

## Asymptotic analysis of higher-order scattering transform of Gaussian processes\*

Gi-Ren Liu<sup>†</sup>    Yuan-Chung Sheu<sup>‡</sup>    Hau-Tieng Wu<sup>§</sup>

### Abstract

We analyze the scattering transform with the quadratic nonlinearity (STQN) of Gaussian processes without depth limitation. STQN is a nonlinear transform that involves a sequential interlacing convolution and nonlinear operators, which is motivated to model the deep convolutional neural network. We prove that with a proper normalization, the output of STQN converges to a chi-square process with one degree of freedom in the finite dimensional distribution sense, and we provide a total variation distance control of this convergence at each time that converges to zero at an exponential rate. To show these, we derive a recursive formula to represent the intricate nonlinearity of STQN by a linear combination of Wiener chaos, and then apply the Malliavin calculus and Stein’s method to achieve the goal.

**Keywords:** scaling limits; wavelet transform; scattering transform; Wiener-Itô decomposition; Stein’s method; Malliavin calculus.

**MSC2020 subject classifications:** Primary 60G60; 60H05; 62M15, Secondary 35K15.

Submitted to EJP on August 12, 2021, final version accepted on March 17, 2022.

## 1 Introduction

The scattering transform (ST) is motivated by establishing a mathematical foundation of the convolutional neural network [20], and it has been applied to various signals, for example, fetal heart rate [8], brain waves [15, 14], respiration [29], marine bioacoustics [4], and audio [1, 13]. It provides a variety of representations for a given function  $X$  through a sequential interlacing convolution and nonlinear operators:

$$U[j_1, j_2, \dots, j_M]X(t) = A(\cdots A(A(X \star \psi_{j_1}) \star \psi_{j_2}) \cdots \star \psi_{j_M}(t)), \quad t \in \mathbb{R},$$

---

\*This work benefited from support of the National Center for Theoretical Science (NCTS, Taiwan) and the Ministry of Science and Technology (MOST, Taiwan).

<sup>†</sup>Department of Mathematics, National Cheng-Kung University, Tainan, Taiwan. E-mail: girenliu@mail.ncku.edu.tw

<sup>‡</sup>Department of Applied Mathematics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. E-mail: sheu@math.nctu.edu.tw

<sup>§</sup>Department of Mathematics and Department of Statistical Science, Duke University, Durham, NC, USA. E-mail: hauwu@math.duke.edu

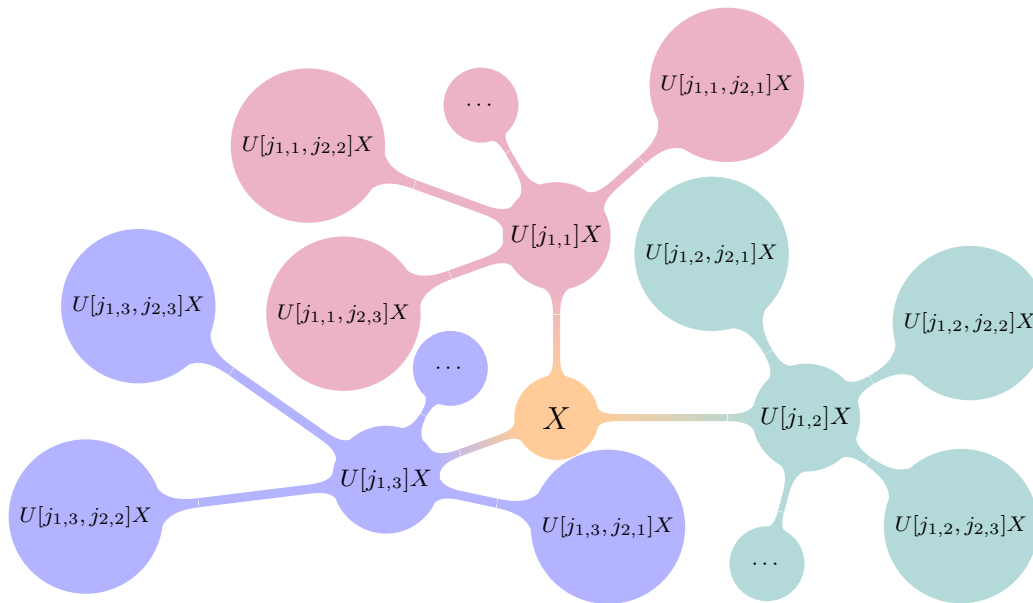


Figure 1: Tree structure of the scattering transform. Here,  $j_{k,1}, j_{k,2}, j_{k,3}, \dots$  are the scale parameters for the  $k$ -th layer of the tree, where  $k \in \mathbb{N}$ .

where  $M \in \mathbb{N}$  is the depth of ST,  $\{j_1, j_2, \dots, j_M\} \subset \{\dots, J - 2, J - 1, J\}$  is a set of scale parameters,  $J \in \mathbb{Z}$  determines the range of interest in the frequency domain,  $\{\psi_{j_1}, \dots, \psi_{j_M}\}$  is a family of wavelets generated from a selected mother wavelet  $\psi$ , and  $A : \mathbb{C} \rightarrow \mathbb{C}$  is the chosen activation function. This construction is called the *scattering network*. See Figure 1 for its structure. The associated  $M$ -th order ST coefficients are then computed through the pooling process

$$S_J[j_1, j_2, \dots, j_M]X(t) := U[j_1, j_2, \dots, j_M]X \star \phi_J(t), \quad t \in \mathbb{R},$$

where  $\phi_J$  is a low-pass filter, and it is usually chosen to be the father wavelet associated with  $\psi$  and  $J$  so that the Littlewood-Paley condition is satisfied. We mention that ST can also be defined with the Gabor transform [9, 12] and others [27, 7], but we focus on the wavelet transform considered in [20] in this paper. In most applications, only the first- and second-order ST coefficients are used because extracting higher-order information brings additional computational costs, except [1, 2], which shows the potential benefit of considering higher order ST coefficients. There are various choices for the activation  $A$ ; for example,  $A(\cdot) = |\cdot|$  [20, 2, 17],  $A(\cdot) = |\cdot|^2$  [4, 3], and more general functions like Lipschitz-continuous functions [28, 16]. We mention that  $A(\cdot) = |\cdot|^2$  is considered to speed up the computation of deeper scattering networks [4], since the computation could be carried out without leaving the Fourier domain. When  $A(\cdot) = |\cdot|^2$ , we call the resulting transform the *ST with the quadratic nonlinearity* (STQN).

There have been several theoretical supports established for ST. If the mother wavelet and the low-pass filter  $\phi_J$  satisfy the Littlewood-Paley condition and  $A(\cdot) = |\cdot|$ , the ST coefficients are approximately invariant to time shifts and stable to small time-warping deformation [20]. These properties have been generalized to Lipschitz-continuous activation functions [16]. Similar time invariance and stability to deformation results are discussed in [28] under a different setting. The authors of [3] proved that the STQN coefficients satisfy the Lipschitz continuous property; that is,  $\|S_J[j_1, j_2, \dots, j_M]X - S_J[j_1, j_2, \dots, j_M]Y\|_{L^2}$  can be controlled by  $\|X - Y\|_{L^2}$  for any  $X, Y \in L^2$ , if some conditions on the wavelets, the low-pass filter and the  $L^\infty$  norm of inputs hold. The authors of [18, Figure 4] showed

that the decay of the STQN coefficients depend on the number of sinusoidal components contained in the signal and STQN has the ability to disentangle multiple factors of variability in the spectral envelope. In addition to the above results in the deterministic setup, ST has also been studied from the probabilistic point of view. The authors of [6] found that the second-order ST with  $A(\cdot) = |\cdot|$  can be applied to characterize the self-similarity and intermittency properties of random processes with stationary increments by computing the second moment of the second-order ST of  $X$  and observing its behavior with respect to the scale parameters  $j_1$  and  $j_2$ . This result has been extended to general Lipschitz-continuous activation functions [16].

Despite the above results, however, due to the nonlinearity of the activation function and its subsequent convolution, the behavior of higher order ST with a random process input is still open. Precisely, when  $M > 2$ , since the input  $U[j_1, j_2, \dots, j_{M-1}]X$  to the  $M$ -th layer of the scattering network is neither a Gaussian process nor a subordinated Gaussian process even when  $X$  is a stationary Gaussian process or a Gaussian process with stationary increments, it is challenging to study ST with random processes when  $M > 2$ . Motivated by its practical usefulness and theoretical challenges, in this work we focus on exploring STQN with depth  $M \geq 2$  from the probabilistic point of view, in which the inputs are the stationary Gaussian processes, or Gaussian processes with stationary increments.

To this end, several analysis techniques are needed. First, we derive a recursive formula to handle the intricate nonlinearity of STQN by representing  $U[j_1, j_2, \dots, j_M]X$  by a linear combination of Wiener chaos that comes from Wiener-Itô integrals of functions defined iteratively from the spectral density function of the Gaussian random process. Then, we apply the Stein's method [23] and the product formula for Wiener-Itô integrals [19] to prove that the total variation (TV) distance between  $\sigma_{j_M}^{-2}U[j_1, j_2, \dots, j_M]X(2^{j_M}t)$  and the square of the standard normal random variable is  $O(2^{-j_M/2})$  for each fixed  $t \in \mathbb{R}$ , where  $\sigma_{j_M}^2 = \mathbb{E}[U[j_1, j_2, \dots, j_M]X(\cdot)]$ . Moreover, we prove that  $2^{j_M}U[j_1, j_2, \dots, j_M]X(2^{j_M}\cdot)$  converges to a weakly dependent chi-square process with one degree of freedom in the finite dimensional distribution sense as  $j_M \rightarrow \infty$ . We shall mention that the proof in this work is inspired by [21], where the Malliavin calculus and Stein's method are combined to derive explicit bounds in the Gaussian and Gamma approximations of random variables in a fixed Wiener chaos of a general Gaussian process. The works [21, Proposition 3.7] and [22, Theorem 5.1.3 and Remark 5.1.4] provide error bounds for the Gaussian approximation of a finite sum of multiple Wiener-Itô integrals. For applying them to our work and ensuring that the approximation error converges to zero when  $j_M \rightarrow \infty$ , we need to analyze the intricate covariance structure between different orders of Wiener chaos in a slightly different way.

Last but not the least, we should mention that the theoretical properties mentioned above are interesting from the machine learning perspective because the time invariance and stability to small local deformation are desirable when data analysts handle real-world data. Aside from having these invariant properties, the ST coefficients keeps multi-scale information [2, 5] within the signal  $X$  by applying  $\{U[j_{m,1}], U[j_{m,2}], \dots\}$  on the outputs of the previous neurons  $\{U[j_{m-1,1}], U[j_{m-1,2}], \dots\}$ , where  $m = 2, 3, \dots, M$ , before taking the convolution with the low-pass filter  $\phi_J$  as shown in Figure 1. In the appendix, these properties are described in more concrete terms for the possible interest of readers. On the other hand, note that there are many time series that can be modeled by random processes; for example, fetal heart rate [8], brain waves [15, 14], respiration [29], marine bioacoustics [4], and audio [1, 13]. While ST has been successfully applied to these signals that can be modeled by random processes, to our knowledge, there are limited theoretical results available to guide data analysts. The theoretical justification of how ST acts on random processes we provide in this paper would fill in this gap. With this

relationship in mind, the results of ST could connect data analysts and probabilists. In sum, a theoretical understanding of ST, particularly its behavior in stochastic processes, would help guide us to better understand how to apply ST to analyze datasets that can be modeled by random processes.

The rest of the paper is organized as follows. In Section 2, we summarize necessary material for ST and present some preliminaries about the Wiener-Itô integrals and the Malliavin calculus. In Section 3, we state our main results, including Theorems 1, 2, and 3. The proofs of our main results and some technical lemmas are given in Section 4. Table 1 below contains a list of frequently used symbols and abbreviations.

Table 1: List of frequently used symbols and abbreviations

Symbol	Description
$\star$	Convolution operator
$M$	Depth of the scattering network
$\pm\lambda_{1:p}$	Abbreviation for $\pm(\lambda_1, \lambda_2, \dots, \lambda_p) \in \mathbb{R}^p$
$\lambda_{1:p}^+, d\lambda_{1:p}$	Abbreviation for $\lambda_1 + \lambda_2 + \dots + \lambda_p$ and $d\lambda_1 d\lambda_2 \dots d\lambda_p$
$j_{1:m}$	Vector of scale parameters for the wavelet transform in the first $m$ layers
$\psi, \hat{\psi}, \psi_j(\cdot)$	Mother wavelet, its Fourier transform, and its scaled version $2^{-j}\psi(2^{-j}\cdot)$
$X$	Stationary Gaussian process
$U[j]X$	Modulus-squared wavelet transform of $X$ defined by $U[j]X =  \psi_j \star X ^2$
$U[j_{1:m}]$	The $m$ -th order scattering transform defined by $U[j_m]U[j_{m-1}] \dots U[j_1]$
$T[j_{1:m+1}]X$	$(U[j_{1:m}]X) \star \psi_{j_{m+1}}$
$\alpha$	Vanishing-moment parameter for $\psi$ defined by $\hat{\psi}(\cdot) = C_{\hat{\psi}}(\cdot) \cdot  \cdot ^\alpha$
$f_Z$	Spectral density of the covariance function of a process $Z$
$\beta$	Singularity parameter for $f_X$ defined through $f_X(\cdot) = C_X(\cdot) \cdot  \cdot ^{\beta-1}$
$\overline{H}$	Hilbert space $\{f \in L^2(\mathbb{R}) \mid f(-\lambda) = f(\lambda) \text{ for all } \lambda \in \mathbb{R}\}$
$\overline{H}^{\otimes p} (\overline{H}^{\odot p})$	The $p$ -th (symmetric) tensor product of the Hilbert space $\overline{H}$
$W$	Complex-valued Gaussian random measure
$I_p$	$p$ -fold Wiener-Itô integrals for $p \in \mathbb{N}$ ( $I_0$ : the identity map)
$f$	Canonical symmetrization of function $f$
$f \otimes_r g, f \tilde{\otimes}_r g$	The $r$ -th contraction of $f$ and $g$ , and its canonical symmetrization
$L^2(\Omega, \overline{H}^{\otimes p})$	Random functions $f$ in $\overline{H}^{\otimes p}$ with $\mathbb{E}\ f\ _{L^2(\mathbb{R}^p)}^2 < \infty$
$M_r$	$2^{M-1} - 2r$ for $r = 0, 1, \dots, 2^{M-2} - 1$
$\{^M F_t^{(M_r)}\}_{r=0}^{2^{M-2}-1}$	Integrands in the Wiener-Itô decomposition of $T[j_{1:M}]X$
$\{^M \tilde{G}_t^{(2^M-2\ell)}\}_{\ell=0}^{2^{M-1}}$	Integrands in the Wiener-Itô decomposition of $U[j_{1:M}]X$
$a_n = O(b_n)$	For sequences $\{a_n\}$ and $\{b_n\}$ , there exists a constant $C > 0$ such that $ a_n  \leq C b_n $ when $n$ is sufficiently large.

## 2 Preliminaries

### 2.1 Scattering transform

Let  $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  be the mother wavelet, which satisfies  $\int_{\mathbb{R}} \psi(t) dt = 0$ . A family of real-valued functions  $\{\psi_j(t) \mid j \in \mathbb{Z}, t \in \mathbb{R}\}$  is called a wavelet family, where

$$\psi_j(t) = \frac{1}{2^j} \psi\left(\frac{t}{2^j}\right).$$

Denote the Fourier transform of  $\psi$  by  $\hat{\psi}$ , which is defined as

$$\hat{\psi}(\lambda) = \int_{\mathbb{R}} e^{-i\lambda t} \psi(t) dt.$$

Because  $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  and  $\int_{\mathbb{R}} \psi(t) dt = 0$ ,  $\hat{\psi} \in C^0(\mathbb{R}) \cap L^2(\mathbb{R})$  with  $\hat{\psi}(0) = 0$ . We make the following assumption.

**Assumption 1.** We assume that  $\hat{\psi} \in L^1(\mathbb{R})$  and there exists a bounded and continuous complex-valued function  $C_{\hat{\psi}}$ , which is positive at the origin (i.e.,  $C_{\hat{\psi}}(0) > 0$ ), such that  $\hat{\psi}$  can be expressed as

$$\hat{\psi}(\lambda) = C_{\hat{\psi}}(\lambda) |\lambda|^\alpha$$

for some  $\alpha > 0$ .

**Remark 1.** Consider the vanishing moment  $N$  of  $\psi$  defined as

$$N = \max \left\{ n \in \mathbb{N} \mid \int_{\mathbb{R}} t^\ell \psi(t) dt = 0 \text{ for } \ell = 0, 1, \dots, n-1, \text{ and } \int_{\mathbb{R}} t^n \psi(t) dt \neq 0 \right\}.$$

The parameter  $\alpha$  in Assumption 1 satisfies  $\alpha \geq N$  by the Taylor expansion. Assumption 1 with  $\alpha \geq 1$  holds for commonly used wavelets, including the real part of complex Morlet wavelet ( $\alpha = 1$ ), the Mexican hat wavelet ( $\alpha = 2$ ), and the Daubechies- $K$  wavelet ( $\alpha = K/2$ ), where  $K = 2, 4, \dots, 20$ . We shall mention that for those wavelet functions whose Fourier transform  $\hat{\psi}$  vanishes in a neighborhood of the origin, e.g., the Meyer wavelet, all results in this paper hold.

Given a locally bounded function  $X : \mathbb{R} \rightarrow \mathbb{R}$ , the wavelet transform of  $X$  is defined as

$$X \star \psi_j(t) = \int_{\mathbb{R}} X(s) \psi_j(t-s) ds, \quad j \in \mathbb{Z}, \quad t \in \mathbb{R}. \tag{2.1}$$

The first-order STQN of  $X$  is obtained by composing the wavelet transform of  $X$  with the nonlinear operator  $|\cdot|^2$ ; that is,

$$U[j_1]X(t) = |X \star \psi_{j_1}(t)|^2, \quad j_1 \in \mathbb{Z}.$$

Fix  $M \geq 1$  to be the order (or depth) of ST. The  $M$ -th order STQN of  $X$  is defined iteratively as

$$U[j_1, j_2, \dots, j_M]X(t) := U[j_M]U[j_{M-1}] \cdots U[j_2]U[j_1]X(t) =: U[j_{1:M}]X(t),$$

where  $j_1, j_2, \dots, j_M \in \mathbb{Z}$  and  $j_{1:M} := [j_1, \dots, j_M] \in \mathbb{Z}^M$  is introduced to simplify the heavy notation for the upcoming proof. Denote the wavelet transform of the output of the  $M$ -th order STQN of  $X$  by

$$T[j_{1:M+1}]X = U[j_{1:M}]X \star \psi_{j_{M+1}}. \tag{2.2}$$

We define  $T[j_1]X = X \star \psi_{j_1}$ . By definition, the operators  $U$  and  $T$  have the relationship

$$U[j_{1:M}]X = |T[j_{1:M}]X|^2. \tag{2.3}$$

### 2.2 Stationary Gaussian processes

In this section, we prepare materials about stationary Gaussian random processes. Let  $\overline{H} = \{f \in L^2(\mathbb{R}) \mid f(-\lambda) = \overline{f(\lambda)} \text{ for all } \lambda \in \mathbb{R}\}$  be a complex Hilbert space with the inner product  $\langle f, g \rangle = \int_{\mathbb{R}} f(\lambda)\overline{g(\lambda)}d\lambda$ , and let  $W$  be a complex-valued Gaussian random measure on  $\mathbb{R}$  satisfying

$$W(\Delta_1) = \overline{W(-\Delta_1)}, \quad \mathbb{E}[W(\Delta_1)] = 0, \quad \text{and} \quad \mathbb{E}\left[W(\Delta_1)\overline{W(\Delta_2)}\right] = \text{Leb}(\Delta_1 \cap \Delta_2) \quad (2.4)$$

for any  $\Delta_1, \Delta_2 \in \mathcal{B}(\mathbb{R})$ , where  $\text{Leb}$  is the Lebesgue measure on  $\mathbb{R}$  and  $\mathcal{B}(\mathbb{R})$  is the Borel  $\sigma$ -algebra on  $\mathbb{R}$ . For every  $f \in \overline{H}$ , we define  $W(f) = \int_{\mathbb{R}} f(\lambda)W(d\lambda)$ . Due to (2.4),  $W(f) \in \mathbb{R}$  and  $\mathbb{E}[W(f)W(g)] = \langle f, g \rangle$ . Hence,  $W = \{W(f) \mid f \in \overline{H}\}$  can be viewed as a centered isonormal Gaussian process over  $\overline{H}$  [22, Section 2.1]. Let  $(\Omega, \mathcal{F}_W, \mathcal{P})$  be a probability space, where the  $\sigma$ -algebra  $\mathcal{F}_W$  is generated by  $W$ .

Let  $X$  be a mean-square continuous and stationary real Gaussian random process with the constant mean  $\mathbb{E}[X(s)] = \mu \in \mathbb{R}$  and the covariance function  $R_X$ :

$$R_X(t_1 - t_2) = \mathbb{E}[X(t_1)X(t_2)] - \mathbb{E}[X(t_1)]\mathbb{E}[X(t_2)], \quad t_1, t_2 \in \mathbb{R}.$$

Note that  $R_X$  is continuous by the assumption of  $X$ . Since  $\int_{\mathbb{R}} \psi(s)ds = 0$ ,  $X \star \psi_j(t) = [X - \mu] \star \psi_j(t)$ , and hence we can assume that  $\mathbb{E}[X(s)] = 0$  without loss of generality. By the Bochner-Khinchin theorem, there exists a unique nonnegative measure  $F_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty)$  such that  $F_X(\Delta) = F_X(-\Delta)$  for any  $\Delta \in \mathcal{B}(\mathbb{R})$  and

$$R_X(t) = \int_{\mathbb{R}} e^{i\lambda t} F_X(d\lambda), \quad t \in \mathbb{R}. \quad (2.5)$$

The measure  $F_X$  is called the spectral measure of the covariance function  $R_X$ .

**Assumption 2.** The spectral measure  $F_X$  has the density  $f_X$  and

$$f_X(\lambda) = \frac{C_X(\lambda)}{|\lambda|^{1-\beta}},$$

where  $\beta \in (0, 1)$  is the *Hurst index* of long-range dependence and  $C_X$  is a bounded and continuous function from  $\mathbb{R}$  to  $[0, \infty)$  such that  $C_X(\lambda)$  decays faster than  $|\lambda|^{-\beta-\varepsilon}$  for some  $\varepsilon > 0$  when  $|\lambda| \rightarrow \infty$ .

If the function  $C_X$  in Assumption 2 satisfies  $C_X(0) > 0$ , then  $X$  is a long-range dependent process [10, 26] because  $f_X$  has a singularity at 0. Note that the spectral density function  $f_X$  is even and nonnegative with  $\int_{\mathbb{R}} f_X(\lambda)d\lambda = R_X(0)$ .

Under Assumption 2, (2.5) can be rewritten as

$$R_X(t) = \int_{\mathbb{R}} e^{i\lambda t} f_X(\lambda)d\lambda, \quad t \in \mathbb{R}.$$

By the Karhunen theorem, the Gaussian process  $X$  has the representation

$$X(t) = \int_{\mathbb{R}} e^{i\lambda t} \sqrt{f_X(\lambda)}W(d\lambda), \quad t \in \mathbb{R}. \quad (2.6)$$

By substituting (2.6) into (2.1), the continuous wavelet transform of  $X$  can be represented as a Wiener integral

$$\begin{aligned} X \star \psi_j(t) &= \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} e^{i\lambda s} \sqrt{f_X(\lambda)}W(d\lambda) \right] \psi_j(t-s)ds \\ &= \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} e^{i\lambda s} \psi(t-s)ds \right] \sqrt{f_X(\lambda)}W(d\lambda) = \int_{\mathbb{R}} e^{i\lambda t} \hat{\psi}_j(\lambda) \sqrt{f_X(\lambda)}W(d\lambda), \end{aligned} \quad (2.7)$$

where the second equality follows from the stochastic Fubini theorem [25, Theorem 2.1]. From (2.7), we know that  $X \star \psi_j$  has the spectral density

$$f_{X \star \psi_j}(\lambda) = f_X(\lambda) |\hat{\psi}_j(\lambda)|^2, \lambda \in \mathbb{R},$$

and

$$U[j_1]X(t) = \left( \int_{\mathbb{R}} e^{it\lambda} \hat{\psi}_{j_1}(\lambda) \sqrt{f_X(\lambda)} W(d\lambda) \right)^2. \tag{2.8}$$

**Remark 2.** It is known since the works of Itô [11] and Yaglom [30] that if a Gaussian random process  $X$  is increment-stationary in the sense that for any  $s, s', t, t'$  and  $h \in \mathbb{R}$ ,

$$\mathbb{E}[(X(t+h) - X(s+h))(X(t'+h) - X(s'+h))] = \mathbb{E}[(X(t) - X(s))(X(t') - X(s'))]$$

and if the covariance between the increments over time intervals  $I$  and  $I'$  tends to zero when  $\min\{|a-b| \mid a \in I, b \in I'\} \rightarrow \infty$ , then  $X$  admits a spectral representation

$$X(t) = c + \int_{\mathbb{R}} (e^{it\lambda} - 1) Z(d\lambda), t \in \mathbb{R}, \tag{2.9}$$

where  $c \in \mathbb{R}$  and  $Z(d\lambda)$  is a complex-valued Gaussian random measure on  $\mathbb{R}$  with spectral measure  $F_X$ . If the spectral measure  $F_X$  has a density  $f_X$ , then (2.9) can be rewritten as

$$X(t) = c + \int_{\mathbb{R}} (e^{it\lambda} - 1) \sqrt{f_X(\lambda)} W(d\lambda), t \in \mathbb{R}.$$

By the mean-zero property of the wavelet,

$$\begin{aligned} X \star \psi_j(t) &= \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} (e^{it'\lambda} - 1) \sqrt{f_X(\lambda)} W(d\lambda) \right] \psi_j(t-t') dt' \\ &= \int_{\mathbb{R}} e^{it\lambda} \hat{\psi}_j(\lambda) \sqrt{f_X(\lambda)} W(d\lambda). \end{aligned} \tag{2.10}$$

From the consistency between (2.7) and (2.10), we know that the all results below hold for Gaussian random processes with stationary increments, including the fractional Brownian motions, when the associated conditions are fulfilled.

### 2.3 Wiener-Itô integrals and Wiener chaos expansion

In this section, we describe how to express the STQN of stationary Gaussian random processes in terms of the Wiener chaos expansion that comes from Wiener-Itô integrals of functions defined iteratively from the spectral density function of the Gaussian random process. Given an integer  $p \geq 2$ , the  $p$ -th tensor product of the Hilbert space  $\overline{H}$  is denoted by  $\overline{H}^{\otimes p}$ . According to [19, p. 27],  $f \in \overline{H}^{\otimes p}$  if and only if  $f = f(\lambda_1, \dots, \lambda_p)$ , where  $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ , is a complex valued function satisfying  $f(-\lambda_1, \dots, -\lambda_p) = f(\lambda_1, \dots, \lambda_p)$  and

$$\|f\|_{\overline{H}^{\otimes p}}^2 = \int_{\mathbb{R}^p} |f(\lambda_{1:p})|^2 d\lambda_{1:p} < \infty,$$

where  $\lambda_{p_1:p_2}$  means  $(\lambda_{p_1}, \lambda_{p_1+1}, \dots, \lambda_{p_2}) \in \mathbb{R}^{p_2-p_1+1}$  and  $d\lambda_{p_1:p_2}$  means  $d\lambda_{p_1} d\lambda_{p_1+1} \dots d\lambda_{p_2}$  when  $p_2 > p_1$ . For any  $f, g \in \overline{H}^{\otimes p}$ , their inner product is defined as

$$\langle f, g \rangle_{\overline{H}^{\otimes p}} = \int_{\mathbb{R}^p} f(\lambda_{1:p}) \overline{g(\lambda_{1:p})} d\lambda_{1:p} = \int_{\mathbb{R}^p} f(\lambda_{1:p}) g(-\lambda_{1:p}) d\lambda_{1:p}.$$

The subscripts of  $\|f\|_{\overline{H}^{\otimes p}}^2$  and  $\langle f, g \rangle_{\overline{H}^{\otimes p}}$  will be ignored as they can be inferred from the number of variables of  $f$  and  $g$ . The  $p$ -th symmetric tensor product of  $\overline{H}$  is denoted by  $\overline{H}^{\odot p}$ , which contains those functions  $f \in \overline{H}^{\otimes p}$  satisfying  $f(\lambda_{\pi(1)}, \dots, \lambda_{\pi(p)}) =$

$f(\lambda_1, \dots, \lambda_p)$ , where  $\pi$  is any permutation of the set  $\{1, 2, \dots, p\}$ . Because  $\overline{H}^{\odot p}$  is a Hilbert subspace of  $\overline{H}^{\otimes p}$ , the norm and inner product in  $\overline{H}^{\odot p}$  are defined in the same way as in  $\overline{H}^{\otimes p}$ . For any  $f \in \overline{H}^{\otimes p}$ , where  $p \in \mathbb{N}$ , the  $p$ -fold Wiener-Itô integrals of  $f$  with respect to the random measure  $W$  is defined by

$$I_p(f) = \int_{\mathbb{R}^p}' f(\lambda_{1:p}) W(d\lambda_1) \cdots W(d\lambda_p),$$

where  $\int'$  means that the integral excludes the diagonal hyperplanes  $\lambda_k = \mp \lambda_{k'}$  for  $k, k' \in \{1, \dots, p\}$  and  $k \neq k'$ . Note that  $I_p(f) = I_p(\tilde{f})$ , where  $\tilde{f} \in \overline{H}^{\odot p}$  is the canonical symmetrization of  $f$  defined as

$$\tilde{f}(\lambda_{1:p}) = \frac{1}{p!} \sum_{\pi} f(\lambda_{\pi(1)}, \dots, \lambda_{\pi(p)}), \tag{2.11}$$

where the sum runs over all permutations  $\pi$  of  $\{1, \dots, p\}$ . By default,  $I_0$  is the identity map, i.e.,  $I_0(f) = f$  for any function  $f$ .

For  $p \in \mathbb{N} \cup \{0\}$ , let  $h_p$  be the Hermite polynomial of degree  $p$ , which is defined by

$$h_p(y) = (-1)^p e^{\frac{y^2}{2}} \frac{d^p}{dy^p} e^{-\frac{y^2}{2}}, \quad y \in \mathbb{R}.$$

We write  $\mathcal{H}_p$  to denote the closed linear subspace of  $L^2(\Omega, \mathcal{F}_W, \mathcal{P})$  generated by the random variables of type  $h_p(W(g))$ ,  $g \in \overline{H}$ ,  $\|g\| = 1$ . The space  $\mathcal{H}_p$  is called the  $p$ -th Wiener chaos of  $W$ . According to [19, 24, 21, 22],  $I_p(\tilde{f}) \in \mathcal{H}_p$  for any  $f \in \overline{H}^{\otimes p}$ . Furthermore, every  $F \in L^2(\Omega, \mathcal{F}_W, \mathcal{P})$  admits a unique expansion of the type

$$F = \mathbb{E}[F] + \sum_{p=1}^{\infty} \text{Proj}(F | \mathcal{H}_p),$$

where  $\text{Proj}(F | \mathcal{H}_p) \in \mathcal{H}_p$  and the above series converges in  $L^2(\Omega, \mathcal{F}_W, \mathcal{P})$ . The following product formula allows us to express the output of STQN in terms of a linear combination of Wiener chaos, which is the first tool used in this paper.

**Lemma 1** (Product Formula [19, 22]). Let  $p, q \geq 1$ . If  $f \in \overline{H}^{\odot p}$  and  $g \in \overline{H}^{\odot q}$ , then

$$I_p(f)I_q(g) = \sum_{r=0}^{p \wedge q} r! \binom{p}{r} \binom{q}{r} I_{p+q-2r}(f \otimes_r g),$$

where  $f \otimes_r g$  is the  $r$ th contraction of  $f$  and  $g$  defined as

$$f \otimes_r g(\lambda_{1:p+q-2r}) = \int_{\mathbb{R}^r} f(\tau_{1:r}, \lambda_{1:p-r}) g(-\tau_{1:r}, \lambda_{p-r+1:p+q-2r}) d\tau_{1:r}$$

for  $r = 1, 2, \dots, p \wedge q$ . When  $r = 0$ , set  $f \otimes_0 g = f \otimes g$ .

Note that  $f \otimes_r g \in \overline{H}^{\otimes p+q-2r}$ , but it may not be symmetric. Let  $\tilde{f} \otimes_r g$  be the canonical symmetrization of  $f \otimes_r g$ . For any  $f \in \overline{H}^{\odot p}$ , by noting that the expectation of any Wiener-Itô integral equals to zero, Lemma 1 implies

$$\mathbb{E} [|I_p(f)|^2] = \mathbb{E} \left[ \sum_{r=0}^p r! \binom{p}{r}^2 I_{2p-2r}(f \otimes_r f) \right] = p! f \otimes_p f = p! \|f\|_{L^2(\mathbb{R}^p)}^2,$$

which is the so-called isometry property of the multiple Wiener-Itô integrals.



We now express the STQN of  $X$  using the product formula. By applying Lemma 1 to (2.8),

$$U[j_1]X(t) = \int_{\mathbb{R}^2} e^{i(\lambda_1+\lambda_2)t} \hat{\psi}_{j_1}(\lambda_1) \sqrt{f_X(\lambda_1)} \hat{\psi}_{j_1}(\lambda_2) \sqrt{f_X(\lambda_2)} W(d\lambda_1) W(d\lambda_2) + \|f_{X \star \psi_{j_1}}\|_1. \tag{2.12}$$

Denote

$$\begin{cases} {}^1\tilde{G}_t^{(0)} = \|f_{X \star \psi_{j_1}}\|_1, \\ {}^1\tilde{G}_t^{(2)}(\lambda_1, \lambda_2) = e^{i(\lambda_1+\lambda_2)t} \hat{\psi}_{j_1}(\lambda_1) \sqrt{f_X(\lambda_1)} \hat{\psi}_{j_1}(\lambda_2) \sqrt{f_X(\lambda_2)}. \end{cases} \tag{2.13}$$

The first-order STQN of  $X$  can be rewritten as

$$U[j_1]X(t) = I_2 \left( {}^1\tilde{G}_t^{(2)} \right) + {}^1\tilde{G}_t^{(0)}.$$

By the property  $\int_{\mathbb{R}} \psi_{j_2}(t-t') dt' = 0$  and (2.12),

$$T[j_{1:2}]X(t) = \int_{\mathbb{R}} I_2 \left( {}^1\tilde{G}_{t'}^{(2)} \right) \psi_{j_2}(t-t') dt'. \tag{2.14}$$

Because (2.13) shows that

$${}^1\tilde{G}_t^{(2)}(\lambda_1, \lambda_2) = {}^1\tilde{G}_0^{(2)}(\lambda_1, \lambda_2) e^{i(\lambda_1+\lambda_2)t}, \tag{2.15}$$

$${}^2F_t^{(2)}(\lambda_1, \lambda_2) := \int_{\mathbb{R}} {}^1\tilde{G}_{t'}^{(2)}(\lambda_1, \lambda_2) \psi_{j_2}(t-t') dt' = {}^1\tilde{G}_t^{(2)}(\lambda_1, \lambda_2) \hat{\psi}_{j_2}(\lambda_1 + \lambda_2), \tag{2.16}$$

which depends on  $t$  in the same way. More concretely,

$${}^2F_t^{(2)}(\lambda_1, \lambda_2) = e^{i(\lambda_1+\lambda_2)t} \hat{\psi}_{j_1}(\lambda_1) \sqrt{f_X(\lambda_1)} \hat{\psi}_{j_1}(\lambda_2) \sqrt{f_X(\lambda_2)} \hat{\psi}_{j_2}(\lambda_1 + \lambda_2).$$

By substituting (2.16) into (2.14), we get

$$T[j_{1:2}]X(t) = I_2 \left( {}^2F_t^{(2)} \right). \tag{2.17}$$

By the product formula again with  $p = q = 2$  and  $f = g = {}^2F_t^{(2)}$ ,

$$U[j_{1:2}]X(t) = [T[j_{1:2}]X(t)]^2 = \left[ I_2 \left( {}^2F_t^{(2)} \right) \right]^2 = \sum_{r=0}^2 I_{4-2r} \left( {}^2\tilde{G}_t^{(4-2r)} \right), \tag{2.18}$$

where

$${}^2\tilde{G}_t^{(4-2r)} = r! \binom{2}{r} \binom{2}{r} {}^2F_t^{(2)} \tilde{\otimes}_r {}^2F_t^{(2)} \in \overline{H}^{\odot 4-2r} \tag{2.19}$$

for each  $t \in \mathbb{R}$ . From (2.15) and (2.16), we have

$${}^2\tilde{G}_t^{(4-2r)}(\lambda_{1:4-2r}) = e^{i\lambda_{1:4-2r}^+ t} {}^2\tilde{G}_0^{(4-2r)}(\lambda_{1:4-2r}) \tag{2.20}$$

for  $r = 0, 1$ , where  $\lambda_{p_1:p_2}^+$  means  $\lambda_{p_1} + \lambda_{p_1+1} + \dots + \lambda_{p_2} \in \mathbb{R}$  for  $p_2 > p_1 \geq 1$ . For general cases, we have the following result.

**Lemma 2.** For every integer  $M \geq 2$  and stationary Gaussian process  $X$ , the random processes  $T[j_{1:M}]X$  and  $U[j_{1:M}]X$  arising from STQN of  $X$  have the Wiener-Itô decomposition as follows

$$T[j_{1:M}]X(t) = \sum_{r=0}^{2^{M-2}-1} I_{2^{M-1}-2r} \left( {}^M F_t^{(2^{M-1}-2r)} \right) \tag{2.21}$$

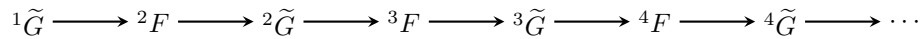


Figure 2: Flow chart for the recursive computation in (2.23) and (2.24).

and

$$U[j_{1:M}]X(t) = \sum_{\ell=0}^{2^{M-1}} I_{2^{M-2\ell}} \left( {}^M\tilde{G}_t^{(2^M-2\ell)} \right), \tag{2.22}$$

where the integrands,  ${}^M F_t^{(2^{M-1}-2r)}$  and  ${}^M\tilde{G}_t^{(2^M-2\ell)}$ , are defined by using the following recursive formula starting with (2.13). For  $\ell = 0, 1, \dots, 2^{M-1}$ ,

$$\begin{aligned} {}^M\tilde{G}_t^{(2^M-2\ell)}(\lambda_{1:2^M-2\ell}) &:= \sum_{r,r'=0}^{2^{M-2}-1} (\ell - r - r')! \binom{2^{M-1} - 2r}{\ell - r - r'} \binom{2^{M-1} - 2r'}{\ell - r - r'} \\ &\times {}^M F_t^{(2^{M-1}-2r)} \otimes_{\ell-r-r'} {}^M F_t^{(2^{M-1}-2r')}(\lambda_{1:2^M-2\ell}), \end{aligned} \tag{2.23}$$

where for  $r = 0, 1, \dots, 2^{M-2} - 1$ ,

$${}^M F_t^{(2^{M-1}-2r)}(\lambda_{1:2^M-2r}) := {}^{M-1}\tilde{G}_t^{(2^{M-1}-2r)}(\lambda_{1:2^{M-1}-2r}) \hat{\psi}_{j_M}(\lambda_{1:2^M-2r}^+). \tag{2.24}$$

For the binomial coefficients in (2.23), we use the following rule: for any  $A, B, C \in \mathbb{Z}$ ,

$$\binom{A}{C} \binom{B}{C} = 0 \text{ if } C < 0 \text{ or } \min\{A, B\} < C.$$

The proof of Lemma 2 is based on the mathematical induction method, which can be found in Section 4.1. For the functions generated by the recursive formula (2.23) and (2.24) with the initial term (2.13), we have the following observation, which will be used in the proof of Lemma 2 and Proposition 2. The first one is

$${}^M\tilde{G}_t^{(2^M-2\ell)}(\lambda_{1:2^M-2\ell}) = \exp\left(it\lambda_{1:2^M-2\ell}^+\right) {}^M\tilde{G}_0^{(2^M-2\ell)}(\lambda_{1:2^M-2\ell}) \tag{2.25}$$

for  $\ell = 0, 1, \dots, 2^{M-1} - 1$  and the second one is

$${}^M F_t^{(2^{M-1}-2r)}(\lambda_{1:2^M-2r}) = \exp\left(it\lambda_{1:2^M-2r}^+\right) {}^M F_0^{(2^{M-1}-2r)}(\lambda_{1:2^M-2r}) \tag{2.26}$$

for  $r = 0, 1, \dots, 2^{M-2} - 1$ , which are generalization of, for instance, (2.16) and (2.20).

### 2.4 Elements of Malliavin calculus

Let  $\mathcal{S}$  denote the set of all random variables  $S$  of the form

$$s(W(f_1), \dots, W(f_n)),$$

where  $n \geq 1$ ,  $s : \mathbb{R}^n \rightarrow \mathbb{R}$  is a  $C^\infty$ -function such that  $s$  and its partial derivatives have at most polynomial growth, and  $f_i \in \overline{H}$ ,  $i = 1, \dots, n$ . We denote by  $L^2(\Omega)$  the set of  $\mathcal{F}_W$ -measurable random variables whose second moments exist. The space  $\mathcal{S}$  is dense in  $L^2(\Omega)$  [22, Lemma 2.3.1].

**Definition 1.** For every integers  $n, p \geq 1$ , the  $p$ -th Malliavin derivative of

$$S = s(W(f_1), \dots, W(f_n)) \in \mathcal{S}$$

with respect to  $W$  is defined by

$$D^p S = \sum_{i_1, \dots, i_p=1}^n \frac{\partial^p s}{\partial x_{i_1} \dots \partial x_{i_p}}(W(f_1), \dots, W(f_n)) f_{i_1} \otimes \dots \otimes f_{i_p}. \tag{2.27}$$

Because the sum in (2.27) runs over all partial derivatives,  $D^p S$  belongs to  $L^2(\Omega, \overline{H}^{\odot p})$ . The Malliavin derivative of multiple Wiener-Itô integrals have the following property. The proof can be found in [22, Proposition 2.7.4].

**Lemma 3** ([22]). For every integer  $p \geq 1$  and  $u \in \overline{H}^{\otimes p}$ ,

$$DI_p(u) = pI_{p-1}(\tilde{u}), \tag{2.28}$$

where  $\tilde{u}$  is the canonical symmetrization of  $u$  defined through (2.11) and

$$I_{p-1}(\tilde{u})(\cdot) = \int_{\mathbb{R}^{p-1}} \tilde{u}(\lambda_1, \dots, \lambda_{p-1}, \cdot) W(d\lambda_1) \cdots W(d\lambda_{p-1}).$$

On the other hand, for all  $r \in \mathbb{N}$ ,  $p \in \mathbb{N}$  and  $q \in [1, \infty]$ ,

$$\|I_p(f)\|_{\mathbb{D}^{r,q}} \leq c_{r,p,q} \|f\|_{\overline{H}^{\otimes p}}, \quad f \in \overline{H}^{\odot p}, \tag{2.29}$$

where  $c_{r,p,q} > 0$  is an universal constant and

$$\|I_p(f)\|_{\mathbb{D}^{r,q}} := \left( \mathbb{E}[|I_p(f)|^q] + \mathbb{E}[\|DI_p(f)\|_{\overline{H}}^q] + \cdots + \mathbb{E}[\|D^r I_p(f)\|_{\overline{H}^{\otimes r}}^q] \right)^{1/q}. \tag{2.30}$$

We note that (2.29) is a special case of the Meyer’s inequality [24, Proposition 1.5.7]. Finally, for applying the normal approximation error bounds in [21, 22] to solve our problem, we need the following definition.

**Definition 2.** For any  $F \in L^2(\Omega, \mathcal{F}_W, \mathcal{P})$ , the pseudo-inverse of the infinitesimal generator  $L$  of the Ornstein-Uhlenbeck semigroup, denoted as  $L^{-1}$ , is defined as

$$L^{-1}F = -\sum_{p=1}^{\infty} \frac{1}{p} \text{Proj}(F | \mathcal{H}_p). \tag{2.31}$$

The details about the semigroup of the Ornstein-Uhlenbeck semigroup and its infinitesimal generator can be found in [22, Section 2.8].

### 3 Main results

As mentioned in Section 1, we will present large-scale limit theorems about the non-Gaussian random process  $U[j_{1:M}]X$  arising from STQNs of stationary Gaussian processes.

#### 3.1 Convergence rate of STQNs with random inputs

Let  $\mathcal{N}$  be a standard normal random variable and  $\sigma_{j_M}^2 = \mathbb{E}U[j_{1:M}]X(t)$ . The total variation (TV) distance between the distributions of  $\frac{1}{\sigma_{j_M}^2}U[j_{1:M}]X(2^{j_M}t)$  and the square of  $\mathcal{N}$  is defined by

$$d_{\text{TV}} \left( \frac{1}{\sigma_{j_M}^2}U[j_{1:M}]X(2^{j_M}t), \mathcal{N}^2 \right) = \sup_{h: \mathbb{R} \rightarrow [0,1]} \left| \mathbb{E} \left[ h \left( \frac{1}{\sigma_{j_M}^2}U[j_{1:M}]X(2^{j_M}t) \right) \right] - \mathbb{E} [h(\mathcal{N}^2)] \right|,$$

where the supremum is taken over Borel functions  $h$  taking values in  $[0, 1]$ . Because the composition function  $h(| \cdot |^2)$  also takes values in  $[0, 1]$ ,

$$\begin{aligned} d_{\text{TV}} \left( \frac{1}{\sigma_{j_M}^2}U[j_{1:M}]X(2^{j_M}t), \mathcal{N}^2 \right) &\leq \sup_{h: \mathbb{R} \rightarrow [0,1]} \left| \mathbb{E} \left[ h \left( \frac{1}{\sigma_{j_M}^2}T[j_{1:M}]X(2^{j_M}t) \right) \right] - \mathbb{E} [h(\mathcal{N})] \right| \\ &= d_{\text{TV}} \left( \frac{1}{\sigma_{j_M}^2}T[j_{1:M}]X(2^{j_M}t), \mathcal{N} \right). \end{aligned} \tag{3.1}$$

**Lemma 4** ([22]). Let  $F$  be a  $\mathcal{F}_W$ -measurable random variable such that  $\mathbb{E}[|F|^2] + \mathbb{E}[\|DF\|_H^2] < \infty$ ,  $\mathbb{E}[F] = 0$  and  $\mathbb{E}[F^2] = 1$ . Then

$$d_{TV}(F, \mathcal{N}) \leq 2\sqrt{\text{Var}(\langle DF, -DL^{-1}F \rangle)}.$$

In the proposition below, we apply Lemma 4 (see also Theorem 5.1.3 and Remark 5.1.4 in [22]) to the right hand side of (3.1) to get an upper bound for  $d_{TV}(\frac{1}{\sigma_{j_M}^2}U[j_{1:M}]X(2^{j_M}t), \mathcal{N}^2)$ .

**Proposition 1.** For every stationary Gaussian process  $X$ , integer  $M \geq 2$ , and  $j_{1:M} = [j_1, \dots, j_M] \in \mathbb{Z}^M$ , the TV distance between the distributions of the non-Gaussian variable  $\frac{1}{\sigma_{j_M}^2}U[j_{1:M}]X(2^{j_M}t)$  and the square of the standard normal random variable  $\mathcal{N}$  has an upper bound

$$d_{TV}\left(\frac{1}{\sigma_{j_M}^2}U[j_{1:M}]X(2^{j_M}t), \mathcal{N}^2\right) \leq \frac{2}{\sigma_{j_M}^2}\mathcal{U}_{j_M}(t, t), \tag{3.2}$$

where

$$\begin{aligned} \mathcal{U}_{j_M}(s, t) &= \sum_{r=0}^{2^{M-2}-1} (2^{M-1} - 2r) \sum_{\ell=1}^{2^{M-1}-2r-1} (\ell - 1)! \binom{2^{M-1} - 2r - 1}{\ell - 1}^2 \sqrt{(2^M - 4r - 2\ell)!} \\ &\times \|M_{2^{j_M} s}^{(2^{M-1}-2r)} \otimes_{\ell} M_{2^{j_M} t}^{(2^{M-1}-2r)}\|_{\overline{H}^{\otimes 2^M - 4r - 2\ell}} \\ &+ \sum_{\substack{r, r'=0 \\ r \neq r'}}^{2^{M-2}-1} (2^{M-1} - 2r') \sum_{\ell=1}^{2^{M-1}-2(r \vee r')} (\ell - 1)! \binom{2^{M-1} - 2r - 1}{\ell - 1} \binom{2^{M-1} - 2r' - 1}{\ell - 1} \\ &\times \sqrt{(2^M - 2r - 2r' - 2\ell)!} \|M_{2^{j_M} s}^{(2^{M-1}-2r)} \otimes_{\ell} M_{2^{j_M} t}^{(2^{M-1}-2r')}\|_{\overline{H}^{\otimes 2^M - 2r - 2r' - 2\ell}} \end{aligned} \tag{3.3}$$

for  $s, t \in \mathbb{R}$ , and the family of functions  $\{M_{2^{j_M} t}^{(2^{M-1}-2r)} \mid r = 0, 1, \dots, 2^{M-2} - 1\}$  is defined through the recursive formula (2.23) and (2.24).

In practice, the scale parameters are set with  $j_{m+1} \geq j_m$ , where  $m \in \mathbb{N}$ , for the continuous wavelet transform in the scattering network in order to use the neuron  $U[j_{m+1}]$  in the  $(m + 1)$ -th layer to extract larger-scale features from the outputs of the neuron  $U[j_m]$  in the  $m$ -th layer. This is supported by the fact that the spectrum of the magnitude of the continuous wavelet transform is mainly supported on the lower frequency area. See [20] for more discussion. For fixed  $j_1, j_2, \dots, j_{M-1} \in \mathbb{Z}$ , the behavior of  $\mathcal{U}_{j_M}$  and  $\sigma_{j_M}^2$  when  $j_M$  is sufficiently large will be discussed in Proposition 2, whose proof relies on the following estimate, and its proof is given in Section 4.3.

**Lemma 5.** For every  $M \in \mathbb{N}$ , there exists a constant  $C_M > 0$ , which depends on  $j_1, \dots, j_{M-1}$  but is independent of  $j_M$  and  $t$ , such that

$$\left| M_{\tilde{G}_t}^{(2^M-2\ell)}(\lambda_{1:2^M-2\ell}) \right| \leq C_M \prod_{k=1}^{2^M-2\ell} \sqrt{f_{X \star \psi_{j_1}}(\lambda_k)} \tag{3.4}$$

for  $\ell = 0, 1, \dots, 2^{M-1} - 1$ .

**Proposition 2.** (a) If the parameter  $\alpha$  in Assumption 1 and the parameter  $\beta$  in Assumption 2 satisfy  $2\alpha + \beta \geq 1$ , then for any  $j_1, \dots, j_{M-1} \in \mathbb{Z}$ , integer  $M \geq 2$ ,  $r, r' \in \{0, 1, \dots, 2^{M-2} - 1\}$ , and  $s, t \in \mathbb{R}$ ,

$$2^{\frac{3}{2}j_M} \left\| M_{2^{j_M} s}^{(2^{M-1}-2r)} \otimes_{\ell} M_{2^{j_M} t}^{(2^{M-1}-2r')} \right\|_{\overline{H}^{\otimes 2^M - 2r - 2r' - 2\ell}}$$

converges when  $j_M \rightarrow \infty$ , where  $\ell \in \{1, 2, \dots, 2^{M-1} - 2r - 1\}$  for the case  $r = r'$  and  $\ell \in \{1, 2, \dots, 2^{M-1} - 2(r \vee r')\}$  for the case  $r \neq r'$ . It further implies that  $\lim_{j_M \rightarrow \infty} 2^{\frac{3}{2}j_M} \mathcal{U}_{j_M}(s, t)$  exists and is finite, where  $\mathcal{U}_{j_M}$  is defined in (3.3).

(b) Under the same condition, i.e.,  $2\alpha + \beta \geq 1$ ,

$$\lim_{j_M \rightarrow \infty} 2^{j_M} \sigma_{j_M}^2 = \sum_{r=0}^{2^{M-2}-1} M_r! c^{(M_r)} \|\hat{\psi}\|^2,$$

where  $M_r = 2^{M-1} - 2r$  and

$$c^{(M_r)} = \int_{\mathbb{R}^{M_r-1}} \left| M^{-1} \tilde{G}_0^{(M_r)}(u_{1:M_r-1}, -u_{1:M_r-1}^+) \right|^2 du_{1:M_r-1}. \tag{3.5}$$

The condition  $2\alpha + \beta \geq 1$  is used for ensuring  $\|f_{X \star \psi_{j_1}}\|_\infty < \infty$ , which allows us to apply the Lebesgue dominated convergence theorem in the proof. Note that this condition could be easily achieved for widely used wavelets, including the Daubechies wavelets, the Morlet wavelet, the Mexican hat wavelet, etc, since  $\alpha \geq \frac{1}{2}$  holds for those wavelets. The proofs of Lemma 5 and Proposition 2 can be found in Section 4.3 and Section 4.4, respectively. By Propositions 1 and 2, we get the first main result as follows.

**Theorem 1.** Let  $\psi$  be a real-valued mother wavelet satisfying Assumption 1, and  $X$  be a stationary Gaussian process satisfying Assumption 2. If the parameter  $\alpha$  in Assumption 1 and the parameter  $\beta$  in Assumption 2 satisfy  $2\alpha + \beta \geq 1$ , then for each fixed  $t \in \mathbb{R}$ , integer  $M \geq 2$ , and  $j_1, \dots, j_{M-1} \in \mathbb{Z}$ , when  $j_M$  is sufficiently large,

$$d_{TV} \left( \frac{1}{\sigma_{j_M}^2} U[j_{1:M}] X(t), \mathcal{N}^2 \right) = O(2^{-\frac{1}{2}j_M}).$$

The theorem above implies that for each fixed  $t \in \mathbb{R}$ ,

$$\frac{1}{\sigma_{j_M}^2} U[j_{1:M}] X(t) \Rightarrow \mathcal{N}^2$$

in the distribution sense when  $j_M \rightarrow \infty$ , and hence together with Proposition 2 we have

$$2^{j_M} U[j_{1:M}] X(t) \Rightarrow \left[ \sum_{r=0}^{2^{M-2}-1} M_r! c^{(M_r)} \|\hat{\psi}\|^2 \right] \mathcal{N}^2$$

when  $j_M \rightarrow \infty$ .

### 3.2 Multivariate normal approximation of $2^{\frac{j_M}{2}} T[j_{1:M}] X(2^{j_M} t)$

In this section, we further show that  $2^{j_M} U[j_{1:M}] X$  converges in the finite dimensional distribution sense and derive the covariance function of the limiting process. Our approach is based on the following lemma coming from the multidimensional Stein’s method [22, Theorem 6.1.2].

**Lemma 6** ([22]). For any integer  $d \geq 2$ , let  $\mathbf{F} = (F_1, \dots, F_d)^\top$  be a  $\mathcal{F}_W$ -measurable random vector such that  $\mathbb{E}[|F_i|^4] + \mathbb{E}[\|DF_i\|_H^4] < \infty$  and  $\mathbb{E}[F_i] = 0$  for  $i = 1, \dots, d$ . Let  $[\mathbf{C}(i, j)]_{1 \leq i, j \leq d}$  be a symmetric non-negative definite matrix in  $\mathbb{R}^{d \times d}$ , and let  $\mathcal{N}_{\mathbf{C}}$  be a  $d$ -dimensional normal random vector with mean zero and covariance matrix  $\mathbf{C}$ . Then, for any twice differentiable function  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  with

$$\|h''\|_\infty := \sup_{1 \leq i, j \leq d} \sup_{x \in \mathbb{R}^d} \left| \frac{\partial^2 h}{\partial x_i \partial x_j}(x_1, \dots, x_d) \right| < \infty,$$

we have

$$|\mathbb{E}[h(\mathbf{F})] - \mathbb{E}[h(\mathcal{N}_{\mathbf{C}})]| \leq \frac{1}{2} \|h''\|_{\infty} \sqrt{\sum_{i,j=1}^d \mathbb{E} \left[ \left( \mathbf{C}(i,j) - \langle DF_j, -DL^{-1}F_i \rangle \right)^2 \right]}. \quad (3.6)$$

With Lemma 6, we proceed to study  $2^{j_M} U[j_{1:M}]X$  in the finite dimensional distribution sense. We start from checking conditions needed for Lemma 6. For any  $t_1, t_2, \dots, t_d \in \mathbb{R}$ , let

$$F_i := 2^{\frac{j_M}{2}} T[j_{1:M}]X(2^{j_M} t_i) = 2^{\frac{j_M}{2}} \left[ \sum_{r=0}^{2^{M-2}-1} I_{2^{M-1}-2r} \left( {}^M F_{2^{j_M} t_i}^{(2^{M-1}-2r)} \right) \right] \quad (3.7)$$

for  $i = 1, 2, \dots, d$ . By (2.29) and (2.30), the condition  $\mathbb{E}[|F_i|^4] + \mathbb{E}[\|DF_i\|_H^4] < \infty$  holds. On the other hand, because  $0 \leq r \leq 2^{M-2} - 1$  in the summation (3.7),  $F_i$  is a linear combination of Wiener chaos of order greater than or equal to 2, which implies that  $\mathbb{E}[F_i] = 0$ . Set the covariance matrix of  $(F_1, \dots, F_d)^{\top}$  as  $\mathbf{C}_{j_M} \in \mathbb{R}^{d \times d}$ ; that is,

$$\mathbf{C}_{j_M}(i, j) := \mathbb{E}[F_i F_j], \quad (3.8)$$

where  $1 \leq i, j \leq d$ , which satisfies the requirement of symmetric and non-negative definite for the matrix  $\mathbf{C}$  in Lemma 6. Moreover, we claim that it can be expressed as

$$\mathbf{C}_{j_M}(i, j) = 2^{j_M} \sum_{r=0}^{2^{M-2}-1} M_r! {}^M F_{2^{j_M} t_i}^{(M_r)} \otimes_{M_r} {}^M F_{2^{j_M} t_j}^{(M_r)}, \quad (3.9)$$

where  $M_r = 2^{M-1} - 2r$ . The proof of (3.9) can be found in Section 4.5. Because all conditions in Lemma 6 are satisfied, the inequality (3.6) holds with  $\mathbf{C} = \mathbf{C}_{j_M}$ . Next, we deal with the right hand side of (3.6). To this end, we apply the Malliavin calculus, the equality (3.9) and Proposition 2. The result is summarized as follows, and its proof is relegated to Section 4.6.

**Theorem 2.** Let  $\psi$  be a real-valued mother wavelet satisfying Assumption 1, and  $X$  be a stationary Gaussian process satisfying Assumption 2. Given integers  $d, M \geq 2$ ,  $j_1, \dots, j_M \in \mathbb{Z}$ , and  $t_1, t_2, \dots, t_d \in \mathbb{R}$ , define

$$\mathbf{F} = \left( 2^{\frac{j_M}{2}} T[j_{1:M}]X(2^{j_M} t_1), \dots, 2^{\frac{j_M}{2}} T[j_{1:M}]X(2^{j_M} t_d) \right)^{\top} \in \mathbb{R}^d$$

and let  $\mathbf{C}_{j_M}$  be its covariance matrix. If the parameter  $\alpha$  in Assumption 1 and the parameter  $\beta$  in Assumption 2 satisfy  $2\alpha + \beta \geq 1$ , then for any twice differentiable function  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  with  $\|h''\|_{\infty} < \infty$ , we have

$$\left| \mathbb{E}[h(\mathbf{F})] - \mathbb{E}[h(\mathcal{N}_{\mathbf{C}_{j_M}})] \right| \leq \frac{1}{2} \|h''\|_{\infty} \sum_{i,j=1}^d 2^{j_M} \mathcal{U}_{j_M}(t_j, t_i) = \|h''\|_{\infty} O(2^{-\frac{1}{2}j_M}) \quad (3.10)$$

when  $j_M$  is sufficiently large, where  $\mathcal{N}_{\mathbf{C}_{j_M}}$  is a  $d$ -dimensional normal random vector with mean zero and covariance matrix  $\mathbf{C}_{j_M}$ .

By considering the special cases  $h(\mathbf{x}) = \cos(\langle \zeta, \mathbf{x} \rangle)$  and  $h(\mathbf{x}) = \sin(\langle \zeta, \mathbf{x} \rangle)$ , where  $\mathbf{x}, \zeta \in \mathbb{R}^d$ , (3.10) implies that the distance between the characteristic function of the random vector  $\mathbf{F}$  and that of  $\mathcal{N}_{\mathbf{C}_{j_M}}$  converges to zero as  $j_M \rightarrow \infty$ . More precisely,

$$\left| \mathbb{E} \left[ e^{i\langle \zeta, \mathbf{F} \rangle} \right] - \mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_{j_M}} \rangle} \right] \right| \leq |\zeta|^2 O(2^{-\frac{1}{2}j_M}). \quad (3.11)$$

The following proposition shows that the covariance matrix  $\mathbf{C}_{j_M}$  converges as  $j_M \rightarrow \infty$ .

**Proposition 3.** Let the assumptions and notation of Theorem 2 prevail. Then, for any  $i, j \in \{1, 2, \dots, d\}$ , we have

$$\mathbf{C}_\infty(i, j) := \lim_{j_M \rightarrow \infty} \mathbf{C}_{j_M}(i, j) = \left[ \sum_{r=0}^{2^{M-2}-1} M_r! c^{(M_r)} \right] \int_{\mathbb{R}} e^{i(t_i-t_j)z} |\hat{\psi}(z)|^2 dz, \quad (3.12)$$

where  $M_r = 2^{M-1} - 2r$ , the constant  $c^{(M_r)}$  is defined in (3.5).

The proof of Proposition 3 is in Section 4.7. Let  $\mathcal{N}_{\mathbf{C}_\infty}$  be a  $d$ -dimensional normal random vector with mean zero and covariance matrix  $\mathbf{C}_\infty$ . Because both  $\mathcal{N}_{\mathbf{C}_{j_M}}$  and  $\mathcal{N}_{\mathbf{C}_\infty}$  are normal random vectors with mean 0, Proposition 3 implies that for any  $\zeta = [\zeta_1, \dots, \zeta_d] \in \mathbb{R}^d$

$$\mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_{j_M}} \rangle} \right] \rightarrow \mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_\infty} \rangle} \right] \quad (3.13)$$

when  $j_M \rightarrow \infty$ . By combining (3.11) and (3.13), we get

$$\begin{aligned} & \left| \mathbb{E} \left[ e^{i\langle \zeta, \mathbf{F} \rangle} \right] - \mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_\infty} \rangle} \right] \right| \\ & \leq \left| \mathbb{E} \left[ e^{i\langle \zeta, \mathbf{F} \rangle} \right] - \mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_{j_M}} \rangle} \right] \right| + \left| \mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_{j_M}} \rangle} \right] - \mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_\infty} \rangle} \right] \right| \rightarrow 0 \end{aligned}$$

when  $j_M \rightarrow \infty$ . Therefore, when  $j_M \rightarrow \infty$ ,

$$\mathbb{E} \left[ e^{i\langle \zeta, \mathbf{F} \rangle} \right] \rightarrow \mathbb{E} \left[ e^{i\langle \zeta, \mathcal{N}_{\mathbf{C}_\infty} \rangle} \right]$$

for all  $\zeta \in \mathbb{R}^d$ , where  $\mathcal{N}_{\mathbf{C}_\infty}$  is a  $d$ -dimensional normal random vector with mean zero and covariance matrix  $\mathbf{C}_\infty$  given in (3.12). It implies that the rescaled random process  $\{2^{\frac{j_M}{2}} T[j_{1:M}]X(2^{j_M}t) \mid t \in \mathbb{R}\}$  converges to a Gaussian process, denoted by  $\{G(t) \mid t \in \mathbb{R}\}$ , in the finite dimensional distribution sense and

$$\text{Cov}(G(t_1), G(t_2)) = \left[ \sum_{r=0}^{2^{M-2}-1} M_r! c^{(M_r)} \right] \int_{\mathbb{R}} e^{i(t_1-t_2)z} |\hat{\psi}(z)|^2 dz, \quad t_1, t_2 \in \mathbb{R}.$$

By the continuous mapping theorem and the well-known equality  $\text{Cov}(U^2, V^2) = 2\text{Cov}(U, V)^2$  when  $U$  and  $V$  are jointly Gaussian, we get a corresponding result for the rescaled random process  $\{2^{j_M} U[j_{1:M}]X(2^{j_M}t) \mid t \in \mathbb{R}\}$  as follows.

**Theorem 3.** Let the assumptions and notation of Theorem 2 prevail. The rescaled random process  $\{2^{j_M} U[j_{1:M}]X(2^{j_M}t) \mid t \in \mathbb{R}\}$  converges to the square of the stationary Gaussian process  $\{G(t) \mid t \in \mathbb{R}\}$  in the finite dimensional distribution sense. Moreover, the covariance function of the limiting process  $G^2$  has the following expression

$$\text{Cov}(G^2(t_1), G^2(t_2)) = 2\text{Cov}(G(t_1), G(t_2))^2$$

for  $t_1, t_2 \in \mathbb{R}$ .

Theorem 3 implies that when  $j_M \rightarrow \infty$ ,

$$2^{j_M} U[j_{1:M}]X(2^{j_M}t) \Rightarrow \left[ \sum_{r=0}^{2^{M-2}-1} M_r! c^{(M_r)} \|\hat{\psi}\|^2 \right] \chi^2(t)$$

in the finite dimensional distribution sense, where  $\{\chi^2(t) \mid t \in \mathbb{R}\}$  is a chi-square process with one degree of freedom and

$$\text{Cov}(\chi^2(t_1), \chi^2(t_2)) = 2 \left[ \int_{\mathbb{R}} e^{i(t_1-t_2)z} \left( \frac{|\hat{\psi}(z)|}{\|\hat{\psi}\|_2} \right)^2 dz \right]^2.$$

**Remark 3.** It is worth mentioning that in our previous work [16], we considered the second-order ST with stationary Gaussian processes and Gaussian processes with stationary increments as inputs, where we apply the Wiener chaos expansion and the Feynman diagram method to derive the central and non-central limit theorem, and the delicate interaction between mother wavelets and activation functions are explored. Different from [16], the analysis techniques in this work allows us to get a general form for the Wiener-Itô decomposition of the outputs of STQNs without depth limitation, and the current approach is non-asymptotic, i.e., the Stein’s method is applied to estimate the distance between the outputs of STQNs and their scaling limits. For commonly used wavelet functions, the current work provides information about the convergence speed of the rescaled random processes arising from STQNs, which cannot be obtained from our previous work. Furthermore, the technique used in the current work can be applied to study the same issue for polynomial activated STs by using more complicated product formula to obtain the Wiener-Itô decomposition of the corresponding outputs, which is not explored in this work.

## 4 Proofs

### 4.1 Proof of Lemma 2

In the following, we prove (2.21) and (2.22) by the mathematical induction method.

- When  $M = 2$ , not only (2.23) and (2.24) can be simplified as (2.19) and (2.16) respectively, but also (2.21) and (2.22) become (2.17) and (2.18). Hence, the statement in Lemma 2 holds for  $M = 2$ .

- Suppose that the statement in Lemma 2 holds for  $M = m$ , where  $m \geq 2$ .

- Now we start to prove the statement in Lemma 2 for  $M = m + 1$ . By the definition (2.2),

$$\begin{aligned} T[j_{1:m+1}]X(t) &= U[j_{1:m}]X \star \psi_{j_{m+1}}(t) \\ &= \int_{\mathbb{R}} \sum_{\ell=0}^{2^m-1} I_{2^m-2\ell} \left( {}^m\tilde{G}_{t'}^{(2^m-2\ell)} \right) \psi_{j_{m+1}}(t-t') dt', \end{aligned} \tag{4.1}$$

where the second equality follows from the induction hypothesis. By (2.25),

$$\int_{\mathbb{R}} {}^m\tilde{G}_{t'}^{(2^m-2\ell)}(\lambda_{1:2^m-2\ell}) \psi_{j_{m+1}}(t-t') dt' = {}^m\tilde{G}_t^{(2^m-2\ell)}(\lambda_{1:2^m-2\ell}) \hat{\psi}_{j_{m+1}}(\lambda_{1:2^m-2\ell}^+). \tag{4.2}$$

We recognize that the right hand side of (4.2) is just the right hand side of (2.24), i.e.  ${}^{m+1}F_t^{(2^m-2\ell)}(\lambda_{1:2^m-2\ell})$ . Hence, (4.1) and (4.2) imply that

$$T[j_{1:m+1}]X(t) = \sum_{\ell=0}^{2^m-1} I_{2^m-2\ell} \left( {}^{m+1}F_t^{(2^m-2\ell)} \right) \tag{4.3}$$

and we get (2.21) with  $M = m + 1$ . Next, by the relationship (2.3) and (4.3),

$$\begin{aligned} U[j_{1:m+1}]X(t) &= [T[j_{1:m+1}]X(t)]^2 = \left[ \sum_{r=0}^{2^m-1} I_{2^m-2r} \left( {}^{m+1}F_t^{(2^m-2r)} \right) \right]^2 \\ &= \sum_{r,r'=0}^{2^m-1} I_{2^m-2r} \left( {}^{m+1}F_t^{(2^m-2r)} \right) I_{2^m-2r'} \left( {}^{m+1}F_t^{(2^m-2r')} \right) \\ &= \sum_{r,r'=0}^{2^m-1} \sum_{p=0}^{2^m-2(r \vee r')} p! \binom{2^m-2r}{p} \binom{2^m-2r'}{p} I_{2^{m+1}-2(r+r'+p)} \left( {}^{m+1}F_t^{(2^m-2r)} \otimes_p {}^{m+1}F_t^{(2^m-2r')} \right), \end{aligned}$$



where the last equality follows from the product formula, i.e., Lemma 1. Denote  $\ell = r + r' + p$ . Because  $0 \leq p \leq 2^m - 2(r \vee r')$ , we have  $r + r' \leq \ell \leq 2^m - |r - r'|$ . Hence,

$$U[j_{1:m+1}]X(t) = \sum_{r,r'=0}^{2^{m-1}} \sum_{\ell=0}^{2^m} 1_{\{r+r' \leq \ell \leq 2^m - |r-r'|\}} (\ell - r - r')! \binom{2^m - 2r}{\ell - r - r'} \binom{2^m - 2r'}{\ell - r - r'} \times I_{2^{m+1}-2\ell} \left( {}^{m+1}F_t^{(2^m-2r)} \otimes_{\ell-r-r'} {}^{m+1}F_t^{(2^m-2r')} \right),$$

where the indicator function  $1_{\{r+r' \leq \ell \leq 2^m - |r-r'|\}}$  can be ignored by setting

$$\binom{2^m - 2r}{\ell - r - r'} \binom{2^m - 2r'}{\ell - r - r'} = 0 \text{ if } \ell - r - r' < 0 \text{ or } \min\{2^m - 2r, 2^m - 2r'\} < \ell - r - r'.$$

By the linearity of the Wiener integrals,

$$U[j_{1:m+1}]X(t) = \sum_{\ell=0}^{2^m} I_{2^{m+1}-2\ell} \left( \sum_{r,r'=0}^{2^{m-1}} (\ell - r - r')! \binom{2^m - 2r}{\ell - r - r'} \binom{2^m - 2r'}{\ell - r - r'} \times {}^{m+1}F_t^{(2^m-2r)} \otimes_{\ell-r-r'} {}^{m+1}F_t^{(2^m-2r')} \right).$$

We recognize that the function inside the integrand of the  $(2^{m+1} - 2\ell)$ -fold Wiener integrals above coincides with the right hand side of (2.23) with  $M = m + 1$ , i.e.,  ${}^{m+1}\tilde{G}_t^{(2^{m+1}-2\ell)}$ . Hence,

$$U[j_{1:m+1}]X(t) = \sum_{\ell=0}^{2^m} I_{2^{m+1}-2\ell} \left( {}^{m+1}\tilde{G}_t^{(2^{m+1}-2\ell)} \right),$$

which is just (2.22) with  $M = m + 1$ . Therefore, we finish the inductive step and the proof.

### 4.2 Proof of Proposition 1

From (2.21), we know that

$$T[j_{1:M}]X(2^{j_M}t) = \sum_{r=0}^{2^{M-2}-1} I_{2^{M-1}-2r} \left( {}^M F_{2^{j_M}t}^{(2^{M-1}-2r)} \right), \tag{4.4}$$

which satisfies  $\mathbb{E}[T[j_{1:M}]X(2^{j_M}t)] = 0$  and  $\mathbb{E}\left[|T[j_{1:M}]X(2^{j_M}t)|^2\right] = \sigma_{j_M}^2 < \infty$ . On the other hand, by (4.4) and Lemma 3,

$$DT[j_{1:M}]X(2^{j_M}t) = \sum_{r=0}^{2^{M-2}-1} (2^{M-1} - 2r) I_{2^{M-1}-2r-1} \left( {}^M F_{2^{j_M}t}^{(2^{M-1}-2r)} \right). \tag{4.5}$$

By (2.29) and (2.30),  $\mathbb{E}\left[\|DT[j_{1:M}]X(2^{j_M}t)\|_{\frac{2}{H}}^2\right] < \infty$ . The observations above show that  $F := \frac{1}{\sigma_{j_M}} T[j_{1:M}]X(2^{j_M}t)$  satisfies all conditions in Lemma 4. Hence, we have  $d_{TV}(F, \mathcal{N}) \leq 2\sqrt{\text{Var}(\langle DF, -DL^{-1}F \rangle)}$ . By the definition of  $L^{-1}$  in (2.31),

$$\begin{aligned} -DL^{-1}F &= D \left[ \frac{1}{\sigma_{j_M}} \sum_{r=0}^{2^{M-2}-1} \frac{1}{2^{M-1} - 2r} I_{2^{M-1}-2r} \left( {}^M F_{2^{j_M}t}^{(2^{M-1}-2r)} \right) \right] \\ &= \frac{1}{\sigma_{j_M}} \sum_{r=0}^{2^{M-2}-1} I_{2^{M-1}-2r-1} \left( {}^M F_{2^{j_M}t}^{(2^{M-1}-2r)} \right). \end{aligned} \tag{4.6}$$

By (4.5) and (4.6),

$$\text{Var}(\langle DF, -DL^{-1}F \rangle) = \text{Var}\left(\frac{1}{\sigma_{j_M}^2} \mathcal{I}\right), \tag{4.7}$$

where

$$\mathcal{I} := \left\langle \sum_{r=0}^{2^{M-2}-1} M_r I_{M_r-1} \left( M F_{2^{j_M} t}^{(M_r)} \right), \sum_{r=0}^{2^{M-2}-1} I_{M_r-1} \left( M F_{2^{j_M} t}^{(M_r)} \right) \right\rangle$$

and  $M_r = 2^{M-1} - 2r$ . Before computing  $\text{Var}(\mathcal{I})$ , we expand  $\mathcal{I}$  by Lemma 1 as follows

$$\begin{aligned} \mathcal{I} &= \sigma_{j_M}^2 + \sum_{r=0}^{2^{M-2}-1} M_r \sum_{\ell=1}^{M_r-1} (\ell-1)! \binom{M_r-1}{\ell-1}^2 I_{2M_r-2\ell} \left( M F_{2^{j_M} t}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t}^{(M_r)} \right) \\ &\quad + \sum_{\substack{r,r'=0 \\ r \neq r'}}^{2^{M-2}-1} M_{r'} \sum_{\ell=1}^{M_r \wedge M_{r'}} (\ell-1)! \binom{M_r-1}{\ell-1} \binom{M_{r'}-1}{\ell-1} I_{M_r+M_{r'}-2\ell} \left( M F_{2^{j_M} t}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t}^{(M_{r'})} \right). \end{aligned}$$

By the Minkowski inequality,

$$\begin{aligned} \sqrt{\text{Var}(\mathcal{I})} &= \mathbb{E} \left\{ \left[ \left| \mathcal{I} - \sigma_{j_M}^2 \right|^2 \right] \right\}^{\frac{1}{2}} \\ &\leq \sum_{r=0}^{2^{M-2}-1} M_r \sum_{\ell=1}^{M_r-1} (\ell-1)! \binom{M_r-1}{\ell-1}^2 \left\{ \mathbb{E} \left[ \left| I_{2M_r-2\ell} \left( M F_{2^{j_M} t}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t}^{(M_r)} \right) \right|^2 \right] \right\}^{\frac{1}{2}} \\ &\quad + \sum_{\substack{r,r'=0 \\ r \neq r'}}^{2^{M-2}-1} M_{r'} \sum_{\ell=1}^{M_r \wedge M_{r'}} (\ell-1)! \binom{M_r-1}{\ell-1} \binom{M_{r'}-1}{\ell-1} \left\{ \mathbb{E} \left[ \left| I_{M_r+M_{r'}-2\ell} \left( M F_{2^{j_M} t}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t}^{(M_{r'})} \right) \right|^2 \right] \right\}^{\frac{1}{2}}. \end{aligned} \tag{4.8}$$

For the expectations above, the isometry property of the Wiener-Itô integrals implies that

$$\mathbb{E} \left[ \left| I_{M_r+M_{r'}-2\ell} \left( M F_{2^{j_M} t}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t}^{(M_{r'})} \right) \right|^2 \right] = (M_r + M_{r'} - 2\ell)! \left\| M F_{2^{j_M} t}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t}^{(M_{r'})} \right\|^2.$$

Therefore, the upper bound in (3.2) follows from combining (4.7) and (4.8).

### 4.3 Proof of Lemma 5

We prove this lemma by mathematical induction. For  $M = 1$ , (2.13) implies that (3.4) holds with  $C_1 = 1$ . Suppose that (3.4) holds for  $M = m - 1$ , where  $m$  is an integer greater than or equal to 2. By (2.23),

$$m \tilde{G}_t^{(2^m-2\ell)} = \sum_{r,r'=0}^{2^{m-2}-1} (\ell-r-r')! \binom{2^{m-1}-2r}{\ell-r-r'} \binom{2^{m-1}-2r'}{\ell-r-r'} m F_t^{(2^{m-1}-2r)} \tilde{\otimes}_{\ell-r-r'} m F_t^{(2^{m-1}-2r')}. \tag{4.9}$$

From (2.24), we know that for  $r = 0, 1, \dots, 2^{m-2} - 1$ ,

$$\begin{aligned} & \left| mF_t^{(2^{m-1}-2r)}(x_1, \dots, x_{2^{m-1}-2r}) \right| \\ &= \left| m^{-1} \tilde{G}_t^{(2^{m-1}-2r)}(x_1, \dots, x_{2^{m-1}-2r}) \hat{\psi}_{j_m}(x_1, \dots, x_{2^{m-1}-2r}) \right| \\ &\leq \|\hat{\psi}\|_\infty \left| m^{-1} \tilde{G}_t^{(2^{m-1}-2r)}(x_1, \dots, x_{2^{m-1}-2r}) \right| \\ &\leq \|\hat{\psi}\|_\infty \left[ C_{m-1} \prod_{k=1}^{2^{m-1}-2r} \sqrt{f_{X^* \psi_{j_1}}(x_k)} \right], \end{aligned}$$

where the last inequality is obtained from the assumption, i.e., (4.9) with  $M = m - 1$ . Hence,

$$\begin{aligned} & \left| mF_t^{(2^{m-1}-2r)} \tilde{\otimes}_{\ell-r-r'} mF_t^{(2^{m-1}-2r')}(\lambda_{1:2^m-2\ell}) \right| \\ &\leq \|\hat{\psi}\|_\infty^2 C_{m-1}^2 \|f_{X^* \psi_{j_1}}\|_1^{\ell-r-r'} \prod_{k=1}^{2^m-2\ell} \sqrt{f_{X^* \psi_{j_1}}(\lambda_k)}. \end{aligned}$$

By substituting the inequality above into (4.9), we get

$$\begin{aligned} \left| m \tilde{G}_t^{(2^m-2\ell)}(\lambda_{1:2^m-2\ell}) \right| &\leq \left[ \sum_{r,r'=0}^{2^{m-2}-1} (\ell-r-r')! \binom{2^{m-1}-2r}{\ell-r-r'} \binom{2^{m-1}-2r'}{\ell-r-r'} \|f_{X^* \psi_{j_1}}\|_1^{\ell-r-r'} \right] \\ &\quad \times \left[ \|\hat{\psi}\|_\infty^2 C_{m-1}^2 \prod_{k=1}^{2^m-2\ell} \sqrt{f_{X^* \psi_{j_1}}(\lambda_k)} \right]. \end{aligned}$$

By the principle of mathematical induction, (3.4) holds for all  $M \in \mathbb{N}$ .

#### 4.4 Proof of Proposition 2

(a) For  $r, r' \in \{0, 1, \dots, 2^{M-2} - 1\}$ , by the recursive relationship (2.24),

$$\begin{aligned} & \left\| M F_{2^{j_M} s}^{(M_r)} \otimes_\ell M F_{2^{j_M} t}^{(M_{r'})} \right\|_{\overline{H}^{\otimes M_r+M_{r'}-2\ell}}^2 \\ &= \int_{\mathbb{R}^{M_r+M_{r'}-2\ell}} d\lambda_{1:M_r+M_{r'}-2\ell} \left[ \int_{\mathbb{R}^\ell} M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(\tau_{1:\ell}, \lambda_{1:M_r-\ell}) \hat{\psi}_{j_M}(\tau_{1:\ell}^+ + \lambda_{1:M_r-\ell}^+) \right. \\ &\quad \times M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-\tau_{1:\ell}, \lambda_{M_r-\ell+1:M_r+M_{r'}-2\ell}) \hat{\psi}_{j_M}(-\tau_{1:\ell}^+ + \lambda_{M_r-\ell+1:M_r+M_{r'}-2\ell}^+) d\tau_{1:\ell} \left. \right] \\ &\quad \times \left[ \int_{\mathbb{R}^\ell} M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(\eta_{1:\ell}, \lambda_{1:M_r-\ell}) \hat{\psi}_{j_M}(\eta_{1:\ell}^+ + \lambda_{1:M_r-\ell}^+) \right. \\ &\quad \left. \times M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-\eta_{1:\ell}, \lambda_{M_r-\ell+1:M_r+M_{r'}-2\ell}) \hat{\psi}_{j_M}(-\eta_{1:\ell}^+ + \lambda_{M_r-\ell+1:M_r+M_{r'}-2\ell}^+) d\eta_{1:\ell} \right], \end{aligned}$$

where  $\ell \in \{1, \dots, M_r - 1\}$  when  $r = r'$ , and  $\ell \in \{1, \dots, M_r \wedge M_{r'}\}$  when  $r \neq r'$ . By considering the change of variables

$$\begin{cases} u_i = \tau_i, & i = 1, \dots, \ell - 1 \text{ for } \ell > 1, \\ v_i = \eta_i, & i = 1, \dots, \ell - 1 \text{ for } \ell > 1, \\ w_i = \lambda_i, & i = 1, \dots, M_r + M_{r'} - 2\ell - 1, \\ x = 2^{j_M} (\tau_{1:\ell}^+ + \lambda_{1:M_r-\ell}^+), \\ y = 2^{j_M} (-\tau_{1:\ell}^+ + \lambda_{M_r-\ell+1:M_r+M_{r'}-2\ell}^+), \\ z = 2^{j_M} (\eta_{1:\ell}^+ + \lambda_{1:M_r-\ell}^+) \end{cases}$$

and noting that  $\hat{\psi}_{j_M}(\cdot) = \hat{\psi}(2^{j_M}\cdot)$ , we have

$$\begin{aligned}
 & 2^{3j_M} \left\| M F_{2^{j_M} s}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t}^{(M_{r'})} \right\|_{\overline{H}^{\otimes M_r + M_{r'} - 2\ell}}^2 \\
 = & \int_{\mathbb{R}^{M_r + M_{r'}}} M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(u_{1:\ell-1}, u_{\ell}^*, w_{1:M_r-\ell}) \\
 & \times M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-u_{1:\ell-1}, -u_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*) \\
 & \times \overline{M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(v_{1:\ell-1}, v_{\ell}^*, w_{1:M_r-\ell})} \\
 & \times \overline{M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-v_{1:\ell-1}, -v_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*)} \\
 & \times \hat{\psi}(x)\hat{\psi}(y)\hat{\psi}(z)\hat{\psi}(x+y-z) du_{1:\ell-1} dv_{1:\ell-1} dw_{1:M_r+M_{r'}-2\ell-1} dx dy dz, \tag{4.10}
 \end{aligned}$$

where  $u_{\ell}^* = 2^{-j_M}x - u_1 - \dots - u_{\ell-1} - w_1 - \dots - w_{M_r-\ell}$ ,  $v_{\ell}^* = 2^{-j_M}z - v_1 - \dots - v_{\ell-1} - w_1 - \dots - w_{M_r-\ell}$ , and  $w_{M_r+M_{r'}-2\ell}^* = 2^{-j_M}x + 2^{-j_M}y - w_1 - \dots - w_{M_r+M_{r'}-2\ell-1}$ . In the following, we show that the limit of the integrand on the right hand side of the equality (4.10) exists when  $j_M \rightarrow \infty$ .

- For the second component of the integrand in (4.10), by (2.25),

$$\begin{aligned}
 & M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-u_{1:\ell-1}, -u_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*) \\
 = & \exp \left\{ i2^{j_M} t \left[ -u_{1:\ell-1}^+ - u_{\ell}^* + w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}^+ + w_{M_r+M_{r'}-2\ell}^* \right] \right\} \\
 & \times M^{-1} \tilde{G}_0^{(M_{r'})}(-u_{1:\ell-1}, -u_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*) \\
 = & e^{ity} \times M^{-1} \tilde{G}_0^{(M_{r'})}(-u_{1:\ell-1}, -u_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*) \\
 \rightarrow & e^{ity} \times M^{-1} \tilde{G}_0^{(M_{r'})}(-u_{1:\ell-1}, u_{1:\ell-1}^+ + w_{1:M_r-\ell}^+, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1} \\
 & , -w_{1:M_r+M_{r'}-2\ell-1}^+)
 \end{aligned}$$

when  $j_M \rightarrow \infty$ .

- For the fourth component of the integrand in (4.10),

$$\begin{aligned}
 & M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-v_{1:\ell-1}, -v_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*) \\
 = & \exp \left\{ i2^{j_M} t \left[ -v_{1:\ell-1}^+ - v_{\ell}^* + w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}^+ + w_{M_r+M_{r'}-2\ell}^* \right] \right\} \\
 & \times M^{-1} \tilde{G}_0^{(M_{r'})}(-v_{1:\ell-1}, -v_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*) \\
 = & e^{it(x+y-z)} M^{-1} \tilde{G}_0^{(M_{r'})}(-v_{1:\ell-1}, -v_{\ell}^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*) \\
 \rightarrow & e^{it(x+y-z)} M^{-1} \tilde{G}_0^{(M_{r'})}(-v_{1:\ell-1}, v_{1:\ell-1}^+ + w_{1:M_r-\ell}^+, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1} \\
 & , -w_{1:M_r+M_{r'}-2\ell-1}^+)
 \end{aligned}$$

when  $j_M \rightarrow \infty$ .

- Similarly, for the first and third components of the integrand in (4.10), we have

$$M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(u_{1:\ell-1}, u_{\ell}^*, w_{1:M_r-\ell}) \rightarrow e^{isx} M^{-1} \tilde{G}_0^{(M_r)}(u_{1:\ell-1}, -u_{1:\ell-1}^+ - w_{1:M_r-\ell}^+, w_{1:M_r-\ell})$$

and

$$M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(v_{1:\ell-1}, v_{\ell}^*, w_{1:M_r-\ell}) \rightarrow e^{isz} M^{-1} \tilde{G}_0^{(M_r)}(v_{1:\ell-1}, -v_{1:\ell-1}^+ - w_{1:M_r-\ell}^+, w_{1:M_r-\ell})$$

respectively when  $j_M \rightarrow \infty$ .

On the other hand, by Lemma 5, for the integrand in (4.10), there exists a constant  $C > 0$ , which is independent of  $j_M$ , such that

$$\begin{aligned} & \sup_{x,y,z \in \mathbb{R}} \left| \begin{aligned} & M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(u_{1:\ell-1}, u_\ell^*, w_{1:M_r-\ell}) \\ & \times \frac{M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-u_{1:\ell-1}, -u_\ell^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*)}{M^{-1} \tilde{G}_{2^{j_M} s}^{(M_r)}(v_{1:\ell-1}, v_\ell^*, w_{1:M_r-\ell})} \\ & \times \frac{M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-v_{1:\ell-1}, -v_\ell^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*)}{M^{-1} \tilde{G}_{2^{j_M} t}^{(M_{r'})}(-v_{1:\ell-1}, -v_\ell^*, w_{M_r-\ell+1:M_r+M_{r'}-2\ell-1}, w_{M_r+M_{r'}-2\ell}^*)} \end{aligned} \right| \\ & \leq C \left[ \sup_{x,y,z \in \mathbb{R}} f_{X^* \psi_{j_1}}(u_\ell^*) f_{X^* \psi_{j_1}}(v_\ell^*) f_{X^* \psi_{j_1}}(w_{M_r+M_{r'}-2\ell}^*) \right] \\ & \quad \times \left[ \prod_{k=1}^{\ell-1} f_{X^* \psi_{j_1}}(u_k) f_{X^* \psi_{j_1}}(v_k) \right] \left[ \prod_{k=1}^{M_r+M_{r'}-2\ell-1} f_{X^* \psi_{j_1}}(w_k) \right]. \end{aligned} \tag{4.11}$$

If  $2\alpha + \beta \geq 1$ , then  $f_{X^* \psi_{j_1}}(\cdot) = 2^{2\alpha j_1} |C_{\hat{\psi}}(2^{j_1} \cdot)|^2 C_X(\cdot) \cdot |^{2\alpha+\beta-1}$  is not only integrable, but also bounded, which implies that the right hand side of (4.11) is integrable with respect to  $(u_{1:\ell-1}, v_{1:\ell-1}, w_{1:M_r+M_{r'}-2\ell-1})$ . By the Lebesgue dominated convergence theorem, we know that  $\lim_{j_M \rightarrow \infty} 2^{3j_M} \|M F_{2^{j_M} s}^{(M_r)} \otimes_\ell M F_{2^{j_M} t}^{(M_{r'})}\|_{\overline{H}^{\otimes M_r+M_{r'}-2\ell}}^2$  exists and satisfies

$$\begin{aligned} & \lim_{j_M \rightarrow \infty} 2^{3j_M} \|M F_{2^{j_M} s}^{(M_r)} \otimes_\ell M F_{2^{j_M} t}^{(M_{r'})}\|_{\overline{H}^{\otimes M_r+M_{r'}-2\ell}}^2 \\ & \leq C \|f_{X^* \psi_{j_1}}\|_1^{M_r+M_{r'}-3} \|f_{X^* \psi_{j_1}}\|_\infty^3 \int_{\mathbb{R}^3} |\hat{\psi}(x) \hat{\psi}(y) \overline{\hat{\psi}(z)} \hat{\psi}(x+y-z)| dx dy dz. \end{aligned}$$

The proof of the statement (a) is completed.

**(b)** For  $r \in \{0, 1, \dots, 2^{M-2} - 1\}$ , by (2.21) and the orthogonal property of multiple Wiener-Itô integrals,

$$\sigma_{j_M}^2 = \mathbb{E} \left| T[j_{1:M}] X(2^{j_M} t) \right|^2 = \mathbb{E} \left| \sum_{r=0}^{2^{M-2}-1} I_{M_r} \left( M F_{2^{j_M} t}^{(M_r)} \right) \right|^2 = \sum_{r=0}^{2^{M-2}-1} M_r! \|M F_{2^{j_M} t}^{(M_r)}\|_{\overline{H}^{\otimes M_r}}^2. \tag{4.12}$$

From (2.26), we know that the value of  $\|M F_{2^{j_M} t}^{(M_r)}\|_{\overline{H}^{\otimes M_r}}^2$  does not change along with the time variable. Hence, we choose  $t = 0$ . From (2.24), for each  $r \in \{0, 1, \dots, 2^{M-2} - 1\}$ ,

$$\|M F_0^{(M_r)}\|_{\overline{H}^{\otimes M_r}}^2 = \int_{\mathbb{R}^{M_r}} \left| M^{-1} \tilde{G}_0^{(M_r)}(\lambda_{1:M_r}) \hat{\psi}_{j_M}(\lambda_{1:M_r}^+) \right|^2 d\lambda_{1:M_r}.$$

By changing of variables  $u_i = \lambda_i$  for  $i = 1, \dots, M_r - 1$  and  $z = 2^{j_M}(\lambda_1 + \dots + \lambda_{M_r})$ ,

$$2^{j_M} \|M F_0^{(M_r)}\|_{\overline{H}^{\otimes M_r}}^2 = \int_{\mathbb{R}^{M_r}} \left| M^{-1} \tilde{G}_0^{(M_r)}(u_{1:M_r-1}, 2^{-j_M} z - u_{1:M_r-1}^+) \right|^2 |\hat{\psi}(z)|^2 du_{1:M_r-1} dz.$$

For the integrand of the integral above, by Lemma 5, there exists a constant  $C_M$ , which is independent of  $j_M$ , such that

$$\left| M^{-1} \tilde{G}_0^{(M_r)}(u_{1:M_r-1}, 2^{-j_M} z - u_{1:M_r-1}^+) \right|^2 \leq C_M^2 f_{X^* \psi_{j_1}}(2^{-j_M} z - u_{1:M_r-1}^+) \prod_{k=1}^{M_r-1} f_{X^* \psi_{j_1}}(u_k).$$

Under the assumption  $2\alpha + \beta \geq 1$ ,  $f_{X^* \psi_{j_1}}$  is not only integrable, but also bounded. Hence, we can again apply the Lebesgue dominated convergence theorem to get

$$\lim_{j_M \rightarrow \infty} 2^{j_M} \|M F_0^{(M_r)}\|_{\overline{H}^{\otimes M_r}}^2 = c^{(M_r)} \|\hat{\psi}\|^2, \tag{4.13}$$

where  $c^{(M_r)}$  is defined in (3.5). The proof is finished by combining (4.12) and (4.13).

**4.5 Proof of the equality (3.9)**

From the definition of  $C_{j_M}(i, j)$  in (3.8), i.e.,  $C_{j_M}(i, j) = \mathbb{E}[F_i F_j]$  and the definition of  $F_i$  in (3.7), i.e.,  $F_i = 2^{\frac{j_M}{2}} T[j_{1:M}]X(2^{j_M} t_i)$ , we have

$$C_{j_M}(i, j) = 2^{j_M} \mathbb{E} \left[ T[j_{1:M}]X(2^{j_M} t_i) T[j_{1:M}]X(2^{j_M} t_j) \right].$$

By replacing  $T[j_{1:M}]X$  with its Wiener-Itô decomposition in (2.21),

$$\begin{aligned} C_{j_M}(i, j) &= 2^{j_M} \mathbb{E} \left\{ \left[ \sum_{r=0}^{2^{M-2}-1} I_{M_r} \left( M_{F_{2^{j_M} t_i}}^{(M_r)} \right) \right] \left[ \sum_{r'=0}^{2^{M-2}-1} I_{M_{r'}} \left( M_{F_{2^{j_M} t_j}}^{(M_{r'})} \right) \right] \right\} \\ &= 2^{j_M} \sum_{r=0}^{2^{M-2}-1} \mathbb{E} \left[ I_{M_r} \left( M_{F_{2^{j_M} t_i}}^{(M_r)} \right) I_{M_r} \left( M_{F_{2^{j_M} t_j}}^{(M_r)} \right) \right], \end{aligned} \tag{4.14}$$

where the last equality follows from the orthogonal property of the multiple Wiener integrals. For  $r \in \{0, 1, \dots, 2^{M-2} - 1\}$ , by Lemma 1,

$$I_{M_r} \left( M_{F_{2^{j_M} t_i}}^{(M_r)} \right) I_{M_r} \left( M_{F_{2^{j_M} t_j}}^{(M_r)} \right) = \sum_{\ell=0}^{M_r} \ell! \binom{M_r}{\ell} \binom{M_r}{\ell} I_{2M_r-2\ell} \left( M_{F_{2^{j_M} t_i}}^{(M_r)} \otimes_{\ell} M_{F_{2^{j_M} t_j}}^{(M_r)} \right).$$

Because the expectation of the Wiener integrals is equal to zero except the constant term,

$$\mathbb{E} \left[ I_{M_r} \left( M_{F_{2^{j_M} t_i}}^{(M_r)} \right) I_{M_r} \left( M_{F_{2^{j_M} t_j}}^{(M_r)} \right) \right] = M_r! M_{F_{2^{j_M} t_i}}^{(M_r)} \otimes_{M_r} M_{F_{2^{j_M} t_j}}^{(M_r)}. \tag{4.15}$$

The equality (3.9) follows from substituting (4.15) into (4.14).

**4.6 Proof of Theorem 2**

As mentioned before the statement of Theorem 2, all conditions in Lemma 6 are satisfied when the random vector  $\mathbf{F}$  and the matrix  $\mathbf{C}$  are given by (3.7) and (3.8). Hence, it suffices to study the behavior of  $\mathbb{E} \left[ \left( C_{j_M}(i, j) - \langle DF_j, -DL^{-1}F_i \rangle \right)^2 \right]$ .

By (2.31) and (2.28), we have

$$DF_j = \sum_{r=0}^{2^{M-2}-1} M_r I_{M_r-1} \left( M_{F_{2^{j_M} t_j}}^{(M_r)} \right)$$

and

$$-DL^{-1}F_i = 2^{\frac{j_M}{2}} D \left[ \sum_{r=0}^{2^{M-2}-1} \frac{1}{M_r} I_{M_r} \left( M_{F_{2^{j_M} t_i}}^{(M_r)} \right) \right] = 2^{\frac{j_M}{2}} \left[ \sum_{r=0}^{2^{M-2}-1} I_{M_r-1} \left( M_{F_{2^{j_M} t_i}}^{(M_r)} \right) \right].$$

Therefore,

$$\langle DF_j, -DL^{-1}F_i \rangle = 2^{j_M} \left\langle \sum_{r=0}^{2^{M-2}-1} M_r I_{M_r-1} \left( M_{F_{2^{j_M} t_j}}^{(M_r)} \right), \sum_{r'=0}^{2^{M-2}-1} I_{M_{r'}-1} \left( M_{F_{2^{j_M} t_i}}^{(M_{r'})} \right) \right\rangle. \tag{4.16}$$

By Lemma 1, (4.16) can be rewritten as follows

$$\langle DF_j, -DL^{-1}F_i \rangle = 2^{j_M} \sum_{r=0}^{2^{M-2}-1} M_r! \left[ M_{F_{2^{j_M} t_j}}^{(M_r)} \otimes_{M_r} M_{F_{2^{j_M} t_i}}^{(M_r)} \right] + \mathcal{I}_1 + \mathcal{I}_2, \tag{4.17}$$

where

$$\mathcal{I}_1 = 2^{j_M} \sum_{r=0}^{2^{M-2}-1} M_r \sum_{\ell=1}^{M_r-1} (\ell-1)! \binom{M_r-1}{\ell-1}^2 I_{2M_r-2\ell} \left( M F_{2^{j_M} t_j}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t_i}^{(M_r)} \right)$$

and

$$\mathcal{I}_2 = 2^{j_M} \sum_{\substack{r,r'=0 \\ r \neq r'}}^{2^{M-2}-1} M_r \sum_{\ell=1}^{M_r \wedge M_{r'}} (\ell-1)! \binom{M_r-1}{\ell-1} \binom{M_{r'}-1}{\ell-1} I_{M_r+M_{r'}-2\ell} \left( M F_{2^{j_M} t_j}^{(M_r)} \otimes_{\ell} M F_{2^{j_M} t_i}^{(M_{r'})} \right).$$

We recognize that the first term in (4.17) coincides with the expression (3.9) of  $\mathbf{C}_{j_M}(i, j)$ . By the Minkowski inequality,

$$\begin{aligned} & \sqrt{\mathbb{E} \left[ \left( \mathbf{C}_{j_M}(i, j) - \langle DF_j, -DL^{-1}F_i \rangle \right)^2 \right]} = \sqrt{\mathbb{E} [|\mathcal{I}_1 + \mathcal{I}_2|^2]} \\ & \leq \sqrt{\mathbb{E} [|\mathcal{I}_1|^2]} + \sqrt{\mathbb{E} [|\mathcal{I}_2|^2]} = 2^{j_M} \mathcal{U}_{j_M}(t_j, t_i), \end{aligned}$$

where  $\mathcal{U}_{j_M}(\cdot, \cdot)$  is defined in (3.3). By Proposition 2, we know that  $\lim_{j_M \rightarrow \infty} 2^{\frac{3}{2}j_M} \mathcal{U}_{j_M}(\cdot, \cdot)$  exists. Hence, for  $i, j \in \{1, 2, \dots, d\}$ , we get

$$\sqrt{\mathbb{E} \left[ \left( \mathbf{C}_{j_M}(i, j) - \langle DF_j, -DL^{-1}F_i \rangle \right)^2 \right]} = O(2^{-\frac{1}{2}j_M})$$

when  $j_M$  is sufficiently large.

### 4.7 Proof of Proposition 3

By the equality (3.9), it suffices to prove that  $2^{j_M} M F_{2^{j_M} t_i}^{(M_r)} \otimes_{M_r} M F_{2^{j_M} t_j}^{(M_r)}$  converges when  $j_M \rightarrow \infty$  for each  $r \in \{0, 1, \dots, 2^{M-2} - 1\}$ . By (2.24),

$$\begin{aligned} & M F_{2^{j_M} t_i}^{(M_r)} \otimes_{M_r} M F_{2^{j_M} t_j}^{(M_r)} = \int_{\mathbb{R}^{M_r}} M F_{2^{j_M} t_i}^{(M_r)}(\lambda_{1:M_r}) \overline{M F_{2^{j_M} t_j}^{(M_r)}(\lambda_{1:M_r})} d\lambda_{1:M_r} \\ & = \int_{\mathbb{R}^{M_r}} \left[ M^{-1} \tilde{G}_{2^{j_M} t_i}^{(M_r)}(\lambda_{1:M_r}) \hat{\psi}_{j_M}(\lambda_{1:M_r}^+) \right] \overline{\left[ M^{-1} \tilde{G}_{2^{j_M} t_j}^{(M_r)}(\lambda_{1:M_r}) \hat{\psi}_{j_M}(\lambda_{1:M_r}^+) \right]} d\lambda_{1:M_r} \\ & = \int_{\mathbb{R}^{M_r}} \left[ M^{-1} \tilde{G}_{2^{j_M} t_i}^{(M_r)}(\lambda_{1:M_r}) \overline{M^{-1} \tilde{G}_{2^{j_M} t_j}^{(M_r)}(\lambda_{1:M_r})} \right] \left| \hat{\psi}_{j_M}(\lambda_{1:M_r}^+) \right|^2 d\lambda_{1:M_r}. \end{aligned}$$

From the recursive formula (2.23) and (2.24) with the initial term (2.13), we know that  $M^{-1} \tilde{G}_t^{(M_r)}(\lambda_{1:M_r})$  satisfies

$$M^{-1} \tilde{G}_t^{(M_r)}(\lambda_{1:M_r}) = M^{-1} \tilde{G}_0^{(M_r)}(\lambda_{1:M_r}) \exp(it\lambda_{1:M_r}^+).$$

Hence,

$$\begin{aligned} & M F_{2^{j_M} t_i}^{(M_r)} \otimes_{M_r} M F_{2^{j_M} t_j}^{(M_r)} \\ & = \int_{\mathbb{R}^{M_r}} \exp(i2^{j_M}(t_i - t_j)\lambda_{1:M_r}^+) \left| M^{-1} \tilde{G}_0^{(M_r)}(\lambda_{1:M_r}) \right|^{2M} \left| \hat{\psi}_{j_M}(\lambda_{1:M_r}^+) \right|^2 d\lambda_{1:M_r}. \end{aligned}$$

By changing of variables

$$\begin{cases} u_k = \lambda_k \text{ for } k = 1, 2, \dots, M_r - 1, \\ z = 2^{j_M} \lambda_{1:M_r}^+, \end{cases}$$

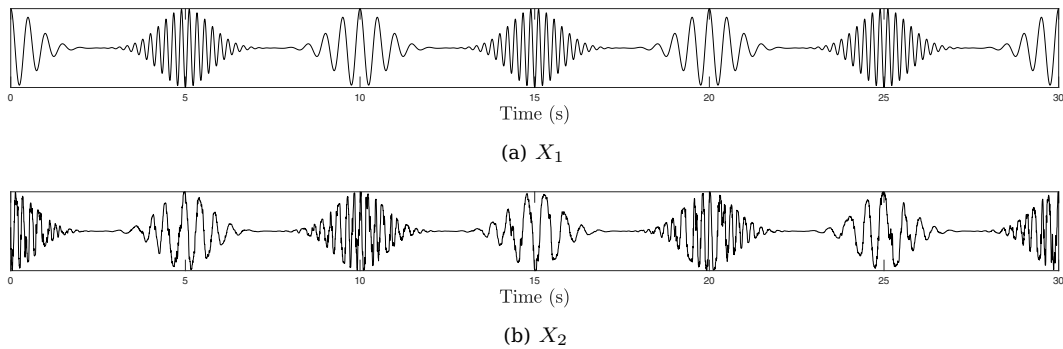


Figure 3: Testing signals

we get

$$2^{j_M} M_{F_{2^{j_M} t_i}}^{(M_r)} \otimes_{M_r} M_{F_{2^{j_M} t_j}}^{(M_r)} = \int_{\mathbb{R}^{M_r}} e^{i(t_i - t_j)z} \left| {}^{M-1}\tilde{G}_0^{(M_r)}(u_{1:M_r-1}, 2^{-j_M} z - u_{1:M_r-1}^+) \right|^2 |\hat{\psi}(z)|^2 du_{1:M_r-1} dz.$$

Under the condition  $2\alpha + \beta \geq 1$ , we apply Lemma 5 and the Lebesgue dominated convergence theorem to get

$$\lim_{j_M \rightarrow \infty} 2^{j_M} M_{F_{2^{j_M} t_i}}^{(M_r)} \otimes_{M_r} M_{F_{2^{j_M} t_j}}^{(M_r)} = \int_{\mathbb{R}} e^{i(t_i - t_j)z} \left\{ \int_{\mathbb{R}^{M_r-1}} \left| {}^{M-1}\tilde{G}_0^{(M_r)}(u_{1:M_r-1}, -u_{1:M_r-1}^+) \right|^2 du_{1:M_r-1} \right\} |\hat{\psi}(z)|^2 dz.$$

Note that the integral inside the curly brackets above is just the constant  $c^{(M_r)}$  defined in (3.5).

### Appendix: illustration of the stability of the multi-scale information extracted by STQN with respect to time shifting and small deformations

For the purpose of showing the effect of time shifting and distortion on the outputs of STQN, we consider two testing signals  $X_1$  and  $X_2$ , where

$$X_1(t) = \left[ \sum_{k \in \mathbb{Z}} e^{-(t-10k)^2} \right] \cos(2\pi\omega_1 t) + \left[ \sum_{k \in \mathbb{Z}} e^{-(t-10k+5)^2} \right] \cos(2\pi\omega_2 t), \quad t \in \mathbb{R},$$

$\omega_1 = 2$  and  $\omega_2 = 4.5$  (units: Hz), while  $X_2$  is obtained by time warping  $X_1$  and shifting the deformed signal by 5 seconds. These two signals are displayed in Figure 3. In Figure 4, the magnitudes of outputs of  $U[j_1]$ ,  $S_J[j_1]$ ,  $U[50, j_2]$ , and  $S_J[50, j_2]$  at different scales and time are represented by colors. The similarity between Figure 4(c) and Figure 4(d) (see also Figure 4(i) and Figure 4(j)), shows that the STQN coefficients are stable to time shifting and small deformations. Moreover, the location of peak scales in Figure 4(c) and 4(d) show that  $X_1$  and  $X_2$  have frequency components around 2 Hz and 4.5 Hz (corresponding to  $j_1 \approx 50$  and  $j_1 \approx 7$ ) during some periods within a 30-second interval. Figure 4(e) and Figure 4(f) were obtained by setting  $j_1 = 50$  and computing the values of  $U[50]X_1$  and  $U[50]X_2$  for each time point. These two figures show that the 2-Hz activities in  $X_1$  and  $X_2$  occur per 10 second interval approximately, i.e., at the frequency 0.1 Hz. This observation can also be obtained from the visualization of the second-order STQN coefficients in Figure 4(i) and Figure 4(j) by finding the peak scales. Hence, the STQN has the ability on extracting multi-scale information from signals.



## Asymptotic analysis of scattering transform of Gaussian processes

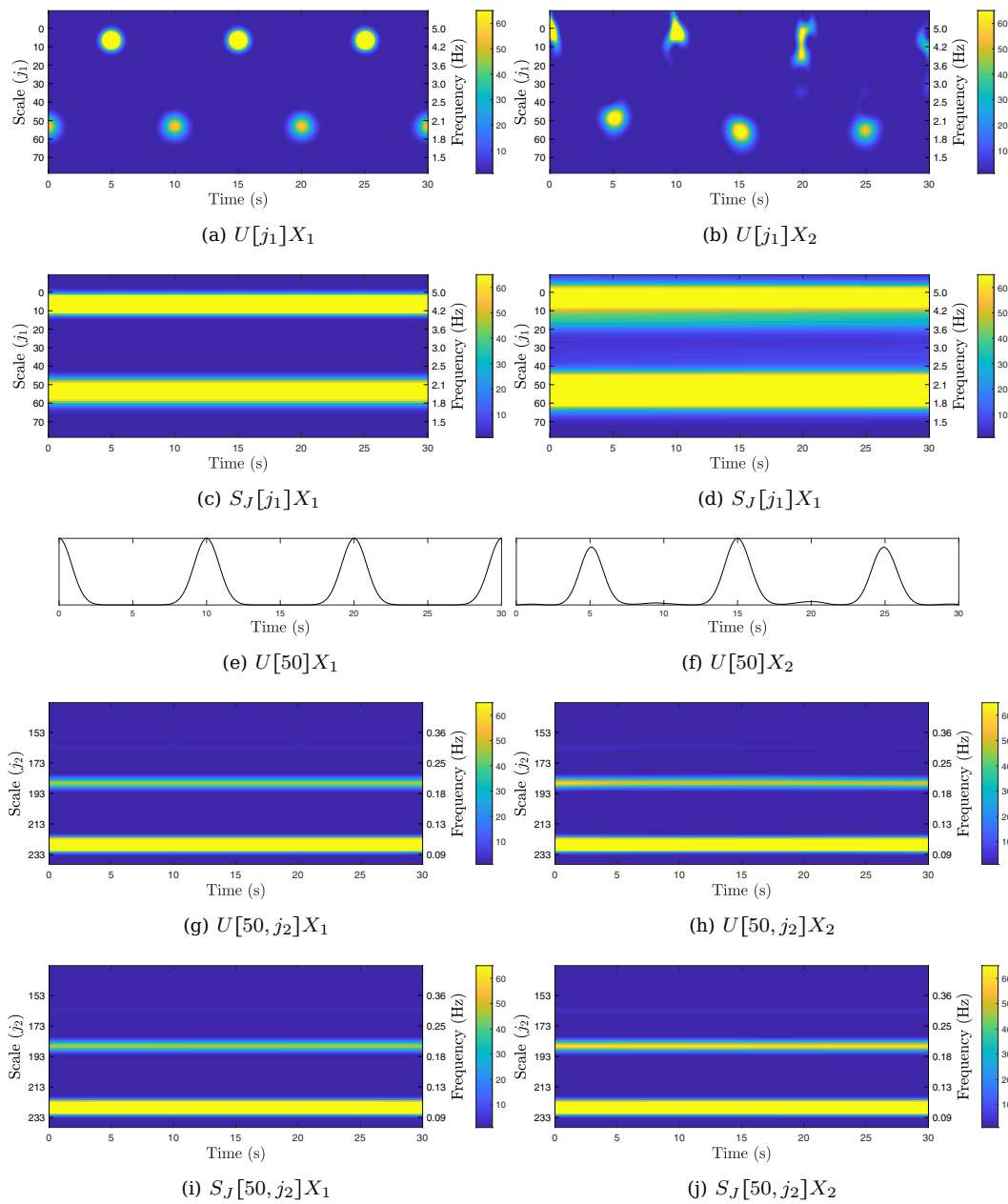


Figure 4: Illustration of the stability of the STQN coefficients to time shifting and small deformations.

### References

- [1] Andén, J., and Mallat, S. Multiscale scattering for audio classification. In *ISMIR* (2011), Miami, FL, pp. 657–662.
- [2] Andén, J., and Mallat, S. Deep scattering spectrum. *IEEE Transactions on Signal Processing* 62, 16 (2014), 4114–4128. MR3260414
- [3] Balan, R., Singh, M., and Zou, D. Lipschitz properties for deep convolutional networks. *Contemporary Mathematics* 706 (2018), 129–151. MR3796635
- [4] Balestriero, R., and Glotin, H. Linear time complexity deep Fourier scattering network and extension to nonlinear invariants. *arXiv preprint arXiv:1707.05841* (2017).

- [5] Bruna, J. *Scattering representations for recognition*. PhD thesis, Ecole Polytechnique X, 2013.
- [6] Bruna, J., Mallat, S., Bacry, E., and Muzy, J. F. Intermittent process analysis with scattering moments. *The Annals of Statistics* 43, 1 (2015), 323–351. MR3311862
- [7] Cheng, X., Chen, X., and Mallat, S. Deep Haar scattering networks. *Information and Inference: A Journal of the IMA* 5, 2 (2016), 105–133. MR3516855
- [8] Chudáček, V., Andén, J., Mallat, S., Abry, P., and Doret, M. Scattering transform for intra-partum fetal heart rate variability fractal analysis: a case-control study. *IEEE Transactions on Biomedical Engineering* 61, 4 (2013), 1100–1108.
- [9] Czaja, W., and Li, W. Analysis of time-frequency scattering transforms. *Applied and Computational Harmonic Analysis* 47, 1 (2019), 149–171. MR3958697
- [10] Doukhan, P., Oppenheim, G., and Taqqu, M. *Theory and applications of long-range dependence*. Springer Science & Business Media, 2002. MR1956041
- [11] Itô, K. Stationary random distributions. *Memoirs of the College of Science, University of Kyoto. Series A: Mathematics* 28, 3 (1954), 209–223. MR0065060
- [12] Kavalarov, I., Li, W., Czaja, W., and Chellappa, R. 3-D Fourier scattering transform and classification of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing* (2020).
- [13] Li, J., Ke, L., Du, Q., Ding, X., Chen, X., and Wang, D. Heart sound signal classification algorithm: A combination of wavelet scattering transform and twin support vector machine. *IEEE Access* 7 (2019), 179339–179348.
- [14] Liu, G. R., Lin, T. Y., Wu, H. T., Sheu, Y. C., Liu, C. L., Liu, W. T., Yang, M. C., Ni, Y. L., Chou, K. T., Chen, C. H., et al. Large-scale assessment of consistency in sleep stage scoring rules among multiple sleep centers using an interpretable machine learning algorithm. *Journal of Clinical Sleep Medicine* 17, 2 (2021), 159–166.
- [15] Liu, G. R., Lo, Y. L., Malik, J., Sheu, Y. C., and Wu, H. T. Diffuse to fuse EEG spectra–intrinsic geometry of sleep dynamics for classification. *Biomedical Signal Processing and Control* 55 (2020), 101576.
- [16] Liu, G. R., Sheu, Y. C., and Wu, H. T. Central and non-central limit theorems arising from the scattering transform and its neural activation generalization. *arXiv preprint arXiv:2011.10801* (2020).
- [17] Liu, G. R., Sheu, Y. C., and Wu, H. T. When scattering transform meets non-Gaussian random processes, a double scaling limit result. *arXiv preprint arXiv:2112.13475* (2021).
- [18] Lostanlen, V., Cohen-Hadria, A., and Bello, J. P. One or two frequencies? The scattering transform answers. In *2020 28th European Signal Processing Conference (EUSIPCO)* (2021), IEEE, pp. 2205–2209.
- [19] Major, P. *Muliple Wiener-Itô Integrals*, vol. 849. Springer, 2014. MR3155040
- [20] Mallat, S. Group invariant scattering. *Communications on Pure and Applied Mathematics* 65, 10 (2012), 1331–1398. MR2957703
- [21] Nourdin, I., and Peccati, G. Stein’s method on Wiener chaos. *Probability Theory and Related Fields* 145, 1-2 (2009), 75–118. MR2520122
- [22] Nourdin, I., and Peccati, G. *Normal approximations with Malliavin calculus: from Stein’s method to universality*. No. 192. Cambridge University Press, 2012. MR2962301
- [23] Nourdin, I., Peccati, G., and Reinert, G. Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. *The Annals of Probability* 38, 5 (2010), 1947–1985. MR2722791
- [24] Nualart, D. *The Malliavin calculus and related topics*, vol. 1995. Springer, 2006. MR1344217
- [25] Pipiras, V., and Taqqu, M. S. Regularization and integral representations of Hermite processes. *Statistics & Probability Letters* 80, 23-24 (2010), 2014–2023. MR2734275
- [26] Pipiras, V., and Taqqu, M. S. *Long-range dependence and self-similarity*, vol. 45. Cambridge University press, 2017. MR3729426
- [27] Wiatowski, T., and Bölcskei, H. Deep convolutional neural networks based on semi-discrete frames. In *2015 IEEE International Symposium on Information Theory (ISIT)* (2015), IEEE, pp. 1212–1216.

- [28] Wiatowski, T., and Bölcskei, H. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Transactions on Information Theory* 64, 3 (2017), 1845–1866. MR3766318
- [29] Wu, H. T., Talmon, R., and Lo, Y. L. Assess sleep stage by modern signal processing techniques. *IEEE Transactions on Biomedical Engineering* 62, 4 (2014), 1159–1168.
- [30] Yaglom, A. M. Some classes of random fields in n-dimensional space, related to stationary random processes. *Theory of Probability & Its Applications* 2, 3 (1957), 273–320. MR0094844

**Acknowledgments.** The authors would like to thank the editors for handling their paper and the anonymous reviewers for their valuable comments. Their comments have significantly improved the quality of their paper. G. R. Liu’s work was supported by the Ministry of Science and Technology under Grant 110-2628-M-006-003-MY3. Y. C. Sheu’s work was supported by the Ministry of Science and Technology under Grant 110-2115-M-A49 -007.

---

# Electronic Journal of Probability

## Electronic Communications in Probability

---

### Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS<sup>1</sup>)
- Easy interface (EJMS<sup>2</sup>)

### Economical model of EJP-ECP

- Non profit, sponsored by IMS<sup>3</sup>, BS<sup>4</sup>, ProjectEuclid<sup>5</sup>
- Purely electronic

### Help keep the journal free and vigorous

- Donate to the IMS open access fund<sup>6</sup> (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

---

<sup>1</sup>LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

<sup>2</sup>EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

<sup>3</sup>IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

<sup>4</sup>BS: Bernoulli Society <http://www.bernoulli-society.org/>

<sup>5</sup>Project Euclid: <https://projecteuclid.org/>

<sup>6</sup>IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>