GENERALIZED MODULI OF CONTINUITY UNDER IRREGULAR OR RANDOM DEFORMATIONS VIA MULTISCALE ANALYSIS

FABIO NICOLA AND S. IVAN TRAPASSO

ABSTRACT. Motivated by the problem of robustness to deformations of the input for deep convolutional neural networks, we identify signal classes which are inherently stable to irregular deformations induced by distortion fields $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, to be characterized in terms of a generalized modulus of continuity associated with the deformation operator.

Resorting to ideas of harmonic and multiscale analysis, we prove that for signals in multiresolution approximation spaces U_s at scale s, stability in L^2 holds in the regime $\|\tau\|_{L^\infty}/s \ll 1$ — essentially as an effect of the uncertainty principle. Instability occurs when $\|\tau\|_{L^\infty}/s \gg 1$, and we provide a sharp upper bound for the asymptotic growth rate. The stability results are then extended to signals in the Besov space $B^{d/2}_{2,1}$ tailored to the given multiresolution approximation. We also consider the case of more general time-frequency deformations.

Finally, we provide stochastic versions of the aforementioned results, namely we study the issue of stability in mean when $\tau(x)$ is modeled as a random field (not bounded, in general) with identically distributed variables $|\tau(x)|$, $x \in \mathbb{R}^d$.

1. Introduction

1.1. **The problem of stability to deformations.** In this note we consider a mathematical problem motivated by the theory and practice of machine learning, that is the robustness of the output of a neural network under modifications of the input datum. Let us briefly illustrate this issue by considering a function $f\colon \mathbb{R}^d \to \mathbb{R}$. Some basic transformations to be taken into account involve *intensity perturbations*, that is $\tilde{f}(x) = f(x) + h(x)$ for some $h\colon \mathbb{R}^d \to \mathbb{R}$, or *signal deformations*, namely $\tilde{f}(x) = F_\tau f(x) \coloneqq f(x - \tau(x))$ for some distortion field $\tau\colon \mathbb{R}^d \to \mathbb{R}^d$. We stress that this model encompasses natural transformations such as translations or rotations.

Regardless of the variety of the architectures, the network under our attention can be represented by a map Φ from $L^2(\mathbb{R}^d)$ to some Banach space with norm $|||\cdot|||$. In order to better appreciate the relevant phenomena, let us consider the classification setting where Φ acts as a feature extractor. A fair degree of stability of Φ to small

2020 *Mathematics Subject Classification.* 94A12, 42B35, 42C15, 68T05.

Key words and phrases. Stability, generalized modulus of continuity, deformations, amalgam spaces, multiresolution analysis, convolutional neural networks.

transformations of the input signal is a naturally desirable property in several contexts. For example, consider the classic learning task of digit recognition from images of handwritten symbols, where the input signals suffer from both intra-class and inter-class variance, due for instance to differences in the position of the digit with respect to the background or handwriting styles. As a rule of thumb, it is expected that a small distortion of f into $\tilde{f}$ should correspond to small norm discrepancy $|||\Phi(\tilde{f}) - \Phi(f)|||$ at the level of features.

The previous remarks thus lead us to require that Φ enjoys a Lipschitz regularity condition:

$$|||\Phi(\tilde{f}) - \Phi(f)||| \le C\|\tilde{f} - f\|_{L^2}, \quad f, \tilde{f} \in L^2(\mathbb{R}^d). \tag{1.1}$$

The smallest constant $C > 0$ for which such an estimate holds will be denoted by $\mathrm{Lip}(\Phi)$. Moreover, in the particular case of a deformation $\tilde{f} = F_\tau f$ of f, it would be desirable for $|||\Phi(F_\tau f) - \Phi(f)|||$ to be small whenever τ is small with respect to some distortion metric. We can distinguish at least two different angles on the matter:

- In keeping with the spirit of geometric deep learning [5], *structural stability* guarantees are inferred from global and local invariance requirements that are *a priori* embedded in the design of the network. A prominent example in this connection is provided by the analysis of the scattering transform introduced in [18] (see also [6]: if Φ is a scattering transform with fixed wavelets filters, modulus nonlinearity and no pooling stages, it was proved in [18, Proposition 2.5] that Φ is a non-expansive transform (i.e., $\mathrm{Lip}(\Phi) = 1$), and in [18, Theorem 2.12] that, for every $\tau \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ with $\|\nabla\tau\|_{L^\infty} \le 1/2$,

$$|||\Phi(F_\tau f) - \Phi(f)||| \le C(2^{-J}\|\tau\|_{L^\infty} + \max\{J, 1\}\|\nabla\tau\|_{L^\infty} + \|H\tau\|_{L^\infty})\|f\|_{\mathrm{scatt}}, \tag{1.2}$$

  where $\|f\|_{\mathrm{scatt}}$ is a mixed $\ell^1(L^2)$ scattering norm (which is finite for functions with a logarithmic Sobolev-type regularity), $H\tau$ denotes the Hessian of τ and 2^J is the coarsest scale in the dyadic multiscale analysis associated with the network filters.
- In the case where little information on the architecture of the network is available or exploitable, one can only assume to satisfy a Lipschitz condition as in (1.1). In such cases, stability results for Φ can be possibly *inherited* from the inherent robustness to deformations of certain input signal classes. This amount to determine a subset $\mathcal{E} \subset L^2(\mathbb{R}^d)$ such that bounds for $\|F_\tau f - f\|_{L^2}$ in terms of some complexity metric of τ can be proved if $f \in \mathcal{E}$. This is the essence of the *decoupling method* introduced in [25, 26, 15] to obtain stability results for generalized scattering networks by exploiting *sensitivity estimates* of the form $\|F_\tau f - f\|_{L^2} \le C_{\mathcal{E}}\|\tau\|_{L^\infty}^{\alpha_{\mathcal{E}}}\|f\|_{L^2}$, which are proved for several classes

of interest (including Lipschitz, band-limited and cartoon functions) and deformations $\tau \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ with $\|\nabla\tau\|_{L^\infty}$ sufficiently small[1].

A detailed comparison between Mallat's scattering transform and generalized scattering networks would lead us too far. For our purposes, we just stress that in both cases the results are proved for regular (i.e., at least C^1) deformations. In the case of the scattering transform, one is ultimately confronted with the interplay between the network multiscale architecture and the deformation regularity. Consider the case where f is a band-pass function; roughly speaking, the condition $\|\nabla\tau\|_{L^\infty} \le 1/2$ guarantees that $F_\tau f$ is still localized in frequency, essentially in the same band of f, therefore a stability result as in (1.2) is reasonable (although highly non-trivial to prove) since the network separates scales by design.

On the other hand, the scope of the decoupling method goes beyond the analysis of generalized scattering transforms: the weak requirement that Φ is Lipschitz stable as in (1.1) allows us to virtually encompass any neural network where detailed information on structural stability is merely not available. Actually, while most of real-life neural networks are empirically observed to enjoy Lipschitz stability [22], assuming solely this condition about the feature extractor is a worst-case scenario, since other elusive forms of regularity are heuristically expected to occur as well — such as regularization and cancellation phenomena across hidden layers. In fact, the mathematical literature in this respect is quite limited (see e.g., [2, 27]) and the available provable bounds for $\mathrm{Lip}(\Phi)$ are usually quite pessimistic, as they do not exploit further structural information on the network.

Let us also highlight that, as observed in [18], the condition $\|\nabla\tau\|_{L^\infty} \le 1/2$ can be relaxed to $\|\nabla\tau\|_{L^\infty} < 1$ but then the constant blows up when $\|\nabla\tau\|_{L^\infty} \to 1$. The same remark applies to the constants $C_{\mathcal{E}}$ of sensitivity bounds proved in [25] for band-limited functions and in [17] for functions in the Sobolev space[2] $H^1(\mathbb{R}^d)$. It is thus natural to wonder whether stability results can be derived if $\|\nabla\tau\|_{L^\infty} \ge 1$ (therefore $x \mapsto x - \tau(x)$ is no longer invertible) or even for less regular deformations, such as discontinuous ones. Broadly speaking, irregular perturbations such as local pixel shuffling of an image proved to be involved in sophisticate adversarial models such as pixel deflection [21]. They could also be used to model local distortion errors arising in signal encoding, where robustness of classification is naturally expected, as well as to compare contiguous frames of a video where pixels locally move in an irregular fashion (i.e., discontinuous optical flows, pose estimation).

[1]Precisely, $\|\nabla\tau\|_{L^\infty} \le 1/2d$ in [25] and $\|\nabla\tau\|_{L^\infty} \le 1/2$ in [18]. This discrepancy is due to the definition $\|\nabla\tau\|_{L^\infty} := \||\nabla\tau|\|_{L^\infty}$ where $|\nabla\tau|$ is the Frobenius norm of the matrix $\nabla\tau(x)$ in [18] and the ℓ^∞ norm of its entries in [25].

[2]Actually, the result in [17] is stated for functions in the Sobolev space $H^2(\mathbb{R}^d)$. Inspection of the proof and an easy density argument show that it actually holds for functions in the Sobolev space $H^1(\mathbb{R}^d)$ of functions $f \in L^2(\mathbb{R}^d)$ such that $\|\nabla f\|_{L^2} < \infty$.

1.2. **Robustness to irregular deformations.** The previous discussion suggests that the interplay between the deformation regularity and the network structure is a subtle issue. In fact, it turns out that, unless a network is purposefully designed to be stable to irregular deformations, stability results for Φ at this low-regularity level can only be obtained via the decoupling methods, hence passing on the robustness issue to the input signal class. Indeed, in the context of irregular deformations, even for well structured networks such as the wavelet scattering ones, it may happen that $|||\Phi(F_\tau f) - \Phi(f)||| \approx \|F_\tau f - f\|_{L^2}$.

To be more precise, let us illustrate two kinds of peculiar phenomena that could occur when dealing with irregular deformations — see also [20] for further details.

(a) Consider a band-pass function f oscillating at frequency $1/s$ ($s > 0$ being the scale); even if $\|\tau\|_{L^\infty}$ is small, it may very well happen that the energy of $F_\tau f$ is amplified by a factor $(\|\tau\|_{L^\infty}/s)^{d/2}$; see Figure 1A. Hence, if Φ is any energy preserving map ($\|f\|_{L^2} \lesssim |||\Phi(f)||| \lesssim \|f\|_{L^2}$) then it follows from the triangle inequality that $|||\Phi(F_\tau f) - \Phi(f)|||/\|f\|_{L^2} \gtrsim (\|\tau\|_{L^\infty}/s)^{d/2}$ when $\|\tau\|_{L^\infty}$ is large *compared to* s.

(b) Let f be a band-pass function, as above, oscillating at frequency $1/s$; even if $\|\tau\|_{L^\infty}$ is small, when $\|\tau\|_{L^\infty}$ is comparable to s it may happen that f and $F_\tau f$ are localized in different dyadic frequency bands, see Figure 1B. In particular, if Φ is a wavelet scattering network, their energy will propagate along separate frequency paths and thus the error $|||\Phi(F_\tau f) - \Phi(f)|||^2 \approx |||\Phi(F_\tau f)|||^2 + |||\Phi(f)|||^2$ will not be small if Φ is energy preserving.

These phenomena are evident sources of instability in the case where $\|\tau\|_{L^\infty}/s \gg 1$ and $\|\tau\|_{L^\infty}/s \approx 1$ respectively. In passing, note that in order for $F_\tau f$ to be well defined as an element of $L^2(\mathbb{R}^d)$ for every $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, independently of the representative of f in $L^2(\mathbb{R}^d)$, f must be assumed continuous at least — see again Figure 1A for a concrete reference.

We thus conclude that for irregular deformations one is forced to shift the robustness problem from the network architecture to the signal class. In keeping with the spirit of mathematical analysis, let us emphasize that proving bounds for $\|F_\tau f - f\|_{L^2}$ in terms of the deformation size $\|\tau\|_{L^\infty}$ and $\|f\|_{L^2}$ for suitable signal classes can be thought of as a generalization of a typical problem of harmonic analysis, where the differentiability properties of certain function spaces are quantitatively measured in terms of the magnitude of some L^p modulus of continuity $\omega_p[f](t) := \|f(x+t) - f(x)\|_{L^p_x}$, $t \in \mathbb{R}^d$, as $|t| \to 0$ — cf. for instance [23, Chapter V] for a classic reference on the topic. This approach allows one to fine tune the regularity scale of a signal in a very precise way.

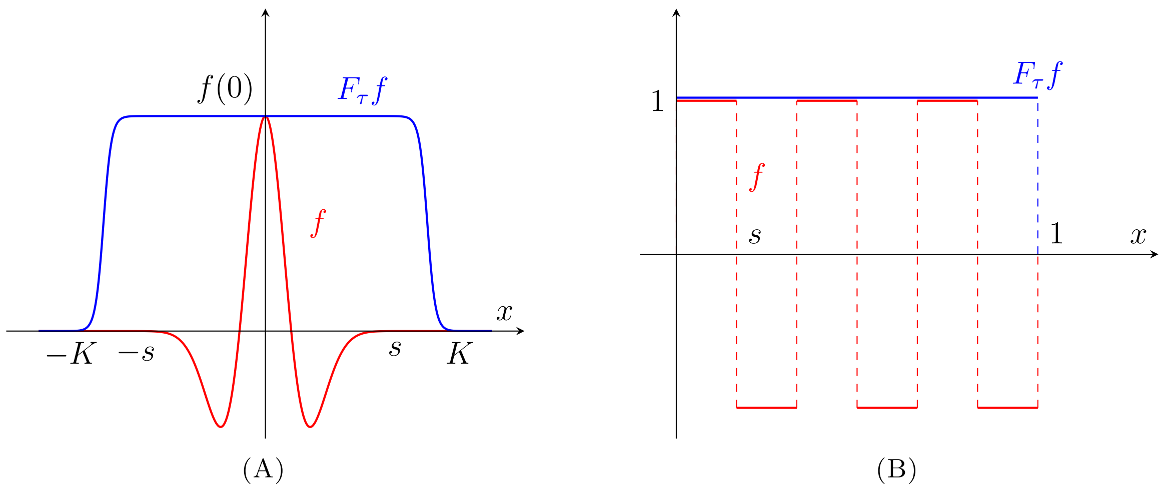

FIGURE 1. *(A) A signal f supported on $[-s,s]$ and its deformation $F_\tau f$, where $\tau(x)=x$ for $|x|<K$, with $K>s$. The plateau level corresponds to the value $f(0)$. The operator F_τ (with the choice of τ specified above) performs a single-point sampling of f, hence it does not make sense on discontinuous signals.*

(B) A signal f localized in frequency where $|\omega| \approx s^{-1}$. With the choice of the deformation $\tau = s\mathbb{1}_{\{f=-1\}}$, the signal $F_\tau f$ is low-pass (a similar example with f continuous is easily obtained by smoothing the steps).

The previous remarks motivate focusing on a family of spaces where a precise tuning of the scale is available, in order to elucidate the relationship with the deformation size. We resort again to ideas and tools of modern harmonic analysis, namely we consider multiresolution approximation spaces $U_s \subset L^2(\mathbb{R}^d)$, $s>0$ [19], with a Riesz basis given by a sequence of functions of the type $\phi_{s,n}(x) := s^{-d/2}\phi((x-ns)/s)$, $n \in \mathbb{Z}^d$, where ϕ is a fixed filter satisfying certain mild regularity and decay conditions (cf. Assumptions A, B and C in Section 5 below). Different choices of ϕ result in diverse multiresolution approximations, including band-limited functions and polynomial splines of order $n \geq 1$ — see the discussion in Example 5.1 below for more details. In general, the introduction of a fixed resolution scale is also natural as a mathematical model of a concrete signal capture system — cf. the general A/D and D/A conversion schemes in [19, Section 3.1.3], and also [3, 4] for a similar limited-resolution assumption in a discrete setting. The scale s (or rather s^{-1}) can also be viewed as a rough measure of the complexity of the input signal, and the previous discussion suggests that the ratio $\|\tau\|_{L^\infty}/s$ should appear in sensitivity bounds rather than just $\|\tau\|_{L^\infty}$, which is also expected in order to have dimensionally consistent estimates.

1.3. **Generalized moduli of continuity for multiresolution spaces.** The core of our first result can be presented as follows. Under suitable assumptions on ϕ there exists a constant $C>0$ such that, for every $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $s>0$,

$$\|F_\tau f - f\|_{L^2} \le \begin{cases} C(\|\tau\|_{L^\infty}/s)\|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \le 1) \\ C(\|\tau\|_{L^\infty}/s)^{d/2}\|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \ge 1) \end{cases}, \quad f \in U_s. \tag{1.3}$$

Stability guarantees for any Lipschitz network Φ can thus be inferred by the fact that $|||\Phi(F_\tau f) - \Phi(f)||| \le \mathrm{Lip}(\Phi)\|F_\tau f - f\|_{L^2}$. We refer to Theorem 5.7 for precise statements. The estimate (1.3) for $\|\tau\|_{L^\infty}/s \le 1$ recovers and extends the results proved in [25] for band-limited functions, now without any regularity assumption on the deformation. In Section 7 we show the sharpness of the estimate (1.3) in both regimes $\|\tau\|_{L^\infty}/s \gg 1$ and $\|\tau\|_{L^\infty}/s \ll 1$.

In short, whenever we have a Lipschitz bound, we have a stability result in the regime $\|\tau\|_{L^\infty}/s \ll 1$, which can be explained in heuristic terms as one of the manifold forms of the uncertainty principle — see below for further comments in this connection. Observe also that the rate of instability agrees with that of the previous discussion in *(a)* when small-size oscillations, compared with the size of the deformation (namely, if $\|\tau\|_{L^\infty}/s \gg 1$), are allowed.

Interestingly, for fixed f, we have in any case $|||\Phi(F_\tau f) - \Phi(f)||| = O(\|\tau\|_{L^\infty})$ as $\|\tau\|_{L^\infty} \to 0$, although this asymptotic estimate is not uniform with respect to s. In fact, in sharp contrast with (1.2), the factor $1/s$ in front of $\|\tau\|_{L^\infty}$ associates with a feature of the input signal (i.e., the resolution of f), whereas the invariance resolution 2^{-J} in (1.2) is a fixed quantity that depends on the architecture of the network. However, the example in Figure 1B above shows that in the framework of irregular deformations, even for a fixed wavelet scattering network, we cannot hope for an estimate whose quality does not deteriorate when $\|\tau\|_{L^\infty}$ becomes comparable to the size of the oscillations of f. We thus infer that while the choice of wavelet filters is crucial in [18] to manufacture a transform that is Lipschitz stable to the action of small diffeomorphisms, robustness under *small and irregular* deformations obeys a more general rule, as already anticipated above. In this connection, we address the reader to the aforementioned paper [20], where instability results are proved for wavelet scattering networks and deformations at low regularity levels, namely for distortion fields $\tau \in C^\alpha(\mathbb{R}^d;\mathbb{R}^d)$ with $0 \le \alpha < 1$.

The assumption that the input signal f belongs to U_s could be judged not realistic in practice. Rather, we often deal with signals that can be well approximated in low-complexity spaces. For such signal classes we have again a stability result, which is briefly outlined here in low-dimensional settings for simplicity — we refer to Theorem 5.9 for a general and precise statement. Let $V_j := U_{2^j}$, $j \in \mathbb{Z}$, be a multiresolution analysis of $L^2(\mathbb{R}^d)$. There exists a constant $C>0$ such that, for every $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$,

$$\|F_\tau f - f\|_{L^2} \le C\|\tau\|_{L^\infty}^{d/2}\|f\|_{\dot{B}^{d/2}_{2,1}}, \qquad d=1,2,$$

for any $f \in L^2(\mathbb{R}^d)$ such that $\|f\|_{\dot{B}^{d/2}_{2,1}} < \infty$, where $\dot{B}^{d/2}_{2,1}$ denotes the homogeneous Besov space tailored to the given multiresolution analysis [19, Section 9.2.3]. This regularity level looks optimal in general — as already observed, f should be at least continuous, and therefore in $B^{d/2}_{2,1}(\mathbb{R}^d)$ if we consider the scale of L^2-based Besov spaces as a reference.

In Section 6 we prove estimates in the same spirit for more general time-frequency deformations of the type $F_{\tau,\omega}f(x) = e^{i\omega(x)}f(x-\tau(x))$. Modulation deformations are relevant in case of spectral distortions of input signals. These deformations are approached here in a "perturbative" way — that is, by reducing to the results already proved for the case $\omega \equiv 0$.

The main technical tools behind our results are the properties of certain spaces $X^{p,q}_r$, tailored to the deformation scale $r>0$. Such function spaces are usually referred to as Wiener amalgam spaces and were introduced by Feichtinger in the '80s [10, 11]. As the name suggests, they are obtained by means of a norm that amalgamates a local summability of L^p type on balls of radius r with an L^q behaviour at infinity. They are of current use in harmonic analysis and PDEs, possibly under slightly different names and forms — see for instance [8, 24].

In Section 3 we collect the main properties of these spaces, while in Section 4 we focus on the space $X^{\infty,2}_r$ of locally bounded functions, uniformly at the scale r, with L^2 decay. This choice should not be intended as a mere technical workaround: in Proposition 4.1 we prove that this class is indeed the optimal choice when dealing with arbitrary bounded deformations, since for functions $f \in X^{\infty,2}_r \cap C(\mathbb{R}^d)$ we have the clear-cut characterization

$$\|f\|_{X^{\infty,2}_r} = \max\{\|F_\tau f\|_{L^2} : \tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d),\ \|\tau\|_{L^\infty} \le r\}.$$

Moreover, the local control offered by $X^{p,q}_r$ can be effectively exploited to prove a crucial embedding, cf. Theorem 5.3, which can be heuristically referred to as a reverse Hölder-type inequality for signals in U_s in the spirit of [24, Lemma 2.2], which can be regarded as a novel form of the already mentioned uncertainty principle. Intuitively, if a function f is localized in a low-frequency ball of radius R^{-1} centered at the origin, then f is approximately constant on balls of radius R. As a result, deliberately ignoring the effect of the tails, its L^∞ norm on a ball of radius $r<R$ can be roughly bounded by the L^2 norm on the same ball (up to a factor $(R/r)^{d/2}$). Strictly speaking, amalgam spaces are needed to put these heuristic remarks on a rigorous ground, leading precisely to the reverse Hölder-type inequality stated in Theorem 5.3.

We adopted so far a deterministic model for the deformation, namely a ball in $L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, without any additional structure, and therefore we provided stability guarantees in a worst-case scenario. In Section 8 we assume instead that τ is a random field with identically distributed variables $|\tau(x)|$, $x\in\mathbb{R}^d$. We accordingly study the issue of stability in mean, providing stochastic versions of the above results. For example, we prove that

$$\mathbb{E}\|F_\tau f - f\|^2_{L^2} \le C\mathbb{E}[|\tau|^d]\|f\|^2_{\dot{B}^{d/2}_{2,1}}, \qquad d=1,2, \tag{1.4}$$

see Theorem 8.1 for the precise statement in any dimension, and for similar results when f belongs to limited-resolution spaces U_s as above. Here we set $\mathbb{E}[|\tau|^d]$ for $\mathbb{E}[|\tau(x)|^d]$, the latter being in fact independent of x. We also emphasize that the field τ is no longer assumed to be bounded.

2. Notation

The open unit ball of $\mathbb{R}^d$ with radius $r>0$ and centered at the origin is denoted by B_r.

We introduce a number of operators acting on $f\colon\mathbb{R}^d\to\mathbb{C}$:

- the dilation D_λ by $\lambda\neq 0$: $D_\lambda f(y)=f(\lambda y)$;
- the translation T_x by $x\in\mathbb{R}^d$: $T_x f(y)=f(y-x)$;
- the modulation M_ξ by $\xi\in\mathbb{R}^d$: $M_\xi f(y)=e^{iy\cdot\xi}f(y)$;
- the reflection: $\mathcal{I}f(y)=f(-y)$;
- the Fourier transform (whenever meaningful, e.g. if $f\in L^1(\mathbb{R}^d)$), normalized here as
$$\widehat{f}(\omega)=\mathcal{F}(f)(\omega)=\int_{\mathbb{R}^d} e^{-i\omega\cdot y}f(y)dy.$$

The space $L^\infty(\mathbb{R}^d;\mathbb{R}^d)$ contains all the measurable vector fields $\tau\colon\mathbb{R}^d\to\mathbb{R}^d$ such that
$$\|\tau\|_{L^\infty} := \operatorname*{ess\,sup}_{y\in\mathbb{R}^d}|\tau(y)|<\infty.$$

We introduce the inhomogeneous magnitude $\langle y\rangle$ of $y\in\mathbb{R}^d$, that is $\langle y\rangle := (1+|y|^2)^{1/2}$.

The symbol $\mathbb{1}_E$ will be used to denote the characteristic function of a set E.

While in the statements of the results we will keep track of absolute constants in the estimates, in the proofs we will heavily make use of the symbol $X\lesssim Y$, meaning that the underlying inequality holds up to a universal positive constant factor, namely
$$X\lesssim Y \quad\Longrightarrow\quad \exists C>0\,:\,X\le CY.$$

Moreover, $X\asymp Y$ means that X and Y are *equivalent quantities*, that is both $X\lesssim Y$ and $X\lesssim Y$ hold.

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

3. Multiscale Wiener amalgam spaces

The following family of function spaces will play a key role in the following.

Definition 3.1. *For $1 \le p, q \le \infty$ and $r > 0$, we denote by $X_r^{p,q}$ the space of all the complex-valued measurable functions in $\mathbb{R}^d$ such that*

$$\|f\|_{X_r^{p,q}} := \left(\int_{\mathbb{R}^d} \|T_{-x} f\|^q_{L^p(B_r)} dx \right)^{1/q} < \infty, \tag{3.1}$$

with obvious modifications if $q = \infty$. In the case where $r = 1$ we write $X^{p,q}$ for $X_1^{p,q}$.

Let us emphasize that $X^{\infty,1}$ coincides with the well known Wiener space of harmonic analysis (cf. e.g. [13, Section 6.1]). More generally, $X^{p,q}$ coincides with the Wiener amalgam space $W(L^p, L^q)$ of functions with local regularity of L^p type and global decay of L^q type, first introduced by Feichtinger in the '80s [10, 11]; recall that the latter is a Banach space provided with the norm

$$\|f\|_{W(L^p,L^q)} = \left(\int_{\mathbb{R}^d} \|T_{-x} f\|^q_{L^p(Q)} dx \right)^{1/p},$$

where $Q \subset \mathbb{R}^d$ is an arbitrary compact set with non-empty interior. In fact, different choices of Q yield equivalent norms; typical choices include $Q = B_1$ and $Q = [0,1]^d$. Moreover, the following equivalent discrete-type norm can be used to measure the amalgamated regularity:

$$\|f\|_{X^{p,q}} \asymp \left(\sum_{k \in \mathbb{Z}^d} \|T_{-k} f\|^q_{L^p(Q)} \right)^{1/q}, \quad Q = [0,1]^d. \tag{3.2}$$

We also highlight that $X_r^{p,p}$ coincides with $L^p(\mathbb{R}^d)$ as set for any $1 \le p \le \infty$, but the norm is rescaled:

$$\|f\|_{X_r^{p,p}} = r^{d/p} \|f\|_{L^p}.$$

A similar change-of-scale property holds with respect to $X^{p,q}$, in the sense of the following result.

Lemma 3.2. *For any $1 \le p, q \le \infty$ and $r > 0$, we have that $X_r^{p,q} = X^{p,q}$ as sets, and*

$$\|f\|_{X_r^{p,q}} = r^{d\left(\frac{1}{p}+\frac{1}{q}\right)} \|D_r f\|_{X^{p,q}}.$$

Proof. Let us consider the case $p,q<\infty$ for conciseness, the other cases following easily. A straightforward computation shows that

$$\begin{aligned}
\|f\|_{X_r^{p,q}} &= \left(\int_{\mathbb{R}^d}\left(\int_{B_r}|f(x+y)|^p dy\right)^{q/p} dx\right)^{1/q}\\
&= r^{d/p}\left(\int_{\mathbb{R}^d}\left(\int_{B_1}|f(x+rz)|^p dz\right)^{q/p} dx\right)^{1/q}\\
&= r^{d/p}\left(\int_{\mathbb{R}^d}\left(\int_{B_1}|D_r f(r^{-1}x+z)|^p dz\right)^{q/p} dx\right)^{1/q}\\
&= r^{d(1/p+1/q)}\left(\int_{\mathbb{R}^d}\left(\int_{B_1}|D_r f(x+z)|^p dz\right)^{q/p} dx\right)^{1/q},
\end{aligned}$$

that is the claim. $\square$

For future reference let us examine some properties of the spaces $X_r^{p,q}$. First, we prove an embedding result that will be often used below.

Proposition 3.3. *For any $1\le p_1,p_2,q\le\infty$ with $p_1\le p_2$, and $r>0$, we have*

$$\|f\|_{X_r^{p_1,q}} \le C r^{d\left(\frac{1}{p_1}-\frac{1}{p_2}\right)}\|f\|_{X_r^{p_2,q}},$$

where the constant $C>0$ depends only on d.

Proof. Fix $x\in\mathbb{R}^d$ and consider the mapping $h_x\colon y\mapsto |f(x+y)|$. The standard Hölder inequality on the ball B_r yields, with ρ such that $1/p_1=1/p_2+1/\rho$,

$$\begin{aligned}
\|T_{-x}f\|_{L^{p_1}(B_r)} &= \|h_x\cdot \mathbb{1}_{B_r}\|_{L^{p_1}(B_r)}\\
&\le \|T_{-x}f\|_{L^{p_2}(B_r)}\|1_{B_r}\|_{L^\rho}\\
&\le (Cr^d)^{\left(\frac{1}{p_1}-\frac{1}{p_2}\right)}\|T_{-x}f\|_{L^{p_2}(B_r)},
\end{aligned}$$

where C is the volume of the d-ball with radius 1. The claim thus follows. $\square$

In the following results we illustrate the behaviour of the spaces $X_r^{p,q}$ under convolution and dilations. In fact, the case with $r=1$ is covered by the standard theory of amalgam spaces (cf. [10, 16] and [7, Proposition 2.2] respectively), hence the result for $r\ne 1$ follows by rescaling the norms in accordance with Lemma 3.2.

Proposition 3.4. *For any $r>0$ and $1\le p_1,p_2,p,q_1,q_2,q\le\infty$ such that*

$$\frac{1}{p_1}+\frac{1}{p_2}=1+\frac{1}{p},\qquad \frac{1}{q_1}+\frac{1}{q_2}=1+\frac{1}{q},$$

we have

$$\|f * g\|_{X_r^{p,q}} \leq C r^{-d} \|f\|_{X_r^{p_1,q_1}} \|g\|_{X_r^{p_2,q_2}},$$

for a constant $C > 0$ that depends only on d.

Proposition 3.5. *For any $r, s > 0$ and $1 \leq p, q \leq \infty$ we have*

$$\|D_s f\|_{X_r^{p,q}} \leq \begin{cases} C s^{-d\max(1/p,1/q)} \|f\|_{X_r^{p,q}} & (0 < s \leq 1) \\ C s^{-d\min(1/p,1/q)} \|f\|_{X_r^{p,q}} & (s \geq 1) \end{cases},$$

for a constant $C > 0$ that depends only on d.

4. L^∞ deformations and the space $X_r^{\infty,2}$

Let us consider the class of deformation mappings F_τ associated with distortion functions $\tau \colon \mathbb{R}^d \to \mathbb{R}^d$ by setting

$$F_\tau f(x) := f(x - \tau(x)),$$

where $f \colon \mathbb{R}^d \to \mathbb{C}$.

We prove that the class $X_r^{\infty,2}$ is the optimal choice as far as sensitivity bounds for arbitrary bounded deformations are concerned. The second part of the following result can be regarded as a linearization of a maximal operator (cf. [12, Section 6.1.3]).

Proposition 4.1. *We have*

$$\|F_\tau f\|_{L^2} \leq \|f\|_{X_r^{\infty,2}}, \quad r = \|\tau\|_{L^\infty}, \tag{4.1}$$

for every $f \in X_r^{\infty,2} \cap C(\mathbb{R}^d)$ and $\tau \in L^\infty(\mathbb{R}^d; \mathbb{R}^d)$.

More precisely, for every function $f \in X_r^{\infty,2} \cap C(\mathbb{R}^d)$, we have the characterization

$$\|f\|_{X_r^{\infty,2}} = \max\{\|F_\tau f\|_{L^2} : \tau \in L^\infty(\mathbb{R}^d; \mathbb{R}^d),\ \|\tau\|_{L^\infty} \leq r\}. \tag{4.2}$$

Remark 4.2. *Note that the continuity assumption on $f \in X_r^{\infty,2}$ is essential in the statement, otherwise $f(x - \tau(x))$ may not even be well defined in L^2 (i.e., independent of the representative f), as evidenced by the case $\tau(x) = x$ for $x \in B_R$ and small $R > 0$. See also Figure 1A in this connection.*

Proof of Proposition 4.1. It is clear that, for almost every $x \in \mathbb{R}^d$,

$$|f(x - \tau(x))| \leq \sup\{|f(x - y)| : y \in \mathbb{R}^d,\ |y| \leq \|\tau\|_{L^\infty}\},$$

and thus (4.1) follows after taking the L^2 norm (the above supremum is the same as the essential supremum because f is continuous).

For what concerns (4.2), it is enough to prove that

$$\|f\|_{X_r^{\infty,2}} \leq \max\{\|F_\tau f\|_{L^2} : \tau \in L^\infty(\mathbb{R}^d; \mathbb{R}^d),\ \|\tau\|_{L^\infty} \leq r\}.$$

To this aim, notice that if we could design a measurable correspondence τ between $x \in \mathbb{R}^d$ and a point $y^* = \tau(x) \in \overline{B_r}$ where the function $\overline{B_r} \ni y \mapsto |f(x-y)|$ attains its maximum, then

$$\max_{|y|\le r} |f(x-y)| = |f(x-\tau(x))| = |F_\tau(x)|,$$

and the desired conclusion would follow once taking the L^2 norm. The existence of such a measurable selector is a consequence of the measurable maximum theorem [1, Theorem 18.19] (in fact, an easier argument would give (4.2) with the supremum in place of the maximum, cf. [12, Section 6.1.3]). $\square$

The following result provides a sensitivity bound for L^2 functions which are locally (i.e., on every compact subset) Lipschitz continuous, uniformly at the deformation scale. It should be compared with the result in [17], valid for functions in the Sobolev space $H^1(\mathbb{R}^d)$ and deformations $\tau \in C^1(\mathbb{R}^d;\mathbb{R}^d)$ with $\|\nabla\tau\|_{L^\infty} \le 1/2$, hence regular.

Proposition 4.3. *There exists a constant $C>0$ such that*

$$\|F_\tau f - f\|_2 \le C\|\tau\|_{L^\infty}\|\nabla f\|_{X_r^{\infty,2}}, \quad r = \|\tau\|_{L^\infty}, \tag{4.3}$$

for every $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$ and every function $f \in X_r^{\infty,2}$ such that $\|\nabla f\|_{X_r^{\infty,2}} < \infty$.

Observe that the condition $\|\nabla f\|_{X_r^{\infty,2}} < \infty$ implies that $\nabla f \in L^\infty_{\mathrm{loc}}(\mathbb{R}^d)$, and therefore f is locally Lipschitz continuous after possibly being redefined on a set of measure zero (cf. [9, Theorem 4, page 294]), in particular f is continuous. In the following we will always identify f with its continuous version. Also, we set

$$\|\nabla f\|_{X_r^{\infty,2}} := \||\nabla f|\|_{X_r^{\infty,2}}. \tag{4.4}$$

Proof of Proposition 4.3. For $x \in \mathbb{R}^d$, $r>0$ let $B(x,r)$ be the open ball in $\mathbb{R}^d$ of radius r and center x. By the Poincaré inequality for a ball[3] (cf. [9, Theorem 2, page 291]) we see that there exists a constant $C>0$ such that, for every $r>0$ and $x \in \mathbb{R}^d$,

$$|f(x-y) - f(x)| \le Cr\|\nabla f\|_{L^\infty(B(x,r))}, \quad |y| \le r. \tag{4.5}$$

Setting $y = \tau(x)$, $r = \|\tau\|_{L^\infty}$ and taking the L^2 norm lead to the desired conclusion. $\square$

[3] That is $\|f - \overline{f}_{x,r}\|_{L^\infty(B(x,r))} \le Cr\|\nabla f\|_{L^\infty(B(x,r))}$ where $\overline{f}_{x,r}$ is the average of f over $B(x,r)$. Since under our assumption f is continuous in $\mathbb{R}^d$, we can replace the L^∞ norm in the left-hand side by the supremum of $|f|$, and then one obtains (4.5) from the triangle inequality (by adding and subtracting $\overline{f}_{x,r}$).

5. Multiresolution approximation spaces

Fix $\phi \in L^2(\mathbb{R}^d)$ and recall [19] that the associated approximation space U_s at scale $s > 0$ is defined as follows:

$$U_s := \overline{\operatorname{span}\{\phi_{s,n}\}_{n\in\mathbb{Z}^d}}, \qquad \phi_{s,n}(x) := s^{-d/2} T_{ns} D_{1/s}\phi(x) = s^{-d/2}\phi\left(\frac{x-ns}{s}\right).$$

In the rest of the paper we are going to deal with the following assumptions on ϕ.

Assumption A. There exist constants $A, B > 0$ such that

$$A \leq \sum_{k\in\mathbb{Z}^d} |\hat{\phi}(\omega - 2\pi k)|^2 \leq B \quad \text{for a.e. } \omega \in \mathbb{R}^d. \tag{5.1}$$

This is *equivalent* to assuming that $\{\phi_{s,n}\}$ is a Riesz basis for U_s (cf. [19, Theorem 3.4] in the case where $d = 1$, while the result for $d > 1$ follows by direct extension of the one-dimensional one).

We further assume one of the following regularity/decay conditions on ϕ.

Assumption B. *At least one* of the following conditions holds.

(i) ϕ belongs to the Wiener space:

$$\phi \in X^{\infty,1}, \tag{5.2}$$

in particular ϕ is locally bounded and has a L^1 decay.

(ii) There exist $\alpha > 1/2$ and $B' > 0$ such that

$$\sum_{k\in\mathbb{Z}^d} |(v^\alpha \hat{\phi})(\omega - 2\pi k)|^2 \leq B' \quad \text{for a.e. } \omega \in [0, 2\pi]^d, \tag{5.3}$$

where we introduced the weight function $v(\omega) = \langle\omega_1\rangle \cdots \langle\omega_d\rangle$, $\omega \in \mathbb{R}^d$.

Assumption C. At least one of the conditions (5.2) and (5.3) of Assumption B is satisfied for all $\partial_j\phi$, $j = 1, \ldots, d$, in place of ϕ.

Example 5.1. *This is a convenient stage where to present some examples of functions satisfying the assumptions. Generally speaking,* (5.2) *is satisfied by any function $\phi \in L^\infty(\mathbb{R}^d)$ with compact support, while the same condition on the Fourier side (i.e., $\hat{\phi} \in L^\infty(\mathbb{R}^d)$ with compact support) guarantees that 5.3 holds. To be more concrete, let us provide some standard examples in dimension $d = 1$ — Assumption A will be satisfied in all cases (cf.* [19, Section 3.1.3, pages 69,70]*).*

- *The choice $\phi = \mathbb{1}_{[0,1]}$, leading to piecewise constant approximations (block sampling), is easily seen to satisfy* (5.2) *but not* (5.3) *for any $\alpha > 1/2$, nor Assumption C.*

- *The normalized sinc function* $\phi(x) = \dfrac{\sin(\pi x)}{\pi x}$*, corresponding to Shannon approximations (i.e., band-limited functions), satisfies* (5.3) *for every* $\alpha > 0$*, as well as Assumption C, but not* (5.2)*.*
- *The B-spline* ϕ *of degree* n*, obtained by* $n+1$ *convolutions of* $\mathbb{1}_{[0,1]}$ *with itself and centering at* 0 *or* 1/2*, can be characterized by its Fourier transform:*

$$\hat{\phi}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{n+1} e^{-i\varepsilon\omega/2}, \quad \varepsilon = \begin{cases} 1 & (n \text{ is even}) \\ 0 & (n \text{ is odd}) \end{cases}.$$

We see that if $n \geq 1$ *then both* (5.2) *and* (5.3) *are satisfied (for* $\alpha < n + 1/2$*), as well as Assumption C (the case* $n = 0$ *is covered by the previous case of* $\phi = \mathbb{1}_{[0,1]}$*).*

In Assumption B we introduced the weight function v. Let us now define a companion Sobolev space, for $\alpha \in \mathbb{R}$, $\alpha \geq 0$:

$$H^{\alpha}_{\otimes}(\mathbb{R}^d) \coloneqq \{f \in L^2(\mathbb{R}^d) : \ \|f\|_{H^{\alpha}_{\otimes}} \coloneqq \|v^{\alpha}\hat{f}\|_{L^2} < \infty\}.$$

Roughly speaking, $H^{\alpha}_{\otimes}(\mathbb{R}^d)$ consists of functions in $L^2(\mathbb{R}^d)$ which have at least α (possibly fractional) derivatives in the directions of the axes in $L^2(\mathbb{R}^d)$. It is easy to realize that this space contains functions in the usual Sobolev space $H^{d\alpha}(\mathbb{R}^d)$ as well as tensor products $\phi_1 \otimes \ldots \otimes \phi_d$, with $\phi_j \in H^{\alpha}(\mathbb{R})$, $j = 1, \ldots, d$.

Proposition 5.2. *If* $\alpha > 1/2$ *we have the embedding* $H^{\alpha}_{\otimes}(\mathbb{R}^d) \hookrightarrow L^{\infty}(\mathbb{R}^d) \cap C(\mathbb{R}^d)$*, as well as*

$$H^{\alpha}_{\otimes}(\mathbb{R}^d) \hookrightarrow X^{\infty,2}.$$

Proof. The embedding in L^{∞} follows at once from the chain of inequalities

$$\|f\|_{L^{\infty}} \lesssim \|\hat{f}\|_{L^1} \leq \|v^{-\alpha}\|_{L^2}\|\hat{f}v^{\alpha}\|_{L^2}$$

and the fact that $v^{-\alpha} \in L^2(\mathbb{R}^d)$ if $\alpha > 1/2$. The embedding in $C(\mathbb{R}^d)$ is then clear because the space of Schwartz functions is easily seen to be dense in $H^{\alpha}_{\otimes}(\mathbb{R}^d)$.

Concerning the embedding in $X^{\infty,2}$, let $g \in C^{\infty}_c(\mathbb{R}^d)$, with $g = 1$ on B_1. Then

$$\|f\|_{X^{\infty,2}} \leq \|\|T_{-x}f \cdot g\|_{L^{\infty}}\|_{L^2_x} \lesssim \|\|f \cdot T_x g\|_{H^{\alpha}_{\otimes}}\|_{L^2_x} \lesssim \|f\|_{H^{\alpha}_{\otimes}},$$

where the last inequality is proved in [13, Propositon 11.3.1(c)]. ☐

We now establish a crucial reverse Hölder-type inequality for functions in U_s.

Theorem 5.3. *Let* $\phi \in L^2(\mathbb{R}^d)$ *be such that Assumption A is satisfied.*

(i) If Assumption B holds then there exists $C > 0$ *such that, for every* $r, s > 0$*,*

$$\|f\|_{X^{\infty,2}_r} \leq C(1 + r/s)^{d/2}\|f\|_{L^2}, \quad f \in U_s. \tag{5.4}$$

(ii) If Assumption C holds then there exists $C > 0$ such that, for every $r, s > 0$,

$$\|\nabla f\|_{X_r^{\infty,2}} \leq Cs^{-1}(1 + r/s)^{d/2}\|f\|_{L^2}, \quad f \in U_s. \tag{5.5}$$

Remark 5.4. *Let P_{U_s} be the orthogonal projection operator on U_s. Since $\|P_{U_s}\|_{L^2\to L^2} = 1$, (5.4) is equivalent to*

$$\|P_{U_s} f\|_{X_r^{\infty,2}} \leq C(1 + r/s)^{d/2}\|f\|_{L^2}, \quad f \in L^2(\mathbb{R}^d).$$

Proof of Theorem 5.3. Let us commence with the proof of (5.4). Let $\{\tilde{\phi}_{s,n}\}_{n\in\mathbb{Z}^d}$ be the dual basis to $\{\phi_{s,n}\}_{n\in\mathbb{Z}^d}$. If $f \in U_s$ then

$$f = \sum_{n\in\mathbb{Z}^d} a_n\phi_{s,n}, \quad a_n := \langle f, \tilde{\phi}_{s,n}\rangle,$$

and by Lemma 3.2 we have

$$\begin{aligned}
\|f\|_{X_r^{\infty,2}} &= r^{d/2}\|D_r f\|_{X^{\infty,2}} \\
&= \left\| \sum_{n\in\mathbb{Z}^d} a_n \phi_{s/r,n} \right\|_{X^{\infty,2}} \\
&= \left(\frac{r}{s}\right)^{d/2} \left\| \sum_{n\in\mathbb{Z}^d} a_n \phi\left(\frac{r}{s}\cdot - n\right) \right\|_{X^{\infty,2}} \\
&= \left(\frac{r}{s}\right)^{d/2} \left\| \left\| \sum_{n\in\mathbb{Z}^d} a_n \phi\left(\frac{r}{s}(x+y) - n\right) \mathbb{1}_{B_1}(y) \right\|_{L_y^\infty} \right\|_{L_x^2} \\
&= \left\| \left\| \sum_{n\in\mathbb{Z}^d} a_n \phi(x + y - n)\mathbb{1}_{B_{r/s}}(y) \right\|_{L_y^\infty} \right\|_{L_x^2} \\
&= \left\| \sum_{n\in\mathbb{Z}^d} a_n T_n\phi \right\|_{X_{r/s}^{\infty,2}} \\
&\lesssim \left(1 + \frac{r}{s}\right)^{d/2} \left\| \sum_{n\in\mathbb{Z}^d} a_n T_n\phi \right\|_{X^{\infty,2}}, \qquad (5.6)
\end{aligned}$$

where in the last step we used Lemma 3.2 and Proposition 3.5.

Assume now (5.2), namely $\phi \in X^{\infty,1}$. Then the conclusion follows from (5.6) using the equivalent discrete-type norm in (3.2) (with $Q = [0, 1]^d$):

$$\|f\|_{X_r^{\infty,2}} \lesssim \left(1 + \frac{r}{s}\right)^{d/2} \left\| \sum_{n\in\mathbb{Z}^d} a_n T_n\phi \right\|_{X^{\infty,2}}$$

$$\lesssim \left(1+\frac{r}{s}\right)^{d/2} \left\| \sum_{n\in\mathbb{Z}^d} |a_n| \|\phi(k+y-n)\mathbb{1}_Q\|_{L^\infty_y} \right\|_{\ell^2_k}$$
$$\lesssim \left(1+\frac{r}{s}\right)^{d/2} \left(\sum_{k\in\mathbb{Z}^d} \|\phi(k+y)\mathbb{1}_Q\|_{L^\infty_y} \right) \left(\sum_{n\in\mathbb{Z}^d} |a_n|^2 \right)^{1/2}$$
$$\lesssim \left(1+\frac{r}{s}\right)^{d/2} \|\phi\|_{X^{\infty,1}} \|f\|_{L^2},$$

where we used that $\ell^1 * \ell^2 \hookrightarrow \ell^2$ and $\|a_n\|_{\ell^2} \lesssim \|f\|_{L^2}$.

Let us assume (5.3) instead. By (5.6), it is enough to show that

$$\left\| \sum_{n\in\mathbb{Z}^d} a_n T_n \phi \right\|_{X^{\infty,2}} \lesssim \|f\|_{L^2}.$$

Using the embedding in Proposition 5.2 we obtain

$$\begin{aligned}
\left\| \sum_{n\in\mathbb{Z}^d} a_n T_n \phi \right\|_{X^{\infty,2}} &\lesssim \left\| \sum_{n\in\mathbb{Z}^d} a_n T_n \phi \right\|_{H^\alpha_\otimes} \\
&= \left\| \sum_{n\in\mathbb{Z}^d} a_n e^{-in\omega} \hat{\phi}(\omega) v^\alpha(\omega) \right\|_{L^2} \\
&= \left(\int_{[0,2\pi]^d} \sum_{k\in\mathbb{Z}^d} |F(\omega)(v^\alpha \hat{\phi})(\omega - 2\pi k)|^2 d\omega \right)^{1/2} \\
&\le \|F\|_{L^2([0,2\pi]^d)} \left(\operatorname*{ess\,sup}_{\omega\in[0,2\pi]} \sum_{k\in\mathbb{Z}^d} |(v^\alpha \hat{\phi})(\omega - 2\pi k)|^2 \right)^{1/2} \\
&\lesssim \|f\|_{L^2},
\end{aligned}$$

where we set $F(\omega) \coloneqq \sum_{n\in\mathbb{Z}^d} a_n e^{-in\omega}$ (which is a 2π-periodic, square integrable on $[0,2\pi]$, function), and then used (5.3) and

$$\|F\|^2_{L^2([0,2\pi]^d)} \asymp \sum_{n\in\mathbb{Z}^d} |a_n|^2 \asymp \|f\|^2_{L^2}.$$

The proof of (5.5) goes along the same lines after differentiation in the representation $f = \sum_{n\in\mathbb{Z}^d} a_n \phi_{s,n}$; the details are left to the interested reader. □

Remark 5.5. *(i) It is easy to realize that if ϕ satisfies* (5.3) *then $\phi \in H^\alpha_\otimes(\mathbb{R}^d)$ (it is enough to integrate both sides of* (5.3) *on* $[0,2\pi]$*); as a result, if $\alpha > 1/2$ then ϕ is continuous by Proposition 5.2.*

(ii) *If $\phi \in L^2(\mathbb{R}^d)$ satisfies Assumption C then ϕ has first order partial derivatives locally in L^∞, hence ϕ is locally Lipschitz, therefore continuous.*

(iii) *If $\phi \in L^2(\mathbb{R}^d)$ satisfies the assumption A and B and is continuous, then $U_s \hookrightarrow C(\mathbb{R}^d)$ since the truncated sums $\sum_{|n|\le N} a_n T_{sn}\phi$ are continuous and (5.4) shows that convergence in L^2 implies convergence in $X_r^{\infty,2} \hookrightarrow L^\infty(\mathbb{R}^d)$ for functions in U_s.*

(iv) *If $s \ll r$ then the occurrence of the factor r/s in (5.4) can be heuristically explained by the presence of highly oscillating functions in U_s, which are not stable under deformations of "size" r.*

We are ready to provide deformation sensitivity bounds for functions in U_s.

Theorem 5.6. *Let $\phi \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ satisfy Assumptions A and B. There exists a constant $C > 0$ such that, for every $\tau \in L^\infty(\mathbb{R}^d; \mathbb{R}^d)$ and $s > 0$*

$$\|F_\tau f\|_{L^2} \le C(1 + \|\tau\|_{L^\infty}/s)^{d/2}\|f\|_{L^2}, \quad f \in U_s. \tag{5.7}$$

Proof. The desired estimate follows by a straightforward concatenation of Proposition 4.1, since the assumptions on ϕ imply that $U_s \hookrightarrow C(\mathbb{R}^d)$ (cf. Remark 5.5), and Theorem 5.3 with $r = \|\tau\|_{L^\infty}$. ☐

Theorem 5.7. *Let $\phi \in L^2(\mathbb{R}^d)$ be such that Assumptions A, B and C are satisfied. There exists a constant $C > 0$ such that*

$$\|F_\tau f - f\|_{L^2} \le \begin{cases} C(\|\tau\|_{L^\infty}/s)\|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \le 1) \\ C(\|\tau\|_{L^\infty}/s)^{d/2}\|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \ge 1) \end{cases}, \tag{5.8}$$

for every $\tau \in L^\infty(\mathbb{R}^d; \mathbb{R}^d)$, $s > 0$ and $f \in U_s$.

Proof. Let us consider first the case $\|\tau\|_{L^\infty}/s \le 1$. Combining Proposition 4.3 with Theorem 5.3 with $r = \|\tau\|_{L^\infty}$ we infer, for $f \in U_s$,

$$\begin{aligned} \|F_\tau f - f\|_{L^2} &\lesssim \|\tau\|_{L^\infty}\|\nabla f\|_{X_r^{\infty,2}} \\ &\lesssim (\|\tau\|_{L^\infty}/s)(1 + \|\tau\|_{L^\infty}/s)^{d/2}\|f\|_{L^2} \\ &\lesssim (\|\tau\|_{L^\infty}/s)\|f\|_{L^2}, \end{aligned}$$

that is the claim.

The case $\|\tau\|_{L^\infty}/s \ge 1$ can be approached via the triangle inequality, that is $\|F_\tau f - f\|_{L^2} \le \|F_\tau f\|_{L^2} + \|f\|_{L^2}$, and Theorem 5.6. ☐

Remark 5.8. *More generally, the same result of Theorem 5.7 holds if f is replaced on the left-hand side by $P_{U_s} f$ for $f \in L^2(\mathbb{R}^d)$, cf. Remark 5.4. Moreover, taking into account the examples in Example 5.1 we see that Theorem 5.7 applies when U_s are approximation spaces of polynomial splines of degree $n \ge 1$, as well of band-limited functions — which can be regarded as splines of infinite order.*

We conclude this section by extending the above stability bounds to signal classes with minimal regularity. In addition to the assumptions of Theorem 5.7, we suppose that $V_j := U_{2^j}$, $j \in \mathbb{Z}$, define a multiresolution approximation of $L^2(\mathbb{R}^d)$, so that $V_{j+1} \subset V_j$. Let W_{j+1} be the orthogonal complement of V_{j+1} in V_j and P_{W_j} be the corresponding orthogonal projection; for $s \in \mathbb{R}$, the corresponding homogeneous Besov norm [19, Section 9.2.3] is given by

$$\|f\|_{\dot{B}^s_{2,1}} = \sum_{j\in\mathbb{Z}} 2^{-js}\|P_{W_j} f\|_{L^2}. \tag{5.9}$$

Theorem 5.9. *Under the same assumptions of Theorem 5.7, suppose in addition that $V_j := U_{2^j}$, $j \in \mathbb{Z}$, define a multiresolution approximation of $L^2(\mathbb{R}^d)$.*

There exists $C > 0$ such that for every $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$ and $f \in L^2(\mathbb{R}^d)$ with $\|f\|_{\dot{B}^{d/2}_{2,1}} < \infty$,

$$\|F_\tau f - f\|_{L^2} \le C(\|\tau\|_{L^\infty}\|f\|_{\dot{B}^1_{2,1}} + \|\tau\|^{d/2}_{L^\infty}\|f\|_{\dot{B}^{d/2}_{2,1}}), \qquad d \ge 2, \tag{5.10}$$

and

$$\|F_\tau f - f\|_{L^2} \le C\|\tau\|^{1/2}_{L^\infty}\|f\|_{\dot{B}^{1/2}_{2,1}}, \qquad d = 1. \tag{5.11}$$

Proof. We consider the decomposition

$$f = \sum_{\|\tau\|_{L^\infty}\le 2^j} P_{W_j} f + \sum_{2^j<\|\tau\|_{L^\infty}} P_{W_j} f$$

and apply (5.8) to each term, hence we obtain

$$\|F_\tau f - f\|_{L^2} \lesssim \|\tau\|_{L^\infty} \sum_{\|\tau\|_{L^\infty}\le 2^j} 2^{-j}\|P_{W_j} f\|_{L^2} + \|\tau\|^{d/2}_{L^\infty} \sum_{2^j<\|\tau\|_{L^\infty}} 2^{-jd/2}\|P_{W_j} f\|_{L^2}, \tag{5.12}$$

which implies the desired result if $d \ge 2$.

For $d = 1$ it is sufficient to continue the estimate in (5.12) using

$$\sum_{\|\tau\|_{L^\infty}\le 2^j} 2^{-j}\|P_{W_j} f\|_{L^2} \le \sum_{\|\tau\|_{L^\infty}\le 2^j} 2^{-j/2}2^{-j/2}\|P_{W_j} f\|_{L^2} \le \|\tau\|^{-1/2}_{L^\infty}\|f\|_{\dot{B}^{1/2}_{2,1}}.$$

∎

Remark 5.10. *From the very definition (5.9) of the Besov norm, it follows that if $d \ge 2$ and $f \in L^2(\mathbb{R}^d)$ with $\|f\|_{\dot{B}^{d/2}_{2,1}} < \infty$ then $\|f\|_{\dot{B}^1_{2,1}} < \infty$.*

Also, note that even in dimension 1 we have $\|F_\tau f - f\|_{L^2} = O(\|\tau\|_{L^\infty})$ as $\|\tau\|_{L^\infty} \to 0$ for every fixed $f \in U_s$ and every $s > 0$, as a consequence of Theorem 5.7. However this asymptotic estimate is not uniform in the ball $\|f\|_{L^2} + \|f\|_{\dot{B}^{1/2}_{2,1}} \le 1$, and the factor $\|\tau\|^{1/2}_{L^\infty}$ in (5.11) is instead optimal when looking for uniform estimates; see the examples in Section 7 below. In dimension $d \ge 2$ it follows easily from (5.10) that

$\|F_\tau f - f\|_{L^2} = O(\|\tau\|_{L^\infty})$ *as* $\|\tau\|_{L^\infty} \to 0$ *uniformly for* f *in the ball* $\|f\|_{L^2} + \|f\|_{\dot{B}^{d/2}_{2,1}} \leq 1$.

6. Frequency-modulated deformations

In this section we extend some results proved so far to the class of time-frequency deformation mappings $F_{\tau,\omega}$ associated with distortion functions $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $\omega \in L^\infty(\mathbb{R}^d;\mathbb{R})$ by setting

$$F_{\tau,\omega} f(x) \coloneqq e^{i\omega(x)} f(x - \tau(x)),$$

where $f\colon \mathbb{R}^d \to \mathbb{C}$. In case of trivially null distortions we write $F_{0,\omega}$ and $F_{\tau,0}$ with obvious meaning.

While most of the results above can be stated and proved with minor updates for general deformations $F_{\tau,\omega}$, we prefer to offer here a different perspective that allows one to reduce to the results for F_τ in a straightforward way. Indeed, note that $F_{\tau,\omega} = F_{0,\omega} F_{\tau,0}$ and $F_{\tau,0}$ coincides with the deformation F_τ considered in the previous sections. Moreover, for every $f \in L^2(\mathbb{R}^d)$ we have that $\|F_{\tau,\omega} f\|_{L^2} = \|F_{\tau,0} f\|_{L^2}$ for arbitrary measurable ω, and

$$\|F_{\tau,\omega} f - f\|_{L^2} \leq \|F_{\tau,\omega} f - F_{\tau,0} f\|_{L^2} + \|F_{\tau,0} f - f\|_{L^2}.$$

The second addend is already covered, while for the first one we have

$$\|F_{\tau,\omega} f - F_{\tau,0} f\|_{L^2} \leq \|e^{i\omega} - 1\|_{L^\infty} \|F_{\tau,0} f\|_{L^2} \leq \|\omega\|_{L^\infty} \|F_{\tau,0} f\|_{L^2}.$$

As a result, the bounds in Propositions 4.1 and 4.3 generalize as follow.

Theorem 6.1. *We have*

$$\|F_{\tau,\omega} f\|_{L^2} \leq \|f\|_{X^{\infty,2}_r}, \quad r = \|\tau\|_{L^\infty}, \tag{6.1}$$

for every $f \in X^{\infty,2}_r \cap C(\mathbb{R}^d)$ *and* $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $\omega \in L^\infty(\mathbb{R}^d;\mathbb{R})$.

Moreover, there exists $C > 0$ *such that*

$$\|F_{\tau,\omega} f - f\|_{L^2} \leq C(\|\tau\|_{L^\infty} \|\nabla f\|_{X^{\infty,2}_r} + \|\omega\|_{L^\infty} \|f\|_{X^{\infty,2}_r}), \quad r = \|\tau\|_{L^\infty}, \tag{6.2}$$

for every $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $\omega \in L^\infty(\mathbb{R}^d;\mathbb{R})$ *and* $f \in X^{\infty,2}_r$ *with* $\|\nabla f\|_{X^{\infty,2}_r} < \infty$.

With the same arguments of the proofs of Theorems 5.7, using the bounds in Theorem 6.1 whenever appropriate, we obtain the following generalization.

Theorem 6.2. *Let* $\phi \in L^2(\mathbb{R}^d)$ *be such that Assumptions A, B and C in Section 5 hold. There exists a constant* $C > 0$ *such that*

$$\|F_{\tau,\omega} f - f\|_{L^2} \leq \begin{cases} C(\|\tau\|_{L^\infty}/s + \|\omega\|_{L^\infty})\|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \leq 1) \\ C(\|\tau\|_{L^\infty}/s)^{d/2} \|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \geq 1) \end{cases}, \tag{6.3}$$

for every $s > 0$, $f \in U_s$ *and* $\tau \in L^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $\omega \in L^\infty(\mathbb{R}^d;\mathbb{R})$.

We remark that for band-limited functions $U_s = \mathrm{PW}_R$ with $s = \pi/R$ and in the relevant case where $R\|\tau\|_{L^\infty} \le 1$ we recover the same bounds proved in [25] without extra regularity conditions on τ or ω. Similarly, one could generalize the estimates in Besov spaces of the previous section.

7. Sharpness of the estimates

We now study the problem of the sharpness of some estimates proved so far, focusing in particular on the case of band-limited functions.

For $R > 0$ consider the space of band-limited functions

$$\mathrm{PW}_R \coloneqq \{f \in L^2(\mathbb{R}^d) : \operatorname{supp}\hat{f} \subset [-R, R]^d\}.$$

We already commented in Example 5.1 that such a space of low-frequency functions can be equivalently designed as a multiresolution space; precisely, we have $\mathrm{PW}_R = U_s$ with $s = \pi/R$ after choosing the normalized low-pass sinc filter $\phi = \phi_0 \otimes \cdots \otimes \phi_0$ (d times), with $\phi_0(t) = \pi^{-1/2} \sin t/t$, $t \in \mathbb{R}$, which satisfies Assumptions A, B, C.

Theorems 5.6 and 5.7 above thus cover the case of band-limited approximations. Precisely, (5.7) now reads

$$\|F_\tau f\|_{L^2} \le C(1 + R\|\tau\|_{L^\infty})^{d/2}\|f\|_{L^2}, \quad f \in \mathrm{PW}_R, \tag{7.1}$$

while (5.8) becomes

$$\|F_\tau f - f\|_{L^2} \le \begin{cases} CR\|\tau\|_{L^\infty}\|f\|_{L^2} & (R\|\tau\|_{L^\infty} \le 1) \\ C(R\|\tau\|_{L^\infty})^{d/2}\|f\|_{L^2} & (R\|\tau\|_{L^\infty} \ge 1) \end{cases}, \quad f \in \mathrm{PW}_R. \tag{7.2}$$

We claim that the exponent $d/2$ appearing in the previous estimates is optimal. For what concerns (7.1), it suffices to consider $f_R \in \mathrm{PW}_R$ given by $f_R = R^{d/2} D_R \phi$, so that $\|f_R\|_{L^2} = 1$ and $\widehat{f_R} = (\pi/R)^{d/2}\mathbb{1}_{[-R,R]^d}$. Now, for $K > 0$ set

$$\tau(x) = \begin{cases} x & (|x| \le K) \\ 0 & (|x| > K) \end{cases},$$

so that $\|\tau\|_{L^\infty} = K$. Then, for $|x| \le K$ we have

$$F_\tau f_R(x) = f_R(0) = (R/\pi)^{d/2},$$

and thus

$$\|F_\tau f_R\|_{L^2} \gtrsim (R\|\tau\|_{L^\infty})^{d/2}.$$

By the triangle inequality we also deduce

$$\|F_\tau f_R - f_R\|_{L^2} \gtrsim (R\|\tau\|_{L^\infty})^{d/2}, \quad R\|\tau\|_{L^\infty} \gg 1, \tag{7.3}$$

which shows the sharpness of the exponent $d/2$ in (7.2) as well.

Concerning the sharpness of the estimate (7.2) in the regime $R\|\tau\|_{L^\infty} \ll 1$ we see that if $f = f_R$ as above and $\tau(x) = (c, 0, \dots, 0) \in \mathbb{R}^d$ (constant), for $|c|R$ small enough we have

$$\|F_\tau f_R - f_R\|_{L^2}^2 = \left(\frac{1}{2R}\right)^d \int_{[-R,R]^d} |e^{-ic\omega_1} - 1|^2 \, d\omega \gtrsim R^{-d} \int_{[-R,R]^d} (c\omega_1)^2 \, d\omega \gtrsim (cR)^2.$$

8. Random deformations

We now model the deformation $\tau(x)$ as a measurable random field, i.e. $\tau(x) = \tau(x, \omega)$ depends on an additional variable[4] $\omega \in \mathcal{U}$, where the sample space $\mathcal{U}$ is equipped with a probability measure $\mathbb{P}$, and the function $\tau(x, \omega)$ is jointly measurable (see for instance [14, Chapter 3] for further details).

It is easy to realize that the results of the previous sections hold for almost every realization of $\tau(x)$ if, e.g., $\|\tau\|_{L^\infty} < \infty$, which must be intended hereinafter as the essential supremum jointly in x, ω. However, it turns out that some results hold, in fact, in a *maximal* sense[5]. Precisely, an inspection of the proof of the formula (4.1) shows that we have

$$\| \|F_\tau f\|_{L^\infty(\mathcal{U})} \|_{L^2} \leq \|f\|_{X_r^{\infty,2}}, \quad r = \|\tau\|_{L^\infty}, \tag{8.1}$$

and similarly (4.3) becomes

$$\| \|F_\tau f - f\|_{L^\infty(\mathcal{U})} \|_{L^2} \leq C \|\tau\|_{L^\infty} \|\nabla f\|_{X_r^{\infty,2}}, \quad r = \|\tau\|_{L^\infty}. \tag{8.2}$$

As a consequence, under the assumptions of Theorem 5.7 we have, for $f \in U_s$,

$$\| \|F_\tau f - f\|_{L^\infty(\mathcal{U})} \|_{L^2} \leq \begin{cases} C(\|\tau\|_{L^\infty}/s) \|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \leq 1) \\ C(\|\tau\|_{L^\infty}/s)^{d/2} \|f\|_{L^2} & (\|\tau\|_{L^\infty}/s \geq 1) \end{cases}, \tag{8.3}$$

while arguing as in the proof of Theorem 5.9 we get

$$\| \|F_\tau f - f\|_{L^\infty(\mathcal{U})} \|_{L^2} \leq C(\|\tau\|_{L^\infty} \|f\|_{\dot{B}^1_{2,1}} + \|\tau\|_{L^\infty}^{d/2} \|f\|_{\dot{B}^{d/2}_{2,1}}), \qquad d \geq 2, \tag{8.4}$$

and

$$\| \|F_\tau f - f\|_{L^\infty(\mathcal{U})} \|_{L^2} \leq C \|\tau\|_{L^\infty}^{1/2} \|f\|_{\dot{B}^{1/2}_{2,1}}, \quad d = 1. \tag{8.5}$$

We are now ready to state our result concerning the stability in mean under random deformations.

---

[4]In this section we do not consider frequency-modulated deformations, nor we use the notation ω for the frequency, hence there is not risk of confusion with the notation of previous sections.

[5]Actually, we could equivalently reformulate the main estimates of the previous sections as results for the maximal operators $\sup_{|y| \leq r} |f(x - y)|$ and $\sup_{|y| \leq r} |f(x - y) - f(x)|$. However, the above presentation in terms of their linearized versions F_τ and $F_\tau - I$ seems closer to the spirit of the intended applications.

Theorem 8.1. *Under the assumption A, B and C in Section 5, there exists a constant $C > 0$ such that, for every $s > 0$ and $f \in U_s$,*

$$\mathbb{E}\|F_\tau f - f\|_{L^2}^2 \le C\mathbb{E}[(|\tau|/s)^2 + (|\tau|/s)^d]\|f\|_{L^2}^2, \quad d \ge 2, \tag{8.6}$$

and

$$\mathbb{E}\|F_\tau f - f\|_{L^2}^2 \le C\mathbb{E}[\min\{(|\tau|/s)^2, (|\tau|/s)^d\}]\|f\|_{L^2}^2, \quad d = 1, \tag{8.7}$$

for every measurable random function τ such that the random variables $|\tau(x)|$, $x \in \mathbb{R}^d$, are identically distributed and the above moments are finite.

Moreover, if the spaces U_{2^j}, $j \in \mathbb{Z}$, define a multiresolution approximation of $L^2(\mathbb{R}^d)$, for the same deformations $\tau(x)$ and every $f \in L^2(\mathbb{R}^d)$ with $\|f\|_{\dot{B}^{d/2}_{2,1}} < \infty$ we have

$$\mathbb{E}\|F_\tau f - f\|_{L^2}^2 \le C(\mathbb{E}[|\tau|^2]\|f\|^2_{\dot{B}^1_{2,1}} + \mathbb{E}[|\tau|^d]\|f\|^2_{\dot{B}^{d/2}_{2,1}}) \qquad d \ge 2 \tag{8.8}$$

and

$$\mathbb{E}\|F_\tau f - f\|_{L^2}^2 \le C\mathbb{E}[|\tau|]\|f\|^2_{\dot{B}^{1/2}_{2,1}} \quad d = 1. \tag{8.9}$$

For the sake of brevity, we wrote $\mathbb{E}[|\tau|^2]$ in place of $\mathbb{E}[|\tau(x)|^2]$, and similarly for the other moments, since the variables $|\tau(x)|$, $x \in \mathbb{R}^d$, are assumed to be identically distributed. However, observe that the field $\tau(x)$ is not assumed to be bounded.

Proof of Theorem 8.1. Let us prove (8.6) and (8.7) first. Let us set

$$\tau_j(x) := \begin{cases} \tau(x) & (2^{j-1} < |\tau(x)| \le 2^j) \\ 0 & (\text{otherwise}) \end{cases}, \quad j \in \mathbb{Z}.$$

Then we can write

$$\begin{aligned} \|F_\tau f - f\|_{L^2}^2 &= \sum_{j\in\mathbb{Z}} \int_{\mathbb{R}^d} |F_{\tau_j} f(x) - f(x)|^2 \mathbb{1}_{\{2^{j-1}<|\tau|\le 2^j\}}(x)\, dx \\ &\le \sum_{j\in\mathbb{Z}} \int_{\mathbb{R}^d} \mathbb{1}_{\{2^{j-1}<|\tau|\le 2^j\}}(x) \|F_{\tau_j} f(x) - f(x)\|^2_{L^\infty(\mathcal{U})}\, dx. \end{aligned}$$

Taking the expectation and setting $p_j = \mathbb{P}(\{2^{j-1} < |\tau(x)| \le 2^j\})$ (note that p_j is independent of x) we get

$$\mathbb{E}\|F_\tau f - f\|_{L^2}^2 \le \sum_{j\in\mathbb{Z}} p_j \|\|F_{\tau_j} f - f\|_{L^\infty(\mathcal{U})}\|^2_{L^2}.$$

We use the estimate (8.3) to bound each term and we obtain

$$\mathbb{E}\|F_\tau f - f\|_{L^2}^2 \lesssim \left(\sum_{2^j \le s} p_j (2^j/s)^2 + \sum_{s<2^j} p_j (2^j/s)^d \right) \|f\|_{L^2}^2, \quad f \in U_s.$$

We now observe that, for every $x \in \mathbb{R}^d$,

$$\sum_{2^j \le s} p_j (2^j/s)^2 = \sum_{2^j \le s} \mathbb{E}[(2^j/s)^2 \mathbb{1}_{\{2^{j-1} < |\tau(x)| \le 2^j\}}] \lesssim \mathbb{E}[(|\tau(x)|/s)^2 \mathbb{1}_{\{|\tau(x)|/s \le 1\}}]$$

and similarly

$$\begin{aligned}\sum_{s < 2^j} p_j (2^j/s)^d &\lesssim \mathbb{E}[(|\tau(x)|/s)^d \mathbb{1}_{\{|\tau(x)|/s > 1/2\}}] \\ &\lesssim \mathbb{E}[(|\tau(x)|/s)^2 \mathbb{1}_{\{1/2 < |\tau(x)|/s \le 1\}} + (|\tau(x)|/s)^d \mathbb{1}_{\{|\tau(x)|/s > 1\}}].\end{aligned}$$

Hence we have proved the estimate

$$\mathbb{E}\|F_\tau f - f\|_{L^2}^2 \lesssim \mathbb{E}[(|\tau(x)|/s)^2 \mathbb{1}_{\{|\tau(x)|/s \le 1\}} + (|\tau(x)|/s)^d \mathbb{1}_{\{|\tau(x)|/s > 1\}}]\|f\|_{L^2}^2,$$

which gives (8.6) and (8.7).

Similar arguments lead to the proof of (8.8) and (8.9), now using (8.4) and (8.5). $\square$

Acknowledgements

The authors wish to express their gratitude to Giovanni S. Alberti, Enrico Bibbona and Matteo Santacesaria for fruitful conversations on the topics of the manuscript, as well as for valuable comments on preliminary drafts.

The present research has been partially supported by the MIUR grant Dipartimenti di Eccellenza 2018-2022, CUP: E11G18000350001, DISMA, Politecnico di Torino.

S. Ivan Trapasso was member of the Machine Learning Genoa (MaLGa) Center, Università di Genova, when this study was performed. This material is based upon work supported by the Air Force Office of Scientific Research under award number FA8655-20-1-7027.

F. Nicola is a fellow of the Accademia delle Scienze di Torino, and a member of the Società Italiana di Scienze e Tecnologie Quantistiche (SISTEQ).

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).

References

[1] Charalambos D. Aliprantis, and Kim Border. *Infinite dimensional Analysis. A Hitchhiker's Guide.* Springer, 2006.

[2] Radu Balan, Maneesh Singh and Dongmian Zou. Lipschitz properties for deep convolutional networks. *Contemporary Mathematics* **706** (2018), 129–151.

[3] Alberto Bietti and Julien Mairal. Invariance and stability of deep convolutional representations. In: *Advances in Neural Information Processing Systems (NIPS)*, 2017.

[4] Alberto Bietti and Julien Mairal. Group invariance, stability to deformations, and complexity of deep convolutional representations. *Journal of Machine Learning Research (JMLR)* **20**(25) (2019):1–49.

[5] Michael M. Bronstein, Joan Bruna, Taco Cohen, Petar Veličković. *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.* To appear — MIT Press, 2025. arXiv:2104.13478.

[6] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **35**(8) (2013), 1872–1886.

[7] Elena Cordero and Fabio Nicola. Sharpness of some properties of Wiener amalgam and modulation spaces. *Bull. Aust. Math. Soc.* **80** (2009), no. 1, 105–116.

[8] Piero D'Ancona and Fabio Nicola. Sharp L^p estimates for Schrödinger groups. *Rev. Mat. Iberoam.* **32** (2016), no. 3, 1019–1038.

[9] Lawrence C. Evans. Partial differential equations. Second edition. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 2010.

[10] Hans G. Feichtinger. Banach convolution algebras of Wiener type. In: *Functions, series, operators, Vol. I, II (Budapest, 1980)*, Colloq. Math. Soc. János Bolyai, 35, North-Holland, Amsterdam, 1983, 509–524.

[11] Hans G. Feichtinger. Banach spaces of distributions of Wiener's type and interpolation. In: *Functional analysis and approximation (Oberwolfach, 1980)*, Internat. Ser. Numer. Math., 60, Birkhäuser, Basel-Boston, Mass., 1981, 153–165.

[12] Loukas Grafakos. *Modern Fourier analysis.* Third edition. Graduate Texts in Mathematics. Springer, New York, 2014.

[13] Karlheinz Gröchenig. *Foundations of time-frequency analysis.* Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[14] Iosif I. Gihman and Anatolij V. Skorohod. *The theory of stochastic processes I* . Translated from the Russian by Samuel Kotz. Corrected reprint of the first English edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 210. Springer-Verlag, Berlin-New York, 1980.

[15] Philipp Grohs, Thomas Wiatowski and Helmut Bölcskei. Deep convolutional neural networks on cartoon functions. In: *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, 1163–1167.

[16] Christopher Heil. An introduction to weighted Wiener amalgams. In: *Wavelets and their applications.* Ed. by S. Thangavelu, M. Krishna, R. Radha. Allied Publishers, New Dehli, 2003, 183–216.

[17] Michael Koller, Johannes Großmann, Ullrich Monich and 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, 6872–6876.

[18] Stéphane Mallat. Group invariant scattering. *Comm. Pure Appl. Math.* **65** (2012), no. 10, 1331–1398.

[19] Stéphane Mallat. *A wavelet tour of signal processing. The sparse way.* Third edition. With contributions from Gabriel Peyré. Elsevier/Academic Press, Amsterdam, 2009.

[20] Fabio Nicola, S. Ivan Trapasso. Stability of the scattering transform for deformations with minimal regularity. *J. Math. Pures Appl. (9)* **180** (2023), 122–150.

[21] Aaditya Prakash, Nick Moran, Solomon Garber, Antonella DiLillo and James Storer. Deflecting adversarial attacks with pixel deflection. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8571–8580, doi: 10.1109/CVPR.2018.00894.

[22] Kevin Scaman and Aladin Virmaux. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS 2018).*

[23] Elias M. Stein. *Singular Integrals and Differentiability Properties of Functions*. Princeton Mathematical Series, No. 30 Princeton University Press, Princeton, N.J. 1970 xiv+290 pp.
[24] Terence Tao. Low regularity semi-linear wave equations. *Comm. Partial Differential Equations* **24** (1999), no. 3-4, 599–629.
[25] Thomas Wiatowski and Helmut Bölcskei. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Trans. Inform. Theory* **64** (2018), no. 3, 1845–1866.
[26] Thomas Wiatowski and Helmut Bölcskei. Deep convolutional neural networks based on semi-discrete frames. In: *2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China*, 2015, 1212–1216.
[27] Dongmian Zou, Radu Balan and Maneesh Singh. On Lipschitz bounds of general convolutional neural networks. *IEEE Trans. on Info. Theory* **66**(3) (2020), 1738–1759.

Dipartimento di Scienze Matematiche, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy.
Email address: `fabio.nicola@polito.it`

Dipartimento di Scienze Matematiche, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy.
Email address: `salvatore.trapasso@polito.it`