

A. Group Transform: A Group Representation Approach

A.1. Background

For further details on the group theoretical aspects described in this section, the reader should refer to [Vilenkin \(1978\)](#).

Definition 1. A group is a set \mathbf{G} with a multiplicative operation \odot that respects enclosure, identity element, inverse element, and associativity.

The representation of the group determines its action on a function space and bridges the gap between group theory and linear algebra, allowing to compute the transformation of a function following the rules induced by the specific group at hand. The representation of a group can be thought as a far-reaching generalization of the exponential function property, $\exp(x + y) = \exp(x)\exp(y)$, $\forall x, y \in \mathbb{R}$ ([Baraniuk, 1993](#)). In fact, it is defined as,

Definition 2. A linear continuous representation ρ of a group \mathbf{G} on the linear space \mathbb{H} is defined as

$$\rho : \mathbf{G} \rightarrow GL(\mathbb{H}), \quad (1)$$

where $GL(\mathbb{H})$ is the the group of linear map in \mathbb{H} such that $\forall g, g' \in \mathbf{G}$

$$\rho(g \odot g') = \rho(g)\rho(g'). \quad (2)$$

For instance, let \mathbb{H} be a vector space such as \mathbb{R}^3 , the representation of the group is induced by 3×3 matrices. In this case, the operation on the right of (2) is a matrix multiplication, where each matrix depends on the group elements g and g' . This concept extends to linear operators acting on functional spaces.

As such, multiple transformations of a function by different elements of the group is equal to the representation of the combination of the group elements applied to the function.

This structure-preserving map defines the action of a group on elements of function spaces. Group transforms such as STFT and CWT can be expressed in such a way by selecting a mother filter space and a group. The representation of the group in the mother filter space provides an operator that takes as input an element of the group and acts on the filter to transform it. A filter-bank can thus be created by iterating this process with different group elements. Therefore, the selected group carries the characteristics of the filter-bank and consequently, the group transform and its time-frequency tiling. The building blocks of the WT through representation theory is provided in [Appendix A.2](#). Notice that further properties such as the invariant measure of the group and the resolution of the identity can be develop using the representation of the group.

A.2. Example: The Wavelet Transform

As an introductory example, we consider the creation of a wavelet filter-bank utilizing transformation group. Let's denote by \mathbf{G}_{aff} the affine group, the so called " $ax + b$ " group, where the elements $(\lambda, \tau) \in \mathbb{R}_+^* \times \mathbb{R}$, where $\mathbb{R}_+^* = (0, +\infty)$, where the multiplicative operation of the group \odot is defined by

$$(\lambda, \tau) \odot (\lambda', \tau') = (\lambda\lambda', \tau + \lambda\tau') \quad (3)$$

Let's define by ρ_{aff} the representation of the affine group in $\mathbb{L}_2(\mathbb{R})$, i.e., $\rho_{\text{aff}} : \mathbf{G}_{\text{aff}} \rightarrow GL(\mathbb{L}_2(\mathbb{R}))$, such that ρ_{aff} is a homomorphism as per [Definition 2](#). Its action on square integrable function ψ is defined as

$$[\rho_{\text{aff}}(g)\psi](t) = \frac{1}{\sqrt{\lambda}}\psi\left(\frac{t - \tau}{\lambda}\right), \quad t \in \mathbb{R}, \quad (4)$$

where (a, b) are respectively the dilation and translation parameters. The wavelet filter-bank is built by transforming a mother filter, ψ by the representation ρ_{aff} for specific elements of the group. A visualization of this approach for a Morlet wavelet filter can be seen in [Figure \(3\)](#). The wavelet transform of a signal $s_i \in \mathbb{L}_2(\mathbb{R})$ is achieved by

$$\mathcal{W}(s_i, \psi)(g_{(\lambda, \tau)}) = \langle s_i, \rho_{\text{aff}}(g_{(\lambda, \tau)})\bar{\psi} \rangle, \quad \forall g_{(\lambda, \tau)} \in \mathbf{G}_{\text{aff}}, \quad (5)$$

$$= (s_i \star \rho_{\text{aff}}(g_{(\lambda, 0)})\psi), \quad \forall g_{(\lambda, 0)} \in \mathbf{G}_{\text{aff}}, \quad (6)$$

where $\bar{\psi}(t) = \psi(-t)$, $\langle ., . \rangle$ denotes the inner product, \star the convolution, and $\rho_{\text{aff}}(g_{(\lambda, \tau)})\psi$ the action of the operator ρ_{aff} , evaluated at the group element $g_{(\lambda, \tau)}$, on the mother filter ψ as per (4). In practice, the filter-bank is generated by sampling

a few elements of the group. For instance, in the case of the dyadic wavelet transform, the dilation parameters follow a geometric progression of common ratio equals to 2. In general, the translation parameter is sampled according to the scaling one (Daubechies, 1992). Notice that in the convolution expression (6), the translation parameter $\tau = 0$, in fact the convolution operator \star acts as the translation one. In the case where the translation parameter depends on the scaling one, a specific stride is used to perform the discrete convolution.

Note that the STFT can be constructed similarly utilizing the Weyl-Heisenberg group (Feichtinger et al., 2009), whose representation on $\mathbb{L}_2(\mathbb{R})$ consists of frequency modulations and translations. More intricated group representations can be built as in Torrésani (1991) where the combination of the affine group and Weyl-Heisenberg group is considered.

B. Architecture Details

B.1. Artificial Data

```
Group Transform + Complex Modulus + Log
Dense Layer (1 sigmoid)
```

After the Group Transform, a batch-normalization is applied.

B.2. Supervised Bird Detection

```
Group Transform + Complex Modulus + Log + Average-Pooling (stride:(1, 512) size:(1, 1024))
Conv2D. layer (16 filters 3 × 3) and Max-Pooling (3 × 3) and ReLU
Conv2D. layer (16 filters 3 × 3) and Max-Pooling (3 × 3) and ReLU
Conv2D. layer (16 filters 3 × 1) and Max-Pooling (3 × 1) and ReLU
Conv2D. layer (16 filters 3 × 1) and Max-Pooling (3 × 1) and ReLU
    Dense layer (256) and ReLU
    Dense layer (32) and ReLU
    Dense layer (1 sigmoid)
```

At each layer a batch-normalization is applied and for the last three layers a 50% dropout is applied as in ((Grill & Schlüter, 2017)). The dimension of the input of the DNN presented is the same for the different benchmarks.

B.3. Haptics Data

```
Group Transform + Complex Modulus + Log + Average-Pooling (stride:(1, 64) size:(1, 128))
Dense Layer (5 softmax)
```

After the Group Transform, a batch-normalization is applied. For this dataset, we perform early stopping as in the benchmarks, not providing any standard deviation.

C. Additional Figures

C.1. Artificial Data

C.1.1. DATA

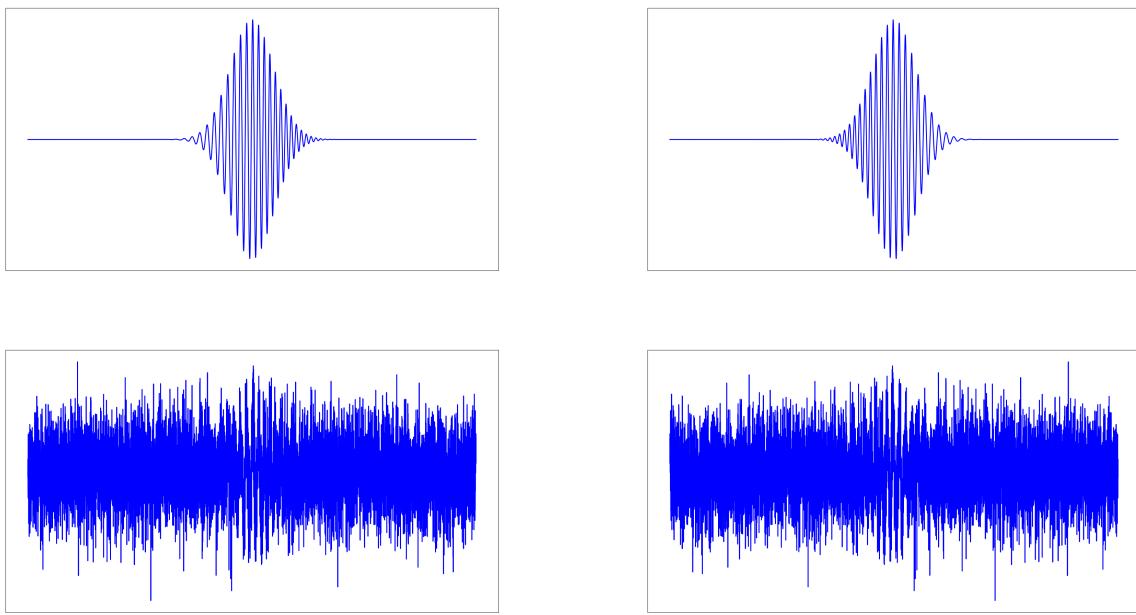


Figure 8. Artificial Dataset: (Top Left) Ascending Chirp, (Top Right) Descending Chirp, i.e. class 0, (Bottom Left) Ascending Chirp plus Gaussian noise, (Bottom Right) Descending Chirp plus Gaussian noise, i.e., class 1. The samples contained in the training and testing set are higher in frequency and close to the Nyquist frequency.

C.1.2. GROUP TRANSFORM

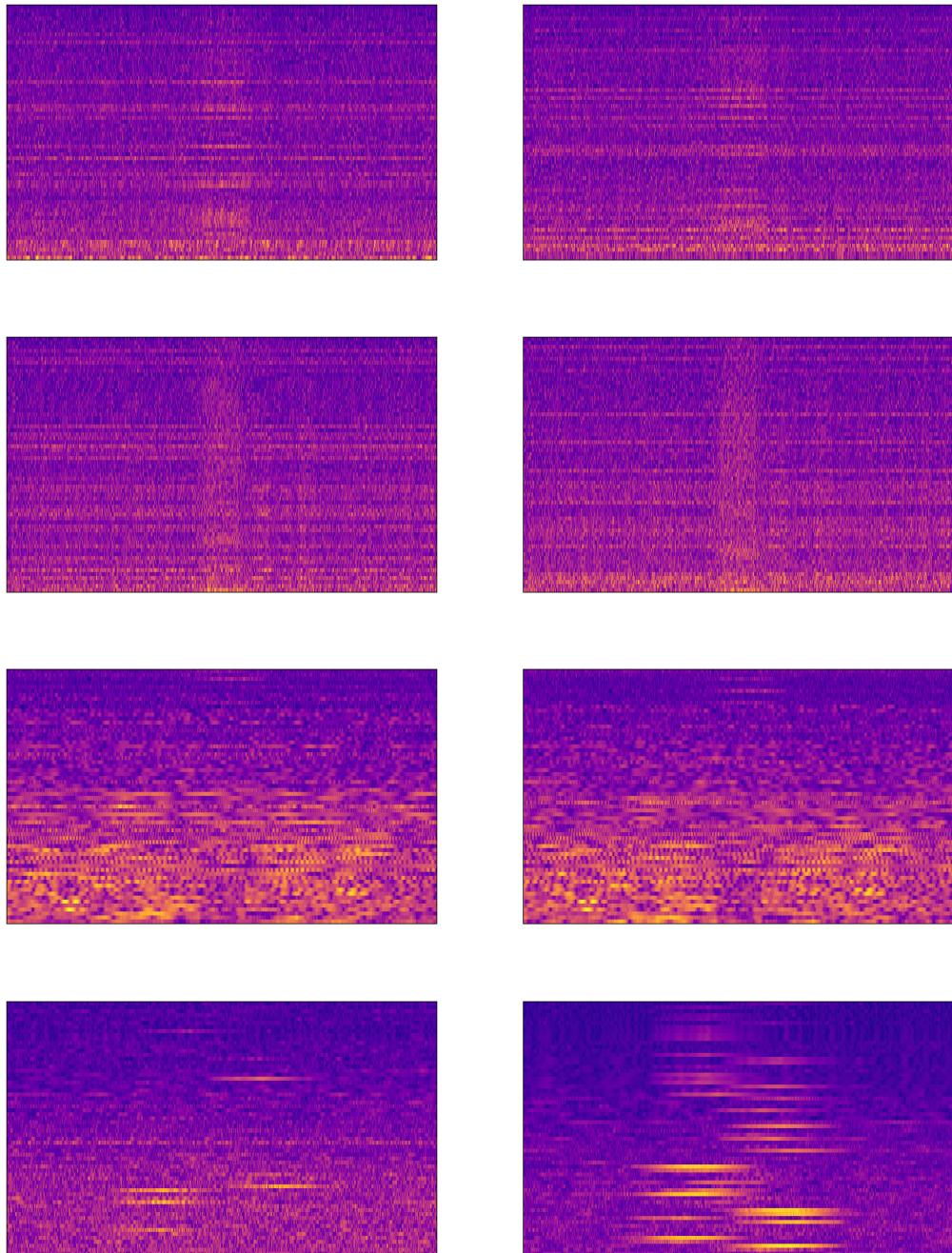


Figure 9. Learnable Group Transform - Visualisation of an ascending chirp sample, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT, cnLGT.

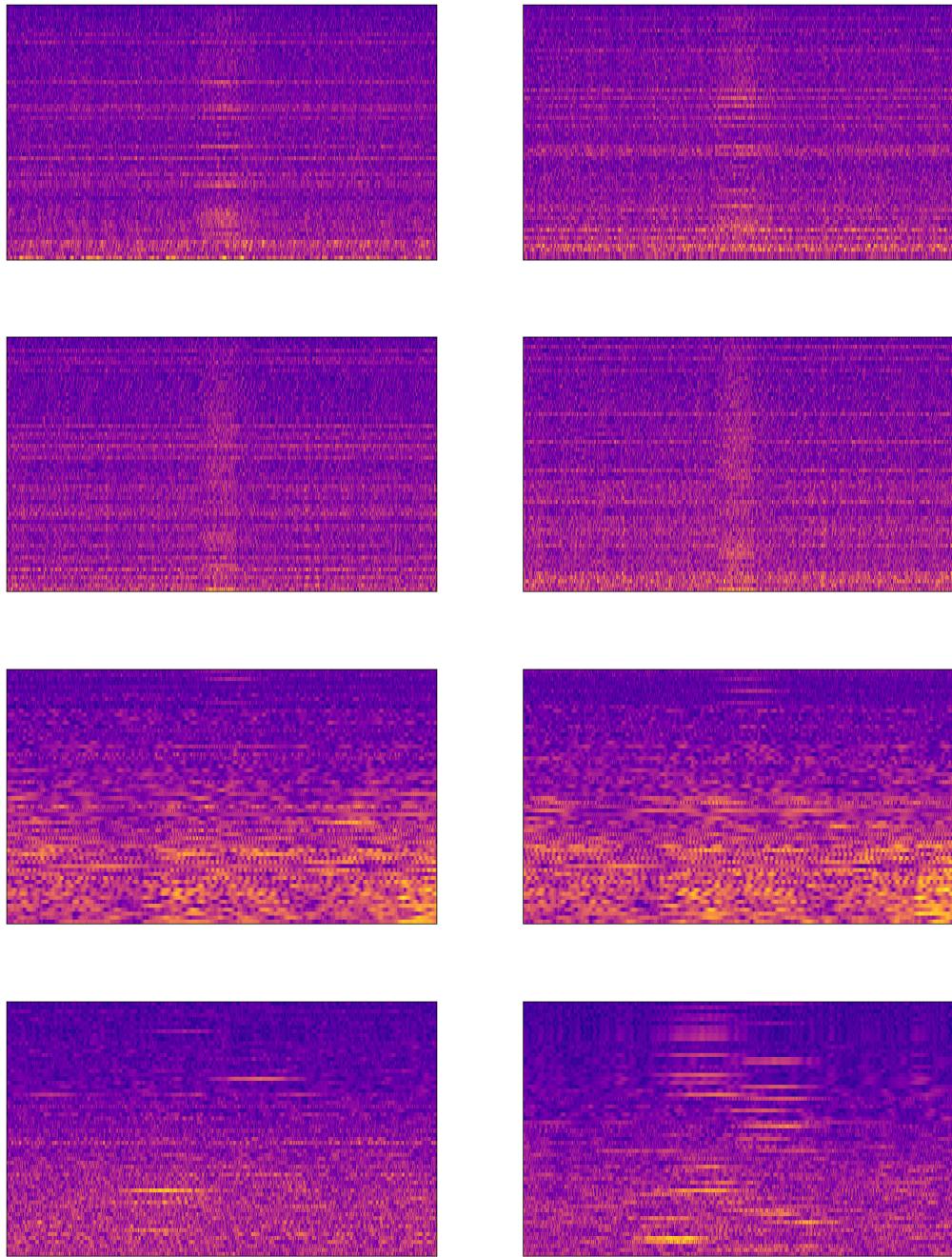


Figure 10. Learnable Group Transform - Visualisation of a descending chirp sample, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT, cnLGT.

C.2. Supervised Bird Detection

C.2.1. DATA

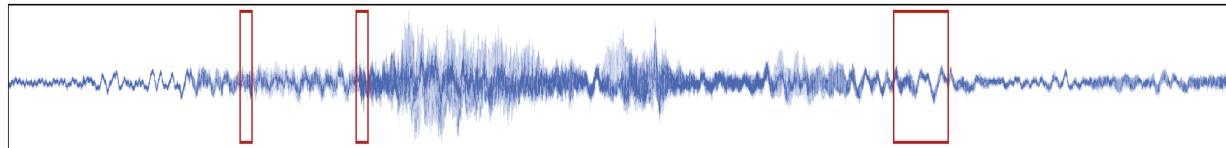


Figure 11. Bird Detection Dataset - Sample containing a bird song. The x -axis is the time and the y -axis the amplitude. The red boxes are the locations of the bird song. Each data sample are normalized, centered and subsampled by two before experiment.

C.2.2. GROUP TRANSFORM

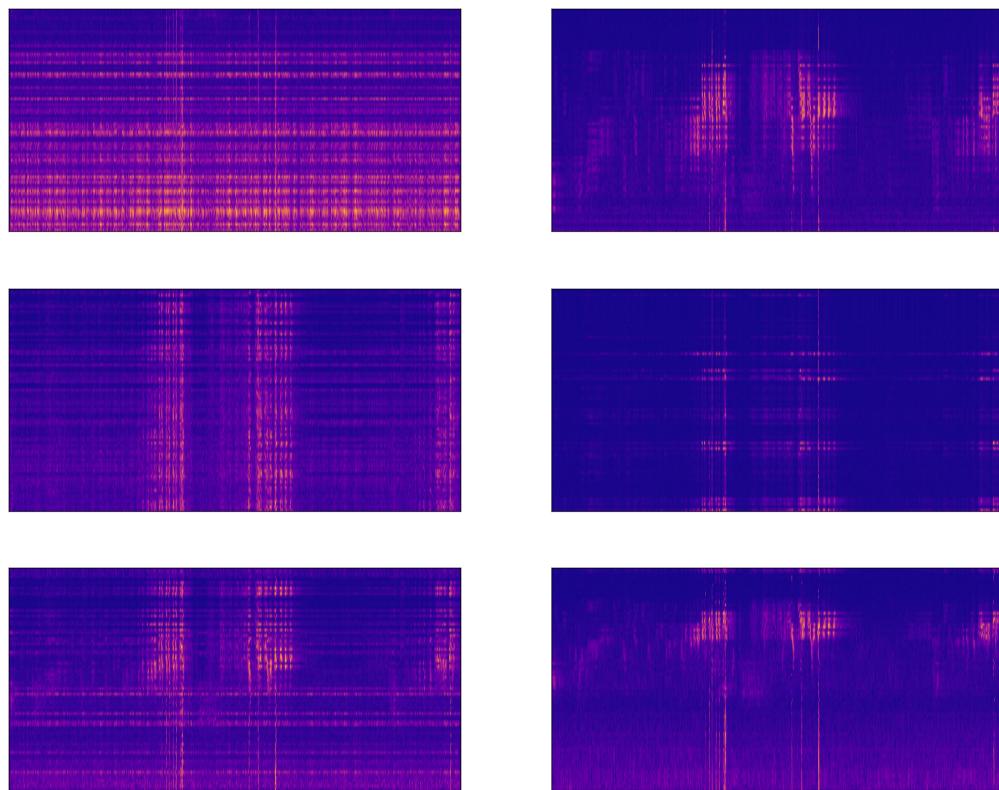


Figure 12. Learnable Group Transform - Visualisation of a sample containing a bird song, where for each row (left) at the initialization and (right) after learning. Each row displays a different setting: (from top to bottom): LGT, nLGT, cLGT.

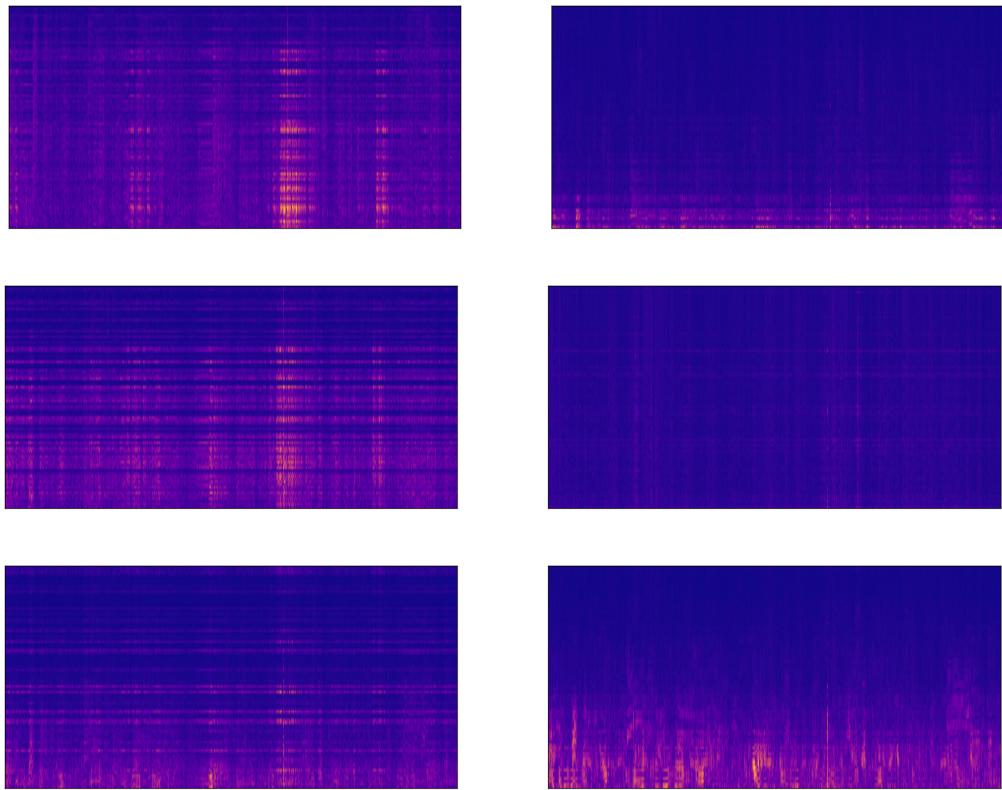


Figure 13. Learnable Group Transform - Visualisation of a sample without a bird song, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT.

C.3. Haptics Data

C.3.1. DATA

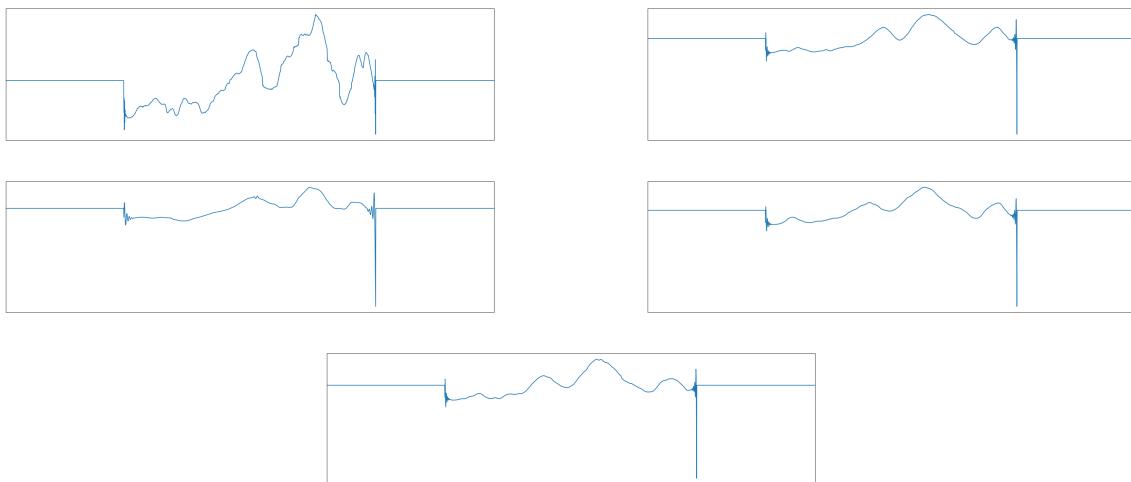


Figure 14. Haptic Dataset - Sample of each class of the Haptic dataset.

Each data is centered and normalized. For the experiments, the number of epochs is set to 1000 and we perform early-stopping and obtain the testing accuracy at this specific epoch as in Khan & Yener (2018), the batch size was set to 64. In order to avoid overfitting, we perform different asymmetric zeros-paddings on the training samples. For the testing samples, we perform a symmetric zeros-padding (512 zeros on each side of the signals).

C.3.2. GROUP TRANSFORM

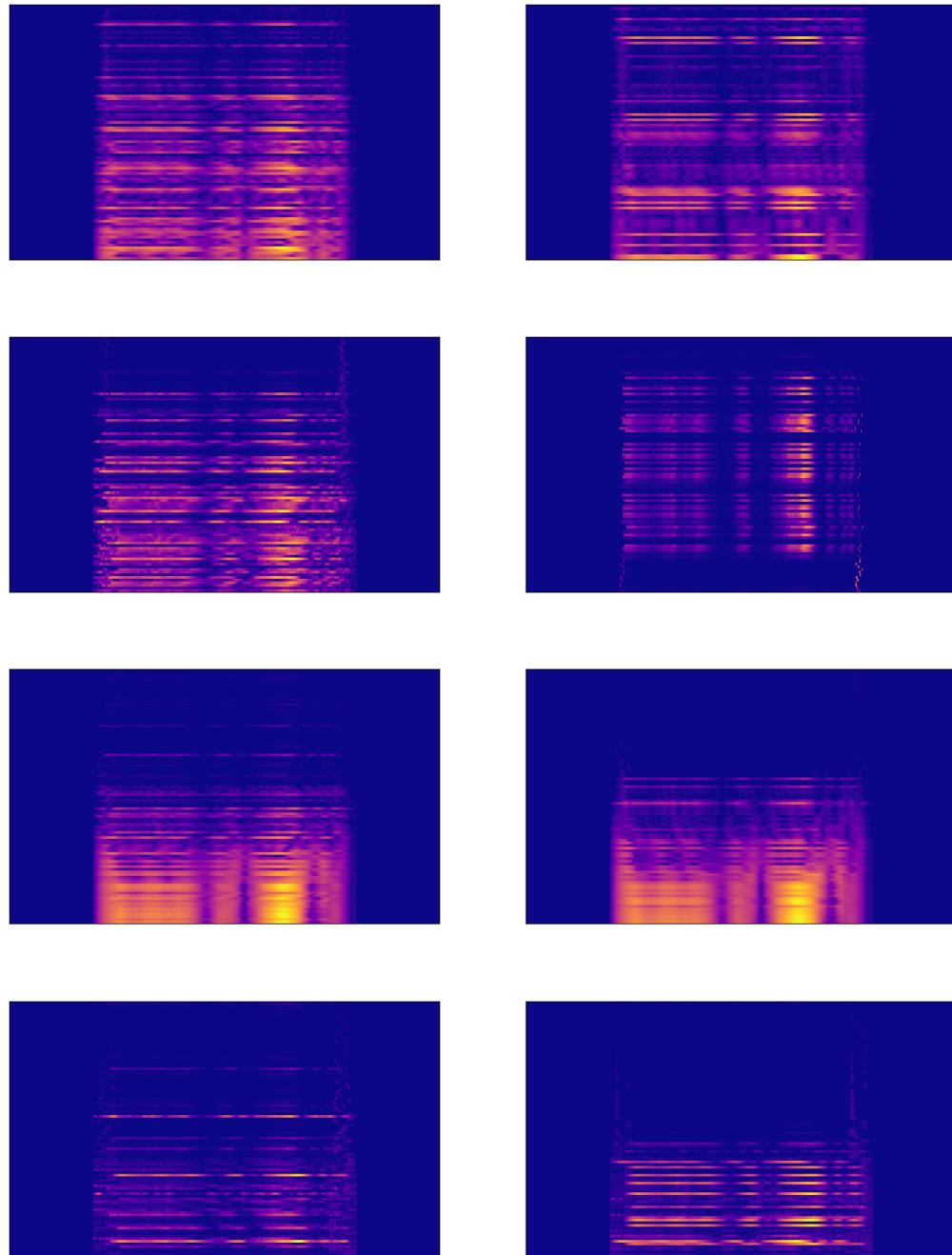


Figure 15. Learnable Group Transform - Visualisation of a sample belonging to class 1, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT, cnLGT.

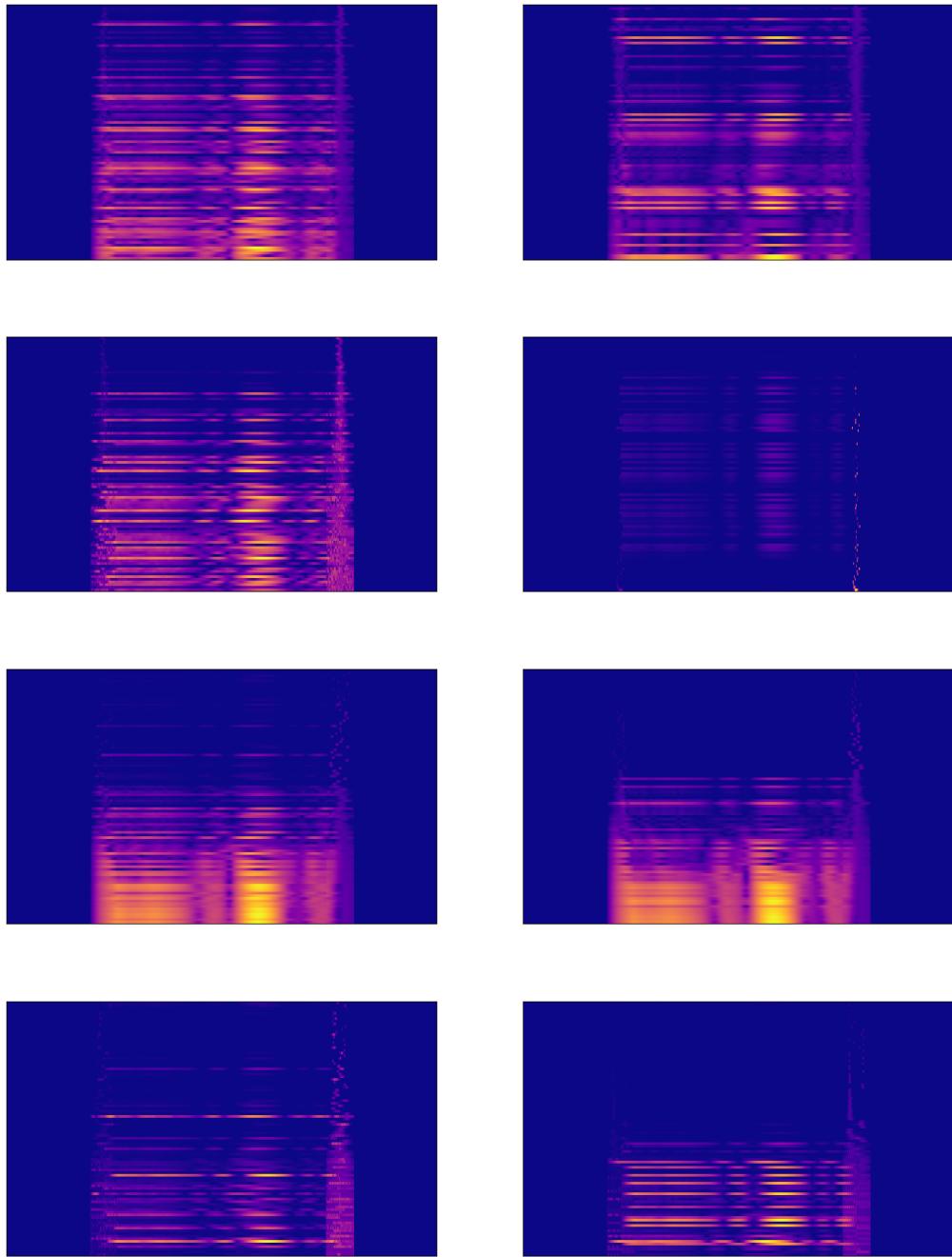


Figure 16. Learnable Group Transform - Visualisation of sample belonging to class 2, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT, cnLGT.

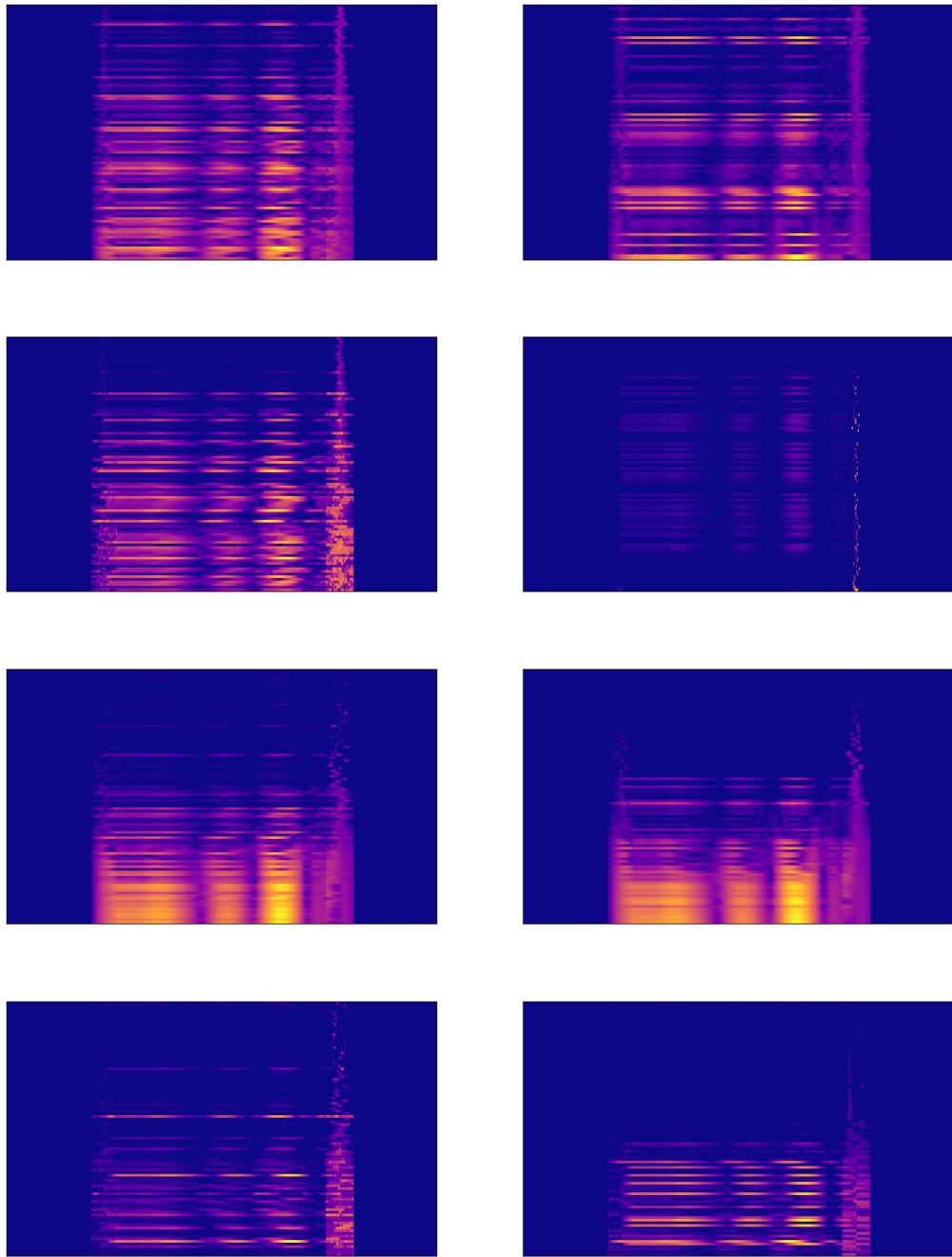


Figure 17. Learnable Group Transform - Visualisation of sample belonging to class 3, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT, cnLGT.

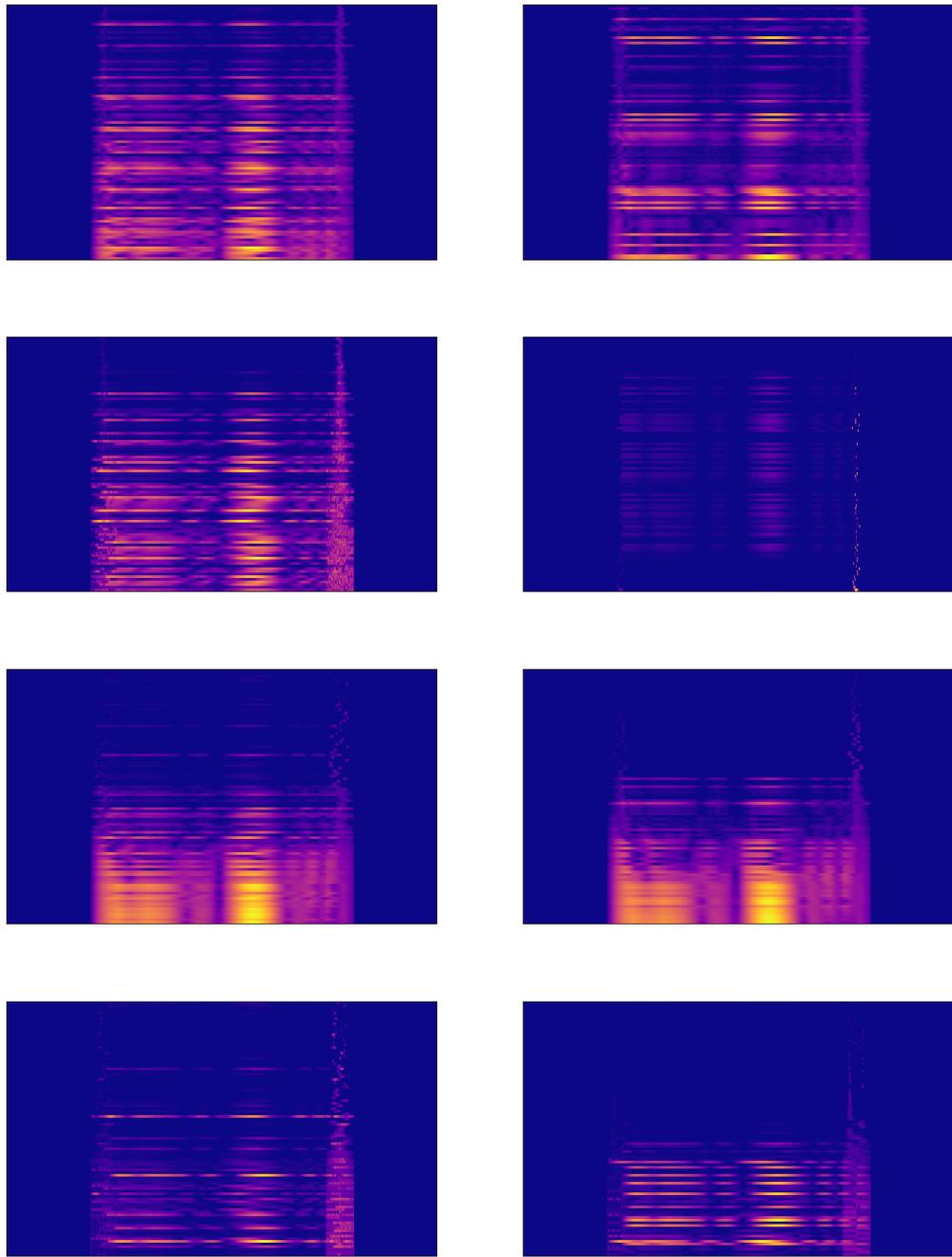


Figure 18. Learnable Group Transform - Visualisation of a sample belonging to class 4, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT, cnLGT.

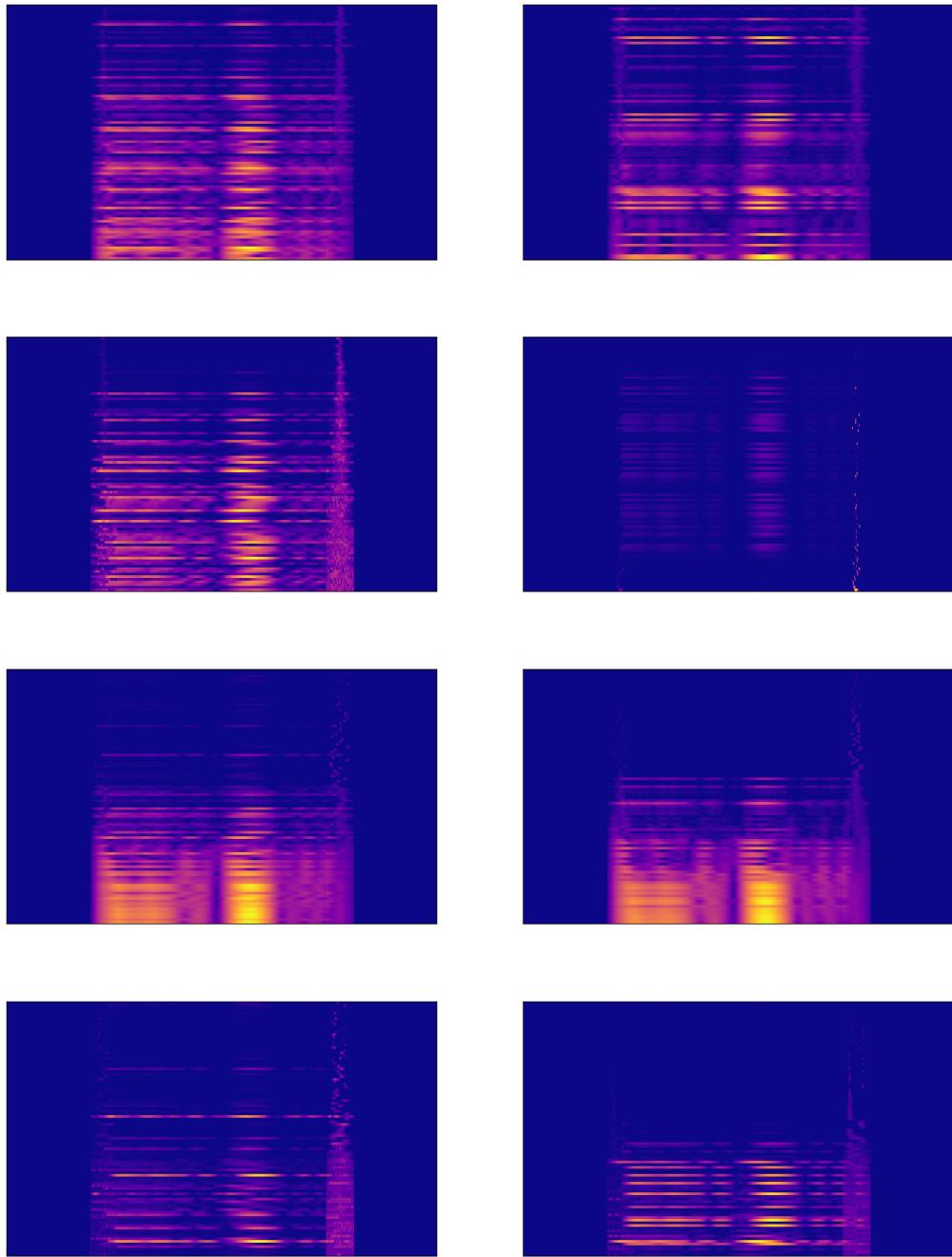


Figure 19. Learnable Group Transform - Visualisation of a sample belonging to class 5, where for each row (*left*) at the initialization and (*right*) after learning. Each row displays a different setting: (*from top to bottom*): LGT, nLGT, cLGT, cnLGT.

D. Proofs

Proposition 2. ρ_{inc} is a group representation of G_{inc} on $\mathbb{L}_2(\mathbb{R})$.

D.1. Proof Representation Property

Proof. Let $g, g' \in \mathbf{G}_{\text{inc}}$, then

$$\begin{aligned} [\rho_{\text{inc}}(g' \circledast g)\psi](t) &= \psi((g' \circledast g)(t)) \\ &= \psi(g'(g(t))) \end{aligned}$$

and,

$$\begin{aligned} [\rho_{\text{inc}}(g')\rho_{\text{inc}}(g)\psi](t) &= [\rho_{\text{inc}}(g')\psi](g(t)) \\ &= \psi(g'(g(t))) \end{aligned}$$

which verifies the homogeneity property. The linearity is implied by,

$$[\rho_{\text{inc}}(g)(\kappa\psi_1 + \psi_2)](t) = (\kappa\psi_1 + \psi_2)(g(t)) = \kappa\psi_1(g(t)) + \psi_2(g(t)), \forall t \in \mathbb{R}.$$

where $\psi_1, \psi_2 \in \mathbb{L}_2(\mathbb{R})$ and $\kappa \in \mathbb{R}$. It is in fact a Koopman operator (Korda & Mezić, 2018). \square

D.2. Proof Proposition 1

Proof. Let $\tau \in \mathbb{R}$ and $g, g' \in \mathbf{G}_{\text{inc}}$,

$$\begin{aligned} \mathcal{W}[\rho_{\text{inc}}(g')s_i, \psi](g, \tau) &= \langle \rho_{\text{inc}}(g')s_i, \rho_{\text{inc}}(g)\overline{\psi_\tau} \rangle \\ &= \langle s_i, \rho_{\text{inc}}(g')^{-1}\rho_{\text{inc}}(g)\overline{\psi_\tau} \rangle \\ &= \langle s_i, \rho_{\text{inc}}(g'^{-1})\rho_{\text{inc}}(g)\overline{\psi_\tau} \rangle \\ &= \langle s_i, \rho_{\text{inc}}(g'^{-1} \odot g)\overline{\psi_\tau} \rangle \\ &= \mathcal{W}[s_i, \psi](g'^{-1} \odot g, \tau), \end{aligned}$$

where $\overline{\psi_\tau}(t) = \psi_\tau(-t)$ denotes the filter ψ centered at position τ . Then, there is no guarantee that this can be extrapolated to all $\tau \in \mathbb{R}$, i.e., in the convolution case, except in the affine case where the global transformation matches the iteration of a local one.

\square