Deep Convolutional Neural Networks Based on Semi-Discrete Frames

Thomas Wiatowski and Helmut Bölcskei
Dept. IT & EE, ETH Zurich, Switzerland
Email: {withomas, boelcskei}@nari.ee.ethz.ch

Abstract—Deep convolutional neural networks have led to breakthrough results in practical feature extraction applications. The mathematical analysis of these networks was pioneered by Mallat [1]. Specifically, Mallat considered so-called scattering networks based on identical semi-discrete wavelet frames in each network layer, and proved translation-invariance as well as deformation stability of the resulting feature extractor. The purpose of this paper is to develop Mallat’s theory further by allowing for different and, most importantly, general semi-discrete frames (such as, e.g., Gabor frames, wavelets, curvelets, shearlets, ridgelets) in distinct network layers. This allows to extract wider classes of features than point singularities resolved by the wavelet transform. Our generalized feature extractor is capable of being translation-invariant, and we develop deformation stability results for a larger class of deformations than those considered by Mallat. For Mallat’s wavelet-based feature extractor, we get rid of a number of technical conditions. The mathematical engine behind our results is continuous frame theory, which allows us to completely detach the invariance and deformation stability proofs from the particular algebraic structure of the underlying frames.

I. INTRODUCTION

A central task in signal classification is feature extraction [2]. For example, we may want to detect whether an image contains a certain handwritten digit [3]. Moreover, this should be possible independently of the feature’s spatial (or temporal) location within the signal, which motivates the use of translation-invariant feature extractors. In addition, sticking to the example of handwritten digits, we want the feature extractor to be robust with respect to different handwriting styles. This is typically accounted for by asking for stability with respect to non-linear deformations of the feature to be extracted.

Spectacular success in many practical classification tasks has been reported for feature extractors generated by deep convolutional neural networks [4], [5]. The mathematical analysis of such networks was initiated by Mallat in [1]. Mallat’s theory applies to so-called scattering networks, where signals are propagated through layers that compute the modulus of the wavelet coefficients. The resulting feature extractor is provably translation-invariant and stable with respect to certain non-linear deformations. Moreover, it leads to state-of-the-art results in various image classification tasks [6], [7].

The wavelet transform resolves signal features characterized by point singularities, but is not very effective in dealing with signals dominated by anisotropic features, such as, e.g., edges in images [8]. It thus seems natural to ask whether Mallat’s theory on scattering networks can be extended to general signal transformations. Moreover, certain audio classification problems [9] suggest that scattering networks with different signal transformations in different layers would be desirable in practice.

Contributions: The goal of this paper is to extend Mallat’s theory to cope with general signal transformations (e.g., Gabor frames, wavelets, curvelets, shearlets, ridgelets), as well as to allow different signal transformations in different layers of the network, all that while retaining translation-invariance and deformation stability. Our second major contribution is a new deformation stability bound valid for a class of non-linear deformations that is wider than that considered by Mallat in [1]. The proofs in [1] all hinge critically on the wavelet transform’s structural properties, whereas the technical arguments in our proofs are completely detached from the particular structure of the signal transforms. This leads to simplified and shorter proofs for translation-invariance and deformation stability. Moreover, in the case of Mallat’s wavelet-based feature extractor we show that the admissibility condition for the mother wavelet (defined in [1, Theorem 2.6]) is not needed. The mathematical engine behind our results is the theory of continuous frames [10].

Notation and preparatory material: The complex conjugate of $z \in \mathbb{C}$ is denoted by $\bar{z}$. The Euclidean inner product of $x, y \in \mathbb{C}^d$ is $\langle x, y \rangle := \sum_{i=1}^{d} x_i \overline{y_i}$, with associated norm $||x|| := \sqrt{\langle x, x \rangle}$. The supremum norm of a matrix $M \in \mathbb{R}^{d \times d}$ is defined by $|M|_{\infty} := \sup_{i,j} |M_{i,j}|$, and the supremum norm of a tensor $T \in \mathbb{R}^{d \times d \times d}$ is $|T|_{\infty} := \sup_{i,j,k} |T_{i,j,k}|$. We write $B_R(x) \subseteq \mathbb{R}^d$ for the open ball of radius $R > 0$ centered at $x \in \mathbb{R}^d$. The Borel $\sigma$-algebra of $\mathbb{R}^d$ is denoted by $\mathcal{B}$. For a $\mathcal{B}$-measurable function $f : \mathbb{R}^d \to \mathbb{C}$, we write $\int_{\mathbb{R}^d} f(x) dx$ for the integral of $f$ with respect to Lebesgue measure $\mu_L$. For $p \in [1, \infty)$, $L^p(\mathbb{R}^d)$ denotes the space of all $\mathcal{B}$-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ such that $\|f\|_p := (\int_{\mathbb{R}^d} |f(x)|^p dx)^{1/p} < \infty$. For $f, g \in L^2(\mathbb{R}^d)$ we set $\langle f, g \rangle := \int_{\mathbb{R}^d} f(x) \overline{g(x)} dx$. The operator norm of the linear bounded operator $A : L^p(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is designated by $\|A\|_{p,q}$. Id : $L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ stands for the identity operator on $L^p(\mathbb{R}^d)$. For a countably infinite set $\mathcal{Q}$, $(L^2(\mathbb{R}^d))^\mathcal{Q}$ denotes the space of sets $s := \{f_q\}_{q \in \mathcal{Q}}$, $f_q \in L^2(\mathbb{R}^d)$, $\forall q \in \mathcal{Q}$, such that $|||s||| := (\sum_{q \in \mathcal{Q}} |||f_q|||_2^2)^{1/2} < \infty$. 
We write \( S(\mathbb{R}^d) \) for the Schwartz space, i.e., the space of functions \( f : \mathbb{R}^d \to \mathbb{C} \) whose derivatives along with the function itself are rapidly decaying [11, Section 7.3]. We denote the Fourier transform of \( f \in L^1(\mathbb{R}^d) \) by \( \hat{f}(\omega) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i(x,\omega)}dx \), and extend it in the usual way to \( L^2(\mathbb{R}^d) \) [11, Theorem 7.9]. The convolution of \( f \in L^2(\mathbb{R}^d) \) and \( g \in L^1(\mathbb{R}^d) \) is \( (f * g)(y) = \int_{\mathbb{R}^d} f(x)g(y-x)dx \). We write \( T_t f(x) := f(x-t), t \in \mathbb{R}^d \), for the translation operator, and \( M_{\omega} f(x) := e^{2\pi i(x,\omega)}f(x), \omega \in \mathbb{R}^d \), for the modulation operator. Inversion is defined by \( (I_f)(x) := \hat{f}(-x) \).

We denote the gradient of a function \( f : \mathbb{R}^d \to \mathbb{C} \) as \( \nabla f \). It is shown in [1] that the feature extractor \( \Phi \) is unambiguously defined as the output of the architecture illustrated in Figure 1, to be applicable in the construction of a wide range of wavelet frames and directions. The wavelets \( \{ \psi_{\lambda} \}_{\lambda \in \Lambda} \) along with the low-pass filter \( \phi \) are atoms of a semi-discrete Parseval wavelet frame \( \Psi_{\Lambda} \) and hence satisfy

\[
\| \phi \ast f \|^2 + \sum_{\lambda \in \Lambda} \| \psi_{\lambda} \ast f \|^2 = \| f \|^2, \quad \forall f \in L^2(\mathbb{R}^d).
\]

We refer the reader to Appendix A for a short review of the theory of semi-discrete frames. The architecture corresponding to (1), illustrated in Figure 1, is known as scattering network [6], and uses the same wavelets \( \{ \psi_{\lambda} \}_{\lambda \in \Lambda} \) in every network layer.

It is shown in [1] that the feature extractor \( \Phi_M \) in (1) is translation-invariant, in the sense that

\[
\Phi_M(T_t f) = T_t \Phi_M(f), \quad \forall t \in \mathbb{R}^d, \forall f \in L^2(\mathbb{R}^d),
\]

where \( T_t \) is applied element-wise in \( T_t \Phi_M(f) \). Further, it is proved in [1] that \( \Phi_M \) is stable with respect to deformations of the form

\[
F_{\tau} f(x) := f(x - \tau(x)).
\]

Specifically, for the normed function space \((H_M, \| \cdot \|_{H_M})\) defined in (8) below, Mallat proved that there exists a constant \( C > 0 \) such that for all \( f \in H_M \) and every \( \tau \in C^2(\mathbb{R}^d, \mathbb{R}^d) \) with \( \| \nabla \tau \|_\infty \leq \frac{1}{2C^2} \), the deformation error satisfies

\[
\| |\Phi_M(f) - \Phi_M(F_{\tau} f)|| \leq C(2^{-J} \| \tau \|_\infty + J \| D\tau \|_\infty + \| D^2 \tau \|_\infty) \| f \|_{H_M}.
\]

Remark 1. Examples of structured frames that satisfy the general semi-discrete frame condition (9) and will hence be seen, in Theorem 1, to be applicable in the construction of generalized feature extractors are, e.g., Gabor frames [12], curvelets [13, 14], shearlets [8], ridgelets [15, 16], and, of course, wavelets [17] as considered by Mallat in [1].

We now introduce our generalized scattering network. To this end, we generalize the multi-stage filtering technique underlying Mallat’s scattering network to allow for general semi-discrete frames that can, in addition, be different in different layers. This requires the definition of a general modulus-convolution operator, and of paths on index sets.

**Definition 2.** Let \( \Psi = (\Psi_n)_{n \in \mathbb{N}} \) be a frame collection with atoms \( \{ \phi_n \} \cup \{ f_{\lambda_n} \}_{\lambda_n \in \Lambda_n} \). For \( 1 \leq m < \infty \), define the

\[
\| f \|_{H_M} := \| f \|_{L^m(\mathbb{R}^d)}
\]
set \( \Omega^m_1 := \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_m \). An ordered sequence \( q = (\lambda_1, \lambda_2, \ldots, \lambda_m) \in \Omega^m_1 \) is called a path. The empty path, \( e := \emptyset \), defines the set \( \Omega^0 := \{ e \} \). The modulus-convolution operator is defined as \( U : (\bigcup_{n=1}^\infty \Lambda_k) \times L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d) \), \( U(\lambda_{m},f) := U[\lambda_{m}]f := |f \ast f_{\lambda_{m}}| \), where \( f_{\lambda_{m}} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \) are the atoms of the semi-discrete frame \( \Psi_n \) associated with the \( n \)-th layer in the network.

We also need to extend the operator \( U \) to paths \( q \in \Omega^m_1 \) and do that according to

\[
U[q]f := (U[\lambda_m] \cdots U[\lambda_2]U[\lambda_1]f) = |f \ast f_{\lambda_{1}}| \ast |f \ast f_{\lambda_{2}}| \ast \cdots \ast |f \ast f_{\lambda_{m}}|,
\]

where we set \( U[e]f = f \). Note that the multi-stage filtering operation (4) is well-defined, as \( \|U[q]f\|_2 \leq (\prod_{n=1}^m \|f_{\lambda_{n}}\|_2)\|f\|_2 \), thanks to Young’s inequality [18, Theorem 1.2.12]. Figure 2 illustrates the generalized scattering network architecture based on general multi-stage filtering (4). The function \( f_{\lambda_{n}}^{(k)} \) is the \( k \)-th atom of the semi-discrete frame \( \Psi_n \) associated with the \( n \)-th layer.

**Definition 3.** Let \( \Psi = (\Psi_n)_{n \in \mathbb{N}} \) be a frame collection, and define \( Q := \bigcup_{n=\mathbb{N}} \Omega^m_1 \). Given a path \( q \in \Omega^m_1 \), \( n \geq 0 \), we write \( \phi[q] := \phi_{\lambda_{n+1}} \) for the output-generating atom of the semi-discrete frame \( \Psi_{n+1} \). The feature extractor \( \Phi_\Psi \) with respect to the frame collection \( \Psi \) is defined as

\[
\Phi_\Psi(f) := \{U[q]f \ast \phi[q] \mid q \in Q\}. \tag{5}
\]

**IV. MAIN RESULT**

The main result of this paper is the following theorem, stating that the feature extractor \( \Phi_\Psi \) defined in (5) is translation-invariant and stable with respect to time-frequency deformations of the form

\[
F_{\tau,\omega}f(x) := e^{2\pi i \omega(x)}f(x - \tau(x)). \tag{6}
\]

The class of deformations we consider is wider than the one in Mallat’s theory, who considered translation-like deformations of the form \( f(x - \tau(x)) \) only. Modulation-like deformations \( e^{2\pi i \omega(x)}f(x) \) occur, e.g., if we have access only to a band-pass version of the signal \( f \in L^2(\mathbb{R}^d) \).

**Theorem 1.** Let \( \Psi \) be a frame collection with upper frame bound \( B \leq 1 \). The feature extractor \( \Phi_\Psi \) defined in (5) is translation-invariant. Further, for \( R > 0 \), define the space of \( R \)-band-limited functions

\[
H_R := \{ f \in L^2(\mathbb{R}^d) \mid \text{supp}(\hat{f}) \subseteq B_R(0) \}.
\]

Then, the feature extractor \( \Phi_\Psi \) is stable on \( H_R \) with respect to non-linear deformations (6), i.e., there exists \( C > 0 \) (that does not depend on \( \Psi \)) such that for all \( f \in H_R \) and all \( \omega \in C(\mathbb{R}^d,\mathbb{R}) \), \( \tau \in C(\mathbb{R}^d,\mathbb{R}^d) \) with \( \|D\tau\|_{\infty} \leq \frac{1}{2C} \), it holds that

\[
\|\Phi_\Psi(f) - \Phi_\Psi(F_{\tau,\omega}f)\| \leq C(R\|\tau\|_{\infty} + \|\omega\|_{\infty})\|f\|_2. \tag{7}
\]

The proof of Theorem 1 can be found in Appendix B. Our main result shows that translation-invariance and deformation stability are retained for the generalized feature extractor \( \Phi_\Psi \). The strength of this result derives itself from the fact that the only condition on \( \Psi \) for this to hold is \( B \leq 1 \). This condition is easily met by normalizing the frame elements accordingly. Such a normalization impacts neither translation-invariance nor the constant \( C \) in (7) which is seen, in (14), to be independent of \( \Psi \). All this is thanks to our proof techniques, unlike those in [1], being independent of the algebraic structure of the underlying frames. This is accomplished through a generalization of a Lipschitz-continuity result by Mallat [1, Proposition 2.5] for the feature extractor \( \Phi_\Psi \) (stated in Proposition 2 in Appendix B), and by employing a partition of unity argument [19] for band-limited functions.

**V. RELATION TO MALLAT’S RESULTS**

To see how Mallat’s wavelet-based architecture is covered by our Theorem 1, simply note that by [1, Eq. 2.7] the atoms \( \{\phi_\lambda\} \cup \{\psi_\lambda\}_{\lambda \in \Lambda_0} \) satisfy (10) with \( A = B = 1 \). Since Mallat’s construction uses the same wavelet frame in each layer, this trivially implies \( \sup_{n \in \mathbb{N}} B_n \leq 1 \).

Mallat imposes additional technical conditions on the atoms \( \{\phi_\lambda\} \cup \{\psi_\lambda\}_{\lambda \in \Lambda_0} \), one of which is the so-called scattering admissibility condition for the mother wavelet, defined in [1, Theorem 2.6]. To the best of our knowledge, no wavelet in \( \mathbb{R}^d \), \( d \geq 2 \), satisfying this condition has been reported in the literature.

Mallat’s stability bound (3) applies to signals \( f \in L^2(\mathbb{R}^d) \) satisfying

\[
\|f\|_{H_m} := \sum_{m=0}^{\infty} \sum_{q \in \Lambda_0^m \cap \mathbb{P}} \|U[q]f\|_2 < \infty. \tag{8}
\]

While [1, Section 2.5] cites numerical evidence on (8) being finite for a large class of functions \( f \in L^2(\mathbb{R}^d) \), it seems difficult to establish this analytically.

Finally, the stability bound (3) depends on the parameter \( J \), which determines the coarsest scale resolved by the wavelets \( \{\psi_\lambda\}_{\lambda \in \Lambda_0} \). For \( J \to \infty \) the term \( 2^{-J}\|\tau\|_{\infty} \) vanishes; however, the term \( J\|D\tau\|_{\infty} \) tends to infinity.

Our main result shows that i) the scattering admissibility condition in [1] is not needed, ii) instead of the signal class

\[f \in L^2(\mathbb{R}^d)\]

is only required that

\[
\|f\|_{H_m} := \sum_{m=0}^{\infty} \sum_{q \in \Lambda_0^m \cap \mathbb{P}} \|U[q]f\|_2 < \infty. \tag{8}
\]
characterized by (8) our result applies provably to the space of $R$-band-limited functions $H_R$, and iii) our deformation stability bound (7), when particularized to wavelets, besides applying to a wider class of non-linear deformations, namely (6) instead of (2), is independent of $J$.

The proof technique used in [1] to establish (3) makes heavy use of structural specifics of the atoms $\{\phi_J\} \cup \{\psi_{\lambda}\}_{\lambda \in \Lambda_W}$, namely isotropic dilations, vanishing moment conditions, and a constant number $K \in \mathbb{N}$ of directional wavelets across scales.

**APPENDIX A**

**SEMI-DISCRETE FRAMES**

This appendix gives a short review of semi-discrete frames [10].

**Definition 4.** Let $\{f_\lambda\}_{\lambda \in \Lambda} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ be a set of functions indexed by a countable set $\Lambda$. The set of translated and involuted functions

$$\Psi_\Lambda = \{T_bI f_\lambda\}_{(b,\lambda) \in \Lambda \times \mathbb{R}^d}$$

is called a semi-discrete frame, if there exist constants $A, B > 0$ such that

$$A\|f\|_2^2 \leq \sum_{\lambda \in \Lambda} \|f * f_\lambda\|_2^2 \leq B\|f\|_2^2$$

for all $f \in L^2(\mathbb{R}^d)$. The functions $\{f_\lambda\}_{\lambda \in \Lambda}$ are called the atoms of the semi-discrete frame $\Psi_\Lambda$. When $A = B$ the semi-discrete frame is said to be tight. A tight semi-discrete frame with frame bound $A = 1$ is called a semi-discrete Parseval frame.

The frame operator associated with the semi-discrete frame $\Psi_\Lambda$ is defined in the weak sense by $S_\Lambda : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$,

$$S_\Lambda f = \left(\sum_{\lambda \in \Lambda} f_\lambda * I f_\lambda\right) * f.$$  

$S_\Lambda$ is a bounded, positive, and boundedly invertible operator [10].

The reader might want to think of semi-discrete frames as shift-invariant frames [20], where the translation parameter is left unsampled. The discrete index set $\Lambda$ typically labels a collection of scales, directions, or frequency-shifts. For instance, as illustrated in Section II, Mallat’s scattering network is based on a semi-discrete Parseval frame of directional wavelet structure, where the atoms $\{\phi_J\} \cup \{\psi_{\lambda}\}_{\lambda \in \Lambda_W}$ are indexed by the set $\Lambda_W = \{(j, k) \mid j > -J, k \in \{1, \ldots, K\}\}$, labeling a collection of scales and directions.

For shift-invariant frames it is often convenient to work with a unitarily equivalent representation of the frame operator.

**Proposition 1.** [17, Theorem 5.11] Let $\Lambda$ be a countable index set. The functions $\{f_\lambda\}_{\lambda \in \Lambda} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ are atoms of the semi-discrete frame $\Psi_\Lambda = \{T_bI f_\lambda\}_{(b,\lambda) \in \Lambda \times \mathbb{R}^d}$ with frame bounds $A, B > 0$ if and only if

$$A \leq \sum_{\lambda \in \Lambda} |f_\lambda(\omega)|^2 \leq B, \quad \text{a.e. } \omega \in \mathbb{R}^d.$$  

**APPENDIX B**

**PROOF OF THEOREM 1**

We first prove translation-invariance. Fix $f \in L^2(\mathbb{R}^d)$ and define $C[q]f := U[q]f * \phi[q]$, $\forall q \in \mathfrak{Q}$. By (5) it follows that $\Phi_\Psi$ is translation-invariant if and only if

$$C[q](T_t f) = T_t(C[q]f), \quad \forall t \in \mathbb{R}^d, \quad \forall q \in \mathfrak{Q}.$$  

(11)

Due to $C[q](T_t f) = U[q](T_t f) * \phi[q]$ and

$$T_t(C[q]f) = T_t(U[q]f * \phi[q]) = (T_t(U[q]f)) * \phi[q],$$

(11) holds if $U[q](T_t f) = T_t(U[q]f)$, $\forall t \in \mathbb{R}^d, \forall q \in \mathfrak{Q}$. The proof is concluded by noting that $U[q]$ is translation-invariant thanks to (4) and

$$U[\lambda_n](T_t f) = |(T_t f) * f_{\lambda_n}| = |T_t(f * f_{\lambda_n})| = T_t(U[\lambda_n]f),$$

for all $t \in \mathbb{R}^d$, $\lambda_n \in \bigcup_{k=1}^{\infty} \Lambda_k$.

Let us now turn to the proof of deformation stability, which is based on two key ingredients, the first being a generalization of a Lipschitz-continuity result by Mallat [1, Proposition 2.5]:

**Proposition 2.** Let $\Psi$ be a frame collection with upper frame bound $B \leq 1$. The feature extractor $\Phi_\Psi : L^2(\mathbb{R}^d) \rightarrow (L^2(\mathbb{R}^d))^{Q}$ is a bounded, Lipschitz-continuous operator with Lipschitz constant $L = \sqrt{B}$, i.e.,

$$||\Phi_\Psi(f) - \Phi_\Psi(h)|| \leq \sqrt{B}\|f - h\|_2$$

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

The proof of Proposition 2 is not given here, as it essentially follows that of [1, Proposition 2.5] with minor changes. We now apply Proposition 2 with $h := F_{\tau,\omega}f$ and get

$$||\Phi_\Psi(f) - \Phi_\Psi(F_{\tau,\omega}f)|| \leq \|f - F_{\tau,\omega}f\|_2$$

for all $f \in L^2(\mathbb{R}^d)$. Here, we used $\sqrt{B} \leq 1$, due to $B \leq 1$, as well as $h = F_{\tau,\omega}f \in L^2(\mathbb{R}^d)$, which is thanks to

$$\|h\|^2_2 = \|F_{\tau,\omega}(f)\|^2_2 = \int_{\mathbb{R}^d} |f(x - \tau(x))|^2 dx \leq 2\|f\|^2_2,$$

obtained through the change of variables $u = x - \tau(x)$, together with

$$\frac{du}{dx} = |\det(\text{Id} - D\tau(x))| \geq 1 - d||D\tau||_{\infty} \geq 1/2.$$  

(12)
The inequalities in (12) hold thanks to [21, Corollary 1] and \( \|D\tau\|_\infty \leq \frac{1}{2\lambda} \), respectively. The second key ingredient of our proof is a partition of unity argument [19] for band-limited functions used to derive an upper bound on \( \|f - F_{\tau,\omega}\|_2 \). We first determine a function \( \gamma \) such that \( f = f * \gamma \) for all \( f \in H_R \). Consider \( \eta \in S(\mathbb{R}^d) \) such that \( \hat{\eta}(\omega) = 1, \forall \omega \in B_1(0) \). Setting 
\[
\gamma(x) := R^d\eta(Rx) \text{ yields } \hat{\gamma}(\omega) = \hat{\eta}(\omega)/R. \text{ Thus, } \hat{\gamma}(\omega) = 1, \forall \omega \in B_R(0) \text{, as well as } \hat{f} = \hat{f} \hat{\gamma} \text{ and } f = f * \gamma \text{ for all } f \in H_R. \text{ Then, we define the operator } A_{\gamma} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d), \ A_{\gamma}(f) = f * \gamma. \text{ Note that } A_{\gamma} \text{ is well-defined as } \gamma \in S(\mathbb{R}^d) \subseteq L^1(\mathbb{R}^d). \text{ We now get}
\[
\|f - F_{\tau,\omega}\|_2 = \|A_{\gamma}f - F_{\tau,\omega}A_{\gamma}f\|_2 \
\leq \|A_{\gamma}f - F_{\tau,\omega}A_{\gamma}\|_{2,2}\|f\|_2
\]
for all \( f \in H_R \). In order to bound the norm \( \|A_{\gamma}f - F_{\tau,\omega}A_{\gamma}\|_{2,2} \), we apply Schur’s Lemma to the integral operator \( F_{\tau,\omega}A_{\gamma} \).

**Schur’s Lemma.** [18, App. 1.1] Let \( k : \mathbb{R}^d \times \mathbb{R}^d \to C \) be a locally integrable function satisfying \( \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} |k(x,u)|\,du \leq C \) and \( \sup_{u \in \mathbb{R}^d} \int_{\mathbb{R}^d} |k(x,u)|\,dx \leq C \). Then, the integral operator \( K \) given by \( K(f)(x) = \int_{\mathbb{R}^d} k(x,u)f(u)\,du \), is a bounded operator from \( L^2(\mathbb{R}^d) \) to \( L^2(\mathbb{R}^d) \) with norm \( \|K\|_{2,2} \leq C \).

From the identity
\[
F_{\tau,\omega}A_{\gamma}(f)(x) = e^{2\pi i \omega(x)} \int_{\mathbb{R}^d} \gamma(x - \tau(x) - u)f(u)\,du,
\]

it follows that \( F_{\tau,\omega}A_{\gamma} - A_{\gamma} \) has the kernel function \( k(x,u) := e^{2\pi i \omega(x)}\gamma(x - \tau(x) - u) - \gamma(x - u) \), which is locally integrable thanks to \( \gamma \in S(\mathbb{R}^d) \) and \( \tau \in C(\mathbb{R}^d, \mathbb{R}^d) \). We next use a first-order Taylor expansion in order to bound \( |k(x,u)| \).

To this end, let \( x, u \in \mathbb{R}^d \), and define \( h(x,u) : \mathbb{R} \to \mathbb{C} \), as \( h(x,u)(t) = e^{2\pi i \omega(x)}\gamma(x - \tau(x) - u) - \gamma(x - u) \). It follows that \( h(x,u)(0) = 0 \) and \( h(x,u)'(1) = k(x,u) \). Therefore, we have \( h(x,u)(t) = h(x,u)(0) + \int_0^t \frac{d}{dt} h(x,u)(\lambda)\,d\lambda \), \( \forall t \in \mathbb{R} \).

The special choice \( t = 1 \) yields \( |k(x,u)| = |h(x,u)(1)| \leq \int_0^1 \left| \frac{d}{dt} h(x,u)(\lambda) \right| d\lambda \) with
\[
\left| \frac{d}{dt} h(x,u)(\lambda) \right| \leq \langle \nabla \gamma(x - \lambda \tau(x) - u), \tau(x) \rangle + 2\pi |\omega(x)| \gamma(x - \lambda \tau(x) - u) \leq 0
\]

Thanks to \( \gamma, \nabla \gamma \in S(\mathbb{R}^d) \), and \( \mu_L([0,1]) = 1 < \infty \), we can apply Fubini’s Theorem to get
\[
\int_{\mathbb{R}^d} |k(x,u)|\,du \leq \|\tau\|_{\infty} \int_0^1 \int_{\mathbb{R}^d} \left| \nabla \gamma(x - \lambda \tau(x) - u) \right| d\lambda d\lambda
\]

Similarly, we obtain
\[
\int_{\mathbb{R}^d} |f(x,u)|\,dx \leq \|\tau\|_{\infty} \int_0^1 \int_{\mathbb{R}^d} \left| \nabla \gamma(x - \lambda \tau(x) - u) \right| d\lambda d\lambda
\]

The inequalities in (13) hold thanks to [21, Corollary 1], \( \|D\tau\|_{\infty} \leq \frac{1}{2\lambda} \), and \( \lambda \in (0,1] \). The proof is completed by setting
\[
C := \max \left\{ 2\|\nabla \eta\|_1, 4\pi \|\eta\|_1 \right\} \|R\|_{\infty} + \|\omega\|_{\infty} \right\}
\]