D–1992] I. Daubechies. *Ten Lectures on Wavelets*. USA: Society for Industrial and Applied Mathematics, 1992.

[D–1997] W. Dahmen. Wavelet and multiscale methods for operator equations. In: *Acta Numerica* 6 (1997), pp. 55–228.

[D–1999] L.T. Dechevsky. Atomic decomposition of function spaces and fractional integral and differential operators. TMSF, AUBG ’99, Part A (Blagoevgrad). In: *Fract. Calc. Appl. Anal.* 2.4 (1999), pp. 367–381.

[DBB+–2011] L.T Dechevsky, J. Bratlie, B. Bang, A. Lakså, and J. Gundersen. Wavelet-based lossless one- and two-dimensional representation of multidimensional geometric data. In: *AIP Conf. Proc.* Vol. 1410. Amer. Inst. Phys., Melville, NY, 2011, pp. 83–97.

[DBG–2012] L.T Dechevsky, J. Bratlie, and J. Gundersen. Index mapping between tensor-product wavelet bases of different number of variables, and computing multivariate orthogonal discrete wavelet transforms on graphics processing units. In: *Lecture Notes in Comput. Sci.* 7116 (2012), pp. 402–410.

[DG–2006a] L.T. Dechevsky and L.M Gulliksen. A multirigid dynamical programming algorithm for discrete dynamical systems and its applications to numerical computation of global geodesics. In: *Int. J. Pure Appl. Math.* 33.2 (2006), pp. 257–286.

[DG–2006b] L.T. Dechevsky and L.M Gulliksen. Application of a multirigid dynamical programming algorithm to optimal parametrization , and a model solution of an industrial problem. In: *Int. J. Pure Appl. Math.* 33.3 (2006), pp. 381–406.

[DP–1997] L. Dechevsky and S. Penev. On shape-preserving probabilistic wavelet approximators. In: *Stochastic Analysis and Applications* 15.2 (1997), pp. 187–215.

[DP–1998] L. Dechevsky and S. Penev. On shape-preserving wavelet estimators of cumulative distribution functions and densities. In: *Stochastic Analysis and Applications* 16.3 (1998), pp. 423–462.

[DPS+–2022a] L.T Dechevsky, L.-E. Person, H. Singh, and K.M. Tangrand. *Learning non-parametric regression-functions and densities by univariate wavelet-based neural networks*. Tech. rep. 2022.

[DPS+–2022b] L.T Dechevsky, L.-E. Person, H. Singh, and K.M. Tangrand. *Learning of multidimensional geometric manifolds with wavelet-based neural networks*. Tech. rep. 2022.

[DRP–1999] L.T. Dechevsky, J.O. Ramsay, and S.I. Penev. Penalized wavelet estimation with Besov regularity constraints. In: *Math. Balkanica (N.S.)* 13.3-4 (1999), pp. 257–376.

[FJW–1991] M. Frazier, B. Jawerth, and G. Weiss. *Littlewood-Paley Theory and the Study of Function Spaces*. Vol. 79. CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence, RI, 1991.

[H–1957] F. Hoyle. *The Black Cloud*. William Heinemann Ltd., 1957.

[HL–2017] M. Hairer and C. Labbé. The reconstruction theorem in Besov spaces. In: *Journal of Functional Analysis* 273.8 (2017), pp. 2578–2618.

[I–2018] H. Iba. *Evolutionary Approach to Machine Learning and Deep Neural Networks. Neuro-evolution and Gene Regulatory Networks*. Springer, Singapore, 2018.

[I–2022] H. Iba. *Swarm Intelligence and Deep Evolution. Evolutionary Approach to Artificial Intelligence*. Taylor & Francis CRC Press, 2022.

[JSZ–2008] K. Jetten, S. Smale, and D.-X Zhou. Learning Theory and Approximation. In: *Mathematisches Forschungsintitut Oberwolfach Workshop, June 29th - July 5th. Oberfolfach Report 30/2008*. 2008, pp. 1655–1705.

[L–1964] S. Lem. *Niezwyciezony*. Wydawnictwo MON, 1964.

[L–1992] F. T. Leighton. *Introduction to Parallel Algorithms and Architectures. Arrays, Trees, Hypercubes*. Morgan Kaufman Publishers, Inc., San Mateo, CA, 1992.

[L–2020] S. Lem. *The Invincible*. The MIT Press, 2020.

[LBH–2015] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. In: *Nature* 521.7553 (2015), pp. 436–444.

[LPW+–2017] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang. The expressive power of neural networks: A view from the width. In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[OUN+–2017] R. Okuta, Y. Unno, D. Nishino, S. Hido, and C. Loomis. CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations. In: *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*. 2017. URL: `http://learningsys.org/nips17/assets/papers/paper_16.pdf`.

[PP–1987] P.P. Petrushev and V.A. Popov. *Rational Approximation of Real Functions*. Encyclopedia of Mathematics and its Applications, No. 28. Cambridge University Press, Cambridge, 1987.

[RS–1980] M. Reed and B. Simon. *Methods of Modern Mathematical Physics, Vol. 1: Functional Analysis. 2nd edition.* 2nd ed. Academic Press [Harcourt Brace Jovanovich, Publishers], New York, 1980.

[S–2015] Jürgen Schmidhuber. Deep learning in neural networks: An overview. In: *Neural Networks* 61 (2015), pp. 85–117.

[SKM–1993] S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers, Yverdon, 1993.

[T–1983] H. Triebel. *Theory of Function Spaces*. Vol. 78. Monographs in Mathematics. Birkhäuser Verlag, Basel, 1983.

[ZB–1992] Q. Zhang and A. Benveniste. Wavelet networks. In: *IEEE Trans Neural Networks* 3.6 (1992), pp. 889–898.

Paper D

Analysis of Civil Engineering Infrastructure in Norway With Solutions Based on Structural Health Monitoring and Artificial Intelligence

Kristoffer Tangrand, Harpal Singh

Technical report, UiT The Arctic University of Norway, 2022.

D

Analysis of Civil Engineering Infrastructure in Norway with Solutions Based on Structural Health Monitoring and Artificial Intelligence

Kristoffer Tangrand, Harpal Singh

Department of Computer Science and Computational Engineering,
UIT - The Arctic University of Norway, Campus Narvik.

* *Corresponding Author*. E-mail: kristoffer.tangrand@uit.no

Abstract

Ageing civil infrastructures such as bridges and building causes many consequences from practical and economical point of view. Especially in northern Norway, the impact of extreme arctic conditions is intense on civil engineering infrastructures (see [39]). With the increased loading due to the prosperous seafood industry and increased cargo activity is putting additional pressure on the aged infrastructures. Research and development of new methods is needed for the damage detection in these structures. In this paper we present, discuss and analyze the situation concerning bridges in Norway with a special focus on northern Norway. Moreover, based on the research in [40], [41] and [42] we describe and emphasise the importance of structural health monitoring methods, artificial intelligence and machine learning when trying to solve these serious problems of structural damage detection especially in arctic regions.

Keywords: Structural health monitoring, Vibration analysis, Operational modal analysis, Damage detection, Bridges, Arctic conditions, Finite element model, Signal processing, Wavelet transform, Artificial intelligence, Machine learning, Neural network, Wavelet network and Statistical methods.

AMS Classification (2010): 35A22, 45A05, 44A15, 65T50, 65T60

1 Introduction

This paper is based on the recent Ph.D. thesis [40], especially the papers [39], [41] and [42], where we highlighted the importance of using artificial intelligence and machine learning for damage detection in structures such as bridges and high-rise buildings. In this paper, we investigate the scale of ageing bridge

infrastructure in Norway with focus on the special problems appearing in arctic regions.

Signal processing plays a very important role in extracting the features from the data generated by the sensors to find damages. Therefore, signal processing is like the heart of structural damage detection methods. With the continuous advancement in technology new techniques are under development for structural health monitoring (SHM)

Some new artificial intelligence (AI) and machine learning (ML) algorithms that are of importance for structural health monitoring (SHM), operation modal analysis (OMA) and finite element (FE) model updating are discussed in this paper. In fact, AI algorithms/techniques such as Deep Learning, Long Short-Term Memory and Ant Colony Optimization were briefly discussed in [9], [45], [46] and [47] for various smart city applications involving time series analysis and flow distribution. These algorithms can be crucial for further development of smart SHM solutions in the future.

This paper is organised as follows: In Section 2 we present and briefly discuss the huge problems caused by the ageing infrastructures in Norway with a special focus on the situation of all bridges in northern Norway. Section 3 covers the topics of SHM, OMA, FE model updating and damage detection. In Section 4 an overview of the current state-of-the-art of SHM with AI are presented. Machine learning, neural network and recurrent neural network are presented with a focus on structural damage detection. Finally, in Section 5 we present some concluding remarks including some suggestions of future research in this important area.

2 Aged civil engineering infrastructure

Damages in structures occur during its operational lifetime due to various environmental or human factors. Lack of maintenance and monitoring can lead to accumulation of damages with time that can significantly decrease the performance of the structures, change in natural symmetry or even destruction. In general, civil engineering structures are designed with a lifetime of 50 to 100 years. In this lifetime, structures are assumed to meet the expected structural integrity. But in general, the structures are prone to unpredictable and unexpected damages arising due to various factors in the lifetime of a structure.

Ageing of civil engineering infrastructures such as bridges, tunnels and buildings cause many problems with great consequences, both from practical and economical points of view. Governments and municipalities around the world have to spend more time and budget for maintenance, repairs or construction of new structures in place of deteriorated or damaged ones, so the citizens can have a decent service.

Infrastructure maintenance costs for the governments around the world are on the rise, as a lot of infrastructures around the globe are approaching towards the end of its life cycle. Moreover, due to scarcity of expert work force to analyze such challenges, it is adding up to the problem. Analysis of such problems is

important e.g. in northern Scandinavia, since such problems are even more serious due to the fact that the impact of extreme arctic conditions is quite intense (see e.g. [40]).

A Norwegian newspaper Verdens Gang (VG) got access to a report published by Statens Vegvesen (The Norwegian Public Roads Administration) in 2017. According to this report there are approximately 16,791 bridges in Norway and Statens Vegvesen have been violating inspection rules for many of them. It was discovered that for one of every two bridges, proper inspection is lacking. Moreover, approximately 1087 bridges in Norway have damages that are described as serious or critical according the internal classification system of Statens Vegvesen (see [48]). In April 2022 it was revealed by the government-

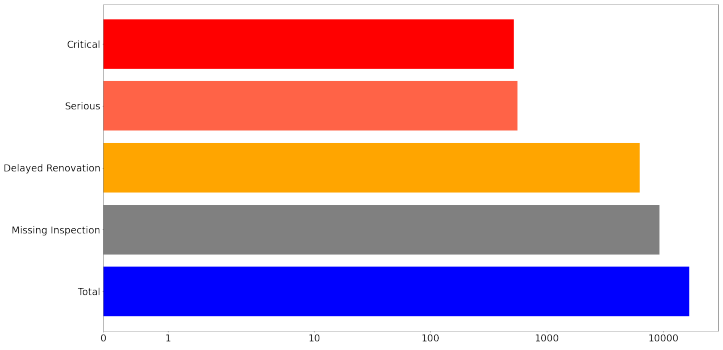


Figure 1: Histogram of state of Norwegian bridges in logarithmic scale.

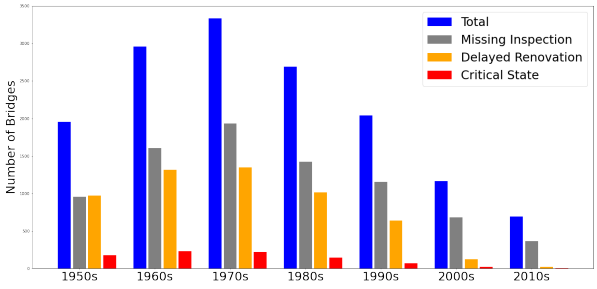


Figure 2: Histogram of Norwegian bridges, grouped by build decade.

owned broadcasting company NRK that around 1000 bridges in Norway were still not up to standard (see [31]).

The newspaper VG also created a map of bridges in Norway, using data acquired from the internal database Brutus used by Statens Vegvesen, which uses classifications such as seriously injured, delayed renovation and lacking inspection (see [49]). From the analysis of this data that is available on the VG website, an analysis of all the bridges in Norway is done and presented in logarithmic scale in Figure 1. Classification is made with respect to bridges that are missing inspection, delayed maintenance action, serious and critical that need action. Moreover, in Figure 2 a histogram of Norwegian bridges classified over different decades, is presented.

In northern Norway, large amounts of seafood cargo is exported along the public roads. The seafood industry of Norway, as of 2021, exported for 12 billion euros and contributed to around 10 percents of Norwegian export earnings. The seafood industry has seen 7 percent year on year growth since the year 2000, essentially doubling every ten years (see [13] and [29]). The seafood industry is expected to keep growing at the same rate, and has already in the first part of 2022 seen record growths of 20 percent year on year (see [34]). Especially in the sparsely populated northern Norway this is expected to put ever increasing loads on already struggling infrastructure. Thus the ageing infrastructure has to be tested and maintained with respect to the increased loading. A down time or failure of any such infrastructure can lead to substantial economic losses and even human lives.

A detailed study is conducted in this paper that maps the clusters of bridges in Nordland, Troms and Finnmark county where the large parts of the seafood industry is concentrated. A graph of all the bridges in northern Norway is presented in Figure 3 and Figure 4. As we know from geographical constraints,

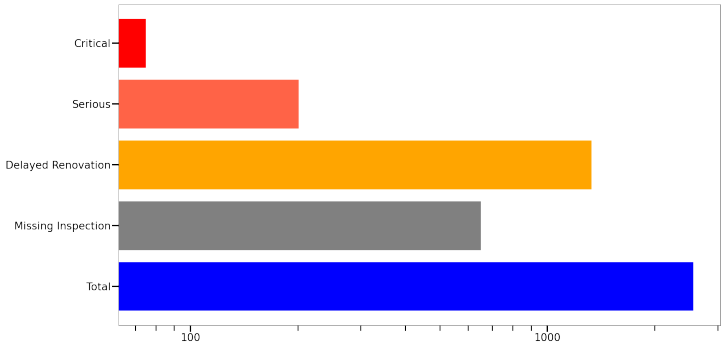


Figure 3: The current state of bridges in the two northernmost Norwegian counties – Nordland, as well as Troms and Finnmark.

Norway is very long, so in order to have better visualization the area of focus that is from Nordland to Finnmark is split into 3 regions. These figures illustrate the extent of problem in northern Norway. Figure 5 is from latitude 57 degree (border of Nordland and Trøndelag) till latitude 66 degree passing Bodø. Figure 6 covers the geographical area from from latitude 66 degree from Bodø till latitude 68 degree. Finally, Figure 7 cover Finnmark from latitude 68 degree Tromsø till 69 degree that is Nordkapp. The hexagon blocks indicate the density of bridges that lack inspection while red circles represents the bridges in critical or serious state.

With this in focus Statens Vegvesen has put a higher priority to investigate and do maintenance of bridges that are critical or are seriously damaged. In the year 2019, a major damage was found in the construction of the Herøysund bridge located on the west-coast in Nordland county in Norway (see Figure 8). As a result concerned authorities decided that the special transport was no longer allowed to drive over the bridge (see [3]).

Later in 2020, Nordland county and the Norwegian public road administration decided to work on building a new bridge that would be located just south of the current bridge. The new Herøysund bridge is expected to cost about 270 million NOK and is expected to be finished in the summer of 2023. Moreover, it was decided that the maintenance and reinforcements will be carried out on the Herøysund bridge so it is safe to use until the new Herøysund bridge opens.

Remark 2.1 A detailed study of the Herøysund bridge will be presented in our forthcoming article. This is possible because we have the concrete data for this case. By doing this we can do a similar analysis for all other bridges in the region of northern Norway. For a more detailed description see Remark 5.1.

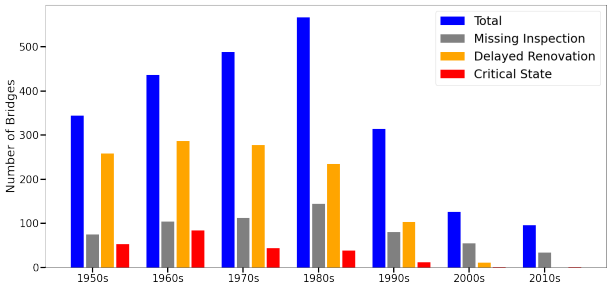


Figure 4: The same data as in Figure 3, ordered by the decade the bridges were built during.

The remaining part of this paper is inspired by the recent research presented in the PhD thesis [40] where some methods presented can be used to overcome these serious problems in northern Norway. In the next section we present structural health monitoring (SHM) along with the importance of artificial intelligence in SHM.

3 Structural health monitoring

Operators/owners of civil engineering infrastructure such as bridges, dams and tunnels are mostly municipalities or government owned enterprises in Norway. As for now, infrastructure assets management decisions are based on visual inspections, which could be aided by localized diagnosis techniques such as the

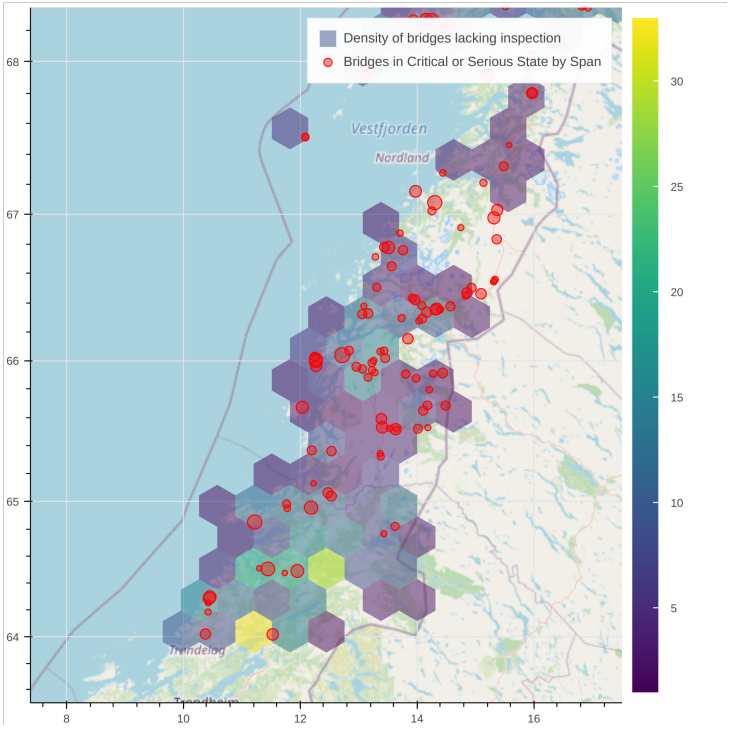


Figure 5: The bridges in southern and middle part of Nordland county.

use of acoustic, ultrasonic or magnetic field non-destructive testing methodologies. Nevertheless, these testing methodologies have several limitations such as, inaccessibility to some parts of the structure, inability to detect internal damage, localization of the damage, and it is challenging to carry out continuous monitoring with such techniques.

With the advancement in technology, new techniques are under continuous development for the monitoring of structures. These techniques are commonly called structural health monitoring (SHM) techniques. SHM refers to the process of systematizing, implementing and characterizing a damage detection strategy in civil, mechanical and aerospace engineering structures (see [12]). The process involves the observation of structure over the course of time with periodically spaced static and dynamic response measurements, extraction of

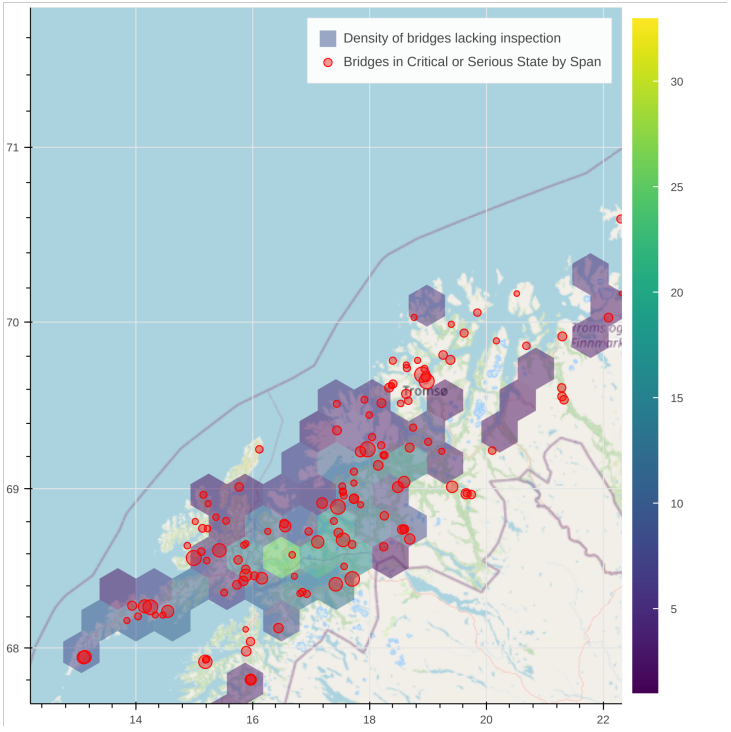


Figure 6: The bridges in the northern Nordland county and southern Troms and Finnmark county.

damage sensitive features from measurements and finally statistical analysis of these features to estimate the current state or health of the structure.

In a typical SHM system sensors are distributed throughout the structure, that are used to estimate the condition of the structure. A damage is defined as an intentional or unintentional change to the material or geometric properties of the structures, including the changes in the boundary conditions or system connectivity which adversely affect current or future performance of the structures (see [12]).

In order to do damage detection and localization, the raw data generated by sensors is processed for extraction of damage sensitive features. For example in a vibration based SHM system, accelerometers are used to find the key parameters: mode shapes, mode frequencies and mode damping. Once these parameters have been estimated, damage detection algorithms can be utilized

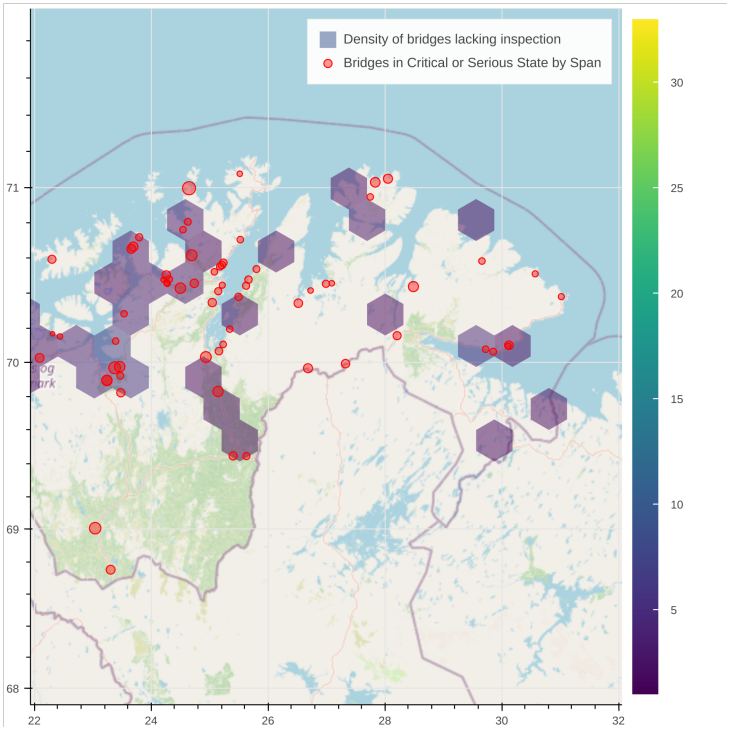


Figure 7: The bridges in the northern Troms and Finnmark county.

to figure out the magnitude of damage occurred, if any (see [17] and [21]).

The ordinary differential equation that represents a linear dynamical model of a vibrating system is as below (see [33] and [42]):

$$M\frac{d^2u(t)}{dt^2} + C_2\frac{du}{dt} + Ku(t) = B_2f(t). \tag{1}$$

The different terms in the equation are described as: M is the mass matrix, C_2 is the damping matrix, K is the stiffness matrix, B_2 is the selection matrix (input matrix), $f(t)$ is a vector with nodal forces and the solution $u(t)$ of this differential equation is the vector with nodal displacements (see [33]). A mathematical model to compute the modal parameters is described in detail in [33] and [42].

In general traditional forced vibration tests with artificial excitation forces can be performed on large structures, but such tests are costly and complicated. Moreover, other vibration sources such as wind and traffic are treated as noise.

Finite element (FE) model updating is one of the most popular methods nowadays to improvise the numerical models for various civil engineering structures such as bridges, high-rise buildings, and mechanical structures such as steel bridges, wind mills, off-shore structures, etc. Moreover, the FE model is updated and improved by updating the numerical response with respect to the

Figure 8: Herøysund bridge.

observed experimental behaviour of the structure (see [18]). FE updated model is also known as a digital twin. It is very crucial to have correct information about the structure. In this context OMA and FE updated model both play very important roles.

Operational modal analysis (OMA) is gaining popularity where ambient vibrations from the wind and traffic are considered as unknown input, and output-only analysis is done to determine the resulting vibration modes. OMA technique has been tested on a steel truss structure bridge in Luleå (see [41]), a high-rise fire tower building in Luleå (see [42]) and various other structures around the world, see also the PhD thesis [40]. OMA are Multiple Input and Multiple Output (MIMO) techniques, so these techniques can estimate the closely shape modes and repeated modes for a high degree of accuracy. Single Input Multiple Output (SIMO), Multiple Input Single Output (MISO) and Single Input and Single Output (SISO) are traditional testing procedures that are not able to find repeated poles due to lack of mode separation.

Remark 3.1 In the new upcoming project the plan is to test this OMA technology on the old Herøysund bridge that is going to be demolished in 2024, and the new Herøysund bridge that is going to become operational in 2023 in the county of Nordland in Norway.

Statistical pattern recognition paradigm for SHM is a model where a comparison between two different states of the structure, one being initial/normal or undamaged state and the other being the damaged or a state with defects is made. For example in case of a bridge, a label of critical damage, need of inspection or need of maintenance can be assigned by comparing the bridge to a database of healthy bridge. This database can be accumulated over the time and be used for training a mathematical framework of machine learning algorithms. This will further be discussed and described in the forthcoming paper.

The sensors used for SHM generate lots of data, thus signal processing techniques makes the heart of SHM. Various signal processing techniques that are of great importance for SHM and OMA have been discussed and compared in our previous article (see [42]). Wavelet analysis is an effective mathematical and signal processing tool that is based on time frequency analysis and overcomes some of the limitations of conventional Fourier analysis based methods. In our forthcoming article we will focus our discussion on wavelets in the context of damage detection and artificial intelligence. However, for the readers convenience already here in Appendix A we give some historical remarks and newest development of wavelets from the first Haar wavelets till the remarkable development described in more the 50 books.

Remark 3.2 SHM systems have very many sensors installed, so the challenge of synchronization of data due to various sampling rates appears naturally. Further, the issue of missing data can appear due to various factors such as sensitivity of sensors or other environmental factors such as low wind or failure to record data. In a SHM system problem of data synchronization and missing

data appears naturally due to various types of sensors that are involved (see [18]). Recent work in machine learning aims to alleviate such issues, which will be addressed in a forthcoming article.

4 Artificial Intelligence and Machine Learning

While the terms are often used interchangeably, Artificial Intelligence (AI) and Machine Learning(ML) have different meanings, and an important relation between each other. The term AI refers to any algorithm that can be used to make a machine (or computer) perform a task. This range of tasks is enormous, and encompasses anything from simple path-finding algorithms to advanced autonomous drones. As such, the term ML is included under the AI umbrella. Where ML differs from other types of AI can be suggested from the name. ML algorithms are able to learn from data. This data can be collected from many sources, including, but not limited to, real-world sensors or images (see [26]), simulated worlds (see [38]) or data created by humans such as text (see [6]). Typically, machine learning is divided into the three paradigms supervised, unsupervised and reinforcement learning. For the purposes of this paper, we present supervised and unsupervised learning.

4.1 Supervised learning

Supervised learning is a machine learning paradigm which aims to learn a function that maps a set of input data points to a set of target data points. This function should generalize well and also should be able to make good predictions for unseen data points. The problem of supervised learning is often solved by finding the closest points in the input space to the target points using a distance function, i.e., by finding the nearest neighbor of each data point. The intuition is that the predicted target point will be the nearest neighbor of the data point, which is closest to the data point.

4.2 Unsupervised learning

Contrary to supervised learning, unsupervised learning algorithms do not require labeled data. The goal of unsupervised learning is to learn the structure in the data, such as grouping some examples together, or finding similar examples. For instance, an unsupervised clustering algorithm, such as K-means, can automatically partition a dataset into different groups. It is important to note that the quality of the results produced by unsupervised learning algorithms are typically lower than those of supervised learning algorithms.

Next, we introduce neural networks, which is a very important and common machine learning model. Countless variations and improvements exist, but we explain the basic version which is the foundation for more advanced models.

4.3 Neural networks

Neural networks are models of computation inspired by the structure and function of biological nervous systems, including the brain. They are a paradigm of machine learning and are used in the development of artificial intelligence. A common choice for the structure of a neural network is the layered feed-forward network, in which a set of input nodes receives data from a set of previous nodes, which receive data from another set of nodes, and so on; the last layer is called the output layer. In the layered feed-forward network, information flows only in one direction, which is called the "feed-forward" direction. Information can spread out to the output layers by activating all the nodes in the layers between the input and output layers. Each of the nodes in the network has a strength and all the nodes are connected with each other. The networks are trained by calculating the error of the network, which is the difference between the desired output and the output of the network, using back propagation (see [35]).

Consider a network N with w connections, x inputs and y outputs. NNs are function approximators, and as such can be expressed as a function $y = f_N(w, x)$. The weights w maps the inputs x to the outputs y. The weights w are usually provided from a random distribution, while the inputs x are the data the network is trying to learn from, such as sensor data, images or signals. The outputs y is then the variable or classification the network is optimizing towards.

Given an input sample $p_j(t)$, for each neuron j in the network, its contribution to to the outputs $o_i(t)$ can be described as:

$$p_j(t) = \sum_i o_i(t) w_{ij}. \tag{2}$$

where the elements w_{ij} in the matrix $[w_{ij}]$ represents the intermediate product between each layer.

Next, we describe some variations and improvements on neural networks that are of importance for upcoming section describing AI in the SHM field.

4.3.1 Recurrent neural networks

Recurrent Neural Networks (RNNs) are a variation of neural networks that include a sequential – or looping property. As such, they can be used to predict the next element of a sequence. They are especially useful for modelling sequences. The most common and famous RNN variation is called the Long Short-Term Memory (LSTM) (see [20]), which was designed to address the vanishing gradient problem found in earlier versions of the RNN. The vanishing gradient problem appears in cases where the weights of the network become so small that they effectively will not change. As such, the network will not train. This problem is improved by improving learnable gating mechanisms in the network, which allows better control over the information flow.

4.3.2 Residual Networks

Inspired by the architecture of the LSTM, one of the motivating factors for the Residual Network (ResNet) was to avoid the vanishing gradient problem. However, it is not the preservation of sequentiality that is the main goal, but rather to make it easier to train and optimize very deep neural networks. This is achieved by utilizing skip connections – basically meaning that the neurons do not only have to communicate between directly neighbouring layers, but can also communicate between distant layers. Like with the LSTM, there are also gating mechanisms to control information flow (see [19] and [44]).

4.3.3 Generative Adversarial Networks

The Generative Adversarial Network((GAN) is in fact not a network architecture, but rather a protocol that two neural networks use to generate new data. The training data is then used as a statistical baseline for the generation of new data. Indeed, there are two networks, the *generator* (generative network) and the *discriminator* (discriminative network), both which having different, competing goals. The generative network creates data samples (by guessing from a sample distribution) that is then evaluated by the discriminative network (which knows the full distribution). As such, the generator will incrementally get better at emulating the true data distribution. This will lead to the generated samples becoming more and more like the true distribution. This technique can be especially effective for generating synthetic data (see [15]).

4.4 AI techniques for SHM and vibration analysis

With the recent technological advances in computer vision, artificial intelligence (AI), and machine learning (ML), we are now witnessing a new era of computer-based automated systems in the damage detection of facilities, infrastructure, and vehicles. In this subsection, we review the use of AI and ML for automated detection of damage in the SHM and OMA spaces.

Combinations of wavelets and ML approaches such as NNs have been explored in the SHM space, with the most basic approach being to first transform the vibration signal with DWT, and then using the NN to train on the transformed signal (see [37]).

A recent study presents a novelty-classification framework applicable to SHM problems. LSTMs are utilized to perform the classification. Then, a GAN and its generated data objects are used to improve the low-sampled data class classification (see [43]). Similarly, various deep learning approaches are explored for automated Structural Damage Detection (SDD) during extreme events. Among the approaches are ResNet for classification. ResNet is also combined with a segmentation network for categorizing and locating structural damage (see [4]). The applicability of Transfer Learning (pre-trained image models) to SHM problems shows both promise and concern (see [7]). Unmanned Aerial Vehicles combined with computer vision and deep learning has been shown to be a fast,

cheap and effective means of SHM for civil infrastructures (see [32]). Data sets from the construction industry is also used to benchmark machine learning architectures, such as in the paper [25] where a large amount (56,000) of images of cracks in concrete were used for training a novel algorithm for crack detection. A Deep Belief Network (a NN that only has connections between layers, but not between neurons) was used, and the configuration of neurons and layers in the network was self-organized using Adaptive Restricted Boltzmann Machine.

Clearly, recent years has seen much work in the SHM space with regards to techniques involving AI and computer vision. However, less work can be found involving AI and vibration analysis from measured sensor data. The work in this field has been more concerned with traditional statistical and mathematical models.

However, some work combining ML with vibration analysis in the SHM space has appeared, recently. An ensemble deep learning technique that combines a Convolutional NN with Dempster-Shafer theory (DST) is proposed, and called CNN-DST. The framework shows robust performance compared to other state-of-the-art classification methods (see [50]). GANs have also been used for synthetic data generation in the context of vibration analysis for SHM (see [27]).

5 Concluding Remarks

Remark 5.1 In the new upcoming project the plan is to test this OMA technology on the old Herøysund bridge that is going to be demolished in 2024, and the new Herøysund bridge that is going to become operational in 2023 in Nordland county in Norway. The main aim is that this can essentially help us for the better understanding of bridges with similar issues and take precautionary steps before the damage in bridges can become serious. In this forthcoming article we describe more details concerning this important motivation.

Remark 5.2 In the recent PhD thesis [40] some new statistical and mathematical results were stated and proved, which hopefully can be useful in the required improvements of the traditional methods in this area of structural health monitoring and artificial intelligence. So far a lot of development has been done in the bounded systems for Fourier analysis and inequalities. Further development of new Fourier analysis techniques (see [5]) and inequalities also in unbounded orthogonal systems (see [1] and [2]) and signal processing problems in non-separable function spaces (see [36]) can provide or help in the improvement of the signal processing techniques used for the damage detection in suspension bridges and related structures.

Remark 5.3 In our new paper we aim to analyze the bridge by using the methods above and also the new theoretical findings in the Ph.D. Thesis [40]. It is especially important to note that the wavelet system (see Appendix A) is unbounded and the traditional theory of Fourier inequalities do not cover this case. In the recent Ph.D. thesis also some new statistical methods were stated and

applied. In particular, non-separable function spaces were used (see [36]), which can further help to tackle similar problems for bridges, especially in northern Norway.

Remark 5.4 The literature search reveals that the intersection of AI/ML and SHM has a long history and many important studies has been conducted in this field. However, the use of AI/ML combined with vibration analysis applied to SHM still needs further research.

Acknowledgements

- We thank Professor Lars-Erik Persson for several generous suggestions that have improved the final version of this paper.
- A special thank to Roy Eivind Antonsen (Project Manager, Statens Vegvesen, Nord) for several fruitful discussions and good suggestions.
- We are also grateful to Espen Dahl-Mortensen (Construction Manager, Norland Fylkeskommune) and Per Ove Ravatsås for providing the picture of the Herøysund Bridge that is used in this paper.
- We also thank the careful referee for comments and remarks that improved the quality of the paper.

References

[1] Gabdolla Akishev, Lars-Erik Persson, and Harpal Singh. Inequalities for the Fourier coefficients in unbounded orthogonal systems in generalized Lorentz spaces. *Nonlinear Studies*, 27(4):1137–1155, 2020.

[2] Gabdolla Akishev, Lars-Erik Persson, and Harpal Singh. Some new Fourier inequalities and Jackson–Nikol´skii type inequalities in unbounded orthonormal systems. *Constructive Mathematical Analysis*, 4(3):291–304, 2021.

[3] Bjørn Olav Amundsen. Denne brua synger på absolutt siste verset: Nå må hus rives for å gi plass til ny(in Norwegian). https://www.veier24.no/artikler/denne-brua-synger-pa-absolutt-siste-verset-na-ma-hus-rives-for-a-gi-plass-til-ny/516133 (accessed: 16.06.2022).

[4] Yongsheng Bai, Bing Zha, Halil Sezen, and Alper Yilmaz. Engineering deep learning methods on automatic detection of damage in infrastructure due to extreme events. *Structural Health Monitoring*. 15 pages, 2022.

[5] David Baramidze, Lars-Erik Persson, Harpal Singh, and George Tephnadze. Some new results and inequalities for subsequences of Nörlund

logarithmic means of Walsh–Fourier series. *Journal of Inequalities and Applications*, 2022(30). 13 pages, 2022.

[6] Tom Brown et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020.

[7] Zaharah Allah Bukhsh, N. Jansen, and Aaqib Saeed. Damage detection using in-domain and cross-domain transfer learning. *Neural Computing and Applications*, 33:16921–16936, 2021.

[8] C. S. Burrus, R. A. Gopinath, , and H. Guo. *Introduction to Wavelets and Wavelet Transforms: A Primer*. Upper Saddle River, N.J: Prentice Hall, 1998.

[9] Shayan Dadman, Bernt Bremdal, and Kristoffer Tangrand. The role of electric snowmobiles and rooftop energy production in the arctic: The case of longyearbyen. *Journal of Clean Energy Technologies*, 9(4):46–53, 2021.

[10] L. T. Dechevsky and P. Zanaty. Triangular beta-function b-spline finite elements: Evaluation and graphical comparisons. In *Large-Scale Scientific Computing*. Springer Berlin Heidelberg, 2012.

[11] L.T. Dechevsky, N. Grip, and L.E. Persson. Sharp error estimates for approximation by wavelet frames in Lebesgue spaces. *Journal of Analysis and Applications*, 1(1):11–31, 2003.

[12] Charles R Farrar and Keith Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, 2012.

[13] Fisk.no. Sjømateksporten passerte 120 milliarder kroner i 2021. https://fisk.no/fiskeri/7553-sjomateksporten-passerte-120-milliarder-kroner-i-2021, (accessed: 06.06.2022).

[14] D. L. Fugal. *Conceptual Wavelets in Digital Signal Processing: An In-depth, Practical Approach for the Non-mathematician*. Space & Signals Technical Publications, 2009.

[15] Ian Goodfellow et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27. 9 pages, 2014.

[16] A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2:50–61, 1995.

[17] Niklas Grip, Natalia Sabourova, and Yongming Tu. Sensitivity-based model updating for structural damage identification using total variation regularization. *Mechanical Systems and Signal Processing*, 84:365–383, 2017.

[18] Niklas Grip, Natalia Sabourova, Yongming Tu, and Lennart Elfgren. Vibrationsanalys för tillståndsbedömning av byggkonstruktioner: Tillämpningsexempel:(main results and summary in swedish. detailed results in english appendices.). Technical report. Luleå University of Technology , 2017.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[20] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[21] Zheng Huang, Yong-Ming Tu, Niklas Grip, Natalia Sabourova, Niklas Bagge, Thomas Blanksvärd, Ulf Ohlsson, and Lennart Elfgren. Modelling of damage and its use in assessment of a pre-stressed concrete bridge. In *19th IABSE Congress, Stockholm 2016, September 21-23*, pages 2093–2108, 2016.

[22] B. B. Hubbard. *The World According to Wavelets: The Story of a Mathematical Technique in the Making*. A. K. Peters, Ltd., USA, 1996.

[23] S. Jaffard, Y. Meyer, and R.D Ryan. *Wavelets: Tools for Science and Technology*. Society for Industrial and Applied Mathematics, 1987.

[24] G. Kaiser. *A Friendly Guide to Wavelets*. Birkhäuser Boston Inc., USA, 2011.

[25] Shin Kamada, Takumi Ichimura, and Takashi Iwasaki. An adaptive structural learning of deep belief network for image-based crack detection in concrete structures using sdnet2018. In *2020 International Conference on Image Processing and Robotics (ICIP)*. 6 pages, 2020.

[26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[27] Furkan Luleci, F. Necati Catbas, and Onur Avci. Generative adversarial networks for data generation in structural health monitoring. *Frontiers in Built Environment*. 17 pages, 2022.

[28] S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, Inc., USA, 3rd edition, 2008.

[29] Menon Economics. Eksportmeldingen 2021. `https://www.menon.no/wp-content/uploads/2021-58-Eksportmeldingen-2021.pdf`, (accessed: 06.06.2022).

[30] Y. Meyer. *Wavelets and Operators*. Cambridge University Press, 1992.

[31] Norsk Rikskringkasting (NRK). Skader trugar bereevna på 1000 norske bruer. https://www.nrk.no/vestland/1000-norske-bruer-har-skader-som-trugar-bereevna-1.15819652, (accessed: 03.06.2022).

[32] Alon Oring. Fast and robust structural damage analysis of civil infrastructure using uav imagery. In Carlo Pellegrino, Flora Faleschini, Mariano Angelo Zanini, José C. Matos, Joan R. Casas, and Alfred Strauss, editors, *Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures*, pages 1251–1260, Cham, 2022. Springer International Publishing.

[33] Edwin Reynders and Guido De Roeck. Reference-based combined deterministic–stochastic subspace identification for experimental and operational modal analysis. *Mechanical Systems and Signal Processing*, 22(3):617–637, 2008.

[34] Riksrevisjonen. Riksrevisjonens undersøkelse av overføring av godstransport fra vei til sjø og bane. `https://www.riksrevisjonen.no/globalassets/rapporter/no-2017-2018/godstransport.pdf`, (accessed: 06.06.2022).

[35] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[36] Natasha Samko and Harpal Singh. Some new contributions concerning non-separable spaces with applications on signal processing techniques with in Bayesian framework. *Mathematical Methods in Applied Sciences*, 2022. 12 pages, (To appear).

[37] Arthur Shi and Xiao-Hua Yu. Structural damage detection using artificial neural networks and wavelet transform. In *2012 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*, pages 7–11, 2012.

[38] David Silver, Thomas Hubert, and Julian Schrittwieser et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[39] Harpal Singh. The Hålogaland bridge - descriptions, challenges and related research under arctic condition. In *9th International Operational Modal Analysis Conference, IOMAC 2022 Vancouver*. 15 pages (To appear).

[40] Harpal Singh. *Some new mathematical and engineering results connected to structural problems*. PhD thesis, UiT The Arctic University of Norway, 2022.

[41] Harpal Singh and Niklas Grip. Recent trends in operation modal analysis techniques and its application on a steel truss bridge. *Nonlinear Studies*, 26(4):911–927, 2019.

[42] Harpal Singh, Niklas Grip, and Per Johan Nicklasson. A comprehensive study of signal processing techniques of importance for operation modal analysis (OMA) and its application to a high-rise building. *Nonlinear Studies*, 28(2):389–412, 2021.

[43] Mohammad Hesam Soleimani-Babakamali, Roksana Soleimani-Babakamali, and Rodrigo Sarlo. A general framework for supervised structural health monitoring and sensor output validation mitigating data imbalance with generative adversarial networks-generated high-dimensional features. *Structural Health Monitoring*, 21(3):1167–1182, 2022.

[44] Rupesh K Srivastava, Klaus Greff, and Jürgen Schmidhuber. Training very deep networks. In *Advances in Neural Information Processing Systems*, volume 28. 9 pages, 2015.

[45] Kristoffer Tangrand and Bernt Bremdal. The flexnett simulator. *IOP Conference Series: Earth and Environmental Science*. Volume 352, 8 pages, 2019.

[46] Kristoffer Tangrand and Bernt Bremdal. Using deep learning methods to monitor non-observable states in a building. *Proceedings of the Northern Lights Deep Learning Workshop*. Volume 1, 6 pages, 2020.

[47] Kristoffer Tangrand and Bernt A. Bremdal. Using ant colony optimization to determine influx of evs and charging station capacities. In *2016 IEEE International Energy Conference (ENERGYCON)*. 6 pages, 2016.

[48] Verdens Gang. The neglected bridges. `https://www.vg.no/spesial/2017/de-forsomte-broene/inspeksjoner/`, (accessed: 05.09.2021).

[49] Verdens Gang. The neglected bridges map. `https://www.vg.no/spesial/2017/de-forsomte-broene/kart/inspeksjon,{(accessed: 05.09.2021)}`.

[50] Vahid Yaghoubi, Liangliang Cheng, Wim Van Paepegem, and Mathias Kersemans. Cnn-dst: Ensemble deep learning based on dempster–shafer theory for vibration-based fault recognition. *Structural Health Monitoring*. 20 pages, 2022.

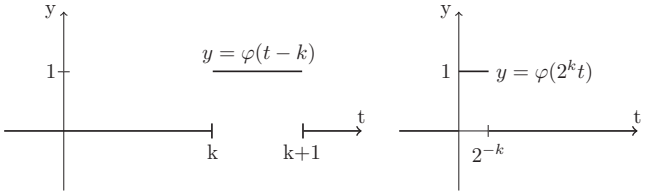
A On Wavelet Theory

There exists both a discrete version (comparable with Fourier series) and a continuous version (comparable with Fourier transforms) of Wavelets Theory. The discrete version can be described as follows:

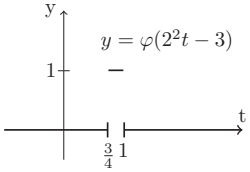
The Classical Haar Mother Wavelet φ and first Haar Wavelet ψ are defined as follows:

$$\varphi(t) = \begin{cases} 1, & 0 \leq t \leq 1 \\ 0, & \text{elsewhere} \end{cases} \qquad \psi(t) = \begin{cases} 1, & 0 \leq t \leq \frac{1}{2} \\ -1, & \frac{1}{2} < t \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

The translations of $\varphi(t-k), k \in \mathbb{R}$ and $k : t$ dilations of φ can be represented as follows, respectively:



We can also combine Dilation and Translation, as follows:



The series $\sum_0^\infty \alpha_k \varphi_k(t)$ is called the Haar series of $f(t)$. And since the system $\{\varphi_n\}$ is orthonormal, from the general Fourier theory it follows that $f(t)$ can be reconstructed exactly as follows:

$$f(t) = \sum_0^\infty \alpha_k \varphi_k(t)$$

from its "basis functions"

$$\varphi_k(t) = 2^{\frac{n}{2}} \varphi(2^n t - k)$$

and the corresponding Haar(-Fourier) coefficients

$$\alpha_k = \int_0^1 f(s) 2^{\frac{n}{2}} \varphi(2^n s - k) ds \ .$$

Remark 1 Wavelets are functions that slice data into differing frequency components. As such, the scale and resolution will match for each component. This

means that wavelets can accommodate both large and small features, depending on the scale and resolution. Wavelets are better at handling signals containing discontinuities and sharp spikes compared to traditional Fourier methods.

Remark 2 These original Haar wavelets can be variated in various ways e.g. involving different mother wavelets. The different wavelet families make different trade-offs between how compactly the basis functions are localized in space and how smooth they are. The *Daubechies wavelet* family is one such example. Usually, each wavelet in a family is named after the number of vanishing moments it contains. A vanishing moment is a rigorous mathematical term that relates to the number of coefficients a wavelet has. The more vanishing moments, the higher complexity can be represented by the scaling function. For applications even more general discrete wavelets are used in different programs. In the basic cases the function space L^2 is used. But for applications it is sometimes important to consider more general function spaces like Besov spaces.

Remark 3 In recent decades, wavelet methods have shown themselves to be of considerable use in Fourier analysis and related applications. The strength of wavelet methods lies in their ability to describe local phenomena more accurately than the traditional expansions in sinus and cosinus can. This is because wavelet functions are localized in space. Thus, wavelets are ideal in many fields where an approach to transient behavior is required, for example, in considering acoustic or seismic signals, image processing, damage detection in bridges (see [40]). For applications that are even more general, discrete wavelets are used in even more different applied fields such as astronomy, nuclear engineering, subband coding, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, radar, human vision, and pure mathematics applications such as solving partial differential equations (see e.g. [16]).

Remark 4 As mentioned above, there exists also the continuous wavelet transform, which is an integral transform, comparable with the Fourier transform. Both of these transforms are very important for various types of applications. For more information, see also the books referred to in the next remark.

Remark 5 From the first discoveries of Alfred Haar (1885–1933) it has been an almost unbelievable development of the wavelet theory. The reasons are both the interest from the mathematical point of view and the applications described above. In particular, more than 50 books on the subject has been written. Here we just mention [8],[14],[22],[23],[24],[28] and [30], as well as the papers [10] and [11], which illustrates various aspects of this broad science and also how many well-known authors have been involved.

Paper E

A comprehensive study of wavelets and artificial intelligence algorithms for SHM and its application on a concrete railway arch bridge

Kristoffer Tangrand, Harpal Singh

Technical report, UiT The Arctic University of Norway, 2022.

E

A Comprehensive Study of Wavelets and Artificial Intelligence Algorithms for SHM and its Application to a Concrete Railway Arch Bridge

Kristoffer Tangrand[*], Harpal Singh[*], Niklas Grip[**]
[*]UiT – The Arctic University of Norway
[**]Luleå University of Technology

July, 2022

[*] *Corresponding Author*. E-mail: ktangrand@gmail.com

Abstract

In this paper, we present a comprehensive study of wavelet theory with a focus on structural damage detection. A new example of the application of operational modal analysis (OMA) techniques to a concrete railway arch bridge located over the Kalix river in Långforsen is presented. Results from the OMA techniques are used for finite element model (FEM) updating. Further, artificial intelligence algorithms that can be useful for addressing the problem of missing data sets in structural health monitoring technologies are reviewed.

Keywords: Structural health monitoring, Vibration analysis, Operational modal analysis, Damage detection, Bridge, Arctic conditions, Finite element model, Signal processing, Wavelet transform, Multi-resolution analysis, Artificial intelligence, Neural network, Wavelet network and Statistical methods.

AMS Classification (2010): 35A22, 45A05, 44A15, 65T50, 65T60

1 Introduction

As described in the previous article [45], bridge infrastructure in Norway is in a stressed situation due to various challenges. Quite recently, in August 2022, the Tretten bridge catastrophically collapsed in Norway, where a truck and a car became stuck. This bridge is located across the Gudbrandsdalslågen river in the Oyer municipality. Luckily, there have been no causalities (see [8]). In another incident in May 2022, a bridge in Kvænangen municipality, Nord-Troms had serious damage, leading to the closure of a bridge over the important E6 European road [49]. More research is needed in this area so that precautionary measures can be taken to prevent such incidences in the future.

The discussions in the paper [45] gave rise to the notion of writing this paper. The idea to do this paper originated during the discussions in papers [41], [43] and [44], where we emphasized the significance of utilising OMA techniques for structural health monitoring (SHM). In this paper, we do a comprehensive study of wavelet theory with a focus on structural damage detection. Moreover, we present a new example where the results from OMA of a concrete railway arch bridge located over the Kalix river in Långforsen are used for FEM updating.

This paper is organized as follows: In Section 2, some basic definitions of wavelet theory are given and an overview of using wavelets for structural damage detection is described, along with the idea of wavelet networks to resolve some of the challenges that arise in this process. In Section 3, a new example of the application of OMA techniques to a concrete railway arch bridge located over the Kalix river in Långforsen is presented. Results from the OMA techniques are used to develop an FEM-updated model of the bridge. Further, Section 4 is dedicated to artificial intelligence algorithms that can be useful for addressing the problem of missing data sets in SHM technologies. Finally, in Section 5 we present some concluding remarks.

2 Wavelet Theory and Applications

Wavelet transforms are one of the most important transforms that have been developed in order to overcome the shortcomings of the Fourier transforms. The basic concepts of wavelet theory that are crucial for the applications discussed in this paper can be found in (see [18], [19]), and the appendix of the paper [45]. More detailed studies can be found in the book [27]. With the use of the wavelet transform, one can cut data, functions, or operators into different frequency components with a resolution that is matched to their scale. The continuous wavelet transform (CWT) W for a time signal $x(t)$ is defined as follows:

$$W(s,\tau) = \int_{-\infty}^{\infty} x(t)\psi(s,\tau,t)dt, \tag{1}$$

where the wavelet function $\psi(s,\tau,t)$ includes a scaling variable s, a translation variable τ and the time t. Starting with a mother wavelet $\psi_0 = \psi_0(u)$ with the property $\int \psi_0(u)du = 0$, the wavelet building blocks $\psi = \psi(s,\tau,t)$ are defined by (see e.g. [27])

$$\psi(s,\tau,t) = \frac{1}{\sqrt{s}}\psi_0(\frac{t-\tau}{s}). \tag{2}$$

The amplitude of the wavelet function $\psi(s,\tau,t)$ can be controlled by the variation in the scaling variable s, whereas the translation factor τ controls the location of the wavelet function in time. Therefore, depending upon the specific problem, wavelet transforms can be designed accordingly to fit the specific problem.

Since the building blocks ψ builds an orthonormal system (only for the discrete wavelet transform), the general Fourier theory is applicable; in particular, the inverse transform W^{-1} of W exists. This gives the re-construction of the signal $x(t)$ is defined as follows:

$$x(t) = C_\psi \iint_{-\infty}^{\infty} W(s,\tau)\psi(s,\tau,t)dsd\tau. \tag{3}$$

The wavelet transform described above is an integral transform which can be compared with the usual Fourier transform. There is also a discrete wavelet transform, which in a sense, is comparable with the usual Fourier series (for further details, see also Appendix in [45]). Both are very important for various types of applications, e.g. those discussed in this paper.

In the following sub-sections we cover the topics of damage detection methods in structures, data compression, and multi-resolution analysis that are based on the further developments of wavelets, which are relevant to this paper.

2.1 Wavelets in structural damage detection

Most of the signals in structural damage detection methods are time-based signals that are recorded by the sensors. Structures vibrate as a result of natural or artificial excitations, such as earthquakes, wind, or artificial excoriations. The output signals from these vibrations, such as accelerations, strains, or displacements, can be recorded. These signals have a non-stationary nature, which means that with time, their features change. But these signals can contain lots of information, which can be useful for stating the health of the structure. Mostly, the signals in the time domain are converted to the frequency domain because damage identification is more effective in the frequency domain (see [11]).

Windowing the signal can provide the time localization with Fourier transforms, but since the window lengths are the same, it does not matter what the frequency component in the signal is. On the other hand, depending on the frequency components, wavelet transforms enable varied time resolutions. In contrast to Fourier Transforms, this results in high accuracy in numerical differentiation and flexible implementation of boundary conditions(see [11]).

Moreover, Fourier sine and cosine functions are not localized in space whereas the wavelet functions are localized, which makes such functions using wavelets "sparse". This feature makes many functions and operators using wavelets "sparse" (see [18]). In applications like data compression, feature detection, and noise removal from temporal signals, wavelets thus become very helpful tools. Wavelets are extremely helpful for non-stationary and non-linear problems that arise in the identification of structural damage since there are no strict limitations on frequency and time resolution.

For example, wavelet transformations are a useful approach for detecting structural degradation (see [23] and [25]). Using wavelet transformations, modal parameters for a structure like mode frequency and modal damping can be calculated (see [35]). The authors of [33] used wavelet transform of the vibration

signals from damaged and undamaged structures , and it was possible to detect and locate the damage.

Another study by the authors employed continuous wavelet transform (CWT) and discrete wavelet transform (DWT) techniques to find damage to a tall airport traffic control tower caused by seismic activity. It was found that CWT could detect seismically caused damage even if the vibration signals had noise in them. In contrast, DWT is more effective than CWT at detecting changes in the stiffness of a structure despite being more sensitive to noise.

In the direction of damage detection using wavelets, researchers around the globe are working to develop new methods based on wavelets such as adaptive wavelet transform (AWT) (see [7] and [32]), synchrosqueezed wavelet transform (SWT) (see [4], [14] and [28]), and the empirical wavelet transform (EWT) (see [17]). Brief descriptions of these new wavelet transform methods along with examples of applications on structures can be found in [44].

Moreover, continuous development is needed for the development of new mathematical methods based of Fourier theory that can address the challenges that appear in signal processing for damage detection problems in civil engineering infrastructure (see [1], [2], [5] and [37]) especially under arctic conditions (see [41]). A significant amount of data is produced in the field of structural damage detection from numerous sensors that are used to look for damage. As a result, a lot of data storage space and processing power are needed. We use wavelet theory to address this problem in the subsection that follows.

2.2 Wavelets in data compression

The Fourier transform is an integral transform, which means that it is defined as the integral of the transform function over the entire time or space domain (see [44]). As we have seen, the continuous wavelets are integral but consist of other types of building blocks, which gives great advantages for various applications. The wavelet transform is also reversible, which means that the original signal can be reconstructed from the wavelet transform by taking the inverse transform. In addition, the wavelet transform is localized in time and space. This means that the transform is zero outside of a small neighborhood around the point in time or space where the transform is being evaluated. This property is especially useful for compression, since it allows us to focus on the local features of the signal. In contrast, the Fourier transform is non-local since it is defined over the entire time or space domain. As such, wavelets have an inherent sequential property, which makes them useful for handling time series. Another strength is that the wavelet structure can be directly related to the shape of the underlying features of a signal.

2.3 Multi-resolution analysis

The DWT described earlier in this section is a multiresolution, band-pass representation of a signal. Wavelet multiresolution analysis(MRA) is a technique that mathematically describes the transition between information density levels.

Transitioning between levels reduces or increases the information density. As such, it can be used to reduce the number of measurements required to represent a signal, which basically amounts to compression. Wavelet neural networks (WNN) are a type of artificial neural network (ANN) that uses Wavelet MRA instead of sigmoid activation functions to represent the weights of the network.

As described earlier, wavelets are well known for data compression. Yet for many applications, such as image processing, a multi-resolution representation of the data is required. A multi-resolution representation consists of a set of data representations at different scales, or resolutions. A representation at one scale is generated by low-pass filtering the data at a coarser scale. The multi-resolution representation can then be used to recover the original data by interpolating between the representations.

Wavelets are, as such, a mathematical representation of a signal in which it is represented as a set of basis functions. Each basis function can be defined by a set of coefficients. Each set of coefficients can be obtained by the discrete wavelet transform for the corresponding dyadic dimension (see [15]).

2.4 Wavelet networks

Earlier work has shown that wavelet MRA can be used to perform multiscale analysis of time series data. Multi-scale analysis is a set of methods for studying the statistical properties of a signal at different scales. Different scalings are often provided by decomposing a signal into a set of wavelet components. This representation can be used as a replacement for the sigmoid activation functions in neural networks (see [52]).

The basic model of a wavelet network takes form as a three-layer network, with an input, middle (hidden) and output layer. The computing units of the hidden layer are referred to as wavelons, as a contrast to neurons in regular neural networks. These computing units dilate and translate the input variables using the mother wavelet.

The equation $f(x) = \sum_i w_i \phi_{n_i}(x)$ describes how the wavelet MRA decomposition provides the weights w_i for the network. These weights are changed when the network learns, or adapts, to the training data. The wavelet basis is thus modified according to the training data.

The result is that the weights of the network are intrinsically tied to the signal it is representing (see [3]). The output of the network can be mathematically represented as

$$g_\lambda(\mathbf{x}; \mathbf{w}) = \hat{y}(\mathbf{x}) = \omega_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} \omega_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} \omega_i^{[0]} \cdot \chi_i \tag{4}$$

where $\Psi_j(\mathbf{x})$ is a multidimensional wavelet which is constructed by the product of m scalar wavelets, $\mathbf{x}$ is the input vector, m is the number of input nodes, while λ is the number of hidden units and ω are network weights, or trainable variables. The notations [0] and [2] denote the input and output layers, respec-

tively. For further details, see Figure 1 in [3]. The wavelet network is then trained using back-propagation.

A lot of research is going on in the field of wavelet transforms for damage detection. Vibrational data from accelerometers is used to estimate parameters like mode shapes, mode frequency and modal damping. In SHM of structures, OMA approaches are utilized to identify the modal characteristics of the structure (see [9], [34] and [43]). In the following section, we present a new example of operational modal analysis (OMA) that is used for finite element model (FEM) updating of a concrete railway arch bridge.

3 Application of OMA techniques on a concrete railway arch bridge for FEM updating

In this section, we describe some modal analysis results for a concrete railway arch bridge. The bridge is situated over Långforsen outside Kalix in Sweden. It is a 177 meter long and 60 meter high bridge, built in 1960 (see Figure 1). The bridge's owner, Trafikverket, was interested and wanted to increase the allowed speed and also wanted to increase the maximum allowed axis load from 225 to 300kN. Therefore, vibration measurements were done, as well as measurements of strain and displacements with trains passing at different speeds.

Figure 1: The concrete arch railway bridge over Kalix river in Långforsen.

3.1 Design and construction

The railway arch bridge over the Kalix River at Långforsen has a length of 177.3 m. The central arch is 89.5 m, while the two side spans are 42 m each. The free spans in the bridge have lengths of 13.0 + 12.8 +12.6 + 87.92 +12.6 + 12.8 + 13.0 m = 164.7 m. The bridge was built in 1960. The arch of the bridge has a reinforced concrete slab via underlying longitudinal and transversal concrete beams that are connected through fixed columns. A reinforced concrete hollow box girder with two hollow chambers makes up the arch. The cross section is highest at the connection to the arch abutment and lowest at the arc's crown. The actual train load corresponds to the locomotive's axle load of 250 kN and a distributed load of 72 kN/m.

3.2 Measurement plan

It was unclear whether there was sufficient wind to excite ambient vibrations sufficiently for modal analysis. Therefore, before each measurement, a T43 ra 240 railway engine was driven across the bridge at speeds ranging from 35 to 63 km/h. The engine's weight, which is divided among its eight wheels, is 72 tonnes, while its dynamic weight is 79 tonnes. To have the same linear system (i.e. bridge alone) as in our FE models and to remove nonlinear effects, such as noises from the wheels clattering against the rails and bridge terminals clattering against the foundation, the measurements started after the engine passed the bridge.

During the measuring days, the wind suddenly decreased, which steadily

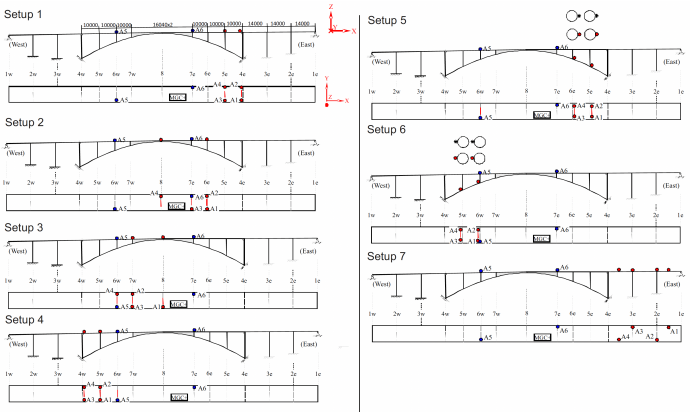


Figure 2: Measurement plan showing the placement of accelerometers for the concrete arch railway bridge across the Kalix river in Långforsen.

reduced the signal-to-noise ratio (SNR). Although the SNR was not as good as in earlier measurements from 2009, there was still a clear consistency between the modeled and measured modal data (mode shapes and frequencies).

The Triaxial Colibrys SF 3000L accelerometers were securely fastened to the bridge using expander bolts. Accelerometers were connected with six-wire twisted pair cables to an MGCplus data acquisition system with an ML801B amplifier module. All measurements were made at a sample rate of 1200 Hz. The six-parameter calibration approach described in [16] and [20] was used to calibrate the measurements. Figure 2 shows the measurement plan for the placement of accelerometers over the railway concrete arch bridge.

The Stochastic Subspace Identification (SSI) method was used to carry out modal analysis in the ARTeMIS 4.0 software.

3.3 Finite element model updating results

To employ structural control and SHM techniques, the FEM must be highly accurate. The type of FEM used to represent the structural members and the attributes given to these elements both affect how accurate the model is. The FEM of the railway arch bridge over the Kalix River at Långforsen along with its dynamic properties is presented in [36]. The boundary conditions, geometry, or material qualities that change as a material ages are not precisely determined by the finite element model. Depending on the loading conditions, material qualities might cause non-linearity. As a result, the FEM model needs to be calibrated using data from actual structures. The key parameters in the finite element model of the structure are calibrated to minimize the smallest feasible discrepancy between the observed vibrations and the simulated vibrations using a numerical optimization technique called FEM updating [43].

Over the years, a lot of effort has been put into developing various FEM models. Two different bridge model types were created using Abaqus/Brigade in 2011: A comprehensive model with foundations (Type I) and a simplified beam element model where the foundations have been exchanged for springs (Type II). The advantage of the earlier model is that the anticipated results from it could be more reliable and closer to the "actual results," but the drawback is that the scale of the problem is quite large. Type II has 47,438 components compared to Type I, which has 93,910. Moreover, Type II has 282,808 variables in comparison to the 438,800 variables in Type I.

The Långforsen Bridge's FEM update is computed using default settings in the main file located in the FEMU/Impl/LangforsenBridge directory and the results are shown in Figure 3. It can update the elasticity modulus and boundary conditions (which are modeled as springs) at various locations along the bridge.

The primary issues were that the FEM update did not improve the top right bending mode shape in Figure 3. The top right bending mode shape in Figure 3 has a mode frequency of 2.78 Hz, while the measured mode frequency was 3.06 Hz. The beam element model, which was created to reduce size and increase computation efficiency for updating finite element models and dynamic response

simulations, may be one of the possible reasons. However, this model had the downside that the arch could not undergo torsion. Therefore, continued refining of the FEM model and FEM updating software is needed for updating shear modulus with some degrees of freedom [46].

One of the major problems while doing the analysis of this railway arch bridge and similar projects like steel truss bridges (see [43]) and high-rise buildings (see [44]) was the calibration of sensors with respect to seasonal differences between temperature measurements (winter vs summer measurements). Because of extremely low wind conditions, missing data sets were also a problem. One possible way to deal with such problems could be the use of artificial intelligence, as mentioned in our previous article. In the following section, we will describe this in detail.

4 Artificial intelligence and the problem of missing data

The fields of contemporary machine learning and artificial intelligence are fast-moving fields. New models such as transformers, residual networks (ResNet), and Generative Adversarial Networks (GAN) have been quickly introduced to application fields such as SHM. For more details on recent advances, see [45].

Missing and irregular data are a problem in most fields that use real-life sensor data. This can be caused by sensors that are not synchronized, events that happen irregularly or network latency. In contrast, most machine learning models expect evenly sampled data. This introduces the need for interpolation or imputation of missing or irregular data. This can, in turn, introduce unwanted bias into the model's latent space.

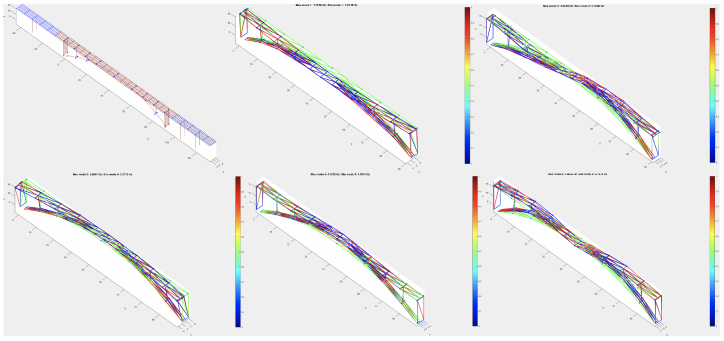


Figure 3: FEM updating results for the concrete arch railway bridge over the Kalix river in Långforsen

The SHM and OMA spaces naturally rely on sensor data. As such, the problem is present here (see [21]). In recent years, there has been an influx of new machine learning models trying to handle such data problems. We review those models of special interest for the applications in this paper.

A time series is temporal data collected along the axis of time. This implies that the data is ordered temporally. A time series is usually a real-valued variable. It can be univariate or multivariate, corresponding to one or multiple dimensions, respectively. Time series can be regular or irregular, depending on whether the values of the variables are collected at equal or unequal time gaps. If a time series can be predicted exactly, it is deterministic. If it can only be determined probabilistically, it is stochastic.

Most datasets collected from sensors measuring some real-world phenomena exhibit some form of irregularity. This can have several causes, such as different sample rates, network latency occurring when data is stored, or simply bugs or errors leading to missing or corrupted values. The literature often refers to this as missing or corrupted data. However, it is clearly a special case of an irregular time series.

Multi-modal data, such as the case with multiple similar or different sensors or data sources, can likewise introduce related problems, such as measurements that are not timed or synchronized. Skewed synchronization can lead to decision-making algorithms misinterpreting correlation or causality between variables or events.

The literature traditionally identifies two distinct strategies for handling irregular time series (see e.g. [48]). Imputation methods, such as interpolation or forward filling, can be used to create a complete time series. This is the more traditional and commonly used approach. Second, and contrarily, are models that inherently handle irregular data.

4.1 Statistical methods

Interpolation between irregular time spans assumes that missing data between recorded points behaves predictably. Such an assumption can obviously not always hold. Even small spans of missing data can introduce serious amounts of unpredictable bias. Other common approaches include padding training samples with zero values where data is missing, but this also influences the parameters of an algorithm to varying degrees. The time stamp itself can also be used as an input feature for the machine learning model.

More sophisticated approaches are described in the literature, such as interpolation networks (see [40]). This model uses neural networks to learn the interpolation features for a specific time series in latent space.

Another sophisticated approach is to generate so-called synthetic data. Synthetic data is basically a form of simulated data that can be combined with real-world data. Such data can be advantageous in situations when real data samples are scarce, missing or unavailable. In fact, this approach has been successful in applications such as damage detection in electric grids (see [30]) and self-driving cars (see [50]). In recent years, the aforementioned GAN model has

become a benchmark in many fields for generating synthetic data, especially image data augmentation (see [38]). GANs are also fast becoming the preferred generative model for time series (see [24]). Generating data for photovoltaic solar energy generation has also shown to be a very promising method (see [53]).

4.2 Gated RNNs

A number of recent developments in gated RNN architectures for time series are reviewed. Gated RNNs, such as LSTM and GRU, still stand as the most resilient and performant models for time series. However, vanilla versions such as LSTM and GRU are not designed to handle irregular data, possibly introducing bias and suboptimality.

4.2.1 t-LSTM

In particular, some architectures are designed to work with irregular time series naturally, without imputation. This design involves modifying the structure of the RNN to accommodate the irregular information supply. In [6], the LSTM memory cell, rather than the forget gate, is modified. The model is thus able to take the time span into consideration, according to the authors, hence the name time-aware LSTM(t-LSTM). The input dynamics of the model are not complex. In addition to the sample itself, another vector includes the elapsed time as a weight transformed by a time decay function.

4.2.2 Phased LSTM

In contrast, a more advanced structural change is presented in [29]. This model, called Phased LSTM, introduces two new oscillatory gates to the vanilla LSTM model, which reduces the number of updates required by the memory cell and hidden state. The authors explicitly state that this architecture is intended to handle irregular sampling rates. The model has been proven to be more effective than vanilla LSTM on long, irregular time series (see [54]).

4.2.3 Bistable recurrent cell

In [47], Vecoven et al. took inspiration from how biological neurons can store information for arbitrary time spans through a process called bistability. This lets the cells themselves remember information from their own past states and inputs. This leads to long-term memory that is potentially very useful for long, sparse, and irregular time series.

4.3 ODE-RNN and derivatives

Neural Ordinary Differential Equation(ODE) with an RNN lets the model have continuous-time hidden dynamics. ODE-RNN can naturally handle arbitrary time gaps between observations (see [12]).

The vanishing gradient problem is inherent to the Backpropagation Through Time(BPTT) algorithm, which is used in basic RNNs. As shown in [26], ODE-RNNs also suffer from this shortcoming. The same authors also show that additional gates found in the LSTM can be included to alleviate the problem.

The innovations that came with Neural ODEs (see [12]) are utilized to write the standard GRU as a difference equation (see [13]). This is further combined with a network that updates the current hidden state with incoming observations, making the combined model GRU-ODE-Bayes (see [10]).

4.4 Transformer

The performant and efficient Transformer architecture has in recent years outperformed various neural network-based models in many different fields. Transformers have been adopted to model time series and have been shown to be performing in the top tier. As for irregular time series, in [22], a Transformer model was used to compare with their SEFT-Attn model. It was adapted to a classification task by mean-aggregating the final representation, and it was then fed into a Softmax layer to predict classes.

4.5 Attention networks

The attention mechanism, usually connected to transformer models, was investigated for irregular time series in [39]. The approach is described as representing the irregular time series at a fixed set of reference points. An encoder-decoder framework is then utilized, such that the encoder produces a fixed-length latent representation over the aforementioned reference points. The decoder then uses the latent space to reconstruct the set of observed time points. The training is done by a Variational Auto-encoder(VAE). The main advantage compared to other cutting-edge techniques is that it gives faster training (1-2 orders of magnitude). This is especially noticeable compared to ODE-RNN methods, since they require an ODE solver.

4.6 Set Functions for time series

Recent advances in differentiable set function learning allow for models that are able to handle irregular and unaligned multivariate time series (see [22]). The multivariate time series is encoded as a set. Each data point is treated as having three values; the time it was collected, which variable it belongs to, and its value. As such, the data points lose their sequentiality, but the time information itself is still preserved. Furthermore, the sets are fed into the deep sets architecture with an attention mechanism (see [51]).

5 Concluding Remarks

Remark 5.1 The research discussed in this paper is also related to some recent research in more theoretical Fourier analysis [1], [2] and the PhD thesis

[42]. We aim to further develop the investigations in this paper as wavelets are unbounded systems.

Remark 5.2 In the forthcoming article, we further aim to develop the theory of wavelet neural networks. We strongly believe that this work will help to address the issue of missing data sets for structural damage detection in structures similar to the concrete railway arch over the Kalix River at Långforsen as discussed in this paper.

Remark 5.3 Several experiments have been conducted on the concrete railway arch over the Kalix River at Långforsen over time. In this paper, we present the results for the finite element model updating in Figure 3. Continuous refining and development of FEM (see [36]) and FEM updating models is needed for updating shear modulus with some degrees of freedom.

Remark 5.4 Multiple promising techniques and models for handling or mitigating missing data have appeared in the AI/ML space recently (see [48]). However, investigating performance requires access to both relevant data and domain-specific knowledge as well as machine learning knowledge. It is thus clear that much work remains in this direction.

Remark 5.5 The research groups at UiT The Arctic University of Norway and Luleå University of Technology Technology are interested in the new research problems that result from this difficult challenge of SHM and OMA in extreme arctic conditions [41] and [45].

Remark 5.6 It is a close connection between wavelet theory and the type of modern Fourier analysis that is described in the new book [31]. We aim to further investigate this fact in relation to the problems described above in a forthcoming paper.

Acknowledgement

We thank Professor Lars-Erik Persson for several generous suggestions that have improved the final version of this paper.

References

[1] Gabdolla Akishev, Lars-Erik Persson, and Harpal Singh. Inequalities for the Fourier coefficients in unbounded orthogonal systems in generalized Lorentz spaces. *Nonlinear Studies*, 27(4):1137–1155, 2020.

[2] Gabdolla Akishev, Lars-Erik Persson, and Harpal Singh. Some new Fourier inequalities and Jackson–Nikolskii type inequalities in unbounded orthonormal systems. *Constructive Mathematical Analysis*, 4(3):291–304, 2021.

[3] Antonios Alexandridis and Achilleas Zapranis. Wavelet neural networks: a practical guide. *Neural Networks*, 42:1–27, 2013.

[4] Juan P Amezquita-Sanchez and Hojjat Adeli. Synchrosqueezed wavelet transform-fractality model for locating, detecting, and quantifying damage in smart highrise building structures. *Smart Materials and Structures*, 24(6):15, 2015.

[5] David Baramidze, Lars-Erik Persson, Harpal Singh, and George Tephnadze. Some new results and inequalities for subsequences of Nörlund logarithmic means of Walsh–Fourier series. *Journal of Inequalities and Applications*, 2022(1):1–13, 2022.

[6] Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 65–74, 2017.

[7] Ales Belsak and Joze Flasker. Adaptive wavelet transform method to identify cracks in gears. *EURASIP Journal on Advances in Signal Processing*, 2010:1–8, 2010.

[8] Mette Finborud Børresen, Marte Iren Noreng Trøen, Line Fosser Vogt, and Kristina Hågensen. Vil fjerne brua i morgen - E6 fortsatt stengt (in norwegian). https://www.nrk.no, (accessed: 16.08.2022).

[9] Rune Brincker and Carlos Ventura. *Introduction to Operational Modal Analysis*. John Wiley & Sons, 2015.

[10] Edward De Brouwer, Jaak Simm, Adam Arany, and Yves Moreau. *GRU-ODE-Bayes: Continuous Modeling of Sporadically-Observed Time Series*. 2019.

[11] Hua-Peng Chen and Yi-Qing Ni. *Structural Health Monitoring of Large Civil Engineering Structures*. Wiley Online Library, 2018.

[12] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[13] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, volume abs/1406.1078, 2014.

[14] Ingrid Daubechies, Jianfeng Lu, and Hau-Tieng Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and Computational Harmonic Analysis*, 30(2):243–261, 2011.

[15] Eduardo Martin Moraud). Wavelet networks. "https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV0809/martinmoraud.pdf".

[16] Thomas Forsberg, Niklas Grip, and Natalia Sabourova. Non-iterative calibration for accelerometers with three non-orthogonal axes, reliable measurement setups and simple supplementary equipment. *Measurement Science and Technology*, 24(3), 2013. 14 pages.

[17] Jerome Gilles. Empirical wavelet transform. *IEEE Transactions on Signal Processing*, 61(16):3999–4010, 2013.

[18] Amara Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2):50–61, 1995.

[19] Niklas Grip. *Wavelet and Gabor frames and bases: Approximation, Sampling and Applications*. PhD thesis, Luleå University of Technology, 2002.

[20] Niklas Grip and Natalia Sabourova. Simple non-iterative calibration for triaxial accelerometers. *Measurement Science and Technology*, 22(12), 2011. 12 pages.

[21] Niklas Grip, Natalia Sabourova, Yongming Tu, and Lennart Elfgren. *Vibrationsanalys för tillståndsbedömning av byggkonstruktioner: Tillämpningsexempel:(Main results and summary in Swedish. Detailed results in English Appendices.)*. 2017.

[22] Max Horn, Michael Moor, Christian Bock, Bastian Rieck, and Karsten Borgwardt. Set functions for time series. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 4353–4363. PMLR, 13–18 Jul 2020.

[23] Zhikun Hou, Mohammad Noori, and R St Amand. Wavelet-based approach for structural damage detection. *Journal of Engineering mechanics*, 126(7):677–683, 2000.

[24] Paul Jeha, Michael Bohlke-Schneider, Pedro Mercado, Rajbir Singh Nirwan, Shubham Kapoor, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. PSA-GAN: Progressive self attention GANs for synthetic time series. In *ICML 2021 Time Series Workshop, ICLR 2022*, 2021.

[25] Hansang Kim and Hani Melhem. Damage detection of structures by wavelet analysis. *Engineering Structures*, 26(3):347–362, 2004.

[26] Mathias Lechner and Ramin Hasani. Learning long-term dependencies in irregularly-sampled time series, 2020. "https://arxiv.org/abs/2006.04418".

[27] Stéphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition*. Elsevier, 2008.

[28] Marko Mihalec, Janko Slavič, and Miha Boltežar. Synchrosqueezed wavelet transform for damping identification. *Mechanical Systems and Signal Processing*, 80:324–334, 2016.

[29] Daniel Neil, Michael Pfeiffer, and Shih-Chii Liu. Phased LSTM: Accelerating recurrent network training for long or event-based sequences. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29, 2016.

[30] Van Nhan Nguyen, Robert Jenssen, and Davide Roverso. Intelligent monitoring and inspection of power line components powered by UAVs and deep learning. *IEEE Power and Energy Technology Systems Journal*, 6(1):11–21, 2019.

[31] Lars-Erik Persson, George Tephnadze, and Ferenc Weisz. *Hardy Spaces and Summability of One-Dimensional Vilenkin-Fourier Series*. Birkhäuser/Springer, October 2022.

[32] Hongya Qu, Tiantian Li, and Genda Chen. Adaptive wavelet transform: Definition, parameter optimization algorithms, and application for concrete delamination detection from impact echo responses. *Structural Health Monitoring*, 18(4):1022–1039, 2019.

[33] Mohammad Hossein Rafiei and Hojjat Adeli. A novel machine learning-based algorithm to detect damage in high-rise building structures. *The Structural Design of Tall and Special Buildings*, 26(18):1–11, 2017.

[34] Carlo Rainieri and Giovanni Fabbrocino. *Operational Modal Analysis of Civil Engineering Structures*. Springer, 2014.

[35] M Ruzzene, A Fasana, L Garibaldi, and B Piombo. Natural frequencies and dampings identification using wavelet transform: application to real data. *Mechanical Systems and Signal Processing*, 11(2):207–218, 1997.

[36] Natalia Sabourova, Niklas Grip, Arto Puurula, Ola Enochsson, Yongming Tu, Ulf Ohlsson, Martin Nilsson, Lennart Elfgren, Anders Carolin, and Håkan Thun. The railway concrete arch bridge over kalix river: dynamic properties and load carrying capacity. In *FIB Symposium: Concrete Structures for Sustainable Community 11/06/2012-14/06/2012*, pages 609–612. Swedish Concrete Association, 2012.

[37] Natasha Samko and Harpal Singh. Some new contributions concerning non-separable spaces with applications on signal processing techniques with in Bayesian framework. *Mathematical Methods in the Applied Sciences*, 2022. 7 pages.

[38] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, July 2019.

[39] Satya Narayan Shukla and Benjamin Marlin. Multi-time attention networks for irregularly sampled time series. In *International Conference on Learning Representations*, 2021.

[40] Satya Narayan Shukla and Benjamin M. Marlin. Interpolation-prediction networks for irregularly sampled time series. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9*.

[41] Harpal Singh. The Hålogaland bridge - descriptions, challenges and related research under arctic condition. In *9th International Operational Modal Analysis Conference, IOMAC 2022*, pages 46–60. International Operational Modal Analysis Conference (IOMAC), 2022.

[42] Harpal Singh. *Some New Mathematical and Engineering Results Connected to Structural Problems*. PhD thesis, UiT The Arctic University of Norway, 2022.

[43] Harpal Singh and Niklas Grip. Recent trends in operation modal analysis techniques and its application on a steel truss bridge. *Nonlinear Studies*, 26(4):911–927, 2019.

[44] Harpal Singh, Niklas Grip, and Per Johan Nicklasson. A comprehensive study of signal processing techniques of importance for operation modal analysis (OMA) and its application to a high-rise building. *Nonlinear Studies*, 28(2):389–412, 2021.

[45] Kristoffer Tangrand and Harpal Singh. Analysis of civil engineering infrastructure in norway with solutions based on structural health monitoring and artificial intelligence. 2022. 21 pages, To appear in the journal Nonlinear Studies.

[46] Anne Teughels, Johan Maeck, and Guido De Roeck. Damage assessment by FE model updating using damage functions. *Computers & structures*, 80(25):1869–1879, 2002.

[47] Nicolas Vecoven, Damien Ernst, and Guillaume Drion. A bio-inspired bistable recurrent cell allows for long-lasting memory. *PLoS ONE*, 16, 2021.

[48] Philip B. Weerakody, Kok Wai Wong, Guanjin Wang, and Wendell Ela. A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, 441:161–178, 2021.

[49] Hanne Wilhelms and Gunnar Grindstein. Fare for at bro på E6 i kvænangen kollapser: – bare snakk om tid (in norwegian). https://www.nrk.no, (accessed: 16.08.2022).

[50] Zhenpei Yang et al. SurfelGAN: Synthesizing realistic sensor data for autonomous driving. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11115–11124, Los Alamitos, CA, USA, 2020. IEEE Computer Society.

[51] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[52] Qinghua Zhang and Albert Benveniste. Wavelet networks. *IEEE Trans Neural Networks*, 3(6):889–898, 1992.

[53] Yufei Zhang, Arno Schlüter, and Christoph Waibel. SolarGAN: Synthetic annual solar irradiance time series on urban building facades via deep generative networks, 2022. `"https://arxiv.org/abs/2206.00747"`.

[54] Jingguang Zhou and Zili Huang. Recover missing sensor data with iterative imputing network. In *The Workshops of the The Thirty-Second AAAI Conference on Artificial Intelligence*, volume WS-18 of *AAAI Technical Report*, pages 209–216, 2018.

Paper F

$H_p - L_p$ Type Inequalities for Subsequences of Nörlund Means of Walsh-Fourier Series

David Baramidze, Lars-Erik Persson, Kristoffer Tangrand, George Tephndadze

Submitted.

$(H_p - L_p)$ TYPE INEQUALITIES FOR SUBSEQUENCES OF NÖRLUND MEANS OF WALSH-FOURIER SERIES

DAVID BARAMIDZE[1,2], LARS-ERIK PERSSON[2,3*], KRISTOFFER TANGRAND[2], GEORGE TEPHNADZE[1]

[1] The University of Georgia, School of science and technology, 77a Merab Kostava St, Tbilisi 0128, Georgia.

[2] UiT The Arctic University of Norway, P.O. Box 385, N-8505, Narvik, Norway.

[3] Department of Mathematics and Computer Science, Karlstad University, 65188 Karlstad, Sweden.

* Corresponding author (email: larserik6pers@gmail.com)

Abstract. We investigate subsequence $\{t_{2^n} f\}$ of Nörlund means with respect to the Walsh system generated by non-increasing and convex sequences and prove that some big class of such summability methods are not bounded from the martingale Hardy spaces H_p to the space $weak - L_p$ for $0 < p < 1/(1+\alpha)$, where $0 < \alpha < 1$. Moreover, some new related inequalities are derived. As application, some well-known and new results are pointed out for well-known summability methods-Nörlund logarithmic means and Cesàro means.

2000 Mathematics Subject Classification. 26015, 42C10, 42B30.

Key words and phrases: Walsh system, Nörlund means, Cesàro means, Nörlund logarithmic means, martingale Hardy space, convergence, divergence, inequalities.

1. Introduction

The terminology and notations used in this introduction can be found in Section 2.

It is well-known that Walsh systems do not form bases in the space L_1. Moreover, there is a function in the Hardy space H_1, such that the partial sums of f are not bounded in the L_1-norm. Moreover, (see Tephnadze [26]) there exists a martingale $f \in H_p\,(0 < p < 1)$, such that

$$\sup_{n\in\mathbb{N}} \|S_{2^n+1} f\|_{weak-L_p} = \infty.$$

On the other hand, (for details see e.g. the books [22] and [28]) the subsequence $\{S_{2^n}\}$ of partial sums is bounded from the martingale

Hardy space H_p to the space H_p, for all $p > 0$, that is the following inequality holds:

$$\|S_{2^n} f\|_{H_p} \leq c_p \|f\|_{H_p}, \quad n \in \mathbb{N}, \quad p > 0. \tag{1}$$

Weisz [29] proved that Fejér means of Vilenkin-Fourier series are bounded from the martingale Hardy space H_p to the space H_p, for $p > 1/2$. Goginava [8] (see also [21], [14, 15, 16, 17]) proved that there exists a martingale $f \in H_{1/2}$ such that

$$\sup_{n\in\mathbb{N}} \|\sigma_n f\|_{1/2} = +\infty.$$

However, Weisz [29] (see also [18]) proved that for every $f \in H_p$, there exists an absolute constant c_p, such that the following inequality holds:

$$\|\sigma_{2^n} f\|_{H_p} \leq c_p \|f\|_{H_p}, \quad n \in \mathbb{N}, \quad p > 0. \tag{2}$$

Móricz and Siddiqi [12] investigated the approximation properties of some special Nörlund means of Walsh-Fourier series of L_p functions in norm. Approximation properties for general summability methods can be found in [3, 4]. Fridli, Manchanda and Siddiqi [6] improved and extended the results of Móricz and Siddiqi [12] to martingale Hardy spaces. The case when $\{q_k = 1/k : k \in \mathbb{N}\}$ was excluded, since the methods are not applicable to Nörlund logarithmic means. In [7] Gát and Goginava proved some convergence and divergence properties of the Nörlund logarithmic means of functions in the Lebesgue space L_1. In particular, they proved that there exists an function in the space L_1, such that

$$\sup_{n\in\mathbb{N}} \|L_n f\|_1 = \infty.$$

In [5] (see also [13]) it was proved that there exists a martingale $f \in H_p$, $(0 < p < 1)$ such that

$$\sup_{n\in\mathbb{N}} \|L_{2^n} f\|_p = \infty.$$

Counterexample for $p = 1$ was proved in [20] However, Goginava [9] proved that for every $f \in H_1$, there exists an absolute constant c, such that the inequality holds

$$\|L_{2^n} f\|_1 \leq c \|f\|_{H_1}, \quad n \in \mathbb{N}. \tag{3}$$

In [2] and prove that for any $0 < p < 1$, there exists a martingale $f \in H_p$ such that

$$\sup_{n\in\mathbb{N}} \|L_{2^n} f\|_{weak-L_p} = \infty.$$

In [19] is was proved that for any non-decreasing sequence $(q_k, k \in \mathbb{N})$ satisfying the conditions

$$\frac{1}{Q_n} = O\left(\frac{1}{n^{\alpha}}\right) \quad \text{where} \quad Q_n = \sum_{k=0}^{n-1} q_k \tag{4}$$

$$\text{and} \quad q_n - q_{n+1} = O\left(\frac{1}{n^{2-\alpha}}\right), \quad \text{as} \quad n \to \infty, \tag{5}$$

then, for every $f \in H_p$, where $p > 1/(1+\alpha)$, there exists an absolute constant c_p, depending only on p, such that

$$\|t_n f\|_{H_p} \leq c_p \|f\|_{H_p}, \quad n \in \mathbb{N}. \tag{6}$$

Boundedness does not hold from H_p to $weak-L_p$, for $0 < p < 1/(1+\alpha)$. As a consequence, (for details see [30]) we get that Cesàro means σ_n^{α} is bounded from H_p to L_p, for $p > 1/(1+\alpha)$, but they are not bounded from H_p to $weak - L_p$, for $0 < p < 1/(1+\alpha)$. In the endpoint case $p = 1/(1+\alpha)$, Weisz and Simon [24] proved that the maximal operator $\sigma^{\alpha,*}$ of Cesàro means define by

$$\sigma^{\alpha,*} f := \sup_{n\in\mathbb{N}} |\sigma_n^{\alpha} f|$$

is bounded from the Hardy space $H_{1/(1+\alpha)}$ to the space $weak-L_{1/(1+\alpha)}$. Goginava [10] gave counterexample, which shows that boundedness does not hold for $0 < p \leq 1/(1+\alpha)$.

In this paper we develop some methods considered in [1, 2, 11] and prove that for any $0 < p < 1$, there exists a martingale $f \in H_p$ such that

$$\sup_{n\in\mathbb{N}} \|t_{2^n} f\|_{weak-L_p} = \infty.$$

Moreover, we prove that a class of $\{t_{2^n} f\}$ of Nörlund means with respect to the Walsh system generated by non-increasing and convex sequences are not bounded from the martingale Hardy spaces H_p to the space $weak - L_p$ for $0 < p < 1/(1+\alpha)$, where $0 < \alpha < 1$. Moreover, some new related inequalities are derived. As application, some well-known and new results are pointed out for well-known summability methods-Nörlund logarithmic means and Cesàro means.

The main results in this paper are presented and proved in Section 4. Section 3 is used to present some auxiliary lemmas, where, in particular, Lemma 2 is new and of independent interest. In order not to disturb our discussions later on some definitions and notations are given in Section 4.

2. Definitions and Notations

Let $\mathbb{N}_+$ denote the set of the positive integers, $\mathbb{N} := \mathbb{N}_+ \cup \{0\}$. Denote by Z_2 the discrete cyclic group of order 2, that is $Z_2 := \{0, 1\}$, where the group operation is the modulo 2 addition and every subset is open. The Haar measure on Z_2 is given so that the measure of a singleton is $1/2$.

Define the group G as the complete direct product of the group Z_2, with the product of the discrete topologies of Z_2's.

The elements of G are represented by sequences

$$x := (x_0, x_1, ..., x_j, ...), \quad \text{where} \quad x_k = 0 \vee 1.$$

It is easy to give a base for the neighborhood of $x \in G$ namely:

$$I_0(x) := G, \quad I_n(x) := \{y \in G : y_0 = x_0, ..., y_{n-1} = x_{n-1}\} \ (n \in \mathbb{N}).$$

Denote $I_n := I_n(0)$, $\overline{I_n} := G \setminus I_n$ and

$$e_n := (0, ..., 0, x_n = 1, 0, ...) \in G, \quad \text{for } n \in \mathbb{N}.$$

If $n \in \mathbb{N}$, then every n can be uniquely expressed as $n = \sum_{k=0}^{\infty} n_j 2^j$, where $n_j \in Z_2$ $(j \in \mathbb{N})$ and only a finite numbers of n_j differ from zero. Let

$$|n| := \max\{k \in \mathbb{N} : \ n_k \neq 0\}.$$

The norms (or quasi-norms) of the spaces $L_p(G)$ and $weak - L_p(G)$, $(0 < p < \infty)$ are, respectively, defined by

$$\|f\|_p^p := \int_G |f|^p \, d\mu \quad \text{and} \quad \|f\|_{weak-L_p}^p := \sup_{\lambda>0} \lambda^p \mu(f > \lambda).$$

The k-th Rademacher function is defined by

$$r_k(x) := (-1)^{x_k} \qquad (x \in G, \ k \in \mathbb{N}).$$

Now, define the Walsh system $w := (w_n : n \in \mathbb{N})$ on G as:

$$w_n(x) := \prod_{k=0}^{\infty} r_k^{n_k}(x) = r_{|n|}(x)(-1)^{\sum_{k=0}^{|n|-1} n_k x_k} \qquad (n \in \mathbb{N}).$$

It is well-known that (see e.g. [22]) Walsh system is orthonormal and complete in $L_2(G)$. Moreover, for any $n \in \mathbb{N}$,

$$w_n(x+y) = w_n(x) w_n(y). \tag{7}$$

If $f \in L_1(G)$ let us define Fourier coefficients, partial sums and Dirichlet kernel by

$$\begin{aligned} \widehat{f}(k) &:= \int_G f w_k d\mu \quad (k \in \mathbb{N}), \\ S_n f &:= \sum_{k=0}^{n-1} \widehat{f}(k) w_k, \quad D_n := \sum_{k=0}^{n-1} w_k \quad (n \in \mathbb{N}_+). \end{aligned}$$

Recall that (for details see e.g. [22]):

$$D_{2^n}(x) = \begin{cases} 2^n, & \text{if } x \in I_n \\ 0, & \text{if } x \notin I_n \end{cases} \tag{8}$$

and

$$D_n = w_n \sum_{k=0}^{\infty} n_k r_k D_{2^k} = w_n \sum_{k=0}^{\infty} n_k \left(D_{2^{k+1}} - D_{2^k}\right), \text{ for } n = \sum_{i=0}^{\infty} n_i 2^i. \tag{9}$$

Let $\{q_k,\ k \geq 0\}$ be a sequence of nonnegative numbers. The Nörlund means for the Fourier series of f are defined by

$$t_n f := \frac{1}{Q_n} \sum_{k=1}^{n} q_{n-k} S_k f, \quad \text{where} \quad Q_n := \sum_{k=0}^{n-1} q_k.$$

In this paper we consider convex $\{q_k,\ k \geq 0\}$ sequences, that is

$$q_{n-1} + q_{n+1} - 2q_n \geq 0, \quad \text{for all} \quad n \in \mathbb{N}.$$

If function $\psi(x)$ is any real valued and convex function (for example $\psi(x) = x^{\alpha-1},\ 0 \leq \alpha \leq 1$), then sequence $\{\psi(n),\ n \in \mathbb{N}\}$ is convex.

Since

$$q_{n-2} - q_{n-1} \geq q_{n-1} - q_n \geq q_n - q_{n+1} \geq q_{n+1} - q_{n+2}$$

we find that

$$q_{n-2} + q_{n+2} \geq q_{n-1} + q_{n+1}$$

an we also get that

$$q_{n-2} + q_{n+2} - 2q_n \geq 0, \quad \text{for all} \quad n \in \mathbb{N}. \tag{10}$$

In the special case when $\{q_k = 1,\ k \in \mathbb{N}\}$, we get the Fejér means

$$\sigma_n f := \frac{1}{n} \sum_{k=1}^{n} S_k f.$$

If $q_k = 1/(k+1)$, then we get the Nörlund logarithmic means:

$$L_n f := \frac{1}{l_n} \sum_{k=1}^{n} \frac{S_k f}{n+1-k}, \qquad \text{where} \quad l_n := \sum_{k=1}^{n} \frac{1}{k}. \tag{11}$$

The Cesàro means σ_n^α (sometimes also denoted (C,α)) is also well-known example of Nörlund means defined by

$$\sigma_n^\alpha f =: \frac{1}{A_n^\alpha}\sum_{k=1}^{n} A_{n-k}^{\alpha-1} S_k f$$

where

$$A_0^\alpha := 0, \qquad A_n^\alpha := \frac{(\alpha+1)\dots(\alpha+n)}{n!}, \qquad \alpha \neq -1,-2,\dots$$

It is well-known that

$$A_n^\alpha = \sum_{k=0}^{n} A_{n-k}^{\alpha-1}, \quad A_n^\alpha - A_{n-1}^\alpha = A_n^{\alpha-1} \quad \text{and} \quad A_n^\alpha \sim n^\alpha. \tag{12}$$

We also define U_n^α means as

$$U_n^\alpha f := \frac{1}{Q_n}\sum_{k=1}^{n}(n+1-k)^{(\alpha-1)} S_k f \quad \text{where} \quad Q_n := \sum_{k=1}^{n} k^{\alpha-1}.$$

Let us define V_n^α means as

$$V_n f := \frac{1}{Q_n}\sum_{k=1}^{n}\ln(n+1-k) S_k f \quad \text{where} \quad Q_n := \sum_{k=1}^{n}\frac{1}{\ln(k+1)}.$$

Let $f := \left(f^{(n)}, n \in \mathbb{N}\right)$ be a martingale with respect to $F_n\,(n \in \mathbb{N})$, which are generated by the intervals $\{I_n(x) : x \in G\}$ (for details see e.g. [28]).

We say that martingale belongs to Hardy martingale spaces $H_p(G)$ where $0 < p < \infty$ if

$$\|f\|_{H_p} := \|f^*\|_p < \infty, \quad \text{where} \quad f^* := \sup_{n\in\mathbb{N}} \left|f^{(n)}\right|.$$

In case $f \in L_1(G)$, the maximal functions are also be given by

$$M(f)(x) := \sup_{n\in\mathbb{N}} \left(\frac{1}{\mu(I_n(x))}\left|\int_{I_n(x)} f(u)\,d\mu(u)\right|\right).$$

If $f \in L_1(G)$, then it is easy to show that the sequence $F = (S_{2^n} f : n \in \mathbb{N})$ is a martingale and $F^* = M(f)$.

If $f = \left(f^{(n)},\ n \in \mathbb{N}\right)$ is a martingale, then the Walsh-Fourier coefficients must be defined in a slightly different manner:

$$\widehat{f}(i) := \lim_{k\to\infty}\int_G f^{(k)}(x)\, w_i(x)\, d\mu(x).$$

A bounded measurable function a is p-atom, if there exists an interval I, such that

$$\operatorname{supp}(a) \subset I, \quad \int_I a d\mu = 0, \quad \|a\|_\infty \leq \mu(I)^{-1/p}.$$

3. Auxiliary Results

The Hardy martingale space $H_p(G)$ has an atomic characterization (see Weisz [28], [29]):

Lemma 1. *A martingale $f = \left(f^{(n)},\ n \in \mathbb{N}\right)$ is in $H_p\,(0 < p \leq 1)$ if and only if there exist a sequence $(a_k, k \in \mathbb{N})$ of p-atoms and a sequence $(\mu_k, k \in \mathbb{N})$ of real numbers such that for every $n \in \mathbb{N}$:*

$$\sum_{k=0}^{\infty} \mu_k S_{2^n} a_k = f^{(n)}, \qquad \text{where} \qquad \sum_{k=0}^{\infty} |\mu_k|^p < \infty. \tag{13}$$

Moreover,

$$\|f\|_{H_p} \backsim \inf \left(\sum_{k=0}^{\infty} |\mu_k|^p \right)^{1/p},$$

where the infimum is taken over all decompositions of f of the form (13).

We also state and prove new lemma of independent interest:

Lemma 2. *Let $k \in \mathbb{N}$, $\{q_k : k \in \mathbb{N}\}$ is any convex and non-increasing sequence and $x \in I_2(e_0 + e_1) \in I_0 \backslash I_1$. Then*

$$\left| \sum_{j=2^{2\alpha_k}}^{2^{2\alpha_k+1}} q_{2^{2\alpha_k+1}-j} D_j \right| \geq q_1 - \frac{3}{2} q_3.$$

Proof. Let $x \in I_2(e_0+e_1) \in I_0 \backslash I_1$. According to (8) and (9) we get that

$$D_j(x) = \begin{cases} w_j, & \text{if } j \text{ is odd number,} \\ 0, & \text{if } j \text{ is even number,} \end{cases}$$

and

$$\begin{aligned} \sum_{j=2^{2\alpha_k}}^{2^{2\alpha_k+1}-1} q_{2^{2\alpha_k+1}-j} D_j &= \sum_{j=2^{2\alpha_k-1}}^{2^{2\alpha_k}-1} q_{2^{2\alpha_k+1}-2j-1} w_{2j+1} \\ &= w_1 \sum_{j=2^{2\alpha_k-1}}^{2^{2\alpha_k}-1} q_{2^{2\alpha_k+1}-2j-1} w_{2j}. \end{aligned}$$

By using (10) we find that

$$\begin{aligned}
&\sum_{j=2^{2\alpha_k-2}+1}^{2^{2\alpha_k-1}-1}\left|q_{2^{2\alpha_k+1}-4j+3}-q_{2^{2\alpha_k+1}-4j+1}\right|\\
=&\sum_{j=2^{2\alpha_k-2}+1}^{2^{2\alpha_k-1}-1}\left(q_{2^{2\alpha_k+1}-4j+1}-q_{2^{2\alpha_k+1}-4j+3}\right)\\
=&\ (q_{2^{2\alpha_k-3}}-q_{2^{2\alpha_k-1}})+(q_{2^{2\alpha_k-7}}-q_{2^{2\alpha_k-5}})+\ldots+(q_5-q_7)\\
\leq&\ \frac{1}{2}\left(q_{2^{2\alpha_k-3}}-q_{2^{2\alpha_k-1}}\right)+\frac{1}{2}\left(q_{2^{2\alpha_k-5}}-q_{2^{2\alpha_k-3}}\right)\\
+&\ \frac{1}{2}\left(q_{2^{2\alpha_k-7}}-q_{2^{2\alpha_k-5}}\right)+\frac{1}{2}\left(q_{2^{2\alpha_k-9}}-q_{2^{2\alpha_k-7}}\right)\\
+&\ \ldots+\frac{1}{2}\left(q_5-q_7\right)+\frac{1}{2}\left(q_3-q_5\right)\\
\leq&\ \frac{1}{2}q_3-\frac{1}{2}q_{2^{2\alpha_k-1}}.
\end{aligned}$$

Hence, if we apply

$$w_{4k+2}=w_2w_{4k}=-w_{4k},\quad \text{for}\quad x\in I_2(e_0+e_1),$$

we find that

$$\begin{aligned}
&\left|\sum_{j=2^{2\alpha_k}}^{2^{2\alpha_k+1}-1}q_{2^{2\alpha_k+1}-j}D_j\right|\\
=&\left|q_0w_{2^{2\alpha_k+1}-2}+q_3w_{2^{2\alpha_k+1}-4}+\sum_{j=2^{2\alpha_k-1}}^{2^{2\alpha_k}-1}q_{2^{2\alpha_k+1}-2j-1}w_{2j}\right|\\
=&\left|(q_3-q_1)2w_{2^{2\alpha_k+1}-4}+\sum_{j=2^{2\alpha_k-2}+1}^{2^{2\alpha_k-1}}\left(q_{2^{2\alpha_k+1}-4j+3}w_{4j-4}-q_{2^{2\alpha_k+1}-4j+1}w_{4j-4}\right)\right|\\
\geq&\ q_1-q_3-\sum_{j=2^{2\alpha_k-2}+1}^{2^{2\alpha_k-1}}\left|q_{2^{2\alpha_k+1}-4j+3}-q_{2^{2\alpha_k+1}-4j+1}\right|\\
\geq&\ q_1-q_3-\frac{1}{2}(q_3-q_{2^{2\alpha_k-1}})\\
\geq&\ q_1-(3/2)q_3.
\end{aligned}$$

The proof is complete. □

4. Main results

Our first main result reads:

Theorem 1. *Let $0 \leq \alpha \leq 1$, β be any non-negative real number and t_n be Nörlund means with convex and non-increasing sequence $\{q_k : k \in \mathbb{N}\}$ satisfying condition*

$$\frac{q_1 - (3/2)q_3}{Q_n} \geq \frac{C}{n^{\alpha} \log^{\beta} n}, \tag{14}$$

for some positive constant C. Then, for any $0 < p < 1/(1+\alpha)$ there exists a martingale $f \in H_p$ such that

$$\sup_{n\in\mathbb{N}} \|t_{2^n} f\|_{weak-L_p} = \infty.$$

Proof. Let $0 < p < 1/(1+\alpha)$. Under condition (14) there exists sequence $\{n_k : k \in \mathbb{N}\}$ such that

$$\frac{2^{2n_k(1/p-1)}}{n_k Q_{2^{2n_k+1}}} \geq \frac{2^{2n_k(1/p-1-\alpha)}}{n_k^{\beta+1}} \to \infty, \quad \text{as} \quad k \to \infty.$$

Let $\{\alpha_k : k \in \mathbb{N}\} \subset \{n_k : k \in \mathbb{N}\}$ be an increasing sequence of the positive integers such that

$$\sum_{k=0}^{\infty} \alpha_k^{-p/2} < \infty, \tag{15}$$

$$\sum_{\eta=0}^{k-1} \frac{\left(2^{2\alpha_\eta}\right)^{1/p}}{\sqrt{\alpha_\eta}} < \frac{\left(2^{2\alpha_k}\right)^{1/p}}{\sqrt{\alpha_k}} \tag{16}$$

and

$$\frac{\left(2^{2\alpha_{k-1}}\right)^{1/p}}{\sqrt{\alpha_{k-1}}} < \frac{q_1 - q_3 - (3/2)q_5}{Q_{2^{2\alpha_k+1}}} \frac{2^{2\alpha_k(1/p-1)-3}}{\alpha_k}. \tag{17}$$

Let

$$f^{(n)} := \sum_{\{k;\ 2\alpha_k < n\}} \lambda_k a_k,$$

where

$$\lambda_k = \frac{1}{\sqrt{\alpha_k}} \quad \text{and} \quad a_k = 2^{2\alpha_k(1/p-1)} \left(D_{2^{2\alpha_k+1}} - D_{2^{2\alpha_k}}\right).$$

From (15) and Lemma 1 we find that $f \in H_p$.

It is easy to show that

$$\widehat{f}(j)=\begin{cases}\frac{2^{2\alpha_k(1/p-1)}}{\sqrt{\alpha_k}}, & \text{if } j\in\{2^{2\alpha_k},...,\ 2^{2\alpha_k+1}-1\},\ k\in\mathbb{N},\\ 0, & \text{if } j\notin\bigcup\limits_{k=1}^{\infty}\{2^{2\alpha_k},...,\ 2^{2\alpha_k+1}-1\}\ .\end{cases}\tag{18}$$

Moreover,

$$\begin{aligned}&t_{2^{2\alpha_k+1}}f\\ &=\frac{1}{Q_{2^{2\alpha_k+1}}}\sum_{j=1}^{2^{2\alpha_k}-1}q_{2^{2\alpha_k+1}-j}S_jf+\frac{1}{Q_{2^{2\alpha_k+1}}}\sum_{j=2^{2\alpha_k}}^{2^{2\alpha_k+1}-1}q_{2^{2\alpha_k+1}-j}S_jf\\ &:=I+II.\end{aligned}\tag{19}$$

Let $j<2^{2\alpha_k}$. By combining (16), (17) and (18) we can conclude that

$$\begin{aligned}|S_jf(x)| &\le \sum_{\eta=0}^{k-1}\sum_{v=2^{2\alpha_\eta}}^{2^{2\alpha_\eta+1}-1}\left|\widehat{f}(v)\right|\\ &\le \sum_{\eta=0}^{k-1}\sum_{v=2^{2\alpha_\eta}}^{2^{2\alpha_\eta+1}-1}\frac{2^{2\alpha_\eta(1/p-1)}}{\sqrt{\alpha_\eta}}\le\sum_{\eta=0}^{k-1}\frac{2^{2\alpha_\eta/p}}{\sqrt{\alpha_\eta}}\le\frac{2^{2\alpha_{k-1}/p+1}}{\sqrt{\alpha_{k-1}}}.\end{aligned}$$

Hence,

$$\begin{aligned}|I| &\le \frac{1}{Q_{2^{2\alpha_k+1}}}\sum_{j=1}^{2^{2\alpha_k}-1}q_{2^{2\alpha_k+1}-j}|S_jf(x)|\\ &\le \frac{1}{Q_{2^{2\alpha_k+1}}}\frac{2^{2\alpha_{k-1}/p}}{\sqrt{\alpha_{k-1}}}\sum_{j=1}^{M_{2\alpha_k+1}-1}q_j\le\frac{2^{2\alpha_{k-1}/p}}{\sqrt{\alpha_{k-1}}}.\end{aligned}\tag{20}$$

Let $2^{2\alpha_k}\le j\le 2^{2\alpha_k+1}-1$. Since

$$\begin{aligned}S_jf &= \sum_{\eta=0}^{k-1}\sum_{v=2^{2\alpha_\eta}}^{2^{2\alpha_\eta+1}-1}\widehat{f}(v)w_v+\sum_{v=2^{2\alpha_k}}^{j-1}\widehat{f}(v)w_v\\ &= \sum_{\eta=0}^{k-1}\frac{2^{2\alpha_\eta(1/p-1)}}{\sqrt{\alpha_\eta}}\left(D_{2^{2\alpha_\eta+1}}-D_{2^{2\alpha_\eta}}\right)+\frac{2^{2\alpha_k(1/p-1)}}{\sqrt{\alpha_k}}\left(D_j-D_{2^{2\alpha_k}}\right),\end{aligned}$$

for II we can conclude that

$$(21)\ II=\frac{1}{Q_{2^{2\alpha_k+1}}}\sum_{j=2^{2\alpha_k}}^{2^{2\alpha_k+1}}q_{2^{2\alpha_k+1}-j}\left(\sum_{\eta=0}^{k-1}\frac{2^{2\alpha_\eta(1/p-1)}}{\sqrt{\alpha_\eta}}\left(D_{2^{2\alpha_\eta+1}}-D_{2^{2\alpha_\eta}}\right)\right)$$
$$+\frac{1}{Q_{2^{2\alpha_k+1}}}\frac{2^{2\alpha_k(1/p-1)}}{\sqrt{\alpha_k}}\sum_{j=2^{2\alpha_k}}^{2^{2\alpha_k+1}-1}q_{2^{2\alpha_k+1}-j}(D_j-D_{2^{2\alpha_k}}).$$

Let $x\in I_2(e_0+e_1)\in I_0\backslash I_1$. According to $\alpha_0\geq 1$ we get that $2\alpha_k\geq 2$, for all $k\in\mathbb{N}$ and if use (8) we get that $D_{2^{2\alpha_k}}=0$ and if we use Lemma 2 we can also conclude that

$$(22)\qquad II = \frac{1}{Q_{2^{2\alpha_k+1}}}\frac{2^{2\alpha_k(1/p-1)}}{\sqrt{\alpha_k}}\sum_{j=2^{2\alpha_k}}^{2^{2\alpha_k+1}-1}q_{2^{2\alpha_k+1}-j}D_j$$
$$\geq \frac{q_1-(3/2)q_3}{Q_{2^{2\alpha_k+1}}}\frac{2^{2\alpha_k(1/p-1)}}{\sqrt{\alpha_k}}.$$

By combining (17), (19)-(22) for $x\in I_2(e_0+e_1)$ we have that

$$\begin{aligned}|t_{2^{2\alpha_k+1}}f(x)| &\geq II-I\\ &\geq \frac{q_1-(3/2)q_3}{Q_{2^{2\alpha_k+1}}}\frac{2^{2\alpha_k(1/p-1)}}{\sqrt{\alpha_k}}-\frac{q_1-(3/2)q_3}{Q_{2^{2\alpha_k+1}}}\frac{2^{2\alpha_k(1/p-1)-3}}{\alpha_k}\\ &\geq \frac{q_1-(3/2)q_3}{Q_{2^{2\alpha_k+1}}}\frac{2^{2\alpha_k(1/p-1)-3}}{\sqrt{\alpha_k}}\geq\frac{C2^{2\alpha_k(1/p-1-\alpha)-3}}{\left(\ln 2^{2\alpha_k+1}+1\right)^{\beta}\sqrt{\alpha_k}}\\ &\geq \frac{C2^{2\alpha_k(1/p-1-\alpha)-3}}{\alpha_k^{\beta+1}}.\end{aligned}$$

Hence, we can conclude that

$$\begin{aligned}&\|t_{2^{2\alpha_k+1}}f\|_{weak-L_p}\\ \geq\ &\frac{C2^{2\alpha_k(1/p-1-\alpha)-3}}{\alpha_k^{\beta+1}}\mu\left\{x\in G:|t_{2^{2\alpha_k+1}}f|\geq\frac{C2^{2\alpha_k(1/p-1)-3}}{\alpha_k^{\beta+1}}\right\}^{1/p}\\ \geq\ &\frac{C2^{2\alpha_k(1/p-1-\alpha)-3}}{\alpha_k^{\beta+1}}\mu\left\{x\in I_2(e_0+e_1):|t_{2^{2\alpha_k+1}}f|\geq\frac{C2^{2\alpha_k(1/p-1)-6}}{\alpha_k^{\beta+1}}\right\}^{1/p}\\ \geq\ &\frac{C2^{2\alpha_k(1/p-1-\alpha)-3}}{\alpha_k^{\beta+1}}\left(\mu\left(I_2(e_0+e_1)\right)\right)^{1/p}\\ >\ &\frac{c2^{2\alpha_k(1/p-1-\alpha)}}{\alpha_k^{\beta+1}}\to\infty,\quad \text{as}\quad k\to\infty.\end{aligned}$$

The proof is complete. □

In the concrete case we get result for Nörlund logarithmic means $\{L_n\}$ proved in [2]:

Corollary 1. *Let $0 < p < 1$. Then there exists a martingale $f \in H_p$ such that*

$$\sup_{n\in\mathbb{N}} \|L_{2^n} f\|_{weak-L_p} = \infty.$$

Proof. It is easy to show that

$$q_1 - (3/2)q_3 = \frac{1}{2} - \frac{3}{2}\cdot\frac{1}{4} = \frac{1}{8} > 0,$$

and condition (14) holds true for $\alpha = \beta = 0$. □

We also get similar result for V_n means:

Corollary 2. *Let $0 < p < 1$. Then there exists a martingale $f \in H_p$ such that*

$$\sup_{n\in\mathbb{N}} \|V_{2^n} f\|_{weak-L_p} = \infty.$$

Proof. It is easy to show that

$$q_1 - (3/2)q_3 = \frac{1}{\ln 2} - \frac{3}{2}\cdot\frac{1}{\ln 4} = \log_2^e - (3/2)\frac{\log_2^e}{\log_2^4} = \log_2^e\left(1 - \frac{3}{4}\right) > 0,$$

and condition (14) holds true for $\alpha = \beta = 0$. □

We also get some new result for the Cesàro means:

Corollary 3. *Let $0 < p < 1/(1+\alpha)$, for some $0 < \alpha \le 0.56$. Then there exists a martingale $f \in H_p$ such that*

$$\sup_{n\in\mathbb{N}} \|\sigma^{\alpha}_{2^n} f\|_{weak-L_p} = \infty.$$

Proof. By routine calculation we find that

$$q_1 - (3/2)q_3 = \alpha - \frac{\alpha(\alpha+1)(\alpha+2)}{4} = \alpha\cdot\frac{2 - 3\alpha - \alpha^2}{4}.$$

It is easy to show that when $0 < \alpha < 0.56$ this expression is positive. Hence, condition (14) holds true for $\beta = 0$ and $0 < \alpha < 1$. □

Corollary 4. *Let $0 < p < 1/(1+\alpha)$, for some $0 < \alpha \le 0.41$. Then there exists a martingale $f \in H_p$ such that*

$$\sup_{n\in\mathbb{N}} \|U^{\alpha}_{2^n} f\|_{weak-L_p} = \infty.$$

Proof. By routine calculation we find that

$$q_1 - (3/2)q_3 = 2^{1-\alpha} - (3/2)4^{1-\alpha} = 2^{1-\alpha}\left(1 - 3/2^{2-\alpha}\right).$$

It is easy to show that when $0 < \alpha < 0.41$ this expression is positive. Hence, condition (14) holds true for $\beta = 0$ and $0 < \alpha < 1$. □

5. Open questions

Open Problem 1: Let $0 < p < 1/(1+\alpha)$, for some $0.56 < \alpha < 1$. Then there exists a martingale $f \in H_p$ such that

$$\sup_{n\in\mathbb{N}} \|\sigma^{\alpha}_{2^n} f\|_{weak-L_p} = \infty$$

Open Problem 2: Let $0 < p < 1/(1+\alpha)$, for some $0.41 < \alpha < 1$. Then there exists a martingale $f \in H_p$ such that

$$\sup_{n\in\mathbb{N}} \|U^{\alpha}_{2^n} f\|_{weak-L_p} = \infty.$$

We also can investigate similar problems for more general summability methods:

Open Problem 3: Let $0 < p < 1/(1+\alpha)$, for some $0.56 < \alpha < 1$ and t_n be Nörlund means of Walsh-Fourier series with non-increasing and convex sequence $\{q_k : k \in \mathbb{N}\}$, satisfying the condition (14).

Does there exist a martingale $f \in H_{1/(1+\alpha)} (0 < p < 1)$, such that

$$\sup_{n\in\mathbb{N}} \|t_{2^n} f\|_{H_{1/(1+\alpha)}} = \infty?$$

Availability of data and material

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

The publication charges for this manuscript is supported by the publication fund at UiT The Arctic University of Norway under code IN-1096130.

Authors' contributions

DB and GT gave the idea and initiated the writing of this paper. LEP and KT followed up this with some complementary ideas. All authors read and approved the final manuscript.

Acknowledgements

The work of George Tephnadze was supported by Shota Rustaveli National Science Foundation grant FR-19-676. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway. The authors also would like to thank the two referees for helpful suggestions.

References

[1] *L. Baramidze, L. E. Persson, G. Tephnadze and P. Wall,* Sharp $H_p - L_p$ type inequalities of weighted maximal operators of Vilenkin-Nörlund means and its applications, J. Inequal. Appl., 2016, DOI: 10.1186/s13660-016-1182-1.

[2] *D. Baramidze, L.-E. Persson, H. Singh and G. Tephnadze,* Some new results and inequalities for subsequences of Nörlund logarithmic means of Walsh-Fourier series, J. Inequal. Appl., 2022, DOI: https://doi.org/10.1186/s13660-022-02765-5.

[3] *I. Blahota, K. Nagy and G. Tephnadze,* Approximation by Θ-Means of Walsh-Fourier Series, Anal. Math., 44 (1), 57-71.

[4] *I. Blahota, K. Nagy and G. Tephnadze,* Approximation by Marcinkiewicz Θ-means of double Walsh-Fourier series, Math. Inequal. Appl., 22 (2019), no. 3, 837–853.

[5] *I. Blahota, L.-E. Persson and G Tephnadze,* On the Nörlund means of Vilenkin-Fourier series, Czech. Math. J., 65 (4), 983-1002.

[6] *S. Fridli, P. Manchanda and A.H. Siddiqi,* Approximation by Walsh-Nörlund means, Acta Sci. Math.(Szeged) 74 (2008), no. 3-4, 593-608.

[7] *G. Gàt and U. Goginava,* Uniform and L-convergence of logarithmic means of Walsh-Fourier series, Acta Math. Sin. 22 (2006), no. 2, 497–506.

[8] *U. Goginava,* The maximal operator of the (C, α) means of the Walsh-Fourier series, Ann. Univ. Sci. Budapest. Sect. Comput. 26 (2006), 127–135.

[9] *U. Goginava,* Almost everywhere convergence of subsequence of logarithmic means of Walsh-Fourier series, Acta Math. Paed. Nyíreg., 21 (2005), 169-175.

[10] *U. Goginava,* The maximal operator of the (C, α) means of the Walsh-Fourier series, Ann. Univ. Sci. Budapest. Sect. Comput., 26 (2006), 127–135.

[11] *D. Lukkassen, L.E. Persson, G. Tephnadze and G. Tutberidze,* Some inequalities related to strong convergence of Riesz logarithmic means of Vilenkin-Fourier series, J. Inequal. Appl., 2020, DOI: https://doi.org/10.1186/s13660-020-02342-8.

[12] *F. Móricz and A. Siddiqi*, Approximation by Nörlund means of Walsh-Fourier series, J. Approx. Theory 70 (1992), no. 3, 375–389.

[13] *N. Memiœ, L. E. Persson and G. Tephnadze,* A note on the maximal operators of Vilenkin-Nörlund means with non-increasing coefficients, Stud. Sci. Math. Hung., 53, 4, (2016) 545-556.

[14] *K. Nagy and G. Tephnadze,* On the Walsh-Marcinkiewicz means on the Hardy space, Cent. Eur. J. Math., 12, 8 (2014), 1214-1228.

[15] *K. Nagy and G. Tephnadze,* Kaczmarz-Marcinkiewicz means and Hardy spaces, Acta math. Hung., 149, 2 (2016), 346-374.

[16] *K. Nagy and G. Tephnadze,* Strong convergence theorem for Walsh-Marcinkiewicz means, Math. Inequal. Appl., 19, 1 (2016), 185-195.

[17] *K. Nagy and G. Tephnadze,* Approximation by Walsh-Marcinkiewicz means on the Hardy space , Kyoto J. Math., 54, 3 (2014), 641-652.

[18] *L. E. Persson and G. Tephnadze,* A sharp boundedness result concerning some maximal operators of Vilenkin-Fejér means, Mediterr. J. Math., 13, 4 (2016) 1841-1853.

[19] *L. E. Persson, G. Tephnadze, P. Wall,* On the maximal operators of Vilenkin-Nörlund means, J. Fourier Anal. Appl., 21, 1 (2015), 76-94.

[20] *L. E. Persson, G. Tephnadze and P. Wall,* On the Nörlund logarithmic means with respect to Vilenkin system in the martingale Hardy space H_1, Acta Math. Hung., 154 (2018), no 2, 289-301.

[21] *L. E. Persson, G. Tephnadze and G. Tutberidze,* On the boundedness of subsequences of Vilenkin-Fejér means on the martingale Hardy spaces, Operators and matrices, 14 (2020), no. 1, 283-294.

[22] *F. Schipp, W.R. Wade, P. Simon and J. Pál,* Walsh series, An Introduction to Dyadic Harmonic Analysis, Akadémiai Kiadó, (Budapest-Adam-Hilger (Bristol-New-York)), 1990.

[23] *P. Simon,* Strong Convergence Theorem for Vilenkin-Fourier Series, J. Math. Anal. Appl., 245 (2000), 52-68.

[24] *P. Simon and F. Weisz,* Weak inequalities for Cesáro and Riesz summability of Walsh-Fourier series, J. Approx. Theory, 151 (2008), no. 1, 1–19.

[25] *G. Tephnadze,* The maximal operators of logarithmic means of one-dimensional Vilenkin-Fourier series, Acta Math. Acad. Paed. Nyíreg., 27 (2011), 245-256.

[26] *G. Tephnadze,* On the partial sums of Vilenkin-Fourier series, J. Contemp. Math. Anal. 49 (2014), no. 1, 23-32.

[27] *G. Tephnadze and G. Tutberidze,* A note on the maximal operators of the Nörlund logaritmic means of Vilenkin-Fourier series, Trans. A. Razmadze Math. Inst., 174 (2020), no. 1, 1070-112.

[28] *F. Weisz,* Martingale Hardy Spaces and their Applications in Fourier Analysis, Springer, Berlin-Heidelberg-New York, 1994.

[29] *F. Weisz,* Hardy spaces and Cesàro means of two-dimensional Fourier series, Bolyai Soc. Math. Studies, (1996), 353-367.

[30] *F. Weisz,*, (C, α) summability of Walsh-Fourier series, Anal. Math., 27 (2001), 141-156.