

# **Sparse Signal Processing**

## **Lecture 4 — Fourier interpolation and the linear model**

Nicolas Keriven (slides by Clément Elvira)

# Outline

N. Keriven: CNRS researcher at IRISA, SIROCCO team.

<https://nkeriven.github.io>

# Outline

N. Keriven: CNRS researcher at IRISA, SIROCCO team.

<https://nkeriven.github.io>

In today's class we will

- ▶ particularize the linear model to an infinite-dimensional setting ;
- ▶ introduce a linear target set ;
- ▶ derive a decoder achieving exact recovery ;

# Outline

N. Keriven: CNRS researcher at IRISA, SIROCCO team.

<https://nkeriven.github.io>

In today's class we will

- ▶ particularize the linear model to an infinite-dimensional setting ;
- ▶ introduce a linear target set ;
- ▶ derive a decoder achieving exact recovery ;

Material at <https://github.com/nkeriven/ENS-signal> (subject to changes, check every now and then)

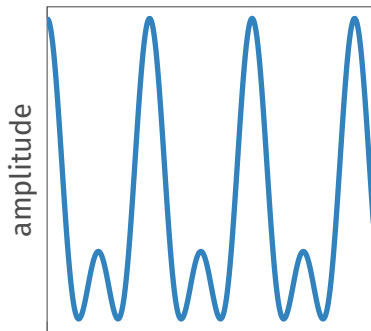
# References

- ▶ Mallat, 2012. *A Wavelet Tour of Signal Processing: the Sparse Way*
- ▶ F Golse polycopié: <http://www.math.polytechnique.fr/~golse/MAT431-10/POLY431.pdf>
- ▶ Wikipédia !

# Considered Task

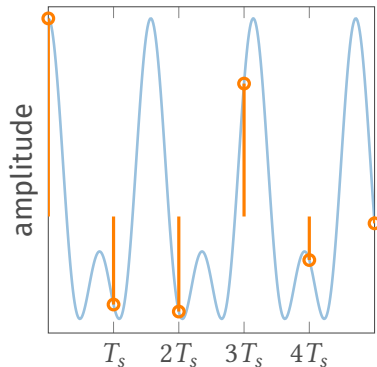
Recovering a signal  $x$  from **infinitely** many function evaluations

$x$



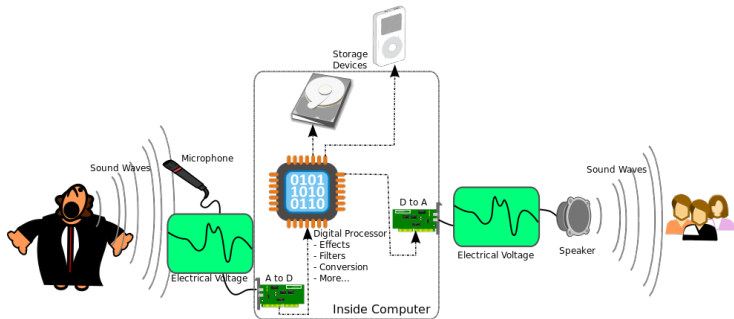
time ( $t$ )

$y$



time ( $t$ )

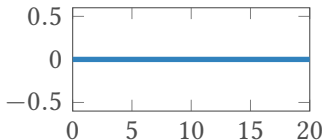
# Application 1: sampling and reconstructing audio recording



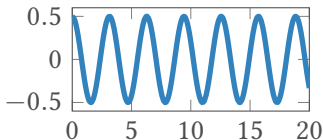
**Important component of sound perception:** the frequency

## Application 2: (simplified) transmitters

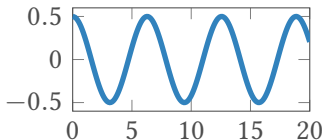
Sequence "00"



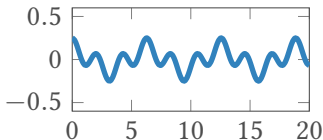
Sequence "01"



Sequence "10"



Sequence "11"



**Support of information:** electromagnetic (sinusoidal at a given frequency) waves



# Wrapping up: main ingredients

1. **The sensing operator:** The (linear) operator defined by

$$\begin{aligned} M: \mathcal{L}^2(\mathbf{R}) &\longrightarrow \mathbf{R}^{\mathbf{Z}} \\ x &\longmapsto \{x(jT_s)\}_{j \in \mathbf{Z}} \end{aligned}$$

2. **The target set:** *To be defined*

*Related to some notion of frequency*

3. **The decoder:** *To be defined*

4. **The accuracy criterion:** “noiseless exact recovery”, i.e.

$$\forall x \in \mathcal{X}_{\text{target}}, \quad D(M(x)) = x$$



Which mathematical tool can be used to emphasize the frequencies characterizing a signal?



Which mathematical tool can be used to emphasize the frequencies characterizing a signal?



The Fourier transform

# Rappels: The Fourier transform

## Definition

**The Fourier transform** is a linear operator defined by

$$\begin{aligned}\mathcal{F}: \mathcal{L}^1(\mathbf{R}) &\longrightarrow \mathcal{L}^1(\mathbf{R}) \\ x &\longmapsto \left( f \mapsto \int_{-\infty}^{+\infty} x(t) e^{-i2\pi ft} dt \right)\end{aligned}$$

# Rappels: The Fourier transform

## Definition

**The Fourier transform** is a linear operator defined by

$$\begin{aligned}\mathcal{F}: \mathcal{L}^1(\mathbf{R}) &\longrightarrow \mathcal{L}^1(\mathbf{R}) \\ x &\longmapsto \left( f \mapsto \int_{-\infty}^{+\infty} x(t) e^{-i2\pi f t} dt \right)\end{aligned}$$

- Sometimes defined without the  $2\pi$  (but less clean, multiplicative constants everywhere)
- $\mathcal{F}x$  is often written  $\hat{x}$

# An example

## Exercise

Evaluate the Fourier transform of

$$\Pi_{[-\frac{a}{2}, \frac{a}{2}]} : t \longmapsto \begin{cases} 1 & \text{if } t \in [-\frac{a}{2}, +\frac{a}{2}] \\ 0 & \text{otherwise} \end{cases}$$

# An example

## Exercise

Evaluate the Fourier transform of

$$\Pi_{[-\frac{a}{2}, \frac{a}{2}]} : t \longmapsto \begin{cases} 1 & \text{if } t \in [-\frac{a}{2}, +\frac{a}{2}] \\ 0 & \text{otherwise} \end{cases}$$

## Answer

$$\mathcal{F}\Pi(f) = a \frac{\sin(\pi af)}{\pi af} \stackrel{\text{def.}}{=} a \operatorname{sinc}(af)$$

# Rappels: Inverse Fourier transform

## Definition

**The conjugate Fourier transform** is the linear operator defined by

$$\begin{aligned}\overline{\mathcal{F}}: \mathcal{L}^1(\mathbf{R}) &\longrightarrow \mathcal{L}^1(\mathbf{R}) \\ x &\longmapsto \left( f \mapsto \int_{-\infty}^{+\infty} x(t) e^{i2\pi ft} dt \right)\end{aligned}$$



# Rappels: Inverse Fourier transform

## Definition

The **conjugate Fourier transform** is the linear operator defined by

$$\begin{aligned}\overline{\mathcal{F}}: \mathcal{L}^1(\mathbf{R}) &\longrightarrow \mathcal{L}^1(\mathbf{R}) \\ x &\longmapsto \left( f \mapsto \int_{-\infty}^{+\infty} x(t) e^{i2\pi ft} dt \right)\end{aligned}$$

## Theorem: Inversion

If  $x \in \mathcal{L}^1(\mathbf{R})$  is **such that**  $\mathcal{F}x \in \mathcal{L}^1(\mathbf{R})$  then

$$\overline{\mathcal{F}}[\mathcal{F}x] \stackrel{\text{a.e.}}{=} x$$

with equality at every point of continuity.

## Extension to $L^2$



**Pbm:**  $x \in \mathcal{L}^1$  does **not** imply  $\mathcal{F}x \in \mathcal{L}^1$ ! (ex: door function and cardinal sinus)

## Extension to $L^2$



**Pbm:**  $x \in \mathcal{L}^1$  does **not** imply  $\mathcal{F}x \in \mathcal{L}^1$ ! (ex: door function and cardinal sinus)

→ **Solution:** extension to  $\mathcal{L}^2$  and Parseval identity.

## Extension to $L^2$



**Pbm:**  $x \in \mathcal{L}^1$  does **not** imply  $\mathcal{F}x \in \mathcal{L}^1$ ! (ex: door function and cardinal sinus)

→ **Solution:** extension to  $\mathcal{L}^2$  and Parseval identity.



**Pbm:** the classical Fourier integral is not defined for  $\mathcal{L}^2$ ...

## Extension to $L^2$



**Pbm:**  $x \in \mathcal{L}^1$  does **not** imply  $\mathcal{F}x \in \mathcal{L}^1$ ! (ex: door function and cardinal sinus)

→ **Solution:** extension to  $\mathcal{L}^2$  and Parseval identity.



**Pbm:** the classical Fourier integral is not defined for  $\mathcal{L}^2$ ...

→ **Solution:** extension by *density*.

# Parseval's identity

## Parseval's identity

If  $f, g \in \mathcal{L}^1 \cap \mathcal{L}^2$ ,

$$\langle f, g \rangle_{L^2} = \langle \mathcal{F}f, \mathcal{F}g \rangle_{L^2}$$

(recall that  $\langle f, g \rangle_{L^2} = \int f(t) \overline{g(t)} dt$ )

# Parseval's identity

## Parseval's identity

If  $f, g \in \mathcal{L}^1 \cap \mathcal{L}^2$ ,

$$\langle f, g \rangle_{L^2} = \langle \mathcal{F}f, \mathcal{F}g \rangle_{L^2}$$

(recall that  $\langle f, g \rangle_{L^2} = \int f(t) \overline{g(t)} dt$ )

Consequence:  $\|f\|_{L^2} = \|\mathcal{F}f\|_{L^2}$  and  $f \in \mathcal{L}^1 \cap \mathcal{L}^2 \Rightarrow \mathcal{F}f \in \mathcal{L}^2$ . Now we just need to extend the domain of definition of  $\mathcal{F}$

# Parseval's identity

## Parseval's identity

If  $f, g \in \mathcal{L}^1 \cap \mathcal{L}^2$ ,

$$\langle f, g \rangle_{L^2} = \langle \mathcal{F}f, \mathcal{F}g \rangle_{L^2}$$

(recall that  $\langle f, g \rangle_{L^2} = \int f(t) \overline{g(t)} dt$ )

Consequence:  $\|f\|_{L^2} = \|\mathcal{F}f\|_{L^2}$  and  $f \in \mathcal{L}^1 \cap \mathcal{L}^2 \Rightarrow \mathcal{F}f \in \mathcal{L}^2$ . Now we just need to extend the domain of definition of  $\mathcal{F}$

We are going to extend Fourier to  $\mathcal{L}^2$  by *density* of  $\mathcal{L}^1 \cap \mathcal{L}^2$  within  $\mathcal{L}^2$

→ i.e.,  $\forall \epsilon > 0, f \in \mathcal{L}^2, \exists g \in \mathcal{L}^1 \cap \mathcal{L}^2$  s.t.  $\|f - g\|_{L^2} \leq \epsilon$



## Extension to $\mathcal{L}^2$

- ▶ Let  $f \in \mathcal{L}^2$ .
- ▶ By density of  $\mathcal{L}^2 \cap \mathcal{L}^1$  in  $\mathcal{L}^2$ , there is a sequence  $f_n \in \mathcal{L}^2 \cap \mathcal{L}^1$  such that  $\|f_n - f\|_{\mathcal{L}^2} \rightarrow 0$

## Extension to $\mathcal{L}^2$

- ▶ Let  $f \in \mathcal{L}^2$ .
- ▶ By density of  $\mathcal{L}^2 \cap \mathcal{L}^1$  in  $\mathcal{L}^2$ , there is a sequence  $f_n \in \mathcal{L}^2 \cap \mathcal{L}^1$  such that  $\|f_n - f\|_{\mathcal{L}^2} \rightarrow 0$
- ▶  $f_n$  is a Cauchy sequence, and by Parseval's identity  $\|f_n - f_p\|_{\mathcal{L}^2} = \|\mathcal{F}(f_n - f_p)\|_{\mathcal{L}^2}$ , hence  $\mathcal{F}f_n \in \mathcal{L}^2$  is also a Cauchy sequence

# Extension to $\mathcal{L}^2$

- ▶ Let  $f \in \mathcal{L}^2$ .
- ▶ By density of  $\mathcal{L}^2 \cap \mathcal{L}^1$  in  $\mathcal{L}^2$ , there is a sequence  $f_n \in \mathcal{L}^2 \cap \mathcal{L}^1$  such that  $\|f_n - f\|_{\mathcal{L}^2} \rightarrow 0$
- ▶  $f_n$  is a Cauchy sequence, and by Parseval's identity  $\|f_n - f_p\|_{\mathcal{L}^2} = \|\mathcal{F}(f_n - f_p)\|_{\mathcal{L}^2}$ , hence  $\mathcal{F}f_n \in \mathcal{L}^2$  is also a Cauchy sequence
- ▶ Since  $\mathcal{L}^2$  is a Hilbert space and **is therefore complete**,  $\mathcal{F}f_n$  is a convergent sequence.

# Extension to $\mathcal{L}^2$

- ▶ Let  $f \in \mathcal{L}^2$ .
- ▶ By density of  $\mathcal{L}^2 \cap \mathcal{L}^1$  in  $\mathcal{L}^2$ , there is a sequence  $f_n \in \mathcal{L}^2 \cap \mathcal{L}^1$  such that  $\|f_n - f\|_{\mathcal{L}^2} \rightarrow 0$
- ▶  $f_n$  is a Cauchy sequence, and by Parseval's identity  $\|f_n - f_p\|_{\mathcal{L}^2} = \|\mathcal{F}(f_n - f_p)\|_{\mathcal{L}^2}$ , hence  $\mathcal{F}f_n \in \mathcal{L}^2$  is also a Cauchy sequence
- ▶ Since  $\mathcal{L}^2$  is a Hilbert space and **is therefore complete**,  $\mathcal{F}f_n$  is a convergent sequence.
- ▶ We call its limit the “Fourier transform of  $f$ ” denoted by  $\mathcal{F}f$ . It satisfies all the usual properties of the Fourier transform (see next slide)

# Properties of Fourier Transform (...easy, or admitted)

- ▶  $f$  symmetric  $\Rightarrow \mathcal{F}f$  real
- ▶  $f$  anti-symmetric  $\Rightarrow \mathcal{F}f$  imaginary
- ▶  $f$  real  $\Rightarrow \mathcal{F}f$  Hermitian
- ▶  $f$  imaginary  $\Rightarrow \mathcal{F}f$  (conjugate) anti-symmetric

# Properties of Fourier Transform (...easy, or admitted)

- ▶  $f$  symmetric  $\Rightarrow \mathcal{F}f$  real
- ▶  $f$  anti-symmetric  $\Rightarrow \mathcal{F}f$  imaginary
- ▶  $f$  real  $\Rightarrow \mathcal{F}f$  Hermitian
- ▶  $f$  imaginary  $\Rightarrow \mathcal{F}f$  (conjugate) anti-symmetric
  
- ▶  $\mathcal{F}(f \star g)(t) = \mathcal{F}f(t)\mathcal{F}g(t)$  (convolution thm)
- ▶  $\langle f, g \rangle_{L^2} = \langle \mathcal{F}f, \mathcal{F}g \rangle_{L^2}$  (Parseval identity)
- ▶  $\mathcal{F}\frac{df}{dt}(t) = 2i\pi t\mathcal{F}f(t) \Rightarrow$  if  $f$  is  $k$  times differentiable,  $\mathcal{F}f$  decrease faster than  $1/t^k$

# Properties (summary)

Property	Function	Fourier transform
	$x(t)$	$\hat{x}(f)$
Inverse	$\hat{x}(t)$	$x(-f)$
Convolution	$x_1 \star x_2(t)$	$\hat{x}_1 \hat{x}_2(f)$
Multiplication	$x_1(t)x_2(t)$	$\hat{x}_1 \star \hat{x}_2(f)$
Translation	$x(t - u)$	$e^{-2i\pi uf} \hat{x}(f)$
Modulation	$e^{2i\pi \xi t} x(t)$	$\hat{x}(f - \xi)$
Scaling	$x(t/s)$	$ s  \hat{x}(sf)$
Time derivatives	$x^{(p)}(t)$	$(2i\pi f)^p \hat{x}(f)$
Freq. derivatives	$(-2i\pi t)^p x(t)$	$\hat{x}^{(p)}(f)$
Complex conjugate	$\bar{x}(t)$	$\bar{\hat{x}}(-f)$

# Uncertainty principle



Can we construct a function that is **highly localised in space**, whose Fourier transform is **highly localised in frequency**?



# Uncertainty principle



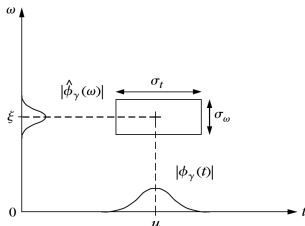
Can we construct a function that is **highly localised in space**, whose Fourier transform is **highly localised in frequency**?

## Heisenberg's uncertainty principle

For  $x \in \mathcal{L}^2(\mathbb{R})$ , define the two density functions

$p_x = \frac{|x(t)|^2}{\|x\|_{L^2}^2}$ ,  $p_{\hat{x}} = \frac{|\hat{x}(f)|^2}{\|\hat{x}\|_{L^2}^2}$  and their respective **variance**  $\sigma_x^2, \sigma_{\hat{x}}^2$ . Then

$$\sigma_x^2 \sigma_{\hat{x}}^2 \geq 1/4$$





Are we done?



Are we done?



The function of interest, namely  $t \mapsto \cos(2\pi ft)$  does not belong to  $\mathcal{L}^1(\mathbf{R})$ !

*nor  $\mathcal{L}^2(\mathbf{R})$*



Are we done?



The function of interest, namely  $t \mapsto \cos(2\pi ft)$  does not belong to  $\mathcal{L}^1(\mathbb{R})$ !

*nor  $\mathcal{L}^2(\mathbb{R})$*

We need a “bigger” space to work with!





Are we done?



The function of interest, namely  $t \mapsto \cos(2\pi ft)$  does not belong to  $\mathcal{L}^1(\mathbf{R})!$

*nor  $\mathcal{L}^2(\mathbf{R})$*

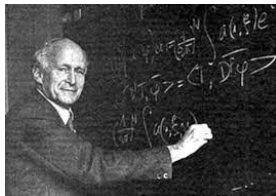
We need a “bigger” space to work with!



→ Space of “distributions”, defined by *duality*.

We need the smallest dual space as possible, to make the biggest distribution space as possible!

# Reminder (?) on distributions



# Tests functions

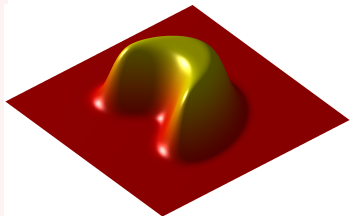
## Definition

A test function refers to any **infinitely differentiable** function

$$\varphi: \mathbf{R} \longrightarrow \mathbf{C}$$

with is **compactly supported**.

The set of test functions is denoted  $\mathcal{D}(\mathbf{R})$ .



# Tests functions

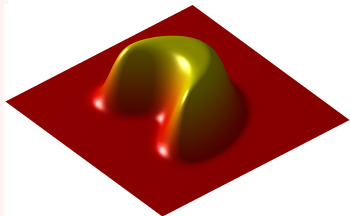
## Definition

A test function refers to any **infinitely differentiable** function

$$\varphi: \mathbf{R} \longrightarrow \mathbf{C}$$

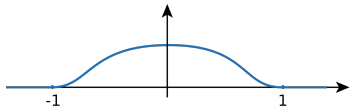
with is **compactly supported**.

The set of test functions is denoted  $\mathcal{D}(\mathbf{R})$ .



**Example:** (*standard one*) the function

$$t \longmapsto \begin{cases} \exp\left(-\frac{1}{1-t^2}\right) & \text{if } t \in [-1, 1] \\ 0 & \text{otherwise} \end{cases}$$



can be proved to be a test function.



# Topology on $\mathcal{D}(\mathbf{R})$

## Definition: convergence in $\mathcal{D}(\mathbf{R})$

Let  $\{\varphi_n\}_{n \in \mathbf{N}}$  be a family of test function.

We say that  $\varphi_n \rightarrow \varphi \in \mathcal{D}(\mathbf{R})$  if and only if

- ▶ there exists an integer  $N$  and a compact set  $K$  such  $n \geq N \Rightarrow \text{support}(\varphi_n - \varphi) \subset K$
- ▶ for all  $r \in \mathbf{N}$ ,

$$\left\| \varphi_n^{(r)} - \varphi^{(r)} \right\|_{\infty} \rightarrow 0$$

- ▶ i.e., supremum norm for  $\varphi$  and all its derivatives.

# Distributions

## Definition

We refer to “distribution” as any continuous linear functional on  $\mathcal{D}(\mathbf{R})$ .

# Distributions

## Definition

We refer to “distribution” as any continuous linear functional on  $\mathcal{D}(\mathbf{R})$ .

## Remarks

- Said differently, the space of distributions is the (topological) dual space of  $\mathcal{D}(\mathbf{R})$ , denoted by  $\mathcal{D}'(\mathbf{R})$ ;

# Distributions

## Definition

We refer to “distribution” as any continuous linear functional on  $\mathcal{D}(\mathbf{R})$ .

## Remarks

- ▶ Said differently, the space of distributions is the (topological) dual space of  $\mathcal{D}(\mathbf{R})$ , denoted by  $\mathcal{D}'(\mathbf{R})$ ;
- ▶ When evaluating a distribution  $T$  at  $\varphi$ , we write

$$\langle T, \varphi \rangle$$

- ▶ By extension, we say that two distributions  $T, T'$  are equal iff

$$\langle T, \varphi \rangle = \langle T', \varphi \rangle \quad \forall \varphi \in \mathcal{D}(\mathbf{R})$$



Distributions are often referred to as “generalized functions”.

This can nevertheless be a misleading terminology, as distributions **do not take scalar as inputs**. (...but *test functions*)



Distributions are often referred to as “generalized functions”.

This can nevertheless be a misleading terminology, as distributions **do not take scalar as inputs**. (...but *test functions*)



Besides mathematical benefits, there is a physical intuition behind this reference

A compactly supported test function can be seen as a model of a physical system measuring some quantity

# Regular distributions

## Definition/Proposition

If  $g$  is **any** locally integrable function, then the linear form  $T_g$  defined as

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle T_g, \varphi \rangle \stackrel{\text{def.}}{=} \int g(t) \varphi(t) dt$$

is a distribution.

Such a distribution is referred to as **“regular distribution”**.

# Regular distributions

## Definition/Proposition

If  $g$  is **any** locally integrable function, then the linear form  $T_g$  defined as

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle T_g, \varphi \rangle \stackrel{\text{def.}}{=} \int g(t) \varphi(t) dt$$

is a distribution.

Such a distribution is referred to as **“regular distribution”**.

*In short: Linearity is fine*

*Continuity: letting  $\varphi_n \rightarrow \varphi$ , there exists a compact  $K$  that contains all supports*

*Then  $|\langle T_g, \varphi_n - \varphi \rangle| \leq \int g(t) dt \|\varphi_n - \varphi\|_\infty$*



# Regular distributions

## Definition/Proposition

If  $g$  is **any** locally integrable function, then the linear form  $T_g$  defined as

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle T_g, \varphi \rangle \stackrel{\text{def.}}{=} \int g(t) \varphi(t) dt$$

is a distribution.

Such a distribution is referred to as **“regular distribution”**.

*In short: Linearity is fine*

*Continuity: letting  $\varphi_n \rightarrow \varphi$ , there exists a compact  $K$  that contains all supports*

*Then  $|\langle T_g, \varphi_n - \varphi \rangle| \leq \int g(t) dt \|\varphi_n - \varphi\|_\infty$*

## Remark

- ▶ Hence distributions is a “bigger” space...
- ▶ If  $T_{g_1} = T_{g_2}$  then  $g_1 \stackrel{\text{a.e.}}{=} g_2$

# Singular distributions



The terminology “singular distributions” refers to any distribution that is not regular

# Singular distributions



The terminology “singular distributions” refers to any distribution that is not regular

**Example 1:** The dirac function  $\delta_t$  defined  $\forall t \in \mathbf{R}$  as

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle \delta_t, \varphi \rangle = \varphi(t)$$

$$\text{since } |\langle \delta_t, \varphi_n - \varphi \rangle| \leq \|\varphi_n - \varphi\|_\infty \rightarrow 0$$

# Singular distributions



The terminology “singular distributions” refers to any distribution that is not regular

**Example 1:** The dirac function  $\delta_t$  defined  $\forall t \in \mathbf{R}$  as

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle \delta_t, \varphi \rangle = \varphi(t)$$

*since  $|\langle \delta_t, \varphi_n - \varphi \rangle| \leq \|\varphi_n - \varphi\|_\infty \rightarrow 0$*

**Example 2:** The dirac comb  $\text{III}_{T_s}$  defined as

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle \text{III}_{T_s}, \varphi \rangle = \sum_{j \in \mathbf{Z}} \langle \delta_{jT_s}, \varphi \rangle = \sum_{j \in \mathbf{Z}} \varphi(jT_s)$$

*Letting  $\varphi_n \rightarrow \varphi$ , there exists a compact  $K$  that contains all supports, and*

*$|\langle \delta_t, \varphi_n - \varphi \rangle| \leq K_{T_s} \|\varphi_n - \varphi\|_\infty \rightarrow 0$ , where  $K_{T_s}$  is the number of  $jT_s$  contained in  $K$*



**Regular distributions** provide intuition on how to define many concepts related to distributions

*derivation, convolution, Fourier transform to name but a few*

# Derivation of a distribution

**For regular distributions:** let  $g$  be a differentiable locally integrable function. Then  $g'$  is also locally integrable and

$$\begin{aligned}\langle T_{g'}, \varphi \rangle &= \int g'(t) \varphi(t) dt \\ &= - \int g(t) \varphi'(t) dt && \text{(IBP)} \\ &= - \langle T_g, \varphi' \rangle\end{aligned}$$

# Derivation of a distribution

**For regular distributions:** let  $g$  be a differentiable locally integrable function. Then  $g'$  is also locally integrable and

$$\begin{aligned}\langle T_{g'}, \varphi \rangle &= \int g'(t) \varphi(t) dt \\ &= - \int g(t) \varphi'(t) dt && \text{(IBP)} \\ &= - \langle T_g, \varphi' \rangle\end{aligned}$$

## Definition

Let  $T$  be a distribution. The derivative of  $T$  is the distribution  $T^{(1)}$  defined by

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle T^{(1)}, \varphi \rangle \stackrel{\text{def.}}{=} -\langle T, \varphi' \rangle$$

## A second example: Fourier transform

Let  $g$  be an integrable function (so that  $T_g$  is a regular distribution). Then  $\mathcal{F}g$  is also integrable and

$$\begin{aligned}\langle T_{\mathcal{F}g}, \varphi \rangle &= \int \mathcal{F}g(f) \varphi(f) df \\ &= \int \left( \int g(t) e^{-i2\pi ft} dt \right) \varphi(f) df \\ &= \int g(t) \left( \int \varphi(f) e^{-i2\pi ft} df \right) dt \\ &\stackrel{(?)}{=} \langle T_g, \mathcal{F}\varphi \rangle\end{aligned}$$



# Fourier transform of a distribution (tentative)

## Definition

Let  $T$  be a distribution. The Fourier of  $T$  is the distribution  $\mathcal{F}T$  defined by

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle \mathcal{F}T, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, \mathcal{F}\varphi \rangle$$

# Fourier transform of a distribution (tentative)

## Definition

Let  $T$  be a distribution. The Fourier of  $T$  is the distribution  $\mathcal{F}T$  defined by

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle \mathcal{F}T, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, \mathcal{F}\varphi \rangle$$

- (need complex-valued test functions to technically make sense)

# Fourier transform of a distribution (tentative)

## Definition

Let  $T$  be a distribution. The Fourier of  $T$  is the distribution  $\mathcal{F}T$  defined by

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle \mathcal{F}T, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, \mathcal{F}\varphi \rangle$$

► (need complex-valued test functions to technically make sense)

Unfortunately, in general  $\mathcal{F}\varphi$  is not compactly supported, it may not be properly defined



# Fourier transform of a distribution (tentative)

## Definition

Let  $T$  be a distribution. The Fourier of  $T$  is the distribution  $\mathcal{F}T$  defined by

$$\forall \varphi \in \mathcal{D}(\mathbf{R}), \quad \langle \mathcal{F}T, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, \mathcal{F}\varphi \rangle$$

► (need complex-valued test functions to technically make sense)

Unfortunately, in general  $\mathcal{F}\varphi$  is not compactly supported, it may not be properly defined



The set of distributions is too big a space!!

# Tempered distribution

We will **extend** the set of test functions...



# Tempered distribution

We will **extend** the set of test functions...



... which will **reduce** the set of distributions (by duality).

# Tempered distributions

## Definition: Schwartz space

We say that a function  $g$  is a Schwartz function if it is infinitely differentiable and for all  $k, q \in \mathbb{N}$

$$\lim_{t \rightarrow \pm\infty} |t^k \varphi^{(q)}(t)| = 0$$

The set of Schwartz functions is denoted  $\mathcal{S}(\mathbb{R})$ .

- Instead of compact support, we take **fast decrease** (faster than any polynomial).

# Tempered distributions

## Definition: Schwartz space

We say that a function  $g$  is a Schwartz function if it is infinitely differentiable and for all  $k, q \in \mathbb{N}$

$$\lim_{t \rightarrow \pm\infty} |t^k \varphi^{(q)}(t)| = 0$$

The set of Schwartz functions is denoted  $\mathcal{S}(\mathbb{R})$ .

- Instead of compact support, we take **fast decrease** (faster than any polynomial).

## Definition: Tempered distribution

We call “**tempered distribution**” any element of the (topological) dual space of the Schwartz space  $\mathcal{S}'(\mathbb{R})$ .

“Topological”: **continuous** linear functionals



# Properties

- Convergence in  $\mathcal{S}(\mathbf{R})$  is defined through the family of semi-norms  $\mathcal{N}_p(\varphi) = \sum_{k,q \leq p} \sup_t |t^q \varphi^{(q)}(t)|$ , i.e.  $\varphi_n$  converge to  $\varphi$  if  $\forall p, \mathcal{N}_p(\varphi_n - \varphi) \rightarrow 0$

# Properties

- Convergence in  $\mathcal{S}(\mathbf{R})$  is defined through the family of semi-norms  $\mathcal{N}_p(\varphi) = \sum_{k,q \leq p} \sup_t |t^q \varphi^{(q)}(t)|$ , i.e.  $\varphi_n$  converge to  $\varphi$  if  $\forall p, \mathcal{N}_p(\varphi_n - \varphi) \rightarrow 0$
- For a tempered distribution  $T$ , there is  $p, C$  such that  $\forall \varphi \in \mathcal{S}(\mathbf{R}), \langle T, \varphi \rangle \leq C \mathcal{N}_p(\varphi)$  (continuity)

# Properties

- ▶ Convergence in  $\mathcal{S}(\mathbf{R})$  is defined through the family of semi-norms  $\mathcal{N}_p(\varphi) = \sum_{k,q \leq p} \sup_t |t^q \varphi^{(q)}(t)|$ , i.e.  $\varphi_n$  converge to  $\varphi$  if  $\forall p, \mathcal{N}_p(\varphi_n - \varphi) \rightarrow 0$
- ▶ For a tempered distribution  $T$ , there is  $p, C$  such that  $\forall \varphi \in \mathcal{S}(\mathbf{R}), \langle T, \varphi \rangle \leq C \mathcal{N}_p(\varphi)$  (continuity)
- ▶ Any test function is a Schwartz function:  $\mathcal{D}(\mathbf{R}) \subset \mathcal{S}(\mathbf{R})$

# Properties

- ▶ Convergence in  $\mathcal{S}(\mathbf{R})$  is defined through the family of semi-norms  $\mathcal{N}_p(\varphi) = \sum_{k,q \leq p} \sup_t |t^q \varphi^{(q)}(t)|$ , i.e.  $\varphi_n$  converge to  $\varphi$  if  $\forall p, \mathcal{N}_p(\varphi_n - \varphi) \rightarrow 0$
- ▶ For a tempered distribution  $T$ , there is  $p, C$  such that  $\forall \varphi \in \mathcal{S}(\mathbf{R}), \langle T, \varphi \rangle \leq C \mathcal{N}_p(\varphi)$  (continuity)
- ▶ Any test function is a Schwartz function:  $\mathcal{D}(\mathbf{R}) \subset \mathcal{S}(\mathbf{R})$
- ▶ (hence) any tempered distribution is a distribution:  $\mathcal{S}'(\mathbf{R}) \subset \mathcal{D}'(\mathbf{R})$

# Fourier and tempered distributions

## Thm

- ▶ Thm: All tempered distribution **admit a Fourier transform** defined by

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle \mathcal{F}T, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, \mathcal{F}\varphi \rangle$$

- ▶ The Fourier transform of a Schwartz function is **also** a Schwartz function

*Hint: link between differentiability and polynomial decrease of the Fourier transform...*

# Classical tempered distribution



Unlike regular distributions, all functions do not correspond to tempered distributions! Need all derivative to grows “as fast as” polynomials...

# Classical tempered distribution



Unlike regular distributions, all functions do not correspond to tempered distributions! Need all derivative to grows “as fast as” polynomials...

Still:

- ▶ The Dirac function  $\delta_t$  is a tempered distribution
- ▶ The Dirac comb is a tempered distribution
- ▶ the function  $t \mapsto \cos(2\pi ft)$  is a tempered distribution (as all bounded functions)

# Classical tempered distribution



Unlike regular distributions, all functions do not correspond to tempered distributions! Need all derivative to grows “as fast as” polynomials...

Still:

- ▶ The Dirac function  $\delta_t$  is a tempered distribution
- ▶ The Dirac comb is a tempered distribution
- ▶ the function  $t \mapsto \cos(2\pi ft)$  is a tempered distribution (as all bounded functions)

→ that's all we need in this course...?







In signal processing, it is folklore to believe that distributions are always well-behaved

*That it all operations are legitimate*

It it nevertheless not always true



In this lecture (and companion lab) it will nevertheless be true

*Some justifications will be out of scope and left in exercise*

# **Some properties of (tempered) distributions**

## Some Fourier transform (1)

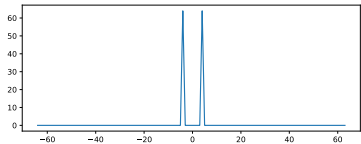
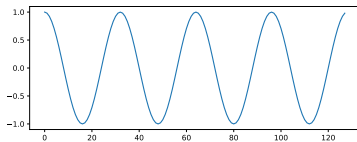
$$\begin{aligned}\langle \mathcal{F}\delta_a, \varphi \rangle &= \langle \delta_a, \mathcal{F}\varphi \rangle \\ &= (\mathcal{F}\varphi)(a) \\ &= \int \varphi(t) e^{-2i\pi at} dt \\ &= \langle T_{t \mapsto e^{-2i\pi at}}, \varphi \rangle\end{aligned}$$

## Some Fourier transform (2)

$$\begin{aligned}\langle \mathcal{F}T_{t \mapsto e^{2i\pi at}}, \varphi \rangle &\stackrel{\text{def.}}{=} \langle T_{t \mapsto e^{2i\pi at}}, \mathcal{F}\varphi \rangle \\ &\stackrel{\text{def.}}{=} \int e^{2i\pi af} \mathcal{F}\varphi(f) df \\ &\stackrel{\text{def.}}{=} \mathcal{F}^{-1} \mathcal{F}\varphi(a) \\ &= \langle \delta_a, \varphi \rangle\end{aligned}$$

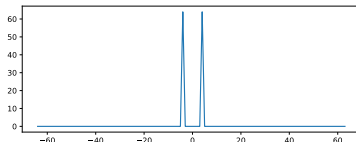
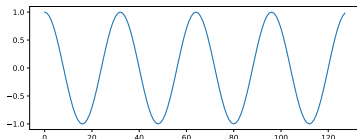
## Some Fourier transform (3)

$$\begin{aligned}\langle \mathcal{F}T_{t \rightarrow \cos(2\pi at)}, \varphi \rangle &= \langle \mathcal{F}T_{t \rightarrow 0.5(e^{2\pi at} + e^{-2\pi at})}, \varphi \rangle \\ &= \frac{1}{2} \langle \mathcal{F}T_{t \rightarrow e^{2\pi at}}, \varphi \rangle + \frac{1}{2} \langle \mathcal{F}T_{t \rightarrow e^{-2\pi at}}, \varphi \rangle \\ &= \frac{1}{2} (\langle \delta_a, \varphi \rangle + \langle \delta_{-a}, \varphi \rangle)\end{aligned}$$



## Some Fourier transform (3)

$$\begin{aligned}\langle \mathcal{F}T_{t \rightarrow \cos(2\pi at)}, \varphi \rangle &= \langle \mathcal{F}T_{t \rightarrow 0.5(e^{2\pi at} + e^{-2\pi at})}, \varphi \rangle \\ &= \frac{1}{2} \langle \mathcal{F}T_{t \rightarrow e^{2\pi at}}, \varphi \rangle + \frac{1}{2} \langle \mathcal{F}T_{t \rightarrow e^{-2\pi at}}, \varphi \rangle \\ &= \frac{1}{2} (\langle \delta_a, \varphi \rangle + \langle \delta_{-a}, \varphi \rangle)\end{aligned}$$



Recall that  $\mathcal{F}[Re(f)](\xi) = \frac{1}{2}(\mathcal{F}f(\xi) + \mathcal{F}f(-\xi))$



Have we achieved our initial objective?



Have we achieved our initial objective?



It is folklore to provide a graphical representation of a Fourier transform



## Another example: the “door” distribution

We define  $\Pi_{[-\frac{a}{2}, \frac{a}{2}]}$  such that

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle \Pi_{[-\frac{a}{2}, \frac{a}{2}]}, \varphi \rangle = \int_{-\frac{a}{2}}^{\frac{a}{2}} \varphi(t) dt$$

**Question:** Evaluate the Fourier transform of  $\Pi_{[-\frac{a}{2}, \frac{a}{2}]}$ .

## Another example: the “door” distribution

We define  $\Pi_{[-\frac{a}{2}, \frac{a}{2}]}$  such that

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle \Pi_{[-\frac{a}{2}, \frac{a}{2}]}, \varphi \rangle = \int_{-\frac{a}{2}}^{\frac{a}{2}} \varphi(t) dt$$

**Question:** Evaluate the Fourier transform of  $\Pi_{[-\frac{a}{2}, \frac{a}{2}]}$ .

$$\begin{aligned} \langle \mathcal{F}\Pi_{[-\frac{a}{2}, \frac{a}{2}]}, \varphi \rangle &= \langle \Pi_{[-\frac{a}{2}, \frac{a}{2}]}, \mathcal{F}\varphi \rangle \\ &\stackrel{\text{Fub.}}{=} \int \varphi(t) \int_{-\frac{a}{2}}^{\frac{a}{2}} e^{-2ift} df dt \\ &= \langle T_{a \operatorname{sinc}(a \cdot)}, \varphi \rangle \end{aligned}$$



Does it ring a bell?

# Compatibility with the “standard” Fourier transform

## Theorem (admitted)

If  $x \in \mathcal{L}^1(\mathbf{R})$  (or  $\mathcal{L}^2(\mathbf{R})$ ) then

$$\mathcal{F}T_x = T_{\mathcal{F}x}$$

# Towards an inverse Fourier transform

## Definition: Conjugate Fourier transform

Let  $T$  be a tempered distribution. The conjugate Fourier of  $T$  is the distribution  $\overline{\mathcal{F}}T$  defined by

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle \overline{\mathcal{F}}T, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, \overline{\mathcal{F}}\varphi \rangle$$

# Towards an inverse Fourier transform

## Definition: Conjugate Fourier transform

Let  $T$  be a tempered distribution. The conjugate Fourier of  $T$  is the distribution  $\overline{\mathcal{F}}T$  defined by

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle \overline{\mathcal{F}}T, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, \overline{\mathcal{F}}\varphi \rangle$$

## Theorem: Inversion formula

Let  $T$  be a tempered distribution.  
Then, its Fourier transform satisfies

$$T = \overline{\mathcal{F}}[\mathcal{F}T]$$

In other words, the standard inversion formula holds in the sense of tempered distributions

# Convolution

## Definition

Let  $T_1, T_2$ .

Their convolution (if it exists) is the distribution  $T_1 * T_2$  defined as

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle T_1 * T_2, \varphi \rangle \stackrel{\text{def.}}{=} \langle T_1, \varphi_{T_2} \rangle$$

where

$$\forall t \in \mathbf{R}, \quad \varphi_{T_2}(t) = \langle T_2, \varphi(\cdot + t) \rangle$$

*( $T_1, T_2$  can be exchanged in the definition)*

**Example:** of  $T_1 = \delta_a$

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle \delta_a * T_2, \varphi \rangle = \langle T_2, \varphi(\cdot - a) \rangle$$

→ Translation of the distribution

# Support of a distribution

## Definition

Let  $T \in \mathcal{S}(\mathbf{R})'$ .

Define  $\Omega$  as the largest open set of  $\mathbf{R}$  such that

$$\text{support}(\varphi) \subseteq \Omega \implies \langle T, \varphi \rangle = 0$$

By extension, the (closed) set  $\text{support}(T) \stackrel{\text{def.}}{=} \Omega^c$  is called the **support** of the distribution



# Support of a distribution

## Definition

Let  $T \in \mathcal{S}(\mathbf{R})'$ .

Define  $\Omega$  as the largest open set of  $\mathbf{R}$  such that

$$\text{support}(\varphi) \subseteq \Omega \implies \langle T, \varphi \rangle = 0$$

By extension, the (closed) set  $\text{support}(T) \stackrel{\text{def.}}{=} \Omega^c$  is called the **support** of the distribution

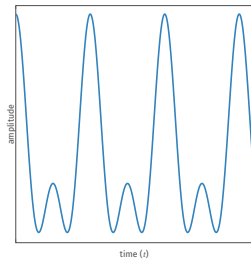
## Examples:

- ▶  $\text{support}(\delta_a) = \{a\}$
- ▶  $\text{support}(\delta_a + \delta_{-a}) = \{-a, a\}$

# Back to our sampling problem

*(with new ingredients)*

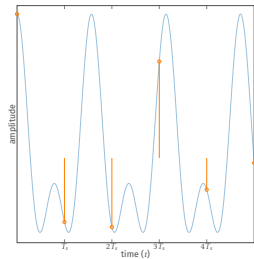
# Ideal Sampling



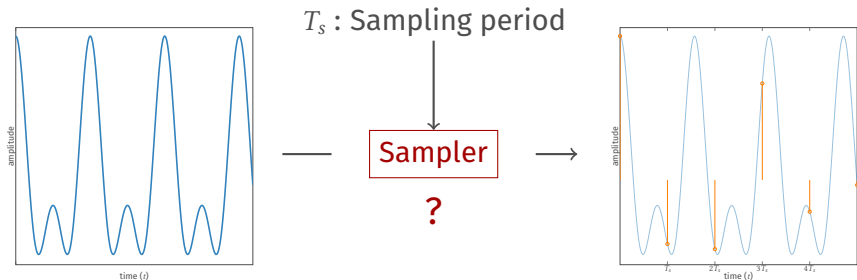
$T_s$  : Sampling period

Sampler

?



# Ideal Sampling



**Observation model:**  $\equiv \mathbf{x} \times \text{Dirac comb}$  with period  $T_s$ :

$$x_e \stackrel{\text{def.}}{=} \sum_{j \in \mathbb{Z}} x(jT_s) \delta_{jT_s} \stackrel{?}{=} \text{III}_{T_s} x$$

## Such a definition is sounded

Is the distribution  $\sum_{j \in \mathbb{Z}} x(jT_s) \delta_{jT_s}$  sounded?



Said differently, does the “distribution-function product”  $\text{III}_{T_s} x$  mean something?

## Such a definition is sounded



Is the distribution  $\sum_{j \in \mathbb{Z}} x(jT_s) \delta_{jT_s}$  sounded?

Said differently, does the “distribution-function product”  $\text{III}_{T_s} x$  mean something?

### Theorem

Let  $T \in \mathcal{S}(\mathbf{R})'$  and  $x \in \mathcal{C}^\infty(\mathbf{R})$  be such that

$$\forall r \geq 0, \exists C_r > 0 \text{ and } n_r \in \mathbf{N} : \quad |x^{(r)}(t)| \leq C_r(1 + |t|)^{n_r}.$$

Then the product  $Tx$  defined as

$$\forall \varphi \in \mathcal{S}(\mathbf{R}), \quad \langle Tx, \varphi \rangle \stackrel{\text{def.}}{=} \langle T, x\varphi \rangle$$

is a tempered distribution

Since  $x\varphi \in \mathcal{S}(\mathbf{R})$ .

The set of such functions is denoted  $\mathcal{O}(\mathbf{R})$

# Wrapping up: main ingredients

1. **The sensing operator:** The (linear) operator defined by

$$\begin{array}{ccc} M: \mathcal{O}(\mathbf{R}) \longrightarrow \mathbf{R}^{\mathbf{Z}} & & M: \mathcal{O}(\mathbf{R}) \longrightarrow \mathcal{S}(\mathbf{R})' \\ \textcolor{blue}{x} \longmapsto \{\textcolor{brown}{x}(jT_s)\}_{j \in \mathbf{Z}} & \text{or} & \textcolor{blue}{x} \longmapsto \textcolor{brown}{\mathbb{I}\mathbb{I}\mathbb{I}}_{T_s} \textcolor{brown}{x} \end{array}$$

2. **The target set:**

3. **The decoder:**

4. **The accuracy criterion:** “noiseless exact recovery”, i.e.

$$\forall \textcolor{blue}{x} \in \mathcal{X}_{\text{target}}, \quad D(M(\textcolor{blue}{x})) = \textcolor{blue}{x}$$

# Spectrum of a sampled signal (following)

**Question:** Does the sampling process affect the spectrum?

$$\mathcal{F}[\text{III}_{T_s} x] =$$



## Two (rather technical) admitted results

### Fourier transform of a product (*admitted*)

Let  $x$  be such that  $T_x$  has a compactly supported distribution.  
Then

$$\mathcal{F}[\text{III}_{T_s} x] = \mathcal{F}[\text{III}_{T_s}] * \mathcal{F}[T_x]$$

(The result holds under broader assumptions)

## Two (rather technical) admitted results

### Fourier transform of a product (*admitted*)

Let  $x$  be such that  $T_x$  has a compactly supported distribution.  
Then

$$\mathcal{F}[\text{III}_{T_s} x] = \mathcal{F}[\text{III}_{T_s}] * \mathcal{F}[T_x]$$

(The result holds under broader assumptions)

### Fourier transform of a dirac comb (*admitted*)

$$\mathcal{F}[\text{III}_{T_s}] = \frac{1}{T_s} \text{III}_{\frac{1}{T_s}}$$

# Spectrum of a sampled signal (following)

**Question:** Does the sampling process affect the spectrum?

$$\mathcal{F}[\text{III}_{T_s} x] \stackrel{\text{Result 1.}}{=} \mathcal{F}[\text{III}_{T_s}] * \mathcal{F}[T_x]$$

# Spectrum of a sampled signal (following)

**Question:** Does the sampling process affect the spectrum?

$$\begin{aligned}\mathcal{F}[\text{III}_{T_s} x] &\stackrel{\text{Result 1.}}{=} \mathcal{F}[\text{III}_{T_s}] * \mathcal{F}[T_x] \\ &\stackrel{\text{Result 2.}}{=} F_s \text{III}_{F_s} * \mathcal{F}[T_x]\end{aligned}$$

where  $F_s = \frac{1}{T_s}$  is the sampling frequency

# Spectrum of a sampled signal (following)

**Question:** Does the sampling process affect the spectrum?

$$\begin{aligned}\mathcal{F}[\text{III}_{T_s} x] &\stackrel{\text{Result 1.}}{=} \mathcal{F}[\text{III}_{T_s}] * \mathcal{F}[T_x] \\ &\stackrel{\text{Result 2.}}{=} F_s \text{III}_{F_s} * \mathcal{F}[T_x] \\ &\stackrel{\text{Distribu.}}{=} F_s \sum_{j \in \mathbb{Z}} (\delta_{jF_s} * \mathcal{F}[T_x])\end{aligned}$$

where  $F_s = \frac{1}{T_s}$  is the sampling frequency

# Spectrum of a sampled signal (following)

**Question:** Does the sampling process affect the spectrum?

$$\begin{aligned}\mathcal{F}[\text{III}_{T_s} x] &\stackrel{\text{Result 1.}}{=} \mathcal{F}[\text{III}_{T_s}] * \mathcal{F}[T_x] \\ &\stackrel{\text{Result 2.}}{=} F_s \text{III}_{F_s} * \mathcal{F}[T_x] \\ &\stackrel{\text{Distribu.}}{=} F_s \sum_{j \in \mathbb{Z}} (\delta_{jF_s} * \mathcal{F}[T_x])\end{aligned}$$

where  $F_s = \frac{1}{T_s}$  is the sampling frequency

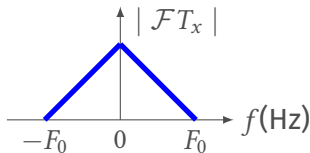
Sampling in the time domain  $\implies$  Periodic spectrum

# Spectrum of band-limited signals

## Band-limited signals

A band-limited signal  $x$  is a signal whose Fourier transform density has **bounded support**

Let  $\mathbf{x}$  be a band-limited signal whose TF has support  $[-F_0, F_0]$

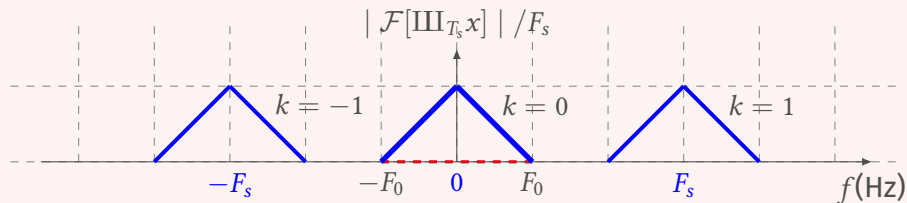


**Question** : Spectrum of  $\text{III}_{T_s} \mathbf{x}$ ?

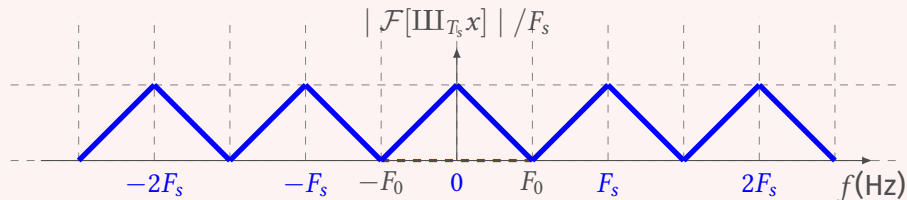
$$\mathcal{F}[\text{III}_{T_s} \mathbf{x}] = F_s \sum_{j=-\infty}^{+\infty} \delta_{jF_s} * \mathcal{F}T_x$$

# Spectrum of a band-limited signal

Case 1:  $F_s > 2F_{\max}$



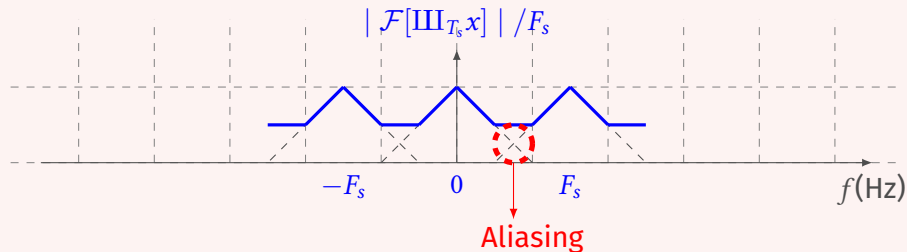
Case 2:  $F_s = 2F_{\max}$





# Spectrum of a band-limited signal

Case 3:  $F_s < 2F_{\max}$



The signal is **deteriorated**  $\implies$  **Loss** of information

# Shannon-Nyquist sampling theorem

**Question:** Necessary condition on  $F_s$  such that no information is lost?

## Nyquist-Shannon sampling theorem

If a function  $x(t)$  contains no frequencies higher than  $F_{\max}$  hertz, it is **completely determined** by giving its ordinates at a series of points spaced  $T_s \geq \frac{1}{2F_{\max}}$  seconds apart.

—→ **Sufficient condition**

$F_s = 2F_{\max}$  is dubbed “**Nyquist frequency**”

# Interpolation formula

## Interpolation formula

If  $F_s \geq 2F_{\max}$ , then  $T_x$  is equal to the regular distribution spanned by

$$t \longmapsto \sum_{j \in \mathbb{Z}} x(jT_s) \operatorname{sinc}(F_s(t - jT_s))$$

**(Informal) proof:** if  $F_s > 2F_{\max}$  the spectrum of  $x$  can be recovered with an ideal **low pass filter**



This operation (operating on the spectrum) is common in signal processing and often referred to as “filtering”

# Interpolation formula

## Interpolation formula

If  $F_s \geq 2F_{\max}$ , then  $T_x$  is equal to the regular distribution spanned by

$$t \longmapsto \sum_{j \in \mathbb{Z}} x(jT_s) \operatorname{sinc}(F_s(t - jT_s))$$

**(Informal) proof:** if  $F_s > 2F_{\max}$  the spectrum of  $x$  can be recovered with an ideal **low pass filter**

$$\mathcal{F}T_x \stackrel{\mathcal{S}(\mathbb{R})'}{=} \mathcal{F}[\text{III}_{T_s}x] \cdot \frac{1}{F_s} \Pi_{\frac{F_s}{2}}(f)$$

$\Rightarrow$

$$T_x \stackrel{\mathcal{S}(\mathbb{R})'}{=} \text{III}_{T_s}x * T_{\operatorname{sinc}(2\pi F_s \cdot)}$$

$$\stackrel{\mathcal{S}(\mathbb{R})'}{=} \sum_{j \in \mathbb{Z}} x[jT_s] \delta_{jT_s} * T_{\operatorname{sinc}(2\pi F_s \cdot)}$$

# Wrapping up: main ingredients

1. **The sensing operator:** The (linear) operator defined by

$$\begin{aligned} M: \mathcal{O}(\mathbf{R}) &\longrightarrow \mathbf{R}^{\mathbf{Z}} & M: \mathcal{O}(\mathbf{R}) &\longrightarrow \mathcal{S}(\mathbf{R})' \\ \textcolor{blue}{x} &\longmapsto \{\textcolor{brown}{x}(jT_s)\}_{j \in \mathbf{Z}} & \text{or} & \textcolor{blue}{x} \longmapsto \textcolor{brown}{\mathbb{I}\hspace{-1pt}\mathbb{I}}_{T_s} \textcolor{brown}{x} \end{aligned}$$

2. **The target set:** The set of functions which admit a Fourier transform which is compactly supported in  $[-F_s, F_s]$

3. **The decoder:**

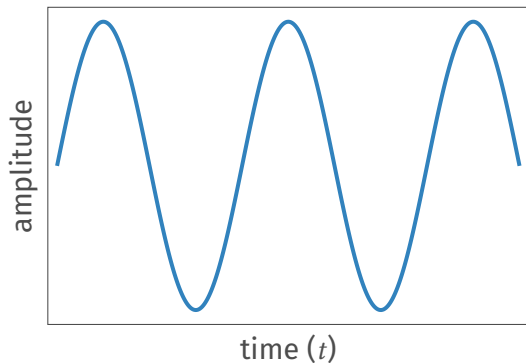
$$\begin{aligned} M: \quad \textcolor{brown}{\mathbf{R}^{\mathbf{Z}}} &\longrightarrow \mathcal{S}' \\ \{\textcolor{brown}{y}_j\}_{j \in \mathbf{Z}} &\longmapsto \sum_{j \in \mathbf{Z}} y_j \text{sinc}(F_s(\cdot - jT_s)) \quad \text{if} \quad F_s \geq 2F_{\max} \end{aligned}$$

4. **The accuracy criterion:** “noiseless exact recovery”, i.e.

$$\forall \textcolor{blue}{x} \in \mathcal{X}_{\text{target}}, \quad D(M(\textcolor{blue}{x})) \stackrel{\mathcal{D}(\mathbf{R})'}{=} T_x$$

# Conclusion

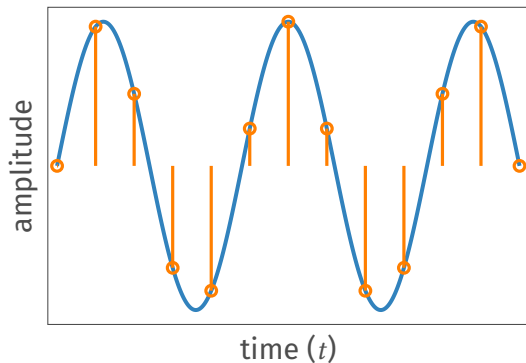
# Numerical illustration



- $x$ :  $F_{\max} = 0.5$  and  
 $x(t) = \cos(2\pi F_{\max} t)$



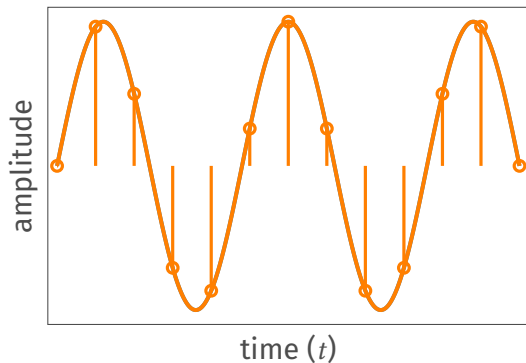
# Numerical illustration



►  $x$ :  $F_{\max} = 0.5$  and  
 $x(t) = \cos(2\pi F_{\max} t)$

►  $T_s = 0.4 \implies$   
 $F_s = 2.5 > 2F_{\max}$

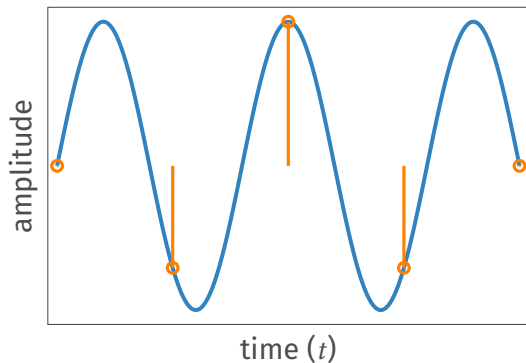
# Numerical illustration



►  $x$ :  $F_{\max} = 0.5$  and  
 $x(t) = \cos(2\pi F_{\max} t)$

►  $T_s = 0.4 \implies$   
 $F_s = 2.5 > 2F_{\max}$

# Numerical illustration

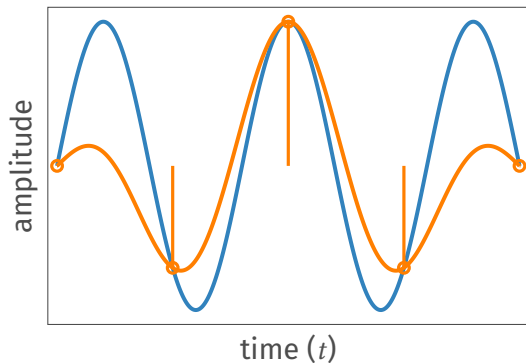


►  $x : F_{\max} = 0.5$  and  
 $x(t) = \cos(2\pi F_{\max} t)$

►  $T_s = 0.4 \implies$   
 $F_s = 2.5 > 2F_{\max}$

►  $T_s = 1.2 \implies$   
 $F_s = .8 < 2F_{\max}$  ⚠

# Numerical illustration



►  $x : F_{\max} = 0.5$  and  
 $x(t) = \cos(2\pi F_{\max} t)$

►  $T_s = 0.4 \implies$   
 $F_s = 2.5 > 2F_{\max}$

►  $T_s = 1.2 \implies$   
 $F_s = .8 < 2F_{\max}$  ⚠



Are we really particularizing the linear model?

*This is a legitimate question*



Are we really particularizing the linear model?

*This is a legitimate question*

**Rappels:** a linear model is characterized by

✓ a linear observation operator operator  $M$  ✓; *Already done*





Are we really particularizing the linear model?

*This is a legitimate question*

**Rappels:** a linear model is characterized by

- ✓ a linear observation operator operator  $M$  ✓; *Already done*
- ✓ a vector space as choices for  $\mathcal{X}_{\text{target}}$ ; *By linearity of the Fourier transform*





Are we really particularizing the linear model?

*This is a legitimate question*



**Rappels:** a linear model is characterized by

- ✓ a linear observation operator operator  $M$  ✓; *Already done*
- ✓ a vector space as choices for  $\mathcal{X}_{\text{target}}$ ; *By linearity of the Fourier transform*
- (an orthogonal projection as “best” decoder)  
*In the finite-dimensional case*



# A concluding remark

If we restrict ourself to signals  $x \in \mathcal{L}^2(\mathbf{R})$

- $\mathcal{L}^2(\mathbf{R})$  is a Hilbert space equipped with the inner product

$$\langle x_1, x_2 \rangle = \int x_1(t) \overline{x_2(t)} dt$$

- The family of function

$$\forall j, \quad m_j : t \mapsto \frac{1}{\sqrt{T_s}} \operatorname{sinc}\left(\frac{t - jT_s}{T_s}\right)$$

is an orthonormal basis of the set of functions with Fourier transform supported in  $[-F_s, F_s]$

- For all signals  $x$  with Fourier transform supported in  $[-F_s, F_s]$ ,

$$\langle m_j, x \rangle = x(jT_s)$$

- The proposed decoder thus rewrites

$$t \mapsto \sum_{j \in \mathbf{Z}} \langle m_j, x \rangle m_j(t)$$



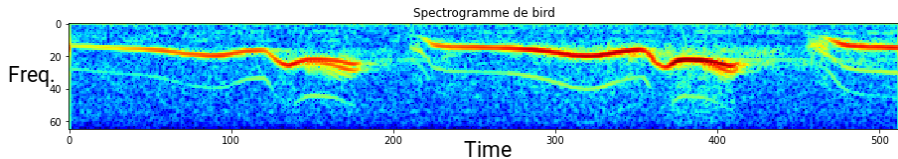
It is an orthogonal projection!

# **Bonus: Time-Frequency analysis, wavelets**

# Time-frequency analysis

- In most domains (audio processing...), the frequency spectrum *itself* evolves in time.

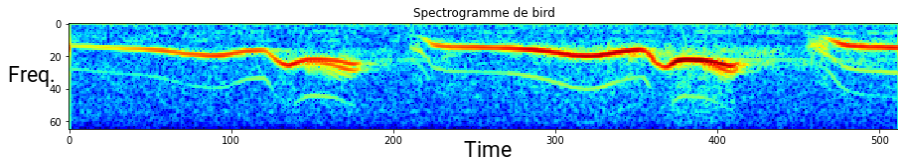
Example: audio *spectrogram*



# Time-frequency analysis

- ▶ In most domains (audio processing...), the frequency spectrum *itself* evolves in time.
- ▶ We want to analyze the frequency component *localized in time*, hence “time-frequency analysis”

Example: audio spectrogram



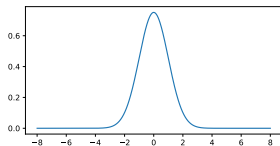
# Windowed Fourier transform

- Popular tool in audio: **Windowed Fourier Transform**, also called **Short Time Fourier Transform** (STFT) in its discrete version.

# Windowed Fourier transform

- Popular tool in audio: **Windowed Fourier Transform**, also called **Short Time Fourier Transform** (STFT) in its discrete version.

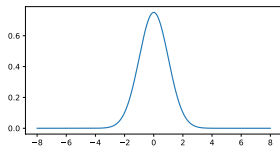
Define a **symmetric window**  $g(t) = g(-t)$   
with  $\|g\|_{L^2} = 1$



# Windowed Fourier transform

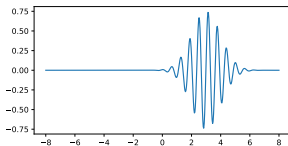
- Popular tool in audio: **Windowed Fourier Transform**, also called **Short Time Fourier Transform (STFT)** in its discrete version.

Define a **symmetric window**  $g(t) = g(-t)$   
with  $\|g\|_{L^2} = 1$



**Time-frequency atoms** are localized in **time and frequency**:

$$g_{u,\xi}(t) = \underbrace{e^{2i\pi\xi t}}_{\text{Freq. } \xi} \underbrace{g(t-u)}_{\text{Time } u}$$





# Windowed Fourier transform

The **Windowed Fourier Transform** is defined as

$$Sx(u, \xi) = \langle x, g_{u, \xi} \rangle_{L^2} = \int x(t) g(t - u) e^{-2i\pi \xi t} dt$$

# Windowed Fourier transform

The **Windowed Fourier Transform** is defined as

$$Sx(u, \xi) = \langle x, g_{u, \xi} \rangle_{L^2} = \int x(t) g(t - u) e^{-2i\pi \xi t} dt$$

- Analyze the signal  $x$  around time  $u$  and frequency  $f$

# Windowed Fourier transform

The **Windowed Fourier Transform** is defined as

$$Sx(u, \xi) = \langle x, g_{u, \xi} \rangle_{L^2} = \int x(t) g(t - u) e^{-2i\pi \xi t} dt$$

- Analyze the signal  $x$  **around time  $u$  and frequency  $f$**

By Parseval's identity,

$$Sx(u, \xi) = \langle \hat{x}, \hat{g}_{u, \xi} \rangle_{L^2}$$

and by the shift formula

$$\hat{g}_{u, \xi}(f) = \hat{g}(f - \xi) e^{-2i\pi u(f - \xi)}$$

- since  $g$  is real symmetric,  $\hat{g}$  is real symmetric (ex: for  $g$  Gaussian,  $\hat{g}$  is Gaussian!)
- Also performs a windowed transform of  $\hat{x}$  around  $\xi$  (and "inverse frequency"  $u$ )

# Heisenberg boxes

$$Sx(u, \xi) = \langle x, g_{u, \xi} \rangle_{L^2} = \langle \hat{x}, \hat{g}_{u, \xi} \rangle_{L^2}$$

# Heisenberg boxes

$$Sx(u, \xi) = \langle x, g_{u, \xi} \rangle_{L^2} = \langle \hat{x}, \hat{g}_{u, \xi} \rangle_{L^2}$$

Considering  $\mathbb{P} = |g_{u, \xi}|^2$  and  $\hat{\mathbb{P}} = |\hat{g}_{u, \xi}|^2$  as **probabilities densities** (since  $\|g_{u, \xi}\|_{L^2} = \|\hat{g}_{u, \xi}\|_{L^2} = 1$ ):

$$\mathbb{E}(\mathbb{P}) = u$$

$$\text{Var}(\mathbb{P}) = \sigma_{\mathbb{P}}^2$$

$$\mathbb{E}(\hat{\mathbb{P}}) = \xi$$

$$\text{Var}(\hat{\mathbb{P}}) = \sigma_{\hat{\mathbb{P}}}^2$$

# Heisenberg boxes

$$Sx(u, \xi) = \langle x, g_{u, \xi} \rangle_{L^2} = \langle \hat{x}, \hat{g}_{u, \xi} \rangle_{L^2}$$

Considering  $\mathbb{P} = |g_{u, \xi}|^2$  and  $\hat{\mathbb{P}} = |\hat{g}_{u, \xi}|^2$  as **probabilities densities** (since  $\|g_{u, \xi}\|_{L^2} = \|\hat{g}_{u, \xi}\|_{L^2} = 1$ ):

$$\mathbb{E}(\mathbb{P}) = u$$

$$\text{Var}(\mathbb{P}) = \sigma_{\mathbb{P}}^2$$

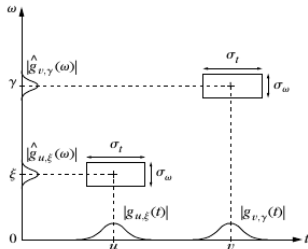
$$\mathbb{E}(\hat{\mathbb{P}}) = \xi$$

$$\text{Var}(\hat{\mathbb{P}}) = \sigma_{\hat{\mathbb{P}}}^2$$

and by Heisenberg's thm:

$$\sigma_{\mathbb{P}}^2 \sigma_{\hat{\mathbb{P}}}^2 \geq 1/4$$

with equality **if and only if**  $g$  is (any) **Gaussian**.



# Reconstruction

## Thm: inversion formula

If  $x \in \mathcal{L}^2$ ,

$$x(t) = \iint Sx(u, \xi) g(t - u) e^{2i\pi\xi t} d\xi du$$

and

$$\|x\|_{L^2}^2 = \iint |Sx(u, \xi)|^2 d\xi du$$

# Reconstruction

## Thm: inversion formula

If  $x \in \mathcal{L}^2$ ,

$$x(t) = \iint Sx(u, \xi) g(t - u) e^{2i\pi\xi t} d\xi du$$

and

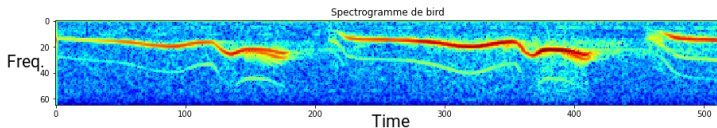
$$\|x\|_{L^2}^2 = \iint |Sx(u, \xi)|^2 d\xi du$$

- ▶ no information is lost (true "transform")
- ▶  $Sx \in \mathcal{L}^2(\mathbf{R}^2)$



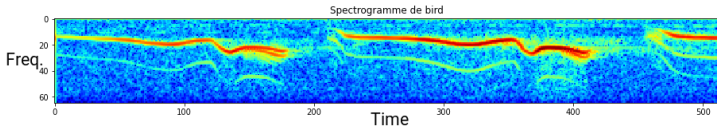
# Example of applications

**Spectrogram**  $|Sx(u, \xi)|^2$  used a lot in audio.



# Example of applications

**Spectrogram**  $|Sx(u, \xi)|^2$  used a lot in audio.



Ex: wav2midi with sparse matrix factorization.

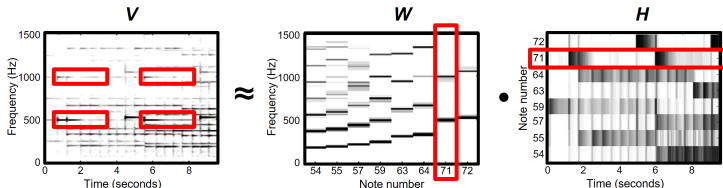


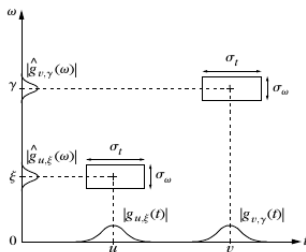
Figure 8.21c-d from [Müller, FMP, Springer 2015]

More precisely, "Non-negative matrix factorization" (NMF), a low-rank factorization of a non-neg. matrix (the spectrogram) into non-neg. components, frequency/activation. See <https://www.audiolabs-erlangen.de/resources/MIR/FMP/C8/C8S3-NMFbasic.html>

# Problem with WFT



- ▶ The WFT analyzes content with the same window length
- ▶ Element can have different durations... (especially low-frequency)
- ▶ Human hearing is spread “logarithmically” in frequency



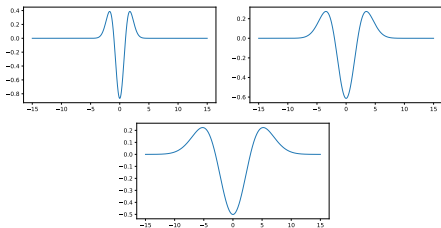
# Wavelet transform

A wavelet is a normalized function  $\psi \in \mathcal{L}^2(\mathbf{R})$  with a zero average:

$$\int \psi(t) dt = 0, \quad \|\psi\| = 1$$

Time-frequency atoms are obtained by *scaling* the wavelet:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$$



Wavelet transform:

$$Wx(u, s) = \langle x, \psi_{u,s} \rangle$$

# Heisenberg boxes

Assume  $\psi$  is analytic for simplicity (i.e. zero negative frequency, see Mallat book).

# Heisenberg boxes

Assume  $\psi$  is analytic for simplicity (i.e. zero negative frequency, see Mallat book).

Considering  $\mathbb{P} = |\psi_{u,s}|^2$  and  $\hat{\mathbb{P}} = |\hat{\psi}_{u,s}|^2$  as **probabilities densities**, and  $\mathbb{P}_1, \hat{\mathbb{P}}_1$  the same for reference  $s = 1$  ( $u$  does not matter). Then:

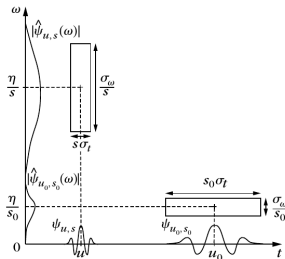
$$\mathbb{E}(\mathbb{P}) = u$$

$$\text{Var}(\mathbb{P}) = s^2 \text{Var}(\mathbb{P}_1)$$

$$\mathbb{E}(\hat{\mathbb{P}}) = \mathbb{E}(\hat{\mathbb{P}}_1)/s$$

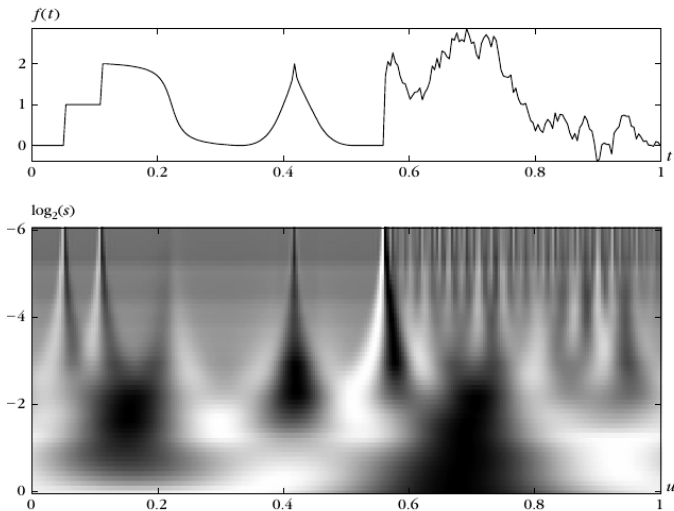
$$\text{Var}(\hat{\mathbb{P}}) = \text{Var}(\hat{\mathbb{P}}_1)/s^2$$

The “right” multiplicative scale! (for human ears)



# Scalogram

$|W_x|^2$  is called a scalogram.



## Beyond...

- Reconstruction theorem for Wavelets
- Discrete Wavelets
- Fast computation (FFT, aka "**the most important algorithm of the XX<sup>th</sup> century**", fast hierarchical wavelets)
- multi-D wavelets (especially 2D!)
- etc.

