The writing of this book 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 Figure 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. 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. The nontrivial aspect of these generalizations is that
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 Figure 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-
2.1 On the implications of the innovation model

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 \(d_k\). 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 \(L_d\) 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

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

where

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

This suggests that the decoupling effect will be the strongest when the convolution kernel \(\beta_L\) is the most localized (minimum support) and closest to an impulse. 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 \(\{\psi_i(\cdot - r_k)\}\). 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

\[ \psi_i(r) = L^* \phi_i(r). \]

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

\[
\begin{align*}
\langle s, \psi_i(\cdot - r_0) \rangle &= \langle s, L^* \phi_i(\cdot - r_0) \rangle \\
&= \langle Ls, \phi_i(\cdot - r_0) \rangle \\
&= \langle w, \phi_i(\cdot - r_0) \rangle = (\phi_i^w \ast w)(r_0)
\end{align*}
\]

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.
where $\phi_\gamma^r(r) = \phi_\gamma(-r)$ is the reversed version of $\phi_\gamma$. 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 $\phi_l$ are not necessarily rescaled versions of each other.

Note that the "L-compatible" property is relatively robust. For instance, if $L = L'L_0$, then an "L-compatible" wavelet is also $L'$-compatible with $\phi_{l'} = L_0^l \phi_l$. The design challenge in the context of stochastic modeling is thus to come up with a suitable wavelet basis such that $\phi_l$ 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 $\nu(r)$) 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 be unfamiliar with concepts used in this book, 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 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-Khintchine 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 of our theory is that every admissible continuous-domain innovation for the model in Figure 2.1 belongs to the so-called family of white Lévy noises. This implies that an innovation process is completely characterized by its 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 Figure 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). The techniques are quite generic and applicable to any operator \( L \) that admits a proper inverse \( L^{-1} \). Moreover, by writing a generalized B-spline as \( \beta_L = L_0 L^{-1} \delta \), 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 \( \beta_L \), which actually results in simplified and streamlined mathematics. Remarkably, these generalized B-splines are also the key for constructing wavelet-like basis functions that are “\( L \)-compatible.”

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. Based on the characteristic form of these processes, we are able to deduce the transform-domain statistics and to precisely assess residual dependencies. These ideas are il-
Roadmap to the monograph

Illustrated with examples of sparse processes for which operator-like wavelets outperform the classical KLT (or DCT) and result in an independent component analysis.

An implicit property of the innovation model is that the statistical distribution of the inner product between a sparse stochastic process and a particular basis function (e.g., wavelet) is uniquely characterized by a “modified” Lévy exponent. Our main point in Chapter 9 is to use this result to show that the sparsity of the input noise is transferred to the transformed domain. Apart from a shaping effect that can be quantified, the resulting probability density function remains within the same family of infinite-divisible laws.

In the final part of the book, we illustrate the use of these stochastic models (and the corresponding analytical tools) with the formulation of algorithms for the recovery of signals and images from incomplete, noisy measurements. In Chapter 10, we develop a general framework for the discretization of linear inverse problems in a B-spline basis, which is analogous to the finite-element method for solving PDEs. The central element is the “projection” of the continuous-domain stochastic model onto the (finite dimensional) reconstruction space in order to specify the prior statistical distribution of the signal. This naturally yields the maximum a posteriori solution to the signal-reconstruction problem. The framework is illustrated with the derivation of practical algorithms for magnetic resonance imaging, deconvolution microscopy, and tomographic reconstruction. Remarkably, the resulting MAP estimators are compatible with the non-quadratic regularization schemes (e.g., $\ell_1$-minimization, LASSO, and/or non-convex $\ell_p$ relaxation) that are currently in favor in imaging. To get a handle on the quality of the reconstruction, we then rely on the innovation model to investigate the extent to which one is able to “optimally” denoise sparse signals. In particular, we demonstrate the feasibility of MMSE reconstruction when the signal belongs to the class of Lévy processes, which provides us with a gold standard against which to compare other algorithms.

In Chapter 11, we present alternative wavelet-based reconstruction methods that are typically faster than the fixed-scale techniques of Chapter 10. These methods capitalize on the orthonormality of the wavelet basis which provides a direct control of the norm of the signal. We show that the underlying optimization task is amenable to iterative thresholding algorithms (ISTA or FISTA) which are simple to deploy and well-suited for large-scale problems. We also investigate the effect of cycle spinning, which is a fundamental ingredient for making wavelets competitive in terms of image quality. Our closing topic is the use of statistical modeling for the improvement of standard wavelet-based denoising—in particular, the optimization of the wavelet-domain thresholding functions and the search of a consensus solution across multiple wavelet expansions in order to minimize the global estimation error.