POLITECNICO DI TORINO
Repository ISTITUZIONALE

Stability of the scattering transform for deformations with minimal regularity

Original
Stability of the scattering transform for deformations with minimal regularity / Nicola, F.; Trapasso, S. I.. - In: JOURNAL DE MATHÉMATIQUES PURES ET APPLIQUÉES. - ISSN 0021-7824. - 180:(2023), pp. 122-150. [10.1016/j.matpur.2023.10.008]

Availability:
This version is available at: 11583/2984023 since: 2023-11-23T08:21:28Z

Publisher:
Elsevier

Published
DOI:10.1016/j.matpur.2023.10.008

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

29 April 2024

J. Math. Pures Appl. 180 (2023) 122–150

Contents lists available at ScienceDirect

Journal de Mathématiques Pures et Appliquées

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

Stability of the scattering transform for deformations with minimal regularity

Fabio Nicola, S. Ivan Trapasso *

Dipartimento di Scienze Matematiche, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

ARTICLE INFO

Article history:
Received 7 October 2022
Available online 7 November 2023

MSC:
94A12
42C40
42C15
42B35
68T07
68T05

Keywords:
Scattering transform
Stability
Deformations
Multiresolution analysis
Wavelets
Mathematics of artificial intelligence

ABSTRACT

The wavelet scattering transform introduced by Stéphane Mallat is a unique example of how the ideas of harmonic and multiscale analysis can be ingeniously exploited to build a signal representation with provable geometric stability properties, such as the Lipschitz continuity to the action of small C^2 diffeomorphisms – a remarkable result for both theoretical and practical purposes, inherently depending on the choice of the filters and their arrangement into a hierarchical architecture. In this note, we further investigate the intimate relationship between the scattering structure and the regularity of the deformation in the Hölder regularity scale C^α, $\alpha > 0$. We are able to precisely identify the stability threshold, proving that stability is still achievable for deformations of class C^α, $\alpha > 1$, whereas instability phenomena can occur at lower regularity levels modeled by C^α, $0 \leq \alpha < 1$. While the analysis at the threshold given by Lipschitz (or even C^1) regularity remains beyond reach, we are able to prove a stability bound in that case, up to ε losses.

© 2023 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

RÉSUMÉ

La transformée de *scattering* par ondelettes introduite par Stéphane Mallat est un exemple unique de la façon dont les idées d'analyse harmonique et multi-échelle peuvent être ingénieusement exploitées pour construire une représentation du signal avec des propriétés de stabilité géométrique démontrables, telles que la continuité Lipschitzienne par petite difféomorphismes de classe C^2 – un résultat remarquable tant d'un point de vue pratique que théorique, dépendant étroitement du choix des filtres et de leur agencement dans une architecture hiérarchique. Dans cette note, nous explorons davantage la relation intime entre la structure de *scattering* et la régularité de la déformation dans l'échelle de régularité de Hölder C^α, $\alpha > 0$. Nous pouvons identifier précisément le seuil de stabilité, en prouvant que la stabilité est encore atteignable pour les déformations de classe C^α, $\alpha > 1$, tandis que des phénomènes d'instabilité peuvent survenir à des niveaux de régularité plus bas modélisés par C^α, $0 \leq \alpha < 1$. Alors que l'analyse au seuil donné par la régularité

* Corresponding author.
E-mail addresses: fabio.nicola@polito.it (F. Nicola), salvatore.trapasso@polito.it (S.I. Trapasso).

https://doi.org/10.1016/j.matpur.2023.10.008
0021-7824/© 2023 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

de Lipschitz (voire C^1) reste hors de portée, nous prouvons un résultat de stabilité dans ce cas à moins de ε pertes.
© 2023 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Broadly speaking, the last decade was certainly marked by a striking series of successes in several machine learning tasks relying on neural networks [15]. In particular, impressive results in image classification, pattern recognition and feature extraction were achieved by means of deep convolutional neural networks. Borrowing from Wigner, the efforts of many researchers are currently directed to provide explanations for the "unreasonable effectiveness" of these models and related intriguing phenomena, such as the double descent error curve [3,12,21,23] or the instability to adversarial attacks [1,8,10,14,28].

The mathematical analysis of convolutional neural networks is a wide area of current interest in the literature. The present note fits into a line of research pioneered by Stéphane Mallat, ultimately aimed at showing how some fundamental principles of harmonic analysis can be used to obtain theoretical models and guarantees in connection with problems of deep learning. Motivated by some properties naturally expected to be satisfied by a proper feature extractor, in the fundamental contribution [20] it is shown how such conditions essentially force the design of a multiscale signal representation to have a hierarchical architecture that shares many similarities with that of a convolutional neural network.

Let us briefly retrace here the basic ideas behind the construction for the sake of clarity. Motivated by image analysis, the goal is to build up a *feature map* $\Phi\colon L^2(\mathbb{R}^d) \to \mathcal{H}$, with values in a suitable Hilbert space $\mathcal{H}$, such that:

1. Φ *is a nonexpansive transform.*
   This condition ensures stability to additive perturbations, that is
$$\|\Phi(f) - \Phi(h)\| \le \|f - h\|_{L^2}, \quad f, h \in L^2(\mathbb{R}^d).$$
2. Φ *is a translation-invariant transform.*
   Let L_x be the translation operator by $x \in \mathbb{R}^d$, acting on $f \in L^2(\mathbb{R}^d)$ as $L_x f(y) = f(y - x)$. Then
$$\Phi(L_x f) = \Phi(f), \quad f \in L^2(\mathbb{R}^d), \quad x \in \mathbb{R}^d.$$
3. Φ *is stable to the action of small diffeomorphisms.*
   A convenient linearization of the action of a diffeomorphism along the orbits of the translation group leads one to consider deformation operators of the form $L_\tau f(y) \coloneqq f(y - \tau(y))$ with distortion field $\tau\colon \mathbb{R}^d \to \mathbb{R}^d$. Stability is achieved if the feature vectors of $L_\tau f$ and f are close when the underlying diffeomorphism $1 - \tau$ is close to identity, namely if there exists $C > 0$ such that
$$\|\Phi(L_\tau f) - \Phi(f)\| \le CK(\tau)\|f\|_{L^2}, \quad f \in L^2(\mathbb{R}^d)$$
   where $K(\tau)$ is some complexity measure/cost associated with the deformation τ.

1.1. The wavelet scattering transform

The approach in [20] relies on the a priori exploitation of the principles of multiscale analysis in order to satisfy the requirements detailed above. It is indeed well understood that instability to deformations is

mostly attributable to the vulnerability of the high-frequency components of a signal, which however carry fine-structure details and hence cannot be discarded without deteriorating the quality of the representation Φ. A Littlewood-Paley wavelet transform [19,22] can be used to rearrange the frequency content of a signal into dyadic packets. Thanks to inherent redundancy and additional nonlinear operations, this procedure allows one to obtain stability guarantees to relatively small translations up to a certain scale. Recovery of the spectral features discarded by a fixed scale wavelet transform is then achieved by iteration of the same procedure on the outputs of the latter, ultimately leading to a cascade of convolutions with fixed wavelet filters and modulus nonlinearities that eventually has the multilayer architecture of a convolutional neural network. The pooling stage is performed by extracting low-frequency averages of each scattered wavelet coefficient, and actually coincides with output feature generation.

An essential yet more detailed discussion of this construction is provided in Section 2, where we also fix the notation used below. Here we just recall that a low-pass filter ϕ and a mother wavelet ψ on $\mathbb{R}^d$ are primarily chosen in such a way that the collection $\{\phi_{2^J}\} \cup \{\psi_\lambda\}_{\lambda\in\Lambda_J}$ obtained by suitable rotations and dilations up to the scale 2^J, $J \in \mathbb{Z}$ (see (2.4) for the precise definition of the index set Λ_J), allows one to essentially cover the frequency space without holes – as entailed by the Littlewood-Paley condition (2.5) below. The wavelet modulus coefficient corresponding to $\lambda \in \Lambda_J$ is given by $U[\lambda]f = |f * \psi_\lambda|$. The cascading sequence that we mentioned before is obtained by iteration along all the possible paths with finite length, namely $\mathcal{P}_J = \bigcup_{m\geq 0} \Lambda_J^m$, so that given $p = (\lambda_1, \dots, \lambda_m) \in \Lambda_J^m$ we set

$$U[p]f \coloneqq U[\lambda_m] \cdots U[\lambda_1]f.$$

The windowed wavelet scattering transform at scale 2^J is thus the collection (indexed by $\mathcal{P}_J$) of features obtained by averaging with the low-pass filter ϕ_{2^J} at the scale 2^J:

$$S_J[\mathcal{P}_J]f \coloneqq \{S_J[p]f\}_{p\in\mathcal{P}_J}, \quad S_J[p]f \coloneqq U[p]f * \phi_{2^J}.$$

The feature space corresponds to $\mathcal{H} = \ell^2(\mathcal{P}_J; L^2(\mathbb{R}^d))$, hence

$$\|S_J[\mathcal{P}_J]f\|^2 = \sum_{p\in\mathcal{P}_J} \|S_J[p]f\|_{L^2}^2.$$

Concerning the stability to small deformations, it was proved in [20, Theorem 2.12] that, under suitable assumptions on the frequency filters (see Section 2.3 below for details), for every input signal f with finite mixed ℓ^1/L^2 scattering norm, that is

$$\|U[\mathcal{P}_J]f\|_1 \coloneqq \sum_{m\geq 0} \Big(\sum_{p\in\Lambda_J^m} \|U[p]f\|_{L^2}^2 \Big)^{1/2} < \infty,$$

and for every deformation $\tau \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ with $\|D\tau\|_{L^\infty} \leq 1/2$, the following stability estimate holds:

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \leq CK_2(\tau)\|U[\mathcal{P}_J]f\|_1, \tag{1.1}$$

with

$$K_2(\tau) = 2^{-J}\|\tau\|_{L^\infty} + \max\Big\{\log \frac{\|\Delta\tau\|_{L^\infty}}{\|D\tau\|_{L^\infty}}, 1\Big\}\|D\tau\|_{L^\infty} + \|D^2\tau\|_{L^\infty},$$

where $\|\Delta\tau\|_{L^\infty} \coloneqq \sup_{x,y\in\mathbb{R}^d} |\tau(x) - \tau(y)|$ and $D^2\tau$ stands for the Hessian of τ.

Some remarks are in order here. First, this estimate implies stability under small C^2 deformations, as well as approximate invariance to global translations up to the scale 2^J.

Concerning the occurrence of the scattering norm, it is proved in [20, Lemma 2.8] that a similar ℓ^2/L^2 norm is finite for functions with a certain average modulus of continuity in L^2, in particular for functions with logarithmic-Sobolev regularity. It is also worthwhile to point out that numerical evidences of exponential decay of the scattering energy coefficients were rigorously confirmed (at least in dimension $d = 1$) in [32]. The latter results also imply that 1-dimensional signals with a (generalized) logarithmic-Sobolev regularity have indeed finite ℓ^1/L^2 scattering norm (see Proposition 2.5 below).

It should be highlighted that one can also restrict to more regular signal classes, such as Sobolev spaces or band-limited and cartoon functions. The underlying gain in signal regularity usually comes along with some degree of stability to small deformations – namely, L^2 sensitivity bounds of the form $\|L_\tau f - f\|_{L^2} = O(K(\tau))$ are satisfied for suitably small and regular deformations, see e.g. [34,33]. In view of the Lipschitz continuity of the feature extractor, the latter bounds reflect into stability results for the signal representation, in a sense "inherited" from the sensitivity to deformations of the underlying signal class [2,4,9,36]. On the other hand, the estimate (1.1) entails the more difficult problem of deriving "structural" stability guarantees from the very design of the feature extractor, which are thus informative on the invariance of the signal representation rather than the regularity of the signal itself.

Putting aside these complementary views on the issue, let us observe that the condition $\|D\tau\|_{L^\infty} \leq 1/2$ suffices to ensure that $I - \tau$ is a bi-Lipschitz map and $L_\tau \colon L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is well-defined and uniformly bounded.[1] More precisely, if f is concentrated in a certain frequency dyadic band, $L_\tau f$ is essentially concentrated in the same band as well and this suggests that such deformations should interact well with the transform architecture, which is adjusted to such bands by design. Nevertheless, we will see that some instability phenomena may occur if $\|D\tau\|_{L^\infty} \not\to 0$.

1.2. A regularity scale for deformations

The purpose of this note is to elucidate the intimate relationship between the scattering architecture and the regularity of the deformation, lying at the very core of the "structural" stability for the wavelet scattering transform. To this aim, we consider distortion fields in the C^α regularity scale, $\alpha > 0$ (Hölder classes, recalled in Section 2 below), hence encompassing the case $\alpha = 2$ already studied in [20]. The quest for the minimal deformation regularity needed to achieve stability guarantees is an intriguing and natural challenge from a mathematical point of view, further motivated by the current practice in several problems in PDEs and image analysis [26,31,35] where diffeomorphisms with lower regularity are taken into account – for instance, Sobolev deformations $\tau \in H^s(\mathbb{R}^d; \mathbb{R}^d)$ with $s > d/2 + 1$, hence in $C^{s-d/2}(\mathbb{R}^d; \mathbb{R}^d)$. Stability results for the scattering transform under such weaker regularity assumptions for the deformation would then broaden the theoretical and practical scope of this mathematical theory.

As a first result we highlight the following instability phenomenon, when $\tau \to 0$ in C^α, for $0 \leq \alpha < 1$, but not in the C^1 norm. We assume here $d = 1$.

Theorem 1.1. *Suppose that the filters $\phi, \psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ in the definition of the scattering transform satisfy the Littlewood-Paley condition (2.5) below. Assume, in addition, that ψ has Fourier transform $\widehat{\psi}$ compactly supported in $(0, +\infty)$.*

There exist $\tau, f \in C^\infty(\mathbb{R}; \mathbb{R}) \setminus \{0\}$ with compact support and satisfying $\|\tau'\|_{L^\infty} \leq 1/2$ such that the following holds true.

There exists $C > 0$ such that, for every $J \in \mathbb{Z}$, $n \in \mathbb{N}$, setting $f_n(x) = 2^{n/2} f(2^n x)$ and $\tau_n(x) = 2^{-n}\tau(2^n x)$,

[1] Indeed, for every $y \in \mathbb{R}^d$, the map $\mathbb{R}^d \to \mathbb{R}^d$ given by $x \mapsto y + \tau(x)$ is a contraction, with Lipschitz constant $L \leq 1/2$. The map that associates y with the corresponding unique fixed point x has Lipschitz constant $\leq 1/(1 - L) \leq 2$.

$$\|S_J[\mathcal{P}_J](L_{\tau_n} f_n) - S_J[\mathcal{P}_J](f_n)\| \geq C. \tag{1.2}$$

As a consequence, for $0 \leq \alpha < 1$*, there exists* $C > 0$ *such that, for every* $J \in \mathbb{Z}$*,* $n \in \mathbb{N}$*,*

$$\|S_J[\mathcal{P}_J](L_{\tau_n} f_n) - S_J[\mathcal{P}_J](f_n)\| \geq C 2^{n(1-\alpha)} \|\tau_n\|_{C^\alpha} \|f_n\|_{L^2}. \tag{1.3}$$

Results in the same spirit hold as well if the L^2-norm is replaced by the scattering norm $\|U[\mathcal{P}_J]f\|_1$, see Proposition 3.1 below.

Notice that the functions $\tau_n \in C^\infty(\mathbb{R};\mathbb{R})$ are all supported in a fixed compact interval, say $I \subset \mathbb{R}$. Moreover, $\|\tau_n'\|_{L^\infty} \leq 1/2$ for every n, and $\|\tau_n\|_{C^\alpha} \to 0$ as $n \to \infty$ for every $0 \leq \alpha < 1$ (by (2.1) below).

To better frame the previous result, consider the set

$$\mathcal{B}_{1/2} = \{\tau \in C^\infty(\mathbb{R};\mathbb{R}) : \operatorname{supp}(\tau) \subset I,\ \|\tau'\|_{L^\infty} \leq 1/2\}, \tag{1.4}$$

equipped with the C^α metric,[2] $0 \leq \alpha < 1$. By the Taylor formula, if $\tau_1, \tau_2 \in \mathcal{B}_{1/2}$ then $\|L_{\tau_1} f - L_{\tau_2} f\|_{L^2} \lesssim \|\tau_1 - \tau_2\|_\infty \|f'\|_{L^2}$. Hence, since $S_J[\mathcal{P}_J]$ is nonexpansive, for every $J \in \mathbb{Z}$ and every *fixed* $f \in H^1(\mathbb{R})$ (Sobolev space) – in particular for each f_n as above – the map $\mathcal{B}_{1/2} \to \ell^2(\mathcal{P}_J; L^2(\mathbb{R}))$ given by $\tau \mapsto S_J[\mathcal{P}_J](L_\tau f)$ is Lipschitz continuous (cf. also [13]). On the other hand, Theorem 1.1 provides a lower bound for the blow-up rate of the Lipschitz constant, depending on α, when the input data become progressively less regular.

The instability results in Theorem 1.1 can be heuristically explained as follows. Consider a smooth signal f with unit L^2 norm. The deformed signal $L_\tau f$ has a certain low-frequency mass, but a relatively small energy bump in a quite far dyadic frequency band may occur even if $\|\tau'\|_{L^\infty} \leq 1/2$. The latter will propagate along different scattering paths, thus preventing the quantity $\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\|$ from being too small – assuming that $S_J[\mathcal{P}_J]$ preserves the norm, which is a consequence of the assumptions in Theorem 1.1. A simple scaling argument shows that the same phenomenon can happen even when $\|\tau\|_{C^\alpha} \to 0$, $0 \leq \alpha < 1$, along with a corresponding loss of regularity for f.

To summarize, Theorem 1.1 and Proposition 3.1 below show that, as far as the Lipschitz continuity under C^α deformation is concerned, the threshold $\alpha = 1$ is critical, both for functions in L^2 and for functions with finite scattering norm. On the other hand, we have the positive result (1.1) in the case $\alpha = 2$. The following stability result essentially fills this gap – we assume the same condition on the filters as in [20] (see Section 2.3).

Theorem 1.2. *Consider* $0 < \alpha < 1$*. There exists a constant* $C > 0$ *such that, for all* $J \in \mathbb{Z}$*,* $f \in L^2(\mathbb{R}^d)$ *with* $\|U[\mathcal{P}_J]f\|_1 < \infty$*, and* $\tau \in C^{1+\alpha}(\mathbb{R}^d;\mathbb{R}^d)$*, with* $\|D\tau\|_{L^\infty} \leq 1/2$*,*

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \leq C K_{1+\alpha}(\tau) \|U[\mathcal{P}_J]f\|_1, \tag{1.5}$$

with

$$K_{1+\alpha}(\tau) = 2^{-J}\|\tau\|_{L^\infty} + \max\Big\{\log \frac{\|\Delta\tau\|_{L^\infty}}{\|D\tau\|_{L^\infty}}, 1\Big\} \|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}.$$

The definition of the C^α seminorm $|\cdot|_{C^\alpha}$ is recalled in Section 2. This result arises as a refinement of [20, Theorem 2.12], with which it shares the backbone structure of the proof. A careful inspection of the latter suggests that lower levels of deformation complexity (such as logarithmic Hölder regularity) could still give rise to stability results. A substantial rearrangement of some parts of the proof strategy is expected to

[2] On $\mathcal{B}_{1/2}$ the C^α *topology*, $0 \leq \alpha < 1$ (but not the metric) is equivalent to the C^0 topology, because of the interpolation inequalities (2.1) below.

accommodate even lower regularity levels, such as Dini continuous deformations. In any case, we preferred to keep the technicalities at a minimum and to use the more natural C^α scale, also in view of applications.

The combination of the previous results provides us with a substantially complete picture on the interplay between stability and deformation regularity. Notably, the case of Lipschitz (or even $C^1(\mathbb{R}^d;\mathbb{R}^d)$) distortions remains open. A dimensional argument shows that, for $f \in L^2(\mathbb{R}^d)$, the expected bound would have the form

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \le C\big(2^{-J}\|\tau\|_{L^\infty} + \|D\tau\|_{L^\infty}\big)\|f\|_{L^2}. \tag{1.6}$$

While proving this estimate is definitely an ambitious goal, this problem seems to be out of reach at the current time. Interestingly, we are able to show that it holds up to arbitrarily small losses, at least in dimension 1. As customary in harmonic analysis, to accomplish this goal we consider the case of band-limited functions f, with $\widehat{f}$ supported in the frequency ball $|\omega| \le R$, say, and determine the blow-up rate in the above regime as $R \to +\infty$. The following stability result for Lipschitz deformations shows that such a rate is indeed smaller than R^ε for every $\varepsilon > 0$.

First, we assume that there are $C, \beta > 0$ such that

$$\|U[\mathcal{P}_0]f\|_1 \le C\log^\beta(e+R)\|f\|_{L^2} \tag{1.7}$$

for every $f \in L^2(\mathbb{R}^d)$ with $\widehat{f}(\omega)$ supported in the ball $|\omega| \le R$. Such an estimate holds in dimension $d = 1$, for every $\beta > 1$, as a consequence of Proposition 2.5 below under an admissibility condition on the filters detailed in [32]. There is reason to believe that such a logarithmic bound holds in arbitrary dimension (cf. for instance [20, Lemma 2.8] and the related remarks).

Theorem 1.3. *Assume* (1.7). *For every $\varepsilon > 0$ there exists $C > 0$ such that, for every $\tau\colon \mathbb{R}^d \to \mathbb{R}^d$ bounded and globally Lipschitz, with $\|D\tau\|_{L^\infty} \le 1/2$, and every $f \in L^2(\mathbb{R}^d)$ with $\widehat{f}(\omega)$ supported in the ball $|\omega| \le R$, $R > 0$, and every $J \in \mathbb{Z}$, we have*

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \le C\big(\log^\beta(e+2^JR)2^{-J}\|\tau\|_{L^\infty} + (1+2^JR)^\varepsilon\|D\tau\|_{L^\infty}\big)\|f\|_{L^2}. \tag{1.8}$$

The proof is based on a nonlinear interpolation argument in the setting of Besov spaces, which is in turn a refinement of a classical technique that has already been successfully developed in the literature in connection with nonlinear estimates for PDEs [16,18,24,29].

To conclude, we observe that it would be also very interesting to investigate similar stability issues for scattering-type transforms associated with other semi-discrete frames, such as curvelet or shearlet systems [7,11], in view of their prominent role in image processing. Also, from a mathematical perspective, it is natural to wonder whether the above results are robust enough to encompass more general operators than L_τ, Fourier integral operators being the natural candidates [6]. We postpone the study of these problems, that require novel ideas and techniques, to future works.

2. Preliminaries and review of the scattering transform

2.1. Notation

The open ball of $\mathbb{R}^d$ centered at x_0 with radius $r > 0$ is denoted by $B_r(x_0)$. For a differentiable map $\tau\colon \mathbb{R}^d \to \mathbb{R}^d$, we denote by $D\tau(x)$ its derivative as a linear map $\mathbb{R}^d \to \mathbb{R}^d$, hence we write $|D\tau(x)|$ for the operator norm of this map and also set $\|D\tau\|_{L^\infty} = \||D\tau|\|_{L^\infty}$. Similarly, for a scalar-valued function f, $\|\nabla f\|_{L^\infty} = \||\nabla f|\|_{L^\infty}$.

The Fourier transform of f is normalized here as

$$\widehat{f}(\omega) = \mathcal{F}(f)(\omega) = \int_{\mathbb{R}^d} e^{-i\omega\cdot x} f(x)\, dx.$$

Given an index set Ω and a collection of operators $T[p]\colon L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ indexed by $p \in \Omega$, we set

$$T[\Omega] = \{T[p]\}_{p\in\Omega}.$$

Unless otherwise stated, the standard norm in this context is that of $\ell^2(\Omega; L^2(\mathbb{R}^d))$, namely

$$\|T[\Omega]f\|^2 = \sum_{p\in\Omega} \|T[p]f\|_{L^2}^2, \quad f \in L^2(\mathbb{R}^d).$$

In the proofs, for brevity, we will heavily make use of the symbol $A \lesssim B$, meaning that the underlying inequality holds up to a positive constant factor, namely

$$A \lesssim B \quad \Longrightarrow \quad \exists\, C > 0 \,:\, A \le CB.$$

If the constant $C = C(\nu)$ depends on some parameter ν we write $A \lesssim_\nu B$. Moreover, $A \approx B$ means that A and B are *equivalent quantities*, that is both $A \lesssim B$ and $B \lesssim A$ hold.

In the rest of the note, all the derivatives are to be understood in the distribution sense, unless otherwise noted.

2.2. Relevant function spaces

Consider an open subset $A \subseteq \mathbb{R}^d$ and set $Y = \mathbb{R}^n$ or $Y = \mathbb{C}$. Given a nonnegative integer k we introduce the space $C^k(A; Y)$ of all the continuously differentiable functions $f\colon A \to Y$ with bounded derivatives up to order k, with the natural norm $\|f\|_{C^k(A)} := \max_{|\beta|\le k} \sup_{x\in A} |\partial^\beta f(x)|$.

We define the α-Hölder seminorm, $0 < \alpha < 1$, and the Lipschitz seminorm of $f\colon A \to Y$ by

$$|f|_{C^\alpha(A)} := \sup_{\substack{x,y\in A\\ x\neq y}} \frac{|f(x) - f(y)|}{|x-y|^\alpha}, \qquad |f|_{\mathrm{Lip}(A)} := \sup_{\substack{x,y\in A\\ x\neq y}} \frac{|f(x) - f(y)|}{|x-y|}.$$

The space $C^\alpha(A; Y)$, $\alpha > 0$, consists of all the functions $f\colon A \to Y$, continuously differentiable up to the order $[\alpha]$ (integer part of α), such that

$$\|f\|_{C^\alpha(A)} := \|f\|_{C^{[\alpha]}(A)} + \sum_{|\gamma|=[\alpha]} |\partial^\gamma f|_{C^{\alpha-[\alpha]}(A)} < \infty.$$

When there is no risk of confusion we usually omit the codomain Y and also the domain in the case where $A = \mathbb{R}^d$, writing for instance C^α in place of $C^\alpha(\mathbb{R}^d; Y)$ for simplicity. We also recall the elementary interpolation inequality

$$|f|_{C^\alpha} \le 2^{1-\alpha} \|f\|_{L^\infty}^{1-\alpha} \|\nabla f\|_{L^\infty}^{\alpha}. \tag{2.1}$$

We collect here some basic properties that will be used below.

Proposition 2.1. *Fix $A \subset \mathbb{R}^d$ and $0 < \alpha < 1$.*

- *(Fractional Leibniz rule) If $f, g \in C^\alpha(A;\mathbb{C})$, then*

$$|fg|_{C^\alpha(A)} \le |f|_{C^\alpha(A)}\|g\|_{L^\infty(A)} + \|f\|_{L^\infty(A)}|g|_{C^\alpha(A)}. \tag{2.2}$$

- *(Schauder estimates) Assume that $F\colon \mathbb{R}^n \to \mathbb{C}$ is Lipschitz. For $h \in C^\alpha(A;\mathbb{R}^n)$,*

$$|F(h)|_{C^\alpha(A)} \le |F|_{\mathrm{Lip}(h(A))}|h|_{C^\alpha(A)}. \tag{2.3}$$

Proof. The fractional Leibniz rule is readily obtained by noting that for all $x, y \in \mathbb{R}^d$, $x \neq y$,

$$\frac{|f(x)g(x) - f(y)g(y)|}{|x-y|^\alpha} \le \frac{|f(x)-f(y)|}{|x-y|^\alpha}|g(x)| + |f(y)|\frac{|g(x)-g(y)|}{|x-y|^\alpha}.$$

The inequality in (2.3) follows similarly – as long as $h(x) \neq h(y)$,

$$\frac{|F(h(x)) - F(h(y))|}{|x-y|^\alpha} \le \frac{|F(h(x)) - F(h(y))|}{|h(x)-h(y)|}\frac{|h(x)-h(y)|}{|x-y|^\alpha}. \quad \square$$

2.3. A brief review of the wavelet scattering transform

In this section we gather some basic facts and results concerning the mathematical analysis of the scattering transform, mainly in order to fix the notation. More details can be found in [5,20].

The basic ingredient is a complex wavelet $\psi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ with at least one vanishing moment ($\widehat{\psi}(0) = 0$), satisfying appropriate conditions that are stated below.

Let G be a finite subgroup of rotations in $\mathbb{R}^d$, also comprising the reflection operator $-I$. For every $\lambda = 2^j r$ with $j \in \mathbb{Z}$ and $r \in G$ we set

$$\psi_\lambda(x) := 2^{jd}\psi(2^j r^{-1}x).$$

The frequency filtering corresponding to ψ_λ is thus obtained by convolution, namely for $f \in L^2(\mathbb{R}^d)$ we set

$$W[\lambda]f(x) := f * \psi_\lambda(x) = \int_{\mathbb{R}^d} f(y)\psi_\lambda(x-y)dy, \quad x \in \mathbb{R}^d.$$

If both f and $\widehat{\psi}$ are real functions it is easy to realize that $W[-\lambda]f = \overline{W[\lambda]f}$. We thus conveniently consider the quotient $G^+ = G/\{\pm I\}$, so that all the pairs of rotations r and $-r$ are identified.

Remark 2.2. As a concrete reference model one typically considers a Gabor-like wavelet ψ such as

$$\psi(x) = e^{i\eta\cdot x}\theta(x), \quad x \in \mathbb{R}^d,$$

for some $\eta \in \mathbb{R}^d$ and a function θ with real-valued Fourier transform $\widehat{\theta}$ essentially supported in a low-frequency ball centered at the origin with radius of the order of π. Then $\widehat{\psi_\lambda}(\omega) = \widehat{\psi}(2^{-j}r^{-1}\omega) = \widehat{\theta}(\lambda^{-1}\omega - \eta)$ is concentrated in a ball centered at $\lambda\eta$ of approximate size $|\lambda| := 2^j$.

Given $J \in \mathbb{Z}$ we introduce the index set

$$\Lambda_J := \{\lambda = 2^j r : j > -J,\, r \in G^+\}. \tag{2.4}$$

The filter bank $W[\Lambda_J] := \{W[\lambda]\}_{\lambda\in\Lambda_J}$ is thus not able to detect a low-frequency component of a real signal f corresponding to a region of the frequency space with size of the order of 2^{-J}. Nevertheless, the latter

can be captured by a suitable average A_J with a dilated low-pass filter $\phi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ such that ϕ is a non-negative real-valued function with $\widehat{\phi}(0) = 1$ (having in mind a Gaussian function as a model), that is

$$A_J f := f * \phi_{2^J}, \qquad \phi_{2^J}(x) := 2^{-Jd}\phi(2^{-J}x), \quad x \in \mathbb{R}^d.$$

To ensure that the frequency content of f is fully preserved by a wavelet analysis at a scale 2^J it is enough that the supports of the filters obtained by dilations of ϕ and ψ cover the whole frequency space. To be more precise, let $W_J f$ be the wavelet analysis of f associated with ϕ and ψ (Fig. 1), namely the collection of signal components indexed by $\{J\} \cup \Lambda_J$ given by

$$W_J f := \{A_J f, W[\Lambda_J]f\}.$$

It is not difficult to show that W_J is an isometry from $L^2(\mathbb{R}^d;\mathbb{R})$ (real-valued functions) to $\ell^2(\{J\} \cup \Lambda_J; L^2(\mathbb{R}^d))$ if and only if the following Littlewood-Paley condition holds for almost every $\omega \in \mathbb{R}^d$:

$$|\widehat{\phi}(2^J\omega)|^2 + \frac{1}{2}\sum_{\lambda\in\Lambda_J}\left[|\widehat{\psi}(\lambda^{-1}\omega)|^2 + |\widehat{\psi}(-\lambda^{-1}\omega)|^2\right] = 1. \tag{2.5}$$

Hence

$$\|f\|_{L^2}^2 = \|W_J f\|_{\ell^2 L^2}^2 := \|A_J f\|_{L^2}^2 + \sum_{\lambda\in\Lambda_J}\|W[\lambda]f\|_{L^2}^2$$

(see [20, Proposition 2.1] for further details). Slight modifications are needed in the case of complex-valued signals f in order to accommodate all the rotations. The wavelet analysis $W_J f$ is then accordingly defined including $W[-\Lambda_J]f$ as well, and unitarity of $W_J f$ is ensured by the condition

$$|\widehat{\phi}(2^J\omega)|^2 + \sum_{\lambda\in\Lambda_J}\left[|\widehat{\psi}(\lambda^{-1}\omega)|^2 + |\widehat{\psi}(-\lambda^{-1}\omega)|^2\right] = 1.$$

We simply discuss below the case where $f \in L^2(\mathbb{R}^d)$ takes real values to lighten the presentation.

We say that ϕ and ψ are *scattering filters* if:

- Given $J \in \mathbb{Z}$ and a group of rotations G as above, the condition in (2.5) is satisfied.
- $\widehat{\psi}$ is real-valued, and $\widehat{\phi}$ is real-valued and symmetric. Moreover, ϕ is non-negative and $\widehat{\phi}(0) = 1$.
- Both $\phi(x)$ and $\psi(x)$ are twice differentiable and decay like $O((1+|x|)^{-d-3})$ together with their first and second partial derivatives.[3]
- $\hat{\psi}(\xi) = 0$ for all $\xi = (\xi, \xi') \in \mathbb{R} \times \mathbb{R}^{d-1}$ such that $\xi_1 \le |\xi|/2$.

The last condition on the spectral conic support of ψ will be used below in connection with the following result.

Lemma 2.3. *Let $R \in \mathbb{R}^{d,d}$ be a matrix such that $\|R\| \le 1/2$. If $\hat{\psi}(\xi) = 0$ for all $\xi = (\xi_1, \xi') \in \mathbb{R} \times \mathbb{R}^{d-1}$ such that $\xi_1 \le |\xi|/2$, then*

$$\int_{\mathbb{R}} \psi((I+R)(x_1, x'))dx_1 = 0 \quad \textit{for all } x' \in \mathbb{R}^{d-1}.$$

[3] In [20] it is assumed a decay condition $O((1+|x|)^{-d-2})$ instead. Nevertheless, it seems that even in that case the decay of order $-(d+3)$ is needed, for instance in order to make the integral in [20, (E.26)] convergent or to suitably bound the last terms in [20, (E.30)] in such a way to obtain [20, (E.31)].

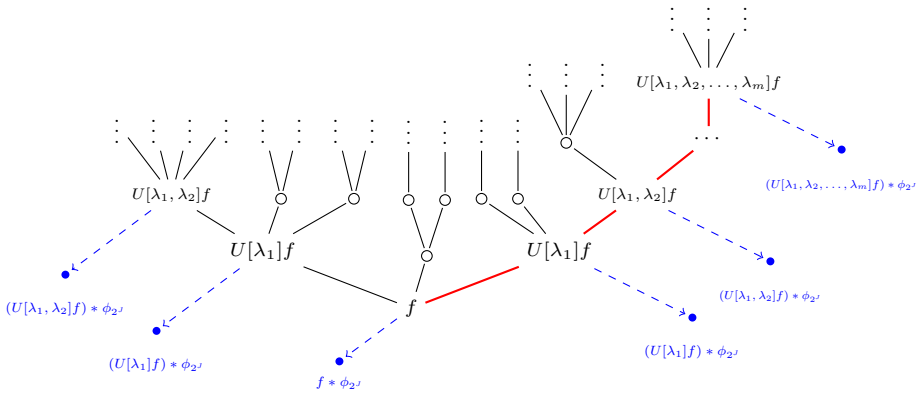

Fig. 1. The scattering network architecture, as described above. The index $\lambda_l \in \Lambda_J$ corresponds to the l-th layer. In blue: some features. In red: an example of a path $q = (\lambda_1, \lambda_2 \ldots, \lambda_m) \in \Lambda_J^m$ of length m.

Proof. In the case where $R = O$ the claim follows directly from the assumption, since $\hat{\psi}(0, \xi') = 0$ for all $\xi' \in \mathbb{R}^{d-1}$. Actually, we have in general $\hat{\psi}((I + R^\top)^{-1}(0, \xi')) = 0$ for all $\xi' \in \mathbb{R}^{d-1}$. By contradiction, if this were not the case, then $(0, \xi') = (I + R^\top)\eta$ for some $\eta = (\eta_1, \eta') \in \mathbb{R} \times \mathbb{R}^{d-1}$ with $\eta_1 > |\eta|/2$, but this cannot happen since

$$(I + R^\top)\eta \cdot e_1 = \eta_1 + R^\top \eta \cdot e_1 \geq \eta_1 - |\eta|/2 > 0,$$

where $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$. $\square$

We now introduce the general index set $\Lambda_\infty := 2^{\mathbb{Z}} \times G^+$ and the corresponding space $\mathcal{P}_\infty := \bigcup_{m \geq 0} \Lambda_\infty^m$ of all the possible finite paths, where it is understood that $\Lambda_\infty^0 = \{\emptyset\}$. The one-step scattering propagator $U[\lambda]$, $\lambda \in \Lambda_\infty$, coincides with a modulus wavelet localization: $U[\lambda]f := |W[\lambda]f| = |f * \psi_\lambda|$. More generally, the path-ordered scattering propagator $U \colon \mathcal{P}_\infty \times L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ acts along a path $p = (\lambda_1, \ldots, \lambda_m) \in \Lambda_\infty^m$ of length $m \geq 1$ by

$$U[p]f := U[\lambda_m] \cdots U[\lambda_2] U[\lambda_1] f.$$

For the empty path $p = \emptyset$ we set $U[\emptyset]f = f$.

The collection of all the paths with finite length and components in Λ_J is $\mathcal{P}_J := \bigcup_{m=0}^\infty \Lambda_J^m$ (again $\Lambda_J^0 = \{\emptyset\}$). The windowed scattering transform $S_J[\mathcal{P}_J]$ is then defined as follows:

$$S_J[\mathcal{P}_J]f = \{S_J[p]f\}_{p \in \mathcal{P}_J}, \quad S_J[p]f := A_J U[p]f.$$

The assumptions satisfied by the underlying scattering wavelets allow one to show that $S_J[\mathcal{P}_J]$ has the desired properties from a feature map with values in $\mathcal{H} = \ell^2(\mathcal{P}_J; L^2(\mathbb{R}^d))$ as discussed in the introduction.

Lipschitz regularity. It is proved in [20, Proposition 2.5] that $S_J[\mathcal{P}_J] \colon L^2(\mathbb{R}^d) \to \mathcal{H}$ is a nonexpansive transform, namely

$$\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\| \leq \|f - h\|_{L^2}, \quad f, h \in L^2(\mathbb{R}^d).$$

Norm preservation. Provided that the filters satisfy additional admissibility conditions (see [20, Theorem 2.6] or [32, Theorem 3.1] in dimension $d = 1$), $S_J[\mathcal{P}_J]$ preserves the norm of the input signal:

$$\|f\|_{L^2}^2 = \|S_J[\mathcal{P}_J]f\|^2 = \sum_{p \in \mathcal{P}_J} \|S_J[p]f\|_{L^2}^2, \quad f \in L^2(\mathbb{R}^d).$$

Translation invariance. It is proved in [20, Proposition 2.9] that the scattering distance $\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|$ is nonincreasing when J increases, and the scattering metric is asymptotically translation invariant, as proved in [20, Theorem 2.10]:

$$\lim_{J\to+\infty} \|S_J[\mathcal{P}_J](T_x f) - S_J[\mathcal{P}_J](f)\| = 0, \quad \forall x \in \mathbb{R}^d,\ f \in L^2(\mathbb{R}^d).$$

Stability to small deformations. As already anticipated in the introduction, the stability bound (1.1) is proved in [20, Theorem 2.12] for functions f such that

$$\|U[\mathcal{P}_J]f\|_1 = \sum_{m\geq 0} \|U[\Lambda_J^m]f\| < \infty.$$

Let us discuss some additional properties of the scattering transform which are used below.

Covariance properties. The joint action of scaling and rotation by $2^l g \in 2^{\mathbb{Z}} \times G$ on a signal f is given by $(2^l g \circ f)(x) := f(2^l gx)$, $x \in \mathbb{R}^d$, while for a path $p = (\lambda_1, \dots, \lambda_m) \in \mathcal{P}_\infty$ of length m we set $2^l gp := (2^l g\lambda_1, \dots, 2^l g\lambda_m)$. It is not difficult to show that the one-step propagator is somehow covariant to scaling and rotations, namely $U[\lambda](2^l g \circ f) = 2^l g \circ U[2^{-l} g\lambda]f$, $\lambda \in \Lambda_\infty$. In view of the cascading structure of the scattering transform, this property reflects into

$$U[p](2^l g \circ f) = 2^l g \circ U[2^{-l} gp]f, \quad p \in \mathcal{P}_\infty, \tag{2.6}$$

and

$$S_J[p](2^l g \circ f) = 2^l g \circ S_{J+l}[2^{-l} gp]f, \quad p \in \mathcal{P}_J. \tag{2.7}$$

Additivity on separated signals. The following simple result shows that $S_J[\mathcal{P}_J]$ is additive on functions that are separated in the wavelet domain.

Lemma 2.4. *Let $f, g \in L^2(\mathbb{R}^d)$ be such that, for every $\lambda \in \Lambda_J$,*

$$f * \psi_\lambda = 0 \quad \text{or} \quad g * \psi_\lambda = 0.$$

Then

$$S_J[\mathcal{P}_J](f + g) = S_J[\mathcal{P}_J](f) + S_J[\mathcal{P}_J](g).$$

Proof. Since the convolution with ϕ_{2^J} is a linear operator, it suffices to prove that

$$U[p](f + g) = U[p]f + U[p]g, \quad p \in \mathcal{P}_J.$$

Consider then $p = (\lambda_1, \dots, \lambda_m) \in \Lambda_J^m$. Since

$$U[\lambda_1, \dots, \lambda_m] = U[\lambda_2, \dots, \lambda_m]U[\lambda_1],$$

it is enough to show that, for every $\lambda \in \Lambda_J$, $U[\lambda](f + g) = U[\lambda]f + U[\lambda]g$ *and* one of the two terms on the right-hand side vanishes. The claim follows at once from the assumption and the definition of the one-step propagator $U[\lambda]f = |f * \psi_\lambda|$. $\square$

We finally present the following embeddings in dimension $d = 1$, obtained by means of the scattering decay results proved in [32]. The formula (2.8) is essentially known, cf. [20, Lemma 2.8] – the latter was proved in arbitrary dimension under a more restrictive admissibility condition on the wavelet ψ (see [20, Theorem 2.6]). The estimate (2.9) seems new.

Proposition 2.5. *Let $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ satisfy the following Littlewood-Paley inequality, for every $\omega \in \mathbb{R}$:*

$$\frac{1}{2}\sum_{j\in\mathbb{Z}}\Big[|\widehat{\psi}(2^{-j}\omega)|^2 + |\widehat{\psi}(-2^{-j}\omega)|^2\Big] \le 1.$$

Moreover, assume that

$$|\widehat{\psi}(-2^{-j}\omega)| \le |\widehat{\psi}(2^{-j}\omega)|$$

for $\omega > 0$ and $j \in \mathbb{Z}$, provided that for every ω the condition holds with strict inequality for at least one value of j.
Finally, assume that $|\widehat{\psi}(\omega)| = O(|\omega|^{1+\varepsilon})$ for some $\varepsilon > 0$, as $\omega \to 0$.

There exists $C > 0$ such that, for every $J \in \mathbb{Z}$,

$$\|U[\mathcal{P}_J]f\|^2 \le C\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \log(e + 2^J|\omega|)\, d\omega. \tag{2.8}$$

Moreover, for every $\beta > 2$ there exists $C > 0$ such that, for every $J \in \mathbb{Z}$,

$$\|U[\mathcal{P}_J]f\|_1 \le C\Big(\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \log^\beta(e + 2^J|\omega|)\, d\omega\Big)^{1/2}. \tag{2.9}$$

Proof of Proposition 2.5. First of all, we note that it suffices to prove the estimates (2.8) and (2.9) for $J = 0$. The claim then follows by a scaling argument. More precisely, consider $f_J(x) := 2^{J/2}f(2^Jx)$ and note that, by (2.6), we have $\|U[\Lambda_0^m]f_J\| = \|U[\Lambda_J^m]f\|$ and therefore $\|U[\mathcal{P}_0]f_J\| = \|U[\mathcal{P}_J]f\|$ as well.

By [32, Theorem 3.1], the assumptions in the statement imply that, for $m \ge 2$,

$$\|U[\Lambda_0^m]f\|^2 \le \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 A_m(\omega)\, d\omega, \tag{2.10}$$

where[4] $A_m(\omega) = 1 - e^{-2\omega^2/(ra^m)^2}$, for suitable $a > 1$, $r > 0$.

Since $\|U[\mathcal{P}_0]f\|^2 = \sum_{m=0}^\infty \|U[\Lambda_0^m]f\|^2$, and $\|U[\emptyset]f\|^2 + \|U[\Lambda_0]f\|^2 \le 2\|f\|_{L^2}^2$ ($U[\Lambda_0]$ being nonexpansive), in order to obtain (2.8) it is enough to verify that

$$\sum_{m=2}^\infty A_m(\omega) \le C\log(e + |\omega|/r) \tag{2.11}$$

for some constant $C > 0$, possibly depending on a only.

A straightforward change of variable shows that we can suppose $r = 1$ without loss of generality. The estimate (2.11) is satisfied if $|\omega| \le a^2$ because $A_m(\omega) \lesssim (\omega/a^m)^2$. On the other hand, if $|\omega| \ge a^2$ we conveniently split the sum in (2.11) in two parts accounting for $m \le N$ and $m > N$, where $N \ge 2$ is such

[4] The factor $1/(2\pi)$ in (2.10) does not appear in [32, Theorem 3.1] because of a different normalization of the Fourier transform. The constant r below will be different from that in [32, Theorem 3.1] as well, for the same reason.

that $a^N \leq |\omega| < a^{N+1}$. Using that $A_m(\omega) \leq 1$ and $A_m(\omega) \lesssim (\omega/a^m)^2 \lesssim a^{2(N-m)}$ in the two regimes, respectively, we obtain

$$\sum_{m=2}^{\infty} A_m(\omega) \lesssim \sum_{2\leq m\leq N} 1 + \sum_{m>N} a^{2(N-m)} \lesssim N \leq \log_a |\omega|,$$

which gives (2.11).

Let us now prove (2.9) with $J = 0$ in light of the previous arguments. By (2.10) we see that it is sufficient to prove the bound

$$\sum_{m=2}^{\infty} \Big(\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 A_m(\omega)\, d\omega \Big)^{1/2} \lesssim \Big(\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \log^{\beta}(e+|\omega|)\, d\omega \Big)^{1/2}. \tag{2.12}$$

Since $\beta > 2$, the latter will follow from the pointwise bound

$$\frac{A_m(\omega)}{\log^{\beta}(e+|\omega|)} \lesssim \frac{1}{m^{\beta}}.$$

This estimate clearly holds if $|\omega| \geq a^{m/2}$, since $A_m(\omega) \leq 1$. If $|\omega| < a^{m/2}$ we have

$$\frac{A_m(\omega)}{\log^{\beta}(e+|\omega|)} \leq A_m(\omega) \lesssim \frac{\omega^2}{a^{2m}} \leq \frac{1}{a^m} \lesssim \frac{1}{m^{\beta}}. \quad \square$$

Remark 2.6. It is worthwhile to point out that (2.12) does not hold for $\beta = 2$, as evidenced by the following example. Consider a function $f \in L^2(\mathbb{R})$ whose Fourier transform is supported in $[ra^2, +\infty) = \cup_{k\geq 2}\Omega_k$, with $\Omega_k = [ra^k, ra^{k+1})$, and takes a constant value on each Ω_k, adjusted so that $\int_{\Omega_k} |\widehat{f}(\omega)|^2 \log^2(e+|\omega|)\, d\omega = 1/(k \log^2 k)$. We then have

$$\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \log^2(e+|\omega|)\, d\omega < \infty.$$

On the other hand, if $k \geq m$, on Ω_k we have $A_m(\omega)/\log^2(e+|\omega|) \gtrsim 1/\log^2(e+|\omega|) \gtrsim 1/k^2$, so that

$$\begin{aligned}\sum_{m=2}^{\infty} \Big(\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 A_m(\omega)\, d\omega \Big)^{1/2} &\geq \sum_{m=2}^{\infty} \Big(\sum_{k\geq m} \int_{\Omega_k} |\widehat{f}(\omega)|^2 A_m(\omega)\, d\omega \Big)^{1/2} \\ &\gtrsim \sum_{m=2}^{\infty} \Big(\sum_{k\geq m} \frac{1}{k^3 \log^2 k} \Big)^{1/2} = \infty.\end{aligned}$$

Indeed, the latter series is readily seen to diverge, since

$$\sum_{k\geq m} \frac{1}{k^3 \log^2 k} \geq \int_m^{+\infty} \frac{1}{x^3 \log^2 x}\, dx \gtrsim \frac{1}{m^2 \log^2 m},$$

where we resorted to integration by parts in the last step.

3. Instability results for C^α regularity, $0 \le \alpha < 1$

This section is devoted to the instability phenomenon occurring for deformations with regularity C^α, $0 \le \alpha < 1$, already illustrated in the introduction. We begin with the proof of Theorem 1.1.

Proof of Theorem 1.1. Let us begin with the proof of (1.2). First of all, under the stated assumptions on ϕ, ψ we infer from [32, Theorem 3.1][5] that

$$\lim_{m\to\infty} \|U[\Lambda_J^m]f\| = 0$$

for every $f \in L^2(\mathbb{R})$. In view of the Littlewood-Paley condition (2.5), the latter result implies that $S_J[\mathcal{P}_J]$ preserves the norm (cf. the proof of [20, Theorem 2.6]).

Consider a compactly supported $f \in C^\infty(\mathbb{R};\mathbb{R})$ with $f(x) = x$ for $0 \le x \le 2\pi$. Let $\varphi \in C^\infty(\mathbb{R};\mathbb{R})$, supported in $[0, 2\pi]$, be such that $0 < \varphi(x) \le 1$ for $0 < x < 2\pi$.

Consider the deformation function defined by

$$\tau(x) = -\frac{A}{N}\sin(Nx)\varphi(x),$$

where $N \in \mathbb{N}$, $N \ge 1$, will be chosen later (large enough) and $A > 0$ is fixed in such a way that $A(1 + \|\varphi'\|_{L^\infty}) \le 1/2$, hence $\|\tau'\|_{L^\infty} \le 1/2$ for every N.

For future reference, we remark that

$$x - \tau(x) \in [0, 2\pi] \qquad \text{for } x \in [0, 2\pi]. \tag{3.1}$$

This follows from the fact that, since $\|\tau\|_{L^\infty} \le AN^{-1} \le \pi N^{-1}$, if x belongs to one of the $2N$ subintervals of $[0, 2\pi]$ where τ has constant sign (note that τ is negative over the first subinterval and positive over the last one), then $x - \tau(x)$ belongs either to the same interval or to an adjacent one.

We finally set $f_n(x) = 2^{n/2} f(2^n x)$ and $\tau_n(x) = 2^{-n}\tau(2^n x)$, $n \in \mathbb{N}$, as in the statement.

A convenient facilitation results from the fact that it suffices to prove the desired estimate (1.2) for $n = 0$, with a constant C_0 independent of $J \in \mathbb{Z}$. This can be readily inferred from the scaling property (2.7), yielding

$$\|S_J[\mathcal{P}_J](L_{\tau_n} f_n) - S_J[\mathcal{P}_J](f_n)\| = \|S_{J+n}[\mathcal{P}_{J+n}](L_\tau f) - S_{J+n}[\mathcal{P}_{J+n}](f)\|. \tag{3.2}$$

Let us thus set $n = 0$ hereafter. We are going to prove that, for $N \in \mathbb{N}$ large enough,

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \gtrsim \frac{1}{N} \tag{3.3}$$

where the hidden constants in the symbols $\gtrsim$, $\lesssim$ and $\approx$ are always independent of $J \in \mathbb{Z}$ and $N \in \mathbb{N}$.

Let us first discuss the strategy. The function f is concentrated in the frequency region where $|\omega| \lesssim 1$, while the function $L_\tau f - f$ will be shown to be concentrated where $|\omega - N| \lesssim 1$ or $|\omega + N| \lesssim 1$. Therefore, if

[5] The assumptions of Theorem [32, Theorem 3.1] are the same as those of Proposition 2.5 and are therefore satisfied here. Indeed, if (2.5) holds for the specified J, by rescaling one sees that it holds for every $J \in \mathbb{Z}$. Letting $J \to -\infty$, since $\widehat{\phi}$ tends to 0 at infinity we see that

$$\frac{1}{2}\sum_{j\in\mathbb{Z}} \left[|\widehat{\psi}(2^{-j}\omega)|^2 + |\widehat{\psi}(-2^{-j}\omega)|^2\right] = 1.$$

N is sufficiently large we have that $S_J[\mathcal{P}_J](L_\tau f)$ approximately coincides with $S_J[\mathcal{P}_J](L_\tau f - f) + S_J[\mathcal{P}_J](f)$ by virtue of Lemma 2.4, hence

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \approx \|S_J[\mathcal{P}_J](L_\tau f - f)\| \gtrsim \|L_\tau f - f\|_{L^2}.$$

Making a rigorous argument out of this clue necessarily comes through suitable bounds for the frequency tails of the functions f and $g := L_\tau f - f$. To this aim, let us start by noting that $g = -\tau$ as a consequence of (3.1) and $f(x) = x$ for $x \in [0, 2\pi]$ by design. Then

$$\widehat{g}(\omega) = -i\frac{A}{2N}(\widehat{\varphi}(\omega - N) - \widehat{\varphi}(\omega + N)). \tag{3.4}$$

For $j, j' \in \mathbb{Z}$ we write $f_{\leq 2^j}$, $f_{\geq 2^{j'}}$, $f_{2^j \leq \cdot \leq 2^{j'}}$, for the projections of f on the subspace of $L^2(\mathbb{R})$ whose Fourier transforms are supported in $|\omega| \leq 2^j$, $|\omega| \geq 2^{j'}$ and $2^j \leq |\omega| \leq 2^{j'}$ respectively, and similarly for the function g.

By assumption, $\widehat{\psi}$ is compactly supported in $(0, +\infty)$, hence there exists $k \in \mathbb{Z}$ such that

$$\operatorname{supp} \widehat{\psi} \subset [2^{-k}, 2^k].$$

As a result, we have that

$$\operatorname{supp} \widehat{\psi}_\lambda \subset [2^{j-k}, 2^{j+k}], \qquad \lambda = 2^j \in 2^{\mathbb{Z}}. \tag{3.5}$$

Let then $j \in \mathbb{Z}$ be such that $2^{j+k+1} < N \leq 2^{j+k+2}$. By (3.4) we have

$$\begin{aligned} \|g_{\leq 2^{j+k}}\|_{L^2}^2 &= \frac{1}{2\pi} \int\limits_{[-2^{j+k}, 2^{j+k}]} |\widehat{g}(\omega)|^2 \, d\omega \\ &= \frac{A^2}{8\pi N^2} \int\limits_{[-2^{j+k}, 2^{j+k}]} |\widehat{\varphi}(\omega - N) - \widehat{\varphi}(\omega + N)|^2 \, d\omega \\ &\leq \frac{A^2}{4\pi N^2} \int\limits_{[-2^{j+k}, 2^{j+k}]} (|\widehat{\varphi}(\omega - N)|^2 + |\widehat{\varphi}(\omega + N)|^2) \, d\omega \\ &\leq \frac{A^2}{4\pi N^2} \int\limits_{\mathbb{R} \setminus [-(N - 2^{j+k}), N - 2^{j+k}]} |\widehat{\varphi}(\omega)|^2 d\omega \\ &\lesssim_\beta \frac{1}{N^\beta} \end{aligned} \tag{3.6}$$

for every $\beta > 0$, because $N - 2^{j+k} > N/2$ and $\widehat{\varphi}$ is a rapidly decreasing function. On the other hand,

$$\|g\|_{L^2}^2 = \|\tau\|_{L^2}^2 = \frac{A^2}{N^2} \int\limits_{[0, 2\pi]} \sin^2(Nx)\varphi(x)^2 \, dx \approx \frac{1}{N^2} \tag{3.7}$$

for N large enough, because the latter integral converges to $\|\varphi\|_{L^2}^2/2$ as $N \to \infty$ due to the Riemann-Lebesgue lemma.

Therefore, we have obtained, for N large enough,

$$\|g_{\geq 2^{j+k}}\|_{L^2}^2 = \|g\|_{L^2}^2 - \|g_{\leq 2^{j+k}}\|_{L^2}^2 \gtrsim \frac{1}{N^2}. \tag{3.8}$$

Furthermore,

$$\|f_{\geq 2^{j-k}}\|_{L^2}^2 = \frac{1}{2\pi} \int\limits_{\mathbb{R}\setminus[-2^{j-k},2^{j-k}]} |\widehat{f}(\omega)|^2 \, d\omega \lesssim \frac{1}{N^\beta} \tag{3.9}$$

for every $\beta > 0$, because $2^{j-k} \geq 2^{-2k-2}N$ and $\widehat{f}$ has rapid decay.

Note that $f = f_{\leq 2^{j-k}} + f_{2^{j-k} \leq \cdot \leq 2^{j+k}} + f_{\geq 2^{j+k}}$ by construction, and a similar decomposition holds for g. Since S_J is nonexpansive, the triangle inequality allows us to write

$$\begin{aligned} \|S_J(L_\tau f) - S_J(f)\| \geq & \big\|S_J(f_{\leq 2^{j-k}} + f_{\geq 2^{j+k}} + g_{\leq 2^{j-k}} + g_{\geq 2^{j+k}}) - S_J(f_{\leq 2^{j-k}} + f_{\geq 2^{j+k}})\big\| \\ & - 2\|f_{2^{j-k} \leq \cdot \leq 2^{j+k}}\|_{L^2} - \|g_{2^{j-k} \leq \cdot \leq 2^{j+k}}\|_{L^2}. \end{aligned} \tag{3.10}$$

We stress that the last two terms are $O(N^{-\beta})$ for every $\beta > 0$ by (3.6) and (3.9).

On the other hand, in view of (3.5) we have that the functions $f_{\leq 2^{j-k}} + g_{\leq 2^{j-k}}$ and $f_{\geq 2^{j+k}} + g_{\geq 2^{j+k}}$ are separated in the wavelet domain, in the sense of Lemma 2.4. Therefore,

$$S_J[\mathcal{P}_J](f_{\leq 2^{j-k}} + g_{\leq 2^{j-k}} + f_{\geq 2^{j+k}} + g_{\geq 2^{j+k}}) = S_J[\mathcal{P}_J](f_{\leq 2^{j-k}} + g_{\leq 2^{j-k}}) + S_J[\mathcal{P}_J](f_{\geq 2^{j+k}} + g_{\geq 2^{j+k}})$$

and similarly,

$$S_J[\mathcal{P}_J](f_{\leq 2^{j-k}} + f_{\geq 2^{j+k}}) = S_J[\mathcal{P}_J](f_{\leq 2^{j-k}}) + S_J[\mathcal{P}_J](f_{\geq 2^{j+k}}).$$

To conclude, since $S_J[\mathcal{P}_J]$ is norm preserving as clarified at the beginning of the proof, we have

$$\begin{aligned} & \big\|S_J[\mathcal{P}_J](f_{\leq 2^{j-k}} + g_{\leq 2^{j-k}} + f_{\geq 2^{j+k}} + g_{\geq 2^{j+k}}) - S_J[\mathcal{P}_J](f_{\leq 2^{j-k}} + f_{\geq 2^{j+k}})\big\| \\ & \qquad \geq \|g_{\geq 2^{j+k}}\|_{L^2} - \|g_{\leq 2^{j-k}}\|_{L^2} - 2\|f_{\geq 2^{j+k}}\|_{L^2} \\ & \qquad \gtrsim \frac{1}{N} \end{aligned}$$

for N large enough, by (3.6), (3.8) and (3.9).

Combining the last bound with (3.10) finally gives (3.3), provided that N is large enough. This concludes the proof of (1.2). The proof of (1.3) turns out to be an immediate consequence of the validity of (1.2), because $\|f_n\|_{L^2} = \|f\|_{L^2}$ for every n and, for $0 \leq \alpha < 1$, $\|\tau_n\|_{C^\alpha} \lesssim 2^{-n(1-\alpha)}$ as a result of (2.1). □

The following result provides lower bounds in the same spirit of (1.3) but for scattering norms instead of the L^2-norm. The factors $\max\{J'+n, 1\}^{1/2}$ and $\max\{J'+n, 1\}^\beta$ occurring in (3.11) and (3.12) below essentially counteract the growth of the norms $\|U[\mathcal{P}_{J'}]f_n\|$ and $\|U[\mathcal{P}_{J'}]f_n\|_1$, respectively (cf. Proposition 2.5).

Proposition 3.1. *Under the same assumption (and notation) as in Theorem 1.1 we have the following lower bounds.*

There exists $C > 0$ such that for every $J, J' \in \mathbb{Z}$, $n \in \mathbb{N}$,

$$\|S_J[\mathcal{P}_J](L_{\tau_n} f_n) - S_J[\mathcal{P}_J](f_n)\| \geq \frac{C 2^{n(1-\alpha)}}{\max\{J'+n, 1\}^{1/2}} \|\tau_n\|_{C^\alpha} \|U[\mathcal{P}_{J'}]f_n\|. \tag{3.11}$$

Moreover, for every $\beta > 1$ there exists $C > 0$ such that, for every $J, J' \in \mathbb{Z}$, $n \in \mathbb{N}$,

$$\|S_J[\mathcal{P}_J](L_{\tau_n} f_n) - S_J[\mathcal{P}_J](f_n)\| \geq \frac{C 2^{n(1-\alpha)}}{\max\{J'+n, 1\}^{\beta}} \|\tau_n\|_{C^\alpha} \|U[\mathcal{P}_{J'}]f_n\|_1. \tag{3.12}$$

Proof. The proof of (3.11) is carried out along the lines of that of (1.3) in Theorem 1.1, now using that

$$\|U[\mathcal{P}_{J'}]f_n\| \lesssim \max\{J'+n,1\}^{1/2}. \tag{3.13}$$

Indeed, as already observed, under the assumptions of Theorem 1.1 the hypotheses of Proposition 2.5 are satisfied as well, so that the latter bound can be inferred from (2.8) and a suitable change of variable:

$$\|U[\mathcal{P}_{J'}]f_n\| \lesssim \Big(\int_{\mathbb{R}} |\widehat{f}(\omega)|^2 \log(e+2^{J'+n}|\omega|)\,d\omega\Big)^{1/2} \lesssim \max\{J'+n,1\}^{1/2},$$

where we also used that $e+2^{J'+n}|\omega| \le (e+2^{J'+n})(e+|\omega|)$.

As far as (3.12) is concerned, it is just enough to replace (2.8) with (2.9) in the aforementioned arguments. □

Remark 3.2. We emphasize that letting $J' \to -\infty$ in (3.11), for fixed J and n, yields (1.2). Indeed, we have that $\|U[\mathcal{P}_{J'}]f\| \to \|f\|_{L^2}$ if $f \in L^2(\mathbb{R}^d)$ and $\|U[\mathcal{P}_{J_0}]f\| < \infty$ for some $J_0 \in \mathbb{Z}$. This follows by an application of the dominated convergence theorem to the map $p \mapsto \|U[p]f\|_{L^2}^2 \mathbb{1}_{\mathcal{P}_{J'}}(p)$ on the set $\mathcal{P}_{J_0}$ (the counting measure being understood), since $\mathbb{1}_{\mathcal{P}_{J'}} \to \mathbb{1}_{\{\emptyset\}}$ pointwise.

In addition, note that the quantity $\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\|$, for fixed f,τ, is nonincreasing when J increases – see Section 2.3 and [20, Proposition 2.9] in this connection. On the contrary, both $\|U[\mathcal{P}_{J'}]f\|$ and $\|U[\mathcal{P}_{J'}]f\|_1$ are increasing with J'. These observations show that the results in Theorem 1.1 and Proposition 3.1 are particularly interesting in the regime $J \gg 1$.

We conclude this section by providing a lower bound for the modulus of continuity of the map $\tau \mapsto S_J[\mathcal{P}_J]L_\tau f$, for signals f with finite scattering norms. The proof is omitted, as it ultimately relies on the same arguments already used for proving Proposition 3.1.

Proposition 3.3. *Consider $0 \le \alpha < 1$. Under the same assumptions (and notation) of Theorem 1.1, there exists a constant $C > 0$ such that, for every $J, J' \in \mathbb{Z}$ and for every n large enough,*

$$\|S_J[\mathcal{P}_J](L_{\tau_n} f_n) - S_J[\mathcal{P}_J](f_n)\| \ge C \log^{-1/2}\big(\|\tau_n\|_{C^\alpha}^{-1}\big)\|U[\mathcal{P}_{J'}]f_n\|.$$

Similarly, for every $\beta > 1$ we have

$$\|S_J[\mathcal{P}_J](L_{\tau_n} f_n) - S_J[\mathcal{P}_J](f_n)\| \ge C \log^{-\beta}\big(\|\tau_n\|_{C^\alpha}^{-1}\big)\|U[\mathcal{P}_{J'}]f_n\|_1.$$

4. Stability results for $C^{1+\alpha}$ regularity, $\alpha > 0$

In this section we provide the proof of Theorem 1.2, where we consider deformation fields $\tau \in C^{1+\alpha}(\mathbb{R}^d;\mathbb{R}^d)$ for $0 < \alpha < 1$. Recall that this implies $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$ and $D\tau \in C^\alpha(\mathbb{R}^d;\mathbb{R}^d)$. We assume that ϕ and ψ are scattering filters in the sense of Section 2.3.

The proof of Theorem 1.2 mostly relies on the structure of [20, Theorem 2.12] and the ancillary results [20, Lemmas 2.13 and 2.14]. For our purposes we only need to isolate a limited number of key steps and elaborate on those as detailed below. Nevertheless, let us briefly discuss the complete roadmap for the sake of clarity.

The goal is to bound the quantity $\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\|$ in terms of $\|\tau\|_{C^\alpha}$ and a scattering norm of f. It is clear that

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \le \|L_\tau(S_J[\mathcal{P}_J]f) - S_J[\mathcal{P}_J](f)\| + \|[S_J[\mathcal{P}_J], L_\tau]\|,$$

where the commutator of two operators A, B is defined by $[A, B] = AB - BA$.

Let us focus on the first term. The stabilizing properties of the average A_J underlying $S_J[\mathcal{P}_J]$ can be exploited to prove that

$$\|L_\tau(S_J[\mathcal{P}_J]f) - S_J[\mathcal{P}_J](f)\| \lesssim 2^{-J}\|\tau\|_{L^\infty}\|U[\mathcal{P}_J]f\|,$$

see [20, (2.42) and (2.51)] – which actually hold assuming only $\tau \in C^1(\mathbb{R}^d; \mathbb{R}^d)$.

Controlling the commutator error $\|[S_J[\mathcal{P}_J], L_\tau]\|$ is actually the main difficulty of this result. First, it is proved in [20, Lemma 2.13] that for any operator L on $L^2(\mathbb{R}^d)$ one has

$$\|[S_J[\mathcal{P}_J], L]f\| \leq \|[U_J, L]\|\|U[\mathcal{P}_J]f\|_1,$$

where $U_J h = \{A_J h, U[\Lambda_J]h\}$ for $h \in L^2(\mathbb{R}^d)$. Note that $U_J = MW_J$, where M is the nonexpansive operator on $\ell^2(\{J\} \cup \Lambda_J; L^2(\mathbb{R}^d))$ given by $M\{h_J, (h_\lambda)_{\lambda\in\Lambda_J}\} = \{h_J, (|h_\lambda|)_{\lambda\in\Lambda_J}\}$, and since $[M, L_\tau] = 0$ the problem ultimately reduces to bounding the commutator error $\|[W_J, L_\tau]\|$ between the Littlewood-Paley wavelet transform W_J at scale 2^J and the deformation operator L_τ. For $\tau \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ it was proved in [20, Lemma 2.14] that

$$\|[W_J, L_\tau]\| \lesssim 2^{-J}\|\tau\|_{L^\infty} + \max\left\{\log\frac{\|\Delta\tau\|_{L^\infty}}{\|D\tau\|_{L^\infty}}, 1\right\}\|D\tau\|_{L^\infty} + \|D^2\tau\|_{L^\infty}.$$

The proof of this result is a technical *tour de force* among quite delicate estimates. Roughly speaking, the operator $[W_J, L_\tau]^*[W_J, L_\tau]$ has a singular kernel along the diagonal, and the standard method of harmonic analysis suggests considering a suitable frequency decomposition. The singular part of the operator is then carried by the high-frequency terms, and the latter are eventually bounded using the Cotlar lemma.

Our contribution in this connection is an improvement of the estimate above.

Proposition 4.1. *Let $0 < \alpha < 1$. There exists a constant $C > 0$ such that for all $J \in \mathbb{Z}$ and $\tau \in C^{1+\alpha}(\mathbb{R}^d; \mathbb{R}^d)$, with $\|D\tau\|_{L^\infty} \leq 1/2$,*

$$\|[W_J, L_\tau]\| \leq C\left(2^{-J}\|\tau\|_{L^\infty} + \max\left\{\log\frac{\|\Delta\tau\|_{L^\infty}}{\|D\tau\|_{L^\infty}}, 1\right\}\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\right). \tag{4.1}$$

Proof. We follow the same pattern of the proof of [20, Lemma 2.14], taking for granted all the estimates proved there under the assumption $\tau \in C^1(\mathbb{R}^d; \mathbb{R}^d)$. We thus confine ourselves to describe the main strategy, while focusing on the necessary modifications. We adhere to the notation used in Mallat's proof for the convenience of the reader.

First, the problem is recast as follows:

$$\begin{aligned}\|[W_J, L_\tau]\| &= \|[W_J, L_\tau]^*[W_J, L_\tau]\|^{1/2} \\ &\leq \underbrace{\sum_{r\in G^+}\left\|\sum_{j=-J+1}^{\infty}[W[2^j r], L_\tau]^*[W[2^j r], L_\tau]\right\|^{1/2}}_{=:I_1} + \underbrace{\|[A_J, L_\tau]^*[A_J, L_\tau]\|^{1/2}}_{=:I_2}.\end{aligned} \tag{4.2}$$

Bounds for the latter quantities are derived in [20, Lemma E.1], that will be improved as well in accordance with our weaker regularity assumptions. The main argument goes as follows. For $j \in \mathbb{Z}$ consider

$$Z_j f = f * h_j, \qquad h_j(x) = 2^{dj}h(2^j x), \tag{4.3}$$

for a twice differentiable function $h\colon \mathbb{R}^d \to \mathbb{C}$ that decays like $O((1+|x|)^{-d-3})$ along with all its first- and second-order partial derivatives. We introduce the companion operators $K_j := Z_j - L_\tau Z_j L_\tau^{-1}$ and note that $K_j = K_{j,1} + K_{j,2}$, where the latter are integral operators with kernels respectively given by

$$k_{j,1}(x,u) = 2^{dj} g(u, 2^j(x-u)), \tag{4.4}$$
$$k_{j,2}(x,u) = \det(I - D\tau(u))\left(h_j((I-D\tau(u))(x-u)) - h_j(x - \tau(x) - u + \tau(u))\right),$$

where we set

$$g(u,v) := h(v) - h\big((I - D\tau(u))v\big)\det(I - D\tau(u)), \quad (u,v) \in \mathbb{R}^{2d}. \tag{4.5}$$

Step 1. Bound for I_2.

In light of the previous assumptions, we have

$$\|[Z_j, L_\tau]\| = \|[Z_j, L_\tau]^*[Z_j, L_\tau]\|^{1/2} \le \|L_\tau\|\|K_j^* K_j\|^{1/2} = \|L_\tau\|\|K_j\| \le 2^{d/2}\|K_j\|.$$

A bound for $I_2 = \|[A_J, L_\tau]\|$ can thus be obtained by bounding $\|K_j\|$ in the case where $h = \phi$ and $j = -J$. In particular, since

$$\|K_j\| \le \|K_{j,1}\| + \|K_{j,2}\|, \tag{4.6}$$

it is enough to separately control the norms of the latter integral operators.

1.1. Bound for $\|K_{j,1}\|$.

This was already proved in [20, Eq. (E.18)]:

$$\|K_{j,1}\| \lesssim \|D\tau\|_{L^\infty}. \tag{4.7}$$

1.2. Bound for $\|K_{j,2}\|$.

Consider the kernel $k_{j,2}$. A Taylor expansion of $\tau(x)$ centered at u gives

$$\begin{aligned}\tau(x) - \tau(u) &= \int_0^1 D\tau(u + t(x-u))(x-u)dt \\ &= D\tau(u)(x-u) + \int_0^1 \big(D\tau(u+t(x-u)) - D\tau(u)\big)(x-u)dt \\ &= D\tau(u)(x-u) + \alpha(u, x-u),\end{aligned}$$

where in view of the assumption on τ the remainder $\alpha(u, x-u)$ defined by the above formula satisfies

$$|\alpha(u, x-u)| \le |D\tau|_{C^\alpha}|x-u|^{1+\alpha}. \tag{4.8}$$

Combining this result with a Taylor expansion of h_j finally gives

$$k_{j,2}(x,u) = -\det(I - D\tau(u))\int_0^1 Dh_j\big((I - tD\tau(u))(x-u) + (1-t)(\tau(u) - \tau(x))\big)\alpha(u, x-u)dt.$$

We now infer a bound for $\|K_{j,2}\|$ by controlling the norm of the kernel in view of Schur's lemma. First, note that the assumption $\|D\tau\|_{L^\infty} \leq 1/2$ implies $|\det(I - D\tau(y))| \leq 2^d$. Moreover, using $Dh_j(y) = 2^{j(d+1)} Dh(2^j y)$ and the substitution $x' = 2^j(x-u)$ we obtain

$$\int_{\mathbb{R}^d} |k_{j,2}(x,u)|dx \leq 2^d \int_{\mathbb{R}^d} \Big| \int_0^1 Dh\big((I - tD\tau(u))x' + (1-t)2^j(\tau(u) - \tau(2^{-j}x' + u))\big) 2^j \alpha(u, 2^{-j}x')dt \Big| dx'.$$

Recall that $|Dh(y)| \lesssim (1+|y|)^{-d-3}$ by assumption, and it is easy to see that for $0 \leq t \leq 1$ we have

$$\begin{aligned} \big|(I - tD\tau(u))x' + (1-t)2^j(\tau(u) - \tau(2^{-j}x' + u))\big| &\geq |x'|(1 - \|D\tau\|_{L^\infty}) \\ &\geq |x'|/2. \end{aligned}$$

Concerning the term $|2^j\alpha(u, 2^{-j}x')|$, on the one hand, by virtue of (4.8) we have

$$|2^j\alpha(u, 2^{-j}x')| \lesssim 2^{-j\alpha}|D\tau|_{C^\alpha}|x'|^{1+\alpha}.$$

Then

$$\int_{\mathbb{R}^d} |k_{j,2}(x,u)|dx \lesssim 2^{-j\alpha}|D\tau|_{C^\alpha} \int_{\mathbb{R}^d} (1 + |x'|/2)^{-d-3}|x'|^{1+\alpha}dx' \lesssim 2^{-j\alpha}|D\tau|_{C^\alpha}.$$

On the other hand, we observe that

$$|2^j\alpha(u, 2^{-j}x')| = 2^j|\tau(2^{-j}x' + u) - \tau(u) - D\tau(u)(2^{-j}x')| \leq 2\|D\tau\|_{L^\infty}|x'|,$$

resulting in

$$\int_{\mathbb{R}^d} |k_{j,2}(x,u)|dx \lesssim \|D\tau\|_{L^\infty} \int_{\mathbb{R}^d} (1 + |x'|/2)^{-d-3}|x'|dx' \lesssim \|D\tau\|_{L^\infty}.$$

Combining the previous estimates gives

$$\int_{\mathbb{R}^d} |k_{j,2}(x,u)|dx \lesssim \min\{2^{-j\alpha}|D\tau|_{C^\alpha}, \|D\tau\|_{L^\infty}\},$$

and the same bound holds for $\int_{\mathbb{R}^d} |k_{j,2}(x,u)|du$ since the previous arguments apply in the same form after the substitution $u' = 2^j(x-u)$. As a consequence of Schur's lemma we thus obtain

$$\|K_{j,2}\| \lesssim \min\{2^{-j\alpha}|D\tau|_{C^\alpha}, \|D\tau\|_{L^\infty}\}. \tag{4.9}$$

The combination of the estimates above shows that the term I_2 in (4.2) is $\lesssim \|D\tau\|_{L^\infty}$.

Step 2. Bound for I_1.

Consider again the convolution operator Z_j introduced in (4.3). By mimicking [20, Lemma E.1], we will prove that if $\int_{\mathbb{R}^d} h(x)dx = 0$ then

$$\Big\| \sum_{j=-\infty}^{+\infty} [Z_j, L_\tau]^*[Z_j, L_\tau] \Big\|^{1/2} \lesssim \max\Big\{\log \frac{\|\Delta\tau\|_{L^\infty}}{\|D\tau\|_{L^\infty}}, 1\Big\}\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}. \tag{4.10}$$

Consider in particular the case where $h(x) = \psi(r^{-1}x)$ for each $r \in G^+$ and replace $-\infty$ with $-J+1$ in the summation ($[Z_j, L_\tau]^*[Z_j, L_\tau]$ is a positive operator). The resulting bound, combined with the one for I_2 proved above, will conclude the proof of (4.1) and thus of Proposition 4.1.

To this aim, let us first remark that the nonsingular part of the commutator has been isolated and bounded in [20, Pages 1386-1387], yielding

$$\Big\| \sum_{j=-\infty}^{+\infty} [Z_j, L_\tau]^*[Z_j, L_\tau] \Big\|^{1/2} \lesssim \max\Big\{\log \frac{\|\Delta\tau\|_{L^\infty}}{\|D\tau\|_{L^\infty}}, 1\Big\} \|D\tau\|_{L^\infty} + \Big\| \sum_{j=0}^{+\infty} K_j^* K_j \Big\|^{1/2}. \tag{4.11}$$

2.1. Bound for $\big\| \sum_{j=0}^{+\infty} K_j^* K_j \big\|^{1/2}$.

Recall that $K_j = K_{j,1} + K_{j,2}$, hence

$$\begin{aligned} \Big\| \sum_{j=0}^{+\infty} K_j^* K_j \Big\|^{1/2} &\le \Big\| \sum_{j=0}^{+\infty} K_{j,1}^* K_{j,1} \Big\|^{1/2} \\ &\quad + \sum_{j=0}^{+\infty} \Big(\|K_{j,2}\| + 2^{1/2} \|K_{j,1}\|^{1/2} \|K_{j,2}\|^{1/2} \Big). \end{aligned} \tag{4.12}$$

Using (4.7) and (4.9) above, we have that

$$\sum_{j=0}^{+\infty} \Big(\|K_{j,2}\| + 2^{1/2} \|K_{j,1}\|^{1/2} \|K_{j,2}\|^{1/2} \Big) \lesssim \|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}. \tag{4.13}$$

2.2. Bound for $\big\| \sum_{j=0}^{+\infty} K_{j,1}^* K_{j,1} \big\|^{1/2}$.

Recall that here we are assuming that $\int_{\mathbb{R}^d} h(x)dx = 0$.

The goal of this section is to provide a bound for $\big\| \sum_{j=-\infty}^{+\infty} Q_j \big\|$, where we set $Q_j = K_{j,1}^* K_{j,1}$ for $j \ge 0$ and $Q_j = 0$ for $j < 0$. We will apply Cotlar's lemma [27, Chapter VII]: if there is a sequence of positive real numbers $\{\beta(j)\}_{j\in\mathbb{Z}}$ such that $\sum_{j\in\mathbb{Z}} \beta(j) < \infty$ and

$$\|Q_j^* Q_l\| \le \beta(j-l)^2, \quad \|Q_j Q_l^*\| \le \beta(j-l)^2,$$

then

$$\Big\| \sum_{j=-\infty}^{+\infty} Q_j \Big\| \le \sum_{j=-\infty}^{+\infty} \beta(j).$$

As a consequence of the self-adjointness of Q_j it is enough to provide a bound for $\|Q_l Q_j\|$ only, hence we resort again to Schur's lemma.

Let $\bar{k}_j$ be the integral kernel of $Q_j = K_{j,1}^* K_{j,1}$. Using (4.4) we infer

$$\begin{aligned} \bar{k}_j(y,z) &= \int_{\mathbb{R}^d} \overline{k_{j,1}(x,y)} k_{j,1}(x,z) dx \\ &= \int_{\mathbb{R}^d} 2^{dj} \overline{g(y, x + 2^j(z-y))} g(z,x) dx \\ &= \int_{\mathbb{R}^d} 2^{dj} \overline{g(y,x')} g(z, x' + 2^j(y-z)) dx'. \end{aligned}$$

The integral kernel $\bar{k}_{l,j}$ of $Q_l Q_j$ is thus given by

$$\bar{k}_{l,j}(y,z) = \int_{\mathbb{R}^{3d}} g(u,x)\overline{g(u,x')}2^{dl}\overline{g(y,x+2^l(u-y))}2^{dj} g(z,x'+2^j(u-z))dxdx'du. \tag{4.14}$$

Setting $H(u,x,x',y) = g(u,x)\overline{g(u,x')}2^{dl}\overline{g(y,x+2^l(u-y))}2^{dj} g(z,x'+2^j(u-z))$, notice that

$$\begin{aligned}\int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} H(u,x,x',y)dxdx'du\Big|dy &= \int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} H(u,x,x',y)dudxdx'\Big|dy\\ &\le \int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} H(u,x,x',y)du\Big|dxdx'dy\\ &= \int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} H(u,x,x',y)du\Big|dydxdx'.\end{aligned}$$

Therefore, it is enough to prove a suitable bound for the functional

$$\varphi(G) = \varphi_{x,x',z,j,l}(G) := \int_{\mathbb{R}^d}\Big|\int_{\mathbb{R}^d} G(u)2^{dl}\overline{g(y,x+2^l(u-y))}2^{dj} g(z,x'+2^j(u-z))du\Big|dy, \tag{4.15}$$

for $G \in C^\alpha(\mathbb{R}^d)$. We restrict to the case $j \ge l \ge 0$ in view of the symmetric role of these parameters. We will prove below that

$$\varphi(G) \lesssim 2^{(l-j)\alpha}\|D\tau\|^2_{L^\infty}\|G\|_{C^\alpha}, \tag{4.16}$$

where the hidden constant does not depend on x, x', z. We are going to apply this estimate to $G(u) := g(u,x)\overline{g(u,x')}$, where $x, x' \in \mathbb{R}^d$ play here the role of parameters. We then will prove that

$$\|G\|_{C^\alpha} \lesssim (1+|x|)^{-d-1}(1+|x'|)^{-d-1}\|D\tau\|_{L^\infty}\big(\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\big). \tag{4.17}$$

The latter bounds allow us to conclude that

$$\begin{aligned}\int_{\mathbb{R}^d}|\bar{k}_{l,j}(y,z)|dy &\le \int_{\mathbb{R}^d}\varphi(g(\cdot,x)g(\cdot,x'))\,dx\,dx'\\ &\lesssim 2^{(l-j)\alpha}\big(\|D\tau\|^4_{L^\infty} + \|D\tau\|^3_{L^\infty}|D\tau|_{C^\alpha}\big)\\ &\lesssim 2^{(l-j)\alpha}\big(\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\big)^4.\end{aligned}$$

The same arguments lead one to the same bound for $\int_{\mathbb{R}^d}|\bar{k}_{l,j}(y,z)|dz$. Therefore, by Schur's lemma we have

$$\|Q_l Q_j\| \lesssim 2^{(l-j)\alpha}\big(\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\big)^4,$$

and Cotlar's lemma with $\beta(j) = C2^{-|j|\alpha/2}\big(\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\big)^2$ for a suitable constant $C > 0$ finally implies that

$$\Big\|\sum_{j=0}^{+\infty} K^*_{j,1}K_{j,1}\Big\| \lesssim \big(\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\big)^2,$$

which combined with (4.11), (4.12) and (4.13) gives (4.10) and concludes the proof.

The proofs of (4.16) and (4.17) are given below. □

Proof of (4.16)**.** A Taylor expansion of h in (4.5) (cf. [20, (E.30)]) implies that

$$g(u,v) = \big(1-\det(I-D\tau(u))\big)h\big((I-D\tau(u))v\big) + \int_0^1 Dh\big((1-t)v + t(I-D\tau(u))v\big)\cdot D\tau(u)v dt. \tag{4.18}$$

Using that $\det(I-D\tau(u)) \geq (1-\|D\tau\|_{L^\infty})^d$, $\|D\tau\|_\infty \leq 1/2$ and the fact that $h(x)$ and $Dh(x)$ by assumption decay like $(1+|x|)^{-d-3}$, we obtain (cf. [20, (E.31)]) that

$$|g(u,v)| \lesssim \|D\tau\|_{L^\infty}(1+|v|)^{-d-2}. \tag{4.19}$$

Using this estimate it is easy to infer that for any $G \in C^0(\mathbb{R}^d)$ we have

$$\varphi(G) \lesssim \|D\tau\|^2_{L^\infty}\|G\|_{C^0}.$$

Recall that the arguments are intended here for $h(x) = \psi(r^{-1}x)$ and every $r \in G^+$. Let us consider first the case where $r = I$. The spectral conic support condition for ψ (see Lemma 2.3) and (4.5) imply that $\int_{\mathbb{R}} g(u,v)\,dv_1 = 0$ for all $u \in \mathbb{R}^d$ and $v' \in \mathbb{R}^{d-1}$ ($v = (v_1, v')$)). Combining the latter result with (4.19), we conclude that $g(u,v) = \partial\overline{g}(u,v)/\partial v_1$ for some $\overline{g}(u,v)$ satisfying (cf. [20, (E.37)])

$$|\overline{g}(u,v)| \leq C\|D\tau\|_\infty(1+|v|)^{-d-1}. \tag{4.20}$$

Also, from (4.18) we have (cf. [20, (E.39)])

$$|\nabla_v g(u,v)| \lesssim \|D\tau\|_\infty(1+|v|)^{-d-1}. \tag{4.21}$$

Integration by parts with respect to u_1 (cf. [20, Page 1391]) therefore yields

$$\begin{aligned}
&\int_{\mathbb{R}^d} G(u)2^{dl}\overline{g(y,x+2^l(u-y))}2^{dj}g(z,x'+2^j(u-z))du \\
=&2^{-j}\int_{\mathbb{R}^d} G(u)2^{dl}\overline{g(y,x+2^l(u-y))}2^{dj}\frac{\partial}{\partial u_1}\Big[\bar{g}(z,x'+2^j(u-z))\Big]du \\
=&-2^{-j}\int_{\mathbb{R}^d}\frac{\partial G}{\partial u_1}(u)2^{dl}\overline{g(y,x+2^l(u-y))}2^{dj}\bar{g}(z,x'+2^j(u-z))du \\
&-2^{l-j}\int_{\mathbb{R}^d} G(u)2^{dl}\overline{\frac{\partial g}{\partial u_1}(y,x+2^l(u-y))}2^{dj}\bar{g}(z,x'+2^j(u-z))du.
\end{aligned}$$

Using the estimates (4.20) and (4.21), along with Young's convolution inequality $L^1 * L^1 \hookrightarrow L^1$, we obtain

$$\begin{aligned}
\varphi(G) &\lesssim 2^{-j}\|D\tau\|^2_{L^\infty}\|G\|_{C^1} + 2^{l-j}\|D\tau\|^2_{L^\infty}\|G\|_{C^0} \\
&\lesssim 2^{l-j}\|D\tau\|^2_{L^\infty}\|G\|_{C^1},
\end{aligned}$$

for every $G \in C^1(\mathbb{R}^d)$. Note that the functional φ is subadditive, namely $|\varphi(f+g)| \leq |\varphi(f)| + |\varphi(g)|$. By real interpolation (see e.g., [18, Theorem 6]), since $[C^0, C^1]_{\alpha,\infty} = C^\alpha$ (see e.g., [17, (1.16)]), we obtain (4.16) for every $G \in C^\alpha(\mathbb{R}^d)$.

Consider now the general case where $h(x) = \psi(r^{-1}x)$ and r is a rotation. Hence (cf. (4.5)) we have

$$g(u,v) = \psi(r^{-1}v) - \psi\big((I - r^{-1}D\tau(u)r)r^{-1}v\big)\det(I - D\tau(u)), \quad (u,v) \in \mathbb{R}^{2d}.$$

Again by Lemma 2.3 we argue that the integral of $g(u,\cdot)$ along all the lines parallel to re_1 ($e_1 = (1,0,\dots,0) \in \mathbb{R}^d$) is zero. Then $g(u,v) = D_{re_1}\bar{g}(u,v)$, where D_{re_1} denotes the directional derivative along the direction re_1 in the space of the v's, for some $\bar{g}(u,v)$ satisfying the same estimates as before. The rest of the proof still works after integrating by parts along the direction re_1 – the transpose of the operator D_{re_1} being $-D_{re_1}$. □

Proof of (4.17)**.** Recall that here $G(u) = g(u,x)\overline{g(u,x')}$, where g is given in (4.18). From the very definition of the C^α norm, using the fractional Leibniz rule (2.2) and (4.19) to bound the norm in $L^\infty(\mathbb{R}^d)$ of G, it suffices eventually to show that

$$|g(\cdot,v)|_{C^\alpha} \lesssim (1+|v|)^{-d-1}\big(\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\big). \tag{4.22}$$

We use the expression in (4.18), and again the fractional Leibniz rule (2.2). Precisely, the desired estimate (4.22) will follow from the bounds given below on the L^∞ norm and C^α seminorm in $\mathbb{R}^d$ (with respect to u) of the factors appearing in (4.18).

- Let $b_1(u) = 1 - \det(I - D\tau(u))$. Since $\det(I - D\tau(u)) \geq (1 - \|D\tau\|_{L^\infty})^d$ we have

  $$\|b_1\|_{L^\infty} \lesssim \|D\tau\|_{L^\infty}.$$

  Moreover, since b_1 is a polynomial in the entries of the matrix $D\tau(u)$ and $\|D\tau(u)\|_{L^\infty} \leq 1/2$, we obtain

  $$|b_1|_{C^\alpha} \lesssim |D\tau|_{C^\alpha}$$

  by a straightforward application of Schauder's estimate (2.3).
- Define $b_2(u,v) = h\big((I - D\tau(u))v\big)$. Clearly

  $$\|b_2(\cdot,v)\|_{L^\infty} \lesssim (1+|v|)^{-d-3}.$$

  A bound for $|b_2(\cdot,v)|_{C^\alpha}$ can be obtained using Schauder's estimate (2.3), in particular

  $$|b_2(\cdot,v)|_{C^\alpha} \lesssim |h|_{\mathrm{Lip}(B_v)}|\widetilde{b_2}(\cdot,v)|_{C^\alpha},$$

  where we introduced the companion map $\widetilde{b_2}(u,v) = (I - D\tau(u))v$ and B_v stands for the range of the map $\tilde{b}_2(\cdot,v)$, for fixed v.

  First, we remark that

  $$|\widetilde{b_2}(\cdot,v)|_{C^\alpha} \lesssim |D\tau|_{C^\alpha}|v|.$$

  Moreover, since $Dh(u) = O((1+|u|)^{-d-3})$ by assumption, and since B_v is contained in the ball $B(v,|v|/2)$, we have

  $$|h|_{\mathrm{Lip}(B_v)} \lesssim (1+|v|)^{-d-3}.$$

  We then conclude that

$$|b_2(\cdot, v)|_{C^\alpha} \lesssim |D\tau|_{C^\alpha}(1+|v|)^{-d-2}$$

as a combination of the previous results.

- Lastly, consider the map b_3 defined for $0 \le t \le 1$ by

$$b_3(u,v) = Dh\big(I - tD\tau(u))v\big) \cdot D\tau(u)v.$$

An application of the fractional Leibniz formula (2.2) combined with the same estimates for b_2 above with Dh in place of h finally gives

$$\begin{aligned} |b_3(\cdot, v)|_{C^\alpha} &\lesssim (1+|v|)^{-d-2}|D\tau|_{C^\alpha} + (1+|v|)^{-d-1}|D\tau|_{C^\alpha}\|D\tau\|_{L^\infty} \\ &\lesssim (1+|v|)^{-d-1}|D\tau|_{C^\alpha}. \end{aligned}$$

This concludes the proof of (4.22). □

Remark 4.2. It was already pointed out in [20] that the term $\max\Big\{\log \frac{\|\Delta\tau\|_{L^\infty}}{\|D\tau\|_{L^\infty}}, 1\Big\}$ in (4.11) can be replaced with $\max\{J,1\}$ – to be precise, it is enough to choose $\gamma = \max\{J,1\}$ in [20, (E.7)–(E.10)]. Therefore, we have also the estimate

$$\|S_J[\mathcal{P}_J](L_\tau f) - S_J[\mathcal{P}_J](f)\| \le C\left(2^{-J}\|\tau\|_{L^\infty} + \max\{J,1\}\|D\tau\|_{L^\infty} + |D\tau|_{C^\alpha}\right)\|U[\mathcal{P}_J]f\|_1. \tag{4.23}$$

Note that the stability bound, in this form, becomes void in the asymptotic regime $J \to +\infty$, hence it fails to recover the asymptotic translation invariance of the scattering transform. It is worth emphasizing that results concerning translation invariance of generalized scattering networks in [33,34] are of "vertical" type, namely they are established through pooling and are asymptotic in the network depth, in accordance with the findings and expectations in the deep learning literature.

5. Stability up to ε losses for Lipschitz deformations

In this section we focus on Lipschitz deformations; in particular, we prove Theorem 1.3. We continue to assume that ϕ and ψ are scattering filters in the sense of Section 2.3.

We need some preliminary results from the theory of real interpolation of Besov spaces (cf. [30]).

Let $\phi_0 \in C^\infty(\mathbb{R}^d)$ be supported in the ball $|\omega| \le 2$, with $\phi_0(\omega) = 1$ for $|\omega| \le 1$. Set $\phi_j(\omega) = \phi_0(2^{-j}\omega)$, $j \in \mathbb{Z}$. The functions $\tilde{\phi}_j := \phi_j - \phi_{j-1}$, $j \in \mathbb{Z}$, are supported in the annuli $2^{j-1} \le |\omega| \le 2^{j+1}$ and induce a Littlewood-Paley partition of unity of $\mathbb{R}^d \setminus \{0\}$. Recall that the Besov norms with $s \in \mathbb{R}$, $1 \le p, q \le \infty$ are accordingly defined, for a temperate distribution f, by

$$\|f\|_{B^s_{p,q}} := \Big(\|\phi_0(D)f\|^q_{L^p} + \sum_{j\ge 1} 2^{jsq}\|\tilde{\phi}_j(D)f\|^q_{L^p}\Big)^{1/q},$$

where $\tilde{\phi}_j(D) = \mathcal{F}^{-1}\tilde{\phi}_j\mathcal{F}$ stands for the Fourier multiplier with symbol $\tilde{\phi}_j$, $j \in \mathbb{Z}$, and similarly $\phi_0(D) = \mathcal{F}^{-1}\phi_0\mathcal{F}$. Obvious changes are needed if $q = \infty$.

Recall that for $s > 0$ *not* integer, the space $B^s_{\infty,\infty}(\mathbb{R}^d)$ coincides with the Hölder class $C^s(\mathbb{R}^d)$ considered above. If $s = 1$ then $B^1_{\infty,\infty}(\mathbb{R}^d)$ *contains* the space $\mathrm{Lip}(\mathbb{R}^d)$ of bounded Lipschitz functions $f\colon \mathbb{R}^d \to \mathbb{C}$, endowed with the norm

$$\|f\|_{\mathrm{Lip}} := \|f\|_{L^\infty} + |f|_{\mathrm{Lip}} = \|f\|_{L^\infty} + \|\nabla f\|_{L^\infty},$$

where ∇f is understood in the sense of distributions or even almost everywhere.

With a temperate distribution f we also associate the set

$$\mathcal{A}_f := \{\phi_j(D)f : \ j \geq 0\}. \tag{5.1}$$

Observe that the operators $\phi_j(D)$, $j \in \mathbb{Z}$, are uniformly bounded on $L^\infty(\mathbb{R}^d)$, since $\mathcal{F}^{-1}\phi_0 \in L^1(\mathbb{R}^d)$.

Lemma 5.1. *Let $0 < \theta < 1$ and $s > 1$ be such that $1 = (1-\theta)s$. For $f \in B^0_{\infty,1}(\mathbb{R}^d)$ and $t > 0$, consider*

$$\tilde{K}(t, f, B^s_{\infty,\infty}, B^0_{\infty,1}) := \inf_{f_0 \in \mathcal{A}_f} \{\|f_0\|_{B^s_{\infty,\infty}} + t\|f - f_0\|_{B^0_{\infty,1}}\}.$$

There exists a constant $C > 0$ such that, for every $f \in B^1_{\infty,\infty}(\mathbb{R}^d)$,

$$\sup_{t>1} t^{-\theta}\tilde{K}(t, f, B^s_{\infty,\infty}, B^0_{\infty,1}) \leq C\|f\|_{B^1_{\infty,\infty}}. \tag{5.2}$$

Proof. The functional $\tilde{K}(t, f, B^s_{\infty,\infty}, B^0_{\infty,1})$ is just a variant of the K-functional in real interpolation theory (cf. [30, Section 2.4.1]), defined for $t > 0$ and $f \in B^0_{\infty,1}(\mathbb{R}^d)$ by

$$K(t, f, B^s_{\infty,\infty}, B^0_{\infty,1}) := \inf\{\|f_0\|_{B^s_{\infty,\infty}} + t\|f - f_0\|_{B^0_{\infty,1}} : \ f_0 \in B^s_{\infty,\infty}\}.$$

It is well known that

$$\sup_{t>0} t^{-\theta}K(t, f, B^s_{\infty,\infty}, B^0_{\infty,1}) \lesssim \|f\|_{B^1_{\infty,\infty}}$$

for every $f \in B^1_{\infty,\infty}(\mathbb{R}^d)$, which amounts to the embedding $B^1_{\infty,\infty} \hookrightarrow (B^s_{\infty,\infty}, B^0_{\infty,1})_{\theta,\infty}$. A proof of this fact can be found in [30, Section 2.4.2], and an accurate inspection of the latter (the part dealing with $t > 1$, to be precise) allows one to realize that (5.2) holds indeed. □

We are now ready to prove Theorem 1.3. We will consider vector fields τ in the Besov space $B^s_{p,q}(\mathbb{R}^d;\mathbb{R}^d)$ (i.e., the components belong to $B^s_{p,q}(\mathbb{R}^d)$), endowed with the norm

$$\|\tau\|_{B^s_{p,q}} := \sum_{k=1}^{d} \|\tau^{(k)}\|_{B^s_{p,q}}, \qquad \tau = (\tau^{(1)}, \dots, \tau^{(d)}).$$

Proof of Theorem 1.3. We can suppose $J = 0$ by virtue of a scaling argument (cf. (3.2)). Indeed, the estimate (1.8) is invariant under the substitutions $J \to J - n$ ($n \in \mathbb{Z}$), $f(x) \to 2^{n/2}f(2^n x)$, $\tau(x) \to 2^{-n}\tau(2^n x)$ and $R \to 2^n R$. We can also suppose that $\|D\tau\|_{L^\infty} \leq \varepsilon_0$, with ε_0 small enough (to be fixed later on), because for $\varepsilon_0 < \|D\tau\|_{L^\infty} \leq 1/2$ the estimate (1.8) with $J = 0$ holds due to the fact that $S_0[\mathcal{P}_0]$ is nonexpansive on $L^2(\mathbb{R}^d)$ and $\|L_\tau\| \leq 2^{d/2}$.

We already know from Remark 4.2 (with $J = 0$) that, for every $s \in (1, 2)$ and $\tau \in C^s(\mathbb{R}^d;\mathbb{R}^d) = B^s_{\infty,\infty}(\mathbb{R}^d;\mathbb{R}^d)$ with $\|D\tau\|_{L^\infty} \leq 1/2$,

$$\|S_0[\mathcal{P}_0](L_\tau f) - S_0[\mathcal{P}_0](f)\| \lesssim \|\tau\|_{B^s_{\infty,\infty}}\|U_0[\mathcal{P}_0]f\|_1.$$

The assumption (1.7) and the fact that $\widehat{f}$ is supported in the ball $|\omega| \leq R$ imply that

$$\|S_0[\mathcal{P}_0](L_\tau f) - S_0[\mathcal{P}_0](f)\| \lesssim \log^\beta(e + R)\|\tau\|_{B^s_{\infty,\infty}}\|f\|_{L^2}. \tag{5.3}$$

On the other hand, if $\tau_0, \tau_1 \in \mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^d)$ satisfy $\|D\tau_0\|_{L^\infty} \leq 1/2$ and $\|D\tau_1\|_{L^\infty} \leq 1/2$, a Taylor expansion yields

$$\|L_{\tau_0} f - L_{\tau_1} f\|_{L^2} \leq \|\tau_0 - \tau_1\|_{L^\infty} \int_0^1 \|L_{(1-t)\tau_1 + t\tau_0}(\nabla f)\|_{L^2}\, dt$$
$$\lesssim \|\tau_0 - \tau_1\|_{L^\infty} \|\nabla f\|_{L^2}$$

where we used that $\|D((1-t)\tau_1 + t\tau_0)\|_{L^\infty} \leq 1/2$. Since $\|\nabla f\|_{L^2} \lesssim (1+R)\|f\|_{L^2}$ and $S_0[\mathcal{P}_0]$ is nonexpansive, we conclude that

$$\begin{aligned} \|S_0[\mathcal{P}_0](L_{\tau_0} f) - S_0[\mathcal{P}_0](L_{\tau_1} f)\| &\lesssim (1+R)\|\tau_0 - \tau_1\|_{L^\infty} \|f\|_{L^2} \\ &\lesssim (1+R)\|\tau_0 - \tau_1\|_{B^0_{\infty,1}} \|f\|_{L^2}, \end{aligned} \tag{5.4}$$

where we used the embedding $B^0_{\infty,1}(\mathbb{R}^d;\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d;\mathbb{R}^d)$ – see for instance [25, Proposition 2.1].

We now resort to a nonlinear interpolation argument between (5.3) and (5.4). Set $\mathcal{H} = \ell^2(\mathcal{P}_0, L^2(\mathbb{R}^d))$ and, for fixed f as above and $\tau \in \mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^d)$ with $\|D\tau\|_{L^\infty} \leq 1/2$, consider

$$T_f(\tau) \coloneqq S_0[\mathcal{P}_0](L_\tau f) - S_0[\mathcal{P}_0](f).$$

Let $s \in (1,2)$ and $\theta \in (0,1)$ be such that $1 = (1-\theta)s$ and consider $\tau \in \mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^d)$ with $\|D\tau\|_{L^\infty} \leq \varepsilon_0$. For any $\tau_0 \in B^s_{\infty,\infty}(\mathbb{R}^d;\mathbb{R}^d)$ with $\|D\tau_0\|_{L^\infty} \leq 1/2$ we have, by the triangle inequality, (5.3) and (5.4), for $t \geq 1$,

$$\begin{aligned} \|T_f(\tau)\|_{\mathcal{H}} &\leq \|T_f(\tau_0)\|_{\mathcal{H}} + t\|T_f(\tau) - T_f(\tau_0)\|_{\mathcal{H}} \\ &\lesssim \log^\beta(e+R)\|\tau_0\|_{B^s_{\infty,\infty}} \|f\|_{L^2} + t(1+R)\|\tau - \tau_0\|_{B^0_{\infty,1}} \|f\|_{L^2} \\ &= \sum_{k=1}^d \big(\log^\beta(e+R)\|\tau_0^{(k)}\|_{B^s_{\infty,\infty}} + t(1+R)\|\tau^{(k)} - \tau_0^{(k)}\|_{B^0_{\infty,1}}\big)\|f\|_{L^2} \end{aligned}$$

where we expanded the Besov norms in terms of the components, that is $\tau = (\tau^{(1)}, \dots, \tau^{(d)})$ and $\tau_0 = (\tau_0^{(1)}, \dots, \tau_0^{(d)})$.

Consider now $\chi \in C^\infty(\mathbb{R}^d)$, supported where $|\omega| \leq 1$, with $\chi(\omega) = 1$ for $|\omega| \leq 1/2$, along with the corresponding Fourier multiplier $\chi(D)$. We write $\tau_0^{(k)} = \chi(D)\tau_0^{(k)} + (1-\chi(D))\tau_0^{(k)}$. By the triangle inequality we obtain

$$\begin{aligned} \|T_f(\tau)\|_{\mathcal{H}} \lesssim \log^\beta(e+R) &\sum_{k=1}^d \|\chi(D)\tau_0^{(k)}\|_{B^s_{\infty,\infty}} \|f\|_{L^2} \\ &+ \sum_{k=1}^d \big(\log^\beta(e+R)\|(1-\chi(D))\tau_0^{(k)}\|_{B^s_{\infty,\infty}} + t(1+R)\|\tau^{(k)} - \tau_0^{(k)}\|_{B^0_{\infty,1}}\big)\|f\|_{L^2}. \end{aligned} \tag{5.5}$$

Since $\|D\tau\|_{L^\infty} \leq \varepsilon_0$, if ε_0 is small enough and assuming now $\tau_0^{(k)} \in \mathcal{A}_{\tau^{(k)}}$ (cf. (5.1)), we have $\|D\tau_0^{(k)}\|_{L^\infty} \leq C\|D\tau^{(k)}\|_{L^\infty} \leq 1/(2\sqrt{d})$, implying in particular that $\|D\tau_0\|_{L^\infty} \leq 1/2$. Moreover

$$\|\chi(D)\tau_0^{(k)}\|_{B^s_{\infty,\infty}} \lesssim \|\tau_0^{(k)}\|_{L^\infty} \lesssim \|\tau^{(k)}\|_{L^\infty}.$$

We also remark that, since the Fourier transform of $\tau^{(k)} - \tau_0^{(k)}$ is supported in the region where $|\omega| \geq 1$ (with reference to (5.1), we have indeed that $\phi_j(\omega) = 1$ in the ball $|\omega| \leq 1$),

$$\tau^{(k)} - \tau_0^{(k)} = (1-\chi(D))\tau^{(k)} - (1-\chi(D))\tau_0^{(k)}.$$

In light of the facts highlighted so far, now we take the infimum of (5.5) over $\tau_0^{(k)} \in \mathcal{A}_{\tau^{(k)}}$. For $t \geq 1$ we obtain

$$\begin{aligned}\|T_f(\tau)\|_{\mathcal{H}} &\lesssim \log^\beta(e+R)\|\tau\|_{L^\infty}\|f\|_{L^2} \\ &\quad + \log^\beta(e+R)\sum_{k=1}^d \tilde{K}(t(1+R)/\log^\beta(e+R), (1-\chi(D))\tau^{(k)}, B^s_{\infty,\infty}, B^0_{\infty,1})\|f\|_{L^2}\end{aligned}$$

where the functional $\tilde{K}$ is defined in Lemma 5.1 and we used the fact that

$$\mathcal{A}_{(1-\chi(D))\tau^{(k)}} = \{(1-\chi(D))f_0 : \ f_0 \in \mathcal{A}_{\tau^{(k)}}\}.$$

Finally, by applying Lemma 5.1,[6] we obtain

$$\begin{aligned}\|T_f(\tau)\|_{\mathcal{H}} &\lesssim \big(\log^\beta(e+R)\|\tau\|_{L^\infty} + \log^{\beta(1-\theta)}(e+R)(1+R)^\theta \sum_{k=1}^d \|(1-\chi(D))\tau^{(k)}\|_{B^1_{\infty,\infty}}\big)\|f\|_{L^2} \\ &= \big(\log^\beta(e+R)\|\tau\|_{L^\infty} + \log^{\beta(1-\theta)}(e+R)(1+R)^\theta \|(1-\chi(D))\tau\|_{B^1_{\infty,\infty}}\big)\|f\|_{L^2}.\end{aligned}$$

On the other hand, we have

$$\|(1-\chi(D))\tau\|_{B^1_{\infty,\infty}} \lesssim \|D\tau\|_{B^0_{\infty,\infty}} \lesssim \|D\tau\|_{L^\infty}.$$

The second inequality is clear from the definition of the $B^0_{\infty,\infty}$ norm. The first inequality follows by a standard argument that we sketch here for the benefit of the reader. Let $\tilde{\phi}'$ be a smooth function in $\mathbb{R}^d$, supported in the annulus $2^{-2} \leq |\omega| \leq 2^4$, with $\tilde{\phi}'(\omega) = 1$ for $2^{-1} \leq |\omega| \leq 2$. Then we write $\tilde{\phi}_j(\omega) = \tilde{\phi}'(2^{-j}\omega)\tilde{\phi}_j(\omega)$ in the definition of the Besov norm, and we observe that the functions $2^j\tilde{\phi}'(2^{-j}\omega)\omega_k/|\omega|^2$, $k = 1, \dots, d$, can be written as $\phi''_k(2^{-j}\omega)$ for some ϕ''_k smooth with compact support. Hence their inverse Fourier transforms have L^1 norm uniformly bounded with respect to j, and the corresponding Fourier multipliers are thus uniformly bounded on $L^\infty(\mathbb{R}^d)$. Similarly, if ϕ'_0 is a smooth function supported in the ball $|\omega| \leq 4$, with $\phi'_0(\omega) = 1$ for $|\omega| \leq 2$ we can write $\phi_0 = \phi'_0\phi_0$ and observe that $(1-\chi(\omega))\phi'_0(\omega)\omega_k/|\omega|^2$ is a smooth function with compact support, and hence defines a Fourier multiplier bound on $L^\infty(\mathbb{R}^d)$.

The desired estimate (1.8) is then proved since s can be chosen arbitrarily close to 1, hence making in turn the exponent θ arbitrarily small. □

Acknowledgements

The authors wish to express their gratitude to Giovanni S. Alberti and Matteo Santacesaria for fruitful conversations on the topics of the manuscript.

The present research was carried out while S. Ivan Trapasso was member of the Machine Learning Genoa (MaLGa) Center, Università di Genova. This material is based upon work supported by the Air Force Office of Scientific Research under award number FA8655-20-1-7027, as well as by Fondazione Compagnia di San Paolo.

The authors are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM).

[6] Precisely, since $(1+R)/\log^\beta(e+R) \geq c_0$ for some $c_0 > 0$ and

$$\tilde{K}(t(1+R)/\log^\beta(e+R), f, B^s_{\infty,\infty}, B^0_{\infty,1}) \leq \tilde{K}(t\max\{c_0^{-1}, 1\}(1+R)/\log^\beta(e+R), f, B^s_{\infty,\infty}, B^0_{\infty,1}),$$

we can resort to (5.2) with t replaced by $t\max\{c_0^{-1}, 1\}(1+R)/\log^\beta(e+R) > 1$.

References

[1] Rima Alaifari, Giovanni S. Alberti, Tandri Gauksson, Adef: an iterative algorithm to construct adversarial deformations, in: 7th International Conference on Learning Representations, ICLR 2019, 2019 (conference paper).
[2] Radu Balan, Maneesh Singh, Dongmian Zou, Lipschitz properties for deep convolutional networks, Contemp. Math. 706 (2018) 129–151.
[3] Mikhail Belkin, Daniel Hsu, Siyuan Ma, Soumik Mandal, Reconciling modern machine-learning practice and the classical bias-variance trade-off, Proc. Natl. Acad. Sci. USA 116 (32) (2019) 15849–15854.
[4] Alberto Bietti, Julien Mairal, Group invariance, stability to deformations, and complexity of deep convolutional representations, J. Mach. Learn. Res. 20 (25) (2019) 1–49.
[5] Joan Bruna, Stéphane Mallat, Invariant scattering convolution networks, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1872–1886.
[6] Emmanuel J. Candès, Laurent Demanet, The curvelet representation of wave propagators is optimally sparse, Commun. Pure Appl. Math. 58 (11) (2005) 1472–1528.
[7] Emmanuel J. Candès, David L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise c^2 singularities, Commun. Pure Appl. Math. 57 (2) (2004) 219–266.
[8] Nicholas Carlini, David Wagner, Towards evaluating the robustness of neural networks, in: 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 2017, pp. 39–57.
[9] Wojciech Czaja, Weilin Li, Analysis of time-frequency scattering transforms, Appl. Comput. Harmon. Anal. 47 (1) (2019) 149–171.
[10] Ian Goodfellow, Jonathon Shlens, Christian Szegedy, Explaining and harnessing adversarial examples, in: 3rd International Conference on Learning Representations, ICLR 2015, 2015 (conference paper).
[11] Kanghui Guo, Gitta Kutyniok, Demetrio Labate, Sparse multidimensional representations using anisotropic dilation and shear operators, in: Wavelets and Splines: Athens 2005, Nashboro Press, Brentwood, TN, 2006, pp. 189–201.
[12] Trevor Hastie, Andrea Montanari, Saharon Rosset, Ryan J. Tibshirani, Surprises in high-dimensional ridgeless least squares interpolation, Ann. Stat. 50 (2) (2022) 949–986.
[13] Michael Koller, Johannes Großmann, Ullrich Monich, Holger Boche, Deformation stability of deep convolutional neural networks on Sobolev spaces, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 6872–6876.
[14] Alexey Kurakin, Ian Goodfellow, Samy Bengio, Adversarial examples in the physical world, in: R.V. Yampolskiy (Ed.), Artificial Intelligence - Safety and Security, Chapman and Hall/CRC, 2018.
[15] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature 521 (2015) 436–444.
[16] Jacques-Louis Lions, Some remarks on variational inequalities, in: Proc. Internat. Conf. on Functional Analysis and Related Topics, Univ. Tokyo Press, Tokyo, 1970, pp. 269–282.
[17] Alessandra Lunardi, Interpolation Theory, third edition, Edizioni della Normale, Pisa, 2018.
[18] Lech Maligranda, On interpolation of nonlinear operators, Comment. Math. Prace Mat. 28 (2) (1989) 253–275.
[19] Stéphane Mallat, A Wavelet Tour of Signal Processing. The Sparse Way, With contributions from Gabriel Peyré, third edition, Elsevier/Academic Press, Amsterdam, 2009.
[20] Stéphane Mallat, Group invariant scattering, Commun. Pure Appl. Math. 65 (10) (2012) 1331–1398.
[21] Song Mei, Andrea Montanari, The generalization error of random features regression: precise asymptotics and the double descent curve, Commun. Pure Appl. Math. 75 (4) (2022) 667–766.
[22] Yves Meyer, Wavelets and Operators, Cambridge University Press, Cambridge, 1992.
[23] Preetum Nakkiran, Gal Kaplun, Yamini Bansal, Tristan Yang, Boaz Barak, Ilya Sutskever, Deep double descent: where bigger models and more data hurt, J. Stat. Mech. Theory Exp. 4003 (12) (2021) 32.
[24] Jaak Peetre, Interpolation of Lipschitz operators and metric spaces, Mathematica 12 (35) (1970) 325–334.
[25] Yoshihiro Sawano, Theory of Besov Spaces, Springer, Singapore, 2018.
[26] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, Frank Lenzen, Variational Methods in Imaging, Springer, New York, 2009.
[27] Elias M. Stein, Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press, Princeton, NJ, 1993.
[28] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, Rob Fergus, Intriguing properties of neural networks, in: 2nd International Conference on Learning Representations, ICLR 2014, 2014 (conference paper).
[29] Luc Tartar, Interpolation non linéaire et régularité, J. Funct. Anal. 9 (1972) 469–489.
[30] Hans Triebel, Theory of Function Spaces, 1983 edition, Reprint of Birkhäuser/Springer, Basel AG, Basel, 2010.
[31] Alain Trouvé, Diffeomorphisms groups and pattern matching in image analysis, Int. J. Comput. Vis. 28 (1998) 213–221.
[32] Irène Waldspurger, Exponential decay of scattering coefficients, in: 2017 International Conference on Sampling Theory and Applications (SampTA), Tallinn, Estonia, 2017, pp. 143–146.
[33] Thomas Wiatowski, Helmut Bölcskei, Deep convolutional neural networks based on semi-discrete frames, in: 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 2015, pp. 1212–1216.
[34] Thomas Wiatowski, Helmut Bölcskei, A mathematical theory of deep convolutional neural networks for feature extraction, IEEE Trans. Inf. Theory 64 (3) (2018) 1845–1866.
[35] Laurent Younes, Shapes and Diffeomorphisms, Springer, Berlin, 2019.
[36] Dongmian Zou, Radu Balan, Maneesh Singh, On Lipschitz bounds of general convolutional neural networks, IEEE Trans. Inf. Theory 66 (3) (2020) 1738–1759.