

# Aliasing in Convnets: A Frame Theoretic Perspective

Daniel Haider<sup>a</sup>, Vincent Lostanlen<sup>b</sup>, Martin Ehler<sup>c</sup>, Nicki Holighaus<sup>a</sup>, and Peter Balazs<sup>a,d</sup>

<sup>a</sup>*Acoustics Research Institute, Austrian Academy of Sciences, Vienna, 1010, AUT*

<sup>b</sup>*Nantes Université, École Centrale Nantes, CNRS, LS2N, Nantes, 44000, FRA*

<sup>c</sup>*Department of Mathematics, University of Vienna, Vienna, 1090, AUT*

<sup>d</sup>*Interdisciplinary Transformation University Austria (IT:U), Linz, 4040, AUT*

---

## Abstract

Using a stride in a convolutional layer inherently introduces aliasing, which has implications for numerical stability and statistical generalization. While techniques such as the parametrizations via paraunitary systems have been used to promote orthogonal convolution and thus ensure Parseval stability, a general analysis of aliasing and its effects on the stability has not been done in this context. In this article, we adapt a frame theoretic approach to describe aliasing in convolutional layers with 1D kernels, leading to practical estimates for stability bounds and characterizations of Parseval stability, that are tailored to take short kernel sizes into account. From this, we derive two computationally very efficient optimization objectives that promote Parseval stability via systematically suppressing aliasing. Finally, for layers with random kernels, we derive closed-form expressions for the expected value and variance of the terms that describe the aliasing effects, revealing fundamental insights into the aliasing behavior at initialization.

*Keywords:* Aliasing, Parseval stability, strided convolution, neural networks, filterbanks, random filters

---

## 1. Introduction

In signal processing, *aliasing* describes the spectral overlap of different frequencies when sampling continuous signals, or applying a downsampling operation to a discrete signal. Fundamentally, aliasing relates to the fact that the samples of pure frequencies, i.e., linear-phase complex exponentials, are indistinguishable if they differ by a multiple of the sampling rate.

Besides analog-to-digital conversion, aliasing is a major concern in *multi-rate signal processing* [1], where convolution operators, usually termed as *filters*, are composed with downsampling (and upsampling) operators to form *multiple-input multiple-output (MIMO)* filterbanks. In the machine learning community, the composition of convolution and downsampling is widely referred to as *strided convolution* and forms the fundamental building block of *convolutional neural networks*: The linear part of any convolutional layer is functionally equivalent to a MIMO filterbank.

Filterbanks, in particular *invertible (or perfect reconstruction)* filterbanks have been intensively studied since the 1980s, with many authors contributing to a large corpus of literature. To date, Vaidyanathan’s book [1] probably remains the most comprehensive reference. His works [2, 3] and works of Vetterli et al. [4, 5, 6, 7] may serve as representative examples of the crucial contributions of a large research community to our understanding of perfect reconstruction (PR) filterbanks. In most of these works, the so-called *polyphase representation* was primarily used for the study of the PR property. In particular, the polyphase representation leads to a convenient characterization of critically decimated PR filterbanks as exactly those filterbanks with a polyphase representation given by a *paraunitary matrix*, i.e., a matrix-valued function  $R: \mathbb{C} \rightarrow \mathbb{C}^{M \times M}$  which satisfies  $R(z)\overline{R^*(z^{-1})} = I_M$  for all  $z \in \mathbb{C}$ . If the downsampling factor  $d$  is smaller than the number of channels  $M$ , then  $R(z) \in \mathbb{C}^{M \times d}$  is a matrix with orthonormal columns. Alternatively, PR filterbanks can be studied using their *alias component matrix*, which is likewise a matrix-valued function  $A: \mathbb{C} \rightarrow \mathbb{C}^{M \times d}$  each column of which is the frequency-domain form of one alias component of the filterbank. A classic equivalence between the polyphase and alias component matrix, see [1, Eq. (5.5.8)], shows that, in fact,  $R(z)$  is paraunitary, if and only if  $A(z)$  is paraunitary (if  $d = M$ ), resp. orthonormal.

In the context of neural networks, these results have been used to characterize linear convolutional layers through paraunitary systems [8, 9], and to establish a relation to the Cayley transform [10], both leading to efficient parametrization and design strategies. However, a filterbank may be stable and invertible, even if it is not PR. In that case, the alias component matrix can be used to explicitly describe the error introduced by aliasing by means of the derived *Walnut representation* [11], thereby providing an intuitive description of a more general class of stable, invertible filterbanks. This is the perspective that we consider here, using tools (and terminology) from frame theory to describe and quantify aliasing in convolutional

layers with 1D kernels [12]. The core idea is the following. Let  $\Theta$  be the operator that applies an undecimated filterbank and  $\mathcal{F}$  the (unitary) Fourier transform. Then  $\hat{S} = \mathcal{F}\Theta^*\Theta\mathcal{F}^*$  is a diagonal operator. After introducing decimation, this is no longer true, and  $\hat{S}$  contains additional non-zero side diagonals, which describe the effect of aliasing on the operator level by means of frequency correlation among the filters:  $\hat{S}$  can be decomposed into a sum of (weighted) translation operators, the *Walnut representation* of  $\hat{S}$ . First described by Walnut [13] for Gabor systems this decomposition extends to arbitrary filterbanks with uniform [14] and non-uniform decimation [15, 16]. It was further shown that quantifying the terms in the representation allows us to estimate stability bounds  $A, B$  in the sense of

$$A\|x\|^2 \leq \|\Theta x\|^2 \leq B\|x\|^2 \quad (1)$$

for all inputs  $x$ . In fact, the existence of  $0 < A \leq B < \infty$  such that the above holds characterizes the invertibility and stability of the filterbank. In this case, we call it a *frame*. Stability in the sense of (1) plays an important role in neural networks, where layer transforms with poor bounds  $A, B$  can cause vanishing or exploding gradients, adversely affect training efficiency and generalization, and create vulnerabilities to adversarial attacks [17, 18, 19]. To avoid instabilities, either a 1-Lipschitz constraint ( $B \leq 1$ ) [20, 21, 22], or even Parseval stability ( $A = B = 1$ ) [23, 24, 25] is imposed in the layers. In our work, we adapt the frame bound estimate from the  $\ell^2(\mathbb{Z})$  setting in [11] to be applicable for convolutional neural networks with stride, and derive an extension that explicitly takes the kernel sizes of the convolutional layers into account. From our results we can derive two optimization objectives that promote tightness ( $A = B$ ) by systematically suppressing aliasing. Directly formulated in the Fourier domain, they are computationally as cheap as the forward pass of a layer. Overall, we provide a new perspective on the numerical stability of strided convolutional layers through aliasing, and indicate new approaches towards regularization via aliasing suppression.

The structure of the paper is as following. In Section 2 we prepare the overall setting for describing single-channel strided convolutional layers and their stability via filterbanks and g-frames in  $\mathbb{C}^L$ . Section 3 introduces the Walnut representation for filterbanks in  $\mathbb{C}^L$  and presents estimates for stability bounds and characterizing conditions for Parseval stability, only based on the aliasing terms. We show that the conditions can be used as optimization

objectives to systematically suppress aliasing and thereby promote Parseval stability as a direct consequence. Section 4 is dedicated to the expected value and variance of the aliasing terms for randomly initialized layers. In Section 5, we describe how all the results in the paper can be applied for multi-channel layers, non-uniform strides, and dilated convolution. Finally, Section 6 contains a concluding discussion, open problems, and an outlook.

## 2. Filterbanks and Generalized Frames

### 2.1. Basics and notation

Throughout this paper, we consider finite complex-valued signals  $x \in \mathbb{C}^L$  and filters  $w \in \mathbb{C}^L$ , both assumed to have periodic extension, i.e.,  $x[n+kL] = x[n]$  for any  $k \in \mathbb{N}$ . This provides a setting where the convolution of a signal with a filter, given by  $(x*w)[n] = \sum_{\ell=0}^{L-1} x[\ell]w[n-\ell]$  for  $n = 0, \dots, L-1$  comes with circular boundary conditions. By time-reversing the filter, denoted by  $(\mathbf{R}w)[n] = w[-n]$  we obtain cross-correlation,  $(x \star w)[n] = (x * \mathbf{R}w)[n] = \sum_{\ell=0}^{L-1} x[\ell]w[n+\ell]$  which is often used instead of convolution. Let the (unitary) discrete Fourier transform (DFT) of length  $L$  be given by

$$\hat{x}[k] = (Fx)[k] = \frac{1}{\sqrt{L}} \sum_{\ell=0}^{L-1} x[\ell]e^{-2\pi i k \ell / L}, \quad (2)$$

where  $F$  is the corresponding DFT matrix. From the circular boundary conditions we have that  $\widehat{x * w} = \hat{x} \odot \hat{w}$ , where  $\odot$  denotes the point-wise product of vectors. For a quadratic matrix  $H \in \mathbb{C}^{L \times L}$ , we propose to call  $\hat{H} = FHF^*$  the matrix Fourier transform of  $H$ . We denote the translation by  $a$  samples by  $(\mathbf{T}_a x)[n] = x[n-a]$ . For applying decimation on a vector by a factor  $d \geq 1$  we write  $(x \downarrow_d)[n] = x[dn]$  and upsampling is denoted by  $x \uparrow_d[n] = x[n/d]$  if  $n/d \in \mathbb{N}$  and 0 otherwise.

### 2.2. Strided convolution and filterbanks

Applying convolution, followed by decimation is equivalent to apply convolution with a hop size, also known as *strided convolution*,

$$(x * w) \downarrow_d [n] = \sum_{\ell=0}^{L-1} x[\ell]w[dn - \ell]. \quad (3)$$

In this context, the decimation factor  $d$  is called the stride. We will assume that  $L/d \in \mathbb{N}$  or, if necessary, extend the signal length to the next multiple of  $d$ , given by  $d\lceil \frac{L}{d} \rceil$ .

A linear single-channel *strided convolutional layer* for  $\mathbb{C}^L$  is a *filterbank* consisting of filters  $(w_j)_{j=1}^M$  in  $\mathbb{C}^L$  that decompose a signal  $x \in \mathbb{C}^L$  into  $M$  channels by strided convolution, i.e., for every  $j = 1, \dots, M$ , the output is  $y_j = (x * w_j) \downarrow_d \in \mathbb{C}^{L/d}$ . We will use the notation

$$\{(w_j)_{j=1}^M, \downarrow_d\} \tag{4}$$

to denote such a filterbank. In practice, the filters in a convolutional layer have a customizable, preferably small, number of trainable parameters, much smaller than the signal length  $L$  (common choices vary between 3 and 32 [26, 27, 28]). We denote this number by  $L_K$ . In the filterbank paradigm this can be interpreted as the  $w_j$  having non-zero entries only at the first  $L_K \ll L$  coordinates. In this spirit, we will refer to the full-length vectors  $w_j \in \mathbb{C}^L$  as the *filters* and to the non-zero part containing the  $L_K$  (trainable) parameters as the *kernels* of  $w_j$ .

To maintain clarity and conciseness, all results in this work will be formulated for the single-channel case. In Section 5, we discuss how they can be extended to multi-channel layers, non-uniform strides, and dilated convolution.

### 2.3. Generalized frames

To study the numerical stability of a filterbank we use a formalism that treats each of the strided convolutions as a separate linear operator. The corresponding notion is that of a generalized frame, or g-frame for short [29].

**Definition 2.1.** *A g-frame is a collection  $(T_j)_{j=1}^M$  of linear operators  $T_j$  which satisfies that there are  $A, B > 0$  such that*

$$A\|x\|^2 \leq \sum_{j=1}^M \|T_j x\|^2 \leq B\|x\|^2 \tag{5}$$

for all  $x \in \mathbb{C}^L$ . The constants  $A, B$  are called the *frame bounds*. A g-frame is said to be *tight* if  $A = B$  and *Parseval* if  $A = B = 1$ .

For the transform  $\Theta$ , given by  $x \mapsto (T_j x)_{j=1}^M$  (also called the analysis operator), the inequality in (5) represents a stability condition which can be interpreted as relaxed variant of the energy preservation condition for Parseval stability of orthogonal operators. Moreover,  $\Theta$  is injective if and only if  $(T_j)_{j=1}^M$  is a g-frame. The most important operator in this paper will be the *frame operator*, defined by

$$S = \Theta^* \Theta : \mathbb{C}^L \rightarrow \mathbb{C}^L \quad (6)$$

$$x \mapsto \sum_{j=1}^M T_j^* T_j x, \quad (7)$$

where  $T_j^*$  is the adjoint operator of  $T_j$  in the sense of  $\langle T_j x, y \rangle_{\mathbb{C}^{L'}} = \langle x, T_j^* y \rangle_{\mathbb{C}^L}$  for  $x \in \mathbb{C}^L$  and  $y \in \mathbb{C}^{L'}$ . In the case where  $(T_j)_{j=1}^M$  is a g-frame for  $\mathbb{C}^L$ , then  $S$  is an invertible and self-adjoint  $L \times L$  matrix whose smallest and largest eigenvalues are always positive and give the optimal values among all possible frame bounds  $A, B$  in (5). Also in terms of tightness ( $A = B$ ), the frame operator is of major importance.

**Lemma 2.2.** *The following are equivalent.*

- (i)  $(T_j)_{j=1}^M$  is a tight g-frame for  $\mathbb{C}^L$ .
- (ii) There is  $A > 0$  such that  $\|\Theta x\|^2 = A \|x\|^2$  for all  $x \in \mathbb{C}^L$ .
- (iii) There is  $A > 0$  such that  $S = \hat{S} = A \cdot I_L$ .

For a filterbank  $\{(w_j)_{j=1}^M, \downarrow_d\}$  in  $\mathbb{C}^L$ , the elements of the associated g-frame and their adjoints are given by

$$\begin{aligned} T_j : \mathbb{C}^L &\rightarrow \mathbb{C}^{L/d} & T_j^* : \mathbb{C}^{L/d} &\rightarrow \mathbb{C}^L \\ x &\mapsto (x * w_j) \downarrow_d, & y &\mapsto y \uparrow_d * \overline{\mathbf{R}w_j}, \end{aligned} \quad (8)$$

and the frame operator applies as  $Sx = \sum_{j=1}^M (((x * w_j) \downarrow_d) \uparrow_d) * \overline{\mathbf{R}w_j}$ . In many applications, it is crucial to use a filterbank that is a tight g-frame. The first main advantage is that we can perfectly reconstruct any  $x$  from its filterbank coefficients using the same filters. This is reflected by  $\Theta^* \Theta x = Sx = Ax$ . Secondly, energy preservation up to a constant in (ii) provides robustness against perturbations [30]. In the neural network context, the Parseval case ( $A = 1$ ) is equivalent to what is referred to as *orthogonal convolution*, and has been leveraged to stabilize the gradient flow, improve training efficiency and generalization, and increase the robustness against adversarial attacks.

#### 2.4. Extension to $\ell^2(\mathbb{Z})$

For proving results that are independent of the signal length  $L$ , we will use that a filterbank in  $\mathbb{C}^L$  is completely determined by its extension to  $\ell^2(\mathbb{Z})$ . We use calligraphic symbols for all operators here. Analogously to the definition of a g-frame for  $\mathbb{C}^L$ , a filterbank with filters  $g_j \in \ell^2(\mathbb{Z})$  and decimation factor  $d$  is a g-frame for  $\ell^2(\mathbb{Z})$  if the frame inequalities from Eq. (5) hold with the corresponding strided convolution and norms in  $\ell^2(\mathbb{Z})$ . It is a tight g-frame if  $A = B$ , which happens if and only if the associated frame operator  $\mathcal{S} : \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$  satisfies that  $\mathcal{S} = A \cdot I_{\ell^2(\mathbb{Z})}$ . For  $g \in \ell^2(\mathbb{Z})$ , let the discrete-time Fourier transform (DTFT) from  $\ell^2(\mathbb{Z})$  to  $L^2([0, 1])$  be given by

$$\hat{g}(\xi) = (\mathcal{F}g)(\xi) = \sum_{\ell \in \mathbb{Z}} g[\ell] e^{-2\pi i \xi \ell}. \quad (9)$$

Analogously, tightness is characterized by  $\hat{\mathcal{S}} = \mathcal{F}\mathcal{S}\mathcal{F}^* = A \cdot I_{L^2([0,1])}$ .

### 3. Stability Through Aliasing Suppression

The aliasing effects in a filterbank that are introduced due to strided convolution can be expressed directly in terms of the matrix Fourier transform of the associated frame operator,  $\hat{\mathcal{S}}$ . In particular, it can be written as a sum of terms that describe the frequency correlation among the filters. The so-called *Walnut representation* of  $\hat{\mathcal{S}}$  makes this structure visible.

#### 3.1. Better call Walnut

The following proposition is a finite dimensional version of Prop. 4 in [11], which was originally formulated for filterbanks in  $\ell^2(\mathbb{Z})$ .

**Proposition 3.1** (Walnut representation for filterbanks in  $\mathbb{C}^L$ ).

Let  $\{(w_j)_{j=1}^M, \downarrow_d\}$  be a filterbank in  $\mathbb{C}^L$  then

$$\hat{\mathcal{S}}\hat{x} = \sum_{n=0}^{d-1} G_n \odot \mathbf{T}_{n\frac{L}{d}}\hat{x}, \quad (10)$$

where

$$G_n = d^{-1} \sum_{j=1}^M \hat{w}_j \odot \overline{\mathbf{T}_{n\frac{L}{d}}\hat{w}_j} \quad (11)$$

are called the *aliasing terms of the filterbank*.

Although the proof is essentially the same as the one for [11, Prop. 4], we include one using the terminology used in this paper.

*Proof.* For  $x \in \mathbb{C}^L$ , we use the following two facts. First, the Fourier transform after down and upsampling is given by

$$(F((x \downarrow_d) \uparrow_d))[k] = \sum_{n=0}^{L/d-1} x[nd] e^{-2\pi i k n d / L}. \quad (12)$$

Second, due to the periodic extension we can interpret  $x$  as  $L$ -periodic sequence, and Poisson's summation formula gives for every  $\ell = 0, \dots, L-1$  that

$$\sum_{k=0}^{L/d-1} x[kd] e^{-2\pi i k \ell / L} = d^{-1} \sum_{n=0}^{d-1} \hat{x}[\ell - n \frac{L}{d}]. \quad (13)$$

For every  $\ell = 0, \dots, L-1$ , we derive

$$\hat{S} \hat{x}[\ell] = F S x[\ell] = F \left( \sum_{j=1}^M (((x * w_j) \downarrow_d) \uparrow_d) * \overline{\mathbf{R} w_j} \right) [\ell] \quad (14)$$

$$= \sum_{j=1}^M F(((x * w_j) \downarrow_d) \uparrow_d) [\ell] \cdot F(\overline{\mathbf{R} w_j}) [\ell] \quad (15)$$

$$= \sum_{j=1}^M \sum_{k=0}^{L/d-1} (x * w_j)[kd] e^{-2\pi i k \ell / L} \cdot \overline{\hat{w}_j[\ell]} \quad \text{by (12)} \quad (16)$$

$$= \sum_{j=1}^M \sum_{k=0}^{L/d-1} F^*(\hat{x} \odot \hat{w}_j)[kd] e^{-2\pi i k \ell / L} \cdot \overline{\hat{w}_j[\ell]} \quad (17)$$

$$= \sum_{j=1}^M d^{-1} \sum_{n=0}^{d-1} \hat{x}[\ell - n \frac{L}{d}] \cdot \hat{w}_j[\ell - n \frac{L}{d}] \cdot \overline{\hat{w}_j[\ell]} \quad \text{by (13)}. \quad (18)$$

Exchanging the complex conjugations of the filters and isolating the aliasing terms yields the final expression.  $\square$

In other words,  $\hat{S}$  is a band-diagonal matrix with  $d$  bands that describe the frequency correlation among the filters. The entries of  $\hat{S}$  are given by

$$\hat{S}[k, \ell] = \begin{cases} G_{\lfloor \frac{\ell d}{L} \rfloor}[k \bmod L/d] & \text{for } k - \ell \equiv 0 \pmod{L/d} \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$



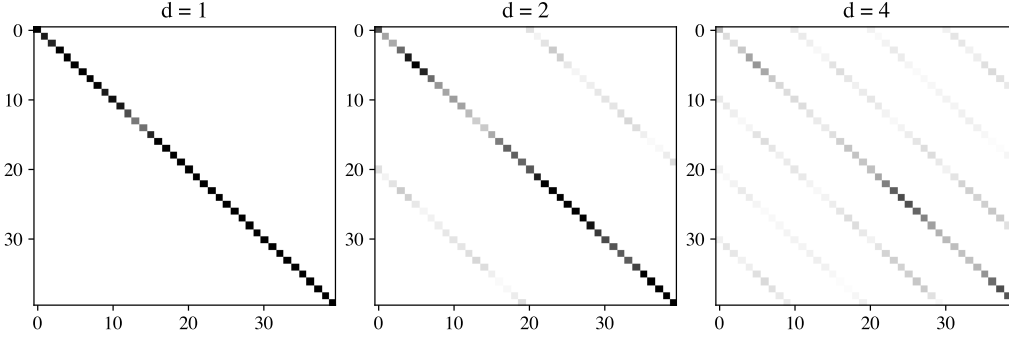


Figure 1: The matrix  $\hat{S} = FSF^*$  for a filterbank for  $\mathbb{C}^{40}$  with  $M = 4$  i.i.d. Gaussian complex random kernels of size  $L_K = 8$  and different strides  $d = 1, 2, 4$ .

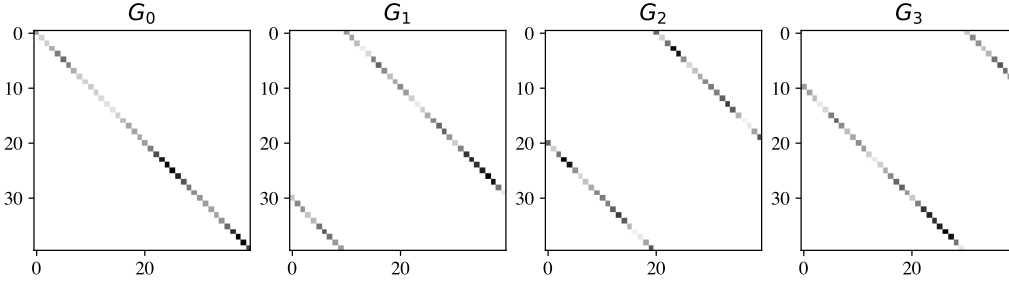


Figure 2: The aliasing terms  $G_0, G_1, G_2, G_3$  isolated as the side diagonals of  $\hat{S}$  for the right matrix ( $d = 4$ ) from above.

Figure 1 and 2 illustrate the structure of  $\hat{S}$  for a filterbank with random kernels and different decimation factors. Setting  $d = 1$ , we have that  $\hat{S} = \text{diag} \left( \sum_{j=1}^M |\hat{w}_j|^2 \right) = \text{diag} (G_0)$ . In particular,  $G_n \equiv 0$  for all  $n > 0$ . Hence, the stability of an undecimated filterbank is determined only by the filterbank response  $G_0$ . In particular, it is a tight g-frame if and only if the response is perfectly flat. The Walnut representation allows us to generalize these statements to  $d > 1$ . By bounding the aliasing terms in different ways, we get frame bound estimates, and see that tightness is equivalent to a perfectly flat response, together with vanishing aliasing terms.

### 3.2. Frame bound estimates and tightness characterizations

For the estimates we are using a result that was originally formulated for filterbanks in  $\ell^2(\mathbb{Z})$  in [11]. In this setting, the aliasing terms are continuous

functions on the torus,  $\mathcal{G}_n \in L^2([0, 1))$ , given by

$$\mathcal{G}_n(\xi) = \sum_{j=0}^M d^{-1} \hat{g}_j(\xi) \overline{\hat{g}_j(\xi - nd^{-1})}. \quad (20)$$

The frame bound estimates are obtained through diagonal dominance.

**Proposition 3.2** (G-frame formulation of [11], Prop. 5). *If there are  $0 < A \leq B < \infty$  with*

$$A \leq \mathcal{G}_0(\xi) \pm \sum_{n=1}^{d-1} |\mathcal{G}_n(\xi)| \leq B \quad (21)$$

*for almost all  $\xi \in [0, 1)$ , then the filterbank with filters  $g_j \in \ell^2(\mathbb{Z})$  and decimation factor  $d$  is a  $g$ -frame for  $\ell^2(\mathbb{Z})$  with frame bounds  $A, B$ .*

In the following theorem, we use Prop. 3.2 to derive statements about the stability of a filterbank in  $\mathbb{C}^L$ . We use the notation  $\|x\|_\infty = \max_n |x[n]|$  and the fact that the average value of any  $x \in \mathbb{C}^L$  is given by  $\frac{1}{\sqrt{L}} \hat{x}[0] = \frac{1}{L} \sum_{\ell=0}^{L-1} x[\ell]$ .

**Theorem 3.3.** *Let  $\{(w_j)_{j=1}^M, \downarrow_d\}$  be a filterbank for  $\mathbb{C}^L$  and let  $G_n$  be the associated aliasing terms. The following holds.*

(i) *Let*

$$A = \min_k (G_0[k] - \sum_{n=1}^{d-1} |G_n[k]|) \quad (22)$$

$$B = \max_k (G_0[k] + \sum_{n=1}^{d-1} |G_n[k]|). \quad (23)$$

*If  $A > 0$  then the filterbank is a  $g$ -frame for  $\mathbb{C}^L$  and  $A, B$  are frame bounds.*

(ii) *The filterbank is a tight  $g$ -frame for  $\mathbb{C}^L$  if and only if  $G_0 \equiv A$  for some  $A > 0$  and  $G_n \equiv 0$  for every  $n = 1, \dots, d-1$ . In particular,  $A = \frac{1}{\sqrt{L}} \hat{G}_0[0]$ .*

(iii) If the filterbank is a  $g$ -frame and  $S$  is the frame operator then

$$\|S - I_L\| \leq \|G_0 - \mathbf{1}_L\|_\infty + \sum_{n=1}^{d-1} \|G_n\|_\infty, \quad (24)$$

where  $\mathbf{1}_L = (1, \dots, 1) \in \mathbb{C}^L$ . Moreover, for  $A, B$  as in (ii) we have

$$\|S - \frac{1}{\sqrt{L}}\hat{G}_0[0] \cdot I_L\| \leq \max(|B - \frac{1}{\sqrt{L}}\hat{G}_0[0]|, |A - \frac{1}{\sqrt{L}}\hat{G}_0[0]|). \quad (25)$$

*Proof.* The frame bound estimates in (i) follow directly from the  $\ell^2(\mathbb{Z})$  version in Prop. 3.2. Point (ii) follows from the Walnut representation (Prop. 3.1) and that  $\frac{1}{\sqrt{L}}\hat{G}_0[0]$  is the average value of  $G_0$ . For the first estimate in (iii) we use the Walnut representation and the triangle inequality to get

$$\|(S - I_L)x\| = \|(\hat{S} - I_L)\hat{x}\| \leq \left( \|G_0 - \mathbf{1}_L\|_\infty + \sum_{n=1}^{d-1} \|G_n\|_\infty \right) \|\hat{x}\|. \quad (26)$$

For the second estimate we use that

$$\|S - \frac{1}{\sqrt{L}}\hat{G}_0[0] \cdot I_L\| = \max(|A^* - \frac{1}{\sqrt{L}}\hat{G}_0[0]|, |B^* - \frac{1}{\sqrt{L}}\hat{G}_0[0]|), \quad (27)$$

where  $A^*, B^*$  are the optimal frame bounds of  $S$ . Let  $A, B$  be as in (ii) then  $A \leq A^* \leq B^* \leq B$  holds. Since  $A \leq \min_k G_0[k] \leq \frac{1}{\sqrt{L}}\hat{G}_0[0]$ , replacing  $A^*, B^*$  by  $A, B$  in (27) remains true, and the claim follows.  $\square$

These results naturally relate aliasing to the stability of a filterbank by quantifying how reduced aliasing leads to improved frame bounds. In particular, they show that tightness is achieved precisely when aliasing is completely absent. However, the results do not take the kernel sizes  $L_K$  of the filters into account, which limits their practicality for long or varying signal lengths. In the following, we present a length-independent version of Theorem 3.3, tailored to the structural constraints of convolutional layers. The conditions that we derive are only required to hold for the minimal signal length of  $L = d \lceil \frac{2L_K - 1}{d} \rceil$ , and maintain valid for all longer signal lengths.

For the proof, we will use the classic result that shifted Dirichlet kernels are an orthonormal basis for the space of trigonometric polynomials.

**Lemma 3.4** ([31], Thm. 5.24). *Let  $p$  be a trigonometric polynomial of degree  $\leq N$  and  $D_n(t) = \sum_{|\ell| \leq n} e^{2\pi i \ell t}$  denote the  $n$ -th Dirichlet kernel. For any  $N' \geq N$ ,  $p$  can be written as the semi-discrete convolution*

$$p(t) = \frac{1}{\sqrt{2N'+1}} \sum_{n=0}^{2N'} p\left(\frac{n}{2N'+1}\right) D_{N'}\left(t - \frac{n}{2N'+1}\right). \quad (28)$$

Moreover, the Fourier coefficients of  $p$  coincide with the DFT coefficients of the vector  $(p(0), p(\frac{1}{2N'+1}), \dots, p(\frac{2N'}{2N'+1}))$ .

The following theorem represents the main result of this paper.

**Theorem 3.5.** *Let  $\{(w_j)_{j=1}^M, \downarrow_d\}$  with kernel size  $L_K$  be a filterbank for  $\mathbb{C}^L$ , where  $L = d \lceil \frac{2L_K-1}{d} \rceil$ . Let  $G_n$  be the associated aliasing terms (of length  $L$ ). For all  $L' \geq L$  the following holds.*

(i) *Let*

$$A = \frac{2}{\sqrt{L}} \hat{G}_0[0] - \sum_{n=0}^{d-1} \|\hat{G}_n\|_1 \quad (29)$$

$$B = \sum_{n=0}^{d-1} \|\hat{G}_n\|_1. \quad (30)$$

*If  $A > 0$  then the filterbank is a  $g$ -frame for  $\mathbb{C}^{L'}$  and  $A, B$  are frame bounds.*

(ii) *The filterbank is a tight  $g$ -frame for  $\mathbb{C}^{L'}$  if and only if  $G_0 \equiv A$  for some  $A > 0$  and  $G_n \equiv 0$  for every  $n = 1, \dots, d-1$ . In particular,  $A = \frac{1}{\sqrt{L}} \hat{G}_0[0]$ .*

(iii) *If the filterbank is a  $g$ -frame for  $\mathbb{C}^L$  and  $S$  the frame operator for  $\mathbb{C}^{L'}$  then*

$$\|S - \frac{1}{\sqrt{L}} \hat{G}_0[0] \cdot I_{L'}\| \leq \left| \frac{1}{\sqrt{L}} \hat{G}_0[0] - \sum_{n=0}^{d-1} \|\hat{G}_n\|_1 \right|. \quad (31)$$

*Proof.* Let us assume that the filters  $w_j$  have kernel size  $L_K = 2N + 1$  for simplicity. Now let  $g_j \in \ell^2(\mathbb{Z})$  be the filters that coincide with the kernels

of the  $w_j$  on  $[-N, N]$ . The  $\hat{g}_j$  are trigonometric polynomials of degree  $\leq N$ , hence, the summands in the associated aliasing terms  $\mathcal{G}_n$  are all of the form

$$h_j(\xi) = \hat{g}_j(\xi) \cdot \overline{\hat{g}_j(\xi - \omega)},$$

with appropriate shifts  $\omega$ . Since the  $h_j$  are trigonometric polynomials of degree  $\leq 2N$ , every  $\mathcal{G}_n$  is likewise a trigonometric polynomial of degree  $\leq 2N$ . The core argument for this proof is now provided by Lemma 3.4, giving us for all  $L' \geq 2N$  and  $n = 0, \dots, d-1$ , that

$$\mathcal{G}_n(\xi) = \frac{1}{\sqrt{2L'+1}} \sum_{\ell=0}^{2L'} \mathcal{G}_n\left(\frac{\ell}{2L'+1}\right) D_{L'}\left(\xi - \frac{\ell}{2L'+1}\right), \quad (32)$$

and that the (non-zero) Fourier coefficients of  $\mathcal{G}_n$  are given by  $\hat{G}_n$ . Assuming that  $L_K = 2N + 1$ , the minimum number of samples needed is  $L = 4N + 1 = 2L_K - 1$ . Taking the divisibility condition for the decimation factors into account gives us  $d \lceil \frac{2L_K - 1}{d} \rceil$  as smallest admissible signal length. In other words, the non-zero Fourier coefficients of the aliasing terms for the  $\ell^2(\mathbb{Z})$  extension of the filterbank, hence, also those for  $\mathbb{C}^{L'}$  where  $L' \geq L$  are already completely determined by  $\hat{G}_n$  (of length  $L$ ).

To prove (i), we approximate the estimates from Lemma 3.2 using that

$$\|\mathcal{G}_n\|_{L^\infty([0,1])} \leq \|\hat{\mathcal{G}}_n\|_{L^1([0,1])} = \|\hat{G}_n\|_1. \quad (33)$$

For the upper bound, we obtain

$$\sup_{\xi \in [0,1]} \left( \mathcal{G}_0(\xi) + \sum_{n=1}^{d-1} |\mathcal{G}_n(\xi)| \right) = \sup_{\xi \in [0,1]} \left( \sum_{n=0}^{d-1} |\mathcal{G}_n(\xi)| \right) \quad (34)$$

$$\leq \sum_{n=0}^{d-1} \|\mathcal{G}_n\|_{L^\infty([0,1])} \leq \sum_{n=0}^{d-1} \|\hat{G}_n\|_1. \quad (35)$$

For the lower bound, we use that  $\|\hat{G}_0\|_1 - \frac{1}{\sqrt{L}} \hat{G}_0[0]$  bounds the maximal

deviation in  $\mathcal{G}_0$ . We get

$$\inf_{\xi \in [0,1]} \left( \mathcal{G}_0(\xi) - \sum_{n=1}^{d-1} |\mathcal{G}_n(\xi)| \right) \geq \inf_{\xi \in [0,1]} (\mathcal{G}_0(\xi)) - \sum_{n=1}^{d-1} \|\mathcal{G}_n\|_{L^\infty([0,1])} \quad (36)$$

$$\geq \frac{1}{\sqrt{L}} \hat{G}_0[0] - \left( \|\hat{G}_0\|_1 - \frac{1}{\sqrt{L}} \hat{G}_0[0] \right) - \sum_{n=1}^{d-1} \|\hat{G}_n\|_1 \quad (37)$$

$$= \frac{2}{\sqrt{L}} \hat{G}_0[0] - \sum_{n=0}^{d-1} \|\hat{G}_n\|_1. \quad (38)$$

To prove (ii), we use that

$$\|\mathcal{G}_n\|_{L^2([0,1])} = \|\hat{G}_n\| = \|G_n\|, \quad (39)$$

where the norms without subscript are Euclidean vector norms. Assuming that  $\{(w_j)_{j=1}^M, \downarrow_d\}$  is tight on  $\mathbb{C}^{L'}$ , then the diagonal entries of  $\hat{S}$  are uniquely determined by the  $L$  samples given by  $G_0$ , hence, they must all be equal. On the other hand, by (39), all the  $G_n$ ,  $n > 0$  must vanish.

For the estimate in (iii) we use that

$$\|S - \frac{1}{\sqrt{L}} \hat{G}_0[0] \cdot I_{L'}\| = \max(|A^* - \frac{1}{\sqrt{L}} \hat{G}_0[0]|, |B^* - \frac{1}{\sqrt{L}} \hat{G}_0[0]|), \quad (40)$$

where  $A^*, B^*$  are the optimal frame bounds of  $S$ . Let  $A, B$  be as in (i) then  $A \leq A^* \leq B^* \leq B$ . We obtain

$$\begin{aligned} \|S - \frac{1}{\sqrt{L}} \hat{G}_0[0] \cdot I_{L'}\| &\leq \max \left( \frac{1}{\sqrt{L}} \hat{G}_0[0] - \sum_{n=0}^{d-1} \|\hat{G}_n\|_1, \sum_{n=0}^{d-1} \|\hat{G}_n\|_1 - \frac{1}{\sqrt{L}} \hat{G}_0[0] \right) \\ &= \left| \frac{1}{\sqrt{L}} \hat{G}_0[0] - \sum_{n=0}^{d-1} \|\hat{G}_n\|_1 \right| \end{aligned} \quad (41)$$

□

While the estimates in Thm. 3.5 are looser than the ones in Thm. 3.3 due to the additional estimation step via the 1-norm, we have been able to detach the frame bounds from the signal length. As a consequence, all the conditions for tightness can be reduced to this minimal signal length of essentially twice the kernel size.

**Corollary 3.6.** Let  $\{(w_j)_{j=1}^M, \downarrow_d\}$  with kernel size  $L_K$  be a filterbank for  $\mathbb{C}^L$ , where  $L = d \lceil \frac{2L_K - 1}{d} \rceil$ . Let  $G_n$  be the associated aliasing terms and  $S$  the frame operator. The following are equivalent.

- (i) The filterbank is tight on  $\mathbb{C}^{L'}$  for all  $L' \geq L$ .
- (ii)  $S = \hat{S} = \frac{1}{\sqrt{L}} \hat{G}_0[0] \cdot I_L$
- (iii)  $\frac{\max_k G_0[k]}{\min_k G_0[k]} + \sum_{n=1}^{d-1} \|G_n\|_\infty = 1$
- (iv)  $\sum_{n=0}^{d-1} \|\hat{G}_n\|_1 = \frac{1}{\sqrt{L}} \hat{G}_0[0]$ .

While (ii) further specifies the standard tightness condition via the actual value of the bound, the conditions in (iii) and (iv) fully leverage the structure of  $\hat{S}$  and only depend on the aliasing terms  $G_n$ . In particular, (iii) concisely characterizes tightness by aliasing cancellation plus response equalization (note that also any other vector norm can be considered instead of  $\|G_n\|_\infty$ ). Condition (iv) further specifies this on the Fourier level of the aliasing terms, thereby, elegantly describing aliasing cancellation in a simple formula. Note that since the  $G_n$  are objects in the Fourier domain, (iv) actually becomes a time-domain condition again. Moreover, since it already involves the tightness bound explicitly, we obtain Parseval stability for free.

### 3.3. Painless hybrid filterbanks

A further generalization of tightness is when  $\hat{S}$  is diagonal matrix, but not necessarily with constant entries. This situation is known as the *painless* case [32], and stability obviously only depends on  $G_0$  anymore, just as in the undecimated case. Classically, this is enforced by designing the filters in a way that they have limited band-pass, defined by the reciprocal decimation factor [33]. A simple example is to use ideal band-pass filters  $\psi_j$  that satisfy

$$\hat{\psi}_j[k] = \begin{cases} 1 & \text{if } k \in [aj, aj + \frac{L}{d} - 1] \\ 0 & \text{else,} \end{cases} \quad (42)$$

for some  $a \leq \frac{L}{d}$ . To make use of this in a strided convolutional layer we can combine the kernels with fixed ideal band-pass filters via pair-wise convolution. Let  $(\psi_j)_{j=1}^M$  and  $(w_j)_{j=1}^M$  be filters. The filterbank given by

$$\{(\psi_j * w_j)_{j=1}^M, \downarrow_d\}. \quad (43)$$

is called a *hybrid filterbank* [34]. It is easy to see that a hybrid filterbank inherits the painless property from the ideal band-pass filters.

**Proposition 3.7.** *Let  $\{(\psi_j * w_j)_{j=1}^M, \downarrow_d\}$  be a hybrid filterbank for  $\mathbb{C}^L$  composed of ideal band-pass filters  $(\psi_j)_{j=1}^M$  satisfying (42) for  $a = \frac{L}{d}$  and filters  $(w_j)_{j=1}^M$ . The hybrid filterbank is painless and*

(i) *a g-frame for  $\mathbb{C}^L$  if and only if  $\min_k |\hat{w}_{\lfloor \frac{kd}{L} \rfloor}[k]| > 0$ .*

(ii) *a tight g-frame for  $\mathbb{C}^L$  if and only if  $|\hat{w}_{\lfloor \frac{kd}{L} \rfloor}[k]| = |\hat{w}_{\lfloor \frac{\ell d}{L} \rfloor}[\ell]| \neq 0$  for all  $k, \ell = 0, \dots, L - 1$ .*

An analog statement can be done for  $a < \frac{L}{d}$ , where the condition depends on the sum of all overlapping  $\hat{w}_j$ .

*Proof.* The aliasing terms for a hybrid filterbank are given by

$$G_n^{(\psi, w)}[k] = \sum_{j=1}^M d^{-1} \hat{\psi}_j[k] \overline{\hat{\psi}_j[k - n\frac{L}{d}]} \hat{w}_j[k] \overline{\hat{w}_j[k - n\frac{L}{d}]}. \quad (44)$$

By (42) we have for all  $j = 1, \dots, M$ ,  $k = 0, \dots, L - 1$ ,

$$\hat{\psi}_j[k] \overline{\hat{\psi}_j[k - n\frac{L}{d}]} = 0. \quad (45)$$

It follows that  $G_n^{(\psi, w)} \equiv 0$  for  $n > 0$ . In the diagonal term, only the non-overlapping squared frequency responses remain such that we get

$$G_0^{(\psi, w)}[k] = d^{-1} |\hat{w}_{\lfloor \frac{kd}{L} \rfloor}[k]|^2.$$

The claims follow directly from this.  $\square$

In other words, we split up the input signal into different frequency regions and process the resulting sub-band signals further, separately. This principle is very useful for audio processing and has been already leveraged for different architectures and applications [34, 35, 36, 37]. In another context, the fixed decomposition into sub-bands has been shown to stabilize the layer in a random setting [38], and to be advantageous for making deconvolution more robust [39].



### 3.4. Numerical experiments on suppressing aliasing

Parseval stability in neural networks has been used against exploding gradients and for enhancing the robustness against noise and adversarial attacks. To maintain this during training, a sequence of works [24, 23, 40, 41] has proposed to minimize

$$\mathcal{L}^S = \|S - I_L\|, \quad (46)$$

(or some variant thereof) for all desired layers, alongside a learning objective. Other possibilities to maintain tightness are to minimize  $\mathcal{L}^\kappa = \frac{B}{A} - 1$  [42] or  $\mathcal{L}^\Theta = \frac{1}{2}(\|\Theta\|_2^2 - \frac{1}{L}\|\Theta\|_F^2)$  [43]. By Corollary 3.6, we can promote tightness also by suppressing aliasing via minimizing

$$\mathcal{L}^G = \frac{\max_k G_0[k]}{\min_k G_0[k]} - 1 + \sum_{n=1}^{d-1} \|G_n\| \quad (47)$$

or

$$\mathcal{L}^{\hat{G}} = \left| \frac{1}{\sqrt{L}} \hat{G}_0[0] - \sum_{n=0}^{d-1} \|\hat{G}_n\|_1 \right|. \quad (48)$$

The above expressions can be computed very efficiently using fast Fourier transforms. The computational complexity for  $\mathcal{L}^G$  is  $\mathcal{O}(ML \log L + MdL)$ , and for  $\mathcal{L}^{\hat{G}}$  it is  $\mathcal{O}((M+d)L \log L + MdL)$ . In comparison, for  $\mathcal{L}^S$  it is  $\mathcal{O}\left(\frac{ML^3}{d} + L^3\right)$ .

As a demonstration, we optimize a linear convolutional layer with  $M$  kernels of size  $L_K$  and stride  $d$  to become approximately Parseval stable for  $\mathbb{C}^L$ ,  $L = d \lceil \frac{2L_K - 1}{d} \rceil$  by minimizing the proposed objectives  $\mathcal{L}^G$  (47) and  $\mathcal{L}^{\hat{G}}$  (48), and compare it to the baseline  $\mathcal{L}^S$  (46). We additionally consider a brute-force method based on the construction of the canonical Parseval g-frame as a reference. Given  $(T_j)_{j=1}^M$ , then  $(T_j S^{-\frac{1}{2}})_{j=1}^M$  is the closest Parseval g-frame in the Frobenius norm among all Parseval g-frames [44]. Note that this corresponds to the filterbank  $\{(S^{-\frac{1}{2}} w_j)_{j=1}^M, \downarrow_d\}$ . Since applying  $S^{-\frac{1}{2}}$  does not maintain the kernel size, we project the filters back onto the first  $L_K$  components and repeat the process. We call this FIR-tightening.

**Definition 3.8** (FIR-tightening). *Let  $\{(w_j^{(0)})_{j=1}^M, \downarrow_d\}$  be a filterbank that forms a g-frame for  $\mathbb{C}^L$  and  $\mathbf{P}_{L_K} : \mathbb{C}^L \rightarrow \mathbb{C}^L$  be the projection operator onto the first*

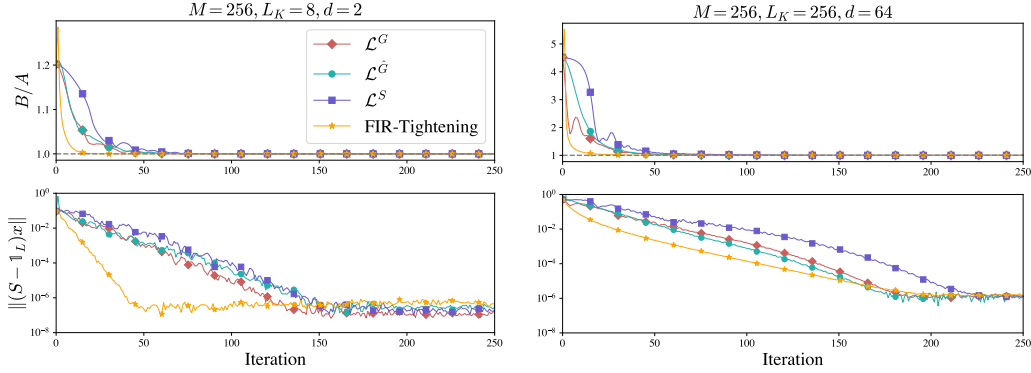


Figure 3: Progress of optimizing randomly initialized strided convolutional layers to be Parseval stable (250 iterations). We compare the three objectives  $\mathcal{L}^G$ ,  $\mathcal{L}^{\hat{G}}$ ,  $\mathcal{L}^S$ , and FIR-tightening. Top: Condition number  $B/A$ . Bottom: Reconstruction error  $\|(S - I_L)x\|$  for random  $x$ , plotted on a log-scale. Left: A common setting for a single-channel layer. Right: The single-channel pendant that corresponds to a multi-channel layer with  $C = 32$  input channels,  $L_K = 8$ , and  $d = 2$ . In general, for any setting where the initialization yields a g-frame, all methods are capable of getting to Parseval stability at machine precision.

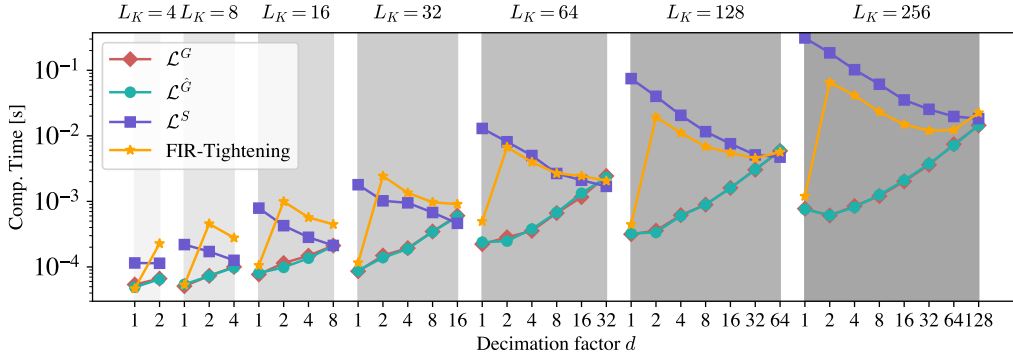


Figure 4: Average computation time for one step on a log scale. We compare different kernel sizes  $L_K$  and decimation factors  $d \leq L_K/2$ . We set  $M = 256$ . Except for  $d = L_K/2$ , the aliasing-based methods  $\mathcal{L}^G$  and  $\mathcal{L}^{\hat{G}}$  are significantly faster than the baseline  $\mathcal{L}^S$  (up to 32 times) and FIR-tightening.

$L_K \leq L$  coordinates. We define the  $n$ -FIR-tightened filterbank  $\{(w_j^{(n)})_{j=1}^M, \downarrow_d\}$  by the filters

$$w_j^{(n)} = \mathbf{P}_{L_K} (S^{(n-1)})^{-\frac{1}{2}} w_j^{(n-1)}, \quad (49)$$

where  $S^{(n-1)}$  is the frame operator for the filterbank  $\{(w_j^{(n-1)})_{j=1}^M, \downarrow_d\}$ .

Clearly, the filters  $w_j^{(n-1)}$  have kernels of size  $L_K$  at any iteration  $n > 0$ . However, we can not guarantee exact Parseval stability due to the projection step. Yet, in many numerical experiments, we observed that FIR-tightening reliably converges to approximately Parseval g-frames in few iterations. It should be noted, however, that this procedure is very invasive due to the projection, and computationally very expansive as it required a matrix inversion in every step.

We initialize the kernels as i.i.d. Gaussian random vectors with zero mean and variance  $(L_K M)^{-1}$ . We use stochastic gradient decent with learning rates between  $10^{-2}$  and  $10^{-4}$  and add an adaptive regulation by scaling the objective with  $(B/A - 1)^{1/10}$  in every iteration. In Figure 3, we show the progress of the conditions numbers  $B/A$  and the reconstruction errors  $\|(S - I_L)x\|$  for random and normalized  $x$  over 250 iterations. All methods were capable of getting to Parseval stability at machine precision ( $B/A \approx 1 + 10^{-7}$  and  $\|(S - I_L)x\| \approx 10^{-7}$ ) in any considered scenario.

When looking at the computation time<sup>1</sup> that the evaluations of the regularizers take, the aliasing-based methods outperform in all relevant settings, especially when the stride is small. In Figure 4 we compare the average speed of computing the different objectives in different settings, and plot the times on a log scale. In very low parameter settings all the methods are approximately equally fast. For any increase in  $M$  or  $L_K$ , the aliasing-based methods take the lead. For the setting  $M = 256, L_K = 64, d = 2$  they are approximately 32 times faster than the baseline.

#### 4. Aliasing at Random Initialization

Initializing neural network weights with random draws from a probability distribution is a standard practice. In convolutional layers, this means that the kernels are random vectors. The statistical behavior of aliasing at this stage can be used to understand how much a feature representation is affected by aliasing before any learning occurs. This insight can guide the design of initialization schemes to mitigate destructive aliasing and potentially enhance learning dynamics.

---

<sup>1</sup>The computations were done on the CPU of a MacBook Pro (M4, 14-core CPU).

#### 4.1. Expected value and variance

We assume that the kernels of the filters  $w_j$  are i.i.d. random vectors of length  $L_K$  and the rest of the  $L - L_K$  filter entries are zero. For the Gaussian or uniform i.i.d. case, the distribution looks like

$$w_j[n] \sim \begin{cases} \mathcal{N}(\mu, \sigma^2) \text{ or } \mathcal{U}_{[a,b]} & \text{if } 0 \leq n \leq L_K - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (50)$$

By the extension with zeros, the  $w_j$  are not independent random vectors. The following theorem shows how the expected value and variance of the aliasing terms of such a filterbank behave. Note that the signal length  $L$  appearing as a factor in all the expression comes from the unitary DFT and would not pop up with another normalization scheme.

**Theorem 4.1.** *Let  $\{(w_j)_{j=1}^M, \downarrow_d\}$  with kernel size  $L_K$  be a filterbank for  $\mathbb{C}^L$ , where the  $w_j$  are distributed as in (50) with zero mean and variance  $\sigma^2$ . The corresponding aliasing terms  $G_n$  are complex random vectors satisfying the following. For any  $0 \leq k \leq L - 1$  the expected value is given by*

$$\mathbb{E}[G_n[k]] = \frac{\sigma^2 M}{Ld} \sum_{\ell=0}^{L_K-1} e^{-2\pi i \ell n/d}. \quad (51)$$

In particular, if  $\frac{L_K}{d} \in \mathbb{N}$ , then

$$\mathbb{E}[G_n[k]] = \begin{cases} \frac{\sigma^2 M L_K}{Ld} & \text{if } n = 0 \\ 0 & \text{if } 1 \leq n \leq d - 1. \end{cases} \quad (52)$$

The variance is given by

$$\mathbb{V}[G_n[k]] = \frac{\sigma^4 M}{L^2 d^2} \left( \frac{\sin^2(L_K \pi (\frac{2k}{L} - \frac{n}{d}))}{\sin^2(\pi (\frac{2k}{L} - \frac{n}{d}))} + L_K^2 \right). \quad (53)$$

Moreover, for any  $0 \leq k \leq L - 1$

$$\frac{\sigma^4 M L_K^2}{L^2 d^2} \leq \mathbb{V}[G_n[k]] \leq \frac{2\sigma^4 M L_K^2}{L^2 d^2}. \quad (54)$$

*Proof.* The proof for both expressions goes by expanding the Fourier transforms of  $\hat{w}_j$  and  $\overline{\mathbf{T}_{nL/d}\hat{w}_j}$  and using the i.i.d. property of the kernels,

$$\text{Cov}(w_j[\ell], w_j[\ell']) = \begin{cases} \sigma^2 \delta[\ell - \ell'] & \text{for } 0 \leq \ell, \ell' \leq T - 1 \\ 0 & \text{for } T \leq \ell, \ell' \leq L - 1. \end{cases} \quad (55)$$

The expression for the mean in (51) is given by

$$\mathbb{E}[G_n[k]] = \frac{1}{Ld} \sum_{j=1}^M \sum_{\ell, \ell'=0}^{L_K-1} e^{-2\pi i \frac{\ell k}{L}} e^{2\pi i \frac{\ell'(k-nL/d)}{L}} \text{Cov}(w_j[\ell], w_j[\ell']) \quad (56)$$

$$= \frac{\sigma^2 M}{Ld} \sum_{\ell=0}^{L_K-1} e^{-2\pi i \ell n/d}. \quad (57)$$

If  $\frac{L_K}{d} \in \mathbb{N}$ , then the exponential terms are roots of unity, which shows (52).

We turn to the variance. Since the terms  $\hat{w}_j[k] \overline{\hat{w}_j[k - n\frac{L}{d}]}$  are independent across  $j$ , we can pull out the sum and get that

$$\mathbb{V}[G_n[k]] = d^{-2} \sum_{j=1}^M \mathbb{E} \left[ \left| \hat{w}_j[k] \overline{\hat{w}_j[k - n\frac{L}{d}]} \right|^2 \right] - \left| \mathbb{E} \left[ \hat{w}_j[k] \overline{\hat{w}_j[k - n\frac{L}{d}]} \right] \right|^2. \quad (58)$$

For the first term, we find that

$$\mathbb{E} \left[ \left| \hat{w}_j[k] \overline{\hat{w}_j[k - n\frac{L}{d}]} \right|^2 \right] \quad (59)$$

$$= \frac{1}{L^2} \sum_{m, m', \ell, \ell'=0}^{L_K-1} e^{-2\pi i \frac{k(m-m')}{L}} e^{-2\pi i \frac{(k-nL/d)(\ell-\ell')}{L}} \mathbb{E}[w_j[m]w_j[\ell]w_j[m']w_j[\ell']]. \quad (60)$$

Applying Isserlis' Theorem [45] and using (55) gives

$$\mathbb{E}[w_j[m]w_j[\ell]w_j[m']w_j[\ell']] = \mathbb{E}[w_j[m]w_j[\ell]] \cdot \mathbb{E}[w_j[m']w_j[\ell']] \quad (61)$$

$$+ \mathbb{E}[w_j[m]w_j[m']] \cdot \mathbb{E}[w_j[\ell]w_j[\ell']] \quad (62)$$

$$+ \mathbb{E}[w_j[m]w_j[\ell']] \cdot \mathbb{E}[w_j[m']w_j[\ell]] \quad (63)$$

$$= \sigma^4 (\delta[m - \ell] \cdot \delta[m' - \ell'] + \delta[m - m'] \cdot \delta[\ell - \ell'] + \delta[m - \ell'] \cdot \delta[m' - \ell]). \quad (64)$$

This reduces the quadruple sum in (60) to double sums for three different configurations of  $m, \ell, m', \ell'$ , determined by the covariances. In the first case ( $m = \ell, m' = \ell'$ ), the sums simplify to  $\sum_{m, m'=0}^{L_K-1} e^{-2\pi i(m-m')(\frac{2k}{L}-\frac{n}{d})}$ . In the second case ( $m = m', \ell = \ell'$ ), all exponents cancel, hence, the sums add up to  $L_K^2$ . The third case ( $m = \ell', m' = \ell$ ), coincides with the second term in the variance,

$$\left| \mathbb{E} \left[ \hat{w}_j[k] \overline{\hat{w}_j[k - n\frac{L}{d}]} \right] \right|^2 = \frac{\sigma^4}{L^2} \sum_{\ell, \ell'=0}^{L_K-1} e^{2\pi i(\ell-\ell')n/d}. \quad (65)$$

Hence, the two terms cancels. In total, this gives us

$$\mathbb{V}[G_n[k]] = \frac{\sigma^4 M}{L^2 d^2} \left( \sum_{m, m'=0}^{L_K-1} e^{-2\pi i(m-m')(\frac{2k}{L}-\frac{n}{d})} + L_K^2 \right). \quad (66)$$

We can further simplify the expression by setting  $\tau = m - m'$  and recognizing the Fourier transform of a triangular function at  $\frac{2k}{L} - \frac{n}{d}$  which has a known closed form expression in terms of  $\sin^2$  [46]. We get

$$\mathbb{V}[G_n[k]] = \frac{\sigma^4 M}{L^2 d^2} \left( \sum_{\tau=-L_K+1}^{L_K-1} (L_K - |\tau|) e^{-2\pi i\tau(\frac{2k}{L}-\frac{n}{d})} + L_K^2 \right) \quad (67)$$

$$= \frac{\sigma^4 M}{L^2 d^2} \left( \frac{\sin^2(L_K \pi(\frac{2k}{L} - \frac{n}{d}))}{\sin^2(\pi(\frac{2k}{L} - \frac{n}{d}))} + L_K^2 \right). \quad (68)$$

Finally, the fact that  $0 \leq \frac{\sin^2(L_K \pi t)}{\sin^2(\pi t)} \leq L_K^2$  for all  $t \in \mathbb{R}$  shows (54).  $\square$

From the expression of the variance we see that the aliasing terms peak for  $\frac{2k}{L} - \frac{n}{d} \in \mathbb{Z}$ . Given that  $k$  takes values from 0 to  $L - 1$  and  $0 \leq n \leq d - 1$ , for every  $G_n$  this happens exactly twice, namely at  $k = \frac{nL}{2d}$  and  $k = \frac{nL}{2d} + \frac{L}{2}$ . Figure 5 shows empirical statistics of the  $G_n$  for a random  $d$ -decimated filterbank, together with the theoretical expressions from Theorem 4.1. We see a behavior that looks like a uniform distribution with mean at one that peaks around the two mentioned values. Note that the peaks comes from the fact that the filters arise by applying rectangular windows of size of the kernels which accumulate energy at their center frequency at zero.

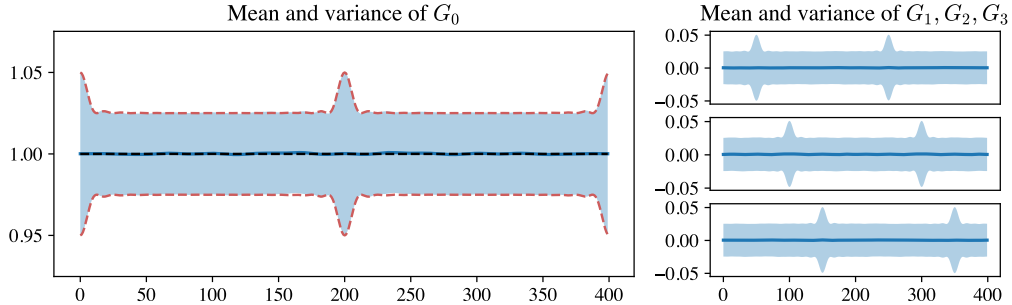


Figure 5: Empirical and theoretical expected values and variances of the aliasing terms of a filterbank with  $M = 40$  random filters in  $\mathbb{C}^{400}$ , kernel size  $L_K = 16$ , and stride  $d = 4$ . The empirical statistics are computed from drawing the random filterbank 100k times. The dashed lines represent the theoretical expressions from Thm. 4.1.

In practice, setting the variance for initializing the kernels in dependence of the total number of parameters in the layer is a common way to balance the energy of the layer. According to our findings in Thm. 4.1, we should set  $\sigma^2 = \frac{Ld}{ML_K}$ .

**Corollary 4.2.** *Let  $\frac{L_K}{d} \in \mathbb{N}$  and  $\sigma^2 = \frac{Ld}{ML_K}$ . Then*

$$\mathbb{E}[\hat{S}] = \mathbb{E}[S] = I_L \quad (69)$$

and

$$M^{-1} \leq \mathbb{V}[G_n] \leq 2M^{-1}. \quad (70)$$

This makes precise that the stability of a random filterbank, distributed with a variance that is inversely proportional to the number of parameters, depends merely on the number of channels. Given that the variance bounds in (70) are very coarse by only considering the two peaks, we find that a random filterbank (with appropriate scaling) seems to behave well in reasonable settings. Even for only  $M = 8$  channels, the variance lies between 0.125 and 0.25. For  $M = 256$ , the bounds already become 0.004 and 0.008.

## 5. Generalizations

In this last section, we describe how more general settings can be converted to the filterbank format  $\{(w_j)_{j=1}^M, \downarrow_d\}$  that we assumed throughout the paper. With these adaption, any result from the previous sections can be applied directly.

### 5.1. Multi-channel layers

Convolutional layers are commonly used to process multiple input channels simultaneously. Let  $\{(w_{1,j})_{j=1}^M, \dots, (w_{C,j})_{j=1}^M, \downarrow_d\}$  denote a multi-channel filterbank for  $\mathbb{C}^L$  with  $C$  input channels and  $M$  output channels, where  $w_{i,j} \in \mathbb{C}^L$  denotes the filter associated with input channel  $i$  and output channel  $j$ . Let  $x_i$  denote the input from channel  $i$  then the transform (analysis operator) for the multi-channel filterbank  $\Theta : \mathbb{C}^{CL} \rightarrow \mathbb{C}^{ML/d}$  is given by

$$(x_1, \dots, x_C) \mapsto \left( \sum_{i=1}^C (x_i * w_{i,j}) \downarrow_d \right)_{j=1}^M. \quad (71)$$

By interlacing the filters across all channels, we obtain a filterbank with equivalent stability properties.

**Lemma 5.1.** *The frame bounds of the filterbank  $\{(w_{1,j})_{j=1}^M, \dots, (w_{C,j})_{j=1}^M, \downarrow_d\}$  in  $\mathbb{C}^L$  are given by the ones of the filterbank*

$$\{(\mathbf{w}_j)_{j=1}^M, \downarrow_{Cd}\} \text{ in } \mathbb{C}^{CL}, \quad (72)$$

with filters  $\mathbf{w}_j \in \mathbb{C}^{CL}$  given by

$$\mathbf{w}_j[nC + i - 1] = w_{i,j}[n]. \quad (73)$$

The kernel size is  $CL_K$ .

*Proof.* Let  $\mathbf{x} \in \mathbb{C}^{CL}$  be given by  $\mathbf{x}[nC + i - 1] = x_i[n]$  then a straightforward calculation shows that for every  $j$ ,

$$\sum_{i=1}^C (x_i * w_{i,j}) \downarrow_d = (\mathbf{x} * \mathbf{w}_j) \downarrow_{Cd}. \quad (74)$$

The claim follows since interlacing is a unitary operation.  $\square$

The corresponding g-frame has elements  $\mathbf{T}_j : \mathbb{C}^{CL} \rightarrow \mathbb{C}^{L/d}$  given by  $x \mapsto (x * \mathbf{w}_j) \downarrow_{Cd}$  and the frame operator is given by  $\mathbf{S} = \Theta^* \Theta = \sum_{j=1}^M \mathbf{T}_j^* \mathbf{T}_j$ . Tightness means  $\mathbf{S} = A \cdot I_{CL}$ .

In the multichannel setting, there are situations where we cannot (and also do not want to) ensure the g-frame property on  $\mathbb{C}^{CL}$  due to specific



architecture designs or dimensionality constraints, i.e.,  $C > M/d$ . In this case, the largest and smallest non-zero eigenvalues of  $\mathbf{S}$  provide the stability bounds on the range of  $\Theta$ . We can express this situation conveniently through the frame bounds of a filterbank with different filters.

**Lemma 5.2.** *Let  $\{(w_{i,j})_{i=1,j=1}^{C,M}, \downarrow_d\}$  be a multi-channel filterbank in  $\mathbb{C}^L$  with  $C > M/d$ . Then the largest and smallest non-zero eigenvalues of  $\mathbf{S}$  are equal to the frame bounds of the filterbank*

$$\{(\mathbf{v}_j)_{j=1}^d, \downarrow_M\} \text{ in } \mathbb{C}^{ML/d}, \quad (75)$$

where the filters  $\mathbf{v}_j \in \mathbb{C}^{ML/d}$  are given by

$$\mathbf{v}_j[nM + k - 1] = \overline{\mathbf{w}_k[nCd + j - 1]}. \quad (76)$$

*Proof.* Through the described row and column rearrangements one obtains that the frame operator for the filterbank  $\{(\mathbf{v}_j)_{j=1}^d, \downarrow_M\}$  is unitarily equivalent to the Gram matrix  $\mathbf{G} = \Theta\Theta^*$  for the filterbank  $\{(\mathbf{w}_j)_{j=1}^M, \downarrow_{Cd}\}$ . Since Gram matrix and frame operator share all non-zero eigenvalues, the claim follows.  $\square$

Tightness in the sense of  $\tilde{\mathbf{S}} = A \cdot I_{ML/d}$  now means that all non-zero eigenvalues of  $\mathbf{S}$  are equal to  $A$ . We can interpret this as the g-frame for the original filterbank being tight on the subspace  $\mathbb{C}^{ML/d}$ . For  $A = 1$ , this is equivalent to the row vectors of the matrix representation of  $\Theta$  being orthonormal [41].

### 5.2. Non-uniform stride

By default, all convolutions in a strided convolutional layer use the same (uniform) stride across the channels. Yet, if a layer architectures requires a non-uniform stride configuration (as e.g., in [37]), we can instead consider an equivalent uniform filterbank that contains shifted copies of the filters.

**Lemma 5.3.** *Let  $\{(w_j)_{j=1}^M, \downarrow_{d_j}\}$  be a non-uniform filterbank in  $\mathbb{C}^L$  with channel-specific strides  $d_1, \dots, d_M \geq 1$ . Then the frame bounds are given by the ones of the filterbank*

$$\{((\mathbf{T}_{nd_j} w_j)_{n=0}^{D_j})_{j=1}^M, \downarrow_{\mathbf{d}}\} \text{ in } \mathbb{C}^L,$$

where  $\mathbf{d} = \text{lcm}(d_1, \dots, d_M)$  and  $D_j = \mathbf{d}/d_j$ .

*Proof.* The statement follows from the  $\ell^2(\mathbb{Z})$  case for which a proof can be found in [11].  $\square$

### 5.3. Dilated convolution

For  $a \geq 1$ , dilated convolution of  $x$  and  $w$  in  $\mathbb{C}^L$  is defined by

$$(x *_a w)[n] = \sum_{\ell=0}^{L-1} x[a\ell] \overline{w[n-\ell]}. \quad (77)$$

This is equivalent to dilate the kernel of the filter by  $a$  directly, which can be realized by upsampling the filter and cropping the result back to length  $L$ . We denote the cropping by  $w_{:L}$ .

**Lemma 5.4.** *Let  $\{(w_j)_{j=1}^M, \downarrow_d, \wedge_a\}$  denote a filterbank in  $\mathbb{C}^L$  that applies dilated convolution with  $a \leq L/L_K$ . The frame bounds are given by the ones of the filterbank*

$$\{((w_j \uparrow_a)_{:L})_{j=1}^M, \downarrow_d\} \text{ in } \mathbb{C}^L, \quad (78)$$

with kernel sizes  $aL_K$ .

*Proof.* It is easy to see that

$$(x *_a w) \downarrow_d [n] = \sum_{\ell=0}^{L-1} x[a\ell] \overline{w[dn-\ell]} = (x * (w \uparrow_a)_{:L}) \downarrow_d [n]. \quad (79)$$

Assuming  $a \leq L/L_K$  guarantees that we only crop zeros. Hence, the frame bounds are not affected.  $\square$

## 6. Discussion and Outlook

This work shows that generalized frames offer a natural formalism to study aliasing in strided convolutional layers and link it to their invertibility and stability. We can elegantly formulate perfect reconstruction conditions, which only depend on the kernel size and the stride, and get tightness characterization with an extension to the painless case via sub-band restriction. With this we can design computationally efficient optimization objectives that are nicely interpretable by means of aliasing suppression. Using the aliasing-based objectives as regularizers to promote Parseval stability in a larger-scale application did not fit into the scope of this work and remains to be tested. In the same context, it is interesting to ask how stability is affected by a learning update in general, and how regularization can intrinsically bias the updates towards a local minimum which yields at least a

comparable performance but with stability guarantees. By a standard frame perturbation result (see e.g. [47]), the worst-case scenario is determined by the upper frame bound of the update filterbank with filters  $u_j$  given by  $u_j[n] = \Phi((w_j)_{j=1}^M)[j, n]$ , where  $\Phi$  represents the update rule (e.g., the gradients).

**Lemma 6.1.** *Let  $\{(w_j)_{j=1}^M, \downarrow_d\}$  be a filterbank with frame bounds  $A, B$  and  $R$  the optimal upper frame bound for  $\{(u_j)_{j=1}^M, \downarrow_d\}$ . If  $\gamma \cdot R < A$  then  $\{(w_j - \gamma \cdot u_j)_{j=1}^M, \downarrow_d\}$  is a  $g$ -frame with frame bounds  $(\sqrt{A} - \gamma\sqrt{R})^2, (\sqrt{B} + \gamma\sqrt{R})^2$ .*

While this means that after each weight update, the frame bounds might worsen by  $\gamma\sqrt{R}$ , the interesting question is if and how regularization can avoid these worst-case scenarios.

In the context of random initialization, this work provides the fundamental statistical properties of the aliasing terms, but leaves some open questions on how to apply them. First, how can we obtain the statistics for the frame bounds from the statistics of the aliasing terms? While it seems to be a natural application, we can not use norm properties of random vectors without knowing their distribution. Such an analysis has to be done in a more comprehensive follow-up work that focuses on randomness. Second, we observed that the peaky behavior of the variance around two entries comes from the accumulation of energy at the center frequency of the rectangular windows that we apply. An approach to better distribute the peaks in the variance could be to apply smooth window functions with different frequencies.

## Acknowledgment

D. Haider is a recipient of a DOC Fellowship (A 26355) of the Austrian Academy of Sciences at the Acoustics Research Institute. The work of V. Lostanlen was supported by the ANR project MuReNN (ANR-23-CE23-0007-01). P. Balazs was supported by the FWF projects LoFT (P 34624), NoMASP (P 34922), Voice Prints (P 36446), and the WWTF project EleCom (LS23-024). The work of N. Holighaus was supported by the FWF project DISCO (PAT4780023).

## References

- [1] P. P. Vaidyanathan, *Multirate systems and filter banks*, Prentice-Hall, Inc., USA, 1993.
- [2] P. Vaidyanathan, “Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques,” *IEEE ASSP Magazine*, vol. 4, no. 3, pp. 4–20, 1987.
- [3] R.D. Koilpillai and P.P. Vaidyanathan, “Cosine-modulated FIR filter banks satisfying perfect reconstruction,” *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 770–783, 1992.
- [4] J. Kovacevic and M. Vetterli, “Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for  $r/\sup n/$ ,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 533–555, 1992.
- [5] Jelena Kovacevic and Martin Vetterli, “Perfect reconstruction filter banks with rational sampling factors,” *IEEE Transactions on signal processing*, vol. 41, no. 6, pp. 2047–2066, 1993.
- [6] Martin Vetterli, “Filter banks allowing perfect reconstruction,” *Signal processing*, vol. 10, no. 3, pp. 219–244, 1986.
- [7] M. Vetterli and D. Le Gall, “Perfect reconstruction FIR filter banks: some properties and factorizations,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 1057–1071, 1989.
- [8] Jiahao Su, Wonmin Byeon, and Furong Huang, “Scaling-up diverse orthogonal convolutional networks by a paraunitary framework,” in *International Conference on Machine Learning (ICML)*. 2022, vol. 162 of *Proceedings of Machine Learning Research*, pp. 20546–20579, PMLR.
- [9] Michael Unser and Stanislas Ducotterd, “Parseval convolution operators and neural networks,” *arXiv: 2408.09981*, 2024.
- [10] Asher Trockman and J. Zico Kolter, “Orthogonalizing convolutional layers with the Cayley transform,” in *International Conference on Learning Representations (ICLR)*, 2021.

- [11] Peter Balazs, Nicki Holighaus, Thibaud Necciari, and Diana Stoeva, “Frame theory for signal processing in psychoacoustics,” in *Excursions in Harmonic Analysis, Volume 5: The February Fourier Talks at the Norbert Wiener Center*, Radu Balan, John J. Benedetto, Wojciech Czaja, Matthew Dellatorre, and Kasso A. Okoudjou, Eds. 2017, pp. 225–268, Springer International Publishing.
- [12] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman, “1D convolutional neural networks and applications: A survey,” *Mechanical Systems and Signal Processing*, vol. 151, pp. 107398, 2021.
- [13] David F Walnut, “Continuity properties of the gabor frame operator,” *Journal of Mathematical Analysis and Applications*, vol. 165, no. 2, pp. 479–504, 1992.
- [14] A. J. E. M. Janssen, *The duality condition for Weyl-Heisenberg frames*, pp. 33–84, Birkhäuser Boston, Boston, MA, 1998.
- [15] Eugenio Hernández, Demetrio Labate, and Guido Weiss, “A unified characterization of reproducing systems generated by a finite family, II,” *The Journal of Geometric Analysis*, vol. 12, pp. 615—662, 2002.
- [16] Nicki Holighaus, “Structure of nonstationary Gabor frames and their dual systems,” *Applied and Computational Harmonic Analysis*, vol. 37, no. 3, pp. 442–463, 2014.
- [17] Xavier Glorot and Yoshua Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *International Conference on Artificial Intelligence and Statistics*. 2010, vol. 9 of *Proceedings of Machine Learning Research*, pp. 249–256, PMLR.
- [18] Claudio Santos and João Papa, “Avoiding overfitting: A survey on regularization methods for convolutional neural networks,” *ACM Computing Surveys*, vol. 54, 2022.
- [19] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, “Explaining and harnessing adversarial examples,” in *International Conference on Learning Representations (ICLR)*, Yoshua Bengio and Yann LeCun, Eds., 2015.

- [20] Louis Béthune, *Deep learning with Lipschitz constraints*, Ph.D. thesis, Université de Toulouse, 2024.
- [21] Yusuke Tsuzuku, Issei Sato, and Masashi Sugiyama, “Lipschitz-margin training: scalable certification of perturbation invariance for deep neural networks,” in *International Conference on Neural Information Processing Systems (NeurIPS)*, 2018, p. 6542–6551.
- [22] Patricia Pauli, Anne Koch, Julian Berberich, Paul Kohler, and Frank Allgöwer, “Training robust neural networks using Lipschitz bounds,” *IEEE Control Systems Letters*, vol. 6, pp. 121–126, 2022.
- [23] Moustapha Cisse, Piotr Bojanowski, Edouard Grave, Yann Dauphin, and Nicolas Usunier, “Parseval networks: Improving robustness to adversarial examples,” in *International Conference on Machine Learning (ICML)*, 2017.
- [24] Andrew Brock, Theodore Lim, J.M. Ritchie, and Nick Weston, “Neural photo editing with introspective adversarial networks,” in *International Conference on Learning Representations (ICLR)*, 2017.
- [25] Marzieh Hasannasab, Johannes Hertrich, Sebastian Neumayer, Gerlind Plonka, Simon Setzer, and Gabriele Steidl, “Parseval proximal neural networks,” *Journal of Fourier Analysis and Applications*, vol. 26, pp. 1–31, 2020.
- [26] Yi Luo and Nima Mesgarani, “Conv-TasNet: Surpassing ideal time–frequency magnitude masking for speech separation,” *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 27, no. 8, pp. 1256–1266, 2019.
- [27] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli, “wav2vec 2.0: A framework for self-supervised learning of speech representations,” *International Conference on Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 12449–12460, 2020.
- [28] Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, Jian Wu, Long Zhou, Shuo Ren, Yanmin Qian, Yao Qian, Jian Wu, Michael Zeng, Xiangzhan Yu, and Furu Wei, “WavLM: Large-scale self-supervised pre-training for full stack speech processing,” *IEEE Journal*

- of *Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1505–1518, 2022.
- [29] Wenchang Sun, “G-frames and G-Riesz bases,” *Journal of Mathematical Analysis and Applications*, vol. 322, no. 1, pp. 437–452, 2006.
- [30] Matthew Fickus and Dustin G. Mixon, “Numerically erasure-robust frames,” *Linear Algebra and its Applications*, vol. 437, no. 6, pp. 1394–1407, 2012.
- [31] Martin Vetterli, Jelena Kovačević, and Vivek K Goyal, *Foundations of Signal Processing*, Cambridge University Press, 2014.
- [32] Ingrid Daubechies, Alexandre Grossmann, and Yves Meyer, “Painless nonorthogonal expansions,” *Journal of Mathematical Physics*, vol. 27, no. 5, pp. 1271–1283, 1986.
- [33] Peter Balazs, Monika Dörfler, Florent Jaillet, Nicki Holighaus, and Gino Angelo Velasco, “Theory, implementation and applications of nonstationary gabor frames,” *Journal of Computational and Applied Mathematics*, vol. 236, no. 6, pp. 1481–1496, October 2011.
- [34] Daniel Haider, Felix Perfler, Vincent Lostanlen, Martin Ehler, and Peter Balazs, “Hold me Tight: Stable encoder/decoder design for speech enhancement,” in *Annual Conference of the International Speech Communication Association (Interspeech)*, 2024.
- [35] Neil Zeghidour, Olivier Teboul, Félix de Chaumont Quitry, and Marco Tagliasacchi, “LEAF: A learnable frontend for audio classification,” in *International Conference on Machine Learning (ICML)*, 2021.
- [36] Mirco Ravanelli and Yoshua Bengio, “Speaker recognition from raw waveform with SincNet,” in *IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2018, pp. 1021–1028.
- [37] Vincent Lostanlen, Daniel Haider, Han Han, Mathieu Lagrange, Peter Balazs, and Martin Ehler, “Fitting auditory filterbanks with multiresolution neural networks,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2023.

- [38] Vincent Lostanlen, Xiran Zhang, Daniel Haider, Mathieu Lagrange, Martin Ehler, and Peter Balazs, “Residual Hybrid Filterbanks,” in *IEEE Statistical Signal Processing Workshop (SSP)*, 2025.
- [39] Rossen Nenov, Vincent Lostanlen, and Peter Balazs, “Robust Deconvolution with Parseval Filterbanks,” in *IEEE International Conference on Sampling Theory and Applications (SampTA)*, 2025.
- [40] Jiayun Wang, Yubei Chen, Rudrasis Chakraborty, and Stella X. Yu, “Orthogonal convolutional neural networks,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11502–11512.
- [41] Estelle Massart, “Orthogonal regularizers in deep learning: how to handle rectangular matrices?,” in *International Conference on Pattern Recognition (ICPR)*, 2022, pp. 1294–1299.
- [42] Peter Balazs, Daniel Haider, Felix Perfler, and Vincent Lostanlen, “Trainable signal encoders that are robust against noise,” in *Internoise Conference*, 2024.
- [43] Rossen Nenov, Daniel Haider, and Peter Balazs, “(Almost) Smooth Sailing: Towards numerical stability of neural networks through differentiable regularization of the condition number,” in *Differentiable Almost Everything Workshop at the International Conference on Machine Learning*, 2024.
- [44] Martin Ehler, “Preconditioning Filter Bank Decomposition Using Structured Normalized Tight Frames,” *Journal of Applied Mathematics*, vol. 2015, pp. 1 – 12, 2015.
- [45] Leon Isserlis, “On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables,” *Biometrika*, vol. 12, no. 1/2, pp. 134–139, 1918.
- [46] Stéphane Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, Inc., USA, 3rd edition, 2008.
- [47] Shayne Waldron, “An introduction to finite tight frames,” *In: Applied and Numerical Harmonic Analysis*, pp. 1–587, 2018.