In this chapter we summarize some of the mathematical preliminaries for the remaining chapters. These concern the function spaces used in the book, duality, generalized functions, probability theory, and generalized random processes. Each of these topics is discussed in a separate section.

For the most part, the theory of function spaces and generalized functions can be seen as an infinite-dimensional generalization of linear algebra (function spaces generalize $\mathbb{R}^n$, and continuous linear operators generalize matrices). Similarly, the theory of generalized random processes involves the generalization of the idea of a finite random vector in $\mathbb{R}^n$ to an element of an infinite-dimensional space of generalized functions.

To give a taste of what is to come, we briefly compare finite- and infinite-dimensional theories in Tables 3.1 and 3.2. The idea, in a nutshell, is to substitute vectors by (generalized) functions. Formally, this extension amounts to replacing some finite sums (in the finite-dimensional formulation) by integrals. Yet, in order for this to be mathematically sound, one needs to properly define the underlying objects as elements of some infinite-dimensional vector space, to specify the underlying notion(s) of convergence (which is not an issue in $\mathbb{R}^n$), while ensuring that some basic continuity conditions are met.

The impatient reader who is not directly concerned by those mathematical issues may skip what follows the tables at first reading and consult these sections later as he may feel the need. Yet, he should be warned that the material on infinite-dimensional probability theory from Subsection 3.4.4 to the end of the chapter is fundamental to our formulation. The mastery of those notions also requires a good understanding of function spaces and generalized functions which are covered in the first part of the chapter.

3.1 Some classes of function spaces

By the term function we shall intend elements of various function spaces. At a minimum, a function space is a set $\mathcal{X}$ along with some criteria for determining, first, whether or not a given “function” $\varphi = \varphi(r)$ belongs to $\mathcal{X}$ (in mathematical notation, $\varphi \in \mathcal{X}$) and, secondly, given $\varphi, \varphi_0 \in \mathcal{X}$, whether or not $\varphi$ and $\varphi_0$ describe the same
### Table 3.1: Comparison of notions of linear algebra with those of functional analysis and the theory of distributions (generalized functions). See Sections 3.1-3.3 for an explanation.

<table>
<thead>
<tr>
<th>finite-dimensional theory (linear algebra)</th>
<th>infinite-dimensional theory (functional analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclidean space $\mathbb{R}^N$, complexification $\mathbb{C}^N$</td>
<td>function spaces such as the Lebesgue space $L_p(\mathbb{R}^d)$ and the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$, among others.</td>
</tr>
<tr>
<td>vector $x = (x_1, \ldots, x_N)$ in $\mathbb{R}^N$ or $\mathbb{C}^N$</td>
<td>function $f(r)$ in $\mathcal{S}'(\mathbb{R}^d)$, $L_p(\mathbb{R}^d)$, etc.</td>
</tr>
<tr>
<td>bilinear scalar product $\langle x, y \rangle = \sum_{n=1}^N x_n y_n$</td>
<td>$\langle \varphi, g \rangle = \int \varphi(r) g(r) , dr \quad \forall \varphi \in \mathcal{S}'(\mathbb{R}^d)$ (test function), $g \in \mathcal{S}'(\mathbb{R}^d)$ (generalized function), or $g \in L_p(\mathbb{R}^d)$, $g \in L_q(\mathbb{R}^d)$ with $\frac{1}{p} + \frac{1}{q} = 1$, for instance.</td>
</tr>
<tr>
<td>equality: $x \equiv y \iff x_n = y_n$</td>
<td>various notions of equality (depends on the space), such as weak equality of distributions: $f = g \in \mathcal{S}'(\mathbb{R}^d) \iff \langle \varphi, f \rangle = \langle \varphi, g \rangle$ for all $\varphi \in \mathcal{S}'(\mathbb{R}^d)$, almost-everywhere equality: $f = g \in L_p(\mathbb{R}^d) \iff \int \text{Prob}(</td>
</tr>
<tr>
<td>$\Rightarrow \quad \langle u, x \rangle = \langle u, y \rangle, \quad \forall u \in \mathbb{R}^N$</td>
<td>continuous linear operators $\mathcal{S}'(\mathbb{R}^d) \rightarrow \mathcal{S}'(\mathbb{R}^d)$</td>
</tr>
<tr>
<td>$\Rightarrow \quad |x - y|^2 = 0$</td>
<td>$y = Ax \Rightarrow y_m = \sum_{n=1}^N a_{mn} x_n$</td>
</tr>
<tr>
<td>$y = Ax \Rightarrow y_m = \sum_{n=1}^N a_{mn} x_n$</td>
<td>$g = A \varphi \Rightarrow g(r) = \int \text{Prob}(a(r,s) \varphi(s)) , ds$ for some $a \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ (Schwartz' kernel theorem)</td>
</tr>
<tr>
<td>transpose $\langle x, Ay \rangle = (A^T x, y)$</td>
<td>adjoint $\langle \varphi, Ag \rangle = (A^* \varphi, g)$</td>
</tr>
</tbody>
</table>

### Table 3.2: Comparison of notions of finite-dimensional statistical calculus with the theory of generalized stochastic processes. See Sections 3.4 for an explanation.

<table>
<thead>
<tr>
<th>finite-dimensional</th>
<th>infinite-dimensional</th>
</tr>
</thead>
<tbody>
<tr>
<td>random variable $X$ in $\mathbb{R}^N$</td>
<td>generalized stochastic process $s$ in $\mathcal{S}'$</td>
</tr>
<tr>
<td>probability measure $\mathcal{P}_X$ on $\mathbb{R}^N$</td>
<td>probability measure $\mathcal{P}_s$ on $\mathcal{S}'$</td>
</tr>
<tr>
<td>$\mathcal{P}_X(E) = \text{Prob}(X \in E) = \int_E p_X(x) , dx \quad (p_X$ is a generalized (i.e., hybrid) pdf)</td>
<td>$\mathcal{P}_s(E) = \text{Prob}(s \in E) = \int_E \mathcal{P}_s(dg)$</td>
</tr>
<tr>
<td>for suitable subsets $E \subset \mathbb{R}^N$</td>
<td>for suitable subsets $E \subset \mathcal{S}'$</td>
</tr>
<tr>
<td>characteristic function $\mathcal{P}<em>X(\omega) = \mathbb{E}[e^{i\omega \cdot X}] = \int</em>{\mathbb{R}^N} e^{i\omega \cdot x} p_X(x) , dx, \quad \omega \in \mathbb{R}^N$</td>
<td>characteristic functional $\mathcal{P}<em>s(\varphi) = \mathbb{E}[e^{i\varphi \cdot s}] = \int</em>{\mathcal{S}'} e^{i\varphi \cdot y} \mathcal{P}_s(dy)$, $\varphi \in \mathcal{S}'$</td>
</tr>
</tbody>
</table>
object in $\mathcal{X}$ (in mathematical notation, $\varphi = \varphi_0$). Most often, in addition to these, the space $\mathcal{X}$ has additional structure (see below).

In this book we shall largely deal with two types of function spaces: complete normed spaces such as Lebesgue $L_p$ spaces, and nuclear spaces such as the Schwartz space $\mathcal{S}$ and the space $\mathcal{D}$ of compactly supported test functions, as well as their duals $\mathcal{X}'$ and $\mathcal{S}'$, which are spaces of generalized functions. These two categories of spaces (complete-normed and nuclear) cannot overlap, except in finite dimensions. Since the function spaces that are of interest to us are infinite-dimensional (they do not have a finite vector-space basis), the two categories are mutually exclusive.

The structure of each of the afore-mentioned spaces has two aspects. First, as a vector space over the real numbers or its complexification, the space has an algebraic structure. Second, with regard to the notions of convergence and taking of limits, the space has a topological structure. The algebraic structure lends meaning to the idea of a linear operator on the space, while the topological structure gives rise to the concept of a continuous operator or map, as we shall see shortly.

All the spaces considered here have a similar algebraic structure. They are either vector spaces over $\mathbb{R}$, meaning for any $\varphi, \varphi_0$ in the space and any $a \in \mathbb{R}$, the operations of addition $\varphi + \varphi_0$ and multiplication by scalars $a \varphi$ are defined and map the space (denoted henceforth by $\mathcal{X}$) into itself. Or, we may take the complexification of a real vector space $\mathcal{X}$, composed of elements of the form $\varphi = \varphi_r + j \varphi_i$ with $\varphi_r, \varphi_i \in \mathcal{X}$ and $j$ denoting the imaginary unit. The complexification is then a vector space over $\mathbb{C}$. In the remainder of the book, we shall denote a real vector space and its complexification by the same symbol. The distinction, when important, will be clear from the context.

For the spaces with which we are concerned in this book, the topological structure is completely specified by providing a criterion for the convergence of sequences.\footnote{This is in contrast with those topological spaces where one needs to consider generalizations of the notion of a sequence involving partially ordered sets (the so-called nets or filters). Spaces in which a knowledge of sequences suffices are called sequential.} By this we mean that, for any given sequence $(\varphi_i)$ in $\mathcal{X}$ and any $\varphi \in \mathcal{X}$, we are equipped with the knowledge of whether or not $\varphi$ is the limit of $(\varphi_i)$. A topological space is a set $\mathcal{X}$ with topological structure. For normed spaces, the said criterion is given in terms of a norm, while in nuclear spaces it is given in terms of a family of seminorms, as we shall discuss below. But before that, let us first define linear and continuous operators.

An operator is a mapping from one vector space to another; that is, a rule that associates an output function $A\varphi \in \mathcal{Y}$ to each input $\varphi \in \mathcal{X}$.

**Definition 3.1 (Linear operator)** An operator $A : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are vector spaces is called **linear** if for any $\varphi, \varphi_0 \in \mathcal{X}$ and $a, b \in \mathbb{R}$ (or $\mathbb{C}$),

$$A(a\varphi + b\varphi_0) = aA\varphi + bA\varphi_0.$$  

**Definition 3.2 (Continuous operator)** Let $\mathcal{X}, \mathcal{Y}$ be topological spaces. An operator $A : \mathcal{X} \to \mathcal{Y}$ is called **sequentially continuous** (with respect to the topologies of
Mathematical context and background

\( X \) and \( Y \) if, for any convergent sequence \( \{\varphi_i\} \) in \( X \) with limit \( \varphi \in X \), the sequence \( \{A\varphi_i\} \) converges to \( A\varphi \) in \( Y \), that is,

\[
\lim_i A\varphi_i = A(\lim_i \varphi_i).
\]

The above definition of continuity coincides with the stricter topological definition for spaces we are interested in.

We shall assume that the topological structure of our vector spaces is such that the operations of addition and multiplication by scalars in \( \mathbb{R} \) (or \( \mathbb{C} \)) are continuous. With this compatibility conditions our object is called a topological vector space.

Having defined the two types of structure (algebraic and topological) and their relation with operators in abstract terms, let us now show concretely how the topological structure is defined for some important classes of spaces.

### 3.1.1 About the notation: mathematics vs. engineering

So far, we have considered a function in abstract terms as an element of a vector space: \( \varphi \in X \). The more conventional view is that of map \( \varphi : \mathbb{R}^d \to \mathbb{R} \) (or \( \mathbb{C} \)) that associates a value \( \varphi(r) \) to each point \( r = (r_1, \ldots, r_d) \in \mathbb{R}^d \). Following the standard convention in engineering, we shall therefore also use the notation \( \varphi(r) \) (instead of \( \varphi(\cdot) \) or \( \varphi \)) to represent the function using \( r \) as our generic \( d \)-dimensional index variable, the norm of which is denoted by \( |r|^2 = \sum_{i=1}^{d} |r_i|^2 \). This is to be contrasted with the point values (or samples) of \( \varphi \) which will be denoted using subscripted index variables; i.e., \( \varphi(r_k) \) stands for the value of \( \varphi \) at \( r = r_k \). Likewise, \( \varphi(r - r_0) = \varphi(\cdot - r_0) \) refers to the function \( \varphi \) shifted by \( r_0 \).

A word of caution is in order here. While the engineering notation has the advantage of being explicit, it can also be felt as being abusive because the point values of \( \varphi \) are not necessarily well defined, especially when the function presents discontinuities, not to mention the case of generalized functions that do not have a pointwise interpretation. \( \varphi(r) \) should therefore be treated as an alternative notation for \( \varphi \) that reminds us of the domain of the function and not interpreted literally.

### 3.1.2 Normed spaces

A norm on \( X \) is a map \( \| \cdot \| : X \to \mathbb{R} \), usually denoted as \( \varphi \mapsto \|\varphi\| \) (with indices used if needed to distinguish between different norms), which fulfills the following properties for all \( a \in \mathbb{R} \) (or \( \mathbb{C} \)) and \( \varphi, \varphi_0 \in X \).

- \( \|\varphi\| \geq 0 \) (nonnegativity).
- \( \|a\varphi\| = |a| \|\varphi\| \) (positive homogeneity).
- \( \|\varphi + \varphi_0\| \leq \|\varphi\| + \|\varphi_0\| \) (triangular inequality).
- \( \|\varphi\| = 0 \) implies \( \varphi = 0 \) (separation of points).

By relaxing the last requirement we obtain a seminorm.

A normed space is a vector space \( X \) equipped with a norm.
A sequence \((q_i)\) in a normed space \(X\) is said to converge to \(q\) (in the topology of \(X\)), in symbols
\[
\lim_{i} q_i = q,
\]
if and only if
\[
\lim_{i} \|q - q_i\| = 0.
\]

Let \((q_i)\) be a sequence in \(X\) such that for any \(\epsilon > 0\) there exists an \(N \in \mathbb{N}\) with
\[
\|q_i - q_j\| < \epsilon \quad \text{for all} \quad i, j \geq N.
\]
Such a sequence is called a Cauchy sequence. A normed space \(X\) is complete if it does not have any holes, in the sense that, for every Cauchy sequence in \(X\), there exists an \(q \in X\) such that \(\lim_i q_i = q\) (in other words if every Cauchy sequence has a limit in \(X\)). A normed space that is not complete can be completed by introducing new points corresponding to the limits of equivalent Cauchy sequences. For example, the real line is the completion of the set of rational numbers with respect to the absolute-value norm.

**Examples**

Important examples of complete normed spaces are the Lebesgue spaces. The Lebesgue spaces \(L^p(\mathbb{R}^d)\), \(1 \leq p \leq \infty\), are composed of functions whose \(L^p(\mathbb{R}^d)\) norm,
\[
\|q(r)\|_{L^p} = \left\{ \begin{array}{ll}
\left( \int_{\mathbb{R}^d} |q(r)|^p \, dr \right)^{\frac{1}{p}} & \text{for} \quad 1 \leq p < \infty \\
\text{ess sup}_{r \in \mathbb{R}^d} |q(r)| & \text{for} \quad p = \infty
\end{array} \right.
\]
and where two functions that are equal almost everywhere are considered to be equivalent.

We may also define weighted \(L^p\) spaces by replacing the shift-invariant Lebesgue measure \(dr\) by a weighted measure \(w(r)dr\) in the above definitions. In that case, \(w(r)\) is assumed to be a measurable function that is (strictly) positive almost everywhere. In particular, for \(w(r) = 1 + |r|^{\alpha}\) with \(\alpha > 0\) (or, equivalently, \(w(r) = (1 + |r|)^{\alpha}\)), we denote the associated norms as \(\| \cdot \|_{p,\alpha}\), and the corresponding normed spaces as \(L^p_{\alpha}(\mathbb{R}^d)\). The latter spaces are useful when characterizing the decay of functions at infinity. For example, \(L^\infty_{\alpha}(\mathbb{R}^d)\) is the space of functions that are bounded by a constant multiple of \(\frac{1}{1+|r|^\alpha}\) almost everywhere.

Some remarkable inclusion properties of \(L^p_{\alpha}(\mathbb{R}^d)\), \(1 \leq p \leq \infty\), \(\alpha > 0\) are

- \(\alpha > \alpha_0\) implies \(L^p_{\alpha}(\mathbb{R}^d) \subseteq L^p_{\alpha_0}(\mathbb{R}^d)\).
- \(L^\infty_{\alpha}(\mathbb{R}^d) \subseteq L^p(\mathbb{R}^d)\) for any \(\epsilon > 0\).

Finally, we define the space of rapidly decaying functions, \(\mathcal{S}(\mathbb{R}^d)\), as the intersection of all \(L^\infty_{\alpha}(\mathbb{R}^d)\) spaces, \(\alpha > 0\), or, equivalently, as the intersection of all \(L^\infty_{\alpha}(\mathbb{R}^d)\) with \(\alpha \in \mathbb{N}\). In other words, \(\mathcal{S}(\mathbb{R}^d)\) contains all bounded functions that essentially decay faster than \(1/|r|^\alpha\) at infinity for all \(\alpha \in \mathbb{N}\). A sequence \((f_i)\) converges in (the topology of) \(\mathcal{S}(\mathbb{R}^d)\) if and only if it converges in all \(L^\infty_{\alpha}(\mathbb{R}^d)\) spaces.
The causal exponential $\rho_\alpha(r) = \mathbb{1}_{[0,\infty)}(r)e^{\alpha r}$ with $\text{Re}(\alpha) < 0$ that is central to linear systems theory is a prototypical example of function included in $\mathcal{S}(\mathbb{R})$.

3.1.3 Nuclear spaces

Defining nuclear spaces is neither easy nor particularly intuitive. Fortunately, for our purpose in this book, knowing the definition is not necessary. We shall simply assert that certain function spaces are nuclear, in order to use certain results that are true for nuclear spaces (specifically, the Minlos-Bochner theorem, see below). For the sake of completeness, a general definition of nuclear spaces is given at the end of this section, but this definition may safely be skipped without compromising the presentation.

Specifically, it will be sufficient for us to know that the spaces $\mathcal{D}(\mathbb{R}^d)$ and $\mathcal{S}(\mathbb{R}^d)$, which we shall shortly define, are nuclear, as are the Cartesian products and powers of nuclear spaces, and their closed subspaces.

To define these spaces, we need to identify their members, as well as the criterion of convergence for sequences in the space.

The space $\mathcal{D}(\mathbb{R}^d)$

The space of compactly supported smooth test functions is denoted by $\mathcal{D}(\mathbb{R}^d)$. It consists of infinitely differentiable functions with compact support in $\mathbb{R}^d$. To define its topology, we provide the following criterion for convergence in $\mathcal{D}(\mathbb{R}^d)$:

A sequence $(\varphi_i)$ of functions in $\mathcal{D}(\mathbb{R}^d)$ is said to converge (in the topology of $\mathcal{D}(\mathbb{R}^d)$) if

1) There exists a compact (here, meaning closed and bounded) subset $K$ of $\mathbb{R}^d$ such that all $\varphi_i$ are supported inside $K$.

2) The sequence $(\varphi_i)$ converges in all of the seminorms

$$\|\varphi\|_{K,n} = \sup_{r \in K} |\partial_1^{n_1} \cdots \partial_d^{n_d} \varphi(r)|$$

for all $n \in \mathbb{N}^d$.

Here, $n = (n_1, \ldots, n_d) \in \mathbb{N}^d$ is what is called a multi-index, and $\partial^n$ is shorthand for the partial derivative $\partial_1^{n_1} \cdots \partial_d^{n_d}$. We take advantage of the present opportunity also to introduce two other notations: $|n|$ for $\sum_{i=1}^d |n_i|$ and $r^n$ for the product $\prod_{i=1}^d r_i^{n_i}$.

The space $\mathcal{D}(\mathbb{R}^d)$ is nuclear (for a proof, see for instance [GV64]).

The Schwartz space $\mathcal{S}(\mathbb{R}^d)$

The Schwartz space $\mathcal{S}(\mathbb{R}^d)$ or the space of so-called smooth and rapidly decaying test functions, denoted as $\mathcal{S}(\mathbb{R}^d)$, consists of infinitely differentiable functions $\varphi$ on $\mathbb{R}^d$, for which all of the seminorms defined below are finite:

$$\|\varphi\|_{m,n} = \sup_{r \in \mathbb{R}^d} |r^m \partial^n \varphi(r)|$$

for all $m, n \in \mathbb{N}^d$.

In other words, $\mathcal{S}(\mathbb{R}^d)$ is populated by functions that, together with all of their derivatives, decay faster than the inverse of any polynomial at infinity.
The topology of $\mathcal{S}(\mathbb{R}^d)$ is defined by positing that a sequence $(\varphi_i)$ converges in $\mathcal{S}(\mathbb{R}^d)$ if and only if it converges in all of the above seminorms.

The Schwartz space has the remarkable property that its complexification is invariant under the Fourier transform. In other words, the Fourier transform, defined by the integral
\[ \hat{\varphi}(\omega) = \mathcal{F}(\varphi)(\omega) = \int_{\mathbb{R}^d} e^{-i \langle r, \omega \rangle} \varphi(r) \, dr, \]
an inverted by
\[ \hat{\varphi}(\omega) = \mathcal{F}^{-1}[\hat{\varphi}](r) = \int_{\mathbb{R}^d} e^{i \langle r, \omega \rangle} \hat{\varphi}(\omega) \frac{d\omega}{(2\pi)^d}, \]
is a continuous map from $\mathcal{S}(\mathbb{R}^d)$ into itself. Our convention here is to use $\omega \in \mathbb{R}^d$ as the generic Fourier-domain index variable.

In addition, both $\mathcal{S}(\mathbb{R}^d)$ and $\mathcal{D}(\mathbb{R}^d)$ are closed and continuous under differentiation of any order and multiplication by polynomials. Lastly, they are included in $\mathcal{R}(\mathbb{R}^d)$ and hence in all the Lebesgue spaces, $L_p(\mathbb{R}^d)$, which do not require any smoothness.

**General definition of nuclear spaces**

Defining a nuclear space requires us to define nuclear operators. These are operators that can be approximated by operators of finite rank in a certain sense (an operator between vector spaces is of finite rank if its range is finite-dimensional).

We first recall the notation $\ell^p(\mathbb{N}), 1 \leq p < \infty$, for the space of $p$-summable sequences; that is, sequences $c = (c_i)_{i \in \mathbb{N}}$ for which
\[ \sum_{i \in \mathbb{N}} |c_i|^p \]
is finite. We also denote by $\ell^\infty(\mathbb{N})$ the space of all bounded sequences.

In a complete normed space $\mathcal{Y}$, let $(\psi_i)_{i \in \mathbb{N}}$ be a sequence with bounded norm (i.e., $\|\psi_i\| \leq M$ for some $M \in \mathbb{R}$ and all $i \in \mathbb{N}$). We then denote by $M_\psi$ the linear operator $\ell_1(\mathbb{N}) \to \mathcal{Y}$ which maps a sequence $c = (c_i)_{i \in \mathbb{N}}$ in $\ell_1$ to the weighted sum
\[ \sum_{i \in \mathbb{N}} c_i \psi_i \]
in $\mathcal{Y}$ (the sum converges in norm by the triangular inequality).

An operator $\Lambda : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y}$ are complete normed spaces, is called nuclear if there exists a continuous linear operator
\[ \tilde{\Lambda} : \mathcal{X} \to \ell_\infty : \varphi \mapsto \{\alpha_i(\varphi)\}, \]
an operator
\[ \Lambda : \ell_\infty \to \ell_1 : (c_i) \mapsto (\lambda_i c_i) \]
where $\sum_i |\lambda_i| < \infty$, and a bounded sequence $\psi = (\psi_i)$ in $\mathcal{Y}$, such that we can write
\[ \Lambda = M_\psi \Lambda \tilde{\Lambda}. \]
This is equivalent to the following decomposition of $A$ into a sum of rank 1 operators:

$$A : \varphi \mapsto \sum_{i \in N} \lambda_i a_i(\varphi) \psi_i$$

The continuous linear operator $X \rightarrow Y : \varphi \mapsto \lambda_i a_i(\varphi) \psi_i$ is of rank 1 because it maps $X$ into the one-dimensional subspace of $Y$ spanned by $\psi_i$; compare $(\psi_i)$ with a basis and $(a_i(\varphi))$ with the coefficients of $A\varphi$ in this basis.

More generally, given an arbitrary topological vector space $X$ and a complete normed space $Y$, the operator $A : X \rightarrow Y$ is said to be nuclear if there exists a complete normed space $X_1$, a nuclear operator $A_1 : X_1 \rightarrow Y$, and a continuous operator $B : X \rightarrow X_1$, such that

$$A = A_1 B.$$ 

Finally, $X$ is a nuclear space if any continuous linear map $X \rightarrow Y$, where $Y$ is a complete normed space, is nuclear.

### 3.2 Dual spaces and adjoint operators

Given a space $X$, a functional on $X$ is a map $f$ that takes $X$ to the scalar field $\mathbb{R}$ (or $\mathbb{C}$). In other words, $f$ takes a function $\varphi \in X$ as argument and returns the number $f(\varphi)$.

When $X$ is a vector space, we may consider linear functionals on it, where linearity has the same meaning as in Definition 3.1. When $f$ is a linear functional, it is customary to use the notation $\langle \varphi, f \rangle$ in place of $f(\varphi)$.

The set of all linear functionals on $X$, denoted as $X^*$, can be given the structure of a vector space in the obvious way by the identity

$$\langle \varphi, af + bf_0 \rangle = a \langle \varphi, f \rangle + b \langle \varphi, f_0 \rangle,$$

where $\varphi \in X$, $f, f_0 \in X^*$, and $a, b \in \mathbb{R}$ (or $\mathbb{C}$) are arbitrary. The resulting vector space $X^*$ is called the algebraic dual of $X$.

The map from $X \times X^*$ to $\mathbb{R}$ (or $\mathbb{C}$) that takes the pair $(\varphi, f)$ to their so-called scalar product $\langle \varphi, f \rangle$ is then bilinear in the sense that it is linear in each of the arguments $\varphi$ and $f$. Note that the reasoning about linear functionals works both ways so that we can also switch the order of the pairing. This translates into the formal commutativity rule $\langle f, \varphi \rangle = \langle \varphi, f \rangle$ with a dual interpretation of the two sides of the equality.

Given vector spaces $X, Y$ with algebraic duals $X^*, Y^*$ and a linear operator $A : X \rightarrow Y$, the adjoint or transpose of $A$, denoted as $A^*$, is the linear operator $Y^* \rightarrow X^*$ defined by

$$A^* f = f \circ A$$

for any linear functional $f : Y \rightarrow \mathbb{R}$ (or $\mathbb{C}$) in $Y^*$, where $\circ$ denotes composition. The motivation behind the above definition is to have the identity

$$\langle A\varphi, f \rangle = \langle \varphi, A^* f \rangle$$

(3.2)
hold for all \( \varphi \in \mathcal{X} \) and \( f \in \mathcal{Y}^* \).

If \( \mathcal{X} \) is a topological vector space, it is of interest to consider the subspace of \( \mathcal{X}^* \) composed of those linear functionals on \( \mathcal{X} \) that are continuous with respect to the topology of \( \mathcal{X} \). This subspace is denoted as \( \mathcal{X}^0 \) and called the \textit{topological or continuous dual} of \( \mathcal{X} \). Note that, unlike \( \mathcal{X}^* \), the continuous dual generally depends on the topology of \( \mathcal{X} \). In other words, the same vector space \( \mathcal{X} \) with different topologies will generally have different continuous duals.

As a general rule, in this book we shall adopt some standard topologies and only work with the corresponding continuous dual space, which we shall call simply the dual. Also, henceforth, we shall assume the scalar product \( \langle \cdot , \cdot \rangle \) to be restricted to \( \mathcal{X} \times \mathcal{X}' \). There, the space \( \mathcal{X} \) may vary but is necessarily paired with its continuous dual.

Following the restrictions of the previous paragraph, we sometimes say that the adjoint of \( A : \mathcal{X} \to \mathcal{Y} \) \textit{exists}, to mean that the algebraic adjoint \( A^* : \mathcal{Y}^* \to \mathcal{X}^* \), when restricted to \( \mathcal{Y}' \), maps into \( \mathcal{X}' \), so that we can write

\[
\langle A \varphi, f \rangle = \langle \varphi, A^* f \rangle,
\]

where the scalar products on the two sides are now restricted to \( \mathcal{Y} \times \mathcal{Y}' \) and \( \mathcal{X} \times \mathcal{X}' \), respectively.

One can define different topologies on \( \mathcal{X}' \) by providing various criteria for convergence. The only one we shall need to deal with is the \textit{weak-* topology}, which indicates (for a sequential space \( \mathcal{X} \)) that \( (f_i) \) converges to \( f \) in \( \mathcal{X}' \) if and only if

\[
\lim_{i} \langle \varphi, f_i \rangle = \langle \varphi, f \rangle \quad \text{for all } \varphi \in \mathcal{X}.
\]

This is precisely the topology of pointwise convergence for all “points” \( \varphi \in \mathcal{X} \).

We shall now mention some examples.

### 3.2.1 The dual of \( L_p \) spaces

The dual of the Lebesgue space \( L_p(\mathbb{R}^d) \), \( 1 \leq p < \infty \), can be identified with the space \( L_{p'}(\mathbb{R}^d) \) with \( 1 < p' \leq \infty \) satisfying \( 1/p + 1/p' = 1 \), by defining

\[
\langle \varphi, f \rangle = \int_{\mathbb{R}^d} \varphi(r) f(r) \, dr
\]

for \( \varphi \in L_p(\mathbb{R}^d) \) and \( f \in L_{p'}(\mathbb{R}^d) \). In particular, \( L_2(\mathbb{R}^d) \), which is the only Hilbert space of the family, is its own dual.

To see that linear functionals described by the above formula with \( f \in L_{p'} \) are continuous on \( L_p \), we can rely on Hölder’s inequality, which states that

\[
|\langle \varphi, f \rangle| \lesssim \int_{\mathbb{R}^d} |\varphi(r) f(r)| \, dr \leq \| \varphi \|_{L_p} \| f \|_{L_{p'}}
\]

for \( 1 \leq p, p' \leq \infty \) and \( 1/p + 1/p' = 1 \). The special case of this inequality for \( p = 2 \) yields the Cauchy-Schwarz inequality.
3.2.2 The duals of $\mathcal{D}$ and $\mathcal{S}$

In this subsection, we give the mathematical definition of the duals of the nuclear spaces $\mathcal{D}$ and $\mathcal{S}$. A physical interpretation of these definitions is postponed until the next section.

The dual of $\mathcal{D}(\mathbb{R}^d)$, denoted as $\mathcal{D}'(\mathbb{R}^d)$, is the so-called space of distributions over $\mathbb{R}^d$ (although we shall use the term distribution more generally to mean any generalized function in the sense of the next section). Ordinary locally integrable functions\(^2\) (in particular, all $L_p$ functions and all continuous functions), can be identified with elements of $\mathcal{D}'(\mathbb{R}^d)$ by using (3.3). By this, we mean that any locally integrable function $f$ defines a continuous linear functional on $\mathcal{D}(\mathbb{R}^d)$ where, for $\varphi \in \mathcal{D}(\mathbb{R}^d)$, $\langle \varphi, f \rangle$ is given by (3.3). However, not all elements of $\mathcal{D}'(\mathbb{R}^d)$ can be characterized in this way. For instance, the Dirac functional $\delta$ (a.k.a. Dirac impulse), which maps $\varphi \in \mathcal{D}(\mathbb{R}^d)$ to the value $\langle \varphi, \delta \rangle = \varphi(0)$, belongs in $\mathcal{D}'(\mathbb{R}^d)$ but cannot be written as an integral à la (3.3). Even in this and similar cases, we may sometimes write $\int_{\mathbb{R}^d} \varphi(r) f(r) \, dr$, keeping in mind that the integral is no longer a true (i.e., Lebesgue) integral, but simply an alternative notation for $\langle \varphi, f \rangle$.

In similar fashion, the dual of $\mathcal{S}(\mathbb{R}^d)$, denoted as $\mathcal{S}'(\mathbb{R}^d)$, is defined and called the space of tempered (or Schwartz) distributions. Since $\mathcal{D} \subset \mathcal{S}$ and any sequence that converges in the topology of $\mathcal{D}$ also converges in $\mathcal{S}$, it follows that $\mathcal{S}'(\mathbb{R}^d)$ is (can be identified with) a smaller space (i.e., a subspace) of $\mathcal{D}'(\mathbb{R}^d)$. In particular, not every locally-integrable function belongs in $\mathcal{S}'$. For example, locally-integrable functions of exponential growth have no place in $\mathcal{S}'$ as their scalar product with Schwartz test functions via (3.3) is not in general finite (much less continuous). Once again, $\mathcal{S}'(\mathbb{R}^d)$ contains objects that are not functions on $\mathbb{R}^d$ in the true sense of the word. For example, $\delta$ also belongs in $\mathcal{S}'(\mathbb{R}^d)$.

3.2.3 Distinction between Hermitian and duality products

We use the notation $\langle f, g \rangle_{L_2} = \int_{\mathbb{R}^d} f(r) \overline{g(r)} \, dr$ to represent the usual (Hermitian-symmetric) $L_2$ inner product. The latter is defined for $f, g \in L_2(\mathbb{R}^d)$ (the Hilbert space of complex finite-energy functions); it is equivalent to Schwartz' duality product only when the second argument is real-valued (due to the presence of complex conjugation). The corresponding Hermitian adjoint of an operator $A$ is denoted by $A^\dagger$. It is defined as $\langle A^\dagger f, g \rangle_{L_2} = \langle f, Ag \rangle_{L_2} = \langle f, \overline{A} g \rangle$ which implies that $A^\dagger = \overline{A}^\top$. The distinction between both types of adjoints is only relevant when considering signal expansions or analyses in terms of complex basis functions.

The classical Fourier transform is defined as

$$\hat{f}(\omega) = \mathcal{F}\{f\}(\omega) = \int_{\mathbb{R}^d} f(r) e^{-i(r, \omega)} \, dr$$

for any $f \in L_1(\mathbb{R}^d)$. This definition admits a unique extension, $\mathcal{F} : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$, which is an isometry map (Plancherel's theorem). The fact that the Fourier transform preserves the $L_2$ norm of a function (up to a normalization factor) is a direct

\(^2\) A function on $\mathbb{R}^d$ is called locally integrable if its integral over any closed bounded set is finite.
consequence of Parseval’s relation

\[
\langle f, g \rangle_{L^2} = \frac{1}{(2\pi)^d} \langle \hat{f}, \hat{g} \rangle_{L^2},
\]

whose duality product equivalent is \( \langle f, \hat{g} \rangle = \langle \hat{f}, g \rangle \).

3.3 Generalized functions

3.3.1 Intuition and definition

We begin with some considerations regarding the modeling of physical phenomena. Let us suppose that the object of our study is some physical quantity \( f \) that varies in relation to some parameter \( r \in \mathbb{R}^d \) representing space and/or time. We assume that our way of obtaining information about \( f \) is by making measurements that are localized in space-time using sensors (\( \varphi, \psi, \ldots \)). We shall denote the measurement of \( f \) procured by \( \varphi \) as \( \langle \varphi, f \rangle \). Let us suppose that our sensors form a vector space, in the sense that for any two sensors \( \varphi, \psi \) and any two scalars \( a, b \in \mathbb{R} \) (or \( \mathbb{C} \)), there is a real or virtual sensor \( a\varphi + b\psi \) such that

\[
\langle a\varphi + b\psi, f \rangle = a\langle \varphi, f \rangle + b\langle \psi, f \rangle.
\]

In addition, we may reasonably suppose that the phenomenon under observation has some form of continuity, meaning that

\[
\lim_{i} \langle \varphi_i, f \rangle = \langle \varphi, f \rangle,
\]

where \( \langle \varphi_i \rangle \) is a sequence of sensors that tend to \( \varphi \) in a certain sense. We denote the set of all sensors by \( \mathcal{X} \). In the light of the above notions of linear combinations and limits defined in \( \mathcal{X} \), mathematically, the space of sensors then has the structure of a topological vector space.

Given the above properties and the definitions of the previous sections, we conclude that \( f \) represents an element of the continuous dual \( \mathcal{X}' \) of \( \mathcal{X} \). Given that our sensors, as previously noted, are assumed to be localized in \( \mathbb{R}^d \), we may model them as compactly supported or rapidly decaying functions on \( \mathbb{R}^d \), denoted by the same symbols (\( \varphi, \psi, \ldots \)) and, in the case where \( f \) also corresponds to a function on \( \mathbb{R}^d \), relate the observation \( \langle \varphi, f \rangle \) to the functional form of \( \varphi \) and \( f \) by the identity

\[
\langle \varphi, f \rangle = \int_{\mathbb{R}^d} \varphi(r) f(r) \, dr.
\]

We exclude from consideration those functions \( f \) for which the above integral is undefined or infinite for some \( \varphi \in \mathcal{X} \).

However, we are not limited to taking \( f \) to be a true function of \( r \in \mathbb{R}^d \). By requiring our sensor or test functions to be smooth, we can permit \( f \) to become singular; that is, to depend on the value of \( \varphi \) and/or of its derivatives at isolated points/crves

\footnote{The connection with previous sections should already be apparent from this choice of notation.}
inside \( \mathbb{R}^d \). An example of a singular generalized function \( f \), which we have already noted, is the Dirac distribution (or impulse) \( \delta \) that measures the value of \( \varphi \) at the single point \( r = 0 \) (i.e., \( \langle \varphi, \delta \rangle = \varphi(0) \)).

Mathematically, we define generalized functions as members of the continuous dual \( X' \) of a nuclear space \( X \) of functions, such as \( \mathcal{D}(\mathbb{R}^d) \) or \( \mathcal{S}(\mathbb{R}^d) \).

3.3.2 Operations on generalized functions

Following (3.2), any continuous linear operator \( \mathcal{D} \rightarrow \mathcal{D} \) or \( \mathcal{S} \rightarrow \mathcal{S} \) can be transposed to define a continuous linear operator \( \mathcal{D}^* \rightarrow \mathcal{D}^* \) or \( \mathcal{S}^* \rightarrow \mathcal{S}^* \). In particular, since \( \mathcal{D}(\mathbb{R}^d) \) and \( \mathcal{S}(\mathbb{R}^d) \) are closed under differentiation, we can define derivatives of distributions.

First, note that, formally,

\[
\langle \partial^n \varphi, f \rangle = \langle \varphi, \partial^n f \rangle.
\]

Now, using integration by parts in (3.3), for \( \varphi, f \) in \( \mathcal{D}(\mathbb{R}^d) \) or \( \mathcal{S}(\mathbb{R}^d) \) we see that \( \partial^n = (-1)^m \partial^n \). In other words, we can write

\[
\langle \varphi, \partial^n f \rangle = (-1)^n \langle \varphi, \partial^n f \rangle.
\]  

(3.5)

The idea is then to use (3.5) as the defining formula in order to extend the action of the derivative operator \( \partial^n \) for any \( f \in \mathcal{D}'(\mathbb{R}^d) \) or \( \mathcal{S}'(\mathbb{R}^d) \).

Formulas for scaling, shifting (translation), rotation, and other geometric transformations of distributions are obtained in a similar manner. For instance, the translation by \( r_0 \) of a generalized function \( f \) is defined via the identity

\[
\langle \varphi, f(\cdot - r_0) \rangle = \langle \varphi(\cdot + r_0), f \rangle.
\]

More generally, we give the following definition.

**Definition 3.3 (Dual extension principle)** Given operators \( U, U^* : \mathcal{S}(\mathbb{R}^d) \rightarrow \mathcal{S}(\mathbb{R}^d) \) that form an adjoint pair on \( \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \), we extend their action to \( \mathcal{S}'(\mathbb{R}^d) \rightarrow \mathcal{S}'(\mathbb{R}^d) \) by defining \( Uf \) and \( U^* f \) so as to have

\[
\langle \varphi, Uf \rangle = \langle U^* \varphi, f \rangle,
\]

\[
\langle \varphi, U^* f \rangle = \langle U \varphi, f \rangle,
\]

for all \( f \). A similar definition gives the extension of adjoint pairs \( \mathcal{D}(\mathbb{R}^d) \rightarrow \mathcal{D}(\mathbb{R}^d) \) to operators \( \mathcal{D}'(\mathbb{R}^d) \rightarrow \mathcal{D}'(\mathbb{R}^d) \).

Examples of operators \( \mathcal{S}(\mathbb{R}^d) \rightarrow \mathcal{S}(\mathbb{R}^d) \) that can be extended in the above fashion include derivatives, rotations, scaling, translation, time-reversal, and multiplication by smooth functions of slow growth in the space-time domain. The other fundamental operation is the Fourier transform which is treated in the next section.
3.3 Generalized functions

3.3.3 The Fourier transform of generalized functions

We have already noted that the Fourier transform \( \mathcal{F} \) is a reversible operator that maps the (complexified) space \( \mathcal{S}(\mathbb{R}^d) \) into itself. The additional relevant property is that \( \mathcal{F} \) is self-adjoint: \( \langle \phi, \mathcal{F} \psi \rangle = \langle \mathcal{F} \phi, \psi \rangle \), for all \( \phi, \psi \in \mathcal{S}(\mathbb{R}^d) \). This helps us specifying the generalized Fourier transform of distributions in accordance with the general extension principle in Definition 3.3.

**Definition 3.4** The generalized Fourier transform of a distribution \( f \in \mathcal{S}'(\mathbb{R}^d) \) is the distribution \( \hat{f} = \mathcal{F}\{f\} \in \mathcal{S}'(\mathbb{R}^d) \) that satisfies

\[
\langle \phi, \hat{f} \rangle = \langle \hat{\phi}, f \rangle
\]

for all \( \phi \in \mathcal{S} \), where \( \hat{\phi} = \mathcal{F}\{\phi\} \) is the classical Fourier transform of \( \phi \) given by the integral

\[
\hat{\phi}(\omega) = \int_{\mathbb{R}^d} e^{-i(\omega, r)} \phi(r) \, dr.
\]

For example, since we have

\[
\int_{\mathbb{R}^d} \phi(r) \, dr = \langle \phi, 1 \rangle = \hat{\phi}(0) = \langle \hat{\phi}, \delta \rangle,
\]

we conclude that the (generalized) Fourier transform of \( \delta \) is the constant function 1.

The fundamental property of the generalized Fourier transform is that it maps \( \mathcal{S}'(\mathbb{R}^d) \) into itself and that it is invertible with \( \mathcal{F}^{-1} = \frac{1}{(2\pi)^d} \mathcal{F} \) where \( \mathcal{F}\{f\} = \mathcal{F}\{f^\#\} \). This quasi self-reversibility—also expressed by the first row of Table 3.3—implies that any operation on generalized functions that is admissible in the space/time domain has its counterpart in the Fourier domain, and vice versa. For instance, the multiplication with a smooth function in the Fourier domain corresponds to a convolution in the signal domain. Consequently, the familiar functional identities concerning the

<table>
<thead>
<tr>
<th>Temporal or spatial domain</th>
<th>Fourier domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{f}(r) = \mathcal{F}{f}(r) )</td>
<td>( (2\pi)^d f(-\omega) )</td>
</tr>
<tr>
<td>( f^#(r) = f(-r) )</td>
<td>( \hat{f}(-\omega) = \hat{f}^#(\omega) )</td>
</tr>
<tr>
<td>( \hat{f}(r) )</td>
<td>( \frac{1}{</td>
</tr>
<tr>
<td>( f(A^T r) )</td>
<td>( \hat{f}(-\omega) )</td>
</tr>
<tr>
<td>( f(r - r_0) )</td>
<td>( e^{-i(r_0, \omega)} \hat{f}(\omega) )</td>
</tr>
<tr>
<td>( e^{j(r, \omega_0)} f(r) )</td>
<td>( \hat{f}(\omega - \omega_0) )</td>
</tr>
<tr>
<td>( \partial^n f(r) )</td>
<td>( (j\omega)^n \hat{f}(\omega) )</td>
</tr>
<tr>
<td>( r^n f(r) )</td>
<td>( j^n \partial^n \hat{f}(\omega) )</td>
</tr>
<tr>
<td>( (g * f)(r) )</td>
<td>( \hat{g}(\omega) \hat{f}(\omega) )</td>
</tr>
<tr>
<td>( g(r) f(r) )</td>
<td>( (2\pi)^{-d} (\hat{g} * \hat{f})(\omega) )</td>
</tr>
</tbody>
</table>

Table 3.3 Basic properties of the (generalized) Fourier transform.
classical Fourier transform such as the formulas for change of variables, differentiation, among others, also hold true for this generalization. These are summarized in Table 3.3.

In addition, the reader can find in Appendix A a table of Fourier transforms of some important singular generalized functions in one and several variables.

3.3.4 The kernel theorem

The kernel theorem provides a characterization of continuous operators $\mathcal{X} \to \mathcal{X}'$ (with respect to the nuclear topology on $\mathcal{X}$ and the weak-* topology on $\mathcal{X}'$). We shall state a version of the theorem for $\mathcal{X} = \mathcal{S}(\mathbb{R}^d)$, which is the one we shall use. The version for $\mathcal{D}(\mathbb{R}^d)$ is obtained by replacing the symbol $\mathcal{S}$ with $\mathcal{D}$ everywhere in the statement of the theorem.

**THEOREM 3.1 (Schwartz’ kernel theorem: first form)** Every continuous linear operator $A : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ can be written in the form

$$\varphi(r) \mapsto A(\varphi)(r) = \int_{\mathbb{R}^d} \varphi(s)a(r, s) \, ds,$$

(3.6)

where $a(\cdot, \cdot)$ is a generalized function in $\mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$.

We can interpret the above formula as some sort of continuous-domain matrix-vector product, where $r, s$ play the role of the row and column indices, respectively (see the list of analogies in Table 3.1). This characterization of continuous linear operators as infinite-dimensional matrix-vector products partly justifies our earlier statement that nuclear spaces “resemble” finite-dimensional spaces in fundamental ways.

The kernel $a \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ associated with the linear operator $A$ can be identified as

$$a(\cdot, r') = A[\delta(\cdot - r')],$$

(3.7)

which corresponds to making the formal substitution $\varphi = \delta(\cdot - r')$ in (3.6). One can therefore view $a(\cdot, r')$ as the generalized impulse response of $A$.

An equivalent statement of Theorem 3.1 is as follows.

**THEOREM 3.2 (Schwartz’s kernel theorem: second form)** Every continuous bilinear form $l : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$ (or $\mathbb{C}$) can be written as

$$l(\varphi_1, \varphi_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi_1(r) \varphi_2(s) a(r, s) \, ds \, dr,$$

(3.8)

where the kernel $a$ is some generalized function in $\mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$.

One may argue that the signal-domain notation that is used in both (3.6) and (3.8) is somewhat abusive since $A(\varphi)$ and $a$ do not necessarily have an interpretation as
3.3 Generalized functions

classical functions (see statement on the notation in Section 3.1.1). The purists therefore prefer to denote (3.8) as

\[ l(\varphi_1, \varphi_2) = (\varphi_1 \otimes \varphi_2, a) \]  

(3.9)

with \((\varphi_1 \otimes \varphi_2)(r, s) = \varphi_1(r)\varphi_2(s)\) for all \(\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)\).

The connection between the representations (3.6)-(3.9) is clarified by relating the continuous bilinear form \(l\) to the underlying continuous linear operator \(A : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)\) by means of the identity

\[ l(\varphi_1, \varphi) = (\varphi_1, A\varphi), \]

where \(A\varphi \in \mathcal{S}'(\mathbb{R}^d)\) is the generalized function specified by (3.6) or, equivalently, by the inner "integral" (duality product) with respect to \(s\) in (3.8).

3.3.5 Linear shift-invariant (LSI) operators and convolutions

Let \(S_{r_0}\) denote the shift operator \(\varphi \mapsto \varphi(-r_0)\). We call an operator \(U\) shift-invariant if \(US_{r_0} = S_{r_0}U\) for all \(r_0 \in \mathbb{R}^d\).

As a corollary of the kernel theorem, we have the following characterization of linear shift-invariant (LSI) operators \(\mathcal{F} \to \mathcal{F}'\) (and a similar characterization for those \(\mathcal{D} \to \mathcal{D}'\)).

**Corollary 3.3** Every continuous linear shift-invariant operator \(\mathcal{F}(\mathbb{R}^d) \to \mathcal{F}'(\mathbb{R}^d)\) can be written as a convolution

\[ \varphi(r) \mapsto (\varphi \ast h)(r) = \int_{\mathbb{R}^d} \varphi(s)h(r-s) \, ds \]

with some generalized function \(h \in \mathcal{S}'(\mathbb{R}^d)\).

The idea there is that the kernel (or generalized impulse response) in (3.6) is a function of the relative displacement only: \(a(r, s) = h(r-s)\) (shift-invariance property).

Moreover, in this case we have the convolution-multiplication formula

\[ \mathcal{F}(h \ast \varphi) = \hat{\varphi} \hat{h}. \]  

(3.10)

Note that the convolution of a test function and a distribution is in general a distribution. The latter is smooth (and therefore equivalent to an ordinary function), but not necessarily rapidly decaying. However, \(\varphi \ast h\) will once again belong continuously to \(\mathcal{F}\) if \(\hat{h}\), the Fourier transform of \(h\), is a smooth (infinitely differentiable) function with at most polynomial growth at infinity because the smoothness of \(\hat{h}\) translates into \(h\) having rapid decay in the spatio-temporal domain, and vice versa. In particular, we note that the condition is met when \(h \in \mathcal{S}(\mathbb{R}^d)\) (since \(r^n h(r) \in L_1(\mathbb{R}^d)\) for any \(n \in \mathbb{N}^d\)). A classical situation in dimension \(d = 1\) where the decay is guaranteed to be exponential is when the Fourier transform of \(h\) is a rational transfer function of the form

\[ \hat{h}(\omega) = C_0 \prod_{m=1}^M (\omega - z_m) / \prod_{n=1}^N (\omega - p_n) \]
with no purely imaginary pole (i.e., with \( \text{Re}(p_n) \neq 0, 1 \leq n \leq N \)).\textsuperscript{4}

Since any sequence that converges in some \( L_p \) space, with \( 1 \leq p \leq \infty \), also converges in \( \mathcal{S}' \), the kernel theorem implies that any continuous linear operator \( \mathcal{S} (\mathbb{R}^d) \rightarrow L_p(\mathbb{R}^d) \) can be written in the form specified by (3.6).

In defining the convolution of two distributions, some caution should be exerted. To be consistent with the previous definitions, we can view convolutions as continuous linear shift-invariant operators. The convolution of two distributions will then correspond to the composition of two LSI operators. To fix ideas, let us take two distributions \( f \) and \( h \), with corresponding operators \( A_f \) and \( A_h \). We then wish to identify \( f \ast h \) with the composition \( A_f A_h \). However, note that, by the kernel theorem, \( A_f \) and \( A_h \) are initially defined \( \mathcal{S} \rightarrow \mathcal{S}' \). Since the codomain of \( A_h \) (the space \( \mathcal{S}' \)) does not match the domain of \( A_f \) (the space \( \mathcal{S} \)), this composition is a priori undefined.

There are two principal situations where we can get around the above limitation. The first is where the range of \( A_h \) is limited to \( \mathcal{S} \) (i.e., \( A_h \) maps \( \mathcal{S} \) to itself instead of the much larger \( \mathcal{S}' \)). This is the case for the distributions with a smooth Fourier transform that we discussed previously.

The second situation where we may define the convolution of \( f \) and \( h \) is when the range of \( A_h \) can be restricted to some space \( \mathcal{X} \) (i.e, \( A_h : \mathcal{S} \rightarrow \mathcal{X} \)), and furthermore, \( A_f \) has a continuous extension to \( \mathcal{X} \); that is, we can extend it as \( A_f : \mathcal{X} \rightarrow \mathcal{S}' \).

An important example of the second situation is when the distributions in question belong to the spaces \( L_p(\mathbb{R}^d) \) and \( L_q(\mathbb{R}^d) \) with \( 1 \leq p, q \leq \infty \) and \( 1/p + 1/q \leq 1 \). In this case, their convolution is well-defined and can be identified with a function in \( L_r(\mathbb{R}^d), 1 \leq r \leq \infty \), with

\[
1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}.
\]

Moreover, for \( f \in L_p(\mathbb{R}^d) \) and \( h \in L_q(\mathbb{R}^d) \), we have

\[
\|f \ast h\|_{L_r} \leq \|f\|_{L_p} \|h\|_{L_q}.
\]

This result is Young’s inequality for convolutions. An important special case of this identity, most useful in derivations, is obtained for \( q = 1 \) and \( p = r \): 

\[
\|h \ast f\|_{L_p} \leq \|h\|_{L_1} \|f\|_{L_p}.
\]

The latter formula indicates that \( L_p(\mathbb{R}^d) \) spaces are “stable” under convolution with elements of \( L_1(\mathbb{R}^d) \) (stable filters).

3.3.6 Convolution operators on \( L_p(\mathbb{R}^d) \)

While the condition \( h \in L_1(\mathbb{R}^d) \) in (3.11) is very useful in practice and plays a central role in the classical theory of linear systems, it does not cover the entire range of bounded convolution operators on \( L_p(\mathbb{R}^d) \). Here we shall be more precise and characterize the complete class of such operators for the cases \( p = 1, 2, \infty \). In harmonic

\textsuperscript{4} For \( M \) or \( N = 0 \), we shall take the corresponding product to be equal to 1.
3.3 Generalized functions

analysis, these operators are commonly referred to as $L_p$ Fourier multipliers using (3.10) as starting point for their definition.

**Definition 3.5 (Fourier multiplier)** An operator $T : L_p(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$ is called a $L_p$ Fourier multiplier if it is continuous on $L_p(\mathbb{R}^d)$ and can be represented as $Tf = \mathcal{F}^{-1} \{ \hat{f} H \}$. The function $H : \mathbb{R}^d \to \mathbb{C}$ is the frequency response of the underlying filter.

The first observation is that the definition guarantees linearity and shift-invariance. Moreover, since $\mathcal{S}(\mathbb{R}^d) \subset L_p(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d)$, the multiplier operator can be written as a convolution $Tf = h \ast f$ (see Corollary 3.3) where $h \in \mathcal{S}'(\mathbb{R}^d)$ is the impulse responses of the operator $T$: $h = \mathcal{F}^{-1} \{ H \} = T\delta$. Conversely, we also have that $H = \hat{h} = \mathcal{F} \{ h \}$.

Since we are dealing with a linear operator on a normed vector space, we can rely on the equivalence between continuity (in accordance with Definition 3.2) and the boundedness of the operator.

**Definition 3.6 (Operator norm)** The norm of the linear operator $T : L_p(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$ is given by

$$
\|T\|_{L_p} = \sup_{f \in L_p(\mathbb{R}^d) \setminus \{0\}} \frac{\|Tf\|_{L_p}}{\|f\|_{L_p}}.
$$

The operator is said to be bounded if its norm is finite.

It practice, it is often sufficient to work out bounds for the extreme cases (e.g., $p = 1, +\infty$) and to then invoke the Riesz-Thorin interpolation theorem to extend the results to the $p$ values in-between.

**Theorem 3.4 (Riesz-Thorin)** Let $T$ be a linear operator that is bounded on $L_{p_1}(\mathbb{R}^d)$ as well as on $L_{p_2}(\mathbb{R}^d)$ with $1 \leq p_1 \leq p_2$. Then, $T$ is also bounded for any $p \in [p_1, p_2]$ in the sense that there exist constants $C_p = \|T\|_{L_p} < \infty$ such that

$$
\|Tf\|_{L_p} \leq C_p \|f\|_{L_p}
$$

for all $f \in L_p(\mathbb{R}^d)$.

The next theorem summarizes the main results that are available on the characterization of convolution operators on $L_p(\mathbb{R}^d)$.

**Theorem 3.5 (Characterization of $L_p$ Fourier multipliers)** Let $T$ be a Fourier-multiplier operator with frequency response $H : \mathbb{R}^d \to \mathbb{C}$ and (generalized) impulse response $h = \mathcal{F}^{-1} \{ H \} = T\delta$. Then, the following statements apply:

1) The operator $T$ is an $L_1$ Fourier multiplier if and only if there exists of a finite complex-valued Borel measure denoted by $\mu_h$ such that $H(\omega) = \int_{\mathbb{R}^d} e^{-i\omega \cdot r} \mu_h(dr)$.

2) The operator $T$ is an $L_\infty$ Fourier multiplier if and only if $H$ is the Fourier transform of a finite complex-valued Borel measure, as stated in 1).

3) The operator $T$ is an $L_2$ Fourier multiplier if and only if $H = \hat{h} \in L_\infty(\mathbb{R}^d)$.
The corresponding operator norms are
\[ \|T\|_{L_2} = \|T\|_{L_\infty} = \|\mu_h\|_{TV} \]
\[ \|T\|_{L_p} = \frac{1}{(2\pi)^{d/2}} \|H\|_{L_\infty}, \]
where \( \|\mu_h\|_{TV} \) is the total variation of the underlying measure. Finally, \( T \) is an \( L_p \) Fourier multiplier for the whole range \( 1 \leq p \leq +\infty \) if the condition on \( H \) in 1) or 2) is met with
\[ \|T\|_{L_p} \leq \|\mu_h\|_{TV} = \sup_{\|\varphi\|_{L_\infty} \leq 1} \langle \varphi, h \rangle. \] (3.12)

We note that the above theorem is an extension upon (3.11) since being a finite Borel measure is less restrictive a condition than \( h \in L_1(\mathbb{R}^d) \). To see this, we invoke Lebesgue’s decomposition theorem stating that a finite measure \( \theta \) admits a unique decomposition as
\[ \theta = \theta_{ac} + \theta_{sing}, \]
where \( \theta_{ac} \) is an absolutely-continuous measure and \( \theta_{sing} \) a singular measure whose mass is concentrated on a set whose Lebesgue measure is zero. If \( \theta_{sing} = 0 \), then there exists a unique function \( h \in L_1(\mathbb{R}^d) \)—the Radon-Nikodym derivative of \( \theta \) with respect to the Lebesgue measure—such that
\[ \int_{\mathbb{R}^d} \varphi(r) \mu_h(dr) = \int_{\mathbb{R}^d} \varphi(r) h(r) \, dr = \langle \varphi, h \rangle. \]

We then recall Hölder’s inequality (3.4) with \( (p, p') = (\infty, 1) \),
\[ |\langle \varphi, h \rangle| \leq \|\varphi\|_{L_\infty} \|h\|_{L_1}, \]
to see that the total variation norm defined by (3.12) reduces to the \( L_1 \) norm: \( \|\mu_h\|_{TV} = \|h\|_{L_1} \). Under those circumstances, there is an equivalence between (3.11) and (3.12).

More generally, when \( \theta_{sing} \neq 0 \), we can make the same kind of association between \( \mu_h \) and a generalized function \( h \) which is no longer in \( L_1(\mathbb{R}^d) \). The typical case is when \( \theta_{sing} \) is a discrete measure which results in a generalized function \( h_{sing} = \sum_k h_k \delta(\cdot - r_k) \) that is a sum of Dirac impulses. The total variation of \( \mu_h \) is then given by \( \|\mu_h\|_{TV} = \|h_{ac}\|_{L_1} + \sum_k |h_k| \).

Statement 3) in Theorem 3.5 is a consequence of Parseval’s identity. It is consistent with the intuition that a “stable” filter should have a bounded frequency response, as a minimal requirement. The class of convolution kernels that satisfy this condition are sometimes called pseudo-measures. These are more-general entities than measures because the Fourier transform of a finite measure is necessarily uniformly continuous in addition to being bounded.

The last result in Theorem 3.5 is obtained by interpolation between Statements 1) and 2) using the Riesz-Thorin theorem. The extent to which the TV condition on \( h \) can be relaxed for \( p \neq 1, 2, \infty \) is not yet settled and considered to be a difficult mathematical problem. A limit example of a 1-D convolution operator that is bounded
for $1 < p < \infty$ (see Theorem 3.6 below), but fails to meet the necessary and sufficient TV condition for $p = 1, \infty$, is the Hilbert transform. Its frequency response is $H_{\text{Hilbert}}(\omega) = -\text{sign}(\omega)$, which is bounded since $|H_{\text{Hilbert}}(\omega)| = 1$ (all-pass filter), but which is not uniformly continuous because of the jump at $\omega = 0$. Its impulse response is the generalized function $h(r) = \frac{1}{\pi r}$, which is not included in $L^1(\mathbb{R})$ for two reasons: the singularity at $r = 0$ and the lack of sufficient decay at infinity.

The case of the Hilbert transform is covered by Mikhlin’s multiplier theorem which provides a sufficient condition on the frequency response of a filter for $L^p$ stability.

**Theorem 3.6 (Mikhlin)** A Fourier-multiplier operator is bounded in $L^p(\mathbb{R}^d)$ for $1 < p < \infty$ if its frequency response $H : \mathbb{R}^d \to \mathbb{C}$ satisfies the differential estimate

$$
|\omega^n \partial^n H(\omega)| \leq C_n \quad \text{for all } |n| \leq (d/2) + 1.
$$

Mikhlin’s condition, which can absorb some degree of discontinuity at the origin, is easy to check in practice. It is stronger than the minimal boundedness requirement for $p = 2$.

### 3.4 Probability theory

#### 3.4.1 Probability measures

Probability measures are mathematical constructs that permit us to assign numbers (probabilities) between 0 (almost impossible) to 1 (almost sure) to events. An event is modeled by a subset $A$ of the universal set $\Omega_X$ of all outcomes of a certain experiment $X$, which is assumed to be known. The symbol $\mathcal{P}_X(A)$ then gives the probability that some element of $A$ occurs as the outcome of experiment $X$. Note that, in general, we may assign probabilities only to some subsets of $\Omega_X$. We shall denote the collection of all subsets of $\Omega_X$ for which $\mathcal{P}_X$ is defined as $\mathcal{S}_X$.

The probability measure $\mathcal{P}_X$ then corresponds to a function $\mathcal{S}_X \to [0,1]$. The triple $(\Omega_X, \mathcal{S}_X, \mathcal{P}_X)$ is called a probability space.

Frequently, the collection $\mathcal{S}_X$ contains open and closed sets, as well as their countable unions and intersections, collectively known as Borel sets. In this case we call $\mathcal{P}_X$ a Borel probability measure.

An important application of the notion of probability is in computing the “average” value of some (real- or complex-valued) quantity $f$ that depends on the outcome in $\Omega_X$. This quantity, the computation of which we shall discuss shortly, is called the expected value of $f$, and is denoted as $E[f(X)]$.

An important context for probabilistic computations is when the outcome of $X$ can be encoded as a finite-dimensional numerical sequence, which implies that we can identify $\Omega_X$ with $\mathbb{R}^n$ (or a subset thereof). In this case, within the proper mathematical setting, we can find a (generalized) function $p_X$, called the probability dis-
distribution\textsuperscript{5} or density function (pdf) of $X$, such that
\[ p_X(A) = \int_A p_X(x) \, dx \]
for suitable subsets $A$ of $\mathbb{R}^n$. \textsuperscript{6}

More generally, the expected value of $f : X \to \mathbb{C}$ is here given by
\[ \mathbb{E}(f(X)) = \int_{\mathbb{R}^n} f(x) p_X(x) \, dx. \tag{3.13} \]
We say “more generally” because $p_X(A)$ can be seen as the expected value of the indicator function $\mathbb{1}_A(X)$. Since the integral of complex-valued $f$ can be written as the sum of its real and imaginary parts, without loss of generality we shall consider only real-valued functions where convenient.

When the outcome of the experiment is a vector with infinitely many coordinates (for instance a function $\mathbb{R} \to \mathbb{R}$), it is typically not possible to characterize probabilities with probability distributions. It is nevertheless still possible to define probability measures on subsets of $\Omega_X$, and also to define the integral (average value) of many a function $f : \mathbb{R}^n \to \mathbb{R}$. In effect, a definition of the integral of $f$ with respect to probability measure $\mathcal{P}_X$ is obtained using a limit of “simple” functions (finite weighted sums of indicator functions) that approximate $f$. For this general definition of the integral we use the notation
\[ \mathbb{E}(f(X)) = \int_{\Omega_X} f(x) \, \mathcal{P}_X(dx), \]
which we may also use, in addition to (3.13), in the case of a finite-dimensional $\Omega_X$.

In general, given a function $f : \Omega_X \to \Omega_Y$ that defines a new outcome $y \in \Omega_Y$ for every outcome $x \in \Omega_X$ of experiment $X$, one can see the result of applying $f$ to the outcome of $X$ as a new experiment $Y$. The probability of an event $B \subset \Omega_Y$ is the same as the combined probability of all outcomes of $X$ that generate an outcome in $B$. Thus, mathematically,
\[ \mathcal{P}_Y(B) = \mathcal{P}_X(f^{-1}(B)) = \mathcal{P}_X \circ f^{-1}(B), \]
where the inverse image $f^{-1}(B)$ is defined as
\[ f^{-1}(B) = \{ x \in \Omega_X : f(x) \in B \}. \]
$\mathcal{P}_Y = \mathcal{P}_X(f^{-1})$ is called the push-forward of $\mathcal{P}_X$ through $f$.

\textsuperscript{5} Probability distributions should not be confused with the distributions in the sense of Schwartz (i.e., generalized functions) that were introduced in Section 3.3. It is important to distinguish the two usages, in part because, as we describe here, in finite dimensions a connection can be made between probability distributions and positive generalized functions.

\textsuperscript{6} In classical probability theory, pdfs are defined as the Radon-Nikodym derivative of a probability measure with respect to some other measure, typically the Lebesgue measure (as we shall assume). This requires the probability measure to be absolutely continuous with respect to the latter measure. The definition of the generalized pdf given here is more permissive, and also includes measures that are singular with respect to the Lebesgue measure (for instance the Dirac measure of a point, for which the generalized pdf is a Dirac distribution). This generalization relies on identifying measures on the Euclidean space with positive linear functionals.
3.4.2 Joint probabilities and independence

When two experiments $X$ and $Y$ with probabilities $\mathcal{P}_X$ and $\mathcal{P}_Y$ are considered simultaneously, one can imagine a joint probability space $(\Omega_{X,Y}, \mathcal{G}_{X,Y}, \mathcal{P}_{X,Y})$ that supports both $X$ and $Y$, in the sense that there exist functions $f : \Omega_{X,Y} \to \Omega_X$ and $g : \Omega_{X,Y} \to \Omega_Y$ such that

$$\mathcal{P}_X(A) = \mathcal{P}_{X,Y}(f^{-1}(A)) \quad \text{and} \quad \mathcal{P}_Y(B) = \mathcal{P}_{X,Y}(g^{-1}(B))$$

for all $A \in \mathcal{G}_X$ and $B \in \mathcal{G}_Y$.

The functions $f, g$ above are assumed to be fixed, and the joint event that $A$ occurs for $X$ and $B$ for $Y$, is given by

$$f^{-1}(A) \cap g^{-1}(B).$$

If the outcome of $X$ has no bearing on the outcome of $Y$ and vice-versa, then $X$ and $Y$ are said to be independent. In terms of probabilities, this translates into the probability factorization rule

$$\mathcal{P}_{X,Y}(f^{-1}(A) \cap g^{-1}(B)) = \mathcal{P}_X(A) \cdot \mathcal{P}_Y(B) = \mathcal{P}_{X,Y}(f^{-1}(A)) \cdot \mathcal{P}_{X,Y}(g^{-1}(B)).$$

The above ideas can be extended to any finite collection of experiments $X_1, \ldots, X_M$ (and even to infinite ones, with appropriate precautions and adaptations).

3.4.3 Characteristic functions in finite dimensions

In finite dimensions, given a probability measure $\mathcal{P}_X$ on $\Omega_X = \mathbb{R}^n$, for any vector $\omega \in \mathbb{R}^n$, we can compute the expected value (integral) of the bounded function $x \mapsto e^{i \langle \omega, x \rangle}$. This permits us to define a complex-valued function on $\mathbb{R}^n$ by the formula

$$\hat{\mathcal{P}}_X(\omega) = E[e^{i \langle \omega, x \rangle}] = \int_{\mathbb{R}^n} e^{i \langle \omega, x \rangle} p_X(x) \, dx = \mathcal{F}(p_X)(\omega), \quad (3.14)$$

which corresponds to a slightly different definition of the Fourier transform of the (generalized) probability distribution $p_X$. The convention in probability theory is to define the forward Fourier transform with a positive sign for $\langle \omega, x \rangle$, which is the opposite of the convention used in analysis.

One can prove that $\hat{\mathcal{P}}_X$, as defined above, is always continuous at $0$ with $\hat{\mathcal{P}}_X(0) = 1$, and that it is positive-definite (see Definition B.1 in Appendix B).

Remarkably, the converse of the above fact is also true. We record the latter result, which is due to Bochner, together with the former observation, as Theorem 3.7.

**Theorem 3.7 (Bochner)** Let $\hat{\mathcal{P}}_X : \mathbb{R}^n \to \mathbb{C}$ be a function that is positive-definite, fulfills $\hat{\mathcal{P}}_X(0) = 1$, and is continuous at $0$. Then, there exists a unique Borel probability measure $\mathcal{P}_X$ on $\mathbb{R}^n$, such that

$$\hat{\mathcal{P}}_X(\omega) = \int_{\mathbb{R}^n} e^{i \langle \omega, x \rangle} \mathcal{P}_X(dx) = E[e^{i \langle \omega, x \rangle}].$$

Conversely, the function specified by (3.14) with $p_X(r) \geq 0$ and $\int_{\mathbb{R}^n} p_X(r) \, dr = 1$ is positive-definite, uniformly continuous, and such that $|\hat{\mathcal{P}}_X(\omega)| \leq \hat{\mathcal{P}}_X(0) = 1$. 

The interesting twist (which is due to Lévy) is that the positive-definiteness of $\hat{p}_X$ and its continuity at 0 implies continuity everywhere (as well as boundedness).

Since, by the above theorem, $\hat{p}_X$ uniquely identifies $\mathcal{P}_X$. It is called the characteristic function of probability measure $\mathcal{P}_X$ (recall that the probability measure $\mathcal{P}_X$ is related to the density $p_X$ by $\mathcal{P}_X(E) = \int_E p_X(x) \, dx$ for sets $E$ in the $\sigma$-algebra over $\mathbb{R}^n$).

The next theorem characterizes weak convergence of measures on $\mathbb{R}^n$ in terms of their characteristic functions.

**Theorem 3.8 (Lévy’s continuity theorem)** Let $(\mathcal{P}_X_i)$ be a sequence of probability measures on $\mathbb{R}^n$ with respective sequence of characteristic functions $(\hat{p}_X_i)$. If there exists a function $\hat{p}_X$ such that

$$\lim i \hat{p}_{X_i}(\omega) = \hat{p}_X(\omega)$$

pointwise on $\mathbb{R}^n$, and if, in addition, $\hat{p}_X$ is continuous at 0, then $\hat{p}_X$ is the characteristic function of a probability measure $\mathcal{P}_X$ on $\mathbb{R}^n$. Moreover, $\mathcal{P}_X_i$ converges weakly to $\mathcal{P}_X$, in symbols

$$\mathcal{P}_{X_i} \xrightarrow{w} \mathcal{P}_X,$$

meaning for any continuous function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\lim i \mathbb{E}_{X_i}(f) = \mathbb{E}_X(f).$$

The reciprocal of the above theorem is also true; namely, if $\mathcal{P}_{X_i} \xrightarrow{w} \mathcal{P}_X$, then $\hat{p}_{X_i}(\omega) \to \hat{p}_X(\omega)$ pointwise.

**3.4.4 Characteristic functionals in infinite dimensions**

Given a probability measure $\mathcal{P}_X$ on the continuous dual $\mathcal{X}'$ of some test function space $\mathcal{X}$, one can define an analogue of the finite-dimensional characteristic function, dubbed the characteristic functional of $\mathcal{P}_X$ and denoted as $\hat{\mathcal{P}}_X$, by means of the identity

$$\hat{\mathcal{P}}_X(\varphi) = \mathbb{E}[(e^{i\varphi, X})].$$

(3.15)

Comparing the above definition with (3.14), one notes that $\mathbb{R}^n$, as the domain of the characteristic function $\hat{p}_X$, is now replaced by the space $\mathcal{X}'$ of test functions.

As was the case in finite dimensions, the characteristic functional fulfills two important conditions:

- Positive-definiteness: $\hat{\mathcal{P}}_X$ is positive-definite, in the sense that for any $N$ (test) functions $\varphi_1, \ldots, \varphi_N$, for any $N$, the $N \times N$ matrix with entries $p_{ij} = \hat{\mathcal{P}}_X(\varphi_i - \varphi_j)$ is nonnegative definite.

- Normalization: $\hat{\mathcal{P}}_X(0) = 1$.

In view of the finite-dimensional result (Bochner’s theorem), it is natural to ask if a condition in terms of continuity can be given also in the infinite-dimensional case, so that any functional $\hat{\mathcal{P}}_X$ fulfilling this continuity condition in addition to the above two, uniquely identifies a probability measure on $\mathcal{X}'$. In the case where $\mathcal{X}$ is a
3.5 Generalized random processes and fields

In this section, we present an introduction to the theory of generalized random processes, which is concerned with defining probabilities on function spaces, that is, infinite-dimensional vector spaces with some notion of limit and convergence. We have made the point before that the theory of generalized functions is a natural extension of finite-dimensional linear algebra. The same kind of parallel can be drawn between the theory of generalized stochastic processes and conventional probability calculus (which deals with finite-dimensional random vector variables). Therefore, before getting into more detailed explanations, it is instructive to have a look back at Table 3.2, which provides a side-by-side summary of the primary probabilistic concepts that have been introduced so far. The reader is then referred to Table 3.4, which presents a comparison of finite- and infinite-dimensional “innovation models”. To give the basic idea, in finite dimensions, an “innovation” is a vector in \( \mathbb{R}^n \) of independent identically distributed (i.i.d.) random variables. An “innovation model” is obtained by transforming such a vector by means of a linear operator (a matrix), which embodies the structure of dependencies of the model. In infinite dimensions, the notion of an i.i.d. vector is replaced by that of a random process with independent values at every point (which we shall call an “innovation process”). The transformation is achieved by applying a continuous linear operator which constitutes the generalization of a matrix. The characterization of such models is made
possible by their characteristic functionals, which, as we saw in the previous section, are the infinite-dimensional equivalents of characteristic functions of random variables.

### Table 3.4 Comparison of innovation models in finite- and infinite-dimensional settings.

<table>
<thead>
<tr>
<th>Finite-dimensional</th>
<th>Infinite-dimensional</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard Gaussian i.i.d. vector $W = (W_1, \ldots, W_N)$</td>
<td>standard Gaussian white noise $w$ (\mathcal{F}_w(q) = e^{-\frac{1}{2}q^2}e^{i\langle \psi \rangle}, q \in \mathcal{F})</td>
</tr>
<tr>
<td>multivariate Gaussian vector $X = AW$</td>
<td>Gaussian generalized process $s$ (s = \lambda w \text{ (for continuous } \lambda : \mathcal{F} \to \mathcal{F})) (\mathcal{F}_s(q) = e^{-\frac{1}{2}q^2}q^2)</td>
</tr>
<tr>
<td>general i.i.d. vector $W = (W_1, \ldots, W_N)$ with exponent $f$ (\hat{p}<em>W(\omega) = e^{\sum</em>{n=1}^{N} f(\omega_n)})</td>
<td>general white noise $w$ with Lévy exponent $f$ (\mathcal{F}<em>w(q) = e^{\int</em>{0}^{\infty} f(\psi(r)) dr})</td>
</tr>
<tr>
<td>linear transformation of general i.i.d. random vector $W$ (innovation model) $X = \lambda W$ (\hat{p}_X(\omega) = \hat{p}_W(\lambda^T \omega))</td>
<td>linear transformation of general white noise $s$ (innovation model) $s = \lambda w$ (\mathcal{F}_s(q) = \mathcal{F}_w(\lambda^* q))</td>
</tr>
</tbody>
</table>

3.5.1 Generalized random processes as collections of random variables

A generalized stochastic process\(^7\) is essentially a randomization of the idea of a generalized function (Section 3.3) in much the same way as an ordinary stochastic process is a randomization of the concept of a function.

At a minimum, the definition of a generalized stochastic process $s$ should permit us to associate probabilistic models with observations made using test functions. In other words, to any test function $\varphi$ in some suitable test-function space $\mathcal{X}$ is associated a random variable $s(\varphi)$, also often denoted as $(\varphi, s)$. This is to be contrasted with an observation $s(t)$ at time $t$, which would be modeled by a random variable in the case of an ordinary stochastic process. We shall denote the probability measure of the random variable $s(\varphi)$ as $\mathcal{P}_{s, \varphi}$. Similarly, to any finite collection of observations $(\varphi_n, s)$, $1 \leq n \leq N$, $N \in \mathbb{N}$, corresponds a joint probability measure $\mathcal{P}_{s, \varphi_1, \varphi_N}$ on $\mathbb{R}^N$ (we shall only consider real-valued processes here, and therefore assume the observations to be real-valued).

Moreover, finite families of observations $(\varphi_n, s)$, $1 \leq n \leq N$, and $(\psi_m, s)$, $1 \leq m \leq M$, need to be consistent or compatible to ensure that all computations of the probability

\(^7\) We shall use the terms random/stochastic process and field almost interchangeably. The distinction, in general, lies in the fact that for a random process, the parameter is typically interpreted as time, while for a field, the parameter is typically multi-dimensional and interpreted as spatial or spatio-temporal location.
of an event involving finite observations yield the same value for the probability. In modeling physical phenomena, it is also reasonable to assume some weak form of continuity in the probability of \( \langle \psi, s \rangle \) as a function of \( \psi \).

Mathematically, these requirements are fulfilled by the kind of probabilistic model induced by a cylinder-set probability measure. In other words, a cylinder-set probability measure provides a consistent probabilistic description for all finite sets of observations of some phenomenon \( s \) using test functions \( \psi \in \mathcal{H} \). Furthermore, a cylinder-set probability measure can always be specified via its characteristic functional \( \mathcal{P}_s(\psi) = \mathbb{E}[e^{i \langle \psi, s \rangle}] \), which makes it amenable to analytic computations.

The only conceptual limitation of such a probability model is that, at least a priori, it does not permit us to associate the sample paths of the process with (generalized) functions. Put differently, in this framework, we are not allowed to interpret \( s \) as a random entity belonging to the dual \( \mathcal{H}' \) of \( \mathcal{H} \), since we have not yet defined a proper probability measure on \( \mathcal{H}' \). Doing so involves some additional steps.

### 3.5.2 Generalized random processes as random generalized functions

Fortunately, the above existence and interpretation problem is fully resolved by taking \( \mathcal{H} \) to be a nuclear space, thanks to the Minlos-Bochner theorem (Theorem 3.9). This allows for the extension of the underlying cylinder-set probability measure to a proper (by which here we mean countably additive) probability measure on \( \mathcal{H}' \) (the topological dual of \( \mathcal{H} \)).

In this case, the joint probabilities \( \mathcal{P}_{s_1, \psi_1; \psi_N} \), corresponding to the random variables \( \langle \psi_n, s \rangle = (s, \psi_n) \) for all possible choices of test functions, collectively define a probability measure \( \mathcal{P}_s \) on the infinite-dimensional dual space \( \mathcal{H}' \). This means that we can view \( s \) as an element drawn randomly from \( \mathcal{H}' \) according to the probability law \( \mathcal{P}_s \).

In particular, if we take \( \mathcal{H} \) to be either \( \mathcal{S}(\mathbb{R}^d) \) or \( \mathcal{D}(\mathbb{R}^d) \), then our generalized random process/field will have realizations that are distributions in \( \mathcal{S}'(\mathbb{R}^d) \) or \( \mathcal{D}'(\mathbb{R}^d) \), respectively. We can then also think of \( \langle \psi, s \rangle \) as the measurement of this random object \( s \) by means of some sensor (test function) \( \psi \) in \( \mathcal{S} \) or \( \mathcal{D} \).

Since we shall rely on this fact throughout the book, we reiterate once more that a complete probabilistic characterization of \( s \) as a probability measure on the space \( \mathcal{H}' \) (dual to some nuclear space \( \mathcal{H} \)) is provided by its characteristic functional. The truly powerful aspect of the Minlos-Bochner theorem is that the implication goes both ways: any continuous positive-definite functional \( \mathcal{P}_s : \mathcal{H} \to \mathbb{C} \) with proper normalization identifies a unique probability measure \( \mathcal{P}_s \) on \( \mathcal{H}' \). Therefore, to define a generalized random process \( s \) with realizations in \( \mathcal{H}' \), it suffices to produce a functional \( \mathcal{P}_s : \mathcal{H} \to \mathbb{C} \) with the noted properties.

\(^8\) In fact, \( \mathcal{H}' \) may very well be too small to support such a description (while the algebraic dual, \( \mathcal{H}^* \), can support the measure—by Kolmogorov’s extension theorem—but is too large for many practical purposes). An important example is that of white Gaussian noise, which one may conceive of as associating a Gaussian random variable with variance \( \| \psi \|^2 \) to any test function \( \psi \in L_2 \). However, the “energy” of white Gaussian noise is clearly infinite. Therefore it cannot be modeled as a randomly chosen function in \( (L_2)^\prime = L_2 \).
3.5.3 Determination of statistics from the characteristic functional

The characteristic functional of the generalized random process $s$ contains complete information about its probabilistic properties, and can be used to compute all probabilities, and to derive or verify the probabilistic properties related to $s$.

Most importantly, it can yield the $N$th-order joint probability density of any set of linear observations of $s$ by suitable $N$-dimensional inverse Fourier transformation. This follows from a straightforward manipulation in the domain of the (joint) characteristic function and is recorded for further reference.

**Proposition 3.10** Let $y = (Y_1, \ldots, Y_N)$ with $Y_n = \langle \varphi_n, s \rangle$ where $\varphi_1, \ldots, \varphi_N \in \mathcal{X}$ be a set of linear measurements of the generalized stochastic process $s$ with characteristic functional $\mathcal{P}_s(\varphi) = \mathbb{E}[e^{i\langle \varphi, s \rangle}]$ that is continuous over the function space $\mathcal{X}$. Then,

$$\hat{p}_{(Y_1; Y_N)}(\omega) = \widehat{\mathcal{P}}_{s, \varphi_1; \varphi_N}(\omega) = \mathcal{P}_s \left( \sum_{n=1}^{N} \omega_n \varphi_n \right)$$

and the joint pdf of $y$ is given by

$$p_{(Y_1; Y_N)}(y) = \mathcal{F}^{-1} \{ \hat{p}_{(Y_1; Y_N)} \}(y) = \int_{\mathbb{R}^N} \mathcal{P}_s \left( \sum_{n=1}^{N} \omega_n \varphi_n \right) e^{-j\langle y, \omega \rangle} \frac{d\omega}{(2\pi)^N},$$

where the observation functions $\varphi_n \in \mathcal{X}$ are fixed and $\omega = (\omega_1, \ldots, \omega_N)$ plays the role of the $N$-dimensional Fourier variable.

**Proof** The continuity assumption over the function space $\mathcal{X}$ (which need not be nuclear) ensures that the manipulation is legitimate. Starting from the definition of the characteristic function of $y = (Y_1, \ldots, Y_N)$, we have

$$\hat{p}_{(Y_1; Y_N)}(\omega) = \mathbb{E} \{ \exp \{ j\langle \omega, y \rangle \} \} = \mathbb{E} \left\{ \exp \left\{ j \left( \sum_{n=1}^{N} \omega_n \langle \varphi_n, s \rangle \right) \right\} \right. = \mathbb{E} \left\{ \exp \left\{ j \left( \sum_{n=1}^{N} \omega_n \varphi_n, s \right) \right\} \right\} = \mathcal{P}_s \left( \sum_{n=1}^{N} \omega_n \varphi_n \right)$$

(by linearity of duality product)

(by definition of $\mathcal{P}_s(\varphi)$)

The density $p_{(Y_1; Y_N)}$ is then obtained by inverse (conjugate) Fourier transformation.

Similarly, the formalism allows one to retrieve all first- and second-order moments of the generalized stochastic process $s$. To that end, one considers the mean and
3.5 Generalized random processes and fields

correlation functionals defined and computed as

\[ M_s(\cdot) = \mathbb{E}\{ h(\cdot) , s \} \]

\[ B_s(\cdot_1 , \cdot_2) = \mathbb{E}\{ (\cdot_1 , s) (\cdot_2 , s) \} \]

When the space of test functions is nuclear (\( X = \mathcal{S}(\mathbb{R}^d) \) or \( D(\mathbb{R}^d) \)) and the above quantities are well defined, we can find generalized functions \( m_s \) (the generalized mean) and \( R_s \) (the generalized autocorrelation function) such that

\[ M_s(\cdot) = \int_{\mathbb{R}^d} \varphi(r) m_s(r) \, dr \]

\[ B_s(\cdot_1 , \cdot_2) = \int_{\mathbb{R}^d} \varphi_1(r) \varphi_2(s) R_s(r,s) \, dr \]

The first identity is simply a consequence of \( M_s \) being a continuous linear functional on \( X \), while the second is an application of Schwartz’ kernel theorem (Theorem 3.2).

3.5.4 Operations on generalized stochastic processes

In constructing stochastic models, it is of interest to separate the essential randomness of the models (the “innovation”) from their deterministic structure. Our way of approaching this objective is by encoding the random part in a characteristic functional \( \mathcal{P}_w \), and the deterministic structure of dependencies in an operator \( U \) (or, equivalently, in its adjoint \( U^* \)). In the following paragraphs, we first review the mathematics of this construction, before we come back to, and clarify, the said interpretation. The concepts presented here in an abstract form are illustrated and made intuitive in the remainder of the book.

Given a continuous linear operator \( U : X \rightarrow Y \) with continuous adjoint \( U^* : Y^* \rightarrow X^* \), where \( X,Y \) need not be nuclear, and a functional

\[ \mathcal{P}_w : Y \rightarrow \mathbb{C} \]

that satisfies the three conditions of Theorem 3.9 (continuity, positive-definiteness, and normalization), we obtain a new functional

\[ \tilde{\mathcal{P}}_s : X \rightarrow \mathbb{C} \]

fulfilling the same properties by composing \( \mathcal{P}_w \) and \( U \) as per

\[ \tilde{\mathcal{P}}_s(\varphi) = \mathcal{P}_w(U\varphi) \quad \text{for all } \varphi \in X. \]

Writing

\[ \tilde{\mathcal{P}}_s(\omega \varphi) = \mathbb{E}(e^{i\omega \varphi}) = \hat{\mu}(\varphi, \omega) \]
Figure 3.1 Definition of linear transformation of generalized stochastic processes using characteristic functionals. In this book, we shall focus on innovation models where \( w \) is a white noise process. The operator \( L = U^{-1} \) (if it exists) is called the whitening operator of \( s \) since \( Ls = w \).

\[
P_s(U) = P_w(U) \quad \text{with} \quad U = L^{-1} w
\]

and

\[
\mathcal{F}_w(\omega U \varphi) = \mathbb{E}[e^{i \omega \langle U \varphi, w \rangle}] = \mathcal{F}_s(\varphi)
\]

for generalized processes \( s \) and \( w \), we deduce that the random variables \( \langle \varphi, s \rangle \) and \( \langle U \varphi, w \rangle \) have the same characteristic functions and therefore follow

\[
\langle \varphi, s \rangle = \langle U \varphi, w \rangle \quad \text{in probability law.}
\]

The manipulation that led to Proposition 3.10 shows that a similar relation exists, more generally, for any finite collection of observations \( \langle \varphi_n, s \rangle \) and \( \langle U \varphi_n, w \rangle \), \( 1 \leq n \leq N, N \in \mathbb{N} \).

Therefore, symbolically at least, by the definition of the adjoint \( U^* : \mathcal{Y} \to \mathcal{X}^* \) of \( U \), we may write

\[
\langle \varphi, s \rangle = \langle \varphi, U^* w \rangle.
\]

This seems to indicate that, in a sense, the random model \( s \), which we have defined using (3.18), can be interpreted as the application of \( U^* \) to the original random model \( w \). However, things are complicated by the fact that, unless \( \mathcal{X} \) and \( \mathcal{Y} \) are nuclear spaces, we may not be able to interpret \( w \) and \( s \) as random elements of \( \mathcal{Y} \) and \( \mathcal{X} \), respectively. Therefore the application of \( U^* : \mathcal{Y} \to \mathcal{X}^* \) to \( s \) should be understood to be merely a formal construction.

On the other hand, by requiring \( \mathcal{X} \) to be nuclear and \( \mathcal{Y} \) to be either nuclear or completely normed, we see immediately that \( \mathcal{F}_s : \mathcal{X} \to \mathbb{C} \) fulfills the requirements of the Minlos-Bochner theorem, and thereby defines a generalized random process with realizations in \( \mathcal{X} \).

The previous discussion suggests the following approach to defining generalized random processes: take a continuous positive-definite functional \( \mathcal{F}_w : \mathcal{Y} \to \mathbb{C} \) on
some (nuclear or completely normed) space $\mathcal{Y}$. Then, for any continuous operator $U$ defined from a nuclear space $\mathcal{X}$ into $\mathcal{Y}$, the composition
\[
\mathcal{P}_s = \mathcal{P}_w(U)
\]
is the characteristic functional of a generalized random process $s$ with realizations in $\mathcal{X}'$.

In subsequent chapters, we shall mostly focus on the situation where $U = L^{-1}$ and $U^* = L^{-1}$ for some given (whitening) operator $L$ that admits a continuous inverse in the suitable topology, the typical choice of spaces being $\mathcal{X} = \mathcal{S}(\mathbb{R}^d)$ and $\mathcal{Y} = L_p(\mathbb{R}^d)$. The underlying hypothesis is that one is able to invert the linear operator $U$ and to recover $w$ from $s$, which is formally written as $w = Ls$; that is,
\[
\langle \varphi, w \rangle = \langle \varphi, Ls \rangle, \text{ for all } \varphi \in \mathcal{Y}.
\]

The above ideas are summarized in Figure 3.1.

### 3.5.5 Innovation processes

In a certain sense, the most fundamental class of generalized random processes we can use to play the role of $w$ in the construction of Section 3.5.4 are those with independent values at every point in $\mathbb{R}^d$ [GV64, Chap. 4, pp. 273-288]. The reason is that we can then isolate the spatiotemporal dependency of the probabilistic model in the mixing operator ($U^*$ in Figure 3.1), and attribute randomness to independent contributions (innovations) at geometrically distinct points in the domain. We call such a construction an innovation model.

Let us attempt to make the notion of independence at every point more precise in the context of generalized stochastic processes, where the objects of study are, more accurately, not pointwise observations, but rather observations made through scalar products with test functions. To qualify a generalized process $s$ as having independent values at every point, we therefore require that the random variables $h_1^i, w_i$ and $h_2^i, w_i$ be independent whenever the test functions $\varphi_1^i$ and $\varphi_2^i$ have disjoint supports.

Since the joint characteristic function of independent random variables factorizes (is separable), we can formulate the above property in terms of the characteristic functional $\mathcal{P}_w$ of $w$ as
\[
\mathcal{P}_w(\varphi_1 + \varphi_2) = \mathcal{P}_w(\varphi_1) \mathcal{P}_w(\varphi_2).
\]

An important class of characteristic functionals fulfilling this requirement are those that can be written in the form
\[
\mathcal{P}_w(\varphi) = e^{\int_{\mathbb{R}^d} f(\varphi(r)) \, dr}. \tag{3.19}
\]

To have $\mathcal{P}_w(0) = 1$ (normalization), we require that $f(0) = 0$. The requirement of positive-definiteness narrows down the class of admissible functions $f$ much further, practically to those identified by the Lévy-Khinchine formula. This will be the subject of the greater part of our next chapter.
3.5.6 Example: Filtered white Gaussian noise

In the above framework, we can define *white Gaussian noise* or *innovation* on $\mathbb{R}^d$ as a random element of the space of Schwartz generalized functions, $\mathcal{S}'(\mathbb{R}^d)$, whose characteristic functional is given by

$$c_{P_w}(\varphi) = e^{-\frac{1}{2} \| \varphi \|_2^2}.$$  

Note that this functional is a special instance of (3.19) with $f(\varphi) = -\frac{1}{2} \varphi^2$. The Gaussian appellation is justified by observing that, for any $N$ test functions $\varphi_1, \ldots, \varphi_N$, the random variables $\langle \varphi_1, w \rangle, \ldots, \langle \varphi_N, w \rangle$ are jointly Gaussian. Indeed, we can apply Proposition 3.10 to obtain the joint characteristic function

$$c_{\varphi_1 \ldots \varphi_N}(\omega) = \exp \left( -\frac{1}{2} \sum_{n=1}^{N} \omega_n \langle \varphi_n \rangle \right).$$

By taking the inverse Fourier transform of the above expression, we find that the random variables $\langle \varphi_n, w \rangle$, $n = 1, \ldots, N$, have a multivariate Gaussian distribution with mean $0$ and covariance matrix with entries

$$C_{mn} = \langle \varphi_m, \varphi_n \rangle.$$

The independence of $\langle \varphi_1, w \rangle$ and $\langle \varphi_2, w \rangle$ is obvious whenever $\varphi_1$ and $\varphi_2$ have disjoint support. This justifies calling the process *white*.\footnote{Our notion of *whiteness* in this book goes further than having a white spectrum. By whiteness, we mean that the process is stationary and has truly independent (not merely uncorrelated) values over disjoint sets.}

In this special case, even mere orthogonality of $\varphi_1$ and $\varphi_2$ is enough for independence, since for $\varphi_1 \perp \varphi_2$ we have $C_{mn} = 0$.

From Formulas (3.16) and (3.17), we also find that $w$ has 0 mean and "correlation function" $R_w(r, s) = \delta(r - s)$, which should also be familiar. In fact, this last expression is sometimes used to formally "define" white Gaussian noise.

A *filtered* white Gaussian noise is obtained by applying a continuous convolution (i.e., LSI) operator $U^* : \mathcal{S}' \rightarrow \mathcal{S}'$ to the Gaussian innovation in the sense described in Section 3.5.4.

Let us denote the convolution kernel of the operator $U : \mathcal{S} \rightarrow \mathcal{S}$ (the adjoint of $U^*$) by $h$.\footnote{Recall that, for the convolution to map back into $\mathcal{S}$, $h$ needs to have a smooth Fourier transform, which implies rapid decay in the temporal or spatial domain. This is the case, in particular, for any rational transfer function that lacks purely imaginary poles.} The convolution kernel of $U^* : \mathcal{S}' \rightarrow \mathcal{S}'$ is then $h^\vee$. Following Section 3.5.4, we find the following characteristic functional for the filtered process $U^* w = h^\vee * w$:

$$c_{U^* w}(\varphi) = e^{-\frac{1}{2} \| h * \varphi \|_2^2}.$$

In turn, it yields the following mean and correlation functions

$$m_{h^\vee * w}(r) = 0,$$

$$R_{h^\vee * w}(r, s) = (h * h^\vee)(r - s),$$

as expected.
3.6 Bibliographical pointers and historical notes

Sections 3.1 and 3.2

Recommended references on functional analysis, topological vector spaces, and duality are the books by Schaefer [Sch99] and Rudin [Rud73].

Much of the theory of nuclear spaces was developed by Grothendieck [Gro55] in his thesis work under the direction of Schwartz. For detailed information, we refer to Pietsch [Pie72].

Section 3.3

For a comprehensive treatment of generalized functions, we recommend the books of Gelfand and Shilov [GS64] and Schwartz [Sch66] (the former being more accessible while maintaining rigor). The results on Fourier multipliers are covered by Hörmander [Hör80] and Mikhlin et al. [MP86].

A historical precursor to the theory of generalized functions is the "operational method" of Heaviside, appearing in his collected works in the last decade of the 19th century [Hea71]. The introduction of the Lebesgue integral was a major step that gave a precise meaning to the concept of the almost-everywhere equivalence of functions. Dirac introduced his eponymous distribution as a convenient notation in the 1920s. Sobolev [Sob36] developed a theory of generalized functions in order to define weak solutions of partial differential equations. But it was Laurent Schwartz [Sch66] who put forth the formal and comprehensive theory of generalized functions (distributions) as we use it today (first edition published in 1950). His work was further developed and exposed by the Russian school of Gelfand et al.

Section 3.4

Kolmogorov is the founding father of the modern axiomatic theory of probability which is based on measure theory. We still recommend his original book [Kol56] as the main reference for the material presented here. Newer and more advanced results can be found in the encyclopedic works of Bogachev [Bog07] and Fremlin [Fre03, Fre04a, Fre04b, Fre06, Fre08] on measure theory.

Paul Lévy defined the characteristic function in the early 1920s and is responsible for turning the Fourier-Stieltjes apparatus into one of the most useful tools of probability theory [Lév25, Tay75]. The foundation of the finite-dimensional Fourier approach is Bochner's theorem, which appeared in 1932 [Boc32].

Interestingly, it was Kolmogorov himself who introduced the characteristic functional in 1935 as an equivalent (infinite-dimensional) Fourier-based description of a measure on a Banach space [Kol35]. This tool then laid dormant for many years. The theoretical breakthrough came when Minlos proved the equivalence between this functional and the characterization of probability measures on duals of nuclear spaces (Theorem 3.9)—as hypothesized by Gelfand [Min63, Kol59]. This powerful framework now constitutes the infinite-dimensional counterpart of the traditional Fourier approach to probability theory.

What is lesser known is that Laurent Schwartz, who also happened to be Paul
Lévy’s son in law, revisited the theory of probability measures on infinite-dimensional topological vector spaces, including developments from the French school, in the final years of his career [Sch73b, Sch81b]. These later works are highly abstract, as one may expect from their author. This makes for an interesting contrast with Paul Lévy who had a limited interest in axioms and whose research was primarily guided by an extraordinary intuition.

Section 3.5

The concept of generalized stochastic processes, including the characterization of continuously-defined white noises, was introduced by Gelfand in 1955 [Gel55]. Itô contributed to the topic by formulating the correlation theory of such processes [Itô54]; see also, [Itô84]. The basic reference for the material presented here is [GV64, Chapter 3].

The first applications of the characteristic functional to the study of stochastic processes have been traced back to 1947; they are due to Le Cam [LC47] and Bochner [Boc47], who both appear to have (re)discovered the tool independently. Le Cam was concerned with the practical problem of modeling the relation between rainfall and riverflow, while Bochner was aiming at a fundamental characterization of stochastic processes. Another early promoter is Bartlett who, in collaboration with Kendall, determined the characteristic functional of several Poisson-type processes that are relevant to biology and physics [BK51]. The framework was consolidated by Gelfand [Gel55] and Minlos in the 1960s. They provided the extension to generalized functions and also addressed the fundamental issue of the uniqueness and consistency of this infinite-dimensional description.