An Introduction to Sparse Stochastic Processes

Michael Unser and Pouya D. Tafti

An Introduction to Sparse Stochastic Processes

MICHAEL UNSER and POUYA D. TAFTI

École Polytechnique Fédérale de Lausanne

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107058545

First published 2014

Printed in the United Kingdom by Clays, St Ives plc

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication Data
Unser, Michael A., author.
An introduction to sparse stochastic processes / Michael Unser and Pouya Tafti, École polytechnique fédérale, Lausanne.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-05854-5 (Hardback)
1. Stochastic differential equations. 2. Random fields. 3. Gaussian processes. I. Tafti, Pouya, author. II. Title.
QA274.23.U57 2014
519.2′3–dc23 2014003923

ISBN 978-1-107-05854-5 Hardback

An Introduction to Sparse Stochastic Processes

Providing a novel approach to sparsity, this comprehensive book presents the theory of stochastic processes that are ruled by linear stochastic differential equations and that admit a parsimonious representation in a matched wavelet-like basis.

Two key themes are the statistical property of infinite divisibility, which leads to two distinct types of behavior – Gaussian and sparse – and the structural link between linear stochastic processes and spline functions, which is exploited to simplify the mathematical analysis. The core of the book is devoted to investigating sparse processes, including a complete description of their transform-domain statistics. The final part develops practical signal-processing algorithms that are based on these models, with special emphasis on biomedical image reconstruction.

This is an ideal reference for graduate students and researchers with an interest in signal/image processing, compressed sensing, approximation theory, machine learning, or statistics.

Michael Unser is Professor and Director of the Biomedical Imaging Group at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He is a member of the Swiss Academy of Engineering Sciences, a Fellow of EURASIP, and a Fellow of the IEEE.

Pouya D. Tafti is a data scientist currently residing in Germany, and a former member of the Biomedical Imaging Group at EPFL, where he conducted research on the theory and applications of probabilistic models for data.

"Over the last twenty years, sparse representation of images and signals became a very important topic in many applications, ranging from data compression, to biological vision, to medical imaging. The book *Sparse Stochastic Processes* by Unser and Tafti is the first work to systematically build a coherent framework for non-Gaussian processes with sparse representations by wavelets. Traditional concepts such as Karhunen-Loéve analysis of Gaussian processes are nicely complemented by the wavelet analysis of Levy Processes which is constructed here. The framework presented here has a classical feel while accommodating the innovative impulses driving research in sparsity. The book is extremely systematic and at the same time clear and accessible, and can be recommended both to engineers interested in foundations and to mathematicians interested in applications."

David Donoho, *Stanford University*

"This is a fascinating book that connects the classical theory of generalised functions (distributions) to the modern sparsity-based view on signal processing, as well as stochastic processes. Some of the early motivations given by I. Gelfand on the importance of generalised functions came from physics and, indeed, signal processing and sampling. However, this is probably the first book that successfully links the more abstract theory with modern signal processing. A great strength of the monograph is that it considers both the continuous and the discrete model. It will be of interest to mathematicians and engineers having appreciations of mathematical and stochastic views of signal processing."

Anders Hansen, *University of Cambridge*

To Gisela, Lucia, Klaus, and Murray

Preface

In the years since 2000, there has been a significant shift in paradigm in signal processing, statistics, and applied mathematics that revolves around the concept of sparsity and the search for "sparse" representations of signals. Early signs of this (r)evolution go back to the discovery of wavelets, which have now superseded classical Fourier techniques in a number of applications. The other manifestation of this trend is the emergence of data-processing schemes that minimize an ℓ_1 norm as opposed to the squared ℓ_2 norm associated with the traditional linear methods. A highly popular research topic that capitalizes on those ideas is compressed sensing. It is the quest for a statistical framework that would support this change of paradigm that led us to the writing of this book.

The cornerstone of our formulation is the classical innovation model, which is equivalent to the specification of stochastic processes as solutions of linear stochastic differential equations (SDE). The non-standard twist here is that we allow for non-Gaussian driving terms (white Lévy noise) which, as we shall see, has a dramatic effect on the type of signal being generated. A fundamental property, hinted in the title of the book, is that the non-Gaussian solutions of such SDEs admit a sparse representation in an adapted wavelet-like basis. While a sizable part of the present material is an outgrowth of our own research, it is founded on the work of Lévy (1930) and Gelfand (arguably, the second most famous Soviet mathematician after Kolmogorov), who derived general functional tools and results that are hardly known by practitioners but, as we argue in the book, are extremely relevant to the issue of sparsity. The other important source of inspiration is spline theory and the observation that splines and stochastic processes are ruled by the same differential equations. This is the reason why we opted for the innovation approach which facilitates the transposition of analytical techniques from one field to the other. While the formulation requires advanced mathematics that are carefully explained in the book, the underlying model has a strong engineering appeal since it constitutes the natural extension of the traditional filtered-white-noise interpretation of a Gaussian stationary process.

The book assumes that the reader has a good understanding of linear systems (ordinary differential equations, convolution), Hilbert spaces, generalized functions (i.e., inner products, Dirac impulses, linear operators), the Fourier transform, basic statistical signal processing, and (multivariate) statistics (probability density and characteristic functions). By contrast, there is no requirement for prior knowledge of splines, stochastic differential equations, or advanced functional analysis (function spaces, Bochner's theorem, operator theory, singular integrals) since these topics are treated in a self-contained fashion.

Several people have had a crucial role in the genesis of this book. The idea of defining sparse stochastic processes originated during the preparation of a talk for Martin Vetterli's 50th birthday (which coincided with the anniversary of the launching of Sputnik) in an attempt to build a bridge between his signals with a finite rate of innovation and splines. We thank him for his long-time friendship and for convincing us to undertake this writing project. We are grateful to our former collaborator, Thierry Blu, for his precious help in the elucidation of the functional link between splines and stochastic processes. We are extremely thankful to Arash Amini, Julien Fageot, Pedram Pad, Qiyu Sun, and John-Paul Ward for many helpful discussions and their contributions to mathematical results. We are indebted to Emrah Bostan, Ulugbek Kamilov, Hagai Kirshner, Masih Nilchian, and Cédric Vonesch for turning the theory into practice and for running the signal- and image-processing experiments described in Chapters 10 and 11. We are most grateful to Philippe Thévenaz for his intelligent editorial advice and his spotting of multiple errors and inconsistencies, while we take full responsibility for the remaining ones. We also thank Phil Meyler, Sarah Marsh and Gaja Poggiogalli from Cambridge University Press, as well as John King for his careful copy-editing.

The authors also acknowledge very helpful and stimulating discussions with Ben Adcock, Emmanuel Candès, Volkan Cevher, Robert Dalang, Mike Davies, Christine De Mol, David Donoho, Pier-Luigi Dragotti, Michael Elad, Yonina Eldar, Jalal Fadili, Mario Figueiredo, Vivek Goyal, Rémy Gribonval, Anders Hansen, Nick Kingsbury, Gitta Kutyniok, Stamatis Lefkimmiatis, Gabriel Peyré, Robert Novak, Jean-Luc Stark, and Dimitri Van De Ville, as well as a number of other researchers involved in the field.

The European Research Commission (ERC) and the Swiss National Science Foundation provided partial support throughout the writing of the book.

Notation

Abbreviations

ADMM	Alternating-direction method of multipliers
AL	Augmented Lagrangian
AR	Autoregressive
ARMA	Autoregressive moving average
AWGN	Additive white Gaussian noise
BIBO	Bounded input, bounded output
CAR	Continuous-time autoregressive
CARMA	Continuous-time autoregressive moving average
CCS	Consistent cycle spinning
DCT	Discrete cosine transform
fBm	Fractional Brownian motion
FBP	Filtered backprojection
FFT	Fast Fourier transform
FIR	Finite impulse response
FISTA	Fast iterative shrinkage/thresholding algorithm
ICA	Independent-component analysis
id	Infinitely divisible
i.i.d.	Independent identically distributed
IIR	Infinite impulse response
ISTA	Iterative shrinkage/thresholding algorithm
JPEG	Joint Photographic Experts Group
KLT	Karhunen–Loève transform
LMMSE	Linear minimum-mean-square error
LPC	Linear predictive coding
LSI	Linear shift-invariant
MAP	Maximum a posteriori
MMSE	Minimum-mean-square error
MRI	Magnetice resonance imaging
PCA	Principal-component analysis
pdf	Probability density function
PSF	Point-spread function
ROI	Region of interest

SαS	Symmetric-alpha-stable
SDE	Stochastic differential equation
SNR	Signal-to-noise ratio
WSS	Wide-sense stationary

Sets

$\mathbb{N}, \mathbb{Z}^+$	Non-negative integers, including 0
$\mathbb{Z}$	Integers
$\mathbb{R}$	Real numbers
$\mathbb{R}^+$	Non-negative real numbers
$\mathbb{C}$	Complex numbers
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{Z}^d$	d-dimensional integers

Various notation

j	Imaginary unit such that $\mathrm{j}^2 = -1$
$\lceil x \rceil$	Ceiling: smallest integer at least as large as x
$\lfloor x \rfloor$	Floor: largest integer not exceeding x
$(x_1 : x_n)$	n-tuple $(x_1, x_2, \ldots, x_n)$
$\\|f\\|$	Norm of the function f (see Section 3.1.2)
$\\|f\\|_{L_p}$	L_p-norm of the function f (in the sense of Lebesgue)
$\\|a\\|_{\ell_p}$	ℓ_p-norm of the sequence a
$\langle \varphi, s \rangle$	Scalar (or duality) product
$\langle f, g \rangle_{L_2}$	L_2 inner product
$f^\vee$	Reversed signal: $f^\vee(\boldsymbol{r}) = f(-\boldsymbol{r})$
$(f * g)(\boldsymbol{r})$	Continuous-domain convolution
$(a * b)[\boldsymbol{n}]$	Discrete-domain convolution
$\widehat{\varphi}(\boldsymbol{\omega})$	Fourier transform of φ: $\int_{\mathbb{R}^d} \varphi(\boldsymbol{r}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}, \boldsymbol{r} \rangle} \, \mathrm{d}\boldsymbol{r}$
$\widehat{f} = \mathscr{F}\{f\}$	Fourier transform of f (classical or generalized)
$f = \mathscr{F}^{-1}\{\widehat{f}\}$	Inverse Fourier transform of $\widehat{f}$
$\overline{\mathscr{F}}\{f\}(\boldsymbol{\omega}) = \mathscr{F}\{f\}(-\boldsymbol{\omega})$	Conjugate Fourier transform of f

Signals, functions, and kernels

$f, f(\cdot)$, or $f(\boldsymbol{r})$	Continuous-domain signal: function $\mathbb{R}^d \to \mathbb{R}$
φ	Generic test function in $\mathscr{S}(\mathbb{R}^d)$
$\psi_{\mathrm{L}} = \mathrm{L}^* \phi$	Operator-like wavelet with smoothing kernel ϕ
$s, \ \langle \varphi, s \rangle$	Generalized function $\mathscr{S}(\mathbb{R}^d) \to \mathbb{R}$
μ_h	Measure associated with h: $\langle \varphi, h \rangle = \int_{\mathbb{R}^d} \varphi(\boldsymbol{r}) \mu_h(\mathrm{d}\boldsymbol{r})$
δ	Dirac impulse: $\langle \varphi, \delta \rangle = \varphi(\boldsymbol{0})$
$\delta(\cdot - \boldsymbol{r}_0)$	Shifted Dirac impulse
β_{L}	Generalized B-spline associated with the operator L
φ_{int}	Spline interpolation kernel

$\beta_+^n = \beta_{\mathrm{D}^{n+1}}$	Causal polynomial B-spline of degree n
$x_+^n = \max(0, x)^n$	One-sided power function
β_α	First-order exponential B-spline with pole $\alpha \in \mathbb{C}$
$\beta_{(\alpha_1:\alpha_N)}$	Nth-order exponential B-spline: $\beta_{\alpha_1} * \cdots * \beta_{\alpha_N}$
$a, a[\cdot]$, or $a[\boldsymbol{n}]$	Discrete-domain signal: sequence $\mathbb{Z}^d \to \mathbb{R}$
$\delta[\boldsymbol{n}]$	Discrete Kronecker impulse

Spaces

$\mathscr{X}, \mathscr{Y}$	Generic vector spaces (normed or nuclear)
$L_2(\mathbb{R}^d)$	Finite-energy functions $\int_{\mathbb{R}^d} \lvert f(\boldsymbol{r})\rvert^2 \, \mathrm{d}\boldsymbol{r} < \infty$
$L_p(\mathbb{R}^d)$	Functions such that $\int_{\mathbb{R}^d} \lvert f(\boldsymbol{r})\rvert^p \, \mathrm{d}\boldsymbol{r} < \infty$
$L_{p,\alpha}(\mathbb{R}^d)$	Functions such that $\int_{\mathbb{R}^d} \left\lvert f(\boldsymbol{r})\left(1 + \lvert\boldsymbol{r}\rvert\right)^\alpha \right\rvert^p \, \mathrm{d}\boldsymbol{r} < \infty$
$\mathscr{D}(\mathbb{R}^d)$	Smooth and compactly supported test functions
$\mathscr{D}'(\mathbb{R}^d)$	Distributions or generalized functions over $\mathbb{R}^d$
$\mathscr{S}(\mathbb{R}^d)$	Smooth and rapidly decreasing test functions
$\mathscr{S}'(\mathbb{R}^d)$	Tempered distributions (generalized functions)
$\mathscr{R}(\mathbb{R}^d)$	Bounded functions with rapid decay
$\ell_2(\mathbb{Z}^d)$	Finite-energy sequences $\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lvert a[\boldsymbol{k}]\rvert^2 < \infty$
$\ell_p(\mathbb{Z}^d)$	Sequences such that $\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \lvert a[\boldsymbol{k}]\rvert^p < \infty$

Operators

Id	Identity
$\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}$	Derivative
$\mathrm{D_d}$	Finite difference (discrete derivative)
D^N	Nth-order derivative
$\partial^{\boldsymbol{n}}$	Partial derivative of order $\boldsymbol{n} = (n_1, \ldots, n_d)$
L	Whitening operator (LSI)
$\widehat{L}(\boldsymbol{\omega})$	Frequency response of L (Fourier multiplier)
ρ_{L}	Green's function of L
L^*	Adjoint of L such that $\langle\varphi_1, \mathrm{L}\varphi_2\rangle = \langle \mathrm{L}^*\varphi_1, \varphi_2\rangle$
L^{-1}	Right inverse of L such that $\mathrm{LL}^{-1} = \mathrm{Id}$
$h(\boldsymbol{r}_1, \boldsymbol{r}_2)$	Generalized impulse response of L^{-1}
L^{-1*}	Left inverse of L^* such that $(\mathrm{L}^{-1*})\mathrm{L}^* = \mathrm{Id}$
$\mathrm{L_d}$	Discrete counterpart of L
$\mathcal{N}_{\mathrm{L}}$	Null space of L
P_α	First-order differential operator: $\mathrm{D} - \alpha\mathrm{Id}, \alpha \in \mathbb{C}$
$\mathrm{P}_{(\alpha_1:\alpha_N)}$	Differential operator of order N: $\mathrm{P}_{\alpha_1} \circ \cdots \circ \mathrm{P}_{\alpha_N}$
Δ_α	First-order weighted difference
$\Delta_{(\alpha_1:\alpha_N)}$	Nth-order weighted differences: $\Delta_{\alpha_1} \circ \cdots \circ \Delta_{\alpha_N}$
∂_τ^γ	Fractional derivative of order $\gamma \in \mathbb{R}^+$ and phase τ
$(-\Delta)^{\frac{\gamma}{2}}$	Fractional Laplacian of order $\gamma \in \mathbb{R}^+$
$\mathrm{I}_p^{\gamma*}$	L_p-stable left inverse of $(-\Delta)^{\frac{\gamma}{2}}$

Probability

X, Y	Generic scalar random variables
$\mathscr{P}_X$	Probability measure on $\mathbb{R}$ of X
$p_X(x)$	Probability density function (univariate)
$\Phi_X(x)$	Potential function: $-\log p_X(x)$
$\mathrm{prox}_{\Phi_X}(x, \lambda)$	Proximal operator
$p_{\mathrm{id}}(x)$	Infinitely divisible probability law
$\mathbb{E}\{\cdot\}$	Expected value operator
m_n	nth-order moment: $\mathbb{E}\{X^n\}$
κ_n	nth-order cumulant
$\widehat{p}_X(\omega)$	Characteristic function of X: $\mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X}\}$
$f(\omega)$	Lévy exponent: $\log \widehat{p}_{\mathrm{id}}(\omega)$
$v(a)$	Lévy density
$p_{(X_1:X_N)}(\boldsymbol{x})$	Multivariate probability density function
$\widehat{p}_{(X_1:X_N)}(\boldsymbol{\omega})$	Multivariate characteristic function
$m_{\boldsymbol{n}}$	Moment with multi-index $\boldsymbol{n} = (n_1, \ldots, n_N)$
$\kappa_{\boldsymbol{n}}$	Cumulant with multi-index $\boldsymbol{n}$
$H_{(X_1:X_N)}$	Differential entropy
$I(X_1, \ldots, X_N)$	Mutual information
$D(p\\|q)$	Kullback–Leibler divergence

Generalized stochastic processes

w	Continuous-domain white noise (innovation)
$\langle \varphi, w \rangle$	Generic scalar observation of innovation process
$f_\varphi(\omega)$	Modified Lévy exponent: $\log \widehat{p}_{\langle \varphi, w \rangle}(\omega)$
$v_\varphi(a)$	Modified Lévy density
s	Generalized stochastic process: $\mathrm{L}^{-1}w$
u	Generalized increment process: $\mathrm{L_d}s = \beta_{\mathrm{L}} * w$
W	1-D Lévy process with $\mathrm{D}W = w$
B_H	Fractional Brownian motion with Hurst index H
$\widehat{\mathscr{P}}_s(\varphi)$	Characteristic functional: $\mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \varphi, s \rangle}\}$
$\mathscr{B}_s(\varphi_1, \varphi_2)$	Correlation functional: $\mathbb{E}\{\langle \varphi_1, s \rangle \langle \varphi_2, s \rangle\}$
$R_s(\boldsymbol{r}_1, \boldsymbol{r}_2)$	Autocorrelation function: $\mathbb{E}\{s(\boldsymbol{r}_1)\overline{s(\boldsymbol{r}_2)}\}$

Contents