Journal of Machine Learning Research 26 (2025) 1-31 Submitted 5/24; Published 1/25

Random ReLU Neural Networks as Non-Gaussian Processes

Rahul Parhi RAHUL@UCSD.EDU
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093, USA

Pakshal Bohra PAKSHALBOHRA@GMAIL.COM
Ayoub El Biari AYOUBELBIARI@GMAIL.COM
Mehrsa Pourya MEHRSA.POURYA@EPFL.CH
Michael Unser MICHAEL.UNSER@EPFL.CH
Biomedical Imaging Group
École polytechnique fédérale de Lausanne
CH-1015 Lausanne, Switzerland

Editor: Mohammad Emtiyaz Khan

Abstract

We consider a large class of shallow neural networks with randomly initialized parameters and rectified linear unit activation functions. We prove that these random neural networks are well-defined non-Gaussian processes. As a by-product, we demonstrate that these networks are solutions to stochastic differential equations driven by impulsive white noise (combinations of random Dirac measures). These processes are parameterized by the law of the weights and biases as well as the density of activation thresholds in each bounded region of the input domain. We prove that these processes are isotropic and wide-sense self-similar with Hurst exponent 3/2. We also derive a remarkably simple closed-form expression for their autocovariance function. Our results are fundamentally different from prior work in that we consider a non-asymptotic viewpoint: The number of neurons in each bounded region of the input domain (i.e., the width) is itself a random variable with a Poisson law with mean proportional to the density parameter. Finally, we show that, under suitable hypotheses, as the expected width tends to infinity, these processes can converge in law not only to Gaussian processes, but also to non-Gaussian processes depending on the law of the weights. Our asymptotic results provide a new take on several classical results (wide networks converge to Gaussian processes) as well as some new ones (wide networks can converge to non-Gaussian processes).

Keywords: Gaussian processes, non-Gaussian processes, random initialization, random neural networks, stochastic processes.

1. Introduction

A shallow (single-hidden-layer) neural network is a function of the form

$$\boldsymbol{x} \mapsto \sum_{k=1}^{N} v_k \sigma(\boldsymbol{w}_k^{\mathsf{T}} \boldsymbol{x} - b_k), \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{1}$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is the activation function, N is the width of the network, and, for $k = 1, \ldots, N$, $v_k \in \mathbb{R}$ and $\boldsymbol{w}_k \in \mathbb{R}^d \setminus \{\boldsymbol{0}\}$ are the weights and $b_k \in \mathbb{R}$ are the biases of the network. It is well-known that, as $N \to \infty$, several such networks with i.i.d. random weights and biases are equivalent to a Gaussian process (Neal, 1996). This result was extended to deep neural networks with i.i.d. random parameters by Lee et al. (2018). This correspondence enables exact Bayesian inference for regression using wide neural networks (Williams, 1996; Lee et al., 2018).

Motivated by the tight link between wide neural networks and stochastic processes, we study properties of shallow rectified linear unit (ReLU) neural networks with randomly initialized parameters, henceforth referred to as *random (ReLU) neural networks*. We study Poisson-type random functions of the form

$$s_{\mathrm{ReLU}}(\boldsymbol{x}) = \sum_{k \in \mathbb{Z}} v_k \Big[\mathrm{ReLU}(\boldsymbol{w}_k^{\mathsf{T}} \boldsymbol{x} - b_k) + \boldsymbol{c}_k^{\mathsf{T}} \boldsymbol{x} + c_{0,k} \Big], \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{2}$$

where $\mathrm{ReLU}(t) \coloneqq t_+ = \max\{0, t\}$, the v_k are drawn i.i.d. with respect to the law $\mathbf{P}_V$ and the $(\boldsymbol{w}_k, b_k)$ are drawn such that

1. the activation thresholds[1] are mutually independent;
2. in expectation, the number of thresholds that intersect a finite volume in $\mathbb{R}^d$ is a constant (proportional to the product of a parameter $\lambda > 0$ and a property related to the geometry of the volume); and
3. for every finite volume in $\mathbb{R}^d$, the thresholds are i.i.d. uniformly in the volume.

The randomness that generates the $(\boldsymbol{w}_k, b_k)$ motivates the denomination *Poisson* as it mimics the randomness in the jumps found in a unit interval of a compound Poisson process (Daley and Vere-Jones, 2007). The parameter $\lambda > 0$ plays the role of the rate parameter of a compound Poisson process and controls the density of activation thresholds in each finite volume. The correction terms $(\boldsymbol{x} \mapsto \boldsymbol{c}_k^{\mathsf{T}} \boldsymbol{x} + c_{0,k})_{k \in \mathbb{Z}}$ that appear in the sum are affine functions that ensure that the sum in (2) converges almost surely. This is equivalent to imposing boundary conditions on s_{ReLU}. These boundary conditions are crucial in proving that, under suitable hypotheses on $\mathbf{P}_V$, s_{ReLU} is a well-defined stochastic process. This is one of the primary technical contributions of this paper. Similar correction terms/boundary conditions appear in the definition of fractional Brownian motion (Mandelbrot and Van Ness, 1968) and Lévy processes (Sato, 1999; Jacob and Schilling, 2001).

By restricting our attention to compact subsets $\Omega \subset \mathbb{R}^d$, say, to the unit ball $\mathbb{B}_1^d = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \leq 1\}$, we have that (see Section 4.1) the process (2) is realized by a random Poisson sum of the form

$$s_{\mathrm{ReLU}}\big|_{\mathbb{B}_1^d}(\boldsymbol{x}) = \boldsymbol{w}_0^{\mathsf{T}} \boldsymbol{x} + b_0 + \sum_{k=1}^{N_\lambda} v_k \, \mathrm{ReLU}(\boldsymbol{w}_k^{\mathsf{T}} \boldsymbol{x} - b_k), \tag{3}$$

where the width N_λ is a Poisson random variable with mean λS, where S is proportional to the surface area of $\mathbb{B}_1^d$, and $\boldsymbol{w}_0^{\mathsf{T}} \boldsymbol{x} + b_0$ is an affine function. Thus, the form in the right-hand

1. The *activation threshold* of the neuron $\boldsymbol{x} \mapsto \mathrm{ReLU}(\boldsymbol{w}^{\mathsf{T}} \boldsymbol{x} - b)$ is the hyperplane $H_{\boldsymbol{w},b} = \{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{w}^{\mathsf{T}} \boldsymbol{x} = b\}$.

side of (3) is a *finite-width* neural network with random parameters (including the width). The affine function $\boldsymbol{x} \mapsto \boldsymbol{w}_0^{\mathsf{T}}\boldsymbol{x} + b_0$ is a *skip connection* in neural network parlance. As $\lambda \to \infty$, we have that the expected value of the width satisfies $\mathbf{E}[N_\lambda] \to \infty$. Therefore, this limiting scenario corresponds to the asymptotic (i.e., infinite-width) regime.

1.1 Contributions

The purpose of this paper is to study the properties of random neural networks as in (2) and (3) for the class of admissible laws $\mathbf{P}_V$ (in the sense of Definition 5) which, for example, includes the Gaussian law. As these networks are completely specified by the law $\mathbf{P}_V$ and the rate parameter $\lambda > 0$, we let

$$s_{\mathrm{ReLU}}(\cdot) \sim \mathcal{RP}(\lambda; \mathbf{P}_V) \tag{4}$$

denote that s_{ReLU} is generated according to the randomness described above, where $\mathcal{RP}$ stands for *ReLU process*. The main contributions of this paper are outlined below.

Random ReLU Networks as Stochastic Processes In Section 4, we prove that s_{ReLU} is a well-defined stochastic process. In doing so, we derive the so-called *characteristic functional*[2] of the process, which provides us with a complete characterization of its statistical distribution. Further, we show that s_{ReLU} is the unique continuous piecewise linear (CPwL) solution to the stochastic differential equation (SDE)

$$\mathrm{T}_{\mathrm{ReLU}}\, s \overset{\mathscr{L}}{=} w \quad \text{s.t.} \quad \partial^{\boldsymbol{m}} s(\mathbf{0}) = 0, |\boldsymbol{m}| \leq 1, \tag{5}$$

where $\overset{\mathscr{L}}{=}$ denotes equality in law and $\mathrm{T}_{\mathrm{ReLU}} = \mathrm{K}\,\mathscr{R}\,\Delta$ is the *whitening operator* for ReLU neurons. The driving term w of the SDE is an *impulsive white noise process* which is constructed from combinations of random Dirac measures. The boundary conditions $\partial^{\boldsymbol{m}} s(\mathbf{0}) = 0$, $|\boldsymbol{m}| \leq 1$, are crucial in guaranteeing the existence of solutions to this SDE. In the form of the whitening operator, K is the filtering operator of computed tomography, $\mathscr{R}$ is the Radon transform, and Δ is the Laplacian (see Section 3 for a precise definition of these operators). The operator $\mathrm{T}_{\mathrm{ReLU}}$ was proposed by Ongie et al. (2020) to study the capacity of bounded-norm infinite-width ReLU networks.

Properties of Random ReLU Networks In Section 5, we derive the first- and second-order statistics of s_{ReLU}. Specifically, we present a remarkably simple closed-form expression for its autocovariance function. With the help of these statistics and the characteristic functional, we show that s_{ReLU} is a non-Gaussian process. We then show that s_{ReLU} is isotropic and wide-sense self-similar with Hurst exponent $H = 3/2$.

Asymptotic Results In Section 6, we show that in the infinite-width regime ($\lambda \to \infty$), s_{ReLU} converges in law to a Gaussian process when $\mathbf{P}_V$ is a Gaussian law with a variance that is inversely proportional to λ. On the other hand, when $\mathbf{P}_V$ is a symmetric α-stable (SαS) law with $\alpha \in (1, 2)$ and scaling parameter proportional to $\lambda^{-1/\alpha}$, s_{ReLU} converges in law to a non-Gaussian process.

2. The characteristic functional of a stochastic process is analogous to the characteristic function of a random variable. See Section 2 for a detailed discussion.

1.2 Related Work

There is a large body of work that investigates the connections between neural networks with random initialization and stochastic processes. Early work in this direction is due to Neal (1996) who proved that wide limits of shallow neural networks with bounded activation functions are Gaussian processes when the $(\boldsymbol{w}_k, b_k)$ are drawn i.i.d. with respect to any law and the v_k are drawn i.i.d. with respect to a law that has zero mean and finite variance. More recently, it has been argued by many authors, with varying degrees of mathematical rigor, that deep neural networks with i.i.d. random initialization are Gaussian processes in wide limits (Lee et al., 2018; Matthews et al., 2018; Garriga-Alonso et al., 2019; Novak et al., 2019; Yang, 2019; Dyer and Gur-Ari, 2020; Hanin, 2023).

Another line of work that is closely related to our setting is that of Yaida (2020), who studies the stochastic processes realized by *finite-width* random neural networks and shows that such processes are *non-Gaussian*. The results of this paper are complementary to that of Yaida (2020) in that our finite-width networks as in (3) also correspond to non-Gaussian processes. However, our work is fundamentally different as we use the framework of generalized stochastic processes (see Section 2). This allows us to derive the characteristic functional of the random neural network, which provides a complete description of its statistical distribution (i.e., the law of the process). The characteristic functional also allows us to easily study the limiting processes as the expected width $\mathbf{E}[N_\lambda] \to \infty$, which Yaida (2020) does not investigate.

In particular, we derive a novel and remarkably simple closed-form expression of the autocovariance function of the ReLU processes. Another important distinction of our asymptotic results compared to prior work on wide networks is that, in the asymptotic regime ($\lambda \to \infty$), the neural networks as in (2) and (3) can converge not only to Gaussian processes, but also to non-Gaussian processes, depending on the specific choice of $\mathbf{P}_V$. This type of result was alluded to by Neal (1996) in the case of SαS initialization, although theoretical arguments were not carried out. Thus, this paper is the first, to the best of our knowledge, to carry out a rigorous investigation of the convergence of wide networks to non-Gaussian processes.

2. Generalized Stochastic Processes

The mathematical framework used in this paper is based on the theory of *generalized stochastic processes* (Itô, 1954; Gelfand, 1955; Gelfand and Vilenkin, 1964; Itô, 1984) as opposed to the more common "time-series" approach to studying stochastic processes. In this section, we present the relevant background on generalized stochastic processes. We also refer the reader to the book of Unser and Tafti (2014) for further background. While this theory relies on some rather heavy concepts from functional analysis, it allows for elegant arguments to investigate the properties of the stochastic processes realized by the random neural networks in (2) and (3).

Throughout this paper, we fix a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Before we introduce this theory, we first recall some results from classical probability theory. A real-valued random vector $\boldsymbol{X}$ is a measurable function from the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where

$\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra on $\mathbb{R}^d$. The *law* of $\boldsymbol{X}$ is the *pushforward measure*

$$\mathbf{P}_{\boldsymbol{X}}(A) \coloneqq (\boldsymbol{X}_\sharp \mathbf{P})(A) \coloneqq \mathbf{P}\big(\boldsymbol{X}^{-1}(A)\big) = \mathbf{P}(\{\omega \in \Omega : \ \boldsymbol{X}(\omega) \in A\}) = \mathbf{P}(\boldsymbol{X} \in A), \quad (6)$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$. Consequently, the *characteristic function* of $\boldsymbol{X}$ is the (conjugate) Fourier transform of $\mathbf{P}_{\boldsymbol{X}}$, given by

$$\widehat{\mathbf{P}}_{\boldsymbol{X}}(\boldsymbol{\xi}) = \mathbf{E}[\mathrm{e}^{\mathrm{i}\boldsymbol{X}^\mathsf{T}\boldsymbol{\xi}}], \quad \boldsymbol{\xi} \in \mathbb{R}^d, \quad (7)$$

where $\mathrm{i}^2 = -1$.

Generalized stochastic processes are random variables that take values in the (continuous) dual of a *nuclear space*. In the remainder of this section, let $\mathcal{N}$ denote a nuclear space and $\mathcal{N}'$ denote its dual. If $u \in \mathcal{N}'$ and $\varphi \in \mathcal{N}$, we let $\langle u, \varphi \rangle_{\mathcal{N}' \times \mathcal{N}}$ denote the the *duality pairing* of u and φ (i.e., the evaluation of u at φ). A prototypical example of a nuclear space is the Schwartz space $\mathcal{S}(\mathbb{R}^d)$ of smooth and rapidly decreasing test functions. Its dual $\mathcal{S}'(\mathbb{R}^d)$ is the space of tempered generalized functions.[3] In order to discuss random variables that take values in the dual of a nuclear space, we must equip that space with a σ-algebra.

Definition 1 *The* cylindrical σ-algebra on $\mathcal{N}'$*, denoted by* $\mathcal{B}_c(\mathcal{N}')$*, is the* σ*-algebra generated by cylinders of the form* $\{u \in \mathcal{N}' : \ (\langle u, \varphi_1 \rangle_{\mathcal{N}' \times \mathcal{N}}, \dots, \langle u, \varphi_N \rangle_{\mathcal{N}' \times \mathcal{N}}) \in A\}$*, where* $N \in \mathbb{N} \setminus \{0\}$*,* $\varphi_1, \dots, \varphi_N \in \mathcal{N}$*, and* $A \in \mathcal{B}(\mathbb{R}^N)$*.*

We remark that when $\mathcal{N}$ is not only nuclear, but also Fréchet, such as $\mathcal{S}(\mathbb{R}^d)$, the cylindrical σ-algebra $\mathcal{B}_c(\mathcal{N}')$ coincides with the Borel σ-algebra $\mathcal{B}(\mathcal{N}')$ (see Fernique, 1967; Itô, 1984).

Definition 2 *A* generalized stochastic process *is a measurable mapping*

$$s : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathcal{N}', \mathcal{B}_c(\mathcal{N}')). \quad (8)$$

The law *of* s *is then the probability measure* $\mathbf{P}_s \coloneqq s_\sharp \mathbf{P}$ *which is defined on* $\mathcal{B}_c(\mathcal{N}')$*. The* characteristic functional[4] *of* s *is the (conjugate) Fourier transform of* $\mathbf{P}_s$*, given by*

$$\widehat{\mathbf{P}}_s(\varphi) = \mathbf{E}[\mathrm{e}^{\mathrm{i}\langle s, \varphi \rangle_{\mathcal{N}' \times \mathcal{N}}}], \quad \varphi \in \mathcal{N}. \quad (9)$$

Observe that this definition recovers the classical characteristic function for random vectors that take values in $\mathbb{R}^d$. Indeed, $\mathbb{R}^d$ is a nuclear space whose dual is $\mathbb{R}^d$. Furthermore, for any $(\boldsymbol{x}, \boldsymbol{\xi}) \in \mathbb{R}^d \times \mathbb{R}^d$, we have that $\langle \boldsymbol{x}, \boldsymbol{\xi} \rangle_{\mathbb{R}^d \times \mathbb{R}^d} = \boldsymbol{x}^\mathsf{T}\boldsymbol{\xi}$. The characteristic functional of a generalized stochastic process contains all statistical information of the process in the same way that the characteristic function of a classical random variable contains all statistical information of that random variable. Analogous to the finite-dimensional case, the Bochner–Minlos theorem (see Minlos (1959)) says that a functional $\widehat{\mathbf{P}} : \mathcal{N} \to \mathbb{C}$ is the characteristic functional of a generalized stochastic process if and only if $\widehat{\mathbf{P}}$ is continuous, positive definite, and satisfies $\widehat{\mathbf{P}}(0) = 1$.

The attractive feature of the framework of generalized stochastic processes is that it covers not only classical stochastic processes, but also processes that do not admit a pointwise interpretation such as white noise processes. For example, a generalized Gaussian process is defined as follows.

3. This space is often referred to as the space of tempered *distributions*. We adopt the nomenclature of tempered *generalized functions* in this paper so as to not cause confusion with probability distributions.
4. The characteristic functional of a generalized stochastic process was introduced by Kolmogorov (1935).

Definition 3 *A* generalized stochastic process *s that takes values in $\mathcal{N}'$ is said to be* Gaussian *if its characteristic functional is of the form*

$$\widehat{\mathbf{P}}_s(\varphi) = \exp\left(\mathrm{i}\mu_s(\varphi) - \frac{1}{2}\Sigma_s(\varphi,\varphi)\right), \tag{10}$$

where $\varphi \in \mathcal{N}$, $\mu_s : \mathcal{N} \to \mathbb{R}$ is the mean functional of the process, given by

$$\mu_s(\varphi) = \mathbf{E}[\langle s, \varphi\rangle_{\mathcal{N}'\times\mathcal{N}}], \tag{11}$$

and $\Sigma_s : \mathcal{N}\times\mathcal{N} \to \mathbb{R}$ is the covariance functional of the process, given by

$$\Sigma_s(\varphi_1,\varphi_2) = \mathbf{E}[(\langle s,\varphi_1\rangle_{\mathcal{N}'\times\mathcal{N}} - \mu_s(\varphi_1))(\langle s,\varphi_2\rangle_{\mathcal{N}'\times\mathcal{N}} - \mu_s(\varphi_2))]. \tag{12}$$

The above definition is backwards compatible with classical Gaussian processes that are space-indexed, as shown by Duttweiler and Kailath (1973), yet it also includes Gaussian white noise (Hida and Ikeda, 1967).

With this machinery in hand, the primary technical contributions of this paper are (i) to prove that, for any $\lambda > 0$ and any admissible $\mathbf{P}_V$ (in the sense of Definition 5), the random neural network $s_{\mathrm{ReLU}} \sim \mathcal{RP}(\lambda; \mathbf{P}_V)$ is a generalized stochastic process that takes values in $\mathcal{S}'(\mathbb{R}^d)$, and (ii) to provide an explicit form of its (non-Gaussian) characteristic functional (Section 4). With the help of the latter, we then derive various properties of the stochastic process in the non-asymptotic regime (Section 5) and also study its asymptotic ($\lambda \to \infty$) behavior (Section 6) for various $\mathbf{P}_V$.

3. The Radon Transform and Related Operators

Our characterization of random ReLU neural networks as stochastic processes hinges on the whitening operator that appears in the SDE (5). This operator is based on the Radon transform. In this section we introduce the relevant background on the Radon transform and related operators. We refer the reader to the books of Ramm and Katsevich (1996) and Helgason (2011) for an in depth treatment of the Radon transform. The Radon transform of $\varphi \in L^1(\mathbb{R}^d)$ is given by

$$\mathscr{R}\{\varphi\}(\boldsymbol{u}, t) = \int_{\boldsymbol{u}^\mathsf{T}\boldsymbol{x}=t} \varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \quad (\boldsymbol{u}, t) \in \mathbb{S}^{d-1}\times\mathbb{R}, \tag{13}$$

where $\mathrm{d}\boldsymbol{x}$ denotes the integration against the $(d-1)$-dimensional Lebesgue measure on the hyperplane $\{\boldsymbol{x}\in\mathbb{R}^d : \boldsymbol{u}^\mathsf{T}\boldsymbol{x} = t\}$ and $\mathbb{S}^{d-1} = \{\boldsymbol{x}\in\mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$ denotes the unit sphere in $\mathbb{R}^d$. Observe that the Radon transform of φ is a *even* since $(\boldsymbol{u}, t)$ and $(-\boldsymbol{u}, -t)$ parametrize the same hyperplane. The adjoint operator, or dual Radon transform, applied to $\phi \in L^\infty(\mathbb{S}^{d-1}\times\mathbb{R})$ is given by

$$\mathscr{R}^*\{\phi\}(\boldsymbol{x}) = \int_{\mathbb{S}^{d-1}} \phi(\boldsymbol{u}, \boldsymbol{u}^\mathsf{T}\boldsymbol{x})\,\mathrm{d}\boldsymbol{u}, \quad \boldsymbol{x}\in\mathbb{R}, \tag{14}$$

where $\mathrm{d}\boldsymbol{u}$ denotes integration against the surface measure of $\mathbb{S}^{d-1}$.

Let $\mathcal{S}(\mathbb{S}^{d-1}\times\mathbb{R})$ denote the Schwartz space of smooth and rapidly decreasing functions on $\mathbb{S}^{d-1}\times\mathbb{R}$. The range of the Radon transform on $\mathcal{S}(\mathbb{R}^d)$, defined by $\mathcal{S}_{\mathscr{R}} \coloneqq \mathscr{R}\big(\mathcal{S}(\mathbb{R}^d)\big)$, is

a closed subspace of $\mathcal{S}(\mathbb{S}^{d-1} \times \mathbb{R})$ (Helgason, 2011, p. 60). Therefore, since $\mathcal{S}(\mathbb{S}^{d-1} \times \mathbb{R})$ is nuclear, $\mathcal{S}_{\mathscr{R}}$ is also nuclear. The next proposition summarizes the continuity and invertibility of the Radon transform.

Proposition 4 (Ludwig 1966; Gelfand et al. 1966; Helgason 2011) *The operator $\mathscr{R}$ continuously maps $\mathcal{S}(\mathbb{R}^d)$ into $\mathcal{S}(\mathbb{S}^{d-1} \times \mathbb{R})$. Moreover,*

$$\mathscr{R}^* \, \mathrm{K} \, \mathscr{R} = \frac{1}{2(2\pi)^{d-1}} (-\Delta)^{\frac{d-1}{2}} \, \mathscr{R}^* \, \mathscr{R} = \frac{1}{2(2\pi)^{d-1}} \, \mathscr{R}^* \, \mathscr{R} (-\Delta)^{\frac{d-1}{2}} = \mathrm{Id} \tag{15}$$

on $\mathcal{S}(\mathbb{R}^d)$. The underlying operators[5] *are the Laplacian $\Delta = \sum_{n=1}^{d} \partial_{x_n}^2$ and the filtering operator $\mathrm{K} = \frac{1}{2(2\pi)^{d-1}} (-\partial_t^2)^{\frac{d-1}{2}}$. Furthermore, $\mathscr{R} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}_{\mathscr{R}}$ is a homeomorphism with inverse $\mathscr{R}^{-1} = \mathscr{R}^* \, \mathrm{K} : \mathcal{S}_{\mathscr{R}} \to \mathcal{S}(\mathbb{R}^d)$.*

4. Random ReLU Neural Networks as Stochastic Processes

In this section, we will prove that, for any $\lambda > 0$ and admissible $\mathbf{P}_V$, the random neural network $s \sim \mathcal{RP}(\lambda; \mathbf{P}_V)$ is a well-defined stochastic process and derive its characteristic functional on $\mathcal{S}(\mathbb{R}^d)$. The admissibility conditions in Definition 5 are rather mild and most choices of $\mathbf{P}_V$ (e.g., Gaussian, SαS for $1 < \alpha \leq 2$, uniform, etc.) satisfy these hypotheses.

Definition 5 *We say that the probability measure $\mathbf{P}_V$ is* admissible *if*

1. *it is a Lévy measure, i.e., it satisfies $\mathbf{P}_V(\{0\}) = 0$ and $\int_{\mathbb{R}} \min\{1, v^2\} \, \mathrm{d}\mathbf{P}_V(v) < \infty$, and*
2. *it has a first absolute moment, i.e., if $V \sim \mathbf{P}_V$, then $\mathbf{E}[|V|] < \infty$.*

Given a ReLU neuron $\boldsymbol{x} \mapsto \mathrm{ReLU}(\boldsymbol{w}^\mathsf{T} \boldsymbol{x} - b)$ with $\boldsymbol{w} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ and $b \in \mathbb{R}$, we observe that, thanks to the homogeneity of the ReLU,

$$\mathrm{ReLU}(\boldsymbol{w}^\mathsf{T} \boldsymbol{x} - b) = \|\boldsymbol{w}\|_2 \, \mathrm{ReLU}(\widetilde{\boldsymbol{w}}^\mathsf{T} \boldsymbol{x} - \widetilde{b}), \tag{16}$$

where $\widetilde{\boldsymbol{w}} = \boldsymbol{w}/\|\boldsymbol{w}\|_2$ and $\widetilde{b} = b/\|\boldsymbol{w}\|_2$. Therefore, the space of functions representable by shallow ReLU neural networks with input weights constrained to be unit norm is the same as the space of functions representable by shallow ReLU neural networks without constraints on the weights (Parhi and Nowak, 2023b; Shenouda et al., 2024). To that end, we focus on neurons of the form $\mathrm{ReLU}(\boldsymbol{w}^\mathsf{T} \boldsymbol{x} - b)$ with $(\boldsymbol{w}, b) \in \mathbb{S}^{d-1} \times \mathbb{R}$.

An important property of the operator $\mathrm{T}_{\mathrm{ReLU}} = \mathrm{K} \mathscr{R} \Delta$ is that it "whitens" ReLU neurons. This result was implicitly proven by Ongie et al. (2020, Example 1), explicitly proven by Parhi and Nowak (2021, Lemma 17), and then further investigated by, e.g., Bartolucci et al. (2023, Lemma 5.6) and Unser (2023, Corollary 11). The whitening property is summarized in the following proposition.

Proposition 6 *For any ReLU neuron*

$$r_{(\boldsymbol{w},b)}(\boldsymbol{x}) = \mathrm{ReLU}(\boldsymbol{w}^\mathsf{T} \boldsymbol{x} - b) \tag{17}$$

5. Non-integer powers of $(-\Delta)$ and $(-\partial_t^2)$ are understood in the Fourier domain.

with $(\boldsymbol{w}, b) \in \mathbb{S}^{d-1} \times \mathbb{R}$*, we have that*

$$\mathrm{T}_{\mathrm{ReLU}}\, r_{(\boldsymbol{w},b)} = \delta^{\mathsf{e}}_{(\boldsymbol{w},b)}, \tag{18}$$

where $\delta^{\mathsf{e}}_{\boldsymbol{z}} = (\delta_{\boldsymbol{z}} + \delta_{-\boldsymbol{z}})/2$ *denotes the even symmetrization of the Dirac measure* $\delta_{\boldsymbol{z}}$ *supported at* $\boldsymbol{z} \in \mathbb{S}^{d-1} \times \mathbb{R}$*.*

The equality in (18) is understood in $\mathcal{M}^{\mathsf{e}}(\mathbb{S}^{d-1} \times \mathbb{R})$, the subspace of even finite (Radon) measures on $\mathbb{S}^{d-1} \times \mathbb{R}$. The arguments of the proof are based on duality. Indeed, observe that the adjoint of $\mathrm{T}_{\mathrm{ReLU}}$ takes the form $\mathrm{T}^*_{\mathrm{ReLU}} = \Delta \mathscr{R}^* \mathrm{K}$ (since Δ and K are self-adjoint). Furthermore, from Proposition 4 combined with the fact that $\Delta : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ is continuous, we see that $\mathrm{T}^*_{\mathrm{ReLU}} : \mathcal{S}_{\mathscr{R}} \to \mathcal{S}(\mathbb{R}^d)$ is continuous. Therefore, by duality, $\mathrm{T}_{\mathrm{ReLU}} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'_{\mathscr{R}}$ is continuous. Since $r_{(\boldsymbol{w},b)} \in \mathcal{S}'(\mathbb{R}^d)$, we have that $\mathrm{T}_{\mathrm{ReLU}}\, r_{(\boldsymbol{w},b)}$ is indeed well-defined. Finally, $\mathcal{M}^{\mathsf{e}}(\mathbb{S}^{d-1} \times \mathbb{R})$ is continuously embedded in $\mathcal{S}'_{\mathscr{R}}$ and so any finite measure in the range of $\mathrm{K}\mathscr{R}$ can be concretely identified to have even symmetries (see Unser, 2023; Parhi and Unser, 2024, for a detailed discussion). These symmetries are evidenced by the fact that the Radon transform of a "classical" function is necessarily even from the integral form in (13).

Proposition 6 motivates us to study Radon-domain impulsive white noises that are realized by Poisson-type random measures of the form

$$w_{\mathrm{Poi}} = \sum_{k \in \mathbb{Z}} v_k\, \delta^{\mathsf{e}}_{(\boldsymbol{w}_k, b_k)}, \tag{19}$$

where $v_k \overset{\text{i.i.d.}}{\sim} \mathbf{P}_V$ for some admissible $\mathbf{P}_V$ (in the sense of Definition 5) and the collection of random variables $((\boldsymbol{w}_k, b_k))_{k \in \mathbb{Z}}$ is a (homogeneous) Poisson point process[6] on $\mathbb{S}^{d-1} \times \mathbb{R}$ with rate parameter $\lambda > 0$. This point process satisfies the following properties.

1. The $(\boldsymbol{w}_k, b_k)$ are mutually independent.

2. For any measurable subset $\Pi \subset \mathbb{S}^{d-1} \times \mathbb{R}$, if we define the random variable

$$N_\Pi = |\{(\boldsymbol{w}_k, b_k) : (\boldsymbol{w}_k, b_k) \in \Pi\}|, \tag{20}$$

   then

$$\mathbf{P}(N_\Pi = n) = \frac{(\lambda |\Pi|)^n}{n!} \mathrm{e}^{-\lambda |\Pi|}, \tag{21}$$

   where $|\Pi|$ denotes the d-dimensional Hausdorff measure of Π. That is to say, N_Π is a Poisson random variable with mean $\lambda |\Pi|$.

3. For any measurable subset $B \subset \mathbb{S}^{d-1} \times \mathbb{R}$,

$$\mathbf{P}((\boldsymbol{w}_k, b_k) \in B \mid (\boldsymbol{w}_k, b_k) \in \Pi) = \frac{|B \cap \Pi|}{|\Pi|}. \tag{22}$$

   That is to say, if a point lies in Π, then its location will be uniformly distributed on Π.

6. For a general treatment of point processes, we refer the reader to the book of Daley and Vere-Jones (2007).

Next, if we suppose that there exists a "suitable" right-inverse $\mathrm{T}_{\mathrm{ReLU}}^{\dagger}$ of $\mathrm{T}_{\mathrm{ReLU}}$ that satisfies $\mathrm{T}_{\mathrm{ReLU}}\,\mathrm{T}_{\mathrm{ReLU}}^{\dagger} = \mathrm{Id}$ on $\mathcal{S}'_{\mathscr{R}}$, then, intuitively, we could "invert" the result of Proposition 6 to find that $\mathrm{T}_{\mathrm{ReLU}}^{\dagger}\{w_{\mathrm{Poi}}\}$ is precisely a random ReLU neural network generated in (2). It turns out that such a family of right-inverses exist. These inverses were first proposed by Parhi and Nowak (2021, Lemma 21) in order to prove representer theorems for neural networks. Some further properties of these operator were identified by Parhi and Nowak (2022) and Unser (2023). We summarize the properties from Parhi and Nowak (2021, Lemma 21) and Unser (2023, Theorem 13) that are required for our investigation in the next proposition.

Proposition 7 *For any $\varepsilon > 0$, there exists an operator $\mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}$ defined on $\mathcal{S}'_{\mathscr{R}}$ such that, for any $w \in \mathcal{S}'_{\mathscr{R}}$,*

$$\mathrm{T}_{\mathrm{ReLU}}\,\mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}\, w = w, \tag{23}$$

$$\Big[(\partial^{\boldsymbol{m}} g_d^{\varepsilon}) * \mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}\{w\}\Big](\mathbf{0}) = 0, |\boldsymbol{m}| \leq 1, \tag{24}$$

where $g_d^{\varepsilon} : \mathbb{R}^d \to \mathbb{R}$ is the multivariate Gaussian probability density function with mean $\mathbf{0}$ and covariance matrix $\mathrm{diag}(\varepsilon, \ldots, \varepsilon)$. The restriction of $\mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}$ to the subspace $\mathcal{M}^{\mathsf{e}}(\mathbb{S}^{d-1} \times \mathbb{R}) \subset \mathcal{S}'_{\mathscr{R}}$ continuously maps $\mathcal{M}^{\mathsf{e}}(\mathbb{S}^{d-1} \times \mathbb{R})$ to $\mathcal{S}'(\mathbb{R}^d)$. This mapping is realized by the integral operator

$$\mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}\Big|_{\mathcal{M}^{\mathsf{e}}(\mathbb{S}^{d-1}\times\mathbb{R})}\{\mu\}(\boldsymbol{x}) = \int_{\mathbb{S}^{d-1}\times\mathbb{R}} k_{\boldsymbol{x}}^{\varepsilon}(\boldsymbol{u}, t)\,\mathrm{d}\mu(\boldsymbol{u}, t) \tag{25}$$

whose kernel is given by

$$\begin{aligned} k_{\boldsymbol{x}}^{\varepsilon}(\boldsymbol{u}, t) &= \mathrm{ReLU}(\boldsymbol{u}^{\mathsf{T}}\boldsymbol{x} - t) - \frac{(\boldsymbol{u}^{\mathsf{T}}\boldsymbol{x} - t)}{2} - \left(g_1^{\varepsilon} * \frac{|\cdot|}{2}\right)(t) + (\boldsymbol{u}^{\mathsf{T}}\boldsymbol{x})\left(g_1^{\varepsilon} * \frac{\mathrm{sgn}}{2}\right)(t) \\ &= \mathrm{ReLU}(\boldsymbol{u}^{\mathsf{T}}\boldsymbol{x} - t) + \boldsymbol{u}_0^{\varepsilon\mathsf{T}}\boldsymbol{x} + t_0^{\varepsilon}, \end{aligned} \tag{26}$$

where sgn *is the signum function. Furthermore, there exists a universal constant $C > 0$ such that*

$$|k_{\boldsymbol{x}}^{\varepsilon}(\boldsymbol{u}, t)| \leq C(1 + \|\boldsymbol{x}\|_2) \quad \textit{for all } (\boldsymbol{u}, t) \in \mathbb{S}^{d-1} \times \mathbb{R}. \tag{27}$$

Remark 8 *The purpose of introducing the ε-indexed right-inverse operators is for a mollification argument. We will eventually consider the limit $\varepsilon \to 0$ (see the proof of Theorem 9 in Appendix A).*

With this inverse operator, we observe that, if w_{Poi} is an impulsive Poisson noise with rate $\lambda > 0$ and weights drawn i.i.d. according to $\mathbf{P}_V$ (as in (19)), then, for any $\varepsilon > 0$,

$$\begin{aligned} \mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}\{w_{\mathrm{Poi}}\} &= \mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}\left\{\sum_{k\in\mathbb{Z}} v_k\, \delta^{\mathsf{e}}_{(\boldsymbol{w}_k, b_k)}\right\} \\ &= \sum_{k\in\mathbb{Z}} v_k\, \mathrm{T}_{\mathrm{ReLU}}^{\dagger\varepsilon}\left\{\delta^{\mathsf{e}}_{(\boldsymbol{w}_k, b_k)}\right\} \\ &= \sum_{k\in\mathbb{Z}} v_k \Big[\mathrm{ReLU}(\boldsymbol{w}_k^{\mathsf{T}}(\cdot) - b_k) + \boldsymbol{c}_k^{\varepsilon\mathsf{T}}(\cdot) + c_{0,k}^{\varepsilon}\Big], \end{aligned} \tag{28}$$

where the second line is justified due to the uniform bound in (27), and the third line follows from (25). Therefore, $\mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}\{w_{\mathrm{Poi}}\}$ is a random neural network as in (2) that satisfies the boundary conditions in (24). We write

$$s^{\varepsilon}_{\mathrm{ReLU}} \sim \mathcal{RP}^{\varepsilon}(\lambda; \mathbf{P}_V) \tag{29}$$

to denote that $s^{\varepsilon}_{\mathrm{ReLU}}$ is such a random neural network. Furthermore, we let

$$s_{\mathrm{ReLU}} \sim \mathcal{RP}(\lambda; \mathbf{P}_V), \tag{30}$$

as introduced in Section 1, correspond to a random ReLU neural network that satisfies the the limiting boundary conditions as $\varepsilon \to 0$. That is to say, $\partial^{\boldsymbol{m}} s_{\mathrm{ReLU}}(\mathbf{0}) = 0$, $|\boldsymbol{m}| \leq 1$, with the convention that the value of a piecewise constant function at a jump is the middle value.

In the next theorem, we prove that these random neural networks are well-defined stochastic process that take values in $\mathcal{S}'(\mathbb{R}^d)$ and provide a complete statistical characterization through their characteristic functional.

Theorem 9 *For any $\varepsilon > 0$, $\lambda > 0$, and admissible $\mathbf{P}_V$ (in the sense of Definition 5), the random neural network $s^{\varepsilon}_{\mathrm{ReLU}} \sim \mathcal{RP}^{\varepsilon}(\lambda; \mathbf{P}_V)$ is a measurable mapping*

$$s^{\varepsilon}_{\mathrm{ReLU}} : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathcal{S}'(\mathbb{R}^d), \mathcal{B}(\mathcal{S}'(\mathbb{R}^d)) \tag{31}$$

with characteristic functional given by

$$\widehat{\mathbf{P}}_{s^{\varepsilon}_{\mathrm{ReLU}}}(\varphi) = \exp\left(\lambda \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} \left(\mathrm{e}^{\mathrm{i}v\, \mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u},t)} - 1\right) \mathrm{d}\boldsymbol{u}\, \mathrm{d}t\, \mathrm{d}\mathbf{P}_V(v)\right), \quad \varphi \in \mathcal{S}(\mathbb{R}^d), \tag{32}$$

where $\mathrm{d}\boldsymbol{u}$ denotes integration against the surface measure on $\mathbb{S}^{d-1}$ and

$$\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}} : \varphi \mapsto \int_{\mathbb{R}^d} k^{\varepsilon}_{\boldsymbol{x}}(\cdot)\varphi(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{33}$$

is the adjoint[7] *of $\mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}$. Furthermore, $s^{\varepsilon}_{\mathrm{ReLU}}$ is the unique CPwL solution to the SDE*

$$\mathrm{T}_{\mathrm{ReLU}}\, s \overset{\mathscr{L}}{=} w_{\mathrm{Poi}} \quad \text{s.t.} \quad [(\partial^{\boldsymbol{m}} g^{\varepsilon}_d) * s](\mathbf{0}) = 0, |\boldsymbol{m}| \leq 1, \tag{34}$$

among all tempered weak solutions,[8] *where w_{Poi} is an impulsive Poisson noise with rate λ and weights drawn i.i.d. according to $\mathbf{P}_V$ (as in (19)). All other tempered weak solutions to the SDE take the form $s^{\varepsilon}_{\mathrm{ReLU}} + h$, where h is a harmonic polynomial of degree ≥ 2.*[9]

Finally, in the limiting scenario ($\varepsilon \to 0$), we have that $s_{\mathrm{ReLU}} \sim \mathcal{RP}(\lambda; \mathbf{P}_V)$ is a measurable mapping $(\Omega, \mathcal{F}, \mathbf{P}) \to (\mathcal{S}'(\mathbb{R}^d), \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$ with characteristic functional given by

$$\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi) = \exp\left(\lambda \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} \left(\mathrm{e}^{\mathrm{i}v\, \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u},t)} - 1\right) \mathrm{d}\boldsymbol{u}\, \mathrm{d}t\, \mathrm{d}\mathbf{P}_V(v)\right), \quad \varphi \in \mathcal{S}(\mathbb{R}^d), \tag{35}$$

7. Observe that $\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}$ is well-defined on $\mathcal{S}(\mathbb{R}^d)$ thanks to (27).
8. A tempered weak solution to the SDE is any random tempered generalized function $s^{\star} \in \mathcal{S}'(\mathbb{R}^d)$ that satisfies (34). Such a solution is referred to as "tempered" as it lies in $\mathcal{S}'(\mathbb{R}^d)$ and "weak" since the action of $\mathrm{T}_{\mathrm{ReLU}}$ on $s^{\star}$ is understood by duality.
9. A harmonic polynomial h is a polynomial defined on $\mathbb{R}^d$ such that $\Delta h = 0$ on all of $\mathbb{R}^d$.

where $\mathrm{T}_{\mathrm{ReLU}}^{\dagger *}$ *is the limiting operator as* $\varepsilon \to 0$ *whose kernel is* $k_{\boldsymbol{x}} \coloneqq \lim_{\varepsilon \to 0} k_{\boldsymbol{x}}^{\varepsilon}$ *(pointwise limit). This random neural network is the unique CPwL solution to the SDE*

$$\mathrm{T}_{\mathrm{ReLU}}\, s \overset{\mathscr{L}}{=} w_{\mathrm{Poi}} \quad \text{s.t.} \quad \partial^{\boldsymbol{m}} s(\boldsymbol{0}) = 0, |\boldsymbol{m}| \leq 1. \tag{36}$$

While the proof of the theorem is rather technical, the main ingredients can be divided into two steps. The first is to prove that w_{Poi} is a well-defined stochastic process that takes values in $\mathcal{S}'_{\mathscr{R}}$. The second is to invoke the computation in (28) which linearly and continuously transforms w_{Poi} into a random ReLU neural network. This transformation allows us to derive the characteristic functional of s_{ReLU} in terms of the characteristic functional of w_{Poi}. The proof appears in Appendix A.

4.1 Restrictions to Compact Domains

Recall from (3) that, for any $\lambda > 0$ and admissible $\mathbf{P}_V$ (in the sense of Definition 5), the restriction of the random neural network $s_{\mathrm{ReLU}} \sim \mathcal{RP}(\lambda; \mathbf{P}_V)$ to a compact domain, say, the unit ball $\mathbb{B}_1^d$ is a random Poisson sum of the form

$$s_{\mathrm{ReLU}}\big|_{\mathbb{B}_1^d}(\boldsymbol{x}) = \boldsymbol{w}_0^{\mathsf{T}} \boldsymbol{x} + b_0 + \sum_{k=1}^{N_\lambda} v_k \,\mathrm{ReLU}(\boldsymbol{w}_k^{\mathsf{T}} \boldsymbol{x} - b_k), \tag{37}$$

where the width N_λ is a Poisson random variable. The reader can quickly check that the activation thresholds that intersect $\mathbb{B}_1^d$ correspond to Poisson points that lie in $\mathbb{S}^{d-1} \times [-1, 1]$. Thus, the number of neurons N_λ is a Poisson random variable with mean $\lambda |\mathbb{S}^{d-1} \times [-1, 1]|$, which is λ multiplied by twice the surface area of the $(d-1)$-sphere. For general compact domains $\Omega \subset \mathbb{R}^d$, following Parhi and Nowak (2023a, Section IV), we define

$$\mathcal{Z}_\Omega \coloneqq \{(\boldsymbol{w}, b) \in \mathbb{S}^{d-1} \times \mathbb{R} : \ \{\boldsymbol{x} : \ \boldsymbol{w}^{\mathsf{T}} \boldsymbol{x} = b\} \cap \Omega \neq \varnothing\}. \tag{38}$$

Then, the restriction $s_{\mathrm{ReLU}}\big|_\Omega$ is a random neural network whose width $N_{\lambda,\Omega}$ is a Poisson random variable with mean $\lambda |\mathcal{Z}_\Omega|$. As $\lambda \to \infty$, we see that $\mathbf{E}[N_{\lambda,\Omega}] \to \infty$. Therefore, the asymptotic setting ($\lambda \to \infty$) corresponds to the infinite-width regime.

5. Properties of Random ReLU Neural Networks

The characteristic functional (35) allows us to derive the first- and second-order statistics of s_{ReLU} as well as infer some of its other properties such as isotropy and wide-sense self-similarity. We summarize these properties in Theorem 10.

Theorem 10 *For* $\lambda > 0$ *and admissible* $\mathbf{P}_V$ *(in the sense of Definition 5), let* $s_{\mathrm{ReLU}} \sim \mathcal{RP}(\lambda; \mathbf{P}_V)$*. Then, the following statements hold.*

1. *The mean of* s_{ReLU} *is given by*

$$\mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})] = \lambda \mathbf{E}[V] \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} k_{\boldsymbol{x}}(\boldsymbol{u}, t) \,\mathrm{d}\boldsymbol{u} \,\mathrm{d}t, \tag{39}$$

where $k_{\boldsymbol{x}}$ *is defined in Theorem 9.*

2. *If* $\mathbf{P}_V$ *has a finite second moment, then the autocovariance of* s_{ReLU} *is given by*

$$\begin{aligned} C_{s_{\mathrm{ReLU}}}(\boldsymbol{x},\boldsymbol{y}) &= \mathbf{E}[(s_{\mathrm{ReLU}}(\boldsymbol{x}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})])(s_{\mathrm{ReLU}}(\boldsymbol{y}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{y})])] \\ &= A\lambda\mathbf{E}[V^2]\Big(\|\boldsymbol{x}-\boldsymbol{y}\|_2^3 - \|\boldsymbol{x}\|_2^3 - \|\boldsymbol{y}\|_2^3 + 3\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}(\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)\Big), \end{aligned} \tag{40}$$

where $A = \frac{\Gamma(-3/2)}{2^{d+3}\pi^{d/2}\Gamma((d+3)/2)}$ *and* $\Gamma(\cdot)$ *is Euler's gamma function.*

3. *The process* s_{ReLU} *is isotropic, i.e., it has the same probability law as its rotated version* $s_{\mathrm{ReLU}}(\mathbf{U}^{\mathsf{T}}\cdot)$*, where* $\mathbf{U}$ *is any* $(d \times d)$ *rotation matrix.*

4. *If* $\mathbf{P}_V$ *has zero mean and a finite second moment, then* s_{ReLU} *is wide-sense self-similar with Hurst exponent* $H = 3/2$*, i.e., it has the same second-order moments as its scaled and renormalized version* $a^H s_{\mathrm{ReLU}}(\cdot/a)$ *with* $a > 0$*.*

5. *The process* s_{ReLU} *is non-Gaussian.*

The proof of Theorem 10 can be found in Appendix B. We mention that the expression of the autocovariance in (40) is remarkably simple. This is in contrast to prior works that either (i) do not provide a closed-form expression (Lee et al., 2018; Yaida, 2020; Hanin, 2023), or (ii) provide a closed-form expression, but do not consider the ReLU activation function (Williams, 1996). Furthermore, other than the work of Yaida (2020), these prior works only consider the infinite-width regime.

6. Asymptotic Results

In the literature, there has been a lot of work on studying the wide limits of random neural networks. Here, we present an asymptotic result for random ReLU neural networks with i.i.d. weights drawn from an SαS law. The proof appears in Appendix C.

Theorem 11 *For* $n \in \mathbb{N}$*, let* $s^n_{\mathrm{ReLU}} \sim \mathcal{RP}(\lambda = n; \mathbf{P}_V)$ *with* $\mathbf{P}_V$ *being a symmetric* α*-stable law with scale parameter* $bn^{(-1/\alpha)}$*,*[10] *where* $\alpha \in (1, 2]$ *and* $b \in \mathbb{R}_+$*, that is,* $\widehat{\mathbf{P}}_V(\xi) = \exp\Big(-\frac{|b\xi|^\alpha}{n}\Big)$*. Then, we have*

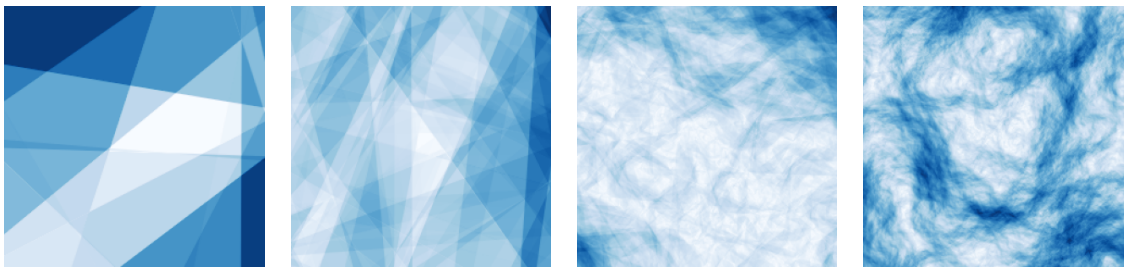
$$s^n_{\mathrm{ReLU}} \xrightarrow[n\to\infty]{\mathscr{L}} s^\infty_{\mathrm{ReLU}}, \tag{41}$$

where s^∞_{ReLU} *is a well-defined generalized stochastic process that takes values in* $\mathcal{S}'(\mathbb{R}^d)$ *and has the characteristic functional*

$$\widehat{\mathbf{P}}_{s^\infty_{\mathrm{ReLU}}}(\varphi) = \exp\Big(-|b|^\alpha \| \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}\|^\alpha_{L^\alpha}\Big), \quad \varphi \in \mathcal{S}(\mathbb{R}^d). \tag{42}$$

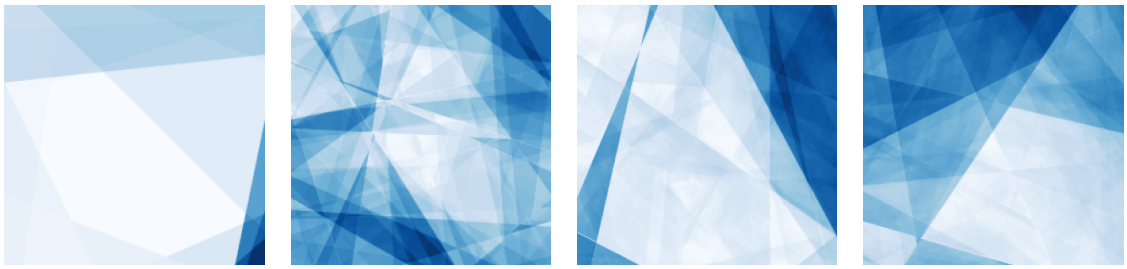
When $\alpha = 2$, the SαS law is the Gaussian law. In this case, we can deduce that s^∞_{ReLU} is indeed a Gaussian process (see Appendix C). On the other hand, for $\alpha \in (1, 2)$, we can readily see that s^∞_{ReLU} is non-Gaussian. Therefore, we have rigorously shown that wide limits of random neural networks are not necessarily Gaussian processes.

10. Similar to Neal (1996); Lee et al. (2018), the scale parameter inversely depends on the expected width of the network.



(a) $\lambda = 1$ (b) $\lambda = 10$ (c) $\lambda = 100$ (d) $\lambda = 1000$

Figure 1: $\mathbf{P}_V$ is Gaussian.



(a) $\lambda = 1$ (b) $\lambda = 10$ (c) $\lambda = 100$ (d) $\lambda = 1000$

Figure 2: $\mathbf{P}_V$ is symmetric ($\alpha = 1.25$)-stable.

We illustrate these observations numerically in Figures 1 and 2, where we generated random neural networks with $\mathbf{P}_V$ being Gaussian ($\alpha = 2$) and non-Gaussian ($\alpha = 1.25$), respectively. There, we plot a top-down view of realizations of random neural networks for $\lambda \in \{1, 10, 100, 1000\}$ where we color the linear regions with the magnitude of the gradient of the function. Figure 1(d) looks like a two-dimensional Gaussian process, while Figure 2(d) remains to look CPwL (non-Gaussian). Discussion on how we generated the random neural networks numerically along with some additional figures appear in Appendix D.

7. Conclusion

We have investigated the statistical properties of random ReLU neural networks. We proved that these networks are well-defined non-Gaussian processes in the non-asymptotic regime. We showed that these processes are isotropic and wide-sense self-similar with Hurst exponent 3/2. Remarkably, the autocovariances of these processes have simple closed-form expressions. Finally, we showed that, under suitable hypotheses, as the expected width tends to infinity, these processes can converge in law not only to Gaussian processes, but also to non-Gaussian processes depending on the law of the weights. These asymptotic results recover the classical observation that wide networks converge to Gaussian processes as well as prove that wide networks can converge to non-Gaussian processes. Although the presented investigation only considered shallow random ReLU neural networks, an important direction of future work would be to generalize our exact characterizations to deeper networks. To that end,

the techniques developed by Zavatone-Veth and Pehlevan (2021) could provide a starting point for that investigation.

Acknowledgments

The authors would like to thank the anonymous reviewers and the action editor for their careful reading of the manuscript. This work was supported in part by the Swiss National Science Foundation under Grant 200020_219356 / 1 and in part by the European Research Council (ERC Project FunLearn) under Grant 101020573.

Appendix A. Proof of Theorem 9

As preparation before the proof of Theorem 9, we collect and prove some intermediary results. To begin, we shall first prove that w_{Poi} is a well-defined stochastic process taking values in $\mathcal{S}'_{\mathscr{R}}$. Recall that w_{Poi} is an impulsive white noise that is realized by a Poisson-type random measure of the form

$$w_{\text{Poi}} = \sum_{k\in\mathbb{Z}} v_k \, \delta^{\mathsf{e}}_{(\boldsymbol{w}_k, b_k)}, \tag{43}$$

where $v_k \overset{\text{i.i.d.}}{\sim} \mathbf{P}_V$ for some admissible $\mathbf{P}_V$ (in the sense of Definition 5) and the collection of random variables $((\boldsymbol{w}_k, b_k))_{k\in\mathbb{Z}}$ is a (homogeneous) Poisson point process on $\mathbb{S}^{d-1}\times\mathbb{R}$ with rate parameter $\lambda > 0$. This point process satisfies the following properties.

1. The $(\boldsymbol{w}_k, b_k)$ are mutually independent.
2. For any measurable subset $\Pi \subset \mathbb{S}^{d-1}\times\mathbb{R}$, if we define the random variable

   $$N_\Pi = |\{(\boldsymbol{w}_k, b_k) : \, (\boldsymbol{w}_k, b_k) \in \Pi\}|, \tag{44}$$

   then

   $$\mathbf{P}(N_\Pi = n) = \frac{(\lambda|\Pi|)^n}{n!}\mathrm{e}^{-\lambda|\Pi|}, \tag{45}$$

   where $|\Pi|$ denotes the d-dimensional Hausdorff measure of Π. That is to say, N_Π is a Poisson random variable with mean $\lambda|\Pi|$.
3. For any measurable subset $B \subset \mathbb{S}^{d-1}\times\mathbb{R}$,

   $$\mathbf{P}((\boldsymbol{w}_k, b_k) \in B \mid (\boldsymbol{w}_k, b_k) \in \Pi) = \frac{|B\cap\Pi|}{|\Pi|}. \tag{46}$$

   That is to say, if a point lies in Π, then its location will be uniformly distributed on Π.

Lemma 12 *The random measure w_{Poi} can be viewed as a measurable mapping*

$$w_{\text{Poi}} : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathcal{S}'_{\mathscr{R}}, \mathcal{B}(\mathcal{S}'_{\mathscr{R}})) \tag{47}$$

with characteristic functional given by

$$\widehat{\mathbf{P}}_{w_{\text{Poi}}}(\psi) = \exp\left(\lambda \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \left(\mathrm{e}^{\mathrm{i}v\psi(\boldsymbol{u},t)} - 1\right) \mathrm{d}\boldsymbol{u}\, \mathrm{d}t\, \mathrm{d}\mathbf{P}_V(v)\right), \tag{48}$$

where $\mathrm{d}\boldsymbol{u}$ denotes integration against the surface measure on $\mathbb{S}^{d-1}$.

Proof Let $\mathcal{D}(\mathbb{R}^d) \subset \mathcal{S}(\mathbb{R}^d)$ denote the space of infinitely differentiable and compactly supported functions on $\mathbb{R}^d$. Let $\mathcal{D}_{\mathscr{R}} \coloneqq \mathscr{R}\big(\mathcal{D}(\mathbb{R}^d)\big)$ denote the range of the Radon transform on $\mathcal{D}(\mathbb{R}^d)$. We now summarize the properties of $\mathcal{D}_{\mathscr{R}}$ that are relevant for our problem (cf., Ludwig, 1966). First, $\mathcal{D}_{\mathscr{R}}$ is a closed subspace of $\mathcal{D}(\mathbb{S}^{d-1} \times \mathbb{R})$, the nuclear space of infinitely differentiable and compactly supported functions on $\mathbb{S}^{d-1} \times \mathbb{R}$ and is therefore nuclear. Furthermore, $\mathcal{D}_{\mathscr{R}}$ is dense in $\mathcal{S}_{\mathscr{R}}$, which implies that $\mathcal{S}'_{\mathscr{R}}$ is continuously embedded in $\mathcal{D}'_{\mathscr{R}}$. In particular, $\mathcal{D}_{\mathscr{R}}$ is the subspace of compactly supported functions in $\mathcal{S}_{\mathscr{R}}$.

Next, we shall prove that w_{Poi} can be viewed as a measurable mapping

$$w_{\text{Poi}} : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathcal{D}'_{\mathscr{R}}, \mathcal{B}_c(\mathcal{D}'_{\mathscr{R}})) \tag{49}$$

by computing its characteristic functional $\widehat{\mathbf{Q}}_{w_{\text{Poi}}}$ on $\mathcal{D}_{\mathscr{R}}$.[11] Let $\psi \in \mathcal{D}_{\mathscr{R}}$ and let

$$N_\psi = |\{(\boldsymbol{w}_k, b_k) : (\boldsymbol{w}_k, b_k) \in \operatorname{supp} \psi\}|. \tag{50}$$

We have, by definition, that

$$\langle w_{\text{Poi}}, \psi \rangle_{\mathcal{D}'_{\mathscr{R}} \times \mathcal{D}_{\mathscr{R}}} = \sum_{k=1}^{N_\psi} v'_k \psi(\boldsymbol{w}'_k, b'_k), \tag{51}$$

where we use an appropriate relabeling of $\{v_k, \boldsymbol{w}_k, b_k : (\boldsymbol{w}_k, b_k) \in \operatorname{supp} \psi\}$. Therefore,

$$\begin{aligned}
\widehat{\mathbf{Q}}_{w_{\text{Poi}}}(\psi) &= \mathbf{E}[\mathrm{e}^{\mathrm{i}\langle w_{\text{Poi}}, \psi \rangle_{\mathcal{D}'_{\mathscr{R}} \times \mathcal{D}_{\mathscr{R}}}}] \\
&= \mathbf{E}[\mathrm{e}^{\mathrm{i} \sum_{k=1}^{N_\psi} v'_k \psi(\boldsymbol{w}'_k, b'_k)}] \\
&= \mathbf{E}\left[\mathbf{E}\left[\prod_{k=1}^{N_\psi} \mathrm{e}^{\mathrm{i} v'_k \psi(\boldsymbol{w}'_k, b'_k)} \,\middle|\, N_\psi\right]\right] \\
&= \mathbf{E}\left[\prod_{k=1}^{N_\psi} \mathbf{E}\left[\mathrm{e}^{\mathrm{i} v'_k \psi(\boldsymbol{w}'_k, b'_k)} \,\middle|\, N_\psi\right]\right] \qquad (52) \\
&= \mathbf{E}\left[\prod_{k=1}^{N_\psi} \mathbf{E}\left[\mathrm{e}^{\mathrm{i} v'_k \psi(\boldsymbol{w}'_k, b'_k)}\right]\right] \\
&= \mathbf{E}\left[\prod_{k=1}^{N_\psi} \mathbf{E}\left[\mathbf{E}\left[\mathrm{e}^{\mathrm{i} v'_k \psi(\boldsymbol{w}'_k, b'_k)} \,\middle|\, v_k\right]\right]\right] \\
&= \mathbf{E}\left[\prod_{k=1}^{N_\psi} \mathbf{E}\left[\frac{1}{|\operatorname{supp} \psi|} \int_{\operatorname{supp} \psi} \mathrm{e}^{\mathrm{i} v'_k \psi(\boldsymbol{w}, b)} \, \mathrm{d}(\boldsymbol{w}, b)\right]\right] \qquad (53) \\
&= \mathbf{E}\left[\prod_{k=1}^{N_\psi} \frac{1}{|\operatorname{supp} \psi|} \int_{\mathbb{R}} \int_{\operatorname{supp} \psi} \mathrm{e}^{\mathrm{i} v \psi(\boldsymbol{w}, b)} \, \mathrm{d}(\boldsymbol{w}, b) \, \mathrm{d}\mathbf{P}_V(v)\right], \qquad (54)
\end{aligned}$$

11. Note that $\mathcal{D}_{\mathscr{R}}$ is not Fréchet so we use the cylindrical σ-algebra as opposed to the Borel σ-algebra.

where (52) holds by the mutual independence of the $(\boldsymbol{w}_k, b_k)$ and (53) holds since the random variables

$$(\boldsymbol{w}_k', b_k') \mid (\boldsymbol{w}_k', b_k') \in \operatorname{supp} \psi \tag{55}$$

are uniformly distributed on $\operatorname{supp} \psi$. Next, define the auxiliary functional

$$M(\psi) = \int_{\mathbb{R}} \int_{\operatorname{supp} \psi} \mathrm{e}^{\mathrm{i}v\psi(\boldsymbol{w},b)} \, \mathrm{d}(\boldsymbol{w}, b) \, \mathrm{d}\mathbf{P}_V(v). \tag{56}$$

We have that

$$\begin{aligned}
\widehat{\mathbf{Q}}_{w_{\mathrm{Poi}}}(\psi) &= \mathbf{E}\left[\prod_{k=1}^{N_\psi} \frac{M(\psi)}{|\operatorname{supp} \psi|}\right] \\
&= \mathbf{E}\left[\left(\frac{M(\psi)}{|\operatorname{supp} \psi|}\right)^{N_\psi}\right] \\
&= \sum_{n=0}^{\infty} \left(\frac{M(\psi)}{|\operatorname{supp} \psi|}\right)^n \frac{(\lambda|\operatorname{supp} \psi|)^n}{n!} \mathrm{e}^{-\lambda|\operatorname{supp} \psi|} \qquad (57) \\
&= \mathrm{e}^{-\lambda|\operatorname{supp} \psi|} \sum_{n=0}^{\infty} \frac{(\lambda M(\psi))^n}{n!} \\
&= \mathrm{e}^{-\lambda|\operatorname{supp} \psi|} \mathrm{e}^{\lambda M(\psi)} \qquad (58) \\
&= \exp(\lambda(M(\psi) - |\operatorname{supp} \psi|) \qquad (59) \\
&= \exp\left(\lambda \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1} \times \mathbb{R}} \left(\mathrm{e}^{\mathrm{i}v\psi(\boldsymbol{z})} - 1\right) \mathrm{d}\boldsymbol{z} \, \mathrm{d}\mathbf{P}_V(v)\right), \qquad (60)
\end{aligned}$$

where (57) holds since N_ψ is a Poisson random variable with mean $\lambda|\operatorname{supp} \psi|$, (58) holds by the Taylor series expansion of $t \mapsto \mathrm{e}^t$, (59) holds since $|\operatorname{supp} \psi| = \int_{\operatorname{supp} \psi} 1 \, \mathrm{d}\boldsymbol{z}$, and (60) holds since $\boldsymbol{z} \mapsto \mathrm{e}^{\mathrm{i}v\psi(\boldsymbol{z})} - 1$ vanishes outside $\operatorname{supp} \psi$. At this point, we remark that, since $\mathbf{P}_V$ is a Lévy measure (Definition 5), it is well-known that the form of (60) is continuous, positive definite, and satisfies $\widehat{\mathbf{Q}}_{w_{\mathrm{Poi}}}(0) = 1$ (see, e.g., Gelfand and Vilenkin, 1964, Theorem 2, p. 275). This implies that w_{Poi} is indeed a generalized stochastic process that takes values in $\mathcal{D}'_{\mathscr{R}}$.

To prove the lemma, it remains to extend the domain of $\widehat{\mathbf{Q}}_{w_{\mathrm{Poi}}}$ to $\mathcal{S}_{\mathscr{R}}$. To that end, let

$$\widehat{\mathbf{P}}_{w_{\mathrm{Poi}}}(\psi) = \exp\left(\lambda \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1} \times \mathbb{R}} \left(\mathrm{e}^{\mathrm{i}v\psi(\boldsymbol{z})} - 1\right) \mathrm{d}\boldsymbol{z} \, \mathrm{d}\mathbf{P}_V(v)\right), \quad \psi \in \mathcal{S}_{\mathscr{R}}. \tag{61}$$

We now invoke an adaption of Fageot et al. (2014, Theorem 3) which investigates impulsive white noise defined on $\mathbb{R}^d$ as a special case. Their theorem implies that, thanks to the admissibility conditions on $\mathbf{P}_V$ (Definition 5), the probability measures $\mathbf{Q}_{w_{\mathrm{Poi}}}$ and $\mathbf{P}_{w_{\mathrm{Poi}}}$ are compatible on $\mathcal{B}(\mathcal{S}'_{\mathscr{R}}) = \mathcal{B}_c(\mathcal{S}'_{\mathscr{R}}) \subset \mathcal{B}_c(\mathcal{D}'_{\mathscr{R}})$ in the sense that

$$\mathbf{Q}_{w_{\mathrm{Poi}}}(B) = \mathbf{P}_{w_{\mathrm{Poi}}}(B), \quad \text{for all } B \in \mathcal{B}(\mathcal{S}'_{\mathscr{R}}) \tag{62}$$

and $\mathbf{Q}_{w_{\mathrm{Poi}}}(\mathcal{D}'_{\mathscr{R}} \setminus \mathcal{S}'_{\mathscr{R}}) = 0$, which proves the lemma. ∎

Let $\mathcal{S}_\Delta(\mathbb{R}^d) \coloneqq \Delta\big(\mathcal{S}(\mathbb{R}^d)\big)$ denote the range of the Laplacian operator on $\mathcal{S}(\mathbb{R}^d)$. This is a closed subspace of $\mathcal{S}(\mathbb{R}^d)$. Observe that its dual $\mathcal{S}'_\Delta(\mathbb{R}^d)$ can be identified with the quotient space $\mathcal{S}'(\mathbb{R}^d)/\mathcal{N}_\Delta$, where

$$\mathcal{N}_\Delta = \{f \in \mathcal{S}'(\mathbb{R}^d) : \ \Delta f = 0 \Leftrightarrow \langle f, \phi\rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} = 0 \text{ for all } \phi \in \mathcal{S}_\Delta(\mathbb{R}^d)\}. \tag{63}$$

is the null space of the Laplacian operator. It is well-known that $\mathcal{N}_\Delta$ is infinite-dimensional and that its members are necessarily polynomials, the so-called *harmonic polynomials*. Therefore, the members of $\mathcal{S}'_\Delta(\mathbb{R}^d)$ are actually equivalence classes of the form

$$[f] = \{f + h : \ h \in \mathcal{N}_\Delta\} \in \mathcal{S}'_\Delta(\mathbb{R}^d), \tag{64}$$

where $f \in \mathcal{S}'(\mathbb{R}^d)$. With this notation, we now prove Theorem 9.

Proof [Proof of Theorem 9] Recall that $\mathrm{T}_{\mathrm{ReLU}} = \mathrm{K}\,\mathscr{R}\,\Delta$ and so $\mathrm{T}^*_{\mathrm{ReLU}} = \Delta\,\mathscr{R}^*\,\mathrm{K}$. Observe that, by Proposition 4,

$$\mathrm{T}^*_{\mathrm{ReLU}} : \mathcal{S}_{\mathscr{R}} \to \mathcal{S}_\Delta(\mathbb{R}^d) \tag{65}$$

is a continuous bijection, where we equip the closed subspaces $\mathcal{S}_{\mathscr{R}} \subset \mathcal{S}(\mathbb{S}^{d-1}\times\mathbb{R})$ and $\mathcal{S}_\Delta(\mathbb{R}^d) \subset \mathcal{S}(\mathbb{R}^d)$ with the subspace topology from their respective parent Fréchet spaces. By the open mapping theorem for Fréchet spaces (see, e.g., Rudin, 1991, Theorem 2.11), there exists a continuous inverse operator

$$\mathrm{T}^{*-}_{\mathrm{ReLU}} : \mathcal{S}_\Delta(\mathbb{R}^d) \to \mathcal{S}_{\mathscr{R}} \tag{66}$$

with the properties that that $\mathrm{T}^*_{\mathrm{ReLU}}\,\mathrm{T}^{*-}_{\mathrm{ReLU}} = \mathrm{Id}$ on $\mathcal{S}_\Delta(\mathbb{R}^d)$ and $\mathrm{T}^{*-}_{\mathrm{ReLU}}\,\mathrm{T}^*_{\mathrm{ReLU}} = \mathrm{Id}$ on $\mathcal{S}_{\mathscr{R}}$. Therefore, by duality, we have the continuous bijections

$$\begin{aligned} \mathrm{T}_{\mathrm{ReLU}} &: \mathcal{S}'_\Delta(\mathbb{R}^d) \to \mathcal{S}_{\mathscr{R}} \\ \mathrm{T}^{-}_{\mathrm{ReLU}} &: \mathcal{S}'_{\mathscr{R}} \to \mathcal{S}'_\Delta(\mathbb{R}^d), \end{aligned} \tag{67}$$

where we recall that $\mathcal{S}'_\Delta(\mathbb{R}^d) \cong \mathcal{S}'(\mathbb{R}^d)/\mathcal{N}_\Delta$.

Next, we note that the operator

$$\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}} : \varphi \mapsto \int_{\mathbb{R}^d} k^\varepsilon_{\boldsymbol{x}}(\cdot)\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{68}$$

specified in (33) continuously maps $\mathcal{S}_\Delta(\mathbb{R}^d) \to \mathcal{S}_{\mathscr{R}}$ (cf., Parhi and Unser, 2025, Equation (A.3)). Observe that, by Proposition 7, its extension by duality $\mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}} : \mathcal{S}'_{\mathscr{R}} \to \mathcal{S}'_\Delta(\mathbb{R}^d)$ coincides with $\mathrm{T}^{-}_{\mathrm{ReLU}}$. In particular, $\mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}$ imposes the boundary conditions from (24) on the affine component of the harmonic polynomials in the equivalence classes in $\mathcal{S}'_\Delta(\mathbb{R}^d)$. Said differently, the range space $\mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}\big(\mathcal{S}'_{\mathscr{R}}\big)$ is the closed subspace of $\mathcal{S}'_\Delta(\mathbb{R}^d)$ whose equivalence class members $[s] \in \mathcal{S}'_\Delta(\mathbb{R}^d)$ additionally satisfy

$$[(\partial^{\boldsymbol{m}} g^\varepsilon_d) * s_0](\mathbf{0}) = 0, |\boldsymbol{m}| \le 1 \tag{69}$$

for all $s_0 \in [s]$. Therefore, we can rewrite the SDE (34) as

$$s \overset{\mathscr{L}}{=} \mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}\, w_{\mathrm{Poi}}, \tag{70}$$

where the equality is understood in $\mathcal{S}'_\Delta(\mathbb{R}^d)$, i.e.,

$$\langle s, \phi \rangle_{\mathcal{S}'_\Delta(\mathbb{R}^d)\times\mathcal{S}_\Delta(\mathbb{R}^d)} \overset{\mathscr{L}}{=} \langle \mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}\, w_{\mathrm{Poi}}, \phi \rangle_{\mathcal{S}'_\Delta(\mathbb{R}^d)\times\mathcal{S}_\Delta(\mathbb{R}^d)} = \langle w_{\mathrm{Poi}}, \mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\, \phi \rangle_{\mathcal{S}'_{\mathscr{R}}\times\mathcal{S}_{\mathscr{R}}}, \tag{71}$$

for all $\phi \in \mathcal{S}_\Delta(\mathbb{R}^d)$. The above equality implies that the characteristic functional of *any solution* s to (70) (and, subsequently, the original SDE (34)) takes the form

$$\widehat{\mathbf{P}}_s(\phi) = \widehat{\mathbf{P}}_{\mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}\, w_{\mathrm{Poi}}}(\phi) = \widehat{\mathbf{P}}_{w_{\mathrm{Poi}}}(\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\, \phi). \tag{72}$$

This characteristic functional is well-defined for any $\phi \in \mathcal{S}_\Delta(\mathbb{R}^d)$ since $\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\, \phi \in \mathcal{S}_{\mathscr{R}}$, which ensures that the right-hand side is well-defined by Lemma 12.

Since $s^\varepsilon_{\mathrm{ReLU}} \coloneqq \mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}\, w_{\mathrm{Poi}}$ via the computation in (28), we see that $s^\varepsilon_{\mathrm{ReLU}}$ is one member in an equivalence class in $\mathcal{S}'(\mathbb{R}^d)/\mathcal{N}_\Delta$. In particular, this implies that $s_{\mathrm{ReLU}} \in \mathcal{S}'(\mathbb{R}^d)$ and that the equivalence class $[s^\varepsilon_{\mathrm{ReLU}}] = \{s^\varepsilon_{\mathrm{ReLU}} + h : \, h \in \mathcal{N}_\Delta\}$ is a well-defined stochastic process that takes values in $\mathcal{S}'_\Delta(\mathbb{R}^d) \cong \mathcal{S}'(\mathbb{R}^d)/\mathcal{N}_\Delta$ whose characteristic functional on $\mathcal{S}_\Delta$ is given by (72). Equivalently stated, the full set of tempered weak solutions of the SDE (34) has members that necessarily take the form $s^\varepsilon_{\mathrm{ReLU}} + h$, where $h \in \mathcal{N}_\Delta$ is a harmonic polynomial of degree ≥ 2 (since boundary conditions of the SDE, imposed by $\mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}$, force the affine component of all solutions to be the same). Consequently, from these boundary conditions, we readily see that the only CPwL solution to the SDE is $s^\varepsilon_{\mathrm{ReLU}}$.

To complete the proof we need to derive the form of the characteristic functional of $s^\varepsilon_{\mathrm{ReLU}}$ on the larger space $\mathcal{S}(\mathbb{R}^d) \supset \mathcal{S}_\Delta(\mathbb{R}^d)$. For any $\varphi \in \mathcal{S}(\mathbb{R}^d)$, we have that

$$\langle s^\varepsilon_{\mathrm{ReLU}}, \varphi \rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} \overset{\mathscr{L}}{=} \langle \mathrm{T}^{\dagger\varepsilon}_{\mathrm{ReLU}}\, w_{\mathrm{Poi}}, \varphi \rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} \tag{73}$$

From the expression of the kernel $(\boldsymbol{u}, t) \mapsto k^\varepsilon_{\boldsymbol{x}}(\boldsymbol{u}, t)$ in (26) we see that (i) it is continuous in the variables $(\boldsymbol{u}, t) \in \mathbb{S}^{d-1} \times \mathbb{R}$ and (ii) it decays faster than any polynomial in the t-variable. Therefore, for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$, the function $\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\{\varphi\}$ is a continuous function in $(\boldsymbol{u}, t) \in \mathbb{S}^{d-1} \times \mathbb{R}$ that decays faster than any polynomial in the t-variable. In particular, this ensures that, for any $1 \leq p \leq \infty$, the map

$$\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}} : \mathcal{S}(\mathbb{R}^d) \to L^p(\mathbb{S}^{d-1} \times \mathbb{R}) \tag{74}$$

is continuous.

The right-hand side of (73) is, by definition, the integration of $\mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\{\varphi\}$ against the locally finite Radon measure w_{Poi}, i.e., for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ we have that

$$\begin{aligned}
\langle s^\varepsilon_{\mathrm{ReLU}}, \varphi \rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} &\overset{\mathscr{L}}{=} \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\{\varphi\}\, \mathrm{d}\left(\sum_{k\in\mathbb{Z}} v_k\, \delta^{\mathsf{e}}_{(\boldsymbol{w}_k, b_k)} \right) \\
&= \sum_{k\in\mathbb{Z}} v_k \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\{\varphi\}\, \mathrm{d}\delta^{\mathsf{e}}_{(\boldsymbol{w}_k, b_k)} \\
&= \sum_{k\in\mathbb{Z}} v_k\, \mathrm{T}^{\dagger\varepsilon *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{w}_k, b_k),
\end{aligned} \tag{75}$$

where interchanging of the integral and sum in the second line is well-defined due to the regularity of $\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}}\{\varphi\}$ and the third line uses the fact that the range of $\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}}$ on $\mathcal{S}(\mathbb{R}^d)$ is a space of even functions. This proves that

$$\begin{aligned}&\widehat{\mathbf{P}}_{s^{\varepsilon}_{\mathrm{ReLU}}}(\varphi)\\ &= \widehat{\mathbf{P}}_{w_{\mathrm{Poi}}}(\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}}\,\varphi) = \exp\biggl(\lambda \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Bigl(\mathrm{e}^{\mathrm{i}v\,\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u},t)} - 1\Bigr)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}t\,\mathrm{d}\mathbf{P}_V(v)\biggr),\end{aligned} \tag{76}$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$, where the last equality comes from Lemma 12. We shall now verify that $\widehat{\mathbf{P}}_{s^{\varepsilon}_{\mathrm{ReLU}}}$ is a valid characteristic functional on $\mathcal{S}(\mathbb{R}^d)$. This then implies that $s^{\varepsilon}_{\mathrm{ReLU}} : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathcal{S}'(\mathbb{R}^d), \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$ is a measurable mapping and therefore a well-defined stochastic process.

Observe that the second admissibility condition on $\mathbf{P}_V$ (Item 2 in Definition 5) states that $\mathbf{P}_V$ has a finite absolute moment. This is a sufficient condition to ensure that this characteristic functional (76) is well-defined for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Indeed, we have that

$$\Psi(\xi) \coloneqq \lambda \int_{\mathbb{R}} \Bigl(\mathrm{e}^{\mathrm{i}v\xi} - 1\Bigr)\,\mathrm{d}\mathbf{P}_V(v) \leq \lambda\,|\xi|\,\mathbf{E}[|V|], \tag{77}$$

where $V \sim \mathbf{P}_V$ (cf., Unser et al., 2014, p. 1952). Therefore,

$$\begin{aligned}\widehat{\mathbf{P}}_{s^{\varepsilon}_{\mathrm{ReLU}}}(\varphi) &= \exp\biggl(\lambda \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Bigl(\mathrm{e}^{\mathrm{i}v\,\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u},t)} - 1\Bigr)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}t\,\mathrm{d}\mathbf{P}_V(v)\biggr)\\ &= \exp\biggl(\int_{\mathbb{S}^{d-1}\times\mathbb{R}} \Psi\Bigl(\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}}\{\varphi\}\Bigr)\,\mathrm{d}\boldsymbol{z}\biggr)\\ &\leq \exp\Bigl(C\,\|\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}}\{\varphi\}\|_{L^1}\Bigr)\\ &< \infty,\end{aligned} \tag{78}$$

for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$, where $C = \lambda\,\mathbf{E}[|V|] < \infty$, where in the last line we used (74) with $p = 1$. Since $\mathrm{T}^{\dagger\varepsilon*}_{\mathrm{ReLU}} : \mathcal{S}(\mathbb{R}^d) \to L^1(\mathbb{S}^{d-1}\times\mathbb{R})$ linearly and continuously, Proposition 3.1 of Fageot and Unser (2019) then guarantees that $\widehat{\mathbf{P}}_{s^{\varepsilon}_{\mathrm{ReLU}}}$ is continuous, positive definite, and satisfies $\widehat{\mathbf{P}}_{s^{\varepsilon}_{\mathrm{ReLU}}}(0) = 1$. Therefore, the Bochner–Minlos theorem ensures that $\widehat{\mathbf{P}}_{s^{\varepsilon}_{\mathrm{ReLU}}}$ is the characteristic functional of the well-defined stochastic process $s^{\varepsilon}_{\mathrm{ReLU}}$.

In the limiting scenario of $\varepsilon \to 0$, we see that the random neural network $s_{\mathrm{ReLU}} \sim \mathcal{RP}(\lambda; \mathbf{P}_V)$ is a measurable mapping $(\Omega, \mathcal{F}, \mathbf{P}) \to (\mathcal{S}'(\mathbb{R}^d), \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$ whose characteristic functional is

$$\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi) = \exp\biggl(\lambda \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Bigl(\mathrm{e}^{\mathrm{i}v\,\mathrm{T}^{\dagger*}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u},t)} - 1\Bigr)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}t\,\mathrm{d}\mathbf{P}_V(v)\biggr), \quad \varphi \in \mathcal{S}(\mathbb{R}^d), \tag{79}$$

where we observe that this limiting characteristic functional remains to be valid in the sense of the Bochner–Minlos theorem since the property that, for any $1 \leq p \leq \infty$, the map

$$\mathrm{T}^{\dagger*}_{\mathrm{ReLU}} : \mathcal{S}(\mathbb{R}^d) \to L^p(\mathbb{S}^{d-1}\times\mathbb{R}) \tag{80}$$

is continuous, remains to be true since $k_{\boldsymbol{x}} = \lim_{\varepsilon \to 0} k_{\boldsymbol{x}}^{\varepsilon}$ (pointwise limt) is compactly supported. Consequently, s_{ReLU} is the unique CPwL solution to SDE (36).[12] ■

Appendix B. Proof of Theorem 10

Proof

1. Thanks to the moment generating properties of the characteristic functional (Gelfand and Vilenkin, 1964), the mean functional of s_{ReLU} can be obtained as

$$\mu_{s_{\mathrm{ReLU}}}(\varphi) = \mathbf{E}[\langle s_{\mathrm{ReLU}}, \varphi \rangle_{\mathcal{S}'(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d)}] = (-\mathrm{i}) \frac{\mathrm{d}}{\mathrm{d}\xi} \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi \varphi) \Big|_{\xi = 0}, \tag{81}$$

where $\varphi \in \mathcal{S}(\mathbb{R}^d)$. First, observe that the characteristic functional of s_{ReLU} can be written as

$$\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi) = \exp\left(\int_{\mathbb{S}^{d-1} \times \mathbb{R}} \Psi\left(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{z}) \right) \mathrm{d}\boldsymbol{z} \right) \tag{82}$$

with Ψ defined as in (77). Here, note that we have

$$\Psi'(x) = \mathrm{i}\lambda \int_{\mathbb{R}} v \mathrm{e}^{\mathrm{i}vx} \, \mathrm{d}\mathbf{P}_V(v). \tag{83}$$

Let us denote $h(\boldsymbol{z}) = \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{z})$. By applying the chain rule, we can write

$$\frac{\mathrm{d}}{\mathrm{d}\xi} \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi \varphi) = \exp\left(\int_{\mathbb{R} \times \mathbb{S}^{d-1}} \Psi(\xi h(\boldsymbol{z})) \, \mathrm{d}\boldsymbol{z} \right) \cdot \int_{\mathbb{R} \times \mathbb{S}^{d-1}} \Psi'(\xi h(\boldsymbol{z})) h(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}. \tag{84}$$

On setting $\xi = 0$, we get

$$\frac{\mathrm{d}}{\mathrm{d}\xi} \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi \varphi) \Big|_{\xi = 0} = \Psi'(0) \int_{\mathbb{R} \times \mathbb{S}^{d-1}} h(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \tag{85}$$

as $\Psi(0) = 0$. Therefore, the mean functional is

$$\begin{aligned} \mu_{s_{\mathrm{ReLU}}}(\varphi) &= (-\mathrm{i}) \frac{\mathrm{d}}{\mathrm{d}\xi} \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi \varphi) \Big|_{\xi = 0} \\ &= \lambda \mathbf{E}[V] \int_{\mathbb{S}^{d-1} \times \mathbb{R}} \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \\ &= \lambda \mathbf{E}[V] \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} k_{\boldsymbol{x}}(\boldsymbol{u}, t) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{u} \, \mathrm{d}t \, \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{86}$$

Next, we establish a link between the mean functional of s_{ReLU} and the quantity $\mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})]$. Since s_{ReLU} has a pointwise interpretation, we have

$$\langle s_{\mathrm{ReLU}}, \varphi \rangle_{\mathcal{S}'(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d)} = \int_{\mathbb{R}^d} s_{\mathrm{ReLU}}(\boldsymbol{x}) \varphi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \tag{87}$$

12. The mollifier argument is necessary in order to make sense of the boundary conditions (36) for elements of $\mathcal{S}'(\mathbb{R}^d)$ that are not regular enough for the derivatives to exist.

Consequently, the mean functional can also be computed as

$$\mu_{s_{\mathrm{ReLU}}}(\varphi) = \mathbf{E}[\langle s_{\mathrm{ReLU}}, \varphi\rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)}] = \mathbf{E}\bigg[\int_{\mathbb{R}^d} s_{\mathrm{ReLU}}(\boldsymbol{x})\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\bigg] = \int_{\mathbb{R}^d} \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})]\varphi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \tag{88}$$

where exchanging the expectation and the integral is justified by the Fubini–Tonelli theorem since the integrand in (86) is absolutely integrable by (80) with $p = 1$. On comparing (88) with (86), we see that

$$\mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})] = \lambda\mathbf{E}[V]\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} k_{\boldsymbol{x}}(\boldsymbol{u}, t)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}t. \tag{89}$$

2. The covariance functional of s_{ReLU} is given by

$$\begin{aligned}&\Sigma_{s_{\mathrm{ReLU}}}(\varphi_1, \varphi_2)\\ &= \mathbf{E}\Big[\Big(\langle s_{\mathrm{ReLU}}, \varphi_1\rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} - \mu_{s_{\mathrm{ReLU}}}(\varphi_1)\Big)\Big(\langle s_{\mathrm{ReLU}}, \varphi_2\rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} - \mu_{s_{\mathrm{ReLU}}}(\varphi_2)\Big)\Big]\\ &= \mathcal{R}_{s_{\mathrm{ReLU}}}(\varphi_1, \varphi_2) - \mu_{s_{\mathrm{ReLU}}}(\varphi_1)\mu_{s_{\mathrm{ReLU}}}(\varphi_2),\end{aligned} \tag{90}$$

where $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$ and

$$\mathcal{R}_{s_{\mathrm{ReLU}}}(\varphi_1, \varphi_2) = \mathbf{E}\Big[\langle s_{\mathrm{ReLU}}, \varphi_1\rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)}\ \langle s_{\mathrm{ReLU}}, \varphi_2\rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)}\Big] \tag{91}$$

is the correlation functional of s_{ReLU}. This quantity can be computed from its characteristic functional (cf., Gelfand and Vilenkin, 1964) as

$$\mathcal{R}_{s_{\mathrm{ReLU}}}(\varphi_1, \varphi_2) = -\frac{\mathrm{d}^2}{\mathrm{d}\xi_1\mathrm{d}\xi_2}\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi_1\varphi_1 + \xi_2\varphi_2)\Big|_{\xi_1=0,\xi_2=0}. \tag{92}$$

Let us first define the quantity $f(\xi_1, \xi_2)$ as

$$\begin{aligned}f(\xi_1, \xi_2) &= \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \Psi\Big(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\xi_1\varphi_1 + \xi_2\varphi_2\}(\boldsymbol{z})\Big)\,\mathrm{d}\boldsymbol{z}\\ &= \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \Psi\Big(\xi_1\,\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_1\}(\boldsymbol{z}) + \xi_2\,\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_2\}(\boldsymbol{z})\Big)\,\mathrm{d}\boldsymbol{z}.\end{aligned} \tag{93}$$

Further, let us denote $h_1(\boldsymbol{z}) = \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_1\}(\boldsymbol{z})$ and $h_2(\boldsymbol{z}) = \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_2\}(\boldsymbol{z})$. By applying the chain rule twice, we write

$$\begin{aligned}&\frac{\mathrm{d}^2}{\mathrm{d}\xi_1\mathrm{d}\xi_2}\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi_1\varphi_1 + \xi_2\varphi_2)\\ &\qquad = \exp(f(\xi_1, \xi_2))\bigg(\frac{\mathrm{d}^2}{\mathrm{d}\xi_1\mathrm{d}\xi_2}f(\xi_1, \xi_2) + \frac{\mathrm{d}}{\mathrm{d}\xi_1}f(\xi_1, \xi_2)\frac{\mathrm{d}}{\mathrm{d}\xi_2}f(\xi_1, \xi_2)\bigg),\end{aligned} \tag{94}$$

where

$$\frac{\mathrm{d}}{\mathrm{d}\xi_k} f(\xi_1, \xi_2) = \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \Psi'(\xi_1 h_1(\boldsymbol{z}) + \xi_2 h_2(\boldsymbol{z})) h_k(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} \tag{95}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}\xi_1 \mathrm{d}\xi_2} f(\xi_1, \xi_2) = \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \Psi''(\xi_1 h_1(\boldsymbol{z}) + \xi_2 h_2(\boldsymbol{z})) h_1(\boldsymbol{z}) h_2(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z}. \tag{96}$$

On setting $\xi_1 = 0$ and $\xi_2 = 0$, we get

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}\xi_1 \mathrm{d}\xi_2} \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi_1\varphi_1 + \xi_2\varphi_2)\Big|_{\xi_1=0,\xi_2=0} = \ & \Psi''(0) \int_{\mathbb{S}^{d-1}\times\mathbb{R}} h_1(\boldsymbol{z}) h_2(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} \\ & + \left(\Psi'(0) \int_{\mathbb{S}^{d-1}\times\mathbb{R}} h_1(\boldsymbol{z}) \mathrm{d}\boldsymbol{z}\right)\left(\Psi'(0) \int_{\mathbb{S}^{d-1}\times\mathbb{R}} h_2(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z}\right).\end{aligned} \tag{97}$$

Note that we have

$$\Psi''(x) = -\lambda \int_{\mathbb{R}} v^2 \mathrm{e}^{\mathrm{i}vx}\, \mathrm{d}\mathbf{P}_V(v). \tag{98}$$

Thus, the correlation functional is of the form

$$\begin{aligned}\mathcal{R}_{s_{\mathrm{ReLU}}}(\varphi_1, \varphi_2) &= -\frac{\mathrm{d}^2}{\mathrm{d}\xi_1 \mathrm{d}\xi_2} \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\xi_1\varphi_1 + \xi_2\varphi_2)\Big|_{\xi_1=0,\xi_2=0} \\ &= \lambda \mathbf{E}[V^2] \left(\int_{\mathbb{S}^{d-1}\times\mathbb{R}} \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_1\}(\boldsymbol{z})\, \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_2\}(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} \right) \\ &\quad + \mu_{s_{\mathrm{ReLU}}}(\varphi_1) \mu_{s_{\mathrm{ReLU}}}(\varphi_2).\end{aligned} \tag{99}$$

Consequently, the covariance functional is given by

$$\begin{aligned}&\Sigma_{s_{\mathrm{ReLU}}}(\varphi_1, \varphi_2) \\ &= \lambda \mathbf{E}[V^2] \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_1\}(\boldsymbol{z})\, \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_2\}(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} \\ &= \lambda \mathbf{E}[V^2] \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} k_{\boldsymbol{x}}(\boldsymbol{u}, t) k_{\boldsymbol{y}}(\boldsymbol{u}, t) \varphi_1(\boldsymbol{x}) \varphi_2(\boldsymbol{y})\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}t\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}.\end{aligned} \tag{100}$$

Next, we derive the connection between the covariance functional of s_{ReLU} and the autocovariance $\mathbf{E}[(s_{\mathrm{ReLU}}(\boldsymbol{x}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})])(s_{\mathrm{ReLU}}(\boldsymbol{y}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{y})])]$. Since s_{ReLU} has a pointwise interpretation, the covariance functional can also be computed as

$$\begin{aligned}&\Sigma_{s_{\mathrm{ReLU}}}(\varphi_1, \varphi_2) \\ &= \mathbf{E}\Big[\Big(\langle s_{\mathrm{ReLU}}, \varphi_1 \rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} - \mu_{s_{\mathrm{ReLU}}}(\varphi_1) \Big) \Big(\langle s_{\mathrm{ReLU}}, \varphi_2 \rangle_{\mathcal{S}'(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d)} - \mu_{s_{\mathrm{ReLU}}}(\varphi_2) \Big) \Big] \\ &= \mathbf{E}\left[\left(\int_{\mathbb{R}^d} (s_{\mathrm{ReLU}}(\boldsymbol{x}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})]) \varphi_1(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \right) \left(\int_{\mathbb{R}^d} (s_{\mathrm{ReLU}}(\boldsymbol{y}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{y})]) \varphi_2(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y} \right) \right] \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{E}[(s_{\mathrm{ReLU}}(\boldsymbol{x}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})])(s_{\mathrm{ReLU}}(\boldsymbol{y}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{y})])] \varphi_1(\boldsymbol{x}) \varphi_2(\boldsymbol{y})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y},\end{aligned} \tag{101}$$

where exchanging the expectation and the integral is justified by the Fubini–Tonelli theorem since the integrand in (100) is absolutely integrable from (80) with $p = 2$. If we compare (101) with (100), we see that

$$\begin{aligned} C_{s_{\mathrm{ReLU}}}(\boldsymbol{x}, \boldsymbol{y}) &= \mathbf{E}[(s_{\mathrm{ReLU}}(\boldsymbol{x}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})])(s_{\mathrm{ReLU}}(\boldsymbol{y}) - \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{y})])] \\ &= \lambda \mathbf{E}[V^2] \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} k_{\boldsymbol{x}}(\boldsymbol{u}, t) k_{\boldsymbol{y}}(\boldsymbol{u}, t) \, \mathrm{d}\boldsymbol{u} \, \mathrm{d}t. \end{aligned} \tag{102}$$

To simplify the double integral in (102), we first observe that, by definition,

$$(\boldsymbol{x}, \boldsymbol{y}) \mapsto \int_{\mathbb{R}} \int_{\mathbb{S}^{d-1}} k_{\boldsymbol{x}}(\boldsymbol{u}, t) k_{\boldsymbol{y}}(\boldsymbol{u}, t) \, \mathrm{d}\boldsymbol{u} \, \mathrm{d}t, \quad (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^d \times \mathbb{R}^d, \tag{103}$$

is the (Schwartz) kernel of the operator $\mathrm{T}^{\dagger}_{\mathrm{ReLU}} \, \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}$. Next, we note that the right-inverse operator can be equivalently specified as the composition of operators $\mathrm{T}^{\dagger}_{\mathrm{ReLU}} = (\mathrm{Id} - \mathrm{P}) \Delta^{-1} \mathscr{R}^*$ (cf. Unser, 2023, Equation (57)), where Δ^{-1} is the Riesz potential of order 2, i.e., it is the Fourier multiplier

$$\widehat{(-\Delta)^{-\frac{\gamma}{2}} f}(\boldsymbol{\omega}) = \|\boldsymbol{\omega}\|_2^{-\gamma} \, \widehat{f}(\boldsymbol{\omega}), \quad \boldsymbol{\omega} \in \mathbb{R}^d, \tag{104}$$

with $\gamma = 2$, and P is the projection onto the space of affine functions adapted to the boundary conditions of the SDE (36). Concretely,

$$\mathrm{P}\{f\} = \sum_{n=0}^{d} \langle \phi_n, f \rangle p_n, \tag{105}$$

where $p_0(\boldsymbol{x}) = 1$ and $p_n(\boldsymbol{x}) = x_n$, $n = 1, \ldots, d$ is a basis for the space of affine function on $\mathbb{R}^d$ and $\phi_0 = \delta$ (Dirac distribution) and $\phi_n = -\delta'_n \coloneqq -\partial_{x_n} \delta$, $n = 1, \ldots, d$, is the linear functional that evaluates the partial derivative in the nth component at $\mathbf{0}$, i.e., $\langle \phi_n, f \rangle = \partial_{x_n} f(\mathbf{0})$, $n = 1, \ldots, d$. Consequently, the adjoint projector is given by

$$\mathrm{P}^*\{f\} = \sum_{n=0}^{d} \langle p_n, f \rangle \phi_n. \tag{106}$$

With this notation, we have that

$$\begin{aligned} \mathrm{T}^{\dagger}_{\mathrm{ReLU}} \, \mathrm{T}^{\dagger *}_{\mathrm{ReLU}} &= (\mathrm{Id} - \mathrm{P}) \Delta^{-1} \mathscr{R}^* \mathscr{R} \, \Delta^{-1} (\mathrm{Id} - \mathrm{P}^*) \\ &= (\mathrm{Id} - \mathrm{P}) \Delta^{-1} (-\Delta)^{-\frac{d-1}{2}} \Delta^{-1} (\mathrm{Id} - \mathrm{P}^*) \\ &= (\mathrm{Id} - \mathrm{P}) (-\Delta)^{-\frac{d+3}{2}} (\mathrm{Id} - \mathrm{P}^*), \end{aligned} \tag{107}$$

where the second line follows from Proposition 4. The (Schwartz) kernel of the operator (generalized impulse response) can be identified with $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \mathrm{T}^{\dagger}_{\mathrm{ReLU}} \, \mathrm{T}^{\dagger *}_{\mathrm{ReLU}} \{\delta(\cdot - \boldsymbol{y})\}(\boldsymbol{x})$. We have that

$$(\mathrm{Id} - \mathrm{P}^*)\{\delta(\cdot - \boldsymbol{y})\} = \delta(\cdot - \boldsymbol{y}) - \sum_{k=0}^{d} \langle p_k, \delta(\cdot - \boldsymbol{y}) \rangle \phi_n = \delta(\cdot - \boldsymbol{y}) - \delta + \sum_{n=1}^{d} y_1 \delta'_n, \tag{108}$$

where we used the property that the shifted Dirac distribution is the sampling functional. Next,

$$(-\Delta)^{-\frac{d+3}{2}}(\mathrm{Id}-\mathrm{P}^*)\{\delta(\cdot-\boldsymbol{y})\}(\boldsymbol{x}) = A\left(\|\boldsymbol{x}-\boldsymbol{y}\|_2^3 - \|\boldsymbol{x}\|_2^3 + \sum_{n=1}^{d} y_1(3x_n\|\boldsymbol{x}\|_2)\right)$$
$$= A\Big(\|\boldsymbol{x}-\boldsymbol{y}\|_2^3 - \|\boldsymbol{x}\|_2^3 + 3\boldsymbol{x}^\mathsf{T}\boldsymbol{y}\|\boldsymbol{x}\|_2\Big), \tag{109}$$

where $A = \frac{\Gamma(-3/2)}{2^{d+3}\pi^{d/2}\Gamma((d+3)/2)}$ and we used the fact that $\boldsymbol{x} \mapsto A\|\boldsymbol{x}\|_2^3$ is the radially symmetric Green's function of $(-\Delta)^{\frac{d+3}{2}}$ (Gelfand and Shilov, 1964). Finally,

$$\begin{aligned}
&(\mathrm{Id}-\mathrm{P})(-\Delta)^{-\frac{d+3}{2}}(\mathrm{Id}-\mathrm{P}^*)\{\delta(\cdot-\boldsymbol{y})\}(\boldsymbol{x}) \\
&= A\Big(\|\boldsymbol{x}-\boldsymbol{y}\|_2^3 - \|\boldsymbol{x}\|_2^3 + 3\boldsymbol{x}^\mathsf{T}\boldsymbol{y}\|\boldsymbol{x}\|_2\Big) - A\left(\|\boldsymbol{y}\|_2^3 - \sum_{n=1}^{d} 3y_n\|\boldsymbol{y}\|_2 x_n\right) \\
&= A\Big(\|\boldsymbol{x}-\boldsymbol{y}\|_2^3 - \|\boldsymbol{x}\|_2^3 - \|\boldsymbol{y}\|_2^3 + 3\boldsymbol{x}^\mathsf{T}\boldsymbol{y}(\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)\Big).
\end{aligned} \tag{110}$$

Putting everything together, we find that the autocovariance takes the form

$$C_{s_{\mathrm{ReLU}}}(\boldsymbol{x},\boldsymbol{y}) = \lambda A\mathbf{E}[V^2]\Big(\|\boldsymbol{x}-\boldsymbol{y}\|_2^3 - \|\boldsymbol{x}\|_2^3 - \|\boldsymbol{y}\|_2^3 + 3\boldsymbol{x}^\mathsf{T}\boldsymbol{y}(\|\boldsymbol{x}\|_2 + \|\boldsymbol{y}\|_2)\Big). \tag{111}$$

3. In order to show that s_{ReLU} is isotropic, we will show that its characteristic functional satisfies

$$\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi) = \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi(\mathbf{U}\cdot)) \tag{112}$$

for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and any $(d \times d)$ rotation matrix $\mathbf{U}$. First, we note that the kernel of $\mathrm{T}^\dagger_{\mathrm{ReLU}}$ can be written as

$$\begin{aligned}
k_{\boldsymbol{x}}(\boldsymbol{u},t) &= \mathrm{ReLU}(\boldsymbol{u}^\mathsf{T}\boldsymbol{x}-t) - \frac{(\boldsymbol{u}^\mathsf{T}\boldsymbol{x}-t)}{2} - \frac{|t|}{2} + (\boldsymbol{u}^\mathsf{T}\boldsymbol{x})\frac{\mathrm{sgn}(t)}{2} \\
&= \mathrm{ReLU}(\boldsymbol{u}^\mathsf{T}\boldsymbol{x}-t) + (\boldsymbol{u}^\mathsf{T}\boldsymbol{x})h_1(t) + h_2(t),
\end{aligned} \tag{113}$$

where $h_1(t) = \frac{\mathrm{sgn}(t)-1}{2}$ and $h_2(t) = \frac{t-|t|}{2}$. Let $\mathbf{U}$ be a $(d \times d)$ rotation matrix. Then, we have

$$\begin{aligned}
\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi(\mathbf{U}\cdot)\}(\boldsymbol{u},t) &= \int_{\mathbb{R}^d} k_{\boldsymbol{x}}(\boldsymbol{u},t)\varphi(\mathbf{U}\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^d} k_{\mathbf{U}^\mathsf{T}\widetilde{\boldsymbol{x}}}(\boldsymbol{u},t)\varphi(\widetilde{\boldsymbol{x}})\,\mathrm{d}\widetilde{\boldsymbol{x}} &(114)\\
&= \int_{\mathbb{R}^d} k_{\widetilde{\boldsymbol{x}}}(\mathbf{U}\boldsymbol{u},t)\varphi(\widetilde{\boldsymbol{x}})\,\mathrm{d}\widetilde{\boldsymbol{x}} &(115)\\
&= \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\mathbf{U}\boldsymbol{u},t). &(116)
\end{aligned}$$

The transition from (114) to (115) is possible because

$$k_{\mathbf{U}^\mathsf{T}\widetilde{\boldsymbol{x}}}(\boldsymbol{u},t) = \mathrm{ReLU}(\boldsymbol{u}^\mathsf{T}\mathbf{U}^\mathsf{T}\widetilde{\boldsymbol{x}}-t) + (\boldsymbol{u}^\mathsf{T}\mathbf{U}^\mathsf{T}\widetilde{\boldsymbol{x}})h_1(t) + h_2(t)$$

$$= \mathrm{ReLU}((\mathbf{U}\boldsymbol{u})^\mathsf{T}\widetilde{\boldsymbol{x}} - t) + ((\mathbf{U}\boldsymbol{u})^\mathsf{T}\widetilde{\boldsymbol{x}})h_1(t) + h_2(t)$$
$$= k_{\widetilde{\boldsymbol{x}}}(\mathbf{U}\boldsymbol{u}, t).$$

From Theorem 9, the characteristic functional of s_{ReLU} is given by

$$\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi) = \exp\left(\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Psi\left(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u}, t)\right) \mathrm{d}\boldsymbol{u}\,\mathrm{d}t\right) \tag{117}$$

with Ψ defined as in (77). Thus, based on (116) and (117), we can write

$$\begin{aligned}\widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi(\mathbf{U}\cdot)) &= \exp\left(\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Psi\left(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi(\mathbf{U}\cdot)\}(\boldsymbol{u}, t)\right) \mathrm{d}\boldsymbol{u}\,\mathrm{d}t\right)\\ &= \exp\left(\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Psi\left(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\mathbf{U}\boldsymbol{u}, t)\right) \mathrm{d}\boldsymbol{u}\,\mathrm{d}t\right)\\ &= \exp\left(\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Psi\left(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\widetilde{\boldsymbol{u}}, t)\right) \mathrm{d}\widetilde{\boldsymbol{u}}\,\mathrm{d}t\right)\\ &= \widehat{\mathbf{P}}_{s_{\mathrm{ReLU}}}(\varphi).\end{aligned} \tag{118}$$

4. In order to show that s_{ReLU} (when $\mathbf{P}_V$ has zero mean and a finite second moment) is wide-sense self-similar with Hurst exponent $H = 3/2$, we will show that for $a > 0$,

$$a^{2H}\mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x}/a)s_{\mathrm{ReLU}}(\boldsymbol{y}/a)] = \mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})s_{\mathrm{ReLU}}(\boldsymbol{y})]. \tag{119}$$

Since $\mathbf{P}_V$ has zero mean, based on (39), we have that $\mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})] = 0$. Thus, using (40), we immediately see that

$$\mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x}/a)s_{\mathrm{ReLU}}(\boldsymbol{y}/a)] = a^{-3}\mathbf{E}[s_{\mathrm{ReLU}}(\boldsymbol{x})s_{\mathrm{ReLU}}(\boldsymbol{y})]. \tag{120}$$

5. From the mean and covariance functionals in (86) and (100), respectively, and the form of the characteristic functional (35), we deduce from Definition 3 that s_{ReLU} is *non-Gaussian*, even when $\mathbf{P}_V$ has a finite second moment

■

Appendix C. Asymptotic Results

To prove Theorem 11, we rely on a generalized version of the Lévy continuity theorem from Biermé et al. (2018, Theorem 2.3), which we state below.

Theorem 13 (Generalized Lévy continuity theorem) *Let $(s_n)_{n\in\mathbb{N}}$ be a sequence of generalized stochastic processes that take values in $\mathcal{S}'(\mathbb{R}^d)$ with characteristic functionals $(\widehat{\mathbf{P}}_{s_n})_{n\in\mathbb{N}}$. If $\widehat{\mathbf{P}}_{s_n}$ converges pointwise to a functional $\widehat{\mathbf{Q}} : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ that is continuous at $\mathbf{0}$, then there exists a generalized stochastic process s such that its characteristic functional satisfies $\widehat{\mathbf{P}}_s = \widehat{\mathbf{Q}}$ and $s_n \xrightarrow[n\to\infty]{\mathscr{L}} s$.*

Proof [Proof of Theorem 11] By Theorem 13, we need to show that

1. for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$, the sequence $\left(\widehat{\mathbf{P}}_{s^n_{\mathrm{ReLU}}}(\varphi)\right)_{n\in\mathbb{N}}$ converges to

$$\widehat{\mathbf{P}}_{s^\infty_{\mathrm{ReLU}}}(\varphi) \coloneqq \exp\Big(-|b|^\alpha \| \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}\|^\alpha_{L^\alpha}\Big) \tag{121}$$

   and

2. the functional $\widehat{\mathbf{P}}_{s^\infty_{\mathrm{ReLU}}}$ is continuous on $\mathcal{S}(\mathbb{R}^d)$.

We first show that, for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$,

$$\lim_{n\to\infty} \widehat{\mathbf{P}}_{s^n_{\mathrm{ReLU}}}(\varphi) = \widehat{\mathbf{P}}_{s^\infty_{\mathrm{ReLU}}}(\varphi). \tag{122}$$

Our derivation is inspired from the proof of Lemma 2 of Fageot et al. (2020). Since s^n_{ReLU} is a *bona fide* generalized stochastic process that takes values in $\mathcal{S}'(\mathbb{R}^d)$, the functional $\widehat{\mathbf{P}}_{s^n_{\mathrm{ReLU}}}(\varphi)$ is well-defined for $\varphi \in \mathcal{S}(\mathbb{R}^d)$. On the other hand, we observe that $\widehat{\mathbf{P}}_{s^\infty_{\mathrm{ReLU}}}(\varphi)$ is also well-defined for $\varphi \in \mathcal{S}(\mathbb{R}^d)$ due to (80). Next, we prove the convergence. The characteristic functional of s^n_{ReLU} is

$$\widehat{\mathbf{P}}_{s^n_{\mathrm{ReLU}}}(\varphi) = \exp\left(\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Psi_n\Big(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u},t)\Big)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}t\right), \tag{123}$$

where

$$\Psi_n(\xi) \coloneqq n\Big(\mathrm{e}^{-\frac{|b\xi|^\alpha}{n}} - 1\Big). \tag{124}$$

For a fixed $\boldsymbol{z} \in \mathbb{S}^{d-1}\times\mathbb{R}$, we have that

$$\Psi_n(\phi(\boldsymbol{z})) = n\Big(\mathrm{e}^{-\frac{|b\phi(\boldsymbol{z})|^\alpha}{n}} - 1\Big) \xrightarrow[n\to\infty]{} -|b\phi(\boldsymbol{z})|^\alpha, \tag{125}$$

where $\phi = \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}$. Thus, we need to show that

$$\int_{\mathbb{S}^{d-1}\times\mathbb{R}} \Psi_n(\phi(\boldsymbol{z}))\,\mathrm{d}\boldsymbol{z} \xrightarrow[n\to\infty]{} \int_{\mathbb{S}^{d-1}\times\mathbb{R}} -|b\phi(\boldsymbol{z})|^\alpha\,\mathrm{d}\boldsymbol{z}. \tag{126}$$

From p. 1058 in Fageot et al. (2020), we have that

$$|\Psi_n(\phi(\boldsymbol{z}))| \le \sqrt{2}|b\phi(\boldsymbol{z})|^\alpha. \tag{127}$$

The function $\boldsymbol{z} \mapsto |b\phi(\boldsymbol{z})|^\alpha$ is in $L^1(\mathbb{S}^{d-1}\times\mathbb{R})$ due to (80). Thus, we can apply the Lebesgue dominated convergence theorem to show that (126), and consequently (122), holds. Finally, the continuity of $\widehat{\mathbf{P}}_{s^\infty_{\mathrm{ReLU}}}(\varphi)$ on $\mathcal{S}(\mathbb{R}^d)$ follows from the fact that the operator $\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}$ continuously maps $\mathcal{S}(\mathbb{R}^d)$ to $L^p(\mathbb{S}^{d-1}\times\mathbb{R})$ for $p \in [1,2]$ (cf., Equation (80)). ■

Gaussianity of $s^{\infty}_{\mathrm{ReLU}}$ When $\alpha = 2$, the characteristic functional of $s^{\infty}_{\mathrm{ReLU}}$ can be written as

$$\widehat{\mathbf{P}}_{s^{\infty}_{\mathrm{ReLU}}}(\varphi) = \exp\left(\int_{\mathbb{R}}\int_{\mathbb{S}^{d-1}} \Psi_{\infty}\Big(\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi\}(\boldsymbol{u},t)\Big)\,\mathrm{d}\boldsymbol{u}\,\mathrm{d}t\right), \tag{128}$$

where $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $\Psi_{\infty}(\xi) = -|b\xi|^2$ for $\xi \in \mathbb{R}$. Using the moment generating properties of the characteristic functional (as in Appendix B), we get that the mean functional is

$$\mu_{s^{\infty}_{\mathrm{ReLU}}}(\varphi) = 0, \quad \varphi \in \mathcal{S}(\mathbb{R}^d), \tag{129}$$

as $\Psi'_{\infty}(0) = 0$, and the covariance functional is

$$\Sigma_{s^{\infty}_{\mathrm{ReLU}}}(\varphi_1,\varphi_2) = 2|b|^2 \int_{\mathbb{S}^{d-1}\times\mathbb{R}} \mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_1\}(\boldsymbol{z})\,\mathrm{T}^{\dagger *}_{\mathrm{ReLU}}\{\varphi_2\}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}, \quad \varphi_1,\varphi_2 \in \mathcal{S}(\mathbb{R}^d), \tag{130}$$

as $\Psi''_{\infty}(0) = -2|b|^2$. Thus, from (128)–(130) and Definition 3, we see that $s^{\infty}_{\mathrm{ReLU}}$ is a Gaussian process when $\alpha = 2$.

Appendix D. Discussion of the Numerical Examples

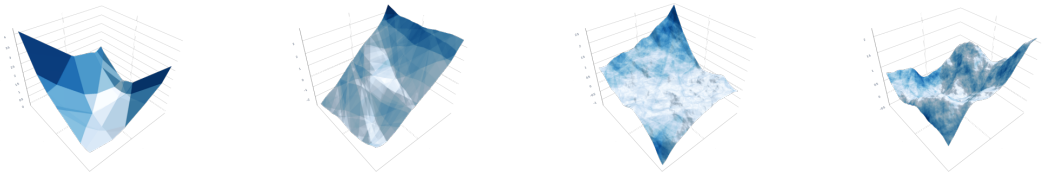
We generated realizations of the random neural networks by taking advantage of the property that Poisson points are uniformly distributed in each finite volume (cf., Equation (22)) combined with the fact that the width of a random neural network observed on a compact domain is a Poisson random variable with mean proportional to the rate parameter λ multiplied by a property related to the geometry of the domain (cf., Section 4.1). In particular, the random neural network realizations in Figures 1 and 2 were plotted on the compact domain $\Omega = [-1,+1]^d$ and were generated according to the following procedure.

1. Generate a Poisson random variable $N_{\lambda,\Omega}$ with mean $\lambda|\mathcal{Z}_{\Omega}|$, where $\mathcal{Z}_{\Omega}$ was defined in (38).
2. Generate $N_{\lambda,\Omega}$ points i.i.d. uniformly on the finite volume $\mathcal{Z}_{\Omega} \subset \mathbb{S}^{d-1}\times\mathbb{R}$, which we denote by $\{(\boldsymbol{w}_k, b_k)\}_{k=1}^{N_{\lambda,\Omega}}$.
3. Generate $N_{\lambda,\Omega}$ i.i.d. random variables according to the law $\mathbf{P}_V$, which we denote by $\{v_k\}_{k=1}^{N_{\lambda,\Omega}}$.
4. Construct the random neural network

$$s^{\mathrm{numeric}}_{\mathrm{ReLU}}(\boldsymbol{x}) = \sum_{k=1}^{N_{\lambda,\Omega}} v_k\Big[\mathrm{ReLU}(\boldsymbol{w}_k^{\mathsf{T}}\boldsymbol{x} - b_k) + \boldsymbol{c}_k^{\mathsf{T}}\boldsymbol{x} + c_{0,k}\Big] \tag{131}$$

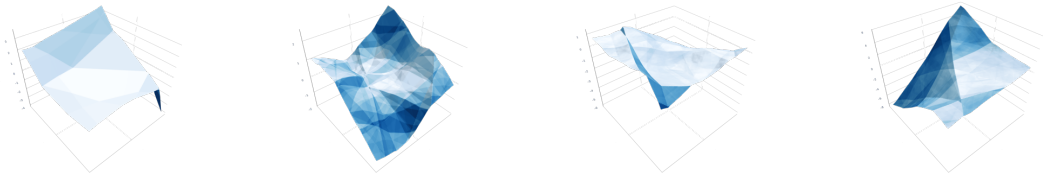
   according to the computation in (28) with $\varepsilon \to 0$.

The resulting random neural network $s^{\mathrm{numeric}}_{\mathrm{ReLU}}$ is, up to an affine function, a realization of $\mathcal{RP}(\lambda;\mathbf{P}_V)$. Finally, in order to highlight the linear regions of the generated networks, we color the top-down plots in Figures 1 and 2 according to the magnitude of the gradient of $s^{\mathrm{numeric}}_{\mathrm{ReLU}}$. As the color map choice is arbitrary, the resulting plots are thus realizations of $\mathcal{RP}(\lambda;\mathbf{P}_V)$ (since the magnitude of the gradient of an affine function is a constant, and therefore simply shifts the color map). We include some additional plots of the random neural networks in Figures 3 and 4. These figures are surface plots of the random neural networks in Figures 1 and 2, respectively.



(a) $\lambda = 1$ (b) $\lambda = 10$ (c) $\lambda = 100$ (d) $\lambda = 1000$

Figure 3: $\mathbf{P}_V$ is Gaussian.



(a) $\lambda = 1$ (b) $\lambda = 10$ (c) $\lambda = 100$ (d) $\lambda = 1000$

Figure 4: $\mathbf{P}_V$ is symmetric ($\alpha = 1.25$)-stable.

References

Francesca Bartolucci, Ernesto De Vito, Lorenzo Rosasco, and Stefano Vigogna. Understanding neural networks with reproducing kernel Banach spaces. *Applied and Computational Harmonic Analysis*, 62:194–236, 2023.

Hermine Biermé, Olivier Durieu, and Yizao Wang. Generalized random fields and Lévy's continuity theorem on the space of tempered distributions. *Communications on Stochastic Analysis*, 12(4):4, 2018.

Daryl J. Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and Its Applications. Springer New York, 2007.

Donald L. Duttweiler and Thomas Kailath. RKHS approach to detection and estimation problems–IV: Non-Gaussian detection. *IEEE Transactions on Information Theory*, 19(1): 19–28, 1973.

Ethan Dyer and Guy Gur-Ari. Asymptotics of wide networks from Feynman diagrams. In *International Conference on Learning Representations*, 2020.

Julien Fageot and Michael Unser. Scaling limits of solutions of linear stochastic differential equations driven by Lévy white noises. *Journal of Theoretical Probability*, 32(3):1166–1189, 2019.

Julien Fageot, Arash Amini, and Michael Unser. On the continuity of characteristic functionals and sparse stochastic modeling. *Journal of Fourier Analysis and Applications*, 20:1179–1211, 2014.

Julien Fageot, Virginie Uhlmann, and Michael Unser. Gaussian and sparse processes are limits of generalized Poisson processes. *Applied and Computational Harmonic Analysis*, 48(3):1045–1065, 2020.

Xavier Fernique. Processus linéaires, processus généralisés. *Annales de l'institut Fourier*, 17 (1):1–92, 1967.

Adrià Garriga-Alonso, Carl Edward Rasmussen, and Laurence Aitchison. Deep convolutional networks as shallow Gaussian processes. In *International Conference on Learning Representations*, 2019.

Izrail M. Gelfand. Generalized random processes. *Dokl. Akad. Nauk SSSR (N.S.)*, 100: 853–856, 1955.

Izrail M. Gelfand and Georgiy E. Shilov. *Generalized functions. Vol. I: Properties and operations*. Academic Press, 1964.

Izrail M. Gelfand and Naum Ya. Vilenkin. *Generalized functions, Vol. 4: Applications of harmonic analysis*. Academic Press, 1964.

Izrail M. Gelfand, Mark I. Graev, and Naum Ya. Vilenkin. *Generalized functions. Vol. 5: Integral geometry and representation theory*. Academic Press, 1966.

Boris Hanin. Random neural networks in the infinite width limit as Gaussian processes. *The Annals of Applied Probability*, 33(6A):4798–4819, 2023.

Sigurdur Helgason. *Integral Geometry and Radon Transforms*. Springer New York, 2011.

Takeyuki Hida and Nobuyuki Ikeda. Analysis on Hilbert space with reproducing kernel arising from multiple Wiener integral. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability*, pages 117–143. Univ. California Press, Berkeley, CA, 1967.

Kiyosi Itô. Stationary random distributions. *Memoirs of the College of Science. University of Kyoto. Series A. Mathematics*, 28:209–223, 1954.

Kiyosi Itô. *Foundations of stochastic differential equations in infinite dimensional spaces*, volume 47. SIAM, 1984.

Niels Jacob and René L. Schilling. *Lévy-Type Processes and Pseudodifferential Operators*, pages 139–168. Birkhäuser Boston, Boston, MA, 2001. ISBN 978-1-4612-0197-7.

Andrei N. Kolmogorov. La transformation de Laplace dans les espaces linéaires. *CR Acad. Sci. Paris*, 200:1717–1718, 1935.

Jaehoon Lee, Yasaman Bahri, Roman Novak, Samuel S. Schoenholz, Jeffrey Pennington, and Jascha Sohl-Dickstein. Deep neural networks as Gaussian processes. In *International Conference on Learning Representations*, 2018.

Donald Ludwig. The Radon transform on Euclidean space. *Communications on Pure and Applied Mathematics*, 19:49–81, 1966.

Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

Alexander G. de G. Matthews, Jiri Hron, Mark Rowland, Richard E. Turner, and Zoubin Ghahramani. Gaussian process behaviour in wide deep neural networks. In *International Conference on Learning Representations*, 2018.

Robert A. Minlos. Generalized random processes and their extension in measure. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 8:497–518, 1959.

Radford M. Neal. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics. Springer New York, 1996.

Roman Novak, Lechao Xiao, Yasaman Bahri, Jaehoon Lee, Greg Yang, Daniel A. Abolafia, Jeffrey Pennington, and Jascha Sohl-Dickstein. Bayesian deep convolutional networks with many channels are Gaussian processes. In *International Conference on Learning Representations*, 2019.

Greg Ongie, Rebecca Willett, Daniel Soudry, and Nathan Srebro. A function space view of bounded norm infinite width ReLU nets: The multivariate case. In *International Conference on Learning Representations*, 2020.

Rahul Parhi and Robert D. Nowak. Banach space representer theorems for neural networks and ridge splines. *Journal of Machine Learning Research*, 22(43):1–40, 2021.

Rahul Parhi and Robert D. Nowak. What kinds of functions do deep neural networks learn? Insights from variational spline theory. *SIAM Journal on Mathematics of Data Science*, 4 (2):464–489, 2022.

Rahul Parhi and Robert D. Nowak. Near-minimax optimal estimation with shallow ReLU neural networks. *IEEE Transactions on Information Theory*, 69(2):1125–1140, 2023a.

Rahul Parhi and Robert D. Nowak. Deep learning meets sparse regularization: A signal processing perspective. *IEEE Signal Processing Magazine*, 40(6):63–74, 2023b.

Rahul Parhi and Michael Unser. Distributional extension and invertibility of the k-plane transform and its dual. *SIAM Journal on Mathematical Analysis*, 56(4):4662–4686, 2024.

Rahul Parhi and Michael Unser. Function-space optimality of neural architectures with multivariate nonlinearities. *SIAM Journal on Mathematics of Data Science*, 7(1):110–135, 2025.

Alexander G. Ramm and Alexander I. Katsevich. *The Radon transform and local tomography*. CRC Press, Boca Raton, FL, 1996.

Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

Ken-Iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

Joseph Shenouda, Rahul Parhi, Kangwook Lee, and Robert D. Nowak. Variation spaces for multi-output neural networks: Insights on multi-task learning and network compression. *Journal of Machine Learning Research*, 25(231):1–40, 2024.

Michael Unser. Ridges, neural networks, and the Radon transform. *Journal of Machine Learning Research*, 24(37):1–33, 2023.

Michael Unser and Pouya D. Tafti. *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.

Michael Unser, Pouya D. Tafti, and Qiyu Sun. A unified formulation of Gaussian versus sparse stochastic processes—Part I: Continuous-domain theory. *IEEE Transactions on Information Theory*, 60(3):1945–1962, 2014.

Christopher Williams. Computing with infinite networks. *Advances in Neural Information Processing Systems*, 9, 1996.

Sho Yaida. Non-Gaussian processes and neural networks at finite widths. In *Mathematical and Scientific Machine Learning*, pages 165–192. PMLR, 2020.

Greg Yang. Tensor programs I: Wide feedforward or recurrent neural networks of any architecture are Gaussian processes. *Advances in Neural Information Processing Systems*, 32, 2019.

Jacob Zavatone-Veth and Cengiz Pehlevan. Exact marginal prior distributions of finite Bayesian neural networks. *Advances in Neural Information Processing Systems*, 34: 3364–3375, 2021.