On Universal Approximation Theory for Minimizing the ACF of a Stochastic Process

A Preprint

Andre Zapico
Quantitative Consultant
Likely LLC
likelyandre@gmail.com

December 5, 2025

Abstract

Building filters for MIMO radar systems using artificial neural networks has received recent attention. For MIMO waveform design problems, typically engineers will use some form of convex optimization. Since using convex optimization algorithms requires intense mathematical expertise, there has been a push to use a machine learning approach using artificial neural networks in machine learning to build filters, due to ease of implementation and flexibility. However, it is still questionable that artificial neural networks are a good approach for filter design. To my knowledge, there is currently no theory that justifies the use in waveform design. In this paper, I take a mathematical theoretical approach and draw from universal approximation theory to determine whether artificial neural networks are an appropriate tool for waveform design. In particular, I determine theoretically whether or not the MLP neural network architecture is an appropriate approach for the constant modulus waveform design problem with good correlation properties.

1 Introduction

This paper is organized as follows. First, we prove that the autocorrelation function of a discrete time stochastic process is discrete. Next, drawing from approximation theory of multilayer perceptron function approximation theory, we investigate whether or not I arrive at a logical contradiction when the function of interest is a discrete function instead of continuous.

2 Theoretical Background

First, I prove that the autocorrelation function of a discrete time stochastic process is discrete. We use the the definition of a discrete time stochastic process as in [Knill, 2021].

Theorem 1 *The autocorrelation function of a discrete time stochastic process is also discrete.*

Proof. Let X_t be a discrete time stochastic process. That is, X_t is a collection of random variables and $t \in T$, a discrete subset of $\mathbf{R}$. Next, its autocorrelation function is then, for $s, t \in T$, $s \neq t$ is

$$\sum_{t \in \{1,2,...,T-s\}} X_t X_{s+t} = \Gamma(s,t)$$

For s in $\{1, 2, ..., T-(s+t)\}$. Since X_t, X_{s+t} are random variables in $\mathbf{R}$, then $\Gamma(s,t) \in \mathbf{R}^s$. Clearly, there exists a surjective function $f : \mathbf{R}^s \to \mathbf{N}^s$, then the autocorrelation function for a stochastic process is discrete. To see that there exists a surjective function, consider the trivial map, where we map each element of the domain to the first $T-(s+t)$ elements of the set of natural numbers. Since $\Gamma(s,t)$ is countable, it is discrete.

Next, we show that ACF is not dense on some open interval on the real numbers. We use the definition of density from the definition in the appendix of [lee].

Definition 1 *A set S is said to be* ***dense*** *in X is every nonempty open subset of X contains at least one point of X.*

Theorem 2 *Let* $(a, b) \in \mathbf{R}$ *be an open subset of R. Then any discrete time sochastic process that takes on values in* (a, b) *is not dense in* (a, b).

Proof. Let $X = (a, b)$ and $S = X_i$, where each $X_i \sim uniform(a, b)$, for $i = 1, 2, 3, ..., N$ is a discrete uniform sample of (a, b), which is a stochastic process. Since each X_i is a discrete uniform sample of (a, b), the cardinatiy of the set $\{X_i\}$, the set containing only point X_i is 1. Since (a, b) is uncountable, the probability of selecting point X_i from the interval (a, b) is $\frac{|(\{X_i\}|}{|(a,b)|}$, where $| \cdot |$ represents the cardinality of a set. But since $\mid (a, b) \mid$ is infinite, the fraction above evaluates to 0. Then, for any X_i, X_j, $i \neq j$, $p(X_i = X_j) = 0$. Then, for any sequence of RV's, X_i, $i = 1, 2, 3, ..., N$, $X_i = X_j$ for $i \neq j$ is entirely improbable. Then, we have $X_i < X_j$ for any $i \neq j \in \{1, 2, 3, ..., N\}$. But also, since (a, b) is continuous, there is some nonempty open interval $(X_i, X_j) \subset (a, b) \subset \mathbf{R}$, s.t. $\beta \neq X_i$ for $i = 1, 2, 3, ..., N$, by the cardinality argument before. Then, since $\beta \notin \{X_1, X_2, ..., X_N\}$ and $\beta \in (a, b)$, then $\{X_1, X_2, ..., X_N\}$ is not dense in (a, b).

We've now proved that any discrete time stochastic process takes on values on an arbitrary open interval in $\mathbf{R}$ is not dense in that open interval. With this fact, we look at the work of [Korkov and Sciences] to examine the case where Kolmogorov's Universal approximation theory is applied to composition of linear functions. This is in fact the case of the multilayer perceptron neural network type architecture in computer science. They prove the following theorem.

Theorem 3 *Let* $n \in \mathbf{N}$ *with* $n \geq 2$, $\sigma : \mathbf{R} \to I$ *be a sigmoidal function,* $f \in C(I^n)$ *and* $\epsilon > 0$ *a real number. Then there exists* $k \in \mathbf{N}$ *and functions* ϕ_i, $\psi_i \in S(\sigma)$ *such that*

$$\mid f(x_1, ..., x_n) - \sum_{i=1}^{k} \phi_i(\sum_{p=1}^{n} \psi_{pj} x_p) \mid < \epsilon$$

for every $(x_1, ..., x_n) \in I^n$.

I omit a proof of the theorem since it's easily found in [Korkov and Sciences]. In the theorem note that the approximand is assumed to be continuous over and unit-n cube. That is, $f \in C(I^n)$. Instead, consider the case where approximand is the autocorrelation function of a discrete time stochastic process, $\Gamma(s, t)$, which we've proved above is not dense on an open interval in $\mathbf{R}$, and hence, clearly not dense in a continuous unit cube. If we then replace the approximand in **Theorem 3** with the function $\Gamma(s, t)$, note that, since there is this lack of density, we cannot find approximating functions to minimize the distance between the approximator and approximand.

To see the case that violates in the inequality above, first recall that ϵ is any real number not less than 0. Take any element in the codomain of $\Gamma(s, t)$. Any of these elements, or selection of more or one thereof, is not dense in R. But since, the cardinality of the set that ϵ is a member of is infinitely greater than $\Gamma(s, t)$'s membership, we can find infinitely many numbers for which the inequality in **Theorem 3** doesn't hold. More formally,

since the cardinality of set A is not dense in set B, and B is uncountable, then there exists infinitely many values that violate the inequality in **Theorem 3**. To see this, consider a cardinality argument.

to prove
Suppose we have setss A and B such that $A \subset B$ and A is not dense in B, and that B is an uncountably infinite set. Then, there exists an ϵ such that $\epsilon \in B$ and $\epsilon \notin A$.

Proof. For purpose of contradiction, supposed $A \subset B$, A is not dense in B, B is uncountably infinite and there does not exist any points ϵ s.t. $\epsilon \in B$ and $\epsilon \notin A$. But since A is not dense in B, we can find a point, namely, an accumulating point, ϵ such that $\epsilon \in A$ and ϵ is not in B.

Now, we've shown that, more generally, that an autocorrelation function will not be dense on some open subset of $\mathbf{R}$, and that the current universal approximation theory does not support the fact that we use an MLP as a universal approximator for an autocorrelation objective function.

At this point, we've proven that Korkov's theorem does not support the fact that composite linear affine functions, and hence, a multilayer perceptron, can work as a universal approximator. Although in this work, we've shown that, above theories do not give mathematical proof that the MLP cannot be a universal approximator.

To show whether or not **Theorem 3** hold for the ACF, or a non-dense codomain, we investigate where **Theorem 3** does not hold when there is no density in the codomain of the approximand.

Proof.

TODO: ϕ, ψ, continuous functions? What assumptions are needed for this proof?

Here's where we left off: Approximation theory of the MLP model in neural networks, Pinkus, Page 151

References

Oliver Knill. *Probability Theory and Stochastic Processes with Applications*. WORLD SCIENTIFIC, 2 edition, November 2021. ISBN 978-981-310-948-3 978-981-310-950-6. doi:10.1142/10029. URL `https://www.worldscientific.com/worldscibooks/10.1142/10029`.

Inroduction to Smooth Manifolds. 2 edition.

Vra Korkov and CzechoslovakAcademyof Sciences. Kolmogorov's Theorem and Multilayer Neural Networks. page 6.