Springer Proceedings in Mathematics & Statistics

Gerold Alsmeyer
Matthias Löwe *Editors*

Random Matrices and Iterated Random Functions

Springer

Springer Proceedings in Mathematics and Statistics

Volume 53

Springer Proceedings in Mathematics and Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including OR and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

For further volumes:
http://www.springer.com/series/10533

Gerold Alsmeyer • Matthias Löwe
Editors

Random Matrices and Iterated Random Functions

Münster, October 2011

Springer

Editors
Gerold Alsmeyer
Institute of Mathematics and Informatics
Westphalian Wilhelms University Münster
Münster, Germany

Matthias Löwe
Institute for Mathematical Statistics
Westphalian Wilhelms University Münster
Münster, Germany

ISSN 2194-1009 ISSN 2194-1017 (electronic)
ISBN 978-3-642-38805-7 ISBN 978-3-642-38806-4 (eBook)
DOI 10.1007/978-3-642-38806-4
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013947798

Mathematics Subject Classification (2010): 60B20, 46L54, 60B15, 60F05, 60J05, 60J80, 60K05

Printed on acid-free paper

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

Preface

Random matrices are of central importance in many areas of probability theory, and the analysis of their theoretical aspects has been the object of very active research over the past 20 years and will certainly continue in the future because of many still open questions. Applications can be found in, for example, mathematical statistics, mathematical physics, random graphs, or telecommunications.

The present collection of research papers focuses on two of their most relevant aspects:

- The spectra of high-dimensional random matrices

and

- Iterated random functions driven by random matrices

The contributions to this volume are based on talks given at the workshop "Random matrices and iterated random functions" organized as part of the scientific program of the Collaborative Research Center 878 from October 4 to October 7, 2011, at the University of Münster.

While the contributions to Random Matrix Theory are centered around questions about universality of the limiting behavior of random matrices and their relation to free probability, the (larger) section on iterated functions systems focuses on questions concerning their long-time behavior (ergodicity) and information about their stationary distributions (tail behavior).

Discussions among the participants of the workshop were also concerned with possible connections between these two fields of probability. Such connections would doubtlessly stimulate the research in both areas.

We thank all authors who contributed to this volume and hope that we have been able to gather an inspiring collection of papers. We are grateful to the CRC 878 for financial support of the workshop and to Olga Friesen and Sebastian Mentemeier for helping us with the production of this book.

Münster, Germany

Gerold Alsmeyer
Matthias Löwe

Contents

Part I
Random Matrices

On the Limiting Spectral Density of Symmetric Random Matrices with Correlated Entries

Olga Friesen* **and Matthias Löwe**

Abstract We analyze the spectral distribution of two different models of symmetric random matrices with correlated entries. While we assume that the diagonals of these random matrices are stochastically independent, the elements of the diagonals are taken to be correlated. Depending on the strength of correlation the limiting spectral distribution is either the famous semicircle law known for the limiting spectral density of symmetric random matrices with independent entries, or some other law related to that derived for Toeplitz matrices by Bryc W, Dembo A, Jiang T (2006) Spectral measure of large random Hankel, Markov and Toeplitz matrices. Ann Probab 34(1):1–38.

1 Introduction

The study of random matrices started in the 1920s with the seminal work of Wishart [16]. His basic motivation was the analysis of data. On the other hand, Wigner used the eigenvalues of random matrices to model the spectra of heavy-nuclei atoms [15]. Nowadays, random matrix theory is a field with many applications from telecommunications to random graphs and with many interesting and surprising results.

A central role in the study of random matrices with growing dimension is played by their eigenvalues. To introduce them let, for any $n \in \mathbb{N}$,

*Research supported by Deutsche Forschungsgemeinschaft via SFB 878 at University of Münster.

O. Friesen (✉) · M. Löwe
Westfälische Wilhelms-Universität Münster, Fachbereich Mathematik, Einsteinstraße 62, 48149 Münster, Germany
e-mail: olga.friesen@uni-muenster.de; maloewe@math.uni-muenster.de

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_1, © Springer-Verlag Berlin Heidelberg 2013

$\{a_n(p,q), 1 \leq p \leq q \leq n\}$ be a real-valued random field. Define the symmetric random $n \times n$ matrix $\mathbf{X}_n$ by

$$\mathbf{X}_n(q,p) = \mathbf{X}_n(p,q) = \frac{1}{\sqrt{n}} a_n(p,q), \qquad 1 \leq p \leq q \leq n.$$

We will denote the (real) eigenvalues of $\mathbf{X}_n$ by $\lambda_1^{(n)} \leq \lambda_2^{(n)} \leq \ldots \lambda_n^{(n)}$. Let μ_n be the empirical eigenvalue distribution, i.e.

$$\mu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k^{(n)}}.$$

Wigner proved in his fundamental work [15] that, if $a_n(p,q)$ are independent, normally distributed with mean 0 and variance 1, for off-diagonal elements, and variance 2 on the diagonal, the empirical eigenvalue distribution μ_n converges weakly (in probability) to the so-called semicircle distribution (or law), i.e. the probability distribution ν on $\mathbb{R}$ with density

$$\frac{1}{2\pi} \sqrt{4 - x^2}\, \mathbb{1}_{[-2,2]}(x)\, dx.$$

An important step to show the universality of this result was taken by Arnold [1], who verified that the convergence to the semicircle law also is true, if one replaces the Gaussian distributed random variables by independent and identically distributed (i.i.d.) random variables with a finite fourth moment. Also the identical distribution may be replaced by some other assumptions (see e.g. [8]). There are various ways to prove such a result. Among others, large deviations techniques as developed in [4] can be applied as well as Stieltjes transforms [2] (the latter method can also be applied to obtain results on the speed of convergence, see [12]). A still very powerful instrument is the moment method, originally employed by Wigner. To this end, it is useful to notice that, if Y is some random variable distributed according to this semicircle distribution, its moments are given by

$$\mathbb{E}[Y^k] = \begin{cases} 0, & \text{if } k \text{ is odd,} \\ C_{\frac{k}{2}}, & \text{if } k \text{ is even,} \end{cases}$$

where C_k, $k \in \mathbb{N}$, are the Catalan numbers defined by $C_k = (2k)!/(k!(k+1)!)$.

Recently, it was observed by Erdös et al. [9] that the convergence of the spectral measure towards the semicircle law holds in a local sense. More precisely, this can be proved on intervals with width going to zero sufficiently slowly.

However, the assumption of the entries being independent cannot be renounced without any replacement. Bryc, Dembo and Jiang [5] studied random symmetric Toeplitz matrices and obtained a different limiting distribution. To be more precise, they considered a family $\{X_i, 0 \leq i \leq n-1\}$, $n \in \mathbb{N}$, of independent and identically distributed real-valued random variables, and assumed that $\mathrm{Var}(X_1) = 1$. Then, the scaled symmetric Toeplitz matrix $\mathbf{T}_n$ was defined by $\mathbf{T}_n(i, j) = 1/\sqrt{n}\ X_{|i-j|}$, $1 \le i, j \le n$, i.e.

$$\mathbf{T}_n = \frac{1}{\sqrt{n}} \begin{pmatrix} X_0 & X_1 & X_2 & \dots & X_{n-2} & X_{n-1} \\ X_1 & X_0 & X_1 & \dots & X_{n-3} & X_{n-2} \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ X_{n-2} & X_{n-3} & X_{n-4} & \dots & X_0 & X_1 \\ X_{n-1} & X_{n-2} & X_{n-3} & \dots & X_1 & X_0 \end{pmatrix}. \tag{1}$$

In this situation the empirical spectral distribution of $\mathbf{T}_n$ converges weakly almost surely to some non-random probability measure γ_T as $n \to \infty$. This measure does not depend on the distribution of X_1. Moreover, it has existing moments of all orders, is symmetric, and has an unbounded support. The aim of the present note is, to investigate the borderline between convergence to the semicircle law and the convergence to γ_T. To this end we will study random matrices with independent diagonals, where the elements on the diagonals may be correlated. If they are independent, we are, of course, back in the Wigner case, while for complete correlation the matrix is a random Toeplitz matrix.

We will see that depending on the strength of the correlation, the empirical spectral distribution either converges to the semicircle law, or to some mixture of γ_T and the semicircle distribution. We hence have a sort of phase transition. Similar results were obtained in [10] for the case of weak correlations and in [11] for stronger correlations. A particularly nice example is borrowed from statistical mechanics. There the Curie-Weiss model is the easiest model of a ferromagnet. Here a magnetic substance has little atoms that carry a magnetic spin, that is either $+1$ or -1. These spins interact in cooperative way, the strength of the interaction being triggered by a parameter, the so-called inverse temperature. The model exhibits phase transition from paramagnetic to magnetic behavior (the standard reference for the Curie-Weiss model is [7]). We will see that this phase transition can be recovered on the level of the limiting spectral distribution of random matrices, if we fill their diagonals independently with the spins of Curie-Weiss models. For small interaction parameter, this limiting spectral distribution is the semicircle law, while for a large interaction parameter we obtain a distribution which shows the influence of γ_T.

The rest of this article is organized in the following way. In the two following sections we will fix our notation. Section 2 contains a description of the measures γ_T introduced above, while Sect. 3 describes the kind of matrices we will deal with in a general framework. From here we follow two paths. Section 4 contains our results for convergence towards the semicircle law, while Sect. 5 is devoted to the case of strong correlations along the diagonals. The basic ideas of the proofs, however are so similar, that we can treat them in a unified way. This is done in Sect. 6.

2 The Measure γ_T

The limiting spectral distribution γ_T can be defined by its moments which are described with the help of *Toeplitz volumes*.

Compared to [5], we will use a slightly different notation. This will make it easier to understand the arguments of the following sections. Thus, denote by $\mathcal{PP}(k)$, $k \in \mathbb{N}$, the set of all pair partitions of $\{1, \ldots, k\}$. For any $\pi \in \mathcal{PP}(k)$, we write $i \sim_\pi j$ if i and j are in the same block of π. To introduce Toeplitz volumes, we associate to any $\pi \in \mathcal{PP}(k)$ the following system of equations in unknowns $x_0, \ldots, x_k$:

$$\begin{aligned}
x_1 - x_0 + x_{l_1} - x_{l_1-1} &= 0, \quad \text{if } 1 \sim_\pi l_1, \\
x_2 - x_1 + x_{l_2} - x_{l_2-1} &= 0, \quad \text{if } 2 \sim_\pi l_2, \\
&\vdots \\
x_i - x_{i-1} + x_{l_i} - x_{l_i-1} &= 0, \quad \text{if } i \sim_\pi l_i, \\
&\vdots \\
x_k - x_{k-1} + x_{l_k} - x_{l_k-1} &= 0, \quad \text{if } k \sim_\pi l_k.
\end{aligned} \tag{2}$$

Since π is a pair partition, we in fact have only $k/2$ equations although we have listed k. However, we have $k+1$ variables. If $\pi = \{\{i_1, j_1\}, \ldots, \{i_{k/2}, j_{k/2}\}\}$ with $i_l < j_l$ for any $l = 1, \ldots, k/2$, we solve (2) for $x_{j_1}, \ldots, x_{j_{k/2}}$, and leave the remaining variables undetermined. We further impose the condition that all variables $x_0, \ldots, x_k$ lie in the interval $I = [0, 1]$. Solving the equations above in this way determines a cross section of the cube $I^{k/2+1}$. The volume of this will be denoted by $p_T(\pi)$. To give an example, consider the partition $\pi = \{\{1, 3\}, \{2, 4\}\}$. Solving (2) for $x_3 = x_0 - x_1 + x_2$ and $x_4 = x_1 - x_2 + x_3 = x_0$, we obtain a cross section of I^3 given by

$$\{x_0 - x_1 + x_2 \in I\} \cap \{x_0 \in I\}.$$

This set has the volume $p_T(\pi) = 2/3$.

Returning to the measure γ_T, it is shown in [5] that all odd moments are zero, and for any even $k \in \mathbb{N}$, the k-th moment is given by

$$\int x^k d\gamma_T(x) = \sum_{\pi \in \mathcal{PP}(k)} p_T(\pi).$$

Since $|p_T(\pi)| \leq 1$ for any $\pi \in \mathcal{PP}(k)$ and $\#\mathcal{PP}(k) = (k-1)!!$, we have

$$\left| \int x^k d\gamma_T(x) \right| \leq (k-1)!!.$$

In particular, Carleman's condition holds implying that γ_T is uniquely determined by its moments. The results for the independent as well as the Toeplitz case will follow directly from Theorems 1 and 2 in case we assume the uniform boundedness of the moments of all orders.

3 Matrices with Independent Processes on the Diagonals

We want to study two different models of symmetric matrices with dependent entries. Both models have the common property that entries from different diagonals are independent while on each diagonal we have a stochastic process with a given covariance structure. Therefore, consider for any $n \in \mathbb{N}$ a family $\{a_n(p,q), 1 \leq p \leq q \leq n\}$ of real-valued random variables. Introduce the symmetric random $n \times n$ matrix $\mathbf{X}_n$ with

$$\mathbf{X}_n(p,q) = \mathbf{X}_n(q,p) = \frac{1}{\sqrt{n}} a_n(p,q), \quad 1 \leq p \leq q \leq n.$$

Put $a_n(p,q) = a_n(q,p)$ if $1 \leq q < p \leq n$. Since we will resort to the method of moments, we first of all want to assume that

(A1) $\mathbb{E}\left[a_n(p,q)\right] = 0, \mathbb{E}\left[a_n(p,q)^2\right] = 1$, and

$$m_k := \sup_{n \in \mathbb{N}} \max_{1 \leq p \leq q \leq n} \mathbb{E}\left[|a_n(p,q)|^k\right] < \infty, \quad k \in \mathbb{N}. \tag{3}$$

Note that the assumption of centered entries can be made without loss of generality if the family $\{a_n(p,q), 1 \leq p \leq q \leq n\}$ consists of identically distributed random variables. Indeed, assuming $\mathbb{E}\left[a_n(p,q)\right] = b_n$ for any $1 \leq p \leq q \leq n$, $n \in \mathbb{N}$, and some sequence $(b_n)_{n \in \mathbb{N}}$ such that $b_n = o(n)$ yields the same limiting spectral distribution as in the centered case, if it exists. This follows from the rank inequality for Hermitian matrices (cf. [3], Lemma 2.2). Changing the variance, however, provides a different limit which is a scaled version of that we obtain with assumption (A1). To make the condition of independent diagonals more precise, we suppose that

(A2) For any $n \in \mathbb{N}$, $j \in \{1, \ldots, n\}$, and distinct integers $r_1, \ldots, r_j \in \{0, \ldots, n-1\}$, the families $\{a_n(p, p + r_1), 1 \leq p \leq n - r_1\}, \ldots, \{a_n(p, p + r_j), 1 \leq p \leq n - r_j\}$ are independent.

So far, we know that if we also have independence among the entries on the same diagonal, we will obtain the semicircle law as the limiting spectral distribution. Although we will violate this assumption in our first model, the quickly decaying dependency structure will ensure that nevertheless, we get the same limiting distribution. In our second model, we will basically assume that the covariance is the same for any two entries on the same diagonal. If it is equal to the variance, it is not surprising that the resulting limit is the same as in the Toeplitz case. In general, we will find that we have a combination of the Toeplitz distribution and the semicircle law.

4 Quickly Decaying Covariances

In our first model, the dependency structure within the diagonals is determined by the conditions

(A3) The covariance of two entries on the same diagonal can be bounded by some constant depending only on their distance, i.e. for any $n \in \mathbb{N}$ and $0 \leq \tau \leq n-1$, there is a constant $c_n(\tau) \geq 0$ such that

$$|\mathrm{Cov}(a_n(p,q), a_n(p+\tau, q+\tau))| \leq c_n(\tau), \quad 1 \leq p \leq q \leq n-\tau,$$

(A4) The entries on the diagonals have a quickly decaying dependency structure, which will be expressed in terms of the condition

$$\sum_{\tau=0}^{n-1} c_n(\tau) = o(n).$$

Theorem 1. *Assume that the symmetric random matrix $\mathbf{X}_n$ satisfies the conditions (A1)–(A4). Then, with probability* 1*, the empirical spectral distribution of $\mathbf{X}_n$ converges weakly to the standard semicircle distribution.*

Remark 1. Note that in order for the semicircle law to hold, it is not possible to renounce condition (A4) without any replacement. To understand this, consider a Toeplitz matrix. We clearly have $c_n(\tau) = \mathcal{O}(1)$, and indeed, the empirical distribution of a sequence of Toeplitz matrices tends with probability 1 to a nonrandom probability measure with unbounded support.

4.1 Examples

We want to give some examples of processes that satisfy the assumptions of Theorem 1. Obviously, this is the case if the entries $\{a_n(p,q), 1 \leq p \leq q \leq n\}$ are independent satisfying (A1). The following three examples deal with finite Markov chains, Gaussian Markov processes and m-dependent processes.

(i) Assume that $\{x(p), p \in \mathbb{N}\}$ is a stationary Markov chain on a finite state space $S = \{s_1, \ldots, s_N\}$, $N \geq 2$. Denote by $\varrho = (\varrho_1, \ldots, \varrho_N)$ the stationary distribution, and suppose that

$$\mathbb{E}[x(p)] = \sum_{j=1}^{N} s_j \varrho(j) = 0, \quad \mathbb{E}[x(p)^2] = \sum_{j=1}^{N} s_j^2 \varrho(j) = 1.$$

If $\{x(p), p \in \mathbb{N}\}$ is aperiodic and irreducible, we have that for some constant $C > 0$ and some $\alpha \in (0, 1)$,

$$\max_{i,j \in \{1,\ldots,N\}} |\mathbb{P}(x(p) = s_i \mid x(1) = s_j) - \varrho(i)| \leq C\alpha^{p-1}, \quad p \in \mathbb{N}.$$

For more details, see [13], Theorem 4.9. In particular, we obtain

$$|\text{Cov}(x(p), x(1))| = \left| \sum_{i,j=1}^{N} s_i s_j \Big(\mathbb{P}(x(p) = s_i \mid x(1) = s_j) - \varrho(i)\Big)\varrho(j) \right| \leq C\alpha^{p-1}.$$

Now assume that the processes $\{a(p, p + r), p \in \mathbb{N}\}$, $r \in \mathbb{N}_0$, are independent copies of $\{x(p), p \in \mathbb{N}\}$, and put $a_n(p, q) := a(p, q)$ for any $n \in \mathbb{N}$, $1 \leq p \leq q \leq n$. Condition (A2) then holds by definition, and the uniform moment bound in (A1) is given since we have a bounded support. Furthermore,

$$|\text{Cov}(a_n(p, q), a_n(p + \tau, q + \tau))| \leq c_n(\tau),$$

where $c_n(\tau) = c(\tau) = C\alpha^\tau$. This is assumption (A3). Finally, (A4) follows since $\sum_{\tau=0}^{\infty} c(\tau) < \infty$, implying $\sum_{\tau=0}^{n-1} c(\tau) = o(n)$.

(ii) Let $\{y(p), p \in \mathbb{N}\}$ be a stationary Gaussian Markov process with mean 0 and variance 1. In addition to this, assume that the process is non-degenerate in the sense that $\mathbb{E}[y(p)|y(q), q \leq p - 1] \neq y(p)$. In this case, we can represent $y(p)$ as

$$y(p) = b_p \sum_{j=1}^{p} d_j \xi_j,$$

where $\{\xi_j, j \in \mathbb{N}\}$ is a family of independent standard Gaussian variables and $b_p, d_1, \ldots, d_p \in \mathbb{R}\backslash\{0\}$. We can now calculate

$$\bar{c}(\tau) := \text{Cov}(y(p + \tau), y(p)) = b_{p+\tau} b_p \sum_{i=1}^{p+\tau} \sum_{j=1}^{p} d_i d_j \mathbb{E}[\xi_i \xi_j] = b_{p+\tau} b_p \sum_{j=1}^{p} d_j^2.$$

Note that $1 = \mathbb{E}[y(p)^2] = b_p^2 \sum_{j=1}^{p} d_j^2$. As a consequence, we have

$$\bar{c}(\tau) = \frac{b_{p+\tau}}{b_p} = \frac{b_{p+\tau}}{b_{p+\tau-1}} \frac{b_{p+\tau-1}}{b_{p+\tau-2}} \cdots \frac{b_{p+1}}{b_p} = \bar{c}(1)^\tau.$$

To see that $|\bar{c}(\tau)| < 1$, we first compute $\bar{c}(1)y(1) = b_2 b_1^2 d_1^3 \xi_1 = b_2 d_1 \xi_1$, implying that $y(2) = b_2 d_2 \xi_2 + \bar{c}(1)y(1)$. Using this identity to calculate the variance, we can take account of the independence of ξ_2 and $y(1)$ to obtain

$$1 = \mathbb{E}[y(2)^2] = b_2^2 d_2^2 + \bar{c}(1)^2.$$

Since $b_2, d_2 \neq 0$, we conclude that $|\bar{c}(1)| < 1$. In analogy to the first example, we assume that the processes $\{a(p, p+r), p \in \mathbb{N}\}$, $r \in \mathbb{N}_0$, are independent copies of $\{y(p), p \in \mathbb{N}\}$, and put $a_n(p,q) := a(p,q)$ for any $n \in \mathbb{N}$, $1 \leq p \leq q \leq n$. Conditions (A1) and (A2) obviously hold. Defining $c(\tau) := |\bar{c}(\tau)|$, we further obtain

$$|\mathrm{Cov}(a_n(p,q), a_n(p+\tau, q+\tau))| \leq c(\tau).$$

Since $|\bar{c}(1)| < 1$, we have $\sum_{\tau=0}^{\infty} c(\tau) < \infty$, implying $\sum_{\tau=0}^{n-1} c(\tau) = o(n)$. Thus assumptions (A3) and (A4) are satisfied.

(iii) Assume that $\{z(p), p \in \mathbb{N}\}$ is a stationary process of m-dependent random variables, i.e. $z(p)$ and $z(q)$ are stochastically independent whenever $|p-q| > m$. Moreover, suppose that $z(1)$ is centered with unit variance, and has existing moments of all orders. Define

$$c(\tau) := |\mathrm{Cov}(z(1), z(\tau+1))|, \quad \tau \in \mathbb{N}_0.$$

Then, $c(\tau) = 0$ for any $\tau > m$. Thus, $\sum_{\tau=0}^{n-1} c(\tau) = \sum_{\tau=0}^{m} c(\tau) = o(n)$ for any $n \geq m+1$. Let $\{a(p, p+r), p \in \mathbb{N}\}$, $r \in \mathbb{N}_0$, be independent copies of $\{z(p), p \in \mathbb{N}\}$, and $a_n(q,p) := a(p,q)$ for any $n \in \mathbb{N}$, $1 \leq p \leq q \leq n$. Then, (A1)–(A4) are satisfied.

5 Constant Covariances

For our second model, we assume that

(A3′) The covariance of two distinct entries on the same diagonal depends only on n, i.e. for any $1 \leq \tau \leq n-1$ and $1 \leq p, q \leq n-\tau$, we can define

$$c_n := \mathrm{Cov}(a_n(p,q), a_n(p+\tau, q+\tau)),$$

(A4′) The limit $c := \lim_{n\to\infty} c_n$ exists.

To describe the limiting spectral distribution in this case, we want to resort to pair partitions. However, we need a further notion which proved to be useful in [5] when considering the limiting spectral distribution of Markov matrices.

Definition 1. Let $k \in \mathbb{N}$ be even, and fix $\pi \in \mathcal{PP}(k)$. The *height* $h(\pi)$ of π is the number of elements $i \sim_\pi j$, $i < j$, such that either $j = i + 1$ or the restriction of π to $\{i + 1, \dots, j - 1\}$ is a pair partition.

Note that the property that the restriction of π to $\{i + 1, \dots, j - 1\}$ is a pair partition in particular requires that the distance $j - i - 1 \geq 1$ is even. To give an example how to calculate the height of a partition, take $\pi = \{\{1, 6\}, \{2, 4\}, \{3, 5\}\}$. Considering the block $\{1, 6\}$, we see that the restriction of π to $\{2, 3, 4, 5\}$ is a pair partition, namely $\{\{2, 4\}, \{3, 5\}\}$. However, this is not true for both remaining blocks. Hence, $h(\pi) = 1$.

In the following, we will say that a pair partition $\pi \in \mathcal{PP}(k)$ is *crossing* if there are $i < j < l < m$ such that $i \sim_\pi l$ and $j \sim_\pi m$. Otherwise, we call the pair partition *non-crossing*. The set of all crossing pair partitions of $\{1, \dots, k\}$ is denoted by $\mathcal{CPP}(k)$, and the set of all non-crossing pair partitions by $\mathcal{NPP}(k)$. Recall that for even $k \in \mathbb{N}$, the Catalan number $C_{k/2}$ is given by $C_{k/2} = \#\mathcal{NPP}(k)$.

We can now state the main result of this section.

Theorem 2. *Assume that the symmetric random matrix* $\mathbf{X}_n$ *satisfies the conditions (A1), (A2), (A3′), and (A4′). Then, with probability* 1*, the empirical spectral distribution of* $\mathbf{X}_n$ *converges weakly to a deterministic probability distribution* ν_c *with* k*-th moment*

$$\int x^k d\nu_c(x) = \begin{cases} C_{\frac{k}{2}} + \sum\limits_{\pi \in \mathcal{CPP}(k)} p_T(\pi) c^{\frac{k}{2} - h(\pi)}, & \textit{if } k \textit{ is even}, \\ 0, & \textit{if } k \textit{ is odd}. \end{cases}$$

If k *is even, we can also write* $\int x^k d\nu_c(x) = \sum_{\pi \in \mathcal{PP}(k)} p_T(\pi) c^{k/2 - h(\pi)}$.

Remark 2. As for the limiting distribution in the Toeplitz case, we can verify the Carleman condition to see that ν_c is uniquely determined by its moments.

Remark 3. If $c = 0$, Theorem 2 states that the limiting distribution is the semicircle law since $h(\pi) < k/2$ for any crossing partition $\pi \in \mathcal{CPP}(k)$. This result can also be deduced from Theorem 1. Indeed, choose $c_n(\tau) = |c_n|$ for any $\tau \geq 1$, and $c_n(0) = 1$. We then have for any $1 \leq p \leq q \leq n - \tau$,

$$|\mathrm{Cov}(a_n(p, q), a_n(p + \tau, q + \tau))| = c_n(\tau).$$

Furthermore, we obtain $\sum_{\tau=0}^{n-1} c_n(\tau) = 1 + (n - 1)|c_n| = o(n)$ since $\lim_{n\to\infty} c_n = c = 0$. Consequently, (A3) and (A4) are satisfied.

5.1 Examples

We want to give some examples of processes satisfying the assumptions of Theorem 2.

(i) Consider a symmetric Toeplitz matrix as in (1). The limiting spectral distribution can be deduced from Theorem 2 as well. Indeed, assuming that the entries are centered with unit variance and have existing moments of any order, we see that all conditions are satisfied with $c = c_n = 1$. Thus, we get

$$\int x^k \, d\nu_1(x) = \begin{cases} C_{\frac{k}{2}} + \sum_{\pi \in \mathcal{CPP}(k)} p_T(\pi) = \sum_{\pi \in \mathcal{PP}(k)} p_T(\pi), & \text{if } k \text{ is even,} \\ 0, & \text{if } k \text{ is odd,} \end{cases}$$

as proven in [5].

(ii) Suppose that for any $n \in \mathbb{N}$, $\{x_n(p), 1 \le p \le n\}$ is a family of exchangeable random variables, i.e. the distribution of the vector $(x_n(1), \dots, x_n(n))$ is the same as that of $(x_n(\sigma(1)), \dots, x_n(\sigma(n)))$ for any permutation σ of $\{1, \dots, n\}$. In this case, we can conclude that for any $1 \le p < q \le n$, we have

$$\mathrm{Cov}(x_n(p), x_n(q)) = \mathrm{Cov}(x_n(1), x_n(2)) =: c_n.$$

Now assume that $c_n \to c \in \mathbb{R}$ as $n \to \infty$. Define for any $n \in \mathbb{N}$, $r \in \{0, \dots, n-1\}$, the process $\{a_n(p, p+r), 1 \le p \le n-r\}$ to be an independent copy of $\{x_n(p), 1 \le p \le n-r\}$. Then, all conditions of Theorem 2 are satisfied if we ensure that the moment condition (A1) holds. The resulting limiting distribution for different choices of c is depicted in Fig. 1.

An example for a process with exchangeable variables is the Curie-Weiss model with inverse temperature $\beta > 0$. Here, the vector $x_n = (x_n(1), \dots, x_n(n))$ takes values in $\{-1, 1\}^n$, and for any $\omega = (\omega(1), \dots, \omega(n)) \in \{-1, 1\}^n$, we have

$$\mathbb{P}(x_n = \omega) = \frac{1}{Z_{n,\beta}} \exp\left(\frac{\beta}{2n} \left(\sum_{i=1}^n \omega(i) \right)^2 \right),$$

where $Z_{n,\beta}$ is the normalizing constant. Since $\mathbb{P}(x_n(1) = -1) = \mathbb{P}(x_n(1) = 1) = 1/2$, we obtain $\mathbb{E}[x_n(1)] = 0$. Further, we clearly have $\mathbb{E}[x_n(1)^2] = 1$. It remains to determine $c = \lim_{n\to\infty} c_n$. Therefore, we want to make use of the identity

$$c_n = \mathrm{Cov}(x_n(1), x_n(2)) = \mathbb{E}[x_n(1)x_n(2)] = \frac{n}{n-1}\mathbb{E}[m_n^2] - \frac{1}{n-1},$$

where $m_n := 1/n \sum_{i=1}^n x_n(i)$ is the so-called magnetization of the system. Since $|m_n| \le 1$, we see that m_n is uniformly integrable. Thus, m_n converges in $\mathscr{L}^2$ to some random variable m if and only if $m_n \to m$ in probability. In [6], it was verified that $m_n \to 0$ in probability if $\beta \le 1$, and $m_n \to m$ with $m \sim 1/2\, \delta_{m(\beta)} + 1/2\, \delta_{-m(\beta)}$ for some $m(\beta) > 0$ if $\beta > 1$. The function $m(\beta)$ is

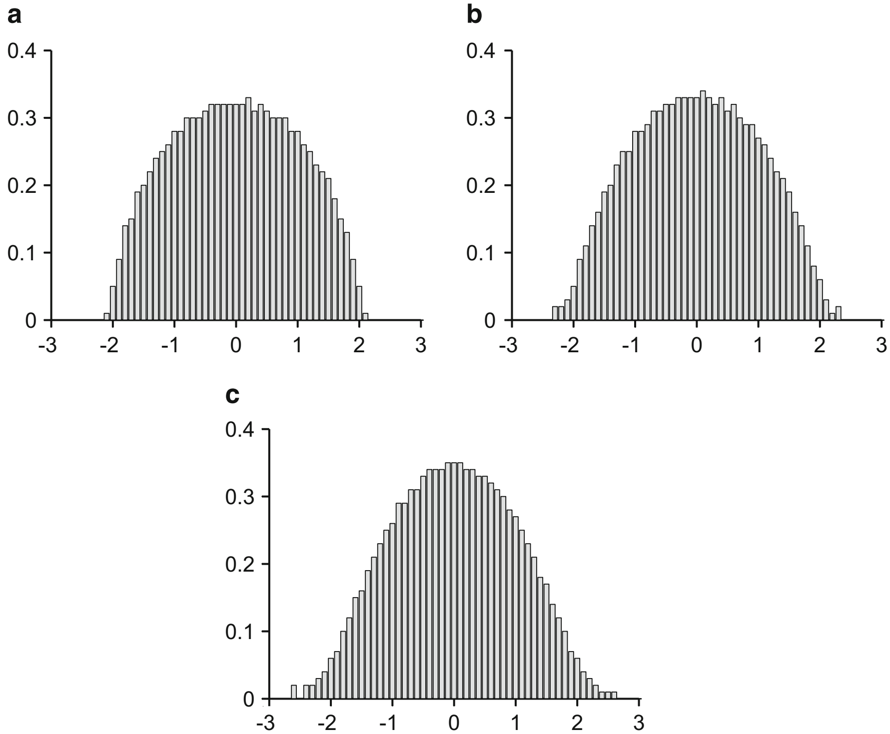


Fig. 1 Histograms of the empirical spectral distribution of 100 realizations of 1,000 × 1,000 matrices $\mathbf{X}_{1{,}000}$ with standard Gaussian entries. (**a**) $c = 0.25$. (**b**) $c = 0.5$. (**c**) $c = 0.75$

monotonically increasing on $(1, \infty)$, and satisfies $m(\beta) \to 0$ as $\beta \searrow 1$ and $m(\beta) \to 1$ as $\beta \to \infty$. We now obtain

$$c = \lim_{n\to\infty} c_n = \begin{cases} 0, & \text{if } \beta \le 1, \\ m(\beta)^2, & \text{if } \beta > 1. \end{cases}$$

Thus, the limiting spectral distribution of $\mathbf{X}_n$ is the semicircle law if $\beta \le 1$, and approximately the Toeplitz limit if β is large. This is insofar not surprising as the different sites in the Curie-Weiss model show little interaction, i.e. behave almost independently, if the temperature is high, or, in other words, β is small. However, if the temperature is low, i.e. β is large, the magnetization of the sites strongly depends on each other. The phase transition at the critical inverse temperature $\beta = 1$ in the Curie-Weiss model is thus reflected in the limiting spectral distribution of $\mathbf{X}_n$ as well.

6 Proof of Theorems 1 and 2

The proofs of both theorems start with the same idea. We need to distinguish them only as soon as the covariances have to be calculated. The main technique we want to apply is the method of moments. The idea is to first determine the weak limit of the expected empirical distribution. Afterwards, concentration inequalities can be used to obtain almost sure convergence.

6.1 The Expected Empirical Spectral Distribution

To determine the limit of the k-th moment of the expected empirical spectral distribution μ_n of $\mathbf{X}_n$, we write

$$\mathbb{E}\left[\int x^k d\mu_n(x)\right] = \frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right]$$
$$= \frac{1}{n^{\frac{k}{2}+1}} \sum_{p_1,\dots,p_k=1}^{n} \mathbb{E}\left[a(p_1,p_2)a(p_2,p_3)\cdots a(p_{k-1},p_k)a(p_k,p_1)\right].$$

The main task is now to compute the expectations on the right hand side. However, we have to face the problem that some of the entries involved are independent and some are not. To be more precise, $a(p_1,q_1),\dots,a(p_j,q_j)$ are independent whenever they can be found on different diagonals of $\mathbf{X}_n$, i.e. the distances $|p_1 - q_1|,\dots,|p_j - q_j|$ are distinct. Hence, a first step in our proof is to consider the expectation $\mathbb{E}\left[a(p_1,p_2)a(p_2,p_3)\cdots a(p_{k-1},p_k)a(p_k,p_1)\right]$, and to identify entries with the same distance of their indices. Therefore, we want to adapt some concepts of [14] and [5] to our situation.

To start with, fix $k \in \mathbb{N}$, and define $\mathcal{T}_n(k)$ to be the set of k-tuples of *consistent pairs*, that is multi-indices $(P_1,\dots,P_k)$ satisfying for any $j = 1,\dots,k$,

1. $P_j = (p_j,q_j) \in \{1,\dots,n\}^2$,
2. $q_j = p_{j+1}$, where $k+1$ is cyclically identified with 1.

With this notation, we find that

$$\frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\dots,P_k)\in\mathcal{T}_n(k)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right].$$

To reflect the dependency structure among the entries $a_n(P_1)\dots a_n(P_k)$, we want to make use of the set $\mathcal{P}(k)$ of partitions of $\{1,\dots,k\}$. Thus, take $\pi \in \mathcal{P}(k)$. We say that an element $(P_1,\dots,P_k) \in \mathcal{T}_n(k)$ is a π*-consistent sequence* if

$$|p_i - q_i| = |p_j - q_j| \iff i \sim_\pi j.$$

According to condition (A2), this implies that $a_n(P_{i_1}), \dots, a_n(P_{i_l})$ are stochastically independent if $i_1, \dots, i_l$ belong to l different blocks of π. The set of all π-consistent sequences $(P_1, \dots, P_k) \in \mathcal{T}_n(k)$ is denoted by $S_n(\pi)$. Note that the sets $S_n(\pi)$, $\pi \in \mathcal{P}(k)$, are pairwise disjoint, and $\bigcup_{\pi \in \mathcal{P}(k)} S_n(\pi) = \mathcal{T}_n(k)$. Consequently, we can write

$$\frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi \in \mathcal{P}(k)} \sum_{(P_1,\dots,P_k)\in S_n(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right]. \tag{4}$$

In a next step, we want to exclude partitions that do not contribute to (4) as $n \to \infty$. These are those partitions satisfying either $\#\pi > k/2$ or $\#\pi < k/2$, where $\#\pi$ denotes the number of blocks of π. We want to treat the two cases separately.

First case: $\#\pi > k/2$. Since π is a partition of $\{1, \dots, k\}$, there is at least one singleton, i.e. a block containing only one element i. Consequently, $a_n(P_i)$ is independent of $\{a_n(P_j), j \neq i\}$ if $(P_1, \dots, P_k) \in S_n(\pi)$. Since we assumed the entries to be centered, we obtain

$$\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \mathbb{E}\Big[\prod_{i \neq l} a_n(P_i)\Big]\mathbb{E}\left[a_n(P_l)\right] = 0.$$

This yields

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\dots,P_k)\in S_n(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = 0.$$

Second case: $r := \#\pi < k/2$. Here, we want to argue that π gives vanishing contribution to (4) as $n \to \infty$ by calculating $\#S_n(\pi)$. To fix an element $(P_1, \dots, P_k) \in S_n(\pi)$, we first choose the pair $P_1 = (p_1, q_1)$. There are at most n possibilities to assign a value to p_1, and another n possibilities for q_1. To fix $P_2 = (p_2, q_2)$, note that the consistency of the pairs implies $p_2 = q_1$. If now $1 \sim_\pi 2$, the condition $|p_1 - q_1| = |p_2 - q_2|$ allows at most two choices for q_2. Otherwise, if $1 \not\sim_\pi 2$, we have at most n possibilities. We now proceed sequentially to determine the remaining pairs. When arriving at some index i, we check whether i is in the same block as some preceding index $1, \dots, i-1$. If this is the case, then we have at most two choices for P_i and otherwise, we have n. Since there are exactly $r = \#\pi$ different blocks, we can conclude that

$$\#S_n(\pi) \le n^2 n^{r-1} 2^{k-r} \le C\, n^{r+1} \tag{5}$$

with a constant $C = C(r, k)$ depending on r and k.

Now the uniform boundedness of the moments (3) and the Hölder inequality together imply that for any sequence $(P_1, \dots, P_k)$,

$$\left|\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right]\right| \le \left[\mathbb{E}\left|a_n(P_1)\right|^k\right]^{\frac{1}{k}} \cdots \left[\mathbb{E}\left|a_n(P_k)\right|^k\right]^{\frac{1}{k}} \le m_k. \tag{6}$$

Consequently, taking account of the relation $r < k/2$, we get

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\ldots,P_k)\in S_n(\pi)} |\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right]| \le C\,\frac{\# S_n(\pi)}{n^{\frac{k}{2}+1}} \le C\,\frac{1}{n^{\frac{k}{2}-r}} = o(1).$$

Combining the calculations in the first and the second case, we can conclude that

$$\frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\substack{\pi\in\mathcal{P}(k),\\ \#\pi=\frac{k}{2}}} \sum_{(P_1,\ldots,P_k)\in S_n(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] + o(1).$$

Now assume that k is *odd*. Then the condition $\#\pi = k/2$ cannot be satisfied, and the considerations above immediately yield

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = 0.$$

It remains to determine the even moments. Thus, let $k \in \mathbb{N}$ be *even*. Recall that we denoted by $\mathcal{PP}(k) \subset \mathcal{P}(k)$ the set of all pair partitions of $\{1,\ldots,k\}$. In particular, $\#\pi = k/2$ for any $\pi \in \mathcal{PP}(k)$. On the other hand, if $\#\pi = k/2$ but $\pi \notin \mathcal{PP}(k)$, we can conclude that π has at least one singleton and hence, as in the first case above, the expectation corresponding to the π-consistent sequences will become zero. Consequently,

$$\frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi\in\mathcal{PP}(k)} \sum_{(P_1,\ldots,P_k)\in S_n(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] + o(1). \tag{7}$$

We have now reduced the original set $\mathcal{P}(k)$ to the subset $\mathcal{PP}(k)$. Next we want to fix a $\pi \in \mathcal{PP}(k)$ and concentrate on the set $S_n(\pi)$. The following lemma will help us to calculate that part of (7) which involves non-crossing partitions.

Lemma 1 (cf. [5], Proposition 4.4.). *Let $S_n^*(\pi) \subseteq S_n(\pi)$ denote the set of π-consistent sequences $(P_1,\ldots,P_k)$ satisfying*

$$i \sim_\pi j \quad\Longrightarrow\quad q_i - p_i = p_j - q_j$$

for all $i \neq j$. Then, we have

$$\#\left(S_n(\pi)\backslash S_n^*(\pi)\right) = o\left(n^{1+\frac{k}{2}}\right).$$

Proof. We call a pair (P_i, P_j) with $i \sim_\pi j, i \neq j$, *positive* if $q_i - p_i = q_j - p_j > 0$ and *negative* if $q_i - p_i = q_j - p_j < 0$. Since $\sum_{i=1}^k q_i - p_i = 0$ by consistency, the existence of a negative pair implies the existence of a positive one. Thus, we can assume that any $(P_1, \dots, P_k) \in S_n(\pi)\backslash S_n^*(\pi)$ contains a positive pair (P_l, P_m). To fix such a sequence, we first determine the positions of l and m, and then fix the signs of the remaining differences $q_i - p_i$. The number of possibilities to accomplish this depends only on k and not on n. Now we choose one of n possible values for p_l, and continue with assigning values to the differences $|q_i - p_i|$ for all P_i except for P_l and P_m. Since π is a pair partition, we have at most $n^{k/2-1}$ possibilities for that. Then, $\sum_{i=1}^k q_i - p_i = 0$ implies that

$$0 < 2(q_l - p_l) = q_l - p_l + q_m - p_m = \sum_{\substack{i\in\{1,\dots,k\},\\ i\neq l,m}} p_i - q_i .$$

Since we have already chosen the signs of the differences $|q_i - p_i|$, $i \neq l, m$, as well as their absolute values, we know the value of the sum on the right hand side. Hence, the difference $q_l - p_l = q_m - p_m$ is fixed. We now have the index p_l, all differences $|q_i - p_i|$, $i \in \{1, \dots, k\}$, and their signs. Thus, we can start at P_l and go systematically through the whole sequence $(P_1, \dots, P_k)$ to see that it is uniquely determined. Consequently, our considerations lead to

$$\#\left(S_n(\pi)\backslash S_n^*(\pi)\right) \leq C n^{\frac{k}{2}} = o\left(n^{1+\frac{k}{2}}\right). \qquad \square$$

A consequence of Lemma 1 and relation (6) is the identity

$$\frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi\in\mathcal{PP}(k)} \sum_{(P_1,\dots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] + o(1). \tag{8}$$

As already mentioned, the sets $S_n^*(\pi)$ help us to deal with the set $\mathcal{NPP}(k)$ of non-crossing pair partitions.

Lemma 2. *Let $\pi \in \mathcal{NPP}(k)$. For any $(P_1, \dots, P_k) \in S_n^*(\pi)$, we have*

$$\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = 1.$$

Proof. Let $l < m$ with $l \sim_\pi m$. Since π is non-crossing, the number $l - m - 1$ of elements between l and m must be even. In particular, there is $l \leq i < j \leq m$ with $i \sim_\pi j$ and $j = i + 1$. By the properties of $S_n^*(\pi)$, we have $a_n(P_i) = a_n(P_j)$, and the sequence $(P_1, \dots, P_l, \dots, P_{i-1}, P_{i+2}, \dots, P_m, \dots, P_k)$ is still consistent. Applying this argument successively, all pairs between l and m can be eliminated and we see that the sequence $(P_1, \dots, P_l, P_m, \dots, P_k)$ is consistent, that is $q_l = p_m$. Then, the identity $p_l = q_m$ also holds. In particular, $a_n(P_l) = a_n(P_m)$. Since this argument applies for arbitrary $l \sim_\pi m$, we obtain

$$\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \prod_{\substack{l<m,\\ l\sim_\pi m}} \mathbb{E}\left[a_n(P_l)a_n(P_m)\right] = 1. \qquad \square$$

By Lemma 2, we can conclude that

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi\in\mathcal{NPP}(k)} \sum_{(P_1,\ldots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi\in\mathcal{NPP}(k)} \# S_n^*(\pi).$$

The following lemma allows us to finally calculate the term on the right hand side.

Lemma 3. *For any* $\pi \in \mathcal{NPP}(k)$, *we have*

$$\lim_{n\to\infty} \frac{\# S_n^*(\pi)}{n^{\frac{k}{2}+1}} = 1.$$

Proof. Since π is non-crossing, we can find a nearest neighbor pair $i \sim_\pi i+1$. Now fix $(P_1,\ldots,P_k) \in S_n^*(\pi)$, and write $P_l = (p_l, p_{l+1})$, $l = 1,\ldots,k$, where $k+1$ is identified with 1. Then the properties of $S_n^*(\pi)$ ensure that $(p_i, p_{i+1}) = (p_{i+2}, p_{i+1})$. Hence, we can eliminate the pairs P_i, P_{i+1} to obtain a sequence $(P_1^{(1)},\ldots,P_{k-2}^{(1)}) := (P_1,\ldots,P_{i-1},P_{i+2},\ldots,P_k)$ which is still consistent. Denote by π' the partition obtained from π by deleting the block $\{i,i+1\}$, and relabeling any $l \geq i+2$ to $l-2$. Since π is non-crossing, we have $\pi' \in \mathcal{NPP}(k-2)$. Moreover, $(P_1^{(1)},\ldots,P_{k-2}^{(1)}) \in S_n^*(\pi')$. Thus we see that any $(P_1,\ldots,P_k) \in S_n^*(\pi)$ can be reconstructed from a tuple $(P_1^{(1)},\ldots,P_{k-2}^{(1)}) \in S_n^*(\pi')$ and a choice of p_{i+1}. The latter admits $n-(k-2)/2$ possibilities since $\{i,i+1\}$ forms a block on its own in π. Consequently,

$$\frac{\# S_n^*(\pi)}{n^{\frac{k}{2}+1}} = \frac{\# S_n^*(\pi')}{n^{\frac{k}{2}}} + o(1). \tag{9}$$

Now if $k=2$, we get $S_n^*(\pi) = \{((p,q),(q,p)) : p,q \in \{1,\ldots,n\}\}$, implying $\# S_n^*(\pi)/n^2 = 1$. For arbitrary even $k \in \mathbb{N}$, the statement of Lemma 3 follows then by induction using the identity in (9). $\square$

Taking account of the relation $\#\mathcal{NPP}(k) = C_{k/2}$, we now arrive at

$$\frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right]$$
$$= C_{\frac{k}{2}} + \frac{1}{n^{\frac{k}{2}+1}} \sum_{\pi\in\mathcal{CPP}(k)} \sum_{(P_1,\ldots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] + o(1), \tag{10}$$

with $\mathcal{CPP}(k)$ being the set of all crossing pair partitions of $\{1, \ldots, k\}$. At this point, we have to distinguish between Theorems 1 and 2. Indeed, to obtain the semicircle law, we need to show that the sum over all crossing partitions is negligible in the limit. However, the limiting distribution in Theorem 2 indicates that we do have a contribution.

6.2 Convergence of the Expected Empirical Spectral Distribution in Theorem 1

The convergence of the expected empirical spectral distribution to the semicircle distribution follows directly from relation (10) and

Lemma 4. *For any crossing $\pi \in \mathcal{CPP}(k)$, we have*

$$\sum_{(P_1,\ldots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = o\left(n^{\frac{k}{2}+1}\right).$$

Proof. Let π be crossing and consider a sequence $(P_1, \ldots, P_k) \in S_n^*(\pi)$. Write $P_l = (p_l, p_{l+1})$. Note that if there is an $l \in \{1, \ldots, k\}$ with $l \sim_\pi l+1$, where $k+1$ is identified with 1, we immediately have $a_n(P_l) = a_n(P_{l+1})$. In particular,

$$\mathbb{E}\left[a_n(P_l)a_n(P_{l+1})\right] = 1.$$

The sequence $(P_1^{(1)}, \ldots, P_{k-2}^{(1)}) := (P_1, \ldots, P_{l-1}, P_{l+2}, \ldots, P_k)$ is still consistent, and

$$\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \mathbb{E}\left[a_n(P_1^{(1)})\cdots a_n(P_{k-2}^{(1)})\right].$$

Define $\pi^{(1)} \in \mathcal{CPP}(k-2)$ to be the pair partition induced by π after eliminating the indices l and $l+1$. In particular, $(P_1^{(1)}, \ldots, P_{k-2}^{(1)}) \in S_n^*(\pi^{(1)})$. Since there are at most n choices for p_{l+1} when $(P_1^{(1)}, \ldots, P_{k-2}^{(1)})$ is fixed, we have for any $(Q_1, \ldots, Q_{k-2}) \in S_n^*(\pi^{(1)})$,

$$\#\{(P_1, \ldots, P_k) \in S_n^*(\pi) : (P_1^{(1)}, \ldots, P_{k-2}^{(1)}) = (Q_1, \ldots, Q_{k-2})\} \leq n.$$

Let r denote the maximum number of pairs of indices that can be eliminated in this way. Since π is crossing, there are at least two pairs left and hence, $r \leq k/2-2$. Define $\pi^{(r)} \in \mathcal{CPP}(k-2r)$ and $(P_1^{(r)}, \ldots, P_{k-2r}^{(r)}) \in S_n^*(\pi^{(r)})$ to be the partition and the sequence left after this elimination. By induction, we conclude that for any $(Q_1, \ldots, Q_{k-2r}) \in S_n^*(\pi^{(r)})$, we have the estimate

$$\#\{(P_1, \ldots, P_k) \in S_n^*(\pi) : (P_1^{(r)}, \ldots, P_{k-2r}^{(r)}) = (Q_1, \ldots, Q_{k-2r})\} \leq n^r.$$

Since $\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \mathbb{E}[a_n(P_1^{(r)})\cdots a_n(P_{k-2r}^{(r)})]$, we obtain

$$\sum_{(P_1,\dots,P_k)\in S_n^*(\pi)} \left|\mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right]\right| \leq n^r \sum_{(Q_1,\dots,Q_{k-2r})\in S_n^*(\pi^{(r)})} \left|\mathbb{E}\left[a_n(Q_1)\cdots a_n(Q_{k-2r})\right]\right|. \tag{11}$$

Choose $i \sim_{\pi^{(r)}} i+j$ such that j is minimal. For any sequence $(Q_1,\dots,Q_{k-2r}) \in S_n^*(\pi^{(r)})$, put $Q_l = (q_l, q_{l+1})$, $l = 1,\dots,k-2r$. We want to count the number of such sequences given that q_i and q_{i+j+1} are fixed. Therefore, we start with choosing one of n possible values for q_{i+1}. Then, the fact that i is equivalent to $i+j$ ensures that we can also deduce the value of

$$q_{i+j} = q_{i+1} - q_i + q_{i+j+1}.$$

Hence, Q_i and Q_{i+j} are fixed. Since j is minimal, any element in $\{i+1,\dots,i+j-1\}$ is equivalent to some element outside the set $\{i,\dots,i+j\}$. There are n possibilities to fix $Q_{i+1} = (q_{i+1}, q_{i+2})$ because q_{i+1} is already fixed. Proceeding sequentially, we have n possibilities for the choice of any pair Q_l with $l \in \{i+2,\dots,i+j-2\}$, and there is only one choice for Q_{i+j-1} since q_{i+j} is already chosen. We thus made n^{j-2} choices to fix all pairs Q_l, $l \in \{i+1,\dots,i+j-1\}$. For any Q_l with $l \in \{1,\dots,k-2r\}\backslash\{i,\dots,i+j\}$, there are at most n possibilities if it is not equivalent to one pair that has already been chosen. Otherwise, there is only one possibility. Since there were $k/2-r-j$ new equivalence classes left, we have at most $n^{k/2-r-j}$ choices for those pairs. Hence, assuming that the elements q_i and q_{i+j+1} are fixed, we have at most

$$n n^{j-2} n^{\frac{k}{2}-r-j} = n^{\frac{k}{2}-r-1}$$

possibilities to choose the rest of the sequence $(Q_1,\dots,Q_{k-2r}) \in S_n^*(\pi^{(r)})$. Note that $|\mathbb{E}[a_n(Q_l)a_n(Q_m)]| \leq (\mathbb{E}[a_n(Q_l)^2])^{1/2}(\mathbb{E}[a_n(Q_m)^2])^{1/2} = 1$. Since $\pi^{(r)}$ is a pair partition, we thus get

$$\left|\mathbb{E}\left[a_n(Q_1)\cdots a_n(Q_{k-2r})\right]\right| \leq |\mathbb{E}[a_n(Q_i)a_n(Q_{i+j})]|.$$

By assumption (A3), the expectation on the right hand side depends only on the absolute value of the difference

$$\min\{q_i, q_{i+1}\} - \min\{q_{i+j}, q_{i+j+1}\} = \max\{q_i, q_{i+1}\} - \max\{q_{i+j}, q_{i+j+1}\}.$$

Now the definition of $S_n^*(\pi^{(r)})$ ensures that $q_i - q_{i+1} = q_{i+j+1} - q_{i+j}$. In particular, $\min\{q_i, q_{i+1}\} - \min\{q_{i+j}, q_{i+j+1}\} = q_i - q_{i+j+1} = q_{i+j} - q_{i+1}$, and

$$\mathbb{E}[a_n(Q_i)a_n(Q_{i+j})] = c_n(|q_i - q_{i+j+1}|).$$

Consequently, estimating the term in (11) further, we obtain

$$\sum_{(P_1,\dots,P_k)\in S_n^*(\pi)} |\mathbb{E}\,[a_n(P_1)\cdots a_n(P_k)]| \le n^{\frac{k}{2}-1} \sum_{q_i,q_{i+j+1}=1}^{n} c_n(|q_{i+j+1}-q_i|)$$
$$\le C\, n^{\frac{k}{2}} \sum_{\tau=0}^{n-1} c_n(\tau) = o\left(n^{\frac{k}{2}+1}\right),$$

since $\sum_{\tau=0}^{n-1} c_n(\tau) = o(n)$ by condition (A4). □

6.3 *Convergence of the Expected Empirical Spectral Distribution in Theorem 2*

We again start with the identity in (10). Since we consider only pair partitions, we know that the expectation on the right hand side is of the form

$$\mathbb{E}\,[a_n(p_1,q_1)a_n(p_1+\tau_1,q_1+\tau_1)]\cdots\mathbb{E}\,[a_n(p_r,q_r)a_n(p_r+\tau_r,q_r+\tau_r)]\,,$$

for $r := k/2$ and some choices of $p_1, q_1, \tau_1, \dots, p_r, q_r, \tau_r \in \mathbb{N}$. In order to calculate this expectation, assumption (A3′) indicates that we only need to distinguish for any $i = 1,\dots,k$, whether we have $\tau_i = 0$ or not. In the first case, we get the identity $\mathbb{E}\,[a_n(p_i,q_i)a_n(p_i+\tau_i,q_i+\tau_i)] = 1$, in the second we can conclude that $\mathbb{E}\,[a_n(p_i,q_i)a_n(p_i+\tau_i,q_i+\tau_i)] = c_n$. Fix some pair partition $\pi \in \mathcal{PP}(k)$, and take $(P_1,\dots,P_k) \in S_n^*(\pi)$. Motivated by these considerations, we put $P_i = (p_i,q_i)$, and define

$$m\,(P_1,\dots,P_k) : = \#\{1 \le i < j \le k : a_n(P_i) = a_n(P_j)\}$$
$$= \#\{1 \le i < j \le k : (p_i,q_i) = (q_j,p_j)\}.$$

Obviously, we have $0 \le m\,(P_1,\dots,P_k) \le k/2$. With this notation, we find that

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\dots,P_k)\in S_n^*(\pi)} \mathbb{E}\,[a_n(P_1)\cdots a_n(P_k)] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{l=0}^{k/2} c_n^{\frac{k}{2}-l}\,\#A_n^{(l)}\,(\pi)\,, \qquad (12)$$

where

$$A_n^{(l)}\,(\pi) := \{(P_1,\dots,P_k) \in S_n^*(\pi) : m\,(P_1,\dots,P_k) = l\}.$$

The following lemma states that if a pair P_i, P_j contributes to $m(P_1,\dots,P_k)$, then we can assume that the block $\{i,j\}$ in π is not crossed by any other block.

Lemma 5. *Let $\pi \in \mathcal{PP}(k)$ and fix $i \sim_\pi j$, $i < j$. Define*

$$S_n^*(\pi; i, j) := \{(P_1, \ldots, P_k) \in S_n^*(\pi) : P_i = (p_i, q_i), P_j = (p_j, q_j), p_i = q_j, q_i = p_j\}.$$

Assume that there is some $i' \sim_\pi j'$ such that $i < i' < j$, and either $j' < i$ or $j < j'$. Then,

$$\# S_n^*(\pi; i, j) = o\left(n^{\frac{k}{2}+1}\right).$$

Proof. To fix some $(P_1, \ldots, P_k) \in S_n^*(\pi; i, j)$, we first choose a value for $p_i = q_j$ and $q_i = p_j$. This allows for at most n^2 possibilities. Hence, P_i and P_j are fixed. Now consider the pairs $P_{i+1}, \ldots, P_{i'-1}$. p_{i+1} is uniquely determined by consistency. For q_{i+1}, there are at most n choices. Then, $p_{i+2} = q_{i+1}$. If $i + 2 \sim_\pi i + 1$, we have one choice for q_{i+2}. Otherwise, there are at most n. Proceeding in the same way, we see that we have n possibilities whenever we start a new equivalence class. Similarly, we can assign values to the pairs $P_j, \ldots, P_{i'+1}$ in this order. Now $P_{i'}$ is determined by consistency. When fixing $P_{i-1}, \ldots, P_1, P_k, \ldots, P_{j+1}$, we again have n choices for any new equivalence class. To sum up, we are left with at most

$$n^2 n^{\frac{k}{2}-2} = n^{\frac{k}{2}}$$

possible values for an element in $S_n^*(\pi; i, j)$. □

Recall Definition 1 where we introduced the notion of the *height* $h(\pi)$ of a pair partition π. Lemma 5 in particular implies that only those $(P_1, \ldots, P_k) \in S_n^*(\pi)$ with

$$0 \le m(P_1, \ldots, P_k) \le h(\pi)$$

contribute to the limit of (12). Indeed, if $m(P_1, \ldots, P_k) > h(\pi)$, we can find some $i \sim_\pi j$, $i < j$, such that $(P_1, \ldots, P_k) \in S_n^*(\pi; i, j)$ and neither $j = i + 1$ nor is the restriction of π to $\{i+1, \ldots, j-1\}$ a pair partition. Hence, the crossing property in Lemma 5 is satisfied, and $(P_1, \ldots, P_k)$ is contained in a set that is negligible in the limit. The identity in (12) thus becomes

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\ldots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \frac{1}{n^{\frac{k}{2}+1}} \sum_{l=0}^{h(\pi)} c_n^{\frac{k}{2}-l} \# B_n^{(l)}(\pi) + o(1),$$

where

$$B_n^{(l)}(\pi) := \big\{(P_1, \ldots, P_k) \in S_n^*(\pi) : m(P_1, \ldots, P_k) = l;$$
$$a_n(P_i) = a_n(P_j), i < j \;\Rightarrow\; j = i + 1 \text{ or } \pi|_{\{i+1,\ldots,j-1\}} \text{ is a pair partition}\big\}.$$

In the next step, we want to simplify the expression above further by showing that $B_n^{(l)}(\pi) = \emptyset$ whenever $0 \leq l < h(\pi)$. This is ensured by

Lemma 6. *Let $\pi \in \mathcal{PP}(k)$. For any $(P_1, \ldots, P_k) \in S_n^*(\pi)$, we have*

$$m(P_1, \ldots, P_k) \geq h(\pi).$$

Proof. If $h(\pi) = 0$, there is nothing to prove. Thus, suppose that $h(\pi) \geq 1$ and take some $i \sim_\pi j$, $i < j$, such that either $j = i + 1$ or $j - i - 1 \geq 2$ is even and the restriction of π to $\{i + 1, \ldots, j - 1\}$ is a pair partition. Fix $(P_1, \ldots, P_k) \in S_n^*(\pi)$, and write $P_l = (p_l, p_{l+1})$ for any $l = 1, \ldots, k$. We need to verify that $p_{i+1} = p_j$. If we achieve this, the definition of $S_n^*(\pi)$ will also ensure that $p_i = p_{j+1}$. As a consequence, the π-block $\{i, j\}$ will contribute to $m(P_1, \ldots, P_k)$. Since there are $h(\pi)$ such blocks, we will obtain $m(P_1, \ldots, P_k) \geq h(\pi)$ for any choice of $(P_1, \ldots, P_k) \in S_n^*(\pi)$.

If $j = i + 1$, we immediately obtain $p_{i+1} = p_j$. To show this property in the second case, note that the sequence $(P_{i+1}, \ldots, P_{j-1})$ solves the following system of equations:

$$\begin{aligned}
p_{i+2} - p_{i+1} + p_{l_1+1} - p_{l_1} &= 0, \quad \text{if } i + 1 \sim_\pi l_1, \\
p_{i+3} - p_{i+2} + p_{l_2+1} - p_{l_2} &= 0, \quad \text{if } i + 2 \sim_\pi l_2, \\
&\vdots \\
p_{i+m+1} - p_{i+m} + p_{l_m+1} - p_{l_m} &= 0, \quad \text{if } i + m \sim_\pi l_m, \\
&\vdots \\
p_j - p_{j-1} + p_{l_{j-i-1}+1} - p_{l_{j-i-1}} &= 0, \quad \text{if } j - 1 \sim_\pi l_{j-i-1}.
\end{aligned}$$

Start with solving the first equation for p_{i+2} which yields

$$p_{i+2} = p_{i+1} - p_{l_1+1} + p_{l_1}.$$

Then, insert this in the second equation, and solve it for p_{i+3} to obtain

$$p_{i+3} = p_{i+1} - p_{l_1+1} + p_{l_1} - p_{l_2+1} + p_{l_2}.$$

In the $j - i - 1$-th step, we substitute $p_{j-1} = p_{i+(j-i-1)}$ in the $j - i - 1$-th equation, and solve it for $p_j = p_{i+(j-i-1)+1}$. We then have

$$p_j = p_{i+1} - \sum_{m=1}^{j-i-1} (p_{l_m+1} - p_{l_m}).$$

Since the restriction of π to $\{i+1,\ldots,j-1\}$ is a pair partition, we can conclude that the sets $\{l_1,\ldots,l_{j-i-1}\}$ and $\{i+1,\ldots,j-1\}$ are equal. Hence, we obtain $\sum_{m=1}^{j-i-1}(p_{l_m+1}-p_{l_m}) = p_j - p_{i+1}$, implying $p_j = p_{i+1}$. □

With the help of Lemma 6, we thus arrive at

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\ldots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \frac{\#B_n^{(h(\pi))}(\pi)}{n^{\frac{k}{2}+1}}\, c_n^{\frac{k}{2}-h(\pi)} + o(1).$$

Note that any element $(P_1,\ldots,P_k)\in S_n^*(\pi)$ satisfying the condition

$$a_n(P_i) = a_n(P_j),\ i<j \quad \Rightarrow \quad j=i+1 \text{ or } \pi|_{\{i+1,\ldots,j-1\}} \text{ is a pair partition,} \tag{13}$$

fulfills the condition $m(P_1,\ldots,P_k) = h(\pi)$ as well. Indeed, (13) guarantees that $m(P_1,\ldots,P_k) \le h(\pi)$, and Lemma 6 ensures that $m(P_1,\ldots,P_k) \ge h(\pi)$. Thus, we can write

$$\begin{aligned} B_n^{(h(\pi))}(\pi) = \{&(P_1,\ldots,P_k)\in S_n^*(\pi) : \\ &a_n(P_i)=a_n(P_j), i<j \;\Rightarrow\; j=i+1 \text{ or } \pi|_{\{i+1,\ldots,j-1\}} \text{ is a pair partition}\big\}. \end{aligned}$$

Now any element in the complement of $B_n^{(h(\pi))}(\pi)$ satisfies for some $i\sim_\pi j$ the crossing assumption in Lemma 5. This yields

$$\frac{\#\left(B_n^{(h(\pi))}(\pi)\right)^c}{n^{\frac{k}{2}+1}} = o(1).$$

Since $B_n^{(h(\pi))}(\pi) \cup \left(B_n^{(h(\pi))}(\pi)\right)^c = S_n^*(\pi)$, we obtain that

$$\frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\ldots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}}\, c_n^{\frac{k}{2}-h(\pi)} + o(1). \tag{14}$$

To calculate the limit on the right-hand side, we have

Lemma 7 (cf. [5], Lemma 4.6). *For any $\pi\in\mathcal{PP}(k)$, it holds that*

$$\lim_{n\to\infty} \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} = p_T(\pi),$$

where $p_T(\pi)$ is the Toeplitz volume defined by solving the system of equation (2).

Proof. Fix $\pi \in \mathcal{PP}(k)$. Note that if $P = \{(p_i, p_{i+1}), i = 1, \ldots, k\} \in S_n^*(\pi)$, then $x_0, x_1, \ldots, x_k$ with $x_i = p_{i+1}/n$ is a solution of the system of equation (2). On the other hand, if $x_0, x_1, \ldots, x_k \in \{1/n, 2/n, \ldots, 1\}$ is a solution of (2) and $p_{i+1} = nx_i$, then either $\{(p_i, p_{i+1}), i = 1, \ldots, k\} \in S_n^*(\pi)$ or $\{(p_i, p_{i+1}), i = 1, \ldots, k\} \in S_n(\eta)$ for some partition $\eta \in \mathcal{P}(k)$ such that $i \sim_\pi j \Rightarrow i \sim_\eta j$, but $\#\eta < \#\pi$.

In (2), we have $k + 1$ variables and only $k/2$ equations. Denote the $k/2 + 1$ undetermined variables by $y_1, \ldots, y_{k/2+1}$. We thus need to assign values from the set $\{1/n, 2/n, \ldots, 1\}$ to $y_1, \ldots, y_{k/2+1}$, and then to calculate the remaining $k/2$ variables from the equations. Since the latter are also supposed to be in the range $\{1/n, 2/n, \ldots, 1\}$, it might happen that not all values for the undetermined variables are admissible. Let $p_n(\pi)$ denote the admissible fraction of the $n^{k/2+1}$ choices for $y_1, \ldots, y_{k/2+1}$. By our remark at the beginning of the proof and estimate (5), we have that

$$\lim_{n\to\infty} \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}} = \lim_{n\to\infty} p_n(\pi),$$

if the limits exist. Now we can interpret $y_1, \ldots, y_{k/2+1}$ as independent random variables with a uniform distribution on $\{1/n, 2/n, \ldots, 1\}$. Then, $p_n(\pi)$ is the probability that the computed values stay within the interval (0, 1]. As $n \to \infty$, $y_1, \ldots, y_{k/2+1}$ converge in law to independent random variables uniformly distributed on [0, 1]. Hence, $p_n(\pi) \to p_T(\pi)$. □

Applying Lemma 7 and assumption (A4′) to Eq. (14), we arrive at

$$\lim_{n\to\infty} \frac{1}{n^{\frac{k}{2}+1}} \sum_{(P_1,\ldots,P_k)\in S_n^*(\pi)} \mathbb{E}\left[a_n(P_1)\cdots a_n(P_k)\right] = p_T(\pi)c^{\frac{k}{2}-h(\pi)}.$$

Substituting this result in (10), we find that for any even $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = C_{\frac{k}{2}} + \sum_{\pi\in\mathcal{CPP}(k)} p_T(\pi)c^{\frac{k}{2}-h(\pi)}.$$

To obtain the alternative expression for the even moments in Theorem 2, note that the considerations above were not restricted to crossing partitions. In particular, we can start from identity (8) instead of (10) to see that

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] = \lim_{n\to\infty} \sum_{\pi\in\mathcal{PP}(k)} \frac{\#S_n^*(\pi)}{n^{\frac{k}{2}+1}}\, c_n^{\frac{k}{2}-h(\pi)} = \sum_{\pi\in\mathcal{PP}(k)} p_T(\pi)c^{\frac{k}{2}-h(\pi)}.$$

6.4 Almost Sure Convergence

The aim of this part of the proof is to show almost sure convergence in both situations, that of Theorem 1 and that of Theorem 2. Therefore, we want to follow the ideas used in [5], Proposition 4.9, to verify the same for Toeplitz matrices. This is possible since the arguments only rely on the fact that the diagonals are independent. The particular dependency structure can be neglected. We want to start with the proof of a concentration inequality for the moments of the spectral measure. The obtained bound will enable us to use the Borel-Cantelli lemma.

Lemma 8. *Suppose that conditions (A1) and (A2) hold. Then, for any* $k, n \in \mathbb{N}$*,*

$$\mathbb{E}\left[\left(\operatorname{tr}\left(X_n^k\right) - \mathbb{E}\left[\operatorname{tr}\left(X_n^k\right)\right]\right)^4\right] \leq C\, n^2.$$

Proof. Fix $k, n \in \mathbb{N}$. Using the notation

$$P = (P_1, \ldots, P_k) = ((p_1, q_1), \ldots, (p_k, q_k)), \qquad a_n(P) = a_n(P_1) \cdots a_n(P_k),$$

we have that

$$\mathbb{E}\left[\left(\operatorname{tr}\left(\mathbf{X}_n^k\right) - \mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right]\right)^4\right]$$
$$= \frac{1}{n^{2k}} \sum_{\pi^{(1)},\ldots,\pi^{(4)} \in \mathcal{P}(k)} \sum_{\substack{P^{(i)} \in \mathcal{S}_n\left(\pi^{(i)}\right), \\ i=1,\ldots,4}} \mathbb{E}\Big[\prod_{j=1}^{4} \left(a_n(P^{(j)}) - \mathbb{E}\left[a_n(P^{(j)})\right]\right)\Big]. \tag{15}$$

Now consider a partition $\boldsymbol{\pi}$ of $\{1, \ldots, 4k\}$. We say that a sequence $(P^{(1)}, \ldots, P^{(4)})$ is π-consistent if each $P^{(i)}, i = 1, \ldots, 4$, is a consistent sequence and

$$\left|q_l^{(i)} - p_l^{(i)}\right| = \left|q_m^{(j)} - p_m^{(j)}\right| \quad \Longleftrightarrow \quad l + (i-1)k \sim_{\boldsymbol{\pi}} m + (j-1)k.$$

Let $\mathcal{S}_n(\pi)$ denote the set of all π-consistent sequences with entries in $\{1, \ldots, n\}$. Then, (15) becomes

$$\mathbb{E}\left[\left(\operatorname{tr}\left(\mathbf{X}_n^k\right) - \mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right]\right)^4\right]$$
$$= \frac{1}{n^{2k}} \sum_{\pi \in \mathcal{P}(4k)} \sum_{(P^{(1)},\ldots,P^{(4)}) \in \mathcal{S}_n(\pi)} \mathbb{E}\Big[\prod_{j=1}^{4} \left(a_n(P^{(j)}) - \mathbb{E}\left[a_n(P^{(j)})\right]\right)\Big]. \tag{16}$$

We want to analyze the expectation on the right hand side. Therefore, fix a $\pi \in \mathcal{P}(4k)$. We call π a *matched partition* if

1. Any block of π contains at least two elements,
2. For any $i \in \{1, \dots, 4\}$, there is a $j \neq i$ and $l, m \in \{1, \dots, k\}$ with

$$l + (i-1)k \ \sim_{\boldsymbol{\pi}} \ m + (j-1)k.$$

In case π does not satisfy (i), we have a singleton $\{l + (i-1)k\}$, implying that $\mathbb{E}[a_n(P^{(i)})] = 0$. As a consequence,

$$\sum_{(P^{(1)},\dots,P^{(4)})\in\mathcal{S}_n(\pi)} \mathbb{E}\Big[\prod_{j=1}^{4}\big(a_n(P^{(j)}) - \mathbb{E}\big[a_n(P^{(j)})\big]\big)\Big] = 0. \tag{17}$$

If π does not satisfy (ii), we can conclude that for some $i \in \{1, \dots, 4\}$, the term $a_n(P^{(i)}) - \mathbb{E}\big[a_n(P^{(i)})\big]$ is independent of $a_n(P^{(j)}) - \mathbb{E}\big[a_n(P^{(j)})\big]$, $j \neq i$. Thus, (17) holds in this case as well. To sum up, we only have to consider matched partitions to evaluate the sum in (16). Let $\boldsymbol{\pi}$ be such a partition and denote by $r = \#\pi$ the number of blocks of π. Note that condition (i) implies $r \leq 2k$. We want to count all π consistent sequences $(P^{(1)}, \dots, P^{(4)})$. Therefore, first choose one of at most n^r possibilities to fix the r different equivalence classes. Afterwards, we fix the elements $p_1^{(1)}, \dots, p_1^{(4)}$, which can be done in n^4 ways. Since now the differences $|q_l^{(i)} - p_l^{(i)}|$ are uniquely determined by the choice of the corresponding equivalence classes, we can proceed sequentially to see that there are at most two choices left for any pair $P_l^{(i)}$. To sum up, we have at most

$$2^{4k} n^4 n^r = C\ n^{r+4}$$

possibilities to choose $(P^{(1)}, \dots, P^{(4)})$. If now $r \leq 2k - 2$, we can conclude that

$$\#\mathcal{S}_n(\pi) \leq C\ n^{2k+2}. \tag{18}$$

It remains to consider the cases in which $r = 2k - 1$ and $r = 2k$, respectively. To begin with, let $r = 2k - 1$. Then, we have either two equivalence classes with three elements or one equivalence class with four. Since π is matched, there must exist an $i \in \{1, \dots, 4\}$ and an $l \in \{1, \dots, k\}$ such that $P_l^{(i)}$ is not equivalent to any other pair in the sequence $P^{(i)}$. Without loss of generality, we can assume that $i = 1$. In contrast to the construction of $(P^{(1)}, \dots, P^{(4)})$ as above, we now alter our procedure as follows: We fix all equivalence classes except of that $P_l^{(1)}$ belongs to. There are n^{r-1} possibilities to accomplish that. Now we choose again one of n^4 possible values for $p_1^{(1)}, \dots, p_1^{(4)}$. Hereafter, we fix $q_m^{(1)}$, $m = 1, \dots, l-1$, and then start from $q_k^{(1)} = p_1^{(1)}$ to go backwards and obtain the values of $p_k^{(1)}, \dots, p_{l+1}^{(1)}$. Each of these steps leaves at most two choices to us, that is 2^{k-1} choices in total. But now, $P_l^{(1)}$ is uniquely determined since $p_l^{(1)} = q_{l-1}^{(1)}$ and $q_l^{(1)} = p_{l+1}^{(1)}$ by consistency. Thus, we had to make one choice less than before, implying (18).

Now, let $r = 2k$. In this case, each equivalence class has exactly two elements. Since we consider a matched partition, we can find here as well an $l \in \{1, \dots, k\}$ such that $P_l^{(1)}$ is not equivalent to any other pair in the sequence $P^{(1)}$. But in addition to that, we also have an $m \in \{1, \dots, k\}$ such that, possibly after relabeling, $P_m^{(2)}$ is neither equivalent to any element in $P^{(1)}$ nor to any other element in $P^{(2)}$. Thus, we can use the same argument as before to see that this time, we can reduce the number of choices to at most $C\, n^{r+2} = C\, n^{2k+2}$. In conclusion, (18) holds for any matched partition π. To sum up our results, we obtain that

$$\mathbb{E}\left[\left(\operatorname{tr}\left(\mathbf{X}_n^k\right) - \mathbb{E}\left[\operatorname{tr}\left(\mathbf{X}_n^k\right)\right]\right)^4\right]$$
$$= \frac{1}{n^{2k}} \sum_{\substack{\pi \in \mathcal{P}(4k), \\ \pi \text{ matched}}} \sum_{(P^{(1)},\dots,P^{(4)}) \in \mathcal{S}_n(\pi)} \mathbb{E}\Big[\prod_{j=1}^{4} \left(a_n(P^{(j)}) - \mathbb{E}\left[a_n(P^{(j)})\right]\right)\Big] \leq C\, n^2,$$

which is the statement of Lemma 8. □

From Lemma 8 and Chebyshev's inequality, we can now conclude that for any $\varepsilon > 0$ and any $k, n \in \mathbb{N}$,

$$\mathbb{P}\left(\left|\frac{1}{n}\operatorname{tr}\left(\mathbf{X}_n^k\right) - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}\left(\mathbf{X}_n^k\right)\right]\right| > \varepsilon\right) \leq \frac{C}{\varepsilon^4 n^2}.$$

Applying the Borel-Cantelli lemma, we see that

$$\frac{1}{n}\operatorname{tr}\left(\mathbf{X}_n^k\right) - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}\left(\mathbf{X}_n^k\right)\right] \to 0, \quad \text{a.s..} \tag{19}$$

Let Y be a random variable distributed according to the semicircle law or according to the distribution given by the moments in Theorem 2, depending on whether we want to prove Theorems 1 or 2. The convergence of the moments of the expected empirical distributions and relation (19) yield

$$\frac{1}{n}\operatorname{tr}\left(\mathbf{X}_n^k\right) \to \mathbb{E}[Y^k], \quad \text{a.s..}$$

Since the distribution of Y is uniquely determined by its moments, we obtain almost sure weak convergence of the empirical spectral distribution of $\mathbf{X}_n$ to the distribution of Y.

References

1. Arnold, L.: On wigner's semicircle law for the eigenvalues of random matrices. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **19**, 191–198 (1971)
2. Bai, Z.D.: Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. Ann. Probab. **21**(2), 625–648 (1993)
3. Bai, Z.: Methodologies in spectral analysis of large-dimensional random matrices, a review. Stat. Sin. **9**(3), 611–677 (1999). With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author
4. Ben Arous, G., Guionnet, A.: Large deviations for Wigner's law and Voiculescu's non-commutative entropy. Probab. Theory Relat. Fields **108**(4), 517–542 (1997)
5. Bryc, W., Dembo, A., Jiang, T.: Spectral measure of large random Hankel, Markov and Toeplitz matrices. Ann. Probab. **34**(1), 1–38 (2006)
6. Ellis, R., Newman, C.: Fluctuationes in Curie-Weiss exemplis. In: Osterwalder, K., Seiler, E. (eds.) Mathematical Problems in Theoretical Physics. Lecture Notes in Physics, vol. 80, pp. 313–324. Springer, Berlin/Heidelberg (1978)
7. Ellis, R.S.: Entropy, Large Deviations, and Statistical Mechanics. Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 271. Springer, New York (1985)
8. Erdös, L.: Universality of Wigner random matrices: a survey of recent results. Uspekhi Mat. Nauk **66**(3(399)), 67–198 (2011)
9. Erdős, L., Schlein, B., Yau, H.T.: Local semicircle law and complete delocalization for Wigner random matrices. Commun. Math. Phys. **287**(2), 641–655 (2009)
10. Friesen, O., Löwe, M.: The semicircle law for matrices with independent diagonals. J. Theor. Probab. (2011, Preprint, To appear)
11. Friesen, O., Löwe, M.: A phase transition for the limiting spectral distribution of random matrices. Electron. J. Probab. **114**(17), 1-17 (2013)
12. Götze, F., Tikhomirov, A.: Rate of convergence to the semi-circular law. Probab. Theory Relat. Fields **127**, 228–276 (2003)
13. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov Chains and Mixing Times. American Mathematical Society, Providence (2006)
14. Schenker, J., Schulz-Baldes, H.: Semicircle law and freeness for random matrices with symmetries or correlations. Math. Res. Lett. **12**, 531–542 (2005)
15. Wigner, E.P.: On the distribution of the roots of certain symmetric matrices. Ann. Math. **67**, 325–328 (1958)
16. Wishart, J.: The generalized product moment distribution in samples from a normal multivariate population. Biometrika **20**, 32–52 (1928)

Asymptotic Eigenvalue Distribution of Random Matrices and Free Stochastic Analysis

Roland Speicher*

Abstract This is a survey on some recent work on free stochastic calculus and free Malliavin calculus. It is hoped that these theories will in the long run provide us with tools for qualitative descriptions of the asymptotic eigenvalue distribution of selfadjoint polynomials of independent Gaussian random matrices. The main concrete results center around the free Fourth Moment Theorem, which says that for a sequence of random variables which are constrained to live in a fixed free chaos, the convergence to the semicircle distribution can be controlled by the convergence of the second and the fourth moments.

1 Random Matrices

Our main interest lies in distributions of polynomials of free semicircular elements. Since those distributions arise as limits of eigenvalue distributions of quite natural random matrix models we will first look, to motivate the kind of questions, on those random matrices.

1.1 Wigner's Semicircle Law

Let $X_N = (X_N(i,j))_{i,j=1}^N$ be an $N \times N$ *Gaussian random matrix*. This means that

*I thank Octavio Arizmendi for the preparation of the figures.

R. Speicher (✉)
Saarland University, Fachrichtung Mathematik, Postfach 151150, 66041, Germany
e-mail: speicher@math.uni-sb.de

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, 31
Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_2, © Springer-Verlag Berlin Heidelberg 2013

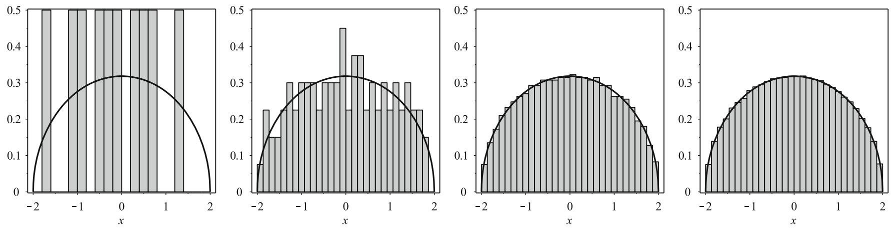


Fig. 1 Eigenvalue distribution of a $N \times N$ Gaussian random matrix, for $N = 10, 100, 1{,}000, 4{,}000$; the *solid curve* is the semicircular distribution

- X_N is selfadjoint, $X_N^* = X_N$
- $\{X_N(i,j) \mid 1 \leq i \leq j \leq N\}$ are i.i.d. random variables (say real-valued), with normal distribution of mean zero and variance $1/N$

Then, Wigner's famous semicircle law [1, 15] says that, for $N \to \infty$, we have convergence of the empirical eigenvalue distribution of X_N to the semicircle distribution. The latter lives on the interval $[-2,2]$ and has there the density $1/(2\pi)\sqrt{4-x^2}\mathrm{d}x$.

Numerical simulations for this approximation of the semicircle are shown in Fig. 1.

One way to capture the essence of Wigner's semicircle law is the statement that the moments $\mathrm{tr}[X_N^n]$ of the Gaussian random matrices converge, for $N \to \infty$, almost surely to the corresponding moments $\frac{1}{2\pi}\int_{-2}^{+2} x^n \sqrt{4-x^2}\mathrm{d}x$ of the semicircular distribution. With tr we denote the normalized trace on matrices,

$$\mathrm{tr}\big((a_{ij})_{i,j=1}^N\big) := \frac{1}{N}\sum_{i=1}^N a_{ii}.$$

We write $X_N \to s$ to indicate the above kind of convergence and say that X_N *converges in distribution* to s. The limiting element s is a semicircular element living in an abstract non-commutative probability space $(\mathscr{A}, \varphi)$, where $\mathscr{A}$ is a unital algebra and $\varphi : \mathscr{A} \to \mathbb{C}$ a unital linear functional; s is determined by the fact that its moments $\varphi(s^n)$ are the moments of the semicircular distribution. The notation $X_N \to s$ means here, by definition, the almost sure convergence

$$\lim_{N\to\infty} \mathrm{tr}[X_N^n] = \varphi(s^n) \qquad \forall n \in \mathbb{N}.$$

See, e.g., [9, 14].

(Note that for the case of one random matrix the use of a non-commutative probability space is quite artificial; we can just take a classical random variable s with semicircle distribution and put $\mathscr{A}$ equal to the algebra generated by s, and take as φ the expectation. Since the semicircular distribution is determined by its moments, $X_N \to s$ is then the same as weak convergence of the empirical eigenvalue distribution of X_N to the semicircular distribution. However, in the multivariate case, where we consider several matrices, we will be leaving the commutative world, and the above non-commutative frame will then be unavoidable.)

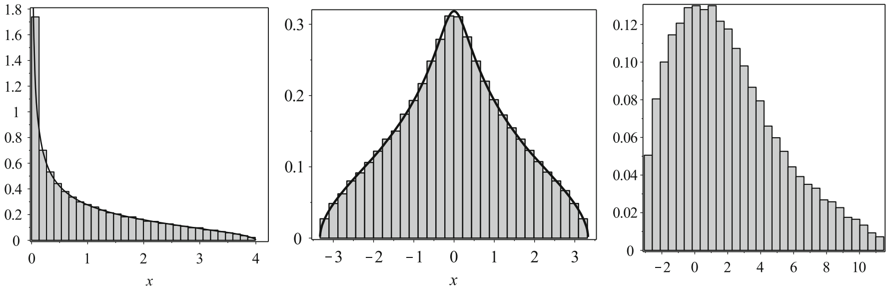


Fig. 2 Eigenvalue distributions of $p(x_1) = x_1^2$ (*left*), $p(x_1, x_2) = x_1x_2 + x_2x_1$ (*middle*) and $p(x_1, x_2, x_3) = x_1x_2^2x_1 + x_1x_3^2x_1 + 2x_2$ (*right*), where x_1, x_2, and x_3 are independent 4,000×4,000 Gaussian random matrices; in the first two cases the asymptotic eigenvalue distribution can be calculated by free probability tools as the *solid curve*, for the third case no explicit solution exists

1.2 Polynomials in Independent Gaussian Random Matrices

Consider now m independent Gaussian random matrices $X_N^{(1)}, \ldots, X_N^{(m)}$. Then, by Voiculescu's multivariate generalization [12] of Wigner's semicircular law, we know that those converge, for $N \to \infty$, to m free semicircular elements $s_1, \ldots, s_m$, living in some non-commutative probability space $(\mathscr{A}, \varphi)$. We will define later the exact meaning of "freeness"; for the moment it suffices to say that this means that it allows in principle to calculate all mixed moments

$$\varphi(s_{i(1)} \cdots s_{i(n)}) = \lim_{N\to\infty} \mathrm{tr}[X_N^{i(1)} \cdots X_N^{i(n)}].$$

In particular, for any selfadjoint polynomial p in m non-commuting variables we have that

$$p(X_N^{(1)}, \ldots, X_N^{(m)}) \to p(s_1, \ldots, s_m).$$

(A selfadjoint polynomial $p(x_1, \ldots, x_m)$ in m non-commuting variables $x_1, \ldots, x_m$ is one for which we have $p^* = p$, if we declare the variables to be selfadjoint, $x_i^* = x_i$ for all $i = 1, \ldots, m$. This guarantees that the application of this polynomial to any m-tuple of selfadjoint operators yields a selfadjoint operator.)

So, in principle, we know everything about the limiting eigenvalue distribution of $p(X_N^{(1)}, \ldots, X_N^{(m)})$. However, in practice, unless the polynomial is very simple (like the sum of one-variable polynomials), we cannot say anything concrete about this distribution. (For more information on the methods and results in the simple cases one might consult, e.g., [2, 9].)

In Fig. 2 we present histograms for eigenvalue distributions for various polynomials. One dimensional cases like $p(x) = x^2$ can be calculated directly by classical means, and some special multivariate cases like $p(x_1, x_2) = x_1x_2 + x_2x_1$ are simple enough to be calculated by free probability methods. However, for more generic polynomials, like $p(x_1, x_2, x_3) = x_1 x_2^2 x_1 + x_1 x_3^2 x_1 + 2x_2$, there is no hope to derive any explicit formula for the limiting eigenvalue distribution.

Where we have hope and what the following will be about, are qualitative features of the distributions which arise in this way.

2 Distribution of Non-commuting Variables

2.1 How Can We Deal with Multivariate Situations?

Let us start with some general remarks on understanding the distribution of non-commuting variables $X_1, \ldots, X_m$? In particular, we want to point out that there is quite a difference between the commutative and the non-commutative situation in the multivariate case $m > 1$.

In the case of commuting variables, the distribution of one variable is a probability measure on $\mathbb{R}$, the distribution of m commuting variables is a probability measure on $\mathbb{R}^m$ – which is not so much of a difference.

In the case of non-commuting variables, the distribution of one variable is still a probability measure on $\mathbb{R}$, but the distribution of $m > 1$ variables is something quite different: analytically, it is a state on the algebra generated by the variables; or combinatorially, it is the collection of all their mixed moments, and there are much more mixed moments in non-commuting variables than in commuting ones.

At the moment we are still lacking the deeper analytic tools to deal directly with the distribution of several non-commuting variables. (Voiculescu's *free analysis* is aiming in this direction, see [13]; see also the Introduction of [5] for a discussion of the problem of multivariate non-commutative distributions.) What we will consider in the following is a kind of reduction to the one-dimensional classical case. Namely, in order to understand our non-commuting variables $X_1, \ldots, X_m$, we will be satisfied with understanding all selfadjoint polynomials in $X_1, \ldots, X_m$. The distribution of each such polynomial is just a probability measure on $\mathbb{R}$, so we might be able to use classical tools to deal with those. However, we should somehow understand all such polynomials.

2.2 Conjectures on Distributions of Several Non-commuting Variables?

There is no hope to have explicit formulas for the asymptotic eigenvalue distributions of general polynomials p in independent Gaussian random matrices or, equivalently, the distribution of polynomials p in free semicircular variables,

$$\lim_{N \to \infty} p(X_N^{(1)}, \ldots, X_N^{(m)}) = p(s_1, \ldots, s_m).$$

However, from the one-variable case and numerical simulations we expect that the distributions for such situations share some common qualitative features. In particular, we believe that for an arbitrary selfadjoint polynomial p the following should be true for the distribution of $p(s_1, \dots, s_m)$.

- It should have no atoms.
- It should have a density with respect to Lebesgue measure.
- This density should also have some nice regularity properties.

There is at least one statement which we can infer from general operator algebraic results. Namely, the facts that polynomials in semicircular elements can be realized in the C^*-algebra of the free group and that the latter has no non-trivial projections imply that the support of the distribution of $p(s_1, \dots, s_m)$ has to be an interval; see, e.g., [7]. The above mentioned regularity properties can then be stated a bit more precisely by conjecturing that the density is continous or even analytic on the interior of this interval. (The example of the square of the semicircle, $p(s) = s^2$, shows that the density can go to infinity at the boundary of the support. see Fig. 2.)

2.3 A Side-Remark: The Linearization Trick

One possible simplification in this business is the following linearization trick. Instead of looking on all polynomials, it might be enough to consider just linear polynomials in $X_1, \dots, X_m$. In the commutative case this is actually enough; knowing the distribution of all linear combinations of the random variables $X_1, \dots, X_m$ is equivalent to the knowledge of the joint distribution of the random variables (and thus to the knowledge of the distribution of arbitrary polynomials in the variables). In the non-commutative case this is not true any more; what one needs there are all operator-valued linear polynomials, i.e., matrices have to be allowed as coefficients in the linear combinations. This is much more general than just scalar-valued linear polynomials, but on the other hand for many calculations the nature of the coefficients is not important, so much of the theory has still the flavour of the usual linear theory.

A striking example of how far this linearization trick can be pushed is the following result of Haagerup and Thorbjornsen [6] (see also [7]) about the largest eigenvalue of any selfadjoint polynomial in independent Gaussian random matrices:

$$\lambda_{\max}\big(p(X_N^{(1)}, \dots, X_N^{(m)})\big) \to \|p(s_1, \dots, s_m)\| \qquad \text{almost surely,}$$

where $\|p(s_1, \dots, s_m)\|$ denotes the operator norm of the operator $p(s_1, \dots, s_m)$ (for example, in a concrete realization of $s_1, \dots, s_m$ as operators on some Hilbert space).

We will, however, not follow this direction, but will try to understand the distribution of general polynomials $p(s_1, \dots, s_m)$ in free semicircular elements by invoking free stochastic calculus.

3 Free Stochastic Calculus

Our approach to the question of qualitative features of distributions of polynomials in free semicircular elements will rely on stochastic calculus, the rough idea being to built a polynomial $p(s_1, \dots, s_m)$ in a differential way and control the differential changes along the way by free stochastic calculus and free Malliavin calculus. So, in the following we will first recall a bit of free stochastic calculus before we come back to our original problem.

As a general reference to basic free probability notions we refer to [9, 14]. Free stochastic calculus was introduced by Biane and Speicher in [3], a good summary can also be found in [8].

3.1 Freeness

Definition 1. Let $(\mathscr{A}, \varphi)$ be a non-commutative probability space; i.e., $\mathscr{A}$ is a unital algebra and $\varphi : \mathscr{A} \to \mathbb{C}$ is a unital linear functional. Unital subalgebras $(\mathscr{A}_i)_{i \in I}$ (for some index set I) are *free* if $\varphi(a_1 \cdots a_n) = 0$ whenever for each $k = 1, \dots, n$ we have $a_k \in \mathscr{A}_{i(k)}$ for some $i(k) \in I$ such that $i(k) \neq i(k+1)$ for all $k = 1, \dots, n-1$ and such that $\varphi(a_k) = 0$ for all $k = 1, \dots, n$.

Random variables $x_i \in \mathscr{A}$ $(i \in I)$ are free, if their generated unital subalgebras are free.

3.2 Free Brownian Motion

Definition 2. A *free Brownian motion* is given by a family $(S(t))_{t \geq 0}$ of random variables $S(t) \in \mathscr{A}$ $(t \geq 0)$, living in a non-commutative probability space $(\mathscr{A}, \varphi)$, where $\mathscr{A}$ is a von Neumann algebra and φ a faithful trace, such that

- $S(0) = 0$
- Each increment $S(t) - S(s)$ $(0 \leq s < t)$ is semicircular with mean 0 and variance $t - s$, i.e.,

$$d\mu_{S(t)-S(s)}(x) = \frac{1}{2\pi(t-s)} \sqrt{4(t-s) - x^2} dx$$

- Disjoint increments are free: for $0 < t_1 < t_2 < \cdots < t_n$, the increments $S(t_1)$, $S(t_2) - S(t_1)$, …, $S(t_n) - S(t_{n-1})$ are free.

A free Brownian motion can also be realized concretely: by the sum of creation and annihilation operators on the full Fock space; and as the limit of matrix-valued (Dyson) Brownian motions. This latter realization is as follows.

Let $(X_N(t))_{t\geq 0}$ be a symmetric $N \times N$-matrix-valued Brownian motion, i.e.,

$$X_N(t) = \begin{pmatrix} B_{11}(t) & \dots & B_{1N}(t) \\ \vdots & \ddots & \vdots \\ B_{N1}(t) & \dots & B_{NN}(t) \end{pmatrix}$$

where

- B_{ij} are, for $i \geq j$, independent classical Brownian motions
- $B_{ij}(t) = B_{ji}(t)$.

Then, $\big(X_N(t)\big)_{t\geq 0} \to \big(S(t)\big)_{t\geq 0}$, in the sense that almost surely

$$\lim_{N\to\infty} \mathrm{tr}\big(X_N(t_1)\cdots X_N(t_n)\big) = \varphi\big(S(t_1)\cdots S(t_n)\big) \qquad \forall\, 0 \leq t_1, t_2, \dots, t_n$$

3.3 *Stochastic Analysis on "Wigner" Space*

3.3.1 Definition of Free Stochastic Integrals

Starting from a free Brownian motion $\big(S(t)\big)_{t\geq 0}$ we define multiple *Wigner integrals*

$$I(f) = \int \cdots \int f(t_1, \dots, t_n) \mathrm{d}S(t_1) \dots \mathrm{d}S(t_n)$$

for scalar-valued functions $f \in L^2(\mathbb{R}_+^n)$, by avoiding the diagonals, i.e. we understand in the above that all times t_i are different. The definition is made rigorous, by first defining $I(f)$ for step functions f which have no support on the diagonals; for those one establishes an Ito isometry

$$\|I(f)\|_2 = \|f\|_{L^2(\mathbb{R}^n)},$$

and then uses this to extend the definition to all $f \in L^2(\mathbb{R}_+^n)$. The integral $I(f)$ is then an element in the L^2-space of the free Brownian motion. The 2-norm on $\mathscr{A}$ is as usual given by $\|A\|_2^2 := \varphi(AA^*)$, and the L^2-space is then the completion of $\mathscr{A}$ with respect to $\|\cdot\|_2$.

3.3.2 A Side Remark: Free Stochastic Integrals are Usually Bounded Operators

We want to remark that free stochastic integrals are not just elements in the L^2-space, but usually are bounded operators and thus their distribution has compact support. Note that the fact that all our $S(t)$ and their increments are bounded does not imply the same property for the integrals (as they are limits of sums), but one needs the freeness of the increments as the essential input to prove this. More precisely, we have the following estimate (which is the analogue of a Haagerup inequality) [3,4]:

$$\left\| \int \cdots \int f(t_1,\ldots,t_n)\mathrm{d}S(t_1)\ldots\mathrm{d}S(t_n) \right\| \le (n+1)\|f\|_{L^2(\mathbb{R}_+^n)}.$$

The operator norm in the above can be recovered from the moments of the variables even without having the free Brownian motion concretely realized as operators on a Hilbert space in the following way:

$$\|A\| = \lim_{m\to\infty} \sqrt[2m]{\varphi((AA^*)^m)}.$$

3.3.3 Ito Formula and Multiplication of Multiple Wigner Integrals

The Ito formula shows up in the multiplication of two multiple Wigner integrals.

As first example, consider the product of two first order Wigner integrals.

$$\begin{aligned}\int f(t_1)\mathrm{d}S(t_1)\cdot\int g(t_2)\mathrm{d}S(t_2) &= \iint f(t_1)g(t_2)\mathrm{d}S(t_1)\mathrm{d}S(t_2) + \int f(t)g(t)\underbrace{\mathrm{d}S(t)\mathrm{d}S(t)}_{\mathrm{d}t}\\ &= \iint f(t_1)g(t_2)\mathrm{d}S(t_1)\mathrm{d}S(t_2) + \int f(t)g(t)\mathrm{d}t.\end{aligned}$$

We have invoked here the free Ito formula $\mathrm{d}S(t)\mathrm{d}S(t) = \mathrm{d}t$. This is the same as the Ito formula for classical Brownian motion $\mathrm{d}B(t)\mathrm{d}B(t) = \mathrm{d}t$. In order to see a difference between the classical and the free situation, we consider a more complicated case, the product of a first order with a second order integral.

$$\begin{aligned}\iint f(t_1,t_2)\mathrm{d}S(t_1)\mathrm{d}S(t_2)\cdot\int g(t_3)\mathrm{d}S(t_3) &= \iiint f(t_1,t_2)g(t_3)\mathrm{d}S(t_1)\mathrm{d}S(t_2)\mathrm{d}S(t_3)\\ &+ \iint f(t_1,t)g(t)\mathrm{d}S(t_1)\underbrace{\mathrm{d}S(t)\mathrm{d}S(t)}_{\mathrm{d}t} + \iint f(t,t_2)g(t)\underbrace{\mathrm{d}S(t)\mathrm{d}S(t_2)\mathrm{d}S(t)}_{\mathrm{d}t\varphi[\mathrm{d}S(t_2)]=0}.\end{aligned}$$

In the classical case, the last term would also contribute as $\mathrm{d}B(t)\mathrm{d}B(t_2)\mathrm{d}B(t) = \mathrm{d}B(t_2)\mathrm{d}B(t)\mathrm{d}B(t) = \mathrm{d}B(t_2)\mathrm{d}t$, whereas in the free case this term vanishes according to the general version of the free Ito formula

$$\mathrm{d}S(t)A\mathrm{d}S(t) = \varphi(A)\mathrm{d}t \quad \text{for } A \text{ adapted.}$$

(Adapted means here as in the classical case that A is only a function of the free Brownian motion up to time t.)

3.3.4 Contractions

For a systematic description of the product of two Wigner integrals it is useful to introduce the notion of contraction of functions. This is defined as follows. Consider $f \in L^2(\mathbb{R}_+^n)$ and $g \in L^2(\mathbb{R}_+^m)$. Then we define, for $0 \le p \le \min(n,m)$, the p-th contraction of f and g, $f \overset{p}{\frown} g \in L^2(\mathbb{R}_+^{n+m-2p})$, by

$$f \overset{p}{\frown} g(t_1, \dots, t_{m+n-2p}) :=$$
$$\int f(t_1, \dots, t_{n-p}, s_1, \dots, s_p) g(s_p, \dots, s_1, t_{n-p+1}, \dots, t_{n+m-2p}) ds_1 \cdots ds_p$$

With this definition we have the following general formula for the product of two Wigner integrals.

$$I(f) \cdot I(g) = \sum_{p=0}^{\min(n,m)} I(f \overset{p}{\frown} g) \qquad (f \in L^2(\mathbb{R}_+^n), g \in L^2(\mathbb{R}_+^m))$$

3.3.5 Free Chaos Decomposition

As for classical Brownian motion one has now the canonical isomorphism

$$L^2(\{S(t) \mid t \ge 0\}) \hat{=} \bigoplus_{n=1}^{\infty} L^2(\mathbb{R}_+^n), \qquad f \hat{=} \bigoplus_{n=0}^{\infty} f_n,$$

via

$$f = \sum_{n=0}^{\infty} I(f_n) = \sum_{n=0}^{\infty} \int \cdots \int f_n(t_1, \dots, t_n) \mathrm{d}S(t_1) \dots \mathrm{d}S(t_n).$$

This decomposition of the L^2-space of free Brownian motion as a direct sum of Wigner integrals is called *free chaos decomposition* and the set $\{I(f_n) \mid f_n \in L^2(\mathbb{R}_+^n)\}$ is called the *n-th free chaos*.

4 Distributions in Chaoses

4.1 Conjectures on Distributions in Fixed or Finite Chaoses

Since free semicircular elements $s_1, \dots, s_m$ can in this frame be realized as elements in the first chaos (namely, as stochastic integrals over characteristic functions of disjoint intervals of length 1), polynomials in our free semicircular elements, $p(s_1, \dots, s_m)$, are then, by the Ito formula, realized as elements from finite chaos

$$\{I(f) \mid f \in \bigoplus_{\text{finite}} L^2(\mathbb{R}_+^n)\}$$

Thus our conjectures on qualitative features of the distribution of $p(s_1, \dots, s_m)$ can now also be rephrased as follows.

The distribution of selfadjoint elements from a finite chaos with no constant ($n = 0$) contribution should have no atoms and they should have a density with respect to Lebesgue measure.

Motivated by the corresponding situation from classical chaos decomposition we conjecture also that selfadjoint variables $I(f)$ and $I(g)$ from different chaoses cannot have the same distribution.

It might not be clear to the reader why our original problem (on densities of polynomials in free semicircular elements) should be more accessible if reformulated in the language of stochastic integrals. As an answer to this we can only offer the observation that stochastic analysis offers some additional tools to attack this question. In particular, it is our hope that free Malliavin calculus, when sufficiently developed, should be able to answer our questions. It should be remarked that classical Malliavin calculus is indeed answering the classical analogue of our question – i.e., it allows to prove that selfadjoint elements from finite chaos of classical Brownian motion (like polynomials in independent normal variables) have a smooth density.

4.2 The Free Fourth Moment Theorem

Most of the above conjectures are still open, but we have at least some positive results which show that looking on elements from a fixed chaos can give some quite unexpected constraints. Namely, we have the following free analogue of a striking classical result of Nualart and Peccati [11]. There, f^*, for $f \in L^2(\mathbb{R}_+^n)$, is defined by

$$f^*(t_1, \dots, t_n) := \overline{f(t_n, \dots, t_1)}.$$

The condition $f^* = f$ just ensures that $I(f)$ is a selfadjoint operator.

Theorem 1 ([8]). *Consider, for fixed n, a sequence* $f_1, f_2, \dots \in L^2(\mathbb{R}_+^n)$ *with* $f_k^* = f_k$ *and* $\|f_k\|_2 = 1$ *for all* $k \in \mathbb{N}$*. Then the following statements are equivalent.*

1. *We have* $\lim_{k\to\infty} \varphi[I(f_k)^4] = 2$.
2. *We have for all* $p = 1, 2, \dots, n-1$ *that*

$$\lim_{k\to\infty} f_k \overset{p}{\frown} f_k = 0 \qquad \textit{in } L^2(\mathbb{R}_+^{2n-2p}).$$

3. *The selfadjoint variable $I(f_k)$ converges in distribution to a semicircular variable of variance 1.*

The striking direction of these equivalences is that (1) implies (3); i.e., if we constrain our variables to live in one fixed chaos then, under fixed second moment, the convergence of the fourth moment implies the convergence of all other moments. The second condition is the key for the proof of this. Convergence of the fourth moments implies the vanishing of the non-trivial contractions; but all moments can be calculated in terms of contractions and the vanishing of the non-trivial ones implies then that this calculations reproduces just the Catalan numbers, which are the moments of a semicircular variable.

As a consequence of the above theorem we get that we can at least distinguish elements from the first chaos (all of which are semicircular variables) from all the other chaoses.

Corollary 1 ([8]). *For $n \geq 2$ and $f \in L^2(\mathbb{R}_+^n)$, the law of $I(f)$ is not semicircular*

4.3 Quantitative Estimates

In the classical case one can refine the qualitative statement of convergence to a normal variable to quantitative estimates of the distance to a normal variable in terms of differences of the fourth moments. Such estimates exist for various distances (like Kolmogorov or Wasserstein distance). In the free case, those "classical" distances do not fit too well with the non-commutative nature of the problem (i.e., even though in the end one wants to compare just two single variable distributions, i.e., two probability distributions, the constraint that they live in some fixed free chaos puts them into a quite non-commutative context). The following distance seems to be more appropriate for our free setting.

4.3.1 Distance Between Operators

Given two self-adjoint random variables X, Y, we define their distance

$$d_{\mathscr{C}_2}(X,Y) := \sup\{|\varphi[h(X)] - \varphi[h(Y)]| \; : \; \mathscr{I}_2(h) \leq 1\};$$

where $\mathscr{I}_2(h)$ is formally the norm $\|\partial h'\|$ of the non-commutative derivative of the classical derivative of h; the non-commutative derivative ∂ is defined by

$$\partial X^n = \sum_{k=0}^{n-1} X^k \otimes X^{n-1-k}.$$

One way to make this rigorous is as follows. If h is the Fourier transform of a complex measure ν on $\mathbb{R}$,

$$h(x) = \hat{\nu}(x) = \int_{\mathbb{R}} e^{ix\xi}\, \nu(\mathrm{d}\xi)$$

then we define

$$\mathscr{I}_2(h) = \int_{\mathbb{R}} \xi^2\, |\nu|(\mathrm{d}\xi).$$

The above definitions are motivated by the fact that $\partial h'$ is essentially the Ito correction term showing up in the free Ito formula applied to $\mathrm{d}h(S_t)$.

4.3.2 The Free Gradient Operator

As in the classical setting we have a free version of a Malliavin calculus. The basic definitions and properties were given in [3]; however, the more advanced features of this theory still have to be developed.

Here is the definition of the free Malliavin gradient operator. Since it is related with the non-commutative derivative, it will, like ∂, also map into the tensor product.

Definition 3. For $f \in L^2(\mathbb{R}_+^n)$ we define the *free (Malliavan) gradient* $\nabla I(f)$ of $I(f)$ by

$$\nabla_t \left(\int f(t_1,\dots,t_n)\, \mathrm{d}S_{t_1} \cdots \mathrm{d}S_{t_n} \right) :=$$
$$\sum_{k=1}^{n} \int f(t_1,\dots,t_{k-1},t,t_{k+1},\dots,t_n) \mathrm{d}S_{t_1} \cdots \mathrm{d}S_{t_{k-1}} \otimes \mathrm{d}S_{t_{k+1}} \cdots \mathrm{d}S_{t_n}.$$

$\nabla_t I(f)$ is to be considered as a function in t, taking on values in the tensor product $L^2 \otimes L^2$.

The adjoint of ∇ is the *free Skorohod integral*, denoted by δ. It maps bi-processes, i.e., functions which take values in the tensor product into the L^2-space of our free Brownian motion.

4.3.3 Estimate of the Distance Against the Gradient

Let us denote by N the number operator, which acts just by multiplication by n in the n-the chaos, i.e.,

$$NI(f) = nI(f), \qquad \text{for } f \in L^2(\mathbb{R}_+^n).$$

Since one has, as in the classical case, $\delta\nabla = N$, one can derive the following estimate.

Theorem 2 ([8]). *Let $F = F^*$ be an element from a finite chaos with mean 0 and variance 1. Then we have the following estimate for the distance between F and a semicircular variable S of mean 0 and variance 1:*

$$d_{\mathscr{C}_2}(F, S) \leq \frac{1}{2}\varphi \otimes \varphi \left(\left| \int \nabla_s (N^{-1} F) \sharp (\nabla_s F)^* \, \mathrm{d}s - 1 \otimes 1 \right| \right).$$

The symbol $\sharp$ denotes here the following product in the tensor product:

$$(a_1 \otimes b_1) \sharp (a_2 \otimes b_2) := a_1 a_2 \otimes b_2 b_1.$$

In the classical case one can estimate the corresponding expression of the above gradient, for F living in some n-th chaos, in terms of the fourth moment of the considered variable, thus giving a quantitative estimate for the distance between the considered variable (from a fixed chaos) and a normal variable in terms of the difference between their fourth moments. In the free case such a general estimate does not seem to exist; at the moment we are only able to do this for elements F from the second chaos.

Corollary 2 ([8]). *Let $F = I(f) = I(f)^*$ ($f \in L^2(\mathbb{R}_+^2)$) be an element from the second chaos with variance 1, i.e., $\|f\|_2 = 1$, and let S be a semicircular variable with mean 0 and variance 1. Then we have*

$$d_{\mathscr{C}_2}(F, S) \leq \frac{1}{2}\sqrt{\frac{3}{2}}\sqrt{\varphi(F^4) - 2}.$$

4.3.4 A Multivariate Fourth Moment Theorem

In the long run, we are looking for multivariate versions of the above results. On the qualitative level, this extension is quite straightforward and does not require new ideas. Here we have the following result.

Theorem 3 ([10]). *Let, for each $i = 1, \ldots, d$, $(f_k^{(i)})_{k \in \mathbb{N}}$ be a sequence in $L^2(\mathbb{R}_+^{n_i})$, such that for all i, j*

$$\lim_{k \to \infty} \varphi[I(f_k^{(i)}) \cdot I(f_k^{(j)})] = \delta_{ij}.$$

The following statements are equivalent.

1. *$((I(f_k^{(1)}), \ldots, I(f_k^{(d)}))$ converges in distribution to a free semicircular family $(s_1, \ldots, s_d)$ as $k \to \infty$.*
2. *For each $i = 1, \ldots, d$, $I(f_k^{(i)})$ converges to s_i as $k \to \infty$.*

3. *For each* $i = 1, \ldots, d$, *the fourth moments of* $I(f_k^{(i)})$ *converge to 2:*

$$\lim_{k\to\infty} \varphi[I(f_k^{(i)})^4] = 2.$$

However, when it comes to quantitative results, nothing is known at the moment for the multivariate situation. In particular, we still don't know what a reasonable multivariate version for the distance $d_{\mathscr{C}_2}$ should be.

References

1. Akemann, G., Baik, J., Di Francesco, P. (eds.): The Oxford Handbook on Random Matrix Theory. Oxford University Press, New York (2011)
2. Belinschi, S.: The Lebuesgue decomposition of the free additive convolution of two probability distributions. Probab. Theory Relat. Fields **142**, 125–150 (2008)
3. Biane, P., Speicher, R.: Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. Probab. Theory Relat. Fields **112**, 373–409 (1998)
4. Bozejko, M.: A q-deformed probability, Nelson's inequality and central limit theorems. In: Garbecaki, P., Popowci, Z. (eds.) Nonlinear Fields, Classical, Random, Semiclassical, pp. 312–335. World Scientific, Singapore (1991)
5. Guionnet, A., Shlyakhtenko, D.: Free monotone transport (Preprint). arXiv:1204.2182
6. Haagerup, U., Thorbjørnsen, S.: A new application of random matrices: Ext($C^*_{\mathrm{red}}(F_2)$) is not a group. Ann. Math. **162**, 711–775 (2005)
7. Haagerup, U., Schultz, H., Thorbjornsen, S.: A random matrix approach to the lack of projections in $C^*_{\mathrm{ref}}(F_2)$. Adv. Math. **204** , 1–83 (2006)
8. Kemp, T., Nourdin, I., Peccati, G., Speicher, R.: Wigner chaos and the fourth moment. Ann. Probab. **40**(4), 1577–1635 (2012)
9. Nica, A., Speicher, R.: Lectures on the Combinatorics of Free Probability. London Mathematical Society Lecture Note Series, vol. 335, xvi+417pp. Cambridge University Press, Cambridge (2006)
10. Nourdin, I., Peccati, G., Speicher, R.: Multidimensional semicircular limits on the free Wigner chaos. In: Dalang, R., Dozzi, M., Russo, F. (eds.) Seminar on Stochastic Analysis, Random Fields and Applications VII. Progress in Probability, Vol. 67, pp. 211–221. Springer, Basel (2013). arXiv:1107:5135
11. Nualart, D., Peccati, G.: Central limit theorems for sequences of multiple stochastic integrals. Ann. Probab. **33**, 177–193 (2005)
12. Voiculescu, D.: Limit laws for random matrices and free products. Invent. Math. **104**, 267–271 (1991)
13. Voiculescu, D.: Free analysis questions. I. Duality transform for the coalgebra of $\partial_{X:B}$. Int. Math. Res. Not. **16**, 793–822 (2004)
14. Voiculescu, D., Dykema, K., Nica, A.: Free Random Variables. A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras and Harmonic Analysis on Free Groups. CRM Monograph Series, vol. 1, vi+70pp. American Mathematical Society, Providence (1992)
15. Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. **62**, 548–564 (1955)

Spacings: An Example for Universality in Random Matrix Theory

Thomas Kriecherbauer and Kristina Schubert

Abstract Universality of local eigenvalue statistics is one of the most striking phenomena of Random Matrix Theory, that also accounts for a lot of the attention that the field has attracted over the past 15 years. In this paper we focus on the empirical spacing distribution and its Kolmogorov distance from the universal limit. We describe new results, some analytical, some numerical, that are contained in Schubert K (2012) On the convergence of the nearest neighbour eigenvalue spacing distribution for orthogonal and symplectic ensembles. PhD thesis, Ruhr-Universität Bochum, Germany. A large part of the paper is devoted to explain basic definitions and facts of Random Matrix Theory, culminating in a sketch of the proof of a weak version of convergence for the empirical spacing distribution σ_N (see (23)).

1 Introduction

The roots of the theory of random matrices reach back more than a century. They can be found, for example, in the study of the Haar measure on classical groups [15] and in statistics [36]. The field experienced a first boost in the 1950s due to a remarkable idea of E. Wigner. He suggested to model the statistics of highly excited energy levels of heavy nuclei by the spectrum of random matrices. Arguably the most striking aspect of his investigations was how well the random eigenvalues described the distribution of spacings between neighbouring energy levels. Even more surprising were the subsequent discoveries that the eigenvalue spacing distributions are also relevant in a number of different areas of physics (e.g. as a signature for quantum chaos) and somewhat exotically also in number theory for the description of zeros of zeta functions (see [1, Chap. 2 and Part III] for recent reviews). Due to these developments Random Matrix Theory became

T. Kriecherbauer (✉) · K. Schubert
University of Bayreuth, 95440 Bayreuth, Germany
e-mail: thomas.kriecherbauer@uni-bayreuth.de; kristina.schubert@uni-bayreuth.de

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_3,

an active area of research that was prospering for many years mainly in the realm of physics. It was only about 15 years ago that random matrices started to attract broader interest also within mathematics stretching over a variety of different areas. The reason for this second boost was the discovery [3] that after appropriate rescaling the length of the longest increasing subsequence of a random permutation on N letters displays for large N the same fluctuations as the largest eigenvalue of a $N \times N$ random matrix (from a particular set of ensembles). Again it turned out that the distribution of the largest eigenvalue (in the limit $N \to \infty$) defines a fundamental distribution that comes up in a number of seemingly unrelated models of combinatorics and statistical mechanics (e.g. growth models, interacting particle systems; see [11] and [19] for recent reviews).

In summary, we have seen that the statistics of eigenvalues of random matrices display a certain degree of universality by describing the fluctuations in a varied list of stochastic, combinatorial and even deterministic (zeros of zeta functions) settings. In this paper, however, we will be concerned with a second aspect of universality that is known as the universality conjecture in random matrix theory. It states that in the limit of large matrix dimensions local eigenvalue statistics (see the beginning of Sect. 3 for an explanation of the meaning of this term) only depend on the symmetry class (cf. Sect. 2) of the matrix ensemble but not on other details of the probability measure. We will discuss this conjecture in the context of the nearest neighbour spacing distribution that has received much less attention in the literature than other statistical quantities such as k-point correlations or gap probabilities. We focus on the question of convergence of the empirical spacing distribution of eigenvalues.

Besides the standard monograph [25], a number of books have appeared recently [2, 5, 6, 12, 30], which present nice introductions into various aspects of Random Matrix Theory. An impressive collection of topics from Random Matrix Theory and its applications can be found in [1]. However, in all these books the information on the convergence of the empirical spacing distribution is somewhat sparse, except for [5] and [18] in the case of unitary ensembles ($\beta = 2$). It is one of the goals of this paper to give a concise and largely self-contained update of [5, 18] w.r.t. spacing distributions including also orthogonal ($\beta = 1$) and symplectic ($\beta = 4$) ensembles.

The paper is organised as follows. First, we introduce in Sect. 2 three important types of matrix ensembles that generalize the classical Gaussian ensembles. In order to define our prime object of study, the empirical spacing distribution (Sect. 3.2), we first discuss the spectral limiting density for all three types of ensembles in Sect. 3.1. From Sects. 4 to 6 we only treat invariant ensembles. We first state how k-point correlations are related to orthogonal polynomials and recall what is known about their convergence (Sect. 4). These results are used in Sect. 5 to sketch the proof of a first convergence result (23) for the empirical spacing distribution. Our main new result Theorem 4, that is proved in [27], is stated in Sect. 6 together with related results for circular ensembles from the literature. We close by mentioning numerical results of [27]. They indicate that a version of the Central Limit Theorem, similar to the one proved in [29] for COE and CUE, should also hold for the ensembles discussed in this paper.

2 Random Matrix Ensembles

Starting with the classical Gaussian ensembles we introduce in the present section three different types of generalisations, Wigner ensembles, Invariant ensembles, and β-ensembles. Together they constitute a large part of the ensembles studied in Random Matrix Theory. Some references are provided where the reader can learn about the main techniques to analyse these ensembles. The concept of symmetry classes is briefly discussed.

We begin by defining one of the most prominent matrix ensembles, the Gaussian Unitary Ensemble (GUE). GUE is a collection of probability measures on Hermitian $N \times N$ matrices X, $N \in \mathbb{N}$, where the diagonal entries x_{jj} and the real and imaginary parts of the upper triangular entries $x_{jk} = u_{jk} + iv_{jk}$, $j < k$ are all independent and normally distributed with $x_{jj} \sim \mathcal{N}(0, 1/\sqrt{N})$, $u_{jk}, v_{jk} \sim \mathcal{N}(0, 1/\sqrt{2N})$. GUE has the following useful properties.

1. The entries are independent as far as the Hermitian symmetry permits.
2. The probability measure is invariant under conjugation by matrices of the unitary group, i.e. under change of orthonormal bases. In fact, this explains why the ensemble is called “unitary” (and the reference to Gauss is due to the normal distribution). Moreover, one can compute the joint distribution of the eigenvalues explicitly. The vector of eigenvalues $(\lambda_1, \ldots, \lambda_N)$ is distributed on $\mathbb{R}^N$ with Lebesgue-density

$$Z_{N,\beta}^{-1} \prod_{j<k} |\lambda_k - \lambda_j|^\beta \prod_i e^{-\beta N \lambda_i^2/4} \, d\lambda_i, \quad \beta = 2, \tag{1}$$

where $Z_{N,\beta}$ denotes some norming constant.

Each of these properties comes with a set of techniques to analyse statistics of eigenvalues. In turn, these techniques can be applied to a large number of matrix ensembles that share this particular property. More precisely:

1. *Wigner ensembles* have independent entries as far as the symmetry of the matrix permits. The distributions of the entries do not need to be normal or identical, but must satisfy some conditions on the moments. Except for the Gaussian case, Wigner ensembles are not unitarily invariant and the joint distribution of eigenvalues is generally not known. Many results for such ensembles (e.g. Wigner semi circle law, distribution of the largest eigenvalue) can be obtained via the *method of moments*, i.e. by analysing the moments of the empirical measure of the eigenvalues (see e.g. [14], cf. Sect. 3). More recently, very powerful new techniques have been introduced by Erdös et al. and independently by Tao and Vu (see e.g. [10, 31] and references therein).
2. *Invariant ensembles* keep the property of invariance under conjugation by unitary matrices. The ensembles considered in this class all have in common that the joint distribution of eigenvalues is given by a measure of the form

$$Z_{N,\beta}^{-1} \prod_{j<k} |\lambda_k - \lambda_j|^\beta \prod_i d\mu_N(\lambda_i), \quad \beta = 2 \tag{2}$$

where μ_N denotes some (positive) finite measure on $\mathbb{R}$ with sufficient decay at infinity to guarantee finiteness of the measure on $\mathbb{R}^N$. As we explain in Sect. 4 it is exactly the structure of (2), i.e. a product measure with dependencies introduced by the square of the Vandermonde determinant, for which the *method of orthogonal polynomials* can be applied. Note that such measures, with μ_N supported on discrete sets, were also central for proving the appearance of local eigenvalue statistics in some of the models from statistical mechanics described in the Introduction (see e.g. [19] for an elementary exposition in the case of interacting particle systems).

Using these two types of generalisations of GUE we may already generate a great number of matrix ensembles. These consist of Hermitian matrices only and we say that they belong to the same symmetry class. By the universality conjecture we expect that in the limit $N \to \infty$ all these ensembles display the same local spectral statistics.

If one replaces in the definition of GUE above the Hermitian matrices by real symmetric resp. by quaternion self-dual matrices, keeping the independence of the entries as well as their normal distributions (with appropriately chosen variances), one obtains the Gaussian Orthogonal Ensemble (GOE) resp. the Gaussian Symplectic Ensemble (GSE). These ensembles can be generalised as above, yielding again Wigner ensembles or invariant ensembles and the only difference compared to the discussion above is that in (1) and (2) we have to choose $\beta = 1$ resp. $\beta = 4$. In this way we have introduced two more symmetry classes which then together constitute all classes from Dyson's threefold way. As it was discovered some 30 years after Dyson's classification result from 1962, it is useful and natural to enlarge this list to a grand total of 10 symmetry classes, thus providing a significant increase in applications of Random Matrix Theory in physics, statistics and mathematics alike (see [37] for a recent survey). It should be noted that from the perspective of invariant matrix ensembles the resulting joint distributions of the eigenvalues are of the form (2) with $\beta \in \{1, 2, 4\}$ for all ten symmetry classes. As we will argue in Sect. 6 there exists a large class of invariant ensembles from all ten symmetry classes for which the localised and appropriately rescaled empirical spacing distributions (see Sect. 3) converge to universal limits that only depend on the value of β. The *method of orthogonal polynomials* mentioned above can also be applied for $\beta = 1, 4$. However, it is more technical and its range of applicability is less general than in the case $\beta = 2$, see e.g. [6].

There is a third property of Gaussian ensembles that leads to a different type of generalisation. The basic observation is the following. If one applies the Householder transformation to GOE in a suitable way, one obtains probability measures on $N \times N$ Jacobi matrices (i.e. real symmetric, tridiagonal matrices with positive off-diagonal entries). By construction they induce the same joint distributions of eigenvalues as (1) with $\beta = 1$. For this ensemble, the entries are again independent (as symmetry permits) with normal distributions on the diagonal and some χ-distributions for the off-diagonal entries. General *β-ensembles* are now generated by modifying the variances of the χ-distributions on the off-diagonals. For any $\beta > 0$ this can be done in such a way that the joint distribution of eigenvalues is given by (1) with the prescribed value of β. A key insight into the analysis of these ensembles is that for large matrix dimensions eigenvalues of the Jacobi matrices may be approximated by the spectrum of a specific stochastic Schrödinger operator, see e.g. [26]. Note that the local eigenvalue statistics of β-ensembles are different for each value of β. Obviously, they reduce to the classical Gaussian ensembles if and only if $\beta \in \{1, 2, 4\}$.

3 The Empirical Spacing Distribution: Localised and Rescaled

In this section we define the empirical spacing distribution as one prime example for local eigenvalue statistics. By the latter we mean, firstly, that the spectrum is localised by considering only some part of the spectrum and, secondly, that the spectrum is being rescaled such that the average distance between neighbouring eigenvalues is constant and of order 1 in the considered spectral region. In order to perform such operations we must first understand the *limiting spectral density* of the ensemble.

3.1 The Limiting Spectral Density

We denote the ordered eigenvalues of a matrix H from one of the ensembles described in Sect. 2 by $\lambda_1^{(N)}(H) \leq \lambda_2^{(N)}(H) \leq \ldots \leq \lambda_N^{(N)}(H)$. The corresponding N-tuple $\lambda^{(N)}(H)$ thus defines a point in the Weyl chamber that we denote by $\mathscr{W}_N := \{x \in \mathbb{R}^N : x_1 \leq \ldots \leq x_N\}$. Moreover, we abbreviate $\lambda_j^{(N)}(H)$ by λ_j from now on to keep the notation manageable.

We associate to each $\lambda \in \mathscr{W}_N$ its counting measure $\delta_\lambda := \frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j}$ which defines a probability measure on $\mathbb{R}$. By the *limiting spectral density* we mean a function $\psi : \mathbb{R} \to [0, \infty)$ satisfying for all $s \in \mathbb{R}$ that

$$\mathbb{E}_{N,\beta}\left(\int_{-\infty}^{s} d\delta_\lambda\right) \to \int_{-\infty}^{s} \psi(t)\,dt \quad \text{as } N \to \infty\,.$$

It is known for ample classes of both Wigner ensembles and invariant ensembles as well as for β-ensembles that the spectral density exists. For Wigner ensembles one can show mainly by combinatorial methods that on average the moments of δ_λ converge to the moments of the semi-circle distribution (Wigner semi-circle law).

The first steps of the proof are provided by the simple observation that for $k \in \mathbb{N}$ one has

$$\mathbb{E}_{N,\beta}\left(\int_{\mathbb{R}} t^k \, d\delta_\lambda(t)\right) = \frac{1}{N}\mathbb{E}_{N,\beta}(\mathrm{tr}\,(H^k))$$

together with an expansion of the right hand side as a sum of expectations of products of entries of H that can be simplified by using the independence of the entries (method of moments, see e.g. [14]).

Next we turn to β-ensembles. Here the limiting spectral density is again given by the Wigner semi-circle law. The proof, however, follows a different path. Recall that the joint distribution of eigenvalues is given by (1). Its density can therefore be rewritten in the form

$$Z_{N,\beta}^{-1} \exp\left[-\beta N^2 I(\delta_\lambda)\right] \quad \text{with} \quad I(\nu) := -\frac{1}{2}\int_{x \neq y} \log|x-y| d\nu(x) d\nu(y) + \int \frac{x^2}{4} d\nu(x)\,. \tag{3}$$

We may think of I as a functional defined on all probability measures on $\mathbb{R}$. It is a well known fact in logarithmic potential theory that I has a unique minimizer that is given by the semi-circle law. Since we have the factor N^2 in the exponent in (3) it is intuitively clear that for large N only those vectors λ will be relevant for which the corresponding counting measure δ_λ is close to the minimizer of I. This idea can be used to prove the Wigner semi-circle law for β-ensembles. Moreover, this idea can also be applied to prove the existence of the limiting spectral density for a large class of invariant ensembles (see e.g. [17]). Indeed, let us assume that in (2) the measure $d\mu_N$ has a Lebesgue-density of the form

$$d\mu_N(x) = e^{-NV(x)} dx \quad \text{satisfying} \quad \lim_{|x|\to\infty} \frac{V(x)}{\log|x|} = \infty\,, \tag{4}$$

in order to guarantee that the measure (2) is finite. Under mild regularity assumptions on V one can proof that the functional

$$I_V(\nu) := -\frac{1}{2}\int_{x\neq y} \log|x-y| d\nu(x) d\nu(y) + \int V(x) d\nu(x)\,, \tag{5}$$

defined on the probability measures on $\mathbb{R}$ has an unique minimizer ν_V with a Lebesgue-density $\psi = \psi_V$. As argued above, one can show that ψ is the limiting spectral density of the ensemble (see e.g. [5, Chap. 6] for an elementary exposition). In the literature on invariant ensembles one also finds a slightly more general setting where in the formula (4) for the density of $d\mu_N$ the function V is replaced by N-dependent functions V_N that converge to some function V satisfying the growth condition (4).

Note, that for invariant ensembles the limiting spectral density depends on V and is therefore not an universal quantity. This is not a contradiction to the universality conjecture of Dyson since the limiting spectral density is a global quantity whereas the universality conjecture only refers to local eigenvalue statistics.

3.2 The Empirical Spacing Distribution

We use the limiting spectral density in order to rescale the eigenvalues. Let a denote a point in the interior of the support of ψ where the limiting density is positive, i.e. $\psi(a) > 0$. We assume further that a is a point of continuity for ψ. For eigenvalues λ_i that are close to a the expected distance of neighbouring eigenvalues is given to leading order by $(N\psi(a))^{-1}$. Therefore we introduce the rescaled and centred eigenvalues

$$\tilde{\lambda}_i := (\lambda_i - a)N\psi(a). \tag{6}$$

Considering only eigenvalues λ_i that lie in an (N-dependent) interval I_N that is centred at a and has vanishing length $|I_N| \to 0$ for $N \to \infty$, we expect that their rescaled versions $\tilde{\lambda}_i$ have a spacing that is close to 1 on average. We introduce

$$A_N := N\psi(a)(I_N - a) = \{N\psi(a)(t-a) \mid t \in I_N\}$$

and observe that $\lambda_i \in I_N$ if and only if $\tilde{\lambda}_i \in A_N$. Therefore and by the expected unit spacing of the rescaled eigenvalues we conclude that the length of A_N gives the average of the number of eigenvalues λ_i that lie in I_N to leading order. For our considerations we assume that this number and hence $N|I_N| = |A_N|/\psi(a)$ tends to infinity for $N \to \infty$. We summarize our assumptions on the length of I_N.

$$|I_N| \to 0\,, \qquad N|I_N| \to \infty \qquad \text{for } N \to \infty. \tag{7}$$

Finally, we define our main object of interest, the empirical spacing distribution. As above we denote the eigenvalues of a random matrix H by $\lambda_1 \leq \ldots \leq \lambda_N$ and their rescaled versions (6) by $\tilde{\lambda}_1 \leq \ldots \leq \tilde{\lambda}_N$. Furthermore, let I_N be an interval centred at a and satisfying (7). Then the empirical spacing distribution for H, localised in I_N, is given by

$$\sigma_N(H) := \frac{1}{|A_N|} \sum_{\lambda_{i+1}, \lambda_i \in I_N} \delta_{\tilde{\lambda}_{i+1} - \tilde{\lambda}_i}. \tag{8}$$

Recall from the discussion above that the expected number of spacings considered in $\sigma_N(H)$ is given by $|A_N| - 1$. This explains the pre-factor $1/|A_N|$ in the definition of $\sigma_N(H)$, which is asymptotically the same as $1/(|A_N| - 1)$.

4 Universality of the k-Point Correlation Functions for Invariant Ensembles

In this section we state results on the convergence of k-point correlation functions for invariant ensembles, as well as their connection to orthogonal polynomials.

We recall that we consider invariant ensembles where the joint distribution of the eigenvalues has a density of the form (see (2) and (4))

$$P_N^{(\beta)}(\lambda_1, \dots, \lambda_N) := \frac{1}{Z_{N,\beta}} \prod_{i<j} |\lambda_j - \lambda_i|^\beta \prod_{k=1}^N w_N^{(\beta)}(\lambda_k), \quad \lambda \in \mathbb{R}^N \tag{9}$$

with $w_N^{(\beta)}(x) = e^{-NV(x)}$. In the proof of the main theorem (Theorem 4) we will use asymptotic results for the marginal densities of $P_N^{(\beta)}$ with respect to k variables. The latter are called the k-point correlation functions, for which we will now give a precise definition.

Definition 1.

(i) For $k \in \mathbb{N}, k \leq N, \beta \in \{1, 2, 4\}$ and $(\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ we set

$$R_{N,k}^{(\beta)}(\lambda_1, \dots, \lambda_k) := \frac{N!}{(N-k)!} \int_{\mathbb{R}^{N-k}} P_N^{(\beta)}(\lambda_1, \dots, \lambda_N)\, d\lambda_{k+1} \dots d\lambda_N.$$

(ii) For $k \in \mathbb{N}, k \leq N$ and $\beta \in \{1, 2, 4\}$ the rescaled k-point correlation functions are given by

$$B_{N,k}^{(\beta)}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_k) := (N\psi(a))^{-k}\, R_{N,k}^{(\beta)}\left(a + \frac{\tilde{\lambda}_1}{N\psi(a)}, \dots, a + \frac{\tilde{\lambda}_k}{N\psi(a)}\right)$$
$$= (N\psi(a))^{-k}\, R_{N,k}^{(\beta)}\,(\lambda_1, \dots, \lambda_k)\,.$$

We observe that $R_{N,k}^{(k)}(t_1, \dots, t_k)$ and $B_{N,k}^{(\beta)}(t_1, \dots, t_k)$ are invariant under permutations of the indices $\{1, \dots, k\}$.

We now sketch how the k-point correlation functions can be analysed using the method of orthogonal polynomials. We start with the simplest case $\beta = 2$. Define $K_{N,2}: \mathbb{R}^2 \to \mathbb{R}$ with

$$K_{N,2}(x, y) := \sum_{j=0}^{N-1} \varphi_j^{(N)}(x)\varphi_j^{(N)}(y), \tag{10}$$

$$\varphi_j^{(N)}(x) := p_j^{(N)}(x)\sqrt{w_N^{(2)}(x)},$$

and $p_j^{(N)}(x) = \gamma_j^{(N)} x^j + \ldots$ with $\gamma_j^{(N)} > 0$ denotes the j-th normalised orthogonal polynomial with respect to the measure $w_N^{(2)}(x)dx$ on $\mathbb{R}$, i.e.

$$\int_{\mathbb{R}} p_j^{(N)}(x) p_k^{(N)}(x) w_N^{(2)}(x) dx = \delta_{jk}.$$

The convergence of the appropriately rescaled kernel $K_{N,2}$

$$\lim_{N\to\infty} \frac{1}{N\psi(a)} K_{N,2}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) = \frac{\sin(\pi(x-y))}{\pi(x-y)} =: K_2(x,y) \tag{11}$$

has by now been proved in quite some generality (see e.g. [23] and references therein). Usually uniform convergence of (11) is only shown for x, y in bounded sets. For our purposes it is convenient to extend this result for x, y in the growing set A_N.

Theorem 1 (cf. [8, 27]). *Let $V: \mathbb{R} \to \mathbb{R}$ be real analytic such that (4) holds and let V be regular in the sense of [8, (1.12),(1.13)]. Moreover, we assume $a \in \mathbb{R}$ with $\psi(a) > 0$ (ψ being defined as the density of the minimizer of I_V, see (5)). Let $(c_N)_{N\in\mathbb{N}}$ be a sequence satisfying $c_N \to \infty$, $\frac{c_N}{N} \to 0$ as $N \to \infty$. Then we have for $N \to \infty$*

$$\sup_{x,y\in[-c_N,c_N]} \left| \frac{1}{N\psi(a)} K_{N,2}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - K_2(x,y) \right| = \mathcal{O}\left(\frac{c_N}{N}\right) \tag{12}$$

$$\sup_{x,y\in[-c_N,c_N]} \left| \frac{\partial}{\partial x} \left(\frac{1}{N\psi(a)} K_{N,2}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - K_2(x,y) \right) \right| = \mathcal{O}\left(\frac{c_N}{N}\right). \tag{13}$$

Remark 1.

(i) The estimate in (13) will be needed to treat the cases $\beta = 1$ and $\beta = 4$.
(ii) The proof of Theorem 1 is essentially contained in [8] although not stated explicitly (a formula somewhat close is presented in [8, (6.18)]). In particular, there is no information on the derivatives in (13). Nevertheless the underlying Riemann-Hilbert analysis also provides (12) and (13), where we use an efficient path, which we have taken from [34]. A sketch of the required refinements and the extension to unbounded sets can be found in [27].
(iii) To unify the notation with the cases $\beta = 1$ and $\beta = 4$ treated below we set

$$\hat{K}_{N,2}(x,y) := \frac{1}{N\psi(a)} K_{N,2}(x,y) \tag{14}$$

and hence (12) reads

$$\hat{K}_{N,2}\left(a+\frac{x}{N\psi(a)},a+\frac{y}{N\psi(a)}\right)=K_2(x,y)+\mathscr{O}\left(\kappa_N\right) \tag{15}$$

with $\kappa_N = \frac{c_N}{N} \to 0$ as $N \to \infty$. The error term is uniform for $x, y \in [-c_N, c_N]$.

Theorem 1 can be used to derive some results about the rescaled correlation function, where one uses a well known determinantal formula expressing $B_{N,k}^{(2)}$ in terms of $K_{N,2}$ (see e.g. [5] and Lemma 1 below). Observe that in the considered setting the term $\mathscr{O}\left(\frac{c_N}{N}\right)$ in the asymptotic behaviour of $K_{N,2}$ (see Theorem 1) is replaced by $\mathscr{O}(|I_N|)$ in statement (*iii*) of Lemma 1.

Lemma 1. *Let the assumptions of Theorem 1 be satisfied. Furthermore, let a, ψ, I_N, A_N be defined as in Sect. 3. Then the following holds*

(i) For $t_1, \ldots, t_k \in \mathbb{R}$ we have

$$B_{N,k}^{(2)}(t_1,\ldots,t_k)=(N\psi(a))^{-k}\det\left(K_{N,2}\left(a+\frac{t_i}{N\psi(a)},a+\frac{t_j}{N\psi(a)}\right)\right)_{1\le i,j\le k},$$

where $K_{N,2}$ is given in (10).

(ii) For N sufficiently large we have for all $k \le N$

$$|B_{N,k}^{(2)}(t_1,\ldots,t_k)| \le 2^k \quad \text{for } t_1,\ldots,t_k \in A_N.$$

(iii) For $t_1, \ldots, t_k \in A_N$ we have

$$B_{N,k}^{(2)}(t_1,\ldots,t_k)=W_k^{(2)}(t_1,\ldots,t_k)+k!\cdot k\cdot 2^k\mathscr{O}(|I_N|), \tag{16}$$

with

$$W_k^{(2)}(t_1,\ldots,t_k):=\det\left(\frac{\sin(\pi(t_i-t_j))}{\pi(t_i-t_j)}\right)_{1\le i,j\le k} \tag{17}$$

and the constant implicit in the error term in (16) is uniform in k, $t_1, \ldots, t_k$ and in N.

It should be noted that with Lemma 1 we have derived all information on the convergence of the k-point correlation functions that is needed to prove the main result Theorem 4 for $\beta = 2$.

We now turn to the cases $\beta = 1, 4$. For technical reasons we restrict the discussion of the case $\beta = 1$ to even values of N. Our presentation follows closely the monograph [6]. Similar to statement (i) of Lemma 1 the k-point correlation function for $\beta = 1$ and $\beta = 4$ can be represented in terms of functions $S_{N,\beta}$, which are related to $K_{N,2}$. It is convenient to express the correlation functions in terms of the Pfaffian. We remind the reader that for real skew-symmetric $2m \times 2m$ matrices the determinant is a perfect square. Consequently, the Pfaffian which is defined to be the square-root of the determinant for such matrices can be expressed as a polynomial in the entries. Indeed,

$$\mathrm{Pf}(A) = \frac{1}{2^m m!} \sum_{\sigma \in S_{2m}} (\mathrm{sgn}\sigma) a_{\sigma_1 \sigma_2} a_{\sigma_3 \sigma_4} \dots a_{\sigma_{2m-1} \sigma_{2m}},$$

where S_{2m} denotes the permutation group on $\{1, \dots, 2m\}$. See also [6] for an elementary exposition on the use of Pfaffians in Random Matrix Theory. According to [6, (4.128),(4.135)] the correlation functions can be expressed via

$$R_{N,k}^{(\beta)}(\lambda_1, \dots, \lambda_k) = \mathrm{Pf}(K\,J), \quad \text{with } K := (K_{N,\beta}(\lambda_i, \lambda_j))_{i,j=1,\dots,k} \tag{18}$$

and

$$J := \mathrm{diag}(\sigma, \dots, \sigma) \in \mathbb{R}^{2N \times 2N}, \quad \sigma := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In (18) the terms $K_{N,\beta}(x, y), \beta = 1, 4$ denote 2×2 matrices with

$$K_{N,4}(x, y) := \begin{pmatrix} S_{N,4}(x, y) & \frac{\partial}{\partial y} S_{N,4}(x, y) \\ -\int_x^y S_{N,4}(t, y)\,dt & S_{N,4}(y, x) \end{pmatrix}$$

and

$$K_{N,1}(x, y) := \begin{pmatrix} S_{N,1}(x, y) & \frac{\partial}{\partial y} S_{N,1}(x, y) \\ -\int_x^y S_{N,1}(t, y)\,dt - \frac{1}{2}\,\mathrm{sgn}(x - y) & S_{N,1}(y, x) \end{pmatrix}.$$

The convergence of the (rescaled) matrix kernels $K_{N,\beta}$ is e.g. considered in [6], but as in the case $\beta = 2$ the known results only apply to the convergence on compact sets and need to be refined to uniform convergence on A_N (recall $|A_N| \to \infty$ as $N \to \infty$). Before we can state Theorem 2 we introduce some more notation (in analogy to (14) for $\beta = 2$). For $\beta = 1, 4$ let $\widehat{K_{N,\beta}}(x, y) \in \mathbb{R}^{2\times 2}$ denote a rescaled version of $K_{N,\beta}(x, y)$ given by

$$\widehat{K_{N,\beta}}(x, y) := \frac{1}{N\psi(a)} \begin{pmatrix} \frac{1}{\sqrt{N\psi(a)}} & 0 \\ & \sqrt{N\psi(a)} \end{pmatrix} K_{N,\beta}(x, y) \begin{pmatrix} \sqrt{N\psi(a)} & 0 \\ 0 & \frac{1}{\sqrt{N\psi(a)}} \end{pmatrix}. \tag{19}$$

We denote the components of the rescaled matrices $\widehat{K_{N,\beta}}(x,y)$ by

$$\begin{pmatrix} \widehat{S_{N,\beta}}(x,y) & \widehat{D_{N,\beta}}(x,y) \\ \widehat{I_{N,\beta}}(x,y) & \widehat{S_{N,\beta}}(y,x) \end{pmatrix} := \widehat{K_{N,\beta}}(x,y).$$

Theorem 2 ([6, 27, 28]). *Let V be a polynomial of even degree with positive leading coefficient and let V be regular in the sense of [8, (1.12),(1.13)]. Moreover, we assume $a \in \mathbb{R}$ with $\psi(a) > 0$ (ψ is defined as in Theorem 1) and let K_2 be given in (11). Let $(c_N)_{N\in\mathbb{N}}$ be a sequence satisfying $c_N \to \infty$, $\frac{c_N}{\sqrt{N}} \to 0$ as $N \to \infty$. Then we have*

(i) For $\beta = 1$ and N even

$$\sup_{x,y\in[-c_N,c_N]} \left| \widehat{S_{N,1}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - K_2(x,y) \right| = \mathscr{O}\left(\frac{1}{\sqrt{N}}\right)$$

$$\sup_{x,y\in[-c_N,c_N]} \left| \widehat{D_{N,1}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - \frac{\partial}{\partial x} K_2(x,y) \right| = \mathscr{O}\left(\frac{1}{\sqrt{N}}\right)$$

$$\sup_{x,y\in[-c_N,c_N]} \left| \widehat{I_{N,1}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - \int_0^{x-y} K_2(t,0)dt - \frac{1}{2}\operatorname{sgn}(x-y) \right| = \mathscr{O}\left(\frac{c_N}{\sqrt{N}}\right)$$

(ii) For $\beta = 4$ and N even

$$\sup_{x,y\in[-c_N,c_N]} \left| \widehat{S_{N/2,4}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - K_2(2(x-y)) \right| = \mathscr{O}\left(\frac{1}{\sqrt{N}}\right)$$

$$\sup_{x,y\in[-c_N,c_N]} \left| \widehat{D_{N/2,4}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - \frac{\partial}{\partial x} K_2(2(x-y)) \right| = \mathscr{O}\left(\frac{1}{\sqrt{N}}\right)$$

$$\sup_{x,y\in[-c_N,c_N]} \left| \widehat{I_{N/2,4}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) - \int_0^{x-y} K_2(2t)dt \right| = \mathscr{O}\left(\frac{c_N}{\sqrt{N}}\right)$$

Remark 2. The proof of Theorem 2 can be derived from [6, 28] and Theorem 1 as follows (details will be given in a later publication): We use the notation $\hat{x} = a + \frac{x}{N\psi(a)}$, $\hat{y} = a + \frac{y}{N\psi(a)}$ and set

$$\Delta_{N,\beta}(\hat{x},\hat{y}) := \frac{1}{N\psi(a)}\left(S_{N,\beta}(\hat{x},\hat{y}) - K_{N,2}(\hat{x},\hat{y})\right) = \hat{S}_{N,\beta}(\hat{x},\hat{y}) - \frac{1}{N\psi(a)} K_{N,2}(\hat{x},\hat{y}).$$

As V is a polynomial, we can apply Widom's formalism [35] to derive a representation of $\Delta_{N,\beta}$ in terms of orthogonal polynomials. Together with the estimates contained in [28] and [6, Sect. 6.3.1] (generalised to the case of varying weights) we obtain

$$\sup_{x,y\in[-c_N,c_N]} \left|\Delta_{N,\beta}(\hat{x},\hat{y})\right| = \mathscr{O}\left(N^{-\frac{1}{2}}\right)$$

$$\sup_{x,y\in[-c_N,c_N]} \left|\frac{1}{N\psi(a)}\frac{\partial}{\partial\hat{y}}\Delta_{N,\beta}(\hat{x},\hat{y})\right| = \mathscr{O}\left(N^{-\frac{1}{2}}\right).$$

The claim of Theorem 2 then follows from Theorem 1 and from the assumption $\frac{c_N}{\sqrt{N}} \to 0$ for $N \to \infty$, which implies $\frac{c_N}{N} = \mathscr{O}\left(\frac{1}{\sqrt{N}}\right)$.

Finally, we introduce some more notation (recall that K_2 was introduced in (11)):

$$S_1(x,y) := K_2(x,y), \quad D_1(x,y) := \frac{\partial}{\partial x}K_2(x,y),$$
$$I_1(x,y) := \int_0^{x-y} K_2(t,0)dt - \frac{1}{2}\ \mathrm{sgn}(x-y)$$

$$S_4(x,y) := K_2(2x,2y), \quad D_4(x,y) := \frac{\partial}{\partial x}K_2(2x,2y), \quad I_4(x,y) := \int_0^{x-y} K_2(2t,0)dt$$

and

$$K_\beta(x,y) := \begin{pmatrix} S_\beta(x,y) & D_\beta(x,y) \\ I_\beta(x,y) & S_\beta(y,x) \end{pmatrix}.$$

Remark 3. (i) With this notation the result of Theorem 2 reads: There exists a sequence κ_N such that $\kappa_N \to 0$ for $N \to \infty$ and

$$\widehat{K_{N,\beta}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) = K_\beta(x,y) + \mathscr{O}(\kappa_N) \qquad (20)$$

uniformly for $x, y \in A_N$.

(ii) Theorems 1 and 2 have been stated for invariant matrix ensembles satisfying (2) and (4) and do not cover all ten symmetry classes (cf. Sect. 2). However, the statements of Theorems 1 and 2 hold mutatis mutandis for all invariant ensembles for which universality has been proved using a Riemann Hilbert analysis in the analytic setting (see e.g. [9, 21, 34] for varying and non-varying Laguerre-type ensembles, [20] for Jacobi-type ensembles and [7] for non-varying Hermite-type ensembles). In this way all symmetry classes are covered. The work of McLaughlin and Miller [24] shows that one can expect that some finite regularity assumption on V combined e.g. with the convexity of V should also suffice.

From (20) one can deduce the analogue of Lemma 1 for $\beta = 1, 4$ using e.g. the formulae in [32]. In particular, one can derive the convergence of the rescaled correlation functions $B_{N,k}^{(\beta)}$. For $\beta = 1, 4$ we set (analogue to (17) for $\beta = 2$, see also (18))

$$W_k^{(\beta)}(t_1, \ldots, t_k) := \operatorname{Pf}(K\,J) \quad \text{with } K := (K_\beta(t_i, t_j))_{1 \le i,j \le k}, \quad t_1, \ldots, t_k \in \mathbb{R}. \tag{21}$$

Lemma 2. *Suppose that the assumptions of Theorem 2 hold. Then the following holds for $\beta \in \{1, 4\}$.*

(i) There exists $C > 0$ such that for all $1 \le k \le N$, $t_1, \ldots, t_k \in A_N$ we have

$$B_{N,k}^{(\beta)}(t_1, \ldots, t_k) = W_k^{(\beta)}(t_1, \ldots, t_k) + k! \cdot C^k\, \mathcal{O}(\kappa_N). \tag{22}$$

The constant implicit in the $\mathcal{O}$-term is uniform in k, N and in $t_1, \ldots, t_k$ and $\kappa_N \to 0$ as $N \to \infty$ as in Remark 3 (i).

(ii) The function $W_k^{(\beta)}$ is a symmetric function on $\mathbb{R}^k$ for all $k \in \mathbb{N}$.

(iii) For $k \in \mathbb{N}$, $t_1, \ldots, t_k$ and $c \in \mathbb{R}$: $\quad W_k^{(\beta)}(t_1+c, \ldots, t_k+c) = W_k^{(\beta)}(t_1, \ldots, t_k)$.

(iv) There exists a constant $C > 1$ such that for all $1 \le k \le N$ we have

$$\left| B_{N,k}^{(\beta)}(t_1, \ldots, t_k) \right| \le C^k\, k^{\frac{k}{2}} \quad \text{for } t_1, \ldots, t_k \in A_N$$
$$\left| W_k^{(\beta)}(t_1, \ldots, t_k) \right| \le C^k\, k^{\frac{k}{2}} \quad \text{for } t_1, \ldots, t_k \in \mathbb{R}.$$

(v) For all $t \in \mathbb{R}$: $\quad W_1^{(\beta)}(t) = 1$.

Remark 4. We note that for the results presented in Sect. 5.1 it is not necessary to keep track of the k-dependence of the error in (22). However, this estimate is needed in the proof of Theorem 4.

5 The Expected Empirical Spacing Distribution and Gap Probabilities

The basic result that we want to explain in this section is the convergence of the expected spacing distribution, i.e.

$$\lim_{N \to \infty} \mathbb{E}_{N,\beta} \left(\int_0^s d\sigma_N(H) \right) = \int_0^s d\mu_\beta \tag{23}$$

for some probability measures μ_β. The limiting spacing distributions μ_β depend on β, but are universal otherwise (see Remark 5 at the end of this section). In our exposition we restrict ourselves to prove the convergence of $\mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right)$ for $N \to \infty$. This is the content of Sect. 5.1. It is not entirely obvious to show that the limit actually defines a probability measure. One way to prove this is to make a connection between $\mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right)$ and the gap probabilities and to use that the latter can be expressed in terms of Painlevé V transcendents. We will discuss this connection in Sect. 5.2.

5.1 Convergence of the Expected Empirical Spacing Distribution

In this section we will show the existence of

$$\lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right)$$

and derive a representation for this limit. As $\int_0^s d\sigma_N(H)$ is a function of the ordered eigenvalues of H the expectation is obtained by integration over the Weyl chamber with respect to $B_{N,N}^{(\beta)}(t)dt$ (see Definition 1).

The first step in the proof is the introduction of related counting measures $\gamma_N(k,H)$ for $k \geq 2$. Recall that the eigenvalues of the random matrix H are denoted by $\lambda_1 \leq \ldots \leq \lambda_N$ and their rescaled versions by $\tilde{\lambda}_1 \leq \ldots \leq \tilde{\lambda}_N$ (see (6)). We define

$$\gamma_N(k,H) := \frac{1}{|A_N|} \sum_{\substack{i_1<\ldots<i_k,\\ \lambda_{i_1},\lambda_{i_k}\in I_N}} \delta_{(\tilde{\lambda}_{i_k}-\tilde{\lambda}_{i_1})}, \quad k \geq 2. \tag{24}$$

Observe that the normalizing factor $\frac{1}{|A_N|}$ corresponds to the fact that we expect $|A_N| = N\psi(a)|I_N|$ eigenvalues $\lambda_i \in I_N$ (see discussion below (8)). The measures $\gamma_N(k,H)$ are related to σ_N (see Lemma 3 below) and the main advantage of $\int_0^s d\gamma_N(k,H)$ over $\int_0^s d\sigma_N(H)$ is that it is a symmetric function of the eigenvalues of H, if we replace $\widetilde{\lambda_{i_k}} - \widetilde{\lambda_{i_1}}$ in (24) by $\max_{1\leq j\leq k}\widetilde{\lambda_{i_j}} - \min_{1\leq j\leq k}\widetilde{\lambda_{i_j}}$. This allows us to calculate the expectation of $\int_0^s d\gamma_N(k,H)$ by integration over $\mathbb{R}^N$ (instead of $\mathscr{W}_N$) with respect to $\frac{1}{N!}B_{N,N}^{(\beta)}(t)dt$ (see (9)) . Thus we can exploit the invariance of the k-point correlation functions under permutations of the arguments together with their uniform convergence given in Lemma 1 resp. in Lemma 2.

By combinatorial arguments (see e.g. Corollary 2.4.11, Lemmas 2.4.9 and 2.4.12 in [18]) one can show the following connection between $\sigma_N(H)$ and $\gamma_N(k,H)$.

Lemma 3 (cf. Chap. 2 in [18]).

(i) For $N \in \mathbb{N}$ we have

$$\int_0^s d\sigma_N(H) = \sum_{k=2}^{N}(-1)^k \int_0^s d\gamma_N(k,H). \tag{25}$$

(ii) For $N \in \mathbb{N}$ and $m \leq N$ we have

$$\int_0^s d\sigma_N(H) \geq \sum_{k=2}^{m} (-1)^k \int_0^s d\gamma_N(k, H) \quad \text{for } m \text{ odd}$$

$$\int_0^s d\sigma_N(H) \leq \sum_{k=2}^{m} (-1)^k \int_0^s d\gamma_N(k, H) \quad \text{for } m \text{ even.}$$

We can use Lemma 3 to prove the following theorem, which states the point wise convergence of the empirical spacing distribution.

Theorem 3 (cf. [5] for $\beta = 2$). *Suppose that the assumptions of Theorem 1 ($\beta = 2$) resp. of Theorem 2 ($\beta = 1, 4$) are satisfied. Then we have for $\beta = 1, 2, 4$, $s \in \mathbb{R}$ and $W_k^{(\beta)}$ as in (21) ($\beta = 1, 4$) resp. in (17) ($\beta = 2$)*

$$\lim_{N\to\infty} \mathbb{E}_{N,\beta} \left(\int_0^s d\sigma_N(H) \right) = \sum_{k\geq 2} (-1)^k \int_{0\leq z_2 \leq \ldots \leq z_k \leq s} W_k^{(\beta)}(0, z_2, \ldots, z_k) dz_2 \ldots dz_k. \tag{26}$$

In particular, we claim that the series on the right hand side of the equation converges.

Proof. The proof is in the spirit of [18, Chap. 5]. Taking expectations in (25) leads to

$$\mathbb{E}_{N,\beta} \left(\int_0^s d\sigma_N(H) \right) = \sum_{k=2}^{N} (-1)^k \mathbb{E}_{N,\beta} \left(\int_0^s d\gamma_N(k, H) \right). \tag{27}$$

We start with the calculation of the expectation on the right hand side of (27). Observe that we can rewrite

$$\int_0^s d\gamma_N(k, H) = \frac{1}{|A_N|} \sum_{T \subset \{1,\ldots,N\}, |T|=k} \chi(\tilde{\lambda}_T)$$

with

$$\tilde{\lambda}_T := (\tilde{\lambda}_{i_1}, \ldots, \tilde{\lambda}_{i_k}) \quad \text{for } T = \{i_1, \ldots, i_k\} \text{ with } 1 \leq i_1 < \ldots < i_k \leq N$$

and

$$\chi(t_t, \ldots, t_k) := \chi_{(0,s)} \left(\max_{i=1,\ldots,k} t_i - \min_{i=1,\ldots,k} t_i \right) \prod_{i=1}^{k} \chi_{A_N}(t_i),$$

where $\chi_{(0,s)}$ resp. χ_{A_N} denote the characteristic functions on $(0, s)$ resp. on A_N.

Using the symmetry of the joint density of the eigenvalues with respect to permutations of the variables (see (9) and Definition 1) we conclude

$$\begin{aligned}\mathbb{E}_{N,\beta}\left(\int_0^s d\gamma_N(k,H)\right) &= \frac{1}{N!}\int_{\mathbb{R}^N}\left(\frac{1}{|A_N|}\sum_{T\subset\{1,\dots,N\},|T|=k}\chi(t_T)\right)B_{N,N}^{(\beta)}(t)dt\\ &= \frac{1}{|A_N|}\frac{1}{N!}\binom{N}{k}\int_{\mathbb{R}^N}\chi(t_1,\dots,t_k)B_{N,N}^{(\beta)}(t)dt\\ &= \frac{1}{|A_N|}\int_{\mathscr{W}_k\cap A_N^k}\chi(t_1,\dots,t_k)B_{N,k}^{(\beta)}(t_1,\dots,t_k)dt_1\dots dt_k.\end{aligned} \tag{28}$$

It is straightforward to prove the following bound

$$\frac{1}{|A_N|}\int_{\mathscr{W}_k\cap A_N^k}\chi(t_1,\dots,t_k)dt_1\dots dt_k \le \frac{s^{k-1}}{(k-1)!}$$

which, together with the uniform convergence of Lemmas 1 (*iii*) and 2 (*i*), leads to

$$\begin{aligned}&\frac{1}{|A_N|}\int_{\mathscr{W}_k\cap A_N^k}\chi(t_1,\dots,t_k)B_{N,k}^{(\beta)}(t_1,\dots,t_k)dt_1\dots dt_k\\ =&\frac{1}{|A_N|}\int_{\mathscr{W}_k\cap A_N^k}\chi(t_1,\dots,t_k)W_k^{(\beta)}(t_1,\dots,t_k)dt_1\dots dt_k + \mathscr{O}_{s,k}(\kappa_N),\end{aligned}$$

where the constant implicit in the $\mathscr{O}$-notation may depend on s and k as indicated by the subscripts. Using the translation invariance of $W_k^{(\beta)}$ (see (17) for $\beta = 2$ and Lemma 2 (*iii*) for $\beta = 1, 4$) together with the change of variables $z_1 = t_1, z_i = t_i - t_1, i = 2,\dots,k$ and the definition of χ we have

$$\begin{aligned}&\frac{1}{|A_N|}\int_{\mathscr{W}_k\cap A_N^k}\chi(t_1,\dots,t_k)W_k^{(\beta)}(t_1,\dots,t_k)dt_1\dots dt_k\\ =&\int_{0\le z_2\le\dots\le z_k\le s}W_k^{(\beta)}(0,z_2,\dots,z_k)dz_2\dots dz_k\\ &-\frac{1}{|A_N|}\int_{A_N}\left(\int_{0\le z_2\le\dots\le z_k\le s}W_k^{(\beta)}(0,z_2,\dots,z_k)\left(1-\prod_{j=2}^k\chi_{A_N}(z_1+z_j)\right)dz_2\dots dz_k\right)dz_1\\ =&\int_{0\le z_2\le\dots\le z_k\le s}W_k^{(\beta)}(0,z_2,\dots,z_k)dz_2\dots dz_k + \mathscr{O}_{s,k}\left(\frac{1}{|A_N|}\right).\end{aligned}$$

Hence we obtain

$$\lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\gamma_N(k,H)\right) = \int_{0\le z_2\le\ldots\le z_k\le s} W_k^{(\beta)}(0,z_2,\ldots,z_k)dz_2\ldots dz_k. \tag{29}$$

For later reference we observe that by the upper bounds on $W_k^{(\beta)}$ provided in Lemma 1 ($\beta = 2$) and in Lemma 2 ($\beta = 1, 4$) we have

$$\lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\gamma_N(k,H)\right) \le C^k s^{k-1}\left(\frac{1}{\sqrt{k-1}}\right)^{k-1}. \tag{30}$$

It remains to show that in (27) the limit $N \to \infty$ may be interchanged with the infinite summation over k. Taking expectations in Lemma 3 (*ii*) together with the convergence in (29) implies for m odd

$$\sum_{k=2}^{m}(-1)^k \lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\gamma_N(k,H)\right) \le \liminf_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right) \tag{31}$$

$$\limsup_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right) \le \sum_{k=2}^{m+1}(-1)^k \lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\gamma_N(k,H)\right). \tag{32}$$

Inequality (30) ensures the convergence of the series in (31) and (32) if we take $m \to \infty$. Sending $m \to \infty$ in (31) and (32) implies that the limit $\mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right)$ exists for $N \to \infty$. We obtain

$$\lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right) = \sum_{k=2}^{\infty}(-1)^k \lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\gamma_N(k,H)\right)$$

which, together with (29), completes the proof. □

Remark 5. In the above theorem we have obtained a representation for the limit of the expected spacing distribution in terms of $W_k^{(\beta)}$, which is hence universal in the sense that the limit does neither depend on V nor on the details of the localisations, i.e. on the point a or on the interval I_N as long as the assumptions of Theorem 1 resp. Theorem 2 are satisfied.

However, formula (26) is somewhat complicated. In the next section we show how it is related to the so-called gap probabilities that have an explicit representation in terms of particular Painlevé V functions.

5.2 Spacing Distributions and Gap Probabilities

A gap probability is the probability of having no eigenvalues in a given interval. Observe that for finite N and $\beta = 2$ we have (see e.g. [5, p. 108])

$$\mathbb{P}_{N,2}(\{\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N\} \cap (0,s) = \emptyset) = \frac{1}{N!} \int_{(\mathbb{R}\setminus(0,s))^N} B_{N,N}^{(2)}(t_1, \ldots, t_N)\, dt_1 \ldots dt_N$$

$$= \sum_{k=0}^{N} \frac{(-1)^k}{k!} \int_0^s \ldots \int_0^s \det\left(K_{N,2}(t_i, t_j)\right)_{1 \le i,j \le k} dt_1 \ldots dt_k$$

$$= \det(1 - K_{N,2}|_{L^2(0,s)}).$$

Here $K_{N,2}|_{L^2(0,s)}$ denotes the integral operator on $L^2(0,s)$ with kernel $K_{N,2}$ and the last equality is just a standard expansion for the corresponding Fredholm determinant. Recall that $K_{N,2} \to K_2$ for $N \to \infty$ (Theorem 1). Furthermore, one can also show that the corresponding Fredholm determinants converge (see [6] and also [33]). This motivates that the large N-limit

$$G_2(s) := \det(1 - K_2|_{L^2(0,s)})$$

is called the gap probability (for $\beta = 2$). For $\beta = 1$ and 4 the gap probabilities G_β are defined as square roots of determinants of operators on $L^2(0,s) \times L^2(0,s)$ (see e.g. [6, Corollary 6.12]).

By the standard expansion of the Fredholm determinant we have

$$G_2(s) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^s \ldots \int_0^s W_k^{(2)}(t_1, \ldots, t_k) dt_1 \ldots dt_k. \tag{33}$$

Using more involved arguments the analogue of Eq. (33) (with $W_k^{(2)}$ replaced by $W_k^{(\beta)}$ and G_2 replaced by G_β) can also be shown for $\beta = 1$ and $\beta = 4$ (see e.g. [27, Sect. 7.1]). The following theorem relates the derivatives of the gap probabilities to the limiting spacing distributions for all $\beta \in \{1, 2, 4\}$ (see e.g. [5, p. 126] for $\beta = 2$).

Lemma 4. *For $\beta = 1, 2, 4$ we have*

$$-G'_\beta(s) = 1 - \lim_{N \to \infty} \mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right)$$

Proof. We introduce the function

$$\tilde{G}_\beta(\varepsilon, s) := \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_\varepsilon^s \dots \int_\varepsilon^s W_k^{(\beta)}(t_1, \dots, t_k) dt_1 \dots dt_k$$
$$= 1 - s + \varepsilon + \sum_{k=2}^{\infty} (-1)^k \int_\varepsilon^s \left(\int_{t_1 \le t_2 \le \dots \le t_k \le s} W_k^{(\beta)}(t_1, \dots, t_k) dt_2 \dots dt_k \right) dt_1. \tag{34}$$

Here we have used $W_1^{(\beta)}(t) = 1$ for all $t \in \mathbb{R}$ (cf. Lemmas 1 and 2). Then the translation invariance of $W_k^{(\beta)}$ implies $\tilde{G}_\beta(\varepsilon, s) = \widetilde{G_\beta}(0, s-\varepsilon) = G_\beta(s-\varepsilon)$ and hence $\frac{\partial}{\partial \varepsilon}\big|_{\varepsilon=0} \tilde{G}_\beta(\varepsilon, s) = -G'_\beta(s)$. Differentiating each term of the series in (34) (which is absolutely convergent, see (30)) we obtain the desired result from Theorem 3. □

Remark 6. As mentioned above there is a remarkable identity that allows to express the gap probabilities $G_\beta, \beta \in \{1, 2, 4\}$ in terms of Painlevé V functions. More precisely, let σ be the solution of

$$(s\sigma'')^2 + 4(s\sigma' - \sigma)(s\sigma' - \sigma + (\sigma')^2)$$

with boundary condition

$$\sigma(s) \sim -\frac{s}{\pi} - \frac{s^2}{\pi^2} - \frac{s^3}{\pi^3} + \mathcal{O}(s^4) \quad \text{for } s \to 0.$$

Then in the case $\beta = 2$ we have (see [16])

$$G_2(s) = \exp\left(\int_0^{\pi s} \frac{\sigma(t)}{t} dt \right).$$

For $\beta = 1$ and $\beta = 4$ see [2] and [13] for analogue formulae.

We recall that Lemma 4 together with the Painlevé representations for G_β are useful to verify that μ_β as defined through (23) is indeed a probability measure (see [27, Chaps. 6 and 7]).

6 Results

In this section we state our new result (Theorem 4) for the expected empirical spacing distribution for invariant orthogonal and symplectic ensembles. We include a brief discussion of related results that can be found in the literature.

Except for the point wise convergence for $\beta = 2$ presented in Theorem 3 (cf. [8]) the empirical spacing distribution has so far only been considered for circular ensembles. In the case $\beta = 2$ the circular unitary ensemble (CUE) is given by the unitary group $U(N)$ with the normalised translation invariant Haar measure. The joint distribution of the complex eigenvalues $e^{i\theta_1}, \ldots, e^{i\theta_N}$ in the CUE and in the related orthogonal and symplectic ensembles ($\beta = 1, 4$) is given by

$$d\mathbb{P}_{N,\beta}(\theta) = \frac{1}{Z_{N,\beta}} \prod_{j<k} \left| e^{i\theta_k} - e^{i\theta_j} \right|^{\beta} d\theta_1 \ldots d\theta_N, \quad \beta = 1, 2, 4. \tag{35}$$

As the expected spectral density is constant for these ensembles, the eigenvalues can be normalised to have mean spacing one by the linear rescaling

$$\widetilde{\theta_i} := \frac{N\theta_i}{2\pi}.$$

Observe that we do not need to localise the spectrum in these cases.

It is a result of Katz and Sarnak in [18, Chaps. 1 and 2] that for circular ensembles with $\beta = 2$ the expected empirical spacing distribution converges to the same measure μ_2 that we have defined in (23). Moreover, they show a stronger version of convergence, i.e. the vanishing of the expected Kolmogorov distance

$$\lim_{N\to\infty} \mathbb{E}_{N,\beta} \left(\sup_{s\in\mathbb{R}} \left| \int_0^s d\sigma_N(H) - \int_0^s d\mu_2 \right| \right) = 0. \tag{36}$$

Here σ_N is defined as the (normalised) counting measure of the nearest neighbour spacings between $\widetilde{\theta}_j$'s. The definition of σ_N is similar to (8) with the pre-factor altered to $1/(N-1)$ and without the restriction to I_N in the sum.

In [29] the convergence in (36) is sharpened, proving almost sure convergence, and generalised to COE ($\beta = 1$), but not to $\beta = 4$. Moreover, Soshnikov shows in [29] for both CUE and COE a central limit theorem for spacings. For example, he proves that the appropriately normalised random variables

$$\xi_N(s) = \frac{\int_0^s d\sigma_N(H) - \mathbb{E}_{N,\beta}\left(\int_0^s d\sigma_N(H)\right)}{\sqrt{N}}$$

converge to a Gaussian process ξ with $\mathbb{E}(\xi(s)) = 0$ and for which $\mathbb{E}(\xi(s)\xi(t))$ can be expressed in terms of the k-point correlations of (35).

Another interesting result [2, Sect. 4.2] concerns the theory of determinantal point processes. In [2] it is shown that for such point processes with constant intensities generated by a suitable class of kernels (including in particular the sine-kernel K_2) the linear statistics of the empirical spacing distribution converge almost surely to the linear statistics of μ_2 as the number of considered points tends to infinity. This result does not deal with the distribution of the eigenvalues of random matrices for finite N. Nevertheless, it is conceivable that this result might be useful for proving the convergence of the empirical spacing distribution.

We now turn to our recent results. We show in [27] that the analogue version of (36) is valid for orthogonal and symplectic invariant ensembles satisfying (2) and (4). In fact, we reduce the question of convergence of the expected Kolmogorov distance of the empirical spacing distribution to the convergence of the corresponding kernel functions. All information that is needed on the convergence $K_{N,\beta} \to K_\beta$ is summarised in the following

Assumption 1. *We consider invariant ensembles with joint distribution of the eigenvalues given by (2) and (4) for $\beta \in \{1, 2, 4\}$. We assume that the limiting spectral density exists and we choose a, I_N and the rescaling of the eigenvalues as in Sect. 3.2. We assume that there exists a sequence κ_N such that $\kappa_N \to 0$ for $N \to \infty$, such that for the rescaled (matrix) kernels $\hat{K}_{N,\beta}$ (see (14) and (19)) we have*

$$\widehat{K_{N,\beta}}\left(a + \frac{x}{N\psi(a)}, a + \frac{y}{N\psi(a)}\right) = K_\beta(x, y) + \mathcal{O}(\kappa_N) \tag{37}$$

uniformly for $x, y \in A_N$.

Our main theorem then reads

Theorem 4 ([27]). *Under Assumption 1 we have*

$$\lim_{N\to\infty} \mathbb{E}_{N,\beta}\left(\sup_{s\in\mathbb{R}}\left|\int_0^s d\sigma_N(H) - \int_0^s d\mu_\beta\right|\right) = 0, \tag{38}$$

where μ_β is defined through (23).

In particular, our theorem covers all invariant ensembles for which the convergence of $K_{N,\beta}$ to K_β has been proved using a Riemann-Hilbert approach (see e.g. Theorems 1 and 2). Observe that our formulation of Theorem 4 also includes all ensembles for which (37) will be established in the future.

The proof follows the path devised by Katz and Sarnak in [18] for $\beta = 2$ and extends their methods in two ways. On the one hand, we have to consider the additional localisation that is needed in our setting. We can use the same methods as in [18] to express the expected empirical distribution of the spacings in terms of the rescaled k-point correlation functions $B_{N,k}^{(\beta)}$ (see (27) and (28)). On the other hand, we generalise their methods to $\beta = 1$ and $\beta = 4$. Here the relation between the matrix-kernel functions $K_{N,\beta}$ and the expected empirical spacing distribution is more involved. Moreover, for $\beta = 4$ subtle cancellations have to be used to establish convergence.

The proof of the main theorem comes in three steps: The first step is the point wise convergence as shown in Theorem 3. This convergence is well known although it seems that the details have so far only been worked out in the case $\beta = 2$ (see e.g. [2, 6]). In order to obtain the convergence of

$$\mathbb{E}_{N,\beta}\left(\left|\int_0^s d\sigma_N(H) - \int_0^s d\mu_\beta\right|\right) \tag{39}$$

for any given $s \in \mathbb{R}$, we bound the variance of $\int_0^s d\gamma_N(k, H)$ in the second step. As stated above, this is the most challenging part in generalising the method of Katz and Sarnak to $\beta = 1, 4$. Here we found the representation of the k-point correlation functions in terms of $K_{N,\beta}$ as provided in [32] useful. Finally, the desired result is obtained by controlling the s-dependence of the bound on (39) together with tail estimates on μ_β. Details of the proof can be found in [27].

7 Numerical Results

In addition to the analytical considerations that led to Theorem 4, the work [27] also contains numerical experiments in MATLAB in order to determine the rate of convergence in (38). We summarise some of the findings of [27] in the present section.

We conduct our experiments for the three classical Gaussian ensembles GOE, GUE, GSE and for general β-ensembles with $\beta \in \{7, 15.5, 20\}$. We also include real, complex and quaternionic Wigner matrices with i.i.d. entries that are drawn e.g. from beta, poisson, exponential, uniform or chi-squared distributions. Observe (see Sect. 3.1) that in all these cases the limiting spectral density ψ is given by the Wigner semi-circle law. We may adapt the parameters such that the support of ψ is the interval $[-1, 1]$.

A little thought shows that the localisation and rescaling procedure to define $d\sigma_N$ (see (8)) will not lead to an optimal and natural rate of convergence. Firstly, the rate will depend on the number of eigenvalues, i.e. on the length of I_N. Secondly, the linear rescaling (6) is not optimal since the density ψ is approximated on all of I_N by the constant $\psi(a)$. A far better rescaling in this respect (but less suitable for analytical considerations) is the so-called unfolding, that we explain now. Let $I \subset [-1, 1] = \text{supp}(\psi)$ be an interval. Denote by

$$F(t) := \frac{2}{\pi}\int_0^t \sqrt{1-s^2}\chi_{[-1,1]}ds$$

the distribution function of the semi-circle law. The rescaling is then given by

$$\tilde{\lambda}_i := NF(\lambda_i).$$

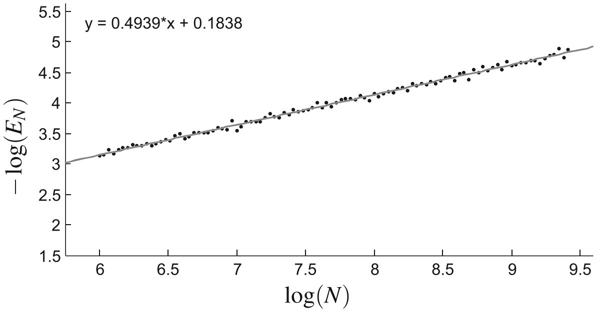


Fig. 1 Data set and best linear fit for β-ensemble with $\beta = 4$ and $I = [-0.5, 0.5]$

Observe that in an average sense

$$\tilde{\lambda}_{i+1} - \tilde{\lambda}_i \approx NF'(\lambda_i)(\lambda_{i+1} - \lambda_i) \approx N\psi(\lambda_i)\frac{1}{N\psi(\lambda_i)} = 1.$$

The spacing distribution corresponding to the unfolded statistics is given by

$$\sigma_N^{(\text{unf})}(H) := \frac{1}{|\{i : \lambda_i \in I\}| - 1} \sum_{\lambda_i, \lambda_{i+1} \in I} \delta_{\tilde{\lambda}_{i+1} - \tilde{\lambda}_i}.$$

We restrict our attention to intervals $I = [-0.1; 0.1]$, $I = [-0.5; 0.5]$, $I = [-0.75; 0.75]$ and some non-centred intervals such as $I = [0.4; 0.6]$

We provide numerical evidence for the claim that the leading asymptotic of the considered expected Kolmogorov distance is $CN^{-1/2}$, i.e.

$$E_N := \mathbb{E}_{N,\beta}\left(\sup_{s\in\mathbb{R}}\left|\int_{-\infty}^{s} d\sigma_N^{(\text{unf})}(H) - \int_{-\infty}^{s} d\mu_\beta\right|\right) \sim CN^{-1/2} \tag{40}$$

for some constant C that depends mildly on the chosen ensemble and on the choice of I. In Figs. 1–3 we have plotted $y := -\log E_N$ against $x := \log N$. We see in all three cases that our numerical approximations to y, obtained by Monte Carlo simulations, cluster impressively close to a straight line

$$y(x) \sim ax + b, \quad \text{i.e.} \quad E_N \sim e^{-b} N^{-a}.$$

In all our experiments [27] we found $a \in [0.48; 0.53]$ and $b \in [-1, 0.5]$, see also Table 1.

One important issue in the numerical experiments is the approximation of the liming measures μ_β resp. their densities p_β. For $\beta = 1, 2, 4$ we use the MATLAB toolbox by Bornemann (cf. [4]) for a fast and precise evaluation of the related gap probabilities. Then we obtain the limiting densities p_β by numerical differentiation.

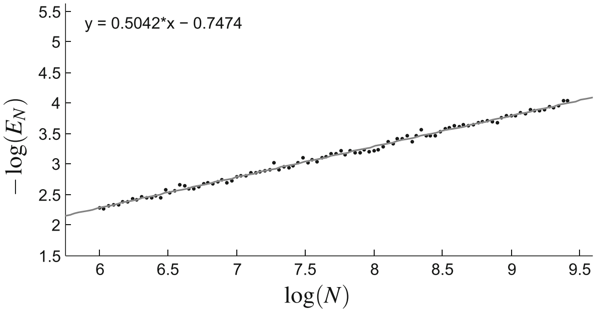


Fig. 2 Data set and best linear fit for β-ensemble with $\beta = 20$ and $I = [0.4, 0.6]$

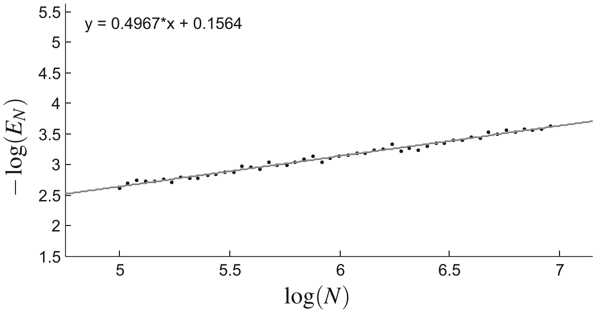


Fig. 3 Data set and best linear fit for real Wigner matrices with unfolded statistics and beta (2,5)-distributed entries and $I = [-0.5, 0.5]$

Table 1 Best fit straight lines for β-ensembles with unfolded statistics

Ensemble	Interval I	Best linear fir
$\beta = 1$	$I = [0.7; 0.9]$	$y = 0.5077x - 0.9758$
$\beta = 2$	$I = [-0.1; 0.1]$	$y = 0.4958x - 0.6325$
$\beta = 4$	$I = [-0.75; 0.75]$	$y = 0.4965x + 0.3356$
$\beta = 7$	$I = [-0.1; 0.1]$	$y = 0.4991x - 0.6357$
$\beta = 15.5$	$I = [-0.5; 0.5]$	$y = 0.4945x + 0.1765$
$\beta = 20$	$I = [0.4; 0.6]$	$y = 0.5042x - 0.7474$

For $\beta \in \mathbb{R}_+ \setminus \{1, 2, 4\}$ no such precise numerical schemes for the evaluation of p_β are available. Instead, we use the generalised Wigner surmise (see [22]), which is only an approximation to the limiting distribution. One may wonder how one can test numerically a limiting law without knowing its precise form. Looking at (40) one notes that the numerics will not detect the replacement of $\int_{-\infty}^{s} d\mu_\beta$ by an approximation $\int_{-\infty}^{s} d\hat{\mu}_\beta$ as long as their deviation is small compared to E_N. As it turns out, the Wigner surmise approximates the true limiting law well enough to confirm (40) for the range of N and β that we have tested. Moreover, since in all our experiments E_N took values below 0.02 we may safely infer that the difference between the distribution functions of the Wigner surmise and the true distribution is less than 0.01 for all values of β that we have investigated, i.e. $\beta \in \{7, 15.5, 20\}$.

Acknowledgements Both authors acknowledge support from the Deutsche Forschungsgemeinschaft in the framework of the SFB/TR 12 “Symmetries and Universality in Mesoscopic Systems”. We are grateful to Peter Forrester for useful remarks.

References

1. Akemann, G., Baik, J., Di Francesco, P.: The Oxford Handbook of Random Matrix Theory. Oxford Handbooks in Mathematics Series. Oxford University Press, New York (2011)
2. Anderson, G.W., Guionnet, A., Zeitouni, O.: An Introduction to Random Matrices, 1st edn. Cambridge University Press, Cambridge (2009)
3. Baik, J., Deift, P., Johansson, K.: On the distribution of the length of the longest increasing subsequence of random permutations. J. Am. Math. Soc. **12**(4), 1119–1178 (1999)
4. Bornemann, F.: On the numerical evaluation of distributions in random matrix theory: a review. Markov Process. Relat. Fields **16**(4), 803–866 (2010)
5. Deift, P.: Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach. Volume 3 of Courant Lecture Notes in Mathematics. New York University Courant Institute of Mathematical Sciences, New York (1999)
6. Deift, P., Gioev, D.: Random Matrix Theory: Invariant Ensembles and Universality. Volume 18 of Courant Lecture Notes in Mathematics. Courant Institute of Mathematical Sciences, New York (2009)
7. Deift, P., Kriecherbauer, T., McLaughlin, K.T.-R., Venakides, S., Zhou, X.: Strong asymptotics of orthogonal polynomials with respect to exponential weights. Commun. Pure Appl. Math. **52**(12), 1491–1552 (1999)
8. Deift, P., Kriecherbauer, T., McLaughlin, K.T.-R., Venakides, S., Zhou, X.: Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. Commun. Pure Appl. Math. **52**(11), 1335–1425 (1999)
9. Deift, P., Gioev, D., Kriecherbauer, T., Vanlessen, M.: Universality for orthogonal and symplectic Laguerre-type ensembles. J. Stat. Phys. **129**(5–6), 949–1053 (2007)
10. Erdős, L.: Universality of Wigner random matrices: a survey of recent results. Uspekhi Mat. Nauk. **66**(3(399)), 67–198 (2011)
11. Ferrari, P.L., Spohn, H.: Random growth models. In: Akemann, G., Baik, J., Di Francesco, P. (eds.) The Oxford Handbook of Random Matrix Theory. Oxford University Press, New York (2011)
12. Forrester, P.J.: Log-Gases and Random Matrices. Volume 34 of London Mathematical Society Monographs Series. Princeton University Press, Princeton (2010)
13. Forrester, P.J., Witte, N.S.: Exact Wigner surmise type evaluation of the spacing distribution in the bulk of the scaled random matrix ensembles. Lett. Math. Phys. **53**(3), 195–200 (2000)
14. Hiai, F., Petz, D.: The Semicircle Law, Free Random Variables and Entropy. Volume 77 of Mathematical Surveys and Monographs. American Mathematical Society, Providence (2000)
15. Hurwitz, A.: Über die Composition der quadratischen Formen von beliebig vielen Variablen. Nachr. Ges. Wiss. Göttingen, 309–316 (1898)
16. Jimbo, M., Miwa, T., Môri, Y., Sato, M.: Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent. Phys. D **1**(1), 80–158 (1980)
17. Johansson, K.: On fluctuations of eigenvalues of random Hermitian matrices. Duke Math. J. **91**(1), 151–204 (1998)

18. Katz, N.M., Sarnak, P.: Random Matrices, Frobenius Eigenvalues, and Monodromy. Volume of 45 American Mathematical Society Colloquium Publications. American Mathematical Society, Providence (1999)
19. Kriecherbauer, T., Krug, J.: A pedestrian's view on interacting particle systems, KPZ universality and random matrices. J. Phys. A **43**(40), 403001, 41 (2010)
20. Kuijlaars, A.B.J., Vanlessen, M.: Universality for eigenvalue correlations from the modified Jacobi unitary ensemble. Int. Math. Res. Not. **30**, 1575–1600 (2002)
21. Kuijlaars, A.B.J., Vanlessen, M.: Universality for eigenvalue correlations at the origin of the spectrum. Commun. Math. Phys. **243**(1), 163–191 (2003)
22. Le Caër, G., Male, C., Delannay, R.: Nearest-neighbour spacing distributions of the β-Hermite ensemble of random matrices. Phys. A Stat. Mech. Appl. **383**, 190–208 (2007)
23. Levin, E., Lubinsky, D.S.: Universality limits in the bulk for varying measures. Adv. Math. **219**(3), 743–779 (2008)
24. McLaughlin, K.T.-R., Miller, P.D.: The $\overline{\partial}$ steepest descent method for orthogonal polynomials on the real line with varying weights. Int. Math. Res. Not. IMRN Art. ID rnn 075, 66 (2008). doi:10.1093/imrn/rnn075
25. Mehta, M.L.: Random matrices, Pure and Applied. Volume 142 of Mathematics (Amsterdam), 3rd edn. Elsevier/Academic, Amsterdam (2004)
26. Ramírez, J.A., Rider, B., Virág, B.: Beta ensembles, stochastic Airy spectrum, and a diffusion. J. Am. Math. Soc. **24**(4), 919–944 (2011)
27. Schubert, K.: On the convergence of the nearest neighbour eigenvalue spacing distribution for orthogonal and symplectic ensembles. PhD thesis, Ruhr-Universität Bochum (2012)
28. Shcherbina, M.: Orthogonal and symplectic matrix models: universality and other properties. Commun. Math. Phys. **307**(3), 761–790 (2011)
29. Soshnikov, A.: Level spacings distribution for large random matrices: Gaussian fluctuations. Ann. Math. **148**(2), 573–617 (1998)
30. Tao, T.: Topics in Random Matrix Theory. Volume of 132 Graduate Studies in Mathematics. American Mathematical Society, Providence (2012)
31. Tao, T., Vu, V.: Random matrices: universality of local eigenvalue statistics. Acta Math. **206**(1), 127–204 (2011)
32. Tracy, C.A., Widom, H.: Correlation functions, cluster functions, and spacing distributions for random matrices. J. Stat. Phys. **92**(5–6), 809–835 (1998)
33. Tracy, C.A., Widom, H.: Matrix kernels for the Gaussian orthogonal and symplectic ensembles. Ann. Inst. Fourier (Grenoble) **55**(6), 2197–2207 (2005)
34. Vanlessen, M.: Strong asymptotics of Laguerre-type orthogonal polynomials and applications in random matrix theory. Constr. Approx. **25**(2), 125–175 (2007)
35. Widom, H.: On the relation between orthogonal, symplectic and unitary matrix ensembles. J. Stat. Phys. **94**(3–4), 47–363 (1999)
36. Wishart, J.: The generalised product moment distribution in samples from a normal multivariate population. Biometrika **20A**(1/2), 32–52 (1928)
37. Zirnbauer, M.R.: Symmetry classes. In: Akemann, G., Baik, J., Di Francesco, P. (eds.) The Oxford Handbook of Random Matrix Theory. Oxford University Press, New York (2011)

Stein's Method and Central Limit Theorems for Haar Distributed Orthogonal Matrices: Some Recent Developments

Michael Stolz

Abstract In recent years, Stein's method of normal approximation has been applied to Haar distributed orthogonal matrices by several authors. We give an introduction to the relevant aspects of the method, highlight a few results thus obtained, and finally argue that the quantitative multivariate central limit theorem for traces of powers that was recently obtained by Döbler and the author for the special orthogonal group remains true for the full orthogonal group.

1 Introduction

The observation that random matrices, picked according to Haar measure from orthogonal groups of growing dimension, give rise to central limit theorems, dates back at least to Émile Borel, whose 1905 result on random elements of spheres can be read as saying that if the upper left entry of a Haar orthogonal $n \times n$ matrix is scaled by $\sqrt{n}$, it converges to a standard normal distribution as n tends to infinity. See [5] for more historical background. Borel's observation may be seen as an early result in random matrix theory, but it must be emphasized that from this point of view it is rather atypical. In the best known random matrix models, such as the Gaussian Unitary Ensemble (GUE) or Wigner matrices, the distributions of the individual matrix entries are either known or subject to certain assumptions, and one is interested in various global and local features of the eigenvalues of the random matrix. On the other hand, for Haar orthogonal matrices or, more generally, for Haar distributed elements of a compact matrix group, properties of the distributions of the individual entries have to be inferred from the distribution of the matrix as a whole.

M. Stolz (✉)
Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany
e-mail: michael.stolz@ruhr-uni-bochum.de

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_4,

Nevertheless, GUE matrices and Wigner matrices give rise to central limit theorems in a different way: If M_n is an $n \times n$ GUE matrix, say, with (necessarily real) eigenvalues $\lambda_1, \dots, \lambda_n$, then the empirical eigenvalue distribution

$$L_n := \mathbb{L}_n(M_n) := \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j}$$

is known to converge to Wigner's semicircular distribution σ in various senses. Then for suitable real valued test functions f the fluctuation

$$n\left(\int f dL_n - \mathbb{E}\left(\int f dL_n\right)\right)$$

tends to a Gaussian limit as $n \to \infty$, see, e.g., [17].

This type of question also makes sense for Haar distributed matrices from a compact group. For the unitary group, the uniform distribution on the unit circle $\mathbb{T}$ of the complex plane replaces the semicircular distribution. If $f : \mathbb{T} \to \mathbb{R}$ is continuous and of bounded variation, then the fluctuation on the unitary group has a pointwise expression

$$\begin{aligned} n(L_n(f) - \mathbb{E}(L_n(f))) &= f(\lambda_{n1}) + \ldots + f(\lambda_{nn}) - n\hat{f}(0) \\ &= \sum_{j=1}^{\infty} \hat{f}(j) \mathrm{Tr}(M_n^j) + \sum_{j=1}^{\infty} \overline{\hat{f}(j) \mathrm{Tr}(M_n^j)}. \end{aligned} \tag{1}$$

This expansion shows that if f is a trigonometric polynomial, a CLT for fluctuations will be equivalent to a CLT for random vectors of the form

$$(\mathrm{Tr}(M_n), \mathrm{Tr}(M_n^2), \dots, \mathrm{Tr}(M_n^d)).$$

This CLT was established in the famous paper of Diaconis and Shahshahani [7] from 1994 that turned traces of powers into a popular subject in the theory of Haar distributed matrices. It was used as a stepping stone for the treatment of more general test functions by Diaconis and Evans [6] in 2001.

Diaconis and Shahshahani proved their theorem using the method of moments. In the orthogonal case, in which the reasoning above remains true with some caveats, it turned out that the moment

$$\mathbb{E}\left((\mathrm{Tr}(M_n))^{a_1} (\mathrm{Tr}(M_n^2))^{a_2} \dots (\mathrm{Tr}(M_n^d))^{a_d}\right)$$

actually coincided with the corresponding moment of the Gaussian limit distribution (to be described in Lemma 4 below) as soon as

$$2n \geq k_a := \sum_{j=1}^{d} j a_j \tag{2}$$

(see [29] for the threshold given here). This led Diaconis to conjecture that the speed of convergence should be rather fast. Subsequently, only a few years later, Stein [28] proved superpolynomial, and Johansson [15] finally exponential convergence.

During the last decade several authors, certainly inspired by Stein's paper, have turned to the broader approach to normal approximation that bears the name "Stein's method" to investigate the speed of convergence in various CLTs for Haar orthogonals (and Haar distributed elements of other compact matrix groups), obtaining worse rates of convergence, but a wider range of results. It is the aim of this survey paper to introduce the relevant techniques, present some results on the linear combinations and traces of powers problems, and extend the multivariate traces of powers result from the special orthogonal to the full orthogonal group.

2 Univariate Normal Approximation via Stein's Method

Consider random variables W and Z, with distributions P and Q, respectively. A useful recipe to quantify the distance between P and Q is to choose a family $\mathscr{H}$ of test functions and define

$$d_{\mathscr{H}}(P, Q) := \sup_{h \in \mathscr{H}} |\mathbb{E}(h(W)) - \mathbb{E}(h(Z))|.$$

Well-known examples are $\mathscr{H} = \{1_{]-\infty,z]} \mid z \in \mathbb{R}\}$, giving rise to the Kolmogorov distance

$$d_{\mathscr{H}}(P, Q) = \sup_{z \in \mathbb{R}} |\mathbb{P}(W \leq z) - \mathbb{P}(Z \leq z)|,$$

and $\mathscr{H} = \{h : \mathbb{R}^d \to \mathbb{R}, \text{Lipschitz with constant} \leq 1\}$, which defines the Wasserstein distance.

Stein's method, developed by Charles Stein since the early 1970s (see [26]), serves to bound distances of this type. Stein himself developed his method for normally distributed Z, his student L.H.Y. Chen developed a parallel theory for the Poisson distribution, see [1] for a monographic treatment. Nowadays, the methods for normal and Poisson limits are still the best developed instances of Stein's approach, but progress has been made on other distributions as well (see, e.g., [3] and [9]). In accordance with the nature of the limit theorems to be discussed in this survey, we will focus on the normal case and start with a sketch of the case of a univariate normal distribution. A much more detailed picture of the fundamentals (and a lot more) of Stein's method of normal approximation can be found in the recent textbook [4].

Write φ for the density of the univariate standard normal distribution. Since φ is strictly positive, for h measurable and $\mathbb{E}|h(Z)| < \infty$ we may define

$$f_h(x) := \frac{1}{\varphi(x)} \int_{-\infty}^{x} (h(y) - \mathbb{E}(h(Z)))\, \varphi(y)\, dy.$$

Then it can be verified by partial integration that f_h solves the *Stein equation*

$$f'(x) = xf(x) + h(x) - \mathbb{E}(h(Z)).$$

For Z a standard normal random variable and W such that $h(W)$ is integrable for all $h \in \mathscr{H}$, this implies that

$$|\mathbb{E}(h(W)) - \mathbb{E}(h(Z))| = |\mathbb{E}(f_h'(W)) - \mathbb{E}(W f_h(W))|. \tag{3}$$

So to bound the distance, defined by the class $\mathscr{H}$ of test functions, between the law of W and the standard normal distribution, which is the law of Z, it suffices to bound the right hand side of the last equation for all Stein solutions f_h, where h runs over $\mathscr{H}$. Note that this right-hand side involves only W, not Z. A crucial fact to be used in what follows is that estimates on f_h and its first and second derivatives are available that require only little information about h. To be specific, one has that if h is absolutely continuous, then

(i) $\|f_h\|_\infty \leq 2\|h'\|_\infty$.
(ii) $\|f_h'\|_\infty \leq \sqrt{2/\pi}\ \|h'\|_\infty$.
(iii) $\|f_h''\|_\infty \leq 2\|h'\|_\infty$.

Actually there are several approaches to bound the right-hand side of (3), see, e.g., [23]. The orthogonal group examples will use "exchangeable pairs", a device that was introduced by Stein in his monograph [27] of 1986. To illustrate the main ideas of this variant of the method, we will extract a few steps from an argument that Stein provided in his book.

An exchangeable pair is a pair (W, W') of random variables, defined on the same probability space and taking values in the same state space, such that (W, W') and (W', W) have the same distribution. An elementary, but crucial, consequence is that

$$\mathbb{E}\, g(W, W') = 0$$

for any antisymmetric function g defined on pairs of elements of the state space. For concreteness, we assume for now that W and W' are real-valued. In later applications they will be elements of a finite dimensional real vector space.

One further condition that has to be imposed on (W, W') is that there exist $0 < \lambda < 1$ such that

$$\mathbb{E}(W'|W) = (1 - \lambda)W.$$

This "regression condition" is quite natural in the context of normal approximation, since it is known to hold if (W, W') has a bivariate normal distribution. Actually, it is desirable to weaken the condition to the effect that the regression property needs to hold only approximately, and indeed this may be done, as shown by Rinott and Rotar in [25]. But for our purely illustrative purposes, we assume the condition as it stands.

Since λ is assumed to lie strictly between 0 and 1, W, W' must be centered, as

$$\mathbb{E}(W) = \mathbb{E}(W') = \mathbb{E}(\mathbb{E}(W'|W)) = \mathbb{E}((1-\lambda)W) = (1-\lambda)\mathbb{E}(W).$$

Making the specific choice

$$g(x, y) := (x - y)(f(x) + f(y))$$

of an antisymmetric function, where f is a function that will be specialized to a Stein solution later on, one obtains that

$$\begin{aligned}
0 &= \mathbb{E}((W - W')(f(W) + f(W'))\\
&= \mathbb{E}((W - W')(f(W') - f(W)) + 2\mathbb{E}((W - W')f(W))\\
&= \mathbb{E}((W - W')(f(W') - f(W)) + 2\mathbb{E}(f(W)\mathbb{E}((W - W')|W))\\
&= \mathbb{E}((W - W')(f(W') - f(W)) + 2\lambda\mathbb{E}(Wf(W)).
\end{aligned}$$

From this one concludes that

$$\begin{aligned}
\mathbb{E}(Wf(W)) &= \frac{1}{2\lambda}\mathbb{E}((W - W')(f(W) - f(W'))\\
&= \frac{1}{2\lambda}\mathbb{E}\left[\int_{W'}^{W}(W - W')f'(t)dt\right]\\
&= \frac{1}{2\lambda}\mathbb{E}\left[\int_{-(W-W')}^{0} f'(W + t)(W - W')dt\right]\\
&= \frac{1}{2\lambda}\mathbb{E}\left[\int_{\mathbb{R}} f'(W + t)K(t)dt\right],
\end{aligned}$$

where

$$K(t) = (W - W')\left(1_{\{-(W-W')\leq t\leq 0\}} - 1_{\{0<t\leq -(W-W')\}}\right).$$

On the other hand, a similar argument yields

$$\mathbb{E}(f'(W)) = \mathbb{E}\left[f'(W)(1 - \frac{1}{2\lambda}\mathbb{E}((W - W')^2|W)\right] + \frac{1}{2\lambda}\mathbb{E}\left[\int_{\mathbb{R}} f'(W)K(t)dt\right].$$

Assume h Lipschitz with minimal constant $\|h'\|_\infty$, and choose the Stein solution f_h in the place of f. Then it follows from the above that

$$\begin{aligned}|\mathbb{E}(h(W)) - \mathbb{E}(h(Z))| = &\ |\mathbb{E}(f_h'(W)) - \mathbb{E}(W f_h(W))| \\ \leq &\ \mathbb{E}\left| f_h'(W)\left(1 - \frac{1}{2\lambda}\mathbb{E}((W - W')^2|W)\right)\right| \\ &+ \frac{1}{2\lambda}\mathbb{E}\left[\int_{\mathbb{R}} |f_h'(W) - f_h'(W+t)|K(t)dt\right].\end{aligned}$$

Observing that

$$|f_h'(W) - f_h'(W+t)| \leq \|f_h''\|_\infty |t| \leq 2\|h'\|_\infty |t|$$

and recalling the bound

$$\|f_h'\|_\infty \leq (2/\pi)\|h'\|_\infty$$

on the solutions of Stein's equation, one finally arrives at a bound

$$\begin{aligned}|\mathbb{E}(h(W)) &- \mathbb{E}(h(Z))| \\ &\leq \frac{2}{\pi}\|h'\|_\infty\, \mathbb{E}\left[\left|1 - \frac{1}{2\lambda}\mathbb{E}((W - W')^2|W)\right|\right] + \frac{\|h'\|_\infty}{2\lambda}\, \mathbb{E}(|W - W'|^3).\end{aligned}$$

Since this bound in particular holds for all 1-Lipschitz functions h, this means that the Wasserstein distance between the distribution of W and the standard normal law has been bounded from above by the expression

$$\frac{2}{\pi}\, \mathbb{E}\left[\left|1 - \frac{1}{2\lambda}\mathbb{E}((W - W')^2|W)\right|\right] + \frac{1}{2\lambda}\, \mathbb{E}(|W - W'|^3). \tag{4}$$

This is a crude version of a bound. In the proofs of the orthogonal group results to be presented in what follows, more elaborate results will be required. In particular, in a situation which exhibits continuous rather than discrete symmetries, such as in a Lie group context, it may be an advantage to consider a continuous family of exchangeable pairs simultaneously, yielding theorems of the type given in Proposition 1 below. Nonetheless, the proof of the present crude version illustrates how the exchangeability condition and the regression condition fit together.

It should be noted that there is no guarantee at all that (4) will yield a reasonable bound. The true challenge is to find an exchangeable pair such that the moments of $W - W'$ which appear in (4) get small in the relevant limit, satisfying a regression condition for which λ does not become too small in this limit.

3 Stein's Method in the Multivariate Case

Against the backdrop of the sketch of univariate normal approximation that has been provided above, it does not seem straightforward to extend Stein's approach to the multivariate case. For instance, it is not obvious which differential operator should be used to construct a Stein equation. The most popular choice is the Ornstein-Uhlenbeck (OU) generator $L = \Delta - x \cdot \nabla$. To see that it serves this purpose, denote by (T_t) the operator semigroup corresponding to the OU process in $\mathbb{R}^d$, and by ν_d the d-dimensional standard normal distribution. It is known that the OU process is stationary w.r.t. ν_d. Hence, for f from a suitable class of test functions, one has that

$$\frac{d}{dt} \int T_t f d\nu_d = 0,$$

hence

$$\int L f d\nu_d = 0.$$

This observation was exploited by Götze [13] in 1991 in his treatment of the multivariate CLT in euclidean space. On the other hand, a multivariate version of the exchangeable pairs method is only a recent achievement. The handy version that will be presented below, due to E. Meckes [20], builds upon her previous joint work with Chatterjee [2] from 2008 as well as on a paper of Reinert and Röllin [24] that appeared in 2009.

For a vector $x \in \mathbb{R}^d$ let $\|x\|_2$ denote its euclidean norm induced by the standard scalar product on $\mathbb{R}^d$ that will be denoted by $\langle \cdot, \cdot \rangle$. For $A, B \in \mathbb{R}^{d \times k}$ let $\langle A, B \rangle_{\mathrm{HS}} := \mathrm{Tr}(A^T B) = \mathrm{Tr}(B^T A) = \mathrm{Tr}(AB^T) = \sum_{i=1}^d \sum_{j=1}^k a_{ij} b_{ij}$ be the usual Hilbert-Schmidt scalar product on $\mathbb{R}^{d \times k}$ and denote by $\| \cdot \|_{\mathrm{HS}}$ the corresponding norm. For random matrices $M_n, M \in \mathbb{R}^{k \times d}$, defined on a common probability space $(\Omega, \mathscr{A}, \mathbb{P})$, we will say that M_n converges to M in $\mathrm{L}^1(\| \cdot \|_{\mathrm{HS}})$ if $\|M_n - M\|_{\mathrm{HS}}$ converges to 0 in $\mathrm{L}^1(\mathbb{P})$.

For $A \in \mathbb{R}^{d \times d}$ let $\|A\|_{\mathrm{op}}$ denote the operator norm induced by the euclidean norm, i.e., $\|A\|_{\mathrm{op}} = \sup\{\|Ax\|_2 : \|x\|_2 = 1\}$. We now state a multivariate normal approximation theorem, due to E. Meckes ([20, Theorem 4]) that has been used in [10] to treat the multivariate CLT for traces of powers of Haar orthogonals. $Z = (Z_1, \ldots, Z_d)^T$ denotes a standard d-dimensional normal random vector, $\Sigma \in \mathbb{R}^{d \times d}$ a positive definite matrix and $Z_\Sigma := \Sigma^{1/2} Z$ with distribution $\mathrm{N}(0, \Sigma)$.

Proposition 1. *Let $W, W_t (t > 0)$ be $\mathbb{R}^d$-valued $L^2(\mathbb{P})$ random vectors on the same probability space $(\Omega, \mathscr{A}, \mathbb{P})$ such that for any $t > 0$ the pair (W, W_t) is exchangeable. Suppose there exist an invertible non-random matrix Λ, a positive definite matrix Σ, a random vector $R = (R_1, \ldots, R_d)^T$, a random $d \times d$-matrix S, a sub-σ-field $\mathscr{F}$ of $\mathscr{A}$ such that W is measurable w.r.t. $\mathscr{F}$ and a non-vanishing deterministic function $s :]0, \infty[\to \mathbb{R}$ such that the following three conditions are satisfied:*

$$\frac{1}{s(t)}\mathbb{E}[W_t - W|\mathscr{F}] \xrightarrow{t\to 0} -\Lambda W + R \; in \; L^1(\mathbb{P}). \tag{i}$$

$$\frac{1}{s(t)}\mathbb{E}[(W_t - W)(W_t - W)^T|\mathscr{F}] \xrightarrow{t\to 0} 2\Lambda\Sigma + S \; in \; L^1(\|\cdot\|_{\mathrm{HS}}). \tag{ii}$$

$$\lim_{t\to 0}\frac{1}{s(t)}\mathbb{E}\left[\|W_t - W\|_2^2 \, 1_{\{\|W_t - W\|_2^2 > \epsilon\}}\right] = 0 \; for \; each \; \epsilon > 0. \tag{iii}$$

Then

$$d_{\mathscr{W}}(W, Z_\Sigma) \le \|\Lambda^{-1}\|_{\mathrm{op}}\left(\mathbb{E}[\|R\|_2] + \frac{1}{\sqrt{2\pi}}\|\Sigma^{-1/2}\|_{\mathrm{op}}\,\mathbb{E}[\|S\|_{\mathrm{HS}}]\right). \tag{5}$$

It should be remarked that the more complete statement of this theorem given in [20, Theorem 4] also treats the case that Σ is only positive semidefinite.

4 Exchangeable Pairs and Quantitative Borel Type Theorems

As mentioned in the introduction, the historical precursor of CLTs for Haar distributed orthogonal matrices is Borel's result about the first coordinate of a random unit vector in euclidean space. This result is a special case (for $A^{(n)}$ specialized to a matrix with 1 in the $(1, 1)$ coordinate and 0 elsewhere) of the following result, due to D'Aristotile, Diaconis, and Newman [5]:

Theorem 1. *For $n \in \mathbb{N}$ choose (deterministic) $A^{(n)} \in \mathbb{R}^{n\times n}$ such that $Tr(A^{(n)}(A^{(n)})') = n$ and let $M_n \in O_n$ be distributed according to Haar measure. Then $Tr(A^{(n)}M_n)$ converges in distribution to $N(0, 1)$ as n tends to infinity.*

Quantitative versions of this result, both with a rate of order $\frac{1}{n-1}$ in total variation distance and with only slightly different constants, have been proven by E. Meckes in [19] and by Fulman and Röllin in [12]. In both papers the method of exchangeable pairs is applied, but the specific exchangeable pairs are quite different. Meckes uses a family (W, W_ϵ), where $W = \mathrm{Tr}(AM)$ and $W_\epsilon = \mathrm{Tr}(AM_\epsilon.)$ Here, for any $\epsilon > 0$, the matrix M_ϵ is defined by $M_\epsilon = HB_\epsilon H^T M$, where H is a Haar orthogonal independent of M, and

$$B_\epsilon = \begin{pmatrix} \sqrt{1-\epsilon^2} & \epsilon & & & \\ -\epsilon & \sqrt{1-\epsilon^2} & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{pmatrix}.$$

Fulman and Röllin, on the other hand, obtain a family $(\mathrm{Tr}(AM_0), \mathrm{Tr}(AM_t))$ (t >0) of exchangeable pairs from a Brownian motion on O_n that is started in Haar measure, which is the stationary distribution of this process. This construction will be explained more carefully below in the context of the multivariate CLT for traces of powers.

5 Exchangeable Pairs and Vectors of Traces of Powers

Let $M = M_n$ be distributed according to Haar measure on $K_n = \mathrm{SO}_n$ or $K_n = \mathrm{O}_n$. For $d \in \mathbb{N}, r = 1, \ldots, d$, consider the r-dimensional real random vector

$$W := W(d, r, n) := (f_{d-r+1}(M), f_{d-r+2}(M), \ldots, f_d(M)),$$

where

$$f_j(M) = \begin{cases} \mathrm{Tr}(M^j), & j \text{ odd}, \\ \mathrm{Tr}(M^j) - 1, & j \text{ even} \end{cases}$$

Theorem 2. *If $K_n = SO_n$ and $n \geq 4d + 1$ or $K_n = O_n$ and $n \geq 2d$, the Wasserstein distance between W and Z_Σ is*

$$d_{\mathscr{W}}(W, Z_\Sigma) = O\left(\frac{\max\left\{\frac{r^{7/2}}{(d-r+1)^{3/2}}, \ (d-r)^{3/2}\sqrt{r}\right\}}{n} \right). \quad (6)$$

In particular, for $r = d$ we have

$$d_{\mathscr{W}}(W, Z_\Sigma) = O\left(\frac{d^{7/2}}{n}\right),$$

and for $r \equiv 1$

$$d_{\mathscr{W}}(W, Z_\Sigma) = O\left(\frac{d^{3/2}}{n}\right).$$

If $1 \leq r = \lfloor cd \rfloor$ for $0 < c < 1$, then

$$d_{\mathscr{W}}(W, Z_\Sigma) = O\left(\frac{d^2}{n}\right).$$

For the special orthogonal group this result is proven in [10] (where the conditions on n in the special orthogonal and symplectic cases have been interchanged in the statement of the main result). The main steps of this proof will be indicated in Sect. 6 below, where it will also be shown how to adapt this strategy of proof to O_n in the place of SO_n. The univariate version of this theorem, including the construction of the exchangeable pair that will be explained below, is due to Fulman [11].

Remark 1. In the case of a single fixed power, the rate of convergence in Theorem 2 is clearly significantly worse than the exponential rate that was obtained by Johansson [15] in the context of limit theorems for Toeplitz determinants. The merit of Theorem 2 may be seen in the fact that it is multivariate and that the powers under consideration may grow with n. That the latter property yields practical benefits is demonstrated by Döbler and the author in [8]. There this property is used to prove, actually in the case of the unitary group, that the fluctuation of the linear eigenvalue statistic in (1) will converge to a normal limit with a rate of $\mathrm{O}(n^{-(1-\epsilon)})$ for any $\epsilon > 0$ if the test function f is of class C^∞.

6 On the Proof of the Multivariate Traces of Powers Result

The aim of this section is to summarize the main steps of the proof of Theorem 2, as provided in [10], for the special orthogonal group, and indicate how this argument can be supplemented to yield a proof of the full orthogonal case.

The overall strategy is to apply Proposition 1 to the traces of powers problem. To do so, one has to find a suitable family of exchangeable pairs. The following construction has been proposed by Fulman in [11] to treat the univariate case. See [14, Sect. V.4] for the relevant facts about diffusions on manifolds.

Let $(M_t)_{t\geq 0}$ be Brownian motion on the compact connected Lie group $K = \mathrm{SO}_n$, started in the Haar measure λ_K on K, which is known to be its stationary distribution. What is more, (M_t) is reversible w.r.t. λ_K. In particular, for any $t > 0$ and measurable f, $(f(M_0), f(M_t))$ is an exchangeable pair. Let $(T_t)_{t\geq 0}$ be the associated semigroup of transition operators on $\mathrm{C}^2(K)$ corresponding to (M_t). Its infinitesimal generator is the Laplace-Beltrami operator Δ, and the map $(t, g) \mapsto (T_t f)(g)$ satisfies the heat equation on K. Hence

$$\begin{aligned} T_t f(g) &= T_0 f(g) + t \left.\frac{d}{dt}\right|_{t=0} T_t f(g) + \mathrm{O}(t^2) \\ &= f(g) + t(\Delta f)(g) + \mathrm{O}(t^2), \end{aligned} \tag{7}$$

and basic Markov process theory yields an expansion that will be useful to establish the regression property that is fundamental for applying the method of exchangeable pairs:

$$\mathbb{E}[f(M_t)|M_0] = (T_t f)(M_0) = f(M_0) + t(\Delta f)(M_0) + \mathrm{O}(t^2). \tag{8}$$

To study traces of powers within this framework, it is useful to express them via power sum symmetric polynomials. To this end, consider $g \in \mathbb{C}^{n\times n}$ with eigenvalues $c_1, \ldots, c_n$ (with multiplicities). Then

$$\mathrm{Tr}(g^k) = c_1^k + \ldots + c_n^k,$$

i.e., the power sum symmetric polynomial $p_k = X_1^k + \ldots + X_n^k$ evaluated at $(c_1, \ldots, c_n) \in \mathbb{C}^n$. As p_k is symmetric in its arguments, we may unambiguously consider p_k as a function on $\mathbb{C}^{n\times n}$. For $k, l \in \mathbb{N}$ we write

$$p_{k,l}(g) := p_k(g)p_l(g) = \mathrm{Tr}(g^k)\mathrm{Tr}(g^l),$$

which is but a special instance of the general definition of power sum symmetric polynomials, as in [18]. Recalling the notation introduced in Sect. 5, we have that

$$f_j(M) = \begin{cases} p_j(M), & j \text{ odd}, \\ p_j(M) - 1, & j \text{ even}. \end{cases}$$

Setting

$$W := (f_{d-r+1}(M), f_{d-r+2}(M), \ldots, f_d(M))$$

and

$$W_t := (f_{d-r+1}(M_t), f_{d-r+2}(M_t), \ldots, f_d(M_t)),$$

we see from the discussion above that for any $t > 0$ the pair (W, W_t) is exchangeable.

We will have to verify that this family of exchangeable pairs satisfies the regression property in the refined form of (i), (ii) from Proposition 1. Obviously, the expansion (8) may be used to this end, as soon as one is able to describe the action of the Laplacian on the polynomials p_j in an explicit way. Fortunately, such formulae are available from work of Rains [22] and Lévy [16]. The latter reference provides a conceptual account of how they follow from an extension of Schur-Weyl duality to the universal enveloping algebra of the Lie algebra of K, hence to invariant differential operators on K. In concrete terms, we have the following lemma:

Lemma 1. *For the Laplacian Δ_{SO_n} on SO_n,*

$$\Delta_{SO_n} p_j = -\frac{(n-1)}{2} j p_j - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l} + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}. \tag{i}$$

$$\begin{aligned} \Delta_{SO_n} p_{j,k} = &-\frac{(n-1)(j+k)}{2} p_{j,k} - \frac{j}{2} p_k \sum_{l=1}^{j-1} p_{l,j-l} - \frac{k}{2} p_j \sum_{l=1}^{k-1} p_{l,k-l} \\ &- jkp_{j+k} + \frac{j}{2} p_k \sum_{l=1}^{j-1} p_{j-2l} + \frac{k}{2} p_j \sum_{l=1}^{k-1} p_{k-2l} + jkp_{j-k}. \end{aligned} \tag{ii}$$

The expansion (8) and Lemma 1 make it possible to identify the vector R and the matrices Λ and S in Proposition 1. By way of illustration, one may argue as follows:

Lemma 2. *For all* $j = d - r + 1, \dots, d$

$$\mathbb{E}[W_{t,j} - W_j | M] = \mathbb{E}[f_j(M_t) - f_j(M)|M] = t \cdot \left(-\frac{(n-1)j}{2} f_j(M) + R_j + O(t) \right),$$

where

$$R_j = -\frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \quad \textit{if } j \textit{ is odd,}$$

$$R_j = -\frac{(n-1)j}{2} - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \quad \textit{if } j \textit{ is even.}$$

Proof. First observe that always $f_j(M_t) - f_j(M) = p_j(M_t) - p_j(M)$, no matter what the parity of j is. By (8) and Lemma 1

$$\begin{aligned}\mathbb{E}[p_j(M_t) - p_j(M)|M] &= t(\Delta p_j)(M) + \mathrm{O}(t^2) \\ &= t \left(-\frac{(n-1)j}{2} p_j(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) \right. \\ &\quad \left. + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right) + \mathrm{O}(t^2),\end{aligned}$$

and the claim follows from the definition of f_j in the even and odd cases. □

From Lemma 2 and the compactness of the group K we conclude

$$\frac{1}{t}\mathbb{E}[W_t - W|M] \xrightarrow{t \to 0} -\Lambda W + R \text{ almost surely and in } \mathrm{L}^1(\mathbb{P}),$$

where $\Lambda = \mathrm{diag}\left(\frac{(n-1)j}{2},\ j = d - r + 1, \dots, d\right)$ and $R = (R_{d-r+1}, \dots, R_d)^T$. Thus, Condition (i) of Proposition 1 is satisfied, and we have identified Λ and R. The verification of (ii), and identification of Σ and S, is based on the following lemma, which is proven along the same lines as Lemma 2.

Lemma 3. *For all* $j, k = d - r + 1, \dots, d$

$$\mathbb{E}[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M))|M] = t\left(jkp_{j-k}(M) - jkp_{j+k}(M)\right) + O(t^2).$$

With Lemma 3 in hand, we obtain that

$$
\begin{aligned}
\frac{1}{t}\mathbb{E}[(W_{t,j}-W_j)(W_{t,k}-W_k)|M] &= \frac{1}{t}\mathbb{E}[(f_j(M_t)-f_j(M))(f_k(M_t)-f_k(M))|M] \\
&= \frac{1}{t}\mathbb{E}[(p_j(M_t)-p_j(M))(p_k(M_t)-p_k(M))|M] \\
&= jkp_{j-k}(M)-jkp_{j+k}(M)+\mathrm{O}(t^2) \\
&\overset{t\to 0}{\to} jkp_{j-k}(M)-jkp_{j+k}(M) \text{ a.s. and in } \mathrm{L}^1(\mathbb{P}),
\end{aligned}
$$

for all $j, k = 1, \ldots, d$. Noting that for $j = k$ the last expression is $j^2 n - j^2 p_{2j}(M)$ and that $2\Lambda\Sigma = \mathrm{diag}((n-1)j^2,\ j = d-r+1, \ldots, d)$ we see that Condition (ii) of Proposition 1 is satisfied with the matrices $\Sigma = \mathrm{diag}(d-r+1, \ldots, d)$ and $S = (S_{j,k})_{j,k=d-r+1,\ldots,d}$ given by

$$
S_{j,k} = \begin{cases} j^2(1-p_{2j}(M)), & j = k \\ jkp_{j-k}(M)-jkp_{j+k}(M), & j \neq k\,. \end{cases}
$$

To proceed further, i.e., to verify Condition (*iii*) of Proposition 1 and bound the right hand side of (5), one has to be able to integrate products of traces of powers with respect to Haar measure. Such formulae are available from the moment-based proof of the CLT for vectors of traces of powers given by Diaconis and Shahshahani in [7], and subsequent work. A version for the special orthogonal group, due to Pastur and Vasilchuk [21], is as follows:

Lemma 4. *If $M = M_n$ is a Haar-distributed element of SO_n, $n-1 \geq k_a$, $Z_1, \ldots, Z_r$ iid real standard normals, then*

$$
\mathbb{E}\left(\prod_{j=1}^{r}(Tr(M^j))^{a_j}\right) = \mathbb{E}\left(\prod_{j=1}^{r}(\sqrt{j}Z_j+\eta_j)^{a_j}\right) = \prod_{j=1}^{r} f_a(j), \tag{9}
$$

where

$$
f_a(j) := \begin{cases} 1 & \text{if } a_j = 0, \\ 0 & \text{if } ja_j \text{ is odd}, a_j \geq 1, \\ j^{a_j/2}(a_j-1)!! & \text{if } j \text{ is odd and } a_j \text{ is even}, a_j \geq 2, \\ 1 + \sum_{d=1}^{\lfloor a_j/2 \rfloor} j^d \binom{a_j}{2d}(2d-1)!! & \text{if } j \text{ is even}, a_j \geq 1. \end{cases}
$$

Here we have used the notations $(2m-1)!! = (2m-1)(2m-3)\cdot\ldots\cdot 3\cdot 1$,

$$
k_a := \sum_{j=1}^{r} ja_j, \quad \text{and} \quad \eta_j := \begin{cases} 1, \text{ if } j \text{ is even}, \\ 0, \text{ if } j \text{ is odd}. \end{cases}
$$

Using Lemma 4, it is tedious, but straightforward, to complete the proof of Theorem 2 in the special orthogonal case.

It should be evident from this sketch that an extension of Theorem 2 to the full orthogonal group will be proven once one has extended the construction of the exchangeable pair in a way that preserves the expansion (8), and verified the validity of Lemma 1 for the full orthogonal group. Lemma 4 for the full orthogonal group is due to Diaconis and Shahshahani [7], and the condition on n can be even weakened to $2n \geq k_a$ as a consequence of the invariant-theoretic proof given in [29] (which does not directly carry over to the special orthogonal group).

In a nutshell, the arguments involving the Laplacian extend to the full orthogonal group because the special orthogonal group is the connected component of the full orthogonal group that contains the identity. Consequently, both groups share the same Lie algebra, and the action of one-parameter semigroups, hence of differential operators, can be extended from SO_n to O_n in a canonical way. Although this has already been briefly discussed by Fulman and Röllin [12] in the context of linear functions of matrix entries, it is perhaps useful to close this survey by expanding a bit on this argument in the present situation.

The full orthogonal group has two connected components, consisting of orthogonal matrices of determinant 1 and -1, respectively. Writing J for the diagonal matrix $\mathrm{diag}(-1, 1, 1, \ldots, 1)$, the connected components of the group $K := \mathrm{O}_n$ are the cosets $K_+ := \mathrm{SO}_n$ and $K_- := J\,\mathrm{SO}_n$. For any $f \in \mathrm{C}(K)$ denote by $f_+ \in \mathrm{C}(K_+)$ and $f_- \in \mathrm{C}(K_-)$ its restrictions to K_+ and K_-, respectively. Then we may extend the family (T_t) of transition operators from $\mathrm{C}(K_+)$ to $\mathrm{C}(K)$ by requiring that for $f \in \mathrm{C}(K)$ there hold $(T_t f)_+ = T_t(f_+)$ and $(T_t f)_- = T_t(f_- \circ \tau_J) \circ \tau_J$, where τ_J is the left translation $(x \mapsto Jx)$. To verify that the process that corresponds to the extended semigroup is reversible w.r.t. Haar measure, one deduces from the invariance of Haar measure under translations and from reversibiliy of the process on the special orthogonal group that for $f, g \in \mathrm{C}(K)$ one has

$$\begin{aligned}
\int_{K_-} (T_t f)_-(x)\, g_-(x)\, \lambda_K(dx) &= \int_{K_-} ((T_t(f_- \circ \tau_J)) \circ \tau_J)(x)\, g_-(x)\, \lambda_K(dx) \\
&= \int 1_{K_+}(Jx)\, ((T_t(f_- \circ \tau_J))(Jx)\, g_-(x)\, \lambda_K(dx) \\
&= \int_{K_+} (T_t(f_- \circ \tau_J))(x)\, g_-(Jx)\, \lambda_K(dx) \\
&= \int_{K_+} (f_- \circ \tau_J)(x)\, T_t(g_- \circ \tau_J)(x)\, \lambda_K(dx) \\
&= \int_{K_-} f_-(x)\, (T_t(g_- \circ \tau_J)) \circ \tau_J)(x)\, \lambda_K(dx) \\
&= \int_{K_-} f_-(x)\, (T_t g)_-(x)\, \lambda_K(dx).
\end{aligned}$$

Since a Laplacian is invariant under translations, the action of $\Delta = \Delta_{K_+}$ on $C^2(K)$ in particular satisfies

$$\Delta(f \circ \tau_J) = (\Delta f) \circ \tau_J. \tag{10}$$

So we have that for any $x \in K_-$

$$\frac{d}{dt}(T_t f)_-(x) = \frac{d}{dt} T_t(f_- \circ \tau_J)(Jx) = \Delta(f_- \circ \tau_J)(Jx) = \Delta(f_-)(x).$$

That Lemma 1 extends to the full orthogonal group is a direct consequence of (10).

Acknowledgements The author has been supported by Deutsche Forschungsgemeinschaft via SFB-TR 12.

References

1. Barbour, A.D., Holst, L., Janson, S.: Poisson Approximation. Oxford Studies in Probability, vol. 2. The Clarendon Press/Oxford University Press, New York (1992). Oxford Science Publications
2. Chatterjee, S., Meckes, E.: Multivariate normal approximation using exchangeable pairs. ALEA Lat. Am. J. Probab. Math. Stat. **4**, 257–283 (2008)
3. Chatterjee, S., Fulman, J., Röllin, A.: Exponential approximation by Stein's method and spectral graph theory. ALEA Lat. Am. J. Probab. Math. Stat. **8**, 197–223 (2011)
4. Chen, L.H.Y., Goldstein, L., Shao, Q.M.: Normal Approximation by Stein's Method. Probability and Its Applications (New York). Springer, Heidelberg (2011)
5. D'Aristotile, A., Diaconis, P., Newman, C.M.: Brownian motion and the classical groups. In: Athreya, K., Majumdar, M., Puri, M., Waymire, E. (eds.) Probability, Statistics and Their Applications: Papers in Honor of Rabi Bhattacharya. IMS Lecture Notes Monograph Series, vol. 41, pp. 97–116. Institute of Mathematical Statistics, Beachwood (2003)
6. Diaconis, P., Evans, S.N.: Linear functionals of eigenvalues of random matrices. Trans. Am. Math. Soc. **353**(7), 2615–2633 (2001)
7. Diaconis, P., Shahshahani, M.: On the eigenvalues of random matrices. J. Appl. Probab. **31A**, 49–62 (1994). Studies in applied probability
8. Döbler, C.: A quantitative central limit theorem for linear statistics of random matrix eigenvalues. J. Theor. Probab. (2012+). http://link.springer.com/article/10.1007%2Fs10959-012-0451-2
9. Döbler, C.: Stein's method for exchangeable pairs for absolutely continuous, univariate distributions with applications to the polya urn model. Available on arXiv.org (2012, Preprint)
10. Döbler, C., Stolz, M.: Stein's method and the multivariate CLT for traces of powers on the classical compact groups. Electron. J. Probab. **16**(86), 2375–2405 (2011)
11. Fulman, J.: Stein's method, heat kernel, and traces of powers of elements of compact lie groups. Electron. J. Probab. **17**, 1–16 (2012)
12. Fulman, J., Röllin, A.: Stein's method, heat kernel, and linear functions on the orthogonal groups. Available on arXiv.org (2011, Preprint)
13. Götze, F.: On the rate of convergence in the multivariate CLT. Ann. Probab. **19**(2), 724–739 (1991)

14. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. North-Holland Mathematical Library, vol. 24, 2nd edn. North-Holland, Amsterdam (1989)
15. Johansson, K.: On random matrices from the compact classical groups. Ann. Math. (2) **145**(3), 519–545 (1997)
16. Lévy, T.: Schur-Weyl duality and the heat kernel measure on the unitary group. Adv. Math. **218**(2), 537–575 (2008)
17. Lytova, A., Pastur, L.: Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. Ann. Probab. **37**(5), 1778–1840 (2009)
18. Macdonald, I.G.: Symmetric Functions and Hall Polynomials. Oxford Mathematical Monographs, 2nd edn. The Clarendon Press/Oxford University Press, New York (1995). With contributions by A. Zelevinsky, Oxford Science Publications
19. Meckes, E.: Linear functions on the classical matrix groups. Trans. Am. Math. Soc. **360**(10), 5355–5366 (2008)
20. Meckes, E.: On Stein's method for multivariate normal approximation. In: Houdré, C., Koltchinskii, V., Mason, D.M., Peligrad, M. (eds.) High Dimensional Probability V: The Luminy Volume. Institute of Mathematical Statistics Collections, vol. 5, pp. 153–178. Institute of Mathematical Statistics, Beachwood (2009)
21. Pastur, L., Vasilchuk, V.: On the moments of traces of matrices of classical groups. Commun. Math. Phys. **252**(1–3), 149–166 (2004)
22. Rains, E.M.: Combinatorial properties of Brownian motion on the compact classical groups. J. Theor. Probab. **10**(3), 659–679 (1997)
23. Reinert, G.: Three general approaches to Stein's method. In: Barbour, A.D., Chen, L.H.Y. (eds.) An Introduction to Stein's Method. Lecture Notes Series, Institute for Mathematical Science, National University of Singapore, vol. 4, pp. 183–221. Singapore University Press, Singapore (2005)
24. Reinert, G., Röllin, A.: Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. Ann. Probab. **37**(6), 2150–2173 (2009)
25. Rinott, Y., Rotar, V.: On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. Ann. Appl. Probab. **7**(4), 1080–1105 (1997)
26. Stein, C.: A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (University of California, Berkeley, California, 1970/1971). Volume II: Probability Theory, pp. 583–602. University of California Press, Berkeley (1972)
27. Stein, C.: Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 7. Institute of Mathematical Statistics, Hayward (1986)
28. Stein, C.: The accuracy of the normal approximation to the distribution of the traces of powers of random orthogonal matrices. Technical Report 470, Stanford University, Department of Statistics (1995)
29. Stolz, M.: On the Diaconis-Shahshahani method in random matrix theory. J. Algebraic Comb. **22**(4), 471–491 (2005)

Part II
Iterated Random Functions

Large Deviation Tail Estimates and Related Limit Laws for Stochastic Fixed Point Equations

Jeffrey F. Collamore and Anand N. Vidyashankar

Abstract We study the forward and backward recursions generated by a stochastic fixed point equation (SFPE) of the form $V \stackrel{d}{=} A \max\{V, D\} + B$, where $(A, B, D) \in (0, \infty) \times \mathbb{R}^2$, for both the stationary and explosive cases. In the stationary case (when $\mathbf{E}[\log\ A] < 0$), we present results concerning the precise tail asymptotics for the random variable V satisfying this SFPE. In the explosive case (when $\mathbf{E}[\log\ A] > 0$), we establish a central limit theorem for the forward recursion generated by the SFPE, namely the process $V_n = A_n \max\{V_{n-1}, D_n\} + B_n$, where $\{(A_n, B_n, D_n) : n \in \mathbb{Z}_+\}$ is an i.i.d. sequence of random variables. Next, we consider recursions where the driving sequence of vectors, $\{(A_n, B_n, D_n) : n \in \mathbb{Z}_+\}$, is modulated by a Markov chain in general state space. We demonstrate an asymmetry between the forward and backward recursions and develop techniques for estimating the exceedance probability. In the process, we establish an interesting connection between the regularity properties of $\{V_n\}$ and the recurrence properties of an associated ξ-shifted Markov chain. We illustrate these ideas with several examples.

J.F. Collamore (✉)
Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
e-mail: collamore@math.ku.dk
Research supported in part by Danish Research Council (SNF) Grant No. 09-092331.

A.N. Vidyashankar
Department of Statistics, Volgeneau School of Engineering, George Mason University, Fairfax, VA 22030, USA
e-mail: avidyash@gmu.edu
Research supported in part by NSF grant DMS 1107108.

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_5,

1 Introduction

In this article, we consider stochastic fixed point equations (SFPE) of the form

$$V \stackrel{d}{=} f(V), \tag{1}$$

where f is a known random function and V is an unknown random variable in $\mathbb{R}$, independent of f. Such equations arise in a variety of applications, ranging from collective risk theory, queuing theory, financial time series modeling, and life insurance mathematics, to problems in branching processes and computer science. In these applications, it is often of interest to describe the tail behavior of the random variable V in (1).

Early work on this problem can be traced to the celebrated paper of Kesten [22], who considered the linear recursion

$$V \stackrel{d}{=} AV + B, \quad (A, B) \in \mathbb{R}^2, \tag{2}$$

in a higher dimensional setting, and applied the results to describe the tail behavior of certain martingale limits that arise in multi-type branching processes in random environments. In this context, he showed that if $\mathbf{E}[\log A] < 0$ (hereafter referred to as the stationary case) and appropriate regularity conditions are satisfied, then

$$\mathbf{P}\{V > u\} \sim C u^{-\xi} \quad \text{as} \quad u \to \infty, \tag{3}$$

where ξ is the nonzero solution to the equation $\mathbf{E}\big[A^{\xi}\big] = 1$. This result was later extended in $\mathbb{R}^1$ to more general recursions by Goldie [19]; see Sect. 2.1 below.

Identifying and characterizing the constant C in (3) is much more of a delicate affair compared to the problem of characterizing ξ. While ξ only depends on the multiplicative factor A of the given recursion, the value of the constant C depends on the *pair* (A, B) in (2) or, more generally, on the function f in (1). For the linear recursion (2), a nonrigorous approach—following earlier work by Yakin and Pollak [34] on likelihood ratio testing and sequential change point problems in statistics—was suggested in [32]. Quite recently, a rigorous solution was provided in [17] for the linear recursion (2) and independent random variables A and B using a coupling argument. A rigorous probabilistic solution—which holds for a general class that subsumes the models considered in Goldie [19]—was recently developed by the authors in [9].

In this article, we begin by giving a characterization of the constant C in the stationary case for the SFPE $V \stackrel{d}{=} A \max\{V, D\} + B$ and its extension to random maps. Next, we study the forward recursion $V_n := A_n \max\{V_{n-1}, D_n\} + B_n$ in the explosive case; that is, when $\mathbf{E}[\log A] > 0$. We show that $(V_n/P_n) \to \mathscr{W}$ as $n \to \infty$ w.p. 1 for a certain random variable P_n and establish a central limit theorem for V_n. Finally, we provide a nontrivial extension of the results in [9] to the Markov case; that is, the case when $\{(A_n, B_n, D_n) : n = 1, 2, \ldots\}$ is a Markov sequence and $V_n := A_n \max\{V_{n-1}, D_n\} + B_n$ (or a related backward recursion as described in Sects. 2.1 and 3.1 below). While Markovian extensions of Goldie's [19] result have been considered in [30] for the linear recursion (2) and in [8] for a wider class of backward recursions, a unified treatment encompassing an estimate for the *pair* (C, ξ) has not been systematically given. Here we present a unified approach, which builds upon work developed by the authors in [9] and earlier work of one of the authors in [8]. A key idea that facilitates this unification is the observation that, if the sequence $\{(A_n, B_n) : n = 1, 2, \ldots\}$ possesses a regenerative structure, then the process $\{V_n : n = 0, 1, \ldots\}$ inherits this regenerative property and the original forward recursion of the SFPE can be expressed as a forward recursion of another SFPE (but belonging to the same class of SFPEs under investigation); i.e., a forward recursion with a different driving sequence. A similar result also holds for the case of backward sequences. Expectedly, the driving sequence will now involve the regeneration times of the modulating Markov sequence.

We illustrate our results with a variety of examples drawn from insurance, financial mathematics, branching processes, and statistical inference.

2 Recursions Driven by i.i.d. Sequences

2.1 The Stationary Case

Our starting point is the SFPE

$$V \stackrel{d}{=} f(V) \equiv F_Y(V), \tag{4}$$

where $F_Y(V) \equiv F(V, Y)$ for some deterministic function $F : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$, assumed throughout the article to be measurable and to be continuous in its first component. In this representation, the random function f is determined by an environmental random vector Y and independent of V. Moreover, we implicitly assume the *shape condition*

$$F_Y(v) = Av + o(v) \quad \text{a.s.} \quad \text{as} \quad v \to \infty, \tag{5}$$

where A takes values on the positive real axis. In the following discussion, we will assume without loss of generality that $Y = (\log\, A, Y')$ for some $Y' \in \mathbb{R}^{d-1}$.

To assure that (4) has a stationary solution, we need the multiplicative factor A in (5) to be contracting; that is, $\mathbf{E}[\log A] < 0$. Let $\lambda(\alpha) := \mathbf{E}[A^\alpha]$ and $\Lambda(\alpha) := \log \lambda(\alpha)$ denote the moment generating function and the cumulant generating function of the random variable (r.v.) $\log\, A$, respectively, where $\alpha \in \mathbb{R}$. Also, let μ_A denote the probability distribution of A. For any function g, let $\mathrm{dom}(g)$ denote the domain of g. We assume

$$\mathbf{E}\big[A^\xi\big] = 1, \quad \text{for some } \xi \in (0, \infty) \cap \mathrm{dom}\, \Lambda'. \tag{6}$$

To determine the tail behavior of V, one approach is to observe that the SFPE (4) induces a renewal equation. Namely,

$$e^{\xi v}\mathbf{P}\{V > e^v\} = e^{\xi v}\Big\{\mathbf{P}\{V > e^v\} - \mathbf{P}\{AV > e^v\}\Big\} + e^{\xi v}\int_{\mathbb{R}} \mathbf{P}\{V > e^{v-x}\}\,\mu_A(dx).$$

Hence, setting $Z(v) = e^{\xi v}\mathbf{P}\{V > e^v\}$ and $z(v)$ equal to the first term on the right-hand side of the previous equation, we obtain that

$$Z(v) = z(v) + Z * \mu_{A,\xi}(v), \quad \text{where} \quad \mu_{A,\xi}(dx) = e^{\xi x}\mu_A(dx), \tag{7}$$

which is closely related to the renewal equation. Thus, if we *knew* that the function z were directly Riemann integrable, then the renewal theorem could be invoked to obtain that $Z(v) \to C$ as $v \to \infty$ and, consequently, $\mathbf{P}\{V > u\} \sim Cu^{-\xi}$ as $u \to \infty$.

Typically, it is impossible to verify that z is directly Riemann integrable. However, this assumption can be avoided by using a smoothing argument introduced in [19]. This techniques yields the following very general theorem, proved by Goldie ([19], Theorem 2.3), building upon the previous work of Kesten [22].

Theorem 1. *Assume that there exists a nonnegative random variable A which is nonarithmetic and satisfies* (6)*, and assume that*

$$\mathbf{E}\left[\left|\left(f(V)^+\right)^\xi - \left((AV)^+\right)^\xi\right|\right] < \infty.$$

Then

$$\lim_{u\to\infty} u^\xi \mathbf{P}\{V > u\} = C, \tag{8}$$

where

$$C = \frac{1}{\xi\lambda'(\xi)}\mathbf{E}\left[\left(f(V)^+\right)^\xi - \left((AV)^+\right)^\xi\right]. \tag{9}$$

While this estimate is easily obtained from the renewal theorem under weak assumptions, this approach has certain limitations. For instance, it is not possible to establish the finiteness and positivity of the constant C without further assumptions. Furthermore, the expression for C in (9) is defined in terms of V and, thus, is not particularly fruitful in practical problems. A useful characterization in terms of the forward process $\{V_n\}$ in (10) below would, in particular, facilitate statistical inference and numerical procedures such as importance sampling.

To address the above difficulties associated with (9), an alternative approach was recently developed in Collamore and Vidyashankar [9]. This approach is based on the observation that the process $V_n := F_{Y_n}(V_{n-1})$ (obtained via forward iterations of the SFPE, see below) is Markovian and behaves like a multiplicative random walk for large values of V_{n-1}. Thus, we may use nonlinear renewal theory to characterize this process for "large" V_{n-1}, and then adapt methods from Markov chain theory to quantify the discrepancy between these two processes.

To describe this approach, we first need to introduce the forward and backward sequences generated by a given SFPE. Let $\{Y_n : n = 1, 2, \ldots\}$ be an i.i.d. sequence having the same probability law as Y in (4). The *forward sequence* $\{V_n\}$ is defined by

$$V_n(v) = F_{Y_n} \circ F_{Y_{n-1}} \circ \cdots \circ F_{Y_1}(v), \quad n = 1, 2, \ldots, \quad V_0 = v; \tag{10}$$

while the *backward sequence* $\{Z_n\}$ is defined by

$$Z_n(z) = F_{Y_1} \circ F_{Y_2} \circ \cdots \circ F_{Y_n}(z), \quad n = 1, 2, \ldots, \quad Z_0 = z. \tag{11}$$

The Furstenberg-Letac principle states that—although the sample paths of the forward and backward sequences are manifestly different—the forward sequence converges in distribution to a random variable V provided that the backward sequence converges a.s. to a random variable Z and is independent of the initial value; furthermore, the distributions of V and Z are the same [18, 23]. This leads to the issue of determining which sequence—the forward or backward sequence—is more amenable for analysis, and this, of course, is problem-dependent. In Collamore and Vidyashankar [9], it is suggested that the forward sequence is preferable for understanding the tail behavior of V described by Theorem 1, and this approach also appears advantageous for the Monte Carlo simulation of these probabilities (cf. [10]). Generally speaking, the advantage of the forward sequence is that this process is a recurrent Markov chain and hence has useful ergodic properties (while the backward sequence converges a.s., which is useful when analytic, rather than probabilistic, methods are employed).

We now specialize to the quasi-linear recursion

$$V \stackrel{d}{=} F_Y(V), \quad F_Y(v) = A \max\{v, D\} + B, \tag{12}$$

where $Y \equiv (\log A, B, D) \in \mathbb{R}^3$. This SFPE is often called "Letac's Model E" and, as we will see in Sect. 2.3, has wide applied relevance. This class of models is roughly equivalent to the class considered by Goldie in [19].

First introduce the following regularity conditions.

Hypotheses:

(H_0) The random variable A has an absolutely continuous component with respect to Lebesgue measure with a nontrivial density in a neighborhood of $\mathbb{R}$.
(H_1) $\Lambda(\xi) = 0$ for some $\xi \in (0, \infty) \cap \operatorname{dom}(\Lambda')$.
(H_2) $\mathbf{E}\big[|B|^{\xi}\big] < \infty$ and $\mathbf{E}\big[(A|D|)^{\xi}\big] < \infty$.
(H_3) $\mathbf{P}\{A > 1, B > 0\} > 0$ or $\mathbf{P}\{A > 1, B \geq 0, D > 0\} > 0$.

Let P_V denote the transition kernel of $\{V_n\}$. Also, let $\mathscr{B}(\mathbb{R}^d)$ denote the Borel sets on $\mathbb{R}^d$, $d \geq 1$. Then under Hypotheses (H_0), (H_1), and (H_2), the forward process $\{V_n\}$ is a Markov chain satisfying the minorization condition

$$\delta \mathbf{1}_{\mathscr{C}}(v)\nu(E) \leq P_V(v, E), \quad v \in \mathbb{R}, \ E \in \mathscr{B}(\mathbb{R}), \qquad (\mathscr{M})$$

where δ is a positive constant, $\mathscr{C}$ a nontrivial set in $\mathbb{R}$, and ν a probability measure; see [9], Lemma 5.1. Hence by a classical result of Athreya and Ney [4] and Nummelin [28], it follows that the forward process $\{V_n\}$ admits a regeneration structure. More precisely, we can find a sequence of independent times $0 \leq T_0 < T_1 < \cdots$ such that:

(i) $\tau_i := T_i - T_{i-1}$ is an i.i.d. sequence, $i \in \mathbb{Z}_+$;
(ii) $\{V_{T_{i-1}}, \ldots, V_{T_i - 1}\}$ form independent blocks, $i \in \mathbb{Z}_+$;
(iii) $V_{T_i} \sim \nu$, independent of its past.

Let τ denote a typical regeneration time, that is to say, the first regeneration time assuming that regeneration has occurred at time zero. Then by Nummelin [29], p. 75, it follows as a consequence of the regeneration lemma that

$$\mathbf{P}\{V > u\} = \frac{\mathbf{E}[N_u]}{\mathbf{E}[\tau]}, \quad \text{where} \quad N_u := \sum_{n=0}^{\tau-1} \mathbf{1}_{(u,\infty)}(V_n). \qquad (13)$$

In particular, N_u counts the number of exceedances of $\{V_n\}$ occurring over a single regeneration cycle, and this number tends to zero as $u \to \infty$. Thus $\{N_u > u\}$ is a *rare event*, and quantifying $\mathbf{E}[N_u]$ for large u is a large deviation problem. It is natural to characterize this probability using a change of measure of the driving sequence $\{Y_n\}$ in (12).

Let μ denote the probability law of $Y \equiv (\log A, B, D)$, and let ξ be given as in (6), and define

$$\mu_\xi(E) = \int_E e^{\xi x} d\mu(x, y, z), \quad E \in \mathscr{B}(\mathbb{R}^3). \qquad (14)$$

Then μ_ξ is itself a probability measure and, with respect to this measure, we easily obtain that the process $\{V_n\}$ is *transient* ([9], Lemma 5.2). Set $\mathscr{T}_u = \inf\{n : Y_n > u\}$, and consider the dual change of measure:

$$\mathfrak{L}\big(\log A_n, B_n, D_n\big) = \begin{cases} \mu_\xi & \text{for } n = 1, \ldots, \mathscr{T}_u, \\ \mu & \text{for } n > \mathscr{T}_u. \end{cases} \qquad (15)$$

Let $\mathbf{E}_{\mathfrak{D}}[\cdot]$ denote the expectation with respect to this dual measure and $\mathbf{E}_\xi[\cdot]$ denote the expectation with respect to μ_ξ.

To estimate $\mathbf{E}_{\mathfrak{D}}[N_u]$, it is helpful to observe that this expectation splits into two parts, one describing the “short term” behavior over a regeneration cycle, and the other describing the "long term" behavior. By Collamore and Vidyashankar [9], Proposition 6.1, conditional on $V_0 \sim \nu$, we have that as $u \to \infty$,

$$\mathbf{E}\left[N_u\right] \sim \mathbf{E}_\xi\left[W^\xi \mathbf{1}_{\{\tau=\infty\}}\right] \cdot u^{-\xi} \mathbf{E}_{\mathfrak{D}}\left[N_u\left(\frac{V_{T_u}}{u}\right)^{-\xi}\right], \tag{16}$$

where $W := \lim_{n\to\infty} V_n/(A_1 \cdots A_n)$. It is worth noting that in a wide variety of examples, including all of the examples in Sect. 2.3 below, the random variable W reduces to the *perpetuity sequence*

$$W = V_0 + \frac{B_1}{A_1} + \frac{B_2}{A_1 A_2} + \cdots,$$

which, in (16), is *killed upon regeneration* of the transient process $\{V_n\}$.

Notice that under the shape condition (5), the process $\{V_n\}$ resembles a multiplicative random walk when this process is far away from the origin. Moreover, the exceedance probabilities for the multiplicative random walk are well-known from classical risk theory (cf. [2]) and for perturbed random walks from nonlinear renewal theory (cf. [33]). Trivially, in (16),

$$\left(V_{T_u}/u\right)^{-\xi} = \exp\left\{-\xi\left(\log V_{T_u} - \log u\right)\right\},$$

where the exponent on the right-hand side describes the overjump of the perturbed random walk $\{\log(V_n \vee 1)\}$ over a barrier at level $\log u$. Consequently, using extensions of results from [33], the second quantity on the right-hand side of (16) can be identified, as $u \to \infty$, as $u^{-\xi}\mathbf{E}[N_u^*]$, where N_u^* denotes the number of exceedances above level $\log u$ which occur for the random walk $S_n = \sum_{i=1}^n \log A_i$ over its regeneration cycle, that is to say, over a cycle starting at the origin and continuing until time $\tau^* = \inf\{n : S_n \le 0\}$. Thus, the first term on the right of (16) describes the *discrepancy* between the decay constant arising for the process $\{V_n\}$ and that arising for the corresponding multiplicative random walk.

To make these ideas rigorous, set $A_0 = 1$ and $B_0 \sim \nu$ (where ν is given as in ($\mathscr{M}$)). Now define the *perpetuity sequence associated with* $\{V_n\}$ by

$$Z_n^{(p)} = \sum_{i=0}^{n} \frac{B_i}{A_0 \cdots A_i}, \quad n = 0, 1, \ldots, \tag{17}$$

and define the *conjugate sequence associated with* $\{V_n\}$ by

$$Z_n^{(c)} = \min\left\{Z_n^{(p)}, 0, \bigwedge_{k=1}^{n} \sum_{i=0}^{k-1} \frac{B_i}{A_0 \cdots A_i} - \frac{D_k}{A_0 \cdots A_{k-1}}\right\}. \tag{18}$$

It is easily seen that both of these quantities are backward sequences in the sense of (11); specifically, (17), resp. (18) are generated by the recursions

$$F_Y^{(p)}(v) = \frac{v}{A} + \frac{B}{A} \quad \text{and} \quad F_Y^{(c)} = \frac{1}{A} \min \left\{ \check{D}, v \right\} + \frac{B}{A},$$

where $\check{D}_0 := -B_0$ and $\check{D}_i := -A_i D_i - B_i$ for $i = 1, 2, \ldots$.

Our main result in this section is the following:

Theorem 2. *Assume* (12)*, and suppose that* (H_0)*,* (H_1)*,* (H_2)*, and* (H_3) *are satisfied. Then*

$$\lim_{u\to\infty} u^{\xi} \mathbf{P}\{V > u\} = C \tag{19}$$

for a finite positive constant C*. Moreover,* $C = \lim_{n\to\infty} C_n$*, where*

$$C_n = \frac{1}{\xi \lambda'(\xi) \mathbf{E}[\tau]} \mathbf{E}_{\xi} \left[\left(\left(Z_n^{(p)} - Z_n^{(c)} \right)^+ \right)^{\xi} \mathbf{1}_{\{\tau > n\}} \right], \tag{20}$$

and $\mathfrak{R}_n := C - C_n = o(e^{-\epsilon n})$ *as* $n \to \infty$*, for some* $\epsilon > 0$.

For the proof of Theorem 2, see [9]. In particular, nonnegativity of the constant follows immediately from (H_3) and the fact that zero is contained in the collection minimized on the right-hand side of (18).

As demonstrated in [9], this method generalizes to a number of related problems. For example, it is shown that the method yields a useful upper bound, akin to the Lundberg inequality from insurance mathematics. Moreover, the method provides a simple characterization for the extremal index of $\{V_n\}$ (thus producing a considerable simplification of that developed for the special case of the ARCH(1) process in [15]). For details, see [9], Sect. 2.3.

Finally, the results can be generalized to a wider class of SFPEs. The main assumptions needed are the presence of a *cancellation condition*, namely

$$A \max\{v, D^*\} + B^* \le f(v) \le A \max\{v, D\} + B,$$

together with the Lipschitz condition

$$\sup_{v \neq w} \frac{|f(v) - f(w)|}{|v - w|} = L, \qquad \mathbf{E}[\log L] < 0,$$

where the approximating Letac models appearing in the cancellation condition are assumed to satisfy (H_0)–(H_3). Then we obtain a complete analog of Theorem 2, although the constant C is now expressed recursively and hence does not assume a simple analytical form. For details, see [9], Sect. 2.4.

2.2 The Explosive Case

While the previous section was concerned with the stationary case, namely when $\mathbf{E}[\log\ A] < 0$, in this section we present new results for the case $\mathbf{E}[\log\ A] \geq 0$. Note by Jensen's inequality that $\mathbf{E}[A] \geq 1$, and by the non-degeneracy of A it follows that $\mathbf{E}[A] > 1$. Our first result concerns the distributional behavior of $\log V_n$ when $1 < \mathbf{E}[A] < \infty$. In the following, we denote the variance of a random variable A by $\mathrm{Var}(A)$.

Theorem 3. *Let $\{V_n\}$ denote the forward process generated by Letac's Model E, as given in (10) and (12). Further assume that $\mathbf{E}[\log(|B|/A)] < \infty$, $\mathbf{E}[\log|D|] < \infty$, and $\mathbf{E}[\log\ A] > 0$. Then:*

(i) V_n diverges to infinity w.p. 1.
(ii) Setting $\mu = \mathbf{E}[\log\ A]$, $\sigma^2 = \mathrm{Var}(\log\ A)$, and $\mathscr{R}_n = n^{-\frac{1}{2}}\sigma^{-1}\{\log(V_n) - n\mu\}$, then $\{\mathscr{R}_n\}$ converges in distribution to a standard normal distribution as $n \to \infty$.

Proof. Using the forward recursion, we can express V_n as follows:

$$V_n = (A_1 \cdots A_n) \max\{J_{1,n}, J_{2,n}\}, \tag{21}$$

where

$$J_{1,n} = \sum_{i=0}^{n} \frac{B_i}{A_1 \cdots A_i}$$

and

$$J_{2,n} = \bigvee_{k=1}^{n} \left[\sum_{i=k}^{n} \frac{B_i}{A_1 \cdots A_i} + \frac{D_k}{A_1 \cdots A_{k-1}} \right] \equiv \bigvee_{k=1}^{n} J_{2,n,k}.$$

Now, by taking logarithms on both sides of (21), we get that

$$\mathscr{R}_n = \frac{\sum_{j=1}^{n} \left(\log\ A_j - \mu\right)}{\sqrt{n}\sigma} + \frac{\log\max\{J_{1,n}, J_{2,n}\}}{\sqrt{n}\sigma}. \tag{22}$$

It is worthwhile to notice that the second term is well-defined since, by our assumptions, $V_n > 0$ and $\prod_{i=1}^{n} A_i > 0$. To complete the proof, we need to show that the second term converges to zero in probability.

To this end, we begin by noticing that

$$J_{1,n} = \sum_{i=0}^{n} \frac{B_i}{A_i} \cdot \frac{1}{A_1 \cdots A_{i-1}}. \tag{23}$$

Under our assumptions, $J_{1,n}$ is a perpetuity generated by the driving sequence $(B/A, 1/A)$. Also $\mathbf{E}[\log(1/A)] < 0$. Hence $J_{1,n}$ converges to $J_{1,\infty}$ with probability one, by Theorem 1.3 of [1]. Furthermore, the random variable $J_{1,\infty}$ does not have an atom at zero. Next consider the term $J_{2,n}$. To this end, notice that

$$J_{2,n,k} \le J_{2,n,k}^{+} \equiv \sum_{i=k}^{n} \frac{|B_i|}{A_1 \cdots A_i} + \frac{|D_k|}{A_1 \cdots A_{k-1}}$$
$$\le \sum_{i=1}^{\infty} \frac{|B_i|}{A_1 \cdots A_i} + \sum_{k=1}^{\infty} \frac{|D_k|}{A_1 \cdots A_{k-1}} \equiv J,$$

where $J < \infty$ w.p. 1 by another application of Theorem 1.3 in [1]. Then

$$\sup_{n \in \mathbb{Z}_+} J_{2,n} \le J. \tag{24}$$

One can strengthen this bound to a convergence result for $J_{2,n}$ by utilizing

$$\max_{1 \le k \le n} \{J_{2,n,k} - J_{2,\infty,k}\} \to 0 \quad \text{w.p. } 1. \tag{25}$$

That is, using standard arguments, one can show that $J_{2,n}$ converges to $J_{2,\infty}$ w.p. 1, where $J_{2,\infty} = \max_k J_{2,k}$ and $J_{2,\infty}$ does not have an atom at 0. Thus it follows that

$$\lim_{n \to \infty} \frac{\log \max\{J_{1,n}, J_{2,n}\}}{\sqrt{n}} = 0 \tag{26}$$

in probability, which completes the proof of the theorem. □

From the proof of the above theorem, we can extract the path properties of $\{V_n\}$. We state this as a theorem.

Theorem 4. *Let $\{V_n\}$ denote the forward process generated by Letac's Model E, as given in* (10) *and* (12)*. Further assume that* $\mathbf{E}[\log(|B|/A)] < \infty$*,* $\mathbf{E}[\log|D|] < \infty$*, and* $\mathbf{E}[\log A] > 0$*. Then*

$$\lim_{n \to \infty} \frac{V_n}{A_1 \cdots A_n} = V_\infty \quad \text{w.p. } 1, \tag{27}$$

where the limit V_∞ is nondegenerate.

The above theorem studies the asymptotic behavior of $\{V_n\}$ under a random normalization instead of a deterministic normalization. Under further strong assumptions, [20] studies the path properties of $\{V_n\}$ under a deterministic normalization. (Consult [20] and references therein for further central limit theorems related to the explosive case.) Extensions of this idea for Letac's Model E are currently being investigated by the authors.

2.3 Examples and Applications

We now turn to a few examples.

Example 1. The simplest example is the reflected random walk,

$$W_n = (X_n + W_{n-1})^+, \quad n = 1, 2, \ldots, \quad W_0 = 0, \tag{28}$$

where $\{X_i\}$ is an i.i.d. sequence of random variables, which is equivalent to the multiplicative process $V_n = A_n \max\{V_{n-1}, 1\}$, where $A_n = \exp X_n$ and $V_n = \exp W_n$.

Extremes of these processes play a prominant role in queuing theory (cf. [21]) and in collective risk theory. In the classical ruin problem of Lundberg [24] and Cramér [12], one lets u denote the initial capital of the company, c the constant rate of premiums income, and $\{\zeta_i\}$ the i.i.d. claims losses, which are assumed to arise according to a Poisson process, $\{N_t\}$. Then the total capital of the company at time t is given by

$$Y_t = u + ct - \sum_{i=1}^{N_t} \zeta_i. \tag{29}$$

Now consider the probability of ruin, namely $\Psi(u) := \mathbf{P}\{Y_t < 0, \text{ for some } t \geq 0\}$. Using Sparre-Andersen's random walk representation of the risk process together with classical duality, this probability may be equated to $\mathbf{P}\{W > u\}$, where $W := \lim_{n\to\infty} W_n$ and $X_i := \zeta_i - c\tau_i$ in (28). For details, see [2].

Example 2. Consider a modification of the previous example, where the insurance company invests its excess capital, earning i.i.d. returns $\{R_n\}$ on these investments. The total capital of the company is then the solution to the recursive sequence of equations

$$\tilde{Y}_n = R_n \tilde{Y}_{n-1} - L_n, \quad n = 1, 2, \ldots, \quad \tilde{Y}_0 = u, \tag{30}$$

where $L_n := -(Y_n - Y_{n-1})$ are the discrete-time losses of the insurance business, governed by the Cramér-Lundberg process described in (29). Next, define the discounted loss process at time n to be the perpetuity sequence

$$\mathscr{L}_n := \frac{L_1}{R_1} + \frac{L_2}{R_1 R_2} + \cdots + \frac{L_n}{R_1 \cdots R_n}. \tag{31}$$

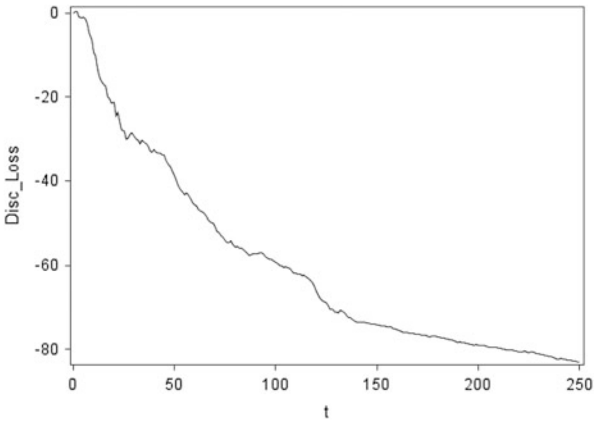


Fig. 1 A sample path of the cumulative loss process. Ruin occurs when this negative-drift process reaches a positive barrier at u, where u is the initial capital. This contrasts with the backward process in Example 1, which is a multiplicative random walk process

Then by a simple argument, the probability of ruin is equivalent to $\tilde{\Psi}(u) := \mathbf{P}\{\mathscr{L}_n > u, \text{ for some } n\}$. The process $\{\mathscr{L}_n\}$ is illustrated in Fig. 1 and is a backward recursion generated by the function $F_Y(v) = v/R + L/R$. In contrast with the previous example—whose backward recursion can be shown to be a random walk—the backward process appearing here has dependent increments and is evidently not Markovian.

To determine the probability of ruin, we need to solve for the tail of the r.v. $\mathscr{L} \equiv \sup_{n\in\mathbb{Z}_+} \mathscr{L}_n$. To this end, observe that $R_2^{-1}L_2 + \cdots + (R_2 \cdots R_{n+1})^{-1}L_{n+1} \stackrel{d}{=} \mathscr{L}_n$, and hence by (31),

$$\mathscr{L}_n \stackrel{d}{=} B + A\mathscr{L}_{n-1}, \quad \text{where} \quad A = \frac{1}{R_1} \text{ and } B = \frac{L_1}{R_1}. \tag{32}$$

Now setting $\tilde{\mathscr{L}} = \left(\sup_{n\in\mathbb{Z}_+} \mathscr{L}_n\right) \vee 0$ yields the SFPE

$$\tilde{\mathscr{L}} \stackrel{d}{=} \left(A\tilde{\mathscr{L}} + B\right)^+. \tag{33}$$

(Alternatively, by a slight variation of this argument, one can also show that $\mathscr{L}$ satisfies the SFPE $\mathscr{L} \stackrel{d}{=} A\max\{\mathscr{L}, 0\} + B$. However, the tail behavior of $\mathscr{L}$ is identical to that of $\tilde{\mathscr{L}}$.)

As with the previous example, there exists an interesting duality in the sense of Siegmund [31] or Asmussen and Sigman [3]. Namely the process $\{\tilde{Y}_n\}$ is dual to the forward process generated by the SFPE (33); cf. [3], p. 12. While the forward process is Markovian, the process $\{\tilde{Y}_n\}$ is actually studied via the complicated backward process $\{\mathscr{L}_n\}$. As the forward process is simpler than this backward process, it is convenient to first convert the backward process, via its SFPE, into a forward process.

Example 3. Consider a single-type branching process in a random environment. Then the population size at time n is given by

$$Z_n = \left(\sum_{i=1}^{Z_{n-1}} \eta_{n,i} \right) + Q_n,$$

where $\{\eta_{n,i} : i = 1, \ldots, Z_{n-1}\}$ represents the number of children in the nth generation, and Q_n represents the number of immigrants in the nth generation. Assume that the probability laws of these quantities are random, modulated by an i.i.d. environmental sequence $\{\zeta_n\}$. Thus $\eta_{n,i} \sim \mathbf{p}(\zeta)$ for all i, and $Q_n \sim \mathbf{q}(\zeta)$ independent of $\{\eta_{n,i} : i \geq 1, n \geq 1\}$. Let $\mathfrak{F}_n$ denote the σ-field generated by $\{\zeta_0, \ldots, \zeta_n\}$, and let $\mathfrak{F}_\infty$ denote the σ-field generated by $\{\zeta_0, \zeta_1, \ldots\}$, and consider $V_n := \mathbf{E}[Z_n | \mathfrak{F}_\infty]$. It is easily seen using the branching property that

$$V_n = \mathbf{E}[\eta_{n,i} | \mathfrak{F}_\infty] V_{n-1} + \mathbf{E}[Q_n | \mathfrak{F}_\infty].$$

Assuming that $\mathbf{E}[\log \mathbf{E}[\eta_{n,1} | \mathfrak{F}_n]] < 0$ and letting $n \to \infty$ in the above equation, one obtains the linear recursion

$$V \stackrel{d}{=} AV + B, \tag{34}$$

where $V := \lim_{n\to\infty} V_n$. A multidimensional extension of this model was the focus of the well-known paper of Kesten [22].

The recursion (34) also appears in many other settings, including the ARCH(1) and GARCH(1,1) processes used for financial time series modeling (cf. [5, 16]), or the perpetuity sequences used for modeling the future liabilities of a life insurance company. For details, see [8], Sect. 3.

It is worth noting that in all of the above examples, it can be shown that the conjugate sequence in Theorem 2 may be taken to be zero, and thus the constant C is determined solely by a perpetuity sequence which is killed in the event that $\{V_n\}$ returns, in the ξ-shifted measure, to its regeneration set; cf. [9], Corollary 2.1.

We conclude this section by remarking that in several real applications in the stationary case, simulation methods are typically used to obtain the tail probabilities, and this can be computationally expensive. Thus, a precise description of the tails of V facilitates inference concerning the extreme percentiles of the distribution of V. Such estimates are of considerable interest in risk management. Estimation of ξ has received much attention in the literature in risk theory, and detailed information concerning the Edgeworth expansion is available (see for instance [6]). However, inference concerning C and the pair (C, ξ) are not addressed in the literature. In an ongoing work, we use the change of measure arguments developed in [9] to describe a method-of-moments approach to the joint estimation of (C, ξ). It turns out that in finite samples, the correlation between estimates of C and ξ is negative, and a detailed mathematical description is currently being carried out in [11].

3 Recursions Driven by Markov-Dependent Sequences

Characterizing the constant C in the Markov case is much more challenging than the i.i.d. case. As explained in Sect. 1, the values of the constant C depend on the entire driving sequence $\{(\log A_n, B_n, D_n) : n = 1, 2, \ldots\}$ and their inherent dependence structure. In the i.i.d. case, the recursions simplify, but in the Markov case, an important asymmetry is introduced between the forward and backward sequences, which we explain below in Sect. 3.1.

In spite of this complication, it is possible to adopt some of the principles from the i.i.d. case by utilizing the regeneration technique of Athreya-Ney-Nummelin [4, 28], which states that the process contains independent blocks of random length which are i.i.d. Using this observation, we may derive appropriate SFPEs for Markov recursions and apply the results of the previous section. This is possible since, somewhat unexpectedly, the k-step composition of a recursion driven by Letac's Model E *retains* the general form of Letac's Model E, but with a new driving sequence (which here will be indexed by the regeneration times of the Markov chain $\{X_n\}$).

While the form of this constant will necessarily be complicated, we note in Remark 1 below that this constant reduces to the same general form as in the i.i.d. case in some important examples.

3.1 Forward and Backward Markov Sequences

Consider the forward and backward sequences associated with the SFPE $F_Y(v) = A \max\{v, D\} + B$, where $Y \equiv (\log A, B, D)$. By (10), the forward sequence is given by

$$V_n = A_n \max\{V_{n-1}, D_n\} + B_n, \quad n = 1, 2, \ldots, \quad V_0 = v, \tag{35}$$

while by (11), the backward sequence is given by

$$Z_n = A_1 \max\left\{Z_{n-1}^{(1)}, D_1\right\} + B_1, \quad n = 1, 2, \ldots, \quad Z_0 = z, \tag{36}$$

where $Z_{n-1}^{(1)}$ is defined as Z_{n-1}, but with $\{Y_1, \ldots, Y_{n-1}\}$ replaced with $\{Y_2, \ldots, Y_n\}$, i.e., the driving sequence is shifted forward by one unit of time. In contrast to the previous sections, we now assume that this driving sequence is Markov-dependent, that is, $Y_n = g(X_n)$, where $\{X_n\}$ is a Markov chain taking values in a state space

$(\mathbb{S}, \mathscr{S})$ and $g : \mathbb{S} \to \mathbb{R}^3$. We will assume throughout this discussion that $\{X_n\}$ is aperiodic, irreducible (with respect to its maximal irreducibility measure φ), and countably generated. Thus, we adopt the basic set-up described in [29] or [25].

Markov-dependent forward and backward sequences arise widely in applications. Natural examples in the forward case include the GARCH(1,1) or ARCH(1) processes or branching processes with Markov-dependent innovations. While such examples are easily motivated in the case of the forward sequence (35), the utility of the backward sequence (36) is less transparent but can be motivated by a couple of elementary examples.

To this end, we return to the ruin problem with investments described in Example 2. In that example, we observed that the probability of ruin is $\mathbf{P}\{\mathscr{L} > u\}$, where $\mathscr{L} \equiv \sup_{n \in \mathbb{Z}_+} \mathscr{L}_n$ and $\mathscr{L}_n$ denotes the discounted losses of the company which accumulate by time n. Now by iterating the sequence $\{\mathscr{L}_n\}$, we obtain after an elementary argument that

$$\mathscr{L}_n = A \max\left\{\mathscr{L}_{n-1}^{(1)}, 0\right\} + AL,$$

where $\mathscr{L}_{n-1}^{(1)}$ denotes that the driving sequence has been shifted forward by one unit in time; cf. the discussion following (33). Now if we set the initial state $z = 0$ and $(B, D) = (AL, 0)$, then this last equation assumes the same form as (36), and our objective is to determine the maximum of the backward sequence $\{\mathscr{L}_n\}$.

A second example is the classical ruin problem mentioned in Example 1. In that example, the corresponding backward process is the multiplicative random walk $\mathscr{W}_n := A_1 \cdots A_n$ (where $A_i = \exp X_i$ is defined as in Example 1), and ruin can be shown to occur when $\tilde{\mathscr{L}} > e^u$, where

$$\tilde{\mathscr{L}} := \sup_{n \in \mathbb{Z}_+} \tilde{\mathscr{L}}_n \quad \text{and} \quad \tilde{\mathscr{L}}_n := \max\{\mathscr{W}_1, \ldots, \mathscr{W}_n\}.$$

By repeating the above argument, we obtain that

$$\tilde{\mathscr{L}}_n = A \max\left\{\tilde{\mathscr{L}}_{n-1}^{(1)}, 1\right\},$$

which has the same form as (36) after setting $z = 1$ and $(B, D) = (0, 1)$.

In the above examples we see, rather generally, that forward sequences often arise in problems involving the steady-state limit of a given recursion, while backward sequences typically arise in problems involving maxima. Heuristically, these can be viewed as dual problems in the sense of Siegmund [31] or Asmussen and Sigman [3].

To analyze the processes (35) and (36), we again utilize the regeneration technique of Athreya-Ney-Nummelin, *applied to the Markov chain* $\{X_n\}$ (rather than to $\{V_n\}$), to derive an SFPE having the same form as (12); thus, in particular, Theorem 2 can be applied to describe the stationary limiting behavior, also in the Markov-modulated case.

Let P denote the transition kernel of $\{X_n\}$, and introduce the minorization condition

$$h(x)\eta(E) \le P(x, E), \quad x \in \mathbb{S}, \; E \in \mathscr{S}, \tag{$\mathscr{M}_0$}$$

where $h(x) = \delta_0 \mathbf{1}_{\mathscr{C}_0}(x)$ for some nontrivial set $\mathscr{C}_0$ and positive constant δ_0, where η is a probability measure on $(\mathbb{S}, \mathscr{S})$.

The regeneration lemma [4, 28] yields the existence of a sequence of stopping times $K_0, K_1, \ldots$ such that

(i) $\kappa_i := K_i - K_{i-1}$ is an i.i.d. sequence, $i \ge 1$;
(ii) $\{(X_{K_{i-1}}, Y_{K_{i-1}}), \ldots, (X_{K_i-1}, Y_{K_i-1})\}$ form independent blocks, $i \ge 1$.
(iii) $X_{K_i} \sim \eta$, independent of the past.

A standard calculation shows that both the recursions (35) and (36) have a nice compositional property, namely, if we calculate the k-step evolution of the process, then it can be viewed as a recursion involving the function $F_Y(v) = A \max\{v, D\} + B$, but with $Y_n \equiv (\log A_n, B_n, D_n)$ replaced with a new driving sequence. Specifically, after a tedious computation, we obtain that the k-step evolution satisfies

$$V_k = \max\left\{\hat{\mathscr{A}}\, V_0, \hat{\mathscr{D}}\right\} + \hat{\mathscr{B}}, \quad k \in \mathbb{Z}_+, \tag{37}$$

where

$$\hat{\mathscr{A}} := A_1 \cdots A_k,$$

$$\hat{\mathscr{B}} := \sum_{i=1}^{k} B_i (A_{i+1} \cdots A_k),$$

$$\hat{\mathscr{D}} := \bigvee_{j=1}^{k} \left[D_j (A_j \cdots A_k) - \sum_{i=1}^{j-1} B_i (A_{i+1} \cdots A_k) \right]$$

(where $A_1 \cdots A_{j-1} = 1$ when $j = 1$). Next observe that (37) has the same general form as (35), but with (A, B, AD) replaced with $(\hat{\mathscr{A}}, \hat{\mathscr{B}}, \hat{\mathscr{D}})$. A similar expression is also obtained for the backward recursion.

This compositional property now carries over to the stopping times $K_i - 1$. Thus, in the case of the forward recursion, we obtain for $\mathscr{V}_i := V_{K_i-1}$ that

$$\mathscr{V}_i = \max\{\mathscr{A}_i \mathscr{V}_{i-1}, \mathscr{D}_i\} + \mathscr{B}_i, \quad i = 1, 2, \ldots. \tag{38}$$

In this recursion, the driving sequence $\mathscr{Y}_i := (\log \mathscr{A}_i, \mathscr{B}_i, \mathscr{D}_i)$ is defined as $\hat{\mathscr{Y}} := (\log \hat{\mathscr{A}}, \hat{\mathscr{B}}, \hat{\mathscr{D}})$ in (37), except that the deterministic interval [1, k] in (37) must be replaced with the random interval $[K_{i-1}, K_{i-1}]$ in (38). (When $i = 0$, $K_{i-1} \equiv 1$ in these definitions.) Thus,

$$\begin{aligned}\mathscr{A}_i &:= A_{K_{i-1}} \cdots A_{K_i - 1},\\ \mathscr{B}_i &:= \sum_{j=K_{i-1}}^{K_i - 1} B_j (A_{j+1} \cdots A_{K_i - 1}),\\ \mathscr{D}_i &:= \bigvee_{j=K_{i-1}}^{K_i - 1} \left[D_j (A_j \cdots A_{K_i - 1}) - \sum_{k=K_{i-1}}^{j-1} B_k (A_{k+1} \cdots A_{K_i - 1}) \right]. \end{aligned} \tag{39}$$

From (38), we obtain that $\mathscr{V} := \lim_{i\to\infty} \mathscr{V}_i$ satisfies the SFPE

$$\mathscr{V} \stackrel{d}{=} \max \{\mathscr{A}\mathscr{V}, \mathscr{D}\} + \mathscr{B}, \tag{40}$$

where $(\mathscr{A}, \mathscr{B}, \mathscr{D}) \stackrel{d}{=} (\mathscr{A}_i, \mathscr{B}_i, \mathscr{D}_i)$, and, as we will observe more formally below, $\{\mathscr{V}_i\}$ has the same steady-state limit as the original process $\{V_n\}$ and, thus, this steady-state limit is characterized as the solution to (40).

In the case of the backward sequence, the regeneration technique works similarly. It is just a matter of writing down the iterates, but now backward in time, to obtain for $\mathscr{Z}_i := \max\{Z_n : 0 \le n \le K_i - 1\}$ that

$$\mathscr{Z}_i = \max \left\{ \mathscr{A}_1 \mathscr{Z}_{i-1}^{(1)}, \mathscr{D}_1 \right\} + \mathscr{B}_1, \quad i = 1, 2, \ldots, \tag{41}$$

where, following our usual convention, $\mathscr{Z}_{i-1}^{(1)}$ has the same distribution as $\mathscr{Z}_{i-1}$ but with the relevant driving sequence shifted forward by one unit in time, and for each positive integer i,

$$\begin{aligned}\mathscr{A}_i &:= A_{K_{i-1}} \cdots A_{K_i - 1},\\ \mathscr{B}_i &:= \sum_{j=K_{i-1}}^{K_i - 1} (A_{K_{i-1}} \cdots A_{j-1}) B_j,\\ \mathscr{D}_i &:= \bigvee_{j=K_{i-1}}^{K_i - 1} \left[(A_{K_{i-1}} \cdots A_j) D_j - \sum_{k=j+1}^{K_i - 1} (A_{K_{i-1}} \cdots A_{k-1}) B_k \right]. \end{aligned} \tag{42}$$

(Once again, when $i = 0$, $K_{i-1} \equiv 1$ in these definitions.)

Setting $\mathscr{Z} = \sup_i \mathscr{Z}_i$ in (41), we obtain the SFPE

$$\mathscr{Z} \stackrel{d}{=} \max \{\mathscr{A}\mathscr{Z}, \mathscr{D}\} + \mathscr{B}, \tag{43}$$

where $(\mathscr{A}, \mathscr{B}, \mathscr{D}) \stackrel{d}{=} (\mathscr{A}_i, \mathscr{B}_i, \mathscr{D}_i)$.

It is important to observe that we obtain *different* distributions for $(\mathscr{B}, \mathscr{D})$ in the forward and backward cases, even though we have started with the same recursion, $F_Y(v) = A \max\{v, D\} + B$, to generate the sequences (35) and (36). Thus, there is an asymmetry between the forward and backward sequences, although the multiplicative term $\mathscr{A}$—which determines the polynomial rate of decay—remains the same. (We note that this feature appears even in the polynomial models with i.i.d. recursions, as described in [9], Example 3.5).

We may now apply Theorems 2 and 3 directly to the SFPEs (38) and (43), but before doing so we will need to verify that the required moment conditions are satisfied. Since our random variables are formed over regeneration cycles, this issue is somewhat subtle and we address it in the next section.

3.2 *Characterizing Moments over Regeneration Cycles*

3.2.1 Moment Properties of $\mathscr{A}$

Let $g_A(X_n) = \log \ A_n$, and for each $\alpha \in \mathbb{R}$ define

$$\hat{P}_\alpha(x, E) := \int_E e^{\alpha g_A(x)} P(x, dy) \quad \text{and} \quad \hat{P}_\alpha^k = \hat{P}_\alpha \circ \hat{P}_\alpha^{k-1}, \quad k > 1.$$

Let $(\lambda(\alpha))^{-1}$ denote the convergence parameter of the kernel $\hat{P}_\alpha$ (for the definition, see [29], p. 27), and let $\Lambda(\alpha) = \log \lambda(\alpha)$. Set $S_n = \sum_{i=1}^n \log \ A_i$, and define

$$\Gamma(\alpha) = \limsup_{n \to \infty} \frac{1}{n} \log \mathbf{E}\left[e^{\alpha S_n}\right], \quad \alpha \in \mathbb{R}.$$

Roughly, the convergence parameter measures the growth rate of $\hat{P}_\alpha^k(x, E)$ as $k \to \infty$, where E is a "small set" satisfying ($\mathscr{M}_0$), while the "Gärtner-Ellis" limit $\Gamma(\alpha)$ measures the growth rate of this quantity when $E = \mathbb{S}$. It is well known that $\Lambda(\alpha) \le \Gamma(\alpha)$ for all α [8, 27]. Now assume that

$$\Gamma(\xi) = 0, \quad \text{for some } \xi \in (0, \infty) \cap \text{dom } \Lambda'. \tag{44}$$

Then it follows after a short argument that $\Lambda(\xi) = 0$; see [8], p. 1426. Then by [27], we have under appropriate conditions that $1 = \mathbf{E}\big[(A_{K_{i-1}} \cdots A_{K_i - 1})^\xi\big] := \mathbf{E}\big[\mathscr{A}^\xi\big]$. (See, in particular, Lemma 4.1 of [27] and its proof.) Thus, the solution ξ to (44) yields the polynomial decay rate in Theorem 2, provided that appropriate moment conditions are satisfied.

Consequently, we obtain an explicit characterization of the decay constant in the Markov case, which is now the solution to the equation $\Lambda(\xi) = 0$, but where the cumulant generating function is replaced with the function Λ now derived from the convergence parameter or, alternatively, with the Gärtner-Ellis limit (as would be expected from the large deviation theory for Markov chains).

3.2.2 Moment Properties of $(\mathscr{B}, \mathscr{D})$: Preliminary Considerations

For notational convenience, assume that regeneration occurs at time zero, and let κ denote the subsequent regeneration time. Then, using the above expressions for $(\mathscr{B}, \mathscr{D})$ we see that in the forward case, the required moment conditions (H_2) will be satisfied provided that $\mathbf{E}\big[\mathscr{B}_f^{\xi}\big] < \infty$, where

$$\mathscr{B}_f := \sum_{i=0}^{\kappa-1} \tilde{B}_i (A_{i+1} \cdots A_{\kappa-1}), \quad \tilde{B}_i := |B_i| + |A_i D_i| \tag{45}$$

and in this expression, we take (A_0, B_0, D_0) to have the distribution of this triplet upon regeneration. Similarly, in the backward case, it is sufficient to verify that $\mathbf{E}\big[\mathscr{B}_b^{\xi}\big] < \infty$, where

$$\mathscr{B}_b := \sum_{i=0}^{\kappa-1} (A_0 \cdots A_{i-1}) \tilde{B}_i, \quad \tilde{B}_i = |B_i| + |A_i D_i|. \tag{46}$$

These last two equations are manifestations of nearly the same mathematical quantity, as can be seen by constructing the time-reversed Markov process (whose existence is assured by [26]). Thus we extend $\{V_n : n \in \mathbb{N}\}$ to a doubly-infinite sequence $\{V_n : n \in \mathbb{Z}\}$, where these two sequences are identical for $n \in \mathbb{N}$. Then, by comparing (45) to this same quantity but over its prior regeneration cycle—that is, a cycle commencing at time $\tilde{\kappa} < 0$ and terminating at time 0—we obtain that

$$\mathscr{B}_f \stackrel{d}{=} \sum_{i=1}^{-\tilde{\kappa}} \tilde{B}_i (A_{i+1} \cdots A_1). \tag{47}$$

As with $\mathscr{B}_b$, this quantity may be viewed as a perpetuity sequence, but now computed *backward* in time (and shifted by one time unit compared with (46)). Consequently, the mathematical analysis is similar for the forward and backward sequences and, to avoid repetition, we will focus on verifying moment conditions for backward sequences in the sequel.

3.2.3 The Moments of $\mathscr{B}_b$ Under Some Simplifying Assumptions

Our next objective is to relate the moments of $\mathscr{B}_b$ to the moments of the regeneration times of the ξ-shifted Markov chain, whose finiteness would then be assured if the Markov chain were $(\lambda(\alpha))^{-1}$-geometrically recurrent with $\alpha = \xi$. In the interest of simplicity, we will first develop this correspondence under a number of simplifying assumptions and later indicate how these assumptions can be removed.

Assume, for the moment, that $D_i \equiv 0$ for all i and that $\{B_i\}$ is an i.i.d. sequence which is independent of the Markov-dependent sequence $\{A_i\}$. Next, introduce the strong minorization condition

$$a\eta(E) \le P(x,E), \quad x \in \mathbb{S}, \;\; E \in \mathscr{S}, \tag{$\mathscr{M}_1$}$$

where $a > 0$ and η is a probability measure.

If the kernel $\hat{P}_\alpha$ is $(\lambda(\alpha))^{-1}$-recurrent, then there exists a right invariant function r_α satisfying the equation $\hat{P}_\alpha r_\alpha = \lambda(\alpha) r_\alpha$. Otherwise, $\hat{P}_\alpha$ is $(\lambda(\alpha))^{-1}$-transient and there exists a right subinvariant function r_α ([29], Sect. 5.1). For any $\alpha \in \mathrm{dom}\,(\lambda)$, introduce the α-shifted transition kernel

$$Q_\alpha(x,E) = \int_E \frac{e^{\alpha g_A(x)} r_\alpha(y)}{\lambda(\alpha) r_\alpha(x)} P(x,dy), \quad x \in \mathbb{S}, \;\; E \in \mathscr{S}.$$

Then Q_α is a probability kernel when $\hat{P}_\alpha$ is $(\lambda(\alpha))^{-1}$-recurrent (and a subprobability kernel in the transient case). Let $\mathbf{E}_\alpha[\cdot]$ denote expectation with respect to this shifted measure.

Observe that the minorization $(\mathscr{M}_0)$ (or the stronger condition $(\mathscr{M}_1)$) induces a minorization for Q_α; in particular, using the definition of Q_α together with $(\mathscr{M}_0)$, we obtain

$$h_\alpha(x)\eta_\alpha(E) \le Q_\alpha(x,E), \quad x \in C, \;\; E \in \mathscr{S}, \tag{$\mathscr{M}_\alpha$}$$

where, for some normalizing constant L,

$$h_\alpha(x) = \frac{Lh(x)}{\lambda(\alpha) r_\alpha(x)} e^{\alpha g_A(x)} \wedge 1 \quad \text{and} \quad \eta_\alpha(dy) = \frac{1}{L} r_\alpha(y)\eta(dy).$$

Here L is a normalizing constant, chosen such that η_α is a probability measure. (We may assume that η has been selected in a suitable way so that $L < \infty$.) Thus, a minorization exists, and hence a regeneration structure for the ξ-shifted chain. Also set $\hat{h}_\alpha(x) = h(x) e^{\alpha g_A(x)}/\lambda(\alpha)$.

Lemma 1. *Assume $\{B_i\}$ is i.i.d. and independent of $\{A_i\}$ with $\mathbf{E}\big[B_i^\xi\big] < \infty$, and assume that $D_i \equiv 0$ for all i and $(\mathscr{M}_1)$ is satisfied. Then*

$$\mathbf{E}_\xi[\kappa] < \infty \Longrightarrow \mathbf{E}\big[\mathscr{B}_b^\xi\big] < \infty.$$

Proof. For the proof, introduce the notation $h \otimes \eta(x,dy) := h(x)\eta(dy)$.

Using the series representation for a regeneration cycle (as in [28], p. 313 or Lemma 4.1 of [27]), we obtain

$$\begin{aligned}
\mathbf{E}_\xi\left[\kappa - 1 | \, X_0 = x\right] &= \mathbf{E}_\xi\left[\sum_{n=1}^{\infty} \mathbf{1}_{\{\kappa > n\}} \middle| X_0 = x\right] \\
&= \sum_{n=1}^{\infty} \int_{\mathbb{S}} \left(Q_\xi - h_\xi \otimes \eta_\xi\right)^n (x,dy)
\end{aligned}$$

$$\geq \frac{1}{r_\xi(x)} \sum_{n=1}^{\infty} \left(\hat{P}_\xi - \hat{h}_\xi \otimes \eta \right)^n r_\xi(x)$$

$$= \frac{1}{r_\xi(x)} \sum_{n=1}^{\infty} \mathbf{E}\left[e^{\xi S'_{n-1}} \mathbf{1}_{\{\kappa > n\}} r_\xi(X_n) \middle| X_0 = x \right], \qquad (48)$$

where $S'_n := \sum_{i=0}^{n} \log A_i$. In the previous display, the inequality follows directly from the definitions of h_ξ, $\hat{h}_\xi$, η_ξ, $\hat{P}_\xi$, and Q_ξ (and we obtain an inequality here due to the additional term "$\wedge 1$" appearing in the definition of h_ξ).

Next observe that under the strong minorization ($\mathscr{M}_1$), the function r_ξ is bounded below by a constant ([7], Remark 2.3). It follows that

$$\mathbf{E}\left[\sum_{n=1}^{\infty} (A_0 \cdots A_{n-1})^\xi \mathbf{1}_{\{\kappa > n\}} \middle| X_0 = x \right] \leq M r_\xi(x), \quad \text{for some } M < \infty.$$

Using independence and the moment assumption on $\{B_i\}$, we conclude that this expression also holds with $(A_0 \cdots A_{n-1} B_n)$ in place of $(A_0 \cdots A_{n-1})$ and M replaced with some finite constant M'. (Since $\{B_i\}$ is i.i.d. and independent of $\{A_i\}$, the B-sequence is independent of the regeneration times.) Consequently,

$$\mathbf{E}\left[\mathscr{B}_b^\xi \right] \leq M' \int r_\xi(x) \eta(dx).$$

In the minorization ($\mathscr{M}_1$), we may assume that the measure η has been chosen such that the integral on the right-hand side of the last expression is finite. Thus we obtain $\mathbf{E}\left[\mathscr{B}_b^\xi \right] < \infty$, as required. ☐

3.2.4 The Moments of $\mathscr{B}_b$ in the General Case

The previous argument can be modified to incorporate a nontrivial sequence $\{D_i\}$ and Markov dependence in the entire driving sequence $\{(\log A_i, B_i, D_i)\}$. Following Collamore ([8], Sect. 6.1), one approach is to replace the kernel $\hat{P}_\xi$ in the above argument with $\hat{R}_\xi$, where for any α,

$$\hat{R}_\alpha(x, E) := \int_E e^{\alpha F(x,y)} P(x, dy), \quad x \in \mathbb{S},\ E \in \mathscr{S},$$

for $(f_A(X_n), f_B(X_n)) = \left(\log A_n, \log(\tilde{B}_n + 1)\right)$ and

$$F(x, y) = f_A(x) + (f_B(y) - f_B(x)).$$

If the minorization has been chosen so that $\tilde{B}_0$ is deterministically bounded from above by a constant, then the previous lemma can be repeated to obtain the same result as before, although the Q-shifted measure is formed with respect to the kernel $\hat{R}_\xi$ rather than $\hat{P}_\xi$. That is to say, we now define

$$\tilde{Q}_\alpha(x,E) = \int_E \frac{e^{\alpha F(x,y)}\tilde{r}_\alpha(y)}{\tilde{\lambda}(\alpha)\tilde{r}_\alpha(x)} P(x,dy), \quad x \in \mathbb{S}, \ E \in \mathscr{S},$$

where $\tilde{r}_\alpha$ and $\tilde{\lambda}(\alpha)$ are the eigenvectors and eigenvalues corresponding to the kernel $\hat{R}_\alpha$. Let $\tilde{\mathbf{E}}_\alpha[\cdot]$ denote expectation with respect to this measure. In addition, assume that $\Gamma(\xi) = \tilde{\Gamma}(\xi) = 0$, where

$$\tilde{\Gamma}(\alpha) := \limsup_{n\to\infty} \frac{1}{n} \log \mathbf{E}\left[e^{\alpha S_n}(\tilde{B}_n + 1)^\alpha\right].$$

Then we obtain:

Lemma 2. *Assume* $(\mathscr{M}_1)$. *Then*

$$\tilde{\mathbf{E}}_\xi[\kappa] < \infty \Longrightarrow \mathbf{E}\big[\mathscr{B}_b^\xi\big] < \infty.$$

Finally, we observe that $(\mathscr{M}_1)$ may be weakened to $(\mathscr{M}_0)$ by first introducing the "augmented kernel"

$$P_a(x,E) := P(x,E) + a\eta(E),$$

and then computing the ξ-shifted measure using this kernel in place of P; cf. [7, 8]. Under this construction, the right invariant function $r_{\xi,a}$ is uniformly positive, as required in the proofs of Lemmas 1 and 2, and the eigenvalue $\lambda_a(\xi) \downarrow 1$ as $a \downarrow 0$.

3.2.5 Toward a Complete Result in the Markov Case

The moment assumptions in Lemmas 1 and 2, expressed in terms of the ξ-shifted measures, are not particularly natural to verify in practice, where it would be preferable to express these conditions in terms of the transition kernel of the original process. Moreover, there is also a need to verify the further moment assumption on $\mathscr{A}$, equivalent to the assumption that $\mathbf{E}\left[\mathscr{A}^\xi(\log \mathscr{A})\right] < \infty$. Now under appropriate conditions, it is known ([27], pp. 581–582) that

$$\mathbf{E}\left[\mathscr{A}^\xi(\log \mathscr{A})\right] = \mathbf{E}_\xi\left[\log \mathscr{A}\right] = \mathbf{E}_\xi\left[\kappa\right]\mathbf{E}_\xi[\log\ A | X \sim \pi],$$

where π denotes the stationary measure of $\{X_n\}$.

Roughly speaking, a sufficient condition for the above results to hold is the geometric ξ-recurrence of the kernels $\hat{P}_\xi$ and $\hat{R}_\xi$ (cf. [29], Proposition 5.25). Thus, it is of some theoretical and applied interest to understand how geometric ξ-recurrence relates to the underlying properties of the given Markov chain.

An effort to draw this connection has been given in [8]. Let $\mathfrak{h} : \mathbb{S} \to [0, \infty)$, and define:

$$\begin{aligned}
\mathscr{L}_a \mathfrak{h} &= \{x \in \mathbb{S} : \mathfrak{h}(x) \le a\}, \quad a \ge 0; \\
\tilde{S}_n &= \{\log\, A_1 + \cdots + \log\, A_{n-1}\} + \log(\tilde{B}_n + 1), \quad n = 1, 2, \ldots; \\
S_n^{\mathfrak{h}} &= \mathfrak{h}(X_1) + \cdots + \mathfrak{h}(X_n), \quad n = 1, 2, \ldots; \\
\Gamma_{\mathfrak{h}}(\alpha, \beta) &= \limsup_{n \to \infty} \frac{1}{n} \log \mathbf{E}\left[e^{\alpha \tilde{S}_n + \beta S_n^{\mathfrak{h}}}\right], \quad (\alpha, \beta) \in \mathbb{R}^2.
\end{aligned}$$

Assume the existence of a nonnegative $\mathfrak{h}$-function such that the following condition holds.

Minorization:

($\mathfrak{M}$) For any $a > 0$ sufficiently large, there exist a constant $\delta_a > 0$ and a probability measure η_a on $(\mathbb{S}, \mathscr{S})$ with $\eta_a(\mathscr{L}_a \mathfrak{h}) > 0$ such that

$$\delta_a \mathbf{1}_{\mathscr{L}_a \mathfrak{h}}(x) \eta_a(E) \le P(x, E), \quad x \in \mathbb{S}, \; E \in \mathscr{S}.$$

Also impose the following additional assumptions on the process.

Hypotheses:

($\mathscr{H}_1$) For the function $\mathfrak{h}$ given in ($\mathfrak{M}$), there exist points $\alpha > r$ and $\beta > 0$ such that $\Gamma_{\mathfrak{h}}(\alpha, \beta) < \infty$.

($\mathscr{H}_2$) For any $a > 0$, there exist nontrivial sets $E_1, \ldots, E_l \subset \mathbb{S}$, possibly dependent on a, and a finite constant J_a such that

$$P(x, E) \le J_a \inf \left\{ \sum_{i=1}^{l} P(x_i, E) : x_i \in E_i, \; 1 \le i \le l \right\}, \quad x \in \mathscr{L}_a \mathfrak{h}, \; E \in \mathscr{S}.$$

We also need to assume the usual regularity assumptions, now with respect the random variables formed over a regeneration cycle. We collect these assumptions as an additional hypothesis:

($\mathscr{H}_3$) Hypotheses (H_0) and (H_3) hold with respect to $(\mathscr{A}, \mathscr{B}, \mathscr{D})$.

In Collamore [8], it is shown that if ($\mathfrak{M}$), ($\mathscr{H}_1$), and ($\mathscr{H}_2$) hold, then

$$\mathbf{E}\left[\mathscr{A}^{\alpha}\right] < \infty \quad \text{and} \quad \mathbf{E}\left[\mathscr{B}_b^{\alpha}\right] < \infty, \quad \text{for some } \alpha > \xi.$$

Moreover, $\mathbf{E}\big[\mathscr{B}_f^{\alpha}\big] < \infty$, provided that these conditions hold with respect to the time-reversed Markov chain (which we implicitly assume in the following development when the forward recursion is considered).

Note that the property described in ($\mathfrak{M}$) always holds for Harris chains when P is replaced by P^{k_a}; see [25]. Also, we note that a regeneration structure we need would still exist if ($\mathscr{M}_0$) were weakened to a condition on P^k rather than P, that is, to Harris recurrence; see [29], p. 134. Finally, it seems plausible that condition ($\mathscr{H}_2$) could possibly be removed, since this is mainly used in [8] to assure that the relevant eigenvectors are bounded from above by a constant.

Combining the results of [9] and [8], we obtain an extension of Theorem 3. First let $\{\mathscr{Z}_n^{(p)}\}$ and $\{\mathscr{Z}_n^{(c)}\}$ denote the perpetuity and conjugate sequences, defined as in (17) and (18), but with (A_i, B_i, D_i) replaced with $(\mathscr{A}_i, \mathscr{B}_i, \mathscr{D}_i)$. Then set $\mathscr{Z}^{(p)} = \lim_{n\to\infty} \mathscr{Z}_n^{(p)}$ and $\mathscr{Z}^{(c)} = \lim_{n\to\infty} \mathscr{Z}_n^{(c)}$. Also, let $\{\tilde{\mathscr{V}}_n\}$ denote the forward process generated by the sequence $\{\mathscr{A}_i, \mathscr{B}_i, \mathscr{D}_i\}$ according to Letac's Model E, and set $\tau = \inf\{i : \tilde{\mathscr{V}}_i \in \mathscr{C}\}$, that is, the first passage time of this Markov chain into its $\mathscr{C}$-set. ($\tilde{\mathscr{V}}_i \equiv \mathscr{V}_i$ if it is a forward process we study, but not if it is a backward process.) Finally, recall that $\lambda(\alpha)$ denotes the convergence parameter associated with the kernel $\hat{P}(\alpha)$, as defined in Sect. 3.2.1.

Theorem 5. *Assume that* (44) *holds and that* ($\mathfrak{M}$), ($\mathscr{H}_1$), ($\mathscr{H}_2$), *and* ($\mathscr{H}_3$) *are satisfied. Then*

$$\lim_{u\to\infty} u^{\xi} \mathbf{P}\{W > u\} = C \tag{49}$$

for a finite positive constant C*, where* $W := \lim_{n\to\infty} V_n$ *in the forward case and* $W := \sup_n Z_n$ *in the backward case. Moreover, the constant* C *may be identified as*

$$C = \frac{1}{\xi\lambda'(\xi)\mathbf{E}[\tau]} \mathbf{E}_{\xi}\left[(\mathscr{Z}^{(p)} - \mathscr{Z}^{(c)})^{\xi} \mathbf{1}_{\{\tau=\infty\}}\right]. \tag{50}$$

Remark 1. In specific examples, the constant C can be identified more explicitly and put into the same general form as in the i.i.d. case studied in Theorem 2. In particular, if $\{V_n\}$ is obtained from the forward recursion $V_n = A_n V_{n-1} + B_n$ and the sequence $\{B_n\}$ is supported on $[0, \infty)$, then the conjugate term in (50) is zero (since all nonzero terms in (18) would then be positive), and so it follows from (50) and (39) that

$$\mathscr{Z}^{(p)} = V_0 + \frac{B_1}{A_1} + \frac{B_2}{A_1 A_2} + \cdots, \tag{51}$$

where, as in Theorem 2, the initial distribution (corresponding to the distribution of the random variable V_0) is obtained from the regeneration measure of the Markov chain $\{V_{K_i} : i = 0, 1, \ldots\}$. We remind the reader that the sequence $\{K_i\}$ represents the regeneration times of the Markov chain $\{X_n\}$, and that (A_n, B_n) is modulated by this chain $\{X_n\}$, that is, $(A_n, B_n) = g(X_n)$.

Moreover, the stopping time τ can also be viewed—as in Theorem 2—as a first passage time. Specifically, in Nummelin's split-chain construction [28, 29], regeneration of the Markov chain $\{X_n\}$ occurs when $Y_n := (X_n, \gamma_n) \in \mathscr{C}_0 \times \{1\}$, where $\{\gamma_n\}$ is an i.i.d. Bernoulli sequence with $\mathbf{P}\{\gamma_n = 1\} = \delta_0$ and $(\mathscr{C}_0, \delta_0)$ appear in the minorization $(\mathscr{M}_0)$ of the Markov chain $\{X_n\}$. Now, $\{\tau = \infty\}$ corresponds to the event that $\{V_{K_i}\}$ never returns to its $\mathscr{C}$-set (namely, the interval $\mathscr{C} = [-M, M]$); that is to say, $\{V_n\}$ never returns to $\mathscr{C}$ at a regeneration time of $\{X_n\}$. But this is the same as the condition that $\{(V_n, Y_n) : n = 0, 1, \ldots\}$ never returns to the set $\mathscr{C} \times \mathscr{C}_0 \times \{1\}$.

Now if $\{V_n\}$ is a perpetuity sequence (thus obtained from backward recursion of the SFPE $f(v) = Av + B$ rather than forward recursion of this SFPE), then (42) must be employed in place of (39), which does not simplify in the same way as (51). However, in this case, it is plausible to employ [26] to obtain the time-reversed process of $\{(V_n, A_n, B_n) : n = 0, 1, \ldots\}$, and to observe that this time-reversed process assumes the form of the forward sequence $V_n = \tilde{A}_n V_{n-1} + \tilde{B}_n$, where $(\tilde{A}_n, \tilde{B}_n) = (A_{-n}, B_{-n})$ for the extended process $\{(A_n, B_n) : n \in \mathbb{Z}\}$. Since the limiting distribution of this forward process agrees with the limiting distribution of the original perpetuity sequence, we expect an expression of the form (51), also for the case of perpetuities.

Remark 2. In [8], Sect. 3, the conditions of Theorem 5 are verified for a variety of problems which are of applied interest. One application considered in [8] is the ruin problem with investments (described above in Example 2), but where the investment returns are Markov-dependent, governed by any one of the following:

(i) A discrete-time Black-Scholes model under Markov regime switching, where the regime switching is determined by an underlying finite-state or uniformly recurrent Markov chain.
(ii) The logarithmic returns $\{-\log\ A_i\}$ are modeled as an AR(p) process or, with slight modifications of the assumptions, an ARMA(p, q) process.
(iii) The insurance company invests a fixed fraction of its surplus capital in a stock and a fixed fraction in a bank account, where the returns on the bank investment are at a deterministic rate $r > 1$, while the returns on the stock investment follow the stochastic volatility model suggested in [13, 14]. Specifically, the investment returns are modeled as $R_n = \sigma_n \zeta_n$, where $\{\zeta_n\}$ is an i.i.d. Gaussian sequence and $\{\log \sigma_n\}$ is modeled, say, as an ARMA(p, q) process.

Another application considered in [8] is a GARCH(1,1) process with regime switching, where the regime shifts are (as in (i)) modulated by an underlying finite-state Markov chain.

The proof of Theorem 3.1 is a direct consequence of Theorem 2.1 of [9] and Theorem 4.1 of [8]. In the forward case, it also needs to be observed that the limit over regeneration cycles agrees with the steady-state limit of the original sequence. But this equivalence is obtained along the lines of [8], p. 1428.

References

1. Alsmeyer, G., Iksanov, A., Rösler, U.: On distributional properties of perpetuities. J. Theor. Probab. **22**, 666–682 (2009)
2. Asmussen, S., Albrecher, H.: Ruin Probabilities, 2nd edn. World Scientific, River Edge (2010)
3. Asmussen, S., Sigman, K.: Monotone stochastic recursions and their duals. Probab. Eng. Inf. Sci. **10**, 1–20 (1996)
4. Athreya, K.B., Ney, P.: A new approach to the limit theory of recurrent Markov chains. Trans. Am. Math. Soc. **245**, 493–501 (1978)
5. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econ. **31**, 307–327 (1986)
6. Brito, M., Freitas, A.: Edgeworth expansion for an estimator of the adjustment coefficient. Insur.: Math. Econ. **43**, 203–208 (2008)
7. Collamore, J.F.: Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random variables. Ann. Appl. Probab. **12**, 382–421 (2002)
8. Collamore, J.F.: Random recurrence equations and ruin in a Markov-dependent stochastic economic environment. Ann. Appl. Probab. **19**, 1404–1458 (2009)
9. Collamore, J.F., Vidyashankar, A.N.: Tail estimates for stochastic fixed point equations via nonlinear renewal theory. Stoch. Process. Appl. **123**, 3378–3429 (2013)
10. Collamore, J.F., Diao, G., Vidyashankar, A.N.: Rare event simulation for processes generated via stochastic fixed point equations. (2011, Preprint). Math arXiv PR/1107.3284
11. Collamore, J.F., Diao, G., Vidyashankar, A.N.: Joint inference for tail parameters of SFPE. (2013, in preparation)
12. Cramér, H.: On the Mathematical Theory of Risk. Försäkringsaktiebolaget Skandia 1855–1930, Parts I and II, Stockholm (1930)
13. Davis, R., Mikosch, T.: Extremes of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.P., Mikosch, T. (eds.) Handbook of Financial Time Series. Springer, Berlin (2008)
14. Davis, R., Mikosch, T.: Probabilistic properties of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.P., Mikosch,T. (eds.) Handbook of Financial Time Series. Springer, Berlin (2008)
15. de Haan, L., Resnick, S., Rootzén, H., de Vries, C.G.: Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. Stoch. Process. Appl. **32**, 213–224 (1989)
16. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica **50**, 987–1007 (1982)
17. Enriquez, N., Sabot, C., Zindy, O.: A probabilistic representation of constants in Kesten's renewal theorem. Probab. Theory Relat. Fields **144**, 581–613 (2009)
18. Furstenberg, H.: Noncommuting random products. Trans. Am. Math. Soc. **108**, 377–428 (1963)
19. Goldie, C.M.: Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. **1**, 126–166 (1991)
20. Hitczenko, P., Wesołowski, J.: Renorming divergent perpetuities. Bernoulli **17**, 880–894 (2011)
21. Iglehart, D.L.: Extreme values in the GI/G/1 queue. Ann. Math. Stat. **43**, 627–635 (1972)
22. Kesten, H.: Random difference equations and renewal theory for products of random matrices. Acta Math. **131**, 207–248 (1973)
23. Letac, G.: A contraction principle for certain Markov chains and its applications. Random matrices and their applications. Proceedings of AMS-IMS-SIAM Joint Summer Research Conference 1984. Contemp. Math. **50**, 263–273 (1986)
24. Lundberg, F.: Approximerad Framställning av Sannolikhetsfunktionen. Återförsäkring av Kollektivrisker. Akad. Afhandling. Almqvist och Wiksell, Uppsala (1903)
25. Meyn, S., Tweedie, R.: Markov Chains and Stochastic Stability. Springer, Berlin (1993)

26. Nelson, E.: The adjoint Markoff process. Duke Math. J. **25**, 671–690 (1958)
27. Ney, P., Nummelin, E.: Markov additive processes I. Eigenvalue properties and limit theorems. Ann. Probab. **15**, 561–592 (1987)
28. Nummelin, E.: A splitting technique for Harris recurrent Markov chains. Z. Wahrsch. Verw. Gebiete **43**, 309–318 (1978)
29. Nummelin, E.: General Irreducible Markov Chains and Non-negative Operators. Cambridge University Press, Cambridge (1984)
30. Roitershtein, A.: One-dimensional linear recursions with Markov-dependent coefficients. Ann. Appl. Probab. **17**, 572–608 (2007)
31. Siegmund, D.: The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. Ann. Probab. **4**, 914–924 (1976)
32. Siegmund, D.: Note on a stochastic recursion. State of the Art in Probability and Statistics (Leiden, 1999), IMS Lecture Notes Monograph Series, vol. 36, pp. 547–554. Institute of Mathematical Statistics, Beachwood, OH (2001)
33. Woodroofe, M.: Nonlinear Renewal Theory in Sequential Analysis. SIAM, Philadelphia (1982)
34. Yakin, B., Pollak, M.: A new representation for a renewal-theoretic constant appearing in asymptotic approximations of large deviations. Ann. Appl. Probab. **8**, 749–774 (1998)

Homogeneity at Infinity of Stationary Solutions of Multivariate Affine Stochastic Recursions

Yves Guivarc'h and Émile Le Page

Abstract We consider a d-dimensional affine stochastic recursion of general type corresponding to the relation

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad X_0 = x. \tag{S}$$

Under natural conditions, this recursion has a unique stationary solution R, which is unbounded. If $d > 2$, we sketch a proof of the fact that R belongs to the domain of attraction of a stable law which depends essentially of the linear part of the recursion. The proof is based on renewal theorems for products of random matrices, radial Fourier analysis in the vector space $\mathbb{R}^d$, and spectral gap properties for convolution operators on the corresponding projective space. We state the corresponding simpler result for $d = 1$.

1 Notations and Main Result

Let $V = \mathbb{R}^d$ be the d-dimensional Euclidean space endowed with the scalar product $< x, y >= \sum_{i=1}^{d} x_i y_i$ and the corresponding norm $|x| = (\sum_{i=1}^{d} |x_i|^2)^{1/2}$. We denote by $G = GL(V)$ (resp. H =Aff(V)) the linear (resp. affine) group of V and we fix a probability measure μ (resp. λ) on G (resp. H) such that μ is the projection of λ. We consider the affine stochastic recursion (S) on V defined by

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad X_0 = x, \tag{S}$$

Y. Guivarc'h (✉)
Université de Rennes 1, UFR de mathématiques, Institut de Recherche Mathématiques de Rennes (IRMAR), 263 Avenue du General Leclerc CS 74205, 35042 Rennes, France
e-mail: yves.guivarch@univ-rennes1.fr

É. Le Page
Université de Bretagne Sud, LMBA, Campus de Tohannic, BP 573, 56017 Vannes, France
e-mail: lepage@univ-ubs.fr

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_6,

where (A_n, B_n) are i.i.d. random variables with law λ, hence A_n (resp. B_n) are i.i.d. random matrices (resp. vectors). We assume that (S) has a stationary solution R which satisfies in distribution

$$R = AR^1 + B$$

where R^1 has the same law as R and is independent of (A, B). We are interested in the "asymptotic shape" of the law ρ of R. Our focus will be on the case $d > 1$. For $d = 1$, corresponding results are described in Sect. 4.

We denote by $\eta * \theta$ the convolution of a probability measure η on H with a positive Radon measure θ on V i.e. $\eta * \theta = \int \delta_{hx}\, d\eta(h)\, d\theta(x)$. Also η^n denotes the nth convolution iterate of η. With these notations, the law ρ_n of X_n is given by $\rho_n = \lambda^n * \delta_x$, and a λ-stationary (probability) measure ρ satisfies $\lambda * \rho = \rho$.

We denote by Ω the product space $H^{\otimes \mathbb{N}}$, by $\mathbb{P}$ the product measure $\lambda^{\otimes \mathbb{N}}$ on Ω, and by $\mathbb{E}$ the corresponding expectation operator. Provided that

$$\mathbb{E}(|\log|A||) + \mathbb{E}(|\log|B||) < \infty,$$

it is well known (see [12] for example) that a λ-stationary measure ρ exists and is unique if the top Lyapunov exponent

$$L_\mu = \lim_{n\to\infty} \frac{1}{n}\, \mathbb{E}(\log|A_n \dots A_1|)$$

is negative. For informations on products of random matrices we refer to [2, 5, 9].

Since the properties of μ play a dominant role for the "shape" of ρ, we give now a few corresponding notations. Let S (resp. T) be the closed subsemigroup of G (resp. H) generated by the support suppμ (resp. suppλ) of μ (resp. λ) and write

$$S_n = A_n \dots A_1, \quad k(s) = \lim_{n\to\infty} (\mathbb{E}(|S_n|^s))^{1/n} \ (s \geq 0).$$

Then $\log k(s)$ is a convex function on $I_\mu = \{s \geq 0; k(s) < \infty\}$ and we write $s_\infty = \sup\{s \geq 0; k(s) < \infty\}$.

We denote by $\mathbb{S}^{d-1}$ (resp. $\mathbb{P}^{d-1}$) the unit sphere (resp. projective space) of V and observe that in polar coordinates:

$$V\backslash\{0\} = \mathbb{S}^{d-1} \times \mathbb{R}_+^*.$$

If $\dot{V}$ denotes the factor space of $V\backslash\{0\}$ by the group $\{\pm Id\}$, we have also

$$\dot{V} = \mathbb{P}^{d-1} \times \mathbb{R}_+^*.$$

For $x \in V\backslash\{0\}$, $g \in G$, we write $\tilde{x}$ (resp. $\bar{x}$) for the projection of $x \in V$ on $\mathbb{S}^{d-1}$ (resp. $\mathbb{P}^{d-1}$), $g \cdot \tilde{x}$ (resp. $g \cdot \bar{x}$) for the projection of gx on $\mathbb{S}^{d-1}$ (resp. $\mathbb{P}^{d-1}$).

For some moment conditions on μ, the quantity $\gamma(g) = \sup(|g|, |g^{-1}|)$ will be used. The dual map g^* of $g \in GL(V)$ is defined by $< g^*x, y >=< x, gy >$ and the push-forward of μ by $g \to g^*$ will be denoted μ^*.

From now on, we assume $d > 1$. An element $g \in G$ is said to be *proximal* if g has a unique simple dominant eigenvalue $\lambda_g \in \mathbb{R}$ with $|\lambda_g| = \lim_{n\to\infty} |g^n|^{1/n}$. In this case we have the decomposition $V = \mathbb{R}w_g \oplus V_g^<$ where w_g is a dominant eigenvector and $V_g^<$ a g-invariant hyperplane. We say that a semigroup of G satisfies *condition* $i - p$ if this semigroup contains a proximal element and does not leave any finite union of subspaces invariant.

One can observe that, if $d > 1$, the set of probability measures μ on G such that S satisfies condition $i - p$ is open and dense in the weak topology. Also, condition $i - p$ is satisfied for S if and only if it is satisfied for the closed subgroup $Zc(S)$, the Zariski closure of S, which is a Lie group with a finite number of components. Thus condition $i - p$ is in particular satisfied, if $Zc(S) = G$.

It is known that, if the probability measure μ satisfies $\mathbb{E}(|\log |A||) < \infty$ and suppμ generates a closed semigroup S satisfying condition $i - p$, then the top Lyapunov exponent of μ is simple (see [2]). In this case $\log k(s)$ is strictly convex and analytic on $[0, s_\infty[$ (see [9]). Also the set S^{prox} of proximal elements in S is open and the set of corresponding positive dominant eigenvalues generates a dense subgroup of $\mathbb{R}_+^*$. Furthermore, the action of μ on $\mathbb{P}^{d-1}$ has a unique μ-stationary measure ν and suppν is the unique S-minimal subset of $\mathbb{P}^{d-1}$; the set $\Lambda(S) = \text{supp}\nu$ is the closure of $\{\bar{w}_g; g \in S^{\text{prox}}\}$ and has positive Hausdorff dimension.

Under condition $i - p$ and for the action of S on $\mathbb{S}^{d-1}$, there are two cases I, II, say, regarding the existence of a convex S-invariant cone in V. In case I (non-existence), the inverse image $\tilde{\Lambda}(S)$ of $\Lambda(S)$ in $\mathbb{S}^{d-1}$ is the unique S-minimal invariant set in $\mathbb{S}^{d-1}$. In case II (existence), $\tilde{\Lambda}(S)$ splits into two symmetric S-minimal subsets $\tilde{\Lambda}_+(S)$ and $\tilde{\Lambda}_-(S)$.

Returning to the affine situation, we need to consider the compactification $V \cup \mathbb{S}_\infty^{d-1}$ of V by the sphere at infinity $\mathbb{S}_\infty^{d-1}$ and we identify $\tilde{\Lambda}(S)$ (resp. $\tilde{\Lambda}_+(S), \tilde{\Lambda}_-(S)$) with the corresponding subset $\tilde{\Lambda}^\infty(S)$ (resp. $\tilde{\Lambda}_+^\infty(S), \tilde{\Lambda}_-^\infty(S)$) of $\mathbb{S}_\infty^{d-1}$. We observe that $\mathbb{S}_\infty^{d-1}$ is H-invariant and the corresponding H-action reduces to the G-action.

If $h = (g, b)$ is such that $|g| < 1$, then h has a fixed point $x \in V$, and this point is attractive, i.e. for any $y \in V$, $\lim_{n\to\infty} h^n y = x$. The set $\Delta_a(T)$ of such attractive fixed points of elements $h \in T$ plays an important role in the description of supp ρ, for ρ λ-stationary.

On the other hand, if for some $s > 0$ we have $k(s) > 1$ and condition $i - p$ is satisfied, then one can show the existence of $g \in S$ with $\lim_{n\to\infty} |g^n|^{1/n} > 1$. This implies that supp$\rho$ is unbounded, if suppλ has no fixed point in V.

We have the following basic (see [10], Proposition 5.1)

Proposition 1. *Assume* $\mathbb{E}(\log \gamma(A)) + \log |B|) < \infty$ *and*

$$L_\mu = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}(\log |S_n|) < 0.$$

Then $R_n = \sum_1^n A_1 \dots A_{k-1} B_k$ *converges* $\mathbb{P}$*-a.e. to*

$$R = \sum_1^\infty A_1 \dots A_{k-1} B_k,$$

and for any $x \in V$, X_n *converges in law to* R. *If* $\beta \in I_\mu$ *satisfies* $k(\beta) < 1$, $\mathbb{E}(|B|^\beta) < \infty$, *then* $\mathbb{E}(|R|^\beta) < \infty$.

The law ρ *of* R *is the unique* λ*-stationary measure on* V. *The closure* $\overline{\Delta_a(T)} = \Lambda_a(T)$ *in* V *is the unique* T*-minimal subset of* V *and* $\Lambda_a(T) = \operatorname{supp}\rho$. *If the semigroup* S *satisfies condition* $i - p$ *and* $\operatorname{supp}\lambda$ *has no fixed point in* V, *then* $\rho(W) = 0$ *for any affine subspace* W. *Furthermore, if* T *contains an element* (g, b) *with* $\lim_{n\to\infty} |g^n|^{1/n} > 1$, *then* $\Lambda_a(T)$ *is unbounded.*

The first part of the proposition is well known (see for example [12]).

For $s \geq 0$, we denote by l^s (resp. h^s) the s-homogeneous measure (resp. function) on $\mathbb{R}_+^*$ given by $l^s(dt) = t^{-(s+1)}dt$, $l^0 = l$ (resp. $h^s(t) = t^s$). We observe that the cone of Radon measures on $\dot{V}$ which are of the form $\eta \otimes l^s$ with η a positive measure on $\mathbb{P}^{d-1}$ is G-invariant. Also $g(\eta \otimes l^s) = (\rho_s(g)\eta) \otimes l^s$ with

$$\rho_s(g)\eta = \int |gx|^s \delta_{g\cdot x} d\eta(x).$$

One can show that if the subsemigroup S associated to μ satisfies condition $i - p$ and $s \in I_\mu$, there exists a unique probability measure ν^s on $\mathbb{P}^{d-1}$ such that equation $\mu * (\nu^s \otimes l^s) = k(s)\nu^s \otimes l^s$ is satisfied. Furthermore ν^s gives mass zero to any projective hyperplane and $\operatorname{supp}\nu^s = \Lambda(S)$.

We denote by $\tilde{\nu}^s$ the unique symmetric positive measure on $\mathbb{S}^{d-1}$ with projection ν^s on $\mathbb{P}^{d-1}$ and (in case II) by $\tilde{\nu}^s_+$, $\tilde{\nu}^s_-$ its normalized restrictions to $\tilde{\Lambda}_+(S)$, $\tilde{\Lambda}_-(S)$ hence $\tilde{\nu}^s = \frac{1}{2}(\tilde{\nu}^s_+ + \tilde{\nu}^s_-)$. Then we have

$$\mu * (\tilde{\nu}^s \otimes l^s) = k(s)\tilde{\nu}^s \otimes l^s$$

and

$$\mu * (\tilde{\nu}^s_+ \otimes l^s) = k(s)\tilde{\nu}^s_+ \otimes l^s, \quad \mu * (\tilde{\nu}_- \otimes l^s) = k(s)\tilde{\nu}^s_- \otimes l^s.$$

If there exists $\alpha \in I_\mu$ such that $k(\alpha) = 1$, the measures $\tilde{\nu}^\alpha \otimes l^\alpha$, $\tilde{\nu}^\alpha_+ \otimes l^\alpha$, $\tilde{\nu}^\alpha_- \otimes l^\alpha$ enter in an essential way in the description of the "shape" of ρ. We need first to discuss the action of S on $\mathbb{S}^{d-1}_\infty$, if supp ρ is unbounded. In this case $\overline{\operatorname{supp}\rho} \cap \mathbb{S}^{d-1}_\infty$ is a non trivial closed S-invariant set, hence three cases can occur, in view of the above discussion of minimality.

CASE I: S has no proper convex invariant cone and $\overline{\Lambda_a(T)} \supset \tilde{\Lambda}^\infty(S)$.

CASE II': S has a proper convex invariant cone and $\overline{\Lambda_a(T)} \supset \tilde{\Lambda}^\infty(S)$.

CASE II”: S has a proper convex invariant cone and $\overline{\Lambda_a(T)}$ contains only one of the sets $\tilde{\Lambda}_+^\infty(S)$, $\tilde{\Lambda}_-^\infty(S)$, say $\tilde{\Lambda}_+^\infty(S)$ hence $\tilde{\Lambda}_-^\infty(S) \cap \overline{\Lambda_a(T)} = \emptyset$.

The push-forward of a measure η on V by the dilation $x \to tx$ $(t > 0)$ will be denoted $t.\eta$. For $d > 1$, our main result in [10] is the following

Theorem 1. *With the above notations, assume that S satisfies condition $i - p$, that T has no fixed point in V, that $L_\mu < 0$, and that there exists $\alpha > 0$ with $k(\alpha) = 1$ and $\mathbb{E}(|A|^\alpha \gamma^\delta(A)) < \infty$, $\mathbb{E}(|B|^{\alpha+\delta}) < \infty$ for some $\delta > 0$. Then* $\operatorname{supp} \rho$ *is unbounded and we have the following vague convergence on $V \setminus \{0\}$:*

$$\lim_{t \to 0+} t^{-\alpha}(t.\rho) = C(\sigma^\alpha \otimes l^\alpha) = \Lambda,$$

where $C > 0$, σ^α is a probability on $\mathbb{S}^{d-1}$ and the Radon measure Λ satisfies $\mu * \Lambda = \Lambda$. Moreover,

$$\sigma^\alpha = \begin{cases} \tilde{\nu}^\alpha & \text{in Case I}, \\ C_+\tilde{\nu}_+^\alpha + C_-\tilde{\nu}_-^\alpha \text{ for some } C_+, C_- > 0 & \text{in Case II'}, \\ \tilde{\nu}_+^\alpha & \text{in Case II''}. \end{cases}$$

The measures $\tilde{\nu}^\alpha \otimes l^\alpha$ (case I), $\tilde{\nu}_+^\alpha \otimes l^\alpha$ and $\tilde{\nu}_-^\alpha \otimes l^\alpha$ (cases II’, II”) are minimal μ-harmonic measures.

The above convergence is valid on any Borel function f with $\sigma^\alpha \otimes l^\alpha$-negligible set of discontinuities such that $|w|^{-\alpha}|\log|w||^{1+\varepsilon}|f(w)|$ is bounded for some $\epsilon > 0$, hence

$$\lim_{t \to 0+} t^{-\alpha}\mathbb{E}(f(tR)) = \Lambda(f).$$

The theorem shows that ρ belongs to the domain of attraction of a stable law, a fact conjectured by F. Spitzer. It plays a basic role in the study of slow diffusion for random walk in a random medium on $\mathbb{Z}$ (see [7]), and also in extreme value theory for GARCH processes. The proof of the theorem shows that the above convergence is valid on the sets $H_w^+ = \{x \in V; < x, w > 1\}$ for $w \in V \setminus \{0\}$ under the weaker hypothesis $\mathbb{E}(|A|^\alpha \log \gamma(A) + |B|^{\alpha+\delta}) < \infty$. Then, using [1], it follows that the theorem is valid if $\alpha \notin \mathbb{N}$. Actually, [1] implies also the validity of the theorem under the above weaker hypothesis, in the following situations:

CASE I and $\alpha \notin 2\mathbb{N}$,
CASE II” and $\alpha > 0$,
CASE II’ and $C_+ = C_-$, $\alpha \notin 2\mathbb{N}$.

As observed in [14], the condition $C_+ = C_-$ occurs if ρ is symmetric, in particular if the law of B is symmetric (for example if B is Gaussian). In the context of extreme value theory the convergence stated in the theorem says that

ρ has "multivariate regular variation". This property is basic for the development of the theory for "ARCH processes" (see [14]).

The proof given in [10] (Theorem 6) is long. For a short survey of earlier work, see [8]. Here we will give a sketch of a few main points of the proof.

2 Some Tools for the Proof of the Theorem

2.1 The Renewal Theorem for Products of Random Matrices ($d > 1$)

We use the notations already introduced above: μ is a probability measure on $G = GL(V)$, S the closed subsemigroup of G generated by supp μ, L_μ the top Lyapunov exponent of μ, ν the μ-stationary measure on $\mathbb{P}^{d-1}$ etc. Under condition $i-p$, the following is the d-dimensional analog of the classical renewal theorem (see [4]) and follows from the general renewal theorem of Kesten [13] for Markov random walks on $\mathbb{R}$.

Theorem 2. *Assume that the semigroup S associated with μ satisfies condition $i-p$, that $\log \gamma(g)$ is μ-integrable, and that $L_\mu = \lim_{n\to\infty} \frac{1}{n}\int \log|g| d\mu^n(g) > 0$. Then, for any $w \in V$, $\sum_0^\infty \mu^k * \delta_w$ is a Radon measure on $\dot{V}$ and we have*

$$\lim_{w\to 0} \sum_0^\infty \mu^k * \delta_w = \frac{1}{L_\mu}\nu^0 \otimes l.$$

in the sense of vague convergence. This convergence is also valid on any bounded continuous function f on $\dot{V}$ with $\sum_{-\infty}^{\infty} \sup\{|f(w)|; 2^l \le |w| \le 2^{l+1}\} < \infty$.

As proved in [10], if S satisfies $i-p$, $s \in I_\mu$ and $\int |g|^s \log \gamma(g) d\mu(g) < \infty$, then the top Lyapunov exponent $L_\mu(s) = \lim_{n\to\infty} \frac{1}{n}\int |g|^s \log|g| d\mu^n(g)$ exists, is simple and satisfies $L_\mu(s) = \frac{k'(s_-)}{k(s)} < \infty$. Also there exists a unique positive function e^s on $\mathbb{P}^{d-1}$ such that $\nu^s(e^s) = 1$ and

$$\mu * \delta_w(e^s \otimes h^s) = k(s)(e^s \otimes h^s)(w).$$

Then, using [13] again, we have the following result which includes information on the fluctuations of $S_n w$:

Theorem 3. *Assume that $L_\mu < 0$, $\alpha \in I_\mu$ exists with $\alpha > 0$, $k(\alpha) = 1$, $\int |g|^\alpha \log \gamma(g) d\mu(g) < \infty$, and S satisfies condition $i-p$. Then we have the following vague convergence on $\dot{V}$, for any $w \in \dot{V}$*

$$\lim_{t\to 0+} t^{-\alpha} \sum_{0}^{\infty} \mu^k * \delta_{tw} = \frac{(e^\alpha \otimes h^\alpha)(w)}{L_\mu(\alpha)} \nu^\alpha \otimes l^\alpha.$$

This convergence is actually valid on any continuous function f on $\dot{V}$ such that $f_\alpha(w) = |w|^{-\alpha} f(w)$ is bounded and $\sum_{-\infty}^{\infty} \sup\{f_\alpha(w); 2^l \le |w| \le 2^{l+1}\} < \infty$.

In particular for some $A > 0$ and any $w \in V$

$$\lim_{t\to\infty} t^\alpha \, \mathbb{P}\{\sup_{n\ge 1} |S_n w| > t\} = A(e^\alpha \otimes h^\alpha)(w).$$

The last formula is the so-called Cramér estimate of ruin in collective risk systems if $d = 1$ [4].

For the convergence proof in Theorem 1, we will need an analogue of Theorem 3 with $\dot{V}$ replaced by $V\setminus\{0\}$. For $u \in \mathbb{S}^{d-1}$, the function $e^\alpha(u)$ can be lifted to $\mathbb{S}^{d-1}$ and we have

$$\int |gu|^\alpha \frac{e^\alpha(g.u)}{e^\alpha(u)} \, d\mu^n(g) = 1$$

for any $n \in \mathbb{N}$. Hence the family of probability measures $|gu|^\alpha \frac{e^\alpha(g.u)}{e^\alpha(u)} d\mu^{\otimes n}(g)$ with $g = g_1 \dots g_n$ defines a projective system on the spaces $G^{\otimes n}$ and one can consider the projective limit $\mathbb{Q}_u^\alpha$ on $G^{\otimes \mathbb{N}}$. Referring again to [13], we get the following

Theorem 4. *Assume μ and α are as in Theorem 3. Then, for any $u \in \mathbb{S}^{d-1}$, we have the vague convergence*

$$\lim_{t\to 0+} t^{-\alpha} \sum_{0}^{\infty} \mu^k * \delta_{tu} = \frac{1}{L_\mu(\alpha)} e^\alpha(u) \, \tilde{\nu}_u^\alpha \otimes l^\alpha,$$

where $\tilde{\nu}_u^\alpha$ is a probability measure on $\mathbb{S}^{d-1}$ and $\tilde{\nu}_u^\alpha \otimes l^\alpha$ is a μ-harmonic Radon measure on $V\setminus\{0\}$. The convergence is valid on any continuous function f such that $f_\alpha(w) = |w|^{-\alpha} f(w)$ is bounded and satisfies

$$\sum_{-\infty}^{\infty} \sup\{|f_\alpha(w)|; 2^l \le |w| \le 2^{l+1}\} < \infty.$$

There are two cases:

Case I: $\tilde{\nu}_u^\alpha = \tilde{\nu}$ has support $\tilde{\Lambda}(S)$.

Case II: $\tilde{\nu}_u^\alpha = p_+^\alpha(u)\tilde{\nu}_+^\alpha + p_-^\alpha(u)\tilde{\nu}_-^\alpha$, where $p_+^\alpha(u)$ (resp. $p_-^\alpha(u)$) is the entrance probability under $\mathbb{Q}_u^\alpha$ of $S_n \cdot u$ into the convex envelope of $\tilde{\Lambda}_+(S)$ (resp. $\tilde{\Lambda}_-(S)$).

These results improve earlier ones by Kesten [12] and Le Page [16].

2.2 A Spectral Gap Property for Convolution Operators $(d > 1)$

As above we consider the operator P on $\dot{V}$ defined by $Pf(w) = (\mu * \delta_w)(f)$ and its action on s-homogeneous functions. The Euclidean norm on V extends to a norm on the wedge product $\bigwedge^2 V$: For $x, y, x', y' \in V$, we put

$$< x \wedge y, x' \wedge y' > := \det \begin{pmatrix} < x, x' > & < x, y' > \\ < y, x' > & < y, y' > \end{pmatrix}.$$

This allows to consider the distance δ on $\mathbb{P}^{d-1}$ defined by $\delta(x, y) = |x \wedge y|$, where x, y correspond to unit vectors $\tilde{x}, \tilde{y}$ in $\mathbb{S}^{d-1}$. We will denote by $H_\varepsilon(\mathbb{P}^{d-1})$ the space of ε-Hölder functions on $\mathbb{P}^{d-1}$ with respect to the distance δ. We write

$$[\varphi]_\varepsilon = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{\delta(x, y)^\varepsilon}, \quad |\varphi| = \sup_x |\varphi(x)|, \quad |\varphi|_\varepsilon = [\varphi]_\varepsilon + |\varphi|,$$

and we observe that $\varphi \to |\varphi|_\varepsilon$ defines a norm on $H_\varepsilon(\mathbb{P}^{d-1})$.

If $z \in \mathbb{C}$, $z = s + it$, and the z-homogeneous function f on $\dot{V}$ is of the form $f = \varphi \otimes h^z$, with $\varphi \in H_\varepsilon(\mathbb{P}^{d-1})$, the action of P on f defines an operator P^z on φ by

$$Pf = P^z \varphi \otimes h^z, \quad \text{i.e.} \quad P^z \varphi(x) = \int \varphi(g \cdot x)\, |gx|^z \, d\mu(g).$$

Then we have the following (see [10], Theorem A)

Theorem 5. *Let $d > 1$ and assume that the closed subsemigroup S generated by* $\operatorname{supp} \mu$ *satisfies condition $i - p$. For $s \in I_\mu$, assume $\int |g|^s \gamma^\delta(g) d\mu(g) < \infty$ for some $\delta > 0$. Then, for any $\varepsilon > 0$ sufficiently small, the operator P^s on $H_\varepsilon(\mathbb{P}^{d-1})$ has a spectral gap, with dominant eigenvalue $k(s)$:*

$$P^s = k(s)(\nu^s \otimes e^s + U_s),$$

where $\nu^s \otimes e^s$ is the projection on $\mathbb{C}e^s$ defined by ν^s, e^s and U_s is an operator with spectral radius less than 1 which commutes with $\nu^s \otimes e^s$. Furthermore, if $\Im z = t \neq 0$, $z = s + it$, then the spectral radius of P^z is less than $k(s)$.

If $s = 0$, P^s reduces to convolution by μ on $\mathbb{P}^{d-1}$ and convergence to the unique μ-stationary measure $\nu^0 = \nu$ was a basic property studied in [5]. In this case the spectral gap property is a consequence of the simplicity of the top Lyapunov exponent of μ (see [2, 9]). The spectral gap properties of P^s are basic ingredients for the study of precise large deviations for the product of random matrices $S_n = A_n \dots A_1$ (see [16, 18]). Here the theorem will be used for the study of s-homogeneous P-eigenmeasures on $\dot{V}$ and $V\backslash\{0\}$. In the context of $V\backslash\{0\}$ we need to replace $\mathbb{P}^{d-1}$ by $\mathbb{S}^{d-1}$ and to use an analogous theorem (see [10]).

2.3 A Choquet-Deny Property for Markov Walk

Here (S, δ) is a compact metric space and P is a Markov kernel on $S\times\mathbb{R} = Y$ which commutes with the $\mathbb{R}$-translations and acts continuously on the space $C_b(S\times\mathbb{R})$ of continuous bounded functions on $S\times\mathbb{R}$. Such a set of datas will be called a Markov walk on $\mathbb{R}$. We define for $t \in \mathbb{R}$ the Fourier operator P^{it} on $C(S)$ by

$$P^{it}\varphi(x) = P(\varphi \otimes e_{it})(x, 0),$$

where e_{it} is the Fourier exponential on $\mathbb{R}$, $e_{it}(r) = e^{itr}$. For $t = 0$, $P^{it} = P^0$ is equal to $\bar{P}$, the factor operator on S defined by P. We assume that for $\varepsilon > 0$ P^{it} preserves the space of ε-Hölder functions $H_\varepsilon(S)$ on (S, δ) and is a bounded operator on $H_\varepsilon(S)$.

We denote $[\varphi]_\varepsilon = \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{\delta(x,y)^\varepsilon}$, $|\varphi| = \sup_x |\varphi(x)|$ for $\phi \in C(S)$. Moreover, we assume that P^{it} and P satisfy the following condition D:

1. For any $t \in \mathbb{R}$, one can find $n_0 \in \mathbb{N}$, $\rho(t) \in [0, 1[$ and $C(t) \geq 0$ for which
$$[(P^{it})^{n_0}\varphi]_\varepsilon \leq \rho(t)[\varphi]_\varepsilon + C(t)|\varphi|.$$
2. For any $t \in \mathbb{R}$, the equation $P^{it}\varphi = e^{i\theta}\varphi$, $\varphi \in H_\varepsilon(S)$, $\varphi \neq 0$, has only the trivial solution $e^{i\theta} = 1$, $t = 0$, $\varphi =$ constant.
3. For some $\delta > 1$: $M_\delta = \sup_{x\in S} \int |a|^\delta P((x, 0), d(y, a)) < \infty$.

Conditions 1 and 2 above imply that $\bar{P}$ has a unique stationary measure π and the spectrum of $\bar{P}$ in $H_\varepsilon(S)$ is of the form $\{1\} \cup \Delta$, where Δ is a compact subset of the open unit disk (see [11]). They imply also that for any $t \neq 0$, the spectral radius of P^{it} is less than one.

If $Y = \dot{V}$, P is the convolution operator by μ on $\dot{V} = \mathbb{P}^{d-1} \times \mathbb{R}_+$ $(d > 1)$, hence $S = \mathbb{P}^{d-1}$ and $\mathbb{R}_+^* = \exp\mathbb{R}$. Theorem 5 implies that condition D is satisfied if $I_\mu \neq 0$ and condition $i - p$ is valid.

Furthermore, for $s \in I_\mu$ one can also consider the Markov operator Q_s on $\dot{V}$ defined by

$$Q_s f = \frac{1}{k(s)e^s \otimes h^s} P(fe^s \otimes h^s).$$

If for some $\delta > 0$, $\int |g|^s \gamma^\delta(g) d\mu(g) < \infty$, Theorem 5 implies that conditions D are also satisfied by Q_s.

We will say that a Radon measure θ on $Y = S \times \mathbb{R}$ is translation-bounded if for any compact $K \subset Y$ there exists $C(K) > 0$ such that $\theta(K + t) \leq C(K)$ for any $t \in \mathbb{R}$, where $K + t$ is the set obtained from K by translation with t. Then we have the following Choquet-Deny type property

Theorem 6. *With the above notations if the Markov operator P on $Y = S \times \mathbb{R}$ satisfies the condition D. Then any translation-bounded P-harmonic measure on Y is proportional to $\pi \otimes l$ with $l = dt$.*

This theorem can be used for $Y = \dot{V}$ and $P = Q_\alpha$ if $0 < \alpha < s_\infty$.

2.4 A Weak Renewal Theorem

As in the Sect. 2.3, we consider a Markov walk P on $\mathbb{R}$ with compact factor space S, a probability ν on S such that $\nu \otimes l$ is P-invariant. A path starting from S for this Markov chain will be denoted (X_n, V_n) with $X_n \in S,\ V_n \in \mathbb{R}$ and the canonical probability measure on the paths starting from $x \in S$ will be denoted by ${}^a\mathbb{P}_x$. We write also ${}^a\mathbb{P}_\nu = \int {}^a\mathbb{P}_x d\nu(x)$.

For a non negative Borel function on $S \times \mathbb{R}$, we write $U\psi = \sum_0^\infty P^k\psi$. We observe that if $(x,t) \in S \times \mathbb{R}$, $\psi = 1_K$, then $U\psi(x,t)$ is the expected number of visits to K starting from $(x,t) \in S \times \mathbb{R}$. In other words $U\psi(x,t) = \mathbb{E}_x\left(\sum_0^\infty \psi(X_k, t + V_k)\right)$. Then we have the following weak analogue of the renewal theorem.

Proposition 2. *Suppose that ψ is a bounded, non-negative and compactly supported Borel function on $S \times \mathbb{R}$. Further suppose that the potential $U\psi = \sum_0^\infty P^k\psi$ is locally bounded and that, for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} {}^a\mathbb{P}_\nu \left\{ \left| \frac{V_n}{n} - \gamma \right| > \varepsilon \right\} = 0 \quad \textit{with} \quad \gamma < 0$$

holds true. Then

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t ds \int_s U\psi(x,s)d\nu(x) = \frac{1}{|\gamma|} \int \int_{-\infty}^{\infty} \psi(x,s)d\nu(x)ds.$$

If ψ is a non-negative Borel function on S such that $\lim_{t\to\infty} U\psi(x,t) = 0$ ν-a.e., then $\psi = 0$ $\nu \otimes l$-a.e.

3 Elements of Proof of Theorem 1

3.1 Convergence for Radon Transforms

For a finite measure η on V we write $\hat{\eta}(w) = \eta(H_w^+)$ where $u = tw,\ t > 0,\ u \in \mathbb{S}^{d-1},\ H_w^+ = \{x \in V; < x, w >> 1\}$. We observe that $\hat{\eta}$ can be considered as an integrated form of the Radon transform of η. Observe that $\widehat{\mu * \eta}(w) = (\mu^* * \delta_w)(\hat{\eta})$, hence convolution equations on $G \times V$ can be transformed to functional equations for Radon transforms.

We will not be able to apply directly the renewal Theorem 4 to the convolution equation $\lambda * \rho = \rho$ corresponding to $R = AR^1 + B$ but rather to functional equations for $\hat{\rho}$ and μ^*. We denote by ρ_1 the law of $R - B$ and we begin with the

Proposition 3. *With the hypothesis of Theorem 4, we denote by* ${}^*\tilde{\nu}_u^\alpha$ *the positive kernel on* $\mathbb{S}^{d-1}$ *given by Theorem 4 and associated with* μ^**. Then one has the equations on* $V \setminus \{0\}$

$$\rho = \sum_0^\infty \mu^k * (\rho - \rho_1), \qquad \hat{\rho}(w) = \sum_0^\infty ((\mu^*)^k * \delta_w)(\hat{\rho} - \hat{\rho}_1).$$

For $u \in \mathbb{S}^{d-1}$*, if* $\alpha \in]0, s_\infty[,\ k(\alpha) = 1$*, the function* $t \to t^{\alpha-1}(\hat{\rho} - \hat{\rho}_1)(u,t)$ *is Riemann-integrable on* $]0,\infty[$ *and one has, with* $r_\alpha(u) = \int_0^\infty t^{\alpha-1}(\hat{\rho} - \hat{\rho}_1)(u,t)dt$

$$\lim_{t\to\infty} t^\alpha \hat{\rho}(u,t) = \frac{{}^*e^\alpha(u)}{L_\mu(\alpha)}\, {}^*\tilde{\nu}_u^\alpha(r_\alpha) = C(\sigma^\alpha \otimes l^\alpha)(H_u^+),$$

where $C \geq 0$ *and the probability* σ^α *on* $\tilde{\Lambda}(S)$ *satisfies* $\mu * (\sigma^\alpha \otimes l^\alpha) = \sigma^\alpha \otimes l^\alpha$*. There exists* $b > 0$ *such that* $\mathbb{P}\{|R| > t\} \leq bt^{-\alpha}$*. Furthermore* $\operatorname{supp} \rho$ *is unbounded and: In case I:* $\sigma^\alpha = \tilde{\nu}^\alpha$*, in case II:* $C\sigma^\alpha = C_+\tilde{\nu}_+^\alpha + C_-\tilde{\nu}_-^\alpha,\ C_+, C_- \geq 0$.

Sketch of Proof

Since $|g^*| = |g|$, the function $k(s)$ is equal to the corresponding function for μ^*, condition $i - p$ is satisfied for μ^* and $L_{\mu^*}(\alpha) = L_\mu(\alpha)$. We observe that the stationarity equation $R - B = AR^1$ can be written in distribution as $\rho - \rho_1 = \rho - \mu * \rho$. Also $\rho(\{0\}) = 0$, hence we get

$$\rho = \sum_0^\infty \mu^k * (\rho - \rho_1), \qquad \hat{\rho}(w) = \sum_0^\infty ((\mu^*)^k * \delta_w)(\hat{\rho} - \hat{\rho}_1)$$

on $V \setminus \{0\}$.

In order to use Theorem 4, we need to regularize $\hat{\rho} - \hat{\rho}_1$ by multiplicative convolution on $\mathbb{R}_+^*$ with $1_{[0,1]}$, hence to consider

$$r^\alpha(u,t) = \frac{1}{t}\int_0^t x^{\alpha-1}(\hat{\rho} - \hat{\rho}_1)(u,x)dx.$$

Clearly $|r^\alpha(u,t)| \leq \alpha^{-1}t^{\alpha-1}$. By using the conditions $\mathbb{E}(|A|^{\alpha+\delta}) < \infty$ and $\mathbb{E}(|B|^{\alpha+\delta}) < \infty$, one can show the existence of $\delta' > 0$, $c(\delta') > 0$ such that for $t \geq 1$,

$$|r^\alpha(u,t)| \leq c(\delta')t^{-\delta'}.$$

Then Theorem 4 can be applied to $f_\alpha(w) = r^\alpha(u,t)$, whence, by a Tauberian argument as in [6], we get the convergence of $t^\alpha \hat{\rho}(u,t)$ towards $\frac{1}{L_\mu(\alpha)}{}^*e^\alpha(u)^*\tilde{\nu}_u^\alpha(r_\alpha)$. From the existence of $\alpha \in]0, s_\infty[$ with $k(\alpha) = 1$, one can deduce the existence of $g \in S$ with $|g| > 1$, hence supp ρ is unbounded.

The above formulae and the description of ${}^*e^\alpha$, σ^α in terms of $\tilde{\nu}^\alpha, \tilde{\nu}_+^\alpha, \tilde{\nu}_-^\alpha$ give the harmonicity equation $\mu*(\sigma^\alpha \otimes l^\alpha) = \sigma^\alpha \otimes l^\alpha$. The boundedness of $t^\alpha P\{|R| > t\}$ follows from the convergence of $t^\alpha \hat{\rho}(u,t)$.

3.2 Homogeneity at Infinity of ρ

The boundedness of $t^\alpha P\{|R| > t\}$ stated in Proposition 3 implies that the family of Radon measures $\{t^{-\alpha}(t.\rho); t \in \mathbb{R}_+\}$ is relatively compact in the vague topology.

Proposition 4. *Given the situation of Theorem 1, assume that η is a vague limit of a sequence $t_n^{-\alpha}(t_n \cdot \rho)$ as $t_n \to \infty$. Then η is translation-bounded and satisfies $\mu * \eta = \eta$. If η and $\sigma \otimes l^\alpha$ satisfy*

$$\eta(H_u^+) = (\sigma \otimes l^\alpha)(H_u^+),$$

for any $u \in \mathbb{S}^{d-1}$ and some positive measure σ on $\mathbb{S}^{d-1}$, then $\eta = \sigma \otimes l^\alpha$.

This proposition is based on the moment conditions satisfied by R, A, B, and on Theorem 6. Using furthermore Propositions 4 and 3, we get the

Theorem 7. *With the hypothesis of Theorem 1, we have the following vague convergence*

$$\lim_{t\to 0+} t^{-\alpha}(t.\rho) = \Lambda = C(\sigma^\alpha \otimes l^\alpha),$$

where $C \geq 0$.

The above convergence is also valid on any Borel function f such that the set of discontinuities of f is $(\sigma^\alpha \otimes l^\alpha)$-negligible and such that for some $\varepsilon > 0$, the function $|w|^{-\alpha}|\log|w||^{1+\varepsilon}|f(w)|$ is bounded.

3.3 Positivity of C_+, C_-

We need to consider processes (dual to X_n) and taking values in $(V\backslash\{0\}) \times \mathbb{R}$ or $\mathbb{S}^{d-1} \times \mathbb{R}$ and we write

$$S'_n = A^*_n \dots A^*_1.$$

Let M be a S^*-minimal subset of $\mathbb{S}^{d-1}$ i.e. $M = \tilde{\Lambda}(S^*)$ in case I and $M = \tilde{\Lambda}_+(S^*)$ (or $(\tilde{\Lambda}_-(S^*))$ in case II. We denote by $\Lambda^*_a(T)$ the set of $u \in \mathbb{S}^{d-1}$ such that the projection of ρ on the line $\mathbb{R}u(u \in \mathbb{S}^{d-1})$ is unbounded in direction u. The following is the essential step in the discussion of positivity.

Proposition 5. *With the hypothesis of Theorem 1, if $\Lambda^*_a(T) \supset M$, then for any $u \in M$*

$$C_M(u) = \lim_{t\to\infty} t^\alpha \mathbb{P}\{< R, u >> t\} > 0.$$

In order to explain the main points of the proof, we need to introduce some notations. We observe that R_n satisfies the recursion

$$< R_{n+1}, w >=< R_n, w > + < B_{n+1}, S'_n w >,$$

hence $(S'_n w, r+ < R_n, w >)$ is a Markov walk on $V\backslash\{0\} \times \mathbb{R}$ based on $\mathbb{S}^{d-1} \times \mathbb{R}$. If we write

$$t' = r^{-1},\ w = |w|u,\ p = r|w|^{-1}$$

with $u \in \mathbb{S}^{d-1}$ this Markov walk can be expressed on $(\mathbb{S}^{d-1} \times \mathbb{R}) \times \mathbb{R}^*$ as

$$u_{n+1} = g^*_{n+1}.u_n,\ p_{n+1} = \frac{p_n+ < b_{n+1}, u_n >}{|g^*_{n+1}u_n|},\ t'_{n+1} = t'_n(|g^*_{n+1}u_n|p_{n+1}p_n^{-1})^{-1}.$$

We denote by ${}^*\hat{P}$ the corresponding Markov kernel. Since $(S'_n w, r+ < R_n, w >)$ has equivariant projection $S'_n w$ on $V\backslash\{0\}$, we have ${}^*\hat{P}({}^*e^\alpha \otimes h^\alpha) ={}^* e^\alpha \otimes h^\alpha$, hence we can consider the new relativized kernel ${}^*\hat{P}_\alpha$ and the corresponding Markov walk (u_n, p_n, t'_n) over the chain $(u_n, p_n) \in X = M \times \mathbb{R}$.

We denote

$${}^*q^\alpha(u, g) = |g^*u|^\alpha \frac{{}^*e^\alpha(g^*.u)}{{}^*e^\alpha(u)}$$

and for $h = (g, b) \in H$,

$$h^u p = \frac{1}{|g^*u|}(p+ < b, u >);$$

then the Markov kernel ${}^*\hat{Q}^\alpha$ of the chain (u_n, p_n) is given by

$$ {}^*\hat{Q}^\alpha\varphi(u,p) = \int \varphi(g^*\cdot u, h^u p)^* q^\alpha(u,g)d\lambda(h). $$

We have $L_\mu(\alpha) > 0$ and M is minimal, hence it is easy to show that ${}^*\hat{Q}^\alpha$ has a unique stationary measure κ on X, and with respect to the Markov measure ${}^*\hat{\mathbb{Q}}_u^\alpha$ on $X \times \Omega$ we have $\mathbb{E}_u^\alpha(\log^+|p|) < \infty$ and $\limsup_{n\to\infty}|S_n'u||p_n| = \infty$. We observe that, since $L_\mu(\alpha) > 0$, the Markov walk (u_n, p_n, t_n') on $X \times \mathbb{R}^*$ has negative drift, in additive notation.

The condition $\Lambda_a^*(T) \supset M$ implies

$$ \kappa(M\times]0,\infty[) > 0,\ \limsup_{n\to\infty}|S_n'u|p_n = \infty, $$

for $p > 0$.

We now consider the following $\mathbb{N}\cup\{\infty\}$-valued stopping time τ on $X \times \Omega$ defined by

$$ \tau = Inf\{n > 1; p^{-1} < R_n, u >> 0\}, $$

and we observe that, by definition of p_n :

$$ \tau = Inf\{n > 1; p^{-1}p_n|S_n'u| > 1\}, $$

hence $p^{-1}p_\tau > 0$. Hence τ (resp. $p^{-1}p_\tau|S_\tau'u||$) can be interpreted as the first ladder epoch (resp. height) of the Markov walk $p^{-1}p_n|S_n'u|$ (see [4]).

Using Poincaré's recurrence theorem and $\lim_{n\to\infty}|S_n'u| = \infty$ ${}^*\hat{\mathbb{Q}}_u^\alpha$-a.e. we infer that $\tau < \infty$ ${}^*\hat{\mathbb{Q}}_\kappa^\alpha$-a.e.

Let ${}^*\hat{P}^\tau$, ${}^*\hat{Q}^\tau$ be the stopped kernels of ${}^*\hat{P}$, ${}^*\hat{Q}$, respectively, defined by τ and let ${}^*\hat{P}_\alpha^\tau$, ${}^*\hat{Q}^{\alpha,\tau}$ be the corresponding relativised Markovian kernels. Then we have the

Lemma 1. *With $tw = u \in \mathbb{S}^{d-1}$, $t > 0$, we write on $X \times \mathbb{R}^*$*

$$ \psi(w,p) = \mathbb{P}\{p^{-1} < R, u >> t\},\ \psi_\tau(v,p) = \mathbb{P}\{t < p^{-1} < R, u >< t + p^{-1} < R_\tau, u >\} $$

$$ \psi^\alpha = ({}^*e^\alpha \otimes h^\alpha)^{-1}\psi,\ \psi_\tau^\alpha = ({}^*e^\alpha \otimes h^\alpha)^{-1}\psi_\tau. $$

Then $\psi = \sum_0^\infty ({}^*\hat{P}^\tau)^k\psi_\tau$, $\psi^\alpha = \sum_0^\infty ({}^*\hat{P}_\alpha^\tau)^k\psi_\tau^\alpha$.

The proof is analogous to the first part of Proposition 3, in order to get the Poisson equation $\psi_\tau = \psi -{}^* \hat{P}^\tau\psi$. Since $p^{-1}p_\tau > 0$, the operator ${}^*\hat{Q}^{\alpha,\tau}$ preserves $X_+ = M\times]0,\infty[$. If $\Lambda_a^*(T) \supset M$, then $\kappa(X_+) > 0$. Since $\mathbb{E}_\kappa^\alpha(\log^+|p|) < \infty$, one can show that the Markov kernel ${}^*\hat{Q}_x^{\alpha,\tau}$ has an ergodic stationary measure κ_+^τ which is absolutely continuous with respect to $1_{X_+}\kappa$. Also we have, using the interpretation of τ as a return time in the dynamical system associated with ${}^*\hat{\mathbb{Q}}_x^\alpha$ and the bilateral shift,

$$\mathbb{E}_0^\alpha(\tau) = \int \mathbb{E}_u^\alpha(\tau) d\kappa_+^\tau(u, p) < \infty, \ \gamma_\tau^\alpha = \mathbb{E}_{\kappa_+^\tau}^\alpha(\log(p^{-1}p_\tau |S'_\tau u|)) \in]0, \infty[$$

with $\gamma_\tau^\alpha = L_\mu(\alpha)\mathbb{E}_0^\alpha(\tau)$.

Now we can consider the Markov walk defined by ${}^*\hat{P}_\alpha^\tau$ on $X_+ \times \mathbb{R}_+^*$. In view of the above observations we can apply Proposition 2 to ${}^*\hat{P}_\alpha^\tau$ and $\kappa_+^\tau \otimes l$. We recall that, in additive notation, this Markov walk has negative drift $-\gamma_\tau^\alpha < 0$. If for some $u \in M$ we have $C_M(u) = 0$, then for $p > 0$ and $u = tw$ $(t > 0)$ $\lim_{t\to\infty} \psi^\alpha(w, p) = 0$.

Using Proposition 3 we get $\lim_{t\to\infty} \psi^\alpha(w, p) = 0$ for any $u = tw \in M$. In particular, this is valid κ_+^τ-a.e., hence Proposition 2 implies $\psi_\tau^\alpha = 0$ $\kappa_+^\tau \otimes l$-a.e., i.e.

$$\mathbb{P}\{t < p^{-1} < R, u >< t + p^{-1} < R_\tau, u >\} = 0.$$

Since $p^{-1} < R_\tau, u >> 0$, we get $p^{-1} < R, u > \leq 0$ $\kappa_+^\tau \otimes \mathbb{P}$-a.e., i.e. $< R, u > \leq 0$ $\mathbb{P}$-a.e. This contradicts $\Lambda_a^*(T) \supset M$. One can show that $\Lambda_a^*(T) = \mathbb{S}^{d-1}$ in cases I, II' and $\Lambda_a^*(T) \supset \tilde{\Lambda}_+(S^*)$ in case II", hence $C_+ > 0$.

4 The One-Dimensional Case

If $d = 1$, the notations and definitions introduced in Sect. 1 make sense. Then $G = \mathbb{R}^*$ and $H = H_1$ is the affine group "$ax + b$" of the line. Condition $i - p$ is always satisfied for any probability μ on $\mathbb{R}^*$, and the analogue of Proposition 1 is valid verbatim. For the analogue of Theorem 1 one needs to consider the possibility that S resp. μ are arithmetic, i.e. S is contained in a subset of $\mathbb{R}^*$ of the form $\{\pm a^n\}$ for some $a > 0$. The function $k(s)$ has the explicit form

$$k(s) = \int |a|^s d\mu(a).$$

Also $L_\mu = \int \log|a| d\mu(a) = k'(0)$. Then, Theorem 1 has the following analogue, with weaker moment conditions.

Theorem 8. *Assume that the probability measure λ on H_1 and μ on $\mathbb{R}^*$ satisfy the following conditions*

(a) $\mathbb{E}(\log|A|) < 0$, $k(\alpha) = 1$, *for some* $\alpha > 0$.
(b) S *is non arithmetic and* T *has no fixed point.*
(c) $\mathbb{E}(|B|^\alpha) < \infty$ *and* $\mathbb{E}|A|^\alpha |\log|A|| < \infty$.

Then one has the following convergences:

$$\lim_{t\to\infty} t^{\alpha}\mathbb{P}\{R > t\} = C_{+}$$

$$\lim_{t\to\infty} |t|^{\alpha}\mathbb{P}\{R < -t\} = C_{-}.$$

Either $\mathrm{supp}\rho = \mathbb{R}$ *and then* $C_{+}, C_{-} > 0$ *or* $\mathrm{supp}\rho$ *is a half-line* $[c, \infty[$ *(resp.* $]-\infty, c]$*) and then* $C_{+} > 0,\ C_{-} = 0$ *(resp.* $C_{-} > 0, C_{+} = 0$*).*

With respect to [6], the main new situation occurs for the discussion of positivity of C_{+}, if $A_n > 0$ and the r.v. B_n may have arbitrary sign. The proof [17] uses only the classical renewal theorem and a spectral gap property for the Markov chain p_n on $\mathbb{R}$. If $\mathrm{supp}\lambda$ does not preserve a half-line $]-\infty, c]$, one considers τ as the entrance time of p_n into $]0, \infty[$. The spectral gap property gives the finiteness of $\mathbb{E}^{\alpha}_{p}(\tau)$ for any $p \in \mathbb{R}$; using Wald's identity for the random walk $\log|S_n|$, one gets the finiteness and positivity of $\log|S_\tau|$ and then one concludes as for $d > 1$. Under stronger assumptions, the positivity of C_{+} has been obtained also in the more general context of [3], using a complex analytic method for Mellin transform due to E. Landau, and familiar in analytic number theory. The positivity of $C_{+} + C_{-}$ was obtained in [6], using P. Levy's symmetrisation method. For an analytic proof of these facts, using also Wiener-Ikehara theorem, see ([10], Appendix). In contrast to Theorem 1 and due to the Diophantine character of the hypothesis, the convergences stated in Theorem 8 are not robust under perturbation of λ in the weak topology. From that point of view, the respective roles of stable laws and of the Gaussian law are different for $d = 1$ and for $d > 1$.

References

1. Boman, J., Lindskog, F.: Support theorems for the Radon transform and Cramér-Wold theorems. J. Theor. Probab. **22**(3), 683–710 (2009)
2. Bougerol, P., Lacroix, J.: Products of Random Matrices with Applications to Schrödinger Operators. Birkhäuser, Boston (1985)
3. Buraczewski, D., Damek, E., Guivarc'h, Y., Hulanicki, A., Urban, R.: Tail-homogeneity of stationary measures for some multidimensional stochastic recursions. Probab. Theory Relat. Fields **145**(3–4), 385–420 (2009)
4. Feller, W.: An Introduction to Probability Theory and its Applications, vol. II, 2nd edn. Wiley, New York (1971)
5. Furstenberg, H.: Boundary theory and stochastic processes on homogeneous spaces. In: Harmonic Analysis on Homogeneous Spaces. Proceedings of Symposia in Pure Mathematics, Williamstown, 1972, vol. XXVI, pp. 193–229. American Mathematical Society, Providence, R.I. (1973)
6. Goldie, C.M.: Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. **1**(1), 126–166 (1991)
7. Goldsheid, I.Y.: Linear and sub-linear growth and the CLT for hitting times of a random walk in random environment on a strip. Probab. Theory Relat. Fields **141**(3–4), 471–511 (2008)

8. Guivarc'h, Y.: Heavy tail properties of stationary solutions of multidimensional stochastic recursions. Dynamics & Stochastics. IMS Lecture Notes Monograph Series, vol. 48, pp. 85–99. Institute of Mathematical Statistics, Beachwood (2006)
9. Guivarc'h, Y.: On contraction properties for products of Markov driven random matrices. Zh. Mat. Fiz. Anal. Geom. **4**(4), 457–489, 573 (2008)
10. Guivarc'h, Y., Le Page, E.: Spectral gap properties and asymptotics of stationary measures for affine random walks (2013). arXiv 1204-6004v3
11. Ionescu Tulcea, C.T., Marinescu, G.: Théorie ergodique pour des classes d'opérations non complètement continues. Ann. Math. **52**(2), 140–147 (1950)
12. Kesten, H.: Random difference equations and renewal theory for products of random matrices. Acta Mathematica **131**, 207–248 (1973)
13. Kesten, H.: Renewal theory for functionals of a Markov chain with general state space. Ann. Probab. **2**(3), 355–386 (1974)
14. Klüppelberg, C., Pergamenchtchikov, S.: Extremal behaviour of models with multivariate random recurrence representation. Stoch. Process. Appl. **117**(4), 432–456 (2007)
15. Le Page, É.: Théorèmes limites pour les produits de matrices aléatoires. Probability Measures on Groups (Oberwolfach, 1981). Lecture Notes in Mathematics, vol. 928, pp. 258–303. Springer, Berlin/New York (1982)
16. Le Page, É.: Théorèmes de renouvellement pour les produits de matrices aléatoires. Séminaires de probabilités Rennes. Publication des Séminaires de Mathématiques, Univ. Rennes I, pp. 1–116 (1983)
17. Le Page, É.: Queues des probabilités stationnaires pour les marches aléatoires affines sur la droite. (2010, preprint)
18. Lëtchikov, A.V.: Products of unimodular independent random matrices. Uspekhi Mat. Nauk **51**(1(307)), 51–100 (1996)

On Solutions of the Affine Recursion and the Smoothing Transform in the Critical Case

Sara Brofferio, Dariusz Buraczewski, and Ewa Damek

Abstract In this paper we present a new result concerning description of asymptotics of the invariant measure of the affine recursion in the critical case. We discuss also relations of this model with the smoothing transform.

1 The Affine Recursion

We consider the random difference equation:

$$X =_d AX + B, \tag{1}$$

where $(A, B) \in \mathbb{R}^+ \times \mathbb{R}$ and X are independent random variables. This equation appears both in numerous applications outside mathematics (in economy, physics, biology) and in purely theoretical problems in other branches of mathematics. It is used to study e.g. some aspects of financial mathematics, fractals, random walks in random environment, branching processes, Poisson and Martin boundaries.

It is well known that if $\mathbb{E}[\log A] < 0$ and $\mathbb{E}[\log^+ |B|] < \infty$, then there exists a unique solution to (1). The solution is the limit in distribution of the Markov chain

$$\begin{aligned} X_0^x &= 0, \\ X_n^x &= A_n X_{n-1}^x + B_n, \end{aligned} \tag{2}$$

S. Brofferio
Laboratoire de Mathématiques, Université Paris-Sud, 91405 Orsay Cedex, France
e-mail: sara.brofferio@math.u-psud.fr

D. Buraczewski (✉) · E. Damek
Instytut Matematyczny, Uniwersytet Wroclawski, pl. Grunwaldzki 2/4, 50-384 Wroclaw, Poland
e-mail: dbura@math.uni.wroc.pl; edamek@math.uni.wroc.pl

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_7,

which is called the affine recursion (since the formula reflects the action of (A_n, B_n), an element of the affine group, on the real line). To simplify our notation we will write $X_n = X_n^0$.

The most celebrated result is due to Kesten [19] (see also Goldie [13]), who proved that if $\mathbb{E}A^\alpha = 1$ for some $\alpha > 0$ (and some other assumptions are satisfied), then

$$\lim_{t\to\infty} t^\alpha \mathbb{P}[|X| > t] = C_+,$$

i.e. if ν is the law of X, then $\nu(dx) \sim \frac{dx}{x^{1+\alpha}}$ at infinity.

We are interested here in the critical case, when $\mathbb{E}\log A = 0$. Then, Eq. (1) has no stochastic solutions. Nevertheless this equation can be written in terms of measures:

$$\mu * \nu = \nu, \tag{3}$$

where μ is the distribution of (A, B), and $\mu * \nu$ is defined as follows

$$\mu * \nu(f) = \int\int f(ax + b)\nu(dx)\mu(da, db).$$

In 1997 Babillot et al. [4] proved that under the following hypotheses

$$\mathbb{E}\big[(|\log A| + \log^+ |B|)^{2+\varepsilon}\big] < \infty,\ \mathbb{P}[Ax + B = x] < 1 \text{ for all } x \in \mathbb{R} \text{ and } \mathbb{P}[A = 1] < 1 \tag{4}$$

there exists a unique (up to a constant factor) Radon measure ν, which is a solution to (3). The measure ν is an invariant measure of the process (2).

Recently we studied behavior of ν at infinity and we proved that for any $c_2 > c_1 > 0$,

$$\lim_{x\to\infty} \nu(c_1 x, c_2 x) = C_+ \log(c_2/c_1),$$

for some strictly positive constant C_+, [5, 7]. In other words we proved that the measure ν behaves at infinity like $C_+\frac{dx}{x}$. Unfortunately this result was proved under very strong hypotheses. We assumed that exponential moments are finite, i.e.

$$\mathbb{E}\big[A^\delta + A^{-\delta} + |B|^\delta\big] < \infty \qquad \text{for some } \delta > 0, \tag{5}$$

moreover in [7] we needed also absolute continuity of the measure $\overline{\mu}$, the law of $\log A$.

In this paper we consider the affine recursion (2), when B is strictly positive, that implies also that the support of ν must be contained in $(0, \infty)$. It turns out, in these settings the assumptions of our previous results can be weakened and exponential moments are not really needed. We restrict ourself to the aperiodic case, i.e. we assume the law of $\log A$ is not contained in any set of the form $p\mathbb{Z}$ for some positive p. Our main result is the following

Theorem 1. *Assume that (4) is satisfied, the measure $\overline{\mu}$ is aperiodic and the following holds*

$$\mathbb{E}\big[(|\log A| + \log^+ B)^{4+\varepsilon}\big] < \infty, \tag{6}$$

and

$$\mathbb{E}\big[|\log B|\big] < \infty, \qquad B \geq 0, \ a.s. \tag{7}$$

Then for every function $\phi \in C_c(\mathbb{R}^+)$

$$\lim_{z\to+\infty} \int_{\mathbb{R}^+} \phi(uz^{-1})\nu(du) = C_+ \int_{\mathbb{R}^+} \phi(u)\frac{du}{u}$$

for some strictly positive constant C_+.

Moreover for every $c_1 < c_2$

$$\lim_{z\to\infty} \nu\big(u: \ c_1 z < u < c_2 z\big) = C_+ \log \frac{c_2}{c_1}. \tag{8}$$

Notice that comparing with the main result of [5] we replace requirements of exponential moments (5) by much weaker assumption (6) and we assume additionally positivity of B. The integral condition in (7) is needed to control behavior of B and of the invariant measure in some small neighborhood of 0, and it is unnecessary if $B > \delta$ a.s. for some $\delta > 0$.

A complete proof of this result will be given in Sect. 3. The idea is the following. First one has to find some preliminary estimates of the measure ν under the hypothesis (4). Here we will just deduce from results contained in [5], that there exists a slowly varying function $L(z)$ such that the family of measures $\frac{\delta_{z^{-1}} * \nu}{L(z)}$ converges weakly to $C\frac{dx}{x}$, i.e. the measure $\nu(dx)$ behaves at infinity like $L(x)\frac{dx}{x}$ (Proposition 1). Next applying the duality principle ([12], p. 609), thanks to positivity of B, we prove that the measure ν is indeed bounded by the logarithm, more precisely we will show $\nu(0,z) \leq C(1 + \log z)$ (Proposition 2). Finally for an arbitrary compactly supported function ϕ on $\mathbb{R}^+$ we consider the function

$$f_\phi(x) = \int_{\mathbb{R}^+} \phi(ue^{-x})\nu(du),$$

defined on $\mathbb{R}$, as a solution of the Poisson equation

$$\overline{\mu} *_{\mathbb{R}} f_\phi = f_\phi + \psi_\phi \tag{9}$$

where ψ_ϕ is defined by the formula above, i.e. $\psi_\phi = \overline{\mu} *_{\mathbb{R}} f_\phi - f_\phi$. Then knowing already some estimates of the function f_ϕ (and our preliminary estimates are sufficient for that purpose) one can describe its asymptotics. There are two different methods. The first one bases on the classical results of Port and Stone [24, 25], who just solved explicitly the Poisson equation in the case when $\overline{\mu}$ is absolutely continuous. Nevertheless for our purpose much less is needed and the appropriate argument was given in [5]. The second method was introduced by Durrett and Liggett [11]. Thanks to the duality lemma they reduce the Poisson equation to the classical renewal equation, i.e. to an equation of the form (9), but with $\overline{\mu}$ replaced by a measure with drift and ψ_ϕ replaced by some other function. In order to prove Theorem 1 we follow here the arguments given in [5]. The second method in the context of the affine recursion was considered by Kolesko [20] and in more general settings of Lipschitz recursions by two of the authors [6].

2 The Smoothing Transform

The measure ν described in Theorem 1 is not a probability measure, but only a Radon measure. However it turns out that this measure appears in a natural way while studying purely probability objects. Here we will shortly present how this result and the methods can be used to study the smoothing transform.

To define the (inhomogeneous) smoothing transform take $(B, A_1, A_2, \dots)$ to be a sequence of positive random variables and let N be a random natural number. On the set $P(\mathbb{R})$ of probability measures on the real line the smoothing transform is defined as follows

$$\mu \mapsto \mathscr{L}\bigg(\sum_{j=1}^{N} A_j X_j + B\bigg),$$

where $X_1, X_2, \dots$ is a sequence of i.i.d random variables with common distribution μ, independent of $(B, A_1, A_2, ..)$ and N. $\mathscr{L}(X)$ denotes the law of the random variable X. A fixed point of the smoothing transform is given by any $\mu \in P(\mathbb{R})$ such that, if X has distribution μ, the equation

$$X =_d \sum_{j=1}^{N} A_j X_j + B, \tag{10}$$

holds true. Notice that if N and A_i, B are constants, the equation above characterizes stable laws as a particular case of (10).

We are interested also in a more specific case of (10). Taking $B = 0$ we obtain the homogeneous smoothing transform, i.e.

$$X =_d \sum_{i=1}^{N} A_i X_i. \tag{11}$$

Both stochastic equations described above are important from the point of view of applications. Equation (11) plays it role in description of e.g. interacting particle systems [11] and the branching random walk [1, 16]. In recent years, from very practical reasons, the inhomogeneous equation has gained importance. This equation appears e.g. in the stochastic analysis of the Pagerank algorithm (which in the heart of the Google engine) [17, 18] as well as in the analysis of a large class of divide and conquer algorithms including the Quicksort algorithm [23, 26].

Although (10) and (11) look similar to (1), often they turn out to have completely different properties. While studying Eqs. (10) and (11) main concern is to describe the right hypotheses for the following issues: existence of solutions, characterization of all the solutions and finally, description of their properties.

2.1 Homogeneous Smoothing Transform

We start first with description of the homogeneous smoothing transform. The properties of fixed points of Eq. (11) are governed by the convex function

$$m(\theta) = \mathbb{E}\Big[\sum_{j=1}^{N} A_j^{\theta}\Big]. \tag{12}$$

To exclude the trivial case we make the assumption $\mathbb{E}N > 1$. The first question that can be asked here is about existence of solutions of (11) and if there are any, what are all of them. The most important results are contained in the work of Durrett and Liggett [11] and in a series of papers of Liu e.g. [21]. They proved that the set of solutions of (11) is nonempty if and only if there is $\alpha \le 1$ such that $m(\alpha) = 1$ and $m'(\alpha) \le 0$. Moreover the parameter α describes the asymptotic of the Laplace transform of solutions. Their proofs goes via the Poisson equation as described in the previous section (of course some additional assumptions are needed). All their results are formulated in terms of the Laplace transform, but applying the Tauberian theorem for $\alpha < 1$ they give the correct asymptotics of X, a solution of (11). Namely they imply

$$\lim_{t\to\infty} t^{\alpha}\mathbb{P}[X > t] = C_1 \text{ if } m'(\alpha) < 0 \qquad \text{and} \qquad \lim_{t\to\infty} \frac{t^{\alpha}}{\log t}\mathbb{P}[X > t] = C_2 \text{ if } m'(\alpha) = 0.$$

Unfortunately the Tauberian theorem does not give the optimal answer when $\alpha = 1$, e.g. if $m'(\alpha) = 0$ one can only deduce the weaker asymptotics

$$\int_0^x \mathbb{P}[X > t]dt \sim C_2 \log x \qquad \text{as } x \to \infty.$$

Thus, the results of [11, 21] are sharp only for $\alpha < 1$.

It turns out that to study the case $\alpha = 1$ one has to reduce the problem to the random difference equation (1). For reader's convenience we sketch here the arguments due to Guivarc'h [15], which work in the case when N is constant and A_i are i.i.d. For the general case see [8, 22].

Let X be a solution to (11). We introduce probability measures: let η be the law of X, θ the law of $\sum_{i=2}^{N} A_i X_i$, ρ the law of A. We define new measures: $\nu(dx) = x\eta(dx)$, $\tilde{\rho}(da) = a\rho(da)$. Then, it turns out that the measure ν is μ invariant for $\mu(da\,db) = N\tilde{\rho}(da) \otimes \theta(db)$ defined on $\mathbb{R}^+ \times \mathbb{R}^+$, i.e. μ and ν satisfy (3). Indeed for any compactly supported function on $\mathbb{R}^+$ we have

$$
\begin{aligned}
\nu(f) &= \int_{\mathbb{R}^+} f(x)\nu(dx) = \int_{\mathbb{R}^+} f(x)x\eta(dx) = \mathbb{E}\big[f(X)X\big] \\
&= \mathbb{E}\Bigg[f\Bigg(\sum_{i=1}^{N} A_i X_i\Bigg)\sum_{i=1}^{N} A_i X_i\Bigg] = N\mathbb{E}\Bigg[f\Bigg(A_1X_1 + \sum_{i=2}^{N} A_i X_i\Bigg)A_1X_1\Bigg] \\
&= N\int\int\int f(ax+b)ax\rho(da)\eta(dx)\theta(db) \\
&= \int\int f(ax+b)\big(N\tilde{\rho}(da)\otimes\theta(db)\big)\nu(dx) \\
&= \int\int f(ax+b)\mu(da\,db)\nu(dx)
\end{aligned}
$$

Assume now that $m(1) = 1$, $m'(1) < 0$ and there exists $\beta > 1$ such that $m(\beta) = 1$. Then observe, that μ is a probability measure and moreover

$$
\begin{aligned}
&\int \log a\,\mu(da\,db) = N\int \log a\,\tilde{\rho}(da) = N\int a\log a\,\rho(da) = m'(1) < 0, \\
&\int a^{\beta-1}\mu(da\,db) = N\int a^{\beta}\rho(da) = m(\beta) = 1.
\end{aligned}
$$

One can easily check also other assumptions of the Kesten theorem, thus $\nu(dx) \sim C_+\frac{dx}{x^{(\beta-1)+1}} = C_+\frac{dx}{x^{\beta}}$, $\eta(dx) \sim C_+\frac{dx}{x^{\beta+1}}$ and finally $P[X > t] \sim C_+t^{-\beta}$ (we refer to [15, 22] for all the details).

Exactly the same argument is valid in the critical case when $m(1) = 1$ and $m'(1) = 0$. In fact this is the case which appear in the literature in the context of branching random walks [1, 16]. Then we reduce the problem to the affine recursion in the critical case and applying Theorem 1 one proves that $P[X > t] \sim C_+t^{-1}$ (see [8] for more details).

2.2 Inhomogeneous Smoothing Transform

The inhomogeneous smoothing transform has been studied for a relatively short time. The problem of existence of solutions was investigated in recent papers of

Alsmeyer and Meiners [2, 3]. Their results are similar to those described above (and also formulated in terms of the function m). They proved that if $m(\alpha) = 1$ and $m'(\alpha) < 0$ for some $\alpha \le 1$ (the contracting case) or $m(\alpha) = 1$ and $m'(\alpha) = 0$ for some $\alpha < 1$ (the critical case) then the set of solutions of (10) is not empty.

To study asymptotics one cannot reduce the problem as in the homogeneous case to the affine recursion. Nevertheless one can apply exactly the same methods, which give results for the affine recursion. This problem was studied by Jelenkovic and Olvera-Cravioto [17, 18] in the contracting case. Assuming that for some $\beta > \alpha$: $m(\beta) = 1$ and $m'(\beta) > 0$ and extending the Goldie's implicit renewal theory [13], they proved that $\mathbb{P}[X > t] \sim C_+ t^{-\beta}$. Positivity of the limiting constant C_+ was recently proved in [9]. The critical case is the subject of the forthcoming paper [10].

3 Proof of Theorem 1

3.1 Preliminary Estimates

In order to prove that the sequence $\delta_{z^{-1}} * \nu$ has a limit, one has to prove first that divided by an appropriately chosen slowly varying function it is weakly convergent.

Proposition 1. *Suppose that* (4) *is satisfied and* $\log A$ *is aperiodic. Let* ν *be an invariant Radon measure not reduced to a mass point at* 0*. Then there exists a positive slowly varying function* L *on* $\mathbb{R}^+$ *such that the family of measures* $\frac{\delta_{z^{-1}} * \nu}{L(z)}$ *converges weakly to* $C \frac{da}{a}$ *for some strictly positive constant* C.

Proof. This proposition was indeed proved in [5] (Theorem 2.1). However the result stated there was written in the multidimensional settings and for this reason was slightly weaker than we need here. More precisely, it was proved in [5] that the family of measures is weakly compact and all accumulation points are invariant under the action of the group generated by the support of A. Nevertheless notice that in our settings this group is just $\mathbb{R}^+$, thus any accumulation point η must be of the form $\eta(da) = C_\eta \frac{da}{a}$. Moreover the slowly varying function is of the form $L(z) = \delta_{z^{-1}} * \nu(\Phi)$, where Φ a compactly supported Lipschitz function (for the precise definition of L see [5]). Since

$$\lim_{z \to \infty} \frac{\delta_{z^{-1}} * \nu(\Phi)}{L(z)} = 1 = \eta(\Phi),$$

the constant C_η must be equal $(\int \Phi(a) \frac{da}{a})^{-1}$ and does not depend on η. □

3.2 Logarithmic Estimates

Proposition 1 implies in particular that the function $z \mapsto \nu(0,z)$ is bounded by some slowly varying function. Now we are going to prove that thanks to our addition assumptions this function is bounded just by a multiple of the logarithm.

For this purpose, let us recall the following [4] explicit construction of the measure ν. Define a random walk on $\mathbb{R}$

$$\begin{aligned} S_0 &= 0, \\ S_n &= \log(A_1 \dots A_n), \quad n \geq 1, \end{aligned} \tag{13}$$

and consider the downward ladder times of S_n:

$$\begin{aligned} L_0 &= 0, \\ L_n &= \inf\{k > L_{n-1}; S_k < S_{L_{n-1}}\}. \end{aligned} \tag{14}$$

Let $L = L_1$. The Markov process $\{X^x_{L_n}\}$ satisfies the recursion

$$X^x_{L_n} = M_n X^x_{L_{n-1}} + Q_n,$$

where (Q_n, M_n) is a sequence of i.i.d. random variables. Notice that $\{X_{L_n}\}$ is a contracting affine recursion possessing a stationary measure. Indeed since $\mathbb{E}[\log^2 A] < \infty$, we have $-\infty < \mathbb{E}S_L < 0$. Moreover $\mathbb{E}[\log^+(Q_n)] < \infty$ (see [14]). Therefore there exists a unique stationary measure ν_L of the process $\{X_{L_n}\}$. Next we define the measure ν_0 putting

$$\nu_0(f) = \int_{\mathbb{R}^+} \mathbb{E}\Big[\sum_{n=0}^{L-1} f(X^x_n)\Big]\nu_L(dx), \tag{15}$$

for any continuous compactly supported function f.

One can easily prove that ν_0 is μ invariant. At this point we cannot deduce that $\nu_0 = C\nu$ for some positive constant C, since we don't know whether ν_0 is a Radon measure. However this will be proved below.

Proposition 2. *Assume that (4) and (7) are satisfied. Then ν_0 is a multiple of ν. Moreover there exists a constant C such that for every bounded nonincreasing nonnegative function f on $\mathbb{R}^+$*

$$\int_{\mathbb{R}^+} f(u)\nu(du) < C\Big(\|f\|_\infty + \int_{1/e}^{\infty} f(y)\frac{dy}{y}\Big)$$

In particular for every $\varepsilon > 0$

$$\int_{\mathbb{R}+} \frac{1}{\log^{1+\varepsilon}(2+u)} \nu(du) < \infty \tag{16}$$

and for $z > 1/e$

$$\nu(0,z) < C(2 + \log z). \tag{17}$$

Proof. Notice that since $X_n^x \geq A_1 \dots A_n x$

$$\nu_0(f) = \int_{\mathbb{R}+} \mathbb{E}\Big[\sum_{n=0}^{L-1} f(X_n^x)\Big]\nu_L(dx) \leq \int_{\mathbb{R}+} \mathbb{E}\Big[\sum_{n=0}^{L-1} f\big(e^{S_n}x\big)\Big]\nu_L(dx).$$

Define the stopping time $T' = \inf\{n : \ S_n \geq 0\}$, where $S_n = \sum_{k=1}^n \log A_i$. Let $\{W_i\}$ be a sequence of i.i.d. random variables with the same distribution as the random variable $S_{T'}$ (recall $0 < \mathbb{E}S_{T'} < \infty$). Using the duality principle [12] we obtain

$$\nu_0(f) \leq \int_{\mathbb{R}+} \mathbb{E}\Big[\sum_{n=0}^{L-1} f\big(e^{S_n}x\big)\Big]\nu_L(dx) = \int_{\mathbb{R}+} \mathbb{E}\Big[\sum_{n=0}^{\infty} f\big(e^{W_1+\dots+W_n}x\big)\Big]\nu_L(dx). \tag{18}$$

Let U be the potential associated with the random walk $W_1 + \ldots + W_n$, i.e.

$$U(a,b) = \mathbb{E}\big[\#n : \ a < W_1 + \ldots + W_n \leq b\big].$$

By the renewal theorem $U(k, k+1)$ is bounded, thus we have

$$\begin{aligned}
\nu_0(f) &\leq \int_{\mathbb{R}+} \mathbb{E}\Big[\sum_{n=0}^{\infty} f\big(e^{W_1+\dots+W_n}x\big)\Big]\nu_L(dx) \\
&\leq \sum_{k=0}^{\infty} \int_{\mathbb{R}+} U(k, k+1) f\big(e^k x\big)\nu_L(dx) \\
&\leq C \sum_{k=0}^{\infty} \int_{\mathbb{R}+} f\big(e^k x\big)\nu_L(dx).
\end{aligned}$$

Next we divide the integral into two parts. First we estimate the integral over $(1,\infty)$

$$\begin{aligned}
\sum_{k=0}^{\infty} \int_1^{\infty} f\big(e^k x\big)\nu_L(dx) &\leq \sum_{k=0}^{\infty} f\big(e^k\big) \leq \sum_{k=-1}^{\infty} \int_k^{k+1} f\big(e^y\big)dy \\
&= \int_{-1}^{\infty} f\big(e^y\big)dy = \int_{1/e}^{\infty} f(y)\frac{dy}{y}.
\end{aligned}$$

Secondly, for $0 < x < 1$ we write

$$\sum_{k=0}^{\infty}\int_0^1 f\big(e^k x\big)\nu_L(dx) \le \int_0^1 \Bigg(\sum_{k=0}^{|\log x|} + \sum_{k=|\log x|}^{\infty}\Bigg) f\big(e^k x\big)\nu_L(dx)$$
$$\le C\|f\|_\infty \int_0^1 \big|\log x\big|\nu_L(dx) + \sum_{k=0}^{\infty} f\big(e^k\big)$$
$$\le C\|f\|_\infty \int_0^1 \big|\log x\big|\nu_L(dx) + \int_{1/e}^{\infty} f(y)\frac{dy}{y}.$$

We will justify that the first term above is finite. Notice that if $x, y \in \mathbb{R}^+$ and $x + y < 1$ then $\big|\log(x+y)\big| < \big|\log x\big|$. Observe also that $X^x_{L_n} \le X^y_{L_n}$ for $x \le y$. We write

$$\int_0^1 \big|\log x\big|\nu_L(dx) = \int_{\mathbb{R}+} \mathbb{E}\Big[\big|\log X^x_L\big| \cdot \mathbf{1}_{\{X^x_L<1\}}\Big]\nu_L(dx)$$
$$\le \int_{\mathbb{R}+} \mathbb{E}\Big[\big|\log X^0_L\big| \cdot \mathbf{1}_{\{X^0_L<1\}}\Big]\nu_L(dx)$$
$$\le \mathbb{E}\Bigg[\bigg|\log\Big(\frac{A_1A_2\dots A_L B_1}{A_1}\Big)\bigg|\Bigg]$$
$$\le \mathbb{E}\big[|S_L| + |\log B_1| + |\log A_1|\big] < \infty.$$

Therefore

$$\nu_0(f) \le C\Bigg(\|f\|_\infty + \int_{1/e}^{\infty} f(y)\frac{dy}{y}\Bigg).$$

Taking $f = \mathbf{1}_{[0,x]}$ we prove that ν_0 is a Radon measure. Since ν_0 is also μ invariant, ν must be just a multiple of ν_0 (recall that in the class of Radon measures the solution of (3) is unique up to a multiplicative constant, [4]). In particular the last inequality is valid for ν instead of ν_0. Putting $f(u) = \frac{1}{\log^{1+\varepsilon}(2+u)}$ and next $f(u) = \mathbf{1}_{[0,z]}(u)$ we complete the proof. □

3.3 Translation of the Invariant Measure ν.

It will be convenient for our purpose to change slightly the measure ν and to consider the measure $\tilde{\nu}$ defined by

$$\tilde{\nu}(f) = \int_{\mathbb{R}+} f(x-1)\nu(dx).$$

The crucial property of $\tilde{\nu}$ is that its support is contained in $(1, \infty)$, so it does not contain 0, that allows us to avoid some technical problems. Let $\tilde{\mu}$ be the law of the random pair $(A, A + B - 1)$, then $\tilde{\nu}$ is $\tilde{\mu}$ invariant:

$$\begin{aligned}\tilde{\mu} * \tilde{\nu}(f) &= \mathbb{E}\Big[\int_{\mathbb{R}+} f\big(A(x+1) + B - 1\big)\tilde{\nu}(dx)\Big] \\ &= \mathbb{E}\Big[\int_{\mathbb{R}+} f\big(Ax + B - 1\big)\nu(dx)\Big] = \int_{\mathbb{R}+} f(x-1)\nu(dx) = \tilde{\nu}(f).\end{aligned}$$

Notice that both measures ν and $\tilde{\nu}$ have the same behavior at infinite, and the family of measure $\delta_{z^{-1}} * \nu$ and $\delta_{z^{-1}} * \tilde{\nu}$ converge to the same limit (of course assuming that they really converge, what we still have to prove). Thus, for our purpose it is sufficient to consider $\tilde{\nu}$. However notice that although both measures μ and $\tilde{\mu}$ are similar they satisfy slightly different hypotheses. The projections on the A-part of μ and $\tilde{\mu}$ coincide and one can easily prove that $\tilde{\mu}$ fulfills hypotheses (4) and (6). But the random variable $A + B - 1$ may happen to be negative with positive probability, thus $\tilde{\mu}$ may not satisfy assumption (7). Nevertheless, we are only interested in behaviour of ν and $\tilde{\nu}$ at infinity, so we will use the fact, that we already know, that $\tilde{\nu}$ satisfies both (16) and (17).

From now we consider measures $\tilde{\nu}$ and $\tilde{\mu}$ instead ν and μ, but to simplify our notation we will just write ν and μ. However the reader should be aware that we are in a slightly different settings and from now instead of (7) we assume:

- Hypothesis (4) and (6) are satisfied;
- The measure ν satisfy (16) and (17).

3.4 The Poisson Equation

In order to understand the asymptotic behavior of the measure ν one has to consider the function

$$f_\phi(x) = \int_{\mathbb{R}^d} \phi(ue^{-x})\nu(du)$$

that is a solution of the Poisson equation

$$\overline{\mu} *_{\mathbb{R}} f_\phi = f_\phi + \psi_\phi \tag{19}$$

for a peculiar choice of the function ψ_ϕ, that is

$$\psi_\phi = \overline{\mu} *_{\mathbb{R}} f_\phi - f_\phi.$$

Under a number of assumptions concerning ψ_ϕ one can describe asymptotic behavior f_ϕ. Here we formulate the known results, based on the methods introduced by Port and Stone [24, 25], which we are going to use. For proofs we refer to [5, 24].

Let $\overline{\mu}$ be a centered aperiodic probability measure on $\mathbb{R}$ with the second moment $\sigma^2 = \int_{\mathbb{R}} x^2 \overline{\mu}(dx)$. The Fourier transform of $\overline{\mu}$, $\hat{\overline{\mu}}(\theta) = \int_{\mathbb{R}} e^{ix\theta} \overline{\mu}(dx)$ is a continuous bounded function, whose Taylor expansion near zero is $\hat{\overline{\mu}}(\theta) = 1 + O(\theta^2)$ and such that $|1 - \hat{\overline{\mu}}(\theta)| > 0$ for all $\theta \in \mathbb{R} \setminus \{0\}$. We consider the set $\mathscr{F}(\overline{\mu})$ of functions ψ that can be written as $\psi(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\theta} \hat{\psi}(\theta) d\theta$ for some bounded, integrable, complex valued function $\hat{\psi}$ verifying the following hypothesis

- Its Taylor expansion near 0 is
$$\hat{\psi}(\theta) = J(\psi) + i\theta K(\psi) + O(\theta^2)$$
for two constants $J(\psi)$ and $K(\psi)$,
- The function $\theta \mapsto \frac{\hat{\psi}(-\theta)}{1-\hat{\overline{\mu}}(\theta)} \cdot \mathbf{1}_{[-a,a]^c}(\theta)$ is integrable for some $a \in \mathbb{R}$.

The following result was proved in [5].

Theorem 2. *There exists a potential A, that is well defined on $\mathscr{F}(\overline{\mu})$ and such that $A\psi(x)$ is a continuous solution of the Poisson equation (19). Furthermore if $J(\psi) \geq 0$ then $A\psi$ is bounded from below and*

$$\lim_{x \to \pm\infty} \frac{A\psi(x)}{x} = \pm\sigma^{-2} J(\psi). \tag{20}$$

If additionally $J(\psi) = 0$, then $A\psi$ is bounded and has a limit at infinity

$$\lim_{x \to \pm\infty} A\psi(x) = \mp\sigma^{-2} K(\psi). \tag{21}$$

Corollary 1. *If $J(\psi) = 0$, then every continuous solution of the Poisson equation bounded from below is of the form*

$$f = A\psi + C_0$$

for some constant C_0. Thus every continuous solution of the Poisson equation is bounded and the limit of $f(x)$ exists when x goes to $+\infty$.

Conversely if there exists a bounded solution of the Poisson equation, then $A\psi$ is bounded and $J(\psi) = 0$. In particular the first part of corollary is valid.

The next lemma describes a class of functions in $\mathscr{F}(\overline{\mu})$ that we will be used later on and that have the same type of decay at infinity as $\overline{\mu}$. In particular we see that if $\overline{\mu}$ has exponential moment then $\mathscr{F}(\overline{\mu})$ contains functions with exponential decay.

Lemma 1. *Let Y a random variable with the law $\overline{\mu}$, then the function*

$$r(x) = \mathbb{E}\left[|Y - x| - |x|\right]$$

is nonnegative and

$$\hat{r}(\theta) = C \cdot \frac{\hat{\overline{\mu}}(\theta) - 1}{\theta^2}$$

for $\theta \neq 0$. *Moreover if* $\mathbb{E}|Y|^{4+\varepsilon} < \infty$ *for some* $\varepsilon > 0$ *then*

$$r(x) \leq \frac{C}{1 + |x|^{3+\varepsilon}},$$

r *is in* $\mathscr{F}(\overline{\mu})$ *and for every function* $\zeta \in L^1(\mathbb{R})$ *such that* $x^2\zeta$ *is integrable the convolution* $r *_{\mathbb{R}} \zeta$ *is in* $\mathscr{F}(\overline{\mu})$.

Proof. The first part of the Lemma follows from the formula

$$r(x) = \begin{cases} -2\mathbb{E}[(Y+x)\mathbf{1}_{Y+x\leq 0}] & \text{for } x \geq 0 \\ 2\mathbb{E}[(Y+x)\mathbf{1}_{Y+x>0}] & \text{for } x < 0 \end{cases} \tag{22}$$

and was proved in [5]. For the second part we just notice, that the last formula implies for positive x:

$$\begin{aligned} |r(x)| &= 2\int_{y<-x} |y+x|\overline{\mu}(dy) = 2\cdot\sum_{m=1}^{\infty}\int_{-(m+1)x\leq y<-mx} |y+x|\overline{\mu}(dy) \\ &\leq 2\cdot\sum_{m=1}^{\infty} mx\int_{|y|>mx}\overline{\mu}(dy) \leq 2\cdot\sum_{m=1}^{\infty} mx\int_{\mathbb{R}}\frac{|y|^{\chi}}{m^{4+\varepsilon}x^{4+\varepsilon}}\overline{\mu}(dy) \leq \frac{C}{x^{3+\varepsilon}}. \end{aligned}$$

It is clear that if $\mathbb{E}|Y|^{4+\varepsilon} < \infty$ then $r \in \mathscr{F}(\overline{\mu})$. If $\psi = r * \zeta$ with ζ and $x^2\zeta$ in $L^1(\mathbb{R})$ then it is easily checked that both ψ and $x^2\psi$ are integrable. Since $\hat{\psi} = \hat{r}\hat{\zeta} = C\frac{\hat{\overline{\mu}}-1}{\theta^2}\hat{\zeta}$ and $\hat{\zeta}$ vanish at infinity then $\psi \in \mathscr{F}(\overline{\mu})$ □

Lemma 2. *If* ϕ *is a continuous function on* $\mathbb{R}^+$ *such that for* $\beta > 2$

$$|\phi(u)| \leq \frac{C}{(1+\log^+ u)^{\beta}},$$

then the functions f_ϕ *and* $\overline{\mu} * f_\phi$ *are well defined. Furthermore if* ϕ *is Lipschitz and* $\beta > 4$, *then*

$$\int_{\mathbb{R}}\int_G\int_{\mathbb{R}^+}\Big|\phi(e^{-x}(au+b)) - \phi(e^{-x}au)\Big|\nu(du)\mu(db\,da)dx < \infty. \tag{23}$$

and

$$|\psi_\phi(x)| \leq \frac{C}{1+|x|^{\chi}},$$

for $\chi = \min\{\beta - 1, 3+\varepsilon\}$.

Proof. Assume first $x < -1$. In view of (17) we have

$$
\begin{aligned}
\left|f_\phi(x)\right| &= \int_{u>1} \left|\phi(e^{-x}u)\right|\nu(du) \le \int_{u>1} \frac{C}{\log^\beta(e^{-x}u)}\nu(du)\\
&\le C\sum_{n=0}^{\infty}\int_{e^n\le u<e^{n+1}} \frac{1}{(n-x)^\beta}\nu(du)\\
&\le C\sum_{n>|x|}^{\infty}\frac{1}{n^\beta}\int_{e^{n+x}\le u<e^{n+x+1}}\nu(du)\\
&\le C\sum_{m=1}^{\infty}\sum_{m|x|\le n<(m+1)|x|}\frac{1}{m^\beta|x|^\beta}\int_{e^{n+x}\le u<e^{n+x+1}}\nu(du)\\
&\le C\sum_{m=1}^{\infty}\frac{1}{m^\beta|x|^\beta}\int_{u<e^{(m+1)|x|}}\nu(du) \le \frac{C}{|x|^{\beta-1}}\sum_{m=1}^{\infty}\frac{1}{m^{\beta-1}}\\
&\le \frac{C}{|x|^{\beta-1}}.
\end{aligned}
$$

To proceed with positive x notice that, by (17), for every $y \in \mathbb{R}^+$ and $\beta' > 2$, arguing as above, we obtain:

$$
\begin{aligned}
\int_{\mathbb{R}^d}\frac{1}{1+\left(\log^+(y|u|)\right)^{\beta'}}\nu(du) &\le \int_{y|u|<1}\nu(du) + \sum_{n=0}^{\infty}\int_{e^n\le y|u|<e^{n+1}}\frac{1}{1+n^{\beta'}}\nu(du)\\
&\le C + C|\log y| + C\sum_{n=1}^{\infty}\frac{1}{1+n^{\beta'-1}} \le C(1+|\log y|)
\end{aligned}
\tag{24}
$$

Hence $|f_\phi(x)| \le C(1+x)$ if $x > 0$.

Finally f_ϕ is continuous, hence for $x \in (-1,0)$ is bounded. Thus

$$
|f_\phi(x)| \le C\left((1+|x|)\mathbf{1}_{x>0} + \frac{1}{1+|x|^{\beta-1}}\mathbf{1}_{x\le 0}\right)
$$

Consider now the convolution of f_ϕ with $\overline{\mu}$. First if $x > 0$, then

$$
\left|\overline{\mu} * f_\phi(x)\right| \le C\int_{\mathbb{R}}\left(1+|x+y|\right)\overline{\mu}(dy) \le C\left(1+|x|\right).
$$

Next if $x < -1$, then since $\mathbb{E}|\log A|^{4+\varepsilon} < \infty$, we have

$$\begin{aligned}\left|\overline{\mu} * f_\phi(x)\right| &\leq \int_{\mathbb{R}} \frac{C}{1+|x+y|^{\beta-1}}\overline{\mu}(dy)\\ &\leq \int_{2|y|<|x|} \frac{C}{1+|x+y|^{\beta-1}}\overline{\mu}(dy) + \frac{C}{|x|^{4+\varepsilon}}\int_{2|y|\geq|x|}|y|^{4+\varepsilon}\overline{\mu}(dy)\\ &\leq \frac{C}{1+|x|^{\chi_0}},\end{aligned}$$

for $\chi_0 = \min\{\beta-1, 4+\varepsilon\}$. The function $\overline{\mu} * f_\phi$ is also continuous, hence finally we obtain

$$|\overline{\mu} * f_\phi(x)| \leq C\left((1+|x|)\mathbf{1}_{x>0} + \frac{1}{1+|x|^{\chi_0}}\mathbf{1}_{x\leq 0}\right).$$

Next we have

$$\begin{aligned}&\int_{-\infty}^{0}\int_G\int_{\mathbb{R}^d}\left|\phi(e^{-x}(au+b)) - \phi(e^{-x}au)\right|\nu(du)\mu(db\,da)dx\\ &\qquad\leq \int_{-\infty}^{0}\int_G\int_{\mathbb{R}^d}\left|\phi(e^{-x}(au+b))\right|\nu(du)\mu(db\,da)dx\\ &\qquad\quad + \int_{-\infty}^{0}\int_G\int_{\mathbb{R}^d}\left|\phi(e^{-x}au)\right|\nu(du)\mu(db\,da)dx\\ &\leq \int_{-\infty}^{0}\int_{\mathbb{R}^d}\left|\phi(e^{-x}u)\right|\nu(du)dx + \int_{-\infty}^{0}\int_G\int_{\mathbb{R}^d}\left|\phi(e^{-x}au)\right|\nu(du)\mu(db\,da)dx\\ &\qquad\qquad\leq \int_{-\infty}^{0}\left|f_{|\phi|}(x)\right|dx + \int_{-\infty}^{0}\left|\overline{\mu} * f_{|\phi|}(x)\right|dx\end{aligned}$$

and in view of our previous estimates both integrals above are finite.

For $x > 0$ we divide the integral of $\left|\phi(e^{-x}au) - \phi(e^{-x}(b+au))\right|$ into several parts and we use the following inequality, being a consequence of the Lipschitz property of ϕ:

$$|\phi(s) - \phi(r)| \leq C|s-r|^\theta \max_{\xi\in\{|s|,|r|\}} \frac{1}{1+(\log^+\xi)^{\beta'}},$$

where $\theta < 1 - 2/\beta$ and $\beta' = \beta(1-\theta) > 2$. We denote by μ_A the law of A.

Case 1. First we assume $|b| \leq e^{\frac{x}{2}}$. Then by (24)

$$\int_{|b|\leq e^{\frac{x}{2}}} \int_{\mathbb{R}+} \Big|\phi(e^{-x}au) - \phi(e^{-x}(b+au))\Big| \nu(du)\mu(db\,da)$$

$$\leq C \int_{|b|\leq e^{\frac{x}{2}}} \int_{\mathbb{R}+} e^{-\theta x}|b|^{\theta} \bigg(\frac{1}{1+(\log^+(e^{-x}a|u|))^{\beta'}} + \frac{1}{1+(\log^+(e^{-x}|au+b|))^{\beta'}} \bigg) \nu(du)\mu(db\,da)$$

$$\leq Ce^{-\theta x/2} \bigg(\int_{\mathbb{R}+} \int_{\mathbb{R}+} \frac{1}{1+(\log^+(e^{-x}a|u|))^{\beta'}} \nu(du)\mu_A(da) + \int_{\mathbb{R}+} \frac{1}{1+(\log^+(e^{-x}|u|))^{\beta'}} \nu(du) \bigg)$$

$$\leq Ce^{-\theta x/2}\Big[1 + x + \int_{\mathbb{R}+} |\log a|\mu_A(da)\Big] < Ce^{-\theta x/4}.$$

Case 2. We assume $au < 2|au+b|$ and $|b| > e^{\frac{x}{2}}$. Notice first

$$\int_{|b|>e^{\frac{x}{2}}} \mu(db\,da) \leq \frac{C}{1+x^{4+\varepsilon}} \int_{\mathbb{R}+} \Big(1 + \big(\log^+|b|\big)^{4+\varepsilon}\Big)\mu(db\,da) \leq \frac{C}{1+x^{4+\varepsilon}}.$$

and

$$\int_{|b|>e^{\frac{x}{2}}} (|\log a| + \log|b|)\,\mu(db\,da)$$

$$\leq \frac{C}{1+x^{3+\varepsilon}} \int_G (1+\big(|\log a|+\log^+|b|\big)^{3+\varepsilon}(\log^+|b|+|\log a|)\mu(db\,da) \leq \frac{C}{1+x^{3+\varepsilon}}.$$

Then, proceeding as previously, we have

$$\int \int_{\substack{a|u|<2|au+b| \\ |b|>e^{\frac{x}{2}}}} \Big|\phi(e^{-x}au) - \phi(e^{-x}(b+au))\Big| \nu(du)\mu(db\,da)$$

$$\leq 2 \int \int_{\substack{a|u|<2|au+b| \\ |b|>e^{\frac{x}{2}}}} \max\Big\{\big|\phi(e^{-x}au)\big|, \big|\phi(e^{-x}(b+au))\big|\Big\} \nu(du)\mu(db\,da)$$

$$\leq C \int_{|b|>e^{\frac{x}{2}}} \int_{\mathbb{R}^d} \frac{1}{1 + (\log^+(e^{-x}a|u|))^\beta} \nu(du)\mu(db\,da)$$

$$\leq C \int_{|b|>e^{\frac{x}{2}}} \big(x + |\log a| + 1\big)\mu(da\,db) \leq \frac{C}{1 + x^{3+\varepsilon}}.$$

Case 3. The last case is $a|u| \geq 2|au + b|$ and $|b| > e^{\frac{x}{2}}$. Then $|u| < \frac{2|b|}{a}$ and we obtain

$$\int \int_{\substack{a|u|\geq 2|au+b| \\ |b|>e^{\frac{x}{2}}}} \Big|\phi(e^{-x}au) - \phi(e^{-x}(b + au))\Big| \nu(du)\mu(db\,da)$$

$$\leq C \int_{|b|>e^{\frac{x}{2}}} \int_{|u|<\frac{2|b|}{a}} \nu(du)\mu(db\,da)$$

$$\leq C \int_{|b|>e^{\frac{x}{2}}} \big(1 + \log|b| + |\log a|\big)\mu(db\,da) \leq \frac{C}{1 + x^{3+\varepsilon}}.$$

We conclude (23) and the required estimates for ψ_ϕ. □

Proof (Proof of Theorem 1). First, we are going to prove that the limit

$$\lim_{x\to+\infty} \int_{\mathbb{R}^+} \phi(ue^{-x})\nu(du) = T(\phi) := -2\sigma^{-2}K(\psi_\phi) \tag{25}$$

exists and is finite for a class of very particular functions, namely for functions of the form

$$\phi(u) = \int_{\mathbb{R}} r(t)\zeta(e^t u)dt, \tag{26}$$

where

$$r(t) = \mathbb{E}\,[|-\log A_1 - t| - |t|] \tag{27}$$

and ζ is a nonnegative Lipschitz function on $\mathbb{R}^+$ such that $\zeta(u) \leq e^{-\gamma|\log|u||}$ for some $\gamma > 0$.

For this purpose we are going to prove that ψ_ϕ is an element of $\mathscr{F}(\overline{\mu})$ and $J(\psi_\phi) = 0$. Then, by Corollary 1, the function $f_\phi(x)$, is a solution of the corresponding Poisson equation, and thus it is bounded and has a limit when x converge to $+\infty$.

In view of (7),

$$
\begin{aligned}
|\phi(u)| &\leq C \int_{\mathbb{R}} \frac{1}{1+|t-\log|u||^{3+\varepsilon}} e^{-\gamma|t|} dt \\
&\leq \frac{C}{1+|\log|u||^{3+\varepsilon}} \int_{\mathbb{R}} \frac{1+|t-\log|u||^{3+\varepsilon}+|t|^{3+\varepsilon}}{1+|t-\log|u||^{3+\varepsilon}} e^{-\gamma|t|} dt \\
&\leq \frac{C}{1+|\log|u||^{3+\varepsilon}} \int_{\mathbb{R}} (1+|t|^{3+\varepsilon}) e^{-\gamma|t|} dt \leq \frac{C}{1+|\log|u||^{3+\varepsilon}}.
\end{aligned}
$$

Thus by Lemma 2, f_ϕ, f_ζ, $\overline{\mu} * f_\phi$ and $\overline{\mu} * f_\zeta$ are well defined. Furthermore since ζ is Lipschitz ψ_ζ is bounded, and $x^2\psi_\zeta(x)$ is integrable on $\mathbb{R}$. We cannot guarantee that ϕ is Lipschitz, but we can observe that

$$
f_\phi(x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} r(t)\zeta(e^{-x+t}u) dt \nu(du) = \int_{\mathbb{R}^d} \int_{\mathbb{R}} r(t+x)\zeta(e^t u) dt \, \nu(du) = r *_{\mathbb{R}} f_\zeta(x)
$$

and

$$
\overline{\mu} * f_\phi(x) = r *_{\mathbb{R}} (\overline{\mu} * f_\zeta)(x).
$$

Hence

$$
\psi_\phi = f_\phi - \overline{\mu} * f_\phi = r * (f_\zeta - \overline{\mu} * f_\zeta) = r *_{\mathbb{R}} \psi_\zeta.
$$

Therefore, by Lemma 1, $\psi_\phi \in \mathscr{F}(\overline{\mu})$.

Furthermore $J(\psi_\phi) = 0$. In fact,

$$
\begin{aligned}
\int_{\mathbb{R}} \psi_\zeta(x) dx &= \int_G \int_{\mathbb{R}^+} \int_{\mathbb{R}} \left[\zeta\left(e^{-x+\log(|au|)}\right) - \zeta_r\left(e^{-x+\log|au+b|}\right) \right] dx \nu(du) \mu(db\, da) \\
&= \int_G \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}} \zeta(e^{-x}) dx - \int_{\mathbb{R}} \zeta(e^{-x}) dx \right) \nu(du) \mu(db\, da) = 0.
\end{aligned}
$$

Observe that we can apply the Fubini theorem since ζ is Lipschitz and, by Lemma 2, the absolute value of the integrand in the second line above is integrable. Hence

$$
J(\psi_\phi) = \int_{\mathbb{R}} \psi_\phi(x) dx = \int_{\mathbb{R}} r * \psi_\zeta(x) dx = \int_{\mathbb{R}} r(x) dx \cdot \int_{\mathbb{R}} \psi_\zeta(x) dx = 0.
$$

By Corollary 1, we have

$$
f_\phi = A\psi_\phi + C_\phi \tag{28}
$$

where C_ϕ is a constant. Thus, f_ϕ is bounded.

In particular the same holds for f_{Φ_γ}, where

$$\Phi_\gamma(u) = \int_{\mathbb{R}} r(t) e^{-\gamma|t+\log|u||} dt.$$

Since zero does not belong to the support of ν, $\lim_{x\to-\infty} f_\phi(x) = 0$ and by Theorem 2

$$-C_\phi = \lim_{x\to-\infty} A\psi_\phi(x) = \sigma^{-2} K(\psi_\phi).$$

Thus when x goes to $-\infty$ the limit of h_ϕ exists which is possible only if h_ϕ is constant and is equal to $-\sigma^{-2}K(\psi_\phi)$. Finally

$$\lim_{x\to+\infty} f_\phi(x) = \lim_{x\to+\infty} A\psi_\phi(x) - \sigma^{-2}K(\psi_\phi) = -2\sigma^{-2}K(\psi_\phi)$$

and we obtain (25).

Fix a $\gamma > 0$. Since $\Phi_\gamma > 0$ for every function $\phi \in C_c(\mathbb{R}^+)$ there exists a constant C_ϕ such that $|\phi| \le C_\phi \Phi_\gamma$. Thus the family of measures on $\mathbb{R}^+$

$$\delta_{(0,e^{-x})} *_G \nu(\phi) = \int_{\mathbb{R}^+} \phi(e^{-x}u)\nu(du)$$

is bounded, hence it is relatively compact in the weak topology. Let η be an accumulation point for a subsequence $\{x_n\}$ that is

$$\lim_{n\to\infty} \delta_{(0,e^{-x_n})} *_G \nu(\phi) = \eta(\phi) \quad \forall \phi \in C_c(\mathbb{R}^+). \tag{29}$$

The measure η is $\mathbb{R}^+$ invariant [5], thus η must be of the form $\eta(da) = C_\eta \frac{dx}{x}$. A standard argument proves indeed that for any continuous non negative function such that $\phi \le C_\phi \Phi_\gamma$, not necessarily compactly supported,

$$\eta(\phi) = \lim_{n\to\infty} \delta_{(0,e^{-x_n})} *_G \nu(\phi).$$

In particular the last formula holds for $\Phi_\gamma(u) = \int_{\mathbb{R}} r(t) e^{-\gamma|t+\log|u||} dt$, since $\eta(\Phi_\gamma) = C_\eta \int_{\mathbb{R}^*_+} \Phi_\gamma(u) \frac{du}{u}$. Then:

$$C_\eta = \frac{T(\Phi_\gamma)}{\int_{\mathbb{R}^*_+} \Phi_\gamma(u) \frac{du}{u}}$$

does not depend on η. Thus, finally, we deduce that the limit

$$\lim_{z\to+\infty} \int_{\mathbb{R}^+} \phi(uz^{-1})\nu(du)$$

exists for every function $\phi \in C_c(\mathbb{R}^+)$ and defines a Radon measure Λ on $\mathbb{R}^+$. This limiting measure must be $\mathbb{R}^+$ invariant, therefore is of the form $C\,\frac{du}{u}$, that by a standard argument implies also (8). For the proof of strict positivity of C see [5] (Theorem 5.1). □

Acknowledgements Dariusz Buraczewski and Ewa Damek were supported by the NCN grant DEC-2012/05/B/ST1/00692.

References

1. Addario-Berry, L., Reed, B.: Minima in branching random walks. Ann. Probab. **37**(3), 1044–1079 (2009)
2. Alsmeyer, G., Meiners, M.: Fixed points of inhomogeneous smoothing transforms. J. Diff. Equ. Appl. **18**, 1287–1304 (2012)
3. Alsmeyer, G., Meiners, M.: Fixed points of the smoothing transform: two-sided solutions. Probab. Theory Relat. Fields (2012)
4. Babillot, M., Bougerol, P., Elie, L.: The random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. Ann. Probab. **25**(1), 478–493 (1997)
5. Brofferio, S., Buraczewski, D., Damek, E.: On the invariant measure of the random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. Ann. Inst. Henri Poincaré Probab. Stat. **48**(2), 377–395 (2012)
6. Brofferio, S., Buraczewski, D.: On unbounded invariant measures of stochastic dynamical systems, preprint, arxiv.org/abs/1304.7145
7. Buraczewski, D.: On invariant measures of stochastic recursions in a critical case. Ann. Appl. Probab. **17**(4), 1245–1272 (2007)
8. Buraczewski, D.: On tails of fixed points of the smoothing transform in the boundary case. Stoch. Process. Appl. **119**(11), 3955–3961 (2009)
9. Buraczewski, D., Damek, E., Zienkiewicz, J.: Precise tail asymptotics of fixed points of the smoothing transform with general weights. (preprint)
10. Buraczewski, D., Kolesko, K.: Linear stochastic equations in the critical case. J. Diff. Equ. Appl. (To appear)
11. Durrett, R., Liggett, T.M.: Fixed points of the smoothing transformation. Z. Wahrsch. Verw. Gebiete **64**(3), 275–301 (1983)
12. Feller, W.: An Introduction to Probability Theory and its Applications, vol. II, 2nd edn. Wiley, New York (1971)
13. Goldie, C.M.: Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. **1**(1), 126–166 (1991)
14. Grincevičjus, A.K.: Limit theorem for products of random linear transformations of the line. Litovsk. Mat. Sb. **15**(4), 61–77, 241 (1975)
15. Guivarc'h, Y.: Sur une extension de la notion de loi semi-stable. Ann. Inst. H. Poincaré Probab. Stat. **26**(2), 261–285 (1990)
16. Hu, Y., Shi, Z.: Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. Ann. Probab. **37**(2), 742–789 (2009)
17. Jelenković, P.R., Olvera-Cravioto, M.: Implicit renewal theory and power tails on trees. Adv. Appl. Probab. **44**(2), 528-561 (2012). www.arxiv.org: 1006.3295v3
18. Jelenković, P.R., Olvera-Cravioto, M.: Information ranking and power laws on trees. Adv. Appl. Probab. **42**(4), 1057–1093 (2010)
19. Kesten, H.: Random difference equations and renewal theory for products of random matrices. Acta Math. **131**, 207–248 (1973)

20. Kolesko, K.: Tail homogeneity of invariant measures of multidimensional stochastic recursions in a critical case. Probab. Theory Relat. Fields (to appear)
21. Liu, Q.: Fixed points of a generalized smoothing transformation and applications to the branching random walk. Adv. Appl. Probab. **30**(1), 85–112 (1998)
22. Liu, Q.: On generalized multiplicative cascades. Stoch. Process. Appl. **86**(2), 263–286 (2000)
23. Neininger, R., Rüschendorf, L.: A general limit theorem for recursive algorithms and combinatorial structures. Ann. Appl. Probab. **14**(1), 378–418 (2004)
24. Port, S.C., Stone, C.J.: Hitting time and hitting places for non-lattice recurrent random walks. J. Math. Mech. **17**, 35–57 (1967)
25. Port, S.C., Stone, C.J.: Potential theory of random walks on Abelian groups. Acta Math. **122**, 19–114 (1969)
26. Rösler, U.: On the analysis of stochastic divide and conquer algorithms. Algorithmica **29**(1–2), 238–261 (2001). (Average-case Analysis of Algorithms, 1998, Princeton)

Power Laws on Weighted Branching Trees

Predrag R. Jelenković and Mariana Olvera-Cravioto

Abstract Consider distributional fixed-point equations of the form

$$R \stackrel{\mathcal{D}}{=} f(Q, C_i, R_i, 1 \leq i \leq N),$$

where $f(\cdot)$ is a possibly random real-valued function, $N \in \{0, 1, 2, 3, \ldots\} \cup \{\infty\}$, $\{C_i\}_{i \in \mathbb{N}}$ are real-valued random weights and $\{R_i\}_{i \in \mathbb{N}}$ are iid copies of R, independent of $(Q, N, C_1, C_2, \ldots)$; $\stackrel{\mathcal{D}}{=}$ represents equality in distribution. Fixed-point equations of this type are important for solving many applied probability problems, ranging from the average case analysis of algorithms to statistical physics. In this paper we present some of our recent work from [26–28, 36] that studies the power tail asymptotics of such solutions. We exemplify our techniques primarily on the nonhomogeneous equation, $R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q$, for which the power tail of the solution, $P(R > t)$, can be determined by three different factors: the multiplicative effect of the weights C_i; the sum of the weights $\sum C_i$; and the innovation variable Q.

1 Introduction

Our recent work on the analysis of recursions on weighted branching trees is motivated by the study of the nonhomogeneous linear fixed-point equation

P.R. Jelenković
Department of Electrical Engineering, Columbia University, New York, NY, USA
e-mail: predrag@ee.columbia.edu

M. Olvera-Cravioto (✉)
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, USA
e-mail: molvera@ieor.columbia.edu

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_8,

$$R \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} C_i R_i + Q, \tag{1}$$

where $(Q, N, C_1, C_2, \dots)$ is a real-valued random vector with $N \in \mathbb{N} \cup \{\infty\}$, $\mathbb{N} = \{0, 1, 2, 3, \dots\}$, $P(|Q| > 0) > 0$, and $\{R_i\}_{i \in \mathbb{N}}$ is a sequence of iid random variables independent of $(Q, N, C_1, C_2, \dots)$ having the same distribution as R. This recursion has been proposed as a stochastic approximation of Google's PageRank algorithm and possibly other ranking schemes of large information sets, e.g., the World Wide Web (WWW); see [26, 46, 47] and the references therein. There it is argued that the stochastic approach is helpful in understanding the qualitative behavior of PageRank given the large scale nature of the WWW. These types of weighted recursions, also studied in the literature on weighted branching processes [42] and branching random walks [10], are found in the probabilistic analysis of other algorithms as well [1, 40, 43]. The homogeneous ($Q \equiv 0$) version of (1) has been studied extensively in the literature of weighted branching processes and multiplicative cascades, see [2, 6, 10, 17, 23, 24, 30, 33, 34, 38, 48] and the references therein.

We now give some more details on the PageRank motivation that was mentioned above. PageRank assigns to each page a numerical weight that measures its relative importance with respect to other pages. We think of the Web as a very large interconnected graph where nodes correspond to pages. The Google trademarked algorithm PageRank defines the page rank as:

$$R(p_i) = \frac{1-d}{n} + d \sum_{p_j \in M(p_i)} \frac{R(p_j)}{L(p_j)}, \tag{2}$$

where, using Google's notation, $p_1, p_2, \dots, p_n$ are the pages under consideration, $M(p_i)$ is the set of pages that link to p_i, $L(p_j)$ is the number of outbound links on page p_j, n is the total number of pages on the Web, and d is a damping factor, usually $d = 0.85$. While in principle the solution to (2) reduces to the solution of a large system (possibly billions) of linear equations, we believe that finding page ranks in such a way is unlikely to be insightful.

In particular, the division by the out-degree, $L(p_j)$ in Eq. (2), was meant to decrease the contribution of pages with highly inflated referencing, i.e., those pages that basically point/reference possibly indiscriminately to other documents. However, the stochastic approach reveals that highly ranked pages are essentially insensitive to the parameters of the out-degree distribution, and high ranks most likely occur either due to a pointer by a very highly ranked neighbor, or by pointers of a very large number of neighbors. Hence, PageRank may not reduce the effects of overly inflated referencing.

A stochastic approach to analyze (2) is to multiply it by n and consider a typical node on the graph

$$R \stackrel{\mathscr{D}}{=} (1-d) + d\sum_{i=1}^{N} \frac{R_i}{D_i}, \tag{3}$$

where $d > 0$, $dE[1/D] < 1$, N is a random variable independent of the R_i's and D_i's, the D_i's are iid random variables satisfying $D_i \geq 1$, and the R_i's are iid random variables having the same distribution as R. In terms of recursion (2), R is the scale-free rank of a random page, N corresponds to the in-degree of that node, the R_i's are the ranks of the pages pointing to it, and the D_i's correspond to the so-called *effective* degrees of each of these pages. The experimental justification of these independence assumptions can be found in [45]. This stochastic setup was first introduced in [47], where the process resulting after a finite number of iterations of (3) was analyzed. Further generalization of (3) leads to (1), which was recently analyzed in [26, 46].

Furthermore, in computer science, a well known divide-and-conquer paradigm is used for designing efficient algorithms, where a problem is recursively divided into two or more sub-problems, until the sub-problems become simple enough to be solved directly. Such approach naturally leads to a recursive analysis, which in the case of randomized algorithms, often results in stochastic recursions of the type in (1). Among these, the most widely analyzed algorithm is Quicksort, whose analysis, after an appropriate normalization introduced in [41], reduces to the stochastic fixed point equation

$$R \stackrel{\mathscr{D}}{=} UR_1 + (1-U)R_2 + Q,$$

where U is uniform, $Q = Q(U)$, $\{R_1, R_2\}$ are independent copies of R and independent of (U, Q); for recent work see [18, 39] and the references therein. Similar binary equations also appear in the analysis of sequential absorption (packing) problems on a line, see Eq. (19) in [9]. Such problems are used for modeling memory fragmentation, advance reservation, particle absorption, e.g., see [15] and the references therein. Multidimensional versions of the fixed-point equation (1) have been considered in [39, 40] and more recently in [13].

In general, many applied probability problems, appearing in the average case analysis of algorithms and statistical physics, reduce to distributional fixed-point equations of the form

$$R \stackrel{\mathscr{D}}{=} f(Q, C_i, R_i, 1 \leq i < N+1), \tag{4}$$

where $f(\cdot)$ is a possibly random real-valued function, $N \in \mathbb{N} \cup \{\infty\}$, the $\{C_i\}_{i\in\mathbb{N}}$ are real-valued random weights and $\{R_i\}_{i\in\mathbb{N}}$ are iid copies of R, independent of $(Q, N, C_1, C_2, \ldots)$. For example, as discussed in [27], one can study the following distributional equations

$$R \stackrel{\mathcal{D}}{=} \left(\bigvee_{i=1}^{N} C_i R_i \right) \vee Q, \quad R \stackrel{\mathcal{D}}{=} \left(\bigvee_{i=1}^{N} C_i R_i \right) + Q, \quad R \stackrel{\mathcal{D}}{=} \left(\sum_{i=1}^{N} C_i R_i \right) \vee Q. \tag{5}$$

The solutions to equations of this type can be recursively constructed on a weighted branching tree, where N represents the generic branching variable and the $\{C_i\}_{i \in \mathbb{N}}$ are the branching weights. For this reason, we also refer to (4) as recursions on weighted branching trees. The maximum recursion, the first one in (5), was previously studied in [5] under the assumption that $Q \equiv 0$, $N = \infty$, and the $\{C_i\}$ are real-valued deterministic constants, and the case of $Q \equiv 0$ and $\{C_i\} \geq 0$ being random was studied earlier in [25]. Furthermore, these max-plus type stochastic recursions appear in a wide variety of applications, see [1] for a recent survey.

Special cases of the preceding recursions are important in many applied probability areas. For example, selecting $N = 1$ in (1) yields the fixed-point equation satisfied by the first order autoregressive process. When $Q = 1, C_i \equiv 1$, the steady state solution to (1) represents the total number of individuals born in an ordinary branching process. Similarly, by setting $N = Q \equiv 1$ and $X_i = \log C_i$ in the first equation in (5), one obtains the well studied supremum of a random walk and in particular the waiting time in the GI/GI/1 queue. By choosing the distributions of N, Q or C_i appropriately, all of these recursions can lead to heavy-tailed solutions.

In this paper we present some of our recent work from [26–28, 36] that studies the power-tail asymptotics of the solution R to the preceding distributional fixed-point equations; all the omitted proofs can be found in these references. We will exemplify our techniques primarily on the nonhomogeneous equation (1), for which the tail of the solution, $P(R > t)$, can be determined by three different factors: the multiplicative effect of the weights C_i; the sum of the weights $\sum C_i$; and the innovation variable Q. In addition, to simplify the exposition, we only present the results for the case where $(Q, N, C_1, C_2, \dots)$ is *nonnegative* and, when appropriate, we comment on the corresponding real-valued extensions.

First, we study the multiplicative effect of the weights by extending the implicit renewal theory of Goldie [19], which was derived for equations of the form $R \stackrel{\mathcal{D}}{=} f(Q, C, R)$ (equivalently $N \equiv 1$ in our case), to cover recursions on weighted branching trees. The extension of Goldie's theorem is presented in Theorem 1 of Sect. 3, and it enables the characterization of the power-tail behavior of the solutions R to many equations of the form in (4), e.g., those stated in (1) and (5). One of the observations that allows this extension is that an appropriately constructed measure on a weighted branching tree is a renewal measure, see Lemma 1.

Then, in Sect. 4, we develop the necessary large deviations techniques that will enable us to study the tail behavior of $P(R > t)$ when it is determined by the sum of the weights, $\sum C_i$, or the innovation variable Q. The key technical contribution is the derivation of uniform bounds (in n and x) for the distribution of the sum of the weights in the nth generation of a weighted branching tree, $P(W_n > x)$, given in Propositions 1 and 2. These uniform bounds are used to establish the geometric rate of convergence of the iterations of the fixed-point equation (1) to the solution R constructed in Sect. 5.

Next, we exemplify the techniques we have developed on the nonhomogeneous linear recursion (1). In this regard, in Sect. 5, we first construct an explicit solution (13) to (1) on a weighted branching tree and then provide sufficient conditions for the finiteness of moments and the uniqueness of this solution under iterations in Lemmas 5 and 6, respectively. However, the fixed-point equation (1) can have additional stable solutions that do not satisfy Lemma 6, as it was recently discovered in [3]. Earlier work for the case when $\{C_i\}, Q$ are deterministic real-valued constants can be found in [4, 42]. Furthermore, it is worth noting that our moment estimates are explicit, see Lemma 4, which may be of independent interest. Our first main result about the constructed solution R to (1) is given in Theorem 2, where through the extension of the Implicit Renewal Theorem it is shown that the multiplicative nature of the weights can lead to a power-tail behavior. Informally, our result shows, under some moment conditions, that

$$P(R > x) \sim \frac{H}{x^\alpha} \quad \text{as} \quad x \to \infty,$$

where α is a solution to $E\left[\sum_{i=1}^N C_i^\alpha\right] = 1$. In addition, for integer power exponent ($\alpha \in \{1, 2, 3, \ldots\}$) the constant H can be explicitly computed as stated in Corollary 1. Furthermore, for non integer α, we will explain how Lemma 2 can be used to obtain an explicit bound on H.

When the conditions for the Implicit Renewal Theorem fail, the tail behavior of R can be determined by $P\left(\sum_{i=1}^N C_i > x\right)$ or $P(Q > x)$. Using our work on the large deviations of weighted random sums we give the corresponding results in Theorems 3 and 4, respectively. In particular, it is shown that if $P\left(\sum_{i=1}^N C_i > x\right)$ or $P(Q > x)$, are regularly varying with index $\alpha > 1$, and certain moment conditions are satisfied, then, respectively,

$$P(R > x) \sim H_S P\left(\sum_{i=1}^N C_i > x\right) \quad \text{or} \quad P(R > x) \sim H_Q P(Q > x)$$

as $x \to \infty$, for some explicit constants $H_S, H_Q > 0$. Lastly, we point out that we focus here only on the heavy-tailed solutions to (1), but it is known that (1) can also have light-tailed solutions, see [20] for the $N \equiv 1$ case and the discussion after Theorem 2.2 in [34] for the general branching case.

We conclude the paper with a brief analysis of other non-linear recursions, e.g., those stated in (5), that could be studied using the extension of the Implicit Renewal Theorem. The main difficulty in applying Theorem 1 is in verifying the conditions of the theorem for a specific fixed-point equation. In this regard, we argue that the two technical lemmas, Lemmas 7 and 8, can be helpful for this purpose.

The rest of the paper is organized as follows. Section 2 contains the construction of the weighted branching tree. In Sect. 3 we present the extension of the implicit renewal theorem to trees and, in Sect. 4, we derive the uniform large deviation bounds for $P(W_n > x)$. Section 5 exemplifies our techniques on the nonhomogeneous linear equation (1), and Sect. 6 briefly discusses how the developed tools can be applied to other fixed-point equations, e.g., those in (5).

2 Weighted Branching Tree

First we construct a random tree $\mathscr{T}$. We use the notation $\emptyset$ to denote the root node of $\mathscr{T}$, and A_n, $n \geq 0$, to denote the set of all individuals in the nth generation of $\mathscr{T}$, $A_0 = \{\emptyset\}$. Let Z_n be the number of individuals in the nth generation, that is, $Z_n = |A_n|$, where $|\cdot|$ denotes the cardinality of a set; in particular, $Z_0 = 1$.

Next, let $\mathbb{N}_+ = \{1, 2, 3, \dots\}$ be the set of positive integers and let $U = \bigcup_{k=0}^{\infty} (\mathbb{N}_+)^k$ be the set of all finite sequences $\mathbf{i} = (i_1, i_2, \dots, i_n)$, where by convention $\mathbb{N}_+^0 = \{\emptyset\}$ contains the null sequence $\emptyset$. To ease the exposition, for a sequence $\mathbf{i} = (i_1, i_2, \dots, i_k) \in U$ we write $\mathbf{i}|n = (i_1, i_2, \dots, i_n)$, provided $k \geq n$, and $\mathbf{i}|0 = \emptyset$ to denote the index truncation at level n, $n \geq 0$. Also, for $\mathbf{i} \in A_1$ we simply use the notation $\mathbf{i} = i_1$, that is, without the parenthesis. Similarly, for $\mathbf{i} = (i_1, \dots, i_n)$ we will use $(\mathbf{i}, j) = (i_1, \dots, i_n, j)$ to denote the index concatenation operation, if $\mathbf{i} = \emptyset$, then $(\mathbf{i}, j) = j$.

We iteratively construct the tree as follows. Let N be the number of individuals born to the root node $\emptyset$, $N_\emptyset = N$, and let $\{N_\mathbf{i}\}_{\mathbf{i} \in U, \mathbf{i} \neq \emptyset}$ be iid copies of N. Define now

$$A_1 = \{i \in \mathbb{N} : 1 \leq i \leq N\}, \quad A_n = \{(\mathbf{i}, i_n) \in U : \mathbf{i} \in A_{n-1}, 1 \leq i_n \leq N_\mathbf{i}\}. \tag{6}$$

It follows that the number of individuals $Z_n = |A_n|$ in the nth generation, $n \geq 1$, satisfies the branching recursion

$$Z_n = \sum_{\mathbf{i} \in A_{n-1}} N_\mathbf{i}.$$

Now, we construct the weighted branching tree $\mathscr{T}_{Q,C}$ as follows. We start by assigning the vector $(Q_\emptyset, N_\emptyset, C_{(\emptyset,1)}, C_{(\emptyset,2)}, \dots) \equiv (Q, N, C_1, C_2, \dots)$ to the root node $\emptyset$. Next, let $\{(Q_\mathbf{i}, N_\mathbf{i}, C_{(\mathbf{i},1)}, C_{(\mathbf{i},2)}, \dots)\}_{\mathbf{i} \in U, \mathbf{i} \neq \emptyset}$ be a sequence of iid copies of $(Q, N, C_1, C_2, \dots)$. Recall that $N_\emptyset$ determines the number of nodes in the first generation of $\mathscr{T}$ according to (6), and assign to each node in the first generation its corresponding vector $(Q_i, N_i, C_{(i,1)}, C_{(i,2)}, \dots)$ from the preceding iid sequence. In general, for $n \geq 2$, to each node $\mathbf{i} \in A_{n-1}$ we assign its corresponding vector $(Q_\mathbf{i}, N_\mathbf{i}, C_{(\mathbf{i},1)}, C_{(\mathbf{i},2)}, \dots)$ from the sequence and construct $A_n = \{(\mathbf{i}, i_n) \in U : \mathbf{i} \in A_{n-1}, 1 \leq i_n \leq N_\mathbf{i}\}$. For each node in $\mathscr{T}_{Q,C}$ we also define the weight $\Pi_{(i_1,\dots,i_n)}$ via the recursion

$$\Pi_{i_1} = C_{i_1}, \qquad \Pi_{(i_1,\dots,i_n)} = C_{(i_1,\dots,i_n)} \Pi_{(i_1,\dots,i_{n-1})}, \quad n \geq 2,$$

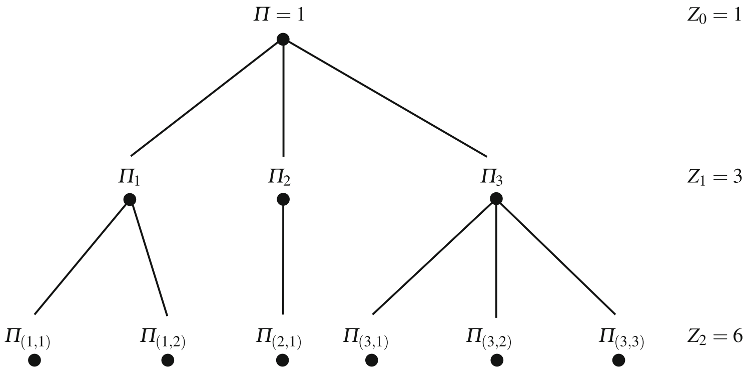


Fig. 1 Weighted branching tree

where $\Pi = 1$ is the weight of the root node. Note that the weight $\Pi_{(i_1,\ldots,i_n)}$ is equal to the product of all the weights $C_{(\cdot)}$ along the branch leading to node $(i_1,\ldots,i_n)$, as depicted in Fig. 1. In some places, e.g., in the following section, the value of Q may be of no importance, and thus we will consider a weighted branching tree defined by the smaller vector $(N, C_1, C_2, \ldots)$. This tree can be obtained from $\mathscr{T}_{Q,C}$ by simply disregarding the values for $Q_{(\cdot)}$ and is denoted by $\mathscr{T}_C$.

The objective of this paper is to present a variety of results that analyze recursions and fixed-point equations embedded in this weighted branching tree.

3 Implicit Renewal Theorem on Trees

In this section we present an extension of Goldie's Implicit Renewal Theorem [19] to weighted branching trees. The observation that facilitates this generalization is the following lemma which shows that a certain measure on a tree is actually a product measure; a similar measure was used in a different context in [11]. Throughout the paper we use the standard convention $0^\alpha \log 0 = 0$ for all $\alpha > 0$, and the notation $x^+ = \max\{x, 0\}$, $x^- = -\min\{x, 0\}$.

Lemma 1. *Let $\mathscr{T}_C$ be the weighted branching tree defined by the nonnegative vector $(N, C_1, C_2, \ldots)$, where $N \in \mathbb{N} \cup \{\infty\}$. For any $n \in \mathbb{N}$ and $\mathbf{i} \in A_n$, let $V_{\mathbf{i}} = \log \Pi_{\mathbf{i}}$. For $\alpha > 0$ define the measure*

$$\mu_n(dt) = e^{\alpha t} E\left[\sum_{\mathbf{i} \in A_n} 1(V_{\mathbf{i}} \in dt)\right], \quad n = 1, 2, \ldots,$$

and let $\eta(dt) = \mu_1(dt)$. Suppose that there exists $j \geq 1$ with $P(N \geq j, C_j > 0) > 0$ such that the measure $P(\log C_j \in du, C_j > 0, N \geq j)$ is nonarithmetic,

$E\left[\sum_{i=1}^{N} C_i^{\gamma}\right] < \infty$ *for some* $0 \le \gamma < \alpha$, *and* $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] = 1$. *Then,* $\eta(\cdot)$ *is a nonarithmetic probability measure on* $\mathbb{R}$ *that places no mass at* $-\infty$ *and has mean*

$$\mu \triangleq \int_{-\infty}^{\infty} u\, \eta(du) = E\left[\sum_{j=1}^{N} C_j^{\alpha} \log C_j\right].$$

Furthermore, $\mu_n(dt) = \eta^{*n}(dt)$, *where* η^{*n} *denotes the nth convolution of* η *with itself.*

Note that $E\left[\sum_{i=1}^{N} C_i^{\gamma}\right] < \infty$ and $E\left[\sum_{i=1}^{N} C_i^{\alpha}\right] < \infty$ for $0 \le \gamma < \alpha$ implies $E\left[\sum_{i=1}^{N} C_i^{\alpha} (\log C_i)^{-}\right] < \infty$, and therefore the mean of $\eta(\cdot)$ is well defined.

We now present Theorem 3.1 of [27], which is a generalization of Goldie's Implicit Renewal Theorem [19] that enables the analysis of recursions on weighted branching trees. Note that except for the independence assumption, the random variable R and the vector $(N, C_1, C_2, \ldots)$ are arbitrary, and therefore the applicability of this theorem goes beyond the recursions that we study here. When this theorem is applied to specific recursions, one can use the nature of the recursion to verify the conditions of the theorem. Typically, it is the absolute integrability in (7) that requires the most work. Throughout the paper we use $g(x) \sim f(x)$ as $x \to \infty$ to denote $\lim_{x\to\infty} g(x)/f(x) = 1$.

Theorem 1. *Let* $(N, C_1, C_2, \ldots)$ *be a nonnegative random vector, where* $N \in \mathbb{N} \cup \{\infty\}$. *Suppose that there exists* $j \ge 1$ *with* $P(N \ge j, C_j > 0) > 0$ *such that the measure* $P(\log C_j \in du, C_j > 0, N \ge j)$ *is nonarithmetic. Assume further that* $0 < E\left[\sum_{j=1}^{N} C_j^{\alpha} \log C_j\right] < \infty$, $E\left[\sum_{j=1}^{N} C_j^{\alpha}\right] = 1$, $E\left[\sum_{j=1}^{N} C_j^{\gamma}\right] < \infty$ *for some* $0 \le \gamma < \alpha$, *and that* $R \ge 0$ *is independent of* $(N, C_1, C_2, \ldots)$ *with* $E[R^{\beta}] < \infty$ *for any* $0 < \beta < \alpha$. *If*

$$\int_0^{\infty} \left| P(R > t) - E\left[\sum_{j=1}^{N} 1(C_j R > t)\right]\right| t^{\alpha-1} dt < \infty, \tag{7}$$

then

$$P(R > t) \sim H t^{-\alpha}, \qquad t \to \infty,$$

where $0 \le H < \infty$ *is given by*

$$H = \frac{1}{E\left[\sum_{j=1}^{N} C_j^{\alpha} \log C_j\right]} \int_0^{\infty} v^{\alpha-1} \left(P(R > v) - E\left[\sum_{j=1}^{N} 1(C_j R > v)\right]\right) dv.$$

Remark 1. (i) As pointed out in [19], the statement of the theorem only has content when R has infinite moment of order α, since otherwise the constant H is zero. (ii) This theorem was recently generalized in Theorem 3.4 of [28] to incorporate real-valued weights $\{C_i\}$ and real-valued R. The proof utilizes the matrix analogue of the renewal theorem from [44]. (iii) When the $\{\log C_i\}$ are lattice-valued, a similar version of the theorem was derived by using the corresponding Renewal Theorem for lattice random walks, see Theorem 3.7 in [28]. (iv) To see that the condition $E\left[\sum_{j=1}^{N} C_j^{\gamma}\right] < \infty$ for some $0 \leq \gamma < \alpha$ is needed, consider the following example. Fix $k \geq 2$ to be such that $A = \sum_{j=k}^{\infty} 1/(j(\log j)^3)$ and $B = \sum_{j=k}^{\infty} (\log j + 3 \log\log j)/(j(\log j)^3)$ are both smaller than 1/2, and choose $C = e^X$ where X is exponentially distributed with mean $(1 - A)$. Now set $C_j = C/(j(\log j)^3)$ for $j \geq k$ and $C_j = 0$ otherwise ($N = \infty$). Then, $E\left[\sum_{j=k}^{\infty} C_j\right] = 1$ and $E\left[\sum_{j=k}^{\infty} C_j \log C_j\right] = A^{-1}(1 - A - B) > 0$, but $E\left[\sum_{j=k}^{\infty} C_j^{\gamma}\right] = \infty$ for any $0 \leq \gamma < 1$. (v) As noted in [19], the early ideas of applying renewal theory to study the power tail asymptotics of autoregressive processes (perpetuities) is due to [31] and [22].

Sketch of the proof of Theorem 1. Let $\mathscr{T}_C$ be the weighted branching tree defined by the nonnegative vector $(N, C_1, C_2, \ldots)$. For each $\mathbf{i} \in A_n$ and all $k \leq n$ define $V_{\mathbf{i}|k} = \log \Pi_{\mathbf{i}|k}$; note that $\Pi_{\mathbf{i}|k}$ is independent of $N_{\mathbf{i}|k}$ but not of $N_{\mathbf{i}|s}$ for any $0 \leq s \leq k-1$. Also note that $\mathbf{i}|n = \mathbf{i}$ since $\mathbf{i} \in A_n$. Let $\mathscr{F}_k$, $k \geq 1$, denote the σ-algebra generated by $\left\{(N_{\mathbf{i}}, C_{(\mathbf{i},1)}, C_{(\mathbf{i},2)}, \ldots) : \mathbf{i} \in A_j, 0 \leq j \leq k-1\right\}$, and let $\mathscr{F}_0 = \sigma(\emptyset, \Omega)$, $\Pi_{\mathbf{i}|0} \equiv 1$. Assume also that R is independent of the entire weighted tree, $\mathscr{T}_C$. Then, for any $t \in \mathbb{R}$, we can write $P(R > e^t)$ via a telescoping sum as follows

$$
\begin{aligned}
&P(R > e^t) \\
&= \sum_{k=0}^{n-1} \left(E\left[\sum_{(\mathbf{i}|k) \in A_k} 1(\Pi_{\mathbf{i}|k} R > e^t) \right] - E\left[\sum_{(\mathbf{i}|k+1) \in A_{k+1}} 1(\Pi_{\mathbf{i}|k+1} R > e^t) \right] \right) \\
&\quad + E\left[\sum_{(\mathbf{i}|n) \in A_n} 1(\Pi_{\mathbf{i}|n} R > e^t) \right] \\
&= \sum_{k=0}^{n-1} E\left[\sum_{(\mathbf{i}|k) \in A_k} \left(1(\Pi_{\mathbf{i}|k} R > e^t) - \sum_{j=1}^{N_{\mathbf{i}|k}} 1(\Pi_{\mathbf{i}|k} C_{(\mathbf{i}|k,j)} R > e^t) \right) \right] \\
&\quad + E\left[\sum_{(\mathbf{i}|n) \in A_n} 1(\Pi_{\mathbf{i}|n} R > e^t) \right]
\end{aligned}
$$

$$= \sum_{k=0}^{n-1} E\left[\sum_{(\mathbf{i}|k)\in A_k} E\left[1(R > e^{t-V_{\mathbf{i}|k}}) - \sum_{j=1}^{N_{\mathbf{i}|k}} 1(C_{(\mathbf{i}|k,j)} R > e^{t-V_{\mathbf{i}|k}}) \middle| \mathscr{F}_k \right]\right]$$

$$+ E\left[\sum_{(\mathbf{i}|n)\in A_n} 1(\Pi_{\mathbf{i}|n} R > e^t) \right].$$

Now, define the measures μ_n according to Lemma 1 and let

$$\nu_n(dt) = \sum_{k=0}^{n} \mu_k(dt), \qquad g(t) = e^{\alpha t}\left(P(R > e^t) - E\left[\sum_{j=1}^{N} 1(C_j R > e^t) \right]\right),$$

$$r(t) = e^{\alpha t} P(R > e^t) \qquad \text{and} \qquad \delta_n(t) = e^{\alpha t} E\left[\sum_{(\mathbf{i}|n)\in A_n} 1(\Pi_{\mathbf{i}|n} R > e^t) \right].$$

Recall that R and $(N_{\mathbf{i}|k}, C_{(\mathbf{i}|k,1)}, C_{(\mathbf{i}|k,2)}, \dots)$ are independent of $\mathscr{F}_k$, from where it follows that

$$E\left[1(R > e^{t-V_{\mathbf{i}|k}}) - \sum_{j=1}^{N_{\mathbf{i}|k}} 1(C_{(\mathbf{i}|k,j)} R > e^{t-V_{\mathbf{i}|k}}) \middle| \mathscr{F}_k \right] = e^{\alpha(V_{\mathbf{i}|k}-t)} g\left(t - V_{\mathbf{i}|k}\right).$$

Then, for any $t \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$r(t) = \sum_{k=0}^{n-1} E\left[\sum_{(\mathbf{i}|k)\in A_k} e^{\alpha V_{\mathbf{i}|k}} g(t - V_{\mathbf{i}|k}) \right] + \delta_n(t) = (g * \nu_{n-1})(t) + \delta_n(t).$$

Next, using the assumptions of the theorem, one can show that $\delta_n(t) \to 0$ as $n \to \infty$, and furthermore,

$$r(t) = g * \nu(t),$$

where $\nu(dt) = \sum_{k=0}^{\infty} \eta^{*k}(dt)$; see [27, 28] for more details. Now, the result would follow from the key renewal theorem for two-sided random walks if it were not for the fact that g is not necessarily directly Riemann integrable. To overcome this difficulty one can introduce a smoothing transform, similarly as it was done in [19], and apply the two-sided key renewal theorem [8] to the transformed equation to show that

$$e^{-t} \int_0^{e^t} v^{\alpha} P(R > v)dv \to H, \qquad t \to \infty.$$

Finally, by a version of the monotone density theorem (see Lemma 9.3 in [19]), one derives

$$P(R > t) \sim Ht^{-\alpha}, \qquad t \to \infty,$$

where

$$\begin{aligned} H &= \frac{1}{\mu} \int_{-\infty}^{\infty} g(t)\, dt \\ &= \frac{1}{\mu} \int_{0}^{\infty} v^{\alpha-1} \left(P(R > v) - E\left[\sum_{j=1}^{N} 1(C_j R > v) \right] \right) dv \end{aligned}$$

and μ was defined in Lemma 1.

4 Large Deviations Analysis

In this section we give the main technical result that allows the analysis of the solutions to recursions on weighted branching trees when the conditions for the Implicit Renewal Theorem do not apply, but either the sum of the weights, $\sum_{i=1}^{N} C_i$, or the innovation, Q, has a heavy-tailed distribution. The analysis in these cases is based on a uniform bound for the tail distribution of the sum of the weights on the nth generation of a weighted branching tree, which we formally define below.

Let $\{W_n : n \geq 0\}$ be the process constructed on $\mathscr{T}_{Q,C}$ via

$$W_0 = Q, \quad W_n = \sum_{\mathbf{i} \in A_n} Q_{\mathbf{i}} \Pi_{\mathbf{i}}, \qquad n \geq 1. \tag{8}$$

Since the tree structure repeats itself after the first generation, W_n satisfies

$$W_n \stackrel{\mathscr{D}}{=} \sum_{k=1}^{N} C_k W_{(n-1),k}, \tag{9}$$

where $\{W_{(n-1),k}\}$ is a sequence of iid random variables independent of $(N, C_1, C_2, \ldots)$ and having the same distribution as W_{n-1}.

We now proceed to compute explicit moment bounds for W_n. The next lemma is the key to this analysis; a generalization to real-valued random variables can be found in [28].

Lemma 2. *For any $k \in \mathbb{N} \cup \{\infty\}$ let $\{C_i\}_{i=1}^{k}$ be a sequence of nonnegative random variables and let $\{Y_i\}_{i=1}^{k}$ be a sequence of nonnegative iid random variables, independent of the $\{C_i\}$, having the same distribution as Y. For $\beta > 1$ set*

$p = \lceil \beta \rceil \in \{2, 3, 4, \ldots\}$, *and if* $k = \infty$ *assume that* $\sum_{i=1}^{\infty} C_i Y_i < \infty$ *a.s. Then,*

$$E\left[\left(\sum_{i=1}^{k} C_i Y_i\right)^{\beta} - \sum_{i=1}^{k} (C_i Y_i)^{\beta}\right] \leq \left(E\left[Y^{p-1}\right]\right)^{\beta/(p-1)} E\left[\left(\sum_{i=1}^{k} C_i\right)^{\beta}\right].$$

Remark 2. Note that the preceding lemma does not exclude the case when $E\left[\left(\sum_{i=1}^{k} C_i Y_i\right)^{\beta}\right] = \infty$ but $E\left[\left(\sum_{i=1}^{k} C_i Y_i\right)^{\beta} - \sum_{i=1}^{k} (C_i Y_i)^{\beta}\right] < \infty$.

We now give estimates for the β-moments of W_n for $\beta \in (0, 1]$ and $\beta > 1$ in Lemmas 3 and 4, respectively; their proofs can be found in [27]. Throughout the rest of the paper define $\rho_\beta = E\left[\sum_{i=1}^{N} C_i^{\beta}\right]$ for any $\beta > 0$, and $\rho \equiv \rho_1$.

Lemma 3. *For* $0 < \beta \leq 1$ *and all* $n \geq 0$,

$$E[W_n^{\beta}] \leq E[Q^{\beta}] \rho_\beta^n.$$

Lemma 4. *For* $\beta > 1$ *suppose* $E[Q^{\beta}] < \infty$, $E\left[\left(\sum_{i=1}^{N} C_i\right)^{\beta}\right] < \infty$, *and* $\rho \vee \rho_\beta < 1$. *Then, there exists a constant* $K_\beta < \infty$ *such that for all* $n \geq 0$,

$$E[W_n^{\beta}] \leq K_\beta (\rho \vee \rho_\beta)^n.$$

The main technical result of this section provides a uniform bound (uniform in n and x) for $P(W_n > x)$ under the assumption that either $P\left(\sum_{i=1}^{N} C_i > x\right) \in \mathscr{R}_{-\alpha}$ or $P(Q > x) \in \mathscr{R}_{-\alpha}$, where $\mathscr{R}_{-\alpha}$ is the family of regularly varying functions with index $-\alpha$. For completeness we give the definition below.

Definition 1. A function f is regularly varying at infinity with index λ, denoted $f \in \mathscr{R}_\lambda$, if $f(x) = x^{\lambda} L(x)$ for some slowly varying function L; and $L : [0, \infty) \to (0, \infty)$ is slowly varying if $\lim_{x \to \infty} L(tx)/L(x) = 1$ for any $t > 0$.

We now state the two main results of this section; their proofs are given in [36].

Proposition 1. *Let* $Z_N = \sum_{i=1}^{N} C_i$ *and suppose* $P(Z_N > x) \in \mathscr{R}_{-\alpha}$ *with* $\alpha > 1$. *Assume further that* $E[Q^{\alpha+\epsilon}] < \infty$ *and* $\rho_{\alpha+\epsilon} < \infty$ *for some* $\epsilon > 0$. *Fix* $\rho \vee \rho_\alpha < \eta < 1$. *Then, there exists a finite constant* $K = K(\eta, \epsilon) > 0$ *such that for all* $n \geq 1$ *and all* $x \geq 1$,

$$P(W_n > x) \leq K \eta^n P(Z_N > x). \tag{10}$$

Remark 3. Note that we can easily obtain a weaker uniform bound by applying the moment estimate on $E[W_n^{\beta}]$ from Lemma 4, i.e., $P(W_n > x) \leq E[W_n^{\beta}] x^{-\beta} \leq$

$K_\beta(\rho \vee \rho_\beta)^n x^{-\beta}$ for some $0 < \beta < \alpha$, so the tradeoff in (10) is a slightly larger geometric term for a lighter tail distribution. However, the assertion in (10) is considerably more difficult to prove.

The corresponding result for the case when $P(Q > x) \in \mathscr{R}_{-\alpha}$ is given below.

Proposition 2. *Suppose $P(Q > x) \in \mathscr{R}_{-\alpha}$, with $\alpha > 1$, $E[Z_N^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$, and let $\rho \vee \rho_\alpha < \eta < 1$. Then, there exists a finite constant $K = K(\eta, \epsilon) > 0$ such that for all $n \geq 1$ and all $x \geq 1$,*

$$P(W_n > x) \leq K\eta^n P(Q > x).$$

5 The Linear Recursion $R \stackrel{\mathscr{D}}{=} \sum_{i=1}^N C_i R_i + Q$

This section focuses on the analysis of the linear nonhomogeneous equation (1), and it is further divided into the three possible sources of power-law tails of the solution R. Before we proceed with the analysis, we give below an explicit construction of R on the weighted branching tree $\mathscr{T}_{Q,C}$ and show that, under appropriate conditions, this solution is the unique limit under iterations of (1). Recall that throughout the paper we assume that the vector $(Q, N, C_1, C_2, \ldots)$ is nonnegative.

Define the process $\{R^{(n)}\}_{n\geq 0}$ according to

$$R^{(n)} = \sum_{k=0}^{n} W_k, \qquad n \geq 0, \tag{11}$$

that is, $R^{(n)}$ is the sum of the weights of all the nodes on the tree up to the nth generation. It is not hard to see that $R^{(n)}$ satisfies the recursion

$$R^{(n)} = \sum_{j=1}^{N_\emptyset} C_{(\emptyset,j)} R_j^{(n-1)} + Q_\emptyset = \sum_{j=1}^{N} C_j R_j^{(n-1)} + Q, \qquad n \geq 1, \tag{12}$$

where $\{R_j^{(n-1)}\}$ are independent copies of $R^{(n-1)}$ corresponding to the tree starting with individual j in the first generation and ending on the nth generation; note that $R_j^{(0)} = Q_j$.

Next, define the random variable R according to

$$R \triangleq \lim_{n\to\infty} R^{(n)} = \sum_{k=0}^{\infty} W_k, \tag{13}$$

where the limit is properly defined by (11) and monotonicity. Hence, it is easy to verify, by applying monotone convergence in (12), that R must solve

$$R = \sum_{j=1}^{N_\emptyset} C_{(\emptyset,j)} R_j^{(\infty)} + Q_\emptyset = \sum_{j=1}^{N} C_j R_j^{(\infty)} + Q,$$

where $\{R_j^{(\infty)}\}_{j\in\mathbb{N}}$ are iid, have the same distribution as R, and are independent of $(Q, N, C_1, C_2, \dots)$.

The derivation provided above implies in particular the existence of a solution in distribution to (1). Moreover, under additional technical conditions, R is the unique solution under iterations as it will be defined and shown in the following section. The constructed R, as defined in (13), is the main object of study in the remainder of this section. Note that, in view of the very recent work in [3], (1) may have other stable-law solutions that are not considered here. The lemma below gives sufficient conditions for the finiteness of moments of R, see [27] for a proof.

Lemma 5. *Assume that $E[Q^\beta] < \infty$ for some $\beta > 0$. In addition, suppose that either (i) $\rho_\beta < 1$ if $0 < \beta < 1$, or (ii) $(\rho \vee \rho_\beta) < 1$ and $E\left[\left(\sum_{i=1}^N C_i\right)^\beta\right] < \infty$ if $\beta \geq 1$. Then, $E[R^\gamma] < \infty$ for all $0 < \gamma \leq \beta$, and in particular, $R < \infty$ a.s. Moreover, if $\beta \geq 1$, $R^{(n)} \stackrel{L_\beta}{\to} R$, where L_β denotes β-norm convergence.*

Remark 4. It is interesting to observe that for $\beta > 1$ the conditions $\rho_\beta < 1$ and $E\left[\left(\sum_{i=1}^N C_i\right)^\beta\right] < \infty$ are consistent with Theorem 3.1 in [2], Proposition 4 in [24] and Theorem 2.1 in [34], which give the conditions for the finiteness of the β-moment of the solution to the related critical ($\rho_1 = 1$) homogeneous ($Q \equiv 0$) equation.

Next, we show that under some technical conditions, the iteration of recursion (1) results in a process that converges in distribution to R for any initial condition R_0^*. To this end, consider a weighted branching tree $\mathcal{T}_{Q,C}$, as defined in Sect. 2. Now, define

$$R_n^* \triangleq R^{(n-1)} + W_n(R_0^*), \qquad n \geq 1,$$

where $R^{(n-1)}$ is given by (11),

$$W_n(R_0^*) = \sum_{\mathbf{i}\in A_n} R_{0,\mathbf{i}}^* \Pi_{\mathbf{i}}, \tag{14}$$

and $\{R_{0,\mathbf{i}}^*\}_{\mathbf{i}\in U}$ are iid copies of an initial value R_0^*, independent of the entire weighted tree $\mathcal{T}_{Q,C}$. It follows from (12) and (14) that, for $n \geq 0$,

$$
\begin{aligned}
R_{n+1}^* &= \sum_{j=1}^{N} C_j R_j^{(n-1)} + Q + W_{n+1}(R_0^*) \\
&= \sum_{j=1}^{N} C_j \left(R_j^{(n-1)} + \sum_{\mathbf{i} \in A_{n,j}} R_{0,\mathbf{i}}^* \prod_{k=2}^{n} C_{(j,\ldots,i_k)} \right) + Q, \qquad (15)
\end{aligned}
$$

where $\{R_j^{(n-1)}\}$ are independent copies of $R^{(n-1)}$ corresponding to the tree starting with individual j in the first generation and ending on the nth generation, and $A_{n,j}$ is the set of all nodes in the $(n+1)$th generation that are descendants of individual j in the first generation. It follows that

$$
R_{n+1}^* = \sum_{j=1}^{N} C_j R_{n,j}^* + Q,
$$

where $\{R_{n,j}^*\}$ are the expressions inside the parenthesis in (15). Clearly, $\{R_{n,j}^*\}$ are iid copies of R_n^*, thus we have shown that R_n^* is equal in distribution to the process derived by iterating (1) with an initial condition R_0^*. The following lemma shows that $R_n^* \Rightarrow R$ for any initial condition R_0^* satisfying a moment assumption, where $\Rightarrow$ denotes convergence in distribution; see [27] for a proof.

Lemma 6. *For any initial condition $R_0^* \geq 0$, if $E[Q^\beta], E[(R_0^*)^\beta] < \infty$ and $\rho_\beta = E\left[\sum_{i=1}^N C_i^\beta\right] < 1$ for some $0 < \beta \leq 1$, then*

$$
R_n^* \Rightarrow R,
$$

with $E[R^\beta] < \infty$. Furthermore, under these assumptions, the distribution of R is the unique solution with finite β-moment to recursion (1).

Remark 5. (i) Note that when $E[N] < 1$ the branching tree is a.s. finite and no conditions on the $\{C_i\}$ are necessary for $R < \infty$ a.s. This corresponds to the second condition in Theorem 1 of [12]. (ii) In view of the same theorem from [12], one could possibly establish the convergence of $R_n^* \Rightarrow R < \infty$ under milder conditions. However, since in this paper we only study the power tails of R, the assumptions of Lemma 6 are not restrictive. (iii) Note that if $E\left[\sum_{i=1}^N C_i^\alpha\right] = 1$ with $\alpha \in (0, 1]$, then there might not be a $0 < \beta < \alpha$ for which $E\left[\sum_{i=1}^N C_i^\beta\right] < 1$, e.g., the case of deterministic C_i's that was studied in [42].

5.1 The Case When the Weights $\{C_i\}$ Dominate

In this section we characterize the tail behavior of the distribution of the solution R to the nonhomogeneous equation (1), as defined by (13), when its power-law tail behavior is due to the multiplicative effect of the weights $\{C_i\}$. The main result is given in the following theorem, which is an application of Theorem 1; see the proof of Theorem 4.1 in [27] and the remark at the end of this subsection. A generalization to real-valued weights can be found in Theorem 4.6 in [28].

Theorem 2. *Let $(Q, N, C_1, C_2, \ldots)$ be a nonnegative random vector, with $N \in \mathbb{N} \cup \{\infty\}$, $P(Q > 0) > 0$, and let R be the solution to (1) given by (13). Suppose that there exists $j \geq 1$ with $P(N \geq j, C_j > 0) > 0$ such that the measure $P\left(\log C_j \in du, C_j > 0, \quad N \geq j\right)$ is nonarithmetic, and that for some $\alpha > 0$, $0 < E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right] < \infty$, $E\left[\sum_{i=1}^N C_i^\alpha\right] = 1$, and $E[Q^\alpha] < \infty$. In addition, assume*

1. $E\left[\sum_{i=1}^N C_i\right] < 1$ *and* $E\left[\left(\sum_{i=1}^N C_i\right)^\alpha\right] < \infty$, *if* $\alpha > 1$*; or,*
2. $E\left[\left(\sum_{i=1}^N C_i^{\alpha/(1+\epsilon)}\right)^{1+\epsilon}\right] < \infty$ *for some* $0 < \epsilon < 1$, *if* $0 < \alpha \leq 1$.

Then,

$$P(R > t) \sim H t^{-\alpha}, \qquad t \to \infty,$$

where $0 \leq H < \infty$ *is given by*

$$H = \frac{1}{E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right]} \int_0^\infty v^{\alpha-1} \left(P(R > v) - E\left[\sum_{i=1}^N 1(C_i R > v)\right]\right) dv$$

$$= \frac{E\left[\left(\sum_{i=1}^N C_i R_i + Q\right)^\alpha - \sum_{i=1}^N (C_i R_i)^\alpha\right]}{\alpha E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right]}.$$

Remark 6. (i) The nonhomogeneous equation has been previously studied for the special case when Q and the $\{C_i\}$ are deterministic constants. In particular, Theorem 5 of [42] analyzes the solutions to (1) when Q and the $\{C_i\}$ are nonnegative deterministic constants, which, when $\sum_{i=1}^N C_i^\alpha = 1$, $\alpha > 0$, implies that $C_i \leq 1$ for all i and $\sum_i C_i^\alpha \log C_i \leq 0$, falling outside of the scope of this theorem. As previously mentioned, the additional stable-law solutions found recently in [3] for Q and $\{C_i\}$ random also fall outside of the scope of this theorem and do not satisfy the conditions of Lemma 6. (ii) When $\alpha > 1$, the condition $E\left[\left(\sum_{i=1}^N C_i\right)^\alpha\right] < \infty$ is needed to ensure that the tail of R is not dominated by $\sum_{i=1}^N C_i$. In particular, if the $\{C_i\}$ are iid and independent of N, the condition reduces to $E[N^\alpha] < \infty$ since $E[C^\alpha] < \infty$ is implied by the other conditions; see

Theorems 4.2 and 5.4 in [26]. Furthermore, when $0 < \alpha \leq 1$ the condition $E\left[\left(\sum_{i=1}^N C_i\right)^\alpha\right] < \infty$ is redundant since $E\left[\left(\sum_{i=1}^N C_i\right)^\alpha\right] \leq E\left[\sum_{i=1}^N C_i^\alpha\right] = 1$, and the additional condition $E\left[\left(\sum_{i=1}^N C_i^{\alpha/(1+\epsilon)}\right)^{1+\epsilon}\right] < \infty$ is needed. When the $\{C_i\}$ are iid and independent of N, the latter condition reduces to $E[N^{1+\epsilon}] < \infty$ (given the other assumptions), which is consistent with Theorem 4.2 in [26]. (iii) Note that the second expression for H is more suitable for actually computing it, especially in the case of α being an integer, as will be stated in the forthcoming Corollary 1, after which we will also explain how Lemma 2 can be used to derive an explicit upper bound on H when $\alpha > 1$ is not an integer. Regarding the lower bound, the elementary inequality $\left(\sum_{i=1}^k x_i\right)^\alpha \geq \sum_{i=1}^k x_i^\alpha$ for $\alpha \geq 1$ and $x_i \geq 0$, (see Exercise 4.2.1, p. 102 in [14]), yields

$$H \geq \frac{E\left[Q^\alpha\right]}{\alpha E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right]} > 0.$$

Similarly, for $0 < \alpha < 1$, using the corresponding inequality $\left(\sum_{i=1}^k x_i\right)^\alpha \leq \sum_{i=1}^k x_i^\alpha$ for $0 < \alpha \leq 1$, $x_i \geq 0$, we obtain $H \leq E\left[Q^\alpha\right] / \left(\alpha E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right]\right)$. (iv) Let us also observe that the solution R, given by (13), to equation (1) may be a constant (non power law) $R = r > 0$ when $P(r = Q + r\sum_{i=1}^N C_i) = 1$. However, similarly as in remark (i), such a solution is excluded from the theorem since $P(r = Q + r\sum_{i=1}^N C_i) = 1$ implies $E[\sum_i C_i^\alpha \log C_i] \leq 0, \alpha > 0$. (iv) The strict positivity of the constant H for the real-valued case has very recently been established in [7], and a version where the weights $\{C_i\}$ are positive matrices and Q is a positive vector can be found in [35].

As indicated earlier, when $\alpha \geq 1$ is an integer, we can obtain the following explicit expression for H.

Corollary 1. *For integer $\alpha \geq 1$, and under the same assumptions of Theorem 2, the constant H can be explicitly computed as a function of $E[R^k]$, $0 \leq k \leq \alpha - 1$, and the mixed moments of order up to α of $(Q, N, C_1, C_2, \dots)$ according to the following expression*

$$H = \frac{1}{\alpha E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right]} E\left[Q^\alpha + \sum_{\mathbf{j} \in B_{\alpha-1}(N)} \binom{\alpha}{j_0, j_1, j_2, \dots} Q^{j_0} \prod_{i=1}^N C_i^{j_i} E[R^{j_i}]\right],$$

where $\mathbf{j} = (j_0, j_1, j_2, \dots)$, $B_p(n) = \{(j_0, j_1, j_2, \dots) \in \mathbb{N}_+^{n+1} : \sum_{k=0}^n j_k = p, 0 \leq j_i < p\}$. In particular, for $\alpha = 1$,

$$H = \frac{E[Q]}{E\left[\sum_{i=1}^{N} C_i \log C_i\right]},$$

and for $\alpha = 2$,

$$H = \frac{E[Q^2] + 2E[R]E\left[Q\sum_{i=1}^{N} C_i\right] + 2(E[R])^2 E\left[\sum_{i=1}^{N}\sum_{j=i+1}^{N} C_i C_j\right]}{2E\left[\sum_{i=1}^{N} C_i^2 \log C_i\right]},$$

$$E[R] = \frac{E[Q]}{1 - E\left[\sum_{i=1}^{N} C_i\right]}.$$

Proof. Using the multinomial expansion we obtain for any $k \in \mathbb{N}$

$$E\left[\left(\sum_{i=1}^{N} C_i R_i + Q\right)^{\alpha} - \sum_{i=1}^{N} (C_i R_i)^{\alpha}\right]$$
$$= E\left[Q^{\alpha} + \sum_{\mathbf{j} \in B_{\alpha-1}(N)} \binom{\alpha}{j_0, j_1, j_2, \ldots} Q^{j_0} \prod_{i=1}^{N} (C_i R_i)^{j_i}\right].$$

Next, condition on $\mathscr{F} = \sigma(Q, N, C_1, C_2, \ldots)$ to obtain

$$E\left[\sum_{\mathbf{j} \in B_{\alpha-1}(N)} \binom{\alpha}{j_0, j_1, j_2, \ldots} Q^{j_0} \prod_{i=1}^{N} (C_i R_i)^{j_i}\right]$$
$$= E\left[\sum_{\mathbf{j} \in B_{\alpha-1}(N)} \binom{\alpha}{j_0, j_1, j_2, \ldots} Q^{j_0} \prod_{i=1}^{N} C_i^{j_i} E\left[R_i^{j_i} \middle| \mathscr{F}\right]\right]$$
$$= E\left[\sum_{\mathbf{j} \in B_{\alpha-1}(N)} \binom{\alpha}{j_0, j_1, j_2, \ldots} Q^{j_0} \prod_{i=1}^{N} C_i^{j_i} E[R^{j_i}]\right]. \qquad \square$$

For the case when $\alpha > 1$ is not an integer, the same arguments used in the proof of Lemma 2 (see Lemma 4.1 in [27]) lead to

$$E\left[\left(\sum_{i=1}^{N} C_i R_i + Q\right)^{\alpha} - \sum_{i=1}^{N} (C_i R_i)^{\alpha}\right]$$
$$\leq E\left[Q^{\alpha} + \left(\sum_{\mathbf{j} \in B_p(N)} \binom{p}{j_0, j_1, j_2, \ldots} Q^{j_0} \prod_{i=1}^{N} (C_i R_i)^{j_i}\right)^{\alpha/p}\right]$$

where $\mathbf{j}$ and $B_p(n)$ are defined as in Corollary 1 and $p = \lceil \alpha \rceil$. Then condition on $\mathscr{F} = \sigma(Q, N, C_1, C_2, \dots)$ and use Jensen's inequality to obtain

$$H \leq \frac{1}{\alpha E\left[\sum_{i=1}^{N} C_i^\alpha \log C_i\right]} E\left[Q^\alpha + \left(\sum_{\mathbf{j} \in B_p(N)} \binom{p}{j_0, j_1, j_2, \dots} Q^{j_0} \prod_{i=1}^{N} C_i^{j_i} E[R^{j_i}] \right)^{\alpha/p} \right].$$

Note that the moments $E[R^{j_i}]$, $j_i \leq \lceil \alpha \rceil - 1$, can be computed recursively as in Corollary 1. For the case $N \equiv 1$ the recent work in [16] provides a computable expression for H.

Two results that facilitate the verification of the conditions in Theorem 1 are given below in Lemmas 7 and 8 (Lemmas 4.6 and 4.7 in [27]). The first of these results transforms the integral conditions in Theorem 1 into an expression that can be verified by using the specific recursion being analyzed. These lemmas can be directly applied to analyze other max-plus recursions as well, such as those mentioned in (5). We illustrate the use of these results by giving a heuristic proof of Theorem 2 at the end of this section.

Lemma 7. *Suppose* $(N, C_1, C_2, \dots)$ *is a nonnegative random vector, with* $N \in \mathbb{N} \cup \{\infty\}$ *and let* $\{R_i\}_{i \in \mathbb{N}}$ *be a sequence of iid nonnegative random variables independent of* $(N, C_1, C_2, \dots)$ *having the same distribution as* R. *For* $\alpha > 0$, *suppose that* $\sum_{i=1}^{N} (C_i R_i)^\alpha < \infty$ *a.s. and* $E[R^\beta] < \infty$ *for any* $0 < \beta < \alpha$. *Furthermore, assume that* $E\left[\left(\sum_{i=1}^{N} C_i^{\alpha/(1+\epsilon)}\right)^{1+\epsilon}\right] < \infty$ *for some* $0 < \epsilon < 1$. *Then,*

$$\begin{aligned} 0 &\leq \int_0^\infty \left(E\left[\sum_{i=1}^{N} 1(C_i R_i > t) \right] - P\left(\max_{1 \leq i \leq N} C_i R_i > t \right) \right) t^{\alpha-1}\, dt \\ &= \frac{1}{\alpha} E\left[\sum_{i=1}^{N} (C_i R_i)^\alpha - \left(\max_{1 \leq i \leq N} C_i R_i \right)^\alpha \right] < \infty. \end{aligned}$$

Lemma 8. *Let* $(Q, N, C_1, C_2, \dots)$ *be a nonnegative vector with* $N \in \mathbb{N} \cup \{\infty\}$ *and let* $\{R_i\}$ *be a sequence of iid random variables, independent of* $(Q, N, C_1, C_2, \dots)$. *Suppose that for some* $\alpha > 1$ *we have* $E[Q^\alpha] < \infty$, $E\left[\left(\sum_{i=1}^{N} C_i\right)^\alpha\right] < \infty$, $E[R^\beta] < \infty$ *for any* $0 < \beta < \alpha$, *and* $\sum_{i=1}^{N} C_i R_i < \infty$ *a.s. Then*

$$E\left[\left(\sum_{i=1}^{N} C_i R_i + Q \right)^\alpha - \sum_{i=1}^{N} (C_i R_i)^\alpha \right] < \infty.$$

We now give some of the key high-level heuristics behind the proof of Theorem 2. The first technical condition to check is that $E[R^\beta] < \infty$ for all $0 < \beta < \alpha$, which follows from the assumptions in the theorem and our moment estimates from Lemmas 3 and 4. The majority of the work in the proof goes into verifying the absolute integrability condition (7). In order to do this, we first observe that

$$P(R > t) = P\left(\sum_{i=1}^{N} C_i R_i + Q > t\right) \approx P\left(\sum_{i=1}^{N} C_i R_i > t\right)$$

for large t since $E[Q^\alpha] < \infty$ and the weighted sum $\sum C_i R_i$ is expected to have infinite α-moment. Then, under the condition $E\left[\left(\sum_{i=1}^{N} C_i\right)^\alpha\right] < \infty$ ($\alpha > 1$), we expect the weighted sum to behave as the maximum according to the well-known heavy-tailed one-big-jump principle, i.e.,

$$P\left(\sum_{i=1}^{N} C_i R_i > t\right) \approx P\left(\max_{1\leq i\leq N} C_i R_i > t\right).$$

The last observation is that

$$P\left(\max_{1\leq i\leq N} C_i R_i > t\right) \approx E\left[\sum_{i=1}^{N} 1(C_i R_i > t)\right],$$

for large t, which is made rigorous in Lemma 7. Hence, the proof is enabled by adding and subtracting the term $P\left(\max_{1\leq i\leq N} C_i R_i > t\right)$ inside the integrand in (7). The rigorous justification of these ideas is quite involved, in part because one has to understand the second order properties of the preceding approximations, i.e., the error term; we refer the reader to [26–28] for the details.

5.2 The Case When the Sum of the Weights $\sum_{i=1}^{N} C_i$ Dominates

In this section we focus on the case where $P(Z_N > x) \in \mathscr{R}_{-\alpha}$ for some $\alpha > 1$, and $\rho \vee \rho_\alpha < 1$. The approach we follow is to first describe the asymptotic behavior of finitely many iterations of (1), those given by $R^{(n)}$, and then use the uniform bound given in Proposition 1 to control the difference $|R - R^{(n)}|$. The first lemma given below is based on the use of some asymptotic limits for randomly stopped and randomly weighted sums recently developed in [37].

Lemma 9. *Let $Z_N = \sum_{i=1}^{N} C_i$ and suppose $P(Z_N > x) \in \mathscr{R}_{-\alpha}$ with $\alpha > 1$, $E[Q^{\alpha+\epsilon}] < \infty$, $\rho_{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$, and $\rho < 1$. Then, for any fixed $n \in \{1, 2, 3, \ldots\}$,*

$$P(R^{(n)} > x) \sim \frac{(E[Q])^\alpha}{(1-\rho)^\alpha} \sum_{k=0}^{n-1} \rho_\alpha^k (1-\rho^{n-k})^\alpha \, P(Z_N > x) \tag{16}$$

as $x \to \infty$, where $R^{(n)}$ was defined in (11).

Remark 7. In terms of the ranking example given in the introduction, Q usually refers to a nonnegative personalization parameter that determines what page to go to in case the algorithm reaches a page with no outbound links (see [46] for more details).

From Lemma 9 one can already guess that, provided $\rho \vee \rho_\alpha < 1$, the tail behavior of R will be

$$P(R > x) \sim \frac{(E[Q])^\alpha}{(1-\rho)^\alpha} \sum_{k=0}^{\infty} \rho_\alpha^k P(Z_N > x)$$

as $x \to \infty$, assuming that the exchange of limits is justified. As mentioned above, this exchange represents the main technical difficulty. This result was proved in [26] for the case where $Q, N, \{C_i\}$ are all independent and the $\{C_i\}$ are iid using sample-path arguments, and in [46] for the case where (Q, N) is independent of $\{C_i\}$ and the $\{C_i\}$ are iid, using transform methods and Tauberian theorems. A version of the results presented here that can be applied when Q is a real-valued random variable can be found in [36].

The uniform bound given by Proposition 1 is the key to establishing that $|R - R^{(n)}|$ goes to zero geometrically fast, which is more precisely stated in the following lemma.

Lemma 10. *Let $Z_N = \sum_{i=1}^N C_i$ and suppose $P(Z_N > x) \in \mathscr{R}_{-\alpha}$ with $\alpha > 1$, $E[Q^{\alpha+\epsilon}] < \infty$ and $\rho_{\alpha+\epsilon} < \infty$, for some $\epsilon > 0$. Assume $\rho \vee \rho_\alpha < 1$, then, for any fixed $0 < \delta < 1$, $n_0 \in \{1, 2, \ldots\}$ and $\rho \vee \rho_\alpha < \eta < 1$, there exists a finite constant $K > 0$ that does not depend on δ or n_0 such that*

$$\lim_{x\to\infty} \frac{P\left(|R - R^{(n_0)}| > \delta x\right)}{P(Z_N > x)} \le \frac{K\eta^{n_0+1}}{\delta^{\alpha+1} n_0}.$$

Combining Lemmas 9 and 10 one can obtain the following result. The proofs of all the results in this subsection can be found in [36].

Theorem 3. *Let $Z_N = \sum_{i=1}^N C_i$ and suppose $P(Z_N > x) \in \mathscr{R}_{-\alpha}$ with $\alpha > 1$, $E[Q^{\alpha+\epsilon}] < \infty$ and $\rho_{\alpha+\epsilon} < \infty$, for some $\epsilon > 0$. Assume $\rho \vee \rho_\alpha < 1$, then,*

$$P(R > x) \sim \frac{(E[Q])^\alpha}{(1-\rho)^\alpha (1-\rho_\alpha)} P(Z_N > x)$$

as $x \to \infty$, where R was defined in (13).

Remark 8. (i) For the case where the $\{C_i\}$ are iid and independent of N, and $P(N > x) \in \mathscr{R}_{-\alpha}$, Lemma 3.7(2) in [29] gives

$$P(Z_N > x) \sim (E[C_1])^\alpha P(N > x) \qquad \text{as } x \to \infty.$$

(ii) Given the previous remark, it follows that Theorem 3 generalizes both Theorem 5.1 in [26] (for $Q, N, \{C_i\}$ all independent and $\{C_i\}$ iid) and the corresponding result from Sect. 3.4 in [46] (for (Q, N) independent of $\{C_i\}$, $\{C_i\}$ iid, $E[Q] < 1$ and $E[C] = (1 - E[Q])/E[N]$). (iii) In view of Lemma 9, the theorem shows that the limits $\lim_{x\to\infty} \lim_{n\to\infty} P(R^{(n)} > x)/P(Z_N > x)$ are interchangeable.

5.3 The Case When *Q* Dominates

This section of the paper treats the case where the heavy-tailed behavior of R arises from the $\{Q_{\mathbf{i}}\}$, known in the autoregressive processes literature as innovations. This setting is well known in the special case $N \equiv 1$, since then the linear fixed-point equation (1) reduces to

$$R \stackrel{\mathscr{D}}{=} CR + Q,$$

where (C, Q) are generally dependent. This fixed-point equation is the one satisfied by the steady state of the autoregressive process of order one with random coefficients, RCA(1) (see [12, 19, 21, 31]).

That the innovations $\{Q_{\mathbf{i}}\}$ can give rise to heavy tails when the α mentioned above does not exist is also well known, see, e.g., [21, 32]; the main theorem of this subsection provides an alternative derivation of the forward implication in Theorem 1 from [21] (see also Proposition 2.4 in [32]) for $Q, N \geq 0$.

The results presented here are very similar to those in Sect. 5.2, and so are their proofs, which can also be found in [36] and include the case where Q is real-valued.

Lemma 11. *Suppose $P(Q > x) \in \mathscr{R}_{-\alpha}$, with $\alpha > 1$, and $E[Z_N^{\alpha+\epsilon}] < \infty$, for some $\epsilon > 0$. Then, for any fixed $n \in \{1, 2, 3, \ldots\}$,*

$$P(R^{(n)} > x) \sim \sum_{k=0}^{n} \rho_\alpha^k P(Q > x)$$

as $x \to \infty$, where $R^{(n)}$ was defined in (13).

As for the case when $Z_N = \sum_{i=1}^{N} C_i$ dominates the asymptotic behavior of R, we can expect that,

$$P(R > x) \sim (1 - \rho_\alpha)^{-1} P(Q > x),$$

and the technical difficulty is in justifying the exchange of the limits. The same techniques used in Sect. 5.2 can be used in this case as well. We point out that even though the condition $\rho < 1$ is not necessary for the proportionality constant in Lemma 11 to be finite, it is required for the finiteness of $E[R]$.

The corresponding version of Lemma 10 is given below.

Lemma 12. *Let $Z_N = \sum_{i=1}^N C_i$ and suppose $P(Q > x) \in \mathscr{R}_{-\alpha}$ with $\alpha > 1$, $E[Z_N^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$, and $E[Q^\beta] < \infty$ for all $0 < \beta < \alpha$. Assume $\rho \vee \rho_\alpha < 1$, then, for any fixed $0 < \delta < 1$, $n_0 \in \{1, 2, \ldots\}$ and $\rho \vee \rho_\alpha < \eta < 1$, there exists a finite constant $K > 0$ that does not depend on δ or n_0 such that*

$$\lim_{x\to\infty} \frac{P\left(|R - R^{(n_0)}| > \delta x\right)}{P(Q > x)} \le \frac{K\eta^{n_0+1}}{\delta^{\alpha+1} n_0}.$$

The main theorem of this section is given below.

Theorem 4. *Suppose $P(Q > x) \in \mathscr{R}_{-\alpha}$, with $\alpha > 1$, $E[Q^\beta] < \infty$ for all $0 < \beta < \alpha$. Assume $\rho \vee \rho_\alpha < 1$, and $E[Z_N^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$. Then,*

$$P(R > x) \sim (1 - \rho_\alpha)^{-1} P(Q > x)$$

as $x \to \infty$, where R was defined in (13).

Remark 9. (i) This result generalizes Theorem 1 in [21] for the case $N \equiv 1$ (the forward implication, $\alpha > 1$) to the weighted branching tree when $Q \ge 0$. It also generalizes the results in [26, 46] in the same way as Theorem 3 does for the case where Z_N dominates. (ii) It is also worth pointing out that the same sample-path techniques used here can be used to study the intermediate case where $P(Q > x) \sim KP(Z_N > x)$ for some constant $K > 0$, which is also analyzed in [46] under stronger conditions than those stated above.

6 Other Recursions

In this section we show how our techniques can be applied to study other recursions on trees, e.g., those stated in (5). In particular, we start with the following non-linear equation

$$R \stackrel{\mathscr{D}}{=} \left(\bigvee_{i=1}^{N} C_i R_i \right) \vee Q, \tag{17}$$

where $(Q, N, C_1, C_2, \ldots)$ is a nonnegative random vector with $N \in \mathbb{N} \cup \{\infty\}$, $P(Q > 0) > 0$ and $\{R_i\}_{i\in\mathbb{N}}$ is a sequence of iid random variables that have the same distribution as R and is independent of $(Q, N, C_1, C_2, \ldots)$. Note that in the case of page ranking applications, where the $\{R_i\}$ represent the ranks of the neighboring pages, the potential ranking algorithm defined by the preceding recursion, determines the rank of a page as a weighted version of the most highly ranked neighboring page. In other words, the highest ranked reference has the dominant impact. Similarly to the homogeneous linear case, this recursion was previously studied in [5] under the assumption that $Q \equiv 0$, $N = \infty$, and the $\{C_i\}$ are real-valued deterministic constants. The more closely related case of $Q \equiv 0$ and $\{C_i\} \geq 0$ being random was studied earlier in [25].

Using standard arguments, we start by constructing a solution to (17) on a tree and then we show that this solution is finite a.s. and unique under iterations (under some moment conditions), similarly to what was done for the nonhomogeneous linear recursion in Sect. 5. Our main result of this section is stated in Theorem 5.

Following the same notation as in Sect. 5, define the process

$$V_n = \bigvee_{\mathbf{i} \in A_n} Q_{\mathbf{i}} \Pi_{\mathbf{i}}, \qquad n \geq 0, \tag{18}$$

on the weighted branching tree $\mathscr{T}_{Q,C}$, as constructed in Sect. 2. Recall that the convention is that $(Q, N, C_1, C_2, \ldots) = (Q_\emptyset, N_\emptyset, C_{(\emptyset,1)}, C_{(\emptyset,2)}, \ldots)$ denotes the random vector corresponding to the root node.

With a possible abuse of notation relative to Sect. 5, define the process $\{R^{(n)}\}_{n\geq 0}$ according to

$$R^{(n)} = \bigvee_{k=0}^{n} V_k, \qquad n \geq 0.$$

Just as with the linear recursion from Sect. 5, it is not hard to see that $R^{(n)}$ satisfies the recursion

$$R^{(n)} = \left(\bigvee_{j=1}^{N_\emptyset} C_{(\emptyset,j)} R_j^{(n-1)} \right) \vee Q_\emptyset = \left(\bigvee_{j=1}^{N} C_j R_j^{(n-1)} \right) \vee Q, \tag{19}$$

where $\{R_j^{(n-1)}\}$ are independent copies of $R^{(n-1)}$ corresponding to the subtree starting with individual j in the first generation and ending on the nth generation. One can also verify that

$$V_n = \bigvee_{k=1}^{N_\emptyset} C_{(\emptyset,k)} \bigvee_{(k,\ldots,i_n)\in A_n} Q_{(k,\ldots,i_n)} \prod_{j=2}^{n} C_{(k,\ldots,i_j)} \stackrel{\mathscr{D}}{=} \bigvee_{k=1}^{N} C_k V_{(n-1),k},$$

where $\{V_{(n-1),k}\}$ is a sequence of iid random variables independent of $(N, C_1, C_2, \ldots)$ and having the same distribution as V_{n-1}.

We now define the random variable R according to

$$R \triangleq \lim_{n\to\infty} R^{(n)} = \bigvee_{k=0}^{\infty} V_k. \tag{20}$$

Note that $R^{(n)}$ is monotone increasing sample-pathwise, so R is well defined. Also, by monotonicity of $R^{(n)}$, (19) and monotone convergence, we obtain that R solves

$$R = \left(\bigvee_{j=1}^{N_\emptyset} C_{(\emptyset,j)} R_j^{(\infty)} \right) \vee Q_\emptyset = \left(\bigvee_{j=1}^{N} C_j R_j^{(\infty)} \right) \vee Q,$$

where $\{R_j^{(\infty)}\}_{j\in\mathbb{N}}$ are iid copies of R, independent of $(Q, N, C_1, C_2, \dots)$. Clearly this implies that R, as defined by (20), is a solution in distribution to (17). However, this solution might be ∞. Now, we establish the finiteness of moments of R, and in particular that $R < \infty$ a.s., in the following lemma.

Lemma 13. *Assume that* $\rho_\beta = E\left[\sum_{i=1}^{N} C_i^\beta\right] < 1$ *and* $E[Q^\beta] < \infty$ *for some* $\beta > 0$. *Then,* $E[R^\gamma] < \infty$ *for all* $0 < \gamma \le \beta$, *and in particular,* $R < \infty$ *a.s. Moreover, if* $\beta \ge 1$, $R^{(n)} \stackrel{L_\beta}{\to} R$, *where* L_β *stands for convergence in* $(E|\cdot|^\beta)^{1/\beta}$ *norm.*

Just as with the linear recursion from Sect. 5, we can define the process $\{R_n^*\}$ as

$$R_n^* \triangleq R^{(n-1)} \vee V_n(R_0^*), \qquad n \ge 1,$$

where

$$V_n(R_0^*) = \bigvee_{\mathbf{i}\in A_n} R_{0,\mathbf{i}}^* \Pi_{\mathbf{i}}, \tag{21}$$

and $\{R_{0,\mathbf{i}}^*\}_{\mathbf{i}\in U}$ are iid copies of an initial value R_0^*, independent of the entire weighted tree $\mathscr{T}_{Q,C}$. It follows from (19) and (21) that

$$R_{n+1}^* = \bigvee_{j=1}^{N} C_j \left(R_j^{(n-1)} \vee \bigvee_{\mathbf{i}\in A_{n,j}} R_{0,\mathbf{i}}^* \prod_{k=2}^{n} C_{(j,\dots,i_k)} \right) \vee Q = \bigvee_{j=1}^{N} C_j R_{n,j}^* \vee Q,$$

where $\{R_j^{(n-1)}\}$ are independent copies of $R^{(n-1)}$ corresponding to the subtree starting with individual j in the first generation and ending on the nth generation, and $A_{n,j}$ is the set of all nodes in the $(n+1)$th generation that are descendants of individual j in the first generation. Moreover, $\{R_{n,j}^*\}$ are iid copies of R_n^*, and thus, R_n^* is equal in distribution to the process obtained by iterating (17) with an initial condition R_0^*. This process can be shown to converge in distribution to R for any initial condition R_0^* satisfying the following moment condition.

Lemma 14. *For any* $R_0^* \geq 0$, *if* $E[Q^\beta], E[(R_0^*)^\beta] < \infty$ *and* $\rho_\beta < 1$ *for some* $\beta > 0$, *then*

$$R_n^* \Rightarrow R,$$

with $E[R^\beta] < \infty$. *Furthermore, under these assumptions, the distribution of* R *is the unique solution with finite* β*-moment to recursion* (17).

We now state the main result of this section; see Theorem 5.1 in [27].

Theorem 5. *Let* $(Q, N, C_1, C_2, \dots)$ *be a nonnegative random vector, with* $N \in \mathbb{N} \cup \{\infty\}$, $P(Q > 0) > 0$ *and* R *be the solution to* (17) *given by* (20). *Suppose that there exists* $j \geq 1$ *with* $P(N \geq j, C_j > 0) > 0$ *such that the measure* $P(\log C_j \in du, C_j > 0, N \geq j)$ *is nonarithmetic, and that for some* $\alpha > 0$, $E[Q^\alpha] < \infty$, $0 < E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right] < \infty$ *and* $E\left[\sum_{i=1}^N C_i^\alpha\right] = 1$. *In addition, assume*

1. $E\left[\left(\sum_{i=1}^N C_i\right)^\alpha\right] < \infty$, *if* $\alpha > 1$; *or,*
2. $E\left[\left(\sum_{i=1}^N C_i^{\alpha/(1+\epsilon)}\right)^{1+\epsilon}\right] < \infty$ *for some* $0 < \epsilon < 1$, *if* $0 < \alpha \leq 1$.

Then,

$$P(R > t) \sim H t^{-\alpha}, \qquad t \to \infty,$$

where $0 \leq H < \infty$ *is given by*

$$\begin{aligned} H &= \frac{1}{E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right]} \int_0^\infty v^{\alpha-1} \left(P(R > v) - E\left[\sum_{i=1}^N 1(C_i R > v)\right]\right) dv \\ &= \frac{E\left[\left(\bigvee_{i=1}^N C_i R_i\right)^\alpha \vee Q^\alpha - \sum_{i=1}^N (C_i R_i)^\alpha\right]}{\alpha E\left[\sum_{i=1}^N C_i^\alpha \log C_i\right]}. \end{aligned}$$

As an illustration of the generality of the developed techniques, we now discuss another example that is closely related to recursion (17), which is given by

$$R \stackrel{\mathcal{D}}{=} \left(\bigvee_{i=1}^N C_i R_i\right) + Q, \tag{22}$$

where $(Q, N, C_1, C_2, \dots)$ is a nonnegative vector with $N \in \mathbb{N} \cup \{\infty\}$, $P(Q > 0) > 0$, and $\{R_i\}_{i \in \mathbb{N}}$ is a sequence of iid random variables independent of $(Q, N, C_1, C_2, \dots)$ having the same distribution as R. Its analysis could follow very closely the steps used for the linear and maximum recursions, except that the constructed solution R would be less explicit. More specifically, one could iterate (22), similarly as it was done in (19) for the maximum recursion. To this end, an iteration $R^{(n)}$ could be constructed as a function of the weights of the first n generations of the tree, and would solve

$$R^{(n)} = \left(\bigvee_{j=1}^{N} C_j R_j^{(n-1)} \right) + Q,$$

where $\{R_j^{(n-1)}\}$ is the corresponding iteration obtained on a subtree that starts on node j in the first generation and ends on the nth generation; clearly $\{R_j^{(n-1)}\}$ is a sequence of iid random variables. Furthermore, it appears that $R^{(n)}$ is monotonically increasing in n, see Eq. (37) in [1], and thus its limit $R = R^{(\infty)} = \lim_{n\to\infty} R^{(n)}$ is properly defined. In addition, by using monotonicity arguments, one can show that

$$R^{(\infty)} = \left(\bigvee_{j=1}^{N} C_j R_j^{(\infty)} \right) + Q,$$

where $\{R_j^{(\infty)}\}$ is the corresponding iterative solution constructed on the infinite subtree that starts at node j in the first generation. Hence, such a constructed R is a solution to (22). Also, since R is bounded from above by the solution (13) to the nonhomogeneous linear equation, sufficient conditions for the finiteness of its moments can be obtained from the corresponding results for the solution in (13). Recursion (22) was termed "discounted tree sums" in [1]; for additional details on the existence and uniqueness of its solution see Sect. 4.4 in [1].

Similarly one could study the third equation from (5),

$$R \stackrel{\mathcal{D}}{=} \left(\sum_{i=1}^{N} C_i R_i \right) \vee Q,$$

by first constructing iteratively an endogenous solution on the weighted branching tree and then develop the conditions for the finiteness of its moments, etc.

Acknowledgements This work is supported by NSF Grant CMMI-1131053.

References

1. Aldous, D.J., Bandyopadhyay, A.: A survey of max-type recursive distributional equation. Ann. Appl. Probab. **15**(2), 1047–1110 (2005)
2. Alsmeyer, G., Kuhlbusch, D.: Double martingale structure and existence of ϕ-moments for weighted branching processes. Münster J Math. **3**, 163–212 (2010)

3. Alsmeyer, G., Meiners, M.: Fixed points of the smoothing transform: two-sided solutions. Probab. Theory Relat. Fields (2012). arXiv:1009.2412
4. Alsmeyer, G., Rösler, U.: A stochastic fixed point equation related to weighted branching with deterministic weights. Electron. J. Probab. **11**, 27–56 (2005)
5. Alsmeyer, G., Rösler, U.: A stochastic fixed point equation for weighted minima and maxima. Annales de l'Institiut Henri Poincaré **44**(1), 89–103 (2008)
6. Alsmeyer, G., Biggins, J.D., Meiners, M.: The functional equation of the smoothing transform. Ann. Probab. (2012). arXiv:0906.3133
7. Alsmeyer, G., Damek, E., Mentemeier, S.: Tails of fixed points of the two-sided smoothing transform. In: Springer Proceedings in Mathematics & Mathematics: Random Matrices and Iterated Random Functions. Springer, Berlin (2013)
8. Athreya, K.B., McDonald, D., Ney, P.: Limit theorems for semi-Markov processes and renewal theory for Markov chains. Ann. Probab. **6**(5), 788–797 (1978)
9. Baryshnikov, Y., Gnedin, A.: Counting intervals in the packing process. Ann. Appl. Probab. **11**(3), 863–877 (2001)
10. Biggins, J.D.: Martingale convergence in the branching random walk. J. Appl. Prob. **14**(1), 25–37 (1977)
11. Biggins, J.D., Kyprianou, A.E.: Seneta-heyde norming in the branching random walk. Ann. Probab. **25**(1), 337–360 (1997)
12. Brandt, A.: The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients. Adv. Appl. Probab. **18**(1), 211–220 (1986)
13. Buraczewski, D., Damek, E., Mentemeier, S., Mirek, M.: Heavy tailed solutions of multivariate smoothing transforms (2012). arXiv:1206.1709
14. Chow, Y.S., Teicher, H.: Probability Theory. Springer, New York (1988)
15. Coffman Jr., E.G., Flatto, L., Jelenković, P.R., Poonen, B.: Packing random intervals on-line. Algorithmica **22**(4), 448–476 (1998)
16. Collamore, J.F., Vidyashankar, A.N.: Tail estimates for stochastic fixed point equations via nonlinear renewal theory. Stoch. Process. Appl **123**(9), 3378–3429 (2013)
17. Durret, R., Liggett, T.: Fixed points of the smoothing transformation. Z. Wahrsch. verw. Gebeite **64**, 275–301 (1983)
18. Fill, J.A., Janson, S.: Approximating the limiting Quicksort distribution. Random Struct. Algorithms **19**(3–4), 376–406 (2001)
19. Goldie, C.M.: Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. **1**(1), 126–166 (1991)
20. Goldie, C.M., Grübel, R.: Perpetuities with thin tails. Adv. Appl. Prob. **28**(2), 463–480 (1996)
21. Grey, D.R.: Regular variation in the tail behaviour of solutions of random difference equations. Ann. Appl. Probab. **4**(1), 169–183 (1994)
22. Grincevičius, A.K.: One limit distribution for a random walk on the line. Lith. Math. J. **15**, 580–589 (1975)
23. Holley, R., Liggett, T.: Generalized potlatch and smoothing processes. Z. Wahrsch. verw. Gebeite **55**, 165–195 (1981)
24. Iksanov, A.M.: Elementary fixed points of the BRW smoothing transforms with infinite number of summands. Stoch. Process. Appl. **114**, 27–50 (2004)
25. Jagers, P., Rösler, U.: Stochastic fixed points involving the maximum. In: Drmota, M., Flajolet, P., Gardy, D., Gittenberger, B. (eds.) Mathematics and Computer Science, vol. III, pp. 325–338. Birkhäuser, Basel (2004)
26. Jelenković, P.R., Olvera-Cravioto, M.: Information ranking and power laws on trees. Adv. Appl. Prob. **42**(4), 1057–1093 (2010)
27. Jelenković, P.R., Olvera-Cravioto, M.: Implicit renewal theory and power tails on trees. Adv. Appl. Prob. **44**(2), 528–561 (2012)
28. Jelenković, P.R., Olvera-Cravioto, M.: Implicit renewal theory for trees with general weights. Stoch. Process. Appl. **122**(9), 3209–3238 (2012)
29. Jessen, A.H., Mikosch, T.: Regularly varying functions. Publications de L'Institut Mathematique, Nouvelle Serie **80**(94), 171–192 (2006)

30. Kahane, J.P., Peyrière, J.: Sur certaines martingales de benoit mandelbrot. Adv. Math. **22**, 131–145 (1976)
31. Kesten, H.: Random difference equations and renewal theory for products of random matrices. Acta Math. **131**(1), 207–248 (1973)
32. Konstantinides, D.G., Mikosch, T.: Large deviations and ruin probabilities for solutions to stochastic recurrence equations with heavy-tailed innovations. Ann. Probab. **33**(5), 1992–2035 (2005)
33. Liu, Q.: Fixed points of a generalized smoothing transformation and applications to the branching random walk. Adv. Appl. Prob. **30**, 85–112 (1998)
34. Liu, Q.: On generalized multiplicative cascades. Stoch. Process. Appl. **86**, 263–286 (2000)
35. Mirek, M.: On fixed points of a generalized multidimensional affine recursion. Probab. Theory Relat. Fields (2011). arXiv:1111.1756v1
36. Olvera-Cravioto, M.: Tail behavior of solutions of linear recursions on trees. Stoch. Process. Appl. **122**(4), 1777–1807 (2012)
37. Olvera-Cravioto, M.: Asymptotics for weighted random sums. Adv. Appl. Prob. **44**(4), 528–561 (2012)
38. Negadailov, P.: Limit theorems for random recurrences and renewal-type processes. PhD Thesis (2010). Available at http://igitur-archive.library.uu.nl/dissertations/
39. Neininger, R.: On a multivariate contraction method for random recursive structures with applications to Quicksort. Random Struct. Algorithm. **19**, 498–524 (2001)
40. Neininger, R., Rüschendorf, L.: A general limit theorem for recursive algorithms and combinatorial structures. Ann. Appl. Prob. **14**(1), 378–418 (2004)
41. Rösler, U.: A limit theorem for Quicksort. RAIRO Inform. Theor. Appl. **25**(1), 85–100 (1991)
42. Rösler, U.: The weighted branching process. In: Dynamics of Complex and Irregular Systems (Bielefeld, 1991). Bielefeld Encounters in Mathematics and Physics VIII, pp. 154–165. World Science, River Edge (1993)
43. Rösler, U., Rüschendorf, L.: The contraction method for recursive algorithms. Algorithmica **29**(1–2), 3–33 (2001)
44. Sgibnev, M.S.: The matrix analogue of the Blackwell renewal theorem on the real line. Sb: Math. **197**(3), 369–386 (2006)
45. Volkovich, Y.: Stochastic analysis of web page ranking. Ph.D. Thesis, University of Twente (2009)
46. Volkovich, Y., Litvak, N.: Asymptotic analysis for personalized web search. Adv. Appl. Prob. **42**(2), 577–604 (2010)
47. Volkovich, Y., Litvak, N., Donato, D.: Determining factors behind the pagerank log-log plot. In: Proceedings of the 5th International Workshop on Algorithms and Models for the Web-Graph, WAW 2007, San Diego (2007)
48. Waymire, E.C., Williams, S.C.: Multiplicative cascades: dimension spectra and dependence. J. Fourier Anal. Appl. **1**, 589–609 (1995). Kahane Special Issue

The Smoothing Transform: A Review of Contraction Results

Gerold Alsmeyer

Abstract Given a sequence $(C,T) = (C,T_1,T_2,\ldots)$ of real-valued random variables, the associated so-called smoothing transform $\mathscr{S}$ maps a distribution F from a subset Γ of distributions on $\mathbb{R}$ to the distribution of $\sum_{i\ge 1} T_i X_i + C$, where $X_1, X_2, \ldots$ are iid with common distribution F and independent of (C,T). This review aims at providing a comprehensive account of contraction properties of $\mathscr{S}$ on subsets Γ specified by the existence of moments up to a given order like, for instance, $\mathscr{P}^p(\mathbb{R}) = \{F : \int |x|^p\, F(dx) < \infty\}$ for $p > 0$ or $\mathscr{P}_c^p(\mathbb{R}) = \{F \in \mathscr{P}^p(\mathbb{R}) : \int x\, F(dx) = c\}$ for $p \ge 1$. The metrics used here are the minimal ℓ_p-metric and the Zolotarev metric ζ_p, both briefly introduced in Sect. 3.

1 Introduction

Any temporally homogeneous Markov chain on the real line or a subset thereof may be described via a random recursive equation *with no branching*, viz.

$$X_n = \Psi_n(X_{n-1}) \tag{1}$$

for $n \ge 1$ and iid random functions $\Psi_1, \Psi_2, \ldots$ independent of X_0. Namely, if P denotes the one-step transition kernel of the chain and

$$G(x,u) := \inf\{y \in \mathbb{R} : P(x,(-\infty,y]) \ge u\}, \quad x \in \mathbb{R},\ u \in (0,1)$$

its associated pseudo-inverse, then one can choose $\Psi_n(x) := G(x,U_n)$ for $n \ge 1$, where $U_1, U_2, \ldots$ are iid *Unif*(0,1) random variables. Provided that the Ψ_n have

G. Alsmeyer (✉)
Westfälische Wilhelms-Universität Münster, Institut für Mathematische Statistik, Einsteinstraße 62, 48149, Münster, Germany
e-mail: gerolda@math.uni-muenster.de

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_9,

additional smoothness properties, for instance, to be (a.s.) globally Lipschitz continuous and contractive in a suitable stochastic sense, stability properties of $(X_n)_{n\ge 0}$ may be studied within the framework of *iterated random functions*, see [17] for a survey and [16, 34] for two more recent contributions of interest. Moreover, any stationary distribution π of the chain is then characterized by the distributional identity

$$X \stackrel{d}{=} \Psi(X) \tag{2}$$

where X has law π, Ψ denotes a generic copy of the Ψ_n independent of X, and $\stackrel{d}{=}$ means equality in distribution. Equation (2) is called a *stochastic fixed-point equation (SFPE)* and π (and also X) a solution to it. The case when Ψ is a random affine linear function and solutions are called perpetuities has received particular interest in the literature, see e.g. [5, 20, 44] and further references therein.

A random recursive equation *with branching* occurs if the right-hand side of (1) involves multiple copies of X_{n-1}, i.e.

$$X_n \;=\; \Psi_n(X_{n-1,1}, X_{n-1,2}, \ldots)$$

for $n \ge 1$, where $(X_{n-1,k})_{k\ge 1}$ is a sequence of iid copies of X_{n-1} and further independent of Ψ_n. Again, of particular interest and also the topic of this article is the situation when the Ψ_n are random affine linear functions, a generic copy thus being of the form

$$\Psi(x_1, x_2, \ldots) = \sum_{k\ge 1} T_k x_k + C$$

for a sequence of real-valued random variables $(C, T_1, T_2, \ldots)$. This leads to the so-called (going back to Durrett and Liggett [18]) *smoothing transform(ation)*

$$\mathscr{S}: \quad F \;\mapsto\; \mathscr{L}\left(\sum_{k\ge 1} T_k X_k + C\right) \tag{3}$$

which maps a distribution $F \in \mathscr{P}(\mathbb{R})$ to the law of $\sum_{k\ge 1} T_k X_k + C$, where $X_1, X_2, \ldots$ are independent of $(C, T_1, T_2, \ldots)$ with common distribution F. It has been studied by many authors due to its occurrence in various fields of applied probability: probabilistic combinatorial optimization [1], stochastic geometry and random fractals [21, 33, 37], the analysis of recursive algorithms and data structures [22, 36, 39, 41] and branching particle systems [10, 25].

On the event where

$$N := \sum_{k\ge 1} \mathbf{1}_{\{T_k \ne 0\}}$$

is infinite, the sum $\sum_{k\ge 1} T_k X_k$ in (3) is understood as the limit of the finite partial sums $\sum_{k=1}^{n} T_k X_k$ in the sense of convergence in probability. Then $\mathscr{S}(F)$ is indeed defined *for all* $F \in \mathscr{P}(\mathbb{R})$ if

$$\mathbb{P}(N < \infty) = 1, \tag{A0}$$

but exists only for F from a subset of $\mathscr{P}(\mathbb{R})$ (always containing δ_0) otherwise. Subsets of interest here are typically characterized by the existence of moments of certain order, viz.

$$\mathscr{P}^p(\mathbb{R}) \ := \ \left\{ F \in \mathscr{P}(\mathbb{R}) : \int |x|^p \ F(dx) < \infty \right\},$$

for any $p > 0$ or, more specifically, the sets of all centered, respectively centered and standardized distributions on $\mathbb{R}$, that is

$$\mathscr{P}_0^1(\mathbb{R}) \ := \ \left\{ F \in \mathscr{P}^1(\mathbb{R}) : \int x \ F(dx) = 0 \right\},$$

$$\mathscr{P}_{0,1}^2(\mathbb{R}) \ := \ \left\{ F \in \mathscr{P}_0^2(\mathbb{R}) : \int x \ F(dx) = 0 \text{ and } \int x^2 \ F(dx) = 1 \right\}.$$

Section 4 will provide conditions for $\mathscr{S}$ to be a self-map on some $\Gamma \subseteq \mathscr{P}^p(\mathbb{R})$, and these do not necessarily include (A0). Under the standing assumption that

$$\mathbb{P}(N \ge 2) > 0, \tag{A1}$$

our goal is then to give a systematic account of conditions under which $\mathscr{S}$ is, in some sense, contractive on Γ with respect to a suitable complete metric ρ and therefore possessing a unique fixed point in Γ, characterized by the SFPE

$$X \ \stackrel{d}{=} \ \sum_{k\ge 1} T_k X_k + C \tag{4}$$

when stated in terms of random variables, where $X_1, X_2, \ldots$ are iid copies of X and independent of $(C, T_1, T_2, \ldots)$. Three types of contraction on (Γ, ρ) will be discussed:

- *Contraction*, i.e. $\rho(\mathscr{S}(F), \mathscr{S}(G)) \le \alpha\, \rho(F, G)$ for all $F, G \in \Gamma$ and some $\alpha \in (0, 1)$.
- *Quasi-contraction*, which holds if $\mathscr{S}^n$ is a contraction for some $n \in \mathbb{N}$.
- *Local contraction*, i.e. $\rho(\mathscr{S}^n(F), \mathscr{S}^{n+1}(F)) \le c\, \alpha^n$ for some $F \in \Gamma, \alpha \in (0, 1)$ and $c \in \mathbb{R}_>$.

The metrics to be considered here because of their good performance in connection with $\mathscr{S}$ are the *minimal L^p-metric ℓ_p* and the *Zolotarev metric ζ_p* for $p > 0$, both briefly introduced in Sect. 3.

Our review draws on results in [35, 38, 40, 42] supplemented by a number of extensions so as to provide a more complete picture. The last two references may also be consulted for multivariate extensions not discussed here. Further information on the set of solutions to (4), especially for the homogeneous case ($C = 0$), has been obtained by many authors, see [2, 3, 6, 9, 11, 14, 15, 18, 26, 31], but will not either be an issue here. The same goes for results on the tail behavior of solutions, see [7, 23, 27–30, 32].

The rest of this paper is organized as follows. In Sect. 2, a brief introduction of the weighted branching model associated with $\mathscr{S}$ is given. It provides the appropriate framework to study the iterates $\mathscr{S}^n$ of $\mathscr{S}$ (Sect. 2). As already mentioned, Sect. 3 collects useful information on the probability metrics ℓ_p and ζ_p and Sect. 4 gives conditions for $\mathscr{S}$ to be a self-map of $\mathscr{P}^p(\mathbb{R})$ or subsets thereof. An auxiliary result on the behavior of the mean values of $\mathscr{S}^n(F)$ for $F \in \mathscr{P}^1(\mathbb{R})$ and as $n \to \infty$ is stated in Sect. 5. After these preliminaries, all contraction results for $\mathscr{S}$ are presented in the main Sect. 6, with proofs for some of these results included. Finally, an Appendix provides a short survey of some useful results in connection with Banach's fixed-point theorem, the latter being stated there as well. It also lists some well-known martingale inequalities which form an essential tool for the proofs of the contraction results and are included here to make the presentation more self-contained.

2 The Iterates of $\mathscr{S}$ and Weighted Branching

In order to study contraction properties of $\mathscr{S}$, a representation of $(\mathscr{S}^n(F))_{n\geq 1}$, the sequence of iterates of $\mathscr{S}$ applied to some $F \in \mathscr{P}(\mathbb{R})$, in terms of random variables is needed. The weighted branching model to be introduced next and taken from [40] provides an appropriate framework.

Consider the infinite *Ulam-Harris tree*

$$\mathbb{T} := \bigcup_{n\geq 0} \mathbb{N}^n, \quad \mathbb{N}^0 := \{\varnothing\},$$

of finite integer words having the empty word $\varnothing$ as its root. As common, we write $\mathsf{v}_1 \ldots \mathsf{v}_n$ as shorthand for $(\mathsf{v}_1, \ldots, \mathsf{v}_n)$, $|\mathsf{v}|$ for the length of v, and uv for the concatenation of u and v. If $\mathsf{v} = \mathsf{v}_1 \ldots \mathsf{v}_n$, put further $\mathsf{v}|0 := \varnothing$ and $\mathsf{v}|k := \mathsf{v}_1 \ldots \mathsf{v}_k$ for $1 \leq k \leq n$. The unique shortest path (geodesic) from the root $\varnothing$ to v, or the ancestral line of v when using a genealogical interpretation, is then given by

$$\mathsf{v}|0 = \varnothing \to \mathsf{v}|1 \to \ldots \to \mathsf{v}|n-1 \to \mathsf{v}|n = \mathsf{v}.$$

The tree $\mathbb{T}$ is now turned into a *weighted (branching) tree* by attaching a *random weight* to each of its edges. Let $T_i(\mathsf{v})$ denote the weight attached to the edge $(\mathsf{v}, \mathsf{v}i)$ and assume that the $T(\mathsf{v}) := (T_i(\mathsf{v}))_{i\geq 1}$ for $\mathsf{v} \in \mathbb{T}$ form a family of iid copies of $T = (T_i)_{i\geq 1}$. The number of nonzero weights $T_i(\mathsf{v})$ is denoted $N(\mathsf{v})$, thus

$$N(\mathsf{v}) \;:=\; \sum_{i\ge 1} \mathbf{1}_{\{T_i(\mathsf{v})\neq 0\}} \;\stackrel{d}{=}\; N.$$

Put further $L(\varnothing) := 1$ and then recursively

$$L(\mathsf{vi}) \;:=\; L(\mathsf{v})T_i(\mathsf{v})$$

for any $\mathsf{v} \in \mathbb{T}$ and $\mathsf{i} \in \mathbb{N}$, which is equivalent to

$$L(\mathsf{v}) \;=\; T_{\mathsf{v}_1}(\varnothing)T_{\mathsf{v}_2}(\mathsf{v}|1)\cdot\ldots\cdot T_{\mathsf{v}_n}(\mathsf{v}|n-1)$$

for any $\mathsf{v} = \mathsf{v}_1\ldots\mathsf{v}_n \in \mathbb{T}$. Hence, $L(\mathsf{v})$ equals the total weight of the minimal path from $\varnothing$ to v obtained upon multiplication of the edge weights along this path.

With the help of a weighted branching model as just introduced, we are now able to describe the iterations of the homogeneous smoothing transform in a convenient way. Namely, if $\mathscr{S}$ is given by (3) with $C = 0$, $\boldsymbol{X} := \{X(\mathsf{v}) : \mathsf{v} \in \mathbb{T}\}$ denotes a family of iid random variables independent of $\boldsymbol{T} := (T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ with common distribution F, and

$$Y_n \;:=\; \sum_{|\mathsf{v}|=n} L(\mathsf{v})X(\mathsf{v})$$

for $n \ge 0$, then $\mathscr{S}^n(F) = \mathscr{L}(Y_n)$ holds true for each $n \ge 0$. We call $(Y_n)_{n\ge 0}$ *weighted branching process (WBP) associated with* $\boldsymbol{T}\otimes\boldsymbol{X} := (T(\mathsf{v}), X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$. In the special case when $X(\mathsf{v}) = 1$ for $\mathsf{v} \in \mathbb{T}$, it is simply called *weighted branching process associated with* $\boldsymbol{T}$.

It is not difficult to extend the previous weighted branching model so as to describe the iterations of $\mathscr{S}$ in the nonhomogeneous case when $\mathbb{P}(C = 0) < 1$. To this end, let $\boldsymbol{C}\otimes\boldsymbol{T} = (C(\mathsf{v}), T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ denote a family of iid copies of (C, T), $T := (T_i)_{i\ge 1}$, and $\boldsymbol{X}$ be independent of $\boldsymbol{C}\otimes\boldsymbol{T}$. Then defining $Y(\varnothing) = X(\varnothing)$ and

$$Y_n \;:=\; \sum_{k=0}^{n-1}\sum_{|\mathsf{v}|=k} L(\mathsf{v})C(\mathsf{v}) + \sum_{|\mathsf{v}|=n} L(\mathsf{v})X(\mathsf{v})$$

for $n \ge 1$, it is readily verified that $\mathscr{S}^n(F) = \mathscr{L}(Y_n)$ holds true for each $n \ge 0$. In this case, we call $\boldsymbol{Y} := (Y_n)_{n\ge 0}$ the *weighted branching process associated with* $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X} := (C(\mathsf{v}), T(\mathsf{v}), X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$.

We proceed to a description of the recursive structure of WBPs after the following useful definition of the *shift operators* $[\cdot]_{\mathsf{v}}$, $\mathsf{v} \in \mathbb{T}$. Given any function Ψ of $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}$ and any $\mathsf{v} \in \mathbb{T}$, put

$$[\Psi(\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X})]_{\mathsf{v}} \;:=\; \Psi\left((C(\mathsf{vw}), T(\mathsf{vw}), X(\mathsf{vw}))_{\mathsf{w}\in\mathbb{T}}\right),$$

which particularly implies

$$[\Psi(\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X})]_{\mathsf{v}} \;=\; \Psi\left([\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}]_{\mathsf{v}}\right).$$

If we think of $\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X}$ as the family of random variables associated with $\mathbb{T}$, then $[\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X}]_{\mathsf{v}}$ equals its subfamily and copy associated with the subtree $\mathbb{T}(\mathsf{v})$ rooted at v which is isomorphic to $\mathbb{T}$. Obviously, $\boldsymbol{L} := (L(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ is a function of $\boldsymbol{T}$, and one can easily verify that $[\boldsymbol{L}]_{\mathsf{v}} = ([L(\mathsf{w})]_{\mathsf{v}})_{\mathsf{w}\in\mathbb{T}}$ with

$$[L(\mathsf{w})]_{\mathsf{v}} := T_{\mathsf{w}_1}(\mathsf{v})T_{\mathsf{w}_2}(\mathsf{v}\mathsf{w}_1)\cdot\ldots\cdot T_{\mathsf{w}_n}(\mathsf{v}\mathsf{w}_1\ldots\mathsf{w}_{n-1})$$

if $\mathsf{w} = \mathsf{w}_1\ldots\mathsf{w}_n$. Hence, $[L(\mathsf{w})]_{\mathsf{v}}$ gives the total weight of the minimal path from v to $\mathsf{v}\mathsf{w}$. Notice that, for all $\mathsf{v},\mathsf{w} \in \mathbb{T}$,

$$L(\mathsf{v}\mathsf{w}) = L(\mathsf{v})\cdot[L(\mathsf{w})]_{\mathsf{v}}$$

and therefore

$$[L(\mathsf{w})]_{\mathsf{v}} = \frac{L(\mathsf{v}\mathsf{w})}{L(\mathsf{v})}$$

for all $\mathsf{w} \in \mathbb{T}$ if $L(\mathsf{v}) \neq 0$. For later use, we put

$$\mathscr{F}_n := \sigma\left(T(\mathsf{v}) : |\mathsf{v}| \le n-1\right) \tag{5}$$

for $n \ge 1$ and let $\mathscr{F}_0$ be the trivial σ-field. Observe that $\mathscr{F}_n \supset \sigma(L(\mathsf{v}) : |\mathsf{v}| \le n)$ for each $n \ge 0$.

Finally, we define

$$\mathfrak{m}(\theta) := \mathbb{E}\left(\sum_{i\ge1}|T_i|^\theta\right) \tag{6}$$

for $\theta \ge 0$ which plays an important role in the study of $\mathscr{S}$. For instance, it is well-known that, if $C = 0$ (homogeneous case), $T \ge 0$ and N is bounded, then $\mathscr{S}$ has nontrivial fixed points in $\mathscr{P}(\mathbb{R}_{\ge})$ iff $m(\alpha) = 1$ and

$$\mathfrak{m}'(\alpha) = \mathbb{E}\left(\sum_{i\ge1}|T_i|^\theta\log|T_i|\right) \le 0$$

for some $\alpha \in (0,1]$, see [18]. The function $\mathfrak{m}$ is convex on $\{\theta : \mathfrak{m}(\theta) < \infty\}$, satisfies $\mathfrak{m}(0) = \mathbb{E}N$ and possesses at most two values $\alpha < \beta$ such that $\mathfrak{m}(\alpha) = \mathfrak{m}(\beta) = 1$. If this is the case, then $\mathfrak{m}'(\alpha) < 0$ and $\mathfrak{m}'(\beta) > 0$. The value α is called *characteristic exponent* of T, owing to its role in connection with the existence of fixed points of $\mathscr{S}$. Under appropriate regularity assumptions, the value β determines the tail index of fixed points of $\mathscr{S}$, see [7,27–29]. As for the contractive behavior of $\mathscr{S}$ on $\mathscr{P}^p(\mathbb{R})$ or subsets thereof, we will see that $\mathfrak{m}(p) < 1$ constitutes a minimal requirement.

3 Probability Metrics

3.1 The Minimal L^p-Metric

Given a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, let $L^p(\mathbb{P}) = L^p(\Omega, \mathfrak{A}, \mathbb{P})$ for $p > 0$ denote the vector space of p times integrable random variables on $(\Omega, \mathfrak{A}, \mathbb{P})$. Then $\|X\|_p := (\mathbb{E}|X|^p)^{1\wedge(1/p)}$ defines a complete (pseudo-)norm on $L^p(\mathbb{P})$ if $p \geq 1$, but fails to do so if $0 < p < 1$. On the other hand,

$$\ell_p(X, Y) \ := \ \|X - Y\|_p$$

provides us with a complete (pseudo-)metric on $L^p(\mathbb{P})$ for each $p > 0$.

A pair (X, Y) of real-valued random variables defined on $(\Omega, \mathfrak{A}, \mathbb{P})$ is called (F, G)-*coupling* if $\mathscr{L}(X) = F$ and $\mathscr{L}(Y) = G$. In this case, we will use the shorthand notation $(X, Y) \sim (F, G)$ hereafter. For a distribution function F on $\mathbb{R}$, let F^{-1} denote its pseudo-inverse, thus $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$ for $u \in (0, 1)$. Then $F^{-1}(U)$ has distribution F if $\mathscr{L}(U) = \mathit{Unif}(0, 1)$. Now, for each $p > 0$, the mapping $\ell_p : \mathscr{P}^p(\mathbb{R}) \times \mathscr{P}^p(\mathbb{R}) \to \mathbb{R}_{\geq}$, defined by

$$\ell_p(F, G) \ := \inf_{(X,Y)\sim(F,G)} \|X - Y\|_p, \tag{7}$$

is a metric on $\mathscr{P}^p(\mathbb{R})$, called *minimal* L^p*-metric* (also *Mallows metric* in [40]). Moreover, the infimum in (7) is attained, namely

$$\ell_p(F, G) \ = \ \|F^{-1}(U) - G^{-1}(U)\|_p$$

for any $\mathit{Unif}(0, 1)$ random variable U. The following characterization of convergence with respect to ℓ_p is easily verified.

Proposition 1. *Let $p > 0$ and $(F_n)_{n\geq 0}$ be a sequence of distributions in $\mathscr{P}^p(\mathbb{R})$. Then the following assertions are equivalent:*

(a) $F_n \xrightarrow{\ell_p} F$, *i.e.* $\lim_{n\to\infty} \ell_p(F_n, F) = 0$.
(b) $F_n \xrightarrow{w} F$ *and* $\lim_{n\to\infty} \int |x|^p \, F_n(dx) = \int |x|^p \, F(dx) < \infty$.
(c) $F_n \xrightarrow{w} F$ *and* $x \mapsto |x|^p$ *is ui with respect to the* F_n*, that is*

$$\lim_{a\to\infty} \sup_{n\geq 1} \int_{(-a,a)^c} |x|^p \ F_n(dx) \ = \ 0.$$

Moreover, the space $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ is complete for each $p > 0$.

For any distribution $F \in \mathscr{P}^1(\mathbb{R})$ with mean value $\mathbb{E}F := \int x \, F(dx)$, let F^0 denote its centering, thus $F^0(t) := F(t + \mathbb{E}F)$ for $t \in \mathbb{R}$. The next lemma provides information about the relation between $\ell_p(F, G)$ and $\ell_p(F^0, G^0)$ for $p \geq 1$.

Lemma 1. *Given $p \geq 1$, distributions $F, G \in \mathscr{P}^p(\mathbb{R})$ with mean values $\mathbb{E}F, \mathbb{E}G$ and a Unif$(0,1)$ random variable U, it holds true that*

$$\ell_p(F^0, G^0) \;=\; \|(F^{-1}(U) - \mathbb{E}F) - (G^{-1}(U) - \mathbb{E}G)\|_p, \tag{8}$$

$$\ell_p(F, G) \;=\; \|((F^0)^{-1}(U) + \mathbb{E}F) - ((G^0)^{-1}(U) + \mathbb{E}G)\|_p, \tag{9}$$

and therefore

$$|\ell_p(F,G) - \ell_p(F^0, G^0)| \;\leq\; |\mathbb{E}F - \mathbb{E}G|. \tag{10}$$

If $p = 2$, then furthermore

$$\ell_2^2(F,G) \;=\; \ell_2^2(F^0, G^0) + (\mathbb{E}F - \mathbb{E}G)^2. \tag{11}$$

Proof. For (8) and (9), it suffices to note that $F^0(t) = F(t + \mathbb{E}F)$ obviously implies $(F^0)^{-1}(t) = F^{-1}(t) - \mathbb{E}F$ for all $t \in \mathbb{R}$. If $p = 2$, then (9) with $X := (F^0)^{-1}(U)$ and $Y := (G^0)^{-1}(U)$ yields

$$\begin{aligned} \ell_2^2(F,G) \;&=\; \mathbb{E}\big((X - Y) + (\mathbb{E}F - \mathbb{E}G)\big)^2 \\ &=\; \mathbb{E}(X-Y)^2 + 2(\mathbb{E}F - \mathbb{E}G)\,\mathbb{E}(X-Y) + (\mathbb{E}F - \mathbb{E}G)^2 \\ &=\; \ell_2^2(F^0, G^0) + (\mathbb{E}F - \mathbb{E}G)^2, \end{aligned}$$

where $\mathbb{E}X = \mathbb{E}Y = 0$ has been utilized. □

3.2 The Zolotarev Metric

We now turn to an alternative probability metric which is better tailored to situations where $\mathscr{S}$ is contractive on subsets of $\mathscr{P}^p(\mathbb{R})$ with specified moments of integral order $\leq p$.

Let $\mathscr{C}^0(\mathbb{R})$ denote the space of continuous functions $f : \mathbb{R} \to \mathbb{R}$ and $\mathscr{C}^m(\mathbb{R})$ for $m \in \mathbb{N}$ the subspace of m times continuously differentiable complex-valued functions. For $p = m + \alpha$ with $m \in \mathbb{N}_0$ and $0 < \alpha \leq 1$, put

$$\mathfrak{F}_p \;:=\; \left\{f \in \mathscr{C}^m(\mathbb{R}) : |f^{(m)}(x) - f^{(m)}(y)| \leq |x-y|^\alpha \text{ for all } x, y \in \mathbb{R}\right\}.$$

which obviously contains the monomials $x \mapsto x^k$ for $k = 1, \ldots, m$ as well as $x \mapsto sign(x)|x|^p/c_p$ and $x \mapsto |x|^p/c_p$ for some $c_p \in \mathbb{R}_>$. Finally, if $p > 1$ and thus $m \in \mathbb{N}$, then denote by $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$, $\mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{R}^m$, the set of distributions on $\mathbb{R}$ having k^{th} moment z_k for $k = 1, \ldots, m$.

Zolotarev [46] introduced the metric ζ_p on $\mathscr{P}^p(\mathbb{R})$, defined by

$$\zeta_p(F,G) \;:=\; \sup_{f \in \mathfrak{F}_p,\,(X,Y)\sim(F,G)} \left|\mathbb{E}\big(f(X) - f(Y)\big)\right| \tag{12}$$

and nowadays named after him. Via a Taylor expansion of the functions $f \in \mathfrak{F}_p$ in (12), it can be shown that $\zeta_p(F,G)$ is finite for all $F, G \in \mathscr{P}^p(\mathbb{R})$ if $0 < p \le 1$, and for all $F, G \in \mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ and $\mathbf{z} \in \mathbb{R}^m$ if $p > 1$. On the other hand, in the last case $\zeta_p(F,G) = \infty$ for distributions $F, G \in \mathscr{P}^p(\mathbb{R})$ that do not have the same integral moments up to order m. We thus see that ζ_p defines a proper probability metric on $\mathscr{P}^p(\mathbb{R})$ only for $0 < p \le 1$ and on $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for any $\mathbf{z} \in \mathbb{R}^m$, otherwise. Here we should add that $\zeta_p(F,G) = 0$ implies $F = G$ because $\mathscr{C}^m_b(\mathbb{R}) := \{f \in \mathscr{C}^m(\mathbb{R}) : f^{(m)}$ is bounded$\}$ is a measure determining class for each $m \in \mathbb{N}_0$.

Given a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, ζ_p can also be defined on $L^p = L^p(\mathbb{P})$, viz.

$$\zeta_p(X,Y) \;:=\; \sup_{f \in \mathfrak{F}_p} \left|\mathbb{E}\big(f(X) - f(Y)\big)\right|, \tag{13}$$

and constitutes a *pseudo-metric* there if $0 < p \le 1$. If $p > 1$, then this is true only on $L^p_{\mathbf{z}} = L^p_{\mathbf{z}}(\mathbb{P}) := \{X \in L^p(\mathbb{P}) : \mathbb{E}X^k = z_k$ for $k = 1, \ldots, m\}$ for any $\mathbf{z} \in \mathbb{R}^m$. Recall that a pseudo-metric has the same properties as a metric with one exception: $\zeta_p(X,Y) = 0$ does not necessarily imply $X = Y$ (here not even with probability one: just take two iid X, Y which are not a.s. constant).

A pseudo-metric ρ on a set of random variables is called *simple* if it depends only on the marginals of the random variables being compared, and *compound* otherwise. It is called $(p,+)$-*ideal* if

$$\rho(cX, cY) \;=\; |c|^p\, \rho(X,Y) \tag{14}$$

for all $c \in \mathbb{R}$ and

$$\rho(X+Z, Y+Z) \;\le\; \rho(X,Y) \tag{15}$$

for any Z independent of X, Y and with well-defined $\rho(X+Z, Y+Z)$. Obviously, ζ_p is simple, namely

$$\zeta_p(X,Y) \;=\; \zeta_p(F,G)$$

for any random variables X, Y with respective laws F, G, whereas the L^p-pseudo-metrics ℓ_p are compound. It will be shown in Proposition 2(a) below that ζ_p is also $(p,+)$-ideal on any $L^p_{\mathbf{z}}$ for $\mathbf{z} \in \mathbb{R}^m$. As for the minimal L^p-metric, one can easily see that it is $(r,+)$-ideal for $r = p \wedge 1$.

In the following, $\mathscr{P}^p_*(\mathbb{R})$, L^p_* stand for $\mathscr{P}^p(\mathbb{R})$, L^p if $0 < p \le 1$, and for $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$, $L^p_{\mathbf{z}}$ for arbitrary $\mathbf{z} \in \mathbb{R}^m$ if $p > 1$. The subsequent propositions gather some useful properties of ζ_p. For a proof we refer to Zolotarev's original work [46]

Proposition 2. *Let $p = m + \alpha$ for some $m \in \mathbb{N}_0$ and $0 < \alpha \le 1$. Then ζ_p, defined by* (12) *or* (13)*, has the following properties:*

(a) ζ_p is a $(p,+)$-ideal pseudo-metric on L^p_*.
(b) For any $X, Y \in L^p_*$,

$$\zeta_p(X,Y) \;\le\; \frac{\Gamma(1+\alpha)}{\Gamma(1+p)}\,\Theta_p(X,Y), \tag{16}$$

where $\Theta_p(X,Y) := \ell_p(X,Y)$ *if* $0 < p = \alpha \le 1$, *and*

$$\Theta_p(X,Y) \;:=\; \ell_p(X,Y)^{\alpha}\,\|X\|_p^m + m\,\ell_p(X,Y)\left(\ell_p(X,Y) + \|Y\|_p\right)^{m-1}$$

if $s \ge 1$.

Convergence with respect to the Zolotarev metric is characterized by a second proposition which may be deduced with the help of the previous one. It particularly shows that ζ_p-convergence and ℓ_p-convergence are equivalent.

Proposition 3. *Under the same assumptions as in the previous result, the following properties hold true for* ζ_p*:*

(a) $\zeta_p(F_n,F) \to 0$ *implies* $\ell_p(F_n,F) \to 0$ *and thus particularly* $F_n \stackrel{w}{\to} F$ *for any* $F, F_1, F_2, \ldots \in \mathscr{P}_*^p(\mathbb{R})$.
(b) Conversely, $\ell_p(F_n,F) \to 0$ *implies* $\Theta_p(F_n,F) \to 0$ *and therefore, by* (16), $\zeta_p(F_n,F) \to 0$ *for any* $F, F_1, F_2, \ldots \in \mathscr{P}_*^p(\mathbb{R})$.
(c) The metric space $(\mathscr{P}_*^p(\mathbb{R}), \zeta_p)$ *is complete.*

4 Conditions for $\mathscr{S}$ to Be a Self-Map of $\mathscr{P}^p(\mathbb{R})$

In order to study the contractive behavior of $\mathscr{S}$ on $\mathscr{P}^p(\mathbb{R})$ for $p > 0$, we must first provide conditions that ensure that $\mathscr{S}$ is a self-map on this subset of distributions on $\mathbb{R}$. In other words, we need conditions on $(C,T) = (C,(T_i)_{i\ge1})$ such that

$$\sum_{i\ge1} T_i X_i + C \;\in\; L^p$$

whenever the iid $X_1, X_2, \ldots$ are in L^p. Choosing $X_1 = X_2 = \ldots = 0$, we see that $C \in L^p$ is necessary, so that we are left with the problem of finding conditions on T such that $\sum_{i\ge1} T_i X_i \in L^p$ if this is true for the X_i. The main result is stated as Proposition 4 below and does not need $N = \sum_{i\ge1} \mathbf{1}_{\{T_i \ne 0\}}$ to be a.s. finite. Therefore, $\sum_{i\ge1} T_i X_i \in L^p$ is generally to be understood in the sense of L^p-convergence of the finite partial sums $\sum_{i=1}^n T_i X_i$, which particularly implies convergence in probability. Before stating the result let us define

$$\mathscr{P}_c^p(\mathbb{R}) \;:=\; \left\{ F \in \mathscr{P}^p(\mathbb{R}) : \int x\, F(dx) = c \right\}$$

and also $L_c^p := \{X \in L^p : \mathbb{E}X = c\}$ for $p \ge 1$ and $c \in \mathbb{R}$.

Proposition 4. *Let $T = (T_i)_{i\ge 1}$ and $(X_i)_{i\ge 1}$ be independent sequences on a given probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ such that $X_1, X_2, \ldots$ are iid and in L^p. Then each of the following set of conditions implies $\sum_{i\ge 1} T_i X_i \in L^p$:*

(i) $0 < p \le 1$ *and* $\sum_{i\ge 1} |T_i|^p \in L^1$.
(ii) $1 < p \le 2$, $\sum_{i\ge 1} T_i \in L^p$ *and* $\sum_{i\ge 1} |T_i|^p \in L^1$.
(iii) $2 \le p < \infty$, $\sum_{i\ge 1} T_i \in L^p$ *and* $\sum_{i\ge 1} T_i^2 \in L^{p/2}$.
(iv) $1 < p \le 2$, $\sum_{i\ge 1} |T_i|^p \in L^1$ *and* $\mathbb{E}X_1 = 0$.
(v) $2 \le p < \infty$, $\sum_{i\ge 1} T_i^2 \in L^{p/2}$ *and* $\mathbb{E}X_1 = 0$.

Conversely, if $1 < p < \infty$, *then*

(a) $\sum_{i\ge 1} T_i X_i \in L^p$ *for any choice of T-independent and iid $X_1, X_2, \ldots$ in L^p implies* $\sum_{i\ge 1} T_i \in L^p$ *and* $\sum_{i\ge 1} T_i^2 \in L^{p/2}$.
(b) $\sum_{i\ge 1} T_i X_i \in L^p$ *for any choice of T-independent and iid $X_1, X_2, \ldots \in L_0^p$ implies* $\sum_{i\ge 1} T_i^2 \in L^{p/2}$.

It should be observed that, in view of (iii) and (v), the implications in the converse parts (a) and (b) are in fact equivalences if $p \ge 2$. It is tacitly understood there that the underlying probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ is rich enough to carry T-independent iid $X_1, X_2, \ldots$ with arbitrary distribution in $\mathscr{P}^p(\mathbb{R})$, which is obviously the case if it carries a sequence of iid *Unif*$(0,1)$ variables. Our proof will show that it is even enough if there exist T-independent iid $X_1, X_2, \ldots$ taking values ± 1 with probability $1/2$ each.

Proof. (i) If $0 < p \le 1$, the subadditivity of $x \mapsto x^p$ for $x \ge 0$ immediately implies under the given assumptions that

$$\mathbb{E}\left(\sum_{i\ge 1} |T_i X_i|\right)^p \le \sum_{i\ge 1} \mathbb{E}|T_i X_i|^p = \mathbb{E}|X_1|^p \sum_{i\ge 1} \mathbb{E}|T_i|^p < \infty$$

and thus the almost sure absolute convergence of $\sum_{i\ge 1} T_i X_i$ as well as its integrability of order p.

(ii) Here we argue that $(\sum_{i=1}^n T_i X_i)_{n\ge 1}$ forms a Cauchy sequence in $(L^p(\mathbb{P}), \|\cdot\|_p)$ and is therefore L^p-convergent. First note that $\mathbb{E}(\sum_{i\ge 1} |T_i|^p) = \sum_{i\ge 1} \mathbb{E}|T_i|^p$ implies $T_i \in L^p$ for each $i \ge 1$, which in combination with $X_i \in L^p$ for each $i \ge 1$ ensures that $\sum_{i=m}^n T_i X_i \in L^p$ for all $n \ge m \ge 1$. Denoting by μ the expectation of the X_i, we have that $(\sum_{i=m}^k T_i(X_i - \mu))_{m\le k\le n}$ conditioned upon T forms an L^p-martingale, for T and $(X_i)_{i\ge 1}$ are independent. Since $1 < p \le 2$, the even function $x \mapsto |x|^p$ is convex with concave derivative on $\mathbb{R}_\ge$ which allows us to make use of the Topchiĭ-Vatutin inequality (see (44) in the Appendix). This yields

$$\mathbb{E}\left(\left|\sum_{i=m}^n T_i(X_i - \mu)\right|^p \,\middle|\, T\right) \le 2\,\mathbb{E}|X_1 - \mu|^p \sum_{i=m}^n |T_i|^p \quad \text{a.s.}$$

and then by taking unconditional expectations

$$\left\| \sum_{i=m}^{n} T_i (X_i - \mu) \right\|_p \;\le\; 2\, \|X_1 - \mu\|_p \left\| \sum_{i=m}^{n} |T_i|^p \right\|_1^{1/p}.$$

Since $\sum_{i\ge 1} |T_i|^p \in L^1$, the right-hand side converges to zero as $m,n \to \infty$. By using the second assumption $\sum_{i\ge 1} T_i \in L^p$, we infer that $\lim_{m,n\to\infty} \| \sum_{i=m}^{n} T_i \|_p = 0$ as well, whence finally

$$\left\| \sum_{i=m}^{n} T_i X_i \right\|_p \;\le\; \left\| \sum_{i=m}^{n} T_i (X_i - \mu) \right\|_p + |\mu| \left\| \sum_{i=m}^{n} T_i \right\|_p \;\to\; 0 \qquad (17)$$

as $m,n \to \infty$.

(iii) Here we use the same Cauchy sequence argument as in (ii), but make use of Burkholder's inequality (see (18) in the Appendix). This yields

$$\mathbb{E}\left(\left| \sum_{i=m}^{n} T_i (X_i - \mu) \right|^p \middle| T \right) \;\le\; b_p^p\, \mathbb{E}\left(\left(\sum_{i=m}^{n} T_i^2 (X_i - \mu)^2 \right)^{p/2} \middle| T \right) \quad \text{a.s.}$$

for a constant $b_p \in \mathbb{R}_>$ which only depends on p. Next, put $\Sigma_{m:n} := (\sum_{i=m}^{n} T_i^2)^{1/2}$ for $n \ge m \ge 1$. Given T and $\Sigma_{m:n} \ne 0$, the vector

$$\left(\frac{T_m^2}{\Sigma_{m:n}^2}, \ldots, \frac{T_n^2}{\Sigma_{m:n}^2} \right)$$

defines a discrete probability distribution on $\{m,\ldots,n\}$, which in combination with the independence of T and $(X_i)_{i\ge 1}$, the convexity of $x \mapsto x^{p/2}$ for $x \ge 0$ and $p \ge 2$ and an appeal to Jensen's inequality yields

$$\begin{aligned}
\mathbb{E}\left(\left(\sum_{i=m}^{n} T_i^2 (X_i - \mu)^2 \right)^{p/2} \middle| T \right) &= \mathbb{E}\left(\left(\sum_{i=m}^{n} \frac{T_i^2}{\Sigma_{m:n}^2} \Sigma_{m:n}^2 (X_i - \mu)^2 \right)^{p/2} \middle| T \right) \\
&\le \mathbb{E}\left(\sum_{i=m}^{n} \frac{T_i^2}{\Sigma_{m:n}^2} \Sigma_{m:n}^p |X_i - \mu|^p \middle| T \right) \\
&= \left(\Sigma_{m:n}^p \sum_{i=m}^{n} \frac{T_i^2}{\Sigma_{m:n}^2} \right) \mathbb{E}|X_1 - \mu|^p \\
&= \Sigma_{m:n}^p\, \mathbb{E}|X_1 - \mu|^p \quad \text{a.s. on } \{\Sigma_{m:n} > 0\}.
\end{aligned}$$

But if $\Sigma_{m:n} = 0$, the inequality is trivially satisfied. Since, by assumption, $\mathbb{E}\Sigma_{m,n}^p \to 0$ as $m, n \to \infty$, we now obtain by taking unconditional expectations and letting m, n tend to infinity that

$$\lim_{m,n\to\infty} \mathbb{E}\left|\sum_{i=m}^{n} T_i(X_i - \mu)\right|^p \le b_p^p\, \mathbb{E}|X_1 - \mu|^p \lim_{m,n\to\infty} \mathbb{E}\Sigma_{m,n}^p = 0.$$

The remaining argument via (17) is identical to the one in the previous case and thus not repeated here.

(iv), (v) If $\mu = \mathbb{E}X_1 = 0$, the assumption in $\sum_{i\ge1} T_i \in L^p$ can be dropped because then the second term on the right-hand side in (17) vanishes.

The converse part:

(a) By choosing $X_i = 1$ for $i \ge 1$, we find that $\sum_{i\ge1} T_i \in L^p$ and are thus left with a proof of $\sum_{i\ge1} T_i^2 \in L^{p/2}$. Let now $X_1, X_2, \ldots$ be iid random variables taking values ±1 with probability $1/2$ each. Then $\mathbb{E}X_1 = 0$, $X_1 \in L^p$ for any $p > 1$, and $(\sum_{i=1}^n T_i X_i)_{n\ge0}$ conditioned on T forms a L^p-bounded martingale. By another appeal to Burkholder's inequality (49) (lower bound) and observing $X_1^2 = 1$, it follows that

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} T_i X_i\right|^p \,\middle|\, T\right) \ge a_p^p \left(\sum_{i=1}^{n} T_i^2\right)^{p/2} \quad \text{a.s.}$$

for a constant $a_p \in \mathbb{R}_>$ which only depends on p. Consequently,

$$\mathbb{E}\left(\sum_{i\ge1} T_i^2\right)^{p/2} \le \frac{1}{a_p^p}\, \mathbb{E}\left|\sum_{i\ge1} T_i X_i\right|^p < \infty$$

which proves the remaining assertion.

(b) Here it suffices to refer to the last argument. □

In the following, we say that *the smoothing transform $\mathcal{S}$ exists in L^p-sense* if $\mathcal{S}$ is a self-map on $\mathcal{P}^p(\mathbb{R})$. As a direct consequence of Proposition 4, one can easily deduce:

Corollary 1. *The smoothing transform $\mathcal{S}$ exists*

- *In L^p-sense for $0 < p \le 1$ if $C \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *In L^p-sense for $1 < p < 2$ if C, $\sum_{i\ge1} T_i \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *From $\mathcal{P}_0^p(\mathbb{R}) \to \mathcal{P}^p(\mathbb{R})$ for $1 < p < 2$ if $C \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *From $\mathcal{P}_0^p(\mathbb{R}) \to \mathcal{P}_0^p(\mathbb{R})$ for $1 < p \le 2$ if $C \in L_0^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$.*
- *In L^p-sense for $2 \le p < \infty$ iff $C, \sum_{i\ge1} T_i \in L^p$ and $\sum_{i\ge1} T_i^2 \in L^{p/2}$.*

- *From* $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ *for* $2 \le p < \infty$ *iff* $C \in L_0^p$ *and* $\sum_{i\ge1} T_i^2 \in L^{p/2}$.
- *From* $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ *for* $2 \le p < \infty$ *iff* $C \in L^p$ *and* $\sum_{i\ge1} T_i^2 \in L^{p/2}$.

Conversely, if $\mathscr{S}$ *exists*

- *In* L^p*-sense for* $1 < p < 2$*, then* $C, \sum_{i\ge1} T_i \in L^p$ *and* $\sum_{i\ge1} T_i^2 \in L^{p/2}$.
- *From* $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ *for* $1 < p < 2$*, then* $C \in L_0^p$ *and* $\sum_{i\ge1} T_i^2 \in L^{p/2}$.

In the particularly important case when $T_1, T_2, \ldots$ are nonnegative, a necessary and sufficient condition for $\mathscr{S}$ to exist in L^p-sense can be given for *all* $p > 0$ and follows directly from the previous result if $p > 0$.

Corollary 2. *Let* $T_1, T_2, \ldots$ *be nonnegative and* $0 < p < \infty$*. Then the smoothing transform* $\mathscr{S}$ *exists in* L^p*-sense iff* $C, \sum_{i\ge1} T_i \in L^p$.

Proof. We must only consider the case $0 < p \le 1$ and verify that $C, \sum_{i\ge1} T_i \in L^p$ is necessary for $\mathscr{S}$ to exist in L^p-sense. But choosing $X_i = 0$, we find $C \in L^p$, while choosing $X_i = 1$ for all $i \ge 1$ then further implies $\sum_{i\ge1} T_i \in L^p$. □

5 Convergence of Iterated Mean Values

By Theorem 15 in the Appendix, the convergence of $\mathscr{S}^n(F)$ to a fixed point in $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ follows if $\mathscr{S}$ is a continuous locally contractive self-map of this space, thus

$$\ell_p(\mathscr{S}^{n+1}(F), \mathscr{S}^n(F)) \;\le\; c\,\alpha^n \tag{18}$$

for suitable $c \ge 0, \alpha \in [0,1)$ and all $n \ge 0$. In order to infer uniqueness of the fixed point, one may consider expected values if $p \ge 1$, which provides the motivation behind the subsequent lemma (see [40, Lemma 1]). Recall that $\mathbb{E}F := \int x\,F(dx)$ for a distribution $F \in \mathscr{P}^1(\mathbb{R})$.

Lemma 2. *Suppose that* $\mathscr{S}$ *exists in* L^p*-sense for some* $p \ge 1$ *and let* $F \in \mathscr{P}^p(\mathbb{R})$*. Then*

(a) $\mathbb{E}(\sum_{i\ge1} T_i) \in (-1,1)$ *implies*

$$\lim_{n\to\infty} \mathbb{E}\mathscr{S}^n(F) \;=\; \frac{\mathbb{E}C}{1-\mathbb{E}(\sum_{i\ge1} T_i)},$$

and the convergence rate is geometric.

(b) $|\mathbb{E}(\sum_{i\ge1} T_i)| > 1$ *and* $\mathbb{E}F + (\mathbb{E}(\sum_{i\ge1} T_i) - 1)^{-1}\,\mathbb{E}C \ne 0$ *imply*

$$\lim_{n\to\infty} |\mathbb{E}\mathscr{S}^n(F)| \;=\; \infty.$$

(c) $|\mathbb{E}(\sum_{i\ge1} T_i)| > 1$ *and* $\mathbb{E}F + (\mathbb{E}(\sum_{i\ge1} T_i) - 1)^{-1}\,\mathbb{E}C = 0$ *imply*

$$\lim_{n\to\infty} \mathbb{E}\mathscr{S}^n(F) \;=\; \mathbb{E}F \;=\; \frac{\mathbb{E}C}{1-\mathbb{E}(\sum_{i\ge1} T_i)}.$$

(d) $\mathbb{E}(\sum_{i\ge1} T_i) = 1$ *and* $\mathbb{E}C \neq 0$ *imply*

$$\lim_{n\to\infty} |\mathbb{E}\mathscr{S}^n(F)| \;=\; \infty.$$

(e) $\mathbb{E}(\sum_{i\ge1} T_i) = 1$ *and* $\mathbb{E}C = 0$ *imply* $\mathbb{E}\mathscr{S}^n(F) = \mathbb{E}F$ *for all* $n \ge 0$.
(f) $\mathbb{E}(\sum_{i\ge1} T_i) = -1$ *implies*

$$\mathbb{E}S^{2n}(F) = \mathbb{E}F \quad \textit{and} \quad \mathbb{E}\mathscr{S}^{2n+1} = \mathbb{E}C - \mathbb{E}F$$

for all $n \ge 0$.

Proof. Fix any $n \ge 1$ and let $(C, T), X_1, X_2, \ldots$ be independent such that $\mathscr{L}(X_i) = \mathscr{S}^{n-1}(F)$ for each $i \ge 1$. Since $\sum_{i\ge1} T_i \in L^1$ by Corollary 1, we infer upon setting $\beta := \mathbb{E}(\sum_{i\ge1} T_i)$ that

$$\mathbb{E}\mathscr{S}^n(F) \;=\; \mathbb{E}C + \mathbb{E}\left(\sum_{i\ge1} T_i X_i\right) \;=\; \mathbb{E}C + \beta\,\mathbb{E}X_1 \;=\; \mathbb{E}C + \beta\,\mathbb{E}\mathscr{S}^{n-1}(F) \tag{19}$$

and then inductively

$$\mathbb{E}\mathscr{S}^n(F) \;=\; \mathbb{E}C \sum_{k=0}^{n-1} \beta^k + \beta^n \mathbb{E}F.$$

All assertions are easily derived from this equation. □

6 Contraction Results for $\mathscr{S}$

In view of the results in Sect. 4, Banach's fixed-point theorem (see the Appendix for a statement of this result along with some generalizations) ensures existence and uniqueness of a fixed point of $\mathscr{S}$ on any of

- $\mathscr{P}^p(\mathbb{R})$ for $p > 0$,
- $\mathscr{P}^p_0(\mathbb{R})$ (a closed subset of $\mathscr{P}^p(\mathbb{R})$) for $p \ge 1$,
- $\mathscr{P}^p_{0,1}(\mathbb{R})$ (a closed subset of $\mathscr{P}^p(\mathbb{R})$) for $p \ge 2$,
- ℓ^p-neighborhoods of a fixed distribution $F \in \mathscr{P}(\mathbb{R})$,
- $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for $p = m + \alpha > 1$ ($m \in \mathbb{N}, \alpha \in (0,1]$) and $\mathbf{z} \in \mathbb{R}^m$,

provided that $\mathscr{S}$ is contractive there with respect to ℓ_p (or ζ_p in the last case).

Conditions on (C, T) for this to happen will now be presented in a systematic way. Section 6.1 provides a condition on T, different for the cases $0 < p \le 1$ and $p > 1$, under which $\mathscr{S}$ is a contraction on $\mathscr{P}^p(\mathbb{R})$ for $p > 0$ (besides the canonical assumption $C \in L^p$). Situations when $\mathscr{S}$ is still a quasi-contraction on $\mathscr{P}^p(\mathbb{R})$ or $\mathscr{P}^p_c(\mathbb{R})$ for $p > 1$ and $c \in \mathbb{R}$ are discussed in Sect. 6.2. An even weaker property, namely local contractive behavior of $\mathscr{S}$, which still entails existence and uniqueness of a geometrically attracting fixed point, is studied for the case $p > 2$ in Sect. 6.3. All results presented this far are based on the minimal L^p-metric and mainly based on [40]. In Sect. 6.4, ℓ^p-neighborhoods of a fixed distribution $F \in \mathscr{P}(\mathbb{R})$ to be defined there are considered. Drawing on [42], we provide conditions ensuring contraction or quasi-contraction of $\mathscr{S}$ on such neighborhoods, an interesting feature being here that F does not need to be an element of $\mathscr{P}^p(\mathbb{R})$. Finally, Sect. 6.5 deals with the contractive behavior of $\mathscr{S}$ with respect to the Zolotarev metric ζ_p, $p > 1$, on subsets of $\mathscr{P}^p(\mathbb{R})$ with specified moments of integral order is shown under a simple condition on T. The contraction lemma used there is from [38, Proposition 1] (see also [35, Lemma 3.1] for an extension).

6.1 Contraction on $\mathscr{P}^p(\mathbb{R})$

Suppose first that $0 < p \le 1$. Due to the fact that the function $x \mapsto x^p$ is then subadditive on $\mathbb{R}_\ge$, this case is the simplest one.

Theorem 1. *Let $0 < p \le 1$. If*

$$C \in L^p \quad \textit{and} \quad \mathfrak{m}(p) < 1,$$

then $\mathscr{S}$ defines a contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Proof. By virtue of the subsequent lemma, $\mathscr{S}$ forms an $\mathfrak{m}(p)$-contraction. Hence, the assertions follow from Banach's fixed-point theorem (Theorem 13 in the Appendix) in combination with (20). □

Lemma 3. *Let $0 < p \le 1$, $C \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$. Then*

$$\ell_p(\mathscr{S}(F), \mathscr{S}(G)) \le \mathfrak{m}(p)\, \ell_p(F, G) \tag{20}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$.

Proof. Pick any $F, G \in \mathscr{P}^p(\mathbb{R})$ and let $(X_1, Y_1), (X_2, Y_2), \ldots$ be iid and (C, T)-independent random variables with $\mathscr{L}(X_1) = F$, $\mathscr{L}(Y_1) = G$ and $\|X_1 - Y_1\|_p = \ell_p(F, G)$. We note that $\mathscr{S}$ exists in L^p-sense by Corollary 1. Since $x \mapsto x^p$ is subadditive for $x \geq 0$ and $(\sum_{i\geq 1} T_i X_i + C, \sum_{i\geq 1} T_i Y_i + C) \sim (\mathscr{S}(F), \mathscr{S}(G))$, we infer

$$\ell_p(\mathscr{S}(F), \mathscr{S}(G)) \leq \left\| \sum_{i\geq 1} T_i X_i - \sum_{i\geq 1} T_i Y_i \right\|_p = \mathbb{E}\left| \sum_{i\geq 1} T_i (X_i - Y_i) \right|^p$$
$$\leq \|X_1 - Y_1\|_p \, \mathbb{E}\left(\sum_{i\geq 1} |T_i|^p \right) = \mathfrak{m}(p)\, \ell_p(F, G),$$

which is the assertion. □

Turning to the case $p > 1$, the result corresponding to Theorem 1 is due to Rösler [40, Theorem 8] (for the case $p = 2$, see also [38, Proposition 3]).

Theorem 2. *Let $p \geq 1$. If*

$$C \in L^p \quad \textit{and} \quad \left\| \sum_{i\geq 1} |T_i| \right\|_p < 1,$$

then $\mathscr{S}$ is a contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point in this space.

Since $\mathfrak{m}(p) \leq \| \sum_{i\geq 1} |T_i| \|_p$ for $p \geq 1$, we see that in general it takes a stronger condition for contraction of $\mathscr{S}$ than in the case $0 < p \leq 1$.

Proof. Pick any $F, G \in \mathscr{P}^p(\mathbb{R})$ and then as usual iid and (C, T)-independent random variables $(X_1, Y_1), (X_2, Y_2), \ldots$ such that $(X_1, Y_1) \sim (F, G)$ and $\|X_1 - Y_1\|_p = \ell_p(F, G)$. Setting $\Sigma_n := \sum_{i=1}^n |T_i|$, it follows by a similar argument as in the proof of Proposition 4(iii) that

$$\mathbb{E}\left(\left(\sum_{i=1}^n |T_i (X_i - Y_i)| \right)^p \middle| T \right) \leq \Sigma_n^p \, \mathbb{E}|X_1 - Y_1|^p = \Sigma_n^p \, \ell_p^p(F, G) \quad \text{a.s.}$$

for all $n \geq 1$ and therefore upon taking expectations, letting $n \to \infty$ and using the monotone convergence theorem

$$\ell_p(\mathscr{S}(F), \mathscr{S}(G)) \leq \left\| \sum_{i\geq 1} |T_i (X_i - Y_i)| \right\|_p \leq \left\| \sum_{i\geq 1} |T_i| \right\|_p \ell_p(F, G).$$

which proves that $\mathscr{S}$ is a contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and thus possesses a unique geometrically attracting fixed point in this set by Banach's fixed-point theorem. □

6.2 Conditions for Quasi-contraction if $p > 1$

Having settled the case $0 < p \le 1$ with just one condition, viz. $\mathfrak{m}(p) < 1$, giving contraction of $\mathscr{S}$ and a unique fixed point on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$, the case $1 < p < \infty$ exhibits a more complex picture as shown by three subsequent theorems, which for $p = 2$ are all from [40]. The afore-mentioned contraction condition, which figured in the previous subsection, is now replaced with

$$\mathscr{C}_p(T) \ := \ \mathfrak{m}(p) \vee \mathbb{E}\left(\sum_{i\ge 1} T_i^2\right)^{p/2} \tag{21}$$

which is still $\mathfrak{m}(p)$ if $1 < p \le 2$, but equals $\|\sum_{i\ge 1} T_i^2\|_{p/2}^{p/2}$ if $p \ge 2$. Plainly, the conditions collapse into one if $p = 2$.

Theorem 3. *Let $p > 1$. If*

$$C \in L_0^p \quad \textit{and} \quad \mathscr{C}_p(T) < 1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}_0^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Theorem 4. *Let $p > 1$. If*

$$C, \sum_{i\ge 1} T_i \in L^p, \quad \mathscr{C}_p(T) < 1 \quad \textit{and} \quad \left|\mathbb{E}\left(\sum_{i\ge 1} T_i\right)\right| < 1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Theorem 5. *Let $p > 1$ and $c \in \mathbb{R}$. If*

$$C \in L_0^p, \quad \sum_{i\ge 1} T_i \in L^p, \quad \mathscr{C}_p(T) < 1, \quad \textit{and} \quad \mathbb{E}\left(\sum_{i\ge 1} T_i\right) = 1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}_c^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_c in this space. Moreover, if even $\sum_{i\ge 1} T_i = 1$ a.s. holds true, then the G_c form a translation family, i.e. $G_c = \delta_c * G_0$ for all $c \in \mathbb{R}$.

We proceed to the statement of two contraction lemmata, treating the cases

- $p = 2$ and $\mathscr{C}_p(T) = \mathfrak{m}(p) = \|\sum_{i\ge 1} T_i^2\|_{p/2}^{p/2} < 1$.
- $p > 1$ and $\mathscr{C}_p(T) < 1$.

The proofs of the previous theorems require only the last of these lemmata, but we have included the other one because the provided contraction constant is better for $p = 2$. Recall that F^0 denotes the centering of F if $F \in \mathscr{P}^1(\mathbb{R})$.

Lemma 4. *Assuming $C \in L^2$ and $\sum_{i\geq 1} T_i^2 \in L^1$, the following assertions hold true:*

(a) $\mathscr{S}$ exists from $\mathscr{P}_0^2(\mathbb{R}) \to \mathscr{P}^2(\mathbb{R})$ and

$$\ell_2^2(\mathscr{S}(F^0), \mathscr{S}(G^0)) \leq \left\| \sum_{i\geq 1} T_i^2 \right\|_1 \ell_2^2(F^0, G^0) \tag{22}$$

for all $F, G \in \mathscr{P}^2(\mathbb{R})$.

(b) If also $\sum_{i\geq 1} T_i \in L^2$, then $\mathscr{S}$ exists in the L^2-sense and

$$\ell_2^2(\mathscr{S}(F), \mathscr{S}(G)) \leq \left\| \sum_{i\geq 1} T_i^2 \right\|_1 \ell_2^2(F^0, G^0) + \left\| \sum_{i\geq 1} T_i \right\|_2^2 \left(\mathbb{E}F - \mathbb{E}G\right)^2 \tag{23}$$

for all $F, G \in \mathscr{P}^2(\mathbb{R})$.

Proof. See [40, Lemma 2] □

The corresponding lemma for $p > 1$, which appears to be new to our best knowledge (however, see [38, Eq. (2.10)] for part (a) in the case $1 < p \leq 2$), is technically more difficult to prove because p^{th} powers of sums can be written out term-wise only for integral p.

Lemma 5. *Let $1 < p < \infty$, $C \in L^p$ and $\sum_{i\geq 1} |T_i|^p \in L^1$. Then the following assertions hold true:*

(a) $\mathscr{S}$ exists from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ and

$$\ell_p(\mathscr{S}^n(F^0), \mathscr{S}^n(G^0)) \leq b_p \mathscr{C}_p(T)^{n/p} \ell_p(F^0, G^0) \tag{24}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$ and $n \geq 1$.

(b) If also $\sum_{i\geq 1} T_i \in L^p$, then $\mathscr{S}$ exists in L^p-sense and

$$\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G))$$

$$\leq b_p \left[\mathscr{C}_p(T)^{n/p} \ell_p(F^0, G^0) + n\lambda_p \kappa_p^{n-1} \left|\mathbb{E}F - \mathbb{E}G\right|\right] \tag{25}$$

$$\leq b_p \left(\frac{n\lambda_p}{\kappa_p} + 2\right) \kappa_p^n \ell_p(F, G) \tag{26}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$ and $n \geq 1$, where

$$\kappa_p := \left| \mathbb{E}\left(\sum_{i\ge 1} T_i\right)\right| \vee \mathscr{C}_p(T)^{1/p}$$

$$\textit{and} \quad \lambda_p := \left\| \sum_{i\ge 1}(T_i - \mathbb{E}T_i)\right\|_p + b_p^{-1} \left\|\sum_{i\ge 1} T_i\right\|_p .$$

If $1 < p \le 2$, *we can choose* $b_p = 2^{1/p}$ *in both parts.*

Proof. The existence of $\mathscr{S}$ in the claimed sense is again guaranteed by Corollary 1.

(a) Given any $F, G \in \mathscr{P}^p(\mathbb{R})$, let $(X(\mathsf{v}), Y(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ be a family of iid random vectors which is independent of $\boldsymbol{C} \otimes \boldsymbol{T} = (C(\mathsf{v}), T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ (having the usual meaning) and satisfies $(X(\mathsf{v}), Y(\mathsf{v})) \sim (F^0, G^0)$ and $\|X(\mathsf{v}) - Y(\mathsf{v})\|_p = \ell_p(F^0, G^0)$. Consider two WBP $(Z_n')_{n\ge 0}$ and $(Z_n'')_{n\ge 0}$ associated with $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X} = (C(\mathsf{v}), T(\mathsf{v}), X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ and $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{Y}$, respectively, so that $\mathscr{L}(Z_n') = \mathscr{S}^n(F^0)$ and $\mathscr{L}(Z_n'') = \mathscr{S}^n(G^0)$ for each $n \ge 0$ (see Sect. 2). Furthermore,

$$Z_n := Z_n' - Z_n'' = \sum_{|\mathsf{v}|=n} L(\mathsf{v})(X(\mathsf{v}) - Y(\mathsf{v})), \quad n \ge 0$$

defines a WBP associated with $\boldsymbol{T}\otimes\boldsymbol{X}-\boldsymbol{Y} = (T(\mathsf{v}), X(\mathsf{v}) - Y(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ such that

$$\ell_p(\mathscr{S}^n(F^0), \mathscr{S}^n(G^0)) \le \|Z_n' - Z_n''\|_p = \|Z_n\|_p$$

for all $n \ge 0$, because $(Z_n', Z_n'') \sim (\mathscr{S}^n(F^0), \mathscr{S}^n(G^0))$. Write Z_n as

$$Z_n = L^p\text{-}\lim_{k\to\infty} \sum_{j=1}^{k} L(\mathsf{v}^j)(X(\mathsf{v}^j) - Y(\mathsf{v}^j))$$

for a suitable enumeration $\mathsf{v}^1, \mathsf{v}^2, \ldots$ of $\mathbb{N}^n$ and observe that, conditioned on $\boldsymbol{T}$, the right-hand sum forms an L^p-martingale in $k \ge 1$. As in the proof of Proposition 4, we must distinguish the cases $1 < p \le 2$ and $p \ge 2$ to complete our argument.

Case 1: $1 < p \le 2$. Then we infer with the help of the Topchiĭ-Vatutin inequality (44) in the Appendix that

$$\begin{aligned}
\mathbb{E}(|Z_n|^p|\boldsymbol{T}) &\le 2 \lim_{k\to\infty} \sum_{j=1}^{k} |L(\mathsf{v}^j)|^p\, \mathbb{E}|X(\mathsf{v}^j) - Y(\mathsf{v}^j)|^p \\
&= 2 \sum_{j\ge 1} |L(\mathsf{v}^j)|^p\, \mathbb{E}|X(\mathsf{v}^j) - Y(\mathsf{v}^j)|^p \\
&= 2\, \ell_p(F^0, G^0)^p \sum_{|\mathsf{v}|=n} |L(\mathsf{v})|^p \quad \text{a.s.}
\end{aligned}$$

One can easily verify that $\mathbb{E}(\sum_{|\mathsf{v}|=n}|L(\mathsf{v})|^p) = \|\sum_{i\geq 1}|T_i|^p\|_1^n$. Hence, we obtain (24) by taking unconditional expectation in the previous estimation.

Case 2: $p \geq 2$. Put $\Sigma_1^2 := \sum_{i\geq 1} T_i(\varnothing)^2$. By proceeding as in the proof of Proposition 4(iii), but with $X(i)-Y(i)$ instead of $X_i-\mu$ and $m=1, n=\infty$, it then follows by use of Burkholder's inequality and Jensen's inequality that

$$\begin{aligned}
\mathbb{E}&\left(\left|\sum_{i\geq 1} T_i(\varnothing)(X(i)-Y(i))\right|^p \middle| \boldsymbol{T}\right) \\
&\leq b_p^p\,\mathbb{E}\left(\left(\sum_{i\geq 1} T_i(\varnothing)^2(X(i)-Y(i))^2\right)^{p/2} \middle| \boldsymbol{T}\right) \\
&\leq b_p^p\,\Sigma_1^p\,\mathbb{E}\left(\left(\sum_{i\geq 1} \frac{T_i(\varnothing)^2}{\Sigma_1^2}(X(i)-Y(i))^2\right)^{p/2} \middle| \boldsymbol{T}\right) \\
&\leq b_p^p\,\Sigma_1^p\,\mathbb{E}|X(1)-Y(1)|^p \\
&\leq b_p^p\,\Sigma_1^p\,\ell_p^p(F^0,G^0) \quad \text{a.s.}
\end{aligned}$$

and thereby

$$\ell_p(\mathscr{S}(F^0),\mathscr{S}(G^0)) \leq \left\|\sum_{i\geq 1} T_i(X(i)-Y(i))\right\|_p \leq b_p\,\|\Sigma\|_p\,\ell_p(F^0,G^0),$$

where b_p only depends on p. This proves (24) for $n=1$. But in the same manner, we obtain for general n

$$\begin{aligned}
\ell_p(\mathscr{S}^n(F),\mathscr{S}^n(G)) &\leq \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(X(\mathsf{v})-Y(\mathsf{v}))\right\|_p \\
&\leq b_p\,\|\Sigma_n\|_p\,\ell_p(F^0,G^0),
\end{aligned}$$

where $\Sigma_n^2 := \sum_{|\mathsf{v}|=n} L(\mathsf{v})^2$. Hence, the proof of (24) will be complete once we have shown that

$$\|\Sigma_n\|_p \leq \|\Sigma\|_p^n \tag{27}$$

for all $n \geq 1$. To this end put $\Sigma(\mathsf{v}) := \sum_{i\geq 1} T_i(\mathsf{v})^2$ for $\mathsf{v} \in \mathbb{T}$ and recall from (5) that $\mathscr{F}_k = \sigma(T(\mathsf{v}) : |\mathsf{v}| \leq k-1)$ for $k \geq 1$. Then

$$\begin{aligned}\mathbb{E}(\Sigma_n^p|\mathscr{F}_{n-1}) &= \mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n-1} L(\mathsf{v})^2\Sigma(\mathsf{v})^2\right)^{p/2}\middle|\mathscr{F}_{n-1}\right)\\ &= \Sigma_{n-1}^p\,\mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n-1}\frac{L(\mathsf{v})^2}{\Sigma_{n-1}^2}\Sigma(\mathsf{v})^2\right)^{p/2}\middle|\mathscr{F}_{n-1}\right)\\ &\le \Sigma_{n-1}^p\,\mathbb{E}\left(\sum_{|\mathsf{v}|=n-1}\frac{L(\mathsf{v})^2}{\Sigma_{n-1}^2}\Sigma(\mathsf{v})^p\middle|\mathscr{F}_{n-1}\right)\\ &= \Sigma_{n-1}^p\,\|\Sigma\|_p^p \quad \text{a.s.}\end{aligned}$$

for each $n \ge 2$, which clearly gives (27) upon taking expectations and iteration.

(b) Let us first note that it suffices to show (25) because then (26) can be easily deduced with the help of (10) and the obvious inequality $|\mathbb{E}F - \mathbb{E}G| \le \ell_p(F,G)$, namely

$$\begin{aligned}&\mathscr{C}_p(T)^{n/p}\,\ell_p(F^0,G^0) + n\lambda_p\kappa_p^{n-1}\left|\mathbb{E}F-\mathbb{E}G\right|\\ &\qquad\le \mathscr{C}_p(T)^{n/p}\,\ell_p(F,G) + \left(\frac{n\lambda_p}{\kappa_p}+1\right)\kappa_p^n\left|\mathbb{E}F-\mathbb{E}G\right|\\ &\qquad\le \left(\frac{n\lambda_p}{\kappa_p}+2\right)\kappa_p^n\,\ell_p(F,G)\end{aligned}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$.

Similar to the proof of part (b) of the previous lemma, we obtain with the help of part (a) and Minkowski's inequality that

$$\begin{aligned}\ell_p(\mathscr{S}^n(F),\mathscr{S}^n(G)) &\le \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})\Big(\big(X(\mathsf{v})-Y(\mathsf{v})\big)+(\mathbb{E}F-\mathbb{E}G)\Big)\right\|_p\\ &\le \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})\big(X(\mathsf{v})-Y(\mathsf{v})\big)\right\|_p + |\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})\right\|_p\\ &= \|Z_n\|_p + |\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})\right\|_p\\ &\le b_p\,\mathscr{C}_p(T)^{n/p}\,\ell_p(F^0,G^0) + |\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})\right\|_p \qquad (28)\end{aligned}$$

for all $n \geq 1$, where b_p can be chosen as $2^{1/p}$ if $1 < p \leq 2$. This leaves us with the task to give an estimate for $a_n := \|\sum_{|\mathsf{v}|=n} L(\mathsf{v})\|_p$, which will be accomplished by another martingale argument involving the Topchiĭ-Vatutin inequality if $1 < p \leq 2$, and the Burkholder inequality if $p \geq 2$.

Case 1: $1 < p \leq 2$. We put $U(\mathsf{v}) := \sum_{i\geq 1} T_i(\mathsf{v})$, $\alpha := \mathscr{C}_p(T)^{1/p}$, $\beta := \mathbb{E}U(\mathsf{v})$ and $\gamma := \|U(\mathsf{v})-\beta\|_p = \|\sum_{i\geq 1} T_i - \beta\|_p$. Since $\sum_{i\geq 1} T_i \in L^p$ and $p > 1$, we have $|\beta| \leq a_1 < \infty$. By a similar argument as in (a), we see that $\sum_{|v|=n} L(\mathsf{v})(U(\mathsf{v}) - \beta)$ conditioned on $\mathscr{F}_n$ is the limit of an L^p-martingale (use that $U(\mathsf{v})$ is independent of $\mathscr{F}_n$), whence the Topchiĭ-Vatutin inequality yields

$$\mathbb{E}\left(\left|\sum_{|v|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right|^p \middle| \mathscr{F}_n\right) \leq 2\gamma^p \sum_{|\mathsf{v}|=n} |L(\mathsf{v})|^p \quad \text{a.s.}$$

As a consequence,

$$\begin{aligned} a_{n+1} &= \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})U(\mathsf{v})\right\|_p \\ &\leq \left\|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right\|_p + |\beta|a_n \\ &\leq 2^{1/p}\gamma \left\|\sum_{|\mathsf{v}|=n} |L(\mathsf{v})|^p\right\|_1^{1/p} + |\beta|a_n \\ &= 2^{1/p}\gamma\,\alpha^n + |\beta|a_n \end{aligned} \tag{29}$$

for all $n \geq 1$, which leads to

$$\begin{aligned} a_{n+1} &\leq 2^{1/p}\gamma \sum_{k=0}^{n-1} |\beta|^k \alpha^{n-k} + |\beta|^n a_1 \\ &\leq (n+1)(2^{1/p}\gamma + a_1)(|\beta| \vee \alpha)^n = (n+1)2^{1/p}\lambda_p \kappa_p^n \end{aligned} \tag{30}$$

for all $n \geq 1$. Since this inequality trivially holds for $n = 0$, we finally obtain the asserted inequality (25) from (28) and (30).

Case 2: $p \geq 2$. In this case, we obtain with the Burkholder inequality that

$$\begin{aligned} \mathbb{E}\left(\left|\sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta)\right|^p \middle| \mathscr{F}_n\right) &\leq b_p^p\, \mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n} L(\mathsf{v})^2(U(\mathsf{v})-\beta)^2\right)^{p/2} \middle| \mathscr{F}_n\right) \\ &\leq b_p^p \gamma^p \Sigma_n^p \quad \text{a.s.} \end{aligned}$$

which upon taking expectations on both sides and using (29) provides us with

$$\left\| \sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta) \right\|_p \le b_p\, \gamma\, \|\Sigma_1\|_p^n = b_p\, \gamma\, \alpha^n$$

and thus (see also (29))

$$a_{n+1} \le \left\| \sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta) \right\|_p + |\beta| a_n \le b_p\, \gamma\, \alpha^n + |\beta| a_n \tag{31}$$

for all $n \ge 1$. For the remaining arguments we can refer to the previous case. □

Now we can turn to the proofs of the theorems stated above.

Proof (of Theorem 3). As $\mathbb{E}C = 0$ is assumed, $\mathscr{S}$ defines a self-map of $\mathscr{P}_0^p(\mathbb{R})$ by Corollary 1. It is also an α-contraction on $(\mathscr{P}_0^p(\mathbb{R}), \ell_p)$ with $\alpha := \| \sum_{i\ge 1}(T_i)^2\|_{p/2}^{1/p}$ if $p = 2$ [by Lemma 4(a)], and an α_m-quasi-contraction with $\alpha_m := b_p\, \alpha^m$ for suitable $m \ge 1$ if $p > 1$ [by Lemma 5(a)]. Therefore, the assertion follows from Banach's fixed-point Theorem 13 or its generalization Theorem 14 in combination with the contraction inequality (22) or (24), respectively. □

Proof (of Theorem 4). The existence of $\mathscr{S}$ in L^p-sense follows again from Corollary 1, while contraction inequality (26) shows that $\mathscr{S}$ is a quasi-contraction on $\mathscr{P}^p(\mathbb{R})$, viz.

$$\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \le c\, \kappa^n\, \ell_p(F,G)$$

for any $\kappa \in (0, \kappa_p)$, $F, G \in \mathscr{P}^p(\mathbb{R})$, $n \ge 1$ and a suitable $c = c(\kappa) > 0$. All assertions now follow from Banach's fixed-point Theorem 14 for quasi-contractions. □

Proof (of Theorem 5). First note that $\mathbb{E}\,C = 0$ and $\mathbb{E}(\sum_{i\ge 1} T_i) = 1$ entail $\mathbb{E}\mathscr{S}(F) = \mathbb{E}\,F = c$ for all $F \in \mathscr{P}_c^p(\mathbb{R})$. Hence, $\mathscr{S}$ is a self-map of $\mathscr{P}_c^p(\mathbb{R})$ for any $c \in \mathbb{R}$. Moreover, (25) simplifies to

$$\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \le b_p \left\| \sum_{i\ge 1} T_i^2 \right\|_{p/2}^{n/2} \ell_p(F,G)$$

for all $n \ge 1$ and $F, G \in \mathscr{P}_c^p(\mathbb{R})$ because $\ell_p(F,G) = \ell_p(F^0, G^0)$. Hence $\mathscr{S}$ is also a quasi-contraction on $\mathscr{P}_c^p(\mathbb{R})$ and therefore has a unique fixed point G_c by Theorem 14. It remains to verify that $G_c = \delta_c * G_0$ in the case when $\sum_{i\ge 1} T_i = 1$

a.s. By the uniqueness property of G_c, it suffices to verify that $\mathscr{S}(\delta_c * G_0) = \delta_c * G_0$. Choose iid (C, T)-independent random variables $X_1, X_2, \ldots$ with law G_0. Then

$$\mathscr{S}(\delta_c * G_0) \;=\; \mathscr{L}\left(\sum_{i\geq 1} T_i(X_i + c) + C\right) \;=\; \mathscr{L}\left(\sum_{i\geq 1} T_i X_i + c + C\right)$$
$$\;=\; \delta_c * \mathscr{L}\left(\sum_{i\geq 1} T_i X_i + C\right) \;=\; \delta_c * \mathscr{S}(G_0) \;=\; \delta_c * G_0$$

yields the desired conclusion. □

6.3 Conditions for Local Contraction if $p > 2$

If $p > 2$ and $\sum_{i\geq 1} T_i \in L^p$ is replaced by the generally stronger condition $\sum_{i\geq 1} |T_i| \in L^p$, then we can trade in the contraction condition $\| \sum_{i\geq 1} T_i^2 \|_{p/2} < 1$ for a weaker one and still obtain local contraction in the sense that

$$\lim_{n\to\infty} \rho^{-n} \ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;=\; 0$$

for some $\rho \in (0, 1)$ and all $F, G \in \mathscr{P}^p(\mathbb{R})$ or $\in \mathscr{P}_0^p(\mathbb{R})$. As a consequence, existence and uniqueness of a geometrically attractive fixed point in these sets still holds. For integral $p > 2$, the following two theorems are again due to Rösler [40, Theorems 9 and 10]. Note that $\mathfrak{m}(q) \vee \mathfrak{m}(p) < 1$ for $0 < q < p < \infty$ implies $\mathfrak{m}(r) < 1$ for any $r \in [q, p]$ because $\mathfrak{m}$ is convex on $[2, p]$.

Theorem 6. *Let $p > 2$. If*

$$C \in L_0^p, \quad \sum_{i\geq 1} |T_i| \in L^p \quad \textit{and} \quad \mathfrak{m}(2) \vee \mathfrak{m}(p) < 1,$$

then $\mathscr{S}$ is a self-map of $\mathscr{P}_0^p(\mathbb{R})$ with a unique geometrically ℓ_p-attracting fixed point G_0 in this set.

Theorem 7. *Let $p > 2$. If*

$$C, \sum_{i\geq 1} |T_i| \in L^p, \quad \mathfrak{m}(2) \vee \mathfrak{m}(p) < 1 \quad \textit{and} \quad \left| \mathbb{E}\left(\sum_{i\geq 1} T_i\right) \right| < 1,$$

then $\mathscr{S}$ exists in L^p-sense and has a unique geometrically ℓ_p-attracting fixed point G_0 in $(\mathscr{P}^p(\mathbb{R}), \ell_p)$.

Proof (of Theorem 6). Here we will proceed in a different way than before and prove that $\mathscr{S}$ is locally contractive on $(\mathscr{P}_0^p(\mathbb{R}), \ell_p)$ in the sense of Theorem 15 (see (32) below). In particular, we will not make use of the Contraction Lemma 5. The first step is to show the result for integral $p > 2$ (as in [40]).

So let $2 < p \in \mathbb{N}$. We prove by induction that, for each $q \in \{1, \ldots, p\}$, there exists $\rho_q \in (0, 1)$ such that

$$\ell_q^q(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; c_q\, \rho_q^n \tag{32}$$

for all $F, G \in \mathscr{P}_0^p(\mathbb{R})$, $n \ge 1$ and a suitable $c_q \in \mathbb{R}_>$ which may depend on F, G. Observe that this corresponds to (42) when choosing $F = \mathscr{S}(G)$.

Hereafter, $K \in \mathbb{R}_>$ shall denote a generic constant which may differ from line to line but does not depend on n. Recall from above that $\mathfrak{m}(2) \vee \mathfrak{m}(p) < 1$ entails $\mathfrak{m}(q) < 1$ for all $q \in [2, p]$.

If $q = 1$ or $= 2$, we may invoke Lemma 4 to find

$$\ell_1^2(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; \ell_2^2(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; \mathfrak{m}(2)^n\, \ell_2^2(F, G)$$

for all $n \ge 1$ and $F, G \in \mathscr{P}_0^2(\mathbb{R})$, which clearly shows (32) in this case. We further see that $\mathscr{S}$ forms a contraction on $(\mathscr{P}_0^2(\mathbb{R}), \ell_2)$ and hence possesses a unique fixed point G_0 in this space. Since $\mathscr{P}_0^2(\mathbb{R}) \supset \mathscr{P}_0^p(\mathbb{R})$, it follows that G_0 is also the unique fixed point in $\mathscr{P}^p(\mathbb{R})$ once (32) has been verified for $q = p$.

For the inductive step suppose that (32) holds for any $r \in \{1, \ldots, q-1\}$ and let $(U_i)_{i \ge 1}$ be a sequence of iid *Unif*$(0, 1)$ random variables which are further independent of (C, T). Fixing any $F, G \in \mathscr{P}_0^q(\mathbb{R})$ throughout the rest of the proof, put

$$Y_{n,i} \;:=\; \mathscr{S}^n(F)^{-1}(U_i) - \mathscr{S}^n(G)^{-1}(U_i), \quad n \ge 1$$

and note that $\|Y_{n,i}\|_r = \ell_r(\mathscr{S}^n(F), \mathscr{S}^n(G))$ for all $i \ge 1, n \ge 0$ and $r \in [1, q]$. Since

$$\ell_q^q(\mathscr{S}^{n+1}(F), \mathscr{S}^{n+1}(G)) \;\le\; \mathbb{E}\left|\sum_{i \ge 1} T_i Y_{n,i}\right|^q \;\le\; \lim_{m \to \infty} \mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|\right)^q$$

we will further estimate the last expectation for arbitrary $m \in \mathbb{N}$ by making use of the multinomial formula which provides us with

$$\mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|\right)^q \;=\; \mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|^q\right) + \mathbb{E}\left(\sum_{\substack{0 \le r_1, \ldots, r_m < q, \\ r_1 + \ldots + r_m = q}} \frac{q!}{r_1! \cdot \ldots \cdot r_m!} \prod_{j=1}^m |T_j Y_{n,j}|^{r_j}\right).$$

The first term on the right-hand side obviously equals $\mathfrak{m}(q)\,\ell_q^q(\mathcal{S}^n(F),\mathcal{S}^n(G))$, while the second may be further computed as follows by conditioning upon T and using the fact that the $Y_{n,i}$ for any fixed n are iid:

$$\begin{aligned}
&\mathbb{E}\left(\sum_{\substack{0\le r_1,\dots,r_m<q,\\ r_1+\dots+r_m=q}} \frac{q!}{r_1!\cdot\ldots\cdot r_m!}\prod_{j=1}^m |T_jY_{n,j}|^{r_j}\right)\\
&\quad= \mathbb{E}\left(\sum_{\substack{0\le r_1,\dots,r_m<q,\\ r_1+\dots+r_m=q}} \frac{q!\,\mathbb{E}|Y_{n,1}|^{r_1}\cdot\ldots\cdot\mathbb{E}|Y_{n,1}|^{r_m}}{r_1!\cdot\ldots\cdot r_m!}\prod_{j=1}^m |T_j|^{r_j}\right)\\
&\quad= \left(\prod_{j=1}^m \ell_{r_j}^{r_j}(\mathcal{S}^n(F),\mathcal{S}^n(G))\right)\mathbb{E}\left(\sum_{\substack{0\le r_1,\dots,r_m<q,\\ r_1+\dots+r_m=q}} \frac{q!}{r_1!\cdot\ldots\cdot r_m!}\prod_{j=1}^m |T_j|^{r_j}\right)\\
&\quad\le K\,\rho^n\,\mathbb{E}\left(\sum_{i=1}^m |T_i|\right)^q
\end{aligned}$$

where the inductive hypothesis has been utilized to give the last estimate with $\rho := \max_{1\le s\le q-1}\rho_s$. The reader should notice that the constant K is not only independent of n but of m as well. Hence, by taking the limit $m\to\infty$, we find that

$$\ell_q^q(\mathcal{S}^{n+1}(F),\mathcal{S}^{n+1}(G)) \;\le\; \mathfrak{m}(q)\,\ell_q^q(\mathcal{S}^n(F),\mathcal{S}^n(G)) + K\,\rho^n$$

for all $n\ge 0$ and thereupon

$$\begin{aligned}
\ell_q^q(\mathcal{S}^{n+1}(F),\mathcal{S}^{n+1}(G)) \;&\le\; \mathfrak{m}(q)^{n+1}\,\ell_q^q(F,G) + K\sum_{k=1}^n \rho^k\mathfrak{m}(q)^{n-k}\\
&\le\; \Big(\ell_q^q(F,G) + Kn\Big)\,(\mathfrak{m}(q)\vee\rho)^{n+1}
\end{aligned}$$

for all $n\ge 0$ which implies (32) for any $\rho_q \in (\mathfrak{m}(q)\vee\rho, 1)$. By an appeal to Theorem 15, we conclude that, for any $F\in\mathcal{P}_0^p(\mathbb{R})$, $\mathcal{S}^n(F)$ converges to a fixed point in this set which must be unique by what has been stated above.

We turn to the second step which aims at an extension of the assertion to general $p>2$ with integer part $\hat{p}$, say. Let $r\in\mathbb{N}$ be such that $2^r<p\le 2^{r+1}$ and $s := p/2^{r+1}\in(0,1]$. From the first part of the proof, we know that (32) holds true for every $q\in\{1,\dots,\hat{p}\}$, and since $\ell_\alpha(\cdot,\cdot)$ is nondecreasing in α, this readily extends to all $q\in[1,\hat{p}\,]$. We will show hereafter that (32) is also true for $q=p$ (and thus for all $q\in[1,p]$) which finally proves the theorem in full generality.

Let us introduce the following operator Δ and its k-fold iterations Δ^k: For any nonnegative random variable W define

$$\Delta W := (W - \mathbb{E}W)^2, \quad \Delta^2 W = \Big((W - \mathbb{E}W)^2 - \mathbb{V}\mathrm{ar}\, W\Big)^2, \quad \text{etc.}$$

and $\Delta^0 W := W$. Naturally, $\Delta W = \infty$ is stipulated if $\mathbb{E}W = \infty$. We note that

$$\mathbb{E}\Delta^k W \le \mathbb{E}(\Delta^{k-1}W)^2 \le 2\,\mathbb{E}(\Delta^{k-2}W)^4 \le \ldots \le 2^{k-1}\mathbb{E}W^{2^k} \tag{33}$$

holds true for any $k \ge 1$.

By repeated use of the Burkholder inequality (49) (in the by now familiar manner after conditioning on T) and the subadditivity of $x \mapsto x^\alpha$ for $x \ge 0$ and $0 < \alpha \le 1$, we now obtain

$$\begin{aligned}
\Big\|\sum_{i\ge1} T_i Y_{n,i}\Big\|_p &\le K \Big\|\sum_{i\ge1} T_i^2 Y_{n,i}^2\Big\|_{p/2}^{1/2} \\
&\le K\left(\Big\|\sum_{i\ge1} T_i^2(Y_{n,i}^2 - \mathbb{E}Y_{n,i}^2)\Big\|_{p/2}^{1/2} + \big(\mathbb{E}Y_{n,1}^2\big)^{1/2}\Big\|\sum_{i\ge1} T_i^2\Big\|_{p/2}^{1/2}\right) \\
&\le K\left(\Big\|\sum_{i\ge1} T_i^4 \Delta Y_{n,i}^2\Big\|_{p/4}^{1/4} + \big(\mathbb{E}Y_{n,1}^2\big)^{1/2}\Big\|\sum_{i\ge1} |T_i|\Big\|_{p}^{1/2}\right) \\
&\ \ \vdots \\
&\le K\left(\Big\|\sum_{i\ge1} T_i^{2^{r+1}} \Delta^r Y_{n,i}^2\Big\|_{s}^{1/2^{r+1}} + \sum_{j=0}^{r-1}\big(\mathbb{E}\Delta^j Y_{n,1}^2\big)^{1/2^{j+1}}\Big\|\sum_{i\ge1} |T_i|\Big\|_{p}^{1/2}\right) \\
&\le K\left(\Big\|\sum_{i\ge1} |T_i|^p \big(\Delta^r Y_{n,i}^2\big)^s\Big\|_{1}^{1/2^{r+1}} + \sum_{j=0}^{r-1}\big(\mathbb{E}\Delta^j Y_{n,1}^2\big)^{1/2^{j+1}}\Big\|\sum_{i\ge1} |T_i|\Big\|_{p}^{1/2}\right) \\
&\le K\left(\|\Delta^r Y_{n,1}^2\|_s^{1/2^{r+1}}\Big\|\sum_{i\ge1} |T_i|^p\Big\|_{1}^{1/2^{r+1}} + \sum_{j=0}^{r-1}\big(\mathbb{E}\Delta^j Y_{n,1}^2\big)^{1/2^{j+1}}\Big\|\sum_{i\ge1} |T_i|\Big\|_{p}^{1/2}\right)
\end{aligned}$$

for all $n \ge 1$. Use (33), the definition of $Y_{n,1}$, and (32) for $\hat{p}$ to infer

$$\begin{aligned}
\big(\mathbb{E}\Delta^j Y_{n,1}^2\big)^{1/2^{j+1}} &\le \Big(2^{j-1}\,\mathbb{E}Y_{n,1}^{2^{j+1}}\Big)^{1/2^{j+1}} \le 2\,\|Y_{n,1}\|_{2^{j+1}} \\
&\le 2\,\|Y_{n,1}\|_{\hat{p}} = 2\,\ell_{\hat{p}}(\mathscr{S}^n(F), \mathscr{S}^n(G)) \le 2\,c_{\hat{p}}\,\rho_{\hat{p}}^n
\end{aligned}$$

for any $j \in \{0, \ldots, r-1\}$ and $n \geq 0$. By combining this with $\| \sum_{i\geq 1} |T_i|^p \|_1 = \mathfrak{m}(p) < 1$, the above estimation finally provides us with

$$\ell_p(\mathscr{S}^{n+1}(F), \mathscr{S}^{n+1}(G)) \;\leq\; \left\| \sum_{i\geq 1} T_i Y_{n,i} \right\|_p \;\leq\; K\,\rho^{n+1}$$

for all $n \geq 0$ and a suitable $\rho \in (0, 1)$. □

Proof (of Theorem 7). We are now in a more comfortable situation because the bulk of work has already been carried out in the previous proof. First note that all assumptions of Theorem 4 with $p = 2$ are fulfilled which allows us to infer the existence of a unique fixed point $G_0 \in \mathscr{P}^2(\mathbb{R})$. By Lemma 2(a), its mean value equals $c := \mathbb{E}G_0 = (1-\beta)^{-1}\mathbb{E}C$ with $\beta := \mathbb{E}(\sum_{i\geq 1} T_i)$. One can easily check that, if $F \in \mathscr{P}^p_c(\mathbb{R})$, then $\mathbb{E}S^n(F) = c$ for all $n \geq 0$ and that this further implies $\mathscr{S}^n(F)^c = \mathscr{S}^n(F^c)$ (recall that $F^c = F^0(\cdot - c)$) and thereupon

$$\begin{aligned} \ell_p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) \;&=\; \ell_p(\mathscr{S}^{n+1}(F)^c, \mathscr{S}^n(F)^c) \\ &=\; \ell_p(\mathscr{S}^{n+1}(F)^0, \mathscr{S}^n(F)^0) \end{aligned} \tag{34}$$

for all $F \in \mathscr{P}^p(\mathbb{R})$ and $n \geq 0$.

Now fix any $F \in \mathscr{P}^p(\mathbb{R})$, define $Y_{n,i}$ as in the previous proof, but for the pair $(\mathscr{S}(F^c), F^c)$. Then (32) for $q = p$ can be shown as in the previous proof, giving

$$\ell_p^p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) \;\leq\; \left\| \sum_{i\geq 1} T_i Y_{n,i} \right\|_p^p \;\leq\; c_p \rho_p^n$$

for all $n \geq 0$ and suitable constants $c_p \in \mathbb{R}_>$ and $\rho_p \in (0, 1)$. Note further that

$$\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F) \;=\; \beta^n \;(\mathbb{E}\mathscr{S}(F) - \mathbb{E}F)$$

for all $n \geq 0$, as has been shown in the proof of Lemma 2 (see (19)). By combining these facts with (10) and (34), we finally obtain

$$\begin{aligned} &\ell_p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) \\ &\quad\leq\; \ell_p(\mathscr{S}^{n+1}(F)^0, \mathscr{S}^n(F)^0) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\ &\quad=\; \ell_p(\mathscr{S}^{n+1}(F)^0, \mathscr{S}^n(F)^0) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\ &\quad=\; \ell_p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\ &\quad\leq\; c_p^{1/p} \rho_p^{n/p} + \beta^n \;|\mathbb{E}\mathscr{S}(F) - \mathbb{E}F| \end{aligned}$$

for all $n \geq 0$, that is geometric contraction of every iteration sequence in $\mathscr{P}^p(\mathbb{R})$. By invoking Theorem 15, we conclude that G_0 is the unique geometrically ℓ_p-attracting fixed point in this set. □

6.4 Contraction on ℓ_p-Neighborhoods of Fixed Distributions

A somewhat different approach than before is taken by Rüschendorf [42] who provides conditions for contraction of $\mathscr{S}$ in ℓ_p-neighborhoods of a fixed distribution $F \in \mathscr{P}(\mathbb{R})$, namely

$$\mathscr{U}^p(F) := \{G \in \mathscr{P}(\mathbb{R}) : \ell_p(F, G) < \infty\}$$

for $p > 0$, and

$$\mathscr{U}_c^p(F) := \{G \in \mathscr{P}_c^1(\mathbb{R}) : \ell_p(F, G) < \infty\}$$

for $p \geq 1$ and $c \in \mathbb{R}$. He embarks on the observation that, for $\ell_p(F, G)$ to be finite, it only takes to find an (F, G)-coupling (X, Y) such that $X - Y \in L^p$ but *not* that X, Y are themselves in L^p. Of course, if $F \in \mathscr{P}^p(\mathbb{R})$, then $\mathscr{U}^p(F) = \mathscr{P}^p(\mathbb{R})$. Besides the contraction condition $\mathscr{C}_p(T) = \mathfrak{m}(p) < 1$, familiar from previous results, he requires a *bounded jump-size condition*, namely

$$\ell_p(F, \mathscr{S}(F)) < \infty, \tag{35}$$

which is quite common in the study of iterated function systems on complete separable metric spaces. In that context, F is an arbitrary reference point and $\mathscr{S}$ a generic copy of the iid random Lipschitz functions to be iterated, see e.g. [19, Theorem 3]. Here the condition serves to ensure that $\mathscr{S}$ is a self-map of $\mathscr{U}^p(F)$ as the following proposition shows.

Proposition 5. *Let $p > 0$ and $F \in \mathscr{P}(\mathbb{R})$ be such that (35) holds true. Then $\mathscr{S}$ defines a self-map of $\mathscr{U}^p(F)$. Moreover, if $F \in \mathscr{P}^1(\mathbb{R})$, $C \in L^1$ and $p \geq 1$, then $\mathscr{S}$ defines a self-map of $\mathscr{U}_c^p(F)$ for any c such that $c = c\,\mathbb{E}(\sum_{i\geq 1} T_i) + \mathbb{E}C$, thus for all $c \in \mathbb{R}$ if $\kappa := \mathbb{E}(\sum_{i\geq 1} T_i) = 1$ and $\mathbb{E}C = 0$, and for $c = (1 - \mathbb{E}(\sum_{i\geq 1} T_i))^{-1}\mathbb{E}C$ if $\kappa \neq 1$.*

Proof. The following choices of random variables may take to enlarge the underlying probability space. Let (X, Y) be a $(F, \mathscr{S}(F))$-coupling such that $\ell_p(F, \mathscr{S}(F)) = \|X - Y\|_p$. Then pick iid copies $X_1, X_2, \ldots$ of X which are further independent of (C, T) and put $Y' := \sum_{i\geq 1} T_i X_i + C$. Finally, let X' be such that the conditional law of X' given $Y' = y$ is the same as the conditional law of X given $Y = y$ for all $y \in \mathbb{R}$, thus (X', Y') is a copy of (X, Y). Now, if $G \in \mathscr{U}^p(\mathbb{R})$, we can choose the X_i along with iid Z_i, independent of (C, T) and with common distribution G, such that the (X_i, Z_i) are iid as well and $\mu := \|X_i - Z_i\|_p < \infty$. It follows that

$$\begin{aligned}\ell_p(F, \mathscr{S}(G)) &\le \ell_p(F, \mathscr{S}(F)) + \ell_p(\mathscr{S}(F), \mathscr{S}(G)) \\ &\le \ell_p(F, \mathscr{S}(F)) + \left\| \sum_{i \ge 1} T_i (X_i - Z_i) \right\|_p \\ &= \ell_p(F, \mathscr{S}(F)) + \left\| \sum_{i \ge 1} T_i \right\|_p \mu < \infty\end{aligned}$$

and therefore that $\mathscr{S}$ is a self-map of $\mathscr{U}^p(F)$. The second assertion follows in a similar manner. □

The following results, containing those for $0 < p \le 2$ and N a fixed integer stated in [42], are the "local" counterparts of Theorems 1 and 3–5 and proved in the same way once having observed that Contraction Lemma 3 remains valid for $F, G \in \mathscr{U}_c^p(F_0)$ with $F_0 \in \mathscr{P}(\mathbb{R})$, and the Contraction Lemmata 4, and 5 remain valid for $F, G \in \mathscr{U}_c^p(F_0)$ with $F_0 \in \mathscr{P}^1(\mathbb{R})$ and $c \in \mathbb{R}$ (see also [42, Lemma 2.1]). We therefore refrain from giving proofs again.

Theorem 8. *If $0 < p \le 1$ and $\mathfrak{m}(p) < 1$, and if $F \in \mathscr{P}(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a contraction on $(\mathscr{U}^p(F), \ell_p)$ with a unique geometrically attracting fixed point.*

Theorem 9. *If $p > 1$, $C \in L_0^1$ and $\mathscr{C}_p(T) < 1$, and if $F \in \mathscr{P}^1(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a quasi-contraction on $(\mathscr{U}_0^p(F), \ell_p)$ with a unique geometrically attracting fixed point.*

Theorem 10. *If $p > 1$, $C \in L^1$, $\sum_{i \ge 1} T_i \in L^p$, $\mathscr{C}_p(T) < 1$ and $|\mathbb{E}(\sum_{i \ge 1} T_i)| < 1$, and if $F \in \mathscr{P}^1(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a quasi-contraction on $(\mathscr{U}^p(F), \ell_p)$ with a unique geometrically attracting fixed point.*

Theorem 11. *If $p > 1$, $C \in L_0^1$, $\sum_{i \ge 1} T_i \in L^p$, $\mathscr{C}_p(T) < 1$ and $\mathbb{E}(\sum_{i \ge 1} T_i) = 1$, and if $F \in \mathscr{P}^1(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a quasi-contraction on $(\mathscr{U}_c^p(F), \ell_p)$ with a unique geometrically attracting fixed point for any $c \in \mathbb{R}$.*

If $1 < p \le 2$, then $\mathscr{C}_p(T) = \mathfrak{m}(p)$ should be recalled. Moreover, if N is a fixed integer, then $\sum_{i \ge 1} T_i = \sum_{i=1}^N T_i \in L^p$ follows from $\mathfrak{m}(p) < 1$. With these observations, one can readily check that the results in [42] are really contained in the ones stated before.

Validity of the bounded jump-size condition (35) is usually difficult to check. In fact, it trivially holds whenever F is fixed point of $\mathscr{S}$. Since, furthermore, $\mathscr{U}^p(F) = \mathscr{U}^p(G)$ for all $G \in \mathscr{U}^p$ as well as $\mathscr{U}_c^p(F) = \mathscr{U}_c^p(G)$ for all $G \in \mathscr{U}_c^p$, the previous results may also be interpreted as follows: Under the respective conditions on p and (C, T), condition (35) holds true for some F only if

F is in finite ℓ^p-distance to a fixed point of $\mathscr{S}$. In contrast to the results from the previous subsections, this fixed point and thus F do not need to be elements of L^p.

Let us finally note that Rüschendorf, as an interesting consequence of his results, provides conditions which entail a certain one-to-one correspondence between the fixed points of a nonhomogeneous smoothing transform $\mathscr{S}$ and its homogeneous counterpart $\mathscr{S}_0$ (same T, but $C = 0$), see [42, Theorem 3.1] for details.

6.5 Contraction on Subsets of $\mathscr{P}^p(\mathbb{R})$ with Specified Integral Moments ($p > 1$)

Let $p = m + \alpha > 1$ hereafter, where $m \in \mathbb{N}$ and $\alpha \in (0, 1]$, and assume that $\mathscr{S}$ exists in L^p-sense so that, by Corollary 1, $C, \sum_{i\ge1} T_i \in L^p$. This final subsection is devoted to situations when $\mathscr{S}$, while not necessarily an ℓ_p-(quasi-)contraction on $\mathscr{P}^p(\mathbb{R})$, turns out to be contractive with respect to the Zolotarev metric ζ_p on subsets with specified integral moments. Recall that $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for $\mathbf{z} = (z_1, \ldots, z_m) \in \mathbb{R}^m$ equals the set of distributions $F \in \mathscr{P}^p(\mathbb{R})$ such that $\int x^k \, F(dx) = z_k$ for $k = 1, \ldots, m$.

In order for $\mathscr{S}$ to be a self-map of $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$, we must have that, given any iid $X_1, X_2, \ldots$ with law in $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$,

$$
\begin{aligned}
z_k &= \mathbb{E}\left(\sum_{i\ge1} T_i X_i + C\right)^k \\
&= \sum_{j_0+j_1+\ldots=k} \frac{k!}{\prod_{i\ge0} j_i!} \left(\prod_{i\ge1} z_{j_i}\right) \mathbb{E}\left(C^{j_0} \prod_{i\ge1} T_i^{j_i}\right) \\
&= z_k\, \mathbb{E}\left(\sum_{i\ge1} T_i^k\right) + \mathbb{E}C^k + \sum_{\substack{j_0+j_1+\ldots=k \\ j_0\vee j_1\vee\ldots<k}} \frac{k!}{\prod_{i\ge0} j_i!} \left(\prod_{i\ge1} z_{j_i}\right) \mathbb{E}\left(C^{j_0} \prod_{i\ge1} T_i^{j_i}\right)
\end{aligned}
$$

for $k = 1, \ldots, m$, because $X_1, X_2, \ldots$ and (C, T) are independent. In other words, $\mathbf{z}$ must satisfy a – for $m \ge 2$ nonlinear – system of equations, and one can easily see that this system may have a unique solution as well as infinitely many.

Theorem 12. *Suppose that $\mathfrak{m}(p) < 1$ and that $\mathscr{S}$ exists in L^p-sense. Then $\mathscr{S}$ is a ζ_p-contraction on $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for any $\mathbf{z} \in \mathbb{R}^m$ such that $\mathscr{S}$ is a self-map of $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$. In particular, it has a unique geometrically ζ_p-attracting fixed-point in this set.*

Proof. Since $(\mathscr{P}^p_{\mathbf{z}}(\mathbb{R}), \zeta_p)$ is a complete metric space (see Proposition 3), the result follows directly with the help of the Contraction Lemma 6 below and Banach's fixed-point theorem. □

Lemma 6. *Let* $(C,T) = (C,T_1,T_2,\ldots)$*,* $(X_n)_{n\ge 1}$ *and* $(Y_n)_{n\ge 1}$ *be independent sequences of real-valued random variables in* L^p *such that*

(A1) $X_1,X_2,\ldots$ *are independent with* $\mathscr{L}(X_n) = F_n$ *for* $n \ge 1$*.*
(A2) $Y_1,Y_2,\ldots$ *are independent with* $\mathscr{L}(Y_n) = G_n$ *for* $n \ge 1$*.*
(A3) For each $n \ge 1$*,* $F_n, G_n \in \mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ *for some* $\mathbf{z} \in \mathbb{R}^m$*.*
(A4) $\sum_{i\ge 1} T_i X_i + C,\ \sum_{i\ge 1} T_i Y_i + C \in L^p$*.*

Then

$$\zeta_s\left(\sum_{i\ge 1} T_i X_i + C, \sum_{i\ge 1} T_i Y_i + C\right) \le \sum_{i\ge 1} \mathbb{E}|T_i|^p\, \zeta_s(F_i,G_i). \tag{36}$$

In particular, if $\mathbf{z} \in \mathbb{R}^m$*, then*

$$\zeta_p(\mathscr{S}(F),\mathscr{S}(G)) \le \mathfrak{m}(p)\,\zeta_p(F,G). \tag{37}$$

for all $F,G \in \mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$*, whenever* $\mathscr{S}$*, the smoothing transform associated with* (C,T)*, exists in* L^p*-sense and is a self-map of* $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$*.*

Proof. First note that $\zeta_p(\sum_{i\ge 1} T_i X_i + C, \sum_{i\ge 1} T_i Y_i + C) < \infty$ because (A3) and (A4) ensure that $\sum_{i\ge 1} T_i X_i + C, \sum_{i\ge 1} T_i Y_i + C \in L^p_{\mathbf{z}}$. Denote by Λ the distribution of (C,T) and let $t = (t_1,t_2,\ldots)$ in the subsequent integration with respect to Λ. Then, by multiple use of properties (14) and (15) of ζ_p (in lines 5, 8 and 9), we infer for each $n \in \mathbb{N}$ that

$$\begin{aligned}
&\zeta_p\left(\sum_{i=1}^n T_i X_i + C, \sum_{i=1}^n T_i Y_i + C\right)\\
&= \sup_{f\in\mathfrak{F}_p}\left|\mathbb{E}\left(f\left(\sum_{i=1}^n T_i X_i + C\right) - f\left(\sum_{i=1}^n T_i Y_i + C\right)\right)\right|\\
&\le \int \sup_{f\in\mathfrak{F}_p}\left|\mathbb{E}\left(f\left(\sum_{i=1}^n t_i X_i + c\right) - f\left(\sum_{i=1}^n t_i Y_i + c\right)\right)\right| \Lambda(dc,dt)\\
&= \int \zeta_p\left(\sum_{i=1}^n t_i X_i + c, \sum_{i=1}^n t_i Y_i + c\right) \Lambda(dc,dt)\\
&\le \int \zeta_p\left(\sum_{i=1}^n t_i X_i, \sum_{i=1}^n t_i Y_i\right) \Lambda(dc,dt)\\
&\le \int \sum_{k=1}^n \zeta_p\left(\sum_{i=k}^n t_i X_i + \sum_{j=1}^{k-1} t_j Y_j, \sum_{i=k+1}^n t_i X_i + \sum_{j=1}^{k} t_j Y_j\right) \Lambda(dc,dt)
\end{aligned}$$

$$
\begin{aligned}
&= \int \sum_{k=1}^{n} \zeta_p\,(t_k X_k + S_k, t_k Y_k + S_k)\ \Lambda(dc, dt) \\
&\left[\text{where } S_k := \sum_{i=k+1}^{n} t_i X_i + \sum_{j=1}^{k-1} t_j Y_j \text{ and is independent of } X_k, Y_k\right] \\
&\le \int \sum_{k=1}^{n} \zeta_p\,(t_k X_k, t_k Y_k)\ \Lambda(dc, dt) \\
&= \int \sum_{k=1}^{n} |t_k|^p\, \zeta_p\,(X_k, Y_k)\ \Lambda(dc, dt) \\
&= \sum_{i=1}^{n} \mathbb{E}|T_i|^p\, \zeta_p(F_i, G_i)
\end{aligned}
$$

which proves (36) by letting n tend to infinity and using

$$
\lim_{n\to\infty} \zeta_p \left(\sum_{i=1}^{n} T_i X_i + C, \sum_{i=1}^{n} T_i Y_i + C \right) = \zeta_p \left(\sum_{i\ge 1} T_i X_i + C, \sum_{i\ge 1} T_i Y_i + C \right).
$$

The second inequality (37) follows from the first one when choosing $F_i = F$ and $G_i = G$ for all $i \ge 1$. □

7 Concluding Remarks

Having provided a comprehensive account of results describing the contractive behavior of the smoothing transform on $\mathscr{P}^p(\mathbb{R})$ or subsets thereof for $p > 0$, we would like to finish this review with some remarks on what has not been covered.

Naturally, other metrics than ℓ_p and ζ_p could have been studied as well. For instance, with $\hat{F}(t) := \int e^{itx}\, F(dx)$ denoting the Fourier transform of F, the Fourier metric

$$
r_p(F, G) := \int_0^\infty \frac{|\hat{F}(t) - \hat{G}(t)|}{t^{1+p}}\, dt, \qquad F, G \in \mathscr{P}_c^p(\mathbb{R})
$$

for $p \in (1, 2)$ was introduced and shown to be complete on $\mathscr{P}_c^p(\mathbb{R})$ by Baringhaus and Grübel [8, Lemma 2.1]. For homogeneous $\mathscr{S}$ with a.s. finite N, they further showed that it is a contraction on $(\mathscr{P}_c^p(\mathbb{R}), r_p)$ if $\mathfrak{m}(p) < 1$ and $\mathbb{E}(\sum_{i\ge 1} T_i) = 1$. The result was later extended by Iksanov [26, Proposition 6] to the case of general N (see also [8, Sect.]). As one can easily see, the result further extends to the nonhomogeneous case with $C \in L_0^1$.

Since contraction (with respect to ℓ_p or ζ_p) on subsets Γ of $\mathscr{P}^p(\mathbb{R})$ for some $p > 0$ particularly entails that, for some fixed point of $\mathscr{S}$, the set Γ is attracting with respect to weak convergence, one may ask about more general results describing such sets without moment assumptions, thus within $\mathscr{P}(\mathbb{R})$. As an example in this direction, we mention the following result obtained by Durrett and Liggett [18, Theorem 2(b)]: If $C = 0$, $T \geq 0$, N is a.s. bounded and T has characteristic exponent $\alpha \in (0, 1]$ (see at the end of Sect. 2), then, given any fixed point $F \in \mathscr{P}(\mathbb{R}_{\geq})$ of $\mathscr{S}$ with Laplace transform $\tilde{F}$, $\mathscr{S}^n(G)$ converges weakly to F whenever

$$\lim_{t \downarrow 0} \frac{1 - \tilde{F}(t)}{1 - \tilde{G}(t)} = 1.$$

An extension of their result under relaxed conditions on N appears in [31, Theorem 1.3]. Results of this type could also be formulated for the general smoothing transform and fixed points on the whole real line when substituting Fourier transforms for Laplace transforms. However, we refrain from supplying any further details.

8 Appendix

8.1 Banach's Fixed-Point Theorem

Let $f : \mathbb{X} \to \mathbb{X}$ be a continuous self-map of a metric space $(\mathbb{X}, \rho)$ and denote by $f^n = f \circ \ldots \circ f$ (n-times) its n-fold composition for $n \geq 1$. If there exists an initial value $x_0 \in \mathbb{X}$ such that the sequence $x_n := f(x_{n-1}) = f^n(x_0)$, $n \geq 1$, converges to some $x_\infty \in \mathbb{X}$, then the continuity of f implies that x_∞ is a fixed point of f, for

$$x_\infty = \lim_{n \to \infty} x_n = f\left(\lim_{n \to \infty} x_{n-1}\right) = f(x_\infty). \tag{38}$$

The map f is called a *contraction* or more specifically α-*contraction* if there exists $\alpha \in [0, 1)$ such that

$$\rho(f(x), f(y)) \leq \alpha\, \rho(x, y) \tag{39}$$

for all $x, y \in \mathbb{X}$. If (39) holds true when replacing f with f^n for some $n \geq 2$, then f is called *quasi-contraction* or α-*quasi-contraction*.

Under a contraction, the distance between two iteration sequences $(f^n(x))_{n\geq 1}$ and $(f^n(y))_{n\geq 1}$ is therefore decreasing geometrically fast, viz.

$$\rho(f^n(x), f^n(y)) \leq \alpha^n\, \rho(x, y)$$

for all $n \geq 1$. If the space $(\mathbb{X}, \rho)$ is complete, then this entails geometric convergence to a unique fixed point of f as the following classic result shows.

Theorem 13 (Banach's fixed-point theorem). *Every contraction $f : \mathbb{X} \to \mathbb{X}$ on a complete metric space $(\mathbb{X}, \rho)$ possesses a unique fixed point $\xi \in \mathbb{X}$. Moreover,*

$$\rho(\xi, f^n(x)) \ \leq \ \frac{\alpha^n}{1-\alpha}\, \rho(f(x), x) \tag{40}$$

holds true for all $x \in \mathbb{X}$ and $n \geq 1$, where α denotes the contraction parameter of f.

The next result shows that Banach's fixed-point theorem essentially remains valid for quasi-contractions.

Theorem 14 (Banach's fixed-point theorem for quasi-contractions). *Every quasi-contraction $f : \mathbb{X} \to \mathbb{X}$ on a complete metric space $(\mathbb{X}, \rho)$ possesses a unique fixed point $\xi \in \mathbb{X}$, and*

$$\rho(\xi, f^n(x)) \ \leq \ \frac{\alpha^n}{1-\alpha} \max_{0 \leq r < m} \rho(f^{m+r}(x), f^m(x)) \tag{41}$$

for some $m \geq 1$, $\alpha \in [0, 1)$ and all $x \in \mathbb{X}$, $n \geq 1$.

Proof. Pick m, α such that f^m forms an α-contraction on $(\mathbb{X}, \rho)$ with unique fixed point ξ. Writing $n \in \mathbb{N}$ in the form $km+r$ with unique $k \in \mathbb{N}_0$ and $r \in \{0, \ldots, m-1\}$, we infer with the help of (40)

$$\rho(\xi, f^n(x)) \ \leq \ \max_{0 \leq j < m} \rho(\xi, f^{km+j}(x)) \ \leq \ \frac{\alpha}{1-\alpha} \max_{0 \leq j < m} \rho(f^{m+j}(x), f^j(x))$$

and thus (41), in particular $\rho(\xi, f^n(x)) \to 0$. Since f is continuous, the latter implies that ξ is also the (necessarily unique) fixed point of f. ☐

Replacing the global by a local contraction property along an iteration sequence, existence of a fixed point still follows, but it needs no longer be unique.

Theorem 15. *Let $(\mathbb{X}, \rho)$ be a complete metric space and $f : \mathbb{X} \to \mathbb{X}$ an arbitrary self-map. Suppose there exist $x_0 \in \mathbb{X}$ and constants $c \geq 0$ and $\alpha \in [0, 1)$ such that*

$$\rho(f^{n+1}(x_0), f^n(x_0)) \ \leq \ c\alpha^n \tag{42}$$

for all $n \geq 1$. Then $\xi = \lim_{n\to\infty} f^n(x_0)$ exists and it is a fixed point of f if the map is continuous. Moreover,

$$\rho(\xi, f^n(x_0)) \ \leq \ \frac{c\alpha^n}{1-\alpha} \tag{43}$$

for all $n \geq 1$.

Proof. Putting $x_n := f^n(x_0)$ and using (42), we obtain

$$\rho(x_{m+n}, x_m) \;\le\; \sum_{k=m}^{m+n-1} \rho(x_{k+1}, x_k) \;\le\; \sum_{k=m}^{m+n-1} c\alpha^k \;\le\; \frac{c\alpha^m}{1-\alpha}$$

for all $m, n \ge 1$, that is, $(x_n)_{n\ge 0}$ is a Cauchy sequence in $\mathbb{X}$ and thus convergent to some $\xi \in \mathbb{X}$, for $(\mathbb{X}, \rho)$ is complete. If f is continuous, then $f(\xi) = \xi$ (see (38)). Finally, (43) follows from (42) when observing that

$$\rho(\xi, f^n(x_0)) \;=\; \rho(\xi, x_n) \;\le\; \sum_{k\ge n} \rho(x_k, x_{k+1}). \qquad \square$$

8.2 Convex Function Inequalities for Martingales and Their Maxima

Let $(M_n)_{n\ge 0}$ be a martingale with natural filtration $(\mathscr{F}_n)_{n\ge 0}$ and increments $D_n = M_n - M_{n-1}$ for $n \ge 1$. In the following, we list some powerful martingale inequalities that provide bounds for the *ϕ-moments* $\mathbb{E}\phi(M_n)$, when $\phi : \mathbb{R} \to \mathbb{R}_\ge$ denotes an even convex function with $\phi(0) = 0$ and some additional properties. This includes the standard class $\phi(x) = |x|^p$ for $p \ge 1$. Setting $M_\infty := \liminf_{n\to\infty} M_n$, all provided upper bounds remain valid for $n = \infty$ when observing that Fatou's lemma implies

$$\mathbb{E}\phi(M_\infty) \;\le\; \liminf_{n\to\infty} \mathbb{E}\phi(M_n).$$

We begin with the class of ϕ that have a concave derivative in $\mathbb{R}_>$ and thus encompasses $\phi(x) = |x|^p$ for $1 \le p \le 2$. The subsequent result is cited from [4] and an improvement (with regard to the appearing constant) of a version due to Topchiĭ and Vatutin [43]. In the more general framework of Banach spaces of a given type, the inequality (with a non-specified constant) is actually due to Woyczynski [45, Proposition 2.1].

Theorem 16 (Topchiĭ-Vatutin inequality). *Let $\phi : \mathbb{R} \to \mathbb{R}_\ge$ be an even convex function with concave derivative on $\mathbb{R}_>$ and $\phi(0) = 0$. Then*

$$\mathbb{E}\phi(M_n) - \mathbb{E}\phi(M_0) \;\le\; c \sum_{k=1}^{n} \mathbb{E}\phi(D_k), \tag{44}$$

for all $n \in \overline{\mathbb{N}}_0$ and $c = 2$. The constant may be chosen as $c = 1$ if $(M_n)_{n\ge 0}$ is nonnegative or has a.s. symmetric conditional increment distributions, and the same holds generally true, if $\phi(x) = |x|$ or $\phi(x) = x^2$, in the last case even with equality sign in (44)*.*

We continue with two famous convex function inequalities by Burkholder, Davis, and Gundy [13] which are valid for a much larger class of convex functions ϕ.

Theorem 17 (Burkholder-Davis-Gundy inequalities). *Let* $\phi:\mathbb{R}\to\mathbb{R}_{\geq}$ *be an even convex function satisfying* $\phi(0)=0$ *and* $\phi(2t)\leq\gamma\,\phi(t)$ *for all* $t\geq 0$ *and some* $\gamma>0$. *Put* $E_n(\phi):=\mathbb{E}\left(\max_{0\leq k\leq n}\phi(M_k)\right)$. *Then*

$$a_\gamma\,\mathbb{E}\phi\left(\left(\sum_{k=1}^n D_k^2\right)^{1/2}\right)\ \leq\ E(\phi)\ \leq\ b_\gamma\,\mathbb{E}\phi\left(\left(\sum_{k=1}^n D_k^2\right)^{1/2}\right) \tag{45}$$

and

$$E_n(\phi)\ \leq\ c_\gamma\left[\mathbb{E}\phi\left(\left(\sum_{k=1}^n \mathbb{E}\big(D_k^2|\mathscr{F}_{k-1}\big)\right)^{1/2}\right)+\mathbb{E}\left(\max_{0\leq k\leq n}\phi(D_k)\right)\right] \tag{46}$$

for all $n\in\overline{\mathbb{N}}_0$ *and constants* $a_\gamma,b_\gamma,c_\gamma\in\mathbb{R}_>$ *depending only on* γ. *The last inequality actually remains valid if, ceteris paribus,* ϕ *is merely nondecreasing instead of convex on* $\mathbb{R}_{\geq}$.

Of special importance in connection with the smoothing transform is the case when M_n is a weighted sum of iid zero-mean random variables and $\phi(x)=|x|^p$ for some $p>0$. We therefore note:

Corollary 3. *If* $\phi(x)=|x|^p$ *(thus* $\gamma=2^p$*) for some* $p>0$ *and* $M_n=\sum_{k=1}^n t_kX_k$ *for* $t_1,t_2,\ldots\in\mathbb{R}$ *and iid* $X_1,X_2,\ldots\in L_0^p$, *then* (46) *takes the form*

$$E_n(\phi)\ \leq\ c_p\left[\|X_1\|_2^p\left(\sum_{k=1}^n t_k^2\right)^{p/2}+\mathbb{E}\left(\max_{1\leq k\leq n}|t_kX_k|^p\right)\right], \tag{47}$$

for all $n\in\overline{\mathbb{N}}_0$ *and a constant* c_p *only depending on* p, *giving in particular*

$$\mathbb{E}|M_n|^p\ \leq\ c_p\left[\|X_1\|_2^p\left(\sum_{k=1}^n t_k^2\right)^{p/2}+\|X_1\|_p^p\sum_{k=1}^n|t_k|^p\right]. \tag{48}$$

Finally, we state the classical L^p-inequality by Burkholder [12], valid for $p>1$ only. The case $p=1$ is different but will not be considered here.

Theorem 18 (Burkholder inequality). *Let* $p>1$. *Then*

$$a_p\left\|\left(\sum_{k=1}^n D_k^2\right)^{1/2}\right\|_p\ \leq\ \|M_n\|_p\ \leq\ b_p\left\|\left(\sum_{k=1}^n D_k^2\right)^{1/2}\right\|_p \tag{49}$$

for $n \in \overline{\mathbb{N}}_0$ *and constants* $a_p, B_p \in \mathbb{R}_>$ *only depending on* p*. Admissible choices are* $a_p = (18p^{3/2}/(p-1))^{-1}$ *and* $b_p = 18p^{3/2}/(p-1)^{1/2}$ *(see [24, Theorem 2.10]).*

Acknowledgements G.A. was supported by Deutsche Forschungsgemeinschaft (SFB 878).

References

1. Aldous, D., Steele, J.M.: The objective method: probabilistic combinatorial optimization and local weak convergence. Probab. Discret. Struct. **110**, 1–72 (2004)
2. Alsmeyer, G., Meiners, M.: Fixed points of inhomogeneous smoothing transforms. J. Differ. Equ. Appl. **18**, 1287–1304 (2012)
3. Alsmeyer, G., Meiners, M.: Fixed points of the smoothing transform: two-sided solutions. Probab. Theory Relat. Fields **155**(1–2), 165–199 (2013)
4. Alsmeyer, G., Rösler, U.: The best constant in the Topchii-Vatutin inequality for martingales. Stat. Probab. Lett. **65**(3), 199–206 (2003)
5. Alsmeyer, G., Iksanov, A., Rösler, U.: On distributional properties of perpetuities. J. Theor. Probab. **22**(3), 666–682 (2009)
6. Alsmeyer, G., Biggins, J.D., Meiners, M.: The functional equation of the smoothing transform. Ann. Probab. **40**(5), 2069–2105 (2012)
7. Alsmeyer, G., Damek, E., Mentemeier, S.: Precise tail index of fixed points of the two-sided smoothing transform. In: Alsmeyer, G., Löwe, M. (eds.) Random Matrices and Iterated Random Functions. Springer Proceedings in Mathematics & Statistics, vol. 53. Springer, Heidelberg (2013)
8. Baringhaus, L., Grübel, R.: On a class of characterization problems for random convex combinations. Ann. Inst. Stat. Math. **49**(3), 555–567 (1997)
9. Biggins, J.D.: Martingale convergence in the branching random walk. J. Appl. Probab. **14**(1), 25–37 (1977)
10. Biggins, J.D., Kyprianou, A.E.: Seneta-Heyde norming in the branching random walk. Ann. Probab. **25**(1), 337–360 (1997)
11. Biggins, J.D., Kyprianou, A.E.: Fixed points of the smoothing transform: the boundary case. Electron. J. Probab. **10**, 609–631 (2005). (electronic)
12. Burkholder, D.L.: Martingale transforms. Ann. Math. Stat. **37**, 1494–1504 (1966)
13. Burkholder, D.L., Davis, B.J., Gundy, R.F.: Integral inequalities for convex functions of operators on martingales. In: Probability Theory, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, University of California, Berkeley, 1970/1971, vol. II, pp. 223–240. University of California Press, Berkeley (1972)
14. Caliebe, A.: Symmetric fixed points of a smoothing transformation. Adv. Appl. Probab. **35**(2), 377–394 (2003)
15. Caliebe, A., Rösler, U.: Fixed points with finite variance of a smoothing transformation. Stoch. Process. Appl. **107**(1), 105–129 (2003)
16. Collamore, J.F., Vidyashankar, A.N.: Large deviation tail estimates and related limit laws for stochastic fixed point equations. In: Alsmeyer, G., Löwe, M. (eds.) Random Matrices and Iterated Random Functions. Springer Proceedings in Mathematics & Statistics, vol. 53. Springer, Heidelberg (2013)
17. Diaconis, P., Freedman, D.: Iterated random functions. SIAM Rev. **41**(1), 45–76 (1999)
18. Durrett, R., Liggett, T.M.: Fixed points of the smoothing transformation. Z. Wahrsch. Verw. Gebiete **64**(3), 275–301 (1983)
19. Elton, J.H.: A multiplicative ergodic theorem for Lipschitz maps. Stoch. Process. Appl. **34**(1), 39–47 (1990)
20. Goldie, C.M., Maller, R.A.: Stability of perpetuities. Ann. Probab. **28**(3), 1195–1218 (2000)

21. Graf, S., Mauldin, R.D., Williams, S.C.: The exact Hausdorff dimension in random recursive constructions. Mem. Am. Math. Soc. **71**(381), x+121 (1988)
22. Grübel, R., Rösler, U.: Asymptotic distribution theory for Hoare's selection algorithm. Adv. Appl. Probab. **28**(1), 252–269 (1996)
23. Guivarc'h, Y.: Sur une extension de la notion de loi semi-stable. Ann. Inst. Henri Poincaré Probab. Stat. **26**(2), 261–285 (1990)
24. Hall, P., Heyde, C.C.: Martingale Limit Theory and Its Application. Probability and Mathematical Statistics. Academic [Harcourt Brace Jovanovich Publishers], New York (1980)
25. Hu, Y., Shi, Z.: Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. Ann. Probab. **37**(2), 742–789 (2009)
26. Iksanov, A.M.: Elementary fixed points of the BRW smoothing transforms with infinite number of summands. Stoch. Process. Appl. **114**(1), 27–50 (2004)
27. Jelenković, P.R., Olvera-Cravioto, M.: Implicit renewal theory and power tails on trees. Adv. Appl. Probab. **44**(2), 528–561 (2012)
28. Jelenković, P.R., Olvera-Cravioto, M.: Information ranking and power laws on trees. Adv. Appl. Probab. **42**(4), 1057–1093 (2010)
29. Jelenković, P.R., Olvera-Cravioto, M.: Implicit renewal theorem for trees with general weights. Stoch. Proc. Appl. **122**(9), 3209–3238 (2012)
30. Liu, Q.: The growth of an entire characteristic function and the tail probabilities of the limit of a tree martingale. In: Trees, Versailles, 1995. Progress in Probability, vol. 40, pp. 51–80. Birkhäuser, Basel (1996)
31. Liu, Q.: Fixed points of a generalized smoothing transformation and applications to the branching random walk. Adv. Appl. Probab. **30**(1), 85–112 (1998)
32. Liu, Q.: On generalized multiplicative cascades. Stoch. Process. Appl. **86**(2), 263–286 (2000)
33. Mauldin, R.D., Williams, S.C.: Random recursive constructions: asymptotic geometric and topological properties. Trans. Am. Math. Soc. **295**(1), 325–346 (1986)
34. Mirek, M.: Heavy tail phenomenon and convergence to stable laws for iterated Lipschitz maps. Probab. Theory Relat. Fields **151**(3–4), 705–734 (2011)
35. Neininger, R., Rüschendorf, L.: A general limit theorem for recursive algorithms and combinatorial structures. Ann. Appl. Probab. **14**(1), 378–418 (2004)
36. Neininger, R., Rüschendorf, L.: Analysis of algorithms by the contraction method: additive and max-recursive sequences. In: Deuschel, J.-D., Greven, A. (eds.) Interacting Stochastic Systems, pp. 435–450. Springer, Berlin (2005)
37. Penrose, M.D., Wade, A.R.: On the total length of the random minimal directed spanning tree. Adv. Appl. Probab. **38**(2), 336–372 (2006)
38. Rachev, S.T., Rüschendorf, L.: Probability metrics and recursive algorithms. Adv. Appl. Probab. **27**(3), 770–799 (1995)
39. Rösler, U.: A limit theorem for "Quicksort". RAIRO Inform. Théor. Appl. **25**(1), 85–100 (1991)
40. Rösler, U.: A fixed point theorem for distributions. Stoch. Process. Appl. **42**(2), 195–214 (1992)
41. Rösler, U., Rüschendorf, L.: The contraction method for recursive algorithms. Algorithmica **29**(1–2), 3–33 (2001). Average-case analysis of algorithms (Princeton, NJ, 1998)
42. Rüschendorf, L.: On stochastic recursive equations of sum and max type. J. Appl. Probab. **43**(3), 687–703 (2006)
43. Vatutin, V.A., Topchiĭ, V.A.: The maximum of critical Galton-Watson processes, and left-continuous random walks. Theory Probab. Appl. **42**(1), 17–27 (1998)
44. Vervaat, W.: On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. Adv. Appl. Probab. **11**(4), 750–783 (1979)
45. Woyczyński, W.A.: On Marcinkiewicz-Zygmund laws of large numbers in Banach spaces and related rates of convergence. Probab. Math. Stat. **1**(2), 117–131 (1980)
46. Zolotarev, V.M.: Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. Theory Probab. Appl. **21**(4), 721–737 (1976)

Precise Tail Index of Fixed Points of the Two-Sided Smoothing Transform

Gerold Alsmeyer, Ewa Damek, and Sebastian Mentemeier

Abstract We consider real-valued random variables R satisfying the distributional equation

$$R \stackrel{d}{=} \sum_{k=1}^{N} T_k R_k + Q,$$

where $R_1, R_2, \ldots$ are iid copies of R and independent of $\mathbf{T} = (Q, (T_k)_{k\geq 1})$. N is the number of nonzero weights T_k and assumed to be a.s. finite. Its properties are governed by the function

$$m(s) := \mathbb{E} \sum_{k=1}^{N} |T_k|^s .$$

There are at most two values $\alpha < \beta$ such that $m(\alpha) = m(\beta) = 1$. We consider solutions R with finite moment of order $s > \alpha$. We review results about existence and uniqueness. Assuming the existence of β and an additional mild moment condition on the T_k, our main result asserts that

$$\lim_{t\to\infty} t^{\beta}\mathbb{P}(|R| > t) \ = \ K \ > \ 0.$$

G. Alsmeyer (✉) · S. Mentemeier
Westfälische Wilhelms-Universität Münster, Institut für Mathematische Statistik, Einsteinstraße 62, 48149 Münster, Germany
e-mail: gerolda@math.uni-muenster.de; mentemeier@uni-muenster.de

E. Damek
Uniwersytet Wroclawski, Instytut Matematyczny, Pl. Grunwaldzki 2/4, 50–384 Wroclaw, Poland
e-mail: edamek@math.uni.wroc.pl

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_10,

the main contribution being that K is indeed positive and therefore β the precise tail index of $|R|$, for the convergence was recently shown by Jelenkovic and Olvera-Cravioto [10].

1 Introduction

Given a sequence $\mathbf{T} := (Q, T_k)_{k\geq 1}$ of real-valued random variables such that

$$N := \sum_{k\geq 1} \mathbb{1}_{\{T_k\neq 0\}} \tag{1}$$

is a.s. finite and (w.l.o.g.) $|T_1| \geq \ldots \geq |T_N| > 0 = |T_{N+1}| = \ldots$, we consider the associated two-sided *smoothing transform* (homogeneous, if $Q \equiv 0$, nonhomogeneous otherwise)

$$\mathscr{S} : F \mapsto \mathscr{L}\left(\sum_{k=1}^{N} T_k R_k + Q\right) \tag{2}$$

which maps a distribution F on $\mathbb{R}$ to the law of $\sum_{k=1}^{N} T_k R_k + Q$, where $R_1, R_2, \ldots$ are iid random variables with distribution F and independent of $\mathbf{T}$. If $\mathscr{S}(F) = F$, then F as well as any random variable R with this distribution is called a fixed point of $\mathscr{S}$. In terms of random variables the fixed-point property may be expressed as

$$R \overset{d}{=} \sum_{k=1}^{N} T_k R_k + Q \tag{3}$$

where $\overset{d}{=}$ means equality in distribution. (3) is called a *stochastic fixed point equation (SFPE)*.

It is well known that properties of fixed points of $\mathscr{S}$ are intimately related to the behavior of the convex function

$$m(s) := \mathbb{E}\left(\sum_{k=1}^{N} |T_k|^s\right). \tag{4}$$

There are at most two values $0 < \alpha < \beta$ such that $m(\alpha) = m(\beta) = 1$. Assuming that both values exist, we are interested in nonzero solutions R to (3) with finite moment of order $s > \alpha$. For a statement about existence and essential uniqueness of solutions with finite α-moment see Lemma 4 below. On the other hand, there are in general also solutions which have infinite α-moment. For the case of nonnegative weights, they were first studied by Durrett and Liggett [7]. In the case of the two-sided smoothing transform, they are studied in recent work by Meiners [15] containing also a characterization of the complete set of fixed points.

For the particular solution with finite moment of order $s > \alpha$, the main result of the present paper is that, under natural assumptions,

$$\lim_{t\to\infty} t^{\beta}\mathbb{P}(|R| > t) \;=\; K \;>\; 0, \tag{5}$$

the main contribution actually being that the constant K is positive and thus β the precise tail index of $|R|$. The convergence was recently derived by Jelenkovic and Olvera-Cravioto [10] via an extension of Goldie's implicit renewal theorem [8] to the branching case ($\mathbb{P}(N > 1) > 0$), see Theorem 1 below. In the homogeneous case with nonnegative i.i.d. weights, (5) was first shown by Guivarc'h [9], see also [14] and the references therein.

Our result is obtained by extending $r \mapsto \mathbb{E}\,|R|^r$ as a holomorphic function and by showing that it has a singularity at β if and only if $K > 0$. This technique was first used in [5] in the study of solutions to multidimensional affine recursions. We are grateful to Mariusz Mirek (personal communication) for drawing our attention to it in the context of the branching equation (3) considered here.

We have organized this work as follows. Section 2 introduces notation and basic assumptions, provides information about the chosen setup and reviews preliminary results. Our main results are stated in Sect. 3. Proofs are given in Sect. 4 with some more technical calculations deferred to Sect. 5.

2 Preliminaries

2.1 Notations and Assumptions

For $m(s)$ defined in (4), note that $m(0) = \mathbb{E}N$ may be infinite. We put

$$\mathfrak{D} := \{s \geq 0 : m(s) < \infty\}, \quad s_0 := \inf \mathfrak{D} \quad \text{and} \quad s_1 := \sup \mathfrak{D}.$$

If $\mathfrak{D}$ is nonempty, then m is a convex function on $\mathfrak{D}$. Since m can be seen as the Laplace transform of an intensity measure (see [3, (1.8)]), we further have the following result (with $\Re z$ denoting the real part of a complex number z):

Lemma 1. *Suppose $\mathfrak{D} = \{s \geq 0 : m(s) < \infty\}$ has inner points, i.e. $s_0 < s_1$. Then the function m extends holomorphically to the strip $s_0 < \Re z < s_1$.*

Our standing assumption throughout this paper is that

$$\exists\, s_0 < \alpha < \beta < s_1 : \quad m(\alpha) = m(\beta) = 1. \tag{A}$$

Then $m'(\alpha) < 1$ and $m'(\beta) > 1$ (Fig. 1).

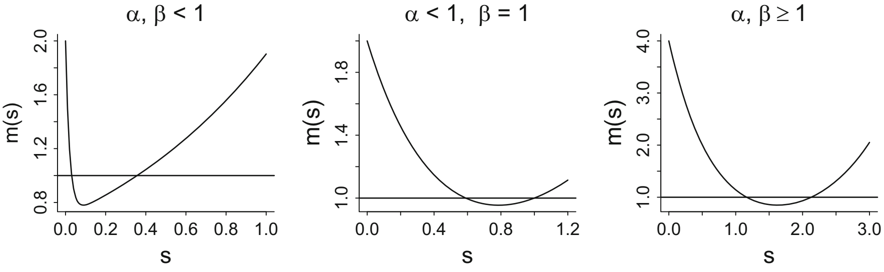


Fig. 1 Three distinct cases for the values of α and β

The existence of a solution R to the SFPE (3) with finite moment of order $s > \alpha$ and a power law behavior of type (5) imposes some restrictions on the range of α to be discussed below. Our assumptions are:

- $1 \le \alpha < 2$ if $Q = 0$ a.s. (homogeneous case),
- $0 < \alpha < 2$ if $\mathbb{P}(Q = 0) < 1$ (nonhomogeneous case).

Next are further moment conditions imposed on Q and $\sum_{k=1}^{N} T_k$. There are situations, in which these quantities govern the tail behavior of R, see [12] for a detailed discussion and references. The condition on Q is quite obvious,viz.

$$\mathbb{E}\,|Q|^s < \infty \quad \text{for all } s < s_1. \tag{B}$$

We will impose conditions on the weight sums $\sum_{k=1}^{N} |T_k|^s$ and $(\sum_{k=1}^{N} |T_k|)^s$ by introducing two functions that dominate $m(s)$ for $s \ge 1$ and $s \le 1$, respectively. Define

$$\mu(s) := \mathbb{E}\left(\sum_{k=1}^{N} |T_k|\right)^s \tag{6}$$

and, for $\epsilon > 0$,

$$m_\epsilon(s) := \mathbb{E}\left(\sum_{k=1}^{N} |T_k|^s\right)^{1+\epsilon}. \tag{7}$$

Then $m(s) \le \mu(s)$ for $s \ge 1$, while $m(s) \le 1 + m_\epsilon(s)$ for $s \le 1$. These functions appear quite naturally in existence theorems for solutions, see below and [1]. Put

$$\mathfrak{D}_\mu := \{s \ge 0 \ : \ \mu(s) < \infty\}, \quad s_\infty := \sup \mathfrak{D}_\mu;$$
$$\mathfrak{D}_\epsilon := \{s \ge 0 : m_\epsilon(s) < \infty\}, \quad s_\epsilon := \sup \mathfrak{D}_\epsilon.$$

Our analysis will often require the study of certain moments of order $s \geq \beta$ and the distinction between the cases when $s > 1, = 1, < 1$. Corresponding to these cases are three different sets of assumptions we introduce now, namely:

$$\beta < s_\infty; \tag{C}$$

$$[\beta - \delta_0, \beta] \subset \mathfrak{D}_{\epsilon_0} \quad \text{for some } \delta_0, \epsilon_0 > 0; \tag{D}$$

$$[\beta - \delta_0, 1] \subset \mathfrak{D}_{\epsilon_0} \quad \text{for some } \delta_0, \epsilon_0 > 0. \tag{D*}$$

Since $\mathfrak{D}_{\epsilon_1} \subset \mathfrak{D}_{\epsilon_2}$ for $\epsilon_1 \geq \epsilon_2$, we may define

$$\hat{s}_\infty := 1 \wedge \lim_{\epsilon \to 0} s_\epsilon$$

and then note that condition (B) implies

$$\mathbb{E}\,|Q|^s < \infty \quad \text{for all } s < \max\{s_\infty, \hat{s}_\infty\}. \tag{8}$$

Finally, if $\alpha \geq 1$, we have to assume that the mean version of the SFPE (3) has a solution, viz.

$$r \;=\; r\,\mathbb{E}\left(\sum_{k=1}^{N} T_k\right) \;+\; \mathbb{E}Q \tag{E}$$

for some $r \in \mathbb{R}$. Note that r is unique, unless $\mathbb{E}(\sum_{k=1}^{N} T_k) = 1$ and $\mathbb{E}Q = 0$.

2.2 *Discussion of the Restrictions on α*

The afore-stated restrictions on the range of α, called *characteristic exponent of* **T** *or* $\mathscr{S}$ in [2–4], will now be justified by a number of lemmata. The first one settles the restriction $\alpha \geq 1$ in the homogeneous case.

Lemma 2. *Suppose that* $\alpha < 1$, $Q = 0$ *and let* R *be a solution to* (3) *with finite moment of order* $s > \alpha$. *Then* $R = 0$ *a.s.*

Proof. Plainly, we may assume $s \in (\alpha, 1)$ and $m(s) < 1$, for such s exists by (A). Then (3) in combination with the subadditivity of $x \mapsto x^s$ for $x \geq 0$ provides us with

$$\begin{aligned}\mathbb{E}|R|^s \;&=\; \mathbb{E}\left|\sum_{k=1}^{N} T_k R_k\right|^s \;\leq\; \sum_{k=1}^{\infty} \mathbb{E}|T_k R_k|^s \\ &=\; \mathbb{E}|R|^s \sum_{k=1}^{\infty} \mathbb{E}|T_k|^s \;=\; \mathbb{E}|R|^s\, m(s)\end{aligned}$$

and thus $\mathbb{E}|R|^s = 0$. ∎

Lemma 3. *Let R be a nonzero solution to (3) with finite moment of order $s \geq 2$. Then $m(2) \leq 1$ and thus $\alpha \leq 2$.*

Proof. Let $\mathbb{E}_{\mathbf{T}}$ and $\mathrm{Var}_{\mathbf{T}}$ denote conditional expectation and conditional variance with respect to $\mathbf{T}$. Introduce i.i.d. copies $R_1, R_2, \ldots$ of R, which are independent of $\mathbf{T}$ and defined on the same probability space. Put $R^* := \sum_{k=1}^{N} T_k R_k + Q$. Then $R^* \stackrel{d}{=} R$, for R is a solution to (3), and R^* satisfies (3) not only in distribution, but even a.s. Since the results we obtain are distributional properties, we may w.l.o.g. assume that R itself satisfies (3) a.s. Then

$$\mathrm{Var}\, R \;=\; \mathbb{E}(\mathrm{Var}_{\mathbf{T}} R) + \mathrm{Var}(\mathbb{E}_{\mathbf{T}} R). \tag{9}$$

Moreover,

$$\mathrm{Var}_{\mathbf{T}} R \;=\; \mathrm{Var}_{\mathbf{T}} \left[\sum_{k=1}^{N} T_k R_k + Q \right] \;=\; \sum_{k=1}^{N} T_k^2 \,\mathrm{Var}(R), \tag{10}$$

whence, upon taking unconditional expectation, we obtain

$$\infty \;>\; \mathrm{Var}\, R \;\geq\; \mathbb{E}(\mathrm{Var}_{\mathbf{T}} R) \;=\; \mathbb{E}\left[\sum_{k=1}^{N} T_k^2 \right] \mathrm{Var}\, R \;=\; m(2)\,\mathrm{Var} R \;>\; 0 \tag{11}$$

and thus $m(2) \leq 1$ as claimed. □

If $\alpha = 2$, then (10) implies $\mathbb{E}\mathrm{Var}_{\mathbf{T}} R = \mathrm{Var}\, R$ and thus, by (9), $\mathrm{Var}(\mathbb{E}_{\mathbf{T}} R) = 0$. Consequently, $\mathbb{E}_{\mathbf{T}} R$ is a.s. constant, in fact

$$\mathbb{E} R \;=\; \mathbb{E}_{\mathbf{T}} R \;=\; \sum_{k=1}^{N} T_k \mathbb{E} R + Q \quad \mathbb{P}\text{-a.s.}$$

or, equivalently,

$$Q \;=\; \left(1 - \sum_{k=1}^{N} T_k \right) \mathbb{E}\, R \quad \mathbb{P}\text{-a.s.}$$

In the homogeneous case, we infer that $\mathbb{E}\, R = 0$ or $\sum_{k=1}^{N} T_k = 1$ a.s. must hold. The first subcase, studied by Caliebe and Rösler [6], leads to mixtures of centered normal distributions, the mixing distribution being the law of a positive constant times the nonnegative, mean one solution to the SFPE

$$W \stackrel{d}{=} \sum_{k=1}^{N} T_k^2 W_k. \tag{12}$$

The latter solution exists and is unique if $\mu(s) < \infty$ for some $s > 2$ (see also [3, Theorem 2.1]). In the second subcase, viz. $\sum_{k=1}^{N} T_k = 1$ a.s., we have that if R is a solution, then so are all $R + m$, $m \in \mathbb{R}$. Consequently, all shifts of solutions in the first subcase are solutions in the second subcase. If $m(s) < 1$ for $s \in (2, 3]$, then [1, Theorem 6.16] gives that there is a unique solution with fixed first and second moment whence, under our standing assumption (A), solutions are mixtures of normal distributions.

The following result provides the extension to the nonhomogeneous case (see also [3, Theorem 2.3] for the case of nonnegative T_k).

Proposition 1. *Suppose that $\mathbb{P}(Q \neq 0) > 0$, $\alpha = 2$ and that, for some $s > 2$, $m(s) < 1$ and $\mu(s) < \infty$. Suppose further that*

$$Q = r(1 - \sum_{k=1}^{N} T_k) \quad \mathbb{P}\text{-a.s.} \tag{13}$$

for some $r \neq 0$. Then, for any $v \geq 0$, there is a unique solution R to the SFPE (3) with mean r and variance v^2. It is symmetric about r and has characteristic function given by

$$\phi_R(t) \;=\; \mathbb{E}\left[\exp\left(irt - \frac{v^2 t^2 W}{2}\right)\right], \tag{14}$$

where W is the unique mean one solution to (12).

Proof. If $m(s) < 1$, $\mu(s) < \infty$ for some $s > 2$, then the smoothing transform is a contraction with respect to the Zolotarev metric ζ_s as defined in [16] on the subsets of probability measures with fixed first and second moment. This fact is easily derived from (a straightforward extension of) Lemma 3.1 in the afore-mentioned reference. Hence we conclude that $\mathscr{S}$ has a unique fixed point with mean r and arbitrary variance $v^2 \geq 0$. Therefore, it remains to verify that R with characteristic function given by (14) does indeed solve our SFPE (3). To this end, let F be the law of R. Then, with $R_1, R_2, \ldots$ and $W_1, W_2, \ldots$ being i.i.d. copies of R and W, respectively, and also independent of $\mathbf{T}$, we obtain

$$\begin{aligned}
\phi_{\mathscr{S}(F)}(t) \;&=\; \mathbb{E}\left[\exp\left(it \sum_{k=1}^{N} T_k R_k + Q\right)\right] \\
&=\; \mathbb{E}\left[\exp(itQ)\, \mathbb{E}_{\mathbf{T}}\left(\prod_{k=1}^{N} \exp(itT_k R_k)\right)\right] \\
&=\; \mathbb{E}\left[\exp(itQ) \prod_{k=1}^{N} \phi_R(tT_k)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\exp\left(itr\left(1-\sum_{k=1}^{N}T_k\right)\right)\prod_{k=1}^{N}\mathbb{E}_{\mathbf{T}}\left[\exp\left(irtT_k - \frac{v^2t^2T_k^2W_k}{2}\right)\right]\right] \\
&= \mathbb{E}\left[\exp\left(itr\left(1-\sum_{k=1}^{N}T_k\right) + irt\sum_{k=1}^{N}T_k - \frac{v^2t^2}{2}\sum_{k=1}^{N}T_k^2W_k\right)\right] \\
&= \mathbb{E}\left[\exp\left(irt - \frac{v^2t^2W}{2}\right)\right] \\
&= \phi_R(t),
\end{aligned}
$$

where we have used assumption (13) on Q in line four and the fixed-point property (12) for W in the last line. □

Summarizing the situation in the case $\alpha = 2$ and $\mathbb{P}(\sum_{k=1}^{N} T_k = 1) < 1$, the law of R in (14) is a W-mixture of normal laws with some fixed mean $r \in \mathbb{R}$ and variance v^2w. It exhibits a power law behavior only if this is true for (the law of) W which in turn is a fixed point of the smoothing transform pertaining to $(T_k^2)_{k\geq 1}$, the latter having characteristic exponent 1. With regard to (5), it is therefore no loss of generality to assume $\alpha < 2$ hereafter.

2.3 Existence and Uniqueness of a Fixed Point with Finite α-Moment

The following lemma compiles results about existence and uniqueness of a solution to (3) with finite moment of order α and may be deduced from results in [17, Sect. 3] and [16, Sect. 3]; see also [1] for a review. Our Lemma corresponds to [1, Theorem 6.16].

Lemma 4. *Assume* (A), (B) *and* $\alpha < 2$.

(a) If $\alpha < 1$, *then there exists a unique solution* R *to* (3) *such that* $\mathbb{E}\,|R|^s < \infty$ *for all* $s < \beta$. *It is nonzero iff* $\mathbb{P}(Q \neq 0) > 0$.

(b) If $\alpha \geq 1$ *and* (C), (E) *are valid, then there is a unique solution* R *to* (3) *with* $\mathbb{E}R = r$ *(determined by* (E)*) and* $\mathbb{E}\,|R|^s < \infty$ *for all* $s < \beta$. *For the nonhomogeneous equation* R *is always nonzero, and for the homogeneous one* R *is nonzero iff* $r \neq 0$.

Remark 1. Since for the homogeneous equation two nonzero solutions with distinct means are proportional, we may in fact speak of the unique nonzero solution with the property $\mathbb{E}|R|^s < \infty$ for $s < \beta$ when stipulating $\mathbb{E}\,R = 1$.

The following lemma sheds some light on the role of the function $\mu(s)$. As before, $\mathbb{E}_{\mathbf{T}}$ denotes conditional expectation with respect to $\mathbf{T} = (Q, (T_k)_{k\geq 1})$.

Lemma 5. *Let* $s \geq 1$ *and* $\mu(s) < \infty$. *Then* $\mathbb{E}\,|R|^s < \infty$ *implies*

$$\mathbb{E}_{\mathbf{T}}\left(\sum_{k=1}^{N} |T_k R_k|\right)^s \leq C\left(\sum_{k=1}^{N} |T_k|\right)^s \mathbb{E}\,|R|^s \tag{15}$$

for some $C > 0$ *which only depends on* s.

Proof. This follows by an application of one of the Burkholder-Davis-Gundy inequalities (see e.g. [1, Theorem 8.5]) when observing that, given $\mathbf{T}$ and with $r = \mathbb{E}\,R$,

$$\sum_{k=1}^{N} T_k(R_k - r) = \sum_{k\geq 1} T_k(R_k - r)\mathbb{1}_{\{N>n\}}$$

is the limit of the zero-mean martingale $(\sum_{k=1}^{n} T_k(R_k - r))_{n\geq 0}$, which in fact consists of finite weighted sums of i.i.d. random variables; see also the proof of [1, Proposition 4.1]. □

Note that for $s \leq 1$, we have the bound

$$\mathbb{E}_{\mathbf{T}}\left(\sum_{k=1}^{N} |T_k R_k|\right)^s \leq \left(\sum_{k=1}^{N} |T_k|^s\right)\mathbb{E}\,|R|^s \tag{16}$$

due to subadditivity of $x \mapsto x^s$ for $x \geq 0$. Taking unconditional expectation in (15) and (16), we arrive

$$\mathbb{E}\left(\sum_{k=1}^{N} |T_k R_k|\right)^s \leq \begin{cases} C\mu(s)\mathbb{E}|R|^s, & \text{if } s \geq 1, \\ m(s)\mathbb{E}|R|^s, & \text{if } s \leq 1. \end{cases} \tag{17}$$

2.4 The Implicit Renewal Theorem by Jelenkovic and Olvera-Cravioto

Our analysis embarks on the following result about the tails of fixed points of two-sided smoothing transforms due to Jelenkovic and Olvera-Cravioto [10]:

Theorem 1. *Suppose that* (A) *holds, that* $\mathbb{P}(T_j < 0) > 0$ *for some* $j \geq 1$ *and that* $\mathbb{P}(\log|T_k| \in \cdot, N \geq k)$ *is nonlattice for some* $k \geq 1$. *Further assume* (C) *if* $\beta > 1$, *and* (D) *if* $\beta \leq 1$. *Let* R *be the unique solution to* (3). *Then*

$$\lim_{t\to\infty} t^{\beta}\mathbb{P}(|R| > t) = \frac{K(\beta)}{m'(\beta)},$$

where

$$K(\beta) := \int_0^\infty \left(\mathbb{P}(|R| > t) - \sum_{k\geq 1} \mathbb{P}(|T_k R_k| > t) \right) t^{\beta-1}\, dt.$$

Proof. As will be explained at the beginning of Sect. 5, conditions (C) and (D) imply the finiteness of

$$\int_0^\infty \left| \mathbb{P}(R > t) - \sum_{k\geq 1} \mathbb{P}(T_k R_k > t) \right| t^{\beta-1}\, dt \tag{18}$$

and

$$\int_0^\infty \left| \mathbb{P}(R < -t) - \sum_{k\geq 1} \mathbb{P}(T_k R_k < -t) \right| t^{\beta-1}\, dt, \tag{19}$$

respectively. Taking this for granted here, the stated result is [10, Theorem 3.4]. □

3 Main Result

We are now ready for our main result which, loosely speaking, asserts that either R has power tails of order β, or a finite moment of order $s > \beta$.

Theorem 2. *Under the assumptions of Theorem 1, the following assertions hold true:*

(a) If $\beta > 1$ and (A), (B), (C) hold true, then either $K(\beta) > 0$, or $\mathbb{E}\,|R|^s < \infty$ for all $s < s_\infty$.
(b) If $\beta \leq 1$ and (A), (B), (C), (D) hold true, then either $K(\beta) > 0$, or $\mathbb{E}\,|R|^s < \infty$ for all $s < s_\infty$.*
(c) If $\beta < 1$ and (A), (B), (D) hold true, then either $K(\beta) > 0$, or $\mathbb{E}\,|R|^s < \infty$ for all $s < \hat{s}_\infty$.

The following proposition provides a sufficient condition for $K(\beta) > 0$.

Proposition 2. *Keeping the assumptions of Theorem 2, let $k \in \mathbb{N}$ be such that $\mathbb{P}(\log|T_k| \in \cdot, N \geq k)$ is nonlattice and assume that $\mathbb{E}\,|T_k|^\gamma = 1$ for some $\beta < \gamma < s_\infty$ in parts (a), (b), resp. $\beta < \gamma < \hat{s}_\infty$ in part (c). Then*

$$K(\beta) > 0 \quad \textit{iff} \quad \mathbb{P}\left(r \sum_{k=1}^N T_k + Q = r \right) < 1 \textit{ for all } r \neq 0,$$

the latter condition being equivalent to

$$\mathbb{P}\left(\sum_{k=1}^{N} T_k = 1\right) < 1$$

in the homogeneous case.

Note that the existence of γ is only a mild condition because

$$\mathbb{P}\left(\max_{1\le k\le N} |T_k| > 1\right) = \mathbb{P}(|T_1| > 1) > 0.$$

Namely, if the latter failed to hold, then $m(s)$ would be decreasing function and thus $m(s) < 1$ for all $s > \alpha$. But this is impossible as $m(\beta) = 1$.

Note further that $\mathbb{P}(r \sum_{k=1}^{N} T_k + Q = r) < 1$ is obviously necessary for heavy tail behaviour, for otherwise $R \equiv r$ would be the unique solution with $\mathbb{E}R = r$.

4 Proof of the Main Theorem

We start with two lemmata about holomorphic functions, the first one giving a basic property of the so-called Mellin transform of a measurable function and being proved in the Appendix.

Lemma 6. *Let $f : \mathbb{R}_{\ge} \to \mathbb{R}$ be a measurable function such that*

$$\int_0^\infty t^{s-1} |f(t)| \, dt < \infty$$

for $s \in \{\sigma_0, \sigma_1\} \subset \mathbb{R}_>$. Then its Mellin transform

$$g(z) := \int_0^\infty t^{z-1} f(t) dt \tag{20}$$

is well defined and holomorphic in the strip $\sigma_0 < \Re z < \sigma_1$.

The next lemma, a proof of which may for instance be found in [18, Theorem II.5b], will play a crucial role in the proof of our main result and is historically due to Landau. Its first application in the given context appears in [5].

Lemma 7. *Given the situation of Lemma 6, suppose further that f is monotonic. Let $\sigma_1 := \sup\{s > 0 : g(s) < \infty\}$ denote the* ***a****bscissa of convergence of g. Then g cannot be extended holomorphically onto any neighborhood of σ_1.*

Defining $G(s) := s^{-1}\mathbb{E}|R|^s$ and $\sigma := \sup\{s > 0 : G(s) < \infty\}$, we have as an immediate consequence:

Corollary 1. *The function G cannot be extended holomorphically onto any neighbourhood of σ.*

Put

$$K(s) := \int_0^\infty \left(\mathbb{P}(|R| > t) - \sum_{k=1}^\infty \mathbb{P}(|T_k R_k| > t)\right) t^{s-1}\, dt \tag{21}$$

and suppose that (C) is valid. In order to show with the help of Lemma 6 that K has a holomorphic extension onto a neighborhood of β, the following proposition is crucial. Its proof will be given in Sect. 5.

Proposition 3. *Assuming (A), (B) and $\sigma \geq \beta$, it follows that $K(\sigma + \delta) < \infty$ for some $\delta > 0$ provided that, furthermore,*

- *(C) holds true and $\sigma < s_\infty$ if $\sigma > 1$,*
- *(C), (D*) hold true and $\sigma < s_\infty$ if $\sigma = 1$,*
- *(D) holds true and $\sigma < \hat{s}_\infty$ if $\sigma < 1$.*

Proof (of Theorem 2). Our proof consists of two steps (tacitly assuming the respective assumptions of the theorem for the cases $\beta >, =, < 1$):

STEP 1. $K(\beta) = 0$ iff $\sigma > \beta$.
STEP 2. If $K(\beta) = 0$ and $\sigma > \beta$, then $\sigma = s_\infty$, resp. $= \hat{s}_\infty$

Before proceeding with these steps, we make the following observation (under the assumptions of the theorem): Lemmata 4 and 5 ensure that $\mathbb{E}\sum_{k=1}^N |T_k R_k|^s$ and $\mathbb{E}|R|^s$ are both finite for $\alpha < s < \beta$. Therefore, we may compute

$$\begin{aligned} K(s) &= \int_0^\infty \left(\mathbb{P}(|R| > t) - \sum_{k=1}^\infty \mathbb{P}(|T_k R_k| > t)\right) t^{s-1}\, dt \\ &= \int_0^\infty \mathbb{P}(|R| > t) t^{s-1}\, dt - \int_0^\infty \sum_{k=1}^\infty \mathbb{P}(|T_k R_k| > t) t^{s-1}\, dt \\ &= \frac{1}{s}\mathbb{E}|R|^s - \frac{1}{s}\mathbb{E}\left[\sum_{k=1}^N |T_k R_k|^s\right] \\ &= \frac{1}{s}(1 - m(s))\,\mathbb{E}|R|^s \end{aligned}$$

giving

$$\frac{K(s)}{1 - m(s)} = G(s) \tag{22}$$

for all $s \in (\alpha, \beta)$. By Lemma 6 (with $f(t) = \mathbb{P}(|R| > t) + \sum_{k=1}^{\infty} \mathbb{P}(|T_k R_k| > t)$) and Lemma 1, both sides extend to holomorphic functions onto the strip $\alpha < \Re z < \beta$, and (22) remains valid in this strip by the identity theorem for holomorphic functions.

STEP 1. Notice that $(1 - m(z))^{-1}$ has a pole of order 1 at β, for $m'(\beta) > 0$, but is holomorphic otherwise in a neighborhood of β. By Proposition 3 and Lemma 6, $K(z)$ is holomorphic in the strip $\alpha < \Re z < \beta + \delta$ for some $\delta > 0$. Hence, if $K(\beta) = 0$, then the left-hand side (LHS) of (22) has a holomorphic extension to a neighborhood of β, and this is also an extension of the RHS, giving $\sigma > \beta$. On the other hand, $K(\beta) > 0$ entails $G(\beta) = \infty$, i.e. $\sigma \leq \beta$.

STEP 2. Now assume $K(\beta) = 0$ and $\sigma > \beta$, but $\sigma < s_\infty$, resp. $\sigma < \hat{s}_\infty$. Then, for all $\beta < \Re z < \sigma$, we have

$$\frac{K(z)}{1 - m(z)} = G(z),$$

whence by another appeal to Proposition 3 together with Lemmata 1 and 6, the LHS extends holomorphically onto $\beta < \Re z < \sigma + \delta$ for some $\delta > 0$, giving an holomorphic extension of the RHS. But this is a contradiction to Corollary 1.

□

We finish this section with the proof of Proposition 2.

Proof (of Proposition 2). First of all, if $K(\beta) > 0$, then the uniqueness of R as a solution to (3) implies that $\mathbb{P}(r \sum_{k=1}^{N} T_k + Q = r) < 1$ for any $r \neq 0$. In order to show the converse, suppose that $K(\beta) = 0$ and thus $\mathbb{E}|R|^s < \infty$ for any $s < s_\infty$, resp. $< \hat{s}_\infty$. W.l.o.g. let $k = 1$, so that $\mathbb{E}\,|T_1|^\gamma = 1$ is assumed. Putting $B := \sum_{k=2}^{N} T_k R_k + Q$, the random variable R satisfies the SFPE

$$R \stackrel{d}{=} T_1 R_1 + B. \tag{23}$$

Since $\mathbb{E}\,|R|^\gamma < \infty$, $m(\gamma) < \infty$, and (if $\gamma > 1$) $\mu(\gamma) < \infty$, we find that the following conditions are fulfilled:

$\mathbb{E}\,|B|^\gamma < \infty$ (by Lemma 5);
$\mathbb{E}\,|T_1|^\gamma = 1$;
$\mathbb{P}(\log|T_1| \in \cdot)$ is nonarithmetic;
$\mathbb{E}\,|T_1|^\gamma \log^+ |T_1| < \infty$.

These conditions render uniqueness of R as a solution to (23) and allow to invoke the results by Kesten [13, Theorem 5] and Goldie [8, Theorem 4.1] to infer that $\mathbb{E}|R|^\gamma < \infty$ and thus $t^\gamma\, \mathbb{P}(|R| > t) = o(1)$ as $t \to \infty$ can only hold if

$$T_1 r + B = r \quad \text{a.s. for some } r \in \mathbb{R}$$

or, equivalently, $R = r$ a.s. (by uniqueness) which in turn is equivalent to

$$r\sum_{k=1}^{N} T_k + Q = r \quad \text{a.s.}$$

This completes our proof of the proposition. □

5 Bounds for *K(s)*

We proceed to a proof of Proposition 3. This proof with $r = \beta$ and $(\cdot)^{\pm}$ instead of $|\cdot|$ also shows the finiteness of (18) and (19), thus completing the argument in the proof of Theorem 1.

Proof (of Proposition 3). By using [8, Lemma 9.4] (in corrected form), and upon defining

$$\begin{aligned}
H(s) &:= \mathbb{E}\left|\left|\sum_{k=1}^{N} T_k R_k + Q\right|^s - \left|\sum_{k=1}^{N} T_k R_k\right|^s\right|,\\
I(s) &:= \mathbb{E}\left|\left|\sum_{k=1}^{N} T_k R_k\right|^s - \sum_{k=1}^{N}|T_k R_k|^s\right|,\\
J(s) &:= \mathbb{E}\left[\sum_{k=1}^{N}|T_k R_k|^s - \sup_{1\le k\le N}|T_k R_k|^s\right],
\end{aligned}$$

we obtain the following estimate for $K(s)$:

$$\begin{aligned}
K(s) &= \int_0^\infty s t^{s-1}\left|\mathbb{P}\left(\left|\sum_{k=1}^{N} T_k R_k + Q\right| > t\right) - \sum_{k=1}^{\infty}\mathbb{P}(|T_k R_k| > t)\right| dt\\
&\le \int_0^\infty s t^{s-1}\left|\mathbb{P}\left(\left|\sum_{k=1}^{N} T_k R_k + Q\right| > t\right) - \mathbb{P}(\left|\sum_{k=1}^{N} T_k R_k\right| > t)\right| dt\\
&\quad + \int_0^\infty s t^{s-1}\left|\mathbb{P}\left(\left|\sum_{k=1}^{N} T_k R_k\right| > t\right) - \mathbb{P}\left(\sup_{1\le k\le N}|T_k R_k| > t\right)\right| dt\\
&\quad + \int_0^\infty s t^{s-1}\left(\sum_{k=1}^{\infty}\mathbb{P}(|T_k R_k| > t) - \mathbb{P}\left(\sup_{1\le k\le N}|T_k R_k| > t\right)\right) dt
\end{aligned}$$

$$= H(s) + \mathbb{E}\left|\left|\sum_{k=1}^{N} T_k R_k\right|^s - \sup_{1\le k\le N} |T_k R_k|^s\right| + J(s)$$
$$\le H(s) + I(s) + 2J(s).$$

As for the second to last line, we note that the appearing integrand is indeed nonnegative because it is equal to $st^{s-1}\sum_{k\ge 2}\mathbb{P}(Y_k > t)$ where $(Y_k)_{k\ge 1}$ denotes the decreasing order statistic of $(|T_k R_k|)_{k\ge 1}$. Then use Fubini's theorem as in [11, Lemma 4.6] to see that the pertinent integral equals $J(s)$. The proof is completed by the next three lemmata which will show that, for some $\delta > 0$, $H(s)$, $I(s)$ and $J(s)$ are bounded for all $\sigma < s < \sigma + \delta$. □

Lemma 8. *Suppose that* (B) *holds and* $\sigma \ge \beta$. *If* $\sigma \ge 1$, *suppose further* (C) *be true and* $\sigma < s_\infty$. *Then*

$$H(s) := \mathbb{E}\left|\left|\sum_{k=1}^{N} T_k R_k + Q\right|^s - \left|\sum_{k=1}^{N} T_k R_k\right|^s\right| < \infty$$

for all $\sigma \le s < \sigma + \delta$ *and some* $\delta > 0$.

Proof. If $s \le 1$, then (recalling (8))

$$H(s) \le \mathbb{E}|Q|^s < \infty.$$

If $s > 1$, choose $\delta \in (0, 1]$ such that $\sigma + \delta < s_\infty$. Now for $1 < s < \sigma + \delta$, use the inequalities

$$|a^s - b^s| \le s(a \vee b)^{s-\delta}|a - b|^\delta,$$
$$(a + b)^s \le 2^{s-1}(a^s + b^s),$$

valid for $a, b \ge 0$, to infer (with $a = |\sum_{k=1}^N T_k R_k + Q|$ and $b = |\sum_{k=1}^N T_k R_k|$)

$$H(s) \le s(1 \vee 2^{s-\delta-1})\,\mathbb{E}\left[|Q|^s + \left|\sum_{k=1}^{N} T_k R_k\right|^{s-\delta} |Q|^\delta\right].$$

The last expectation is finite because, by Lemma 5 and Hölder's inequality,

$$\mathbb{E}\left[\left|\sum_{k=1}^{N} T_k R_k\right|^{s-\delta} |Q|^\delta\right] \le C\,\mathbb{E}|R|^{s-\delta}\,\mathbb{E}\left[\left|\sum_{k=1}^{N} |T_k|\right|^{s-\delta} |Q|^\delta\right]$$
$$\le C\,\mu(s)^{(s-\delta)/s}(\mathbb{E}|Q|^s)^{\delta/s}$$

for some constant $C \in \mathbb{R}_>$ □

Lemma 9. *Let $\sigma \geq \beta$. Suppose that* (C) *holds and $\sigma < s_\infty$ if $\sigma > 1$, that* (D*) *holds if $\sigma = 1$, and that* (D) *holds and $\sigma < \hat{s}_\infty$ if $\sigma < 1$. Then $J(s) < \infty$ for all $0 < s < \sigma + \delta$ and some $\delta > 0$.*

Proof. If $\sigma \leq 1$, pick $\delta \in (0, \delta_0)$ such that $\frac{\sigma+\delta}{\sigma-\delta} < 1 + \varepsilon_0$ (and $\sigma + \delta < \hat{s}_\infty$ if $\sigma < 1$). If $\sigma > 1$, pick $\delta > 0$ such that $[\sigma - \delta, \sigma + \delta] \subset (1, s_\infty)$.

If $0 < s < \sigma - \delta$, then $J(s) < \infty$ follows from the obvious estimate

$$J(s) \;\leq\; \sum_{k\geq 1} \mathbb{E}|T_k|^s \,\mathbb{E}|R|^s \;=\; m(s)\,\mathbb{E}|R|^s.$$

So let $s \in (\sigma-\delta, \sigma+\delta)$ hereafter. Then one can follow the proof of [11, Lemma 4.6] (replacing (α, β) and $C_i R_i$ there with $(s, \sigma - \delta)$ and $|T_k R_k|$, respectively) to obtain the bound

$$\begin{aligned} J(s) \;&\leq\; C\left(\mathbb{E}\,|R|^{\sigma-\delta}\right)^{s/(\sigma-\delta)} \mathbb{E}\left[\left(\sum_{k=1}^{N} |T_k|^{\sigma-\delta}\right)^{s/(\sigma-\delta)}\right] \;<\; \infty \\ &=\; C\left(\mathbb{E}\,|R|^{\sigma-\delta}\right)^{s/(\sigma-\delta)} m_{\varepsilon_0}\left(\frac{s}{1+\varepsilon_0}\right) \;<\; \infty \end{aligned}$$

for some constant $C \in \mathbb{R}_>$. Here we should note that, if $\sigma - \delta < 1$, the second expectation on the right-hand side is indeed finite because $s/(\sigma - \delta) < 1 + \epsilon_0$ and $\sigma - \delta < \hat{s}_\infty$ ensures $m_{\varepsilon_0}(\sigma - \delta) < \infty$. If $\sigma - \delta \geq 1$ then we arrive at the same conclusion, for $\sum_{k=1}^N |T_k|^{\sigma-\delta} \leq \left(\sum_{k=1}^N |T_k|\right)^{\sigma-\delta}$. □

Lemma 10. *Let $\sigma \geq \beta$. Assume* (C) *and $\sigma < s_\infty$ if $\sigma > 1$,* (D*) *if $\sigma = 1$, and* (D) *and $\sigma < \hat{s}_\infty$ if $\sigma < 1$. Then $I(s) < \infty$ for all $0 < s < \sigma + \delta$ and some $\delta > 0$.*

Proof. The first part of the proof follows the argument given for [10, Lemmata 4.8 and 4.9]. Put $S := \sum_{k=1}^N T_k R_k$, $S^\pm = \left(\sum_{k=1}^N T_k R_k\right)^\pm$, $S_\pm := \sum_{k=1}^N (T_k R_k)^\pm$ and $S_\pm(s) := \sum_{k=1}^N \left((T_k R_k)^\pm\right)^s$. Then

$$\begin{aligned} I(s) \;&=\; \mathbb{E}\,||S|^s - S_+(s) - S_-(s)| \\ &=\; \mathbb{E}\left|(S^+)^s + (S^-)^s - S_+(s) - S_-(s)\right| \\ &\leq\; \mathbb{E}\left|(S^+)^s - S_+(s)\right| + \mathbb{E}\,|(S^-)^s - S_-(s)|\,, \end{aligned}$$

whence it suffices to show $\mathbb{E}\left|(S^\pm)^s - S_\pm(s)\right| < \infty$ and, by an obvious reflection argument, only $\mathbb{E}\left|(S^+)^s - S_+(s)\right| < \infty$. As in [10], we estimate

$$\begin{aligned} \mathbb{E}\left|(S^+)^s - S_+(s)\right| \;\leq\; \mathbb{E}S_+(s)\mathbb{1}_{\{S_+ \leq S_-\}} \;+\; \mathbb{E}\left(S_+^s - (S_+ - S_-)^s\right)\mathbb{1}_{\{S_+ > S_-\}} \\ +\; \mathbb{E}|S_+^s - S_+(s)| \end{aligned} \tag{24}$$

The first two expectations on the right-hand side can be bounded by a constant times

$$\left(\mathbb{E}|R|^{s/(1+\epsilon)}\right)^{1+\epsilon}\,\mathbb{E}\left(\sum_{k=1}^{N}|T_k|^{s/(1+\epsilon)}\right)^{1+\epsilon}$$

if $\sigma<1$ (choose $a=s/(1+\epsilon)$ and $b=s\epsilon/(1+\epsilon)$ in the proof of [10, Lemma 4.9]), and by a constant times

$$\mathbb{E}|R|\,\mathbb{E}|R|^{s-1}\mathbb{E}\left(\sum_{k=1}^{N}|T_k|\right)^{s}$$

if $\sigma\geq 1$. These bounds are finite if $0<s<\sigma+\delta$ for sufficiently small $\delta>0$ and $\epsilon<\epsilon_0$ with ϵ_0 given by (D) or (D*).

It remains to show finiteness of the final expectation in (24), viz. of

$$L(s)\;:=\;\mathbb{E}\left|\left(\sum_{k=1}^{N}(T_kR_k)^+\right)^{s}-\sum_{k=1}^{N}\left((T_kR_k)^+\right)^{s}\right|$$

for all $0<s<\sigma+\delta$ and some $\delta>0$. We will do so by distinguishing the cases

(i) $\sigma<1$, (ii) $\sigma=1$, (iii) $1<\sigma\leq 2$ and (iv) $\sigma>2$.

(i) If $\sigma<1$, then for each $0<s\leq 1$ (see also [10, proof of Lemma 4.9])

$$\begin{aligned}
L(s)\;&=\;\mathbb{E}\left[\sum_{k=1}^{N}\left((T_kR_k)^+\right)^{s}-\left(\sum_{k=1}^{N}(T_kR_k)^+\right)^{s}\right]\\
&\leq\;\mathbb{E}\left[\sum_{k=1}^{N}\left((T_kR_k)^+\right)^{s}-\max_{1\leq k\leq N}\left((T_kR_k)^+\right)^{s}\right]\\
&\leq\;\mathbb{E}\left[\sum_{k=1}^{N}|T_kR_k|^{s}-\max_{1\leq k\leq N}\left((T_kR_k)^+\right)^{s}-\max_{1\leq k\leq N}\left((T_kR_k)^-\right)^{s}\right]\\
&\leq\;\mathbb{E}\left[\sum_{k=1}^{N}|T_kR_k|^{s}-\max_{1\leq k\leq N}|T_kR_k|^{s}\right]\;=\;J(s),
\end{aligned}$$

and the latter function is finite by Lemma 9.

(ii) Next, let $\sigma=1$. Fix ζ such that $1-\delta_0<\zeta<1$ and $(1+\epsilon_0)\zeta>1$, where δ_0,ϵ_0 are given by condition (D*). Then choose $\delta<\min\{(1+\epsilon_0)\zeta-1,\zeta,2\zeta-1\}=(1+\epsilon_0)\zeta-1$. Let $1<s<1+\delta$ and note that $s-\zeta<1$. Applying Lemma 11 to $f(x) = x^s$ (thus $\xi = s-1$) and the ζ chosen above, we infer for a suitable constant $C \in \mathbb{R}_>$

$$\begin{aligned}
L(s) &= \mathbb{E}\left|\left(\sum_{k=1}^{N}(T_kR_k)^+\right)^s - \sum_{k=1}^{N}\left((T_kR_k)^+\right)^s\right| \\
&\le C\,\mathbb{E}\left[\sum_{j=1}^{N-1}\left(\sum_{k=1}^{j}|T_kR_k|\right)^{s-\zeta}|T_{j+1}R_{j+1}|^\zeta\right] \\
&= C\,\mathbb{E}\left[\sum_{j=1}^{N-1}\mathbb{E}_{\mathbf{T}}\left(\left(\sum_{k=1}^{j}|T_kR_k|\right)^{s-\zeta}|T_{j+1}R_{j+1}|^\zeta\right)\right] \\
&= C\,\mathbb{E}\,|R|^\zeta\;\mathbb{E}\left[\sum_{j=1}^{N-1}|T_{j+1}|^\zeta\,\mathbb{E}_{\mathbf{T}}\left(\sum_{k=1}^{j}|T_kR_k|\right)^{s-\zeta}\right] \\
&\le C\,\mathbb{E}\,|R|^\zeta\;\mathbb{E}\left[\sum_{j=1}^{N-1}|T_{j+1}|^\zeta\left(\mathbb{E}_{\mathbf{T}}(\sum_{k=1}^{j}|T_kR_k|)^\zeta\right)^{(s-\zeta)/\zeta}\right] \\
&\le C\,\mathbb{E}\,|R|^\zeta\;\mathbb{E}\left[\sum_{j=1}^{N-1}|T_{j+1}|^\zeta\left(\mathbb{E}_{\mathbf{T}}\sum_{k=1}^{j}|T_kR_k|^\zeta\right)^{(s-\zeta)/\zeta}\right] \\
&= C\,(\mathbb{E}\,|R|^\zeta)^{s/\zeta}\,\mathbb{E}\left[\left(\sum_{k=1}^{N}|T_k|^\zeta\right)^{s/\zeta}\right] \;<\; \infty
\end{aligned}$$

where Jensen's inequality and then subadditivity have been utilized in line 5. Finiteness of the final expectation is guaranteed by (D*).

(iii) Turning to the case $1 < \sigma < 2$, we proceed in the same manner. Applying again Lemma 11 to $f(x) = x^s$ for $0 < s < s_\infty \wedge 2$, but now with $\zeta = 1$, we obtain for some $C \in \mathbb{R}_>$

$$\begin{aligned}
L(s) &\le C\,\mathbb{E}|R|\,\mathbb{E}\left[\sum_{j=1}^{N}|T_j|\,\mathbb{E}_{\mathbf{T}}\left(\sum_{k=1}^{N}|T_kR_k|\right)^{s-1}\right] \\
&\le C\,\mathbb{E}|R|\,\mathbb{E}\left[\sum_{j=1}^{N}|T_j|\left(\mathbb{E}_{\mathbf{T}}\sum_{k=1}^{N}|T_kR_k|\right)^{s-1}\right] \\
&\le C(\mathbb{E}|R|)^s\,\mathbb{E}\left(\sum_{k=1}^{N}|T_k|\right)^s \;<\; \infty
\end{aligned}$$

where finiteness of the last expectation is guaranteed by (C).

(iv) Finally left with the case $\sigma \geq 2$, we fix again $\delta < 1$ sufficiently small such that $s + \delta < s_\infty$. For $s \in (\sigma, \sigma + \delta)$ and small $\theta > 0$, define

$$p(\theta) := \frac{\sigma}{s-2} - \theta \quad \text{and} \quad q(\theta) := \frac{p(\theta)}{p(\theta) - 1} = \frac{\sigma - \theta(s-2)}{2 + 2\theta - (s-\sigma) - \theta s}.$$

As one can readily check, $\lim_{\theta\to 0} p(\theta) > 1$ and $1 < \lim_{\theta\to 0} q(\theta) < \sigma$. So we may fix $\theta > 0$ so small (depending on δ) that $p = p(\theta)$ and $q = q(\theta)$ for this θ satisfy

$$1 < p < \infty, \quad 1 < q < \sigma \quad \text{and} \quad (s-2)p < \sigma.$$

In the following estimation, C denotes a generic finite positive constant which may differ from line to line. Using Lemma 12 from the Appendix with $f(x) = x^s$, we obtain

$$\begin{aligned}
L(s) &\leq C\,\mathbb{E}\left[\left(\sum_{i=1}^N (T_i R_i)^+\right)^{s-2} \sum_{1\leq j\neq k\leq N} (T_j R_j)^+ (T_k R_k)^+\right] \\
&\leq C\,\mathbb{E}\left[\left(\sum_{i=1}^N |T_i R_i|\right)^{s-2} \sum_{1\leq j\neq k\leq N} |T_j R_j|\,|T_k R_k|\right] \\
&= C\,\mathbb{E}\left(\mathbb{E}_{\mathbf{T}}\left[\left(\sum_{i=1}^N |T_i R_i|\right)^{s-2} \sum_{1\leq j\neq k\leq N} |T_j R_j|\,|T_k R_k|\right]\right) \\
&= C\,\mathbb{E}\left(\sum_{1\leq k\neq l\leq N} \mathbb{E}_{\mathbf{T}}\left[\left(\sum_{i=1}^N |T_i R_i|\right)^{s-2} |T_k R_k|\,|T_l R_l|\right]\right) \\
&\leq C\,\mathbb{E}\left(\sum_{1\leq k\neq l\leq N} \left(\mathbb{E}_{\mathbf{T}}\left(\sum_{i=1}^N |T_i R_i|\right)^{p(s-2)}\right)^{1/p} \left(\mathbb{E}_{\mathbf{T}}\,|T_k R_k|^q\,|T_l R_l|^q\right)^{1/q}\right) \\
&\leq C\,\mathbb{E}\left[\sum_{1\leq k\neq l\leq N} \left(\left(\sum_{i=1}^N |T_i|\right)^{p(s-2)} \mathbb{E}\,|R|^{p(s-2)}\right)^{1/p} \left(\mathbb{E}\,|R|^q\right)^{2/q} |T_k|\,|T_l|\right] \\
&= C\left(\mathbb{E}\,|R|^{p(s-2)}\right)^{1/p}\left(\mathbb{E}\,|R|^q\right)^{2/q}\mathbb{E}\left[\left(\sum_{i=1}^N |T_i|\right)^{s-2}\left(\sum_{1\leq k\neq l\leq N} |T_k|\,|T_l|\right)\right] \\
&\leq C\left(\mathbb{E}\,|R|^{p(s-2)}\right)^{1/p}\left(\mathbb{E}\,|R|^q\right)^{2/q}\mathbb{E}\left[\left(\sum_{i=1}^N |T_i|\right)^{s-2}\left(\sum_{j=1}^N |T_j|\right)^2\right] \\
&= C\left(\mathbb{E}\,|R|^{p(s-2)}\right)^{1/p}\left(\mathbb{E}\,|R|^q\right)^{2/q}\mathbb{E}\left(\sum_{k=1}^N |T_k|\right)^s < \infty
\end{aligned}$$

where Lemma 5 has been used for line 6. □

The previous proof gives rise to a Corollary which may be interesting in its own right:

Corollary 2. *Let* $(R_k)_{k\geq 1}$ *be a sequence iid random variables independent of the random weights* $(T_k)_{k\geq 1}$*. Let* $\sigma > 1$*,* $0 < \delta < 1$ *and suppose that* $\mathbb{E}|R_1|^s < \infty$ *for* $s < \sigma$ *and* $\mathbb{E}(\sum_{k=1}^N |T_k|)^{\sigma+\delta} < \infty$*. Then*

$$\mathbb{E}\left|\left(\sum_{k=1}^{N}(T_k R_k)^+\right)^s - \sum_{k=1}^{N}((T_k R_k)^+)^s\right| < \infty$$

for all $\sigma < s \leq \sigma + \delta$*.*

Proof. If $\sigma \geq 2$ or $\sigma + \delta \leq 2$, then the result is contained in the proof of Lemma 10. If $\sigma < 2$, but $s := \sigma + \delta > 2$, then observe that case (iv) also works when $\sigma < 2 < s$. □

Remark 2. In the case when $\sigma > 1$ is not an integer, the finiteness of $L(s)$ for $0 < s < \sigma + \delta$ and some $\delta > 0$ sufficiently small may alternatively be inferred by the same arguments as in [10, Proof of Lemma 5.2].

Appendix

Proof (of Lemma 6). We have the uniform bound

$$\int_0^\infty \left|t^{z-1} f(t)\right| dt = \int_0^\infty t^{\Re z - 1} |f(t)|\, dt$$
$$\leq \int_0^1 t^{\sigma_0 - 1} |f(t)|\, dt + \int_1^\infty t^{\sigma_1 - 1} |f(t)|\, dt < \infty.$$

In order to show holomorphicity, take any closed path c in the strip $\sigma_0 < \Re z < \sigma_1$, then we may use Fubini's theorem to infer

$$\int_c g(z) dz = \int_c \left(\int_0^\infty t^{z-1} f(t) dt\right) dz$$
$$= \int_0^\infty \left(\int_c t^{z-1} dz\right) f(t) dt = 0.$$

In fact, g is the Mellin-Transform of the measure $f(t)dt$. □

Lemma 11. *Let* $f : \mathbb{R}_\geq \to \mathbb{R}_\geq$ *be a differentiable function such that* $f(0) = 0$ *and* f' *is Hölder continuous of order* $\xi \in (0, 1]$*, i.e.*

$$|f'(x_1) - f'(x_2)| \leq C|x_1 - x_2|^\xi$$

for some $C \in \mathbb{R}_>$ *and all* $x_1, x_2 \in \mathbb{R}_\geq$. *Then*

$$\left| f(s_n) - \sum_{k=1}^{n} f(x_k) \right| \leq C \sum_{j=1}^{n-1} s_j^{1+\xi-\zeta} x_{j+1}^{\zeta} \tag{A.1}$$

for any $\frac{1+\xi}{2} \leq \zeta \leq 1$ *and* $x_1, \ldots, x_n \in \mathbb{R}_\geq$, *where* $s_n := \sum_{j=1}^{n} x_j$.

Proof. We will use induction over $n \geq 2$. For $n = 2$, use $f(0) = 0$ to obtain

$$|f(x+y) - f(x) - f(y)| = \left| \int_0^1 \left[f'(x+sy) - f'(sy) \right] y \, ds \right| \leq C x^{\xi} y, \tag{A.2}$$

for all $x, y \in \mathbb{R}_\geq$ which gives the result if $\zeta = 1$. Otherwise, pick any $0 < \sigma < 1$. Then (A.2) provides us with

$$\begin{aligned} |f(x+y) - f(x) - f(y)|^2 &\leq (C x^{\xi} y)^{1+\sigma} (C x y^{\xi})^{1-\sigma} \\ &= C^2 x^{\xi(1+\sigma)+1-\sigma} y^{\xi(1-\sigma)+1+\sigma}, \end{aligned}$$

which proves (A.1) for $n = 2$ with $\zeta = \frac{\xi(1-\sigma)+1+\sigma}{2}$. For the inductive step $n-1 \to n$, we note that

$$\begin{aligned} \left| f(s_n) - \sum_{j=1}^{n} f(x_j) \right| &\leq |f(s_n) - f(s_{n-1}) - f(x_n)| + \left| f(s_{n-1}) - \sum_{j=1}^{n-1} f(x_j) \right| \\ &\leq C \left(s_{n-1}^{1+\xi-\zeta} x_n^{\zeta} + \sum_{j=1}^{n-2} s_j^{1+\xi-\zeta} x_{j+1}^{\zeta} \right) \\ &= C \sum_{j=1}^{n-1} s_j^{1+\xi-\zeta} x_{j+1}^{\zeta} \end{aligned}$$

which proves our claim. □

Lemma 12. *Let* $f : \mathbb{R}_\geq \to \mathbb{R}_\geq$ *be a twice continuously differentiable function such that* f'' *is nonnegative and increasing. Then*

$$\left| f(s_n) - \sum_{k=1}^{n} f(x_k) \right| \leq f''(s_n) \sum_{i \neq j} x_i x_j. \tag{A.3}$$

for all $x_1, \ldots, x_n \in \mathbb{R}_\geq$, *where* $s_n := \sum_{j=1}^{n} x_j$.

Proof. We will use induction over $n \geq 2$. For $n = 2$, use $f(0) = 0$ to obtain

$$\begin{aligned} f(x+y) - f(x) - f(y) &= \int_0^1 \left[f'(x+sy) - f'(sy)\right] y\, ds \\ &= \int_0^1 \left(\int_0^1 \frac{d}{dr} f'(rx+sy)\, dr\right) y\, ds \\ &= \int_0^1 \int_0^1 f''(rx+sy)\, xy\, dr\, ds \end{aligned}$$

By assumption $f''(rx + sy) \leq f''(x + y)$ for all $r, s \in [0, 1]$, whence

$$0 \leq f(x+y) - f(x) - f(y) \leq f''(x+y)\, xy$$

as asserted. For the inductive step $n - 1 \to n$, we note that

$$\begin{aligned} \left| f(s_n) - \sum_{j=1}^{n} f(x_j) \right| &\leq |f(s_n) - f(s_{n-1}) - f(x_n)| + \left| f(s_{n-1}) - \sum_{j=1}^{n-1} f(x_j) \right| \\ &\leq f''(s_n)\, x_n\, s_{n-1} + f''(s_{n-1}) \sum_{1 \leq i \neq j \leq n-1} x_i x_j \\ &\leq f''(s_n) \sum_{1 \leq i \neq j \leq n} x_i x_j . \end{aligned}$$

which proves our claim for general $n \geq 2$. □

Acknowledgements We thank the referee for a very careful reading of the manuscript. G.A. and S.M. were supported by Deutsche Forschungsgemeinschaft (SFB 878). E. D. was supported by MNiSW grant N N201 393937.

References

1. Alsmeyer, G.: The smoothing transform: a review of contraction results. In: Alsmeyer, G., Löwe, M. (eds.) Random Matrices and Iterated Random Functions. Springer, Heidelberg (2013)
2. Alsmeyer, G., Meiners, M.: Fixed points of inhomogeneous smoothing transforms. J. Differ. Equ. Appl. **18**(8), 1287–1304 (2012)
3. Alsmeyer, G., Meiners, M.: Fixed points of the smoothing transform: two-sided solutions. Probab. Theory Relat. Fields **155**(1–2), 165–199 (2013)
4. Alsmeyer, G., Biggins, J.D., Meiners, M.: The functional equation of the smoothing transform. Ann. Probab. **40**(5), 2069–2105 (2012)

5. Buraczewski, D., Damek, E., Guivarc'h, Y., Hulanicki, A., Urban, R.: Tail-homogeneity of stationary measures for some multidimensional stochastic recursions. Probab. Theory Relat. Fields **145**, 385–420 (2009)
6. Caliebe, A., Rösler, U.: Fixed points with finite variance of a smoothing transformation. Stoch. Process. Appl. **107**(1), 105–129 (2003)
7. Durrett, R., Liggett, T.M.: Fixed points of the smoothing transformation. Z. Wahrsch. Verw. Gebiete **64**(3), 275–301 (1983)
8. Goldie, C.M.: Implicit renewal theory and tails of solutions of random equations. Ann. Appl. Probab. **1**, 126–166 (1991)
9. Guivarc'h, Y.: Sur une extension de la notion de loi semi-stable. Ann. Inst. H. Poincaré Probab. Statist. **26**(2), 261–285 (1990)
10. Jelenkovic, P.R., Olvera-Cravioto, M.: Implicit renewal theorem for trees with general weights. Stoch. Process. Appl. **122**(9), 3209–3238 (2012)
11. Jelenkovic, P.R., Olvera-Cravioto, M.: Implicit Renewal Theory and Power Tails on Trees. Adv. Appl. Probab. **44**(2), 528–561 (2012)
12. Jelenkovic, P.R., Olvera-Cravioto, M.: Power laws on weighted branching trees. In: Alsmeyer, G., Löwe, M. (eds.) Random Matrices and Iterated Random Functions. Springer, Heidelberg (2013)
13. Kesten, H.: Random difference equations and renewal theory for products of random matrices. Acta Math. **131**, 207–248 (1973)
14. Liu, Q.: On generalized multiplicative cascades. Stoch. Process. Appl. **86**(2), 263–286 (2000)
15. Meiners, M.: Fixed points of multivariate smoothing transforms and generalized equations of stability for continuous-time stochastic processes. (2012, Preprint)
16. Neininger, R., Rüschendorf, L.: A general limit theorem for recursive algorithms and combinatorial structures. Ann. Appl. Prob. **14**(1), 378–418 (2004)
17. Rösler, U.: A fixed point theorem for distributions. Stoch. Process. Appl. **42**(2), 195–214 (1992)
18. Widder, D.V.: The Laplace Transform. Princeton Mathematical Series, vol. 6. Princeton University Press, Princeton (1946)

Conditioned Random Walk in Weyl Chambers and Renewal Theory

C. Lecouvey, E. Lesigne, and M. Peigné

Abstract We present here the main result from [8] and explain how to use Kashiwara crystal basis theory to associate a random walk to each minuscule irreducible representation V of a simple Lie algebra; the generalized Pitman transform defined in [10] for similar random walks with uniform distributions yields yet a Markov chain when the crystal attached to V is endowed with a probability distribution compatible with its weight graduation. The main probabilistic argument in our proof is a quotient version of a renewal theorem that we state in the context of general random walks in a lattice [8]. We present some explicit examples, which can be computed using insertion schemes on tableaux described in [9].

1 Introduction

1.1 The Pitman Transform for the Brownian Motion

Let $(B(t))_{t\geq 0}$ be a standard Brownian motion on $\mathbb{R}$ starting at 0. We denote by $m(t)$ the minimum process defined by $m(t) := \inf_{0\leq s\leq t} B(s)$. The *Pitman transform* of $(B(t))_{t\geq 0}$ is given by

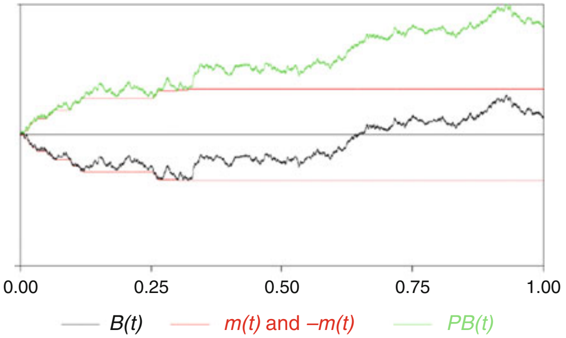
$$\mathcal{P}B(t) := B(t) - 2m(t).$$

The reader will find a proof of the following statement in [11] and references therein:

C. Lecouvey (✉) · E. Lesigne · M. Peigné
Laboratoire de Mathématiques et Physique Théorique, Parc de Grandmont 37200 Tours, France
e-mail: lecouvey@lmpt.univ-tours.fr; lesigne@lmpt.univ-tours.fr; peigne@lmpt.univ-tours.fr
The figures have been drawn with the help of J.R. Licois

G. Alsmeyer and M. Löwe (eds.), *Random Matrices and Iterated Random Functions*, Springer Proceedings in Mathematics & Statistics 53,
DOI 10.1007/978-3-642-38806-4_11,

Theorem 1. *The process $\mathcal{P}B(t)$ is a 3-dimensional Bessel process; in particular, it has the same law as the Brownian motion on $]0, +\infty[$ conditioned to stay positive.*



A Brownian motion trajectory *B(t)* and its Pitman transform *PB(t)*

There exists a multi-dimensional generalization of this theorem, called the *generalized Pitman transform* (see [1]): it corresponds for instance to the motion of the eigenvalues of some Hermitian Brownian motion in $SU(2)$.

1.2 The Pitman Transform for the Simple Random Walk

We consider the simple random walk $S_k := X_1 + \cdots + X_k$ on the set $\mathbb{Z}$ with increments ± 1:

$$\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = \frac{1}{2}.$$

The *Pitman transform* of this random walk is the process $(\mathcal{P}_k := S_k - 2m_k)_{k\geq 0}$ where $m_k = \min(0, S_1, \cdots, S_n)$; it is a Markov chain on $\mathbb{N}$ with transition probabilities

$$\forall a \in \mathbb{N} \qquad p(a, a+1) = \frac{a+2}{2(a+1)} \quad \text{and} \quad p(a, a-1) = \frac{a}{2(a+1)}.$$

By a straightforward computation, one gets

$$\forall a \in \mathbb{N} \qquad p(a, a \pm 1) = \lim_{k\to+\infty} \mathbb{P}(S_1 = a \pm 1/S_0 = a, S_1 \geq 0, \cdots, S_k \geq 0).$$

To obtain this equality, one may notice for instance that for any $a, k \geq 0$ one gets

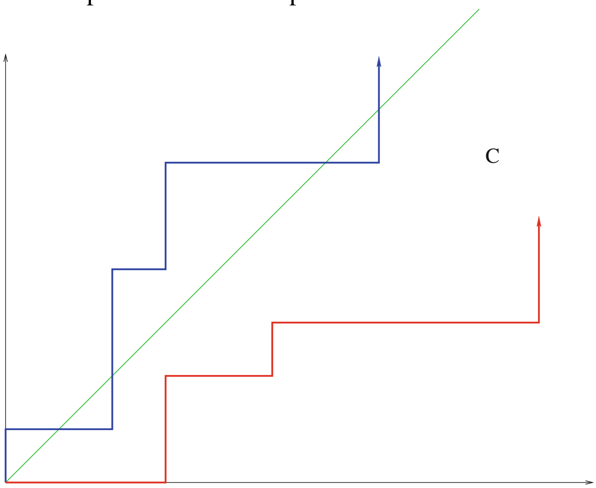
$$\mathbb{P}(S_1 = a+1/S_0 = a, S_1, \cdots, S_k \geq 0) = \frac{\mathbb{P}(m_{k-1} \geq -a-1)}{\mathbb{P}(m_{k-1} \geq -a-1) + \mathbb{P}(m_{k-1} \geq -a+1)}$$

and use classical estimations of the probability $\mathbb{P}(m_{k-1} \geq -a - 1)$ for the simple random walk.

We may represent the process $(\mathcal{P}_k)_{k\geq 0}$ as a process in the plane. We fix the standard basis $\{\overrightarrow{i}, \overrightarrow{j}\}$ in $\mathbb{R}^2$; the vector $\overrightarrow{i}$ corresponds to the step $+1$ and $\overrightarrow{j}$ to the step -1. We consider for instance the following trajectory

time k	0	1	2	3	4	5	6	7	⋯
X_k		-1	1	1	-1	-1	-1	1	⋯
S_k	0	-1	0	1	0	-1	-2	-1	⋯
path in $\mathbb{Z}^2$	$\vec{0}$	$\vec{j}$	$\vec{i}+\vec{j}$	$2\vec{i}+\vec{j}$	$2\vec{i}+2\vec{j}$	$2\vec{i}+3\vec{j}$	$2\vec{i}+4\vec{j}$	$3\vec{i}+4\vec{j}$	⋯
m_k	0	-1	-1	-1	-1	-1	-2	-2	⋯
$\mathcal{P}_k = S_k - 2m_k$	0	1	2	3	2	1	2	3	⋯
Pitman path in $\mathbb{N}^2$	$\vec{0}$	$\vec{i}$	$2\vec{i}$	$3\vec{i}$	$3\vec{i}+\vec{j}$	$3\vec{i}+2\vec{j}$	$4\vec{i}+2\vec{j}$	$5\vec{i}+2\vec{j}$	⋯

and its geometrical representation in the plane



2 The Ballot Problem in $\mathbb{R}^n, n \geq 2$

The Pitman transform of the simple random walk on $\mathbb{Z}$ can be seen as a transform of some process on $\mathbb{N}^2$, the so-called "Bertrand's ballot problem" in combinatorics. We generalize here this correspondence in any dimension.

2.1 Cones and Paths

We fix a basis $\mathcal{B} = \{\overrightarrow{e}_1, \cdots, \overrightarrow{e}_n\}$ of $\mathbb{R}^n$ and introduce the following cone

$$\mathcal{C} = \{x \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n \geq 0\};$$

we denote by $\mathring{\mathcal{C}} := \{x \in \mathbb{R}^n \mid x_1 > \cdots > x_n > 0\}$ its interior. We will consider the collection of paths in $\mathbb{Z}^n = \bigoplus_{i=1}^{n} \mathbb{Z}\overrightarrow{e}_i$ starting at 0 and with increments in $\{\overrightarrow{e}_1, \dots, \overrightarrow{e}_n\}$.

A *path* of length ℓ in $\mathbb{Z}^n$ will be a *word* $w = x_1 \cdots x_\ell$ on the alphabet $\{1, \dots, n\}$ and its *weight* the n-uple $\mathrm{wt}(w) = (\mu_1, \dots, \mu_n)$ where μ_i is the number of letters i in w. We are interested with (infinite) paths which remain inside $\mathcal{C}$, that is to say (infinite) words $w = x_1 x_2 \cdots$ such that, for any $\ell \geq 1$ and $i \in \{1, \cdots, n-1\}$ the number of i in $\{x_1, \dots, x_\ell\}$ is greater or equal to the number of $i+1$.

Example. The word $w = 112321231$ has weight $(4, 3, 2)$ and the corresponding path remains in $\mathcal{C}$.

2.2 The Simple Random Walk on $\mathbb{N}^n$

We fix a probability vector $p = (p_1, \cdots, p_n)$ in $\mathbb{R}^n$ (that is $p_i \geq 0$ for any $1 \leq i \leq n$ and $p_1 + \cdots + p_n = 1$) and consider a sequence $(X_\ell)_{\ell \geq 1}$ of i. i. d. random variables defined on a probability space $(\Omega, \mathcal{T}, \mathbb{P})$ such that

$$\forall i \in \{1, \cdots, n\} \quad \mathbb{P}(X_\ell = \overrightarrow{e}_i) = p_i.$$

The random walk $(S_\ell = X_1 + \cdots + X_\ell)_{\ell \geq 0}$ has the transition probability matrix

$$\Pi(\alpha, \beta) = \begin{cases} p_i & \text{if} \qquad \beta - \alpha = \overrightarrow{e}_i \in \mathcal{B}, \\ 0 & \text{otherwise.} \end{cases}$$

If $\beta := \alpha + \ell_1 \overrightarrow{e}_1 + \cdots + \ell_n \overrightarrow{e}_n$, with $\alpha \in \mathbb{N}^n, \ell_1 \cdots, \ell_n \geq 0$, all the paths joining α to β have length $\ell = \ell_1 + \cdots + \ell_n$ and the same probability $p_1^{\ell_1} \times \cdots \times p_n^{\ell_n}$; then

$$\Pi^\ell(\alpha, \beta) = \frac{\ell!}{\ell_1! \cdots \ell_n!} p_1^{\ell_1} \times \cdots \times p_n^{\ell_n}.$$

2.3 The Conditioned Random Walk in $\mathcal{C}$

Let $\Pi_{\mathcal{C}}$ be the restriction of Π to the cone $\mathcal{C}$. One gets the

Proposition 1. *If* $m := \mathbb{E}(X_\ell) \in \mathring{\mathcal{C}}$ *(or equivalently* $p_1 > \cdots > p_n$*) then*

$$\forall \lambda \in \mathcal{C} \quad \mathbb{P}_\lambda\Big(S_\ell \in \mathcal{C}, \forall \ell \geq 0\Big) > 0.$$

Moreover, the function $h : \lambda \mapsto \mathbb{P}_\lambda\Big(S_\ell \in \mathcal{C}, \forall \ell \geq 0\Big)$ *is* $\Pi_{\mathcal{C}}$*-harmonic.*

The transition matrix $P_{\mathcal{C}}$ of the random walk $(S_\ell)_{\ell \geq 0}$ conditioned to stay inside $\mathcal{C}$ is the h-Doob transform of $\Pi_{\mathcal{C}}$ given by:

$$\forall \lambda, \mu \in \mathcal{C} \quad P_{\mathcal{C}}(\lambda, \mu) = \frac{h(\mu)}{h(\lambda)} \Pi_{\mathcal{C}}(\lambda, \mu).$$

The aim of this work is to explain how to compute $P_{\mathcal{C}}$ and the value of $h(\lambda), \lambda \in \mathcal{C}$, when $m = (p_1, \cdots, p_n) \in \overset{\circ}{\mathcal{C}}$. Following N. O'Connell [10], we will use the representation theory and generalize the Pitman transform in this discrete context.

2.4 The Representation Theory of $\mathfrak{sl}_n(\mathbb{C})$

2.4.1 Weights of $\mathfrak{sl}_n(\mathbb{C})$

We consider the *weight lattice* $P := \mathbb{Z}^n = \bigoplus_{i=1}^n \mathbb{Z}\overrightarrow{e}_i$; the cone of *dominant weights* is $P_+ := \bigoplus_{i=1}^n \mathbb{N}\overrightarrow{e}_i$. The *roots* of $\mathfrak{sl}_n(\mathbb{C})$ are the vectors $\pm(\overrightarrow{e}_i - \overrightarrow{e}_j)$ with $1 \leq i < j \leq n$; the set of *positive roots* is $R_+ := \{\overrightarrow{e}_i - \overrightarrow{e}_j, 1 \leq i < j \leq\}$ and the *simple roots* are the $n-1$ vectors $\overrightarrow{e}_i - \overrightarrow{e}_{i+1}, 1 \leq i \leq n-1$.
We denote by $\mathcal{I}$ the set of irreducible finite dimensional representations of $\mathfrak{sl}_n(\mathbb{C})$. It is a classical fact that the elements of $\mathcal{I}$ are labelled by the dominant weights: for any $\lambda \in P_+$, we denote by $V(\lambda)$ the corresponding irreducible finite dimensional representation and the map $\lambda \longleftrightarrow V(\lambda)$ is a one-to-one correspondence between P_+ and $\mathcal{I}$. For instance, the natural representation $\mathbb{C}^n$ of $\mathfrak{sl}_n(\mathbb{C})$ is labelled $V(1, \underbrace{0, \cdots, 0}_{n-1 \text{ times}})$, or simply $V(1)$.[1]
We now introduce some usual quantities in representation theory: the multiplicities f_μ and $f_{\mu/\lambda}$ related to the decompositions of the representations $V(1)^{\otimes \ell}$ and $V(\lambda) \otimes V(1)^{\otimes \ell}, \ell \geq 1, \lambda \in P_+$, in direct sum of elements of $\mathcal{I}$. Namely, for any $\ell \in \mathbb{N}$ and $\lambda \in P_+$, one has the decomposition

$$V(1)^{\otimes \ell} = \bigoplus_{\mu \in P_+} V(\mu)^{\oplus f_\mu} \quad \text{and} \quad V(\lambda) \otimes V(1)^{\otimes \ell} = \bigoplus_{\mu \in P_+} V(\mu)^{\oplus f_{\mu/\lambda}}.$$

This leads to the

[1] in order to simplify the notations, we will omit the (last) coordinates 0 which appear in $\lambda \in P_+$

Proposition 2. *For any $\lambda = (\lambda_1, \cdots, \lambda_n)$ and $\mu = (\mu_1, \cdots, \mu_n)$ in P_+ such that $\lambda \leq \mu$ (i-e $\lambda_i \leq \mu_i$ for any $i = 1, \cdots, n$), the number of paths between λ and μ which stay inside $\mathcal{C}$ is equal to $f_{\mu/\lambda}$; in particular, the number of paths between* 0 *and μ which stay inside $\mathcal{C}$ is equal to f_μ.*

Consequently, the representation theory is a powerful tool to compute the exact number of $paths$ staying inside $\mathcal{C}$; it will be also useful to estimate the probability of the set of such trajectories for a large class of random walks.

2.4.2 The Notion of Crystal

One may associate to each $V(\lambda) \in \mathcal{I}$ its *Kashiwara crystal* $B(\lambda)$. This is the combinatorial skeleton of the $U_q(\mathfrak{sl}_n(\mathbb{C}))$-module with dominant weight $\lambda \in P_+$: it has a structure of a colored and oriented graph (see [5, 6]).

Example. The crystal of $V(1) = \mathbb{C}^n$ is

$$B(1) : 1 \overset{1}{\to} 2 \overset{2}{\to} 3 \overset{3}{\to} \cdots \overset{n-2}{\to} n-1 \overset{n-1}{\to} n.$$

The crystal $B(\lambda) \otimes B(\mu)$ associated with $V(\lambda) \otimes V(\mu)$ may be constructed from $B(\lambda)$ and $B(\mu)$; its set of vertices is the direct product of the ones of $B(\lambda)$ and $B(\mu)$, the crystal structure (that is the choice of the arrows between vertices) being given by some technical rules presented for instance in [8], Theorem 5.1. One important property of the crystal theory is that the irreducible components of $V(\lambda) \otimes V(\mu)$ are in one-to-one correspondence with the connected components of $B(\lambda) \otimes B(\mu)$.

Example. The crystals $B(1)$ and $B(1)^{\otimes 2}$ for $\mathfrak{sl}_3(\mathbb{C})$

The crystal $B(1)$ of $V(1) = \mathbb{C}^3$ is $1 \overset{1}{\to} 2 \overset{2}{\to} 3.$

The crystal $B(1)^{\otimes 2}$ associated with $V(1)^{\otimes 2}$ is

$$\begin{array}{cccccc}
 & 1 & \overset{1}{\to} & 2 & \overset{2}{\to} & 3 \\
1 & 1\otimes 1 & \overset{1}{\to} & 2\otimes 1 & \overset{2}{\to} & 3\otimes 1 \\
1\downarrow & & & 1\downarrow & & 1\downarrow \\
2 & 1\otimes 2 & & 2\otimes 2 & \overset{2}{\to} & 3\otimes 2 \\
2\downarrow & 2\downarrow & & & & 2\downarrow \\
3 & 1\otimes 3 & \overset{1}{\to} & 2\otimes 3 & & 3\otimes 3
\end{array}$$

The two connected components are labelled by their source vertex, namely $1 \otimes 1$ and $1 \otimes 2$.

The letters which appear in the source vertex $1 \otimes 1$ are both equal to 1, this vertex corresponds to the irreducible component $V(2,0,0) \simeq V(2)$; in the same way, the source vertex $1 \otimes 2$ corresponds to $V(1,1,0) \simeq V(1,1)$; so

$$V(1)^{\otimes 2} \simeq V(2) \oplus V(1,1).$$

This corresponds to the fact that the two words (11) and (12) are the only ones "allowed" paths of length 2 which remain inside $\mathcal{C} \subset \mathbb{R}^3$.

2.4.3 Relation Between the Crystal and the Set of Words

The word $w = x_1 \cdots x_\ell$ on the alphabet $\{1, \ldots, n\}$ may be identified with the vertex

$$b = x_1 \otimes \cdots \otimes x_\ell \in B(1)^{\otimes \ell}$$

We denote by $B(b)$ the connected component of $B(1)^{\otimes \ell}$ which contains b. The *Pitman transform* will be the map $\mathcal{P}$ defined by

$$\begin{aligned} \mathcal{P} \; : \; B(1)^{\otimes \ell} &\to \mathcal{C} \\ b &\mapsto \text{ weight of the source vertex of } B(b). \end{aligned}$$

2.4.4 The Probability Distribution on the Crystal

The probability of the letter i is p_i; it will be the probability of the vertex $i \in B(1)$. The word $x_1 \cdots x_\ell$ has probability $p_1^{\mu_1} \cdots p_n^{\mu_n}$ where $(\mu_1, \cdots, \mu_n)$ is the weight of this word; this is also the probability of the vertex $b = x_1 \otimes \cdots \otimes x_\ell \in B(1)^{\otimes \ell}$. Finally, we have fixed a probability p on $B(1)$, endowed $B(1)^{\otimes \mathbb{N}}$ with $p^{\otimes \mathbb{N}}$ and set $(S_\ell) :=$ the sequence of weights of the corresponding process on $B(1)^{\otimes \mathbb{N}}$. The *Pitman process* $(\mathcal{H}_\ell)_\ell$ is the sequence of weights defined as the images by $\mathcal{P}$ of the k-vectors $(S_\ell)_{1 \leq \ell \leq k}, k \geq 1$.

2.4.5 The Character and the Schur Functions

Let $\mathfrak{h}$ be the sub-algebra of diagonal matrices of $\mathfrak{sl}_n(\mathbb{C})$; any representation M of $\mathfrak{sl}_n(\mathbb{C})$ may be decomposed in *weight spaces*

$$M := \bigoplus_{\mu \in P} M_\mu$$

with $M_\mu := \{v \in M / h(v) = \mu(v)v$ for any $h \in \mathfrak{h}\}$. The *character function of* M is the Laurent polynomial s_M defined by

$$\forall x \in \mathbb{C}^n \; s_M(x) := \sum_{\mu \in P} \dim M_\mu \; x^\mu$$

When M is an irreducible representation $V(\lambda)$, the character function is called the *Schur function* and denoted s_λ.

Example. The Schur function of the natural representation of $\mathfrak{sl}(n,\mathbb{C})$.

For any $\lambda = (\lambda_1, \cdots, \lambda_n) \in P_+$ and $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, we denote by $a_\lambda(x)$ the Vandermonde function

$$a_\lambda(x) := \det(x_i^{\lambda_j}) = \begin{vmatrix} x_1^{\lambda_1} & x_1^{\lambda_2} & \cdots & x_1^{\lambda_n} \\ x_2^{\lambda_1} & x_2^{\lambda_2} & \cdots & x_2^{\lambda_n} \\ \vdots & \vdots & \vdots & \vdots \\ x_n^{\lambda_1} & x_n^{\lambda_2} & \cdots & x_n^{\lambda_n} \end{vmatrix}.$$

For $\delta = (n-1, n-2, \cdots, 0)$, one gets

$$a_\delta(x) := \begin{vmatrix} x_1^{n-1} & x_1^{n-2} & \cdots & 1 \\ x_2^{n-1} & x_2^{n-2} & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^{n-1} & x_n^{n-2} & \cdots & 1 \end{vmatrix} = \prod_{1 \le i < j \le n} (x_i - x_j).$$

For any $\lambda \in P_+$, the Schur function s_λ of $V(\lambda)$ is given by

$$s_\lambda(x) := \frac{a_{\lambda+\delta}(x)}{a_\delta(x)}; \tag{1}$$

in particular, the Schur function of $V(1) = V(1, 0, \cdots, 0) = \mathbb{C}^n$ is

$$s_1(x) := \frac{a_{(1,0,\cdots,0)+\delta}(x)}{a_\delta(x)} = \frac{1}{a_\delta(x)} \times \begin{vmatrix} x_1^n & x_1^{n-2} & \cdots & 1 \\ x_2^n & x_2^{n-2} & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^n & x_n^{n-2} & \cdots & 1 \end{vmatrix} = x_1 + \cdots + x_n. \tag{2}$$

One may now state the following

Theorem 2 ([10]).

- *The process* $(\mathcal{H}_\ell)_{\ell \ge 0}$ *is a Markov chain on* $\mathcal{C}$ *with transition probability*

$$P_{\mathcal{H}}(\lambda, \mu) = \frac{s_\lambda(p_1, \cdots, p_n)}{s_\mu(p_1, \cdots, p_n)} 1_B(\mu - \lambda).$$

- *The transition matrix* $P_{\mathcal{C}}$ *of the r.w.* $(S_\ell)_{\ell \ge 0}$ *conditioned to stay inside* $\mathcal{C}$ *is equal to* $P_{\mathcal{H}}$.

- *In particular, one gets*

$$\mathbb{P}_0(S_\ell \in \mathcal{C}, \forall \ell \geq 0) = \prod_{\alpha \in R_+} (1 - p^{-\alpha}) = \prod_{1 \leq i < j \leq n} \left(1 - \frac{p_i}{p_j}\right).$$

3 The Probabilistic Argument

We present here the probabilistic ingredients which allow to link up the random walk conditioned to stay inside $\mathcal{C}$ and the Pitman process. The details are given for the ballot problem, as in [10], and remain valid in a more general context (see Sect. 4).

3.1 The Markov Chain $(\mathcal{H}_\ell)_{\ell \geq 0}$

Using the crystal basis theory, one may check that $(\mathcal{H}_\ell)_{\ell \geq 0}$ is a Markov chain with transition matrix

$$P_{\mathcal{H}}(\lambda, \mu) = f_{\mu/\lambda} \frac{s_\mu(p)}{s_\lambda(p) s_1(p)}$$

with $f_{\mu/\lambda} \in \{0, 1\}$; for the ballot problem, we have $s_1(p) = p_1 + \cdots + p_n = 1$ and so $P_{\mathcal{H}}(\lambda, \mu) = f_{\mu/\lambda} s_\mu(p)/s_\lambda(p)$.

We denote by $\Pi_{\mathcal{C}}$ the restriction of Π to the cone $\mathcal{C}$; the matrix $P_{\mathcal{H}}$ is the ψ-Doob transform of the substochastic matrix $\Pi_{\mathcal{C}}$ with $\psi(\lambda) := \frac{s_\lambda(p)}{p^\lambda}$. We are going to prove that ψ coincides up to a multiplicative constant with the function h given in Proposition 1.

3.2 The Doob Theorem

Let E be a countable set and Q sub-stochastic matrix transition on E. Let G be the Green kernel associated with Q. Fix an origin $x^* \in E$ such that $0 < G(x^*, y) < +\infty$ for any $y \in E$ and let K be the Martin kernel defined by

$$\forall x, y \in E \quad K(x, y) = \frac{G(x, y)}{G(x^*, y)}.$$

Let $\mathbf{h}$ be an strictly positive and Q-harmonic function on E, let $Q_{\mathbf{h}}$ be the $\mathbf{h}$-Doob transform of Q and consider a Markov chain $(Y_\ell^{\mathbf{h}})_{\ell \geq 0}$ on E with transition matrix $Q_{\mathbf{h}}$. One gets the classical following result:

Theorem 3 (Doob, [3]). *Let* $\mathbf{f} : E \to \mathbb{R}$ *such that*

$$\forall x \in E \qquad \lim_{\ell \to +\infty} K(x, Y_\ell^{\mathbf{h}}(\omega)) = \mathbf{f}(x) \quad \mathbb{P}(d\omega) - \text{a.s.}$$

Then there exits $c > 0$ *such that* $\mathbf{f} = c\mathbf{h}$.

In our case, we take $E = \mathcal{C}$ with origin $x^* = 0$, the sub-stochastic matrix Q is $\Pi_{\mathcal{C}}$ and $\mathbf{h}(\lambda) = h(\lambda) = \mathbb{P}_\lambda(S_\ell \in \mathcal{C}, \forall \ell \geq 0)$. By the Strong Law of Large Numbers, one gets

$$S_\ell \sim \ell m + o(\ell) \quad \mathbb{P} - \text{a.s.}$$

N. O' Connell directly checks, using the explicit expression of the Schur function s_λ given in (1), that for m inside the cone $\mathcal{C}$ and any sequence $\mu_\ell = \ell m + o(\ell)$

$$K(\lambda, \mu_\ell) = p^{-\lambda} \frac{f_{\mu_\ell/\lambda}}{f_{\mu_\ell}} \to \frac{s_\lambda(p)}{p^\lambda} \quad \text{as} \quad \ell \to +\infty.$$

Unfortunately, such an explicit formula for the Schur function does not exists in the more general situation we want to consider and we avoid his approach as follows: using the theory of crystal bases, we may decompose the Martin kernel and write

$$K(\lambda, \mu_\ell) = \frac{1}{p^\lambda} \sum_{\gamma \text{ weight of } V(\lambda)} f_{\gamma/\lambda} \times p^\gamma \times \frac{G(0, \mu_\ell - \gamma)}{G(0, \mu_\ell)}$$

for any λ and $\mu_\ell = \ell m + o(\ell) \in \mathcal{C}$ with ℓ large enough. It remains to prove that, for any $\gamma \in \mathcal{C}$

$$\frac{G(0, \mu_\ell - \gamma)}{G(0, \mu_\ell)} \to 1 \quad \text{when} \quad \ell \to +\infty.$$

3.3 A Quotient Renewal Theorem in the Cone

We consider here a sequence $(X_\ell)_{\ell \geq 1}$ of independent and identically distributed $\mathbb{Z}^n$-valued random variables with law μ and set $S_\ell := X_1 + \cdots + X_\ell$ for any $\ell \geq 1$. The central argument of our approach is the following:

Theorem 4 ([8, 9]). *Assume that the law* μ *is aperiodic on* $\mathbb{Z}^n$, *its support is bounded and the mean vector* $m := \mathbb{E}(X_\ell]$ *lies inside the cone* $\mathcal{C}$. *Let* $\alpha < 2/3$ *and* $(\mu_\ell)_\ell, (h_\ell)_\ell$ *be two sequences in* $\mathbb{Z}^n$ *such that* $\lim \ell^{-\alpha}\|\mu_\ell - \ell m\| = 0$ *and* $\lim \ell^{-1/2}\|h_\ell\| = 0$. *Then, when* ℓ *tends to infinity, we have*

$$\sum_{k\geq 1} \mathbb{P}\left(S_1 \in \mathcal{C}, \ldots, S_k \in \mathcal{C}, S_k = \mu_\ell + h_\ell\right) \sim \sum_{k\geq 1} \mathbb{P}\left(S_1 \in \mathcal{C}, \ldots, S_k \in \mathcal{C}, S_k = \mu_\ell\right).$$

The first ingredient of the proof is the following

Lemma 1 (R. Garbit (2008) [4]). *Under the same hypotheses as above, for any $\alpha > \frac{1}{2}$, there exists $c = c_\alpha > 0$ such that, for all ℓ large enough and $\mu \in \mathcal{C}$*

$$\mathbb{P}\left(S_1 - m \in \mathcal{C}, \ldots, S_\ell - \ell m \in \mathcal{C}, S_\ell = \ell m + \mu\right) \geq \exp\left(-c\ell^\alpha\right).$$

The second ingredient is a version of the renewal limit theorem due to H. Carlson and S. Wainger [2]. We assume that $m := \mathbb{E}(X_\ell)$ is nonzero. Let $(\overrightarrow{\epsilon_1}, \ldots, \overrightarrow{\epsilon_{n-1}})$ be an orthonormal basis of the hyperplane $m^\perp$. If $x \in \mathbb{R}^n$, denote by x' its orthogonal projection on $m^\perp$ expressed in this basis and let B be the covariance matrix of the random vector X'_ℓ. Let $\mathcal{N}_B$ be the $(n-1)$-dimensional Gaussian density with covariance matrix B and V be the n-dimensional volume of the fundamental domain of the group generated by the support of the law of X_ℓ. The following result may be deduced from [2], the proof of the present statement is detailed in [7]:

Theorem 5. *We assume the random variables X_ℓ have an exponential moment. Fix $\alpha < 2/3$ and let (μ_ℓ) be a sequence of real numbers such that $\mu_\ell = m\ell + o(\ell^\alpha)$. Then, when ℓ goes to infinity, we have*

$$\sum_{k\geq 0} \mathbb{P}(S_k = \mu_\ell) \sim \frac{V}{\|m\|} \ell^{-(n-1)/2} \mathcal{N}_B\left(\frac{1}{\sqrt{\ell}}\mu'\right).$$

We will apply this result along the sequences $(\mu_\ell)_\ell = (S_\ell(\omega))_\ell$ for almost all $\omega \in \Omega$, which is possible since, for any $\epsilon > 0$, one gets $S_\ell \sim \ell m + o(\ell^{\frac{1}{2}+\epsilon})$ a.s.

4 Generalization: The Pitman Transform for Minuscule Representations

We consider in [8] a representation $V(\delta)$ of a simple Lie algebra $\mathfrak{g}$ over $\mathbb{C}$ and endow the associated crystal $B(\delta)$ with a probability distribution $p = (p_b)_{b\in B(\delta)}$ which is compatible with the weight graduation of $B(\delta)$. We consider a random walk $(S_\ell)_{\ell\geq 0}$ with independent increments of law p and transition matrix Π; this random walk will take values in the weight lattice P associated with $\mathfrak{g}$, we will have $P \subset \frac{1}{r}\mathbb{Z}^n$ for some $\in \mathbb{N}^*$ depending on $\mathfrak{g}$ (see the table below).

As in the previous section, we construct a Markov chain $(\mathcal{H}_\ell)_\ell$ in the Weyl chamber $\mathcal{C} \subset P$, with transition matrix $P_\mathcal{H}$, which will play the role of the Pitman process. We prove that $(\mathcal{H}_\ell)_\ell$ coincides with the ψ-Doob transform of the restriction to $\mathcal{C}$ of the transition matrix of $(S_\ell)_\ell$ (for some explicit function ψ expressed in terms of characters) if and only if $V(\delta)$ is *minuscule*, that is when the orbit of δ under the action of the Weyl group of $\mathfrak{g}$ contains all the weights of $V(\delta)$. The minuscule representations are given in the following table

Type	Minuscule weights	N	Decomposition on the basis B
A_n	$\omega_i, i = 1,\ldots,n$	$n+1$	$\omega_i = \varepsilon_1 + \cdots + \varepsilon_i$
B_n	ω_n	n	$\omega_n = \frac{1}{2}(\varepsilon_1 + \cdots + \varepsilon_n)$
C_n	ω_1	n	$\omega_1 = \varepsilon_1$
D_n	$\omega_1, \omega_{n-1}, \omega_n$	n	$\omega_1 = \varepsilon_1, \omega_{n+t} = \frac{1}{2}(\varepsilon_1 + \cdots + \varepsilon_n) + t\varepsilon_n, t \in \{-1, 0\}$
E_6	ω_1, ω_6	8	$\omega_1 = \frac{2}{3}(\varepsilon_8 - \varepsilon_7 - \varepsilon_6), \omega_6 = \frac{1}{3}(\varepsilon_8 - \varepsilon_7 - \varepsilon_6) + \varepsilon_5$
E_7	ω_7	8	$\omega_7 = \varepsilon_6 + \frac{1}{2}(\varepsilon_8 - \varepsilon_7)$.

When $V(\delta)$ is minuscule, we also prove that for any m in the interior $\mathring{\mathcal{C}}$ of $\mathcal{C}$, one may choose the probability $p = (p_b)_{b \in B(\delta)}$ on the crystal $B(\delta)$ (and so the random walk $(S_\ell)_{l \geq 0}$ on $\mathbb{Z}^n$) in such a way its drift is m.

The main result of [8] and [9] may be stated as follows

Theorem 6 ([8, 9]). *If the representation $V(\delta)$ is minuscule and $m = \mathbb{E}(X) \in \mathring{\mathcal{C}}$, then the transition matrix of the r.w. $(S_\ell)_{\ell \geq 0}$ conditioned to stay inside $\mathcal{C}$ is equal to $P_{\mathcal{H}}$. In particular, for any $\lambda \in P_+$, one gets*

$$\mathbb{P}_\lambda(S_\ell \in C, \forall \ell \geq 0) = p^{-\lambda} s_\lambda(p) \prod_{\alpha \in R_+} (1 - p^{-\alpha}).$$

Furthermore, when $\mu^{(\ell)} = \ell m + o(\ell^\alpha)$ with $\alpha < 2/3$, one gets

$$\lim_{\ell \to \infty} \frac{f^{\ell}_{\mu^{(\ell)}/\lambda}}{f^{\ell}_{\mu^{(\ell)},\lambda}} = s_\lambda(p).$$

The same result holds for direct sums of distinct minuscule representations and also for some *super Lie algebras*, for instance $\mathfrak{g}(m, n)$ (see [7]).

Example. Case of a C_2 representation: $\mathfrak{sp}(4, \mathbb{C})$.

We consider the representation $V = V(\omega_1)$. The corresponding crystal is

$$B(\omega_1) : 1 \xrightarrow{1} 2 \xrightarrow{1} \overline{2} \xrightarrow{1} \overline{1}.$$

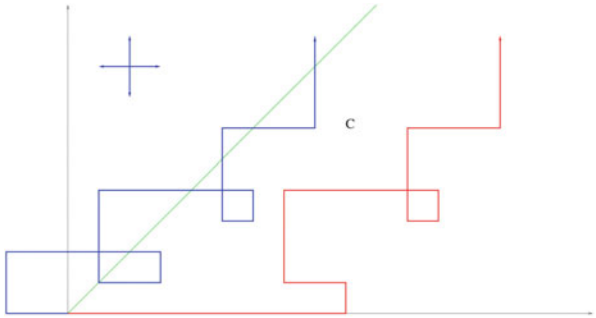
The probability $p = (p_{\vec{e}_1}, p_{-\vec{e}_1}, p_{\vec{e}_2}, p_{-\vec{e}_2})$ is such that

$$p_{\vec{e}_1} \times p_{-\vec{e}_1} = p_{\vec{e}_2} \times p_{-\vec{e}_2}.$$

In this case, one fixes $0 < p_2 < p_1 < 1$ with $p_1 + p_2 < 1$ and sets

$$p_{\vec{e}_1} = p_1, p_{-\vec{e}_1} = \frac{c}{p_1}, p_{\vec{e}_2} = p_2 \text{ and } p_{-\vec{e}_2} = \frac{c}{p_2}$$

with $c = p_1 p_2 (\frac{1}{p_1 + p_2} - 1)$ (so that $p_1 + p_2 + \frac{c}{p_1} + \frac{c}{p_2} = 1$).



A random path in the plane and its Pitman transform, for the vectorial representation of $\mathfrak{sp}(4, \mathbb{C})$

$$\mathbb{P}_0\Big(S_\ell \in \mathcal{C}, \forall \ell \geq 1\Big) = \Big(1 - \frac{p_2}{p_1}\Big)\Big(1 - \frac{c}{p_1 p_2}\Big)\Big(1 - \frac{c}{p_1}\Big)\Big(1 - \frac{c}{p_2}\Big).$$

Acknowledgements The authors thank the referee for his very careful reading and many helpful remarks and comments on a preliminary version of this text. M. Peigné especially acknowledges G. Alsmeyer and M. Löwe for their warm hospitality during the conference *Random matrices and iterated random functions* they organized in Münster, October 4–7, 2011 and where he presented this work.

References

1. Biane, P., Bougerol, P., O'Connell, N.: Littelmann paths and Brownian paths. Duke Math. J. **130**(1), 127–167 (2005)
2. Carlsson, H., Wainger, S.: On the multi-dimensional renewal theorem. J. Math. Anal. Appl. **100**(1), 316–322 (1984)
3. Doob, J.L.: Classical Potential Theory and Its Probabilistic Counterpart. Springer, New York (1984)
4. Garbit, G.: Temps de sortie d'un cône pour une marche aléatoire centrée, [Exit time of a centered random walk from a cone] C. R. A.S. Paris **345**(10), 587–591 (2007)
5. Kashiwara, M.: On crystal bases. Can. Math. Soc. Conf. Proc. **16**, 155–197 (1995)
6. Lecouvey, C.: Combinatorics of crystal graphs for the root systems of types A_n, B_n, C_n, D_n and G_2. Combinatorial Aspect of Integrable Systems. MSJ Memoirse, vol. 17, pp. 11–41. Mathametical Society of Japan, Tokyo (2007)
7. Lecouvey, C., Lesigne, E., Peigné, M.: Le théorème de renouvellement multi-dimensionnel dans le "cas lattice" (2010). http://hal.archives-ouvertes.fr,hal-00522875.2010
8. Lecouvey, C., Lesigne, E., Peigné, M.: Random walks in Weyl chambers and crystals. Proc. London Math. Soc. **104**(2), 323–358 (2012)
9. Lecouvey, C., Lesigne, E., Peigné, M.: Conditionned ballot problems and combinatorial representation theory (2012). http://arxiv.org/abs/1202.3604
10. O' Connell, N.: A path-transformation for ramdom walks and the Robison-Schensted correspondence. Trans. Amer. Math. Soc. **355**, 669–3697 (2003)
11. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion, Corrected 3rd printing. Springer, New York (2005)