Applied and Computational Harmonic Analysis 79 (2025) 101790

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

journal homepage: www.elsevier.com/locate/acha

Full Length Article

Energy propagation in scattering convolution networks can be arbitrarily slow

Hartmut Führ, Max Getter *

Chair for Geometry and Analysis, RWTH Aachen University, Pontdriesch 10–12, Aachen, 52062, North Rhine-Westphalia, Germany

ARTICLE INFO

Communicated by Radu Balan

MSC:
42C15
68T07
42C40
42B35

Keywords:
Deep convolutional neural networks
Scattering transform
Energy propagation
Filter banks
Wavelets

ABSTRACT

We analyze energy decay for deep convolutional neural networks employed as feature extractors, including Mallat's wavelet scattering transform. For time-frequency scattering transforms based on Gabor filters, previous work has established that energy decay is exponential for arbitrary square-integrable input signals. In contrast, our main results allow proving that this is false for wavelet scattering in any dimension. Specifically, we show that the energy decay of wavelet and wavelet-like scattering transforms acting on generic square-integrable signals can be arbitrarily slow. Importantly, this slow decay behavior holds for dense subsets of $L^2(\mathbb{R}^d)$, indicating that rapid energy decay is generally an unstable property of signals. We complement these findings with positive results that allow us to infer fast (up to exponential) energy decay for generalized Sobolev spaces tailored to the frequency localization of the underlying filter bank. Both negative and positive results highlight that energy decay in scattering networks critically depends on the interplay between the respective frequency localizations of both the signal and the filters used.

1. Introduction

1.1. Motivation and related work

Up to this day, the tremendous success of convolutional neural networks (CNNs) in computer vision tasks [1] like image classification is only partially understood. This outstanding performance of CNNs is generally believed to be due to their ability to capture information at multiple scales through the use of convolutions and subsequently aggregating the information observed at these scales by means of further operations such as pooling. Consequently, CNNs are seemingly able to extract semantic content from images while discarding irrelevant information. However, the available rigorous analytical explanations for their success are still incomplete.

In his pioneering work [2], Stéphane Mallat introduced the scattering transform, a nonlinear operator that cascades convolutions followed by the modulus non-linearity. The scattering transform can be viewed as an infinitely deep and infinitely wide CNN with the modulus as non-linearity, predetermined filters, and no pooling [3]. Due to the use of hand-crafted filters based on a wavelet construction, there is no need to train the network. The windowed scattering transform was extended in [3] to allow for the use of other (optionally learned) filters that are not wavelet-generated, with potentially different filters for different layers, other non-

* Corresponding author.
E-mail addresses: fuehr@mathga.rwth-aachen.de (H. Führ), getter@mathga.rwth-aachen.de (M. Getter).

https://doi.org/10.1016/j.acha.2025.101790
Received 28 June 2024; Received in revised form 27 December 2024; Accepted 15 June 2025

Available online 20 June 2025
1063-5203/© 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

linearities than the modulus, and pooling operations between layers. The present work studies the scattering transform in a more classical fashion, with, for simplicity of exposition, identical (not necessarily wavelet-generated) filters across network layers, the modulus as non-linearity, and no pooling.

Despite the absence of a training phase and pooling, Mallat-type scattering networks are still performing comparably well to other state-of-the-art models, even after more than a decade of progress in machine learning since their invention (e.g., [4,5]). They are used in various applications such as vision and audio tasks, quantum chemistry, medicine, astronomy, and manifold learning [6–13]. In many of these, the scattering transform (or a variant thereof) is employed in the preprocessing phase of the input data in that it acts as a feature extractor, intended to improve the overall performance of the machine learning model.

Such scattering networks satisfy many of the nice properties that are generally believed to be essential in a feature extractor [2,14,3,15]. One of these properties, which is of particular importance both from a practical as well as a theoretical point of view, is the question of fast energy propagation across the network layers: Under mild assumptions on the filters, the energy of any input signal decomposes, for every $N \in \mathbb{N}$, into the aggregated energy that is contained in the first N layers of the scattering network and some energy remainder (see Proposition 1.3 for details). Overall, the scattering transform preserves the energy of an input signal if and only if the corresponding energy remainder converges to zero as $N \to \infty$. In this sense, all of the original information about the input signal is then still contained in the features generated by the scattering network. In practice, however, only finitely many features of the infinite scattering network can be computed. When restricting the network to finitely many features of the first N layers, the corresponding energy remainder is a lower bound for the loss of information. Hence, fast decay of this quantity is desirable.

Energy conservation in the above sense was first established in [2, Theorem 2.6] for wavelet-generated filter banks under a rather restrictive and technical admissibility assumption. Under less restrictive assumptions on the wavelets, but only in dimension $d = 1$, the decay of the energy remainder was shown [16, Theorem 3.1] to be at least of exponential order $\mathcal{O}(\alpha^N)$ for Sobolev functions with some *unspecified*, signal-independent $\alpha \in (0, 1)$. In recent years, quantitative results in a similar fashion were also derived for other types of filter banks. Energy decay was proven (in arbitrary dimension) to be at least of exponential order $\mathcal{O}(\alpha^N)$ for all finite-energy input signals, with some unspecified, signal-independent $\alpha \in (0, 1)$, if the filters of the underlying scattering network form a so-called uniform covering frame [17, Proposition 3.1]. In particular, this class of filter banks includes certain Weyl-Heisenberg (Gabor) frames [17, Proposition 2.3]. However, the uniform covering property requires the Fourier supports of the filters to be uniformly bounded, so that a broad range of commonly used filter banks in signal processing (e.g., wavelet-generated filter banks) is not addressed by this result.

With refined estimates, but based on an idea similar to that of [16, Theorem 3.1], energy decay was proven to be at least of exponential order $\mathcal{O}(\alpha^N)$ for Sobolev functions if the filter bank is generated by a *bandlimited* wavelet in dimension $d = 1$, with a concretely specified basis α that depends on the bandwidth of the generator [18, Theorem 2], [19, Theorem 3.1]. If the filters do not necessarily have an underlying structure as in the case of wavelet-generated filters, but instead satisfy certain mild analyticity assumptions in arbitrary dimension $d \in \mathbb{N}$, energy decay was shown to be at least of polynomial order $\mathcal{O}(N^{-m_d})$ for Sobolev functions, where $m_d \in (0, 1]$ and $m_d \to 0$ as $d \to \infty$ [18, Theorem 1].

So far, an important question has remained unanswered:

Does exponential energy decay hold for the wavelet scattering transform, globally on $L^2(\mathbb{R}^d)$?

1.2. Contributions

Our first main result, Theorem 2.2, provides a negative answer to the question of global exponential energy decay for the wavelet scattering transform. In fact, we show that energy decay can even become arbitrarily slow over $L^2(\mathbb{R}^d)$ when employing filter banks that have an underlying structure similar to wavelet-generated filter banks (including the latter). It turns out that, for given decay rate, the set of signals that do not obey this rate in terms of scattering energy decay is dense in $L^2(\mathbb{R}^d)$. Informally, energy decay can thus be considered instable over $L^2(\mathbb{R}^d)$ for these types of scattering networks.

We complement these negative findings with our second main result, Theorem 3.9, which is of positive nature. By exploiting the interplay between Fourier decay of the input signal and frequency concentration of the underlying filters of the scattering network, we provide explicit upper bounds on the convergence rates of the energy remainder for large filter-dependent subclasses of $L^2(\mathbb{R}^d)$, thereby generalizing, unifying, and partially improving the findings of the previous works [16–19].

In view of selecting wavelet-generated filters for the scattering network (as is the case in many applications), these results have two main implications:

- Fast energy propagation in wavelet scattering networks can only be guaranteed if a priori knowledge about the global Fourier decay of the input data is available, with sufficient conditions specified by Corollary 4.7.
- The set of those signals in $L^2(\mathbb{R}^d)$, for which the mixed $(\ell^1, L^2(\mathbb{R}^d))$ scattering norm considered in [2, Section 2.5] is infinite, is dense in $L^2(\mathbb{R}^d)$. Results in the spirit of [2, Theorem 2.12], which guarantee stability of the windowed scattering transform under the action of small diffeomorphisms for signals with finite mixed $(\ell^1, L^2(\mathbb{R}^d))$ scattering norm, therefore do not apply to such (in this sense) ill-behaved input signals. On the other hand, finiteness of this mixed scattering norm can be guaranteed by means of our positive results, e.g., for all $f \in H^s_{\log}(\mathbb{R}^d)$, $s > 1$, which generalizes [15, Proposition 2.4] to arbitrary dimension $d \in \mathbb{N}$ if the generating wavelet is bandlimited.

Table 1
Overview of notation used in the paper.

Notation	Meaning	Reference
d	Dimension of the domain of the input signal	
$S_{\geq x}$	Vectors in $S \subseteq \mathbb{R}^d$ that satisfy the inequality component-wise for fixed $x \in \mathbb{R}^d$	Sec. 1.4
$L^p(\mathbb{R}^d)$	Lebesgue space of complex-valued functions with exponent p	Sec. 1.4
$\ell^p(\Lambda;\mathcal{H}) = \ell^p(\mathcal{H})$	Space of p-summable maps from a countable set Λ into a Hilbert space $\mathcal{H}$	Eq. (3)
$\langle \cdot, \cdot \rangle$	Standard inner product on $\mathbb{C}^d$, $L^2(\mathbb{R}^d)$, or $\ell^2(\mathcal{H})$	Sec. 1.4
$\|\cdot\|_p$	Norm on $L^p(\mathbb{R}^d)$ or $\ell^p(\mathcal{H})$	Sec. 1.4
$\mathrm{GL}_d(\mathbb{R})$	General linear group in dimension d over the field $\mathbb{R}$	
$O_d(\mathbb{R})$	Orthogonal group in dimension d over the field $\mathbb{R}$	
I_d	Identity matrix in dimension d	
$\sigma_{\min}(A), \sigma_{\max}(A)$	The smallest (respectively largest) singular value of a matrix A	
$B_r(x)$	Open Euclidean ball with center x and radius r	
$S_{r,R}$	Closed spherical shell in $\mathbb{R}^d$ with center 0, inner radius r, and outer radius R	Sec. 1.4
C^ρ, C^ρ_A	Closed cone with tip at the origin, opening angle ρ, and orientation $A \in O_d(\mathbb{R})$	Eq. (32)
$\mathbb{1}_M$	Indicator function on the set M	
$\mathrm{vol}(M)$	d-dimensional Lebesgue-measure of the set M	
$\mathbb{S}^{d-1}$	$d-1$-dimensional sphere in d-dimensional Euclidean space	Sec. 1.4
$W_N \in \mathcal{O}(E_N)$	Bachmann-Landau asymptotic big O notation for sequences E, W	Sec. 1.4
$\mathrm{supp}(f)$	Support of a function f	Sec. 1.4
$D^p_A f$	L^p-normalized dilation of a function f by $A \in \mathrm{GL}_d(\mathbb{R})$, $p \in \{1,2\}$	Sec. 1.4
f^*	Reflection and conjugation of a function f	Sec. 1.4
$\mathcal{F}f = \widehat{f}$	Fourier transform of a function f with normalization $2\pi i$	Eq. (1)
$\mathcal{F}L^2_\omega(\mathbb{R}^d)$	Sobolev space with weight ω	Eq. (2)
$H^s(\mathbb{R}^d)$	Fractional Sobolev space of class regularity s	Sec. 1.4
$H^s_{\log}(\mathbb{R}^d)$	Fractional Log-Sobolev space of class regularity s	Sec. 1.4
$\mathfrak{F} = \{\chi\} \cup \Psi$	Semi-discrete Parseval frame with low-pass filter χ	Sec. 1.5
d_ψ	Diameter of the spectral support of a filter ψ	Eq. (39)
D_Ψ	Supremum of the diameters of the spectral supports of filters in Ψ	Definition 4.1
$D_\omega(\Psi; L^2(\mathbb{R}^d))$	Generalized Sobolev space with weight ω and filters Ψ	Eq. (40)
Y_E	Signals whose energy remainder decays at least at the rate of E	Corollary 2.6
$U[\psi]$	Single step scattering propagator with filter ψ	Eq. (4)
$U[p]$	Scattering propagator along filter path p	Eq. (5)
$S[p;\chi]$	Windowed scattering propagator along path p and with window χ	Eq. (6)
$U[P]$, $S[P;\chi]$	Collection of (windowed) scattering propagators along paths in P	Eq. (7), (8)
$S[\mathfrak{F}]$	Windowed scattering transform with filters $\mathfrak{F}$	Eq. (13)
$W_N[\Phi](f)$	Energy remainder associated with depth N, filters $\Phi \subseteq \mathfrak{F}$, and signal f	Eq. (17)
$W_N(f)$	Quantity of main interest; $W_N(f) := W_N[\Psi](f)$	Eq. (17)

1.3. Structure of the paper

We begin our exposition in Section 1.5 by reviewing the windowed scattering transform and clarifying our base assumptions on the underlying filters of the scattering network.

Our results in Section 2 concern filter banks with an inherent structure similar to those generated by wavelets. The starting point of our analysis is Lemma 2.1, which states an approximate version of super-additivity of the energy remainder for signals that are nearly separated in frequency domain. We build on this result by constructing an adversarial signal by means of an infinite series of appropriately chosen, nearly pairwise separated (in frequency domain) signals.

In Section 3, we derive upper bounds on the decay rate of the energy remainder, which are based on the interplay between Fourier decay of the input signal and the frequency concentration of the filters. Our results apply to a broad range of filter banks and their corresponding signal classes. The filters must satisfy a mild analyticity assumption, which can be seen as a refinement of the setting considered in [18, Assumption 1].

Finally, we illustrate our results in Section 4 by outlining the consequences for a certain class of filters with uniformly bounded bandwidth, and for wavelet-generated filters. In doing so, we recover some existing results from the literature, and we improve and generalize others. These results are of considerable independent interest, which is further enhanced by the contrast to the negative findings from Section 2.

1.4. Notation

This paper has a significant amount of notation, which we summarize here categorically. Table 1 provides an overview thereof.

Fix a dimension $d \in \mathbb{N}$. The sets $\mathbb{N}^d \subseteq \mathbb{N}_0^d \subseteq \mathbb{Z}^d \subseteq \mathbb{R}^d \subseteq \mathbb{C}^d$ have their usual meanings. For $S \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$, we define $S_{\geq x} := \{s \in S \mid s_k \geq x_k \text{ for all } k \in \{1,\dots,d\}\}$. We denote the Euclidean inner product of two vectors $z, w \in \mathbb{C}^d$ by $\langle z, w \rangle := \sum_{k=1}^d z_k \overline{w_k}$, with associated norm $\|\cdot\|_2$, and the respective open ball in $\mathbb{R}^d$ with center $x \in \mathbb{R}^d$ and radius $r > 0$ by $B_r(x) \subseteq \mathbb{R}^d$. Further, we write $S_{r,R} := \overline{B_R(0)} \setminus B_r(0)$ for the closed spherical shell with center at the origin, inner radius $r > 0$, and outer radius $R > 0$. By $\mathbb{1}_M$, we refer to the indicator function of a set $M \subseteq \mathbb{R}^d$. If M is Lebesgue-measurable, its measure is denoted by $\mathrm{vol}(M) \in [0, \infty]$. The sphere in $\mathbb{R}^d$ is the set $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d \mid \|x\|_2 = 1\}$.

As usual, $\mathrm{GL}_d(\mathbb{R})$ is the set of $d \times d$ invertible matrices with entries from $\mathbb{R}$, and $O_d(\mathbb{R})$ is the subgroup of orthogonal matrices. It is well known that any complex matrix A has a singular value decomposition. We denote the smallest (respectively largest) singular value of A by $\sigma_{\min}(A)$ (respectively $\sigma_{\max}(A)$). We denote the identity matrix in dimension d by I_d.

For sequences $E, W : \mathbb{N} \to \mathbb{R}$, we write $W_N \in \mathcal{O}(E_N)$ if there exist $C > 0$ and $N_0 \in \mathbb{N}$ such that $W_N \leq C \cdot E_N$ holds for all $N \in \mathbb{N}_{\geq N_0}$. We denote the support of a function $f : \mathbb{R}^d \to \mathbb{C}$ by $\operatorname{supp}(f) := \overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}}$, where the closure is taken with respect to standard Euclidean topology on $\mathbb{R}^d$. For $A \in \mathrm{GL}_d(\mathbb{R})$, $p \in \{1, 2\}$, we also define $D_A^p f := |\det(A)|^{1/p} f(A\,\cdot)$. Moreover, we set $f^* := \overline{f(-\cdot)}$. By $L^p(\mathbb{R}^d)$, for $p \in [1, \infty)$, we refer to the standard Lebesgue space, i.e., the set of Lebesgue-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ such that $\|f\|_p^p := \int_{\mathbb{R}^d} |f(x)|^p \,\mathrm{d}\, x < \infty$, where functions that agree almost everywhere (a.e.) with respect to the Lebesgue-measure are identified. With the same identification, the space $L^\infty(\mathbb{R}^d)$ is the set of Lebesgue-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ that are bounded almost everywhere, i.e., $\|f\|_\infty := \inf\{c \in \mathbb{R}_{\geq 0} \ : \ |f(x)| \leq c \text{ a.e. } x \in \mathbb{R}^d\} < \infty$. The inner product of two functions f, g from the Hilbert space $L^2(\mathbb{R}^d)$ is simply denoted by $\langle f, g \rangle := \int_{\mathbb{R}^d} f(x) \cdot \overline{g(x)} \,\mathrm{d}\, x$.

For $f \in L^1(\mathbb{R}^d)$, we define the Fourier transform of f by

$$\mathcal{F} f(\cdot) := \widehat{f}(\cdot) := \int_{\mathbb{R}^d} f(x)\, e^{-2\pi i \langle x, \cdot \rangle} \,\mathrm{d}\, x\,. \tag{1}$$

The Fourier transform extends uniquely to a unitary automorphism of $L^2(\mathbb{R}^d)$. Its inverse is the unique operator extension of the inverse Fourier transform given by

$$\mathcal{F}^{-1} f(\cdot) := \widehat{f}(-\cdot) = \int_{\mathbb{R}^d} f(x)\, e^{2\pi i \langle x, \cdot \rangle} \,\mathrm{d}\, x\,.$$

If $\omega : (0, \infty) \to (0, \infty)$ is a nondecreasing function, we define

$$\mathcal{F}L^2_\omega(\mathbb{R}^d) := \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; \omega(\|\cdot\|_2)\widehat{f} \in L^2(\mathbb{R}^d) \right\}. \tag{2}$$

For $s > 0$, setting $\omega_s(t) := (1 + t^2)^{\frac{s}{2}}$, the fractional Sobolev space of class regularity s is $H^s(\mathbb{R}^d) := \mathcal{F}L^2_{\omega_s}(\mathbb{R}^d)$. We define the s-fractional Log-Sobolev space by $H^s_{\log}(\mathbb{R}^d) := \mathcal{F}L^2_{\omega_{s,\log}}(\mathbb{R}^d)$, where $\omega_{s,\log}(t) := \ln^s(e + t)$. Furthermore, let $\mathcal{S}(\mathbb{R}^d)$ denote the Schwartz space on $\mathbb{R}^d$. For a Hilbert space $\mathcal{H}$, a countable set Λ, and $p \in [1, \infty)$, we define

$$\ell^p(\mathcal{H}) := \ell^p(\Lambda; \mathcal{H}) := \left\{ f = (f_\lambda)_{\lambda \in \Lambda} \in \mathcal{H}^\Lambda \;\middle|\; \sum_{\lambda \in \Lambda} \|f_\lambda\|_{\mathcal{H}}^p < \infty \right\}. \tag{3}$$

If $p = 2$, this is a Hilbert space, equipped with the inner product given by $\langle f, g \rangle := \sum_{\lambda \in \Lambda} \langle f_\lambda, g_\lambda \rangle$ for $f, g \in \ell^2(\mathcal{H})$ (with unconditional convergence of the series).

1.5. A brief review of the windowed scattering transform

A scattering network computes cascades of convolutions followed by the application of the modulus as non-linearity, with no pooling between the layers. Given any $\psi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, we denote by

$$U[\psi] : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \quad f \mapsto |f * \psi| \tag{4}$$

the *single step scattering propagator* associated to filtering with the function ψ (which in this context is thus commonly called a *filter*). Moreover, we refer to an ordered sequence $p = (\psi_1, \ldots, \psi_N) \in (L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d))^N$ as a *path* of length $N \in \mathbb{N}$. Iterating the single step scattering propagators along this path yields the *scattering propagator*

$$U[p] : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \quad f \mapsto U[\psi_N] \ldots U[\psi_1] f = |\ldots|f * \psi_1| \ldots * \psi_N|. \tag{5}$$

For notational convenience, we define the scattering propagator along the empty path e (the unique path of length zero) to be the identity operator $U[e] := \mathrm{id}_{L^2(\mathbb{R}^d)}$.

A designated function $\chi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ induces a *windowed scattering propagator* along a path p of length $N \in \mathbb{N}$ and window χ (referred to as output-generating filter), given by

$$S[p; \chi] : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \quad f \mapsto (U[p] f) * \chi = |\ldots|f * \psi_1| \ldots * \psi_N| * \chi. \tag{6}$$

By Young's convolution inequality, the operators $U[p]$ and $S[p; \chi]$ are well-defined.

Our main objective is to study the properties of the operators (Fig. 1)

$$U[P] : L^2(\mathbb{R}^d) \to \ell^2(P; L^2(\mathbb{R}^d)), \quad f \mapsto (U[p] f)_{p \in P} \tag{7}$$

and

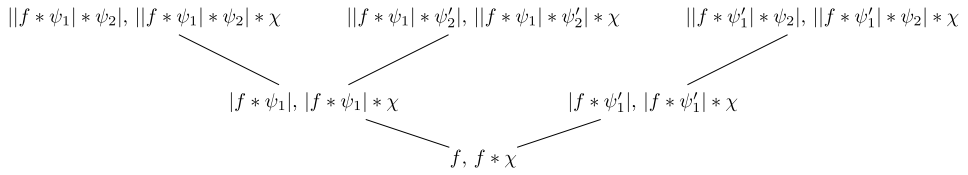


Fig. 1. Illustration of the operators $U[P], S[P;\chi]$ for the path set $P = \{e, \psi_1, \psi_1', (\psi_1,\psi_2), (\psi_1,\psi_2'), (\psi_1',\psi_2)\}$ and output-generating filter χ.

$$S[P;\chi] : L^2(\mathbb{R}^d) \to \ell^2(P; L^2(\mathbb{R}^d)), \quad f \mapsto (S[p;\chi]f)_{p\in P} \tag{8}$$

for certain path sets P. In particular, this requires the operators to be well-defined in the sense that, for any $f \in L^2(\mathbb{R}^d)$, the series $\sum_{p\in P} \|U[p]f\|_2^2$ converges unconditionally. In this respect, let us now state our main requirement on the filters, which are the basic building blocks of the windowed scattering transform.

Throughout this paper, we assume that $\mathfrak{F} := \{\chi\} \cup \Psi \subset L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ is a countably infinite collection of functions so that the corresponding filter transform

$$L^2(\mathbb{R}^d) \to \ell^2(\mathfrak{F}; L^2(\mathbb{R}^d)), \ f \mapsto (f*\chi) \cup (f*\psi)_{\psi\in\Psi} \tag{9}$$

is an isometry. This is precisely the case if $\mathfrak{F}$ forms a semi-discrete Parseval frame [20, Section 5.1.5], i.e., if for all $f \in L^2(\mathbb{R}^d)$,

$$\|f*\chi\|_2^2 + \sum_{\psi\in\Psi} \|f*\psi\|_2^2 = \|f\|_2^2. \tag{10}$$

We frequently use the fact (which follows directly from Parseval's theorem and the convolution theorem) that $\mathfrak{F}$ forms a semi-discrete Parseval frame if and only if $\mathfrak{F}$ satisfies the Littlewood-Paley condition

$$|\widehat{\chi}(\xi)|^2 + \sum_{\psi\in\Psi} |\widehat{\psi}(\xi)|^2 = 1 \quad \text{a.e. } \xi \in \mathbb{R}^d. \tag{11}$$

If any of the latter equivalent conditions holds, then every $f \in L^2(\mathbb{R}^d)$ admits a decomposition

$$f = f*\chi*\chi^* + \sum_{\psi\in\Psi} f*\psi*\psi^*. \tag{12}$$

Remark 1.1. To simplify the notation, we do not assign the filters Ψ an explicit (countably infinite) index set, unless we work in a more concrete setting. This precludes the use of multiple occurrences of the same filter, which is often allowed in the definition of the scattering transform. In principle, this can be accommodated, primarily at notational cost, by extending the subsequent discussion to multisets. To us the benefit of such an extension seems quite limited, which is why we have refrained from making the necessary adjustments.

We mainly consider the operators $U[P]$ and $S[P;\chi]$ on nonempty subsets P of the set of all finite length scattering paths $\mathcal{P}_\Psi := \bigcup_{N=0}^\infty \Psi^N$, where we set $\Psi^0 := \{e\}$. The *windowed scattering transform associated with the filters* $\mathfrak{F}$ is the operator

$$S[\mathfrak{F}] := S[\mathcal{P}_\Psi;\chi] : L^2(\mathbb{R}^d) \to \ell^2(\mathcal{P}_\Psi; L^2(\mathbb{R}^d)), \quad f \mapsto (S[p;\chi]f)_{p\in\mathcal{P}_\Psi}. \tag{13}$$

In the interest of generality, we do not make further assumptions on the Parseval frame $\mathfrak{F}$ for now. Even in this very general setting, the windowed scattering transform $S[\mathfrak{F}]$ (Fig. 2) is always a well-defined operator into the Hilbert space $\ell^2(\mathcal{P}_\Psi; L^2(\mathbb{R}^d))$. We address this in more detail in Proposition 1.3.

Typically, the filters are concentrated in different areas in frequency domain, with χ taking on the special role of a low-pass filter and the remaining filters Ψ being of high-pass nature. For the sake of concreteness and later use, let us give one particular example of an admissible collection $\mathfrak{F} = \{\chi\} \cup \Psi$ in dimension $d = 1$, where in this case Ψ is generated from a bandlimited wavelet.

Example 1.2. Let $\psi, \phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ satisfy $\operatorname{supp}(\widehat{\psi}) \subseteq [\frac{1}{2}, 2]$ and

$$|\widehat{\phi}(\xi)|^2 + \sum_{j=1}^\infty \left(|\widehat{\psi}(2^{-j}\cdot\xi)|^2 + |\widehat{\psi}(-2^{-j}\cdot\xi)|^2\right) = 1, \quad \text{a.e. } \xi \in \mathbb{R}.$$

Such a pair of functions arises, e.g., from the construction of an orthogonal wavelet basis, with father wavelet (scaling function) ϕ and mother wavelet ψ. Then, $\mathfrak{F} := \{\chi\} \cup \{\psi_j \mid j \in \mathbb{Z}\setminus\{0\}\}$ is a semi-discrete Parseval frame, where $\chi := \phi$ and

$$\psi_j := \begin{cases} D^1_{2^j}\psi = 2^j\psi(2^j\,\cdot) & \text{for } j \in \mathbb{N} \\ D^1_{2^{|j|}}\psi(-\,\cdot) = 2^{|j|}\psi(-2^{|j|}\,\cdot) & \text{for } j \in \mathbb{Z}_{<0} \end{cases}.$$

In this case, the filter transform is the associated wavelet transform.

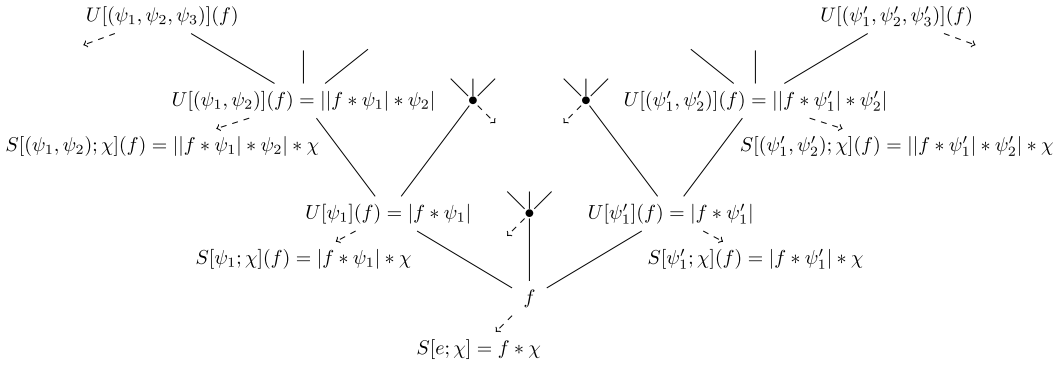


Fig. 2. Indication of the architecture of a scattering network as described above, with filters $\psi_N, \psi_N' \in \Psi$ corresponding to the Nth network layer, $N \in \{1,2,3\}$. The function χ is the output-generating filter, which is (in our setup) the same across all layers. The residual energy $W_N(f)$ is the aggregated energy of the Nth horizontal layer of the network, e.g., $W_2(f) = W_2[\Psi](f) = \big\| |f * \psi_1| * \psi_2 \big\|_2^2 + \| |f * \psi_1'| * \psi_2' \|_2^2 + \cdots$.

The following proposition, which was first stated in (the proof of) [2, Theorem 2.6] and later generalized in [3, Proposition 1] and [18, Proposition 1], forms the basis for all further analysis in this paper.

Proposition 1.3. *Every $f \in L^2(\mathbb{R}^d)$ satisfies the energy decomposition, for all $N \in \mathbb{N}_0$,*

$$\sum_{n=0}^{N-1} \|S[\Psi^n;\chi]f\|^2_{\ell^2(\Psi^n;L^2(\mathbb{R}^d))} + \left\|U[\Psi^N]f\right\|^2_{\ell^2(\Psi^N;L^2(\mathbb{R}^d))} = \|f\|_2^2 . \tag{14}$$

Consequently,

$$\|S[\mathfrak{F}]f\|_{\ell^2(\mathcal{P}_\Psi;L^2(\mathbb{R}^d))} \le \|f\|_2 ,$$

which ensures the well-definedness of the operator $S[\mathfrak{F}] : L^2(\mathbb{R}^d) \to \ell^2(\mathcal{P}_\Psi; L^2(\mathbb{R}^d))$.

Moreover,

$$\lim_{N\to\infty} \left\|U[\Psi^N]f\right\|^2_{\ell^2(\Psi^N;L^2(\mathbb{R}^d))} = 0 \tag{15}$$

holds if and only if

$$\|S[\mathfrak{F}]f\|^2_{\ell^2(\mathcal{P}_\Psi;L^2(\mathbb{R}^d))} = \|f\|_2^2 . \tag{16}$$

The energy decomposition identity (14) quantifies the loss of information about the original input signal f if the scattering network is restricted to its first N layers. Motivated by this important relationship, let us define the quantity, for any $\emptyset \neq \Phi \subseteq \mathfrak{F}$ and $N \in \mathbb{N}_0$,

$$W_N[\Phi](f) := \left\|U[\Phi^N]f\right\|^2_{\ell^2(\Phi^N;L^2(\mathbb{R}^d))} = \sum_{p\in\Phi^N} \|U[p](f)\|_2^2 . \tag{17}$$

Our main objective in this paper is to study the asymptotic behavior of the quantity $W_N[\Psi](f)$ as $N \to \infty$, which, as we show, distinctively depends on the interplay between the choice of filters Ψ and the input signal f. If there is no ambiguity about the underlying filters Ψ, we sometimes only write $W_N(f)$ to refer to $W_N[\Psi](f)$.

As already pointed out in Section 1.1, there are several sufficient criteria for $\mathfrak{F}$ available that guarantee the asymptotic energy decay (15). More precisely, [16, Theorem 3.1] and [18, Theorem 1, Theorem 2] state an upper bound on $W_N[\Psi](f)$, for all $f \in L^2(\mathbb{R}^d)$ and $N \in \mathbb{N}_{\ge 2}$, which is of the type

$$W_N[\Psi](f) \le \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot K_N(\xi)\, \mathrm{d}\xi \tag{18}$$

for a family of integral kernels $(K_N)_{N\in\mathbb{N}_{\ge 2}}$ (independent of f) that satisfy the following properties:

- For all $N \in \mathbb{N}_{\ge 2}$, K_N is continuous.
- For all $N \in \mathbb{N}_{\ge 2}$, $0 \le K_{N+1} \le K_N \le 1$.
- $\lim_{N\to\infty} K_N(\xi) = 0$ holds pointwise for all $\xi \in \mathbb{R}^d$.

By Dini's theorem [21, Theorem 7.3], this upper integral bound entails, for all $f \in L^2(\mathbb{R}^d)$, the convergence (15) (cf. [18, Proposition 2]). In particular, the scattering transform $S[\mathfrak{F}]$ is norm-preserving in these settings by Proposition 1.3. Specifically the assumptions

of [18, Theorem 1] allow for some flexibility in the choice of the filters while guaranteeing norm-preservation. Clearly, this implies that the single signal that produces zero-features in every layer by means of the scattering transform is the trivial null-signal $f = 0$.

In order to produce many informative (in the sense of nontrivial) features by means of the scattering transform, one typically chooses an output-generating filter χ that is of low-pass type. By the Littlewood-Paley condition (11), the remaining filters Ψ are automatically of high-pass nature. Let us briefly illustrate how informativeness of $S[\mathfrak{F}]$ can be achieved by imposing a low-pass condition on χ: Let $0 \neq g, \chi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and suppose that $\widehat{\chi}(0) \neq 0$. By continuity of $\widehat{\chi}$ and $\mathcal{F}(|g|)$, and since $\mathcal{F}(|g|)(0) = \|g\|_1 > 0$, there exist $\varepsilon, \delta > 0$ such that $|\widehat{\chi}| \cdot |\mathcal{F}(|g|)| \geq \varepsilon$ on $B_\delta(0)$. Thus, we have

$$\| |g| * \chi \|_2^2 = \left\| \mathcal{F}(|g|) \cdot \widehat{\chi} \right\|_2^2 \geq \int\limits_{B_\delta(0)} |\mathcal{F}(|g|)(\xi)|^2 \cdot |\widehat{\chi}(\xi)|^2 \, \mathrm{d}\xi \geq \varepsilon^2 \cdot \mathrm{vol}(B_\delta(0)) > 0.$$

Hence, any feature of this type is nontrivial. The assumption that $\widehat{\chi}(0) \neq 0$ is of vital importance for the argument, as is shown in [18, Appendix A] by the example of a scattering network (feature extractor) with nontrivial kernel.

For the remainder of this section, we collect some basic auxiliary statements about the windowed scattering transform that will prove useful later. We include their proofs for self-containment.

Lemma 1.4. *Let $N \in \mathbb{N}$. Then $U[\mathfrak{F}^N] : L^2(\mathbb{R}^d) \to \ell^2(\mathfrak{F}; L^2(\mathbb{R}^d))$ is well-defined and norm-preserving, i.e., for all $f \in L^2(\mathbb{R}^d)$,*

$$W_N[\mathfrak{F}](f) = \sum_{p \in \mathfrak{F}^N} \|U[p]f\|_2^2 = \|f\|_2^2 .$$

Proof. This is easily shown by induction: The base case $N = 1$ is exactly one of the characterizations of a semi-discrete Parseval frame, hence true for $\mathfrak{F}$. The induction step is established as follows. Suppose that the statement is true for some $N \in \mathbb{N}$. Then, for all $f \in L^2(\mathbb{R}^d)$, we have

$$W_{N+1}[\mathfrak{F}](f) = \sum_{\psi \in \mathfrak{F}} W_N[\mathfrak{F}](|f * \psi|) = \sum_{\psi \in \mathfrak{F}} \| |f * \psi| \|_2^2 = \sum_{\psi \in \mathfrak{F}} \|f * \psi\|_2^2 = \|f\|_2^2 . \quad \square$$

Despite its simplicity and brevity, the argument illustrates how the underlying inductive structure of the scattering network can be exploited in proofs. The following Proposition (cf. [2, Proposition 2.5]) states the non-expansiveness of the operators U and S, which is also due to their inductive definition. Considering the desirable properties of the scattering transform as a feature extractor, this property can be considered as stability against additive noise.

Proposition 1.5. *Let $\emptyset \neq P \subseteq \Psi^N$ for some $N \in \mathbb{N}_0$. Then, for all $f, g \in L^2(\mathbb{R}^d)$,*

$$\|S[P;\chi]f - S[P;\chi]g\|_{\ell^2(P;L^2(\mathbb{R}^d))} \leq \|U[P]f - U[P]g\|_{\ell^2(P;L^2(\mathbb{R}^d))} \leq \|f - g\|_2 .$$

Furthermore, if $S[\mathfrak{F}] = S[\mathcal{P}_\Psi;\chi]$ is norm-preserving, then it is also nonexpansive.

Proof. By the Littlewood-Paley condition (11), we have $\left\|\widehat{\chi}\right\|_\infty \leq 1$. Thus, for all $p \in P$, and all $f, g \in L^2(\mathbb{R}^d)$,

$$\|S[p;\chi]f - S[p;\chi]g\|_2 = \|(U[p]f - U[p]g) * \chi\|_2 \leq \|U[p]f - U[p]g\|_2 ,$$

which proves the first inequality.

Observe that it suffices to prove the second inequality for the case $P = \Psi^N$, which we do by induction on $N \in \mathbb{N}_0$. The base case $N = 0$ is trivial. For the purpose of the induction step, suppose that the hypothesis is true for some $N \in \mathbb{N}_0$. By the reverse triangle inequality, we have, for all $f, g \in L^2(\mathbb{R}^d)$,

$$\begin{aligned}
&\| |U[\Psi^{N+1}]f - U[\Psi^{N+1}]g| \|_{\ell^2(\Psi^{N+1};L^2(\mathbb{R}^d))}^2 \\
&\quad = \sum_{p \in \Psi^N} \sum_{\psi \in \Psi} \| |(U[p]f) * \psi| - |(U[p]g) * \psi| \|_2^2 \\
&\quad \leq \sum_{p \in \Psi^N} \sum_{\psi \in \Psi} \| |(U[p]f - U[p]g) * \psi| \|_2^2 \\
&\quad = \sum_{p \in \Psi^N} \sum_{\psi \in \Psi} \|(U[p]f - U[p]g) * \psi\|_2^2 \\
&\quad = \sum_{p \in \Psi^N} \left(\|U[p]f - U[p]g\|_2^2 - \|S[p;\chi]f - S[p;\chi]g\|_2^2 \right) \\
&\quad = \left\| U[\Psi^N]f - U[\Psi^N]g \right\|_{\ell^2(\Psi^N;L^2(\mathbb{R}^d))}^2 - \left\| S[\Psi^N;\chi]f - S[\Psi^N;\chi]g \right\|_{\ell^2(\Psi^N;L^2(\mathbb{R}^d))}^2 ,
\end{aligned}$$

where the second last step is due to $\mathfrak{F}$ being a frame. Applying the induction hypothesis to the right-hand side of the latter inequality concludes this part of the proof.

Finally, if $\mathcal{S}[\mathfrak{F}]$ is norm-preserving, the non-expansiveness of $\mathcal{S}[\mathfrak{F}]$ follows directly from the above estimates (which were derived independent of the induction hypothesis) using a telescoping series argument. In fact, by Proposition 1.3, we have $\lim_{N\to\infty} W_N[\Psi](h) = 0$ for all $h \in L^2(\mathbb{R}^d)$, which entails that, for all $f, g \in L^2(\mathbb{R}^d)$,

$$\lim_{N\to\infty} \left\| U[\Psi^N]f - U[\Psi^N]g \right\|^2_{\ell^2(\Psi^N; L^2(\mathbb{R}^d))} = 0.$$

Consequently,

$$\begin{aligned}
&||\mathcal{S}[\mathfrak{F}]f - \mathcal{S}[\mathfrak{F}]g||^2_{\ell^2(\mathcal{P}_\Psi; L^2(\mathbb{R}^d))} \\
&\quad = \sum_{N=0}^{\infty} \left\| S[\Psi^N; \chi]f - S[\Psi^N; \chi]g \right\|^2_{\ell^2(\Psi^N; L^2(\mathbb{R}^d))} \\
&\quad \leq \sum_{N=0}^{\infty} \left(\left\| U[\Psi^N]f - U[\Psi^N]g \right\|^2_{\ell^2(\Psi^N; L^2(\mathbb{R}^d))} - \left\| U[\Psi^{N+1}]f - U[\Psi^{N+1}]g \right\|^2_{\ell^2(\Psi^{N+1}; L^2(\mathbb{R}^d))} \right) \\
&\quad = \|f - g\|_2^2. \quad \square
\end{aligned}$$

We exploit the non-expansiveness of U to derive some useful estimates concerning the energy remainder.

Lemma 1.6. *Let $\emptyset \neq \Phi \subseteq \Psi$, and let $N \in \mathbb{N}$. Then, we have, for all $f, g \in L^2(\mathbb{R}^d)$,*

a) $|W_N[\Phi](f) - W_N[\Phi](g)| \leq \sqrt{2} \cdot \|f - g\|_2 \cdot \sqrt{W_N[\Phi](f) + W_N[\Phi](g)}$,
b) $W_N[\Phi](f) \geq \frac{1}{2} \cdot W_N[\Phi](g) - \|f - g\|_2^2$.

Proof. Let $f, g \in L^2(\mathbb{R}^d)$. Applying the reverse triangle inequality yields, for any $p \in \Phi^N$,

$$\begin{aligned}
| \, \|U[p]f\|_2^2 - \|U[p]g\|_2^2 \, | &= | \, \|U[p]f\|_2 - \|U[p]g\|_2 \, | \cdot (\|U[p]f\|_2 + \|U[p]g\|_2) \\
&\leq \|U[p]f - U[p]g\|_2 \cdot (\|U[p]f\|_2 + \|U[p]g\|_2).
\end{aligned}$$

Together with the Cauchy-Schwarz inequality and the non-expansiveness of $U[\Phi^N]$, we obtain

$$\begin{aligned}
|W_N[\Phi](f) - W_N[\Phi](g)|^2 &= \left| \sum_{p\in\Phi^N} \|U[p]f\|_2^2 - \|U[p]g\|_2^2 \right|^2 \\
&\leq \left(\sum_{p\in\Phi^N} \|U[p]f - U[p]g\|_2 \cdot \left(\|U[p]f\|_2 + \|U[p]g\|_2 \right) \right)^2 \\
&\leq \left(\sum_{p\in\Phi^N} \|U[p]f - U[p]g\|_2^2 \right) \cdot \left(\sum_{p\in\Phi^N} \left(\|U[p]f\|_2 + \|U[p]g\|_2 \right)^2 \right) \\
&\leq \|f - g\|_2^2 \cdot \sum_{p\in\Phi^N} \left(2\,\|U[p]f\|_2^2 + 2\,\|U[p]g\|_2^2 \right) \\
&= 2\,\|f - g\|_2^2 \cdot (W_N[\Phi](f) + W_N[\Phi](g)).
\end{aligned}$$

Taking square roots on both sides of the latter inequality proves a).

We again use the non-expansiveness of $U[\Phi^N]$ and find that

$$\begin{aligned}
W_N[\Phi](g) &= \left\| U[\Phi^N](g) - U[\Phi^N](f) + U[\Phi^N](f) \right\|^2_{\ell^2(\Phi^N; L^2(\mathbb{R}^d))} \\
&\leq \left(\left\| U[\Phi^N](g) - U[\Phi^N](f) \right\|_{\ell^2(\Phi^N; L^2(\mathbb{R}^d))} + \left\| U[\Phi^N](f) \right\|_{\ell^2(\Phi^N; L^2(\mathbb{R}^d))} \right)^2 \\
&\leq \left(\|g - f\|_2 + \left\| U[\Phi^N](f) \right\|_{\ell^2(\Phi^N; L^2(\mathbb{R}^d))} \right)^2 \\
&\leq 2\,\|f - g\|_2^2 + 2 \left\| U[\Phi^N](f) \right\|^2_{\ell^2(\Phi^N; L^2(\mathbb{R}^d))} \\
&= 2\,\|f - g\|_2^2 + 2W_N[\Phi](f).
\end{aligned}$$

By rearranging, we conclude the proof of b). $\square$

It is now easy to derive the continuity of $W_N[\Phi]$, which we use to prove the density result, Proposition 2.3, in Section 2.

Corollary 1.7. *Let $\emptyset \neq \Phi \subseteq \Psi$, and let $N \in \mathbb{N}$. Then, $W_N[\Phi] : L^2(\mathbb{R}^d) \to \mathbb{R}$ is continuous.*

Proof. As the composition of continuous maps, $W_N[\Phi] = \|\cdot\|^2_{\ell^2(\Phi^N; L^2(\mathbb{R}^d))} \circ U[\Phi^N]$ is continuous. Alternatively, one can use the estimate from Lemma 1.6a) to derive the continuity of $W_N[\Phi]$ more explicitly: Continuity at $f = 0$ is obvious, since we have, for all $g \in L^2(\mathbb{R}^d)$, $W_N[\Phi](g) \leq \|g\|_2^2$. Now, fix $\varepsilon > 0$ and $f \in L^2(\mathbb{R}^d) \setminus \{0\}$. Define

$$\delta = \delta(f, \varepsilon) := \min\left\{ \frac{\varepsilon}{\sqrt{10}\,\|f\|_2}, \|f\|_2 \right\} > 0.$$

Applying Lemma 1.6a) to $g \in L^2(\mathbb{R}^d)$ with $\|f - g\|_2 < \delta$ gives

$$\begin{aligned} |W_N[\Phi](f) - W_N[\Phi](g)|^2 &\leq 2\,\|f-g\|_2^2 \cdot (W_N[\Phi](f) + W_N[\Phi](g)) \\ &\leq 2\delta^2 \cdot (\|f\|_2^2 + \|g\|_2^2) \\ &\leq 2\delta^2 \cdot (\|f\|_2^2 + (\|f\|_2 + \delta)^2) \leq 10\delta^2 \cdot \|f\|_2^2 \leq \varepsilon^2. \quad \square \end{aligned}$$

We conclude this section with a lemma describing the interplay of the scattering propagator with dilations, which is similar to covariance (cf. [2, Equation (20)]). Despite its elementary proof, this property forms the basis for our main result in Section 2.

Lemma 1.8. *Let $N \in \mathbb{N}$, let $p \in (L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d))^N$, and let $A \in \mathrm{GL}_d(\mathbb{R})$. Then, for all $f \in L^2(\mathbb{R}^d)$,*

$$U[p] D^1_A f = D^1_A U[D^1_{A^{-1}} p] f,$$

where $D^1_{A^{-1}} p := (D^1_{A^{-1}} p_1, \ldots, D^1_{A^{-1}} p_N)$. Moreover, if $\emptyset \neq \Phi \subseteq \mathfrak{F}$, then we have, for all $f \in L^2(\mathbb{R}^d)$,

$$W_N[\Phi](D^2_A f) = W_N[D^1_{A^{-1}} \Phi](f).$$

Proof. Let $f \in L^2(\mathbb{R}^d)$, and let $g \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. A direct computation shows that

$$(D^1_A f) * g = D^1_A (f * (D^1_{A^{-1}} g)).$$

The modulus operator

$$M : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \quad h \mapsto |h|$$

commutes with D^1_A, which proves $U[g] D^1_A f = D^1_A U[D^1_{A^{-1}} g] f$. Iterating this relation along the path p concludes the first identity.

Concerning the second, we note that $D^1_A = |\det(A)|^{\frac{1}{2}} \cdot D^2_A$, which gives $U[p] D^2_A = D^2_A U[D^1_{A^{-1}} p]$. Since D^2_A is unitary on $L^2(\mathbb{R}^d)$, summing over all paths in Φ^N entails the second identity. $\square$

2. Slow scattering propagation

In this section, we show that energy propagation can be arbitrarily slow in scattering networks that employ filters $\mathfrak{F} = \{\chi\} \cup \Psi$ with an underlying structure similar to those generated by wavelets.

As some of the details of this section are quite technical, we begin by illustrating the conceptual idea behind them. Let us consider the particularly simple setting from Example 1.2 again, i.e., suppose that the high-pass filters $\Psi = \{\psi_j \mid j \in \mathbb{Z} \setminus \{0\}\}$ are generated by dilations of $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, $\mathrm{supp}(\widehat{\psi}) \subseteq [\frac{1}{2}, 2]$, according to

$$\psi_j := \begin{cases} D^1_{2^j} \psi = 2^j \psi(2^j \cdot) & \text{for } j \in \mathbb{N} \\ D^1_{2^{|j|}} \psi(-\cdot) = 2^{|j|} \psi(-2^{|j|} \cdot) & \text{for } j \in \mathbb{Z}_{<0} \end{cases}.$$

We want to show that for any nonincreasing null-sequence $E \in \mathbb{R}^{\mathbb{N}}_{>0}$, there exists a function $f_E \in L^2(\mathbb{R})$ that satisfies the following conditions:

a) We have $\|f_E\|_2^2 = E_1$.
b) For all $N \in \mathbb{N}$, we have that $W_N[\Psi](f_E) \geq \frac{1}{2} \cdot E_N$.

Proof sketch. Define

$$(a_k)_{k \in \mathbb{N}} := \left(\sqrt{E_k - E_{k+1}} \right)_{k \in \mathbb{N}},$$

and note that a telescoping argument gives

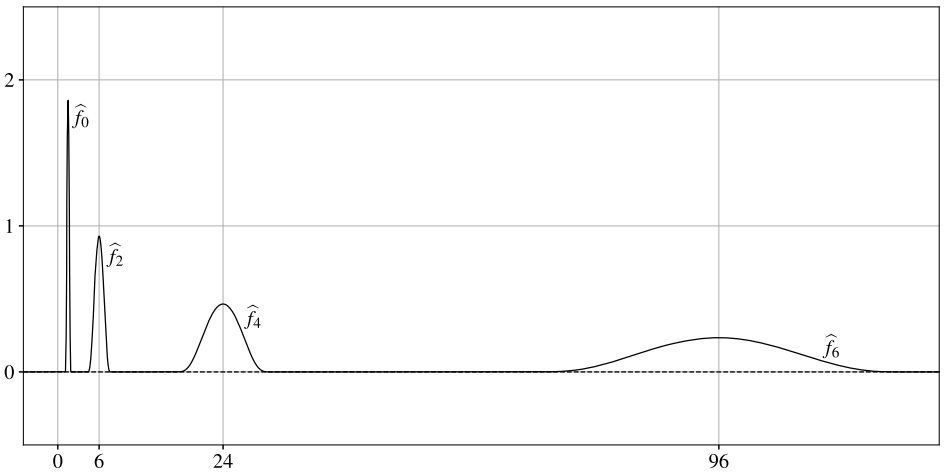


Fig. 3. Fourier transform of a sample function f_0, and of its dilations f_2, f_4, f_6, illustrating the frequency separation between the individual components of f_E resulting from the dilation. Frequency separation is key to our construction of adversarial signals. In this illustration, $f_0 = \frac{h}{\|h\|_2}$, and $\widehat{h}(\xi) = \mathbb{1}_{(1,2)} \cdot \exp\left(-\frac{4}{1-(3-2\xi)^2}\right)$.

$$\lim_{N\to\infty} \sum_{k=1}^{N} a_k^2 = \lim_{N\to\infty} E_1 - E_{N+1} = E_1.$$

Fix any $f_0 \in L^2(\mathbb{R})$ that satisfies $\|f_0\|_2 = 1$ and $\operatorname{supp}(\widehat{f_0}) \subseteq [1,2]$. We will later (see Corollary 2.8 and Corollary 4.7) show that, for all $N \in \mathbb{N}$,

$$\lim_{m\to\infty} W_N\left(f_m\right) = 1,$$

where $f_m := 2^{m/2} f_0(2^m \cdot)$, $m \in \mathbb{N}$. This is a consequence of Lemma 1.8.

Therefore, we can find a strictly increasing sequence $(m_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ such that, for all $k \in \mathbb{N}$, $m_{k+1} \geq 2 + m_k$ and $W_k(f_{m_k}) \geq \frac{1}{2}$. We claim that

$$f_E := \sum_{k=1}^{\infty} a_k f_{m_k} \in L^2(\mathbb{R})$$

does the job (Figs. 3, 4). To this end, first note that, for all $m \in \mathbb{N}$, $\operatorname{supp}(\widehat{f_m}) \subseteq [2^m, 2^{m+1}]$. In particular, the $(f_m)_{m\in\mathbb{N}}$ are orthonormal, hence $\|f_E\|_2^2 = \sum_{k=1}^{\infty} |a_k|^2 = E_1$. Moreover, since

$$\operatorname{supp}(\widehat{\psi_j}) \subseteq [\operatorname{sgn}(j) \cdot 2^{|j|-1}, \operatorname{sgn}(j) \cdot 2^{|j|+1}],$$

we have

$$f_E * \psi_j = \begin{cases} a_k f_{m_k} * \psi_{m_k} & \text{if } j = m_k \text{ for some } k \in \mathbb{N} \\ a_k f_{m_k} * \psi_{m_k+1} & \text{if } j = m_k + 1 \text{ for some } k \in \mathbb{N} \\ 0 & \text{else} \end{cases}, \quad j \in \mathbb{Z} \setminus \{0\}.$$

Altogether, we obtain, for all $N \in \mathbb{N}$,

$$\begin{aligned} W_N(f_E) &= \sum_{j\in\mathbb{Z}\setminus\{0\}} W_{N-1}(|f_E * \psi_j|) \\ &= \sum_{k=1}^{\infty} \left(W_{N-1}(|a_k f_{m_k} * \psi_{m_k}|) + W_{N-1}(|a_k f_{m_k} * \psi_{m_k+1}|) \right) \\ &= \sum_{k=1}^{\infty} |a_k|^2 \cdot \left(W_{N-1}(|f_{m_k} * \psi_{m_k}|) + W_{N-1}(|f_{m_k} * \psi_{m_k+1}|) \right) \\ &= \sum_{k=1}^{\infty} |a_k|^2 \cdot W_N(f_{m_k}) \geq \sum_{k=N}^{\infty} |a_k|^2 \cdot W_k(f_{m_k}) \geq \frac{1}{2} \cdot \sum_{k=N}^{\infty} |a_k|^2 = \frac{1}{2} \cdot E_N. \quad \square \end{aligned}$$

To summarize, the central idea of the proof is as follows: Given $\delta > 0$, construct a sequence of signals $(f_{m_k})_{k\in\mathbb{N}} \subseteq L^2(\mathbb{R})$ such that (i) the energy remainder W_N behaves (super-)additively due to separation in the frequency domain and (ii) the energy decays slowly for

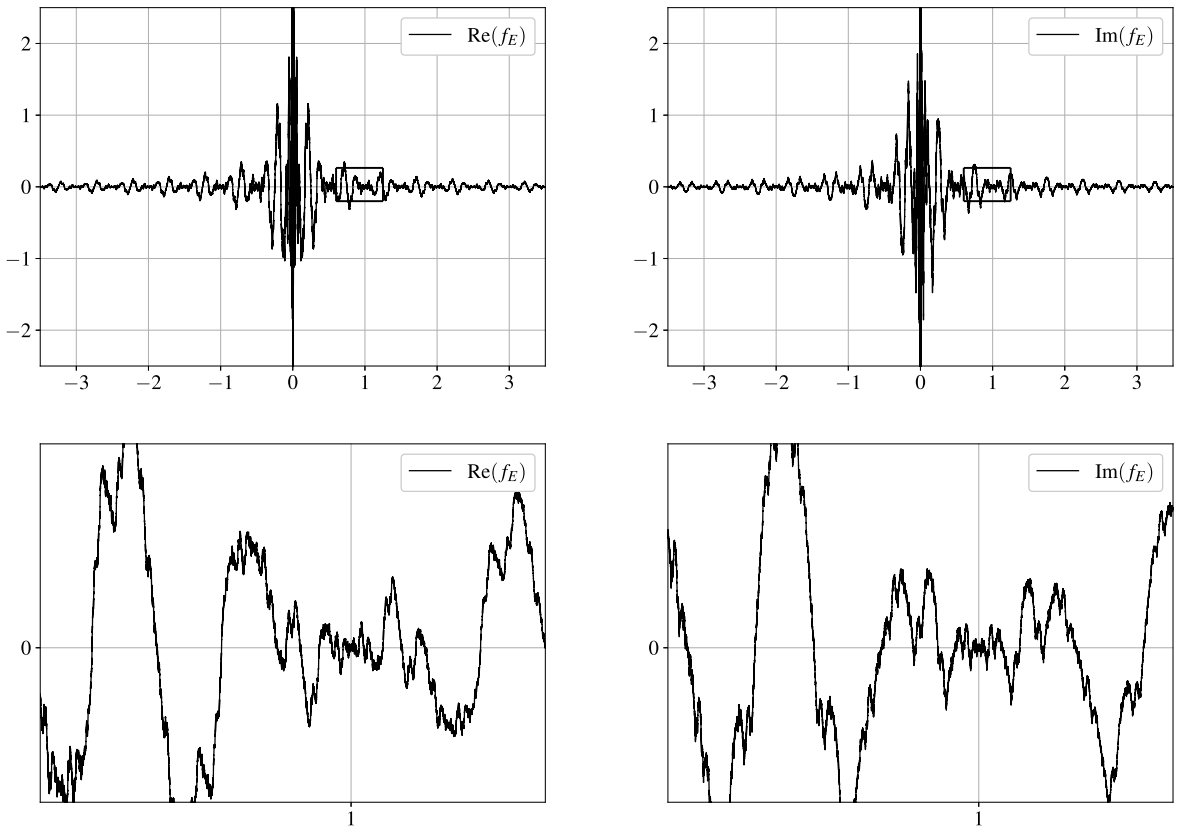


Fig. 4. Approximation of the real part (left), and imaginary part (right) of f_E, where $E = (1/k)_{k\in\mathbb{N}}$, and $f_0 = \mathcal{F}^{-1}\left(\mathbb{1}_{[1,\frac{3}{2}]} - \mathbb{1}_{[\frac{3}{2},2]}\right)$. To avoid numerical issues arising from numbers being close to zero, and for the sake of illustration, we have explicitly chosen $m_k = 2k$, $k \in \mathbb{N}$. The top figures indicate the global behavior of f_E, while the zoomed-in figures below suggest the wild local behavior of our semi-explicit constructions.

f_{m_k} across the first k layers of the scattering network, according to $W_k(f_{m_k}) \geq \delta$. Weighting the signals with an ℓ^2-sequence ensures square-integrability of f_E. The energy that is contained in the scattering layers of f_E of depth greater than $N-1$ (i.e., $W_N(f_E)$) is controlled by the weighted energy remainders of the signals $(f_{m_k})_{k\geq N}$. This results in a slow propagation of energy across all network layers. The speed of convergence is bounded from below by the speed of convergence of the ℓ^2-series of the chosen weights.

In order to generalize this idea to larger classes of filter banks and higher dimensions, we need to establish a much more general version of super-additivity for the family of nonlinear operators $(W_N : L^2(\mathbb{R}^d) \to \mathbb{R})_{N\in\mathbb{N}}$. In the proof sketch above, we have implicitly used their additivity on separated signals (cf. [15, Lemma 2.3]). However, if the generator ψ fails to be bandlimited, we cannot generally expect exact additivity on separated signals. Moreover, we would like to allow for more flexibility in the choice of our filter bank. The following lemma establishes an *approximate* version of super-additivity for the operators $(W_N)_{N\in\mathbb{N}}$ if the signals are nearly separated in Fourier domain. We use the high-pass filters Ψ to measure the degree of separation.

Lemma 2.1. *Let $F = (f_k)_{k\in\mathbb{N}_0} \subset L^2(\mathbb{R}^d)$, let $(\eta_k)_{k\in\mathbb{N}_0} \in \mathbb{R}_{>0}^{\mathbb{N}_0}$, and let $a = (a_k)_{k\in\mathbb{N}_0} \in \ell^2(\mathbb{N}_0;\mathbb{C})$. Assume that there exist finite sets $\Psi_k \subseteq \Psi$, $k \in \mathbb{N}_0$, such that the following hold:*

(i) The series $\sum_{k,j=0}^{\infty} \left| a_k a_j \left\langle f_k, f_j \right\rangle \right|$ converges.
(ii) The series $\sum_{k=0}^{\infty} \varepsilon_k$ converges, where $\varepsilon_k := (k+1)\cdot\eta_k + \sum_{j=k+1}^{\infty} \eta_j$.
(iii) We have $\Psi_{k_1} \cap \Psi_{k_2} = \emptyset$ whenever $k_1 \neq k_2$.
(iv) We have, for every $k \in \mathbb{N}_0$,

$$\sum_{\psi\in\Psi\setminus\Psi_k} \|f_k * \psi\|_2^2 \leq \eta_k.$$

(v) We have, for every $k \in \mathbb{N}_0$ and $0 \leq j < k$,

$$\sum_{\psi\in\Psi_k} \|f_j * \psi\|_2^2 \leq \eta_k.$$

Then, the series

$$f := \sum_{k=0}^{\infty} a_k f_k$$

converges in $L^2(\mathbb{R}^d)$ *with* $\|f\|_2^2 \leq \sum_{k,j=0}^{\infty} \left| a_k a_j \left\langle f_k, f_j \right\rangle \right|$, *and we have, for every* $N \in \mathbb{N}$, $n \in \mathbb{N}_0$,

$$W_N(f) \geq \sum_{k=n}^{\infty} \left(\frac{|a_k|^2}{2} \cdot W_N(f_k) - 2 \|a\|_{\ell^2(\mathbb{C})}^2 \cdot \varepsilon_k \right). \tag{19}$$

Moreover, suppose that in addition to our standing assumptions (i)-(v) there exists $\delta > 0$ *such that*

(vi) for all $k \in \mathbb{N}$,

$$W_k(f_k) \geq 4\delta.$$

(vii) for all $k \in \mathbb{N}$,

$$a_k = 0 \qquad \text{or} \qquad \varepsilon_k \leq \frac{\delta \cdot |a_k|^2}{2 \|a\|_{\ell^2(\mathbb{C})}^2}.$$

Then we even have, for all $N \in \mathbb{N}$,

$$W_N(f) \geq \delta \cdot \sum_{k=N}^{\infty} |a_k|^2. \tag{20}$$

Proof. The series defining f converges if

$$\left\| \sum_{k=n}^{m} a_k f_k \right\|_2 \xrightarrow{n,m\to\infty} 0.$$

Expanding the latter norm squared yields

$$\left\| \sum_{k=n}^{m} a_k f_k \right\|_2^2 = \sum_{k,j=n}^{m} a_k \overline{a_j} \cdot \left\langle f_k, f_j \right\rangle \leq \sum_{k,j=n}^{\infty} \left| a_k a_j \cdot \left\langle f_k, f_j \right\rangle \right|. \tag{21}$$

Being the remainder of a convergent series by assumption (i), the right-hand side of (21) converges to 0, so that f is in fact a well-defined element of $L^2(\mathbb{R}^d)$. Setting $n = 0$ in (21), we also obtain the claimed bound for the (squared) norm of f as $m \to \infty$.

Let us now turn to the proof of (19). Here, we first note that the recursive nature of W_N and Lemma 1.6 b) imply

$$\begin{aligned}
W_N(f) &= \sum_{\psi\in\Psi} W_{N-1}(|f * \psi|) \\
&\geq \sum_{k=n}^{\infty} \sum_{\psi\in\Psi_k} W_{N-1}(|f * \psi|) \\
&\geq \sum_{k=n}^{\infty} \sum_{\psi\in\Psi_k} \left(\frac{1}{2} \cdot W_{N-1}(|(a_k f_k) * \psi|) - \big\| |f * \psi| - |(a_k f_k) * \psi| \big\|_2^2 \right) \\
&\geq \sum_{k=n}^{\infty} \sum_{\psi\in\Psi_k} \left(\frac{|a_k|^2}{2} \cdot W_{N-1}(|f_k * \psi|) - \big\| (f - a_k f_k) * \psi \big\|_2^2 \right) \\
&= \sum_{k=n}^{\infty} \left(\frac{|a_k|^2}{2} \sum_{\psi\in\Psi_k} W_{N-1}(|f_k * \psi|) - \mathrm{Rest}_k \right).
\end{aligned} \tag{22}$$

As a consequence of our assumptions (iii)-(v), the remainder is bounded by

$$\begin{aligned}
\mathrm{Rest}_k &= \sum_{\psi\in\Psi_k} \left\| \left(\sum_{j=0, j\neq k}^{\infty} a_j f_j \right) * \psi \right\|_2^2 \\
&\leq \sum_{\psi\in\Psi_k} \left(\sum_{j=0, j\neq k}^{\infty} |a_j| \cdot \left\| f_j * \psi \right\|_2 \right)^2
\end{aligned}$$

$$\leq 2 \sum_{\psi\in\Psi_k} \left[\left(\sum_{j=0}^{k-1} |a_j| \cdot \left\| f_j * \psi \right\|_2 \right)^2 + \left(\sum_{j=k+1}^{\infty} |a_j| \cdot \left\| f_j * \psi \right\|_2 \right)^2 \right]$$

$$\leq 2 \|a\|_{\ell^2(\mathbb{C})}^2 \cdot \sum_{\psi\in\Psi_k} \left(\sum_{j=0}^{k-1} \left\| f_j * \psi \right\|_2^2 + \sum_{j=k+1}^{\infty} \left\| f_j * \psi \right\|_2^2 \right)$$

$$\leq 2 \|a\|_{\ell^2(\mathbb{C})}^2 \cdot \left(k \cdot \eta_k + \sum_{j=k+1}^{\infty} \eta_j \right).$$

Moreover, for every $k \geq n$,

$$\begin{aligned} \sum_{\psi\in\Psi_k} W_{N-1}(|f_k * \psi|) &= \sum_{\psi\in\Psi} W_{N-1}(|f_k * \psi|) - \sum_{\psi\in\Psi\setminus\Psi_k} W_{N-1}(|f_k * \psi|) \\ &= W_N(f_k) - \sum_{\psi\in\Psi\setminus\Psi_k} W_{N-1}(|f_k * \psi|) \\ &\geq W_N(f_k) - \eta_k. \end{aligned}$$

Substituting this and the bound for Rest_k into (22) yields

$$\begin{aligned} W_N(f) &\geq \sum_{k=n}^{\infty} \left(\frac{|a_k|^2}{2} \sum_{\psi\in\Psi_k} W_{N-1}(|f_k * \psi|) - \mathrm{Rest}_k \right) \\ &\geq \sum_{k=n}^{\infty} \left[\frac{|a_k|^2}{2} \cdot \left(W_N(f_k) - \eta_k \right) - 2 \|a\|_{\ell^2(\mathbb{C})}^2 \cdot \left(k \cdot \eta_k + \sum_{j=k+1}^{\infty} \eta_j \right) \right] \\ &\geq \sum_{k=n}^{\infty} \left[\frac{|a_k|^2}{2} \cdot W_N(f_k) - 2 \|a\|_{\ell^2(\mathbb{C})}^2 \cdot \left((k+1) \cdot \eta_k + \sum_{j=k+1}^{\infty} \eta_j \right) \right] \\ &= \sum_{k=n}^{\infty} \left(\frac{|a_k|^2}{2} \cdot W_N(f_k) - 2 \|a\|_{\ell^2(\mathbb{C})}^2 \cdot \varepsilon_k \right). \end{aligned}$$

Finally, let us derive (20) from (19) under the additional assumptions (vi) and (vii). It suffices to show the statement under the assumption $a_k \neq 0$ for all $k \in \mathbb{N}_0$. The general statement then follows by throwing out all zero-terms from the input. Choosing $n = N$ in (19) and exploiting the upper bound for ε_k, $k \in \mathbb{N}$, we find that

$$W_N(f) \geq \sum_{k=N}^{\infty} \left(\frac{|a_k|^2}{2} \cdot W_N(f_k) - 2 \|a\|_{\ell^2(\mathbb{C})}^2 \cdot \varepsilon_k \right) \geq \sum_{k=N}^{\infty} \left(\frac{|a_k|^2}{2} \cdot W_N(f_k) - \delta \cdot |a_k|^2 \right). \tag{23}$$

Since $W_N(f)$ is nonincreasing in N for fixed $f \in L^2(\mathbb{R}^d)$, we have, for all $k \geq N$,

$$W_N(f_k) \geq W_k(f_k) \geq 4\delta.$$

Substituting this into (23) concludes the proof. □

We are now ready to state and prove a first theorem about arbitrarily slow energy propagation in scattering networks. This is a generalization of the idea outlined at the beginning of this section.

Theorem 2.2. *Let $F = (f_m)_{m\in\mathbb{N}_0} \subset L^2(\mathbb{R}^d)$. Suppose that for all tolerances $(\eta_k)_{k\in\mathbb{N}_0} \in \mathbb{R}_{>0}^{\mathbb{N}_0}$ there exist finite sets $\Psi_k \subseteq \Psi$, $k \in \mathbb{N}_0$, and a subsequence $(f_{m_k})_{k\in\mathbb{N}_0}$ of F with $m_0 = 0$ such that the following hold:*

(i) The expression $C_F := \sup_{m\in\mathbb{N}_0} \|f_m\|_2^2$ is finite, i.e., F is a bounded sequence.
(ii) We have, for all $j, k \in \mathbb{N}_0$ with $j < k$,

$$\left| \left\langle f_{m_k}, f_{m_j} \right\rangle \right| \leq \eta_k.$$

(iii) We have $\Psi_{k_1} \cap \Psi_{k_2} = \emptyset$ whenever $k_1 \neq k_2$.
(iv) We have, for every $k \in \mathbb{N}_0$,

$$\sum_{\psi\in\Psi\setminus\Psi_k} \left\| f_{m_k} * \psi \right\|_2^2 \leq \eta_k.$$

(v) We have, for every $k \in \mathbb{N}_0$ and $0 \le j < k$,

$$\sum_{\psi \in \Psi_k} \left\| f_{m_j} * \psi \right\|_2^2 \le \eta_k.$$

(vi) There exists $\delta > 0$ such that, for all $k \in \mathbb{N}$, $W_k(f_{m_k}) \ge 4\delta$.

If $E = (E_N)_{N\in\mathbb{N}} \in \mathbb{R}_{>0}^{\mathbb{N}}$ is a nonincreasing null-sequence, then there exists $f_E \in L^2(\mathbb{R}^d)$ that satisfies

a) $\|f_E\|_2^2 \le 2C_F \cdot (1 + E_1)$,
b) $\|f_E - f_0\|_2^2 \le 2C_F \cdot E_1$, *and*
c) for all $N \in \mathbb{N}$, $W_N(f_E) \ge \delta \cdot E_N$.

Proof. Define

$$(a_k)_{k\in\mathbb{N}} := \left(\sqrt{E_k - E_{k+1}}\right)_{k\in\mathbb{N}},$$

and set $a_0 := 1$. A telescoping argument gives, for every $N \in \mathbb{N}$,

$$\sum_{k=N}^{\infty} |a_k|^2 = \sum_{k=N}^{\infty} (E_k - E_{k+1}) = E_N.$$

In particular, $a = (a_k)_{k\in\mathbb{N}_0} \in \ell^2(\mathbb{N}_0;\mathbb{C})$ with $\|a\|_{\ell^2(\mathbb{C})}^2 = 1 + E_1$.

We want to apply Lemma 2.1 to a subsequence $(f_{m_k})_{k\in\mathbb{N}_0}$ of F, which implies that f_E can be chosen of the type

$$f_E := \sum_{k=0}^{\infty} a_k f_{m_k}.$$

Here, $(f_{m_k})_{k\in\mathbb{N}_0}$ is the subsequence of F from the prerequisites of this theorem that depends on tolerance levels $(\eta_k)_{k\in\mathbb{N}_0}$, which we now specify.

Recall from Lemma 2.1 that f_E converges in $L^2(\mathbb{R}^d)$ with an upper bound for its squared norm given by

$$\|f_E\|_2^2 \le \sum_{k,j=0}^{\infty} \left| a_k a_j \cdot \left\langle f_{m_k}, f_{m_j} \right\rangle \right|.$$

Likewise,

$$\|f_E - f_0\|_2^2 = \left\| \sum_{k=1}^{\infty} a_k f_{m_k} \right\|_2^2 \le \sum_{k,j=1}^{\infty} \left| a_k a_j \cdot \left\langle f_{m_k}, f_{m_j} \right\rangle \right|.$$

We bound the latter series using the approximate orthogonality relation (ii). Repeatedly applying the Cauchy-Schwarz inequality yields, for all $n \in \mathbb{N}_0$,

$$\begin{aligned}
&\sum_{k,j=n}^{\infty} \left| a_k a_j \cdot \left\langle f_{m_k}, f_{m_j} \right\rangle \right| \\
&\quad = \sum_{k=n}^{\infty} |a_k| \cdot \left(\sum_{j=n}^{k-1} |a_j| \cdot \underbrace{\left| \left\langle f_{m_k}, f_{m_j} \right\rangle \right|}_{\le \eta_k} + |a_k| \cdot \underbrace{\left| \left\langle f_{m_k}, f_{m_k} \right\rangle \right|}_{\le C_F} + \sum_{j=k+1}^{\infty} |a_j| \cdot \underbrace{\left| \left\langle f_{m_k}, f_{m_j} \right\rangle \right|}_{\le \eta_j} \right) \\
&\quad \le C_F \cdot \sum_{k=n}^{\infty} |a_k|^2 + \left(\sum_{j=n, j\neq k}^{\infty} |a_j|^2 \right)^{\frac{1}{2}} \cdot \left(\sum_{k=n}^{\infty} |a_k| \cdot \left[(k-n) \cdot \eta_k^2 + \sum_{j=k+1}^{\infty} \eta_j^2 \right]^{\frac{1}{2}} \right) \\
&\quad \le C_F \cdot \sum_{k=n}^{\infty} |a_k|^2 + \left(\sum_{j=n}^{\infty} |a_j|^2 \right) \cdot \left(\sum_{k=n}^{\infty} \left[(k-n) \cdot \eta_k^2 + \sum_{j=k+1}^{\infty} \eta_j^2 \right] \right)^{\frac{1}{2}}.
\end{aligned}$$

Thus, if we choose the tolerances $(\eta_k)_{k\in\mathbb{N}_0} \in \mathbb{R}_{>0}^{\mathbb{N}_0}$ small enough such that

$$\sum_{k=0}^{\infty} \left[k \cdot \eta_k^2 + \sum_{j=k+1}^{\infty} \eta_j^2 \right] \le C_F^2,$$

then we obtain

$$\sum_{k,j=n}^{\infty} \left| a_k a_j \cdot \left\langle f_{m_k}, f_{m_j} \right\rangle \right| \leq 2C_F \cdot \left(\sum_{j=n}^{\infty} |a_j|^2 \right) = \begin{cases} 2C_F \cdot (1+E_1) & \text{if } n=0 \\ 2C_F \cdot E_1 & \text{if } n=1 \end{cases}.$$

This proves the norm bounds in a) and b).

As in Lemma 2.1, define $\varepsilon_k := (k+1)\cdot \eta_k + \sum_{j=k+1}^{\infty} \eta_j$ for all $k \in \mathbb{N}_0$. Since the tolerances $(\eta_k)_{k\in\mathbb{N}_0} \in \mathbb{R}_{>0}^{\mathbb{N}_0}$ can be chosen arbitrarily small, we can additionally assume that the series $\sum_{k=0}^{\infty} \varepsilon_k$ converges and that it holds, for all $k \in \mathbb{N}_0$,

$$a_k = 0 \qquad \text{or} \qquad \varepsilon_k \leq \frac{\delta \cdot |a_k|^2}{2\,\|a\|_{\ell^2(\mathbb{C})}^2}.$$

By applying Lemma 2.1, we conclude that, for all $N \in \mathbb{N}$,

$$W_N(f_E) \geq \delta \cdot \sum_{k=N}^{\infty} |a_k|^2 = \delta \cdot E_N. \quad \square$$

Under certain assumptions on the high-pass filters Ψ, the latter theorem guarantees that for any nonincreasing null-sequence $E = (E_N)_{N\in\mathbb{N}} \in \mathbb{R}_{>0}^{\mathbb{N}}$ there exists at least one input signal $f_E \in L^2(\mathbb{R}^d)$ whose network energy propagates slower across the network layers than E. Specifically, this implies that the set

$$\left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; W_N(f) \notin \mathcal{O}(E_N) \right\}$$

is nonempty. Our next result provides a sufficient criterion under which this set is even dense in $L^2(\mathbb{R}^d)$.

Proposition 2.3. *Let $E = (E_N)_{N\in\mathbb{N}} \in \mathbb{R}_{>0}^{\mathbb{N}}$ be a nonincreasing null-sequence. Assume that for all $g \in L^2(\mathbb{R}^d)$ there are $C_g, \delta_g > 0$ such that the following holds: For every nonincreasing null-sequence $E' = (E'_N)_{N\in\mathbb{N}} \in \mathbb{R}_{>0}^{\mathbb{N}}$ with $\liminf_{N\to\infty} E_N/E'_N = 0$ there exists $f_{g,E'} \in L^2(\mathbb{R}^d)$ such that*

(i) $\left\| f_{g,E'} - g \right\|_2^2 \leq C_g \cdot E'_1$, *and*
(ii) $W_N(f_{g,E'}) \geq \delta_g \cdot E'_N$ *for all* $N \in \mathbb{N}$.

Then,

$$Y_E := \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; W_N(f) \in \mathcal{O}(E_N) \right\}$$

is a countable union of nowhere dense sets in $L^2(\mathbb{R}^d)$. In particular, $L^2(\mathbb{R}^d) \setminus Y_E$ is dense in $L^2(\mathbb{R}^d)$.

Proof. Define, for all $p \in \mathbb{N}$,

$$M_p := \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; \forall N \in \mathbb{N}:\ W_N(f) \leq p \cdot E_N \right\},$$

and observe that Y_E can be written as countable union of these sets,

$$Y_E = \bigcup_{p\in\mathbb{N}} M_p.$$

Our goal is to show that M_p is nowhere dense in $L^2(\mathbb{R}^d)$ for every $p \in \mathbb{N}$. To this end, let us first recall from Corollary 1.7 that the nonlinear operator $W_N : L^2(\mathbb{R}^d) \to \mathbb{R}$ is continuous. Thus,

$$M_p = \bigcap_{N\in\mathbb{N}} \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; W_N(f) \leq p \cdot E_N \right\}$$

is closed and it only remains to show that the interior of M_p is empty: Fix $g \in M_p$, $\varepsilon > 0$, and define

$$E'_N := \begin{cases} \frac{\varepsilon^2}{2C_g} & \text{if } N = 1 \\ E'_1 \cdot \sqrt{\frac{E_N}{E_1}} & \text{if } N \geq 2 \end{cases}, \quad \text{for } N \in \mathbb{N}.$$

Clearly, $E' = (E'_N)_{N\in\mathbb{N}} \in \mathbb{R}_{>0}^{\mathbb{N}}$ is a nonincreasing null-sequence that satisfies $\liminf_{N\to\infty} E_N/E'_N = 0$. Hence, there is $f_{g,E'} \in L^2(\mathbb{R}^d)$ that satisfies (i) and (ii) from the assumptions of this proposition. Thus, there is $N_0 \in \mathbb{N}$ such that

$$W_{N_0}(f_{g,E'}) \geq \delta_g \cdot E'_{N_0} > p \cdot E_{N_0},$$

which means

$$f_{g,E'} \in B_\epsilon(g) \setminus M_p,$$

where $B_\epsilon(g)$ denotes the open ball with radius ϵ and center g in $L^2(\mathbb{R}^d)$ with respect to $\|\cdot\|_2$. Since this entails that the interior of M_p is empty, we have shown in total that Y_E is a countable union of nowhere dense sets in $L^2(\mathbb{R}^d)$. Therefore, its complement $L^2(\mathbb{R}^d) \setminus Y_E$ is dense in $L^2(\mathbb{R}^d)$ by the Baire category theorem. □

For the remainder of this section, we examine the assumptions of Theorem 2.2 and Proposition 2.3. While the number of conditions on the family F and the high-pass filters Ψ that must be met simultaneously for our results to apply is considerable, we demonstrate that these conditions can in fact be easily satisfied. To provide a concrete reference model, let us next establish that F can be obtained by L^2-normalized dilations of a single function.

Lemma 2.4. *Let $f \in L^2(\mathbb{R}^d)$. If $(A_m)_{m\in\mathbb{N}} \in \mathrm{GL}_d(\mathbb{R})^{\mathbb{N}}$ satisfies $\lim_{m\to\infty} \sigma_{min}(A_m) = \infty$, then*

$$\lim_{m\to\infty} \left| \left\langle f, D^2_{A_m} f \right\rangle \right| = 0.$$

Proof. The statement is obvious if $f = 0$. For the other case, fix $f \in L^2(\mathbb{R}^d) \setminus \{0\}$ and $\epsilon > 0$. By Lebesgue's dominated convergence theorem there is $R > 0$ so that $f_R := f \cdot \mathbb{1}_{B_R(0)}$ satisfies

$$\|f - f_R\|_2 \le \frac{\epsilon}{2\|f\|_2}.$$

For $m \in \mathbb{N}$, splitting the inner product gives

$$\left| \left\langle f, D^2_{A_m} f \right\rangle \right| \le \left| \left\langle f, D^2_{A_m} f_R \right\rangle \right| + \left| \left\langle f, D^2_{A_m} (f - f_R) \right\rangle \right| \tag{24}$$

We obtain an upper bound for the second term by the Cauchy-Schwarz inequality,

$$\left| \left\langle f, D^2_{A_m} (f - f_R) \right\rangle \right| \le \|f\|_2 \cdot \left\| D^2_{A_m} (f - f_R) \right\|_2 = \|f\|_2 \cdot \|f - f_R\|_2 \le \frac{\epsilon}{2}. \tag{25}$$

We now turn our attention to the first term in (24). As can be seen from a singular value decomposition of A_m, we have, for all $x \in \mathbb{R}^d$,

$$\|A_m x\|_2 \ge \sigma_{\min}(A_m) \cdot \|x\|_2 .$$

Since by assumption $\lim_{m\to\infty} \sigma_{\min}(A_m) = \infty$, we find that, for all $x \in \mathbb{R}^d \setminus \{0\}$,

$$\mathbb{1}_{B_R(0)}(A_m x) \le \mathbb{1}_{B_{\frac{R}{\sigma_{\min}(A_m)}}(0)}(x) \xrightarrow{m\to\infty} 0.$$

Hence, by Lebesgue's dominated convergence theorem there is $m_0 \in \mathbb{N}$ so that, for all $m \ge m_0$,

$$\left\| f \cdot \mathbb{1}_{B_{\frac{R}{\sigma_{\min}(A_m)}}(0)} \right\|_2 \le \frac{\epsilon}{2\|f\|_2}.$$

Thus, for all $m \ge m_0$,

$$\begin{aligned} \left| \left\langle f, D^2_{A_m} f_R \right\rangle \right| &\le \int_{\mathbb{R}^d} |f(x)| \cdot |D^2_{A_m} f(x)| \cdot \mathbb{1}_{B_R(0)}(A_m x) \,\mathrm{d}x \\ &\le \int_{\mathbb{R}^d} |f(x)| \cdot |D^2_{A_m} f(x)| \cdot \mathbb{1}_{B_{\frac{R}{\sigma_{\min}(A_m)}}(0)}(x) \,\mathrm{d}x \\ &\le \left\| f \cdot \mathbb{1}_{B_{\frac{R}{\sigma_{\min}(A_m)}}(0)} \right\|_2 \cdot \left\| D^2_{A_m} f \right\|_2 \le \frac{\epsilon}{2}. \end{aligned} \tag{26}$$

Plugging (25) and (26) into (24) concludes the proof. □

The following lemma shows that F satisfies assumptions (i)–(v) from Theorem 2.2 if it is generated by L^2-normalized matrix-dilations of a fixed function for matrices that asymptotically become increasingly expansive.

Lemma 2.5. *Let $(A_m)_{m\in\mathbb{N}} \in \mathrm{GL}_d(\mathbb{R})^{\mathbb{N}}$ be so that $\lim_{m\to\infty} \sigma_{min}(A_m) = \infty$. Let $f \in L^2(\mathbb{R}^d)$, and define $f_m := D^2_{A_m} f$, $m \in \mathbb{N}$. Finally, choose $f_0 \in \{0, f\}$.*

Then, $F = (f_m)_{m\in\mathbb{N}_0}$ satisfies assumptions (i)–(v) from Theorem 2.2.

Proof. The statement is trivial if $f = 0$. Hence, suppose $f \in L^2(\mathbb{R}^d) \setminus \{0\}$ in the following. Furthermore, assumptions (i)-(v) of Theorem 2.2 are clearly easier to meet if we are in the case that $f_0 = 0$ (instead of $f_0 = f$). Therefore, we only spell out the proof for $f_0 = f$. For notational convenience, we set $A_0 := I_d$, so we can write $f_0 = D^2_{A_0} f$.

By the normalization of the dilations, $C_F := \sup_{m\in\mathbb{N}_0} \|f_m\|_2^2 = \|f\|_2^2$ is finite, i.e., assumption (i) of Theorem 2.2 is satisfied. We now have to show that for all tolerances $(\eta_k)_{k\in\mathbb{N}_0} \in \mathbb{R}_{>0}^{\mathbb{N}_0}$ there exist finite sets $\Psi_k \subseteq \Psi$, $k \in \mathbb{N}_0$, and a subsequence $(f_{m_k})_{k\in\mathbb{N}_0}$ of F with $m_0 = 0$ such that assumptions (ii)-(v) of Theorem 2.2 hold. We proceed by choosing the sets $\Psi_k \subseteq \Psi$, $k \in \mathbb{N}_0$, and $(m_k)_{k\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ inductively.

Starting with $k = 0$ and $m_0 = 0$, we only need to show that condition (iv) can be satisfied. Since

$$\sum_{\psi\in\Psi} \left\| f_{m_0} * \psi \right\|_2^2 \le \|f\|_2^2 < \infty,$$

there is a finite subset $\Psi_0 \subset \Psi$ such that

$$\sum_{\psi\in\Psi\setminus\Psi_0} \left\| f_{m_0} * \psi \right\|_2^2 \le \eta_0.$$

For the induction step, let $k \in \mathbb{N}$ and assume that $\Psi_0, \dots, \Psi_{k-1}$ and $m_0, \dots, m_{k-1}$ are such that (ii)-(v) hold. Our goal is now to obtain Ψ_k and m_k with the desired properties.

It is easy to meet condition (ii): For all $m \in \mathbb{N}_0$ and $0 \le j < k$, we have

$$\left| \left\langle f_m, f_{m_j} \right\rangle \right| = \left| \left\langle D^2_{A_{m_j}^{-1}} D^2_{A_m} f, f \right\rangle \right| = \left| \left\langle D^2_{A_m A_{m_j}^{-1}} f, f \right\rangle \right|.$$

Moreover,

$$\sigma_{\min}\left(A_m A_{m_j}^{-1}\right) \ge \sigma_{\min}\left(A_m\right) \cdot \sigma_{\min}\left(A_{m_j}^{-1}\right) = \frac{\sigma_{\min}\left(A_m\right)}{\sigma_{\max}\left(A_{m_j}\right)} \ge \frac{\sigma_{\min}\left(A_m\right)}{\max_{0\le l<k} \sigma_{\max}\left(A_{m_l}\right)}.$$

Since the latter bound is uniformly in j for $0 \le j < k$, and $\lim_{m\to\infty} \sigma_{\min}\left(A_m\right) = \infty$ by assumption, Lemma 2.4 implies that there exists $m'_k \in \mathbb{N}_0$ such that $\left| \left\langle f_{m_k}, f_{m_j} \right\rangle \right| < \eta_k$ holds for all $j \in \{0, \dots, k-1\}$ whenever we choose m_k larger than m'_k.

The harder part is to satisfy the remaining conditions (iii)-(v) simultaneously. We begin with a simple observation about the impact of matrix-dilations on the frequency content of f. For $R > r > 0$, let

$$P_{r,R} f := \mathcal{F}^{-1}(\mathcal{F}(f) \cdot \mathbb{1}_{S_{r,R}})$$

be the projection of f on the subspace of $L^2(\mathbb{R}^d)$ whose Fourier transform is supported in $S_{r,R}$. We introduce $\varepsilon \in (0, 1/2)$, which we will choose later in the proof (small enough). By Lebesgue's dominated convergence theorem, $\widehat{f}$ is concentrated in a spherical shell $S_{r(\varepsilon),R(\varepsilon)}$ up to ε-error for some $R(\varepsilon) > r(\varepsilon) > 0$, meaning that

$$\left\| P_{r(\varepsilon),R(\varepsilon)} f \right\|_2^2 \ge (1-\varepsilon) \cdot \|f\|_2^2.$$

Now, if $A \in \mathrm{GL}_d(\mathbb{R})$, then $\widehat{D^2_A f}$ is concentrated in $S_{\sigma_{\min}(A)\cdot r(\varepsilon),\sigma_{\max}(A)\cdot R(\varepsilon)}$ up to ε-error, since

$$\begin{aligned}
\left\| P_{\sigma_{\min}(A)\cdot r(\varepsilon),\sigma_{\max}(A)\cdot R(\varepsilon)} D^2_A f \right\|_2^2 &= \left\| \widehat{D^2_A f} \cdot \mathbb{1}_{S_{\sigma_{\min}(A)\cdot r(\varepsilon),\sigma_{\max}(A)\cdot R(\varepsilon)}} \right\|_2^2 \\
&= \int_{\mathbb{R}^d} |\det(A)|^{-1} \cdot |\widehat{f}(A^{-T}\xi)|^2 \cdot \mathbb{1}_{S_{\sigma_{\min}(A)\cdot r(\varepsilon),\sigma_{\max}(A)\cdot R(\varepsilon)}}(\xi)\, \mathrm{d}\xi \\
&= \int_{\mathbb{R}^d} |\widehat{f}(z)|^2 \cdot \mathbb{1}_{S_{\sigma_{\min}(A)\cdot r(\varepsilon),\sigma_{\max}(A)\cdot R(\varepsilon)}}(A^T z)\, \mathrm{d}z \\
&\ge \int_{\mathbb{R}^d} |\widehat{f}(z)|^2 \cdot \mathbb{1}_{S_{r(\varepsilon),R(\varepsilon)}}(z)\, \mathrm{d}z \\
&= \left\| \mathcal{F} P_{r(\varepsilon),R(\varepsilon)} f \right\|_2^2 \ge (1-\varepsilon) \cdot \|f\|_2^2.
\end{aligned}$$

Specifically, setting

$$r_{<k} := \min_{0\le j<k} \sigma_{\min}(A_{m_j}) \cdot r(\varepsilon) \quad \text{and} \quad R_{<k} := \max_{0\le j<k} \sigma_{\max}(A_{m_j}) \cdot R(\varepsilon),$$

the above frequency concentration estimate entails

$$\left\|(\mathrm{id}_{L^2(\mathbb{R}^d)} - P_{r_{<k},R_{<k}}) f_{m_j}\right\|_2^2 \leq \varepsilon \cdot \|f\|_2^2 \quad \text{for } 0 \leq j < k.$$

By the Littlewood-Paley condition (11) and the continuity of the Fourier transform of the filters, there is a finite subset $\widetilde{\Psi} \subseteq \Psi$ containing $\bigcup_{j=0}^{k-1} \Psi_j$, and such that

$$|\widehat{\chi}(\xi)|^2 + \sum_{\psi \in \widetilde{\Psi}} |\widehat{\psi}(\xi)|^2 > 1 - \varepsilon \quad \text{for all } \xi \in S_{r_{<k},R_{<k}}. \tag{27}$$

By the Riemann-Lebesgue lemma, there exists $T \geq R_{<k}$ such that

$$|\widehat{\chi}(\xi)|^2 + \sum_{\psi \in \widetilde{\Psi}} |\widehat{\psi}(\xi)|^2 < \varepsilon \quad \text{whenever } |\xi| \geq T. \tag{28}$$

Since $\lim_{m\to\infty} \sigma_{\min}(A_m) = \infty$ by assumption, there is $m_k'' \in \mathbb{N}$ with $\sigma_{\min}(A_m) \cdot r(\varepsilon) \geq T$ for all $m \geq m_k''$. Define $m_k := \max\{m_k', m_k''\}$, $r_k := \sigma_{\min}(A_{m_k}) \cdot r(\varepsilon)$, and $R_k := \sigma_{\max}(A_{m_k}) \cdot R(\varepsilon)$. Thus,

$$\left\|(\mathrm{id}_{L^2(\mathbb{R}^d)} - P_{r_k,R_k}) f_{m_k}\right\|_2^2 \leq \varepsilon \cdot \|f\|_2^2.$$

Moreover, because of (28), the Littlewood-Paley condition (11), and the continuity of the Fourier transform of the filters, there exists a finite subset $\Psi(k,\varepsilon) \subseteq \Psi \setminus \widetilde{\Psi}$ so that

$$\sum_{\psi \in \Psi(k,\varepsilon)} |\widehat{\psi}(\xi)|^2 > 1 - \varepsilon \quad \text{for all } \xi \in S_{r_k,R_k},$$

which itself entails

$$\sum_{\psi \in \Psi \setminus \Psi(k,\varepsilon)} |\widehat{\psi}(\xi)|^2 < \varepsilon \quad \text{for all } \xi \in S_{r_k,R_k}. \tag{29}$$

At the same time, (27) and the fact that $\Psi(k,\varepsilon) \subseteq \Psi \setminus \widetilde{\Psi}$, imply

$$\sum_{\psi \in \Psi(k,\varepsilon)} |\widehat{\psi}(\xi)|^2 < \varepsilon \quad \text{for all } \xi \in S_{r_{<k},R_{<k}}. \tag{30}$$

We claim that $\Psi_k := \Psi(k,\varepsilon)$ with $\varepsilon := \frac{\eta_k}{2\|f\|_2^2}$ does the job.

In fact, first note that conditions (i)-(iii) are immediately satisfied by the above construction. Concerning condition (iv), we have

$$\begin{aligned}
\sum_{\psi \in \Psi \setminus \Psi_k} \left\|f_{m_k} * \psi\right\|_2^2 &= \sum_{\psi \in \Psi \setminus \Psi_k} \left\|(P_{r_k,R_k} f_{m_k}) * \psi\right\|_2^2 + \sum_{\psi \in \Psi \setminus \Psi_k} \left\|((\mathrm{id}_{L^2(\mathbb{R}^d)} - P_{r_k,R_k}) f_{m_k}) * \psi\right\|_2^2 \\
&\leq \int_{\mathbb{R}^d} |\widehat{f_{m_k}}(\xi)|^2 \cdot \sum_{\psi \in \Psi \setminus \Psi_k} |\widehat{\psi}(\xi)|^2 \cdot \mathbb{1}_{S_{r_k,R_k}}(\xi)\, \mathrm{d}\xi + \left\|(\mathrm{id}_{L^2(\mathbb{R}^d)} - P_{r_k,R_k}) f_{m_k}\right\|_2^2 \\
&\leq \varepsilon \cdot \int_{\mathbb{R}^d} |\widehat{f_{m_k}}(\xi)|^2\, \mathrm{d}\xi + \varepsilon \cdot \|f\|_2^2 \\
&= \varepsilon \cdot \left\|f_{m_k}\right\|_2^2 + \varepsilon \cdot \|f\|_2^2 = \eta_k.
\end{aligned}$$

Finally, we obtain condition (v) by proceeding analogously to the previous step. Indeed, for $0 \leq j < k$, we have

$$\begin{aligned}
\sum_{\psi \in \Psi_k} \left\|f_{m_j} * \psi\right\|_2^2 &= \sum_{\psi \in \Psi_k} \left\|(P_{r_{<k},R_{<k}} f_{m_j}) * \psi\right\|_2^2 + \sum_{\psi \in \Psi_k} \left\|((\mathrm{id}_{L^2(\mathbb{R}^d)} - P_{r_{<k},R_{<k}}) f_{m_j}) * \psi\right\|_2^2 \\
&\leq \int_{\mathbb{R}^d} |\widehat{f_{m_j}}(\xi)|^2 \cdot \sum_{\psi \in \Psi_k} |\widehat{\psi}(\xi)|^2 \cdot \mathbb{1}_{S_{r_{<k},R_{<k}}}(\xi)\, \mathrm{d}\xi + \left\|(\mathrm{id}_{L^2(\mathbb{R}^d)} - P_{r_{<k},R_{<k}}) f_{m_j}\right\|_2^2 \\
&\leq \varepsilon \cdot \int_{\mathbb{R}^d} |\widehat{f_{m_j}}(\xi)|^2\, \mathrm{d}\xi + \varepsilon \cdot \|f\|_2^2 \\
&= \varepsilon \cdot \left\|f_{m_j}\right\|_2^2 + \varepsilon \cdot \|f\|_2^2 = \eta_k. \quad \square
\end{aligned}$$

We summarize our previous results in the case where F is generated by L^2-normalized matrix-dilations.

Corollary 2.6. *Let $(A_m)_{m\in\mathbb{N}} \in \mathrm{GL}_d(\mathbb{R})^{\mathbb{N}}$ be so that $\lim_{m\to\infty} \sigma_{min}(A_m) = \infty$. Suppose that*

$$\delta := \inf_{\substack{f\in L^2(\mathbb{R}^d),\\ \|f\|_2=1}} \inf_{k\in\mathbb{N}} \limsup_{m\to\infty} W_k(D^2_{A_m} f) > 0.$$

Then, for every $g \in L^2(\mathbb{R}^d)$ there exists a constant $C_g > 0$ with the following property:
For any nonincreasing null-sequence $E = (E_N)_{N\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_{>0}$ there is $f_{g,E} \in L^2(\mathbb{R}^d)$ such that

a) $\left\|f_{g,E}\right\|_2^2 \le C_g \cdot (1 + E_1)$,
b) $\left\|f_{g,E} - g\right\|_2^2 \le C_g \cdot E_1$, *and*
c) $W_N(f_{g,E}) \ge \frac{\|g\|_2^2}{8} \cdot \delta \cdot E_N$ *for all* $N \in \mathbb{N}$.

Furthermore,

$$Y_E := \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; W_N(f) \in \mathcal{O}(E_N) \right\}$$

is a countable union of nowhere dense sets in $L^2(\mathbb{R}^d)$. In particular, $L^2(\mathbb{R}^d) \setminus Y_E$ is dense in $L^2(\mathbb{R}^d)$.

Proof. Let us first assume that $g \neq 0$. Define $g_m := D^2_{A_m} g$ for $m \in \mathbb{N}_0$, where $A_0 := I_d$. From the assumption that $\delta > 0$, we conclude that there exists a strictly increasing sequence $(m_k)_{k\in\mathbb{N}_0} \subseteq \mathbb{N}_0$ with $m_0 = 0$, and such that $W_k(g_{m_k}) \ge \frac{\|g\|_2^2}{2} \cdot \delta$ for all $k \in \mathbb{N}$. The existence of $f_{g,E}$ that satisfies the properties a)–c) then follows from Theorem 2.2 and Lemma 2.5.

For $g = 0$, fix any $h \in L^2(\mathbb{R}^d)$ with $\|h\|_2 = 1$. Set $h_0 := 0 \in L^2(\mathbb{R}^d)$ and $h_m := D^2_{A_m} h$ for $m \in \mathbb{N}$. Again from the assumption that $\delta > 0$, we conclude that there is a strictly increasing sequence $(m_k)_{k\in\mathbb{N}_0} \subseteq \mathbb{N}_0$ with $m_0 = 0$, and such that, for all $k \in \mathbb{N}$, $W_k(h_{m_k}) \ge \frac{1}{2} \cdot \delta$. Applying Theorem 2.2 and Lemma 2.5 to the sequence $(h_{m_k})_{k\in\mathbb{N}}$ then yields $f_{g,E} := f_{h,E} \in L^2(\mathbb{R}^d)$ that satisfies $\left\|f_{g,E}\right\|_2^2 \le 2E_1$ and, for all $N \in \mathbb{N}$, $W_N(f_{g,E}) \ge \frac{1}{8} \cdot \delta \cdot E_N$.

Overall, this also shows that we can apply Proposition 2.3, which completes the proof. □

We are now ready to present the main theorem of this section. It states that the prerequisites of Corollary 2.6 are satisfied if the high-pass filters Ψ admit an inclusive structure when L^1-dilated by the matrix sequence $(A_m^{-1})_{m\in\mathbb{N}}$. Specifically, this includes filter banks, where the high-pass filters Ψ arise from the L^1-normalized dilations by the matrix sequence $(A_m)_{m\in\mathbb{N}}$ applied to a finite number of generators, as is the case for wavelet-generated filter banks. In these cases, both the result of arbitrarily slow energy propagation and its consequence concerning the density of signals that fail to meet a given decay rate apply.

Theorem 2.7. *Let $(A_m)_{m\in\mathbb{N}} \in \mathrm{GL}_d(\mathbb{R})^{\mathbb{N}}$ be such that the following are true:*

(i) We have $\lim_{m\to\infty} \sigma_{min}(A_m) = \infty$.
(ii) The inclusion $\Psi_m \subseteq \Psi_{m+1}$ holds for all $m \in \mathbb{N}$, where $\Psi_m := D^1_{A_m^{-1}} \Psi$.

Then, $\Psi_\infty := \bigcup_{m\in\mathbb{N}} \Psi_m$ is a semi-discrete Parseval frame. Moreover, for all $f \in L^2(\mathbb{R}^d)$, $k \in \mathbb{N}$,

$$\lim_{m\to\infty} W_k[\Psi]\left(D^2_{A_m} f\right) = \|f\|_2^2.$$

In particular, Corollary 2.6 applies with $\delta = 1$.

Proof. By the Littlewood-Paley condition (11), we have, for every $m \in \mathbb{N}$,

$$|\hat\chi(A_m^T\xi)|^2 + \sum_{\psi\in\Psi_m} |\hat\psi(\xi)|^2 = |\hat\chi(A_m^T\xi)|^2 + \sum_{\psi\in\Psi} |\hat\psi(A_m^T\xi)|^2 = 1 \quad \text{a.e. } \xi \in \mathbb{R}^d.$$

In particular, for all $f \in L^2(\mathbb{R}^d)$, $m \in \mathbb{N}$,

$$\sum_{\psi\in\Psi_m} \|f * \psi\|_2^2 \le \|f\|_2^2,$$

which implies

$$\sum_{\psi\in\Psi_\infty} \|f * \psi\|_2^2 \le \|f\|_2^2.$$

On the other hand, as a consequence of $\lim_{m\to\infty} \sigma_{\min}(A_m^T) = \infty$, the Riemann-Lebesgue lemma, and Lebesgue's dominated convergence theorem, we find that

$$\sum_{\psi\in\Psi_\infty} \|f * \psi\|_2^2 \geq \sum_{\psi\in\Psi_m} \|f * \psi\|_2^2 = \|f\|_2^2 - \left\| f * D^1_{A_m^{-1}} \chi \right\|_2^2 \xrightarrow{m\to\infty} \|f\|_2^2 .$$

Thus, Ψ_∞ is a semi-discrete Parseval frame.

Fix $f \in L^2(\mathbb{R}^d)$. Recall from Lemma 1.8 that we have, for all $k \in \mathbb{N}$, $m \in \mathbb{N}$,

$$W_k[\Psi](D^2_{A_m} f) = W_k[D^1_{A_m^{-1}}\Psi](f) = W_k[\Psi_m](f).$$

Let $\varepsilon > 0$. Since $W_k[\Psi_\infty](f) = \|f\|_2^2$ by Lemma 1.4, there is a finite subset $\widetilde{\Psi} = \widetilde{\Psi}(\varepsilon) \subset \Psi_\infty$ such that

$$W_k[\widetilde{\Psi}](f) \geq \|f\|_2^2 - \varepsilon.$$

Since $\widetilde{\Psi}$ is finite, there exists $M = M(\varepsilon) \in \mathbb{N}$ with $\widetilde{\Psi} \subset \Psi_M$. Thus, for all $m \geq M$,

$$\|f\|_2^2 \geq W_k[\Psi_m](f) \geq W_k[\Psi_M](f) \geq W_k[\widetilde{\Psi}](f) \geq \|f\|_2^2 - \varepsilon,$$

which concludes the second part of the theorem. □

In our main application - concerning the wavelet scattering transform; see Section 4.2 - of the previous theorem, the matrices $(A_m)_{m\in\mathbb{N}}$ are generated by a single expansive matrix immediately implying that condition (i) from Theorem 2.7 is fulfilled. We conclude this section by explicitly (re-)stating our main negative findings for this particular instance of our assumptions, which is much easier to verify than the assumptions of Theorem 2.2.

Corollary 2.8. *Let $A \in \mathrm{GL}_d(\mathbb{R})$ be such that $\sigma_{\min}(A) > 1$. If $\Psi_m \subseteq \Psi_{m+1}$ holds for all $m \in \mathbb{N}$, where $\Psi_m := D^1_{A^{-m}}\Psi$, then $\Psi_\infty := \bigcup_{m\in\mathbb{N}} \Psi_m$ is a semi-discrete Parseval frame. Moreover, for all $f \in L^2(\mathbb{R}^d)$, $k \in \mathbb{N}$,*

$$\lim_{m\to\infty} W_k[\Psi]\left(D^2_{A^m} f\right) = \|f\|_2^2 .$$

In particular, Corollary 2.6 applies with $\delta = 1$.

3. Convergence rates for scattering propagation

In this section, we show that under mild analyticity assumptions on the scattering filters Ψ, fast energy propagation can still be achieved for large signal classes in $L^2(\mathbb{R}^d)$. These signal classes depend on the required speed of energy propagation, i.e., the faster the energy propagation, the smaller the signal class for which we can *guarantee* this speed. This is reflected by an interplay between the decay behavior of the Fourier transform of the input signal relative to the filters and the frequency localization of the filters regarding the size and shape of their Fourier supports.

Notably, our results in this section are general, since

- they hold in any dimension $d \in \mathbb{N}$;
- our assumptions can be easily satisfied; and
- the setup provides several parameters that can be used for fine-tuning.

In particular, Section 4.2 demonstrates that these results apply to filter banks generated by bandlimited wavelets, thereby complementing the negative results from the previous section about arbitrarily slow energy propagation in such scattering networks.

Our analysis begins with a similar approach to that used in [16, Theorem 3.1] and [18, Theorem 1, Theorem 2], establishing an upper bound on $W_N[\Psi](f)$, for all $f \in L^2(\mathbb{R}^d)$ and $N \in \mathbb{N}_{\geq 2}$, which is of the type

$$W_N[\Psi](f) \leq \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot K_N(\xi)\, \mathrm{d}\xi \tag{31}$$

for a family $(K_N)_{N\in\mathbb{N}_{\geq 2}}$ of integral kernels (independent of f) that satisfy additional properties. In parts, we combine the techniques from [16,18] in the proof of Theorem 3.5, which allows for more flexibility in our assumptions on the filters. On the basis of this approach, we then leverage the inductive structure of the scattering network to involve the filters Ψ in the integral in (31).

To clarify our assumptions for this section, we introduce some additional notation. Let $v^* \in \mathbb{S}^{d-1}$ be the vector with entries $v_k^* := 1/\sqrt{d}$ for each $k \in \{1, \dots, d\}$. For $\rho \in [0, 1)$, we denote by

$$C^\rho := \left\{ x \in \mathbb{R}^d_{\geq 0} \;\middle|\; \langle x, v^* \rangle \geq (1-\rho) \cdot \|x\|_2 \right\} \tag{32}$$

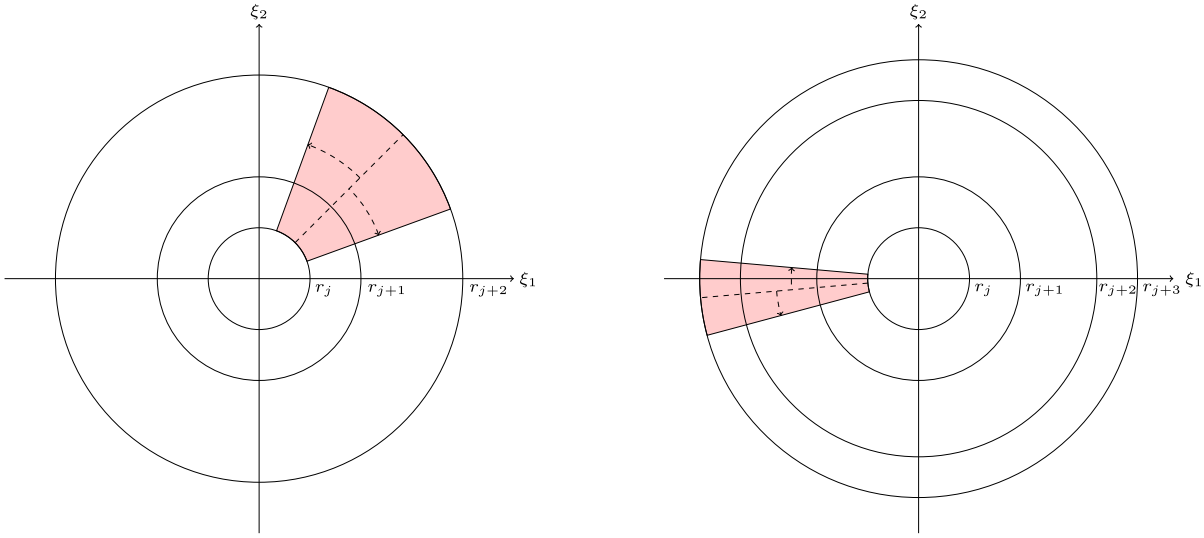


Fig. 5. Illustration of the allowed regions for the frequency supports of two different filters from (possibly) two different families of high-pass filters. The parameter κ (left $\kappa = 2$, right $\kappa = 3$) reflects the number of scales the filter may interfere with, and ρ parameterizes the opening angle of the colored segment.

the closed cone in the first canonical orthant with its tip at the origin and opening angle relative to v^* parameterized by ρ. Note that C^ρ is precisely the first canonical orthant $\mathbb{R}^d_{\geq 0}$ if $\rho \geq 1 - \frac{1}{\sqrt{d}}$. In particular, in dimension $d = 1$, we have, for all $\rho \in [0, 1)$, $C^\rho = [0, \infty)$. The following assumptions are motivated by [18, Assumption 1].

Assumption 3.1. Let $\kappa \in \mathbb{N}_{\geq 2}$. Let $(r_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_{>0}$ be a strictly increasing sequence of scales and suppose that

$$\gamma := \gamma(\kappa) := \inf_{j\in\mathbb{N}} \frac{r_j}{r_{j+\kappa}} > 0.$$

In addition to the Littlewood-Paley condition (11), we assume that there exists a partition $\Psi = \bigcup_{j\in\mathbb{N}} \Psi_j$ such that for every $\psi \in \Psi_j$, $j \in \mathbb{N}$, there is an orthogonal matrix $A_\psi \in O_d(\mathbb{R})$ such that

$$\operatorname{supp}(\widehat{\psi}) \subseteq C^\rho_{A_\psi} \cap S_{r_j, r_{j+\kappa}},$$

where $C^\rho_{A_\psi} := \{ A_\psi x \mid x \in C^\rho \}$.

The present setting may be regarded as a refinement of [18, Assumption 1], in that it incorporates additional information on the frequency localization of the filters, which ultimately affects the decay behavior of the energy remainder. We think of this as follows: By the Littlewood-Paley condition (11), the filters Ψ induce a decomposition of the frequency space $\mathbb{R}^d$ according to

$$\mathbb{R}^d = B_{r_1}(0) \cup \bigcup_{j\in\mathbb{N}} \bigcup_{\psi\in\Psi_j} (C^\rho_{A_\psi} \cap S_{r_j, r_{j+\kappa}}). \tag{33}$$

The filters Ψ are of high-pass nature in the sense that they leave a frequency gap around zero,

$$\sum_{\psi\in\Psi} |\widehat{\psi}(\xi)|^2 = 0 \quad \text{for all } \xi \in B_{r_1}(0). \tag{34}$$

By the Littlewood-Paley condition (11), this gap is filled by $\widehat{\chi}$. Moreover, any high-pass filter is frequency-localized in a region that is the intersection of a spherical shell, and a (rotated) cone with tip at the origin and angular spread parameterized by ρ. The maximum number of scale-interferences is specified by κ. The parameter γ reflects the (inverse of the) maximal relative spread between the inner and outer radius of any such shell. Clearly, we have $\gamma \in (0, 1)$, since the scales are strictly increasing. The closer γ is to 1 and the closer ρ is to 0, the more localized the high-pass filters are in frequency space (Fig. 5).

These additional parameters are reflected in the decay rates that we can guarantee for the energy remainder of certain signals, offering new insight into the interplay between the localization of the filters (in terms of size and shape of their Fourier supports) and the frequency decay behavior of the signals. We note however that our assumptions implicitly require the filters to be bandlimited, which is slightly more restrictive than the setting considered in [18, Assumption 1].

We begin our analysis with two auxiliary results about the geometry of the allowed domain for the Fourier supports of the filters. The first one provides an upper bound on the maximum distance from v^* to any other point in the set $C^\rho \cap \mathbb{S}^{d-1}$.

Lemma 3.2. *If $t > 0$, then* $\max_{x \in C^\rho \cap \mathbb{S}^{d-1}} \|x - tv^*\|_2^2 \leq t^2 - 2t \cdot (1-\rho) + 1$.

Proof. A direct computation shows that, for all $x \in C^\rho \cap \mathbb{S}^{d-1}$,

$$\left\|x - tv^*\right\|_2^2 = \|x\|_2^2 - 2t \cdot \langle x, v^* \rangle + t^2 \cdot \left\|v^*\right\|_2^2 \leq t^2 - 2t \cdot (1-\rho) \cdot \|x\|_2 + 1 = t^2 - 2t \cdot (1-\rho) + 1. \quad \square$$

The next lemma is crucial for the induction step in the proof of Theorem 3.5.

Lemma 3.3. *There is a family $(v_\psi)_{\psi \in \Psi} \subseteq \mathbb{R}^d$ so that, for all $\psi \in \Psi$, and all $\xi \in \operatorname{supp}(\widehat{\psi})$,*

$$\left\|\xi - v_\psi\right\|_2 \leq \alpha \cdot \|\xi\|_2,$$

where $\alpha = \alpha(\gamma, \rho) := \sqrt{1 - \frac{4\gamma}{(1+\gamma)^2} \cdot (1-\rho)^2} \in (0,1)$.

Proof. Since the function $(0,1) \to \mathbb{R},\ \gamma \mapsto \frac{4\gamma}{(1+\gamma)^2}$ is strictly increasing, we have

$$0 < \frac{4\gamma}{(1+\gamma)^2} \cdot (1-\rho)^2 \leq \frac{4\gamma}{(1+\gamma)^2} < \lim_{\gamma \uparrow 1} \frac{4\gamma}{(1+\gamma)^2} = 1.$$

Thus, $\alpha \in (0,1)$. Now let, for every $\psi \in \Psi_j$, $j \in \mathbb{N}$,

$$v_\psi := \frac{2r_j}{1+\gamma} \cdot (1-\rho) \cdot A_\psi v^* \in \mathbb{R}^d.$$

By Assumption 3.1, any $\xi \in \operatorname{supp}(\widehat{\psi})$ can be written as $A_\psi \omega = \xi$ for some $\omega \in C^\rho \cap S_{r_j, r_{j+\kappa}}$. Since A_ψ is an orthogonal matrix, it preserves the euclidean distances $\|\xi\|_2 = \|\omega\|_2$ and

$$\|\xi - v_\psi\|_2 = \left\|\omega - \frac{2r_j \cdot (1-\rho)}{1+\gamma} \cdot v^*\right\|_2.$$

It therefore suffices to show that the latter term is bounded by $\alpha \cdot \|\omega\|_2$. By Lemma 3.2,

$$\begin{aligned}
\frac{\left\|\omega - \frac{2r_j \cdot (1-\rho)}{(1+\gamma)} \cdot v^*\right\|_2}{\|\omega\|_2} &\leq \max_{s \in [r_j, r_{j+\kappa}]} \left\|\frac{\omega}{\|\omega\|_2} - \frac{2r_j \cdot (1-\rho)}{s \cdot (1+\gamma)} \cdot v^*\right\|_2 \\
&\leq \max_{s \in [r_j, r_{j+\kappa}]} \max_{x \in C^\rho \cap \mathbb{S}^{d-1}} \left\|x - \underbrace{\frac{2r_j \cdot (1-\rho)}{s \cdot (1+\gamma)}}_{=: t(s)} \cdot v^*\right\|_2 \\
&\leq \max_{s \in [r_j, r_{j+\kappa}]} \sqrt{t(s)^2 - 2t(s) \cdot (1-\rho) + 1} \\
&\leq \max_{s \in [r_j, \gamma^{-1} r_j]} \sqrt{t(s)^2 - 2t(s) \cdot (1-\rho) + 1} \\
&= \max_{t \in \left[\frac{2\gamma \cdot (1-\rho)}{1+\gamma}, \frac{2(1-\rho)}{1+\gamma}\right]} \sqrt{t^2 - 2t \cdot (1-\rho) + 1}. \qquad (35)
\end{aligned}$$

Since the polynomial

$$P : \mathbb{R} \to \mathbb{R}, \quad t \mapsto t^2 - 2t \cdot (1-\rho) + 1$$

is convex, the maximum in (35) is attained at one of the boundary points of the interval. A direct computation yields

$$\max_{r \in \left[\frac{2\gamma \cdot (1-\rho)}{1+\gamma}, \frac{2(1-\rho)}{1+\gamma}\right]} \sqrt{r^2 - 2r \cdot (1-\rho) + 1} = \sqrt{1 - \frac{4\gamma}{(1+\gamma)^2} \cdot (1-\rho)^2}. \quad \square$$

The following lemma is due to Mallat (cf. [2, Lemma 2.7]) and turns out to be essential for proving the main result of this section. For the sake of completeness, we reproduce its (elementary) proof.

Lemma 3.4. *Let $f \in L^2(\mathbb{R}^d)$, $g \in L^1(\mathbb{R}^d)$, and $v \in \mathbb{R}^d$. If $g \geq 0$, then*

$$\int_{\mathbb{R}^d} |\mathcal{F}(|f|)(\xi)|^2 \cdot \left(1 - \left|\widehat{g}(\xi)\right|^2\right) \mathrm{d}\xi \leq \int_{\mathbb{R}^d} \left|\widehat{f}(\xi)\right|^2 \cdot \left(1 - \left|\widehat{g}(\xi - v)\right|^2\right) \mathrm{d}\xi.$$

Proof. Clearly, since $g \geq 0$, we have

$$\| |f| * g \|_2^2 = \| |f| * |M_v g| \|_2^2 \geq \| f * M_v g \|_2^2 .$$

By the convolution theorem and Parseval's theorem, we obtain

$$\begin{aligned} \int_{\mathbb{R}^d} |\mathcal{F}(|f|)(\xi)|^2 \cdot \left(1 - \left|\widehat{g}(\xi)\right|^2\right) \mathrm{d}\xi &= \| |f| \|_2^2 - \| |f| * g \|_2^2 \\ &\leq \|f\|_2^2 - \|f * M_v g\|_2^2 = \int_{\mathbb{R}^d} \left|\widehat{f}(\xi)\right|^2 \cdot \left(1 - \left|\widehat{g}(\xi - v)\right|^2\right) \mathrm{d}\xi . \quad \square \end{aligned}$$

We are now ready to establish an upper bound on the energy remainder that is of the type (31). Previous work in this direction [16,18,19] showed that, in dimension $d = 1$, for certain wavelet-generated filter banks Ψ_{wav}, there exist constants $C = C(\Psi_{\mathrm{wav}}) > 0$ and $\alpha = \alpha(\Psi_{\mathrm{wav}}) \in (0,1)$ such that, for all (real-valued, cf. [16]) signals $f \in L^2(\mathbb{R})$, and for all $N \in \mathbb{N}_{\geq 2}$,

$$W_N[\Psi_{\mathrm{wav}}](f) \leq \int_{\mathbb{R}} |\widehat{f}(\xi)|^2 \cdot \left(1 - \left|\widehat{\vartheta}\left(C \cdot \alpha^{N-1} \cdot \xi\right)\right|^2\right) \mathrm{d}\xi .$$

Here, $\widehat{\vartheta}$ denotes a concretely specified positive definite and even function that ensures asymptotic decay of the right-hand side (a Gaussian in [16], and a truncated power function in [18]). Our result is an extension of this previous work, since it applies to the fairly large class of filter banks satisfying Assumption 3.1, including filter banks that are not necessarily structured as in the case of wavelet-generated filter banks. At the same time, our result holds in arbitrary dimension $d \in \mathbb{N}$. Moreover, for filters that satisfy our Assumption 3.1, which is slightly more restrictive than [18, Assumption 1], we improve the bound given in [18, Theorem 1] from polynomial decay of order m_d (where $m_d \in (0,1]$ and $m_d \to 0$ as $d \to \infty$) in the argument of $\widehat{\vartheta}$ to exponential decay.

Theorem 3.5. *Let $\vartheta \in L^1(\mathbb{R}^d)$, $\vartheta \geq 0$, and suppose that there exists a nonincreasing function $\eta : [0,\infty) \to [0,1]$ so that $\eta(0) = 1$ and $|\widehat{\vartheta}| = \eta(\|\cdot\|_2)$. Then, there exists a constant $C_{\chi,\vartheta} > 0$ such that, for all $f \in L^2(\mathbb{R}^d)$, and all $N \in \mathbb{N}_{\geq 2}$,*

$$W_N(f) \leq \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot \left(1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta} \cdot \alpha^{N-1} \cdot \xi\right)\right|^2\right) \mathrm{d}\xi , \tag{36}$$

where $\alpha = \alpha(\gamma,\rho) := \sqrt{1 - \frac{4\gamma}{(1+\gamma)^2} \cdot (1-\rho)^2}$.

Moreover, if there is $\widetilde{C} > 0$ so that $|\widehat{\vartheta}(\widetilde{C}\,\cdot)| \leq |\widehat{\chi}|$, then (36) holds for all $N \in \mathbb{N}$ with $C_{\chi,\vartheta} = \widetilde{C}$.

Proof. We perform induction on N.

First, let us assume that there is $\widetilde{C} > 0$ so that $|\widehat{\vartheta}(\widetilde{C}\,\cdot)| \leq |\widehat{\chi}|$. In this scenario, it is easy to prove the base case, which works analogously to the proof of the base case in [18, Theorem 1]. In fact, by the Littlewood-Paley condition (11), for all $f \in L^2(\mathbb{R}^d)$,

$$\begin{aligned} W_1(f) = \sum_{\psi \in \Psi} \|f * \psi\|_2^2 &= \int_{\mathbb{R}^d} \sum_{\psi \in \Psi} |\widehat{f}(\xi)|^2 \cdot |\widehat{\psi}(\xi)|^2 \, \mathrm{d}\xi \\ &= \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot (1 - |\widehat{\chi}(\xi)|^2) \, \mathrm{d}\xi \leq \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot (1 - |\widehat{\vartheta}(\widetilde{C} \cdot \xi)|^2) \, \mathrm{d}\xi . \end{aligned}$$

The proof of the base case is harder if there is no $\widetilde{C} > 0$ such that $|\widehat{\vartheta}(\widetilde{C}\,\cdot)| \leq |\widehat{\chi}|$ holds true. Here we proceed in three steps, motivated by [16, Section 6.3]. In the first two steps we construct certain auxiliary functions, whose properties we exploit to eventually prove the base case in the last step.

Step 1: Let $r_1 > 0$ be as in Assumption 3.1, corresponding to the radius of the frequency gap of the high-pass filters Ψ. Choose an even function $h \in C^\infty(\mathbb{R}^d)$ that is supported in $\overline{B_{\frac{r_1}{2}}(0)}$ and that satisfies $h(\xi) > 0$ for all $\xi \in B_{\frac{r_1}{2}}(0)$. Set $g := \mathcal{F}(h * h)$. Then, $g \in S(\mathbb{R}^d)$, $g = |\widehat{h}|^2 \geq 0$, $\mathrm{supp}(\widehat{g}) \subseteq \overline{B_{r_1}(0)}$, and $\widehat{g}(\xi) > 0$ for all $\xi \in \overline{B_{\frac{r_1}{2}}(0)}$. In particular,

$$m_g := \min_{\xi \in \overline{B_{\frac{r_1}{2}}(0)}} |\widehat{g}(\xi)|^2 > 0.$$

Finally, we may assume that $\|\widehat{g}\|_\infty \leq 1$ by imposing $\|h\|_2 \leq 1$ if necessary.

Step 2: Let $\tilde{c} : [0,\infty) \to (0,\infty)$ be nonincreasing and such that

$$\int_{\mathbb{R}^d} \tilde{c}(\|v\|_2) \, \mathrm{d}v = 1.$$

Defining the functions $c : \mathbb{R}^d \to [0,\infty)$, $\nu \mapsto \tilde{c}(\|\nu\|_2)$ and $F := |\widehat{g}|^2 * c$, we find that, for all $\xi \in \mathbb{R}^d$,

$$F(\xi) = \int_{\mathbb{R}^d} |\widehat{g}(\nu)|^2 \cdot c(\xi - \nu)\,\mathrm{d}\nu \geq m_g \int_{B_{\frac{r_1}{2}}(0)} c(\xi-\nu)\,\mathrm{d}\nu \geq m_g \cdot \mathrm{vol}\left(B_{\frac{r_1}{2}}(0)\right) \cdot \tilde{c}\left(\|\xi\|_2 + \frac{r_1}{2}\right).$$

Thus, if we choose $c = \tilde{c}(\|\cdot\|_2)$ so that

$$\lim_{s\to\infty} \frac{\eta^2(s)}{\tilde{c}\left(s + \frac{r_1}{2}\right)} = 0,$$

then there is $M > 0$ such that, for all $\xi \in \mathbb{R}^d \setminus B_M(0)$, it holds $F(\xi) \geq |\widehat{\vartheta}(\xi)|^2$. Furthermore, F is strictly positive and continuous. Hence,

$$m_F := \min_{\xi \in \overline{B_M(0)}} F(\xi) > 0.$$

Since $\lim_{s\to\infty} \eta(s) = 0$, there exists $C_{\chi,\vartheta} > \alpha^{-1}$ such that, for all $\xi \in \mathbb{R}^d \setminus B_{r_1}(0)$,

$$m_F \geq |\widehat{\vartheta}(C_{\chi,\vartheta} \cdot \alpha \cdot \xi)|^2.$$

Finally, we deduce $\|F\|_\infty \leq 1$ from $\|\widehat{g}\|_\infty \leq 1$ and $\|c\|_1 = 1$. Altogether, for all $\xi \in \mathbb{R}^d \setminus B_{r_1}(0)$,

$$0 \leq 1 - F(\xi) \leq 1 - |\widehat{\vartheta}(C_{\chi,\vartheta} \cdot \alpha \cdot \xi)|^2. \tag{37}$$

Step 3: Recall from (34) that $\sum_{\psi\in\Psi} |\widehat{\psi}|^2$ vanishes on $B_{r_1}(0)$. Since $\widehat{g}$ is supported in $\overline{B_{r_1}(0)}$ and since $\|\widehat{g}\|_\infty \leq 1$, we have

$$\sum_{\psi\in\Psi} |\widehat{\psi}(\xi)|^2 \leq 1 - |\widehat{g}(\xi)|^2 \quad \text{a.e. } \xi \in \mathbb{R}^d.$$

Hence, for all $f \in L^2(\mathbb{R}^d)$,

$$\begin{aligned} W_2(f) &= \sum_{\psi\in\Psi} \sum_{\psi'\in\Psi} \big\| |\,|f * \psi| * \psi'\,| \big\|_2^2 = \sum_{\psi\in\Psi} \sum_{\psi'\in\Psi} \big\| |f*\psi| * \psi' \big\|_2^2 \\ &\leq \sum_{\psi\in\Psi} \int_{\mathbb{R}^d} |\mathcal{F}(|f*\psi|)(\xi)|^2 \cdot \left(1 - |\widehat{g}(\xi)|^2\right)\,\mathrm{d}\xi\,. \end{aligned}$$

Using (34) and Lemma 3.4, we conclude that, for all $\nu \in \mathbb{R}^d$,

$$\begin{aligned} W_2(f) &\leq \sum_{\psi\in\Psi} \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot |\widehat{\psi}(\xi)|^2 \cdot \left(1 - |\widehat{g}(\xi-\nu)|^2\right)\,\mathrm{d}\xi \\ &\leq \int_{\mathbb{R}^d \setminus B_{r_1}(0)} |\widehat{f}(\xi)|^2 \cdot \left(1 - |\widehat{g}(\xi-\nu)|^2\right)\,\mathrm{d}\xi\,. \end{aligned}$$

Finally, we average this inequality with the function c and insert the upper bound from (37) to obtain

$$\begin{aligned} W_2(f) &\leq \int_{\mathbb{R}^d} c(\nu) \cdot \int_{\mathbb{R}^d \setminus B_{r_1}(0)} |\widehat{f}(\xi)|^2 \cdot \left(1 - |\widehat{g}(\xi-\nu)|^2\right)\,\mathrm{d}\xi\,\mathrm{d}\nu \\ &= \int_{\mathbb{R}^d \setminus B_{r_1}(0)} |\widehat{f}(\xi)|^2 \cdot \int_{\mathbb{R}^d} c(\nu) - c(\nu) \cdot |\widehat{g}(\xi-\nu)|^2\,\mathrm{d}\nu\,\mathrm{d}\xi \\ &= \int_{\mathbb{R}^d \setminus B_{r_1}(0)} |\widehat{f}(\xi)|^2 \cdot (1 - F(\xi))\,\mathrm{d}\xi \\ &\leq \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot \left(1 - |\vartheta(C_{\chi,\vartheta} \cdot \alpha \cdot \xi)|^2\right)\,\mathrm{d}\xi\,. \end{aligned}$$

This concludes the proof of the base case.

The induction step goes along the lines of [18, Theorem 1]: Suppose that (36) holds for some $N \in \mathbb{N}$ for all $f \in L^2(\mathbb{R}^d)$. Our goal is then to establish the analogous upper bound for $W_{N+1}(f)$. For $\psi \in \Psi$, let $\nu_\psi \in \mathbb{R}^d$ be as in Lemma 3.3. Note that the inverse Fourier transform of $\widehat{\vartheta}\left(C_{\chi,\vartheta} \cdot \alpha^{N-1} \cdot\right)$ is nonnegative by the non-negativity of ϑ. Thus, applying the induction hypothesis to $|f*\psi|$, as well as employing Lemma 3.4, we find that

$$W_N(|f*\psi|) \le \int_{\mathbb{R}^d} |\mathcal{F}(|f*\psi|)(\xi)|^2 \cdot \left(1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot\xi\right)\right|^2\right) \mathrm{d}\xi$$
$$\le \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot |\widehat{\psi}(\xi)|^2 \cdot \left(1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi - v_\psi)\right)\right|^2\right) \mathrm{d}\xi\,.$$

Summing over $\psi\in\Psi$ on both sides of the above inequality yields

$$W_{N+1}(f) = \sum_{\psi\in\Psi} W_N(|f*\psi|) \le \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot h_{N-1}(\xi)\,\mathrm{d}\xi\,, \tag{38}$$

where

$$h_{N-1}(\xi) := \sum_{\psi\in\Psi} |\widehat{\psi}(\xi)|^2 \cdot \left(1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi - v_\psi)\right)\right|^2\right).$$

Now, Lemma 3.3 comes into play to establish a pointwise upper bound for h_{N-1} that suffices to conclude the proof. Since $\sum_{\psi\in\Psi}|\widehat{\psi}(\xi)|^2 \le 1$ holds for all $\xi\in\mathbb{R}^d$, it is enough to show that we have, for all $\psi\in\Psi$, and all $\xi\in\operatorname{supp}(\widehat{\psi})$,

$$1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi - v_\psi)\right)\right|^2 \le 1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N}\cdot\xi\right)\right|^2.$$

In fact, $1 - |\widehat{\vartheta}|^2 = 1 - \eta^2(\|\cdot\|_2)$, where $1-\eta^2$ is nondecreasing, and Lemma 3.3 implies that

$$\left\|C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi - v_\psi)\right\|_2 \le \left\|C_{\chi,\vartheta}\cdot\alpha^{N}\cdot\xi\right\|_2.$$

Consequently, we have, for all $\xi\in\mathbb{R}^d$,

$$h_{N-1}(\xi) \le 1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N}\cdot\xi\right)\right|^2.$$

Inserting this into (38) concludes the induction step and thus the proof of the theorem. □

Remark 3.6. Smaller values of α lead to faster convergence to 0 of the integral on the right side of (36) as $N\to\infty$. Hence, let us briefly comment on the impact of the parameters on the value of α. To this end, we note that the function

$$\alpha : (0,1)\times[0,1) \to (0,1), \quad (\gamma,\rho) \mapsto \sqrt{1 - \frac{4\gamma}{(1+\gamma)^2}\cdot(1-\rho)^2}$$

is strictly decreasing in γ, while it is strictly increasing in ρ. Moreover, we have the asymptotic behavior

$$\lim_{(\gamma,\rho)\to(1,0)} \alpha(\gamma,\rho) = 0.$$

We interpret this as follows: α is small if the size of the Fourier supports of the filters is approximately constant across all scales and if the filters are well localized in terms of their maximum angular frequency spread. This result appears to be consistent with the earlier findings stated in [18, Theorem 1], which indicate a faster decay in lower dimensions: A reduction of the underlying dimension implicitly forces a higher concentration of angular frequency, with a maximum concentration achieved in dimension $d = 1$ (Fig. 6).

Next, we show how Theorem 3.5 entails fast (up to exponential) energy decay for signals belonging to generalized Sobolev spaces that are tailored to the frequency localization of the high-pass filters Ψ. We introduce certain weights to define these spaces.

Definition 3.7. Let $k>0$. We say that a function $\omega : (0,\infty)\to(0,\infty)$ is a **weakly t^k-dominated weight** if ω is nondecreasing and if there exists $T>0$ so that the auxiliary function

$$h_{k,\omega} : (0,\infty)\to(0,\infty),\ t\mapsto t^k\cdot\omega^{-2}(t)$$

is bounded on $(0,T)$, nondecreasing on $[T,\infty)$, and satisfies $\lim_{t\to\infty} h_{k,\omega}(t) = \infty$.

If, in addition to the above, $h_{k,\omega}$ is nondecreasing on the entire interval $(0,\infty)$, then we say that ω is a **strongly t^k-dominated weight**.

Remark 3.8. If ω is differentiable, we can easily establish a sufficient criterion for the (simultaneous) monotonicity of ω and $h_{k,\omega}$, relying on the non-negativity of their derivatives. In fact, for all $t>0$,

$$h'_{k,\omega}(t) = k\cdot t^{k-1}\cdot\omega^{-2}(t) - 2t^k\cdot\omega^{-3}(t)\cdot\omega'(t) = \omega^{-2}(t)\cdot t^{k-1}\cdot\left(k - 2t\cdot\omega^{-1}(t)\cdot\omega'(t)\right).$$

Thus, $h'_{k,\omega}(t)\ge 0$ is equivalent to $k\cdot\omega(t)\ge 2t\cdot\omega'(t)$. Altogether, if there is $T\ge 0$ so that for all $t>T$,

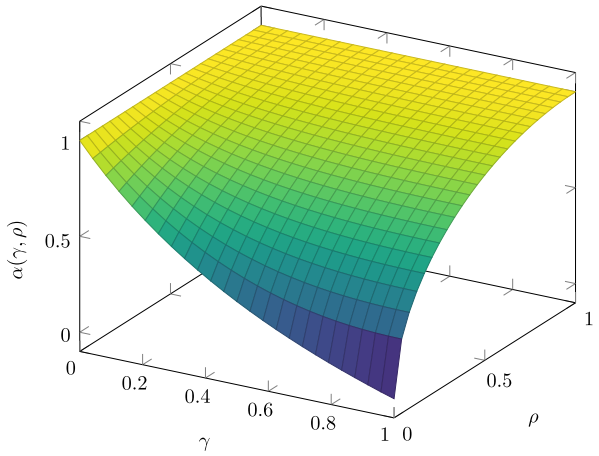


Fig. 6. Plot of α as a function of the parameters γ and ρ. Smaller values of α guarantee faster decay.

$$k \cdot \omega(t) \geq 2t \cdot \omega'(t) \geq 0,$$

then both ω and $h_{k,\omega}$ are nondecreasing on (T, ∞).

We define, for all $\psi \in \Psi$,

$$d_\psi := \inf_{\xi' \in \mathbb{R}^d} \sup_{\xi \in \operatorname{supp}(\widehat{\psi})} \left\| \xi - \xi' \right\|_2 = \min_{\xi' \in \mathbb{R}^d} \max_{\xi \in \operatorname{supp}(\widehat{\psi})} \left\| \xi - \xi' \right\|_2 . \tag{39}$$

For a weakly t^k-dominated weight, we consider the generalized Sobolev space

$$D_\omega(\Psi; L^2(\mathbb{R}^d)) := \left\{ f \in L^2(\mathbb{R}^d) \;\middle|\; \sum_{\psi \in \Psi} \omega^2(d_\psi) \cdot \|f * \psi\|_2^2 < \infty \right\}. \tag{40}$$

Theorem 3.9. *Let $\vartheta \in L^1(\mathbb{R}^d)$ be as in Theorem 3.5, and assume that there are $k = k_\vartheta > 0$, $C = C_\vartheta > 0$ such that, for all $\xi \in \mathbb{R}^d$,*

$$1 - |\widehat{\vartheta}(\xi)|^2 \leq C \cdot \|\xi\|_2^k . \tag{41}$$

If $\omega : (0, \infty) \to (0, \infty)$ is a weakly t^k-dominated weight, then we have, for all $f \in D_\omega(\Psi; L^2(\mathbb{R}^d))$,

$$W_N(f) \in \mathcal{O}\left(\omega^{-2}\left(\alpha^{-N}\right)\right),$$

where $\alpha = \alpha(\gamma, \rho) := \sqrt{1 - \frac{4\gamma}{(1+\gamma)^2} \cdot (1-\rho)^2} \in (0, 1)$.

Moreover, if there is $\widetilde{C} > 0$ so that $|\widehat{\vartheta}(\widetilde{C} \cdot)| \leq |\widehat{\chi}|$, and if ω is a strongly t^k-dominated weight, then the following explicit upper bound holds: For all $N \in \mathbb{N}_{\geq 2}$,

$$W_N(f) \leq \max\left\{1, C \cdot \widetilde{C}^k \cdot \alpha^{-k}\right\} \cdot \left(\sum_{\psi \in \Psi} \omega^2(d_\psi) \cdot \|f * \psi\|_2^2 \right) \cdot \omega^{-2}(\alpha^{-N}).$$

Remark 3.10. By the non-negativity of ϑ, the condition (41) can only be satisfied if $k = k_\vartheta \leq 2$. Let us briefly sketch the reason for this. By definition of the L^1-Fourier transform, we have, for all $\xi \in \mathbb{R}^d$,

$$\begin{aligned} 1 - |\widehat{\vartheta}(\xi)|^2 &= \operatorname{Re}\left(|\widehat{\vartheta}(0)|^2 - |\widehat{\vartheta}(\xi)|^2 \right) \\ &= \operatorname{Re}\left(\int_{\mathbb{R}^{2d}} \vartheta(x) \cdot \vartheta(y) \cdot \left(1 - e^{-2\pi \cdot i \cdot \langle \xi, x-y \rangle}\right) \, \mathrm{d}\lambda(x,y) \right) \\ &= \int_{\mathbb{R}^{2d}} \vartheta(x) \cdot \vartheta(y) \cdot (1 - \cos(2\pi \cdot \langle \xi, x-y \rangle)) \, \mathrm{d}\lambda(x,y) . \end{aligned}$$

For all $z \in [-1, 1]$,

$$1 - \cos(z) \geq \frac{z^2}{3}.$$

Hence, if $\xi = s \cdot e_1$, where $s > 0$, and e_1 denotes the first standard unit vector in $\mathbb{R}^d$, we obtain

$$\begin{aligned}\frac{1-|\widehat{\vartheta}(\xi)|^2}{\|\xi\|_2^2} &\geq \int\limits_{\{(x,y)\in\mathbb{R}^{2d}\,:\,|2\pi\cdot s\cdot(x_1-y_1)|\leq 1\}} \vartheta(x)\cdot\vartheta(y)\cdot(x_1-y_1)^2\,\mathrm{d}\lambda(x,y)\\ &\xrightarrow{s\to 0} \int\limits_{\mathbb{R}^{2d}} \vartheta(x)\cdot\vartheta(y)\cdot\big(x_1-y_1\big)^2\,\mathrm{d}\lambda(x,y) > 0.\end{aligned}$$

We conclude that $k_\vartheta \leq 2$.

Now, observe that larger values for k in (41) allow more flexibility in the choice of ω, as the assumptions regarding $h_{k,\omega}$ are easier to fulfill. However, as justified above, the largest possible value for any $\vartheta \in L^1(\mathbb{R}^d)$ is $k_\vartheta = 2$. Finally, note that, while we can show an analogous result to Theorem 3.5 for $\vartheta \in L^2(\mathbb{R}^d)$, we would not gain a qualitative improvement to Theorem 3.9, since the same threshold order of convergence $k_\vartheta = 2$ in (41) would still hold.

Proof of Theorem 3.9. Fix $f \in \mathcal{D}_\omega(\Psi; L^2(\mathbb{R}^d))$. Let, for every $\psi \in \Psi$,

$$\xi_\psi \in \operatorname*{argmin}_{\xi'\in\mathbb{R}^d} \max_{\xi\in\operatorname{supp}(\widehat{\psi})} \big\|\xi - \xi'\big\|_2 .$$

We start in a similar fashion to the induction step of the proof of Theorem 3.5. By the same theorem, by the path structure of the scattering network, and by Lemma 3.4, we find that, for all sufficiently large $N \in \mathbb{N}$,

$$\begin{aligned}W_N(f) &= \sum_{\psi\in\Psi} W_{N-1}(|f*\psi|)\\ &\leq \sum_{\psi\in\Psi} \int\limits_{\mathbb{R}^d} |\mathcal{F}(|f*\psi|)(\xi)|^2 \cdot \left(1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot\xi\right)\right|^2\right)\mathrm{d}\xi\\ &\leq \sum_{\psi\in\Psi} \int\limits_{\mathbb{R}^d} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2 \cdot \left(1 - \left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi-\xi_\psi)\right)\right|^2\right)\mathrm{d}\xi\,. \end{aligned} \tag{42}$$

Our strategy is to split the integral into small scales and large scales compared with α^{-N}, and to establish upper bounds on those terms separately. Note that our conditions on $h_{k,\omega}$ guarantee that, for all sufficiently large $N \in \mathbb{N}$,

$$\sup_{t\in(0,\alpha^{-N}]} h_{k,\omega}(t) = h_{k,\omega}(\alpha^{-N}) = \alpha^{-kN}\cdot\omega^{-2}(\alpha^{-N}). \tag{43}$$

For the small scales, we use (41) and (43) to derive

$$\begin{aligned}&\int\limits_{B_{\alpha^{-N}}(\xi_\psi)} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\left(1-\left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi-\xi_\psi)\right)\right|^2\right)\mathrm{d}\xi\\ &\quad\leq \int\limits_{B_{\alpha^{-N}}(\xi_\psi)} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot C\cdot\left\|C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi-\xi_\psi)\right\|_2^k\mathrm{d}\xi\\ &\quad= C\cdot C_{\chi,\vartheta}^k\cdot\alpha^{k(N-1)}\cdot\int\limits_{B_{\alpha^{-N}}(\xi_\psi)} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(\|\xi-\xi_\psi\|_2)\cdot h_{k,\omega}(\|\xi-\xi_\psi\|_2)\,\mathrm{d}\xi\\ &\quad\leq C\cdot C_{\chi,\vartheta}^k\cdot\alpha^{-k}\cdot\omega^{-2}(\alpha^{-N})\cdot\int\limits_{B_{\alpha^{-N}}(\xi_\psi)} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(d_\psi)\,\mathrm{d}\xi\,.\end{aligned}$$

The monotonicity of ω suffices to control the large scales,

$$\begin{aligned}&\int\limits_{\mathbb{R}^d\setminus B_{\alpha^{-N}}(\xi_\psi)} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\left(1-\left|\widehat{\vartheta}\left(C_{\chi,\vartheta}\cdot\alpha^{N-1}\cdot(\xi-\xi_\psi)\right)\right|^2\right)\mathrm{d}\xi\\ &\quad\leq \int\limits_{\mathbb{R}^d\setminus B_{\alpha^{-N}}(\xi_\psi)} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(\|\xi-\xi_\psi\|_2)\cdot\omega^{-2}(\|\xi-\xi_\psi\|_2)\,\mathrm{d}\xi\\ &\quad\leq \omega^{-2}(\alpha^{-N})\cdot\int\limits_{\mathbb{R}^d\setminus B_{\alpha^{-N}}(\xi_\psi)} |\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(d_\psi)\,\mathrm{d}\xi\,.\end{aligned}$$

Inserting those estimates into (42) yields, for all sufficiently large $N \in \mathbb{N}$,

$$W_N(f) \le \max\left\{1, C\cdot C_{\chi,\vartheta}^k\cdot\alpha^{-k}\right\}\cdot\omega^{-2}(\alpha^{-N})\cdot\sum_{\psi\in\Psi}\int_{\mathbb{R}^d}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(d_\psi)\,\mathrm{d}\xi$$

$$= \max\left\{1, C\cdot C_{\chi,\vartheta}^k\cdot\alpha^{-k}\right\}\cdot\left(\sum_{\psi\in\Psi}\omega^2(d_\psi)\cdot\|f*\psi\|_2^2\right)\cdot\omega^{-2}(\alpha^{-N}).$$

Now, if there is $\widetilde{C}>0$ so that $|\widehat{\vartheta}(C_{\chi,\vartheta}\cdot)|\le|\widehat{\chi}|$ and if $h_{k,\omega}$ is nondecreasing on the entire interval $(0,\infty)$, then (42) and (43), and thus also the latter bound for $W_N(f)$, are in fact valid for all $N\in\mathbb{N}_{\ge 2}$. This concludes the proof of the theorem. □

The following lemma shows that fast decay (with respect to the weight ω) of the Fourier transform suffices, independent of Ψ, for a signal to belong to the generalized Sobolev space $\mathcal{D}_\omega(\Psi;L^2(\mathbb{R}^d))$.

Lemma 3.11. *Let $k>0$. If $\omega:(0,\infty)\to(0,\infty)$ is a weakly t^k-dominated weight, then*

$$\mathcal{F}L^2_\omega(\mathbb{R}^d)\subseteq\mathcal{D}_\omega(\Psi;L^2(\mathbb{R}^d)).$$

Proof. Recall from Assumption 3.1 that Ψ can be decomposed into $\Psi=\bigcup_{j\in\mathbb{N}}\Psi_j$ such that $\mathrm{supp}(\widehat{\psi})\subseteq S_{r_j,r_{j+\kappa}}$ whenever $\psi\in\Psi_j$. In particular, $d_\psi\le r_{j+\kappa}\le\gamma^{-1}\cdot r_j$. By assumption on ω, there exists $J\in\mathbb{N}$ such that $h_{k,\omega}$ is nondecreasing on (r_J,∞). Hence, for all $j\in\mathbb{N}_{>J}$,

$$\omega^2(d_\psi)\le\omega^2(\gamma^{-1}\cdot r_j)\le\gamma^{-k}\cdot\omega^2(r_j). \tag{44}$$

Let $f\in\mathcal{F}L^2_\omega(\mathbb{R}^d)$. We aim to show that the sum $\sum_{\psi\in\Psi}\omega^2(d_\psi)\cdot\|f*\psi\|_2^2$ converges. To do so, we split the sum into small scales and large scales compared with J,

$$\sum_{j=1}^{J}\sum_{\psi\in\Psi_j}\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(d_\psi)\,\mathrm{d}\xi+\sum_{j=J+1}^{\infty}\sum_{\psi\in\Psi_j}\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(d_\psi)\,\mathrm{d}\xi,$$

so that it suffices to bound these expressions individually. We can easily control the small scales by the monotonicity of ω,

$$\sum_{j=1}^{J}\sum_{\psi\in\Psi_j}\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(d_\psi)\,\mathrm{d}\xi$$

$$\le\omega^2(r_{J+\kappa})\cdot\sum_{j=1}^{J}\sum_{\psi\in\Psi_j}\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\,\mathrm{d}\xi$$

$$\le\frac{\omega^2(r_{J+\kappa})}{\omega^2(r_1)}\cdot\sum_{j=1}^{J}\sum_{\psi\in\Psi_j}\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(\|\xi\|_2)\,\mathrm{d}\xi.$$

For the large scales, in addition to the monotonicity of ω, we also use (44) to find that

$$\sum_{j=J+1}^{\infty}\sum_{\psi\in\Psi_j}\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(d_\psi)\,\mathrm{d}\xi$$

$$\le\sum_{j=J+1}^{\infty}\sum_{\psi\in\Psi_j}\omega^2(\gamma^{-1}\cdot r_j)\cdot\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\,\mathrm{d}\xi$$

$$\le\sum_{j=J+1}^{\infty}\sum_{\psi\in\Psi_j}\gamma^{-k}\cdot\omega^2(r_j)\cdot\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\,\mathrm{d}\xi$$

$$\le\sum_{j=J+1}^{\infty}\sum_{\psi\in\Psi_j}\gamma^{-k}\cdot\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(\|\xi\|_2)\,\mathrm{d}\xi.$$

Altogether, letting $C=\max\left\{\frac{\omega^2(r_{J+\kappa})}{\omega^2(r_1)},\gamma^{-k}\right\}$, we obtain

$$\sum_{\psi\in\Psi}\omega^2(d_\psi)\cdot\|f*\psi\|_2^2\le C\cdot\sum_{j=1}^{\infty}\sum_{\psi\in\Psi_j}\int_{S_{r_j,r_{j+\kappa}}}|\widehat{f}(\xi)|^2\cdot|\widehat{\psi}(\xi)|^2\cdot\omega^2(\|\xi\|_2)\,\mathrm{d}\xi$$

$$\leq C \cdot \kappa \cdot \sum_{\psi \in \Psi} \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot |\widehat{\psi}(\xi)|^2 \cdot \omega^2(\|\xi\|_2)\, \mathrm{d}\xi$$

$$\leq C \cdot \kappa \cdot \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 \cdot \omega^2(\|\xi\|_2)\, \mathrm{d}\xi \; < \infty. \quad \square$$

There are choices of the weight ω and the filter bank Ψ, for which the inclusion $\mathcal{F}L^2_\omega(\mathbb{R}^d) \subseteq \mathcal{D}_\omega(\Psi; L^2(\mathbb{R}^d))$ is strict, i.e., the spaces $\mathcal{F}L^2_\omega(\mathbb{R}^d)$ and $\mathcal{D}_\omega(\Psi; L^2(\mathbb{R}^d))$ do not coincide, see Remark 4.3.

As a consequence of Theorem 3.9 and Lemma 3.11, we obtain asymptotic convergence rates for signals of (logarithmic) Sobolev regularity.

Corollary 3.12. *Suppose that Assumption 3.1 holds for Ψ with parameters γ and ρ. As before, let*

$$\alpha := \sqrt{1 - \frac{4\gamma}{(1+\gamma)^2} \cdot (1-\rho)^2} \in (0,1).$$

Then, the following hold:

a) If $f \in H^s(\mathbb{R}^d)$ for some $s > 0$, then

$$W_N(f) \in \mathcal{O}(\alpha^{2\min\{s,1\}\cdot N}).$$

b) If $f \in H^s_{\log}(\mathbb{R}^d)$ for some $s > 0$, then

$$W_N(f) \in \mathcal{O}(N^{-2s}).$$

Proof. Let

$$\vartheta : \mathbb{R}^d \to \mathbb{R}, \quad x \mapsto e^{-\pi \cdot \|x\|_2^2}.$$

Then, ϑ satisfies the prerequisites of Theorem 3.5. From the series expansion of the one-dimensional Gaussian at 0 we can see that, for all $\xi \in \mathbb{R}^d$,

$$1 - |\widehat{\vartheta}(\xi)|^2 \leq 2\pi \cdot \|\xi\|_2^2.$$

Now, part a) follows immediately from Theorem 3.9 and Lemma 3.11 if we take

$$\omega : (0,\infty) \to (0,\infty), \quad t \mapsto (1+t^2)^{\frac{s}{2}}.$$

In fact, using the criterion from Remark 3.8, it is straightforward to see that ω is a strongly t^2-dominated weight if $s \in (0,1]$, since for all $t > 0$,

$$2\omega(t) \geq 2s \cdot \frac{t^2}{1+t^2} \cdot (1+t^2)^{\frac{s}{2}} = 2t \cdot \omega'(t) \geq 0.$$

We draw the conclusion by noting that $\mathcal{F}L^2_\omega(\mathbb{R}^d) = H^s(\mathbb{R}^d)$ and that, for all $N \in \mathbb{N}$,

$$\omega^{-2}(\alpha^{-N}) \leq \alpha^{2s\cdot N}.$$

For $s > 1$, the statement simply follows from the inclusion $H^s(\mathbb{R}^d) \subseteq H^1(\mathbb{R}^d)$.

In order to establish part b), let us consider

$$\omega : (0,\infty) \to (0,\infty), \quad t \mapsto \ln^s(e+t).$$

We verify the condition from Remark 3.8 again. In general, ω is not a strongly t^2-dominated weight (if s is too large). However, ω is a weakly t^2-dominated weight. Indeed, for all $t \geq e^s - e$, we have $\ln(e+t) \geq s$, which entails

$$2\omega(t) = 2\ln(e+t) \cdot \ln^{s-1}(e+t) \geq 2s \cdot \frac{t}{e+t} \cdot \ln^{s-1}(e+t) = 2t \cdot \omega'(t) \geq 0.$$

Finally, by definition, $\mathcal{F}L^2_\omega(\mathbb{R}^d) = H^s_{\log}(\mathbb{R}^d)$, and we have, for all $N \in \mathbb{N}$,

$$\omega^{-2}(\alpha^{-N}) \leq \ln^{-2s}(\alpha^{-N}) = \ln^{-2s}(\alpha^{-1}) \cdot N^{-2s}.$$

This concludes the proof of part b) by Theorem 3.9. $\square$

Remark 3.13. The Sobolev spaces contain many practically relevant signal classes, such as the class of bandlimited signals and the class of cartoon functions. In the mathematical signal processing literature (e.g., [18,22,23] in the context of scattering transforms),

cartoon functions often serve as a reference model for natural images (e.g., images of handwritten digits). If defined as in [18, Definition 1], the cartoon functions indeed belong to $H^s(\mathbb{R}^d)$ for all $s \in (0, \frac{1}{2})$; see [18, Lemma 1]. Thus, by Corollary 3.12, their scattering coefficients admit exponential energy decay if the scattering filters satisfy Assumption 3.1.

Using arguments very similar to those in the proof of Theorem 3.9 together with the **compactly** supported positive definite radial basis function

$$\widehat{\vartheta} : \mathbb{R}^d \to \mathbb{R}, \quad \xi \mapsto \left(1 - \|\xi\|_2\right)_+^{\left\lfloor \frac{d}{2} \right\rfloor + 1} \tag{45}$$

in the above proof, one can derive the following **explicit** upper bounds (as opposed to the asymptotic bounds in the above corollary):

a) If $f \in H^s(\mathbb{R}^d)$ for some $s > 0$, then we have, for all $N \in \mathbb{N}$,

$$W_N(f) \leq \max\left\{1, \frac{2\left(\left\lfloor \frac{d}{2} \right\rfloor + 1\right)}{\alpha \cdot r_1}\right\} \cdot \|f\|^2_{H^s(\mathbb{R}^d)} \cdot \alpha^{\min\{2s,1\} \cdot N}.$$

b) If $f \in H^s_{\log}(\mathbb{R}^d)$ for some $s > 0$, then we have, for all $N \in \mathbb{N}$,

$$W_N(f) \leq \max\left\{1, \frac{2\left(\left\lfloor \frac{d}{2} \right\rfloor + 1\right)}{\alpha \cdot r_1}\right\} \cdot \ln^{-\min\{2s,1\}}(\alpha^{-1}) \cdot \|f\|^2_{H^s_{\log}(\mathbb{R}^d)} \cdot N^{-\min\{2s,1\}}.$$

The differences in the asymptotic rates of decay (compared with the above corollary) are due to the fact that

- the largest possible values k_ϑ in (41) differ if ϑ is a Gaussian ($k_\vartheta = 2$) or if ϑ is defined as in (45) ($k_\vartheta = 1$).
- the weight $\omega : (0, \infty) \to (0, \infty), t \mapsto \ln^s(e + t)$ is not strongly t-dominated if s is too large. However, it is easy to see that ω is strongly t-dominated if $s \in (0, \frac{1}{2}]$.

4. Applications

4.1. Scattering with UFC filters

In this section, we apply our results from Section 3 to a large class of filter banks that is closely related to the class of uniform covering frames introduced in [17].

Definition 4.1. We say that the filters Ψ have **uniform frequency concentration** (UFC) if they fulfill Assumption 3.1, and if

$$D_\Psi := \sup_{\psi \in \Psi} d_\psi = \sup_{\psi \in \Psi} \inf_{\xi_\psi \in \mathbb{R}^d} \sup_{\xi \in \operatorname{supp}(\widehat{\psi})} \left\| \xi - \xi_\psi \right\|_2 < \infty.$$

Starting from [17, Proposition 2.3], one can see that the class of uniform frequency concentration filters includes certain Weyl-Heisenberg (Gabor) frames. More generally, UFC filters are easy to construct as follows: Departing from a covering of the frequency space $\mathbb{R}^d$ that is compatible with Assumption 3.1, one chooses a suitable partition of unity (typically a bounded uniform partition of unity [24, Definition 11.6.1]) subordinate to this covering, as indicated by (33). Renormalizing (cf. [3, Proposition 3]) this partition of unity yields the desired UFC filters.

It turns out that for such scattering networks, energy decay is exponential, globally on $L^2(\mathbb{R}^d)$. Thereby, we complement the earlier result [17, Proposition 3.3], which states global exponential energy decay for scattering networks based on uniform covering frames. We have several comments concerning the similarities and differences between our following result and the result in [17]:

- In [17], the output-generating low-pass filter is assumed to be bandlimited. Our result also applies if that is not the case.
- In [17], the frequency support of each filter $\psi \in \Psi$ is assumed to be connected. Our result does not require connectedness of the frequency supports. However, as is pointed out in [17, Remark 2.2], the motivation to require connectedness of the frequency supports is to preclude certain pathological behavior such as $\operatorname{supp}(\widehat{\psi})$ having two connected components, where one component is near the origin and the other is far from the origin. In our setting, we implicitly preclude such pathological behavior by means of Assumption 3.1, which imposes frequency localization of the filters.
- If $D_\Psi < \infty$, then the frequency support of each filter $\psi \in \Psi$ is contained in a closed ball of radius D_Ψ. In particular, Ψ satisfies the uniform covering property introduced in [17], i.e., for any $R > 0$, there exists $N \in \mathbb{N}$ such that for each $\psi \in \Psi$, the set $\operatorname{supp}(\widehat{\psi})$ can be covered by N cubes of side length $2R$. In turn, if Ψ satisfies the uniform covering property, and if the frequency supports of the filters are connected, then $D_\Psi < \infty$.

- The findings in [17, Proposition 3.3] guarantee exponential decay of the energy remainder: For all $f \in L^2(\mathbb{R}^d)$, and all $N \in \mathbb{N}$,

$$W_N(f) \leq \alpha^N \cdot \left(\|f\|_2^2 - \|f * \chi\|_2^2 \right),$$

for an *unspecified* constant $\alpha \in (0,1)$ depending only on the uniform covering frame. The main advantage of our result is that we can explicitly specify the values for all occurring quantities in the upper bound on the energy remainder, including a precise value for α.

Corollary 4.2. *Suppose that the filters Ψ have uniform frequency concentration with parameters γ, ρ, and r_1 from Assumption 3.1. Then we have, for all $f \in L^2(\mathbb{R}^d)$, and for all $N \in \mathbb{N}$,*

$$W_N(f) \leq \max\left\{1, \frac{2\left(\left\lfloor \frac{d}{2} \right\rfloor + 1\right)}{\alpha \cdot r_1}\right\} \cdot D_\Psi \cdot \left(\|f\|_2^2 - \|f * \chi\|_2^2 \right) \cdot \alpha^N,$$

where, as before, $\alpha := \sqrt{1 - \frac{4\gamma}{(1+\gamma)^2} \cdot (1-\rho)^2}$. Asymptotically, we even have

$$W_N(f) \in \mathcal{O}\left(\alpha^{2N}\right).$$

Proof. Both statements (the specific bound and the asymptotic behavior) are immediate consequences of Theorem 3.9. To derive the specific upper bound, we choose (in the notation of the theorem)

$$\omega : (0,\infty) \to (0,\infty), \quad t \mapsto \sqrt{t} \qquad \text{and} \qquad \widehat{\vartheta} : \mathbb{R}^d \to \mathbb{R}, \quad \xi \mapsto \left(1 - \|\xi\|_2\right)_+^{\left\lfloor \frac{d}{2} \right\rfloor + 1}.$$

Clearly, ω is a strongly t-dominated weight. Further, ϑ satisfies the prerequisites of Theorem 3.5 with $|\widehat{\vartheta}(r_1^{-1}\cdot)| \leq |\widehat{\chi}|$, i.e., $C_{\chi,\vartheta} = r_1^{-1}$. Moreover, by Bernoulli's inequality we have, for all $\xi \in \mathbb{R}^d$,

$$1 - |\widehat{\vartheta}(\xi)|^2 \leq 2\left(\left\lfloor \frac{d}{2} \right\rfloor + 1\right) \cdot \|\xi\|_2 .$$

Finally, note that

$$\sum_{\psi \in \Psi} \omega^2(d_\psi) \cdot \|f * \psi\|_2^2 \leq \omega^2(D_\Psi) \cdot \sum_{\psi \in \Psi} \|f * \psi\|_2^2 = D_\Psi \cdot \left(\|f\|_2^2 - \|f * \chi\|_2^2 \right),$$

which concludes the first part of the corollary. The asymptotic rate also follows directly from Theorem 3.9, if we take

$$\omega : (0,\infty) \to (0,\infty), \quad t \mapsto t \qquad \text{and} \qquad \vartheta : \mathbb{R}^d \to \mathbb{R}, \quad x \mapsto e^{-\pi \cdot \|x\|_2^2}. \qquad \square$$

Remark 4.3. In the context of the above corollary, it holds that

$$\mathcal{F}L^2_\omega(\mathbb{R}^d) \subsetneq L^2(\mathbb{R}^d) = \mathcal{D}_\omega(\Psi; L^2(\mathbb{R}^d)).$$

This again shows the strength of Theorem 3.9, and stresses the role of the filter bank specificity in energy propagation in scattering networks.

4.2. Scattering with wavelets

In this section, we apply our results from Section 2 and Section 3 to filter banks that are wavelet-generated.

We begin by briefly reviewing scattering with directional wavelets as introduced in [2]. Let $a > 1$, and let $G < O_d(\mathbb{R})$ be a finite subgroup of rotations in $\mathbb{R}^d$ that comprises the reflection operator $-I \in G$.

Definition 4.4. We say that a pair of functions $(\psi, \phi) \in \left(L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)\right)^2$ is **admissible for (a,G)-wavelet scattering** if

$$|\widehat{\phi}(\xi)|^2 + \sum_{j=1}^{\infty} \sum_{M \in G} |\widehat{\psi}(a^{-j} \cdot M^{-1}\xi)|^2 = 1 \quad \text{a.e. } \xi \in \mathbb{R}^d. \tag{46}$$

Note that (46) entails that

$$\sum_{j \in \mathbb{Z}} \sum_{M \in G} |\widehat{\psi}(a^{-j} \cdot M^{-1}\xi)|^2 = 1 \quad \text{a.e. } \xi \in \mathbb{R}^d,$$

which also implicitly requires ψ to have at least one vanishing moment, i.e., $\widehat{\psi}(0) = 0$. Fix $J \in \mathbb{Z}$, and set

$$\Lambda(a,G,J) := \{a^j \cdot M \mid j \in \mathbb{Z}_{>-J}, M \in G\} \subseteq \mathrm{GL}_d(\mathbb{R}).$$

Let, for each $\lambda = a^j \cdot M \in \Lambda(a,G,J)$,

$$\psi_\lambda := D^1_{\lambda^T}\psi = a^{dj}\psi(a^j \cdot M^{-1}\,\cdot).$$

Moreover, set

$$\phi_J := D^1_{a^{-J}\cdot I}\phi = a^{-dJ}\phi(a^{-J}\,\cdot).$$

Finally, define

$$\mathrm{Wav}(\psi,a,G,J) := \left\{\psi_\lambda \;\middle|\; \lambda \in \Lambda(a,G,J)\right\}.$$

By construction, the filters $\mathrm{Wav}(\psi,a,G,J)$ together with the corresponding low-pass filter $\chi := \phi_J$ form a semi-discrete Parseval frame. In fact, we obtain the Littlewood-Paley condition (11) from the admissibility condition (46) by an index shift,

$$\begin{aligned}|\widehat{\phi}_J(\xi)|^2 + \sum_{\lambda\in\Lambda(a,G,J)} |\widehat{\psi}(\lambda^{-1}\cdot\xi)|^2 &= |\widehat{\phi}(a^J\cdot\xi)|^2 + \sum_{j=1-J}^{\infty}\sum_{M\in G} |\widehat{\psi}(a^{-j}\cdot M^{-1}\xi)|^2 \\ &= |\widehat{\phi}(a^J\cdot\xi)|^2 + \sum_{j=1}^{\infty}\sum_{M\in G} |\widehat{\psi}(a^{-j}\cdot M^{-1}(a^J\cdot\xi))|^2 = 1 \quad \text{a.e. } \xi\in\mathbb{R}^d.\end{aligned}$$

Definition 4.5. We say that the filters Ψ are **wavelet-generated** if there exist a finite subgroup $G < O_d(\mathbb{R})$ with $-I \in G$, $a > 1$, $J \in \mathbb{Z}$, and a pair $(\psi,\phi) \in \left(L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)\right)^2$ that is admissible for (a,G)-wavelet scattering such that $\Psi = \mathrm{Wav}(\psi,a,G,J)$.

Our main application of the negative findings from Section 2 is that energy propagation can be arbitrarily slow in wavelet scattering networks.

Corollary 4.6. *If the scattering filters Ψ are wavelet-generated, then for every $g \in L^2(\mathbb{R}^d)$ there exists a constant $C_g > 0$ with the following property:*

For any nonincreasing null-sequence $E = (E_N)_{N\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_{>0}$ there exists $f_{g,E} \in L^2(\mathbb{R}^d)$ such that

a) $\left\|f_{g,E}\right\|_2^2 \le C_g\cdot(1+E_1)$,
b) $\left\|f_{g,E}-g\right\|_2^2 \le C_g\cdot E_1$, *and*
c) $W_N(f_{g,E}) \ge \frac{\|g\|_2^2}{8}\cdot E_N$ *for all* $N\in\mathbb{N}$.

Furthermore,

$$Y_E := \left\{ f\in L^2(\mathbb{R}^d) \;\middle|\; W_N(f)\in\mathcal{O}(E_N)\right\}$$

is a countable union of nowhere dense sets in $L^2(\mathbb{R}^d)$. In particular, $L^2(\mathbb{R}^d)\setminus Y_E$ is dense in $L^2(\mathbb{R}^d)$.

Proof. The proof is a straightforward application of Corollary 2.8. Suppose that $\Psi = \mathrm{Wav}(\psi,a,G,J)$. Setting $A := a\cdot I_d$ gives $\sigma_{\min}(A) = a > 1$. Moreover, for all $m\in\mathbb{N}$, applying $D^1_{A^{-m}}$ to the filters Ψ results in an index shift, which leads to the required inclusion relation

$$\begin{aligned}\Psi_m := D^1_{A^{-m}}\Psi &= \left\{ D^1_{A^{-m}}\psi_\lambda \;\middle|\; \lambda\in\Lambda(a,G,J)\right\} \\ &= \left\{ D^1_{\lambda^T\cdot A^{-m}}\psi \;\middle|\; \lambda\in\Lambda(a,G,J)\right\} \\ &= \left\{\psi_\lambda \;\middle|\; \lambda\in\Lambda(a,G,J+m)\right\} \\ &\subseteq \left\{\psi_\lambda \;\middle|\; \lambda\in\Lambda(a,G,J+m+1)\right\} = \Psi_{m+1}. \quad \square\end{aligned}$$

The fact that for any nonincreasing null-sequence $E\in\mathbb{R}^{\mathbb{N}}_{>0}$ the set $L^2(\mathbb{R}^d)\setminus Y_E$, which contains those signals whose associated energy propagates *not* at the order of E through the scattering network, is dense in $L^2(\mathbb{R}^d)$ indicates that wavelet scattering energy propagation is not a stable property of signals. This is further emphasized by the contrast that fast energy propagation provably also holds for dense signal classes of $L^2(\mathbb{R}^d)$ if the generating wavelet is bandlimited.

Corollary 4.7. *Suppose that the scattering filters are wavelet-generated according to $\Psi = \mathrm{Wav}(\psi,a,G,J)$. Moreover, assume that there exist $M_\psi \in O_d(\mathbb{R})$, $\rho\in[0,1)$, and $\kappa\in\mathbb{N}_{\ge 2}$ such that*

$$\operatorname{supp}(\widehat{\psi}) \subseteq C^{\rho}_{M_\psi} \cap \mathcal{S}_{a^{-1},a^{\kappa-1}}.$$

Let

$$\alpha := \sqrt{1 - \frac{4a^\kappa}{(1+a^\kappa)^2} \cdot (1-\rho)^2} \in (0,1).$$

If $\omega : (0,\infty) \to (0,\infty)$ is a weakly t^2-dominated weight, then we have, for all $f \in \mathcal{D}_\omega(\Psi; L^2(\mathbb{R}^d))$,

$$W_N(f) \in \mathcal{O}(\omega^{-2}(\alpha^{-N})).$$

Specifically, we can guarantee the following convergence rates:

a) If $f \in H^s(\mathbb{R}^d)$ for some $s > 0$, then

$$W_N(f) \in \mathcal{O}(\alpha^{2\min\{s,1\}\cdot N}).$$

b) If $f \in H^s_{\log}(\mathbb{R}^d)$ for some $s > 0$, then

$$W_N(f) \in \mathcal{O}(N^{-2s}).$$

Proof. We only need to show that Assumption 3.1 is satisfied in this setting. To this end, we note that we have, for all $\lambda = a^j \cdot M \in \Lambda(a,G,J)$,

$$\operatorname{supp}(\widehat{\psi_\lambda}) \subseteq \lambda \cdot \operatorname{supp}(\widehat{\psi}) \subseteq \lambda \cdot \left(C^{\rho}_{M_\psi} \cap \mathcal{S}_{a^{-1},a^{\kappa-1}} \right) = C^{\rho}_{M\cdot M_\psi} \cap \mathcal{S}_{a^{j-1},a^{\kappa+j-1}}.$$

Thus, we may apply Theorem 3.9 and Corollary 3.12, where $\gamma = a^{-\kappa}$. □

Remark 4.8. Particularizing part a) of Corollary 4.7 for an overlap of $\kappa = 2$ and $\rho = 0$ in dimension $d = 1$ improves the currently best known upper bound on the asymptotic convergence rate for $f \in H^s(\mathbb{R})$ from

$$W_N(f) \in \mathcal{O}\left(\left(\frac{a^2-1}{a^2+1} \right)^{\min\{2s,1\}\cdot N} \right)$$

given by [19, Theorem 3.1] to

$$W_N(f) \in \mathcal{O}\left(\left(\frac{a^2-1}{a^2+1} \right)^{2\min\{s,1\}\cdot N} \right).$$

Remark 4.9. The semantic content of a generic signal occurring in classification tasks is typically stable under the action of small diffeomorphisms that deform signals [2,14]. The windowed scattering transform was introduced as a model for feature extraction, building translation invariant representations of signals in $L^2(\mathbb{R}^d)$ that take this form of stability into account.

One approach [23,3,25] to prove stability of the scattering transform under the action of small diffeomorphisms relies on the non-expansiveness of $\mathcal{S}[\mathfrak{F}]$. In this case, the stability ultimately results from the deformation sensitivity of the class of signals being considered, and thus naturally only applies to strict subclasses of $L^2(\mathbb{R}^d)$.

A different approach [2,15] explicitly takes into account the network architecture of scattering networks. Here, stability of the wavelet scattering transform under the action of small diffeomorphisms can be guaranteed for signals that have a finite mixed $(\ell^1, L^2(\mathbb{R}^d))$ scattering norm, i.e., for signals $f \in L^2(\mathbb{R}^d)$ that satisfy

$$\sum_{N=0}^{\infty} \left\| U[\Psi^N] f \right\|_{\ell^2(\Psi^N; L^2(\mathbb{R}^d))} = \sum_{N=0}^{\infty} \left(W_N[\Psi](f) \right)^{\frac{1}{2}} < \infty. \tag{47}$$

In general, it is not easy to decide whether a signal satisfies this condition. In light of Corollary 4.6, there exist dense subsets of $L^2(\mathbb{R}^d)$ for which the mixed $(\ell^1, L^2(\mathbb{R}^d))$ scattering norm (47) does not converge. On the other hand, if the generating wavelet of the filter bank is bandlimited, then Corollary 4.7 implies that the mixed $(\ell^1, L^2(\mathbb{R}^d))$ scattering norm is finite at least for all signals belonging to $H^s_{\log}(\mathbb{R}^d)$ for any $s > 1$. This generalizes [15, Proposition 2.4] to arbitrary dimension $d \in \mathbb{N}$.

5. Concluding remarks

Our paper focuses on the study of scattering transforms for the *continuous-domain setting*, i.e., transforms acting on the infinite-dimensional Hilbert space $L^2(\mathbb{R}^d)$. This raises the question of how relevant the results of our paper are for practical applications, necessarily dealing with finite implementations acting on finite signals or images.

At the outset of this discussion, we wish to point out that the use of continuous domains for the development and analysis of signal and image processing methods has a long-established tradition, and underlies many existing algorithms in the field. As an example of

a rather large direction of research founded on such ideas, let us mention variational signal and image processing, and name [26] as a particularly influential representative of a huge body of literature. Another case in point, which is somewhat closer to the origins of the scattering transform, is wavelet signal and image processing, in particular in the context of data compression [27,28].

We see at least two ways in which the continuous-domain setting can prove useful in understanding the finite signal setting. The first approach uses the continuous domain as a laboratory to develop ideas and intuitions, often based on mathematical theorems that exploit the rich analytical theory that is available for the continuous domain. The transfer to the finite, discrete-signal domain is then done mostly by analogy, rather than in the form of rigorous, quantitative results. The second, more rigorous approach tries to translate theorems about objects in the continuous-domain setting to theorems about suitable discretizations of these objects.

As an illustration of the first approach, we cite the invention of the scattering transform itself, which was initially conceived for the continuous domain. It is beneficial to recall the rationale for the introduction of the scattering transform as a designed (rather than learned) convolutional neural network feature extractor. The analytical properties of the continuous-domain scattering transform serve as a template for desirable properties one wishes to exploit in applications of its discrete-domain counterpart to finite signals. However, most of the relevant properties of the scattering features, such as robustness with respect to deformations, would have very cumbersome formulations in a strictly discrete, finite-domain setting. Nonetheless, one expects that the analytical properties derived in the continuous-domain setting are relevant for the discrete setting as well, and this expectation has been corroborated in case studies for a variety of applications [6,10,29].

We claim that some of the results and arguments in our paper have similarly useful interpretations in the discrete-domain setting as well. More precisely, we expect the intuitions that led to the construction of our adversarial signal class exhibiting arbitrarily slow energy decay to be of some use in the construction of similar signals in finite domain. The fundamental ideas behind our construction are that higher frequency components take longer to propagate, and that the scattering transform behaves almost additively on sums of signal components that are separated in frequency. It is conceivable (although beyond the scope of this paper) to use these intuitions to identify general signal classes with somewhat slow energy decay behavior for a finite, fixed number of layers.

However, the finite domain limits the extent to which both relevant techniques (dilating signal components to shift them to higher frequencies, and component separation) can be employed. In fact, we do not expect the negative results of our paper to carry over verbatim to the finite-domain setting. To mention a related (if not directly applicable) recent result, [30, Proposition 3.3] guarantees exponential energy decay for scattering transforms over finite graphs, globally for all input signals. In some sense, our Corollary 4.2 seems more directly relevant to the finite-domain setting: Recall that this result concludes exponential decay, as soon as the filterbank underlying the scattering transform has uniformly limited bandwidth. One can argue that over finite domains, this limited bandwidth condition is always fulfilled, and therefore the principle behind the proof of Corollary 4.2 (if not the proof itself) leads one to expect exponential decay.

We next sketch the second, more rigorous approach of making the improved understanding of the scattering transform provided by our results useful in the discrete context. The underlying idea is that discrete signals and images are understood as discretizations of continuous-domain objects, and one expects to employ well-established analytical tools to predict the result of signal processing algorithms acting on the discretization of a given continuous-domain signal, based on the analytical properties of the latter. The main benefit is that one expects to control the impact of data resolution; e.g., to explain how the processing of different discretized versions of the same continuous-domain signal is affected by their resolution. Examples of such approaches in the context of wavelet data compression (with different rationales) can be found in [27,28].

Transferring this reasoning to the question of energy decay for scattering transforms, one can speculate under what conditions a continuous-domain signal, for which the continuous-domain theory predicts a certain decay behavior in a given scattering transform, exhibits a comparable decay after discretizing and processing by a corresponding discrete-domain scattering transform, in a way that is *independent of the resolution*. It is important to realize that this last feature is not provided by the finite-domain decay result cited above, namely [30, Proposition 3.3]. This result formulates global exponential decay, but the basis of the exponential decay converges to one as the dimension of the signal space (which essentially corresponds to the resolution) becomes large. In other words, this decay behavior depends on the resolution.

From this point of view, our counterexamples serve as a caveat regarding the use of finite wavelet scattering transforms with a resolution-independent fixed depth. Note, however, that one possible way of justifying this choice consists in making the assumption that all practically relevant discrete signals are discretizations of continuous-time *signals from suitable function spaces*, such as the generalized Sobolev spaces that appear in our positive results. This assumption can be understood as an implicit bias imposed on the relevant signal classes. We point out that in the already mentioned sources on wavelet signal compression, the continuous-domain Besov spaces play a comparable role. Understanding whether such a bias exists, and how it is incurred by the choice of filter bank, has been one of the initial impulses of this work.

Acknowledgment

The authors acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 442047500 through the Collaborative Research Center “Sparsity and Singular Structures” (SFB 1481).

Data availability

No data was used for the research described in the article.

References

[1] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, M. Parmar, A review of convolutional neural networks in computer vision, Artif. Intell. Rev. 57 (4) (2024) 1–43.
[2] S. Mallat, Group invariant scattering, Commun. Pure Appl. Math. 65 (10) (2012) 1331–1398.
[3] T. Wiatowski, H. Bölcskei, A mathematical theory of deep convolutional neural networks for feature extraction, IEEE Trans. Inf. Theory 64 (3) (2017) 1845–1866.
[4] J. Su, J. Kempe, Wavelets beat monkeys at adversarial robustness, arXiv preprint, arXiv:2304.09403, 2023.
[5] X. Zhao, S. Zuo, Y. Mao, 3d scatternet: inference from 21 cm light-cones, in: 40th International Conference on Machine Learning, 2023.
[6] J. Andén, V. Lostanlen, S. Mallat, Joint time-frequency scattering for audio classification, in: 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2015, pp. 1–6.
[7] J. Chew, M. Hirn, S. Krishnaswamy, D. Needell, M. Perlmutter, H. Steach, S. Viswanath, H.-T. Wu, Geometric scattering on measure spaces, Appl. Comput. Harmon. Anal. (2024) 101635.
[8] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, L. Thiry, Solid harmonic wavelet scattering for predictions of molecule properties, J. Chem. Phys. 148 (24) (2018).
[9] F. Gao, G. Wolf, M. Hirn, Geometric scattering for graph data analysis, in: International Conference on Machine Learning, PMLR, 2019, pp. 2122–2131.
[10] M. Hirn, S. Mallat, N. Poilvert, Wavelet scattering regression of quantum chemical energies, Multiscale Model. Simul. 15 (2) (2017) 827–863.
[11] P. Sinz, M.W. Swift, X. Brumwell, J. Liu, K.J. Kim, Y. Qi, M. Hirn, Wavelet scattering networks for atomistic systems with extrapolation of material properties, J. Chem. Phys. 153 (8) (2020).
[12] E. Tolley, Wavelet scattering networks for identifying radio galaxy morphologies, arXiv preprint, arXiv:2403.03016, 2024.
[13] M. Tschannen, T. Kramer, G. Marti, M. Heinzmann, T. Wiatowski, Heart sound classification using deep structured features, in: 2016 Computing in Cardiology Conference (CinC), IEEE, 2016, pp. 565–568.
[14] S. Mallat, Understanding deep convolutional networks, Philos. Trans. R. Soc. A, Math. Phys. Eng. Sci. 374 (2065) (2016) 20150203.
[15] F. Nicola, S.I. Trapasso, Stability of the scattering transform for deformations with minimal regularity, J. Math. Pures Appl. 180 (2023) 122–150.
[16] I. Waldspurger, Exponential decay of scattering coefficients, in: 2017 International Conference on Sampling Theory and Applications (SampTA), IEEE, 2017, pp. 143–146.
[17] W. Czaja, W. Li, Analysis of time-frequency scattering transforms, Appl. Comput. Harmon. Anal. 47 (1) (2019) 149–171.
[18] T. Wiatowski, P. Grohs, H. Bölcskei, Energy propagation in deep convolutional neural networks, IEEE Trans. Inf. Theory 64 (7) (2017) 4819–4842.
[19] T. Wiatowski, P. Grohs, H. Bölcskei, Topology reduction in deep convolutional feature extraction networks, in: Wavelets and Sparsity XVII, vol. 10394, SPIE, 2017, pp. 269–280.
[20] S. Mallat, A Wavelet Tour of Signal Processing, Elsevier, 1999.
[21] E. DiBenedetto, Real Analysis, Springer, 2002.
[22] T. Wiatowski, M. Tschannen, A. Stanic, P. Grohs, H. Boelcskei, Discrete deep feature extraction: a theory and new architectures, in: M.F. Balcan, K.Q. Weinberger (Eds.), Proceedings of the 33rd International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 48, PMLR, New York, New York, USA, 2016, pp. 2149–2158.
[23] P. Grohs, T. Wiatowski, H. Bölcskei, Deep convolutional neural networks on cartoon functions, in: 2016 IEEE International Symposium on Information Theory (ISIT), IEEE, 2016, pp. 1163–1167.
[24] C. Heil, An Introduction to Weighted Wiener Amalgams, 2003, pp. 183–216.
[25] M. Koller, J. Großmann, U. Monich, H. Boche, Deformation stability of deep convolutional neural networks on Sobolev spaces, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 6872–6876.
[26] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, 60 (1992) 259–268, experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).
[27] R.A. DeVore, B. Jawerth, B.J. Lucier, Image compression through wavelet transform coding, IEEE Trans. Inf. Theory 38 (2) (1992) 719–746.
[28] D.L. Donoho, M. Vetterli, R.A. DeVore, I. Daubechies, Data compression and harmonic analysis, 44 (1998) 2435–2476, information theory: 1948–1998.
[29] J. Bruna, S. Mallat, Invariant scattering convolution networks, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1872–1886.
[30] D. Zou, G. Lerman, Graph convolutional neural networks via scattering, Appl. Comput. Harmon. Anal. 49 (3) (2020) 1046–1074.