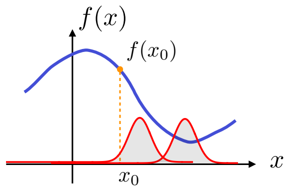
Introduction to functional analysis and distribution theory

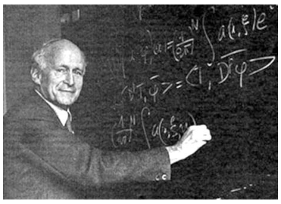
Michael Unser
Biomedical Imaging Group
EPFL, Lausanne, Switzerland

Supplementary material for Signals and Systems—October 2022

TABLE OF CONTENT

Functions vs. distributions





Laurent Schwartz, Fields Medal 1950

- Classical perspective

  Signal as a function $f : \mathbb{R} \to \mathbb{R},\ x \mapsto f(x)$

- Distributional perspective $\boxed{f : \mathcal{S} \to \mathbb{R},\ \varphi \mapsto f(\varphi)}$

  Physical measurement: $\langle f, \varphi \rangle = \int_{\mathbb{R}} f(x)\varphi(x)\mathrm{d}x \quad \in \mathbb{R}$

  Signal seen through a family of **linear** sensors $f : \varphi_i \mapsto f(\varphi_i) = \langle f, \varphi_i \rangle$

  $$\langle f, a_1\varphi_1 + a_2\varphi_2 \rangle = a_1\langle f, \varphi_1 \rangle + a_2\langle f, \varphi_2 \rangle$$

  Continuity of the measurements with respect to variations in φ_i:

  $$\lim_i \varphi_i = \varphi \quad \Rightarrow \quad \lim_i \langle f, \varphi_i \rangle = \langle f, \lim_i \varphi_i \rangle = \langle f, \varphi \rangle$$

- Notion of weak equality

  $f = g \quad \Leftrightarrow \quad \langle f, \varphi \rangle = \langle g, \varphi \rangle$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$

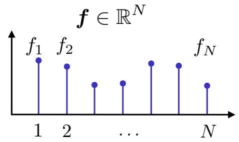
  "If it looks like a duck, swims like a duck, and quacks like a duck, then it is (weakly) a duck"

2. TOPOLOGICAL VECTOR SPACES

- Norms and semi-norms
- Cauchy/converging sequences
- Normed/Banach spaces
- Schwartz' space of test functions
- Linear operators

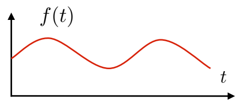
Vector space: algebraic structure

A vector space is a set of elements (vectors) that can be added and multiplied by arbitrary scalars.



$$\alpha f + \beta g \in \mathcal{X} \quad \text{for all } f, g \in \mathcal{X} \text{ and } \alpha, \beta \in \mathbb{R} \text{ (or } \mathbb{C})$$

Addition and scalar multiplication satisfy the usual rules of commutativity, associativity, and distributivity.

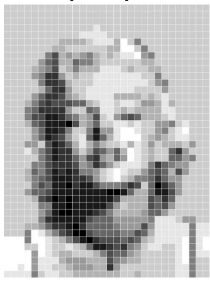


The elements of a vector space are functions $f : \mathbb{E} \to \mathbb{R}$ (or $\mathbb{C}$) defined on a **common index set** $\mathbb{E}$. Depending on the nature of $\mathbb{E}$, these can be:

- Vectors ($\mathbb{E} = \{1, \dots, N\}$) often denoted by $\mathbf{f} = (f_1, \cdots, f_N)$
- Discrete signals (with $\mathbb{E} = \mathbb{Z}$) denoted by $f[k]$ or $f[\cdot]$
- Digital images (with $\mathbb{E} = \mathbb{Z}^2$) denoted by $f[\boldsymbol{k}] = f[k_1, k_2]$ or $f[\cdot]$
- Analog signals (with $\mathbb{E} = \mathbb{R}$) denoted by $f(t)$ or $f(\cdot)$
- Continuous-domain images (with $\mathbb{E} = \mathbb{R}^2$) denoted by $f(\boldsymbol{x}) = f(x_1, x_2)$ or $f(\cdot)$

$\vdots$

$f[k_1, k_2]$



Topological structure: norms and semi-norms

The topology of a vector space specifies the neighborhood(s) of a point; it is a component that allows one to assess the proximity of points and to define a corresponding notion of convergence (continuity).

The topology of the primary vector spaces (Hilbert and Banach) is specified by a norm.

Definition

A **semi-norm** on the vector space $\mathcal{X}$ is a map $\|\cdot\| : \mathcal{X} \to \mathbb{R}^+$ satisfying

1) $\|\alpha f\| = |\alpha| \, \|f\|$ (Homogeneity)

2) $\|f + g\| \le \|f\| + \|g\|$ (Triangle inequality)

for all $f, g \in \mathcal{X}$ and $\alpha \in \mathbb{R}$ (or $\mathbb{C}$). If, in addition, it holds that

3) $\|f - g\| = 0 \Leftrightarrow f = g$ (Ability to separate points)

then $\|\cdot\|$ is called a **norm**.

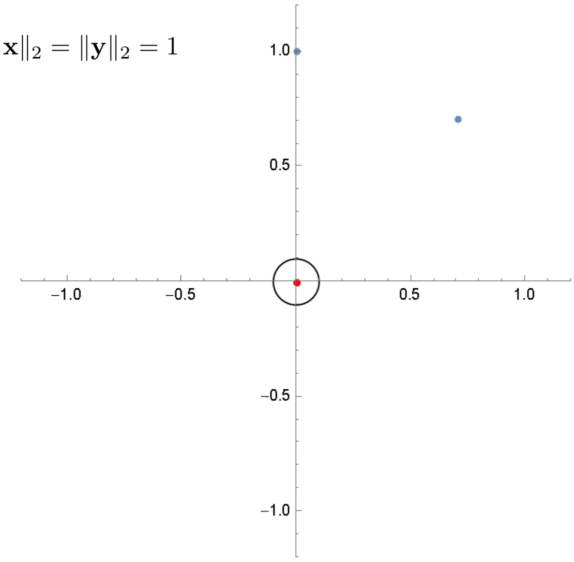
- Examples for $\mathcal{X} = \mathbb{R}^N$ with $\mathbf{x} = (x_1, \dots, x_N)$
  - Euclidean (or ℓ_2) norm: $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_N^2}$
  - Absolute (or ℓ_1) norm: $\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_N|$

Comparison of topologies

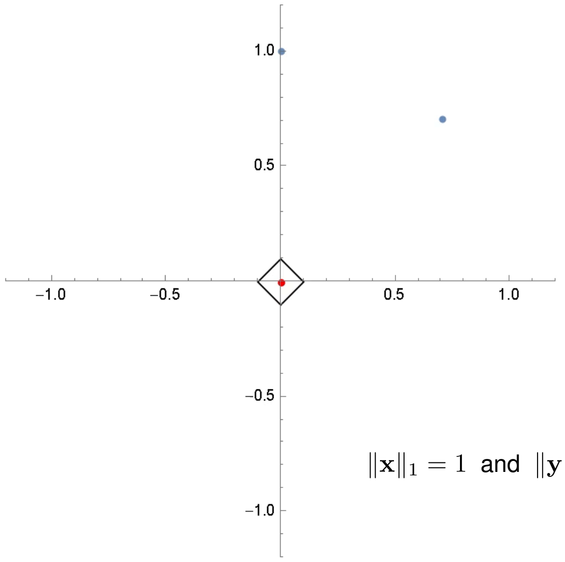
$\mathbf{x} = (0, 1)$ and $\mathbf{y} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$

$\forall \mathbf{x} \in \mathbb{R}^d,\ \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$

$\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$



ℓ_2 (or Euclidean) norm

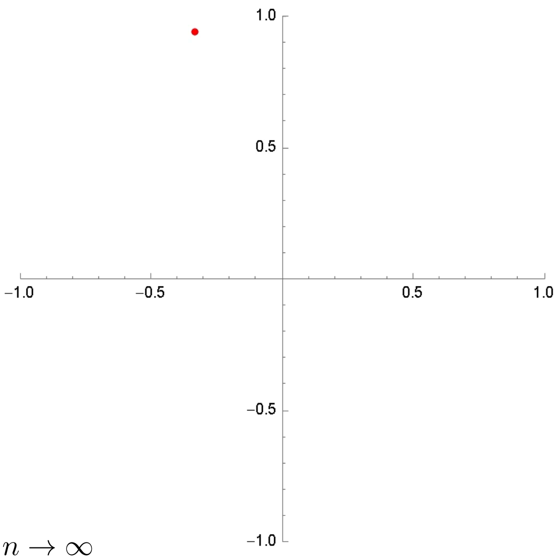


$\|\mathbf{x}\|_1 = 1$ and $\|\mathbf{y}\|_1 = \sqrt{2}$

ℓ_1 norm

■ Example of Cauchy (converging) sequence in $\mathbb{R}^2$

$(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ with $\boldsymbol{x}_n \in \mathbb{R}^2$ and $\boldsymbol{x}_n \to \mathbf{0}$ as $n \to \infty$



■ Assessment of continuity

$\lim_n f(x_n) = f(\lim_n x_n) = f(x)$ with $x_n \to x$ as $n \to \infty$

Cauchy and converging sequences

Normed space: vector space $\mathcal{X}$ equipped with a norm, which is denoted by $(\mathcal{X}, \|\cdot\|)$.

Definition
Let $\mathcal{X} = (\mathcal{X}, \|\cdot\|)$ be a normed vector space.
A sequence $(x_n) = (x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ **converges** to a limit $x \in \mathcal{X}$ if

$$\forall \epsilon > 0, \exists N_\epsilon \in \mathbb{N} \text{ such that } \|x - x_n\| < \epsilon \text{ for all } n > N_\epsilon$$

One then writes $x_n \to x$ or $x = \lim_n x_n$.
Likewise, (x_n) is a **Cauchy sequence** in $\mathcal{X}$ if

1) $x_n \in \mathcal{X}$ for all $n \in \mathbb{N}$

2) $\forall \epsilon > 0, \exists N_\epsilon \in \mathbb{N}$ such that $\|x_m - x_n\| < \epsilon$ for all $m, n > N_\epsilon$.

- The limits of converging and Cauchy sequences are unique
- If (x_n) converges in $\mathcal{X}$, then it is necessarily a Cauchy sequence
- If (x_n) converges or is Cauchy in $\mathcal{X}$, then it is bounded in the sense that $\sup_n \|x_n\| < \infty$
- The notion of Cauchy (resp., of a converging) sequence also exists for more general topologies; e.g., for a Fréchet space, which is equipped with a **countable family of semi-norms**.

Normed/Banach spaces

Topological vector space (TVS): a vector space equipped with a topology (e.g, the normed space $(\mathcal{X}, \|\cdot\|)$).

Definition
A TVS $\mathcal{X}$ is (sequentially) **complete** if the limit points of all Cauchy sequences (x_n) in $\mathcal{X}$ are included in $\mathcal{X}$.

Important case: **Banach space** = complete normed space.

- Properties
  - Every finite-dimensional normed vector space is complete
  - Every normed space $(\mathcal{X}, \|\cdot\|)$ has a **unique** (Banach) **completion**, which is denoted by $\overline{(\mathcal{X}, \|\cdot\|)}$.
    (In essence, this amounts to augmenting $\mathcal{X}$ by adding the limit points of all Cauchy sequences to it.) Another way to put it: $(\mathcal{X}, \|\cdot\|)$ is isometrically and densely embedded in $\overline{(\mathcal{X}, \|\cdot\|)}$.

Some useful Banach spaces for engineers

- Finite-dimensional vector spaces

$$(\mathbb{R}^N, \|\cdot\|_p) \text{ with } \|\mathbf{x}\|_p \triangleq \left(\sum_{n=1}^{N} |x_n|^p\right)^{1/p} \quad \text{is a Banach space for } p = 1, 2.$$

- Function spaces

Space of continuous functions equipped with the sup norm: $\|f\|_{\sup} \triangleq \sup_{t\in\mathbb{R}} |f(t)|$

$$C_{\mathrm{b}}(\mathbb{R}) = \big\{f \text{ continuous } \mathbb{R} \to \mathbb{R} : \|f\|_{\sup} < \infty\big\} \text{ is a Banach space.}$$

Lebesgue spaces (under implicit assumption of measurability):

$$L_p\text{-norm: } \|f\|_{L_p} \triangleq \left(\int_{\mathbb{R}} |f(t)|^p \mathrm{d}t\right)^{1/p}$$

$$L_p(\mathbb{R}) = \big\{f : \mathbb{R} \to \mathbb{R} \text{ s.t. } \|f\|_{L_p} < \infty\big\} \text{ is a Banach space for } p \in [1, \infty).$$

Schwartz' functions

Motivation

Schwartz's functions (a.k.a. test functions) are extremely well-behaved so that you can manipulate them without further precaution: differentiation, multiplication by polynomials, integrations, and more are always valid.

Definition: A function $\varphi : \mathbb{R} \to \mathbb{R}$ is in Schwartz' class $\big(\text{i.e., } \varphi \in \mathcal{S}(\mathbb{R})\big)$ if

- φ is infinitely differentiable (**smoothness**)
- for all $m, n \in \mathbb{N}$, there exists some constant $C_{m,n}$ such that

$$\forall t \in \mathbb{R} : \quad |\mathrm{D}^m\{\varphi\}(t)| \le C_{m,n} \frac{1}{1 + |t|^n} \qquad \textbf{(rapid decay)}$$

$$\mathrm{D}^m = \frac{\mathrm{d}^m t}{\mathrm{d}t^m}$$

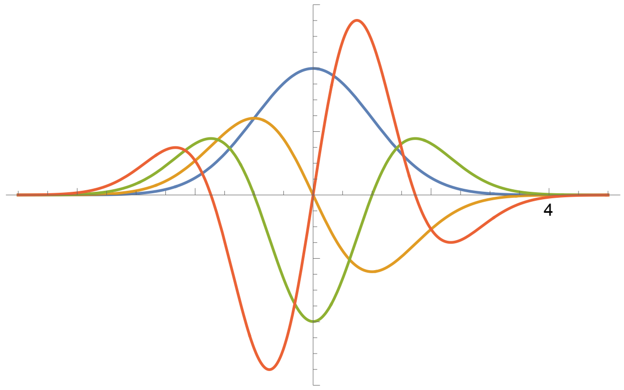
- Examples

The standardized Gaussian $g_0(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-|t|^2/2} \in \mathcal{S}(\mathbb{R})$
and, more generally, the Hermite functions $g_n(t) = \mathrm{D}^n\{g_0\}(t)$ for $n \in \mathbb{N}$.

Gauss-Hermite functions

Hermite functions: $g_n(t) = \mathrm{D}^n\{g_0\}(t)$ for $n \in \mathbb{N}$ where g_0 is the standardized Gaussian

- $g_0(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-|t|^2/2}$
- $g_1(t) = \frac{-t}{\sqrt{2\pi}} \mathrm{e}^{-|t|^2/2}$
- $g_2(t) = \frac{-1+t^2}{\sqrt{2\pi}} \mathrm{e}^{-|t|^2/2}$
- $g_3(t) = \frac{3t-t^3}{\sqrt{2\pi}} \mathrm{e}^{-|t|^2/2}$
- ...



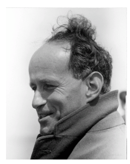
Schwartz' space of test functions

$\mathcal{S}(\mathbb{R})$ is a Fréchet space = a **complete** TVS equipped with a **countable family of semi-norms**

- $\mathcal{S}(\mathbb{R})$: Schwartz' space of smooth and rapidly decreasing functions

  Separating family of semi-norms: $\|\varphi\|_{m,n} = \sup_{t\in\mathbb{R}} \left|t^m \mathrm{D}^n\{\varphi\}(t)\right|$ for all $m, n \in \mathbb{N}$.

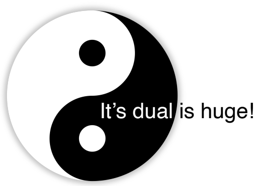
  Super strict notion of **convergence**: $\lim_i \varphi_i = \varphi \quad \Leftrightarrow \quad \lim_i \|\varphi_i - \varphi\|_{m,n} = 0, \forall m, n \in \mathbb{N}$



$$\mathcal{S}(\mathbb{R}) = \left\{\varphi : \mathbb{R} \to \mathbb{R} \text{ (or } \mathbb{C}) \text{ such that } \|\varphi\|_{m,n} < \infty, \forall m, n \in \mathbb{N}\right\}$$

Why should we care about $\mathcal{S} = \mathcal{S}(\mathbb{R})$?

- Classical interpretation of functions. Convergence in $\mathcal{S}$ conforms with our perception of what a **strict equality** $\lim_i f_i = \varphi$ should be: the functions on either side as well as all their derivatives match at every location $t \in \mathbb{R}$.
- $\mathcal{S}$ is small and yet large enough to represent any signal with an arbitrary degree of precision.
- $\mathcal{S}$ is dense in most of the classical function spaces of analysis. In particular, $L_1(\mathbb{R}) = \overline{(\mathcal{S}(\mathbb{R}), \|\cdot\|_{L_1})}$ and $L_2(\mathbb{R}) = \overline{(\mathcal{S}(\mathbb{R}), \|\cdot\|_{L_2})}$.
- $\mathcal{S}$ work beautifully with the **Fourier transform**: $\mathcal{F}$ bijectively maps $\mathcal{S}$ (complex version) into itself.



Linear operators: continuity property

Definition
An operator $\mathrm{A}:\mathcal{X}\to\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$, are vector spaces is ***linear*** if, for any $\varphi_1,\varphi_2\in\mathcal{X}$ and $a_1,a_2\in\mathbb{R}$ (or $\mathbb{C}$),

$$\mathrm{A}\{a_1\varphi_1+a_2\varphi_2\}=a_1\mathrm{A}\{\varphi_1\}+a_2\mathrm{A}\{\varphi_2\}$$

Definition
Let $\mathcal{X},\mathcal{Y}$ be topological vector spaces. An operator $\mathrm{A}:\mathcal{X}\to\mathcal{Y}$ is (sequentially) ***continuous*** (with respect to the topologies of $\mathcal{X}$ and $\mathcal{Y}$) if, for any convergent sequence (φ_i) in $\mathcal{X}$ with limit $\varphi\in\mathcal{X}$, the sequence $\big(\mathrm{A}\{\varphi_i\}\big)$ converges to $\mathrm{A}\{\varphi\}$ in $\mathcal{Y}$; i.e.,

$$\lim_i \mathrm{A}\{\varphi_i\}=\mathrm{A}\{\lim_i\varphi_i\}.$$

Example: The linear operator D^{m_0} is continuous $\mathcal{S}(\mathbb{R})\to\mathcal{S}(\mathbb{R})$.

Schwartz' semi-norms: $\|\varphi\|_{m,n}\triangleq\sup_{t\in\mathbb{R}}|t^m\mathrm{D}^n\{\varphi\}(t)|$ with $m,n\in\mathbb{N}$

$$\|\mathrm{D}^{m_0}\{\varphi_i\}-\mathrm{D}^{m_0}\{\varphi\}\|_{m,n}=\|\mathrm{D}^{m_0}\{\varphi_i-\varphi\}\|_{m,n}=\sup_{t\in\mathbb{R}}|t^m\mathrm{D}^{n+m_0}\{\varphi_i-\varphi\}(t)|=\|\varphi_i-\varphi\|_{m,n+m_0}\to 0$$

Examples of continuous operators on $\mathcal{S}(\mathbb{R})$

$\forall\varphi,\phi\in\mathcal{S}(\mathbb{R})$ and $n_0,m_0\in\mathbb{N}$

- n_0-fold derivative $\mathrm{D}^{n_0}:\mathcal{S}(\mathbb{R}^d)\to\mathcal{S}(\mathbb{R}^d)$

  $\mathrm{D}^{n_0}\{\varphi\}(t)=\frac{\mathrm{d}^{n_0}}{\mathrm{d}t^{n_0}}\varphi(t)\in\mathcal{S}(\mathbb{R})$

- Multiplication by a monomial of degree m_0, $\mathrm{M}_{m_0}:\mathcal{S}(\mathbb{R})\to\mathcal{S}(\mathbb{R})$

  $\mathrm{M}_{m_0}\{\varphi\}(t)=t^{m_0}\varphi(t)\in\mathcal{S}(\mathbb{R})$

- Shift by t_0, $\mathrm{S}_{t_0}:\mathcal{S}(\mathbb{R})\to\mathcal{S}(\mathbb{R})$

  $\mathrm{S}_{t_0}\{\varphi\}(t)=\varphi(t-t_0)\in\mathcal{S}(\mathbb{R})$

- Fourier transform $\mathcal{F}:\mathcal{S}(\mathbb{R}^d)\to\mathcal{S}(\mathbb{R}^d)$

  $$\mathcal{F}\{\varphi\}(\omega)=\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{j}\omega t}\varphi(t)\,\mathrm{d}t=\hat{\varphi}(\omega)\in\mathcal{S}(\mathbb{R})$$

Different notions of equality for functions

Three forms are not equivalent ...

- Classical equality (strict)

  $f(t) = g(t), \quad \forall t \in \mathbb{R}$

  Context: Functions should be continuous.

$$\Downarrow$$

- Egality in the norm (or "almost everywhere")

  $$\|f - g\|^2 = \int_{-\infty}^{+\infty} |f(t) - g(t)|^2 \mathrm{d}t = 0 \quad \Leftrightarrow \quad f = g \quad a.e.$$

  Context: Finite-energie functions, measure theory (Lebesgue, 1901)

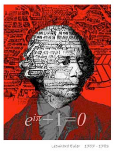
$$\Downarrow$$

- Weak equality (or "in the sense of distributions")

  $$\forall \phi \in \mathcal{S}, \quad \langle f, \phi \rangle = \langle g, \phi \rangle \quad \Leftrightarrow \quad f = g \quad \text{(in the sense of distributions)}$$

  - Notation for "duality product": $\langle f, \phi \rangle = \int_{-\infty}^{+\infty} f(t)\phi(t)\,\mathrm{d}t$
  - $\mathcal{S}$: Schwartz's space of "test" functions (smooth with rapid decay)
  - Context: Generalized fonctions, distribution theory (Schwartz, 1950)

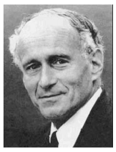
degree of generality



$C_b(\mathbb{R})$
or $\mathcal{S}(\mathbb{R})$
(if one cares about derivatives)

$L_2(\mathbb{R})$ and $L_1(\mathbb{R})$

with $L_p(\mathbb{R}) = \overline{(\mathcal{S}(\mathbb{R}), \|\cdot\|_{L_p})}, \quad p = 1, 2$



$\mathcal{S}'(\mathbb{R})$ (space of distributions)

with $\mathcal{S}(\mathbb{R}) \subset L_p(\mathbb{R}) \subset \mathcal{S}'(\mathbb{R})$

3. LINEAR FUNCTIONALS AND DUAL SPACE

- Linear functionals
- Continuous dual of a vector space
- Finite-dimensional scenario
- Adjoint operator

Linear functionals

Context: $\mathcal{X}$ is a complete topological vector space with some associated notion of sequential convergence; that is, if $(\varphi_i)_{i\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{X}$, then $\lim_{i\to\infty}\varphi_i = \varphi \in \mathcal{X}$.

Definition

- A **functional** on a TVS $\mathcal{X}$ is a mapping $\varphi \mapsto f(\varphi)$ that associates the **real value** $f(\varphi)$ to any element $\varphi \in \mathcal{X}$.
- The functional $f : \mathcal{X} \to \mathbb{R}$ is **linear** if

  $f(a_1\varphi_1 + a_2\,\varphi_2) = a_1 f(\varphi_1) + a_2 f(\varphi_2)$ for all $\varphi_1, \varphi_2 \in \mathcal{X}$ and $a_1, a_2 \in \mathbb{R}$ (or $\mathbb{C}$).
- The functional $f : \mathcal{X} \to \mathbb{R}$ is **continuous** if $\lim_i f(\varphi_i) = f(\lim_i \varphi_i) = f(\varphi)$ for any converging sequence (φ_i) in $\mathcal{X}$.

Proposition

A **linear functional** $f : \mathcal{X} \to \mathbb{R}, \varphi \mapsto f(\varphi) = \langle f, \varphi\rangle$ is **continuous** if and only if it is continuous at the origin where it takes the value 0.

$$\lim_i f(\varphi_i) = f(\lim_i \varphi_i) = f(\varphi) \Leftrightarrow \lim_i f(\varphi_i - \varphi) = f(0) = 0 \quad \text{(by linearity)}$$

Continuous dual of a vector space

Context: $\mathcal{X}$ is a complete topological vector space with some associated notion of sequential convergence; that is, if $(\varphi_i)_{i\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{X}$, then $\lim_{i\to\infty}\varphi_i = \varphi \in \mathcal{X}$.

Notation for a **linear functional** on $\mathcal{X}$. $\quad f : \varphi \mapsto \langle f, \varphi\rangle \quad$ with $\langle f, \varphi\rangle \in \mathbb{R}$

Observation: The set of all linear functionals on $\mathcal{X}$ is a vector space denoted by $\mathcal{X}^*$ (algebraic dual)

$$\langle a_1 f_1 + a_2 f_2, \varphi\rangle = a_1\langle f_1, \varphi\rangle + a_2\langle f_2, \varphi\rangle, \qquad \forall a_1, a_2 \in \mathbb{R}$$

Likewise: The set of all ***continuous linear functionals*** on $\mathcal{X}$ is a vector space $\mathcal{X}' \subseteq \mathcal{X}^*$

$\mathcal{X}'$: topological or ***continuous dual*** of $\mathcal{X}$

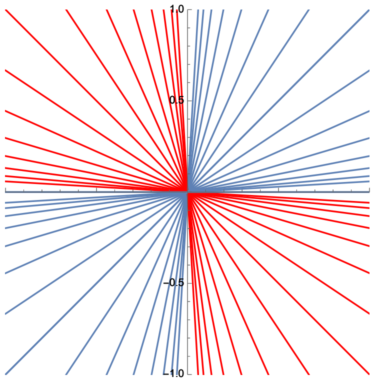
Scalar (or duality) product $\langle\cdot,\cdot\rangle$ is a bilinear functional: $\mathcal{X} \times \mathcal{X}' \to \mathbb{R}$ continuous in both arguments

- Weak-$*$ topology on $\mathcal{X}'$: $\quad (f_i)$ converges to f in $\mathcal{X}'$

  $\Leftrightarrow \quad \lim_i \langle f_i, \varphi\rangle = \langle f, \varphi\rangle$ for all $\varphi \in \mathcal{X}$

Most basic example: $(\mathbb{R}, |\cdot|)'$, which is isomorphic to $(\mathbb{R}, |\cdot|)$

- $\mathcal{X} = \mathbb{R}$ as a complete normed vector space
  - Algebraic structure: $a_1x_1 + a_2x_2 \in \mathbb{R}$ for all $x_1, x_2 \in \mathcal{X}$ and $a_1, a_2 \in \mathbb{R}$
  - Norm: $x \mapsto |x|$ (the choice here is unique)
  - Completeness: $\mathbb{R}$ is complete;
    i.e., all Cauchy sequences in $\mathbb{R}$ have a limit in $\mathbb{R}$.



- The linear functional on $\mathcal{X} = \mathbb{R}$
  - They are all of the form $f_i : x \mapsto \alpha_i x$ with $\alpha_i \in \mathbb{R}$
  - They are automatically continuous in the topology of $\mathbb{R}$.
  - Each of them is uniquely identified by an element of $\mathbb{R}$: $\alpha_i = f_i(1) = \mathrm{D}f_i(0) \in \mathbb{R}$

$\Rightarrow$ $(\mathbb{R}, |\cdot|)'$ is isomorphic to $(\mathbb{R}, |\cdot|)$ (often written as $\mathbb{R}' = \mathbb{R}$)

Finite-dimensional scenario

- Identification of linear functionals on $\mathcal{X} = \mathbb{R}^N$

$f(\mathbf{x}) : \mathbb{R}^N \to \mathbb{R}$ is a linear functional on $\mathbb{R}^N$ $\quad \forall \mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$ and $\forall a_1, a_2 \in \mathbb{R}$

$\Leftrightarrow \quad f(a_1\mathbf{x}_1 + a_2\mathbf{x}_2) = a_1 f(\mathbf{x}_1) + a_2 f(\mathbf{x}_2)$

$\Leftrightarrow \quad f(\mathbf{x}) = \mathbf{f}^T\mathbf{x} = \langle \mathbf{f}, \mathbf{x}\rangle$ for some $\mathbf{f} \in \mathbb{R}^N$

> Association between $f : \mathbb{R}^N \to \mathbb{R}$
> and $\mathbf{f} = (f_n) \in \mathbb{R}^N$ is one-to-one with $f_n = f(\mathbf{e}_n)$

or, alternatively, $\mathbf{f} = \boldsymbol{\nabla} f(\mathbf{0})$ (slope)

- Topological consideration for $\mathcal{X} = (\mathbb{R}^N, \|\cdot\|_2)$ (Euclidean space)

Continuity: $|f(\mathbf{x}_i) - f(\mathbf{x}_j)| = |f(\mathbf{x}_i - \mathbf{x}_j)| = |\langle \mathbf{f}, \mathbf{x}_i - \mathbf{x}_j\rangle| \le \|\mathbf{f}\|_2 \, \|\mathbf{x}_i - \mathbf{x}_j\|_2$ (Cauchy-Schwarz)

> **Proposition**
> The set of all continuous linear functionals on $\mathbb{R}^N$ is a vector space $(\mathbb{R}^N)'$ isomorphic to $\mathbb{R}^N$
> $$\langle a_1\mathbf{f}_1 + a_2\mathbf{f}_2, \mathbf{x}\rangle = a_1\langle \mathbf{f}_1, \mathbf{x}\rangle + a_2\langle \mathbf{f}_2, \mathbf{x}\rangle$$

Adjoint operator

Let $\mathcal{X}$ be a complete topological vector space with continuous dual $\mathcal{X}'$.

Duality pairing: $(f,\varphi) \mapsto \langle f,\varphi\rangle_{\mathcal{X}'\times\mathcal{X}} = \langle f,\varphi\rangle$ (for short)

- Adjoint of a linear operator $\mathrm{A} : \mathcal{X} \to \mathcal{X}$

Theorem
If A is a continuous linear operator on $\mathcal{X}$, then there exists a unique continuous linear operator $\mathrm{A}^* : \mathcal{X}' \to \mathcal{X}'$ (the adjoint of A) such that, for all $(f,\varphi) \in \mathcal{X}' \times \mathcal{X}$,

$$\langle f, \mathrm{A}\{\varphi\}\rangle = \langle \mathrm{A}^*\{f\}, \varphi\rangle .$$

- Finite-dimensional scenario : $\mathcal{X} = (\mathbb{R}^N, \|\cdot\|_2)$

$\mathrm{A} : \mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ with $\mathbf{A} \in \mathbb{R}^{N\times N}$
$\mathrm{A}^* : \mathbf{y} \mapsto \mathbf{A}^T\mathbf{y}$ (transpose)

Indeed, $\forall \mathbf{x},\mathbf{y} \in \mathbb{R}^N : \langle \mathbf{y}, \mathbf{A}\mathbf{x}\rangle = \mathbf{y}^T\mathbf{A}\mathbf{x} = (\mathbf{A}^T\mathbf{y})^T\mathbf{x} = \langle \mathbf{A}^T\mathbf{y}, \mathbf{x}\rangle$

4. DISTRIBUTIONS

- Mathematical notion of distribution
- Examples of distributions
- Schwartz space of tempered distrubutions
- Operations on distributions
- Examples of calculations
- Conclusion

What is a distribution (a.k.a. generalized function)?

It is a ***continuous linear functional*** on $\mathcal{S}(\mathbb{R})$

$f \in \mathcal{S}'(\mathbb{R})$: Schwartz' space of tempered distributions

It is a rule $\mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ that associates a real number $\langle \varphi, f \rangle$ to every test function φ

Example: $\langle \delta(\cdot - t_0), \varphi \rangle = \varphi(t_0)$

It is an extension of the classical notion of function.

If f is (slowly increasing and) locally integrable, then $\langle \varphi, f \rangle = \displaystyle\int_{\mathbb{R}} f(t)\varphi(t)\mathrm{d}t$

Otherwise, when f cannot be identified as a function, the "integral" notation $\displaystyle\int_{\mathbb{R}} f(t)\varphi(t)\mathrm{d}t = \langle f, \varphi \rangle$ is **only formal**

Example: $\displaystyle\int_{\mathbb{R}} \varphi(t)\delta(t - t_0)\mathrm{d}t = \varphi(t_0)$ (sampling property of Dirac)

Examples of distributions

Specification of a distribution $f : \varphi \mapsto \langle f, \varphi \rangle$

Properties to check

1. Linearity: $\langle f, a_1\varphi_1 + a_2\varphi_2 \rangle = a_1\langle f, \varphi_1 \rangle + a_2\langle f, \varphi_2 \rangle$ for all $a_1, a_2 \in \mathbb{R}$ and $\varphi_1, \varphi_2 \in \mathcal{S}$
2. Continuity: $\lim_i \langle f, \varphi_i \rangle = \langle f, \lim_i \varphi_i \rangle = 0$ for any sequence (φ_i) in $\mathcal{S}$ that converges to 0.

N.B.: $\varphi_i \to 0$ in $\mathcal{S}$ is equivalent to $\|\varphi_i\|_{m,n} \triangleq \sup_{t\in\mathbb{R}} |t^m \mathrm{D}^n\{\varphi_i\}(t)| \to 0$ for all $m, n \in \mathbb{N}$

- Shifted Dirac impulse

  $\delta(\cdot - t_0) : \varphi \mapsto \langle \delta(\cdot - t_0), \varphi \rangle \triangleq \varphi(t_0)$

  Linearity: obvious

  Continuity: $|\langle \delta(\cdot - t_0), \varphi \rangle| \le \|\varphi\|_{0,0}$

- Absolutely-integrable function

  $f : \varphi \mapsto \langle f, \varphi \rangle \triangleq \displaystyle\int_{\mathbb{R}} f(t)\varphi(t)\mathrm{d}t$ with $\displaystyle\int_{\mathbb{R}} |f(t)|\mathrm{d}t = \|f\|_{L_1} < \infty$

  Linearity: obvious

  Continuity: $|\langle f, \varphi \rangle| \le \|f\|_{L_1}\|\varphi\|_{0,0}$

- Functions with no-faster-than-algebraic growth

  $\exists m_0$ such $\displaystyle\int_{\mathbb{R}} |f(t)|(1 + |t|)^{-m_0}\mathrm{d}t = \|f\|_{L_1,-m_0} < \infty$

  Continuity: $|\langle f, \varphi \rangle| \le C\|f\|_{L_1,-m_0}\|\varphi\|_{m_0,0}$

Schwartz' space $\mathcal{S}'(\mathbb{R})$ of tempered distributions

$\mathcal{S}'(\mathbb{R})$ is the topological dual of $\mathcal{S}(\mathbb{R})$, which is itself a *nuclear* Fréchet space*.
This implies that $\mathcal{S}'(\mathbb{R})$ is also a *nuclear space*, which fully justifies the use of the weak topology.

*: *A* **nuclear space** *is a TVS with the remarkable property that strong and weak=weak-* convergence are equivalent.*

- Convergence in $\mathcal{S}'(\mathbb{R})$

$\mathcal{S}' = \mathcal{S}'(\mathbb{R})$ equipped with the weak topology is complete.
A sequence $(f_n) = (f_n)_{n\in\mathbb{N}}$ of distributions in $\mathcal{S}'$ **converges** to a limit $f \in \mathcal{S}'$ if

$$\forall \varphi \in \mathcal{S}: \lim_{n\to\infty} \langle f_n, \varphi\rangle = \langle \lim_{n\to\infty} f_n, \varphi\rangle = \langle f, \varphi\rangle$$

- Example

Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous function such that $\|f\|_{L_1} < \infty$.

Then, $f_n(t) = \sum_{k=-n^2}^{n^2} f(\frac{k}{n})\delta(t-\frac{k}{n})\frac{1}{n}$ converges to f in $\mathcal{S}'$ as $n\to\infty$.

Further examples of weak convergence

Convergence in the sense of distributions: $\lim_{n\to\infty} f_n = f$ in $\mathcal{S}'$ $\Leftrightarrow$ $\forall\varphi\in\mathcal{S}: \lim_{n\to\infty}\langle f_n,\varphi\rangle = \langle f,\varphi\rangle$

- Construction of Dirac impulse

The functions $f_n(t) = n\,\mathrm{rect}\big(nt\big)$ converge to δ in $\mathcal{S}'$ as $n\to\infty$.

- Unit integral: $\int_{\mathbb{R}} f_n(t) = 1$ for all $n\in\mathbb{N}$
- $$\lim_{n\to\infty}\langle f_n,\varphi\rangle = \lim_{n\to\infty}\int_{\mathbb{R}} f_n(t)\varphi(t)\mathrm{d}t \underset{t=\frac{\tau}{n}}{=} \lim_{n\to\infty}\int_{-\frac{1}{2}}^{\frac{1}{2}}\varphi\big(\tfrac{\tau}{n}\big)\mathrm{d}\tau = \int_{-\frac{1}{2}}^{\frac{1}{2}}\underbrace{\lim_{n\to\infty}\varphi\big(\tfrac{\tau}{n}\big)}_{\varphi(0)}\mathrm{d}\tau = \varphi(0)$$

The argument caries through for any $g:\mathbb{R}\to\mathbb{R}$ that is continuous at $t=0$ with $\int_{\mathbb{R}} g(t)\mathrm{d}t = 1$.

Remark: While $f_n \in L_1(\mathbb{R})$ with $\|f\|_{L_1} = 1$ for all $n\in\mathbb{N}$, $\lim_{n\to\infty} f_n = \delta \notin L_1(\mathbb{R})$ because (f_n) fails to be a Cauchy sequence in $L_1(\mathbb{R})$. In other words, δ **cannot be identified as a function**.

Operations on distributions

Dual extension principle
Let $\mathrm{U}, \mathrm{U}^* : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ be two operators that form an adjoint pair on $\mathcal{S}(\mathbb{R}) \times \mathcal{S}(\mathbb{R})$.
Their action to $\mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ is extended by defining $\mathrm{U}\{f\}$ and $\mathrm{U}^*\{f\}$ such that

$$\langle \mathrm{U}\{f\}, \varphi\rangle = \langle f, \mathrm{U}^*\{\varphi\}\rangle,$$
$$\langle \mathrm{U}^*\{f\}, \varphi, \rangle = \langle f, \mathrm{U}\{\varphi\}\rangle.$$

- Examples of definition of operations on distributions
  - Shift by $t_0 \in \mathbb{R}$: $\langle f(\cdot - t_0), \varphi\rangle \triangleq \langle f, \varphi(\cdot + t_0)\rangle$
  - nth-order derivative: $\langle \mathrm{D}^n\{f\}, \varphi\rangle \triangleq (-1)^n \langle f, \mathrm{D}^n\{\varphi\}\rangle$
  - Multiplication by monomial: $\langle \mathrm{M}_{m_0}\{f\}, \varphi\rangle \triangleq \langle f, \mathrm{M}_{m_0}\{\varphi\}\rangle \triangleq \langle f, (\cdot)^{m_0}\varphi\rangle$
  - Convolution: $\langle h * f, \varphi\rangle \triangleq \langle f, h^\vee * \varphi\rangle$ with $h \in \mathcal{S}(\mathbb{R})$ and $h^\vee(t) = h(-t)$
  - Fourier transform: $\langle \mathcal{F}\{f\}, \hat{\varphi}\rangle \triangleq \langle f, \mathcal{F}^*\{\hat{\varphi}\}\rangle = 2\pi\langle f, \varphi\rangle$ for all $\hat{\varphi} \in \mathcal{S}(\mathbb{R})$

Justification: Continuity and adjoint relations on $\mathcal{S}(\mathbb{R})$

$\forall \varphi, \phi \in \mathcal{S}(\mathbb{R})$ and $n_0, m_0 \in \mathbb{N}$

- n_0-fold derivative $\mathrm{D}^{n_0} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$

  $\mathrm{D}^{n_0}\{\varphi\}(t) = \frac{\mathrm{d}^{n_0}}{\mathrm{d}t^{n_0}}\varphi(t) \in \mathcal{S}(\mathbb{R})$

  Adjoint relation: $(\mathrm{D}^{n_0})^* = (-1)^{n_0}\mathrm{D}^{n_0}$; i.e., $\langle \mathrm{D}^{n_0}\{\varphi\}, \phi\rangle = (-1)^{n_0}\langle \varphi, \mathrm{D}^{n_0}\{\phi\}\rangle$ (integration by part)

- Multiplication by a monomial of degree $m_0 \in \mathbb{N}$, $\mathrm{M}_{m_0} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$

  $\mathrm{M}_{m_0}\{\varphi\}(t) = t^{m_0}\varphi(t) \in \mathcal{S}(\mathbb{R})$

  Adjoint relation: $\mathrm{M}_{m_0}^* = \mathrm{M}_{m_0} \quad \Leftrightarrow \quad \langle \mathrm{M}_{m_0}\{\varphi\}, \phi\rangle = \langle \varphi, \mathrm{M}_{m_0}\{\phi\}\rangle$

- Fourier transform $\mathcal{F} : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$

  $$\mathcal{F}\{\varphi\}(\omega) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{j}\omega t}\varphi(t)\,\mathrm{d}t = \hat{\varphi}(\omega) \in \mathcal{S}(\mathbb{R})$$

  Adjoint relation $\mathcal{F}^* = (2\pi)\mathcal{F}^{-1}$ (Parseval) with $\mathcal{F}^{-1} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$

Examples of calculations

- Differentiation: $\langle \mathrm{D}\{f\}, \varphi\rangle \stackrel{\triangle}{=} -\langle f, \mathrm{D}\{\varphi\}\rangle$ for all $\varphi \in \mathcal{S}(\mathbb{R})$
  - $\langle \delta', \varphi\rangle = -\langle \delta, \mathrm{D}\{\varphi\}\rangle = -\varphi'(0)$
  - $\mathrm{D}\{u\} = \delta.$
    Indeed: $\langle \mathrm{D}\{u\}, \varphi\rangle = -\langle u, \mathrm{D}\{\varphi\}\rangle = -\int_0^{+\infty} \varphi'(t)\mathrm{d}t = \varphi(0) - \varphi(\infty) = \varphi(0)$

- Convolution with $h \in \mathcal{S}$: $\langle f * h, \varphi\rangle \stackrel{\triangle}{=} \langle f, h^\vee * \varphi\rangle$ for all $\varphi \in \mathcal{S}(\mathbb{R})$
  - $\delta * h = h$
    $\forall \varphi \in \mathcal{S} : \langle \delta * h, \varphi\rangle = \langle \delta, h^\vee * \varphi\rangle = h^\vee * \varphi(0) = \langle h, \varphi\rangle$
  - $\delta' * h = \mathrm{D}\{h\} = h'$

Conclusion: Advantages of working with $\mathcal{S}'(\mathbb{R})$

- Direct **analogy with linear algebra**
  - Fundamental elements (not included in $L_2(\mathbb{R})$ = classical Hilbert space):
    (i) $\delta(\cdot - t_0) \in \mathcal{S}'(\mathbb{R})$ (continuous-domain analog of canonical basis),
    (ii) $\mathrm{e}^{-\mathrm{j}\omega_0 \cdot} \in \mathcal{S}'(\mathbb{R})$ (eigenfunctions of LSI operators)
  - Schwartz' kernel theorem:
    Every linear operator is characterized by its generalized **impulse response** (analog of bi-infinite matrix)

- Rigorous extensions of **basic signal processing operators**
  - Weak derivative(s), which may produce Dirac impulses
  - Generalized Fourier transform, which may not have an explicit representation as a "classical Fourier integral"
  - Multiplication with smooth, slowly-increasing functions (e.g. modulation, multiplication with monomial)

  (These operations are not necessarily valid in the classical sense for ordinary (e.g. piecewise continuous) functions $f : \mathbb{R} \to \mathbb{R}$)

- Largest mathematical framework that supports use of the **Fourier transform**
  - $\mathcal{F} : \mathcal{S}'(\mathbb{R}) \to \mathcal{S}'(\mathbb{R})$ (continuous bijection)
  - Distributional calculus is self-consistent ("tous les coups, ou presque, sont permis")
    Liberal use of properties in Tables (without worrying about underlying mathematical hypotheses).

Summary: infinite-dimensional extension of linear algebra

Vector algebra		*Functional counterpart*
$\mathbf{x} = (x_1, \cdots, x_N)$	⇒	$f(t)$
$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^{N} x_n y_n$	⇒	$\langle f, \varphi \rangle = \int_{\mathbb{R}} f(t)\varphi(t)\mathrm{d}t$ (duality product)
$\mathbf{e}_n$ (canonical basis)	⇒	Shifts of the Dirac distribution: $\delta(\cdot - \tau) \in \mathcal{S}'(\mathbb{R})$
$\mathbf{y} = \mathbf{A}\mathbf{x}$	⇒	$\mathrm{A}\{f\}(t) = \int_{\mathbb{R}} a(t, \tau) f(t) \mathrm{d}\tau$
Transpose of a matrix: $\mathbf{A}^T$	⇒	Adjoint operator: $\mathrm{A}^*\{f\}(t) = \int_{\mathbb{R}} a(\tau, t) f(t) \mathrm{d}\tau$