An introduction to sparse stochastic processes

Michael Unser and Pouya Tafti
Summary

Sparse stochastic processes are continuous-domain processes that admit a parsimonious representation in some matched wavelet-like basis. Such models are relevant for image compression, compressed sensing, and, more generally, for the derivation of statistical algorithms for solving ill-posed inverse problems.

This book introduces an extended family of sparse processes that are specified by a generic (non-Gaussian) innovation model or, equivalently, as solutions of linear stochastic differential equations driven by white Lévy noise. It presents the mathematical tools for their characterization. The two leading threads that underly the exposition are

- the statistical property of infinite divisibility, which induces two distinct types of behavior—Gaussian vs. sparse—at the exclusion of any other;
- the structural link between linear stochastic processes and spline functions which is exploited to simplify the mathematics.

The last chapter is devoted to the use of these models for the derivation of algorithms that recover sparse signals. This leads to a Bayesian reinterpretation of popular sparsity-promoting processing schemes—such as total-variation denoising, LASSO, and wavelet shrinkage—as MAP estimators for specific types of Lévy processes.

The book, which is mostly self-contained, is targeted to an audience of graduate students and researchers with an interest in signal/image processing, compressed sensing, approximation theory, machine learning, or statistics.
Chapter 1

Introduction

1.1 Sparsity: Occam's razor of modern signal processing?

The hypotheses of Gaussianity and stationarity play a central role in the standard statistical formulation of signal processing. They fully justify the use of the Fourier transform as the optimal signal representation and naturally lead to the derivation of optimal linear filtering algorithms for a large variety of statistical estimation tasks. This classical view of signal processing is elegant and reassuring, but it not at the forefront of research anymore.

Starting with the discovery of the wavelet transform in the late 80s [Dau88, Mal89], researchers in signal processing have progressively moved away from the Fourier transform and have uncovered powerful alternatives. Consequently, they have ceased modeling signals as Gaussian stationary processes and have adopted a more deterministic, approximation-theoretic point of view. The key developments that are presently reshaping the field, and which are central to the theory presented in this monograph, are summarized below.

- **Novel transforms and dictionaries for the representation of signals**: New redundant and non-redundant representations of signals (wavelets, local cosine, curvelets) have emerged during the past two decades and have led to better algorithms for data compression, data processing, and feature extraction. The most prominent example is the wavelet-based JPEG-2000 standard for image compression [CSE00], which outperforms the widely-used JPEG method based on the DCT (discrete cosine transform). Another illustration is wavelet-domain image denoising which provides a good alternative to more traditional linear filtering [Don95]. The various dictionaries of basis functions that have been proposed so far are tailored to specific types of signals; there does not appear to be one that fits all.

- **Sparsity as a new paradigm for signal processing**: At the origin of this new trend is the key observation that many naturally-occurring signals and images—in particular, the ones that are piecewise-smooth—can be accurately reconstructed from a “sparse” wavelet expansion that involves much fewer terms than the original number of samples [Mal98]. The concept of sparsity has been systematized and extended to other transforms, including redundant representations (a.k.a. frames); it is at the heart of recent developments in signal processing. Sparse signals are easy to compress and to denoise by simple pointwise processing (e.g., shrinkage) in the transformed domain. Sparsity provides an equally-powerful framework for dealing with more difficult, ill-posed signal-reconstruction problems [CW08, BDE09]. Promoting sparse solutions in linear models is also of interest in statistics: a popular regression shrinkage estimator is LASSO, which imposes an upper bound on the $\ell_1$-norm of the model coefficients [Tib96].
1. INTRODUCTION

- **New sampling strategies with fewer measurements**: The theory of compressed sensing deals with the problem of the reconstruction of a signal from a minimal, but suitably chosen, set of measurements [Don06, CW08, BDE09]. The strategy there is as follows: among the multitude of solutions that are consistent with the measurements, one should favor the “sparsest” one. In practice, one replaces the underlying $\ell_0$-norm minimization problem, which is NP hard, by a convex $\ell_1$-norm minimization which is computationally much more tractable. Remarkably, researchers have shown that this simplification does yield the correct solution under suitable conditions (e.g., restricted isometry) [CW08]. Similarly, it has been demonstrated that signals with a finite rate of innovation (the prototypical example being a stream of Dirac impulses with unknown locations and amplitudes) can be recovered from a set of uniform measurements at twice the “innovation rate” [VMB02], rather than twice the bandwidth, as would otherwise be dictated by Shannon’s classical sampling theorem.

- **Superiority of nonlinear signal-reconstruction algorithms**: There is increasing empirical evidence that nonlinear variational methods (non-quadratic or sparsity-driven regularization) outperform the classical (linear) algorithms (direct or iterative) that are being used routinely for solving bioimaging reconstruction problems [CBFAB97, FN03]. So far, this has also been demonstrated for the problem of image deconvolution and for the reconstruction of non-Cartesian MRI [LDP07]. The considerable research effort in this area has also resulted in the development of novel algorithms (ISTA, FISTA) for solving convex optimization problems that were previously considered out of numerical reach [FN03, DDDM04, BT09].

1.2 Sparse stochastic models: The next step beyond Gaussianity

While the recent developments listed above are truly remarkable and have resulted in significant algorithmic advances, the overall picture and understanding is still far from being complete. One limiting factor is that the current formulations of compressed sensing and sparse-signal recovery are fundamentally deterministic. By drawing on the analogy with the classical linear theory of signal processing, where there is an equivalence between quadratic energy-minimization techniques and minimum-mean-square-error (MMSE) estimation under the Gaussian hypothesis, there are good chances that further progress is achievable by adopting a complementary statistical-modeling point of view\(^1\). The cru-

---

\(^1\) It is instructive to recall the fundamental role of statistical modeling in the development of traditional signal processing. The standard tools of the trade are the Fourier transform, Shannon-type sampling, linear filtering, and quadratic energy-minimization techniques. These methods are widely used in practice: They are powerful, easy to deploy, and mathematically convenient. The important conceptual point is that they are justifiable based on the theory of Gaussian stationary processes (GSP). Specifically, one can invoke the following optimality results:

- **The Fourier transform as well as several of its real-valued variants (e.g., DCT) are asymptotically equivalent to the Karhunen-Loève transform (KLT) for the whole class of GSP. This supports the use of sinusoidal transforms for data compression, data processing, and feature extraction. The underlying notion of optimality here is energy compaction, which implies decorrelation. Note that the decorrelation is equivalent to independence in the Gaussian case only.**

- **Optimal filters**: Given a series of linear measurements of a signal corrupted by noise, one can readily specify its optimal reconstruction (LMSE estimator) under the general Gaussian hypothesis. The corresponding algorithm (Wiener filter) is linear and entirely determined by the covariance structure of the signal and noise. There is also a direct connection with variational reconstruction techniques since the Wiener solution can also be formulated as a quadratic energy-minimization problem (Gaussian MAP estimator).

- **Optimal sampling/interpolation strategies**: While this part of the story is less known, one can also invoke estimation-theoretic arguments to justify a Shannon-type, constant-rate sampling, which ensures a minimum loss of information for a large class of predominantly-lowpass GSP [PM62, Uns93]. This is not totally surprising since the basis functions of the KLT are inherently bandlimited. One can also derive minimum mean-square-error interpolators for GSP in general. The optimal signal-reconstruction algorithm takes the form of a hybrid Wiener filter whose input is discrete (signal samples) and whose output is a continuously-
cial ingredient that is required to guide such an investigation is a sparse counterpart to the classical family of Gaussian stationary processes (GSP). This monograph focuses on the formulation of such a statistical framework, which may be aptly qualified as the next step after Gaussianity under the functional constraint of linearity.

In light of the elements presented in the introduction, the basic requirements for a comprehensive theory of sparse stochastic processes are as follows:

- **Backward compatibility**: There is a large body of literature and methods based on the modeling of signals as realizations of GSP. We would like the corresponding identification, linear filtering, and reconstruction algorithms to remain applicable, even though they obviously become suboptimal when the Gaussian hypothesis is violated. This calls for an extended formulation that provides the same control of the correlation structure of the signals (second-order moments, Fourier spectrum) as the classical theory does.

- **Continuous-domain formulation**: The proper interpretation of qualifying terms such as “piecewise-smooth”, “translation-invariant”, “scale-invariant”, “rotation-invariant” calls for continuous-domain models of signals that are compatible with the conventional (finite-dimensional) notion of sparsity. Likewise, if we intend to optimize or possibly redesign the signal-acquisition system as in generalized sampling and compressed sensing, the very least is to have a model that characterizes the information content prior to sampling.

- **Predictive power**: Among other things, the theory should be able to explain why wavelet representations can outperform the older Fourier-related types of decompositions, including the KLT, which is optimal from the classical perspective of variance concentration.

- **Ease of use**: To have practical usefulness, the framework should allow for the derivation of the (joint) probability distributions of the signal in any transformed domain. This calls for a linear formulation with the caveat that it needs to accommodate non-Gaussian distributions. In that respect, the best thing beyond Gaussianity is infinite divisibility, which is a general property of random variables that is preserved under arbitrary linear combinations.

- **Stochastic justification and refinement of current reconstruction algorithms**: A convincing argument for adopting a new theory is that it must be compatible with the state of the art, while it also ought to suggest new directions of research. In the present context, it is important to be able to establish the connection with deterministic recovery techniques such as $\ell_1$-norm minimization.

The good news is that the foundations for such a theory exist and can be traced back to the pioneering work of Paul Lévy, who defined a broad family of “additive” stochastic processes, now called Lévy processes. Brownian motion (a.k.a. the Wiener process) is the only Gaussian member of this family, and, as we shall demonstrate, the only representative that does not exhibit any degree of sparsity. The theory that is developed in this monograph constitutes the full linear, multidimensional extension of those ideas where the essence of Paul Lévy’s construction is embodied in the definition of Lévy innovations (or white Lévy noise), which can be interpreted as the derivative of a Lévy process in the sense of distributions (a.k.a. generalized functions). The Lévy innovations are then linearly transformed to generate a whole variety of processes whose spectral characteristics are controlled by a linear mixing operator, while their sparsity is governed by the innovations. The latter can also be viewed as the driving term of some corresponding linear stochastic differential equation (SDE). Another way of describing the extent of this generalization is to consider the representation of a general continuous-domain Gaussian

defined signal that can be represented in terms of generalized B-spline basis functions [UB05b].
1. Introduction

process by a stochastic Wiener integral:

\[ s(x) = \int_{\mathbb{R}} h(x, x') \, dW(x') \]  

(1.1)

where \( h(x, x') \) is the kernel—that is, the infinite-dimensional analog of the matrix representation of a transformation in \( \mathbb{R}^n \)—of a general, \( L_2 \)-stable linear operator. \( dW \) is the so-called random Wiener measure which is such that \( W(x) = \int_{-\infty}^{x} dW(x') \)

is the Wiener process, where the latter equation constitutes a special case of (1.1) with \( h(x, x') = \mathbb{1}_{\{x > x' \geq 0\}} \). There, \( \mathbb{1}_{\Omega} \) denotes the indicator function of the set \( \Omega \). If \( h(x, x') = h(x - x') \) is a convolution kernel, then (1.1) defines the whole class of Gaussian stationary processes. The essence of the present formulation is to replace the Wiener measure by a more general non-Gaussian, multidimensional Lévy measure. The catch, however, is that we shall not work with measures but rather with generalized functions and generalized stochastic processes. These are easier to manipulate in the Fourier domain and better suited for specifying general linear transformations. In other words, we shall rewrite (1.1) as

\[ s(x) = \int_{\mathbb{R}} h(x, x') w(x') \, dx' \]  

(1.2)

where the entity \( w \) (the continuous-domain innovation) needs to be given a proper mathematical interpretation. The main advantage of working with innovations is that they provide a very direct link with the theory of linear systems, which allows for the use of standard engineering notions such as the impulse and frequency responses of a system.

1.3 From splines to stochastic processes, or when Schoenberg meets Lévy

We shall start our journey by making an interesting connection between splines, which are deterministic objects with some inherent sparsity, and Lévy processes with a special focus on the compound Poisson process, which constitutes the archetype of a sparse stochastic process. The key observation is that both categories of signals—namely, deterministic and random—are ruled by the same differential equation. They can be generated via the proper integration of an “innovation” signal that carries all the necessary information. The fun is that the underlying differential system is only marginally stable, which requires the design of a special anti-derivative operator. We then use the close relationship between splines and wavelets to gain insight on the ability of wavelets to provide sparse representations of such signals. Specifically, we shall see that most non-Gaussian Lévy processes admit a better \( M \)-term representation in the Haar wavelet basis than in the classical Karhunen-Loève transform (KLT) which is usually believed to be optimal for data compression. The explanation for this counter-intuitive result is that we are breaking some of the assumptions that are implicit in the proof of optimality of the KLT.

1.3.1 Splines and Legos revisited

Splines constitute a general framework for converting series of data points (or samples) into continuously-defined signals or functions. By extension, they also provide a powerful mechanism for translating tasks that are specified in the continuous domain into efficient numerical algorithms (discretization).
1.3. From splines to stochastic processes, or when Schoenberg meets Lévy

![Figure 1.1: Examples of spline signals. (a) Cardinal spline interpolant of degree 0 (piecewise-constant). (b) Cardinal spline interpolant of degree 1 (piecewise-linear). (c) Nonuniform D-spline or compound Poisson process, depending on the interpretation (deterministic vs. stochastic).](image)

The cardinal setting corresponds to the configuration where the sampling grid is on the integers. Given a sequence of sample values \( f[k], k \in \mathbb{Z} \), the basic cardinal interpolation problem is to construct a continuously-defined signal \( f(x), x \in \mathbb{R} \) that satisfies the interpolation condition \( f(x) \big|_{x=k} = f[k] \), for all \( k \in \mathbb{Z} \). Since the general problem is obviously ill-posed, the solution is constrained to live in a suitable reconstruction subspace (e.g., a particular space of cardinal splines) whose degrees of freedom are in one-to-one correspondence with the data points. The most basic concretization of those ideas is the construction of the piecewise-constant interpolant

\[
f_1(x) = \sum_{k \in \mathbb{Z}} f[k] \beta_0^0(x-k) \tag{1.3}
\]

which involves rectangular basis functions (informally described as Legos) that are shifted replicates of the causal \( ^0 \text{B-spline} \) of degree zero

\[
\beta_0^0(x) = \begin{cases} 
1, & \text{for } 0 \leq x < 1 \\
0, & \text{otherwise.} 
\end{cases} \tag{1.4}
\]

Observe that the basis functions \( \{\beta_0^0(x-k)\}_{k \in \mathbb{Z}} \) are non-overlapping, orthonormal, and that their linear span defines the space of cardinal polynomial splines of degree 0. Moreover, since \( \beta_0^0(x) \) takes the value one at the origin and vanishes at all other integers, the expansion coefficients in (1.3) coincide with the original samples of the signal. Equation (1.3) is nothing but a mathematical representation of the sample-and-hold method of interpolation which yields the type of "Lego-like" signal shown in Fig. 1.1a.

A defining property of piecewise-constant signals is that they exhibit "sparse" first-order derivatives that are zero almost everywhere, except at the points of transition where differentiation is only meaningful in the sense of distributions. In the case of the cardinal

---

2. A function \( f_+ (x) \) is said to be causal if \( f_+ (x) = 0 \), for all \( x < 0 \).
spline specified by (1.3), we have that

\[ Df_1(x) = \sum_{k \in \mathbb{Z}} a_1[k] \delta(x - k) \quad (1.5) \]

where the weights of the integer-shifted Dirac impulses \( \delta(-k) \) are given by the corresponding jump size of the function: \( a_1[k] = f[k] - f[k-1] \). The main point is that the application of the operator \( D = \frac{d}{dx} \) uncovers the spline discontinuities (a.k.a. knots) which are located on the integer grid: Its effect is that of a mathematical A-to-D conversion since the r.h.s. of (1.5) corresponds to the continuous-domain representation of a discrete signal commonly used in the theory of linear systems. In the nomenclature of splines, we say that \( f_1(x) \) is a cardinal \( D \)-spline\(^3\), which is a special case of a general non-uniform \( D \)-spline where the knots can be located arbitrarily (cf. Fig. 1.1c).

The next fundamental observation is that the expansion coefficients in (1.5) are obtained via a finite-difference scheme which is the discrete counterpart of differentiation. To get some further insight, we define the finite-difference operator

\[ D_d f(x) = f(x) - f(x - 1). \]

The latter turns out to be a smoothed version of the derivative

\[ D_d f(x) = (\beta^0_+ \ast Df)(x), \]

where the smoothing kernel is precisely the B-spline generator for the expansion (1.3). An equivalent manifestation of this property can be found in the relation

\[ \beta^0_+ (x) = D_d D^{-1} \delta(x) = D_d 1_+(x) \quad (1.6) \]

where the unit step \( 1_+(x) = \frac{x}{0, +\infty} \) (a.k.a. the Heaviside function) is the causal Green function \(^4\) of the derivative operator. This formula is illustrated in Fig. 1.2a. Its Fourier-domain counterpart is

\[ \hat{\beta}^0_+ (\omega) = \int_{\mathbb{R}} \beta^0_+(x) e^{-j\omega x} \, dx = \frac{1 - e^{-j\omega}}{j\omega} \quad (1.7) \]

which is recognized as being the ratio of the frequency responses of the operators \( D_d \) and \( D \), respectively.

---

\(^3\) Other brands of splines are defined in the same fashion by replacing the derivative \( D \) by some other differential operator generically denoted by \( L \).

\(^4\) We say that \( \rho(x) \) is the causal Green function of the shift-invariant operator \( L \) if \( \rho \) is causal and satisfies \( L \rho = \delta \). This can also be written as \( L^{-1}\delta = \rho \), meaning that \( \rho \) is the causal impulse response of the shift-invariant inverse operator \( L^{-1} \).
Thus, the basic Lego component, $\beta_0^0$, is much more than a mere building block: it is also a kernel that characterizes the approximation that is made when replacing a continuous-domain derivative by its discrete counterpart. This idea (and its generalization for other operators) will prove to be one of the key ingredient in our formulation of sparse stochastic processes.

### 1.3.2 Higher-degree polynomial splines

A slightly more sophisticated model is to select a piecewise-linear reconstruction which admits the similar B-spline expansion

$$f_2(x) = \sum_{k \in \mathbb{Z}} f[k+1] \beta_1^1(x-k) \quad (1.8)$$

where

$$\beta_1^1(x) = (\beta_0^0 \ast \beta_0^0)(x) = \begin{cases} x, & \text{for } 0 \leq x < 1 \\ 2-x, & \text{for } 1 \leq x < 2 \\ 0, & \text{otherwise} \end{cases} \quad (1.9)$$

is the causal B-spline of degree 1, a triangular function centered at $x = 1$. Note that the use of a causal generator is compensated by the unit shifting of the coefficients in (1.8), which is equivalent to re-centering the basis functions on the sampling locations. The main advantage of $f_2$ in (1.8) over $f_1$ in (1.3) is that the underlying function is now continuous, as illustrated in Fig. 1b.

In an analogous manner, one can construct higher-degree spline interpolants that are piecewise-polynomials of degree $n$ by considering B-splines atoms of degree $n$ obtained from the $(n+1)$-fold convolution of $\beta_0^0(x)$ (cf. Fig. 1.2b). The generic version of such a higher-order spline model is

$$f_{n+1}(x) = \sum_{k \in \mathbb{Z}} c[k] \beta_n^1(x-k) \quad (1.10)$$

with

$$\beta_n^1(x) = (\beta_0^0 \ast \beta_0^0 \cdots \ast \beta_0^0)_{n+1}(x).$$

The catch though is that, for $n > 1$, the expansion coefficients $c[k]$ in (1.10) are not identical to the sample values $f[k]$ anymore. Yet, they are in a one-to-one correspondence with them and can be determined efficiently by solving a linear system of equations that has a convenient band-diagonal Toeplitz structure [Uns99].

The higher-order counterparts of relations (1.7) and (1.6) are

$$\hat{\beta}_n^1(\omega) = \left(1 - \frac{1}{j\omega}\right)^{n+1}$$

and

$$\beta_n^1(x) = D_{d+1}^{-n-1} D_n(x)^n \quad (1.11)$$

where

$$D_n(x)^n = \frac{D_{d+1}(x)^n}{n!} = \sum_{m=0}^{n+1} (-1)^m \binom{n+1}{m} \frac{(x-k)^n}{n!}.$$
1. Introduction

with \( (x)_+ = \max(0, x) \). The latter explicit time-domain formula follows from the fact that the impulse response of the \((n+1)\)-fold integrator (or, equivalently, the causal Green function of \(D^{n+1}\)) is the one-sided power function \(D^{-(n+1)}\delta(x) = \frac{x^+}{n!} \). This elegant formula is due to Schoenberg, the father of splines [Sch46]. He also proved that the polynomial B-spline of degree \(n\) is the shortest cardinal \(D^{n+1}\)-spline and that its integer translates form a Riesz basis of such polynomial splines. In particular, he showed that the B-spline representation (1.10) is unique and stable, in the sense that

\[
\|f_n\|_{l_2}^2 = \int_R |f_n(x)|^2 \, dx \leq \|c\|_{l_2}^2 = \sum_{k \in \mathbb{Z}} |c[k]|^2.
\]

Note that the inequality above becomes an equality for \(n = 0\) since the squared \(L_2\)-norm of the corresponding piecewise-constant function is easily converted into a sum. This also follows from Parseval’s identity because the B-spline basis \(\{\beta_n^k\}\) is orthonormal.

One last feature is that polynomial splines of degree \(n\) are inherently smooth, in the sense that they are \(n\)-times differentiable everywhere with bounded derivatives—that is, Hölder continuous of order \(n\). In the cardinal setting, this follows from the property that

\[
D^n \beta_n^k(x) = D^n D^{n+1}_d D^{-(n+1)} \delta(x) = D^n D_d D^{-1} \delta(x) = D^n \beta_0^0(x),
\]

which indicates that the \(n\)th-order derivative of a B-spline of degree \(n\) is piecewise-constant and bounded.

1.3.3 Random splines, innovations, and Lévy processes

To make the link with Lévy processes, we now express the random counterpart of (1.5) as

\[
Ds(x) = \sum_n a_n \delta(x - x_n) = w(x)
\]

(1.12)

where the locations \(x_n\) of the Dirac impulses are uniformly distributed over the real line (Poisson distribution with rate parameter \(\lambda\)) and the weights \(a_n\) are i.i.d. with amplitude distribution \(p_\lambda(a)\). For simplicity, we are also assuming that \(p_\lambda\) is symmetric with finite variance \(\sigma_\lambda^2 = \int_\mathbb{R} a^2 p_\lambda(a) \, da\). We shall refer to \(w\) as the innovation of the signal \(s\) since it contains all the parameters that are necessary for its description. Clearly, \(s\) is a signal with a finite rate of innovation, a term that was coined by Vetterli et al.[VMB02].

The idea now is to reconstruct \(s\) from its innovation \(w\) by integrating (1.12). This requires the specification of some boundary condition to fix the integration constant. Since the constraint in the definition of Lévy processes is \(s(0) = 0\) (with probability one), we first need to find a suitable antiderivative operator, which we shall denote by \(D_0^{-1}\). In the event when the input function is Lebesgue integrable, the relevant operator is readily specified as

\[
D_0^{-1} \varphi(x) = \int_{-\infty}^{x} \varphi(t) \, dt - \int_{-\infty}^{0} \varphi(t) \, dt = \begin{cases} 
\int_{0}^{x} \varphi(t) \, dt, & \text{for } x \geq 0 \\
- \int_{x}^{0} \varphi(t) \, dt, & \text{for } x < 0
\end{cases}
\]

It is the corrected version (subtraction of the proper signal-dependent constant) of the conventional shift-invariant integrator \(D^{-1}\) for which the integral runs from \(-\infty\) to \(x\). The Fourier counterpart of this definition is

\[
D_0^{-1} \varphi(x) = \int_{\mathbb{R}} \frac{\phi^{(x)} - 1}{j\omega} \varphi(\omega) \, d\omega / 2\pi
\]
1.3. From splines to stochastic processes, or when Schoenberg meets Lévy

which can be extended, by duality, to a much larger class of generalized functions (cf. Chapter 5). This is feasible because the latter expression is a regularized version of an integral that would otherwise be singular, since the division by \( j\omega \) is tempered by a proper correction in the numerator: \( e^{j\omega x} - 1 = j\omega + O(\omega^2) \). It is important to note that \( D_0^{-1} \) is scale-invariant (in the sense that it commutes with scaling), but not shift-invariant, unlike \( D^{-1} \). Our reason for selecting \( D_0^{-1} \) over \( D^{-1} \) is actually more fundamental than just imposing the “right” boundary conditions. It is guided by stability considerations: \( D_0^{-1} \) is a valid right inverse of \( D \) in the sense that \( DD_0^{-1} = \text{Id} \) over a large class of generalized functions, while the use of the shift-invariant inverse \( D^{-1} \) is much more constrained. Other than that, both operators share most of their global properties. In particular, since the finite-difference operator has the convenient property of annihilating the constants that are in the null space of \( D \), we see that

\[
\beta_0^s(x) = D_0D_0^{-1} \delta(x) = D_0D^{-1} \delta(x). \tag{1.13}
\]

Having the proper inverse operator at our disposal, we can apply it to formally solve the stochastic differential equation (1.12). This yields the explicit representation of the sparse stochastic process:

\[
s(x) = D_0^{-1} w(x) = \sum_n d_n D_0^{-1} \delta(x - x_n)(x) \\
= \sum_n d_n \{ 1_+ (x - x_n) - 1_+ (-x_n) \} \tag{1.14}
\]

where the second term \( 1_+ (-x_n) \) in the last parenthesis ensures that \( s(0) = 0 \). Clearly, the signal defined by (1.14) is piecewise-constant (random spline of degree 0) and its construction is compatible with the classical definition of a compound Poisson process, which is a special type of Lévy process. A representative example is shown in Fig. 1.1c.

It can be shown that the innovation \( w \) specified by (1.12), made of random impulses, is a special type of continuous-domain white noise with the property that

\[
\mathbb{E}[w(x)w(x')] = R_w(x-x') = \sigma_0^2 \delta(x-x') \tag{1.15}
\]

where \( \sigma_0^2 = \lambda \sigma_n^2 \) is the product of the Poisson rate parameter \( \lambda \) and the variance \( \sigma_n^2 \) of the amplitude distribution. More generally, we can determine the correlation form of the innovation, which is given by

\[
\mathbb{E}\langle \varphi_1, w \rangle \langle \varphi_2, w \rangle = \sigma_0^2 \langle \varphi_1, \varphi_2 \rangle \tag{1.16}
\]

for any real-valued functions \( \varphi_1, \varphi_2 \in L_2(\mathbb{R}) \) and \( \langle \varphi_1, \varphi_2 \rangle = \int_\mathbb{R} \varphi_1(x)\varphi_2(x) \, dx \).

This suggests that we can apply the same operator-based synthesis to other types of continuous-domain white noise, as illustrated in Fig. 1.3. In doing so, we are able to generate the whole family of Lévy processes. In the case where \( w \) is a white Gaussian noise, the resulting signal is a Brownian motion which has the property of being continuous almost everywhere. A more extreme case arises when \( w \) is an alpha-stable noise which yields a stable Lévy process whose sample path has a few really large jumps and is rougher than a Brownian motion.

In the classical literature on stochastic processes, Lévy processes are usually defined in terms of their increments, which are i.i.d. and infinitely-divisible random variables (cf. Chapter 7). Here, we shall consider the so-called increment process \( u(x) = s(x) - s(x - 1) \), which has a number of remarkable properties. The key observation is that \( u \), in its continuous-domain version, is the convolution of a white noise (innovation) with the B-spline kernel \( \beta_0^s \). Indeed, the relation (1.13) leads to

\[
u(x) = D_0 s(x) = D_0D_0^{-1} w(x) = (\beta_0^s * w)(x). \tag{1.17}
\]
Figure 1.3: Synthesis of different brands of Lévy processes by integration of a corresponding continuous-domain white noise. The alpha-stable excitation in the bottom example is such that the increments of the Lévy process have a symmetric Cauchy distribution.

This implies, among other things, that \( u \) is stationary, while the original Lévy process \( s \) is not (since \( D_{\beta_1}^{-1} \) is not shift-invariant). It also suggests that the samples of the increment process \( u \) are independent if they are taken at a distance of 1 or more apart, the limit corresponding to the support of the rectangular convolution kernel \( \beta_1 \). When the autocorrelation function \( R_w(y) \) of the driving noise is well-defined and given by (1.15), we can easily determine the autocorrelation of \( u \) as

\[
R_u(y) = E\{u(x)u(x+y)\} = (\beta_1^0 \ast (\beta_1^0)\upsim) \ast R_w(y) = \sigma_w^2 \beta_1^1 (y-1)
\]

where \( (\beta_1^0)\upsim (x) = \beta_1^0 (-x) \). It is proportional to the autocorrelation of a rectangle, which is a triangular function (centered B-spline of degree 1).

Of special interest to us are the samples of \( u \) on the integer grid, which are characterized for \( k \in \mathbb{Z} \) as

\[
u[k] = s(k) - s(k-1) = \langle u, \beta_1^k \rangle (\cdot - k).
\]

The r.h.s. relation can be used to show that the \( u[k] \) are i.i.d. because \( w \) is white, stationary, and the supports of the analysis functions \( \beta_1^k \) are non-overlapping. We shall refer to \( \{u[k]\}_{k \in \mathbb{Z}} \) as the discrete innovation of \( s \). Its determination involves the sampling of \( s \) at the integers and a discrete differentiation (finite differences), in direct analogy with the generation of the continuous-domain innovation \( w(x) = Ds(x) \).

The discrete innovation sequence \( u[\cdot] \) will play a fundamental role in signal processing because it constitutes a convenient tool for extracting the statistics and characterizing the samples of a stochastic process. It is probably the best practical way of presenting the information because

1. we never have access to the full signal \( s(x) \), which is a continuously-defined entity, and
2. we cannot implement the whitening operator (derivative) exactly, not to mention that the continuous-domain innovation \( w(x) \) does not admit an interpretation as an ordinary function of \( x \). For instance, Brownian motion is not differentiable anywhere in the classical sense.

This points to the fact that the continuous-domain innovation model is a theoretical construct. Its primary purpose is to facilitate the determination of the joint probability distributions of any series of linear measurements of a wide class of sparse stochastic pro-
1.3. From splines to stochastic processes, or when Schoenberg meets Lévy

Figure 1.4: Dual pair of multiresolution bases where the first kind of functions (wavelets) are the derivatives of the second (hierarchical basis functions): (a) (unnormalized) Haar wavelet basis. (b) Faber-Schauder basis (a.k.a. Franklin system).

cesses, including the discrete version of the innovation which has the property of being maximally decoupled.

1.3.4 Wavelet analysis of Lévy processes and $M$-term approximations

Our purpose so far has been to link splines and Lévy processes to the derivative operator $D$. We shall now exploit this connection in the context of wavelet analysis. To that end, we consider the Haar basis \( \{ \psi_{i,k} \}_{i \in \mathbb{Z}, k \in \mathbb{Z}} \), which is generated by the Haar wavelet

\[
\psi_{\text{Haar}}(x) = \begin{cases} 
1, & \text{for } 0 < x < \frac{1}{2} \\
-1, & \text{for } \frac{1}{2} \leq x < 1 \\
0, & \text{otherwise.}
\end{cases} \tag{1.19}
\]

The basis functions, which are orthonormal, are given by

\[
\psi_{i,k}(x) = 2^{-i/2} \psi_{\text{Haar}} \left( \frac{x - 2^i k}{2^i} \right) \tag{1.20}
\]

where $i$ and $k$ are the scale (dilation of $\psi_{\text{Haar}}$ by $2^i$) and location (translation of $\psi_{i,0}$ by $2^i k$) indices, respectively. A closely related system is the Faber-Schauder basis $\{ \phi_{i,k} \}_{i \in \mathbb{Z}, k \in \mathbb{Z}}$, which is made up of B-splines of degree 1 in a wavelet-like configuration (cf. Fig. 1.4).

Specifically, the hierarchical triangle basis functions are given by

\[
\phi_{i,k}(x) = \beta_i^1 \left( \frac{x - 2^i k}{2^i-1} \right). \tag{1.21}
\]

While these functions are orthogonal within any given scale (because they are non-overlapping), they fail to be so across scales. Yet, they form a Schauder basis, which is a somewhat weaker property than being a Riesz basis of $L_2(\mathbb{R})$.

The fundamental observation for our purpose is that the Haar system can be obtained by differentiating the Faber-Schauder one, up to some amplitude factor. Specifically, we
have the relations
\[ \psi_{i,k} = 2^{j/2-1} D \phi_{i,k} \]
(1.22)
\[ D_0^{-1} \psi_{i,k} = 2^{j/2-1} \phi_{i,k}. \]
(1.23)

Let us now apply (1.22) to the formal determination of the wavelet coefficients of the Lévy process \( s = D_0^{-1} w \). The crucial manipulation, which will be justified rigorously within the framework of generalized stochastic processes (cf. Chapter 3), is \( \langle s, D \phi_{i,k} \rangle = \langle D^* s, \phi_{i,k} \rangle = -\langle w, \phi_{i,k} \rangle \) where we have used the adjoint relation \( D^* = -D \) and the right inverse property of \( D_0^{-1} \). This allows us to express the wavelet coefficients as
\[ W_{i,k} = \langle s, \psi_{i,k} \rangle = -2^{j/2-1} \langle w, \phi_{i,k} \rangle \]
which, up to some scaling factors, amounts to a Faber-Schauder analysis of the innovation \( w = Ds \). Since the triangle functions \( \phi_{i,k} \) are non-overlapping within a given scale and the innovation is independent at every point, we immediately deduce that the corresponding wavelet coefficients are also independent. However, the decoupling is not perfect across scales due to the parent-to-child overlap of the triangle functions. The residual correlation can be determined from the correlation form (1.16) of the noise, according to
\[ E[W_{i,j} W_{i',k}] = 2^{(j+j')/2-2} \mathbb{E}\{\langle w, \phi_{i,k} \rangle \langle w, \phi_{i',k} \rangle\} \propto \langle \phi_{i,k}, \phi_{i',k} \rangle. \]
Since the triangle functions are non-negative, the residual correlation is zero iff. \( \phi_{i,k} \) and \( \phi_{i',k} \) are non-overlapping, in which case the wavelet coefficients are independent as well. We can also predict that the wavelet transform of a compound Poisson process will be sparse (i.e., with many vanishing coefficients) because the random Dirac impulses of the innovation will intersect only few Faber-Schauder functions, an effect that becomes more and more pronounced as the scale gets finer. The level of sparsity can therefore be expected to be directly dependent upon \( \lambda \) (the density of impulses per unit length).

To quantify this behavior, we applied Haar wavelets to the compression of sampled realizations of Lévy processes and compared the results with those of the “optimal” textbook solution for transform coding. In the case of a Lévy process with finite variance, the Karhunen-Loève transform (KLT) can be determined analytically from the knowledge of the covariance function \( E[s(x)s(y)] = C(|x| + |y| - |x - y|) \) where \( C \) is an appropriate constant. The KLT is also known to converge to the discrete cosine transform (DCT) as the size of the signal increases. The present compression task is to reconstruct a series of 4096-point signals from their \( M \) largest transform coefficients, which is the minimum-error selection rule dictated by Parseval’s relation. Fig. 1.5 displays the graph of the relative quadratic \( M \)-term approximation errors for the three types of Lévy processes shown in Fig. 1.3. We also considered the identity transform as baseline, and the DCT as well, whose results were found to be indistinguishable from those of the KLT. We observe that the KLT performs best in the Gaussian scenario, as expected. It is also slightly better than wavelets at large compression ratios for the compound Poisson process (piecewise-constant signal with Gaussian-distributed jumps). In the latter case, however, the situation changes dramatically as \( M \) increases since one is able to reconstruct the signal perfectly from a fraction of the wavelet coefficients, in reason of the sparse behavior explained above. The advantage of wavelets over the KLT/DCT is striking for the Lévy flight (StS) distribution with \( \alpha = 1 \). While these findings are surprising at first, they do not contradict the classical theory which tells us that the KLT has the minimum basis-restriction error for the given class of processes. The twist here is that the selection of the \( M \) largest transform coefficients amounts to some adaptive reordering of the basis functions, which is not accounted for in the derivation of the KLT. The other point is that the KLT solution is not defined for the third type of StS process whose theoretical covariances are unbounded—this does not
1.3. From splines to stochastic processes, or when Schoenberg meets Lévy

Brownian Motion
Compound Poisson
\( P(x=0) = 0 \).

Cauchy-Lévy

Figure 1.5: Haar wavelets vs. KLT: \( M \)-term approximation errors for different brands of Lévy processes. (a) Gaussian (Brownian motion). (b) Compound Poisson with Gaussian jump distribution and \( e^{-\lambda} = 0.9 \). (c) Alpha-stable (symmetric Cauchy). The results are averages over 1000 realizations.

prevent us from applying the Gaussian solution/DCT to a finite-length realization whose \( \ell_2 \)-norm is finite (almost surely). This simple experiment with various stochastic models corroborates the results obtained with image compression where the superiority of wavelets over the DCT (e.g., JPEG2000 vs. JPEG) is well-established.

1.3.5 Lévy’s wavelet-based synthesis of Brownian motion

We close this introductory chapter by making the connection with a multiresolution scheme that Paul Lévy developed in the 1930s to characterize the properties of Brownian motion. To do so, we adopt a point of view that is the dual of the one in Section 1.3.4: it essentially amounts to interchanging the analysis and synthesis functions. As first step, we expand the innovation \( w \) in the orthonormal Haar basis and obtain

\[
 w = \sum_{i \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} Z_{i,k} \psi_{i,k} \text{ with } Z_{i,k} = \langle w, \psi_{i,k} \rangle.
\]

This is acceptable\(^5\) under the finite-variance hypothesis on \( w \). Since the Haar basis is orthogonal, the coefficients \( Z_{i,k} \) in the above expansion are fully decorrelated, but not necessarily independent, unless the white noise is Gaussian or the corresponding basis

\(^5\) The convergence in the sense of distributions is ensured since the wavelet coefficients of a rapidly-decaying test function \( \varphi \) are rapidly-decaying as well.
functions do not overlap. We then construct the Lévy process \( s = D^{-1}_0 w \) by integrating the wavelet expansion of the innovation, which yields

\[
s(x) = \sum_{i \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} Z_{i,k} D^{-1}_0 \psi_{i,k}(x)
= \sum_{i \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} 2^{i/2-1} Z_{i,k} \phi_{i,k}(x).
\quad \text{(1.24)}
\]

The representation (1.24) is of special interest when the noise is Gaussian, in which case the coefficients \( Z_{i,k} \) are i.i.d. and follow a standardized Gaussian distribution. The formula then maps into Lévy’s recursive mid-point method of synthesizing Brownian motion which Yves Meyer singles out as the first use of wavelets to be found in the literature. The Faber-Schauder expansion (1.24) stands out as a localized, practical alternative to Wiener’s original construction of Brownian motion which involves a sum of harmonic cosines (KLT-type expansion).

### 1.4 Historical notes
Chapter 2

Road map to the monograph

The writing of this monograph was motivated by our desire to formalize and extend the ideas presented in Section 1.3 to a class of differential operators much broader than the derivative D. Concretely, this translates into the investigation of the family of stochastic processes specified by the general innovation model that is summarized in Fig. 2.1. The corresponding generator of random signals (upper part of the diagram) has two fundamental components: (1) a continuous-domain noise excitation $w$, which may be thought of as being composed of a continuum of i.i.d. random atoms (innovations), and (2) a deterministic mixing procedure (formally described by $L^{-1}$) which couples the random contributions and imposes the correlation structure of the output. The concise description of the model is $Ls = w$ where $L$ is the whitening operator. The term “innovation” refers to the fact that $w$ represents the unpredictable part of the process. When the inverse operator $L^{-1}$ is linear shift-invariant (LSI), the signal generator reduces to a simple convolutional system which is characterized by its impulse response (or, equivalently, its frequency response). Innovation modeling has a long tradition in statistical communication theory and signal processing; it is the basis for the interpretation of a Gaussian stationary process as a filtered version of a white Gaussian noise [Kai70, Pap91].

In the present context, the underlying objects are continuously-defined. The innovation model then results from defining a stochastic process (or random field when the index variable $r$ is a vector in $\mathbb{R}^d$) as the solution of a stochastic differential equation (SDE) driven by a particular brand of noise. The nonstandard aspect here is that we are considering the innovation model in its greatest generality, allowing for non-Gaussian inputs and differential systems that are not necessarily stable. We shall argue that these extensions are essential for making this type of modeling compatible with the latest developments in signal processing pertaining to the use of wavelets and sparsity-promoting reconstruction algorithms. Specifically, we shall see that it is possible to generate a wide variety of sparse processes by replacing the traditional Gaussian input by some more general brand of (Lévy) noise, within the limits of mathematical admissibility [UTAKed, UTsed]. We shall also demonstrate that such processes admit a sparse representation in a wavelet basis under the assumption that $L$ is scale-invariant. The difficulty there is that scale-invariant SDEs are inherently unstable (due to the presence of poles at the origin); yet, we shall see that they can still result in a proper specification of fractal-type processes, albeit not within the usual framework of stationary processes[TVDVU09, UT11]. The nontrivial aspect of these generalizations is that they necessitate the resolution of instabilities—in the form of singular integrals. This is required not only at the system level, to allow for non-stationary processes, but also at the stochastic level because the most interesting sparsity patterns are associated with unbounded Lévy measures.
2. Road map to the monograph

Figure 2.1: Innovation model of a generalized stochastic process. The process is generated by application of the (linear) inverse operator $L^{-1}$ to a continuous-domain white-noise process $w$. The generation mechanism is general in the sense that it applies to the complete family of Lévy noises, including Gaussian noise as the most basic (non-sparse) excitation. The output process $s$ is stationary iff. $L^{-1}$ is shift-invariant.

Before proceeding with the statistical characterization of sparse stochastic processes, we shall highlight the central role of the operator $L$ and make a connection with spline theory and the construction of signal-adapted wavelet bases.

2.1 On the implications of the innovation model

To motivate our approach, we start with an informal discussion, leaving the technicalities aside. The stochastic process $s$ in Fig. 2.1 is constructed by applying the (integral) operator $L^{-1}$ to some continuous-domain white noise $w$. In most cases of interest, $L^{-1}$ has an infinitely-supported impulse response which introduces long-range dependencies. If we are aiming at a concise statistical characterization of $s$, it is essential that we somehow invert this integration process, the ideal being to apply the operator $L$ which would give back the innovation signal $w$ that is fully decoupled. Unfortunately, this is not feasible in practice because we do not have access to the signal $s(r)$ over the entire domain $r \in \mathbb{R}^d$, but only to its sampled values on a lattice or, more generally, to a series of coefficients in some appropriate basis. Our analysis options are essentially two fold, as described in Sections 2.1.1 and 2.1.2.

2.1.1 Linear combination of sampled values

Given the sampled values $s(k), k \in \mathbb{Z}^d$, the best we can aim at is to implement a discrete version of the operator $L$, which is denoted by $L_d$. In effect, $L_d$ will act on the sampled version of the signal as a digital filter. The corresponding continuous-domain description of its impulse response is

$$L_d \delta(r) = \sum_{k \in \mathbb{Z}^d} d[k] \delta(r - k)$$

with some appropriate weights $d$. To fix ideas, $L_d$ may correspond to the numerical version of the operator provided by the finite-difference method of approximating derivatives.

The interest is now to characterize the (approximate) decoupling effect of this discrete version of the whitening operator. This is quite feasible when the continuous-domain composition of the operators $L$ and $L^{-1}$ is shift-invariant with impulse response $\beta_L(r)$ which is assumed to be absolutely integrable (BIBO stability). In that case, one readily
finds that

\[ u(r) = L_d s(r) = (\beta_L * w)(r) \]  

(2.1)

where

\[ \beta_L(r) = L_d L^{-1} \delta(r). \]  

(2.2)

This suggests that the decoupling effect will be the strongest when the convolution kernel \( \beta_L \) is the most localized and closest to an impulse \(^1\). We call \( \beta_L \) the generalized B-spline associated with the operator \( L \). For a given operator \( L \), the challenge will be to design the most localized kernel \( \beta_L \), which is the way of approaching the discretization problem that best matches our statistical objectives. The good news is that this is a standard problem in spline theory, meaning that we can take advantage of the large body of techniques available in this area, even though they have been hardly applied to the stochastic setting so far.

### 2.1.2 Wavelet analysis

The second option is to analyze the signal \( s \) using wavelet-like functions \( \{ \sqrt{i} \psi_i(\cdot - r_0) \} \). For that purpose, we assume that we have at our disposal some real-valued “\( L \)-compatible” generalized wavelets which, at a given resolution level \( i \), are such that

\[ \sqrt{i} \psi_i(r) = L^* \phi_i(r). \]  

(2.3)

Here, \( L^* \) is the adjoint operator of \( L \) and \( \phi_i \) is some smoothing kernel with good localization properties. The interpretation is that the wavelet transform provides some kind of multiresolution version of the operator \( L \) with the effective width of the kernels \( \phi_i \) increasing in direct proportion to the scale; typically, \( \phi_i(r) \propto \phi_0(r/2^i) \). Then, the wavelet analysis of the stochastic process \( s \) reduces to

\[ \langle s, \psi_i(\cdot - r_0) \rangle = \langle s, L^* \phi_i(\cdot - r_0) \rangle \]

\[ = \langle Ls, \phi_i(\cdot - r_0) \rangle \]

\[ = \langle w, \phi_i(\cdot - r_0) \rangle = (\phi_i^* * w)(r_0) \]  

(2.4)

where \( \phi_i^*(r) = \phi_i(-r) \) is the reversed version of \( \phi_i \). The remarkable aspect is that the effect is essentially the same as in (2.1) so that it makes good sense to develop a common framework to analyze white noise.

This is all nice in principle as long as one can construct “\( L \)-compatible” wavelet bases. For instance, if \( L \) is a pure \( n \)th-order derivative operator—or by extension, a scale-invariant differential operator—then the above reasoning is directly applicable to conventional wavelets bases. Indeed, these are known to behave like multiscale versions of derivatives due to their vanishing-moment property [Mey90, Dau92, Mal09]. In prior work, we have linked this behavior, as well as a number of other fundamental wavelet properties, to the polynomial B-spline convolutional factor that is necessarily present in every wavelet that generates a multiresolution basis of \( L_2(\mathbb{R}) \) [UB03]. What is not so widely known is that the spline connection extends to a much broader variety of operators—not necessarily scale-invariant—and that it also provides a general recipe for constructing wavelet-like basis functions that are matched to some given operator \( L \). This has been demonstrated in 1D for the entire family of ordinary differential operators [KU06]. The only significant difference with the conventional theory of wavelets is that the smoothing kernels \( \phi_i \) are not necessarily rescaled versions of each other.

---

^1: One may be tempted to pretend that \( \beta_L \) is a Dirac impulse, which amounts to neglecting all discretization effects. Unfortunately, this is incorrect and most likely to result in false statistical conclusions. In fact, we shall see that the localization deteriorates as the order of the operator increases, inducing higher (Markov) orders of dependencies.
Note that the “L-compatible” property is relatively robust. For instance, if \( L = L\delta \), then an “L-compatible” wavelet is also \( L'\)-compatible with \( \phi'_i = L\phi_i \). The design challenge in the context of stochastic modeling is thus to come up with a suitable wavelet basis such that \( \phi_i \) in (2.3) is most localized—possibly, of compact support.

### 2.2 Organization of the monograph

The reasoning of Section 2.1 is appealing because of its conceptual simplicity and generality. Yet, the precise formulation of the theory requires some special care because the underlying stochastic objects are infinite-dimensional and possibly highly singular. For instance, we are faced with a major difficulty at the onset because the continuous-domain input of our model (the innovation \( w \)) does not admit a conventional interpretation as a function of the domain variable \( r \). This entity can only be probed indirectly by forming scalar products with test functions in accordance with Laurent Schwartz’ theory of distributions, so that the use of advanced mathematics is unavoidable.

For the benefit of readers who would not be familiar with some the concepts used in this monograph, we provide the relevant mathematical background in Chapter 3, which also serves the purpose of introducing the notation. The first part is devoted to the definition of the relevant function spaces with special emphasis on generalized functions (a.k.a. tempered distributions) which play a central role in our formulation. The second part reviews the classical, finite-dimensional tools of probability theory and shows how some of the concepts (e.g., characteristic function, Bochner’s theorem) are extendable to the infinite-dimensional setting within the framework of Gelfand’s theory of generalized stochastic processes [GV64].

Chapter 4 is devoted to the mathematical specification of the innovation model. Since the theory gravitates around the notion of Lévy exponents, we start with a systematic investigation of such functions, denoted by \( f(\omega) \), which are fundamental to the (classical) study of infinitely-divisible probability laws. In particular, we discuss their canonical representation given by the Lévy-Khinchine formula. In Section 4.4, we make use of the powerful Minlos-Bochner theorem to transfer those representations to the infinite-dimensional setting. The fundamental result for our theory is that the class of admissible continuous-domain innovations for the model in Fig. 2.1 is constrained to the so-called family of white Lévy noises, each brand being uniquely characterized by a Lévy exponent \( f(\omega) \). We conclude the chapter with the presentation of mathematical criteria for the existence of solutions of Lévy-driven SDEs (stochastic differential equations) and provide the functional tools for the complete statistical characterization of these processes. Interestingly, the classical Gaussian processes are covered by the formulation (by setting \( f(\omega) = -\frac{1}{2}|\omega|^2 \)), but they turn out to be the only non-sparse members of the family.

Besides the random excitation \( w \), the second fundamental component of the innovation model in Fig. 2.1 is the inverse \( L^{-1} \) of the whitening operator \( L \). It must fulfill some continuity/boundedness condition in order to yield a proper solution of the underlying SDE. The construction of such inverses (shaping filters) is the topic of Chapter 5, which presents a systematic catalog of the solutions that are currently available, including recent constructs for scale-invariant/unstable SDEs.

In Chapter 6, we review the tools that are available from the theory of splines in relation to the specification of the analysis kernels in Equations (2.1) and (2.3) above. Remarkably, the techniques are quite generic and applicable for any operator \( L \) that admits a proper inverse \( L^{-1} \). This is not too surprising because we have taken advantage of our expert knowledge of splines to engineer/derive the solutions that are presented in Chapter 5. Indeed, by writing a generalized B-spline as \( \beta_\lambda = L_\lambda L^{-1} \delta \), one can appreciate that the
construction of a B-spline for some operator $L$ implicitly provides the solution of two innovation-related problems at once: 1) the formal inversion of the operator $L$ (for solving the SDE), and 2) the proper discretization of $L$ through a finite-difference scheme. The leading thread in our formulation is that these two tasks should not be dissociated—this is achieved formally via the identification of $\tilde{\beta}_L$, which actually results in simplified and streamlined mathematics.

In Chapter 7, we apply our framework to the functional specification of a variety of generalized stochastic processes, including the classical family of Gaussian stationary processes and their sparse counterparts. We also characterize non-stationary processes that are solutions of unstable SDEs. In particular, we describe higher-order extensions of Lévy processes, as well as a whole variety of fractal-type processes.

In Chapter 8, we rely on our functional characterization to obtain a maximally-decoupled representation of sparse stochastic processes by application of the discretized version of the whitening operator or by suitable wavelet expansion.

While the transformed domain statistics can be worked out explicitly, our main point in Chapter 9 is to show that the sparsity pattern of the input noise is essentially preserved. Apart from a shaping effect that can be quantified, the resulting PDF remains within the same family of infinite-divisible laws.

Chapter 10 is devoted to the application of the theory to the general problem of recovering signals from incomplete, noisy measurements, which is highly relevant to signal processing and biomedical imaging. To that end, we develop a general framework for the discretization of linear inverse problems using a suitable set of basis functions (e.g., B-splines or wavelets) which is analogous to the finite-element method for solving PDEs. The central element is the "projection" of the continuous-domain stochastic model of the signal onto the (finite dimensional) reconstruction space in order to specify the prior statistical distribution of the signal. We then apply Bayes’ rule to derive corresponding signal estimators (MAP or MMSE). We present examples of imaging applications including wavelet-domain signal denoising, deconvolution, and the reconstruction of MRI data.
Chapter 3

Mathematical context and background

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 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 $\mathcal{S}'$ and $\mathcal{D}'$, which are spaces of generalized functions. These two categories of spaces (complete-normed
3. **Mathematical Context and Background**

<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 $\mathbf{x} = (x_1, \ldots, x_N)$ 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 $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^{N} x_n y_n$</td>
<td>$\langle \varphi, g \rangle = \int \varphi(r) g(r) , dr$ $\varphi \in \mathcal{S}'(\mathbb{R}^d)$ (test function), $g \in \mathcal{S}'(\mathbb{R}^d)$ (generalized function), or $\varphi \in L_p(\mathbb{R}^d)$, $g \in L_q(\mathbb{R}^d)$ with $\frac{1}{p} + \frac{1}{q} = 1$, for instance.</td>
</tr>
<tr>
<td>equality: $\mathbf{x} = \mathbf{y} \iff x_n = y_n$</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 $\varphi \in \mathcal{S}'(\mathbb{R}^d)$,</td>
</tr>
<tr>
<td>$\iff \langle \mathbf{u}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{y} \rangle$, $\forall \mathbf{u} \in \mathbb{R}^N$</td>
<td>almost-everywhere equality: $f = g \in L_p(\mathbb{R}^d)$ $\iff \int_{\mathbb{R}^d}</td>
</tr>
<tr>
<td>$\iff |\mathbf{x} - \mathbf{y}|^2 = 0$</td>
<td>linear operators $\mathbb{R}^N \rightarrow \mathbb{R}^M$ continuous linear operators $\mathcal{S}'(\mathbb{R}^d) \rightarrow \mathcal{S}'(\mathbb{R}^d)$</td>
</tr>
<tr>
<td>$\mathbf{y} = \mathbf{Ax} \Rightarrow y_m = \sum_{n=1}^{N} a_{mn} x_n$</td>
<td>$\varphi = A\varphi \Rightarrow g(r) = \int_{\mathbb{R}^d} a(r, s) \varphi(s) , ds$ for some $a \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ (Schwartz' kernel theorem)</td>
</tr>
<tr>
<td>transpose $\langle \mathbf{x}, \mathbf{Ay} \rangle = (\mathbf{A}^\top \mathbf{x}, \mathbf{y})$</td>
<td>adjoint $\langle \varphi, A\varphi \rangle = \langle \mathbf{A}^\dagger \varphi, g \rangle$</td>
</tr>
</tbody>
</table>

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.

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

...
3.1. Some classes of function spaces

<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>$\mathcal{P}_X(E) = \text{Prob}(X \in E) = \int_E p_X(x) , dx$ (pdf $p_X$)</td>
<td>$\mathcal{P}_s(E) = \text{Prob}(s \in E) = \int_E \mathcal{P}_s(, dg\rangle$ for suitable subsets $E \subset \mathbb{R}^N$</td>
</tr>
<tr>
<td>characteristic function $\widetilde{\mathcal{P}}<em>X(\xi) = \mathbb{E}(e^{i\langle \xi, x \rangle}) = \int</em>{\mathbb{R}^N} e^{i\langle \xi, x \rangle} p_X(x) , dx$, $\xi \in \mathbb{R}^N$</td>
<td>characteristic functional $\widetilde{\mathcal{P}}<em>s(\varphi) = \mathbb{E}(e^{i\langle \varphi, x \rangle}) = \int</em>{\mathcal{S}} e^{i\langle \varphi, x \rangle} \mathcal{P}_s(, dg\rangle$, $\varphi \in \mathcal{S}$ for suitable subsets $E \subset \mathcal{S}'$</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.

$\mathcal{X}'$, 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 $(\varphi_i)$. \textit{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 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 $\mathbb{C}$),

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

(3.1)

**Definition 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 $\mathcal{X}$ and $\mathcal{Y}$) if, for any convergent sequence $(\varphi_i)$ in $\mathcal{X}$ with limit $\varphi \in \mathcal{X}$, the sequence $(A\varphi_i)$ converges to $A\varphi$ in $\mathcal{Y}$, 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.
3. **Mathematical Context and Background**

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

- \( \| \varphi \| \geq 0 \) (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 \( \mathcal{X} \) equipped with a norm.

A sequence \( (\varphi_i) \) in a normed space \( \mathcal{X} \) is said to **converge to** \( \varphi \) (in the topology of \( \mathcal{X} \)), in symbols

\[
\lim_{i} \varphi_i = \varphi,
\]

if and only if

\[
\lim_{i} \| \varphi - \varphi_i \| = 0.
\]

Let \( (\varphi_i) \) be a sequence in \( \mathcal{X} \) such that for any \( \epsilon > 0 \) there exists an \( N \in \mathbb{N} \) with

\[
\| \varphi_i - \varphi_j \| < \epsilon \quad \text{for all} \ i, j \geq N.
\]

Such a sequence is called a **Cauchy sequence**. A normed space \( \mathcal{X} \) is **complete** if it does not have any holes, in the sense that, for every Cauchy sequence in \( \mathcal{X} \), there exists an \( \varphi \in \mathcal{X} \) such that \( \lim_{i} \varphi_i = \varphi \) (in other words if every Cauchy sequence has a limit in \( \mathcal{X} \)).

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.
3.1. Some classes of function spaces

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, denoted as $\|f\|_{L^p}$, is finite, where

$$
\|f\|_{L^p} := \begin{cases} 
\left( \int_{\mathbb{R}^d} |f(r)|^p \, dr \right)^{\frac{1}{p}} & \text{for } 1 \leq p < \infty, \\
\operatorname{ess sup}_{r \in \mathbb{R}^d} |f(r)| & \text{for } p = \infty
\end{cases}
$$

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 $\alpha > 0$, we denote the associated norms as $\|f\|_{L_{p,\alpha}}$ 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 $L_{p,\alpha}(\mathbb{R}^d) \subset L_{p,\alpha_0}(\mathbb{R}^d)$.
- $L_{\infty,\alpha}(\mathbb{R}^d) \subset L_p(\mathbb{R}^d)$ for any $\alpha > 0$.

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 $\alpha \in \mathbb{R}^+$. 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 $\rho_\alpha(r) = \frac{1}{1+(\rho(r))e^{\alpha r}}$ with $\operatorname{Re}(\alpha) < 0$ 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 $\mathcal{D}(\mathbb{R}^d)$:
3. MATHETMATICAL CONTEXT AND BACKGROUND

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
   \[
   \|\varphi\|_{K,n} := \sup_{r \in K} |\partial^n \varphi(r)| \quad \text{for all } n \in \mathbb{N}^d.
   \]

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 \(\partial^n_{n_1} \cdots \partial^n_{n_d}\). 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 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:

\[
\|\varphi\|_{m,n} := \sup_{r \in \mathbb{R}^d} |r^m \partial^n \varphi(r)| \quad \text{for all } m, n \in \mathbb{N}^d.
\]

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

\[
\varphi(r) \mapsto \hat{\varphi}(\omega) = \mathcal{F}(\varphi)(\omega) := \int_{\mathbb{R}^d} e^{-i(r,\omega)} \varphi(r) \, dr,
\]

and inverted by

\[
\hat{\varphi}(\omega) \mapsto \varphi(r) = \mathcal{F}^{-1}(\hat{\varphi})(r) := \int_{\mathbb{R}^d} e^{i(r,\omega)} \hat{\varphi}(\omega) \, d\omega, \quad (2\pi)^d,
\]

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 \(\mathcal{S}(\mathbb{R}^d)\) 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 \(l_p(\mathbb{N})\), \(1 \leq p < \infty\), for the space of \(p\)-summable sequences; that is, sequences \(c = (c_i)_{i \in \mathbb{N}}\) for which

\[
\sum_{i \in \mathbb{N}} |c_i|^p
\]
3.2 Dual spaces and adjoint operators

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 \( \mathcal{M}_\mathcal{Y} \) the linear operator \( \ell_1(\mathbb{N}) \to \mathcal{Y} \) which maps a sequence \( \mathbf{c} = (c_i)_{i \in \mathbb{N}} \) 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 \( \mathcal{A} : \mathcal{X} \to \mathcal{Y} \), where \( \mathcal{X}, \mathcal{Y} \) are complete normed spaces, is called nuclear if there exists a continuous linear operator \( \mathcal{M} \) such that

\[
\mathcal{A} = \mathcal{M} \mathcal{A}
\]

This is equivalent to the following decomposition of \( \mathcal{A} \) into a sum of rank 1 operators:

\[
\mathcal{A} : \phi \mapsto \sum_{i \in \mathbb{N}} \lambda_i a_i(\phi) \psi_i
\]

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

More generally, given an arbitrary topological vector space \( \mathcal{X} \) and a complete normed space \( \mathcal{Y} \), the operator \( \mathcal{A} : \mathcal{X} \to \mathcal{Y} \) is said to be nuclear if there exists a complete normed space \( \mathcal{X}_1 \), a nuclear operator \( \mathcal{A}_1 : \mathcal{X}_1 \to \mathcal{Y} \), and a continuous operator \( \mathcal{B} : \mathcal{X} \to \mathcal{X}_1 \), such that

\[
\mathcal{A} = \mathcal{A}_1 \mathcal{B}
\]

Finally, \( \mathcal{X} \) is a nuclear space if any continuous linear map \( \mathcal{X} \to \mathcal{Y} \), where \( \mathcal{Y} \) is a complete normed space, is nuclear.

3.2 Dual spaces and adjoint operators

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

When \( \mathcal{X} \) is a vector space, we may consider linear functionals on it, where linearity has the same meaning as in Definition 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 \( \mathcal{X} \), denoted as \( \mathcal{X}^* \), can be given the structure of a vector space in the obvious way by the identity

\[
\langle \varphi, a f + b f_0 \rangle = a \langle \varphi, f \rangle + b \langle \varphi, f_0 \rangle,
\]

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

The map from \( \mathcal{X} \times \mathcal{X}^* \) 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 \( A : X \to Y \), the adjoint or transpose of \( A \), denoted as \( A^* \), is the linear operator \( Y^* \to X^* \) defined by
\[
A^* f = f \circ A
\]
for any linear functional \( f : Y \to \mathbb{R} \) (or \( \mathbb{C} \)) in \( Y^* \), where \( \circ \) denotes composition. The motivation behind the above definition is to have the identity
\[
\langle A \varphi, f \rangle = \langle \varphi, A^* f \rangle \tag{3.2}
\]
hold for all \( \varphi \in X \) and \( f \in Y^* \).

If \( X \) is a topological vector space, it is of interest to consider the subspace of \( X^* \) composed of those linear functionals on \( X \) that are continuous with respect to the topology of \( X \). This subspace is denoted as \( X' \) and called the topological or continuous dual of \( X \). Note that, unlike \( X^* \), the continuous dual generally depends on the topology of \( X \). In other words, the same vector space \( X \) 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 \( X \times X' \). There, the space \( X \) may vary but is necessarily paired with its continuous dual.

Following the restrictions of the previous paragraph, we sometimes say that the adjoint of \( A : X \to Y \) exists, to mean that the algebraic adjoint \( A^* : Y^* \to X^* \), when restricted to \( Y' \), maps into \( X' \), 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 \( Y \times Y' \) and \( X \times X' \), respectively.

One can define different topologies on \( X' \) by providing various criteria for convergence. The only one we shall need to deal with is the weak-* topology, which indicates (for a sequential space \( X \)) that \( (f_i) \) converges to \( f \) in \( X' \) if and only if
\[
\lim_i \langle \varphi, f_i \rangle = \langle \varphi, f \rangle \quad \text{for all } \varphi \in X.
\]
This is precisely the topology of pointwise convergence for all “points” \( \varphi \in X \).

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
\[
\langle \varphi, f \rangle = \int_{\mathbb{R}^d} \varphi(r) f(r) \, dr \tag{3.3}
\]
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.
3.2. Dual spaces and adjoint operators

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

\[
|\langle \varphi, f \rangle| = \int \varphi(r) f(r) \, dr \leq \|\varphi\|_{L^p} \|f\|_{L^{p'}}
\]

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 \emph{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 \emph{Dirac functional} \( \delta \), 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 \emph{space of tempered} (or \emph{Schwartz}) \textit{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 \( A^H \). It is defined as \( \langle A^H f, \xi \rangle_{L_2} = \langle f, A \xi \rangle_{L_2} = \langle f, \overline{A} \xi \rangle \) which implies that \( A^H = \overline{A}^\dagger \). 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

\[
\hat{f}(\omega) = \mathcal{F}(f)(\omega) := \int_{\mathbb{R}^d} f(r) e^{-j r \cdot \omega} \, dr
\]

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

\(^2\) A function on \( \mathbb{R}^d \) is called locally integrable if its integral over any closed bounded set is finite.
3. Mathematical context and background

the $L_2$ norm of a function (up to a normalization factor) is a direct 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 $\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle$.

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 $\{\varphi_i\}$ 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/curves inside $\mathbb{R}^d$. An example of a singular function $f$, which we have already noted, is the Dirac distribution $\delta$ 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 $\mathcal{X}'$ of a nuclear space $\mathcal{X}$ of functions, such as $\mathcal{D}'(\mathbb{R}^d)$ or $\mathcal{S}'(\mathbb{R}^d)$.

3. The connection with previous sections should already be apparent from this choice of notation.
3.3.2 Operations on generalized functions

Following (3.2), any continuous linear operator $\mathcal{D} \to \mathcal{D}$ or $\mathcal{F} \to \mathcal{F}$ can be transposed to define a continuous linear operator $\mathcal{D}^* \to \mathcal{D}^*$ or $\mathcal{F}^* \to \mathcal{F}^*$. In particular, since $\mathcal{D}^0(\mathbb{R}^d)$ and $\mathcal{F}(\mathbb{R}^d)$ 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 $\mathcal{D}^0(\mathbb{R}^d)$ or $\mathcal{F}(\mathbb{R}^d)$ we see that $\partial^n \varphi = (-1)^{|n|} \partial^n f$. In other words, we can write

$$\langle \varphi, \partial^n f \rangle = (-1)^{|n|} \langle \partial^n \varphi, f \rangle. \quad (3.4)$$

The idea is then to use (3.4) as the defining formula in order to extend the action of the derivative operator $\partial^n$ for any $f \in \mathcal{D}^0(\mathbb{R}^d)$ or $\mathcal{F}(\mathbb{R}^d)$.

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.** Given operators $U, U^* : \mathcal{F}(\mathbb{R}^d) \to \mathcal{F}(\mathbb{R}^d)$ that form an adjoint pair on $\mathcal{F}(\mathbb{R}^d) \times \mathcal{F}(\mathbb{R}^d)$, we extend their action to $\mathcal{F}'(\mathbb{R}^d) \to \mathcal{F}'(\mathbb{R}^d)$ 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 $\mathcal{D}^0(\mathbb{R}^d) \to \mathcal{D}^0(\mathbb{R}^d)$ to operators $\mathcal{D}'(\mathbb{R}^d) \to \mathcal{D}'(\mathbb{R}^d)$.

Examples of operators $\mathcal{F}(\mathbb{R}^d) \to \mathcal{F}(\mathbb{R}^d)$ 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.

3.3.3 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 $\mathcal{G}(\mathbb{R}^d)$ into itself. The additional relevant property is that $\mathcal{F}$ is self-adjoint: $\langle \varphi, \mathcal{F} \psi \rangle = \langle \mathcal{F} \varphi, \psi \rangle$, for all $\varphi, \psi \in \mathcal{F}(\mathbb{R}^d)$. This helps us specifying the generalized Fourier transform of distributions in accordance with the general extension principle in Definition 3.

**Definition 4.** The generalized Fourier transform of a distribution $f \in \mathcal{F}'(\mathbb{R}^d)$ is the distribution $\hat{f} = \mathcal{F}(f) \in \mathcal{F}'(\mathbb{R}^d)$ that satisfies

$$\langle \varphi, \hat{f} \rangle = \langle \hat{\varphi}, f \rangle$$

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

$$\hat{\varphi}(\omega) = \int_{\mathbb{R}^d} e^{-i(r, \omega)} \varphi(r) \, dr.$$
3. **Mathematical context and background**

<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^\vee(r) = f(-r) )</td>
<td>( \hat{f}(-\omega) = \hat{f}^\vee(\omega) )</td>
</tr>
<tr>
<td>( f(r) )</td>
<td>( \hat{f}(-\omega) )</td>
</tr>
<tr>
<td>( f(\Lambda^\top r) )</td>
<td>( \frac{1}{</td>
</tr>
<tr>
<td>( f(r - r_0) )</td>
<td>( e^{-i&lt;\omega_0, \omega&gt;} \hat{f}(\omega) )</td>
</tr>
<tr>
<td>( e^{i\omega_0 \cdot r} f(r) )</td>
<td>( \hat{f}(\omega - \omega_0) )</td>
</tr>
<tr>
<td>( \partial^n f(r) )</td>
<td>( (i\omega)^n \hat{f}(\omega) )</td>
</tr>
<tr>
<td>( r^n f(r) )</td>
<td>( i^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} \ast \hat{f}(\omega))</td>
</tr>
</tbody>
</table>

Table 3.3: Basic properties of the (generalized) Fourier transform.

For example, since we have

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

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 \( \mathcal{F}(f) = \mathcal{F}(f^\vee) \). 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 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 1** (Schwartz’ kernel theorem: first form). Every continuous linear operator \( \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d) \) can be written in the form

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

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

We can interpret the above formula as some sort of continuous-domain matrix-vector product, where \( r, s \) play the role of vector indices. 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 some important ways.

An equivalent statement of the above theorem is as follows.

**Theorem 2** (Schwartz’s kernel theorem: second form). Every continuous bilinear form \( l: \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \rightarrow \mathbb{R} \) (or \( \mathbb{C} \)) can be written as

\[
l(\varphi_1, \varphi_2) = \langle \varphi_1(\cdot) \varphi_2(\cdot), h(\cdot, \cdot) \rangle = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi_1(r) \varphi_2(s) h(r, s) \, ds \, dr.
\]

The connection between the two statements is clarified by relating the continuous bilinear form \( l \) to a continuous linear operator \( A : \mathcal{S}(\mathbb{R}^d) \rightarrow \mathcal{S}'(\mathbb{R}^d) \) by an identity of the form

\[
l(\varphi_1, \varphi_2) = \langle \varphi_1, A \varphi_2 \rangle.
\]

3.3.5 Linear shift-invariant (LSI) operators and convolutions

Let \( S_{\tau_0} \) denote the shift operator \( \varphi \mapsto \varphi(\cdot - \tau_0) \). We call an operator \( U \) shift-invariant if

\[
US_{\tau_0} = S_{\tau_0} U \quad \text{for all} \quad \tau_0 \in \mathbb{R}^d.
\]

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

**Corollary 1.** Every continuous linear shift-invariant operator \( \mathcal{F}(\mathbb{R}^d) \rightarrow \mathcal{F}'(\mathbb{R}^d) \) can be written as a convolution

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

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

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

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

(3.6)

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, \( \varphi * h \) 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

\[
\hat{h}(\omega) = C_0 \prod_{n=1}^{M} \frac{i(\omega - z_m)}{\Pi_{n=1}^{N} (i(\omega - p_n))}
\]

with no purely imaginary pole (i.e., with \( \text{Re}(p_n) \neq 0, 1 \leq n \leq N \)).

\[4.\] For \( M \) or \( N = 0 \), we shall take the corresponding product to be equal to 1.
3. MATHEMATICAL CONTEXT AND BACKGROUND

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 $\mathcal{S}(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$ can be written in the form specified by (3.5).

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 $\mathcal{S} \to \mathcal{S}'$. 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 $\mathcal{S} \subset \mathcal{S}'$ (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., $A_h : \mathcal{S} \to \mathcal{X}$), and furthermore, $A_f$ has a continuous extension to $\mathcal{X}$; that is, we can extend it as $A_f : \mathcal{X} \to \mathcal{S}'$.

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

$$\frac{1}{r} = 1 - \frac{1}{p} - \frac{1}{q}.$$  

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 $p = r$:

$$\|h \ast f\|_{L_p} \leq \|h\|_{L_1} \|f\|_{L_p}. \quad (3.7)$$

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 (3.7) 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 analysis, these operators are commonly referred to as $L_p$ Fourier multipliers using (3.6) as starting point for their definition.

**Definition 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
3.3. Generalized functions

convolution \( Tf = h \ast f \) (see Corollary 1) where \( h \in \mathcal{S}'(\mathbb{R}^d) \) is the impulse response 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 2) and the boundedness of the operator.

**Definition 6 (Operator norm).** The norm of the linear operator \( T : L_p(\mathbb{R}^d) \to L_p(\mathbb{R}^d) \) is given by

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

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 (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 \( C_p = \|T\|_p < \infty \) such that

\[
\|Tf\|_p \leq C_p \|f\|_p
\]

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 4 (Characterization of \( L_p \) Fourier multipliers).** Let \( T \) be a Fourier-multiplier operator with frequency response \( H : \mathbb{R}^d \to \mathbb{C} \). Then, the following statements apply:

1) The operator \( T \) is an \( L_1 \) Fourier multiplier if and only if \( \hat{H} = \hat{h} \) is the Fourier transform of a finite complex-valued Borel measure.

2) The operator \( T \) is an \( L_\infty \) Fourier multiplier if and only if \( \hat{H} = \hat{h} \) is the Fourier transform of a finite complex-valued Borel measure.

3) The operator \( T \) is an \( L_2 \) Fourier multiplier if and only if \( \hat{h} \in L_\infty(\mathbb{R}^d) \).

The corresponding operator norms are

\[
\|T\|_{L_1} = \|T\|_{L_\infty} = \|h\|_{TV}
\]

\[
\|T\|_{L_2} = \frac{1}{(2\pi)^{d/2}} \|H\|_{L_\infty},
\]

where \( \|h\|_{TV} \) 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 \( H \) (or \( h \)) in 1) or 2) is met with

\[
\|T\|_{L_p} \leq \|h\|_{TV}.
\]

We note that the above theorem is an extension upon (3.7) 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 \( \mu \) admits a unique decomposition as

\[
\mu = \mu_{ac} + \mu_{sing},
\]

where \( \mu_{ac} \) is an absolutely-continuous measure and \( \mu_{sing} \) a singular measure whose mass is concentrated on a set whose Lebesgue measure is zero. If \( \mu_{sing} = 0 \), then there exists
3. **Mathematical Context and Background**

A unique function $h \in L_1(\mathbb{R}^d)$—the Radon-Nikodym derivative of $\mu$ with respect to the Lebesgue measure—such that

$$\int_{\mathbb{R}^d} \varphi(r) \mu(dr) = \int_{\mathbb{R}^d} \varphi(r) h(r) dr,$$

so that the total variation in Theorem 4 reduces to the $L_1$ norm $\|h\|_{TV} = \|h\|_{L_1}$. Under those circumstances, there is an equivalence between (3.7) and (3.8).

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

Statement 3) in Theorem 4 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 4 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 5 below), but fails to meet the necessary and sufficient TV condition for $p = 1, \infty$, is the Hilbert transform. Its frequency response is $H_{\text{Hilbert}}(\omega) = -j\text{sign}(\omega)$, 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 Mihlin’s multiplier theorem which provides a sufficient condition on the frequency response of a filter for $L_p$ stability.

**Theorem 5** (Mihlin). *A Fourier-multiplier operator is bounded in $L_p(\mathbb{R}^d)$ for $1 < p < \infty$ if its frequency response $H: \mathbb{R}^d \rightarrow \mathbb{C}$ satisfies the differential estimate

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

Mihlin’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

In this section we give an informal introduction to probability theory. A more precise presentation can be found in Appendix D.

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

Frequently, the collection $\mathcal{G}_X$ contains open and closed sets, as well as their countable unions and intersections, collectively known as Borel sets. In this case we call $P_X$ 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 distribution (pdf) or density function, of $X$, such that

$$P_X(A) = \int_A p_X(x) \, dx$$

for suitable subsets $A$ of $\mathbb{R}^n$.  

More generally, the expected value of $f : X \rightarrow \mathbb{C}$ is here given by

$$E(f(X)) = \int_{\mathbb{R}^n} f(x) p_X(x) \, dx. \quad (3.9)$$

We say “more generally” because $P(X)$ can be seen as the expected value of the indicator function $\mathbb{I}_A(X)$. 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 $\mathbb{R} \rightarrow \mathbb{R}$), 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 $f : \Omega_X \rightarrow \mathbb{R}$. In effect, a definition of the integral of $f$ with respect to probability measure $P_X$ 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

$$E(f(X)) = \int_{\Omega_X} f(x) P_X(dx),$$

which we may also use, in addition to (3.9), in the case of a finite-dimensional $\Omega_X$.

In general, given a function $f : \Omega_X \rightarrow \Omega_Y$ 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,

$$P_Y(B) = P_X(f^{-1}(B)) = P_X \circ f^{-1}(B),$$

5. Probability distributions should not be confused with the distributions in the sense of Schwartz (i.e., generalised 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.

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. Mathematical context and background

where the inverse image \( f^{-1}(B) \) is defined as
\[
 f^{-1}(B) = \{ x \in \Omega_X : f(x) \in B \}.
\]

\( \mathcal{P}_Y = \mathcal{P}_X(f^{-1} \cdot) \) is called the push-forward of \( \mathcal{P}_X \) through \( f \).

### 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 \( \xi \in \mathbb{R}^n \), we can compute the expected value (integral) of the bounded function \( x \mapsto e^{i\langle \xi, x \rangle} \). This permits us to define a complex-valued function on \( \mathbb{R}^n \) by the formula
\[
 \hat{\mathcal{P}}_X(\xi) = E[e^{i\langle \xi, X \rangle}] = \int_{\mathbb{R}^n} e^{i\langle \xi, x \rangle} p_X(x) \, dx = \mathcal{F}(p_X)(\xi), \tag{3.10}
\]
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 \( \langle \xi, x \rangle \), which is the opposite of the convention used in analysis. To avoid confusion, we use the variable \( \xi \) as the Fourier variable in probabilistic Fourier transforms (characteristic functions), and \( \omega \) in the analytic definition.

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 32 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 6.

**Theorem 6** (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
\[
 \hat{\mathcal{P}}_X(\xi) = \int_{\mathbb{R}^n} e^{i\langle \xi, x \rangle} \mathcal{P}_X(dx) = E[e^{i\langle \xi, X \rangle}] = \mathcal{F}(p_X)(\xi). \tag{3.10}
\]

Conversely, the function specified by (3.10) with \( p_X(r) \geq 0 \) and \( \int_{\mathbb{R}^n} p_X(r) \, dr = 1 \) is positive-definite, uniformly continuous, and such that \( |\hat{\mathcal{P}}_X(\xi)| \leq \hat{\mathcal{P}}_X(0) = 1 \).
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 $P_X$, it is called the characteristic function of probability measure $P_X$ (recall that the probability measure $P_X$ is related to the density $p_X$ by $P_X(E) = \int_E p_X(x) \, dx$ 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 7** (Lévy’s continuity theorem). *Let $(P_{X_i})$ 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 \[ \lim_{i} \hat{p}_{X_i}(\xi) = \hat{p}_X(\xi) \] 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 $P_X$ on $\mathbb{R}^n$. Moreover, $P_{X_i}$ converges weakly to $P_X$, in symbols\[ P_{X_i} \overset{w}{\to} P_X, \]
meaning for any continuous function $f : \mathbb{R}^n \to \mathbb{R}$, \[ \lim_{i} E_{X_i} \{ f \} = E_X \{ f \}. \]

The reciprocal of the above theorem is also true; namely, if $P_{X_i} \overset{w}{\to} P_X$, then $\hat{p}_{X_i}(\xi) \to \hat{p}_X(\xi)$ pointwise.

### 3.4.4 Characteristic functionals in infinite dimensions

Given a probability measure $P_X$ 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 $P_X$ and denoted as $\mathcal{P}_X$, by means of the identity

\[ \mathcal{P}_X(\varphi) = E[\exp(i\varphi,X)]. \quad (3.11) \]

Comparing the above definition with (3.10), 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**: $\mathcal{P}_X$ is positive-definite, in the sense that for any $N$ (test) functions $\varphi_1,\ldots,\varphi_N$, for any $N \times N$ matrix with entries $p_{ij} = \mathcal{P}_X(\varphi_i - \varphi_j)$ is non-negative definite.
- **Normalization**: $\mathcal{P}_X(0) = 1$.

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 $\mathcal{P}_X$ 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 nuclear space (and, in particular, for $\mathcal{X} = C(\mathbb{R}^d)$ or $\mathcal{S}(\mathbb{R}^d)$, cf. Subsection 3.1.3) such a condition is given by the Minlos-Bochner theorem.

**Theorem 8** (Minlos-Bochner). *Let $\mathcal{X}$ be a nuclear space and let $\mathcal{P}_X : \mathcal{X} \to \mathbb{C}$ be a functional that is positive-definite in the sense discussed above, fulfills $\mathcal{P}_X(0) = 1$, and is con-
3. Mathematical context and background

\begin{align*}
\mathcal{X} & \rightarrow \mathbb{C}. \text{ Then, there exists a unique probability measure } \mathcal{P}_\mathcal{X} \text{ on } \mathcal{X}' \text{ (the continuous dual of } \mathcal{X}), \text{ such that } \\
\mathcal{P}_\mathcal{X}(\varphi) &= \int_{\mathcal{X}'} e^{i\varphi,x} \mathcal{P}_\mathcal{X}(dx) = E[e^{i\varphi,X}].
\end{align*}

Conversely, the characteristic functional associated to some probability measure \( \mathcal{P}_\mathcal{X} \) on \( \mathcal{X}' \) is positive-definite, continuous over \( \mathcal{X} \), and such that \( \mathcal{P}_\mathcal{X}(0) = 1 \).

The practical implication of this result is that one can rely on characteristic functionals to indirectly specify infinite-dimensional measures (most importantly, probabilities of stochastic processes)—which are difficult to pin down otherwise. Operationally, the characteristic functional \( \mathcal{P}_\mathcal{X}(\varphi) \) is nothing but a mathematical rule (e.g., \( \mathcal{P}_\mathcal{X}(\varphi) = e^{-\frac{1}{2} \varphi_2^2} \)) that returns a value in \( \mathbb{C} \) for any given function \( \varphi \in \mathcal{F} \). The truly powerful aspect is that this rule condenses all the information about the statistical distribution of some underlying infinite-dimensional random object \( X \). When working with characteristic functionals, we shall see that computing probabilities and deriving various properties of the said processes are all reduced to analytical derivations.

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.

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,

\(^7\) 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.
Important example is that of white Gaussian noise, which one may conceive of as associating a Gaussian random variable with exponent 2 to any test function \( \varphi \) in some suitable test-function space \( \mathcal{X} \) is associated a random variable \( s(\varphi) \), also often denoted as \( \langle \varphi, s \rangle \). 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 \( \langle \varphi, s \rangle \) as \( \mathcal{P}_{s,\varphi} \). Similarly, to any finite collection of observations \( \langle \varphi_i, s \rangle, 1 \leq i \leq N, N \in \mathbb{N} \), 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 \( \langle \varphi_i, s \rangle, 1 \leq i \leq N \), and \( \langle \psi_j, s \rangle, 1 \leq j \leq M \), need to be consistent or compatible, as explained in Appendix D.4, to ensure that all computations of the probability 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 \( \langle \varphi, s \rangle \) as a function of \( \varphi \).

Mathematically, these requirements are fulfilled by the kind of probabilistic model induced by a cylinder-set probability measure, as discussed in Appendix D.4. 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 \( \varphi \in \mathcal{X} \). Furthermore, a cylinder-set probability measure can always be specified via its characteristic functional \( \mathcal{P}_s(\varphi) = \mathbb{E}(e^{i \langle \varphi, s \rangle}) \), 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 \( \mathcal{X}' \) of \( \mathcal{X} \), since we have not yet defined a proper probability measure on \( \mathcal{X}' \). Doing so involves some additional steps.

---

8. In fact, \( \mathcal{X}' \) may very well be too small to support such a description (while the algebraic dual, \( \mathcal{X}^* \), 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 \( \| \varphi \|_2^2 \) to any test function \( \varphi \in L_2 \). However, the “energy” of white Gaussian noise is clearly
3.5.2 Generalized random processes as random generalized functions

Fortunately, the above existence and interpretation problem is fully resolved by taking $\mathcal{X}$ to be a nuclear space, thanks to the Minlos-Bochner theorem (Theorem 8). 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 $\mathcal{X}'$ (the topological dual of $\mathcal{X}$), as sketched out in Appendix D.4.

In this case, the joint probabilities $\mathcal{P}_{s\varphi_1,\ldots,\varphi_N}$, $\varphi_1,\ldots,\varphi_N \in \mathcal{X}$, $N \in \mathbb{N}$, corresponding to the random variables $(\varphi_i, s) = (s, \varphi_i)$ for all possible choices of test functions, collectively define a probability measure $\mathcal{P}_s$ on the infinite-dimensional dual space $\mathcal{X}'$. This means that we can view $s$ as an element drawn randomly from $\mathcal{X}'$ according to the probability law $\mathcal{P}_s$.

In particular, if we take $\mathcal{X}$ 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 $(\varphi, s)$ as the measurement of this random object $s$ by means of some sensor (test function) $\varphi$ 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 $\mathcal{X}'$ (dual to some nuclear space $\mathcal{X}$) 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 $\hat{\mathcal{P}}_s : \mathcal{X} \to \mathbb{C}$ with proper normalization identifies a unique probability measure $\mathcal{P}_s$ on $\mathcal{X}'$. Therefore, to define a generalized random process $s$ with realizations in $\mathcal{X}'$, it suffices to produce a functional $\hat{\mathcal{P}}_s : \mathcal{X} \to \mathbb{C}$ with the noted properties.

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 characteristic function and is recorded for further reference.

**Proposition 1.** Let $Y = ((\varphi_1, s),\ldots,(\varphi_N, s))$ with $\varphi_1,\ldots,\varphi_N \in \mathcal{X}$ be a set of linear measurements of the generalized stochastic process $s$ with characteristic functional $\hat{\mathcal{P}}_s(\varphi) = \mathbb{E}[e^{i\varphi \cdot s}]$ that is continuous over the function space $\mathcal{X}$. Then,

$$\hat{p}_Y(\xi) = \hat{\mathcal{P}}_{s;\varphi_1,\ldots,\varphi_N}(\xi) = \hat{\mathcal{P}}_s \left( \sum_{n=1}^{N} \xi_n \varphi_n \right)$$

and the joint pdf of $Y$ is given by

$$p_Y(y) = \mathcal{F}^{-1}(\hat{p}_Y)(y) = \int_{\mathbb{R}^N} \hat{\mathcal{P}}_s \left( \sum_{n=1}^{N} \xi_n \varphi_n \right) e^{-i\langle y, \xi \rangle} \frac{d\xi}{(2\pi)^N},$$

where the observation functions $\varphi_n \in \mathcal{X}$ are fixed and $\xi = (\xi_1,\ldots,\xi_N)$ plays the role of the $N$-dimensional Fourier variable.
3.5. Generalized random processes and fields

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 \( Y \), we have

\[
\hat{p}_Y(\xi) = \mathbb{E} \left\{ \exp \left\{ j \sum_{n=1}^{N} \xi_n \psi_n(s) \right\} \right\}
\]

(by linearity of duality product)

\[
= \mathbb{P}_s \left( \sum_{n=1}^{N} \xi_n \psi_n \right) \tag{by definition of \( \mathbb{P}_s(\psi) \)}
\]

The density \( p_Y \) 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 correlation functionals defined and computed as

\[
\mathcal{M}_s(\psi) := \mathbb{E}(\langle \psi, s \rangle) = (-j) \frac{d}{d\xi} \mathbb{P}_s,\psi(\xi) \big|_{\xi=0}
\]

\[
= (-j) \frac{d}{d\xi} \mathbb{P}_s,\psi(\xi) \big|_{\xi=0}.
\]

\[
\mathcal{R}_s(\psi_1, \psi_2) := \mathbb{E}(\langle \psi_1, s \rangle \langle \psi_2, s \rangle) = (-j)^2 \frac{\partial^2}{\partial \xi_1 \partial \xi_2} \mathbb{P}_s,\psi_1,\psi_2(\xi_1, \xi_2) \big|_{\xi_1, \xi_2=0}
\]

\[
= (-j)^2 \frac{\partial^2}{\partial \xi_1 \partial \xi_2} \mathbb{P}_s(\xi_1 \psi_1 + \xi_2 \psi_2) \big|_{\xi_1, \xi_2=0}.
\]

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

\[
\mathcal{M}_s(\psi) = \int_{\mathbb{R}^d} \psi(r) m_s(r) \, dr, \tag{3.12}
\]

\[
\mathcal{R}_s(\psi_1, \psi_2) = \int_{\mathbb{R}^d} \psi_1(r) \psi_2(s) c_s(r, s) \, dr. \tag{3.13}
\]

The first identity is simply a consequence of \( \mathcal{M}_s \) being a continuous linear functional on \( \mathcal{X} \), while the second is an application of Schwartz' kernel theorem (Theorem 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 \( \mathbb{P}_w \), 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 \( U : \mathcal{X} \to \mathcal{Y} \) with continuous adjoint \( U^* : \mathcal{Y}^* \to \mathcal{X}^* \), where \( \mathcal{X}, \mathcal{Y} \) need not be nuclear, and a functional

\[
\mathbb{P}_w : \mathcal{Y} \to \mathbb{C}
\]
that satisfies the three conditions of Theorem 8 (continuity, positive-definiteness, and normalization), we obtain a new functional

$$\mathcal{P}_s : \mathcal{X} \to \mathbb{C}$$

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

$$\mathcal{P}_s = \mathcal{P}_w \circ U$$

for all $\varphi \in \mathcal{X}$. (3.14)

Writing

$$\mathcal{P}_s(\xi \varphi) = \mathbb{E}\{e^{i\xi(\varphi, s)}\} = \hat{p}(\varphi, s)(\xi)$$

and

$$\mathcal{P}_w(\xi U \varphi) = \mathbb{E}\{e^{i\xi(U \varphi, w)}\} = \hat{p}(U \varphi, w)(\xi)$$

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

$$\langle \varphi, s \rangle = \langle U \varphi, w \rangle$$

in probability law.

The manipulation that led to Proposition 1 shows that a similar relation exists, more generally, for any finite collection of observations $\langle \varphi_1, s \rangle$ and $\langle U \varphi_i, w \rangle$, $1 \leq i \leq N$, $N \in \mathbb{N}$.

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 (3.14), 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 $\mathcal{P}_s : \mathcal{X} \to \mathbb{C}$ 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 $\mathcal{P}_w : \mathcal{Y} \to \mathbb{C}$ 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

$$\mathcal{P}_s = \mathcal{P}_w \circ U$$

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 $\mathcal{X} = \mathcal{H}(\mathbb{R}^d)$ 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 $w = L s$; that is,

$$\langle \varphi, w \rangle = \langle \varphi, L s \rangle,$$

for all $\varphi \in \mathcal{Y}$.

The above ideas are summarized in Figure 3.1.
3.5. Generalized random processes and fields

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 $φ_1$, $w_i$ and $φ_2$, $w_i$ be independent whenever the test functions $φ_1$ and $φ_2$ 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 $c_{P_w}$ of $w$ as

$$c_{P_w}(φ_1 + φ_2) = c_{P_w}(φ_1) c_{P_w}(φ_2).$$

An important class of characteristic functionals fulfilling this requirement are those that can be written in the form

$$c_{P_w}(φ) = e^{∫_R R_f(r) dr}.$$  \hspace{1cm} (3.15)

To have $c_{P_w}(0) = 1$ (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 charac-
teristic functional is given by
\[ \mathcal{P}_w(\varphi) = e^{-\frac{1}{2} \|\varphi\|_2^2}. \]
Note that this functional is a special instance of (3.15) with \( f(\xi) = \frac{1}{2} \xi^2 \). 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 1 to obtain the joint characteristic function
\[ \mathcal{P}_{\varphi_1, \varphi_N}(\xi) = \exp \left( -\frac{1}{2} \sum_{i=1}^{N} \xi_i \varphi_i^2 \right). \]
By taking the inverse Fourier transform of the above expression, we find that the random variables \( \langle \varphi_1, w \rangle, \varphi_i, \ldots, \langle \varphi_N, w \rangle \) have a multivariate Gaussian distribution with mean \( 0 \) and covariance matrix with entries
\[ C_{ij} = \langle \varphi_i, \varphi_j \rangle. \]
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. 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 \( C_{ij} = 0 \).

From Formulas (3.12) and (3.13), we also find that \( w \) has 0 mean and “correlation function” \( c_w(r, s) = \delta(r-s) \), 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 \( U \) 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 \( h \). The convolution kernel of \( U^* : \mathcal{S} \rightarrow \mathcal{S} \) is then \( h^\ast \). Following Section 3.5.4, we find the following characteristic functional for the filtered process \( U^* w = h^\ast \ast w \):
\[ \mathcal{P}_{U^* w}(\varphi) = e^{-\frac{1}{2} \|h \ast \varphi\|_2^2}. \]
In turn, it yields the following mean and correlation functions
\[ m_{U^* w}(r) = 0, \]
\[ c_{U^* w}(r, s) = (h \ast h^\ast)(r-s), \]
as expected.

---
9. 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.
10. 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.
Chapter 4

Continuous-domain innovation models

The stochastic processes that we wish to characterize are those generated by linear transformation of non-Gaussian white noise. If we were operating in the discrete domain and restricting ourselves to a finite number of dimensions, we would be able to use any sequence of i.i.d. random variables \( w_n \) as system input and rely on conventional multivariate statistics to characterize the output. This strongly suggests that the specification of the mixing matrix \( (L^{-1}) \) and the probability density function (pdf) of the innovation is sufficient to obtain a complete description of a linear stochastic process, at least in the discrete setting.

But our goal is more ambitious since we place ourselves in the context of continuously-defined processes. Then, the situation is not quite as straightforward because: 1) we are dealing with infinite dimensional objects, 2) it is much harder to properly define the notion of continuous-domain white noise, and 3) there are theoretical restrictions on the class of admissible innovations. While this calls for an advanced mathematical machinery, the payoff is that the continuous-domain formalism lends itself better to analytical computations, by virtue of the powerful tools of functional and harmonic analysis. Another benefit is that the non-Gaussian members of the family are necessarily sparse as a consequence of the theory which rests upon the powerful characterization and existence theorems by Lévy-Khinchine, Minlos-Bochner, and Gelfand-Vilenkin.

As in the subsequent chapters, we start by providing some intuition in the first section and then proceed with a more formal characterization. Section 4.2 is devoted to an in-depth investigation of Lévy exponents which are intimately tied to the family of infinitely divisible distributions in the classical (scalar) theory of probability. What is nonstandard here and fundamental to our argumentation is the link that is made between infinite divisibility and sparsity in 4.2.3. In Section 4.3, we apply those results to the Fourier-domain characterization of a multivariate linear model driven by an infinitely divisible noise vector, which primarily serves as preparation for the subsequent infinite-dimensional generalization. In Section 4.4, we extend the formulation to the continuous domain which results in the proper specification of white Lévy noise \( w \) (or non-Gaussian innovations) as a generalized stochastic process (in the sense of Gelfand and Vilenkin) within independent "values" at every point. The fundamental result is that a given brand of noise (or innovations) is uniquely specified by its Lévy exponent \( f(\omega) \) via its characteristic form \( \mathcal{P}_w(\varphi) \). Finally, in Section 4.5, we characterize the statistical effect of the mixing operator \( L^{-1} \) (general linear model) and provide mathematical conditions on \( f \) and \( L \) that ensure that the resulting process \( s = L^{-1}w \) is well-defined mathematically.
4. Continuous-domain innovation models

Figure 4.1: Examples of canonical, infinitely divisible probability density functions and corresponding observations of a continuous-domain white-noise process through an array of non-overlapping rectangular integration windows: (a) Gaussian distribution (not sparse), (b) Laplace distribution (moderately sparse), (c) Compound-Poisson distribution (finite rate of innovation), (d) Cauchy distribution (ultra-sparse = heavy-tailed with unbounded variance).

4.1 Introduction: From Gaussian to sparse probability distributions

Intuitively, a continuous-domain white-noise process is formed by the juxtaposition of a continuum of i.i.d. random contributions. Since these atoms of randomness are infinitesimal, the realizations (a.k.a. sample paths) of such processes are highly singular (discontinuous), meaning that they do not admit a classical interpretation as (random) functions of the index variable \( r \in \mathbb{R}^d \). Consequently, the random variables associated with the sample values \( w(r_0) \) are undefined. The only concrete way of observing such noises is by probing them through some localized analysis window \( \varphi(-r_0) \) centered around some location \( r_0 \). This produces some scalar quantity \( X = \langle \varphi(-r_0), w \rangle \) which is a conventional random variable with some pdf \( p_X(\varphi) \). Note that \( p_X(\varphi) \) is independent upon the position \( r_0 \), which reflects the fact that \( w \) is stationary. In order to get some sense about the variety of achievable random patterns, we propose to convert the continuous-domain process \( w \) into some corresponding i.i.d. sequence \( X_k \) (discrete white noise) by selecting a sequence of non-overlapping rectangular windows:

\[
X_k = \langle \text{rect}(\cdot - k), w \rangle.
\]

The concept is illustrated in Figure 4.1. The main point that will be made clearer in the sequel is that there is a one-to-one correspondence between the pdf of \( X_k \)—the so-called
4.2. Lévy exponents and infinitely divisible distributions

The investigation of sparse stochastic processes requires a solid understanding of the classical notions of Lévy exponents and infinite divisibility, which constitute the pillars of our formulation. This section provides a self-contained presentation of the required mathematical background. It also brings out the link with sparsity.

**Definition 5** (Lévy exponent). A continuous, complex-valued function \( f : \mathbb{R} \to \mathbb{C} \) such that \( f(0) = 0 \) is a valid Lévy exponent iff. it is conditionally positive-definite of order one, so that

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} f(\omega_m - \omega_n) \xi_m \bar{\xi}_n \geq 0
\]

under the condition \( \sum_{m=1}^{N} \xi_m = 0 \) for every possible choice of \( \omega_1, \ldots, \omega_N \in \mathbb{R}, \xi_1, \ldots, \xi_N \in \mathbb{C} \) and \( N \in \mathbb{Z}^+ \).

The importance of Lévy exponents in mathematical statistics is that they are tightly linked with the property of infinite divisibility.

**Definition 6** (Infinite divisibility). A random variable \( X \) with generic pdf \( p_\text{id} \) is infinitely divisible (id) iff., for any \( N \in \mathbb{Z}^+ \), there exist i.i.d. random variables \( X_1, \ldots, X_N \) such that \( X \) has the same distribution as \( X_1 + \cdots + X_N \).

The foundation of the theory of such random variables is that their characteristic functions are in one-to-one correspondence with Lévy exponents. While the better known formulation of this equivalence is provided by the Lévy-Khintchine theorem (Theorem 7), we first like to express it in functional terms, building upon the work of three giants in harmonic analysis: Lévy, Bochner, and Schoenberg.

**Theorem 6** (Lévy-Schoenberg). Let \( \hat{p}_\text{id}(\omega) = \mathbb{E}[e^{j\omega X}] = \int_{\mathbb{R}} e^{j\omega x} p_\text{id}(x) \, dx \) be the characteristic function of an infinitely divisible (id) random variable \( X \). Then,

\[
f(\omega) = \log \hat{p}_\text{id}(\omega)
\]
4. CONTINUOUS-DOMAIN INNOVATION MODELS

is a Lévy exponent in the sense of Definition 5. Conversely, if \( f \) is a valid Lévy exponent, then the inverse Fourier integral

\[
p_{\text{id}}(x) = \int_{\mathbb{R}} e^{f(\omega)} e^{-j\omega x} \, \frac{d\omega}{2\pi}
\]

yields the pdf of an id random variable.

The proof is given in the supplementary material in Section 4.2.4. As for the latter implication, we observe that the condition \( f(0) = 0 \Leftrightarrow \hat{p}_{\text{id}}(0) = 1 \) implies that \( \int_{\mathbb{R}} p_{\text{id}}(x) \, dx = 1 \), while the positive-definiteness ensures that \( p_{\text{id}}(x) \geq 0 \) so that it is a valid pdf.

4.2.1 Canonical Lévy-Khinchine representation

The second, more explicit statement of the announced equivalence with id distributions capitalizes on the property that Lévy exponents admit a canonical representation in terms of their so-called Lévy density.

**Definition 7** (Lévy density). A nonnegative function \( v : \mathbb{R} \rightarrow \mathbb{R}^+ \) is called a Lévy density if it satisfies the admissibility condition

\[
\int_{\mathbb{R}} \min(a^2, 1) v(a) \, da < \infty. \tag{4.1}
\]

As in the case of a pdf, the density \( v \) is implicitly linked to a measure which allows for the inclusion of isolated Dirac impulses (discrete part of the measure).

**Theorem 7** (Lévy-Khinchine). A probability distribution \( p_{\text{id}} \) is infinitely divisible (id) if and only if its characteristic function can be written as

\[
\hat{p}_{\text{id}}(\omega) = \int_{\mathbb{R}} p_{\text{id}}(x) e^{j\omega x} \, dx = \exp \{ f(\omega) \} \tag{4.2}
\]

with

\[
f(\omega) = b_1^2 \omega^2 \frac{\omega^2}{2} + \int_{\mathbb{R} \setminus \{0\}} \left( e^{i\omega a} - 1 - j\omega \mathbb{1}_{|a|<1}(a) \right) v(a) \, da \tag{4.3}
\]

where \( b_1 \in \mathbb{R} \) and \( b_2 \in \mathbb{R}^+ \) are some arbitrary constants, and where \( v \) is an admissible Lévy density; \( \mathbb{1}_{|a|<1}(a) \) is the indicator function of the set \( \Omega = \{ a \in \mathbb{R} : |a| < 1 \} \) (i.e., \( \mathbb{1}_{\Omega}(a) = 1 \) if \( a \in \Omega \) and \( \mathbb{1}_{\Omega}(a) = 0 \) otherwise).

The admissibility condition (4.1) guarantees that the right-hand-side integral in (4.3) is well-defined; this follows from the bounds \( |\exp \{ i\omega a \} - 1| < a^2 \omega^2 \) and \( |\exp \{ i\omega a \} - 1| < \min(|\omega a|, 2) \). An important aspect of the theory is that this allows for (non-integrable) Lévy densities with a singular behavior around the origin; for instance, \( v(a) = O(1/|a|^2+\epsilon) \) with \( \epsilon \in (0,1) \) as \( a \to 0 \).

The connection with Theorem 6 is that the Lévy-Khinchine expansion (4.3) provides a complete characterization of the conditionally positive-definite functions of order one, as specified in Definition 5. This theme is further developed in Appendix B, which contains

---

1. In most mathematical texts, the Lévy-Khinchine decomposition is formulated in terms of a Lévy measure rather than a Lévy density. Even though Lévy measures need not always have a density in the sense of the Radon-Nikodym derivative with respect to the Lebesgue measure (i.e., as an ordinary function), following Bourbaki we may still identify them with positive linear functionals, which we represent notationally as integrals against a "generalized" density: \( V(E) = \int_{\mathbb{R}} v(a) \, da \) for any set in the Borel algebra on \( \mathbb{R} \setminus \{0\} \).
4.2. Lévy exponents and infinitely divisible distributions

the proof of the above statement and also makes interesting links with theoretical results that are fundamental to machine learning and approximation theory.

In Section 4.4, we shall indicate how id distributions (or, equivalently, Lévy exponents) can be used to specify an extended family of continuous-domain white-noise processes. In that context, we shall typically require that \( p_\text{id} \) has a well-defined first-order absolute moment and/or that it is symmetric with respect to the origin, which leads to the following simplifications of the canonical representation.

**Corollary 2.** Let \( p_\text{id} \) be an infinitely divisible pdf whose characteristic function is given by \( \hat{p}_\text{id}(\omega) = e^{f(\omega)} \). Then, depending on the properties of \( p_\text{id} \) (or, equivalently, on the Lévy measure \( \nu \)), the Lévy exponent \( f \) admits the following Lévy-Khintchine-type representations:

1. \( p_\text{id} \) id symmetric (i.e., \( p_\text{id}(x) = p_\text{id}(-x) \)) if and only if
   \[
   f(\omega) = -\frac{b_2 \omega^2}{2} + \int_{\mathbb{R}} (\cos(\omega a) - 1) \nu(a) \, da
   \]  
   with \( \nu(a) = \nu(-a) \).

2. \( p_\text{id} \) id with \( \int_{\mathbb{R}} |x| p_\text{id}(x) \, dx < \infty \) if and only if
   \[
   f(\omega) = \int_{\mathbb{R}} \left( e^{i\omega a} - 1 - i\omega a \right) \nu(a) \, da
   \]  
   with \( \int_{|a| > 1} |a| \nu(a) \, da < \infty \).

3. \( p_\text{id} \) id with \( \int_{\mathbb{R}_+} |a| \nu(a) \, da < \infty \) if and only if
   \[
   f(\omega) = \int_{\mathbb{R}} \left( e^{i\omega a} - 1 - i\omega a \right) \nu(a) \, da
   \]  
   where \( b_1 \in \mathbb{R}, b_2 \in \mathbb{R}^+, b_1' = b_1 - \int_{\mathbb{R}_+} a \, \nu(a) \, da \) and \( \nu(a) \geq 0 \) is an admissible Lévy density.

These are obtained by direct manipulation of (4.3) with \( b_1' = b_1 + \int_{|a| > 1} a \, \nu(a) \, da \) Equation (4.4) is valid in all generality, provided that we interpret the integral as a Cauchy principal-value limit to handle potential (symmetric) singularities around the origin. The Lévy-Khintchine formulas (4.4) and (4.5) are fundamental because they give an explicit, constructive characterization of the noise functionals that are central to our formulation. From now on, we rely on Corollary 2 to specify admissible Lévy exponents: The parameters \( (b_1, b_2, v) \) will be referred to as the Lévy triplet of \( f(\omega) \).

Below is a summary of known criteria for identifying admissible Lévy exponents, some being more operational than others [GV64, pp. 275-282]. These are all consequences of Bochner’s theorem, which provides a Fourier-domain equivalence between continuous, positive-definite functions and probability density functions (or positive Borel measures). See Appendix B for an overview and discussion of the functional notion of positive definiteness and corresponding mathematical tools.

**Proposition 2.** The following statements on \( f \) are equivalent:

1. \( f(\omega) \) is a continuous, conditionally positive-definite function of order one.
2. \( p_\text{id}(x) = \mathcal{F}^{-1} e^{f(\omega)}(x) \) is an infinitely divisible distribution.
3. \( f \) admits a Lévy-Khintchine representation as in Theorem 7.
4. Let \( p_{X_\tau}(x) = \mathcal{F}^{-1} e^{f(\omega)} \) for \( \tau \geq 0 \). Then, \( \{p_{X_\tau}\}_{\tau \in \mathbb{R}_+} \) is a family of valid pdfs; that is, \( p_{X_\tau}(x) \geq 0 \) and \( \int_{\mathbb{R}} p_{X_\tau}(x) \, dx = 1 \) for all \( \tau \geq 0 \).
5. \( \hat{p}_{X_\tau}(\omega) = e^{f(\omega)} \) is a continuous, positive-definite function of \( \omega \in \mathbb{R} \) with \( \hat{p}_{X_0}(0) = 1 \) for any \( \tau \in [0, \infty) \).
Interestingly, it was Schoenberg (the father of splines) who first established the equivalence between Statements 1) and 5) (see proof of the direct part in Section 4.2.4). The equivalence between 4) and 5) follows from Bochner’s theorem (Theorem 3). The fact that 2) implies 4) is a side product of the proof in Appendix 4.2.4, while the converse implication is a direct consequence of 3). Indeed, if \( f(\omega) \) has a Lévy-Khinchine representation, then the same is true for \( \tau f(\omega) \), which also implies that the whole family of pdfs \( \{p_X(\omega)\}_{\tau \in \mathbb{R}^+} \) is infinitely divisible. The latter is uniquely specified by \( f \) and therefore in one-to-one correspondence with the canonical id distribution \( p_{\text{id}}(X) = \left. p_X(x) \right|_{x=1} \). Another important observation is that \( \tilde{p}_X(\omega) = (e^{if(\omega)})^T = (\tilde{p}_{\text{id}}(\omega))^T \) so that \( p_X(\omega) \) in Statement 4) may be interpreted as the \( \tau \)-fold (possibly, fractional) convolution of \( p_{\text{id}} \).

In our work, we sometimes need to limit ourselves to some particular subset of Lévy exponents.

**Definition 8.** A Lévy exponent \( f \) with derivative \( f' \) is called \( p \)-admissible if it satisfies the inequality

\[
|f(\omega)| + |\omega| \cdot |f'(\omega)| \leq C|\omega|^p
\]

for some constant \( C > 0 \) and \( 0 < p \leq 2 \).

**Proposition 3.** The generic Lévy exponents

1. \( f_1(\omega) = \int_{\mathbb{R}\setminus\{0\}} (e^{j\omega a} - 1) v_1(a) \, da \) with \( \int_{\mathbb{R}} |a| v_1(a) \, da < \infty \)
2. \( f_2(\omega) = \int_{\mathbb{R}\setminus\{0\}} (\cos(\omega a) - 1) v_2(a) \, da \) with \( \int_{\mathbb{R}} a^2 v_2(a) \, da < \infty \)
3. \( f_3(\omega) = \int_{\mathbb{R}\setminus\{0\}} (e^{j\omega a} - 1 - ja\omega) v_3(a) \, da \) with \( \int_{\mathbb{R}} a^2 v_3(a) \, da < \infty \)

are \( p \)-admissible with \( p_1 = 1 \), \( p_2 = 2 \), and \( p_3 = 2 \), respectively.

**Proof.** The first result follows from the bounds \( |e^{j\omega a} - 1| \leq |a| \cdot |\omega| \) and \( \frac{\partial}{\partial a} \left| e^{j\omega a} - 1 \right| < |a| \). The second is based on \( |\cos(\omega a) - 1| \leq |\omega a|^2 \) and \( |\sin(\omega a)| \leq |\omega a| \). Specifically,

\[
|f_2(\omega)| \leq \int_{\mathbb{R}} |a|^2 v_2(a) \, da = |\omega|^2 \int_{\mathbb{R}} a^2 v_2(a) \, da
\]

\[
|\omega| |f_2'(\omega)| = |\omega| \left| \int_{\mathbb{R}\setminus\{0\}} a \sin(\omega a) v_2(a) \, da \right|
\]

\[
\leq |\omega| \int_{\mathbb{R}} |a| |\omega a| v_2(a) \, da = |\omega|^2 \int_{\mathbb{R}} a^2 v_2(a) \, da.
\]

As for the third exponent, we also use the inequality \( |e^{j\omega a} - 1 - ja\omega| \leq |\omega a|^2 \), which yields

\[
|f_3(\omega)| \leq \int_{\mathbb{R}\setminus\{0\}} |\omega a|^2 v_3(a) \, da = |\omega|^2 \int_{\mathbb{R}} a^2 v_3(a) \, da
\]

\[
|\omega| |f_3'(\omega)| = |\omega| \left| \int_{\mathbb{R}\setminus\{0\}} ja(e^{j\omega a} - 1) v_3(a) \, da \right|
\]

\[
\leq |\omega| \int_{\mathbb{R}} |a| |\omega a| v_3(a) \, da = |\omega|^2 \int_{\mathbb{R}} a^2 v_3(a) \, da,
\]

which completes the proof.

Since the \( p \)-admissibility property is preserved through summation, this covers a large portion of the Lévy exponents specified in Corollary 2.
4.2. Lévy exponents and infinitely divisible distributions

<table>
<thead>
<tr>
<th>Type</th>
<th>$p_A(x)$</th>
<th>$\text{Var}(X)$</th>
<th>$\beta(x) = \int p_A(x)^{\alpha-1} , dx$</th>
<th>Lévy density $\nu(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>$\frac{1}{\sqrt{2\pi x^2}} e^{-\frac{x^2}{2x^2}}$</td>
<td>$\sigma^2$</td>
<td>$e^{-\frac{\lambda^2}{2\sigma^2}} x$</td>
<td>N/A</td>
</tr>
<tr>
<td>Compound Poisson</td>
<td>$e^{-\lambda} \delta(x) + (1 - e^{-\lambda})p_{A+1}(x)$</td>
<td>$\exp\left(\lambda p_B(a) - \lambda p_B(0)\right)$</td>
<td>$\exp\left(\lambda \int_{0}^{\infty} \lambda(a) , da\right)$</td>
<td>$\lambda p_B(a) \in L_1(\mathbb{R})$</td>
</tr>
<tr>
<td>Laplace ($\lambda \in \mathbb{R}^+$)</td>
<td>$\frac{\lambda}{\pi x^2} e^{-\frac{\lambda^2}{x^2}}$</td>
<td>$\frac{\lambda^2}{\pi x^2 + \sigma^2}$</td>
<td>$e^{-\frac{\lambda x}{\sigma^2}}$</td>
<td>$\frac{\lambda x}{\sigma^2}$</td>
</tr>
<tr>
<td>Sym Gamma $\mathcal{G}$</td>
<td>$2 \Gamma \left(\frac{1}{2}\right)^{-\frac{1}{r}} K_{\frac{1}{r}}(\lambda x)$</td>
<td>$2r$</td>
<td>$\left(\frac{1}{1 + x^2}\right)^{\frac{r}{2}}$</td>
<td>$\frac{r}{2}$</td>
</tr>
<tr>
<td>Hyperbolic cosine $\mathcal{C}$</td>
<td>$\frac{1}{2 \pi x \cosh \left(\frac{\lambda x}{\sigma}\right)}$</td>
<td>$\sigma^2$</td>
<td>$\frac{1}{\cosh(\lambda x)}$</td>
<td>$\frac{1}{2 \cosh \left(\frac{\lambda x}{\sigma}\right)}$</td>
</tr>
<tr>
<td>Meixner $\mathcal{M}$</td>
<td>$2^{-\frac{1}{2} - \frac{1}{2} \beta x^2} \frac{1}{\beta(1+z)} x^{\beta z}$</td>
<td>$x^2 r$</td>
<td>$\left(\frac{1}{\cosh(\lambda x)}\right)^{\frac{r}{2}}$</td>
<td>$\frac{1}{2 \cosh \left(\frac{\lambda x}{\sigma}\right)}$</td>
</tr>
<tr>
<td>Cauchy $\mathcal{C}$</td>
<td>$\frac{1}{\pi x^2 + \sigma^2}$</td>
<td>N/A</td>
<td>$e^{-\frac{\lambda x}{\sigma^2}}$</td>
<td>$\frac{\lambda x}{\sigma^2}$</td>
</tr>
<tr>
<td>Sym Student $\mathcal{S}$</td>
<td>$\frac{1}{\pi x^2} \left(\frac{1}{x^2 + 1}\right)^{\frac{r}{2}}$</td>
<td>$r \leq 1$: N/A</td>
<td>$\frac{1}{\Gamma(r + \frac{1}{2}) \Gamma(\frac{1}{2})}$</td>
<td>unknown</td>
</tr>
<tr>
<td>$\mathcal{S}$, $a \in (0,2)$, $\mu \in \mathbb{R}^+$</td>
<td>$p_A(x; a, \mu)$</td>
<td>N/A</td>
<td>$e^{-\frac{\lambda x}{\mu} x^2}$</td>
<td>$\frac{\lambda x}{\mu}$</td>
</tr>
<tr>
<td>$\mathcal{S}$, $a \in (0,2)$, $\mu \in \mathbb{R}^+$</td>
<td>$p_A(x; a, \mu)$</td>
<td>N/A</td>
<td>$e^{-\frac{\lambda x}{\mu} x^2}$</td>
<td>$\frac{\lambda x}{\mu}$</td>
</tr>
</tbody>
</table>

Table 4.1: Primary families of symmetric, infinitely divisible distributions.

**Examples:** The power law $f_\alpha(x) = -|x|^\alpha$ with $0 < \alpha \leq 2$ is Lévy $\alpha$-admissible; it generates the symmetric $\alpha$-stable (SaS) id distributions [Fel71]. Note that $f_\alpha(x)$ fails to be conditionally positive-definite for $\alpha > 2$, meaning that the inverse Fourier transform of $e^{-|x|^\alpha}$ exhibits negative values and is therefore not a valid pdf. The upper acceptable limit is $\alpha = 2$ and corresponds to the Gaussian law. More generally, a Lévy exponent that is symmetric and twice-differentiable at the origin (which is equivalent to the variance of the corresponding id distribution being finite) is $p$-admissible with $p = 2$; this follows as a direct consequence of Corollary 2 and Proposition 3.

Another fundamental instance, which generates the complete family of id compound Poisson distributions, is

$$f_{\text{Poisson}}(\omega) = \lambda \int_{R^1(0)} (e^{i\omega a} - 1) p_A(a) \, da$$

where $\lambda > 0$ is the Poisson rate and $p_A(a) \geq 0$ the amplitude pdf with $f_{\text{Poisson}}(\omega) \, da = 1$. In general, $f_{\text{Poisson}}(\omega)$ is $p$-admissible with $p = 1$ provided that $\mathbb{E}[|a|] = \frac{\lambda}{\gamma} p_A(a) \, da < \infty$ (cf. Proposition 3). If, in addition, $p_A$ is symmetric with a bounded variance, then the Poisson range of admissibility extends to $p \in [1,2]$. Further examples of symmetric id distributions are documented in Table 4.1. Their Lévy exponent is simply obtained by taking $f(\omega) = \log \beta_X(\omega)$.

The relevance of id distributions for signal processing is that any linear combination of independent id random variables is id as well. Indeed, let $X_1$ and $X_2$ be two independent id random variables with Lévy exponents $f_1$ and $f_2$, respectively; then, it is not difficult to show that $a_1 X_1 + a_2 X_2$, where $a_1$ and $a_2$ are arbitrary constants, is id with Lévy exponent $f(\omega) = f_1(a_1 \omega) + f_2(a_2 \omega)$.  

---

An introduction to Sparse Stochastic Processes

Copyright © M. Unser and P.D. Tafti
4. CONTINUOUS-DOMAIN INNOVATION MODELS

4.2.2 Deciphering the Lévy-Khinchine formula

From a harmonic-analysis perspective, the Lévy-Khinchine representation is closely related to Bochner’s theorem stating that a positive-definite function $g$ can always be expressed as the Fourier transform of a positive finite Borel measure; i.e., $g(\omega) = \int g(x) e^{i\omega x} dx$ with $\omega = 0$ and $g(0) = \int g(x) dx < \infty$. Here, the additional requirement is that $f(0) = 0$ (conditional positive-definiteness of order one), which is enforced by proper subtraction of a linear correction in $j\omega$ within the integral, the latter being partly compensated by the addition of the component $j\omega b_1$. The side benefit of this regularization is that it enlarges the set of admissible densities to those satisfying $\int_a \min(\omega^2, 1) |\hat{v}(\omega)| d\omega < \infty$, which allows for a singular behavior around the origin. As for the linear and quadratic terms outside the integral, they map into the singular point distribution $b_1 \delta'(x) + b_2 \delta''(x)$ (weighted derivatives of Dirac impulses) that is concentrated at the origin $a = 0$ and excluded from the (classical) Lebesgue integral. For the complete details, we refer the reader to the second half of Appendix B. The proposed treatment relies on Gelfand and Vilenkin’s distributional characterization of conditionally positive-definiteness of order one in Theorem 28. Despite the greater generality of the result, we find its proof more enlightening and of lesser technical nature than the traditional derivation of the Lévy-Khintchine formula, which is summarized in Section 4.2.4-B, for completeness.

From a statistical perspective, the exponent $f$ specified by the Lévy-Khinchine formula is the logarithm of the characteristic function of an id random variable. This means that breaking $f$ into additive subparts is in fact equivalent to factorizing the pdf into convolutional factors. Specifically, let $\hat{\rho}_X(\omega) = e^{\sum_n f_n(\omega)}$ be the characteristic function of a (compound) id distribution where the $f_n$ are valid Lévy exponents. Then, $\hat{\rho}_X(\omega) = \prod_n \hat{\rho}_{X_n}(\omega)$ with $\hat{\rho}_{X_n}(\omega) = e^{f_n(\omega)} = E[e^{i\omega X_n}]$, which translates into the convolution relation $p(x) = (p_{X_1} * p_{X_2} * \cdots * p_{X_N})(x)$.

The statistical interpretation is that $X = X_1 + \cdots + X_N$ where the $X_n$ are independent with id pdf $p_{X_n}$. The infinitely divisible property simply translates into the fact that, for a given $f(\omega)$ and any $N > 0$, $X$ can be always be broken down into $N$ independent and identically distributed components with Lévy exponent $f(\omega)/N$. Indeed, it is easy to see from the Lévy representation that the admissibility of $f(\omega)$ implies that $\tau f(\omega)$ is a valid Lévy exponent as well for any $\tau \geq 0$.

To further our understanding of id distributions, it is instructive to characterize the atoms of the Lévy-Khintchine representation. Focusing on the simplest form (4.6), we identify three types of elementary constituents with the third type being motivated by the decomposition of a (continuous) Lévy density into a weighted “sum” of Dirac impulses:

$v(a) = \int_R v(\tau) \delta(a - \tau) d\tau = \sum_n \lambda_n \delta(a - \tau_n)$ with $\lambda_n = v(\tau_n)(\tau_n - \tau_{n-1})$:

1. Linear term $f_1(\omega) = b_1 \omega$. This corresponds to the (degenerate) pdf of a constant $X_1 = b_1$ with $p_{X_1}(x) = \delta(x - b_1)$.

2. Quadratic term $f_2(\omega) = -\frac{b_2 \omega^2}{2}$. As already mentioned, this leads to the centered Gaussian with variance $b_2$ given by $p_{X_2}(x) = \mathcal{N}^{-1}(e^{-\frac{b_2 x^2}{2}})(x) = \frac{1}{\sqrt{2\pi b_2}} \exp\left( -\frac{x^2}{2b_2} \right)$.

3. Exponential (or Poisson) term $f_3(\omega) = \lambda (\exp{\omega} - 1)$, which is associated with the elementary Lévy triplet $(0, 0, \lambda \delta(\omega - 1))$. Based on the Taylor-series expansion $\hat{\rho}_{X_3}(\omega) = e^{\lambda(\omega - 1)} = e^{\lambda(\omega - 1) + \frac{\lambda^2}{2} \omega^2} = e^{\lambda \omega + \frac{\lambda^2}{2} \omega^2}$ with $z = e^{\lambda \omega}$, we readily obtain the pdf by (generalized) in-
verse Fourier transformation:

\[ p_X(x) = \sum_{m=0}^{\infty} \frac{e^{-\lambda} \lambda^m}{m!} \delta(x - mt). \]

This formula coincides with the continuous-domain representation of a Poisson distribution\(^2\) with Poisson parameter \(\lambda\) and gain factor \(t\); that is, \(\text{Prob}(X_t = r m) = e^{-\lambda} \frac{\lambda^m}{m!}\). More generally, when \(v(a) = \lambda p(a)\) where \(p(a) \geq 0\) is some arbitrary pdf with \(\int_R p(a) \, da = 1\), we can make a compound-Poisson\(^3\) interpretation with

\[ f_{\text{Poisson}}(\omega) = \int_R (e^{i\alpha x} - 1) p(a) \, da = \lambda(\hat{\rho}(\omega) - 1), \]

where \(\hat{\rho}(\omega) = \int_R e^{i\alpha x} p(a) \, da\) is the characteristic function of \(p = p_A\). Using the fact that \(\hat{\rho}(\omega)\) is bounded, we apply the same type of Taylor-series argument and express the characteristic function as

\[ e^{f_{\text{Poisson}}(a)} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{(\lambda \hat{\rho}(\omega))^m}{m!} = \hat{\rho}_Y(\omega). \]

Finally, by using the property that \(\hat{\rho}(\omega)^m\) is the characteristic function of the \(m\)-fold convolution of \(p\), we get the general formula of the compound-Poisson pdf with Poisson parameter \(\lambda\) and amplitude distribution \(p\) as

\[ p_Y(x) = e^{-\lambda} \left( \delta(x) + \frac{\lambda}{1!} p(x) + \frac{\lambda^2}{2!} (p \ast p)(x) + \frac{\lambda^3}{3!} (p \ast p \ast p)(x) + \cdots \right). \quad (4.8) \]

Thus, in essence, the Lévy-Khinchine formula is a description of the Fourier transform of a distribution that is the convolution of three components: an impulse \(\delta(-b_1)\) (shifting), a Gaussian of variance \(b_2\) (smoothing), and a compound-Poisson distribution (spreading). The effect of the first term is a simple re-centering of the pdf around \(b_1\). The third compound-Poisson component is itself obtained via a suitable composition of \(m\)-fold convolutions of some primary pdf \(p\). It is much more concentrated at the origin than the Gaussian, because of the presence of the Dirac distribution with weight \(e^{-\lambda}\), but also heavier-tailed because of the spreading effect the \(m\)-fold convolution.

The additional linear correction terms in (4.3) and (4.5) allow for a wider variety of distributions that have the common property of being limits of compound-Poisson distributions.

**Proposition 4.** Every id distribution is the weak limit of a sequence of Poisson distributions.

**Proof.** Let \(\hat{\rho}_{id}\) be the characteristic function of some id distribution \(p_{id}\) and consider an arbitrary sequence \(\tau_n \downarrow 0\). Then,

\[ \hat{\rho}_{X_{\tau_n}}(\omega) = \exp \{ \tau_n^{-1}(\hat{\rho}_{id}(\omega)^{\tau_n} - 1) \}. \]

---

2. The standard form of the discrete Poisson probability model is \(\text{Prob}(N = n) = \frac{e^{-\lambda} \lambda^n}{n!}\) with \(n \in \mathbb{N}\). It provides the probability of a given number of independent events \((n)\) occurring in a fixed space/time interval when the average rate of occurrence is \(\lambda\). The Poisson parameter is equal to the expected value of \(N\), but also to its variance: \(\lambda = \text{E}(N) = \text{Var}(N)\).

3. The compound Poisson probability model has two components: The first is a random variable \(N\) that follows a Poisson distribution with parameter \(\lambda\), and the second a series of i.i.d. random variables \(A_1, A_2, A_3, \ldots\) with pdf \(p_A\) which are drawn at each trial of \(N\). Then, \(Y = \sum_{i=1}^N A_i\) is a compound-Poisson random variable with Poisson parameter \(\lambda\) and amplitude pdf \(p_A\). Its mean and variance are given by \(\text{E}(Y) = \lambda \text{E}(A)\) and \(\text{Var}(Y) = \lambda \text{Var}(A)\), respectively.
4. CONTINUOUS-DOMAIN INNOVATION MODELS

is the characteristic function of a compound-Poisson distribution with parameter $\lambda = r_n^{-1}$ and amplitude distribution $p(x) = \mathcal{F}^{-1} (\hat{p}_{\text{id}}(\omega)^{r_n})$. Moreover, we have that

$$\hat{p}_{X_n}(\omega) = \exp \left( r_n^{-1} (e^{r_n \log \hat{p}_{\text{id}}(\omega)} - 1) \right) = \exp \left( \log \hat{p}_{\text{id}}(\omega) + O(\tau_n) \right)$$

for every $\omega$ as $n \to \infty$. Hence, $\hat{p}_{X_n}(\omega) \to \exp(\log \hat{p}_{\text{id}}(\omega)) = \hat{p}_{\text{id}}(\omega)$ so that $p_{X_n}$ converges weakly to $p_{\text{id}}$ by Lévy’s continuity theorem (Theorem 4).

The id pdfs for which $\nu \in L_1(\mathbb{R})$ are generally smoother than the compound-Poisson ones for they not display a singularity (Dirac impulse) at the origin, unlike (4.8). Yet, depending of the degree of concentration (or singularity) of $\nu$ around the origin, they will typically exhibit a peaky behavior around the mean. While this class of distributions is responsible for the additional level of complication in the Lévy-Khinchine formula—as compared to the simpler Poisson version (4.6)—, we argue that it is highly relevant for applications because of the many possibilities that it offers. Somewhat surprisingly, there are many families of id distributions with singular Lévy density that are more tractable mathematically than their compound-Poisson cousins found in Table 4.1; at least, when considering their pdf.

4.2.3 Gaussian vs. sparse categorization

The family of id distributions allows for a range of behaviors that varies between the purely Gaussian and sparse extremes. In the context of Lévy processes, these are often referred to as the diffusive and jump modes. To make our point, we consider two distinct scenarios.

Finite variance case

We first assume that the second moment $m_2 = \int \sigma^2 \nu(a) \, da$ of the Lévy density is finite, which also implies that $\int_{|a|>1} |a| \nu(a) \, da < \infty$ because of the admissibility condition. Hence, the corresponding Lévy-Khinchine representation is (4.5). An interesting non-Poisson example of infinitely-divisible probability laws that falls into this category (with non-integrable $\nu$) is the Laplace density with Lévy triplet $(0, 0, \nu(a) = e^{-|a|})$ and $p(x) = \frac{1}{2} e^{-|x|}$ (see Figure 4.1b). This model is particularly relevant in the context of sparse signal recovery because it provides a tight connection between Lévy processes and total-variation regularization [UT11, Section VI].

Now, if the Lévy density is measurable (i.e., $\nu \in L_1(\mathbb{R})$), we can pull the linear correction out of the Lévy-Khinchine integral and represent $f$ using Expansion (4.6) with $\nu(a) = \lambda p_{A}(a)$ and $\int \lambda p_{A}(a) \, da = 1$. This implies that we can decompose $X$ into the sum of two independent Gaussian and compound-Poisson random variables. The variances of the Gaussian and Poisson components are $\sigma^2 = b_2$ and $\lambda \mathbb{E}[a^2]$, respectively. The Poisson component is sparse because its probability density function exhibits the mass distribution $e^{-\lambda} \delta(x)$ at the origin shown in Figure 4.1c, meaning that the chances for a continuous amplitude distribution of getting zero are overwhelmingly higher than any other value, especially for smaller values of $\lambda > 0$. It is therefore justifiable to use $0 \leq e^{-\lambda} < 1$ as our Poisson sparsity index.

Infinite variance case

We now turn our attention to the case where the second moment of the Lévy density is unbounded, which we like to label as “super-sparse”. To justify this terminology, we invoke
the Ramachandran-Wolfe theorem which states that the $p$th moment $\mathbb{E}[|x|^p]$ with $p \in \mathbb{R}^+$ of an infinitely divisible distribution is finite if and only if $\int_{|a| \geq 2} |a|^p \nu(a) \, da < \infty$ [Ram69, Wol71]. For $p \geq 2$, the latter is equivalent to $\int_{a \geq 0} |a|^p \nu(a) \, da < \infty$ because of the Lévy admissibility condition. It follows that the cases that are not covered by the previous scenario (including the Gaussian + Poisson model) necessarily give rise to distributions whose moments of order $p$ are unbounded for $p \geq 2$. The prototypical representatives of such heavy-tail distributions are the alpha-stable ones (see Figure 4.1d) or, by extension, the broad family of infinitely divisible probability laws that are in their domain of attraction. It has been shown that these distributions precisely fulfill the requirement for $\ell_p$ compressibility [AUMew], which is a stronger form of sparsity than the presence of a mass probability density at the origin.

4.2.4 Proofs of Theorems 6 and 7

For completeness, we end this section on Lévy exponents with the proofs of the two key theorems in the theory of infinitely divisible distributions. The Lévy-Schoenberg theorem is central to our formulation because it makes the link between the id property and the fundamental notion of positive definiteness. In the case of the Lévy-Khintchine theorem, we have opted for a sketch of proof which is adapted from the literature. The main intent there was to provide additional insights on the nature of the singularities of the Lévy density and their effect on the form of the exponent.

A. Proof of Theorem 6 (Lévy-Schoenberg)

Let $\hat{\rho}_{id}(\omega) = \int_{\mathbb{R}} e^{ix\omega} p_{id}(x) \, dx$ be the characteristic function of an id random variable. Then, by definition, $(\hat{\rho}_{id}(\omega))^{1/n}$ is a valid characteristic function for any $n \in \mathbb{Z}^+$. Since the convolution of two pdfs is a pdf, we can also take integer powers, which results into $(\hat{\rho}_{id}(\omega))^{m/n}$ being a characteristic function. For any irrational number $\tau > 0$, we can specify a sequence of rational numbers $m/n$ that converges to $\tau$ so that $(\hat{\rho}_{id}(\omega))^{m/n} \to (\hat{\rho}_{id}(\omega))^\tau$ with the limit function being continuous. This implies that $\hat{\rho}_{X_{\tau}}(\omega) = (\hat{\rho}_{id}(\omega))^\tau$ is a characteristic function for any $\tau \geq 0$ by Lévy’s continuity theorem (Theorem 4). Moreover, $\hat{\rho}_{X_{\tau}}(\omega) = (\hat{\rho}_{X_{1/\tau}}(\omega))^\tau$ must be non-zero for any finite $\tau$, owing to the fact that $\lim_{\tau \to \infty} \hat{\rho}_{X_{1/\tau}}(\omega) = 1$. In particular, $\hat{\rho}_{id}(\omega) = \hat{\rho}_{X_{1}}(\omega)$ at $\tau = 1$ so that we can write it as $\hat{\rho}_{id}(\omega) = e^{f(\omega)}$ where $f(\omega)$ is continuous with $\text{Re}(f(\omega)) \leq 0$ and $f(0) = 0$. Hence, $\hat{\rho}_{X_{\tau}}(\omega) = (\hat{\rho}_{id}(\omega))^\tau = e^{\tau f(\omega)} = \int_{\mathbb{R}} e^{ix\omega} p_{X_{\tau}}(x) \, dx$, where $p_{X_{\tau}}(x)$ is a valid pdf for any $\tau \in [0, \infty)$, which is Statement 4 in Proposition 2. By Bochner’s theorem (Theorem 3), this is equivalent to $e^{\tau f(\omega)}$ being positive-definite for any $\tau \geq 0$ with $f(\omega)$ continuous and $f(0) = 0$. The first part of Theorem 6 then follows as a corollary of the next fundamental result, which is due to Schoenberg.

Theorem 8 (Schoenberg correspondence). The function $f(\omega)$ is conditionally positive-definite of order one if and only if $e^{\tau f(\omega)}$ is positive-definite for any $\tau > 0$.

Proof. We only give the easy part and refer to [Sch38, Joh66] for the complete details. The property that $e^{\tau f(\omega)}$ is positive-definite for every $\tau > 0$ is expressed as

$$\sum_{m=1}^{N} \sum_{n=1}^{N} \xi_m \bar{\xi}_n e^{\tau f(\omega_m - \omega_n)} \geq 0,$$
4. Continuous-domain innovation models

for every possible choice of \(\omega_1, \ldots, \omega_N \in \mathbb{R}, \xi_1, \ldots, \xi_N \in \mathbb{C}\) and \(N \in \mathbb{Z}^+\). In the more restricted setup of Definition 5 where \(\sum_{n=1}^{N} \xi_n = 0\), this can also be restated as

\[
\frac{1}{\tau} \sum_{m=1}^{N} \sum_{n=1}^{N} \xi_m \xi_n (e^{\tau f(\omega_m - \omega_n)} - 1) \geq 0.
\]

The next step is to take the limit

\[
\lim_{\tau \to 0} \sum_{m=1}^{N} \sum_{n=1}^{N} \xi_m \xi_n (e^{\tau f(\omega_m - \omega_n)} - 1) - \frac{1}{\tau} = N \sum_{m=1}^{N} \sum_{n=1}^{N} \xi_m \xi_n f(\omega_m - \omega_n) \geq 0,
\]

which implies that \(f(\omega)\) is conditionally positive-definite of order one. \(\square\)

This also makes the second part of Theorem 6 easy because \(\hat{p}_{\text{id}}(\omega) = e^{f(\omega)}\) can be factorized into a product of \(N\) identical positive-definite subparts with Lévy exponent \(\frac{1}{N} f(\omega)\).

B. Sketch of proof of Theorem 7 (Lévy-Khinchine)

We start from the equivalence of the id property and Statement 4) in Proposition 2 established above. This result is restated as

\[
e^{\tau f(\omega)} - 1 = \int_{\mathbb{R}} (e^{i\omega x} - 1) \frac{p_X(x)}{\tau} \, dx,
\]

the limit of which as \(\tau \to 0\) exists and is equal to \(f(\omega)\). Next, we define the measure

\[K_\tau(\, dt) = \frac{x^2}{x^2 + 1} \frac{p_X(x)}{\tau} \, dx,\]

which is bounded for all \(\tau > 0\) because \(\frac{x^2}{x^2 + 1} \leq 1\) and \(p_X\) is a valid pdf. We then express the Lévy exponent as

\[
f(\omega) = \lim_{\tau \to 0} \left( \frac{e^{\tau f(\omega)} - 1}{\tau} \right) = \lim_{\tau \to 0} \int_{\mathbb{R}} \left( e^{i\omega x} - 1 \right) \frac{p_X(x)}{\tau} \, dx
\]

\[
= \lim_{\tau \to 0} \int_{\mathbb{R}} \left( e^{i\omega x} - 1 - \frac{ix\omega}{1 + x^2} \right) \frac{x^2 + 1}{x^2} K_\tau(\, dx) + i\omega \lim_{\tau \to 0} \alpha(\tau)
\]

where

\[
\alpha(\tau) = \int_{\mathbb{R}} \frac{x}{1 + x^2} \frac{p_X(x)}{\tau} \, dx.
\]

The technical part of the work, which is quite tedious and not included here, is to show that the above integrals are bounded and that the two limits are well defined in the sense that \(a(\tau) \to a_0\) and \(K_\tau \to K\) (weakly) as \(\tau \to 0\) with \(K\) being a finite measure. This ultimately yields Khinchine's canonical representation

\[
f(\omega) = i\omega a_0 + \int_{\mathbb{R}} \left( e^{i\omega x} - 1 - \frac{ix\omega}{1 + x^2} \right) \frac{x^2 + 1}{x^2} K(\, dx)
\]

where \(a_0 \in \mathbb{R}\) and \(K\) is some bounded Borel measure. A potential advantage of Khinchine's representation is that the corresponding measure \(K\) is not singular. The connection with the standard Lévy-Khinchine formula is \(b_2 = K(0^+ - K(0^-)\) and \(v(x) \, dx = \frac{x^2}{x^2 + 1} K(\, dx)\) for \(x \neq 0\). It is also possible to work out a relation between \(a\) and \(b_1\), which depends upon the type of linear compensation in the canonical representation.

The above manipulation shows that the coefficients of the linear and quadratic terms of the Lévy-Khinchine formula (4.3) are primarily due to the non-integrable part of \(g(x) = \lim_{\tau \to 0} \frac{p_X(x)}{\tau} = \frac{x^2 + 1}{x^2} k(x)\) where \(k(x) \, dx = K(\, dx)\).
4.3 Finite-dimensional innovation model

By convention, the classical Lévy density $\nu$ is assumed to be zero at the origin so that it differs from $g$ by a point distribution that is concentrated at the origin. By invoking a basic theorem in distribution theory stating that a distribution entirely localized at the origin can always be expressed as a linear combination of the Dirac impulse and its derivatives, we can write that $g(x) - \nu(x) = b_0\delta(x) + b_1\delta'(x) + b_2\delta''(x)$, where the higher-order derivatives of $\delta$ are excluded because of the admissibility condition.

For the indirect part of the proof, we start from the integral of the Lévy-Khinchine formula and consider the sequence of distributions whose exponent is

$$f_n(\omega) = \int_{|\omega|<1/n} e^{i\omega u} - 1 - j\omega \nu_1(\omega) \, d\omega - j\omega \int_{1/n<|\omega|<1} \nu(\omega) \, d\omega + \int_{|\omega|>1/n} e^{i\omega u} - 1 \, d\omega.$$

Since the leading constant $a_n$ is finite and $\int_{|\omega|>1/n} \nu(\omega) \, d\omega < \infty$ for any fixed $n$ (due to the admissibility condition on $\nu$), this corresponds to the exponent of a shifted compound-Poisson distribution whose characteristic function is $e^{f_n(\omega)}$. This allows us to deduce that $\hat{\nu}_n(\omega) = e^{-b_1j\omega - \frac{1}{2}a_2\omega^2 + f_n(\omega)}$ is a valid characteristic function for any $n \in \mathbb{Z}^+$. Finally, we have the convergence of the sequence $\hat{\nu}_n(\omega) \to e^{f(\omega)}$ as $n \to \infty$ where $f(\omega)$ is given by (4.3). Since $f(\omega)$—and therefore, $e^{f(\omega)}$—is continuous around $\omega = 0$, we infer that $e^{f(\omega)}$ is a valid pdf (by Lévy’s continuity theorem). The continuity of $f$ is established by bounding the Lévy-Khinchine integral and invoking Lebesgue’s dominated convergence theorem. The id part is obvious.

4.3 Finite-dimensional innovation model

To set the stage for the infinite-dimensional extension to come in Section 4.4, it is instructive to investigate the structure of a purely discrete innovation model whose input is the random vector $U = (U_1, \ldots, U_N)$ of i.i.d. infinitely divisible random variables. The generic $N$th-order pdf of the discrete innovation variable $U$ is

$$p_U(u_1, \ldots, u_N) = \prod_{n=1}^{N} p_{\text{id}}(u_n)$$

(4.9)

where $p_{\text{id}}(x) = \mathcal{F}^{-1}\{e^{f(\omega)}\}(x)$ and $f$ is the Lévy exponent of the underlying id distribution. Since $p_U(u_1, \ldots, u_N)$ is separable due to the independence assumption, we can write its characteristic function as the product of individual id factors

$$\hat{p}_U(\omega) = E_U[e^{i\omega U}] = \prod_{n=1}^{N} e^{f(\omega_n)}$$

(4.10)

where $\omega = (\omega_1, \ldots, \omega_N)$ is the frequency variable. The $N$-dimensional output signal $X = (X_1, \ldots, X_N)$ is then specified as the solution of the matrix-vector innovation equation

$$W = LX$$

where the $N \times N$ whitening matrix $L$ is assumed to be invertible. This implies that $X = AU$ is a linear transformation of the excitation noise with $A = L^{-1}$. Its $N$th-order characteristic
function is obtained by simple (linear) change of variable
\[
\hat{p}_X(\omega) = \mathbb{E}_{U} \{ e^{i \omega U^T} \} = \mathbb{E}_{U} \{ e^{i \lambda U^T} \}
\]
\[
= \hat{p}_U(\lambda)
\]
\[
= \exp \left( \sum_{n=1}^{N} f \left( |\lambda^T|_n \right) \right). \tag{4.11}
\]

Based on this equation, we can determine any marginal distribution by setting the appropriate frequency variables to zero. For instance, we find that the first-order pdf of \(X_n\), the \(n\)th component of \(X\), is given by
\[
p_{X_n}(x) = F^{-1}_X \left\{ e^{f_n(x)} \right\}(x)
\]
where
\[
f_n(\omega) = \sum_{m=1}^{N} f \left( a_{nm} \omega \right)
\]
with weighting coefficients \(a_{nm} = |A|_{nm} = [L^{-1}]_{mn}\). The key observation here is that \(f_n\) is an admissible Lévy exponent, which implies that the underlying distribution is infinitely divisible (by Theorem 6), with the same being true for all the marginals and, by extension, the distribution of any linear measurement(s) of \(X\). While this provides a general mechanism for characterizing the probability law(s) of the discrete signal \(X\) within the classical framework of multivariate statistics, it is a priori difficult to perform the required computations (matrix inverse and inverse Fourier transforms) analytically, except in the Gaussian case where the exponent is quadratic. Indeed, in this latter situation, (4.11) simplifies to
\[
\hat{p}_X(\omega) = e^{-\frac{1}{2} |\lambda^T|_2^2},
\]
which is the Fourier transform of a multivariate Gaussian distribution with zero mean and covariance matrix \(\mathbb{E}\{XX^T\} = AA^T\).

As we shall see in the next two sections, these results are transposable to the infinite-dimensional setting (cf. Table 3.4). While this may look as an unnecessary complication at first sight, the payoff is a theory that lends itself better to an analytical treatment using the powerful tools of harmonic analysis. The essence of the generalization is to replace the frequency variable \(\omega\) by a generic test function \(\varphi \in \mathcal{S}(\mathbb{R}^d)\), the sums in (4.10) and (4.11) by Lebesgue integrals and the matrix inverses by Green's functions which can often be specified explicitly. To make an analogy, it is conceptually and practically easier to formulate a comprehensive (deterministic) theory of linear systems using Fourier analysis and convolution operators than by relying on linear algebra, with the same applying here. At the end of the exercise, it is still possible to come back to an finite-dimensional signal representation by projecting the continuous-domain model onto a suitable set of basis functions, as will be shown in Chapter 10.

### 4.4 White Lévy noises or innovations

Having gained a solid understanding of Lévy exponents, we can now move to the specification of a corresponding family of continuous-domain white-noise processes to drive the innovation model in Figure 2.1. To that end, we rely of Gelfand’s theory of generalized stochastic processes [GV64], which was briefly summarized in Section 3.5. This powerful formalism allows for the complete and remarkably concise description of a generalized stochastic process by its characteristic form. While the latter is not widely used in the standard formulation of stochastic processes, it lends itself quite naturally to the specification of generalized white-noise processes in terms of Lévy exponents, in direct analogy with what we have done before for id distributions.
4.4. White Lévy noises or innovations

**Definition 9** (White Lévy noise or innovation). A generalized stochastic process \( w \) over \( \mathcal{D}'(\mathbb{R}^d) \) is called a white Lévy noise (or innovation) if its characteristic form is given by

\[
\mathcal{P}_w(\varphi) = \mathbb{E}[e^{i\varphi(w,\cdot)}] = \exp\left(\int_{\mathbb{R}^d} f(\varphi(r)) \, dr\right)
\]  

(4.12)

where \( f \) is a valid Lévy exponent and \( \varphi \) is a generic test function in \( \mathcal{D}(\mathbb{R}^d) \) (the space of infinitely differentiable functions of compact support).

Equation (4.12) is very similar to (4.2) and its multivariate extension (4.10). The key difference is that the frequency variable is now replaced by the generic test function \( \varphi \in \mathcal{D}(\mathbb{R}^d) \) (which is a more general infinite-dimensional entity) and that the sum inside the exponential in (4.10) is substituted by an integral over the domain of \( \varphi \). The fundamental point is that \( \mathcal{P}_w(\varphi) \) is a continuous, positive-definite functional on \( \mathcal{D}(\mathbb{R}^d) \) with the key property that \( \mathcal{P}_w(\varphi_1 + \varphi_2) = \mathcal{P}_w(\varphi_1) \mathcal{P}_w(\varphi_2) \) whenever \( \varphi_1 \) and \( \varphi_2 \) have non-overlapping support (i.e., \( \varphi_1(r)\varphi_2(r) = 0 \)). The first part of the statement ensures that these generalized processes are well-defined (by the Minlos-Bochner theorem), while the separability property implies that they take independent values at all points, which partially justifies the “white noise” nomenclature. Remarkably, Gelfand and Vilenkin have shown that there is also a converse implication [GV64, Theorem 6, p. 283]: Equation (4.12) specifies a continuous, positive-definite functional on \( \mathcal{D} \) (and hence defines an admissible white-noise process) if and only if \( f \) is a Lévy exponent. This ensures that the Lévy family constitutes the broadest possible class of acceptable white-noise inputs for our innovation model.

4.4.1 Specification of white noise over Schwartz’ space \( \mathcal{S} \)

In the present work, which relies a lot on convolution operators and Fourier analysis, we find it more convenient to define generalized stochastic processes with respect to test functions in the nuclear space \( \mathcal{S}'(\mathbb{R}^d) \), rather than the smaller space \( \mathcal{D}'(\mathbb{R}^d) \) used by Gelfand and Vilenkin. This requires a minimal restriction on the class of admissible Lévy densities in reference to Definition 7 to compensate for the lack of compact support of the functions in \( \mathcal{S}'(\mathbb{R}^d) \).

**Theorem 9.** A white Lévy noise specified by (4.12) with \( \varphi \in \mathcal{S}(\mathbb{R}^d) \) is a generalized stochastic process over \( \mathcal{S}'(\mathbb{R}^d) \) provided that \( f(\cdot) \) is characterized by the Lévy-Khintchine formula (4.3) with Lévy triplet \( (\alpha, \sigma, \nu) \) where the Lévy density \( \nu(a) \geq 0 \) satisfies

\[
\int_{\mathbb{R}} \min\{a^2, |a|^\epsilon\} \nu(a) \, da < \infty \quad \text{for some} \quad \epsilon > 0.
\]  

(4.13)

**Proof.** The material in this section is currently under review. It is available from the authors upon request.
4. Continuous-domain innovation models

Note that (4.13), which will be referred to as Lévy-Schwartz admissibility, is a very slight restriction on the classical condition ($c = 0$) for id laws (see (4.1) in Definition 7). The fact that $c$ can be chosen arbitrarily small reflects the property that the functions in $\mathcal{F}$ have a faster-than-algebraic decay. From now on, we implicitly assume that the Lévy-Schwartz admissibility condition is met.

To exemplify the procedure, we select a quadratic exponent which is trivially admissible (since $\nu(a) = 0$). This results in

$$\hat{w}_{\text{Gauss}}(\varphi) = \exp \left( -\frac{b_2}{2} \| \varphi \|_{L_2}^2 \right),$$

which is the functional that completely characterizes the white Gaussian noise of the classical theory of stationary processes.

4.4.2 Impulsive Poisson noise

We have already alluded to the fact that the continuous-domain white-noise processes $w$ are highly singular and generally too rough to admit an interpretation as conventional functions of the index variable $r \in \mathbb{R}^d$. The realizations (or sample paths) are generalized functions that can only be probed indirectly through their scalar products $\langle \varphi, w \rangle$ with test functions or observation windows, as illustrated in Section 4.1. While the use of such
4.4. White Lévy noises or innovations

an indirect approach is unavoidable in the mathematical formulation, it is possible to provide an explicit pointwise description of a noise realization in the special case where the Lévy exponent $f$ is associated with a compound-Poisson distribution [UT11]. The corresponding impulsive Poisson noise model is

$$w(r) = \sum_{k \in \mathbb{Z}} A_k \delta(r - r_k)$$  \hspace{1cm} (4.14)

where $r_k$ are random point locations in $\mathbb{R}^d$ and where the $A_k$ are i.i.d. random variables with pdf $p_A$. The random events are indexed by $k$ (using some arbitrary ordering); they are mutually independent and follow a spatial Poisson distribution. Specifically, let $\Pi$ be any finite-measure subset of $\mathbb{R}^d$, then the probability of observing $N(\Pi) = n$ events in $\Pi$ is

$$\text{Prob}(N(\Pi) = n) = \frac{e^{-\lambda \text{Vol}(\Pi)} (\lambda \text{Vol}(\Pi))^n}{n!}$$

where $\text{Vol}(\Pi)$ is the measure (or spatial volume) of $\Pi$. This is to say that the Poisson parameter $\lambda$ represents the average number of random impulses per unit hyper-volume. The link with the formal specification of Lévy noise in Definition 9 is as follows.

**Theorem 10.** The characteristic form of the impulsive Poisson noise specified by (4.14) is

$$\hat{F}_{w\text{Poisson}}(\varphi) = \mathbb{E}[e^{i\varphi(w)}] = \exp \left( \int_{\mathbb{R}^d} f_{\text{Poisson}}(\varphi(r)) \, dr \right)$$  \hspace{1cm} (4.15)

with

$$f_{\text{Poisson}}(\omega) = \lambda \int_{\mathbb{R}} (e^{ia\omega} - 1)p_A(a) \, da = \lambda (\hat{p}_A(\omega) - 1),$$  \hspace{1cm} (4.16)

where $\lambda$ is the Poisson density parameter, $p_A$ the amplitude pdf of the Dirac impulses and $\hat{p}_A$ the corresponding characteristic function.

**Proof.** We select an arbitrary test function $\varphi \in \mathcal{D}(\mathbb{R}^d)$ of compact support, with its support included in the centered cube $\Pi_{\varphi} = [-c_{\varphi}, c_{\varphi}]^d$. We denote by $N_{w,\varphi}$ the number of Poisson points of $w$ in $\Pi_{\varphi}$; by definition, it is a Poisson random variable with parameter $A\text{Vol}(\Pi_{\varphi})$. The restriction of $w$ to $\Pi_{\varphi}$ corresponds to the random sum

$$\sum_{n=1}^{N_{w,\varphi}} d_n \delta(r - r_n'),$$

where we used an appropriate relabeling of the variables $\{(a_k, r_k) | r_k \in \Pi_{\varphi}\}$ in (4.14). Correspondingly, we have $\langle \varphi, w \rangle = \sum_{n=1}^{N_{w,\varphi}} d_n' \varphi(r_n')$.

By the order-statistics property of Poisson processes, the $r_n'$ are independent and all equivalent in distribution to a random variable $r'$ that is uniform on $\Pi_{\varphi}$.

Using the law of total expectation, we expand the characteristic functional of $w$, $\hat{F}_w(\varphi) = \cdots$
4. CONTINUOUS-DOMAIN INNOVATION MODELS

$$\mathcal{E}\{e^{i\langle \varphi, w \rangle}\},$$
as

$$\mathcal{T}_w(\varphi) = \mathbb{E}\{\mathbb{E}\{e^{i\langle \varphi, w \rangle} | N_{w, \varphi}\}\}$$

$$= \mathbb{E}\left\{ \prod_{n=1}^{N_{w, \varphi}} e^{i\varphi(r_n)} | N_{w, \varphi}\right\}$$

$$= \mathbb{E}\left\{ \prod_{n=1}^{N_{w, \varphi}} e^{i\varphi(r_n)} \right\}$$

(by independence)

$$= \mathbb{E}\left\{ \prod_{n=1}^{N_{w, \varphi}} e^{i\varphi(r_n)} \right\}$$

(total expectation)

$$= \mathbb{E}\left\{ \prod_{n=1}^{N_{w, \varphi}} \frac{\int_{\Pi_{\varphi}} e^{i\varphi(r')} dr'}{\text{Vol}(\Pi_{\varphi})} \right\}$$

(as $r'$ is uniform in $\Pi_{\varphi}$)

$$= \mathbb{E}\left\{ \prod_{n=1}^{N_{w, \varphi}} \frac{\int_{\Pi_{\varphi}} e^{i\varphi(r')} dr'}{\text{Vol}(\Pi_{\varphi})} \right\}.$$  

(4.17)

The last expression has the inner expectation expanded in terms of the pdf $p_A(a)$ of the random variable $a$. Defining the auxiliary functional

$$M(\varphi) = \int_{\Pi_{\varphi}} \int_{\mathbb{R}} e^{i\varphi(r')} p_A(a) da dr,$$

we rewrite (4.17) as

$$\mathbb{E}\left\{ \prod_{n=1}^{N_{w, \varphi}} \frac{M(\varphi)}{\text{Vol}(\Pi_{\varphi})} \right\} = \mathbb{E}\left\{ \left( \frac{M(\varphi)}{\text{Vol}(\Pi_{\varphi})} \right)^{N_{w, \varphi}} \right\}.$$  

Next, we use the fact that $N_{w, \varphi}$ is a Poisson random variable to compute the above expectation directly:

$$\mathbb{E}\left\{ \left( \frac{M(\varphi)}{\text{Vol}(\Pi_{\varphi})} \right)^{N_{w, \varphi}} \right\} = \sum_{n=0}^{\infty} \left( \frac{M(\varphi)}{\text{Vol}(\Pi_{\varphi})} \right)^n \frac{e^{-\text{Vol}(\Pi_{\varphi})} (\lambda \text{Vol}(\Pi_{\varphi}))^n}{n!}$$

$$= e^{-\text{Vol}(\Pi_{\varphi})} \sum_{n=0}^{\infty} \frac{(\lambda M(\varphi))^n}{n!}$$

(Taylor)

$$= e^{-\text{Vol}(\Pi_{\varphi})} e^{\lambda M(\varphi)}$$

We now replace $M(\varphi)$ by its integral equivalent, noting also that $\text{Vol}(\Pi_{\varphi}) = \int_{\Pi_{\varphi}} \int_{\mathbb{R}} 1 \times p_A(a) da dr$, whereupon we obtain the expression

$$\mathcal{T}_w(\varphi) = \exp\left( \lambda \int_{\Pi_{\varphi}} \int_{\mathbb{R}} (e^{i\varphi(r')} - 1) p_A(a) da dr \right).$$

As $(e^{i\varphi(r')} - 1)$ vanishes outside the support of $\varphi$ (and, therefore, outside $\Pi_{\varphi}$), we may enlarge the domain of the inner integral to all of $\mathbb{R}^d$, which yields (4.15). Finally, we use the fact that the derived Poisson functional is part of the Lévy family and invoke Theorem 9 to extend the domain of $\mathcal{T}_w$ from $\mathcal{D}(\mathbb{R}^d)$ to $\mathcal{F}(\mathbb{R}^d)$.

The interest of this result is twofold. First, it gives a concrete meaning to the compound Poisson scenario in Figure 4.1c, allowing for a description in terms of conventional point processes. Along the same vein, we can propose a physical analogy for the elementary
4.4. White Lévy noises or innovations

Poisson term \( f_3(\omega) = \lambda(e^{\lambda a_0} - 1) \) in Section 4.2.2 with \( p_3(a) = \delta(a - a_0) \): the counting of photons impinging on the detectors of a CCD camera with the photon density being constant over \( \mathbb{R}^d \) and the integration time proportional to \( \lambda \). The corresponding process is usually termed "photon noise" in optical imaging. Second, the explicit noise model (4.14) suggests a practical mechanism for generating generalized Poisson processes as a weighted sum of shifted Green functions of \( L \), each Dirac impulse being replaced by the response of the inverse operator in the innovation model in Figure 2.1.

4.4.3 Properties of white noise

To emphasize the parallel with the scalar formulation in Section 4.2, we start by introducing the functional counterpart of Definition 5.

**Definition 10 (Generalized Lévy exponent).** A continuous complex-valued functional \( F \) on the nuclear space \( \mathcal{S}(\mathbb{R}^d) \) such that \( F(0) = 0 \), \( F(\varphi) = \overline{F(-\varphi)} \) is called a generalized Lévy exponent if it is conditionally positive-definite of order one; i.e.,

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} F(\varphi_m - \varphi_n) \xi_m \overline{\xi_n} \geq 0,
\]

under the condition \( \sum_{n=1}^{N} \xi_n = 0 \) for every possible choice \( \varphi_1, \ldots, \varphi_N \in \mathcal{S}(\mathbb{R}^d) \), \( \xi_1, \ldots, \xi_N \in \mathbb{C} \), and \( N \in \mathbb{Z}^+ \).

This definition is motivated by the infinite-dimensional counterpart of Schoenberg's correspondence theorem (Theorem 8) [PR70].

**Theorem 11 (Prakasa Rao).** Let \( F \) be a complex-valued functional on the nuclear space \( \mathcal{S}(\mathbb{R}^d) \) such that \( F(0) = 0 \), \( F(\varphi) = \overline{F(-\varphi)} \). Then, the following conditions are equivalent.

1. The functional \( F \) is continuously positive definite of order one.
2. For every choice \( \varphi_1, \ldots, \varphi_N \in \mathcal{S}(\mathbb{R}^d) \), \( \xi_1, \ldots, \xi_N \in \mathbb{C} \), \( \tau > 0 \) and \( N \in \mathbb{Z}^+ \),

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} \exp(\tau F(\varphi_m - \varphi_n)) \xi_m \overline{\xi_n} \geq 0.
\]

3. For every choice \( \varphi_1, \ldots, \varphi_N \in \mathcal{S}(\mathbb{R}^d) \), \( \xi_1, \ldots, \xi_N \in \mathbb{C} \) and \( N \in \mathbb{Z}^+ \),

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} (F(\varphi_m - \varphi_n) - F(\varphi_m) - F(-\varphi_n)) \xi_m \overline{\xi_n} \geq 0.
\]

In the white Lévy noise scenario of Definition 9, we have that

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

(4.18)

It then comes as no surprise that the generalized Lévy exponent \( F(\varphi) \) inherits the relevant properties of \( f \), including conditional positive definiteness, with

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} F(\varphi_m - \varphi_n) \xi_m \overline{\xi_n} = \int_{\mathbb{R}^d} \sum_{m=1}^{N} \sum_{n=1}^{N} f(\varphi_m(r) - \varphi_n(r)) \xi_m \overline{\xi_n} \, dr \geq 0
\]

subject to the constraint \( \sum_{n=1}^{N} \xi_n = 0 \).

The simple additive nature of the mapping (4.18) between generalized Lévy exponents and the classical ones translates into the following white-noise properties which are central to our formulation.
4. CONTINUOUS-DOMAIN INNOVATION MODELS

1) Independent atoms and stationarity

**Proposition 5.** A white Lévy noise is stationary and independent at every point.

**Proof.** The stationarity property is expressed by \( \mathcal{F}_w(\varphi) = \mathcal{F}_w(\varphi(-r_0)) \) for all \( r_0 \in \mathbb{R}^d \).
It is established by simple change of variable in the defining integral. To investigate
the independence at every point, we determine the joint characteristic function of the
random variables \( X_1 = \langle \varphi_1, w \rangle \) and \( X_2 = \langle \varphi_2, w \rangle \), which is given by
\[
\hat{\rho}_{X_1,X_2}(\omega_1,\omega_2) = \exp \{ F(\omega_1\varphi_1 + \omega_2\varphi_2) \}
\]
where \( F \) is defined by (4.18). When \( \varphi_1 \) and \( \varphi_2 \) have non-overlapping
support, we use the fact that \( f(0) = 0 \) and decompose the exponent as
\[
F(\omega_1\varphi_1 + \omega_2\varphi_2) = F(\omega_1\varphi_1) + F(\omega_2\varphi_2),
\]
which implies that
\[
\hat{\rho}_{X_1,X_2}(\omega_1,\omega_2) = \hat{\rho}_{X_1}(\omega_1) \times \hat{\rho}_{X_2}(\omega_2)
\]
where \( \hat{\rho}_X(\omega) = \exp \{ F(\omega\varphi) \} \), which proves that \( X_1 \) and \( X_2 \) are independent. The inde-
pendence at every point follows from the fact that one can consider contracting, Dirac-
like sequences of functions \( \varphi_1 \) and \( \varphi_2 \) that are non-overlapping and whose support gets
arbitrarily small. \( \square \)

2) Infinite divisibility

**Proposition 6.** A white Lévy noise is uniquely specified by a canonical id distribution
\( p_{id}(x) = \int_{\mathbb{R}} e^{i\omega x} \frac{d\mathcal{F}_w(\omega)}{d\omega} \) where \( f \) is the defining Lévy exponent in (4.18). The latter corre-
responds to the pdf of the observation \( X = \langle \text{rect}(\cdot - r_0), w \rangle \) through a rectangular window
at some arbitrary location \( r_0 \).

The first part is just a restatement of the functional equivalence between Lévy noises and
id distributions on the one hand, and Lévy exponents on the other. As for the second part,
we recall that the characteristic function of the variable \( X = \langle \varphi, w \rangle \) is given by
\[
E[e^{i\omega X}] = E[e^{i\omega\varphi}\exp(\varphi w)] = E[e^{i\omega\varphi}] = \mathcal{F}_w(\varphi) = \exp \{ F(\omega\varphi) \},
\]
By choosing \( \varphi(r) = \text{rect}(r - r_0) \) with \( \text{rect}(r) = 1 \) for \( r \in \left(-\frac{1}{2},\frac{1}{2}\right)^d \) and zero otherwise,
we formally resolve the integral \( \int_{\mathbb{R}} f(\omega \varphi) \text{rect}(r) \, dr = f(\omega) \); this implies that \( p_{id}(\omega) = \exp \{ f(\omega) \} \), which is the desired result.
Along the same line, we can show that the use of an arbitrary, non-rectangular analysis
window does not fundamentally change the situation in the sense that the pdf
infinitely divisible.

**Proposition 7.** The observation \( X = \langle \varphi, w \rangle \) of a white Lévy noise with Lévy exponent \( f \)
through an arbitrary observation window \( \varphi \)—not necessarily in \( \mathcal{S}(\mathbb{R}^d) \)—yields an infi-
nitely divisible random variable \( X \) whose characteristic function is
\( \hat{\rho}_{X}(\omega) = e^{\int f(\omega r) \, dr} \) where \( f_{\varphi}(\omega) = \int_{\mathbb{R}^d} f(\omega \varphi(r)) \, dr \). The validity of \( f_{\varphi} \) requires some (mild) technical condition on \( f \)
when \( \varphi \in L_p(\mathbb{R}^d) \) is not rapidly decaying.

This property is investigated in full depth in Chapter 9 and exploited for deriving
transform-domain statistics. The precise statement of this id result for \( \varphi \in L_p(\mathbb{R}^d) \) is given
in Theorem 21.

Another manifestation of the id property is that a continuous-domain Lévy noise can al-
ways be broken down into an arbitrary number of independent and identically distrib-
uted (i.i.d.) components.
4.4. White Lévy noises or innovations

**Proposition 8.** A white Lévy noise $w$ is infinitely divisible in the sense that it can be decomposed as $w = w_1 + \cdots + w_N$ for any $N \in \mathbb{Z}^+$ where the $w_n$ are i.i.d. white-noise processes.

This simply follows from the property that the characteristic form of the sum of two independent processes is the product of their individual characteristic forms. Specifically, we can write

$$\mathcal{P}_w(\varphi) = \left( \mathcal{P}_{w_n}(\varphi) \right)^N$$

where $\mathcal{P}_{w_n}(\varphi) = \exp\left( \int_{\mathbb{R}^d} f(\varphi(r))/N \, dr \right)$ is the characteristic form of a Lévy noise. The justification is that $f(\omega)/N$ is a valid Lévy exponent for any $N \geq 1$. In the impulsive Poisson case, this simply translates into the Poisson density parameter $\lambda$ being divided by $N$.

Interestingly, there is also a converse to the statement in Proposition 8 [PR70]: a generalized process $s$ over $\mathcal{S}(\mathbb{R}^d)$ is infinitely divisible if and only if its characteristic form can be written as $\mathcal{P}_s(\varphi) = \exp\{F(\varphi)\}$ where $F(\varphi)$ is a continuous, conditional positive-definite functional over $\mathcal{S}(\mathbb{R}^d)$ (or generalized Lévy exponent) as specified in Definition 10 [PR70, Main theorem]. While this general characterization is nice conceptually, it is hardly discriminative since the underlying notion of infinite divisibility applies to all concrete families of generalized stochastic processes that are known to us. In particular, it does not require the “whiteness” property that is fundamental for defining proper innovations.

3) Flat power spectrum

Strictly speaking, the properties of stationarity and independence at every point are not sufficient for specifying “white” noise. There is also some implicit idea of enforcing a flat Fourier spectrum. A simple example that satisfies the two first properties but fails to meet the latter is the weak derivative of a Lévy noise whose generalized power spectrum (when defined) is not flat but proportional to $|\omega|^2$.

The notion of power spectrum is based on second-order moments and does only make sense when the stochastic process is stationary with a well-defined autocorrelation. In Gelfand’s theory, the second-order dependencies are captured by the correlation form $\mathcal{B}_w(\varphi_1, \varphi_2) = \mathbb{E}\{\langle \varphi_1, w \rangle \cdot \langle \varphi_2, w \rangle\}$, where it is assumed that the generalized noise process $w$ is real-valued and that its second-order moments are well-defined. The latter second-order requirement is equivalent to imposing that the Lévy exponent $f$ should be twice differentiable at the origin or, equivalently, that the canonical id distribution $p_{id}$ of the process has a finite second-order moment.

**Proposition 9.** Let $w$ be a (second-order) white Lévy noise with Lévy exponent $f$ such that $f'(0) = 0$ (zero-mean assumption) and $\sigma_w^2 = -f''(0) < +\infty$ (finite-variance assumption). Then,

$$\mathcal{B}_w(\varphi_1, \varphi_2) = \sigma_w^2 \langle \varphi_1, \varphi_2 \rangle. \quad (4.19)$$

Formally, this corresponds to the statement that the autocorrelation of a second-order Lévy noise is a Dirac impulse; i.e.,

$$R_w(r) = \mathbb{E}\{w(r_0) w(r_0 - r)\} = \mathcal{B}_w(\delta(\cdot - r_0), \delta(\cdot - r_0 - r)) = \sigma_w^2 \delta(r)$$

4. In the statistical literature, a second-order process usually designates a stochastic process whose second-order moments are all well defined. In the case of generalized processes, the property refers to the existence of the correlation form.
Similarly, we get
\[ \Phi_R(\omega) = \mathcal{F} \{ R_w(r) \} (\omega) = \sigma_w^2. \]

We recall that the term “white” is used in reference to white light, whose electromagnetic spectrum is distributed over the visible band in a way that stimulates all color receptors of the eye equally. This is in opposition with “colored” noise whose spectral content is not equally distributed.

**Proof.** We have that \( \mathcal{B}_w(\varphi_1, \varphi_2) = \mathbb{E}[X_1 X_2] \) where \( X_1 = \langle \varphi_1, w \rangle \) and \( X_2 = \langle \varphi_2, w \rangle \) are real-valued random variables with joint characteristic function \( \hat{p}_{X_1, X_2}(\omega) = \exp \{ F(\omega_1 \varphi_1 + \omega_2 \varphi_2) \} \) with \( \omega = (\omega_1, \omega_2) \). We then invoke the moment-generating property of the Fourier transform which translates into

\[ \mathcal{B}_w(\varphi_1, \varphi_2) = \mathbb{E}[X_1 X_2] = (-j)^2 \frac{\partial^2 \hat{p}_{X_1, X_2}(\omega)}{\partial \omega_1 \partial \omega_2} \bigg|_{\omega_1 = 0, \omega_2 = 0}. \]

By applying the chain rule twice, we obtain

\[ \frac{\partial^2 \hat{p}_{X_1, X_2}(\omega)}{\partial \omega_1 \partial \omega_2} = e^{f_{X_1, X_2}(\omega_1, \omega_2)} \left( \frac{\partial f_{X_1, X_2}(\omega)}{\partial \omega_1} \frac{\partial f_{X_1, X_2}(\omega)}{\partial \omega_2} + \frac{\partial^2 f_{X_1, X_2}(\omega)}{\partial \omega_1 \partial \omega_2} \right) \]

where \( f_{X_1, X_2}(\omega) = \log \hat{p}_{X_1, X_2}(\omega_1, \omega_2) = \int_{\mathbb{R}^d} f(\omega_1 \varphi_1(r) + \omega_2 \varphi_2(r)) \, dr \) is the cumulant generating function of \( p_{X_1, X_2} \). The required first derivative with respect to \( \omega_i \), \( i = 1, 2 \) is given by

\[ \frac{\partial f_{X_1, X_2}(\omega)}{\partial \omega_i} = \int_{\mathbb{R}^d} f'(\omega_1 \varphi_1(r) + \omega_2 \varphi_2(r)) \varphi_i(r) \, dr, \]

which, when evaluated at the origin, simplifies to

\[ \frac{\partial f_{X_1, X_2}(0)}{\partial \omega_i} = f'(0) \int_{\mathbb{R}^d} \varphi_i(r) \, dr = -j \mathbb{E} \{ X_i \}. \]

(4.20)

Similarly, we get

\[ \frac{\partial^2 f_{X_1, X_2}(0)}{\partial \omega_1 \partial \omega_2} = f''(0) \int_{\mathbb{R}^d} \varphi_1(r) \varphi_2(r) \, dr. \]

By combining those results and using the property that \( f_{X_1, X_2}(0) = 0 \), we conclude that

\[ \mathbb{E}[X_1 X_2] = -f''(0) \langle \varphi_1, \varphi_2 \rangle - (f'(0))^2 \langle \varphi_1, 1 \rangle \langle \varphi_2, 1 \rangle, \]

which is equivalent to (4.19) under the hypothesis that \( f'(0) = 0 \). It is also clear from (4.20) that this latter condition is equivalent to the zero-mean property of the noise; that is, \( \mathbb{E} \{ \varphi, w \} = 0 \) for all \( \varphi \in \mathcal{F}(\mathbb{R}^d) \). Finally, we note that (4.19) is compatible with the more general cumulant formula (9.20) if we set \( n = (1, 1) \), \( n = 2 \), and \( \kappa_2 = (-j)^2 f''(0) \). \( \square \)

Since \( f(0) = 0 \) by definition, another way of writing the hypotheses in Proposition 9 is \( f(\omega) = -\frac{\sigma_w^2}{2} \omega^2 + O(|\omega|^3) \), which expresses an asymptotic equivalence with the symmetric Gaussian scenario (purely quadratic Lévy exponent). This *second-order assumption* ensures that the noise has zero-mean and a finite variance \( \sigma_w^2 \), so that its correlation form (4.19) is well defined. In the sequel, it is made implicitly whenever we are talking of correlations or power spectra.
4.4. White Lévy noises or innovations

4) Stochastic counterpart of the Dirac impulse

From an engineering perspective, white noise is often viewed as the stochastic analog of the Dirac distribution $\delta$ whose spectrum is flat in the literal sense (i.e., $\mathcal{F}(\delta) = 1$). The fundamental difference, of course, is that the generalized function $\delta$ is a deterministic entity. The simplest way of introducing randomness is by considering a shifted and weighted impulse $\lambda \delta(-r_0)$ whose location $r_0$ is uniformly distributed over some compact subset of $\mathbb{R}^d$ and whose amplitude $\lambda$ is a random variable with pdf $p_\lambda$. A richer form of excitation is obtained through the summation of such i.i.d. elementary contributions which results in the construction of impulsive Poisson noise, as specified by (4.14). Theorem 10 ensures that this explicit way of representing noise is legitimate in the case where the Lévy density $\nu = \lambda p_\lambda$ is integrable and the Gaussian part absent. We shall now see that this constructive approach can be pushed to the limit for the non-Poisson brands of innovations, including the Gaussian ones.

**Proposition 10.** A white Lévy noise is the limit of a sequence of impulsive Poisson-noise processes in the sense of the weak convergence of the underlying infinite-dimensional measures.

**Proof.** The technical part of the proof uses an infinite-dimensional generalization of Lévy’s continuity theorem and will be reported elsewhere. The key idea is to consider the following sequence of Lévy exponents

$$f_n(\omega) = n(e^{\frac{1}{n}f(\omega)} - 1) = f(\omega) + O\left(\frac{f^2(\omega)}{n}\right),$$

which are of the compound-Poisson type with $\lambda_n = n$ and $e^{\frac{1}{n}f(\omega)}$ and which converge to $f(\omega)$ as $n$ goes to infinity. This suggests forming the corresponding sequence of characteristic functionals

$$\mathcal{T}_{\omega_n}(\varphi) = \exp \left[ \int_{\mathbb{R}^d} n \left( e^{\frac{1}{n}f(\varphi(r))} - 1 \right) \, dr \right],$$

which are expected to converge to $\mathcal{T}_\omega(\varphi) = \exp \left( \int_{\mathbb{R}^d} f(\varphi(r)) \, dr \right)$ as $n \to +\infty$. The main point for the argument is that these are all of the impulsive-Poisson type for $n$ fixed (see Theorem 10). The crux of the proof is to control the convergence by specifying some appropriate bound and to verify that some basic equicontinuity conditions are met.

The result is interesting because it gives us some insight into the nature of continuous-domain noise. The limit process involves random Dirac impulses that get denser as $n$ increases. When the variance of the noise $\sigma_w^2 = -f''(0)$ is finite, the increase of the average number of impulses per unit volume $\lambda_n = O(n)$ is compensated by a decrease of the variance of the amplitude distribution in inverse proportion: $\text{Var}(A_n) = \sigma_w^2 / n$. While any of the generalized noise processes in the sequence is as rough as a Dirac impulse, this picture suggests that the degree of singularity of the limit object in the non-Poisson scenario can be potentially reduced due to the accumulation of impulses and the fact that the variance of their amplitude distribution converges to zero. The particular example that we have in mind is the Gaussian white noise, which can be obtained as a limit of compound Poisson processes with contracting Gaussian amplitude distributions.
4. CONTINUOUS-DOMAIN INNOVATION MODELS

4.5 Generalized stochastic processes and linear models

As already mentioned, the class of generalized stochastic processes that are of interest to us are those defined through the generic innovation model $Ls = w$ (linear stochastic differential equation) where the differential operator $L$ is shift-invariant and where the driving term $w$ is a continuous-domain white Lévy noise. Having made sense of the latter, we can now proceed with the specification of the class of admissible whitening operators $L$. The key requirement there is that the model be invertible (in the sense of generalized functions) which, by duality, translates into some boundedness constraint on the adjoint operator $L^\dagger$. For the time being, we shall limit ourselves to making some general statements about $L$ and its inverse that ensure existence while deferring to Chapter 5 for concrete examples of admissible operators.

4.5.1 Innovation models

The interpretation of the above continuous-domain linear model in the sense of generalized functions is

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d), \quad \langle \varphi, Ls \rangle = \langle \varphi, w \rangle. \quad (4.21)$$

The generalized stochastic process $s = L^{-1}w$ is generated by solving this equation, which amounts to a linear transformation of the Lévy innovation $w$. Formally, this translates into

$$\forall \varphi \in \mathcal{S}(\mathbb{R}^d), \quad \langle \varphi, s \rangle = \langle \varphi, L^{-1}w \rangle = \langle L^{-1} \varphi, w \rangle \quad (4.22)$$

where $L^{-1}$ is an appropriate right inverse of $L$. The above manipulation obviously only makes sense if the action of the adjoint operator $L^\dagger$ is well-defined over Schwartz’ class $\mathcal{S}(\mathbb{R}^d)$ of test functions—ideally, a continuous mapping from $\mathcal{S}(\mathbb{R}^d)$ into itself or, possibly, $L^p(\mathbb{R}^d)$ (or some variant) if one imposes suitable restrictions of the Lévy exponent $f$ to maintain continuity.

We like to refer to (4.21) as the analysis statement of the model, and to (4.22)—or its shorthand $s = L^{-1}w$—as the synthesis description. Of course, this will only work properly if we have an exact equivalence, meaning that there is a proper and unique definition of $L^\dagger$.

The latter will need to be made explicit on a case-by-case basis with the possible help of boundary conditions.

4.5.2 Existence and characterization of the solution

We shall now see that, under suitable conditions on $L^{-1\dagger}$ (see Theorem 12 below), one can completely specify such processes via their characteristic form and ensure their existence as solutions of (4.21). We recall that the characteristic form of a generalized stochastic process $s$ is defined as

$$\hat{\mathcal{P}}_s(\varphi) = \mathbb{E}[e^{i\langle \varphi, s \rangle}] = \int_{\mathcal{S}'(\mathbb{R}^d)} e^{i\langle \varphi, s \rangle} \mathcal{P}_s(\,ds),$$

where the latter expression involves an abstract infinite-dimensional integral over the space of tempered distributions and provides the connection with the defining measure $\mathcal{P}_s$ on $\mathcal{S}'(\mathbb{R}^d)$. $\hat{\mathcal{P}}_s$ is a functional $\mathcal{S}'(\mathbb{R}^d) \to \mathbb{C}$ that associates the complex number $\hat{\mathcal{P}}_s(\varphi)$ to each test function $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and which is endowed with three fundamental properties: positive definiteness, continuity, and normalization (i.e., $\hat{\mathcal{P}}_s(0) = 1$). It can also be
4.5. Generalized stochastic processes and linear models

specified using the more concrete formula

\[
\tilde{\mathcal{P}}_s(\varphi) = \int_\mathbb{R} e^{iy} d\mathcal{P}_{(\varphi,s)}(y),
\]

(4.23)

which involves a classical Stieltjes integral with respect to the probability law \(\mathcal{P}_{Y=\langle \varphi, s \rangle} = \text{Prob}(Y < y)\), where \(Y = \langle \varphi, s \rangle\) is a conventional scalar random variable, once \(\varphi\) is fixed.

For completeness, we also recall the meaning of the underlying terminology in the context of a generic (normed or nuclear) space \(\mathcal{X}\) of test functions.

**Definition 11** (Positive-definite functional). A complex-valued functional \(G : \mathcal{X} \rightarrow \mathbb{C}\) defined over the function space \(\mathcal{X}\) is said to be positive-definite if

\[
\sum_{m=1}^{N} \sum_{n=1}^{N} G(\varphi_m - \varphi_n) \xi_m \overline{\xi_n} \geq 0
\]

for every possible choice of \(\varphi_1, \ldots, \varphi_N \in \mathcal{X}\), \(\xi_1, \ldots, \xi_N \in \mathbb{C}\), and \(N \in \mathbb{N}^+\).

**Definition 12** (Continuous functional). A functional \(G : \mathcal{X} \rightarrow \mathbb{R}\) (or \(\mathbb{C}\)) is said to be continuous (with respect to the topology of the function space \(\mathcal{X}\)) if, for any convergent sequence \((\varphi_i)_i\) in \(\mathcal{X}\) with limit \(\varphi \in \mathcal{X}\), the sequence \(G(\varphi_i)\) converges to \(G(\varphi)\); that is,

\[
\lim_{i} G(\varphi_i) = G(\lim_{i} \varphi_i).
\]

An essential element of our formulation is that Schwartz’ space of test function \(\mathcal{S}(\mathbb{R}^d)\) is nuclear (see Section 3.1.3), as required by the Minlos-Bochner theorem (Theorem 5). The latter expresses the one-to-one correspondence (in the form of an infinite-dimensional Fourier pair) between the characteristic form \(\tilde{\mathcal{P}}_s : \mathcal{S}(\mathbb{R}^d) \rightarrow \mathbb{C}\) and the measure \(\mathcal{P}_s\) on \(\mathcal{S}'(\mathbb{R}^d)\) that uniquely characterizes the generalized process \(s\). The truly powerful aspect of the theorem is that it suffices to check that \(\tilde{\mathcal{P}}_s\) satisfies the three defining conditions—positive definiteness, continuity over \(\mathcal{S}(\mathbb{R}^d)\), and normalization—to prove that it is a valid characteristic form, which then automatically ensures the existence of the process since the corresponding measure over \(\mathcal{S}'(\mathbb{R}^d)\) is well defined.

The formulation of our generative model (4.22) in that context is

\[
\tilde{\mathcal{P}}_s(\varphi) = \tilde{\mathcal{P}}_{L^{-1}s}(\varphi) = \tilde{\mathcal{P}}_{W}(L^{-1} \varphi),
\]

(4.24)

where \(\tilde{\mathcal{P}}_w(\varphi)\) is the characteristic form of the innovation process \(w\).

**Theorem 12** (Generalized innovation model). Let \(U = L^{-1}s\) be a linear operator that satisfies the two conditions

1. Left inverse property: \(UL^* \varphi = \varphi\) for all \(\varphi \in \mathcal{S}(\mathbb{R}^d)\) where \(L^*\) is the adjoint of some given (whitening) operator \(L\);
2. Stability: \(U\) is a continuous linear map from \(\mathcal{S}(\mathbb{R}^d)\) into itself.

Then, the generalized stochastic process \(s\) that is fully characterized by \(\tilde{\mathcal{P}}_s(\varphi) = \exp\left(\int \varphi (L^{-1} \varphi(w)) \, dw\right)\) is well-defined and satisfies the innovation model \(L_s = w\), where \(w\) is a Lévy innovation with exponent \(f\).

Moreover, when \(f\) is \(p\)-admissible (cf. Definition 8), the second condition can be substituted by the much weaker \(L_p\)-stability requirement

\[
\|U\varphi\|_{L_p} < C_2 \|\varphi\|_{L_p}, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).
\]

(4.25)
4. **Continuous-domain innovation models**

*Proof.* First, we prove that $s$ is a bona fide generalized stochastic process over $\mathcal{S}'(\mathbb{R}^d)$ by showing that $\tilde{\mathcal{F}}_s(\varphi)$ is a continuous, positive-definite functional on $\mathcal{S}(\mathbb{R}^d)$ such that $\tilde{\mathcal{F}}_s(0) = 1$ (by the Minlos-Bochner theorem).

As for the first condition, we rely on the property that the Lévy noise functional $\tilde{\mathcal{F}}_w(\varphi) = \exp\left(\int_{\mathbb{R}^d} f(\varphi(r)) \, dr\right)$ is continuous over $\mathcal{S}(\mathbb{R}^d)$ (Theorem 9). Moreover, when the Lévy exponent $f$ is $p$-admissible, the positive definiteness and continuity properties extend to $L_p(\mathbb{R}^d)$ by Theorem 20, which will be proven in Chapter 8. By construction, $\tilde{\mathcal{F}}_w$ is a continuous functional on $\mathcal{S}(\mathbb{R}^d)$ (resp., $L_p(\mathbb{R}^d)$). This, together with the assumption that $L^{-1*}$ is a continuous operator on $\mathcal{S}(\mathbb{R}^d)$, implies that the composed functional $\tilde{\mathcal{F}}_s(\varphi) = \tilde{\mathcal{F}}_w(L^{-1*} \varphi)$ is continuous on $\mathcal{S}(\mathbb{R}^d)$. Likewise, when $f$ is $p$-admissible, the $L_p$ stability condition implies that $L^{-1*}$ is a continuous operator $\mathcal{S}(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$, so that the same reasoning is applicable as well—see the triangular diagram in Figure 3.1 with $\mathcal{X} = \mathcal{S}(\mathbb{R}^d)$ and $\mathcal{Y} = L_p(\mathbb{R}^d)$.

Next, for any given set of functions $\varphi_1, \ldots, \varphi_N \in \mathcal{S}(\mathbb{R}^d)$ and coefficients $\xi_1, \ldots, \xi_N \in \mathbb{C}$, we have

$$
\sum_{m=1}^{N} \sum_{n=1}^{N} \tilde{\mathcal{F}}_s(\varphi_m - \varphi_n) \xi_m \xi_n = \sum_{m=1}^{N} \sum_{n=1}^{N} \tilde{\mathcal{F}}_w(L^{-1*} (\varphi_m - \varphi_n)) \xi_m \xi_n
$$

$$
= \sum_{m=1}^{N} \sum_{n=1}^{N} \tilde{\mathcal{F}}_w(L^{-1} \varphi_m - L^{-1} \varphi_n) \xi_m \xi_n
$$

(by linearity)

$$
\geq 0, \quad \text{(by the positivity of } \tilde{\mathcal{F}}_w\text{)}
$$

which shows that $\tilde{\mathcal{F}}_s$ is positive definite on $\mathcal{S}(\mathbb{R}^d)$. Finally, $\tilde{\mathcal{F}}_s(0) = \tilde{\mathcal{F}}_w(L^{-1*} 0) = \tilde{\mathcal{F}}_w(0) = 1$, which completes the first part of the proof.

The above result and stability conditions ensure that the action of the inverse operator $L^{-1}$ is well-defined over the relevant subset of tempered distributions, which justifies the formal manipulation made in (4.22). Now, if $L^{-1*}$ is a proper left inverse of $L^*$, we have that

$$
(\varphi, w) = (L^{-1*} L^* \varphi, w) = (L^* \varphi, L^{-1*} w) = (\varphi, Ls),
$$

which proves that the generalized process $s = L^{-1} w$ satisfies (4.21).

The next chapters are devoted to the investigation of specific instances of this model and to making sure that the conditions for existence in Theorem 12 are met. We shall also discuss the conceptual connection with splines and wavelets. This connection is fundamental to our purpose. In Chapter 9, we shall then use the generalized innovation model to show that the primary statistical features of the input noise are essentially transferred to the signal as well as to the transform domain. The main point is that the marginal distributions are all part of the same infinitely divisible family as long as the composition of the mixing procedure ($L^{-1}$) and the signal analysis remains linear. On the other hand, the amount of coupling and level of interdependence will strongly depend on the nature of the transformation.

### 4.6 Notes

Infinitely divisible distributions were introduced by de Finneti and their primary properties established by Kolmogorov, Lévy, and Khintchine in the 1930s [BDR02, SVH03].
4.6. Notes

We have chosen to name Theorem 6 after Lévy and Schoenberg because it essentially results from the combination of two fundamental theorems in harmonic analysis named after these authors [Sch38]. While their groundwork dates back to the 1930s, it took until the late 1960s to reformulate the Lévy-Khintchine characterization in terms of the (conditional) positive definiteness of the exponent [Joh66].
Chapter 5

Operators and their inverses

In this chapter we review three classes of linear shift-invariant (LSI) operators: convolution operators with stable LSI inverses, operators that are linked with ordinary differential equations, and fractional operators.

The first class, considered in Section 5.2, is composed of the broad family of multidimensional operators whose inverses are stable convolution operators—or filters. Convolution operators play a central role in signal processing. They are easy to characterize mathematically via their impulse response. The corresponding generative model for stochastic processes amounts to LSI filtering of a white noise, which automatically yields stationary processes.

Our second class is the 1-D family of ordinary differential operators with constant coefficients, which is relevant to a wide range of modeling applications. In the “stable” scenario, reviewed in Section 5.3, these operators admit stable LSI inverses on $\mathcal{S}$ and are therefore included in the previous category. On the other hand, when the differential operators in question have zeros on the imaginary axis (the marginally stable case), they find a non-trivial null-space in $\mathcal{S}$, which consists of (exponential) polynomials. This implies that they are no longer unconditionally invertible on $\mathcal{S}$, and that we can at best identify left- or right-side inverses, which should additionally fulfill appropriate “boundedness” requirements in order to be usable in the definition of stochastic processes. However, as we shall see in Section 5.4, obtaining an inverse with the required boundedness properties is feasible but requires giving up shift-invariance. As a consequence, stochastic processes defined by these operators are generally non-stationary.

The third class of LSI operators, investigated in Section 5.5, consists of fractional derivatives and/or Laplacians in one and several dimensions. Our focus is on the family of linear operators on $\mathcal{S}$ that are simultaneously homogeneous (scale-invariant up to a scalar coefficient) and invariant under shifts and rotations. These operators are intimately linked to self-similar processes and fractals [BU07, TVDVU09]. Once again, finding a stable inverse operator to be used in the definition of self-similar processes poses a mathematical challenge since the underlying system is inherently unstable. The difficulty is evidenced by the fact that statistical self-similarity is generally not compatible with stationarity, which means that a non-shift-invariant inverse operator needs to be constructed. Here again, a solution may be found by extending the approach used for the previous class of operators. From our first example in Section 5.1, we shall actually see that one is able to reconcile the classical theory of stationary processes with that of self-similar ones by viewing the latter as a limit case of the former.

Before we begin our discussion of operators, let us formalize some notions of invariance.
5. Operators and their inverses

**Definition 15** (Translation invariance). An operator $T$ is shift- (or translation-) invariant if and only if, for any function $\varphi$ in its domain and any $r_0 \in \mathbb{R}^d$,

$$T[\varphi(\cdot - r_0)](r) = T[\varphi(\cdot - r_0)](r).$$

**Definition 16** (Scale invariance). An operator $T$ is scale-invariant (homogeneous) of order $\alpha$ if and only if, for any function $\varphi$ in its domain,

$$T[\varphi(r/a)] = |a|^{-\alpha} T[\varphi(r)],$$

where $a \in \mathbb{R}^+$ is the dilation factor.

**Definition 17** (Rotation invariance). An operator $T$ is scalarly rotation-invariant if and only if, for any function $\varphi$ in its domain,

$$T[\varphi(R^T r)] = T[\varphi(R^T r)],$$

where $R$ is any orthogonal matrix in $\mathbb{R}^{d \times d}$ (by using orthogonal matrices in the definition, we take into account both proper and improper rotations, with respective determinants 1 and $-1$).

5.1 Introductory example: first-order differential equation

To fix ideas, let us consider the generic first-order differential operator $L = D - \alpha \text{Id}$ with $\alpha \in \mathbb{C}$, where $D = \frac{d}{dr}$ and $\text{Id}$ are the derivative and identity operators, respectively. Clearly, $L$ is LSI, but generally not scale-invariant unless $\alpha = 0$. The corresponding linear system with (deterministic or stochastic) output $s$ and input $w$ is defined by the differential equation:

$$\frac{d}{dr} s(r) - \alpha s(r) = w(r).$$

Under the classical stability assumption $\text{Re}(\alpha) < 0$ (pole in the left half of the complex plane), its impulse response is given by

$$\rho_\alpha(r) = \mathcal{F}^{-1} \left\{ \frac{1}{j\omega - \alpha} \right\}(r) = \mathbb{1}_+(r) e^{\alpha r}$$

and is rapidly decaying. This provides us with an explicit characterization of the inverse operator, which reduces to a simple convolution with a decreasing causal exponential:

$$(D - \alpha \text{Id})^{-1} \varphi = \rho_\alpha * \varphi.$$

Likewise, it is easy to see that the corresponding adjoint inverse operator $L^{-1*}$ is specified by

$$(D - \alpha \text{Id})^{-1*} \varphi = \rho_\alpha^* * \varphi,$$

where $\rho_\alpha^*(r) = \rho_\alpha(-r)$ is the time-reversed version of $\rho_\alpha$. Thanks to its rapid decay, $\rho_\alpha^*$ defines a continuous linear translation-invariant map from $\mathcal{F}(\mathbb{R})$ into itself.

This allows us to express the solution of the first-order SDE as a filtered version of the input noise $s_\alpha = \rho_\alpha * w$. It follows that $s_\alpha$ is a stationary process that is completely specified by its characteristic form $\mathcal{P}_{s_\alpha}(\varphi) = \mathcal{P}_w(\rho_\alpha^* * \varphi)$, where $\mathcal{P}_w$ is the Lévy-white-noise functional of the input (see Section 4.4).

Let us now focus our attention on the limit case $\alpha = 0$, which yields an operator $L = D$ that is scale-invariant. Here too, it is possible to specify the LSI inverse (integrator)

$$D^{-1} \varphi(r) = \int_{-\infty}^{r} \varphi(t) \, dt = (\mathbb{1}_+ * \varphi)(r).$$
5.1. Introductory example: first-order differential equation

\[ D^{-1} \varphi(t) = \int_{-\infty}^{t} \varphi(\tau) d\tau \]

\[ I_0 \varphi(t) = \int_{0}^{t} \varphi(\tau) d\tau \]

Figure 5.1: Comparison of antiderivative operators. (a) Input signal. (b) Result of shift-invariant integrator and its adjoint. (c) Result of scale-invariant integrator \( I_0 \) and its \( L_p \)-stable adjoint \( I_0^* \); the former yields a signal that vanishes at the origin, while the latter enforces the decay of the output as \( t \to -\infty \) at the cost of a jump discontinuity at the origin.

whose output is well-defined pointwise when \( \varphi \in \mathcal{S}(\mathbb{R}) \). The less-favorable aspect is that the classical LSI integrator does not fulfill the usual stability requirement due to non-integrability of its impulse response \( \mathds{1}_+ \not\in L_1(\mathbb{R}) \). This implies that \( D^{-1} \varphi = \mathds{1}_+^* \ast \varphi \) is generally not in \( L_p(\mathbb{R}) \). Thus, we are no longer fulfilling the admissibility condition in Theorem 15. The source of the problem is the lack of decay of \( D^{-1} \varphi(r) \) as \( r \to -\infty \) when \( \int_{\mathbb{R}} \varphi(r) \, dr = \hat{\varphi}(0) \neq 0 \) (see Figure 5.1b). Fortunately, this can be compensated by defining the modified antiderivative operator

\[ I_0^* \varphi(r) = \int_{-\infty}^{r} \varphi(\tau) \, d\tau - 1_+ (-r) \hat{\varphi}(0) = (\mathds{1}_+^* \ast \varphi)(r) - \int_{\mathbb{R}} \varphi(r) \, dr \mathds{1}_+^*(r) \]

\[ = D^{-1} \varphi(r) - (D^{-1} \varphi)(-\infty) \mathds{1}_+^*(r) \]

which happens to be the only left inverse of \( D^* = -D \) that is both scale-invariant and \( L_p \)-stable for any \( p > 0 \). The adjoint of \( I_0^* \) specifies the adjusted \(^1\) integrator

\[ I_0 \varphi(r) = \int_{0}^{r} \varphi(r) \, dr = D^{-1} \varphi(r) - (D^{-1} \varphi)(0), \]

which is the correct scale-invariant inverse of \( D \) for our formulation, to be applied to elements of \( \mathcal{S}'(\mathbb{R}) \). It follows that the solution of the corresponding (unstable) SDE can be

\(^1\) Any two valid right inverses can only differ by a component (constant) that is in the null space of the operator. The scale-invariant solution is the one that forces the output to vanish at the origin.
expressed as \( s = I_0 w \), which is a well-defined stochastic process as long as the input noise is Lévy \( p \)-admissible with \( p \geq 1 \) (by Theorem 15). We note that the price to pay for the stabilization of the solution is to give up on shift invariance. Indeed, the adjusted integrator is such that it imposes the boundary condition \( s(0) = (I_0 w)(0) = 0 \) (see Figure 5.1c), which is incompatible with stationarity, but a necessary condition for self-similarity. The so-constructed processes are fully specified by their characteristic form \( \exp\left\{ \int_0^t f(l_0 \varphi(r)) \, dt \right\} \), where \( f \) is a Lévy exponent. Based on this representation, we can show that these are equivalent to the Lévy processes that are usually defined for \( r \geq 0 \) only [Sat94]. While this connection with the classical theory of Lévy processes is already remarkable, it turns out that the underlying principle is quite general and applicable to a much broader family of operators, provided that we can ensure \( L_p \)-stability.

5.2 Shift-invariant inverse operators

The association of LSI operators with convolution integrals will be familiar to most readers. In effect, we saw in Section 3.3.5 that, as a consequence of the Schwartz kernel theorem, every continuous LSI operator \( L : \mathcal{S}(\mathbb{R}^d) \rightarrow \mathcal{S}'(\mathbb{R}^d) \) corresponds to a convolution

\[
L : \varphi \mapsto L[\delta] * \varphi
\]

with a kernel (impulse response) \( L[\delta] \in \mathcal{S}'(\mathbb{R}^d) \). Moreover, by the convolution-product rule, we may also characterize \( L \) by a multiplication in the Fourier domain:

\[
L : \varphi \mapsto \mathcal{F}^{-1}\{ \hat{L}(\omega) \hat{\varphi}(\omega) \}. \tag{5.4}
\]

We call \( \hat{L} \) the Fourier multiplier or symbol associated with the operator \( L \).

From the Fourier-domain characterization of \( L \), we see that, if \( \hat{L} \) is smooth and does not grow too fast, then \( L \) maps \( \mathcal{S}(\mathbb{R}^d) \) back into \( \mathcal{S}(\mathbb{R}^d) \). This is in particular true if \( L[\delta] \) (the impulse response) is an ordinary locally integrable function with rapid decay. It is also true if \( L \) is a 1-D linear differential operator with constant coefficients, in which case \( L[\delta] \) is a finite sum of derivatives of the Dirac distribution and the corresponding Fourier multiplier \( \hat{L}(\omega) \) is a polynomial in \( j \omega \).

For operators with smooth Fourier multipliers that are nowhere zero in \( \mathbb{R}^d \) and not decaying (or decaying slowly) at \( \infty \), we can define the inverse \( L^{-1} : \mathcal{S}(\mathbb{R}^d) \rightarrow \mathcal{S}(\mathbb{R}^d) \) of \( L \) by

\[
L^{-1} : \varphi \mapsto \mathcal{F}^{-1}\left\{ \frac{\hat{\varphi}(\omega)}{\hat{L}(\omega)} \right\}.
\]

This inverse operator is also linear and shift-invariant, and has the convolution kernel

\[
\rho_L = \mathcal{F}^{-1}\left\{ \frac{1}{\hat{L}(\omega)} \right\}, \tag{5.4}
\]

which is in effect the Green’s function of the operator \( L \). Thus, we may write

\[
L^{-1} : \varphi \mapsto \rho_L * \varphi.
\]

For the cases in which \( \hat{L}(\omega) \) vanishes at some points, its inverse \( 1/\hat{L}(\omega) \) is not in general a locally-integrable function, but even in the singular case we may still be able to regularize the singularities at the zeros of \( \hat{L}(\omega) \) and obtain a singular “generalized function” whose inverse Fourier transform, per (5.4), once again yields a convolution kernel \( \rho_L \) that is a

---

2. A similar result holds for continuous LSI operators \( \mathcal{D} \rightarrow \mathcal{D}' \), which will have a convolution kernel in \( \mathcal{D}' \).
Green’s function of L. The difference is that in this case, for an arbitrary \( \varphi \in \mathcal{S}(\mathbb{R}^d) \), \( \rho_L \ast \varphi \) may no longer belong to \( \mathcal{S}(\mathbb{R}^d) \).

As in our introductory example, the simplest scenario occurs when the inverse operator \( L^{-1} \) is shift-invariant with an impulse response \( \rho_L \) that has sufficient decay for the system to be BIBO-stable (bounded input, bounded output).

**Proposition 11.** Let \( L^{-1} \varphi(r) = (\rho_L \ast \varphi)(r) = \int_{\mathbb{R}^d} \rho_L(r - r') \varphi(r') \, dr' \) with \( \rho_L \in L_1(\mathbb{R}^d) \) (or, more generally, where \( \rho_L \) is a complex-valued Borel measure of bounded variation). Then, \( L^{-1} \) and its adjoint specified by \( L^{-1*} \varphi(r) = (\rho_L^* \ast \varphi)(r) = \int_{\mathbb{R}^d} \rho_L(-r') \varphi(r - r') \, dr' \) are both \( L_p \)-stable in the sense that

\[
\|L^{-1}\varphi\|_{L_p} \leq \|\rho_L\|_{L_1} \|\varphi\|_{L_p}
\]

\[
\|L^{-1*}\varphi\|_{L_p} \leq \|\rho_L\|_{L_1} \|\varphi\|_{L_p}
\]

for all \( p \geq 1 \).

The result follows from Theorem 4. For the sake of completeness, we shall establish the bound based on the two extreme cases \( p = 1 \) and \( p = +\infty \).

**Proof.** To obtain the \( L_1 \) bound, we manipulate the norm of the convolution integral as

\[
\|\rho_L \ast \varphi\|_{L_1} = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \rho_L(r) \varphi(r' - r) \, dr' \right| \, dr
\]

\[
\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\rho_L(r)| |\varphi(s)| \, dr \, ds \quad \text{(change of variable \( s = r' - r \))}
\]

\[
= \int_{\mathbb{R}^d} |\rho_L(r)| \, dr \int_{\mathbb{R}^d} |\varphi(s)| \, ds = \|\rho_L\|_{L_1} \|\varphi\|_{L_1},
\]

where the exchange of integrals is justified by Fubini’s theorem. The corresponding \( L_\infty \) bound follows from the simple pointwise estimate

\[
|\rho_L \ast \varphi(r)| \leq \|\varphi\|_{L_\infty} \int_{\mathbb{R}^d} |\rho_L(r')| \, dr' = \|\rho_L\|_{L_1} \|\varphi\|_{L_\infty}
\]

for any \( r \in \mathbb{R}^d \). The final step is to invoke the Riesz-Thorin theorem (Theorem 3) which yields Young’s inequality for \( L_p \) functions (see (3.7)), and hence proves the desired result for \( 1 \leq p \leq \infty \). The inequality also applies to the adjoint operator since the latter amounts to a convolution with the reversed impulse response \( \rho_L^*(r) = \rho(-r) \in L_1(\mathbb{R}^d) \).

Note that the \( L_1 \) condition in Proposition 11 is the standard hypothesis that is made in the theory of linear systems to ensure the BIBO stability of an analog filter. It is slightly stronger than the TV condition in Theorem 4, which is necessary and sufficient for both BIBO \((p = \infty)\) and \( L_1 \) stabilities.

If, in addition, \( \rho_L(r) \) decays faster than any polynomial (e.g., is compactly supported or decays exponentially), then we can actually ensure \( \mathcal{F} \)-continuity so that there is no restriction on the class of corresponding stochastic processes.

**Proposition 12.** Let \( L^{-1} \varphi(r) = (\rho_L \ast \varphi)(r) \) with \( |\rho_L(r)| \leq \frac{C_n}{1 + |r|^n} \) for all \( n \in \mathbb{Z}^+ \) and \( r \in \mathbb{R}^d \). Then, \( L^{-1} \) and \( L^{-1*} \) are \( \mathcal{F} \)-continuous in the sense that \( \varphi \in \mathcal{F} \Rightarrow L^{-1} \varphi, L^{-1*} \varphi \in \mathcal{F} \) with both operators being bounded.

The key here is that the convolution with \( \rho_L \) preserves the rapid decay of the test function \( \varphi \). The degree of smoothness of the output is not an issue because, for non-constant functions, the convolution operation commutes with differentiation.
5. OPERATORS AND THEIR INVERSES

The good news is that the entire class of stable 1D differential systems with rational transfer functions and poles in the left half of the complex plane falls into the category of Proposition 12. The application of such operators provides us with a convenient mechanism for solving ordinary differential equations, as detailed in the next section.

The $\mathcal{F}$-continuity property is important for our formulation. It also holds for all shift-invariant differential operators whose impulse response is a point distribution; e.g., $(D - \text{Id})[\delta] = \delta' - \delta$. It is preserved under convolution, which justifies the factorization of operators into simpler constituents.

5.3 Stable differential systems in 1-D

The generic form of a linear shift-invariant differential equation in 1-D with (deterministic or random) output $s$ and driving term $w$ is

$$\sum_{n=0}^{N} a_n D^n s = \sum_{m=0}^{M} b_m D^m w$$

(5.5)

where the $a_n$ and $b_m$ are arbitrary complex coefficients with the normalization constraint $a_N = 1$. Equation (5.5) thus covers the general 1-D case of $Ls = w$ where $L$ is a shift-invariant operator with the rational transfer function

$$\tilde{L}(\omega) = \frac{(j\omega)^N + a_{N-1}(j\omega)^{N-1} + \cdots + a_1 (j\omega) + a_0}{b_M (j\omega)^M + \cdots + b_1 (j\omega) + b_0} = \frac{p_N(j\omega)}{q_M(j\omega)}.$$  (5.6)

The poles of the system, which are the roots of the characteristic polynomial $p_N(\zeta) = \zeta^N + a_{N-1} \zeta^{N-1} + \cdots + a_0$ with Laplace variable $\zeta \in \mathbb{C}$, are denoted by $\{a_n\}_{n=1}^N$. In the standard causal-stable scenario where $\text{Re}(a_n) < 0$ for $n = 1, \ldots, N$, the solution is obtained as

$$s(r) = L^{-1} w(r) = (\rho_L \ast w)(r)$$

where $\rho_L$ is the causal Green function of $L$ specified by (5.4).

In practice, the determination of $\rho_L$ is based on the factorization of the transfer function of the system as

$$\frac{1}{\tilde{L}(\omega)} = q_M(j\omega) \prod_{n=1}^{N} \frac{1}{j\omega - a_n}$$

(5.7)

$$= b_M \frac{\prod_{m=1}^{M} (j\omega - \gamma_m)}{\prod_{n=1}^{N} (j\omega - a_n)},$$

(5.8)

which is then broken into simple constituents, either by serial composition of first-order factors or by decomposition into simple partial fractions. We are providing the fully factored form (5.8) of the transfer function to recall the property that a stable $N$th-order system is completely characterized by its poles $\{a_n\}_{n=1}^N$ and zeros $\{\gamma_m\}_{m=1}^M$, up to the proportionality factor $b_M$.

Since the $\mathcal{F}$-continuity property is preserved through the composition of convolution operators, we shall rely on (5.8) to factorize $L^{-1}$ into elementary operators. To that end, we shall study the effect of simple constituents (first-order differential operators with stable inverses) before considering their composition into higher-order operators. We shall also treat the leading polynomial factor $q_M(j\omega)$ in (5.7) separately because it corresponds to a convolution operator whose impulse response is the point distribution $\sum_{m=0}^{M} b_m \delta^{(m)}$. The latter is $\mathcal{F}$-continuous, irrespective of the choice of coefficients $b_m$ (or, equivalently, the zeros $\gamma_m$ in (5.8)).
5.3. Stable differential systems in 1-D

5.3.1 First-order differential operators with stable inverses

The first-order differential operator

\[ P_a = D - \alpha I_d \]

corresponds to the convolution kernel (impulse response) \( (\delta' - \alpha \delta) \) and Fourier multiplier \( (j \omega - \alpha) \). For \( \text{Re}(\alpha) \neq 0 \) (the stable case in signal-processing parlance), the inverse of the Fourier multiplier, \( (j \omega - \alpha)^{-1} \), is non-zero for all \( \omega \in \mathbb{R} \), with the well-defined Fourier inverse

\[ \rho_a(r) = \mathcal{F}^{-1} \left\{ \frac{1}{j \omega - \alpha} \right\}(r) = \begin{cases} e^{\alpha r} \mathbb{1}_{[0, \infty)}(r) & \text{if } \text{Re}(\alpha) < 0, \\ -e^{\alpha r} \mathbb{1}_{[-\infty, 0]}(r) & \text{if } \text{Re}(\alpha) > 0. \end{cases} \]

Clearly, in either of the causal (\( \text{Re}(\alpha) < 0 \)) or anti-causal (\( \text{Re}(\alpha) > 0 \)) cases, these functions decay rapidly at infinity, and convolutions with them map \( \mathcal{S}(\mathbb{R}) \) both from the left and the right, a reason for which we can write

\[ P_a^{-1} \psi = \rho_a \ast \psi \tag{5.9} \]

for \( \text{Re}(\alpha) \neq 0 \).

Since \( P_a \) and \( P_a^{-1} \), \( \text{Re}(\alpha) \neq 0 \), are both \( \mathcal{S} \)-continuous (continuous from \( \mathcal{S}(\mathbb{R}) \) into \( \mathcal{S}(\mathbb{R}) \)), their action can be transferred to the space \( \mathcal{S}'(\mathbb{R}) \) of Schwartz distributions by identifying the adjoint operator \( P_a^* \) with \( (-P_a) \), in keeping with the identity

\[ (P_a \psi, \phi) = -(\psi, P_a \phi) \]

on \( \mathcal{S}(\mathbb{R}) \). We recall that, in this context, \( (\psi, \phi) \) denotes the bilinear form \( \int_{\mathbb{R}} \psi(r) \phi(r) \, dr \), not the Hermitian product.

5.3.2 Higher-order differential operators with stable inverses

As noted earlier, the preservation of continuity by composition facilitates the study of differential systems by permitting us to decompose higher-order operators into first-order factors. Specifically, let us consider the equivalent factorized representation of the \( N \)-th order differential equation (5.5) given by

\[ P_{a_1} \cdots P_{a_N} s(r) = q_M(D) \{ w \}(r) \tag{5.10} \]

where \( \{a_n\}_{n=1}^N \) are the poles of the system and where \( q_M(D) = \sum_{m=0}^M b_m D^m \) is the \( M \)-th order differential operator acting on the right-hand side of (5.5). Under the assumption that \( \text{Re}(a_n) \neq 0 \) (causal or anti-causal stability), we can invert the operators acting on the left-hand side of (5.10), which allows us to express the solution of the differential equation as

\[ s(r) = \underbrace{P_{a_N}^{-1} \cdots P_{a_1}^{-1}}_{L^{-1}} q_M(D) \{ w \}(r). \tag{5.11} \]

This translates into the following formulas for the corresponding inverse operator \( L^{-1} \):

\[ L^{-1} = P_{a_N}^{-1} \cdots P_{a_1}^{-1} q_M(D) = b_M P_{a_N}^{-1} \cdots P_{a_1}^{-1} P_{r_1} \cdots P_{r_N}, \]

which are consistent with (5.7) and (5.8), respectively. These operator-based manipulations are legitimate since all the elementary constituents in (5.10) and (5.11) are \( \mathcal{S} \)-continuous convolution operators. We also recall that all \( L_{\mu} \)-stable and, a fortiori \( \mathcal{S} \)-continuous, convolution operators satisfy the properties of commutativity, associativity,
and distributivity, so that the ordering of the factors in (5.10) and (5.11) is immaterial. Interestingly, this divide-and-conquer approach to the problem of inverting a differential operator is also extendable to the unstable scenarios (with \( \text{Re}(\alpha_n) = 0 \) for some values of \( n \)), the main difference being that the ordering of operators becomes important (partial loss of commutativity).

5. Operators and Their Inverses

5.4 Unstable \( N \)th-order differential systems

Classically, a differential system is categorized as being unstable when some of its poles are in the right complex half-plane, including the imaginary axis. Mathematically, there is no fundamental reason for excluding the cases \( \text{Re}(\alpha_n) > 0 \) because one can simply switch to an anti-causal configuration which preserves the exponential decay of the response, as we did in defining these inverses in Section 5.3.1.

The only tricky situation occurs for purely-imaginary poles of the form \( \alpha_n = j\omega_0 \) with \( \omega_0 \in \mathbb{R} \), to which we now turn our attention.

5.4.1 First-order differential operators with unstable shift-invariant inverses

Once again, we begin with first-order operators. Unlike the case of \( \text{Re}(\alpha) \neq 0 \), for purely imaginary \( \alpha = j\omega_0 \), \( P_{j\omega_0} \) is not a surjective operator from \( \mathscr{S}(\mathbb{R}) \) to \( \mathscr{S}(\mathbb{R}) \) (meaning it maps \( \mathscr{S}(\mathbb{R}) \) into a proper subset of itself, not into the entire space), because it introduces a frequency-domain zero at \( \omega_0 \). This implies that we cannot find an operator \( U : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R}) \) that is a right inverse of \( P_{j\omega_0} \). In other words, we cannot fulfill \( P_{j\omega_0}U \varphi = \varphi \) for all \( \varphi \in \mathscr{S}(\mathbb{R}) \) subject to the constraint that \( U \varphi \in \mathscr{S}(\mathbb{R}) \).

If we now consider \( P_{j\omega_0} \) as an operator on \( \mathscr{S}'(\mathbb{R}) \) {as the adjoint of \( (-P_{j\omega_0}) \)}, then this operator is not one-to-one. More precisely, it has a non-trivial null-space consisting of multiples of the complex sinusoid \( e^{j\omega_0x} \). Consequently, on \( \mathscr{S}'(\mathbb{R}) \), \( P_{j\omega_0} \) does not have a left inverse \( U' : \mathscr{S}'(\mathbb{R}) \to \mathscr{S}'(\mathbb{R}) \) with \( U'P_{j\omega_0}f = f \) for all \( f \in \mathscr{S}'(\mathbb{R}) \).

The main conclusion of concern to us is that \( P_{j\omega_0} : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R}) \) and its adjoint \( P_{j\omega_0}^* : \mathscr{S}'(\mathbb{R}) \to \mathscr{S}'(\mathbb{R}) \) are not invertible in the usual sense of the word (i.e., from both sides). However, we are able to properly invert \( P_{j\omega_0} \) on its image (range), as we discuss now.

Let \( \mathscr{S}_{j\omega_0} \) denote the image of \( \mathscr{S}(\mathbb{R}) \) under \( P_{j\omega_0} \). This is the same as the subspace of \( \mathscr{S}(\mathbb{R}) \) consisting of functions \( \varphi \) for which

\[
\int_{-\infty}^{+\infty} e^{-j\omega_0 r} \varphi(r) \, dr = 0.
\]

In particular, for \( j\omega_0 = 0 \), we obtain \( \mathscr{S}_0 \), the space of Schwartz test functions with vanishing 0th-order moment. We may then view \( P_{j\omega_0} \) as an operator \( \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R}) \), and this operator now has an inverse \( P_{j\omega_0}^{-1} \) from \( \mathscr{S}_{j\omega_0} \to \mathscr{S}(\mathbb{R}) \) defined by

\[
P_{j\omega_0}^{-1} \varphi(r) = (\rho_{j\omega_0} * \varphi)(r), \tag{5.12}
\]

where

\[
\rho_{j\omega_0}(r) = \mathscr{F}^{-1} \left\{ \frac{1}{|r - j\omega_0|} \right\}(r) = \frac{1}{2} \text{sign}(r)e^{j\omega_0 r}. \tag{5.13}
\]

3. To see this, note that \( (D - j\omega_0 \mathbb{I}) \varphi(r) = e^{j\omega_0 r}D[e^{-j\omega_0 r} \varphi(r)] \).
An introduction to Sparse Stochastic Processes

Copyright © M. Unser and P. D. Tafti

5.4. Unstable Nth-order differential systems

Specifically, this LSI operator satisfies the right and left inverse relations

\[ P_{j_0}^{-1} P_{j_0}^{-1} \varphi = \varphi \text{ for all } \varphi \in \mathcal{A}_{j_0} \]
\[ P_{j_0}^{-1} P_{j_0} \varphi = \varphi \text{ for all } \varphi \in \mathcal{S}(\mathbb{R}). \]

In order to be able to use such inverse operators for defining stochastic processes, we need to extend \( P_{j_0}^{-1} \) to all of \( \mathcal{S}(\mathbb{R}) \).

Note that unlike the case of \( P_a^{-1} \) with \( \text{Re}(\alpha) \neq 0 \), here the extension of \( P_{j_0}^{-1} \) to an operator \( \mathcal{S}(\mathbb{R}) \to \mathcal{S}'(\mathbb{R}) \) is in general not unique. For instance, \( P_{j_0}^{-1} \) may also be specified as

\[ P_{j_0}^{-1},+ \varphi(r) = (\rho_{j_0},+ * \varphi)(r), \quad (5.14) \]

with causal impulse response

\[ \rho_{j_0},+(r) = e^{i \omega_0 r} I_+(r), \quad (5.15) \]

which defines the same operator on \( \mathcal{A}_{j_0} \), but not on \( \mathcal{S}(\mathbb{R}) \). In fact, we could as well have taken any impulse response of the form \( \rho_{j_0}(r) + p_0(r) \), where \( p_0(r) = \xi_0 e^{i \omega_0 r} \) is an oscillatory component that is in the null space of \( P_{j_0} \). By contrast, the \( L_p \)-continuous inverses that we define below remain the same for all of these extensions. To convey the idea, we shall first consider the extension based on the causal operator \( P_{j_0}^{-1},+ \) defined by (5.14). Its adjoint is denoted by \( P_{j_0}^{-1,*} \), and amounts to an (anti-causal) convolution with \( \rho_{j_0}',+ \).

To solve the stochastic differential equation \( P_{j_0} s = w \), we need to find a left inverse of the adjoint \( P_{j_0}' \) acting on the space of test functions, which maps \( \mathcal{S}(\mathbb{R}) \) into the required \( L_p \) space (Theorem 15). The problem with the “shift-invariant”\(^4\) extensions of the inverse defined above is that their image inside \( \mathcal{S}'(\mathbb{R}) \) is not contained in arbitrary \( L_p \) spaces. For this reason, we now introduce a different, “corrected”, extension of the inverse of \( P_{j_0}' \) that maps \( \mathcal{S}(\mathbb{R}) \) to \( \mathcal{S}(\mathbb{R}) \)—therefore, a fortiori, also into all \( L_p \) spaces. This corrected left inverse, which we shall denote as \( I_{\omega_0}^* \), is constructed as

\[ I_{\omega_0}^* \varphi(r) = \lim_{y \to -\infty} P_{j_0}^{-1,*} \varphi(y) \rho_{j_0},+ (r) \]
\[ = (\rho_{j_0},+ * \varphi)(r) - \hat{\varphi}(-\omega_0) \rho_{j_0},+ (r) \quad (5.16) \]

in direct analogy with (5.2). As in our introductory example, the role of the second term is to remove the tail of \( P_{j_0}^{-1,*} \varphi(r) \). This ensures that the output decays fast enough to belong to \( \mathcal{S}(\mathbb{R}) \). It is not difficult to show that the convolutional definition of \( I_{\omega_0}^* \) given by (5.16) does not depend on the specific choice of the impulse response within the class of admissible LSI inverses of \( P_{j_0} \) on \( \mathcal{A}_{j_0} \). Then, we may simplify the notation by writing

\[ I_{\omega_0}^* \varphi(r) = (\rho_{j_0}',+ * \varphi)(r) - \hat{\varphi}(-\omega_0) \rho_{j_0},(r), \quad (5.17) \]

for any Green’s function \( \rho_{j_0} \) of the operator \( P_{j_0} \). While the left inverse \( I_{\omega_0}^* \) fixes the decay, it fails to be a right inverse of \( P_{j_0}^* \), unless \( \varphi \in \mathcal{S}_{j_0} \) or, equivalently, \( \hat{\varphi}(-\omega_0) = 0 \).

The corresponding right inverse of \( P_{j_0} \) is provided by the adjoint of \( I_{\omega_0}^* \). It is identified via the scalar product manipulation

\[ \langle \varphi, I_{\omega_0}^* \varphi \rangle = \langle \varphi, P_{j_0}^{-1,*} \varphi \rangle - \hat{\varphi}(-\omega_0) \langle \varphi, \rho_{j_0}' \rangle \quad \text{(by linearity)} \]
\[ = (P_{j_0}^{-1} \varphi, \varphi) - (e^{i \omega_0 r}, \varphi) (P_{j_0}^{-1} \varphi)(0) \quad \text{(using (5.12))} \]
\[ = (P_{j_0}^{-1} \varphi, \varphi) - (e^{i \omega_0 r} P_{j_0}^{-1} \varphi)(0), \varphi \).

\(^4\) These operators are shift-invariant because they are defined by means of convolutions.
5. OPERATORS AND THEIR INVERSES

Since the above is equal to \( \langle I_{\omega_0} \phi, \phi \rangle \) by definition, we find that
\[
I_{\omega_0} \phi(r) = P_{\omega_0}^{-1} \phi(r) - e^{i\omega_0 r} (P_{\omega_0}^{-1} \phi)(0) \\
= (\rho_{\omega_0} \ast \phi)(r) - e^{i\omega_0 r} (\rho_{\omega_0} \ast \phi)(0),
\]
where \( \rho_{\omega_0} \) is defined by (5.13). The specificity of \( I_{\omega_0} \) is to impose the boundary condition \( s(0) = 0 \) on the output \( s = I_{\omega_0} \phi \), irrespective of the input function \( \phi \). This is achieved by the addition of a component that is in the null space of \( P_{\omega_0} \). This also explains why we may substitute \( \rho_{\omega_0} \) in (5.18) by any other Green’s function of \( P_{\omega_0} \), including the causal one given by (5.15).

In particular, for \( j\omega_0 = 0 \) (that is, for \( P_{\omega_0} = D \)), we have
\[
I_0 \phi(r) = \int_{-\infty}^{r} \phi(t) \, dt - \int_{-\infty}^{0} \phi(t) \, dt = \begin{cases} \int_{-\infty}^{r} \phi(t) \, dt & r \geq 0 \\ -\int_{-\infty}^{r} \phi(t) \, dt & r < 0, \end{cases}
\]
while the adjoint is given by
\[
I_0^* \phi(r) = \int_{-\infty}^{r} \phi(t) \, dt - \mathbb{I}_{[-\infty,0]}(r) \int_{-\infty}^{\infty} \phi(t) \, dt = \begin{cases} \int_{-\infty}^{\infty} \phi(t) \, dt & r \geq 0 \\ -\int_{-\infty}^{\infty} \phi(t) \, dt & r < 0. \end{cases}
\]
These are equivalent to the solution described in Section 5.1 (see Figure 5.1).

The Fourier-domain counterparts of (5.18) and (5.17) are
\[
\mathcal{I}_{\omega_0} \phi(r) = \int_{\mathbb{R}} \hat{\phi}(\omega) e^{i\omega r} \, d\omega \quad (5.19)
\]
\[
\mathcal{I}_{\omega_0}^* \phi(r) = \int_{\mathbb{R}} \hat{\phi}(\omega) - \hat{\phi}(\omega_0) e^{i\omega r} \, d\omega \quad (5.20)
\]
respectively. One can observe that the form of the numerator in both integrals is such that it tempers the singularity of the denominator at \( \omega = \omega_0 \). The relevance of these corrected inverse operators for the construction of stochastic processes is due to the following theorem.

**Theorem 16.** The operator \( \mathcal{I}_{\omega_0}^* \) defined by (5.17) is continuous from \( \mathcal{S}(\mathbb{R}) \) to \( \mathcal{S}(\mathbb{R}) \) and extends continuously to a linear operator \( \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R}) \). It is the dual of the operator \( \mathcal{I}_{\omega_0} \) defined by (5.18) in the sense that \( \langle \mathcal{I}_{\omega_0} \phi, \psi \rangle = \langle \phi, \mathcal{I}_{\omega_0}^* \psi \rangle \). Moreover,
\[
\mathcal{I}_{\omega_0} \phi(0) = 0 \quad \text{(zero boundary condition)}
\]
\[
\mathcal{I}_{\omega_0}^* (D - j\omega_0 I_d) \psi = \phi \quad \text{(left-inverse property)}
\]
\[
(D - j\omega_0 I_d) \mathcal{I}_{\omega_0}^* \phi = \phi \quad \text{(right-inverse property)}
\]
\[
\|\mathcal{I}_{\omega_0}^* \phi\|_{L^p} \leq C_p \|\phi\|_{L^p} \quad \text{\( (L^p)\)-stability)}
\]
for all \( \phi, \psi \in \mathcal{S}(\mathbb{R}) \).

The first part of the theorem follows from Proposition 13 below, which indicates that \( \mathcal{I}_{\omega_0}^* \) preserves rapid decay. The statements in the second part have either already been discussed or, as in the case of the penultimate one, are a direct consequence of the first part. The left-inverse property, for instance, follows from the fact that \( (D - j\omega_0 I_d)^* \psi \in \mathcal{S}(\mathbb{R}) \), which is the subspace of \( \mathcal{S}(\mathbb{R}) \) for which all the inverses of \( P_{\omega_0}^* \) are equivalent. The right-inverse property of \( \mathcal{I}_{\omega_0} \) is easily verified by applying \( P_{\omega_0} \) to (5.18).

To qualify the rate of decay of functions, we rely on the \( L_{\infty,\alpha} \) norm, defined as
\[
\|\phi\|_{L_{\infty,\alpha}} = \text{esssup}_{r \in \mathbb{R}} \left| \phi(r) (1 + |r|^\alpha) \right|.
\]
5.4. Unstable $N$th-order differential systems

Hence, the inclusion $\varphi \in L_{\infty,a}(\mathbb{R})$ is equivalent to

$$|\varphi(r)| \leq \frac{\|\varphi\|_{L_{\infty,a}}}{1+|r|^a} \quad \text{a.e.,}$$

which is to say that $\varphi$ has an algebraic decay of order $a$. We also recall that $\mathcal{S}(\mathbb{R})$ is the space of rapidly-decaying functions, which is the intersection of all $L_{\infty,a}(\mathbb{R})$ spaces with $a > 0$. The relevant embedding relations are $\mathcal{S}(\mathbb{R}) \subset \mathcal{S}_p(\mathbb{R}) \subset L_{\infty,a}(\mathbb{R}) \subset L_p(\mathbb{R})$ for any $a > 1$ and $p \geq 1$. Moreover, since $\mathcal{S}(\mathbb{R})$ has the strictest topology in the chain, a sequence that converges in $\mathcal{S}(\mathbb{R})$ is also convergent in $\mathcal{S}_p(\mathbb{R})$, $L_{\infty,a}(\mathbb{R})$, or $L_p(\mathbb{R})$.

**Proposition 13.** Let $I_{\omega_0}^*$ be the linear operator defined by (5.16). Then, for all $\varphi \in L_{\infty,a}(\mathbb{R})$ with $a > 1$, there exists a constant $C$ such that

$$\|I_{\omega_0}^*\varphi\|_{L_{\infty,a-1}} \leq C\|\varphi\|_{L_{\infty,a}}.$$  

Hence, $I_{\omega_0}^*$ is a continuous operator from $L_{\infty,a}(\mathbb{R})$ into $L_{\infty,a-1}(\mathbb{R})$ and, by restriction of its domain, from $\mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ or $\mathcal{S}_p(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$.

**Proof.** For $r < 0$, we rewrite (5.16) as

$$I_{\omega_0}^*\varphi(r) = I_{\omega_0}^*\varphi(r) - e^{-j\omega_0 r} \varphi(-\omega_0)$$

$$= \int_{-\infty}^{r} e^{-j\omega_0 (r-t)} \varphi(t) \, dt - e^{-j\omega_0 r} \int_{-\infty}^{\infty} e^{j\omega_0 t} \varphi(t) \, dt$$

$$= -e^{-j\omega_0 r} \int_{-\infty}^{r} e^{j\omega_0 t} \varphi(t) \, dt.$$

This implies that

$$\|I_{\omega_0}^*\varphi(r)\| = \left| \int_{-\infty}^{r} e^{j\omega_0 t} \varphi(t) \, dt \right|$$

$$\leq \int_{-\infty}^{r} |\varphi(t)| \, dt$$

$$\leq \int_{-\infty}^{r} \frac{\|\varphi\|_{L_{\infty,a}}}{1+|r|^a} \, dt \leq \left( \frac{2\alpha}{\alpha - 1} \right) \frac{\|\varphi\|_{L_{\infty,a}}}{1+|r|^a}$$

for all $r < 0$. For $r > 0$, $I_{\omega_0}^*\varphi(r) = \int_{-\infty}^{r} e^{-j\omega_0 (r-t)} \varphi(t) \, dt$ so that the above upper bound remains valid.

While $I_{\omega_0}^*$ is continuous over $\mathcal{S}(\mathbb{R})$, it is not shift-invariant. Moreover, it will generally spoil the global smoothness of the functions in $\mathcal{S}(\mathbb{R})$ to which it is applied due to the discontinuity at the origin that is introduced by the correction. By contrast, its adjoint $I_{\omega_0}$ preserves the smoothness of the input but fails to return functions that are rapidly decaying at infinity. This lack of shift-invariance and the slow growth of the output at infinity appear to be the price to pay for being able to solve unstable differential systems.

5.4.2 Higher-order differential operators with unstable shift-invariant inverses

Given that the operators $I_{\omega_0}^*$, $\omega_0 \in \mathbb{R}$, defined in Section 5.4.1 are continuous $\mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$, they may be composed to obtain higher-order continuous operators $\mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ that serve as left inverses of the corresponding higher-order differential operators in the sense
of the previous section. More precisely, given \((\omega_1, \ldots, \omega_K) \in \mathbb{R}^K\), we define the composite integration operator

\[
I_{(\omega_1, \ldots, \omega_K)} = I_{\omega_1} \circ \cdots \circ I_{\omega_K}
\]

whose adjoint is given by

\[
I^*_{(\omega_1, \ldots, \omega_K)} = (I_{\omega_1} \circ \cdots \circ I_{\omega_K})^* = I^*_{\omega_K} \circ \cdots \circ I^*_{\omega_1}.
\]

Putting everything together, with the definitions of Section 5.3.2, we arrive at the following corollary of Theorem 16:

**Corollary 3.** For \(\alpha = (\alpha_1, \ldots, \alpha_M) \in \mathbb{C}^M\) with \(\text{Re}(\alpha_n) \neq 0\) and \((\omega_1, \ldots, \omega_K) \in \mathbb{R}^K\), the \((M + K)\)th-order operator \(P^*_{(\alpha_1; \alpha_M)} I^*_{(\omega_1, \ldots, \omega_K)}\) maps \(\mathcal{S}\) continuously into \(\mathcal{S}\). It is a left inverse of \(P^*_{(\alpha_1; \alpha_M)} P^*_{(\alpha_1; \alpha_M)}\) in the sense that

\[
P^{-1*}_{(\alpha_1; \alpha_M)} I^*_{(\omega_1, \ldots, \omega_K)} P^*_{(\alpha_1; \alpha_M)} P^*_{(\alpha_1; \alpha_M)} \varphi = \varphi
\]

for all \(\varphi \in \mathcal{S}\).

We shall now use this result to solve the differential equation (5.10) in the non-stable scenario. To that end, we order the poles in such a way that the unstable ones come last with \(\alpha_{N-K+m} = \omega_m, 1 \leq m \leq K\), where \(K\) is the number of purely imaginary poles. We thus specify the right-inverse operator

\[
L^{-1} = I_{(\omega_K, \ldots, \omega_1)} P^{-1}_{(\alpha_{N-K}; \alpha_1)} q_M(D),
\]

which we then apply to \(w\) to obtain the solution \(s = L^{-1} w\). In effect, by applying \(I_{(\omega_K, \ldots, \omega_1)}\) last, we are also enforcing the \(K\) linear boundary conditions

\[
\begin{align*}
& s(0) = 0 \\
& P_{\omega_2} \cdots P_{\omega_K} s(0) = 0 \quad \text{etc.} \\
& \vdots \\
& P_{\omega_2} \cdots P_{\omega_K} s(0) = 0.
\end{align*}
\]

To show that \(s = L^{-1} w\) is a consistent solution, we proceed by duality and write

\[
\langle \varphi, P_{(\alpha_1; \alpha_{N-K})} P_{(\omega_1; \omega_K)} s \rangle = \langle \varphi, P_{(\alpha_1; \alpha_{N-K})} P_{(\omega_1; \omega_K)} L^{-1} w \rangle = \langle P^*_{(\omega_1; \omega_K)} P^*_{(\alpha_1; \alpha_{N-K})} \varphi, \quad I_{(\omega_K, \ldots, \omega_1)} P^{-1}_{(\alpha_{N-K}; \alpha_1)} q_M(D) w \rangle
\]

\[
= \langle P^{-1*}_{(\alpha_{N-K}; \alpha_1)} I^*_{(\omega_K, \ldots, \omega_1)} P^*_{(\omega_1; \omega_K)} P^*_{(\alpha_1; \alpha_{N-K})} \varphi, \quad q_M(D) w \rangle
\]

\[
= \langle \varphi, q_M(D) w \rangle,
\]

where we have made use of Corollary 3. This proves that \(s\) satisfies the differential equation (5.10) with driving term \(w\), subject to the boundary conditions (5.23).
5.4. Unstable Nth-order differential systems

5.4.3 Generalized boundary conditions

In the resolution method presented so far, the inverse operator \( I_{0,0} \) was designed to impose zero boundary conditions at the origin. In more generality, one may consider inverse operators \( I_{\omega_0, \varphi_0} \) that incorporate conditions of the form

\[
\langle \varphi_0, I_{\omega_0, \varphi_0} w \rangle = 0
\]  

(5.24)

on the solution \( s = I_{\omega_0, \varphi_0} w \). This leads to the definition of the right-inverse operator (see [UTAKed, Appendix A])

\[
I_{\omega_0, \varphi_0} \varphi(r) = (\rho_{\omega_0} * \varphi)(r) - e^{i \omega_0 r} \left( \frac{\rho_{\omega_0} * \varphi \varphi_0}{\varphi_0(-\omega_0)} \right),
\]

(5.25)

where \( \rho_{\omega_0} \) is a Green’s function of \( P_{\omega_0} \) and \( \varphi_0 \) is some given rapidly-decaying function such that \( \varphi_0(-\omega_0) \neq 0 \). In particular, if we set \( \omega_0 = 0 \) and \( \varphi_0 = \delta \), we recover the scale-invariant integrator \( I_0 = I_{0,\delta} \) that was used in our introductory example (Section 5.1) to provide the connection with the classical theory of Lévy processes. The Fourier-domain counterpart of (5.25) is

\[
I_{\omega_0, \varphi_0} \varphi(r) = \int_{\mathbb{R}} \hat{\varphi}(\omega) \frac{e^{i \omega_0 r} \varphi_0(-\omega_0)}{\varphi_0(-\omega_0)} \frac{d\omega}{2\pi},
\]

(5.26)

The above operator is well-defined pointwise for any \( \varphi \in L_1(\mathbb{R}) \). Moreover, it is a right inverse of \( (D - \omega_0 I_d) \) on \( \mathcal{S}(\mathbb{R}) \) because the regularization in the numerator amounts to a sinusoidal correction that is in the null space of the operator. The adjoint of \( I_{\omega_0, \varphi_0} \) is specified by the Fourier-domain integral

\[
I^*_{\omega_0, \varphi_0} \varphi(r) = \int_{\mathbb{R}} \left( \hat{\varphi}(\omega) - \frac{\hat{\varphi}(-\omega_0)}{\varphi_0(-\omega_0)} \hat{\varphi}_0(\omega) \right) e^{i \omega_0 r} \frac{d\omega}{2\pi},
\]

(5.27)

which is non-singular, thanks to the regularization in the numerator. The beneficial effect of this adjustment is that \( I^*_{\omega_0, \varphi_0} \) is \( \mathcal{S} \)-continuous and \( L_p \)-stable, unlike its more conventional shift-invariant counterpart \( P_{\omega_0}^{-1} \). The time-domain counterpart of (5.27) is

\[
I^*_{\omega_0, \varphi_0} \varphi(r) = (\rho^*_{\omega_0} * \varphi)(r) - \frac{\hat{\varphi}(-\omega_0)}{\varphi_0(-\omega_0)} \varphi_0 \varphi^*_{\omega_0}(r),
\]

(5.28)

where \( \rho_{\omega_0} \) is a Green’s function of \( P_{\omega_0} \). This relation is very similar to (5.17), with the notable difference that the second term is convolved by \( \varphi_0 \). This suggests that we can restore the smoothness of the output by picking a kernel \( \varphi_0 \) with a sufficient degree of differentiability. In fact, by considering a sequence of such kernels in \( \mathcal{S}(\mathbb{R}) \) that converge to the Dirac distribution (in the weak sense), we can specify a left-inverse operator that is arbitrarily close to \( I^*_{\omega_0} \) and yet \( \mathcal{S} \)-continuous.

While the imposition of generalized boundary conditions of the form (5.24) has some significant implication on the statistical properties of the signal (non-stationary behavior), it is less of an issue for signal processing because of the use analysis tools (wavelets, finite-difference operators) that stationarize these processes—in effect, filtering out the null-space components—so that the traditional tools of the trade remain applicable. Therefore, to simplify the presentation, we shall only consider boundary conditions at zero in the sequel, and work with the operators \( I^*_{\omega_0} \) and \( I_{\omega_0} \).
5. Operators and their inverses

5.5 Fractional-order operators

5.5.1 Fractional derivatives in one dimension

In one dimension, we consider the general class of all homogeneous and shift-invariant operators and their inverses. These were studied in [UB07, BU07]. To motivate their definition, let us recall that the $n$th-order derivative $D^n$ corresponds to the Fourier multiplier $(j\omega)^n$. This suggests the following fractional extension, going back to Liouville, whereby the exponent $n$ is replaced by a nonnegative real number $\gamma$:

$$D^\gamma \varphi(r) = \int_{\mathbb{R}} (j\omega)^\gamma \hat{\varphi}(\omega) e^{j\omega r} \frac{d\omega}{2\pi}.$$  

This definition is further generalized in the next proposition, which gives a complete characterization of scale-invariant convolution operators in 1-D.

**Proposition 14** (see [UB07, Proposition 2]). The complete family of 1-D scale-invariant convolution operators of order $\gamma \geq 0$ is given by the fractional derivative $D^\gamma$, whose Fourier-based definition is

$$D^\gamma \varphi(r) = \int_{\mathbb{R}} (j\omega)^{\gamma+i\tau} (-j\omega)^{\gamma-i\tau} \hat{\varphi}(\omega) e^{j\omega r} \frac{d\omega}{2\pi},$$

where $\hat{\varphi}(\omega)$ is the 1-D Fourier transform of $\varphi(r)$.

These operators are endowed with a semi-group property: they satisfy the composition rule $D^\gamma D^\tau = D^{\gamma + \tau}$ for $\gamma, \tau \in \mathbb{R}$ and $\tau, \tau' \in \mathbb{R}$. The parameter $\tau$ is a phase factor that yields a progressive transition between the purely causal derivative $D^\gamma = \partial_x^\gamma$ and its anti-causal counterpart $D^\tau = \partial_x^{-\tau}$, which happens to be the adjoint of the former. We also note that $D^0$ is equivalent to the fractional Hilbert transform operator $\mathcal{H}_\tau$ investigated in [CU10]. This implies that the fractional derivatives of order $\gamma$, $D^\gamma = \partial_x^{\gamma+i\tau} = D^\gamma \mathcal{H}_\tau^{-\gamma}$, are all related to $D^\tau$ via a fractional Hilbert transform, which is a unitary operator that essentially acts as a shift operator on the oscillatory part of a wavelet.

The Fourier multiplier associated with $D^\gamma$ is equal in amplitude to $|\omega|^\gamma$. This observation, in combination with Theorem 19, indicates that $L_p$-stable inverses of fractional derivatives are the ones given by Proposition 15.

**Proposition 15.** The $L_p$-stable left and right inverses of $D^\gamma$, applied to test and generalized functions, respectively, are given by

$$D^{-\gamma}_{-\tau,p} \varphi(r) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\check{\phi}(\omega) - \sum_{k \leq \gamma - 1 + \frac{1}{p} \frac{|j\omega|^{k}}{k!}}}{(-j\omega)^{\gamma - \tau} (j\omega)^{\gamma + \tau}} e^{j\omega r} d\omega,$$

$$D^{-\gamma}_{-\tau,p} \varphi(r) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{j\omega r} - \sum_{k \leq \gamma - 1 + \frac{1}{p} \frac{|j\omega|^{k}}{k!}}}{(-j\omega)^{\gamma - \tau} (j\omega)^{\gamma + \tau}} \hat{\varphi}(\omega) d\omega,$$

where $\check{\phi}^{(k)}(0) = \frac{d^k \hat{\varphi}(\omega)}{d\omega^k} \big|_{\omega=0}$.

5.5.2 Fractional Laplacians

The fractional Laplacian of order $\gamma \geq 0$ is defined by the inverse Fourier integral

$$(-\Delta)^{\gamma/2} \varphi(r) = \int_{\mathbb{R}^d} \|\omega\|^\gamma \hat{\varphi}(\omega) e^{j\omega (r,r)} \frac{d\omega}{(2\pi)^d}.$$
where $\hat{\varphi}(\omega)$ is the $d$-dimensional Fourier transform of $\varphi(r)$. For $\gamma = 2$, this characterization coincides with the classical definition of the negative Laplacian; $-\Delta = -\sum \partial_i^2$.

In slightly more generality, we obtain a complete characterization of homogeneous shift- and rotation-invariant operators and their inverses in terms of convolutions with homogeneous rotation-invariant distributions, as given in Theorem 17. The idea and definitions may be traced back to [Duc77, GS68, Hör80].

**Theorem 17** ([Taf11], Corollary 2.ac). Any continuous linear operator $\mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ that is simultaneously shift- and rotation-invariant, and homogeneous or scale-invariant of order $\gamma$ in the sense of Definition 7.5, is a multiple of the operator

$$L^\gamma : \varphi \mapsto \rho^{-1-d} \ast \varphi = (2\pi)^\frac{d}{2} \mathcal{F}^{-1} \{ \rho^\gamma \hat{\varphi} \}$$

where $\rho^\gamma, \gamma \in \mathbb{C}$, is the distribution defined by

$$\rho^\gamma(\omega) = \frac{\|\omega\|^\gamma}{2^\frac{\gamma}{2} \Gamma\left(\frac{\gamma+d}{2}\right)}$$

with the property that

$$\mathcal{F}\{\rho^\gamma\}(\omega) = (2\pi)^\frac{d}{2} \rho^{-1-d}(\omega)$$

($\Gamma$ denotes the Gamma function.)

Note that $L^\gamma$ is self-adjoint, with $L^{\gamma^*} = L^\gamma$.

As is clear from the definition of $\rho^\gamma$, for $\gamma \neq -d, -d-2, -d-4, \ldots$, $L^\gamma$ is simply a renormalized version of the fractional Laplacian introduced earlier. For $\gamma = -d - 2m, m = 0, 1, 2, 3, \ldots$, where the $\Gamma$ function in the denominator of $\rho^\gamma$ has a pole, $\rho^\gamma$ is proportional to $(-\Delta)^{m}\delta$, while the previous definition of the fractional Laplacian without normalization does not define a scale-invariant operator.

Also note that when $\text{Re}(\gamma) > -d$ (respectively, $\text{Re}(\gamma) \leq -d$), $\rho^\gamma$ (its Fourier transform, respectively) is singular at the origin. This singularity is resolved in the manner described in Appendix A, which is equivalent to the analytical continuation of the formula

$$\Delta \rho^\gamma = \gamma \rho^{\gamma-2},$$

initially valid for $\text{Re}(\gamma) - 2 > -d$, in the variable $\gamma$. For the details of the previous two observations we refer the reader to Appendix A and Tafti [Taf11, Section 2.2].

Finally, it is important to remark that unlike integer-order operators, unless $\gamma$ is a positive even integer, the image of $\mathcal{S}(\mathbb{R}^d)$ under $L^\gamma$ is not contained in $\mathcal{S}(\mathbb{R}^d)$. Specifically, while for any $\varphi \in \mathcal{S}$, $L^\gamma \varphi$ is always an infinitely differentiable regular function, in the case of $\gamma \neq 2m, m = 1, 2, \ldots$, it may have slow (polynomial) decay or growth.

Put more simply, the fractional Laplacian of a Schwartz’ function is not in general a Schwartz’ function.

### 5.5.3 $L^p$-stable inverses

From the identity

$$\rho^\gamma(\omega) \rho^{-\gamma}(\omega) = \frac{1}{\Gamma\left(\frac{d+\gamma}{2}\right) \Gamma\left(\frac{d-\gamma}{2}\right)} \text{ for } \omega \neq 0,$$

5. The difference in factors of $(2\pi)^\frac{d}{2}$ and $(2\pi)^d$ between the formulas given here and in Tafti [Taf11] is due to different normalizations used in the definition of the Fourier transform.
we conclude that up to normalization, $L^{-\gamma}$ is the inverse of $L^\gamma$ on the space of Schwartz’ test functions with vanishing moments, in particular for $\text{Re}(\gamma) > -d$. This inverse for $\text{Re}(\gamma) > -d$ can further be extended to a shift-invariant left inverse of $L^\gamma$ acting on Schwartz’ functions. However, as was the case in Section 5.4.1, this shift-invariant inverse generally does not map $\mathcal{S}(\mathbb{R}^d)$ into a given $L_p(\mathbb{R}^d)$ space, and is therefore not suitable for defining generalized random fields in $\mathcal{S}'(\mathbb{R}^d)$.

The problem exposed in the previous paragraph is once again overcome by defining a “corrected” left inverse. However, unlike previously, it is not possible here to have a single left-inverse operator that maps $\mathcal{S}(\mathbb{R}^d)$ into the intersection of all $L_p(\mathbb{R}^d)$ spaces, $p \geq 1$. Instead, we shall need to define a separate left-inverse operator for each $L_p(\mathbb{R}^d)$ space we are interested in. Under the constraints of homogeneity and rotation invariance, such “non-shift-invariant” left inverses are identified in the following theorem.

**Theorem 18** ([Taf11], Theorem 2.aq and Corollary 2.am). The operator

$$L_p^{-\gamma*} : \varphi \mapsto \rho^{\gamma-d} \ast \varphi - \sum_{|k| \leq |\text{Re}(\gamma) + \frac{d}{p}|} \frac{\partial_k \rho^{\gamma-d}}{k!} \int_{\mathbb{R}^d} y^k \varphi(y) \, dy \quad (5.30)$$

with $\text{Re}(\gamma) + \frac{d}{p} \neq 1, 2, 3, \ldots$ is rotation-invariant and homogeneous of order $(-\gamma)$ in the sense of Definition 7.5. It maps $\mathcal{S}(\mathbb{R}^d)$ continuously into $L_p(\mathbb{R}^d)$.

The adjoint of $L_p^{-\gamma*}$ is given by

$$L_p^{-\gamma} : \varphi \mapsto \rho^{\gamma-d} \ast \varphi - \sum_{|k| \leq |\text{Re}(\gamma) + \frac{d}{p}|} r^k \frac{\partial_k L^{-\gamma} \varphi(0)}{k!}.$$

If we exclude the cases where the denominator of (5.29) has a pole, we may normalize the above operators to find left and right inverses corresponding to the fractional Laplacian $(-\Delta)^{\gamma/2}$. The next theorem gives an equivalent Fourier-domain characterization of these operators.

**Theorem 19** (cf. [SU12, Theorem 3.7]). Let $L_p^{\gamma*}$ with $p \geq 1$ and $\gamma > 0$ be the isotropic fractional integral operator defined by

$$L_p^{\gamma*} \varphi(r) = \int_{\mathbb{R}^d} \frac{\partial_k \hat{\varphi}(0) \omega^k}{||\omega||^\gamma} c_{(\omega,r)} \, d\omega \quad (5.31)$$

Then, under the condition that $\gamma \neq 2, 4, \ldots$ and $\gamma + \frac{d}{p} \neq 1, 2, 3, \ldots$, $L_p^{\gamma*}$ is the unique left inverse of $(-\Delta)^{\gamma/2}$ that is scale-invariant and $L_p$-stable in the sense of (4.25). The adjoint operator $L_p^{\gamma}$, which is the proper scale-invariant right inverse of $(-\Delta)^{\gamma/2}$, is given by

$$L_p^{\gamma} \varphi(r) = \int_{\mathbb{R}^d} \frac{\partial_k \varphi(0) \omega^k}{||\omega||^\gamma} c_{(\omega,r)} \, d\omega \quad (5.32)$$

The fractional integral operators $L_p^{\gamma}$ and $L_p^{\gamma*}$ are both scale-invariant of order $(-\gamma)$, but they are not shift-invariant.

---

6. Here we exclude the cases where the Gamma functions in the denominator have poles, namely where their argument is a negative integer. For details see [Taf11, 2.ah].
5.6 Discrete convolution operators

We conclude this chapter by providing a few basic results on discrete convolution operators and their inverses. These will turn out to be helpful for establishing the existence of certain spline interpolators which are required for the construction of operator-like wavelet bases in Chapter 6 and for the representation of autocorrelation functions in Chapter 7.

The convention in this book is to use square brackets to index sequences. This allows one to distinguish them from functions of a continuous variable. In other words, \( h(\cdot) \) or \( h(r) \) stands for a function defined on a continuum, while \( h[\cdot] \) or \( h[k] \) denotes some discrete sequence. The notation \( h[\cdot] \) is often simplified to \( h \) when the context is clear.

The discrete convolution between two sequences \( h[\cdot] \) and \( a[\cdot] \) over \( \mathbb{Z}^d \) is defined as

\[
(h * a)[n] = \sum_{m \in \mathbb{Z}^d} h[m]a[n - m].
\]

This convolution describes how a digital filter with discrete impulse response \( h \) acts on some input sequence \( a \). If \( h \in \ell_1(\mathbb{Z}^d) \), then the map \( a[\cdot] \rightarrow (h * a)[\cdot] \) is a continuous operator \( \ell_p(\mathbb{Z}^d) \rightarrow \ell_p(\mathbb{Z}^d) \). This follows from Young’s inequality for sequences:

\[
\|h * a\|_{\ell_p} \leq \|h\|_{\ell_1} \|a\|_{\ell_p},
\]

where

\[
\|a\|_{\ell_p} = \begin{cases} 
\left( \sum_{n \in \mathbb{Z}^d} |a(n)|^p \right)^{\frac{1}{p}} & \text{for } 1 \leq p < \infty \\
\sup_{n \in \mathbb{Z}^d} |a(n)| & \text{for } p = \infty.
\end{cases}
\]

Note that the condition \( h \in \ell_1(\mathbb{Z}^d) \) is the discrete counterpart of the (more involved) TV condition in Theorem 4. As in the continuous formulation, it is necessary and sufficient for stability in the extreme cases \( p = 1, +\infty \). Establishing (5.34) for those two cases is a routine calculation—one can then invoke Theorem 3 (Riesz-Thorin) to interpolate the bound for \( 1 < p < +\infty \). Another way of understanding the result is by observing that \( \ell_\infty \)-stability implies all the other forms because of the embedding \( \ell_p(\mathbb{Z}^d) \subset \ell_q(\mathbb{Z}^d) \) for any \( 1 \leq p < q \leq \infty \), which goes hand-in-hand with the basic norm inequality

\[
\|a\|_{\ell_p} \geq \|a\|_{\ell_q} \geq \|a\|_{\ell_\infty}
\]

for all \( a \in \ell_p(\mathbb{Z}^d) \subset \ell_q(\mathbb{Z}^d) \subset \ell_\infty(\mathbb{Z}^d) \).

In the Fourier domain, (5.33) maps into the multiplication of the discrete Fourier transforms of \( h \) and \( a \) as

\[
\hat{b}[n] = (h * a)[n] \quad \Leftrightarrow \quad B(\omega) = H(\omega)A(\omega)
\]

where we are using capitalized letters to denote the discrete-time Fourier transforms of the underlying sequences. Specifically, we have that

\[
B(\omega) = \mathcal{F}_d[b](\omega) = \sum_{k \in \mathbb{Z}^d} b[k]e^{-j\omega k},
\]

which is \( 2\pi \)-periodic, while the corresponding inversion formula is

\[
b[n] = \mathcal{F}_d^{-1}[B][n] = \int_{[-\pi,\pi]^d} B(\omega)e^{j\omega n} \frac{d\omega}{(2\pi)^d}.
\]

The stability condition \( h \in \ell_1(\mathbb{Z}^d) \) ensures that the frequency response \( H(\omega) \) of the digital filter \( h \) is bounded and continuous.

The task of specifying discrete inverse filters is greatly facilitated by a theorem, known as Wiener’s lemma, which ensures that the inverse convolution operator is \( \ell_p \)-stable whenever the frequency response of the original filter is non-vanishing.
5. OPERATORS AND THEIR INVERSES

**Theorem 20** (Wiener’s lemma). Let \( H(\omega) = \sum_{k \in \mathbb{Z}^d} h[k] e^{-j(\omega, k)} \), with \( h \in \ell_1(\mathbb{Z}^d) \), be a stable discrete Fourier multiplier such that \( H(\omega) \neq 0 \) for \( \omega \in [-\pi, \pi]^d \). Then, \( G(\omega) = 1 / H(\omega) \) has the same property in sense that it can be written as \( 1 / H(\omega) = \sum_{k \in \mathbb{Z}^d} g[k] e^{-j(\omega, k)} \) with \( g \in \ell_1(\mathbb{Z}^d) \).

The so-defined filter \( g \) identifies a stable convolution inverse of \( h \) with the property that

\[
(g * h)[n] = (h * g)[n] = \delta[n]
\]

where

\[
\delta[n] = \begin{cases} 
1 & \text{for } n = 0 \\
0 & \text{for } n \in \mathbb{Z}^d \setminus \{0\}
\end{cases}
\]

is the Kronecker unit impulse.

5.7 Notes

The specification of \( L_p \)-stable convolution operators is a central topic in harmonic analysis [SW71, Gra08]. The basic result in Proposition 11 relies on Young’s inequality with \( q = r = 1 \). The complete class of functions that result in \( \mathcal{S} \)-continuous convolution kernels is provided by the inverse Fourier transform of the space of smooth slowly-increasing Fourier multipliers which play a crucial role in the theory of generalized functions [Sch66]. They are an extension upon \( \mathcal{S}(\mathbb{R}^d) \) in the sense that they also contain point distributions such as \( \delta \) and its derivatives.

The operational calculus that is used for solving ordinary differential equations (ODEs) can be traced to Heaviside who also introduced the symbol \( D \) for the derivative operator. It was initially met with skepticism because Heaviside’s exposition lacked rigor. Nowadays, the preferred method for solving ODEs is based on the Laplace transform or the Fourier transform. The operator-based formalism that is presented Section 5.3 is a standard application of distribution theory and Green’s functions [Zem10].

The main advantage of the framework of generalized stochastic processes is that it greatly facilitates the transposition of the deterministic methods to the solution of linear stochastic differential equations (SDEs). The procedure is fairly straightforward in the stable scenario since all underlying operators are \( \mathcal{S} \)-continuous. As we saw in sections 5.4 and 5.5.1, the extension to the unstable regime is possible as well, but more challenging mathematically, which explains why the developments in this area are more recent. The operator-based approach for solving unstable SDEs was initiated in [BU07] in an attempt to link splines and fractals. It was further refined in another series of papers [UTAKed, UTsed, TVDVU09, UT11, SU12].
Chapter 6

Splines and wavelets

A fundamental aspect of our formulation is that the whitening operator $L$ is naturally tied to some underlying B-spline function, which will play a crucial role in the sequel. The spline connection also provides a strong link with wavelets \cite{UB03}.

In this chapter, we review the foundations of spline theory and show how one can construct B-spline basis functions and wavelets that are tied to some specific operator $L$. The chapter starts with a gentle introduction to wavelets that exploits the analogy with Legos blocks. This naturally leads to the formulation of a multiresolution analysis of $L^2(\mathbb{R})$ using piecewise-constant functions and a \textit{de visu} identification of Haar wavelets. We then proceed in Section 6.2 with a formal definition of our generalized brand of splines—the cardinal L-splines—followed by a detailed discussion of the fundamental notion of Riesz basis. We then systematically cover the first-order operators with the construction of exponential B-splines and wavelets, which have the convenient property of being orthogonal. We then address the theory in its full generality and present two generic methods for constructing B-spline basis functions (Section 6.4) and semi-orthogonal wavelets (Section 6.5). The pleasing aspect is that these results apply to the whole class of shift-invariant differential operators $L$ whose null space is finite-dimensional (possibly trivial), which are precisely those that can be safely inverted to specify sparse stochastic processes.

6.1 From Legos to wavelets

It is instructive to get back to our introductory example of piecewise-constant splines in Chapter 1 (§1.3) and to show how these are naturally connected to wavelets. The fundamental idea in wavelet theory is to construct a series of fine-to-coarse approximations of a function $s(r)$ and to exploit the structure of the multiresolution approximation errors, which are orthogonal across scale. Here, we shall consider a series of approximating signals $\{s_i\}_{i \in \mathbb{Z}}$, where $s_i$ is a piecewise-constant spline with knots positioned on $2^i \mathbb{Z}$. These multiresolution splines are represented by their B-spline expansion

$$s_i(r) = \sum_{k \in \mathbb{Z}} c_i(k) \phi_{i,k}(r),$$

where the B-spline basis functions (rectangles) are dilated versions of the cardinal ones by a factor of $2^i$

$$\phi_{i,k}(r) = \beta_i\left(\frac{r - 2^i k}{2^i}\right) = \begin{cases} 1, & \text{for } r \in [2^i k, 2^i (k + 1)) \\ 0, & \text{otherwise.} \end{cases}$$

93
The variable \( i \) is the scale index that specifies the resolution (or knot spacing) \( a = 2^i \), while the integer \( k \) encodes for the spatial location. The B-spline of degree 0, \( \phi = \phi_{0,0} = \beta_0^2 \), is the scaling function of the representation. Interestingly, it is the identification of a proper scaling function that constitutes the most fundamental step in the construction of a wavelet basis of \( L_2(\mathbb{R}) \).

**Definition 18** (Scaling function). \( \phi \in L_2(\mathbb{R}) \) is a valid scaling function if and only if it satisfies the following three properties:

- Two-scale relation

\[
\phi(r/2) = \sum_{k \in \mathbb{Z}} h(k)\phi(r-k),
\]

where the sequence \( h \in \ell_1(\mathbb{Z}) \) is the so-called refinement mask.

- Partition of unity

\[
\sum_{k \in \mathbb{Z}} \phi(r-k) = 1
\]

- The set of functions \( \{\phi(-k)\}_{k \in \mathbb{Z}} \) forms a Riesz basis.

In practice, a given brand of (orthogonal) wavelets (e.g., Daubechies or splines) is often summarized by its refinement filter \( h \) since the latter uniquely specifies \( \phi \), subject to the admissibility constraints (6.4) and \( \phi \in L_2(\mathbb{R}) \). In the case of the B-spline of degree 0, we have that \( h(k) = \delta(k) + \delta(k-1) \), where

\[
\delta(k) = \begin{cases} 
1, & \text{for } k = 0 \\
0, & \text{otherwise}
\end{cases}
\]

is the discrete Kronecker impulse. This translates into what we jokingly refer to as the Lego-Duplo relation

\[
\beta_0^2(r/2) = \beta_0^2(r) + \beta_1^2(r-1).
\]

The fact that \( \beta_0^2 \) satisfies the partition of unity is obvious. Likewise, we already observed in Chapter 1 that \( \beta_0^2 \) generates an orthogonal system which is a special case of a Riesz basis.

By considering the rescaled version of such a basis, we specify the subspace of splines at scale \( i \) as

\[
V_i = \left\{ s_i(r) = \sum_{k \in \mathbb{Z}} c_{i,k} \phi_{i,k}(r) : c_{i,k} \in \ell_2(\mathbb{Z}) \right\} \subset L_2(\mathbb{R})
\]

which, in our example, contains all the finite-energy functions that are piecewise-constant on the intervals \( [2^i k, 2^i(k+1)] \) with \( k \in \mathbb{Z} \). The two-scale relation (6.3) implies that the basis functions at scale \( i = 1 \) are contained in \( V_0 \) (the original space of cardinal splines) and, by extension, in \( V_i \) for \( i \leq 0 \). This translates into the general inclusion property \( V_i \subset V_j \) for any \( i' > i \), which is fundamental to the theory. A subtler point is that the closure of \( \bigcup_{i \in \mathbb{Z}} V_i \) is equal to \( L_2(\mathbb{R}) \), which follows from the fact that any finite-energy function can be approximated arbitrarily well by a piecewise-constant spline when the sampling step \( 2^i \) tends to zero (\( i \to -\infty \)). The necessary and sufficient condition for this asymptotic convergence is the partition of unity (6.4), which ensures that the representation is complete.

---

1. The Duplos are the larger-scale versions of the Lego building blocks and are more suitable for smaller children to play with. The main point of the analogy with wavelets is that Legos and Duplos are compatible; they can be combined to build more complex shapes. The enabling property is that a Duplo is equivalent to two smaller Legos placed next to each other, as expressed by (6.5)
6.1. From Legos to wavelets

Wavelets: Haar transform revisited

\[ r_i(x) = s_{i-1}(x) - s_i(x) \]

Figure 6.1: Multiresolution signal analysis using piecewise-constant splines with a dyadic scale progression. Left: multiresolution pyramid. Right: error signals between two successive levels of the pyramid.

Having set the notation and specified the underlying hierarchy of function spaces, we now proceed with the multiresolution approximation procedure starting from the fine-scale signal \( s_0(x) \), as illustrated in Figure 6.1. Given the sequence \( c_0[\cdot] \) of fine-scale coefficients, our task is to construct the best spline approximation at scale 1 which is specified by its B-spline coefficients \( c_1[\cdot] \) in (6.1) with \( i = 1 \). It is easy to see that the minimum-error solution (orthogonal projection of \( s_0 \) onto \( V_1 \)) is obtained by taking the mean of two consecutive samples. The procedure is then repeated at the next coarser scale and so forth until one reaches the bottom of the pyramid, as shown on the left-hand side of Figure 6.1. The description of this coarsening algorithm is

\[
c_i[k] = \frac{1}{2} c_{i-1}[2k] + \frac{1}{2} c_{i-1}[2k+1] = (c_{i-1} * \hat{h})[2k]. \tag{6.6}
\]

It is run recursively for \( i = 1, \ldots, i_{\text{max}} \) where \( i_{\text{max}} \) denotes the bottom level of the pyramid. The outcome is a multiresolution analysis of our input signal \( s_0 \).

In order to uncover the wavelets, it is enlightening to look at the residual signals \( r_i(r) = s_{i-1}(r) - s_i(r) \in V_{i-1} \) on the right of Figure 6.1. While these are splines that live at the same resolution as \( s_{i-1} \), they actually have half the apparent degrees of freedom. These error signals exhibit a striking sign-alternation pattern due to the fact that two consecutive samples \( (c_{i-1}[2k], c_{i-1}[2k+1]) \) are at an equal distance from their mean value \( (c_i[k]) \). This suggests rewriting the residuals more concisely in terms of oscillating basis functions (wavelets) at scale \( i \), like

\[
r_j(r) = s_{j-1}(r) - s_j(r) = \sum_{k \in \mathbb{Z}} d_j[k] \psi_{j,k}(r), \tag{6.7}
\]

where the (non-normalized) Haar wavelets are given by

\[
\psi_{j,k}(r) = \psi_{\text{Haar}} \left( \frac{r - 2^j k}{2^i} \right)
\]
with the Haar wavelet being defined by (1.19). The wavelet coefficients \( d_i[k] \) are given by the consecutive half differences

\[
d_i[k] = \frac{1}{2} c_{i-1}[2k] - \frac{1}{2} c_{i-1}[2k+1] = (c_{i-1} * g)[2k].
\] (6.8)

More generally, since the wavelet template at scale \( i = 1 \), \( \psi_{1,0} \in V_0 \), we can write

\[
\psi(r/2) = \sum_{k \in Z} g[k] \phi(r-k)
\] (6.9)

which is the wavelet counterpart of the two-scale relation (6.3). In the present example, we have \( g[k] = (-1)^k h[k] \), which is a general relation that is characteristic of an orthogonal design. Likewise, in order to gain in generality, we have chosen to express the decomposition algorithms (6.6) and (6.8) in terms of discrete convolution (filtering) and downsampling operations where the corresponding Haar analysis filters are \( \tilde{h}[k] = \frac{1}{2} h[-k] \) and \( \tilde{g}[k] = \frac{1}{2} (-1)^k h[-k] \). The Hilbert-space interpretation of this approximation process is that \( r_i \in W_i \), where \( W_i \) is the orthogonal complement of \( V_i \) in \( V_{i-1} \); that is, \( V_{i-1} = V_i + W_i \) with \( V_i \perp W_i \) (as a consequence of the orthogonal-projection theorem).

Finally, we can close the loop by observing that

\[
s_0(r) = s_{\text{max}}(r) + \sum_{i=1}^{\text{max}} \left( s_{i-1}(r) - s_i(r) \right)_r(r)
\]

\[
= \sum_{k \in Z} c_{\text{max}}[k] \phi_{\text{max},k}(r) + \sum_{i=1}^{\text{max}} \sum_{k \in Z} d_i[k] \psi_{i,k}(r),
\] (6.10)

which provides an equivalent, one-to-one representation of the signal in an orthogonal wavelet basis, as illustrated in Figure 6.2. More generally, we can push the argument to the limit and apply the decomposition to any finite-energy function

\[
\forall s \in L_2(\mathbb{R}), \quad s = \sum_{i \in Z} \sum_{k \in Z} d_i[k] \psi_{i,k},
\] (6.11)

where \( d_i[k] = \langle s, \psi_{i,k} \rangle_{L_2} \) and \( \{\psi_{i,k}\}_{i,k} \in L_2^2 \) is a suitable (bi-)orthogonal wavelet basis with the property that \( \langle \psi_{i,k}, \psi_{r,k} \rangle_{L_2} = \delta_{k-k',i-i'} \).

Remarkably, the whole process described above—except the central expressions in (6.6) and (6.8), and the equations explicitly involving \( \rho_0 \)—is completely generic and applicable to any other wavelet basis of \( L_2(\mathbb{R}) \) that is specified in terms of a wavelet filter \( g \) and a scaling function \( \phi \) (or, equivalently, an admissible refinement filter \( h \)). The bottom line is that the wavelet decomposition and reconstruction algorithm is fully described by the four digital filters \( (h, g, \tilde{h}, \tilde{g}) \) that form a perfect reconstruction filterbank. The Haar transform is associated with the shortest-possible filters. Its less favorable aspects are that the basis functions are discontinuous and that the scale-truncated error decays only like the first power of the sampling step \( a = 2^l \) (first order of approximation).

The fundamental point of our formulation is that the Haar wavelet is matched to the pure derivative operator \( D = \frac{d}{dr} \), which goes hand in hand with Lévy processes (see Chapter 1). In that respect, the critical observations relating to spline and wavelet theory are as follows:

- all piecewise-constant functions can be interpreted as \( D \)-splines;
- the Haar wavelet acts as a smoothed version of the derivative in the sense that \( \Psi_{\text{Haar}} = D \phi \), where \( \phi \) is an appropriate kernel (triangle function);
- the B-spline of degree 0 can be expressed as \( \rho_0 = \rho_D = D_0 D^{-1} \delta \), where the finite-difference operator \( D_0 \) is the discrete counterpart of \( D \).

We shall now show how these ideas are extendable to a much broader class of differential operators \( L \).
6.2 Basic concepts and definitions

6.2.1 Spline-admissible operators

Let $L : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ be a generic Fourier-multiplier operator with frequency response $\hat{L}(\omega)$. We further assume that $L$ has a continuous extension $L : \mathcal{X}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ to some larger space of functions $\mathcal{X}(\mathbb{R}^d)$ with $\mathcal{S}(\mathbb{R}^d) \subset \mathcal{X}(\mathbb{R}^d)$.

The null space of $L$ is denoted by $\mathcal{N}_L$ and defined as

$$\mathcal{N}_L = \{ p_{0}(r) : Lp_{0}(r) = 0 \}.$$  

The immediate consequence of $L$ being LSI is that $\mathcal{N}_L$ is shift-invariant as well, in the sense that $p_{n}(r) \in \mathcal{N}_L$ implies $p_{0}(r-r_{0}) \in \mathcal{N}_L$ for any $r_{0} \in \mathbb{R}^d$. We shall use this property to argue that $\mathcal{N}_L$ generally consists of generalized functions whose Fourier transforms are point distributions. In the space domain, they correspond to modulated polynomials, which are linear combinations of exponential monomials of the form $e^{i\omega_0 \cdot r} r^n$ with $\omega_0 \in \mathbb{R}^d$ and multi-index $n = (n_1, \ldots, n_d) \in \mathbb{N}^d$. It actually turns out that the existence of a single such element in $\mathcal{N}_L$ has direct implications on the structure and dimensionality of the underlying function space.

**Proposition 16** (Characterization of null space). If $L$ is LSI and $p_{n}(r) = e^{i\omega_0 \cdot r} r^n \in \mathcal{N}_L$ with $\omega_0 \in \mathbb{C}^d$, then $\mathcal{N}_L$ does necessarily include all exponential monomials of the form $p_{m}(r) = e^{i\omega_0 \cdot r} r^m$ with $m \leq n$. In addition, if $\mathcal{N}_L$ is finite-dimensional, it can only consist of atoms of that particular form.

**Proof.** The LSI property implies that $p_{n}(r-r_{0}) \in \mathcal{N}_L$ for any $r_{0} \in \mathbb{R}^d$. To make our point about the inclusion of the lower-order exponential polynomials in $\mathcal{N}_L$, we start by ex-
6. SPLINES AND WAVELETS

Expanding the scalar term \((r_i - r_{0,i})^n\) as

\[
(r_i - r_{0,i})^n = \sum_{m=0}^{n} \binom{n}{m} r_{0,i}^{m} (-1)^{n-m} r_{0,i}^{n-m} = \sum_{m+k=n} \frac{n!}{m! k!} (-1)^k r_{0,i}^k r_{0,i}^{n-k}.
\]

By proceeding in a similar manner with the other monomials and combining the results, we find that

\[
(r - r_0)^n = \sum_{m+k=n} \frac{n!}{m! k!} (-1)^k r_{0}^m r^k = \sum_{m+n} b_m(r_0) r^m
\]

with polynomial coefficients \(b_m(r_0)\) that depend upon the multi-index \(m\) and the shift \(r_0\).

Finally, we note that the exponential factor \(e^{(\omega_0 \cdot r)}\) can be shifted by \(r_0\) by simple multiplication with a constant (see (6.12) below). These facts taken together establish the structure of the underlying vector space. As for the last statement, we rely on the theory of Lie group that tells us that the only finite-dimensional collection of functions that are translation-invariant is made of exponential polynomials. The pure exponentials \(e^{(\omega_0 \cdot r)}\) with \(n = 0\) are special in that respect: They are the eigenfunctions of the shift operator in the sense that

\[
e^{(\omega_0 \cdot r - r_0)} = \lambda(r_0) e^{(\omega_0 \cdot r)} \quad \text{(6.12)}
\]

with the (complex) eigenvalue \(\lambda(r_0) = e^{i(\omega_0 \cdot r_0)}\), and hence the only elements that specify shift-invariant subspaces of dimension 1.

Since our formulation relies on the theory of generalized functions, we shall focus on the restriction of \(\mathcal{L}_1\) to \(\mathcal{S}'(\mathbb{R}^d)\). This rules out the exponential factors \(\omega_0 = \alpha_0 + j \omega_0\) in Proposition 16 with \(\alpha_0 \in \mathbb{R}^d \setminus \{0\}\) for which the Fourier-multiplier operator is not necessarily well-defined. We are then left with null-space atoms of the form \(e^{i(\omega_0 \cdot r)} r^n\) with \(\omega_0 \in \mathbb{R}^d\), which are functions of slow growth.

The next important ingredient is the Green’s function \(\rho_L\) of the operator \(L\). Its defining property is \(L \rho_L = \delta\), where \(\delta\) is the \(d\)-dimensional Dirac distribution. Since there are many equivalent Green’s functions of the form \(\rho_L + \rho_0\) where \(\rho_0 \in \mathcal{L}_1\) is an arbitrary component of the null space, we resolve the ambiguity by defining the (primary) Green’s function of \(L\) as

\[
\rho_L(r) = \mathcal{F}^{-1} \left\{ \frac{1}{L(\omega)} \right\}(r), \quad \text{(6.13)}
\]

with the requirement that \(\rho_L \in \mathcal{S}'(\mathbb{R}^d)\) is an ordinary function of slow growth. In other words, we want \(\rho_L(r)\) to be defined pointwise for any \(r \in \mathbb{R}^d\) and to grow no faster than a polynomial. The existence of the generalized inverse Fourier transform (6.13) imposes some minimal continuity and decay conditions on \(1/L(\omega)\) and also puts some restrictions on the number and nature of its singularities (e.g., the zeros of \(L(\omega)\)).

**Definition 19** (Spline admissibility). The Fourier-multiplier operator \(L : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)\) with frequency response \(L(\omega)\) is called spline admissible if (6.13) is well-defined and \(\rho_L(r)\) is an ordinary function of slow growth.

An important characteristic of spline-admissible operators is the rate of growth of their frequency response at infinity.

**Definition 20** (Order of a Fourier multipler). The Fourier-multiplier \(L(\omega)\) is of (asymptotic) order \(\gamma \in \mathbb{R}^+\) if there exists a radius \(R \in \mathbb{R}^+\) and a constant \(C\) such that

\[
C |\omega|^\gamma \leq |L(\omega)| \quad \text{for all } |\omega| \geq R\]

where \(\gamma\) is critical in the sense that the condition fails for any larger value.
The order is in direct relation with the degree of smoothness of the Green’s function \( \rho_L \).

In the case of a scale-invariant operator, it also coincides with the scaling order (or degree of homogeneity) of \( L(\omega) \). For instance, the fractional derivative operator \( D^\gamma \), which is defined via the Fourier multiplier \( (j\omega)\gamma \), is of order \( \gamma \). Its Green’s function is given by (see Table A.1 in Appendix A)

\[
\rho_{D^\gamma}(r) = |\mathcal{F}^{-1}\left\{ \frac{1}{(j\omega)^\gamma} \right\} (r) = \frac{r^{\gamma-1}}{\Gamma(\gamma)},
\]

where \( \Gamma \) is Euler’s gamma function (see Appendix C.2) and \( r^{\gamma-1} = \max(0, r^{\gamma-1}) \). Clearly, the latter is a function of slow growth. It has a single singularity at the origin whose Hölder exponent is \( (\gamma - 1) \), and is infinitely differentiable everywhere else. It follows that \( \rho_{D^\gamma} \) is uniformly Hölder-continuous of degree \( (\gamma - 1) \). This is one less than the order of the operator. On the other hand, the null space of \( D^\gamma \) consists of the polynomials of degree \( N = [\gamma - 1] \) since \( \frac{d^{|\omega|}}{|\omega|^\gamma} \propto (j\omega)^{-n} \) is vanishing at the origin up to order \( N \) with \( (\gamma - 1) \leq N < \gamma \) (see argumentation in Section 6.4.1).

A fundamental result is that all partial differential operators with constant coefficients are spline-admissible. This follows from the Malgrange-Ehrenpreis theorem which guarantees the existence of their Green’s function [Mal, Wag09]. The generic form of such operators is

\[
L_N = \sum_{|n| < N} a_n \partial^n
\]

with \( a_n \in \mathbb{R}^d \), where \( \partial^n \) is the multi-index notation for \( \frac{\partial^n_1 \cdots \partial^n_d}{\partial x_1^{n_1} \cdots \partial x_d^{n_d}} \). The corresponding Fourier multiplier is \( \hat{L}_N(\omega) = \sum_{|n| < N} a_n |n|! \omega^n \), which is a polynomial of degree \( N = |n| \). The operator is elliptic if \( \hat{L}_N(\omega) \) vanishes at the origin and nowhere else. More generally, it is called quasi-elliptic of order \( \gamma \) if \( \hat{L}_N(\omega) \) fulfills the growth condition in Definition 20. For \( d = 1 \), it is fairly easy to determine \( \rho_L \) using standard Fourier-inversion techniques (see Chapter 5). Moreover, the condition for quasi-ellipticity of order \( N \) is automatically satisfied. When moving to higher dimensions, the study of partial differential operators and the determination of their Green’s function becomes more challenging because of the absence of a general multidimensional factorization mechanism. Yet, it is possible to treat special cases in full generality such as the scale-invariant operators (with homogeneous, but not necessarily rotation-invariant, Fourier multipliers), or the class of rotation-invariant operators that are polynomials of the Laplacian \((-\Delta)\).

### 6.2.2 Splines and operators

The foundation of our formulation is the direct correspondence between a spline-admissible operator \( L \) and a particular brand of splines.

**Definition 21** (Cardinal L-spline). A function \( s(r) \) (possibly of slow growth) is called a cardinal L-spline if and only if

\[
Ls(r) = \sum_{k \in \mathbb{Z}^d} a[k] \delta(r - k).
\]

The location of the Dirac impulses specifies the spline discontinuities (or knots). The term “cardinal” refers to the particular setting where these are located on the Cartesian grid \( \mathbb{Z}^d \).

The remarkable aspect in this definition is that the operator \( L \) has the role of a mathematical A-to-D converter since it maps a continuously defined signal \( s \) into a discrete sequence \( a = (a[k]) \). Also note that the weighted sum of Dirac impulses in the r.h.s. of
the above equation can be interpreted as the continuous-domain representation of the discrete signal $a$—it is a hybrid-type representation that is commonly used in the theory of linear systems to model ideal sampling (multiplication with a train of Dirac impulses). The underlying concept of spline is fairly general and it naturally extends to nonuniform grids.

**Definition 22** (Nonuniform spline). Let $\{r_k\}_{k \in \mathbb{S}}$ be a set of points (not necessarily finite) that specifies a (nonuniform) grid in $\mathbb{R}^d$. Then, a function $s(r)$ (possibly of slow growth) is a nonuniform $L$-spline with knots $\{r_k\}_{k \in \mathbb{S}}$ if and only if

$$Ls(r) = \sum_{k \in \mathbb{S}} a_k \delta(r - r_k).$$

The direct implication of this definition is that a (nonuniform) $L$-spline with knots $\{r_k\}$ can generally be expressed as

$$s(r) = p_0(r) + \sum_{k \in \mathbb{S}} a_k \rho_L(r - r_k),$$

where $\rho_L = L^{-1} \delta$ is the Green's function of $L$ and $p_0 \in \mathfrak{N}_L$ is an appropriate null-space component that is typically selected to fulfill some boundary conditions.

In the case where the grid is uniform, it is usually more convenient to express splines in terms of localized $B$-spline functions which are shifted replicates of a simple template $\Phi_L$, or some other equivalent generator. An important requirement is that the set of $B$-spline functions constitutes a Riesz basis.

### 6.2.3 Riesz bases

To quote Ingrid Daubechies [Dau92], a Riesz basis is the next best thing after an orthogonal basis. The reason for not enforcing orthogonality is to leave more room for other desirable features such as simplicity of the construction, maximum localization of the basis function (e.g., compact support), and, last but not least, fast computational solutions.

**Definition 23** (Riesz basis). A sequence of functions $\{\phi_k(r)\}_{k \in \mathbb{Z}}$ in $L_2(\mathbb{R}^d)$ forms a Riesz basis if and only if there exist two constants $A$ and $B$ such that

$$A \|c\|_{\ell_2} \leq \left\| \sum_{k \in \mathbb{Z}} c_k \phi_k(r) \right\|_{L_2(\mathbb{R}^d)} \leq B \|c\|_{\ell_2}$$

for any sequence $c = (c_k) \in \ell_2$. More generally, the basis is $L_p$-stable if there exist two constants $A_p$ and $B_p$ such that

$$A_p \|c\|_{\ell_p} \leq \left\| \sum_{k \in \mathbb{Z}} c_k \phi_k(r) \right\|_{L_p(\mathbb{R}^d)} \leq B_p \|c\|_{\ell_p}.$$

This definition imposes an equivalence between the $L_2$ ($L_p$, resp.) norm of the continuously defined function $s(r) = \sum_{k \in \mathbb{Z}} c_k \phi_k(r)$ and the $\ell_2$ ($\ell_p$, resp.) norm of its expansion coefficients $(c_k)$. This ensures that the representation is stable in the sense that a small perturbation of the expansion coefficients results in a perturbation of comparable magnitude on $s(r)$ and vice versa. Also note that the lower inequality implies that the functions $\{\phi_k\}$ are linearly independent (by setting $s(r) = 0$), which is the defining property of a basis in finite dimensions—but which, on its own, does not ensure stability in infinite dimensions. When $A = B = 1$, we have a perfect norm equivalence which translates into the
6.2. Basic concepts and definitions

basis being orthonormal (Parseval’s relation). Finally, we like to point out that the existence of the bounds \( A \) and \( B \) ensures that the (infinite) Gram matrix is positive-definite so that it can be readily diagonalized to yield an equivalent orthogonal basis.

In the (multi-)integer shift-invariant case where the basis functions are given by \( \phi_k(r) = \phi(r - k), k \in \mathbb{Z}^d \), there is a simpler equivalent reformulation of the Riesz basis requirement of Definition 23.

**Theorem 21.** Let \( \phi(r) \in L_2(\mathbb{R}^d) \) be a B-spline-like generator whose Fourier transform is denoted by \( \hat{\phi}(\omega) \). Then, \( |\phi(r - k)|_{k \in \mathbb{Z}^d} \) forms a Riesz basis with Riesz bounds \( A \) and \( B \) if and only if

\[
0 < A^2 \leq \sum_{n \in \mathbb{Z}^d} |\hat{\phi}(\omega + 2\pi n)|^2 \leq B^2 < \infty \tag{6.17}
\]

for almost every \( \omega \). Moreover, the basis is \( L_p \)-stable for all \( 1 \leq p \leq +\infty \) if, in addition,

\[
\sup_{r \in [0,1]^d} \sum_{k \in \mathbb{Z}^d} |\phi(r - k)| = A_{2,\infty} < +\infty. \tag{6.18}
\]

Under such condition(s), the induced function space

\[
V_\phi = \left\{ s(r) = \sum_{k \in \mathbb{Z}^d} c(k) \phi(r - k) : c \in \ell_p(\mathbb{Z}^d) \right\}
\]

is a closed subspace of \( L_p(\mathbb{R}^d) \), including the standard case \( p = 2 \).

Observe that the central quantity in (6.17) corresponds to the discrete-domain Fourier transform of the Gram sequence \( a_\phi[k] = \langle \phi(-k), \phi \rangle_{L_2} \). Indeed, we have that

\[
A_\phi(e^{j\omega}) = \sum_{k \in \mathbb{Z}^d} a_\phi[k] e^{-j(\omega,k)} = \sum_{n \in \mathbb{Z}^d} |\hat{\phi}(\omega + 2\pi n)|^2 \tag{6.19}
\]

where the r.h.s. follows from Poisson’s summation formula applied to the sampling at the integers of the autocorrelation function \( \langle \hat{\phi}^* \phi \rangle(r) \). Formula (6.19) is especially advantageous in the case of compactly supported B-splines for which the autocorrelation is often known explicitly (as a B-spline of twice the order) since it reduces the calculation to a finite sum over the support of the Gram sequence (discrete-domain Fourier transform).

Theorem 21 is a fundamental result in sampling and approximation theory [Uns00]. It is instructive here to briefly run through the \( L_2 \) part of the proof which also serves as a refresher on some of the standard properties of the Fourier transform. In particular, we like to emphasize the interaction between the continuous and discrete aspects of the problem.

**Proof.** We start by computing the Fourier transform of \( s(r) = \sum_{k \in \mathbb{Z}^d} c[k] \phi(r - k) \), which gives

\[
\mathcal{F}[s](\omega) = \sum_{k \in \mathbb{Z}^d} c[k] e^{-j(\omega,k)} \hat{\phi}(\omega) \quad \text{(by linearity and shift property)}
\]

\[
= C(e^{j\omega}) \cdot \hat{\phi}(\omega),
\]
6. SPLINES AND WAVELETS

where \( C(e^{i\omega}) \) is recognized as the discrete-domain Fourier transform of \( c[\cdot] \). Next, we invoke Parseval’s identity and manipulate the Fourier-domain integral as follows

\[
\| s \|_{L_2}^2 = \sum_{n \in \mathbb{Z}^d} \int_{[0,2\pi]^d} |C(e^{i\omega})|^2 \frac{d\omega}{(2\pi)^d} \left| \hat{\phi}(\omega + 2\pi n) \right|^2 \frac{d\omega}{(2\pi)^d} = \int_{[0,2\pi]^d} |C(e^{i\omega})|^2 \sum_{n \in \mathbb{Z}^d} \left| \hat{\phi}(\omega + 2\pi n) \right|^2 \frac{d\omega}{(2\pi)^d} = \int_{[0,2\pi]^d} |C(e^{i\omega})|^2 A_\phi(e^{i\omega}) \frac{d\omega}{(2\pi)^d}.
\]

There, we have used the fact that \( C(e^{i\omega}) \) is \( 2\pi \)-periodic and the nonnegativity of the integrand to interchange the summation and the integral (Fubini). This naturally leads to the inequality

\[
\inf_{\omega \in [0,2\pi]^d} A_\phi(e^{i\omega}) \cdot \| c \|_{L_2}^2 \leq \| s \|_{L_2}^2 = \sup_{\omega \in [0,2\pi]^d} A_\phi(e^{i\omega}) \cdot \| c \|_{L_2}^2
\]

where we are now making use of Parseval’s identity for sequences, so that

\[
\| c \|_{L_2}^2 = \int_{[0,2\pi]^d} |C(e^{i\omega})|^2 \frac{d\omega}{(2\pi)^d}.
\]

The final step is to show that these bounds are sharp. This can be accomplished through the choice of some particular (bandlimited) sequence \( c[\cdot] \).

Note that the almost everywhere part of (6.17) can be dropped when \( \phi \in L_1(\mathbb{R}^d) \) because the Fourier transform of such a function is continuous (Riemann-Lebesgue Lemma).

While the result of Theorem 21 is restricted to the classical \( L_p \) spaces, there is no fundamental difficulty in extending it to wider classes of weighted (with negative powers) \( L_p \) spaces by imposing some stricter condition than (6.18) on the decay of \( \phi \). For instance, if \( \phi \) has exponential decay, then the definition of the function space \( V_p \) can be extended for all sequences \( c \) that are growing no faster than a polynomial. This happens to be the appropriate framework for sampling generalized stochastic processes which do not live in the \( L_p \) spaces since they are not decaying at infinity.

6.2.4 Admissible wavelets

The other important tool for analyzing stochastic processes is the wavelet transform whose basis functions must be “tuned” to the object under investigation.

Definition 24. A wavelet function \( \psi \) is called \( L \)-admissible if it can be expressed as \( \psi = L^H \phi \) with \( \phi \in L_1(\mathbb{R}^d) \).

Observe that we are now considering the Hermitian transpose operator \( L^H = \Gamma^* \) which is distinct from the adjoint operator \( L^* \) when the impulse response has some imaginary component. The reason for this is that the wavelet-analysis step involves a Hermitian inner product \( \langle \cdot, \cdot \rangle_{L_2} \) whose definition differs by a complex conjugation from that of the distributional scalar product \( \langle \cdot, \cdot \rangle \) used in our formulation of stochastic processes when the second argument is complex valued; specifically, \( \langle f, g \rangle_{L_2} = \langle f, \overline{g} \rangle = \int_{\mathbb{R}^d} f(r) \overline{g(r)} \, dr \).

The best matched wavelet is the one for which the wavelet kernel \( \phi \) is the most localized—ideally, the shortest-possible support assuming that it is at all possible to construct a
6.3 First-order exponential B-splines and wavelets

Rather than aiming for the highest level of generality right away, we propose to first examine the 1-D first-order scenario in some detail. First-order differential models are important theoretically because they go hand in hand with the Markov property. In that respect, they constitute the next level of generalization just beyond the Lévy processes. Mathematically, the situation is still quite comparable to that of the derivative operator in the sense that it leads to a nice and self-contained construction of (exponential) B-splines and wavelets. The interesting aspect, though, is that the underlying basis functions are no longer conventional wavelets that are dilated versions of a single prototype: they now fall into the lesser-known category of non-stationary wavelets.

The (causal) Green’s function of our canonical first-order operator \( P_\alpha = (D - \alpha I) \) is identical to the impulse response \( \rho_\alpha \) of the corresponding differential system, while the (one-dimensional) null space of the operator is given by \( \mathcal{N}_\alpha = \{ a_0 e^{\alpha r} : a_0 \in \mathbb{R} \} \). Some examples of such Green’s functions are shown in Figure 6.3. The case \( \alpha = 0 \) (darker/red curve) is the classical one already treated in Section 6.1.

Figure 6.3: Comparison of basis functions related to the first-order differential operator \( P_\alpha = D - \alpha I \) for \( \alpha = 0, -1, -2, -4 \). (a) Green’s functions \( \rho_\alpha(r) \). (b) Exponential B-splines \( \beta_\alpha(r) \) (c) Augmented spline interpolators \( \psi_{\text{int}}(r) \). (d) Orthonormalized versions of the exponential spline wavelets \( \psi_\alpha(r) = P_\alpha \psi_{\text{int}}(r) \).

compactly-supported wavelet basis. The very least is that \( \phi \) should be concentrated around the origin and exhibit a sufficient rate of decay; for instance, \( |\phi(r)| \leq \frac{C}{1 + |r|^a} \) for some \( a > d \).

A direct implication of Definition 24 is that the wavelet \( \psi \) will annihilate all the components (e.g., polynomials) that are in the null space of \( L \) because \( \langle \psi(-r_0), \rho_0 \rangle = \langle \phi(-r_0), L \rho_0 \rangle = 0 \), for all \( \rho_0 \in \mathcal{N}_\alpha \) and \( r_0 \in \mathbb{R}^d \). In conventional wavelet theory, this behavior is achieved by designing “\( N \)th-derivative-like” wavelets with vanishing moments up to the order \( N \).

6.3 First-order exponential B-splines and wavelets

2. In the terminology of wavelets, the term “non-stationary” refers to the property that the shape of the wavelet changes with scale, but not with respect to the location, as the more usual statistical meaning of the term would suggest.
6. SPLINES AND WAVELETS

6.3.1 B-spline construction

The natural discrete approximation of the differential operator \( P_a = (D - a \text{Id}) \) is the first-order weighted difference operator

\[
\Delta_a s(r) = s(r) - e^{a} s(r - 1).
\]

(6.20)

Observe that \( \Delta_a \) annihilates the exponentials \( a_0 e^{ar} \) so that its null space includes \( \mathcal{N}_a \). The corresponding B-spline is obtained by applying \( \Delta_a \) to \( \Omega_a \), which yields

\[
\beta_a(r) = \mathcal{F}^{-1} \left\{ \frac{1 - e^{a} e^{-j\omega}}{j\omega - a} \right\}(r) = \begin{cases} e^{ar}, & \text{for } 0 \leq r < 1 \\ 0, & \text{otherwise.} \end{cases}
\]

(6.21)

In effect, the localization by \( \Delta_a \) results in a “chopped off” version of the causal Green's function that is restricted to the interval \([0,1)\) (see Figure 6.3b). Importantly, the scheme remains applicable in the unstable scenario \( \text{Re}(a) < 0 \). It always results in a well-defined Fourier transform due to the convenient pole-zero cancellation in the central expression of (6.21). The marginally unstable case \( a = 0 \) results in the rectangular function shown in red in Figure 6.3, which is the standard basis function for representing piecewise-constant signals. Likewise, \( \beta_a \) generates an orthogonal basis for the space of cardinal \( P_a \)-splines in accordance with Definition 21. This allows us to specify our prototypical exponential spline space as \( V_0 = \text{span}\{\beta_a(-k)\}_{k \in \mathbb{Z}} \) with knot spacing \( 2^0 = 1 \).

6.3.2Interpolator in augmented-order spline space

The second important ingredient is the interpolator for the “augmented-order” spline space generated by the autocorrelation \( (\widehat{\beta}_a \ast \beta_a)(r) \) of the B-spline. Constructing it is especially easy in the first-order case because it involves the simple normalization

\[
\varphi_{\text{int},a}(r) = \frac{1}{(\widehat{\beta}_a \ast \beta_a)(0)} (\widehat{\beta}_a \ast \beta_a)(r)
\]

(6.22)

Specifically, \( \varphi_{\text{int},a} \) is the unique cardinal \( P_a^H P_a \)-spline function that vanishes at all the integers except at the origin where it takes the value one (see Figure 6.3c). Its classical use is to provide a sinc-like kernel for the representation of the corresponding family of splines, and also for the reconstruction of spline-related signals, including special brands of stochastic processes, from their integer samples [UB05b]. Another remarkable and lesser known property is that this function provides the proper smoothing kernel for defining an operator-like wavelet basis.

6.3.3 Differential wavelets

In the generalized spline framework, instead of specifying a hierarchy of multiresolution subspaces of \( L_2(\mathbb{R}) \) (the space of finite-energy functions) via the dilation of a scaling function, one considers the fine-to-coarse sequence of L-spline spaces

\[
V_i = \{s(r) \in L_2(\mathbb{R}) : Ls(r) = \sum_{k \in \mathbb{Z}} a_i[k] \delta(r - 2^i k)\},
\]

where the embedding \( V_i \supseteq V_j \) for \( i \leq j \) is obvious from the (dyadic) hierarchy of spline knots, so that \( s_j \in V_j \) implies that \( s_j \in V_i \) with an appropriate subset of its coefficients \( a_i[k] \) being zero.
We now detail the construction of a wavelet basis at resolution 1 such that $W_1 = \text{span}\{\psi_{1,k}\}_{k \in \mathbb{Z}}$ with $W_1 \perp V_1$ and $V_1 + W_1 = V_0 = \text{span}\{\beta_a(-k)\}_{k \in \mathbb{Z}}$. The recipe is to take
\[
\psi_a(r) = \psi_a(r - 1 - 2k)/\|\psi_a\|_{L_2} \quad \text{where } \psi_a \text{ is the mother wavelet given by}
\]
\[
\psi_a(r) = \mathcal{F}_a[X_a^H \varphi_{\text{int},a}(r) \propto \Delta_H a \beta_a(r)].
\]
There, $\Delta_H a$ is the Hermitian adjoint of the finite-difference operator $\Delta a$. Examples of such exponential-spline wavelets are shown in Figure 6.3d, including the classical Haar wavelet (up to a sign change) which is obtained for $a = 0$ (red curve). The basis functions $\psi_{1,k}$ are shifted versions of $\psi_a$ that are centered at the odd integers and normalized to have a unit norm. Since these wavelets are non-overlapping, they form an orthonormal basis. Moreover, the basis is orthogonal to the coarser spline space $V_1$ as a direct consequence of the interpolating property of $\varphi_{\text{int},a}$ (Proposition 19 in Section 6.5). Finally, based on fact that $\psi_{1,k} \in V_0$ for all $k \in \mathbb{Z}$, one can show that these wavelets span $W_1$, which translates into
\[
W_1 = \left\{ v(r) = \sum_{k \in \mathbb{Z}} v_1[k] \psi_{1,k}(r) : v_1 \in L_2(\mathbb{Z}) \right\}.
\]
This method of construction extends to the other wavelet subspaces $W_i$ provided that the interpolating kernel $\varphi_{\text{int},a}$ is substituted by its proper counterpart at resolution $a = 2^{i-1}$ and the sampling grid adjusted accordingly. Ultimately, this results into a wavelet basis of $L_2(\mathbb{R})$ whose members are all $P_a$-splines—that is, piecewise-exponential with parameter $a$—but not dilates of the same prototype unless $a = 0$. Otherwise, the corresponding decomposition is not fundamentally different from a conventional wavelet expansion. The basis functions are equally well localized and the scheme admits the same type of fast reversible filterbank algorithm, albeit with scale-dependent filters [KU06].

### 6.4 Generalized B-spline basis

The procedure of Section 6.3.1 remains applicable for the broad class of spline-admissible operators (see Definition 19) in one or multiple dimensions. The two ingredients for constructing a generalized B-spline basis are: 1) the knowledge of the Green’s function $\rho_L$ of the operator $L$, and 2) the availability of a discrete approximation (finite-difference-like) of the operator of the form
\[
L_d s(r) = \sum_{k \in \mathbb{Z}^d} d_k |k| s(r - k) \quad (6.23)
\]
with $d \in \ell_1(\mathbb{Z}^d)$ that fulfills the null-space matching constraint$^\text{3}$
\[
L_d p_0(r) = L p_0(r) = 0 \quad \text{for all } p_0 \in \mathcal{M}_L. \quad (6.24)
\]
The generalized B-spline associated with the operator $L$ is then given by
\[
\beta_L(r) = L_d \rho_L(r) = \mathcal{F}^{-1}\left\{ \sum_{k \in \mathbb{Z}^d} d_k |k| e^{-i(k,\omega)} / \hat{L}(\omega) \right\}(r), \quad (6.25)
\]
where the numerator and denominator in the r.h.s. expression correspond to the frequency responses of $L_d$ and $L$, respectively. The null-space matching constraint is especially helpful for the unstable cases where $\rho_L \notin L_1(\mathbb{R}^d)$: it ensures that the zeros of $\hat{L}(\omega)$ (singularities) are cancelled by some corresponding zeros of $\hat{L}_d(\omega)$ so that the Fourier transform of $\beta_L$ remains bounded.

---

$^3$ We want the null space of $L_d$ to include $\mathcal{M}_L$ and to remain the smallest possible. In that respect, it is worth noting that the null space of a discrete operator will always be much larger than that of its continuous-domain counterpart. For instance, the derivative operator $D$ suppresses constant signals, while its finite-difference counterpart annihilates all 1-periodic functions, including the constants.
Definition 25. The function $\beta_L$ specified by (6.25) is an admissible B-spline for $L$ if and only if 1) $\beta_L \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$, and 2) it generates a Riesz basis of the space of cardinal $L$-splines.

In light of Theorem 21, the latter property requires the existence of the two Riesz bounds $A$ and $B$ such that

$$0 < A^2 \leq \sum_{n \in \mathbb{Z}^d} |\hat{\beta}_L(\omega + 2\pi n)|^2 = \frac{|\sum_{k \in \mathbb{Z}^d} d_k |e^{-i|k,\omega}|^2}{\sum_{n \in \mathbb{Z}^d} |L(\omega + 2\pi n)|^2} \leq B^2. \quad (6.26)$$

A direct consequence of (6.25) is that

$$L\beta_L(r) = \sum_{k \in \mathbb{Z}^d} d_k |\delta(r - k)| \quad (6.27)$$

so that $\beta_L$ is itself a cardinal L-spline in accordance with Definition 21. The bottom line in Definition 25 is that any cardinal L-spline admits a unique representation in the B-spline basis $\{\beta_L(\leq k)\}_{k \in \mathbb{Z}^d}$ as

$$s(r) = \sum_{k \in \mathbb{Z}^d} c[k] \beta_L(r - k) \quad (6.28)$$

where the $c[k]$ are the B-spline coefficients of $s$.

While (6.25) provides us with a nice recipe for constructing B-splines, it does not guarantee that the Riesz-basis condition (6.26) is satisfied. This needs to be established on a case-by-case basis. The good news for the present theory of stochastic processes is that B-splines are available for virtually all the operators that have been discussed so far.

6.4.1 B-spline properties

To motivate the use of B-splines, we shall first restrict our attention to the space $V_L$ of cardinal L-splines with finite energy, which is formally defined as

$$V_L = \left\{ s(r) \in L_2(\mathbb{R}^d) : s(r) = \sum_{k \in \mathbb{Z}^d} a[k] \delta(r - k) \right\}. \quad (6.29)$$

The foundation of spline theory is that there are two complementary ways of representing splines using different types of basis functions: Green’s functions versus B-splines. The first representation follows directly from the Definition 21 (see also (6.16)) and is given by

$$s(r) = p_0(r) + \sum_{k \in \mathbb{Z}^d} a[k] \rho_L(r - k), \quad (6.30)$$

where $p_0 \in \mathcal{N}_L$ is a suitable element of the null space of $L$ and where $\rho_L = L^{-1}\delta$ is the Green’s function of the operator. The functions $\rho_L(\leq k)$ are nonlocal and very far from being orthogonal. In many cases, they are not even part of $V_L$, which raises fundamental issues concerning the $L_2$ convergence of the infinite sum in (6.30) and the conditions that must be imposed upon the expansion coefficients $a[\cdot]$. The second type of B-spline expansion (6.28) does not have such stability problems. This is the primary reason why it is favored by practitioners.

4. Without further assumptions on $\rho_L$ and $a$, (6.30) is only valid in the weak sense of distributions.
Stable representation of cardinal L-splines

The equivalent B-spline specification of the space \( V_L \) of cardinal splines is

\[
V_L = \left\{ s(r) = \sum_{k \in \mathbb{Z}^d} c[k] \beta_L(r - k) : c[\cdot] \in \ell_2(\mathbb{Z}^d) \right\},
\]

where the generalized B-spline \( \beta_L \) satisfies the conditions in Definition 25. The Riesz-basis property ensures that the representation is stable in the sense that, for all \( s \in V_L \), we have that

\[
A \|c\|_{\ell_2} \leq \|s\|_{\ell_2} \leq B \|c\|_{\ell_2}. \tag{6.31}
\]

There, \( \|c\|_{\ell_2} = (\sum_{k \in \mathbb{Z}^d} |c[k]|^2)^{\frac{1}{2}} \) is the \( \ell_2 \)-norm of the B-spline coefficients \( c \). The fact that the underlying functions are cardinal L-splines is a simple consequence of the atoms being splines themselves. Moreover, we can easily make the link with (6.30) by using (6.27), which yields

\[
Ls(r) = \sum_{k \in \mathbb{Z}^d} c[k] L \beta_L(r - k) = \sum_{k \in \mathbb{Z}^d} \left( c * d \right)[k] \delta(r - k).
\]

The less obvious aspect, which is implicit in the definition of the B-spline, is the completeness of the representation in the sense that the B-spline basis spans the space \( V_L \) defined by (6.29). We shall establish this by showing that the B-splines are capable of reproducing \( \Omega_L \) as well as any component \( p_0 \in \mathcal{N}_L \) in the null space of \( L \). The implication is that any function of the form (6.30) admits a unique expansion in a B-spline basis. This is also true when the function is not in \( L_2(\mathbb{R}^d) \), in which case the B-spline coefficients \( c \) are no longer in \( \ell_2(\mathbb{Z}^d) \) due to the discrete-continuous norm equivalence (6.31).

Reproduction of Green's functions

The reproduction of Green's functions follows from the special form of (6.25). To reveal it, we consider the inverse \( L_1^{-1} \) of the discrete localization operator \( L_1 \) specified by (6.23), whose continuous-domain impulse response is written as

\[
L_1^{-1}(r) = \sum_{k \in \mathbb{Z}^d} p[k] \delta(r - k) = \mathcal{F}^{-1} \left\{ \frac{1}{\sum_{k \in \mathbb{Z}^d} d_L[k] |k|^\alpha} \right\}.
\]

The sequence \( p \), which can be determined by generalized inverse Fourier transform, is of slow growth with the property that \( (p * d)[k] = \delta[k] \). The Green's function reproduction formula is then obtained by applying \( L_1^{-1} \) to the B-spline \( \beta_L \) and making use of the left-inverse property of \( L_1^{-1} \). Thus,

\[
L_1^{-1}\beta_L(r) = L_1^{-1}L_1\beta_L(r) = \beta_L(r)
\]

results into

\[
\rho_L(r) = \sum_{k \in \mathbb{Z}^d} p[k] \beta_L(r - k). \tag{6.32}
\]

To illustrate the concept, let us get back to our introductory example in Section in 6.3.1 with \( L = P_\alpha = (D - \alpha \text{Id}) \) where \( \text{Re}(\alpha) < 0 \). The frequency response of this first-order operator is

\[
\hat{P}_\alpha(\omega) = j\omega - \alpha,
\]
6. SPLINES AND WAVELETS

while its Green’s function is given by

\[ \rho_\alpha(r) = \mathbb{1}_+(r)e^{\alpha r} = \mathcal{F}^{-1}\left\{ \frac{1}{\omega - \alpha} \right\}(r). \]

On the discrete side of the picture, we have the finite-difference operator \( \Delta_\alpha \) with

\[ \hat{\Delta}_\alpha(\omega) = 1 - e^{\alpha - j\omega}, \]

and its inverse \( \Delta_\alpha^{-1} \) whose expansion coefficients are

\[ p_\alpha[k] = \mathbb{1}_+[k]e^{\alpha k} = \mathcal{F}^{-1}_d\left\{ \frac{1}{1 - e^{\alpha - j\omega}} \right\}[k], \]

where \( \mathcal{F}^{-1}_d \) denotes the discrete-domain inverse Fourier transform. The application of (6.32) then yields the exponential-reproduction formula

\[ \mathbb{1}_+(r)e^{\alpha r} = \sum_{k=0}^{\infty} e^{\alpha k} \beta_\alpha(t - k) \tag{6.33} \]

where \( \beta_\alpha \) is the exponential B-spline defined by (6.21). Note that the range of applicability of (6.33) extends to \( \text{Re}(\alpha) \leq 0 \).

Reproduction of null-space components

A fundamental property of B-splines is their ability to reproduce the components that are in the null space of their defining operator. In the case of our working example, we can simply extrapolate (6.33) for negative indices, which yields

\[ e^{\alpha r} = \sum_{k \in \mathbb{Z}} e^{\alpha k} \beta_\alpha(r - k). \]

It turns out that this reproduction property is induced by the matching null-space constraint (6.24) that is imposed upon the localization filter. While the reproduction of exponentials is interesting in its own right, we shall focus here on the important case of polynomials and provide a detailed Fourier-based analysis.

We start by recalling that the general form of a multidimensional polynomial of total degree \( N \) is

\[ q_N(r) = \sum_{|n| \leq N} a_n r^n \]

using the multi-index notation with \( n = (n_1, \ldots, n_d) \in \mathbb{N}^d, \ r^n = r_1^{n_1} \cdots r_d^{n_d}, \) and \( |n| = n_1 + \cdots + n_d. \) The generalized Fourier transform of \( q_N \in \mathbb{F}'(\mathbb{R}^d) \) (see Table 3.3 and entry \( r^n f(r) \) with \( f(r) = 1 \)) is given by

\[ \hat{q}_N(\omega) = \sum_{|n| \leq N} (2\pi)^d a_n [n]^{d\delta}(\omega), \]

where \( \partial^n \delta \) denotes the \( n \)th partial derivative of the multidimensional Dirac impulse \( \delta. \)

Hence, the Fourier multiplier \( \hat{L} \) will annihilate the polynomials of order \( N \) if and only if \( \hat{L}(\omega)\partial^n \delta(\omega) = 0 \) for all \( |n| \leq N. \) To understand when this condition is met, we expand \( \hat{L}(\omega)\partial^n \delta(\omega) \) in terms of \( \partial^k \hat{L}(0), \ |k| \leq |n| \) by using the general product rule for the manipulation of Dirac impulses and their derivatives given by

\[ f(r) \partial^n \delta(r - r_0) = \sum_{k+l=n} \frac{n!}{k! l!} (-1)^{|n|+|l|} \partial^k f(r_0) \partial^l \delta(r - r_0). \]

5. Our definition of the inverse discrete Fourier transform in 1-D is

\[ \mathcal{F}^{-1}_d \{ H(\omega) \} = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(\omega) e^{i\omega k} \, d\omega \text{ with } k \in \mathbb{Z}. \]

\[ 108 \]
6.4. Generalized B-spline basis

The latter follows from Leibnitz’ rule for partially differentiating a product of functions as
\[
\partial^n (f \varphi) = \sum_{k+|n|} n! \partial^k f \partial^{|n|} \varphi,
\]
and the adjoint relation \( \langle \varphi, f \partial^n \delta(-r_0) \rangle = (\partial^n (f \varphi), \delta(-r_0)) \) with \( \partial^n = (-1)^{|n|} \partial^{|n|} \). This allows us to conclude that the necessary and sufficient condition for the inclusion of the polynomials of order \( N \) in the null space of \( L \) is
\[
\partial^n \hat{L}(0) = 0, \text{ for all } n \in \mathbb{N}^d \text{ with } |n| \leq N, \tag{6.34}
\]
which is equivalent to \( \hat{L} (\omega) = O(\|\omega\|^{N+1}) \) around the origin. Note that this behavior is prototypical of scale-invariant operators such as fractional derivatives and Laplacians. The same condition has obviously to be imposed upon the localization filter \( \hat{L}_d \) for the Fourier transform of the B-spline in (6.25) to be nonsingular at the origin. Since \( \hat{L}_d (\omega) \) is 2\( \pi \)-periodic, we have that
\[
\partial^n \hat{L}_d (2\pi k) = 0, \quad k \in \mathbb{Z}^d, n \in \mathbb{N}^d \text{ with } |n| \leq N. \tag{6.35}
\]
For practical convenience, we shall assume that the B-spline \( \beta_L \) is normalized to have a unit integral\(^6\) so that \( \hat{\beta}_L (0) = 1 \). Based on (6.35) and \( \hat{\beta}_L (\omega) = \hat{L}_d (\omega)/\hat{L}(\omega) \), we find that
\[
\begin{aligned}
\hat{\beta}_L (0) = 1, \\
\partial^n \hat{\beta}_L (2\pi k) = 0, \quad k \in \mathbb{Z}^d \setminus \{0\}, n \in \mathbb{N}^d \text{ with } |n| \leq N, \tag{6.36}
\end{aligned}
\]
which are the so-called Strang-Fix conditions of order \( N \). Recalling that \( j^{|n|} \partial^n \hat{\beta}_L (\omega) \) is the Fourier transform of \( r^n \beta_L (r) \) and that periodization in the signal domain corresponds to a sampling in the Fourier domain, we finally deduce that
\[
\sum_{k \in \mathbb{Z}^d} (r-k)^n \beta_L (r-k) = j^{|n|} \partial^n \hat{\beta}_L (0) = C_n, \quad n \in \mathbb{N}^d \text{ with } 0 < |n| \leq N, \tag{6.37}
\]
with the implicit assumption that \( \beta_L \) has a sufficient order of algebraic decay for the above sums to be convergent. The special case of (6.37) with \( n = 0 \) reads
\[
\sum_{k \in \mathbb{Z}^d} \beta_L (r-k) = 1 \tag{6.38}
\]
and is called the partition of unity. It reflects the fact that \( \beta_L \) reproduces the constants. More generally, Condition (6.37) (or (6.36)) is equivalent to the existence of sequences \( p_n \) such that
\[
r^n = \sum_{k \in \mathbb{Z}^d} p_n |k| \beta_L (r-k) \quad \text{for all } |n| \leq N, \tag{6.39}
\]
which is a more direct statement of the polynomial-reproduction property. For instance, (6.37) with \( n = (1, \ldots, 1) \) implies that
\[
\begin{aligned}
r \left( \sum_{k \in \mathbb{Z}^d} \beta_L (r-k) \right) - \sum_{k \in \mathbb{Z}^d} k \beta_L (r-k) = C_{(1, \ldots, 1)}, \\
p_{(0, \ldots, 0)} = 1
\end{aligned}
\]
\(^6\) This is always possible thanks to Condition (6.24), which ensures that \( \hat{\beta}_L (0) \neq 0 \) due to a proper cancellation of poles and zeros in the r. h. s. of (6.25).
from which one deduces that $p_{|1,...,1|}[k] = k + C_{|1,...,1|}$. The other sequences $p_{n}$, which are polynomials in $k$, may be determined in a similar fashion by proceeding recursively. Another equivalent way of stating the Strang-Fix conditions of order $N$ is that the sums

$$
\sum_{k \in \mathbb{Z}^d} k^n \hat{\beta}_{L}(r - k) = \sum_{t \in \mathbb{Z}^d} |n| \partial^{|n|}_{\omega} (e^{-j\omega r} \hat{\beta}_{L}(-\omega))|_{\omega=2\pi t}
$$

are polynomials with leading term $r^n$ for all $|n| \leq N$. The left-hand-side expression follows from Poisson's summation formula\(^7\) applied to the function $f(x) = x^n \hat{\beta}_{L}(r - x)$ with $r$ being considered as a constant shift.

**Localization**

The guiding principle for designing B-splines is to produce basis functions that are maximally localized on $\mathbb{R}^d$. Ideally, B-splines should have the smallest possible support which is the property that makes them so useful in applications. When it is not possible to construct compactly supported basis functions, the B-spline should at least be concentrated around the origin and satisfy some decay bound with the tightest possible constants. The primary types of spatial localization, by order of preference, are:

1. **Compact support**: $\hat{\beta}_{L}(r) = 0$ for all $r \in \Omega$ where $\Omega \subset \mathbb{R}^d$ is a convex set with the smallest-possible Lebesgue measure.
2. **Exponential decay**: $|\hat{\beta}_{L}(r - r_0)| \leq C \exp(-\alpha |r|)$ for some $r_0 \in \mathbb{R}^d$ and the largest possible $\alpha \in \mathbb{R}^+$.\(^8\)
3. **Algebraic decay**: $|\hat{\beta}_{L}(r - r_0)| \leq C \frac{1}{(1 + |r|)^N}$ for some $r_0 \in \mathbb{R}^d$ and the largest possible $\alpha \in \mathbb{R}^+$.\(^9\)

By relying on the classical relations that link spatial decay to the smoothness of the Fourier transform, one can get a good estimate of spatial decay based on the knowledge of the Fourier transform $\hat{\beta}_{L}(\omega) = \hat{L}_{d}(\omega)/\hat{L}(\omega)$ of the B-spline. Since the localization filter $\hat{L}_{d}(\omega)$ acts by compensating the (potential) singularities of $\hat{L}(\omega)$, the guiding principle is that the rate of decay is essentially determined by the degree of differentiability of $\hat{L}(\omega)$.

Specifically, if $\hat{\beta}_{L}(\omega)$ is differentiable up to order $N$, then the B-spline $\beta_{L}$ is guaranteed to have an algebraic decay of order $N$. To show this, we consider the Fourier transform pair $r^n \hat{\beta}_{L}(r) \rightarrow j^{n} \partial^{|n|}_{\omega} \hat{\beta}_{L}$ subject to the constraint that $\partial^{|n|}_{\omega} \hat{\beta}_{L} \in L_{1}(\mathbb{R}^d)$ for all $|n| < N$. From the definition of the inverse Fourier integral, it immediately follows that

$$
|r^n \hat{\beta}_{L}(r)| \leq \frac{1}{(2\pi)^d} \|\partial^{|n|}_{\omega} \hat{\beta}_{L}\|_{L_{1}},
$$

which, when properly combined over all multi-indices $|n| < N$, yields an algebraic decay estimate with $\alpha = N$.

By pushing the argument to the limit, we see that exponential decay (which is faster than any order of algebraic decay) requires that $\hat{\beta}_{L} \in C^{\infty}(\mathbb{R}^d)$ (infinite order of differentiability), which is only possible if $\hat{L}(\omega) \in C^{\infty}(\mathbb{R}^d)$ as well.

The ultimate limit in Fourier-domain regularity is when $\hat{\beta}_{L}$ has an analytic extension that is an entire\(^8\) function. In fact, by the Paley-Wiener theorem (Theorem 22 below), one achieves compact support of $\beta_{L}$ if and only if $\hat{\beta}_{L}(\zeta)$ is an entire function of exponential type. To explain this concept, we focus on the one-dimensional case where the B-spline

---

\(^7\) The standard form of Poisson's summation formula is $\sum_{k \in \mathbb{Z}^d} f(k) = \sum_{t \in \mathbb{Z}^d} f(2\pi t)$; it is valid for any Fourier pair $f, \hat{f} = \mathcal{F}(f) \in L_{1}(\mathbb{R}^d)$ with sufficient decay for the two sums to be convergent.

\(^8\) An entire function is a function that is analytic over the whole complex plane $\mathbb{C}$. 

---

110
6.4. Generalized B-spline basis

\( \beta_L \) is supported in the finite interval \([-A, +A] \). We then consider the holomorphic Fourier (or Fourier-Laplace) transform of the B-spline given by

\[
\hat{\beta}_L(\zeta) = \int_{-A}^{+A} \beta_L(r) e^{-\zeta r} \, dr
\]  

(6.40)

with \( \zeta = \sigma + j \omega \in \mathbb{C} \), which formally amounts to substituting \( j \omega \) by \( \zeta \) in the expression of the Fourier transform of \( \beta_L \). In order to obtain a proper analytic extension, we need to verify that \( \hat{\beta}_L(\zeta) \) satisfies the Cauchy-Riemann equation. We shall do so by applying a dominated-convergence argument. To that end, we construct the exponential bound

\[
|\hat{\beta}_L(\zeta)| \leq e^{A|\zeta|} \int_{-A}^{+A} |\beta_L(r)| \, dr
\]

\[
\leq e^{A|\zeta|} \sqrt{\int_{-A}^{+A} 1 \, dr} \sqrt{\int_{-A}^{+A} |\beta_L(r)|^2 \, dr}
\]

\[
= e^{A|\zeta|} \sqrt{2A \|\beta_L\|_{L_2}}
\]

where we have applied Cauchy-Schwarz' inequality to derive the lower inequality. Since \( e^{-\zeta r} \) for \( r \) fixed is itself an entire function and (6.40) is convergent over the whole complex plane, the conclusion is that \( \hat{\beta}_L(\zeta) \) is entire as well, in addition to being a function of exponential type \( A \) as indicated by the bound. The whole strength of the Paley-Wiener theorem is that the implication also works the other way around.

**Theorem 22** (Paley-Wiener). Let \( f \in L_2(\mathbb{R}) \). Then, \( f \) is compactly supported in \([-A, A]\) if and only if its Laplace transform

\[
F(\zeta) = \int_{\mathbb{R}} f(r) e^{-\zeta r} \, dr
\]

is an entire function of exponential type \( A \), meaning that there exists a constant \( C \) such that

\[
|F(\zeta)| \leq C e^{A|\zeta|}
\]

for all \( \zeta \in \mathbb{C} \).

The result implies that one can deduce the support of \( f \) from its Laplace transform. We can also easily extend the result to the case where the support is not centered around the origin by applying the Paley-Wiener theorem to the autocorrelation function \((f * f^*)(r)\). The latter is supported in the interval \([-2A, 2A]\), which is twice the size of the support of \( f \) irrespective of its center. This suggest the following expression for the determination of the support of a B-spline:

\[
\text{support}(\beta_L) = \limsup_{R \to \infty} \frac{\log \left( \sup_{|\zeta| \leq R} |\hat{\beta}_L(\zeta)\hat{\beta}_L(-\zeta)| \right)}{R}.
\]

(6.41)

It returns twice the exponential type of the recentered B-spline which gives \( \text{support}(\hat{\beta}_L) = 2A \). While this formula is only strictly valid when \( \hat{\beta}_L(\zeta) \) is an entire function, it can be used otherwise as an operational measure of localization when the underlying B-spline is not compactly supported. Interestingly, (6.41) provides a measure that is additive with respect to convolution and proportional to the order \( \gamma \). For instance, the support of an (exponential) B-spline associated with an ordinary differential operator of order \( N \) is precisely \( N \), as a consequence of the factorization property of such B-splines (see Sections 6.4.2 and 6.4.4).
To get some insight into (6.41), let us consider the case of the polynomial B-spline of order 1 (or degree 0) with \( \beta_D(r) = \mathbb{I}_{[0,1]}(r) \) and Laplace transform

\[
\hat{\beta}_D(\zeta) = \left( \frac{1 - e^{-\zeta}}{\zeta} \right).
\]

The required product in (6.41) is

\[
\hat{\beta}_D(\zeta) \hat{\beta}_D(-\zeta) = \frac{-e^{\zeta} + 2 - e^{-\zeta}}{\zeta^2},
\]

which is analytic over the whole complex plane because of the pole-zero cancellation at \( \zeta = 0 \). For \( R \) sufficiently large, we clearly have that

\[
\max_{|\zeta| = R} \left| \hat{\beta}_D(\zeta) \hat{\beta}_D(-\zeta) \right| = \frac{e^R + 2 + e^{-R}}{R^2} \to \frac{e^R}{R^2}.
\]

By plugging the above expression in (6.41), we finally get

\[
\text{support}(\beta_D) = \limsup_{R \to \infty} \frac{R - 2 \log R}{R} = 1,
\]

which is the desired result. While the above calculation may look like overkill for the determination of the already-known support of \( \beta_D \), it becomes quite handy for making predictions for higher-order operators. To illustrate the point, we now consider the B-spline of order \( \gamma \) associated with the (possibly fractional) derivative operator \( D^\gamma \) whose Fourier-Laplace transform is

\[
\hat{\beta}_{D^\gamma}(\zeta) = \left( \frac{1 - e^{-\zeta}}{\zeta} \right)^\gamma.
\]

We can then essentially replicate the previous manipulation while moving the order out of the logarithm to deduce that

\[
\text{support}(\beta_{D^\gamma}) = \limsup_{R \to \infty} \frac{\gamma R - 2\gamma \log R}{R} = \gamma.
\]

This shows that the “support” of the B-spline is equal to its order, with the caveat that the underlying Fourier-Laplace transform \( \hat{\beta}_{D^\gamma}(\zeta) \) is only analytic (and entire) when the order \( \gamma \) is a positive integer. This points to the fundamental limitation that a B-spline associated with a fractional operator—that is, when \( \hat{L}(\zeta) \) is not an entire function—cannot be compactly supported.

**Smoothness**

The smoothness of a B-spline refers to its degree of continuity and/or differentiability. Since a B-spline is a linear combination of shifted Green’s functions, its smoothness is the same as that of \( \rho_L \).

Smoothness descriptors come in two flavors—Hölder continuity versus Sobolev differentiability—depending on whether the analysis is done in the signal or Fourier domain. Due to the duality between Fourier decay and order of differentiation, the smoothness of \( \rho_L \) may be predicted from the growth of \( \hat{L}(\omega) \) at infinity without need for the explicit calculation of \( \rho_L \). To that end, one considers the Sobolev spaces \( W^s_2(\mathbb{R}^d) \) which are defined as

\[
W^s_2(\mathbb{R}^d) = \left\{ f : \int_{\mathbb{R}^d} \left( 1 + \| \omega \|^2 \right)^s |\hat{f}(\omega)|^2 \, d\omega < \infty \right\}.
\]
Since the partial differentiation operator $\partial^n$ corresponds to a Fourier-domain multiplication by $(|\omega|^a)^n$, the inclusion of $f$ in $W^a_2(\mathbb{R}^d)$ requires that its (partial) derivatives be well-defined in the $L^2$ sense up to order $\alpha$. The same is also true for the “Bessel potential” operators $(|\text{Id} - \Delta|^{a/2})$ of order $\alpha$, or, alternatively, the fractional Laplacians $(-\Delta)^{a/2}$ with Fourier multiplier $\|\omega\|^a$.

**Proposition 17.** Let $\beta_{L}$ be an admissible B-spline that is associated with a Fourier multiplier $\hat{L}(\omega)$ of order $\gamma$. Then, $\beta_{L} \in W^a_2(\mathbb{R}^d)$ for any $\alpha < \gamma - d/2$.

**Proof.** Because of Parseval’s identity, the statement $\beta_{L} \in W^a_2(\mathbb{R}^d)$ is equivalent to $\beta_{L} \in L_2(\mathbb{R}^d)$ and $(-\Delta)^{a/2} \beta_{L} \in L_2(\mathbb{R}^d)$. Since the first inclusion is part of the definition, it is sufficient to check for the second. To that end, we recall the stability conditions $\beta_{L} \in L_1(\mathbb{R}^d)$ and $\hat{L}(\omega)$, which are implicit to the B-spline construction (6.25). These, together with the order condition (6.14), imply that

$$|\hat{L}_{d}(\omega)| \leq \|d\|_{\ell_1},$$

$$|\beta_{L}(\omega)| = \frac{\hat{L}_{d}(\omega)}{\hat{L}(\omega)} \leq \min \left\{ \|\beta_{L}L_1\|, C \|\ell_1\|_{\ell_{\gamma}} \right\}.$$  

This latter bound allows us to control the $L_2$ norm of $(-\Delta)^{a/2} \beta_{L}$ by splitting the spectral range of integration as

$$\|(-\Delta)^{a/2} \beta_{L}\|_{L_2}^2 = \int_{\mathbb{R}^d} \|\omega\|^{2\alpha} |\hat{\beta}_{L}(\omega)|^2 \frac{d\omega}{(2\pi)^d}$$

$$= \int_{|\omega|<R} \|\omega\|^{2\alpha} |\hat{\beta}_{L}(\omega)|^2 \frac{d\omega}{(2\pi)^d} + \int_{|\omega|>R} \|\omega\|^{2\alpha} |\hat{\beta}_{L}(\omega)|^2 \frac{d\omega}{(2\pi)^d}$$

$$\leq \|\beta_{L}\|_{L_1} \int_{|\omega|<R} \|\omega\|^{2\alpha} \frac{d\omega}{(2\pi)^d} + C^2 \|d\|_{\ell_1}^2 \int_{|\omega|>R} \|\omega\|^{2\alpha-2\gamma} \frac{d\omega}{(2\pi)^d}.$$

The first integral $I_2$ is finite due the boundedness of the domain. As for $I_2$, it is convergent provided that the rate of decay of the argument is faster than $d$, which corresponds to the critical Sobolev exponent $\alpha = \gamma - d/2$.

As final step of the analysis, we invoke the Sobolev embedding theorems to infer that $\beta_{L}$ is Hölder-continuous of order $r$ with $r < \alpha - \frac{d}{2} = (\gamma - d)$, which essentially means that $\beta_{L}$ is differentiable up to order $r$ with bounded derivatives. One should keep in mind, however, that the latter estimate is a lower bound on Hölder continuity, unlike the Sobolev exponent in Proposition 17, which is sharp. For instance, in the case of the 1-D Fourier multiplier $(|\omega|^\gamma)$, we find that the corresponding (fractional) B-spline—if it exists—should have a Sobolev smoothness $(\gamma - \frac{1}{2})$, and a Hölder regularity $r < (\gamma - 1)$. Note that the latter is arbitrarily close (but not equal) to the true estimate $r_0 = (\gamma - 1)$ that is readily deduced from the Green’s function (6.15).

### 6.4.2 B-spline factorization

A powerful aspect of spline theory is that it is often possible to exploit the factorization properties of differential operators to recursively generate whole families of B-splines. Specifically, if the operator can be decomposed as $L = L_1L_2$, where the B-splines associated to $L_1$ and $L_2$ are already known, then $\beta_{L} = \beta_{L_1} * \beta_{L_2}$ is the natural choice of B-spline for $L$.

---

9. The underlying Fourier multipliers are of comparable size in the sense that there exist two constants $c_1$ and $c_2$ such that $c_1(1 + |\omega|^2)^{\alpha} \leq 1 + |\omega|^{2\alpha} \leq c_2(1 + |\omega|^2)^{\alpha}$.
Proposition 18. Let $\beta_{1,1}, \beta_{1,2}$ be admissible B-splines for the operators $L_1$ and $L_2$, respectively. Then, $\beta_1(r) = (\beta_{1,1} * \beta_{1,2})(r)$ is an admissible B-spline for $L = L_1 L_2$ if and only if there exists a constant $A > 0$ such that

$$\sum_{n \in \mathbb{Z}^d} |\hat{\beta}_{1,1}(\omega + 2\pi n)\hat{\beta}_{1,2}(\omega + 2\pi n)|^2 \geq A > 0.$$ 

for all $\omega \in [0, 2\pi]^d$. When $L_1 = L_1^\gamma$ for $\gamma \geq 0$, then the auxiliary condition is automatically satisfied.

Proof. Since $\beta_{1,1}, \beta_{1,2} \in L_1(\mathbb{R}^d)$, the same holds true for $\beta_1$ (by Young’s inequality). From the Fourier-domain definition (6.25) of the B-splines, we have

$$\hat{\beta}_{1,1}(\omega) = \frac{\sum_{k \in \mathbb{Z}^d} d_1[k] \delta(k, \omega)}{L_1(\omega)} = \frac{\hat{L}_{d,1}(\omega)}{L_1(\omega)},$$

which implies that

$$\beta_1 = \mathcal{F}^{-1} \left\{ \frac{\hat{L}_{d,1}(\omega) \hat{L}_{d,2}(\omega)}{L_1(\omega) \hat{L}_2(\omega)} \right\} = \mathcal{F}^{-1} \left\{ \frac{\hat{L}_d(\omega)}{L(\omega)} \right\} = L^{-1} L_d \delta$$

with $L^{-1} = L_2^{-1} L_1^{-1}$ and $L_d = L_{d,1} L_{d,2}$. This translates into the combined localization operator $L_d s(r) = \sum_{k \in \mathbb{Z}^d} d_1[k] s(r-k)$ with $d_1[k] = (d_1 * d_2)[k]$, which is factorizable by construction. To establish the existence of the upper Riesz bound for $\beta_1$, we perform the manipulation

$$A\beta_1(\omega) = \sum_{n \in \mathbb{Z}^d} |\hat{\beta}_{1,1}(\omega + 2\pi n)|^2$$

$$= \sum_{n \in \mathbb{Z}^d} |\hat{\beta}_{1,1}(\omega + 2\pi n)|^2 |\hat{\beta}_{1,2}(\omega + 2\pi n)|^2$$

$$\leq \left( \sum_{n \in \mathbb{Z}^d} |\hat{\beta}_{1,1}(\omega + 2\pi n)| \right)^2 \left( \sum_{n \in \mathbb{Z}^d} |\hat{\beta}_{1,2}(\omega + 2\pi n)| \right)^2$$

$$\leq \sum_{n \in \mathbb{Z}^d} |\hat{\beta}_{1,1}(\omega + 2\pi n)|^2 \sum_{n \in \mathbb{Z}^d} |\hat{\beta}_{1,2}(\omega + 2\pi n)|^2$$

$$\leq B_1^2 B_2^2 < +\infty$$

where the third line follows from the norm inequality $\|a\|_{\ell_2} \leq \|a\|_{\ell_1}$ and the fourth from Cauchy-Schwarz; $B_1$ and $B_2$ are the upper Riesz bounds of $\beta_{1,1}$ and $\beta_{1,2}$, respectively. The additional condition in the proposition takes care of the lower Riesz bound.

6.4.3 Polynomial B-splines

The factorization property is directly applicable to the construction of the polynomial B-splines (we use the equivalent notation $\beta_{n} = \beta_{D_{n+1}}$ in Section 1.3.2) via the iterated convolution of a B-spline of degree 0. Specifically,

$$\beta_{D^{n+1}}(r) = (\beta_D * \beta_D^{n})(r) = (\beta_D * \cdots * \beta_D)(r)_{n+1}$$

with $\beta_D = \beta_0^{+} = \delta_0(1)$ and the convention that $\beta_D^{0} = \beta_{ld} = \delta$.  

6.4. Generalized B-spline basis

6.4.4 Exponential B-splines

More generally, one can consider a generic \( N \)th-order differential operator of the form \( P_\alpha = P_{\alpha_1} \cdots P_{\alpha_N} \) with parameter vector \( \alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{C}^N \) and \( P_{\alpha_n} = D - \alpha_n \text{Id} \). The corresponding basis function is an exponential B-spline of order \( N \) with parameter vector \( \alpha \), which can be decomposed as

\[
\beta_\alpha(r) = \left( \beta_{\alpha_1} \ast \beta_{\alpha_2} \ast \cdots \ast \beta_{\alpha_N} \right)(r)
\]

(6.42)

where \( \beta_\alpha = \beta_{\alpha_n} \) is the first-order exponential spline defined by (6.21). The Fourier-domain counterpart of (6.42) is

\[
\hat{\beta}_\alpha(\omega) = \prod_{n=1}^N \frac{1 - e^{\alpha_n \omega}}{\omega - \alpha_n},
\]

(6.43)

which also yields

\[
\beta_\alpha(r) = \Delta_\alpha \rho_\alpha(r),
\]

(6.44)

where \( \Delta_\alpha = \Delta_{\alpha_1} \cdots \Delta_{\alpha_N} \) (with \( \Delta_\alpha \) defined by (6.20)) is the corresponding \( N \)th-order localization operator (weighted differences) and \( \rho_\alpha \) the causal Green’s function of \( P_\alpha \). Note that the complex parameters \( \alpha_n \), which are the roots of the characteristic polynomial of \( P_\alpha \), are the poles of the exponential B-spline, as seen in (6.43). The actual recipe for localization is that each pole is cancelled by a corresponding \((2\pi\text{-periodic})\) zero in the numerator.

Based on the above equations, one can infer the following properties of the exponential B-splines (see [UB05a] for a complete treatment of the topic):

- They are causal, bounded, and compactly supported in \([0, N]\), simply because all elementary constituents in (6.42) are bounded and supported in \([0, 1]\).
- They are piecewise-exponential with joining points at the integers and a maximal degree of smoothness (spline property). The first part follows from (6.44) using the well-known property that the causal Green’s function of an \( N \)th-order ordinary differential operator is an exponential polynomial restricted to the positive axis. As for the statement about smoothness, the B-splines are Hölder-continuous of order \((N-1)\). In other words, they are differentiable up to order \((N-1)\) with bounded derivatives. This follows from the fact that \( D\beta_{\alpha_n}(r) = \delta(r) - e^{\alpha_n \delta(r-1)} \), which implies that every additional elementary convolution factor in (6.42) improves the differentiability of the resulting B-spline by one.
- They are the shortest elementary constituents of exponential splines (maximally localized kernels) and they each generate a valid Riesz basis (by integer shifting) of the spaces of cardinal \( P_\alpha \)-splines if and only if \( \alpha_n - \alpha_m \neq \pm 2\pi k, k \in \mathbb{Z} \), for all distinct, purely imaginary poles.
- They reproduce the exponential polynomials that are in the null space of the operator \( P_\alpha \), as well as any of its Green’s functions \( \rho_\alpha \), which all happen to be special types of \( P_\alpha \)-splines (with a minimum number of singularities).
- For \( \alpha = (0, \ldots, 0) \), one recovers Schoenberg’s classical polynomial B-splines of degree \((N-1)\) [Sch46, Sch73], as expressed by the notational equivalence

\[
\beta^n_\alpha(r) = \beta^1_{D_{n+1}}(r) = \beta_{(0, \ldots, 0)}(r).
\]

The system-theoretic interpretation is that the classical polynomial spline of degree \( n \) has a pole of multiplicity \((n+1)\) at the origin: It corresponds to an (unstable) linear system that is an \((n+1)\)-fold integrator.
There is also a corresponding B-spline calculus whose main operations are
- Convolution by concatenation of parameter vectors:
  \[(\beta_{\alpha_1} \ast \beta_{\alpha_2})(r) = \beta_{(\alpha_1; \alpha_2)}(r).\]
- Mirroring by sign change
  \[\beta_{\alpha}(-r) = \left( \prod_{n=1}^{N} e^{\alpha_n} \right) \beta_{-\alpha}(r + N).\]
- Complex-conjugation
  \[\bar{\beta}_{\alpha}(r) = \beta_{(\alpha)}.\]
- Modulation by parameter shifting
  \[e^{j\omega r} \beta_{\alpha}(r) = \bar{\beta}_{\alpha + j\omega}(r)\]
  with the convention that \(j = (j, \ldots, j)\).

Finally, we like to point out that exponential B-splines can be computed explicitly on a case-by-case basis using the mathematical software described in [Uns05, Appendix A].

### 6.4.5 Fractional B-splines

The fractional splines are an extension of the polynomial splines for all non-integer degrees \(\alpha > -1\). The most notable members of this family are the causal fractional B-splines \(\beta_{\alpha}^{\nu}\) whose basic constituents are piecewise-power functions of degree \(\alpha\) [UB00]. These functions are associated with the causal fractional derivative operator \(D^{\alpha+1}\) whose Fourier-based definition is

\[
D^{\gamma}\varphi(r) = \int_{\mathbb{R}} (j\omega)^{\gamma} \hat{\varphi}(\omega) e^{j\omega r} \frac{d\omega}{2\pi},
\]

in the sense of generalized functions. The causal Green’s function of \(D^{\gamma}\) is the one-sided power function of degree \((\gamma - 1)\) specified by (6.15). One constructs the corresponding B-splines through a localization process similar to the classical one, replacing finite differences by the fractional differences defined as

\[
\Delta_{\omega}^{\gamma}\varphi(r) = \int_{\mathbb{R}} (1 - e^{-j\omega r})^{\gamma} \hat{\varphi}(\omega) e^{j\omega r} \frac{d\omega}{2\pi}.
\]

In that respect, it is important to note that \((1 - e^{-j\omega r})^{\gamma} = (j\omega)^{\gamma} + O(|\omega|^{2\gamma})\) which justifies this particular choice. By applying (6.25), we readily obtain the Fourier-domain representation of the fractional B–splines

\[
\hat{\beta}_{\alpha}^{\nu}(\omega) = \left( \frac{1 - e^{-j\omega}}{j\omega} \right)^{\alpha+1}
\]

which can then be inverted to provide the explicit time-domain formula

\[
\beta_{\alpha}^{\nu}(r) = \frac{\Delta_{\omega}^{\alpha+1} \omega}{\Gamma(\alpha + 1)}
\]

\[
= \sum_{m=0}^{\infty} (-1)^m \binom{\alpha + 1}{m} \frac{(r - m)^{\gamma}}{\Gamma(\alpha + 1)},
\]

where the generalized fractional binomial coefficients are given by

\[
\binom{\alpha + 1}{m} = \frac{\Gamma(\alpha + 2)}{\Gamma(m + 1)\Gamma(\alpha + 2 - m)} = \frac{(\alpha + 1)!}{m!(\alpha + 1 - m)!}.
\]
6.4. Generalized B-spline basis

What is remarkable with this construction is the way in which the classical B-spline formulas of Section 1.3.2 carry over to the fractional case almost literally by merely replacing \( n \) by \( \alpha \). This is especially striking when we compare (6.47) to (1.11), as well as the expanded versions of these formulas given below, which follow from the (generalized) binomial expansion of \((1 - e^{-j\omega})^{\alpha+1}\).

Likewise, it is possible to construct the \((\alpha, \tau)\) extension of these B-splines. They are associated with the operators \( L = \partial_{\alpha+1}^\tau \rightleftharpoons (j\omega)^{\frac{\alpha+1}{2}} + (-j\omega)^{\frac{\alpha+1}{2}} - \tau \) and \( \tau \in \mathbb{R} \) [BU03]. This family covers the entire class of translation- and scale-invariant operators in 1-D (see Proposition 14).

The fractional B-splines share virtually all the properties of the classical B-splines, including the two-scale relation, and can also be used to define fractional wavelet bases with an order \( \gamma = \alpha + 1 \) that varies continuously. They only lack positivity and compact support. Their most notable properties are summarized below.

- **Generalization:** For \( \alpha \) integer, they are equivalent to the classical polynomial splines. The fractional B-splines interpolate the polynomial ones in very much the same way as the gamma function interpolates the factorials.
- **Stability:** All brands of fractional B-splines satisfy the Riesz-basis condition in Theorem 21.
- **Regularity:** The fractional splines are \( \alpha \)-Hölder continuous; their critical Sobolev exponent (degree of differentiability in the \( L_2 \) sense) is \( \alpha + 1/2 \) (see Proposition 17).
- **Polynomial reproduction:** The fractional B-splines reproduce the polynomials of degree \( N = \lfloor \alpha + 1 \rfloor \) that are in the null space of the operator \( \mathcal{D}^{\alpha+1} \) (see Section 6.2.1).
- **Decay:** The fractional B-splines decay at least like \( |r|^{-\alpha-2} \); the causal ones are compactly supported for \( \alpha \) integer.
- **Order of approximation:** The fractional splines have the non-integer order of approximation \( \alpha + 1 \), a property that is rather unusual in approximation theory.
- **Fractional derivatives:** Simple formulas are available for obtaining the fractional derivatives of B-splines. In addition, the corresponding fractional spline wavelets essentially behave like fractional-derivative operators.

6.4.6 Additional brands of univariate B-splines

To be complete, we briefly mention some additional types of univariate B-splines that have been investigated systematically in the literature.

- The generalized exponential B-splines of order \( N \) that cover the whole class of differential operators with rational transfer functions [Uns05]. These are parameterized by their poles and zeros. Their properties are very similar to those of the exponential B-splines of the previous section, which are included as a special case.
- The Matérn splines of (fractional) order \( \gamma \) and parameter \( \alpha \in \mathbb{R}^+ \) with \( L = (D + \alpha Id)^{\gamma} \rightleftharpoons (j\omega + \alpha)^{\gamma} \) [RU06]. These constitute the fractionalization of the exponential B-spline with a single pole of multiplicity \( N \).

In principle, it is possible to construct even broader families via the convolution of existing components. The difficulty is that it may not always be possible to obtain explicit signal-domain formulas, especially when some of the constituents are fractional.

6.4.7 Multidimensional B-splines

While the construction of B-splines is well understood and covered systematically in 1-D, the task becomes more challenging in multiple dimensions because of the inherent difficulty of imposing compact support. Apart from the easy cases where the operator \( L \)
can be decomposed in a succession of 1-D operators (tensor-product B-splines and box splines), the available collection of multidimensional B-splines is much more restricted than in the univariate case. The construction of B-splines is still considered an art where the ultimate goal is to produce the most localized basis functions. The primary families of multidimensional B-splines that have been investigated so far are

- The polyharmonic B-splines of (fractional) order $\gamma$ with $L = (-\Delta)^{\frac{\gamma}{2}}$ [MN90b, Rab92a, Rab92b, VDVBU05].
- The box splines of multiplicity $N \geq d$ with $L = D_{u_1} \cdots D_{u_N} \mapsto \prod_{n=1}^{N} \langle \omega, u_n \rangle$ with $\|u_n\| = 1$, where $D_{u_n} = \langle \nabla, u_n \rangle$ is the directional derivative along $u_n$ [dBHR93]. The box splines are compactly supported functions in $L_1(\mathbb{R}^d)$ if and only if the set of orientation vectors $\{u_n\}_{n=1}^{N}$ forms a frame of $\mathbb{R}^d$.

We encourage the reader who finds the present list incomplete to work on expanding it. The good news for the present study is that the polyharmonic B-splines are particularly relevant for image-processing applications because they are associated with the class of operators that are scale- and rotation-invariant. They naturally come into play when considering isotropic fractal-type random fields.

The principal message of this section is that B-splines—no matter the type—are localized functions with an equivalent width that increases in proportion to the order. In general, the fractional brands and the non-separable multidimensional ones are not compactly supported. The important issue of localization and decay is not yet fully resolved in higher dimensions. Also, since $L_{d, \beta} = \beta L_s$, it is clear that the search for a “good” B-spline $\beta L$ is intrinsically related to the problem of finding an accurate numerical approximation $L_{d}$ of the differential operator $L$. Looking at the discretization issue from the B-spline perspective leads to new insights and sometimes to nonconventional solutions. For instance, in the case of the Laplacian $L = \Delta$, the continuous-domain localization requirement points to the choice of the 2D discrete operator $\Delta_d$ described by the $3 \times 3$ filter mask

\[
\text{Isotropic discrete Laplacian: } \frac{1}{6} \begin{pmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{pmatrix}
\]

which is not the standard version used in numerical analysis. This particular set of weights produces a much nicer, bell-shaped polyharmonic B-spline than the conventional finite-difference mask which induces significant directional artifacts, especially when one starts iterating the operator [VDVBU05].

### 6.5 Generalized operator-like wavelets

In direct analogy with the first-order scenario in Section 6.3.3, we shall now take advantage of the general B-spline formalism to construct a wavelet basis that is matched to some generic operator $L$.

#### 6.5.1 Multiresolution analysis of $L_2(\mathbb{R}^d)$

The first step is to lay out a fine-to-coarse sequence of (multidimensional) $L$-spline spaces in essentially the same way as in our first-order example. Specifically,

\[
V_l = \{ s(r) \in L_2(\mathbb{R}^d) : L_s(r) = \sum_{k \in \mathbb{Z}^d} a_l[k] \delta(r - D^l k) \}
\]
where $D$ is a proper dilation matrix with integer entries (e.g., $D = 2I$ in the standard dyadic configuration). These spline spaces satisfy the general embedding relation $V_i \supseteq V_j$ for $i \leq j$.

The reference space ($i = 0$) is the space of cardinal $L$ splines which admits the standard B-spline representation

$$V_0 = \{ s(r) = \sum_{k \in \mathbb{Z}^d} c[k] \beta_L(r-k) : c \in \ell_2(\mathbb{Z}^d) \},$$

where $\beta_L$ is given by (6.25). Our implicit assumption is that each $V_i$ admits a similar B-spline representation

$$V_i = \{ s(r) = \sum_{k \in \mathbb{Z}^d} c_i[k] \beta_{L,i}(r-D^i k) : c_i \in \ell_2(\mathbb{Z}^d) \},$$

which involves the multiresolution generators $\beta_{L,i}$ described in the next section.

### 6.5.2 Multiresolution B-splines and the two-scale relation

In direct analogy with (6.25), the multiresolution B-splines $\hat{\beta}_{L,i}$ are localized versions of the Green’s function $\rho_L$ with respect to the grid $D^i \mathbb{Z}^d$. Specifically, we have that

$$\beta_{L,i}(r) = \sum_{k \in \mathbb{Z}^d} d_i[k] \rho_L(r-D^i k) = L_{d,i} L^{-1} \delta(r),$$

where $L_{d,i}$ is the discretized version of $L$ on the grid $D^i \mathbb{Z}^d$. The Fourier-domain counterpart of this equation is

$$\hat{\beta}_{L,i}(\omega) = \frac{\sum_{k \in \mathbb{Z}^d} d_i[k] e^{-i \omega \cdot D^i k}}{L(\omega)}. \quad (6.48)$$

The implicit requirement for the multiresolution decomposition scheme to work is that $\beta_{L,i}$ generates a Riesz basis. This needs to be asserted on a case-by-case basis.

A particularly favorable situation occurs when the operator $L$ is scale-invariant with $L(\omega) = |\omega|^\gamma L(\omega)$. Let $i' > i$ be two multiresolution levels of the pyramid such that $D^{i'-i} = mI$ where $m$ is a proportionality constant. It is then possible to relate the B-spline at resolution $i'$ to the one at the finer level $i$ via the simple dilation relation

$$\beta_{L,i'}(r) \propto \beta_{L,i}(r/m).$$

This is shown by considering the Fourier transform of $\beta_{L,i}(r/m)$, which is written as

$$|m|^d \hat{\beta}_{L,i}(m\omega) = |m|^d \frac{\sum_{k \in \mathbb{Z}^d} d_i[k] e^{-i \omega \cdot m D^i k}}{L(m\omega)}$$

$$= |m|^d \frac{\sum_{k \in \mathbb{Z}^d} d_i[k] e^{-i \omega \cdot D^{i'} k}}{L(\omega)}$$

$$= |m|^d \frac{\sum_{k \in \mathbb{Z}^d} d_i[k] e^{-i \omega \cdot D^i k}}{L(\omega)},$$

and found to be compatible with the form of $\hat{\beta}_{L,i'}(\omega)$ given by (6.48) by taking $d_i/k \propto d_i[k]$. The prototypical scenario is the dyadic configuration $D = 2I$ for which the B-splines at level $i$ are all constructed through the dilation of the single prototype $\beta_L = \hat{\beta}_{L,0}$ subject to the scale-invariance constraint on $L$. This happens, for instance, for the classical polynomial splines which are associated with the Fourier multipliers $(j\omega)^N$.  

119
A crucial ingredient for the fast wavelet-transform algorithm is the two-scale relation that links the B-splines basis functions at two successive levels of resolution. Specifically, we have that

$$\hat{h}_{i+1}(\omega) = \sum_{k \in \mathbb{Z}^d} d_{i+1}[k] e^{-j \omega \cdot D^{i+1} k},$$

where the sequence $h_i$ specifies the scale-dependent refinement filter. The frequency response of $h_i$ is obtained by taking the ratios of the Fourier transforms of the corresponding B-splines as

$$\hat{h}_i(\omega) = \frac{\hat{\beta}_{i+1}(\omega)}{\hat{\beta}_{i+1}(\omega)} = \frac{\sum_{k \in \mathbb{Z}^d} d_{i+1}[k] e^{-j \omega \cdot D^{i+1} k}}{\sum_{k \in \mathbb{Z}^d} d_i[k] e^{-j \omega \cdot D^i k}},$$

which is $2\pi(D^T)^{-i}$ periodic and hence defines a valid digital filter with respect to the spatial grid $D^i \mathbb{Z}^d$.

To illustrate those relations, we return to our introductory example in Section 6.1: the Haar wavelet transform, which is associated with the Fourier multipliers $j \omega$ (derivative) and $(1 - e^{-j \omega})$ (finite-difference operator). The dilation matrix is $D = 2$ and the localization filter is the same at all levels because the underlying derivative operator is scale-invariant. By plugging those entities into (6.48), we obtain the Fourier transform of the corresponding B-spline at resolution $i$ as

$$\hat{\beta}_{D,i}(\omega) = 2^{-i/2} \frac{1 - e^{2j \omega}}{j \omega},$$

where the normalization by $2^{-i/2}$ is included to standardize the norm of the B-splines. The application of (6.49) then yields

$$\hat{h}_i(\omega) = \frac{1}{\sqrt{2}} \frac{1 - e^{2j \omega}}{1 - e^{2j \omega}} = \frac{1}{\sqrt{2}} (1 + e^{2j \omega}),$$

which, up to the normalization by $\sqrt{2}$, is the expected refinement filter with coefficients proportional to $(1,1)$ that are independent upon the scale.

### 6.5.3 Construction of an operator-like wavelet basis

To keep the notation simple, we concentrate on the specification of the wavelet basis at the scale $i = 1$ with $W_1 = \text{span}\{\psi_{1,k}\}_{k \in \mathbb{Z}^d}$ such that $W_1 \perp V_1$ and $V_0 = V_1 + W_1$, where $V_0 = \text{span}\{\beta_{1,k}^{\top} - k\}_{k \in \mathbb{Z}^d}$ is the space of cardinal L-splines.

The relevant smoothing kernel is the interpolation function $\varphi_{\text{int}} = \varphi_{\text{int},0}$ for the space of cardinal $L^H$-L-splines, which is generated by $(\beta_{1}^\top \ast \beta_{1})(r)$ (autocorrelation of the generalized B-spline). This interpolator is best described in the Fourier domain using the formula

$$\varphi_{\text{int}}(r) = \mathcal{F}^{-1} \left\{ \frac{|\hat{\beta}_1(\omega)|^2}{\sum_{n \in \mathbb{Z}^d} |\hat{\beta}_1(\omega + 2\pi n)|^2} \right\}(r),$$

(6.51)
where \( \hat{\beta}_L \) (resp., \( \overline{\beta}_L \)) is the Fourier transform of the generalized B-spline \( \beta_L \) (resp., \( \overline{\beta}_L \)). It satisfies the fundamental interpolation property

\[
q_{\text{int}}(k) = \delta|k| = \begin{cases} 1, & \text{for } k = 0 \\ 0, & \text{for } k \in \mathbb{Z}^d \setminus \{0\}.
\end{cases}
\] (6.52)

The existence of such a function is guaranteed whenever \( \hat{\beta}_L = \hat{\beta}_{L,0} \) is an admissible B-spline. In particular, the Riesz basis condition (6.26) implies that the denominator of \( \hat{q}_{\text{int}}(\omega) \) in (6.51) is non-vanishing.

The sought-after wavelets are then constructed as \( \psi_{1,k}(r) = \psi_L(r - k)/\|\psi_L\|_{L^2} \), where the operator-like mother wavelet \( \psi_L \) is given by

\[
\psi_L(r) = L^H q_{\text{int}}(r),
\] (6.53)

where \( L^H \) is the adjoint of \( L \) with respect to the Hermitian-symmetric \( L_2 \) inner product. Also, note that we are removing the functions located on the next coarser resolution grid \( \mathbb{DZ}^d \) associated with \( V_1 \) (critically sampled configuration).

The proof of the following result is illuminating because it relies heavily on the notion of duality which is central to our whole argumentation.

**Proposition 19.** The operator-like wavelet \( \psi_L = L^H q_{\text{int}} \) satisfies the property \( \langle s, \psi_L(-k) \rangle_{L^2} = \langle s, L^* q_{\text{int}}(-k) \rangle \) for any spline \( s \in V_1 \). Moreover, it can be written as \( \psi_L(r) = L^H \hat{\beta}_L(r) = \sum_{k \in \mathbb{Z}^d} d_L[-k] \hat{\beta}_L(r - k) \) where \( \{\hat{\beta}_L(-k)\}_{k \in \mathbb{Z}^d} \) is the dual basis of \( V_0 \) such that \( \langle \hat{\beta}_L(-k), \hat{\beta}_L(-k') \rangle_{L_2} = \delta[k-k'] \). This implies that \( W_1 = \text{span}\{\psi_1,k\}_{k \in \mathbb{DZ}^d} \subset V_0 \) and \( W_1 \perp V_1 \).

**Proof.** We pick an arbitrary spline \( s_1 \in V_1 \) and perform the inner-product manipulation

\[
\langle s_1, \psi_L(-k_0) \rangle_{L^2} = \langle s_1, L^* q_{\text{int}}(-k_0) \rangle \text{ (by shift-invariance of } L) = \langle L s_1, q_{\text{int}}(-k_0) \rangle \text{ (by duality)} = \langle \sum_{k \in \mathbb{Z}^d} a_1[k] \delta(-Dk), q_{\text{int}}(-k_0) \rangle \text{ (by definition of } V_1) = \sum_{k \in \mathbb{Z}^d} a_1[k] q_{\text{int}}(Dk-k_0). \text{ (by definition of } \delta)\]

Due to the interpolation property of \( q_{\text{int}} \), the kernel values in the sum are vanishing if \( Dk-k_0 \in \mathbb{Z}^d \setminus \{0\} \) for all \( k \in \mathbb{Z}^d \), which proves the first part of the statement.

As for the second claim, we consider the Fourier-domain expression of \( \psi_L \):

\[
\hat{\psi}_L(\omega) = \overline{L(\omega)} q_{\text{int}}(\omega) = \overline{L(\omega)} \hat{\beta}_L(\omega) \hat{\beta}_L(\omega)
\]

where

\[
\hat{\beta}_L(\omega) = \frac{\hat{\beta}_L(\omega)}{\sum_{n \in \mathbb{Z}^d | \hat{\beta}_L(\omega + 2\pi n)|^2}
\]

is the Fourier transform of the dual B-spline \( \hat{\beta}_L \). The above factorization implies that \( q_{\text{int}}(r) = (\hat{\beta}_L * \overline{\hat{\beta}_L})(r) \), which ensures the biorthonormality \( \langle \hat{\beta}_L(-k), \hat{\beta}_L(-k') \rangle_{L_2} = \delta[k-k'] \) of the basis functions.

Finally, by replacing \( \hat{\beta}_L(\omega) \) by its explicit expression (6.25), we show that \( \overline{L(\omega)} \hat{\beta}_L(\omega) = \overline{\hat{L}_d(\omega)} \), where \( \hat{L}_d(\omega) = \sum_{k \in \mathbb{DZ}^d} d_L[k] e^{-ik \cdot \omega} \) is the frequency response of the (discrete) operator \( L_d \). This implies that \( \hat{\psi}_L = \overline{\hat{L}_d} \hat{\beta}_L \), which is the Fourier equivalent of \( \psi_L = L^H \hat{\beta}_L \). \( \square \)
This interpolation-based method of construction is applicable to all the wavelet subspaces \( W_i \) and leads to the specification of operator-like Riesz bases of \( L^2(\mathbb{R}^d) \) under relatively mild assumptions on \( L \) [KUW]. Specifically, we have that \( W_i = \text{span}\{\psi_{i,k}\}_{k \in \mathbb{Z}^d} \) with

\[
\psi_{i,k}(r) \propto L^H \varphi_{\text{int},i-1}(r - D^{i-1} k),
\]

where \( \varphi_{\text{int},i-1} \) is the \( L^H \)-spline interpolator on the grid \( D^{i-1} \mathbb{Z}^d \). The fact that the interpolator is specified with respect to the grid of the next finer spline space \( V_{i-1} = \text{span}\{\beta_{L,i-1}(-D^{i-1} k)\}_{k \in \mathbb{Z}^d} \) is essential to ensure that \( W_i \subset V_{i-1} \). This kernel satisfies the fundamental interpolation property

\[
\varphi_{\text{int},i-1}(D^{i-1} k) = \delta[k],
\]

which results in \( W_i \) being orthogonal to \( V_i = \text{span}\{\beta_{L,i}(-D^i k)\}_{k \in \mathbb{Z}^d} \) (the reasoning is the same as in the proof of Proposition 19 which covers the case \( i = 1 \)). For completeness, we also provide the general expression of the Fourier transform of \( \varphi_{\text{int},i} \),

\[
\hat{\varphi}_{\text{int},i}(\omega) = |\text{det}(D)|^i \left( \sum_{n \in \mathbb{Z}^d} \frac{|\hat{\beta}_{L,i}(\omega + 2\pi(D^T)^{-1} n)|^2}{|\hat{\beta}_{L,i}(\omega)|^2} \right),
\]

which can be used to show that \( L^H \varphi_{\text{int},i}(-D^i k) \propto L^H \hat{\beta}_{L,i}(-D^i k) \in V_i \) for any \( k \in \mathbb{Z}^d \).

While we have seen that this scheme produces an orthonormal basis for the first-order operator \( P_n \) in Section 6.3.3, the general procedure does only guarantee semi-orthogonality. More precisely, it ensures the orthogonality between the wavelet subspaces \( W_i \). If necessary, one can always fix the intra-scale orthogonality a posteriori by forming appropriate linear combinations of wavelets at a given resolution. The resulting orthogonal wavelets will still be \( L \)-admissible in the sense of Definition 24. However, for \( d > 1 \), intra-scale orthogonalization is likely to spoil the simple, convenient structure of the above construction which uses a single generator per scale, irrespective of the number of dimensions. Indeed, the examples of multidimensional orthogonal wavelet transforms that can be found in the literature—either separable, or non-separable—systematically involve \( M = (\text{det}(D) - 1) \) distinct wavelet generators per scale. Moreover, unlike the present operator-like wavelets, they do generally not admit an explicit analytical description.

In summary, wavelets generally behave like differential operators and it is possible to match them to a given class of stochastic processes. The wavelet transforms that are currently most widely used in applications act as multiscale derivatives or Laplacians. They are therefore best suited for the representation of fractal-type stochastic processes that are defined by scale-invariant SDEs [TVDVU09].

The general theme that emerges is that a signal transform will behave appropriately if it has the ability to suppress the signal components (polynomial or sinusoidal trends) that are in the null space of the whitening operator \( L \). This will result in a stationarizing effect that is well-documented in the Gaussian context [Fla89, Fla92]. This is the fundamental reason why vanishing moments are so important.
6.6 Bibliographical notes

Section 6.1

Alfred Haar constructed the orthogonal Haar system as part of his Ph.D. thesis, which he defended in 1909 under the supervision of David Hilbert [Haa10]. From then on, the Haar system remained relatively unnoticed until it was revitalized by the discovery of wavelets nearly one century later. Stéphane Mallat set the foundation of the multiresolution theory of wavelets in [Mal89] with the help of Yves Meyer, while Ingrid Daubechies constructed the first orthogonal family of compactly supported wavelets [Dau88]. Many of the early constructions of wavelets are based on splines [Mal89, CW91, UAE92, UAE93]. The connection with splines is actually quite fundamental in the sense that all multiresolution wavelet bases, including the non-spline brands such as Daubechies', necessarily include a B-spline as a convolution factor—the latter is responsible for their primary mathematical properties such as vanishing moments, differentiability, and order of approximation [UB03]. Further information on wavelets can be found in several textbooks [Dau92, Mey90, Mal09].

Section 6.2

Splines constitute a beautiful topic of investigation in their own right with hundreds of papers specifically devoted to them. The founding father of the field is Schoenberg who, during war time, was asked to develop a computational solution for constructing an analytic function that fits a given set of equidistant noisy data points [Sch88]. He came up with the concept of spline interpolation and proved that polynomial spline functions have a unique expansion in terms of B-splines [Sch46]. While splines can also be specified for nonuniform grids and extended in a variety of ways [dB78, Sch81], the cardinal setting is especially pleasing because it lends itself to systematic treatment with the aid of the Fourier transform [Sch73]. The relation between splines and differential operators was recognized early on and led to the generalization known as L-splines [SV67].

The classical reference on partial differential operators and Fourier multipliers is [Hör80]. A central result of the theory is the Malgrange-Ehrenpreiss theorem [Mal, Ehr54], and its extension stating that the convolution with a compactly supported generalized function is invertible [Hör05].

The concept of a Riesz basis is standard in functional analysis and approximation theory [Chr03]. The special case where the basis functions are integer translates of a single generator is treated in [AU94]. See also [Uns00] for a review of such representations in the context of sampling theory.

Section 6.3

The first-order illustrative example is borrowed from [UB05a, Figure 1] for the construction of the exponential B-spline, and from [KU06, Figure 1] for the wavelet part of the story.

Section 6.4

The 1-D theory of cardinal L-splines for ordinary differential operators with constant coefficients is due to Micchelli [Mic76]. In the present context, we are especially concerned with ordinary differential equations, which go hand-in-hand with the extended family of cardinal exponential splines [Uns05]. The properties of the relevant B-splines...
are investigated in full detail in [UB05a], which constitutes the ground material for Section 6.4. A key property of B-splines is their ability to reproduce polynomials. It is ensured by the Strang-Fix conditions (6.37) which play a central role in approximation theory [dB87, SF71]. While there is no fundamental difficulty in specifying cardinal-spline interpolators in multiple dimensions, it is much harder to construct compactly supported B-splines, except for the special cases of the box splines [DBH82, dBHR93] and exponential box splines [Ron88]. For elliptic operators such as the Laplacian, it is possible to specify exponentially decaying B-splines, with the caveat that the construction is not unique [MN90b, Rab92a, Rab92b]. This calls for some criterion to identify the most-localized solution [VDVBU05]. B-splines, albeit non-compactly supported ones, can also be specified for fractional operators [UB07]. This line of research was initiated by Unser and Blu with the construction of the fractional B-splines [UB00]. As suggested by the name, the (Gaussian) stochastic counterparts of these B-splines are Mandelbrot’s fractional Brownian motions [MVN68], as we shall see in Chapters 7 and 8. The association is essentially the same as the connection between the B-spline of degree 0 (rect) and Brownian motion, or by extension, the whole family of Lévy processes (see Section 1.3).

Section 6.5

de Boor et al. were among the first to extend the notion of multiresolution analysis beyond the idea of dilation and to propose a general framework for constructing “non-stationary” wavelets [dBDR93]. Khalidov and Unser proposed a systematic method for constructing wavelet-like basis functions based on exponential splines and proved that these wavelets behave like differential operators [KU06]. The material in Section 6.5 is an extension of those ideas to the case of a generic Fourier-multiplier operator in multiple dimensions; the full technical details can be found in [KUW]. Operator-like wavelets have also been specified within the framework of conventional multiresolution analysis; in particular, for the Laplacian and its iterates [VDVBU05, TVDVU09] and for the various brands of 1-D fractional derivatives [VDVFHUB10], which have the common property of being scale-invariant. Finally, we mention that each exponential-spline wavelet has a compactly supported Daubechies’ counterpart that is orthogonal and operator-like in the sense of having the same vanishing exponential moments [VBU07].