

# Time-causal and time-recursive wavelets

Tony Lindeberg

**Abstract** This paper presents a framework for time-causal wavelet analysis. It targets real-time processing of temporal signals, where data from the future are not available.

The study builds upon temporal scale-space theory, originating from a complete classification of temporal smoothing kernels that guarantee non-creation of new structures from finer to coarser temporal scale levels. We construct temporal wavelets from the temporal derivatives of a special time-causal smoothing kernel, referred to as the time-causal limit kernel, as arising from the classification of variation-diminishing smoothing transformations with the complementary requirement of temporal scale covariance, to guarantee self-similar handling of structures in the input signal at different temporal scales. This enables decomposition of the signal into different components at different scales, while adhering to temporal causality.

The paper establishes theoretical foundations for these time-causal wavelet representations, and maps structural relationships to the non-causal Ricker wavelets. We also describe how efficient discrete approximations of the presented theory can be performed in terms of first-order recursive filters coupled in cascade, which enables numerically well-conditioned real-time processing with low resource usage. We characterize and quantify how the continuous scaling properties transfer to the discrete implementation, demonstrating how the proposed time-causal wavelet representation can reflect the duration of locally dominant temporal structures in the input signal.

We propose this notion of time-causal wavelet analysis as a generic multi-purpose tool for signal processing tasks, where streams of signals are to be processed in real time, specifically for signals that may contain local variations over

a rich span of temporal scales, or more generally for analysing physical or biophysical temporal phenomena, where a fully time-causal analysis is called for to be physically realistic.

**Keywords** Time · Wavelet · Temporal · Scale · Time-causal · Time-recursive · Scale covariance · Filter · Signal processing

## 1 Introduction

Temporal signals can inherently contain different types of structures at different temporal scales. Making such multi-scale relationships in the data explicit, and exploring these relationships, can often substantially improve the performance of signal processing pipelines. For the purpose of handling such scale variations in not necessarily only temporal signals, the notion of wavelet analysis has therefore been developed by the signal processing community, to decompose signals into different components at different scales (Ricker 1944, Morlet *et al.* 1982a, 1982b, Grossmann and Morlet 1984, Hosken 1988, Mallat 1989, 1999, Heil and Walnut 1989, Strang 1989, Daubechies 1992, Meyer 1992, Chui 1992, Ruskain *et al.* 1992, Donoho and Johnston 1993, Teolis 1998, Walker 1999, Debnath and Shah 2002, Misiti *et al.* 2007). Due to their scale separating properties and their computational efficiency, wavelet representations are used as a powerful generic building block for representing the information contents in image data and time-dependent signals for a very wide range of signal processing methods and applications.

While there have been a few wavelet formulations that respect temporal causality (Szu *et al.* 1992, Vázquez *et al.* 2005), traditional wavelet theory is, however, largely non-causal, making use of filters that are often symmetric or antisymmetric with respect to the temporal origin. Such an approach is, however, not feasible in real-time situations, when the future cannot be accessed. While one could, in principle, apply non-causal wavelet theory to real-time signals by introducing explicit temporal buffers with sufficiently long

---

The support from the Swedish Research Council (contract 2022-02969) is gratefully acknowledged.

Tony Lindeberg, Computational Brain Science Lab, Division of Computational Science and Technology, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. E-mail: tony@kth.se ORCID: 0000-0002-9081-2170

temporal delays, beyond which the wavelets would be truncated to zero, for time-critical applications, where one wants to minimize the temporal delays, it is of interest to develop truly time-causal wavelets, that do not make use of any information from the future.

The subject of this paper is to show how a principled temporal wavelet theory can be developed based on the notion of time-causal temporal scale-space representation, originally developed for video analysis in Lindeberg (2016), and then dedicated for a pure time-causal temporal domain in Lindeberg (2023). This temporal scale-space theory, which determines canonical filter shapes for the temporal wavelets, is based on convolution with a special temporal smoothing kernel, denoted the time-causal limit kernel, and which corresponds to the convolution of an infinite set of truncated exponential kernels in cascade, with especially chosen time constants to obtain temporal scale covariance.

Temporal scale covariance in this respect corresponds to self-similarity over temporal scales, so that the temporal scale-space representation computed for a rescaled signal corresponds to a rescaling of the temporal scale-space representation of the original signal, complemented with a shift in the dimension of the temporal scale parameter. By combining this type of time-causal temporal scale-space representation with temporal derivatives, we can construct mother wavelets for time-causal wavelet representations from temporal derivatives of the time-causal limit kernel. In this way, the resulting time-causal wavelets will have the ability to handle temporal structures at different temporal scales in a provably self-similar manner, and can in this way constitute the foundation for defining time-causal and scale-covariant temporal basis functions, to be used as computational primitives in more composed signal processing or learning-based methods.

Furthermore, by forming the differences between temporal scale-space representations at adjacent temporal scales, we can construct time-causal bandpass wavelet representations, which are here formally shown to up to a scaling factor correspond to the time-causal wavelet representation based on first-order temporal derivatives of the time-causal limit kernel. From such a temporal bandpass representation, reconstruction of the original signal is straightforward, by mere addition over scales of the time-causal bandpass representations, which opens up possibilities for real-time manipulation of the signal as decomposed into a time-causal wavelet representation. This situation is structurally very closely related to the corresponding property of the Ricker or Mexican hat wavelet, based on second-order derivatives of the non-causal Gaussian kernel, where reconstruction, as will be shown here, can be directly performed by integrating the corresponding second-order non-causal wavelet representations over the scale parameter, or the difference-of-Gaussians wavelet, where reconstruction is performed by instead summing up the difference-of-Gaussians wavelet representations over the temporal scales.

By the time-causal limit kernel constituting the composition of a set of truncated exponential kernels in cascade, the temporal smoothing operation equivalently corresponds to applying a set of first-order integrators in cascade. In terms of discrete implementation, such an operation can be approximated by a set of first-order recursive filters coupled in cascade. Thereby, the resulting implementation is fully time-recursive, meaning that no other temporal memory of the past is needed, beyond the information contained in the temporal scale channels themselves. Specifically, this implies straightforward and very efficient implementation schemes on regular signal processing architectures.

By combining such mother wavelets with different types of normalizations with respect to the scale parameter, either in terms of  $L_p$ -norms or the notion of scale-normalized derivatives, scale selective signal processing algorithms can then be designed, that lead to extrema over scales in the magnitude of the resulting time-causal wavelet representations at temporal scale levels proportional to characteristic durations of different types of temporal structures in the input data (Lindeberg 2017, 2018b, 2018a), with close relationships to the previously studied notion of modulus maxima in non-causal wavelet representations (Mallat and Hwang 1992).

In this way, we will demonstrate that significant components in more traditional non-causal wavelet theory can be carried over to a time-causal temporal domain, which opens up for a variety of application areas, where either (i) temporal signals are to be processed in real time, such as for purposes of real-time monitoring, time series analysis and prediction, including learning-based systems as detailed further in Section 8, or as components in control-loop systems, that are to react to a continuous inflow of information, or (ii) analysis of biological or physical signals that needs to be performed in a completely time-causal manner, to be biologically or physically realistic, such as in mathematically based models of biological or physical processes.

In this paper, we develop the underlying fundamentals of the proposed class of time-causal wavelets, with special emphasis on their theoretical properties, combined with characterizations of scaling properties and proofs-of-concepts for different types of temporal signals.

Beyond a very few early approaches, this is the first systematic treatment of time-causal wavelet theory, with structurally close relations to non-causal wavelets based on derivatives of the Gaussian kernel. Specifically, we propose that the time-causal wavelet theory to be presented here constitutes a canonical way to define wavelet representations over a time-causal temporal domain, in a corresponding way as non-causal wavelets based on derivatives of the Gaussian kernel, such as the Ricker or Mexican hat wavelets, can be regarded as canonical wavelet representations over a non-causal temporal domain, according to the theory for variation-diminishing convolution transformations to be described in Section 2.3.

## 1.1 Structure of the presentation

The presentation is organized as follows: Section 2 begins by showing how continuous time-causal wavelets can be constructed based on temporal derivatives of a special time-causal smoothing kernel, referred to as the time-causal limit kernel, with the theoretical results summarized into continuous time-causal wavelet representations in Section 3.

Section 4 then shows how this continuous kernel can be canonically discretized in terms of a set of first-order recursive filters coupled in cascade, complemented with time-causal difference operators, to compute discrete approximations of temporal derivatives of the time-causal limit kernel, thereby leading to a family of discrete time-causal wavelet representations as summarized in Section 5.

Section 6 additionally provides a set of characterizations of scaling properties of the resulting discrete wavelet representations, quantifying how well the ideal scaling properties of the continuous time-causal wavelets carry over to a discrete implementation, specifically as depending on the distribution parameter  $c$  of the time-causal limit kernel, which specifies how densely the temporal scale levels are quantized, and thereby leading to trade-off issues regarding the accuracy of computed features *vs.* the inherent temporal delays, that by necessity will occur due to the time-causal processing of the input signals.

Section 7 complements with examples of results of computing the proposed time-causal wavelet representations for different types of temporal signals. Finally, Section 8 concludes with a summary and discussion, including relations to related multi-scale signal processing and learning approaches.

## 2 Continuous wavelet representations based on temporal derivatives of the time-causal limit kernel

In this section, we will after (i) a first review of main concepts regarding regular, not necessarily time-causal, wavelet theory, describe (ii) the time-causal limit kernel and how this kernel is constrained with respect its shape based on (iii) a very general theoretical result regarding provably variation-diminishing convolution transformations and the desirable property of basing the time-causal wavelet analysis on a (iv) provably scale-covariant convolution kernel.

Then, we will describe (v) how mother wavelets can be defined from temporal derivatives of the time-causal limit kernel, and how such mother wavelets can be used for defining a continuous wavelet transform based on a discrete temporal scale parameter. We also describe (vi) how the normalization of this kernel over scales relates to the notion of scale-normalized derivatives in scale-space theory, including the special role of  $L_1$ -normalization across scales and scale selection properties for local characteristic model signals, as well as (vii) how that scale normalization relates to the normalization over scales in more traditional wavelet representations.

Based on these foundations, we will then describe (viii) how the first- and second-order temporal derivatives of the time-causal limit kernel can be regarded as a time-causal analogues of the Ricker wavelet, also referred to as the Mexican hat wavelet, and corresponding to the second-order derivative of the Gaussian kernel, as well as also analogues of the first-order derivatives of the non-causal Gaussian kernel.

The intention behind this treatment is to lay out the theoretical foundations for using the proposed time-causal wavelets as computational primitives in more composed signal processing and learning approaches. Specifically, since the time-causal limit kernel, that we will use as a fundamental basis for defining the time-causal wavelet representations in terms of derivatives thereof, does not have any compact closed-form expression, we will in this section devote special emphasis on describing its theoretical properties, so as to be able to infer properties of the derived time-causal wavelet representations, and in this way mediate a better theoretical understanding.

An effectively more compact description of some of the main concepts and results will then be given in Section 3.

### 2.1 Continuous wavelet representation

#### 2.1.1 Mother wavelet

The notion of a continuous wavelet representation starts from a mother wavelet  $\chi(t)$ , which is assumed to satisfy the admissibility criterion (Mallat 1999 Equation (4.36))

$$C_\chi = \int_{\omega=0}^{\infty} \frac{|\hat{\chi}(\omega)|^2}{\omega} d\omega < \infty \quad (1)$$

where

$$\hat{\chi}(\omega) = \int_{t \in \mathbb{R}} \chi(t) e^{-it\omega} dt \quad (2)$$

denotes the Fourier transform of the mother wavelet  $\chi(t)$ . Specifically, the admissibility criterion (1) implies that the integral of the mother wavelet should be zero<sup>1</sup>

$$\int_{t \in \mathbb{R}} \chi(t) dt = 0. \quad (3)$$

Furthermore, if the mother wavelet decays sufficiently fast to zero when  $|t| \rightarrow \infty$

$$\int_{t \in \mathbb{R}} (1 + |t|) |\chi(t)| dt < \infty, \quad (4)$$

then it holds that  $C_\chi < \infty$  (Mallat 1999 page 75).

<sup>1</sup> This can be seen from the fact that the admissibility criterion (1) implies that by necessity  $\hat{\chi}(0) = 0$ , which is also the integral of the mother wavelet.

### 2.1.2 Wavelet transform

Given that the mother wavelet satisfies the admissibility criterion and is normalized to unit  $L_2$ -norm

$$\int_{t \in \mathbb{R}} |\chi(t)|^2 dt = 1, \quad (5)$$

the wavelet representation  $\mathcal{W}_\chi f$  of any function  $f \in L_2(\mathbb{R})$  can then be defined as (Heil and Walnut 1989 page 629)

$$(\mathcal{W}_\chi f)(u, v) = \int_{t \in \mathbb{R}} f(t) e^{-u/2} \overline{\chi(e^{-u} t - v)} dt, \quad (6)$$

from which the original function can be recovered as (Heil and Walnut 1989 page 629)

$$f(t) = \int_{u \in \mathbb{R}} \int_{v \in \mathbb{R}} (\mathcal{W}_\chi f)(u, v) e^{-u/2} \chi(e^{-u} t - v) du dv. \quad (7)$$

Basically, in this representation, the variable  $u$  constitutes a logarithmic parameterization of the scaling factor  $a = e^u$  in a scaling transformation of the mother wavelet according to

$$\chi_a(t) = \frac{1}{a^b} \chi\left(\frac{t}{a}\right) \quad (8)$$

for some  $a \in \mathbb{R}_+$  and some  $b \in \mathbb{R}$ .

The normalization of the mother wavelet  $\chi(t)$  in the forward wavelet transform in (6) and the inverse wavelet transform (7) corresponds to setting  $b = 1/2$  in (8). As we will see later, other forms of normalizations of filters across scales are commonly used in the related area of scale-space theory, notably  $b = 1$  to obtain maximum scale invariance, when computing wavelet representations in terms of temporal derivatives of the time-causal limit kernel over multiple temporal scales, as we will consider in Section 2.8.

### 2.2 The time-causal limit kernel

While the above theory applies to very general classes of mother wavelets, we will in this treatment focus solely on mother wavelets that are obtained from temporal derivatives of a special type of time-causal kernel, denoted the time-causal limit kernel (Lindeberg 2016 Section 5; Lindeberg 2023 Section 3)

$$h(t; \tau) = \Psi(t; \tau, c), \quad (9)$$

characterized by having a Fourier transform of the form

$$\hat{\Psi}(\omega; \tau, c) = \prod_{k=1}^{\infty} \frac{1}{1 + i c^{-k} \sqrt{c^2 - 1} \sqrt{\tau} \omega}. \quad (10)$$

This time-causal kernel has been derived from criteria concerning the notion of temporal scale space representation, where any continuous signal  $f: \mathbb{R} \rightarrow \mathbb{R}$  is to be represented

by a temporal scale-space representation  $L: \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$  (see Lindeberg 2023 Equation (26))

$$L(t; \tau, c) = \int_{\xi \in \mathbb{R}} \Psi(\xi; \tau, c) f(t - \xi) d\xi \quad (11)$$

over multiple temporal scale levels  $\tau \in \mathbb{R}_+$  in such a way that no new local extrema or zero-crossings may be created from finer to coarser levels of temporal scales. The parameter  $c$  is specifically a distribution parameter with  $c > 1$ , that determines the discrete scale levels in units of the variance of the time-causal limit kernel according to<sup>2</sup> (Lindeberg 2016 Equation (18))

$$\tau_k = \tau_0 c^{2k}, \quad (12)$$

where  $\tau_0 > 0$  represents a reference scale, and where for common purposes of implementation one often chooses  $\tau_0 = 1$  and either  $c = \sqrt{2}$  or  $c = 2$ . Alternatively, the time-causal limit kernel can also be parameterized in terms of the standard deviation  $\sigma = \sqrt{\tau}$  of the time-causal limit kernel, which is of dimension [time].

Figure 1 shows graphs of the time-causal limit kernel for a few values of the temporal scale parameter  $\sigma = \sqrt{\tau}$ , together with its first- and second-order temporal derivatives.

### 2.3 Characterization of continuous smoothing kernels

A theorem by Schoenberg (1950) characterizes the families of continuous smoothing kernels that are variation-diminishing, in the sense that the number of zero-crossings in a smoothed signal must never exceed the number of zero-crossings in the original signal, to having a bilateral Laplace-Stieltjes transform of the form

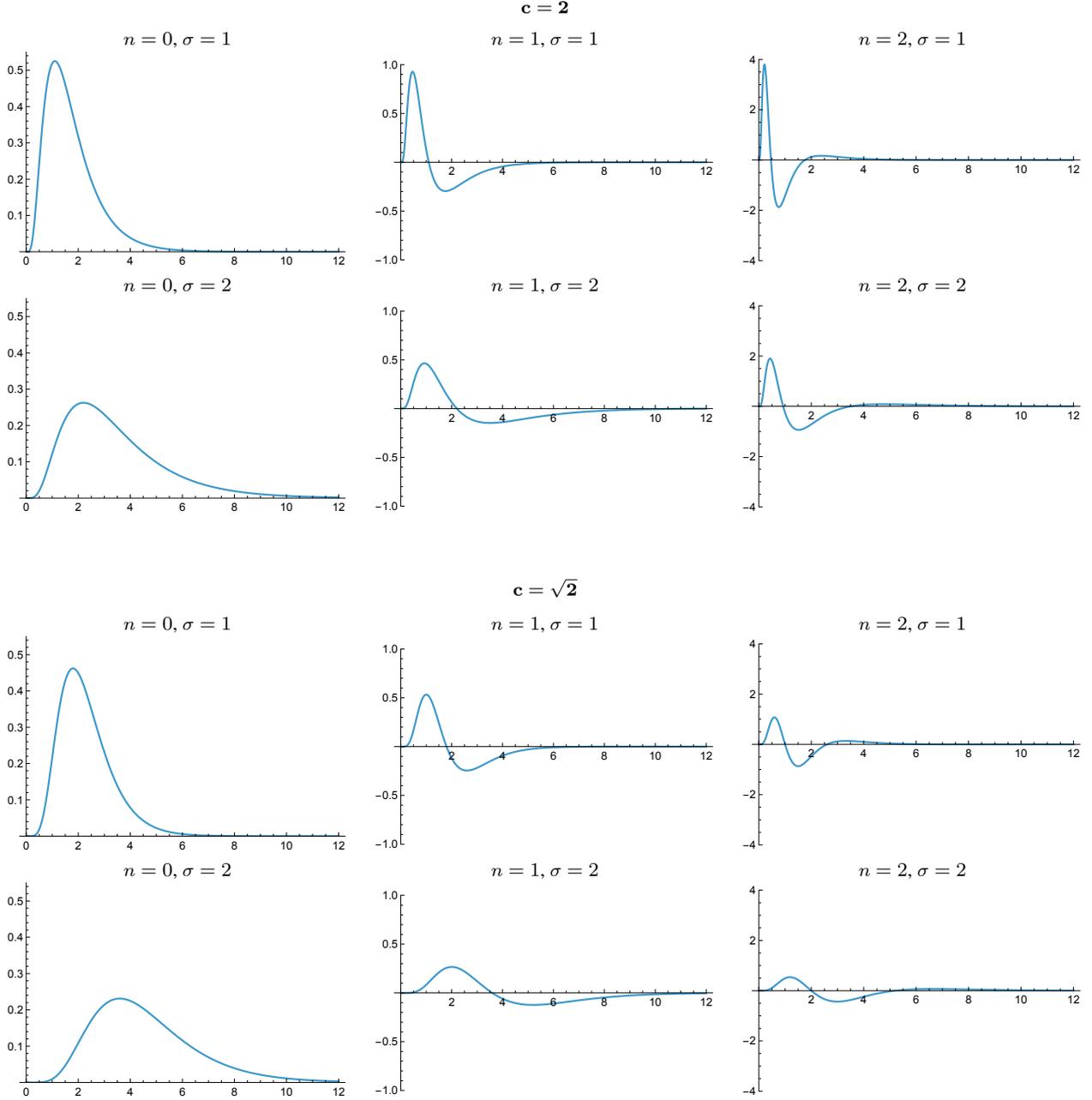
$$\int_{\xi=-\infty}^{\infty} e^{-s\xi} h(\xi) d\xi = C e^{\gamma s^2 + \delta s} \prod_{i=1}^{\infty} \frac{e^{a_i s}}{1 + a_i s} \quad (13)$$

for  $-c < \text{Re}(s) < c$  and some  $c > 0$ , where  $C \neq 0$ ,  $\gamma \geq 0$ ,  $\delta$  and  $a_i$  are real and  $\sum_{i=1}^{\infty} a_i^2$  is convergent.

From this result,<sup>3</sup> it follows that any continuous 1-D smoothing kernel can, beyond trivial rescaling operations and shifts, be decomposed into the following classes of primitive smoothing transformations:

<sup>2</sup> This logarithmic spacing of the temporal scale levels is similar to the spacing generated by a uniform spacing with respect to the canonical logarithmic parameterization of the scale parameter into an effective scale parameter  $\tau_{\text{eff}}$  according to  $\tau_{\text{eff}} = A + B \log \tau$ , as can be derived from axiomatic arguments for a regular spatial scale-space representation (Lindeberg 1993). Compared to a uniform spacing of the scale levels relative to the scale parameter measured in units of the variance  $\tau$  of the temporal smoothing kernel, the logarithmic distribution according to (12) also leads to significantly shorter delays, as further described in Lindeberg (2016) Section 4 and Appendix 1.

<sup>3</sup> In Equation (13), (i) the factor  $e^{\gamma s^2}$  corresponds to the Laplace-Stieltjes transform of the Gaussian kernel, (ii) the factor  $1/(1 + a_i s)$  corresponds to the Laplace-Stieltjes transform of a truncated exponential kernel, while (iii) the factors  $e^{\delta s}$  and  $e^{a_i s}$  correspond to the Laplace-Stieltjes transforms of shift operations. The overall product form of this expression does furthermore correspond to convolutions between these primitive smoothing transformations.



**Fig. 1** The time-causal limit kernel  $\Psi(t; \tau, c)$  according to (18), truncated after the 8 truncated exponential kernels having the longest time constants, with its scale-normalized temporal derivatives  $\Psi_{\zeta^n}(t; \tau, c) = \tau^{n\gamma/2} \Psi_{t^n}(t; \tau, c)$  up to order  $n = 2$ , with the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization across scales, for different combinations of the temporal scales  $\sigma = \sqrt{\tau} \in \{1, 2\}$  and different values to the distribution parameter  $c \in \{\sqrt{2}, 2\}$ . (Horizontal axes: time  $t \in [0, 12]$ . Vertical axes: kernel values with different ranges  $[0, 0.50]$ ,  $[-1.0, 1.0]$  or  $[-4.0, 4.0]$ , depending on the order  $n$  of temporal differentiation.)

– convolution with *Gaussian kernels*

$$h(\xi) = e^{-\gamma\xi^2}, \quad (14)$$

– convolution with *truncated exponential functions*

$$h(\xi) = \begin{cases} e^{-|\lambda|\xi} & \xi \geq 0, \\ 0 & \xi < 0, \end{cases} \quad h(\xi) = \begin{cases} e^{|\lambda|\xi} & \xi \leq 0, \\ 0 & \xi > 0, \end{cases} \quad (15)$$

for some strictly positive  $|\lambda|$ .

With respect to smoothing of temporal signals over multiple temporal scales, this result has the following important implications:

- Over a symmetric non-causal temporal domain, where the future is assumed to be accessible, as it can be for pre-recorded temporal signals, this result singles out the Gaussian kernel as the canonical semi-group of temporal smoothing kernels (Karlin 1968 Theorem 5.2, Lindeberg

1990 Theorem 5) that obey

$$h(\cdot; \tau_1) * h(\cdot; \tau_2) = h(\cdot; \tau_1 + \tau_2) \quad (16)$$

where  $\tau_1 \geq 0$  and  $\tau_2 \geq 0$  are the temporal scale parameters.

- Over a time-causal temporal domain, where the future cannot be accessed, the only possible temporal smoothing kernels are truncated exponential kernels

$$h_{\text{exp}}(t; \mu_k) = \begin{cases} \frac{1}{\mu_k} e^{-t/\mu_k} & t \geq 0, \\ 0 & t < 0, \end{cases} \quad (17)$$

coupled in cascade (Lindeberg and Fagerström 1996, Lindeberg 2023 Section 2.2).

In this way, we can obtain a complete characterization of what idealized temporal smoothing kernels to use when to smooth one-dimensional signals to coarser levels of temporal scales in the cases of either a time-causal or a non-causal temporal domain.

#### 2.4 Scale covariance property of the time-causal limit kernel

The definition of the time-causal limit kernel according to (9), originally performed in Lindeberg (2016) Section 5 and then refined in Lindeberg (2023) Section 3, starts from the above characterization of time-causal temporal smoothing kernels as only possible to express in terms of truncated exponential kernels coupled in cascade, to choose the time constants  $\mu_k$  in an infinite cascade of convolution operations

$$\Psi(\cdot; \tau, c) = *_{k=1}^{\infty} h_{\text{exp}}(\cdot; \mu_k) \quad (18)$$

based on intermediate temporal scale levels  $\tau_k$  according to (12) in such a way that the choice of the temporal scale levels according to

$$\mu_k = c^{-k} \sqrt{c^2 - 1} \sqrt{\tau} \quad (19)$$

implies that the resulting time-causal limit kernel obeys temporal scale covariance.

Temporal scale covariance here means that if we for any temporal scaling factor  $S_t$ , that is an integer power of the distribution parameter  $c > 1$  according to

$$S_t = c^j \quad \text{for} \quad j \in \mathbb{Z}, \quad (20)$$

consider a scaling transformation of the original signal  $f$  into a rescaled signal  $f'$  defined by

$$f'(t') = f(t) \quad \text{for} \quad t' = S_t t, \quad (21)$$

then the resulting temporal scale-space representations

$$L(\cdot; \tau, c) = \Psi(\cdot; \tau, c) * f(\cdot) \quad (22)$$

$$L'(\cdot; \tau', c) = \Psi(\cdot; \tau', c) * f'(\cdot) \quad (23)$$

are equal (Lindeberg 2016 Equation (47))

$$L'(t'; \tau', c) = L(t; \tau, c) \quad (24)$$

for matching values of the temporal moments  $t' = S_t t$  and the temporal scale parameters

$$\tau' = S_t^2 \tau. \quad (25)$$

This result thus means that the time-causal limit kernel is self-similar under temporal scaling transformations, which also, with appropriate complementary temporal scale normalization, carries over to temporal derivatives of the time-causal limit kernel according to

$$L_{t^n}(t; \tau) = \partial_t^n L(t; \tau), \quad (26)$$

as will be later described in Section 2.7.1.

Note, however, that since the temporal scale levels are inherently discrete, as implied by the construction of the temporal scale-space representation from multiple truncated exponential kernels coupled in cascade, the temporal scale covariance property holds exactly only when the temporal scaling factor  $S_t$  is an integer power of the distribution parameter  $c$  according to  $S_t = c^j$  for  $j \in \mathbb{Z}$ . For other values of  $S_t$ , the corresponding result will only be approximate.

#### 2.5 Properties of the time-causal limit kernel

##### 2.5.1 Characterization of the temporal duration

The variance of the time-causal limit kernel, which characterizes the temporal duration, is given by (Lindeberg 2016 Equation (35))

$$V(\Psi(\cdot; \tau, c)) = \sum_{k=1}^{\infty} \mu_k^2 = \tau. \quad (27)$$

##### 2.5.2 Characterizations of the temporal delay

The temporal mean of the time-causal limit kernel, which constitutes one way of characterizing the temporal delay, is (Lindeberg 2016 Equation (34))

$$M(\Psi(\cdot; \tau, c)) = \sum_{k=1}^{\infty} \mu_k = \sqrt{\frac{c+1}{c-1}} \sqrt{\tau}. \quad (28)$$

An alternative characterization of the temporal delay is given by the position of the temporal maximum of the time-causal limit kernel. Unfortunately, it is hard to calculate the derivatives of the time-causal limit kernel in compact closed form. The maximum point can, however, be estimated from an approximation of the time-causal limit kernel in terms of Koenderink's (1988) scale-time model according to (Lindeberg 2023 Equation (39))

$$t_{\text{max}} \approx \frac{(c+1)^2 \sqrt{\tau}}{2\sqrt{2}\sqrt{(c-1)c^3}} = \delta. \quad (29)$$

### 2.5.3 Trade-off issue regarding the choice of the distribution parameter $c$

From these estimates of the temporal delay, in combination with the result (24) concerning exact scale covariance properties only for temporal scaling factors  $S_t$  that are integer powers of the distribution parameter  $c$ , we have a trade-off issue in the respect that:

- A lower value of the distribution parameter  $c$  will imply a denser set of temporal scale levels with numerically better ability to approximate temporal scale covariance for temporal scaling factors  $S_t$  that are not integer powers of the distribution parameter  $c$ .
- A larger value of the distribution parameter  $c$  will lead to shorter temporal delays, as estimated according to (28) and (29), with better ability to respond fast in real-time situations.

### 2.5.4 Differential equation formulation of the time-causal temporal scale-space representation

In terms of differential equations,<sup>4</sup> the relationships between adjacent levels of temporal scales are given by (Lindeberg 2023 Equation (11))

$$\partial_t L(t; \tau_k, c) = -\frac{1}{\mu_k} (L(t; \tau_k, c) - L(t; \tau_{k-1}, c)) \quad (30)$$

with the initial condition  $L(t; 0, c) = f(t)$ .

An important property of this type of temporal scale-space representation is that it is *time-recursive*. The temporal scale-space representations  $L(t; \tau_k, c)$  over the temporal scale channels with indices  $k$  constitute a *sufficient temporal memory of the past*, to compute the temporal scale-space representation and the next temporal moment, given a new input in the input signal  $f(t)$ .

The form of the first-order integrators (30) also explicitly implies that local perturbations in the signal are gradually smoothed out from finer to coarser levels of temporal

<sup>4</sup> The fact that the time-causal scale-space representation  $L(t; \tau, c)$  according to (11) satisfies the recurrence relation (30) between adjacent levels of temporal scales can be formally verified as follows: Let us first rewrite the recurrence relation between the input signal  $f_{in}(t)$  and the output signal  $f_{out}(t)$  between adjacent levels of temporal scales  $\tau_{k-1}$  and  $\tau_k$  as  $\partial_t f_{out}(t) = -1/\mu_k \times (f_{out}(t) - f_{in}(t))$ . Taking the Laplace transform of this relation then gives  $s \mathcal{L}(f_{out})(s) = -1/\mu_k \times (\mathcal{L}(f_{out})(s) - \mathcal{L}(f_{in})(s))$ . Let us next assume that the Laplace transforms  $\mathcal{L}(f_{out})(s)$  and  $\mathcal{L}(f_{in})(s)$  of  $f_{out}(t)$  and  $f_{in}(t)$ , respectively, are related by a transfer function  $H(s)$  according to  $\mathcal{L}(f_{out})(s) = H(s) \mathcal{L}(f_{in})(s)$ . Inserting this expression into the Laplace transform of the recurrence relation, then gives  $s H(s) \mathcal{L}(f_{in})(s) = -1/\mu_k \times (H(s) \mathcal{L}(f_{in})(s) - \mathcal{L}(f_{in})(s))$ , from which we can solve for the transfer function as  $H(s) = 1/(1 + \mu_k s)$ . From tables of the Laplace transforms of standard functions, we have that the Laplace transform of the truncated exponential kernel  $e^{-\alpha t}$  for  $t > 0$ , 0 elsewhere, is  $1/(s + \alpha)$ . From this, we can conclude that the transfer function  $H(s)$  corresponds to the Laplace transform of the truncated exponential kernel  $h_{exp}(t; \mu_k) = 1/\mu_k \times e^{-t/\mu_k}$  for  $t > 0$ , 0 elsewhere, which proves the result.

scales, because of the negative feedback with respect to the differences  $L(t; \tau_k, c) - L(t; \tau_{k-1}, c)$  between the adjacent temporal scale channels.

### 2.5.5 Cascade smoothing property of the time-causal limit kernel

Due to the definition of the time-causal limit kernel according to (18), it follows that it obeys the following cascade smoothing property between adjacent levels of temporal scales  $\tau$  and  $\tau/c^2$  (Lindeberg 2016 Equation (41)):

$$\Psi(\cdot; \tau, c) = h_{exp}(\cdot; \frac{\sqrt{c^2-1}}{c} \sqrt{\tau}) * \Psi(\cdot; \frac{\tau}{c^2}, c). \quad (31)$$

This property means that, when computing the temporal scale-space representation or a temporal derivative of the temporal scale-space representation, it is not necessary to redo all the computations when to compute those representations over multiple temporal scales. Instead, it is sufficient to perform the major amount of temporal smoothing once and for all, and then compute the representations at the coarser levels of temporal scales, by applying a usually much smaller number of truncated exponential kernels to the corresponding representations at finer temporal scales.

Between adjacent temporal scale levels of the smoothed temporal scale-space representations, we therefore have

$$L(\cdot; \tau, c) = h_{exp}(\cdot; \frac{\sqrt{c^2-1}}{c} \sqrt{\tau}) * L(\cdot; \frac{\tau}{c^2}, c), \quad (32)$$

and regarding temporal derivatives

$$L_{t^n}(\cdot; \tau, c) = h_{exp}(\cdot; \frac{\sqrt{c^2-1}}{c} \sqrt{\tau}) * L_{t^n}(\cdot; \frac{\tau}{c^2}, c). \quad (33)$$

Since each truncated exponential kernel is a simplifying kernel, that guarantees non-creation of new local extrema or zero-crossings, this result means that also the temporal derivatives obey a simplifying property from finer to coarser levels of temporal scales.

### 2.5.6 Truncating the infinite convolution kernel for implementation purposes

For purposes of numerical or physical implementation, the time-causal limit kernel can, because of the rapid convergence property of the temporal scale levels, as corresponding to the convergence properties of a geometric series, be truncated for a finite number  $K$  of layers of exponential kernels coupled in cascade.

For this purpose, we can choose the time-constants of the  $K$  selected truncated exponential kernels coupled in cascade

according to (Lindeberg 2016 Equations (19)-(20))<sup>5</sup>

$$\mu_1 = c^{1-K} \sqrt{\tau}, \quad (34)$$

$$\mu_k = \sqrt{\tau_k - \tau_{k-1}} = c^{k-K-1} \sqrt{c^2 - 1} \sqrt{\tau} \quad (2 \leq k \leq K), \quad (35)$$

where the time constant  $\mu_1$  of the first truncated exponential kernel has been determined in such a way that the temporal variance of the resulting composed kernel is exactly  $\tau$ :

$$V(h(\cdot; \mu_1, \mu_2, \dots, \mu_K)) = \sum_{k=1}^K \mu_k^2 = \tau. \quad (36)$$

In this way, the first truncated exponential kernel approximates the temporal smoothing effect of all the truncated exponential kernels that have been truncated away, as quantified in terms of the temporal variances of the involved temporal smoothing kernels.

### 2.5.7 Fulfilment of the admissibility criterion for a mother wavelet

Based on a partial fraction expansion of the Laplace transform of the time-causal limit kernel defined according to (10):

$$H_{\Psi}(q; \tau, c) = \prod_{k=1}^{\infty} \frac{1}{1 + \mu_k q} = \sum_{k=1}^{\infty} \frac{A_k}{1 + \mu_k q}, \quad (37)$$

with the time constants  $\mu_k$  as functions of  $\tau$  and  $c$  according to (35), we can determine the coefficients  $A_k$  according to (Lindeberg 2023 Equation (73))

$$A_k = \prod_{i=1, i \neq k}^{\infty} \frac{1}{1 - \frac{\mu_i}{\mu_k}}, \quad (38)$$

implying that the  $n$ :th order temporal derivative of the time-causal limit kernel will have the following series representation (Lindeberg 2023 Equation (75)):

$$(\partial_{t^n} \Psi)(t; \tau, c) = \sum_{k=1}^{\infty} \left( \frac{-1}{\mu_k} \right)^n \frac{A_k}{\mu_k} e^{-t/\mu_k} \quad (t \geq 0). \quad (39)$$

<sup>5</sup> To derive these relations, we have made use of the facts that (i) for non-negative convolution kernels, the variances are additive under the convolution operation, and (ii) the variance of truncated exponential kernel  $h_{\text{exp}}(t; \mu_k)$  with time constant  $\mu_k$  is  $V(h_{\text{exp}}(\cdot; \mu_k)) = \mu_k^2$ . These properties lead to the relationships in Equation (35), given the logarithmic sampling of the temporal scale levels  $\tau_k = c^{2k}$  according to (12). Concerning Equation (34), we have on the other hand determined the first time constant  $\mu_1$  in such a way that the variance of the first truncated exponential kernel  $V(h_{\text{exp}}(\cdot; \mu_1)) = \mu_1^2$  is equal to the sum of the variances of all the truncated exponential kernels in the infinite convolution operation (18) that have been truncated away. In this way, the variance of the resulting composed convolution kernel is equal to the same value  $\tau$  as would be obtained with the original time-causal limit kernel  $\Psi(t; \tau, c)$  with temporal scale parameter  $\tau$ .

Due to the exponential decrease of this expression as function of  $t$ , it thereby follows that (4)

$$\int_{t \in \mathbb{R}} (1 + |t|) |\Psi(t; \tau, c)| dt < \infty, \quad (40)$$

which formally verifies that the time-causal limit kernel fulfills the admissibility condition (1) for being a mother wavelet.

### 2.6 Time-causal and time-recursive mother wavelets and wavelet representations

Our proposal is that the temporal derivatives of the time-causal limit kernel can with appropriate complementary normalization be used as mother wavelets

$$\chi(t; c) = \frac{\partial_{t^n} \Psi(t; 1, c)}{\|\partial_{t^n} \Psi(t; 1, c)\|_p} \quad (41)$$

for computing time-causal and time-recursive wavelet representations according to (6):

$$(\mathcal{W}_{\chi} f)(u, v; c) = \int_{t \in \mathbb{R}} f(t) e^{-u/2} \chi(e^{-u} t - v; c) dt. \quad (42)$$

A conceptual difference between this formulation of a time-causal wavelet transform and the more traditional formulation (6), however, is that because of the discrete nature of the temporal scale levels  $\tau_k = c^{2k}$  according to (12), the logarithmic variable  $u$  in the time-causal wavelet transform (42) will only assume discrete values  $u_j$  according to

$$e^{u_j} = c^j, \quad (43)$$

that is of the form

$$u_j = j \log c \quad \text{for } j \in \mathbb{Z}. \quad (44)$$

According to traditional wavelet theory, one usually makes use of  $L_2$ -norms when normalizing the mother wavelet. For purposes of computing derivatives over multiple scales in scale-space theory, however, other choices of  $L_p$ -norms are more common for relating image structures over scales (Lindeberg 1998a Table 3), as will be detailed further below.

### 2.7 Relations to scale-normalized derivatives in scale-space theory

For purposes in scale-space theory, the notion of scale-normalized temporal derivatives can be introduced according to (Lindeberg 1998a Equation (6); Lindeberg 2017 Equation (6))

$$\partial_{\zeta} = \tau^{\gamma/2} \partial_t, \quad (45)$$

where  $\gamma \in \mathbb{R}_+$  is a scale normalization parameter.

### 2.7.1 Scale covariance property of scale-normalized temporal derivatives

As shown in Lindeberg (2016) Appendix 3, under a temporal scaling transformation of the form

$$f'(t') = f(t) \quad \text{for} \quad t' = c^{j'-j} t, \quad (46)$$

with the corresponding transformation between corresponding temporal scale levels of the form

$$\tau' = c^{2(j'-j)} \tau, \quad (47)$$

the scale-normalized temporal derivatives defined from temporal differentiation of convolution with the time-causal limit kernel  $\Psi(t; \tau, c)$  do for any temporal input signal relate according to

$$L'_{\zeta'n}(t'; \tau', c) = c^{(j'-j)n(\gamma-1)} L_{\zeta'n}(t; \tau, c). \quad (48)$$

Specifically, local temporal scale estimates  $\hat{\tau}$  and  $\hat{\tau}'$ , as determined from local extrema over temporal scales in the two temporal domains, will be assumed at corresponding temporal scale levels and will thus be transformed in a scale-covariant way for any temporal scaling transformation of the form  $t' = c^{j'-j} t$ , according to

$$\hat{\tau}' = c^{2(j'-j)} \hat{\tau} \quad (49)$$

Furthermore, in the special case when  $\gamma = 1$ , the scale-normalized derivatives are equal

$$L'_{\zeta'n}(t'; \tau', c) = L_{\zeta'n}(t; \tau, c). \quad (50)$$

This means that we can readily compare the magnitudes of scale-normalized derivatives between matching temporal moments  $\{t \leftrightarrow t'\}$  and temporal scales  $\{\tau \leftrightarrow \tau'\}$ .

### 2.7.2 Scale selection properties for characteristic temporal signals

By studying the evolution properties of such temporal scale-normalized derivatives over scales for the *non-causal* temporal scale-space representation defined from convolution with Gaussian kernels according to

$$L(\cdot; \tau, c) = g(\cdot; \tau) * f(\cdot) \quad (51)$$

where the non-causal temporal Gaussian kernel is given by

$$g(t; \tau) = \frac{1}{\sqrt{2\pi}\tau} e^{-t^2/2\tau}, \quad (52)$$

it specifically follows that:

- For a 1-D Gaussian blob

$$f(t) = g(t; \tau_0) \quad (53)$$

the strongest magnitude response of the scale-normalized second-order derivative operator

$$L_{\zeta\zeta}(t; \tau) = \tau L_{tt}(t; \tau) \quad (54)$$

is assumed for the time moment  $t = 0$  and the temporal scale level (Lindeberg 2017 Equation (13))

$$\hat{\tau} = \frac{2\gamma}{3-2\gamma} \tau_0 \quad (55)$$

proportional to the temporal scale  $\tau_0$  in the input signal. For the specific choice of

$$\gamma = \gamma_2 = \frac{3}{4} \quad (56)$$

we have a perfect match to the characteristic temporal scale in the input signal according to

$$\hat{\tau} = \tau_0. \quad (57)$$

- For a 1-D Gaussian edge

$$f(t) = \int_{\xi=-\infty}^t g(\xi; \tau_0) d\xi \quad (58)$$

the strongest maximum response of the scale-normalized first-order derivative operator

$$L_{\zeta}(t; \tau) = \tau^{1/2} L_t(t; \tau) \quad (59)$$

is assumed for the time moment  $t = 0$  and the temporal scale level (Lindeberg 2017 Equation (24))

$$\hat{\tau} = \frac{\gamma}{1-\gamma} \tau_0 \quad (60)$$

proportional to the temporal scale  $\tau_0$  in the input signal. For the specific choice of

$$\gamma = \gamma_1 = \frac{1}{2} \quad (61)$$

we have a perfect match to the characteristic temporal scale in the input signal according to

$$\hat{\tau} = \tau_0. \quad (62)$$

In this way, the scale selection mechanism based on local extrema over temporal scales leads to scale estimates  $\hat{\tau}$  that reflect the inherent temporal scales  $\tau$  in the input signal. Notably, these results bear close relationships to the matched filter theorem (Woodward 1953, Turin 1960), in that the scale selection mechanism will choose filters for detecting the different types of temporal structures in the input signal, that as well as possible match their size.

### 2.7.3 Interpretations in terms of normalization over scale with respect to $L_p$ -norms of Gaussian derivative kernels

In relation to normalization of the filters with respect to  $L_p$ -norms, it can specifically be shown (Lindeberg 1998a Section 9.1) that the notion of  $\gamma$ -normalized derivatives corresponds to normalizing the  $n$ :th order scale-normalized Gaussian derivative kernels  $g_{\zeta^n}(t; \tau)$  over a one-dimensional domain to constant  $L_p$ -norms over scale

$$\|g_{\zeta^n}(\cdot; \tau)\|_p = \left( \int_{t \in \mathbb{R}} |g_{\zeta^n}(t; \tau)|^p dx \right)^{1/p} = G_{n,\gamma} \quad (63)$$

with

$$p = \frac{1}{1 + n(1 - \gamma)}, \quad (64)$$

where the perfectly scale invariant case

$$\gamma = 1 \quad (65)$$

corresponds to  $L_1$ -normalization for all orders  $n$  of temporal differentiation. Such default  $L_1$ -normalization corresponds to scale-normalized temporal derivatives that are dimensionless in units of [time] and will therefore treat temporal structures at all temporal scales in a similar manner.

Compare also with the transformation property of scale-normalized temporal derivatives in the scale covariance property (48), where the choice of  $\gamma = 1$  corresponding to  $p = 1$  makes the scale-normalized temporal derivatives equal at corresponding temporal moments  $\{t \leftrightarrow t'\}$  and temporal scales  $\{\tau \leftrightarrow \tau'\}$ .

With regard to the above detection of the scale of a 1-D Gaussian blob from extrema over scale in the magnitude of the scale-normalized second-order derivative using the scale normalization power  $\gamma_2 = 3/4$ , or the detection of the scale of a 1-D Gaussian edge from extrema over scale in the magnitude of the scale-normalized first-order derivative using the scale normalization power  $\gamma_1 = 1/2$ , it can be noted that both of these scale normalization methods correspond to normalizing the corresponding Gaussian derivative kernels to constant  $L_p$ -norms over scale for the choice of  $p = 2/3$ .

### 2.8 Normalization over scales in wavelet representations

For a rescaled mother wavelet of the form (8)

$$\chi_a(t) = \frac{1}{a^b} \chi\left(\frac{t}{a}\right), \quad (66)$$

where  $a \in \mathbb{R}_+$  and  $b \in \mathbb{R}$ , its  $L_p$ -norm is given by

$$\|\chi_a(\cdot)\|_p = \frac{a^{1/p}}{a^b} \|\chi(\cdot)\|_p, \quad (67)$$

implying that the  $L_p$ -norm of the rescaled mother wavelet (66) is constant over scales if and only if

$$b = \frac{1}{p}. \quad (68)$$

In regular wavelet representations, the choice  $p = 2$  corresponding to  $b = 1/2$  constitutes the default choice. In scale-space theory, the theoretical results described in Sections 2.7.1 and 2.7.3 imply that  $p = 1$  corresponding to  $b = 1$  does, on the other hand, constitutes the default choice.

2.9 The first- and second-order derivatives of the time-causal limit kernel as time-causal analogues of the Ricker wavelet / the Mexican hat wavelet and the wavelet based on the first-order derivative of the Gaussian kernel

In regular wavelet theory, the Ricker wavelet, also known as the Mexican hat wavelet, and corresponding to the second-order derivative of the Gaussian kernel

$$g_{tt}(t; \tau) = \frac{(t^2 - \tau)}{\tau^2} g(t; \tau) = \frac{(t^2 - \tau)}{\tau^2} \frac{1}{\sqrt{2\pi\tau}} e^{-t^2/2\tau} \quad (69)$$

constitutes a natural choice for defining a mother wavelet. With normalization to unit  $L_2$ -norm over scales for

$$\|g_{tt}(\cdot; \tau)\|_2 = \frac{\sqrt{\frac{3}{2}}}{2\sqrt[4]{\pi} \tau^{5/4}}, \quad (70)$$

we obtain the following mother wavelet

$$\chi_{g_{tt}, L_2}(t; \tau) = \frac{g_{tt}(t; \tau)}{\|g_{tt}(\cdot; \tau)\|_2} = \frac{2(t^2 - \tau) e^{-\frac{t^2}{2\tau}}}{\sqrt{3}\sqrt[4]{\pi} \tau^{5/4}}, \quad (71)$$

and with normalization to unit  $L_1$ -norm over scales for

$$\|g_{tt}(\cdot; \tau)\|_1 = \sqrt{\frac{8}{e\pi}} \frac{1}{\tau}, \quad (72)$$

we obtain

$$\chi_{g_{tt}, L_1}(t; \tau) = \frac{g_{tt}(t; \tau)}{\|g_{tt}(\cdot; \tau)\|_1} = \frac{\sqrt{e}(t^2 - \tau) e^{-\frac{t^2}{2\tau}}}{4\tau^{3/2}}. \quad (73)$$

In a corresponding manner, we can also start from the first-order derivative of the Gaussian kernel

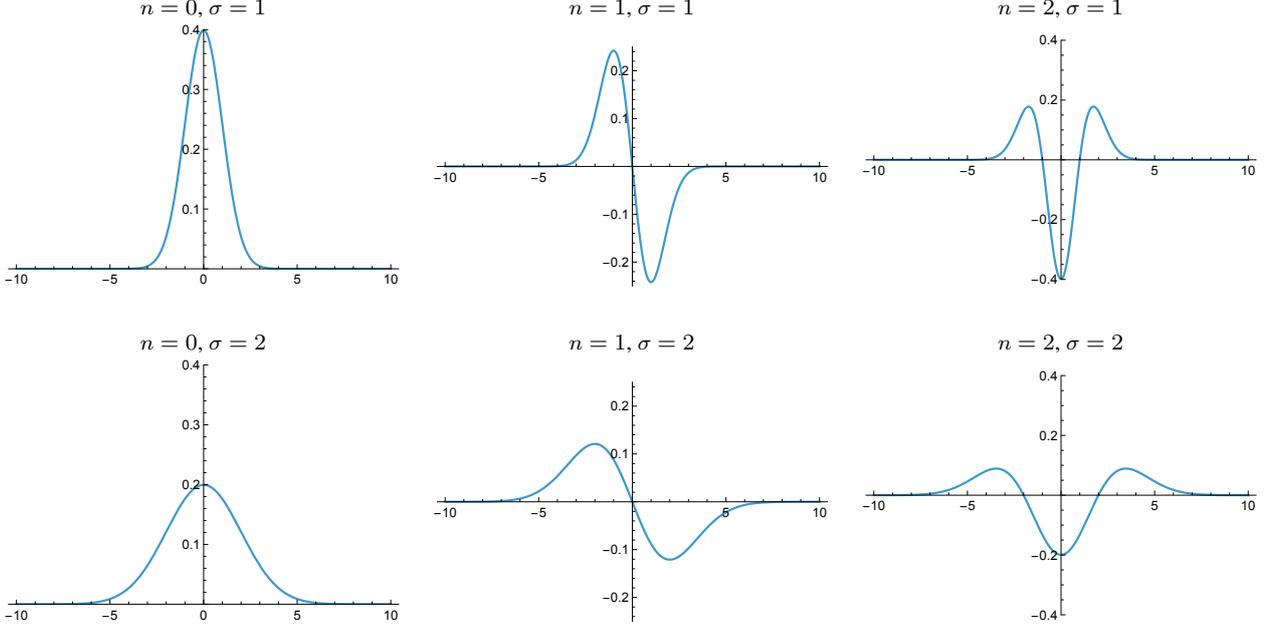
$$g_t(t; \tau) = -\frac{t}{\tau} g(t; \tau) = -\frac{t}{\tau} \frac{1}{\sqrt{2\pi\tau}} e^{-t^2/2\tau} \quad (74)$$

to define a mother wavelet. With normalization to unit  $L_2$ -norm over scales for

$$\|g_t(\cdot; \tau)\|_2 = \frac{1}{2\sqrt[4]{\pi} \tau^{3/4}}, \quad (75)$$

we obtain the following mother wavelet

$$\chi_{g_t, L_2}(t; \tau) = \frac{g_t(t; \tau)}{\|g_t(\cdot; \tau)\|_2} = -\frac{\sqrt{2} t e^{-\frac{t^2}{2\tau}}}{\sqrt[4]{\pi} \tau^{3/4}}, \quad (76)$$



**Fig. 2** The Gaussian kernel  $g(t; \tau)$  according to (52) with its scale-normalized temporal derivatives  $g_{t^n}(t; \tau) = \tau^{n\gamma/2} g_{t^n}(t; \tau)$  up to order  $n = 2$ , with the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization across scales, for different temporal scales  $\sigma = \sqrt{\tau} \in \{1, 2\}$ . (Horizontal axes: time  $t \in [-10, 10]$ . Vertical axes: kernel values with different ranges, depending on the order  $n$  of temporal differentiation.)

and with normalization to unit  $L_1$ -norm over scales for

$$\|g_t(\cdot; \tau)\|_1 = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{\tau}}, \quad (77)$$

we obtain

$$\chi_{g_t, L_1}(t; \tau) = \frac{g_t(t; \tau)}{\|g_t(\cdot; \tau)\|_1} = -\frac{t e^{-\frac{t^2}{2\tau}}}{2\tau}. \quad (78)$$

Figure 2 shows graphs of the Gaussian derivative operators underlying these constructions, here visualized in terms of scale-normalized derivatives for scale normalization power  $\gamma = 1$  and corresponding to  $L_1$ -normalization over scales.

Notably, given the uniqueness of the Gaussian kernel for constructing a scale-space representation over a symmetric one-dimensional domain, as based on the characterization of one-dimensional smoothing kernels in Section 2.3, and the uniqueness of temporal smoothing with truncated exponential kernels over a time-causal temporal domain, as also resulting from the same characterization of one-dimensional smoothing kernels, with the special role of the time-causal limit kernel as leading to temporal scale covariance, as described in Section 2.4, the Gaussian kernel can be regarded as the canonical choice of smoothing kernel over a non-causal temporal domain, whereas the time-causal limit kernel can be regarded as a canonical choice of smoothing kernel over a time-causal temporal domain.

Given this theoretical and conceptual background, we propose to consider the first- and second-order temporal derivatives of the time-causal limit kernel according (9) as the time-causal analogues of the first- and second-order temporal derivatives of the Gaussian kernel.

With complementary normalization, we then obtain the following first-order time-causal mother wavelet based on  $L_2$ -normalization

$$\chi_{\Psi_t, L_2}(t; \tau, c) = \frac{\Psi_t(t; \tau, c)}{\|\Psi_t(\cdot; \tau, c)\|_2} \quad (79)$$

and the following first-order time-causal mother wavelet based on  $L_1$ -normalization

$$\chi_{\Psi_t, L_1}(t; \tau, c) = \frac{\Psi_t(t; \tau, c)}{\|\Psi_t(\cdot; \tau, c)\|_1}, \quad (80)$$

as well as the following second-order time-causal mother wavelet based on  $L_2$ -normalization

$$\chi_{\Psi_{tt}, L_2}(t; \tau, c) = \frac{\Psi_{tt}(t; \tau, c)}{\|\Psi_{tt}(\cdot; \tau, c)\|_2} \quad (81)$$

and the following second-order time-causal mother wavelet based on  $L_1$ -normalization

$$\chi_{\Psi_{tt}, L_1}(t; \tau, c) = \frac{\Psi_{tt}(t; \tau, c)}{\|\Psi_{tt}(\cdot; \tau, c)\|_1}. \quad (82)$$

In a corresponding manner, time-causal analogues of Gaussian-derivative-based mother wavelets for other orders  $n$  of temporal differentiation as well as other exponents  $p$

$$\chi_{g_{t^n}, L_p}(t; \tau) = \frac{g_{t^n}(t; \tau)}{\|g_{t^n}(\cdot; \tau)\|_p} \quad (83)$$

can also be formulated according to

$$\chi_{\Psi_{t^n}, L_p}(t; \tau, c) = \frac{\Psi_{t^n}(t; \tau, c)}{\|\Psi_{t^n}(\cdot; \tau, c)\|_p}. \quad (84)$$

Unfortunately, the norms  $\|\Psi_{t^n}(\cdot; \tau, c)\|_p$  are hard to compute in closed form for the temporal derivatives of the time-causal limit kernel. Therefore, for implementation purposes, these norms can either be computed numerically, or be moved outside of the details of the actual numerical implementation of the wavelet analysis of the signal, by instead basing the corresponding computations on temporal derivatives of the signal that are not explicitly normalized in magnitude with respect to the  $L_p$ -norms of the mother wavelets, such as by using scale-normalized derivatives according to (45).

For such purposes, we provide the following numerical values of the  $L_2$ -norms and  $L_1$ -norms of the temporal derivatives of the time-causal limit kernel up to order 2, computed based on a truncation of the time-causal limit kernel to the 8 truncated exponential kernels having the longest time constants  $\mu_k$  for the case of the temporal scale parameter being  $\tau = 1$  and the distribution parameter set to  $c = 2$ :

$$\|\Psi_t(\cdot; 1, 2)\|_2 \approx 0.635, \quad (85)$$

$$\|\Psi_t(\cdot; 1, 2)\|_1 \approx 0.995, \quad (86)$$

$$\|\Psi_{tt}(\cdot; 1, 2)\|_2 \approx 2.084, \quad (87)$$

$$\|\Psi_{tt}(\cdot; 1, 2)\|_1 \approx 2.385, \quad (88)$$

and for the case of the distribution parameter  $c = \sqrt{2}$ :

$$\|\Psi_t(\cdot; 1, \sqrt{2})\|_2 \approx 0.513, \quad (89)$$

$$\|\Psi_t(\cdot; 1, \sqrt{2})\|_1 \approx 0.924, \quad (90)$$

$$\|\Psi_{tt}(\cdot; 1, \sqrt{2})\|_2 \approx 0.983, \quad (91)$$

$$\|\Psi_{tt}(\cdot; 1, \sqrt{2})\|_1 \approx 1.555. \quad (92)$$

To compute the values of the  $L_2$ -norms and the  $L_1$ -norms of the temporal derivatives of the time-causal limit kernels for other values of the temporal scale parameter  $\tau$ , we can then make use of the following recursive relationship (Lindeberg 2016 Equation (177))

$$\|\Psi_{t^n}(\cdot; c^{2j}, c)\|_p = c^{-j(n+1)+j/p} \|\Psi_{t^n}(\cdot; 1, c)\|_p. \quad (93)$$

to express the  $L_2$ -norms and the  $L_1$ -norms at the scale levels

$$\tau_j = c^{2j} \quad (94)$$

that are related to the base level  $\tau_0 = 1$  according to integer powers of the distribution parameter squared  $c^2$ , in units of the variance of the time-causal limit kernel.

### 3 Time-causal wavelet representations for continuous signals

#### 3.1 Time-causal wavelet representations based on temporal derivatives of the time-causal limit kernel

Given the theory presented above, for any input signal  $f: \mathbb{R} \rightarrow \mathbb{R}$  and given any order  $n > 0$  of temporal differentiation of the time-causal limit kernel  $\Psi(\cdot; \tau_k, c)$ , we can for a suitable

choice of the scale normalization factor  $\gamma$  define the continuous temporal derivative of the time-causal scale-space representation  $L_{\zeta^n}: \mathbb{R} \times \mathbb{Z} \rightarrow \mathbb{R}$  according to

$$\begin{aligned} L_{\zeta^n}(\cdot; \tau_k, c) &= \Psi_{\zeta^n}(\cdot; \tau_k, c) * f(\cdot) = \\ &= \tau^{n\gamma/2} \partial_{t^n} \Psi(\cdot; \tau_k, c) * f(\cdot) \end{aligned} \quad (95)$$

for the temporal scale levels chosen as (Lindeberg 2016 Equation (18))

$$\tau_k = \tau_0 c^{2k} \quad \text{for } k \in \mathbb{Z}. \quad (96)$$

With this parameterization, the discrete parameter  $k \in \mathbb{Z}$  has a similar role as the continuous variable  $u$  has in the parameterization of the scaling group under the exponential map  $e^u$  in the formulation of the continuous wavelet transform according to (6).

If desired, we can then complement this representation by an additional normalization with respect to a suitable  $L_p$ -norm of the  $n$ :th order temporal derivative of the time-causal limit kernel according to the theory above, then, however, also taking into account the scaling properties of the scale-normalized temporal derivatives.

In the following analysis to be performed, we will, however, not explicitly normalize the mother wavelets to constant  $L_p$ -norm, instead fully relying on the up to a uniform scaling factor normalization over scales in terms of scale-normalized temporal derivatives according to (45).

#### 3.2 Time-causal bandpass wavelet representations

If one from an input signal  $f(t)$  constructs a set of successively smoothed temporal representations  $L(t; \tau_k, c)$  by convolution with the time-causal limit kernel according to (22), then the differences between these representations

$$\Delta L(t; \tau_k, c) = L(t; \tau_k) - L(t; \tau_{k-1}, c) \quad (97)$$

will constitute a bandpass representation of the original signal  $f(t)$ , with very close structural relationships to a difference-of-Gaussians representation in the area of image processing (Burt and Adelson 1983, Crowley and Stern 1984).

In terms of convolution kernels, such a representation can equivalently, although less efficiently, be computed with a time-causal convolution kernel of the form

$$\Delta L(\cdot; \tau_k, c) = \Delta \Psi(\cdot; \tau_k, \tau_{k-1}, c) * f(\cdot), \quad (98)$$

where the difference-of-the-time-causal-limit-kernels (DoT) kernel  $\Delta \Psi$  is given by

$$\Delta \Psi(t; \tau_k, \tau_{k-1}, c) = \Psi(t; \tau_k, c) - \Psi(t; \tau_{k-1}, c) \quad (99)$$

for (Lindeberg 2016 Equation (18))

$$\tau_k = c^2 \tau_{k-1}. \quad (100)$$

Specifically, if such a representation is computed between a finest minimum temporal scale level  $\tau_1$ , for the specific choice of defining  $L(t; \tau_0, c) = f(t)$  in the boundary case, and a coarsest maximum temporal scale level  $\tau_K$ , then reconstruction of the original signal or the representation at any other temporal scale level is straightforward from such a representation, by merely adding the corresponding bandpass representations over the temporal scale channels for all the temporal scale levels:

$$L(t; \tau_j, c) = L(t; \tau_K, c) + \sum_{k=j+1}^K \Delta L(t; \tau_k, c), \quad (101)$$

which for the special choice of  $j = 0$  gives

$$f(t) = L(t; \tau_K, c) + \sum_{k=1}^K \Delta L(t; \tau_k, c). \quad (102)$$

### 3.3 Bandpass representations based on the second-order derivatives of the Gaussian kernel

Notably, there is a close structural relationship between the above notion of bandpass wavelets and to the Ricker wavelet, also known as the Mexican hat wavelet, in that the difference-of-Gaussians operator can be seen as an approximation to the Laplacian operator in the 2-D case or the second-order derivative of the Gaussian in the 1-D case.

Since the non-causal Gaussian representation of a 1-D signal defined by convolution with Gaussian kernels

$$L(\cdot; \tau) = g(\cdot; \tau) * f(\cdot) \quad (103)$$

satisfies the 1-D diffusion equation

$$\partial_\tau L = \frac{1}{2} L_{tt}, \quad (104)$$

it follows from the fact that the difference between the non-causal Gaussian-based temporal scale-space representations at the scales  $\tau$  and  $\tau + \Delta\tau$  will constitute an approximation of the derivative of the scale-space representation in the scale direction

$$\partial_\tau L(t; \tau) \approx \frac{L(t; \tau + \Delta\tau) - L(t; \tau)}{\Delta\tau}. \quad (105)$$

This difference in the Gaussian-based temporal scale-space representations can also be approximated by convolving the input image with the second-order derivative of the Gaussian kernel

$$\frac{L(t; \tau + \Delta\tau) - L(t; \tau)}{\Delta\tau} \approx \frac{1}{2} (g_{tt}(\cdot; \tau) * f(\cdot))(t; \tau). \quad (106)$$

Thus, we have that, up to a scalar normalization factor and a normalization with respect to the scale parameter, the Ricker

wavelet, also known as the Mexican hat wavelet, approximately encodes differences in a Gaussian-based scale-space representation between temporal scales.

The relationship (106) also becomes exact, if we replace the second-derivative of the Gaussian kernel with the corresponding difference-of-Gaussian kernel

$$\text{DoG}(t; \tau, \Delta\tau) = g(t; \tau + \Delta\tau) - g(t; \tau), \quad (107)$$

leading to the relationship

$$\frac{L(t; \tau + \Delta\tau) - L(t; \tau)}{\Delta\tau} = \frac{1}{2} (\text{DoG}(\cdot; \tau, \Delta\tau) * f(\cdot))(t; \tau), \quad (108)$$

as corresponding to following relationships between the difference-of-Gaussians kernel and the second-derivative of the Gaussian kernel according to

$$\text{DoG}(t; \tau, \Delta\tau) \approx g_{tt}(t; \tau), \quad (109)$$

with the approximation becoming gradually better for decreasing values of the scale difference  $\Delta\tau$ .

Specifically, regarding the second-order derivative of the non-causal Gaussian kernel, it also follows from the diffusion equation interpretation of the non-causal 1-D scale-space representation (104), that provided that second-order Gaussian-derivative-based responses are computed over a continuum of scale levels over some range  $[t_1, t_2]$ , then the signal at the finer scale level  $t_2$  can be reconstructed from the signal at the coarser scale level  $t_1$  in combination with the second-order derivative responses over this scale range according to

$$\begin{aligned} L(t; \tau_2) - L(t; \tau_1) &= \int_{t=\tau_1}^{\tau_2} \partial_\tau L(t; \tau) d\tau = \\ &= \frac{1}{2} \int_{t=\tau_1}^{\tau_2} L_{tt}(t; \tau) d\tau = \frac{1}{2} \int_{t=\tau_1}^{\tau_2} (g_{tt}(\cdot; \tau) * f(\cdot))(t; \tau) d\tau, \end{aligned} \quad (110)$$

which implies the following explicit relationship regarding non-causal reconstruction:

$$L(t; \tau_1) = L(t; \tau_2) + \frac{1}{2} \int_{t=\tau_1}^{\tau_2} (g_{tt}(\cdot; \tau) * f(\cdot))(t; \tau) d\tau. \quad (111)$$

This expression can be seen as a non-causal analogue of the corresponding reconstruction (101) from a time-causal bandpass representation, again demonstrating a close structural relationship to regard the temporal derivatives of the time-causal limit kernel as time-causal analogues of the derivatives of the regular Gaussian kernel.

3.4 Conceptual difference between the bandpass representations based on the time-causal limit kernel in relation to the bandpass representation based on the non-causal Gaussian kernel

Regarding the difference-of-the-time-causal-limit kernel (99) used in the temporal bandpass wavelets representation in Section 3.2, it follows from the differential equation interpretation (30) in terms of first-order integrators coupled in cascade, that this difference kernel does instead represent the first-order temporal derivative of the time-causal limit kernel according to

$$\Delta\Psi(t; \tau_k, \tau_{k-1}, c) = -\mu_k \Psi_t(t; \tau_k, c), \quad (112)$$

and also with the notable difference that the difference between the temporal scale levels  $\tau_k$  and  $\tau_{k-1}$  now has to be macroscopic of the form

$$\tau_k = c^2 \tau_{k-1}. \quad (113)$$

#### 4 Discrete approximations of temporal derivatives of the time-causal limit kernel

To transfer the above theory over a continuous time-causal temporal domain to a corresponding discrete time-causal domain, there is indeed a corresponding discrete temporal scale-space theory to build upon, which can be used for constructing both theoretically well-founded and numerically well-conditioned discrete implementations.

##### 4.1 Discretization of the first-order integrators

A basic result is that the canonical discretization of each layer in the set of first-order integrators coupled in cascade (30)

$$\partial_t f_{\text{out}}(t) = -\frac{1}{\mu_{k,\text{cont}}} (f_{\text{out}}(t) - f_{\text{in}}(t)) \quad (114)$$

is a first-order recursive filter of the form

$$f_{\text{out}}(t) - f_{\text{out}}(t-1) = -\frac{1}{1 + \mu_{k,\text{disc}}} (f_{\text{out}}(t-1) - f_{\text{in}}(t)), \quad (115)$$

where we here assume that the discrete signals in Equation (115) are sampled with a grid spacing  $\Delta t$  relative to the continuous signals in Equation (114). Then, it can be shown that (Lindeberg 2025b Appendix B.3) the discrete recurrence relation (115) constitutes a true discretization of the continuous differential equation (114) provided that

$$\mu_{k,\text{disc}} = \frac{\mu_{k,\text{cont}}}{\Delta t}. \quad (116)$$

##### 4.2 Classification of discrete temporal smoothing kernels

The result that the time-causal smoothing operation should be discretized on the form (115) also builds upon a classification of discrete variation-diminishing convolution transformations by Schoenberg (1948), which states that a discrete kernel guarantees non-creation of new zero-crossings in the convolved signal if and only if its generating function is of the form

$$\varphi(z) = c z^k e^{(q_{-1}z^{-1} + q_1z)} \prod_{i=1}^{\infty} \frac{(1 + \alpha_i z)(1 + \delta_i z^{-1})}{(1 - \beta_i z)(1 - \gamma_i z^{-1})} \quad (117)$$

where  $c > 0$ ,  $k \in \mathbb{Z}$ ,  $q_{-1}, q_1, \alpha_i, \beta_i, \gamma_i, \delta_i \geq 0$  and  $\sum_{i=1}^{\infty} (\alpha_i + \beta_i + \gamma_i + \delta_i) < \infty$ , from which we can recognize the  $z$ -transform of the discrete recurrence relation (115) as one of the members of the class of possible functions, as occurring in the denominator of the product expression in (117):

$$f_{\text{out}}(x) = f_{\text{in}}(x) + \beta_i f_{\text{out}}(x-1) \quad (0 \leq \beta_i < 1). \quad (118)$$

##### 4.3 Discrete analogue of the time-causal limit kernel

For formulating a discrete analogue of the time-causal limit kernel, one special fact to consider is that while the temporal mean of the first-order recursive filter is structurally similar as for the truncated exponential filter  $V = m_k$ , the expression for the variance of the discrete recursive filter is structurally different

$$\Delta\tau_k = \mu_k^2 + \mu_k. \quad (119)$$

Thus, to determine the time constants  $\mu_k$  from the desired scale increments  $\Delta\tau_k$  between adjacent temporal scale levels according to

$$\Delta\tau_k = \tau_k - \tau_{k-1} = \tau_0 \left( c^{2k} - c^{2(k-1)} \right), \quad (120)$$

when  $k > 1$ , for some suitably sufficiently low<sup>6</sup> selected first temporal level  $\tau_k > 0$ , and the first amount of incremental smoothing emulating a discrete approximation of all the truncated exponential kernels that have been truncated away according to

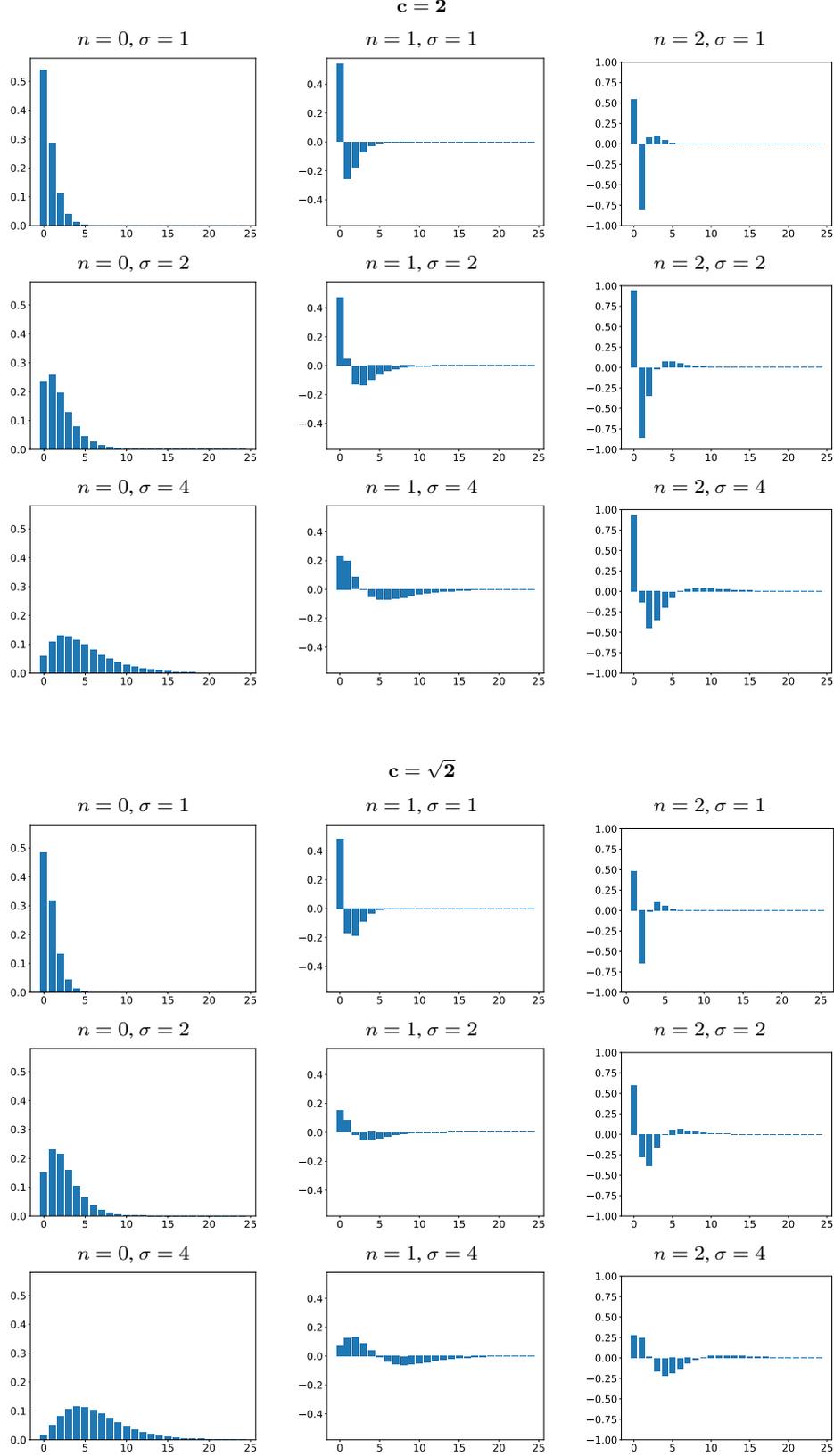
$$\Delta\tau_1 = \tau_0 c^2, \quad (121)$$

we should use (Lindeberg 2023 Equation (55))

$$\mu_k = \frac{\sqrt{1 + 4\Delta\tau_k} - 1}{2}. \quad (122)$$

Algorithm 1 summarizes the different steps involved in implementing a discrete approximation of convolution with the time-causal limit kernel in this way.

<sup>6</sup> In order to compute an actual discrete approximation of the convolution of a discrete signal at a certain level, a sufficient number of initial scale levels needs to be implemented, to approximate the truncation of the infinite convolution sufficiently well. For our purposes, we often use a minimum of 4 to 8 temporal scale levels.



**Fig. 3** Equivalent discrete approximations of the time-causal limit kernel  $\Psi(t; \tau, c)$  as obtained from a set of first-order recursive filters coupled in cascade, with the time constants determined according to (122), and with the infinite cascade truncated after the 8 recursive filters having the longest time constants, together with discrete approximations of its scale-normalized temporal derivatives  $\Psi_{c^n}(t; \tau, c) = \tau^{n\gamma/2} \Psi_{t^n}(t; \tau, c)$  up to order  $n = 2$ , obtained by applying first- and second-order temporal difference operators  $\delta_t$  and  $\delta_{tt}$  according to (123) to the discrete approximations of time-causal limit kernel, with the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization across scales, for different combinations of the temporal scales  $\sigma = \sqrt{\tau} \in \{1, 2, 4\}$  and different values to the distribution parameter  $c \in \{\sqrt{2}, 2\}$ . (Horizontal axes: time  $t \in [0, 25]$ . Vertical axes: kernel values with different ranges  $[0, 0.50]$ ,  $[-0.40, 0.40]$  or  $[-1.00, 1.00]$ , depending on the order  $n$  of temporal differentiation.)

**Algorithm 1** Pseudocode for the discrete time-causal temporal filtering module corresponding to the set of first-order recursive filters of the form (115) coupled in cascade, which implements convolution with a discrete approximation of the time-causal limit kernel  $\Psi(t; \tau, c)$  according to (9). Here, it is assumed that the time constants  $\mu_k > 0$  have been already computed according to (122). The variable  $level\_prev$  represents a memory from the previous frame, necessary to compute the temporal differences that drive the recursive filters. In this respect, the algorithm is strictly time-recursive, since it only makes use of information from the present moment and a very short-term memory from the previous frame.

---

```

procedure DISC-TIME-CAUS-FILT( $f, \mu$ )                                ▷  $f$  input stream,  $\mu$  of size  $K$ 
   $level \leftarrow 0$                                                 ▷ of size  $K$ 
   $level\_prev \leftarrow 0$                                          ▷ of size  $K$ 
   $n \leftarrow 0$                                                   ▷ time counter
  repeat
     $signal \leftarrow f(n \Delta t)$                                 ▷ read the input stream with time increment  $\Delta t$ 
    for  $k \leftarrow 1, K$  do
      if  $k = 1$  then
         $level_k \leftarrow level\_prev_k + (signal - level\_prev_k)/(1 + \mu_k)$   ▷ the first layer
      else
         $level_k \leftarrow level\_prev_k + (level_{k-1} - level\_prev_k)/(1 + \mu_k)$   ▷ the higher layers
      end if
    end for
     $level\_prev \leftarrow level$                                     ▷ update the buffer for the previous time frame
     $n \leftarrow n + 1$                                            ▷ prepare for the next time frame
  until interrupt
end procedure

```

---

#### 4.4 Discrete approximations of temporal derivatives of the time-causal limit kernel

Given that the input signal has been smoothed to a set of temporal scales in this way, discrete approximations of temporal derivatives can then be computed with temporal difference operators of the form

$$\delta_t = (-1, +1), \quad \delta_{tt} = (1, -2, 1), \quad (123)$$

and scale-normalized derivatives according to (45) in turn be computed by multiplying the regular discrete derivative approximations by an appropriate power of the scale parameter  $\tau$  in units of the variance of the underlying equivalent discrete temporal smoothing kernel.

Figure 3 shows examples of such equivalent discrete derivative approximation kernels, obtained by filtering a discrete delta function with the proposed methodology for computing discrete approximation of temporal derivatives at multiple temporal scales. As can be seen from these graphs, the discrete nature of the smoothing kernels and their discrete derivative approximations is significant for lower temporal scale levels, while the shapes of the kernels approach their continuous counterparts in Figure 1 with increasing values of the temporal scale parameter  $\tau$ .

## 5 Time-causal wavelet representations for discrete signals

Given that we use the scheme in Algorithm 1 for smoothing a discrete signal stream  $f: \mathbb{Z} \rightarrow \mathbb{R}$  with an approximation of the discrete analogue of the time-causal limit kernel  $\Psi_{\text{disc}}(t; \tau, c)$ , and then compute discrete approximations of temporal derivatives using the time-causal discrete temporal

derivative operators  $\delta_t$  and  $\delta_{tt}$  according to (123), complemented with scale normalization according to

$$\delta_{\zeta^n} = \delta_{t^n, \text{norm}} = \tau^{n\gamma/2} \delta_{t^n} \quad (124)$$

for some suitably selected value of the scale normalization parameter  $\gamma > 0$ , we thereby obtain numerical approximations of the discrete analogue  $L_{\zeta^n}: \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$  of the temporal derivative of order  $n$  of the time-causal scale-space representation of the signal  $f$  according to

$$\begin{aligned} L_{\zeta^n}(\cdot; \tau_k, c) &= \Psi_{\zeta^n, \text{disc}}(\cdot; \tau_k, c) * f(\cdot) = \\ &= \tau^{n\gamma/2} \delta_{t^n} \Psi_{\text{disc}}(\cdot; \tau_k, c) * f(\cdot) \end{aligned} \quad (125)$$

for the temporal scale levels chosen as

$$\tau_k = \tau_0 c^{2k} \quad \text{for } k \in \mathbb{Z}. \quad (126)$$

With this parameterization, the discrete parameter  $k \in \mathbb{Z}$  again has a similar role as the continuous variable  $u$  has in the parameterization of the scaling group under the exponential map  $e^u$  in the formulation of the continuous wavelet transform according to (6).

Alternatively, given the raw undifferentiated temporal scale-space representation equivalently obtained by convolution with the discrete analogue  $\Psi_{\text{disc}}(t; \tau_k, c)$  of the time-causal limit kernel according to

$$L(\cdot; \tau_k, c) = \Psi_{\text{disc}}(\cdot; \tau_k, c) * f(\cdot), \quad (127)$$

we can compute a bandpass wavelet representation from

$$\Delta L(t; \tau_k, c) = L(t; \tau_k) - L(t; \tau_{k-1}, c), \quad (128)$$

which is computationally equivalent to convolving the input signal  $f$  with the bandpass wavelet filter

$$\Delta \Psi_{\text{disc}}(t; \tau_k, \tau_{k-1}, c) = \Psi_{\text{disc}}(t; \tau_k) - \Psi_{\text{disc}}(t; \tau_{k-1}, c)$$

(129)

between adjacent levels of temporal scales

$$\tau_k = c^2 \tau_{k-1}. \quad (130)$$

A technical difference compared to the continuous case is, however, the following: Whereas for the continuous time-causal wavelets, the bandpass wavelet representation is directly related to the first-order temporal derivative of the time-causal scale-space representation according to (30)

$$L(t; \tau_k, c) - L(t; \tau_{k-1}, c) = -\mu_k \partial_t L(t; \tau_k, c), \quad (131)$$

for the discrete time-causal wavelets, the bandpass wavelet representation is only related to the discrete approximation of the first-order temporal derivative of the time-causal scale-space representation, if we additionally include a temporal shift when defining the bandpass representation according to the recurrence relation (115):

$$\begin{aligned} L(t-1; \tau_k, c) - L(t; \tau_{k-1}, c) &= \\ &= -(1 + \mu_k) (L(t; \tau_k, c) - L(t-1; \tau_k, c)) = \\ &= -(1 + \mu_k) \delta_t L(t; \tau_k, c). \end{aligned} \quad (132)$$

## 6 Characterizations of scaling properties in discretizations of the continuous theory

The theory for scale-covariant time-causal and time-recursive wavelets in Section 2 is developed for fully continuous signals. In this section, we will characterize how well the continuous scaling properties carry over to a discrete implementation, as based on the methodology proposed in Sections 4 and 5, and depending on the relationship between the temporal scale levels in relation to the sampling distance in the signal.

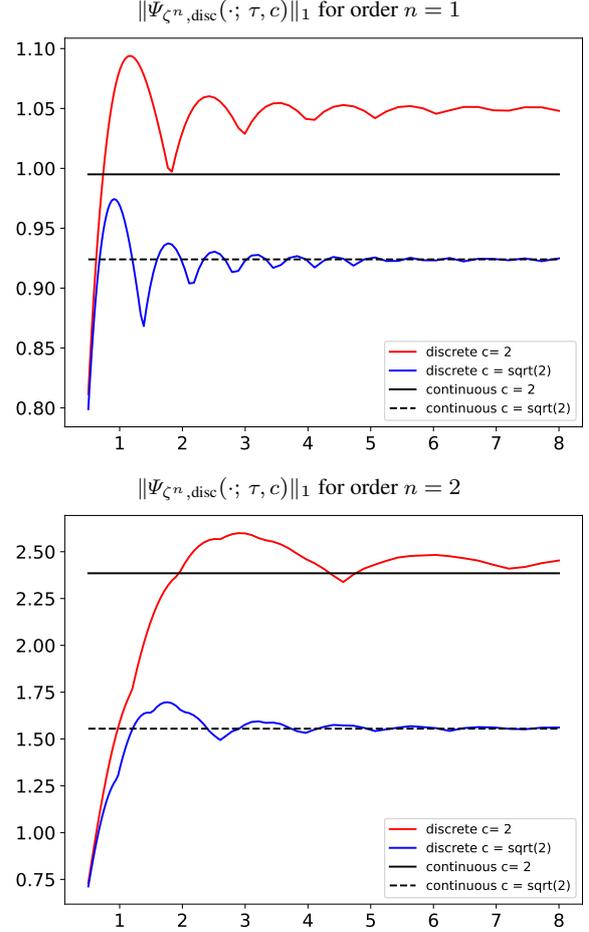
### 6.1 Scale dependencies of the discrete $l_p$ -norms of the equivalent discrete derivative approximation kernels

Figure 4 shows a first characterization in this respect, by showing how the discrete  $l_1$ -norms of the equivalent convolution kernels corresponding to discrete approximations of the time-causal limit kernel

$$\|\Psi_{\zeta^n, \text{disc}}(t; \tau, c)\|_p = \|\tau^{n\gamma/2} \delta_t^n \Psi_{\text{disc}}(t; \tau, c)\|_p \quad (133)$$

for the scale normalization power  $\gamma = 1$  vary as function of the scale parameter  $\tau$ , in the graphs notably parameterized in terms of the standard deviation  $\sigma = \sqrt{\tau}$  instead of the variance  $\tau$ .

For reference, the graphs also show the evolution properties of the  $L_1$ -norms of the temporal derivatives of the continuous time-causal limit kernel, according to (86), (88), (90) and (92), which are constant over scales for  $\gamma = 1$ .



**Fig. 4** Graphs over the dependency on the scale parameter  $\sigma = \sqrt{\tau}$  for the discrete  $l_1$ -norms according to (133) of the equivalent discrete kernels illustrated in Figure 3, that correspond to the discrete approximations of the derivatives of the time-causal limit kernel, as obtained with the discretization methodology described in Section 4 for the choice of the scale normalization power  $\gamma = 1$ . For reference, the graphs also show the evolution properties over scales of the  $L_1$ -norms of the temporal derivatives of the continuous time-causal limit kernel, according to (86), (88), (90) and (92), which are constant over scales for  $\gamma = 1$ . As can be seen from these graphs, there are some transient phenomena at finer levels of scales, where the influence of the discretization effects is strongest. Towards coarser levels of scales, however, the discrete  $l_p$ -norms asymptotically approach constant values. When the distribution parameter  $c = \sqrt{2}$ , there are very good matches between the asymptotic values and the corresponding expressions for the fully continuous theory. When the distribution parameter  $c = 2$ , there are, however, certain deviations. (Horizontal axes: temporal scale parameter in units of  $\sigma = \sqrt{\tau}$ . Vertical values: magnitudes of the discrete  $l_p$ -norms, with red curves showing the results for the distribution parameter  $c = 2$  and blue curves showing the results for the distribution parameter  $c = \sqrt{2}$ .)

As can be seen from these graphs, there are some transient phenomena at finer levels of scales, where the influence of the discretization effects is strongest. Towards coarser levels of scales, however, the discrete  $l_p$ -norms asymptotically approach constant values.

When the distribution parameter is  $c = \sqrt{2}$ , we obtain very good matches between the asymptotic values and

the corresponding expressions for the fully continuous time-causal model. When the distribution parameter is  $c = 2$ , there are, however, certain deviations, of the order of 5 % relative to the continuous theory. This quantification thereby demonstrates that the discretization effects due to the discrete nature of the temporal scale levels may be notable when the distribution parameter  $c = 2$ , while there is a much better match between the continuous and discrete theories when the distribution parameter is  $c = \sqrt{2}$ .

Concerning how to interpret these curves, it should be noted that here we have treated the scale parameter as a continuous variable. Given that the temporal scale levels are to be inherently discrete in a temporal scale-space representation, this corresponds to emulating the effect of using possibly different initial scale values  $\tau_0$  when initiating the temporal scale levels according to (12). In a practical implementation, a fixed value of the initial scale value  $\tau_0$  should, however, always be used, thus implying that the resulting graphs will only assume values for the values of the scale parameter  $\sigma = \sqrt{\tau}$  that are integer multiples of the distribution parameter  $c$ .

The reason why there are certain oscillations in these curves is because of the positive and the negative lobes in the discrete approximations of the derivatives in the time-causal limit kernel interact with the grid spacing, which thereby affects the discrete  $l_1$ -norms.

## 6.2 Scale selection properties for local blob-like and edge-like structures in the signal

To investigate the scale selective properties of the proposed methodology for defining time-causal wavelets, let us next focus on the ability of the time-causal wavelets in terms of scale-normalized derivatives of the time-causal limit kernel to reflect the inherent characteristic scales in the input data.

For this purpose, we generated blob-like and edge-like temporal structures in terms of discrete approximations of the Gaussian kernel in terms of the discrete analogue of the Gaussian kernel defined by (Lindeberg 1990 Equation (19))

$$T(m; \tau) = e^{-\tau} I_m(\tau), \quad (134)$$

where  $I_m(\tau)$  denotes the modified Bessel functions of integer order (Abramowitz and Stegun 1964). More, specifically blob-like temporal structures were generated as (see Figure 5(left))

$$f_{\text{blob}}(t; \tau_{\text{ref}}) = T(t; \tau_{\text{ref}}) \quad (135)$$

and edge-like structures according to (see Figure 5(right))

$$f_{\text{edge}}(t; \tau_{\text{ref}}) = \sum_{m=-\infty}^t T(m; \tau_{\text{ref}}). \quad (136)$$

Then, we computed the discrete time-causal wavelet representations of these signals over the range of temporal scale levels  $\sigma = \sqrt{\tau} \in [1/8, 64]$  (see Figure 5), by computing

- discrete approximations of the second-order scale-normalized temporal derivatives for the scale normalization power  $\gamma = 3/4$  of the time-causal wavelet representation of the idealized blob models  $f_{\text{blob}}(t; \tau_{\text{ref}})$  according to (135), as well as
- discrete approximations of the first-order scale-normalized temporal derivatives for the scale normalization power  $\gamma = 1/2$  of the time-causal wavelet representation of the idealized edge models  $f_{\text{edge}}(t; \tau_{\text{ref}})$  according to (136).

Thereafter, we determined the time moments and the temporal scales at which these discrete approximations of the scale-normalized temporal derivative responses assumed their maximum values over time and temporal scales

$$(\hat{t}_{\text{blob}}, \hat{\tau}_{\text{blob}}) = \operatorname{argmax}_{t, \tau} |(\Psi_{\zeta, \text{disc}}(\cdot; \tau, c) * f_{\text{blob}}(\cdot; \tau_{\text{ref}}))(t)|, \quad (137)$$

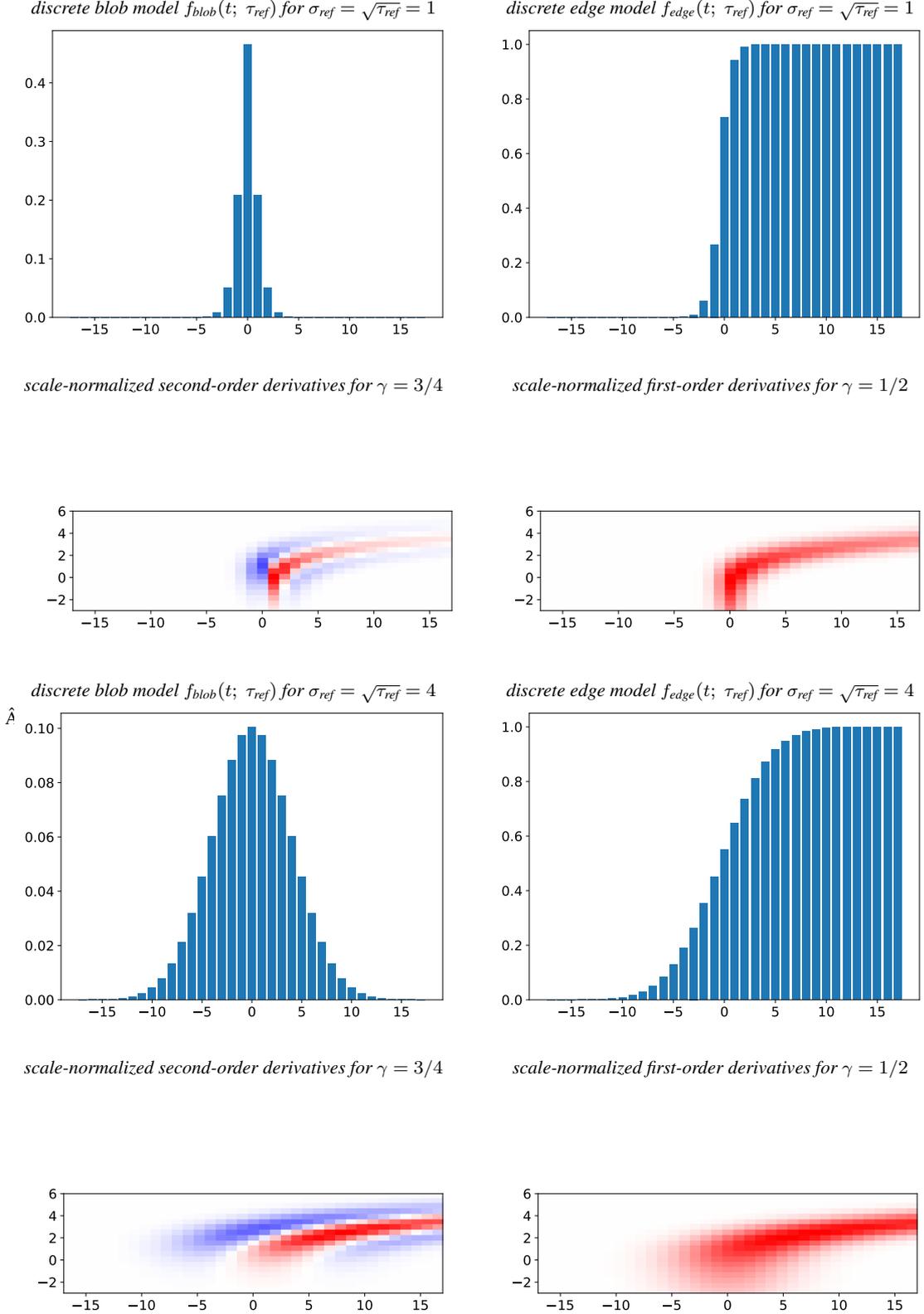
$$(\hat{t}_{\text{edge}}, \hat{\tau}_{\text{edge}}) = \operatorname{argmax}_{t, \tau} |(\Psi_{\zeta, \text{disc}}(\cdot; \tau, c) * f_{\text{edge}}(\cdot; \tau_{\text{ref}}))(t)|. \quad (138)$$

Such points computed in a temporal scale-space representation are referred to as scale-space extrema, and constitute a generalization of a corresponding methodology developed for a spatial scale-space representation (Lindeberg 1998a, 1998b). Specifically, based on the theory described in Section 2.7.2, in case the time-causal limit kernel would be replaced by a non-causal Gaussian kernel, this methodology would in the ideal continuous case determine the inherent temporal duration of the corresponding idealized temporal blob and edge models according to (57) and (62).

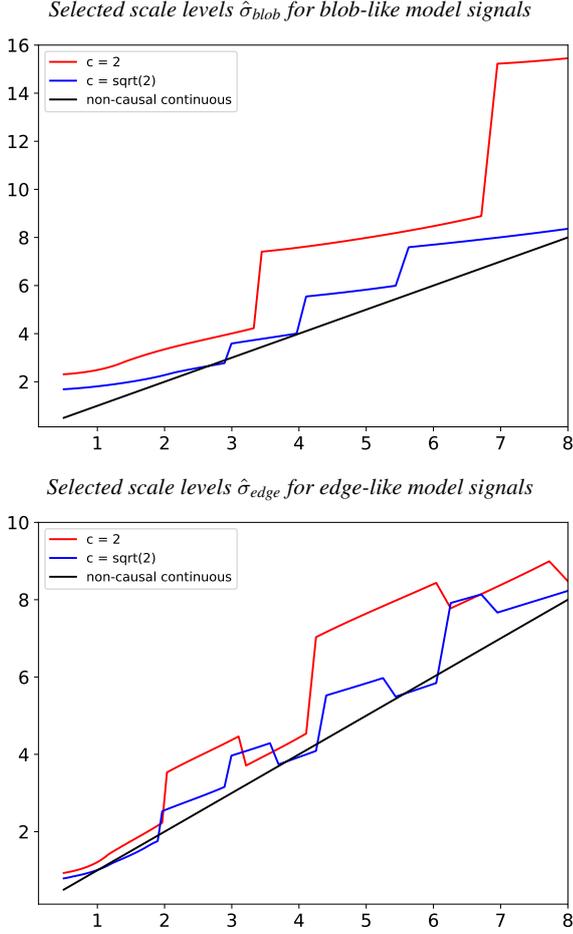
Figure 6 shows graphs of the dependency of the resulting scale estimates for a range of reference scale levels  $\tau_{\text{ref}}$ , when detecting scale-space extrema in this way for the reference blob and edge signals  $f_{\text{blob}}(t; \tau_{\text{ref}})$  and  $f_{\text{edge}}(t; \tau_{\text{ref}}$  according to (135) and (136), when using time-causal wavelet representations for a fixed reference scale  $\tau_0 = 1$  in (12) using the distribution parameter set to either  $c = 2$  or  $c = \sqrt{2}$ . Given the discrete nature of both the temporal scale levels and the temporal sampling, we used local parabolic interpolation around each extremum over temporal scales to increase the accuracy in the resulting estimates of the temporal scales  $\hat{\tau}$ .

As can be seen from the results, the resulting scale estimates largely follow a linear increase in the selected scale levels as function of the reference scales  $\tau_{\text{ref}}$  of the model signals. Due to the discrete nature of the temporal scale levels, and the lack of provable scale covariance properties for temporal scaling factors  $S_t$  that are not integer powers of the distribution parameter  $c$ , we cannot expect the graphs of the selected scale levels to be smooth curves across the discontinuities over the temporal scale levels.

Notably, the quantization effects decrease substantially when decreasing the distribution parameter  $c$  from  $c = 2$  to  $c = \sqrt{2}$ . In this respect, there is a trade-off issue in that decreasing the value of the distribution parameter  $c$  leads

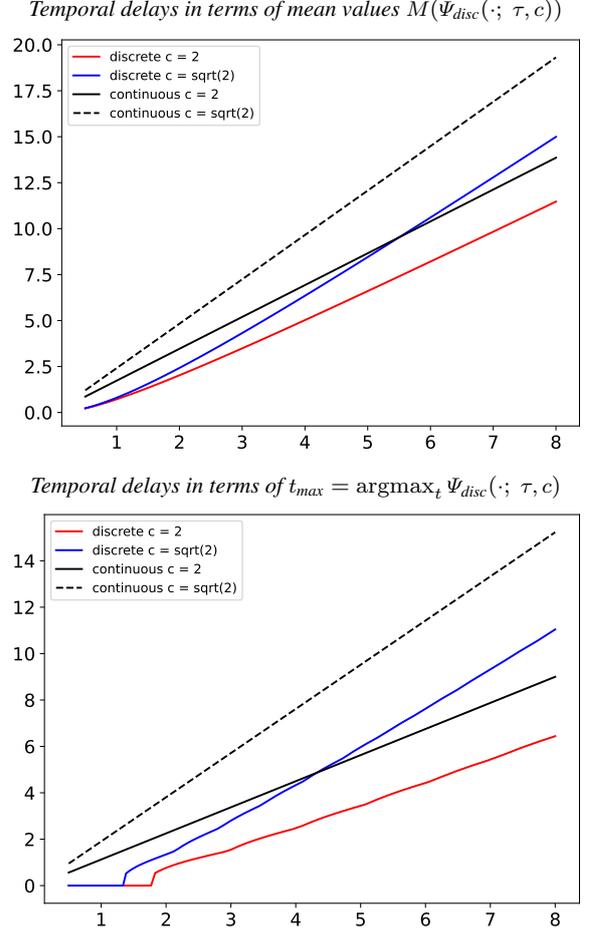


**Fig. 5** Illustration of the conceptual stages involved in characterizing the scaling properties of the scale-normalized temporal derivatives of the time-causal limit kernel. The left column show two discrete blob-like model signals based on the discrete analogue of the Gaussian kernel according to (135) for two different reference scales  $\sigma_{\text{ref}} = \sqrt{\tau_{\text{ref}}} \in \{1, 4\}$ , together with the scale-normalized derivatives of order 2 for the scale normalization power  $\gamma = 3/4$  over the scale range  $\sigma = \sqrt{\tau} \in [1/8, 64]$ . The right column shows two edge-like model signals according to (136) together with the scale-normalized derivatives of order 1 for  $\gamma = 1/2$  over the same scale ranges. All results have been computed with the distribution parameter  $c = \sqrt{2}$  for the discrete analogue of the time-causal limit kernel. (The reason why the slopes of the red and blue stripes are not vertical, but oblique, is because of the different amounts of temporal delay at different temporal scales.) (Horizontal axes: time  $t \in [-17, 17]$ . Vertical axes in the temporal scale-space representations: Effective scale =  $\log_2 \sqrt{\tau}$ .) (In the time-causal wavelet representations, red denotes positive values and blue denotes negative values.)



**Fig. 6** The selected scale levels  $\hat{\sigma}_{\text{blob}} = \sqrt{\hat{\tau}_{\text{blob}}}$  and  $\hat{\sigma}_{\text{edge}} = \sqrt{\hat{\tau}_{\text{edge}}}$  from maxima over scales of scale-normalized temporal derivatives for blob-like structures according to (137) when using the scale normalization power  $\gamma = 3/4$  according to (56) and for edge-like structures according to (138) when using the scale normalization power  $\gamma = 1/2$  according to (61), computed from the time-causal wavelet representations of (top) ideal blob-like model signals of the form (135) and (bottom) ideal edge-like model signals of the form (136) under variations of the reference scale  $\tau_{\text{ref}}$  in units of the standard deviation  $\sigma_{\text{ref}} = \sqrt{\tau_{\text{ref}}}$  of the underlying model functions, and for different values of the distribution parameter  $c = \sqrt{2}$  (blue curves) or  $c = 2$  (red curves). For comparison, the corresponding scale estimates  $\hat{\sigma} = \sigma_{\text{ref}}$  according to (57) and (62) that would be obtained for a non-causal fully continuous Gaussian-derivative-based wavelet representation are also shown (black curves). As can be seen from the graphs, there is largely a linear dependency between the reference scales  $\sigma_{\text{ref}}$  and the resulting scale estimates  $\hat{\sigma}_{\text{ref}}$ . There are, however, also clear discontinuities in the graphs, caused by the discrete nature of the temporal scale levels. Notably, the discretization effects decrease significantly when decreasing the distribution parameter from  $c = 2$  to  $c = \sqrt{2}$ . For time-critical applications, however, a larger value of the distribution parameter leads to shorter temporal delays. (Horizontal axes: reference scale  $\sigma_{\text{ref}}$  in units of the standard deviations of the underlying kernels. Vertical axes: scale estimates  $\hat{\sigma}_{\text{ref}}$ , also in units of the standard deviation of the underlying kernels.)

to more accurate scale estimates from the underlying signal, whereas increasing the distribution parameter  $c$  leads to shorter temporal delays; compare with Equations (28) and (29), and see also the following section.



**Fig. 7** Measures of the temporal delays of the discrete approximations of the time-causal limit kernel in terms of the temporal mean according to (139) and in terms of the temporal maximum point according to (140) for the distribution parameter  $c = \sqrt{2}$  (blue curves) and  $c = 2$  (red curves). For comparison, the results from the corresponding expressions (28) and (29) for the continuous kernels are also shown (black curves). (Horizontal axes in the temporal scale-space representations: scale parameter  $\sigma$  in units of the standard deviation of the underlying kernels. Vertical axes: delay estimates in units of [time].)

### 6.3 Temporal delays of the discrete approximations of the time-causal limit kernel

To characterize the temporal delays for the discrete approximations of the time-causal limit kernel, we can compute corresponding discrete analogues of the temporal delay estimates for the continuous time-causal limit kernel in terms

of the temporal mean according to Equation (28)

$$M(\Psi_{\text{disc}}(\cdot; \tau, c)) = \frac{\sum_{t=0}^{\infty} t \Psi_{\text{disc}}(t; \tau, c)}{\sum_{t=0}^{\infty} \Psi_{\text{disc}}(t; \tau, c)} = \sum_{k=1}^{\infty} \mu_k, \quad (139)$$

with the discrete time constants  $\mu_k$  according to (122) and (120), and the time value at which the discrete approximation of the time-causal limit kernel assumes its maximum value over scales

$$t_{\text{max}} = \operatorname{argmax}_t \Psi_{\text{disc}}(t; \tau, c). \quad (140)$$

Figure 7 shows the result of computing these characterizations of the temporal delay for the distribution parameter  $c \in \{\sqrt{2}, 2\}$ , and with the estimate of  $t_{\text{max}}$  interpolated to subquantization resolution using parabolic interpolation. For comparison, the graphs also show the dependencies on the temporal scale levels for the continuous temporal delay estimates according to Equations (28) and (29). As can be seen from the graphs, the measures of the temporal delays are significantly shorter for the distribution parameter  $c = 2$  compared to the distribution parameter  $c = \sqrt{2}$ . We can also note that the measures of the temporal delays are shorter for the discrete approximations of the time-causal limit kernel compared to the continuous time-causal limit kernel.

The reason for the latter difference between the continuous and the discrete theories is that the time constants  $\mu_k$  are determined in a different way from the differences  $\Delta\tau_k$  in the temporal scale levels for the discrete kernels than for the continuous kernels; compare (122) in the discrete case with the expression  $\mu_k = \sqrt{\Delta\tau_k}$  in the continuous case.

## 7 Experiments

To demonstrate the ability of the proposed time-causal wavelet representations to adaptively respond to local structures at different temporal scales in a signal, let us consider the result of performing a time-causal wavelet analysis of the following model signal with exponentially varying frequency:

$$f(t) = \sin\left(\exp\left(\frac{b-t}{a}\right)\right) \quad (141)$$

for  $a = 200$  and  $b = 1000$ , see Figure 8 for an illustration.

For this purpose, we computed discrete approximations of the first-order and second-order temporal derivatives according to (125) over the scale range  $\sigma = \sqrt{\tau} \in [1/8, 1024]$  for the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization over scales of the underlying temporal derivatives of the continuous time-causal limit kernel. The motivation for this choice, is to treat responses at different scales in a similar manner, so that the magnitudes of the temporal derivative responses can be readily compared across scales, as manifested in the scale covariance property (50).

To define a feature descriptor that is less sensitive to the local phase in the signal near the preferred scales, as determined by the temporal duration of the locally dominant

temporal structures, we also computed the following quasi quadrature measure (Lindeberg 2018b)

$$QL = \sqrt{L_{\zeta}^2 + C L_{\zeta\zeta}^2} \quad (142)$$

for the weighting parameter  $C = 1/\sqrt{2}$ .

As can be seen from the results, the dominant responses in the time-causal wavelet representations across scales move from finer to coarser scales with increasing local wavelengths in the signal, in agreement with the desirable properties of a multi-scale wavelet representation, and also in agreement with the scaling properties. Due to the exponential increase of the local wavelengths in the input signal as function of time, the corresponding temporal scale estimates ought to also increase exponentially over time. With the logarithmic parameterization of the temporal scale parameter used in the figures, we experimentally obtain scale dependencies that are well described as linear, thus with a very good match between the continuous theory and the discrete implementation.

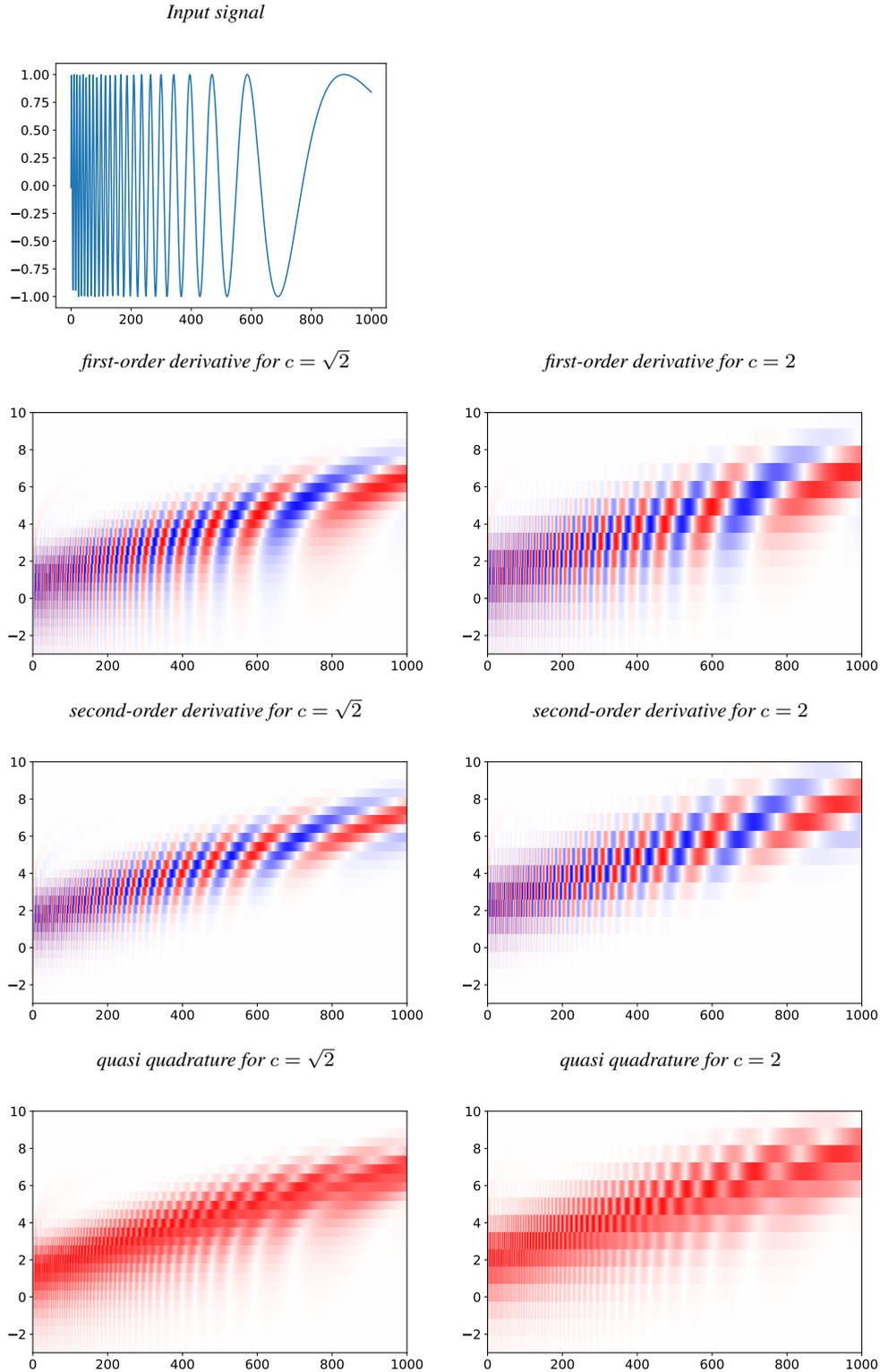
The quantization effects are notably stronger when using a sparser sampling of the temporal scale levels, as induced by setting the distribution parameter to  $c = 2$ , while the quantization effect are much lower when using a denser sampling of the temporal scale levels, as induced by setting the distribution parameter to  $c = \sqrt{2}$ .

In these respects, the results from this experiment are in agreement with the theoretical results presented in the previous sections, and demonstrate how the proposed family of time-causal wavelet representations can locally capture and reflect the temporal scales of locally dominant temporal structures in a time-causal analysis of temporal signals.

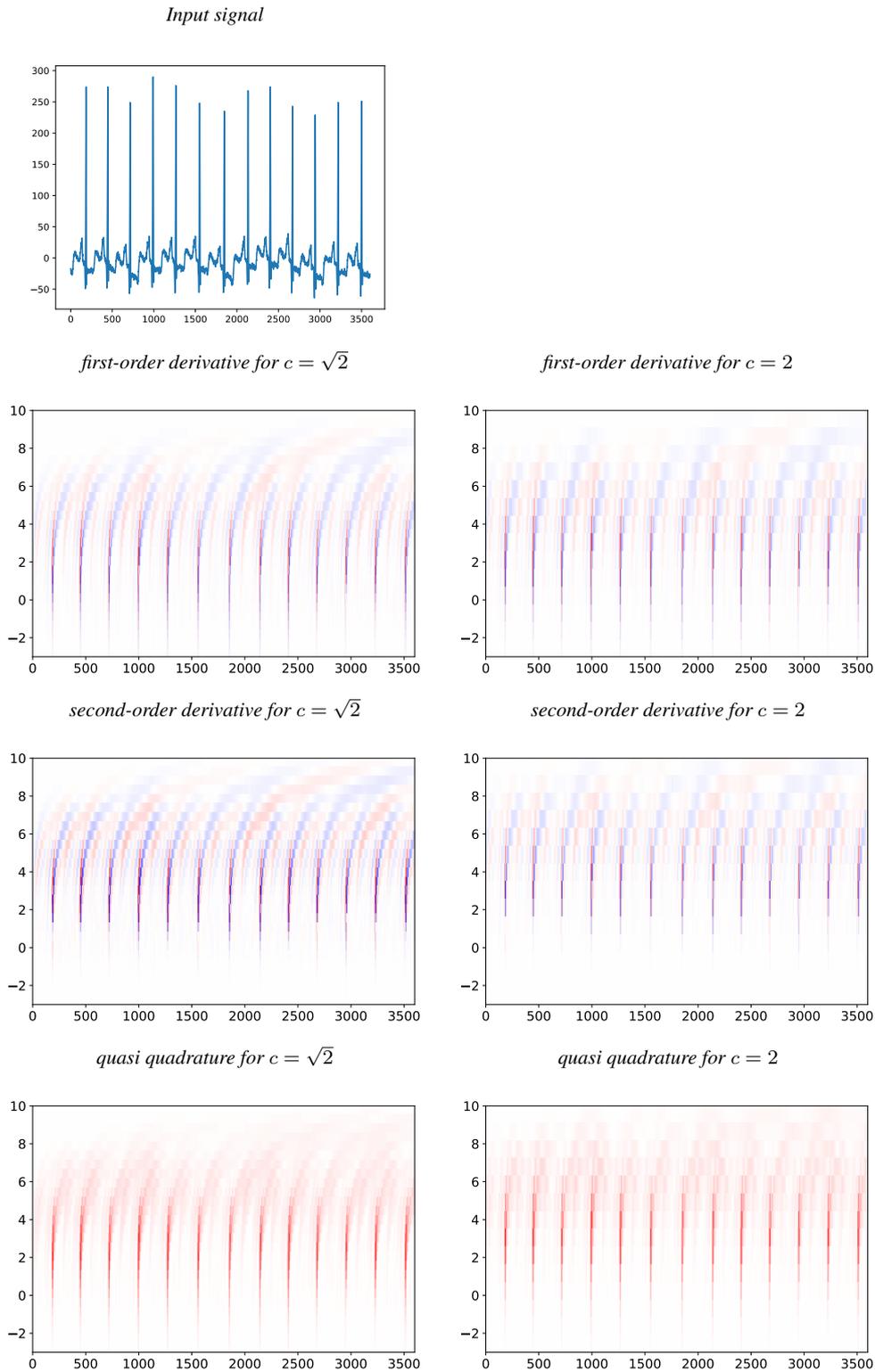
Let us also illustrate the effects of applying the time-causal frequency analysis to three real-world signals: Figure 9 shows such results for an electrocardiography signal from the PhysioNet service (<http://www.physionet.org>) from the MIT-BIH Arrhythmia database, item “NSR/100m (0).mat” from (Plawiak 2017). As can be seen from the results, there are strong values in the wavelet transforms near the transients in the signal. At coarser temporal scales, the temporal delays become longer. The temporal delays are, however, significantly shorter when the distribution parameter is  $c = 2$  compared to using  $c = \sqrt{2}$ .

Figure 10 shows such results from a signal with measurements of vibrations from the NASA Bearing Dataset (<https://www.kaggle.com/datasets/vinayak123tyagi/bearing-dataset>) (1st test recorded 2003.11.25.23.39.56). As can be seen from the results, there are peaks in the energy of the wavelet representations near the temporal moments where there are also major peaks in signal.

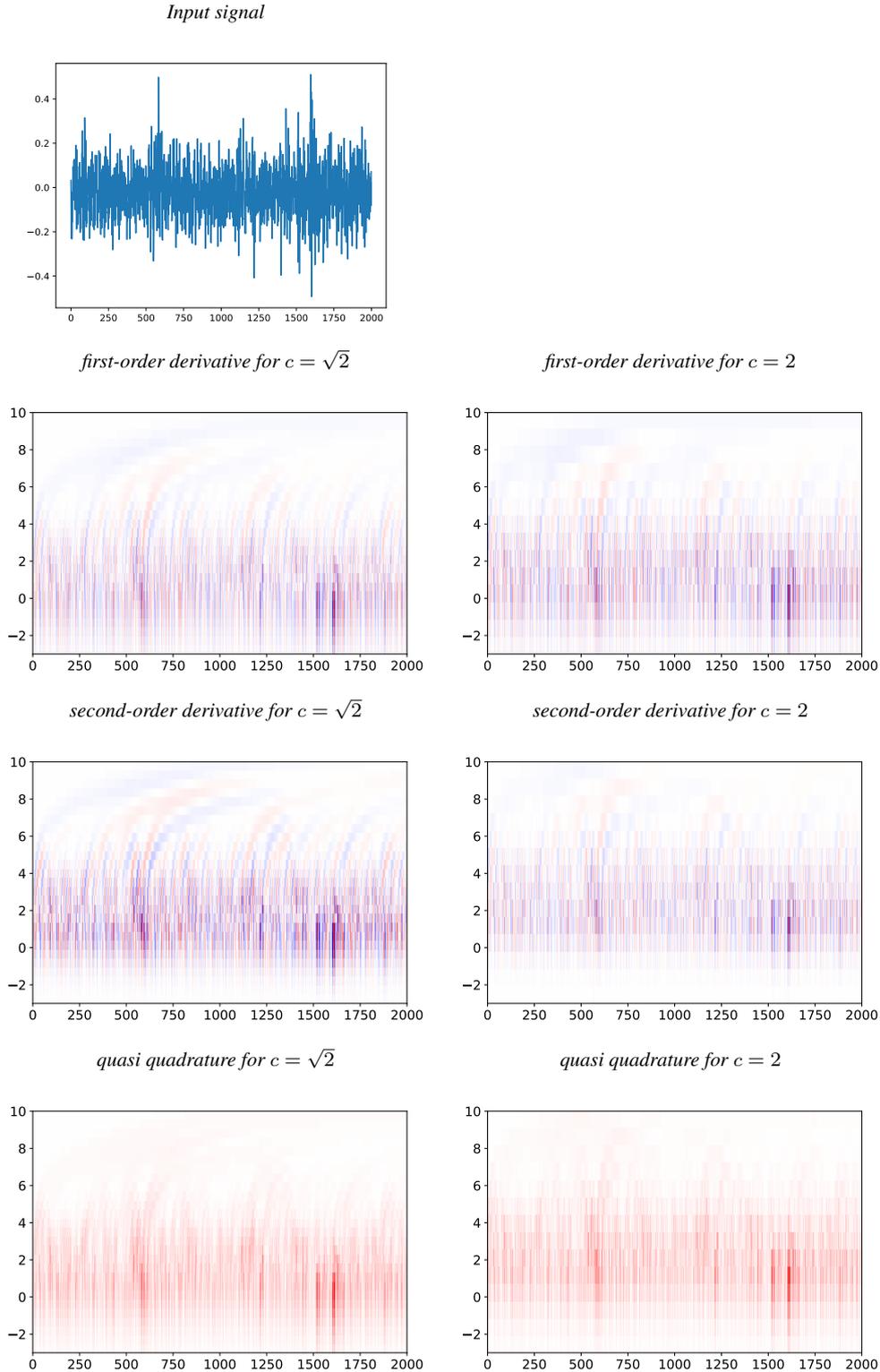
Figure 11 shows corresponding results for an auditory percussion sound of a cymbal (from <https://pixabay.com/sound-effects/search/cymbal/> item “Cymbal Crash”). As we can see, the time-causal wavelet transform largely responds over three ranges of temporal scales, with different durations of



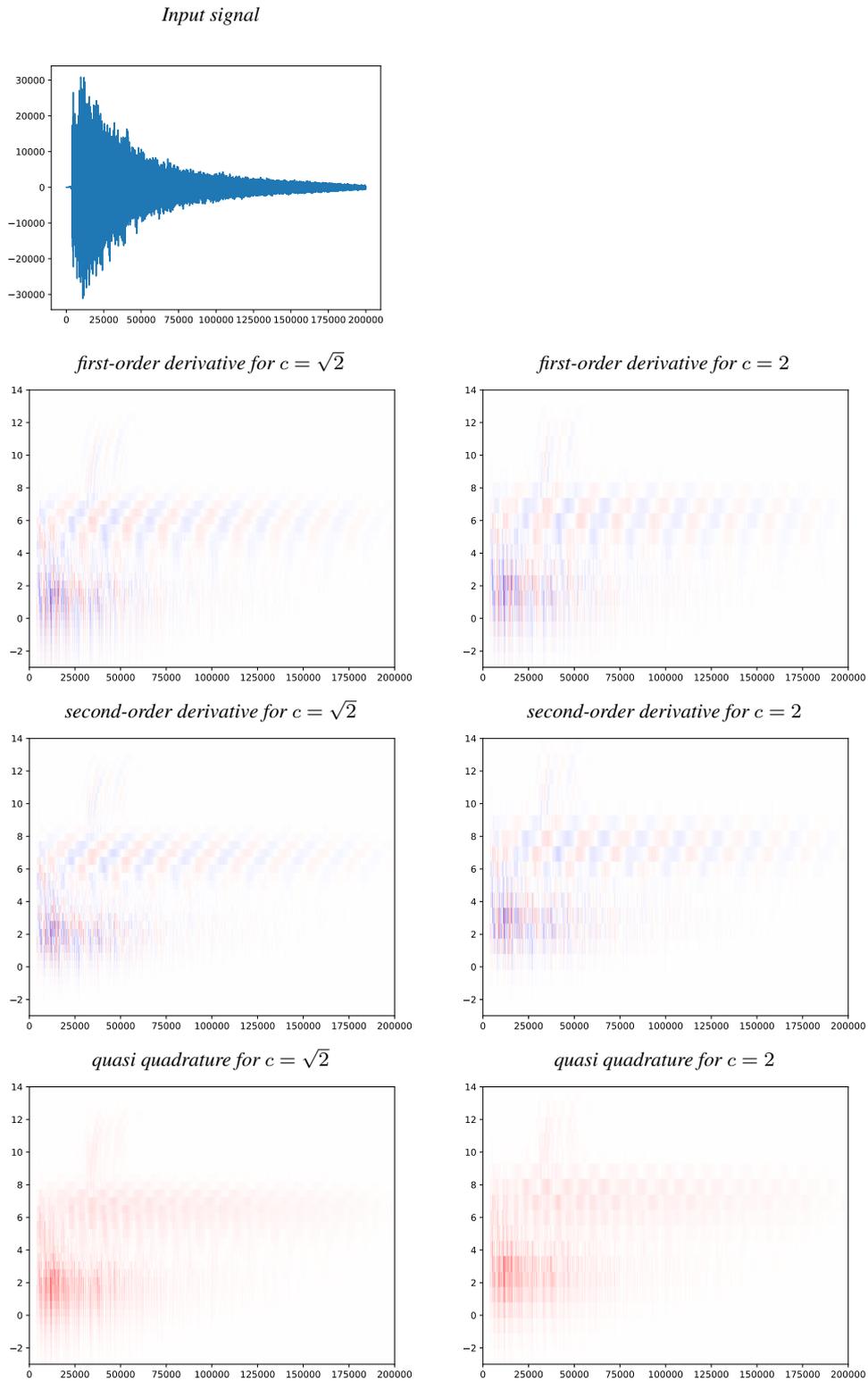
**Fig. 8** The result of computing discrete approximations of convolutions with the temporal derivatives of the time-causal limit kernel  $L_{\zeta^n}$  for orders  $n = 1$  and  $n = 2$  from a sine wave with exponentially varying frequency, here over the scale range  $\sigma = \sqrt{\tau} \in [1/8, 1024]$  and using the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization over scales of the underlying temporal derivatives of the time-causal limit kernel. In the bottom row, we have also shown the result of computing the quasi quadrature measure  $QL = \sqrt{L_{\zeta}^2 + CL_{\zeta\zeta}^2}$  for the relative weighting parameter  $C = 1/\sqrt{2}$ . As can be seen from the results, the dominant responses in the time-causal wavelet representations move from finer to coarser scales with increasing local wavelengths in the signal, in agreement with the desirable properties of a multi-scale wavelet representation, and also in agreement with the scaling properties. (The reason why the slopes of the red and blue stripes are not vertical, but oblique, is because of the different amounts of temporal delay at different temporal scales.) (Horizontal axes: time  $t \in [0, 1000]$ . Vertical axes in the time-causal wavelet representations: Effective scale =  $\log_2 \sqrt{\tau}$ .) (In the time-causal wavelet representations, red denotes positive values and blue denotes negative values.)



**Fig. 9** The result of computing discrete approximations of convolutions with the temporal derivatives of the time-causal limit kernel  $L_{\zeta^n}$  for orders  $n = 1$  and  $n = 2$  from an *electrocardiography* signal, here over the scale range  $\sigma = \sqrt{\tau} \in [1/8, 1024]$  and using the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization over scales of the underlying temporal derivatives of the time-causal limit kernel. In the bottom row, we have also shown the result of computing the quasi quadrature measure  $QL = \sqrt{L_{\zeta}^2 + C L_{\zeta\zeta}^2}$  for the relative weighting parameter  $C = 1/\sqrt{2}$ . (The reason why the slopes of the red and blue stripes are not vertical, but oblique, is because of the different amounts of temporal delay at different temporal scales.) (Horizontal axes: time  $\in [0, 3600]$ . Vertical axes in the time-causal wavelet representations: Effective scale  $= \log_2 \sqrt{\tau}$ .) (In the time-causal wavelet representations, red denotes positive values and blue denotes negative values.) (The values on the vertical axis of the signal have been shifted to make the DC level approximately zero.)



**Fig. 10** The result of computing discrete approximations of convolutions with the temporal derivatives of the time-causal limit kernel  $L_{\zeta^n}$  for orders  $n = 1$  and  $n = 2$  from a vibration signal, here over the scale range  $\sigma = \sqrt{\tau} \in [1/8, 1024]$  and using the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization over scales of the underlying temporal derivatives of the time-causal limit kernel. In the bottom row, we have also shown the result of computing the quasi quadrature measure  $QL = \sqrt{L_{\zeta}^2 + C L_{\zeta\zeta}^2}$  for the relative weighting parameter  $C = 1/\sqrt{2}$ . (Horizontal axes: time  $\in [0, 2000]$ . Vertical axes in the time-causal wavelet representations: Effective scale =  $\log_2 \sqrt{\tau}$ .) (In the time-causal wavelet representations, red denotes positive values and blue denotes negative values.) (The values on the vertical axis of the signal have been shifted to make the DC level approximately zero.)



**Fig. 11** The result of computing discrete approximations of convolutions with the temporal derivatives of the time-causal limit kernel  $L_{\zeta^n}$  for orders  $n = 1$  and  $n = 2$  from an auditory signal with a percussive sound from a cymbal, here over the scale range  $\sigma = \sqrt{\tau} \in [1/8, 16384]$  and using the scale normalization power  $\gamma = 1$  corresponding to  $L_1$ -normalization over scales of the underlying temporal derivatives of the time-causal limit kernel. In the bottom row, we have also shown the result of computing the quasi quadrature measure  $QL = \sqrt{L_{\zeta}^2 + CL_{\zeta\zeta}^2}$  for the relative weighting parameter  $C = 1/\sqrt{2}$ . (Horizontal axes: time  $\in [0, 150000]$  relative to a sampling frequency of 44.1 kHz. Vertical axes in the time-causal wavelet representations: Effective scale =  $\log_2 \sqrt{\tau}$ .) (In the time-causal wavelet representations, red denotes positive values and blue denotes negative values.) (The values on the vertical axis of the signal have been shifted to make the DC level approximately zero.)

the responses over these ranges of temporal scales, and as corresponding to vibrations in the cymbal over different ranges of frequencies. In this way, the proposed time-causal wavelet transform leads to an intuitive separation of the temporal variations in the signal over different temporal scales.

## 8 Summary and discussion

We have presented a theory for how the notions of temporal causality and temporal recursivity can be introduced in wavelet representations.

For this purpose, we have built the time-causal wavelet analysis on a special temporal smoothing kernel, denoted the time-causal limit kernel, and which constitutes the result of coupling an infinite set of truncated exponential kernels in cascade, with specially chosen time constants to obtain temporal scale covariance and self-similarity across scales. Specifically, we have in Sections 2 and 3 proposed to use temporal derivatives of this kernel as mother wavelets, with complementary normalization in a suitable  $L_p$ -norm. We have analyzed the continuous scaling properties of this representation, and also shown that the temporal derivatives of the time-causal limit kernel obey the admissibility condition for a mother wavelet. We have also argued that the temporal derivatives of the time-causal limit kernel can by theoretical arguments be regarded as canonical time-causal analogues of the derivatives of the regular Gaussian kernel, which for example, constitute the foundation for the Ricker or Mexican hat wavelet.

Since the time-causal limit kernel or its temporal derivatives do not have any compact closed-form explicit expressions, we have for this purpose specifically focused on what can be stated theoretically about these kernels, regarding their recurrence relations across scales, their scaling properties across scales, and regarding bandpass wavelet representations defined from differences between time-causal wavelet representation at adjacent discrete temporal scales.

Computationally, the proposed bandpass wavelet representations specifically allow from much more straightforward reconstruction of the original signal from the wavelet transforms. Whereas a standard (usually non-causal) wavelet representation requires reconstruction from the general inversion formula reproduced in Equation (7), when using the bandpass wavelet representation according to Equation (97), the original signal can instead be recovered from a mere summation of the bandpass wavelet channels according to Equation (101). Notably, this could also open up to real-time manipulation of local contents in the signal, by restricting the reconstruction to subsets of temporal subbands, that is over subsets of the temporal scale channels.

As we have shown in Equation (112), there is a direct relationship between the underlying computationally equivalent difference-of-time-causal-limit-kernels (DoT) kernel and the *first-order* temporal derivative of the time-causal limit

kernel. This relationship is specifically qualitatively different from the corresponding relationship for non-causal wavelets based on derivatives of the Gaussian kernel, where the corresponding bandpass kernel does according to Equation (110) instead correspond to the *second-order* derivative of the non-causal Gaussian kernel.

Then, we have in Section 4 turned to the problem of numerically approximating the time-causal limit kernel and its temporal derivatives for discrete signals. In this regard, first-order recursive filters constitute the natural analogue to the first-order integrators, that equivalently describe the temporal smoothing with truncated exponential kernels, and which constitute the main computational primitives in the continuous theory. Thereby, we have describe a canonical way to compute discrete time-causal wavelet representations in Section 5, and also shown how this representation relates to bandpass representations, as can also be defined from continuous wavelets. Notably, the proposed implementation method is fully time-recursive, and does therefore lend itself directly to real-time processing on regular signal processing architectures. Specifically, as described in Algorithm 1, the temporal smoothing can be implemented in terms of a low number (4-8) first-order recursive filters coupled in cascade. The wavelet transforms are then obtained by applying compact temporal difference operators according to Equation (123) to the temporally smoothed data, complemented by complementary scale normalization. Thereby, real-time computations can be achieved on very low-capacity computational architectures.

In Section 6 and Section 7, we have additionally analyzed how well the ideal scaling properties of the continuous theory are transferred to a discrete implementation, with special attention to the ability of the proposed time-causal wavelet representations to reveal locally dominant structures at different temporal scales in the input signals.

In summary, these characterizations clearly demonstrate how the proposed family of time-causal wavelet representations can locally capture and reflect the temporal scales of locally dominant temporal structures based on a fully time-causal analysis of the input signals.

Based on these results, and the above mentioned principled theoretical foundations, we propose that this notion of time-causal wavelet analysis could be a valuable multi-purpose tool for signal processing tasks, where streams of signals are to be processed and analyzed in real time. Beyond real-time monitoring and time-series analysis and prediction, such applications may concern purposes of real-time feedback for *e.g.* control-loop systems, and as time-causal and scale-covariant temporal basis functions in learning-based approaches. Specifically, we argue that the proposed time-causal wavelet representations could be suitable for signals that may contain local variations over a rich span of temporal scales, and more generally for analysing physical or biophysical temporal phenomena, where a fully time-causal analysis is called for to be physically realistic.

Concerning further applications of the theory, we have intentionally in this presentation focused mainly on theoretical and generic properties of the proposed families of time-causal wavelet representations, and not here aimed at developing more application-oriented experiments beyond the proof-of-concept studies in Sections 7 and 6.2. The motivation for this is that we want the results to be as timeless as possible, and not depending on specific properties of the experimental data, which may be different for different application domains, and may furthermore warrant the inclusion of additional more specialized computational mechanisms as well as integration with other signal processing approaches to build full-fledge applications.

Let us, however, emphasize that temporal derivatives of the time-causal limit kernel, as used as the mother wavelets in the here proposed time-causal wavelet representations, have been used for handling the temporal domain in models for spatio-temporal receptive fields for purposes of video analysis in classical computer vision (Jansson and Lindeberg 2018, Lindeberg 2018a, 2018b). Those results clearly demonstrate that the numerical properties of the temporal derivatives of the time-causal limit kernel can be used for computing highly useful results on real-world temporal data.

Temporal derivatives of the time-causal limit kernel have also been used for modelling the temporal component in biological visual receptive fields (Lindeberg 2021), and leading to provable covariance properties under geometric image transformations (Lindeberg 2025a). A precursor to the time-causal limit kernel, in terms of finite sets of truncated exponential kernels coupled in cascade, with logarithmically distributed temporal scale levels, has furthermore been used for modelling the temporal component in auditory receptive fields (Lindeberg and Friberg 2015). In this work, we instead make use of the temporal derivatives of the time-causal limit kernel for developing a time-causal wavelet theory, and do also describe structural relations to non-causal wavelet representations based on derivatives of the Gaussian kernel.

Closely related work of using single truncated exponential kernels for modelling neural computations in biological perception. memory processes and temporal learning has in turn been performed by Jacques (2022), Holt *et al.* (2022), Howard *et al.* (2023), Howard (2024) and Howard *et al.* (2025), with close relations to earlier work on temporal scale-space representations by Lindeberg and Fagerström (1996). Compared to the zero-order truncated exponential functions used as temporal basis functions for learning by Howard *et al.* and Holt *et al.*, the time-causal wavelets in this paper obey simplifying cascade properties over temporal scales and are also DC-balanced. For biological support of the logarithmic distribution of the temporal scale levels used both here, according to Equation (12), and in those works, see Cao *et al.* (2022).

The raw zero-order (non-differentiated) time-causal limit kernel has additionally been successfully used as a replacement of the non-causal Gaussian window function in time-

causal time-frequency analysis. For the corresponding notion of a time-causal and time-recursive analogue of Gabor filtering, corresponding to a time-causal analogue of the Morlet wavelet (Morlet *et al.* 1982a, 1982b), see Lindeberg (2025b). Compared to such time-causal time-frequency analysis, the time-causal wavelets developed in this paper require significantly less computations and lead to shorter temporal delays, and with a better ability to handle rapid transients. For more general notions of a non-causal wavelets based on Hermite transforms, and which constitute extensions of the wavelets based on first- and second-order derivatives of the Gaussian kernel to higher orders of differentiation, see Markett (1993), Glaeske (2000) and Pandey and Phukan (2020).

In this treatment, we have, however, solely focused on temporal derivatives of orders 1 and 2, and leave the problem to future work of characterizing the properties of time-causal wavelets based on temporal derivatives of the time-causal limit kernel for higher orders of temporal differentiation.

## References

- M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Applied Mathematics Series. National Bureau of Standards, US Government Printing Office, 55 edition, 1964. (Reprinted by Dover Publications).
- P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans. Communications*, 9(4):532–540, 1983.
- R. Cao, J. Bladon, S. J. Charczynski, M. E. Hasselmo, and M. W. Howard. Internally generated time in the rodent hippocampus is logarithmically compressed. *eLife*, 11:e75353, 2022.
- C. K. Chui. *An Introduction to Wavelets*, volume 1. Academic Press, 1992.
- J. L. Crowley and R. M. Stern. Fast computation of the difference of low pass transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):212–222, 1984.
- I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.
- L. Debnath and F. A. Shah. *Wavelet Transforms and Their Applications*. Springer, 2002.
- D. Donoho and I. Johnston. Wavelet decomposition of image data: Theory and experience. In K. Mardia, editor, *Statistics and images*, Abington, Oxford, 1993. Carfax Publishing Co.
- H.-J. Glaeske. Convolution structure of (generalized) Hermite transforms. *Banach Center Publications*, 53(1):113–120, 2000.
- A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4):723–736, 1984.
- C. E. Heil and D. F. Walnut. Continuous and discrete wavelet transforms. *SIAM Review*, 31(4):628–666, 1989.
- S. I. Holt, Z. Qian, and M. van der Schaar. Neural Laplace: Learning diverse classes of differential equations in the Laplace domain. In *International Conference on Machine Learning (ICML 2022)*, pages 8811–8832, 2022.
- J. W. J. Hosken. Ricker wavelets in their various guises. *First Break*, 6(1):24–33, 1988.
- M. W. Howard. Memory for time. In *Oxford Handbook of Human Memory*. Oxford University Press, 2024.
- M. W. Howard, Z. G. Esfahani, B. Le, and P. B. Sederberg. Foundations of a temporal RL. *arXiv preprint arXiv:2302.10163*, 2023.
- M. W. Howard, Z. G. Esfahani, B. Le, and P. B. Sederberg. Learning temporal relationships between symbols with Laplace neural manifolds. *Computational Brain & Behavior*, 8(2):211–232, 2025.

- G. B. Jacques, Z. Tiganj, A. Sarkar, M. Howard, and P. Sederberg. A deep convolutional neural network that is invariant to time rescaling. In *International Conference on Machine Learning (ICML 2022)*, pages 9729–9738, 2022.
- Y. Jansson and T. Lindeberg. Dynamic texture recognition using time-causal and time-recursive spatio-temporal receptive fields. *Journal of Mathematical Imaging and Vision*, 60(9):1369–1398, 2018.
- S. Karlin. *Total Positivity*. Stanford Univ. Press, 1968.
- J. J. Koenderink. Scale-time. *Biological Cybernetics*, 58:159–162, 1988.
- T. Lindeberg. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):234–254, Mar. 1990.
- T. Lindeberg. Effective scale: A natural unit for measuring scale-space lifetime. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1068–1074, Oct. 1993.
- T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):77–116, 1998a.
- T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–154, 1998b.
- T. Lindeberg. Time-causal and time-recursive spatio-temporal receptive fields. *Journal of Mathematical Imaging and Vision*, 55(1):50–88, 2016.
- T. Lindeberg. Temporal scale selection in time-causal scale space. *Journal of Mathematical Imaging and Vision*, 58(1):57–101, 2017.
- T. Lindeberg. Spatio-temporal scale selection in video data. *Journal of Mathematical Imaging and Vision*, 60(4):525–562, 2018a.
- T. Lindeberg. Dense scale selection over space, time and space-time. *SIAM Journal on Imaging Sciences*, 11(1):407–441, 2018b.
- T. Lindeberg. Normative theory of visual receptive fields. *Heliyon*, 7(1):e05897:1–20, 2021. doi: 10.1016/j.heliyon.2021.e05897.
- T. Lindeberg. A time-causal and time-recursive scale-covariant scale-space representation of temporal signals and past time. *Biological Cybernetics*, 117(1–2):21–59, 2023.
- T. Lindeberg. Unified theory for joint covariance properties under geometric image transformations for spatio-temporal receptive fields according to the generalized Gaussian derivative model for visual receptive fields. *Journal of Mathematical Imaging and Vision*, 67(4):44:1–49, 2025a.
- T. Lindeberg. A time-causal and time-recursive analogue of the Gabor transform. *IEEE Transactions on Information Theory*, 71(2):1450–1480, 2025b.
- T. Lindeberg and D. Fagerström. Scale-space with causal time direction. In *Proc. European Conf. on Computer Vision (ECCV'96)*, volume 1064 of *Springer LNCS*, pages 229–240, Cambridge, UK, Apr. 1996.
- T. Lindeberg and A. Friberg. Idealized computational models of auditory receptive fields. *PLOS One*, 10(3):e0119032:1–58, 2015.
- S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–694, 1989.
- S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- S. G. Mallat and W. L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. Information Theory*, 38(2):617–643, 1992.
- C. Markett. The product formula and convolution structure associated with the generalized Hermite polynomials. *Journal of Approximation Theory*, 73(2):199–217, 1993.
- Y. Meyer. *Wavelets and Operators: Volume 1*. Cambridge University Press, 1992.
- M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, editors. *Wavelets and Their Applications*. ISTE Ltd., 2007.
- J. Morlet, G. Arens, E. Fourgeau, and D. Giard. Wave propagation and sampling theory—Part I: Complex signal and scattering in multilayered media. *Geophysics*, 47(2):203–221, 1982a.
- J. Morlet, G. Arens, E. Fourgeau, and D. Giard. Wave propagation and sampling theory—Part II: Sampling theory and complex waves. *Geophysics*, 47(2):222–236, 1982b.
- C. P. Pandey and P. Phukan. Continuous and discrete wavelet transforms associated with Hermite transform. *International Journal of Analysis and Applications*, 18(4):531–549, 2020.
- P. Plawiak. ECG signals (1000 fragments). *Mendeley Data*, 2017. doi: 10.17632/7dybx7wyfn.3.
- N. Ricker. Wavelet functions and their polynomials. *Geophysics*, 9(3):314–323, 1944.
- W. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, editors. *Wavelets and Their Applications*. Jones and Barlett Publishers, Boston, Massachusetts, 1992.
- I. J. Schoenberg. Some analytical aspects of the problem of smoothing. In *Courant Anniversary Volume, Studies and Essays*, pages 351–370, New York, 1948.
- I. J. Schoenberg. On Pölya frequency functions. ii. Variation-diminishing integral operators of the convolution type. *Acta Sci. Math. (Szeged)*, 12:97–106, 1950.
- G. Strang. Wavelets and dilation equations: a brief introduction. *SIAM Review*, 31:614–627, 1989.
- H. H. Szu, B. A. Telfer, and A. W. Lohmann. Causal analytical wavelet transform. *Optical Engineering*, 31(9):1825–1829, 1992.
- A. Teolis. *Computational Signal Processing with Wavelets*. Birkhäuser, 1998.
- G. Turin. An introduction to matched filters. *IRE Transactions on Information Theory*, 6(3):311–329, 1960.
- J. M. Vázquez, M. Mazilu, A. Miller, and I. Galbraith. Wavelet transforms for optical pulse analysis. *Journal of the Optical Society of America A*, 22(12):2890–2899, 2005.
- J. S. Walker. *A Primer on Wavelets and their Scientific Applications*. Chapman & Hall/CRC, 1999.
- P. M. Woodward. *Probability and Information Theory, with Applications to Radar*, volume 3. Pergamon Press, 1953.