SPRINGER BRIEFS IN STATISTICS
JSS RESEARCH SERIES IN STATISTICS

Gareth William Peters
Tomoko Matsui
Editors

Theoretical Aspects of Spatial-Temporal Modeling

Springer

SpringerBriefs in Statistics

JSS Research Series in Statistics

The current research of statistics in Japan has expanded in several directions in line with recent trends in academic activities in the area of statistics and statistical sciences over the globe. The core of these research activities in statistics in Japan has been the Japan Statistical Society (JSS). This society, the oldest and largest academic organization for statistics in Japan, was founded in 1931 by a handful of pioneer statisticians and economists and now has a history of about 80 years. Many distinguished scholars have been members, including the influential statistician Hirotugu Akaike, who was a past president of JSS, and the notable mathematician Kiyosi Itô, who was an earlier member of the Institute of Statistical Mathematics (ISM), which has been a closely related organization since the establishment of ISM. The society has two academic journals: the Journal of the Japan Statistical Society (English Series) and the Journal of the Japan Statistical Society (Japanese Series). The membership of JSS consists of researchers, teachers, and professional statisticians in many different fields including mathematics, statistics, engineering, medical sciences, government statistics, economics, business, psychology, education, and many other natural, biological, and social sciences.

The JSS Series of Statistics aims to publish recent results of current research activities in the areas of statistics and statistical sciences in Japan that otherwise would not be available in English; they are complementary to the two JSS academic journals, both English and Japanese. Because the scope of a research paper in academic journals inevitably has become narrowly focused and condensed in recent years, this series is intended to fill the gap between academic research activities and the form of a single academic paper.

The series will be of great interest to a wide audience of researchers, teachers, professional statisticians, and graduate students in many countries who are interested in statistics and statistical sciences, in statistical theory, and in various areas of statistical applications.

More information about this series at http://www.springer.com/series/13497

Gareth William Peters · Tomoko Matsui
Editors

Theoretical Aspects of Spatial-Temporal Modeling

Springer

Editors
Gareth William Peters
Department of Statistical Science
University College London
London
UK

Tomoko Matsui
The Institute of Statistical Mathematics
Tachikawa, Tokyo
Japan

ISSN 2191-544X ISSN 2191-5458 (electronic)
SpringerBriefs in Statistics
ISSN 2364-0057 ISSN 2364-0065 (electronic)
JSS Research Series in Statistics
ISBN 978-4-431-55335-9 ISBN 978-4-431-55336-6 (eBook)
DOI 10.1007/978-4-431-55336-6

Library of Congress Control Number: 2015954809

Springer Tokyo Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer Japan KK is part of Springer Science+Business Media (www.springer.com)

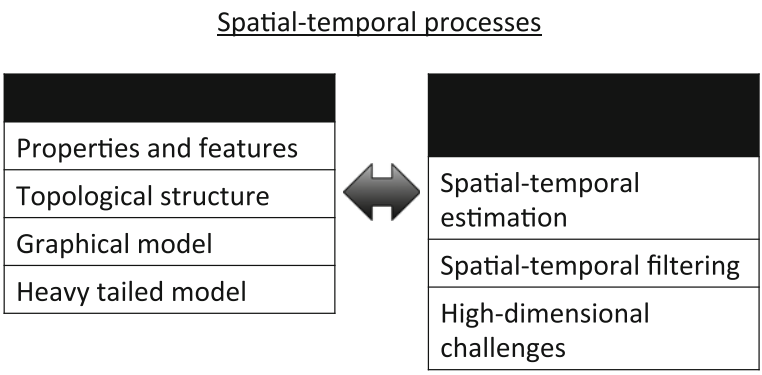
Preface

The idea to create this book arose as a response to the discussions and presentations that took place in the first and second annual international workshops on spatial and temporal modeling (STM2013 and STM2014), both of which were held in the Institute of Statistical Mathematics (ISM), Tokyo, Japan. These workshops were cohosted by Prof. Tomoko Matsui (ISM) and Dr. Gareth W. Peters (UCL). It was apparent after these workshops were completed that the wide range of participants from various backgrounds including probability, statistics, applied mathematics, physics, engineering, and signal processing as well as speech and audio processing had been recently developing a range of new theory, models, and methods for dealing with spatial and temporal problems that would be beneficial to document for a wider scientific audience.

Therefore, this book is intended to bring together a range of new innovations in the area of spatial and temporal modeling in the form of self-contained tutorial chapters on recent areas of research innovations. Since it is based on contributions from a range of world experts in spatial and temporal modeling who participated in the workshop, it reflects a cross section of specialist information on a range of important related topics. It is the aim of such a text to provide a means to motivate further research, discussion, and cross-fertilization of research ideas and directions among the different research fields representative of the authors who contributed.

While this book covers more of the theoretical aspects of spatial–temporal modeling, its companion book, also in the Springer Briefs series, titled *Modern Methodology and Applications in Spatial-Temporal Modeling*, complements this book for practitioners as it covers a range of new innovations in methodology for modeling and applications. This book aims to provide a modern introductory tutorial on specialized theoretical aspects of spatial and temporal modeling. The areas covered involve a range of topics which reflect the diversity of this domain of research across a number of quantitative disciplines. For instance, Chap. 1 provides modern coverage of particle association measures that underpin the theoretical properties of recently developed random set methods in space and time otherwise known as the class of probability hypothesis density framework (Ph.D. filters).

Chapter 2 deals with an overview of recent advances in Monte Carlo methods for Bayesian filtering in high-dimensional spaces. In particular it explains how one may extend classical sequential Monte Carlo methods for filtering and static inference problems to high dimensions and big-data applications. Chapter 3 deals with an overview of generalized families of processes that extend the class of Gaussian process models to heavy-tailed families known as alpha-stable processes. In particular it covers aspects of characterization via the spectral measure of heavy-tailed distributions, and it then provides an overview of their applications in wireless communications channel modeling. Chapter 4 concludes with an overview of analysis for probabilistic spatial percolation methods as would be relevant in the modeling of graphical networks and connectivity applications in sensor networks which also incorporate stochastic geometry features.



We first note that each chapter of this book is intended to be a self-contained research-level tutorial on modern approaches to the theoretical study of some aspect of spatial and temporal statistical modeling. However, to guide the reader in considering the sections of this book we note the following relationships between chapters. Chapters 1 and 2 cover recent advances in spatial tracking and state space modeling settings in high-dimensional contexts. The first arises in multiple target tracking settings and is based on extensions of sequential Monte Carlo methods for such contexts which have become known as probability hypothesis density filters. Chapter 2 deals with the class of high-dimensional state space models and introduces different approaches one can adopt to tackle the curse of dimensionality that the standard SMC method suffers from when the state space is high dimensional. In particular it introduces ideas of blocked particle filters, discusses recent space–time particle filters and studies, and compares these to the recently developed class of methods known as sequential Markov chain Monte Carlo (SMCMC) methods.

Chapters 3 and 4 are not so much focused on the estimation of latent process models in spatial–temporal settings, but instead focus on the study of phenomena that have been developed recently to characterize extremes in spatial–temporal settings. In this regard the fourth chapter discusses new approaches to the characterization of heavy-tailed stochastic processes, focusing specifically on the

α-stable family. The final chapter constructs characterizations of spatial processes from a geometrical perspective, focusing on spatial network structures, random graphs and the study of certain connectivity phenomena for such graphical structures. The chapter introduces ideas that can be used to characterize and understand random graphical models that are growing in popularity in tracking multiple objects and populations, finance, ecology, and social network analysis.

Tokyo, Japan
August 2015

Gareth William Peters
Tomoko Matsui

Acknowledgments

We are very grateful to Research Organization of Information and Systems (ROIS), The Institute of Statistical Mathematics (ISM), UK Royal Society International Exchange Grant, and Ministry of Education, Culture, Sports, Science and Technology (MEXT) undertake project "Cooperation with Math Program" to support the first and second annual international workshops on spatial and temporal modeling (STM2013 and STM2014) and the first workshop on complex systems modeling and estimation challenges in big data (CSM2014). The idea to create this book arose as a response to the discussions and presentations that took place in the workshops.

We would like to express our sincere thanks to all the following presenters in the workshops.

- Prof. Nourddine Azzaoui, Université Blaise Pascal
- Prof. Jen-Tzung Chien, National Chiao Tung University
- Prof. Arnaud Doucet, Oxford University
- Prof. Norikazu Ikoma, KIT
- Prof. Kenji Fukumizu, ISM
- Prof. Konstatin Markov, Aizu University
- Prof. Daichi Mochihashi, ISM
- Prof. Pierre Del Moral, UNSW
- Prof. Tor Andre Myrvoll, SINTEF
- Dr. Ido Nevat, Institute for Infocomm Research, A-Star
- Prof. Yosihiko Ogata, ERI, University of Tokyo and ISM, ROIS
- Dr. Takashi Owada, Technion
- Prof. Daniel P. Palomar, HKUST
- Prof. François Septier, Telecom lille 1
- Prof. Taiji Suzuki, Tokyo Institute of Technology
- Prof. Kazuya Takeda, Nagoya University
- Prof. Mario Wüthrich, ETH Zurich

Contents

Editors and Contributors

About the Editors

Dr. Gareth William Peters Department of Statistical Science, University College London, UK.

He is Assistant Professor in the Department of Statistical Science, Principle Investigator in Computational Statistics and Machine Learning, and Academic Member of the UK Ph.D. Center of Financial Computing at University College London. He is also Adjunct Scientist in the Commonwealth Scientific and Industrial Research Organization, Australia; Associate Member Oxford-Man Institute at the Oxford University; and Associate Member in the Systemic Risk Center at the London School of Economics. Dr. Peters is also a Visiting Professor at the Institute of Statistical Mathematics, Tokyo, Japan, where he has been visiting since 2009. Dr. Peters obtained a B.Sc. (hons 1st) in Mathematics and Physics and a B.Eng. (hons 1st) from the University of Melbourne in 2003, a M.Sc. (research) from the University of Cambridge in 2006, and a Ph.D. in Statistics (by publication) from the University of New South Wales in 2009.

His research interests range over several areas of mathematical statistics, time series and state space modeling, heavy-tailed stochastic processes and Levy processes, dependence structure modeling and a range of applications in insurance, risk management, econometrics, finance, signal processing, ecology, and medical statistics. Dr. Peters has lectured at the Department of Mathematics and Statistics in the University of New South Wales, Sydney, Australia (2008−2012) and where he still holds a an Associated Lecturer position. He has worked as a Lead Quantitative Analyst in the Commonwealth Bank of Australia for 3 years and has more than 5 years industry experience in financial trading in the hedge fund sector and asset management.

Prof. Tomoko Matsui The Institute of Statistical Mathematics, Tokyo, Japan.

She received the Ph.D. degree from the Computer Science Department, Tokyo Institute of Technology, Tokyo, Japan, in 1997. From 1988 to 2002, she was with NTT, where she worked on speaker and speech recognition. From 1998 to 2002,

she was with the Spoken Language Translation Research Laboratory, ATR, Kyoto, Japan, as a Senior Researcher and worked on speech recognition. From January to June 2001, she was an Invited Researcher in the Acoustic and Speech Research Department, Bell Laboratories, Murray Hill, NJ, working on finding effective confidence measures for verifying speech recognition results. She is currently a Professor in the Institute of Statistical Mathematics, Tokyo, working on statistical spatial-temporal modeling for various applications including speech and image recognition. Prof. Matsui received the paper award of the Institute of Electronics, Information, and Communication Engineers of Japan (IEICE) in 1993.

Contributors

Prof. Nourddine Azzaoui Laboratoire de Mathematiques, Université Blaise Pascal, France.

He is a member of the team Probability Analysis and Statistics (PAS) which is a component of the Mathematics Laboratory UMR-CNRS 6620. From 2008 to 2010, he was a Postdoc Researcher in the Charles Delaunay Institute (ICD) in the technology University of Troyes. Before joining the ICD, from 2007 to 2008 he was an Assistant Professor and Member of the statistics and probability team of the Paul Painlevé Laboratory in Lille. From 2006 to 2007 he was an Assistant Professor at the Burgundy Mathematics Institute (IMB) in Dijon, where he defended his Ph.D. thesis in applied mathematics and statistics in 2006.

His research interests include spectral analysis and statistics of infinite variance and nonstationary processes. Simulation techniques, signal processing and modelization of nonlinear and nonstationary phenomena. Channel and interference modeling in the context of ultra wide band and millimeter waves. Nonparametric and semiparametric techniques applied to detection and statistical learning.

Prof. Laurent Clavier Institut Mines-Télécom/Télécom Lille/IRCICA FR CNRS 3024, France.

He conducts his research in IEMN (UMR CNRS 8520—Institut d'Electronique de Microelectronique et de Nanotechnologie) and in IRCICA (USR CNRS 3380—Institut de Recherche sur les composants logiciels et materiels pour l'Information et la Communication Avancee). He is responsible for the research axis "Telecommunication Circuits and Systems" in IEMN.

His research activities concern digital communications and the physical layer of wireless networks, more specifically energy autonomous sensor networks. He is particularly interested in the UWB transmissions up-converted in the 60 GHz band, the radio channel and interference statistical modeling, and in designing robust receivers in non-Gaussian varying environments.

Prof. Pierre Del Moral School of Mathematics and Statistics, University of New South Wales, Australia.

He is a world-class researcher in the field of particle methods, filtering, sequential Monte Carlo, and mean field approximations. He is one of the main

developers of the so-called particle methods that are of increasing importance in simulations of a multidimensional complex system. Pierre's research is on both theoretical and abstract mathematics but he also blends his results with applications in many important research areas: Propagations of chaos, Central limit theorems, Large deviation, Filtering and multiple object tracking, Bayesian Inference, Financial Mathematics, Biology, and Chemistry. He has authored 7 books (4 research monograph) and over 100 research publications (many of them in the top journals). He has also involved in an impressive number of academic and industrial research projects.

Philippe Deprez Ph.D ETH Zurich, RiskLab, Department of Mathematics, Zurich

Prof. Arnaud Guillin Laboratoire de Mathematiques, Université Blaise Pascal, France.

He has been a Professor of Mathematics Laboratory in University Blaise Pascal since September 2008. Before, he was Maitre de conferences at Paris Dauphine University (2001–2005) and Professor at Ecole Centrale Marseille (2006–2008). He was a Junior Member of the Institut "Universitaire de France" from 2010 to 2015. Specialized in probability, he has published more than 50 papers in top probability and statistics journals.

Jeremie Houssineau Ph.D School of Engineering and Physical Sciences, Heriot-Watt University, UK

Prof. François Septier Institut Mines-Télécom/Télécom Lille/CRIStAL UMR 9189, Villeneuve d'Ascq, France.

He received the engineer degree in electrical engineering and signal processing in 2004 from Telecom Lille, France, the M.Sc. degree in digital communications, and a Ph.D. in electrical engineering both from the University of Valenciennes, France, in 2004 and 2008, respectively. From March 2008 to August 2009, he was a Research Associate in the Signal Processing Laboratory, Cambridge University, UK. From September 2009, he is an Associate Professor with the Institut Mines-Telecom/ Telecom Lille/CRIStAL UMR 9189 CNRS, France. His research focuses on Monte Carlo statistical methods for Bayesian inference in complex and/or high-dimensional systems such as multitarget tracking, sensor networks, and source term estimation.

Prof. Mario V. Würthrich Swiss Finance Institute SFI Professor, ETH Zurich, RiskLab, Department of Mathematics, Zurich.

He is Professor in the Department of Mathematics at ETH Zurich, Adjunct Professor at University of Bologna, Honorary Visiting Professor at City University London, Honorary Professor at University College London, and Swiss Finance Institute Professor. He holds a Ph.D. in mathematics from ETH Zurich. From 2000 to 2005, he held an actuarial position at Winterthur Insurance and was responsible for claims reserving in non-life insurance, as well as for developing and implementing the Swiss Solvency Test. He is a fully qualified Actuary SAA, serves on the board of the Swiss Association of Actuaries, and is Editor of ASTIN Bulletin and the European Actuarial Journal.

Chapter 1
Particle Association Measures and Multiple Target Tracking

Pierre Del Moral and Jeremie Houssineau

Abstract In the last decade, the area of multiple target tracking has witnessed the introduction of important concepts and methods, aiming at establishing principled approaches for dealing with the estimation of multiple objects in an efficient way. One of the most successful classes of multi-object filters that have been derived out of these new grounds includes all the variants of the Probability Hypothesis Density (PHD) filter. In spite of the attention that these methods have attracted, their theoretical performances are still not fully understood. In this chapter, we first focus on the different ways of establishing the equations of the PHD filter, using a consistent set of notations. The objective is then to introduce the idea of observation path, upon which association measures are defined. We will see how these concepts highlight the structure of the first moment of the multi-object distributions in time, and how they allow for devising solutions to practical estimation problems.

1.1 Introduction

Multiple target tracking refers to the estimation of the state of an unknown and time-varying number of objects, given a sequence of noisy, incomplete and corrupted collections of observations. These collections of observations are incomplete since some objects might not be consistently detected, and may be corrupted by spurious observations that are not generated by objects in the system of interest. The variation in the number of objects is due to the random appearance/disappearance of objects from the surveillance zone, and the processes associated to these aspects of the population dynamics are referred to as birth and death.

P. Del Moral (✉)
School of Mathematics and Statistics, University of New South Wales, Sydney, Australia
e-mail: p.del-moral@unsw.edu.au

J. Houssineau
School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK
e-mail: j.houssineau@hw.ac.uk

G.W. Peters and T. Matsui (eds.), *Theoretical Aspects of Spatial-Temporal Modeling*,
JSS Research Series in Statistics, DOI 10.1007/978-4-431-55336-6_1

The first solutions [1, 7] to multiple target tracking were built from single-object filters such as the Kalman filter, and were handling birth, death and data association, i.e. the search for the right correspondence between objects and observations, in a heuristic way. These methods have been and still are widely used because they are intuitively appealing and can be easily modified to suit different practical requirements. However, they also show some limitations in terms of responsiveness and robustness in complicated situations.

In order to overcome the limitations of traditional multiple target tracking algorithms, more principled ways of modelling multi-object systems were introduced [8, 11, 14], from which several multi-object estimation algorithms were derived, including the Probability Hypothesis Density (PHD) filter [12] and the cardinalized PHD filter [13, 23]. The algorithms derived from this approach show an improved responsiveness to changes in the probability of detection or in the number of spurious observations, and as far as the PHD filter is concerned, have a low computational cost compared to the usual methods. Yet, because the focus is on the system of objects rather than on the objects themselves, the PHD filter does not naturally provide tracks, i.e. specific information about identified objects, and post-processing algorithms are required to obtain track estimates [17]. Even though the PHD filter was originally designed to track a possibly high number of indistinguishable objects, it is most often used in contexts where the final objective is to obtain individual tracks. The limitation induced by this absence of track estimates is not severe in the Gaussian mixture implementation of the PHD filter [21] since the terms in the mixture can often be interpreted as potential tracks. However, this limitation leads to a noticeable degradation of performance in the sequential Monte Carlo (SMC) implementation [22]. Indeed, the so-called SMC-PHD filter is based on the update and prediction of a single set of particles representing the whole multi-object system and thus requires the use of clustering algorithms to compute track estimates, inducing additional computational costs and inaccuracies.

In this work, the PHD filter is expressed in different ways but with a consistent set of notations drawn from probability theory. The purpose of such an approach is to present different viewpoints including: (*a*) an update prediction scheme that is standard in filtering, (*b*) a one-step filtering scheme that leads to a measure-valued-process formulation, and (*c*) an expression based on the concept of *association measure*, differing significantly from the previous ones, which is based on [16] and [6] and which has been used in a similar context in [9]. These different approaches are detailed in Sect. 1.3, after the introduction of a point-process-based modelling of multi-object systems in Sect. 1.2. We show how the association measure allows for some operations to be performed at the object level, including the extraction of track estimates in a way that does not depend on the considered implementation. Some of the results presented in [6] and dealing with the analysis of the stability of the PHD filter equations are recalled in Sect. 1.4. Finally, two different proofs of the PHD filter are detailed in Sect. 1.5, including the original generating functional-based derivation proposed in [12] and a 'direct proof' subsequently presented in [2]. These two

different approaches were based on different frameworks in their respective publications, and the motivation in this work is to present them in a unified way.

The formalism of probability theory is used in this article to facilitate the statement of the results, even though this choice is not the most usual in the area of multi-target tracking. Henceforth, $\mathbf{M}(\mathbf{E})$ (resp. $\mathbf{P}(\mathbf{E})$) will stand for the set of finite positive measures (resp. probability measures) on a given measurable space $(\mathbf{E}, \mathscr{E})$. Additionally, the Banach space of all bounded and measurable functions equipped with the uniform norm $\|\cdot\|$ will be denoted as $\mathbf{B}(\mathbf{E})$, and we write $\gamma(f) = \int f(x)\gamma(\mathrm{d}x)$ for any measure $\gamma \in \mathbf{M}(\mathbf{E})$ and any measurable function $f \in \mathbf{B}(\mathbf{E})$. In particular, the measurable functions equal everywhere to zero and one are, respectively, denoted $\mathbf{0}$ and $\mathbf{1}$ and the indicator function of a given measurable set $B \in \mathscr{E}$ is denoted $\mathbf{1}_B$.

A bounded positive integral operator Q from a measurable space $\mathbf{E}$ into a measurable space $\mathbf{E}'$ is an operator $f \mapsto Q(f)$ from $\mathbf{B}(\mathbf{E}')$ to $\mathbf{B}(\mathbf{E})$ such that the functions

$$x \mapsto Q(f)(x) = \int_{\mathbf{E}'} Q(x, \mathrm{d}y) f(y)$$

are bounded and measurable for some measure $Q(x, \cdot) \in \mathbf{M}(\mathbf{E}')$. If it holds that $Q(\mathbf{1})(x) = 1$ for any $x \in \mathbf{E}$, then Q is referred to as a Markov kernel or Markov transition from $\mathbf{E}$ to $\mathbf{E}'$.

Let $G : \mathbf{E} \to (0, \infty)$ be a bounded positive potential function. The following change of probability measure is referred to as Boltzmann–Gibbs transformation:

$$\begin{aligned} \Psi_G : \mathbf{M}(\mathbf{E}) &\to \mathbf{P}(\mathbf{E}) \\ \gamma &\mapsto \Psi_G(\gamma) \end{aligned} \tag{1.1}$$

where assuming $\gamma(G) > 0$,

$$\Psi_G(\gamma)(\mathrm{d}x) = \frac{1}{\gamma(G)} G(x)\gamma(\mathrm{d}x).$$

Additionally, if $\mathbf{E}$ is a topological space, then we can consider its Borel σ-algebra, denoted as $\mathscr{B}(\mathbf{E})$. The set of integer-valued measures on $\mathbf{E}$ is denoted as $\mathbf{N}(\mathbf{E})$ and is a subset of $\mathbf{M}(\mathbf{E})$. In particular, if $\mathbf{E}$ is a Polish space, then a point process on $\mathbf{E}$ is a random variable in the set $\mathbf{N}(\mathbf{E})$ equipped with an appropriate σ-algebra [4].

1.2 Modelling Multi-object Systems

The objective is to provide a statistical description of a multi-object system partially observed through time. Without loss of generality, the set $\mathbb{T}$ of time steps is assumed to be equal to the set $\mathbb{N}$ of non-negative integers. The objects of interest are described,

at time $t \in \mathbb{T}$, by their state in the state space $\mathbf{X}_t$ which is assumed to be Polish and which is equipped with its Borel σ-algebra $\mathscr{B}(\mathbf{X}_t)$. The multi-object system is then described in the state space $\mathbf{X}_t$ by a (finite) point process $\mathscr{X}_t$ on $\mathbf{X}_t$ that can be represented as

$$\mathscr{X}_t = \sum_{i=1}^{N_t} \delta_{X_i},$$

where N_t is a $\mathbb{N}$-valued random variable and $X_1, \ldots, X_{N_t}$ is a collection of $\mathbf{X}_t$-valued random variables. In order to model that some phenomena might affect the evolution and the observation of the system without being described in $\mathbf{X}_t$, we augment the state space $\mathbf{X}_t$ by two *empty* states ψ_b and ψ_c and define the corresponding extended state space $\bar{\mathbf{X}}_t = \mathbf{X}_t \cup \{\psi_\mathrm{b}, \psi_\mathrm{c}\}$. The empty state ψ_b is used to model that some objects that are expected to appear at time $t+1$ do not have a state in $\mathbf{X}_t$. The state ψ_c is used to model the *clutter generators*, i.e. the objects/phenomena that are not part of the multi-object system of interest and that will generate a spurious observation at time t. In order to integrate these additional variables in the modelling of the system, an extended version of the point process $\mathscr{X}_t$ can be defined on $\bar{\mathbf{X}}_t$ as

$$\bar{\mathscr{X}}_t(\mathrm{d}x) = \mathbf{1}_{\mathbf{X}_t}(x)\mathscr{X}_t(\mathrm{d}x) + N_{\mathrm{b},t}\delta_{\psi_\mathrm{b}}(\mathrm{d}x) + N_{\mathrm{c},t}\delta_{\psi_\mathrm{c}}(\mathrm{d}x),$$

where $N_{\mathrm{b},t}$ is a $\mathbb{N}$-valued random variable with expected value $n_{\mathrm{b},t} \doteq \mathbb{E}[N_{\mathrm{b},t}]$ which describes the number of objects that will appear at time $t+1$, and $N_{\mathrm{c},t}$ is the random number of clutter generators at time t.

Apart from the birth and the clutter generators that are not modelled in this way in [12], a slight difference between the point-process-based modelling introduced here and the one originally proposed for the PHD filter is that the point process $\mathscr{X}_t$ is not assumed to be simple, i.e. with a maximum of one point per state in $\mathbf{X}_t$. Indeed, the framework used in [12] is based on the concept of random finite set which requires all the objects to have a different state. We will see in Sect. 1.5 that relaxing this assumption does not affect the derivation of the filter. Yet, notice that the way of including the birth and the clutter generators that is considered here could not be used with random finite sets because of the assumption of simplicity.

1.2.1 Observation

At any time $t \in \mathbb{T}$, the extended point process $\bar{\mathscr{X}}_t$ is partially observed and a collection of observations $z_1, \ldots, z_{n'_t}$ in the observation space $\mathbf{Z}_t$ is made available. The space $\mathbf{Z}_t$ is also assumed to be Polish. In order to model the stochasticity underlying the generation of this collection of observations, we introduce another point process $\mathscr{Z}_t$ on $\mathbf{Z}_t$, assumed to be simple, of the form

$$\mathscr{Z}_t = \sum_{i=1}^{N_t'} \delta_{Z_i},$$

where N_t' is a $\mathbb{N}$-valued random variable giving the number of points in the point process, and $Z_1, \ldots, Z_{N_t'}$ is a collection of $\mathbf{Z}_t$-valued random variables corresponding to random observations in the observation space $\mathbf{Z}_t$. Some of the observations do not correspond to objects of interest and are consequences of background or sensor noise. These spurious observations are often referred to as *clutter points* in the target tracking literature. Also, some objects might not be detected and do not have any corresponding observation in $\mathbf{Z}_t$, for this reason, we introduce the empty observation ϕ and we consider the extended observation space $\bar{\mathbf{Z}}_t = \mathbf{Z}_t \cup \{\phi\}$. We also introduce an extended version of the observation point process $\mathscr{Z}_t$ defined as $\bar{\mathscr{Z}}_t = \mathscr{Z}_t + N_\phi \delta_\phi$, where N_ϕ is a $\mathbb{N}$-valued random variable representing the number of undetected objects. A Markov kernel L_t from $\bar{\mathbf{X}}_t$ to $\bar{\mathbf{Z}}_t$, called the likelihood, can then be introduced in order to model the observation in $\bar{\mathbf{Z}}_t$ of objects with states in $\mathbf{X}_t$, as well as the distribution of the spurious observations. The likelihood L_t is defined as follows: for any $x \in \mathbf{X}_t$ and any $B \in \mathscr{B}(\mathbf{Z}_t)$,

(*a*) $L_t(x, B)$ gives the probability for an object that has state x to be detected in the Borel set B,
(*b*) $L_t(x, \{\phi\})$ gives the probability for the detection of an object that has state x to fail,
(*c*) $L_t(\psi_\mathrm{c}, B)$ is the probability for a spurious observation that has been triggered to be in B, and
(*d*) $L_t(\psi_\mathrm{b}, \{\phi\}) = 1$.

As a consequence of (*c*), it holds that $L_t\big(\psi_\mathrm{c}, \mathbf{Z}_t\big) = 1$ since a spurious observation that has been triggered must be in $\mathbf{Z}_t$, and we find that $L_t\big(\psi_\mathrm{c}, \{\phi\}\big) = 0$. We assume that there exists a reference measure λ on $\mathbf{Z}_t$ such that $L_t(x, \cdot)$ is absolutely continuous with respect to (w.r.t.) λ for any $x \in \bar{\mathbf{X}}_t$ and we define the measurable function $\ell_t : \bar{\mathbf{X}}_t \times \bar{\mathbf{Z}}_t \to [0, \infty)$ as the Radon–Nikodym derivative of L_t on $\bar{\mathbf{Z}}_t$, i.e.

$$\ell_t(x, z) = \frac{\mathrm{d}L_t(x, \cdot)}{\mathrm{d}(\lambda + \delta_\phi)}(z), \quad \forall (x, z) \in \bar{\mathbf{X}}_t \times \bar{\mathbf{Z}}_t.$$

The structure and the interpretation of L_t are given in Fig. 1.1, and an alternative formulation of this Markov kernel is detailed in the following remark.

Remark 1 It is possible to consider the point ψ_c separately by introducing a probability measure $\nu_t \in \mathbf{P}(\mathbf{Z}_t)$ such that $\nu_t(A) = L_t(\psi_\mathrm{c}, A)$ for any $A \in \mathscr{B}(\mathbf{Z}_t)$. The measure ν_t is interpreted as the spatial distribution of the spurious observations. Also, since $L_t(x, \cdot)$ is a probability measure on $\bar{\mathbf{Z}}_t$ for any $x \in \mathbf{X}_t$, it holds that

$$L_t(x, \bar{\mathbf{Z}}_t) = L_t(x, \{\phi\}) + L_t(x, \mathbf{Z}_t) = 1, \quad \forall x \in \mathbf{X}_t,$$

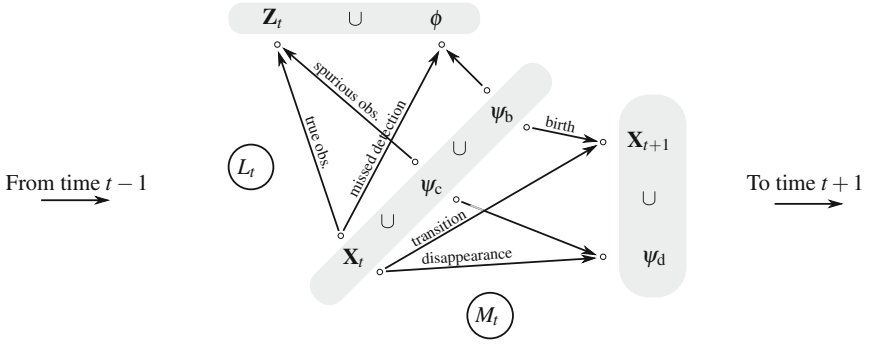


Fig. 1.1 Representation of the possible transitions in the Markov kernels L_t and M_t

so that, denoting $d_t(x)$ the probability $L_t(x, \mathbf{Z}_t)$ for an object at point $x \in \mathbf{X}_t$ to generate a non-empty observation, we find that $L_t(x, \{\phi\}) = 1 - d_t(x)$ and that L_t can be factorised on $\mathbf{Z}_t$ as

$$L_t(x, A) = d_t(x) L'_t(x, A), \quad \forall A \in \mathscr{B}(\mathbf{Z}_t),$$

where $L'_t(x, \cdot)$ is a Markov kernel from $\mathbf{X}_t$ to $\mathbf{Z}_t$ that is uniquely defined whenever $d_t(x) \neq 0$. We also introduce ν_t and $\ell'_t(x, \cdot)$ as the Radon–Nikodym derivatives of ν_t and $L'_t(x, \cdot)$ w.r.t. a reference measure $\lambda \in \mathbf{M}(\mathbf{Z}_t)$.

Example 1 The Markov kernel L'_t introduced in Remark 1 can be interpreted as a state measurement model for an observation equation of the form

$$z_t = h_t(x_t, w'_t),$$

where $x_t \in \mathbf{X}_t$ is the state of a given object in the system of interest and $z_t \in \mathbf{Z}_t$ is the observation it generates under the observation function h_t and the observation noise w'_t. If $\mathbf{Z}_t$ is a subset of $\mathbb{R}^{d'}$ for a given $d' > 0$, if the observation function takes the special form

$$h_t(x_t, w'_t) = h'_t(x_t) + w'_t,$$

and if w'_t is Gaussian with zero mean and with covariance R_t, then ℓ'_t can be defined as the Radon–Nikodym derivative of L'_t w.r.t. the Lebesgue measure on $\mathbb{R}^{d'}$, and can be expressed as

$$\ell'_t(x, z) = \frac{1}{\sqrt{(2\pi)^{d'} |R_t|}} \exp\Big(-\frac{1}{2} \big(z - h'_t(x)\big)^T R_t^{-1} \big(z - h'_t(x)\big) \Big).$$

1.2.2 Motion

When the system evolves from time t to time $t+1$, some objects disappear from the scene. In order to model this behaviour, we consider the cemetery state ψ_{d}, with 'd' for disappearance, on which these objects are represented at time $t+1$. We describe how the system evolves in time by introducing, for any time $t \in \mathbb{T}$, a Markov kernel M_t from $\bar{\mathbf{X}}_t$ to $\mathbf{X}_{t+1} \cup \{\psi_{\mathrm{d}}\}$ which characterises the motion at the object level. Because of the composite nature of these source and target spaces, the kernel M_t can be parted and defined as follows: for any $x \in \mathbf{X}_t$ and any Borel set $B \in \mathscr{B}(\mathbf{X}_{t+1})$,

(*a*) $M_t(x, B)$ gives the probability for an object that had state x at time t to persist to time $t+1$ and to be in B,
(*b*) $M_t(x, \{\psi_{\mathrm{d}}\})$ gives the probability for an object that had state x at time t to disappear from the scene,
(*c*) $M_t(\psi_{\mathrm{b}}, B)$ is the probability for an object that appeared at time $t+1$ to be in B, and
(*d*) $M_t(\psi_{\mathrm{c}}, \{\psi_{\mathrm{d}}\}) = 1$.

As a consequence of (*c*), $M_t(\psi_{\mathrm{b}}, \mathbf{X}_{t+1})$ is interpreted as the probability for an object that appeared at time $t+1$ to be in $\mathbf{X}_{t+1}$, which is equal to 1 by definition, so that the Markov kernel M_t must verify $M_t(\psi_{\mathrm{b}}, \{\psi_{\mathrm{d}}\}) = 0$. Note that defining the distribution of appearing objects as a kernel can be useful whenever this distribution varies according to the duration of the time steps. The reason behind the constraint (*d*) on M_t is that clutter generators are not assumed to have a predictable behaviour in time so that the ones that were considered at time t are assumed to all disappear and will be replaced by new ones at time $t+1$. The structure of M_t is detailed in Fig. 1.1, and an alternative formulation of this Markov kernel is proposed in the following remark.

Remark 2 First, the component related to the birth of new objects can be singled out by setting $\eta_{\mathrm{b},t}(B) = M_t(\psi_{\mathrm{b}}, B)$ for any $B \in \mathscr{B}(\mathbf{X}_{t+1})$. The Markov kernel $M_t(x, \cdot)$ can then be factorised for any $x \in \mathbf{X}_t$ as

$$M_t(x, B) = s_t(x) M'_t(x, B), \quad \forall B \in \mathscr{B}(\mathbf{X}_{t+1}),$$

where $s_t(x) = M_t(x, \mathbf{X}_{t+1})$ is the probability of survival for an object with state $x \in \mathbf{X}_t$ and where $M'_t(x, \cdot)$ is a Markov kernel from $\mathbf{X}_t$ to $\mathbf{X}_{t+1}$, defined uniquely when $s_t(x) \neq 0$.

Example 2 As in Example 1 with the observation kernel L'_t, the Markov kernel M'_t can be interpreted as a state transition model corresponding to the state equation

$$x_{t+1} = f_t(x_t, w_t),$$

where $x_{t+1} \in \mathbf{X}_{t+1}$ and $x_t \in \mathbf{X}_t$ are the states of a given object at times $t+1$ and t and where f_t is the state transition function which additionally depends on the state noise w_t. If we assume that the state transition function f_t takes the form

$$f_t(x_t) = F_t x_t + w_t,$$

and that w_t is Gaussian with zero mean and with covariance Q_t, and if the prior uncertainty on x_t is also Gaussian with mean $x_{t|t}$ and covariance $P_{t|t}$, then the state can be predicted with the Kalman filter prediction as

$$\begin{aligned} x_{t+1|t} &= F_t x_{t|t}, \\ P_{t|t} &= F_t P_{t|t} F_t^T + Q_t. \end{aligned}$$

Equipped with suitable ways of describing the motion and the observation of multi-object systems, we tackle the question of the estimation of these systems in the next section.

1.3 The Probability Hypothesis Density Filter

In this section, the objective is to show how the modelling introduced in Sect. 1.2 can lead, under assumptions, to a simple estimation algorithm for multi-object systems, or *multi-target tracker*, called the Probability Hypothesis Density filter. It is usual in the target tracking literature to express estimation algorithms via two equations, one for propagating the state of the system to the next time step, usually called *prediction*, and one for updating our representation of the system with a realisation of the observation process, often referred to as *update*. After considering these *two-step filtering equations*, we will see that other formulations can be convenient for establishing connections with different existing approaches.

1.3.1 Two-Step Filtering Equations

The motivation behind the Probability Hypothesis Density filter, or PHD filter, is to bypass the complexity of the posterior distribution of $\mathscr{X}_t | \mathscr{Z}_{0:t}$ by only computing its first-moment measure. For the first-moment measure of $\mathscr{X}_t | \mathscr{Z}_{0:t}$ to be expressed as a function of the first-moment measure of $\mathscr{X}_t | \mathscr{Z}_{0:t-1}$ only, we consider the following assumptions:

- the observation occurs independently for each object and no more than one observation is generated for each,
- $\mathscr{X}_t | \mathscr{Z}_{0:t-1}$ is a Poisson point process, and
- $N_{\mathrm{c},t}$ is a Poisson random variable with parameter $n_{\mathrm{c},t}$ and is independent of the point process $\mathscr{X}_t$.

In order to integrate spurious observations and the objects to be born between times t and $t+1$ in this statistical picture, an extended version of the first-moment measure $\gamma_{t|t-1}$ of $\mathscr{X}_t|\mathscr{Z}_{0:t-1}$ on $\bar{\mathbf{X}}_t$, denoted $\bar{\gamma}_{t|t-1}$, is defined by additionally setting $\bar{\gamma}_{t|t-1}(\{\psi_\mathrm{c}\}) = n_{\mathrm{c},t}$ and $\bar{\gamma}_{t|t-1}(\{\psi_\mathrm{b}\}) = n_{\mathrm{b},t}$. The action of augmenting a given measure $\gamma \in \mathbf{M}(\mathbf{X}_t)$ by two additional atoms at ψ_b and ψ_c with respective mass $n_{\mathrm{b},t}$ and $n_{\mathrm{c},t}$, is encoded into the change of measure $\Xi_t : \mathbf{M}(\mathbf{X}_t) \to \mathbf{M}(\bar{\mathbf{X}}_t)$.

Theorem 1 *The first-moment measure $\bar{\gamma}_t$ of the updated point process $\bar{\mathscr{X}}_t|\mathscr{Z}_{0:t}$ on $\bar{\mathbf{X}}_t$ can be expressed as a function of $\bar{\gamma}_{t|t-1}$ as follows:*

$$\bar{\gamma}_t(B) = \int_B g_{t,\bar{\gamma}_{t|t-1}}(x)\bar{\gamma}_{t|t-1}(\mathrm{d}x), \quad \forall B \in \mathscr{B}(\bar{\mathbf{X}}_t), \tag{1.2}$$

where $g_{t,\gamma}$ is a bounded and measurable potential function *on $\bar{\mathbf{X}}_t$, defined for any measure $\gamma \in \mathbf{M}(\bar{\mathbf{X}}_t)$ such that $\gamma(\ell_t(\cdot, Z_i)) > 0$ for every $1 \le i \le N'_t$ as*

$$g_{t,\gamma}(x) = \ell_t(x,\phi) + \int \mathscr{Z}_t(\mathrm{d}z)\frac{\ell_t(x,z)}{\gamma\big(\ell_t(\cdot,z)\big)}, \quad \forall x \in \bar{\mathbf{X}}_t. \tag{1.3}$$

In order to highlight the structure of the update Eq. (1.2), Theorem 1 will be proved in two different ways in Sect. 1.5.

Remark 3 Following the notations of Remark 1, we can rewrite the potential function $g_{t,\gamma}$ defined in (1.3) as follows:

$$g_{t,\gamma}(x) = 1 - d_t(x) + \sum_{i=1}^{N'_t} \frac{d_t(x)\ell'_t(x,Z_i)}{n_\mathrm{c}v_t(Z_i) + \gamma\big(d_t\ell'_t(\cdot,Z_i)\big)}, \quad \forall x \in \mathbf{X}_t,$$

for any given $\gamma \in \mathbf{M}(\mathbf{X}_t)$ such that $n_\mathrm{c}v_t(Z_i) + \gamma(d_t\ell'_t(\cdot,Z_i)) > 0$ for any $1 \le i \le N'_t$, which is closer to the standard formulation of the PHD filter than (1.3). However, (1.2) also holds for $x = \psi_\mathrm{c}$, in which case we find that

$$\bar{\gamma}_t(\{\psi_\mathrm{c}\}) = \sum_{i=1}^{N'_t} \frac{n_\mathrm{c}v_t(Z_i)}{n_\mathrm{c}v_t(Z_i) + \bar{\gamma}_{t|t-1}\big(d_t\ell'_t(\cdot,Z_i)\big)},$$

which is the mean number of clutter generators a posteriori. This result is not usually part of the PHD filter equations, and is obtained here because the clutter generators have been included in the state space via ψ_c.

As far as time prediction is concerned, considering the first-moment measure $\bar{\gamma}_t$ of the point process $\bar{\mathscr{X}}_t|\mathscr{Z}_{0:t}$ together with the assumption that all the objects in the system evolve independently allows for expressing the first-moment measure $\gamma_{t+1|t}$ of $\mathscr{X}_{t+1}|\mathscr{Z}_{0:t}$ as a function of $\bar{\gamma}_t$.

Proposition 1 *The first-moment measure $\gamma_{t+1|t}$ of $\mathscr{X}_{t+1}|\mathscr{Z}_{0:t}$ is found to be*

$$\gamma_{t+1|t}(B) = \bar{\gamma}_t\big(M_t(\cdot, B)\big), \quad \forall B \in \mathscr{B}(\mathbf{X}_{t+1}). \tag{1.4}$$

In its original formulation [12], the prediction of the PHD filter contains a term related to *spawning*, which describes the cases where one object at time t generates several objects at time $t+1$. However, spawning is not very often considered in practical applications and is not taken into account here for the sake of simplicity. As with Theorem 1, the Markov kernel M_t can be split up in order to make Proposition 1 more explicit.

Remark 4 Using the notations introduced in Remark 2, (1.4) becomes

$$\gamma_{t+1|t}(B) = \gamma_{\mathrm{b},t}(B) + \gamma_t\big(s_t M'_t(\cdot, B)\big),$$

where $\gamma_{\mathrm{b},t} = n_{\mathrm{b},t}\eta_{\mathrm{b},t}$ is the first-moment measure of the birth point process. As in Remark 3, this is a more usual way to state the prediction equation of the PHD filter.

Remark 5 A slightly different way of reformulating the Markov kernel M_t can be found when proceeding as in Remark 2, but with a kernel M''_t from $\bar{\mathbf{X}}_t$ to $\mathbf{X}_t$ defined for any $x \in \bar{\mathbf{X}}_t$ as

$$M_t(x, B) = s_t(x)M''_t(x, B), \quad \forall B \in \mathscr{B}(\mathbf{X}_{t+1}),$$

where s_t is also defined on ψ_{b} and ψ_{c} as $s_t(\psi_{\mathrm{b}}) = 1$ and $s_t(\psi_{\mathrm{c}}) = 0$. This formulation will be useful in the next section when expressing the recursion as a Boltzmann–Gibbs transformation joined with a Markov kernel with $\mathbf{X}_{t+1}$ as a target space.

Equation (1.2) together with (1.4) describe a recursive algorithm that can be used at any time $t \in \mathbb{T}$, provided that some prior information is available. These two equations thus form a *filter*, namely, the PHD filter. However, when used recursively, the assumption that $\mathscr{X}_t|\mathscr{Z}_{0:t-1}$ is a Poisson point process becomes an approximation. Indeed, the updated point process $\mathscr{X}_{t-1}|\mathscr{Z}_{0:t-1}$ is not Poisson in general and neither is the predicted point process $\mathscr{X}_t|\mathscr{Z}_{0:t-1}$. In spite of this approximation, it can be shown that the output of the PHD filter is the best Poisson approximation of the actual updated point process in terms of Kullback–Leibler divergence [12, 20].

Standard implementations of the PHD filter include (*a*) a Gaussian mixture-based technique suitable whenever the underlying models are all linear and Gaussian, and (*b*) the sequential Monte Carlo approximation called the SMC-PHD filter. When the objective is to obtain an individual track for each object in the scene, the later implementation suffers from some accuracy-impairing limitations such as the need for applying clustering techniques to the output of the filter in order to obtain these track estimates. The objective is now to reformulate these equations in different ways, in order to highlight their structure and to enable the use of other known approximation schemes.

1.3.2 One-Step Filtering Equation

Under the same assumptions as in the previous section, the update and prediction Eqs. (1.2) and (1.4) can be joined together in order to form a one-step filtering algorithm. We still consider the extended version $\bar{\gamma}_{t|t-1} \doteq \Xi_t(\gamma_{t|t-1})$ of the first-moment measure of the predicted point process $\mathscr{X}_t|\mathscr{Z}_{0:t-1}$ and now express the first-moment measure $\gamma_{t+1|t}$ of $\mathscr{X}_{t+1}|\mathscr{Z}_{0:t}$ directly as a function of $\bar{\gamma}_{t|t-1}$ as follows:

$$\gamma_{t+1|t}(B) = \bar{\gamma}_{t|t-1}\big(Q_{t,\bar{\gamma}_{t|t-1}}(\cdot, B)\big), \quad \forall B \in \mathscr{B}(\mathbf{X}_{t+1}), \tag{1.5}$$

where $Q_{t,\gamma}(x, B) \doteq g_{t,\gamma}(x) M_t(x, B)$ is a bounded integral operator from $\bar{\mathbf{X}}_t$ to $\mathbf{X}_{t+1}$, indexed by $t \in \mathbb{T}$ and by a measure $\gamma \in \mathbf{M}(\bar{\mathbf{X}}_t)$. Equation (1.5) can be further reformulated as a Markov transport equation representing the distribution of the point process $\mathscr{X}_{t+1}|\mathscr{Z}_{0:t}$, together with a mass process which is associated to the expected number of objects in the scene:

$$(m_{t+1|t}, \eta_{t+1|t}) = \Lambda_t(m_{t|t-1}, \eta_{t|t-1}),$$

where, for any $t' \in \mathbb{T}$, the non-negative real number $m_{t'+1|t'} \doteq \gamma_{t'+1|t'}(\mathbf{1})$ is the total mass in $\gamma_{t'+1|t'}$ which is interpreted as the expected number of objects, where the probability measure $\eta_{t'+1|t'}$ in $\mathbf{P}(\mathbf{X}_{t'+1})$ defined as

$$\eta_{t'+1|t'}(B) \doteq \frac{\gamma_{t'+1|t'}(B)}{m_{t'+1|t'}}, \quad \forall B \in \mathscr{B}(\mathbf{X}_{t'+1}),$$

is the common law of the point process $\mathscr{X}_{t'+1}|\mathscr{Z}_{0:t'}$, and where the mapping

$$\begin{aligned} \Lambda_t : \mathbb{R}^+ \times \mathbf{P}(\bar{\mathbf{X}}_t) &\to \mathbb{R}^+ \times \mathbf{P}(\mathbf{X}_{t+1}) \\ (m, \eta) &\mapsto \big(\Lambda_t^{(1)}(m, \eta), \Lambda_t^{(2)}(m, \eta)\big), \end{aligned}$$

can be characterised via its two components $\Lambda_t^{(1)}$ and $\Lambda_t^{(2)}$ by

$$\begin{aligned} \Lambda_t^{(1)}(m, \eta) &= \bar{m}\bar{\eta}(G_{t,\bar{m}\bar{\eta}}) \\ \Lambda_t^{(2)}(m, \eta)(B) &= \Psi_{G_{t,\bar{m}\bar{\eta}}}(\bar{\eta})(\hat{M}_{t,\bar{m}\bar{\eta}}(\cdot, B)), \quad \forall B \in \mathscr{B}(\mathbf{X}_{t+1}), \end{aligned}$$

where $\bar{m}$ and $\bar{\eta}$ are characterised by the relation $\bar{m}\bar{\eta} = \Xi_t(m\eta)$ and where for any $\gamma \in \mathbf{M}(\bar{\mathbf{X}}_t)$ and any $x \in \bar{\mathbf{X}}_t$,

$$G_{t,\gamma}(x) \doteq Q_{t,\gamma}(\mathbf{1})(x), \quad \text{and} \quad \hat{M}_{t,\gamma}(x, B) \doteq \frac{Q_{t,\gamma}(x, B)}{Q_{t,\gamma}(\mathbf{1})(x)}.$$

This way of expressing the PHD filter will be helpful when proceeding to the analysis of the long-time behaviour of this filter in Sect. 1.4. Also, mean-field-related

approximation techniques can then be used for these distributions and mass, but this type of approach would not allow for bypassing the limitations already encountered with the standard SMC-PHD filter.

1.3.3 Association Measure

The SMC-PHD filter is very commonly used as a means of obtaining track estimates for the objects in the system of interest. In this situation, the main source of inaccuracy comes from the inability to extract track estimates directly from the particles, which makes the approach dependent on error-prone clustering algorithms. However, it is possible to use observation histories in order to find a natural partition of the set of particles in a given SMC implementation of the PHD filter. This can be seen as a generalisation of the idea presented in [19] where the SMC-PHD is reformulated in order to distinguish persisting and appearing objects by associating a different label to each of these parts of the considered multi-object system.

Formulation

First, define the random finite set $\mathfrak{Z}_t$ and its extension $\bar{\mathfrak{Z}}_t$ as the respective supports of the point processes $\mathscr{Z}_t$ and $\bar{\mathscr{Z}}_t$. In this way, the random set $\bar{\mathfrak{Y}}_t$ of observation histories, or *observation paths*, can be defined as the set containing all the sequences of observations from time 0 to t, i.e. as

$$\bar{\mathfrak{Y}}_t \doteq \left\{(z_0, \ldots, z_t) : z_{t'} \in \bar{\mathfrak{Z}}_{t'}, 0 \leq t' \leq t\right\}.$$

The set $\bar{\mathfrak{Y}}_t$ is a random finite subset of the space $\bar{\mathbf{Y}}_t \doteq \bar{\mathbf{Z}}_0 \times \ldots \times \bar{\mathbf{Z}}_t$. We consider in particular the element $\boldsymbol{\phi}_t$ referred to as the *empty observation path* and defined as the sequence in $\bar{\mathbf{Y}}_t$ such that $z_{t'} = \phi$, for all $0 \leq t' \leq t$. We also define the random set of non-empty observation paths $\mathfrak{Y}_t \doteq \bar{\mathfrak{Y}}_t \setminus \boldsymbol{\phi}_t$, which is in the space $\mathbf{Y}_t \doteq \bar{\mathbf{Y}}_t \setminus \boldsymbol{\phi}_t$ of observation paths with at least one non-empty observation. Examples of observation paths are given in Fig. 1.2.

Note that for each random observation path $\boldsymbol{y}$ in $\mathbf{Y}_{t-1}$, we can define the probability measure $p^{(\boldsymbol{y})}_{t|t-1} \in \mathbf{P}(\mathbf{X}_t)$ as the predicted law on the random state of a single object given the observation history $\boldsymbol{y}$. Also, a measure α_{t-1} on $\bar{\mathbf{Y}}_{t-1}$ with support $\bar{\mathfrak{Y}}_{t-1}$ is introduced and is assumed to give, for any $\boldsymbol{y}$ in $\mathfrak{Y}_{t-1}$, the probability for the corresponding law $p^{(\boldsymbol{y})}_{t|t-1}$ to represent an object of the system of interest. The measure α_{t-1} is referred to as the *association measure* at time $t-1$.

The empty observation path possibly describes several objects (all the undetected ones), and this part of the system is characterised by the expected number of objects in this situation given by $\alpha_{t-1}(\{\boldsymbol{\phi}_{t-1}\})$ and by their common distribution $p^{(\boldsymbol{y})}_{t|t-1}$ with $\boldsymbol{y} = \boldsymbol{\phi}_{t-1}$. As before, we need to extend the predicted first-moment measure with objects to be born at time $t+1$ and clutter generators. This extension can

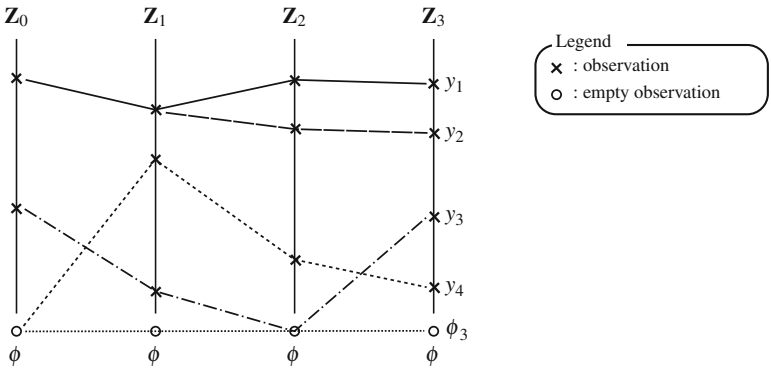


Fig. 1.2 Examples of observation paths at time $t = 3$: $\boldsymbol{y}_1$ represents a case where the corresponding object would be detected at all times, $\boldsymbol{y}_2$ is a variation of $\boldsymbol{y}_1$ where the two last observations differ, $\boldsymbol{y}_3$ shows a case where a missed detection occurred at time $t = 2$, $\boldsymbol{y}_4$ represents a case where the associated object might have been missed detected at the first time step or might not have existed at all at this time, and finally $\boldsymbol{\phi}_3$ is the empty observation path at time $t = 3$

be jointly performed on the association measure α_{t-1} and on the predicted laws $p^{(\boldsymbol{y})}_{t|t-1}$, $\boldsymbol{y} \in \mathbf{Y}_{t-1}$, as follows:

$$\bar{\alpha}_{t-1}\big(\bar{p}^{(\cdot)}_{t|t-1}\big) = \Xi_t\Big(\alpha_{t-1}\big(p^{(\cdot)}_{t|t-1}\big)\Big).$$

Since the objects that are to be born at time $t+1$ and the clutter generators have never been detected (recalling that clutter generators are renewed at every time step), only the law $p^{(\boldsymbol{y})}_{t|t-1}$ with $\boldsymbol{y} = \boldsymbol{\phi}_{t-1}$ and the association measure α_{t-1} at point $\boldsymbol{\phi}_{t-1}$ need to be modified, and we consider that

$$\bar{\alpha}_{t-1}|_{\mathbf{Y}_{t-1}} = \alpha_{t-1}|_{\mathbf{Y}_{t-1}}, \quad \text{and} \quad \bar{p}^{(\boldsymbol{y})}_{t|t-1} = p^{(\boldsymbol{y})}_{t|t-1}, \quad \forall \boldsymbol{y} \in \mathbf{Y}_{t-1}.$$

The next theorem makes use of the notations introduced in Remark 5. In particular, the function s_t describing the probability of survival at any point of the state space $\bar{\mathbf{X}}_t$ and the Markov kernel M''_t from $\bar{\mathbf{X}}_t$ to $\mathbf{X}_{t+1}$ will be required to express the predicted first-moment measure $\gamma_{t+1|t}$ in a concise way.

Theorem 2 *The first-moment measure $\gamma_{t+1|t}$ of the conditional point process $\mathscr{X}_{t+1}|\mathscr{Z}_{0:t}$ can be expressed, for any $B \in \mathscr{B}(\mathbf{X}_{t+1})$, as*

$$\gamma_{t+1|t}(B) = \int_{\bar{\mathbf{Y}}_t} \alpha_t(\mathrm{d}\boldsymbol{y})\, p^{(\boldsymbol{y})}_{t+1|t}(B),$$

where the measure $p^{(\cdot)}_{t+1|t} \in \mathbf{M}(\bar{\mathbf{X}}_{t+1})$ *is found to be*

$$p^{(\boldsymbol{y},z)}_{t+1|t}(B) = \Psi_{s_t \ell_t(\cdot,z)}(\bar{p}^{(\boldsymbol{y})}_{t-1|t})\big(M''_t(\cdot, B)\big), \quad \forall \boldsymbol{y} \in \bar{\mathbf{Y}}_{t-1}, \forall z \in \bar{\mathbf{Z}}_t,$$

and where the association measure α_t *on the space* $\bar{\mathbf{Y}}_t = \bar{\mathbf{Y}}_{t-1} \times \bar{\mathbf{Z}}_t$ *can be expressed as a function of* $\bar{\alpha}_{t-1}$ *as*

$$\begin{aligned}\alpha_t(\mathrm{d}(\boldsymbol{y}, z)) = \bar{\alpha}_{t-1}(\mathrm{d}\boldsymbol{y})\Bigg[&\delta_\phi(\mathrm{d}z)\,\bar{p}^{(\boldsymbol{y})}_{t|t-1}\big(s_t \ell_t(\cdot, \phi)\big) \\ &+ \mathscr{Z}_t(\mathrm{d}z)\frac{\bar{p}^{(\boldsymbol{y})}_{t|t-1}\big(s_t \ell_t(\cdot, z)\big)}{\bar{\alpha}_{t-1}\big(\bar{p}^{(\cdot)}_{t|t-1}(\ell_t(\cdot, z))\big)}\Bigg].\end{aligned} \tag{1.6}$$

The identification of observation paths in the expression of the PHD filter provides a natural way of clustering the first-moment measure $\bar{\gamma}_{t+1|t}$. The advantage in such an identification is twofold: (*a*) in terms of interpretation, since the output of the filter can now be understood as being made of interacting single-object filters, and (*b*) in terms of implementation, since the induced clustering makes unnecessary the use of an additional clustering algorithm for the extraction of track estimates with SMC-PHD filters.

Remark 6 (Forward–backward) Without additional derivations, the fact that the single-object posterior laws $p^{(\boldsymbol{y})}_{t+1|t}$ appear explicitly in the expression of the first-moment measure $\gamma_{t+1|t}$ allows for considering objectwise forward–backward algorithms. More general forward–backward filters based on the equations of the PHD filter have been studied in [15] and [3].

Remark 7 (Parameter estimation) For similar reasons as in Remark 6, a 'local' parameter estimation can be performed for each observation path, e.g. if the variance of the noise of the observation process is unknown and varies from object to object. A global parameter estimation is also possible but requires a reformulation of the filtering equations in order to integrate system-wide uncertainties, as in [18] and [10].

In order to devise approximations for the filtering equation described in Theorem 2, the different terms involved need to be rewritten in a suitable way. First, the recursive expression (1.6) of α_t can be expressed in a more concise way as

$$\alpha_t(\mathrm{d}(\boldsymbol{y}, z)) = \big(\bar{\alpha}_{t-1} \otimes \bar{\mathscr{Z}}_t\big)\big(\mathrm{d}(\boldsymbol{y}, z)\big)\,\bar{p}^{(\boldsymbol{y})}_{t|t-1}\big(G^{(z)}_{t,\bar{\gamma}_{t|t-1}}\big),$$

where $\bar{\alpha}_{t-1} \otimes \bar{\mathscr{Z}}_t$ refers to the product measure of $\bar{\alpha}_{t-1}$ and $\bar{\mathscr{Z}}_t$, and where the potential function $G^{(z)}_{t,\gamma}$ is defined, for any observation $z \in \bar{\mathbf{Z}}_t$ and any measure $\gamma \in \mathbf{M}(\bar{\mathbf{X}}_t)$ such that $\gamma(\ell_t(\cdot, z)) > 0$, as

$$G^{(z)}_{t,\gamma}(x) \doteq \begin{cases} \dfrac{s_t(x)\ell_t(x, z)}{\gamma(\ell_t(\cdot, z))} & \text{if } z \in \mathbf{Z}_t \\ s_t(x)\ell_t(x, \phi) & \text{if } z = \phi. \end{cases}$$

Defining, for any $t \in \mathbb{T}$, the probability measure $\beta_t \in \mathbf{P}(\bar{\mathbf{Y}}_t)$ as the normalised association measure at time t and $a_t \in \mathbb{R}_+$ as the total mass in α_t, i.e.

$$a_t \doteq \alpha_t(\mathbf{1}), \quad \text{and} \quad \beta_t \doteq \frac{1}{a_t}\alpha_t,$$

the recursion (1.6) for the association measure $\alpha_t \in \mathbf{M}(\bar{\mathbf{Y}}_t)$ can be translated into a recursion for the normalised association measure $\beta_t \in \mathbf{P}(\bar{\mathbf{Y}}_t)$ as

$$\beta_t = \Pi_{t-1}(a_{t-1}, \beta_{t-1}),$$

where Π_{t-1} describes concisely the update of the measure β_{t-1}, and is defined as

$$\begin{aligned} \Pi_{t-1} : \mathbb{R}_+ \times \mathbf{P}(\bar{\mathbf{Y}}_{t-1}) &\to \mathbf{P}(\bar{\mathbf{Y}}_t) \\ (a, \beta) &\mapsto \Psi_{G'_{\bar{a},\bar{\beta}}}(\bar{\beta} \otimes \bar{\mathscr{Z}}_t) \end{aligned}$$

where $\bar{a}$ and $\bar{\beta}$ are characterised by

$$\bar{a}\bar{\beta}(\bar{p}^{(\cdot)}_{t|t-1}) = \Xi_t\big(a\beta(p^{(\cdot)}_{t|t-1})\big),$$

and where the potential function $G'_{a,\beta}$ is defined, for any $\boldsymbol{y} \in \bar{\mathbf{Y}}_{t-1}$ and any $z \in \bar{\mathbf{Z}}_t$, as

$$G'_{\bar{a},\bar{\beta}}(\boldsymbol{y}, z) \doteq \bar{p}^{(\boldsymbol{y})}_{t|t-1}\left(G^{(z)}_{t,\bar{a}\bar{\beta}(\bar{p}^{(\cdot)}_{t|t-1})}\right).$$

Describing the recursion of the association measure through the transformation of a probability measure allows for considering approximations as in the following section.

Approximation

If we assume that the single-object posterior laws $p^{(\boldsymbol{y})}_{t+1|t}$ can be computed explicitly for any $\boldsymbol{y} \in \bar{\mathbf{Y}}_t$, as in the linear Gaussian case with the Kalman filter, then the main source of complexity lies in the fast augmentation of the number of observation paths that form the support of α_t when t increases. Indeed, the recursive expression of α_t could be reformulated as a product measure of α_{t-1} and $\bar{\mathscr{Z}}_t$, thus indicating that the number of observation paths is multiplied by the number of observations at each time step. In order to control the number of considered observation paths, one can resort to a mean-field-type approximation consisting in the introduction of an empirical measure in place of the exact association measure α_t. Formally, define

$$\beta^N_0 \doteq \frac{1}{N}\sum_{i=1}^{N} \delta_{\boldsymbol{y}^i_0}$$

as the empirical measure associated with the collection of N independent and identically distributed (i.i.d.) random variables $\boldsymbol{y}_0^1, \ldots, \boldsymbol{y}_0^N$ with the initial normalised association measure β_0 as common distribution. The normalised first-moment measure $\eta_{1|0} \in \mathbf{P}(\mathbf{X}_1)$ can then be approximated by

$$\eta_{1|0}^N(B) = \int \beta_0^N(\mathrm{d}\boldsymbol{y}) p_{1|0}^{(\boldsymbol{y})}(B), \quad \forall B \in \mathscr{B}(\mathbf{X}_1).$$

This approach allows for limiting the number of observation paths at time $t = 1$ but, when applying the filtering step at time $t = 2$, the number of observation paths will be multiplied again. The normalised first-moment measure $\eta_{2|1} \in \mathbf{P}(\mathbf{X}_2)$ is then approximated by

$$\eta_{2|1}^N(B) = \int \beta_1^N(\mathrm{d}\boldsymbol{y}) p_{2|1}^{(\boldsymbol{y})}(B), \quad \forall B \in \mathscr{B}(\mathbf{X}_2),$$

where β_1^N is an empirical measure based on N i.i.d. random variables with common law $\Pi_0(a_0, \beta_0^N)$. This approach can then be used recursively, so that the approximated normalised first-moment measure at time $t + 1$ can be expressed using an empirical measure β_t^N, which is itself based on the update via Π_{t-1} of another empirical measure β_{t-1}^N at time $t - 1$.

When the single-object posterior laws cannot be computed via the Kalman filter, other approximations need to be considered at the object level and two-level particle models can be introduced.

Remark 8 (MCMC exploration) In order to improve the exploration of the space of observation paths by the empirical association measure, Markov chain Monte Carlo (MCMC) algorithms can be used on a suitable time window. This window will depend on the time of stability of the filtering equations since any modification of an association prior to this time will have almost no effect on the current state of the system. Practically, these MCMC steps could be used when the computations for the sequential estimation have been performed and some computational power is available before the acquisition of the next set of observations. This differs from approaches where the full multi-target problem is solved using MCMC algorithms as in [24] and [10].

Similar Association Paths

Very often, several observation paths will have in common many of the observations composing them. This is mainly due to detection uncertainty, to objects being close to each other with respect to the accuracy of the sensor, or to spurious observations being close to an actual object. In consequence, the posterior laws corresponding to these observation paths might be close statistically. Although it does not cause any issues in the usual implementations of the PHD filter—because of the mixture reduction techniques used in Gaussian mixture implementations and because of the absence of natural clustering in the standard SMC implementation—this aspect has to

be dealt with when considering observation paths for two reasons: (*a*) it might reduce the efficiency of the approximate filter detailed above by accumulating redundant information, and (*b*) it makes the extraction of track estimates more difficult since one object might be represented by multiple observation paths.

Even though there exist statistical distances between Gaussian distributions that can be computed explicitly, such as the Mahalanobis or the Hellinger distance, there is no counterpart for empirical measures and alternative ways have to be found in order to obtain a consistent treatment of both cases. One solution lies in the identification of the causes of the statistical similarity between two posterior laws: (*a*) the stability of the considered single-object filter, which reduces the effect of past observations on the current state of the object, and (*b*) the robustness, which relates to the fact that a small change in the recent observations only induces small modifications of the posterior distribution. For instance, if a (possibly statistical) distance $d_{\bar{\mathbf{Z}}'_t}$ is available for observations in $\bar{\mathbf{Z}}_{t'}$ at any time $t' \in \mathbb{T}$, then an example of a distance on observation paths up to time $t \in \mathbb{T}$ would take the form

$$d_{\bar{\mathbf{Y}}_t}(\boldsymbol{y}, \boldsymbol{y}') \propto \sum_{t' \leq t} \exp\big(-c(t-t')\big) d_{\bar{\mathbf{Z}}_{t'}}\big(\boldsymbol{y}_{t'}, \boldsymbol{y}'_{t'}\big),$$

where $c \in \mathbb{R}_+$ is a coefficient that controls the duration required for large deviations of observations to become negligible in the overall distance between observation paths. The fact that $d_{\bar{\mathbf{Y}}_t}$ is a distance follows directly.

1.4 Analysis

The objective in this section is to recall a result of [6] about the long-time behaviour of the measure-valued-process form of the PHD filter introduced in Sect. 1.3.2. Noting that this measure-valued process was factorised into a total mass $m_{t+1|t} \in \mathbb{R}_+$ and a probability measure $\eta_{t+1|t} \in \mathbf{P}(\bar{\mathbf{X}}_{t+1})$, one approach for studying the stability properties of this process is to proceed as follows: (a) find properties related to the behaviour of the mass for a given flow of probability measure, (b) find properties related to the behaviour of the probability measure for a given flow of mass, and (c) find sufficient conditions for the two previous results to be combined into a global analysis of the process.In order to state the results, we consider, for any measurable space $(\mathbf{E}, \mathscr{E})$, the set $\mathrm{Osc}_1(\mathbf{E})$ made of the measurable functions on $(\mathbf{E}, \mathscr{E})$ with oscillations less than one, i.e. the functions $f \in \mathbf{B}(\mathbf{E})$ such that

$$\sup_{x,y \in \mathbf{E}} |f(x) - f(y)| \leq 1.$$

We first assume that there exists a collection $I_0, I_1, \ldots$ of compact subsets of $\mathbb{R}_+$ such that the mass process verifies $m_{t+1|t} \in I_{t+1}$ for any $t \in \mathbb{T}$ as long as it holds that $m_0 \in I_0$. Then, for any $t \in \mathbb{T}$, for any collection $m_0, m_1, \ldots$ of non-negative

real numbers verifying $m_t \in I_t$, for all $t \in \mathbb{T}$, and for any collection $\eta_0, \eta_1, \ldots$ of probability measures such that $\eta_t \in \mathbf{P}(\bar{\mathbf{X}}_t)$, for all $t \in \mathbb{T}$, we introduce the semigroups

$$\Phi^{(1)}_{t',t,\eta} \doteq \Phi^{(1)}_{t,\eta_t} \circ \cdots \circ \Phi^{(1)}_{t',\eta_{t'}} \quad \text{and} \quad \Phi^{(2)}_{t',t,m} \doteq \Phi^{(2)}_{t,m_t} \circ \cdots \circ \Phi^{(2)}_{t',m_{t'}},$$

where the transformations $\Phi^{(1)}_{t,\eta_t} : I_t \to I_{t+1}$ and $\Phi^{(2)}_{t,m_t} : \mathbf{P}(\bar{\mathbf{X}}_t) \to \mathbf{P}(\bar{\mathbf{X}}_{t+1})$ are defined as

$$\Phi^{(1)}_{t,\eta_t}(n) = \Lambda^{(1)}_t(n, \eta_t) \quad \text{and} \quad \Phi^{(1)}_{t,m_t}(\mu) = \Lambda^{(2)}_t(m_t, \mu).$$

Details about these models can also be found in [5].

Two properties of the semigroup transformations $\Phi^{(1)}_{t',t,\eta}$ and $\Phi^{(2)}_{t',t,m}$ will prove to be of importance in the analysis of the long-time behaviour of the underlying measure-valued process:

(**L**) For any time step $t' \leq t$, any masses $n, n' \in I_{t'}$, any probability measures $\mu, \mu' \in \mathbf{P}(\bar{\mathbf{X}}_{t'})$ and any $f \in \mathrm{Osc}_1(\bar{\mathbf{X}}_{t+1})$, the following Lipschitz inequalities hold:

$$\left|\Phi^{(1)}_{t',t,\eta}(n) - \Phi^{(1)}_{t',t,\eta}(n')\right| \leq c^{(1)}_{t',t}|n - n'|$$
$$\left|\left[\Phi^{(2)}_{t',t,m}(\mu) - \Phi^{(2)}_{t',t,m}(\mu')\right](f)\right| \leq c^{(2)}_{t',t} \int |[\mu - \mu'](\varphi)| K_{t',t,\mu'}(f, \mathrm{d}\varphi),$$

for some

- finite constants $c^{(1)}_{t',t}$ and $c^{(2)}_{t',t}$ that only depend on t' and t,
- collection of Markov kernels $K_{t',t,\mu'}$ from $\mathrm{Osc}_1(\bar{\mathbf{X}}_{t+1})$ to $\mathrm{Osc}_1(\bar{\mathbf{X}}_{t'})$ that only depend on t', t and μ'.

(**C**) For any time step $t \in \mathbb{T}$, any masses $n, n' \in I_t$, any probability measures $\mu, \mu' \in \mathbf{P}(\bar{\mathbf{X}}_t)$ and any $f \in \mathrm{Osc}_1(\bar{\mathbf{X}}_{t+1})$, the following continuity inequalities hold:

$$\left|\Phi^{(1)}_{t,\mu}(n) - \Phi^{(1)}_{t,\mu'}(n)\right| \leq c'^{(1)}_t \int |[\mu - \mu'](\varphi)| P_{t,\mu'}(\mathrm{d}\varphi)$$
$$\left|\left[\Phi^{(2)}_{t,n}(\mu) - \Phi^{(2)}_{t,n'}(\mu)\right](f)\right| \leq c'^{(2)}_t |n - n'|,$$

for some

- finite constants $c'^{(1)}_t$ and $c'^{(2)}_t$ that only depend on t,
- collection of probability measures $P_{t,\mu'}$ on $\mathrm{Osc}_1(\bar{\mathbf{X}}_t)$ that only depend on t and μ'.

Property (**L**) is related to the regularity of $\Phi^{(1)}_{t',t,\eta}$ and $\Phi^{(2)}_{t',t,m}$ when their argument is modified, whereas Property (**C**) refers to the stability of the one-step transformations $\Phi^{(1)}_{t,\cdot}$ and $\Phi^{(2)}_{t,\cdot}$ when their parameter is changed. These two properties can be combined in order to analyse the long-time behaviour of the evolution semigroup

$$\Lambda_{t',t} \doteq \Lambda_{t-1} \circ \cdots \circ \Lambda_{t'},$$

for any $t, t' \in \mathbb{T}$ such that $t' \leq t$, where we consider that $\Lambda_{t,t} = \mathrm{Id}$ by convention. In the following theorem, the binary relations $\wedge$ and $\vee$ are used to, respectively, denote the minimum and the maximum of two real numbers.

Theorem 3 *(From [6], Theorem 13.3.3) Assume that the properties (**L**) and (**C**) are verified for the semigroup transformations $\Phi^{(1)}_{t',t,\eta}$ and $\Phi^{(2)}_{t',t,m}$ with constants $c^{(i)}_{t',t}$ and $c'^{(i)}_t$, $i = 1, 2$, such that, for any $t, t' \in \mathbb{T}$ with $t' \leq t$, it holds that*

$$c^{(i)}_{t',t} \leq a_i e^{-b_i(t-t')} \quad \text{and} \quad c'_i = \sup_{t \in \mathbb{T}} c'^{(i)}_t < \infty,$$

where a_i and $b_i > 0$ are some finite constants, $i = 1, 2$, which are assumed to verify $b_1 \neq b_2$ and

$$a_1 a_2 c'_1 c'_2 \leq \left(1 - e^{-(b_1 \wedge b_2)}\right)\left(e^{-(b_1 \wedge b_2)} - e^{-(b_1 \vee b_2)}\right).$$

Then, for any time step $t' \leq t$, any masses $n, n' \in I_{t'}$, and any probability measures $\mu, \mu' \in \mathbf{P}(\bar{\mathbf{X}}_{t'})$, the following Lipschitz inequalities hold:

$$\left|\Lambda^{(1)}_{t',t}(n, \mu) - \Lambda^{(1)}_{t',t}(n', \mu')\right| \leq e^{-b(t-t')}\left(a_{1,1}|n - n'| + a_{1,2}\int \left|[\mu - \mu'](\varphi)\right| P'_{t',t,n',\mu'}(\mathrm{d}\varphi)\right)$$

and, for any $f \in \mathrm{Osc}_1(\bar{\mathbf{X}}_{t+1})$,

$$\left|\Lambda^{(2)}_{t',t}(n, \mu)(f) - \Lambda^{(2)}_{t',t}(n', \mu')(f)\right| \leq e^{-b(t-t')}\left(a_{2,1}|n - n'| + a_{2,2}\int \left|[\mu - \mu'](\varphi)\right| K'_{t',t,n',\mu'}(f, \mathrm{d}\varphi)\right),$$

where $P'_{t',t,n',\mu'}$ is a collection of probability measures on $\mathrm{Osc}_1(\bar{\mathbf{X}}_{t'})$, where $K'_{t',t,n',\mu'}$ is a collection of Markov kernels from $\mathrm{Osc}_1(\bar{\mathbf{X}}_{t+1})$ to $\mathrm{Osc}_1(\bar{\mathbf{X}}_{t'})$, where

$$b = (b_1 \wedge b_2) - \log\left(1 + a_1 a_2 c'_1 c'_2 \frac{e^{b_1 \wedge b_2}}{e^{-(b_1 \wedge b_2)} - e^{-(b_1 \vee b_2)}}\right)$$

and where the constants $a_{i,j}$, $i, j = 1, 2$, are such that

$$\begin{aligned} a_{1,1} &= a_1\left(1 + a_{2,1} c'_1/(e^{-b} - e^{-b_1})\right) \\ a_{1,2} &= a_1 a_2 c'_1/(e^{-b} - e^{-b_1}) \\ a_{2,1} &= a_1 a_2 c'_2/\left(e^{-(b_1 \wedge b_2)} - e^{-(b_1 \vee b_2)}\right) \\ a_{2,2} &= a_2. \end{aligned}$$

This important theorem provides conditions for the semigroup transformations $\Phi^{(1)}_{t',t,\eta}$ and $\Phi^{(2)}_{t',t,m}$ to satisfy (exponential) stability properties. The objective is now to understand under which practical situations these conditions are satisfied for the PHD filter. In order to simplify the presentation, we consider the following assumptions:

- the probability of detection d_t, the probability of survival s_t as well as the distribution ν_t and the expected number $n_{\mathrm{c},t}$ of spurious observations, are constant in time and space and are thus, respectively, denoted as d, s, ν and n_{c},
- the spaces $\mathbf{X}_t$ and $\mathbf{Z}_t$, the Markov transition M_t, the likelihood ℓ_t as well as the distribution $\eta_{\mathrm{b},t}$ and the expected number $n_{\mathrm{b},t}$ of appearing objects, are constant in time and are, respectively, denoted as $\mathbf{X}$, $\mathbf{Z}$, M, ℓ, η_{b} and n_{b}.

We additionally assume that

- it holds that $n_{\mathrm{b}} > 0$ and $s > 0$,
- for any $z \in \mathbf{Z}$, it holds that

$$\ell^{(-)}(z) \doteq \inf_{x \in \mathbf{X}} \ell(x, z) \geq 0 \quad \text{and} \quad \ell^{(+)}(z) \doteq \sup_{x \in \mathbf{X}} \ell(x, z) < \infty.$$

We are now in position to specify the practical situations under which the PHD filter is stable.

Theorem 4 *(From [6], Theorem 13.4.1) If the quantity* $\sup_{t \in \mathbb{T}} \mathscr{Z}_t(f)$ *is finite for* f *equal to* $\ell^{(+)}/\ell^{(-)}$ *and* $\ell^{(+)}/(\ell^{(-)})^2$*, then there exist constants*

$$0 < r_d \leq 1, \quad r_{\mathrm{b}} < \infty, \quad \textit{and} \quad r_{\mathrm{c}} > 0,$$

such that $\Phi^{(1)}_{t',t,\eta}$ *and* $\Phi^{(2)}_{t',t,m}$ *satisfy the conditions of Theorem 3 whenever*

$$d \geq r_d, \quad n_{\mathrm{b}} \geq r_{\mathrm{b}}, \quad \textit{and} \quad n_{\mathrm{c}} \leq r_{\mathrm{c}}.$$

The result of Theorem 4 can be informally stated as: the PHD filter is exponentially stable when the probability of detection is sufficiently high, when the expected number of appearing objects is large enough, and when the number of spurious observations is limited. Although the claims related to the detection and to the spurious observations are natural, it is useful to note that reducing the expected number of appearing objects can have a negative impact on the stability. This is due to the fact that in case of failure, a high birth rate will allow for the filter to recover quickly by reinitialising lost tracks.

1.5 Derivation

Two different proofs of Theorem 1 are first detailed using the notations introduced so far, establishing a link between (*a*) the proof originally proposed in [12], which relies on the concept of probability generating functionals, and (*b*) a 'direct' proof

subsequently proposed in [2], for which only standard probabilistic concepts are used. The association-measure formulation of the PHD filter is demonstrated in the last part of this section.

In the derivation of the update equations, the empty state ψ_{b} is ignored since all objects with such a state will almost surely not be detected, so that this point state is not affected during the update.

1.5.1 With Probability Generating Functionals

This approach is the one that has been originally used in [12] and is here adapted to the measure theoretic notations introduced in the previous section.

Definition and Properties

We follow [4] for the definition of the concept of probability generating functional, and we first introduce $\mathbf{V}(\mathbf{X})$ as the set of measurable functions h on $(\mathbf{X}, \mathscr{B}(\mathbf{X}))$, for any Polish space $\mathbf{X}$, such that $1 - h$ vanishes outside some bounded set and such that $0 \leq h(x) \leq 1$ for any $x \in \mathbf{X}$. The probability generating functional (p.g.fl.) of a given point process $\mathscr{X}$ on $\mathbf{X}$ is defined as

$$G(h) \doteq \mathbb{E}\left(\exp\left(\int \log h(x)\, \mathscr{X}(\mathrm{d}x) \right) \right), \quad \forall h \in \mathbf{V}(\mathbf{X}).$$

The p.g.fl. G characterises the point process $\mathscr{X}$. If the point process $\mathscr{X}$ is written as a sum $\sum_{i=1}^{N} \delta_{X_i}$, then the p.g.fl. can be equivalently expressed as

$$G(h) = \mathbb{E}\left(\prod_{i=1}^{N} h(X_i) \right), \quad \forall h \in \mathbf{V}(\mathbf{X}).$$

The property of p.g.fl.s that will be of particular interest in this proof is that the first-moment measure $\gamma \in \mathbf{M}(\mathbf{X})$ of $\mathscr{X}$ can be recovered from them through functional differentiation as follows:

$$\gamma(f) = \delta G(\mathbf{1}; f), \quad \forall f \in \mathbf{B}(\mathbf{X}),$$

where, for any $h, v \in \mathbf{V}(\mathbf{X})$, the term $\delta G(h; u)$ denotes the functional derivative of the functional G at point h and in the direction u. Higher order derivatives are accordingly denoted as $\delta G(h; u_1, \ldots, u_n)$ for any $h, u_1, \ldots, u_n$ in $\mathbf{V}(\mathbf{X})$.

In the case where $\mathscr{X}$ is a Poisson point process, the p.g.fl. G can be expressed as a function of the first-moment measure γ of $\mathscr{X}$ as

$$G(h) = \exp\big(-\gamma(1 - h) \big), \quad \forall h \in \mathbf{V}(\mathbf{X}).$$

Indeed, a Poisson point process is characterised by both its p.g.fl. and its first-moment measure.

Finally, it is worth noting that if $\mathscr{X}_1, \ldots, \mathscr{X}_n$ is a collection of independent point processes on $\mathbf{X}$, then the p.g.fl. of the superposition $\mathscr{X} = \sum_{i=1}^{n} \mathscr{X}_i$ is found to be

$$G(h) = \prod_{i=1}^{n} G_i(h), \quad \forall h \in \mathbf{V}(\mathbf{X}), \tag{1.7}$$

where G_i denotes the p.g.fl. of the point process $\mathscr{X}_i$, $1 \leq i \leq n$.

p.g.fl. Representation of Multi-object Systems

Now considering the notations introduced in the previous sections, we start by jointly expressing the point processes $\bar{\mathscr{X}}_{t|t-1} \doteq \bar{\mathscr{X}}_t | \mathscr{Z}_{0:t-1}$ and $\bar{\mathscr{Z}}_t$.

Proposition 2 *The point processes $\bar{\mathscr{X}}_{t|t-1}$ and $\bar{\mathscr{Z}}_t$ are jointly characterised by the p.g.fl. H on $\bar{\mathbf{X}}_t \times \bar{\mathbf{Z}}_t$ expressed as*

$$H(g, h) = G_{t|t-1}(hL_t(g)), \quad \forall g \in \mathbf{V}(\bar{\mathbf{Z}}_t), \forall h \in \mathbf{V}(\bar{\mathbf{X}}_t),$$

where $G_{t|t-1}$ is the p.g.fl. of $\bar{\mathscr{X}}_{t|t-1}$, which is found to be

$$G_{t|t-1}(h) = G_{\mathrm{c}}(h) \exp\big(-\gamma_{t|t-1}(1-h)\big),$$

where $G_{\mathrm{c}}(h) = \exp(-n_{\mathrm{c},t}\delta_{\psi_{\mathrm{c}}}(1-h))$ is the p.g.fl. associated to the clutter generators.

Proof First, since both the point process $\mathscr{X}_{t|t-1} \doteq \mathscr{X}_t | \mathscr{Z}_{0:t-1}$ and the random number of clutter generators are Poisson, the corresponding p.g.fl.s can be easily deduced. The p.g.fl. $G_{t|t-1}$ is then found to be characterising the superposition of the point process $\mathscr{X}_{t|t-1}$ with the random numbers of clutter generators, i.e. according to (1.7),

$$G_{t|t-1}(h) = G_{\mathrm{c}}(h) \exp\big(-\gamma_{t|t-1}(1-h)\big).$$

In order to derive the expression of the p.g.fl. H, we first notice that all the objects in the extended population described by $\bar{\mathscr{X}}_{t|t-1}$ generate one (possibly empty) observation through L_t. In consequence, the p.g.fl. describing the generation of an observation for a single object is found to be

$$G'_t(g)(x) = L_t(g)(x), \quad \forall g \in \mathbf{V}(\bar{\mathbf{Z}}_t), \forall x \in \bar{\mathbf{X}}_t.$$

Then, since the observation of each object and the generation of clutter are assumed to happen independently, the p.g.fl. of the observation process $\bar{\mathscr{Z}}_t$ is found to be

$$G_{\bar{\mathscr{Z}}_t}(g)(X) = \prod_{i=1}^{n} G'_t(g)(x_i) = \prod_{i=1}^{n} L_t(g)(x_i), \quad \forall X \doteq \sum_{i=1}^{n} \delta_{x_i} \in \mathbf{N}(\bar{\mathbf{X}}_t).$$

Finally, the p.g.fl. H that jointly characterises $\bar{\mathscr{X}}_{t|t-1}$ and $\bar{\mathscr{Z}}_t$ is expressed, for any $h \in \mathbf{V}(\bar{\mathbf{X}}_t)$, as

$$H(g,h) = p(0) + \sum_{n\geq 1} p(n) \int h(x_1)\dots h(x_n) G_{\bar{\mathscr{Z}}_t}(g) \Big(\sum_{i=1}^{n} \delta_{x_i} \Big) \eta(\mathrm{d}x_1)\dots\eta(\mathrm{d}x_n)$$
$$= p(0) + \sum_{n\geq 1} p(n) \int \Big[\prod_{i=1}^{n} h(x_i) L_t(g)(x_i) \Big] \eta(\mathrm{d}x_1)\dots\eta(\mathrm{d}x_n),$$

where p and $\eta \in \mathbf{P}(\bar{\mathbf{X}}_t)$ are, respectively, the cardinality distribution and the common distribution of $\bar{\mathscr{X}}_{t|t-1}$. The p.g.fl. of the point process $\bar{\mathscr{X}}_{t|t-1}$ with the function $hL_t(g)$ as an argument can then be recognised, i.e.

$$H(g,h) = G_{t|t-1}(hL_t(g)),$$

which ends the proof of the proposition.

We assume that expectations and functional differentiation can always be exchanged, see [20]. The following lemma is related to [12], but adapted to the approach considered here.

Lemma 1 *The p.g.fl. G_t of the updated point process $\bar{\mathscr{X}}_{t|t} \doteq \bar{\mathscr{X}}_t | \mathscr{Z}_{0:t}$ can be expressed as a function of the joint p.g.fl. H of $\bar{\mathscr{X}}_{t|t-1}$ and $\bar{\mathscr{Z}}_t$ as*

$$G_t(h) = \frac{\delta H\big(g,h;\mathbf{1}_{Z_1},\dots,\mathbf{1}_{Z_{N_t'}}\big)\big|_{g=\mathbf{1}_\phi}}{\delta H\big(g,\mathbf{1};\mathbf{1}_{Z_1},\dots,\mathbf{1}_{Z_{N_t'}}\big)\big|_{g=\mathbf{1}_\phi}}, \quad \forall h \in \mathbf{V}(\bar{\mathbf{X}}_t). \tag{1.8}$$

Proof Since the observation point process is only given on the subset $\mathbf{Z}_t$ of $\bar{\mathbf{Z}}_t$, it follows that the point $g \in \mathbf{V}(\bar{\mathbf{Z}}_t)$ at which the derivatives in (1.8) are considered should verify $g|_{\mathbf{Z}_t} = 0$ and $g(\phi) = 1$, that is $g = \mathbf{1}_\phi$. Now let $X \doteq \sum_{i=1}^{n} \delta_{x_i} \in \mathbf{N}(\bar{\mathbf{X}}_t)$ be a counting measure on $\bar{\mathbf{X}}_t$, let $A_1,\dots,A_{n_t'}$ be a given collection of disjoint Borel subsets of $\mathbf{Z}_t$, and $\mathscr{A}$ be a measurable subset of $\mathbf{N}(\bar{\mathbf{Z}}_t)$ defined as

$$\mathscr{A} \doteq \big\{ Z \in \mathbf{N}(\bar{\mathbf{Z}}_t) : \forall 1 \leq i \leq n_t' \big(Z(A_i) = 1\big), \quad Z(\mathbf{Z}_t \setminus (\cup_i A_i)) = 0 \big\}.$$

The measurable subset $\mathscr{A}$ contains the counting measures with exactly 1 point in each A_i, no point elsewhere in $\mathbf{Z}_t$, and any number of points on ϕ. Using the notations introduced in the proof of Proposition 2, we first compute the n_t'th-order functional derivative of $G_{\bar{\mathscr{Z}}_t}(g)(X)$ as follows:

$$\delta G_{\bar{\mathscr{Z}}_t}\big(g;\mathbf{1}_{A_1},\dots,\mathbf{1}_{A_{n_t'}}\big)(X)|_{g=\mathbf{1}_\phi} = n_t'! \Big[\prod_{i=1}^{n_t'} L_t(X_i, A_i) \Big] \Big[\prod_{j=n_t'+1}^{n} L_t(X_j, \{\phi\}) \Big],$$

whenever $n'_t \leq n$ and 0 otherwise. This last result can be rewritten as

$$\delta G_{\bar{\mathscr{Z}}_t}\big(g; \mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_{n'_t}}\big)(X)|_{g=\mathbf{1}_\phi} = n'_t!\, \mathbb{P}\big(\bar{\mathscr{Z}}_t \in \mathscr{A} | \bar{\mathscr{X}}_{t|t-1} = X\big).$$

Using Bayes' theorem, we obtain

$$\mathbb{E}\big(F(\bar{\mathscr{X}}_{t|t-1})\big|\bar{\mathscr{Z}}_t \in \mathscr{A}\big) = \frac{\mathbb{E}\big(F(\bar{\mathscr{X}}_{t|t-1})\mathbb{P}(\bar{\mathscr{Z}}_t \in \mathscr{A}\big|\bar{\mathscr{X}}_{t|t-1})\big)}{\mathbb{E}\big(\mathbb{P}(\bar{\mathscr{Z}}_t \in \mathscr{A}\big|\bar{\mathscr{X}}_{t|t-1})\big)},$$

which can be translated into a p.g.fl. form by setting $F(X) = \prod_{i=1}^n h(x_i)$. Finally, considering the Radon–Nikodym derivative of the Markov kernel L_t and a general conditioning on $\bar{\mathscr{Z}}_t$ proves the desired result.

We are now in position to provide a first proof for Theorem 1.

Proof of Theorem 1

We first need to find a more detailed expression of the p.g.fl. G_t by performing the functional differentiations in (1.8). The p.g.fl. H is rewritten more explicitly as

$$H(g, h) = \exp\Big(-\bar{\gamma}_{t|t-1}\big(1 - hL_t(g)\big)\Big).$$

By differentiating H in the directions $u_1, \ldots, u_{N'_t}$, we find that the numerator of G_t can be expressed as

$$\delta H\big(g, h; u_1, \ldots, u_{N'_t}\big)\big|_{g=\mathbf{1}_\phi} = H(\mathbf{1}_\phi, h)\prod_{i=1}^{N'_t}\bar{\gamma}_{t|t-1}(hL_t(u_i)).$$

An expression of G_t can be deduced as follows by considering that $u_i = \mathbf{1}_{Z_i}$, for all $1 \leq i \leq N'_t$,

$$G_t(h) = \exp\Big(-\bar{\gamma}_{t|t-1}\big((1-h)\ell_t(\cdot, \phi)\big)\Big)\prod_{i=1}^{N'_t}\frac{\bar{\gamma}_{t|t-1}(h\ell_t(\cdot, Z_i))}{\bar{\gamma}_{t|t-1}\big(\ell_t(\cdot, Z_i)\big)}.$$

The first-moment measure $\bar{\gamma}_t$ of the updated point process $\bar{\mathscr{X}}_t|\mathscr{Z}_{0:t}$ can then be found by differentiating G_t at point $\mathbf{1}$ in the direction $f \in \mathbf{B}(\bar{\mathbf{X}}_t)$,

$$\begin{aligned}\bar{\gamma}_t(f) &= \delta G_t(\mathbf{1}; f)\\ &= \bar{\gamma}_{t|t-1}\big(f\,\ell_t(\cdot, \phi)\big) + \int \mathscr{Z}_t(\mathrm{d}z)\frac{\bar{\gamma}_{t|t-1}(f\,\ell_t(\cdot, z))}{\bar{\gamma}_{t|t-1}\big(\ell_t(\cdot, z)\big)},\end{aligned}$$

which terminates the proof of the theorem.

As mentioned above, this p.g.fl.-based approach is the one that was originally used in [12], where the PHD filter was first introduced. However, it might prove challenging to understand the structure of the result via this *p.g.fl. transform domain*. A *direct* proof, which only relies on the standard operations of probability theory, is presented in the next section.

1.5.2 Direct Proof

In order to shed light on the structure of the PHD filter, we study a direct proof of it, mostly based on [2]. The underlying idea is to find the conditional distribution of the predicted point process $\bar{\mathscr{X}}_{t|t-1}$ given the extended observation point process $\bar{\mathscr{Z}}_t$, denoted as $\mathbb{E}(F(\bar{\mathscr{X}}_{t|t-1})|\bar{\mathscr{Z}}_t)$ for any $F \in \mathbf{B}(\mathbf{M}(\bar{\mathbf{X}}_t))$. However, since only the observation point process $\mathscr{Z}_t$ is given, the objective is to find the conditional distribution $\mathbb{E}(F(\bar{\mathscr{X}}_{t|t-1})|\mathscr{Z}_t)$ which can be recovered from the former one through

$$\mathbb{E}\big(F(\bar{\mathscr{X}}_{t|t-1})|\mathscr{Z}_t\big) = \mathbb{E}\Big(\mathbb{E}\big(F(\bar{\mathscr{X}}_{t|t-1})|\bar{\mathscr{Z}}_t\big)|\mathscr{Z}_t\Big).$$

The first-moment measure $\bar{\gamma}_t$ of the updated point process $\bar{\mathscr{X}}_{t|t}$ can then be directly found through the following relation

$$\bar{\gamma}_t(f) = \mathbb{E}\big(\bar{\mathscr{X}}_{t|t-1}(f)|\mathscr{Z}_t\big), \quad \forall f \in \mathbf{B}(\bar{\mathbf{X}}_t).$$

In the remainder of this section the first-moment measure and the common distribution of the points in the predicted point process $\bar{\mathscr{X}}_{t|t-1}$ will be, respectively, denoted as $\bar{\gamma} \in \mathbf{M}(\bar{\mathbf{X}}_t)$ and $\bar{\eta} \in \mathbf{P}(\bar{\mathbf{X}}_t)$ for the sake of compactness.

Lemma 2 *A version of the conditional distribution of the predicted point process $\bar{\mathscr{X}}_{t|t-1}$ given the extended observation point process $\bar{\mathscr{Z}}_t$ is expressed for any $F \in \mathbf{B}(\mathbf{M}(\bar{\mathbf{X}}_t))$ as*

$$\mathbb{E}(F(\bar{\mathscr{X}}_{t|t-1})|\bar{\mathscr{Z}}_t) = \int F\Big(\sum_{i=1}^{M'_t} \delta_{x_i}\Big) \prod_{i=1}^{M'_t} \Psi_{\ell_t(\cdot,Z_i)}(\bar{\eta})(\mathrm{d}x_i),$$

where the $\mathbb{N}$-valued random variable M'_t is defined as $M'_t \doteq N'_t + N_\phi$.

Proof We first observe that $\bar{\mathscr{X}}_{t|t-1}$ and $\bar{\mathscr{Z}}_t$ have almost surely the same total mass because of the one-to-one correspondence between objects and clutter generators on the one hand and (possibly empty) observations on the other hand. Then, we consider the point process $\mathscr{W}_t$ on the space $\bar{\mathbf{X}}_t \times \bar{\mathbf{Z}}_t$ that is characterised by

$$\mathscr{W}_t \doteq \sum_{i=1}^{M'_t} \delta_{(X_i,Z_i)},$$

where the $\mathbb{N}$-valued random variable M'_t is defined by the almost sure relation $M'_t \doteq N'_t + N_\phi = N_t + N_c$. The point process $\mathscr{W}_t$ is Poisson and its underlying spatial distribution $\hat{\eta} \in \mathbf{M}(\bar{\mathbf{X}}_t \times \bar{\mathbf{Z}}_t)$ can be expressed as

$$\hat{\eta}(\mathrm{d}(x,z)) = \bar{\eta}(\mathrm{d}x) L_t(x, \mathrm{d}z).$$

In order to express the common conditional distribution L'_t of the points in $\bar{\mathscr{X}}_{t|t-1}$ given a realisation of $\bar{\mathscr{Z}}_t$, we consider the following reversal formula:

$$\bar{\eta}(\mathrm{d}x) L_t(x, \mathrm{d}z) = \bar{\eta}\big(L_t(\cdot, \mathrm{d}z)\big) L'_t(z, \mathrm{d}x),$$

which can be expressed as Bayes' theorem using the Radon–Nikodym derivative $\ell_t(x, \cdot)$ of $L_t(x, \cdot)$ as

$$L'_t(z, \mathrm{d}x) = \frac{\bar{\eta}(\mathrm{d}x)\ell_t(x,z)}{\bar{\eta}\big(\ell_t(\cdot, z)\big)} = \Psi_{\ell_t(\cdot,z)}(\bar{\eta})(\mathrm{d}x), \quad \forall z \in \bar{\mathbf{Z}}_t.$$

Noting that any realisation of $\bar{\mathscr{Z}}_t$ gives away the total mass of the point process $\bar{\mathscr{X}}_{t|t-1}$, the result of the lemma is obtained easily.

Recalling that the extended point process $\bar{\mathscr{Z}}_t$ can be divided into the point process $\mathscr{Z}_t$ on $\mathbf{Z}_t$ and the atom $N_\phi \delta_\phi$, we obtain the following corollary from Lemma 2.

Corollary 1 *The conditional distribution* $\mathbb{E}(F(\bar{\mathscr{X}}_{t|t-1})|\bar{\mathscr{Z}}_t)$ *can be expressed for any* $F \in \mathbf{B}(\mathbf{M}(\bar{\mathbf{X}}_t))$ *as*

$$\mathbb{E}\big(F(\bar{\mathscr{X}}_{t|t-1})|\bar{\mathscr{Z}}_t\big) = \int F\Big(\sum_{i=1}^{N'_t} \delta_{x_i} + \sum_{j=1}^{N_\phi} \delta_{x'_j}\Big)\Big[\prod_{i=1}^{N'_t} \Psi_{\ell_t(\cdot,Z_i)}(\bar{\eta})(\mathrm{d}x_i)\Big]\Big[\prod_{j=1}^{N_\phi} \Psi_{\ell_t(\cdot,\phi)}(\bar{\eta})(\mathrm{d}x'_j)\Big].$$

The more detailed expression of Corollary 2 allows for integrating in a more direct way the fact that the component $N_\phi \delta_\phi$ of the point process $\bar{\mathscr{Z}}_t$ is not actually observed.

Lemma 3 *A version of the conditional distribution of the predicted point process* $\bar{\mathscr{X}}_{t|t-1}$ *given the observation point process* $\mathscr{Z}_t$ *is expressed, for any measurable function* $F \in \mathbf{B}(\mathbf{M}(\bar{\mathbf{X}}_t))$*, as*

$$\mathbb{E}(F(\bar{\mathscr{X}}_{t|t-1})|\mathscr{Z}_t) = \exp\Big(-\bar{\gamma}\big(\ell_t(\cdot,\phi)\big)\Big)\sum_{k\geq 0}\frac{\bar{\gamma}(\ell_t(\cdot,\phi))^k}{k!} \times \int F\Big(\sum_{i=1}^{N'_t}\delta_{x_i} + \sum_{j=1}^{k}\delta_{x'_j}\Big)\Big[\prod_{i=1}^{N'_t}\Psi_{\ell_t(\cdot,Z_i)}(\bar{\eta})(\mathrm{d}x_i)\Big]\Big[\prod_{j=1}^{k}\Psi_{\ell_t(\cdot,\phi)}(\bar{\eta})(\mathrm{d}x'_j)\Big].$$

Proof To prove the lemma, it is sufficient to note that a version of the conditional distribution of $\mathscr{Z}_t^{\bar{}}$ given the observation point process $\mathscr{Z}_t$ can be expressed, for any measurable function $F' \in \mathbf{B}(\mathbf{M}(\bar{\mathbf{Z}}_t))$, as

$$\mathbb{E}\big(F'(\mathscr{Z}_t^{\bar{}})|\mathscr{Z}_t\big) = \exp\Big(-\bar{\gamma}\big(\ell_t(\cdot,\phi)\big)\Big)\sum_{k\geq 0}\frac{\bar{\gamma}(\ell_t(\cdot,\phi))^k}{k!}F'(\mathscr{Z}_t + k\delta_\phi),$$

and then consider $F'(\mathscr{Z}_t^{\bar{}}) = \mathbb{E}(F(\mathscr{X}_{t|t-1}^{\bar{}})|\mathscr{Z}_t^{\bar{}})$.

We are now in position to propose a second proof for Theorem 1.

Proof (Theorem 1) As mentioned above, we compute the first-moment measure $\bar{\gamma}_t$ of the updated point process $\mathscr{X}_{t|t}^{\bar{}}$ from the conditional distribution $\mathbb{E}(F(\mathscr{X}_{t|t-1}^{\bar{}})|\mathscr{Z}_t^{\bar{}})$ via equation

$$\bar{\gamma}_t(f) = \mathbb{E}\big(\mathscr{X}_{t|t-1}^{\bar{}}(f)|\mathscr{Z}_t\big), \quad \forall f \in \mathbf{B}(\bar{\mathbf{X}}_t).$$

Considering the expression of the conditional distribution of $\mathscr{X}_{t|t-1}^{\bar{}}$ given in Lemma 3 and the fact that measures transformed via $\Psi_{\ell_t(\cdot,z)}$ are probability measures, we find that

$$\begin{aligned}\mathbb{E}\big(\mathscr{X}_{t|t-1}^{\bar{}}(f)|\mathscr{Z}_t\big) = \exp\Big(-\bar{\gamma}\big(\ell_t(\cdot,\phi)\big)\Big)\sum_{k\geq 0}\frac{\bar{\gamma}(\ell_t(\cdot,\phi))^k}{k!}\\ \times\left[\sum_{i=1}^{N_t'}\Psi_{\ell_t(\cdot,Z_i)}(\bar{\eta})(f) + k\Psi_{\ell_t(\cdot,\phi)}(\bar{\eta})(f)\right].\end{aligned}$$

Then, noting that the Boltzmann–Gibbs transformation is left invariant under rescaling of the input measure, i.e. that $\Psi_{\ell_t(\cdot,z)}(\bar{\eta}) = \Psi_{\ell_t(\cdot,z)}(\bar{\gamma})$ for any $z \in \bar{\mathbf{Z}}_t$, and that

$$\exp\Big(-\bar{\gamma}\big(\ell_t(\cdot,\phi)\big)\Big)\sum_{k\geq 0}\frac{\bar{\gamma}(\ell_t(\cdot,\phi))^k}{k!} = 1,$$

we obtain

$$\mathbb{E}\big(\mathscr{X}_{t|t-1}^{\bar{}}(f)|\mathscr{Z}_t\big) = \bar{\gamma}\big(\ell_t(\cdot,\phi)\big)(f) + \sum_{i=1}^{N_t'}\Psi_{\ell_t(\cdot,Z_i)}(\bar{\gamma})(f),$$

which terminates the proof.

1.5.3 Proof of Theorem 2

One way of proving Theorem 2 is to use an induction approach together with the filtering equations already demonstrated in the previous sections. To that purpose, we assume that the first-moment measure $\gamma_{t|t-1}$ of the point process $\mathscr{X}_t|\mathscr{Z}_{0:t-1}$ can be expressed as

$$\gamma_{t|t-1}(B) = \int \alpha_{t-1}(\mathrm{d}\boldsymbol{y})\, p^{(\boldsymbol{y})}_{t|t-1}(B), \quad \forall B \in \mathscr{B}(\mathbf{X}_t),$$

where $p^{(\boldsymbol{y})}_{t|t-1} \in \mathbf{P}(\mathbf{X}_t)$ for any $\boldsymbol{y} \in \mathbf{Y}_t$ and where $p^{(\boldsymbol{y})}_{t|t-1} \in \mathbf{M}(\mathbf{X}_t)$ when $\boldsymbol{y} = \boldsymbol{\phi}_t$. In order to show the result, we need to prove that after augmenting $\gamma_{t|t-1}$ through Ξ_t as well as updating and predicting, the obtained first-moment measure can still be expressed in a similar form.

First we observe, for any $B \in \mathscr{B}(\bar{\mathbf{X}}_t)$, that

$$\begin{aligned}\bar{\gamma}_{t|t-1}(B) &= \Xi_t\big(\gamma_{t|t-1}\big)(B) \\ &= \bar{\alpha}_{t-1}(\{\boldsymbol{\phi}_t\})\bar{p}^{(\boldsymbol{\phi}_t)}_{t|t-1}(B) + \int_{\mathbf{Y}_t} \alpha_{t-1}(\mathrm{d}\boldsymbol{y})\, p^{(\boldsymbol{y})}_{t|t-1}(B) \\ &= \int \bar{\alpha}_{t-1}(\mathrm{d}\boldsymbol{y})\, \bar{p}^{(\boldsymbol{y})}_{t|t-1}(B).\end{aligned}$$

Then, using the update Eq. (1.2), we find that

$$\begin{aligned}\bar{\gamma}_t(B) &= \int_B g_{t,\bar{\gamma}_{t|t-1}}(x)\bar{\gamma}_{t|t-1}(\mathrm{d}x) \\ &= \int \bar{\alpha}_{t-1}(\mathrm{d}\boldsymbol{y})\left[\bar{p}^{(\boldsymbol{y})}_{t|t-1}(\mathbf{1}_B \ell_t(\cdot,\phi)) + \int \mathscr{Z}_t(\mathrm{d}z)\frac{\bar{p}^{(\boldsymbol{y})}_{t|t-1}(\mathbf{1}_B \ell_t(\cdot,z))}{\bar{\alpha}_{t-1}\big(\bar{p}^{(\cdot)}_{t|t-1}(\ell_t(\cdot,z))\big)}\right].\end{aligned}$$

Finally, using the prediction Eq. (1.4), we can conclude that

$$\bar{\gamma}_{t+1|t}(B) = \int \bar{\alpha}_{t-1}(\mathrm{d}\boldsymbol{y})\zeta(\mathrm{d}z)\, p^{(\boldsymbol{y},z)}_{t+1|t}(B), \quad \forall B \in \mathscr{B}(\mathbf{X}_{t+1}),$$

with

$$p^{(\boldsymbol{y},z)}_{t+1|t}(B) = \frac{\bar{p}^{(\boldsymbol{y})}_{t|t-1}\big(M_t(\cdot,B)\ell_t(\cdot,z)\big)}{\bar{p}^{(\boldsymbol{y})}_{t|t-1}\big(M_t(\cdot,\mathbf{X}_{t+1})\ell_t(\cdot,z)\big)} \in \mathbf{P}(\mathbf{X}_{t+1}),$$

and

$$\zeta(\mathrm{d}z) = \delta_\phi(\mathrm{d}z)\bar{p}^{(\boldsymbol{y})}_{t|t-1}\big(M_t(\cdot,\mathbf{X}_{t+1})\ell_t(\cdot,\phi)\big) + \mathscr{Z}_t(\mathrm{d}z)\frac{\bar{p}^{(\boldsymbol{y})}_{t|t-1}\big(M_t(\cdot,\mathbf{X}_{t+1})\ell_t(\cdot,z)\big)}{\bar{\alpha}_{t-1}\big(\bar{p}^{(\cdot)}_{t|t-1}(\ell_t(\cdot,z))\big)}.$$

The expression obtained in Theorem 2 can be directly deduced from this result by recalling that $s_t(x) = M_t(x, \mathbf{X}_t) \in 0, 1$ is the probability of survival at point $x \in \bar{\mathbf{X}}_t$.

1.6 Conclusion

The introduction of the concept of association measure in the formulation of the PHD filter enabled the expression of the first-moment measure of the distribution related to a multi-object system as a mixture of single-object posterior laws, thus highlighting the structure of the corresponding updating-prediction equation. This approach shows the PHD filter from another point of view, that is, as single-object filters in interaction or, in the linear Gaussian case, as interacting Kalman filters. This formulation enabled the derivation of a mean field approximation of the first-moment measure that does not depend on the resolution of the single-object filtering problem, hence showing a certain versatility when compared to the usual Gaussian mixture and sequential Monte Carlo implementations of the PHD filter. This approach also makes different techniques available at the object level, such as objectwise forward–backward algorithms or parameter estimation.

References

1. Blackman, S.S.: Multiple-Target Tracking with Radar Applications, vol. 463 p. 1. Artech House, Inc., Dedham, MA (1986)
2. Caron, F., Del Moral, P., Doucet, A., Pace, M.: On the conditional distributions of spatial point processes. Adv. Appl. Probab. **43**(2), 301–307 (2011)
3. Clark, D.E.: First-moment multi-object forward-backward smoothing. In: 2010 13th Conference on Information Fusion (FUSION), IEEE (2010)
4. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes, vol. II. Springer, New York (2008)
5. Del Moral, P.: Feynman-Kac Formulae. Springer, Berlin (2004)
6. Del Moral, P.: Mean field simulation for Monte Carlo integration. Chapman and Hall/CRC Monographs on Statistics and Applied Probability (2013)
7. Fortmann, T.E., Bar-Shalom, Y., Scheffe, M.: Sonar tracking of multiple targets using joint probabilistic data association. IEEE J. Oceanic Eng. **8**(3), 173–184 (1983)
8. Goodman, I.R., Mahler, R.P.S., Nguyen, H.T.: Mathematics of Data Fusion, vol. 37. Springer Science & Business Media (1997)
9. Houssineau, J., Del Moral, P., Clark, D.E.: General multi-object filtering and association measure. In: 2013 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), (2013)
10. Jiang, L., Singh, S.S., Yıldırım, S.: arXiv preprint arXiv:1410.2046 (2014)
11. Mahler, R.P.S.: An Introduction to Multisource-Multitarget Statistics and Applications. Lockheed Martin (2000)
12. Mahler, R.P.S.: Multitarget Bayes filtering via first-order multitarget moments. IEEE Trans. Aerosp. Electron. Syst. **39**(4), 1152–1178 (2003)
13. Mahler, R.P.S.: PHD filters of higher order in target number. IEEE Trans. Aerosp. Electron. Syst. **43**(4), 1523–1543 (2007)

14. Mahler, R.P.S.: Statistical Multisource-Multitarget Information Fusion. Artech House, Boston (2007)
15. Mahler, R.P.S., Vo, B.T., Vo, B.N.: Forward-backward probability hypothesis density smoothing. IEEE Trans. Aerosp. Electron. Syst. **48**(1), 707–728 (2012)
16. Pace, M., Del Moral, P.: Mean-field PHD filters based on generalized Feynman-Kac flow. J. Sel. Top. Signal Process. Special issue on multi-target tracking (2013)
17. Panta, K., Clark, D.E., Vo, B.N.: Data association and track management for the Gaussian mixture probability hypothesis density filter. IEEE Trans. Aerosp. Electron. Syst. **45**(3), 1003–1016 (2009)
18. Ristic, B., Clark, D.: Particle filter for joint estimation of multi-object dynamic state and multi-sensor bias. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3877–3880. IEEE (2012)
19. Ristic, B., Clark, D., Vo, B.N.: Improved SMC implementation of the PHD filter. In: 2010 13th Conference on Information Fusion (FUSION), pp. 1–8. IEEE (2010)
20. Singh, S.S., Vo, B.N., Baddeley, A., Zuyev, S.: Filters for spatial point processes. SIAM J. Control Optim. **48**(4), 2275–2295 (2009)
21. Vo, B.N., Ma, W.K.: The Gaussian mixture probability hypothesis density filter. IEEE Trans. Signal Process. **54**(11), 4091–4104 (2006)
22. Vo, B.N., Singh, S., Doucet, A.: Sequential Monte Carlo methods for multitarget filtering with random finite sets. IEEE Trans. Aerosp. Electron. Syst. **41**(4), 1224–1245 (2005)
23. Vo, B.T., Vo, B.N., Cantoni, A.: Analytic implementations of the cardinalized probability hypothesis density filter. IEEE Trans. Signal Process. **55**(7), 3553–3567 (2007)
24. Vu, T., Vo, B.N., Evans, R.: A particle marginal Metropolis-Hastings multi-target tracker. IEEE Trans. Signal Process. **62**(15), 3953–3964 (2014)

Chapter 2
An Overview of Recent Advances in Monte-Carlo Methods for Bayesian Filtering in High-Dimensional Spaces

François Septier and Gareth W. Peters

Abstract Nonlinear non-Gaussian state-space models arise in numerous applications in statistics and signal processing. In this context, one of the most successful and popular approximation techniques is the sequential Monte-Carlo (SMC) algorithm, also known as the particle filter. Nevertheless, this method tends to be inefficient when applied to high-dimensional problems. In this chapter, we present, an overview of recent contributions related to Monte-Carlo methods for sequential simulation from ultra high-dimensional distributions, often arising for instance in Bayesian applications.

2.1 Introduction

In many applications, we are interested in estimating a signal from a sequence of noisy observations. Optimal filtering techniques for general nonlinear and non-Gaussian state-space models are consequently of great interest. Except in a few special cases, including linear and Gaussian state-space models (Kalman filter [26]) and hidden finite-state space Markov chains [7], it is impossible to evaluate this filtering distribution analytically. However, linear systems with Gaussian dynamics are generally inappropriate for the accurate modeling of a dynamical system, since they fail to account for the local nonlinearities in the state space or the dynamic changing nature of the system which is under study. It is therefore increasingly common to consider nonlinear or non-Gaussian dynamical systems. In the case of additive Gaussian errors, one could adopt an Extended Kalman filter (EKF) or in the case of non-Gaussian additive errors, an Unscented Kalman filter (UKF) [25].

F. Septier (✉)
Institut Mines-Télécom/Télécom Lille/CRIStAL UMR 9189, Villeneuve D'ascq, France
e-mail: francois.septier@telecom-lille.fr

G.W. Peters
Department of Statistical Science, University College London, London, UK
e-mail: gareth.peters@ucl.ac.uk

G.W. Peters and T. Matsui (eds.), *Theoretical Aspects of Spatial-Temporal Modeling*,
JSS Research Series in Statistics, DOI 10.1007/978-4-431-55336-6_2

Since the 1990s, sequential Monte Carlo (SMC) approaches have become a powerful methodology to cope with nonlinear and non-Gaussian problems [16]. In comparison with standard approximation methods, such as the EKF, the principal advantage of SMC methods is that they do not rely on any local linearization technique or any crude functional approximation. These particle filtering (PF) methods [23], exploit numerical representation techniques for approximating the filtering probability density function of inherently nonlinear non-Gaussian systems. Using these methods for the empirical characterization of sequences of distributions and the resulting estimators formed based on these empirical estimates can be set arbitrarily close to the optimal solution at the expense of computational complexity.

However, due to their importance sampling-based design, classical SMC methods tend to be inefficient when applied to high-dimensional problems [36, 42]. This issue, known as the curse of dimensionality, has rendered traditional SMC algorithms largely useless in high-dimensional applications such as multiple target tracking, weather prediction, and oceanography. In this chapter, we aim at reviewing recent developments in Monte-Carlo-based techniques that have been specifically designed to deal with high-dimensional systems. The chapter is organized as follows. In Sect. 2.2, we describe the model and the different quantities of interest in dynamic settings. Then, Sect. 2.3 discusses the general principle of SMC methods and their limitations in high-dimensional systems. Several recent developments to improve their performance in this specific setting are then presented. Section 2.4 describes another class of sequential inference algorithms based on the use of Markov chain Monte-Carlo methods (SMCMC) as an alternative to SMC methods. Numerical results are shown in Sect. 2.6. Conclusions are given in Sect. 2.7.

2.2 Problem Formulation

A *hidden Markov model* (HMM) corresponds to a $\mathbb{R}^d$-valued discrete-time Markov process, $\{X_n\}_{n\geq 1}$ that is not directly observable but we have only access to another $\mathbb{R}^{d_y}$-valued discrete-time stochastic process, $\{Y_n\}_{n\geq 1}$, which is linked to the *hidden* Markov process of interest. Owing to the Markovian property of the process, the joint distribution of the process $\{X_n\}_{n\geq 1}$,

$$p(x_{1:n}) = \mu(x_1)\prod_{k=1}^{n} f_k(x_k|x_{k-1}) \tag{2.1}$$

is completely defined by an initial probability density function (pdf) $\mu(x_1)$ and the transition density function at any time k, denoted by $f_k(x_k|x_{k-1})$.

In a HMM, the observed process $\{Y_n\}_{n\geq 1}$ is such that the conditional joint density of $Y_{1:n} = y_{1:n}$ given $X_{1:n} = x_{1:n}$ has the following conditional independence (product) form

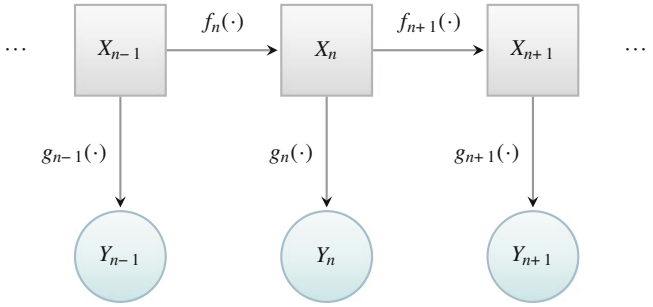


Fig. 2.1 Graphical representation of a hidden Markov model

$$p(y_{1:n}|x_{1:n}) = \prod_{k=1}^{n} g_k(y_k|x_k). \tag{2.2}$$

The dependence structure of an HMM can be represented by a graphical model shown in Fig. 2.1.

Equations (2.1)–(2.2) define a Bayesian model where (2.1) defines the prior distribution of the "state" process of interest $\{X_n\}_{n\geq 1}$ and (2.2) defines the likelihood function of the conditional observations. One of the most common inference problems, known as *optimal filtering*, which occurs with HMMs is the estimation of the current state value based upon the sequence of observations observed so far. Such inference about X_n given a sequence of the observations $Y_{1:n} = y_{1:n}$ relies upon the posterior distribution,

$$p(x_{1:n}|y_{1:n}) = \frac{p(x_{1:n}, y_{1:n})}{p(y_{1:n})} = \frac{p(x_{1:n})p(y_{1:n}|x_{1:n})}{p(y_{1:n})}. \tag{2.3}$$

This posterior distribution, known also as the *smoothing* distribution, satisfies the following recursion

$$p(x_{1:n}|y_{1:n}) = \frac{g_n(y_n|x_n)\, f_n(x_n|x_{n-1})}{p(y_n|y_{1:n-1})} p(x_{1:n-1}|y_{1:n-1}), \tag{2.4}$$

where

$$p(y_n|y_{1:n-1}) = \int g_n(y_n|x_n)\, f_n(x_n|x_{n-1}) p(x_{n-1}|y_{1:n-1}) dx_{n-1:n}. \tag{2.5}$$

In the literature, this recursion is sometimes presented directly in terms of the marginal posterior distribution, $p(x_n|y_{1:n})$, known as the *filtering* distribution:

$$p(x_n|y_{1:n}) = \frac{g_n(y_n|x_n)p(x_n|y_{1:n-1})}{p(y_n|y_{1:n-1})}, \tag{2.6}$$

with

$$p(x_n|y_{1:n-1}) = \int f_n(x_n|x_{n-1})\, p(x_{n-1}|y_{1:n-1}) dx_{n-1}. \tag{2.7}$$

However, most sequential Monte-Carlo-based algorithms rely on a numerical approximation of recursion (2.4) instead of (2.6).

2.3 Sequential Monte-Carlo Methods

2.3.1 General Methodology

SMC methods have several variants sometimes appearing under the names of particle filtering or interacting particle systems, e.g. [10, 15, 37], and their theoretical properties have been extensively studied in [8–10, 29].

The general context of a standard SMC method is that one wants to approximate a (often naturally occurring) sequence of target probability density functions (pdf) $\{\pi_n(x_{1:n})\}_{n\geq 1}$ of increasing dimension, i.e. the support of every function in this sequence is defined as $\text{supp}(\pi_n) = \mathbb{R}^{dn}$ and therefore the dimension of its support forms an increasing sequence with n. We may also assume that π_n is only known up to a normalizing constant,

$$\pi_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{Z_n}. \tag{2.8}$$

SMC methods firstly provide an approximation of $\pi_1(x_1)$ and an unbiased estimate of Z_1, then at the second iteration ("time step" 2) once a new observation is received, an approximation of $\pi_2(x_{1:2})$ is formed as well as an unbiased estimate of Z_2 and this repeats with each distribution in the sequence.

Let us remark at this stage that SMC methods can be used for any sequence of target distributions and therefore the application of SMC to optimal filtering, known as *particle filtering*, is just a special case of this general methodology by choosing $\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n})$ and $Z_n = p(y_{1:n})$.

Procedurally, we initialize the algorithm by sampling a set of N particles, $\left\{X_1^j\right\}_{j=1}^N$, from the distribution π_1 and set the normalized weights as $W_1^j = 1/N$, for all $j = 1, \ldots, N$. If it is not possible to sample directly from π_1, one should sample from an importance distribution q_1 and calculate its weights according to the importance sampling principle, i.e. $W_1^j \propto \pi_1(X_1^j)/q_1(X_1^j)$. Then, the particles are sequentially propagated thorough each distribution π_t in the sequence via two main processes: mutation and correction (incremental importance weighting). In the first step (mutation), we propagate particles from time $t-1$ to time t and in the second one (correction) we calculate the new importance weights of the particles.

This method can be seen as a sequence of *importance sampling* steps, where the target distribution at each step n is $\pi_n(x_{1:n})$ and the importance distribution is given by

$$q_n(x_{1:n}) = q_1(x_1) \prod_{k=2}^{n} q_k(x_k|x_{1:k-1}), \tag{2.9}$$

where $q_k(x_k|x_{1:k-1})$ is the proposal distribution used to propagate particles from time $k-1$ to k. As a consequence, the unnormalized importance weights are computed recursively by:

$$\begin{aligned} W(x_{1:n}) &= \frac{\gamma_n(x_{1:n})}{q_n(x_{1:n})} \\ &= \frac{\gamma_{n-1}(x_{1:n-1})}{q_{n-1}(x_{1:n-1})} \frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1}) q_n(x_n|x_{1:n-1})} \\ &= W(x_{1:n-1}) \tilde{w}(x_{1:n}), \end{aligned} \tag{2.10}$$

where $\tilde{w}(x_{1:n})$ is known as the *incremental importance weight*. When SMC is applied for the optimal filtering problem with $\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n})$, it is straightforward to show by using the recursion of the smoothing distribution in Eq. (2.4) that the incremental importance weight is given by:

$$\tilde{w}(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1}) q_n(x_n|x_{1:n-1})} = \frac{g_n(y_n|x_n) f_n(x_n|x_{n-1})}{q_n(x_n|x_{1:n-1})}. \tag{2.11}$$

At any time n, we obtain an approximation of the target distribution via the empirical measure obtained by the collection of weighted samples, i.e.

$$\widehat{\pi}_n(x_{1:n}) = \sum_{j=1}^{N} W_n^j \delta_{X_{1:n}^j}(dx_{1:n}), \tag{2.12}$$

where W_n^j is the normalized importance weights such that $\sum_{j=1}^{N} W_n^j = 1$. Moreover, an unbiased estimate of the ratio of two successive normalizing constants is also provided as follows:

$$\widehat{\frac{Z_n}{Z_{n-1}}} = \sum_{j=1}^{N} W_{n-1}^j \tilde{w}(X_{1:n}^j). \tag{2.13}$$

The algorithm described above is known as the *Sequential Importance Sampling* (SIS) algorithm. However, direct importance sampling on a very large space is rarely efficient as the importance weights exhibit very high variance. As a consequence, SIS will provide estimates whose variance increases exponentially with time n. Indeed, after only a few iterations, all but a few particles will have negligible weights thus leading to the phenomena known as *weight degeneracy*. A well-known criterion to

quantify in an online manner this degeneracy is the *effective sample size* defined as follows:

$$\mathrm{ESS}_n = \frac{1}{\sum_{j=1}^{N} \left(W_n^i\right)^2} \tag{2.14}$$

with $1 \leq \mathrm{ESS}_n \leq N$. In order to overcome this degeneracy problem, a resampling step is thus added in the basic algorithm when the effective sample size drops below some threshold, which as a rough guide is typically in the range of 30–60 % of the total number of particles. The purpose of resampling is to reduce this degeneracy by eliminating samples which have low importance weights and duplicating samples with large importance weights [15]. It is quite obvious that when one is interested in the filtering distribution $p(x_n|y_{1:n})$, performing a resampling step at the previous time step will lead to a better level of sample diversity as those particles which were already extremely improbable at time $n-1$ are likely to have been eliminated and those which remain have a better chance of representing the situation at time n accurately. Unfortunately, when the smoothing distribution is really the quantity of interest, it is more problematic since the resampling mechanism eliminates some trajectories with every iteration, thus leading to problem known as *path or sample degeneracy*. Indeed, resampling will reduce at every iteration the number of distinct samples representing the first time instant of the hidden Markov process. Since in filtering applications, one is generally only interested in the final filtering posterior distribution, this resampling step is widely used in practice at the expense of further diminishing the quality of the path samples. Some strategy that will be discussed in Sect. 2.3.3.1 is generally employed in practice to increase the diversity of the samples.

This SMC algorithm which incorporates a resampling step is often referred to as *Sequential Importance Resampling* (SIR) or *Sequential Importance Sampling and Resampling* (SIS-R). This approach applied for filtering is summarized in Algorithm 1. By assuming that the cost of computing the product of the prior and the likelihood distribution is $\mathcal{O}(d)$ (i.e., a function of the dimension of the hidden state), the cost of the general SMC algorithm is $\mathcal{O}(nNd)$.

2.3.2 Limitations of SMC Methods

In this section, we will discuss the limitations of SMC methods when applied to high-dimensional problems. The main reason why the SIR algorithm performs poorly when the model dimension is high is essentially the same reason why the SIS algorithm behaves badly when the time-horizon is large, and it has to do with the fact that the importance sampling paradigm is typically very inefficient in high-dimensional models. As discussed previously, the SIS algorithm is designed to approximate the smoothing distribution $p(x_{1:n}|y_{1:n})$, weight degeneracy occurs as n increases since the dimension of this target distribution increases with time. Now, if the hidden Markov process is high-dimensional, weight degeneracy will occur as the dimension

Algorithm 1 SMC algorithm for optimal filtering

if time $n = 1$ **then**
  Sample $X_1^j \sim q_1(x_1), \forall j = 1, \ldots, N$
  Calculate the weights $W_1^j \propto \dfrac{g_1(Y_1|X_1^j)\mu(X_1^j)}{q_1(X_1^{(j)})}, \forall j = 1, \ldots, N$
else if time $n \geq 2$ **then**
  Sample $X_n^j \sim q_n(x_n|X_{1:n-1}^j)$ and set $X_{1:n}^j := (X_{1:n-1}^j, X_n^j), \forall j = 1, \ldots, N$
  Calculate the weights $W_n^j \propto W_{n-1}^j \dfrac{g_n(Y_n|X_n^j) f_n(X_n^j|X_{n-1}^j)}{q_n(X_n^j|X_{1:n-1}^j)}, \forall j = 1, \ldots, N$
end if
if $ESS_n < \Gamma$ **then**
  Resample $\left\{W_n^j, X_{1:n}^j\right\}$ to obtain N equally weighted particles $\left\{W_n^j = 1/N, X_{1:n}^j\right\}$
end if
Output: Approximation of the smoothing distribution via the following empirical measure:

$$\pi(x_{1:n}) \approx \sum_{j=1}^{N} W_n^j \delta_{X_{1:n}^j}(dx_{1:n})$$

of this process increases. As a consequence, this degeneracy is seen even in a single iteration of the algorithm. In [4, 42], a careful analysis shows that the collapse phenomenon occurs unless the sample size N is taken to be exponential in the dimension, which provides a rigorous statement of the curse of dimensionality. Let us remark that a similar weight degeneracy phenomena could be observed in SMC, even in low dimensional models, when for example the noise driving both the dynamics and the observation has very small variance.

The performance of the SMC strongly depends on the choice of the importance distribution. In the literature, the "optimal" proposal distribution in the sense of minimizing the variance of the importance weights is defined as:

$$\begin{aligned} q_n(x_n|x_{n-1}) &= \pi_n(x_n|x_{1:n-1}) \\ &= p(x_n|y_n, x_{n-1}) \quad (\textit{in HMM filtering problems}) \end{aligned} \tag{2.15}$$

which leads to the following incremental weight $\tilde{w}(x_{1:n}) = p(y_n|x_{n-1})$ whose variance conditional upon $x_{1:n-1}$ is zero since it is independent of x_n. Unfortunately, in many scenarios, it is impossible to sample from this "optimal" distribution. Many techniques have been proposed to design "efficient" importance distributions $q_n(x_n|x_{n-1})$ which approximate $p(x_n|y_n, x_{n-1})$. In particular, approximations based on the Extended Kalman Filter or the Unscented Kalman Filter to obtain importance distributions are very popular in the literature [6].

While the practical performance of the SIR algorithm can be largely improved by working with importance distributions that are tailored to the specific model being investigated, the benefit is limited to reducing the constants sitting in front of the error bounds, and this technique does not provide a fundamental solution to the curse of

dimensionality [35, 41]. When the optimal importance distribution is used, the curse of dimensionality would indeed still arise due to the recursive nature of the filtering problem.

In the next section, we will describe several strategies that have been proposed in order to improve the performance of standard particle filter in high-dimensional systems.

2.3.3 SMC Strategies for High-Dimensional Systems

2.3.3.1 MCMC Moves and the Use of Bridging Densities

The use of Markov Chain Monte-Carlo (MCMC) algorithms within SMC methods is a well-known strategy to improve the filter performance. As discussed previously, repeated resampling stages progressively impoverish the set of particles, by decreasing the number of distinct values represented in that set. This degeneracy problem has historically been addressed using the resample-move algorithm [20] which consists in applying one or more times after the resampling stage an MCMC transition kernel, $\mathscr{K}_n(x_{1:n}, x'_{1:n})$, such as a Gibbs sampler or Metropolis–Hastings scheme [38], having $\pi(x_{1:n})$ as its stationary distribution which means that the following property holds:

$$\int \pi(x_{1:n})\mathscr{K}_n(x_{1:n}, x'_{1:n})dx_{1:n} = \pi(x'_{1:n}). \tag{2.16}$$

As a consequence, if the particles $X^j_{1:n}$ are truly drawn from $\pi(x_{1:n})$, then the Markov kernel applied to any of the particles will simply generate new state sequences which are also drawn from the desired distribution. Moreover, even if the particles are not accurately drawn from $\pi(x_{1:n})$, the use of such Markov transition kernel will move the particles so that their distribution is closer to the target one (in total variation norm). The use of such MCMC moves can therefore be very effective in reducing the path degeneracy as well as in improving the accuracy of the empirical measure of the posterior distribution. In practice for filtering problems, in order to keep a truly online algorithm with a computational cost linear in time, the Markov transition kernels will not operate on the entire state history, but rather on some fixed time lag L by only updating the variables $X_{n-L+1:n}$.

An interesting generalization of the combination of SMC and MCMC has been proposed in [21] in which the authors propose to introduce a sequence of bridging densities between the initial sampling distribution (generally, the predictive posterior distribution, i.e., $p(x_{0:n}|y_{0:n-1})$) and the posterior at time n. By introducing gradually the effect of the likelihood function, the MCMC sampler is thus expected to converge faster, especially when the likelihood for the new data point is centered far from the points sampled from the importance distribution. As a consequence, such strategy could be more effective than standard SMC techniques in high-dimensional

problems. More specifically, the following sequence of $M \geq 1$ bridging densities is introduced at time n:

$$\pi_m(x_{1:n}) \propto p(x_{1:n-1}|y_{1:n-1}) f_n(x_n|x_{n-1}) g_n(y_n|x_n)^{\alpha_m} \tag{2.17}$$

with $0 \leq \alpha_1 < \cdots < \alpha_M = 1$. In order to move the particles through this sequence of bridging densities, the authors propose to use the framework of the annealed importance sampling [33] (or its generalization the sequential Monte-Carlo sampler [11, 34]). At time n, the particles are first propagated like in the standard SMC methods using an importance distribution, $q_n(x_n|x_{1:n-1})$, and let us denote them by $X^j_{0:n,0}$ and set $W^j_{n,0} = W^j_{n-1}$ and $\pi_0(x_{1:n}) = q_n(x_n|x_{1:n-1})\pi(x_{1:n-1})$. Then, at step $m = 1, \ldots, M$, the importance weights are computed as follows:

$$W^j_{n,m} \propto W^j_{n,m-1} \frac{\pi_m(X^j_{1:n,m-1})}{\pi_{m-1}(x^j_{1:n,m-1})}. \tag{2.18}$$

Then, a resampling step can be performed if the weights are too degenerate. Finally, each particle is moved independently to obtain $X^j_{1:n,m}$ using a Markov transition kernel, $\mathscr{K}_m(x^j_{1:n,m-1}, \cdot)$ having $\pi_m(x_{1:n})$ as stationary distribution. After the M steps, a set of weighted samples from the posterior distribution $\pi(x_{0:n})$ is therefore obtained by setting $\left\{W^j_n, X^j_{0:n}\right\} = \left\{W^j_{n,M}, X^j_{0:n,M}\right\}$. The algorithm is summarized in Algorithm 2 and its cost is $\mathscr{O}(nNMd)$ by assuming that the cost of computing the product of the prior and the likelihood distribution is $\mathscr{O}(d)$ as well as the MCMC kernel used. Let us notice that the resample-move algorithm [20] is a special case when $M = 1$ and also that this scheme is similar to some strategies known as annealed particle filtering [12, 17].

2.3.3.2 Local SMC Methods or Block Particle Filter

The underlying idea of these local SMC methods is to partition the state space into separate subspaces of small dimensions and run one SMC algorithm on each subspace. Such strategies have been developed in [13, 14, 32, 36]. Generally, the common assumption used in these approaches is that there exists an ensemble of disjoint sets $\left\{D_{n,j}\right\}_{j=1}^{B_n}$ with $\cup_{j=1}^{B_n} D_{n,j} = \{1 : d\}$ and $D_{n,j} \cap D_{n,i} = \emptyset$ for $i \neq j$, for some integer $0 < B_n \leq d$, such that we can factorize:

$$g_n(y_n|x_n) f_n(x_n|x_{n-1}) = \prod_{j=1}^{B_n} \alpha_{n,j}(y_n, x_{n-1}, x_n(D_{n,j})), \tag{2.19}$$

for appropriate functions $\alpha_{n,j}(\cdot)$, where $x_n(D) = \{x_n(j) : j \in D\} \in \mathbb{R}^{|D|}$ (Fig. 2.2).

By running an SMC algorithm on each nonoverlapping subset, the filtering distribution of interest is therefore approximated at the end of each iteration as follows:

Algorithm 2 SMC algorithm with Bridging densities for optimal filtering

Ensure: $n \geq 2$

1: Sample $X_{n,0}^j \sim q_n(x_n|X_{1:n-1}^j)$ and set $X_{1:n,0}^j := (X_{1:n-1}^j, X_{n,0}^j), \forall j = 1, \ldots, N$

2: Set $W_{n,0}^j = W_{n-1}^j, \forall j = 1, \ldots, N$ and define $\pi_0(x_{1:n}) = q_n(x_n|x_{1:n-1})\pi(x_{1:n-1})$

3: **for** $m = 1, \ldots, M$ **do**

4: Calculate the weights $W_{n,m}^j \propto W_{n,m-1}^j \dfrac{\pi_m(X_{1:n,m-1}^j)}{\pi_{m-1}(x_{1:n,m-1}^j)}, \forall j = 1, \ldots, N$

5: **if** $ESS_n < \Gamma$ **then**

6: Resample $\left\{W_{n,m}^j, X_{1:n,m-1}^j\right\}$ to obtain N equally weighted particles $\left\{W_{n,m}^j = 1/N, X_{1:n,m-1}^j\right\}$

7: **end if**

8: Sample $X_{1:n,m}^j \sim \mathscr{K}_m(X_{1:n,m-1}^j, x_{1:n}), \forall j = 1, \ldots, N$ using $\mathscr{K}_m(\cdot)$ a Markov kernel having $\pi_m(\cdot)$ as its stationary distribution.

9: **end for**

10: Set $\left\{W_n^j, X_{1:n}^j\right\} = \left\{W_{n,M}^j, X_{1:n,M}^j\right\}, \forall j = 1, \ldots, N$

11: **Output:** Approximation of the smoothing distribution via the following empirical measure:

$$\pi(x_{1:n}) \approx \sum_{j=1}^{N} W_n^j \delta_{X_{1:n}^j}(dx_{1:n})$$

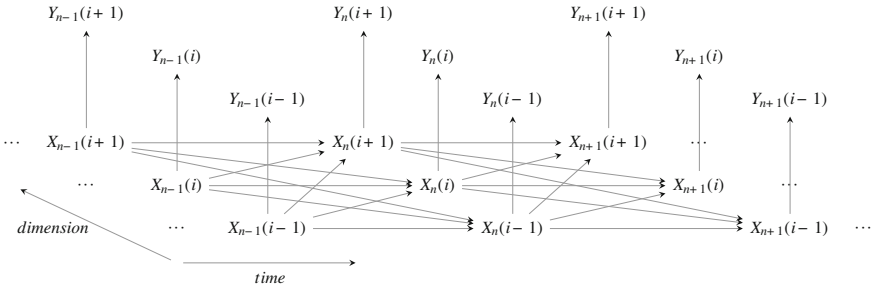


Fig. 2.2 Graphical representation of a hidden Markov model that satisfies an example of a factorization in Eq. (2.19)

$$\pi(x_n) \approx \bigotimes_{j=1}^{B_n} \pi(x_n(D_{n,j})). \tag{2.20}$$

The local SMC method, summarized in Algorithm 3, is well suited to distributed computation as the particles weights are computed locally due to the factorization described in Eq. (2.19). However, this strategy introduces some bias in the algorithm, so that the estimates given by the local SMC method do not converge to the exact filter distributions as the number of particles N goes to infinity. However, the hope is that by introducing a small amount of bias in the algorithm, its variance can be

reduced significantly since the B_n SMC algorithms are running on smaller dimension, i.e., $|D_{n,j}| \leq d$ for $j = 1, \ldots, B_n$. Moreover, the local error induced by the approximation of the target distribution as the product of marginals of the nonoverlapping subset is spatially inhomogeneous, i.e., the error will be larger for elements in x_n closer to subset boundaries. The computational cost of the local particle filter is $\mathcal{O}(nNd)$ by assuming that the cost of computing the product of the prior and likelihood distribution is $\mathcal{O}(d)$ (or equivalently that the cost of computing $\alpha_{n,j}$ is $\mathcal{O}(|D_{n,j}|)$ for each n and j).

Algorithm 3 Local SMC algorithm

1: **if** time $n = 1$ **then**
2: Sample $X_1^j(D_{1,i}) \sim q_1(x_1(D_{1,i}))$, $\forall j = 1, \ldots, N$ and $\forall i = 1, \ldots, B_1$
3: Calculate the weights $W_{1,i}^j \propto \dfrac{\alpha_{1,i}(Y_1, X_1^j(D_{1,i}))}{q_1(X_1^j(D_{1,i}))}$, $\forall j = 1, \ldots, N$ and $\forall i = 1, \ldots, B_1$
4: **else if** time $n \geq 2$ **then**
5: Sample X_{n-1}^j from $\widehat{\pi}(x_{n-1})$ for $j = 1, \ldots, N$
6: Sample $X_n^j(D_{n,i}) \sim q_n(x_n(D_{n,i})|X_{n-1}^j)$ $\forall j = 1, \ldots, N$ and $\forall i = 1, \ldots, B_n$
7: Calculate the weights $W_{n,i}^j \propto \dfrac{\alpha_{n,i}(Y_n, X_{n-1}^j, X_n^j(D_{n,i}))}{q_n(X_n^j(D_{n,i})|X_{n-1}^j)}$, $\forall j = 1, \ldots, N$ and $\forall i = 1, \ldots, B_n$
8: **end if**
9: **Output:** Approximation of the filtering distribution via the following empirical measure:

$$\widehat{\pi}(x_n) = \bigotimes_{i=1}^{B_n} \sum_{j=1}^{N} W_{n,i}^j \delta_{X_n^j(D_{n,i})}(dx_n(D_{n,i}))$$

2.3.3.3 Space-Time Particle Filter

The space–time Particle filter (STPF) has been recently proposed in [3]. As in both previous approaches described in Sects. 2.3.3.1 and 2.3.3.2, the idea is to have a gradual introduction of the likelihood $g_n(y_n|x_n)$ into the successive steps of the algorithm in order to decrease the variance of the importance weights which is the main reason of the collapse of the SMC methods in high-dimensional systems. In this work, the authors assume there exists an increasing sequence of sets $\left\{A_{n,j}\right\}_{j=1}^{B_n}$ with $A_{n,1} \subset A_{n,2} \subset \ldots \subset A_{n,B_n} = \{1 : d\}$, for some integer $0 < B_n \leq d$, such that we can factorize:

$$g_n(y_n|x_n) f_n(x_n|x_{n-1}) = \prod_{j=1}^{B_n} \alpha_{n,j}(y_n, x_{n-1}, x_n(A_{n,j})) \tag{2.21}$$

for appropriate functions $\alpha_{n,j}(\cdot)$, where $x_n(A) = \{x_n(j) : j \in A\} \in \mathbb{R}^{|A|}$. Let us remark that the assumption required for this factorization is weaker than the one

described in Eq. (2.19) for the local SMC methods since some dependencies between elements of x_n from different subsets given x_{n-1} are allowed. As discussed by the authors in their paper, this factorization is not a requirement but in such cases the performance of the filter will be degraded as additional sampling and reweighting steps are necessary.

The underlying idea of the STPF is to exploit the structure in Eq. (2.21) to design a particle filter moving along both the space and time index (as opposed to traditional particle filter that moves only along the time index). This approach can also be viewed as a generalization of the particle island particle filter proposed in [43] since the STPF combines a local filter running B_n space-step using M particles with a global particle filter making time steps using N particles. The authors show that this algorithm is asymptotically consistent and has a subexponential cost in d. Since the local particle filters are running along the space dimension, a patch degeneracy (on the space dimension) effect can be expected as the dimension of the system increases. The authors describe different strategies based on MCMC rejuvenation that could be employed to improve the performance of the algorithm at the expense of additional cost. The algorithm is summarized in Algorithm 4 and its computational cost is $\mathcal{O}(nNMd)$ by assuming that the cost of computing $\alpha_{n,j}$ is $\mathcal{O}(|A_{n,j}|)$ for each n and j. More specifically, the authors present results that, in 1) an i.i.d. scenario both in time and space and 2) a Markovian model along space, the algorithm is stable by setting the number of particles in the local systems equal to the dimension of the system (i.e., $M = d$), thus leading in that case to a cost of $\mathcal{O}(nNd^2)$.

2.4 Sequential Markov Chain Monte Carlo

In this section, another class of sequential Bayesian algorithm based on MCMC sampling (unlike importance sampling as in the previous section) is described. MCMC methods are generally more effective than importance sampling techniques in high-dimensional spaces. Their traditional formulation, however, allows sampling from probability distributions in a nonsequential fashion. Recently, advanced sequential MCMC schemes were proposed in [2, 5, 22, 27, 40] for solving online filtering inference problems. These approaches are distinct from the technique described previously in Sect. 2.3.3.1 where the MCMC algorithm is used to move samples following importance sampling resampling since these sequential MCMC use neither resampling nor importance sampling.

2.4.1 General Principle

Several sequential MCMC (SMCMC) methods have been proposed in the literature recently. In this section, we will describe a general framework that include all of them. The underlying idea of all these SMCMC approaches is to perform a

Algorithm 4 Space-Time Particle Filter

Define the subset $\Omega_i = A_{1,i} \setminus A_{1,i-1}$
if time $n = 1$ **then**
  for $j = 1, \ldots, N$ **do**
    for $i = 1, \ldots, B_1$ **do**
      Sample $X_1^{j,l}(\Omega_i) \sim q_1(x_1(\Omega_i)|X_1^{j,l}(A_{1,i-1})), \forall l = 1, \ldots, M$
      Set $X_1^{j,l}(A_i) = \left[X_1^{j,l}(A_{1,i-1}) \;\; X_1^{j,l}(\Omega_i)\right]$
      Calculate the weights $w_i^{j,l} = \dfrac{\alpha_{1,i}(y_1, \cdot, X_1^{j,l}(A_{1,i}))}{q_1(X_1^{j,l}(\Omega_i)|X_1^{j,l}(A_{1,i-1}))}, \forall l = 1, \ldots, M$
      Resample local particles $\left\{X_1^{j,l}(A_{1,i})\right\}_{l=1}^{M}$ according to their normalized weights $\left\{w_i^{j,l}\left[\sum_{k=1}^{M} w_i^{j,k}\right]^{-1}\right\}_{l=1}^{M}$
    end for
  end for
  Resample the N-particle systems, i.e., $\left\{X_1^{j,1:M}\right\}_{j=1}^{N}$, according to their weights defined as:

$$W_1^j \propto \prod_{i=1}^{B_1} \frac{1}{M} \sum_{l=1}^{M} w_i^{j,l}$$

else if time $n \geq 2$ **then**
  for $j = 1, \ldots, N$ **do**
    for $i = 1, \ldots, B_n$ **do**
      Sample $X_n^{j,l}(\Omega_i) \sim q_1(x_n(\Omega_i)|X_n^{j,l}(A_{n,i-1}), X_{n-1}^{j,l}), \forall l = 1, \ldots, M$
      Set $X_n^{j,l}(A_{n,i}) = \left[X_n^{j,l}(A_{n,i-1}) \; X_n^{j,l}(\Omega_i)\right]$
      Calculate the weights $w_i^{j,l} = \dfrac{\alpha_{n,i}(y_n, X_{n-1}^{j,l}, X_n^{j,l}(A_{n,i}))}{q_n(X_n^{j,l}(\Omega_i)|X_n^{j,l}(A_{n,i-1}), X_{n-1}^{j,l})}, \forall l = 1, \ldots, M$
      Resample local particles $\left\{X_{n-1}^{j,l}, X_n^{j,l}(A_{n,i})\right\}_{l=1}^{M}$ according to their normalized weights $\left\{w_i^{j,l}\left[\sum_{k=1}^{M} w_i^{j,k}\right]^{-1}\right\}_{l=1}^{M}$
    end for
  end for
  Resample the N-particle systems, i.e., $\left\{X_{1:n}^{j,1:M}\right\}_{j=1}^{N}$, according to their weights defined as:

$$W_n^j \propto \prod_{i=1}^{B_n} \frac{1}{M} \sum_{l=1}^{M} w_i^{j,l}$$

end if
Output: Approximation of the filtering distribution via the following empirical measure:

$$\pi(x_n) \approx \frac{1}{NM} \sum_{j=1}^{N} \sum_{l=1}^{M} \delta_{X_n^{j,l}}(dx_n)$$

Metropolis–Hastings (MH) accept-rejection step as a correction for having used a proposal distribution to sample the current state in order to approximate the posterior target distribution as opposed to SMC methods that use a correction based on Importance sampling.

At time step n, the target distribution of interest to be sampled from is

$$\underbrace{p(x_{1:n}|y_{1:n})}_{\pi_n(x_{1:n})} \propto g_n(y_n|x_n)\, f_n(x_n|x_{n-1})\, \underbrace{p(x_{1:n-1}|y_{1:n-1})}_{\pi_{n-1}(x_{1:n-1})}. \tag{2.22}$$

Unfortunately, it is impossible to sample from $p(x_{1:n-1}|y_{1:n-1})$ since this distribution is analytically intractable. The key idea of all existing SMCMC methods is therefore to replace $p(x_{1:n-1}|y_{1:n-1})$ by an empirical approximation obtained from previous iterations of the algorithm. The target distribution of interest at time step n is therefore defined as:

$$\pi_n(x_{1:n}) \propto g_n(y_n|x_n)\, f_n(x_n|x_{n-1}) \widehat{\pi}^i_{n-1}(x_{1:n-1}), \tag{2.23}$$

with

$$\widehat{\pi}^i_{n-1}(x_{1:n-1}) = \frac{1}{i - N_b} \sum_{m=N_b+1}^{i} \delta_{X^m_{n-1,1:n-1}}(dx_{1:n-1}), \tag{2.24}$$

where $\left\{X^m_{n-1,1:n-1}\right\}_{m=1}^{i}$ corresponds to the i samples of the $(n-1)$th Markov chain, whose distribution is $\pi_{n-1}(x_{1:n-1})$ as defined in Eq. (2.23) that has been generated until the current iteration of the MCMC at time step n (N_b represents the length of the burn-in period). By using this empirical approximation of the previous target distribution, an MCMC kernel can be employed in order to obtain a Markov chain, denoted by $\left(X^1_{n,1:n}, X^2_{n,1:n}, \ldots\right)$, with stationary distribution $\pi_n(x_{1:n})$ as defined Eq. (2.23).

As summarized in Algorithm 5, the SMCMC proceeds as follows. At time step $n = 1$, an MCMC kernel $\mathscr{K}_1$ of invariant distribution $\pi_1(x_1) \propto g_1(y_1|x_1)\mu(x_1)$ is employed to generate a Markov chain denoted by $\left(X^1_{1,1}, \ldots, X^{N+N_b}_{1,1}\right)$. At time step n, the $N + N_b$ iterations of the SMCMC aims at producing a Markov chain, denoted by $\left(X^1_{n,1:n}, \ldots, X^{N+N_b}_{n,1:n}\right)$, by using an MCMC kernel $\mathscr{K}_n$ of invariant distribution $\pi_n(x_{1:n})$ as defined in Eq. (2.23). Moreover, samples can be added to the previous L Markov chains, i.e., $X_{n-L,1:n-L}$ with $L > 1$, in order to improve the empirical approximation $\widehat{\pi}_{n-1}(x_{1:n-1})$ required in the posterior distribution of interest at time step n. Once the nth Markov chain has been generated, the last N are extracted to obtain the empirical approximation of the filtering distribution:

$$p(x_n|y_{1:n}) \approx \frac{1}{N} \sum_{m=N_b+1}^{N+N_b} \delta_{X^m_{n,n}}(dx_n). \tag{2.25}$$

By assuming that the computation of the product of likelihood and prior as well as the MCMC kernel used is $\mathcal{O}(d)$, the cost of this algorithm is $\mathcal{O}(nLNd)$ since the length of the burn-in period is generally considered to be a percentage of the useful samples, i.e., $N_b = \beta N$ with $0 \leq \beta \leq 1$.

Algorithm 5 Generic Sequential MCMC algorithm for optimal filtering

Initialization $\{i_n\}_{n\geq 0} = 0$
if time $n = 1$ **then**
  for $j = 1, \ldots, N + N_b$ **do**
    Set $i_1 = i_1 + 1$
    Sample $X_{1,1}^{i_1} \sim \mathscr{K}_1(x_1^{i_1-1}, \cdot)$ with $\mathscr{K}_1$ an MCMC kernel of invariant distribution $\pi_1(x_1) \propto g_1(y_1|x_1)\mu(x_1)$.
  end for
else if time $n \geq 2$ **then**
  for $j = 1, \ldots, N + N_b$ **do**
    for $k = \max(1, n - L + 1), \ldots, n$ **do**
      Set $i_k = i_k + 1$
      Sample $X_{k,1:k}^{i_k} \sim \mathscr{K}_k^{(i_{k-1})}(X_{k,1:k}^{i_k-1}, \cdot)$ with $\mathscr{K}_k^{(i_{k-1})}$ an MCMC kernel of invariant distribution $\pi_k^{(i_{k-1})}$ given by:

$$\pi_k^{(i_{k-1})}(x_{1:k}) \propto g_k(y_k|x_k) f_k(x_k|x_{k-1}) \widehat{\pi}_{k-1}^{(i_{k-1})}(x_{1:k-1})$$

with $\widehat{\pi}_{k-1}^{i_{k-1}}$ being the empirical measure obtained using previous samples, i.e.

$$\widehat{\pi}_{k-1}^{(i_{k-1})}(x_{1:k-1}) = \frac{1}{i_{k-1} - N_b} \sum_{m=N_b+1}^{i_{k-1}} \delta_{X_{k-1,1:k-1}^m}(dx_{1:k-1})$$

    end for
  end for
end if
Output: Approximation of the smoothing distribution with the following empirical measure:

$$\pi(\mathbf{x}_n) \approx \frac{1}{N} \sum_{j=N_b+1}^{N+N_b} \delta_{X_{n,1:n}^j}(dx_{1:n})$$

2.4.2 Algorithm Settings

The overall performance of the SMCMC algorithm applied to optimal filtering depends heavily upon the choice of the MCMC kernel. One of the attractive features of this SMCMC is to be able to employ all the different MCMC methods that have been proposed in the scientific literature. In practical implementation of the SMCMC and more especially for high-dimensional systems, composite kernel

based on joint and conditional draws are generally very efficient [40]. Summarized in Algorithm 6, such a composite kernel is based on the following two main steps:

1. A joint draw in which a Metropolis–Hastings sampler is used to update all the path of states corresponding to $x_{1:n}$
2. A refinement step in which previous history $x_{1:n-1}$ and current state x_n are updated successively. Moreover, if x_n is high-dimensional, an efficient way to update it consists in firstly dividing its space into P disjoint subsets and update them successively either via a random scan or a deterministic scan using a series of block MH-within—Gibbs steps.

The cost of this MCMC kernel is $\mathcal{O}(d)$ if a factorization such as the one defined in Eq. (2.21) is valid.

As a comparison, Berzuini et al. [2] made only use of the individual refinement step described above (with $L = 1$ in Algorithm 5). This can potentially lead to poor mixing in high-dimensional problems due to the highly disjoint predictive density of the particle representation. On the other hand, Golightly and Wilkinson [22] made use of only the joint draw to move the MCMC chain. This can potentially reduce the effectiveness of the MCMC as refinement moves are not employed to explore the structured probabilistic space which is very challenging in high-dimensional systems. Indeed, it could be difficult to design a proposal distribution for the joint draw that does not lead to low acceptance rate. In [5], the authors propose a general framework of SMCMC with the possibility of updating previous Markov chains (i.e., $L > 1$ in Algorithm 5). An independent Metropolis–Hastings sampler as MCMC kernel is employed. By doing so, the ratio of the normalizing constant Z_n/Z_{n-1} can be easily estimated but it could be difficult to design the independent proposal distribution leading to satisfactory performance. Finally, in [39], the authors proposed to incorporate several attractive features of population-based MCMC methods [19, 31] such as genetic moves and simulated annealing in order to improve the mixing of the Markov chain in complex scenarios.

2.5 Assessing Local and Global Sample Effective Sample Size and SMCMC Convergence Diagnostics

Each of the discussed algorithms: SMC with MCMC moves; local SMC; Space–Time SMC; and sequential MCMC will have different features with regard to the effective sample size produced and even how one may consider the effective sample size under each class of algorithm requires further consideration. In this section, we provide a brief overview of two aspects, firstly how to determine the number of independent samples present in the resulting set of samples or particles. Then, we also discuss some convergence diagnostics in standard MCMC that may be adapted for the setting of SMCMC to adaptively modify the past "population" MCMC chains in the sequence.

Algorithm 6 Example of an MCMC Kernel $\mathscr{K}_n^{(i_{n-1})}(X_{n,1:n}^{i_n-1},\cdot)$ for the SMCMC

1: *Joint Draw*
2: Propose $\{X_{n,1:n}^*\} \sim q_1(x_{1:n}|X_{n,1:n}^{i_n-1})$
3: Compute the MH acceptance probability $\rho_1 = \min\left(1, \frac{\pi_n^{(i_{n-1})}(X_{n,1:n}^*)}{q_1(X_{n,1:n}^*|X_{n,1:n}^{i_n-1})} \frac{q_1(X_{n,1:n}^{i_n-1}|X_{n,1:n}^*)}{\pi_n^{(i_{n-1})}(X_{n,1:n}^{i_n-1})}\right)$
4: Accept $X_{n,1:n}^{i_n} = X_{n,1:n}^*$ with probability ρ_1 otherwise set $X_{n,1:n}^{i_n} = X_{n,1:n}^{i_n-1}$
5: *Refinement*
6: Propose $\{X_{n,1:n-1}^*\} \sim q_{R,1}(x_{1:n-1}|X_{n,1:n}^{i_n})$
7: Compute the MH acceptance probability $\rho_{R,1} = \min\left(1, \frac{\pi_n^{(i_{n-1})}(X_{n,1:n-1}^*, X_{n,n}^{i_n})}{q_{R,1}(X_{n,1:n-1}^*|X_{n,1:n}^{i_n})} \frac{q_{R,1}(X_{n,1:n-1}^{i_n}|X_{n,1:n-1}^*, X_n^{i_n})}{\pi_n^{(i_{n-1})}(X_{n,1:n}^{i_n})}\right)$
8: Accept $X_{n,1:n-1}^{i_n} = X_{,1:n-1}^*$ with probability $\rho_{R,1}$.
9: Randomly divide x_n into P disjoint blocks $\{\Omega_p\}_{p=1}^P$ such that $\bigcup_p \Omega_p = \{1:d\}$
10: **for** $p = 1,\ldots,P$ **do**
11: Propose $\{X_{n,n}^*(\Omega_p)\} \sim q_{R,p}(x_n(\Omega_p)|X_{n,1:n}^{i_n})$
12: Compute the MH acceptance probability $\rho_{R,p} = \min\left(1, \frac{\pi_n^{(i_{n-1})}(X_{n,n}^*(\Omega_p), X_{n,n}^{i_n}(\mathbb{R}^d \setminus \Omega_p), X_{n,1:n-1}^{i_n})}{q_{R,p}(X_{n,n}^*(\Omega_p)|X_{n,1:n}^{i_n})} \frac{q_{R,p}(X_{n,n}^{i_n}(\Omega_p)|X_{n,n}^*(\Omega_p), X_{n,n}^{i_n}(\mathbb{R}^d \setminus \Omega_p), X_{n,1:n-1}^{i_n})}{\pi_n^{(i_{n-1})}(X_{n,1:n}^{i_n})}\right)$
13: Accept $X_{n,n}^{i_n}(\Omega_p) = X_{n,n}^*(\Omega_p)$ with probability $\rho_{R,p}$
14: **end for**

2.5.1 Assessing Local and Global Sample Effective Sample Size for SMCMC

We start by briefly recalling the properties of effective sample size in the standard markov chain setting before talking about these in the context of the three classes of algorithms we consider in the high-dimensional state space models discussed in this chapter.

In general for a correlated time series one may define the effective sample size which goes back to early studies such as those by [30] who studies the time between effectively independent samples or the reciprocal effective number of independent samples in a time span which is often referred to as the effective sample size (ESS). A simple definition of such a quantity is to equate the ensemble mean square of a time-averaged mean denoted by $\sigma_{\overline{X}}^2$ which is based on the autocovariance function (acf) to the standard formula for the variance of the mean of independent samples. The solution to the number of independent samples is one measure of ESS. To proceed consider the N values from a time series $X_1,\ldots,X_N$ of a stationary stochastic process with variance σ^2, then one can write the ensemble mean as follows (see [1]) with respect to the mean μ and symmetric covariance between observations lagged by a time interval τ denoted $C(\tau)$ and corresponding lag-correlation function $\rho(\tau)$ according to:

$$\sigma_{\overline{X}}^2 = \frac{1}{N}\sum_{i,j}^{N}\left\langle (X_i-\mu)(X_j-\mu)\right\rangle = \frac{1}{N^2}\sum_{i,j}^{N} C(i-j)$$
$$= \frac{1}{N^2}\sum_{\tau=-(N-1)}^{(N-1)} [N-|\tau|]\, C(\tau) \tag{2.26}$$
$$= \frac{\sigma_2}{N}\sum_{\tau=-(N-1)}^{(N-1)}\left[1-\frac{|\tau|}{N}\right]\rho(\tau).$$

If one then considers the independence case with N' samples then one would have had the variance of the sample mean given instead by σ^2/N', by equating these one obtains the effective sample size

$$N' = \frac{\sigma^2}{\sigma_{\overline{X}}^2} = N\left[\sum_{\tau=-(N-1)}^{(N-1)}\left[1-\frac{|\tau|}{N}\right]\rho(\tau)\right]^{-1}. \tag{2.27}$$

Therefore, such an estimator is typically used for standard MCMC settings where the ESS is defined for a MCMC sample of size N by

$$\mathrm{ESS}^{\mathrm{MCMC}} = \frac{N}{1+2\sum_{k=1}^{\infty}\rho_k}. \tag{2.28}$$

In the context of Markov chain Monte-Carlo methods, this framework can be adopted to study the asymptotic variance of the mean of a Markov chain, with respect to a bounded and integrable test function generically denoted by φ, under the central limit theorem. In this case, one can state the following results. Let $X = \{X_i : i = 0, 1, 2, \ldots\}$ be a Harris ergodic Markov chain on a general space χ with invariant probability distribution π having support χ. Let φ be a Borel function and define $\overline{\varphi}_T := \frac{1}{T}\sum_{i=1}^{T}\varphi(X_i)$ and $\mathbb{E}_\pi[\varphi] := \int_\chi \varphi(x)\pi(dx)$. When $\mathbb{E}_\pi[|\varphi|] < \infty$ the ergodic theorem guarantees that $\overline{\varphi}_T \to \mathbb{E}_\pi[\varphi]$ with probability 1 as $T \to \infty$. The conditions on the Markov chain for this convergence result to hold are stated for a range of MCMC methods in [24] and the following general CLT result applies under these different conditions:

$$\sqrt{T}\,(\overline{\varphi}_T - \mathbb{E}_\pi[\varphi]) \xrightarrow{d} N(0, \sigma_\varphi^2). \tag{2.29}$$

Here, the asymptotic variance of the estimated mean of a test function φ is given by the finite variance given by

$$\sigma_\varphi^2 := \mathbb{V}\mathrm{ar}_\pi[\varphi(X_0)] + 2\sum_{i=1}^{\infty}\mathbb{C}\mathrm{ov}_\pi[\varphi(X_0), \varphi(X_i)]. \tag{2.30}$$

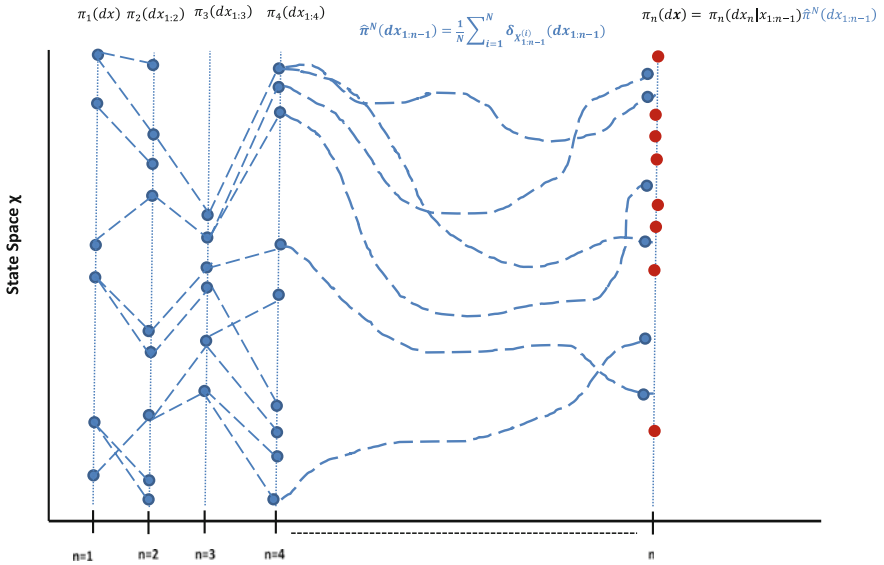


Fig. 2.3 SMCMC sampler construction

If one selects $\varphi(X) = X$ and applies a truncation to the number of Markov chain samples to N, then after renormalization, this is exactly the expression obtained in Eq. (2.26). This is the typical framework used to understand effective sample size and also it acts as a core ingredient in the derivation of the convergence diagnostics for MCMC samplers. In the following, we will explain how to adapt this classical MCMC ESS framework to the case of the SMCMC setting.

In the SMCMC setting, we are sampling sequentially the target distribution sequence $\{\pi_n\}_{n\in\mathbb{N}}$ via a sequence of Markov chains constructed through a Metropolis–Hastings accept reject framework, however, the sequence of chains are constructed based on the previous sequence path-space genealogies. To understand this we refer to Fig. 2.3 where the blue "particle" trajectories correspond to the previously sampled path-space genealogies for the SMCMC algorithm that comprise the empirical measure $\widehat{\pi}_{n-1}(x_{1:n-1})$ for the construction of the sampler at target π_n. The blue trajectories are the previously accepted sequences of Markov chain samples that have been accepted, so that at iteration n we would randomly (with replacement) draw a trajectory path, then construct conditionally on this path a new state sample denoted in red which would be accepted or rejected based on a Metropolis–Hastings accept reject mechanism. We will refer to the previous genealogical paths used in the proposal at time n by the set of path-space branches χ_n (in blue).

We can see from this representation that one needs to develop an effective sample size criterion for the SMCMC algorithm that would adequately reflect the effect of the geneological path-space behaviour used to construct the sequence of distributions

sampled. To achieve this we consider propose the forward and backward SMCMC effective sample size criterions.

Definition 1 (*Forward Efficiency of SMCMC*) Consider the path-space genealogy χ_n at distribution sequence iteration n, the conditional forward efficiency $\eta_n \in [0.1]$ of the algorithm having sampled N MCMC iterations is given, for a bounded integrable test function φ by

$$\eta_n(\varphi;\, \chi_n) := \frac{\mathbb{V}\mathrm{ar}\left(\overline{\varphi}^{\text{i.i.d.}}_{N'}|\chi_n\right)}{\mathbb{V}\mathrm{ar}\left(\overline{\varphi}_N|\chi_n\right)} = \frac{1}{1+2\sum_{k=1}^{\infty}\rho_k(\chi_n)} \tag{2.31}$$

where $\rho_k(\chi_n)$ denotes the autocorrelation which is implicitly dependent on the geneological path-space χ_n, $\overline{\varphi}_N$ is the mean estimated from N correlated MCMC samples from $\pi_n(dx_{1:n}|\chi_n) = \pi_n(dx_n|\chi_n)\widehat{\pi}^N(dx_{1:n-1})$ and $\overline{\varphi}^{\text{i.i.d.}}_{N'}$ is the estimator for the optimal case of i.i.d. samples from $\pi_n(dx_{1:n}|\chi_n)$ with $N' \leq N$.

For this forward measure of efficiency of the SMCMC algorithm, which can be computed online for each target distribution π_n one can approximate this efficiency measure by the estimator given by first constructing from the N correlated MCMC samples $\varphi_i := \varphi(X^i_{n,1:n})$ the autocoviariance function

$$\widehat{\gamma}_{\varphi}(k;\, \chi_n) = \frac{1}{N}\sum_{i=1}^{N-|k|}\left(\varphi_{i+|k|} - \overline{\varphi}_N\right)\left(\varphi_i - \overline{\varphi}_N\right), \quad -N < k < N. \tag{2.32}$$

This would then lead to the estimator for the autocorrelations given by

$$\widehat{\rho}_{\varphi}(k;\, \chi_n) = \frac{\widehat{\gamma}_{\varphi}(k;\, \chi_n)}{\widehat{\gamma}_{\varphi}(0;\, \chi_n)}, \tag{2.33}$$

giving the estimator for the efficiency at stage n in Eq. 2.31 by substitution. However, it may also be of interest to consider the efficiency in another sense, to capture the path-space implicit effect on the SMCMC. To achieve this, we consider also the backward efficiency at stage n conditional on the SMCMC samples at stage n, this is given in the following definition.

Definition 2 (*Backward Efficiency of SMCMC*) Consider, the path-space genealogy χ_n at distribution sequence iteration n decomposed as $\chi_n = \chi_{n-1} \cup \left\{X^i_{n-1,n-1}\right\}_{i=1}^N$ such that $\chi_{n-1} \subseteq \chi_n$, then the conditional one-stage backward efficiency viewed from iteration n is given by $\overleftarrow{\eta}_n \in [0,\, 1]$, for a bounded integrable test function φ according to

$$\overleftarrow{\eta}_n(\varphi;\chi_{n-1}) := \int \eta_n(\varphi;\{X^i_{n-1,n-1}\}_{i=1}^N \cup \chi_{n-1})\pi_{n-1}(dx_{n-1,1:n-1}|\chi_{n-1}). \quad (2.34)$$

This can then be defined recursively for any number of backward looking steps from distribution sequence iteration n.

One can approximate this using the path-space samples obtained in each iteration according to the following estimator

$$\widehat{\overleftarrow{\eta}}_n(\varphi;\chi_{n-1}) = \frac{1}{J}\sum_{j=1}^{J}\eta_n^{(j)}(\varphi;\{X^i_{n-1,n-1}\}_{i=1}^N \cup \chi_{n-1}). \quad (2.35)$$

where $\eta_n^{(j)}\left(\varphi;\left\{X^{(i,j)}_{n-1,n-1}\right\}_{i=1,j=1}^{N,J} \cup \chi_{n-1}\right)$ is obtained using the estimator at time n of efficiency in Eq. 2.31 for the jth population sample of $\left\{X^{(i,j)}_{n-1,n-1}\right\}_{i=1}^N$ conditional on previous genealogical paths in χ_{n-1}, a visual representation of how this estimator is obtained based on resampling of the previous generation at time $n-1$ is provided in Fig. 2.4. This illustration shows in red the resampled path genealogies for iteration n looking back in this case one time step to iteration n's $n-1$ parents and regenerating these $j \in \{1,\ldots,J\}$ times, with the jth regeneration producing the backward

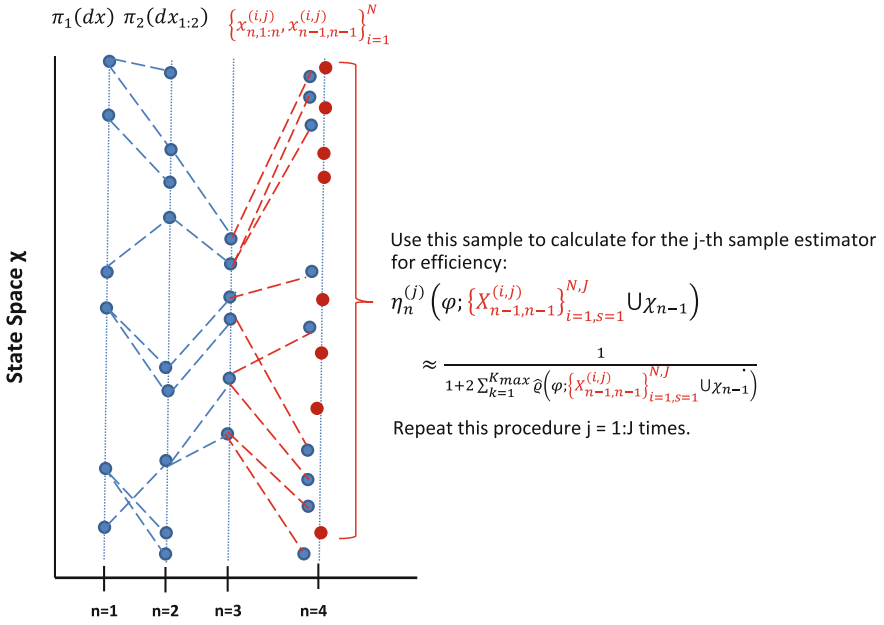


Fig. 2.4 SMCMC sampler backward efficiency measure estimator

looking effective sample size one step back approximation obtained by the estimated autocorrelations.

In addition to these two measure of efficiency, one can also monitor at iteration n a related quantity that can be estimated at each MCMC iteration of the SMCMC algorithm at distribution sequence iteration n to decide if one should perform more local or more global moves. This involves adapting the following well-known Geweke [18] convergence diagnostic for MCMC methods can be adopted in the SMCMC sampler setting at each iteration as follows. If the total chain has length $N + N_b$, the initial burn-in stage will correspond to the first N_b samples. We denote by $\{X_{n,i}^{(t)}\}_{t=1:N}$ the Markov chain of the ith parameter after burn-in. The diagnostics we consider are given by:

- For state $X_{n,i}$ it is calculated as follows:

1. Split the Markov chain samples into two sequences, $\{X_{n,i}^{(t)}\}_{t=1:N_1}$ and $\{X_{n,i}^{(t)}\}_{t=N^*:N}$, such that $N^* = N - N_2 + 1$, and with ratios N_1/N and N_2/N fixed such that $(N_1 + N_2)/N < 1$ for all N.
2. Evaluate $\widehat{\mu}\left(X_{n,i}^{N_1}\right)$ and $\widehat{\mu}\left(X_{n,i}^{N_2}\right)$ corresponding to the sample means on each subsequence.
3. Evaluate consistent spectral density estimates for each subsequence, at frequency 0, denoted $\widehat{SD}(0; N_1, X_{n,i})$ and $\widehat{SD}(0; N_2, X_{n,i})$. The spectral density estimator considered in this paper is the classical nonparametric periodogram or power spectral density estimator. We use Welch's method with a Hanning window.
4. Evaluate convergence diagnostic given by
$$Z_N = \frac{\widehat{\mu}\left(X_{n,i}^{N_1}\right) - \widehat{\mu}\left(X_{n,i}^{N_2}\right)}{N_1^{-1}\widehat{SD}(0; N_1, X_{n,i}) + N_2^{-1}\widehat{SD}(0; N_2, X_{n,i})}.$$
According to the central limit theorem, as $N \to \infty$ one has that $Z_N \to \mathcal{N}(0, 1)$ if the sequence $\{X_{n,i}^{(t)}\}_{t=1:N}$ is stationary.

Note, this can be monitored and tested online for each stage and each parameter subspace of the SMCMC algorithm at iteration n in the distribution sequence to decide if one should sample more local moves or more global moves.

2.5.2 *Effective Sample Size for SMC Methods*

In the SMC literature, the notion of ESS that is typically adopted is based on the approach discussed for standard SMC algorithms of [28]. In this framework, an approximation to the effective sample size of the filtering distribution is obtained at time t. To understand this approximation we first define:

- the estimated sample mean from the filtering distribution weighted particle population given by samples drawn from mutation kernel $q(\cdot)$,

$$\widehat{\mathbb{E}}_q^N = \frac{1}{N}\sum_{i=1}^{N} W\left(X^{(i)}\right)\varphi\left(X^{(i)}\right); \tag{2.36}$$

- the estimated sample mean from the filtering distribution given by samples drawn from the true filtering target distribution $\pi(\cdot)$,

$$\widehat{\mathbb{E}}_\pi^N = \frac{1}{N}\sum_{i=1}^{N} \varphi\left(X^{(i)}\right). \tag{2.37}$$

Then in the standard SMC setting one typically starts by considering the ratio of the following two sample mean variances and applies a Taylor series expansion and applies the Delta method to obtain

$$\begin{aligned}\eta^{SMC} = \frac{\mathbb{V}\mathrm{ar}_\pi\left[\widehat{\mathbb{E}}_\pi^N\right]}{\mathbb{V}\mathrm{ar}_q\left[\widehat{\mathbb{E}}_q^N\right]} &\approx \left(1+\mathbb{V}\mathrm{ar}_q\left[W(X)\right]\right)^{-1} \\ &= \left(\mathbb{E}_q\left[W(X)^2\right]\right)^{-1}. \end{aligned} \tag{2.38}$$

2.6 Numerical Simulations

In this section, we study the empirical performance of the different algorithms that have been previously described, namely: (a) SMC in Algorithm 1—(b) SMC-MCMC in Algorithm 2—(c) Local SMC in Algorithm 3— (d) STPF in Algorithm 4, and (e) SMCMC in Algorithm 5 and 6. For all the different SMC-based algorithms, the resampling step is performed when the effective sample size, ESS, is below $N/2$. All the proposal distributions required in these algorithms are based on the prior distributions. The MCMC kernel used in the SMC-MCMC defined in Algorithm 2 correspond to series of P Metropolis–Hastings within Gibbs samplers used in the SMCMC and described in lines 9–13 of Algorithm 6. The parametric function used for the cooling schedule strategy used to design the sequence of bridging densities within the SMC-MCMC in this section is defined as, for $m = 1, \ldots, M$ by the sequence:

$$\alpha_m = \frac{\exp(\gamma m/M) - 1}{\exp(\gamma) - 1} \tag{2.39}$$

with $\gamma = 5$. In results presented for the SMCMC algorithm, only the current posterior distribution is updated, i.e., $L = 1$.

2.6.1 Linear and Gaussian Dynamical Model

As a first example, a simple linear and Gaussian state space model is considered, i.e., for $n = 1, \ldots, T$:

$$f_n(x_n|x_{n-1}) = \mathcal{N}(x_n; Hx_{n-1}, \Sigma_x),$$
$$g_n(y_n|x_n) = \mathcal{N}(y_n; Gx_n, \Sigma_y). \tag{2.40}$$

Such a model is interesting for the understanding and the study of approximation methods since the posterior distribution can be derived analytically via the use of the Kalman filter [26]. In our simulation results, the matrix H of size $d \times d$ has been obtained by randomly and independently selecting for each row two column indexes for which the value is set to 0.495. The covariance matrices are defined as $\Sigma_x = I_d$ and $\Sigma_y = I_{d_y}$. For a fair comparison between the different algorithms, we decide to set their parameter as shown in Table 2.1 in order to have an equivalent computational cost for all algorithms.

The performances are studied with a scenario in which all the d-dimensions of the hidden state are observed at each time step, i.e., $d_y = d$ and $G = I_d$. From this model, owing to the diagonality of G and the covariance matrix both in the prior and the likelihood distribution, it is obvious that their product can be factorized as in Eq. (2.19) with $|D_{n,j}| = 1$, $\forall n, j$. As a consequence, we define the partitioning of the subset in the STPF such that $\forall n, j$: $|A_{n,j} \setminus A_{n,i-1}| = 1$. In Fig. 2.5, the ESS scaled by the number of particles obtained for the different SMC-based algorithms is depicted. The standard SMC algorithm performs very badly even when the dimension is 10 and completely collapses when $d = 50$. The same remarks hold when the SMC-MCMC is used with only $M = 1$ which corresponds to the resample-move algorithm. However, we can observe that the use of a sequence of bridging densities that gradually introduces the effect of the likelihood distribution (SMC-MCMC with $M = 10$), improves the effective sample size remarkably. The effective sample size of the local SMC filter corresponds to the average of the ESS obtained from the d/B_n SMC filters used in each B_n subsets. As a consequence, the ESS depends quite obviously on the cardinality of each subset. Finally, the performance of the STPF in terms of ESS is deteriorating when the dimension increases due to the path degeneracy effect that we have discussed previously. However, this ESS is obtained

Table 2.1 Value of the different parameters of Monte-Carlo algorithms used in the simulation

	SMC	SMC-MCMC		Local SMC	STPF		SMCMC	
Complexity	$\mathcal{O}(nNd)$	$\mathcal{O}(nNMd)$		$\mathcal{O}(nNd)$	$\mathcal{O}(nNMd)$		$\mathcal{O}(nLNd)$	
Parameter	N	N	M	N	N	M	L	N
Value	1000	100	10	1000	10	100	1	1000

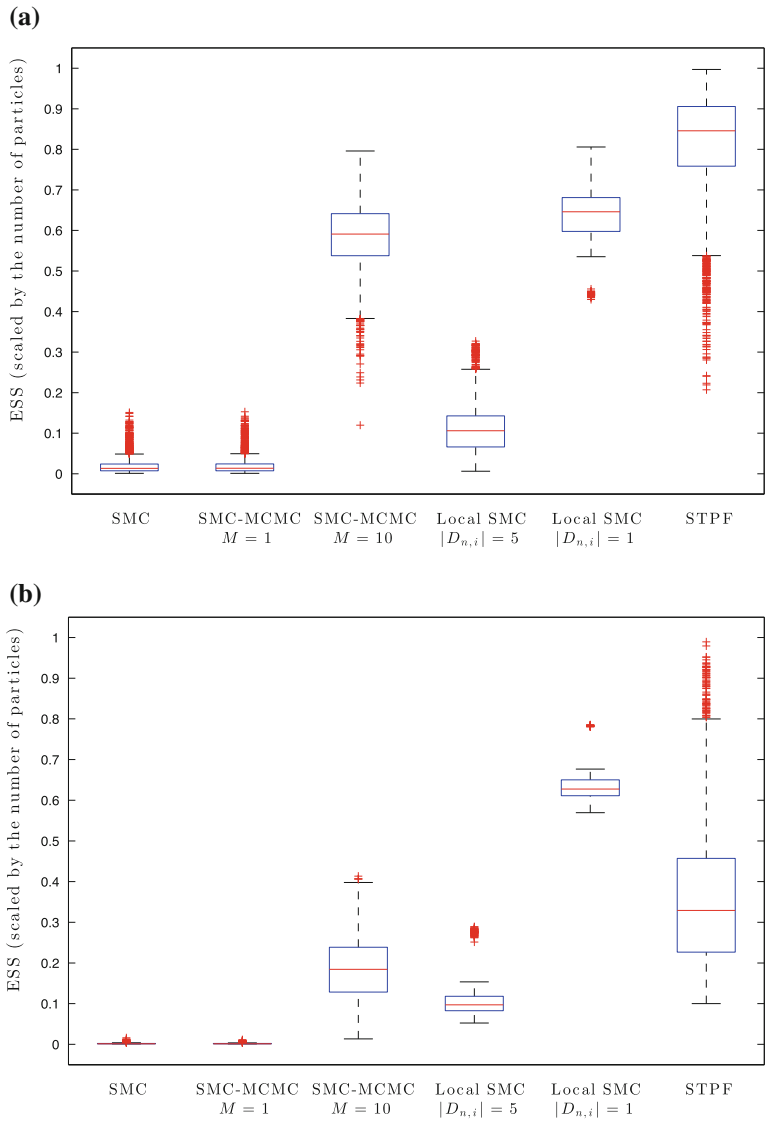


Fig. 2.5 Effective sample size, scaled by the number of particles, obtained with the different algorithms for the linear and Gaussian state-space model at the different time steps using 100 runs. **a** $d = 10$. **b** $d = 50$

using the global weights (line 21 of Algorithm 4) which thus corresponds to the number of local particle systems that contributes to the final estimator.

Figure 2.6 shows the variance for the estimators of the d-dimensional latent states, $X(1), \ldots, X(d)$ (with $d = 100$), averaged over time and obtained using 100 runs.

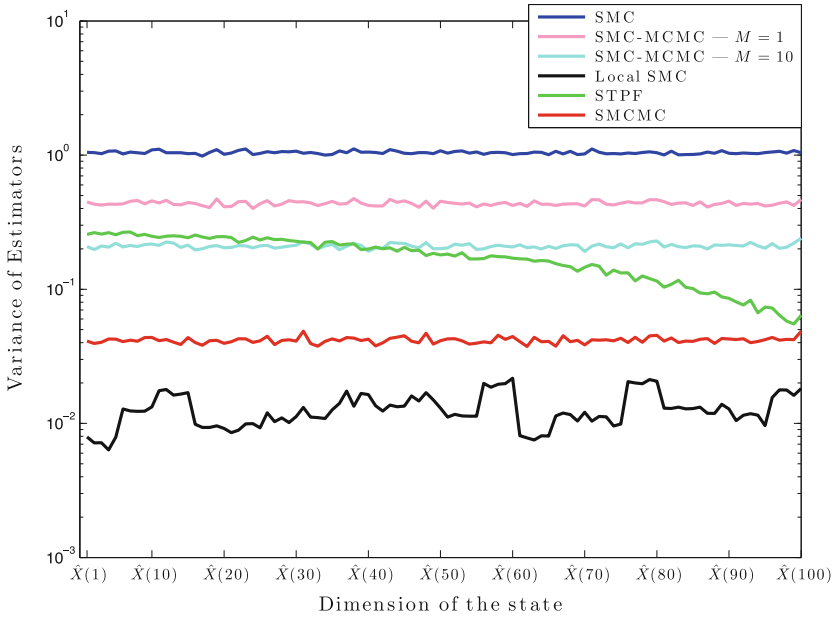


Fig. 2.6 Variance for estimators of $X(1), \ldots, X(d)$ (with $d = 100$) averaged over time and obtained using 100 runs. The MCMC kernel used in both SMC-MCMC and SMCMC partition the space with subsets of dimension 5. The dimension of each subset in the local SMC is also 5

We can clearly see the path degeneracy problem in the STPF which leads to an higher variance for $\hat{X}(1)$ compared to $\hat{X}(d)$. The variance of both SMC-MCMC and SMCMC are quite stable across dimension of the state. The local SMC outperforms, in terms of variance, the other techniques but suffers from a spatially inhomogeneous approximation of the posterior distribution as we can see from unstable variance over space.

Table 2.2 summarizes the bias and the variance for the estimator of the posterior mean for all the algorithms with different parameter configuration. As expected, the performances of the classical SMC algorithm deteriorates quite significantly as d increases. The introduction of the sequence of bridging densities (SMC-MCMC) clearly improves the performances of the algorithm. Moreover, the use of smaller dimension on each subset ($P = d$ vs. $P = d/5$) for the Metropolis–Hastings within Gibbs sampler used within the SMC-MCMC leads in that example to better performance. The STPF performs quite well compared to both SMC and SMC-MCMC, but as previously illustrated, could be subject to path degeneracy effects as d increases. The Local SMC filter with block of dimension 1 (i.e., $B_n = d$) gives the smallest variance, but at the expense of a nonnegligible bias, due to the approximation of the posterior as a product of marginals on each block—Eq. (2.20). As an example, the bias obtained with this technique is higher than the one from the classical SMC when $d = 10$. Finally, the SMCMC algorithm that uses an MCMC kernel with $P = d$ gives the smallest bias and reasonable variance. Let us remark that both the bias and the

Table 2.2 Statistical properties of the estimator of the posterior mean—time and space average of the absolute value of the bias and the variance across 100 MC runs

Algorithm			$d = 10$		$d = 50$		$d = 100$	
			Bias	Var	Bias	Var	Bias	Var
SMC			0.0481	0.0657	0.4460	0.7689	0.5536	1.0430
SMC-MCMC	$M = 1$	$P = d/5$	0.0426	0.0507	0.3362	0.3703	0.4062	0.4369
		$P = d$	0.0267	0.0206	0.2324	0.1885	0.2841	0.2366
	$M = 10$	$P = d/5$	0.0282	0.0368	0.1225	0.1452	0.1660	0.2110
		$P = d$	0.0141	0.0139	0.0602	0.0610	0.0929	0.1009
Local SMC	$B_n = d/5$		0.0541	0.0101	0.0594	0.0133	0.0620	0.0129
	$B_n = d$		0.0759	**0.0009**	0.0584	**0.0012**	0.0603	**0.0012**
STPF			0.0095	0.0114	0.0502	0.1017	0.0730	0.1778
SMCMC	$P = d/5$		0.0074	0.0087	0.0211	0.0235	0.0416	0.0419
	$P = d$		**0.0038**	0.0026	**0.0162**	0.0169	**0.0388**	0.0366

variance tends theoretically to zero asymptotically with the number of particles for all the methods, except for the local SMC filter.

2.6.2 Two-Dimensional Graph Model

In this section, we consider a two-dimensional graph that has been used in both [3, 36] to assess the performances of the different algorithms. Let the components of state x_n be indexed by vertices $v \in V$, where $V = \{1, \ldots, \sqrt{d}\}^2$. The dimension of the model is thus d. At time step n, the prior distribution at vertex v follows the following mixture distribution:

$$f(x_n(v)|x_{n-1}) = \sum_{u \in N(v)} w_u(v) f_u(x_n(v)|x_{n-1}(u)), \tag{2.41}$$

where $N(v) = \{u : D(u, v) \leq r\}$ corresponds to the neighborhood of vertex v with $r \geq 1$ and $D(u, v) = \sqrt{(a-c)^2 + (b-d)^2}$ the Euclidean distance between the two vertices $v = (a, b)$ and $u = (c, d)$. The observations are

$$Y_n(v) = X_n(v) + \eta_n(v), \tag{2.42}$$

for $v \in V$ where $\eta_n(v)$ are i.i.d. t-distributed random variables with degree of freedom ν. In the simulation experiments, a Gaussian mixture is used with component mean $X_{n-1}(u)$ and unity variance. The mixture weights are set to be $w_u(v) \propto 1/(D(u, v) + \delta)$ such that $\sum_{u \in N(v)} w_u(v) = 1$. Finally, the data has been generated by using $r = 1$, $\delta = 1$ and $\nu = 10$.

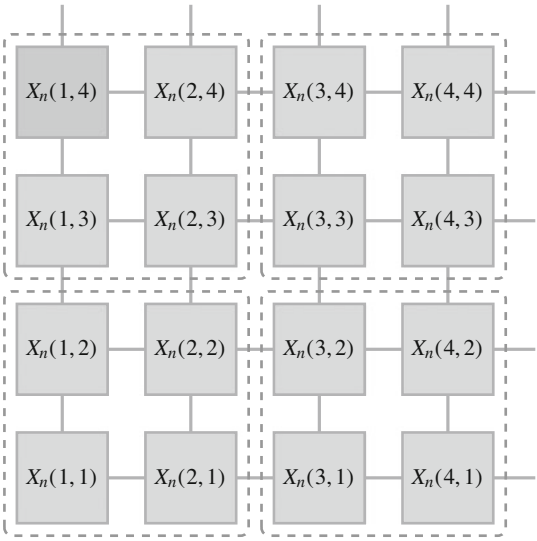


Fig. 2.7 Illustration of the block partitioning of the state for the local SMC in the two-dimensional graph example

Table 2.3 Mean squared error of the posterior mean averaged over time, space and 100 runs with $d = 144$

	SMC-MCMC				Local SMC			SMCMC	
SMC	$M = 1$		$M = 10$		$B_n = d/9$	$B_n = d$	STPF	$P = d/9$	$P = d$
	$P = d/9$	$P = d$	$P = d/9$	$P = d$					
3154	2266	1392	1154	261	617	152	174	344	**151**

From this model description, it is straightforward to see that the product of the likelihood and the prior can be factorized as in Eq. (2.19) with $|D_{n,j}| = 1, \forall n, j$. As a consequence, we define the partitioning of the subset in the STPF such that $\forall n, j$: $|A_{n,j}| = 1$. For the local SMC filter, the space is partitioned such that each block is itself a square as illustrated in Fig. 2.7. The same configuration concerning the number of particles shown in Table 2.1 is used in this example. Table 2.3 shows the mean squared error for the posterior mean obtained using all the Monte-Carlo algorithms under different settings. The SMCMC ($P = d$) and the local SMC ($B_n = d$) give similar performance and outperform slightly the STPF and more significantly the other algorithms. Once again, the introduction of the sequence of bridging densities within the SMC-MCMC clearly improves the performance of the estimators.

2.7 Conclusion

In this chapter, after describing the generic framework of traditional SMC methods for the optimal filtering problem in a general HMM, we discuss their limitation when applied to high-dimensional systems. We thus provide an overview of recent Monte-Carlo-based approaches that have been proposed in order to improve the performance of such approaches in high-dimensional systems. Through two examples, we have shown empirically that the use of these recent developments could lead to a significant improvement. It is however difficult to state that in general case one technique would be better than another. Indeed, the choice will be clearly dependent on the model which is under study and on possible constraints like computing resources available, storage capacity, or desired level of accuracy. A more detailed analysis of all these algorithms with a finite number of samples will be clearly interesting for comparison purpose and could be used to design some automatic strategy to select the "optimal" algorithm and its parameters given the model and the constraints.

References

1. Anderson, T.W.: The Statistical Analysis of Time Series, vol. 19. Wiley, New York (2011)
2. Berzuini, C., Best, N.G., Gilks, W.R., Larizza, C.: Dynamic conditional independence models and Markov chain Monte Carlo methods. J. Am. Stat. Assoc. **92**(440), 1403–1412 (1997)
3. Beskos, A., Crisan, D., Jasra, A., Kamatani, K., Zhou, Y.: A Stable Particle Filter in High-Dimensions (2014). arXiv.org
4. Bickel, P., Li, B., Bengtsson, T.: Sharp failure rates for the bootstrap particle filter in high dimensions. Insti. Math. Stat. Collect. **3**, 318–329 (2008)
5. Brockwell, A., Del Moral, P., Doucet, A.: Sequentially interacting Markov chain Monte Carlo methods. Ann. Statist. **38**(6), 3387–3411 (2010)
6. Cappé, O., Godsill, S., Moulines, E.: An overview of existing methods and recent advances in sequential Monte Carlo. Proc. IEEE **95**(5), 899–924 (2007)
7. Cappe, O., Moulines, E., Ryden, T.: Inference in Hidden Markov Models (2005)
8. Chopin, N.: Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. Ann. Stat. 2385–2411 (2004)
9. Crisan, D., Doucet, A.: A survey of convergence results on particle filtering methods for practitioners. IEEE Trans. Sig. Process. **50**(3), 736–746 (2002)
10. Del Moral, P.: Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications. Springer, New York (2004)
11. Del Moral, P., Doucet, A., Jasra, A.: Sequential Monte Carlo samplers. J. Roy. Stat. Soc. Ser. B (Stat. Methodol.) **68**(3), 411–436 (2006)
12. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. IEEE Conf. Comput. Vis. Patt. Recogn. **2000**, 126–133 (2000)
13. Djuric, P., Bugallo, M.F.: Particle filtering for high-dimensional systems. In: IEEE 5th International Workshop on Computational Avances in Multi-Sensor Adaptive Processing Adaptive Processing (CAMSAP), pp. 352–355 (2013)
14. Djuric, P., Lu, T., Bugallo, M.F.: Multiple particle filtering. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2007, ICASSP (2007)
15. Doucet, A., De Freitas, N., Gordon, N. (eds.).: Sequential Monte Carlo Methods in Practice. Springer, New York (2001)

16. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte-Carlo sampling methods for Bayesian filtering. Stat. Comput. **10**, 197–208 (2000)
17. Gall, J., Potthoff, J., Schnörr, C., Rosenhahn, B., Seidel, H.P.: Interacting and annealing particle filters: mathematics and a recipe for applications. J. Math. Imaging Vis. **28**(1), 1–18 (2014)
18. Geweke, J., et al.: Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments, vol. 196. Federal Reserve Bank of Minneapolis, Research Department (1991)
19. Geyer, C.J.: Markov chain Monte Carlo maximum likelihood. In: Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, pp. 156–163 (1991)
20. Gilks, W.R., Berzuini, C.: Following a moving target-Monte Carlo inference for dynamic Bayesian models. J. Roy. Stat. Soc. Ser. B (Stat. Methodol.) **63**, 127–146 (2001)
21. Godsill, S.J., Clapp, T.: Improvement strategies for Monte Carlo particle filters. In: Doucet, A., De Freitas, N., Gordon, N. (eds.) Sequential Monte Carlo Methods in Practice. Springer, Berlin(2001)
22. Golightly, A., Wilkinson, D.: Bayesian sequential inference for nonlinear multivariate diffusions. Stat. Comput. **16**(4), 323–338 (2006)
23. Gordon, N., Salmond, D., Smith, A.F.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F. Radar. Sig. Process. **140**, 107–113 (1993)
24. Jones, G.L., et al.: On the Markov chain central limit theorem. Probab. Surv. **1**, 299–320 (2004)
25. Julier, S.J., Uhlmann, J.K.: Unscented filtering and nonlinear estimation. In: Proceedings of the IEEE, pp. 401–422 (2004)
26. Kalman, R.E.: A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. **82**, 35–45 (1960)
27. Khan, Z., Balch, T., Dellaert, F.: MCMC-Based particle filtering for tracking a variable number of interacting targets. IEEE Trans. Pattern Anal. Mach. Intell. **27**(11), 1805–1819 (2005)
28. Kong, A., Liu, J.S., Wong, W.H.: Sequential imputations and bayesian missing data problems. J. Am. Stat. Assoc. **89**(425), 278–288 (1994)
29. Künsch, H.R.: Recursive Monte Carlo filters: algorithms and theoretical analysis. Ann. Stat. 1983–2021 (2005)
30. Leith, C.: The standard error of time-average estimates of climatic means. J. Appl. Meteorol. **12**(6), 1066–1069 (1973)
31. Liang, F., Wong, W.H.: Evolutionary monte carlo: applications to C_p Model sampling and change point problem. Stat. Sinica **10**, 317–342 (2000)
32. Mihaylova, L., Hegyi, A., Gning, A., Boel, R.K.: Parallelized particle and gaussian sum particle filters for large-scale freeway traffic systems. IEEE Trans. Intell. Transp. Syst. **13**(1), 36–48 (2012)
33. Neal, R.: Annealed importance sampling. Stat. Comput. 125–139 (2001)
34. Peters, G.W.: Topics in Sequential Monte Carlo Samplers. Master's thesis, University of Cambridge (2005)
35. Rebeschini, P.: Nonlinear Filtering in High Dimension. Ph.D. thesis, Princeton University (2014)
36. Rebeschini, P., van Handel, R.: Can Local Particle Filters Beat the Curse of Dimensionality? arXiv.org (2013)
37. Ristic, B., Arulampalam, S., Gordon, N.: Beyond the Kalman filter: particle filters for tracking applications. Artech House (2004)
38. Robert, C.P., Casella, G.: Monte Carlo Statistical Methods. Springer, Berlin (2004)
39. Septier, F., Carmi, A., Pang, S., Godsill, S.: Multiple object tracking using evolutionary and hybrid MCMC-based particle algorithms. In: 15th IFAC Symposium on System Identification, (SYSID 2009). Saint Malo, France (2009)
40. Septier, F., Pang, S., Carmi, A., Godsill, S.: On MCMC-Based particle methods for bayesian filtering : application to multitarget tracking. In: International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2009). Aruba, Dutch Antilles (2009)

41. Snyder, C.: Particle filters, the"optimal" proposal and high-dimensional systems. In: ECMWF Seminar on Data Assimilation for Atmosphere and Ocean, pp. 1–10 (2011)
42. Snyder, C., Bengtsson, T., Bickel, P., Anderson, J.: Obstacles to high-dimensional particle filtering. Monthly Weather Rev., Spec. Collect.: Math. Adv. Data Assimilation **136**(12), 4629–4640 (2008)
43. Vergé, C., Dubarry, C., Del Moral, P., Moulines, E.: On Parallel Implementation of Sequential Monte Carlo Methods: The Island Particle Model. arXiv.org (2013)

Chapter 3
Spectral Measures of α-Stable Distributions: An Overview and Natural Applications in Wireless Communications

Nourddine Azzaoui, Laurent Clavier, Arnaud Guillin and Gareth W. Peters

Abstract Currently, we are witnessing the proliferation of wireless sensor networks and the superposition of several communicating objects which have a heterogeneous nature. Those are merely the beginnings of an evolution toward the so-called Internet of Things. The advent of these networks as well as the increasing demand for improved quality and services will increase the complexity of communications and put a strain on current techniques and models. Indeed, they must first adapt to the temporal and spatial evolutions and second, they must take into account the rare and unpredictable events that can have disastrous consequences for decision-making. This chapter provides an overview of the various spectral techniques used in signal processing and statistics literature to describe a communication channel having an impulsive behavior. This project is mainly motivated by the historical success of the interaction between probability, statistics and the world of communications, information theory and signal processing. The second motivation is the scarcity of references and literature summarizing mathematical developments on the application of alpha-stable process for channel modeling. This chapter will be divided into two parts: the first is devoted to the synthesis of various developments on alpha-stable variables and processes in a purely mathematical mind. The second part will be devoted to applications in the context of communications. The two sides will

N. Azzaoui (✉) · A. Guillin
Laboratoire de Mathematiques, Université Blaise Pascal, Aubiere, France
e-mail: nourddine.azzaoui@math.univ-bpclermont.fr

A. Guillin
e-mail: guillin@math.univ-bpclermont.fr

L. Clavier
Institut Mines-Télécom/Télécom Lille/IRCICA FR CNRS 3024, Villeneuve-d'Ascq, France
e-mail: laurent.clavier@telecom-lille.fr

G.W. Peters
Department of Statistical Science, University College London, London, UK
e-mail: gareth.peters@ucl.ac.uk

G.W. Peters and T. Matsui (eds.), *Theoretical Aspects of Spatial-Temporal Modeling*,
JSS Research Series in Statistics, DOI 10.1007/978-4-431-55336-6_3

combine two fundamentally linked aspects: first, a theoretical approach, necessary for a good formalization of problems and identifying the best solutions. Second, the use of these models in real work of channel modeling.

3.1 Short Review and Introduction

One of the problems addressed by researchers in probability and statistics during the eighteenth and nineteenth centuries, was to find the best fit of a set of observed data to a given equation (modeling problems). After several failures, this was accomplished through least squares methods; works of Laplace and Legendre were the most influential in this area. For the error distribution, Gauss highlighted the importance of the normal distribution, also subsequently named in his honor the Gaussian distribution. After the development of Fourier theory at the end of 1800, Poisson applied Fourier series and integrals to probability distributions as a new natural tool in analysis. At the end of 1850, Cauchy, who was still a student supervised by Laplace, was interested in the theory of errors and generalized the density of the Gaussian distribution by the following function:

$$f_\alpha(x) = \int_0^\infty e^{-ct^\alpha} \cos(tx)dt; \quad \alpha \in \mathbb{R}^+$$

where t^2 is replaced by t^α. He was able to calculate this integral for $\alpha = 1$: $f_1(x) = \frac{c}{\pi(c^2+x^2)}$, density of the famous Cauchy law. Then, after a long silence until 1919, Bernstein showed that f_α is a probability density only when $0 < \alpha \leq 2$.

Later, in 1924, Lévy [50] updated this research area by introducing the theory of α-stable distributions. He looked for more general conditions on central limit theorem validity. He showed that an α-stable distribution can substitute a Gaussian distribution for modeling infinite variance phenomena. Multivariate α-stable laws were then studied and developed by Lévy and Khinchine [43] and Gnedenko [34], who explored the properties of multivariate α-stable distribution, with particular emphasis on stability properties by product of convolution and the central limit theorem. Another convenient characterisation of multivariate α-stable vectors is given by their characteristic function. With contrast to the probability density, this characteristic function was given under an analytical form involving a spectral density on the $\mathbb{R}^d$ sphere, see for instance Feldheim [29], Lévy [51], Feller [30].

Spectral analysis of stationary processes has a long history with interest in both theory and applications. Wide-ranging applications in various practical problems in engineering, economics, science, and medicine are well documented. In general, spectral analysis is considered as a powerful tool when undertaking a statistical treatment of stochastic processes. Its strengths lie in the fact that it focuses on the repetitive components or frequencies of these processes: this means that, by contrast to temporal processing, it allows to reveal the mixture of repeated information hidden

in the process realizations. The basic theory of this technique dates back to the work of Fourier who, in his famous decomposition (so-called Fourier series), expresses each deterministic function as a linear combination of trigonometric functions. At the beginning of the twentieth century, this technique has been exported for processing of random functions or stochastic processes; that is to say a measurable mapping defined by:

$$X : \begin{array}{ccc} \mathbb{R} \times \Omega & \longrightarrow & \mathbb{C} \\ (t, \omega) & \longmapsto & X(t, \omega) = X_t(\omega) \end{array}$$

where $(\Omega, \mathcal{B}, \mathbb{P})$ is a probability space in σ-algebra $\mathcal{B}$. Early works about this extension were not possible without the pioneering work of Kolmogorov [44, 45]; who gave the foundation of modern probability theory and his famous construction of stochastic processes known as the Kolmogorov theorem. The main idea of spectral representation of some classes of random variables, is to find a correspondence between this class and a functional space. This idea was also introduced by Kolmogorov. The latter found an isometric correspondence between the vector space generated by Gaussian stationary processes and the vector space of real square integrable functions.

Wiener [84] was at the origin of modern spectral analysis theory when he published in the thirties, his paper *Generalized Harmonic Analysis*. Among his contributions he had given the precise statistical definition of the autocovariance function and the spectral density of second-order stationary random processes (stationary in the strong sense). In 1934, Khinchine [42] was the first to define the concept of weak stationarity through the covariance or correlation function. Thus, a continuous time stochastic process X is weakly stationary if its covariance function defined by:

$$r : \begin{array}{ccc} \mathbb{R} \times \mathbb{R} & \longrightarrow & \mathbb{C} \\ (s, t) & \longmapsto & r(s, t) = \operatorname{cov}(X_s, X_t) \end{array} \tag{3.1}$$

is continuous and depends only on the time difference $|s - t|$. In this case, the function r is reduced to a single variable function and is continuous and positive definite. By applying Bochner's theorem [13], this function r can be represented as the Fourier transform of a bounded positive measure F:

$$r(t) = \int_{\mathbb{R}} e^{it\lambda} F(d\lambda). \tag{3.2}$$

The measure F is called the spectral measure of X. From this last representation (3.2), several results concerning the structure of second-order processes were established. Note, for example, the Cramer–Kolmogorov theorem which states that a stationary stochastic process with covariance function satisfying (3.2) has the following stochastic integral representation:

$$X_t = \int_{\mathbb{R}} e^{it\lambda} d\xi, \tag{3.3}$$

where $d\xi$ is a random measure having orthogonal increments[1] defined on the Borel set $\mathcal{B}(\mathbb{R})$, see for example Rao [68]. These weakly stationary processes found important applications in several fields; between others, meteorology, communication, electrical engineering. . ., etc. These applications are cited in several literature references, as examples see [1, 4, 8, 15, 24, 25, 30].

To make the spectral theory accessible to researchers in applied fields, it has been necessary to introduce statistical tools that address the spectral estimation. Tukey [80, 81] is the founder of modern empirical spectral analysis: in 1949 he gave the fundamentals of spectral estimation by developing methods to estimate the autocorrelation function from a sample taken from the observations of a stationary process. It should be noted that most of the terms and spectral estimation techniques such as *aliasing, smoothing, tapering …etc.* are attributed to Tuckey. Among the most commonly used spectral estimation techniques, the periodogram is the most important feature for the estimation of the spectral density: this is a spectral estimation of the Fourier coefficients from observations of a process. The periodogram was introduced at the end of the nineteenth century and was used to detect hidden periodicities of famous sunspots observations. Among the most influential work on spectral estimation of second-order stationary process, we find Parzen [64, 65], Rosenblatt [69], Anderson [4], Masry [55–57], Priestley [66]. For the nonstationary process, we cite among others: Priestley [66, 67], Dahlhaus [25].

This large amount of works on second-order processes, whether stationary or not, was conceivable to thank the Hilbertian structure of the covariance which exhibits several nice algebraic properties. In several practical and theoretical situations, researchers were requested to deal with stochastic processes which are not necessarily the second order. This necessarily implies that the results obtained cannot be used in this case. To try to solve this problem even partially, several probabilistic and statisticians have studied the class of stochastic process α-stable.

An important notion in the analysis of random vectors is the concept of independence or the existence of a statistical link between their elements. For example, the covariance matrix of a Gaussian vector is an inevitable tool used to describe the dependency structure through this matrix. Besides, covariance determines entirely the distribution of a Gaussian centered vector. For this reason, it plays an essential role in several research areas covering both theoretical and applied problems, notably in most theories of the statistical treatment of processes and time series. Of course, covariance is not defined for stable random variables because their second-order moments are infinite. To remedy this problem and to introduce a dependency measure compatible with the α-stable random vectors, Miller [59] was the first to introduce the term *covariation*[2] as a generalized concept of the covariance. He had shown that in spite of its asymmetry and its non-bilinearity it can play, in some cases, the same role as covariance, see also [58].

When considering α-stable random vectors of infinite dimension, one speaks about α-stable stochastic processes. This concept was widely studied in literature

[1] This is to say that ξ is a σ-additive application verifying $\text{cov}(\xi(A), \xi(B)) = F(A \cap B)$.

[2] This measure of dependency had been used by Kanter [41] but not under the name of covariation.

and used in several application domains, notably when it is about the study of the random phenomena which occur in time and that are characterized by a significant variability [1, 2, 5, 9, 40, 54, 76]. Their utility was proven in several domains of application, especially in signal processing [62], in finance [2, 12]. . . Definition and development of the stochastic integral allowed several authors to extend most properties of Gaussian processes to the class of α-stable processes. The stochastic integral with respect to α-stable motions (sometimes said Lévy processes) or more generally with respect to an α-stable stochastic measure, was studied by Hardin [36], Hosoya [38], Cambanis and Miamee [17], Makagon and Manderkar [53], Samorodnitsky and Taqqu [76]. The utility of the stochastic integral is that it allows to link up these processes with a functional space: this notion is known under the name of the spectral analysis of processes. In the α-stable case first works in this sense were those of Bretagnolle et al. [14] then Miller and Cambanis [16, 18].

3.2 α-Stable Random Variables, Vectors, and Processes

3.2.1 Univariate α-Stable Random Variables and Vectors

In this part we recall the classical definitions of a stable distribution and we give the practical interpretation of its parameters to clarify the statistical properties of variables generated from this distribution. First two definitions concern the convolution stability property, which means that the family of stable distribution is preserved by convolution. The third definition explains the role of stable distribution in the context of the central limit theorem, that is a stable distribution can be approached by a normalized sum of independent and identically distributed random variables; this is also known as stable domain of attraction. This property is important for the use of α-stable distributions in modeling. Finally the fourth definition specifies the characteristic function of an α-stable random variable. This characteristic function is explained in an analytical precise manner.

Definitions and preliminary results

Definition 3.2.1 A random variable X is stable (or has a stable distribution) if and only if for any positive reals A and B, there exists a positive real C and a real D such that the distribution of the random variable $AX_1 + BX_2$ is equal to that of $CX + D$; whenever X_1 and X_2 are independent random copies of X.

In order to understand the meaning of the word emphstability in this context, let us recall that the distribution of the sum of two independent random variables is the convolution of respective distributions of these two variables. In fact the word stable in the last Definition 3.2.1 comes from the fact that the set of probability distribution is stable under convolution. It is also shown in Feller [30] that for any stable random variable X, there exists a unique real $\alpha \in]0, 2]$, such as the positive real numbers A, B, and C fulfill the equality,

$$A^{\alpha} + B^{\alpha} = C^{\alpha}. \tag{3.4}$$

The number α is unique and depends only on the probability distribution of the random variable X. This result justifies the use of the prefix α in the denomination α-stable. Property (3.4) can be generalized to the sum of a finite number of independent copies of an α-stable variable.

Definition 3.2.2 A random variable X is stable if and only if for any integer $n \geq 2$ there exists a positive real C_n and a real D_n such as the probability distribution of the sum $X_1 + X_2 + \cdots\cdots + X_n$ is the same as that of $C_n X + D_n$. Whenever X_1, X_2, X_n are independent random copies of X.

Using property (3.4) and proceeding by induction, it is easy to see that the constant, $C_n = n^{\frac{1}{\alpha}}$, where α is defined in relation (3.4). Particularly in the Gaussian case where $\alpha = 2$, we have $C_n = \sqrt{n}$; this quantity is directly linked to the rate of convergence in the classical central limit theorem. One of the remarkable properties of stable variables is the possibility to explain them as a limit of independent and identically distributed (i.i.d) random variables. This is known as the generalized central limit theorem.

Definition 3.2.3 A random variable X has a stable distribution if and only if it has an attraction domain. This means the existence of i.i.d series of random variables $(Y_i)_i$, a set of positive numbers D_n, and a set of real numbers A_n, such as: $\frac{Y_1 + Y_2 + \cdots\cdots + Y_n}{D_n} + A_n$ converges in law toward X.

It is classically known that the Definitions 3.2.1–3.2.3, are equivalents. For more details see for instance Feller [30] or [76]. The last definition gives a stable random variable as a normalized sum of (*i.i.d.*) variables. Let us point out that this result is very useful in practice because it allows to approach a normalized sum $i.i.d$ sequence by a stable random variable. It generalizes the classical central limit theorem that requires a finite variance (Gaussian limit). These three last definitions introduce α-stable variables in an abstract way and give no precision about the likelihood or the probability density function of these variables. The most used characterisation of stable variables, especially in practice, is based on the definition of the characteristic function which is nothing but the Fourier transform of its probability density function.

Definition 3.2.4 A random variable X has a stable distribution if and only if there exist four unique parameters: $0 < \alpha \leq 2, \sigma \geq 0, -1 \leq \beta \leq 1$, and a real μ such that the characteristic function of X has the form:

$$\mathbb{E}e^{\imath\theta X} = \begin{cases} \exp\{-\sigma^{\alpha}|\theta|^{\alpha}(1 - \imath\beta(\text{sign}(\theta))\tan\frac{\pi\alpha}{2}) + \imath\mu\theta\}, & \text{if } \alpha \neq 1, \\ \exp\{-\sigma|\theta|(1 + \imath\beta\frac{2}{\pi}(\text{sign}(\theta))\ln|\theta|) + \imath\mu\theta\}, & \text{if } \alpha = 1, \end{cases} \tag{3.5}$$

with

$$\text{sign}(\theta) = \begin{cases} 1, & if\ \theta > 0, \\ 0, & if\ \theta = 0, \\ -1, & if\ \theta < 0. \end{cases} \tag{3.6}$$

We follow the notations of Samorodnitsky and Taqqu [76], where the distribution will be noted $S_\alpha(\sigma, \beta, \mu)$.

One of the strengths of this characterisation is that the four parameters α, σ, β, and μ are sufficient to characterize a stable distribution in a unique way, meaning that the stable laws admit a parametric representation. For practitioners, it is important to know the statistical meaning of each parameter as well as its influence on the shape of the density or the distribution.

- The parameter α is called the stability index; it measures the way the tail of distribution function will go to zero when x goes to infinity (i.e., like $x^{-\alpha}$). This implies that, when α is smaller, the realizations of an α-stable random variable become more impulsive and more variable.
- The parameter σ is called the scale or dispersion parameter; its role is similar to the standard deviation of a normal variable.
- The skewness parameter β takes values in the interval $[-1, 1]$ and it measures how asymmetric the density of a stable random variable will be. For example, a positive value of β results in gap on the right of the density curve.
- The parameter μ is the position parameter; it is the central points around which most realizations lie. For example, for $1 < \alpha \leq 2$ it is the mean while for $0 < \alpha < 1$ it represents the median.

Probability density function and Likelihood of stable samples

For α-stable random variables, there is no explicit expression for the probability density function in a general case. However, it is possible to get an expression via the inverse Fourier transform of the characteristic function,

$$\begin{aligned} f(x, \alpha, \sigma, \beta, \mu) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\imath t x) \Phi_X(t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{-\imath t(x - \mu) - |\sigma t|^\alpha \psi(t)\} dt \end{aligned} \tag{3.7}$$

where $\psi(t) = 1 - \imath\beta \text{sign}(t) \tan\frac{\pi\alpha}{2}$. The distribution is called symmetric if $(\beta = 0)$ and $(\mu = 0)$; the characteristic function is therefore real and even. These properties are used to simplify the expression of the probability density function in (3.7) that can be written as:

$$f(x, \alpha, \sigma) = \frac{1}{\pi} \int_0^{\infty} \exp(-\sigma^\alpha |t|^\alpha) \cos(tx) dt. \tag{3.8}$$

The exact evaluation of the integral (3.50) is only possible in three special cases: The Lévy distribution corresponding $\alpha = \frac{1}{2}$, Cauchy distribution when $\alpha = 1$, and Gaussian distribution when $\alpha = 2$. Except for these three particular laws, the probability density function of an α-stable random variable has no exact analytical expression. However, using the integral representation (3.50) with $\sigma = 1$ (without loss of generality), a series expansion has been introduced into the literature by Bergström [11],

$$f(x,\alpha) = \frac{1}{\pi\alpha}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k!}\Gamma(\frac{2k+1}{\alpha})x^{2k}, \tag{3.9}$$

where Γ is the common Gamma function defined, for $x > 0$, by

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt. \tag{3.10}$$

The series expansion (3.9) is discussed in several books and works about α-stable laws, we cite among others Feller [30] or [76]. The problem with the infinite sum (3.9) is that it contains increasing gamma terms with alternating signs; this makes it difficult to use in practice or to give an approximation for every x. The first attempt to solve this problem was introduced by, Bergström [11] for SαS random variables with $\alpha > 1$. He gave an asymptotic approximation of (3.9), for $x = 0$ given by

$$f(x,\alpha) = \frac{1}{\pi\alpha}\sum_{k=0}^{n}\frac{(-1)^k}{2k!}\Gamma(\frac{2k+1}{\alpha})x^{2k} + O(|x^{2n+1}|), \tag{3.11}$$

and when x goes to infinity, we have the approximation:

$$f(x,\alpha) = \frac{-1}{\pi}\sum_{k=0}^{n}\frac{(-1)^k}{k!}\Gamma(\alpha k+1)\frac{\sin(k\alpha\pi/2)}{|x|^{\alpha k+1}} + O\left(|x|^{-\alpha(n+1)-1)}\right). \tag{3.12}$$

Calculating these asymptotic series for large values of n is tricky due to the evaluation of the gamma function. These difficulties can be overcome by following the procedure proposed in [[62], p. 17]. As one can see easily from the asymptotic decomposition (3.12), the distribution function $F(x)$ of an α-stable variable decreases to 0 at rate $x^{-\alpha}$ when x goes to infinity. The distribution is slowly decreasing *heavy-tailed distribution*. In the next definition, we recall some relevant notions of heavy-tailed functions.

Definition 3.2.5 The probability distribution of a real random variable is said to be *Heavy-tailed*, of index α if there exist a slowly varying function h, that is to say $\lim_{x\to\infty}\frac{h(bx)}{h(x)} = 1$ for any $b \geq 0$, such that:

$$\mathbb{P}(X \geq x) = x^{-\alpha}h(x). \tag{3.13}$$

As mentioned before, α-stable variables are *heavy-tailed distributions*. In fact, this result is summed up in the following proposition:

Proposition 3.2.1 *Let X be an $S_\alpha(\sigma, \beta, \mu)$ random variable with $0 < \alpha < 2$, then we have the following results:*

$$\lim_{t \longrightarrow \infty} t^\alpha \mathbb{P}(X > t) = \sigma^\alpha . C_\alpha . \frac{1+\beta}{2} \text{ and } \lim_{t \longrightarrow \infty} t^\alpha \mathbb{P}(X < -t) = \sigma^\alpha . C_\alpha . \frac{1-\beta}{2}$$

where C_α is a constant depending only on α and is given by, $\frac{1-\alpha}{\Gamma(2-\alpha)\cos(\frac{\pi\alpha}{2})}$ when $\alpha \neq 1$ and $C_1 = \frac{2}{\pi}$

For the proof of these results, see for instance [76, p. 16].

Algebraic properties and moments of stable distributions

Proposition 3.2.2 *Let X_1 and X_2 be two independent α-stable random variables having, respectively, the distributions $S_\alpha(\sigma_1, \beta_1, \mu_1)$ and $S_\alpha(\sigma_2, \beta_2, \mu_2)$, then we have following properties:*

- *The random variable $Y = X_1 + X_2$ is also α-stable $S_\alpha(\sigma, \beta, \mu)$ with:*

$$\sigma^\alpha = \sigma_1^\alpha + \sigma_2^\alpha, \beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \mu = \mu_1 + \mu_2. \tag{3.14}$$

- *For any real numbers a and b, the random variable $aX_1 + b$ is also α-stable and have the distribution $S_\alpha(|a|.\sigma_1, \text{sign}(a).\beta_1, \mu_1 + b)$.*
- *If X is an α-stable variable then its moments of order greater than α are infinite, that is to say:*

$$\mathbb{E}|X^r| < \infty \quad \text{if} \quad r < \alpha \tag{3.15}$$

$$\mathbb{E}|X|^r = \infty \quad \text{if} \quad r \geq \alpha. \tag{3.16}$$

- *If X is an α-stable random variable $S_\alpha(\sigma, \beta, \mu)$ with $\alpha > 1$ then we have $\mathbb{E}(X) = \mu$.*

The proofs of these results are detailed in most books on α-stable laws, see for example [40, 76]. As we have seen in (3.16), moments of order α are infinite but one can see the way these moments go to infinity when the order of the moments increases to α, we summarize this in the following.

Proposition 3.2.3 *Let X an $S_\alpha(\sigma, \beta, \mu)$ random variable, then we have the following limits:*

$$\lim_{R \nearrow \alpha} (\alpha - r)\mathbb{E}|X|^r = \alpha C_\alpha \sigma^\alpha, \tag{3.17}$$

$$\lim_{R \nearrow \alpha} (\alpha - r)\mathbb{E}X^{<r>} = \alpha\beta C_\alpha \sigma^\alpha, \tag{3.18}$$

where C_α is defined in Proposition (3.2.1) and $x^{<\alpha>} = |x|^\alpha \, sign(x)$.

Lepage series representation of an α-stable random variable

In general the series representation of a random variable that is infinitely divisible[3] without Gaussian component was established by Ferguson and Klass [31], then further developed by Lepage [49]. The extension of this representation in more general situations has been studied by Rosinski [71]. The idea of this representation consists in writing an α-stable random variable as an infinite sum involving independent random variables and arrival times of a Poisson process. The usefulness of this decomposition is that it allows the proof of several theoretical results, for example [76]. It can also be used to generate an α-stable variable; but due to its slow convergence many authors advice to instead use direct simulation methods introduced in [20]. In this chapter, we discuss only the particular case of SαS random variables for their practical use and similarity to Gaussian variables. To state the main result of Lepage, let $(\varepsilon_i)_{i\in\mathbb{N}}$, $(W_i)_{i\in\mathbb{N}}$ and $(\Gamma_i)_{i\in\mathbb{N}}$ denote three independent sequences of random variables defined as follows:

- The random variables $\varepsilon_1, \varepsilon_2, \varepsilon_n \ldots$ are independent and identically distributed having a Rademacher distribution: they are concentrated on 1 and -1, such that $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = \frac{1}{2}$.
- The variables $\Gamma_1, \Gamma_2, \Gamma_n \ldots$. are arrival times of a Poisson process with intensity 1. They follow a gamma distribution parameter i and they are dependent:

$$\Gamma_i = E_1 + \cdots + E_i.$$

  The random variables $(E_k)_k$ are independent and identically distributed of an exponential distribution of parameter 1.
- $W_1, W_2, W_n, \cdots$ are independent and identically distributed random variables.

Proposition 3.2.4 *If the random variable W has a finite moment of order α, that is to say $\mathbb{E}|W|^\alpha < \infty$ then for any $0 \leq \alpha < 2$, the series $\sum_{I=1}^{\infty} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i$ converges almost surely to an SαS random variable of parameters $\sigma = C_\alpha^{-1}[E|W_1|^\alpha]^{\frac{1}{\alpha}}$, with C_α given in Proposition (3.2.1).*

For the proof see [76] or [40]. Conversely, each symmetric α-stable random variable (SαS) admits a series decomposition given in Proposition (3.2.4).

[3]The family of distribution or infinitely divisible random variables is a more general class of distribution generalizing α-stable distribution. This distribution is characterized between others due to the fact that they prove limit central theorem and due to the fact that they have attraction domains.

Proposition 3.2.5 *Let X be an SαS random variable such as $X \sim S_\alpha(\sigma, 0, 0)$ then*

$$X \stackrel{d}{=} (C_\alpha)^{\frac{1}{\alpha}} \sum_{I=1}^{\infty} \varepsilon_i \Gamma_i^{-\frac{1}{\alpha}} W_i, \tag{3.19}$$

with $\sigma = (\mathbb{E}|W|^\alpha)^{\frac{1}{\alpha}}$.

The proof of this result is detailed in [76]. One can notice that representation (3.19) is not unique because the arbitrary choice of W having a moment order α is equal to $\sigma^\alpha = \mathbb{E}|W|^\alpha < \infty$.

3.2.2 Multivariate α-Stable Random Variables

The definition of a multivariate stable distribution is the extension of the definition the univariate case which is naturally the invariance with respect to convolution product. In this section we will focus only on the main result concerning components of the vector, for instance, that the linear combinations of the elements of a stable vector is also stable. We highlight that this result is not true in general; but only when $\alpha > 1$ or if the vector is strictly stable.

Definition 3.2.6 A random vector $\mathrm{X} = (X_1, \ldots, X_d)$ is said to be α-stable distributed in $\mathbb{R}^d$ if for any positive numbers A and B, there is a positive number C and a vector D in $\mathbb{R}^d$ such as

$$AX^{(1)} + BX^{(2)} \stackrel{d}{=} CX + D, \tag{3.20}$$

where $X^{(1)}$ and $X^{(2)}$ are independent copies of the vector X. When $D = 0$ we talk about strictly stable law.

Definition 3.2.6 imposes conditions on the joint distribution of the random vector X. To highlight stability of the components of this vector, that the elements of X are also stable and more generally any combination of elements is also a stable random variable.

Theorem 3.2.6 *Let $X = (X_1, \ldots, X_d)$ be a stable random vector, then we have:*

- *Any linear combination of the components of X of form $Y = \sum_{K=1}^{d} b_k X_k$ is an α-stable variable.*
- *There is a unique $\alpha \in]0, 2]$ such that constants A, B, and C of Definition 3.2.6 satisfy: $A^\alpha + B^\alpha = C^\alpha$.*
- *A random vector $X = (X_1, \ldots, X_d)$ is α-stable in $\mathbb{R}^d$ with $0 < \alpha < 2$ if and only if for any $n \geq 2$ there is a vector D_n of $\mathbb{R}^d$ such as:*

$$X^{(1)} + \cdots + X^{(n)} \stackrel{d}{=} n^{\frac{1}{\alpha}} X + D_n, \tag{3.21}$$

where $X^{(1)}, \ldots, X^{(n)}$ are independent copies of X.

Theorem 3.2.6 shows that any linear combination of the components of a stable vector is also stable. It is known that the inverse is true in the case of Gaussian vectors but not for the general α-stable case.

Proposition 3.2.7 *Let X be a random vector in $\mathbb{R}^d$ then we have the following:*

- *If any linear combination $Y = \sum_{k=1}^{d} b_k X_k$ is strictly stable then the random vector X is also strictly stable.*
- *If any linear combination Y is symmetric stable, then X is also symmetric stable.*
- *If any linear combination Y is α-stable with a stability index greater or equal to 1 then the vector X is also α-stable.*
- *If X is infinitely divisible and if any linear combination of Y is stable then the vector X is also stable.*

Proof of these results are detailed in [76]. These last properties are difficult to use because they give no precision on the analytical expression of the probability distribution of an α-stable vector. Similarly to the univariate case, it is difficult to give an analytical expression of their probability density. But we can give the expression of their characteristic function, defined for all $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$, by:

$$\Phi_X(\theta) = \mathbb{E}\left(\exp(\imath(\theta, X))\right) = \mathbb{E}\left(\exp(\imath \sum_{k=1}^{d} \theta_k X_k)\right) \tag{3.22}$$

The following result was first shown by Feldheim [29]. For more details, see Kuelbs [47].

Theorem 3.2.8 *A random vector $X = (X_1, \ldots, X_d)$ is α-stable if and only if there exists unique measure Γ defined on the unit sphere*

$$S_d = \left\{s \in \mathbb{R}^d, \|s\| = s_1^2 + \cdots + s_d^2 = 1\right\},$$

and a real vector μ of $\mathbb{R}^d$ such that

- *if $\alpha \neq 1$ then,*

$$\Phi_X(\theta) = \exp\left\{-\int_{S_d} |(\theta, s)|^\alpha (1 - \imath \operatorname{sign}((\theta, s)) \tan(\frac{\pi\alpha}{2})) \Gamma(ds) + \imath(\theta, \mu)\right\}, \tag{3.23}$$

- *if $\alpha = 1$ then,*

$$\Phi_X(\theta) = \exp\left\{-\int_{S_d} |(\theta, s)|(1 - \imath \frac{2}{\pi} \operatorname{sign}((\theta, s)) \ln(|(\theta, s)|)) \Gamma(ds) + \imath(\theta, \mu)\right\}. \tag{3.24}$$

Measure Γ is called the spectral measure.

Unlike the univariate α-stable case, random vectors are not parametric but the couple (Γ, μ) is unique and characterizes the probability distribution of the vector X. We give some examples of characteristic functions of some common α-stable vectors.

Example 1

- *Link with univariate stable variables.*
  For $d = 1$ and $\alpha > 1$, the measure Γ is concentrated on the unit sphere of $\mathbb{R}$ which is formed by two points $\{1, -1\}$. On the one hand, using the characteristic function, given in (3.23), we have:

$$\begin{aligned}\Phi_X(\theta) &= \exp\left\{\Gamma(\{1\})|\theta.1|^\alpha(1 - \iota.\text{sign}(\theta)\tan(\frac{\pi\alpha}{2})) \right.\\ &\qquad \left. +\Gamma(\{-1\})|-\theta|^\alpha(1 + \iota.\text{sign}(\theta)\tan(\frac{\pi\alpha}{2}))\right\}\\ &= \exp\left\{|\theta.1|^\alpha(\Gamma(\{1\}) + \Gamma(\{-1\}) - \iota.\text{sign}(\Gamma(\{1\}) - \Gamma(\{-1\})))\right\},\end{aligned}$$

  on the other hand, by comparing this equality to (3.49), the characteristic function Φ_X of an α-stable random variable $S_\alpha(\sigma, \beta, \mu)$ verifies:

$$\sigma = \Gamma(\{1\}) + \Gamma(\{-1\}), \quad \beta = \frac{\Gamma(\{1\}) - \Gamma(\{-1\})}{\Gamma(\{1\}) + \Gamma(\{-1\})}.$$

  For instance, when the random variable X is symmetric ($\beta = 0$) then the spectral measure Γ is symmetric and we have $\Gamma(\{1\}) = \Gamma(\{-1\})$.
- *Symmetric α-stable vectors.*
  An α-stable random vector X is symmetric if X and $-X$ have the same probability distribution. This means that, for any borelian set A of $\mathbb{R}^d$ we have: $\mathbb{P}(X \in A) = \mathbb{P}(-X \in A)$. From this definition, we can give a characterisation of symmetric α-stable vectors from the structure of their spectral measure Γ. A vector X is then symmetric α-stable if and only if there exists a unique symmetric measure Γ defined on the unit sphere S_d such that

$$\Phi_X(\theta) = \exp\left\{-\int_{S_d} |(\theta, s)|^\alpha \Gamma(ds)\right\}, \quad \text{for every} \quad \theta \in \mathbb{R}^d. \tag{3.25}$$

  The support structure of the spectral measure Γ, allows to differentiate several kinds of α-stable vectors. A class of spectral measures which is often used in practice is the family of measures concentrated on countable sets of points on the unit sphere. For example, the independence of components of a stable vector means that its spectral measure is concentrated on special set of points. Indeed this spectral measure Γ is concentrated on the points $(\pm 1, 0, \ldots, 0), \ldots, (0, \ldots, \pm 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, \pm 1)$.

- *Sub-Gaussian random vectors.*
  A random vector X is said to be sub-Gaussian if it can be expressed as, $X = A.(G_1, \ldots, G_n)$ where $G = (G_1, \ldots, G_n)$ is a random Gaussian vector and A is an $\frac{\alpha}{2}$-stable random variable independent of the vector G. This type of stable variable plays an important role because they make a link between Gaussian and α-stable variables. Another suitable property is that their characteristic function is given by the following simple formula:

$$\Phi_X(\theta) = \exp\left\{-(\theta^t \Sigma \theta)^{\frac{\alpha}{2}}\right\}, \tag{3.26}$$

  where Σ is the covariance matrix of the Gaussian vector $G = (G_1, \ldots, G_n)$.

Complex random variables and vectors

Complex random variables have applications in various fields, including signal processing. The study of complex variables was favored by the importance of harmonizable processes that are expressed as a Fourier transform of a random measure or a process (3.3).

Definition 3.2.7 A complex random variable $X = X_1 + \imath.X_2$ is α-stable if and only if the couple of real random variables (X_1, X_2) is an α-stable vector in $\mathbb{R}^2$. More generally, a complex random vector $(X_1, \ldots, X_d)$ with $X_j = X_j^1 + \imath X_j^2$ for $j = 1, \ldots, d$, is α-stable if and only if the vector $(X_1^1, X_1^2, \ldots, X_d^1, X_d^2)$ is an α-stable random vector in $\mathbb{R}^{2d}$.

Definition 3.2.8 A complex random variable SαS, $X = X_1 + \imath.X_2$, is said isotropic (rotationally invariant) if and only if for any $\phi \in [0, 2\pi[$ the random variables X and $e^{\imath\phi}.X$ have the same probability distribution.

Isotropic random variables are frequently encountered in wireless communications where directions or more generally phases do not influence the received signal, that is why they are sometimes called rotationally invariant, see for example [82]. These variables have several properties that we summarize in the following proposition:

Proposition 3.2.9 *Let $X = X_1 + \imath.X_2$ be an isotropic complex random variable then:*

- *The real and imaginary parts X_1 and X_2 are dependent real α-stable variables.*
- *The spectral measure Γ of the vector SαS, (X_1, X_2) is given by uniform law on the unit circle.*
- *The real random vector (X_1, X_2) is sub-Gaussian.*
- *The characteristic function of X is given by:*

$$\Phi_X(\theta) = \mathbb{E}\left[\exp\{\imath.\mathcal{R}e(\theta.\overline{X})\}\right] = \exp\{-c_0.\Gamma_{(X_1,X_2)}(S_2), |\theta|^\alpha\} \tag{3.27}$$

  where c_0 is given by, $c_0 = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} |\cos(\phi)|^\alpha d\phi$.

This last property shows that the characteristic function of isotropic α-stable complex variable is parametric and depends only on $|\theta|^\alpha$ and on the parameter $\Gamma_{(X_1,X_2)}(S_2)$.

The proof of these results are detailed in [58].

Dependence and association measure of SαS random variables

For Gaussian random variables, the covariance matrix characterizes the distribution of a Gaussian vector. Its flexible algebraic properties, such as its bilinearity and the fact that it is positive definite, make the study of the covariance function simpler. This important tool which is naturally associated with Gaussian variables is based on second-order moments. According to (3.16), the latter are infinite in case of α-stable variables with $0 < \alpha < 2$. Miller [59] introduced an equivalent concept for α-stable vectors under the name *covariation*, as a new dependency measure to replace the covariance in the case of symmetric α-stable random variables with $1 < \alpha < 2$.

Covariation

In the following we assume that $1 < \alpha < 2$ unless stated otherwise.

Definition 3.2.9 Let $X = (X_1, X_2)$ be a symmetric α-stable real vector with a characterizing spectral measure Γ defined on the unit circle S_2. The covariation of X_1 on X_2 denoted $[X_1, X_2]_\alpha$ is defined by:

$$[X_1, X_2]_\alpha = \int_{S_2} s_1 . s_2^{<\alpha-1>} d\Gamma(s_1, s_2) \tag{3.28}$$

with $s^{<\beta>} = \text{sign}(s).|s|^\beta$ when s is a real number. In the same way, if $X^1 = X_1^1 + \imath X_2^1$ and $X^2 = X_1^2 + \imath X_2^2$ are two complex jointly symmetric α-stable random variables, then covariation of X^1 on X^2 is given by

$$[X^1, X^2]_\alpha = \int_{S_4} (s_1^1 + \imath s_2^1).(s_1^2 + \imath s_2^2)^{<\alpha-1>} d\Gamma_{X_1^1,X_2^1,X_1^2,X_2^2}(s_1^1, s_2^1, s_1^2, s_2^2), \tag{3.29}$$

where $\Gamma_{X_1^1,X_2^1,X_1^2,X_2^2}$ is the unique spectral measure corresponding to the SαS vector $(X_1^1, X_2^1, X_1^2, X_2^2)$. The notation $z^{<\beta>} = |z|^{\beta-1}\overline{z}$ where $\overline{z}$ is the complex conjugate of z.

Unlike the covariance, we see in this definition that, the covariation is not naturally symmetric. This asymmetry comes from the signed power $< \alpha - 1 >$ which appears in (3.28) and (3.29). There is another definition equivalent to Definition 3.2.9 which involves the exponent of the characteristic function given in (3.23). This simple result is easy to handle in practice and is given by the following proposition:

Proposition 3.2.10 *Let (X_1, X_2) jointly SαS random vector, then the covariation of X_1 on X_2 can be expressed as,*

$$[X_1, X_2]_\alpha = \frac{1}{\alpha} . \frac{\partial \sigma^\alpha(\theta_1, \theta_2)}{\partial \theta_1}|_{\theta_1=0,\theta_2=1} \tag{3.30}$$

where $\sigma(\theta_1, \theta_2)$ is the scale parameter of the real SαS random variable, $\theta_1 X_1 + \theta_2 X_2$.

The equivalence between (3.28) and (3.30) is detailed in [76].

As mentioned at the beginning, the covariation was designed to replace the covariance but unfortunately it does not have some of its suitable properties. The following proposition gives the different characteristics and properties of covariation.

Proposition 3.2.11 *The covariation has the following properties:*

1. *Covariation is linear to the left (with respect to its first component) that is to say, for any jointly SαS vector,* (X_1, X_2, Y) *we have,*

$$[X_1 + X_2, Y]_\alpha = [X_1, Y]_\alpha + [X_2, Y]_\alpha.$$

2. *If* X *and* Y *are two real or complex independent SαS random variables, then* $[X, Y]_\alpha = 0$. *The converse is not always true.*
3. *Covariation is additive with respect to its second component in the case of independence that is to say,* $[X, Y_1 + Y_2]_\alpha = [X, Y_1]_\alpha + [X, Y_2]_\alpha$ *if* Y_1 *and* Y_2 *are independent.*
4. *For any real or complex scalars* a *and* b, *we have,*

$$[a.X, b.Y]_\alpha = ab^{<\alpha-1>}[X, Y]_\alpha.$$

The proof of these results is detailed in most works dealing with α-stable variables. For example, in Cambanis [16] we find the proof of these properties in the general case of complex variables.

3.2.3 α-Stable Stochastic Processes

At the beginning research on α-stable stochastic processes and models were developed, first, to find results similar to those established for Gaussian processes and in a second step to explore the novel characteristics. Recently, many remarkable results have been introduced in the literature. For more details, see for example [40, 70, 76]. Let us recall the definition of α-stable stochastic processes and give some of their elementary properties:

Definition 3.2.10 Let T be any ordered set and $0 < \alpha \leq 2$. A real or (complex) stochastic process $(X_t, t \in T)$ is α-stable if for any finite subset $\{t_1, \ldots, t_n\}$ of T, the real or (complex) random vector, $(X_{t_1}, \ldots, X_{t_n})$ is α-stable random vector.

Theorem 3.2.12 *A stochastic process* $(X_t, t \in T)$ *is symmetric* α*-stable if and only if for all linear combinations,*

$$\sum_{k=1}^{d} b_k X_{t_k} \quad \textit{where,} \quad d \geq 1, t_1, \ldots, t_n \in T \quad \text{and} \quad b_1, \ldots, b_n \in \mathbb{R}, \tag{3.31}$$

is also a symmetric α*-stable variable.*

This theorem shows that the set of all finite linear combinations (3.31) is a vector space that will be denoted $l(X)$. In following paragraph, we will show that the covariation generates a norm for which $l(X)$ is a Banach space.

The covariation Norm

It is known that the vector space generated by a subfamily of second-order random variables is a Hilbert space. This Hilbertian structure is not conceivable for α-stable random variables because of their infinite variance. However, using the covariation we can substitute the Hilbert structure by a Banach one. Indeed, if a α-stable process $(X(t), t \in T)$, then the vector space $l(X)$ is composed of α-stable random variables having the same stability index α, see for instance [19]. The following theorem specifies the Banach structure of $l(X) = l(X(t), t \in T)$.

Proposition 3.2.13 *Let* $1 < \alpha < 2$ *then the application,*

$$\|.\|_\alpha : \begin{array}{ccc} l(X) & \longrightarrow & \mathbb{R}^+ \\ Y & \longmapsto & \|Y\|_\alpha \triangleq ([Y,Y]_\alpha)^{\frac{1}{\alpha}}, \end{array} \tag{3.32}$$

defines a norm called covariation norm.

In that case $(l(X), \|.\|_\alpha)$ *is a Banach space and its topology is equivalent to the topology of convergence in probability. Besides, for any fixed* Z *in* $l(X)$*, the application,*

$$\begin{array}{ccc} l(X) & \longrightarrow & \mathbb{C} \\ Z & \longmapsto & [.,Z]_\alpha, \end{array}$$

is a continuous linear form of norm $\|Z\|_\alpha^{\alpha-1}$*. In* $(l(X), \|.\|_\alpha)$ *the covariation function is also continuous with respect to its second component and for any complex SαS vector* (Z_1, Z_2, Z_3) *we have:*

$$|[Z_1,Z_2]_\alpha - [Z_1,Z_3]_\alpha| \leq 2\|Z_1\|_\alpha . \|Z_2 - Z_3\|_\alpha^{\alpha-1}. \tag{3.33}$$

Proof The proof of the first assumption is detailed in [16]. To show inequality (3.33), let us denote $Z_1 = Z_1^1 + \imath . Z_1^2, Z_2 = Z_2^1 + \imath . Z_2^2, Z_3 = Z_3^1 + \imath . Z_3^2$ and $\Gamma = \Gamma_{(Z_1,Z_2,Z_3)}$ the unique measure defined on the unit sphere S_6 of $\mathbb{R}^6$ corresponds to the vector SαS $(Z_1^1, Z_1^2, Z_2^1, Z_2^2, Z_3^1, Z_3^2)$, then by definition of covariation we have:

$$|[X,Y]_\alpha - [X,Z]_\alpha| \leq \int_{S_6} |z_1||z_2^{<\alpha-1>} - z_3^{<\alpha-1>}| d\Gamma(w_1, \ldots, w_6),$$

where $z_i = z_i^1 + \imath z_i^2$. Using inequality $|z_1^{<\alpha-1>} - z_2^{<\alpha-1>}| \leq 4|z_1 - z_2|^{\alpha-1}$ and according to Hölder inequality we have:

$$
\begin{aligned}
|[X,Y]_\alpha - [X,Z]_\alpha| &\leq 4\int_{S_6} |z_1||z_2-z_3|^{<\alpha-1>} d\Gamma(z_1^1, z_1^2, z_3^1, z_3^2) \\
&\leq 4\left[\int_{S_6} |z_1|^\alpha d\Gamma(z_1^1, z_1^2, z_3^1, z_3^2)\right]^{\frac{1}{\alpha}} \cdot \left[\int_{S_6} |z_2-z_3|^\alpha d\Gamma(z_1^1, z_1^2, z_3^1, z_3^2)\right]^{\frac{\alpha-1}{\alpha}} \\
&\leq 4\|X\|_\alpha \cdot \|Y-Z\|_\alpha^{\alpha-1}
\end{aligned}
\tag{3.34}
$$

We give some results concerning moments of isotropic complex SαS random variables as well as their relation with covariation.

Proposition 3.2.14 *Let Y be a symmetric α-stable random variable, then if Y is real the moment of order p, with $1 < p < \alpha$, of X is given by:*

$$
\mathbb{E}|Y|^p = \mathcal{S}_\alpha(p) \cdot \|Y\|_\alpha^p \quad \text{and} \quad \mathcal{S}_\alpha(p) = 2^p \frac{\Gamma(\frac{1+p}{2}) \cdot \Gamma(1-\frac{p}{\alpha})}{\Gamma(1-\frac{p}{2})\Gamma(\frac{1}{2})}, \tag{3.35}
$$

where Γ is the common gamma function defined in (3.10). If Y is isotropic complex then the moment of order p, with $1 < p < \alpha$ is given by

$$
\mathbb{E}|Y|^p = \widetilde{\mathcal{S}}_\alpha(p) \cdot \|Y\|_\alpha^p \quad \text{and} \quad \widetilde{\mathcal{S}}_\alpha(p) = 2^p \frac{\Gamma(\frac{2+p}{2}) \cdot \Gamma(1-\frac{p}{\alpha})}{\Gamma(1-\frac{p}{2})}. \tag{3.36}
$$

The proof of these results is detailed for example in Feller [30] for the real case and Cambanis and Miamee [17] for the isotropic complex case. In Definition 3.2.9, covariation is given in terms of the spectral measure which is not given in an explicit way and is difficult to manipulate in practice. The following proposition gives the link between the covariation and the fractional moments $\mathbb{E}(XY^{<p-1>})$ where $1 < p < \alpha$. The proof of this result is in Cambanis and Miamee [17].

Proposition 3.2.15 *Let X and Y be two jointly SαS real or complex isotropic random variables. Then for every $1 < p < \alpha$, we have:*

$$
\frac{[X,Y]_\alpha}{\|Y\|_\alpha^\alpha} = \frac{\mathbb{E}(X \cdot Y^{<p-1>})}{\mathbb{E}|Y|^p}. \tag{3.37}
$$

3.2.4 Stochastic Integration and Spectral Representation of Stable Processes

The history of stochastic integration started in 1900 with Bachelier [7] where it was first used in financial modeling. Then Einstein [28] used stochastic integrals in the field of physics when he tried to integrate with respect to a Wiener Process. It was only after the pioneering work of Kolmogorov [44, 45] on probability theory that K. Itô [39], said to be the father of the stochastic integration, developed the foundation of this concept as it is known today. The stochastic integral with respect

to an independent increment α-stable process has been studied by several authors in the 1970s. We cite among others, Schilder [77], Hosoya [38], Hardin [36], Miller [59], Cambanis [16, 18, 19], etc. In Smorodnitsky and Taqqu [76], we find a detailed overview on building as well as the properties of stochastic integration with respect to an indepentally scattered α-stable random measure.

Lebesgue-Stieltjes stochastic integral

In this section we address Lebesgue-Stieltjes integration with respect to an independent increments SαS process. We follow the works detailed in Cambanis [16]. Let us consider a symmetric α-stable process $\xi = (\xi_t, t \in \mathbb{R})$ with $1 < \alpha < 2$ and let us denote $\Delta(\xi)$ the Banach space generated by the increments of ξ that is to say the completion, with respect to the covariation norm $\|.\|_\alpha$, of all finite linear combinations of the increments of ξ. According to Cambanis [16], for the Lebesgue-Stieltjes integral to exist, the process ξ must verify the conditions:

- The process ξ is right continuous with respect the covariation norm.
- For any fixed linear combination ζ of increments of ξ, the application, $v : t \longmapsto [\xi_t, \zeta]_\alpha$ must have bounded variation.

In this case the Lebesgue-Stieltjes stochastic integral is constructed in the following way:

For a step function defined on a bounded interval $[a, b[$ that one can write as, $f = \sum_{k=1}^{n} f_k \mathbb{1}_{]t_{k-1}, t_k]}$ with $a = t_0 < t_1 < \cdots < t_n = b$, Cambanis [16] defined the stochastic integral of the function f by:

$$\int f d\xi = \sum_{k=1}^{n} f_k (\xi_{t_k} - \xi_{t_{k-1}}). \tag{3.38}$$

Since the function v is of bounded variation and right continuous, we can define a norm by $\|f\|_{1,\alpha} \triangleq \left(\int f(t) dv(t) \right)^{\frac{1}{\alpha}}$ with $v(t) = [\xi_t, \int f d\xi]_\alpha$. A simple calculation shows that

$$\begin{aligned} \|f\|_{1,\alpha}^{\alpha} &= \int f(t) dv(t) = \sum_{k=1}^{n} f_k . (v(t_k) - v(t_{k-1})) \\ &= \sum_{k=1}^{n} f_k . \left[\xi_{t_k} - \xi_{t_{k-1}}, \int f d\xi \right]_\alpha \\ &= \left[\sum_{k=1}^{n} f_k (\xi_{t_k} - \xi_{t_{k-1}}), \int f d\xi \right]_\alpha \\ &= \left[\int f d\xi, \int f d\xi \right]_\alpha = \left\| \int f d\xi \right\|_\alpha^\alpha \end{aligned} \tag{3.39}$$

First, using the fact that the convergence with respect to the covariation norm is equivalent to convergence in probability, see Cambanis, for example [16] or Samorodnitsky and Taqqu [76], we deduce that stochastic integrals $\int f_n d\xi$ converge in probability if and only if it converges in the norm $\|.\|_\alpha$. Therefore, according to (3.39), if and only if the function sequence f_n is a Cauchy sequence with respect to $\|.\|_{1,\alpha}$. As in [16], we denote by $\Lambda_\alpha(\Delta\xi)$ the completion with respect to the norm $\|.\|_{1,\alpha}$ of step functions having compact support on $\mathbb{R}$. For any function f of $\Lambda_\alpha(\Delta\xi)$ there exists a sequence f_n of step functions with compact support such that, $\|f_n - f\|_{1,\alpha} \to 0$. The Lebesgue-Stieltjes stochastic integral of f is therefore the limit in probability of $\int f_n d\xi$, given in (3.38). We then get an isomorphism between $\Lambda_\alpha(\Delta\xi)$ and $l(\Delta\xi)$.

Case where ξ have independent increments

We present some results about stochastic integration with respect to a process with independent increments, for more details see [16]. The interest of this case lies in the fact that these processes are right continuous with respect to the topology of convergence in probability. On the other hand, for any finite reals $t_0 < t_1 < \cdots < t_n$, using the ξ independence of increments we can decompose as follows:

$$\|\xi_{t_n} - \xi_{t_0}\|_\alpha^\alpha = \sum_{k=1}^{n} \|\xi_{t_k} - \xi_{t_{k-1}}\|_\alpha^\alpha. \tag{3.40}$$

This additivity property, in the case of independence, allows to construct a Lebesgue-Stieltjes positive measure μ defined by:

$$\mu(]s,t]) = \|\xi_t - \xi_s\|_\alpha^\alpha \qquad \text{for all,} \quad s < t. \tag{3.41}$$

Consider two step functions f and g with compact support that we write as, $f = \sum_{k=1}^{n} f_k \mathbb{1}_{]t_{k-1},t_k]}$ and $g = \sum_{k=1}^{n} g_k \mathbb{1}_{]t_{k-1},t_k]}$, and we denote by $X_k = \xi_{t_k} - \xi_{t_{k-1}}$ the increments of ξ. Using the definition of the integral given in (3.38), we have:

$$\left[\int f d\xi, \int g d\xi\right]_\alpha = \left[\sum_{k=1}^{n} f_k X_k, \sum_{k=1}^{n} g_k X_k\right]_\alpha. \tag{3.42}$$

As the process ξ has independent increments, one can see easily that random variables $X_0, X_1, \ldots, X_n$ are independents. Thanks to the additivity of the covariation Proposition 3.2.11, we have:

$$\begin{aligned}\left[\int f d\xi, \int g d\xi\right]_\alpha &= \left[\sum_{k=1}^{n} f_k X_k, \sum_{k=1}^{n} g_k X_k\right]_\alpha \\ &= \sum_{k=1}^{n}\sum_{k'=1}^{n} f_k . g_{k'}^{<\alpha-1>} [X_k, X_{k'}]_\alpha\end{aligned}$$

$$= \sum_{k=1}^{n} f_k . g_k^{<\alpha-1>} [X_k, X_k]_\alpha, \quad \text{for} \quad k \neq k', \quad [X_k, X_{k'}]_\alpha = 0.$$

$$= \sum_{k=1}^{n} f_k . g_k^{<\alpha-1>} \| \xi_{t_k} - \xi_{t_{k-1}} \|_\alpha^\alpha$$

$$= \sum_{k=1}^{n} f_k . g_k^{<\alpha-1>} \mu((t_{k-1}, t_k])$$

$$= \int f . g^{<\alpha-1>} d\mu. \tag{3.43}$$

This result can be generalized to all functions of $\Lambda_\alpha(\Delta\xi)$. The Lebesgue-Stieltjes positive measure μ is called the spectral measure (sometimes control measure) of the process ξ. Cambanis [16] showed, in the following proposition, that any Lebesgue-Stieltjes measure is a spectral measure of some symmetric α-stable process having independent increments.

Proposition 3.2.16 *For any finite measure defined on $\mathbb{R}$, there exist a real or complex stochastic process ξ with independent increments for which, $\Lambda_\alpha(\Delta\xi) = L^\alpha(\mu)$ and satisfying relation (3.43).*

3.2.5 Harmonisable Processes and Spectral Density in Masry-Cambanis's Sense

The main motivation of Masry and Cambanis's work [58] was to investigate SαS harmonisable stochastic processes $X_t = \int e^{it\lambda} d\xi(\lambda)$ where ξ is a symmetric α-stable process with isotropic independent increments. In that case, relation (3.43) is rewritten as:

$$[X_s, X_t]_\alpha = \int_{-\infty}^{\infty} e^{i(s-t)\lambda} d\mu(\lambda). \tag{3.44}$$

Representation (3.44) is similar to the spectral representation of the covariance function for harmonisable second-order processes (3.2). By analogy with the power spectral density, Masry and Cambanis called μ defined in (3.41) the spectral measure of the process X. When this measure accepts a radon derivative with respect to the Lebesgue measure, this density is called the spectral density of the process X.

Although (3.44) does not have the power and energy signification as the spectral density of second-order processes, Masry and Cambanis pointed out that it can play a similar role in several practical situations. Statistical analysis and estimations of this spectral density were given in [58] for continuous time processes. An estimate of the spectral density from discrete observations was studied in [73, 74].

3.2.6 Estimation of the Spectral Density Using the Characteristic Function

We present in this section the estimation procedure using works of [73] or generally [26] that are based on Masry and Cambanis's work [58] in the case of continuous time harmonisable SαS processes. We will recall some notation used in these papers. Let us first recall the definition of Jackson polynomials:

$$\mathcal{J}^{(N)}(l) = \frac{1}{q_{k,n}} \left(\frac{\sin(\frac{nl}{2})}{\sin(\frac{l}{2})} \right)^{2k} \quad \text{where} \quad q_{k,n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{\sin(\frac{nl}{2})}{\sin(\frac{l}{2})} \right)^{2k} dl,$$

with N a fixed real number such that, $N = 2k(n-1)+1$, where $n \in \mathbb{N}$ and $k \in \mathbb{N} \cup \left\{\frac{1}{2}\right\}$; and if $k = \frac{1}{2}$ then N is an odd integer.
It is known that their exists a function $\mathbf{j}_k$ satisfying:

$$\mathcal{J}^{(N)}(\lambda) = \sum_{m'=-k(n-1)}^{k(n-1)} \mathbf{j}_k(m'/n) \cos(\lambda m').$$

It is shown in [72] that:

$$\mathbf{j}_k(u) = \begin{cases} \dfrac{Q_N^{(k)}(nu + k(n-1))}{Q_N^{(k)}(k(n-1))} & \text{, if } nu \in \mathbb{Z}, \\ 0 & \text{, if } nu \notin \mathbb{Z}, \end{cases}$$

where $Q_N^{(k)}(\tau)$ is the number of integer combinations $(t_1, t_2, \ldots, t_{2k})$ that satisfy: $\forall j \in \{1, 2, \ldots, 2k\}$, $0 \le t_j \le n-1$ and $t_1 + t_2 + \cdots + t_{2k} = \tau$.

From this Jackson polynomial function, we then define the following kernel which is used in the construction of the periodogram,

$$|\mathcal{J}_N(\lambda)|^{\alpha} = \left| A_N \mathcal{J}^{(N)}(\lambda) \right|^{\alpha} \quad \text{where} \quad A_N = \left(\int_{-\pi}^{\pi} \left| \mathcal{J}^{(N)}(\lambda) \right|^{\alpha} d\lambda \right)^{-\frac{1}{\alpha}}.$$

Let $\hat{I}_N$ be the periodogram defined on $]-\Omega, \Omega[$ as follows:

$$\hat{I}_N(\lambda) = C_{p,\alpha} |I_N(\lambda)|^p, \qquad 0 < p < \frac{\alpha}{2}$$

where,

$$I_N(\lambda) = [\tau]^{\frac{1}{\alpha}} A_N \mathcal{R}e \left[\sum_{n'=-k(n-1)}^{n'=k(n-1)} h_k\left(\frac{n'}{n}\right) \exp\left\{-i(n'\tau\lambda)\right\} X\left(n'\tau + k(n-1)\tau\right) \right].$$

The normalization constant $C_{p,\alpha}$ is given by $C_{p,\alpha} = \dfrac{D_p}{F_{p,\alpha}[C_\alpha]^{p/\alpha}}$, with:

$$D_p = \int_{-\infty}^{\infty} \frac{1-\cos(u)}{|u|^{1+p}} du, \tag{3.45}$$

$$F_{p,\alpha} = \int_{-\infty}^{\infty} \frac{1-e^{-|u|^a}}{|u|^{1+p}} du, \tag{3.46}$$

$$C_\alpha = \frac{1}{\pi\alpha} \int_0^{\pi} |\cos(u)|^\alpha \, du. \tag{3.47}$$

In order to obtain a consistent estimate of $[\phi(\lambda)]^{\frac{p}{\alpha}}$, the periodogram is smoothed via a spectral window W_N defined by: $W_N(u) = M_N W(M_N u)$, where M_N satisfies $M_N \to +\infty$ and $\frac{M_N}{N} \to 0$ as $N \to +\infty$, W is a nonnegative, even, continuous function vanishing for $|\lambda| > 1$ such that $\int_{-1}^{1} W(u)du = 1$. The bandwidth of the spectral window is then proportional to $1/M_N$. We consider then the smoothed periodogram f_N defined by:

$$f_N(\lambda) = \int_{\mathbb{R}} W_N(\lambda - u))\hat{I}_N(u)du, \qquad -\Omega < \lambda < \Omega.$$

It is shown in [72] or generally in [26], that $f_N(\lambda)$ is an asymptotically unbiased consistent estimator of $[\phi(\lambda)]^{\frac{p}{\alpha}}$ for $-\Omega < \lambda < \Omega$.

We deduce from this that,

$$\hat{\phi}(\lambda) = (f_N(\lambda))^{\alpha/p}.$$

It is commonly known that the spectral density of a second-order processes gives the energy distribution according to frequencies. In the case of α-stable processes, where this energy is theoretically infinite, the spectral density loses the notion of the power distribution. However, it can still be seen as a fractional relative fractional power distribution.

3.3 Application of Stable Processes to Communication Channel Modeling

Radio transmission is permanently looking for new solutions to increase the data rates. One solution is to find free frequency bands but this is difficult because it is already highly occupied. However, some possibilities exist and one of them is an available spectrum around 60 GHz where a very wide band (several GHz) is available. Developing systems in this band necessitates to characterize and model the radio channel. Many references can be found on channel characterization (see

for example [23, 33, 78, 85]) and models have been developed. Some are based on deterministic modeling and ray tracing [52, 61, 78] with satisfying results or geometric approaches. We are interested in statistical models. Early works, in other frequency bands, rely on the framework proposed by Bello [10]. It relies mainly on two assumptions: the channel is wide sense stationary and has uncorrelated scatterers. This means stationary in the time domain and in the frequency domain.

The channel can then be represented by a linear filter characterized by its impulse response and four random variables:

$$h(t) = \sum_{k=1}^{N} a_k \delta\left(t - \tau_k\right) e^{J\theta_k}, \tag{3.48}$$

where a_k, τ_k, θ_k, and N are, respectively, the amplitude, the delay, the phase of path k and the paths' number. Phases are usually considered as uniformly distributed over $[0, 2\pi]$ because the path length is much larger than the wavelength [27].

As long as the bandwidth is not too big, a regularly spaced taps delay line can be used instead of (3.48), meaning $\tau_k = kT_e$ where T_e is the sampling time. In [35], models for a 6 GHz large channel centered at 5 GHz are compared and regularly spaced taps and arbitrary delays give both good results. However, the regularly spaced taps are problematic for ultra wide bandwidth (meaning a very small T_e). The main reason is the sparsity of the impulse response [60]: paths are not present in each time interval and the distributions of the tap coefficients cannot be based on complex Gaussian random variables. As a consequence, we are interested in models with arbitrary path times of arrival.

The distribution of this variable is a difficult problem [6, 22, 78, 79, 83]. The most widely used solution was proposed by Saleh and Valuenzela [75]. It relies on a doubly stochastic Poisson approach : paths arrive in clusters; the arrival times of clusters and the arrival time of paths in clusters are exponential random variables. It has been used for the 60 GHz indoor channel [32, 48, 63] and adopted by the standardization groups IEEE 802.15.3a for ultra wide band channels, 802.15.4a for sensor networks and IEEE 802.15.3c for the 60 GHz band [86]. However, it remains a complicated model, difficult to handle analytically and necessitating many parameters (delays, amplitudes).

Classical assumptions

To have a statistical representation of the channel, based on second-order statistics, classical works [3, 21, 33, 48, 78], assume the channel is Wide Sense Stationary and with Uncorrelated Scatterers. It appears however that those assumptions are not valid:

- In the context of millimeter waves, one single office room will show various behaviors and characteristics. The local areas where the WSS assumption holds are reduced to a few cm^2 due to the small wavelength.

- Due to the increase in the bandwidth, more paths can be resolved. Consequently, correlation between multipath can appear and the uncorrelated scatterers property does not hold.

One solution is to increase the number of parameters in the model to account for the correlation and the high number of detected paths. Then, the parameters themselves should vary or even be considered as random variables.

New approach

We propose a model that is not based on second-order statistics. Let us consider the channel transfer function $H(\omega)$ which is the Fourier transform of the impulse response $h(\tau)$:

$$H(\omega) = \int e^{J\omega\tau} h(\tau) d\tau. \tag{3.49}$$

We model the transfer function as a stochastic process represented by the stochastic integral:

$$H(\omega, .) = \int_{\mathbb{R}} e^{J\omega\tau} d\xi(\tau). \tag{3.50}$$

The term $d\xi(\tau) = h(\tau, .)d\tau$ is the Lebesgue-Stieltjes random measure constructed from the stochastic process $h(\tau, .)$.

Here, (3.50) shows that the probabilistic properties of $d\xi(\tau)$ are linked to those of $H(\omega, .)$ by an inverse Fourier transform. This means that the knowledge of the random measure $d\xi$ characterizes the transfer function and, consequently, the radio channel. Figure 3.1 represents realizations of the random variables $H(\omega, .)$ (details on the measurement setup can be found in [33]) and the test of infinite variance [62, p. 62–63] on these realizations that confirm the impulsive nature of the channel.

Consequently, we assume that $d\xi$ is an α-stable random measure with independent increments. This assumption [16] allows to characterize $H(\omega, .)$ by a unique measure μ defined on $\mathbb{R}$ as described in Sect. 3.2.3. Then, in contrast with the IEEE 802.15 approach, we are brought back to find a deterministic function in the place of random parameters.

Estimation of the Spectral Density

We first estimate the stability index α based on real and imaginary parts of measured $H(\omega, .)$ for different ω and average the obtained estimates. We then estimate the density $\hat{\phi}$ of the measure μ; one transfer function allows to calculate the periodogram and we average the periodograms obtained from the different observations.

We summarize in the Algorithm 1 the estimation procedure, following the framework described in Sect. 3.2.6, where all detailed expressions can be found.

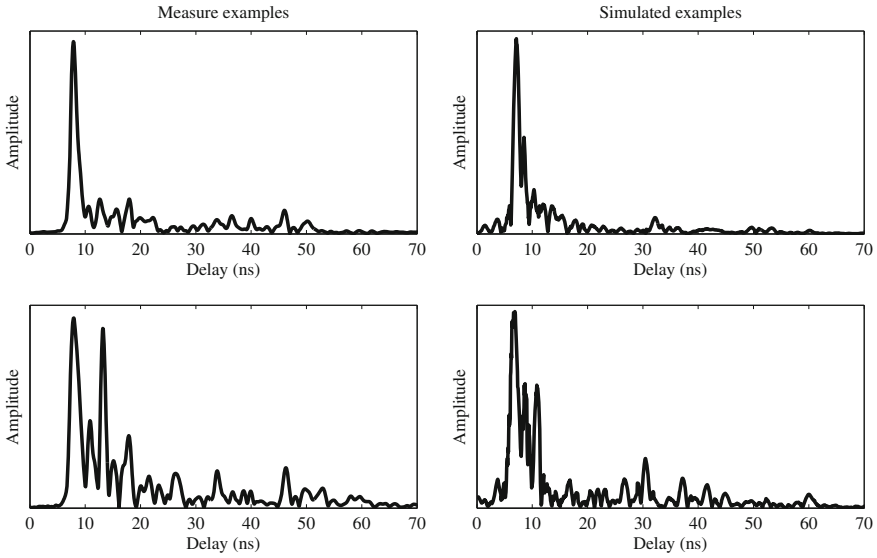


Fig. 3.1 Test of infinite variance for $|H(\omega)|$

Algorithm 1: Nonparametric estimation of the spectral density $\hat{\phi}(\lambda)$ from observed transfer functions.

Input: Measured transfer functions $H_l(\omega_i)_{i=1\ldots N}$ at positions $\ell = 1, \ldots, L$.
Pretreatment: *;*
Estimate index exponent α *from* $\{H_l(\omega)\}_{l=1,\ldots,L}$ *using [46];*
Take $p = \frac{\alpha}{2.5}$ *and calculate,* c_α, D_p *and* $F_{p,\alpha}$ *given in (3.45), (3.46) and (3.47) ;*
Calculate, $C_{p,\alpha}$ *;*
Calculate $\mathbf{j}_k(\frac{n'}{n})$ *;*
foreach $\ell = 1 \ldots K$ **do**
  Calculate $I_{N,\ell}(\lambda)$;
  Calculate $\hat{I}_{N,\ell}(\lambda)$;
Calculate $\bar{\hat{I}}_N$, the mean of $\hat{I}_{N,\ell}(\lambda)$ over $\ell = 1 \ldots L$;
Smooth $\bar{\hat{I}}_N(\lambda)$ to obtain $f_N(\lambda)$;
Output: The spectral density is $\hat{\phi}(\lambda) = (f_N(\lambda))^{\frac{\alpha}{p}}$.

Generation of the impulse responses

The transfer function $H(\omega, .)$ may be decomposed in Lepage type series (see Proposition 3.2.5) as:

$$H(\omega, .) \stackrel{d}{=} (\mu(\mathbb{R})C_\alpha)^{\frac{1}{\alpha}} \sum_{i=1}^{\infty} \gamma_i \Gamma_i^{-\frac{1}{\alpha}} e^{J\omega\vartheta_i}(S_{i,1} + JS_{i,2}). \tag{3.51}$$

- Equality "$\stackrel{d}{=}$" means equality in distribution, and C_α is a constant depending only on α:

$$C_\alpha = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\frac{\pi\alpha}{2})} \quad \text{for} \quad \alpha \neq 1 \quad \text{and} \quad C_1 = \frac{2}{\pi} \tag{3.52}$$

- (γ_i) are independent copies of Rademacher random variable γ, this means that, $P(\gamma = -1) = P(\gamma = 1) = 1/2$,
- (Γ_i) are the arrival times of a Poisson process, they are gamma distributed of rate i,
- $(\vartheta_i)_i$ are independent copies of the random variable ϑ,
- $(S_{i,1}, S_{i,2})$ are independent copies of a continuous random variable uniformly distributed on the unit circle. For convenience, we represent it in complex notation: $S_{i,1} + \jmath S_{i,2} = e^{\jmath\theta_i}$.

From the estimated density $\hat{\phi}$, we generate independent copies of the random variable ϑ. We use the Ahrens algorithm described in [37, Algorithm 5.1, p. 115].

The infinite representation has to be truncated to a positive integer N chosen large enough to have a practically reasonable approximation. We propose a measure of the approximation accuracy by calculating the risk probability to generate a transfer function $H_N(\omega)$ not close to the true trajectory $H(\omega)$. In order to have accuracy η with risk ε, we take N such that

$$R_N(\alpha) \leq \frac{\eta^2\varepsilon}{K}. \tag{3.53}$$

In a last step we adjust our model to the limited time resolution of the experimental setup. We observe the impulse response at discrete times with a step Δt. Let $I\,\Delta t$ be the total duration of the impulse response. We finally generate:

$$h(t) = \sum_{i=0}^{I-1} a_i \delta(i\,\Delta t), \tag{3.54}$$

where

$$a_i = \sum_{j=1}^{N} \gamma_j \Gamma_j^{-\frac{1}{\alpha}} e^{\jmath\theta_j} \mathbb{1}_{\vartheta_j \in [i\,\Delta t, (i+1)\Delta t[}. \tag{3.55}$$

We detail the impulse response generation technique in Algorithm 2.

3.3.1 Illustration

In order to show that our model is appropriate in the case of the 60 GHz channel, we have confronted it to the observed impulse responses.

Algorithm 2: The channel impulse response generation.

Input: Estimated spectral density $\hat{\phi}(\lambda)$ and α from Algorithm 1.
Pretreatment: ;
Calculate C_α from (3.52).;
Approximate the total mass $\mu(\mathbb{R}) = \int \phi(\lambda) d\lambda$.;
Find the threshold N as in (3.53).;
do: ;
Generate $\Gamma_i = e_1 + \cdots + e_i$ where (e_i) are i.i.d. random variables exponential of mean 1.;
Generate γ_i i.i.d. copies of γ.;
Generate ϑ_i from $\hat{\mu}$, for instance with Ahrens algorithm (see [37, Chap. 5]).;
Calculate a_i from (3.55) .;
end do: ;
Output: An impulse response $h(t)$ from (3.54).

In Fig. 3.2, we present simulated impulse responses. Thirty thousand samples were generated per impulse response for a risk $\eta = 0.05$ and an accuracy $\varepsilon = 0.01$.

Figure 3.2 shows the ability of our model to represent very different situations. In contrast, classical approaches based on power delay profile cannot adapt to such a variability. One single measurement is indeed sufficient to represent the channel behavior in the whole room. This results from the fact that the wide sense stationary property is not required for our model. Another important fact is that the model is able to represent the path time of arrival with a great accuracy. This is an important feature when ultra wide band channels are considered.

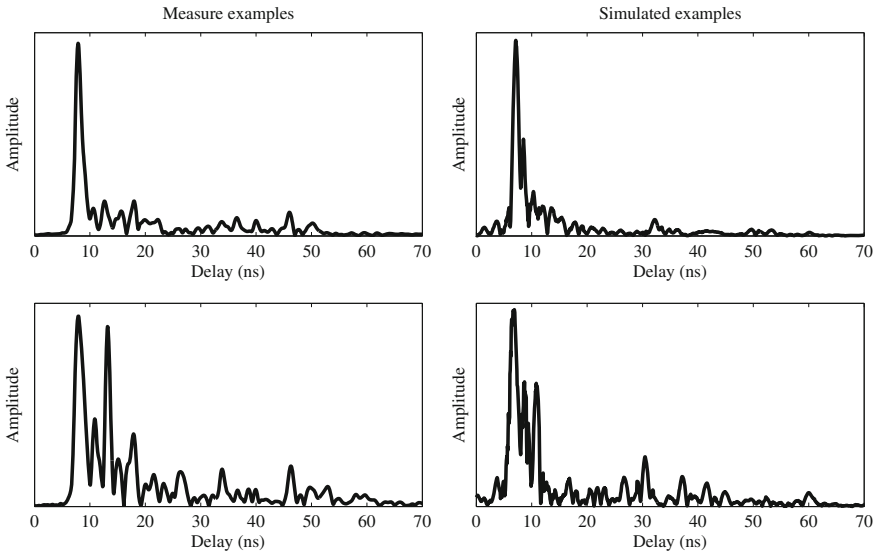


Fig. 3.2 Example of simulated impulse responses

References

1. Adler, R.J., Cambanis, S., Samorodnitsky, G.: On stable markov processes. Stochastic Processes and their Applications **34**, 1–17 (1990)
2. Akgirav, V., Booth, G.: The stable-law model of stock returns. J. Econ. Stat. **6**, 51–57 (1988)
3. Al-Nuaimi, M., Siamarou, A.: Coherence bandwidth and K-factor measurements for indoor wireless radio channels at 62.4 GHz. In: 11th International Conference on Antennas and Propagation, ICAP01, pp. 275–278 (2001)
4. Anderson, T.W.: The Statistical Analyis of Time Series. Wiley, New York (1971)
5. Astrauskas, A., Levy, J., Taqqu, M.S.: The asymptotic dependence structure of the linear fractional levy motion. Lietuvos Matematikos Rinkinys (Lith. Math. J.) **31**(1), 1–28 (1991)
6. Azzaoui, N., Clavier, L., Sabre, R.: Path delay model based on α-stable distribution for the 60 GHz indoor channel, pp. 1638–1643 (2003)
7. Bachelier, L.: Théorie de la spéculation. Ph.D. thesis. Gauthier-Willars, à la Faculté des Sciences de Paris (1900)
8. Banerjee, A., Burlina, P., Chellappa, R.: Adaptive target detection in foliage-penetrating sar images using alpha-stable models. IEEE Trans. Image Process. **8**(12), 1823–1831 (1999)
9. Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S.I.: Lévy Processes : Theory and Applications. Birkhäuser, Boston (2001)
10. Bello, P.: Characterization of randomly time-variant linear channels. IEEE Trans. Commun. **11**(4), 360–393 (1963)
11. Bergström, H.: On distribution functions with a limiting stable distribution function. Arkiv. Mat **2**, 463–474 (1953)
12. Blattberg, R., Gonedes, N.: A comparison of the stable and student distributions as statistical models for stock prices. J. Bus. **47**, 244–280 (1974)
13. Bochner, S.: Vorlesungen über Fouriersche Integrale. Leipzig, Akad. Verlag (1932)
14. Bretagnolle, J., Dacunha-Castelle, D., Krivine, J.: Lois stables et espaces l^p. Annals de l' Institut Henri Poincaré **2**, section B, 231–259 (1966)
15. Brockwell, P., Davis, R.: Time Series: Theory and Methods. Springer, New York (1987)
16. Cambanis, S.: Complex symetric stable variables and processes. Contribution to Statistics: P.K sen editions, Contribution to statistics (Essays in honor of Norman L. Johnson), pp. 63–79 (1983)
17. Cambanis, S., Miamee, A.G.: On prediction of harmonizable stable processes. Sankhya: Indian J. Stat. V **51**, 269–294 (1989)
18. Cambanis, S., Miller, G.: Some path properties of pth order and symmetric stable processes. Ann. Prob. **8**(6), 1148–1156 (1980)
19. Cambanis, S., Miller, G.: Linear problems in pth order and symmetric stable processes. SIAM J. Appl. Math. **41**, 43–69 (1981)
20. Chambers, J., Mallows, C., Stuck, B.: A method for simulating stable random variables. J. Am. Stat. Assoc. **71**(354), 340–344 (1976)
21. Clavier, L., Fryziel, M., Garnier, C., Delignon, Y., Boulinguez, D.: Performance of DS-CDMA on the 60 GHz channel, pp. 2332–2336 (2002)
22. Clavier, L., Rachdi, M., Fryziel, M., Delignon, Y., Thuc, V.L., Garnier, C., Rolland, P.: Wide band 60 GHz indoor channel: characterization and statistical modelling, pp. 2098–2102 (2001)
23. Collonge, S., Zaharia, G., Zein, G.E.: Wideband and dynamic characterization of the 60 GHz indoor radio propagation—future home WLAN architectures. Annales des Télécommunications **58**(3-4) (2003)
24. Cramér, H., Leadbetter, M.: Stationary and Related Stochastic Processes. Wiley, New York (1967)
25. Dahlhaus, R.: Fitting time series models to nonstationary processes. Ann. Stat. **25**(1), 1–37 (1997)
26. Demesh, N.N., Chekhmenok, S.L.: Estimation of the spectral density of a homogeneous random stable discrete time field. SORT (Stat. Oper. Res. Trans.) **29**, 101–118 (2005)

27. Durgin, G.: Space-Time Wireless Channels. Prentice Hall PTR (2002)
28. Einstein, A.: On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. Ann. Phys. **17**(4), 891–921 (1905)
29. Feldheim, E.: Étude de la stabilité des lois de probabilité. Ph.D. thesis. Imprimerie et librairie de la ville (1937)
30. Feller, W.: An Introduction to Probability Theory and its Applications, 2nd edn. Wiley, New York (1971)
31. Ferguson, T., Klass, M.: A representation theorem of independent increment processes without Gaussian component. Ann. Math. Stat. **43**, 1634–1643 (1972)
32. Flament, M., Svensson, A.: Virtual cellular networks for 60 GHz wireless infrastructure. In: IEEE International Conference on Communications, ICC 2003, vol. 2, pp. 1223–1227 (2003)
33. Fryziel, M., Loyez, C., Clavier, L., Rolland, N.: Path loss model of the 60 GHz radio channel. Microw. Opt. Technol. Lett. **34**(3), 158–162 (2002)
34. Gnedenko, V., Kolmogorov, A.N.: Limit distributions for sums of independent random variables. Adisson-Wesley Publ, Cambridge (1954)
35. Greenstein, L.J., Ghassemzadeh, S.S., Hong, S.C., Tarokh, V.: Comparison study of UWB indoor channel models. IEEE Trans. Wirel. Commun. **6**(1), 128–135 (2007)
36. Hardin, C.D.: Spectral representation of symmetric stable processes. J. Multivar. Anal. **12**(3), 385–401 (1982)
37. Hörman, W., Leydold, J., Derflinger, G.: Automatic Nonuniform Random Variate Generation. Springer, Berlin (2005)
38. Hosoya, Y.: Harmonizable stable processes. Prob. Theor. Relat. Fields **60**(4), 517–533 (1982)
39. Ito, K.: Stochastic integral. Proc. Imp. Acad. Tokyo **20**(8), 519–524 (1944)
40. Janiki, A., Weron, A.: Simulation and Chaotic Behavior of Stable Processes. Marcel Dekker, New York (1993)
41. Kanter, M.: Linear sample spaces and stable processes. J. Funct. Anal. **9**(4), 441–459 (1972)
42. Khinchine, A.Y.: Korrelationstheorie der stationären stochastischen prozesse. Mathematische Annalen **109**, 604 (1934)
43. Khinchine, A.Y., Lévy, P.: Sur les lois stables. Comptes rendus hebdomadaires des seances de l'Academie des sciences, Academie des science (Paris) **202**, (Serie A), 374–376 (1936)
44. Kolmogorov, A.N.: über die analytischen methoden in der wahrscheinlichkeitsrechnung. Mathematische Annalen **104**, 415–458 (1931)
45. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin (1933)
46. Koutrouvelis, I.A.: Regression-type estimation of the parameters of stable laws. J. Am. Stat. Assoc. **75**(372), 918–928 (1980)
47. Kuelbs, J.: A representation theorem for symmetric stable processes and stable measures on H. Prob. Theor. Relat. Fields **26**(4), 259–271 (1973)
48. Kunisch, J., Zollinger, E., Pamp, J., Winkelmann, A.: MEDIAN 60 GHz wideband indoor radio channel measurements and model. In: 50th IEEE Vehicular Technology Conference. VTC Fall, vol. 4, pp. 2393–2397 (1999)
49. LePage, R.: Multidimensional Infinitely Divisible Variables and Processes. Part I: Stable Case. Lecture Notes in Mathematics, vol. 1391, pp. 153–163. Springer, Berlin (1990)
50. Lévy, P.: Théorie des erreurs. la loi de gauß et les lois exceptionnelles. Bulletin de la Société Mathématique de France **52**, 49–85 (1924)
51. Lévy, P.: Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris (1954)
52. Lostalen, Y., Corre, Y., Louet, Y., Helloco, Y.L., Collonge, S., Zein, G.E.: Comparison of measurements and simulations in indoor environments for wireless local area networks. In: 55th IEEE Vehicular Technology Conference. VTC Spring, vol. 1, pp. 389–393 (2002)
53. Makagon, A., Mandrekar, V.: The spectral representation of stable processes: harmonizability and regularity. Prob. Theor. Relat. Fields **85**(1), 1–11 (1990)
54. Mandelbrot, B.B., Ness, J.W.V.: Fractional Brownian motions, fractional noises and applications. SIAM Rev. **10**(4), 422–437 (1968)
55. Masry, E.: Alias-free sampling: an alternative conceptualization and its applications. IEEE Trans. Inf. Theor. **24**(3), 317–324 (1978)

56. Masry, E.: The wavelet transform of stochastic processes with stationaryincrements and its application to fractional Brownian motion. IEEE Trans. Inf. Theor. **39**(1), 260–264 (1993)
57. Masry, E., Cambanis, S.: On the reconstruction of the covariance of stationary Gaussian processes observed through zero-memory nonlinearities–Part II (corresp.). IEEE Trans. Inf. Theor. **26**(4), 503–507 (1980)
58. Masry, E., Cambanis, S.: Spectral density estimation for stationary stable processes. Stochastic Process. Appl. **18**(1), 1–31 (1984)
59. Miller, G.W.: Some results on symmetric stable distributions and processes. Ph.D. thesis, University of North Carolina at Chapel Hill (1977)
60. Molisch, A., Tufvesson, F., Karedal, J., Mecklenbräuker, C.: A survey on vehicle-to-vehicle propagation channels. IEEE Wirel. Commun. **16**(6), 12–22 (2009)
61. Moraitis, N., Constantinou, P.: Indoor channel modeling at 60 GHz for wireless LAN applications. In: 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, vol. 3, pp. 1203–1207 (2002)
62. Nikias, C.L., Shao, M.: Signal Processing With Alpha-stable Distributions and Applications. Wiley, New York (1996)
63. Park, J., Kim, Y., Hur, Y., Kim, K., Kim, K.: Analysis of 60 GHz band indoor wireless channels with channel configurations. In: 9th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, vol. 2, pp. 617–620 (1998)
64. Parzen, E.: On consistent estimates of the spectrum of a stationary time series. Ann. Math. Stat. **28**(2), 329–348 (1957)
65. Parzen, E.: On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. J. R. Stat. Soc. Series B (Methodological) **20**(2), 303–322 (1958)
66. Priestley, M.B.: Spectral Analysis and Time Series. Academic Press, New York (1981)
67. Priestley, M.B.: Non-linear and Non-stationary Time Series Analysis. Academic Press, London (1988)
68. Rao, M.M.: Harmonizable processes : structure theory. L'enseignement Mathématiques (Essays in Honor of Prof. S. Bochner) **28**, 295–351 (1982)
69. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. Math. Stat. **27**(3), 832–837 (1956)
70. Rosinski, J.: Minimal integral representations of stable processes. To appear in Probability and Mathematical Statistics **26**(1)
71. Rosinski, J.: On series representations of infinitely divisible random vectors. Ann. Prob. **18**(1), 405–430 (1990)
72. Sabre, R.: Spectral density estimation for statonary stable random fields. Applicationes Mathematicae Journal **2**(32), 107–133 (1995)
73. Sabre, R.: Discrete estimation of spectral density for symmetric stable process. Statistica **60**(3), 497–520 (2000)
74. Sabre, R.: Aliasing free for stable random fields. Egypt Stat. J. **46**(1), 53–75 (2002)
75. Saleh, A.A.M., Valenzuela, R.A.: A statistical model for indoor multipath propagation. IEEE J. Select. Areas Commun. **SAC–5**(2), 128–137 (1987)
76. Samorodnitsky, G., Taqqu, M.S.: Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall, New York (1994)
77. Schilder, M.: Some structure theorems for the symmetric stable laws. Ann. Math. Stat. **41**(2), 412–421 (1970)
78. Smulders, P., Correia, L.: Characterisation of propagation in 60 GHz radio channels. Electron. Commun. Eng. J. **9**(2), 73–80 (1997)
79. Smulders, P., Fernandes, J.: Wide-band simulations and measurements of MM-wave indoor radio channels. In: 5th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, vol. 2, pp. 501–504 (1994)
80. Tukey, J.W.: Non-parametric estimation II. Statistically equivalent blocks and tolerance regions—the continuous case. Ann. Math. Stat. **18**, 529–539 (1947)
81. Tukey, J.W.: The future of data analysis. Ann. Math. Stat. **33**, 1–67 (1962)

82. Uchaikin, V.V., Zolotarev, V.M.: Chance and Stability: Stable Distributions and their Applications. VSP, Utrecht (1999)
83. Wales, S., Rickard, D.: Wideband propagation measurements of short range millimetric radio channels. Race 1043 Electron. Commun. Eng. J. **5**(4), 249–254 (1993)
84. Wiener, N.: Generalized harmonic analysis. Acta Mathematica **55**, 178–258 (1930)
85. Xu, H., Hukshya, V., Rappaport, T.: Spatial and temporal characteristics of 60 GHz indoor channel. IEEE J. Select. Areas Commun. **20**(3), 620–630 (2002)
86. Yong, S.: TG3c channel modeling sub-committee final report. IEEE 802. 15-07-0584-00-003c (2007)

Chapter 4
Networks, Random Graphs and Percolation

Philippe Deprez and Mario V. Wüthrich

Abstract The theory of random graphs goes back to the late 1950s when Paul Erdős and Alfréd Rényi introduced the Erdős-Rényi random graph. Since then many models have been developed, and the study of random graph models has become popular for real-life network modelling such as social networks and financial networks. The aim of this overview is to review relevant random graph models for real-life network modelling. Therefore, we analyse their properties in terms of stylised facts of real-life networks.

4.1 Stylised Facts of Real-Life Networks

A network is a set of particles that may be linked to each other. The particles represent individual network participants and the links illustrate how they interact among each other, for an example see Fig. 4.1. Such networks appear in many real-life situations, for instance, there are virtual social networks with different users that communicate with (are linked to) each other, see Newman et al. [35], or there are financial networks such as the banking system where banks exchange lines of credits with each other, see Amini et al. [3] and Cont et al. [15]. These two examples represent rather recently established real-life networks that originate from new technologies and industries but, of course, the study of network models is much older motivated by studies in sociology or questions about interacting particle systems in physics.

Such real-life networks, in particular social networks, have been studied on many different empirical data sets. These studies have raised several stylised facts about

P. Deprez
Department of Mathematics, ETH Zurich, RiskLab, Zurich, Switzerland
e-mail: philippe.deprez@math.ethz.ch

M.V. Wüthrich (✉)
Department of Mathematics, Swiss Finance Institute SFI, ETH Zurich, RiskLab, Zurich, Switzerland
e-mail: mario.wuethrich@math.ethz.ch

G.W. Peters and T. Matsui (eds.), *Theoretical Aspects of Spatial-Temporal Modeling*,
JSS Research Series in Statistics, DOI 10.1007/978-4-431-55336-6_4

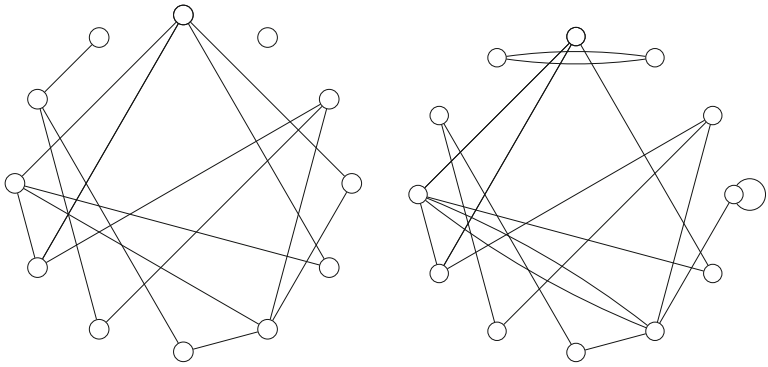


Fig. 4.1 *lhs* ER random graph; *rhs* NSW random graph

large real-life networks that we would briefly like to enumerate, for more details see Newman et al. [35] and Sect. 1.3 in Durrett [19] and the references therein.

1. Many pairs of distant particles are connected by a very short chain of links. This is sometimes called the "small-world" effect. Another interpretation of the small-world effect is the observation that the typical distance of any two particles in real-life networks is at most six links, see Watts [40] and Sect. 1.3 in Durrett [19]. The work of Watts [40] was inspired by the statement of his father saying that "he is only six handshakes away from the president of the United States". For other interpretations of the small-world effect we refer to Newman et al. [35].
2. The clustering property of real-life networks is often observed which means that linked particles tend to have common friends.
3. The distribution of the number of links of a single particle is heavy-tailed, i.e. its survival probability has a power law decay. In many real-life networks the power law constant (tail parameter) τ is estimated between 1 and 2 (finite mean and infinite variance, see also (4.3) below). Section 1.4 in Durrett [19] presents the following examples:

   - number of oriented links on web pages: $\tau \approx 1.5$,
   - routers for e-mails and files: $\tau \approx 1.2$,
   - movie actor network: $\tau \approx 1.3$,
   - citation network Physical Review D: $\tau \approx 1.9$.

   Typical real-life networks are heavy-tailed in particular if maintaining links is free of costs.

Since real-life networks are too complex to be modelled particle by particle and link by link, researchers have developed many models in random graph theory that help to understand the geometry of such real-life networks. The aim of this overview paper is to review relevant models in random graph theory, in particular, we would like to analyse whether these models fulfil the stylised facts mentioned above. The models we consider in Sects. 4.2 and 4.3 are the Erdős-Rényi random graph and the Newman-Strogatz-Watts random graph, respectively. We then look at random graph

models where particles are embedded in Euclidean space. This leads to nearest-neighbour bond percolation treated in Sect. 4.4. In the subsequent two sections, we consider extensions of the nearest-neighbour bond percolation model, the homogeneous and the heterogeneous long-range percolation model. In Sect. 4.7 we consider heterogeneous long-range percolation in continuum space. In the last section, we explain renormalisation techniques which are crucial tools used in various proofs of statements about the clustering property and the small-world effect.

Standard literature on random graph and percolation theory is Bollobás [10], Durrett [19], Franceschetti-Meester [21], Grimmett [24, 25], van der Hofstad [26] and Meester-Roy [31].

4.2 Erdős-Rényi Random Graph

We choose a set of particles $V_n = \{1, \dots, n\}$ for fixed $n \in \mathbb{N}$. Thus, V_n contains n particles. The Erdős-Rényi (ER) random graph introduced in the late 1950s, see [20], attaches to every pair of particles $x, y \in V_n, x \neq y$, independently an edge with fixed probability $p \in (0, 1)$, i.e.

$$\eta_{x,y} = \eta_{y,x} = \begin{cases} 1 \text{ with probability } p, \\ 0 \text{ with probability } 1 - p, \end{cases} \tag{4.1}$$

where $\eta_{x,y} = 1$ means that there is an edge between x and y, and $\eta_{x,y} = 0$ means that there is *no* edge between x and y. Identity $\eta_{x,y} = \eta_{y,x}$ illustrates that we have an undirected random graph. We denote this random graph model by $\mathrm{ER}(n, p)$. In Fig. 4.1(lhs), we provide an example for $n = 12$, observe that this realisation of the ER random graph has one isolated particle and the remaining ones lie in the same connected component.

We say that x and y are *adjacent* if $\eta_{x,y} = 1$. We say that x and y are *connected* if there exists a path of adjacent particles from x to y. We define the *degree* $\mathscr{D}(x)$ of particle x to be the number of adjacent particles of x in V_n. Among others, general random graph theory is concerned with the limiting behaviour of the ER random graph $\mathrm{ER}(n, p_n)$ for $p_n = \vartheta/n$, $\vartheta > 0$, as $n \to \infty$. Observe that for $k \in \{0, \dots, n-1\}$ we have, see for instance Lemma 2.9 in [41],

$$g_k = g_k^{(n)} = \mathbb{P}[\mathscr{D}(x) = k] = \binom{n-1}{k} p_n^k (1 - p_n)^{n-1-k} \quad \to \quad e^{-\vartheta} \frac{\vartheta^k}{k!}, \tag{4.2}$$

as $n \to \infty$. We see that the degree distribution of a fixed particle $x \in V_n$ with edge probability p_n converges for $n \to \infty$ to a Poisson distribution with parameter $\vartheta > 0$. In particular, this limiting distribution is light-tailed and, therefore, the ER graph does not fulfil the stylised fact of having a power law decay of the degree distribution.

The ER random graph has a phase transition at $\vartheta = 1$, reflecting different regimes for the size of the largest connected component in the ER random graph. For $\vartheta < 1$, all connected components are small, the largest being of order $\mathscr{O}(\log n)$, as $n \to \infty$.

For $\vartheta > 1$, there is a constant $\chi(\vartheta) > 0$ and the largest connected component of the ER random graph is of order $\chi(\vartheta)n$, as $n \to \infty$, and all other connected components are small, see Bollobás [10] and Chap. 2 in Durrett [19]. At criticality ($\vartheta = 1$) the largest connected component is of order $n^{2/3}$, however, this analysis is rather sophisticated, see Sect. 2.7 in Durrett [19].

Moreover, the ER random graph has only very few complex connected components such as cycles (see Sect. 2.6 in Durrett [19]): for $\vartheta \neq 1$ most connected components are trees, only a few connected components have triangles and cycles, and only the largest connected component (for $\vartheta > 1$) is more complicated. At criticality the situation is more complex; a few large connected components emerge and finally merge to the largest connected component as $n \to \infty$.

4.3 Newman-Strogatz-Watts Random Graph

The approach of Newman-Strogatz-Watts (NSW) [34, 35] aims at directly describing the degree distribution $(g_k)_{k\geq 0}$ of $\mathscr{D}(x)$ for a given particle $x \in V_n$ ($n \in \mathbb{N}$ being large). The aim is to modify the degree distribution in (4.2) so that we obtain a power law distribution. Assume that any particle $x \in V_n$ has a degree distribution of the form $g_0 = 0$ and

$$g_k = \mathbb{P}[\mathscr{D}(x) = k] \sim ck^{-(\tau+1)}, \quad \text{as } k \to \infty, \tag{4.3}$$

for given tail parameter $\tau > 0$ and $c > 0$. Note that $\sum_{k\geq 1} k^{-(\tau+1)} < 1 + 1/\tau$ which implies that $c > 0$ is admissible. By definition the survival probability of this degree distribution has a power law with tail parameter $\tau > 0$. However, this choice (4.3) does not explain how one obtains an explicit graph from the degrees $\mathscr{D}(x)$, $x \in V_n$. The graph construction is done by the Molloy-Reed [32] algorithm: attach to each particle $x \in V_n$ exactly $\mathscr{D}(x)$ ends of edges and then choose these ends randomly in pairs (with a small modification if the total number of ends is odd). This will provide a random graph with the desired degree distribution. In Fig. 4.1(rhs) we provide an example for $n = 12$, observe that this realisation of the NSW random graph has two connected components. The Molloy-Reed construction may provide multiple edges and self-loops, but if $\mathscr{D}(x)$ has finite second moment ($\tau > 2$) then, there are only a few multiple edges and self-loops, as $n \to \infty$, see Theorem 3.1.2 in Durrett [19]. However, in view of real-life networks we are rather interested into tail parameters $\tau \in (1, 2)$, for which we so far have no control on multiple edges and self-loops.

Newman et al. [34, 35] have analysed this random graph by basically considering cluster growth in a two-step branching process. Define the probability generating function of the first generation by

$$G_0(z) = \mathbb{E}\left[z^{\mathscr{D}(x)}\right] = \sum_{k\geq 1} g_k z^k, \quad \text{for } z \in \mathbb{R}.$$

Note that we have $G_0(1) = 1$ and $\mu = \mathbb{E}[\mathscr{D}(x)] = G_0'(1)$ (supposed that the latter exists). The second generation has then probability generating function given by

$$G_1(z) = \sum_{k\geq 0} \frac{(k+1)g_{k+1}}{\mu} z^k = \sum_{k\geq 1} \frac{kg_k}{\mu} z^{k-1}, \quad \text{for } z \in \mathbb{R},$$

where the probability weights are specified by $\widetilde{g}_k = (k+1)g_{k+1}/\mu$ for $k \geq 0$. For $\tau > 2$ the second generation has finite mean given by

$$\vartheta = \sum_{k\geq 0} k\widetilde{g}_k = \sum_{k\geq 0} k \frac{(k+1)g_{k+1}}{\mu} = \frac{1}{\mu} \sum_{k\geq 1} (k-1)kg_k.$$

Note that the probability generating functions are related to each other by $G_0'(z) = \mu G_1(z) = G_0'(1)G_1(z)$. Similar to the ER random graph there is a phase transition in this model. It is determined by the mean ϑ of the second generation, see (5)–(6) in Newman et al. [35] and Theorems 3.1.3 and 3.2.2 in Durrett [19]: for $\vartheta > 1$ the largest connected component has size of order $\chi(\vartheta)n$, as $n \to \infty$. The fraction $\chi(\vartheta) = 1 - G_0(z_0)$ is found by choosing z_0 to be the smallest fixed point of G_1 in $[0, 1]$. Moreover, no other connected component has size of order larger than $\mathscr{O}(\log n)$. Note that we require finite variance $\tau > 2$ for ϑ to exist.

If $\vartheta < 1$ the distribution of the size of the connected component of a fixed particle converges in distribution to a limit with mean $1 + \mu/(1-\vartheta)$, as $n \to \infty$, see Theorem 3.2.1 in Durrett [19]. The size of the largest connected component in this case ($\tau > 2$ and $\vartheta < 1$) is conjectured to be of order $n^{1/\tau}$: the survival probability of the degree distribution has asymptotic behaviour of order $k^{-\tau}$, and therefore, the largest degree of n independent degrees has size of order $n^{1/\tau}$, which leads to the same conjecture for the largest connected component, see also Conjecture 3.3.1 in Durrett [19].

From a practical point of view the interesting regime is $1 < \tau < 2$ because many real-life networks have such a tail behaviour, see Sect. 1.4 in Durrett [19]. In this case, we have $\vartheta = \infty$ and an easy consequence is that the largest connected component grows proportionally to n (because this model dominates a model with finite second moment and mean of the second generation being bigger than 1). In this regime $1 < \tau < 2$ we can study the graph distance of two randomly chosen particles (counting the number of edges connecting them) in the largest connected component, see Sect. 4.5 in Durrett [19]. In the Chung-Lu model [13, 14], which uses a variant to the Molley-Reed [32] algorithm, it is proved that this graph distance behaves as $\mathscr{O}(\log\log n)$, see Theorem 4.5.2 in Durrett [19]. Van der Hofstadt et al. [27] obtain the same asymptotic behaviour $\mathscr{O}(\log\log n)$ for the NSW random graph in the case $1 < \tau < 2$. Moreover, in their Theorem 1.2 [27] they also state that this graph distance behaves as $\mathscr{O}(\log n)$ for $\tau > 2$. These results on the graph distances can be interpreted as the small-world effect because two randomly chosen particles in V_n are connected by very few edges.

We conclude that NSW random graphs have heavy tails for the degree distribution choices according to (4.3). Moreover, the graph distances have a behaviour that can be interpreted as small-world effect.

Less desirable features of NSW random graphs are that they may have self-loops and multiple edges. Moreover, the NSW random graph is expected to be locally rather sparse leading to locally tree-like structures, see also Hurd-Gleeson [28]. That is, we do not expect to get reasonable local graph geometry and the required clustering property. Variations considered allowing for statistical interpretations in terms of likelihoods include the works of Chung-Lu [13, 14] and Olhede-Wolfe [36].

4.4 Nearest-Neighbour Bond Percolation

In a next step, we would like to embed the previously introduced random graphs and the corresponding particles into Euclidean space. This will have the advantage of obtaining a natural distance function between particles, and it will allow to compare Euclidean distance to graph distance between particles (counting the number of edges connecting two distinct particles). Before giving the general random graph model we restrict ourselves to the nearest-neighbour bond percolation model on the lattice $\mathbb{Z}^d$ because this model is the basis for many derivations. More general and flexible random graph models are provided in the subsequent sections.

Percolation theory was first presented by Broadbent-Hammersley [11]. It was mainly motivated by questions from physics, but these days percolation models are recognised to be very useful in several fields. Key monographs on nearest-neighbour bond percolation theory are Kesten [30] and Grimmett [24, 25].

Choose a fixed dimension $d \in \mathbb{N}$ and consider the square lattice $\mathbb{Z}^d$. The vertices of this square lattice are the particles and we say that two particles $x, y \in \mathbb{Z}^d$ are nearest-neighbour particles if $\|x - y\| = 1$ (where $\|\cdot\|$ denotes the Euclidean norm). We attach at random edges to nearest-neighbour particles $x, y \in \mathbb{Z}^d$, independently of all other edges, with a fixed edge probability $p \in [0, 1]$, that is,

$$\eta_{x,y} = \eta_{y,x} = \begin{cases} 1_{\{\|x-y\|=1\}} & \text{with probability } p, \\ 0 & \text{with probability } 1-p, \end{cases} \tag{4.4}$$

where $\eta_{x,y} = 1$ means that there is an edge between x and y, and $\eta_{x,y} = 0$ means that there is *no* edge between x and y. The resulting graph is called nearest-neighbour (bond) random graph in $\mathbb{Z}^d$, see Fig. 4.2 (lhs) for an illustration. Two particles $x, y \in \mathbb{Z}^d$ are connected if there exists a path of nearest-neighbour edges connecting x and y. It is immediately clear that this random graph does not fulfil the small-world effect because one needs at least $\|x - y\|$ edges to connect x and y, i.e. the number of edges grows at least linearly in the Euclidean distance between particles $x, y \in \mathbb{Z}^d$. The degree distribution is finite because there are at most 2^d nearest-neighbour edges, more precisely, the degree has a binomial distribution with parameters 2^d and p. We present this square lattice model because it is an interesting basis for the development

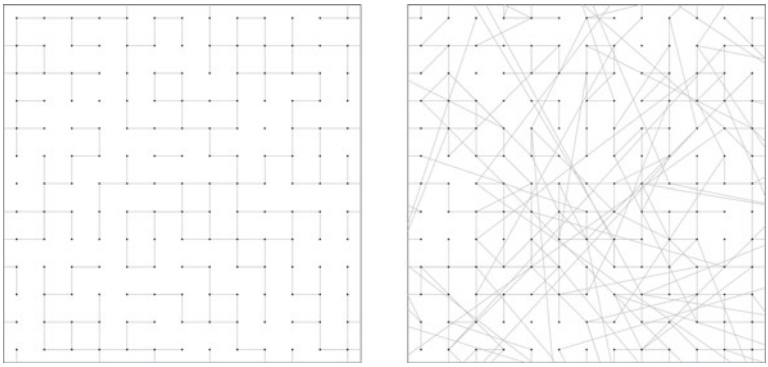


Fig. 4.2 *lhs* nearest-neighbour percolation; *rhs* homogeneous long-range percolation

of more complex models. Moreover, this model is at the heart of many proofs in percolation problems which are based on so-called renormalisation techniques, see Sect. 4.8 below for a concrete example.

In percolation theory, the object of main interest is the connected component of a given particle $x \in \mathbb{Z}^d$ which we denote by

$$\mathscr{C}(x) = \left\{ y \in \mathbb{Z}^d : \ x \text{ and } y \text{ are connected by a path of nearest-neighbour edges} \right\}.$$

By translation invariance it suffices to define the percolation probability at the origin

$$\theta(p) = \mathbb{P}_p \left[|\mathscr{C}(0)| = \infty \right],$$

where $|\mathscr{C}(0)|$ denotes the size of the connected component of the origin and $\mathbb{P}_p$ is the product measure on the possible nearest-neighbour edges with edge probability $p \in [0, 1]$, see Grimmett [24], Sect. 2.2. The *critical probability* $p_c = p_c(\mathbb{Z}^d)$ is then defined by

$$p_c = \inf \{ p \in (0, 1] : \ \theta(p) > 0 \}.$$

Since the percolation probability $\theta(p)$ is non-decreasing, the critical probability is well defined. We have the following result, see Theorem 3.2 in Grimmett [24].

Theorem 1 *For nearest-neighbour bond percolation in* $\mathbb{Z}^d$ *we have*

(a) for $d = 1$*:* $p_c(\mathbb{Z}) = 1$*; and*
(b) for $d \geq 2$*:* $p_c(\mathbb{Z}^d) \in (0, 1)$.

This theorem says that there is a non-trivial phase transition in $\mathbb{Z}^d$, $d \geq 2$. This needs to be considered together with the following result which goes back to Aizenman et al. [1], Gandolfi et al. [22] and Burton-Keane [12]. Denote by $\mathscr{I}$ the number of infinite connected components. Then, we have the following statement, see Theorem 7.1 in Grimmett [24].

Theorem 2 *For any $p \in (0,1)$ either $\mathbb{P}_p[\mathscr{I} = 0] = 1$ or $\mathbb{P}_p[\mathscr{I} = 1] = 1$.*

Theorems 1 and 2 imply that there is a *unique* infinite connected component for $p > p_c(\mathbb{Z}^d)$, a.s. This motivates the notation $\mathscr{C}_\infty$ for the unique infinite connected component for the given edge configuration $(\eta_{x,y})_{x,y}$ in the case $p > p_c(\mathbb{Z}^d)$. $\mathscr{C}_\infty$ may be considered as an infinite (nearest-neighbour) network on the particle system $\mathbb{Z}^d$ and we can study its geometrical and topological properties. Using a duality argument, Kesten [29] proved that $p_c(\mathbb{Z}^2) = 1/2$ and monotonicity then provides $p_c(\mathbb{Z}^{d+1}) \le p_c(\mathbb{Z}^d) \le p_c(\mathbb{Z}^2) = 1/2$ for $d \ge 2$.

One object of interest is the so-called *graph distance* (chemical distance) between $x, y \in \mathbb{Z}^d$, which is for a given edge configuration defined by

$$d(x,y) = \text{minimal length of path connecting } x \text{ and } y \text{ by nearest-neighbour edges } \eta_{z_1,z_2} = 1,$$

where this is defined to be infinite if there is no nearest-neighbour path connecting x and y for the given edge configuration. We have already mentioned that $d(x,y) \ge \|x - y\|$ because this is the minimal number of nearest-neighbour edges we need to cross from x to y. Antal-Pisztora [4] have proved the following upper bound.

Theorem 3 *Choose $p > p_c(\mathbb{Z}^d)$. There exists a positive constant $c = c(p,d)$ such that, a.s.,*

$$\limsup_{\|x\| \to \infty} \frac{1}{\|x\|} d(0,x) \mathbb{1}_{\{0 \text{ and } x \text{ are connected}\}} \le c.$$

4.5 Homogeneous Long-Range Percolation

Long-range percolation is the first extension of nearest-neighbour bond percolation. It allows for edges between any pair of particles $x, y \in \mathbb{Z}^d$. Long-range percolation was originally introduced by Schulman [37] in one dimension. Existence and uniqueness of the infinite connected component in long-range percolation was proved by Schulman [37] and Newman-Schulman [33] for $d = 1$ and by Gandolfi et al. [23] for $d \ge 2$.

Consider again the percolation model on the lattice $\mathbb{Z}^d$, but we now choose the edges differently. Choose $p \in [0,1]$, $\lambda > 0$ and $\alpha > 0$ fixed and define the edge probabilities for $x, y \in \mathbb{Z}^d$ by

$$p_{x,y} = \begin{cases} p & \text{if } \|x - y\| = 1, \\ 1 - \exp(-\lambda \|x - y\|^{-\alpha}) & \text{if } \|x - y\| > 1. \end{cases} \tag{4.5}$$

Between any pair $x, y \in \mathbb{Z}^d$ we attach an edge, independently of all other edges, as follows

$$\eta_{x,y} = \eta_{y,x} = \begin{cases} 1 & \text{with probability } p_{x,y}, \\ 0 & \text{with probability } 1 - p_{x,y}. \end{cases}$$

We denote the resulting product measure on the edge configurations by $\mathbb{P}_{p,\lambda,\alpha}$. Figure 4.2 (rhs) shows part of a realised configuration. We say that the particles x and y are *adjacent* if there is an edge $\eta_{x,y} = 1$ between x and y. We say that x and y are *connected* if there exists a path of adjacent particles in $\mathbb{Z}^d$ that connects x and y. The connected component of x is given by

$$\mathscr{C}(x) = \left\{ y \in \mathbb{Z}^d : \ x \text{ and } y \text{ are connected} \right\}.$$

We remark that the edge probabilities $p_{x,y}$ used in the literature have a more general form. Since for many results only the asymptotic behaviour of $p_{x,y}$ as $\|x - y\| \to \infty$ is relevant, we have decided to choose the explicit (simpler) form (4.5) because this also fits to our next models. Asymptotically, we have the following power law

$$p_{x,y} \ \sim \ \lambda \|x - y\|^{-\alpha}, \quad \text{as } \|x - y\| \to \infty.$$

Theorem 1 (b) immediately implies that we have percolation in $\mathbb{Z}^d$, $d \geq 2$, for p sufficiently close to 1. We have the following theorem, see Theorem 1.2 in Berger [6].

Theorem 4 *For long-range percolation in $\mathbb{Z}^d$ we have, in an a.s. sense,*

(a) for $\alpha \leq d$: there is an infinite connected component;
(b) for $d \geq 2$ and $\alpha > d$: for p sufficiently close to 1 there is an infinite connected component;
(c) for $d = 1$:

- *(1) $\alpha > 2$: there is no infinite connected component;*
- *(2) $1 < \alpha < 2$: for p sufficiently close to 1 there is an infinite connected component;*
- *(3) $\alpha = 2$ and $\lambda > 1$: for p sufficiently close to 1 there is an infinite connected component;*
- *(4) $\alpha = 2$ and $\lambda \leq 1$: there is no infinite connected component.*

The case $\alpha \leq d$ follows from an infinite degree distribution for a given particle, i.e. for $\alpha \leq d$ we have, a.s.,

$$\mathscr{D}(0) = \left| \left\{ x \in \mathbb{Z}^d : \ 0 \text{ and } x \text{ are adjacent} \right\} \right| = \infty, \tag{4.6}$$

and for $\alpha > d$ the degree distribution is light-tailed (we give a proof in the continuum space model in Sect. 4.7, because the proof turns out to be straightforward in continuum space). Interestingly, we now also obtain a non-trivial phase transition in the one dimensional case $d = 1$ once long-range edges are sufficiently likely, i.e. α is sufficiently small. At criticality $\alpha = 2$ also the decay scaling constant $\lambda > 0$ matters. The case $d \geq 2$ is less interesting because it is in line with nearest-neighbour bond percolation. The main interest of adding long-range edges is the study of the resulting geometric properties of connected components $\mathscr{C}(x)$. We will state below that there are three different regimes:

- $\alpha \leq d$ results in an infinite degree distribution, a.s., see (4.6);
- $d < \alpha < 2d$ has finite degrees but is still in the regime of small-world behaviour;
- $\alpha > 2d$ behaves as nearest-neighbour bond percolation.

The critical case $\alpha = 2d$ is not well understood at present. We again focus on the graph distance

$$d(x, y) = \text{minimal number of edges that connect } x \text{ and } y, \tag{4.7}$$

where this is defined to be infinite if x and y do not belong to the same connected component, i.e. $y \notin \mathscr{C}(x)$. For $\alpha < d$ we have infinite degrees and the infinite connected component $\mathscr{C}_\infty$ contains all particles of $\mathbb{Z}^d$, a.s. Moreover, Benjamini et al. [5] prove in Example 6.1 that the graph distance is bounded, a.s., by

$$\left\lceil \frac{d}{d-\alpha} \right\rceil.$$

The case $\alpha \in (d, 2d)$ is considered in Biskup [9], Theorem 1.1, and in Trapman [39]. They have proved the following result:

Theorem 5 *Choose $\alpha \in (d, 2d)$ and assume, a.s., that there exists a unique infinite connected component $\mathscr{C}_\infty$. Then for all $\varepsilon > 0$ we have*

$$\lim_{\|x\|\to\infty} \mathbb{P}_{p,\lambda,\alpha}\left[\Delta - \varepsilon \leq \frac{\log d(0,x)}{\log\log \|x\|} \leq \Delta + \varepsilon \,\middle|\, 0, x \in \mathscr{C}_\infty \right] = 1,$$

where $\Delta^{-1} = \log_2(2d/\alpha)$.

This result says that the graph distance $d(0, x)$ is roughly of order $(\log \|x\|)^\Delta$ with $\Delta = \Delta(\alpha, d) > 1$. Unfortunately, the known bounds are not sufficiently sharp to give more precise asymptotic statements. Theorem 5 can be interpreted as small-world effect since it tells us that long Euclidean distances can be crossed by a few edges. For instance, $d = 2$ and $\alpha = 2.5$ provide $\Delta = 1.47$ and we get $(\log \|x\|)^\Delta = 26.43$ for $\|x\| = 10{,}000$, i.e. a Euclidean distance of 10,000 is crossed in roughly 26 edges.

The case $\alpha > 2d$ is considered in Berger [7].

Theorem 6 *If $\alpha > 2d$ we have, a.s.,*

$$\liminf_{\|x\|\to\infty} \frac{d(0,x)}{\|x\|} > 0.$$

This result proves that for $\alpha > 2d$ the graph distance behaves as in nearest-neighbour bond percolation, because it grows linearly in $\|x\|$. The proof of an upper bound is still open, but we expect a result similar to Theorem 3 in nearest-neighbour bond percolation, see Conjecture 1 of Berger [7].

We conclude that this model has a small-world effect for $\alpha < 2d$. It also has some kind of clustering property because particles that are close share an edge more

commonly, which gives a structure that is locally more dense, see Corollary 3.4 in Biskup [9]. But the degree distribution is light-tailed which motivates to extend the model by an additional ingredient. This is done in the next section.

4.6 Heterogeneous Long-Range Percolation

Heterogeneous long-range percolation extends the previously introduced long-range percolation models on the lattice $\mathbb{Z}^d$. Deijfen et al. [16] have introduced this model under the name of scale-free percolation. The idea is to place additional weights W_x to the particles $x \in \mathbb{Z}^d$ which determine how likely a particle may play the role of a hub in the resulting network.

Consider again the percolation model on the lattice $\mathbb{Z}^d$. Assume that $(W_x)_{x\in\mathbb{Z}^d}$ are i.i.d. Pareto distributed with threshold parameter 1 and tail parameter $\beta > 0$, i.e. for $w \geq 1$

$$\mathbb{P}[W_x \leq w] = 1 - w^{-\beta}. \tag{4.8}$$

Choose $\alpha > 0$ and $\lambda > 0$ fixed. Conditionally given $(W_x)_{x\in\mathbb{Z}^d}$, we consider the edge probabilities for $x, y \in \mathbb{Z}^d$ given by

$$p_{x,y} = 1 - \exp(-\lambda W_x W_y \|x - y\|^{-\alpha}). \tag{4.9}$$

Between any pair $x, y \in \mathbb{Z}^d$ we attach an edge, independently of all other edges, as follows

$$\eta_{x,y} = \eta_{y,x} = \begin{cases} 1 & \text{with probability } p_{x,y}, \\ 0 & \text{with probability } 1 - p_{x,y}. \end{cases}$$

We denote the resulting probability measure on the edge configurations by $\mathbb{P}_{\lambda,\alpha,\beta}$. In contrast to (4.5) we have additional weights W_x and W_y in (4.9). The bigger these weights the more likely is an edge between x and y. Thus, particles $x \in \mathbb{Z}^d$ with a big weight W_x will have many adjacent particles y (i.e. particles $y \in \mathbb{Z}^d$ with $\eta_{x,y} = 1$). Such particles x will play the role of hubs in the network system. Figure 4.3 (lhs) shows part of a realised edge configuration.

The first interesting result is that this model provides a heavy-tailed degree distribution, see Theorems 2.1 and 2.2 in Deijfen et al. [16]. Denote again by $\mathscr{D}(0)$ the number of particles of $\mathbb{Z}^d$ that are adjacent to 0, then we have the following result.

Theorem 7 *Fix $d \geq 1$. We have the following two cases for the degree distribution:*

- *for $\min\{\alpha, \beta\alpha\} \leq d$, a.s., $\mathscr{D}(0) = \infty$;*
- *for $\min\{\alpha, \beta\alpha\} > d$ set $\tau = \beta\alpha/d$, then*

$$\mathbb{P}_{\lambda,\alpha,\beta}[\mathscr{D}(0) > k] = k^{-\tau}\ell(k),$$

for some function $\ell(\cdot)$ that is slowly varying at infinity.

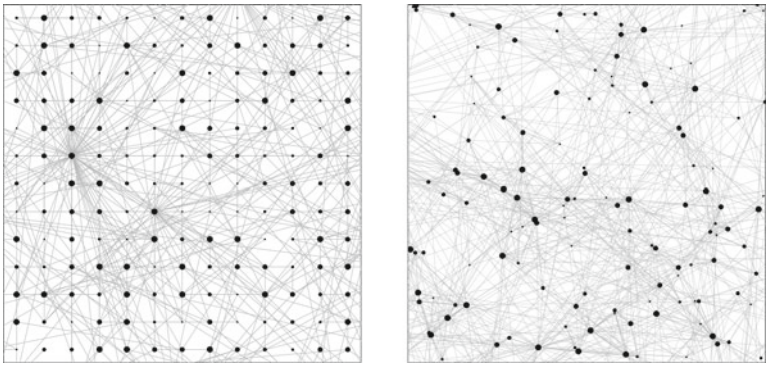


Fig. 4.3 *lhs* heterogeneous long-range percolation; *rhs* continuum space long-range percolation; the size of the particles illustrates the different weights $W_x \geq 1$

We observe that the heavy-tailedness of the weights W_x induces heavy-tailedness in the degree distribution which is similar to choice (4.3) in the NSW random graph model of Sect. 4.3. For $\alpha > d$ there are three different regimes: (i) $\beta\alpha \leq d$ implies infinite degree, a.s.; (ii) for $d < \beta\alpha < 2d$ the degree distribution has finite mean but infinite variance because $1 < \tau < 2$; (iii) for $\beta\alpha > 2d$ the degree distribution has finite variance because $\tau > 2$. We will see that the distinction of the latter two cases has also implications on the behaviour of the percolation properties and the graph distances similar to the considerations in NSW random graphs. Note that from a practical point of view the interesting regime is (ii).

We again consider the connected component of a given particle $x \in \mathbb{Z}^d$ denoted by $\mathscr{C}(x)$ and we define the percolation probability (for given α and β)

$$\theta(\lambda) = \mathbb{P}_{\lambda,\alpha,\beta}\left[|\mathscr{C}(0)| = \infty\right].$$

The *critical percolation value* λ_c is then defined by

$$\lambda_c = \inf\left\{\lambda > 0 : \ \theta(\lambda) > 0\right\}.$$

We have the following result, see Theorem 3.1 in Deijfen et al. [16].

Theorem 8 *Fix $d \geq 1$. Assume* $\min\{\alpha, \beta\alpha\} > d$.

(a) If $d \geq 2$, then $\lambda_c < \infty$.
(b) If $d = 1$ and $\alpha \in (1, 2]$, then $\lambda_c < \infty$.
(c) If $d = 1$ and $\min\{\alpha, \beta\alpha\} > 2$, *then $\lambda_c = \infty$.*

This result is in line with Theorem 4. Since $W_x \geq 1$, a.s., an edge configuration from edge probabilities $p_{x,y}$ defined in (4.9) stochastically dominates an edge configuration with edge probabilities $1 - \exp(-\lambda\|x - y\|^{-\alpha})$. The latter is similar to the homogeneous long-range percolation model on $\mathbb{Z}^d$ and the results of the above theorem directly follow from Theorem 4. For part (c) of the theorem we also refer to

Theorem 3.1 of Deijfen et al. [16]. The next theorem follows from Theorems 4.2 and 4.4 of Deijfen et al. [16].

Theorem 9 *Fix* $d \geq 1$. *Assume* $\min\{\alpha, \beta\alpha\} > d$.

(a) If $\beta\alpha < 2d$, *then* $\lambda_c = 0$.
(b) If $\beta\alpha > 2d$, *then* $\lambda_c > 0$.

Theorems 8 and 9 give the phase transition pictures for $d \geq 1$, see Fig. 4.4 for an illustration. They differ for $d = 1$ and $d \geq 2$ in that the former has a region where $\lambda_c = \infty$ and the latter does not, see also the distinction in Theorem 4. The case $\beta\alpha = 2d$ is not yet well understood. The most interesting case from a practical point of view is the infinite variance case, $1 < \tau < 2$ and $d < \beta\alpha < 2d$, respectively, which provides percolation for any $\lambda > 0$. It follows from Gandolfi et al. [23] that there is only one infinite connected component $\mathscr{C}_\infty$ whenever $\lambda > \lambda_c$, a.s. A difficult question to answer is what happens at criticality for $\lambda_c > 0$. There is the following partial result, see Theorem 3 in Deprez et al. [17]: for $\alpha \in (d, 2d)$ and $\beta\alpha > 2d$, there does not exist an infinite connected component at criticality $\lambda_c > 0$. The case $\min\{\alpha, \beta\alpha\} > 2d$ is still open.

Next, we consider the graph distance $d(x, y)$, see also (4.7). We have the following result, see Deijfen et al. [16] and Theorem 8 in Deprez et al. [17].

Theorem 10 *Assume* $\min\{\alpha, \beta\alpha\} > d$.

(a) (infinite variance of degree distribution $1 < \tau < 2$*). Assume* $d < \beta\alpha < 2d$. *For any* $\lambda > \lambda_c = 0$ *there exists* $\eta_1 > 0$ *such that for all* $\varepsilon > 0$

$$\lim_{\|x\| \to \infty} \mathbb{P}_{\lambda,\alpha,\beta}\left[\eta_1 \leq \frac{d(0,x)}{\log\log \|x\|} \leq \frac{2}{|\log(\beta\alpha/d - 1)|} + \varepsilon \,\middle|\, 0, x \in \mathscr{C}_\infty \right] = 1.$$

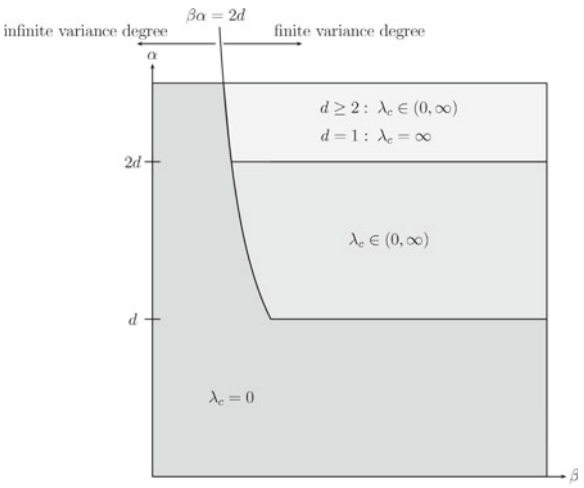


Fig. 4.4 Phase transition picture for $d \geq 1$

(b1) *(finite variance of degree distribution $\tau > 2$ case 1). Assume that $\beta\alpha > 2d$ and $\alpha \in (d, 2d)$. For any $\lambda > \lambda_c$ and any $\varepsilon > 0$*

$$\lim_{\|x\|\to\infty} \mathbb{P}_{\lambda,\alpha,\beta}\left[1-\varepsilon \le \frac{\log d(0,x)}{\log\log\|x\|} \le \Delta+\varepsilon \,\middle|\, 0, x \in \mathscr{C}_\infty\right] = 1,$$

where Δ was defined in Theorem 5.

(b2) *(finite variance of degree distribution $\tau > 2$ case 2). Assume $\min\{\alpha, \beta\alpha\} > 2d$. There exists $\eta_2 > 0$ such that*

$$\lim_{\|x\|\to\infty} \mathbb{P}_{\lambda,\alpha,\beta}\left[\eta_2 < \frac{d(0,x)}{\|x\|}\right] = 1.$$

Compare Theorem 10 (heterogeneous case) to Theorems 5 and 6 (homogeneous case). We observe that in the finite variance cases (b1)–(b2), i.e. for $\tau = \beta\alpha/d > 2$, we obtain the same behaviour for heterogeneous and homogeneous long-range percolation models. The infinite variance case (a) of the degree distribution, i.e. $1 < \tau < 2$ and $d < \beta\alpha < 2d$, respectively, is new. This infinite variance case provides a much slower decay of the graph distance, that is $d(0, x)$ is of order $\log\log\|x\|$ as $\|x\| \to \infty$. This is a pronounced version of the small-world effect, and this behaviour is similar to the NSW random graph model. Recall that empirical studies often suggest a tail parameter τ between 1 and 2 which corresponds to the infinite variance regime of the degree distribution. In Fig. 4.5 we illustrate Theorem 10 and we complete the picture about the chemical distances with the corresponding conjectures.

We conclude that this model fulfils all three stylised facts of small-world effect, the clustering property (which is induced by the Euclidean distance in the probability weights (4.9)) and the heavy-tailedness of the degree distribution.

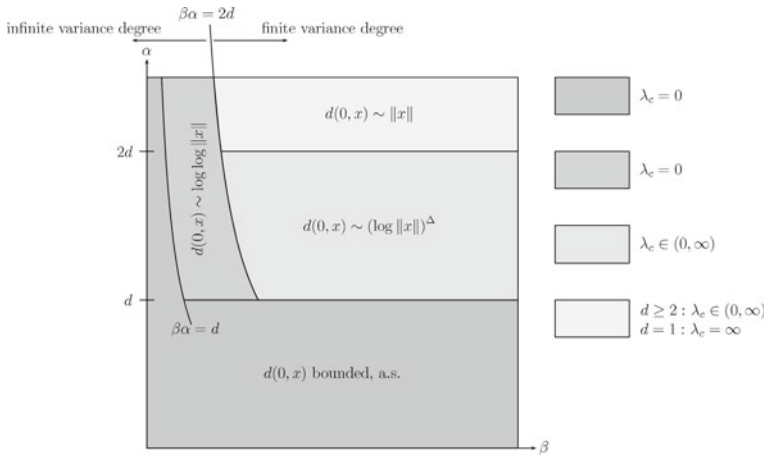


Fig. 4.5 Chemical distances according to Theorem 10 and corresponding conjectures

We end this section by giving an application of the heterogeneous long-range percolation model on $\mathbb{Z}^2$. Namely, we consider the network studied in Soramaki et al. [38] which analyzes the interbank payments transferred between 7,584 banks over the Fedwire Funds Service in the United States. We consider the graph whose particles represent these banks and there is a link between two banks if there is a transaction between them within a given day. We use heterogeneous long-range percolation to model such a daily graph where the observations of Soramaki et al. [38] allow to calibrate model parameters and to verify the stylised facts. The observations of Soramaki et al. [38] show that the stylised fact of having heavy-tailed degree distributions with tail parameter τ between 1 and 2 is fulfiled, note that τ can be directly estimated from the data. Moreover, Soramaki et al. [38] observe that the total number of payments sent from a given bank to its trading partners is heavy-tailed with estimated tail parameter $\hat{\beta} = 0.8$. We interpret this quantity as the size (or weight) of a bank and use $\hat{\beta} = 0.8$ as a calibration of β. To calibrate α we assume that all banks are located within a box of side-length $88 \approx \sqrt{7,584}$. Then, for $\alpha = 3.25$ and $\beta = 0.8$, Theorem 10 (a) states that graph distances are bounded by roughly 2.6 which is in line with the observations of Soramaki et al. [38]. In particular, we have a small-world effect. We conclude that the heterogeneous long-range percolation model on $\mathbb{Z}^2$ with parameters $\alpha = 3.25$ and $\beta = 0.8$ fit to the observations in Soramaki et al. [38]. Note that these parameters provide tail parameter $\tau = \beta\alpha/2 = 1.3$ and we get percolation for any choice of $\lambda > 0$.

4.7 Continuum Space Long-Range Percolation Model

The model of last section is restricted to the lattice $\mathbb{Z}^d$. A straightforward modification is to replace the lattice $\mathbb{Z}^d$ by a homogeneous Poisson point process X in $\mathbb{R}^d$. In comparison to the lattice model, some of the proofs simplify because we can apply classical integration in $\mathbb{R}^d$, other proofs become more complicated because one needs to make sure that the realisation of the Poisson point process is sufficiently regular in space. As in Deprez-Wüthrich [18] we consider a homogeneous marked Poisson point process in $\mathbb{R}^d$, where

- X denotes the spatially homogeneous Poisson point process in $\mathbb{R}^d$ with constant intensity $\nu > 0$. The individual particles of X are denoted by $x \in X \subset \mathbb{R}^d$;
- W_x, $x \in X$, are i.i.d. marks having a Pareto distribution with threshold parameter 1 and tail parameter $\beta > 0$, see (4.8).

Choose $\alpha > 0$ and $\lambda > 0$ fixed. Conditionally given X and $(W_x)_{x \in X}$, we consider the edge probabilities for $x, y \in X$ given by

$$p_{x,y} = 1 - \exp(-\lambda W_x W_y \|x - y\|^{-\alpha}). \tag{4.10}$$

Between any pair $x, y \in X$ we attach an edge, independently of all other edges, as follows

$$\eta_{x,y} = \eta_{y,x} = \begin{cases} 1 & \text{with probability } p_{x,y}, \\ 0 & \text{with probability } 1 - p_{x,y}. \end{cases}$$

We denote the resulting probability measure on the edge configuration by $\mathbb{P}_{\nu,\lambda,\alpha,\beta}$. Figure 4.3 (rhs) shows part of a realised configuration. We have the following result for the degree distribution, see Proposition 3.2 and Theorem 3.3 in Deprez-Wüthrich [18].

Theorem 11 *Fix $d \geq 1$. We have the following two cases for the degree distribution:*

- *for $\min\{\alpha, \beta\alpha\} \leq d$, a.s., $\mathbb{P}_0[\mathscr{D}(0) = \infty | W_0] = 1$;*
- *for $\min\{\alpha, \beta\alpha\} > d$ set $\tau = \beta\alpha/d$, then*

$$\mathbb{P}_0\left[\mathscr{D}(0) > k\right] = k^{-\tau}\ell(k),$$

for some function $\ell(\cdot)$ that is slowly varying at infinity.

Remarks.

- Note that the previous statement needs some care because we need to make sure that there is a particle at the origin. This is not straightforward in the Poisson case and $\mathbb{P}_0$ can be understood as the conditional distribution, conditioned on having a particle at the origin. The formally precise construction is known as the Palm distribution, which considers distributions shifted by the particles in the Poisson cloud X.
- In analogy to the homogeneous long-range percolation model in $\mathbb{Z}^d$ we could also consider continuum space homogeneous long-range percolation in $\mathbb{R}^d$. This is achieved by setting $W_x = W_y = 1$, a.s., in (4.10). In this case, the proof of the statement equivalent to (4.6) becomes rather easy. We briefly give the details in the next lemma, see also proof of Lemma 3.1 in Deprez-Wüthrich [18].

Lemma 1 *Choose $W_x = W_y = 1$, a.s., in (4.10). For $\alpha \leq d$ we have, a.s., $\mathscr{D}(0) = \infty$; for $\alpha > d$ the degree $\mathscr{D}(0)$ has a Poisson distribution.*

Proof of Lemma 1 and (4.6) in continuum space Let X be a Poisson cloud with $0 \in X$ and denote by $X(A)$ the number of particles in $X \cap A$ for $A \subset \mathbb{R}^d$. Every particle $x \in X \setminus \{0\}$ is now independently from the others removed from the Poisson cloud with probability $1 - p_{0,x}$. The resulting process $\widetilde{X}$ is a thinned Poisson cloud having intensity function $x \mapsto \nu p_{0,x} = \nu(1 - \exp(-\lambda\|x\|^{-\alpha})) \sim \nu\lambda\|x\|^{-\alpha}$ as $\|x\| \to \infty$. Since $\mathscr{D}(0) = \widetilde{X}(\mathbb{R}^d \setminus \{0\})$ it follows that $\mathscr{D}(0)$ is infinite, a.s., if $\alpha \leq d$ and that $\mathscr{D}(0)$ has a Poisson distribution otherwise. To see this let μ denote the Lebesgue measure in $\mathbb{R}^d$ and choose a finite Borel set $A \subset \mathbb{R}^d$ containing the origin. Since A contains the origin, we have $X(A) \geq \widetilde{X}(A) \geq 1$. This motivates for $k \in \mathbb{N}_0$ to study

$$\mathbb{P}_0\left[\widetilde{X}(A) = k + 1\right] = \sum_{i \geq k} \mathbb{P}_0\left[\widetilde{X}(A) = k + 1 \middle| X(A) = i + 1\right] \mathbb{P}_0\left[X(A) = i + 1\right].$$

Since A contains the origin, the case $i = 0$ is trivial, i.e. $\mathbb{P}_0[\widetilde{X}(A) = 1 | X(A) = 1] = 1$. There remains $i \geq 1$. Conditionally on $\{X(A) = i + 1\}$, the i particles (excluding the origin) are independent and uniformly distributed in A. The conditional moment generating function for $r \in \mathbb{R}$ is then given by

$$\begin{aligned}
&\mathbb{E}_0\left[\exp\left\{r\left(\widetilde{X}(A) - 1\right)\right\} \middle| X(A) = i + 1\right] \\
&\quad = \frac{1}{\mu(A)^i} \int_{A\times\cdots\times A} \mathbb{E}_0\left[\exp\left\{r \sum_{l=1}^{i} \eta_{0,x_l}\right\}\right] dx_1 \ldots dx_i \\
&\quad = \frac{1}{\mu(A)^i} \int_{A\times\cdots\times A} \prod_{l=1}^{i} \mathbb{E}_0\left[\exp\left\{r\eta_{0,x_l}\right\}\right] dx_1 \ldots dx_i \\
&\quad = \left(\frac{1}{\mu(A)} \int_A \mathbb{E}_0\left[\exp\left\{r\eta_{0,x}\right\}\right] dx\right)^i .
\end{aligned}$$

We calculate the integral for $W_0 = W_x = 1$, a.s., in (4.10)

$$\begin{aligned}
\frac{1}{\mu(A)} \int_A \mathbb{E}_0\left[\exp\left\{r\eta_{0,x}\right\}\right] dx &= \frac{1}{\mu(A)} \int_A e^r p_{0,x} + (1 - p_{0,x})\, dx \\
&= e^r p(A) + (1 - p(A)),
\end{aligned}$$

with $p(A) = \mu(A)^{-1} \int_A p_{0,x} dx \in (0, 1)$. Thus, conditionally on $\{X(A) = i + 1\}$, $\widetilde{X}(A) - 1$ has a binomial distribution with parameters i and $p(A)$. This implies that

$$\begin{aligned}
\mathbb{P}_0\left[\widetilde{X}(A) = k + 1\right] &= \sum_{i\geq k} \binom{i}{k} p(A)^k (1 - p(A))^{i-k}\ \mathbb{P}_0\left[X(A) = i + 1\right] \\
&= \sum_{i\geq k} \frac{p(A)^k}{k!} \frac{(1 - p(A))^{i-k}}{(i - k)!}\ \exp\{-\nu\mu(A)\}(\nu\mu(A))^i \\
&= \frac{(\nu\mu(A)p(A))^k}{k!} \sum_{i\geq k} \frac{(\nu\mu(A)(1 - p(A)))^{i-k}}{(i - k)!} \exp\{-\nu\mu(A)\} \\
&= \exp\left\{-\nu\mu(A)p(A)\right\} \frac{(\nu\mu(A)p(A))^k}{k!} \\
&= \exp\left\{-\nu \int_A p_{0,x} dx\right\} \frac{\left(\nu \int_A p_{0,x} dx\right)^k}{k!}.
\end{aligned}$$

This implies that $\widetilde{X}$ is a non-homogeneous Poisson point process with intensity function

$$x \mapsto \nu p_{0,x} = \nu\left(1 - \exp(-\lambda\|x\|^{-\alpha})\right) \sim \nu\lambda\|x\|^{-\alpha}, \quad \text{as} \|x\| \to \infty.$$

But this immediately implies that the degree distribution $\mathscr{D}(0) = \widetilde{X}(\mathbb{R}^d \setminus \{0\})$ is infinite, a.s., if $\alpha \leq d$, and that it has a Poisson distribution otherwise. This finishes the proof. □

We now switch back to the heterogeneous long-range percolation model (4.10). We consider the connected component $\mathscr{C}(0)$ of a particle in the origin under the Palm distribution $\mathbb{P}_0$. We define the percolation probability

$$\theta(\lambda) = \mathbb{P}_0 \left[|\mathscr{C}(0)| = \infty\right].$$

The *critical percolation value* λ_c is then defined by

$$\lambda_c = \inf \left\{\lambda > 0 : \ \theta(\lambda) > 0\right\}.$$

We have the following results, see Theorem 3.4 in Deprez-Wüthrich [18].

Theorem 12 *Fix $d \geq 1$. Assume $\min\{\alpha, \beta\alpha\} > d$.*

(a) If $d \geq 2$, then $\lambda_c < \infty$.
(b) If $d = 1$ and $\alpha \in (1, 2]$, then $\lambda_c < \infty$.
(c) If $d = 1$ and $\min\{\alpha, \beta\alpha\} > 2$, then $\lambda_c = \infty$.

Theorem 13 *Fix $d \geq 1$. Assume $\min\{\alpha, \beta\alpha\} > d$.*

(a) If $\beta\alpha < 2d$, then $\lambda_c = 0$.
(b) If $\beta\alpha > 2d$, then $\lambda_c > 0$.

These are the continuum space analogues to Theorems 8 and 9, for an illustration see also Fig. 4.4. The work on the graph distances in the continuum space long-range percolation model is still work in progress, but we expect similar results to the ones in Theorem 10, see also Fig. 4.5. However, proofs in the continuum space model are more sophisticated due to the randomness of the positions of the particles.

The advantage of the latter continuum space model (with homogeneous marked Poisson point process) is that it can be extended to non-homogeneous Poisson point processes. For instance, if certain areas are more densely populated than others we can achieve such a non-homogeneous space model by modifying the constant intensity ν to a space-dependent density function $\nu(\cdot) : \mathbb{R}^d \to \mathbb{R}_+$.

4.8 Renormalisation Techniques

In this section, we present a crucial technique that is used in many of the proofs of the previous statements about existences of infinite connected components and about graph distances. These proofs are often based on renormalisation techniques. That is, one collects particles in boxes. These boxes are defined to be either *good*

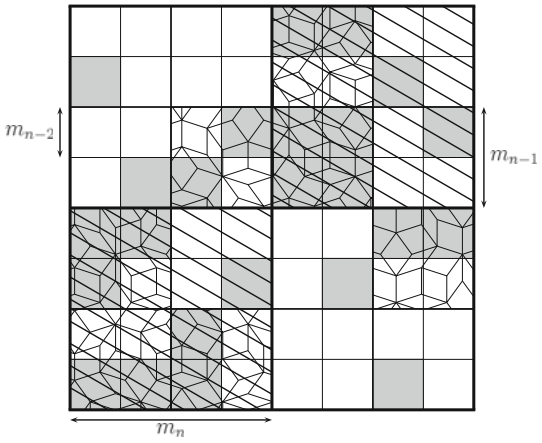


Fig. 4.6 Example of the renormalisation technique. Define inductively the box lengths $m_n = 2m_{n-1}$, $n \in \mathbb{N}$, for some initial $m_0 \in \mathbb{N}$. Call translates of $[0, m_0 - 1]^d$ to be 0-stage boxes and assume that goodness of such boxes is defined. For $n \in \mathbb{N}$ we call the translates of $[0, m_n - 1]^d$ n-stage boxes and we inductively say that an n-stage box is good if it contains at least two good $(n-1)$-stage boxes. Assume that in the illustration good $(n-2)$-stage boxes, $n \geq 2$, are coloured in grey. The good $(n-1)$-stage boxes are then the Penrose-patterned boxes of side-length m_{n-1} and the good n-stage boxes are striped. The illustrated $(n+1)$-stage box is good, because it contains two good n-stage boxes

(having a certain property) or *bad* (not possessing this property). These boxes are then again merged to bigger good or bad boxes. These scalings and renormalisations are done over several generations of box sizes, see Fig. 4.6 for an illustration. The purpose of these rescalings is that one arrives at a certain generation of box sizes that possesses certain characteristics to which classical site-bond percolation results apply. We exemplify this with a particular example.

4.8.1 Site-Bond Percolation

Though we will not directly use site-bond percolation, we start with the description of this model because it is often useful. Site-bond percolation in $\mathbb{Z}^d$ is a modification of homogeneous long-range percolation introduced in Sect. 4.5. Choose a fixed dimension $d \geq 1$ and consider the square lattice $\mathbb{Z}^d$. Assume that every site $x \in \mathbb{Z}^d$ is occupied independently with probability $r^* \in [0, 1]$ and every bond between x and y in $\mathbb{Z}^d$ is occupied independently with probability

$$p^*_{x,y} = 1 - \exp(-\lambda^* \|x - y\|^{-\alpha}), \tag{4.11}$$

for given parameters $\lambda^* > 0$ and $\alpha > 0$. The connected component $\mathscr{C}^*(x)$ of a given site $x \in \mathbb{Z}^d$ is then defined to be the set of all occupied sites $y \in \mathbb{Z}^d$ such that x and y are connected by a path only running through occupied sites and occupied bonds (if x is not occupied then $\mathscr{C}^*(x)$ is the empty set). We can interpret this as follows: we place particles at sites $x \in \mathbb{Z}^d$ at random with probability r^*. This defines a (random) subset of $\mathbb{Z}^d$, and then, we consider long-range percolation on this random subset, i.e. this corresponds to a thinning of homogeneous long-range percolation in $\mathbb{Z}^d$. We can then study the percolation properties of this site-bond percolation model, some results are presented in Lemma 3.6 of Biskup [9] and in the proof of Theorem 2.5 of Berger [6]. The aim in many proofs in percolation theory is to define different generations of box sizes using renormalisations, see Fig. 4.6. We perform these renormalisations until we arrive at a generation of box sizes for which good boxes occur sufficiently often. If this is the case and if all the necessary dependence assumptions are fulfilled we can apply classical site-bond percolation results.

In order to simplify our outline we use a modified version of the homogeneous long-range percolation model (4.5) of Sect. 4.5. We set $p = 1 - \exp(-\lambda)$ and obtain the following model.

Model 14 (modified homogeneous long-range percolation) Fix $d \geq 2$. Choose $\alpha > 0$ and $\lambda > 0$ fixed and define the edge probabilities for $x, y \in \mathbb{Z}^d$ by

$$p_{x,y} = 1 - \exp(-\lambda \|x - y\|^{-\alpha}).$$

Then edges between all pairs of particles $x, y \in \mathbb{Z}^d$ are attached independently with edge probability $p_{x,y}$ and the probability measure of the resulting edge configurations $\eta = (\eta_{x,y})_{x,y \in \mathbb{Z}^d}$ is denoted by $\mathbb{P}_{\lambda,\alpha}$.

Note that this model is a special case of site-bond percolation with $r^* = 1$ and $\lambda^* = \lambda$ in (4.11).

4.8.2 *Largest Semi-clusters*

In order to demonstrate the renormalisation technique we repeat the proof of Lemma 2.3 of Berger [6] in the modified homogeneous long-range percolation Model 14, see Theorem 15 below. This proof is rather sophisticated because it needs a careful treatment of dependence and we revisit the second version of the proof of Lemma 2.3 provided in Berger [8].

Fix $\alpha \in (d, 2d)$ and choose $\lambda > 0$ so large that there exists a unique infinite connected component, a.s., having density $\kappa > 0$ (which exists due to Theorem 4). Choose $M \geq 1$ and $K \geq 0$ integer valued. For $v \in \mathbb{Z}^d$ we define box B_v and its K-enlargement $B_v^{(K)}$ by

$$B_v = Mv + [0, M-1]^d \quad \text{and} \quad B_v^{(K)} = Mv + [-K, M+K-1]^d.$$

For every box B_v we define a *ℓ-semi-cluster* to be a set of at least ℓ sites in B_v which are connected within $B_v^{(K)}$. For any $\varepsilon > 0$ there exists $M' \geq 1$ such that for all $M \geq M'$ and some $K \geq 0$ we have

$$\mathbb{P}_{\lambda,\alpha}[\text{at least } M^d\kappa/2 \text{ sites of } B_v \text{ belong to the infinite connected component and these sites are connected within } B_v^{(K)}] \geq 1 - \varepsilon/2. \tag{4.12}$$

Existence of $M' \geq 1$ follows from the ergodic theorem and existence of K from the fact that the infinite connected component is unique, a.s., and therefore, all sites in B_v belonging to the infinite connected component need to be connected within a certain K-enlargement of B_v. Formula (4.12) says that we have a $(M^d\kappa/2)$-semi-cluster in B_v with at least probability $1 - \varepsilon/2$. We first show uniqueness of large semi-clusters.

Lemma 2 *Choose $\xi \in (\alpha/d, 2)$ and $\gamma \in (0,1)$ with $18\gamma > 16 + \xi$. There exist $\varphi = \varphi(\xi,\gamma) > 0$ and $M' = M'(\xi,\gamma) \geq 1$ such that for all $M \geq M'$ and all $K \geq 0$ we have*

$$\mathbb{P}_{\lambda,\alpha}\left[\textit{there is at most one } M^{d\gamma}\textit{-semi-cluster in } B_v\right] > 1 - M^{-d\varphi},$$

where by "at most one" we mean that there is no second $M^{d\gamma}$-semi-cluster in B_v which is not connected to the first one within $B_v^{(K)}$.

Proof of Lemma 2 The proof uses the notion of inhomogeneous random graphs as defined in Aldous [2]. An inhomogeneous random graph $H(N,\xi)$ with size N and parameter ξ is a set of particles $\{1,\dots,k\}$ and corresponding masses $s_1,\dots,s_k$ such that $N = \sum_{i=1}^k s_i$; and any $i \neq j$ are connected independently with probability $1 - \exp\left(-s_i s_j N^{-\xi}\right)$. From Lemma 2.5 of Berger [8] we know that for any $1 < \xi < 2$ and $0 < \gamma < 1$ with $18\gamma > 16 + \xi$, there exist $\varphi = \varphi(\xi,\gamma) > 0$ and $N' = N'(\xi,\gamma) \geq 1$ such that for all $N \geq N'$ and every inhomogeneous random graph with size N and parameter ξ we have

$$\mathbb{P}\left[H(N,\xi) \text{ contains more than one connected component } C \text{ with } \sum_{i\in C} s_i \geq N^\gamma\right] < N^{-\varphi}. \tag{4.13}$$

We now show uniqueness of $M^{d\gamma}$-semi-clusters in B_v. Choose $\xi \in (\alpha/d, 2)$ and $\gamma \in (0,1)$ such that $18\gamma > 16 + \xi$. For any $x, y \in B_v$ we have $\|x - y\| \leq \sqrt{d}M$. Choose M so large that $\lambda(\sqrt{d}M)^{-\alpha} > M^{-d\xi}$ and choose $K \geq 0$ arbitrarily. Particles $x, y \in B_v$ are then attached with probability $p_{x,y}$ uniformly bounded by

$$\begin{aligned} p_{x,y} &= 1 - \exp(-\lambda\|x-y\|^{-\alpha}) \\ &\geq 1 - \exp\left(-\lambda\left(\sqrt{d}M\right)^{-\alpha}\right) > 1 - \exp\left(-M^{-d\xi}\right) = \nu > 0, \end{aligned}$$

where the last equality defines ν. This allows to decouple the sampling of edges $\eta = (\eta_{x,y})_{x,y\in\mathbb{Z}^d}$ in B_v. For every $x, y \in B_v$, define $p'_{x,y} \in (0,1)$ by

$$p_{x,y} = p'_{x,y} + \nu - \nu p'_{x,y}.$$

We now sample $\eta = (\eta_{x,y})_{x,y\in\mathbb{Z}^d}$ in two steps. We first sample η' according to Model 14 but with edge probabilities $p'_{x,y}$ if $x, y \in B_v$ and with edge probabilities $p_{x,y}$ otherwise. Secondly, we sample η'' as an independent configuration on B_v where there is an edge between x and y with edge probability ν for $x, y \in B_v$. By definition of $p'_{x,y}$ we get that $\eta' \vee \eta'' \overset{(\mathrm{d})}{=} \eta$. Let $S_1, S_2 \subset B_v$ be two disjoint maximal sets of sites in B_v that are η'-connected within $B_v^{(K)}$, i.e. S_1 and S_2 are two disjoint maximal semi-clusters in B_v for given edge configuration η'. Note that by maximality

$$\begin{aligned}
&\mathbb{P}\left[\text{there is an } \eta\text{-edge between } S_1 \text{ and } S_2 \middle| \eta'\right] \\
&\quad = \mathbb{P}\left[\text{there is an } \eta''\text{-edge between } S_1 \text{ and } S_2 \middle| \eta'\right] \\
&\quad = 1 - (1-\nu)^{|S_1||S_2|} \;=\; 1 - \exp\left(-|S_1||S_2|M^{-d\xi}\right).
\end{aligned}$$

If we denote by $S_1, \ldots, S_k$ all disjoint maximal semi-clusters in B_v for given edge configuration η' then we see that these maximal semi-clusters form an inhomogeneous random graph of size $\sum_{i=1}^k |S_i| = M^d$ and parameter ξ. Therefore, there exist $\varphi > 0$ and $M' \geq 1$ such that for all $M \geq M'$ and all $K \geq 0$ we have from (4.13)

$$\mathbb{P}\Bigg[H(M^d, \xi) \text{ contains more than one connected component } C \\ \text{with } \sum_{i\in C} |S_i| \geq M^{d\gamma} \Bigg| \eta' \Bigg] < M^{-d\varphi}.$$

Note that this bound is uniform in η' and $K \geq 0$. Therefore, the probability of having at least two $M^{d\gamma}$-semi-clusters in B_v which are not connected within $B_v^{(K)}$ is bounded by $M^{-d\varphi}$. □

We can now combine (4.12) and Lemma 2. Choose $\varepsilon > 0$. For all M sufficiently large and $K \geq 0$ such that (4.12) holds we have

$$\mathbb{P}_{\lambda,\alpha}\left[\text{there is exactly one } (M^d\kappa/2)\text{-semi-cluster in } B_v\right] \geq 1 - \varepsilon, \tag{4.14}$$

where by "exactly one" we mean that there is no other $(M^d\kappa/2)$-semi-cluster in B_v which is not connected to the first one within $B_v^{(K)}$. This follows because of $\gamma < 1$, which implies that $M^{d\gamma} \leq M^d\kappa/2$ for all M sufficiently large, and because $M^{-d\varphi} < \varepsilon/2$ for all M sufficiently large.

4.8.3 Renormalisation

Choose $\varepsilon > 0$ fixed, and $M > 1$ and $K \geq 0$ such that (4.14) holds. For $v \in \mathbb{Z}^d$ we say that box B_v is *good* if there is exactly one $(M^d\kappa/2)$-semi-cluster in B_v (where exactly one is meant in the sense of above). Therefore, on good boxes there are at least $M^d\kappa/2$ sites in B_v that are connected within $B_v^{(K)}$ and we have

$$\mathbb{P}_{\lambda,\alpha}[B_v \text{ is good}] \geq 1 - \varepsilon. \tag{4.15}$$

Note that the goodness properties of B_{v_1} and B_{v_2} for $v_1 \neq v_2 \in \mathbb{Z}^d$ are not necessarily independent because their K-enlargements $B_{v_1}^{(K)}$ and $B_{v_2}^{(K)}$ may overlap.

Now, we define renormalisation over different generations $n \in \mathbb{N}_0$; terminology *n-stage* is referred to the nth generation. Choose an integer valued sequence $a_n > 1$, $n \in \mathbb{N}_0$, with $a_0 = M$ and define the box lengths $(M_n)_{n\in\mathbb{N}_0}$ as follows: set $M_0 = a_0 = M$ and for $n \in \mathbb{N}$

$$M_n = a_n M_{n-1} = M_0 \prod_{i=1}^{n} a_i = \prod_{i=0}^{n} a_i.$$

Define the n-stage boxes, $n \in \mathbb{N}_0$, by

$$B_{n,v} = M_n v + [0, M_n - 1]^d \quad \text{with } v \in \mathbb{Z}^d.$$

Note that n-stage boxes $B_{n,v}$ have volume $M_n^d = a_n^d M_{n-1}^d = \prod_{i=0}^{n} a_i^d$ and every n-stage box $B_{n,v}$ contains a_n^d of $(n-1)$-stage boxes $B_{n-1,x} \subset B_{n,v}$, and $(M_n/a_0)^d = \prod_{i=1}^{n} a_i^d$ of 0-stage boxes $B_x = B_{0,x} \subset B_{n,v}$, see also Fig. 4.6.

Renormalisation. We define goodness of n-stage boxes $B_{n,v}$ recursively for a given sequence $\kappa_n \in (0, 1)$, $n \in \mathbb{N}_0$, of densities where we initialise $\kappa_0 = \kappa/2$.
(i) *Initialisation* $n = 0$. We say that 0-stage box $B_{0,v}$, $v \in \mathbb{Z}^d$, is good if it contains exactly one $(\kappa_0 a_0^d)$-semi-cluster. Due to our choices of $M > 1$ and $K \geq 0$ we see that the goodness of 0-stage box $B_{0,v}$ occurs with at least probability $1 - \varepsilon$, see (4.15).
(ii) *Iteration* $n - 1 \to n$. Choose $n \in \mathbb{N}$ and assume that goodness of $(n-1)$-stage boxes $B_{n-1,v}$, $v \in \mathbb{Z}^d$, has been defined. For $v \in \mathbb{Z}^d$ we say that n-stage box $B_{n,v}$ is good if the event $A_{n,v} = A_{n,v}^{(a)} \cap A_{n,v}^{(b)}$ occurs, where

(a) $A_{n,v}^{(a)} = \{\text{at least } \kappa_n a_n^d \text{ of the } (n-1)\text{-stage boxes } B_{n-1,x} \subset B_{n,v} \text{ are good; and}$
(b) $A_{n,v}^{(b)} = \{\text{all}(\prod_{i=0}^{n-1} \kappa_i a_i^d)\text{-semi-clusters of all good } (n-1)\text{-stage boxes}$
in $B_{n,v}$ are connected within $B_{n,v}^{(K)}\}$. □

Observe that on event $A_{n,v}$ the n-stage box $B_{n,v}$ contains at least $\prod_{i=0}^{n} \kappa_i a_i^d$ sites that are connected within the K-enlargement $B_{n,v}^{(K)}$ of $B_{n,v}$. We set density $u_n = \prod_{i=0}^{n} \kappa_i$ which gives

$$\prod_{i=0}^{n} \kappa_i a_i^d = M_n^d \prod_{i=0}^{n} \kappa_i = M_n^d u_n.$$

Therefore, good n-stage boxes contain $(M_n^d u_n)$-semi-clusters. Our next aim is to calculate the probability p_n of having a good n-stage box. The case $n = 0$ follows from (4.15), i.e. for any $\varepsilon > 0$ and any M sufficiently large there exists $K \geq 0$ such that

$$p_0 \;=\; \mathbb{P}_{\lambda,\alpha}[B_{0,v} \text{ is good}] \;=\; \mathbb{P}_{\lambda,\alpha}[B_v \text{ is good}] \;\geq\; 1 - \varepsilon.$$

Theorem 15 *Assume $\alpha \in (d, 2d)$. Choose $\lambda > 0$ so large that we have a unique infinite connected component, a.s., having density $\kappa > 0$. For every $\varepsilon' \in (0, 1)$ there exists $N_0 \geq 1$ such that for all $N \geq N_0$*

$$\mathbb{P}_{\lambda,\alpha}\left[|C_N| \geq N^{\alpha/2}\right] \geq 1 - \varepsilon',$$

where C_N is the largest connected component in $[0, N-1]^d$.

Note that for density $\kappa > 0$ of the infinite connected component we expect roughly κN^d sites in box $[0, N-1]^d$ belonging to the infinite connected component. The above lemma however says that at least $N^{\alpha/2}$ sites in $[0, N-1]^d$ are connected *within that box*. That is, here we do not need any K-enlargements as in (4.12) and, therefore, this event is independent for different disjoint boxes $vN + [0, N-1]^d$ and we may apply classical site-bond percolation results.

Proof of Theorem 15 Choose $\alpha \in (d, 2d)$ and $\varepsilon' \in (0, 1)$ fixed. As in Lemma 2.3 of Berger [8] we now make a choice of parameters and sequences which will provide the statement of Theorem 15. Choose $\xi \in (\alpha/d, 2)$ and $\gamma \in (0, 1)$ such that $18\gamma > 16 + \xi$. Choose $\delta > \vartheta > 1$ with $2\vartheta < \delta(2d - \alpha)$ and $d\delta - \vartheta > d\gamma\delta$. Note that this is possible because it requires that $\delta \min\{1, (2d - \alpha)/2, d(1 - \gamma)\} > \vartheta > 1$. Define for $n \in \mathbb{N}$

$$\kappa_n = (n+1)^{-\vartheta} \quad \text{and} \quad a_n = (n+1)^{\delta}. \tag{4.16}$$

For simplicity, we assume that δ is an integer which implies that also $a_n > 1$ is integer valued, and $\kappa_n \in (0, 1)$ will play the role of densities introduced above. Observe that for $\vartheta > 1$ we have for all $n \geq 1$

$$\prod_{l=1}^{n}(1 + 3\kappa_l) \;\leq\; \lim_{n\to\infty}\prod_{l=1}^{n}(1 + 3\kappa_l) \;=\; c_1 \in (1, \infty). \tag{4.17}$$

Choose $\varepsilon \in (0, \varepsilon'/c_1) \subset (0, 1)$ fixed.

There still remains the choice of $a_0 = M \geq 1$ and $\kappa_0 \in (0,1)$. We set $\kappa_0 = \kappa/2$. Note that choices (4.16) imply

$$(2M_{n-1}^d)^\gamma = 2^\gamma M^{d\gamma}(n!)^{d\gamma\delta} \quad \text{and} \quad M_{n-1}^d u_{n-1} = \frac{\kappa}{2} M^d (n!)^{d\delta-\vartheta}.$$

Therefore,

$$\frac{M_{n-1}^d u_{n-1}}{(2M_{n-1}^d)^\gamma} = \frac{\kappa}{2^{1+\gamma}} M^{d(1-\gamma)}(n!)^{d\delta-\vartheta-d\gamma\delta}. \tag{4.18}$$

Because of $d\delta - \vartheta > d\gamma\delta$ the right-hand side of (4.18) is uniformly bounded from below in $n \geq 1$ and for M sufficiently large the right-hand side of (4.18) is strictly bigger than 1 for all $n \geq 1$. Therefore, there exists $m_1 \geq 1$ such that for all $M \geq m_1$ and all $n \geq 1$ we have

$$(2M_{n-1}^d)^\gamma < M_{n-1}^d u_{n-1}. \tag{4.19}$$

Next we are going to bound for $n \in \mathbb{N}_0$ the probabilities

$$p_n = \mathbb{P}_{\lambda,\alpha}[B_{n,v} \text{ is good}] = \mathbb{P}_{\lambda,\alpha}[A_{n,v}].$$

We have for $n \geq 1$

$$\begin{aligned} 1 - p_n &= \mathbb{P}_{\lambda,\alpha}\left[A_{n,v}^{\mathrm{c}}\right] \\ &= \mathbb{P}_{\lambda,\alpha}\left[(A_{n,v}^{(a)} \cap A_{n,v}^{(b)})^{\mathrm{c}}\right] \leq \mathbb{P}_{\lambda,\alpha}\left[(A_{n,v}^{(a)})^{\mathrm{c}}\right] + \mathbb{P}_{\lambda,\alpha}\left[(A_{n,v}^{(b)})^{\mathrm{c}}\right]. \end{aligned} \tag{4.20}$$

For the first term in (4.20) we have, using Markov's inequality and translation invariance,

$$\begin{aligned} \mathbb{P}_{\lambda,\alpha}\left[(A_{n,v}^{(a)})^{\mathrm{c}}\right] &= \mathbb{P}_{\lambda,\alpha}\left[\sum_{B_{n-1,x}\subset B_{n,v}} 1_{A_{n-1,x}} < \kappa_n a_n^d\right] \\ &= \mathbb{P}_{\lambda,\alpha}\left[\sum_{B_{n-1,x}\subset B_{n,v}} 1_{A_{n-1,x}^{\mathrm{c}}} > (1-\kappa_n) a_n^d\right] \\ &\leq \frac{1}{(1-\kappa_n)a_n^d} \sum_{B_{n-1,x}\subset B_{n,v}} \mathbb{P}_{\lambda,\alpha}\left[A_{n-1,x}^{\mathrm{c}}\right] \\ &= \frac{1}{1-\kappa_n}\, \mathbb{P}_{\lambda,\alpha}\left[A_{n-1,v}^{\mathrm{c}}\right] = \frac{1-p_{n-1}}{1-\kappa_n}. \end{aligned}$$

The second term in (4.20) is more involved due to possible dependence in the K-enlargements. Choose $\varphi = \varphi(\xi,\gamma) > 0$ and $M'(\xi,\gamma) \geq 1$ as in Lemma 2. On event $(A_{n,v}^{(b)})^{\mathrm{c}}$ there exist at least two $(M_{n-1}^d u_{n-1})$-semi-clusters in good $(n-1)$-stage boxes B_{n-1,v_1} and B_{n-1,v_2} in $B_{n,v}$ that are not connected within the K-enlargement $B_{n,v}^{(K)}$.

Define $B = B_{n-1,v_1} \cup B_{n-1,v_2}$. Note that B has volume $2M_{n-1}^d$ and that any $x, y \in B_{n,v}$ have maximal distance $\sqrt{d}M_n$. We analyse the following ratio

$$\frac{\lambda(\sqrt{d}M_n)^{-\alpha}}{(2M_{n-1}^d)^{-\xi}} = \lambda d^{-\alpha/2} 2^{\xi}\ M^{d\xi-\alpha}(n!)^{(d\xi-\alpha)\delta}\ (n+1)^{-\alpha\delta}.$$

Note that $d\xi > \alpha$. This implies that the right-hand side of the previous equality is uniformly bounded from below in $n \geq 1$. Therefore, there exists $m_2 \geq m_1$ such that for all $M \geq m_2$ and all $n \geq 1$ inequality (4.19) holds and

$$\lambda(\sqrt{d}M_n)^{-\alpha} > (2M_{n-1}^d)^{-\xi}. \tag{4.21}$$

This choice implies that for any $x, y \in B$ we have

$$\begin{aligned} p_{x,y} &= 1 - \exp(-\lambda\|x-y\|^{-\alpha}) \\ &\geq 1 - \exp\left(-\lambda\left(\sqrt{d}M_n\right)^{-\alpha}\right) > 1 - \exp\left(-(2M_{n-1}^d)^{-\xi}\right) = \nu_n > 0, \end{aligned}$$

where the last equality defines ν_n. We now proceed as in Lemma 2. Decouple the sampling of edges $\eta = (\eta_{x,y})_{x,y\in\mathbb{Z}^d}$ in B. For every $x, y \in B$, define $p'_{x,y} \in (0,1)$ by

$$p_{x,y} = p'_{x,y} + \nu_n - \nu_n p'_{x,y}.$$

We again sample $\eta = (\eta_{x,y})_{x,y\in\mathbb{Z}^d}$ in two steps. We first sample η' according to Model 14 but with edge probabilities $p'_{x,y}$ if $x, y \in B$ and with edge probabilities $p_{x,y}$ otherwise. Second, we sample η'' as an independent configuration on B where there is an edge between x and y with edge probability ν_n for $x, y \in B$. By definition of $p'_{x,y}$ we get that $\eta' \vee \eta'' \overset{(\mathrm{d})}{=} \eta$. Let $S_1, S_2 \subset B$ be two disjoint maximal sets of sites in B that are η'-connected within $B_{n,v}^{(K)}$, i.e. S_1 and S_2 are two disjoint maximal semi-clusters in B for given edge configuration η'. Note that by maximality

$$\begin{aligned} &\mathbb{P}\left[\text{there is an } \eta\text{-edge between } S_1 \text{ and } S_2 \middle|\, \eta'\right] \\ &\quad = \mathbb{P}\left[\text{there is an } \eta''\text{-edge between } S_1 \text{ and } S_2 \middle|\, \eta'\right] \\ &\quad = 1 - (1-\nu_n)^{|S_1||S_2|} = 1 - \exp\left(-|S_1||S_2|(2M_{n-1}^d)^{-\xi}\right). \end{aligned}$$

If we denote by $S_1, \ldots, S_k$ all disjoint maximal semi-clusters in B for given edge configuration η' then we see that these maximal semi-clusters form an inhomogeneous random graph of size $\sum_{i=1}^k |S_i| = 2M_{n-1}^d \geq 2M^d$ and parameter ξ. Therefore, for choices $\varphi = \varphi(\xi,\gamma) > 0$ and $m_3 \geq \max\{m_2, M'(\xi,\gamma)\}$ (where $\varphi(\xi,\gamma)$ and $M'(\xi,\gamma)$ were given by Lemma 2) we have that for all $M \geq m_3$ and all $n \geq 1$ inequality (4.19) holds, and for all $K \geq 0$ we have from (4.13)

$$\mathbb{P}\Big[H(2M_{n-1}^d, \xi) \text{ contains more than one connected component } C \text{ with } \sum_{i \in C} |S_i| \geq (2M_{n-1}^d)^\gamma \Big| \eta' \Big] < (2M_{n-1}^d)^{-\varphi}.$$

Note that this bound is uniform in η' and $K \geq 0$ and holds for all $n \geq 1$. Therefore, the probability of having at least two $(2M_{n-1}^d)^\gamma$-semi-clusters in B which are not connected within $B_{n,v}^{(K)}$ is bounded by $(2M_{n-1}^d)^{-\varphi}$. Next, we use that for all $m \geq m_3$ inequality (4.19) holds. Therefore, we get for all $M \geq m_3$, all $n \geq 1$ and all $K \geq 0$

$$\mathbb{P}_{\lambda,\alpha}\left[\text{there are at least two } (M_{n-1}^d u_{n-1})\text{-semi-clusters in } B\right] < (2M_{n-1}^d)^{-\varphi}.$$

Note that $B_{n,v}$ contains a_n^d disjoint $(n-1)$-stage boxes,and therefore, we get for all $M \geq m_3$, all $n \geq 1$ and all $K \geq 0$

$$\mathbb{P}_{\lambda,\alpha}\left[(A_{n,v}^{(b)})^{\mathrm{c}}\right] \leq \binom{a_n^d}{2} \left(2M_{n-1}^d\right)^{-\varphi} \leq a_n^{2d} M_{n-1}^{-d\varphi}.$$

This implies for all $M \geq m_3$, all $n \geq 1$ and all $K \geq 0$

$$\begin{aligned} 1 - p_n &= \mathbb{P}_{\lambda,\alpha}[B_{n,v} \text{ is not good}] \\ &= \mathbb{P}_{\lambda,\alpha}[A_{n,0}^c] \leq \frac{1 - p_{n-1}}{1 - \kappa_n} + a_n^{2d} M_{n-1}^{-d\varphi} \leq (1 - p_{n-1})(1 + 2\kappa_n) + a_n^{2d} M_{n-1}^{-d\varphi}. \end{aligned}$$

Consider

$$\frac{a_n^{2d} M_{n-1}^{-d\varphi}}{\varepsilon \kappa_n} = \frac{(n+1)^{2d\delta} M^{-d\varphi} (n!)^{-d\delta\varphi}}{\varepsilon (n+1)^{-\vartheta}} = \varepsilon^{-1} M^{-d\varphi} (n+1)^{2d\delta + \vartheta} (n!)^{-d\delta\varphi}.$$

Note that this is uniformly bounded from above in n. Therefore, there exists $m_4 \geq m_3$ such that for all $M \geq m_4$, all $n \geq 1$ and all $K \geq 0$

$$1 - p_n \leq (1 - p_{n-1})(1 + 2\kappa_n) + \varepsilon \kappa_n \leq (1 + 3\kappa_n) \max\{1 - p_{n-1}, \varepsilon\}.$$

Applying induction we obtain for all $M \geq m_4$, all $n \geq 1$ and all $K \geq 0$

$$1 - p_n \leq \max\{\varepsilon, 1 - p_0\} \prod_{i=1}^{n} (1 + 3\kappa_i).$$

Choose $m_5 \geq m_4$ such that for all $M \geq m_5$ there exists $K = K(M) \geq 0$ such that (4.14) and (4.15) hold. These choices imply that $p_0 \geq 1 - \varepsilon$. Therefore, for all $M \geq m_5$, $K(M)$ such that (4.15) holds, and all $n \geq 1$

$$1 - p_n \le \varepsilon \prod_{i=1}^{n} (1 + 3\kappa_i) \le \varepsilon c_1 < \varepsilon',$$

where $c_1 \in (1, \infty)$ was defined in (4.17). Thus, for all $M \ge m_5$, $K(M)$ such that (4.15) holds, and for all $n \ge 0$

$$\begin{aligned} &\mathbb{P}_{\lambda,\alpha}\left[\text{there are at least } M_n^d u_n \text{ sites in box } B_{n,0} \text{ connected within } B_{n,0}^{(K)}\right] \\ &\quad \ge \mathbb{P}_{\lambda,\alpha}[A_{n,0}] \ge 1 - \varepsilon', \end{aligned} \tag{4.22}$$

note that $B_{n,0} = [0, M_n - 1]^d$. Note that the explicit choices (4.16) provide

$$M_n = M((n+1)!)^{\delta} \quad \text{and} \quad u_n = \kappa_0((n+1)!)^{-\vartheta}.$$

The edge length of $B_{n,0}^{(K)}$ is given by $M_n + 2K = M_n + 2K(M) = M((n+1)!)^{\delta} + 2K(M)$. Therefore, for all $M \ge 1$ there exists $n_0 \ge 1$ such that for all $n \ge n_0$

$$M_n + 2K \le N = N(M, n) = 2M((n+1)!)^{\delta},$$

where the last identity is the definition of $N = N(M, n)$. For all $n \ge n_0$ the number of connected vertices in $B_{n,0}^{(K)}$ under (4.22) is at least

$$M_n^d u_n = \kappa_0 M^d ((n+1)!)^{-\vartheta + d\delta} = 2^{\vartheta/\delta - d} \kappa_0 M^{\vartheta/\delta} N^{d - \vartheta/\delta}.$$

Note that the choices of δ and ϑ are such that $d - \vartheta/\delta > \alpha/2 > 0$. Therefore, there exists $n_1 \ge n_0$ such that for all $n \ge n_1$ we have

$$M_n^d u_n \ge N^{\alpha/2}.$$

This implies for all $n \ge n_1$, see (4.22),

$$\begin{aligned} &\mathbb{P}_{\lambda,\alpha}\left[|C_N| \ge N^{\alpha/2}\right] \\ &\quad \ge \mathbb{P}_{\lambda,\alpha}\left[\text{there are at least } N^{\alpha/2} \text{ sites in box } B_{n,0} \text{ connected within } B_{n,0}^{(K)}\right] \\ &\quad \ge \mathbb{P}_{\lambda,\alpha}\left[\text{there are at least } M_n^d u_n \text{ sites in box } B_{n,0} \text{ connected within } B_{n,0}^{(K)}\right] \\ &\quad \ge 1 - \varepsilon', \end{aligned}$$

where $N = N(M, n) \ge N(M, n_1) = 2M((n_1+1)!)^{\delta}$. This proves the claim on the grid $N(M, n_1), N(M, n_1 + 1), \ldots$, with $N(M, n+1) = N(M, n)(n+2)^{\delta}$ for $n \ge n_1$. For $n' \in [N(M, n), N(M, n+1))$ we have on the set $\{|C_{N(M,n)}| \ge \rho_0 N(M, n)^{d - \vartheta/\delta}\}$ with $\rho_0 = 2^{\vartheta/\delta - d} \kappa_0 M^{\vartheta/\delta}$

$$
\begin{aligned}
|C_{n'}| \geq |C_{N(M,n)}| \;&\geq\; \rho_0 N(M,n)^{d-\vartheta/\delta} \\
&= \rho_0 N(M,n)^{d-\vartheta/\delta-\alpha/2} \left(\frac{N(M,n)}{n'}\right)^{\alpha/2} (n')^{\alpha/2} \\
&\geq \rho_0 N(M,n)^{d-\vartheta/\delta-\alpha/2} \left(\frac{N(M,n)}{N(M,n+1)}\right)^{\alpha/2} (n')^{\alpha/2} \\
&= \rho_0 \, \frac{N(M,n)^{d-\vartheta/\delta-\alpha/2}}{(n+2)^{\delta\alpha/2}} \, (n')^{\alpha/2} \;\geq\; (n')^{\alpha/2},
\end{aligned}
$$

for all n sufficiently large. This finishes the proof of Theorem 15. □

Conclusion. Theorem 15 defines good boxes $[0, N-1]^d$ on a new scale, i.e. these are boxes that contain sufficiently large connected components C_N. The latter occurs with probability $1-\varepsilon' \geq r^*$, for small ε'. If we can prove that such large connected components in disjoint boxes are connected by an occupied edge with probability bounded below by (4.11), then we are in the set-up of a site-bond percolation model. This is exactly what is used in Theorem 3.2 of Biskup [9] in order to prove that (i) large connected components are percolating, a.s.; and (ii) $|C_N|$ is even of order ρN^d for an appropriate positive constant $\rho > 0$, which improves Theorem 15.

References

1. Aizenman, M., Kesten, H., Newman, C.M.: Uniqueness of the infinite cluster and continuity of connectivity functions for short- and long-range percolation. Commun. Math. Phys. **111**, 505–532 (1987)
2. Aldous, D.: Brownian excursions, critical random graphs and the multiplicative coalescent. Ann. Probab. **25**(2), 812–854 (1997)
3. Amini, H., Cont, R., Minca, A.: Stress testing the resilience of financial networks. Int. J. Theor. Appl. Finance **15**(1), 1250,006–1250,020 (2012)
4. Antal, P., Pisztora, A.: On the chemical distance for supercritical Bernoulli percolation. Ann. Probab. **24**(2), 1036–1048 (1996)
5. Benjamini, I., Kesten, H., Peres, Y., Schramm, O.: Geometry of the uniform spanning forest: transition in dimensions 4,8,12. Ann. Math. **160**, 465–491 (2004)
6. Berger, N.: Transience, recurrence and critical behavior for long-range percolation. Commun. Math. Phys. **226**(3), 531–558 (2002)
7. Berger, N.: A lower bound for the chemical distance in sparse long-range percolation models. arXiv:math/0409021v1 (2008)
8. Berger, N.: Transience, recurrence and critical behavior for long-range percolation. arXiv:math/0110296v3 (2014)
9. Biskup, M.: On the scaling of the chemical distance in long-range percolation models. Ann. Probab. **32**, 2938–2977 (2004)
10. Bollobás, B.: Random Graphs, 2nd edn. Cambridge University Press, Cambridge (2001)
11. Broadbent, S.R., Hammersley, J.M.: Percolation processes I. Crystals and mazes. Math. Proc. Cambridge Philos. Soc. **53**, 629–641 (1957)
12. Burton, R.M., Keane, M.: Density and uniqueness in percolation. Commun. Math. Phys. **121**, 501–505 (1989)

13. Chung, F., Lu, L.: The average distances in random graphs with given expected degrees. Proc. Natl. Acad. Sci. **99**, 15879–15882 (2002)
14. Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequences. Ann. Comb. **6**(2), 125–145 (2002)
15. Cont, R., Moussa, A., Santos, E.B.: Network structure and systemic risk in banking system. SSRN Server, Manuscript ID 1733528 (2010)
16. Deijfen, M., van der Hofstad, R., Hooghiemstra, G.: Scale-free percolation. Ann. Inst. Henri Poincaré Probab. Stat. **49**(3), 817–838 (2013)
17. Deprez, P., Hazra, R.S., Wüthrich, M.V.: Inhomogeneous long-range percolation for real-life network modeling. Risks **3**(1), 1–23 (2015)
18. Deprez, P., Wüthrich, M.V.: Poisson heterogeneous random-connection model. arXiv:1312.1948 (2013)
19. Durrett, R.: Random Graph Dynamics. Cambridge University Press, Cambridge (2007)
20. Erdős, P., Rényi, A.: On random graphs I. Publ. Math. Debrecen **6**, 290–297 (1959)
21. Franceschetti, M., Meester, R.: Random Networks for Communication. Cambridge University Press, Cambridge (2007)
22. Gandolfi, A., Grimmett, G.R., Russo, L.: On the uniqueness of the infinite open cluster in the percolation model. Commun. Math. Phys. **114**, 549–552 (1988)
23. Gandolfi, A., Keane, M.S., Newman, C.M.: Uniqueness of the infinite component in a random graph with applications to percolation and spin glasses. Probab. Theory Related Fields **92**, 511–527 (1992)
24. Grimmett, G.R.: Percolation and disordered systems. In: Bernard, P. (ed.) Lectures on Probability and Statistics, Lecture Notes in Mathematics, vol. 1665, pp. 153–300. Springer, New York (1997)
25. Grimmett, G.R.: Percolation, 2nd edn. Springer, New York (1999)
26. van der Hofstad, R.: Random graphs and complex networks. http://www.win.tue.nl/~rhofstad/NotesRGCN2013.pdf (2013). Accessed 22 Apr 2015
27. van der Hofstad, R., Hooghiemstra, G., Znamenksi, D.: Distances in random graphs with finite mean and infinite variance degrees. Electron. J. Probab. **12**, 703–766 (2007)
28. Hurd, T.R., Gleeson, J.P.: A framework for analyzing contagion in banking networks. Preprint (2012)
29. Kesten, H.: The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. Commun. Math. Phys. **74**, 41–59 (1980)
30. Kesten, H.: Percolation Theory for Mathematicians. Birkhäuser, Boston (1982)
31. Meester, R., Roy, R.: Continuum Percolation. Cambridge University Press, Cambridge (1996)
32. Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. Random Struct. Algorithms **6**, 161–180 (1995)
33. Newman, C.M., Schulman, L.S.: One dimensional $1/|j-i|^s$ percolation models: the existence of a transition for $s \leq 2$. Commun. Math. Phys. **104**, 547–571 (1986)
34. Newman, M.E.J., Strogatz, S.H., Watts, D.J.: Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E. **64**(2), 026,118 (2001)
35. Newman, M.E.J., Watts, D.J., Strogatz, S.H.: Random graph models of social networks. Proc. Natl. Acad. Sci. **99**, 2566–2572 (2002)
36. Olhede, S.C., Wolfe, P.J.: Degree-based network models. arXiv:1211.6537v2 (2013)
37. Schulman, L.S.: Long-range percolation in one dimension. J. Phys. A **16**(17), L639–L641 (1983)
38. Soramäki, K., Bech, M., Arnold, J., Glass, R., Beyeler, W.: The topology of interbank payment flows. Phys. A **379**(1), 317–333 (2007)
39. Trapman, P.: The growth of the infinite long-range percolation cluster. Ann. Probab. **38**(4), 1583–1608 (2010)
40. Watts, D.J.: Six Degrees: The Science of a Connected Age. W. W. Norton, New York (2003)
41. Wüthrich, M.V.: Non-life insurance: Mathematics and statistics. SSRN Server, Manuscript ID 2319328 (2013)