Chapter 2
The Univariate Case

We consider the stochastic recurrence equation

$$X_t = A_t\, X_{t-1} + B_t\,, \quad t \in T\,, \tag{2.0.1}$$

for the index sets $T = \mathbb{N} = \{0, 1, \ldots\}$ or $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, where $\big((A_t, B_t)\big)_{t\in\mathbb{Z}}$ is an $\mathbb{R}^2$-valued iid sequence. If X_0 and $\big((A_t, B_t)\big)_{t\geq 1}$ are independent then iteration of (2.0.1) generates a Markov chain $(X_t)_{t\geq 0}$. This Markov chain does not necessarily constitute a strictly stationary sequence.[1] Our first task will be to find conditions for strict stationarity of this Markov chain. This is accomplished in Section 2.1. In particular, we elaborate on results by Vervaat [256] and Goldie and Maller [130].

In most applications one focuses on the stationary version of (2.0.1). But it is also of interest to study the properties of the Markov chain $(X_t)_{t\geq 0}$ such as irreducibility, aperiodicity, mixing properties, and absolute continuity of the Markov kernel. This is the task of Section 2.2. There we will also discuss the close relationship between the stationary solution $(X_t)_{t\in\mathbb{Z}}$ of (2.0.1) and the fixed-point equation in law

$$X_0 \stackrel{d}{=} A_1\, X_0 + B_1\,. \tag{2.0.2}$$

Immediately, stationarity of the Markov chain (X_t) implies (2.0.2). On the other hand, if the law P_{X_0} of X_0 satisfies (2.0.2) the Markov chain $(X_t)_{t\in\mathbb{Z}}$ generated by (2.0.1) is stationary. In Section 2.2 we also discuss the notion of contractivity in the context of the stochastic recurrence equation (2.0.1).

Sections 2.3–2.5 are devoted to the distributional properties of X_0 in (2.0.2) or, equivalently, to the properties of the marginal distribution of the stationary Markov chain (X_t). In Section 2.3 we collect some results about the existence and structure of

[1] Throughout, stationarity and strict stationarity are used as synonyms.

D. Buraczewski et al., *Stochastic Models with Power-Law Tails*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-29679-1_2

the moments of X_0. Section 2.4 deals with the complicated problem of the asymptotic tail behavior of P_{X_0}. We present an overview of the theory with focus on the results by Kesten [175] and Goldie [128] about the power-law asymptotics of $\mathbb{P}(|X_0| > x)$ and $\mathbb{P}(X_0 > x)$. The support of P_{X_0}, denoted by supp P_{X_0}, is studied in Section 2.5. Omitting the trivial case, when the distribution P_{X_0} is concentrated at one point, we prove that supp P_{X_0} coincides either with a half-sided infinite interval or with the whole real line, provided $|A_1|$ exceeds 1 with positive probability. Moreover the distribution P_{X_0} is atomless and either absolutely continuous or singular with respect to Lebesgue measure.

To ease notation the symbol Y stands for a generic element of any stationary sequence (Y_t). In particular, we write (A, B) for a generic element of the sequence $\big((A_t, B_t)\big)_{t\in\mathbb{Z}}$ and, if $(X_t)_{t\in\mathbb{Z}}$ is a stationary solution to (2.0.1), X denotes a generic element of this sequence. The identity in law (2.0.2) takes on the form $X \stackrel{d}{=} AX + B$ with the convention that (A, B) and X are independent.

2.1 Stationary Solution

2.1.1 Existence and Uniqueness of the Stationary Solution

Here we search for conditions that ensure the existence and uniqueness of a strictly stationary *causal* solution (X_t) to the stochastic recurrence equation (2.0.1). A solution (X_t) is *causal*[2] if, for every t, X_t is a measurable function of $(A_s, B_s)_{s\le t}$, i.e., it is a function of past and present noise variables (A_s, B_s), $s \le t$. Then (X_t) also constitutes a Markov chain.

Intuitively, a causal solution to (2.0.1) is obtained by backward iteration, i.e., by applying (2.0.1) backward in time. After n iterations one obtains for any $t \in \mathbb{Z}$,

$$\begin{aligned} X_t &= A_t\, X_{t-1} + B_t \\ &= A_t A_{t-1} X_{t-2} + \big(B_t + A_t\, B_{t-1}\big) \\ &= A_t A_{t-1} A_{t-2} X_{t-3} + \big(B_t + A_t\, B_{t-1} + A_t A_{t-1} B_{t-2}\big) \\ &\;\vdots \\ &= \Pi_{t-n+1,t} X_{t-n} + \sum_{i=t-n+1}^{t} \Pi_{i+1,t}\, B_i\,. \end{aligned} \tag{2.1.3}$$

[2] In the literature a causal solution is sometimes called *non-anticipative*; see e.g. Bougerol and Picard [51, 52], Babillot et al. [22]. We follow here the tradition of time series analysis, where it is common to refer to a causal solution if X_t is a function of the past and present noise.

Here and in what follows it will be convenient to use the notation

$$\Pi_{ij} = \begin{cases} A_i \cdots A_j & \text{if } i \le j\,, \\ 1 & \text{if } i > j\,, \end{cases} \quad \text{and} \quad \Pi_j = \Pi_{1j}\,, \quad i, j \in \mathbb{Z}\,.$$

In view of (2.1.3) a natural candidate for a causal stationary solution to (2.0.1) is suggested by the infinite series

$$X_t = \sum_{i=-\infty}^{t} \Pi_{i+1,t}\, B_i\,, \qquad t \in \mathbb{Z}\,. \tag{2.1.4}$$

Indeed, if this series converged a.s. for every t, (X_t) would satisfy (2.0.1). It suffices to show that the series converges a.s. for $t = 0$. Replacing the indices $-i$ by i, $\sum_{i=-\infty}^{0} \Pi_{i+1,0}\, B_i$ turns into $\sum_{i=0}^{\infty} \Pi_{0,i-1}\, B_i$. We will prove that the latter series converges a.s.

A complete solution of this problem can be found in Theorem 2.1 of Goldie and Maller [130]. We cite one part of this result.

Theorem 2.1.1 *Consider an iid $\mathbb{R}^2$-valued sequence $\big((A_t, B_t)\big)_{t\in\mathbb{Z}}$ and assume $\mathbb{P}(B = 0) < 1$ and $\mathbb{P}(A = 0) = 0$. Then the following conditions are equivalent:*

(1) *The infinite series $\sum_{i=1}^{\infty} \Pi_{i-1} B_i$ converges absolutely a.s.;*
(2) $\Pi_{n-1} B_n \stackrel{\text{a.s.}}{\to} 0$ *as* $n \to \infty$.

Each of the relations (1) *or* (2) *implies that*

$$\Pi_n\, Z_0 + \sum_{i=1}^{n} \Pi_{i-1}\, B_i \stackrel{\text{a.s.}}{\to} \sum_{i=1}^{\infty} \Pi_{i-1}\, B_i\,, \qquad n \to \infty\,,$$

for any random variable Z_0 independent of $\big((A_t, B_t)\big)_{t\ge 1}$. Moreover, $\Pi_n \stackrel{\text{a.s.}}{\to} 0$ is necessary for both statements (1) *and* (2)*.*

Conversely, if also $\mathbb{P}(Ax + B = x) < 1$ for all $x \in \mathbb{R}$, and (1) *or* (2) *do not hold then*

$$\Big|\Pi_n\, Z_0 + \sum_{i=1}^{n} \Pi_{i-1}\, B_i\Big| \stackrel{\mathbb{P}}{\to} \infty\,, \qquad n \to \infty\,,$$

for any random variable Z_0 independent of $\big((A_t, B_t)\big)_{t\ge 1}$.

We immediately get the following result.

Corollary 2.1.2 *Consider an iid* $\mathbb{R}^2$*-valued sequence* $\big((A_t, B_t)\big)_{t\in\mathbb{Z}}$ *and assume*

1. $\mathbb{P}(A = 0) = 0$*;*
2. $\mathbb{P}(Ax + B = x) < 1$ *for all* $x \in \mathbb{R}$.[3]

Then the following conditions are equivalent:

(1) *There exists an a.s.-unique causal ergodic strictly stationary solution* (X_t) *to the stochastic recurrence equation* (2.0.1).
(2) $\sum_{i=1}^{\infty} |\Pi_{i-1} B_i| < \infty$ *a.s.*
(3) $\Pi_{n-1} B_n \overset{\text{a.s.}}{\to} 0$ *as* $n \to \infty$.

The solution (X_t) *is given by the infinite series* (2.1.4) *which converges a.s. for every* $t \in \mathbb{Z}$.

Proof The equivalence of (2) and (3) follows from Theorem 2.1.1.

Relation (2) implies the a.s. convergence of the series (2.1.4) which constitutes a causal solution to (2.0.1). The series representation (2.1.4) yields a functional relationship of the type $X_t = f((A_{s+t}, B_{s+t}), s \in \mathbb{Z})$. Then strict stationarity is straightforward and ergodicity follows from standard theory for stationary processes; see e.g. Krengel [186], Proposition 4.3. If there is any other causal strictly stationary sequence $(\widetilde{X}_t)$ satisfying $\widetilde{X}_t = A_t \widetilde{X}_{t-1} + B_t$, $t \in \mathbb{Z}$, we have

$$X_t - \widetilde{X}_t = \Pi_{t-n+1,t}\,(X_{t-n} - \widetilde{X}_{t-n})\,, \qquad n \ge 1\,. \tag{2.1.5}$$

The right-hand side expression converges in probability to zero as $n \to \infty$. Indeed, we know from Theorem 2.1.1 that $\Pi_{t-n+1,t} \overset{d}{=} \Pi_n \overset{\text{a.s.}}{\to} 0$. Causality implies that $\Pi_{t-n+1,t}$ and $X_{t-n} - \widetilde{X}_{t-n}$ are independent for every $n \ge 1$. Since (X_t) and $(\widetilde{X}_t)$ are strictly stationary we finally conclude that the right-hand side of (2.1.5) converges to zero in probability as $n \to \infty$. This is possible only if $X_t - \widetilde{X}_t = 0$ a.s. for every t. Thus we proved that (2) implies (1).

Now assume that (1) holds but (2) does not. Theorem 2.1.1 implies that

$$\begin{aligned} |X_n| &= \Big|\Pi_n\, X_0 + \sum_{i=1}^{n} \Pi_{i+1,t}\, B_i\Big| \\ &\overset{d}{=} \Big|\Pi_n\, X_0 + \sum_{i=1}^{n} \Pi_{i-1}\, B_i\Big| \overset{\mathbb{P}}{\to} \infty\,, \qquad n \to \infty\,. \end{aligned}$$

In view of the stationarity of (X_n) this means that $|X_n| = \infty$ a.s. in contradiction to (1). □

[3]Choosing $x = 0$, this condition also excludes the case $B = 0$ a.s.

Next we give some sufficient conditions for the existence of a solution to the stochastic recurrence equation (2.0.1). These conditions are easily checked. In many situations, they are also necessary.

Theorem 2.1.3 *Consider an iid $\mathbb{R}^2$-valued sequence $\big((A_t, B_t)\big)_{t\in\mathbb{Z}}$ and assume that one of the following conditions holds:*

1. $\mathbb{P}(A = 0) > 0$;
2. $\mathbb{P}(A = 0) = 0$, $-\infty \le \mathbb{E}[\log|A|] < 0$ *and* $\mathbb{E}[\log^+|B|] < \infty$.

Then there exists an a.s.-unique causal ergodic strictly stationary solution to the stochastic recurrence equation (2.0.1)*. The solution* (X_t) *is given by the infinite series* (2.1.4) *which converges a.s. for every* $t \in \mathbb{Z}$.

Moreover, assume one of the following conditions:

3. $\mathbb{P}(A = 0) = 0$, $\mathbb{P}(B = 0) < 1$ *and* $0 \le \mathbb{E}[\log|A|] \le \infty$;
4. $\mathbb{E}[\log|A|] > -\infty$ *and* $\mathbb{E}[\log^+|B|] = \infty$.

Then no strictly stationary causal solution to (2.0.1) *exists.*

Proof First assume that $q_0 = \mathbb{P}(A = 0) > 0$. Then

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{i=1}^{\infty} \Pi_{i-1} B_i\Big| < \infty\Big) &\ge \mathbb{P}(\Pi_i = 0 \text{ for some } i \ge 1)\\ &= \sum_{i=1}^{\infty} \mathbb{P}(A_i = 0, |\Pi_{i-1}| > 0)\\ &= \sum_{i=1}^{\infty} q_0\,(1-q_0)^{i-1} = 1.\end{aligned}$$

Thus the infinite series $\sum_{i=1}^{\infty} \Pi_{i-1} B_i$ collapses into a finite sum with probability 1 and all arguments in the proof of Corollary 2.1.2 apply to ensure the existence of an a.s.-unique causal solution to (2.0.1).

Now assume the set of conditions 2. Write $T_0 = 0$ and $T_n = \sum_{i=1}^{n} \log|A_i|, n \ge 1$. In view of $\mathbb{E}[\log|A|] < 0$ this is a random walk with negative drift and the strong law of large numbers yields $n^{-1} T_n \stackrel{\text{a.s.}}{\to} \mathbb{E}[\log|A|]$ as $n \to \infty$. By the Borel–Cantelli lemma and since $\mathbb{E}[\log^+|B|] < \infty$, $n^{-1} \log^+|B_n| \stackrel{\text{a.s.}}{\to} 0$ as $n \to \infty$. Thus we have

$$\begin{aligned}\limsup_{n\to\infty} |\Pi_{n-1} B_n|^{1/n} &= \limsup_{n\to\infty} \mathrm{e}^{\,n^{-1}T_{n-1} + n^{-1}\log|B_n|}\mathbf{1}(|B_n| > 0)\\ &\le \mathrm{e}^{\,\mathbb{E}[\log|A|]} < 1\,.\end{aligned}$$

An application of the Cauchy root criterion implies that $\sum_{i=1}^{\infty} |\Pi_{i-1} B_i| < \infty$ a.s. The same arguments as in the proof of Corollary 2.1.2 ensure the existence of an a.s.-unique causal solution to (2.0.1).

Now assume the set of conditions 3. If $0 < \mathbb{E}[\log |A|] \leq \infty$ then the strong law of large numbers implies that

$$\Pi_n = \mathrm{e}^{T_n} \overset{\text{a.s.}}{\to} \infty\,, \qquad n \to \infty\,,$$

and therefore the necessary condition $\Pi_n \overset{\text{a.s.}}{\to} 0$ in Theorem 2.1.1 is violated. Hence no causal stationary solution (X_t) exists in this case. If $\mathbb{E}[\log |A|] = 0$ a similar remark applies. In this case, there exists a subsequence along which the random walk (T_n) converges to infinity. This follows from the recurrence of the random walk; see e.g. Feller [120], Chapter XII.

Finally, assume $\mathbb{E}[\log^+ |B|] = \infty$ and $\mathbb{E}[\log |A|] > -\infty$. We follow an argument in Vervaat [256], proof of Lemma 1.7. In view of the Borel–Cantelli lemma, the logarithmic moment condition on B is equivalent to

$$\mathbb{P}(\log |B_n| > n \,\log \varepsilon \quad \text{i.o.}) = \mathbb{P}(|B_n|^{1/n} > \varepsilon \quad \text{i.o.}) = 1 \qquad \text{for any } \varepsilon > 1.$$

Therefore $\limsup_{n\to\infty} |B_n|^{1/n} = \infty$ and we have

$$\begin{aligned}\limsup_{n\to\infty} |\Pi_{n-1} B_n|^{1/n} &= \limsup_{n\to\infty} \mathrm{e}^{\,n^{-1} T_{n-1} + n^{-1} \log |B_n|} \\ &= \mathrm{e}^{\,\mathbb{E}[\log |A|]} \limsup_{n\to\infty} \mathrm{e}^{\,n^{-1} \log |B_n|} \\ &= \infty \qquad \text{a.s.}\end{aligned}$$

An application of the Cauchy root criterion shows that the series $\sum_{i=1}^{\infty} \Pi_{n-1} B_n$ does not converge. This fact excludes the existence of a causal strictly stationary solution to (2.0.1). This concludes the proof. □

2.1.2 A Discussion of the Conditions of the Existence Results

Literature

Theorem 2.1.3 was proved in the literature under similar conditions; see Kesten [175], Vervaat [256], Goldie [128], Bougerol and Picard [51]. The sharpest conditions for the a.s. convergence of the infinite series $\sum_{i=1}^{\infty} \Pi_{i-1} B_i$ can be found in Goldie and Maller [130]; Theorem 2.1.1 above contains a few of these conditions. They provide necessary and sufficient conditions for (2.1.4) to converge a.s. which, in general, are not easy to check. However, the conditions of Theorem 2.1.3 cover many cases of interest and are often easy to verify. Goldie and Maller [130] give explicit credit to Vervaat's [256] paper and ideas in Grincevičius [136, 137].

A Necessary Condition for the Existence of a Stationary Solution

The condition $\mathbb{P}(|A| < 1) > 0$ is *necessary* for the existence of a stationary solution to the stochastic recurrence equation (2.0.1) under the natural condition $\mathbb{P}(B = 0) < 1$. Indeed, if $\mathbb{P}(|A| \geq 1) = 1$ the sequence (Π_n) cannot converge to zero, but $\Pi_n \stackrel{\text{a.s.}}{\rightarrow} 0$ as $n \to \infty$ is necessary for the existence of the stationary solution (X_t); see Theorem 2.1.1.

The Contractive, Critical, and Divergent Cases

The condition $\mathbb{E}[\log|A|] < 0$ describes the *contractive case*. Some intuition for this name will be given in Section 2.2.5. This moment condition is satisfied if $\mathbb{E}[|A|^p] < 1$ for some $p > 0$. Indeed, $\mathbb{E}[\log|A|] = -\infty$ if $\mathbb{P}(A = 0) > 0$ while for $\mathbb{P}(A = 0) = 0$ in view of Jensen's inequality,

$$p\,\mathbb{E}[\log|A|] = \mathbb{E}[\log(|A|^p)] \leq \log\,\mathbb{E}[|A|^p] < 0\,.$$

The intermediate case when $\mathbb{E}[\log|A|] < 0$ but $\mathbb{E}[\log^+|B|] = \infty$ and no stationary solution exists was also considered in the literature; see Buraczewski and Iksanov [80] and references therein.

If $A > 0$ and $\mathbb{E}[\log A] = 0$ one refers to the *critical case*. The second part of Theorem 2.1.3 indicates that, if an invariant measure of the Markov chain $(X_t)_{t\geq 0}$ with the dynamics (2.0.1) existed, it could not be a probability distribution. However, there exists an *infinite invariant Radon measure* for this chain and this measure is unique up to a multiplicative constant; see the discussion in Section 5.1.

In the *divergent case* when $\mathbb{E}[\log|A|] > 0$ one can still construct an infinite invariant Radon measure for the Markov chain $(X_t)_{t\geq 0}$ but such a measure is not unique. In fact, an infinite family of invariant measures exists; see Guivarc'h and Le Page [143].

Degenerate Solutions

There exist some degenerate solutions to (2.0.1). If $B = 0$ a.s. and the conditions on A in the first part of Theorem 2.1.3 are satisfied then $X_t = 0$ a.s., $t \in \mathbb{Z}$, is the only solution to (2.0.1). A rather artificial case appears when $x \in \mathbb{R}$ satisfies the relation $x = Ax + B$ a.s., i.e., $1 - A$ and B are proportional. Then (2.0.1) has the trivial solution $X_t = x$ a.s.

These degenerate cases are not of particular interest; we will typically exclude them from consideration.

An Extension: Stationary Ergodic Noise

Assume that $((A_t, B_t))_{t\in\mathbb{Z}}$ constitutes a strictly stationary ergodic sequence with generic element (A, B), $-\infty < c_0 = \mathbb{E}[\log|A|] < 0$ and $\mathbb{E}[\log^+|B|] < \infty$. Then the infinite series in (2.1.4) converges a.s. for every t and (X_t) constitutes a stationary ergodic sequence (cf. Krengel [186], Proposition 4.3) representing the a.s.-unique causal solution to (2.0.1). Indeed, as in the iid case, we observe that for any $\varepsilon > 0$,

$$|\Pi_{i-1}|\,|B_i| \le \mathrm{e}^{\,T_{i-1}}\mathrm{e}^{\,\log^+|B_i|}\mathbf{1}(|B_i|>0) \le \mathrm{e}^{\,(c_0+\varepsilon)i}\,, \tag{2.1.6}$$

for a.e. $\omega \in \Omega$ and sufficiently large i. Here we used the strong law of large numbers $i^{-1}T_{i-1} \stackrel{\text{a.s.}}{\to} c_0 < 0$ and

$$\frac{1}{i}\log^+|B_i| = \frac{1}{i}\sum_{j=1}^{i}\log^+|B_j| - \frac{1}{i}\sum_{j=1}^{i-1}\log^+|B_j| \stackrel{\text{a.s.}}{\to} \mathbb{E}[\log^+|B|] - \mathbb{E}[\log^+|B|] = 0\,.$$

Thus the right-hand side of (2.1.6) converges to zero exponentially fast as $i \to \infty$ with probability 1, provided $\varepsilon > 0$ is chosen such that $c_0 + \varepsilon < 0$. This fact was already observed by Brandt [57]; see also the monograph Brandt et al. [58] which includes the case of multivariate dependent $\big((\mathbf{A}_t, \mathbf{B}_t)\big)$.

Noncausal Solutions

In time series analysis it is common to consider noncausal solutions of difference equations as well. For example, the autoregressive equation $X_{t+1} = \varphi X_t + Z_{t+1}$, $t \in \mathbb{Z}$, with iid noise (Z_t) satisfying $\mathbb{E}[\log^+|Z|] < \infty$ has the strictly stationary solution $X_t = -\sum_{j=1}^{\infty}\varphi^{-j}Z_{t+j}$, $t \in \mathbb{Z}$, if and only if $|\varphi| > 1$; see Brockwell and Davis [61], p. 81. Indeed, we have $X_t = \varphi^{-1}X_{t+1} - \varphi^{-1}Z_{t+1}$ and forward iteration of this equation yields the desired form of the stationary solution. This solution depends only on the future values Z_{t+j}, $j \ge 1$, which, in a time series context, are not observable at time t. The latter fact is referred to as *noncausality*.

To the best of our knowledge, Theorem 2.1 in Vervaat [256] is the only result where the idea of a *noncausal* strictly stationary solution to the stochastic recurrence equation (2.0.1) was considered. Assuming $\mathbb{P}(A = 0) = 0$ and following the above argument for a noncausal autoregressive process, we can write $X_t = X_{t+1}/A_{t+1} - B_{t+1}/A_{t+1}$ and, iterating forward, an educated guess for the corresponding *noncausal* stationary solution X_t as a function of $((A_s, B_s))_{s>t}$ is given by the infinite series

$$X_t = -\sum_{i=t+1}^{\infty}\frac{B_i}{\Pi_{t+1,i}}\,, \quad t \in \mathbb{Z}\,. \tag{2.1.7}$$

Assuming that $0 < \mathbb{E}[\log|A|] \le \infty$ and $\mathbb{E}[\log^+|B/A|] < \infty$, the same proof as for Theorem 2.1.3 shows that the infinite series (2.1.7) converges a.s. for every t and constitutes a noncausal strictly stationary solution to (2.0.1). Note that the law of X_0 for this stationary process is not an invariant distribution for the Markov chain generated by the recursion (2.0.1).

2.1.3 Examples

In many concrete cases it is not difficult to verify the conditions of Theorem 2.1.3 for (A, B).

Example 2.1.4 Recall the setting of Example 1.0.2. For stationarity of $X_t = \beta^{Y_t} X_{t-1} + E_t,\ t \in \mathbb{Z}$, one has to verify that $\mathbb{E}[\log(\beta^Y)] = \mathbb{E}[Y] \log \beta < 0$ and $\mathbb{E}[\log^+ E_1] < \infty$. These conditions are trivially satisfied in view of $\beta \in [0, 1)$, $Y > 0$ a.s. and the fact that E_1 has an exponential distribution.

Example 2.1.5 Assume that $A = a \in (0, 1)$ a.s. and B is symmetric Bernoulli distributed on $\{-1, 1\}$, i.e., $\mathbb{P}(B = \pm 1) = 0.5$. It is easy to see that the conditions of Theorem 2.1.3 are satisfied. Hence the infinite series $\sum_{i=1}^{\infty} \Pi_{i-1} B_i = \sum_{j=0}^{\infty} a^j B_{j+1}$ converges a.s. The distributional properties of this series have attracted a lot of attention; see Examples 1.0.3 and 2.5.10.

The verification of the stationarity of a GARCH model is much more involved:

Example 2.1.6 (Stationarity of GARCH$(1, 1)$ process). Recall the setting of ARCH and GARCH processes (X_t) from Example 1.0.1. Strict stationarity of (X_t) follows from strict stationarity of the volatility sequence (σ_t). In the general GARCH(p, q) case, conditions for strict stationarity are rather subtle; see Theorem 4.1.9 on p. 146 and Corollary 4.1.12 on p. 148 for seminal results by Bougerol and Picard [52]. In the GARCH$(1, 1)$ case, one can use Theorem 2.1.3 to conclude that the conditions

$$\alpha_0 > 0 \quad \text{and} \quad \mathbb{E}[\log(\alpha_1 Z^2 + \beta_1)] < 0 \tag{2.1.8}$$

are necessary and sufficient for the existence of a non-vanishing a.s.-unique causal strictly stationary solution to the equation

$$\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2 (\alpha_1 Z_{t-1}^2 + \beta_1)\,, \quad t \in \mathbb{Z}\,. \tag{2.1.9}$$

Hence conditions for strict stationarity of (X_t) depend on the distribution of the noise (Z_t). These are in general not easily verified and one needs to involve numerical methods, but in view of Jensen's inequality and since $\mathbb{E}[Z^2] = 1$ by assumption,

$$\mathbb{E}[\log(\alpha_1 Z^2 + \beta_1)] \le \log\big(\mathbb{E}[\alpha_1 Z^2 + \beta_1]\big) = \log(\alpha_1 + \beta_1)\,.$$

Thus if $\alpha_0 > 0$ and $\alpha_1 + \beta_1 < 1$, (2.1.8) is satisfied, and then, taking expectations in (2.1.9),

$$\mathbb{E}[X^2] = \mathbb{E}[\sigma^2] = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} < \infty\,,$$

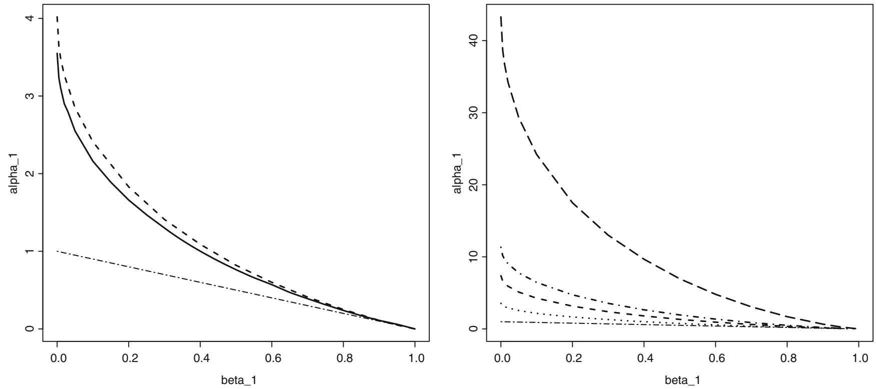

Figure 2.1 The area below the curves describes the region of (β_1, α_1)-values where the GARCH(1, 1) process is strictly stationary, i.e., the condition $\mathbb{E}[\log(\alpha_1 Z^2 + \beta_1)] < 0$ is satisfied. Left: The curves from top to bottom correspond to Student-distributed Z with 10 degrees of freedom, standard normal Z (solid line) and to the line $\alpha_1 + \beta_1 = 1$. For $\alpha_1 + \beta_1 < 1$ ($\alpha_1 + \beta_1 \geq 1$), the GARCH(1, 1) model has finite (infinite) variance. Right: The curves from top to bottom correspond to Student-distributed Z with 2.1, 2.5, 3, 10 degrees of freedom and to the line $\alpha_1 + \beta_1 = 1$. The Student distributions are standardized to variance 1.

while the same procedure yields $\mathbb{E}[\sigma^2] = \infty$ for $\alpha_1 + \beta_1 \geq 1$. Thus the condition $\alpha_1 + \beta_1 < 1$ ensures not only strict but also second-order stationarity of (σ_t) and (X_t), while the parameter choice $\alpha_1 + \beta_1 \geq 1$ goes beyond second-order stationarity.

In Figure 2.1 we visualize the (α_1, β_1)-regions for (standardized to unit variance) normal and Student-distributed Z, where the GARCH(1, 1) process with parameter (α_1, β_1) is strictly stationary.

The case $\alpha_1 + \beta_1 = 1$ has attracted some attention because one often estimates $(\alpha_1 + \beta_1)$-values close to one for real-life return data. This fact was the motivation for Engle and Bollerslev [114] to introduce the notion of integrated GARCH (IGARCH). An integrated ARMA process (Y_t) is nonstationary and becomes stationary after finitely many applications of the difference operation $Y_t - Y_{t-1}$; see Brockwell and Davis [61], Section 9.1. This is in contrast to an IGARCH(1,1) process, i.e., a GARCH(1, 1) process which satisfies the additional condition $\alpha_1 + \beta_1 = 1$. Under mild conditions it still constitutes a strictly stationary process because it can be shown that (2.1.8) remains valid for certain choices of pairs (α_1, β_1) satisfying $\alpha_1 + \beta_1 \geq 1$; see Nelson [224] for IGARCH(1,1) and Bougerol and Picard [52] in the general case. We refer to Section 4.1.2 for a detailed treatment of the stationarity problem for a general GARCH(p, q) process.

2.2 The Markov Chain

In this section we focus on the Markov chain property of the solution to the stochastic recurrence equation (2.0.1). We assume that X_0 is independent of the iid $\mathbb{R}^2$-valued sequence $\big((A_t, B_t)\big)_{t\ge 1}$. Then the dynamics described by (2.0.1) generate a Markov chain $(X_t)_{t\ge 0}$ with state space $E_0 \subset \mathbb{R}$. This Markov chain does not necessarily constitute a strictly stationary process. In what follows, we collect some of the basic properties of the Markov chain $(X_t)_{t\ge 0}$. A general reference to Markov chains with general state space is the monograph by Meyn and Tweedie [205], where one also finds the terminology used below.

2.2.1 Generalities

Transition Probabilities

The 1-*step transition probabilities* of the Markov chain $(X_t)_{t\ge 0}$ are given by the *kernel* or *transition operator*: for any $x \in E_0$ and any $C \in \mathscr{E}_0 = \mathscr{B}(E_0)$, the Borel σ-field of E_0,

$$\begin{aligned} P(x, C) &= \mathbb{P}(X_1 \in C \mid X_0 = x) \\ &= \mathbb{P}_x(X_1 \in C) \\ &= \mathbb{P}(Ax + B \in C)\,. \end{aligned} \tag{2.2.10}$$

Here and in what follows, we write $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot \mid X_0 = x)$ for $x \in E_0$, and $\mathbb{E}_x$ denotes the corresponding expected value. Similarly, for $n \ge 2$ the n-*step transition probability kernel* is given by

$$P^n(x, C) = \mathbb{P}(X_n \in C \mid X_0 = x) = \mathbb{P}_x(X_n \in C)\,.$$

Invariant Distribution, Stationary Distribution

A P-*invariant distribution* is a probability measure P_0 on E_0 satisfying

$$\int_{\mathbb{R}} P(x, C)\, P_0(dx) = P_0(C) \quad \text{for any } C \in \mathscr{E}_0. \tag{2.2.11}$$

This means that P_0 is the *invariant* or *stationary distribution* of the Markov chain (X_t). If X_0 has the distribution P_0 then the Markov chain $(X_t)_{t\ge 0}$ is strictly stationary and can be extended to a strictly stationary Markov chain $(X_t)_{t\in\mathbb{Z}}$; see Lemma 2.2.7.

Irreducibility

The Markov chain (X_t) is ν-*irreducible* for some σ-finite non-null measure ν on $(E_0, \mathscr{E}_0)$ if for any $C \in \mathscr{E}_0$ with $\nu(C) > 0$,

$$\mathbb{P}(X_n \in C \text{ for some } n \geq 1 \mid X_0 = x) > 0\,, \qquad x \in E_0\,. \tag{2.2.12}$$

Relation (2.2.12) turns into

$$\sum_{n=1}^{\infty} P^n(x, C) > 0\,, \qquad x \in E_0\,.$$

Typical choices for ν will be the stationary distribution P_0 or Lebesgue measure.

Harris Chain

The Markov chain $(X_t)_{t\geq 0}$ is *Harris recurrent* if it is ν-irreducible and for every Borel set $C \subset E_0$ with $\nu(C) > 0$,

$$\mathbb{P}_x(X_n \in C \quad \text{i.o.}) = 1\,, \qquad x \in E_0\,. \tag{2.2.13}$$

The chain is *positive* if it is ν-irreducible and admits a stationary distribution. Moreover, $(X_t)_{t\geq 0}$ is a *positive Harris chain* if it is Harris recurrent and positive.

Feller Chain

The Markov chain $(X_t)_{t\geq 0}$ is a *Feller chain* if the function $\mathbb{E}\big[f(X_1) \mid X_0 = x\big]$ is continuous for every bounded continuous function f on E_0. Since

$$\mathbb{E}\big[f(X_1) \mid X_0 = x\big] = \mathbb{E}_x[f(X_1)] = \mathbb{E}[f(Ax + B)]\,, \qquad x \in E_0\,, \tag{2.2.14}$$

this property is immediate by an application of dominated convergence.

In the following result we collect some relevant properties of the Markov chain $(X_t)_{t\geq 0}$ defined by the stochastic recurrence equation (2.0.1).

Proposition 2.2.1 *Assume that the following conditions hold.*

1. $\mathbb{P}(Ax + B = x) < 1$ *for all* $x \in \mathbb{R}$*;*
2. $-\infty \leq \mathbb{E}[\log|A|] < 0$ *and* $\mathbb{E}[\log^+|B|] < \infty$.

Then the Markov chain $(X_t)_{t\geq 0}$ *given by* (2.0.1) *has a unique stationary distribution* P_0.

Assume in addition the following condition:

3. *There exists an open Borel set* $C_0 \subset E_0$ *such that* $P_0(C_0) > 0$*, and for every* $x \in C_0$*,* $P(x, \cdot) = \mathbb{P}(Ax + B \in \cdot)$ *has an absolutely continuous component with*

respect to some σ-finite non-null measure ν on E_0, i.e., there exists a nonnegative measurable function f_x on E_0 such that

$$\int_{E_0} f_x(y)\,\nu(dy) > 0 \text{ and } P(x, C) \geq \int_C f_x(y)\,\nu(dy) \text{ for all Borel sets } C \subset E_0.$$

Then the Markov chain $(X_t)_{t\geq 0}$ is aperiodic, positive Harris and P_0-irreducible on E_0, i.e., for any set $C \in \mathscr{E}_0$ with $P_0(C) > 0$, relation (2.2.13) is satisfied.

Theorem 2.1 in Alsmeyer [5] shows that Proposition 2.2.1 remains valid if $P^n(x, \cdot)$ has an absolutely continuous component with respect to some measure ν on E_0 for some $n \geq 1$.

Proof The existence of a unique stationary distribution P_0 follows from Theorem 2.1.3. We will see in Section 2.5 below that the interior of the support of P_0 is nonempty. Under the Conditions 1.–3., Theorems 2.1, 2.2 and Corollary 2.3 in [5] yield aperiodicity and P_0-irreducibility of the Markov chain on E_0. □

We collect some straightforward sufficient conditions for this proposition.

Lemma 2.2.2 *The following conditions imply that $P(x, \cdot) = \mathbb{P}(A\,x + B \in \cdot)$, $x \in C_0 \in \mathscr{E}_0$, are absolutely continuous with respect to Lebesgue measure:*

1. *(A, B) has Lebesgue density.*
2. *A, B are independent and B has a Lebesgue density.*
3. *A, B are independent, $0 \notin C_0$ and A has a Lebesgue density.*
4. *$A = c\,B$ for some constant c, A has a Lebesgue density and $-1/c \notin C_0$.*

The proof of the following result is less elementary.

Lemma 2.2.3 *Assume that X solves the equation $X \stackrel{d}{=} AX + B$, $\mathbb{P}(X = 0) < 1$ and one of the following two conditions holds.*

1. *There exist intervals $I_1 = (a_1, a_2)$, $I_2 = (b_0 - \varepsilon, b_0 + \varepsilon)$ for some $a_1 < a_2$, b_0, $\varepsilon > 0$, a σ-finite measure ν_0 with b_0 in the support of ν_0 and a constant $c_0 > 0$ such that for any Borel sets $D_1, D_2 \subset \mathbb{R}$,*

$$P_{(A,B)}(D_1 \times D_2) \geq c_0\, \big|D_1 \cap I_1\big|\, \nu_0\big(D_2 \cap I_2\big)\,, \tag{2.2.15}$$

   where $|G|$ is the Lebesgue measure of a Borel set G.
2. *There exist intervals $I_1 = (a_0 - \varepsilon, a_0 + \varepsilon)$, $I_2 = (b_1, b_2)$ for some a_0, $b_1 < b_2$, $\varepsilon > 0$, a σ-finite measure ν_0 with a_0 in the support of ν_0 and a constant $c_0 > 0$ such that for any Borel sets $D_1, D_2 \subset \mathbb{R}$,*

$$P_{(A,B)}(D_1 \times D_2) \geq c_0\, \nu_0\big(D_1 \cap I_1\big)\, \big|D_2 \cap I_2\big|\,. \tag{2.2.16}$$

Then there exist nonempty open intervals C_0, J *such that* $P_X(C_0) > 0$ *and a constant* $c_1 > 0$ *such that for any Borel set* $D \subset \mathbb{R}$,

$$P(x, D) \geq c_1 \, |D \cap J| \, \mathbf{1}_{C_0}(x) \,. \tag{2.2.17}$$

Under the conditions of this lemma, we have for $x \in C_0$ that $|D \cap J| > 0$ implies that $P(x, D) > 0$. Hence $P(x, \cdot)$ has an absolutely continuous component with respect to Lebesgue measure on J. In view of Proposition 2.2.1 the Markov chain (X_t) is aperiodic and P_0-irreducible.

The bound in (2.2.17) is uniform for $x \in C_0$. This is much more than required in condition 3 of Proposition 2.2.1. An important consequence of the uniformity is that it yields a *regeneration scheme* for the Markov chain (X_t) in the sense of Athreya et al. [21].

Proof We prove only the first part of the lemma; the same argument also works in the second case. Fix a Borel set D. In view of (2.2.15) we observe that for all $x \neq 0$,

$$\begin{aligned} P(x, D) &= \int_{\mathbb{R}} \mathbf{1}_D(ax + b) \, P_{(A,B)}(d(a, b)) \\ &\geq c_0 \int_{I_1} \int_{I_2} \mathbf{1}_D(ax + b) \, da \, \nu_0(db) \\ &= c_0 \, x^{-1} \int_D \int_{I_2} \mathbf{1}_{I_1}\big(x^{-1}(z - b)\big) \, dz \, \nu_0(db) \,. \end{aligned}$$

Without loss of generality pick a positive x_0 from the support of P_X and define $C_0 = (x_0 - \varepsilon, x_0 + \varepsilon) \subset \mathbb{R}_+$ for some small $\varepsilon > 0$. Then we also have $P_X(C_0) > 0$. The set

$$J = \big((b_0 + \varepsilon) + (x_0 + \varepsilon) \, a_1 \, , (b_0 - \varepsilon) + (x_0 - \varepsilon) \, a_2\big)$$

is nonempty if ε is so small that $2\varepsilon + \varepsilon(a_1 + a_2) < x_0(a_2 - a_1)$. If $x \in C_0$, $z \in J$ and $b \in I_2$ then $x^{-1}(z - b) \in I_1$ and hence

$$\begin{aligned} P(x, D) &\geq c_0 \, x^{-1} \int_D \int_{I_2} \mathbf{1}_J(z) \, dz \, \nu_0(db) \\ &\geq c_0 \, (x_0 + \varepsilon)^{-1} \, \nu_0(I_2) \, |D \cap J| \,. \end{aligned}$$

The constant $c_1 = c_0 \, (x_0 + \varepsilon)^{-1} \, \nu_0(I_2)$ is positive since b_0 belongs to the support of ν_0 and therefore $\nu_0(I_2) > 0$.

This proves the lemma for positive x_0; the proof for negative x_0 is analogous. □

2.2.2 Mixing Properties

The Markov chain $(X_t)_{t\ge 0}$ is said to be *geometrically ergodic* if there exists some number $\rho_0 \in (0,1)$ such that

$$\big\| P^n(x,\cdot) - P_0(\cdot)\big\|_{\mathrm{TV}} = o(\rho_0^n)\,, \qquad n\to\infty\,,$$

where, as before, P_0 is the stationary distribution of the Markov chain and $\|\cdot\|_{\mathrm{TV}}$ is the *total variation distance* between probability measures.

Geometric ergodicity implies that the stationary Markov chain is *strongly* or *α-mixing with geometric rate*. This means that the stationary Markov chain $(X_t)_{t\ge 0}$ satisfies the relation

$$\frac{1}{4}\sup_{f,g:\|f\|_{L^\infty}\le 1\,,\|g\|_{L^\infty}\le 1}\big|\mathrm{cov}\big(f(X_0),g(X_n)\big)\big| =: \alpha_n \le c_0\,\rho_1^n\,, \qquad n\to\infty\,,$$

for some constants $c_0>0$ and $\rho_1\in(0,1)$. The supremum is taken over all measurable functions f and g with the property $\|f\|_{L^\infty}\le 1$ and $\|g\|_{L^\infty}\le 1$. This inequality follows, for example, from Theorem 16.1.5 in Meyn and Tweedie [205]. The function (α_n) is the *mixing rate function* of (X_t).

For the two-sided extension $(X_t)_{t\in\mathbb{Z}}$ of the stationary Markov chain $(X_t)_{t\ge 0}$ we also have the relation

$$\alpha_n = \frac{1}{4}\sup_{f,g:\|f\|_{L^\infty}\le 1,\|g\|_{L^\infty}\le 1}\big|\mathrm{cov}(f(\ldots,X_{-1},X_0),g(X_n,X_{n+1},\ldots))\big| \qquad (2.2.18)$$

$$= \sup_{C\in\sigma(X_s,s\le 0)\,,D\in\sigma(X_s,s\ge n)}\big|\mathbb{P}(C\cap D)-\mathbb{P}(C)\,\mathbb{P}(D)\big|\,, \qquad (2.2.19)$$

where $\sigma(X_s,\,s\in T)$ for some $T\subset\mathbb{Z}$ is the σ-field generated by $(X_s)_{s\in T}$. The last equality follows from Doukhan [102], p. 3, which we also recommend as a general reference to mixing properties of a stationary sequence. We mention in passing that a standard definition of the strong mixing property for a general stationary (not necessarily Markov) process is provided via the property $\alpha_n\to 0$ as $n\to\infty$, where (α_n) is given by (2.2.18) or (2.2.19).

The following result (Theorem 2.8 in Basrak et al. [27]) gives some sufficient conditions for the strong mixing property of the solution (X_t) to (2.0.1).

Proposition 2.2.4 *Assume that the following conditions hold:*

1. $\mathbb{E}[|A|^\varepsilon]<1$ *and* $\mathbb{E}[|B|^\varepsilon]<\infty$ *for some* $\varepsilon>0$*;*
2. $\mathbb{P}(Ax+B=x)<1$ *for all* $x\in\mathbb{R}$.

Then the stochastic recurrence equation (2.0.1) *has a strictly stationary causal solution* (X_t) *which constitutes a Markov chain. If this chain is* ν*-irreducible for some*

σ-finite non-null measure ν then (X_t) is geometrically ergodic and, hence, strongly mixing with geometric rate.

Proof Jensen's inequality implies that $\varepsilon\, \mathbb{E}[\log|A|] \le \log \mathbb{E}[|A|^\varepsilon] < 0$, and $\mathbb{E}[|B|^\varepsilon] < \infty$ ensures that $\mathbb{E}[\log^+|B|] < \infty$. Then Theorem 2.1.3 yields the existence of the stationary solution to (2.0.1).

To show geometric ergodicity, it suffices to check the three conditions of Theorem 1 in Feigin and Tweedie [119]:

- The chain has the Feller property.
- ν-irreducibility.
- A drift condition holds.

The Feller property was verified in the discussion on p. 20 and we assume ν-irreducibility. So it remains to verify the *drift condition*, i.e., there exist a compact set K and a nonnegative continuous function g such that $\nu(K) > 0$, $g(x) \ge 1$ on K, and for some $\beta \in (0, 1)$,

$$\mathbb{E}_x[g(X_1)] \le \beta\, g(x) \qquad \text{for all } x \in K^c.$$

We choose $g(x) = |x|^\varepsilon + 1$, $x \in \mathbb{R}$, where ε is given in the assumptions, and we also assume without loss of generality that $\varepsilon \in (0, 1]$. Then

$$\begin{aligned}\mathbb{E}_x[g(X_1)] &= \mathbb{E}[|Ax + B|^\varepsilon] + 1\\ &\le \mathbb{E}[|Ax|^\varepsilon] + \mathbb{E}[|B|^\varepsilon] + 1\\ &= \mathbb{E}[|A|^\varepsilon]\, g(x) + \big(\mathbb{E}[|B|^\varepsilon] - \mathbb{E}[|A|^\varepsilon] + 1\big)\,.\end{aligned}$$

Choose $K = [-m, m]$ and $m > 0$ so large that $\nu(K) > 0$ and

$$\mathbb{E}_x[g(X_1)] \le \beta\, g(x)\,, \qquad |x| > m\,,$$

for some constant $\beta \in (\mathbb{E}[|A|^\varepsilon], 1)$. This proves the drift condition and completes the argument. □

Example 2.2.5 In Example 1.0.1 we introduced the GARCH process and we also mentioned that the squares of an ARCH(1) process satisfy the relation

$$X_t^2 = \alpha_1\, Z_t^2\, X_{t-1}^2 + \alpha_0\, Z_t^2\,, \qquad t \in \mathbb{Z}\,, \tag{2.2.20}$$

where (Z_t) is an iid sequence with mean zero and variance one, and $\alpha_0, \alpha_1 > 0$ are chosen such that (X_t^2) is a non-vanishing stationary solution to (2.2.20). In this case, $A_t = \alpha_1 Z_t^2$ and $B_t = \alpha_0 Z_t^2$ are proportional. If P_Z has an absolutely continuous component with respect to Lebesgue measure then $P(x, \cdot) = \mathbb{P}(Z^2(\alpha_1 x + \alpha_0) \in \cdot)$

has an absolutely component for $x \geq 0$ and then Propositions 2.2.1 and 2.2.4 apply: (X_t^2) is an aperiodic, positive Harris and P_X-irreducible Markov chain which is strongly mixing with geometric rate.

In the GARCH(1, 1) case we know that (σ_t^2) satisfies the relation

$$\sigma_t^2 = \big(\alpha_1 Z_{t-1}^2 + \beta_1\big)\,\sigma_{t-1}^2 + \alpha_0\,, \tag{2.2.21}$$

where $\alpha_0, \alpha_1, \beta_1 > 0$ are chosen such that (σ_t^2) is a stationary solution to (2.2.21). Again, if P_Z has an absolutely continuous component then $P(x,\cdot) = \mathbb{P}\big((\alpha_1 Z^2 + \beta_1)\,x + \alpha_0 \in \cdot\big)$ has an absolutely continuous component for every $x > 0$ and then Propositions 2.2.1 and 2.2.4 apply: (σ_t^2) is an aperiodic, positive Harris and P_X-irreducible Markov chain which is strongly mixing with geometric rate.

Example 2.2.6 Andrews [15] showed that the stationary solution to the equation $X_t = a\,X_{t-1} + Z_t$, $t \in \mathbb{Z}$, with iid Bernoulli distributed (Z_t) does not satisfy the strong mixing condition when $a \in (0, 0.5]$.

2.2.3 The Fixed-Point Equation

If X has the stationary distribution P_0 and is independent of (A, B), then (2.2.11) is equivalent to the fixed-point equation

$$X \overset{d}{=} AX + B\,. \tag{2.2.22}$$

The following result is Lemma 3.3 in Bougerol and Picard [51]. It shows the intrinsic relationship between the strictly stationary causal solution to the stochastic recurrence equation (2.0.1) and the invariant distribution of the corresponding Markov chain.

Lemma 2.2.7 *There is a one-to-one correspondence between the strictly stationary causal solution $(X_t)_{t\in\mathbb{Z}}$ to (2.0.1) and the P-invariant distribution of the corresponding Markov chain $(X_t)_{t\geq 0}$.*

Proof First assume that $(X_t)_{t\in\mathbb{Z}}$ is a causal strictly stationary solution to (2.0.1) with generic element X which has the common law P_X. By causality, X_0 is independent of (A_1, B_1) and therefore for any Borel set $C \subset \mathbb{R}$,

$$P_X(C) = \mathbb{P}(X_1 \in C) = \mathbb{P}(A_1 X_0 + B_1 \in C) = \int_{\mathbb{R}} P(x, C)\,P_X(dx)\,.$$

Hence P_X is P-invariant.

On the other hand, if P_X is P-invariant, if X_0 has this distribution and is independent of (A_t, B_t), $t \geq 1$, then one can use the recursion $X_t = A_t X_{t-1} + B_t$ for $t \geq 1$ to construct the Markov chain $(X_t)_{t\geq 0}$ with transition kernel P. Since the law of X_0 is P-invariant this procedure defines a strictly stationary process on the nonnegative integers which is also a causal solution to (2.0.1). Finally, using standard theory (e.g., Krengel [186], Theorem 4.8), the Markov chain can be extended to the solution of (2.0.1) on the index set $\mathbb{Z}$. $\square$

The fixed-point equation (2.2.22) is handy if one wants to study the characteristics of the marginal distribution of a stationary solution (X_t) to (2.0.1), for example, its moments, tails, and support; see Sections 2.3–2.5. The identity in law (2.2.22) also merits its own interest. In various papers the law of X has been determined for a given distribution of (A, B). Some early examples can be found in Vervaat [256]. Dufresne [103, 104, 105, 106] developed an analytical theory for finding the law of X in some special cases. Marc Yor and coworkers have used the structure of the infinite series $\sum_{i=1}^{\infty} \Pi_{i-1} B_i$ to determine the law of X; see Hirsch and Yor [150] for an overview on the topic and some recent results which are closely related to the problem of determining the law of exponential functionals of Lévy processes; see also the discussion and references in Example 2.3.6. Of course, a trivial deterministic solution to (2.2.22) exists if B is proportional to $1 - A$, i.e., $x(1 - A) = B$ a.s. for some real x.

We consider some simple cases, where the distribution of X can be determined by applying different techniques.

Example 2.2.8 Recall the setting of Example 2.1.5: $A = a$ a.s. for some $a \in (0, 1)$ and $\mathbb{P}(B = \pm 1) = 0.5$. The distribution P_X of

$$X \stackrel{d}{=} \sum_{i=0}^{\infty} a^i B_i \tag{2.2.23}$$

is in general unknown. A real exception is the case $a = 0.5$. In this case, it is well known that X has uniform distribution on $(-2, 2)$; see Kallenberg [172], Lemma 2.20. We give a simple proof of this fact.

Relation (2.2.23) implies that $|X| \leq \sum_{i=0}^{\infty} 2^{-i} = 2$. Therefore the distribution of X is concentrated on $[-2, 2]$. If $X \stackrel{d}{=} 0.5X + B$ for independent B, X, the unique law P_X satisfies the following relation for any $x \in [-2, 2]$:

$$\begin{aligned} P_X([-2, x]) &= 0.5\, P_X([-2, 2(x - 1)]) + 0.5\, P_X([-2, 2(x + 1)]) \\ &= 0.5\, P_X([-2, 2(x - 1)])\, \mathbf{1}_{[0,2]}(x) + 0.5\, P_X([-2, 2(x + 1)])\, \mathbf{1}_{[-2,0]}(x)\,. \end{aligned} \tag{2.2.24}$$

On the other hand, if X is uniform on $(-2, 2)$ direct calculation yields (2.2.24) for any $x \in [-2, 2]$. The uniqueness of P_X in (2.2.24) implies that X has a uniform distribution on $(-2, 2)$.

Example 2.2.9 This example was considered by Vervaat [256]: assume that $A \geq 0$ a.s. and B are independent and B has a *strictly α-stable distribution* for some $\alpha \in (0, 2]$; see Feller [120], Samorodnitsky and Taqqu [249]. This means that for any $n \geq 2$ and positive constants c_i, $i = 1, \ldots, n$,

$$c_1\, B_1 + \cdots + c_n\, B_n \stackrel{d}{=} \big(c_1^\alpha + \cdots + c_n^\alpha\big)^{1/\alpha}\, B\,.$$

An application of this identity conditional on (A_t) yields

$$X \stackrel{d}{=} \sum_{i=1}^{\infty} \Pi_{i-1}\, B_i \stackrel{d}{=} B\,\Big(\sum_{i=1}^{\infty} \Pi_{i-1}^{\alpha}\Big)^{1/\alpha}\,,$$

provided $\sum_{i=1}^{\infty} \Pi_{i-1}^{\alpha} < \infty$ a.s. Hence the distribution of X is a scale mixture of the α-stable distribution P_B. However, in general we do not know the distribution of $Y_\alpha = \sum_{i=1}^{\infty} \Pi_{i-1}^{\alpha}$ and therefore it is difficult to determine the exact distribution of X. Notice that Y_α satisfies the identity in law $Y_\alpha \stackrel{d}{=} A^\alpha\, Y_\alpha + 1$ for independent A, Y_α.

Example 2.2.10 We consider another simple example from Vervaat [256] given for independent A, B with $1 - p_0 = \mathbb{P}(A = 0) = 1 - \mathbb{P}(A = 1) \in (0, 1)$. Write

$$\phi_Y(s) = \mathbb{E}[\mathrm{e}^{\,i\,s\,Y}]\,, \qquad s \in \mathbb{R}\,,$$

for the characteristic function of any random variable Y. Then the characteristic function of X satisfies the relation

$$\phi_X(s) = (1 - p_0)\,\phi_B(s)\; + p_0\,\phi_X(s)\,\phi_B(s)\,, \qquad s \in \mathbb{R}\,.$$

Therefore

$$\phi_X(s) = \frac{(1 - p_0)\,\phi_B(s)}{1 - p_0\,\phi_B(s)}\,, \qquad s \in \mathbb{R}\,,$$

and the distribution of X is essentially determined by the distribution of B. For example, if B is exponential with $\mathbb{P}(B \leq x) = 1 - \mathrm{e}^{\,-\lambda\,x}$, $x > 0$, for some $\lambda > 0$, then $\phi_B(s) = (1 - \lambda i s)^{-1}$, $s \in \mathbb{R}$, implying that

$$\phi_X(s) = \frac{1}{1 - (\lambda/(1 - p_0))is}\,, \qquad s \in \mathbb{R}\,. \tag{2.2.25}$$

In turn, $\mathbb{P}(X \leq x) = 1 - \mathrm{e}^{\,-(\lambda/(1-p_0))\,x}$, $x > 0$, i.e., X is exponential as well.

This fact can be checked in an alternative way. We have

$$X \stackrel{d}{=} \sum_{i=1}^{\infty} \Pi_{i-1} B_i \stackrel{d}{=} B_0 + \sum_{i=1}^{N} B_i \,,$$

where $N = \min\{n \geq 1 : A_1 \neq 0, \ldots, A_{n-1} \neq 0\,, A_n = 0\}$. The random variable N is independent of (B_i) and has a geometric distribution:

$$\mathbb{P}(N = n) = \mathbb{P}(A_1 \neq 0, \ldots, A_{n-1} \neq 0, A_n = 0) = (1 - p_0)\, p_0^{n-1}\,, \qquad n \geq 1\,.$$

Then

$$\phi_X(s) = \phi_B(s)\, \mathbb{E}\big[\phi_B(s))^N\big]\,, \qquad s \in \mathbb{R}\,,$$

and direct calculation yields the desired characteristic function (2.2.25).

Example 2.2.11 In this example we explain how the fixed-point equation (2.2.22) is related to harmonic functions on the upper half-plane

$$D = \mathbb{R}_+ \times \mathbb{R} = \{(a, b) : \ a > 0\,, b \in \mathbb{R}\}\,;$$

see Damek [91] and Damek and Hulanicki [92] for relevant references on this topic. On D, we consider the second-order elliptic differential operator

$$L = (a\,\partial_a)^2 - \alpha\, a\, \partial_a + a^2\, \partial_b^2\,, \qquad \alpha > 0\,. \tag{2.2.26}$$

A function g on D is called *L-harmonic* if $L\,g = 0$. Observe that L can also be written in the form

$$L = a^2\, \partial_a^2 + (1 - \alpha)\, a\, \partial_a + a^2\, \partial_b^2\,, \qquad \alpha > 0\,.$$

If $\alpha = 1$ an L-harmonic function is *harmonic* in the classical sense, i.e., with respect to the Laplace operator $\Delta = \partial_a^2 + \partial_b^2$. We aim at exploiting some invariance properties of the operator L.

The set D equipped with the multiplication

$$(a_1, b_1)\,(a_2, b_2) = (a_1 a_2, a_1 b_2 + b_1)\,, \qquad (a_i, b_i) \in D\,, \qquad i = 1, 2\,,$$

can be identified with the group $\mathrm{Aff}(\mathbb{R})$ of the affine transformations of the real line. Thus we may consider D as a group and, thanks to its topological structure, it is also a Lie group. Then L is a *left-invariant operator on D,* i.e.,

$$L \circ \tau_{(a_1, b_1)} = \tau_{(a_1, b_1)} \circ L\,,$$

where

$$(\tau_{(a_1,b_1)}\, g)(a,b) = g((a_1,b_1)\,(a,b))\,, \qquad g \in \mathbb{C}^2(D)\,.$$

According to Theorem 5.1 in Hunt [159], L generates a unique *convolution semigroup* of probability measures $(P_t)_{t\ge 0}$, satisfying the following conditions:

- $P_t * P_s = P_{t+s}$, where the convolution is taken in the group sense, i.e.,

$$(P_t * P_s)(C) = \int_{D\times D} \mathbf{1}\big((a_1,b_1)\,(a_2,b_2) \in C\big)\, P_t\big(d(a_1,b_1)\big)\, P_s\big(d(a_2,b_2)\big)\,;$$

- $P_t \stackrel{w}{\to} \varepsilon_{(0,1)}$ as $t \downarrow 0$, where $\varepsilon_{(0,1)}$ denotes the degenerate measure at $(0,1)$;
- for any $g \in \mathbb{C}^2_C(D)$, where $\mathbb{C}^2_C(D)$ is the space of twice continuously differentiable functions on D with compact support,

$$\lim_{t\to\infty} \|t^{-1}(g * P_t - g) - L\,g\|_{L^\infty} = 0\,.$$

Let $(Y_t)_{t\ge 0}$ be a stochastic process generated by L. Then P_t is the distribution of Y_t and (Y_t) has independent increments in the following sense: if $0 \le t_1 < t_2 < \cdots < t_k$ then[4] $Y_{t_1}, Y_{t_1}^{-1} Y_{t_2}, \ldots, Y_{t_{k-1}}^{-1} Y_{t_k}$ are independent random variables with the corresponding distributions $P_{t_1}, P_{t_2-t_1}, \ldots, P_{t_k-t_{k-1}}$. This makes (Y_t) a Lévy process on $\mathrm{Aff}(\mathbb{R})$; see Applebaum [16].

To find a connection with the fixed-point equation (2.2.22) we choose the measure P_1 from the semigroup and consider an iid sequence $\big((A_n,B_n)\big)_{n\in\mathbb{Z}}$ with marginal distribution

$$\check{P}_1(C) = P_1(C^{-1})\,, \qquad C \subset D\,,$$

where $C^{-1} = \{(a^{-1}, -a^{-1}b) : (a,b) \in C\}$. Using the properties of $\check{P}_1$, direct calculation yields that with $\mathbb{P} = \check{P}_1^{\mathbb{N}}$, $\mathbb{E}[\log^+ |B|] < \infty$ and $\mathbb{E}[\log A] < 0$; see Damek [91], Section 3. Hence there exists a solution X to the fixed-point equation (2.2.22) whose law P_X has the property

$$\int_{\mathbb{R}} \check{P}_1(A\,x + B \in \cdot)\, P_X(dx) = P_X(\cdot)\,; \tag{2.2.27}$$

also compare with (2.2.34) below. Equation (2.2.27) remains valid when $\check{P}_1$ is replaced by $\check{P}_t$ for any $t > 0$. Therefore, for every $f \in L^\infty(\mathbb{R})$, the function

$$g(a,b) = \int_{\mathbb{R}} f(a\,x + b)\, P_X(dx)\,, \qquad (a,b) \in D\,, \tag{2.2.28}$$

[4] For a fixed $\omega \in \Omega$, $Y_t^{-1}(\omega)$ is the inverse element of $Y_t(\omega)$ in the group $\mathrm{Aff}(\mathbb{R})$.

is bounded and L-harmonic. Conversely, every bounded L-harmonic function has the form (2.2.28) for some f; see [91]. Moreover, f is uniquely defined by g; it is called the *boundary value of* g. If f is continuous then it is just the limit (in the topological sense) of $g(b, a)$ as $a \downarrow 0$. The stationary solution X has a nice characterization in terms of a generalized Ornstein–Uhlenbeck process as defined in Behme and Lindner [32] or in Section 5.3 of Behme [31]. If (A_t, V_t) is the diffusion generated by the semigroup $(\check{P}_t)$ then (V_t) is such a process and $V_t \xrightarrow{d} X$ as $t \to \infty$.

The measure P_X is the *Poisson kernel.* It is the only probability measure that yields the representation (2.2.28) and has a smooth Lebesgue density f_X; see [91]. The density f_X can be computed explicitly and corresponds to a Student-t distribution with α degrees of freedom:

$$f_X(x) = \frac{\Gamma((\alpha+1)/2)}{\Gamma(\alpha/2)\sqrt{\pi}} \frac{1}{(1+x^2)^{(1+\alpha)/2}}, \qquad x \in \mathbb{R}. \tag{2.2.29}$$

In particular, if $\alpha = 1$ we obtain the classical *Poisson kernel* or *Cauchy density*

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \qquad x \in \mathbb{R}.$$

The density (2.2.29) can be derived from an ordinary differential equation for f_X. In view of (2.2.28) we have

$$\begin{aligned} g(a,b) &= \int_{\mathbb{R}} f(a\,x+b)\, f_X(x)\, dx \\ &= \int_{\mathbb{R}} f(x)\, a^{-1}\, f_X(a^{-1}(x-b))\, dx\,. \end{aligned}$$

The function g is L-harmonic for every $f \in L^\infty$, hence

$$L\big(a^{-1}\, f_X(a^{-1}\,(x-b))\big) = 0\,,$$

where x is fixed and L is applied to $a^{-1}\, f_X(a^{-1}\,(x-b))$ as a function of (a, b). In particular, for $x = 0$ we have

$$L\big(a^{-1}\, f_X(-a^{-1}b)\big) = 0\,.$$

Now apply the expression (2.2.26) for L, resulting in the ordinary differential equation

$$(1+b^2)\, f_X''(b) + (3+\alpha)\, b\, f_X'(b) + (1+\alpha)\, f_X(b) = 0\,. \tag{2.2.30}$$

For $\alpha = 1$, (2.2.30) turns into $\big((1+b^2)\, f_X(b)\big)'' = 0$, hence $f_X(b) = (1/\pi)\,(1+b^2)^{-1}$. For arbitrary α, we may try to find a solution of the form $f_X(b) = c\,(1+b^2)^{-\gamma}$ for some constants $c, \gamma > 0$. Indeed, (2.2.30) is satisfied for $\gamma = (\alpha + 1)/2$. Moreover, $f_X(b) = c\,(1 + b^2)^{-(\alpha+1)/2}$ is integrable, which proves that it must be the density of the Poisson kernel. In Guivarc'h and Le Page [143] the solution is calculated in a slightly different way.

We note that (2.2.28) is a particular case of the Poisson representation mentioned in the Preface on p. ix. Here $\mathbb{R}$ is an Aff$(\mathbb{R})$–space, i.e., $\mathbb{R}$ is equipped with an action of Aff$(\mathbb{R})$ defined by

$$(a, b) \circ x = a\,x + b\,.$$

By virtue of Raugi's [235] results, Aff$(\mathbb{R})$ can be replaced by a Lie group G and $\mathbb{R}$ by a quotient space of G. More precisely, let μ be a probability measure on G that is spread out. A bounded measurable function H is *μ-harmonic* if

$$H(h) = \int_G H(h\,g)\,\mu(dg)\,. \tag{2.2.31}$$

Raugi's results say that there exist a quotient space M of G and a probability measure ν on M, depending only on μ, such that (2.2.31) is equivalent to

$$H(g) = \int_M f(g\,x)\,\nu(dx) \qquad \text{for some function } f \in L^\infty(d\nu)\,,$$

where $x \to g\,x$ denotes the action of G on M as its quotient space, M is unique under an isomorphism (that is natural in this situation) and often appears to be a Lie subgroup of G.

2.2.4 ⊛-Convolution

Here we introduce the useful notion of *⊛-convolution* of the distribution $P_{(A,B)}$ and a Radon measure ν on $\mathbb{R}$: for any Borel set $C \subset \mathbb{R}$,

$$\begin{aligned}(P_{(A,B)} \circledast \nu)(C) &= \int_{\mathbb{R}} \mathbb{P}(Ax + B \in C)\,\nu(dx)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}^2} \mathbf{1}_C(ax+b)\,P_{(A,B)}(d(a,b))\,\nu(dx)\,.\end{aligned} \tag{2.2.32}$$

Then we can also introduce the *n-fold ⊛-convolution* $P^n_{(A,B)} \circledast \nu$ in a recursive way:

$$P^1_{(A,B)} \circledast \nu = P_{(A,B)} \circledast \nu \quad \text{and} \quad P^n_{(A,B)} \circledast \nu = P_{(A,B)} \circledast \left(P^{n-1}_{(A,B)} \circledast \nu\right), \quad n \ge 2\,.$$

The operation $\circledast$ is motivated by the distribution of the Markov chain $X_t = A_t X_{t-1} + B_t$, $t \geq 0$, where the distribution P_0 of X_0 does not necessarily have the stationary distribution of the Markov chain. Indeed, we have for any Borel set $C \subset \mathbb{R}$,

$$\big(P^n_{(A,B)} \circledast P_0\big)(C) = \mathbb{P}(A_n X_{n-1} + B_n \in C) = P_{X_n}(C)\,, \quad n \geq 1\,. \tag{2.2.33}$$

In particular, if P_0 is the stationary distribution of the Markov chain (X_t), we have $P_0 = P^n_{(A,B)} \circledast P_0$.

In Section 5.1 we will consider the critical case $\mathbb{E}[\log A] = 0$ for a Markov chain $(X_t)_{t\geq 0}$ given by the stochastic recurrence equation $X_t = A_t X_{t-1} + B_t$, $t \geq 0$. In this case, no P-invariant probability distribution P_0 exists. However, if we vary ν in (2.2.32) over some class of Radon measures it may be possible to find a unique (up to constant multiples) infinite Radon measure ν_0 which solves the equation

$$\nu(C) = (P_{(A,B)} \circledast \nu)(C) \quad \text{for any Borel set } C \subset \mathbb{R}. \tag{2.2.34}$$

The solution is again called the *P-invariant* or *invariant measure* of the Markov chain $(X_t)_{t\geq 0}$ which, in this case, does not constitute a strictly stationary process.

A Radon measure ν on $\mathbb{R}$ is uniquely determined by the totality of the integrals

$$\nu(f) = \int_{\mathbb{R}} f(x)\,\nu(dx)\,, \qquad f \in \mathbb{C}_C(\mathbb{R})\,,$$

where $\mathbb{C}_C(\mathbb{R})$ is the space of continuous functions on $\mathbb{R}$ with compact support; see Kallenberg [171], Lemma 1.4. Then an alternative way of writing (2.2.34) is given by

$$\begin{aligned}\nu(f) &= \big(P_{(A,B)} \circledast \nu\big)(f)\\ &= \int_{\mathbb{R}} \int_{\mathbb{R}^2} f(ax+b)\, P_{(A,B)}(d(a,b))\,\nu(dx)\,, \qquad f \in \mathbb{C}_C(\mathbb{R})\,.\end{aligned}$$

2.2.5 The Contractive Case

In Section 2.1 we coined the name *contractive case* for the situation when $\mathbb{E}[\log |A|] < 0$. Assuming this condition and also taking into account $\mathbb{E}[\log^+ |B|] < \infty$, an appeal to the proof of Theorem 2.1.3 shows that

$$X_n \stackrel{d}{=} \Pi_n\, X_0 + \sum_{i=1}^n \Pi_{i-1}\, B_i \ \stackrel{\text{a.s.}}{\to}\ \sum_{i=1}^\infty \Pi_{i-1}\, B_i = X\,, \tag{2.2.35}$$

and the a.s. convergence to the limit X is exponentially fast whatever the distribution of X_0. We observe that P_X is the stationary distribution of the Markov chain; this follows from Theorem 2.1.3.

Similarly, for $x \neq y$ we have

$$|X_n^x - X_n^y| = |\Pi_n| \, |x - y| \to 0 \quad \text{a.s.}\,, \tag{2.2.36}$$

where the right-hand side converges to zero exponentially fast. If we write $\Psi_i(x) = A_i x + B_i$, $x \in \mathbb{R}$, for the sequence of random affine maps generated by the sequence (A_i, B_i), $i = 1, 2, \ldots$, we observe that

$$X_n^x = \Psi_n \circ \cdots \circ \Psi_1(x)\,, \qquad x \in \mathbb{R}\,, \quad n \geq 1.$$

The sequence $(X_n^x)_{n \geq 1}$ is the *forward process* related to the sequence (A_i, B_i), $i = 1, 2, \ldots$. Relation (2.2.36) shows that $\Psi_n \circ \cdots \circ \Psi_1$ is a random Lipschitz function with Lipschitz coefficient $|\Pi_n|$ decaying exponentially as $n \to \infty$. This fact supports the idea of a *contractive map*.

Now consider the corresponding *backward process* for $x \in \mathbb{R}$,

$$\begin{aligned} Y_n^x &= \Psi_1 \circ \cdots \circ \Psi_n(x) \\ &= \Pi_n \, x + \sum_{i=1}^{n} \Pi_{i-1} \, B_i\,, \qquad n \geq 1\,. \end{aligned} \tag{2.2.37}$$

In view of (2.2.35), $Y_n^x \stackrel{\text{a.s.}}{\to} X$ as $n \to \infty$ and the random variable X satisfies the fixed-point equation $X \stackrel{d}{=} A\,X + B$.

Notice that for fixed n, X_n and Y_n have the same distribution. However, the trajectories of both processes behave in a completely different way. While the backward process (Y_n) converges a.s. the forward process (X_n) is ergodic and visits every set of positive P_X-measure infinitely often; see Figure 2.2 for an illustration of these processes.

The convergence in distribution of (X_n) to the stationary distribution of the Markov chain is supplemented by the convergence of the corresponding moments. Assume that $\mathbb{E}[|A|^p] < 1$ and $\mathbb{E}[|B|^p] < \infty$ for some $p > 0$. We know from the discussion on p. 15 that the condition $\mathbb{E}[\log |A|] < 0$ follows. It is not difficult to see that $\mathbb{E}[|X|^p] < \infty$ as well; see also the second part of Lemma 2.3.1 and its proof. As before, we assume that X_0 may have any distribution, but we additionally require that $\mathbb{E}[|X_0|^p] < \infty$.

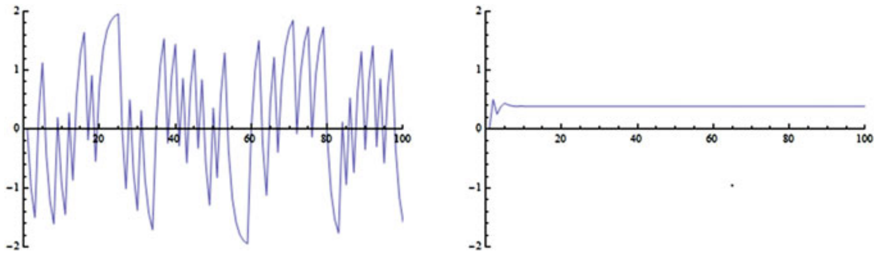

Figure 2.2 The figure illustrates the different behavior of the forward process (X_n) and the backward process (Y_n). Here we assume that $A = 0.5$ a.s. and $\mathbb{P}(B = 1) = \mathbb{P}(B = -1) = 0.5$. Then the stationary distribution P_X is uniform on $(-2, 2)$; see Example 2.2.8. We plot the first 100 elements of the forward process (left) and of the backward process (right), interpolating linearly between consecutive points. The forward process visits every open subset of the interval $[-2, 2]$ infinitely often with probability 1 while the backward process converges to some point.

For any random variable Y with finite pth moment we write

$$\|Y\|_p = \begin{cases} \mathbb{E}[|Y|^p]\,, & p \in (0, 1]\,, \\ (\mathbb{E}[|Y|^p])^{1/p}\,, & p > 1\,. \end{cases} \tag{2.2.38}$$

We make a short excursion to the *L_p-minimal metrics*, also referred to as *Mallows metrics d_p*, $p > 0$. In what follows, we assume that all random variables considered are defined on the same (nonatomic) probability space and have finite pth moment. For a bivariate vector (Y, Z) we define

$$d_p(Y, Z) = \inf \|Y' - Z'\|_p\,,$$

where the infimum is taken over all bivariate vectors (Y', Z') with the property $Y \stackrel{d}{=} Y'$, $Z \stackrel{d}{=} Z'$.[5] The Mallows metric d_p is a *probability metric* in the sense of Zolotarev [261, 262]; see also Rachev [233].

Then we have for any random variables $Y, Y_1, Y_2, \ldots$ defined on the same probability space that $d_p(Y_n, Y) \to 0$ as $n \to \infty$ holds if and only if both limit relations $Y_n \stackrel{d}{\to} Y$ and $\mathbb{E}[|Y_n|^p] \to \mathbb{E}[|Y|^p]$ as $n \to \infty$ are satisfied; see Bickel and Freedman [40]. We will apply this fact to (X_n) and X defined on the right-hand side of (2.2.35).

[5]The appearance of (Y, Z) in the notation $d_p(Y, Z)$ is convenient but incorrect: d_p is a metric on the set of probability measures with finite pth moment and given marginal distributions; it does not depend on a particular pair of random variables (Y, Z).

For $p > 0$, we have

$$
\begin{aligned}
d_p(X_n, X) &\le \Big\| \Big(\Pi_n X_0 + \sum_{i=1}^{n} \Pi_{i-1} B_i \Big) - \sum_{i=1}^{\infty} \Pi_{i-1} B_i \Big\|_p \\
&= \Big\| \Pi_n X_0 - \sum_{i=n+1}^{\infty} \Pi_{i-1} B_i \Big\|_p \\
&\le \| \Pi_n X_0 \|_p + \Big\| \sum_{i=n+1}^{\infty} \Pi_{i-1} B_i \Big\|_p \\
&= \|A\|_p^n \, \|X_0\|_p + \|\Pi_n\|_p \Big\| \sum_{i=1}^{\infty} \Pi_{i-1} B_i \Big\|_p \\
&\le \|A\|_p^n \left(\|X_0\|_p + \|X\|_p \right) .
\end{aligned}
$$

In the first inequality, we used (2.2.35) and the fact that, for any bivariate vectors (Y, Z) and (Y', Z') with $Y \overset{d}{=} Y'$, $Z \overset{d}{=} Z'$, $d_p(Y, Z) \le \|Y' - Z'\|_p$.

In this way we proved that $d_p(X_n, X) \le c \, \|A\|_p^n \to 0$, and since $\|A\|_p < 1$, this convergence happens at a geometric rate. In particular, we conclude that $X_n \overset{d}{\to} X$ (a fact we already know from (2.2.35)) and $\mathbb{E}[|X_n|^p] \to \mathbb{E}[|X|^p]$. The fact that $d_p(X_n, X) \to 0$ at a geometric rate again supports the intuition of a *contractive map*.

We mention in passing that Burton and Rösler [83] proved the d_2-convergence of (X_n) to X, even in the more abstract setting of a Hilbert space-valued stochastic recurrence equation.

2.3 Moments

We mentioned in Section 2.2 that the fixed-point equation

$$
X \overset{d}{=} A\,X + B\,, \tag{2.3.39}
$$

is useful for studying the distributional properties of the marginal distribution of the stationary solution (X_t) to the stochastic recurrence equation (2.0.1). For example, one can check whether certain moments of X exist if the corresponding moments of A, B exist.

Lemma 2.3.1 *Assume that the equation* (2.3.39) *has a solution X and let $p > 0$.*

1. *If $A, B \ge 0$ or if A, B are independent then $\mathbb{E}[|X|^p] < \infty$ implies $\mathbb{E}[|A|^p] + \mathbb{E}[|B|^p] < \infty$.*

2. *If the conditions* $\mathbb{P}(A = 0) = 0$, $\mathbb{P}(B = 0) < 1$ *and* $\mathbb{P}(Ax + B = x) < 1$ *for all* $x \in \mathbb{R}$ *hold then* $\mathbb{E}[|X|^p] < \infty$ *if and only if* $\mathbb{E}[|A|^p] < 1$ *and* $\mathbb{E}[|B|^p] < \infty$.

A proof of the sufficiency of the conditions $\mathbb{E}[|A|^p] < 1$ and $\mathbb{E}[|B|^p] < \infty$ for $\mathbb{E}[|X|^p] < \infty$ can already be found in Vervaat [256]. The second part was proved in Alsmeyer et al. [10]; in the case $A, B \geq 0$ a.s. they attribute the result to H.G. Kellerer in an unpublished technical report.

Proof In view of (2.3.39) we have

$$\mathbb{E}[|X|^p] = \mathbb{E}[|AX + B|^p]\,.$$

If $A, B \geq 0$ then $\mathbb{E}[|X|^p] < \infty$ implies that $\mathbb{E}[(AX)^p] = \mathbb{E}[A^p]\,\mathbb{E}[X^p] < \infty$ and $\mathbb{E}[B^p] < \infty$. The same conclusion holds by virtue of Fubini's theorem if AX and B are independent which is the case if X, A, B are independent.

The proof of "$\mathbb{E}[|A|^p] < 1$ & $\mathbb{E}[|B|^p] < \infty \Rightarrow \mathbb{E}[|X|^p] < \infty$" is not difficult: using the notation (2.2.38), concavity of the function $f(x) = x^p$, $x \geq 0$, for $p \in (0, 1]$ and Minkowski's inequality for $p > 1$, we show that

$$\begin{aligned}
\|X\|_p &= \Big\| \sum_{i=1}^{\infty} \Pi_{i-1}\, B_i \Big\|_p \leq \sum_{i=1}^{\infty} \|\Pi_{i-1}\, B_i\|_p \\
&= \sum_{i=1}^{\infty} \|\Pi_{i-1}\|_p \, \|B_i\|_p = \sum_{i=1}^{\infty} \|A\|_p^{i-1}\, \|B\|_p \\
&= \|B\|_p / (1 - \|A\|_p) < \infty\,.
\end{aligned}$$

Here we also used the independence of Π_{i-1} and B_i.

The converse statement of the second part is much more involved; see the proof of Theorem 1.4 in Alsmeyer et al. [10]. □

Alsmeyer et al. [10] also investigated the exponential moments of X. They proved the following result.

Lemma 2.3.2 *Assume that the equation* (2.3.39) *has a solution* X *and the following conditions hold.*

1. $\mathbb{P}(A = 0) = \mathbb{P}(|A| = 1) = 0$;
2. $\mathbb{P}(B = 0) < 1$;
3. $\mathbb{P}(Ax + B = x) < 1$ *for all* $x \in \mathbb{R}$.

Then for any $s > 0$, $\mathbb{E}[\mathrm{e}^{s\,|X|}] < \infty$ *if and only if* $\mathbb{P}(|A| < 1) = 1$ *and* $\mathbb{E}[\mathrm{e}^{s\,|B|}] < \infty$.

The positive lth integer moments of X can be calculated by using a recursive argument given in Vervaat [256], assuming $\mathbb{E}[|X|^l] < \infty$. First taking the lth power on both sides of (2.3.39) and then expectations, an application of the binomial formula yields

$$\mathbb{E}[X^l] = \mathbb{E}[(A\,X+B)^l] = \mathbb{E}[X^l]\,\mathbb{E}[A^l] + \sum_{k=0}^{l-1}\binom{l}{k}\mathbb{E}[A^k B^{l-k}]\,\mathbb{E}[X^k]\,.$$

Here we also exploited the independence of X and (A, B). We conclude the following.

Lemma 2.3.3 *Assume that equation (2.3.39) has a solution X and that $\mathbb{E}[|X|^l] + \mathbb{E}[|A|^l] + \mathbb{E}[|B|^l] < \infty$ for some integer $l \geq 1$, and $\mathbb{E}[A^l] \neq 1$. Then the following recursive relation holds:*

$$\mathbb{E}[X^l] = \frac{1}{1-\mathbb{E}[A^l]}\sum_{k=0}^{l-1}\binom{l}{k}\mathbb{E}[A^k B^{l-k}]\,\mathbb{E}[X^k]\,. \tag{2.3.40}$$

Of course, (2.3.40) further simplifies if A, B are independent. It follows from Lemma 2.3.1 that in most cases of interest the condition $\mathbb{E}[|X|^l] < \infty$ follows from or is equivalent to $\mathbb{E}[|A|^l] + \mathbb{E}[|B|^l] < \infty$. Moreover, the conditions of the second part of Lemma 2.3.1 also ensure that $\mathbb{E}[|A|^l] < 1$, hence $\mathbb{E}[A^l] \neq 1$. We also mention Boxma et al. [55] who calculate the positive integer moments $\mathbb{E}[X^l]$ for various concrete choices of distributions for (A, B).

The calculation of fractional moments of X or $|X|$ can also be of interest, for example in financial time series analysis. In Example 1.0.1 we introduced the GARCH process as a model for returns of speculative prices. This time series is of the form $X_t = \sigma_t\,Z_t$, where σ_t and Z_t are independent. Moreover, the positive stochastic volatility σ_t is not directly observable and therefore (estimates of) the moments $\mathbb{E}[|X_t|^p] = \mathbb{E}[\sigma_t^p]\,\mathbb{E}[|Z_t|^p]$ for positive p provide a measure of the magnitude of σ_t. In practice, one often considers only the cases $p = 1$ and $p = 2$. In the case $p = 1$, expressions for $\mathbb{E}[\sigma_t]$ are not available while $\mathbb{E}[\sigma_t^2]$ can be calculated for a GARCH(1, 1) process, using Lemma 2.3.3 and the defining equation for σ_t^2; see (1.0.3).

The literature on fractional moments of X satisfying (2.3.39) is rather sparse, showing that it is a hard task to calculate them. Mikosch et al. [211] contains a recent attempt; the reference list therein gives a rather complete picture. The discussion below follows some parts of [211].

The following simple formula is useful for calculating fractional moments of non-negative X. In what follows, we write F_Z for the distribution function and distribution of any random variable Z and $\overline{F}_Z = 1 - F_Z$ for its right tail.

Lemma 2.3.4 *Assume that the equation (2.3.39) has a solution X for some nonnegative independent random variables A, B. Let $p \neq 0$ be any real number. Then*

$$\mathbb{E}[(AX+B)^p - (AX)^p] = p\int_0^\infty \mathbb{E}[(AX+u)^{p-1}]\,\overline{F}_B(u)\,du\,, \tag{2.3.41}$$

where both sides are finite or infinite at the same time.

Formulæ similar to (2.3.41) were applied for calculating the moments of exponential functionals of Lévy processes in Carmona et al. [84], Guillemin et al. [138], Maulik and Zwart [201], Hirsch and Yor [150].

Proof We observe that for any $p \in \mathbb{R}$,

$$(AX + B)^p - (AX)^p = p \int_0^B (AX + u)^{p-1}\, du\,.$$

Hence, by independence of AX and B,

$$\begin{aligned}\mathbb{E}[(AX + B)^p - (AX)^p] &= p\,\mathbb{E}\Big[\int_0^B (AX + u)^{p-1}\, du\Big]\\ &= p \int_0^\infty \Big[\int_0^b \mathbb{E}[(AX + u)^{p-1}]\, du\Big]\, F_B(db)\\ &= p \int_0^\infty \mathbb{E}[(AX + u)^{p-1}]\, \overline{F}_B(u)\, du\,.\end{aligned}$$

Then the statement of the lemma follows. □

Remark 2.3.5 In Section 2.4 we will consider the tails of X. In this context, will discover that we often have the asymptotic relation $\mathbb{P}(X > x) \sim c_+\, x^{-\alpha}$ for some positive α, and the constant c_+ contains the expression $\mathbb{E}[(AX + B)^\alpha - (AX)^\alpha]$; see Theorem 2.4.4. Then $\mathbb{E}[X^\alpha] = \infty$ and formula (2.3.41) shows how the left-hand side can be relaxed to a moment of order $\alpha - 1$ which sometimes can be calculated, for example, if α is a positive integer. This idea was already explained in Goldie [128].

If $0 < \mathbb{E}B < \infty$, (2.3.41) can be written in the form

$$\mathbb{E}[(A\,X + B)^p - (A\,X)^p] = p\,\mathbb{E}B\;\mathbb{E}[(A\,X + B^*)^{p-1}]\,,$$

where A, B^*, X are independent and B^* has the *integrated tail distribution* of B given by

$$F_B^*(b) = \frac{\int_0^b \overline{F}_B(u)\, du}{\mathbb{E}B}\,, \quad b > 0\,. \tag{2.3.42}$$

Also notice that $B \overset{d}{=} B^*$ for standard exponential B and then

$$\mathbb{E}[(AX + B)^\alpha - (AX)^\alpha] = \alpha\,\mathbb{E}[(AX + B)^{\alpha-1}] = \alpha\,\mathbb{E}[X^{\alpha-1}]\,;$$

see also Example 2.3.6.

Relation (2.3.41) does not require that $\mathbb{E}[X^p] < \infty$, but if this moment is finite, $\mathbb{E}B < \infty$ and[6] $\mathbb{E}[A^p] \neq 1$ then the lemma yields the formula

$$\mathbb{E}[X^p] = \frac{p\,\mathbb{E}B}{1-\mathbb{E}[A^p]}\,\mathbb{E}[(AX+B^*)^{p-1}]\,. \tag{2.3.43}$$

Lemma 2.3.4 can be applied iteratively. We explain the approach via an example. Assume the conditions of the lemma are satisfied. Let B^{n*} be a random variable whose distribution is obtained by applying n times the integrated tail operation (2.3.42), and assume that A, B^{n*}, X are independent. Then assuming that all moments involved are finite and $p \neq 0$, (2.3.43) yields

$$\begin{aligned}
\mathbb{E}[X^p] &= \frac{p\,\mathbb{E}B}{1-\mathbb{E}[A^p]}\big[(\mathbb{E}[(AX+B^*)^{p-1}] - \mathbb{E}[(AX)^{p-1}] + \mathbb{E}[(AX)^{p-1}]\big]\\
&= \frac{p\,\mathbb{E}B}{1-\mathbb{E}[A^p]}\big[(p-1)\mathbb{E}[B^*]\,\mathbb{E}[(AX+B^{2*})^{p-2}] + \mathbb{E}[(AX)^{p-1}]\big]\\
&= \frac{p\,(p-1)\,\mathbb{E}B}{1-\mathbb{E}[A^p]}\Big[\mathbb{E}[B^*]\,\mathbb{E}[(AX+B^{2*})^{p-2}] + \frac{\mathbb{E}[A^{p-1}]\,\mathbb{E}B}{1-\mathbb{E}[A^{p-1}]}\,\mathbb{E}[(AX+B^*)^{p-2}]\Big].
\end{aligned}$$

It is in general difficult to evaluate $\mathbb{E}[X^p]$ by using the formulæ above. However, if B has an exponential distribution the calculations simplify.

Example 2.3.6 Assume that B has a standard exponential distribution, i.e., $\overline{F}_B(x) = \mathrm{e}^{-x}$, $x > 0$, and $\mathbb{E}[A^p] < 1$. Then $B^* \stackrel{d}{=} B$ and, assuming A, B^*, X independent, $AX + B^* \stackrel{d}{=} X$. Multiple use of (2.3.43) yields for real p,

$$\mathbb{E}[X^p] = \frac{p\cdots(p-n+1)}{(1-\mathbb{E}[A^p])\cdots(1-\mathbb{E}[A^{p-n+1}])}\,\mathbb{E}[X^{p-n}]\,,\quad n \geq 1\,. \tag{2.3.44}$$

The case of exponential B has been studied in the literature for some time. Recall from Theorem 2.1.3 that

$$X \stackrel{d}{=} \sum_{i=1}^{\infty} \Pi_{i-1}\,B_i = \sum_{i=1}^{\infty} \mathrm{e}^{\,S_{i-1}}\,B_i\,,$$

where $S_i = \sum_{k=1}^{i} \log A_i, i \geq 0$, with the convention that $S_0 = 0$. Since $\mathbb{E}[\log A] < 0$ the random walk (S_i) has negative drift. Writing

$$N_t = \#\{i \geq 1 : B_1 + \cdots + B_i \leq t\}\,,\qquad t \geq 0\,,$$

[6] In view of Lemma 2.3.1 the condition $\mathbb{E}[A^p] < 1$ is necessary for $\mathbb{E}[X^p] < \infty$ under mild conditions.

for the unit rate Poisson process generated by the iid exponential sequence (B_i) and $(\xi_t) = (S_{N_t})$ for the compound Poisson process generated by the sequence $(\log A_i)$, we have

$$X \stackrel{d}{=} \int_0^\infty e^{\xi_t}\, dt\,. \tag{2.3.45}$$

Carmona et al. [84] studied exponential functionals of the type (2.3.45) for Lévy processes (ξ_t) more general than compound Poisson processes. They derived (2.3.44) for positive p. Related results were obtained by Behme et al. [34], Brockwell and Lindner [62], Maulik and Zwart [201]; see also the references therein. Hirsch and Yor [150] provide a survey of results on exponential functionals (2.3.45) for subordinators ξ more general than compound Poisson.

Exponential functionals were also studied in Guillemin et al. [138] in the context of Example 1.0.2. They considered the stochastic recurrence equation $X_t = A_t X_{t-1} + B_t$, $t \in \mathbb{Z}$, assuming independence between A and B, B standard exponential and $A_t = \beta^{Y_t}$, $\beta \in (0,1)$, for an iid sequence (Y_t) of positive random variables. Under these conditions, using Mellin transforms, they derived (2.3.44) and from there

$$\mathbb{E}[X^p] = \Gamma(p+1) \prod_{k=1}^\infty \frac{1-\mathbb{E}[A^{p+k}]}{1-\mathbb{E}[A^k]}\,, \tag{2.3.46}$$

provided $p \in \mathbb{R}$, $-p \notin \mathbb{N}$, $\mathbb{E}[A^{p+1}] < \infty$ and $\mathbb{E}[(1-A)^{-1}] < \infty$. An inspection of the proof shows that the latter conditions are not needed if $p > 0$.

If $Y \equiv 1$ a.s. (X_t) is an autoregressive process of order one and $X_t \stackrel{d}{=} \sum_{i=0}^\infty \beta^i B_i$. Then $\mathbb{E}[A^p] = \beta^p < \infty$ for any real p, $\mathbb{E}[(1-A)^{-1}] = (1-\beta)^{-1} < \infty$ and (2.3.46) turns into

$$\mathbb{E}[X^p] = \Gamma(p+1) \prod_{k=1}^\infty \frac{1-\beta^{p+k}}{1-\beta^k}\,, \qquad p \in \mathbb{R}\,, \quad -p \notin \mathbb{N}\,. \tag{2.3.47}$$

The same formula can be found in Bertoin et al. [38], Theorem 1, and the authors also derived the density and the Laplace transform of X in this case.

We mention in passing that the same arguments as above apply if B has the mixture distribution $\mathbb{P}(B \le x) = (1-a) + a\,(1 - e^{-x})$, $x > 0$, for some $a \in (0,1)$. Then B^* is standard exponential and the moments $\mathbb{E}[X^{p-k}]$, $k = 1, 2, \ldots$, can be calculated by using the recursion (2.3.43).

Mikosch et al. [211] discuss some further special cases, where one can calculate the pth fractional moment explicitly. However, the examples considered are far from exhaustive and further investigations are needed.

2.4 The Tails

2.4.1 *Generalities*

In this section we give an overview of results on the asymptotic tail behavior of the solution X to the univariate fixed-point equation

$$X \stackrel{d}{=} AX + B\,, \tag{2.4.48}$$

where, as usual, X and (A, B) are independent. At the same time we gain information about the tails of the marginal distribution of the stationary causal solution to the stochastic recurrence equation $X_t = A_t X_{t-1} + B_t$, $t \in \mathbb{Z}$; see Theorem 2.1.3.

The behavior of the tail $\mathbb{P}(|X| > x)$ for large x depends on the distribution of (A, B). Of course, $\mathbb{P}(|X| > x)$ converges to zero as x tends to infinity. Thus a natural problem consists of describing the rate at which this happens. In the literature, one distinguishes between three distinct cases of tail behavior of X:

- If $|A| \le 1$ a.s. and B has a moment generating function in some neighborhood of the origin then $\mathbb{P}(|X| > x) \le K\,\mathrm{e}^{-c_0 x}$ for $x > 0$ and constants $K, c_0 > 0$, i.e., the tails of X decrease exponentially. Below we present results of Goldie and Grübel [129] and Hitczenko and Wesołowski [152]; see Theorem 2.4.1.
- If $\mathbb{P}(|A| > 1) > 0$ then there exists $r > 0$ such that $\mathbb{E}[|X|^r] = \infty$; see Theorem 2.4.2. On the other hand, in view of Lemma 2.3.1, if some moments of A and B are finite then $\mathbb{E}[|X|^s] < \infty$ for small s. Below we present two directions of research: Theorem 2.4.3 due to Grincevičius [134] and Grey [131], and the well-known Kesten–Goldie Theorem 2.4.4 in combination with Theorem 2.4.7 proved by Kesten [175] and later on by Goldie [128], using different arguments. In both cases the tails of X are regularly varying, i.e., they decrease polynomially.
- If all (power) moments of A are infinite then, under some additional assumptions such as subexponentiality of $\log(\max(|A|, |B|))$, the tails of X are slowly varying; see Section 5.5.

We may classify the distribution of X as *heavy-tailed* (when certain moments of X are infinite) or *light-tailed* (when all moments of X exist). The distribution of X can be heavy-tailed as a consequence of the tails of B: if $\mathbb{E}[|B|^r] = \infty$ for some $r > 0$ and A has lighter tails than B then $\mathbb{E}[|X|^r] = \infty$.

2.4.2 *The Goldie-Grübel Theorem*

First, we focus on the case when $|A| \le 1$ a.s. The following result is the first part of Theorem 2.1 in Goldie and Grübel [129].

Theorem 2.4.1 *Assume that X satisfies (2.4.48), where X is independent of (A, B). Moreover, assume the conditions:*

1. $|A| \le 1$ *a.s. and* $\mathbb{P}(|A| < 1) > 0$.
2. $\mathbb{E}[e^{\varepsilon |B|}] < \infty$ *for some* $\varepsilon > 0$.

Then

$$\limsup_{x \to \infty} \frac{\log \mathbb{P}(|X| > x)}{x} \le -\rho_0 \,,$$

where $\rho_0 = \sup\{\rho : \mathbb{E}[e^{\rho |A|} |B|] < 1\}$.

This result says that for every $\delta > 0$ and sufficiently large x, $\mathbb{P}(|X| > x) \le e^{-(\rho_0 - \delta)x}$. In particular, $\mathbb{E}[e^{h|X|}] < \infty$ for $0 < h < \rho_0$.

Theorem 2.1 in Goldie and Grübel [129] also deals with one-sided cases, for example, when $0 \le A \le 1$ a.s., $\mathbb{P}(A = 0) > 0$ and $\mathbb{E}[e^{\varepsilon B_+}] < \infty$ for some $\varepsilon > 0$, where for any real x, $x_+ = \max(x, 0)$. Then one gets separate results for the left and right tails of the distribution of X. In the case $B \equiv 1$ a.s. and $\mathbb{P}(A < 1) > 0$ these results imply that

$$\limsup_{x \to \infty} \frac{\log \mathbb{P}(X > x)}{x} \le \log(\mathbb{E}A) \,. \tag{2.4.49}$$

Goldie and Grübel [129] showed that, in general, their bounds cannot be improved. However, X can even be bounded. For example, assume $|A| \le c_A < 1$ and $|B| \le c_B$ a.s. Examples of this kind include the fractal images of Example 1.0.5 and the random binary expansions of Example 1.0.3. Then

$$\Big| \sum_{i=1}^{\infty} \Pi_{i-1} B_i \Big| \le c_B \sum_{i=1}^{\infty} c_A^{i-1} = \frac{c_B}{1 - c_A} < \infty \quad \text{a.s.}$$

Depending on the order of magnitude of the probabilities $\mathbb{P}(|A| \in [1 - \delta, 1])$ as $\delta \downarrow 0$, differing behavior of $\mathbb{P}(|X| > x)$ as $x \to \infty$ may result. Theorems 3.1 and 3.2 in [129] yield a "Poissonian tail" of the form

$$\lim_{x \to \infty} \frac{\log \mathbb{P}(X > x)}{x \log x} = -c_1 \tag{2.4.50}$$

for a certain constant $c_1 > 0$, under various conditions on A; one assumption ensures that $\delta^{-1} \mathbb{P}(1 - \delta \le A \le 1)$ is bounded away from zero and infinity as $\delta \downarrow 0$. Assuming precise rates of decay for $\mathbb{P}(1 - \delta \le A \le 1)$ as $\delta \downarrow 0$, Hitczenko and Wesołowski [152] showed that much lighter tails than prescribed by (2.4.49) or even by Poissonian tails (2.4.50) are possible; see also a recent paper by Kolodziejek [184].

A surprising result is Theorem 4.1 in Goldie and Grübel [129]:

Theorem 2.4.2 *Assume that X satisfies (2.4.48), where X is independent of (A, B). If $\mathbb{P}(|A| > 1) > 0$ then $|X|$ has at least power-law tail:*

$$\liminf_{x\to\infty} \frac{\log \mathbb{P}(|X| > x)}{\log x} > -\infty\,.$$

In particular, there is $r > 0$ such that $\mathbb{E}[|X|^r] = \infty$.

The results by Goldie and Grübel [129] show that for "well-behaved" B intermediate decay rates for $\mathbb{P}(|X| > x)$ between a power-law and an exponential function are not possible. In Section 2.4.4 we will refine Theorem 2.4.2 by showing that the very precise tail behavior $\mathbb{P}(|X| > x) \sim c\,x^{-\alpha}$ as $x \to \infty$ for some constant $c > 0$ can be derived under mild conditions.

2.4.3 The Grincevičius-Grey Theorem

Theorems 2.4.1 and 2.4.2 indicate that the distribution of B has rather marginal influence on the tail behavior of X at least when $|B|$ has finite exponential moments. This changes if B has heavy tails and A has lighter tails than B. Next, we will give a corresponding result if the tails of B are of power-law type. We start with a brief introduction to distributions with regularly varying tails; more details can be found in Appendix B.

Recall from Feller [120] or Bingham et al. [45] that a positive measurable function g on $(0, \infty)$ is *regularly varying with index* $\rho \in \mathbb{R}$ if it has the form

$$g(x) = x^\rho L(x)\,, \quad x > 0\,,$$

where L is a *slowly varying function*, i.e., for $c > 0$, $\lim_{x\to\infty} L(cx)/L(x) = 1$. We will be interested in random variables X with regularly varying tails. In particular, we call a random variable X *regularly varying with index* $\alpha \ge 0$ if there exist constants $p, q \ge 0$ with $p + q = 1$ and a slowly varying function L such that

$$\mathbb{P}(X > x) \sim p\,x^{-\alpha} L(x) \quad \text{and} \quad \mathbb{P}(X < -x) \sim q\,x^{-\alpha} L(x)\,, \qquad x \to \infty\,.$$

This means that both tails are regularly varying functions with index $-\alpha$, possibly degenerate in the case when $p = 0$ or $q = 0$. If $p = 0$ we interpret the assumption on the right tail in the sense that $\mathbb{P}(X > x) = o(\mathbb{P}(|X| > x))$ as $x \to \infty$; the case $q = 0$ is treated correspondingly.

Theorem 2.4.3 *Assume that the following conditions hold:*

1. $A \geq 0$ *a.s.,* $\mathbb{P}(A = 0) < 1$.
2. $\mathbb{E}[A^\alpha] < 1$ *for some* $\alpha > 0$ *and* $\mathbb{E}[A^{\alpha+\delta}] < \infty$ *for some* $\delta > 0$.

Then the following statements hold:

(1) *Assume that* X *solves the equation* $X \stackrel{d}{=} AX + B$, *where* X *is independent of the pair of random variables* (A, B). *If the random variable* X *is regularly varying with index* $\alpha > 0$ *then* B *is regularly varying with the same index.*

(2) *Conversely, if* B *is regularly varying with index* $\alpha > 0$ *then there exists a solution to the equation* $X \stackrel{d}{=} AX + B$, *where* X *is independent of the pair of random variables* (A, B).
If $\lim_{x\to\infty} \mathbb{P}(\pm B > x)/\mathbb{P}(|B| > x) = c_\pm$ *for some positive constants* c_+ *and* c_- *then the following tail equivalence relation holds:*

$$\mathbb{P}(\pm X > x) \sim (1 - \mathbb{E}[A^\alpha])^{-1}\, \mathbb{P}(\pm B > x)\,, \quad x \to \infty\,. \tag{2.4.51}$$

A proof of this theorem can be found in Grey [131]. The second part of the theorem was also proved by Grincevičius [134]. Below we give an independent proof of part (2). A proof of part (1) for the multivariate fixed-point equation $\mathbf{X} \stackrel{d}{=} \mathbf{A}\,\mathbf{X} + \mathbf{B}$ is provided in Theorem 4.4.27 on p. 200; the result in the univariate case is a straightforward consequence.

Proof (*Proof of part* (2)) If B is regularly varying with index α it has moments of order less than α, hence $\mathbb{E}[\log^+ |B|] < \infty$. In view of Theorem 2.1.3 there exists a strictly stationary causal solution (X_t) to the stochastic recurrence equation (2.0.1). A generic element X of this sequence satisfies the fixed-point equation $X \stackrel{d}{=} AX + B$ and has representation in law given by

$$X \stackrel{d}{=} \sum_{i=1}^{\infty} \Pi_{i-1}\, B_i\,.$$

Consider the decomposition

$$\sum_{i=1}^{\infty} \Pi_{i-1} B_i = \Big(\sum_{i=1}^{s} + \sum_{i=s+1}^{\infty}\Big) \Pi_{i-1} B_i = \widetilde{X}_s + \widetilde{X}^s\,, \qquad s \geq 1\,.$$

Assume $i < j$ and write $C_m = \{\Pi_{i-1} \vee \Pi_{j-1} \leq m\}$ for any $m > 0$. Then for large x and some constants $c, c(m) > 0$,

$$
\begin{aligned}
&\mathbb{P}(\Pi_{i-1}|B_i| > x, \Pi_{j-1}|B_j| > x) \\
&\le \mathbb{P}((\Pi_{i-1} \vee \Pi_{j-1})\,(|B_i| \wedge |B_j|) > x) \\
&= \mathbb{P}(\{(\Pi_{i-1} \vee \Pi_{j-1})\,(|B_i| \wedge |B_j|) > x\} \cap C_m) \\
&\quad + \mathbb{P}(\{(\Pi_{i-1} \vee \Pi_{j-1})\,(|B_i| \wedge |B_j|) > x\} \cap C_m^c) \\
&\le \mathbb{P}(m\,(|B_i| \wedge |B_j|) > x) + \mathbb{P}((\Pi_{i-1} \vee \Pi_{j-1})\,\mathbf{1}(C_m^c)\,|B_j| > x) \\
&\le c(m)\,[\mathbb{P}(|B| > x)]^2 + c\,\mathbb{E}\big[(\Pi_{i-1} \vee \Pi_{j-1})\big)^\alpha \mathbf{1}(C_m^c)\big]\,\mathbb{P}(|B| > x)\,.
\end{aligned}
$$

For the first term in the last inequality, we used the regular variation of B, implying that $\mathbb{P}(|B| > x/m) \sim m^\alpha \mathbb{P}(|B| > x)$ as $x \to \infty$. For the second term, we applied Breiman's Lemma B.5.1. It requires independence between $Z = (\Pi_{i-1} \vee \Pi_{j-1})\,\mathbf{1}(C_m^c)$ and $Y = |B_j|$, which is satisfied for $i < j$, and we also need to ensure that $\mathbb{E}[Z^{\alpha+\delta}] < \infty$ for some $\delta > 0$, but this follows from the assumption $\mathbb{E}[A^{\alpha+\delta}] < \infty$. Thus by first letting $x \to \infty$ and then $m \to \infty$, we conclude that

$$
\lim_{x\to\infty} \frac{\mathbb{P}(\Pi_{i-1}|B_i| > x, \Pi_{j-1}|B_j| > x)}{\mathbb{P}(|B| > x)} = 0\,, \quad i \ne j\,. \tag{2.4.52}
$$

An application of Lemma B.6.1 and multiple use of Breiman's Lemma B.5.1 yield as $x \to \infty$,

$$
\begin{aligned}
\mathbb{P}(\widetilde{X}_s > x) &\sim \sum_{i=1}^{s} \mathbb{P}(\Pi_{i-1} B_i > x) \sim \mathbb{P}(B > x) \sum_{i=1}^{s} \mathbb{E}[\Pi_{i-1}^\alpha] \\
&= \mathbb{P}(B > x) \sum_{i=1}^{s} (\mathbb{E}[A^\alpha])^{i-1}\,.
\end{aligned}
$$

Thus $\widetilde{X}_s$ inherits the regular variation from B for every fixed $s \ge 1$. Since for small $\varepsilon \in (0, 1)$,

$$
\mathbb{P}(\widetilde{X}_s > (1+\varepsilon)x) - \mathbb{P}(\widetilde{X}^s \le -\varepsilon x) \le \mathbb{P}(X > x) \le \mathbb{P}(\widetilde{X}_s > (1-\varepsilon)x) + \mathbb{P}(\widetilde{X}^s > \varepsilon x)
$$

and B is regularly varying we have for $s \ge 1$,

$$
\begin{aligned}
&\frac{c_+}{(1+\varepsilon)^\alpha} \sum_{i=1}^{s} (\mathbb{E}[A^\alpha])^{i-1} - \limsup_{x\to\infty} \frac{\mathbb{P}(|\widetilde{X}^s| > \varepsilon x)}{\mathbb{P}(|B| > x)} \\
&\le \liminf_{x\to\infty} \frac{\mathbb{P}(X > x)}{\mathbb{P}(|B| > x)} \le \limsup_{x\to\infty} \frac{\mathbb{P}(X > x)}{\mathbb{P}(|B| > x)} \\
&\le \frac{c_+}{(1-\varepsilon)^\alpha} \sum_{i=1}^{s} (\mathbb{E}[A^\alpha])^{i-1} + \limsup_{x\to\infty} \frac{\mathbb{P}(|\widetilde{X}^s| > \varepsilon x)}{\mathbb{P}(|B| > x)}\,.
\end{aligned}
$$

Therefore (2.4.51) follows by letting $s \to \infty$ and $\varepsilon \downarrow 0$ if we can also show that

$$\lim_{s\to\infty} \limsup_{x\to\infty} \frac{\mathbb{P}(|\widetilde{X}^s| > x)}{\mathbb{P}(|B| > x)} = 0\,. \tag{2.4.53}$$

Fix $\gamma \in (0, 1)$ and observe that

$$\begin{aligned}\mathbb{P}(|\widetilde{X}^s| > x) &\le \sum_{i=s+1}^{\infty} \mathbb{P}\big(\Pi_{i-1}\,|B_i| > x\,(1-\gamma)\gamma^{i-(s+1)}\big)\\ &= \sum_{i=0}^{\infty} \mathbb{P}\big(\Pi_{s+i}\,|B_{s+i+1}| > x\,(1-\gamma)\gamma^{i}\big)\,.\end{aligned}$$

We employ the Potter bounds (B.2.3) on p. 274: for given $\widetilde{\varepsilon} > 0$, there exists a constant $x_0 = x_0(\widetilde{\varepsilon})$ such that for any $c > 0$,

$$\frac{\mathbb{P}(|B| > x/c)}{\mathbb{P}(|B| > x)} \le \begin{cases} (1+\widetilde{\varepsilon})\,c^{\alpha+\widetilde{\varepsilon}}\,, & \text{for } c \ge 1,\ x/c \ge x_0,\\ (1+\widetilde{\varepsilon})\,c^{\alpha-\widetilde{\varepsilon}}\,, & \text{for } c < 1,\ x \ge x_0.\end{cases}$$

Applying these bounds conditional on Π_{s+i}, we obtain for given $\widetilde{\varepsilon} > 0$, some constant $c_0 > 0$ which does not depend on i and s,

$$\begin{aligned}&\frac{\mathbb{P}\big(\Pi_{s+i}\,|B_{s+i+1}| > x(1-\gamma)\gamma^{i}\big)}{\mathbb{P}(|B| > x)}\\ &\le \frac{\mathbb{E}\big[\Pi_{s+i}^{\alpha+\widetilde{\varepsilon}}\mathbf{1}\big(\Pi_{s+i} \ge (1-\gamma)\gamma^{i},\ x/x_0 \ge \Pi_{s+i}/((1-\gamma)\gamma^{i})\big)\big]}{((1-\gamma)\gamma^{i})^{\alpha+\widetilde{\varepsilon}}}\\ &\quad+\frac{\mathbb{E}\big[\Pi_{s+i}^{\alpha-\widetilde{\varepsilon}}\mathbf{1}\big(\Pi_{s+i} < (1-\gamma)\gamma^{i},\ x \ge x_0\big)\big]}{((1-\gamma)\gamma^{i})^{\alpha-\widetilde{\varepsilon}}}\\ &\quad+\frac{\mathbb{P}\big(\Pi_{s+i} > (1-\gamma)\gamma^{i}x/x_0\big)}{\mathbb{P}(|B| > x)}\\ &\le c_0\Big[(\mathbb{E}[A^{\alpha+\widetilde{\varepsilon}}])^{s}\ (\mathbb{E}[(A/\gamma)^{\alpha+\widetilde{\varepsilon}}])^{i} + (\mathbb{E}[A^{\alpha-\widetilde{\varepsilon}}])^{s}\ (\mathbb{E}[(A/\gamma)^{\alpha-\widetilde{\varepsilon}}])^{i}\\ &\qquad+(\mathbb{E}[A^{\alpha+\widetilde{\varepsilon}}])^{s}\,\frac{x^{-(\alpha+\widetilde{\varepsilon})}}{\mathbb{P}(|B| > x)}\,(\mathbb{E}[(A/\gamma)^{\alpha+\widetilde{\varepsilon}}])^{i}\Big]\,.\end{aligned} \tag{2.4.54}$$

In the last step, we used Markov's inequality. By assumption, there exists $\delta > 0$ such that $\mathbb{E}[A^{\alpha+\delta}] < \infty$. Using this fact in combination with $\mathbb{E}[A^{\alpha}] < 1$, there exist $\widetilde{\varepsilon} > 0$ and $\gamma \in (0, 1)$ such that $\mathbb{E}[(A/\gamma)^{\alpha+z}] < 1$ for $z \in [-\widetilde{\varepsilon}, \widetilde{\varepsilon}]$. Combining the inequalities above, we obtain

$$\limsup_{x\to\infty} \frac{\mathbb{P}(|\widetilde{X}^s| > x)}{\mathbb{P}(|B| > x)} \le c_0\Big[\frac{(\mathbb{E}[A^{\alpha+\widetilde{\varepsilon}}])^{s}}{1-\mathbb{E}[(A/\gamma)^{\alpha+\widetilde{\varepsilon}}]} + \frac{(\mathbb{E}[A^{\alpha-\widetilde{\varepsilon}}])^{s}}{1-\mathbb{E}[(A/\gamma)^{\alpha-\widetilde{\varepsilon}}]}\Big],$$

and the right-hand side vanishes as $s \to \infty$. Thus we proved (2.4.53). This finishes the proof. □

2.4.4 The Kesten–Goldie Theorem

In his 1973 paper, Kesten [175] proved a rather astonishing result about the tails of the $\mathbb{R}^d$-valued solution $\mathbf{X}$ to the fixed-point equation $\mathbf{X} \stackrel{d}{=} \mathbf{A}\,\mathbf{X} + \mathbf{B}$. Roughly speaking, he showed that the solution to this equation may have power-law tails in the sense that $\mathbb{P}(|\mathbf{X}| > x) \sim c\,x^{-\alpha}$ as $x \to \infty$ for some positive constant c; a detailed analysis of the tails in the multivariate case is given in Section 4.4. Although Theorem 2.4.2 has prepared us for such a result, the precision of the tail asymptotics is still surprising.

Later, in 1991, Goldie [128] gave an independent proof of this result in the univariate case, based on implicit renewal theory. On p. 131 of his paper [128], he mentioned that he benefitted from ideas in Grincevičius [134]. Grincevičius partly rediscovered Kesten [175] but also developed his own approach which remained incomplete in some arguments.

The following result is a special case of Theorem 5 in Kesten [175] and of Theorem 4.1 in Goldie [128] when $A \ge 0$ a.s.; the complete result (without proof) is given as Theorem 2.4.7 below. Here we present Goldie's proof. He did not only derive the precise power-law tail behavior but he also determined the constants in the tails.

Theorem 2.4.4 *Assume that the following conditions hold.*

1. $A \ge 0$ *a.s. and the law of* $\log A$ *conditioned on* $\{A > 0\}$ *is non-arithmetic.*[7]
2. *There exists* $\alpha > 0$ *such that* $\mathbb{E}[A^\alpha] = 1$, $\mathbb{E}[|B|^\alpha] < \infty$ *and* $\mathbb{E}[A^\alpha \log^+ A] < \infty$.
3. $\mathbb{P}(Ax + B = x) < 1$ *for every* $x \in \mathbb{R}$.

Then the equation $X \stackrel{d}{=} AX + B$ *has a solution* X *which is independent of* (A, B) *and there exist constants* c_+, c_- *such that* $c_+ + c_- > 0$ *and*

$$\mathbb{P}(X > x) \sim c_+\,x^{-\alpha} \quad \textit{and} \quad \mathbb{P}(X < -x) \sim c_-\,x^{-\alpha}\,, \qquad x \to \infty\,, \tag{2.4.55}$$

The constants c_+, c_- *are given by*

$$\begin{cases} c_+ = \dfrac{1}{\alpha\, m_\alpha}\, \mathbb{E}\big[(AX + B)_+^\alpha - (AX)_+^\alpha\big], \\ c_- = \dfrac{1}{\alpha\, m_\alpha}\, \mathbb{E}\big[(AX + B)_-^\alpha - (AX)_-^\alpha\big], \end{cases} \tag{2.4.56}$$

where $m_\alpha = \mathbb{E}\big[A^\alpha \log A\big] > 0$.

[7] A random variable is non-arithmetic if it is not supported on any of the sets $a\mathbb{Z}$, $a \ge 0$.

Some Comments Related to Theorem 2.4.4

- In Section 2.5 we will study in detail the support of the distribution P_X of X. In particular, in Theorem 2.5.5 we will prove under general conditions that the support of P_X is either a half-line of the form (a, ∞) or $(-\infty, a)$ for some real a or the whole real line. In the former cases, we have either $c_+ > 0$ and $c_- = 0$ or $c_- > 0$ and $c_+ = 0$. Both constants c_+ and c_- are positive if, in addition to the conditions of Theorem 2.4.4, we assume that the support of P_X is the real line. This is the content of Theorem 2.4.6 below.
- The function

$$h(p) = \mathbb{E}[A^p] = \mathbb{E}[\mathrm{e}^{\,p \log A}] \tag{2.4.57}$$

has a positive second derivative in $(0, \alpha)$ and, therefore, is strictly convex on $[0, \alpha]$. A graph of the function $h(p)$ is shown in Figure 2.3. If $A \neq 1$ a.s. and a positive solution to the equation $h(p) = 1$ exists it is unique because of strict convexity of h. For the same reason, $h(p) < 1$ on $(0, \alpha)$. By Jensen's inequality,

$$p\,\mathbb{E}[\log A] = \mathbb{E}[\log(A^p)] \leq \log\big(\mathbb{E}[A^p]\big) < 0\,.$$

Since $\mathbb{E}[\log A] < 0$ is the right derivative of h at zero, h decreases in an interval $[0, p_0]$ for some $p_0 > 0$. On the other hand, strict convexity and $h(\alpha) = 1$ imply that h increases in some neighborhood of α and therefore the left first derivative of h at α, given by m_α, is positive.

The condition $h(\alpha) = 1$ appears in numerous applied probability problems. It is often referred to as *Cramér* or *Cramér-Lundberg condition* and plays a fundamental role for determining the *ruin probability* in a non-life portfolio in the presence

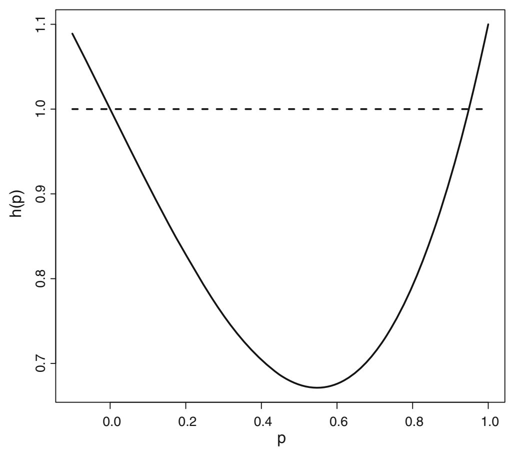

Figure 2.3 An example of the convex function $h(p) = \mathbb{E}[A^p]$; see (2.4.57). If $A \neq 1$ a.s. and the equation $h(p) = 1$ has a positive solution α it is unique. If $A \leq 1$ and $A \neq 1$ a.s. then $h(p) < 1$ for any $p > 0$, hence α does not exist. In the case when $\mathbb{P}(A > 1) > 0$, a positive solution α does not exist if there is $p_0 > 0$ such that $h(p) < 1$ on $(0, p_0)$ and $h(p_0) = \infty$.

of light tails. In this context, $\log A_i$ has interpretation as the difference between claim size and premium per time unit in the portfolio. Then the random walk $S_n = \sum_{i=1}^n \log A_i$ is the difference between the aggregated claim amount and premium in the first n periods. Since $\mathbb{E}[\log A] < 0$, this random walk has a negative drift and, by the strong law of large numbers, it tends to $-\infty$ as $n \to \infty$. We refer to Asmussen [17], Section XIII.5, Asmussen and Albrecher [18], Chapters IV and VI, and Embrechts et al. [112], Section 1.2, for further reading on ruin probabilities and related topics in queuing theory.

- The rate of convergence in (2.4.55) was investigated in Goldie [128] and Buraczewski et al. [78]. Depending on the assumptions, they found convergence rates of the order $x^{-\sigma_1}$ or $(\log x)^{-\sigma_2}$ for positive constants σ_1 and σ_2. More concretely, assuming a Lebesgue density for A and $\mathbb{E}[A^{\alpha+\delta}] < \infty$ for some $\delta > 0$, Goldie [128], Theorem 3.2, proved that there exists a positive constant σ_1 such that

$$\big|(c_+ + c_-) - x^\alpha\, \mathbb{P}(|X| > x)\big| \le \text{const}\; x^{-\sigma_1}\,, \quad x > 0\,.$$

Under much weaker assumptions Buraczewski et al. [78] proved that

$$\big|(c_+ + c_-) - x^\alpha\, \mathbb{P}(|X| > x)\big| \le \text{const}\; (\log x)^{-\sigma_2}\,, \qquad x > 1\,,$$

for some positive constant σ_2.

- In general, formula (2.4.56) does not allow one to evaluate the constants c_+ and c_-. However, there exist other formulæ which can be used to approximate these values in a more efficient way. For example, under the conditions of Theorem 2.4.4 with $Y_n = \sum_{i=1}^n \Pi_{i-1}\, B_i$,

$$\lim_{n\to\infty} \frac{1}{\alpha\, m_\alpha}\, \frac{\mathbb{E}[|Y_n|^\alpha\, \mathbf{1}(Y_n > 0)]}{n} = c_+ \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{\alpha\, m_\alpha}\, \frac{\mathbb{E}[|Y_n|^\alpha\, \mathbf{1}(Y_n < 0)]}{n} = c_-\,,$$

and in turn,

$$\lim_{n\to\infty} \frac{1}{\alpha\, m_\alpha}\, \frac{\mathbb{E}[|Y_n|^\alpha]}{n} = c_+ + c_-\,;$$

see Bartkowiak at al. [26] in the case $B = 1$ and Buraczewski et al. [79] in the general case. Other approximation formulæ for $c_\pm$ can be found in Enriquez et al. [115], Collamore and Vidyashankar [89], Collamore et al. [88].

- Recently, Roitershtein [243] and Collamore [87] independently considered the situation when the sequence $\big((A_t, B_t)\big)_{t\in\mathbb{Z}}$ is Markov dependent (either it constitutes a stationary Markov chain [87] or is a Markov modulated process [243]). Under appropriate assumptions (including Harris-type conditions), applying the techniques used in the proof below, they describe the tails of the solution to the equation $X_t = A_t X_{t-1} + B_t$, $t \in \mathbb{Z}$.

- The arithmetic case when $\log A$ is supported in the set $a\mathbb{Z}$ for some $a > 0$ was studied by Grincevičius [134]. Under conditions 2 and 3 of Theorem 2.4.4 he proved for all but countably many $y \in \mathbb{R}$ that there exists $c(y) > 0$ such that

$$\mathbb{P}\big(X > \mathrm{e}^{\,y+n\,a}\big) \sim c(y)\,\mathrm{e}^{\,-\alpha\,(y+n\,a)}, \qquad n \to \infty\,. \tag{2.4.58}$$

This result implies in particular that there are constants $c_2 > c_1 > 0$ such that

$$c_1 \le \liminf_{x\to\infty}\; x^\alpha\,\mathbb{P}(|X| > x) \le \limsup_{x\to\infty}\; x^\alpha\,\mathbb{P}(|X| > x) \le c_2.$$

If $\mathbb{P}(B \ge 0) = 1$ and $\mathbb{P}(B = 0) < 1$ then the limit in (2.4.58) exists and is positive for every y.

Next we will give the proof of Theorem 2.4.4. Following Goldie [128], we first prove the tail bounds (2.4.55) and then, developing some independent arguments, we show positivity of $c_+ + c_-$.

Proof (*Proof of Theorem* 2.4.4. *The tail asymptotics.*) The tail bounds (2.4.55) will essentially follow from the fixed-point equation $X \stackrel{d}{=} AX + B$ and the key renewal Theorem A.1.1 on p. 268 which yields a description of the behavior at infinity of the solution to an appropriate renewal equation.

Changing variables, we intend to prove the existence of the limit

$$\lim_{x\to\infty} \mathrm{e}^{\,\alpha x}\,\mathbb{P}(X > \mathrm{e}^{\,x}).$$

Keeping in mind the fixed-point equation for X, we write

$$\mathbb{P}(X > \mathrm{e}^{\,x}) = \mathbb{P}(A\,X > \mathrm{e}^{\,x}) + \psi(x)\,, \tag{2.4.59}$$

where

$$\psi(x) = \mathbb{P}(A\,X + B > \mathrm{e}^{\,x}) - \mathbb{P}(A\,X > \mathrm{e}^{\,x}).$$

With the convention that $f(x) = \mathbb{P}(X > \mathrm{e}^{\,x})$ we can interpret (2.4.59) as a *renewal equation*[8]

$$f(x) = \psi(x) + \mathbb{E}[f(x - \log A)]\,. \tag{2.4.60}$$

[8] We refer to Appendix A for a short introduction to renewal theory.

Unfortunately, we cannot apply Smith's key renewal theorem for positive random variables: in view of the condition $\mathbb{E}[A^\alpha] = 1$ the random variable $\log A$ assumes both positive and negative values with positive probability and its mean is negative; see the comments on p. 48. Instead we will apply a renewal theorem proved in Athreya et al. [19]; see Appendix A for more details. This result concerns real-valued random variables but requires positivity of the mean. This condition is not satisfied in our case.

In this situation a classical trick helps: an exponential change of measure, also known as *Esscher transform*, exploiting the Cramér-Lundberg condition $\mathbb{E}[A^\alpha] = 1$. After this change of the probability measure, $\log A$ still assumes positive and negative values but its mean (with respect to the new measure) is positive. Then we will be able to apply the extended version of the classical renewal theorem in Appendix A.

For any Borel set $C \subset \mathbb{R}$, we define the new probability measure

$$\begin{aligned}\mathbb{P}_\alpha(\log A \in C) &= \frac{\mathbb{E}\big[\mathbf{1}(\log A \in C)\, \mathrm{e}^{\alpha \log A}\big]}{\mathbb{E}[\mathrm{e}^{\alpha \log A}]} \\ &= \frac{\mathbb{E}\big[\mathbf{1}(\log A \in C)\, A^\alpha\big]}{\mathbb{E}[A^\alpha]} \\ &= \mathbb{E}\big[\mathbf{1}(\log A \in C)\, A^\alpha\big].\end{aligned}$$

Denoting the corresponding expected value by $\mathbb{E}_\alpha$, we have for integrable g,

$$\mathbb{E}_\alpha[g(\log A)] = \mathbb{E}\big[g(\log A)\, A^\alpha\big].$$

In particular, $\mathbb{E}_\alpha[\log A] = m_\alpha$ which is finite by assumption and we also have $m_\alpha > 0$ by convexity of the function $\mathbb{E}[A^p]$, $p \in [0, \alpha]$; see the comments on p. 48.

To change the measure we multiply both sides of (2.4.60) by $\mathrm{e}^{\alpha x}$:

$$\mathrm{e}^{\alpha x} f(x) = \mathrm{e}^{\alpha x}\, \psi(x) + \mathrm{e}^{\alpha x}\, \mathbb{E}[f(x - \log A)],$$

and we write $f_\alpha(x) = \mathrm{e}^{\alpha x} f(x)$, $\psi_\alpha(x) = \mathrm{e}^{\alpha x} \psi(x)$. Since

$$\mathrm{e}^{\alpha x} \mathbb{E}[f(x - \log A)] = \mathbb{E}_\alpha\big[f(x - \log A)\, \mathrm{e}^{\alpha(x - \log A)}\big],$$

we arrive at a new renewal equation by change of measure:

$$f_\alpha(x) = \psi_\alpha(x) + \mathbb{E}_\alpha[f_\alpha(x - \log A)]. \tag{2.4.61}$$

If we followed the patterns of classical renewal theory (see Appendix A) one would try to solve (2.4.61) for f_α in terms of ψ_α and the renewal measure given by

$$\nu_\alpha(C) = \sum_{i=1}^{\infty} \mathbb{P}_\alpha\big(S_i \in C\big) \quad \text{for any Borel set } C \subset \mathbb{R}, \tag{2.4.62}$$

where $S_0 = 0$, $S_t = \sum_{i=1}^{t} \log A_i$, $t \geq 0$, is the random walk generated by the iid sequence $(\log A_t)$. Then if ψ_α were directly Riemann integrable (dRi) (see p. 268 for the definition), a typical solution of the renewal equation (2.4.61) would assume the form

$$f_\alpha(x) = \sum_{i=0}^{\infty} \mathbb{E}_\alpha[\psi_\alpha(x - S_i)] = \psi_\alpha(x) + \int_{\mathbb{R}} \psi_\alpha(x - y)\, \nu_\alpha(dy)$$

and the key renewal theorem would describe the behavior of f_α at infinity:

$$\lim_{x\to\infty} f_\alpha(x) = \frac{1}{m_\alpha} \int_{\mathbb{R}} \psi_\alpha(y)\, dy\,.$$

After changing the measure $\mathbb{P}_\alpha$ back to $\mathbb{P}$, the latter relation is exactly the upper tail estimate (2.4.55) with the corresponding constant c_+.

However, we are not in the position to show direct Riemann integrability of ψ_α in a direct way. Since for large values of y, $Ay + B$ is comparable with Ay we expect that the symmetric difference of the events $\{AX + B > \mathrm{e}^{x}\}$ and $\{AX > \mathrm{e}^{x}\}$ is of small measure for large x. Thus the function ψ_α should be very small at infinity. We do not get pointwise estimates of ψ_α but we can prove that ψ_α is integrable.

In what follows, we will use a *smoothing operator* of convolution type: for any integrable function g and $K(s) = \mathrm{e}^{-s}\mathbf{1}_{(0,\infty)}(s)$, $s \in \mathbb{R}$, define

$$\breve{g}(s) = (K * g)(s) = \int_{-\infty}^{s} \mathrm{e}^{-(s-y)}\, g(y)\, dy\,, \qquad s \in \mathbb{R}\,. \tag{2.4.63}$$

The smoothed function $\breve{g}$ preserves the integral properties of g, for example, direct calculation shows $\int_{\mathbb{R}} \breve{g}(y)dy = \int_{\mathbb{R}} g(y)dy$, but $\breve{g}$ also has nice local properties. It is continuous and even dRi. Applying the smoothing operator to both sides of (2.4.61), we obtain

$$\breve{f}_\alpha(s) = \breve{\psi}_\alpha(s) + \mathbb{E}_\alpha[\breve{f}_\alpha(s - \log A)]\,. \tag{2.4.64}$$

Applying the key renewal theorem to (2.4.64), one can solve this equation explicitly and gets

$$\lim_{s\to\infty} \breve{f}_\alpha(s) = \frac{1}{m_\alpha} \int_{\mathbb{R}} \breve{\psi}_\alpha(y)\, dy = \frac{1}{m_\alpha} \int_{\mathbb{R}} \psi_\alpha(y)\, dy = c_+. \tag{2.4.65}$$

Finally, one has to "unsmooth" this result and thus proves the desired relation

$$\lim_{x\to\infty} f_\alpha(x) = c_+ \,. \tag{2.4.66}$$

To fill all the gaps in the above argument we divide the proof into consecutive steps and prove:

Step 1. The function ψ_α is integrable on $\mathbb{R}$, i.e., $\int_{\mathbb{R}} |\psi_\alpha(y)|\,dy < \infty$.

Step 2. If ψ_α is integrable then $\breve{\psi}_\alpha$ is dRi.

Step 3. The solution $\breve{f}_\alpha$ to (2.4.64) satisfies (2.4.65).

Step 4. The function $f_\alpha(x)$ converges to c_+ as $x \to \infty$, i.e., (2.4.66) holds.

Step 5. The constant c_+ is given by (2.4.56).

Proof of step 1. Using the elementary inequalities for $a, b \in \mathbb{R}$, some $C_\alpha > 0$,

$$\begin{aligned} \big||a+b|^\alpha - |a|^\alpha\big| &\le |b|^\alpha && \text{for } \alpha \le 1\,,\\ \big||a+b|^\alpha - |a|^\alpha\big| &\le C_\alpha\,(|a|^{\alpha-1} + |b|^{\alpha-1})\,|b| && \text{for } \alpha > 1\,, \end{aligned}$$

we conclude that

$$\mathbb{E}\big[\big||AX+B|^\alpha - |AX|^\alpha\big|\big] < \infty\,. \tag{2.4.67}$$

Indeed, for $\alpha \le 1$,

$$\mathbb{E}\big[\big||AX+B|^\alpha - |AX|^\alpha\big|\big] \le \mathbb{E}[|B|^\alpha] < \infty\,,$$

and for $\alpha > 1$, using the independence of X and (A, B),

$$\begin{aligned} \mathbb{E}\big[\big||AX+B|^\alpha - |AX|^\alpha\big|\big] &\le C_\alpha\,\mathbb{E}\big[\big(|B|^{\alpha-1} + |AX|^{\alpha-1}\big)\,|B|\big]\\ &= C_\alpha\,\big(\mathbb{E}[|B|^\alpha] + \mathbb{E}\big[|A|^{\alpha-1}|B|\big]\,\mathbb{E}[|X|^{\alpha-1}]\big). \end{aligned}$$

While $\mathbb{E}[|B|^\alpha] < \infty$ by assumption, $\mathbb{E}\big[|A|^{\alpha-1}|B|\big] < \infty$ follows by Hölder's inequality. Indeed, $\mathbb{E}\big[|A|^{\alpha-1}|B|\big] \le (\mathbb{E}[|A|^\alpha])^{(\alpha-1)/\alpha}(\mathbb{E}[|B|^\alpha])^{1/\alpha}$. Finally, $\mathbb{E}[|X|]^{\alpha-1} < \infty$ follows from the second part of Lemma 2.3.1, by observing that we also have $\mathbb{E}[|A|^{\alpha-1}] < 1$; see the comments on p. 48. This proves (2.4.67).

Write $R = (AX+B)_+$ and $Q = (AX)_+$. Then

$$\begin{aligned} \int_{\mathbb{R}} |\psi_\alpha(x)|\,dx &= \int_0^\infty \big|\mathbb{P}(AX+B > s) - \mathbb{P}(AX > s)\big|\,s^{\alpha-1}\,ds\\ &\le \int_0^\infty \mathbb{P}(Q \le s < R)\,s^{\alpha-1}\,ds + \int_0^\infty \mathbb{P}(R \le s < Q)\,s^{\alpha-1}\,ds\,. \end{aligned}$$

Both integrals on the right-hand side can be computed explicitly and are finite. For example, by Fubini's theorem, we have

$$\begin{aligned}\int_0^\infty \mathbb{P}(Q \le s < R)\, s^{\alpha-1}\, ds &= \int_0^\infty \mathbb{E}\big[\mathbf{1}(Q \le s < R)\big]\, s^{\alpha-1} ds \\ &= \mathbb{E}\Big[\mathbf{1}(Q < R) \int_Q^R s^{\alpha-1}\, ds\Big] \\ &= \alpha^{-1}\, \mathbb{E}\Big[\mathbf{1}(Q < R)\big[R^\alpha - Q^\alpha\big]\Big] \\ &\le \alpha^{-1}\, \mathbb{E}\Big[\mathbf{1}(Q < R)\big[|AX + B|^\alpha - |AX|^\alpha\big]\Big], \\ &\le \alpha^{-1}\, \mathbb{E}\big[\big||AX + B|^\alpha - |AX|^\alpha\big|\big],\end{aligned}$$

which is finite in view of (2.4.67). The same calculations without absolute value signs yield $m_\alpha^{-1} \int_{\mathbb{R}} \psi_\alpha(y)\, dy = c_+$.

Proof of step 2. We write ψ_α as a difference of its positive and negative parts: $\psi_\alpha = \psi_\alpha^+ - \psi_\alpha^-$. Recalling the definition of direct Riemann integrability from Appendix A, it suffices to prove that $\breve{\psi}_\alpha^+$ and $\breve{\psi}_\alpha^-$ are dRi. We restrict ourselves to the proof for $\breve{\psi}_\alpha^+$. Observe first that for $\delta > 0$,

$$\begin{aligned}\breve{\psi}_\alpha^+(s + \delta) &= \int_{-\infty}^{s+\delta} \mathrm{e}^{-(s+\delta-u)}\, \psi_\alpha^+(u)\, du \\ &\ge \mathrm{e}^{-\delta} \int_{-\infty}^{s} \mathrm{e}^{-(s-u)}\, \psi_\alpha^+(u)\, du \\ &= \mathrm{e}^{-\delta} \breve{\psi}_\alpha^+(s)\,.\end{aligned}$$

Recalling the definition of the lower sum $\underline{U}$ from Appendix A, we have for $\delta \in (0, 1)$,

$$\begin{aligned}\underline{U}\big(\breve{\psi}_\alpha^+, \delta\big) &= \delta \sum_{t\in\mathbb{Z}} \inf_{s\in[t\delta,(t+1)\delta]} \breve{\psi}_\alpha^+(s) \\ &\ge \delta\, \mathrm{e}^{-\delta} \sum_{t\in\mathbb{Z}} \breve{\psi}_\alpha^+(t\delta) \\ &\ge \mathrm{e}^{-2\delta} \sum_{t\in\mathbb{Z}} \int_{(t-1)\delta}^{t\delta} \breve{\psi}_\alpha^+(u)du \\ &= \mathrm{e}^{-2\delta} \int_{\mathbb{R}} \breve{\psi}_\alpha^+(u)\, du\,.\end{aligned}$$

In the same way we prove that the upper sum $\overline{U}$ satisfies the bound

$$\overline{U}\big(\breve{\psi}_\alpha^+, \delta\big) = \delta \sum_{t\in\mathbb{Z}} \sup_{s\in[t\delta,(t+1)\delta]} \breve{\psi}_\alpha^+(s)$$
$$\leq \mathrm{e}^{2\delta} \int_{\mathbb{R}} \breve{\psi}_\alpha^+(u)\, du\,.$$

Therefore the difference of the upper and lower sums converges to zero as $\delta \to 0$. The same argument also shows that for any small $\delta > 0$

$$\overline{U}\big(\breve{\psi}_\alpha^+, \delta\big) \leq \mathrm{e}^{2\delta} \int_{\mathbb{R}} \breve{\psi}_\alpha^+(u)du \leq \mathrm{e}^{2\delta} \int_{\mathbb{R}} |\psi_\alpha(u)|\, du < \infty.$$

Again appealing to Appendix A, we have proved the direct Riemann integrability of $\breve{\psi}_\alpha^+$. In the same way we prove the direct Riemann integrability of $\breve{\psi}_\alpha^-$.

Proof of step 3. We recall the definition of the random walk with negative drift, $S_0 = 0$, $S_t = \sum_{i=1}^t \log A_i$, $t \geq 1$. Iterating the renewal equation (2.4.64) n times, we obtain

$$\breve{f}_\alpha(s) = \mathbb{E}_\alpha[\breve{f}_\alpha(s - S_n)] + \sum_{k=0}^{n-1} \mathbb{E}_\alpha[\breve{\psi}_\alpha(s - S_k)]\,. \tag{2.4.68}$$

By Fubini's theorem,

$$\mathbb{E}_\alpha[\breve{f}_\alpha(s - S_n)] = \int_{-\infty}^{s} \mathrm{e}^{-(s-y)}\, \mathbb{E}_\alpha[f_\alpha(y - S_n)]\, dy$$
$$= \int_{-\infty}^{s} \mathrm{e}^{-(s-y)} \big(\mathrm{e}^{\alpha y}\, \mathbb{E}[f(y - S_n)]\big)\, dy$$
$$= \int_{-\infty}^{s} \mathrm{e}^{-(s-y)}\, \mathrm{e}^{\alpha y}\, \mathbb{P}(\Pi_n X_0 > \mathrm{e}^{y})\, dy\,.$$

Since $\Pi_n \to 0$ $\mathbb{P}$-a.s. we have, for every $y \in \mathbb{R}$, $\lim_{n\to\infty} \mathbb{P}(\Pi_n X_0 > \mathrm{e}^{y}) = 0$. Dominated convergence yields that $\mathbb{E}[\breve{f}_\alpha(s - S_n)] \to 0$ as $n \to \infty$. Therefore, letting $n \to \infty$ in (2.4.68), we obtain

$$\breve{f}_\alpha(s) = \sum_{k=0}^{\infty} \mathbb{E}_\alpha[\breve{\psi}_\alpha(s - S_k)] = \breve{\psi}_\alpha(s) + \int_{\mathbb{R}} \breve{\psi}_\alpha(s - y)\, \nu_\alpha(dy)\,, \tag{2.4.69}$$

where ν_α is the renewal measure introduced in (2.4.62). Finally, an application of the renewal Theorem A.1.1 yields (2.4.65).

Proof of step 4. Fix a constant $b > 1$. Then

$$\begin{aligned}\frac{b^{\alpha+1}-1}{\alpha+1}\,\mathrm{e}^{\alpha s}\,\mathbb{P}\big(X > \mathrm{e}^{s}\big) &= \mathrm{e}^{-s}\int_{\mathrm{e}^{s}}^{b\mathrm{e}^{s}} u^{\alpha}\,\mathbb{P}\big(X > e^{s}\big)\,du\\ &\geq \mathrm{e}^{-s}\int_{\mathrm{e}^{s}}^{b\mathrm{e}^{s}} u^{\alpha}\,\mathbb{P}\big(X > u\big)\,du\\ &= \int_{s}^{s+\log b} \mathrm{e}^{-(s-y)}\,f_{\alpha}(y)\,dy\\ &= b\,\check{f}_{\alpha}(s+\log b) - \check{f}_{\alpha}(s)\\ &\sim c_{+}\,(b-1)\,, \qquad \text{as } s\to\infty\,,\end{aligned}$$

and passing with b to 1 we have

$$\liminf_{s\to\infty} s^{\alpha}\,\mathbb{P}(X > s) \geq c_{+}\,.$$

A similar argument yields

$$\limsup_{s\to\infty} s^{\alpha}\,\mathbb{P}(X > s) \leq c_{+}\,.$$

Thus we proved (2.4.66).

Proof of step 5. The formula for c_+ follows immediately from the definition of ψ and the arguments presented in the proof of step 1. □

Proof (*Proof of Theorem* 2.4.4*: Positivity of the limiting constant.*) Our aim is to prove that $c_+ + c_- > 0$. In the case when B is almost surely positive (or negative) positivity of c_+ (or c_-) follows immediately from formula (2.4.56). However, when B assumes both positive and negative values this argument does not work. The proof of the asymptotic behavior of $\mathbb{P}(X > x)$ does not contain a clue about the positivity of c_+, c_- either. An appeal to the fixed-point equation $X \stackrel{d}{=} AX + B$ is not helpful for our purposes. Instead we use the explicit representation

$$X = \sum_{i=1}^{\infty} \Pi_{i-1} B_i$$

which is inspired by the stationary solution to the stochastic recurrence equation $X_t = A_t X_{t-1} + B_t$, $t \in \mathbb{Z}$; see Theorem 2.1.3.

Intuition tells us that X is large if one of the products Π_i is large. This event is well described in the literature on ruin and queuing theory; see for example the monographs Feller [120], Chapter XII, and Asmussen and Albrecher [18], Chapter VI.

By virtue of these results we have under the assumptions of the theorem, for some $c_0 > 0$,

$$\mathbb{P}\big(\max_{i\ge 1} \Pi_i > s\big) \sim c_0\, s^{-\alpha}\,, \qquad s \to \infty\,; \tag{2.4.70}$$

see also (9.28) in Goldie [128]. We will use this result to show that there exists a constant $c_1 > 0$ such that

$$\liminf_{s\to\infty} s^{\alpha}\, \mathbb{P}(|X| > s) = c_1,$$

ensuring the positivity of $c_+ + c_-$.

We proceed as follows. We choose x from the support of the distribution of X and consider the *backward process* $(Y_t^x)_{t\ge 0}$ *starting at* x*:*

$$Y_t^x = \Pi_t\, x + \sum_{i=1}^{t} \Pi_{i-1} B_i\,, \quad t \ge 0\,. \tag{2.4.71}$$

Then we take a generic element (A, B) which is also independent of the sequence $(A_t, B_t), t \ge 1$, and write $Z = Ax + B$. In the first step we will find positive constants ε, δ and c such that

$$\mathbb{P}\big(|x - (Ax + B)| - \delta(A + 1) > \varepsilon\big) \ge c\,. \tag{2.4.72}$$

Since for any x we have $\mathbb{P}(Ax + B = x) < 1$, we may choose $\varepsilon > 0$ so small that $2c = \mathbb{P}(|x - Z| > 2\varepsilon) > 0$. Let m be so large that $\mathbb{P}(A > m) \le c$. Then

$$\mathbb{P}(|x - (Ax + B)| > 2\varepsilon\,,\, A \le m) \ge \mathbb{P}(|x - Z| > 2\varepsilon) - \mathbb{P}(A > m) \ge c\,,$$

and for $\delta = \varepsilon/(m + 1)$ we have

$$\begin{aligned}\mathbb{P}(|x - Z| > 2\varepsilon\,,\, A \le m) &= \mathbb{P}(|x - Z| - \delta(A + 1) > 2\varepsilon - \varepsilon(A + 1)/(m + 1), A \le m)\\ &\le \mathbb{P}(|x - Z| - \delta(A + 1) > \varepsilon)\,.\end{aligned}$$

This proves (2.4.72).

Next we observe that

$$\big|Y_t^x - Y_t^Z\big| = \Pi_t\, |x - Z|. \tag{2.4.73}$$

Consider the mutually exclusive events

$$D_t = \big\{\Pi_t > 2s/\varepsilon \text{ and } \Pi_r \le 2s/\varepsilon\,, \quad r \le t - 1\big\}\,, \qquad t = 1, 2, \ldots\,.$$

Applying (2.4.70)–(2.4.73) and using the independence of D_t and (A_{t+1}, B_{t+1}), we obtain for large values s,

$$\begin{aligned}c\,[c_0\,(2s/\varepsilon)^{-\alpha}] &\le c\,\mathbb{P}\big(\max_{t\ge1} \Pi_t > 2s/\varepsilon\big) = c\sum_{t=1}^{\infty}\mathbb{P}(D_t)\\ &\le \sum_{t=1}^{\infty}\mathbb{P}(D_t)\,\mathbb{P}\big(\big|x-(A_{t+1}x+B_{t+1})\big| - \delta(A_{t+1}+1) > \varepsilon\big)\\ &\le \sum_{t=1}^{\infty}\mathbb{P}\big(\big\{\big|Y_t^x - Y_{t+1}^x\big| - \delta(\Pi_{t+1}+\Pi_t) > 2s\big\}\cap D_t\big)\,.\end{aligned}$$

In the last inequality we used the fact that $\varepsilon > 2\,s/\Pi_t$ on D_t. Denoting the probabilities in the last line by π_t, we observe that

$$\begin{aligned}\pi_t &\le \mathbb{P}\big(\big[\big\{\big|Y_t^x\big| - \delta\Pi_t > s\big\}\cup\big\{\big|Y_{t+1}^x\big| - \delta\Pi_{t+1} > s\big\}\big]\cap D_t\big)\\ &\le \mathbb{P}\big(\big\{\big|Y_t^x\big| - \delta\Pi_t > s\big\}\cap D_t\big) + \mathbb{P}\big(\big\{\big|Y_{t+1}^x\big| - \delta\Pi_{t+1} > s\big\}\cap D_t\big)\,.\end{aligned}$$

Therefore, again using that the events (D_t) are mutually exclusive,

$$\begin{aligned}\sum_{t=1}^{\infty}\pi_t &\le \mathbb{P}\big(|Y_t^x| - \delta\Pi_t > s \text{ for some } t\ge1\big)\\ &\qquad + \mathbb{P}\big(|Y_{t+1}^x| - \delta\Pi_{t+1} > s \text{ for some } t\ge1\big)\\ &\le 2\,\mathbb{P}\big(|Y_t^x| - \delta\Pi_t > s \text{ for some } t\ge1\big)\\ &\le 2\,\mathbb{P}\big(\inf_{u\in U}|Y_t^u| > s \text{ for some } t\ge1\big)\end{aligned}$$

for $U = [x-\delta, x+\delta]$. For the last inequality, it is sufficient to observe that (2.4.73) implies

$$|Y_t^x| - \delta\Pi_t \le |Y_t^x - Y_t^u| + |Y_t^u| - \delta\Pi_t = \Pi_t\,(|x-u| - \delta) + |Y_t^u| \le |Y_t^u|\,, \qquad u\in U\,.$$

So far we have shown that

$$\liminf_{s\to\infty} s^{\alpha}\,\mathbb{P}\big(\inf_{u\in U}|Y_t^u| > s \text{ for some } t\big) > 0\,. \tag{2.4.74}$$

Theorems 2.1.1 and 2.1.3 ensure that $Y_t^u \stackrel{\text{a.s.}}{\to} X$ as $t\to\infty$ whatever the value of u. Therefore we expect that, if at some instant the trajectory of $(|Y_t^u|)$ exceeds s, the limit X should exceed s as well and in (2.4.74) we should be able to replace $(|Y_t^u|)_{u\in U}$ by X. We make this intuitive argument precise.

We write

$$X = \sum_{i=1}^{\infty} \Pi_{i-1} B_i = \sum_{i=1}^{t} \Pi_{i-1} B_i + \Pi_t X^t = Y_t^0 + \Pi_t X^t = Y_t^{X^t},$$

where $X^t = \sum_{i=t+1}^{\infty} \Pi_{t+1,i-1} B_i$ is a copy of X and independent of Π_t and Y_t^0; compare with (2.4.71). Since x was chosen as an element of the support of the distribution of X, we have for some positive ε_0,

$$\mathbb{P}(X^t \in U) = \mathbb{P}(X \in U) = \varepsilon_0 > 0.$$

Moreover, if $|Y_t^u| > s$ and $X^t = u$ for some $u \in U$ then $|X| = |Y_t^{X^t}| > s$.

Consider the mutually exclusive events

$$E_t = \Big\{ \inf_{u \in U} |Y_t^u| > s \text{ and } \inf_{u \in U} |Y_i^u| \le s \,, i \le t-1 \Big\}, \quad t = 1, 2, \ldots .$$

Then using the independence of X^t and E_t, we obtain

$$\begin{aligned} &\varepsilon_0 \, \mathbb{P}\big(\inf_{u \in U} |Y_t^u| > s \text{ for some } t \big) \\ &= \varepsilon_0 \sum_{t=1}^{\infty} \mathbb{P}(E_t) = \sum_{t=1}^{\infty} \mathbb{P}(E_t) \, \mathbb{P}\big(X^t \in U\big) = \sum_{t=1}^{\infty} \mathbb{P}\big(E_t \cap \{X^t \in U\}\big) \\ &\le \sum_{t=1}^{\infty} \mathbb{P}\big(E_t \cap \{|Y_t^{X^t}| > s\}\big) = \sum_{t=1}^{\infty} \mathbb{P}\big(E_t \cap \{|X| > s\}\big) \le \mathbb{P}(|X| > s) \,. \end{aligned}$$

Combining the latter bound with (2.4.74) and the first part of the proof, we have finally proved the desired positivity of the tail constants

$$\lim_{s \to \infty} s^{\alpha} \, \mathbb{P}(|X| > s) = c_+ + c_- > 0 \,.$$

□

Remark 2.4.5 Positivity of the limiting constants c_+ and c_- can be proved in various ways. The proof presented above is a reminiscent of the arguments in Goldie [128] and it also contains some ideas of Grincevičius [134]. Below we follow an alternative approach due to Guivarc'h and Le Page [143]. Further independent proofs are available in Buraczewski et al. [72] (based on results from complex analysis) and in Buraczewski and Mentemeier [82] (the proofs are provided in more general settings and exploit large deviation results and the Bahadur-Rao theorem).

Next we will deal with the case when the support of P_X is the real line, supp $P_X = \mathbb{R}$. We refer to Section 2.5.3 for sufficient conditions. It turns out that both constants c_+ and c_- are positive, implying that the left and right tails of P_X are equivalent and cannot decay at different rates; cf. also the comments on p. 48.

Guivarc'h and Le Page [143] proved the following result:

Theorem 2.4.6 *Assume the conditions of Theorem 2.4.4 and, additionally,* $A > 0$ *a.s. and* supp $P_X = \mathbb{R}$. *Then both constants* c_+ *and* c_- *in* (2.4.56) *are positive.*

Proof The original argument of Guivarc'h and Le Page is quite difficult; we present a different approach and prove a stronger result: if the support of P_X is unbounded at $+\infty$, then c_+ is positive.

Recall the explicit representation of $X = \sum_{i=1}^{\infty} \Pi_{i-1} B_i$ and the backward process $Y_t = \sum_{i=1}^{t} \Pi_{i-1} B_i$, $t \geq 1$. We have the identity in law

$$X \stackrel{d}{=} Y_t + \Pi_t X',$$

where X' is a copy of X independent of (Π_t, Y_t). Write

$$\overline{B}_t = \max\{|B_t|, 1\} \qquad \text{and} \qquad \overline{Y}_t = \sum_{i=1}^{t} \Pi_{i-1} \overline{B}_i\,, \qquad t \geq 1\,.$$

Obviously, $|Y_t| \leq \overline{Y}_t$. Under our assumptions,

$$\overline{Y}_t \stackrel{\text{a.s.}}{\rightarrow} \overline{Y} = \sum_{i=1}^{\infty} \Pi_{i-1} \overline{B}_i\,, \qquad t \to \infty\,.$$

An application of the Kesten–Goldie Theorem 2.4.4 yields for some $c_1 > 0$,

$$\lim_{s \to \infty} s^{\alpha}\, \mathbb{P}\big(\overline{Y} > s\big) = c_1\,.$$

In view of (2.4.70) we have for large s and any constant $C > 0$,

$$\begin{aligned}
\frac{c_0}{2}\, s^{-\alpha} &\leq \mathbb{P}\big(\max_{t \geq 1} \Pi_t > s\big)\\
&= \mathbb{P}\big(\Pi_t > s \text{ for some } t \text{ and } \overline{Y} > Cs\big)\\
&\quad + \mathbb{P}\big(\Pi_t > s \text{ for some } t \text{ and } \overline{Y} \leq Cs\big)\\
&\leq \frac{2\,c_1}{C^{\alpha}}\, s^{-\alpha} + \mathbb{P}\big(\Pi_t > s \text{ and } -Cs < Y_t \text{ for some } t\big)\,.
\end{aligned}$$

Choosing a large C such that $2c_1/C^{\alpha} < c_0/4$, we obtain for $\delta = c_0/4 > 0$ and large s,

$$\mathbb{P}\big(\Pi_t > s \text{ and } -Cs < Y_t \text{ for some } t\big) \geq \delta\, s^{-\alpha}\,. \tag{2.4.75}$$

We consider the stopping time

$$T_s = \inf\{t :\ \Pi_t > s \text{ and } -Cs < Y_t\}\,.$$

In view of (2.4.75) we have

$$\begin{aligned}\delta\, s^{-\alpha} &\le \mathbb{P}\big(T_s < \infty\big) = \mathbb{P}\Big(\bigcup_{t=1}^{\infty}\{T_s = t\}\Big)\\ &\le \sum_{t=1}^{\infty}\mathbb{P}\big(\big\{\Pi_t > s \text{ and } -Cs < Y_t\big\}\cap\big\{T_s = t\big\}\big)\,.\end{aligned}$$

Since P_X has unbounded support we have $\mathbb{P}(X' > C+1) > 0$. We conclude that

$$\begin{aligned}\delta\,\mathbb{P}(X' > C+1) &\le s^{\alpha}\sum_{t=1}^{\infty}\mathbb{P}\big(\big\{\Pi_t > s \text{ and } -Cs < Y_t\big\}\cap\big\{T_s = t\big\}\big)\,\mathbb{P}(X' > C+1)\\ &\le s^{\alpha}\sum_{t=1}^{\infty}\mathbb{P}\big(\big\{\Pi_t\, X' + Y_t > s\big\}\cap\big\{T_s = t\big\}\big)\\ &\le s^{\alpha}\,\mathbb{P}(X > s)\,.\end{aligned}$$

This proves the result. □

We mentioned before that Theorem 2.4.4 is valid in a more general setting when A may assume negative values with positive probability as well. Then, however, the constants c_+ and c_- coincide. What happens is that, with probability 1, Π_n change sign infinitely often and contribute in the same way both to $\mathbb{P}(X > x)$ and $\mathbb{P}(X < -x)$. We formulate the corresponding result (Goldie [128], Theorem 4.1) but refrain from giving a proof.

Theorem 2.4.7 *Assume that the following conditions hold.*

1. $\mathbb{P}(A < 0) > 0$ *and the conditional law of* $\log|A|$ *given* $\{A \neq 0\}$ *is non-arithmetic.*
2. *There exists* $\alpha > 0$ *such that* $\mathbb{E}[|A|^{\alpha}]=1$, $\mathbb{E}[|B|^{\alpha}] < \infty$ *and* $\mathbb{E}\big[|A|^{\alpha}\log^{+}|A|\big] < \infty$.
3. $\mathbb{P}(Ax + B = x) < 1$ *for every* $x \in \mathbb{R}$.

Then the equation $X \overset{d}{=} AX + B$ *has a solution* X *which is independent of* (A, B) *and there exists a positive constant* c_+ *such that*

$$\mathbb{P}(X > x) \sim c_+\, x^{-\alpha} \quad \textit{and} \quad \mathbb{P}(X < -x) \sim c_+\, x^{-\alpha}\,, \qquad x \to \infty\,, \tag{2.4.76}$$

where

$$c_+ = \frac{1}{2\,\alpha\, m_\alpha}\,\mathbb{E}\big[|AX+B|^\alpha - |AX|^\alpha\big]$$

and $m_\alpha = \mathbb{E}[|A|^\alpha \log|A|] > 0$.

2.5 The Support

In this section we will study the support of the solution X to the fixed-point equation $X \stackrel{d}{=} AX + B$, where the random variable X and the two-dimensional vector (A, B) are independent. As usual, we write P_Y for the distribution of any random element Y.

The content of this section was essentially communicated to us by Yves Guivarc'h. Some parts of the proof of Proposition 2.5.4 were also clarified by Radoslaw Ceszkiel.

2.5.1 Preliminaries

The Support of a Measure

For our convenience, we recall the notion of support of a measure μ on $\mathbb{R}^d$. Let V be the union of all open subsets $U \subset \mathbb{R}^d$ such that $\mu(U) = 0$. The *support of* μ is the set

$$\operatorname{supp}\mu = V^c = \mathbb{R}^d \setminus V\,.$$

$\circledast$-Convolution

In Section 2.2.4 we introduced the $\circledast$-*convolution* for any probability measure P_0 on $\mathbb{R}$:

$$(P_{(A,B)} \circledast P_0)(C) = \int_{\mathbb{R}} \mathbb{P}(Ax + B \in C)\, P_0(dx) \quad \text{for any Borel set } C. \tag{2.5.77}$$

Then the fixed-point equation $X \stackrel{d}{=} AX + B$ turns into $P_X = P_{(A,B)} \circledast P_X$.

The Semigroup of Affine Transformations on $\mathbb{R}$

In what follows, the *affine transformations* of the real line

$$h(x) = ax + b\,, \qquad x \in \mathbb{R}\,, \tag{2.5.78}$$

for real numbers a and b, will play an essential role. It will be convenient to identify h with the pair $(a, b) \in \mathbb{R} \times \mathbb{R}$ and we will also write $h = (a, b)$. The affine transformations constitute a semigroup[9] $\mathrm{Aff}(\mathbb{R})$ with identity $h_0 = (1, 0)$ and multiplication for $h_i = (a_i, b_i) \in \mathbb{R} \times \mathbb{R}$, $i = 1, 2$,

$$h_1\, h_2 = (a_1, b_1)\,(a_2, b_2) = (a_1 a_2, b_1 + a_1 b_2)\,. \tag{2.5.79}$$

We notice that

$$(h_1 h_2)(x) = h_1(h_2(x))\,, \qquad x \in \mathbb{R}\,.$$

This means that (2.5.78) defines an *action* of $\mathrm{Aff}(\mathbb{R})$ on $\mathbb{R}$. We will also identify $\mathrm{Aff}(\mathbb{R})$ with $\mathbb{R} \times \mathbb{R}$. Usually, $\mathrm{Aff}(\mathbb{R})$ is defined for positive a (as we already did in Example 2.2.11 on p. 28). Then it constitutes a group and the inverse of $h = (a, b)$ is given by $h^{-1} = (a^{-1}, -a^{-1}b)$. However most of the presented results are valid also for negative a and it is convenient to extend this definition.

For integer $n \geq 1$ and $h \in \mathrm{Aff}(\mathbb{R})$, we can now define h^n:

$$h^1(x) = h(x) = ax + b\,, \quad h^n(x) = h(h^{n-1}(x))\,, \quad x \in \mathbb{R}\,. \tag{2.5.80}$$

Then calculation yields

$$h^n = (a^n, b_n)\,, \quad \text{where} \quad b_n = \sum_{i=0}^{n-1} a^i\, b\,. \tag{2.5.81}$$

Consider the solution $x_0 = x_0(h)$ to the fixed-point equation $h(x) = ax + b = x$, i.e.,

$$x_0 = \frac{b}{1-a}\,, \qquad a \neq 1\,.$$

We observe that

$$h(x) = a\,(x - x_0) + x_0\,, \qquad x \in \mathbb{R}\,.$$

Iterating the last identity and recalling the definition of h^n from (2.5.80), we get

$$h^n(x) = a^n\,(x - x_0) + x_0\,, \qquad x \in \mathbb{R}\,, \quad n \geq 1\,. \tag{2.5.82}$$

This means that, modulo the fixed point x_0, the action of h is either contracting or expanding, depending on whether $|a| < 1$ or $|a| > 1$.

---

[9] To get some intuition on this semigroup we encourage the reader to verify the property of associativity.

The following result shows that $h(\text{supp } P_X) = a \text{ supp } P_X + b$ is contained in supp P_X provided $h = (a, b) \in \text{supp } P_{(A,B)}$.

Lemma 2.5.1 *Let X be a solution to the equation $X \stackrel{d}{=} AX + B$. Then for every $(a, b) \in \text{supp } P_{(A,B)}$ the following inclusion holds:*

$$\text{supp } P_{aX+b} = a \text{ supp } P_X + b \subset \text{supp } P_X. \tag{2.5.83}$$

If, in addition, $\mathbb{P}(|A| > 1) > 0$ and $\mathbb{P}(Ax + B = x) < 1$ for every $x \in \mathbb{R}$, then supp P_X *is unbounded.*

Here and in what follows, we exclude the case $Ax + B = x$ a.s. for some $x \in \mathbb{R}$. In this case, supp $P_X = \{x\}$.

Proof Suppose that (2.5.83) does not hold for some $(a_0, b_0) \in \text{supp } P_{(A,B)}$. Then supp $P_{a_0X+b_0} \backslash \text{supp } P_X$ is not empty. Hence there is an open set $C \subset \mathbb{R}$ such that

$$P_{a_0X+b_0}(C) > 0 \quad \text{and} \quad P_X(C) = 0.$$

By the portmanteau theorem (Theorem 2.1 in Billingsley [44]) and since C is open,

$$\liminf_{(a,b)\to(a_0,b_0)} P_{aX+b}(C) > 0\,.$$

Therefore $P_{aX+b}(C) > 0$ for (a, b) in some neighborhood U of (a_0, b_0). Since $(a_0, b_0) \in \text{supp } P_{(A,B)}$ we have $P_{(A,B)}(U) > 0$ and hence

$$P_X(C) = \mathbb{P}(AX + B \in C) \geq \int_U \mathbb{P}(aX + b \in C)\, P_{(A,B)}(d(a, b)) > 0\,,$$

contradicting $P_X(C) = 0$. This proves (2.5.83).

Now we prove the second statement. Choose some $h = (a, b) \in \text{supp } P_{(A,B)}$ such that $a \neq 1$. Then the fixed-point equation $ax + b = x$ has the solution $x_0 = (1 - a)^{-1}b$. Since $\mathbb{P}(Ax + b = x) < 1$ there exists a second point $x_1 \in \text{supp } P_X$ such that $x_1 \neq x_0$.

In view of the first part of the proof, $h(x) = ax + b \in \text{supp } P_X$ for any $x \in \text{supp } P_X$ and therefore also $h^n(x_1) \in \text{supp } P_X$. Since we assume that $\mathbb{P}(|A| > 1) > 0$ we can choose $(a, b) \in \text{supp } P_{(A,B)}$ such that $|a| > 1$. In view of (2.5.82) we then have

$$|h^n(x_1) = \left|a^n\,(x_1 - x_0) + x_0\right| \to \infty.$$

Therefore supp P_X is unbounded. □

2.5.2 The Support is Atomless

The following lemma was proved in Alsmeyer et al. [10] and under slightly stronger assumptions in Grincevičius [132]. We give an alternative proof which we learned from Yves Guivarc'h.

Proposition 2.5.2 *Assume that there is a unique solution X to the equation $X \stackrel{d}{=} AX + B$ and that the following conditions hold:*

1. $\mathbb{P}(Ax + B = x) < 1$ *for every* $x \in \mathbb{R}$.
2. $\mathbb{P}(A = 0) = 0$.

Then P_X does not have atoms and is of pure type, i.e., it is either absolutely continuous or singular with respect to Lebesgue measure.

The assumptions of the proposition cannot be weakened. For example if $\mathbb{P}(A = 0) > 0$ then the measure P_X may have atoms; see Example 2.5.15.

Proof Suppose that P_X has atoms. Since P_X is a probability measure, the value $\max_{x\in\mathbb{R}} P_X(\{x\})$ is attained for finitely many atoms $x_1, \dots, x_k$ for some $k \geq 1$. In view of the identity $X \stackrel{d}{=} AX + B$ we have

$$\begin{aligned} 0 &= P_X(\{x_j\}) - \int_{\mathbb{R}\times\mathbb{R}} \mathbb{P}(a\,X + b = x_j)\, P_{(A,B)}(d(a,b)) \\ &= \int_{\mathbb{R}\times\mathbb{R}} \big[P_X(\{x_j\}) - \mathbb{P}(a\,X + b = x_j)\big]\, P_{(A,B)}(d(a,b))\,. \end{aligned}$$

Since the function $P_X(\{x_j\}) - \mathbb{P}(a\,X + b = x_j)$ is nonnegative it must be zero for any $(a, b) \in \operatorname{supp} P_{(A,B)}$. Therefore

$$P_X(\{x_j\}) = P_X\big(\{a^{-1}(x_j - b)\}\big)\,, \qquad (a, b) \in \operatorname{supp} P_{(A,B)}\,, \qquad j = 1, \dots, k\,,$$

but this is possible only if for each j and (a, b) there exists a unique x_i such that $a^{-1}(x_j - b) = x_i$. In other words, the function $h^{-1}(x) = a^{-1}(x - b)$ only permutes the values x_j, $j = 1, \dots, k$. Writing $x_0 = k^{-1}(x_1 + \cdots + x_k)$, we observe that for any $(a, b) \in \operatorname{supp} P_X$,

$$a x_0 + b = \sum_{j=1}^{k} \frac{a x_j + b}{k} = \frac{x_1 + \cdots + x_k}{k} = x_0\,. \tag{2.5.84}$$

This means that $Ax_0 + B = x_0$ a.s. in contradiction to our assumptions. Thus we proved that P_X has no atoms.

Therefore we have the Lebesgue decomposition

$$P_X = p_1\, P_{\text{abs}} + p_2\, P_{\text{sing}}\,, \tag{2.5.85}$$

where P_{abs} and P_{sing} are unique absolutely continuous and singular probability measures, respectively, and p_1, p_2 are unique nonnegative numbers such that $p_1 + p_2 = 1$. If $p_1 = 1$ or $p_2 = 1$ there is nothing to prove and therefore we assume that $p_i \in (0, 1)$, $i = 1, 2$. Recalling the $\circledast$-notation from (2.5.77), we have

$$P_X = P_{(A,B)} \circledast P_X = p_1\, P_{(A,B)} \circledast P_{\text{abs}} + p_2\, P_{(A,B)} \circledast P_{\text{sing}}\,,$$

and $P_{(A,B)} \circledast P_{\text{abs}}$ is absolutely continuous. Indeed, writing ϕ for the density of P_{abs}, Fubini's theorem and the change of measure formula imply for any Borel set $C \subset \mathbb{R}$,

$$\begin{aligned}\big(P_{(A,B)} \circledast P_{\text{abs}}\big)(C) &= \int_{\mathbb{R}} \mathbb{P}(Ax + B \in C)\, P_{\text{abs}}(dx)\\ &= \int_{\mathbb{R}} \mathbb{P}(Ax + B \in C)\, \phi(x)\, dx\\ &= \int_C \mathbb{E}\big[\phi\big(A^{-1}(x - B)\big)\, A^{-1}\big]\, dx\,,\end{aligned}$$

where the integrand in the last expression is the Lebesgue density of $P_{(A,B)} \circledast P_{\text{abs}}$.

We further decompose

$$P_{(A,B)} \circledast P_{\text{sing}} = \rho_{\text{abs}} + \rho_{\text{sing}}\,,$$

where ρ_{abs} and ρ_{sing} are unique absolutely continuous and singular measures, respectively. If $\rho_{\text{abs}}(\mathbb{R}) \neq 0$ the uniqueness of the decomposition (2.5.85) implies that

$$p_1 = p_1\, P_{\text{abs}}(\mathbb{R}) = p_1\, \big(P_{(A,B)} \circledast P_{\text{abs}}\big)(\mathbb{R}) + p_2\, \rho_{\text{abs}}(\mathbb{R}) = p_1 + p_2\, \rho_{\text{abs}}(\mathbb{R})\,, \tag{2.5.86}$$

contradicting the assumption $p_2 > 0$. Therefore $p_2 = 0$ and $P_X = P_{\text{abs}}$. If ρ_{abs} vanishes we conclude that

$$P_{\text{sing}} = P_{(A,B)} \circledast P_{\text{sing}} = \rho_{\text{sing}}\,.$$

But this is only possible if $P_{\text{sing}} = P_X$. Indeed, the law of X solving $X \stackrel{d}{=} AX + B$ is unique. Hence p_1 must vanish. This proves the theorem. □

2.5.3 *The Structure of the Support*

In this section we will study the structure of the support of P_X.

Consider the subsemigroup $G_{(A,B)}$ of $\mathbb{R} \times \mathbb{R}$ generated by supp $P_{(A,B)}$, i.e.,

$$G_{(A,B)} = \{h_1 \cdots h_n : h_i \in \text{supp}\, P_{(A,B)}\,, \quad i = 1, \ldots, n\,, \quad n \geq 1\}\,,$$

and let $\overline{G}_{(A,B)}$ be its closure with respect to the usual topology on $\mathbb{R} \times \mathbb{R}$. We will identify $h \in \mathbb{R} \times \mathbb{R}$ with a pair $(a, b) \in \mathbb{R} \times \mathbb{R}$. A set $S \subset \mathbb{R}$ is said to be $\overline{G}_{(A.B)}$-*invariant* if for every $h \in \overline{G}_{(A,B)}$ and $x \in S$, $h(x) = ax + b \in S$.

Under the conditions of Lemma 2.5.1, for every $h = (a, b) \in G_{(A,B)}$,

$$h(\text{supp } P_X) = a \text{ supp } P_X + b \subset \text{supp } P_X \,,$$

and since supp P_X is closed by definition, the latter relation remains valid for $h \in \overline{G}_{(A,B)}$. Hence supp P_X is a closed $\overline{G}_{(A,B)}$-invariant set.

Proposition 2.5.3 *Assume that the equation* $X \stackrel{d}{=} AX + B$ *has a unique solution. Then* supp P_X *is given by the set*

$$S_0 = \overline{\{(1-a)^{-1}b : h = (a,b) \in G_{(A,B)} \,, |a| < 1\}} \,. \tag{2.5.87}$$

Furthermore, any $\overline{G}_{(A,B)}$*-invariant closed subset of* $\mathbb{R}$ *contains* supp P_X.

Proof We choose $h = (a, b) \in G_{(A,B)}$ such that $|a| < 1$. Then the unique solution $x_0 = x_0(h) = (1 - a)^{-1}b$ to the fixed-point equation $h(x) = ax + b = x$ exists and $x_0 \in S_0$. Appealing to (2.5.82) and using the fact that $|a| < 1$, we observe that $h^n(x) = a^n(x - x_0) + x_0 \to x_0$ as $n \to \infty$ for any $x \in \mathbb{R}$.

Let $S \subset \mathbb{R}$ be any $\overline{G}_{(A,B)}$-invariant and closed set. Then, in particular, $h^n(x) \in S$ for every $x \in S$ and $h \in \overline{G}_{(A,B)}$, and since S is closed, $\lim_{n\to\infty} h^n(x) = x_0 \in S$. Therefore we have $S_0 \subset S$, and, in particular, $S_0 \subset$ supp P_X.

Next we show that S_0 is $\overline{G}_{(A,B)}$-invariant. Choose $h' = (a', b') \in G_{(A,B)}$ and, as before, $h = (a, b) \in G_{(A,B)}$ with $|a| < 1$. Then $x_0(h)$ is a generic element from a dense subset of S_0. Recall the formula for h^n from (2.5.81). Then we also have

$$h'h^n = (a'a^n, a'b_n + b') \,, \qquad n \geq 1 \,.$$

Hence $|a'a^n| < 1$ for sufficiently large n and therefore

$$x_0(h'h^n) = (1 - a'a^n)^{-1}(a'b_n + b') \in S_0 \,.$$

Moreover,

$$x_0(h'h^n) \to a' \sum_{i=0}^{\infty} a^i \, b + b' = a' \, (1-a)^{-1} \, b + b' = a'x_0(h) + b' = h'(x_0(h)) \,.$$

Since the limit is in S_0, we proved that $h'(S_0) \subset S_0$ for any $h' \in \overline{G}_{(A,B)}$.

Finally, we have to prove that supp $P_X \subset S_0$. Assume this is not the case and there is $x_0 \in$ supp $P_X \backslash S_0$. Since S_0 is closed there is $\delta > 0$ such that the ball $B_\delta(x_0)$ of radius δ centered at x_0 and S_0 are disjoint. Moreover, $P_X(U) > 0$ for $U = B_{\delta/2}(x_0)$. Fix any element $y_0 \in S_0$ and consider the stationary solution $(X_n)_{n\ge 0}$ to the stochastic recurrence equation $X_t = A_t X_{t-1} + B_t$, $t \ge 0$. If $(X_t^{y_0})_{n\ge 0}$ denotes the solution to this equation for $X_0 = y_0$ then $|X_n^{y_0} - X_n| = \Pi_n |y_0 - X_0|$. Since $X_n^{y_0} \in S_0$ for every n (here we use the invariance of the set S_0) and $\Pi_n \stackrel{\text{a.s.}}{\to} 0$ as $n \to \infty$, we have $X_n \notin U$ a.s. for large n. Hence, by the dominated converge theorem, we have

$$0 < P_X(U) = \mathbb{E}\big[\mathbf{1}_U(X_0)\big] = \mathbb{E}\big[\mathbf{1}_U(X_n)\big] = \lim_{n\to\infty} \mathbb{E}\big[\mathbf{1}_U(X_n)\big] = \mathbb{E}\big[\lim_{n\to\infty} \mathbf{1}_U(X_n)\big] = 0\,.$$

Thus we are led to a contradiction. This finally implies supp $P_X \subset S_0$. □

If we have additional information we may obtain a much more precise description of the support of P_X as the following result shows.

Proposition 2.5.4 *Let X be a solution to the equation $X \stackrel{d}{=} AX + B$. If there are $h = (a, b)$ and $h' = (a', b')$ in $G_{(A,B)}$, $0 < a < 1$, $a' > 1$, such that*

$$x_0(h') = \frac{b'}{1-a'} < x_0(h) = \frac{b}{1-a},$$

then $[x_0(h), \infty) \subset$ supp P_X.

Proof Define

$$r = \log a / \log a' < 0\,.$$

We consider two cases.

1. *r is an irrational number.* Writing $\{x\}$ and $[x]$ for the fractional and integer parts of a real x, respectively, we have $k\,r + m = \{kr\} + [kr] + m$ for any nonnegative integers k, m. In view of Weyl's theorem [259], $(\{kr\})_{k\ge k_0}$ is uniformly distributed on $(0, 1)$ in the number-theoretic sense for any choice of $k_0 \ge 0$, hence this sequence is dense in $(0, 1)$. Then the sequence $(l + \{kr\})$ for integer l is dense in $(l, l+1)$ and hence $(k\,r + m)_{k\ge k_0, m\ge 0}$ is dense in $\mathbb{R}$. In turn, the set of the numbers

$$I_0 = \big\{a^k a'^m \,:\, k \ge k_0,\ m \ge 0\big\} = \big\{\mathrm{e}^{\,\log a'(k\,r+m)} \,:\, k \ge k_0,\ m \ge 0\big\}$$

is dense in $\mathbb{R}_+$.

Recalling (2.5.82), we have

$$\begin{aligned} h^k h'^m(x_0(h)) &= h^k\big(a'^m(x_0(h) - x_0(h')) + x_0(h')\big) \\ &= a^k\big[a'^m\big(x_0(h) - x_0(h')) + x_0(h')\big) - x_0(h)\big] + x_0(h) \\ &= \big[x_0(h) + a^k a'^m(x_0(h) - x_0(h'))\big] + a^k(x_0(h') - x_0(h)). \end{aligned}$$

Consider the set

$$I = x_0(h) + \big(x_0(h) - x_0(h')\big) I_0 = \{x_0(h) + a^k a'^m \left(x_0(h) - x_0(h')\right) : k \geq k_0\,, m \geq 0\}.$$

Since I_0 is dense in $\mathbb{R}_+$ and $x_0(h) - x_0(h') > 0$, I is dense in $[x_0(h), \infty)$.

By Proposition 2.5.3, $x_0(h) \in \operatorname{supp} P_X$. Since also $h^k, h'^m \in G_{(A,B)}$, Lemma 2.5.1 ensures that $h^k h'^m(x_0(h)) \in \operatorname{supp} P_X$. Together with the previous argument and since $|a^k(x_0(h') - x_0(h))|$ can be made arbitrarily small for $k \geq k_0$ say, we conclude that $[x_0(h), \infty) \subset \operatorname{supp} P_X$.

2. *r is a rational number.* Then we can find $n_0, m_0 \in \mathbb{N}$ such that $a^{n_0} a'^{m_0} = 1$. Applying (2.5.82) again, we have for any $x \in \mathbb{R}$,

$$\begin{aligned} h^{n_0} h'^{m_0}(x) &= h^{n_0}\big(a'^{m_0}(x - x_0(h')) + x_0(h')\big) \\ &= a^{n_0}\big[\big(a'^{m_0}(x - x_0(h')) + x_0(h')\big) - x_0(h)\big] + x_0(h) \\ &= (x - x_0(h')) + a^{n_0}\,(x_0(h') - x_0(h)) + x_0(h) \\ &= x + \big(x_0(h) - x_0(h')\big) + a^{n_0}\,(x_0(h') - x_0(h))\,. \end{aligned} \tag{2.5.88}$$

Fix $p \in \mathbb{N}$ and $\varepsilon > 0$. Since $(a^{n_0} a'^{m_0})^k = 1$ for any integer k, we can choose infinitely many values $n, m \in \mathbb{N}$ such that $a^n a'^m = 1$. We can also assume that n_0 is so large that

$$|a^{n_0}\,(x_0(h') - x_0(h))| < \varepsilon/p\,.$$

We will prove that $u = \big((h^{n_0} h'^{m_0})^p\big)(x_0(h))$ satisfies

$$|x_0(h) + p\,(x_0(h) - x_0(h')) - u| < \varepsilon.$$

By induction on $l \leq p$, we prove the existence of u_l such that

$$|x_0(h) + l\,(x_0(h) - x_0(h')) - u_l| < l\varepsilon/p. \tag{2.5.89}$$

In view of (2.5.88), $u_1 = h^{n_0} h'^{m_0}(x_0(h))$ satisfies this relation. We construct u_l recursively:

$$u_l = h^{n_0} h'^{m_0}(u_{l-1}) = u_{l-1} + (x_0(h) - x_0(h')) + a^{n_0}\,(x_0(h') - x_0(h))\,.$$

Then we have

$$\begin{aligned}
&\left|x_0(h) + l\,(x_0(h) - x_0(h')) - u_l\right| \\
&= \left|x_0(h) + (l-1)\,(x_0(h) - x_0(h')) - u_{l-1} - a^{n_0}\,(x_0(h') - x_0(h))\right| \\
&\leq \left|x_0(h) + (l-1)\,(x_0(h) - x_0(h')) - u_{l-1}\right| + \left|a^{n_0}\,(x_0(h') - x_0(h))\right| \\
&< l\,\varepsilon/p\,.
\end{aligned}$$

This proves (2.5.89).

If we proceed in the same way for all $p \in \mathbb{N}_+$, small $\varepsilon > 0$ and (m, n) such that $a^n a'^m = 1$, we see that the elements of the set $U = \{(h^n h'^m)^l (x_0(h)) : a^n a'^m = 1\,, m, n, l \in \mathbb{N}\}$ and the values $x_0(h) + p\,(x_0(h) - x_0(h'))$ are arbitrarily close. Therefore

$$\{x_0(h) + p\,(x_0(h) - x_0(h')) : p \in \mathbb{N}\} \subset \overline{U} \subset \operatorname{supp} P_X\,.$$

Next we consider the set

$$I = \{h^k(x_0(h) + p(x_0(h) - x_0(h')) : k, p \in \mathbb{N}\} \subset \operatorname{supp} P_X$$

and we prove that it is dense in $[x_0(h), \infty)$. We proceed as before and obtain

$$h^k(x_0(h) + p\,(x_0(h) - x_0(h')) = a^k\,p\,(x_0(h) - x_0(h')) + x_0(h)\,.$$

Since $a \in (0, 1)$ and $x_0(h) - x_0(h') > 0$, the set

$$\left\{a^k\,p\,(x_0(h) - x_0(h')) : k \geq k_0, k, p \in \mathbb{N}\right\}$$

is dense in $\mathbb{R}_+$ for any $k_0 \geq 1$. Let now $x = x_0(h) + y \in [x_0(h), \infty)$. Given $\varepsilon > 0$ we may choose k such that

$$|a^k p(x_0(h) - x_0(h')) - y| < \varepsilon.$$

Then

$$\left|h^k\big(x_0(h) + p\,(x_0(h) - x_0(h')\big) - x\right| = |a^k p(x_0(h) - x_0(h')) - y| < \varepsilon\,.$$

Thus we proved that the set I is dense in $[x_0(h), \infty)$. This concludes the proof. □

The following result was proved by Guivarc'h and Le Page [143].

Theorem 2.5.5 *Let X be a solution to the equation $X \stackrel{d}{=} AX + B$.*

(1) *Assume the following conditions*

1. $\mathbb{P}(Ax + B = x) < 1$ *for every* $x \in \mathbb{R}$.
2. $A \geq 0$ *a.s.*
3. $\mathbb{P}(0 < A < 1) > 0$.
4. $\mathbb{P}(A > 1) > 0$.

Then the support of P_X is either a half-line or $\mathbb{R}$.

(2) *Assume the following conditions:*

1. $\mathbb{P}(Ax + B = x) < 1$ *for every* $x \in \mathbb{R}$.
2. $\mathbb{P}(A < 0) > 0$ *a.s.*
3. $\mathbb{P}(0 < |A| < 1) > 0$.
4. $\mathbb{P}(|A| > 1) > 0$.

Then supp $P_X = \mathbb{R}$.

Remark 2.5.6 Recently, a stronger result was proved by Alsmeyer et al. [7]: if supp P_X is unbounded then this set is either a half-line or $\mathbb{R}$. In particular, the assumptions $\mathbb{P}(A > 1) > 0$ or $\mathbb{P}(|A| > 1) > 0$ are not needed. Thus even if $\mathbb{P}(|A| \leq 1) = 1$ unboundedness of supp P_X implies that it is a connected set, without any holes.

Proof (*Proof of Theorem* 2.5.5) First, we assume the first set of conditions. We consider two cases.

1. Choose $h = (a, b) \in G_{(A,B)}$ and $a < 1$ and assume there are $h_i = (a_i, b_i) \in G_{(A,B)}$, $a_i > 1$, $i = 1, 2$, such that

$$x_0(h_1) < x_0(h) < x_0(h_2)\,.$$

Then by Proposition 2.5.4 both half-lines $[x_0(h), \infty)$ and $(-\infty, x_0(h)]$ are included in the support of P_X. Hence supp $P_X = \mathbb{R}$.

2. Assume that for every $h \in G_{(A,B)}$ with $a < 1$ and $h' \in G_{(A,B)}$ with $a' > 1$ we have

$$x_0(h') \leq x_0(h).$$

Recalling the definition of the set $S_0 =$ supp P_X from (2.5.87), we see that $x_0(h') \leq \inf S_0$ for every $h' \in G_{(A,B)}$ with $a' > 1$. Applying Proposition 2.5.4 to $x_0(h')$ and $x_0(h)$ we see that the half-line $[x_0(h), \infty)$ is contained in supp P_X. Taking $\varepsilon > 0$ arbitrarily small, we see that $[\inf S_0, \infty) \subset$ supp P_X. On the other hand, supp $P_X \subset [\inf S_0, \infty)$. Hence supp $P_X = [\inf S_0, \infty)$.

The situation when

$$x_0(h') \geq x_0(h)$$

for every $h \in G_{(A,B)}$ with $a < 1$ and $h' \in G_{(A,B)}$ with $a' > 1$ can be handled in the same way, ensuring that supp $P_X = (-\infty, \sup S_0]$.

Now we assume the second set of conditions. Assume that

$$x_0(h') < x_0(h)$$

for $h = (a, b) \in G_{(A,B)}$ with $a < 1$ and $h' = (a', b') \in G_{(A,B)}$ with $a' > 1$. The case when $x_0(h') > x_0(h)$ can be treated in an analogous way.

1. Suppose that $a \in (0, 1)$, $a' < -1$ and write $h'' = (h')^2$. Using h'' instead of h' in the proof of Proposition 2.5.4, we conclude that $[x_0(h), \infty) \subset$ supp P_X. Notice that h'' has the same fixed point as h'. Consider any $x > x_0(h)$. Then

$$h'(x) = a'\,(x - x_0(h')) + x_0(h') < x_0(h').$$

Therefore $(-\infty, h'(x_0(h))] \subset$ supp P_X. It remains to prove that $[h'(x_0(h)), x_0(h)] \subset$ supp P_X. For $y < x_0(h)$ and $k \geq 1$, we have

$$h^k(y) = a^k\,(y - x_0(h)) + x_0(h) \to x_0(h)\,, \qquad k \to \infty\,,$$

hence

$$\bigcup_{k=1}^{\infty} h^k(-\infty, h'(x_0(h))) = (-\infty, x_0(h)) \subset \text{supp } P_X.$$

2. If $a \in (-1, 0)$ and $a' > 1$ we proceed as in the proof of Proposition 2.5.4, now replacing h by h^2. Then we have $[x_0(h), \infty) \subset$ supp P_X, and for $x > x_0(h)$,

$$h(x) = a\,(x - x_0(h)) + x_0(h) < x_0(h)\,,$$

and, in fact,

$$\{h(x) : x > x_0(h)\} = (-\infty, x_0(h)) \subset \text{supp } P_X\,.$$

3. If $a \in (-1, 0)$ and $a' < -1$ we first replace h, h' by $h^2, (h')^2$ in the proof of Proposition 2.5.4 and then we may proceed as above. We omit further details. □

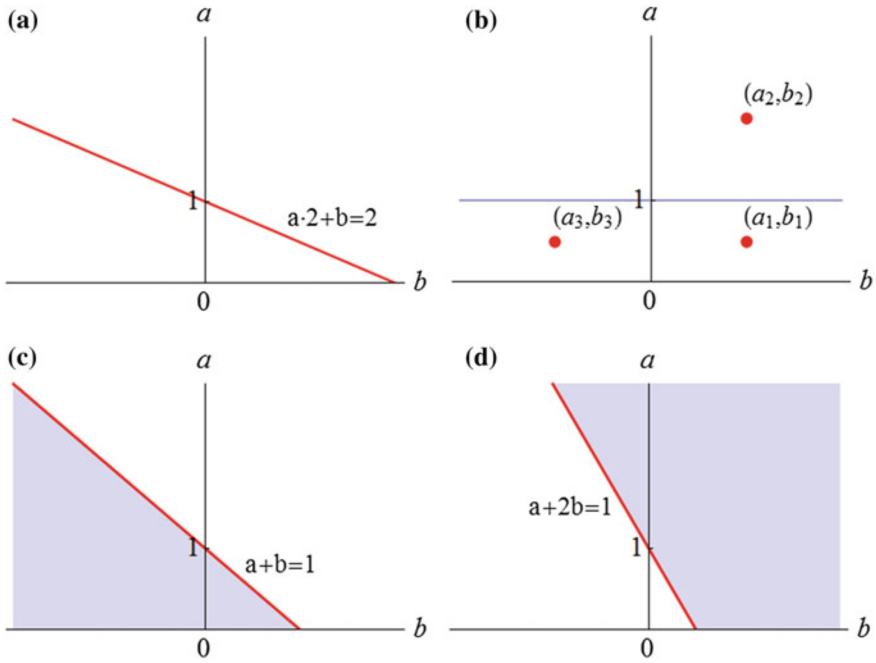

Figure 2.4 The figure illustrates the relation between the supports of $P_{(A,B)}$ and P_X under the assumption $A > 0$ a.s. We identify (A, B) with a point of the upper half-plane $\{(a, b) : a > 0, b \in \mathbb{R}\}$ and assume that a solution X of the equation $X \stackrel{d}{=} AX + B$ exists. Notice that, in contrast to the usual convention, a represents the vertical coordinate. a) If supp $P_{(A,B)}$ is contained in a line passing through $(1, 0)$, i.e., in the set $\{(a, b) : ay_0 + b = y_0\}$ for some y_0, then $Ay_0 + B = y_0$ a.s. and supp $P_X = \{y_0\}$. b) In this example, $(a_1, b_1) = (1/2, 1)$, $(a_2, b_2) = (2, 1)$, $(a_3, b_3) = (1/2, -1)$ with the corresponding fixed points of these mappings $x_1 = 2 > x_2 = -1 > x_3 = -2$. If supp $P_{(A,B)}$ contains these three points Proposition 2.5.4 and Theorem 2.5.5 ensure that supp $P_X = \mathbb{R}$. In this case, no line passes through $(1, 0)$ such that the three points are on the same side of the line. c) If there is a line $\{(a, b) : ay_0 + b = y_0\}$ for some y_0 which contains $(1, 0)$ such that supp $P_{(A,B)}$ is on the left side of this line (i.e., it is a subset of the darker area), then Lemma 2.5.7 yields that supp P_X is bounded from above by y_0 (here $y_0 = 1$). Indeed, for all $h \in$ supp $P_{(A,B)}$ such that $a < 1$, we have $x_0(h) \le y_0$ while for $h \in$ supp $P_{(A,B)}$ with $a > 1$, $x_0(h) \ge y_0$. d) Here we assume that supp $P_{(A,B)}$ is contained in the darker area. The solution X is bounded from below by y_0 (in this case $y_0 = 1/2$).

If $A > 0$ a.s. knowledge of the support of $P_{(A,B)}$ allows one to describe the support of P_X. We present here a simple criterion and illustrate it in Figure 2.4.

Lemma 2.5.7 *Let X be a solution to the equation $X \stackrel{d}{=} AX + B$. Assume $A > 0$ a.s., $\mathbb{P}(A = 1) = 0$, $\mathbb{P}(Ax + B = x) < 1$ for every $x \in \mathbb{R}$, and choose any two distinct points $h_i = (a_i, b_i) \in$ supp $P_{(A,B)}$, $i = 1, 2$.*

(1) *If there exist* h_1 *and* h_2 *such that*

$$a_1 > 1,\ a_2 < 1 \quad \textit{and} \quad x_0(h_1) < x_0(h_2),$$

then supp P_X *is unbounded at* $+\infty$.

(2) *If one has the property*

$$x_0(h_1) \geq x_0(h_2) \quad \textit{for all } h_1, h_2 \textit{ such that } a_2 < 1 < a_1\,,$$

then supp P_X *is bounded from the right side.*

Proof Part (1) is just Proposition 2.5.4. As regards (2) we observe that the half-line $(-\infty, y_0)$ is supp P_X-invariant if we choose $y_0 = \sup\{x_0(h) :\ a < 1\}$. □

2.5.4 Examples

We consider some examples of possible supports of P_X. The structure of supp P_X is completely characterized by the set S_0 in (2.5.87) which sometimes allows one to find subintervals of supp P_X. We give two examples.

Example 2.5.8 Assume that $\{a\} \times (b_1, b_2) \subset \text{supp } P_{(A,B)}$ for some $0 < a < 1$ and $b_1 < b_2$. Then we may immediately conclude from the structure of S_0 that

$$(1-a)^{-1}[b_1, b_2] \subset \text{supp } P_X\,.$$

Example 2.5.9 Assume that $(a_1, a_2) \times \{b\} \subset \text{supp } P_{(A,B)}$ for some $0 < a_1 < a_2 < 1$ and $b > 0$. Then

$$b\left[(1-a_1)^{-1}, (1-a_2)^{-1}\right] \subset \text{supp } P_X\,.$$

In view of the structure of S_0 we may use the same argument if, instead of choosing a subinterval of supp $P_{(A,B)}$, we take it from the nth convolution power given by

$$\{h_1 \cdots h_n : h_i \in \text{supp } P_{(A,B)}\,, i = 1, \ldots, n\}\,.$$

For $n \geq 2$, this is a much richer class than supp $P_{(A,B)}$ and with it, it may be simpler to find a subinterval of supp P_X. It is even easier to find such an interval if one starts from $G_{(A,B)}$.

The aforementioned results are not very surprising in the light of the results about P_X proved in this section. However, if $P_{(A,B)}$ is supported only by a few points and

$0 < A \le 1$ a.s. one will have lower expectations that supp P_X contains an interval or a half-line in its support. A well-known example in this context is the following one:

Example 2.5.10 Choose $A = 0.5$ and B symmetric Bernoulli on $\{-1, 1\}$, i.e., $\mathbb{P}(B = \pm 1) = 0.5$. In Example 2.2.8 we proved that P_X is the uniform distribution on $(-2, 2)$. When $A = a$ for some $a \in (0.5, 1)$ and B is symmetric Bernoulli distributed, Solomyak [251] proved that P_X is absolutely continuous for Lebesgue a.e. $a \in (0.5, 1)$; later a simpler proof was given in Peres and Solomyak [228].

For $a \in (0, 0.5)$, P_X is singularly continuous. We give a short proof. Since $X \stackrel{d}{=} aX + B$ and B, X are independent, we have

$$\text{supp } P_X \subset \big(a \text{ supp } P_X + 1\big) \cup \big(a \text{ supp } P_X - 1\big)\,.$$

Denoting by $|C|$ the Lebesgue measure of any Borel set C, we obtain

$$\begin{aligned} \big|\text{supp } P_X\big| &\le \big|a \text{ supp } P_X + 1\big| + \big|a \text{ supp } P_X - 1\big| \\ &= 2a \,\big|\text{supp } P_X\big|\,. \end{aligned}$$

Since $2a < 1$ this is only possible if supp P_X has Lebesgue measure zero.

Example 2.5.11 Assume that $A = 1/3$ and B assumes both values 0 and $2/3$ with positive probability. Then the support of P_X coincides with the *Cantor set*. This is immediate from the representation

$$X = \sum_{i=1}^{\infty} \Pi_{i-1} B_i = \sum_{i=1}^{\infty} C_i \, 3^{-i}\,,$$

where $C_i = 3B_i$ assumes both values 0 and 2 with positive probability.

Example 2.5.12 In the light of Theorem 2.5.5 the support of P_X always contains a half-line if $A > 0$ a.s. and $\mathbb{P}(A > 1) > 0$, with the one exception when $Ax + B = x$ a.s. for some real x. More precisely, for any two points (a, b), (a', b') in $P_{(A,B)}$ or, more generally, in $G_{(A,B)}$ with $a < 1$ and $a' > 1$, either $[(1-a)^{-1}b, \infty)$ or $(-\infty, (1-a)^{-1}b]$ is included in supp P_X, provided this half-line does not contain $(1-a')^{-1}b'$.

Example 2.5.13 Consider independent A and B such that

$$\mathbb{P}(A = a) = p = 1 - \mathbb{P}(A = a^{-1}) \quad \text{and} \quad \mathbb{P}(B = \pm 1) = 0.5\,,$$

for some $0.5 \le a < 1$ and $0.5 < p \le 1$. Then the condition $\mathbb{E}[\log A] < 0$ is satisfied, the stationary distribution P_X exists and $\mathbb{P}(A > 1) > 0$. Recently, Brieussel and

Tanaka [60] proved that P_X is absolutely continuous. Obviously, $\pm(1-a)^{-1} \in S_0$. Moreover, a simple calculation shows that $(1-a^{-1})^{-1}, -(1-a^{-1})^{-1} \in [-(1-a)^{-1}, (1-a)^{-1}]$. Then Proposition 2.5.4 ensures that $(-\infty, -(1-a)^{-1}]$ and $[(1-a)^{-1}, \infty)$ are subsets of supp P_X. Hence, by Theorem 2.5.5, the support must be the whole real line.

The following example was kindly communicated to us by Aleksander Iksanov.

Example 2.5.14 Recently, Pratsiovytyi and Khvorostina [232] studied recursions based on Lüroth-type alternating expansions. It is known (see Kalpazidou et al. [173]) that for any real number $x \in (0, 1]$ there exists a sequence of positive integers (a_j), finite or infinite, such that

$$x = \frac{1}{a_1} + \sum_{n\geq 2} \frac{-1}{a_1(a_1+1)} \cdots \frac{-1}{a_{n-1}(1+a_{n-1})} \frac{1}{a_n}. \tag{2.5.90}$$

Moreover, each irrational number in (0, 1) has a unique infinite non-periodic representation (2.5.90) and each rational number in (0, 1) has either a finite or a periodic representation (0, 1).

When looking carefully at the above representation, one can recognize a deterministic backward process (Y_t); see (2.2.37) on p. 33. Pratsiovytyi and Khvorostina considered the situation when the sequence (a_n) is random. In particular, they studied the case when the random pairs (A, B) are supported in the set $\big(-1/(k(k+1)), 1/k\big)_{k\in\mathbb{N}}$. Then the forward process has the form

$$X_t = \frac{-1}{a_t(1+a_t)} X_{t-1} + \frac{1}{a_t}$$

for some random natural numbers (a_t). They proved that the solution to this equation is absolutely continuous if and only if

$$\mathbb{P}(a_t = k) = \frac{1}{k(k+1)}\,.$$

In this case, P_X is the uniform distribution on (0, 1).

Throughout, we assumed $|A| > 0$ a.s. If $\mathbb{P}(A = 0) > 0$ and A, B are independent one can easily construct cases where P_X has atoms.

Example 2.5.15 Assume that P_B has an atom $b_0 \neq 0$. Since $X \overset{d}{=} \sum_{i=1}^{\infty} \Pi_{i-1} B_i$ we have

$$\begin{aligned}\mathbb{P}(X = b_0) &= \sum_{k=1}^{\infty} \mathbb{P}\Big(B_1 + \sum_{i=1}^{k-1} \Pi_i\, B_{i+1} = b_0\,, \Pi_{k-1} \neq 0\,, A_k = 0\Big)\\ &\geq \mathbb{P}(B_1 = b_0\,, A_1 = 0) = \mathbb{P}(B_1 = b_0)\,\mathbb{P}(A_1 = 0) > 0\,,\end{aligned}$$

i.e., b_0 is also an atom of P_X. In particular, if $B = 1$ a.s.,

$$\mathbb{P}(X = x) = \sum_{k=1}^{\infty} \mathbb{P}\Big(\sum_{i=1}^{k-1} \Pi_i = x - 1\,,\, \Pi_{k-1} \neq 0\Big)\, \mathbb{P}(A = 0)\,.$$

The right-hand side is positive if $x = 1$ (then $\mathbb{P}(X = 1) \geq \mathbb{P}(A = 0)$) or if $x - 1$ is an atom of the distribution of $\sum_{i=1}^{k} \Pi_i$ for some $k \geq 1$. For example, if there exists $a_0 \neq 0$ such that $\mathbb{P}(A = a_0) = 1 - \mathbb{P}(A = 0)$, P_X has the set of atoms $\{1, 1 + a_0, 1 + a_0 + a_0^2, \ldots\}$.

Springer
http://www.springer.com/978-3-319-29678-4

Stochastic Models with Power-Law Tails
The Equation X = AX + B
Buraczewski, D.; Damek, E.; Mikosch, Th.
2016, XV, 320 p. 9 illus., 5 illus. in color., Hardcover
ISBN: 978-3-319-29678-4