Chapter 1

Introduction

1.1 Intro

Since the early years of deep learning development, researchers have tried to provide a coherent and working extent of those techniques and methodologies also in the complex domain. The first motivation was obviously that it appeared as the most "natural" extent of ordinary machine learning (literally, as Neuronal Synchrony [**?**] guarantees also a biological interpretation), but also thanks to more recent works and fundamental theoretical analyses, suggesting that complex numbers could have richer expressiveness: many deterministic wave signals, such as seismic, electrical, or vibrational, in fact, contained information in their phase, which risked being lost when studied using a real-valued model. In the last thirty years, however, despite the large amount of papers published in this sense, proposing the most disparate theories, the research world still missed a definite complex-valued deep learning framework. The motivations of this very slow growth are various, from the theoretical difficulties of re-interpreting the properties of the complex plane in this sense, to the big shortcomings in the support of complex numbers by hardware accelerators. And all of this still holds, even tho complex data arise in various practical contexts, such as medicine (e.g. Magnetic Resonance Images), communication areas (audio and radar signals) and industrial applications (vibration signals). Furthermore, thanks to a powerful instrument like the Fourier transform, we can study time series in the momentum space, which is inherently complex-valued.

In this thesis, we want to recover the most promising extents of deep learning techniques and operators to the complex domain, re-organize them and build a complex-valued deep learning framework that is both efficient and rigorous from a mathematical point of view. After a first part in which we discuss how to define structures like neurons and layers in the complex domain, considering also an alternative version of the backpropagation algorithm, we go through a practical implementation of the problem: in order to overcome the limited support of complex operators by existing deep learning frameworks (like Tensorflow and Pytorch), we decided to write our own Python library. The code is actually still quite raw, and supports only the most fundamental operations in deep learning, but is written on top of JAX, a Python library recently developed by Google DeepMind, that provides such an high level of optimization that we can partially fulfill the lack of hardware acceleration.

After the technical part, we proceed with the practical implementation and training of complex-valued models on simple datasets, focusing also on the comparison, in term of performances, with equivalent real-valued architectures. The conventional approach foresees, in presence of complex inputs, to drop the phase information in data, keeping only their magnitude, or to split them into real and imaginary parts, considered now as two independent channels: we will see that ignoring inherent relationships in complex data can drastically affect the training behavior in our models, making also an interesting study on the circularity property, peculiar of complex random variables. We will also discuss about the apparent "higher robustness" of complex-valued models to overfitting.

In the end, we will treat the problem of complex-valued deep learning in industrial applications considering the task of condition monitoring: we will show that complex models can effectively work with real world problems, guaranteeing also very nice performances. For the first time, in fact, we

apply complex neural networks in a task that foresees the classification of vibration signals, that effectively, never appeared in the numerous works published in recent years. Furthermore, as final analysis, we will demonstrate that, at least qualitatively, Wasserstein Guided Representation Learning can be a suitable transfer learning algorithm for complex-valued data: if our results and hypothesis would be confirmed by successive studies in this sense, this could represent an important starting point of many future works, since transfer learning is still a widely unexplored brach of complex-valued deep learning.

1.2 Previous work

Several attempts have been performed during the years, in order to develop the complex framework we are looking for. The very first trial of developing a complex backpropagation algorithm even dates back to 1991, by Nitta [?]. This extension was anything but straightforward: moving into the complex domain, in fact, you encounter a notion of differentiability that is far stronger than its counterpart in the real space. And the presence of non-holomorphic functions (non-differentiable, in the complex sense) was a huge problem. He was the first, then, to introduce the idea of exploiting Wirtinger derivatives to limit the large operational complexity required to compute backpropagation of non-holomorphic functions. Several other works ([?, ?, ?]) have succeeded this one, almost all based on Wirtinger calculus, each one adding a slight improvement to the algorithm.

From now on, many papers were published, in order to contribute building this new deep learning area based on complex numbers: from proposing complex multi-layer perceptrons [?], to whole new deep learning models [?], to simple new activation functions [?, ?], suitable to work in this new domain. In particular, a certain attention should be deserved by complex-valued convolutional neural networks [?], that, according to Guberman [?], are the only models able to detect meaningful phase structures in the data.

Even if there is no definitive proof that complex models are better than real, more and more clues are being collected in the most various areas: complex-valued neural networks have been successfully applied to seismic data [?], non-linear signal processing [?], MRI classification [?] and even differential privacy [?] (we did it for vibration signals). Furthermore, in a very recent work by Barrachina et [?], it seemed that the circularity property, peculiar of complex-random variables, could draw a clear line in the eternal comparison among complex and real networks. We achieved something similar in our thesis.

Regarding the last part of our work, in which we will address the problem of domain adaptation with complex-valued data, instead, at least according to our knowledge no papers have been published in this direction, making the analysis very interesting since we can propose it as the starting point for many future works.

1.3 Thesis Organization

After providing a brief explanation of the problem in the introduction above, the thesis is organized into two big parts: part ?? is dedicated to the theoretical analysis and implementation of a complex-valued deep learning framework, while in part ?? we proceed with a real-world application of the methodology developed.

In particular, in chapter ?? we go through a brief theoretical introduction of complex analysis, paying particular attention to the notions of complex differentiability and circularity; we will also introduce the theoretical advantages that complex-valued deep learning should provide over the real counterpart. In chapter ??, instead, we will continue with a more precise formulation of the framework we are trying to build, first with a working complex backpropagation algorithm and then with some possible extents, in the complex domain, of the most common machine learning layers and activations. In chapter ?? we added a practical implementation of the operations defined, together with the comparison with an equivalent real-valued procedure. There is also an interesting section examining the impact of the circularity quotient when training complex models.

In chapter ??, we will finally discuss and implement the framework in a real world situation, that

is the problem of condition monitoring in industrial applications. In particular, we will work on a dataset of vibration signals, provided by Bonfiglioli, an important gearmotors producer, comparing the effective performances achieved with both real and complex-valued models. In the last chapter **??**, instead, we will propose an analysis of a known domain adaptation algorithm that seems to be easily extendable also in the complex domain.