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 is not at the forefront of research anymore.

Starting with the discovery of the wavelet transform in the late 1980s [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 book, 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 since the mid 1990s 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 many 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 ℓ_1-norm of the model coefficients [Tib96].

- *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 ℓ_0-norm minimization problem, which is NP hard, by a convex ℓ_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 non-linear signal-reconstruction algorithms*. There is increasing empirical evidence that non-linear 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 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, BT09b].

1.2 Sparse stochastic models: the 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 crucial ingredient that is required to guide such an investigation is a sparse counterpart to the classical family of Gaussian stationary processes (GSP). This

[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

book 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

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 (LMMSE 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) (see Section 10.2.2).
- *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 defined signal that can be represented in terms of generalized B-spline basis functions [UB05b].

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 ℓ_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 book 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 process by a stochastic Wiener integral,

$$s(t) = \int_{\mathbb{R}} h(t, \tau)\, \mathrm{d}W(\tau), \tag{1.1}$$

where $h(t, \tau)$ 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. W is a random measure which is such that

$$W(t) = \int_0^t \mathrm{d}W(\tau)$$

is the Wiener process, where the latter equation constitutes a special case of (1.1) with $h(t, \tau) = \mathbb{1}_{\{t>\tau\geq 0\}}$. Here, $\mathbb{1}_\Omega$ denotes the indicator function of the set Ω. If $h(t, \tau) = h(t-\tau)$ 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

$$s(t) = \int_{\mathbb{R}} h(t, \tau) w(\tau)\, \mathrm{d}\tau, \tag{1.2}$$

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 antiderivative operator. We then use the close relationship between splines and wavelets to gain insight into 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).

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 $f(t), t \in \mathbb{R}$ that satisfies the interpolation condition $f(t)|_{t=k} = f[k]$, 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

$$f_1(t) = \sum_{k\in\mathbb{Z}} f[k]\beta_+^0(t-k), \tag{1.3}$$

which involves rectangular basis functions (informally described as "Legos") that are shifted replicates of the causal[2] B-spline of degree zero:

$$\beta_+^0(t) = \begin{cases} 1, & \text{for } 0 \le t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{1.4}$$

[2] A function $f_+(t)$ is said to be causal if $f_+(t) = 0$, for all $t < 0$.

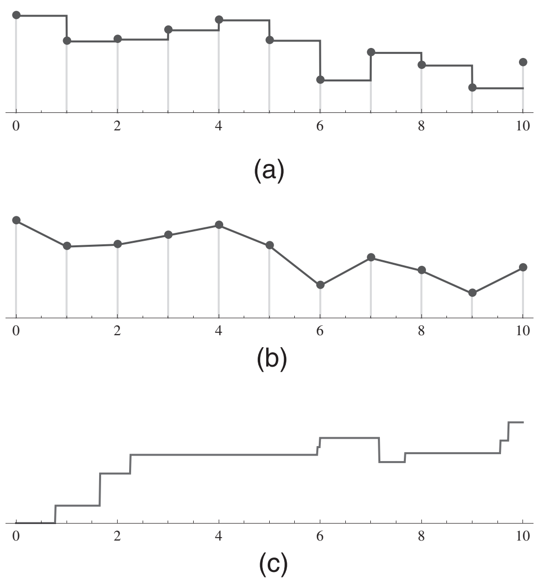

Figure 1.1 Examples of spline signals. (a) Cardinal spline interpolant of degree zero (piecewise-constant). (b) Cardinal spline interpolant of degree one (piecewise-linear). (c) Non-uniform D-spline or compound-Poisson process, depending on the interpretation (deterministic vs. stochastic).

Observe that the basis functions $\{\beta_+^0(t-k)\}_{k\in\mathbb{Z}}$ are non-overlapping and orthonormal, and that their linear span defines the space of cardinal polynomial splines of degree zero. Moreover, since $\beta_+^0(t)$ 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 Figure 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 spline specified by (1.3), we have that

$$\mathrm{D}f_1(t) = \sum_{k\in\mathbb{Z}} a_1[k]\delta(t-k), \tag{1.5}$$

where the weights of the integer-shifted Dirac impulses $\delta(\cdot - k)$ 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 $\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}$ 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 right-hand side 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 $f_1(t)$ is a *cardinal* D-spline,[3] which is a special case

[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.

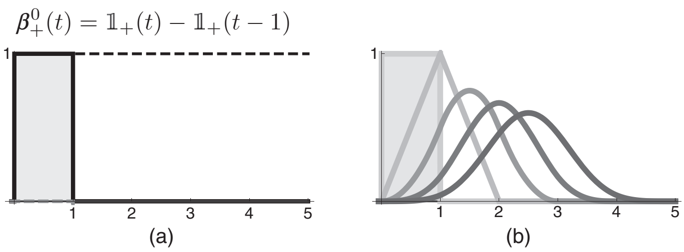

Figure 1.2 Causal polynomial B-splines. (a) Construction of the B-spline of degree zero starting from the causal Green's function of D. (b) B-splines of degree $n = 0, \ldots, 4$ (light to dark), which become more bell-shaped (and beautiful) as n increases.

of a general *non-uniform* D-spline where the knots can be located arbitrarily (see Figure 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

$$\mathrm{D_d} f(t) = f(t) - f(t-1).$$

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

$$\mathrm{D_d} f(t) = (\beta_+^0 * \mathrm{D}f)(t),$$

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

$$\beta_+^0(t) = \mathrm{D_d D}^{-1}\delta(t) = \mathrm{D_d}\mathbb{1}_+(t), \tag{1.6}$$

where the unit step $\mathbb{1}_+(t) = \mathbb{1}_{[0,+\infty)}(t)$ (a.k.a. the Heaviside function) is the causal Green's function[4] of the derivative operator. This formula is illustrated in Figure 1.2a. Its Fourier-domain counterpart is

$$\widehat{\beta}_+^0(\omega) = \int_{\mathbb{R}} \beta_+^0(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \frac{1-\mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}, \tag{1.7}$$

which is recognized as being the ratio of the frequency responses of the operators $\mathrm{D_d}$ and D, respectively.

Thus, the basic Lego component, β_+^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 ingredients in our formulation of sparse stochastic processes.

[4] We say that $\rho(t)$ is the causal Green's function of the shift-invariant operator L if ρ is causal and satisfies $\mathrm{L}\rho = \delta$. This can also be written as $\mathrm{L}^{-1}\delta = \rho$, meaning that ρ is the causal impulse response of the shift-invariant inverse operator L^{-1}.

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

$$f_2(t) = \sum_{k\in\mathbb{Z}} f[k+1]\beta_+^1(t-k), \tag{1.8}$$

where

$$\beta_+^1(t) = \left(\beta_+^0 * \beta_+^0\right)(t) = \begin{cases} t, & \text{for } 0 \leq t < 1 \\ 2-t, & \text{for } 1 \leq t < 2 \\ 0, & \text{otherwise} \end{cases} \tag{1.9}$$

is the causal B-spline of degree one, a triangular function centered at $t = 1$. Note that the use of a causal generator is compensated by the unit shifting of the coefficients in (1.8), which is equivalent to recentering 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 Figure 1b.

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

$$f_{n+1}(t) = \sum_{k\in\mathbb{Z}} c[k]\beta_+^n(t-k), \tag{1.10}$$

with

$$\beta_+^n(t) = \big(\underbrace{\beta_+^0 * \beta_+^0 * \cdots * \beta_+^0}_{n+1}\big)(t).$$

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

$$\widehat{\beta}_+^n(\omega) = \left(\frac{1-\mathrm{e}^{-\mathrm{j}\omega}}{\mathrm{j}\omega}\right)^{n+1}$$

and

$$\begin{aligned} \beta_+^n(t) &= \mathrm{D}_\mathrm{d}^{n+1}\mathrm{D}^{-(n+1)}\delta(t) \\ &= \frac{\mathrm{D}_\mathrm{d}^{n+1}(t)_+^n}{n!} \\ &= \sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}\frac{(t-k)_+^n}{n!} \end{aligned} \tag{1.11}$$

with $(t)_+ = \max(0, t)$. 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's function of D^{n+1}) is the one-sided power function $\mathrm{D}^{-(n+1)}\delta(t) = \frac{t_+^n}{n!}$. 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

$$\|f_n\|_{L_2}^2 = \int_{\mathbb{R}} |f_n(t)|^2 \, \mathrm{d}t \le \|c\|_{\ell_2}^2 = \sum_{k\in\mathbb{Z}} \big|c[k]\big|^2.$$

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 $\{\beta_+^0(\cdot - k)\}_{k\in\mathbb{Z}}$ 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

$$\begin{aligned}\mathrm{D}^n \beta_+^n(t) &= \mathrm{D}^n \mathrm{D}_\mathrm{d}^{n+1} \mathrm{D}^{-(n+1)} \delta(t) \\ &= \mathrm{D}_\mathrm{d}^n \mathrm{D}_\mathrm{d} \mathrm{D}^{-1} \delta(t) = \mathrm{D}_\mathrm{d}^n \beta_+^0(t),\end{aligned}$$

which indicates that the nth-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

$$\mathrm{D}s(t) = \sum_n A_n \delta(t - t_n) = w(t), \tag{1.12}$$

where the locations t_n of the Dirac impulses are uniformly distributed over the real line (Poisson distribution with rate parameter λ) and the weights A_n are independent identically distributed (i.i.d.) with amplitude distribution $p_A(a)$. For simplicity, we are also assuming that p_A is symmetric with finite variance $\sigma_A^2 = \int_{\mathbb{R}} a^2 p_A(a) da$. 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

$$\mathrm{D}_0^{-1}\varphi(t) = \int_{-\infty}^{t} \varphi(\tau)\,\mathrm{d}\tau - \int_{-\infty}^{0} \varphi(\tau)\,\mathrm{d}\tau = \begin{cases} \int_0^t \varphi(\tau)\,\mathrm{d}\tau, & \text{for } t \geq 0 \\ -\int_t^0 \varphi(\tau)\,\mathrm{d}\tau, & \text{for } t < 0. \end{cases}$$

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 t. The Fourier counterpart of this definition is

$$\mathrm{D}_0^{-1}\varphi(t) = \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{j}\omega t} - 1}{\mathrm{j}\omega}\widehat{\varphi}(\omega)\frac{\mathrm{d}\omega}{2\pi},$$

which can be extended, by duality, to a much larger class of generalized functions (see 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 $\mathrm{j}\omega$ is tempered by a proper correction in the numerator: $\mathrm{e}^{\mathrm{j}\omega t} - 1 = \mathrm{j}\omega t + O(\omega^2)$. 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 $\mathrm{D}\mathrm{D}_0^{-1} = \mathrm{Id}$ 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

$$\beta_+^0(t) = \mathrm{D_d}\mathrm{D}_0^{-1}\delta(t) = \mathrm{D_d}\mathrm{D}^{-1}\delta(t). \tag{1.13}$$

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:

$$\begin{aligned} s(t) = \mathrm{D}_0^{-1}w(t) &= \sum_n A_n \mathrm{D}_0^{-1}\{\delta(\cdot - t_n)\}(t) \\ &= \sum_n A_n \big(\mathbb{1}_+(t - t_n) - \mathbb{1}_+(-t_n)\big), \end{aligned} \tag{1.14}$$

where the second term $\mathbb{1}_+(-t_n)$ in the last parenthesis ensures that $s(0) = 0$. Clearly, the signal defined by (1.14) is piecewise-constant (random spline of degree zero) 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 Figure 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

$$\mathbb{E}\{w(t)w(t')\} = R_w(t - t') = \sigma_w^2\delta(t - t'), \tag{1.15}$$