The smoothing transform: a review of contraction results

Gerold Alsmeyer

Abstract Given a sequence $(C,T)=(C,T_1,T_2,...)$ of real-valued random variables, the associated so-called smoothing transform $\mathscr{S}$ maps a distribution F from a subset Γ of distributions on $\mathbb{R}$ to the distribution of $\sum_{i\ge 1}T_iX_i+C$, where $X_1,X_2,...$ are iid with common distribution F and independent of (C,T). This review aims at providing a comprehensive account of contraction properties of $\mathscr{S}$ on subsets Γ specified by the existence of moments up to a given order like, for instance, $\mathscr{P}^p(\mathbb{R})=\{F:\int|x|^p\,F(dx)<\infty\}$ for $p>0$ or $\mathscr{P}_c^p(\mathbb{R})=\{F\in\mathscr{P}^p(\mathbb{R}):\int x\,F(dx)=c\}$ for $p\ge 1$. The metrics used here are the minimal ℓ_p-metric and the Zolotarev metric ζ_p, both briefly introduced in Section 3.

1 Introduction

Any temporally homogeneous Markov chain on the real line or a subset thereof may be described via a random recursive equation *with no branching*, viz.

$$X_n \;=\; \Psi_n(X_{n-1}) \tag{1}$$

for $n\ge 1$ and iid random functions $\Psi_1,\Psi_2,...$ independent of X_0. Namely, if P denotes the one-step transition kernel of the chain and

$$G(x,u) \;:=\; \inf\{y\in\mathbb{R}:P(x,(-\infty,y])\ge u\},\quad x\in\mathbb{R},\,u\in(0,1)$$

its associated pseudo-inverse, then one can choose $\Psi_n(x):=G(x,U_n)$ for $n\ge 1$, where $U_1,U_2,...$ are iid *Unif*(0,1) random variables. Provided that the Ψ_n have additional smoothness properties, for instance, to be (a.s.) globally Lipschitz continuous and contractive in a suitable stochastic sense, stability properties of $(X_n)_{n\ge 0}$ may be

Gerold Alsmeyer
Westfälische Wilhelms-Universität Münster, Institut für Mathematische Statistik, Einsteinstraße 62, 48149 Münster, e-mail: gerolda@math.uni-muenster.de

studied within the framework of *iterated random functions*, see [17] for a survey and [34, 16] for two more recent contributions of interest. Moreover, any stationary distribution π of the chain is then characterized by the distributional identity

$$X \stackrel{d}{=} \Psi(X) \tag{2}$$

where X has law π, Ψ denotes a generic copy of the Ψ_n independent of X, and $\stackrel{d}{=}$ means equality in distribution. (2) is called a *stochastic fixed-point equation (SFPE)* and π (and also X) a solution to it. The case when Ψ is a random affine linear function and solutions are called perpetuities has received particular interest in the literature, see e.g. [44, 20, 4] and further references therein.

A random recursive equation *with branching* occurs if the right-hand side of (1) involves multiple copies of X_{n-1}, i.e.

$$X_n \;=\; \Psi_n(X_{n-1,1}, X_{n-1,2}, ...)$$

for $n \geq 1$, where $(X_{n-1,k})_{k\geq 1}$ is a sequence of iid copies of X_{n-1} and further independent of Ψ_n. Again, of particular interest and also the topic of this article is the situation when the Ψ_n are random affine linear functions, a generic copy thus being of the form

$$\Psi(x_1, x_2, ...) = \sum_{k\geq 1} T_k x_k + C$$

for a sequence of real-valued random variables $(C, T_1, T_2, ...)$. This leads to the so-called (going back to Durrett and Liggett [18]) *smoothing transform(ation)*

$$\mathscr{S}: \quad F \;\mapsto\; \mathscr{L}\left(\sum_{k\geq 1} T_k X_k + C\right) \tag{3}$$

which maps a distribution $F \in \mathscr{P}(\mathbb{R})$ to the law of $\sum_{k\geq 1} T_k X_k + C$, where $X_1, X_2, ...$ are independent of $(C, T_1, T_2, ...)$ with common distribution F. It has been studied by many authors due to its occurrence in various fields of applied probability: probabilistic combinatorial optimization [1], stochastic geometry and random fractals [37, 33, 21], the analysis of recursive algorithms and data structures [39, 22, 41, 36] and branching particle systems [10, 25].

On the event where

$$N := \sum_{k\geq 1} \mathbf{1}_{\{T_k \neq 0\}}$$

is infinite, the sum $\sum_{k\geq 1} T_k X_k$ in (3) is understood as the limit of the finite partial sums $\sum_{k=1}^n T_k X_k$ in the sense of convergence in probability. Then $\mathscr{S}(F)$ is indeed defined *for all* $F \in \mathscr{P}(\mathbb{R})$ if

$$\mathbb{P}(N < \infty) = 1, \tag{A0}$$

but exists only for F from a subset of $\mathscr{P}(\mathbb{R})$ (always containing δ_0) otherwise. Subsets of interest here are typically characterized by the existence of moments of certain order, viz.

$$\mathscr{P}^p(\mathbb{R}) := \left\{F \in \mathscr{P}(\mathbb{R}) : \int |x|^p \, F(dx) < \infty\right\},$$

for any $p > 0$ or, more specifically, the sets of all centered, respectively centered and standardized distributions on $\mathbb{R}$, that is

$$\mathscr{P}_0^1(\mathbb{R}) := \left\{F \in \mathscr{P}^1(\mathbb{R}) : \int x \, F(dx) = 0\right\},$$

$$\mathscr{P}_{0,1}^2(\mathbb{R}) := \left\{F \in \mathscr{P}_0^2(\mathbb{R}) : \int x \, F(dx) = 0 \text{ and } \int x^2 \, F(dx) = 1\right\}.$$

Section 4 will provide conditions for $\mathscr{S}$ to be a self-map on some $\Gamma \subseteq \mathscr{P}^p(\mathbb{R})$, and these do not necessarily include (A0). Under the standing assumption that

$$\mathbb{P}(N \geq 2) > 0, \tag{A1}$$

our goal is then to give a systematic account of conditions under which $\mathscr{S}$ is, in some sense, contractive on Γ with respect to a suitable complete metric ρ and therefore possessing a unique fixed point in Γ, characterized by the SFPE

$$X \stackrel{d}{=} \sum_{k \geq 1} T_k X_k + C \tag{4}$$

when stated in terms of random variables, where $X_1, X_2, \ldots$ are iid copies of X and independent of $(C, T_1, T_2, \ldots)$. Three types of contraction on (Γ, ρ) will be discussed:

- *contraction*, i.e. $\rho(\mathscr{S}(F), \mathscr{S}(G)) \leq \alpha \, \rho(F,G)$ for all $F, G \in \Gamma$ and some $\alpha \in (0,1)$.
- *quasi-contraction*, which holds if $\mathscr{S}^n$ is a contraction for some $n \in \mathbb{N}$.
- *local contraction*, i.e. $\rho(\mathscr{S}^n(F), \mathscr{S}^{n+1}(F)) \leq c \, \alpha^n$ for some $F \in \Gamma$, $\alpha \in (0,1)$ and $c \in \mathbb{R}_>$.

The metrics to be considered here because of their good performance in connection with $\mathscr{S}$ are the *minimal* L^p*-metric* ℓ_p and the *Zolotarev metric* ζ_p for $p > 0$, both briefly introduced in Section 3.

Our review draws on results in [40, 42, 35, 38] supplemented by a number of extensions so as to provide a more complete picture. The last two references may also be consulted for multivariate extensions not discussed here. Further information on the set of solutions to (4), especially for the homogeneous case ($C = 0$), has been obtained by many authors, see [9, 18, 31, 14, 15, 26, 11, 5, 6, 2], but will not either be an issue here. The same goes for results on the tail behavior of solutions, see [23, 30, 32, 27, 28, 29, 3].

The rest of this paper is organized as follows. In Section 2, a brief introduction of the weighted branching model associated with $\mathscr{S}$ is given. It provides the appropriate framework to study the iterates $\mathscr{S}^n$ of $\mathscr{S}$ (Section 2). As already mentioned, Section 3 collects useful information on the probability metrics ℓ_p and ζ_p and Section 4 gives conditions for $\mathscr{S}$ to be a self-map of $\mathscr{P}^p(\mathbb{R})$ or subsets thereof. An auxiliary result on the behavior of the mean values of $\mathscr{S}^n(F)$ for $F\in\mathscr{P}^1(\mathbb{R})$ and as $n\to\infty$ is stated in Section 5. After these preliminaries, all contraction results for $\mathscr{S}$ are presented in the main Section 6, with proofs for some of these results included. Finally, an Appendix provides a short survey of some useful results in connection with Banach's fixed-point theorem, the latter being stated there as well. It also lists some well-known martingale inequalities which form an essential tool for the proofs of the contraction results and are included here to make the presentation more self-contained.

2 The iterates of $\mathscr{S}$ and weighted branching

In order to study contraction properties of $\mathscr{S}$, a representation of $(\mathscr{S}^n(F))_{n\ge1}$, the sequence of iterates of $\mathscr{S}$ applied to some $F\in\mathscr{P}(\mathbb{R})$, in terms of random variables is needed. The weighted branching model to be introduced next and taken from [40] provides an appropriate framework.

Consider the infinite *Ulam-Harris tree*

$$\mathbb{T}\ :=\ \bigcup_{n\ge0}\mathbb{N}^n,\quad \mathbb{N}^0:=\{\varnothing\},$$

of finite integer words having the empty word $\varnothing$ as its root. As common, we write $\mathsf{v}_1\ldots\mathsf{v}_n$ as shorthand for $(\mathsf{v}_1,\ldots,\mathsf{v}_n)$, $|\mathsf{v}|$ for the length of v, and uv for the concatenation of u and v. If $\mathsf{v}=\mathsf{v}_1\ldots\mathsf{v}_n$, put further $\mathsf{v}|0:=\varnothing$ and $\mathsf{v}|k:=\mathsf{v}_1\ldots\mathsf{v}_k$ for $1\le k\le n$. The unique shortest path (geodesic) from the root $\varnothing$ to v, or the ancestral line of v when using a genealogical interpretation, is then given by

$$\mathsf{v}|0=\varnothing\ \to\ \mathsf{v}|1\ \to\ldots\to\ \mathsf{v}|n-1\ \to\ \mathsf{v}|n=\mathsf{v}.$$

The tree $\mathbb{T}$ is now turned into a *weighted (branching) tree* by attaching a *random weight* to each of its edges. Let $T_i(\mathsf{v})$ denote the weight attached to the edge (v,vi) and assume that the $T(\mathsf{v}):=(T_i(\mathsf{v}))_{i\ge1}$ for $\mathsf{v}\in\mathbb{T}$ form a family of iid copies of $T=(T_i)_{i\ge1}$. The number of nonzero weights $T_i(\mathsf{v})$ is denoted $N(\mathsf{v})$, thus

$$N(\mathsf{v})\ :=\ \sum_{i\ge1}\mathbf{1}_{\{T_i(\mathsf{v})\ne0\}}\ \stackrel{d}{=}\ N.$$

Put further $L(\varnothing):=1$ and then recursively

$$L(\mathsf{vi})\ :=\ L(\mathsf{v})T_i(\mathsf{v})$$

for any $\mathsf{v}\in\mathbb{T}$ and $\mathsf{i}\in\mathbb{N}$, which is equivalent to

$$L(\mathsf{v})\ =\ T_{\mathsf{v}_1}(\varnothing)T_{\mathsf{v}_2}(\mathsf{v}|1)\cdot\ldots\cdot T_{\mathsf{v}_n}(\mathsf{v}|n-1)$$

for any $\mathsf{v} = \mathsf{v}_1 \ldots \mathsf{v}_n \in \mathbb{T}$. Hence, $L(\mathsf{v})$ equals the total weight of the minimal path from $\varnothing$ to v obtained upon multiplication of the edge weights along this path.

With the help of a weighted branching model as just introduced, we are now able to describe the iterations of the homogeneous smoothing transform in a convenient way. Namely, if $\mathscr{S}$ is given by (3) with $C = 0$, $\boldsymbol{X} := \{X(\mathsf{v}) : \mathsf{v} \in \mathbb{T}\}$ denotes a family of iid random variables independent of $\boldsymbol{T} := (T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ with common distribution F, and

$$Y_n := \sum_{|\mathsf{v}|=n} L(\mathsf{v}) X(\mathsf{v})$$

for $n \geq 0$, then $\mathscr{S}^n(F) = \mathscr{L}(Y_n)$ holds true for each $n \geq 0$. We call $(Y_n)_{n\geq 0}$ *weighted branching process (WBP) associated with* $\boldsymbol{T} \otimes \boldsymbol{X} := (T(\mathsf{v}), X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$. In the special case when $X(\mathsf{v}) = 1$ for $\mathsf{v} \in \mathbb{T}$, it is simply called *weighted branching process associated with* $\boldsymbol{T}$.

It is not difficult to extend the previous weighted branching model so as to describe the iterations of $\mathscr{S}$ in the nonhomogeneous case when $\mathbb{P}(C = 0) < 1$. To this end, let $\boldsymbol{C} \otimes \boldsymbol{T} = (C(\mathsf{v}), T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ denote a family of iid copies of (C, T), $T := (T_i)_{i\geq 1}$, and $\boldsymbol{X}$ be independent of $\boldsymbol{C} \otimes \boldsymbol{T}$. Then defining $Y(\varnothing) = X(\varnothing)$ and

$$Y_n := \sum_{k=0}^{n-1} \sum_{|\mathsf{v}|=k} L(\mathsf{v}) C(\mathsf{v}) + \sum_{|\mathsf{v}|=n} L(\mathsf{v}) X(\mathsf{v})$$

for $n \geq 1$, it is readily verified that $\mathscr{S}^n(F) = \mathscr{L}(Y_n)$ holds true for each $n \geq 0$. In this case, we call $\boldsymbol{Y} := (Y_n)_{n\geq 0}$ the *weighted branching process associated with* $\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X} := (C(\mathsf{v}), T(\mathsf{v}), X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$.

We proceed to a description of the recursive structure of WBPs after the following useful definition of the *shift operators* $[\cdot]_{\mathsf{v}}$, $\mathsf{v} \in \mathbb{T}$. Given any function Ψ of $\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X}$ and any $\mathsf{v} \in \mathbb{T}$, put

$$[\Psi(\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X})]_{\mathsf{v}} := \Psi\left((C(\mathsf{vw}), T(\mathsf{vw}), X(\mathsf{vw}))_{\mathsf{w}\in\mathbb{T}}\right),$$

which particularly implies

$$[\Psi(\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X})]_{\mathsf{v}} = \Psi\left([\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X}]_{\mathsf{v}}\right).$$

If we think of $\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X}$ as the family of random variables associated with $\mathbb{T}$, then $[\boldsymbol{C} \otimes \boldsymbol{T} \otimes \boldsymbol{X}]_{\mathsf{v}}$ equals its subfamily and copy associated with the subtree $\mathbb{T}(\mathsf{v})$ rooted at v which is isomorphic to $\mathbb{T}$. Obviously, $\boldsymbol{L} := (L(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ is a function of $\boldsymbol{T}$, and one can easily verify that $[\boldsymbol{L}]_{\mathsf{v}} = ([L(\mathsf{w})]_{\mathsf{v}})_{\mathsf{w}\in\mathbb{T}}$ with

$$[L(\mathsf{w})]_{\mathsf{v}} := T_{\mathsf{w}_1}(\mathsf{v}) T_{\mathsf{w}_2}(\mathsf{vw}_1) \cdot \ldots \cdot T_{\mathsf{w}_n}(\mathsf{vw}_1 \ldots \mathsf{w}_{n-1})$$

if $\mathsf{w} = \mathsf{w}_1 \ldots \mathsf{w}_n$. Hence, $[L(\mathsf{w})]_{\mathsf{v}}$ gives the total weight of the minimal path from v to vw. Notice that, for all $\mathsf{v}, \mathsf{w} \in \mathbb{T}$,

$$L(\mathsf{vw}) = L(\mathsf{v}) \cdot [L(\mathsf{w})]_{\mathsf{v}}$$

and therefore

$$[L(\mathsf{w})]_{\mathsf{v}} \;=\; \frac{L(\mathsf{vw})}{L(\mathsf{v})}$$

for all $\mathsf{w} \in \mathbb{T}$ if $L(\mathsf{v}) \neq 0$. For later use, we put

$$\mathscr{F}_n \;:=\; \sigma\left(T(\mathsf{v}) : |\mathsf{v}| \leq n-1\right) \tag{5}$$

for $n \geq 1$ and let $\mathscr{F}_0$ be the trivial σ-field. Observe that $\mathscr{F}_n \supset \sigma(L(\mathsf{v}) : |\mathsf{v}| \leq n)$ for each $n \geq 0$.

Finally, we define

$$\mathfrak{m}(\theta) \;:=\; \mathbb{E}\left(\sum_{i\geq 1} |T_i|^{\theta}\right) \tag{6}$$

for $\theta \geq 0$ which plays an important role in the study of $\mathscr{S}$. For instance, it is well-known that, if $C = 0$ (homogeneous case), $T \geq 0$ and N is bounded, then $\mathscr{S}$ has nontrivial fixed points in $\mathscr{P}(\mathbb{R}_{\geq})$ iff $m(\alpha) = 1$ and

$$\mathfrak{m}'(\alpha) \;=\; \mathbb{E}\left(\sum_{i\geq 1} |T_i|^{\theta} \log |T_i|\right) \;\leq\; 0$$

for some $\alpha \in (0,1]$, see [18]. The function $\mathfrak{m}$ is convex on $\{\theta : \mathfrak{m}(\theta) < \infty\}$, satsifies $\mathfrak{m}(0) = \mathbb{E}N$ and possesses at most two values $\alpha < \beta$ such that $\mathfrak{m}(\alpha) = \mathfrak{m}(\beta) = 1$. If this is the case, then $\mathfrak{m}'(\alpha) < 0$ and $\mathfrak{m}'(\beta) > 0$. The value α is called *characteristic exponent* of T, owing to its role in connection with the existence of fixed points of $\mathscr{S}$. Under appropriate regularity assumptions, the value β determines the tail index of fixed points of $\mathscr{S}$, see [27, 28, 29, 3]. As for the contractive behavior of $\mathscr{S}$ on $\mathscr{P}^p(\mathbb{R})$ or subsets thereof, we will see that $\mathfrak{m}(p) < 1$ constitutes a minimal requirement.

3 Probability metrics

3.1 The minimal L^p-metric

Given a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, let $L^p(\mathbb{P}) = L^p(\Omega, \mathfrak{A}, \mathbb{P})$ for $p > 0$ denote the vector space of p times integrable random variables on $(\Omega, \mathfrak{A}, \mathbb{P})$. Then $\|X\|_p := (\mathbb{E}|X|^p)^{1\wedge(1/p)}$ defines a complete (pseudo-)norm on $L^p(\mathbb{P})$ if $p \geq 1$, but fails to do so if $0 < p < 1$. On the other hand,

$$\ell_p(X,Y) \;:=\; \|X - Y\|_p$$

provides us with a complete (pseudo-)metric on $L^p(\mathbb{P})$ for each $p > 0$.

A pair (X,Y) of real-valued random variables defined on $(\Omega,\mathfrak{A},\mathbb{P})$ is called (F,G)-*coupling* if $\mathscr{L}(X)=F$ and $\mathscr{L}(Y)=G$. In this case, we will use the shorthand notation $(X,Y)\sim(F,G)$ hereafter. For a distribution function F on $\mathbb{R}$, let F^{-1} denote its pseudo-inverse, thus $F^{-1}(u):=\inf\{x\in\mathbb{R}:F(x)\ge u\}$ for $u\in(0,1)$. Then $F^{-1}(U)$ has distribution F if $\mathscr{L}(U)=\mathit{Unif}(0,1)$. Now, for each $p>0$, the mapping $\ell_p:\mathscr{P}^p(\mathbb{R})\times\mathscr{P}^p(\mathbb{R})\to\mathbb{R}_{\ge}$, defined by

$$\ell_p(F,G)\ :=\ \inf_{(X,Y)\sim(F,G)}\|X-Y\|_p, \tag{7}$$

is a metric on $\mathscr{P}^p(\mathbb{R})$, called *minimal L^p-metric* (also *Mallows metric* in [40]). Moreover, the infimum in (7) is attained, namely

$$\ell_p(F,G)\ =\ \|F^{-1}(U)-G^{-1}(U)\|_p$$

for any $\mathit{Unif}(0,1)$ random variable U. The following characterization of convergence with respect to ℓ_p is easily verified.

Proposition 3.1 *Let $p>0$ and $(F_n)_{n\ge0}$ be a sequence of distributions in $\mathscr{P}^p(\mathbb{R})$. Then the following assertions are equivalent:*

(a) $F_n\xrightarrow{\ell_p}F$, *i.e.* $\lim_{n\to\infty}\ell_p(F_n,F)=0$.
(b) $F_n\xrightarrow{w}F$ *and* $\lim_{n\to\infty}\int|x|^p\,F_n(dx)=\int|x|^p\,F(dx)<\infty$.
(c) $F_n\xrightarrow{w}F$ *and* $x\mapsto|x|^p$ *is ui with respect to the* F_n, *that is*

$$\lim_{a\to\infty}\sup_{n\ge1}\int_{(-a,a)^c}|x|^p\,F_n(dx)\ =\ 0.$$

Moreover, the space $(\mathscr{P}^p(\mathbb{R}),\ell_p)$ is complete for each $p>0$.

For any distribution $F\in\mathscr{P}^1(\mathbb{R})$ with mean value $\mathbb{E}F:=\int x\,F(dx)$, let F^0 denote its centering, thus $F^0(t):=F(t+\mathbb{E}F)$ for $t\in\mathbb{R}$. The next lemma provides information about the relation between $\ell_p(F,G)$ and $\ell_p(F^0,G^0)$ for $p\ge1$.

Lemma 3.2 *Given $p\ge1$, distributions $F,G\in\mathscr{P}^p(\mathbb{R})$ with mean values $\mathbb{E}F,\mathbb{E}G$ and a $\mathit{Unif}(0,1)$ random variable U, it holds true that*

$$\ell_p(F^0,G^0)\ =\ \|(F^{-1}(U)-\mathbb{E}F)-(G^{-1}(U)-\mathbb{E}G)\|_p, \tag{8}$$

$$\ell_p(F,G)\ =\ \|((F^0)^{-1}(U)+\mathbb{E}F)-((G^0)^{-1}(U)+\mathbb{E}G)\|_p, \tag{9}$$

and therefore

$$|\ell_p(F,G)-\ell_p(F^0,G^0)|\ \le\ |\mathbb{E}F-\mathbb{E}G|. \tag{10}$$

If $p=2$, then furthermore

$$\ell_2^2(F,G)\ =\ \ell_2^2(F^0,G^0)+(\mathbb{E}F-\mathbb{E}G)^2. \tag{11}$$

Proof. For (8) and (9), it suffices to note that $F^0(t)=F(t+\mathbb{E}F)$ obviously implies $(F^0)^{-1}(t)=F^{-1}(t)-\mathbb{E}F$ for all $t\in\mathbb{R}$. If $p=2$, then (9) with $X:=(F^0)^{-1}(U)$ and

$Y := (G^0)^{-1}(U)$ yields

$$\begin{aligned}\ell_2^2(F,G) &= \mathbb{E}\big((X-Y)+(\mathbb{E}F-\mathbb{E}G)\big)^2\\ &= \mathbb{E}(X-Y)^2+2(\mathbb{E}F-\mathbb{E}G)\,\mathbb{E}(X-Y)+(\mathbb{E}F-\mathbb{E}G)^2\\ &= \ell_2^2(F^0,G^0)+(\mathbb{E}F-\mathbb{E}G)^2,\end{aligned}$$

where $\mathbb{E}X=\mathbb{E}Y=0$ has been utilized. □

3.2 The Zolotarev metric

We now turn to an alternative probability metric which is better tailored to situations where $\mathscr{S}$ is contractive on subsets of $\mathscr{P}^p(\mathbb{R})$ with specified moments of integral order $\le p$.

Let $\mathscr{C}^0(\mathbb{R})$ denote the space of continuous functions $f:\mathbb{R}\to\mathbb{R}$ and $\mathscr{C}^m(\mathbb{R})$ for $m\in\mathbb{N}$ the subspace of m times continuously differentiable complex-valued functions. For $p=m+\alpha$ with $m\in\mathbb{N}_0$ and $0<\alpha\le 1$, put

$$\mathfrak{F}_p := \left\{f\in\mathscr{C}^m(\mathbb{R}) : |f^{(m)}(x)-f^{(m)}(y)|\le|x-y|^\alpha \text{ for all } x,y\in\mathbb{R}\right\}.$$

which obviously contains the monomials $x\mapsto x^k$ for $k=1,\dots,m$ as well as $x\mapsto sign(x)|x|^p/c_p$ and $x\mapsto|x|^p/c_p$ for some $c_p\in\mathbb{R}_>$. Finally, if $p>1$ and thus $m\in\mathbb{N}$, then denote by $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$, $\mathbf{z}=(z_1,\dots,z_m)\in\mathbb{R}^m$, the set of distributions on $\mathbb{R}$ having k^{th} moment z_k for $k=1,\dots,m$.

Zolotarev [46] introduced the metric ζ_p on $\mathscr{P}^p(\mathbb{R})$, defined by

$$\zeta_p(F,G) := \sup_{f\in\mathfrak{F}_p,\,(X,Y)\sim(F,G)}\big|\mathbb{E}\big(f(X)-f(Y)\big)\big| \tag{12}$$

and nowadays named after him. Via a Taylor expansion of the functions $f\in\mathfrak{F}_p$ in (12), it can be shown that $\zeta_p(F,G)$ is finite for all $F,G\in\mathscr{P}^p(\mathbb{R})$ if $0<p\le1$, and for all $F,G\in\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ and $\mathbf{z}\in\mathbb{R}^m$ if $p>1$. On the other hand, in the last case $\zeta_p(F,G)=\infty$ for distributions $F,G\in\mathscr{P}^p(\mathbb{R})$ that do not have the same integral moments up to order m. We thus see that ζ_p defines a proper probability metric on $\mathscr{P}^p(\mathbb{R})$ only for $0<p\le1$ and on $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for any $\mathbf{z}\in\mathbb{R}^m$, otherwise. Here we should add that $\zeta_p(F,G)=0$ implies $F=G$ because $\mathscr{C}^m_b(\mathbb{R}) := \{f\in\mathscr{C}^m(\mathbb{R}) : f^{(m)} \text{ is bounded}\}$ is a measure determining class for each $m\in\mathbb{N}_0$.

Given a probability space $(\Omega,\mathfrak{A},\mathbb{P})$, ζ_p can also be defined on $L^p=L^p(\mathbb{P})$, viz.

$$\zeta_p(X,Y) := \sup_{f\in\mathfrak{F}_p}\big|\mathbb{E}\big(f(X)-f(Y)\big)\big|, \tag{13}$$

and constitutes a *pseudo-metric* there if $0<p\le1$. If $p>1$, then this is true only on $L^p_{\mathbf{z}}=L^p_{\mathbf{z}}(\mathbb{P}) := \{X\in L^p(\mathbb{P}) : \mathbb{E}X^k=z_k \text{ for } k=1,\dots,m\}$ for any $\mathbf{z}\in\mathbb{R}^m$. Recall that a pseudo-metric has the same properties as a metric with one exception: $\zeta_p(X,Y)=0$ does not necessarily imply $X=Y$ (here not even with probability one: just take two iid X,Y which are not a.s. constant).

A pseudo-metric ρ on a set of random variables is called *simple* if it depends only on the marginals of the random variables being compared, and *compound* otherwise. It is called $(p,+)$-*ideal* if

$$\rho(cX,cY) \;=\; |c|^p\,\rho(X,Y) \tag{14}$$

for all $c\in\mathbb{R}$ and

$$\rho(X+Z,Y+Z) \;\le\; \rho(X,Y) \tag{15}$$

for any Z independent of X,Y and with well-defined $\rho(X+Z,Y+Z)$. Obviously, ζ_p is simple, namely

$$\zeta_p(X,Y) \;=\; \zeta_p(F,G)$$

for any random variables X,Y with respective laws F,G, whereas the L^p-pseudo-metrics ℓ_p are compound. It will be shown in Proposition 3.3(a) below that ζ_p is also $(p,+)$-ideal on any $L^p_{\mathbf{z}}$ for $\mathbf{z}\in\mathbb{R}^m$. As for the minimal L^p-metric, one can easily see that it is $(r,+)$-ideal for $r=p\wedge 1$.

In the following, $\mathscr{P}^p_*(\mathbb{R})$, L^p_* stand for $\mathscr{P}^p(\mathbb{R})$, L^p if $0<p\le 1$, and for $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$, $L^p_{\mathbf{z}}$ for arbitrary $\mathbf{z}\in\mathbb{R}^m$ if $p>1$. The subsequent propositions gather some useful properties of ζ_p. For a proof we refer to Zolotarev's original work [46]

Proposition 3.3 *Let $p=m+\alpha$ for some $m\in\mathbb{N}_0$ and $0<\alpha\le 1$. Then ζ_p, defined by* (12) *or* (13)*, has the following properties:*

(a) ζ_p *is a $(p,+)$-ideal pseudo-metric on L^p_*.*

(b) *For any $X,Y\in L^p_*$,*

$$\zeta_p(X,Y) \;\le\; \frac{\Gamma(1+\alpha)}{\Gamma(1+p)}\,\Theta_p(X,Y), \tag{16}$$

where $\Theta_p(X,Y):=\ell_p(X,Y)$ if $0<p=\alpha\le 1$, and

$$\Theta_p(X,Y) \;:=\; \ell_p(X,Y)^\alpha\,\|X\|_p^m + m\,\ell_p(X,Y)\,(\ell_p(X,Y)+\|Y\|_p)^{m-1}$$

if $s\ge 1$.

Convergence with respect to the Zolotarev metric is characterized by a second proposition which may be deduced with the help of the previous one. It particularly shows that ζ_p-convergence and ℓ_p-convergence are equivalent.

Proposition 3.4 *Under the same assumptions as in the previous result, the following properties hold true for ζ_p:*

(a) $\zeta_p(F_n,F)\to 0$ *implies* $\ell_p(F_n,F)\to 0$ *and thus particularly* $F_n\xrightarrow{w}F$ *for any* $F,F_1,F_2,\ldots\in\mathscr{P}^p_*(\mathbb{R})$.

(b) *Conversely,* $\ell_p(F_n,F) \to 0$ *implies* $\Theta_p(F_n,F) \to 0$ *and therefore, by* (16), $\zeta_p(F_n,F) \to 0$ *for any* $F,F_1,F_2,\ldots \in \mathscr{P}^p_*(\mathbb{R})$.

(c) *The metric space* $(\mathscr{P}^p_*(\mathbb{R}),\zeta_p)$ *is complete.*

4 Conditions for $\mathscr{S}$ to be a self-map of $\mathscr{P}^p(\mathbb{R})$

In order to study the contractive behavior of $\mathscr{S}$ on $\mathscr{P}^p(\mathbb{R})$ for $p>0$, we must first provide conditions that ensure that $\mathscr{S}$ is a self-map on this subset of distributions on $\mathbb{R}$. In other words, we need conditions on $(C,T)=(C,(T_i)_{i\ge1})$ such that

$$\sum_{i\ge1} T_iX_i + C \ \in\ L^p$$

whenever the iid $X_1,X_2,\ldots$ are in L^p. Choosing $X_1=X_2=\ldots=0$, we see that $C\in L^p$ is necessary, so that we are left with the problem of finding conditions on T such that $\sum_{i\ge1}T_iX_i\in L^p$ if this is true for the X_i. The main result is stated as Proposition 4.1 below and does not need $N=\sum_{i\ge1}\mathbf{1}_{\{T_i\ne0\}}$ to be a.s. finite. Therefore, $\sum_{i\ge1}T_iX_i\in L^p$ is generally to be understood in the sense of L^p-convergence of the finite partial sums $\sum_{i=1}^n T_iX_i$, which particularly implies convergence in probability. Before stating the result let us define

$$\mathscr{P}^p_c(\mathbb{R}) \ :=\ \left\{F\in\mathscr{P}^p(\mathbb{R}) : \int x\,F(dx)=c\right\}$$

and also $L^p_c:=\{X\in L^p:\mathbb{E}X=c\}$ for $p\ge1$ and $c\in\mathbb{R}$.

Proposition 4.1 *Let* $T=(T_i)_{i\ge1}$ *and* $(X_i)_{i\ge1}$ *be independent sequences on a given probability space* $(\Omega,\mathfrak{A},\mathbb{P})$ *such that* $X_1,X_2,\ldots$ *are iid and in* L^p. *Then each of the following set of conditions implies* $\sum_{i\ge1}T_iX_i\in L^p$*:*

(i) $0<p\le1$ *and* $\sum_{i\ge1}|T_i|^p\in L^1$.
(ii) $1<p\le2$, $\sum_{i\ge1}T_i\in L^p$ *and* $\sum_{i\ge1}|T_i|^p\in L^1$.
(iii) $2\le p<\infty$, $\sum_{i\ge1}T_i\in L^p$ *and* $\sum_{i\ge1}T_i^2\in L^{p/2}$.
(iv) $1<p\le2$, $\sum_{i\ge1}|T_i|^p\in L^1$ *and* $\mathbb{E}X_1=0$.
(v) $2\le p<\infty$, $\sum_{i\ge1}T_i^2\in L^{p/2}$ *and* $\mathbb{E}X_1=0$.

Conversely, if $1<p<\infty$*, then*

(a) $\sum_{i\ge1}T_iX_i\in L^p$ *for any choice of* T*-independent and iid* $X_1,X_2,\ldots$ *in* L^p *implies* $\sum_{i\ge1}T_i\in L^p$ *and* $\sum_{i\ge1}T_i^2\in L^{p/2}$.
(b) $\sum_{i\ge1}T_iX_i\in L^p$ *for any choice of* T*-independent and iid* $X_1,X_2,\ldots\in L^p_0$ *implies* $\sum_{i\ge1}T_i^2\in L^{p/2}$.

It should be observed that, in view of (iii) and (v), the implications in the converse parts (a) and (b) are in fact equivalences if $p\ge2$. It is tacitly understood there that the underlying probability space $(\Omega,\mathfrak{A},\mathbb{P})$ is rich enough to carry T-independent iid $X_1,X_2,...$ with arbitrary distribution in $\mathscr{P}^p(\mathbb{R})$, which is obviously the case if it carries a sequence of iid $Unif(0,1)$ variables. Our proof will show that it is even enough if there exist T-independent iid $X_1,X_2,...$ taking values ± 1 with probability $1/2$ each.

Proof. (i) If $0<p\le 1$, the subadditivity of $x\mapsto x^p$ for $x\ge 0$ immediately implies under the given assumptions that

$$\mathbb{E}\left(\sum_{i\ge 1}|T_iX_i|\right)^p \;\le\; \sum_{i\ge 1}\mathbb{E}|T_iX_i|^p \;=\; \mathbb{E}|X_1|^p\sum_{i\ge 1}\mathbb{E}|T_i|^p \;<\; \infty$$

and thus the almost sure absolute convergence of $\sum_{i\ge 1}T_iX_i$ as well as its integrability of order p.

(ii) Here we argue that $(\sum_{i=1}^n T_iX_i)_{n\ge 1}$ forms a Cauchy sequence in $(L^p(\mathbb{P}),\|\cdot\|_p)$ and is therefore L^p-convergent. First note that $\mathbb{E}(\sum_{i\ge 1}|T_i|^p)=\sum_{i\ge 1}\mathbb{E}|T_i|^p$ implies $T_i\in L^p$ for each $i\ge 1$, which in combination with $X_i\in L^p$ for each $i\ge 1$ ensures that $\sum_{i=m}^n T_iX_i\in L^p$ for all $n\ge m\ge 1$. Denoting by μ the expectation of the X_i, we have that $(\sum_{i=m}^k T_i(X_i-\mu))_{m\le k\le n}$ conditioned upon T forms an L^p-martingale, for T and $(X_i)_{i\ge 1}$ are independent. Since $1<p\le 2$, the even function $x\mapsto |x|^p$ is convex with concave derivative on $\mathbb{R}_{\ge}$ which allows us to make use of the Topchiĭ-Vatutin inequality (see (44) in the Appendix). This yields

$$\mathbb{E}\left(\left|\sum_{i=m}^n T_i(X_i-\mu)\right|^p\middle|\,T\right) \;\le\; 2\,\mathbb{E}|X_1-\mu|^p\sum_{i=m}^n|T_i|^p \quad \text{a.s.}$$

and then by taking unconditional expectations

$$\left\|\sum_{i=m}^n T_i(X_i-\mu)\right\|_p \;\le\; 2\,\|X_1-\mu\|_p\left\|\sum_{i=m}^n|T_i|^p\right\|_1^{1/p}.$$

Since $\sum_{i\ge 1}|T_i|^p\in L^1$, the right-hand side converges to zero as $m,n\to\infty$. By using the second assumption $\sum_{i\ge 1}T_i\in L^p$, we infer that $\lim_{m,n\to\infty}\|\sum_{i=m}^n T_i\|_p=0$ as well, whence finally

$$\left\|\sum_{i=m}^n T_iX_i\right\|_p \;\le\; \left\|\sum_{i=m}^n T_i(X_i-\mu)\right\|_p+|\mu|\left\|\sum_{i=m}^n T_i\right\|_p \;\to\; 0 \tag{17}$$

as $m,n\to\infty$.

(iii) Here we use the same Cauchy sequence argument as in (ii), but make use of Burkholder's inequality (see (8.7) in the Appendix). This yields

$$\mathbb{E}\left(\left|\sum_{i=m}^{n} T_i(X_i-\mu)\right|^p \middle| T\right) \;\le\; b_p^p\,\mathbb{E}\left(\left(\sum_{i=m}^{n} T_i^2(X_i-\mu)^2\right)^{p/2} \middle| T\right) \quad \text{a.s.}$$

for a constant $b_p \in \mathbb{R}_>$ which only depends on p. Next, put $\Sigma_{m:n} := (\sum_{i=m}^n T_i^2)^{1/2}$ for $n \ge m \ge 1$. Given T and $\Sigma_{m:n} \neq 0$, the vector

$$\left(\frac{T_m^2}{\Sigma_{m:n}^2},\dots,\frac{T_n^2}{\Sigma_{m:n}^2}\right)$$

defines a discrete probability distribution on $\{m,\dots,n\}$, which in combination with the independence of T and $(X_i)_{i\ge1}$, the convexity of $x\mapsto x^{p/2}$ for $x\ge0$ and $p\ge2$ and an appeal to Jensen's inequality yields

$$\begin{aligned}
\mathbb{E}\left(\left(\sum_{i=m}^{n} T_i^2(X_i-\mu)^2\right)^{p/2} \middle| T\right) \;&=\; \mathbb{E}\left(\left(\sum_{i=m}^{n} \frac{T_i^2}{\Sigma_{m:n}^2}\Sigma_{m:n}^2(X_i-\mu)^2\right)^{p/2} \middle| T\right)\\
&\le\; \mathbb{E}\left(\sum_{i=m}^{n} \frac{T_i^2}{\Sigma_{m:n}^2}\Sigma_{m:n}^p|X_i-\mu|^p \middle| T\right)\\
&=\; \left(\Sigma_{m:n}^p\sum_{i=m}^{n} \frac{T_i^2}{\Sigma_{m:n}^2}\right)\mathbb{E}|X_1-\mu|^p\\
&=\; \Sigma_{m:n}^p\,\mathbb{E}|X_1-\mu|^p \quad \text{a.s. on } \{\Sigma_{m:n}>0\}.
\end{aligned}$$

But if $\Sigma_{m:n}=0$, the inequality is trivially satisfied. Since, by assumption, $\mathbb{E}\Sigma_{m,n}^p \to 0$ as $m,n\to\infty$, we now obtain by taking unconditional expectations and letting m,n tend to infinity that

$$\lim_{m,n\to\infty}\mathbb{E}\left|\sum_{i=m}^{n} T_i(X_i-\mu)\right|^p \;\le\; b_p^p\,\mathbb{E}|X_1-\mu|^p \lim_{m,n\to\infty}\mathbb{E}\Sigma_{m,n}^p \;=\; 0.$$

The remaining argument via (17) is identical to the one in the previous case and thus not repeated here.

(iv), (v) If $\mu=\mathbb{E}X_1=0$, the assumption in $\sum_{i\ge1}T_i\in L^p$ can be dropped because then the second term on the right-hand side in (17) vanishes.

The converse part:

(a) By choosing $X_i=1$ for $i\ge1$, we find that $\sum_{i\ge1}T_i\in L^p$ and are thus left with a proof of $\sum_{i\ge1}T_i^2\in L^{p/2}$. Let now $X_1,X_2,\dots$ be iid random variables taking values ±1 with probability $1/2$ each. Then $\mathbb{E}X_1=0$, $X_1\in L^p$ for any $p>1$, and $(\sum_{i=1}^n T_iX_i)_{n\ge0}$ conditioned on T forms a L^p-bounded martingale. By another appeal to Burkholder's inequality (49) (lower bound) and observing $X_1^2=1$, it follows that

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} T_i X_i\right|^p \middle| T\right) \geq a_p^p \left(\sum_{i=1}^{n} T_i^2\right)^{p/2} \quad \text{a.s.}$$

for a constant $a_p \in \mathbb{R}_>$ which only depends on p. Consequently,

$$\mathbb{E}\left(\sum_{i\geq 1} T_i^2\right)^{p/2} \leq \frac{1}{a_p^p}\,\mathbb{E}\left|\sum_{i\geq 1} T_i X_i\right|^p < \infty$$

which proves the remaining assertion.

(b) Here it suffices to refer to the last argument. □

In the following, we say that *the smoothing transform $\mathscr{S}$ exists in L^p-sense* if $\mathscr{S}$ is a self-map on $\mathscr{P}^p(\mathbb{R})$. As a direct consequence of Proposition 4.1, one can easily deduce:

Corollary 4.2 *The smoothing transform $\mathscr{S}$ exists*

- *in L^p-sense for $0 < p \leq 1$ if $C \in L^p$ and $\sum_{i\geq 1} |T_i|^p \in L^1$.*
- *in L^p-sense for $1 < p < 2$ if C, $\sum_{i\geq 1} T_i \in L^p$ and $\sum_{i\geq 1} |T_i|^p \in L^1$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ for $1 < p < 2$ if $C \in L^p$ and $\sum_{i\geq 1} |T_i|^p \in L^1$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ for $1 < p \leq 2$ if $C \in L_0^p$ and $\sum_{i\geq 1} |T_i|^p \in L^1$.*
- *in L^p-sense for $2 \leq p < \infty$ iff $C, \sum_{i\geq 1} T_i \in L^p$ and $\sum_{i\geq 1} T_i^2 \in L^{p/2}$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ for $2 \leq p < \infty$ iff $C \in L_0^p$ and $\sum_{i\geq 1} T_i^2 \in L^{p/2}$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ for $2 \leq p < \infty$ iff $C \in L^p$ and $\sum_{i\geq 1} T_i^2 \in L^{p/2}$.*

Conversely, if $\mathscr{S}$ exists

- *in L^p-sense for $1 < p < 2$, then $C, \sum_{i\geq 1} T_i \in L^p$ and $\sum_{i\geq 1} T_i^2 \in L^{p/2}$.*
- *from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}_0^p(\mathbb{R})$ for $1 < p < 2$, then $C \in L_0^p$ and $\sum_{i\geq 1} T_i^2 \in L^{p/2}$.*

In the particularly important case when $T_1, T_2, \ldots$ are nonnegative, a necessary and sufficient condition for $\mathscr{S}$ to exist in L^p-sense can be given for *all* $p > 0$ and follows directly from the previous result if $p > 0$.

Corollary 4.3 *Let $T_1, T_2, \ldots$ be nonnegative and $0 < p < \infty$. Then the smoothing transform $\mathscr{S}$ exists in L^p-sense iff C, $\sum_{i\geq 1} T_i \in L^p$.*

Proof. We must only consider the case $0 < p \leq 1$ and verify that C, $\sum_{i\geq 1} T_i \in L^p$ is necessary for $\mathscr{S}$ to exist in L^p-sense. But choosing $X_i = 0$, we find $C \in L^p$, while choosing $X_i = 1$ for all $i \geq 1$ then further implies $\sum_{i\geq 1} T_i \in L^p$.

5 Convergence of iterated mean values

By Theorem 8.3 in the Appendix, the convergence of $\mathscr{S}^n(F)$ to a fixed point in $(\mathscr{P}^p(\mathbb{R}),\ell_p)$ follows if $\mathscr{S}$ is a continuous locally contractive self-map of this space, thus

$$\ell_p(\mathscr{S}^{n+1}(F),\mathscr{S}^n(F)) \ \le\ c\,\alpha^n \tag{18}$$

for suitable $c\ge 0$, $\alpha\in[0,1)$ and all $n\ge 0$. In order to infer uniqueness of the fixed point, one may consider expected values if $p\ge 1$, which provides the motivation behind the subsequent lemma (see [40, Lemma 1]). Recall that $\mathbb{E}F:=\int x\,F(dx)$ for a distribution $F\in\mathscr{P}^1(\mathbb{R})$.

Lemma 5.1 *Suppose that $\mathscr{S}$ exists in L^p-sense for some $p\ge 1$ and let $F\in\mathscr{P}^p(\mathbb{R})$. Then*

(a) $\mathbb{E}(\sum_{i\ge 1}T_i)\in(-1,1)$ *implies*

$$\lim_{n\to\infty}\mathbb{E}\mathscr{S}^n(F) \ =\ \frac{\mathbb{E}C}{1-\mathbb{E}(\sum_{i\ge 1}T_i)},$$

and the convergence rate is geometric.

(b) $|\mathbb{E}(\sum_{i\ge 1}T_i)|>1$ *and* $\mathbb{E}F+(\mathbb{E}(\sum_{i\ge 1}T_i)-1)^{-1}\,\mathbb{E}C\ne 0$ *imply*

$$\lim_{n\to\infty}|\mathbb{E}\mathscr{S}^n(F)| \ =\ \infty.$$

(c) $|\mathbb{E}(\sum_{i\ge 1}T_i)|>1$ *and* $\mathbb{E}F+(\mathbb{E}(\sum_{i\ge 1}T_i)-1)^{-1}\,\mathbb{E}C=0$ *imply*

$$\lim_{n\to\infty}\mathbb{E}\mathscr{S}^n(F) \ =\ \mathbb{E}F \ =\ \frac{\mathbb{E}C}{1-\mathbb{E}(\sum_{i\ge 1}T_i)}.$$

(d) $\mathbb{E}(\sum_{i\ge 1}T_i)=1$ *and* $\mathbb{E}C\ne 0$ *imply*

$$\lim_{n\to\infty}|\mathbb{E}\mathscr{S}^n(F)| \ =\ \infty.$$

(e) $\mathbb{E}(\sum_{i\ge 1}T_i)=1$ *and* $\mathbb{E}C=0$ *imply* $\mathbb{E}\mathscr{S}^n(F)=\mathbb{E}F$ *for all* $n\ge 0$.

(f) $\mathbb{E}(\sum_{i\ge 1}T_i)=-1$ *implies*

$$\mathbb{E}S^{2n}(F)=\mathbb{E}F \quad \textit{and} \quad \mathbb{E}\mathscr{S}^{2n+1}=\mathbb{E}C-\mathbb{E}F$$

for all $n\ge 0$.

Proof. Fix any $n\ge 1$ and let (C,T), $X_1,X_2,\ldots$ be independent such that $\mathscr{L}(X_i)=\mathscr{S}^{n-1}(F)$ for each $i\ge 1$. Since $\sum_{i\ge 1}T_i\in L^1$ by Corollary 4.2, we infer upon setting $\beta:=\mathbb{E}(\sum_{i\ge 1}T_i)$ that

$$\mathbb{E}\mathscr{S}^n(F) \ =\ \mathbb{E}C+\mathbb{E}\left(\sum_{i\ge 1}T_iX_i\right) \ =\ \mathbb{E}C+\beta\,\mathbb{E}X_1 \ =\ \mathbb{E}C+\beta\,\mathbb{E}\mathscr{S}^{n-1}(F) \tag{19}$$

and then inductively

$$\mathbb{E}\mathscr{S}^n(F) \;=\; \mathbb{E}C\sum_{k=0}^{n-1}\beta^k + \beta^n\mathbb{E}F.$$

All assertions are easily derived from this equation. □

6 Contraction results for $\mathscr{S}$

In view of the results in Section 4, Banach's fixed-point theorem (see the Appendix for a statement of this result along with some generalizations) ensures existence and uniqueness of a fixed point of $\mathscr{S}$ on any of

- $\mathscr{P}^p(\mathbb{R})$ for $p > 0$,
- $\mathscr{P}^p_0(\mathbb{R})$ (a closed subset of $\mathscr{P}^p(\mathbb{R})$) for $p \ge 1$,
- $\mathscr{P}^p_{0,1}(\mathbb{R})$ (a closed subset of $\mathscr{P}^p(\mathbb{R})$) for $p \ge 2$,
- ℓ^p-neighborhoods of a fixed distribution $F \in \mathscr{P}(\mathbb{R})$,
- $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for $p = m+\alpha > 1$ ($m \in \mathbb{N}$, $\alpha \in (0,1]$) and $\mathbf{z} \in \mathbb{R}^m$,

provided that $\mathscr{S}$ is contractive there with respect to ℓ_p (or ζ_p in the last case).

Conditions on (C,T) for this to happen will now be presented in a systematic way. Subsection 6.1 provides a condition on T, different for the cases $0 < p \le 1$ and $p > 1$, under which $\mathscr{S}$ is a contraction on $\mathscr{P}^p(\mathbb{R})$ for $p > 0$ (besides the canonical assumption $C \in L^p$). Situations when $\mathscr{S}$ is still a quasi-contraction on $\mathscr{P}^p(\mathbb{R})$ or $\mathscr{P}^p_c(\mathbb{R})$ for $p > 1$ and $c \in \mathbb{R}$ are discussed in Subsection 6.2. An even weaker property, namely local contractive behavior of $\mathscr{S}$, which still entails existence and uniqueness of a geometrically attracting fixed point, is studied for the case $p > 2$ in Subsection 6.3. All results presented this far are based on the minimal L^p-metric and mainly based on [40]. In Subsection 6.4, ℓ^p-neighborhoods of a fixed distribution $F \in \mathscr{P}(\mathbb{R})$ to be defined there are considered. Drawing on [42], we provide conditions ensuring contraction or quasi-contraction of $\mathscr{S}$ on such neighborhoods, an interesting feature being here that F does not need to be an element of $\mathscr{P}^p(\mathbb{R})$. Finally, Subsection 6.5 deals with the contractive behavior of $\mathscr{S}$ with respect to the Zolotarev metric ζ_p, $p > 1$, on subsets of $\mathscr{P}^p(\mathbb{R})$ with specified moments of integral order is shown under a simple condition on T. The contraction lemma used there is from [38, Prop. 1] (see also [35, Lemma 3.1] for an extension).

6.1 Contraction on $\mathscr{P}^p(\mathbb{R})$

Suppose first that $0 < p \le 1$. Due to the fact that the function $x \mapsto x^p$ is then subadditive on $\mathbb{R}_\ge$, this case is the simplest one.

Theorem 6.1 *Let* $0 < p \le 1$. *If*

$$C \in L^p \quad and \quad \mathfrak{m}(p) < 1,$$

then $\mathscr{S}$ defines a contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Proof. By virtue of the subsequent lemma, $\mathscr{S}$ forms an $\mathfrak{m}(p)$-contraction. Hence, the assertions follow from Banach's fixed-point theorem (Theorem 8.1 in the Appendix) in combination with (20). $\square$

Lemma 6.2 *Let $0 < p \le 1$, $C \in L^p$ and $\sum_{i\ge 1} |T_i|^p \in L^1$. Then*

$$\ell_p(\mathscr{S}(F), \mathscr{S}(G)) \ \le\ \mathfrak{m}(p)\, \ell_p(F,G) \tag{20}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$.

Proof. Pick any $F, G \in \mathscr{P}^p(\mathbb{R})$ and let $(X_1,Y_1),(X_2,Y_2),\ldots$ be iid and (C,T)-independent random variables with $\mathscr{L}(X_1) = F$, $\mathscr{L}(Y_1) = G$ and $\|X_1 - Y_1\|_p = \ell_p(F,G)$. We note that $\mathscr{S}$ exists in L^p-sense by Corollary 4.2. Since $x \mapsto x^p$ is subadditive for $x \ge 0$ and $(\sum_{i\ge 1} T_i X_i + C, \sum_{i\ge 1} T_i Y_i + C) \sim (\mathscr{S}(F), \mathscr{S}(G))$, we infer

$$\begin{aligned} \ell_p(\mathscr{S}(F), \mathscr{S}(G)) \ &\le\ \left\| \sum_{i\ge 1} T_i X_i - \sum_{i\ge 1} T_i Y_i \right\|_p \ =\ \mathbb{E}\left| \sum_{i\ge 1} T_i (X_i - Y_i) \right|^p \\ &\le\ \|X_1 - Y_1\|_p\, \mathbb{E}\left(\sum_{i\ge 1} |T_i|^p \right) \ =\ \mathfrak{m}(p)\, \ell_p(F,G), \end{aligned}$$

which is the assertion. $\square$

Turning to the case $p > 1$, the result corresponding to Theorem 6.1 is due to Rösler [40, Thm. 8] (for the case $p = 2$, see also [38, Prop. 3]].

Theorem 6.3 *Let $p \ge 1$. If*

$$C \in L^p \quad and \quad \left\| \sum_{i\ge 1} |T_i| \right\|_p < 1,$$

then $\mathscr{S}$ is a contraction on $(\mathscr{P}^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point in this space.

Since $\mathfrak{m}(p) \le \| \sum_{i\ge 1} |T_i| \|_p$ for $p \ge 1$, we see that in general it takes a stronger condition for contraction of $\mathscr{S}$ than in the case $0 < p \le 1$.

Proof. Pick any $F, G \in \mathscr{P}^p(\mathbb{R})$ and then as usual iid and (C,T)-independent random variables $(X_1,Y_1),(X_2,Y_2),\ldots$ such that $(X_1,Y_1) \sim (F,G)$ and $\|X_1 - Y_1\|_p = \ell_p(F,G)$. Setting $\Sigma_n := \sum_{i=1}^n |T_i|$, it follows by a similar argument as in the proof of Proposition 4.1(iii) that

$$\mathbb{E}\left(\left(\sum_{i=1}^{n}|T_i(X_i-Y_i)|\right)^p\Bigg|T\right)\ \le\ \Sigma_n^p\,\mathbb{E}|X_1-Y_1|^p\ =\ \Sigma_n^p\,\ell_p^p(F,G)\quad\text{a.s.}$$

for all $n\geq 1$ and therefore upon taking expectations, letting $n\to\infty$ and using the monotone convergence theorem

$$\ell_p(\mathscr{S}(F),\mathscr{S}(G))\ \le\ \left\|\sum_{i\ge 1}|T_i(X_i-Y_i)|\right\|_p\ \le\ \left\|\sum_{i\ge 1}|T_i|\right\|_p\ell_p(F,G).$$

which proves that $\mathscr{S}$ is a contraction on on $(\mathscr{P}^p(\mathbb{R}),\ell_p)$ and thus possesses a unique geometrically attracting fixed point in this set by Banach's fixed-point theorem. □

6.2 Conditions for quasi-contraction if $p>1$

Having settled the case $0<p\leq 1$ with just one condition, viz. $\mathfrak{m}(p)<1$, giving contraction of $\mathscr{S}$ and a unique fixed point on $(\mathscr{P}^p(\mathbb{R}),\ell_p)$, the case $1<p<\infty$ exhibits a more complex picture as shown by three subsequent theorems, which for $p=2$ are all from [40]. The afore-mentioned contraction condition, which figured in the previous subsection, is now replaced with

$$\mathscr{C}_p(T)\ :=\ \mathfrak{m}(p)\vee\mathbb{E}\left(\sum_{i\ge 1}T_i^2\right)^{p/2}\tag{21}$$

which is still $\mathfrak{m}(p)$ if $1<p\leq 2$, but equals $\left\|\sum_{i\ge 1}T_i^2\right\|_{p/2}^{p/2}$ if $p\geq 2$. Plainly, the conditions collapse into one if $p=2$.

Theorem 6.4 *Let $p>1$. If*

$$C\in L_0^p\quad\textit{and}\quad\mathscr{C}_p(T)<1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}_0^p(\mathbb{R}),\ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Theorem 6.5 *Let $p>1$. If*

$$C,\ \sum_{i\ge 1}T_i\in L^p,\quad\mathscr{C}_p(T)<1\quad\textit{and}\quad\left|\mathbb{E}\left(\sum_{i\ge 1}T_i\right)\right|<1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}^p(\mathbb{R}),\ell_p)$ and has a unique geometrically attracting fixed point G_0 in this space.

Theorem 6.6 *Let $p>1$ and $c\in\mathbb{R}$. If*

$$C \in L_0^p, \quad \sum_{i\ge1} T_i \in L^p, \quad \mathscr{C}_p(T) < 1, \quad \textit{and} \quad \mathbb{E}\left(\sum_{i\ge1} T_i\right) = 1,$$

then $\mathscr{S}$ defines a quasi-contraction on $(\mathscr{P}_c^p(\mathbb{R}), \ell_p)$ and has a unique geometrically attracting fixed point G_c in this space. Moreover, if even $\sum_{i\ge1} T_i = 1$ a.s. holds true, then the G_c form a translation family, i.e. $G_c = \delta_c * G_0$ for all $c \in \mathbb{R}$.

We proceed to the statement of two contraction lemmata, treating the cases

- $p = 2$ and $\mathscr{C}_p(T) = \mathfrak{m}(p) = \|\sum_{i\ge1} T_i^2\|_{p/2}^{p/2} < 1$.
- $p > 1$ and $\mathscr{C}_p(T) < 1$.

The proofs of the previous theorems require only the last of these lemmata, but we have included the other one because the provided contraction constant is better for $p = 2$. Recall that F^0 denotes the centering of F if $F \in \mathscr{P}^1(\mathbb{R})$.

Lemma 6.7 *Assuming $C \in L^2$ and $\sum_{i\ge1} T_i^2 \in L^1$, the following assertions hold true:*

(a) *$\mathscr{S}$ exists from $\mathscr{P}_0^2(\mathbb{R}) \to \mathscr{P}^2(\mathbb{R})$ and*

$$\ell_2^2(\mathscr{S}(F^0), \mathscr{S}(G^0)) \;\le\; \left\|\sum_{i\ge1} T_i^2\right\|_1 \ell_2^2(F^0, G^0) \tag{22}$$

for all $F, G \in \mathscr{P}^2(\mathbb{R})$.

(b) *If also $\sum_{i\ge1} T_i \in L^2$, then $\mathscr{S}$ exists in the L^2-sense and*

$$\ell_2^2(\mathscr{S}(F), \mathscr{S}(G)) \;\le\; \left\|\sum_{i\ge1} T_i^2\right\|_1 \ell_2^2(F^0, G^0) + \left\|\sum_{i\ge1} T_i\right\|_2^2 \left(\mathbb{E}F - \mathbb{E}G\right)^2 \tag{23}$$

for all $F, G \in \mathscr{P}^2(\mathbb{R})$.

Proof. See [40, Lemma 2] □

The corresponding lemma for $p > 1$, which appears to be new to our best knowledge (however, see [38, Eq. (2.10)] for part (a) in the case $1 < p \le 2$), is technically more difficult to prove because p^{th} powers of sums can be written out term-wise only for integral p.

Lemma 6.8 *Let $1 < p < \infty$, $C \in L^p$ and $\sum_{i\ge1} |T_i|^p \in L^1$. Then the following assertions hold true:*

(a) *$\mathscr{S}$ exists from $\mathscr{P}_0^p(\mathbb{R}) \to \mathscr{P}^p(\mathbb{R})$ and*

$$\ell_p(\mathscr{S}^n(F^0), \mathscr{S}^n(G^0)) \;\le\; b_p \mathscr{C}_p(T)^{n/p} \ell_p(F^0, G^0) \tag{24}$$

for all $F, G \in \mathscr{P}^p(\mathbb{R})$ and $n \ge 1$.

(b) *If also $\sum_{i\ge 1} T_i \in L^p$, then $\mathscr{S}$ exists in L^p-sense and*

$$\begin{aligned}
&\ell_p(\mathscr{S}^n(F),\mathscr{S}^n(G)) \\
&\quad\le\ b_p\left[\mathscr{C}_p(T)^{n/p}\,\ell_p(F^0,G^0)+n\lambda_p\kappa_p^{n-1}\left|\mathbb{E}F-\mathbb{E}G\right|\right] &(25)\\
&\quad\le\ b_p\left(\frac{n\lambda_p}{\kappa_p}+2\right)\kappa_p^n\,\ell_p(F,G) &(26)
\end{aligned}$$

for all $F,G\in\mathscr{P}^p(\mathbb{R})$ and $n\ge 1$, where

$$\begin{aligned}
\kappa_p\ &:=\ \left|\mathbb{E}\left(\sum_{i\ge 1}T_i\right)\right|\vee\mathscr{C}_p(T)^{1/p}\\
\text{and}\quad \lambda_p\ &:=\ \left\|\sum_{i\ge 1}(T_i-\mathbb{E}T_i)\right\|_p+b_p^{-1}\left\|\sum_{i\ge 1}T_i\right\|_p.
\end{aligned}$$

If $1<p\le 2$, we can choose $b_p=2^{1/p}$ in both parts.

Proof. The existence of $\mathscr{S}$ in the claimed sense is again guaranteed by Corollary 4.2.

(a) Given any $F,G\in\mathscr{P}^p(\mathbb{R})$, let $(X(\mathsf{v}),Y(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ be a family of iid random vectors which is independent of $\boldsymbol{C}\otimes\boldsymbol{T}=(C(\mathsf{v}),T(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ (having the usual meaning) and satisfies $(X(\mathsf{v}),Y(\mathsf{v}))\sim(F^0,G^0)$ and $\|X(\mathsf{v})-Y(\mathsf{v})\|_p=\ell_p(F^0,G^0)$. Consider two WBP $(Z'_n)_{n\ge 0}$ and $(Z''_n)_{n\ge 0}$ associated with $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{X}=(C(\mathsf{v}),T(\mathsf{v}),X(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ and $\boldsymbol{C}\otimes\boldsymbol{T}\otimes\boldsymbol{Y}$, respectively, so that $\mathscr{L}(Z'_n)=\mathscr{S}^n(F^0)$ and $\mathscr{L}(Z''_n)=\mathscr{S}^n(G^0)$ for each $n\ge 0$ [see Section 2]. Furthermore,

$$Z_n\ :=\ Z'_n-Z''_n\ =\ \sum_{|\mathsf{v}|=n}L(\mathsf{v})(X(\mathsf{v})-Y(\mathsf{v})),\quad n\ge 0$$

defines a WBP associated with $\boldsymbol{T}\otimes\boldsymbol{X}-\boldsymbol{Y}=(T(\mathsf{v}),X(\mathsf{v})-Y(\mathsf{v}))_{\mathsf{v}\in\mathbb{T}}$ such that

$$\ell_p(\mathscr{S}^n(F^0),\mathscr{S}^n(G^0))\ \le\ \|Z'_n-Z''_n\|_p\ =\ \|Z_n\|_p$$

for all $n\ge 0$, because $(Z'_n,Z''_n)\sim(\mathscr{S}^n(F^0),\mathscr{S}^n(G^0))$. Write Z_n as

$$Z_n\ =\ L^p\text{-}\lim_{k\to\infty}\sum_{j=1}^{k}L(\mathsf{v}^j)(X(\mathsf{v}^j)-Y(\mathsf{v}^j))$$

for a suitable enumeration $\mathsf{v}^1,\mathsf{v}^2,\ldots$ of $\mathbb{N}^n$ and observe that, conditioned on $\boldsymbol{T}$, the right-hand sum forms an L^p-martingale in $k\ge 1$. As in the proof of Proposition 4.1, we must distinguish the cases $1<p\le 2$ and $p\ge 2$ to complete our argument.

CASE 1: $1<p\le 2$. Then we infer with the help of the Topchiĭ-Vatutin inequality (44) in the Appendix that

$$\begin{aligned}\mathbb{E}(|Z_n|^p|\boldsymbol{T}) &\le 2\lim_{k\to\infty}\sum_{j=1}^{k}|L(\mathsf{v}^j)|^p\,\mathbb{E}|X(\mathsf{v}^j)-Y(\mathsf{v}^j)|^p\\ &= 2\sum_{j\ge1}|L(\mathsf{v}^j)|^p\,\mathbb{E}|X(\mathsf{v}^j)-Y(\mathsf{v}^j)|^p\\ &= 2\,\ell_p(F^0,G^0)^p\sum_{|\mathsf{v}|=n}|L(\mathsf{v})|^p\quad\text{a.s.}\end{aligned}$$

One can easily verify that $\mathbb{E}(\sum_{|\mathsf{v}|=n}|L(\mathsf{v})|^p)=\|\sum_{i\ge1}|T_i|^p\|_1^n$. Hence, we obtain (24) by taking unconditional expectation in the previous estimation.

CASE 2: $p\ge2$. Put $\Sigma_1^2:=\sum_{i\ge1}T_i(\varnothing)^2$. By proceeding as in the proof of Proposition 4.1(iii), but with $X(i)-Y(i)$ instead of $X_i-\mu$ and $m=1$, $n=\infty$, it then follows by use of Burkholder's inequality and Jensen's inequality that

$$\begin{aligned}&\mathbb{E}\left(\left|\sum_{i\ge1}T_i(\varnothing)(X(i)-Y(i))\right|^p\middle|\boldsymbol{T}\right)\\ &\quad\le b_p^p\,\mathbb{E}\left(\left(\sum_{i\ge1}T_i(\varnothing)^2(X(i)-Y(i))^2\right)^{p/2}\middle|\boldsymbol{T}\right)\\ &\quad\le b_p^p\,\Sigma_1^p\,\mathbb{E}\left(\left(\sum_{i\ge1}\frac{T_i(\varnothing)^2}{\Sigma_1^2}(X(i)-Y(i))^2\right)^{p/2}\middle|\boldsymbol{T}\right)\\ &\quad\le b_p^p\,\Sigma_1^p\,\mathbb{E}|X(1)-Y(1)|^p\\ &\quad\le b_p^p\,\Sigma_1^p\,\ell_p^p(F^0,G^0)\quad\text{a.s.}\end{aligned}$$

and thereby

$$\ell_p(\mathscr{S}(F^0),\mathscr{S}(G^0))\ \le\ \left\|\sum_{i\ge1}T_i(X(i)-Y(i))\right\|_p\ \le\ b_p\,\|\Sigma\|_p\,\ell_p(F^0,G^0),$$

where b_p only depends on p. This proves (24) for $n=1$. But in the same manner, we obtain for general n

$$\begin{aligned}\ell_p(\mathscr{S}^n(F),\mathscr{S}^n(G)) &\le \left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})(X(\mathsf{v})-Y(\mathsf{v}))\right\|_p\\ &\le b_p\,\|\Sigma_n\|_p\,\ell_p(F^0,G^0),\end{aligned}$$

where $\Sigma_n^2:=\sum_{|\mathsf{v}|=n}L(\mathsf{v})^2$. Hence, the proof of (24) will be complete once we have shown that

$$\|\Sigma_n\|_p\le\|\Sigma\|_p^n \tag{27}$$

for all $n\ge1$. To this end put $\Sigma(\mathsf{v}):=\sum_{i\ge1}T_i(\mathsf{v})^2$ for $\mathsf{v}\in\mathbb{T}$ and recall from (5) that $\mathscr{F}_k=\sigma(T(\mathsf{v}):|\mathsf{v}|\le k-1)$ for $k\ge1$. Then

$$\begin{aligned}\mathbb{E}(\Sigma_n^p|\mathscr{F}_{n-1}) &= \mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n-1} L(\mathsf{v})^2\Sigma(\mathsf{v})^2\right)^{p/2}\middle|\mathscr{F}_{n-1}\right)\\ &= \Sigma_{n-1}^p\,\mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n-1}\frac{L(\mathsf{v})^2}{\Sigma_{n-1}^2}\Sigma(\mathsf{v})^2\right)^{p/2}\middle|\mathscr{F}_{n-1}\right)\\ &\le \Sigma_{n-1}^p\,\mathbb{E}\left(\sum_{|\mathsf{v}|=n-1}\frac{L(\mathsf{v})^2}{\Sigma_{n-1}^2}\Sigma(\mathsf{v})^p\middle|\mathscr{F}_{n-1}\right)\\ &= \Sigma_{n-1}^p\,\|\Sigma\|_p^p \quad \text{a.s.}\end{aligned}$$

for each $n\ge 2$, which clearly gives (27) upon taking expectations and iteration.

(b) Let us first note that it suffices to show (25) because then (26) can be easily deduced with the help of (10) and the obvious inequality $|\mathbb{E}F-\mathbb{E}G|\le \ell_p(F,G)$, namely

$$\begin{aligned}&\mathscr{C}_p(T)^{n/p}\,\ell_p(F^0,G^0)+n\lambda_p\kappa_p^{n-1}\left|\mathbb{E}F-\mathbb{E}G\right|\\ &\quad\le \mathscr{C}_p(T)^{n/p}\,\ell_p(F,G)+\left(\frac{n\lambda_p}{\kappa_p}+1\right)\kappa_p^n\left|\mathbb{E}F-\mathbb{E}G\right|\\ &\quad\le \left(\frac{n\lambda_p}{\kappa_p}+2\right)\kappa_p^n\,\ell_p(F,G)\end{aligned}$$

for all $F,G\in\mathscr{P}^p(\mathbb{R})$.

Similar to the proof of part (b) of the previous lemma, we obtain with the help of part (a) and Minkowski's inequality that

$$\begin{aligned}\ell_p(\mathscr{S}^n(F),\mathscr{S}^n(G)) &\le \left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\Big(\big(X(\mathsf{v})-Y(\mathsf{v})\big)+(\mathbb{E}F-\mathbb{E}G)\Big)\right\|_p\\ &\le \left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\big(X(\mathsf{v})-Y(\mathsf{v})\big)\right\|_p+|\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\right\|_p\\ &= \|Z_n\|_p+|\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\right\|_p\\ &\le b_p\mathscr{C}_p(T)^{n/p}\,\ell_p(F^0,G^0)+|\mathbb{E}F-\mathbb{E}G|\left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\right\|_p \qquad (28)\end{aligned}$$

for all $n\ge 1$, where b_p can be chosen as $2^{1/p}$ if $1<p\le 2$. This leaves us with the task to give an estimate for $a_n:=\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})\|_p$, which will be accomplished by another martingale argument involving the Topchiĭ-Vatutin inequality if $1<p\le 2$, and the Burkholder inequality if $p\ge 2$.

CASE 1: $1<p\le 2$. We put $U(\mathsf{v}):=\sum_{i\ge 1}T_i(\mathsf{v})$, $\alpha:=\mathscr{C}_p(T)^{1/p}$, $\beta:=\mathbb{E}U(\mathsf{v})$ and $\gamma:=\|U(\mathsf{v})-\beta\|_p=\|\sum_{i\ge 1}T_i-\beta\|_p$. Since $\sum_{i\ge 1}T_i\in L^p$ and $p>1$, we have $|\beta|\le a_1<\infty$. By a similar argument as in (a), we see that $\sum_{|v|=n}L(\mathsf{v})(U(\mathsf{v})-\beta)$ conditioned on $\mathscr{F}_n$ is the limit of an L^p-martingale (use that $U(\mathsf{v})$ is independent of $\mathscr{F}_n$), whence the Topchiĭ-Vatutin inequality yields

$$\mathbb{E}\left(\left|\sum_{|v|=n}L(\mathsf{v})(U(\mathsf{v})-\beta)\right|^p\middle|\mathscr{F}_n\right)\ \le\ 2\,\gamma^p\sum_{|\mathsf{v}|=n}|L(\mathsf{v})|^p\quad\text{a.s.}$$

As a consequence,

$$\begin{aligned}a_{n+1}\ &=\ \left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})U(\mathsf{v})\right\|_p\\ &\le\ \left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})(U(\mathsf{v})-\beta)\right\|_p+|\beta|a_n\\ &\le\ 2^{1/p}\gamma\left\|\sum_{|\mathsf{v}|=n}|L(\mathsf{v})|^p\right\|_1^{1/p}+|\beta|a_n\\ &=\ 2^{1/p}\gamma\alpha^n+|\beta|a_n \end{aligned}\tag{29}$$

for all $n\ge 1$, which leads to

$$\begin{aligned}a_{n+1}\ &\le\ 2^{1/p}\gamma\sum_{k=0}^{n-1}|\beta|^k\alpha^{n-k}+|\beta|^n a_1\\ &\le\ (n+1)(2^{1/p}\gamma+a_1)(|\beta|\vee\alpha)^n\ =\ (n+1)2^{1/p}\lambda_p\kappa_p^n \end{aligned}\tag{30}$$

for all $n\ge 1$. Since this inequality trivially holds for $n=0$, we finally obtain the asserted inequality (25) from (28) and (30).

CASE 2: $p\ge 2$. In this case, we obtain with the Burkholder inequality that

$$\begin{aligned}\mathbb{E}\left(\left|\sum_{|\mathsf{v}|=n}L(\mathsf{v})(U(\mathsf{v})-\beta)\right|^p\middle|\mathscr{F}_n\right)\ &\le\ b_p^p\,\mathbb{E}\left(\left(\sum_{|\mathsf{v}|=n}L(\mathsf{v})^2(U(\mathsf{v})-\beta)^2\right)^{p/2}\middle|\mathscr{F}_n\right)\\ &\le\ b_p^p\,\gamma^p\,\Sigma_n^p\quad\text{a.s.}\end{aligned}$$

which upon taking expectations on both sides and using (29) provides us with

$$\left\|\sum_{|\mathsf{v}|=n}L(\mathsf{v})(U(\mathsf{v})-\beta)\right\|_p\ \le\ b_p\,\gamma\|\Sigma_1\|_p^n\ =\ b_p\,\gamma\alpha^n$$

and thus [see also (29)]

$$a_{n+1} \;\le\; \left\| \sum_{|\mathsf{v}|=n} L(\mathsf{v})(U(\mathsf{v})-\beta) \right\|_p + |\beta| a_n \;\le\; b_p \,\gamma \alpha^n + |\beta| a_n \tag{31}$$

for all $n \ge 1$. For the remaining arguments we can refer to the previous case. □

Now we can turn to the proofs of the theorems stated above.

Proof (of Theorem 6.4). As $\mathbb{E}C = 0$ is assumed, $\mathscr{S}$ defines a self-map of $\mathscr{P}_0^p(\mathbb{R})$ by Corollary 4.2. It is also an α-contraction on $(\mathscr{P}_0^p(\mathbb{R}), \ell_p)$ with $\alpha := \|\sum_{i\ge1}(T_i)^2\|_{p/2}^{1/p}$ if $p = 2$ [by Lemma 6.7(a)], and an α_m-quasi-contraction with $\alpha_m := b_p\,\alpha^m$ for suitable $m \ge 1$ if $p > 1$ [by Lemma 6.8(a)]. Therefore, the assertion follows from Banach's fixed-point theorem 8.1 or its generalization 8.2 in combination with the contraction inequality (22) or (24), respectively. □

Proof (of Theorem 6.5). The existence of $\mathscr{S}$ in L^p-sense follows again from Corollary 4.2, while contraction inequality (26) shows that $\mathscr{S}$ is a quasi-contraction on $\mathscr{P}^p(\mathbb{R})$, viz.

$$\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; c\,\kappa^n \,\ell_p(F,G)$$

for any $\kappa \in (0, \kappa_p)$, $F, G \in \mathscr{P}^p(\mathbb{R})$, $n \ge 1$ and a suitable $c = c(\kappa) > 0$. All assertions now follow from Banach's fixed-point theorem 8.2 for quasi-contractions. □

Proof (of Theorem 6.6). First note that $\mathbb{E}C = 0$ and $\mathbb{E}(\sum_{i\ge1} T_i) = 1$ entail $\mathbb{E}\mathscr{S}(F) = \mathbb{E}F = c$ for all $F \in \mathscr{P}_c^p(\mathbb{R})$. Hence, $\mathscr{S}$ is a self-map of $\mathscr{P}_c^p(\mathbb{R})$ for any $c \in \mathbb{R}$. Moreover, (25) simplifies to

$$\ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; b_p \left\| \sum_{i\ge1} T_i^2 \right\|_{p/2}^{n/2} \ell_p(F,G)$$

for all $n \ge 1$ and $F, G \in \mathscr{P}_c^p(\mathbb{R})$ because $\ell_p(F,G) = \ell_p(F^0, G^0)$. Hence $\mathscr{S}$ is also a quasi-contraction on $\mathscr{P}_c^p(\mathbb{R})$ and therefore has a unique fixed point G_c by Theorem 8.2. It remains to verify that $G_c = \delta_c * G_0$ in the case when $\sum_{i\ge1} T_i = 1$ a.s. By the uniqueness property of G_c, it suffices to verify that $\mathscr{S}(\delta_c * G_0) = \delta_c * G_0$. Choose iid (C,T)-independent random variables $X_1, X_2, \ldots$ with law G_0. Then

$$\begin{aligned}\mathscr{S}(\delta_c * G_0) \;&=\; \mathscr{L}\left(\sum_{i\ge1} T_i(X_i + c) + C\right) \;=\; \mathscr{L}\left(\sum_{i\ge1} T_i X_i + c + C\right)\\ &=\; \delta_c * \mathscr{L}\left(\sum_{i\ge1} T_i X_i + C\right) \;=\; \delta_c * \mathscr{S}(G_0) \;=\; \delta_c * G_0\end{aligned}$$

yields the desired conclusion. □

6.3 Conditions for local contraction if $p > 2$

If $p > 2$ and $\sum_{i\ge 1} T_i \in L^p$ is replaced by the generally stronger condition $\sum_{i\ge 1} |T_i| \in L^p$, then we can trade in the contraction condition $\|\sum_{i\ge 1} T_i^2\|_{p/2} < 1$ for a weaker one and still obtain local contraction in the sense that

$$\lim_{n\to\infty} \rho^{-n} \ell_p(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;=\; 0$$

for some $\rho \in (0,1)$ and all $F,G \in \mathscr{P}^p(\mathbb{R})$ or $\in \mathscr{P}_0^p(\mathbb{R})$. As a consequence, existence and uniqueness of a geometrically attractive fixed point in these sets still holds. For integral $p > 2$, the following two theorems are again due to Rösler [40, Thms. 9 and 10]. Note that $\mathfrak{m}(q) \vee \mathfrak{m}(p) < 1$ for $0 < q < p < \infty$ implies $\mathfrak{m}(r) < 1$ for any $r \in [q,p]$ because $\mathfrak{m}$ is convex on $[2,p]$.

Theorem 6.9 *Let $p > 2$. If*

$$C \in L_0^p, \quad \sum_{i\ge 1} |T_i| \in L^p \quad \text{and} \quad \mathfrak{m}(2) \vee \mathfrak{m}(p) < 1,$$

then $\mathscr{S}$ is a self-map of $\mathscr{P}_0^p(\mathbb{R})$ with a unique geometrically ℓ_p-attracting fixed point G_0 in this set.

Theorem 6.10 *Let $p > 2$. If*

$$C,\ \sum_{i\ge 1} |T_i| \in L^p, \quad \mathfrak{m}(2) \vee \mathfrak{m}(p) < 1 \quad \text{and} \quad \left| \mathbb{E}\left(\sum_{i\ge 1} T_i\right)\right| < 1,$$

then $\mathscr{S}$ exists in L^p-sense and has a unique geometrically ℓ_p-attracting fixed point G_0 in $(\mathscr{P}^p(\mathbb{R}), \ell_p)$.

Proof (of Theorem 6.9). Here we will proceed in a different way than before and prove that $\mathscr{S}$ is locally contractive on $(\mathscr{P}_0^p(\mathbb{R}), \ell_p)$ in the sense of Theorem 8.3 [see (32) below]. In particular, we will not make use of the Contraction Lemma 6.8. The first step is to show the result for integral $p > 2$ (as in [40]).

So let $2 < p \in \mathbb{N}$. We prove by induction that, for each $q \in \{1, ..., p\}$, there exists $\rho_q \in (0,1)$ such that

$$\ell_q^q(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; c_q\, \rho_q^n \tag{32}$$

for all $F,G \in \mathscr{P}_0^p(\mathbb{R})$, $n \ge 1$ and a suitable $c_q \in \mathbb{R}_>$ which may depend on F,G. Observe that this corresponds to (42) when choosing $F = \mathscr{S}(G)$.

Hereafter, $K \in \mathbb{R}_>$ shall denote a generic constant which may differ from line to line but does not depend on n. Recall from above that $\mathfrak{m}(2) \vee \mathfrak{m}(p) < 1$ entails $\mathfrak{m}(q) < 1$ for all $q \in [2,p]$.

If $q = 1$ or $= 2$, we may invoke Lemma 6.7 to find

$$\ell_1^2(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; \ell_2^2(\mathscr{S}^n(F), \mathscr{S}^n(G)) \;\le\; \mathfrak{m}(2)^n\, \ell_2^2(F,G)$$

for all $n \geq 1$ and $F,G \in \mathscr{P}_0^2(\mathbb{R})$, which clearly shows (32) in this case. We further see that $\mathscr{S}$ forms a contraction on $(\mathscr{P}_0^2(\mathbb{R}),\ell_2)$ and hence possesses a unique fixed point G_0 in this space. Since $\mathscr{P}_0^2(\mathbb{R}) \supset \mathscr{P}_0^p(\mathbb{R})$, it follows that G_0 is also the unique fixed point in $\mathscr{P}^p(\mathbb{R})$ once (32) has been verified for $q = p$.

For the inductive step suppose that (32) holds for any $r \in \{1,...,q-1\}$ and let $(U_i)_{i\geq 1}$ be a sequence of iid $Unif(0,1)$ random variables which are further independent of (C,T). Fixing any $F,G \in \mathscr{P}_0^q(\mathbb{R})$ throughout the rest of the proof, put

$$Y_{n,i} := \mathscr{S}^n(F)^{-1}(U_i) - \mathscr{S}^n(G)^{-1}(U_i), \quad n \geq 1$$

and note that $\|Y_{n,i}\|_r = \ell_r(\mathscr{S}^n(F),\mathscr{S}^n(G))$ for all $i \geq 1, n \geq 0$ and $r \in [1,q]$. Since

$$\ell_q^q(\mathscr{S}^{n+1}(F),\mathscr{S}^{n+1}(G)) \leq \mathbb{E}\left|\sum_{i\geq 1} T_i Y_{n,i}\right|^q \leq \lim_{m\to\infty} \mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|\right)^q$$

we will further estimate the last expectation for arbitrary $m \in \mathbb{N}$ by making use of the multinomial formula which provides us with

$$\mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|\right)^q = \mathbb{E}\left(\sum_{i=1}^m |T_i Y_{n,i}|^q\right) + \mathbb{E}\left(\sum_{\substack{0\leq r_1,...,r_m<q,\\ r_1+...+r_m=q}} \frac{q!}{r_1!\cdot\ldots\cdot r_m!}\prod_{j=1}^m |T_j Y_{n,j}|^{r_j}\right).$$

The first term on the right-hand side obviously equals $\mathfrak{m}(q)\,\ell_q^q(\mathscr{S}^n(F),\mathscr{S}^n(G))$, while the second may be further computed as follows by conditioning upon T and using the fact that the $Y_{n,i}$ for any fixed n are iid:

$$\begin{aligned}
&\mathbb{E}\left(\sum_{\substack{0\leq r_1,...,r_m<q,\\ r_1+...+r_m=q}} \frac{q!}{r_1!\cdot\ldots\cdot r_m!}\prod_{j=1}^m |T_j Y_{n,j}|^{r_j}\right)\\
&\quad = \mathbb{E}\left(\sum_{\substack{0\leq r_1,...,r_m<q,\\ r_1+...+r_m=q}} \frac{q!\,\mathbb{E}|Y_{n,1}|^{r_1}\cdot\ldots\cdot\mathbb{E}|Y_{n,1}|^{r_m}}{r_1!\cdot\ldots\cdot r_m!}\prod_{j=1}^m |T_j|^{r_j}\right)\\
&\quad = \left(\prod_{j=1}^m \ell_{r_j}^{r_j}(\mathscr{S}^n(F),\mathscr{S}^n(G))\right)\mathbb{E}\left(\sum_{\substack{0\leq r_1,...,r_m<q,\\ r_1+...+r_m=q}} \frac{q!}{r_1!\cdot\ldots\cdot r_m!}\prod_{j=1}^m |T_j|^{r_j}\right)\\
&\quad \leq K\rho^n\,\mathbb{E}\left(\sum_{i=1}^m |T_i|\right)^q
\end{aligned}$$

where the inductive hypothesis has been utilized to give the last estimate with $\rho := \max_{1\leq s\leq q-1}\rho_s$. The reader should notice that the constant K is not only independent of n but of m as well. Hence, by taking the limit $m \to \infty$, we find that

$$\ell_q^q(\mathscr{S}^{n+1}(F),\mathscr{S}^{n+1}(G)) \ \le\ \mathfrak{m}(q)\,\ell_q^q(\mathscr{S}^n(F),\mathscr{S}^n(G))+K\,\rho^n$$

for all $n\ge 0$ and thereupon

$$\begin{aligned}\ell_q^q(\mathscr{S}^{n+1}(F),\mathscr{S}^{n+1}(G)) \ &\le\ \mathfrak{m}(q)^{n+1}\,\ell_q^q(F,G)+K\sum_{k=1}^n \rho^k\mathfrak{m}(q)^{n-k}\\ &\le\ \Big(\ell_q^q(F,G)+Kn\Big)\,(\mathfrak{m}(q)\vee\rho)^{n+1}\end{aligned}$$

for all $n\ge 0$ which implies (32) for any $\rho_q\in(\mathfrak{m}(q)\vee\rho,1)$. By an appeal to Theorem 8.3, we conclude that, for any $F\in\mathscr{P}_0^p(\mathbb{R})$, $\mathscr{S}^n(F)$ converges to a fixed point in this set which must be unique by what has been stated above.

We turn to the second step which aims at an extension of the assertion to general $p>2$ with integer part $\widehat{p}$, say. Let $r\in\mathbb{N}$ be such that $2^r<p\le 2^{r+1}$ and $s:=p/2^{r+1}\in(0,1]$. From the first part of the proof, we know that (32) holds true for every $q\in\{1,...,\widehat{p}\}$, and since $\ell_\alpha(\cdot,\cdot)$ is nondecreasing in α, this readily extends to all $q\in[1,\widehat{p}]$. We will show hereafter that (32) is also true for $q=p$ (and thus for all $q\in[1,p]$) which finally proves the theorem in full generality.

Let us introduce the following operator Δ and its k-fold iterations Δ^k: For any nonnegative random variable W define

$$\Delta W:=(W-\mathbb{E}W)^2,\quad \Delta^2 W=\Big((W-\mathbb{E}W)^2-\mathbb{V}\mathrm{ar}\,W\Big)^2,\quad \text{etc.}$$

and $\Delta^0 W:=W$. Naturally, $\Delta W=\infty$ is stipulated if $\mathbb{E}W=\infty$. We note that

$$\mathbb{E}\Delta^k W\ \le\ \mathbb{E}(\Delta^{k-1}W)^2\ \le\ 2\,\mathbb{E}(\Delta^{k-2}W)^4\ \le ...\le\ 2^{k-1}\mathbb{E}W^{2^k} \tag{33}$$

holds true for any $k\ge 1$.

By repeated use of the Burkholder inequality (49) (in the by now familiar manner after conditioning on T) and the subadditivity of $x\mapsto x^\alpha$ for $x\ge 0$ and $0<\alpha\le 1$, we now obtain

$$\begin{aligned}&\Bigg\|\sum_{i\ge 1}T_iY_{n,i}\Bigg\|_p\ \le\ K\,\Bigg\|\sum_{i\ge 1}T_i^2Y_{n,i}^2\Bigg\|_{p/2}^{1/2}\\ &\le\ K\left(\Bigg\|\sum_{i\ge 1}T_i^2(Y_{n,i}^2-\mathbb{E}Y_{n,i}^2)\Bigg\|_{p/2}^{1/2}+\big(\mathbb{E}Y_{n,1}^2\big)^{1/2}\Bigg\|\sum_{i\ge 1}T_i^2\Bigg\|_{p/2}^{1/2}\right)\\ &\le\ K\left(\Bigg\|\sum_{i\ge 1}T_i^4\Delta Y_{n,i}^2\Bigg\|_{p/4}^{1/4}+\big(\mathbb{E}Y_{n,1}^2\big)^{1/2}\Bigg\|\sum_{i\ge 1}|T_i|\Bigg\|_p^{1/2}\right)\\ &\ \vdots\end{aligned}$$

$$\begin{aligned}
&\le K\left(\left\|\sum_{i\ge1}T_i^{2^{r+1}}\Delta^r Y_{n,i}^2\right\|_s^{1/2^{r+1}}+\sum_{j=0}^{r-1}\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}}\left\|\sum_{i\ge1}|T_i|\right\|_p^{1/2}\right)\\
&\le K\left(\left\|\sum_{i\ge1}|T_i|^p\left(\Delta^r Y_{n,i}^2\right)^s\right\|_1^{1/2^{r+1}}+\sum_{j=0}^{r-1}\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}}\left\|\sum_{i\ge1}|T_i|\right\|_p^{1/2}\right)\\
&\le K\left(\left\|\Delta^r Y_{n,1}^2\right\|_s^{1/2^{r+1}}\left\|\sum_{i\ge1}|T_i|^p\right\|_1^{1/2^{r+1}}+\sum_{j=0}^{r-1}\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}}\left\|\sum_{i\ge1}|T_i|\right\|_p^{1/2}\right)
\end{aligned}$$

for all $n\ge1$. Use (33), the definition of $Y_{n,1}$, and (32) for $\widehat{p}$ to infer

$$\begin{aligned}
\left(\mathbb{E}\Delta^j Y_{n,1}^2\right)^{1/2^{j+1}} &\le \left(2^{j-1}\,\mathbb{E}Y_{n,1}^{2^{j+1}}\right)^{1/2^{j+1}} \le 2\,\|Y_{n,1}\|_{2^{j+1}}\\
&\le 2\,\|Y_{n,1}\|_{\widehat{p}} = 2\,\ell_{\widehat{p}}(\mathscr{S}^n(F),\mathscr{S}^n(G)) \le 2\,c_{\widehat{p}}\,\rho_{\widehat{p}}^n
\end{aligned}$$

for any $j\in\{0,\dots,r-1\}$ and $n\ge0$. By combining this with $\|\sum_{i\ge1}|T_i|^p\|_1=\mathfrak{m}(p)<1$, the above estimation finally provides us with

$$\ell_p(\mathscr{S}^{n+1}(F),\mathscr{S}^{n+1}(G)) \le \left\|\sum_{i\ge1}T_iY_{n,i}\right\|_p \le K\rho^{n+1}$$

for all $n\ge0$ and a suitable $\rho\in(0,1)$. □

Proof (of Theorem 6.10). We are now in a more comfortable situation because the bulk of work has already been carried out in the previous proof. First note that all assumptions of Theorem 6.5 with $p=2$ are fulfilled which allows us to infer the existence of a unique fixed point $G_0\in\mathscr{P}^2(\mathbb{R})$. By Lemma 5.1(a), its mean value equals $c:=\mathbb{E}G_0=(1-\beta)^{-1}\mathbb{E}C$ with $\beta:=\mathbb{E}(\sum_{i\ge1}T_i)$. One can easily check that, if $F\in\mathscr{P}_c^p(\mathbb{R})$, then $\mathbb{E}S^n(F)=c$ for all $n\ge0$ and that this further implies $\mathscr{S}^n(F)^c=\mathscr{S}^n(F^c)$ (recall that $F^c=F^0(\cdot-c)$) and thereupon

$$\begin{aligned}
\ell_p(\mathscr{S}^{n+1}(F^c),\mathscr{S}^n(F^c)) &= \ell_p(\mathscr{S}^{n+1}(F)^c,\mathscr{S}^n(F)^c)\\
&= \ell_p(\mathscr{S}^{n+1}(F)^0,\mathscr{S}^n(F)^0)
\end{aligned}\tag{34}$$

for all $F\in\mathscr{P}^p(\mathbb{R})$ and $n\ge0$.

Now fix any $F\in\mathscr{P}^p(\mathbb{R})$, define $Y_{n,i}$ as in the previous proof, but for the pair $(\mathscr{S}(F^c),F^c)$. Then (32) for $q=p$ can be shown as in the previous proof, giving

$$\ell_p^p(\mathscr{S}^{n+1}(F^c),\mathscr{S}^n(F^c)) \le \left\|\sum_{i\ge1}T_iY_{n,i}\right\|_p^p \le c_p\rho_p^n$$

for all $n\ge0$ and suitable constants $c_p\in\mathbb{R}_>$ and $\rho_p\in(0,1)$. Note further that

$$\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^{n}(F) \;=\; \beta^n \left(\mathbb{E}\mathscr{S}(F) - \mathbb{E}F\right)$$

for all $n \geq 0$, as has been shown in the proof of Lemma 5.1 [see (19)]. By combining these facts with (10) and (34), we finally obtain

$$\begin{aligned}
&\ell_p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) \\
&\quad\leq\; \ell_p(\mathscr{S}^{n+1}(F)^0, \mathscr{S}^n(F)^0) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\
&\quad=\; \ell_p(\mathscr{S}^{n+1}(F)^0, \mathscr{S}^n(F)^0) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\
&\quad=\; \ell_p(\mathscr{S}^{n+1}(F^c), \mathscr{S}^n(F^c)) + \left|\mathbb{E}\mathscr{S}^{n+1}(F) - \mathbb{E}\mathscr{S}^n(F)\right| \\
&\quad\leq\; c_p^{1/p}\rho_p^{n/p} + \beta^n \left|\mathbb{E}\mathscr{S}(F) - \mathbb{E}F\right|
\end{aligned}$$

for all $n \geq 0$, that is geometric contraction of every iteration sequence in $\mathscr{P}^p(\mathbb{R})$. By invoking Theorem 8.3, we conclude that G_0 is the unique geometrically ℓ_p-attracting fixed point in this set. $\square$

6.4 Contraction on ℓ_p-neighborhoods of fixed distributions

A somewhat different approach than before is taken by Rüschendorf [42] who provides conditions for contraction of $\mathscr{S}$ in ℓ_p-neighborhoods of a fixed distribution $F \in \mathscr{P}(\mathbb{R})$, namely

$$\mathscr{U}^p(F) \;:=\; \left\{G \in \mathscr{P}(\mathbb{R}) : \ell_p(F,G) < \infty\right\}$$

for $p > 0$, and

$$\mathscr{U}_c^p(F) \;:=\; \left\{G \in \mathscr{P}_c^1(\mathbb{R}) : \ell_p(F,G) < \infty\right\}$$

for $p \geq 1$ and $c \in \mathbb{R}$. He embarks on the observation that, for $\ell_p(F,G)$ to be finite, it only takes to find an (F,G)-coupling (X,Y) such that $X - Y \in L^p$ but *not* that X,Y are themselves in L^p. Of course, if $F \in \mathscr{P}^p(\mathbb{R})$, then $\mathscr{U}^p(F) = \mathscr{P}^p(\mathbb{R})$. Besides the contraction condition $\mathscr{C}_p(T) = \mathfrak{m}(p) < 1$, familiar from previous results, he requires a *bounded jump-size condition*, namely

$$\ell_p(F, \mathscr{S}(F)) < \infty, \tag{35}$$

which is quite common in the study of iterated function systems on complete separable metric spaces. In that context, F is an arbitrary reference point and $\mathscr{S}$ a generic copy of the iid random Lipschitz functions to be iterated, see e.g. [19, Thm. 3]. Here the condition serves to ensure that $\mathscr{S}$ is a self-map of $\mathscr{U}^p(F)$ as the following proposition shows.

Proposition 6.11 *Let $p > 0$ and $F \in \mathscr{P}(\mathbb{R})$ be such that* (35) *holds true. Then $\mathscr{S}$ defines a self-map of $\mathscr{U}^p(F)$. Moreover, if $F \in \mathscr{P}^1(\mathbb{R})$, $C \in L^1$ and $p \geq 1$, then $\mathscr{S}$ defines a self-map of $\mathscr{U}_c^p(F)$ for any c such that $c = c\,\mathbb{E}(\sum_{i\geq 1} T_i) + \mathbb{E}C$, thus for* *all $c \in \mathbb{R}$ if $\kappa := \mathbb{E}(\sum_{i\geq 1} T_i) = 1$ and $\mathbb{E}C = 0$, and for $c = (1-\mathbb{E}(\sum_{i\geq 1} T_i))^{-1}\mathbb{E}C$ if $\kappa \neq 1$.*

Proof. The following choices of random variables may take to enlarge the underlying probability space. Let (X,Y) be a $(F,\mathscr{S}(F))$-coupling such that $\ell_p(F,\mathscr{S}(F)) = \|X-Y\|_p$. Then pick iid copies $X_1, X_2, \ldots$ of X which are further independent of (C,T) and put $Y' := \sum_{i\geq 1} T_i X_i + C$. Finally, let X' be such that the conditional law of X' given $Y' = y$ is the same as the conditional law of X given $Y = y$ for all $y \in \mathbb{R}$, thus (X',Y') is a copy of (X,Y). Now, if $G \in \mathscr{U}^p(\mathbb{R})$, we can choose the X_i along with iid Z_i, independent of (C,T) and with common distribution G, such that the (X_i, Z_i) are iid as well and $\mu := \|X_i - Z_i\|_p < \infty$. It follows that

$$\begin{aligned} \ell_p(F,\mathscr{S}(G)) &\leq \ell_p(F,\mathscr{S}(F)) + \ell_p(\mathscr{S}(F),\mathscr{S}(G)) \\ &\leq \ell_p(F,\mathscr{S}(F)) + \left\| \sum_{i\geq 1} T_i(X_i - Z_i) \right\|_p \\ &= \ell_p(F,\mathscr{S}(F)) + \left\| \sum_{i\geq 1} T_i \right\|_p \mu < \infty \end{aligned}$$

and therefore that $\mathscr{S}$ is a self-map of $\mathscr{U}^p(F)$. The second assertion follows in a similar manner. □

The following results, containing those for $0 < p \leq 2$ and N a fixed integer stated in [42], are the “local” counterparts of Theorems 6.1 and 6.4 – 6.6 and proved in the same way once having observed that Contraction Lemma 6.2 remains valid for $F, G \in \mathscr{U}_c^p(F_0)$ with $F_0 \in \mathscr{P}(\mathbb{R})$, and the Contraction Lemmata 6.7, and 6.8 remain valid for $F, G \in \mathscr{U}_c^p(F_0)$ with $F_0 \in \mathscr{P}^1(\mathbb{R})$ and $c \in \mathbb{R}$ (see also [42, Lemma 2.1]). We therefore refrain from giving proofs again.

Theorem 6.12 *If $0 < p \leq 1$ and $\mathfrak{m}(p) < 1$, and if $F \in \mathscr{P}(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a contraction on $(\mathscr{U}^p(F), \ell_p)$ with a unique geometrically attracting fixed point.*

Theorem 6.13 *If $p > 1$, $C \in L_0^1$ and $\mathscr{C}_p(T) < 1$, and if $F \in \mathscr{P}^1(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a quasi-contraction on $(\mathscr{U}_0^p(F), \ell_p)$ with a unique geometrically attracting fixed point.*

Theorem 6.14 *If $p > 1$, $C \in L^1$, $\sum_{i\geq 1} T_i \in L^p$, $\mathscr{C}_p(T) < 1$ and $|\mathbb{E}(\sum_{i\geq 1} T_i)| < 1$, and if $F \in \mathscr{P}^1(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a quasi-contraction on $(\mathscr{U}^p(F), \ell_p)$ with a unique geometrically attracting fixed point.*

Theorem 6.15 *If $p > 1$, $C \in L_0^1$, $\sum_{i\geq 1} T_i \in L^p$, $\mathscr{C}_p(T) < 1$ and $\mathbb{E}(\sum_{i\geq 1} T_i) = 1$, and if $F \in \mathscr{P}^1(\mathbb{R})$ satisfies (35), then $\mathscr{S}$ is a quasi-contraction on $(\mathscr{U}_c^p(F), \ell_p)$ with a unique geometrically attracting fixed point for any $c \in \mathbb{R}$.*

If $1 < p \leq 2$, then $\mathscr{C}_p(T) = \mathfrak{m}(p)$ should be recalled. Moreover, if N is a fixed integer, then $\sum_{i\geq 1} T_i = \sum_{i=1}^N T_i \in L^p$ follows from $\mathfrak{m}(p) < 1$. With these observations, one can readily check that the results in [42] are really contained in the ones stated before.

Validity of the bounded jump-size condition (35) is usually difficult to check. In fact, it trivially holds whenever F is fixed point of $\mathscr{S}$. Since, furthermore, $\mathscr{U}^p(F) = \mathscr{U}^p(G)$ for all $G \in \mathscr{U}^p$ as well as $\mathscr{U}_c^p(F) = \mathscr{U}_c^p(G)$ for all $G \in \mathscr{U}_c^p$, the previous results may also be interpreted as follows: Under the respective conditions on p and (C,T), condition (35) holds true for some F only if F is in finite ℓ^p-distance to a fixed point of $\mathscr{S}$. In contrast to the results from the previous subsections, this fixed point and thus F do not need to be elements of L^p.

Let us finally note that Rüschendorf, as an interesting consequence of his results, provides conditions which entail a certain one-to-one correspondence between the fixed points of a nonhomogeneous smoothing transform $\mathscr{S}$ and its homogeneous counterpart $\mathscr{S}_0$ (same T, but $C = 0$), see [42, Thm. 3.1] for details.

6.5 Contraction on subsets of $\mathscr{P}^p(\mathbb{R})$ with specified integral moments $(p > 1)$

Let $p = m + \alpha > 1$ hereafter, where $m \in \mathbb{N}$ and $\alpha \in (0,1]$, and assume that $\mathscr{S}$ exists in L^p-sense so that, by Corollary 4.2, $C, \sum_{i\ge 1} T_i \in L^p$. This final subsection is devoted to situations when $\mathscr{S}$, while not necessarily an ℓ_p-(quasi-)contraction on $\mathscr{P}^p(\mathbb{R})$, turns out to be contractive with respect to the Zolotarev metric ζ_p on subsets with specified integral moments. Recall that $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for $\mathbf{z} = (z_1, ..., z_m) \in \mathbb{R}^m$ equals the set of distributions $F \in \mathscr{P}^p(\mathbb{R})$ such that $\int x^k F(dx) = z_k$ for $k = 1, ..., m$.

In order for $\mathscr{S}$ to be a self-map of $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$, we must have that, given any iid $X_1, X_2, ...$ with law in $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$,

$$\begin{aligned} z_k &= \mathbb{E}\left(\sum_{i\ge 1} T_i X_i + C\right)^k \\ &= \sum_{j_0+j_1+...=k} \frac{k!}{\prod_{i\ge 0} j_i!} \left(\prod_{i\ge 1} z_{j_i}\right) \mathbb{E}\left(C^{j_0} \prod_{i\ge 1} T_i^{j_i}\right) \\ &= z_k\, \mathbb{E}\left(\sum_{i\ge 1} T_i^k\right) + \mathbb{E}C^k + \sum_{\substack{j_0+j_1+...=k \\ j_0\vee j_1\vee ...<k}} \frac{k!}{\prod_{i\ge 0} j_i!} \left(\prod_{i\ge 1} z_{j_i}\right) \mathbb{E}\left(C^{j_0} \prod_{i\ge 1} T_i^{j_i}\right) \end{aligned}$$

for $k = 1, ..., m$, because $X_1, X_2, ...$ and (C,T) are independent. In other words, $\mathbf{z}$ must satisfy a – for $m \ge 2$ nonlinear – system of equations, and one can easily see that this system may have a unique solution as well as infinitely many.

Theorem 6.16 *Suppose that $\mathfrak{m}(p) < 1$ and that $\mathscr{S}$ exists in L^p-sense. Then $\mathscr{S}$ is a ζ_p-contraction on $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ for any $\mathbf{z} \in \mathbb{R}^m$ such that $\mathscr{S}$ is a self-map of $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$. In particular, it has a unique geometrically ζ_p-attracting fixed-point in this set.*

Proof. Since $(\mathscr{P}^p_{\mathbf{z}}(\mathbb{R}),\zeta_p)$ is a complete metric space (see Proposition 3.4), the result follows directly wityh the help of the Contraction Lemma 6.17 below and Banach's fixed-point theorem. □

Lemma 6.17 *Let $(C,T)=(C,T_1,T_2,\ldots)$, $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be independent sequences of real-valued random variables in L^p such that*

(A1) $X_1,X_2,\ldots$ *are independent with* $\mathscr{L}(X_n)=F_n$ *for* $n\geq 1$.
(A2) $Y_1,Y_2,\ldots$ *are independent with* $\mathscr{L}(Y_n)=G_n$ *for* $n\geq 1$.
(A3) *For each* $n\geq 1$, $F_n,G_n\in\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$ *for some* $\mathbf{z}\in\mathbb{R}^m$.
(A4) $\sum_{i\geq 1}T_iX_i+C,\sum_{i\geq 1}T_iY_i+C\in L^p$.

Then

$$\zeta_s\left(\sum_{i\geq 1}T_iX_i+C,\sum_{i\geq 1}T_iY_i+C\right)\ \leq\ \sum_{i\geq 1}\mathbb{E}|T_i|^p\,\zeta_s(F_i,G_i). \tag{36}$$

In particular, if $\mathbf{z}\in\mathbb{R}^m$*, then*

$$\zeta_p(\mathscr{S}(F),\mathscr{S}(G))\ \leq\ \mathfrak{m}(p)\,\zeta_p(F,G). \tag{37}$$

for all $F,G\in\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$*, whenever* $\mathscr{S}$*, the smoothing transform associated with* (C,T)*, exists in* L^p*-sense and is a self-map of* $\mathscr{P}^p_{\mathbf{z}}(\mathbb{R})$.

Proof. First note that $\zeta_p(\sum_{i\geq 1}T_iX_i+C,\sum_{i\geq 1}T_iY_i+C)<\infty$ because (A3) and (A4) ensure that $\sum_{i\geq 1}T_iX_i+C,\sum_{i\geq 1}T_iY_i+C\in L^p_{\mathbf{z}}$. Denote by Λ the distribution of (C,T) and let $t=(t_1,t_2,\ldots)$ in the subsequent integration with respect to Λ. Then, by multiple use of properties (14) and (15) of ζ_p (in lines 5, 8 and 9), we infer for each $n\in\mathbb{N}$ that

$$\begin{aligned}
&\zeta_p\left(\sum_{i=1}^n T_iX_i+C,\sum_{i=1}^n T_iY_i+C\right)\\
&=\ \sup_{f\in\mathfrak{F}_p}\left|\mathbb{E}\left(f\left(\sum_{i=1}^n T_iX_i+C\right)-f\left(\sum_{i=1}^n T_iY_i+C\right)\right)\right|\\
&\leq\ \int\sup_{f\in\mathfrak{F}_p}\left|\mathbb{E}\left(f\left(\sum_{i=1}^n t_iX_i+c\right)-f\left(\sum_{i=1}^n t_iY_i+c\right)\right)\right|\ \Lambda(dc,dt)\\
&=\ \int\zeta_p\left(\sum_{i=1}^n t_iX_i+c,\sum_{i=1}^n t_iY_i+c\right)\ \Lambda(dc,dt)\\
&\leq\ \int\zeta_p\left(\sum_{i=1}^n t_iX_i,\sum_{i=1}^n t_iY_i\right)\ \Lambda(dc,dt)\\
&\leq\ \int\sum_{k=1}^n\zeta_p\left(\sum_{i=k}^n t_iX_i+\sum_{j=1}^{k-1}t_jY_j,\ \sum_{i=k+1}^n t_iX_i+\sum_{j=1}^{k}t_jY_j\right)\ \Lambda(dc,dt)\\
&=\ \int\sum_{k=1}^n\zeta_p\left(t_kX_k+S_k,t_kY_k+S_k\right)\ \Lambda(dc,dt)
\end{aligned}$$

$$\left[\text{where } S_k := \sum_{i=k+1}^{n} t_iX_i + \sum_{j=1}^{k-1} t_jY_j \text{ and is independent of } X_k, Y_k\right]$$
$$\le \int \sum_{k=1}^{n} \zeta_p\left(t_kX_k, t_kY_k\right) \Lambda(dc,dt)$$
$$= \int \sum_{k=1}^{n} |t_k|^p \zeta_p\left(X_k, Y_k\right) \Lambda(dc,dt)$$
$$= \sum_{i=1}^{n} \mathbb{E}|T_i|^p \zeta_p(F_i, G_i)$$

which proves (36) by letting n tend to infinity and using

$$\lim_{n\to\infty} \zeta_p\left(\sum_{i=1}^{n} T_iX_i + C, \sum_{i=1}^{n} T_iY_i + C\right) = \zeta_p\left(\sum_{i\ge1} T_iX_i + C, \sum_{i\ge1} T_iY_i + C\right).$$

The second inequality (37) follows from the first one when choosing $F_i = F$ and $G_i = G$ for all $i \ge 1$. $\square$

7 Concluding remarks

Having provided a comprehensive account of results describing the contractive behavior of the smoothing transform on $\mathscr{P}^p(\mathbb{R})$ or subsets thereof for $p > 0$, we would like to finish this review with some remarks on what has not been covered.

Naturally, other metrics than ℓ_p and ζ_p could have been studied as well. For instance, with $\widehat{F}(t) := \int e^{itx} F(dx)$ denoting the Fourier transform of F, the Fourier metric

$$r_p(F,G) := \int_0^\infty \frac{|\widehat{F}(t) - \widehat{G}(t)|}{t^{1+p}}\, dt, \quad F,G \in \mathscr{P}_c^p(\mathbb{R})$$

for $p \in (1,2)$ was introduced and shown to be complete on $\mathscr{P}_c^p(\mathbb{R})$ by Baringhaus & Grübel [8, Lemma 2.1]. For homogeneous $\mathscr{S}$ with a.s. finite N, they further showed that it is a contraction on $(\mathscr{P}_c^p(\mathbb{R}), r_p)$ if $\mathfrak{m}(p) < 1$ and $\mathbb{E}(\sum_{i\ge1} T_i) = 1$. The result was later extended by Iksanov [26, Prop. 6] to the case of general N (see also [8, Section]. As one can easily see, the result further extends to the nonhomogeneous case with $C \in L_0^1$.

Since contraction (with respect to ℓ_p or ζ_p) on subsets Γ of $\mathscr{P}^p(\mathbb{R})$ for some $p > 0$ particularly entails that, for some fixed point of $\mathscr{S}$, the set Γ is attracting with respect to weak convergence, one may ask about more general results describing such sets without moment assumptions, thus within $\mathscr{P}(\mathbb{R})$. As an example in this direction, we mention the following result obtained by Durrett & Liggett [18, Thm. 2(b)]: If $C = 0$, $T \ge 0$, N is a.s. bounded and T has characteristic exponent $\alpha \in (0,1]$ (see at the end of Section 2), then, given any fixed point $F \in \mathscr{P}(\mathbb{R}_\ge)$ of $\mathscr{S}$ with

Laplace transform $\widetilde{F}$, $\mathscr{S}^n(G)$ converges weakly to F whenever

$$\lim_{t\downarrow 0}\frac{1-\widetilde{F}(t)}{1-\widetilde{G}(t)} \;=\; 1.$$

An extension of their result under relaxed conditions on N appears in [31, Thm. 1.3]. Results of this type could also be formulated for the general smoothing transform and fixed points on the whole real line when substituting Fourier transforms for Laplace transforms. However, we refrain from supplying any further details.

8 Appendix

8.1 Banach's fixed-point theorem

Let $f:\mathbb{X}\to\mathbb{X}$ be a continuous self-map of a metric space $(\mathbb{X},\rho)$ and denote by $f^n=f\circ\ldots\circ f$ (n-times) its n-fold composition for $n\ge 1$. If there exists an initial value $x_0\in\mathbb{X}$ such that the sequence $x_n:=f(x_{n-1})=f^n(x_0)$, $n\ge 1$, converges to some $x_\infty\in\mathbb{X}$, then the continuity of f implies that x_∞ is a fixed point of f, for

$$x_\infty \;=\; \lim_{n\to\infty}x_n \;=\; f\left(\lim_{n\to\infty}x_{n-1}\right) \;=\; f(x_\infty). \tag{38}$$

The map f is called a *contraction* or more specifically α*-contraction* if there exists $\alpha\in[0,1)$ such that

$$\rho(f(x),f(y))\le\alpha\,\rho(x,y) \tag{39}$$

for all $x,y\in\mathbb{X}$. If (39) holds true when replacing f with f^n for some $n\ge 2$, then f is called *quasi-contraction* or α*-quasi-contraction*.

Under a contraction, the distance between two iteration sequences $(f^n(x))_{n\ge 1}$ and $(f^n(y))_{n\ge 1}$ is therefore decreasing geometrically fast, viz.

$$\rho(f^n(x),f^n(y)) \;\le\; \alpha^n\,\rho(x,y)$$

for all $n\ge 1$. If the space $(\mathbb{X},\rho)$ is complete, then this entails geometric convergence to a unique fixed point of f as the following classic result shows.

Theorem 8.1 [Banach's fixed-point theorem] *Every contraction $f:\mathbb{X}\to\mathbb{X}$ on a complete metric space $(\mathbb{X},\rho)$ possesses a unique fixed point $\xi\in\mathbb{X}$. Moreover,*

$$\rho(\xi,f^n(x)) \;\le\; \frac{\alpha^n}{1-\alpha}\,\rho(f(x),x) \tag{40}$$

holds true for all $x\in\mathbb{X}$ and $n\ge 1$, where α denotes the contraction parameter of f.

The next result shows that Banach's fixed-point theorem essentially remains valid for quasi-contractions.

Theorem 8.2 [Banach's fixed-point theorem for quasi-contractions] *Every quasi-contraction $f:\mathbb{X}\to\mathbb{X}$ on a complete metric space $(\mathbb{X},\rho)$ possesses a unique fixed point $\xi\in\mathbb{X}$, and*

$$\rho(\xi,f^n(x))\ \le\ \frac{\alpha^n}{1-\alpha}\max_{0\le r<m}\rho(f^{m+r}(x),f^m(x)) \tag{41}$$

for some $m\ge 1$, $\alpha\in[0,1)$ and all $x\in\mathbb{X}$, $n\ge 1$.

Proof. Pick m,α such that f^m forms an α-contraction on $(\mathbb{X},\rho)$ with unique fixed point ξ. Writing $n\in\mathbb{N}$ in the form $km+r$ with unique $k\in\mathbb{N}_0$ and $r\in\{0,...,m-1\}$, we infer with the help of (40)

$$\rho(\xi,f^n(x))\ \le\ \max_{0\le j<m}\rho(\xi,f^{km+j}(x))\ \le\ \frac{\alpha}{1-\alpha}\max_{0\le j<m}\rho(f^{m+j}(x),f^j(x))$$

and thus (41), in particular $\rho(\xi,f^n(x))\to 0$. Since f is continuous, the latter implies that ξ is also the (necessarily unique) fixed point of f. $\square$

Replacing the global by a local contraction property along an iteration sequence, existence of a fixed point still follows, but it needs no longer be unique.

Theorem 8.3 *Let $(\mathbb{X},\rho)$ be a complete metric space and $f:\mathbb{X}\to\mathbb{X}$ an arbitrary self-map. Suppose there exist $x_0\in\mathbb{X}$ and constants $c\ge 0$ and $\alpha\in[0,1)$ such that*

$$\rho(f^{n+1}(x_0),f^n(x_0))\ \le\ c\alpha^n \tag{42}$$

for all $n\ge 1$. Then $\xi=\lim_{n\to\infty}f^n(x_0)$ exists and it is a fixed point of f if the map is continuous. Moreover,

$$\rho(\xi,f^n(x_0))\ \le\ \frac{c\alpha^n}{1-\alpha} \tag{43}$$

for all $n\ge 1$.

Proof. Putting $x_n:=f^n(x_0)$ and using (42), we obtain

$$\rho(x_{m+n},x_m)\ \le\ \sum_{k=m}^{m+n-1}\rho(x_{k+1},x_k)\ \le\ \sum_{k=m}^{m+n-1}c\alpha^k\ \le\ \frac{c\alpha^m}{1-\alpha}$$

for all $m,n\ge 1$, that is, $(x_n)_{n\ge 0}$ is a Cauchy sequence in $\mathbb{X}$ and thus convergent to some $\xi\in\mathbb{X}$, for $(\mathbb{X},\rho)$ is complete. If f is continuous, then $f(\xi)=\xi$ (see (38)). Finally, (43) follows from (42) when observing that

$$\rho(\xi,f^n(x_0))\ =\ \rho(\xi,x_n)\ \le\ \sum_{k\ge n}\rho(x_k,x_{k+1}). \qquad\square$$

8.2 Convex function inequalities for martingales and their maxima

Let $(M_n)_{n\ge 0}$ be a martingale with natural filtration $(\mathscr{F}_n)_{n\ge 0}$ and increments $D_n = M_n - M_{n-1}$ for $n \ge 1$. In the following, we list some powerful martingale inequalities that provide bounds for the ϕ-*moments* $\mathbb{E}\phi(M_n)$, when $\phi : \mathbb{R} \to \mathbb{R}_\ge$ denotes an even convex function with $\phi(0) = 0$ and some additional properties. This includes the standard class $\phi(x) = |x|^p$ for $p \ge 1$. Setting $M_\infty := \liminf_{n\to\infty} M_n$, all provided upper bounds remain valid for $n = \infty$ when observing that Fatou's lemma implies

$$\mathbb{E}\phi(M_\infty) \;\le\; \liminf_{n\to\infty} \mathbb{E}\phi(M_n).$$

We begin with the class of ϕ that have a concave derivative in $\mathbb{R}_>$ and thus encompasses $\phi(x) = |x|^p$ for $1 \le p \le 2$. The subsequent result is cited from [7] and an improvement (with regard to the appearing constant) of a version due to Topchiĭ and Vatutin [43]. In the more general framework of Banach spaces of a given type, the inequality (with a non-specified constant) is actually due to Woyczynski [45, Prop. 2.1].

Theorem 8.4 [Topchiĭ-Vatutin inequality] *Let $\phi : \mathbb{R} \to \mathbb{R}_\ge$ be an even convex function with concave derivative on $\mathbb{R}_>$ and $\phi(0) = 0$. Then*

$$\mathbb{E}\phi(M_n) - \mathbb{E}\phi(M_0) \;\le\; c \sum_{k=1}^{n} \mathbb{E}\phi(D_k), \tag{44}$$

for all $n \in \overline{\mathbb{N}}_0$ and $c = 2$. The constant may be chosen as $c = 1$ if $(M_n)_{n\ge 0}$ is nonnegative or has a.s. symmetric conditional increment distributions, and the same holds generally true, if $\phi(x) = |x|$ or $\phi(x) = x^2$, in the last case even with equality sign in (44).

We continue with two famous convex function inequalities by Burkholder, Davis and Gundy [13] which are valid for a much larger class of convex functions ϕ.

Theorem 8.5 [Burkholder-Davis-Gundy inequalities] *Let $\phi : \mathbb{R} \to \mathbb{R}_\ge$ be an even convex function satisfying $\phi(0) = 0$ and $\phi(2t) \le \gamma\phi(t)$ for all $t \ge 0$ and some $\gamma > 0$. Put $E_n(\phi) := \mathbb{E}\left(\max_{0\le k\le n} \phi(M_k)\right)$. Then*

$$a_\gamma \mathbb{E}\phi\left(\left(\sum_{k=1}^{n} D_k^2\right)^{1/2}\right) \;\le\; E(\phi) \;\le\; b_\gamma \mathbb{E}\phi\left(\left(\sum_{k=1}^{n} D_k^2\right)^{1/2}\right) \tag{45}$$

and

$$E_n(\phi) \;\le\; c_\gamma \left[\mathbb{E}\phi\left(\left(\sum_{k=1}^{n} \mathbb{E}\big(D_k^2|\mathscr{F}_{k-1}\big)\right)^{1/2}\right) + \mathbb{E}\left(\max_{0\le k\le n} \phi(D_k)\right)\right] \tag{46}$$

for all $n \in \overline{\mathbb{N}}_0$ and constants $a_\gamma, b_\gamma, c_\gamma \in \mathbb{R}_>$ depending only on γ. The last inequality actually remains valid if, ceteris paribus, ϕ is merely nondecreasing instead of convex on $\mathbb{R}_\geq$.

Of special importance in connection with the smoothing transform is the case when M_n is a weighted sum of iid zero-mean random variables and $\phi(x) = |x|^p$ for some $p > 0$. We therefore note:

Corollary 8.6 *If $\phi(x) = |x|^p$ (thus $\gamma = 2^p$) for some $p > 0$ and $M_n = \sum_{k=1}^n t_k X_k$ for $t_1, t_2, \ldots \in \mathbb{R}$ and iid $X_1, X_2, \ldots \in L_0^p$, then* (46) *takes the form*

$$E_n(\phi) \ \le \ c_p \left[\|X_1\|_2^p \left(\sum_{k=1}^n t_k^2 \right)^{p/2} + \mathbb{E}\left(\max_{1\le k\le n} |t_k X_k|^p \right) \right], \tag{47}$$

for all $n \in \overline{\mathbb{N}}_0$ and a constant c_p only depending on p, giving in particular

$$\mathbb{E}|M_n|^p \ \le \ c_p \left[\|X_1\|_2^p \left(\sum_{k=1}^n t_k^2 \right)^{p/2} + \|X_1\|_p^p \sum_{k=1}^n |t_k|^p \right]. \tag{48}$$

Finally, we state the classical L^p-inequality by Burkholder [12], valid for $p > 1$ only. The case $p = 1$ is different but will not be considered here.

Theorem 8.7 [Burkholder inequality] *Let $p > 1$. Then*

$$a_p \left\| \left(\sum_{k=1}^n D_k^2 \right)^{1/2} \right\|_p \ \le \ \|M_n\|_p \ \le \ b_p \left\| \left(\sum_{k=1}^n D_k^2 \right)^{1/2} \right\|_p \tag{49}$$

for $n \in \overline{\mathbb{N}}_0$ and constants $a_p, B_p \in \mathbb{R}_>$ only depending on p. Admissible choices are $a_p = (18p^{3/2}/(p-1))^{-1}$ and $b_p = 18p^{3/2}/(p-1)^{1/2}$ (see [24, Thm. 2.10]).

Acknowledgements G.A. was supported by Deutsche Forschungsgemeinschaft (SFB 878).

References

1. Aldous, D., Steele, J.M.: The objective method: probabilistic combinatorial optimization and local weak convergence. Probability on Discrete Structures **110**, 1–72 (2004)
2. Alsmeyer, G., Biggins, J.D., Meiners, M.: The functional equation of the smoothing transform. To appear in Ann. Probab. (2012+). `www.arxiv.org`:0906.3133v2
3. Alsmeyer, G., Damek, E., Mentemeier, S.: Precise tail index of fixed points of the two-sided smoothing transform. This volume (2012)
4. Alsmeyer, G., Iksanov, A., Rösler, U.: On distributional properties of perpetuities. J. Theoret. Probab. **22**(3), 666–682 (2009)

5. Alsmeyer, G., Meiners, M.: Fixed points of inhomogeneous smoothing transforms. J. Diff. Equations Appl. **18**, 1287–1304 (2012)
6. Alsmeyer, G., Meiners, M.: Fixed points of the smoothing transform: two-sided solutions. To appear in Probab. Theory Related Fields (2012+). `www.arXiv.org`: 1009.2412v1
7. Alsmeyer, G., Rösler, U.: The best constant in the Topchii-Vatutin inequality for martingales. Statist. Probab. Lett. **65**(3), 199–206 (2003)
8. Baringhaus, L., Grübel, R.: On a class of characterization problems for random convex combinations. Ann. Inst. Statist. Math. **49**(3), 555–567 (1997)
9. Biggins, J.D.: Martingale convergence in the branching random walk. J. Appl. Probab. **14**(1), 25–37 (1977)
10. Biggins, J.D., Kyprianou, A.E.: Seneta-Heyde norming in the branching random walk. Ann. Probab. **25**(1), 337–360 (1997)
11. Biggins, J.D., Kyprianou, A.E.: Fixed points of the smoothing transform: the boundary case. Electron. J. Probab. **10**, 609–631 (electronic) (2005)
12. Burkholder, D.L.: Martingale transforms. Ann. Math. Statist. **37**, 1494–1504 (1966)
13. Burkholder, D.L., Davis, B.J., Gundy, R.F.: Integral inequalities for convex functions of operators on martingales. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory, pp. 223–240. Univ. California Press, Berkeley, Calif. (1972)
14. Caliebe, A.: Symmetric fixed points of a smoothing transformation. Adv. Appl. Probab. **35**(2), 377–394 (2003)
15. Caliebe, A., Rösler, U.: Fixed points with finite variance of a smoothing transformation. Stochastic Process. Appl. **107**(1), 105–129 (2003)
16. Collamore, J.F., Vidyashankar, A.N.: Large deviation tail estimates and related limit laws for stochastic fixed point equations. This volume (2012)
17. Diaconis, P., Freedman, D.: Iterated random functions. SIAM Rev. **41**(1), 45–76 (1999)
18. Durrett, R., Liggett, T.M.: Fixed points of the smoothing transformation. Z. Wahrsch. Verw. Gebiete **64**(3), 275–301 (1983)
19. Elton, J.H.: A multiplicative ergodic theorem for Lipschitz maps. Stochastic Process. Appl. **34**(1), 39–47 (1990)
20. Goldie, C.M., Maller, R.A.: Stability of perpetuities. Ann. Probab. **28**(3), 1195–1218 (2000)
21. Graf, S., Mauldin, R.D., Williams, S.C.: The exact Hausdorff dimension in random recursive constructions. Mem. Amer. Math. Soc. **71**(381), x+121 (1988)
22. Grübel, R., Rösler, U.: Asymptotic distribution theory for Hoare's selection algorithm. Adv. in Appl. Probab. **28**(1), 252–269 (1996)
23. Guivarc'h, Y.: Sur une extension de la notion de loi semi-stable. Ann. Inst. H. Poincaré Probab. Statist. **26**(2), 261–285 (1990)
24. Hall, P., Heyde, C.C.: Martingale limit theory and its application. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York (1980). Probability and Mathematical Statistics
25. Hu, Y., Shi, Z.: Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. Ann. Probab. **37**(2), 742–789 (2009)
26. Iksanov, A.M.: Elementary fixed points of the BRW smoothing transforms with infinite number of summands. Stochastic Process. Appl. **114**(1), 27–50 (2004)
27. Jelenković, P.R., Olvera-Cravioto, M.: Implicit renewal theory and power tails on trees. To appear in Adv. Appl. Probab. (2010). `www.arxiv.org`: 1006.3295v3
28. Jelenković, P.R., Olvera-Cravioto, M.: Information ranking and power laws on trees. Adv. in Appl. Probab. **42**(4), 1057–1093 (2010)
29. Jelenković, P.R., Olvera-Cravioto, M.: Implicit renewal theorem for trees with general weights. To appear in Stoch. Proc. Appl. (2012). `www.arxiv.org`:1012.2165v2
30. Liu, Q.: The growth of an entire characteristic function and the tail probabilities of the limit of a tree martingale. In: Trees (Versailles, 1995), *Progr. Probab.*, vol. 40, pp. 51–80. Birkhäuser, Basel (1996)
31. Liu, Q.: Fixed points of a generalized smoothing transformation and applications to the branching random walk. Adv. Appl. Probab. **30**(1), 85–112 (1998)

32. Liu, Q.: On generalized multiplicative cascades. Stochastic Process. Appl. **86**(2), 263–286 (2000)
33. Mauldin, R.D., Williams, S.C.: Random recursive constructions: asymptotic geometric and topological properties. Trans. Amer. Math. Soc. **295**(1), 325–346 (1986)
34. Mirek, M.: Heavy tail phenomenon and convergence to stable laws for iterated Lipschitz maps. Probab. Theory Related Fields **151**(3-4), 705–734 (2011)
35. Neininger, R., Rüschendorf, L.: A general limit theorem for recursive algorithms and combinatorial structures. Ann. Appl. Probab. **14**(1), 378–418 (2004)
36. Neininger, R., Rüschendorf, L.: Analysis of algorithms by the contraction method: additive and max-recursive sequences. In: Interacting stochastic systems, pp. 435–450. Springer, Berlin (2005)
37. Penrose, M.D., Wade, A.R.: On the total length of the random minimal directed spanning tree. Adv. in Appl. Probab. **38**(2), 336–372 (2006)
38. Rachev, S.T., Rüschendorf, L.: Probability metrics and recursive algorithms. Adv. Appl. Probab. **27**(3), 770–799 (1995)
39. Rösler, U.: A limit theorem for "Quicksort". RAIRO Inform. Théor. Appl. **25**(1), 85–100 (1991)
40. Rösler, U.: A fixed point theorem for distributions. Stochastic Process. Appl. **42**(2), 195–214 (1992)
41. Rösler, U., Rüschendorf, L.: The contraction method for recursive algorithms. Algorithmica **29**(1-2), 3–33 (2001). Average-case analysis of algorithms (Princeton, NJ, 1998)
42. Rüschendorf, L.: On stochastic recursive equations of sum and max type. J. Appl. Probab. **43**(3), 687–703 (2006)
43. Vatutin, V.A., Topchiĭ, V.A.: The maximum of critical Galton-Watson processes, and left-continuous random walks. Theory Probab. Appl. **42**(1), 17–27 (1998)
44. Vervaat, W.: On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. Adv. Appl. Probab. **11**(4), 750–783 (1979)
45. Woyczyński, W.A.: On Marcinkiewicz-Zygmund laws of large numbers in Banach spaces and related rates of convergence. Probab. Math. Statist. **1**(2), 117–131 (1981) (1980)
46. Zolotarev, V.M.: Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. Theory Probab. Appl. **21**(4), 721–737 (1976)