Chapter 6
Bochner Integrals and Neural Networks

Paul C. Kainen and Andrew Vogt

Abstract. A Bochner integral formula \( f = B - \int_Y w(y)\Phi(y)\,d\mu(y) \) is derived that presents a function \( f \) in terms of weights \( w \) and a parametrized family of functions \( \Phi(y), y \in Y \). Comparison is made to pointwise formulations, norm inequalities relating pointwise and Bochner integrals are established, \( G \)-variation and tensor products are studied, and examples are presented.

Keywords: Variational norm, essentially bounded, strongly measurable, Bochner integration, tensor product, \( L^p \) spaces, integral formula.

1 Introduction

A neural network utilizes data to find a function consistent with the data and with further “conceptual” data such as desired smoothness, boundedness, or integrability. The weights for a neural net and the functions embodied in the hidden units can be thought of as determining a finite sum that approximates some function. This finite sum is a kind of quadrature for an integral formula that would represent the function exactly.

This chapter uses abstract analysis to investigate neural networks. Our approach is one of enrichment: not only is summation replaced by integration, but also numbers are replaced by real-valued functions on an input set \( \Omega \), the functions lying in a function space \( \mathcal{X} \). The functions, in turn, are replaced by \( \mathcal{X} \)-valued measurable functions \( \Phi \) on a measure space \( Y \) of parameters. The goal is to understand approximation of functions by neural networks so that one can make effective choices of the parameters to produce a good approximation.

Paul C. Kainen · Andrew Vogt
Department of Mathematics and Statistics, Georgetown University
Washington, D.C. 20057-1233, USA
e-mail: \{kainen, vogta\}@georgetown.edu
To achieve this, we utilize Bochner integration. The idea of applying this tool to neural nets is in Girosi and Anzellotti \[14\] and we developed it further in Kainen and Kůrková \[23\]. Bochner integrals are now being used in the theory of support vector machines and reproducing kernel Hilbert spaces; see the recent book by Steinwart and Christmann \[42\], which has an appendix of more than 80 pages of material on operator theory and Banach-space-valued integrals. Bochner integrals are also widely used in probability theory in connection with stochastic processes of martingale-type; see, e.g., \[8, 39\]. The corresponding functional analytic theory may help to bridge the gap between probabilistic questions and deterministic ones, and may be well-suited for issues that arise in approximation via neural nets.

Training to replicate given numerical data does not give a useful neural network for the same reason that parrots make poor conversationalists. The phenomenon of overfitting shows that achieving fidelity to data at all costs is not desirable; see, e.g., the discussion on interpolation in our other chapter in this book (Kainen, Kůrková, and Sanguineti). In approximation, we try to find a function close to the data that achieves desired criteria such as sufficient smoothness, decay at infinity, etc. Thus, a method of integration which produces functions \emph{in toto} rather than numbers could be quite useful.

Enrichment has lately been utilized by applied mathematicians to perform image analysis and even to deduce global properties of sensor networks from local information. For instance, the Euler characteristic, ordinarily thought of as a discrete invariant, can be made into a variable of integration \[7\]. In the case of sensor networks, such an analysis can lead to effective computations in which theory determines a minimal set of sensors \[40\].

By modifying the traditional neural net focus on training sets of data so that we get to families of functions in a natural way, we aim to achieve methodological insight. Such a framework may lead to artificial neural networks capable of performing more sophisticated tasks.

The main result of this chapter is Theorem 5 which characterizes functions to be approximated in terms of pointwise integrals and Bochner integrals, and provides inequalities that relate corresponding norms. The relationship between integral formulas and neural networks has long been noted; e.g., \[20, 6, 37, 13, 34, 29\] We examine integral formulas in depth and extend their significance to a broader context.

An earlier version of the Main Theorem, including the bounds on variational norm by the $L^1$-norm of the weight function in a corresponding integral formula, was given in \[23\] and it also utilized functional (i.e., Bochner) integration. However, the version here is more general and further shows that if $\phi$ is a real-valued function on $\Omega \times Y$ (the cartesian product of input and parameter spaces), then the associated map $\Phi$ which maps the measure space to the Banach space defined by $\Phi(y)(x) = \phi(x, y)$ is measurable; cf. \[42\] Lemma 4.25, p. 125 where $\Phi$ is the “feature map.”

Other proof techniques are available for parts of the Main Theorem. In particular, Kůrková \[28\] gave a different argument for part (iv) of the