PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 249–269
doi:10.19195/0208-4147.38.2.1

BOUNDARY VALUE PROBLEMS FOR THE DUNKL LAPLACIAN

BY

MOHAMED BEN CHROUDA (MONASTIR),
KHALIFA EL MABROUK (HAMMAM SOUSSE), AND KODS HASSINE (MONASTIR)

Abstract. Let Δ_k be the Dunkl Laplacian on $\mathbb{R}^d$ associated with a reflection group W and a multiplicity function k. The purpose of this paper is to establish the existence and the uniqueness of a positive solution on the unit ball B of $\mathbb{R}^d$ to the following boundary value problem:

$$\Delta_k u = \varphi(u) \text{ in } B \quad \text{and} \quad u = f \text{ on } \partial B.$$

We distinguish two cases of nonnegative perturbation φ: trivial and nontrivial.

2010 AMS Mathematics Subject Classification: Primary: 31B05, 60J45; Secondary: 35J08, 35J61.

Key words and phrases: Dunkl Laplacian, Dirichlet problems, Green operators, semilinear equations.

1. INTRODUCTION

The Dunkl Laplacian is the sum of a second order differential operator and a difference term associated with a multiplicity function k and a reflection group W. An important motivation to study the Dunkl Laplacian rises from its relevance for the analysis of certain exactly solvable models of mechanics, namely the Calogero–Moser–Sutherland type (see [5], [13], [19]). Since its introduction by C. F. Dunkl in [6], the analysis of Dunkl theory has been the subject of many articles and it has deep and fruitful interactions with various mathematics fields, namely Fourier analysis and special functions [15], [28], [29], algebra (double affine Hecke algebras [17]) and probability theory (Feller processes with jumps [11], [4]). The Dunkl Laplacian generates a positive strongly continuous contraction semigroup [25]. This fact gives rise to a Hunt process, called a Dunkl process, and so to a corresponding family of harmonic kernels $(H_V)_V$. If the multiplicity function k is identically zero, then the operator Δ_k reduces to the classical Laplace operator Δ, and so the Dunkl process is the Brownian motion and $H_V(x, \cdot)$ is the classical harmonic measure relative to V and x. If k is not trivial, then paths of the Dunkl pro-

cess are discontinuous (see [11]), and thus it follows from the general theory of balayage spaces [1] that Δ_k generates a balyage space and not a harmonic space. This yields that for every bounded open set V and every $x \in V$ the harmonic measure $H_V(x,\cdot)$ is not necessarily supported by the Euclidean boundary ∂V of V, as in the classical setting $k = 0$, but it may live on the entire complement $V^c := \mathbb{R}^d \backslash V$.

Throughout this paper we assume that k is strictly positive. Our first purpose is to show that, for every bounded open subset V of $\mathbb{R}^d$ and every $x \in V$, the harmonic measure $H_V(x,\cdot)$ is supported by a compact set of V^c and not by the whole V^c. In the particular case where V is invariant under the reflection group W (e.g. V is an open ball of $\mathbb{R}^d$ centered at the origin), we shall prove that the support of $H_V(x,\cdot)$ is contained in ∂V. This fact allows us to investigate, for an open ball B of center zero, the boundary value problem

$$\begin{cases} \Delta_k u = \varphi(u) & \text{in } B, \\ u = f & \text{on } \partial B, \end{cases} \tag{1.1}$$

where f is a nonnegative continuous function on ∂B. We impose that $\varphi : [0,\infty[\to [0,\infty[$ is nondecreasing, continuous and satisfies $\varphi(0) = 0$. Our main goal is to establish the existence and the uniqueness of a positive solution to problem (1.1). We distinguish two cases of perturbation φ (trivial and nontrivial). In the first step, we consider $\varphi = 0$ and we prove that the function $H_B f$ defined on B by

$$H_B f(x) = \int\limits_{\partial B} f(y) H_B(x, dy)$$

is the unique continuous extension u of f on $\overline{B}$ satisfying $\Delta_k u = 0$ in B. That is, $H_B f$ is the unique solution of (1.1) for $\varphi = 0$. Assuming that φ is not trivial, we show that u satisfies (1.1) if and only if

$$u + G_B^k\big(\varphi(u)\big) = H_B f,$$

where G_B^k is the Green operator on B. Then, by a compactness argument of G_B^k, we prove that the map $u \mapsto H_B f - G_B^k\big(\varphi(u)\big)$ admits one and only one fixed point $u \in C(\overline{B})$, and so u is the unique solution of problem (1.1).

2. NOTATION AND PRELIMINARIES

For every subset F of $\mathbb{R}^d$, let $\mathcal{B}(F)$ be the set of all Borel-measurable functions on F and let 1_F be the indicator function of F. Let $C(F)$ be the set of all continuous real-valued functions on F, $C^n(F)$ be the class of all functions that are n times continuously differentiable on F, and $C_0(F)$ be the set of all continuous functions on F such that $u = 0$ on ∂F, which means that $\lim_{x \to z} u(x) = 0$ for all $z \in \partial F$ and $\lim_{x \to \infty} u(x) = 0$ if F is unbounded. We denote by $\mathcal{C}_c^\infty(F)$ the set of all infinitely differentiable functions on F with compact support. If $\mathcal{G}$ is a set

of numerical functions, then $\mathcal{G}^+$ (respectively $\mathcal{G}_b$) will denote the class of all functions in $\mathcal{G}$ which are nonnegative (respectively bounded). The uniform norm will be denoted by $\|\cdot\|$.

For every $\alpha \in \mathbb{R}^d \setminus \{0\}$, let H_α be the hyperplane orthogonal to α and let σ_α be the reflection in H_α, i.e.,

$$\sigma_\alpha(x) := x - 2\frac{\langle \alpha, x\rangle}{|\alpha|^2}\alpha,$$

where $\langle\cdot,\cdot\rangle$ denotes the usual inner product on $\mathbb{R}^d$ and $|\cdot|$ is the associated norm. A finite subset R of $\mathbb{R}^d \setminus \{0\}$ is called a *root system* if $R \cap \mathbb{R}\cdot\alpha = \{\pm\alpha\}$ and $\sigma_\alpha(R) = R$ for all $\alpha \in R$. For a given root system R, the reflection $\sigma_\alpha,\ \alpha \in R$, generates a finite group W called a *reflection group* associated with R. A function $k : R \to \mathbb{R}_+$ is called a *multiplicity function* if it satisfies $k(\sigma_\alpha\beta) = k(\beta)$ for every $\alpha, \beta \in R$. Throughout this paper we fix a root system R and a multiplicity function k. We consider the differential-difference operators $T_i,\ 1 \leqslant i \leqslant d$, defined in [7] for every $u \in C^1(\mathbb{R}^d)$ by

$$T_i u(x) = \frac{\partial u}{\partial x_i}(x) + \frac{1}{2}\sum_{\alpha\in R} k(\alpha)\alpha_i \frac{u(x) - u(\sigma_\alpha x)}{\langle \alpha, x\rangle},$$

and called *Dunkl operators* in the literature. The Dunkl Laplacian Δ_k is the sum of squares of Dunkl operators:

$$\Delta_k := \sum_{i=1}^d T_i^2.$$

It is given explicitly, for $u \in C^2(\mathbb{R}^d)$, by

$$\Delta_k u(x) = \Delta u(x) + \sum_{\alpha\in R} k(\alpha)\left(\frac{\langle \nabla u(x), \alpha\rangle}{\langle \alpha, x\rangle} - \frac{|\alpha|^2}{2}\frac{u(x) - u\big(\sigma_\alpha(x)\big)}{\langle \alpha, x\rangle^2}\right). \tag{2.1}$$

Likewise the classical Laplace operator Δ, the Dunkl Laplacian has the following symmetry property: For $u \in C^2(\mathbb{R}^d)$ and $v \in C_c^2(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \Delta_k u(x)v(x)w_k(x)\,dx = \int_{\mathbb{R}^d} u(x)\Delta_k v(x)w_k(x)\,dx, \tag{2.2}$$

where w_k is the homogeneous weight function defined on $\mathbb{R}^d$ by

$$w_k(x) = \prod_{\alpha\in R} |\langle x, \alpha\rangle|^{k(\alpha)}.$$

A fundamental result in Dunkl theory is the existence of an intertwining operator $V_k : C^\infty(\mathbb{R}^d) \to C^\infty(\mathbb{R}^d)$ between the classical Laplacian Δ and Dunkl Laplacian,

i.e., $\Delta_k V_k = V_k \Delta$. We refer to [8], [26], [28] for more details about the intertwining operator. By means of V_k, there exists a counterpart of the usual exponential function, called a *Dunkl kernel* $E_k(\cdot,\cdot)$, which is defined for every $y \in \mathbb{C}^d$ and $x \in \mathbb{R}^d$ by

$$E_k(x,y) = V_k(e^{\langle \cdot, y\rangle})(x).$$

It is clear from (2.1) that if k vanishes identically, then the Dunkl Laplacian reduces to the classical Laplacian Δ. In this case the intertwining operator V_k is the identity operator, and so E_k reduces to the classical exponential function. Notice that E_k is symmetric and positive on $\mathbb{R}^d \times \mathbb{R}^d$ and satisfies $E_k(\lambda y, x) = E_k(y, \lambda x) =$ for every $\lambda \in \mathbb{C}$.

In all this paper we assume that

$$m := d + \sum_{\alpha \in R} k(\alpha) > 2.$$

Let p_t^k be the Dunkl heat kernel, introduced in [25], defined for every $t > 0$ and every $x, y \in \mathbb{R}^d$ by

$$p_t^k(x,y) = \frac{c_k^2}{2^m} \int_{\mathbb{R}^d} e^{-t|\xi|^2} E_k(-ix,\xi) E_k(iy,\xi) w_k(\xi) d\xi, \tag{2.3}$$

where

$$c_k = \Big(\int_{\mathbb{R}^d} e^{-|y|^2} w_k(y)\, dy \Big)^{-1}.$$

For every $x, y \in \mathbb{R}^d$, $p_t^k(x,y) > 0$, $p_t^k(x,y) = p_t^k(y,x)$ and

$$p_t^k(x,y) \leqslant \frac{c_k}{(4t)^{m/2}} \exp\left(- \frac{(|x| - |y|)^2}{4t} \right). \tag{2.4}$$

Also, for every $x \in \mathbb{R}^d$, the function $(t,y) \mapsto p_t^k(x,y)$ solves the generalized heat equation $\partial_t u - \Delta_k u = 0$ on $]0,\infty[\times\mathbb{R}^d$. More precisely, the following holds:

$$\frac{\partial}{\partial t} p_t^k(x,y) = \Delta_k\big(p_t^k(\cdot,y)\big)(x) = \Delta_k\big(p_t^k(x,\cdot)\big)(y). \tag{2.5}$$

For every $f \in C_0(\mathbb{R}^d)$ and $t > 0$ let

$$P_t^k f(x) = \int_{\mathbb{R}^d} p_t^k(x,y) f(y) w_k(y)\, dy, \quad x \in \mathbb{R}^d.$$

Then $(P_t^k)_{t>0}$ forms a positive strongly continuous contraction semigroup on $C_0(\mathbb{R}^d)$ of generator Δ_k. This fact yields the existence of a Hunt process (X_t, P^x) (see [2], Theorem I.9.4), called the Dunkl process, with state space $\mathbb{R}^d$ and transition kernel

$$P_t^k(x, dy) = p_t^k(x,y) w_k(y)\, dy.$$

3. HARMONIC KERNELS

For every bounded open subset D of $\mathbb{R}^d$, we denote by τ_D the first exit time from D by (X_t), i.e.,

$$\tau_D = \inf\{t > 0; X_t \notin D\}.$$

LEMMA 3.1. *Let D be a bounded open set. Then, for every $x \in D$,*

$$P^x\,(0 < \tau_D < \infty) = 1.$$

P r o o f. Let $x \in D$. Since the Dunkl process has right continuous paths, we immediately conclude that $P^x(0 < \tau_D) = 1$. Let $r > 0$ be such that $D \subset B_r$, the ball of center zero and radius r. Clearly,

$$\begin{aligned} E^x[\tau_D] \leqslant E^x[\tau_{B_r}] &= E^x\big[\int_0^{\tau_{B_r}} 1_{B_r}(X_t)\,dt\big] \\ &\leqslant \int_0^\infty E^x[1_{B_r}(X_t)]dt = \int_0^\infty \int_{B_r} p_t^k(x,y)w_k(y)\,dy\,dt. \end{aligned}$$

So, to prove that $P^x(\tau_D < \infty) = 1$, it will be sufficient to show that

$$\int_0^\infty \int_{B_r} p_t^k(x,y)w_k(y)\,dy\,dt < \infty.$$

Using spherical coordinates and applying the fact that the function w_k is homogeneous of degree $m - d$, we infer from the integral representation (2.3) of p_t^k that, for every $y \in \mathbb{R}^d$,

$$p_t^k(x,y) = \frac{c_k^2}{2^m} \int_0^\infty \int_{S^{d-1}} e^{-ts^2} E_k(-ix, s\xi)E_k(iy, s\xi)w_k(\xi)s^{m-1}\sigma(d\xi)ds,$$

where σ denotes the surface area measure on the unit sphere S^{d-1} of $\mathbb{R}^d$. Therefore,

$$\int_0^\infty p_t^k(x,y)dt = \frac{c_k^2}{2^m} \int_0^\infty \int_{S^{d-1}} E_k(-ix, s\xi)E_k(iy, s\xi)w_k(\xi)s^{m-3}\sigma(d\xi)ds.$$

Using again spherical coordinates and then applying Fubini's theorem, we get

$$\begin{aligned} \int_0^\infty \int_{B_r} p_t^k(x,y)w_k(y)\,dy\,dt &= \int_0^r \int_{S^{d-1}} \big(\int_0^\infty p_t^k(x,uy)dt\big)w_k(y)u^{m-1}\sigma(dy)du \\ &= \frac{c_k^2}{2^m} \int_0^r \int_0^\infty \int_{S^{d-1}} \big(\int_{S^{d-1}} E_k(iuy, s\xi)w_k(y)\sigma(dy)\big) \\ &\quad \times E_k(-ix, s\xi)w_k(\xi)s^{m-3}u^{m-1}\sigma(d\xi)dsdu. \end{aligned}$$

On the other hand, we recall from [27] that

$$\int_{S^{d-1}} E_k(iz,y)w_k(y)\sigma(dy) = 2^{m/2}c_k^{-1}\frac{J_{m/2-1}(|z|)}{|z|^{m/2-1}},$$

where $J_{m/2-1}$ is the Bessel function of index $m/2-1$ given by

$$J_{m/2-1}(z) := \left(\frac{z}{2}\right)^{m/2-1}\sum_{n=0}^{\infty}\frac{(-1)^n z^{2n}}{4^n n!\Gamma(n+m/2)}.$$

Hence

$$\int_0^{\infty}\int_{B_r} p_t^k(x,y)w_k(y)\,dy\,dt = \int_0^r \frac{u^{m-1}}{(u|x|)^{m/2-1}}\Big(\int_0^{\infty} J_{m/2-1}(s|x|)J_{m/2-1}(us)s^{-1}ds\Big)du,$$

and so

$$\int_0^{\infty}\int_{B_r} p_t^k(x,y)w_k(y)\,dy\,dt = \frac{1}{m-2}\int_0^r u^{m-1}\big(\max(u,|x|)\big)^{2-m}\,du = \frac{1}{m-2}\left(\frac{|x|^2}{m}+\frac{r^2-|x|^2}{2}\right). \tag{3.1}$$

To get (3.1), one should use a formula from [21], p. 100. ∎

For every bounded open set D, we define

$${}^{W}D := \bigcup_{w\in W} w(D) \quad \text{and} \quad \Gamma_D := \overline{{}^{W}D}\setminus D.$$

That is, ${}^{W}D$ is the smallest open bounded set containing D which is invariant under the reflection group W. In the following theorem, we show that if the process starts from $x\in D$ then, at the first exit time from D, it should be in the compact Γ_D.

THEOREM 3.1. *Let D be a bounded open subset of $\mathbb{R}^d$. Then, for every $x\in D$,*

$$P^x\left(X_{\tau_D}\in\Gamma_D\right)=1. \tag{3.2}$$

In particular, if D is W-invariant, i.e., ${}^{W}D=D$, then $\Gamma_D=\partial D$, and therefore

$$P^x\left(X_{\tau_D}\in\partial D\right)=1.$$

Proof. Let $x\in D$ and consider the function F defined for every $y,z\in\mathbb{R}^d$ by $F(y,z)=0$ if $z\in\{\sigma_\alpha y;\alpha\in R\}$ and $F(y,z)=1$ otherwise. Let

$$Y_t := \sum_{s<t} 1_{\{X_{s^-}\neq X_s\}}F(X_{s^-},X_s),\quad t>0.$$

It follows from Proposition 3.2 in [11] that for every $t > 0$, $P^x(Y_t = 0) = 1$, and consequently

$$P^x\big(1_{\{X_{s^-} \neq X_s\}} F(X_{s^-}, X_s) = 0;\ \forall s > 0\big) = 1.$$

Then, since $P^x(0 < \tau_D < \infty) = 1$, we deduce that

$$P^x\big(1_{\{X_{\tau_D^-} \neq X_{\tau_D}\}} F(X_{\tau_D^-}, X_{\tau_D}) = 0\big) = 1.$$

On the other hand, since $X_{\tau_D^-} \in \overline{D}$ on $\{0 < \tau_D < \infty\}$, we have

$$\{X_{\tau_D} \notin \Gamma_D, 0 < \tau_D < \infty\} \subset \{1_{\{X_{\tau_D^-} \neq X_{\tau_D}\}} F(X_{\tau_D^-}, X_{\tau_D}) = 1\}.$$

This completes the proof. ∎

For every bounded open set D and every $x \in \mathbb{R}^d$, let $H_D(x, \cdot)$ be the harmonic measure relative to x and D, i.e., for every Borel set A,

$$H_D(x, A) := P^x(X_{\tau_D} \in A).$$

For every $f \in \mathcal{B}_b(\mathbb{R}^d)$, let $H_D f$ be the function defined on $\mathbb{R}^d$ by

$$H_D f(x) = \int f(y) H_D(x, dy).$$

Since, for $x \in D$, the harmonic measure $H_D(x, \cdot)$ is supported by the compact set Γ_D, it will be convenient to put again

$$H_D f(x) = \int f(y) H_D(x, dy), \quad x \in D, \tag{3.3}$$

for every $f \in \mathcal{B}_b(\Gamma_D)$.

Let ${}^*\mathcal{H}^+(\mathbb{R}^d)$ denote the set of all nonnegative lower semicontinuous functions f on $\mathbb{R}^d$ such that

$$H_D f \leqslant f \quad \text{for every bounded open set } D.$$

Because $(\mathbb{R}^d, P^x)$ is a Hunt process, it follows from Theorem IV.8.1 in [1] that $\big(\mathbb{R}^d, {}^*\mathcal{H}^+(\mathbb{R}^d)\big)$ is a balayage space. Hence, it follows from the general theory of balayage spaces that for every $f \in \mathcal{B}_b(\Gamma_D)$

$$H_D f \in C(D) \tag{3.4}$$

and

$$H_V H_D f = H_D f \quad \text{on } V \text{ for all open sets } V \text{ such that } \overline{V} \subset D. \tag{3.5}$$

Furthermore, a function $f \in \mathcal{B}^+(\mathbb{R}^d)$ belongs to ${}^*\mathcal{H}^+(\mathbb{R}^d)$ if and only if

$$\sup_{t>0} P_t^k f = f.$$

Let us now introduce the Green function G^k of the Dunkl Laplacian which will play an important role in our approach. It is defined for every $x, y \in \mathbb{R}^d$ by

$$G^k(x,y) = \int_0^\infty p_t^k(x,y)dt.$$

For every $y \in \mathbb{R}^d$, the function $G_y^k := G^k(\cdot, y) \in {}^*\mathcal{H}^+(\mathbb{R}^d)$. Indeed, by the semigroup property,

$$P_t^k G_y^k(x) = \int_t^\infty p_s^k(x,y)ds \leqslant G^k(x,y).$$

This implies that the map $t \mapsto P_t^k G_y^k$ is decreasing on $]0,\infty[$, and so

$$\sup_{t>0} P_t^k G_y^k = \lim_{t\to 0} P_t^k G_y^k = G_y^k.$$

Hence $G_y^k \in {}^*\mathcal{H}^+(\mathbb{R}^d)$, which means that for every bounded open set D,

$$\int G^k(z,y) H_D(x,dz) \leqslant G^k(x,y). \tag{3.6}$$

Furthermore, it is obvious that G^k is positive and symmetric on $\mathbb{R}^d \times \mathbb{R}^d$. Therefore, it follows from Theorem VI.1.16 in [2] that for every bounded open set D and every $x, y \in \mathbb{R}^d$,

$$\int G^k(x,z) H_D(y,dz) = \int G^k(y,z) H_D(x,dz). \tag{3.7}$$

4. DIRICHLET PROBLEM

Let B be an open ball of $\mathbb{R}^d$ of center zero and radius $r > 0$. We first introduce the following three kinds of harmonicity on B:

A continuous function $h : B \to \mathbb{R}$ is said to be

(i) Δ_k-*harmonic* on B if $h \in C^2(B)$ and $\Delta_k h(x) = 0$ for every $x \in B$.

(ii) X-*harmonic* on B if $H_D h(x) = h(x)$ for every bounded open set D such that $\overline{D} \subset B$ and every $x \in D$.

(iii) Δ_k-*harmonic on B in the distributional sense* if

$$\langle h, \Delta_k \varphi \rangle_k := \int_B h(x) \Delta_k \varphi(x) w_k(x) dx = 0 \quad \text{for all } \varphi \in C_c^\infty(B).$$

LEMMA 4.1. *Let $f \in C_c^2(\mathbb{R}^d)$. For every $x \in \mathbb{R}^d$,*

$$\int_{\mathbb{R}^d} G^k(x,y)\Delta_k f(y) w_k(y)dy = -f(x). \tag{4.1}$$

In particular, for every bounded open set D and every $x \in D$,

$$H_D f(x) - f(x) = E^x\big[\int_0^{\tau_D} \Delta_k f(X_s)ds\big]. \tag{4.2}$$

P r o o f. Let $x \in \mathbb{R}^d$. Using Fubini's theorem and formulas (2.2) and (2.5), we have

$$\begin{aligned}\int_{\mathbb{R}^d} G^k(x,y)\Delta_k f(y)w_k(y)\,dy &= \int_0^\infty \int_{\mathbb{R}^d} p_t^k(x,y)\Delta_k f(y)w_k(y)\,dy\,dt\\ &= \int_0^\infty \int_{\mathbb{R}^d} \Delta_k\big(p_t^k(x,\cdot)\big)(y) f(y)w_k(y)\,dy\,dt\\ &= \int_0^\infty \int_{\mathbb{R}^d} \Delta_k\big(p_t^k(\cdot,y)\big)(x) f(y)w_k(y)\,dy\,dt\\ &= \lim_{t\to\infty} P_t^k f(x) - \lim_{t\to 0} P_t^k f(x) = -f(x).\end{aligned}$$

To get $\lim_{t\to\infty} P_t^k f(x) = 0$, we only use (2.4) and the fact that f has compact support. Formula (4.2) follows from (4.1) and the strong Markov property. In fact, let D be a bounded open set and let $x \in D$. Then

$$\begin{aligned}-f(x) &= \int G^k(x,y)\Delta_k f(y)w_k(y)dy = \int_0^\infty \int p_t^k(x,y)\Delta_k f(y)w_k(y)dydt\\ &= E^x\big[\int_0^\infty \Delta_k f(X_s)ds\big] = E^x\big[\int_0^{\tau_D} \Delta_k f(X_s)ds\big] + E^x\big[\int_{\tau_D}^\infty \Delta_k f(X_s)ds\big]\\ &= E^x\big[\int_0^{\tau_D} \Delta_k f(X_s)ds\big] + E^x\Big[E^{X_{\tau_D}}\big[\int_0^\infty \Delta_k f(X_s)ds\big]\Big]\\ &= E^x\big[\int_0^{\tau_D} \Delta_k f(X_s)ds\big] + E^x\left[-f(X_{\tau_D})\right] = E^x\big[\int_0^{\tau_D} \Delta_k f(X_s)ds\big] - H_D f(x). \quad \blacksquare\end{aligned}$$

LEMMA 4.2. *For every bounded open set D and for every $\varphi, \psi \in C_c^2(\mathbb{R}^d)$,*

$$\langle H_D\psi, \Delta_k\varphi\rangle_k = \langle \Delta_k\psi, H_D\varphi\rangle_k. \tag{4.3}$$

P r o o f. Applying formula (4.1) to ψ, we get

$$\langle H_D\psi, \Delta_k\varphi\rangle_k = -\int\int\int G^k(z,y)\Delta_k\psi(y)w_k(y)dy H_D(x,dz)\Delta_k\varphi(x)w_k(x)dx.$$

Then (4.3) is obtained by Fubini's theorem by using formula (3.7) and formula (4.1) applied to φ. $\blacksquare$

Now, we show that the three kinds of harmonicity on B introduced at the beginning of this section are equivalent.

THEOREM 4.1. *Let $h \in C(B)$. The following three assertions are equivalent*:
(i) *h is Δ_k-harmonic on B.*
(ii) *h is X-harmonic on B.*
(iii) *h is Δ_k-harmonic on B in the distributional sense.*

Proof. (i) Assume that h is Δ_k-harmonic on B. Let D be a bounded open set such that $\overline{D} \subset B$ and let $x \in D$. We claim that

$$H_D h(x) - h(x) = E^x\big[\int_0^{\tau_D} \Delta_k h(X_s)ds\big]. \tag{4.4}$$

Let V be a bounded open set such that $\overline{D} \subset V \subset \overline{V} \subset B$. By C^∞-Uryshon's lemma, there exists $\theta \in C_c^\infty(B)$ such that $\theta = 1$ on V. Let $f := h\theta$ and $\psi := h - f$. Obviously, $h = f$ on V, $\psi = 0$ on V and $f \in C_c^2(B)$. Then, using (4.2), we obtain

$$H_D h(x) - h(x) = E^x\big[\int_0^{\tau_D} \Delta_k f(X_s)ds\big] + H_D\psi(x). \tag{4.5}$$

For every $y \in \mathbb{R}^d$, let $N(y, dz)$ be the Lévy kernel of the Dunkl process X which is given in [11] by the following formula:

$$N(y, dz) = \sum_{\alpha\in\mathbb{R}_+,\langle y,\alpha\rangle\neq 0} \frac{k(\alpha)}{\langle\alpha, y\rangle^2}\delta_{\sigma_\alpha y}(dz). \tag{4.6}$$

Since $\psi = 0$ on V, it follows from Theorem 1 in [14] that

$$H_D\psi(x) = E^x\big[\int_0^{\tau_D}\int \psi(z)N(X_s, dz)ds\big]. \tag{4.7}$$

On the other hand, by (2.1) and (4.6) we easily see that for every $y \in D$,

$$\Delta_k f(y) = \Delta_k h(y) - \int \psi(z)N(y, dz). \tag{4.8}$$

Thus formula (4.4) is obtained by combining (4.5), (4.7) and (4.8). Hence, by (4.4), $H_D h(x) = h(x)$, and so h is X-harmonic on B.

(ii) Assume that h is X-harmonic on B. Let $\varphi \in C_c^\infty(B)$ and let $D \subset \overline{D} \subset B$ be a W-invariant bounded open set which contains the support of φ. Let $(h_n)_{n\geqslant 1} \subset C_c^2(B)$ be a sequence which converges uniformly to h on ∂D. Since $H_D\varphi = 0$ on D, applying (4.3), we obtain

$$\langle H_D h_n, \Delta_k\varphi\rangle_k = 0, \quad n \geqslant 1. \tag{4.9}$$

On the other hand,

$$\sup_{x\in D}|H_D h_n(x) - H_D h(x)| = \sup_{x\in D}\Big|\int_{\partial D}\big(h_n(y)-h(y)\big)H_D(x,dy)\Big| \leqslant \sup_{y\in\partial D}|h_n(y)-h(y)| \to 0 \quad \text{as } n\to\infty.$$

Hence, by letting n tend to infinity in (4.9), we get $\langle H_D h, \Delta_k\varphi\rangle_k = 0$, and therefore $\langle h, \Delta_k\varphi\rangle_k = 0$ since $h = H_D h$ on D.

(iii) Assume that h is Δ_k-harmonic on B in the distributional sense. The hypoellipticity of the Dunkl Laplacian Δ_k on W-invariant open sets [12], [22] yields $h\in C^\infty(B)$. Thus, by (2.2), it follows that, for every $\varphi\in C_c^\infty(B)$,

$$\int_B \Delta_k h(x)\varphi(x)w_k(x)\,dx = 0.$$

Hence $\Delta_k h(x) = 0$ for every $x\in B$, which means that h is Δ_k-harmonic on B. ∎

It is worth noting that the equivalence established in the above theorem remains valid if we replace the ball B by any W-invariant open set, for example, the whole space $\mathbb{R}^d$.

THEOREM 4.2. *For every $f\in C^+(\partial B)$, the problem*

$$\begin{cases}\Delta_k h = 0 & \text{on } B,\\ h = f & \text{on } \partial B\end{cases} \tag{4.10}$$

admits one and only one solution in $C^+(\overline{B})$ which is given by $H_B f$.

Proof. Let $f\in C^+(\partial B)$. By (3.4) and (3.5), the function $H_B f$ is continuous and X-harmonic on B. We shall show that $H_B f$ is a continuous extension of f on $\overline{B}$. Let $z\in\partial B$ and consider $V = \mathbb{R}^d\backslash\{0\}$, and let u be the function defined on V by

$$u(x) = G^k(x,0) - G^k(z,0).$$

Since

$$p_t^k(x,0) = \frac{c_k}{(4t)^{m/2}}e^{-|x|^2/(4t)}, \quad x\in\mathbb{R}^d,$$

it follows that

$$G^k(x,0) = \frac{c_k}{4}\frac{\Gamma(m/2-1)}{|x|^{m-2}}. \tag{4.11}$$

Then, using (3.6) and (4.11), it is easy to verify that u is a barrier of z (with respect to B), i.e.,

(i) u is hyperharmonic on $V\cap B$,

(ii) u is positive on $V \cap B$,

(iii) $\lim_{x\in V\cap B, x\to z} u(x) = 0$.

Hence, by Propositions VII.3.1 and VII.3.3 in [1], we obtain $H_B(z,\cdot) = \delta_z$ and $\lim_{x\in B, x\to z} H_B f(x) = f(z)$. Since z is arbitrary in ∂B, $H_B f$ is a continuous extension of f on $\overline{B}$. So, it remains to prove the uniqueness of the solution. Let h be another continuous extension of f on $\overline{B}$ which is the solution to the problem (4.10). Let $x \in B$ and let $(D_n)_{n\geqslant 1}$ be a sequence of nonempty bounded open sets such that $x \in D_n \subset \overline{D_n} \subset D_{n+1}$ and $B = \bigcup_n D_n$. Then $(\tau_{D_n})_n$ converges to τ_B almost surely. Hence, the continuity of h on $\overline{B}$ together with the quasi-left-continuity of the Dunkl process yield $H_B h(x) = \lim_n H_{D_n} h(x)$, and consequently $H_B h(x) = h(x)$, since $H_{D_n} h(x) = h(x)$ for every $n \geqslant 1$. Thus $h(x) - H_B f(x) = H_B(h-f)(x) = 0$ since $h = f$ on ∂B. So, $h = H_B f$ on B and the uniqueness is proved. ∎

5. GREEN OPERATORS

The Green operator G^k on the whole space $\mathbb{R}^d$ is defined, for every $f \in \mathcal{B}^+(\mathbb{R}^d)$, by the formula

$$G^k f(x) := \int_{\mathbb{R}^d} G^k(x,y) f(y) w_k(y)\, dy, \qquad x \in \mathbb{R}^d.$$

By Fatou's lemma, for each $y \in \mathbb{R}^d$, $G^k(\cdot, y)$ is lower semicontinuous on $\mathbb{R}^d$, and so $G^k f$ is lower semicontinuous on $\mathbb{R}^d$.

In the sequel, B_r denotes the ball of $\mathbb{R}^d$ of center zero and radius $r > 0$, and $A_{t,s}$ denotes the annulus of $\mathbb{R}^d$ of center zero and radius $0 < t < s < \infty$.

LEMMA 5.1. (i) *For every* $0 < r < \infty$,

$$G^k 1_{B_r}(x) = \begin{cases} \frac{1}{m-2}\left(\frac{|x|^2}{m} + \frac{r^2-|x|^2}{2}\right) & \text{if } |x| \leqslant r, \\ \frac{1}{m(m-2)} r^m |x|^{2-m} & \text{if } |x| \geqslant r. \end{cases} \tag{5.1}$$

(ii) *For every* $0 \leqslant t < s < \infty$,

$$0 \leqslant \sup_{x\in A_{t,s}} G^k 1_{A_{t,s}}(x) \leqslant \frac{2}{m-2} s(s-t). \tag{5.2}$$

P r o o f. Formula (5.1) follows immediately from (3.1) because

$$G^k 1_{B_r}(x) = \int_0^\infty \int_{B_r} p_t^k(x,y) w_k(y)\, dy\, dt.$$

Let $0 \leqslant t < s < \infty$. It is clear that $0 \leqslant G^k 1_{A_{t,s}}$ and that

$$G^k 1_{A_{t,s}} = G^k 1_{B_s} - G^k 1_{B_t}.$$

Then, by (5.1), it follows that for every $x \in A_{t,s}$,

$$\begin{aligned} G^k 1_{A_{t,s}}(x) &= \frac{1}{m-2}\left[\frac{|x|^2}{m} + \frac{s^2-|x|^2}{2}\right] - \frac{1}{m(m-2)} t^m |x|^{2-m} \\ &= \frac{1}{m-2}\left[\frac{|x|^2}{m}\left(1-\left(\frac{t}{|x|}\right)^m\right) + \frac{s^2-|x|^2}{2}\right] \\ &\leqslant \frac{1}{m-2}\left[\frac{s^2}{m}\left(1-\left(\frac{t}{s}\right)^m\right) + \frac{s^2-t^2}{2}\right] \\ &\leqslant \frac{1}{m-2}\left[s^2\left(1-\frac{t}{s}\right) + \frac{s^2-t^2}{2}\right] \\ &\leqslant \frac{2}{m-2} s(s-t). \quad \blacksquare \end{aligned}$$

An immediate consequence of the above lemma is that for each $x \in \mathbb{R}^d$ the function $G^k(\cdot, x) w_k$ is locally Lebesgue-integrable on $\mathbb{R}^d$. Thus, by Fubini's theorem, for every $f \in \mathcal{B}_b(\mathbb{R}^d)$ with compact support, we have

$$\begin{aligned} G^k f(x) = \int_{\mathbb{R}^d} G^k(x,y) f(y) w_k(y)\, dy &= \int_0^\infty \int_{\mathbb{R}^d} p_t^k(x,y) f(y) w_k(y)\, dy\, dt \\ &= \int_0^\infty E^x\left[f(X_t)\right] dt = E^x\Big[\int_0^\infty f(X_t)\, dt\Big]. \end{aligned}$$

PROPOSITION 5.1. *Let $f \in \mathcal{B}_b(\mathbb{R}^d)$ with compact support. Then $G^k f \in C_0(\mathbb{R}^d)$ and*

$$\Delta_k G^k f = -f \quad \text{in } \mathbb{R}^d \tag{5.3}$$

in the distributional sense, i.e., for every $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} G^k f(x) \Delta_k \psi(x) w_k(x)\, dx = -\int_{\mathbb{R}^d} f(x)\psi(x) w_k(x)\, dx.$$

Moreover, $G^k f$ is radially symmetric whenever f is.

P r o o f. Let $r > 0$ be such that the support of f is contained in B_r. Let us assume first that $f \geqslant 0$ and put $g = \|f\| 1_{B_r} - f$. Then, applying the Green operator G^k, we obtain

$$G^k f + G^k g = \|f\| G^k 1_{B_r}. \tag{5.4}$$

Since $G^k f$ and $G^k g$ are lower semicontinuous on $\mathbb{R}^d$ and $G^k 1_{B_r} \in C_0(\mathbb{R}^d)$ (see (5.1)), we immediately deduce from (5.4) that $G^k f \in C_0(\mathbb{R}^d)$. For f of arbitrary sign, we write $f = f^+ - f^-$, where $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$. Then

the same reasoning shows that $G^k f^+$ and $G^k f^-$ are in $C_0(\mathbb{R}^d)$. Hence $G^k f = G^k f^+ - G^k f^-$ is in $C_0(\mathbb{R}^d)$, as desired. Let $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$. Then, by (4.1), for every $y \in \mathbb{R}^d$ we have

$$\int_{\mathbb{R}^d} G^k(x,y)\Delta_k\psi(x)w_k(x) = -\psi(y).$$

Hence,

$$\begin{aligned}\int_{\mathbb{R}^d} G^k f(x)\Delta_k\psi(x)w_k(x)\,dx &= \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d} G^k(x,y)f(y)w_k(y)\,dy\Big)\Delta_k\psi(x)w_k(x)\,dx\\ &= \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d} G^k(x,y)\Delta_k\psi(x)w_k(x)\,dx\Big)f(y)w_k(y)\,dy\\ &= -\int_{\mathbb{R}^d} f(y)\psi(y)w_k(y)\,dy.\end{aligned}$$

Formula (5.1) justifies the transformation of the above integrals by Fubini's theorem. Now, assume that f is radially symmetric. Let $(f_n)_n$ be an increasing sequence of functions of the form

$$f_n = \sum_{i=1}^{n} \alpha_i 1_{B_{r_i}},$$

which converges pointwise to f on $\mathbb{R}^d$. Clearly, by formula (5.1), $G^k f_n$ is radially symmetric. On the other hand, using the dominated convergence theorem, we get for every $x \in \mathbb{R}^d$, $\lim_{n\to\infty} G^k f_n(x) = G^k f(x)$. Thus $G^k f$ is radially symmetric. ∎

For every open set D, we define the Green operator G_D^k on $\mathcal{B}_b(D)$ by

$$G_D^k f(x) := E^x\big[\int_0^{\tau_D} f(X_s)\,ds\big], \quad x \in D.$$

For every $f \in \mathcal{B}_b(D)$, we denote by $\widetilde{f}$ the extension of f on $\mathbb{R}^d$ such that $\widetilde{f} = 0$ on $\mathbb{R}^d \setminus D$. Since the Dunkl process satisfies the strong Markov property, for every $x \in D$ we have

$$\begin{aligned}G^k\widetilde{f}(x) &= E^x\big[\int_0^{\infty} \widetilde{f}(X_s)\,ds\big]\\ &= E^x\big[\int_0^{\tau_D} \widetilde{f}(X_s)\,ds\big] + E^x\big[\int_{\tau_D}^{\infty} \widetilde{f}(X_s)\,ds\big]\\ &= E^x\big[\int_0^{\tau_D} f(X_s)\,ds\big] + E^x\Big[E^{X_{\tau_D}}\big[\int_0^{\infty} \widetilde{f}(X_s)ds\big]\Big]\\ &= E^x\big[\int_0^{\tau_D} f(X_s)\,ds\big] + H_D G^k\widetilde{f}(x).\end{aligned}$$

Therefore,

$$G_D^k f = G^k \widetilde{f} - H_D G^k \widetilde{f} \quad \text{on } D. \tag{5.5}$$

Let B be an open ball of $\mathbb{R}^d$ of center zero and radius $r > 0$. Then it follows from (5.5) that, for every $f \in \mathcal{B}_b(B)$, $G_B^k f$ can be represented by

$$G_B^k f(x) = \int_B G_B^k(x,y) f(y) w_k(y)\, dy,$$

where, for every $x, y \in B$,

$$G_B^k(x,y) := G^k(x,y) - \int_{\partial B} G^k(y,z) H_B(x,dz). \tag{5.6}$$

Since, by (2.4), for every $y, z \in \mathbb{R}^d$, we have

$$G^k(y,z) \leqslant \frac{c_k \Gamma(m/2-1)}{4\,(|y|-|z|)^{m-2}}, \tag{5.7}$$

it is immediate to see that, for every $x, y \in B$,

$$\int_{\partial B} G^k(y,z) H_B(x,dz) \leqslant \frac{c_k \Gamma(m/2-1)}{4\,(|y|-r)^{m-2}} < \infty.$$

Therefore, $G_B^k(x,y)$ introduced in (5.6) exists, and so the Green function $G_B^k(\cdot,\cdot)$ is well defined from $B \times B$ into $]0,\infty]$. In the following corollary, we collect some properties of the Green operator G_B^k.

COROLLARY 5.1. *Let $f \in \mathcal{B}_b(B)$. Then $G_B^k f \in C_0(B)$ and*

$$\Delta_k G_B^k f = -f \quad \text{in } B \tag{5.8}$$

in the distributional sense.

P r o o f. Clearly, $G_B^k f$ is continuous on B since $G^k \widetilde{f}$ and $H_B G^k \widetilde{f}$ are. For every $z \in \partial B$,

$$\lim_{x \to z} G_B^k f(x) = 0$$

since $\lim_{x \to z} H_B G^k \widetilde{f}(x) = G^k \widetilde{f}(z)$. Thus $G_B^k f \in C_0(B)$. Formula (5.8) follows immediately from (5.3) and (5.5). ∎

PROPOSITION 5.2. *For every $M > 0$, the family $\{G_B^k f,\ \|f\| \leqslant M\}$ is relatively compact in $C_0(B)$ endowed with the uniform norm.*

P r o o f. In virtue of the Arzelà–Ascoli theorem, we need to show that $\{G_B^k f,\ \|f\| \leqslant M\}$ is uniformly bounded and equicontinuous on B. Let r be the radius of the ball B. Let $f \in \mathcal{B}_b(B)$ be such that $\|f\| \leqslant M$. Obviously, $\|G_B^k f\| \leqslant M\|G_B^k 1\| \leqslant M\|G^k 1_B\|$. Thus, using (5.1), we obtain

$$\|G_B^k f\| \leqslant \frac{r^2 M}{2(m-2)}.$$

This means that the family $\{G_B^k f,\ \|f\| \leqslant M\}$ is uniformly bounded. Next, we claim that the family $\{G_B^k(x,\cdot),\ x \in B\}$ is uniformly integrable with respect to the measure $w_k(y)\,dy$. Let $x \in B$ and $\epsilon > 0$ be small enough. Let $A_{t,s}$ be the annulus of $\mathbb{R}^d$ of center zero and radius $t = \max(0, |x| - \epsilon)$ and $s = |x| + \epsilon$. Then, for every Borel subset D of B, we have

$$\begin{aligned}\int_D G_B^k(x,y)w_k(y)dy &\leqslant \int_D G^k(x,y)w_k(y)dy \\ &= \int_{D\cap A_{t,s}} G^k(x,y)w_k(y)\,dy + \int_{D\setminus A_{t,s}} G^k(x,y)w_k(y)dy \\ &\leqslant G^k 1_{A_{t,s}}(x) + \Big(\sup_{y\in D\setminus A_{t,s}} G^k(x,y)\Big)\int_D w_k(y)dy.\end{aligned}$$

Hence, it follows from (5.7) and (5.2) that

$$\int_D G_B^k(x,y)w_k(y)\,dy \leqslant \frac{4r}{m-2}\epsilon + \frac{c_k\Gamma(m/2-1)}{4\epsilon^{m-2}}\int_D w_k(y)\,dy.$$

Put $\eta = \epsilon^{m-1}$. Then for every Borel subset D of B such that $\int_D w_k(y)\,dy < \eta$, we have

$$\int_D G_B^k(x,y)w_k(y)\,dy \leqslant \left(\frac{4r}{m-2} + \frac{c_k\Gamma(m/2-1)}{4}\right)\epsilon.$$

Thus, the uniform integrability of the family $\{G_B^k(x,\cdot),\ x \in B\}$ is shown. Therefore, in virtue of Vitali's convergence theorem, for $z \in B$,

$$\lim_{x\to z}\int_B |G_B^k(x,y) - G_B^k(z,y)|w_k(y)\,dy = 0.$$

Hence, the family $\{G_B^k f,\ \|f\| \leqslant M\}$ is equicontinuous on B since

$$\begin{aligned}\lim_{x\to z}\sup_{\|f\|\leqslant M} &|G_B^k f(x) - G_B^k f(z)| \\ &\leqslant M\lim_{x\to z}\int_B |G_B^k(x,y) - G_B^k(z,y)|w_k(y)\,dy = 0. \quad \blacksquare\end{aligned}$$

6. SEMILINEAR DIRICHLET PROBLEM

Let B be an open ball of $\mathbb{R}^d$ of center zero. Let $\varphi : [0,\infty[\to [0,\infty[$ be a nondecreasing continuous function such that $\varphi(0) = 0$. By a solution of

$$\Delta_k u = \varphi(u) \quad \text{in } B \tag{6.1}$$

we shall mean every function $u \in C(B)$ such that

$$\int_B u(x)\,\Delta_k\psi(x)\,w_k(x)\,dx = \int_B \varphi\big(u(x)\big)\,\psi(x)\,w_k(x)\,dx$$

holds for every $\psi \in \mathcal{C}_c^\infty(B)$. We recall from Theorem 4.2 that if $\varphi \equiv 0$, then $H_B f$ is the unique solution of (6.1) satisfying $u = f$ on ∂B. In all the following, we assume that φ is not identically zero.

LEMMA 6.1. *Let $u \in C^+(\overline{B})$. Then u is a solution of equation* (6.1) *if and only if $u + G_B^k\big(\varphi(u)\big) = H_B u$ on B.*

P r o o f. Let us note first that $G_B^k\big(\varphi(u)\big) \in C_0(B)$ since the function $\varphi(u)$ is bounded on B. Put $h := u + G_B^k\big(\varphi(u)\big)$. Clearly, $h \in C(\overline{B})$ and $h = u$ on ∂B. On the other hand, using Fubini's theorem and formula (5.8), we obtain for every $\psi \in C_c^\infty(B)$,

$$\begin{aligned}
&\int_B h(x)\Delta_k\psi(x)w_k(x)\,dx \\
&\qquad = \int_B u(x)\Delta_k\psi(x)w_k(x)\,dx + \int_B G_B^k\big(\varphi(u)\big)(x)\Delta_k\psi(x)w_k(x)\,dx \\
&\qquad = \int_B u(x)\Delta_k\psi(x)w_k(x)\,dx - \int_B \varphi\big(u(x)\big)\psi(x)w_k(x)\,dx.
\end{aligned}$$

So, $\Delta_k u = \varphi(u)$ in B if and only if $\Delta_k h = 0$ in B. In this case, since $h = u$ on ∂B, the uniqueness of the solution to problem (4.10) yields $h = H_B u$ on B. This completes the proof. ∎

LEMMA 6.2. *Let $u, v \in C^+(\overline{B})$ be two solutions of equation* (6.1). *If $u \geqslant v$ on ∂B, then $u \geqslant v$ on B.*

P r o o f. Define $w := u - v$ and $\rho := \varphi(u) - \varphi(v)$. By Lemma 6.1, we have

$$w + G_B^k\rho = H_B w \quad \text{on } \overline{B}. \tag{6.2}$$

Suppose that the open set $D := \{x \in B;\ w(x) < 0\}$ is not empty. Since φ is nondecreasing, it follows that $\rho \leqslant 0$ on D, and hence $G_D^k\rho \leqslant 0$ on D. Let $x \in D$.

It is clear that $\overline{B}$ contains the support of the measure $H_D(x,\cdot)$. Now integrate (6.2) with respect to $H_D(x,\cdot)$ to obtain

$$H_D w(x) + H_D(G_B^k \rho)(x) = H_D H_B w(x) = H_B w(x).$$

Consequently,

$$H_D w(x) = H_B w(x) - H_D(G_B^k \rho)(x) = w(x) + \big(G_B^k \rho(x) - H_D G_B^k \rho(x)\big). \tag{6.3}$$

On the other hand, using the strong Markov property, we obtain

$$\begin{aligned} G_B^k \rho(x) - G_D^k \rho(x) &= E^x\Big[\int_{\tau_D}^{\tau_B} \rho(X_s)\,ds\Big] = E^x\Big[E^{X_{\tau_D}}\Big[\int_0^{\tau_B} \rho(X_s)\,ds\Big]\Big] \\ &= H_D G_B^k \rho(x). \end{aligned} \tag{6.4}$$

Thus, it follows from (6.3) and (6.4) that $w(x) + G_D^k \rho(x) = H_D w(x)$. But this is absurd since $w(x) + G_D^k \rho(x) < 0$ and $H_D w(x) \geqslant 0$. Therefore, D is empty, and consequently $u \geqslant v$ on B. ∎

THEOREM 6.1. *For every $f \in C^+(\partial B)$, the semilinear Dirichlet problem*

$$\begin{cases} \Delta_k u = \varphi(u) & \text{in } B, \\ u = f & \text{on } \partial B \end{cases} \tag{6.5}$$

admits one and only one solution $u \in C^+(\overline{B})$.

P r o o f. It follows from Lemma 6.2 that problem (6.5) admits at most one solution. To prove the existence, in virtue of Lemma 6.1, it will be sufficient to establish the existence of $u \in C^+(\overline{B})$ such that

$$u + G_B^k\big(\varphi(u)\big) = H_B f \quad \text{on } B. \tag{6.6}$$

Since $G_B^k 1 \leqslant G^k 1_B$, we immediately deduce by (5.1) that $\sup_{x\in B} G_B^k 1(x) < \infty$. Let $f \in C^+(\partial B)$, $a = \|f\|$ and $M = a + \varphi(a)\|G_B^k 1\|$. Let ϕ be the function defined on $\mathbb{R}$ by

$$\phi(t) = \begin{cases} 0 & \text{if } t \leqslant 0, \\ \varphi(t) & \text{if } 0 \leqslant t \leqslant a, \\ \varphi(a) & \text{if } t \geqslant a. \end{cases}$$

Let $\Lambda := \{u \in C(\overline{B});\ \|u\| \leqslant M\}$ and consider the operator $T : \Lambda \to C(\overline{B})$ defined by

$$Tu(x) = H_B f(x) - G_B^k\big(\phi(u)\big)(x), \quad x \in \overline{B}.$$

Since $\sup_{x\in B}\phi\big(u(x)\big)\leqslant\varphi(a)$, we easily deduce that

$$\|Tu\|\leqslant M$$

for every $u\in\Lambda$. This implies that $T(\Lambda)\subset\Lambda$. Now, let $(u_n)_n$ be a sequence in Λ converging uniformly to $u\in\Lambda$. Let $\varepsilon>0$. Since ϕ is uniformly continuous on the interval $[-M,M]$, we immediately deduce that there exists $n_0\in\mathbb{N}$ such that, for every $n\geqslant n_0$,

$$\|\phi(u_n)-\phi(u)\|\leqslant\varepsilon.$$

Then, for every $n\geqslant n_0$ and every $x\in B$,

$$|Tu_n(x)-Tu(x)|\leqslant G_B^k\big(|\phi(u_n)-\phi(u)|\big)(x)\leqslant\varepsilon\sup_{x\in B}G_B^k1(x).$$

This show that $(Tu_n)_n$ converges uniformly to Tu, and therefore T is continuous. On the other hand, Λ is a closed bounded convex subset of $C(\overline{B})$ and, in virtue of Proposition 5.2, $T(\Lambda)$ is relatively compact. Thus, the Schauder fixed point theorem ensures the existence of a function $u\in\Lambda$ such that

$$u+G_B^k\big(\phi(u)\big)=H_Bf\quad\text{ on }B.$$

Clearly, $u\in C(\overline{B})$ and $u(x)\leqslant H_Bf(x)\leqslant a$ for every $x\in B$. So, to obtain (6.6), we need to show that $\phi(u)=\varphi(u)$ on B, or equivalently, $u\geqslant0$ on B. Assume that the open set $D:=\{x\in B,\ u(x)<0\}$ is not empty. Let $x\in D$. Then,

$$H_Du(x)=H_D\Big(H_Bu-G_B^k\big(\phi(u)\big)\Big)(x)=H_Bu(x)-H_DG_B^k\big(\phi(u)\big)(x).$$

The same reasoning as in (6.4), based on the strong Markov property, shows that

$$H_DG_B^k\big(\phi(u)\big)(x)=G_B^k\big(\phi(u)\big)(x)-G_D^k\big(\phi(u)\big)(x).$$

Thus, because $\phi(u)=0$ on D, we get

$$\begin{aligned}H_Du(x)&=H_Bu(x)-G_B^k\big(\phi(u)\big)(x)+G_D^k\big(\phi(u)\big)(x)\\&=u(x)+G_D^k\big(\phi(u)\big)(x)=u(x)<0.\end{aligned}$$

But, $H_Du(x)\geqslant0$ since $u\geqslant0$ on $\overline{B}\setminus D$, which contains the support of $H_D(x,\cdot)$. So D must be empty, and consequently $u\geqslant0$ on B. This completes the proof. ∎

REFERENCES

[1] J. Bliedtner and W. Hansen, *Potential Theory: An Analytic and Probabilistic Approach to Balayage*, Springer, Berlin 1986.

[2] R. M. Blumenthal and R. K. Getoor, *Markov Processes and Potential Theory*, Academic Press, New York–London 1968.

[3] H. Brezis and S. Kamin, *Sublinear elliptic equations in* R^n, Manuscripta Math. 74 (1) (1992), pp. 87–106.

[4] N. Demni, *First hitting time of the boundary of the Weyl chamber by radial Dunkl processes*, SIGMA Symmetry Integrability Geom. Methods Appl. 4 (2008), paper 074.

[5] J. F. van Diejen and L. Vinet (eds.), *Calogero–Moser–Sutherland Models*, Springer, New York 2000.

[6] C. F. Dunkl, *Reflection groups and orthogonal polynomials on the sphere*, Math. Z. 197 (1) (1988), pp. 33–60.

[7] C. F. Dunkl, *Differential-difference operators associated to reflection groups*, Trans. Amer. Math. Soc. 311 (1) (1989), pp. 167–183.

[8] C. F. Dunkl, *Integral kernels with reflection group invariance*, Canad. J. Math. 43 (6) (1991), pp. 1213–1227.

[9] E. B. Dynkin, *Solutions of semilinear differential equations related to harmonic functions*, J. Funct. Anal. 170 (2) (2000), pp. 464–474.

[10] K. El Mabrouk, *Entire bounded solutions for a class of sublinear elliptic equations*, Nonlinear Anal. 58 (1–2) (2004), pp. 205–218.

[11] L. Gallardo and M. Yor, *A chaotic representation property of the multidimensional Dunkl processes*, Ann. Probab. 34 (4) (2006), pp. 1530–1549.

[12] K. Hassine, *Mean value property of* Δ_k*-harmonic functions on* W*-invariant open sets*, Afr. Mat. 27 (7–8) (2016), pp. 1275–1286.

[13] K. Hikami, *Dunkl operator formalism for quantum many-body problems associated with classical root systems*, J. Phys. Soc. Japan 65 (2) (1996), pp. 394–401.

[14] N. Ikeda and S. Watanabe, *On some relations between the harmonic measure and the Lévy measure for a certain class of Markov processes*, J. Math. Kyoto Univ. 2 (1962), pp. 79–95.

[15] M. F. E. de Jeu, *The Dunkl transform*, Invent. Math. 113 (1) (1993), pp. 147–162.

[16] J. B. Keller, *On solutions of* $\Delta u = f(u)$, Comm. Pure Appl. Math. 10 (1957), pp. 503–510.

[17] T. Khongsap and W. Wang, *Hecke–Clifford algebras and spin Hecke algebras IV: Odd double affine type*, SIGMA Symmetry Integrability Geom. Methods Appl. 5 (2009), paper 012.

[18] A. V. Lair and A. W. Wood, *Large solutions of sublinear elliptic equations*, Nonlinear Anal. 39 (6) (2000), pp. 745–753.

[19] L. Lapointe and L. Vinet, *Exact operator solution of the Calogero–Sutherland model*, Comm. Math. Phys. 178 (2) (1996), pp. 425–452.

[20] A. C. Lazer and P. J. McKenna, *On a singular nonlinear elliptic boundary-value problem*, Proc. Amer. Math. Soc. 111 (3) (1991), pp. 721–730.

[21] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Springer, Berlin 1966.

[22] H. Mejjaoli and K. Trimèche, *Hypoellipticity and hypoanalyticity of the Dunkl Laplacian operator*, Integral Transforms Spec. Funct. 15 (6) (2004), pp. 523–548.

[23] R. Osserman, *On the inequality* $\Delta u \geqslant f(u)$, Pacific J. Math. 7 (1957), pp. 1641–1647.

[24] G. Ren and L. Liu, *Liouville theorem for Dunkl polyharmonic functions*, SIGMA Symmetry Integrability Geom. Methods Appl. 4 (2008), paper 076.

[25] M. Rösler, *Generalized Hermite polynomials and the heat equation for Dunkl operators*, Comm. Math. Phys. 192 (3) (1998), pp. 519–542.

[26] M. Rösler, *Positivity of Dunkl's intertwining operator*, Duke Math. J. 98 (3) (1999), pp. 445–463.

[27] M. Rösler, *A positive radial product formula for the Dunkl kernel*, Trans. Amer. Math. Soc. 355 (6) (2003), pp. 2413–2438.
[28] K. Trimèche, *Paley–Wiener theorems for the Dunkl transform and Dunkl translation operators*, Integral Transforms Spec. Funct. 13 (1) (2002), pp. 17–38.
[29] Y. Xu, *Orthogonal polynomials for a family of product weight functions on the spheres*, Canad. J. Math. 49 (1) (1997), pp. 175–192.

Mohamed Ben Chrouda
Monastir University
Higher Institute of Computer Science
and Mathematics
5000 Monastir, Tunisia
E-mail: Mohamed.BenChrouda@isimm.rnu.tn

Khalifa El Mabrouk
Sousse University
High School of Sciences and Technology
4011 Hammam Sousse, Tunisia
E-mail: khalifa.elmabrouk@fsm.rnu.tn

Kods Hassine
Monastir University
Faculty of Science of Monastir
5000 Monastir, Tunisia
E-mail: hassinekods@gmail.com

Received on 18.5.2016;
revised version on 18.2.2017

Probability and Mathematical Statistics 38, z. 2, 2018

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 271–286
doi:10.19195/0208-4147.38.2.2

CONVERGENCE OF RANDOM OSCILLATORY INTEGRALS IN THE PRESENCE OF LONG-RANGE DEPENDENCE AND APPLICATION TO HOMOGENIZATION

BY

ATEF LECHIHEB (TUNIS), IVAN NOURDIN (LUXEMBOURG), GUANGQU ZHENG (LAWRENCE), AND EZZEDINE HAOUALA (TUNIS)

Abstract. This paper deals with the asymptotic behavior of random oscillatory integrals in the presence of long-range dependence. As a byproduct, we solve the corrector problem in random homogenization of one-dimensional elliptic equations with highly oscillatory random coefficients displaying long-range dependence, by proving convergence to stochastic integrals with respect to Hermite processes.

2010 AMS Mathematics Subject Classification: Primary: 60F05, 80M40; Secondary: 60H05, 60H20, 60G10, 60G18.

Key words and phrases: Elliptic equation, Hermite process, oscillatory integral, corrector, homogenization.

1. MAIN RESULTS

1.1. Convergence of random oscillatory integrals. One of our goals in the paper is to study, once properly normalized, the distributional convergence of some random oscillatory integrals of the form

$$\int_0^1 \Phi[g(x/\varepsilon)]h(x)\,dx, \tag{1.1}$$

where

- $h \in C([0,1])$ is deterministic,
- $\{g(x)\}_{x\in\mathbb{R}_+}$ is a certain centered stationary Gaussian process exhibiting long-range correlation,
- $\Phi \in L^2(\mathbb{R},\nu)$ has Hermite rank $m \geqslant 1$ (with ν the standard Gaussian measure).

As we will see later, the main motivation of this study comes from the random corrector problem studied in [4].

Let us first introduce the Gaussian process $\{g(x)\}_{x\in\mathbb{R}_+}$ we will deal with throughout all this paper. It is constructed as follows:

1. Let $m \in \mathbb{N}^*$ be fixed, let $H_0 \in \big(1-\frac{1}{2m}, 1\big)$, and set $H = 1 + m(H_0 - 1) \in (1/2, 1)$.

2. Fix a slowly varying function $L : (0, +\infty) \to (0, +\infty)$ at $+\infty$, that is, consider a measurable and locally bounded function L such that $L(\lambda x)/L(x) \to 1$ as $x \to +\infty$, for every $\lambda > 0$. Assume furthermore that L is bounded away from 0 and $+\infty$ on every compact subset of $(0, +\infty)$. (See [3] for more details on slowly varying functions.)

3. Let $e : \mathbb{R} \to \mathbb{R}$ be a square-integrable function such that

(3a) $\int_{\mathbb{R}} e(u)^2\, du = 1$;

(3b) $|e(u)| \leqslant C u^{H_0-3/2} L(u)$ for almost all $u > 0$ and for some absolute constant C;

(3c) $e(u) \sim C_0 u^{H_0-3/2} L(u)$, where $C_0 = \big(\int_0^\infty (u+u^2)^{H_0-3/2}\, du\big)^{-1/2}$;

(3d) there exists $0 < \gamma < \min\big\{H_0 - \big(1-\frac{1}{2m}\big), 1 - H_0\big\}$ such that

$$\int_{-\infty}^{0} |e(u)e(xy+u)|\, du = o\big(x^{2H_0-2}L(x)^2\big) y^{2H_0-2-2\gamma}$$

as $x \to \infty$, uniformly in $y \in (0, t]$ for each given $t > 0$.

4. Finally, let W be a two-sided Brownian motion.

Bearing all these ingredients in mind, we can now set, for $x \in \mathbb{R}_+$,

$$g(x) := \int_{-\infty}^{\infty} e(x-\xi) dW_\xi. \tag{1.2}$$

REMARK 1.1. (i) Assumptions (3a) and 4 ensure that $\{g(x)\}_{x\in\mathbb{R}_+}$ is a normalized centered Gaussian process.

(ii) Assumption (3b) controls $|e(u)|$ for small u, while assumption (3d) ensures that the "forward" contribution of $e(u)$ is ultimately negligible due to the following computation:

$$\begin{aligned}
\mathbb{E}[g(s)g(s+x)] &= \int_{-\infty}^{\infty} e(s-\xi)e(s+x-\xi)\, d\xi = \int_{-\infty}^{\infty} e(u)e(u+x)\, du \\
&= \int_{-\infty}^{0} e(u)e(u+x)\, du + \int_{0}^{\infty} e(u)e(u+x)\, du \\
&= o\big(x^{2H_0-2}L(x)^2\big) + x\int_0^\infty e(xu)e(xu+x)\, du.
\end{aligned}$$

(iii) Assumption (3c) ensures that the process $\{g(x)\}_{x\in\mathbb{R}_+}$ exhibits the following asymptotic behavior:

$$R_g(x) := \mathbb{E}[g(s)g(s+x)] \sim x^{2H_0-2}L(x)^2 \quad \text{as } x \to +\infty, \tag{1.3}$$

see [12], equation (2.3).

In Section 3.1, we will show that the random integral given by (1.1) exhibits the following asymptotic behavior as $\varepsilon \to 0$.

THEOREM 1.1. *Let g be the centered stationary Gaussian process defined by* (1.2) *and assume that $\Phi \in L^2(\mathbb{R}, \nu)$ has Hermite rank $m \geqslant 1$. Then, for any $h \in C([0,1])$, the following convergence in law holds*:

$$M_h^\varepsilon := \frac{1}{\varepsilon d(1/\varepsilon)} \int_0^1 \Phi[g(x/\varepsilon)] h(x)\, dx \xrightarrow{\varepsilon \downarrow 0} M_h^0 := \frac{V_m}{m!} \int_0^1 h(x)\, dZ(x), \tag{1.4}$$

where Z is the mth order Hermite process defined by (2.4) *below, and $d(\cdot)$ is defined by*

$$d(x) = \sqrt{\frac{m!}{H(2H-1)}}\, x^H L(x)^m. \tag{1.5}$$

As we already mentioned, the fine analysis of the asymptotic behavior of (1.4) is motivated by the random corrector problem studied in [4]; it will be described below.

1.2. A motivating example. Theorem 1.1 appears to be especially useful and relevant in the study of the following homogenization problem. Consider the following *one-dimensional* elliptic equation displaying random coefficients:

$$\begin{cases} -\dfrac{d}{dx}\left(a(x/\varepsilon, \omega) \dfrac{d}{dx} u^\varepsilon(x, \omega) \right) = f(x), \quad x \in (0,1), \quad \varepsilon > 0, \\ u^\varepsilon(0, \omega) = 0, \quad u^\varepsilon(1, \omega) = b \in \mathbb{R}. \end{cases} \tag{1.6}$$

In (1.6), the random potential $\{a(x)\}_{x \in \mathbb{R}_+}$ is assumed to be a uniformly bounded, positive[1] and stationary stochastic process, whereas the data f is continuous. This model has received a lot of interests in the literature (see, e.g., [5], pp. 13–14).

Taking strong advantage of the fact that the ambient dimension is one, it is immediate to check that the solution to (1.6) is given explicitly by

$$u^\varepsilon(x, \omega) = c^\varepsilon(\omega) \int_0^x \frac{1}{a(y/\varepsilon, \omega)}\, dy - \int_0^x \frac{F(y)}{a(y/\varepsilon, \omega)}\, dy, \tag{1.7}$$

where $F(x) := \int_0^x f(y)\, dy$ is the antiderivative of f vanishing at zero, and where

$$c^\varepsilon(\omega) := \left(b + \int_0^1 \frac{F(y)}{a(y/\varepsilon, \omega)}\, dy \right) \left(\int_0^1 \frac{1}{a(y/\varepsilon, \omega)}\, dy \right)^{-1}.$$

[1] That is, there exists $r \in (0,1)$ such that $r \leqslant a(x) \leqslant r^{-1}$ for every $(x, \omega) \in \mathbb{R}_+ \times \Omega$.

Under suitable ergodic and stationary assumptions on a, the ergodic theorem applied to (1.7) implies that u^ε converges pointwise to $\bar{u}$ as $\varepsilon \to 0$, where

$$\bar{u}(x) = \frac{c^* x}{a^*} - \int_0^x \frac{F(y)}{a^*}\, dy,$$

with $c^* := ba^* + \int_0^1 F(y)\, dy$ and

$$a^* := \frac{1}{\mathbb{E}[1/a(0)]}.$$

The above parameter a^* is usually referred to as the *effective diffusion coefficient* in the literature, see e.g. [10]. It is also immediately checked that $\bar{u}$ is the unique solution to the following *deterministic* equation:

$$\begin{cases} -\dfrac{d}{dx}\left(a^* \dfrac{d}{dx}\bar{u}(x) \right) = f(x), \quad x \in (0,1), \\ \bar{u}(0) = 0, \quad \bar{u}(1) = b. \end{cases} \tag{1.8}$$

Interested readers can refer to [2] for a recent review on models involving more general elliptic equations.

In this work, we address to the random corrector problem for (1.6) in presence of *long-range* media, that is, we analyze the behavior of the random fluctuations between u^ε and $\bar{u}$ when the random potential a is obtained by means of a long-range process (see below for the details). Taking advantage of the explicit expressions for both (1.6) and (1.8), it is easy but crucial to observe that the random corrector $u^\varepsilon(x) - \bar{u}(x)$ can be fully expressed by means of random oscillatory integrals of the form

$$\int_0^1 \left[\frac{1}{a(y/\varepsilon)} - \frac{1}{a^*} \right] h(y)\, dy \tag{1.9}$$

for some function h. Thus, the random corrector problem for (1.6) reduces in a careful analysis of the asymptotic behavior of random quantities of the form (1.9) as $\varepsilon \to 0$. To this aim, we need to give a precise description of the form of the process a.

Let ν denote the standard Gaussian measure on $\mathbb{R}$. Every $\Phi \in L^2(\mathbb{R}, \nu)$ admits the series expansion

$$\Phi = \sum_{q=0}^{\infty} \frac{V_q}{q!} H_q, \quad \text{with } V_q := \int_{\mathbb{R}} \Phi(x) H_q(x) \nu(dx), \tag{1.10}$$

where $H_q(x) = (-1)^q \exp(x^2/2) \frac{d^q}{dx^q} \exp(-x^2/2)$ denotes the qth Hermite polynomial. Recall that the integer $m_\Phi := \inf\{q \geqslant 0 :\ V_q \neq 0\}$ is called the *Hermite*

rank of Φ (with the convention $\inf\emptyset = +\infty$). For any integer $m \geqslant 1$, we define $\mathscr{G}_m$ to be the collection of all square-integrable functions (with respect to the standard Gaussian measure on $\mathbb{R}$) that have Hermite rank m.

Using Theorem 1.1 as the main ingredient, we will prove the following result about the asymptotic behaviour of the random corrector associated with (1.6).

THEOREM 1.2. *Fix an integer $m \geqslant 1$ as well as two real numbers $H_0 \in \big(1 - \frac{1}{2m}, 1\big)$ and $b \in \mathbb{R}$, and let $\{a(x)\}_{x\in\mathbb{R}_+}$ be a uniformly bounded, positive and stationary stochastic process. Assume in addition that $q = \{q(x)\}_{x\in\mathbb{R}_+}$ given by*

$$q(x) = \frac{1}{a(x)} - \frac{1}{a^*}, \quad \text{where } a^* := 1/\mathbb{E}[1/a(0)], \tag{1.11}$$

has the form

$$q(x) = \Phi\big(g(x)\big), \tag{1.12}$$

where $\Phi \in L^2(\mathbb{R}, \nu)$ belongs to $\mathscr{G}_m$ and $\{g(x)\}_{x\in\mathbb{R}_+}$ is the Gaussian process given by (1.2)*. Finally, let $f : [0,1] \to \mathbb{R}$ be continuous and let us consider the solutions u^ε and $\bar{u}$ of* (1.6) *and* (1.8) *respectively. Then, for each $\varepsilon > 0$, the random corrector $u^\varepsilon - \bar{u}$ is a continuous process on $[0,1]$. Moreover, we have the following convergence in law on $C([0,1])$ endowed with the supremum norm as $\varepsilon \to 0$:*

$$\left\{\frac{u^\varepsilon(x) - \bar{u}(x)}{\varepsilon d(1/\varepsilon)}\right\}_{x\in[0,1]} \Longrightarrow \left\{\frac{V_m}{m!}\int_{\mathbb{R}} F(x,y)\,dZ(y)\right\}_{x\in[0,1]},$$

where d is given by (1.5),

$$F(x) = \int_0^x f(y)dy, \quad c^* = a^*b + \int_0^1 F(y)\,dy,$$

$$F(x,y) = [c^* - F(y)]\mathbf{1}_{[0,x]}(y) + x\big(F(y) - \int_0^1 F(z)dz - a^*b\big)\mathbf{1}_{[0,1]}(y),$$

and Z is the Hermite process of order m and self-similar index

$$H := 1 + m(H_0 - 1) \in (1/2, 1).$$

(*The definition of Z is given in Theorem* 2.1 *below.*)

Note that it is not difficult to construct a process a satisfying all the assumptions of Theorem 1.2. Indeed, bearing in mind the notation of Theorem 1.2, we can write

$$a(x) = \left(q(x) + \frac{1}{a^*}\right)^{-1} = \left(\Phi\big(g(x)\big) + \frac{1}{a^*}\right)^{-1}. \tag{1.13}$$

First, we note that since g given by (1.2) is stationary, clearly the same holds for a, whatever the expression of Φ. Second, given any fixed $a^* > 0$, we can construct a bounded measurable function $\Phi \in \mathscr{G}_2$ with $\|\Phi\|_\infty \leqslant 1/(2a^*)$ as follows.

Let h_1, h_2 be two bounded measurable functions; then it is clear that they belong to $L^2(\mathbb{R}, \nu)$ and admit the series expansion

$$h_1 - \int_{\mathbb{R}} h_1 \, d\nu = \sum_{k=1}^{\infty} a_k H_k \quad \text{and} \quad h_2 - \int_{\mathbb{R}} h_2 \, d\nu = \sum_{k=1}^{\infty} b_k H_k,$$

where the coefficients a_k, b_k are defined in an obvious manner. Therefore, the function

$$\Psi := b_1\big(h_1 - \int_{\mathbb{R}} h_1 \, d\nu\big) - a_1\big(h_2 - \int_{\mathbb{R}} h_2 \, d\nu\big)$$

is bounded and belongs to $\mathscr{G}_2$. Then we pick $\Phi = \Psi/(2a^*\|\Psi\|_\infty) \in \mathscr{G}_2$. Therefore, $a(x)$ defined by (1.13) satisfies

$$0 < \frac{2a^*}{3} \leqslant a(x) \leqslant 2a^*. \tag{1.14}$$

Inductively, one can construct a bounded measurable Φ with Hermite rank $m \geqslant 3$ (by starting with two bounded functions in $\mathscr{G}_{m-1}$) such that the process $\{a(x), x \in \mathbb{R}\}$ given in (1.13) satisfies (1.14).

Another possibility of constructing such a process $\{a(x), x \in \mathbb{R}\}$ is stated (more explicitly) as follows: let us fix $0 < t_1 < \ldots < t_m$ and consider the unique $(m+1)$-tuple $(b_0, \ldots, b_m)$ satisfying

$$\begin{cases} \sum\limits_{l=0}^{m} b_l \, e^{-kt_l} = 0 & \text{for all } k \in \{0, \ldots, m-1\}, \\ \sum\limits_{l=0}^{m} b_l \, e^{-mt_l} = 1. \end{cases} \tag{1.15}$$

(The existence and uniqueness of a solution to (1.15) is a consequence of a Vandermonde determinant.) Now, consider any measurable function ψ satisfying

$$0 \leqslant \psi \leqslant \frac{1}{2a^* \sum\limits_{l=0}^{m} |b_l|}. \tag{1.16}$$

Since ψ belongs obviously to $L^2(\mathbb{R}, \nu)$, it may be expanded in Hermite polynomials as $\psi = \sum_{k=0}^{\infty} a_k H_k$. We assume moreover that $a_m \neq 0$. (The existence of ψ satisfying both (1.16) and $a_m \neq 0$ is clear by a contradiction argument.) Now, let

$$\Phi = \sum_{l=0}^{m} b_l P_{t_l} \psi,$$

where $P_t\psi(x) = \int_{\mathbb{R}} \psi(e^{-t}x + \sqrt{1-e^{-2t}}y)\nu(dy)$ is the classical Ornstein–Uhlenbeck semigroup. Due to (1.15), it is readily checked that the expansion of Φ is

$$\Phi = a_m H_m + \sum_{k=m+1}^{\infty} \Big\{ \sum_{l=0}^{m} b_l e^{-kt_l} \Big\} a_k H_k,$$

so that $\Phi \in \mathscr{G}_m$. Moreover,

$$\|\Phi\|_\infty \leqslant \sum_{l=0}^{m} |b_l| \|P_{t_l}\psi\|_\infty \leqslant \|\psi\|_\infty \sum_{l=0}^{m} |b_l| \leqslant \frac{1}{2a^*},$$

and a given by (1.13) is positive and bounded. So, the existence of a process a satisfying all the assumptions of Theorem 1.2 is shown.

Theorem 1.2 should be seen as an extension of and a unified approach to the main results of [4], and it contains these results as particular cases. More precisely, the case where the Hermite rank of Φ is $m = 1$ corresponds to Theorem 2.5 in [1] and involves the fractional Brownian motion in the limit, whereas the case where the Hermite rank of Φ is $m = 2$ corresponds to Theorem 2.2 in [4] and involves the Rosenblatt process in the limit. Also, in their last section (entitled *Conclusions and further discussion*), the authors of [4] pointed out that "it is natural to ask what would happen if the Hermite rank of Φ was greater than 2". Our Theorem 1.2 answers this question by showing (as guessed by the authors of [4]) that, in the case $m \geqslant 3$, the limit takes the form of an integral with respect to the Hermite process of order m. Finally, we would like to emphasize that our Theorem 1.2, even in the cases $m = 1$ and $m = 2$, is a strict extension of the results of [4], as we allow the possibility to deal with a slowly varying function L. That being said, our proof of Theorem 1.2 is exclusively based on the ideas and results contained in the seminal paper [12] and follows the strategy developed in [4]. In higher dimension, it is usually very hard to study the corrector theory due to the lack of the explicit form of the solution. In the recent papers [8], [9], the authors considered the discretized version of the corrector problem in higher dimension and were able to study the scaling limit to some Gaussian fields. For more details, we refer the interested readers to these two papers and the references therein.

The rest of the paper is organized as follows. In Section 2, we give some preliminary results divided into several subsections. Section 3 contains the proof of Theorems 1.2 and 1.1.

2. PRELIMINARY RESULTS

Throughout this section, we let all the notation and assumptions of Sections 1.1 and 1.2 prevail.

2.1. Asymptotic behavior of the covariance function of q. For $x \in \mathbb{R}$, set $R_q(x) = \mathbb{E}[q(0)q(x)]$. Also, recall that m is the Hermite rank of Φ. Then, proceeding in similar lines to those in Lemma 2.1 of [4], one can show that

$$|R_q(x)| = \big(o(1) + V_m^2/m!\big) L(|x|)^{2m} |x|^{-2(1-H)} \tag{2.1}$$

as $|x| \to +\infty$. Here $o(1)$ means that the term converges to zero when $x \to \infty$.

The asymptotic relation (2.1) implies the existence of some absolute constant C satisfying

$$|R_q(x)| \leqslant C\, L(|x|)^{2m} |x|^{-2(1-H)} \tag{2.2}$$

for any $x \neq 0$.

2.2. Taqqu's theorem and convergence to the Hermite process Z. Recall $d(x)$ from (1.5). Its main property is that the variance of $\frac{1}{d(x)} \int_0^x H_m\big(g(y)\big)\, dy$ is asymptotically equal to one as $x \to +\infty$.

The following result, due to Taqqu in 1979, is the key ingredient in our proofs.

THEOREM 2.1 ([12], Lemma 5.3). *Let us assume $\Phi \in \mathscr{G}_m$ and let g be given by* (1.2)*. Then, as $T \to +\infty$, the process*

$$Y_T(x) = \frac{1}{d(T)} \int_0^{Tx} \Phi[g(y)]\, dy, \quad x \in \mathbb{R}_+, \tag{2.3}$$

converges to $(V_m/m!)Z(x)$ in the sense of finite-dimensional distributions, where the mth order Hermite process Z with self-similar index $H = m(H_0 - 1) + 1$ is defined by

$$Z(x) = K(m, H_0)\Big\{ \int_{-\infty}^{\infty} dB_{\xi_1} \int_{-\infty}^{\xi_1} dB_{\xi_2} \dots \int_{-\infty}^{\xi_{m-1}} dB_{\xi_m} \int_0^x \prod_{i=1}^m (s - \xi_i)^{H_0 - 3/2} \mathbf{1}_{(\xi_i < s)}\, ds \Big\}, \tag{2.4}$$

where

$$K(m, H_0) := \sqrt{\frac{m! H(2H-1)}{\big(\int_0^\infty (u + u^2)^{H_0 - 3/2}\, du\big)^m}}$$

is the normalizing constant such that $\mathbb{E}[Z(1)^2] = 1$. (See [12], equation (1.6).)

Note that $Z(x)$ lives in the Wiener chaos of order m, which is non-Gaussian unless $m = 1$ or $x = 0$.

2.3. Wiener integral with respect to Z. Let Z be given as above and let $\mathcal{E}$ be the set of elementary (deterministic) functions, that is, the set of functions h of the form

$$h(x) = \sum_{k=1}^{\ell} a_k \mathbf{1}_{(t_k, t_{k+1}]}(x)$$

with $\ell \in \mathbb{N}^*$, $a_k \in \mathbb{R}$, $t_k < t_{k+1}$. For such h, we define the Wiener integral with respect to Z in the usual way, as a linear functional over $\mathcal{E}$:

$$\int_{\mathbb{R}} h(x)\, dZ(x) = \sum_{k=1}^{\ell} a_k [Z(t_{k+1}) - Z(t_k)].$$

One can easily verify that this definition is independent of choices of representation for elementary functions. Now we introduce the space of (deterministic) integrands for this Wiener integral:

$$\Lambda^H = \Big\{ f : \mathbb{R} \longrightarrow \mathbb{R} \,\Big|\, \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) f(v) |u - v|^{2H-2}\, du\, dv < +\infty \Big\}, \tag{2.5}$$

equipped with the norm

$$\|f\|^2_{\Lambda^H} = H(2H-1) \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) f(v) |u - v|^{2H-2}\, du\, dv. \tag{2.6}$$

When $h \in \mathcal{E}$, it is straightforward to check the following isometry property:

$$\mathbb{E}\Big[\Big(\int_{\mathbb{R}} h(x) dZ(x) \Big)^2\Big] = \|h\|^2_{\Lambda^H}.$$

As a consequence, one can define the Wiener integral $\int_{\mathbb{R}} f(x) dZ(x)$ for any $f \in \Lambda^H$ by a usual approximation procedure.

It is well known by now (thanks to [11]) that $(\Lambda^H, \|\cdot\|_{\Lambda^H})$ is a Hilbert space that contains distributions in the sense of Schwartz. To overcome this problem, we shall restrict ourselves to the proper subspace

$$|\Lambda^H| = \Big\{ f : \mathbb{R} \to \mathbb{R} \,\Big|\, \int_{\mathbb{R}} \int_{\mathbb{R}} |f(u) f(v)| |u - v|^{2H-2}\, du\, dv < +\infty \Big\}$$

equipped with the norm

$$\|f\|^2_{|\Lambda^H|} = H(2H-1) \int_{\mathbb{R}} \int_{\mathbb{R}} |f(u) f(v)| |u - v|^{2H-2}\, du\, dv.$$

We then have (see [11], Proposition 4.2)

$$L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \subset L^{1/H}(\mathbb{R}) \subset |\Lambda^H| \subset \Lambda^H. \tag{2.7}$$

Moreover, $(|\Lambda^H|, \|\cdot\|_{|\Lambda^H|})$ is a Banach space in which the set $\mathcal{E}$ is dense. So for $h \in |\Lambda^H|$, we can define

$$\int_{\mathbb{R}} h(x)\,dZ(x) = \lim_{n\to+\infty} \int_{\mathbb{R}} h_n(x)\,dZ(x), \tag{2.8}$$

where (h_n) is any sequence of $\mathcal{E}$ converging to h in $(|\Lambda^H|, \|\cdot\|_{|\Lambda^H|})$; the convergence in (2.8) holds in $L^2(\Omega)$.

For a detailed account of this integration theory, one can refer to [7], [11].

2.4. Some facts about slowly varying functions. Let $L : (0,+\infty) \to (0,+\infty)$ be a slowly varying function at $+\infty$ and $\alpha > 0$. It is well known (see [3], Proposition 1.3.6(v)) that

$$x^\alpha L(x) \to +\infty \quad \text{and} \quad x^{-\alpha} L(x) \to 0$$

as $x \to +\infty$. In particular, one can deduce that

$$\lim_{\varepsilon\downarrow 0} \varepsilon^{1-H} L(1/\varepsilon)^m = 0. \tag{2.9}$$

The following result is known as Potter's theorem (see [3], Theorem 1.5.6(ii)).

THEOREM 2.2. *Let $L : (0,+\infty) \to (0,+\infty)$ be a slowly varying function at $+\infty$ such that it is bounded away from 0 and $+\infty$ on every compact subset of $(0,+\infty)$. Then for any $\delta > 0$ there exists some constant $C = C(\delta)$ such that*

$$\frac{L(y)}{L(x)} \leqslant C \max\{(x/y)^\delta\,,\,(y/x)^\delta\} \quad \textit{for any } x, y \in (0,+\infty).$$

3. PROOFS OF THE MAIN RESULTS

3.1. Proof of Theorem 1.1. First recall that a typical function h in $\mathcal{E}$ has the form

$$h(x) = \sum_{\ell=1}^{n} a_\ell \mathbf{1}_{(t_\ell, t_{\ell+1}]}(x), \quad t_\ell < t_{\ell+1}, \quad a_\ell \in \mathbb{R}, \quad \ell = 1, \dots, n.$$

For such a simple function h, we deduce from Taqqu's theorem (Theorem 2.1) that

$$\begin{aligned} M_h^\varepsilon &= \frac{1}{\varepsilon d(1/\varepsilon)} \int_{\mathbb{R}} q(x/\varepsilon) \sum_{\ell=1}^{n} a_\ell \mathbf{1}_{(t_\ell, t_{\ell+1}]}(x)\,dx \\ &= \sum_{\ell=1}^{n} a_\ell \frac{1}{d(1/\varepsilon)} \Big(\int_0^{t_{\ell+1}/\varepsilon} \Phi\big(g(x)\big)\,dx - \int_0^{t_\ell/\varepsilon} \Phi\big(g(x)\big)\,dx \Big) \\ &\xrightarrow{\varepsilon\to 0} \frac{V_m}{m!} \sum_{\ell=1}^{n} a_\ell [Z(t_{\ell+1}) - Z(t_\ell)] = \frac{V_m}{m!} \int_{\mathbb{R}} h(x)\,dZ(x). \end{aligned}$$

This proves (1.4) for simple functions $h \in \mathcal{E}$.

Let us now consider $h \in C([0,1])$. It is easy to see that there exists a sequence $(h_n) \subset \mathcal{E}$ such that

$$\lim_{n\to+\infty} \|h_n - h\|_\infty = 0.$$

Let us fix a number $\zeta \in (0,1)$ and show the convergence in $L^2(\Omega)$ of $M^\varepsilon_{h_n}$, uniformly in $\varepsilon \in (0,\zeta)$. First, one can write

$$\begin{aligned}
&\sup_{\varepsilon\in(0,\zeta)} \mathbb{E}[|M^\varepsilon_{h_n} - M^\varepsilon_h|^2] \\
&\qquad = \sup_{\varepsilon\in(0,\zeta)} \frac{1}{\varepsilon^2 d(1/\varepsilon)^2}\, \mathbb{E}\Big[\Big|\int_0^1 q(x/\varepsilon)[h_n(x) - h(x)]\,dx\Big|^2\Big] \\
&\qquad \leqslant \|h_n - h\|^2_\infty \sup_{\varepsilon\in(0,\zeta)} \frac{1}{\varepsilon^2 d(1/\varepsilon)^2} \int_{\mathbb{R}^2\setminus D} \left|R_q\left(\frac{y-x}{\varepsilon}\right)\right| dx\,dy,
\end{aligned}$$

where $D = \{(x,y) \in [0,1]^2 : x = y\}$ is a negligible subset of $\mathbb{R}^2$. By (2.2),

$$\left|R_q\left(\frac{y-x}{\varepsilon}\right)\right| \leqslant \mathrm{Cst}\, L\left(\left|\frac{y-x}{\varepsilon}\right|\right)^{2m} \left|\frac{y-x}{\varepsilon}\right|^{-2(1-H)} \qquad \text{for all } (x,y)\in\mathbb{R}^2\setminus D.$$

Second, with $\beta > 0$ small enough such that $2m\beta + 2(1-H) \in (0,1)$, we have

$$\begin{aligned}
&\sup_{\varepsilon\in(0,\zeta)} \frac{1}{X(\varepsilon)^2} \int_{[0,1]^2\setminus D} \left|R_q\left(\frac{y-x}{\varepsilon}\right)\right| dx\,dy \\
&\qquad \leqslant \mathrm{Cst} \sup_{\varepsilon\in(0,\zeta)} \int_{[0,1]^2\setminus D} \left\{\frac{L\big(|(x-y)/\varepsilon|\big)}{L(1/\varepsilon)}\right\}^{2m} |x-y|^{-2(1-H)}\,dx\,dy \\
&\qquad \leqslant \mathrm{Cst} \int_{[0,1]^2\setminus D} |x-y|^{-2m\beta-2(1-H)}\,dx\,dy \\
&\qquad < +\infty,
\end{aligned} \tag{3.1}$$

where the second inequality follows from Theorem 2.2. It is now clear that, indeed,

$$\lim_{n\to+\infty} \sup_{\varepsilon\in(0,\zeta)} \mathbb{E}\big[|M^\varepsilon_{h_n} - M^\varepsilon_h|^2\big] = 0. \tag{3.2}$$

To conclude, let $d(\cdot,\cdot)$ denote any distance metrizing the convergence in distribution between real-valued random variables (for instance, the Fortet–Mourier distance). For $h \in C([0,1])$ and $(h_n) \subset \mathcal{E}$ converging to h, one can write, for any $\varepsilon > 0$ and $n \in \mathbb{N}$:

$$d(M^\varepsilon_h, M^0_h) \leqslant d(M^\varepsilon_h, M^\varepsilon_{h_n}) + d(M^\varepsilon_{h_n}, M^0_{h_n}) + d(M^0_{h_n}, M^0_h).$$

Fix $\eta > 0$. By (3.2), one can choose n big enough so that, for any $\varepsilon \in (0, \zeta)$, both $d(M_h^\varepsilon, M_{h_n}^\varepsilon)$ and $d(M_{h_n}^0, M_h^0)$ are less than $\eta/3$. It remains to choose $\varepsilon > 0$ small enough so that $d(M_{h_n}^\varepsilon, M_{h_n}^0)$ is less than $\eta/3$ (by (1.4) for the simple function $h_n \in \mathcal{E}$), to conclude that (1.4) holds true for any continuous function h.

REMARK 3.1. Clearly, the above result still holds true for any function h that is continuous except at finitely many points. Note also that the function $\Phi \in \mathscr{G}_m$ is not necessarily bounded in Theorem 1.1.

3.2. Proof of Theorem 1.2. The proof is divided into five steps. We write

$$X(\varepsilon) = \varepsilon d(1/\varepsilon) = \sqrt{\frac{m!}{H(2H-1)}}\varepsilon^{1-H} L(1/\varepsilon)^m.$$

(a) *Preparation*. Following [4], especially identities (5.1) and (5.19) therein, we first rewrite the rescaled corrector as follows:

$$\frac{u^\varepsilon(x) - \bar{u}(x)}{X(\varepsilon)} = \mathcal{U}^\varepsilon(x) + \underbrace{\frac{1}{X(\varepsilon)} r^\varepsilon(x) + \frac{1}{X(\varepsilon)} \rho^\varepsilon \frac{x}{a^*}}_{=:\mathcal{R}^\varepsilon(x)}, \tag{3.3}$$

where

$$\mathcal{U}^\varepsilon(x) = \frac{1}{X(\varepsilon)} \int_{\mathbb{R}} F(x,y) q(y/\varepsilon)\, dy,$$

$$r^\varepsilon(x) = (c^\varepsilon - c^*) \int_0^x q(y/\varepsilon)\, dy,$$

and

$$\rho^\varepsilon := \frac{a^*}{\int_0^1 a(y/\varepsilon)^{-1}\, dy} \Big[\big(a^* b + \int_0^1 F(y) dy \big) \big(\int_0^1 q(y/\varepsilon)\, dy \big)^2 - \int_0^1 F(y) q(y/\varepsilon)\, dy \int_0^1 q(y/\varepsilon)\, dy \Big].$$

Now, let us first show the weak convergence of $\mathcal{U}^\varepsilon$ to $\mathcal{U}$ in $C([0,1])$ and then prove that $\mathcal{R}^\varepsilon$ is a remainder. To prove the first claim, we start by establishing the f.d.d. convergence and then prove the tightness.

(b) *Convergence of finite-dimensional distributions of* $\mathcal{U}^\varepsilon$. For $x_1, \ldots, x_n \in \mathbb{R}$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ ($n \geqslant 1$), we have

$$\sum_{k=1}^n \lambda_k \mathcal{U}^\varepsilon(x_k) = \frac{1}{X(\varepsilon)} \int_{\mathbb{R}} \sum_{k=1}^n \lambda_k F(x_k, y) q(y/\varepsilon)\, dy.$$

Note that the function $\sum_{k=1}^n \lambda_k F(x_k,\cdot)$ has at most finitely many discontinuities. Thus, Theorem 1.1 and Remark 3.1 imply that $\sum_{k=1}^n \lambda_k \mathcal{U}^\varepsilon(x_k)$ converges in distribution to $\sum_{k=1}^n \lambda_k \mathcal{U}(x_k)$, yielding the desired convergence of finite-dimensional distributions.

(c) *Tightness of* $\mathcal{U}^\varepsilon$. We check Kolmogorov's criterion ([6], Corollary 16.9). First observe that $\mathcal{U}^\varepsilon(0)=0$. Now, fix $0 \leqslant u < v \leqslant 1$, and set $F_1(y) = c^* - F(y)$ and $F_2(y) = F(y) - \int_0^1 F(t)\,dt - a^*b$, so that $F(x,y) = F_1(y)\mathbf{1}_{[0,x]}(y) + xF_2(y)\mathbf{1}_{[0,1]}(y)$. Then

$$
\begin{aligned}
(3.4)\quad & \mathbb{E}\big(|\mathcal{U}^\varepsilon(u)-\mathcal{U}^\varepsilon(v)|^2\big) \\
&= \mathbb{E}\bigg[\frac{1}{X(\varepsilon)^2}\Big|\int_0^1 \mathbf{1}_{(u,v]}(y)q(y/\varepsilon)F_1(y)\,dy + (v-u)\int_0^1 q(y/\varepsilon)F_2(y)\,dy\Big|^2\bigg] \\
&\leqslant \frac{2}{X(\varepsilon)^2}\mathbb{E}\Big[\Big|\int_0^1 \mathbf{1}_{(u,v]}(y)q(y/\varepsilon)F_1(y)\,dy\Big|^2 + \Big|(v-u)\int_0^1 q(y/\varepsilon)F_2(y)\,dy\Big|^2\Big] \\
&\leqslant \frac{2}{X(\varepsilon)^2}\int_u^v\int_u^v F_1(x)F_1(y)R_q\bigg(\frac{y-x}{\varepsilon}\bigg)\,dx\,dy \\
&\qquad\qquad + \frac{2(v-u)^2}{X(\varepsilon)^2}\int_0^1\int_0^1 F_2(x)F_2(y)R_q\bigg(\frac{y-x}{\varepsilon}\bigg)\,dx\,dy.
\end{aligned}
$$

Note that F_2 is bounded on $[0,1]$. Therefore, as far as the second term in the last inequality in (3.4) is concerned, one can write, using Potter's theorem as in the proof of Theorem 1.1,

$$
(3.5)\qquad \sup_{\varepsilon\in(0,\zeta)}\bigg|\frac{(v-u)^2}{X(\varepsilon)^2}\int_0^1\int_0^1 F_2(x)F_2(y)R_q\bigg(\frac{y-x}{\varepsilon}\bigg)\,dx\,dy\bigg| \leqslant \mathrm{Cst}(v-u)^2.
$$

Now, let us consider the first term in the last inequality in (3.4). Similarly,

$$
\begin{aligned}
(3.6)\quad & \sup_{\varepsilon\in(0,\zeta)}\frac{1}{X(\varepsilon)^2}\bigg|\int_u^v\int_u^v F_1(x)F_1(y)R_q\bigg(\frac{y-x}{\varepsilon}\bigg)\,dx\,dy\bigg| \\
&\leqslant \mathrm{Cst}\sup_{\varepsilon\in(0,\zeta)}\frac{1}{X(\varepsilon)^2}\int_u^v\int_u^v\bigg|R_q\bigg(\frac{y-x}{\varepsilon}\bigg)\bigg|dx\,dy \quad \text{(since } F_1 \text{ is bounded)} \\
&\leqslant \mathrm{Cst}\sup_{\varepsilon\in(0,\zeta)}\frac{1}{L(1/\varepsilon)^{2m}}\int_u^v\int_u^v L(|y-x|/\varepsilon)^{2m}\frac{dx\,dy}{|y-x|^{2(1-H)}} \\
&\leqslant \mathrm{Cst}\int_u^v\int_u^v |y-x|^{-2(1-H)-2m\beta}\,dy\,dx \quad \text{(as in (3.1))} \\
&= \mathrm{Cst}(v-u)^{2-2m\beta-2(1-H)}.
\end{aligned}
$$

Since $2-2m(1-H_0)-2m\beta > 1$, this proves the tightness of $(\mathcal{U}^\varepsilon)_\varepsilon$ by means of the usual Kolmogorov's criterion.

(d) *Control on the remainder term $\mathcal{R}^\varepsilon$ in* (3.3). We shall prove that the process $\mathcal{R}^\varepsilon$ converges in probability to zero in $C([0,1])$. First we claim that if $G \in C([0,1])$, then there exists some constant $C = C(G)$ such that

$$\sup_{x\in[0,1]} \mathbb{E}\big[\big(\int_0^x q(y/\varepsilon)G(y)\,dy\big)^2\big] \leqslant C\,X(\varepsilon)^2. \tag{3.7}$$

Indeed, the same argument we used for obtaining (3.5) works here as well, so we get

$$\begin{aligned}
&\sup_{x\in[0,1]} \mathbb{E}\big[\big(\int_0^x q(y/\varepsilon)G(y)\,dy\big)^2\big] \\
&\qquad \leqslant \|G\|_\infty^2 \int_{[0,1]^2} \big|R_q(|y-z|/\varepsilon)\big|\,dy\,dz \\
&\qquad \leqslant \|G\|_\infty^2 X(\varepsilon)^2 \bigg(\sup_{\varepsilon\in(0,\zeta)} \frac{1}{X(\varepsilon)^2} \int_{[0,1]^2} \big|R_q(|y-z|/\varepsilon)\big|\,dy\,dz\bigg) \\
&\qquad \leqslant \mathrm{Cst} X(\varepsilon)^2,
\end{aligned}$$

where the last inequality follows from (3.1).

Now, let us consider $\mathcal{R}^\varepsilon$:

(i) Due to the explicit expression of ρ^ε, it follows from (3.7), the fact that a is bounded from below and the Cauchy–Schwarz inequalities that

$$\begin{aligned}
&\mathbb{E}[|\rho^\varepsilon|] \\
&\leqslant \mathrm{Cst}\big\{\big\|\int_0^1 q(y/\varepsilon)\,dy\big\|_{L^2(\Omega)}^2 + \big\|\int_0^1 F(y)q(y/\varepsilon)\,dy\big\|_{L^2(\Omega)}\big\|\int_0^1 q(y/\varepsilon)\,dy\big\|_{L^2(\Omega)}\big\} \\
&\leqslant \mathrm{Cst}\,X(\varepsilon)^2.
\end{aligned}$$

(ii) Observe that

$$\begin{aligned}
c^\varepsilon - c^* &= a^* \int_0^1 \big(F(y) - \int_0^1 F(t)\,dt - ba^*\big)q(y/\varepsilon)\,dy + \rho^\varepsilon \\
&=: \int_0^1 \widehat{F}(y)q(y/\varepsilon)\,dy + \rho^\varepsilon.
\end{aligned}$$

Then

$$\begin{aligned}
&\sup_{x\in[0,1]} \mathbb{E}[|r^\varepsilon(x)|] = \sup_{x\in[0,1]} \mathbb{E}\big[\big|(c^\varepsilon - c^*)\int_0^x q(y/\varepsilon)\,dy\big|\big] \\
&\leqslant \sup_{x\in[0,1]} \mathbb{E}\big[\big|\int_0^1 \widehat{F}(y)q(y/\varepsilon)\,dy \int_0^x q(y/\varepsilon)\,dy\big|\big] + \mathrm{Cst}\,\mathbb{E}[|\rho^\varepsilon|] \leqslant \mathrm{Cst}\,X(\varepsilon)^2.
\end{aligned}$$

Therefore, as $\varepsilon \to 0$ we have, by (2.9),

$$\sup_{x\in[0,1]} \mathbb{E}[|\mathcal{R}^\varepsilon(x)|] \leqslant \mathrm{Cst} X(\varepsilon) \to 0.$$

In particular, $\{\mathcal{R}^\varepsilon(x), x \in [0,1]\}$ converges to zero in the sense of finite-dimensional distributions. Now, let us check the tightness of $(\mathcal{R}^\varepsilon)_\varepsilon$. Note that $\mathcal{R}^\varepsilon(0) = 0$ and that, for $0 \leqslant u < v \leqslant 1$,

$$\begin{aligned}
&\big\|\mathcal{R}^\varepsilon(u) - \mathcal{R}^\varepsilon(v)\big\|^2_{L^2(\Omega)} \\
&\quad\leqslant \frac{2}{X(\varepsilon)^2}\left\{\big\|r^\varepsilon(u) - r^\varepsilon(v)\big\|^2_{L^2(\Omega)} + \frac{2(u-v)^2}{|a^*|^2}\mathbb{E}[|\rho^\varepsilon|^2]\right\} \\
&\quad\leqslant \frac{2}{X(\varepsilon)^2}\big\|r^\varepsilon(u) - r^\varepsilon(v)\big\|^2_{L^2(\Omega)} \\
&\qquad + \mathrm{Cst}\frac{(u-v)^2}{X(\varepsilon)^2}\mathbb{E}[|\rho^\varepsilon|] \quad \text{(since } \rho^\varepsilon \text{ is uniformly bounded)} \\
&\quad\leqslant \frac{2}{X(\varepsilon)^2}\big\|r^\varepsilon(u) - r^\varepsilon(v)\big\|^2_{L^2(\Omega)} + \mathrm{Cst}(u-v)^2 \quad \text{(by point (i) above)} \\
&\quad\leqslant \mathrm{Cst}\frac{1}{X(\varepsilon)^2}\int\limits_{[u,v]^2} \big|R\big((y-z)/\varepsilon\big)\big|\,dy\,dz \\
&\qquad + \mathrm{Cst}(u-v)^2 \quad \text{(since } c^\varepsilon - c^* \text{ is uniformly bounded)} \\
&\quad\leqslant \mathrm{Cst}(v-u)^{2-2(1-H)-2m\beta} + \mathrm{Cst}(v-u)^2,
\end{aligned}$$

where the last inequality follows from the same arguments as in (3.6). Therefore, $\mathcal{R}^\varepsilon$ converges in distribution to zero, as $\varepsilon \downarrow 0$, so it converges in probability to zero.

(e) *Conclusion*. Combining the results of (a) to (d), we conclude the proof of Theorem 1.2 by evoking the Slutsky lemma.

REFERENCES

[1] G. Bal, J. Garnier, S. Motsch, and V. Perrier, *Random integrals and correctors in homogenization*, Asymptot. Anal. 59 (1–2) (2008), pp. 1–26.

[2] G. Bal and Y. Gu, *Limiting models for equations with large random potential: A review*, Commun. Math. Sci. 13 (3) (2015), pp. 729–748.

[3] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*, Cambridge University Press, Cambridge 1989.

[4] Y. Gu and G. Bal, *Random homogenization and convergence to integrals with respect to the Rosenblatt process*, J. Differential Equations 253 (4) (2012), pp. 1069–1087.

[5] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik, *Homogenization of Differential Operators and Integral Functionals*, Springer, Berlin 1994.

[6] O. Kallenberg, *Foundations of Modern Probability*, second edition, Springer, New York 2002.

[7] M. Maejima and C. A. Tudor, *Wiener integrals with respect to the Hermite process and a non-central limit theorem*, Stoch. Anal. Appl. 25 (5) (2007), pp. 1043–1056.

[8] J. Mourrat and J. Nolen, *Scaling limit of the corrector in stochastic homogenization*, Ann. Appl. Probab. 27 (2) (2017), pp. 944–959.
[9] J. Mourrat and F. Otto, *Correlation structure of the corrector in stochastic homogenization*, Ann. Probab. 44 (5) (2016), pp. 3207–3233.
[10] G. C. Papanicolaou and S. R. S. Varadhan, *Boundary value problems with rapidly oscillating random coefficients*, in: *Proceedings of the Colloquium on Random Fields (Esztergom, Hungary, 1979)*, Colloq. Math. Soc. János Bolyai 27, North Holland, 1981, pp. 835–873.
[11] V. Pipiras and M. S. Taqqu, *Integration questions related to fractional Brownian motion*, Probab. Theory Related Fields 118 (2) (2000), pp. 251–291.
[12] M. S. Taqqu, *Convergence of integrated processes of arbitrary Hermite rank*, Z. Wahrsch. Verw. Gebiete 50 (1) (1979), pp. 53–83.

Atef Lechiheb
Université de Tunis El Manar
Faculté des sciences de Tunis
LR11ES13 Laboratoire d'Analyse stochastique
et applications
2092, Tunis, Tunisie
E-mail: atef.lechiheb@gmail.com

Ivan Nourdin
Université du Luxembourg
Unité de Recherche en Mathématiques
Maison du Nombre
6, avenue de la Fonte
L-4364 Esch-sur-Alzette
Grand Duchy of Luxembourg
E-mail: ivan.nourdin@uni.lu

Guangqu Zheng
University of Kansas
Mathematics Department
Snow Hall, 1460 Jayhawk Blvd
Lawrence, Kansas 66045, USA
E-mail: gzheng90@ku.edu

Ezzedine Haouala
Université de Tunis El Manar
Faculté des sciences de Tunis
LR11ES13 Laboratoire d'Analyse stochastique
et applications
2092, Tunis, Tunisie
E-mail: ezdine.haouala@fst.rnu.tn

Received on 14.7.2016;
revised version on 18.2.2017

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 287–298
doi:10.19195/0208-4147.38.2.3

AN EQUIVALENT CHARACTERIZATION OF WEAK BMO MARTINGALE SPACES

BY

DEJIAN ZHOU (CHANGSHA), WEIWEI LI (CHANGSHA), AND YONG JIAO* (CHANGSHA)

Abstract. In this paper, we give an equivalent characterization of weak BMO martingale spaces due to Ferenc Weisz (1998).

2010 AMS Mathematics Subject Classification: Primary: 60G42; Secondary: 60G46.

Key words and phrases: Weak BMO space, martingale, John–Nirenberg inequality.

1. INTRODUCTION

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, and $\{\mathcal{F}_n\}_{n\geqslant 0}$ be an increasing sequence of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F} = \sigma\big(\bigcup_{n\geqslant 0}\mathcal{F}_n\big)$. The expectation operator and the conditional expectation operator relative to $\mathcal{F}_n$ are denoted by $\mathbb{E}$ and $\mathbb{E}_n$, respectively. A sequence $f = (f_n)_{n\geqslant 0}$ of random variables such that f_n is $\mathcal{F}_n$-measurable is said to be a *martingale* if $\mathbb{E}(|f_n|) < \infty$ and $\mathbb{E}_n(f_{n+1}) = f_n$ for every $n \geqslant 0$.

The study of the space BMO (Bounded Mean Oscillation) began with the establishment of the so-called John–Nirenberg theorem in [11]. Basing mainly on the duality and something else, the space BMO plays a remarkable role both in classical analysis and martingale theory. For example, BMO is a good space in operator actions (see e.g. [14], Chapter 4). And the martingale space $BMO_r(\alpha)$ was first introduced by Herz in [4] as the dual of H_p^s $(0 < p \leqslant 1)$ associated with the dyadic filtration (see Example 2.1 below). With the help of atomic decomposition, Weisz extended this result in [15] to a general case. Let $\mathcal{T}$ be the set of all stopping times with respect to $\{\mathcal{F}_n\}_{n\geqslant 0}$. The martingale space $BMO_r(\alpha)$ ([16], p. 8; or [15]) for $1 \leqslant r < \infty$ and $\alpha \geqslant 0$ is defined as

$$BMO_r(\alpha) = \{f = (f_n)_{n\geqslant 0} : \|f\|_{BMO_r(\alpha)} < \infty\},$$

* Research supported by NSFC (11471337) and Hunan Province Natural Science Foundation (14JJ1004).

where

$$\|f\|_{BMO_r(\alpha)} = \sup_{\nu\in\mathcal{T}} \mathbb{P}(\nu<\infty)^{-1/r-\alpha}\|f-f^{\nu}\|_r.$$

We present two well-known results (see [16] or [15]). If $0<p\leqslant 1$ and $\alpha=\frac{1}{p}-1$, then $BMO_2(\alpha)$ is the dual space of the Hardy space H_p^s. If the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 1}$ is regular, then $BMO_r(\alpha)=BMO_1(\alpha)$. And recently, Yi et al. proved in [18] that $BMO_E(\alpha)=BMO_1(\alpha)$, where $\alpha=0$ and E is a rearrangement invariant Banach function space.

In the present paper, we consider a weak BMO martingale space. To characterize the dual of the weak Hardy martingale space $H_{p,\infty}^s$, Weisz in [17] first introduced and studied the weak BMO martingale space. Let us recall the definition. We also refer the reader to [12] and [13] for some new results related to weak BMO martingales spaces.

DEFINITION 1.1. Let $1\leqslant r<\infty, \alpha r+1>0$. The space $w\mathcal{BMO}_r(\alpha)$ is defined as the set of all martingales $f\in L_r$ with the norm

$$\|f\|_{w\mathcal{BMO}_r(\alpha)} = \int_0^{\infty} \frac{t_\alpha^r(x)}{x}dx<\infty,$$

where

$$t_\alpha^r(x) = x^{-1/r-\alpha} \sup_{\nu\in\mathcal{T}:P(\nu<\infty)\leqslant x} \|f-f^{\nu}\|_r.$$

In the very recent paper [8], the generalized BMO martingale space is introduced as the dual of Hardy–Lorentz martingale space. Strongly motivated by [8], Definition 1.1, we introduce the following new weak BMO martingale space by stopping time sequences.

DEFINITION 1.2. Let $1\leqslant r<\infty$ and $\alpha\geqslant 0$. The weak BMO martingale space $wBMO_r(\alpha)$ is defined by

$$wBMO_r(\alpha) = \{f\in L_r : \|f\|_{wBMO_r(\alpha)}<\infty\},$$

where

$$\|f\|_{wBMO_r(\alpha)} = \sup \frac{\sum\limits_{k\in\mathbb{Z}} 2^k\mathbb{P}(\nu_k<\infty)^{1-1/r}\|f-f^{\nu_k}\|_r}{\sup_k 2^k\mathbb{P}(\nu_k<\infty)^{1+\alpha}}$$

and the supremum is taken over all stopping time sequences $\{\nu_k\}_{k\in\mathbb{Z}}$ such that $2^k\mathbb{P}(\nu_k<\infty)^{1+\alpha}\in\ell_\infty$.

It is a very natural question: what is the relationship between $w\mathcal{BMO}_r(\alpha)$ and $wBMO_r(\alpha)$? The paper fully answers this question. Our main result can be described as follows. We simply put $w\mathcal{BMO}=w\mathcal{BMO}(0)$ and $wBMO=wBMO(0)$.

THEOREM 1.1. *Let* $1 \leqslant r < \infty$ *and* $\alpha \geqslant 0$. *If the stochastic basis* $\{\mathcal{F}_n\}_{n\geqslant 0}$ *is regular*, *then*

$$w\mathcal{BMO}_r(\alpha) = wBMO_r(\alpha)$$

with equivalent norms. In particular,

$$w\mathcal{BMO}_r = wBMO_r$$

with equivalent norms.

In this paper, the set of integers and the set of nonnegative integers are always denoted by $\mathbb{Z}$ and $\mathbb{N}$, respectively. We use C to denote a positive constant which may vary from line to line. The symbol $\subset$ means the continuous embedding.

2. PRELIMINARIES

Firstly, we give the definition of Lorentz spaces. We denote by $L_0(\Omega, \mathcal{F}, \mathbb{P})$, or simply $L_0(\Omega)$, the space of all measurable functions on $(\Omega, \mathcal{F}, \mathbb{P})$. For any $f \in L_0(\Omega)$, we define the distribution function of f by

$$\lambda_s(f) = \mathbb{P}\big(\{\omega \in \Omega : |f(\omega)| > s\}\big), \quad s \geqslant 0.$$

Moreover, denote by $\mu_t(f)$ the decreasing rearrangement of f defined by

$$\mu_t(f) = \inf\{s \geqslant 0 : \lambda_s(f) \leqslant t\}, \quad t \geqslant 0,$$

with the convention that $\inf \emptyset = \infty$.

DEFINITION 2.1. Let $0 < p < \infty$ and $0 < q \leqslant \infty$. Then, the *Lorentz space* $L_{p,q}(\Omega)$ consists of measurable functions such that $\|f\|_{p,q} < \infty$, where

$$\|f\|_{p,q} = \left[\int_0^\infty \big(t^{1/p}\mu_t(f)\big)^q \frac{dt}{t}\right]^{1/q}, \quad 0 < q < \infty,$$

and

$$\|f\|_{p,\infty} = \sup_{0\leqslant t<\infty} t^{1/p}\mu_t(f), \quad q = \infty.$$

REMARK 2.1. *We refer the reader to* [2] *for the following basic properties.*

(1) *If* $p = q$, *then* $L_{p,q}(\Omega)$ *becomes* $L_p(\Omega)$.

(2) *If* $0 < p_1 \leqslant p_2 < \infty$ *and* $0 < q \leqslant \infty$, *then* $\|f\|_{p_1,q} \leqslant C\|f\|_{p_2,q}$, *where* C *depends on* p_1, p_2 *and* q. *This is due to* $\mathbb{P}(\Omega) = 1$.

(3) *If* $0 < p < \infty$ *and* $0 < q_1 \leqslant q_2 \leqslant \infty$, *then* $\|f\|_{p,q_2} \leqslant C\|f\|_{p,q_1}$, *where* C *depends on* q_1, q_2 *and* p.

Denote by $\mathcal{M}$ the set of all martingales $f = (f_n)_{n\geqslant 0}$ relative to $\{\mathcal{F}_n\}_{n\geqslant 0}$ such that $f_0 = 0$. For $f \in \mathcal{M}$, denote its martingale difference by $d_n f = f_n - f_{n-1}$ ($n \geqslant 0$, with the convention $f_{-1} = 0$). Then the maximal function and the conditional quadratic variation of a martingale f are respectively defined by

$$f_n^* = \sup_{0\leqslant i\leqslant n} |f_i|, \quad f^* = \sup_{n\geqslant 0} |f_n|,$$

$$s_n(f) = \Big(\sum_{i=1}^{n} \mathbb{E}_{i-1}|d_i f|^2\Big)^{1/2}, \quad s(f) = \Big(\sum_{i=1}^{\infty} \mathbb{E}_{i-1}|d_i f|^2\Big)^{1/2}.$$

Then we define *martingale Hardy–Lorentz spaces* as follows.

DEFINITION 2.2. Let $0 < p < \infty$ and $0 < q \leqslant \infty$. Define

$$H^*_{p,q} = \{f \in \mathcal{M} : \|f\|_{H^*_{p,q}} = \|f^*\|_{p,q} < \infty\},$$
$$H^s_{p,q} = \{f \in \mathcal{M} : \|f\|_{H^s_{p,q}} = \|s(f)\|_{p,q} < \infty\}.$$

If $p = q$, then the martingale Hardy–Lorentz spaces recover the martingale Hardy spaces H^*_p and H^s_p (see [16]).

Recall that the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$ is said to be *regular* if there exists a positive constant $R > 0$ such that

$$f_n \leqslant R f_{n-1}, \quad \forall n > 0, \tag{2.1}$$

holds for all nonnegative martingales $f = (f_n)_{n\geqslant 0}$. Condition (2.1) can be replaced by several other equivalent conditions (see [14], Chapter 7). We refer the reader to [14], p. 265, for examples for regular stochastic basis. Here, we give a special case.

EXAMPLE 2.1. Let $\big((0,1], \mathcal{F}, \mu\big)$ be a probability space such that μ is the Lebesgue measure and subalgebras $\{\mathcal{F}_n\}_{n\geqslant 0}$ are generated as follows:

$$\mathcal{F}_n = \text{a } \sigma\text{-algebra generated by atoms } \left(\frac{j}{2^n}, \frac{j+1}{2^n}\right], j = 0, \ldots, 2^n - 1.$$

Then $\{\mathcal{F}_n\}_{n\geqslant 0}$ is regular. And all martingales with respect to such $\{\mathcal{F}_n\}_{n\geqslant 0}$ are called *dyadic martingales*.

The method of atomic decompositions plays an important role in martingale theory (see, for example, [3]–[5], [16], [17]). The atomic decompositions of Hardy–Lorentz martingale spaces $H^s_{p,q}$ and martingale inequalities are given in [6] and [8]. We also mention that Hardy–Lorentz spaces with variable exponents were investigated very recently in [9] and [10]. Let us first introduce the concept of an atom (see [16], p. 14).

DEFINITION 2.3. Let $0 < p < \infty$ and $p < r \leqslant \infty$. A measurable function a is called a $(1, p, r)$-*atom* (or $(3, p, r)$-*atom*) if there exists a stopping time $\nu \in \mathcal{T}$ such that $a_n = \mathbb{E}_n(a) = 0$ if $\nu \geqslant n$, and

$$\|s(a)\|_r \text{ (or } \|a^*\|_r) \leqslant \mathbb{P}(\nu < \infty)^{1/r-1/p}.$$

REMARK 2.2. *Let $0 < p < r \leqslant \infty$ and $0 < q \leqslant r$. If a is a $(1, p, r)$-atom, then $\|a\|_{H^s_{p,q}} \leqslant C$. Choose p_1, p_2 such that $\frac{1}{p} = \frac{1}{r} + \frac{1}{p_1}$, $\frac{1}{q} = \frac{1}{r} + \frac{1}{q_1}$. By Hölder's inequality, we have (ν is the stopping time corresponding to the atom a)*

$$\begin{aligned}\|a\|_{H^s_{p,q}} &= \|s(a)\chi_{\{\nu<\infty\}}\|_{p,q} \leqslant C\|s(a)\|_{r,r}\|\chi_{\{\nu<\infty\}}\|_{p_1,q_1} \\ &\leqslant C\mathbb{P}(\nu<\infty)^{1/r-1/p}\Big(\int_0^\infty t^{q_1/p_1-1}\chi_{(0,\mathbb{P}(\nu<\infty))}dt\Big)^{1/q_1} \leqslant C.\end{aligned}$$

Similarly, we have $\|a\|_{H^*_{p,q}} \leqslant C$ for a $(3, p, r)$-atom a. If $p = q$, then $C = 1$.

The following result is from [8]. And the result about the Hardy space $H^*_{p,q}$ follows from the combining of Theorem 3.3 and Lemma 5.1 in [8].

THEOREM 2.1. *If $f = (f_n)_{n\geqslant 0} \in H^s_{p,q}$ for $0 < p < \infty, 0 < q \leqslant \infty$, then there exist a sequence $(a^k)_{k\in\mathbb{Z}}$ of $(1, p, \infty)$-atoms and a positive number A satisfying $\mu_k = A \cdot 2^k\mathbb{P}(\nu_k < \infty)^{1/p}$ (where ν_k is the stopping time corresponding to a^k) such that*

$$f_n = \sum_{k\in\mathbb{Z}} \mu_k a^k_n \ \ a.e., \quad n \in \mathbb{N}, \tag{2.2}$$

and

$$\|\{\mu_k\}\|_{l_q} \leqslant C\|f\|_{H^s_{p,q}}.$$

Conversely, if the martingale f has the above decomposition, then $f \in H^s_{p,q}$ and $\|f\|_{H^s_{p,q}} \approx \inf \|\{\mu_k\}\|_{l_q}$, where the infimum is taken over all the above decompositions.

Moreover, if the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$ is regular and if we replace $H^s_{p,q}$, $(1, p, \infty)$-atoms by $H^*_{p,q}$, $(3, p, \infty)$-atoms, then the conclusions above still hold.

LEMMA 2.1 ([1], Lemma 1.2). *Let $0 < p < \infty$ and let the nonnegative sequence $\{\mu_k\}$ be such that $\{2^k\mu_k\} \in l^q, 0 < q \leqslant \infty$. Further, suppose the nonnegative function φ satisfies the following property: there exists $0 < \varepsilon < \min(1, q/p)$ such that, given an arbitrary integer k_0, we have $\varphi \leqslant \psi_{k_0} + \eta_{k_0}$, where ψ_{k_0} and η_{k_0} satisfy*

$$2^{k_0 p}\mathbb{P}(\psi_{k_0} > 2^{k_0})^\varepsilon \leqslant C\sum_{k=-\infty}^{k_0-1}(2^k\mu_k^\varepsilon)^p,$$

$$2^{k_0\varepsilon p}\mathbb{P}(\eta_{k_0} > 2^{k_0}) \leqslant C\sum_{k=k_0}^{\infty}(2^{k\varepsilon}\mu_k)^p.$$

Then $\varphi \in L_{p,q}$ and $\|\varphi\|_{p,q} \leqslant C\|\{2^k\mu_k\}\|_{l_q}$.

3. A JOHN–NIRENBERG THEOREM

In this section, we prove a John–Nirenberg theorem when the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$ is regular. The main idea and method are similar to those of [8]. The following lemma can be found in [5], [16]. In fact, it follows from Theorem 7.14 in [5] and Corollary 5.13 in [16].

LEMMA 3.1. *Suppose that* $0 < q \leqslant \infty$ *and the stochastic basis* $\{\mathcal{F}_n\}_{n\geqslant 0}$ *is regular.*

If $0 < p < \infty$, *then* $H^*_{p,q}$ *and* $H^s_{p,q}$ *are equivalent.*

If $1 < p < \infty$, *then* $H^*_{p,q}$, $H^s_{p,q}$ *and* $L_{p,q}$ *are all equivalent.*

L_p is not dense in $L_{p,\infty}$. This fact is mentioned in [17], p. 143 (see also [2], Remark 1.4.14). Hence, to describe the duality, we need the following definition from [7], Remark 1.7.

DEFINITION 3.1. Let a measurable set $A_k \subset \Omega$ satisfy $\mathbb{P}(A_k) \to 0$ as $k \to \infty$. Define $\mathcal{L}_{p,\infty}$ as the set of all $f \in L_{p,\infty}$ having the absolute continuous quasi-norm defined by

$$\mathcal{L}_{p,\infty} = \{f \in L_{p,\infty} : \lim_{k\to\infty} \|f\chi_{A_k}\|_{p,\infty} = 0\}.$$

$\mathcal{L}_{p,\infty}$ is a closed subspace of $L_{p,\infty}$ and $L_p \subset \mathcal{L}_{p,\infty} \subset L_{p,\infty}$ (see [7]). Now we define

$$\mathcal{H}^s_{p,\infty} = \{f = (f_n)_{n\geqslant 0} : s(f) \in \mathcal{L}_{p,\infty}\},$$

which is a closed subspace of $H^s_{p,\infty}$. Similarly, we define $\mathcal{H}^*_{p,\infty}$.

REMARK 3.1. (1) *According to* [7], *Remark* 2.2, *we can conclude that* $H^s_2 = L_2$ *is dense in* $\mathcal{H}^s_{p,\infty}$.

(2) *If the stochastic basis* $\{\mathcal{F}_n\}_{n\geqslant 0}$ *is regular*, *then*, *by the same argument of Remark* 2.2 *in* [7], L_∞ *is dense in* $\mathcal{H}^*_{p,\infty}$.

LEMMA 3.2. *Let* $0 < p \leqslant 1$. *If the stochastic basis* $\{\mathcal{F}_n\}_{n\geqslant 0}$ *is regular*, *then*

$$(\mathcal{H}^*_{p,\infty})^* = wBMO_1(\alpha), \quad \alpha = \frac{1}{p} - 1.$$

Proof. Let $g \in wBMO_1(\alpha)$. Define

$$\phi_g(f) = \mathbb{E}(fg), \quad f \in L_\infty.$$

Then, by Theorem 2.1, we find that (ν_k is the stopping time corresponding to the

atom a^k for every $k \in \mathbb{Z}$)

$$\begin{aligned}|\phi_g(f)| &\leqslant \sum_{k\in\mathbb{Z}} |\mu_k|\mathbb{E}\big(a^k(g-g^{\nu_k})\big) \leqslant \sum_{k\in\mathbb{Z}} |\mu_k|\|a^k\|_\infty \|g-g^{\nu_k}\|_1 \\ &\leqslant C\sum_{k\in\mathbb{Z}} |\mu_k|\|(a^k)^*\|_\infty \|g-g^{\nu_k}\|_1 \\ &\leqslant C\sum_{k\in\mathbb{Z}} |\mu_k|\mathbb{P}(\nu_k<\infty)^{-1/p}\|g-g^{\nu_k}\|_1 \\ &= C\cdot A\sum_{k\in\mathbb{Z}} 2^k\|g-g^{\nu_k}\|_1.\end{aligned}$$

By the definition of $\|\cdot\|_{wBMO_r(\alpha)}$, we obtain

$$\begin{aligned}|\phi_g(f)| &\leqslant C\cdot A\sup_k 2^k\mathbb{P}(\nu_k<\infty)^{1/p}\|g\|_{wBMO_1(\alpha)} \\ &\leqslant C\|f\|_{H^*_{p,\infty}}\|g\|_{wBMO_1(\alpha)}.\end{aligned}$$

Since the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$ is regular, L_∞ is dense in $\mathcal{H}^*_{p,\infty}$ (see Remark 3.1(2)). Then ϕ_g can be uniquely extended to be a continuous linear functional on $\mathcal{H}^*_{p,\infty}$.

Conversely, let $\phi \in (\mathcal{H}^*_{p,\infty})^*$. Since L_2 is dense in $\mathcal{H}^*_{p,\infty}$ (see Remark 3.1(2)), there exists $g \in L_2 \subset L_1$ such that

$$\phi(f) = \mathbb{E}(fg), \quad f \in L_\infty.$$

Let $\{\nu_k\}_{k\in\mathbb{Z}}$ be a stopping time sequence satisfying $\{2^k\mathbb{P}(\nu_k<\infty)^{1/p}\}_{k\in\mathbb{Z}} \in l_\infty$ and let

$$h_k = \operatorname{sign}(g-g^{\nu_k}), \quad a^k = \frac{1}{2}(h_k - h_k^{\nu_k})\mathbb{P}(\nu_k<\infty)^{-1/p}.$$

Then a^k is a $(3,p,\infty)$-atom. Let $f^N = \sum_{k=-N}^N 2^{k+1}\mathbb{P}(\nu_k<\infty)^{1/p}a^k$, where N is an arbitrary nonnegative integer. By Theorem 2.1, we have $f^N \in H^*_{p,\infty}$ and

$$\|f^N\|_{H^*_{p,\infty}} \leqslant C\sup_k 2^k\mathbb{P}(\nu_k<\infty)^{1/p}.$$

Consequently,

$$\begin{aligned}\sum_{k=-N}^N 2^k\|g-g^{\nu_k}\|_1 &= \sum_{k=-N}^N 2^k\mathbb{E}\big(h_k(g-g^{\nu_k})\big) = \sum_{k=-N}^N 2^k\mathbb{E}\big((h_k-h_k^{\nu_k})g\big) \\ &= \mathbb{E}(f^Ng) = \phi(f^N) \leqslant \|f^N\|_{H^*_{p,\infty}}\|\phi\| \\ &\leqslant C\sup_k 2^k\mathbb{P}(\nu_k<\infty)^{1/p}\|\phi\|.\end{aligned}$$

Thus we have

$$\frac{\sum_{k=-N}^{N} 2^k \|g - g^{\nu_k}\|_1}{\sup_k 2^k \mathbb{P}(\nu_k < \infty)^{1/p}} \leqslant C\|\phi\|.$$

This implies $\|g\|_{wBMO_1(\alpha)} \leqslant C\|\phi\|$. The proof is complete. ∎

LEMMA 3.3. *Let $0 < p \leqslant 1$, $1 < r < \infty$. If the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$ is regular, then*

$$(\mathcal{H}^*_{p,\infty})^* = wBMO_r(\alpha), \quad \alpha = \frac{1}{p} - 1.$$

P r o o f. By Hölder's inequality, we have $\|f\|_{wBMO_1(\alpha)} \leqslant \|f\|_{wBMO_r(\alpha)}$ for any $f \in wBMO_r(\alpha)$. Let $g \in wBMO_r(\alpha) \subset L_r$. We define

$$\phi_g(f) = \mathbb{E}(fg), \quad \forall f \in L_{r'}.$$

Then, by Lemma 3.2, we have

$$|\phi_g(f)| \leqslant C\|f\|_{H^s_{p,\infty}}\|g\|_{wBMO_1(\alpha)} \leqslant C\|f\|_{H^s_{p,\infty}}\|g\|_{wBMO_r(\alpha)}.$$

It follows from Remark 3.1(2) that $L_{r'}$ is dense in $\mathcal{H}^*_{p,\infty}$. Thus ϕ_g can be uniquely extended to be a continuous linear functional on $\mathcal{H}^*_{p,\infty}$.

Conversely, if $\phi \in (\mathcal{H}^*_{p,\infty})^*$, by Doob's maximal inequality, we have $L_{r'} = H^*_{r',r'} \subset \mathcal{H}^*_{p,\infty}$. Then $(\mathcal{H}^*_{p,\infty})^* \subset (L_{r'})^* = L_r$. Thus there exists $g \in L_r$ such that

$$\phi(f) = \phi_g(f) = \mathbb{E}(fg), \quad \forall f \in L_{r'}.$$

Let $\{\nu_k\}_{k\in\mathbb{Z}}$ be a stopping time sequence such that $\{2^k\mathbb{P}(\nu_k < \infty)^{1/p}\}_{k\in\mathbb{Z}} \in l_\infty$ and N be an arbitrary nonnegative integer. Let

$$h_k = \frac{|g - g^{\nu_k}|^{r-1}\operatorname{sign}(g - g^{\nu_k})}{\|g - g^{\nu_k}\|_r^{r-1}}, \quad f = \sum_{k=-N}^{N} 2^k \mathbb{P}(\nu_k < \infty)^{1/r'}(h_k - h_k^{\nu_k}).$$

For an arbitrary integer k_0 which satisfies $-N \leqslant k_0 \leqslant N$ (for $k_0 \leqslant -N$, let $G = 0$ and $H = f$; for $k_0 > N$, let $H = 0$ and $G = f$), let

$$f = G + H,$$

where

$$G = \sum_{k=-N}^{k_0-1} 2^k \mathbb{P}(\nu_k < \infty)^{1/r'}(h_k - h_k^{\nu_k})$$

and

$$H = \sum_{k=k_0}^{N} 2^k \mathbb{P}(\nu_k < \infty)^{1/r'}(h_k - h_k^{\nu_k}).$$

Obviously, $\|h_k\|_{r'} = 1$, and $\|G\|_{r'} \leqslant 2\sum_{k=-N}^{k_0-1} 2^k \mathbb{P}(\nu_k < \infty)^{1/r'}$. By the sublinearity of the maximal operator $*$, we have $f^* \leqslant G^* + H^*$. Let $\varepsilon = p/r'$ $(0<\varepsilon<1)$. By Doob's maximal inequality, we have

$$\begin{aligned}\mathbb{P}(G^* > 2^{k_0}) &\leqslant \frac{1}{2^{k_0 r'}}\|G^*\|_{r'}^{r'} \leqslant C\frac{1}{2^{k_0 r'}}\|G\|_{r'}^{r'} \\ &\leqslant C\frac{1}{2^{k_0 r'}}\Big(\sum_{k=-N}^{k_0-1} 2^k \mathbb{P}(\nu_k < \infty)^{1/r'}\Big)^{r'}.\end{aligned}$$

On the other hand, $\{H^* > 0\} \subset \bigcup_{k=k_0}^{N}\{\nu_k < \infty\}$. Then, for each $0 < \varepsilon < 1$, we have

$$\begin{aligned}2^{k_0\varepsilon p}\mathbb{P}(H^* > 2^{k_0}) &\leqslant 2^{k_0\varepsilon p}\mathbb{P}(H^* > 0) \leqslant 2^{k_0\varepsilon p}\sum_{k=k_0}^{N}\mathbb{P}(\nu_k < \infty) \\ &\leqslant \sum_{k=k_0}^{N} 2^{k\varepsilon p}\mathbb{P}(\nu_k < \infty) = \sum_{k=k_0}^{N}\big(2^{k\varepsilon}\mathbb{P}(\nu_k < \infty)^{1/p}\big)^p \\ &\leqslant \sum_{k=k_0}^{\infty}\big(2^{k\varepsilon}\mathbb{P}(\nu_k < \infty)^{1/p}\big)^p.\end{aligned}$$

By Lemma 2.1, we have $f^* \in L_{p,\infty}$ and $\|f^*\|_{p,\infty} \leqslant C\|\{2^k\mathbb{P}(\nu_k < \infty)^{1/p}\}_{k\in\mathbb{Z}}\|_{l_\infty}$. Thus, $f \in H^*_{p,\infty}$ and

$$\|f\|_{H^*_{p,\infty}} \leqslant C\sup_k 2^k\mathbb{P}(\nu_k < \infty)^{1/p}.$$

Consequently,

$$\begin{aligned}\sum_{k=-N}^{N} 2^k\mathbb{P}(\nu_k < \infty)^{1-1/r}\|g - g^{\nu_k}\|_r &= \sum_{k=-N}^{N} 2^k\mathbb{P}(\nu_k < \infty)^{1/r'}\mathbb{E}\big(h_k(g - g^{\nu_k})\big) \\ &= \sum_{k=-N}^{N} 2^k\mathbb{P}(\nu_k < \infty)^{1/r'}\mathbb{E}\big((h_k - h_k^{\nu_k})g\big) \\ &= \mathbb{E}(fg) = \varphi(f) \leqslant \|f\|_{H^*_{p,q}}\|\varphi\| \\ &\leqslant C\sup_k 2^k\mathbb{P}(\nu_k < \infty)^{1/p}.\end{aligned}$$

Thus we obtain

$$\frac{\sum_{k=-N}^{N} 2^k\mathbb{P}(\nu_k < \infty)^{1-1/r}\|g - g^{\nu_k}\|_r}{\sup_k 2^k\mathbb{P}(\nu_k < \infty)^{1/p}} \leqslant C\|\varphi\|.$$

Taking $N \to \infty$ and the supremum over all stopping time sequences satisfying $\{2^k\mathbb{P}(\nu_k < \infty)^{1/p}\}_{k\in\mathbb{Z}} \in l_\infty$, we get $\|g\|_{wBMO_r(\alpha)} \leqslant C\|\varphi\|$. ∎

Now we formulate the weak version of the John–Nirenberg theorem, which directly follows from Lemmas 3.2 and 3.3.

THEOREM 3.1. *Let $\alpha \geqslant 0$ and $1 \leqslant r < \infty$. If the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$ is regular, then*

$$wBMO_r(\alpha) = wBMO_1(\alpha)$$

with equivalent norms.

According to Lemma 3.1, Lemma 3.3 holds if we replace $\mathcal{H}^*_{p,\infty}$ by $\mathcal{H}^s_{p,\infty}$. Without regularity of stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$, we also get a duality result.

PROPOSITION 3.1. *Let $0 < p \leqslant 1$. Then $\left(\mathcal{H}^s_{p,\infty}\right)^* = wBMO_2(\alpha)$ with $\alpha = 1/p - 1$.*

P r o o f. Note that $H^s_2 = L_2$ is dense in $\mathcal{H}^s_{p,\infty}$ by Remark 3.1(1). The first part of the proof is similar to that of Lemma 3.2, and the converse part is similar to that of Lemma 3.3 with $r = 2$. We omit the proof. ∎

4. PROOF OF THE MAIN THEOREM

In this section we complete the proof of Theorem 1.1.

Let $\overline{H}^s_{p,\infty}$ be the $H^s_{p,\infty}$ closure of H^s_∞. Since $H^s_\infty \subset H^s_2 = L_2$, using Remark 3.1(1), we have $\overline{H}^s_{p,\infty} \subset \mathcal{H}^s_{p,\infty}$. Then $(\mathcal{H}^s_{p,\infty})^* \subset (\overline{H}^s_{p,\infty})^*$.

LEMMA 4.1 ([17], Corollary 6). *Let $0 < p < 2$. Then the dual space of $\overline{H}^s_{p,\infty}$ is $w\mathcal{BMO}_2(\alpha)$ with $\alpha = 1/p - 1$.*

LEMMA 4.2 ([17], Corollary 8). *Suppose that the stochastic basis $\{\mathcal{F}_n\}_{n\geqslant 0}$ is regular and $1 \leqslant r < \infty$. If $\alpha r + 1 > 0$ for a fixed α, then*

$$w\mathcal{BMO}_r(\alpha) = w\mathcal{BMO}_2(\alpha)$$

with equivalent norms.

THEOREM 4.1. *Suppose that $\alpha \geqslant 0$. Then*

$$w\mathcal{BMO}_2(\alpha) = wBMO_2(\alpha)$$

with equivalent norms.

P r o o f. Let $p = \frac{1}{1+\alpha}$. Since $(\mathcal{H}^s_{p,\infty})^* \subset (\overline{H}^s_{p,\infty})^*$, it follows from Proposition 3.1 and Lemma 4.1 that

$$wBMO_2(\alpha) \subset w\mathcal{BMO}_2(\alpha).$$

To obtain

$$wBMO_2(\alpha) \supset w\mathcal{BMO}_2(\alpha),$$

we shall show that

$$C\|f\|_{w\mathcal{BMO}_2(\alpha)} \geqslant \|f\|_{wBMO_2(\alpha)}$$

for any $f \in w\mathcal{BMO}_2(\alpha)$. Suppose that $\{\nu_k\}_{k\in\mathbb{Z}}$ is an arbitrary stopping time sequence such that $\{2^k\mathbb{P}(\nu_k < \infty)^{1/p}\}_{k\in\mathbb{Z}} \in \ell_\infty$. Let

$$B = \sup_k 2^k\mathbb{P}(\nu_k < \infty)^{1/p}.$$

We can claim that

$$\sum_{k=-\infty}^{\infty} t_\alpha^2(B^p 2^{-kp}) \leqslant C\|f\|_{w\mathcal{BMO}_2(\alpha)}.$$

To this end, let $C_k = B2^{-kp}$. Then, for any $x \in (C_{k+1}, C_k)$, we have

$$C_{k+1}^{1/2+\alpha} t_\alpha^2(C_{k+1}) \leqslant x^{1/2+\alpha} t_\alpha^2(x) \leqslant C_k^{1/2+\alpha} t_\alpha^2(C_k).$$

We refer to [17], p. 144, for a more general case of the inequalities above. Hence,

$$\int_0^\infty \frac{t_\alpha^2(x)}{x}dx = \sum_{k=-\infty}^{\infty} \int_{C_{k+1}}^{C_k} \frac{t_\alpha^2(x)}{x}dx \geqslant (1-2^{-p})2^{-p(1/2+\alpha)} \sum_{k=-\infty}^{\infty} t_\alpha^2(B^p 2^{-kp}).$$

On the other hand, since $B^p 2^{-kp} \geqslant \mathbb{P}(\nu_k < \infty)$ for all k, we have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} t_\alpha^2(B^p 2^{-kp}) &\geqslant \sum_{k=-\infty}^{\infty} \frac{2^k (B^p 2^{-kp})^{1/2}\|f - f^{\nu_k}\|_2}{B} \\ &\geqslant \sum_{k=-\infty}^{\infty} \frac{2^k\mathbb{P}(\nu_k < \infty)^{1/2}\|f - f^{\nu_k}\|_2}{B}.\end{aligned}$$

By the definition of $wBMO_2(\alpha)$, we complete the proof. ∎

REMARK 4.1. *If one proves the dual space of $\mathcal{H}^s_{p,\infty}$ is $w\mathcal{BMO}(\alpha)$, then Theorem* 4.1 *holds. If one shows $\mathcal{H}^s_{p,\infty} = \overline{H}^s_{p,\infty}$, then Proposition* 3.1 *implies Theorem* 4.1. *We leave the proofs to the interested reader.*

Now we are ready to prove the main result of the paper.

Proof of Theorem 1.1. It directly follows from Theorems 3.1 and 4.1 and Lemma 4.2. ∎

Acknowledgments. The authors would like to thank the anonymous referee for helpful comments and suggestions to improve the paper.

REFERENCES

[1] W. Abu-Shammala and A. Torchinsky, *The Hardy–Lorentz spaces $H^{p,q}(\mathbb{R}^n)$*, Studia Math. 182 (3) (2007), pp. 283–294.

[2] L. Grafakos, *Classical Fourier Analysis*, second edition, Springer, New York 2008.

[3] Z. Hao, Y. Jiao, F. Weisz, and D. Zhou, *Atomic subspaces of L_1-martingale spaces*, Acta Math. Hungar. 150 (2) (2016), pp. 423–440.

[4] C. Herz, *H_p-spaces of martingales, $0 < p \leqslant 1$*, Z. Wahrsch. Verw. Gebiete 28 (1973/74), pp. 189–205.

[5] K. Ho, *Atomic decompositions, dual spaces and interpolations of martingale Hardy–Lorentz–Karamata spaces*, Q. J. Math. 65 (3) (2014), pp. 985–1009.

[6] Y. Jiao, P. Lihua, and L. Peide, *Atomic decompositions of Lorentz martingale spaces and applications*, J. Funct. Spaces 7 (2) (2009), pp. 153–166.

[7] Y. Jiao, L. Wu, and L. Peng, *Weak Orlicz–Hardy martingale spaces*, Internat. J. Math. 26 (8) (2015), 1550062.

[8] Y. Jiao, L. Wu, A. Yang, and R. Yi, *The predual and John–Nirenberg inequalities on generalized BMO martingale spaces*, Trans. Amer. Math. Soc. 369 (1) (2017), pp. 537–553.

[9] Y. Jiao, D. Zhou, F. Weisz, and L. Wu, *Variable martingale Hardy spaces and their applications in Fourier analysis*, submitted, 2018.

[10] Y. Jiao, Y. Zuo, D. Zhou, and L. Wu, *Variable Hardy–Lorentz spaces $H^{p(\cdot),q}(\mathbb{R}^n)$*, Math. Nachr. (to appear).

[11] F. John and L. Nirenberg, *On functions of bounded mean oscillation*, Comm. Pure Appl. Math. 14 (1961), pp. 415–426.

[12] K. Liu and D. Zhou, *Dual spaces of weak martingale Hardy–Lorentz–Karamata spaces*, Acta Math. Hungar. 151 (1) (2017), pp. 50–68.

[13] K. Liu, D. Zhou, and L. Peng, *A weak type John–Nirenberg theorem for martingales*, Statist. Probab. Lett. 122 (2017), pp. 190–197.

[14] R. L. Long, *Martingale Spaces and Inequalities*, Peking University Press, Beijing 1993.

[15] F. Weisz, *Martingale Hardy spaces for $0 < p \leqslant 1$*, Probab. Theory Related Fields 84 (3) (1990), pp. 361–376.

[16] F. Weisz, *Martingale Hardy Spaces and Their Applications in Fourier Analysis*, Springer, Berlin 1994.

[17] F. Weisz, *Weak martingale Hardy spaces*, Probab. Math. Statist. 18 (1) (1998), pp. 133–148.

[18] R. Yi, L. Wu, and Y. Jiao, *New John–Nirenberg inequalities for martingales*, Statist. Probab. Lett. 86 (2014), pp. 68–73.

Dejian Zhou
School of Mathematics and Statistics
Central South University
Changsha 410075, China
E-mail: zhoudejian@csu.edu.cn

Weiwei Li
School of Mathematics and Statistics
Central South University
Changsha 410075, China
E-mail: liweiweionline@hotmail.com

Yong Jiao
School of Mathematics and Statistics
Central South University
Changsha 410075, China
E-mail: jiaoyong@csu.edu.cn

Received on 8.4.2015;
revised version on 15.3.2017

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 299–315
doi:10.19195/0208-4147.38.2.4

SERIES REPRESENTATION OF TIME-STABLE STOCHASTIC PROCESSES

BY

CHRISTOPH KOPP (BERN) AND ILYA MOLCHANOV* (BERN)

Abstract. A stochastically continuous process $\xi(t)$, $t \geqslant 0$, is said to be *time-stable* if the sum of n i.i.d. copies of ξ equals in distribution the time-scaled stochastic process $\xi(nt)$, $t \geqslant 0$. The paper advances the understanding of time-stable processes by means of their LePage series representations as the sum of i.i.d. processes with the arguments scaled by the sequence of successive points of the unit intensity Poisson process on $[0, \infty)$. These series yield numerous examples of stochastic processes that share one-dimensional distributions with a Lévy process.

2010 AMS Mathematics Subject Classification: Primary: 60G52; Secondary: 60G51.

Key words and phrases: Infinite divisibility, LePage series, Lévy process, point process, time-stable process.

1. INTRODUCTION

The (strict) stability property of stochastic processes is conventionally defined by requiring that the sum of i.i.d. copies of a process is distributed as the space-scaled variant of the original process. An alternative scaling operation applied to the time argument leads to another definition of stability.

DEFINITION 1.1. A stochastically continuous real-valued process $\xi(t)$, $t \geqslant 0$, is said to be *time-stable* if, for each $n \geqslant 2$,

$$\xi_1 + \ldots + \xi_n \stackrel{\mathcal{D}}{\sim} n \circ \xi, \tag{1.1}$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. copies of ξ, $\stackrel{\mathcal{D}}{\sim}$ means the equality of all finite-dimensional distributions and $(n \circ \xi)(t) = \xi(nt)$, $t \geqslant 0$, is the process obtained by time scaling ξ.

* Supported in part by Swiss National Science Foundation grants 200021-137527 and 200021-153597.

Definition 1.1 goes back to Mansuy [18], where processes satisfying (1.1), regardless of their stochastic continuity, are called *infinitely divisible with respect to time* (IDT), see also [4], Section 6.7. Indeed, they are infinitely divisible in the sense that ξ can be represented as the sum of n i.i.d. processes for each $n \geqslant 2$. However, the "time-stable" name better emphasises the particular stability feature of such processes. These processes have recently been investigated in [8], [12], also with a multivariate time argument. Time-stable processes with values in $\mathbb{R}^d$ can be defined similarly to Definition 1.1. Similarly to other stable random elements, time-stable processes naturally appear as limits for time-scaled sums of stochastic processes.

The major difficulty in the analysis of time-stable processes stems from the necessity to work with the whole paths of the processes. The time-stability concept cannot be formulated in terms of finite-dimensional distributions at any given time moments, since the time argument on the right-hand side of (1.1) is scaled.

Definition 1.1 can be modified to introduce α-time-stable processes as

$$\xi_1 + \ldots + \xi_n \stackrel{\mathcal{D}}{\sim} n^{1/\alpha} \circ \xi,$$

where each $\alpha \neq 0$ is possible. This concept appears in Example 8.12 of [7] as an example of the stability property in the cone of continuous functions with scaling applied to the argument. While such processes (for general α) have been considered in [11], the process $\xi(t^{1/\alpha})$, $t \geqslant 0$, obtained by time change is time-stable (with $\alpha = 1$) and so it is not necessary to study α-time stability for general $\alpha \neq 1$.

Another closely related concept is that of a *dilatively stable process* ζ that satisfies the following scaling relation for some $\alpha > 0$, $\delta \in (0, 2\alpha]$, and all $n \geqslant 2$:

$$\zeta_1 + \ldots + \zeta_n \stackrel{\mathcal{D}}{\sim} n^{1/2-\alpha/\delta}(n^{1/\delta} \circ \zeta),$$

see [14], where such processes are also assumed to have moments of all orders and the left-hand side is replaced by the convolution power for finite-dimensional distributions of order n with not necessarily integer n (which however does not alter the family of processes). If ζ is dilatively stable, then $\xi(t) = t^{1/2-\alpha/\delta}\zeta(t^{1/\delta})$, $t \geqslant 0$, satisfies (1.1) and so is a time-stable process if ζ is stochastically continuous.

Barczy et al. [1] extended the setting from [14] by allowing α and δ be arbitrary real numbers and relaxing the moment conditions. They also defined (ρ_1, ρ_2)-*aggregate self-similar processes* ζ for arbitrary real numbers ρ_1 and ρ_2 by the scaling relation

$$\zeta_1 + \ldots + \zeta_n \stackrel{\mathcal{D}}{\sim} n^{\rho_1}(n^{-\rho_2} \circ \zeta),$$

so that for $\rho_1 = \rho_2$ one recovers the aggregate similar process from [16]. It is easy to see that $t^{\rho_1}\zeta(t^{-\rho_2})$, $t \geqslant 0$, satisfies (1.1), so that this and all other above-mentioned generalisations may be obtained by time and scale change from time-stable processes. An exponential time change leads to translatively stable processes, see [13], Definition 2.4.3. A similar concept was introduced by Penrose [21], who

called a non-negative stochastic process ξ *semi-min-stable* if $\min\bigl(\xi_1(t),\dots,\xi_n(t)\bigr)$ shares the finite-dimensional distributions with $n^{-1}\xi(n^\alpha t)$, $t \geqslant 0$.

Section 2 discusses elementary properties of time-stable processes. The infinite divisibility of such processes makes it possible to use their spectral representation obtained in [15] and then show that the Lévy measure is homogeneous with respect to time scaling, see Section 3. The main result of Section 4 and of the whole paper is the LePage representation of time-stable processes whose Lévy measures are supported by the family of right-continuous functions with left limits. In particular, this is the case for non-negative processes. The obtained LePage representation yields the series representations for dilatively stable and aggregate self-similar processes. The structure of pure jump time-stable processes is closely related to the stability property of marked point processes; in this case the LePage representation is similar to the cluster representation of infinitely divisible point processes, see Section 5.

The concept of time stability allows generalisations in various directions. The necessary structure consists of a time set which is invariant under scaling by arbitrary positive real numbers and an associative and commutative binary operation which is applied pointwisely to the values of processes. For instance, the definition applies also to stochastic processes defined on the whole line and on $\mathbb{R}^d$ or with addition replaced by the maximum operation.

While (1.1) actually defines a *strictly* time-stable stochastic process, the stability concept can be relaxed by replacing the right-hand side with $n \circ \xi + f_n$ for deterministic functions $\{f_n\}$. Moreover, it is possible to consider random measures stable with respect to scaling of their argument (see [7], Example 8.23) and also time-stable generalised stochastic processes, i.e. random generalised functions.

2. ELEMENTARY PROPERTIES

The following standard result provides an alternative definition of time-stable processes.

PROPOSITION 2.1. *A stochastically continuous process $\xi(t)$, $t \geqslant 0$, is time-stable if and only if*

$$a \circ \xi_1 + b \circ \xi_2 \overset{\mathcal{D}}{\sim} (a+b) \circ \xi \tag{2.1}$$

for all $a, b > 0$, where ξ_1 and ξ_2 are independent copies of ξ.

Each Lévy process is time-stable, see [4], Section 6.7. If ξ is time-stable, then there exists a unique Lévy process $\tilde{\xi}$, called the *associated Lévy process* of ξ, such that $\tilde{\xi}(t)$ coincides in distribution with $\xi(t)$ for each given $t \geqslant 0$, see [18], Proposition 4.1. Thus, the characteristic function of $\xi(t)$ is given by

$$\mathbf{E}\exp\{\imath\lambda\xi(t)\} = \exp\{-t\Psi(\lambda)\}, \quad t \geqslant 0,\ \lambda \in \mathbb{R}, \tag{2.2}$$

where Ψ denotes the *cumulant* of $\tilde{\xi}$ and also of ξ.

It follows from (2.2) that $\xi(t)$ weakly converges to zero as $t \downarrow 0$, which corresponds to the stochastic continuity of ξ, since $\xi(0) = 0$ a.s. by (1.1). Furthermore, if $\xi(t)$ and $\xi(s)$ share the same distribution for $t \neq s$, then ξ is a.s. zero.

Comparing the one-dimensional distributions shows that if the non-degenerate time-stable process is a.s. non-negative for any $t > 0$, then it is a.s. non-negative everywhere, its one-dimensional distributions are increasing in the stochastic order, and $\sup_{t \geqslant 0} \xi(t)$ is a.s. infinite. In contrast to Lévy processes, non-negative time-stable processes need not be a.s. monotone, for example, $\xi(t) = N(2t) - N(t)$, $t \geqslant 0$, if N is the standard Poisson process.

THEOREM 2.1. *A time-stable process ξ is identically distributed as the sum of a linear function, a centred Gaussian process with covariance function C that satisfies $C(ut, us) = uC(t, s)$ for all $t, s, u \geqslant 0$, and an independent time-stable process without Gaussian component.*

P r o o f. Since ξ is infinitely divisible, its finite-dimensional distributions are infinitely divisible. The rest follows by comparing the Lévy triplets of the n-fold convolution of $\big(\xi(t_1), \dots, \xi(t_k)\big)$ and of $\big(\xi(nt_1), \dots, \xi(nt_k)\big)$ for any $t_1, \dots, t_k \geqslant 0$ and $k, n \geqslant 1$. ∎

Various characterisations of Gaussian time-stable processes are presented in [18]. In the following we only consider time-stable processes without a Gaussian part.

3. LÉVY MEASURES OF TIME-STABLE PROCESSES

Each stochastically continuous process is separable in probability (also is said to satisfy Condition S from [25], Definition 3.11.2), meaning the existence of an at most countable set $T_0 \subset [0, \infty)$ such that for all $t \geqslant 0$, there exists a sequence $t_n \in T_0$, $n \geqslant 1$, such that $\xi(t_n)$ converges to $\xi(t)$ in probability. The spectral representation of infinitely divisible stochastic processes that are separable in probability and do not have a Gaussian component is obtained in [15], Theorem 2.14, using a Poisson process on a certain space $(\Omega, \mathfrak{F})$ with a σ-finite measure μ. Reformulating this result for $(\Omega, \mathfrak{F})$ being the space $\mathbb{R}^{[0,\infty)}$ with the cylindrical σ-algebra $\mathfrak{C}$, we see that an infinitely divisible stochastically continuous process ξ without a Gaussian component admits a spectral representation

$$\xi(t) \stackrel{\mathcal{D}}{\sim} c(t) + \int_{\mathbb{R}^{[0,\infty)} \setminus \{0\}} f(t) d\Pi_Q(f), \tag{3.1}$$

where c is a deterministic function and $\Pi_Q = \{f_i(\cdot) : \ i \geqslant 1\}$ is the Poisson process on $\mathbb{R}^{[0,\infty)} \setminus \{0\}$ with a σ-finite intensity measure Q such that

$$\int_{\mathbb{R}^{[0,\infty)} \setminus \{0\}} \min\big(1, f(t)^2\big) Q(df) < \infty \tag{3.2}$$

for all $t \geqslant 0$. The measure Q is called the *Lévy measure* of ξ. The integral with respect to Π_Q in (3.1) is defined as the a.s. existing limit of the compensated sums

$$\lim_{r \downarrow 0} \Big[\sum_{f_i \in \Pi_Q} f_i(t) \mathbf{1}_{|f_i(t)|>r} - \int\limits_{\{f\colon |f(t)|>r\}} L\big(f(t)\big) Q(df) \Big], \tag{3.3}$$

where

$$L(u) = \begin{cases} u, & |u| \leqslant 1, \\ 1, & u > 1, \\ -1, & u < -1, \end{cases} \tag{3.4}$$

is a Lévy function, see also [19].

Furthermore, Theorem 2.14 in [15] ensures the existence of a *minimal* spectral representation, meaning that the σ-algebra generated by $\{f : \ f(t) \in A\}$ for all $t \geqslant 0$ and Borel $A \subset \mathbb{R}$ coincides with the cylindrical σ-algebra $\mathfrak{C}$ on $\mathbb{R}^{[0,\infty)}$ up to Q-null sets and there is no set $B \in \mathfrak{C}$ with $Q(B) > 0$ such that for every $t \geqslant 0$, $Q\big(\{f \in B : \ f(t) \neq 0\}\big) = 0$. In the following assume that the cylindrical σ-algebra $\mathfrak{C}$ is complete with respect to Q. By [15], Theorem 2.17, the minimal spectral representation is unique up to an isomorphism, and so the Lévy measure is well defined.

The stochastic continuity of ξ implies that ξ has a measurable modification, see [10], Theorem 3.3.1. Then Proposition 2.19 in [15] establishes that the representation (3.1) involves a measurable function $c(t)$, $t \geqslant 0$, and that the functions f from Π_Q can be chosen to be strongly separable. The latter means the existence of a measurable null-set $\Omega_0 \subset \mathbb{R}^{[0,\infty)}$ and a countable set $\mathbb{Q} \subset [0,\infty)$ (called a *separant*) such that, for each open $G \subset [0,\infty)$ and closed $F \subset \mathbb{R}$, we have

$$\{f : \ f(t) \in F \ \forall t \in G \cap \mathbb{Q}\} \setminus \{f : \ f(t) \in F \ \forall t \in G\} \subset \Omega_0. \tag{3.5}$$

If (3.2) is strengthened to require

$$\int\limits_{\mathbb{R}^{[0,\infty)} \setminus \{0\}} \min\big(1, |f(t)|\big) Q(df) < \infty, \tag{3.6}$$

then the integral (3.1) is well defined without taking the limit and without the compensating term in (3.3), so that

$$\xi(t) \stackrel{\mathcal{D}}{\sim} c(t) + \sum_{f_i \in \Pi_Q} f_i(t) \tag{3.7}$$

for a deterministic function c. It is well known that (3.6) holds if $\xi(t)$ is a.s. nonnegative for all $t \geqslant 0$, see e.g. [26], Theorem 51.1.

LEMMA 3.1. *For each $B \in \mathfrak{C}$ and $s > 0$, the set $s \circ B = \{s \circ f : \ f \in B\}$ also belongs to $\mathfrak{C}$.*

P r o o f. If B is a cylinder, then $s \circ B \in \mathfrak{C}$, and the statement follows from the monotone class argument. ∎

The next result follows from the fact that $\xi(0) = 0$ a.s. for a time-stable process ξ.

LEMMA 3.2. *The Lévy measure of a time-stable process is supported by* $\{f \in \mathbb{R}^{[0,\infty)} \setminus \{0\} : \ f(0) = 0\}$.

LEMMA 3.3. *An infinitely divisible stochastically continuous process ξ without a Gaussian component is time-stable if and only if $c(t) = bt$, $t \geqslant 0$, for a constant $b \in \mathbb{R}$ and the Lévy measure Q satisfies*

$$Q(s \circ B) = s^{-1} Q(B), \quad s > 0, \tag{3.8}$$

for all $B \in \mathfrak{C}$.

P r o o f. The sufficiency follows from the expression for the characteristic function of the finite-dimensional distributions of ξ,

$$\begin{aligned} &\mathbf{E} \exp\{\imath \sum_j \theta_j \xi(t_j)\} \\ &= \exp\{\imath b \sum_j \theta_j t_j + \int [\exp(\imath \sum_j \theta_j f(t_j)) - 1 - \imath \sum_j \theta_j L(f(t_j))] Q(df)\}. \end{aligned} \tag{3.9}$$

Now assume that ξ is time-stable. Comparing the characteristic functions of the finite-dimensional distributions for the processes on the left- and right-hand side of (2.1) and using the uniqueness of the Lévy triplets show that the function c is additive and so is linear in view of its measurability.

The Lévy measure corresponding to the minimal spectral representation of the process on the left-hand side of (2.1) is $Q(a^{-1} \circ B) + Q(b^{-1} \circ B)$. In view of the uniqueness of the minimal spectral representation (see [15], Theorem 2.17), the Lévy measures of the processes on the left- and right-hand side of (2.1) coincide. Thus

$$Q(a^{-1} \circ B) + Q(b^{-1} \circ B) = Q((a+b)^{-1} \circ B)$$

for all $a, b > 0$ and all $B \in \mathfrak{C}$. Since Q is non-negative, Theorem 1.1.7 in [3] implies that $Q(s^{-1} \circ B)$ is a linear function of s, i.e. (3.8) holds. ∎

The same scaling property for the Lévy measure appears in [18], Lemma 5.1, and later on was reproduced in [11], Proposition 4.1, for time-stable processes with paths in the Skorokhod space of right-continuous functions with left limits (*càdlàg* functions). The proof there is however incomplete, since it is not shown that the Lévy measure of such a process is supported by *càdlàg* functions.

PROPOSITION 3.1. *If $\xi(t)$, $t \geqslant 0$, is a time-stable* càdlàg *process with a.s. non-negative values, then its Lévy measure Q is supported by* càdlàg *functions.*

P r o o f. In this case the Lévy measure Q satisfies (3.6) and so ξ admits the representation (3.7). If ξ' is an independent copy of ξ, then $\xi - \xi'$ is symmetric and has the series decomposition with the Lévy measure supported by *càdlàg* (free of oscillatory discontinuities) functions, see [22], Theorem 4. The support of Q is a subset of the support of the Lévy measure for $\xi - \xi'$. ∎

4. LEPAGE SERIES REPRESENTATION

In finite-dimensional spaces, Lévy measures of strictly stable laws admit a polar decomposition into the product of a radial and a finite-directional part, and the corresponding sum (if necessary compensated) of points of the Poisson process is known as the LePage series, see [25], Corollary 3.10.4, and [17], [23]. The LePage series can be defined in general spaces [7], where they provide a rich source of stable laws and in many cases characterise stable laws.

The following result provides the LePage series characterisation for time-stable processes without a Gaussian part and whose Lévy measure is supported by the family $\mathbb{D}'$ of not identically vanishing *càdlàg* functions on $[0, \infty)$. We endow the family $\mathbb{D}'$ with the σ-algebra induced by $\mathfrak{C}$. Let $\mathbb{D}'_0$ be the family of not identically vanishing *càdlàg* functions that vanish at the origin.

THEOREM 4.1. *The following statements are equivalent for a stochastically continuous* càdlàg *process $\xi(t)$, $t \geqslant 0$.*

(i) *The process ξ is time-stable without a Gaussian part and with its Lévy measure Q supported by $\mathbb{D}'$.*

(ii) *The stochastic process ξ is infinitely divisible without a Gaussian part, with a deterministic linear part, its Lévy measure Q is supported by $\mathbb{D}'_0$, and*

$$Q(B) = \int_{\mathbb{D}'_0} \int_0^\infty \sigma(t \circ B)dt \tag{4.1}$$

for each measurable $B \subset \mathbb{D}'_0$ and a probability measure σ on $\mathbb{D}'_0$ such that

$$\int_{\mathbb{D}'_0} \int_0^\infty \min\big(1, f(t)^2\big)t^{-2}dt\sigma(df) < \infty. \tag{4.2}$$

(iii) *The stochastic process ξ has the same distribution as*

$$bt + \lim_{r \downarrow 0} \Big[\sum_{i=1}^\infty \varepsilon_i(\Gamma_i^{-1}t)\mathbf{1}_{|\varepsilon_i(\Gamma_i^{-1}t)|>r} - \mathbf{E}\int_0^\infty L\big(\varepsilon(s^{-1}t)\big)\mathbf{1}_{|\varepsilon(s^{-1}t)|>r}\, ds\Big], \quad t \geqslant 0, \tag{4.3}$$

where the limit exists almost surely, $b \in \mathbb{R}$ *is a constant*, L *is defined as in* (3.4), $\{\varepsilon_i, i \geqslant 1\}$ *is a sequence of i.i.d. stochastic processes distributed as* ε, *such that* ε *a.s. takes values in* $\mathbb{D}'_0$,

$$\mathbf{E} \int_0^\infty \min\big(1, \varepsilon(t)^2\big) t^{-2} dt < \infty, \tag{4.4}$$

and $\{\Gamma_i, i \geqslant 1\}$ *is the sequence of successive points of a homogeneous unit intensity Poisson process on* $[0, \infty)$.

P r o o f. By Lemma 3.3, a time-stable process without a Gaussian part can be alternatively described as an infinitely divisible stochastically continuous process whose Lévy measure Q satisfies (3.8) and so is supported by $\mathbb{D}'_0$. It is obvious that Q given by (4.1) satisfies (3.8). It remains to show that the scaling property (3.8) yields (4.1), so that (i) implies (ii).

The following construction is motivated by the argument used to prove Theorem 10.3 in [9]. By Lemma 3.2, Q is supported by $\mathbb{D}'_0$. Decompose $\mathbb{D}'_0$ into the union of disjoint sets

$$\mathbb{X}_0 = \{f : \sup_{t \geqslant 0} |f(t)| > 1\}$$

and

$$\mathbb{X}_k = \{f : \sup_{t \geqslant 0} |f(t)| \in (2^{-k}, 2^{-k+1}], f \notin \mathbb{X}_j, j = 0, \dots, k-1\}, \quad k \geqslant 1.$$

In view of the completeness assumption on the σ-algebra, all sets $\mathbb{X}_k$, $k \geqslant 0$, are measurable. Recall the separant $\mathbb{Q}$ and the exceptional set Ω_0 from (3.5) that holds due to the assumed stochastic continuity and infinite divisibility of ξ. Denote by $\tilde{\mathbb{X}}_k$, $k \geqslant 0$, the analogues of $\mathbb{X}_k$ where the supremum is taken over the set of non-negative rational numbers. Since

$$\mathbb{X}_0^c = \{f : |f(t)| \leqslant 1, t \in [0, \infty)\},$$

we have $\mathbb{X}_0 \setminus \tilde{\mathbb{X}}_0 \subset \Omega_0$. Similarly, $\mathbb{X}_k \setminus \tilde{\mathbb{X}}_k \subset \Omega_0$ for all $k \geqslant 1$.

For each $k \geqslant 0$, define the map $\tau_k : \mathbb{X}_k \to (0, \infty)$ by

$$\tau_k(f) = \inf\{t > 0 : |f(t)| > 2^{-k}\}, \quad f \in \mathbb{X}_k.$$

Since all functions from $\mathbb{D}'_0$ vanish at the origin, $\tau_k(f)$ is strictly positive and finite, and $\tau_k(c \circ f) = c^{-1}\tau_k(f)$ for all $c > 0$. Let

$$\mathbb{S}_k = \{f \in \mathbb{X}_k : \tau_k(f) = 1\}.$$

Then $|f(1)| \geqslant 2^{-k}$ for all $f \in \mathbb{S}_k$, $k \geqslant 0$, and each function $g \in \mathbb{X}_k$ can be uniquely represented as $s \circ f$ for $f \in \mathbb{S}_k$ and $s > 0$. The maps $(f, s) \mapsto s \circ f$ and $g \mapsto$

$\big(\tau_k(g)\circ g, \tau_k(g)^{-1}\big)$ are mutually inverse measurable bijections between the sets $\mathbb{S}_k\times(0,\infty)$ and $\mathbb{X}_k$. This is seen by using the separability assumption (3.5) and Lemma 3.1. The right-continuity of f and (3.5) imply that

$$\Delta_k(f)=\sup\{t\in\mathbb{Q}:\ |f(s)|\geqslant 2^{-k-1}\text{ for all } s\in[1,1+t]\},\quad f\in\mathbb{S}_k,$$

is strictly positive and Borel measurable for each $k\geqslant 0$. Define

$$\begin{aligned}\mathbb{S}_{k0}&=\{f\in\mathbb{S}_k:\ \Delta_k(f)>1\},\\ \mathbb{S}_{kj}&=\{f\in\mathbb{S}_k:\ \Delta_k(f)\in(2^{-j},2^{-j+1}]\},\quad j\geqslant 1.\end{aligned}$$

Then $\mathbb{S}_k$ is the disjoint union of $\mathbb{S}_{kj}$ for $j\geqslant 0$ and $\mathbb{X}_k$ is the disjoint union of

$$\mathbb{X}_{kj}=\{s\circ f:\ f\in\mathbb{S}_{kj},\ s>0\},\quad j\geqslant 0.$$

Fix any $k,j\geqslant 0$. Then

$$\begin{aligned}q_{kj}&=Q(\{s\circ f:\ f\in\mathbb{S}_{kj},\ s\in[1,1+2^{-j}]\})\\ &\leqslant Q\big(\{f\in\mathbb{D}_0':\ |f(1)|\geqslant 2^{-k-1}\}\big)\\ &\leqslant 2^{2k+2}\int\limits_{\{f:\,|f(1)|\geqslant 2^{-k-1}\}}\min\big(1,f(1)^2\big)Q(df)\\ &\leqslant 2^{2k+2}\int\min\big(1,f(1)^2\big)Q(df)<\infty.\end{aligned}$$

By (3.8),

$$\begin{aligned}&Q(\{s\circ f: f\in\mathbb{S}_{kj},\ s\geqslant 1\})\\ &\qquad\leqslant\sum_{i=0}^{\infty}Q\big(\{s\circ f:\ f\in\mathbb{S}_{kj},\ s\in[(1+2^{-j})^i,\,(1+2^{-j})^{i+1}]\}\big)\\ &\qquad=\sum_{i=0}^{\infty}(1+2^{-j})^{-i}q_{kj}<\infty.\end{aligned}$$

Thus, Q restricted onto $\mathbb{X}_{kj}$ is a push-forward under the map $(f,s)\to s\circ f$ of the product $\eta_{kj}\otimes\theta$ of a finite measure η_{kj} supported by $\mathbb{S}_{kj}$ and the measure θ on $(0,\infty)$ with density $s^{-2}ds$. Let c_{kj} be some positive number; then the measure σ_{kj} defined on $\mathbb{D}_0'$ by $\sigma_{kj}(B)=c_{kj}\eta_{kj}(c_{kj}^{-1}\circ B)$ assigns all its mass to the set $c_{kj}\circ\mathbb{S}_{kj}$. Then the push-forward of $\sigma_{kj}\otimes\theta$ under the map $(f,s)\to s\circ f$ is Q restricted on $\mathbb{X}_{kj}$ and the total mass of σ_{kj} equals $c_{kj}\eta_{kj}(\mathbb{S}_{kj})$. By choosing c_{kj} appropriately, it is always possible to achieve that $\sigma=\sum_{k,j\geqslant 0}\sigma_{kj}$ is a probability measure on $\mathbb{D}_0'$. Combining the push-forward representations of Q restricted to $\mathbb{X}_{kj}$, $k,j\geqslant 0$, we see that Q is the push-forward of $\sigma\otimes\theta$ and so (4.1) holds. Given (4.1), (4.2) is equivalent to (3.2).

The equivalence of (ii) and (iii) is immediate by choosing ε to be i.i.d. with distribution σ and noticing that (4.2) is equivalent to (4.4) and that the limit in (4.3) corresponds to the limit in (3.3). Note that $\{\Gamma_i^{-1}, i\geqslant 1\}$ form the Poisson process on $\mathbb{R}_+$ with intensity $s^{-2}ds$. ∎

REMARK 4.1. *There are many probability measures σ that satisfy* (4.1), *and so the distribution of ε in* (4.3) *is not unique. For example*, *it is possible to scale the time arguments of $\{\varepsilon_i, i \geqslant 1\}$ by a sequence of i.i.d. positive random variables of mean one. The distribution of ε is unique if ε is supported by a given measurable set $\mathbb{S}' \subset \mathbb{D}'_0$ such that each $f \in \mathbb{D}'_0$ can be uniquely represented as $c \circ g$ for $c > 0$ and $g \in \mathbb{S}'$.*

REMARK 4.2. *It follows from Theorem* 3.1 *in* [2] *that the LePage series* (4.3) *converges uniformly for t from any compact subset of $(0, \infty)$. If $H(t, r, V) = \varepsilon(t/r)$, then Condition* (3.3) *of* [2] *becomes*

$$\int_0^\infty \mathbf{P}\big\{\big(\varepsilon(t_1/r), \dots, \varepsilon(t_k/r)\big) \in B\big\} dr = Q\Big(\big\{f : \ \big(f(t_1), \dots, f(t_k)\big) \in B\big\}\Big)$$

for all Borel B in $\mathbb{R}^k \setminus \{0\}$, $t_1, \dots, t_k \geqslant 0$, and $k \geqslant 1$.

THEOREM 4.2. *A stochastically continuous stochastic process ξ is time-stable without a Gaussian part and with the Lévy measure Q supported by $\mathbb{D}'$ and satisfying* (3.6) *if and only if*

$$\xi(t) \stackrel{\mathcal{D}}{\sim} bt + \sum_{i=1}^\infty \varepsilon_i(\Gamma_i^{-1} t), \quad t \geqslant 0, \tag{4.5}$$

where the series converges almost surely, $b \in \mathbb{R}$ is a constant, $\{\varepsilon_i, i \geqslant 1\}$ is a sequence of i.i.d. stochastic processes with realisations in $\mathbb{D}'_0$ such that

$$\mathbf{E} \int_0^\infty \min\big(1, |\varepsilon(t)|\big) t^{-2} dt < \infty, \tag{4.6}$$

and $\{\Gamma_i, i \geqslant 1\}$ is the sequence of successive points of the homogeneous unit intensity Poisson process on $[0, \infty)$.

P r o o f. It suffices to note that (4.6) is equivalent to (3.6). ∎

COROLLARY 4.1. *Each a.s. non-negative* càdlàg *time-stable process admits the LePage representation* (4.5).

REMARK 4.3. *Condition* (4.6) (*respectively* (4.4)) *holds if $\int_0^1 \mathbf{E}|\varepsilon(t)| t^{-2} dt < \infty$ (respectively $\int_0^1 \mathbf{E}\big(\varepsilon(t)^2\big) t^{-2} dt < \infty$). For example, this is the case if $\varepsilon(t) = 0$, $t \in [0, \tau)$, for a positive random variable τ such that τ^{-1} is integrable.*

REMARK 4.4. *Analogues of the above results hold for time-stable processes with values in $\mathbb{R}^d$. This can be shown by replacing $\mathbb{S}_{kj}$ from the proof of Theorem* 4.1 *with the Cartesian product of d-tuples of such sets $\mathbb{S}^1_{k_1 j_1} \times \ldots \times \mathbb{S}^d_{k_d j_d}$, $k_i, j_i \geqslant 0$, $i = 1, \dots, d$, constructed for each of the coordinates of the process. In particular*, *Corollary* 4.1 *applies for time-stable processes with values in $\mathbb{R}^d_+$.*

EXAMPLE 4.1 (*Lévy processes*). The spectral representation (3.1) of a Lévy process ξ without a Gaussian part can be obtained by setting $f_i(t) = m_i \mathbf{1}_{t \geqslant \tau_i}$, where $\{(\tau_i, m_i),\ i \geqslant 1\}$ is a marked Poisson process on $(0,\infty) \times (\mathbb{R} \setminus \{0\})$ with intensity measure being the product of the Lebesgue measure and a Lévy measure Λ on $\mathbb{R} \setminus \{0\}$. Indeed, then

$$\xi(t) \overset{\mathcal{D}}{\sim} bt + \lim_{r \downarrow 0} \Big[\sum_{|m_i| > r} m_i \mathbf{1}_{\tau_i \leqslant t} - t \int_{|x|>r} L(x) \Lambda(dx) \Big],$$

which is the classical decomposition of a Lévy process. In view of the uniqueness of the minimal spectral representation, the Lévy measure Q is supported by step functions of type $m\mathbf{1}_{t \geqslant \tau}$. By Theorem 4.1, ξ admits the series decomposition (4.3) with $\varepsilon(t) = \eta \mathbf{1}_{t\zeta \geqslant 1}$, where (4.4) corresponds to $\mathbf{E}[\min(1, \eta^2)\zeta] < \infty$. Following the construction from the proof of Theorem 4.1, the joint distribution of (η, ζ) can be constructed as follows. Write $B_0 = \{x \in \mathbb{R} :\ |x| > 1\}$ and $B_k = \{x \in \mathbb{R} :\ 2^{-k} < |x| \leqslant 2^{-k+1}\}$, $k \geqslant 1$, let $q_k = \Lambda(B_k)$, $k \geqslant 0$, and choose strictly positive $\{c_k, k \geqslant 0\}$ such that $\sum_{k=0}^{\infty} c_k q_k = 1$. Then

$$\mathbf{P}\{\eta \in A, \zeta = c_k^{-1}\} = \Lambda(A \cap B_k) c_k$$

for every Borel $A \subset \mathbb{R} \setminus \{0\}$ and $k \geqslant 0$. It is easy to see that

$$\mathbf{E}[\min(1, \eta^2)\zeta] = \int_{\mathbb{R} \setminus \{0\}} \min(1, x^2) \Lambda(dx).$$

If ξ has bounded variation, then Theorem 4.2 applies and

$$\xi(t) \overset{\mathcal{D}}{\sim} bt + \sum_{i=1}^{\infty} \eta_i \mathbf{1}_{t\zeta_i \geqslant \Gamma_i}$$

provides a LePage representation of ξ on the *whole* $\mathbb{R}_+$, cf. [24] for the LePage representation of Lévy processes on $[0,1]$. The choice of $\varepsilon(t) = \eta \mathbf{1}_{t \geqslant 1}$, $t \geqslant 0$, yields the compound Poisson process $\xi(t)$, which becomes the standard Poisson process if $\eta = 1$ a.s.

The time and the size of the jump of ε may be dependent. For instance, let $\varepsilon(t) = \eta \mathbf{1}_{t \geqslant \eta}$ for a positive random variable η. This random function always satisfies (4.6) and yields the Lévy process

$$\xi(t) = \sum_{i=1}^{\infty} \eta_i \mathbf{1}_{t \geqslant \Gamma_i \eta_i}$$

with the cumulant $\Psi(\lambda) = \mathbf{E}[(1 - e^{\imath \lambda \eta}) \eta^{-1}]$.

EXAMPLE 4.2. If $\varepsilon(t) = \eta t^{1/\alpha}$, where $\alpha \in (0,2)$ and η is a symmetric random variable with $\mathbf{E}|\eta|^{\alpha} < \infty$, then the LePage series (4.5) converges a.s., by Theorem 1.4.2 of [25], to $\xi(t) = bt + \zeta t^{1/\alpha}$ for a symmetric α-stable random variable

ζ, see [18]. If $\alpha = 1$ and $b = 0$, then $\xi(t) = \zeta t$ for the Cauchy random variable ζ. This yields a time-stable process with stationary increments, which is not a Lévy process. If $\alpha < 1$, the symmetry of η is not required for the convergence of the LePage series and ζ is strictly α-stable by Theorem 1.4.5 in [25].

EXAMPLE 4.3. Choosing ε to be a stochastic process with stationary increments yields examples of time-stable processes with stationary increments which are not Lévy processes. For instance, let ε be the fractional Brownian motion with Hurst parameter $H \in \left(\frac{1}{2}, 1\right)$. Then (4.4) holds since

$$\mathbf{E} \int_0^1 \min\left(1, \varepsilon(t)^2\right) t^{-2} dt \leqslant \int_0^1 \mathbf{E}\varepsilon(t)^2 t^{-2} dt = \int_0^1 t^{2H-2} dt < \infty.$$

EXAMPLE 4.4 (*Sub-stable processes*). Let $\varepsilon(t) = \xi(t^{1/\alpha})$, $t \geqslant 0$, for $\alpha \in (0, 1)$ and a time-stable process ξ such that $\mathbf{E}|\xi(1)| < \infty$. Then (4.6) holds since

$$\int_0^1 \mathbf{E}|\xi(t^{1/\alpha})| t^{-2} dt = \mathbf{E}|\xi(1)| \int_0^1 t^{1/\alpha - 2} dt < \infty.$$

By conditioning on $\{\Gamma_i\}$ and using Proposition 2.1, one obtains

$$\sum_{i=1}^{\infty} \varepsilon_i(\Gamma_i^{-1} t) = \sum_{i=1}^{\infty} \xi_i(\Gamma_i^{-1/\alpha} t^{1/\alpha}) \overset{\mathcal{D}}{\sim} \xi\big(t^{1/\alpha} \sum_{i=1}^{\infty} \Gamma_i^{-1/\alpha}\big) = \xi(t^{1/\alpha}\zeta)$$

for a strictly α-stable non-negative random variable ζ independent of ξ. Then the LePage series (4.5) yields the process $X(t) = \xi(t^{1/\alpha}\zeta)$, $t \geqslant 0$, where ξ is time-stable and ζ is a positive strictly α-stable random variable independent of ξ, with $\alpha \in (0, 1)$. The process X is called *sub-stable* in view of the construction of sub-stable random elements in [25], Section 1.3.

EXAMPLE 4.5 (*Subordination by time-stable processes*). Let ξ be a non-decreasing time-stable process that admits the LePage representation (4.5) with $b = 0$. If $\{X_i, i \geqslant 1\}$ are i.i.d. copies of a Lévy process X independent of ξ, then

$$\sum_{i=1}^{\infty} X_i\big(\varepsilon_i(\Gamma_i^{-1} t)\big)$$

is the LePage representation of the time-stable process $X\big(\xi(t)\big)$. This is seen by conditioning upon ε_i and $\{\Gamma_i, i \geqslant 1\}$ and noticing that X is stochastically continuous. The time-stability property of $X\big(\xi(t)\big)$ is proved in [8], Theorem 3.6, directly by computing the characteristic function.

EXAMPLE 4.6 (*Random convex broken lines*). Consider $\varepsilon(t) = (t-1)_+$, i.e. the positive part of $(t-1)$. Then the graph of ξ is the continuous convex broken

line with vertices at $(0,0)$ and at

$$\big(\Gamma_n, \Gamma_n \sum_{i=1}^{n} \Gamma_i^{-1} - n\big), \quad n \geqslant 1.$$

In order to obtain a differentiable curve, it is possible to use $\varepsilon(t) = (t-1)_+^{\beta}$ for $\beta > 1$.

5. TIME-STABLE STEP FUNCTIONS

Assume that ξ is a *pure jump* time-stable process, i.e. its paths are *càdlàg* piecewise constant functions with finitely many jumps in each finite interval in $[0,\infty)$ and a.s. vanishing at zero. In view of the assumed stochastic continuity and Lemma 1.6.2 of [27], the jump times of ξ have non-atomic distributions. The jump part of any *càdlàg* time-stable process is also time-stable by noticing that the jump part of the sum of two independent stochastic processes with non-atomic distribution of jump times is equal to the sum of their jump parts. This also applies to the process of jumps larger than $\delta > 0$ in absolute value.

PROPOSITION 5.1. *The time of the first jump of a non-degenerate* càdlàg *pure-jump time-stable process has an exponential distribution.*

P r o o f. Observe that the time of the first jump of the sum of n independent processes equals the minimum of the first jump times $\tau_1, \ldots, \tau_n$ of all summands. Then (1.1) implies that $n^{-1}\tau$ has the same distribution as the minimum of n i.i.d. copies of τ and so characterises the exponential distribution. ∎

The time of the second jump is not necessarily distributed as the sum of two independent exponential random variables since the times between jumps may be dependent and the waiting time between the first and the second jump is no longer exponentially distributed in general.

Let $\mathcal{M}\big((0,\infty)\times\mathbb{R}\big)$ denote the family of marked point configurations on $(0,\infty)$ with marks from $\mathbb{R}$. A *marked point process* is a random element in the product space $\mathcal{M}\big((0,\infty)\times\mathbb{R}\big)$, see [5], Section 6.4. The successive ordered jump times $\{\tau_k\}$ and the jump heights $\{m_k\}$ of a pure jump time-stable process ξ form the marked point process $M = \{(\tau_k, m_k), k \geqslant 1\}$, so that

$$\xi(t) = \sum_{\tau_k \leqslant t} m_k, \quad t \geqslant 0.$$

The sum is finite for every t since the process is assumed to have only a finite number of jumps in any bounded interval. This construction introduces a correspondence between pure jump processes and marked point processes. Note that M is a random closed (and locally finite) set in $(0,\infty)\times\mathbb{R}$, see [20]. The process

ξ is compound Poisson if and only if M is an independently marked homogeneous Poisson process, i.e. the jump times form a homogeneous Poisson process on $(0,\infty)$, while the jump sizes are i.i.d. random variables independent of the jump times.

Scaling the argument of a pure jump process ξ can be rephrased in terms of scaling the marked point process M corresponding to ξ, so that $a \circ \xi$ corresponds to the marked point process

$$a^{-1} \circ M = \{(a^{-1}\tau_k, m_k) : \ k \geqslant 1\}.$$

The sum of independent pure jump processes corresponds to the superposition of the corresponding marked point processes. The next result relates the time-stability property to the union stability of random sets (see [20], Section 4.1.3); it immediately follows from (1.1).

PROPOSITION 5.2. *A stochastically continuous pure jump process ξ is time-stable if and only if its corresponding marked point process M is a union-stable random closed set in the sense that*

$$M_1 \cup \ldots \cup M_n \overset{\mathcal{D}}{\sim} n^{-1} \circ M \tag{5.1}$$

for each $n \geqslant 2$, where $M_1, \ldots, M_n$ are independent copies of M.

COROLLARY 5.1. *A stochastically continuous pure jump process ξ is time-stable if and only if $\xi = \xi_+ - \xi_-$ for the pair of stochastically continuous pure jump processes (ξ_+, ξ_-) that form a pure jump time-stable process with values in $\mathbb{R}^2_+$.*

Proof. For $(\tau, m) \in (0,\infty) \times \mathbb{R}$, let $f(\tau, m) = (\tau, m_+, m_-)$, with m_+ and m_- being the positive and negative parts of $m \in \mathbb{R}$. Then M satisfies (5.1) if and only if $f(M)$ satisfies the analogue of (5.1) with the scaling along the first coordinate. Finally, this property of $f(M)$ is a reformulation of the time stability of (ξ_+, ξ_-), where ξ_+ is the sum of all positive jumps of ξ and ξ_- is the sum of all negative jumps. ∎

THEOREM 5.1. *A stochastically continuous pure jump process ξ is time-stable if and only if*

$$\xi(t) \overset{\mathcal{D}}{\sim} \sum_{i=1}^{\infty} \varepsilon_i(\Gamma_i^{-1} t), \quad t \geqslant 0, \tag{5.2}$$

where $\{\Gamma_i, i \geqslant 1\}$ form a homogeneous unit intensity Poisson point process on $(0,\infty)$, and $\{\varepsilon_i, i \geqslant 1\}$ are independent copies of a random step function ε defined on $[0,\infty)$ which is independent of $\{\Gamma_i\}$ and satisfies (4.6).

P r o o f. The sufficiency is immediate and follows from Theorem 4.2. For the necessity, consider the map f and the random set M from the proof of Corollary 5.1 and note that $f(M)$ is an infinitely divisible point process on $(0,\infty)\times\mathbb{R}^2_+$. It is well known (see e.g. [6], Theorem 10.2.V) that such infinitely divisible marked point process can be represented as a superposition of point configurations that build a Poisson point process on $\mathcal{M}\big((0,\infty)\times\mathbb{R}^2_+\big)$. The unique intensity measure $\tilde{Q}$ of this Poisson process is called the KLM measure of M. This measure can also be viewed as the Lévy measure, see [7], Corollary 6.9.

Each point configuration from $\mathcal{M}\big((0,\infty)\times\mathbb{R}^2_+\big)$ corresponds to a pure jump function. The push-forward of $\tilde{Q}$ under this correspondence is the Lévy measure of (ξ_+,ξ_-) that is supported by pure jump (and so *càdlàg*) functions. Since the components of (ξ_+,ξ_-) are non-negative, Remark 4.4 yields its representation as

$$\big(\xi_+(t),\xi_-(t)\big)\overset{\mathcal{D}}{\sim}\sum_{i=1}^{\infty}\big(\varepsilon_i'(\Gamma_i^{-1}t),\varepsilon_i''(\Gamma_i^{-1}t)\big),$$

so that ξ admits the series representation (5.2) with $\varepsilon=\varepsilon'-\varepsilon''$. ∎

REMARK 5.1. *In the classical LePage series for random vectors, it is possible to scale the directional component to bring its norm to one. However, it is not possible in general to rescale the argument of* $\{\varepsilon_i, i\geqslant 1\}$ *from* (5.2) *in order to ensure that each function has the first jump at time one.*

REMARK 5.2. *It is possible to derive Theorem* 5.1 *from the LePage representation of the marked point process* M *as the union of clusters corresponding to the Poisson cluster process determined by* $\tilde{Q}$*. The corresponding series representation then becomes*

$$M=\bigcup_{i=1}^{\infty}\Gamma_i\circ E_i,$$

where $\{E_i, i\geqslant 1\}$ *is a point process on* $\mathcal{M}\big((0,\infty)\times\mathbb{R}^2_+\big)$ *with the intensity measure* $\tilde{Q}$*.*

If ε has a single jump only, then (5.2) yields a Lévy process, see Example 4.1.

EXAMPLE 5.1. Let $\varepsilon(t)=[t]$ be the integer part of t. Then

$$\xi(t)=\sum_{k=1}^{\infty}N(t/k),$$

where $N(t)$ is the Poisson process. For every $t\geqslant 0$, the series consists of a finite number of summands and so converges almost surely. Note that $\xi(t)$ is not integrable for $t>0$. The jump sizes of ξ are always one, and the jump times form a point process on $\mathbb{R}_+$ obtained as the superposition of the set of natural numbers scaled by Γ_i, $i\geqslant 1$.

Acknowledgments. The authors are grateful to Zakhar Kabluchko and Rolf Riedi for discussions and hints at various stages of this work, and to the referee for suggesting some changes. The second author is grateful to Steve Evans for insightful discussions concerning general LePage series.

REFERENCES

[1] M. Barczy, P. Kern, and G. Pap, *Dilatively stable stochastic processes and aggregate similarity*, Aequationes Math. 89 (6) (2015), pp. 1485–1507.
[2] A. Basse-O'Connor and J. Rosiński, *On the uniform convergence of random series in Skorohod space and representations of càdlàg infinitely divisible processes*, Ann. Probab. 41 (6) (2013), pp. 4317–4341.
[3] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*, Cambridge University Press, Cambridge 1989.
[4] L. Chaumont and M. Yor, *Exercises in Probability: A Guided Tour from Measure Theory to Random Processes, via Conditioning*, second edition, Cambridge University Press, Cambridge 2012.
[5] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*, second edition, Springer, New York 2003.
[6] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*, second edition, Springer, New York 2008.
[7] Y. Davydov, I. Molchanov, and S. Zuyev, *Strictly stable distributions on convex cones*, Electron. J. Probab. 13 (2008), paper no. 11, pp. 259–321.
[8] K. Es-Sebaiy and Y. Ouknine, *How rich is the class of processes which are infinitely divisible with respect to time?*, Statist. Probab. Lett. 78 (5) (2008), pp. 537–547.
[9] S. N. Evans and I. Molchanov, *The semigroup of metric measure spaces and its infinitely divisible probability measures*, Trans. Amer. Math. Soc. 369 (3) (2017), pp. 1797–1834.
[10] Ĭ. Ī. Gīhman and A. V. Skorohod, *The Theory of Stochastic Processes. I*, Grundlehren Math. Wiss. 210, Springer, New York–Heidelberg 1974.
[11] A. Hakassou and Y. Ouknine, *A note on α-IDT processes*, technical report, arXiv math:1207.0874, 2012.
[12] A. Hakassou and Y. Ouknine, *IDT processes and associated Lévy processes with explicit constructions*, Stochastics 85 (6) (2013), pp. 1073–1111.
[13] E. Iglói, *Dilative Stability*, Ph.D. thesis, University of Debrecen, Debrecen 2008.
[14] E. Iglói and M. Barczy, *Path properties of dilatively stable processes and singularity of their distributions*, Stoch. Anal. Appl. 30 (5) (2012), pp. 831–848.
[15] Z. Kabluchko and S. Stoev, *Stochastic integral representations and classification of sum- and max-infinitely divisible processes*, Bernoulli 22 (1) (2016), pp. 107–142.
[16] I. Kaj, *Limiting fractal random processes in heavy-tailed systems*, in: *Fractals in Engineering: New Trends in Theory and Applications*, J. Lévy-Véhel and E. Lutton (Eds.), Springer, London 2005, pp. 199–217.
[17] R. LePage, M. Woodroofe, and J. Zinn, *Convergence to a stable distribution via order statistics*, Ann. Probab. 9 (4) (1981), pp. 624–632.
[18] R. Mansuy, *On processes which are infinitely divisible with respect to time*, arXiv math:0504408, 2005.
[19] G. Maruyama, *Infinitely divisible processes*, Teor. Verojatn. Primen. 15 (1970), pp. 3–23.
[20] I. Molchanov, *Theory of Random Sets*, Springer, London 2005.
[21] M. D. Penrose, *Semi-min-stable processes*, Ann. Probab. 20 (3) (1992), pp. 1450–1463.
[22] J. Rosiński, *On path properties of certain infinitely divisible processes*, Stochastic Process. Appl. 33 (1) (1989), pp. 73–87.

[23] J. Rosiński, *On series representations of infinitely divisible random vectors*, Ann. Probab. 18 (1) (1990), pp. 405–430.
[24] J. Rosiński, *Series representations of Lévy processes from the perspective of point processes*, in: *Lévy Processes*, O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick (Eds.), Birkhäuser, Boston, MA, 2001, pp. 401–415.
[25] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York 1994.
[26] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge 1999.
[27] D. S. Silvestrov, *Limit Theorems for Randomly Stopped Stochastic Processes*, Springer, London 2004.

Christoph Kopp
School of Agricultural, Forest and Food Sciences
Bern University of Applied Sciences
Länggasse 85, CH-3052 Zollikofen
Switzerland
E-mail: christoph.kopp@bfh.ch

Ilya Molchanov
Institute of Mathematical Statistics
and Actuarial Science
University of Bern
Sidlerstrasse 5, CH-3012 Bern
Switzerland
E-mail: ilya.molchanov@stat.unibe.ch

Received on 7.10.2015;
revised version on 15.3.2017

Probability and Mathematical Statistics 38, z. 2, 2018

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 317–357
doi:10.19195/0208-4147.38.2.5

PREDICTION INTERVALS AND REGIONS FOR MULTIVARIATE TIME SERIES MODELS WITH SIEVE BOOTSTRAP

BY

ROMAN RÓŻAŃSKI (WROCŁAW), GRZEGORZ CHŁAPIŃSKI (WROCŁAW),
MARCIN HŁAWKA (WROCŁAW),
KRZYSZTOF JAMRÓZ (WROCŁAW),
MACIEJ KAWECKI (WROCŁAW), AND ADAM ZAGDAŃSKI (WROCŁAW)

Abstract. In the paper, the construction of unconditional bootstrap prediction intervals and regions for some class of second order stationary multivariate linear time series models is considered. Our approach uses the sieve bootstrap procedure introduced by Kreiss (1992) and Bühlmann (1997). Basic theoretical results concerning consistency of the bootstrap replications and the bootstrap prediction regions are proved. We present a simulation study comparing the proposed bootstrap methods with the Box–Jenkins approach.

2010 AMS Mathematics Subject Classification: Primary: 62M10, 62M20; Secondary: 62G09, 62G15, 62G05.

Key words and phrases: Multivariate time series models, vector of time series, sieve bootstrap, prediction regions, simultaneous prediction intervals.

1. INTRODUCTION

Determination of forecasts of time series future values based on previous observations is an extremely important – from a practical point of view – part of statistical data analysis. Since the high-speed personal computers have appeared, we can even deal with the prediction for a large number of dimensions of data. Methods of determining the prediction for the future and unknown observations are now frequently used in the world around us. They have a wide range of applications, both to predict the behavior of stock prices, stock indices, interest rates, and similar financial market and economic data ([27], [28]) as well in predicting the data of the general nature and geographic scope. For instance, the vector autoregressive models VAR were used by Di Battista et al. [11] in modeling the diversity of the population of some species in their natural environment, and Mirmirani and Li [22] used VAR models to predict the oil prices.

In this paper we present the construction of bootstrap prediction regions for some class of second order stationary multivariate linear time series models. We consider both hybrid bootstrap and bootstrap-t methods. Using the Bonferroni inequality, we can construct a multivariate bootstrap prediction cube, i.e. we construct a prediction interval for each coordinate. We consider also bootstrap simultaneous prediction intervals based on extreme statistics. They are an alternative to the bootstrap prediction intervals based on the Bonferroni correction. The main advantages of the bootstrap methods are nonparametricity (no specific assumptions about the form of the model) and easiness to implement. Thus, the bootstrap methods are a natural alternative to the methods under general asymptotic statistical considerations (e.g. with the popular assumption of the normality of noise distribution).

We consider the most popular and nonparametric method for constructing the replication of time series data, namely the sieve bootstrap. The algorithm was proposed by Kreiss [18] and Bühlmann [9]. Their idea uses a Grenander sieve [14] involving the approximation of infinite-dimensional model by a sequence of finite-dimensional models whose size increases with the number of observations n. For the class of stationary and invertible time series models ($VAR(\infty)$ models), Bühlmann proposed approximation as a sequence of vector autoregressive ($VAR(p)$) models, where $p = p(n)$ increases to infinity at an appropriate rate. In [21] the consistency of sieve bootstrap for general statistics being estimators of parameters in vector autoregressive time series models is considered under assumptions which essentially imply the assumptions imposed on the time series models and the sieve method investigated in our paper (see the assumptions (LA) in Remark 2.1). However, the characterization of asymptotic behavior of parametric estimators obtained by the authors does not cover the problem of asymptotics of VAR sieve bootstrap for predictors and bootstrap prediction regions considered in this article.

In the case of univariate causal linear time series models admitting the $AR(\infty)$ representation, Alonso et al. [1] constructed the sieve bootstrap estimator $X^\star_{T+h}$ of the future value X_{T+h}. They proved that $X^\star_{T+h} \to^{d^\star} X_{T+h}$ in probability, which implies that the bootstrap distribution $F^\star_{X^\star_{T+h}}$ approximates in probability the unknown distribution of $F_{X_{T+h}}$. Further, using the quantiles $Q^\star()$ of the distribution $F^\star_{X^\star_{T+h}}$ as bootstrap estimators of quantiles $Q()$ of the distribution $F_{X_{T+h}}$, the authors construct a prediction interval for the future value X_{T+h}. In general, the distribution $F^\star_{X^\star_{T+h}}$ and the quantiles $Q^\star()$ are not known. So, in simulations, the authors use the Monte Carlo method to approximate $F^\star_{X^\star_{T+h}}$ and $Q^\star()$. Unfortunately, they did not argue that the Monte Carlo approximation of $Q^\star()$ is a consistent estimator of the quantile $Q()$. At least, one should mention the results of Shi et al. [26].

Since the sieve approximation $p(n)$ is charged with serious error as an estimator of possibly finite but unknown order p of the considered univariate time series

model, Alonso et al. [2] joined the sieve bootstrap and the moving block bootstrap to select the order $p^\star$ and to introduce model uncertainty in procedures of resampling. Unfortunately, no proofs of consistency are given and the simulation results are restricted to the Gaussian errors.

In this article, we construct sieve bootstrap prediction regions for causal linear, invertible ($VAR(\infty)$) multivariate time series, approximating the prediction error by bootstrap replications of the prediction error. There are two ways (described in Section 4.2) of the bootstrap replicating of the prediction error. One can see that, in the univariate case, prediction hybrid bootstrap intervals constructed by the first method are identical with bootstrap prediction intervals constructed in [1] but it does not happen when we construct bootstrap-t (studentized) prediction intervals.

In this work, we generalize the results obtained in [13] from $VAR(p)$ models with finite but unknown order p to some class of second order stationary multivariate linear, $VAR(\infty)$ time series models.

It is worth noting that we have constructed consistent unconditional bootstrap prediction regions and the results allow us to use these prediction regions as effective and useful tools for testing and selection of models. Moreover, it follows from the theorems proved in the present article that the constructed bootstrap prediction regions are good approximations of prediction regions constructed on the base of unknown optimal linear predictors.

2. MODEL AND ASSUMPTIONS

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a second order stationary k-dimensional vector process with mean $EX_t = 0$ and the autocovariance function $\gamma(j) = EX_{t+j}X_t^T$, where $\mathbb{Z}$ is the set of all integers. We assume also that the process $\{X_t\}_{t\in\mathbb{Z}}$ is purely stochastic. Thus, using Wold's Decomposition Theorem (see [3] or [20]), we can represent $\{X_t\}_{t\in\mathbb{Z}}$ as an infinite vector moving average process $VMA(\infty)$,

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad \psi_0 = I_k, \tag{2.1}$$

where $\sum_{j=0}^{\infty} \|\psi_j\|^2 < \infty$ and $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a vector white noise process with the covariance matrix $E\epsilon_t\epsilon_t^T = \Sigma$ (Σ is invertible). Additionally, we assume that the process $\{X_t\}_{t\in\mathbb{Z}}$ is invertible. Thus, it can be represented as an infinite vector autoregressive process $VAR(\infty)$:

$$X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \epsilon_t, \tag{2.2}$$

where $\sum_{j=1}^{\infty} \|\phi_j\|^2 < \infty$.

Further, we will use the following assumptions:

(A1) $X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, $\psi_0 = I_k, t \in \mathbb{Z}$, where

$$\{\epsilon_t\}_{t\in\mathbb{Z}} = \{(\epsilon_{t,1}, \ldots, \epsilon_{t,i}, \ldots, \epsilon_{t,k})^T\}_{t\in\mathbb{Z}}$$

is an i.i.d. sequence such that $E\epsilon_t = 0$, $E\epsilon_t\epsilon_t^T = \Sigma$ and $E|\epsilon_{t,i}\epsilon_{t,j}\epsilon_{t,l}\epsilon_{t,s}| < \infty$ for $i, j, l, s = 1, \ldots, k$ and all t.

(A2) $\big|\det\big(\Psi(z)\big)\big| = \big|\det\big(\sum_{j=0}^{\infty} \psi_j z^j\big)\big| > 0$ for $|z| \leqslant 1$ and $\sum_{j=0}^{\infty} j^r \|\psi_j\| < \infty$ for some $r \in \mathbb{N}$.

(B) $p = p(n) \to \infty$ as $n \to \infty$ and $\widehat{\Phi}_p = [\widehat{\phi}_{1,n}, \ldots, \widehat{\phi}_{p,n}]^T$ satisfies the Yule–Walker equations, i.e.

$$\widehat{\Gamma}_p \widehat{\Phi}_p = \widehat{\gamma}_p, \tag{2.3}$$

where $\widehat{\Gamma}_p = [\widehat{\gamma}(i-j)]_{i,j=1}^p$, $\widehat{\gamma}_p = [\widehat{\gamma}(1), \ldots, \widehat{\gamma}(p)]^T$, and $\widehat{\gamma}(\cdot)$ is the sample autocovariance function,

$$\widehat{\gamma}(j) = \frac{1}{n}\sum_{t=1}^{n-j}(X_{t+j} - \overline{X}_n)(X_t - \overline{X}_n)^T, \quad \widehat{\gamma}(-j) = \widehat{\gamma}^T(j),$$

where $\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$.

REMARK 2.1. *In the sequel, the following list of assumptions, called* (LA), *will be also imposed*:

- *assumption* (A1),
- *assumption* (A2) *with* $r \geqslant 1$,
- *assumption* (B) *with* $p(n) = o\big((n/\log n)^{1/(2r+2)}\big)$ *and* $r \geqslant 1$.

Under the assumptions (LA) we prove the main theorems on bootstrap consistency.

3. SIEVE BOOTSTRAP ALGORITHM

The invertibility of the process $\{X_t\}_{t\in\mathbb{Z}}$ implies the $VAR(\infty)$ representation of the process which is crucial in the idea of the sieve bootstrap (see e.g. [9]). The sieve bootstrap algorithm uses the idea of the Grenander method of sieve (see [14]). Namely, we approximate $VAR(\infty)$ given by (2.2) by the sequence of $VAR(p)$ models, where $p = p(n)$ is a sequence growing to infinity sufficiently slow with the sample size n (assumption (B)).

REMARK 3.1. *It is also possible to approximate the process* $\{X_t\}_{t\in\mathbb{Z}}$ *given by* (2.1) *and construct sieve as a sequence of finite* $VMA(q)$, *where* $q = q(n) \to \infty$ (see [8]).

Let $X_1, \ldots, X_n$ be the observations of the process $\{X_t\}_{t\in\mathbb{Z}}$. We describe the sieve bootstrap algorithm in the following steps.

S t e p 1. Choose the approximation order $p = p(n)$ using FPE, the final prediction error (see [20]),

$$FPE(p) = \left(\frac{n+pk+1}{n-pk-1}\right)^k \det(\widehat{\Sigma}_p),$$

where $\widehat{\Sigma}_p$ is an estimator of the covariance matrix Σ of the white noise $\{\epsilon_t\}_{t\in\mathbb{Z}}$ in the model $VAR(p)$. Further, we choose $p = p(n)$, which minimizes $FPE(p)$ and $p \in \{p_{\min}(n), \dots, p_{\max}(n)\}$, where $p_{\min} = \log_{10} n$, $p_{\max} = 10 \log_{10} n$.

S t e p 2. Estimate the coefficients $\phi_{1,n}, \dots, \phi_{p,n}$ of the model $VAR(p)$ using the Yule–Walker method (assumption (B)) and obtain estimators $\widehat{\phi}_{1,n}, \dots, \widehat{\phi}_{p,n}$.

S t e p 3. Compute the residuals

$$\widehat{\epsilon}_t = X_t - \sum_{j=1}^{p} \widehat{\phi}_{j,n} X_{t-j}, \quad t = p+1, \dots, n.$$

S t e p 4. Center the residuals

$$\widetilde{\epsilon}_t = \widehat{\epsilon}_t - \frac{1}{n-p} \sum_{j=p+1}^{n} \widehat{\epsilon}_t, \quad t = p+1, \dots, n,$$

and draw bootstrap residuals ϵ_t^* from the empirical cumulative distribution function $\widehat{F}_{\epsilon,n}$, where

$$\widehat{F}_{\epsilon,n}(x) = \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{1}\{\widetilde{\epsilon}_t \leqslant x\}, \tag{3.1}$$

and $\leqslant$ denotes the relation of product order (partial order) in $\mathbb{R}^k$.

S t e p 5. Define bootstrap replications $X_1^*, \dots, X_n^*$ of $X_1, \dots, X_n$ by

$$X_t^* = \sum_{j=1}^{p} \widehat{\phi}_{j,n} X_{t-j}^* + \epsilon_t^*, \quad t = 1, \dots, n.$$

In practice, we can generate replications $X_1^*, \dots, X_n^*$ starting the recursion from some initial values, e.g. $X_t^* = \epsilon_t^*$, $t = 0, -1, \dots, -p+1$.

S t e p 6. Generate bootstrap replications $\widehat{\phi}_{j,n}^*$ of the Yule–Walker estimators $\widehat{\phi}_{j,n}$, where $j = 1, 2, \dots, p(n)$.

The bootstrap construction induces a conditional probability measure P^* given the sample $X_1, \dots, X_n$. In the sequel, all quantities with respect to P^* will be endowed with asterisk *.

REMARK 3.2. *The FPE criterion, used in Step* 1, *is asymptotically equivalent to the AIC criterion* (see [20]) *used in* [9].

REMARK 3.3. *In Step* 2, *the Yule–Walker estimators can be computed by using Whittle's algorithm, which is a multivariate version of the Durbin–Levinson algorithm* (see [7]).

In the sequel, we will use the following notation for the coefficients of autoregressive models and their related moving average models. For the stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ given by (2.1) or (2.2) the autoregressive coefficients ϕ_j and the moving average coefficients ψ_j are related by

$$\Phi(z) = I_k - \sum_{j=1}^{\infty} \phi_j z^j, \quad \Phi^{-1}(z) = \Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j,$$

$$\psi_0 = I_k, \quad \text{and} \quad \psi_j = \sum_{i=1}^{j} \phi_i \psi_{j-i}, \quad j = 1, 2, \ldots$$

In the model $VAR(p)$, approximating the model (2.2), we denote by $\phi_{j,n}$ the autoregressive coefficients which fulfill the theoretical Yule–Walker equations (2.3) (assumption (B), where the sample autocovariance function is replaced by $\gamma(j) = EX_{t+j}X_t^T$), and by $\psi_{j,n}$ the moving average coefficients:

$$\Phi_n(z) = I_k - \sum_{j=1}^{p} \phi_{j,n} z^j, \quad \Phi_n^{-1}(z) = \Psi_n(z) = \sum_{j=0}^{\infty} \psi_{j,n} z^j, \tag{3.2}$$

$$\psi_{0,n} = I_k, \quad \text{and} \quad \psi_{j,n} = \sum_{i=1}^{j} \phi_{i,n} \psi_{j-i,n}, \quad j = 1, 2, \ldots$$

We write the Yule–Walker estimators given in assumption (B):

$$\widehat{\Phi}_n(z) = I_k - \sum_{j=1}^{p} \widehat{\phi}_{j,n} z^j, \quad \widehat{\Phi}_n^{-1}(z) = \widehat{\Psi}_n(z) = \sum_{j=0}^{\infty} \widehat{\psi}_{j,n} z^j, \tag{3.3}$$

$$\widehat{\psi}_{0,n} = I_k, \quad \text{and} \quad \widehat{\psi}_{j,n} = \sum_{i=1}^{j} \widehat{\phi}_{i,n} \widehat{\psi}_{j-i,n}, \quad j = 1, 2, \ldots$$

Let $\widehat{\phi}^*_{j,n}$ and $\widehat{\psi}^*_{j,n}$ be the bootstrap replications of Yule–Walker estimators $\widehat{\phi}_{j,n}$, $\widehat{\psi}_{j,n}$,

$$\widehat{\Phi}^*_n(z) = I_k - \sum_{j=1}^{p} \widehat{\phi}^*_{j,n} z^j, \quad \widehat{\Phi}^{*-1}_n(z) = \widehat{\Psi}^*_n(z) = \sum_{j=0}^{\infty} \widehat{\psi}^*_{j,n} z^j, \tag{3.4}$$

where $\widehat{\psi}^*_{0,n} = I_k$.

REMARK 3.4. *The assumptions* (A1) *and* (A2) *provide us with the correctness of defined functions* $\Psi(z)$ *and* $\Phi(z)$, *which means that the series* $\sum_{j=0}^{\infty}\psi_j z^j$ *and* $\sum_{j=1}^{\infty}\phi_j z^j$ *are convergent for* $|z|\leqslant 1$ *and* $\Phi^{-1}(z)=\Psi(z)$. *The properties of the Yule–Walker equations ensure the correctness of defined functions* $\Psi_n(z)$, $\widehat{\Psi}_n(z)$ *and* $\widehat{\Psi}_n^*(z)$ (*solutions of the Yule–Walker equations give us a causal model*, see [7]).

4. PREDICTION REGIONS

4.1. Linear predictors and Gaussian prediction regions. Forecasting the future values X_{n+h}, $h=1,2,\ldots$, is a very common task in the statistical analysis of time series. For the second order stationary process we can construct the best linear predictor, in the mean square sense, as an orthogonal projection of X_{n+h} onto a closed subspace $\overline{\mathrm{sp}}\{X_1,\ldots,X_n\}$ of $L^2(\Omega,\mathcal{F},\mathcal{P})$ (see [20]). We can represent the h-step predictor as

$$\mathrm{Proj}_n X_{n+h}=\mathrm{Proj}_{\overline{\mathrm{sp}}\{X_1,\ldots,X_n\}}X_{n+h}$$

and the mean squared prediction error as

$$\Sigma_X(h)=E\left(X_{n+h}-\mathrm{Proj}_n X_{n+h}\right)\left(X_{n+h}-\mathrm{Proj}_n X_{n+h}\right)^T. \tag{4.1}$$

Assuming that (A1) holds, we have also

$$\Sigma_X(h)=\sum_{j=0}^{h-1}\psi_j\Sigma\psi_j^T. \tag{4.2}$$

The mean squared prediction error can be obtained by using the multivariate version of the innovations algorithm (see [7]).

REMARK 4.1. *For the autoregressive model* $VAR\big(p=p(n)\big)$,

$$X_t=\sum_{j=1}^{p}\phi_{j,n}X_{t-j}+\epsilon_{t,n},$$

approximating the model $VAR(\infty)$, *the best linear predictor* (*in the mean square sense*) *and the mean squared error matrix have the forms*

$$\mathrm{Proj}_n X_{n+h,VAR(p)}=\sum_{j=1}^{p}\phi_{j,n}\mathrm{Proj}_n X_{n+h-j,VAR(p)},$$

$$\Sigma_{n,X}(h)=\sum_{j=0}^{h-1}\psi_{j,n}\Sigma\psi_{j,n}^T,$$

where $\mathrm{Proj}_n X_{n+j,VAR(p)}=X_{n+j}$ *for* $j\leqslant 0$, $\Sigma=E\epsilon_t\epsilon_t^T$ *and coefficients* $\psi_{j,n}$ *are given by the recursive equations*

$$\psi_{0,n}=I_k \quad \textit{and} \quad \psi_{j,n}=\sum_{i=1}^{j}\phi_{i,n}\psi_{(j-i),n},\quad j=1,2,\ldots$$

From the continuity of the projection operator in L^2 it follows that

$$\mathrm{Proj}_n X_{n+h} - \mathrm{Proj}_n X_{n+h,VAR(p)} \xrightarrow{L^2} 0 \quad \textit{and} \quad \Sigma_{n,X}(h) \to \Sigma_X(h) \textit{ as } n \to \infty.$$

Since the form of the best linear predictor (in the mean square sense) and the mean squared error matrix depend on unknown parameters of a model considered, we will use an estimator $\widehat{\mathrm{Proj}}_n X_{n+h} = \widehat{X}_{n+h}$ of the best linear predictor $\mathrm{Proj}_n X_{n+h}$ with appropriately chosen estimators of parameters of the model.

The most common method of constructing the prediction regions for X_{n+h}, $h = 1, 2, \ldots$, is the Box–Jenkins method. This method assumes that the prediction error has at least asymptotically normal distribution

$$X_{n+h} - \widehat{X}_{n+h} \sim \mathcal{N}\big(0, \Sigma_X(h)\big).$$

Thus, the quadratic form below has at least the asymptotically χ-square distribution with k degrees of freedom:

$$(X_{n+h} - \widehat{X}_{n+h})^T \Sigma_X^{-1}(h)(X_{n+h} - \widehat{X}_{n+h}) \sim \chi^2(k).$$

So, the prediction region for X_{n+h}, $h = 1, 2, \ldots$, with nominal confidence level $1 - \alpha$ has a shape of k-dimensional ellipse

$$\mathcal{E}_{B-J}(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T \Sigma_X^{-1}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant \chi^2_{1-\alpha}(k)\}, \tag{4.3}$$

where $\chi^2_\alpha(k)$ is an α quantile of χ-square distribution with k degrees of freedom. We can also use the Bonferroni inequality and construct a k-dimensional prediction cube

$$I_{B-J}(h) = \{X_{n+h,j} \in [\widehat{X}_{n+h,j} + \sigma_{X,j}(h) z_{\alpha/(2k)}, \widehat{X}_{n+h,j} + \sigma_{X,j}(h) z_{1-\alpha/(2k)}],\ j = 1, \ldots, k\}, \tag{4.4}$$

where z_α is an α quantile of the normal distribution $\mathcal{N}(0, 1)$ and $\sigma_{X,j}(h)$ is a square root of the jth diagonal element of the mean squared error matrix $\Sigma_X(h)$.

4.2. Bootstrap prediction regions. The bootstrap methods are very common in the problem of constructing the confidence intervals (see e.g. [12]). The same idea can be used in the construction of prediction regions for X_{n+h}, $h = 1, 2, \ldots$ Namely, we may apply the sieve bootstrap method and generate bootstrap replications $X_1^*, \ldots, X_n^*$ using observations $X_1, \ldots, X_n$. Then, we construct the bootstrap replications of the future observations on the base of the $VAR(p)$ approximation

$$\begin{aligned} &\text{for } h = 1\colon \quad X^*_{n+1} = \sum_{j=1}^{p} \widehat{\phi}^*_{j,n} X_{n+1-j} + \epsilon^*_{n+1}, \\ &\text{for } h > 1\colon \quad X^*_{n+h} = \sum_{j=1}^{h-1} \widehat{\phi}^*_{j,n} X^*_{n+h-j} + \sum_{j=h}^{p} \widehat{\phi}^*_{j,n} X_{n+h-j} + \epsilon^*_{n+h}, \end{aligned} \tag{4.5}$$

where $\widehat{\phi}^*_{j,n}$ is a bootstrap replication of the estimator $\widehat{\phi}_{j,n}$. It is worth noting that we applied a modification of a standard procedure, proposed by Bühlmann [9] for the one-dimensional case, to generate the future bootstrap observations in the equation (4.5). Namely, we generated future observations starting recursion from $X_1, \ldots, X_n$, in contrast to the standard method in which the recursion is started from the bootstrap replications $X^*_1, \ldots, X^*_n$. The results of the simulations showed that this type of modification improved the empirical probability of coverage of the bootstrap prediction regions.

Using the $VAR(p)$ approximation, we construct the estimator of the best linear predictor

$$\widehat{X}_{n+h} = \sum_{j=1}^{p(n)} \widehat{\phi}_{j,n} \widehat{X}_{n+h-j}, \tag{4.6}$$

where $\widehat{X}_{n+j} = X_{n+j}, j \leqslant 0$.

We construct the bootstrap prediction regions using two methods: hybrid bootstrap and bootstrap-t.

In the hybrid bootstrap, the unknown distribution of the prediction error

$$H_n(h) = X_{n+h} - \widehat{X}_{n+h} \tag{4.7}$$

can be approximated by two bootstrap variants:

$$H^*_n(h) = X^*_{n+h} - \widehat{X}_{n+h}, \tag{4.8}$$

$$\widetilde{H}^*_n(h) = X^*_{n+h} - \widehat{X}^*_{n+h}, \tag{4.9}$$

where $\widehat{X}^*_{n+h}$ is the bootstrap replication of $\widehat{X}_{n+h}$ given by (4.6). It can be proved that both bootstrap variants of the prediction error (4.7) are consistent. However, we will focus on (4.8) because of its good simulation results. Since the proof of consistency for the bootstrap variant (4.9) goes along the same lines as the proof of consistency for the bootstrap variant (4.8), we will omit it. Thus, the bootstrap prediction cube, constructed by using the Bonferroni inequality, has the form

$$\begin{aligned} I_B(h) = \{X_{n+h,j} \in [&\widehat{X}_{n+h,j} + q^*_{\alpha/(2k),j}, \\ &\widehat{X}_{n+h,j} + q^*_{1-\alpha/(2k),j}],\ j = 1, \ldots, k\}, \end{aligned} \tag{4.10}$$

where $q^*_{\alpha,j}$ is an α quantile of the distribution $X^*_{n+h,j} - \widehat{X}_{n+h,j}$, $j = 1, \ldots, k$.

Using the hybrid bootstrap, we can also create a bootstrap prediction region in the shape of k-dimensional ellipse:

$$\mathcal{E}_B(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T (X_{n+h} - \widehat{X}_{n+h}) \leqslant q^*_{1-\alpha}\}, \tag{4.11}$$

where q^*_α is an α quantile of the distribution $\|H^*_n(h)\|^2$.

In the bootstrap-t method we approximate the studentized unknown distribution of the prediction error

$$T_n(h) = \widehat{\Sigma}_{n,X}^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \tag{4.12}$$

by its bootstrap replication

$$T_n^*(h) = \widehat{\Sigma}_{n,X}^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}), \tag{4.13}$$

where $\widehat{\Sigma}_{n,X}(h)$ is an estimator of the mean squared error matrix $\Sigma_X(h)$, and $\widehat{\Sigma}_X^*(h)$ is a bootstrap replication of the estimator. By Remark 4.1, we define the estimator $\widehat{\Sigma}_{n,X}(h)$ as

$$\widehat{\Sigma}_{n,X}(h) = \sum_{j=0}^{h-1} \widehat{\psi}_{j,n} \widehat{\Sigma} \widehat{\psi}_{j,n}^T, \tag{4.14}$$

where $\widehat{\psi}_{j,n} = \sum_{i=1}^{j} \widehat{\phi}_{i,n} \widehat{\psi}_{(j-i),n}$. Thus, the bootstrap prediction cube, constructed by using the Bonferroni inequality, has the form

$$\begin{aligned} I_{B-t}(h) = \{X_{n+h,j} \in [&\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h) t^*_{\alpha/(2k),j}, \\ &\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h) t^*_{1-\alpha/(2k),j}],\ j = 1, \ldots, k\}, \end{aligned} \tag{4.15}$$

where $t^*_{\alpha,j}$ is an α quantile of the distribution $\widehat{\sigma}_{X,j}^{*-1}(h)(X^*_{n+h,j} - \widehat{X}_{n+h})$ for $j = 1, \ldots, k$.

Using the bootstrap-t, we can also create a bootstrap prediction region in the shape of k-dimensional ellipse:

$$\mathcal{E}_{B-t}(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T \widehat{\Sigma}_{n,X}^{-1}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant t^*_{1-\alpha}\}, \tag{4.16}$$

where t^*_α is an α quantile of the distribution $\|T_n^*(h)\|^2$.

4.3. Bootstrap simultaneous prediction regions based on extreme statistics. In this subsection we present different types of bootstrap confidence regions based on the distribution of minimum and maximum of $X_{n+h} - \widehat{X}_{n+h}$. More precisely, we investigate simultaneous hybrid and studentized confidence regions. For each type we propose left-sided, right-sided and both-sided regions. First, we introduce the following notation for statistics U_{n+h}, V_{n+h} and R_{n+h}:

$$U_{n+h} = \min_{1 \leqslant i \leqslant k} W_{n+h,i} = \min_{1 \leqslant i \leqslant k} (X_{n+h,i} - \widehat{X}_{n+h,i}), \tag{4.17}$$

$$V_{n+h} = \max_{1 \leqslant i \leqslant k} W_{n+h,i} = \max_{1 \leqslant i \leqslant k} (X_{n+h,i} - \widehat{X}_{n+h,i}), \tag{4.18}$$

$$R_{n+h} = \max_{1 \leqslant i \leqslant k} |W_{n+h,i}| = \max_{1 \leqslant i \leqslant k} |X_{n+h,i} - \widehat{X}_{n+h,i}|, \tag{4.19}$$

and for their bootstrap versions:

$$U^*_{n+h} = \min_{1\leqslant i\leqslant k} W^*_{n+h,i} = \min_{1\leqslant i\leqslant k} (X^*_{n+h,i} - \widehat{X}_{n+h,i}), \tag{4.20}$$

$$V^*_{n+h} = \max_{1\leqslant i\leqslant k} W^*_{n+h,i} = \max_{1\leqslant i\leqslant k} (X^*_{n+h,i} - \widehat{X}_{n+h,i}), \tag{4.21}$$

$$R^*_{n+h} = \max_{1\leqslant i\leqslant k} |W^*_{n+h,i}| = \max_{1\leqslant i\leqslant k} |X^*_{n+h,i} - \widehat{X}_{n+h,i}|. \tag{4.22}$$

Let $\widehat{u}^*_{n+h,(\cdot)}$, $\widehat{v}^*_{n+h,(\cdot)}$ and $\widehat{r}^*_{n+h,(\cdot)}$ be Monte Carlo estimators of quantiles of bootstrap distributions of U^*_{n+h}, V^*_{n+h} and R^*_{n+h}. Then, the bootstrap prediction cubes, constructed by using extreme statistics, can be written in the following form:

$$I^{UV}_B(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha/2}, \widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha/2}],\ i = 1,\dots,k\}, \tag{4.23}$$

$$I^{U}_B(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha}, +\infty),\ i = 1,\dots,k\}, \tag{4.24}$$

$$I^{V}_B(h) = \{X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha}],\ i = 1,\dots,k\}, \tag{4.25}$$

$$I^{R}_B(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} - \widehat{r}^*_{n+h,1-\alpha}, \widehat{X}_{n+h,i} + \widehat{r}^*_{n+h,1-\alpha}],\ i = 1,\dots,k\}. \tag{4.26}$$

We can also construct studentized prediction cubes:

$$I^{UV}_{B-t}(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{us}^*_{n+h,\alpha/2}\widehat{\sigma}_{n+h,i}, \widehat{X}_{n+h,i} + \widehat{vs}^*_{n+h,1-\alpha/2}\widehat{\sigma}_{n+h,i}],\ i = 1,\dots,k\}, \tag{4.27}$$

$$I^{U}_{B-t}(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{us}^*_{n+h,\alpha}\widehat{\sigma}_{n+h,i}, +\infty), i = 1,\dots,k\}, \tag{4.28}$$

$$I^{V}_{B-t}(h) = \{X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + \widehat{vs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h,i}], i = 1,\dots,k\}, \tag{4.29}$$

$$I^{R}_{B-t}(h) = \{X_{n+h,i} \in [\widehat{X}_{n+h,i} - \widehat{rs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h,i}, \widehat{X}_{n+h,i} + \widehat{rs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h,i}],\ i = 1,\dots,k\}, \tag{4.30}$$

where $\widehat{us}^*_{n+h,(\cdot)}$, $\widehat{vs}^*_{n+h,(\cdot)}$ and $\widehat{rs}^*_{n+h,(\cdot)}$ are, respectively, Monte Carlo estimators of quantiles of bootstrap distributions of

$$US^*_{n+h} = \min_{1\leqslant i\leqslant k} W^*_{n+h,i} = \min_{1\leqslant i\leqslant k} \left(\frac{X^*_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}^*_{n+h,i}}\right),$$

$$VS^*_{n+h} = \max_{1\leqslant i\leqslant k} W^*_{n+h,i} = \max_{1\leqslant i\leqslant k} \left(\frac{X^*_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}^*_{n+h,i}}\right),$$

$$RS^*_{n+h} = \max_{1\leqslant i\leqslant k} |W^*_{n+h,i}| = \max_{1\leqslant i\leqslant k} \left|\frac{X^*_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}^*_{n+h,i}}\right|,$$

and US^*_{n+h}, VS^*_{n+h} and RS^*_{n+h} are the following bootstrap versions of distributions:

$$US_{n+h} = \min_{1\leqslant i\leqslant k} W_{n+h,i} = \min_{1\leqslant i\leqslant k}\left(\frac{X_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}}\right),$$

$$VS_{n+h} = \max_{1\leqslant i\leqslant k} W_{n+h,i} = \max_{1\leqslant i\leqslant k}\left(\frac{X_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}}\right),$$

$$RS_{n+h} = \max_{1\leqslant i\leqslant k} |W_{n+h,i}| = \max_{1\leqslant i\leqslant k}\left|\frac{X_{n+h,i} - \widehat{X}_{n+h,i}}{\widehat{\sigma}_{n+h,i}}\right|.$$

5. CONSISTENCY OF BOOTSTRAP METHODS

5.1. Representation of the $VAR(\infty)$ model by the moving average model. In this subsection we use the following theorem proved in [15] (see Theorem 7.4.2). The theorem is formulated below but with changed notation, adapted to the present article.

THEOREM 5.1 ([15], Theorem 7.4.2). *If* $\det\big(\Psi(z)\big) \neq 0$ *for* $|z| \leqslant 1$ *and*

$$\sum_{j=0}^{\infty} j^{\lambda}\,\|\psi_j\| < \infty, \quad \lambda > 0,$$

then for $\Phi(z) = \Psi^{-1}(z) = \sum_{j=1}^{\infty}\phi_j z^j$ *we have*

$$\sum_{j=0}^{\infty} j^{\lambda}\,\|\phi_j\| < \infty.$$

By analogical reasoning, we can prove the converse of Theorem 5.1. Namely, we have

COROLLARY 5.1. *The following assertions are equivalent*:
(i) $\det\big(\Phi(z)\big) \neq 0$, $|z| \leqslant 1$, $\sum_{j=0}^{\infty} j^{\lambda}\,\|\phi_j\| < \infty$, $\lambda > 0$.
(ii) $\det\big(\Psi(z)\big) \neq 0$, $|z| \leqslant 1$, $\sum_{j=0}^{\infty} j^{\lambda}\,\|\psi_j\| < \infty$, $\lambda > 0$.

The next lemma is a multivariate version of a lemma given by Bühlmann ([8], Lemma 2.2). For the purpose of this lemma (and only for this lemma) we assume that $\Phi_n(z) = I_k - \sum_{j=1}^{\infty}\phi_{j,n}z^j$ is some deterministic approximation of the function $\Phi(z) = I_k - \sum_{j=1}^{\infty}\phi_j z^j$ (model (2.2)) and we define $\Psi_n(z) = \Phi_n^{-1}(z) = \sum_{j=0}^{\infty}\psi_{j,n}z^j$, where $\psi_{0,n} = I_k$. The lemma gives us conditions under which the function $\Psi_n(z)$ is correctly defined and also asymptotical properties of the coefficients $\psi_{j,n}$.

LEMMA 5.1. *Let* (A2) *with* $r \geqslant 1$ *hold and* $\sum_{j=1}^{\infty} j^r\|\phi_{j,n} - \phi_j\| = o(1)$, *where* $n \to \infty$. *Then*:

(i) *There exists* $n_0 \in \mathbb{N}$ *such that*

$$\sup_{n\geqslant n_0}\sum_{j=1}^{\infty} j^r \|\phi_{j,n}\| < \infty \qquad \textit{and} \qquad \inf_{n\geqslant n_0}\inf_{|z|\leqslant 1} \big|\det\big(\Phi_n(z)\big)\big| > 0.$$

(ii) *There exists* $n_1 \in \mathbb{N}$ *such that, for* $n \geqslant n_1$, $\Psi_n(z)$ *is absolutely convergent for* $|z| \leqslant 1$ *and*

$$\sup_{n\geqslant n_1}\sum_{j=0}^{\infty} j^r \|\psi_{j,n}\| < \infty.$$

P r o o f. (i) is an immediate consequence of the assumptions of the lemma and Corollary 5.1.

(ii) By the formula for the inversion of the matrix we have

$$\Psi_n(z) = \Phi_n^{-1}(z) = \frac{1}{\det\big(\Phi_n(z)\big)}\mathrm{adj}\big(\Phi_n(z)\big).$$

Denoting the element of the matrix by $\phi_{j,n} = [\phi_{(sv),j,n}]_{s,v=1}^{k}$, $j = 1, 2, \ldots$, we get

$$\Phi_n(z) = \Big[\sum_{j=1}^{\infty} \phi_{(sv),j,n} z^j\Big]_{s,v=1}^{k}.$$

From (i) we have $\det\big(\Phi_n(z)\big) \neq 0$ for $|z| \leqslant 1$, $n \geqslant n_0$ and

$$\sup_{n\geqslant n_0}\sum_{j=1}^{\infty} j^r |\phi_{(sv),j,n}| < \infty \quad \text{ for all } s, v = 1, \ldots, k.$$

For some $s, v, u, w = 1, \ldots, k$ we have

$$\sum_{j=1}^{\infty} \phi_{(sv),j,n} z^j \cdot \sum_{j=1}^{\infty} \phi_{(uw),j,n} z^j = \sum_{j=1}^{\infty} \big(\phi_{(sv)} * \phi_{(uw)}\big)_{j,n} z^j,$$

$$\big(\phi_{(sv)} * \phi_{(uw)}\big)_{j,n} = \sum_{l=1}^{\infty} \phi_{(sv),l,n}\phi_{(uw),j-l,n}.$$

Using properties of the convolution, we get

$$\sum_{j=1}^{\infty} j^r |\big(\phi_{(sv)} * \phi_{(uw)}\big)_{j,n}| \leqslant \sum_{j=1}^{\infty} j^r |\phi_{(sv),j,n}| \sum_{j=1}^{\infty} j^r |\phi_{(uw),j,n}|,$$

and further

$$\sup_{n\geqslant n_0}\sum_{j=1}^{\infty} j^r |\big(\phi_{(st)} * \phi_{(uw)}\big)_{j,n}| < \infty.$$

In consequence, we infer that the coefficients of the determinant $\det\big(\Phi_n(z)\big)$ and the coefficients of all elements of the matrix $\mathrm{adj}\big(\Phi_n(z)\big)$ fulfill

$$\det\big(\Phi_n(z)\big) = \sum_{j=1}^{\infty} \check{\phi}_{j,n} z^j, \qquad \sup_{n\geqslant n_0} \sum_{j=1}^{\infty} j^r |\check{\phi}_{j,n}| < \infty,$$

$$\mathrm{adj}\big(\Phi_n(z)\big) = [\phi^0_{(sv),n}(z)]_{s,v=1}^k, \qquad \phi^0_{(sv),n}(z) = \sum_{j=1}^{\infty} \phi^0_{(sv),j,n} z^j,$$

$$\sup_{n\geqslant n_0} \sum_{j=1}^{\infty} j^r |\phi^0_{(sv),j,n}| < \infty \quad \text{for all } s, v = 1, \ldots, k.$$

To complete the proof, we have to show that $1/\det\big(\Phi_n(z)\big)$ is an analytical function for $|z| \leqslant 1$ with coefficients $\underline{\phi}_{j,n}$ satisfying

$$\frac{1}{\det\big(\Phi_n(z)\big)} = \sum_{j=1}^{\infty} \underline{\phi}_{j,n} z^j, \qquad \sup_{n\geqslant n_0} \sum_{j=1}^{\infty} j^r |\underline{\phi}_{j,n}| < \infty.$$

However, this is a consequence of the Wiener theorem (see [30]) and the lemma given by Bühlmann [8] for the one-dimensional case. Finally, we get

$$\begin{aligned}\Psi_n(z) &= \frac{1}{\det\big(\Phi_n(z)\big)} \mathrm{adj}\big(\Phi_n(z)\big) = \sum_{j=1}^{\infty} \underline{\phi}_{j,n} z^j [\phi^0_{(sv),n}(z)]_{s,v=1}^k \\ &= \sum_{j=1}^{\infty} \underline{\phi}_{j,n} z^j \Big[\sum_{j=1}^{\infty} \phi^0_{(sv),j,n} z^j \Big]_{s,v=1}^k = \sum_{j=0}^{\infty} \psi_{j,n} z^j\end{aligned}$$

and

$$\sup_{n\geqslant n_1} \sum_{j=0}^{\infty} j^r \, \|\psi_{j,n}\| < \infty. \quad \blacksquare$$

The next two theorems give us properties of the coefficients $\widehat{\psi}_{j,n}$ from the representation (3.3), in which $\widehat{\Phi}_n(z)$ replaces the function $\Phi_n(z)$ in Lemma 5.1. These two theorems are multivariate versions of the theorems given by Bühlmann for the one-dimensional case ([8], Theorems 3.1 and 3.2).

THEOREM 5.2. *Let* (LA) *hold. Then there exists a random variable* n_1 *for which*

$$\sup_{n\geqslant n_1} \sum_{j=0}^{\infty} j^r \|\widehat{\psi}_{j,n}\| < \infty \text{ almost surely}.$$

P r o o f. We will use Lemma 5.1. Thus, we have to check its assumption:

$$\begin{aligned}\sum_{j=1}^{\infty} j^r \|\widehat{\phi}_{j,n} - \phi_j\| &\leqslant \sum_{j=1}^{p} j^r \|\widehat{\phi}_{j,n} - \phi_{j,n}\| + \sum_{j=1}^{p} j^r \|\phi_{j,n} - \phi_j\| + \sum_{j=p+1}^{\infty} j^r \, \|\phi_j\| \\ &= S_1 + S_2 + S_3.\end{aligned}$$

By Theorem 2.1 from [16] we obtain $S_1 = o(1)$ almost surely. Under the assumption (A2), using the Baxter inequality ([15], Theorem 6.6.12 and p. 271) and Corollary 5.1, we get $S_2 = o(1)$. Similarly we show that $S_3 = o(1)$. Finally, we get

$$\sum_{j=1}^{\infty} j^r \|\widehat{\phi}_{j,n} - \phi_j\| = o(1) \text{ almost surely.}$$

We complete the proof using Lemma 5.1. ∎

THEOREM 5.3. *Let* (LA) *hold. Then*

$$\sup_{1 \leqslant j < \infty} \|\widehat{\psi}_{j,n} - \psi_j\| = O\big((\log n/n)^{1/2}\big) + O(p^{-r}) \textit{ almost surely.}$$

P r o o f. We start from showing that

$$\Sigma - \widehat{\Sigma} = O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}). \tag{5.1}$$

To prove it, we use Theorem 2.1 from [16] and the Hannan and Deistler bound ([15], Theorem 7.4.3). We get

$$\max_{0 \leqslant j < \infty} \|\widehat{\gamma}(j) - \gamma(j)\| = O_{a.s.}\big((\log n/n)^{1/2}\big), \tag{5.2}$$

where $\widehat{\gamma}(j) = 0$ for $|j| \geqslant n$. From the Yule–Walker equations for the models (2.2) and (3.3) we have

$$\Sigma = \gamma(0) - \sum_{j=1}^{\infty} \phi_j \gamma^T(j), \quad \widehat{\Sigma} = \widehat{\gamma}(0) - \sum_{j=1}^{p} \widehat{\phi}_{j,n} \widehat{\gamma}^T(j).$$

Therefore,

$$\begin{aligned}
\|\Sigma - \widehat{\Sigma}\| &\leqslant \|\gamma(0) - \widehat{\gamma}(0)\| + \Big\| \sum_{j=1}^{p} \big(\widehat{\phi}_{j,n} \widehat{\gamma}^T(j) - \phi_j \gamma^T(j)\big) \Big\| + \Big\| \sum_{j=p+1}^{\infty} \phi_j \gamma^T(j) \Big\| \\
&\leqslant \|\gamma(0) - \widehat{\gamma}(0)\| \\
&\quad + \max_{1 \leqslant j \leqslant p} \|\widehat{\phi}_{j,n} - \phi_j\| \big(p \max_{1 \leqslant j \leqslant p} \|\widehat{\gamma}(j) - \gamma(j)\| + \sum_{j=1}^{p} \|\gamma(j)\| \big) \\
&\quad + \max_{1 \leqslant j \leqslant p} \|\widehat{\gamma}(j) - \gamma(j)\| \sum_{j=1}^{\infty} \|\phi_j\| + \|[EX_{t,i}^2 EX_{t,j}^2]_{i,j=1}^k\| \sum_{j=p+1}^{\infty} \|\phi_j\| \\
&= O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r})
\end{aligned}$$

and we get (5.1).

Following the ideas of the proof of Theorem 3.2 from [8] for the one-dimensional case, let $\{Y_t\}_{t\in\mathbb{Z}}$ be a conditional process (given $X_1,\dots,X_n$) in the form

$$Y_t = \sum_{j=1}^{p} \widehat{\phi}_{j,n} Y_{t-j} + \eta_t, \tag{5.3}$$

where $\{\eta_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d. random vectors and $E\eta_t = 0$, $E\eta_t\eta_t^T = \Sigma$. Additionally, we assume that η_t and ϵ_t are independent. According to properties of the Yule–Walker estimators, $\{Y_t\}_{t\in\mathbb{Z}}$ is a causal process

$$Y_t = \sum_{j=0}^{\infty} \widehat{\psi}_{j,n} \eta_{t-j}. \tag{5.4}$$

Denote by E^Y and Cov^Y the conditional expectation and the autocovariance function of the process $\{Y_t\}_{t\in\mathbb{Z}}$ under the condition $X_1,\dots,X_n$. Thus, using (5.3) and (5.4) for Y_{t+i}, we have

$$E^Y Y_{t+i}\eta_t^T \;=\; \widehat{\psi}_{i,n}\Sigma = \gamma_Y(i) - \sum_{j=1}^{p} \gamma_Y(i+j)\widehat{\phi}_{j,n}^T,$$

where $\gamma_Y(j) = \mathrm{Cov}^Y(Y_{t+j}, Y_t) = \sum_{l=0}^{\infty} \widehat{\psi}_{l+j,n}\Sigma\widehat{\psi}_{l,n}^T$. We get a similar equality for ψ_i:

$$\psi_i\Sigma = \gamma(i) - \sum_{j=1}^{\infty} \gamma(i+j)\phi_j^T.$$

Thus, for $i = 0, 1, 2, \dots$ we have

$$\begin{aligned}
\|\widehat{\psi}_{i,n} - \psi_i\| &\leqslant \|\Sigma^{-1}\| \Big\| \gamma_Y(i) - \sum_{j=1}^{p} \gamma_Y(i+j)\widehat{\phi}_{j,n}^T - \gamma(i) + \sum_{j=1}^{\infty} \gamma(i+j)\phi_j^T \Big\| \\
&\leqslant \|\Sigma^{-1}\| \Big\| \sum_{j=1}^{p} \gamma_Y(i+j)(\widehat{\phi}_{j,n}^T - \phi_j^T) - \sum_{j=p+1}^{\infty} \gamma_Y(i+j)\phi_j^T \Big\| \\
&\quad + \|\Sigma^{-1}\| \Big\| \sum_{j=1}^{\infty} \big(\gamma_Y(i+j) - \gamma(i+j)\big)\phi_i^T \Big\| + \|\Sigma^{-1}\| \, \|\gamma_Y(i) - \gamma(i)\| \\
&= \|\Sigma^{-1}\|(S_{1,i} + S_{2,i} + S_{3,i}).
\end{aligned}$$

We bound the components separately. We get

$$\begin{aligned}
S_{1,i} &\leqslant \sum_{j=1}^{p} \|\gamma_Y(i+j)\| \, \|\widehat{\phi}_{j,n} - \phi_j\| + \sum_{j=p+1}^{\infty} \|\gamma_Y(i+j)\| \, \|\phi_j\| \\
&\leqslant \|\Sigma\| \max_{1\leqslant j\leqslant p} \|\widehat{\phi}_{j,n} - \phi_j\| \Big(\sum_{l=0}^{\infty} \|\widehat{\psi}_{l,n}\| \Big)^2 + \|\Sigma\| \Big(\sum_{l=0}^{\infty} \|\widehat{\psi}_{l,n}\| \Big)^2 \sum_{j=p+1}^{\infty} \|\phi_j\| \\
&= O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}) \quad \text{for each } i \in \mathbb{N},
\end{aligned}$$

where the last approximation is obtained by using Theorem 2.1 from [16] and Theorem 5.2. Further, using the convention $\sum_{j=l}^{i} a_j = 0$ for $i < l$, we have

$$\begin{aligned}
S_{2,i} &\leqslant \Big\| \sum_{j=1}^{p-i} \big(\gamma_Y(i+j) - \gamma(i+j)\big)\phi_j^T \Big\| + \Big\| \sum_{j=p-i+1}^{\infty} \big(\gamma_Y(i+j) - \gamma(i+j)\big)\phi_j^T \Big\| \\
&\leqslant \max_{1\leqslant j\leqslant p} \|\widehat{\gamma}(j) - \gamma(j)\| \sum_{j=0}^{\infty} \|\phi_j\| + \|\Sigma - \widehat{\Sigma}\| \Big(\sum_{j=0}^{\infty} \|\widehat{\psi}_{j,n}\| \Big)^2 \sum_{j=0}^{\infty} \|\phi_j\| \\
&\quad + \|\Sigma\| \sum_{j=0}^{\infty} \|\phi_j\| \Big(\sum_{j=0}^{\infty} \|\widehat{\psi}_{j,n}\| \sum_{j=p+1}^{\infty} \|\widehat{\psi}_{j,n}\| + \sum_{j=0}^{\infty} \|\psi_j\| \sum_{j=p+1}^{\infty} \|\psi_j\| \Big) \\
&= O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}) \quad \text{for each } i \in \mathbb{N},
\end{aligned}$$

where the last approximation is due to (5.1), (5.2), Theorem 5.2 and the assumption (A2). The component $S_{3,i}$ can be bounded analogously to the component $S_{2,i}$ and we get $S_{3,i} = O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r})$ for each $i \in \mathbb{N}$. Finally, we have

$$\sup_{0\leqslant i} \|\widehat{\psi}_{i,n} - \psi_i\| = O_{a.s.}\big((\log n/n)^{1/2}\big) + O(p^{-r}). \quad \blacksquare$$

5.2. Consistency of sieve bootstrap. From the construction of the bootstrap replications algorithm we have $E^*\epsilon_t^* = 0$. The following lemma gives us the convergence of the bootstrap covariance matrix $\Sigma^* = E^*\epsilon_t^*\epsilon_t^{*T}$ to the covariance matrix Σ of the white noise $\{\epsilon_t\}_{t\in\mathbb{Z}}$. In the one-dimensional case, this fact was proved in [9] (see Lemma 5.3).

LEMMA 5.2. *Let* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

$$E^*\epsilon_t^*\epsilon_t^{*T} = E\epsilon_t\epsilon_t^T + o_P(1).$$

P r o o f. Although the proof of this lemma is similar to the proof of Lemma 5.3 in [9] for the one-dimensional case, we will give it for completeness. Notice the following equality holds:

$$\widehat{\epsilon}_{t,n} = X_t - \sum_{j=1}^{p} \widehat{\phi}_{j,n} X_{t-j} = \epsilon_t + Q_{t,n} + R_{t,n}, \tag{5.5}$$

where

$$Q_{t,n} = \sum_{j=1}^{p} (\phi_{j,n} - \widehat{\phi}_{j,n}) X_{t-j},$$

$$R_{t,n} = \sum_{j=1}^{\infty} (\phi_j - \phi_{j,n})\, X_{t-j}, \qquad \phi_{j,n} = 0 \text{ for } j > p.$$

The bootstrap covariance matrix has the form

$$E^*\epsilon_t^*\epsilon_t^{*T} = \frac{1}{n-p}\sum_{t=p+1}^{n}(\widehat{\epsilon}_{t,n}-\overline{\epsilon}_n)\,(\widehat{\epsilon}_{t,n}-\overline{\epsilon}_n)^T,$$

where $\overline{\epsilon}_n = \frac{1}{n-p}\sum_{t=p+1}^{n}\widehat{\epsilon}_{t,n}$. First, we show that $\overline{\epsilon}_n = o_P(1)$. We have

$$\overline{\epsilon}_n = \frac{1}{n-p}\sum_{t=p+1}^{n}\widehat{\epsilon}_{t,n} = \frac{1}{n-p}\sum_{t=p+1}^{n}(\epsilon_t + Q_{t,n} + R_{t,n}) = S_1 + S_2 + S_3.$$

From the assumption (A1) and the Markov inequality we get $S_1 = O_P(n^{-1/2})$. Using the Cauchy–Schwarz inequality and the Jensen inequality, we obtain

$$\begin{aligned}\|S_2\| &= \left\|\frac{1}{n-p}\sum_{t=p+1}^{n}\sum_{j=1}^{p}(\phi_{j,n}-\widehat{\phi}_{j,n})X_{t-j}\right\| \\ &\leqslant \Big(\sum_{j=1}^{p}\|\phi_{j,n}-\widehat{\phi}_{j,n}\|^2\Big)^{1/2}\left(\frac{1}{n-p}\sum_{t=p+1}^{n}\sum_{j=1}^{p}\|X_{t-j}\|^2\right)^{1/2}.\end{aligned}$$

By Theorem 2.1 from [16] we have

$$\sum_{j=1}^{p}\|\phi_{j,n}-\widehat{\phi}_{j,n}\|^2 \leqslant p\max_{1\leqslant j\leqslant p}\|\phi_{j,n}-\widehat{\phi}_{j,n}\|^2 = o_{a.s.}\big((\log n/n)^{1/2}\big).$$

Under the assumption (A1) and using the Markov inequality, we have

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\sum_{j=1}^{p}\|X_{t-j}\|^2 = O_P\big(p(n)\big),$$

and further

$$S_2 = o_{a.s.}\big((\log n/n)^{1/4}\big)O_P\big(p(n)^{1/2}\big) = o_P(1).$$

To bound S_3, we use the Baxter inequality (see [15], Theorem 6.6.12 and p. 271) and get

$$E\,\|S_3\| \leqslant \sum_{j=1}^{\infty}\|\phi_j-\phi_{j,n}\|\,E\,\|X_{t-j}\| \leqslant \text{const}\cdot E\,\|X_t\|\sum_{j=p+1}^{\infty}\|\phi_j\|\,.$$

Using Corollary 5.1 and the Markov inequality, we see that $S_3 = o_P(1)$. Thus, we have $\overline{\epsilon}_n = o_P(1)$. By formula (5.5) for $\widehat{\epsilon}_{t,n}$ we can write

$$\begin{aligned}\frac{1}{n-p}\sum_{t=p+1}^{n}\widehat{\epsilon}_{t,n}\widehat{\epsilon}_{t,n}^T = \frac{1}{n-p}\sum_{t=p+1}^{n}&(\epsilon_t\epsilon_t^T + Q_{t,n}Q_{t,n}^T + R_{t,n}R_{t,n}^T + \epsilon_t Q_{t,n}^T \\ &+ \epsilon_t R_{t,n}^T + Q_{t,n}\epsilon_t^T + Q_{t,n}R_{t,n}^T + R_{t,n}\epsilon_t^T + R_{t,n}Q_{t,n}^T).\end{aligned}$$

Under the assumption (A1) we obtain $\frac{1}{n-p}\sum_{t=p+1}^{n}\epsilon_t\epsilon_t^T = E\epsilon_t\epsilon_t^T + O_P(n^{-1/2})$. Using the Cauchy–Schwarz inequality, we have

$$\left\|\frac{1}{n-p}\sum_{t=p+1}^{n}Q_{t,n}Q_{t,n}^T\right\| \leqslant \sum_{j=1}^{p}\|\phi_{j,n}-\widehat{\phi}_{j,n}\|^2\frac{1}{n-p}\sum_{t=p+1}^{n}\sum_{j=1}^{p}\|X_{t-j}\|^2 = o_P(1).$$

From the Baxter inequality we get

$$\begin{aligned}E\left\|\frac{1}{n-p}\sum_{t=p+1}^{n}R_{t,n}R_{t,n}^T\right\| &\leqslant \sum_{i,j=1}^{\infty}\|\phi_j-\phi_{j,n}\|\,\|\phi_i-\phi_{i,n}\|\,E\,\|X_{t-j}\|\,\|X_{t-i}\| \\ &\leqslant \text{const}\cdot E\,\|X_t\|^2\Big(\sum_{j=p+1}^{\infty}\|\phi_j\|\Big)^2 = O(p^{-2r}).\end{aligned}$$

Thus $\frac{1}{n-p}\sum_{t=p+1}^{n}R_{t,n}R_{t,n}^T = o_P(1)$. The remaining components can be bounded analogously, by using the Cauchy–Schwarz inequality. Finally, we obtain

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\widehat{\epsilon}_{t,n}\widehat{\epsilon}_{t,n}^T = E\epsilon_t\epsilon_t^T + o_P(1),$$

which completes the proof of the lemma. ∎

By similar arguments to those given above, we obtain

COROLLARY 5.2. *Let us assume that* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

$$E^*(\epsilon_t^*\epsilon_t^{*T})^2 = E(\epsilon_t\epsilon_t^T)^2 + o_P(1).$$

In the next lemma we prove the consistency of the bootstrap replications ϵ_t^*. It is a multivariate version of Lemma 5.5 given by Bühlmann [9] for the one-dimensional case.

LEMMA 5.3. *Let* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

$$\epsilon_t^* \xrightarrow{D^*} \epsilon_t \text{ in probability.}$$

Proof. Write $F_{\epsilon,n}(x) = \frac{1}{n-p}\sum_{t=p+1}^{n}\mathbf{1}\{\epsilon_t \leqslant x\}$ and $F_\epsilon(x) = P(\epsilon_t \leqslant x)$. Let $\widehat{F}_{\epsilon,n}$ be given by (3.1) and $d_2(\cdot,\cdot)$ be the Mallows metric. Consequently, we have (see [6])

$$d_2(F_{\epsilon,n}, F_\epsilon) = o(1) \text{ almost surely.}$$

To complete the proof, we have to show that

$$d_2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) = o_P(1).$$

By the definition of the Mallows metric and (5.5) we obtain

$$d_2^2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) \leqslant \frac{1}{n-p}\sum_{t=p+1}^{n} \|\widetilde{\epsilon}_{t,n} - \epsilon_t\|^2 = \frac{1}{n-p}\sum_{t=p+1}^{n} \|\widehat{\epsilon}_{t,n} - \epsilon_t - \overline{\epsilon}_n\|^2$$
$$= \frac{1}{n-p}\sum_{t=p+1}^{n} \|Q_{t,n} + R_{t,n} - \overline{\epsilon}_n\|^2 .$$

Using bounds from the proof of Lemma 5.2, we have $d_2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) = o_P(1)$, and as a consequence

$$d_2(\widehat{F}_{\epsilon,n}, F_\epsilon) \leqslant d_2\left(F_{\epsilon,n}, F_\epsilon\right) + d_2(\widehat{F}_{\epsilon,n}, F_{\epsilon,n}) = o_P(1),$$

which completes the proof of the lemma. ∎

The following lemma gives us asymptotic bounds for some sample bootstrap estimators.

LEMMA 5.4. *Let* (A1), (A2) *with* $r \geqslant 1$ *and* (B) *with* $p(n) = o\big((n/\log n)^{1/2}\big)$ *hold. Then*

$$\frac{1}{n}\sum_{t=1}^{n} \epsilon_t^* = O_{P^*}(n^{-1/2}) \text{ in probability}, \tag{5.6}$$

$$\frac{1}{n}\sum_{t=1}^{n} \epsilon_t^* \epsilon_{t+s}^{*T} = O_{P^*}(n^{-1/2}) \text{ in probability}, \tag{5.7}$$

$$\frac{1}{n}\sum_{t=1}^{n} \epsilon_t^* \epsilon_t^{*T} = \Sigma^* + O_{P^*}(n^{-1/2}) \text{ in probability}, \tag{5.8}$$

where $\Sigma^* = E^*\epsilon_t^*\epsilon_t^{*T}$ *and* $s \neq 0$.

Proof. Let $\epsilon, \eta > 0$. According to Lemma 5.2 we have

$$E^* \|\epsilon_t^*\|^2 = E^* \sum_{i=1}^{k} \epsilon_{t,i}^{*2} \xrightarrow{P} E \sum_{i=1}^{k} \epsilon_{t,i}^2 = E\|\epsilon_t\|^2.$$

So, there exists some constant M_η such that

$$P(E^* \|\epsilon_t^*\|^2 > M_\eta) < \eta.$$

Let $\delta_{\epsilon,\eta} = \sqrt{M_\eta/\epsilon}$. Thus, (5.6) is a consequence of the following inequality:

$$P\left(P^*\left(n^{1/2}\left\|\frac{1}{n}\sum_{t=1}^{n} \epsilon_t^*\right\| > \delta_{\epsilon,\eta}\right) > \epsilon\right) \leqslant P\left(\frac{E^*\|\epsilon_t^*\|^2}{\delta_{\epsilon,\eta}^2} > \epsilon\right) < \eta.$$

The bounds (5.7) and (5.8) can be shown analogously, by using Corollary 5.2. ∎

In the next lemma we prove the convergence of the bootstrap replications $\widehat{\phi}_j^*$ of the Yule–Walker estimators. It is a multivariate version of the result given by Alonso et al. [1].

LEMMA 5.5. *Let* (LA) *hold. Then*

$$\max_{1\leqslant j\leqslant p}\|\widehat{\phi}_j^*-\phi_{j,n}\|\xrightarrow{P^*}0 \text{ in probability.}$$

P r o o f. We have the following bounds:

$$\begin{aligned}\|\widehat{\Phi}_p^*-\Phi_{p,n}\|_{kp\times p,\infty}&=\|\widehat{\Gamma}_p^{*-1}\widehat{\gamma}_p^*-\Gamma_p^{-1}\gamma_p\|_{kp\times p,\infty}\\&=\|(\widehat{\Gamma}_p^{*-1}-\Gamma_p^{-1})\widehat{\gamma}_p^*-\Gamma_p^{-1}(\gamma_p-\widehat{\gamma}_p^*)\|_{kp\times p,\infty}\\&\leqslant\|\widehat{\Gamma}_p^{*-1}-\Gamma_p^{-1}\|_{kp\times kp,\infty}\|\widehat{\gamma}_p^*\|_{kp\times p,\infty}\\&\quad+\|\Gamma_p^{-1}\|_{kp\times kp,\infty}\|\gamma_p-\widehat{\gamma}_p^*\|_{kp\times p,\infty},\end{aligned}$$

where $\widehat{\Phi}_p^*=(\widehat{\phi}_1^*,\dots,\widehat{\phi}_p^*)^T$, $\Phi_{p,n}=(\phi_{1,n},\dots,\phi_{p,n})^T$, $\widehat{\Gamma}_p^*=[\widehat{\gamma}^*(i-j)]_{i,j=1}^p$, $\widehat{\gamma}_p^*=\big(\widehat{\gamma}^*(1),\dots,\widehat{\gamma}^*(p)\big)^T$, $\Gamma_p=[\gamma(i-j)]_{i,j=1}^p$, and $\gamma_p=\big(\gamma(1),\dots,\gamma(p)\big)^T$. Hannan and Deistler ([15], Theorem 6.6.11) proved that the norm of the matrix Γ_p^{-1} is uniformly bounded with respect to p, i.e. $\sup_{0<p<\infty}\|\Gamma_p^{-1}\|_{kp\times kp,\infty}\leqslant C<\infty$. Thus, we get

$$\begin{aligned}\|\widehat{\Gamma}_p^{*-1}-\Gamma_p^{-1}\|_{kp\times kp,\infty}&=\|\widehat{\Gamma}_p^{*-1}(\widehat{\Gamma}_p^*-\Gamma_p)\Gamma_p^{-1}\|_{kp\times kp,\infty}\\&\leqslant\|\widehat{\Gamma}_p^{*-1}\|_{kp\times kp,\infty}\|\widehat{\Gamma}_p^*-\Gamma_p\|_{kp\times kp,\infty}\|\Gamma_p^{-1}\|_{kp\times kp,\infty}\\&\leqslant(C+\|\widehat{\Gamma}_p^{*-1}-\Gamma_p^{-1}\|_{kp\times kp,\infty})C\|\widehat{\Gamma}_p^*-\Gamma_p\|_{kp\times kp,\infty}.\end{aligned}$$

Further, we have

$$\frac{\|\widehat{\Gamma}_p^{*-1}-\Gamma_p^{-1}\|_{kp\times kp,\infty}}{(C+\|\widehat{\Gamma}_p^{*-1}-\Gamma_p^{-1}\|_{kp\times kp,\infty})C}\leqslant\|\widehat{\Gamma}_p^*-\Gamma_p\|_{kp\times kp,\infty}.$$

So, we have to show that $\|\widehat{\Gamma}_p^*-\Gamma_p\|_{kp\times kp,\infty}=o_{P^*}(1)$ in probability.

Notice that $\|\widehat{\Gamma}_p^*-\Gamma_p\|_{kp\times kp,\infty}\leqslant p\max_{0\leqslant i\leqslant p}\|\widehat{\gamma}^*(i)-\gamma(i)\|$. We have

$$p\|\widehat{\gamma}^*(i)-\gamma(i)\|\leqslant p\|\widehat{\gamma}^*(i)-\gamma^*(i)\|+p\|\gamma^*(i)-\gamma(i)\|=S_1+S_2.$$

The component S_1 contains a sample bootstrap autocovariance function, thus we are able to bound it only for $|i|\leqslant p$. We have

$$\begin{aligned}S_1&=p\bigg\|\frac{1}{n}\sum_{t=1}^{n-i}X_{t+i}^*X_t^{*T}-\gamma^*(i)\bigg\|\\&\leqslant p\bigg\|\sum_{j,l=0}^{\infty}\widehat{\psi}_{j,n}\frac{1}{n}\sum_{t=1}^{n-i}\epsilon_{t+i-j}^*\epsilon_{t-l}^{*T}\widehat{\psi}_{l,n}^T-\sum_{j=0}^{\infty}\widehat{\psi}_{j+i,n}\Sigma^*\widehat{\psi}_{j,n}^T\bigg\|\\&=o_{P^*}(1)\text{ in probability,}\end{aligned}$$

where the obtained bounds are a consequence of Theorem 5.2 and Lemma 5.4. Notice that with $p(n) = o\big((\log n/n)^{-1/(2r+2)}\big)$ in assumption (B) of Lemma 5.2 we get

$$p\,(\Sigma - \Sigma^*) = o_P(1). \tag{5.9}$$

We bound the component S_2 for each $i \in \mathbb{Z}$. We get

$$\begin{aligned} S_2 &= p\Big\| \sum_{j=0}^{\infty} \widehat{\psi}_{j+i,n}\Sigma^*\widehat{\psi}_{j,n}^T - \sum_{j=0}^{\infty} \psi_{j+i}\Sigma\psi_j^T \Big\| \\ &\leqslant \sum_{j=0}^{\infty} p\|\widehat{\psi}_{j+i,n} - \psi_{j+i}\|\|\Sigma^*\|\|\widehat{\psi}_{j,n}\| + \sum_{j=0}^{\infty} \|\psi_{j+i}\|p\|\Sigma^* - \Sigma\|\|\widehat{\psi}_{j,n}\| \\ &\quad + \sum_{j=0}^{\infty} \|\psi_{j+i}\|\|\Sigma\|p\|\widehat{\psi}_{j,n} - \psi_j\| \\ &= o_P(1), \end{aligned}$$

where the bounds are obtained by Theorems 5.2 and 5.3 and formula (5.9). Finally, we have

$$p \max_{0\leqslant i\leqslant p} \|\widehat{\gamma}^*(i) - \gamma(i)\| = o_{P^*}(1) \text{ in probability}, \tag{5.10}$$

which completes the proof of the lemma. ∎

Moreover, from Lemma 5.5, Theorem 2.1 in [16] and the Baxter inequality ([15], Theorem 6.6.12 and p. 271) we obtain

COROLLARY 5.3. *Let* (LA) *hold. Then*

$$\max_{1\leqslant j\leqslant p} \|\widehat{\phi}_j^* - \phi_j\| \xrightarrow{P^*} 0 \text{ in probability.}$$

In the next lemmas we present multivariate generalizations of the results given by Różański and Zagdański for the one-dimensional case (see [24]) and concerning the prediction error $X_{n+h} - \widehat{X}_{n+h}$. Since the proofs of these lemmas are similar to the proofs for the one-dimensional case, we omit them.

LEMMA 5.6. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$X_{n+h} - \widehat{X}_{n+h} = O_P(1).$$

LEMMA 5.7. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$\begin{aligned} X_{n+h} - \widehat{X}_{n+h} = D_{1,h}(\Phi_{h-1})\epsilon_{n+1} + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{n+h-1} \\ + \epsilon_{n+h} + o_P(1), \end{aligned}$$

where $D_{1,h}(\cdot), \ldots, D_{h-1,h}(\cdot)$ *are some continuous functions and*

$$\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T.$$

In the following, we prove analogous results for the bootstrap prediction error $X^*_{n+h} - \widehat{X}_{n+h}$ approximating the prediction error $X_{n+h} - \widehat{X}_{n+h}$.

LEMMA 5.8. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$X^*_{n+h} - \widehat{X}_{n+h} = O_{P^*}(1) \text{ in probability.}$$

P r o o f. We give the proof by induction on h. For $h = 1$ we have

$$\begin{aligned} X^*_{n+1} - \widehat{X}_{n+1} &= \sum_{j=1}^{p} \widehat{\phi}^*_j X_{n+1-j} + \epsilon^*_{n+1} - \sum_{j=1}^{p} \widehat{\phi}_j X_{n+1-j} \\ &= \sum_{j=1}^{p} X_{n+1-j}(\widehat{\phi}^*_j - \widehat{\phi}_j) + \epsilon^*_{n+1} \\ &= \epsilon^*_{n+1} + o_{P^*}(1) = O_{P^*}(1) \text{ in probability.} \end{aligned}$$

Let us assume that for all l such that $1 \leqslant l \leqslant h-1$ we have $X^*_{n+l} - \widehat{X}_{n+l} = O_{P^*}(1)$ in probability and we prove it for h. We have

$$\begin{aligned} X^*_{n+h} - \widehat{X}_{n+h} &= \Big(\sum_{j=1}^{h-1} \widehat{\phi}^*_j X^*_{n+h-j} + \sum_{j=h}^{p} \widehat{\phi}^*_j X_{n+h-j} + \epsilon^*_{n+h} \Big) \\ &\quad - \Big(\sum_{j=1}^{h-1} \widehat{\phi}_j \widehat{X}_{n+h-j} + \sum_{j=h}^{p} \widehat{\phi}_j X_{n+h-j} \Big) \\ &= \sum_{j=1}^{h-1} \widehat{\phi}^*_j (X^*_{n+h-j} - \widehat{X}_{n+h-j}) + \sum_{j=1}^{h-1} \widehat{X}_{n+h-j}(\widehat{\phi}^*_j - \widehat{\phi}_j) \\ &\quad + \sum_{j=h}^{p} X_{n+h-j}(\widehat{\phi}^*_j - \widehat{\phi}_j) + \epsilon^*_{n+h} \\ &= O_{P^*}(1) \text{ in probability,} \end{aligned}$$

where the last bounds are a consequence of Lemma 5.5, Corollary 5.3, the Baxter inequality ([15], Theorem 6.6.12 and p. 271) and the induction assumption. ∎

LEMMA 5.9. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$

$$\begin{aligned} X^*_{n+h} - \widehat{X}_{n+h} = D_{1,h}(\Phi_{h-1})\epsilon^*_{n+1} + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon^*_{n+h-1} \\ + \epsilon^*_{n+h} + \ o_{P^*}(1) \text{ in probability,} \end{aligned}$$

where $D_{1,h}(\cdot), \ldots, D_{h-1,h}(\cdot)$ *are the same continuous functions as in Lemma* 5.7 *and* $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

P r o o f. We give the proof by induction on h. For $h = 1$ we have

$$\begin{aligned} X^*_{n+1} - \widehat{X}_{n+1} &= \sum_{j=1}^{p} \widehat{\phi}^*_j X_{n+1-j} + \epsilon^*_{n+1} - \sum_{j=1}^{p} \widehat{\phi}_j X_{n+1-j} \\ &= \sum_{j=1}^{p} X_{n+1-j}(\widehat{\phi}^*_j - \widehat{\phi}_j) + \epsilon^*_{n+1} \\ &= \epsilon^*_{n+1} + o_{P^*}(1) \text{ in probability.} \end{aligned}$$

Let us assume that for all l such that $1 \leqslant l \leqslant h-1$ we have

$$\begin{aligned} X^*_{n+l} - \widehat{X}_{n+l} = \widetilde{D}_{1,l}(\Phi_{l-1})\epsilon^*_{n+1} + \ldots + \widetilde{D}_{l-1,l}(\Phi_{l-1})\epsilon^*_{n+l-1} \\ + \epsilon^*_{n+l} + o_{P^*}(1) \text{ in probability.} \end{aligned}$$

We prove the equality for h. We have

$$\begin{aligned} X^*_{n+h} - \widehat{X}_{n+h} &= \Big(\sum_{j=1}^{h-1} \widehat{\phi}^*_j X^*_{n+h-j} + \sum_{j=h}^{p} \widehat{\phi}^*_j X_{n+h-j} + \epsilon^*_{n+h} \Big) \\ &\quad - \Big(\sum_{j=1}^{h-1} \widehat{\phi}_j \widehat{X}_{n+h-j} + \sum_{j=h}^{p} \widehat{\phi}_j X_{n+h-j} \Big) \\ &= \sum_{j=1}^{h-1} \widehat{\phi}^*_j (X^*_{n+h-j} - \widehat{X}_{n+h-j}) + \sum_{j=1}^{h-1} \widehat{X}_{n+h-j}(\widehat{\phi}^*_j - \widehat{\phi}_j) \\ &\quad + \sum_{j=h}^{p} X_{n+h-j}(\widehat{\phi}^*_j - \widehat{\phi}_j) + \epsilon^*_{n+h} \\ &= \sum_{j=1}^{h-1} \phi_j (X^*_{n+h-j} - \widehat{X}_{n+h-j}) + o_{P^*}(1) + \epsilon^*_{n+h} \text{ in probability,} \end{aligned}$$

where the last bounds are obtained by Corollary 5.3, Lemma 5.8 and the Baxter inequality ([15], Theorem 6.6.12 and p. 271). We complete the proof of the lemma using the induction assumption. ∎

In the same way as in the one-dimensional case (see [24]) we can prove the following useful lemma.

LEMMA 5.10. *Assume that*

$$|P^*(X^*_n \leqslant u) - P(X \leqslant u)| \xrightarrow{P} 0$$

for some continuity point u *of the cumulative distribution function* F_X, *where* $\leqslant$ *means the relation of product order* (*partial order*) *in* $\mathbb{R}^k$,

$$\begin{aligned} Y^*_n &\xrightarrow{P^*} 0 \textit{ in probability}, \\ V_n &\xrightarrow{P} 0. \end{aligned}$$

Then

$$|P^*\left(X_n^* + Y_n^* \leqslant u\right) - P\left(X + V_n \leqslant u\right)| \xrightarrow{P} 0.$$

The next theorem gives us consistency of the hybrid bootstrap. An analogous result for the one-dimensional case was given by Różański and Zagdański (see [24]).

THEOREM 5.4. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$ *we have*

$$|P^*(X^*_{n+h} - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \widehat{X}_{n+h} \leqslant u)| = o_P(1)$$

for each u *being a continuity point of the cumulative distribution function of the random vector* $D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h$, *where* $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

P r o o f. Let $h \in \mathbb{N}$. According to Lemmas 5.7 and 5.9 we can write

$$\begin{aligned}
&P^*(X^*_{n+h} - \widehat{X}^*_{n+h} \leqslant u) - P(X_{n+h} - \widehat{X}_{n+h} \leqslant u)\\
&= P^*\big(D_{1,h}(\Phi_{h-1})\epsilon^*_{n+1} + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon^*_{n+h-1} + \epsilon^*_{n+h} + o_{P^*}(1) \leqslant u\big)\\
&\quad - P\big(D_{1,h}(\Phi_{h-1})\epsilon_{n+1} + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{n+h-1} + \epsilon_{n+h} + o_P(1) \leqslant u\big)\\
&= P^*\big(D_{1,h}(\Phi_{h-1})\epsilon^*_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon^*_{h-1} + \epsilon^*_h + o_{P^*}(1) \leqslant u\big)\\
&\quad - P\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h + o_P(1) \leqslant u\big).
\end{aligned}$$

Using the independence of ϵ_t, conditional independence of ϵ^*_t and Lemmas 5.3 and 5.10 we complete the proof of the theorem. ∎

By Theorem 5.4 and Remark 4.1 we have

COROLLARY 5.4. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$ *we have*

$$|P^*(X^*_{n+h} - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \mathrm{Proj}_n X_{n+h} \leqslant u)| = o_P(1)$$

for each u *being a continuity point of the cumulative distribution function of the random vector* $D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h$, *where* $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$ *and* $\mathrm{Proj}_n X_{n+h}$ *is the best linear* h*-step predictor* (*in the mean square sense*) *of* X_{n+h}.

COROLLARY 5.5. *Let* (LA) *hold. Additionally*, *assume that the cumulative distribution function of the distribution* ϵ_t *is continuous. Then for* $h \in \mathbb{N}$ *we have*

$$\sup_{u \in \mathbb{R}^k} |P^*(X^*_{n+h} - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \widehat{X}_{n+h} \leqslant u)| = o_P(1)$$

and

$$\sup_{u \in \mathbb{R}^k} |P^*(X^*_{n+h} - \widehat{X}_{n+h} \leqslant u) - P(X_{n+h} - \mathrm{Proj}_n X_{n+h} \leqslant u)| = o_P(1).$$

To prove the consistency of bootstrap-t, we have to show the consistency of the estimator $\widehat{\Sigma}_{n,X}(h)$ (estimator of the mean squared prediction error (4.1)).

LEMMA 5.11. *Let* (LA) *hold. Then*

$$\widehat{\Sigma}_{n,X}(h) \xrightarrow{P} \Sigma_X(h), \tag{5.11}$$

$$\widehat{\Sigma}^{-1}_{n,X}(h) \xrightarrow{P} \Sigma^{-1}_X(h). \tag{5.12}$$

P r o o f. We consider an estimator of the mean squared prediction error of the form (4.2):

$$\Sigma_X(h) = \sum_{j=0}^{h-1} \psi_j \Sigma \psi_j^T, \quad \widehat{\Sigma}_{n,X}(h) = \sum_{j=0}^{h-1} \widehat{\psi}_{j,n} \widehat{\Sigma} \widehat{\psi}_{j,n}^T.$$

Thus

$$\begin{aligned}\Big\| \sum_{j=0}^{h-1} \widehat{\psi}_{j,n} \widehat{\Sigma} \widehat{\psi}_{j,n}^T - \sum_{j=0}^{h-1} \psi_j \Sigma \psi_j^T \Big\| &\leqslant \sum_{j=0}^{h-1} \|\widehat{\psi}_{j,n} - \psi_j\| \|\widehat{\Sigma}\| \|\widehat{\psi}_{j,n}\| \\ &\quad + \sum_{j=0}^{h-1} \|\psi_j\| \|\widehat{\Sigma} - \Sigma\| \|\widehat{\psi}_{j,n}\| + \sum_{j=0}^{h-1} \|\psi_j\| \|\Sigma\| \|\widehat{\psi}_{j,n} - \psi_j\| \\ &= o_P(1),\end{aligned}$$

where the bounds are obtained by using (5.1) and Theorems 5.2 and 5.3. From the consistency of $\widehat{\Sigma}_{n,X}(h)$ it follows that $P(\widehat{\Sigma}_{n,X}(h)$ is invertible$) \to 1$ as $n \to \infty$, which together with the continuity of $\widehat{\Sigma}^{-1}_X(h)$ as a function of elements of $\widehat{\Sigma}_X(h)$ imply (5.12) (see [25], Theorem 5.18, p. 188). ∎

In the next lemma we prove the convergence of the bootstrap replication $\widehat{\psi}_j^*$. It is a multivariate version of the result given by Zagdański [29].

LEMMA 5.12. *Let* (LA) *hold. Then*

$$\sup_{0 \leqslant j} \|\widehat{\psi}_j^* - \psi_j\| = o_{P^*}(1) \text{ in probability.}$$

P r o o f. We use a similar method to that given in the proof of Theorem 5.3. We can write analogous equations:

$$\psi_j \Sigma = \gamma(j) - \sum_{i=1}^{\infty} \gamma(j+i) \phi_i^T, \quad \widehat{\psi}_j^* \Sigma^* = \gamma^*(j) - \sum_{i=1}^{p} \gamma^*(j+i) \widehat{\phi}_i^{*T}.$$

For $j \in \mathbb{N}$ we have

$$\begin{aligned}\|\widehat{\psi}_j^* - \psi_j\| &\leqslant \|\Sigma^{-1} - \Sigma^{*-1}\|\Big\|\gamma(j) - \sum_{i=1}^{\infty}\gamma(j+i)\phi_i^T\Big\| \\ &\quad + \|\Sigma^{*-1}\|\|\gamma^*(j) - \gamma(j)\| \\ &\quad + \|\Sigma^{*-1}\|\Big\|\sum_{i=1}^{\infty}\gamma(j+i)\phi_i^T - \sum_{i=1}^{p}\gamma^*(j+i)\widehat{\phi}_i^{*T}\Big\| \\ &= S_1 + S_2 + S_3.\end{aligned}$$

We treat the components S_i individually. We have

$$\begin{aligned}\|\Sigma^{-1} - \Sigma^{*-1}\| &= \|\Sigma^{*-1}(\Sigma^* - \Sigma)\Sigma^{-1}\| \\ &\leqslant \|\Sigma^{*-1}\|\|\Sigma^* - \Sigma\|\|\Sigma^{-1}\| \\ &\leqslant (\|\Sigma^{-1}\| + \|\Sigma^{-1} - \Sigma^{*-1}\|)\|\Sigma^* - \Sigma\|\|\Sigma^{-1}\|.\end{aligned}$$

Further we get

$$\frac{\|\Sigma^{-1} - \Sigma^{*-1}\|}{(C + \|\Sigma^{-1} - \Sigma^{*-1}\|)\,C} \leqslant \|\Sigma^* - \Sigma\| = o_P(1), \tag{5.13}$$

where $\|\Sigma^{-1}\| \leqslant C$. Thus $S_1 = o_P(1)$ by Lemma 5.2 and the assumption (A2). We bound the components S_2 and S_3 using the same method. Thus, we present the calculations only for S_3:

$$\begin{aligned}S_3 &= \|\Sigma^{*-1}\|\Big\|\sum_{i=1}^{p}\gamma(j+i)(\phi_i^T - \widehat{\phi}_i^{*T}) + \sum_{i=p+1}^{\infty}\gamma(j+i)\phi_i^T \\ &\qquad + \sum_{i=1}^{p}\big(\gamma(j+i) - \gamma^*(j+i)\big)\widehat{\phi}_i^{*T}\Big\| \\ &\leqslant \|\Sigma^{*-1}\|\Big(\max_{1\leqslant i\leqslant p}\|\phi_i - \widehat{\phi}_i^*\|\sum_{i=1}^{\infty}\|\gamma(i)\| + \Big\|[EX_{t,i}^2EX_{t,j}^2]_{i,j=1}^k\Big\|\sum_{i=p+1}^{\infty}\|\phi_i\| \\ &\qquad + p\max_{1\leqslant i\leqslant p}\|\gamma(i) - \gamma^*(i)\|\max_{1\leqslant i\leqslant p}\|\widehat{\phi}_i^*\|\Big) \\ &= o_{P^*}(1),\end{aligned}$$

where the bounds are a consequence of (5.10), (5.13) and Corollary 5.3. ∎

LEMMA 5.13. *Let* (LA) *hold. Then*

$$\widehat{\Sigma}_X^*(h) \xrightarrow{P^*} \Sigma_X(h) \text{ in probability}, \tag{5.14}$$

$$\widehat{\Sigma}_X^{*-1}(h) \xrightarrow{P^*} \Sigma_X^{-1}(h) \text{ in probability.} \tag{5.15}$$

P r o o f. We consider the bootstrap estimator of the mean squared prediction error in the form

$$\widehat{\Sigma}_X^*(h) = \sum_{j=0}^{h-1} \widehat{\psi}_j^* \Sigma^* \widehat{\psi}_j^{*T}.$$

Thus, using (5.10) and Lemma 5.12, we obtain

$$\begin{aligned}\Big\| \sum_{j=0}^{h-1} \widehat{\psi}_j^* \Sigma^* \widehat{\psi}_j^{*T} - \sum_{j=0}^{h-1} \psi_j \Sigma \psi_j^T \Big\| &\leqslant \sum_{j=0}^{h-1} \|\widehat{\psi}_j^* - \psi_j\| \|\Sigma^*\| \|\widehat{\psi}_j^*\| \\ &\quad + \sum_{j=0}^{h-1} \|\psi_j\| \|\Sigma^* - \Sigma\| \|\widehat{\psi}_j^*\| + \sum_{j=0}^{h-1} \|\psi_j\| \|\Sigma\| \|\widehat{\psi}_j^* - \psi_j\| \\ &= o_{P^*}(1),\end{aligned}$$

which proves (5.14). And (5.15), in the same way as (5.12), follows from the continuity of $\widehat{\Sigma}_X^{*-1}(h)$ as a function of elements of $\widehat{\Sigma}_X^*(h)$ (see [25], Theorem 5.18, p. 188). ∎

REMARK 5.1. *All proofs of the theorems and lemmas above remain valid for* $p(n)$ *chosen as in Step* 1 *of the sieve bootstrap algorithm.*

The next theorem gives us consistency of bootstrap-t. An analogous result for the one-dimensional case was given by Zagdański (see [29]).

THEOREM 5.5. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$ *we have*

$$\begin{aligned}\big| P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X_{n+h}^* - \widehat{X}_{n+h}) \leqslant u\big) \\ - P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\big| = o_P(1)\end{aligned}$$

for each u *being a continuity point of the cumulative distribution function of the random vector* $\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)$, *where* $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

P r o o f. We have the convergence

$$X_{n+h} - \widehat{X}_{n+h} \xrightarrow{D} D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h.$$

Using the multidimensional version of Slutsky's theorem ([20], Proposition C, p. 683) and Lemma 5.11, we have

$$\begin{aligned}&\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \\ &\quad \xrightarrow{D} \Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big).\end{aligned}$$

Similarly, using the conditional independence of ϵ_t^* and Lemma 5.3, we obtain convergence for the bootstrap prediction error

$$X^*_{n+h} - \widehat{X}_{n+h} \xrightarrow{D^*} D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h \text{ in probability.}$$

By the Conditional Slutsky's Theorem ([19], p. 77) and Lemma 5.13 we get

$$\begin{aligned}&\widehat{\Sigma}_X^{*-1/2}(h)(X^*_{n+h} - \widehat{X}_{n+h})\\ &\xrightarrow{D^*} \Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big) \text{ in probability.}\end{aligned}$$

Finally, we show that

$$\begin{aligned}\Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X^*_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\\ - P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\Big| = o_P(1)\end{aligned}$$

for some u being a continuity point of the cumulative distribution function of the random vector $\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)$. ∎

From Corollary 5.4, Lemmas 5.11, 5.13 and Theorem 5.5 we deduce the following corollaries.

COROLLARY 5.6. *Let* (LA) *hold. Then for* $h \in \mathbb{N}$ *we have*

$$\begin{aligned}\Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X^*_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\\ - P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \mathrm{Proj}_n X_{n+h}) \leqslant u\big)\Big| = o_P(1)\end{aligned}$$

for each u *being a continuity point of the cumulative distribution function of the random vector* $\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)$, *where* $\Phi_{h-1} = (\phi_1, \ldots, \phi_{h-1})^T$.

COROLLARY 5.7. *Let* (LA) *hold. Additionally, assume that the cumulative distribution function of the distribution* ϵ_t *is continuous. Then for* $h \in \mathbb{N}$ *we have*

$$\begin{aligned}\sup_{u\in\mathbb{R}^k}\Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X^*_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\\ - P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\Big| = o_P(1)\end{aligned}$$

and

$$\begin{aligned}\sup_{u\in\mathbb{R}^k}\Big|P^*\big(\widehat{\Sigma}_X^{*-1/2}(h)(X^*_{n+h} - \widehat{X}_{n+h}) \leqslant u\big)\\ - P\big(\widehat{\Sigma}_X^{-1/2}(h)(X_{n+h} - \mathrm{Proj}_n X_{n+h}) \leqslant u\big)\Big| = o_P(1).\end{aligned}$$

5.3. Consistency of extreme statistics. By Lemma 5.7 and the continuous mapping theorem we conclude that U_{n+h}, V_{n+h} and R_{n+h} given by formulas (4.17)–(4.19) are convergent in distribution to some random variables.

LEMMA 5.14. *Assume that* (LA) *hold and that ϵ_t has a continuous distribution function. Then, as $n \to \infty$, $U_{n+h} \xrightarrow{D} U$, $V_{n+h} \xrightarrow{D} V$ and $R_{n+h} \xrightarrow{D} R$, where U, V and R are random variables with continuous distribution functions.*

In the next lemma we show that distributions of U^*_{n+h}, V^*_{n+h} and R^*_{n+h} given by formulas (4.20)–(4.22) are close to their corresponding non-bootstrap distributions.

LEMMA 5.15. *Assuming that* (LA) *hold, we have $U^*_{n+h} \xrightarrow{D^*} U$, $V^*_{n+h} \xrightarrow{D^*} V$ and $R^*_{n+h} \xrightarrow{D^*} R$ in probability.*

P r o o f. By Lemma 5.7, Lemma 10 from [5], and Theorem 5.4 we conclude that, as $n \to \infty$,

$$W_{n+h} = X_{n+h} - \widehat{X}_{n+h} \xrightarrow{D} W = D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h,$$

and the sequence of bootstrap distributions of W^*_{n+h} weakly approaches in probability the sequence of distributions of W_{n+h}. In consequence, $W^*_{n+h} \xrightarrow{D^*} W$ in probability. Since the mappings defining U^*_{n+h}, V^*_{n+h}, R^*_{n+h} are continuous functions, the version of the continuous mapping theorem (see [4]) implies $U^*_{n+h} \xrightarrow{D^*} U$ in probability, $V^*_{n+h} \xrightarrow{D^*} V$ in probability and $R^*_{n+h} \xrightarrow{D^*} R$ in probability, which completes the proof. ∎

Using Lemmas 5.11, 5.13, and Theorem 5.5 we can prove analogous results for asymptotic distributions of US_{n+h}, VS_{n+h}, RS_{n+h} and their bootstrap versions US^*_{n+h}, VS^*_{n+h}, RS^*_{n+h} in the same way as in the proofs of the lemmas above.

LEMMA 5.16. *With* (LA) *we have, as $n \to \infty$,*

$$US_{n+h} \xrightarrow{D} US, \quad VS_{n+h} \xrightarrow{D} VS \quad \textit{and} \quad RS_{n+h} \xrightarrow{D} RS.$$

LEMMA 5.17. *With* (LA)

$$US^*_{n+h} \overset{D^*}{\cong} US, \quad VS^*_{n+h} \overset{D^*}{\cong} VS \quad \textit{and} \quad RS^*_{n+h} \overset{D^*}{\cong} RS \quad \textit{in probability}.$$

In the construction of simultaneous confidence intervals we should replace unknown quantiles with their bootstrap equivalences. Therefore, we show that bootstrap quantiles of U^*_{n+h}, V^*_{n+h} and R^*_{n+h} are approximations of the quantiles of U, V and R. This fact has been proved in [24] (see Lemma 5.2). For the clearance let us reformulate their lemma with notation stated in Lemma 5.15.

LEMMA 5.18. *With* (LA)

$$u^*_{n+h,\alpha} = u_\alpha + o_P(1), \quad v^*_{n+h,\alpha} = v_\alpha + o_P(1), \quad r^*_{n+h,\alpha} = r_\alpha + o_P(1).$$

In practice, we also do not know the exact bootstrap distributions of U^*_{n+h}, V^*_{n+h} and R^*_{n+h}. Therefore, using the result by Shi et al. [26], we apply the Monte Carlo method to approximate corresponding quantiles. In consequence, we obtain the following lemmas.

LEMMA 5.19. *With* (LA)

$$\begin{aligned}\widehat{u}^*_{n+h,\alpha} &= u_\alpha + o_P(1),\\ \widehat{v}^*_{n+h,\alpha} &= v_\alpha + o_P(1),\\ \widehat{r}^*_{n+h,\alpha} &= r_\alpha + o_P(1).\end{aligned}$$

LEMMA 5.20. *With* (LA) *and for a continuous distribution function of* ϵ_t

$$\begin{aligned}\widehat{us}^*_{n+h,(\cdot)} &= us_{(\cdot)} + o_P(1),\\ \widehat{vs}^*_{n+h,(\cdot)} &= vs_{(\cdot)} + o_P(1),\\ \widehat{rs}^*_{n+h,(\cdot)} &= rs_{(\cdot)} + o_P(1),\end{aligned}$$

where $\widehat{us}^*_{n+h,(\cdot)}$, $\widehat{vs}^*_{n+h,(\cdot)}$ *and* $\widehat{rs}^*_{n+h,(\cdot)}$ *are Monte Carlo bootstrap estimators of quantiles of* US^*_{n+h}, VS^*_{n+h}, RS^*_{n+h}, *and thus the approximations of quantiles of* US_{n+h}, VS_{n+h}, RS_{n+h}.

6. CONSISTENCY OF BOOTSTRAP PREDICTION REGIONS

In this section we present results about the consistency of the bootstrap prediction regions for the stationary time series models. Theorem 6.1 is a multivariate version of the result given by Różański and Zagdański (see [24]), and Theorem 6.3 is a multivariate version of the result given by Zagdański (see [29]).

To prove the consistency of the prediction regions we will use auxiliary results about convergence of quantiles for a weakly convergent sequence of the cumulative distribution function. The first lemma was given by Politis et al. [23] and the second one, the modification for the conditional case, was given by Różański and Zagdański [24].

LEMMA 6.1 ([23], Lemma 1.2.1). *Let* $\{F_n\}$ *be a sequence of cumulative distribution functions which converges to* F *in a weak sense and assume that* F *is continuous and strictly increasing at* $y = F^{-1}(\alpha)$ (*for* $x \in (0,1)$, $F^{-1}(x) = \inf\{y : x \leqslant F(y)\}$). *Then*

$$F_n^{-1}(\alpha) \to F^{-1}(\alpha).$$

LEMMA 6.2 ([24]). *Let $\{F_n^*\}$ be a sequence of cumulative distribution functions which converges to F in a weak sense (i.e. $F_n^* \Rightarrow F$ in probability) and assume that F is continuous and strictly increasing at $y = F^{-1}(\alpha)$. Then*

$$F_n^{*-1}(\alpha) \xrightarrow{P} F^{-1}(\alpha).$$

In Subsection 4.3 we have presented the hybrid bootstrap prediction cube given by (4.10). In practice we do not know the distribution of the random vector $H_n^*(h)$, and in consequence we are not able to compute its quantiles. Thus, we define the modified hybrid bootstrap prediction cube in the form

$$\widehat{I}_B(h) = \{X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{q}^*_{\alpha/(2k),j}, \widehat{X}_{n+h,j} + \widehat{q}^*_{1-\alpha/(2k),j}],\ j = 1, \ldots, k\}, \tag{6.1}$$

where $\widehat{q}^*_{\alpha/(2k),j}$, $\widehat{q}^*_{1-\alpha/(2k),j}$ are Monte Carlo approximations of quantiles $q^*_{\alpha/(2k),j}$, $q^*_{1-\alpha/(2k),j}$ (computed by using B bootstrap replications).

REMARK 6.1. *The replacement of the quantiles is made due to the result given by Shi et al.* [26], *i.e.*

$$q^*_{\alpha,j} - \widehat{q}^*_{\alpha,j} = o_P(1), \quad j = 1, \ldots, k,\ \alpha \in (0,1).$$

THEOREM 6.1. *Let* (LA) *hold. Additionally, assume that $c_{\alpha/(2k),j}$, $c_{1-\alpha/(2k),j}$ are continuity points of the cumulative distribution function of the random variables H_j, which are the jth coordinates of the random vector $D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h$. Then for $h \in \mathbb{N}$ we have*

$$P\big(X_{n+h} \in \widehat{I}_B(h)\big) \geqslant 1 - \alpha \quad \text{as } n \to \infty.$$

P r o o f. Using Theorem 5.4 we get for $1 \leqslant j \leqslant k$

$$X_{n+h,j} - \widehat{X}_{n+h,j} \xrightarrow{D} H_j.$$

By the result given by Różański and Zagdański [24] we have for $1 \leqslant j \leqslant k$

$$P(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{q}^*_{\alpha/(2k),j}, \widehat{X}_{n+h,j} + \widehat{q}^*_{1-\alpha/(2k),j}]) \to 1 - \alpha/k.$$

Using the Bonferroni inequality, we obtain

$$\begin{aligned} &P\big(X_{n+h} \in \widehat{I}_B(h)\big) \\ &\geqslant 1 - \sum_{j=1}^{k} P(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{q}^*_{\alpha/(2k),j}, \widehat{X}_{n+h,j} + \widehat{q}^*_{1-\alpha/(2k),j}]^c) \to 1 - \alpha. \quad \blacksquare \end{aligned}$$

In a similar way, we define $\widehat{\mathcal{E}}_B(h)$ as an equivalent of $\mathcal{E}_B(h)$ given by (4.11) in the form

$$\widehat{\mathcal{E}}_B(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T (X_{n+h} - \widehat{X}_{n+h}) \leqslant \widehat{q}^*_{1-\alpha}\}, \tag{6.2}$$

where $\widehat{q}^*_{1-\alpha}$ is the Monte Carlo approximation of quantile $q^*_{1-\alpha}$ (computed by using B bootstrap replications). The consistency of this hybrid prediction region is given in the next theorem.

THEOREM 6.2. *Let* (LA) *hold. Additionally, assume that* $c_{1-\alpha}$ *is a continuity point of the cumulative distribution function of the random variable* $\|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2$. *Then for* $h \in \mathbb{N}$

$$P\big(X_{n+h} \in \widehat{\mathcal{E}}_B(h)\big) \to 1 - \alpha.$$

P r o o f. Using continuity of the function $\|\cdot\|^2$ and Theorem 5.4, we have

$$\|H_n(h)\|^2 \xrightarrow{D} \|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2,$$
$$\|H^*_n(h)\|^2 \xrightarrow{D^*} \|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2 \text{ in probability.}$$

Let $q_{1-\alpha}$ be the $1-\alpha$ quantile of the distribution $\|H_n(h)\|^2$. By Lemmas 6.1, 6.2 and the consistency of the bootstrap sample quantiles (Remark 6.1) we have

$$q_{1-\alpha} - c_{1-\alpha} = o(1), \quad q^*_{1-\alpha} - q_{1-\alpha} = o_P(1), \quad \widehat{q}^*_{1-\alpha} - q^*_{1-\alpha} = o_P(1).$$

Notice that

$$\begin{aligned} &P\big(X_{n+h} \in \widehat{\mathcal{E}}_B(h)\big) = P\left(\|H_n(h)\|^2 \leqslant \widehat{q}^*_{1-\alpha}\right) \\ &= P\left(\|H_n(h)\|^2 \leqslant (\widehat{q}^*_{1-\alpha} - q_{1-\alpha}) + (q^*_{1-\alpha} - q_{1-\alpha}) + (q_{1-\alpha} - c_{1-\alpha}) + c_{1-\alpha}\right) \\ &= P\big(\|H_n(h)\|^2 + o_P(1) \leqslant c_{1-\alpha}\big). \end{aligned}$$

We use the Slutsky theorem to complete the proof:

$$\begin{aligned} &P\left(\|H_n(h)\|^2 + o_P(1) \leqslant c_{1-\alpha}\right) \\ &\to P\big(\|D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\|^2 \leqslant c_{1-\alpha}\big) = 1 - \alpha. \quad \blacksquare \end{aligned}$$

We define the modified bootstrap-t prediction cube $\widehat{I}_{B-t}(h)$ and the modified bootstrap-t prediction ellipse $\widehat{\mathcal{E}}_{B-t}(h)$:

$$\begin{aligned} \widehat{I}_{B-t}(h) = \{X_{n+h,j} \in [&\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h)\widehat{t}^*_{\alpha/(2k),j}, \\ &\widehat{X}_{n+h,j} + \widehat{\sigma}_{X,j}(h)\widehat{t}^*_{1-\alpha/(2k),j}],\ j = 1, \ldots, k\}, \end{aligned} \tag{6.3}$$

$$\widehat{\mathcal{E}}_{B-t}(h) = \{(X_{n+h} - \widehat{X}_{n+h})^T \widehat{\Sigma}^{-1}_{n,X}(h)(X_{n+h} - \widehat{X}_{n+h}) \leqslant \widehat{t}^{\,*}_{1-\alpha}\}, \tag{6.4}$$

where $\widehat{t}^{\,*}_{\alpha/(2k),j}, \widehat{t}^{\,*}_{1-\alpha/(2k),j}$ are the Monte Carlo approximations of the quantiles $t^*_{\alpha/(2k),j}$, $t^*_{1-\alpha/(2k),j}$, and $\widehat{t}^{\,*}_{1-\alpha}$ is the Monte Carlo approximation of the quantile $t^*_{1-\alpha}$ (all the approximations are computed by using B bootstrap replications).

THEOREM 6.3. *Let* (LA) *hold. Additionally, assume that $d_{\alpha/(2k),j}$, $d_{1-\alpha/(2k),j}$ are continuity points of the cumulative distribution functions of the random variables $T_j = H_j/\sigma_{X,j}(h)$. Then for $h \in \mathbb{N}$ we have*

$$P\big(X_{n+h} \in \widehat{I}_{B-t}(h)\big) \geqslant 1 - \alpha \quad \text{as } n \to \infty.$$

P r o o f. Using Theorem 5.4, Lemma 5.11 and the Slutsky theorem, we get for $1 \leqslant j \leqslant k$

$$\frac{X_{n+h,j} - \widehat{X}_{n+h,j}}{\widehat{\sigma}_{X,j}(h)} \xrightarrow{D} T_j.$$

By the result given by Zagdański [29] we have for $1 \leqslant j \leqslant k$

$$P\big(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{t}^{\,*}_{\alpha/(2k),j}\widehat{\sigma}_{X,j}(h), \widehat{X}_{n+h,j} + \widehat{t}^{\,*}_{1-\alpha/(2k),j}\widehat{\sigma}_{X,j}(h)]\big) \to 1-\alpha/k.$$

Using the Bonferroni inequality, we obtain

$$\begin{aligned} &P\big(X_{n+h} \in \widehat{I}_{B-t}(h)\big) \\ &\geqslant 1 - \sum_{j=1}^{k} P\big(X_{n+h,j} \in [\widehat{X}_{n+h,j} + \widehat{t}^{\,*}_{\alpha/(2k),j}\widehat{\sigma}_{X,j}(h), \widehat{X}_{n+h,j} + \widehat{t}^{\,*}_{1-\alpha/(2k),j}\widehat{\sigma}_{X,j}(h)]^c\big) \\ &\to 1 - \alpha. \quad \blacksquare \end{aligned}$$

THEOREM 6.4. *Let* (LA) *hold. Assume that $d_{1-\alpha}$ is a continuity point of the cumulative distribution function of the random variable*

$$\big\|\Sigma_X^{-1/2}(h)\big(D_{1,h}(\Phi_{h-1})\epsilon_1 + \ldots + D_{h-1,h}(\Phi_{h-1})\epsilon_{h-1} + \epsilon_h\big)\big\|^2.$$

Then for $h \in \mathbb{N}$

$$P\big(X_{n+h} \in \widehat{\mathcal{E}}_{B-t}(h)\big) \to 1 - \alpha.$$

P r o o f. The proof is analogous to the proof of Theorem 6.2. $\blacksquare$

Now, we will prove the consistency of simultaneous bootstrap prediction intervals. First, assume that we know the distribution of U, V and R. By the Slutsky lemma, we obtain the following theorem.

THEOREM 6.5. *With the assumptions from Lemma* 5.14 *we get*

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} + u_{\alpha/2}, \widehat{X}_{n+h,i} + v_{1-\alpha/2}]\big) \geqslant 1-\alpha, \tag{6.5}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} + u_{\alpha}, +\infty)\big) = 1-\alpha, \tag{6.6}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + v_{1-\alpha}]\big) = 1-\alpha, \tag{6.7}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} - r_{1-\alpha}, \widehat{X}_{n+h,i} + r_{1-\alpha}]\big) = 1-\alpha, \tag{6.8}$$

where $u_{(\cdot)}$, $v_{(\cdot)}$ *and* $r_{(\cdot)}$ *are quantiles of corresponding distributions of* U, V *and* R.

Further, we can formulate the main theorem about consistency of bootstrap prediction simultaneous intervals with theoretical quantiles in formulas (6.5)–(6.8) replaced by their Monte Carlo approximations by using the result given by Shi et al. [26]:

THEOREM 6.6. *Assume that* (LA) *hold and that* ϵ_t *has a continuous distribution. Then for each* u *being the continuity point of limited distribution we have simultaneous prediction intervals satisfying the following*:

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha/2}, \widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha/2}]\big) \geqslant 1-\alpha, \tag{6.9}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha}, +\infty)\big) = 1-\alpha, \tag{6.10}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha}]\big) = 1-\alpha, \tag{6.11}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} \quad X_{n+h,i} \in [\widehat{X}_{n+h,i} - \widehat{r}^*_{n+h,1-\alpha}, \widehat{X}_{n+h,i} + \widehat{r}^*_{n+h,1-\alpha}]\big) = 1-\alpha. \tag{6.12}$$

P r o o f. For the formula (6.9) we have the following relations:

$$\begin{aligned}
1 - P\big(\forall_{i=1,\ldots,k}\, X_{n+h,i} &\in [\widehat{X}_{n+h,i} + \widehat{u}^*_{n+h,\alpha/2}, \widehat{X}_{n+h,i} + \widehat{v}^*_{n+h,1-\alpha/2}]\big) \\
&= 1 - P\big(\min_{1\leqslant i\leqslant k}(X_{n+h,i} - \widehat{X}_{n+h,i}) \geqslant \widehat{u}^*_{n+h,\alpha/2} \\
&\qquad \wedge \max_{1\leqslant i\leqslant k}(X_{n+h,i} - \widehat{X}_{n+h,i}) \leqslant \widehat{v}^*_{n+h,1-\alpha/2}\big) \\
&= P(U_{n+h} < \widehat{u}^*_{n+h,\alpha/2} \vee V_{n+h} > \widehat{v}^*_{n+h,1-\alpha/2}) \\
&\leqslant P(U_{n+h} < \widehat{u}^*_{n+h,\alpha/2}) + P(V_{n+h} > \widehat{v}^*_{1-\alpha/2}) \\
&\approx P\big(U < o_P(1) + u_{\alpha/2}\big) + P\big(V > o_P(1) + v_{1-\alpha/2}\big) \\
&\to \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha \quad \text{as } n \to \infty,
\end{aligned}$$

where the above approximation is a consequence of Lemma 5.19, and the last convergence follows from the Slutsky lemma. For formulas (6.10)–(6.12) the reasoning is analogous. ■

As in the case of hybrid intervals defined in Theorem 6.6 we can construct studentized bootstrap intervals and prove the following theorem.

THEOREM 6.7. *With the assumptions from Lemma* 5.14 *we have the following simultaneous prediction intervals*:

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{us}^*_{n+h,\alpha/2}\widehat{\sigma}_{n+h}, \widehat{X}_{n+h,i} + \widehat{vs}^*_{n+h,1-\alpha/2}\widehat{\sigma}_{n+h}]\big) \geqslant 1-\alpha, \tag{6.13}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} + \widehat{us}^*_{n+h,\alpha}\widehat{\sigma}_{n+h}, +\infty)\big) = 1-\alpha, \tag{6.14}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in (-\infty, \widehat{X}_{n+h,i} + \widehat{vs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h}]\big) = 1-\alpha, \tag{6.15}$$

$$\lim_{n\to\infty} P\big(\forall_{i=1,\ldots,k} X_{n+h,i} \in [\widehat{X}_{n+h,i} - \widehat{rs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h}, \widehat{X}_{n+h,i} + \widehat{rs}^*_{n+h,1-\alpha}\widehat{\sigma}_{n+h}]\big) = 1-\alpha, \tag{6.16}$$

where $\widehat{us}^*_{n+h,(\cdot)}$, $\widehat{vs}^*_{n+h,(\cdot)}$ *and* $\widehat{rs}^*_{n+h,(\cdot)}$ *are Monte Carlo estimators of quantiles of bootstrap distributions of* US^*_{n+h}, VS^*_{n+h} *and* RS^*_{n+h}, *respectively.*

REMARK 6.2. *It is worth noting that by Remark* 4.1 *and Corollaries* 5.4–5.7 *all the constructed bootstrap prediction intervals and regions are asymptotically equivalent to corresponding prediction intervals and regions based on the best linear mean squared prediction of* X_{n+h}.

7. SIMULATIONS

In this section we investigate how the presented procedures work on simulated data. We consider the following $VARMA(5,4)$ model:

$$\begin{aligned} X_t = {} & \begin{bmatrix} -0.91 & 0.01 \\ 0.37 & -0.90 \end{bmatrix} X_{t-1} + \begin{bmatrix} -0.37 & 0.12 \\ 0.42 & -0.49 \end{bmatrix} X_{t-2} \\ & + \begin{bmatrix} -0.18 & 0.10 \\ 0.30 & 0.18 \end{bmatrix} X_{t-3} + \begin{bmatrix} -0.12 & 0.08 \\ 0.14 & 0.24 \end{bmatrix} X_{t-4} \\ & + \begin{bmatrix} 0.17 & -0.02 \\ 0.18 & 0.36 \end{bmatrix} X_{t-5} + \begin{bmatrix} -0.91 & 0.01 \\ 0.37 & -0.90 \end{bmatrix} \epsilon_{t-1} \\ & + \begin{bmatrix} -0.37 & 0.12 \\ 0.42 & -0.49 \end{bmatrix} \epsilon_{t-2} + \begin{bmatrix} -0.18 & 0.10 \\ 0.30 & 0.18 \end{bmatrix} \epsilon_{t-3} \\ & + \begin{bmatrix} -0.12 & 0.08 \\ 0.14 & 0.24 \end{bmatrix} \epsilon_{t-4} + \epsilon_t. \end{aligned}$$

In the model $VARMA(5,4)$ we used the following distributions for the noise process ϵ_t:

(N) normal distribution $\mathcal{N}(0,\Sigma)$,

(T) t-Student distribution $\mathcal{T}(5)$,

(χ^2) χ-square distribution $\chi^2(5)$,

(M) mixture of the normal distributions $0.1\mathcal{N}([9,9]^T,\Sigma)+0.9\mathcal{N}([-1,-1]^T,\Sigma)$.

The observations of the noise process ϵ_t for each of these distributions have been scaled (observations from χ^2 have been centered). Thus for each considered distribution we have the mean $E\epsilon_t=[0,0]^T$ and the covariance matrix

$$E\epsilon_t\epsilon_t^T=\Sigma=\begin{bmatrix}1.0 & 0.5\\ 0.5 & 1.0\end{bmatrix}.$$

On the base of the simulation results, we compare sample coverage of the prediction regions, which were computed by using the Box–Jenkins method (this method assumes normality of ϵ_t), with the sample coverage of the bootstrap prediction regions. We check performance of each method for different distributions of ϵ_t. We use t-Student distribution (T) as a heavy tailed distribution, χ-square distribution (χ^2) as a nonsymmetric distribution, and mixture (M) of the normal distributions as a bimodal distribution.

In simulations we used parameters:

- confidence level $1-\alpha=90\%$,
- forecast horizon $h=1,2,3,4,5$,
- number of observations $n=50,200$,
- number of bootstrap replications $B=1000$,
- number of Monte Carlo repetitions $N=1000$.

We have constructed three types of the prediction regions. The first type of the prediction regions is constructed by using the Bonferroni inequality and the prediction regions have cubical shape. The Box–Jenkins prediction cube is given as in equation (4.4) and the bootstrap prediction cubes have forms of hybrid bootstrap (6.1) and bootstrap-t (6.3).

The prediction regions of the second type have elliptical shape. The Box–Jenkins prediction ellipse is given by (4.3) and the bootstrap prediction ellipses are given by hybrid bootstrap (6.2) and bootstrap-t (6.4).

The third type of prediction regions is constructed by using extreme statistics, and the prediction regions have cubical shape. We have constructed the hybrid bootstrap prediction cubes $I_B^{UV}(h)$ (see (4.23)), $I_B^{R}(h)$ (see (4.26)) and bootstrap-t prediction cubes $I_{B-t}^{UV}(h)$ (see (4.27)), $I_{B-t}^{R}(h)$ (see (4.30)).

In Tables 1 and 3, we present empirical coverage of the prediction cubes, and in Table 2, we present empirical coverage of the prediction ellipses for different number of observations $n=50,200$. In the brackets, next to the empirical coverage, we present the mean area of the prediction regions.

The area of the prediction ellipse, constructed by using the Box–Jenkins method, has been calculated via the formula (see [17])

$$V\big(\mathcal{E}_{B-J}(h)\big) = \frac{\pi^{k/2}}{\Gamma\left(\frac{k}{2}+1\right)} \left(\chi^2_{1-\alpha}(k)\right)^{k/2} \Big(\det\left(\Sigma_X(h)\right)\Big)^{1/2}, \tag{7.1}$$

where $\Sigma_X(h)$ is replaced by $\widehat{\Sigma}_X(h)$.

The area of the bootstrap-t prediction ellipse has been calculated by using (7.1) with $\Sigma_X(h)$ replaced by $\widehat{\Sigma}_X(h)$ and the quantile $\chi^2_{1-\alpha}(k)$ replaced by $\widehat{t}^{\,*}_{1-\alpha}$.

It is worth noting the better performance of the bootstrap prediction regions in comparison with the performance of the prediction regions constructed via the Box–Jenkins method.

In all cases we observe that empirical coverage of bootstrap-t prediction regions is larger than empirical coverage of hybrid bootstrap prediction regions but bootstrap-t prediction regions have larger areas.

The empirical coverage of the bootstrap prediction cubes is similar to empirical coverage of the bootstrap prediction ellipses. However, areas of the bootstrap prediction ellipses are smaller than areas of the bootstrap prediction cubes.

The bootstrap prediction regions constructed with extreme statistics are more stable than the bootstrap prediction regions constructed by using the Bonferroni inequality.

TABLE 1. The empirical coverage of the bootstrap prediction cubes for the model $VARMA(5,4)$.

Distribution	h	$n = 50$			$n = 200$		
		Box–Jenkins	hybrid bootstrap	bootstrap-t	Box–Jenkins	hybrid bootstrap	bootstrap-t
N	1	77.2 (3.1)	83.1 (3.8)	89.0 (4.7)	86.1 (3.1)	88.1 (3.4)	90.3 (3.6)
	2	75.6 (5.2)	80.1 (6.2)	86.6 (8.0)	85.2 (5.5)	86.7 (5.8)	88.9 (6.4)
	3	77.0 (6.0)	82.0 (7.1)	87.8 (9.1)	87.2 (6.3)	88.1 (6.7)	89.9 (7.4)
	4	79.3 (6.7)	82.7 (7.7)	87.4 (9.7)	87.7 (7.0)	89.1 (7.4)	91.3 (8.2)
	5	79.3 (7.2)	82.3 (8.2)	87.5 (10.3)	87.7 (7.7)	89.2 (8.1)	90.6 (8.9)
T	1	78.0 (3.1)	82.6 (3.9)	88.0 (4.9)	86.7 (3.1)	88.4 (3.4)	90.6 (3.8)
	2	77.0 (5.2)	82.7 (6.4)	88.9 (8.3)	85.6 (5.6)	87.3 (6.0)	89.2 (6.7)
	3	77.7 (6.1)	82.6 (7.3)	87.1 (9.4)	85.8 (6.4)	87.6 (6.8)	89.4 (7.6)
	4	80.1 (6.7)	83.4 (7.8)	89.0 (10.1)	87.4 (7.1)	88.4 (7.6)	90.4 (8.5)
	5	80.1 (7.3)	83.7 (8.4)	89.1 (10.7)	87.4 (7.8)	88.3 (8.3)	90.6 (9.2)
χ^2	1	75.6 (3.1)	81.7 (3.9)	86.6 (4.8)	85.4 (3.1)	86.5 (3.5)	88.7 (3.8)
	2	77.7 (5.2)	80.7 (6.3)	87.3 (8.2)	84.2 (5.5)	85.5 (5.9)	87.9 (6.6)
	3	76.0 (6.1)	81.3 (7.2)	87.2 (9.2)	84.3 (6.3)	85.2 (6.8)	87.7 (7.5)
	4	76.9 (6.7)	81.7 (7.8)	86.9 (9.9)	85.9 (7.0)	86.9 (7.5)	88.3 (8.3)
	5	76.9 (7.3)	80.3 (8.3)	85.2 (10.5)	85.9 (7.7)	86.4 (8.2)	88.2 (9.1)
M	1	74.7 (3.0)	80.8 (4.0)	85.6 (5.1)	81.1 (3.1)	86.0 (4.0)	87.6 (4.4)
	2	73.4 (5.2)	77.8 (6.5)	83.6 (8.6)	81.0 (5.5)	83.5 (6.3)	85.2 (7.0)
	3	72.8 (6.1)	77.1 (7.4)	82.3 (9.6)	80.2 (6.4)	83.4 (7.1)	85.2 (7.9)
	4	76.2 (6.7)	79.4 (8.0)	84.9 (10.3)	80.2 (6.4)	83.4 (7.1)	85.2 (7.9)
	5	76.2 (7.4)	78.2 (8.6)	83.6 (11.0)	83.4 (7.8)	85.6 (8.6)	86.5 (9.5)

TABLE 2. The empirical coverage of the bootstrap prediction ellipses for the model $VARMA(5,4)$.

Distribution	h	n = 50			n = 200		
		Box–Jenkins	hybrid bootstrap	bootstrap-t	Box–Jenkins	hybrid bootstrap	bootstrap-t
N	1	75.6 (2.8)	85.3 (5.7)	88.7 (4.4)	84.6 (2.9)	87.6 (5.1)	88.9 (3.3)
	2	75.0 (4.8)	84.5 (8.8)	86.3 (7.6)	84.3 (5.1)	87.9 (7.7)	88.5 (5.9)
	3	74.0 (5.5)	83.9 (9.4)	86.0 (8.5)	85.9 (5.7)	89.4 (8.3)	89.1 (6.7)
	4	78.6 (6.2)	85.0 (10.1)	87.6 (9.2)	86.0 (6.5)	88.9 (9.3)	89.6 (7.5)
	5	78.6 (6.7)	84.8 (11.0)	88.2 (9.8)	86.0 (7.2)	89.7 (10.4)	90.0 (8.3)
T	1	77.2 (2.8)	86.6 (5.7)	88.1 (4.5)	86.7 (2.9)	89.4 (5.0)	90.1 (3.4)
	2	75.7 (4.8)	85.6 (9.0)	88.6 (7.8)	85.8 (5.1)	87.8 (7.8)	89.0 (6.1)
	3	76.9 (5.6)	85.3 (9.7)	87.6 (8.8)	86.4 (5.8)	87.8 (8.4)	88.7 (6.8)
	4	77.7 (6.2)	85.5 (10.3)	89.1 (9.4)	85.8 (6.6)	88.9 (9.4)	89.4 (7.7)
	5	77.7 (6.8)	85.4 (11.3)	89.3 (10.1)	85.8 (7.3)	88.8 (10.5)	88.8 (8.5)
χ^2	1	74.4 (2.8)	83.1 (5.7)	86.8 (4.5)	85.1 (2.9)	87.1 (5.1)	88.6 (3.4)
	2	74.7 (4.8)	86.0 (8.9)	87.3 (7.8)	83.9 (5.1)	87.4 (7.8)	88.3 (6.0)
	3	75.7 (5.5)	84.6 (9.5)	87.6 (8.6)	84.0 (5.7)	88.2 (8.4)	87.8 (6.8)
	4	75.1 (6.2)	82.8 (10.2)	87.1 (9.3)	85.4 (6.5)	86.7 (9.3)	89.0 (7.6)
	5	75.1 (6.8)	83.4 (11.1)	85.5 (9.9)	85.4 (7.2)	86.3 (10.5)	88.7 (8.4)
M	1	73.3 (2.8)	82.8 (5.8)	86.8 (4.8)	80.7 (2.9)	88.3 (5.8)	88.3 (4.0)
	2	71.5 (4.8)	81.7 (9.0)	83.9 (8.1)	81.1 (5.1)	86.0 (8.3)	85.5 (6.6)
	3	72.3 (5.6)	83.1 (9.6)	83.8 (9.0)	79.8 (5.8)	84.0 (8.9)	85.2 (7.3)
	4	75.0 (6.2)	83.5 (10.3)	84.1 (9.6)	82.7 (6.6)	88.1 (9.8)	86.7 (8.1)
	5	75.0 (6.8)	82.7 (11.2)	84.6 (10.3)	82.7 (7.3)	88.3 (10.9)	87.6 (8.8)

TABLE 3. The empirical coverage of the bootstrap prediction cubes constructed by extreme statistics for the model $VARMA(5,4)$.

Distri-bution	h	n = 50				n = 200			
		UV	R	UV-t	R-t	UV	R	UV-t	R-t
N	1	85.6 (6.6)	85.6 (6.7)	88.5 (4.7)	89.1 (4.7)	88.1 (6.0)	88.0 (6.0)	90.2 (3.6)	89.7 (3.5)
	2	85.1 (10.0)	85.1 (10.3)	86.3 (7.9)	87.6 (8.0)	88.2 (8.9)	88.8 (9.0)	88.5 (6.3)	88.8 (6.3)
	3	83.7 (10.6)	84.5 (10.8)	87.6 (8.9)	88.3 (9.0)	89.0 (9.5)	89.6 (9.6)	89.8 (7.2)	89.4 (7.2)
	4	84.9 (11.4)	86.0 (11.7)	87.0 (9.6)	87.6 (9.6)	88.4 (10.7)	88.7 (10.7)	90.9 (8.0)	90.7 (8.0)
	5	84.5 (12.4)	86.1 (12.7)	87.7 (10.2)	88.8 (10.3)	90.6 (12.0)	90.6 (12.1)	90.6 (8.7)	90.5 (8.7)
T	1	86.2 (6.6)	87.4 (6.8)	87.5 (4.8)	87.5 (4.7)	89.0 (5.9)	89.2 (6.0)	90.0 (3.7)	89.3 (3.6)
	2	84.6 (10.3)	85.6 (10.5)	87.4 (8.1)	88.1 (8.2)	87.9 (9.0)	88.1 (9.0)	88.5 (6.4)	88.6 (6.4)
	3	84.7 (10.9)	85.4 (11.2)	86.8 (9.2)	87.7 (9.2)	87.7 (9.5)	87.7 (9.6)	88.8 (7.3)	89.1 (7.3)
	4	85.4 (11.7)	85.8 (11.9)	87.7 (9.9)	88.7 (9.9)	89.4 (10.7)	89.5 (10.8)	90.0 (8.2)	89.8 (8.1)
	5	85.4 (12.8)	85.6 (13.1)	88.7 (10.6)	88.7 (10.6)	89.0 (12.1)	89.6 (12.1)	89.8 (8.9)	89.4 (8.9)
χ^2	1	82.7 (6.6)	83.7 (6.7)	86.2 (4.8)	87.1 (4.8)	87.6 (6.0)	87.9 (6.0)	88.6 (3.7)	88.4 (3.6)
	2	85.0 (10.1)	86.4 (10.4)	87.1 (8.0)	87.0 (8.1)	87.2 (9.0)	87.8 (9.0)	88.0 (6.3)	87.4 (6.3)
	3	83.5 (10.7)	85.1 (10.9)	86.7 (9.0)	87.4 (9.0)	87.8 (9.5)	88.2 (9.5)	87.6 (7.2)	87.6 (7.2)
	4	82.5 (11.5)	84.2 (11.7)	86.7 (9.7)	86.9 (9.8)	86.1 (10.6)	87.3 (10.7)	88.0 (8.1)	88.3 (8.1)
	5	82.7 (12.6)	83.8 (12.8)	85.1 (10.4)	85.8 (10.4)	86.0 (12.0)	86.5 (12.1)	88.2 (8.9)	88.2 (8.8)
M	1	82.8 (6.8)	84.0 (6.9)	84.9 (5.0)	86.1 (5.0)	87.7 (6.7)	88.1 (6.8)	87.4 (4.3)	87.3 (4.2)
	2	82.5 (10.3)	83.0 (10.5)	83.8 (8.4)	85.0 (8.4)	86.8 (9.5)	86.5 (9.5)	84.6 (6.7)	85.0 (6.7)
	3	83.7 (10.9)	84.4 (11.1)	82.3 (9.4)	82.8 (9.4)	84.2 (10.0)	84.2 (10.0)	84.5 (7.6)	85.1 (7.6)
	4	83.1 (11.6)	84.1 (11.8)	84.6 (10.0)	84.9 (10.1)	88.5 (11.1)	88.2 (11.1)	86.1 (8.5)	86.3 (8.4)
	5	82.9 (12.8)	83.7 (13.0)	84.1 (10.8)	83.8 (10.8)	88.4 (12.5)	88.9 (12.6)	86.4 (9.2)	86.6 (9.2)

REFERENCES

[1] A. M. Alonso, D. Peña, and J. Romo, *On sieve bootstrap prediction intervals*, Statist. Probab. Lett. 65 (1) (2003), pp. 13–20.

[2] A. M. Alonso, D. Peña, and J. Romo, *Introducing model uncertainty by moving blocks bootstrap*, Statist. Papers 47 (2) (2006), pp. 167–179.

[3] T. W. Anderson, *The Statistical Analysis of Time Series*, Wiley, New York 1971.

[4] Yu. K. Belyaev and O. Seleznjev, *Approaching in distribution with applications to resampling of stochastic processes*, Scand. J. Stat. 27 (2) (2000), pp. 371–384.

[5] Yu. K. Belyaev and S. Sjöstedt-de Luna, *Weakly approaching sequences of random distributions*, J. Appl. Probab. 37 (3) (2000), pp. 807–822.

[6] P. J. Bickel and D. A. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Statist. 9 (6) (1981), pp. 1196–1217.

[7] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, second edition, Springer, New York 1991.

[8] P. Bühlmann, *Moving-average representation of autoregressive approximations*, Stochastic Process. Appl. 60 (2) (1995), pp. 331–342.

[9] P. Bühlmann, *Sieve bootstrap for time series*, Bernoulli 3 (2) (1997), pp. 123–148.

[10] P. Bühlmann, *Sieve bootstrap for smoothing in nonstationary time series*, Ann. Statist. 26 (1) (1998), pp. 48–83.

[11] T. Di Battista, S. A. Gattone, and M. Granturco, *VAR-models for predicting biodiversity*, J. Math. Stat. 1 (4) (2005), pp. 312–315.

[12] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, New York 1993.

[13] D. Fresoli, E. Ruiz, and L. Pascual, *Bootstrap multi-step forecasts of non-Gaussian VAR models*, Int. J. Forecast. 31 (2015), pp. 834–848.

[14] U. Grenander, *Abstract Inference*, Wiley Ser. Probab. Math. Statist., Wiley, New York 1981.

[15] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*, Wiley, New York 1988.

[16] E. J. Hannan and L. Kavalieris, *Regression, autoregression models*, J. Time Series Anal. 7 (1) (1986), pp. 27–49.

[17] J. H. Kim, *Asymptotic and bootstrap prediction regions for vector autoregression*, Int. J. Forecast. 15 (1999), pp. 393–403.

[18] J.-P. Kreiss, *Bootstrap procedures for AR(∞)-processes*, in: *Bootstrapping and Related Techniques*, K. H. Jöckel, G. Rothe, and W. Sendler (Eds.), Springer, New York 1992, pp. 107–113.

[19] S. N. Lahiri, *Resampling Methods for Dependent Data*, Springer, New York 2003.

[20] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer, Berlin 2005.

[21] M. Meyer and J. Kreiss, *On the vector autoregressive sieve bootstrap*, J. Time Series Anal. 36 (3) (2015), pp. 377–397.

[22] S. Mirmirani and H. C. Li, *A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil*, Adv. Econom. 19 (2004), pp. 203–223.

[23] D. N. Politis, J. P. Romano, and M. Wolf, *Subsampling*, Springer, New York 1999.

[24] R. Różański and A. Zagdański, *On the consistency of sieve bootstrap prediction intervals for stationary time series*, Discuss. Math. Probab. Stat. 24 (1) (2004), pp. 5–40.

[25] J. R. Schott, *Matrix Analysis for Statistics*, third edition, Wiley, Hoboken, NJ, 2017.

[26] X. Q. Shi, C. J. Wu, and J. H. Chen, *Weak and strong representations for quantile processes from finite populations with application to simulation size in resampling inference*, Canad. J. Statist. 18 (2) (1990), pp. 141–148.

[27] R. S. Tsay, *Analysis of Financial Time Series*, third edition, Wiley, Hoboken, NJ, 2010.

[28] R. S. Tsay, *Multivariate Time Series Analysis: With R and Financial Applications*, Wiley, Hoboken, NJ, 2014.

[29] A. Zagdański, *On the construction and properties of bootstrap-t prediction intervals for stationary time series*, Probab. Math. Statist. 25 (1) (2005), pp. 133–153.
[30] A. Zygmund, *Trigonometric Series*, Vol. I, Cambridge University Press, London–New York 1968.

Roman Różański and Adam Zagdański
(corresponding authors)
Faculty of Pure and Applied Mathematics
Wrocław University of Science and Technology
27 Wybrzeże Wyspiańskiego
50-370 Wrocław, Poland
E-mail: Roman.Rozanski@pwr.edu.pl
E-mail: Adam.Zagdanski@pwr.edu.pl

Grzegorz Chłapiński
KRUK S.A.

Marcin Hławka
Krajowy Rejestr Długów
Biuro Informacji Gospodarczej SA

Krzysztof Jamróz
Santander Bank Polska S.A.

Maciej Kawecki
IT Consulting Maciej Kawecki

Received on 11.10.2015;
revised version on 8.4.2017

Probability and Mathematical Statistics 38, z. 2, 2018

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 359–384
doi:10.19195/0208-4147.38.2.6

LIMITING SPECTRAL DISTRIBUTIONS OF SUMS OF PRODUCTS OF NON-HERMITIAN RANDOM MATRICES*

BY

HOLGER KÖSTERS (BIELEFELD) AND ALEXANDER TIKHOMIROV (SYKTYVKAR)

Abstract. For fixed $l \geqslant 0$ and $m \geqslant 1$, let $\mathbf{X}_n^{(0)}, \mathbf{X}_n^{(1)}, \ldots, \mathbf{X}_n^{(l)}$ be independent random $n \times n$ matrices with independent entries, let $\mathbf{F}_n^{(0)} := \mathbf{X}_n^{(0)}(\mathbf{X}_n^{(1)})^{-1} \ldots (\mathbf{X}_n^{(l)})^{-1}$, and let $\mathbf{F}_n^{(1)}, \ldots, \mathbf{F}_n^{(m)}$ be independent random matrices of the same form as $\mathbf{F}_n^{(0)}$. We show that as $n \to \infty$, the matrices $\mathbf{F}_n^{(0)}$ and $m^{-(l+1)/2}(\mathbf{F}_n^{(1)} + \ldots + \mathbf{F}_n^{(m)})$ have the same limiting eigenvalue distribution.

To obtain our results, we apply the general framework recently introduced in Götze, Kösters, and Tikhomirov (2015) to sums of products of independent random matrices and their inverses. We establish the universality of the limiting singular value and eigenvalue distributions, and we provide a closer description of the limiting distributions in terms of free probability theory.

2010 AMS Mathematics Subject Classification: Primary: 60B20; Secondary: 60E07, 60F05, 46L54.

Key words and phrases: Non-Hermitian random matrices, limiting spectral distributions, free probability theory, stable distributions.

1. INTRODUCTION AND MAIN RESULTS

The investigation of the asymptotic spectral distributions of random matrices is a major topic in random matrix theory. In recent years, sums and products of independent *non-Hermitian* random matrices with independent entries have found increasing attention; see e.g. [1], [3], [11], [13]–[18], [20], [25], [27], [29], [31], [35]–[37] for results on *global* spectral distributions, and also the survey paper [2] and the references therein for results on *local* spectral distributions. In particular, the paper [18] provides a general framework for the investigation of the limiting (global) spectral distributions of products of independent random matrices with independent entries. Furthermore, the paper [37] shows that this approach proves useful for the investigation of sums of products as well. The aim of the present

* Research supported by CRC 701.

paper is to show that certain products of independent random matrices give rise to random matrices with *stable* limiting eigenvalue distributions, in the sense that the sums of several independent copies of these products have the same limiting eigenvalue distribution after appropriate rescaling.

Throughout this paper, for each $n \geqslant 1$, let $\mathbf{X}_n^{(1)}, \mathbf{X}_n^{(2)}, \mathbf{X}_n^{(3)}, \ldots$ be independent random matrices of size $n \times n$ with independent entries. More precisely, we assume that

$$\mathbf{X}_n^{(q)} = \Big(\tfrac{1}{\sqrt{n}} X_{jk}^{(q)}\Big)_{j,k=1,\ldots,n}, \tag{1.1}$$

where $(X_{jk}^{(q)})_{j,k,q\in\mathbb{N}}$ is a family of independent real or complex random variables such that

$$\mathbb{E}X_{jk}^{(q)} = 0,\ \mathbb{E}(X_{jk}^{(q)})^2 = 1 \quad \text{in the real case,} \tag{1.2}$$

and

$$\mathbb{E}X_{jk}^{(q)} = 0,\ \mathbb{E}(X_{jk}^{(q)})^2 = 0,\ \mathbb{E}|X_{jk}^{(q)}|^2 = 1 \quad \text{in the complex case,} \tag{1.3}$$

and we additionally assume that this family is uniformly square-integrable, i.e.

$$\lim_{a\to\infty} \sup_{j,k,q\in\mathbb{N}} \mathbb{E}\big(|X_{jk}^{(q)}|^2\, \mathbf{1}_{\{|X_{jk}^{(q)}|\geqslant a\}}\big) = 0. \tag{1.4}$$

In this case we also say the matrices $\mathbf{X}_n^{(q)}$ are independent *Girko–Ginibre matrices*. In the special case where the entries have real or complex Gaussian distributions, we usually write $\mathbf{Y}_n^{(q)} = \big(\frac{1}{\sqrt{n}} Y_{jk}^{(q)}\big)_{j,k=1,\ldots,n}$ instead of $\mathbf{X}_n^{(q)} = \big(\frac{1}{\sqrt{n}} X_{jk}^{(q)}\big)_{j,k=1,\ldots,n}$ and call the matrices $\mathbf{Y}_n^{(q)}$ *Gaussian random matrices* or *Ginibre matrices*. Note that the assumption (1.4) is clearly satisfied in this special case, the random variables $Y_{jk}^{(q)}$ being independent and identically distributed (i.i.d.).

We will be interested in the limiting spectral distributions of random matrices $\mathbf{F}_n$ given by sums of products of the matrices $\mathbf{X}_n^{(q)}$ and their inverses. Let $\mathbf{F}_n$ have the singular values $s_1(\mathbf{F}_n) \geqslant \ldots \geqslant s_n(\mathbf{F}_n)$ and eigenvalues $\lambda_1(\mathbf{F}_n), \ldots, \lambda_n(\mathbf{F}_n)$. Then we write $\nu_n := \nu(\mathbf{F}_n) := \frac{1}{n}\sum_{j=1}^n \delta_{s_j(\mathbf{F}_n)}$ for the (empirical) *singular value distribution* of $\mathbf{F}_n$ and $\mu_n := \mu(\mathbf{F}_n) := \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j(\mathbf{F}_n)}$ for the (empirical) *eigenvalue distribution* of $\mathbf{F}_n$. The corresponding weak limits in probability (if existent) will be denoted by $\nu := \nu_{\mathbf{F}}$ and $\mu := \mu_{\mathbf{F}}$, respectively. Note that μ will in general be a probability measure on the complex plane, the random matrices $\mathbf{F}_n$ being non-Hermitian. The density of μ (if existent) will be denoted by $f(z)$, or by $f(r)$ (with $r = |z|$) in case it is rotation-invariant with respect to the origin.

Let us mention some relevant results from the literature.

EXAMPLES 1.1.

(a) (*Circular law*) Let $\mathbf{F}_n = \mathbf{X}_n^{(1)}$. Then $f(r) = \frac{1}{\pi}\mathbf{1}_{[0,1]}(r)$, i.e. μ is the uniform distribution on the unit disk.

(b) Let $\mathbf{F}_n = \mathbf{X}_n^{(1)} + \ldots + \mathbf{X}_n^{(m)}$. Then $\mathbf{F}_n$ is a random matrix with independent entries of mean zero and variance m/n, so, by simple rescaling, $f(r) = \frac{1}{m\pi}\mathbf{1}_{[0,\sqrt{m}]}(r)$. In particular, for the rescaled matrices $\frac{1}{\sqrt{m}}\mathbf{F}_n$, the limiting eigenvalue distribution is again the uniform distribution on the unit disk.

(c) Let $\mathbf{F}_n = \mathbf{X}_n^{(1)}\mathbf{X}_n^{(2)}$. Then $f(r) = \frac{1}{2\pi r}\mathbf{1}_{[0,1]}(r)$, i.e. μ is the induced distribution of the uniform distribution on the unit disk under the mapping $z \mapsto z^2$. See e.g. [18], Section 8.2.2, for a "simple" derivation.

(d) Let $\mathbf{F}_n = \mathbf{X}_n^{(1)}\mathbf{X}_n^{(2)} + \ldots + \mathbf{X}_n^{(2m-1)}\mathbf{X}_n^{(2m)}$. Then

$$f(r) = \frac{1}{\pi\sqrt{(m-1)^2 + 4r^2}}\,\mathbf{1}_{[0,\sqrt{m}]}(r);$$

see [37], Section 2.

(e) (*Spherical law*) Let $\mathbf{F}_n = \mathbf{X}_n^{(1)}(\mathbf{X}_n^{(2)})^{-1}$. Then

$$f(r) = \frac{1}{\pi(1+r^2)^2},$$

i.e. μ is the spherical distribution on the complex plane.

(f) Let $\mathbf{F}_n = \mathbf{X}_n^{(1)}(\mathbf{X}_n^{(2)})^{-1} + \ldots + \mathbf{X}_n^{(2m-1)}(\mathbf{X}_n^{(2m)})^{-1}$. Then

$$f(r) = \frac{m^2}{\pi(m^2+r^2)^2};$$

see [37], Section 3. Thus, for the rescaled matrices $\frac{1}{m}\mathbf{F}_n$, the limiting eigenvalue distribution is again the spherical distribution on the complex plane.

In view of examples (b) and (f), it seems natural to ask whether there exist further examples of random matrices $\mathbf{F}_n^{(0)}$ such that for any $m \in \mathbb{N}$, the sums of m independent matrices of the same form as $\mathbf{F}_n^{(0)}$ have the same limiting eigenvalue distribution as the original random matrices $\mathbf{F}_n^{(0)}$, after appropriate rescaling. We will answer this question in the affirmative by proving the following result, which contains examples (b) and (f) as special cases:

THEOREM 1.1. *Fix $m \in \mathbb{N}$ and $l \in \mathbb{N}_0$, let*

$$\mathbf{F}_n^{(0)} := (\mathbf{X}_n^{(0)})(\mathbf{X}_n^{(1)})^{-1} \ldots (\mathbf{X}_n^{(l)})^{-1}, \tag{1.5}$$

where $\mathbf{X}_n^{(0)}, \mathbf{X}_n^{(1)}, \ldots, \mathbf{X}_n^{(l)}$ are independent random matrices as in (1.1)–(1.4), *and let $\mathbf{F}_n^{(1)}, \ldots, \mathbf{F}_n^{(m)}$ be independent matrices of the same form as $\mathbf{F}_n^{(0)}$. Then the matrices $m^{-(l+1)/2}(\mathbf{F}_n^{(1)} + \ldots + \mathbf{F}_n^{(m)})$ and $\mathbf{F}_n^{(0)}$ have the same limiting eigenvalue distribution μ. More precisely, we have $\mu = \mathcal{H}\big(\sigma_s\big(\frac{2}{l+1}\big)\big)$, where $\sigma_s\big(\frac{2}{l+1}\big)$ is the symmetric $\boxplus$-stable distribution with parameter $\frac{2}{l+1}$* (see Section 2.3) *and $\mathcal{H}\big(\sigma_s\big(\frac{2}{l+1}\big)\big)$ is the associated rotation-invariant distribution on $\mathbb{C}$* (see Section 2.1).

Moreover, as we will see in Section 3, apart from a possible permutation of the exponents ± 1, the matrices $\mathbf{F}_n^{(0)}$ in Theorem 1.1 are the only examples of products of *independent* Girko–Ginibre matrices and their inverses such that for any $m \in \mathbb{N}$, $\mathbf{F}_n^{(0)}$ and $\mathbf{F}_n^{(1)} + \ldots + \mathbf{F}_n^{(m)}$ have the same limiting eigenvalue distribution after appropriate rescaling. In particular, the matrices

$$\mathbf{F}_n^{(0)} := \mathbf{X}_n^{(1)} \ldots \mathbf{X}_n^{(k)} (\mathbf{X}_n^{(k+1)})^{-1} \ldots (\mathbf{X}_n^{(k+l)})^{-1} \tag{1.6}$$

with $k > 1$ do not share this property.

However, the same limiting eigenvalue distributions may arise for products involving *powers* of random matrices:

THEOREM 1.2. *Fix $m \in \mathbb{N}$, $k \in \mathbb{N}_0$ and $l_1, \ldots, l_k \in \mathbb{N}$, let $l := l_1 + \ldots + l_k$ and define*

$$\mathbf{F}_n^{(0)} := (\mathbf{X}_n^{(0)})(\mathbf{X}_n^{(1)})^{-l_1} \ldots (\mathbf{X}_n^{(k)})^{-l_k}, \tag{1.7}$$

where $\mathbf{X}_n^{(0)}, \mathbf{X}_n^{(1)}, \ldots, \mathbf{X}_n^{(k)}$ are independent random matrices as in (1.1)–(1.4)*, and let $\mathbf{F}_n^{(1)}, \ldots, \mathbf{F}_n^{(m)}$ be independent matrices of the same form as $\mathbf{F}_n^{(0)}$. Then the matrices $m^{-(l+1)/2}(\mathbf{F}_n^{(1)} + \ldots + \mathbf{F}_n^{(m)})$ and $\mathbf{F}_n^{(0)}$ have the same limiting eigenvalue distribution μ, which is the same as in Theorem* 1.1.

Theorem 1.1 will be deduced from a more general result about random matrices of the form

$$\mathbf{F}_n(\mathbf{X}) := \sum_{q=1}^{m} \mathbf{F}_n^{(q)}(\mathbf{X}) := \sum_{q=1}^{m} \prod_{r=1}^{l} (\mathbf{X}_n^{((q-1)l+r)})^{\varepsilon_r}, \tag{1.8}$$

where $m, l \in \mathbb{N}$ and $\varepsilon_1, \ldots, \varepsilon_l \in \{+1, -1\}$ are fixed. (Thus, the matrices $\mathbf{F}_n^{(q)}(\mathbf{X})$ are independent random matrices of the same form as the matrix $\prod_{r=1}^{l} (\mathbf{X}_n^{(r)})^{\varepsilon_r}$.) Let us note that under the assumptions (1.1)–(1.4), each matrix $\mathbf{X}_n^{(r)}$ is invertible with probability $1 + o(1)$ as $n \to \infty$ (see e.g. Lemma 4.9), so that $\mathbf{F}_n(\mathbf{X})$ is defined with probability $1 + o(1)$ as $n \to \infty$. Here we have the following result, which establishes the existence of the limiting singular value and eigenvalue distributions and provides a closer description of them in terms of free probability theory:

THEOREM 1.3. *Let the matrices $\mathbf{F}_n(\mathbf{X})$ be defined as in* (1.8)*. Then there exist non-random probability measures ν and μ on $(0, \infty)$ and $\mathbb{C}$, respectively, such that $\lim_{n\to\infty} \nu\big(\mathbf{F}_n(\mathbf{X})\big) = \nu$ and $\lim_{n\to\infty} \mu\big(\mathbf{F}_n(\mathbf{X})\big) = \mu$ weakly in probability, and the limiting distributions are the same as those for the matrices $\mathbf{F}_n(\mathbf{Y})$ derived from Gaussian random matrices. More precisely, the limiting measures ν and μ are given by*

$$\mathcal{S}\nu = \big(\mathcal{Q}^{-1}(\gamma^{\varepsilon_1} \boxtimes \ldots \boxtimes \gamma^{\varepsilon_l})\big)^{\boxplus m} \quad \text{and} \quad \mu = \mathcal{H}(\mathcal{S}\nu), \tag{1.9}$$

where γ is the Marchenko–Pastur distribution, γ^{-1} is the inverse Marchenko–Pastur distribution, $\boxplus$ and $\boxtimes$ denote the additive and multiplicative free convolution, and $\mathcal{S}$, $\mathcal{Q}$ and $\mathcal{H}$ are the operators described in Section 2.1 *below.*

In particular, this result shows that the limiting spectral distributions are *universal*, i.e. they do not depend on the distributions of the matrix entries apart from a few moment conditions as in (1.2)–(1.4).

REMARK 1.1. *As will follow from the proof, Theorem* 1.3 *extends to* certain *sums of products of powers of independent Girko–Ginibre matrices and their inverses, namely to random matrices of the form*

$$\mathbf{F}_n(\mathbf{X}) := \sum_{q=1}^{m} \mathbf{F}_n^{(q)}(\mathbf{X}) := \sum_{q=1}^{m} \prod_{r=1}^{k} \big((\mathbf{X}_n^{((q-1)k+r)})^{\varepsilon_r}\big)^{l_r}, \tag{1.10}$$

where $m, k \in \mathbb{N}$, $\varepsilon_1, \ldots, \varepsilon_k \in \{-1, +1\}$ and $l_1, \ldots, l_k \in \mathbb{N}$ are fixed, and

$$\text{for some } r = 1, \ldots, k, \text{ we have } l_r = 1. \tag{1.11}$$

Here, with the notation as above, the limiting measures ν and μ are given by

$$\mathcal{S}\nu = \Big(\mathcal{Q}^{-1}\big((\gamma^{\varepsilon_1})^{\boxtimes l_1} \boxtimes \ldots \boxtimes (\gamma^{\varepsilon_k})^{\boxtimes l_k}\big)\Big)^{\boxplus m} \quad \text{and} \quad \mu = \mathcal{H}(\mathcal{S}\nu). \tag{1.12}$$

This will be important for the proof of Theorem 1.2.

To obtain the preceding results, we apply the general framework from [18] for the investigation of (global) limiting spectral distributions to sums of products of independent Girko–Ginibre random matrices and their inverses (see Section 4). Related results for various special cases can be found e.g. in [1], [3], [11], [13], [14], [16], [20], [25], [27], [29], [35], [37]. In particular, in the Gaussian case, the limiting eigenvalue and singular value distributions of the products (1.6) were recently obtained in [1] and [16], respectively.

To apply the framework from [18], we need to verify certain technical conditions, see Conditions A, B and C in Section 4.2 for details. This will be achieved by means of a suitable induction argument, which forms the major part of Section 4 and which represents the main contribution of this work. Furthermore, to identify the limiting spectral distributions, we use tools from free probability theory. Here it is worth emphasizing that for the matrices in Theorems 1.1 and 1.2 the limiting spectral distributions may be described relatively explicitly. It seems that comparable results are available only in a few special cases, see e. g. [10], [21], [22], [26]. Let us mention, however, the very recent work [6], [33], [7] which provides an algorithm for calculating the Brown measures of general polynomials in free non-commutative random variables. This should yield many further examples where the limiting spectral distributions may now be determined.

2. BACKGROUND

In this section we recall some well-known concepts and results from the literature which will be needed later.

2.1. Results from random matrix theory. The derivation of our results on limiting eigenvalue distributions will be based on Girko's *Hermitization method* (see also [12]). Thus, we will first study the limiting eigenvalue distributions of the *Hermitian* matrices

$$\mathbf{V}_n := \begin{bmatrix} \mathbf{O} & \mathbf{F}_n \\ \mathbf{F}_n^* & \mathbf{O} \end{bmatrix} \quad \text{and} \quad \mathbf{W}_n := \mathbf{F}_n \mathbf{F}_n^*. \tag{2.1}$$

Note that if the singular values of $\mathbf{F}_n$ are given by $s_1, \dots, s_n$, then the eigenvalues of $\mathbf{V}_n$ and $\mathbf{W}_n$ are given by $\pm s_1, \dots, \pm s_n$ and $s_1^2, \dots, s_n^2$, respectively. It is easy to see that knowledge of one of the distributions $\nu(\mathbf{F}_n), \mu(\mathbf{V}_n), \mu(\mathbf{W}_n)$ (or its convergence) implies knowledge of the other two (or their convergence). More precisely, if $\mathcal{S}$ denotes the operator which associates with each distribution ν on $(0,\infty)$ its symmetrization on $\mathbb{R}^*$, and $\mathcal{Q}$ denotes the operator which associates with each symmetric distribution μ on $\mathbb{R}^*$ its induced distribution on $(0,\infty)$ under the mapping $x \mapsto x^2$, the operators $\mathcal{S}$ and $\mathcal{Q}$ are one-to-one, and we have

$$\mu(\mathbf{V}_n) = \mathcal{S}\nu(\mathbf{F}_n) \quad \text{and} \quad \mu(\mathbf{W}_n) = \mathcal{Q}\mu(\mathbf{V}_n). \tag{2.2}$$

Furthermore, given a symmetric distribution $\mu_{\mathbf{V}}$ on $\mathbb{R}^*$ such that

$$\int \log^+ |t| \, d\mu_{\mathbf{V}}(t) < \infty, \tag{2.3}$$

we write $\mathcal{H}\mu_{\mathbf{V}}$ for the rotation-invariant distribution on $\mathbb{C}$ (if existent) such that

$$U_{\mathbf{V}}(\alpha) := -\int \log|z-\alpha| \, d(\mathcal{H}\mu_{\mathbf{V}})(z) = -\int \log|x| \, d\big(\mu_{\mathbf{V}} \boxplus B(\alpha)\big)(x) \tag{2.4}$$

for any $\alpha \in \mathbb{C}$. Here, the function $U_{\mathbf{V}}(\alpha)$ is the so-called *logarithmic potential* of the measure $\mathcal{H}\mu_{\mathbf{V}}$, $B(\alpha) := \frac{1}{2}\delta_{-|\alpha|} + \frac{1}{2}\delta_{+|\alpha|}$ denotes the Bernoulli distribution, and $\boxplus$ denotes free additive convolution. It follows from basic results in logarithmic potential theory that such a distribution $\mathcal{H}\mu_{\mathbf{V}}$, if it exists, is uniquely determined by (2.4), see e.g. [32], and also the comments at the end of Section 2.2.

Girko's *Hermitization method* (see also [12]) now states that under appropriate assumptions, the weak convergence of the eigenvalue distributions $\mu(\mathbf{F}_n)$ follows from the weak convergence of the singular value distributions $\nu(\mathbf{F}_n - \alpha\mathbf{I}_n)$ of the *shifted* matrices $\mathbf{F}_n - \alpha\mathbf{I}_n$ for all $\alpha \in \mathbb{C}$. We will only need the following special case:

THEOREM 2.1 ([18], Theorem 7.6). *If the random matrices $\mathbf{F}_n$ satisfy Condition* C *in Section* 4 *below and there exists a non-random probability measure $\nu_{\mathbf{F}}$ on $(0,\infty)$ such that for all $\alpha \in \mathbb{C}$, $\mathcal{S}\nu(\mathbf{F}_n - \alpha\mathbf{I}_n) \to (\mathcal{S}\nu_{\mathbf{F}}) \boxplus B(\alpha)$ weakly in*

probability, then $\mu(\mathbf{F}_n) \to \mu_{\mathbf{F}} := \mathcal{H}(\mathcal{S}\nu_{\mathbf{F}})$ *weakly in probability. Moreover, with the notation from* [18] *and under regularity conditions, the measure* $\mu_{\mathbf{F}}$ *has the Lebesgue density*

$$f(u,v) = \frac{1}{2\pi|\alpha|^2}\left(u\,\frac{\partial\psi}{\partial u} + v\,\frac{\partial\psi}{\partial v}\right), \tag{2.5}$$

where $\alpha = u + \mathrm{i}v$ *and* ψ *is a continuous function on* $\mathbb{C}^*$ *taking values in* $[0,1]$ *and satisfying*

$$\psi(\alpha)\big(1-\psi(\alpha)\big) = -|\alpha|^2\big(1-\psi(\alpha)\big)^2\Big(S_{\mathbf{V}}\Big(-\big(1-\psi(\alpha)\big)\Big)\Big)^2. \tag{2.6}$$

Here, $S_{\mathbf{V}}$ *denotes the* S*-transform of the symmetric probability measure* $\mu_{\mathbf{V}} = \mathcal{S}\nu_{\mathbf{F}}$ (see [30], [4], [18]).

Furthermore, as the starting point for the proof of Theorem 1.3 (which will be by induction on the number of factors and summands in (1.8)), we will rely upon the well-known *Marchenko–Pastur theorem*, which states that when $\mathbf{F}_n = \mathbf{X}_n^{(1)}$, $n \in \mathbb{N}$, then $\mu(\mathbf{W}_n) \to \gamma$ weakly in probability, where

$$\gamma(dx) = \frac{1}{2\pi}\sqrt{\frac{4-x}{x}}\,\mathbf{1}_{(0,4)}(x)\,\lambda(dx) \tag{2.7}$$

is the *Marchenko–Pastur distribution* (with parameter one). Therefore, when $\mathbf{F}_n = (\mathbf{X}_n^{(1)})^{-1}$, $n \in \mathbb{N}$, we have $\mu(\mathbf{W}_n) \to \gamma^{-1}$ weakly in probability, where γ^{-1} is the induced measure of γ under the mapping $x \mapsto x^{-1}$. We will call this measure the *inverse Marchenko–Pastur distribution*. Finally, let us note that the S-transforms of γ and γ^{-1} are given by

$$S_\gamma(z) = \frac{1}{z+1} \quad \text{and} \quad S_{\gamma^{-1}}(z) = -z, \tag{2.8}$$

respectively, see e.g. Section 8.1.1 in [18].

2.2. Results from free probability theory. To describe the limiting singular value distributions of the random matrices $\mathbf{F}_n$ in Theorem 1.3, we will use various concepts and results from free probability theory. See e.g. [38], [28] for a thorough introduction to free probability theory, or Section 5 in [18] for a brief introduction tailored to our purposes. In particular, we will use the free additive and multiplicative convolutions $\boxplus$ and $\boxtimes$, the associated R and S transforms (also for probability measures with *unbounded* support), and the asymptotic freeness of random matrices. Furthermore, we will frequently use the following result:

PROPOSITION 2.1 (Asymptotic freeness). *For each* $n \in \mathbb{N}$, *let* $\mathbf{A}_n$ *and* $\mathbf{B}_n$ *be independent bi-unitary invariant random matrices of size* $n \times n$ *such that*

$$\sup_{n\in\mathbb{N}} \max\left\{\mathbb{E}\big(\tfrac{1}{n}\operatorname{trace}(\mathbf{A}_n\mathbf{A}_n^*)^k\big), \mathbb{E}\big(\tfrac{1}{n}\operatorname{trace}(\mathbf{B}_n\mathbf{B}_n^*)^k\big)\right\} < \infty$$

for all $k \in \mathbb{N}$, and suppose that there exist compactly supported (deterministic) probability measures $\mu_{\mathbf{AA}^*}$ and $\mu_{\mathbf{BB}^*}$ on $(0,\infty)$ such that $\mu(\mathbf{A}_n\mathbf{A}_n^*) \to \mu_{\mathbf{AA}^*}$ and $\mu(\mathbf{B}_n\mathbf{B}_n^*) \to \mu_{\mathbf{BB}^*}$ weakly in probability. Then one has the following:

(a) *The families $\{\mathbf{A}_n, \mathbf{A}_n^*\}$ and $\{\mathbf{B}_n, \mathbf{B}_n^*\}$ are asymptotically free, and*

$$(\mathbf{A}_n\mathbf{B}_n)(\mathbf{A}_n\mathbf{B}_n)^* \to \mu_{\mathbf{AA}^*} \boxtimes \mu_{\mathbf{BB}^*} \text{ in moments.}$$

(b) *For any $k, l \in \mathbb{N}$, the matrices $(\mathbf{A}_n^k)^*\mathbf{A}_n^k$ and $\mathbf{A}_n^l(\mathbf{A}_n^l)^*$ are asymptotically free, and for any $k \in \mathbb{N}$,*

$$\mathbf{A}_n^k(\mathbf{A}_n^k)^* \to \mu_{\mathbf{AA}^*}^{\boxtimes k} \text{ in moments.}$$

(c) *The matrices $\mathbf{V}_n(\mathbf{A}_n)$ and $\mathbf{V}_n(\mathbf{B}_n)$ are asymptotically free, and*

$$\mathbf{V}_n(\mathbf{A}_n) + \mathbf{V}_n(\mathbf{B}_n) \to \mu_{\mathbf{V}(\mathbf{A})} \boxplus \mu_{\mathbf{V}(\mathbf{B})} \text{ in moments.}$$

(d) *The matrices $\mathbf{V}_n(\mathbf{A}_n)$ and $\mathbf{J}_n(\alpha)$ are asymptotically free, and*

$$\mathbf{V}_n(\mathbf{A}_n) + \mathbf{J}_n(\alpha) \to \mu_{\mathbf{V}(\mathbf{A})} \boxplus B(\alpha) \text{ in moments.}$$

Here, $\mathbf{V}(\mathbf{A}_n)$ and $\mathbf{V}(\mathbf{B}_n)$ are defined as in equation (2.1), *$\mu_{\mathbf{V}(\mathbf{A})}$ and $\mu_{\mathbf{V}(\mathbf{B})}$ denote the corresponding limiting distributions, and*

$$\mathbf{J}_n(\alpha) := \begin{bmatrix} \mathbf{O} & -\alpha\mathbf{I}_n \\ -\overline{\alpha}\mathbf{I}_n & \mathbf{O} \end{bmatrix}. \tag{2.9}$$

Parts (a) and (b) follow from the results in Section 4.3 in [23], part (d) is proved in Section 5 in [18], and part (c) follows by similar arguments. Also, let us mention that part (c) is already implicit in [37].

REMARK 2.1. *Observe that Proposition* 2.1 *may be used to establish the weak convergence of the* mean *singular value distributions of the matrices $\mathbf{A}_n\mathbf{B}_n$, $\mathbf{A}_n^k$ and $\mathbf{A}_n + \mathbf{B}_n$. However, in most of the situations in which we will use Proposition* 2.1 *later, this already implies the weak convergence in probability of the singular value distributions of these matrices* (see e.g. Section A.1 in [18]).

It is worth mentioning that there is another description of the limiting density $f(u,v)$ in Theorem 2.1 due to Haagerup and Larsen [21] and Haagerup and Schultz [22]. Actually, in these papers, the density f is shown to describe the *Brown measure* of a so-called *R-diagonal element* in a W^*-probability space. Roughly speaking, an R-diagonal element is a non-commutative random variable of the form uh, where u is Haar unitary and h is a positive element $*$-free from u.

For our purposes, this description of the density f may be summarized as follows. In the situation of Theorem 2.1, let $\nu_{\mathbf{F}}$ be the limiting singular value distribution of the matrices $\mathbf{F}_n$, set $\mu_{\mathbf{V}} = \mathcal{S}\nu_{\mathbf{F}}$ and $\mu_{\mathbf{W}} = \mathcal{Q}\nu_{\mathbf{V}}$ (which are the limiting

eigenvalue distributions of the matrices $\mathbf{V}_n$ and $\mathbf{W}_n$ in (2.1), respectively), and suppose that $\mu_{\mathbf{W}}$ is not a Dirac measure. Let $S_{\mathbf{W}}$ denote the S-transform of $\mu_{\mathbf{W}}$, and set

$$F(t) := \frac{1}{\sqrt{S_{\mathbf{W}}(t-1)}}.$$

Then F is a smooth bijection from the interval $(0,1)$ to the interval

$$(a,b) := \Big(\big(\int x^{-2}\, d\nu_{\mathbf{F}}(x) \big)^{-1/2}, \big(\int x^2\, d\nu_{\mathbf{F}}(x) \big)^{1/2} \Big)$$

(where $1/\infty := 0$ and $1/0 := \infty$), and the limiting eigenvalue distribution $\mu_{\mathbf{F}} = \mathcal{H}\mu_{\mathbf{V}}$ of the matrices $\mathbf{F}_n$ has a rotation-invariant density $f(r)$ given by

$$f(r) = \frac{1}{2\pi r\, F'\big(F^{-1}(r)\big)} \mathbf{1}_{(a,b)}(r) \tag{2.10}$$

(see [21], Section 4, and [22], Section 4). Clearly, the connection to Theorem 2.1 arises from the fact that $\psi = F^{-1}$ on the interval (a,b). Moreover, equation (2.10) shows that $F^{-1}(r) = \int_0^r 2\pi s\, f(s)\, ds$, which implies that $\mu_{\mathbf{W}}$, and hence $\mu_{\mathbf{V}}$, is uniquely determined by $\mu_{\mathbf{F}}$. Thus, the mapping $\mu_{\mathbf{V}} \mapsto \mu_{\mathbf{F}}$ is one-to-one.

Furthermore, it follows from the results in [21], [22] that the measure $\mathcal{H}\mu_{\mathbf{V}}$ exists for any symmetric probability measure $\mu_{\mathbf{V}}$ on $\mathbb{R}^*$ satisfying (2.3) and that the operator $\mathcal{H}$ thus defined furnishes a one-to-one correspondence between the set of these distributions on $\mathbb{R}^*$ and a certain set H of rotation-invariant distributions on $\mathbb{C}$. Finally, it is easy to see that for any symmetric distribution μ on $\mathbb{R}^*$ satisfying (2.3), we have

$$\mathcal{H}(\mathcal{D}_c\mu) = \mathcal{D}_c\mathcal{H}(\mu) \tag{2.11}$$

for all $c > 0$, where $\mathcal{D}_c$ is the scaling operator which maps a probability measure on $\mathbb{R}$ or $\mathbb{C}$ to its induced measure under the mapping $x \mapsto cx$.

2.3. Results on $\boxplus$-stable distributions. Let us collect some results on $\boxplus$-stable distributions which will be needed later. A distribution μ on $\mathbb{R}$ is called (*strictly*) $\boxplus$-*stable* if there exists a constant $\alpha > 0$ such that $\mu^{\boxplus m} = \mathcal{D}_{m^{1/\alpha}}\mu$ for all $m \in \mathbb{N}$. Here, $\mathcal{D}_c$ is defined as in equation (2.11). We will often call the constant α the *stability index* of μ.

The (strictly) $\boxplus$-stable distributions have been investigated in [9], [8] and [4]. First of all, let us recall that for any $\boxplus$-stable distribution, $\alpha \in]0,2]$. We will need the following result, which is contained in Appendix A of [8] and in [4]:

PROPOSITION 2.2. *Fix $\alpha \in]0,2]$. For a symmetric probability measure μ on $\mathbb{R}^*$, the following are equivalent*:

(i) *μ is (strictly) $\boxplus$-stable with stability index α.*

(ii) *$R_\mu(z) = bz^{\alpha-1}$, where $b \in \mathbb{C}^*$ with $\arg b = -\pi + \alpha\pi/2$.*

(iii) $S_\mu(z) = z^{(1/\alpha)-1}/b^{1/\alpha}$, *where* $b \in \mathbb{C}^*$ *with* $\arg b = -\pi + \alpha\pi/2$.

Moreover, in this case, the constants b *in parts* (ii) *and* (iii) *are the same.*

Here, for the S-transform $S_\mu(z)$, we make the convention that we take arguments in $]-\pi, +\pi]$ to define powers of b and arguments in $(-2\pi, 0)$ to define powers of z. Then, with i the imaginary unit, we have $S_\mu(z) \in (0,\infty)\mathrm{i}$ for $z \in (-1, 0)$, in line with the convention in [18].

Henceforward, we write $\sigma_s(\alpha)$ for the (unique) symmetric $\boxplus$-stable distribution with parameters $\alpha \in]0, 2]$ and $b := e^{(-\pi+\alpha\pi/2)\mathrm{i}}$. Note that in the special cases $\alpha = 2$ and $\alpha = 1$, we obtain the standard semicircle and Cauchy distribution, respectively. Furthermore, let us recall from [8], Appendix A, that the distribution $\sigma_s(\alpha)$ has a continuous density f_α such that $f_\alpha(x) = \mathcal{O}(|x|^{-\alpha-1})$ as $|x| \to \infty$. Thus, in particular, the distribution $\sigma_s(\alpha)$ satisfies condition (2.3).

3. PROOF OF THEOREMS 1.1 AND 1.2

In this section, we prove Theorems 1.1 and 1.2 using Theorem 1.3 and Remark 1.1, respectively.

Proof of Theorem 1.1. By Theorem 1.3, the limiting eigenvalue distributions of the matrices $\mathbf{F}_n := \mathbf{F}_n^{(0)}$ and $\widetilde{\mathbf{F}}_n := m^{-(l+1)/2}(\mathbf{F}_n^{(1)} + \ldots + \mathbf{F}_n^{(m)})$ in Theorem 1.1 are given by

$$\mu_{\mathbf{F}} = \mathcal{H}\Big(\mathcal{Q}^{-1}\big(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l}\big)\Big)$$

and

$$\mu_{\widetilde{\mathbf{F}}} = \mathcal{H}\Big(\mathcal{D}_{m^{-(l+1)/2}}\Big(\mathcal{Q}^{-1}\big(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l}\big)\Big)^{\boxplus m}\Big),$$

respectively, where $\mathcal{D}_c$ is defined as in equation (2.11). To obtain the description asserted in the theorem, we calculate the S-transform of $\mathcal{Q}^{-1}\big(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l}\big)$. Using (2.8) and the relation $S_{\nu_1 \boxtimes \nu_2}(z) = S_{\nu_1}(z) S_{\nu_2}(z)$, we find that

$$S_{\mathbf{W}}(z) = S_{\gamma \boxtimes (\gamma^{-1})^{\boxtimes l}}(z) = \frac{(-z)^l}{z+1},$$

and therefore

$$S_{\mathbf{V}}(z) = S_{\mathcal{Q}^{-1}(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l})}(z) = \sqrt{\frac{z+1}{z}\, S_{\mathbf{W}}(z)} = \sqrt{\frac{z+1}{z}\frac{(-z)^l}{z+1}} = \mathrm{i}^l z^{(l-1)/2}.$$

By Proposition 2.2, the corresponding distribution is $\mathcal{Q}^{-1}\big(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l}\big) = \sigma_s\big(\frac{2}{l+1}\big)$, the symmetric $\boxplus$-stable distribution of parameter $\frac{2}{l+1}$. Thus, $\mu_{\mathbf{F}} = \mathcal{H}\big(\sigma_s\big(\frac{2}{l+1}\big)\big)$. Also, using the defining property of a $\boxplus$-stable distribution, we get

$$\mathcal{D}_{m^{-(l+1)/2}}\Big(\Big(\mathcal{Q}^{-1}\big(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l}\big)\Big)^{\boxplus m}\Big) = \mathcal{D}_{m^{-(l+1)/2}}\Big(\Big(\sigma_s\big(\tfrac{2}{l+1}\big)\Big)^{\boxplus m}\Big) = \sigma_s\big(\tfrac{2}{l+1}\big).$$

Thus, $\mu_{\widetilde{\mathbf{F}}} = \mathcal{H}\big(\sigma_s\big(\frac{2}{l+1}\big)\big)$ as well, and the proof of Theorem 1.1 is complete. ∎

P r o o f o f T h e o r e m 1.2. By Theorem 1.3 and Remark 1.1, the limiting eigenvalue distributions of the matrices

$$\mathbf{F}_n := \mathbf{F}_n^{(0)} \quad \text{and} \quad \widetilde{\mathbf{F}}_n := m^{-(l+1)/2}(\mathbf{F}_n^{(1)} + \ldots + \mathbf{F}_n^{(m)})$$

in Theorem 1.2 are given by

$$\mu_{\mathbf{F}} = \mathcal{H}\Big(\mathcal{Q}^{-1}\big(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l_1} \boxtimes \ldots \boxtimes (\gamma^{-1})^{\boxtimes l_k}\big)\Big)$$

and

$$\mu_{\widetilde{\mathbf{F}}} = \mathcal{H}\Big(\mathcal{D}_{m^{-(l+1)/2}}\Big(\mathcal{Q}^{-1}\big(\gamma \boxtimes (\gamma^{-1})^{\boxtimes l_1} \boxtimes \ldots \boxtimes (\gamma^{-1})^{\boxtimes l_k}\big)\Big)^{\boxplus m}\Big),$$

respectively. But $(\gamma^{-1})^{\boxtimes l_1} \boxtimes \ldots \boxtimes (\gamma^{-1})^{\boxtimes l_k} = (\gamma^{-1})^{\boxtimes l}$, so the assertion follows in the same way as in the previous proof. ∎

REMARK 3.1. *In principle, the density of the limiting distribution $\mu_{\mathbf{F}}$ in Theorems* 1.1 *and* 1.2 *can be found by means of Theorem* 7.6 *in* [18]. *In our situation, it is easy to check that equation* (2.6) *reduces to*

$$\psi(\alpha)\big(1-\psi(\alpha)\big) = |\alpha|^2\big(1-\psi(\alpha)\big)^{l+1}.$$

(*Recall from Section* 2.3 *that $S_{\mathbf{V}}(z)$ takes values in $(0,\infty)\mathrm{i}$ when $z \in (-1,0)$.*) *Thus, since $\psi(\alpha)$ is continuous with values in $[0,1]$ and $\psi(\alpha) \neq 1$ for $\alpha \approx 0$* (see Sections 6 and 7 in [18]), *we obtain, for $l = 0,1,2,3$,*

$$\psi_0(r) = 1 \wedge r^2, \quad \psi_1(r) = \frac{r^2}{1+r^2}, \quad \psi_2(r) = 1 - \frac{2}{\sqrt{1+4r^2}+1},$$

$$\psi_3(r) = 1 - \frac{3}{\big(1+v^2(r)+w^2(r)\big)^2},$$

and therefore

$$f_0(r) = \frac{1}{\pi}\mathbf{1}_{(0,1)}(r), \quad f_1(r) = \frac{1}{\pi(1+r^2)^2},$$

$$f_2(r) = \frac{2}{\pi\sqrt{1+4r^2}(1+2r^2+\sqrt{1+4r^2})},$$

$$f_3(r) = \frac{27\big(v(r)+w(r)\big)}{\pi\sqrt{4+27r^2}\big(1+v^2(r)+w^2(r)\big)^3},$$

where we have set

$$v(r) := \big(\tfrac{1}{2}\sqrt{4+27r^2} + \tfrac{1}{2}\sqrt{27}r\big)^{1/3} \quad \textit{and} \quad w(r) := \big(\tfrac{1}{2}\sqrt{4+27r^2} - \tfrac{1}{2}\sqrt{27}r\big)^{1/3}$$

for abbreviation. ∎

REMARK 3.2. *It seems natural to ask whether there exist further examples of random matrices* $\mathbf{F}_n^{(0)}$ *such that for any* $m \in \mathbb{N}$, $\mathbf{F}_n^{(0)}$ *and* $\mathbf{F}_n^{(1)} + \ldots + \mathbf{F}_n^{(m)}$ *have the same limiting eigenvalue distributions after appropriate rescaling. However, it turns out that within the class of products of independent Girko–Ginibre matrices and their inverses, there exist no further examples beyond those mentioned in Theorem* 1.1*, apart from possible permutations of the exponents* ± 1*. Indeed, suppose that* $\mathbf{F}_n^{(0)}$ *is a product of* p *factors* $\mathbf{Y}_n^{(r)}$ *and* q *factors* $(\mathbf{Y}_n^{(r)})^{-1}$ *(all of them independent, and in arbitrary order), and let* $\mathbf{V}_n$ *and* $\mathbf{W}_n$ *be defined as in* (2.1)*. Then, arguing as in the proof of Theorem* 1.1*, we find that the corresponding* S*-transforms* $S_{\mathbf{W}}$ *and* $S_{\mathbf{V}}$ *are given by*

$$S_{\mathbf{W}}(z) = \frac{(-z)^q}{(1+z)^p} \quad \text{and} \quad S_{\mathbf{V}}(z) = \frac{\mathrm{i}^q\, z^{(q-1)/2}}{(1+z)^{(p-1)/2}},$$

respectively, and by Proposition 2.2*, the latter is the* S*-transform of a symmetric* $\boxplus$*-stable distribution if and only if* $p = 1$ *and* $q \in \mathbb{N}_0$*. Now use the observation that, by equation* (2.11)*, if* $\mu^{\boxplus m}$ *is not a rescaled version of* μ*, then* $\mathcal{H}(\mu^{\boxplus m})$ *is not a rescaled version of* $\mathcal{H}(\mu)$*.* ∎

REMARK 3.3. *The limiting eigenvalue distribution in Theorems* 1.1 *and* 1.2 *may be interpreted as a* stable distribution *with respect to an appropriately defined convolution* $\oplus$*. To define this convolution, suppose that* μ_1 *and* μ_2 *are two probability measures which belong to the class* H *introduced above equation* (2.11) *and that* $\mathbf{A}_n$ *and* $\mathbf{B}_n$ *are independent bi-unitary invariant random matrices with limiting eigenvalue distributions* μ_1 *and* μ_2*, respectively. Also, suppose that these matrices satisfy the assumptions of Theorem* 2.1*. Then, if* $\widetilde{\nu}_1$ *and* $\widetilde{\nu}_2$ *are the limiting symmetrized singular value distributions of* $\mathbf{A}_n$ *and* $\mathbf{B}_n$*, we have* $\mu_1 = \mathcal{H}(\widetilde{\nu}_1)$ *and* $\mu_2 = \mathcal{H}(\widetilde{\nu}_2)$ *by Theorem* 2.1*. Furthermore, suppose that the matrix sums* $\mathbf{A}_n + \mathbf{B}_n$ *have the limiting symmetrized singular value distribution* $\widetilde{\nu}_1 \boxplus \widetilde{\nu}_2$ *(which seems natural in view of Proposition* 2.1*) and that they also satisfy the assumptions of Theorem* 2.1*. Then, again by Theorem* 2.1*, the associated limiting eigenvalue distribution is given by* $\mathcal{H}(\widetilde{\nu}_1 \boxplus \widetilde{\nu}_2)$*. This motivates the following definition*:

DEFINITION 3.1. Given two probability measures μ_1 and μ_2 of class H, set $\mu_1 \oplus \mu_2 := \mathcal{H}\big(\mathcal{H}^{-1}(\mu_1) \boxplus \mathcal{H}^{-1}(\mu_2)\big)$.

This convolution $\oplus$ *may also be interpreted in terms of free probability. Indeed, given* μ_1 *and* μ_2 *in* H*, pick* R*-diagonal elements* x_1 *and* x_2 *(in some* W^**-probability space) such that the Brown measure of* x_1 *is* μ_1*, the Brown measure of* x_2 *is* μ_2*, and* x_1 *and* x_2 *are* $*$*-free. Then* $\mu_1 \oplus \mu_2$ *is the Brown measure of* $x_1 + x_2$*, as follows from the results in* [21] *and* [22].

It is now natural to introduce the concept of a (strictly) $\oplus$*-stable distribution: A probability measure* μ *of class* H *is called* (strictly) $\oplus$-stable *if there exists a constant* $\alpha > 0$ *such that* $\mu^{\oplus m} = \mathcal{D}_{m^{1/\alpha}}\mu$ *for all* $m \in \mathbb{N}$*. Then, by equation* (2.11)*,* $\widetilde{\nu}$ *is* $\boxplus$*-stable if and only if* $\mathcal{H}(\widetilde{\nu})$ *is* $\oplus$*-stable. Therefore, the* $\oplus$*-stable distributions in*

H are in one-to-one correspondence with the symmetric $\boxplus$-stable distributions on $\mathbb{R}^*$, and the limiting spectral distributions occurring in Theorems 1.1 *and* 1.2 *are special examples of this type.* ∎

4. PROOF OF THEOREM 1.3

4.1. Overview. In this section we prove Theorem 1.3 using the general framework from [18]. In Subsection 4.2, we summarize the technical conditions and the main universality results from [18] to make the presentation reasonably self-contained. Subsections 4.3–4.5 prepare for the proof of Theorem 1.3 by verifying the technical conditions from [18]. Subsection 4.6 contains the core of the proof of Theorem 1.3, and Subsection 4.7 describes the necessary modifications for Remark 1.1. Some auxiliary results from the literature are collected in Subsection 4.8.

4.2. General framework. A major step in [18] is to prove the *universality* of the limiting singular value and eigenvalue distributions, i.e. to show that these distributions (if existent) do not depend on the distributions of the matrix entries, apart from a few moment conditions as in (1.2)–(1.4). To state this more precisely, we need two sets of random matrices.

To this end, it seems convenient to regard $\mathbf{F}_n$ in (1.8) as a *matrix function* (by slight abuse of notation) and to write

$$\mathbf{F}_n(\mathbf{Z}_n^{(1)},\dots,\mathbf{Z}_n^{(ml)}) := \sum_{q=1}^{m}\prod_{r=1}^{l}(\mathbf{Z}_n^{((q-1)l+r)})^{\varepsilon_r}, \tag{4.1}$$

where $m,l\in\mathbb{N}$ and $\varepsilon_1,\dots,\varepsilon_l\in\{-1,+1\}$ are the same as in (1.8), and $\mathbf{Z}_n^{(q)} = (Z_{jk}^{(q)})_{j,k=1,\dots,n}$ is a matrix in the indeterminates $Z_{jk}^{(q)}$, $q=1,\dots,ml$. Then, we may write $\mathbf{F}_n(\mathbf{X}) := \mathbf{F}_n(\mathbf{X}_n^{(1)},\dots,\mathbf{X}_n^{(ml)})$ for the random matrices built from the random matrices $\mathbf{X}_n^{(q)} := \big(\frac{1}{\sqrt{n}}X_{jk}^{(q)}\big)_{j,k=1,\dots,n}$, and $\mathbf{F}_n(\mathbf{Y}) := \mathbf{F}_n(\mathbf{Y}_n^{(1)},\dots,\mathbf{Y}_n^{(ml)})$ for the corresponding random matrices built from the *Gaussian* random matrices $\mathbf{Y}_n^{(q)} := \big(\frac{1}{\sqrt{n}}Y_{jk}^{(q)}\big)_{j,k=1,\dots,n}$. We always assume that the families $(X_{jk}^{(q)})_{j,k,q\in\mathbb{N}}$ and $(Y_{jk}^{(q)})_{j,k,q\in\mathbb{N}}$ are defined on the same probability space and independent. When the choice of the matrices $\mathbf{X}_n^{(1)},\dots,\mathbf{X}_n^{(ml)}$ is clear from the context, we also write $\mathbf{F}_n$ instead of $\mathbf{F}_n(\mathbf{X})$.

REMARK 4.1. *More generally, using the arguments from this section, we might deal with matrix functions of the form*

$$\mathbf{F}_n(\mathbf{Z}_n^{(1)},\mathbf{Z}_2^{(n)},\mathbf{Z}_3^{(n)},\dots) := \sum_{q=1}^{m}\prod_{r=1}^{l_q}(\mathbf{Z}_n^{(i_{q,r})})^{\varepsilon_{q,r}},$$

where $m,l_1,\dots,l_m\in\mathbb{N}$, $\varepsilon_{q,r}\in\{+1,-1\}$, the indices $i_{q,r}\in\mathbb{N}$ are pairwise diffe-

rent, and all parameters do not depend on n. That is to say, the numbers and the types of the factors in the m summands need not be the same.

In our investigation of the limiting spectral distributions of the matrices $\mathbf{F}_n$, we will also consider the *shifted matrices* $\mathbf{F}_n - \alpha\mathbf{I}_n$, with $\alpha \in \mathbb{C}$, the *regularized matrices* $\mathbf{F}_{n,t}$, with $t > 0$, and their combinations $\mathbf{F}_{n,t} - \alpha\mathbf{I}_n$. Here, the regularized matrices $\mathbf{F}_{n,t}$ arise from the regularized matrix functions

$$\mathbf{F}_{n,t}(\mathbf{Z}_n^{(1)}, \dots, \mathbf{Z}_n^{(ml)}) := \sum_{q=1}^{m} \prod_{r=1}^{l} (\mathbf{Z}_n^{((q-1)l+r)})_t^{\varepsilon_r}, \tag{4.2}$$

where $(\mathbf{Z}_n)_t^{\varepsilon} := \mathbf{Z}_n$ for $\varepsilon = +1$ and $(\mathbf{Z}_n)_t^{\varepsilon} := (\mathbf{Z}_n^* \mathbf{Z}_n + t\mathbf{I}_n)^{-1}\mathbf{Z}_n^*$ for $\varepsilon = -1$. Note that, by definition, the regularization has no effect when $\varepsilon = +1$ and that $\lim_{t\downarrow 0}(\mathbf{Z}_n)_t^{-1} = (\mathbf{Z}_n)^{-1}$ when $\mathbf{Z}_n$ is invertible.

Furthermore, fix a sequence $(\tau_n)_{n\in\mathbb{N}}$ of positive real numbers such that $\tau_n \to 0$ and $\tau_n\sqrt{n} \to \infty$, and for $0 \leqslant \varphi \leqslant \pi/2$, set

$$Z_{jk}^{(q)}(\varphi) := (\cos\varphi)\, X_{jk}^{(q)} \mathbf{1}_{\{|X_{jk}^{(q)}| \leqslant \tau_n\sqrt{n}\}} + (\sin\varphi)\, Y_{jk}^{(q)} \mathbf{1}_{\{|Y_{jk}^{(q)}| \leqslant \tau_n\sqrt{n}\}} \quad (j,k,q \in \mathbb{N}),$$

$$\mathbf{Z}_n^{(q)}(\varphi) := \big(\tfrac{1}{\sqrt{n}} Z_{jk}^{(q)}(\varphi)\big)_{j,k=1,\dots,n} \ (q \in \mathbb{N}),\ \mathbf{F}_n(\varphi) := \mathbf{F}_n\big(\mathbf{Z}_n^{(1)}(\varphi), \dots, \mathbf{Z}_n^{(ml)}(\varphi)\big).$$

Note that the matrices $\mathbf{Z}_n^{(q)}(\varphi)$ provide an interpolation between truncated versions of the matrices $\mathbf{X}_n^{(q)}$ (for $\varphi = 0$) and $\mathbf{Y}_n^{(q)}$ (for $\varphi = \pi/2$).

With this notation, we have to check the following Conditions A, B and C.

CONDITION A. For $\mathbf{F}_n = \mathbf{F}_n(\mathbf{X})$ and $\mathbf{F}_n = \mathbf{F}_n(\mathbf{Y})$, the matrices $\mathbf{F}_n$ satisfy the following condition:

For each $\alpha \in \mathbb{C}$ and $z \in \mathbb{C}^+$, we have $\lim_{t\to 0} \limsup_{n\to\infty} |s_{n,t}(z) - s_n(z)| = 0$ in probability, where $s_n(z)$ and $s_{n,t}(z)$ are the Stieltjes transforms of the Hermitian matrices $(\mathbf{F}_n - \alpha\mathbf{I}_n)(\mathbf{F}_n - \alpha\mathbf{I}_n)^*$ and $(\mathbf{F}_{n,t} - \alpha\mathbf{I}_n)(\mathbf{F}_{n,t} - \alpha\mathbf{I}_n)^*$, respectively.

CONDITION B. For each $t > 0$, $\alpha \in \mathbb{C}$, $z \in \mathbb{C}^+$, we have

$$\sup_{\varphi\in[0;\pi/2]} \sup_{q,j,k} \max_{D} \big\| \mathbb{E}\big\{ D g_{j,k}^{(q)}(\varphi) \big| X_{jk}^{(q)}, Y_{jk}^{(q)} \big\} \big\|_{\infty} \leqslant A < \infty,$$

where the maximum is over all partial derivatives D of orders zero, one and two in the matrix entries $\operatorname{Re} Z_{jk}^{(q)}$ and $\operatorname{Im} Z_{jk}^{(q)}$, $g_{j,k}^{(q)}(\varphi)$ may be either

$$g_{j,k}^{(q)}(\varphi) = \left(\frac{\partial}{\partial \operatorname{Re} Z_{jk}^{(q)}} \operatorname{trace}\big(\mathbf{V}_{n,t}(\alpha;\varphi) - z\mathbf{I}_{2n}\big)^{-1} \right) \Big|_{Z_{jk}^{(q)} \to \theta Z_{jk}^{(q)}}$$

or

$$g_{j,k}^{(q)}(\varphi) = \left(\frac{\partial}{\partial \operatorname{Im} Z_{jk}^{(q)}} \operatorname{trace}\big(\mathbf{V}_{n,t}(\alpha;\varphi) - z\mathbf{I}_{2n}\big)^{-1} \right) \Big|_{Z_{jk}^{(q)} \to \theta Z_{jk}^{(q)}},$$

the $\mathbf{V}_{n,t}(\alpha;\varphi)$ are the Hermitizations of the matrices $\mathbf{F}_n(\varphi)$, but with $\mathbf{F}_n$ replaced by $\mathbf{F}_{n,t} - \alpha\mathbf{I}_n$, and θ (the rescaling parameter in the substitution $Z_{j,k}^{(q)} \to \theta Z_{j,k}^{(q)}$) is a random variable which is uniformly distributed on $[0,1]$ and independent of everything else.

CONDITION C. For $\mathbf{F}_n = \mathbf{F}_n(\mathbf{X})$ and $\mathbf{F}_n = \mathbf{F}_n(\mathbf{Y})$, the matrices $\mathbf{F}_n$ satisfy the following conditions:

(C0) There exists some $p > 0$ such that $\frac{1}{n}\sum_{k=1}^{n} s_k^p(\mathbf{F}_n)$ is bounded in probability as $n \to \infty$.

(C1) For any fixed $\alpha \in \mathbb{C}$, there exists some $Q > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}\big(s_n(\mathbf{F}_n - \alpha\mathbf{I}_n) \leqslant n^{-Q}\big) = 0.$$

(C2) For any fixed $\alpha \in \mathbb{C}$, there exists some $0 < \gamma < 1$ such that for *any* sequence $(\delta_n)_{n\in\mathbb{N}}$ with $\delta_n \to 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(\tfrac{1}{n} \sum_{n_1 \leqslant j \leqslant n_2} |\log s_j(\mathbf{F}_n - \alpha\mathbf{I}_n)| > \varepsilon\big) = 0 \quad \text{for all } \varepsilon > 0,$$

where $n_1 = [n - n\delta_n] + 1$ and $n_2 = [n - n^\gamma]$.

REMARK 4.2 (Condition $\mathrm{C}_{\mathrm{simple}}$). *It will be convenient to consider Condition* C *for more general random matrices* $\mathbf{F}_n$ *(with* $\mathbf{F}_n$ *of dimension* $n \times n$*) than in* (1.8)*. If a sequence of random matrices* $\mathbf{F}_n$ *satisfies Conditions* (C0), (C1) *and* (C2)*, we say that the matrices* $\mathbf{F}_n$ *satisfy* Condition C*. Also, if a sequence of random matrices* $\mathbf{F}_n$ *satisfies Condition* (C0) *as well as Conditions* (C1) *and* (C2) *with* $\alpha = 0$*, we say that the matrices* $\mathbf{F}_n$ *satisfy* Condition $\mathrm{C}_{\mathrm{simple}}$.

The following universality result is implicitly contained in [18]:

THEOREM 4.1 (Universality of singular value and eigenvalue distributions). *Let* $\mathbf{F}_n(\mathbf{X})$*,* $\mathbf{F}_n(\mathbf{Y})$ *be defined as above, and let* $\nu_n(\mathbf{X})$, $\nu_n(\mathbf{Y})$ *and* $\mu_n(\mathbf{X})$, $\mu_n(\mathbf{Y})$ *denote the associated singular value and eigenvalue distributions, respectively.*

(a) *If Conditions* A *and* B *hold,* $\nu_n(\mathbf{X}) - \nu_n(\mathbf{Y}) \to 0$ *weakly in probability.*

(b) *If Conditions* A, B *and* C *hold,* $\mu_n(\mathbf{X}) - \mu_n(\mathbf{Y}) \to 0$ *weakly in probability.*

Proof. (a) Set $\alpha := 0$. For $\mathbf{Z} = \mathbf{X}$ and $\mathbf{Z} = \mathbf{Y}$, let $m_n(z;\mathbf{Z})$ and $s_n(z;\mathbf{Z})$ denote the Stieltjes transforms of the Hermitian matrices

$$\mathbf{V}_n(\mathbf{Z}) := \begin{bmatrix} \mathbf{O} & \mathbf{F}_n(\mathbf{Z}) \\ \mathbf{F}_n^*(\mathbf{Z}) & \mathbf{O} \end{bmatrix} \quad \text{and} \quad \mathbf{W}_n(\mathbf{Z}) := \mathbf{F}_n(\mathbf{Z})\mathbf{F}_n^*(\mathbf{Z}),$$

and let $m_{n,t}(z;\mathbf{Z})$ and $s_{n,t}(z;\mathbf{Z})$ denote the corresponding Stieltjes transforms when $\mathbf{F}_n(\mathbf{Z})$ is replaced with $\mathbf{F}_{n,t}(\mathbf{Z})$. Fix $t > 0$. By Condition B and Theorem 3.2 in [18], we have, for each $z \in \mathbb{C}^+$, $m_{n,t}(z;\mathbf{X}) - m_{n,t}(z;\mathbf{Y}) \to 0$ in probability,

and therefore $s_{n,t}(z;\mathbf{X}) - s_{n,t}(z;\mathbf{Y}) \to 0$ in probability. (Note that the Lindeberg condition in [18] holds by our assumption (1.4), while the rank condition in [18] follows from basic inequalities for the rank of matrix sums and matrix products.) It therefore follows from Condition A that, for each $z \in \mathbb{C}^+$, $s_n(z;\mathbf{X}) - s_n(z;\mathbf{Y}) \to 0$ in probability, which implies the claim.

(b) By the same argument as in (a), the conclusion of (a) holds not only for the singular value distributions of the matrices $\mathbf{F}_n$, but also for the singular value distributions of the shifted matrices $\mathbf{F}_n - \alpha\mathbf{I}_n$ for any fixed $\alpha \in \mathbb{C}$. Thus, the claim follows from Condition C and Remark 4.2 in [18]. ∎

REMARK 4.3. *As follows from the proof, if one is only interested in the limiting singular value distributions of the matrices* $\mathbf{F}_n$*, it suffices to assume that Conditions* A *and* B *hold with* $\alpha = 0$*.*

We will use Theorem 4.1 to establish Theorem 1.3. This requires verifying Conditions A, B and C, of course. For this purpose, we provide some auxiliary results in the next three subsections.

4.3. On Condition A. Let $\mathbf{F}_n = \mathbf{F}_n(\mathbf{X})$ be defined as in (1.8). To obtain a matrix function which is smooth in the matrix entries (as needed for Condition B), we replace all inverses $(\mathbf{X}_n^{(q)})^{-1}$ with regularized inverses $(\mathbf{X}_n^{(q)})_t^{-1}$. We do this in a step-by-step fashion. Hence, fix $t > 0$, fix an index Q such that $\varepsilon_Q = -1$, and for all the other indices q with $\varepsilon_q = -1$, fix a choice between $(\mathbf{X}_n^{(q)})^{-1}$ and $(\mathbf{X}_n^{(q)})_t^{-1}$. Then it suffices to consider random matrices of the form

$$\mathbf{F}_n = \mathbf{A}_n(\mathbf{X}_n)^{-1}\mathbf{B}_n + \mathbf{C}_n, \tag{4.3}$$

where $\mathbf{X}_n \equiv \mathbf{X}_n^{(Q)}$ (we omit the index Q for simplicity) and $\mathbf{A}_n$, $\mathbf{B}_n$ and $\mathbf{C}_n$ depend only on the matrices $\mathbf{X}_n^{(q)}$ with $q \neq Q$.

Fix $\alpha \in \mathbb{C}$, and for $0 \leqslant u \leqslant t$, let

$$\mathbf{F}_{n,u} := \mathbf{A}_n(\mathbf{X}_n)_u^{-1}\mathbf{B}_n + \mathbf{C}_n := \mathbf{A}_n(\mathbf{X}_n^*\mathbf{X}_n + u)^{-1}\mathbf{X}_n^*\mathbf{B}_n + \mathbf{C}_n \tag{4.4}$$

and

$$s_{n,u}(z) := \tfrac{1}{n}\operatorname{trace}\big((\mathbf{F}_{n,u} - \alpha\mathbf{I}_n)(\mathbf{F}_{n,u} - \alpha\mathbf{I}_n)^* - z\mathbf{I}_n\big)^{-1}. \tag{4.5}$$

Note that $\mathbf{F}_{n,0}$ coincides with $\mathbf{F}_n$ if $\mathbf{X}_n$ is invertible. Then, by way of induction, it will suffice to prove the following lemma:

LEMMA 4.1. *For each* $n \in \mathbb{N}$*, let* $\mathbf{X}_n = \big(\frac{1}{\sqrt{n}}X_{jk}\big)_{j,k=1,\dots,n}$ *be as in* (1.1)–(1.4)*. Furthermore, for each* $n \in \mathbb{N}$*, let* $\mathbf{A}_n$*,* $\mathbf{B}_n$ *and* $\mathbf{C}_n$ *be random matrices of dimension* $n \times n$ *such that the singular value distributions of the random matrices* $\mathbf{B}_n$ *and* $\mathbf{C}_n$ *converge weakly in probability to* (*non-random*) *probability measures on*

$(0,\infty)$ *and* $[0,\infty)$, *respectively*, *and let* $\mathbf{F}_{n,u}$ *and* $s_{n,u}(z)$ *be defined as in* (4.4) *and* (4.5). *Then*, *for any* $z = u + \mathrm{i}v \in \mathbb{C}^+$, *we have*

$$\lim_{t\to 0} \limsup_{n\to\infty} |s_{n,t}(z) - s_{n,0}(z)| = 0 \text{ in probability}. \tag{4.6}$$

REMARK 4.4. *Let us emphasize that although the matrices* $\mathbf{A}_n$, $\mathbf{B}_n$, $\mathbf{C}_n$ *and* $\mathbf{X}_n$ *in the decomposition* (4.3) *are independent*, *this is not required in Lemma* 4.1.

REMARK 4.5. *Lemma* 8.16 *in* [18] *contains a similar result for the case* $\mathbf{C}_n = 0$, *although under additional assumptions and with a proof which does not seem to extend to the case* $\mathbf{C}_n \neq 0$. *The main difference in the proof of Lemma* 4.1 (*as compared to that of Lemma* 8.16 *in* [18]) *is that we control the* auxiliary *modifications of the matrices* $\mathbf{B}_n$ *and* $\mathbf{C}_n$ *via the matrix rank*, *and not via the resolvent.*

REMARK 4.6. *Let us illustrate the way Lemma* 4.1 *will be used later. Consider an* l*-fold product* $\mathbf{F}_n(\mathbf{X}) = (\mathbf{X}_n^{(1)})^{\varepsilon_1} \dots (\mathbf{X}_n^{(l)})^{\varepsilon_l}$, *where* $\varepsilon_1, \dots, \varepsilon_l \in \{-1, +1\}$, *and suppose by way of induction that we have weak convergence for any matrix product with less than* l *factors*, *possibly regularized. Then*, *setting* $\mathbf{F}_{n,t_1,\dots,t_l} := (\mathbf{X}_n^{(1)})_{t_1}^{\varepsilon_1} \dots (\mathbf{X}_n^{(l)})_{t_l}^{\varepsilon_l}$ *and*

$$s_n(t_1, \dots, t_l; z) := \tfrac{1}{n} \operatorname{trace}\big((\mathbf{F}_{n,t_1,\dots,t_l} - \alpha\mathbf{I}_n)(\mathbf{F}_{n,t_1,\dots,t_l} - \alpha\mathbf{I}_n)^* - z\mathbf{I}_n\big)^{-1}$$

and writing $\mathbf{t}_k := (t, \dots, t, 0, \dots, 0)$ *for the vector consisting of* k t*'s and* $l - k$ 0*'s*, *we have the estimate*

$$|s_{n,t}(z) - s_{n,0}(z)| \leqslant \sum_{k=1}^{l} |s_n(\mathbf{t}_k; z) - s_n(\mathbf{t}_{k-1}; z)|. \tag{4.7}$$

Now, *for each* $k = 1, \dots, l$, *the* k*th summand on the right-hand side in* (4.7) *satisfies* (4.6), *either trivially* (*when* $\varepsilon_k = +1$) *or by Lemma* 4.1 (*when* $\varepsilon_k = -1$). *Thus*, *the left-hand side in* (4.7) *satisfies* (4.6) *as well*, *and Condition* A *is proved for the* l*-fold product* $\mathbf{F}_n(\mathbf{X})$.

Proof of Lemma 4.1. For the sake of simplicity, we consider only the case $\alpha = 0$ here, the extension to the case $\alpha \neq 0$ being straightforward. We have to show that for any given $\varepsilon > 0$ and $\delta > 0$,

$$\limsup_{t\to 0} \limsup_{n\to\infty} \mathbb{P}\big(|s_{n,t}(z) - s_{n,0}(z)| > \varepsilon\big) < \delta. \tag{4.8}$$

Hence, fix $\varepsilon > 0$ and $\delta > 0$. As in the proof of Lemma 8.16 in [18], we introduce *auxiliary* modifications of the matrices $\mathbf{B}_n$ and $\mathbf{C}_n$ before we regularize the inverse matrices $\mathbf{X}_n^{-1}$.

For an $n \times n$ matrix $\mathbf{M}$, let $s_1(\mathbf{M}) \geqslant \dots \geqslant s_n(\mathbf{M})$ denote the singular values. Since the singular value distributions of $\mathbf{B}_n$ and $\mathbf{C}_n$ converge weakly in probability

to (non-random) probability measures on $(0,\infty)$ and $[0,\infty)$, respectively, we may find $K > 1$ and $N \in \mathbb{N}$ such that for $n \geqslant N$, we have

$$\mathbb{P}\bigg(\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{s_k(\mathbf{B}_n)<K^{-1} \text{ or } s_k(\mathbf{B}_n)>K\}} > \frac{\varepsilon v}{24} \text{ or } \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{s_k(\mathbf{C}_n)>K\}} > \frac{\varepsilon v}{24}\bigg) < \frac{\delta}{2}.$$

Then, the modifications $\widetilde{\mathbf{B}}_n$ and $\widetilde{\mathbf{C}}_n$ are defined as follows. For the matrix $\mathbf{C}_n$, take the singular value decomposition $\mathbf{C}_n = \mathbf{U}\mathbf{\Delta}\mathbf{V}^*$, let $\widetilde{\mathbf{\Delta}}$ be the diagonal matrix obtained from $\mathbf{\Delta}$ by replacing the diagonal elements $\mathbf{\Delta}_{kk}$ with $\widetilde{\mathbf{\Delta}}_{kk} := \mathbf{\Delta}_{kk} \wedge K$, and set $\widetilde{\mathbf{C}}_n := \mathbf{U}\widetilde{\mathbf{\Delta}}\mathbf{V}^*$. For the matrix $\mathbf{B}_n$, take the singular value decomposition $\mathbf{B}_n = \mathbf{U}\mathbf{\Delta}\mathbf{V}^*$, let $\widetilde{\mathbf{\Delta}}$ be the diagonal matrix obtained from $\mathbf{\Delta}$ by replacing the diagonal elements $\mathbf{\Delta}_{kk}$ with $\widetilde{\mathbf{\Delta}}_{kk} := (\mathbf{\Delta}_{kk} \wedge K) \vee K^{-1}$, and set $\widetilde{\mathbf{B}}_n := \mathbf{U}\widetilde{\mathbf{\Delta}}\mathbf{V}^*$. Then we have

$$\|\widetilde{\mathbf{B}}_n\| \leqslant K, \quad \|\widetilde{\mathbf{B}}_n^{-1}\| \leqslant K, \quad \|\widetilde{\mathbf{C}}_n\| \leqslant K, \tag{4.9}$$

and for $n \geqslant N$, with a probability of at least $1 - \delta/2$, we also have

$$\tfrac{1}{n}\operatorname{rank}(\mathbf{B}_n - \widetilde{\mathbf{B}}_n) \leqslant \varepsilon v/24, \quad \tfrac{1}{n}\operatorname{rank}(\mathbf{C}_n - \widetilde{\mathbf{C}}_n) \leqslant \varepsilon v/24. \tag{4.10}$$

Furthermore, let $\widetilde{\mathbf{F}}_{n,u}$ and $\widetilde{s}_{n,u}(z)$ be defined as in (4.4) and (4.5), but with $\mathbf{B}_n$ and $\mathbf{C}_n$ replaced by $\widetilde{\mathbf{B}}_n$ and $\widetilde{\mathbf{C}}_n$. It then follows from (4.10) that for $n \geqslant N$, with a probability of at least $1 - \delta/2$, we have

$$\tfrac{1}{n}\operatorname{rank}(\mathbf{F}_{n,u}\mathbf{F}_{n,u}^* - \widetilde{\mathbf{F}}_{n,u}\widetilde{\mathbf{F}}_{n,u}^*) \leqslant \varepsilon v/6,$$

and therefore, by the rank inequality (compare e.g. [5], Lemma 6.9),

$$|s_{n,u}(z) - \widetilde{s}_{n,u}(z)| \leqslant \varepsilon/3.$$

Thus, we have reduced the proof of (4.8) to showing that

$$\lim_{t\to 0} \limsup_{n\to\infty} |\widetilde{s}_{n,t}(z) - \widetilde{s}_{n,0}(z)| = 0 \text{ in probability.} \tag{4.11}$$

Since we only deal with the modified matrices for the rest of the proof, we omit the tildes and write $\mathbf{B}_n, \mathbf{C}_n, \mathbf{F}_{n,u}$ and $s_{n,u}(z)$ instead of $\widetilde{\mathbf{B}}_n, \widetilde{\mathbf{C}}_n, \widetilde{\mathbf{F}}_{n,u}$ and $\widetilde{s}_{n,u}(z)$, respectively. Moreover, for brevity, we usually omit the index n.

To establish (4.11), we may proceed as in the proof of Lemma 8.16 in [18]. Setting $\mathbf{R}_u := (\mathbf{F}_u\mathbf{F}_u^* - z\mathbf{I})^{-1}$, $0 \leqslant u \leqslant t$, we have the estimates

$$\begin{gathered} \|\mathbf{R}_u\| \leqslant v^{-1}, \quad \|\mathbf{F}_u^*\mathbf{R}_u\mathbf{F}_u\| \leqslant 1 + |z|v^{-1}, \\ \|\mathbf{R}_u\mathbf{F}_u\| \leqslant \big(v^{-1}(1+|z|v^{-1})\big)^{1/2}, \quad \|\mathbf{F}_u^*\mathbf{R}_u\| \leqslant \big(v^{-1}(1+|z|v^{-1})\big)^{1/2} \end{gathered} \tag{4.12}$$

as well as the representation

$$\mathbf{R}_t - \mathbf{R}_0 = \int_0^t \frac{d\mathbf{R}_u}{du}\,du = -\int_0^t \mathbf{R}_u \frac{d(\mathbf{F}_u\mathbf{F}_u^*)}{du}\mathbf{R}_u\,du. \tag{4.13}$$

Thus, it is straightforward to check that

$$\begin{aligned}\left|\tfrac{1}{n}\operatorname{trace}(\mathbf{R}_t - \mathbf{R}_0)\right| &\leqslant \int_0^t \left|\tfrac{1}{n}\operatorname{trace}\left(\mathbf{R}_u\mathbf{F}_u\mathbf{B}^{-1}(\mathbf{X}\mathbf{X}^*+u\mathbf{I})^{-1}\mathbf{B}\mathbf{F}_u^*\mathbf{R}_u\right)\right|du\\ &\quad+\int_0^t \left|\tfrac{1}{n}\operatorname{trace}\left(\mathbf{R}_u\mathbf{C}\mathbf{B}^{-1}(\mathbf{X}\mathbf{X}^*+u\mathbf{I})^{-1}\mathbf{B}\mathbf{F}_u^*\mathbf{R}_u\right)\right|du\\ &\quad+\int_0^t \left|\tfrac{1}{n}\operatorname{trace}\left(\mathbf{R}_u\mathbf{F}_u\mathbf{B}^*(\mathbf{X}\mathbf{X}^*+u\mathbf{I})^{-1}(\mathbf{B}^*)^{-1}\mathbf{F}_u^*\mathbf{R}_u\right)\right|du\\ &\quad+\int_0^t \left|\tfrac{1}{n}\operatorname{trace}\left(\mathbf{R}_u\mathbf{F}_u\mathbf{B}^*(\mathbf{X}\mathbf{X}^*+u\mathbf{I})^{-1}(\mathbf{B}^*)^{-1}\mathbf{C}^*\mathbf{R}_u\right)\right|du.\end{aligned}$$

Using the inequality $|\operatorname{trace}(\mathbf{M}_1\mathbf{M}_2\mathbf{M}_3)| \leqslant \|\mathbf{M}_1\|\|\mathbf{M}_3\|\operatorname{trace}(\mathbf{M}_2)$ (which holds for any $n\times n$ matrices $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$ such that $\mathbf{M}_2$ is positive definite) as well as (4.9) and (4.12), we therefore obtain

$$\left|\tfrac{1}{n}\operatorname{trace}(\mathbf{R}_t - \mathbf{R}_0)\right| \leqslant C(K,z)\int_0^t \tfrac{1}{n}\operatorname{trace}(\mathbf{X}\mathbf{X}^* + u\mathbf{I})^{-1}\,du, \tag{4.14}$$

where $C(K,z)$ is some constant depending only on K and z. Thus, it remains to show that

$$\lim_{t\to 0}\limsup_{n\to\infty}\int_0^t \tfrac{1}{n}\operatorname{trace}(\mathbf{X}_n\mathbf{X}_n^* + u\mathbf{I}_n)^{-1}\,du = 0 \text{ in probability.} \tag{4.15}$$

But this follows from the fact that the random matrices $\mathbf{X}_n$ satisfy Condition C; see the proof of Lemma 8.14 in [18] for details. ∎

4.4. On Condition B. Here we have the following result:

LEMMA 4.2. *With $\mathbf{F}_n$ defined as in equation* (1.8), *Condition* B *holds.*

The proof follows from similar estimates to those in Section 8.1 in [18]. Since the required modifications are relatively straightforward, we omit the details.

4.5. On Condition C. Here we provide a number of lemmas which will be helpful in verifying Conditions C and $\mathrm{C}_{\text{simple}}$. Recall that Condition $\mathrm{C}_{\text{simple}}$ was introduced in Remark 4.2.

LEMMA 4.3. *For each $n\in\mathbb{N}$, let $\mathbf{F}_n$ and $\mathbf{G}_n$ be random matrices of dimension $n\times n$. If the matrices $\mathbf{F}_n$ and $\mathbf{G}_n$ satisfy Condition* $\mathrm{C}_{\text{simple}}$, *then the matrix products $\mathbf{F}_n\mathbf{G}_n$ also satisfy Condition* $\mathrm{C}_{\text{simple}}$.

Since this result follows from similar arguments to those in the proof of Theorem 8.22 in [18] or to those for Lemma 4.6 below, we omit the proof.

LEMMA 4.4. *For each $n \in \mathbb{N}$, let $\mathbf{X}_n = \big(\frac{1}{\sqrt{n}} X_{jk}\big)_{j,k=1,\dots,n}$ be as in the assumptions* (1.1)–(1.4). *Then the matrices $\mathbf{X}_n$ and $\mathbf{X}_n^{-1}$ satisfy Condition* C.

For the matrices $\mathbf{X}_n$, Condition C is checked in [20] (in fact, it follows from the relation $\mathbb{E}\|\mathbf{X}\|_2^2 = n$ and from Lemmas 4.9 and 4.10), and for the matrices $\mathbf{X}_n^{-1}$, Condition C follows from the arguments given in the proof of Theorem 8.22 in [18]. We therefore omit the details.

REMARK 4.7. *A careful analysis of the proof of Theorem* 8.22 *in* [18] *shows that if the matrices $\mathbf{G}_n$ satisfy Condition* $\mathrm{C}_{\mathrm{simple}}$, *then the inverse matrices $\mathbf{G}_n^{-1}$ satisfy Conditions* (C1) *and* (C2) *with $\alpha = 0$.*

LEMMA 4.5. *Let $\mathbf{F}_n = (\mathbf{X}_n^{(i_1)})^{\varepsilon_1} \dots (\mathbf{X}_n^{(i_l)})^{\varepsilon_l}$, where $l \in \mathbb{N}$, $i_1, \dots, i_l \in \mathbb{N}$ (not necessarily different), and $\varepsilon_1, \dots, \varepsilon_l \in \{-1, +1\}$ are fixed. Then $\mathbf{F}_n$ satisfies Condition* $\mathrm{C}_{\mathrm{simple}}$.

P r o o f. By Lemma 4.4, the claim is true (even with the stronger Condition C) for $l = 1$. By Lemma 4.3 and induction, the claim remains true for $l > 1$. ∎

LEMMA 4.6. *For each $n \in \mathbb{N}$, let $\mathbf{X}_n = \big(\frac{1}{\sqrt{n}} X_{jk}\big)_{j,k=1,\dots,n}$ be as in the assumptions* (1.1)–(1.4). *Furthermore, for each $n \in \mathbb{N}$, let $\mathbf{A}_n$, $\mathbf{B}_n$ and $\mathbf{C}_n$ be random matrices of dimension $n \times n$ such that $\mathbf{A}_n$, $\mathbf{B}_n$, $\mathbf{C}_n$ and $\mathbf{X}_n$ are independent.*

(a) *If the matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ satisfy Condition* $\mathrm{C}_{\mathrm{simple}}$ *and the matrices $\mathbf{C}_n$ satisfy Condition* (C0), *then the matrices $\mathbf{A}_n\mathbf{X}_n\mathbf{B}_n + \mathbf{C}_n$ satisfy Condition* C.

(b) *If the matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ satisfy Condition* $\mathrm{C}_{\mathrm{simple}}$ *and the matrices $\mathbf{C}_n$ satisfy Condition* C *or $\mathbf{C}_n = \mathbf{0}$ for all $n \in \mathbb{N}$, then the matrices $\mathbf{A}_n\mathbf{X}_n^{-1}\mathbf{B}_n + \mathbf{C}_n$ satisfy Condition* C.

P r o o f. To shorten the notation, we omit the index n throughout this proof. First of all, let us note that if a sequence of random matrices $\mathbf{G}_n$ (with $\mathbf{G}_n$ of dimension $n \times n$) satisfies Condition (C0), there exists some $L_{\mathbf{G}} > 0$ such that

$$\lim_{n\to\infty} \mathbb{P}(\|\mathbf{G}_n\| \geqslant n^{L_{\mathbf{G}}}) = 0. \tag{4.16}$$

In fact, if $p > 0$ is such that $\frac{1}{n}\sum_{k=1}^n s_k^p(\mathbf{G}_n)$ is bounded in probability as $n \to \infty$ and $\varepsilon > 0$ is arbitrary, it follows that

$$\limsup_{n\to\infty} \mathbb{P}\big(s_1(\mathbf{G}_n) \geqslant n^{(1+\varepsilon)/p}\big) \leqslant \limsup_{n\to\infty} \mathbb{P}\big(\frac{1}{n}\sum_{k=1}^n s_k^p(\mathbf{G}_n) \geqslant n^{\varepsilon}\big) = 0,$$

so that the assertion holds for any $L_{\mathbf{G}} > 1/p$.

(a) Condition (C0) follows from Lemmas 4.8 and 4.7, Hölder's inequality, and the fact that the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{X}$ satisfy Condition (C0). To prove Conditions (C1) and (C2), we use the factorization

$$\mathbf{AXB} + \mathbf{C} - \alpha\mathbf{I} = \mathbf{A}\big(\mathbf{X} + \mathbf{A}^{-1}(\mathbf{C} - \alpha\mathbf{I})\mathbf{B}^{-1}\big)\mathbf{B}.$$

Then it remains to check that for each of the three factors $\mathbf{M}_n$ on the right-hand side, we have, for some $Q > 0$,

$$\mathbb{P}\big(s_n(\mathbf{M}_n) \leqslant n^{-Q}\big) = o(1) \quad \text{and} \quad \tfrac{1}{n} \sum_{n_1 \leqslant n \leqslant n_2} \log^- s_j(\mathbf{M}_n) = o_P(1).$$

For $\mathbf{A}$ and $\mathbf{B}$, this is true by assumption. For $\mathbf{X} + \mathbf{A}^{-1}(\mathbf{C} - \alpha\mathbf{I})\mathbf{B}^{-1}$, this follows from Lemmas 4.9 and 4.10. More precisely, if the matrices $\mathbf{A}$ and $\mathbf{B}$ satisfy Condition (C1) with $\alpha = 0$ and $Q > 0$, and the matrices $\mathbf{C}$ satisfy (4.16) with $L_{\mathbf{C}} > 0$, we have $\mathbb{P}\big(s_1\big(\mathbf{A}^{-1}(\mathbf{C} - \alpha\mathbf{I})\mathbf{B}^{-1}\big) > 2n^{2Q+L_{\mathbf{C}}}\big) \to 0$ by Lemma 4.7. Thus, we may use Lemmas 4.9 and 4.10 conditionally on $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and on the set of probability $1 + o(1)$ where $s_1\big(\mathbf{A}^{-1}(\mathbf{C} - \alpha\mathbf{I})\mathbf{B}^{-1}\big) \leqslant 2n^{2Q+L_{\mathbf{C}}}$.

(b) We consider only the case that the matrices $\mathbf{C}$ satisfy Condition C, leaving the simpler case $\mathbf{C} = \mathbf{0}$ to the reader. By reasoning as above, we see that Condition (C0) follows from Lemmas 4.8 and 4.7, Hölder's inequality, and the fact that the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{X}^{-1}$ satisfy Condition (C0). To prove Conditions (C1) and (C2), we use the factorization

$$\mathbf{AX}^{-1}\mathbf{B} + \mathbf{C} - \alpha\mathbf{I} = \mathbf{AX}^{-1}\big(\mathbf{B}(\mathbf{C} - \alpha\mathbf{I})^{-1}\mathbf{A} + \mathbf{X}\big)\mathbf{A}^{-1}(\mathbf{C} - \alpha\mathbf{I}).$$

Then it remains to check that for each of the five factors $\mathbf{M}_n$ on the right-hand side, we have, for some $Q > 0$,

$$\mathbb{P}\big(s_n(\mathbf{M}_n) \leqslant n^{-Q}\big) = o(1) \quad \text{and} \quad \tfrac{1}{n} \sum_{n_1 \leqslant n \leqslant n_2} \log^- s_j(\mathbf{M}_n) = o_P(1).$$

But this is true (i) by assumption, (ii) by Lemma 4.4, (iii) by Lemmas 4.9 and 4.10 (applied conditionally on $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$), (iv) by Remark 4.7, and (v) again by assumption. ∎

4.6. Proof of Theorem 1.3. After the preparations above, we may turn to the proof of Theorem 1.3. Given a sequence of random matrices $(\mathbf{G}_n)_{n\in\mathbb{N}}$, we write $\nu(\mathbf{G}_n)$ for the singular value distributions, $\mu(\mathbf{G}_n\mathbf{G}_n^*)$ for the squared singular value distributions, $\mathcal{S}\nu(\mathbf{G}_n)$ for the symmetrized singular value distributions, and $\nu_{\mathbf{G}}$, $\mu_{\mathbf{G}\mathbf{G}^*}$ and $\mathcal{S}\nu_{\mathbf{G}}$ for the corresponding weak limits in probability (if existent). Furthermore, for $t > 0$, let $\gamma_t^{+1} := \gamma^{+1} := \gamma$, and let γ_t^{-1} be the induced measure of γ under the mapping $x \mapsto (x + t)^{-1}x(x + t)^{-1}$. These notions are motivated by our regularization procedure in equation (4.2).

Let us start with the singular value distributions. We will first use induction on l to prove the claim for the case $m = 1$ and then use induction on m to prove the claim for the case $m > 1$. More precisely, we will show the following:

(4.17) The matrices $\mathbf{F}_n(\mathbf{X})$ from (1.8) satisfy Conditions A and B, and for any $t > 0$, the singular value distributions of the matrices $\mathbf{F}_{n,t}(\mathbf{X})$ converge weakly in probability to the probability measure ν_t on $(0, \infty)$ with symmetrization $\mathcal{S}\nu_t = \big(\mathcal{Q}^{-1}(\gamma_t^{\varepsilon_1} \boxtimes \ldots \boxtimes \gamma_t^{\varepsilon_l})\big)^{\boxplus m}$.

Indeed, by Condition A, we may then let $t \to 0$ get the limiting singular value distribution of the matrices $\mathbf{F}_n(\mathbf{X})$. Note that Condition B has already been established in Lemma 4.2, so that it remains to check Condition A as well as the claim about the limiting distribution.

Products of independent random matrices. For $\mathbf{F}_n(\mathbf{X}) = \mathbf{X}_n$, Condition A holds trivially, and for $\mathbf{F}_n(\mathbf{X}) = \mathbf{X}_n^{-1}$, Condition A holds by Lemma 4.1. Furthermore, the Marchenko–Pastur theorem implies that, for any $t > 0$ and $\varepsilon \in \{-1, +1\}$, we have $\mu\big(\mathbf{X}_{n,t}^{\varepsilon}(\mathbf{X}_{n,t}^{\varepsilon})^*\big) \to \gamma_t^{\varepsilon}$. Thus, (4.17) is true for $l = 1$.

Now let $l > 1$, let $\mathbf{F}_n$ be an l-fold product of independent random matrices, and suppose that (4.17) holds for any product $\mathbf{G}_n$ with less than l factors. It then follows from Lemma 4.1 that the matrices $\mathbf{F}_n(\mathbf{X})$ satisfy Condition A; see Remark 4.6 for details. Now consider the particular decomposition $\mathbf{F}_n(\mathbf{X}) = \mathbf{X}_n^{\varepsilon}\mathbf{G}_n(\mathbf{X})$, where $\varepsilon = +1$ or $\varepsilon = -1$ and $\mathbf{X}_n$ and $\mathbf{G}_n(\mathbf{X})$ are independent. Then, for any $t > 0$, the matrices $\mathbf{Y}_{n,t}^{\varepsilon}$ and $\mathbf{G}_{n,t}(\mathbf{Y})$ are independent bi-unitary invariant matrices with

$$\mu\big(\mathbf{Y}_{n,t}^{\varepsilon}(\mathbf{Y}_{n,t}^{\varepsilon})^*\big) \to \gamma_t^{\varepsilon} \quad \text{and} \quad \mu\big(\mathbf{G}_{n,t}(\mathbf{Y})\mathbf{G}_{n,t}^*(\mathbf{Y})\big) \to \mu_{\mathbf{G}(t)\mathbf{G}(t)^*},$$

by the inductive hypothesis in the latter case. Therefore, by asymptotic freeness (see Proposition 2.1 (a)),

$$\mu\big(\mathbf{F}_{n,t}(\mathbf{Y})\mathbf{F}_{n,t}^*(\mathbf{Y})\big) \to \gamma_t^{\varepsilon} \boxtimes \mu_{\mathbf{G}(t)\mathbf{G}(t)^*}.$$

Thus, by Theorem 4.1 (a), (4.17) holds for the matrices $\mathbf{F}_n(\mathbf{X})$ as well.

Hence, by induction on l, we come to the conclusion that (4.17) holds for any product of independent matrices (i.e. for the case $m = 1$).

Sums of products of independent random matrices. We have just proved (4.17) for $m = 1$. Now let $m > 1$, let $\mathbf{F}_n$ be an m-fold sum of products of independent random matrices, and suppose that (4.17) holds for any such sum $\mathbf{C}_n$ with less than m summands. It then follows by Lemma 4.1 and a similar argument to that in Remark 4.6 that the matrices $\mathbf{F}_n(\mathbf{X})$ satisfy Condition A. Now consider the particular decomposition $\mathbf{F}_n(\mathbf{X}) = \mathbf{G}_n(\mathbf{X}) + \mathbf{C}_n(\mathbf{X})$, where $\mathbf{G}_n(\mathbf{X})$ is a product, $\mathbf{C}_n(\mathbf{X})$ is an $(m-1)$-fold sum of products, and $\mathbf{G}_n(\mathbf{X})$ and $\mathbf{C}_n(\mathbf{X})$ are independent. Then, for any $t > 0$, the matrices $\mathbf{G}_{n,t}(\mathbf{Y})$ and $\mathbf{C}_{n,t}(\mathbf{Y})$ are independent bi-unitary invariant matrices with

$$\mathcal{S}\nu\big(\mathbf{G}_{n,t}(\mathbf{Y})\big) \to \mathcal{S}\nu_{\mathbf{G}(t)} \quad \text{and} \quad \mathcal{S}\nu\big(\mathbf{C}_{n,t}(\mathbf{Y})\big) \to \mathcal{S}\nu_{\mathbf{C}(t)}$$

by the result for the case $m = 1$ and the inductive hypothesis, respectively. Therefore, by asymptotic freeness (see Proposition 2.1 (c)),

$$\mathcal{S}\nu\big(\mathbf{F}_{n,t}(\mathbf{Y})\big) \to \mathcal{S}\nu_{\mathbf{C}(t)} \boxplus \mathcal{S}\nu_{\mathbf{G}(t)}.$$

Thus, by Theorem 4.1 (a), (4.17) holds for the matrices $\mathbf{F}_n(\mathbf{X})$ as well.

Hence, by induction on m, we come to the conclusion that (4.17) holds for any sum of products of independent matrices (i.e. for the case $m > 1$).

Let us now consider the eigenvalue distributions. To begin with, by Lemma 4.6, we may check by induction on m that the matrices $\mathbf{F}_n(\mathbf{X})$ satisfy Condition C, too. Therefore, we may use Theorem 4.1 (b), and it remains to determine the limiting eigenvalue distributions in the Gaussian case, i.e. for the matrices $\mathbf{F}_n(\mathbf{Y})$. Here, it follows by asymptotic freeness (see Proposition 2.1 (d)) that $\mathcal{S}\nu\big(\mathbf{F}_{n,t}(\mathbf{Y}) - \alpha\mathbf{I}_n\big)$ $\to \mathcal{S}\nu_{t,\alpha} := (\mathcal{S}\nu_t) \boxplus B(\alpha)$, with $B(\alpha)$ as in Theorem 2.1. Letting $t \to 0$ and using Condition A, it further follows that $\mathcal{S}\nu\big(\mathbf{F}_n(\mathbf{Y}) - \alpha\mathbf{I}_n\big) \to \mathcal{S}\nu_\alpha := (\mathcal{S}\nu) \boxplus B(\alpha)$, where ν is the probability measure described in the theorem. Now apply Theorem 2.1. ∎

4.7. Proof of Remark 1.1. A slight variation of the preceding arguments shows that Conditions A, B and C continue to hold for random matrices $\mathbf{F}_n$ of the form (1.10), *provided that* the extra condition (1.11) holds:

C o n d i t i o n A. Here we can regularize the matrices $(\mathbf{X}_n^{-1})^l$ by means of $\big((\mathbf{X}_n)_t^{-1}\big)^l$ (i.e. each factor in the power is regularized individually) and invoke Lemma 4.1. For this, it is important that the matrices $\mathbf{A}_n$, $\mathbf{B}_n$ and $\mathbf{C}_n$ in Lemma 4.1 need not be independent of $\mathbf{X}_n$; see Remark 4.4.

C o n d i t i o n B. Here we may extend Lemma 4.2 to products of powers of independent Girko–Ginibre matrices, using similar arguments to those in Sections 8.1.3 and 8.1.4 in [18].

C o n d i t i o n C. Under the extra condition (1.11), it follows from Lemma 4.6 (applied with $\mathbf{X} = \mathbf{X}^{(r)}$) and by induction on m that the matrices $\mathbf{F}_n$ satisfy Condition C. (Unfortunately, without the extra condition (1.11), Lemma 4.6 does not allow us to draw this conclusion in general, even though we would expect that Condition C continues to hold in this case.)

Now, the proof of Remark 1.1 is quite similar to that of Theorem 1.3, which is why we omit the details. ∎

4.8. Auxiliary results. In this subsection we collect several auxiliary results from the literature which we have used to verify Condition C. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ matrices, and recall that $s_1(\mathbf{M}) \geqslant \ldots \geqslant s_n(\mathbf{M})$ denote the singular values of the $n \times n$ matrix $\mathbf{M}$.

LEMMA 4.7 ([24], Theorem 3.3.14). *For all $p > 0$ and all $k = 1, \ldots, n$, we have $\sum_{j=1}^k \big(s_j(\mathbf{AB})\big)^p \leqslant \sum_{j=1}^k \big(s_j(\mathbf{A})s_j(\mathbf{B})\big)^p$.*

LEMMA 4.8 ([24], Theorem 3.3.16). *For all $p>0$, we have $\sum_{j=1}^n s_j^p(\mathbf{A}+\mathbf{B}) \leqslant C_p\big(\sum_{j=1}^n s_j^p(\mathbf{A}) + \sum_{j=1}^n s_j^p(\mathbf{B})\big)$, where C_p is a constant depending only on p.*

Since $s_j(\mathbf{M}^{-1}) = s_{n-j+1}^{-1}(\mathbf{M})$, $j = 1, \dots, n$, it is clear that similar results hold for the smallest singular values.

LEMMA 4.9 ([20], Section 5). *Suppose that the conditions* (1.1)–(1.4) *hold. Then, for any fixed $K > 0$ and $L > 0$, there exist positive constants A and B such that for any non-random matrices $\mathbf{M}_n$ with $\|\mathbf{M}_n\|_2 \leqslant Kn^L$, we have*

$$\mathbb{P}\big(s_n(\mathbf{X}_n - \mathbf{M}_n) \leqslant n^{-A}\big) \leqslant n^{-B}.$$

LEMMA 4.10 ([20], Section 5). *Suppose that the conditions* (1.1)–(1.4) *hold. Then, for any fixed $K > 0$ and $L > 0$, there exists a constant $0 < \gamma < 1$ such that for any non-random matrices $\mathbf{M}_n$ with $\|\mathbf{M}_n\|_2 \leqslant Kn^L$ and for any sequence $\delta_n \to 0$, we have*

$$\lim_{n\to\infty} \tfrac{1}{n} \sum_{n_1 \leqslant j \leqslant n_2} \log^- s_j(\mathbf{X}_n - \mathbf{M}_n) = 0 \text{ almost surely},$$

where $n_1 = [n - n\delta_n] + 1$ and $n_2 = [n - n^\gamma]$.

Acknowledgments. We thank an anonymous referee for careful reading.

REFERENCES

[1] K. Adhikari, N. Kishore Reddy, T. Ram Reddy, and K. Saha, *Determinantal point processes in the plane from products of random matrices*, Ann. Inst. Henri Poincaré Probab. Stat. 52 (1) (2016), pp. 16–46.

[2] G. Akemann and J. R. Ipsen, *Recent exact and asymptotic results for products of independent random matrices*, Acta Phys. Polon. B 46 (9) (2015), pp. 1747–1784.

[3] N. Alexeev, F. Götze, and A. Tikhomirov, *Asymptotic distribution of singular values of powers of random matrices*, Lith. Math. J. 50 (2) (2010), pp. 121–132.

[4] O. Arizmendi E. and V. Pérez-Abreu, *The S-transform of symmetric probability measures with unbounded supports*, Proc. Amer. Math. Soc. 137 (9) (2009), pp. 3057–3066.

[5] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, second edition, Springer, New York 2010.

[6] S. T. Belinschi, T. Mai, and R. Speicher, *Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem*, J. Reine Angew. Math. 732 (2017), pp. 21–53.

[7] S. T. Belinschi, P. Śniady, and R. Speicher, *Eigenvalues of non-Hermitian random matrices and Brown measure of non-normal operators: Hermitian reduction and linearization method*, Linear Algebra Appl. 537 (2018), pp. 48–83.

[8] H. Bercovici and V. Pata, *Stable laws and domains of attraction in free probability theory*, Ann. of Math. (2) 149 (3) (1999), pp. 1023–1060.

[9] H. Bercovici and D. Voiculescu, *Free convolution of measures with unbounded support*, Indiana Univ. Math. J. 42 (3) (1993), pp. 733–773.

[10] P. Biane and F. Lehner, *Computation of some examples of Brown's spectral measure in free probability*, Colloq. Math. 90 (2) (2001), pp. 181–211.

[11] C. Bordenave, *On the spectrum of sum and product of non-Hermitian random matrices*, Electron. Commun. Probab. 16 (2011), pp. 104–113.
[12] C. Bordenave and D. Chafaï, *Around the circular law*, Probab. Surv. 9 (2012), pp. 1–89.
[13] Z. Burda, *Free products of large random matrices – a short review of recent developments*, J. Phys.: Conf. Ser. 473 (2013), 012002.
[14] Z. Burda, R. A. Janik, and B. Waclaw, *Spectrum of the product of independent random Gaussian matrices*, Phys. Rev. E (3) 81 (4) (2010), 041132.
[15] Z. Burda, A. Jarosz, G. Livan, M. A. Nowak, and A. Swiech, *Eigenvalues and singular values of products of rectangular Gaussian random matrices*, Phys. Rev. E (3) 82 (6) (2010), 061114.
[16] P. J. Forrester, *Eigenvalue statistics for product complex Wishart matrices*, J. Phys. A 47 (2014), 345202.
[17] P. J. Forrester and D. Liu, *Raney distributions and random matrix theory*, J. Stat. Phys. 158 (5) (2015), pp. 1051–1082.
[18] F. Götze, H. Kösters, and A. Tikhomirov, *Asymptotic spectra of matrix-valued functions of independent random matrices and free probability*, Random Matrices Theory Appl. 4 (2) (2015), 1550005.
[19] F. Götze and A. Tikhomirov, *The circular law for random matrices*, Ann. Probab. 38 (4) (2010), pp. 1444–1491.
[20] F. Götze and A. N. Tikhomirov, *On the asymptotic spectrum of products of independent random matrices*, preprint, arXiv:1012.2710, 2010.
[21] U. Haagerup and F. Larsen, *Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras*, J. Funct. Anal. 176 (2) (2000), pp. 331–367.
[22] U. Haagerup and H. Schultz, *Brown measures of unbounded operators affiliated with a finite von Neumann algebra*, Math. Scand. 100 (2) (2007), pp. 209–263.
[23] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, American Mathematical Society, Providence, RI, 2000.
[24] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge 1991.
[25] A. Jarosz, *Summing free unitary random matrices*, Phys. Rev. E 84 (2011), 011146.
[26] F. Lehner, *On the computation of spectra in free probability*, J. Funct. Anal. 183 (2) (2001), pp. 451–471.
[27] W. Młotkowski, M. A. Nowak, K. A. Penson, and K. Życzkowski, *Spectral density of generalized Wishart matrices and free multiplicative convolution*, Phys. Rev. E (3) 92 (1) (2015), 012121.
[28] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, Cambridge University Press, Cambridge 2006.
[29] S. O'Rourke and A. Soshnikov, *Products of independent non-Hermitian random matrices*, Electron. J. Probab. 16 (2011), paper no. 81, pp. 2219–2245.
[30] N. R. Rao and R. Speicher, *Multiplication of free random variables and the S-transform: The case of vanishing mean*, Electron. Commun. Probab. 12 (2007), pp. 248–258.
[31] T. Rogers, *Universal sum and product rules for random matrices*, J. Math. Phys. 51 (9) (2010), 093304.
[32] E. B. Saff and V. Totik, *Logarithmic Potentials with External Fields* (with Appendix B by Thomas Bloom), Springer, Berlin 1997.
[33] R. Speicher, *Polynomials in asymptotically free random matrices*, Acta Phys. Polon. B 46 (9) (2015), pp. 1611–1624.
[34] T. Tao and V. Vu, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab. 38 (5) (2010), pp. 2023–2065.
[35] A. N. Tikhomirov, *Asymptotic distribution of the singular numbers for spherical ensemble matrices*, Mat. Tr. 16 (2013), pp. 169–200.
[36] D. A. Timushev and A. N. Tikhomirov, *On the asymptotic distribution of singular val-*

ues of powers products of sparse random matrices, Izvestia Komi Science Center of Ural Division of RAS 13 (2013), pp. 10–17.

[37] D. A. Timushev and A. N. Tikhomirov, *Limit theorems for spectra of sums of random matrices*, Proc. Komi Science Center of Ural Division of RAS 187 (2014), pp. 24–32.

[38] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables: A Noncommutative Probability Approach to Free Products with Applications to Random Matrices, Operator Algebras, and Harmonic Analysis on Free Groups*, American Mathematical Society, Providence, RI, 1992.

Holger Kösters
Department of Mathematics
Bielefeld University, Germany
E-mail: hkoesters@math.uni-bielefeld.de

Alexander Tikhomirov
Institute of Physics and Mathematics
Komi Science Center of Ural Division of RAS
Syktyvkar State University, Russia
E-mail: antikhom51@gmail.com

Received on 14.8.2016;
revised version on 8.4.2017

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 385–405
doi:10.19195/0208-4147.38.2.7

SUPERMODULAR ORDERING OF POISSON AND BINOMIAL RANDOM VECTORS BY TREE-BASED CORRELATIONS*

BY

BÜNYAMIN KIZILDEMIR (SINGAPORE) AND NICOLAS PRIVAULT (SINGAPORE)

Abstract. We construct a dependence structure for binomial, Poisson and Gaussian random vectors, based on partially ordered binary trees and sums of independent random variables. Using this construction, we characterize the supermodular ordering of such random vectors via the componentwise ordering of their covariance matrices. For this, we apply Möbius inversion techniques on partially ordered trees, which allow us to connect the Lévy measures of Poisson random vectors on the discrete d-dimensional hypercube to their covariance matrices.

2010 AMS Mathematics Subject Classification: Primary: 60E15; Secondary: 62H20, 05C05, 06A11, 60E07.

Key words and phrases: Stochastic ordering, supermodular functions, Möbius transform, Möbius inversion, binary trees, Poisson random vectors, binomial random vectors.

1. INTRODUCTION

A d-dimensional random vector $X = (X_1, \dots, X_d)$ is said to be *dominated* by another random vector $Y = (Y_1, \dots, Y_d)$ in the supermodular order, and one writes $X \leqslant_{\rm sm} Y$, if

$$E[\Phi(X)] \leqslant E[\Phi(Y)]$$

for all integrable *supermodular* functions, i.e., for all functions $\Phi : \mathbb{R}^d \to \mathbb{R}$ such that

$$\Phi(x) + \Phi(y) \leqslant \Phi(x \wedge y) + \Phi(x \vee y), \quad x, y \in \mathbb{R}^d,$$

where the maximum $\vee$ and the minimum $\wedge$ are defined with respect to the componentwise order of $x, y \in \mathbb{R}^d$. The supermodular stochastic ordering is used in particular to capture a preference for greater interdependence in economic variables. In other words, we have $X \leqslant_{\rm sm} Y$ if the (positive) dependence among the

* This research was supported by Singapore MOE Tier 1 Grant MOE2015-T1-2-130 RG122/15. We thank two anonymous referees for useful suggestions.

components of Y is greater than the (positive) dependence among the components of X. See, for example, [2] and [3] for applications of supermodular ordering in insurance, and [9] for applications to portfolio risk management, cf. also [8] and references therein for applications in economics.

In the case where X and Y are multivariate Gaussian vectors, the supermodular ordering of X and Y has been characterized by the componentwise ordering of their covariance matrices in [10]. Sufficient conditions for the supermodular ordering of general random vectors have been given in [4] for general random vectors, including Poisson and gamma vectors, cf. Section 4.2 therein. We note that our recursive update of Bernoulli random vectors in (5.8) below consists in an implementation on binary trees of the formulas in Section 4.2 of [4] for Poisson and gamma vectors.

In this paper, we construct a tree-based covariance structure for binomial and Poisson random vectors, under which the supermodular ordering can be characterized by the ordering of covariance matrices, cf. Theorems 4.1 and 5.1. This approach uses Möbius inversion techniques which allow us to connect partially ordered binary trees on the discrete unit hypercube $\{0,1\}^d$ to supermodular ordering. We also show the necessity of dependence structure of this type in Counterexample 4.1. Other types of tree-based dependence structures in the setting of Bernoulli random vectors have been developed in [5] and references therein.

We proceed as follows. In Section 2 we construct a general dependence structure based on independent variables arranged according to a binary tree on the vertices of the d-dimensional hypercube. In Section 3 we describe the Möbius inversion that allows one to recover the parameters of individual random variables from the covariance matrix of the considered random vector. In Section 4 we deal with the case of Poisson random vectors via the use of Lévy measures on the vertices of the discrete unit hypercube $\{0,1\}^d$, cf. Theorem 4.1. In Section 5 we apply this dependence structure to the characterization of the binomial supermodular ordering via the componentwise ordering of covariances, cf. Theorem 5.1. This result naturally extends to the supermodular ordering of sums of binomial, multivariate Gaussian and Poisson random vectors.

2. TREE-BASED CORRELATION STRUCTURES

In this section we introduce the general dependence structure used in this paper. Let $(e_1, \ldots, e_d)$ denote the canonical basis of $\mathbb{R}^d$, and let

$$C_d := \{0,1\}^d = \big\{x = (x_1, \ldots, x_d) \ : \ x_i \in \{0,1\},\ i = 1, \ldots, d\big\}$$

denote the discrete set of vertices of the d-dimensional unit hypercube.

Every $x = (x_1, \ldots, x_d) \in \{0,1\}^d$ is identified with its index set

$$S_x := \big\{i \in \{1, \ldots, d\} \ : \ x_i = 1\big\},$$

and we endow $C_d = \{0,1\}^d$ with the partial inclusion ordering of index sets, i.e., we write

$$x \preceq y \quad \text{when } 0 \leqslant x_i \leqslant y_i \leqslant 1,\ i = 1, \dots, d,$$

and $x \prec y$ when $x \preceq y$ and $x \neq y$; we also let $x \setminus \{a\}$ denote $(x_i \mathbf{1}_{\{i \neq a\}})_{i=1,\dots,d}$ for $x \in C_d$.

2.1. Random vectors. Given $(X_{i,j})_{1 \leqslant i \leqslant j \leqslant d}$ a family of independent random variables and $(e_{k,l})_{1 \leqslant k \leqslant l \leqslant d} \subset C_d$ with $e_{k,k} = e_k$, $k = 1, \dots, d$, we define the random vector $X = (X_1, \dots, X_d)$ by

$$X_i := \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}}} X_{k,l}, \quad i = 1, \dots, d.$$

In other words, we have

$$\begin{aligned} X = \sum_{i=1}^{d} e_i X_i &= \sum_{i=1}^{d} e_i \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}}} X_{k,l} \\ &= \sum_{1 \leqslant k \leqslant l \leqslant d} X_{k,l} \sum_{\substack{1 \leqslant i \leqslant d \\ e_i \preceq e_{k,l}}} e_i = \sum_{1 \leqslant k \leqslant l \leqslant d} X_{k,l} e_{k,l}, \end{aligned} \tag{2.1}$$

which implies

$$E[X_i] = \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}}} E[X_{k,l}], \quad i = 1, \dots, d,$$

and

$$\operatorname{Cov}(X_i, X_j) = \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}, e_j \preceq e_{k,l}}} \sigma_{k,l}^2, \quad 1 \leqslant i \leqslant j \leqslant d, \tag{2.2}$$

where $\sigma_{k,l}^2 := \operatorname{Var}[X_{k,l}]$, $1 \leqslant k \leqslant l \leqslant d$.

EXAMPLE 2.1. If we take $d = 5$, the subset $(e_{k,l})_{1 \leqslant k \leqslant l \leqslant 5}$ of C_5 given by

$$\begin{cases} e_{1,2} = (1,1,0,0,1), \\ e_{1,3} = (1,1,1,0,1), \\ e_{1,4} = (1,0,0,1,0), \\ e_{1,5} = (1,0,0,0,1), \\ e_{2,3} = (0,1,1,0,0), \\ e_{2,4} = (0,1,0,1,0), \\ e_{2,5} = (0,1,0,0,1), \\ e_{3,4} = (0,1,1,1,0), \\ e_{3,5} = (0,1,1,0,1), \\ e_{4,5} = (1,0,0,1,1) \end{cases}$$

corresponds, under (2.1), to the random vector

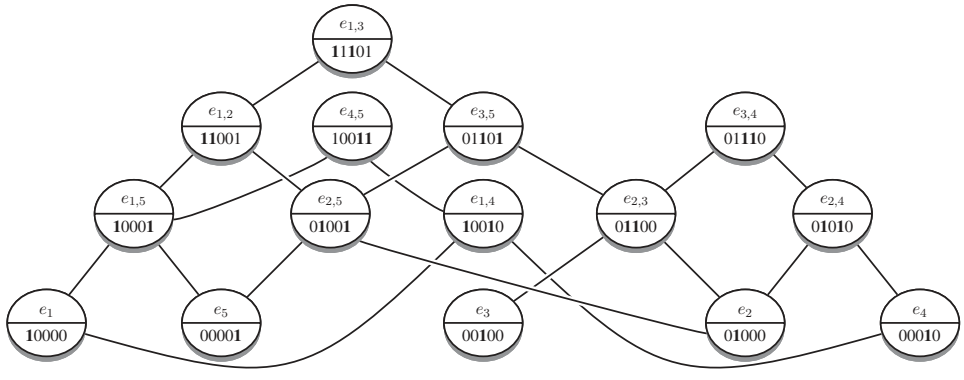
$$(2.3)\quad \begin{cases} X_1 = X_{1,1} + X_{1,2} + X_{1,3} + X_{1,4} + X_{1,5} & & & & & & & & & + X_{4,5} \\ X_2 = X_{2,2} + X_{1,2} + X_{1,3} & & & & + X_{2,3} + X_{2,4} + X_{2,5} & & + X_{3,4} + X_{3,5} & & & \\ X_3 = X_{3,3} & & + X_{1,3} & & & & + X_{2,3} & & & + X_{3,4} + X_{3,5} & & \\ X_4 = X_{4,4} & & & & + X_{1,4} & & & & + X_{2,4} & & + X_{3,4} & & + X_{4,5} \\ X_5 = X_{5,5} + X_{1,2} + X_{1,3} & & & & & & + X_{1,5} & & & + X_{2,5} & & + X_{3,5} & + X_{4,5}. \end{cases}$$

2.2. Binary tree structure. From now on, we work under the following Hypothesis (H) that builds a tree on the set $(e_{k,l})_{1\leqslant k<l\leqslant d}$. Note that not all random vectors admit a tree-based representation according to Hypothesis (H), see Examples 3.4, 3.5 and Counterexample 4.1 below.

(H) The family $(e_{k,l})_{1\leqslant k\leqslant l\leqslant d} \subset \{0,1\}^d$ forms an ordered binary tree for the partial order $\preceq$, in which every node $e_{k,l}$, $k<l$, has exactly two children $e_{k,l}\backslash\{k\}$ and $e_{k,l}\backslash\{l\}$.

We note that the tree $(e_{k,l})_{1\leqslant k\leqslant l\leqslant d}$ has size $d(d+1)/2$ and height at most d.

The random vector (2.3) of Example 2.1 satisfies Hypothesis (H) with the following tree structure:



LEMMA 2.1. *Under Hypothesis* (H) *we have the equivalence*

$$e_{i,j} \preceq e_{k,l} \Longleftrightarrow (e_i \preceq e_{k,l} \text{ and } e_j \preceq e_{k,l})$$

for all $1 \leqslant i \leqslant j \leqslant d$ *and* $1 \leqslant k \leqslant l \leqslant d$.

P r o o f. (i) Assume that $e_{i,j} \preceq e_{k,l}$. Since both children $e_{i,j}\backslash\{i\}$ and $e_{i,j}\backslash\{j\}$ of $e_{i,j}$ satisfy $e_{i,j}\backslash\{i\} \prec e_{i,j}$ and $e_{i,j}\backslash\{j\} \prec e_{i,j}$, we have $e_i \preceq e_{i,j}$ and $e_j \preceq e_{i,j}$, which implies $e_i \preceq e_{k,l}$ and $e_j \preceq e_{k,l}$ since $e_{i,j} \preceq e_{k,l}$.

(ii) Assume that $e_i \preceq e_{k,l}$ and $e_j \preceq e_{k,l}$. We work by decreasing induction on the height of nodes in the tree. If $e_{k,l} = e_k$ is a leaf, i.e. $k = l$, then ($e_i \preceq e_{k,l}$ and $e_j \preceq e_{k,l}$) implies $i = j = k = l$, hence $e_{i,j} = e_i = e_k = e_{k,l}$. Next, assuming that the conclusion holds for all nodes of height at least $h \geqslant 2$, consider a node $e_{k,l}$ of

height $h-1$, with $k \neq l$. If ($e_i \preceq e_{k,l}$ and $e_j \preceq e_{k,l}$) and $\{i,j\} \neq \{k,l\}$, we must have either $i \neq l$ and $j \neq l$, or $i \neq k$ and $j \neq k$. In the first case, ($e_i \preceq e_{k,l}$ and $e_j \preceq e_{k,l}$) implies $e_i \preceq e_{k,l}\backslash\{l\}$ and $e_j \preceq e_{k,l}\backslash\{l\}$, where $e_{k,l}\backslash\{l\}$ has height h, hence $e_i \preceq e_{k,l}\backslash\{l\} \preceq e_{k,l}$ and $e_j \preceq e_{k,l}\backslash\{l\} \preceq e_{k,l}$ by the induction hypothesis. The conclusion is similar in the second case, by replacing l with k. ∎

Based on Lemma 2.1, for all $1 \leqslant i \leqslant j \leqslant d$ we can now rewrite (2.2) as the sum

$$\operatorname{Cov}(X_i, X_j) = \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_{i,j} \preceq e_{k,l}}} \sigma^2_{k,l}, \quad 1 \leqslant i \leqslant j \leqslant d, \tag{2.4}$$

over all couples (k,l) with $e_{i,j} \preceq e_{k,l}$. In other words, $\big(\operatorname{Cov}(X_i, X_j)\big)_{1 \leqslant i \leqslant j \leqslant d}$ is the *Möbius transform* of $(\sigma^2_{k,l})_{1 \leqslant k \leqslant l \leqslant d}$ on the partially ordered set $\big((e_{k,l})_{1 \leqslant k \leqslant l \leqslant d}, \preceq\big)$, cf. [13] or Section 2.5 of [12] for details.

3. MÖBIUS INVERSION

By Möbius inversion (cf. Proposition 2.6.3 of [12]), we can recover the coefficients $(\sigma^2_{k,l})_{1 \leqslant k \leqslant l \leqslant d}$ in (2.2) using the covariances $\big(\operatorname{Cov}(X_i, X_j)\big)_{1 \leqslant i \leqslant j \leqslant d}$ as the sum

$$\sigma^2_{k,l} = \sum_{\substack{1 \leqslant i \leqslant j \leqslant d \\ e_{k,l} \preceq e_{i,j}}} \mu(e_{i,j}, e_{k,l}) \operatorname{Cov}(X_i, X_j), \quad 1 \leqslant k \leqslant l \leqslant d, \tag{3.1}$$

over all couples (i,j) such that $e_{k,l} \preceq e_{i,j}$, $1 \leqslant i \leqslant j \leqslant d$, where $\mu(x,y)$ is the *Möbius function* defined recursively by $\mu(x,x) := 1$ and

$$\mu(x,y) := -\sum_{y \prec z \preceq x} \mu(x,z), \quad x, y \in \{0,1\}^d, \tag{3.2}$$

cf. Proposition 2.6.1 of [12].

PROPOSITION 3.1. *The Möbius function $\mu(x,y)$ on the tree $\big((e_{k,l})_{1 \leqslant k \leqslant l \leqslant d}, \preceq\big)$ is given by*

$$\begin{cases} \mu(e_{k,l}, e_{k,l}) = 1, & \text{(3.3a)} \\ \mu(e_{k,l}, e_{k,l}\backslash\{k\}) = -1, & \\ \mu(e_{k,l}, e_{k,l}\backslash\{l\}) = -1, & \\ \mu(e_{k,l}, e_{k,l}\backslash\{k,l\}) = 1, \quad 1 \leqslant k \leqslant l \leqslant d, & \text{(3.3b)} \end{cases}$$

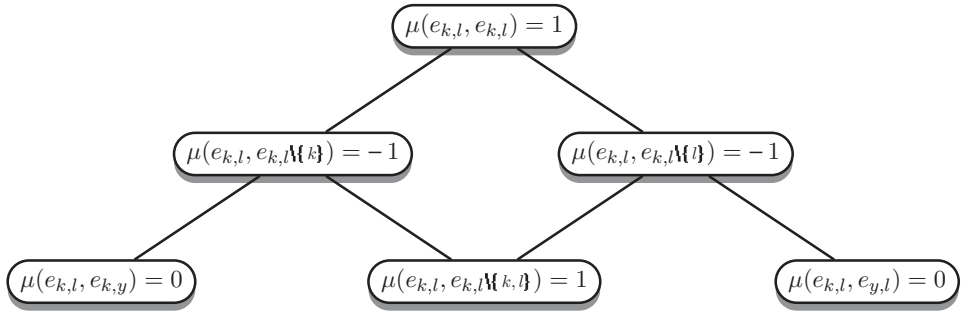
with $\mu(e_{k,l}, e_{i,j}) = 0$ in all other cases.

P r o o f. Given $e_{k,l} \in \{0,1\}^d$, we clearly have

$$\mu(e_{k,l}, e_{k,l}) = 1, \quad \mu(e_{k,l}, e_{k,l}\backslash\{k\}) = -1, \quad \text{and} \quad \mu(e_{k,l}, e_{k,l}\backslash\{l\}) = -1.$$

Next, since the two children $e_{k,l}\backslash\{k\}$ and $e_{k,l}\backslash\{l\}$ of $e_{k,l}$ have themselves a unique common child $e_{k,l}\backslash\{k,l\}$, (3.2) yields $\mu(e_{k,l}, e_{k,l}\backslash\{k,l\}) = 1$. ∎

The next graph, in which $y \notin \{k,l\}$, summarizes the result of Proposition 3.1.



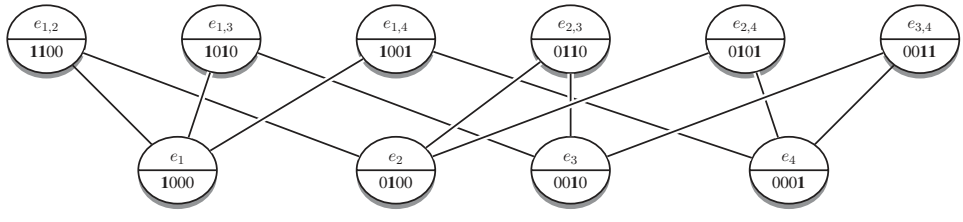
Using formula (3.1), we can now solve (2.2) for $(\sigma^2_{k,l})_{1\leqslant k\leqslant l\leqslant d}$ starting from $\big(\mathrm{Cov}\,(X_i, X_j)\big)_{1\leqslant i\leqslant j\leqslant d}$. However, not all these covariance matrices may lead to a positive solution $(\sigma^2_{k,l})_{1\leqslant k\leqslant l\leqslant d}$, meaning that not all random vectors admit a representation of the form (2.1), see Example 3.4 below.

EXAMPLE 3.1 (*Comonotonic vectors*). The comonotonic vector $(X_{k,l}, X_{k,l}, \ldots, X_{k,l})$ can be represented by using a binary tree with a single node $e_{k,l} = 111\ldots111$ and letting $\sigma^2_{i,j} = 0$ for $(i,j) \neq (k,l)$, since $\mathrm{Cov}\,(X_i, X_j) = \sigma^2_{k,l}$ for all (i,j).

EXAMPLE 3.2 (*Pairwise dependence*). The binary tree is reduced to the d leaves $e_1, \ldots, e_d$, and to their parents $(d-1)d/2$,

$$e_{k,l} = (0, \ldots, 0, \underset{\uparrow}{1}_{k}, 0, \ldots, 0, \underset{\uparrow}{1}_{l}, 0, \ldots, 0), \quad 1 \leqslant k \leqslant l \leqslant d,$$

as in the following example with $d = 4$:



Here, the vector $(X_i)_{i=1,\dots,d}$ is given by

$$\text{(3.4)}\qquad \begin{cases} X_1 = X_{1,1} + X_{1,2} + X_{1,3} + X_{1,4} \\ X_2 = X_{2,2} + X_{1,2} + X_{2,3} + X_{2,4} \\ X_3 = X_{3,3} + X_{1,3} + X_{2,3} + X_{3,4} \\ X_4 = X_{4,4} + X_{1,4} + X_{2,4} + X_{3,4}, \end{cases}$$

and for any $d \geqslant 1$, by (2.2) we have

$$\operatorname{Cov}(X_i, X_j) = \sigma^2_{i,j}, \quad 1 \leqslant i < j \leqslant d,$$

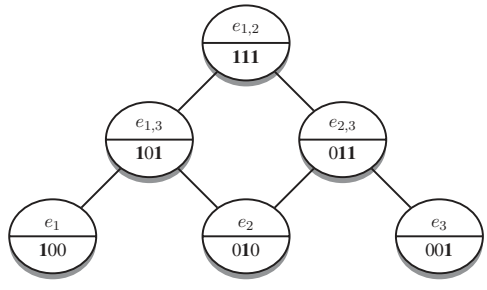
and

$$\text{(3.5)}\qquad \operatorname{Var}[X_i] = \sum_{j=1}^{i-1} \sigma^2_{j,i} + \sum_{j=i}^{d} \sigma^2_{i,j}, \quad i = 1, \dots, d.$$

Here, the inversion of (3.5) by the Möbius transform (3.1) reads

$$\sigma^2_{k,k} = \operatorname{Var}[X_k] - \sum_{l=1,\ l\neq k}^{d} \operatorname{Cov}(X_k, X_l), \quad k = 1, \dots, d.$$

EXAMPLE 3.3 (*Recombining trees*). In dimension $d = 3$, the only available tree structure in addition to the pairwise dependence of Example 3.2 is the recombining (or binomial) full tree



which is associated with the random vector

$$\begin{cases} X_1 = X_{1,1} + X_{1,3} + X_{1,2} \\ X_2 = X_{2,2} + X_{2,3} + X_{1,2} \\ X_3 = X_{3,3} + X_{1,3} + X_{2,3} + X_{1,2}, \end{cases}$$

with the inversion formula (3.1) written as

$$\text{(3.6)}\qquad \begin{cases} \sigma^2_{1,1} = \operatorname{Cov}(X_1, X_1) - \operatorname{Cov}(X_1, X_3), \\ \sigma^2_{2,2} = \operatorname{Cov}(X_2, X_2) - \operatorname{Cov}(X_1, X_3) - \operatorname{Cov}(X_2, X_3) + \operatorname{Cov}(X_1, X_2), \\ \sigma^2_{3,3} = \operatorname{Cov}(X_3, X_3) - \operatorname{Cov}(X_2, X_3), \\ \sigma^2_{1,3} = \operatorname{Cov}(X_1, X_3) - \operatorname{Cov}(X_1, X_2), \\ \sigma^2_{2,3} = \operatorname{Cov}(X_2, X_3) - \operatorname{Cov}(X_1, X_2), \\ \sigma^2_{1,2} = \operatorname{Cov}(X_1, X_2). \end{cases}$$

EXAMPLE 3.4 (*Multivariate Gaussian vectors*). If $(X_{i,j})_{1\leqslant i\leqslant j\leqslant d}$ is a family of independent Gaussian random variables, then $X=(X_1,\ldots,X_d)$ in (2.1) is a multivariate Gaussian vector with matrix $\big(\mathrm{Cov}\,(U_i,U_j)\big)_{1\leqslant i\leqslant j\leqslant d}$ of nonnegative covariances given by (2.4). However, not all Gaussian vectors can fit into a tree-based structure under Hypothesis (H) above. For example, when $d=3$, consider the multivariate Gaussian vector

$$\begin{cases} X_1 = X_{1,1}+X_{1,3} \qquad +X_{1,2}+Z \\ X_2 = X_{2,2} \qquad +X_{2,3}+X_{1,2}+Z \\ X_3 = X_{3,3}+X_{1,3}+X_{2,3}+X_{1,2}, \end{cases} \tag{3.7}$$

where $(X_{k,l})_{1\leqslant k\leqslant l\leqslant d}$ are standard normal random variables and Z is an independent Gaussian random variable with variance four. Here, (X_1,X_2,X_3) has the (positive definite) covariance matrix

$$\begin{pmatrix} \mathrm{Cov}\,(X_1,X_1) & \mathrm{Cov}\,(X_1,X_2) & \mathrm{Cov}\,(X_1,X_3) \\ \mathrm{Cov}\,(X_1,X_2) & \mathrm{Cov}\,(X_2,X_2) & \mathrm{Cov}\,(X_2,X_3) \\ \mathrm{Cov}\,(X_1,X_3) & \mathrm{Cov}\,(X_2,X_3) & \mathrm{Cov}\,(X_3,X_3) \end{pmatrix} = \begin{pmatrix} 7 & 5 & 2 \\ 5 & 7 & 2 \\ 2 & 2 & 4 \end{pmatrix},$$

in which case (3.6) cannot yield a nonnegative solution $(\sigma^2_{k,l})_{1\leqslant k\leqslant l\leqslant 3}$, e.g. when $(k,l)=(1,3)$. In this case, the multivariate Gaussian vector (X_1,X_2,X_3) given by (3.7) admits no binary tree-based representation as the inversion formula (3.6) is based on a full tree.

EXAMPLE 3.5. As in Example 3.4 above, binomial, Poisson and gamma random vectors having a given matrix of nonnegative covariances can be constructed on a binary tree, provided that (3.1) admits a nonnegative solution $(\sigma^2_{k,l})_{1\leqslant k\leqslant l\leqslant 3}$ since their marginals are characterized by their variance parameters and they are stable by summation. However, in this case the construction may not be unique, depending on the chosen binary tree structure, as their joint distribution is not characterized by their covariance matrices.

EXAMPLE 3.6. The particular dependence structure considered in [7] for Poisson random vectors corresponds to the binary tree built on the $d(d-1)/2$ nodes

$$e_{i,j}=(1,\ldots,1,\underset{\substack{\uparrow\\ i}}{1},0,\ldots,0,\underset{\substack{\uparrow\\ j}}{1},0,\ldots,0), \qquad 1\leqslant i<j\leqslant d,$$

and on the d leaves $e_1,\ldots,e_d$.

4. POISSON RANDOM VECTORS

In this section we provide a characterization of the supermodular ordering of Poisson random vectors, based on their covariance matrices in Theorem 4.1. This extends the results of [7] (cf. Example 3.6 above) to more general dependence structures.

Recall that any d-dimensional infinitely divisible Poisson random vector $X = (X_1, \dots, X_d)$ is defined by its characteristic function

$$E[e^{i\langle \bar{t}, X\rangle}] = \exp\big(\int_{\mathbb{R}^d} (e^{i\langle \bar{t}, x\rangle} - 1)\mu(dx)\big),$$

where $\bar{t} = (t_1, \dots, t_d) \in \mathbb{R}^d$, $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^d$, and the Lévy measure

$$\mu(dx) := \sum_{y \in \{0,1\}^d} a_y \delta_y(dx)$$

is supported on $C_d = \{0,1\}^d$. Here δ_y denotes the Dirac measure at the point $y \in \{0,1\}^d$, and $(a_y)_{y \in C_d}$ is a family of nonnegative coefficients with $a_{(0,\dots,0)} = 0$.

Equivalently, $X = (X_1, \dots, X_d)$ can be represented as

$$X_i = \sum_{y \in \{0,1\}^d} \mathbf{1}_{\{i \in y\}} Z_y = \sum_{y \in C_d,\ e_i \preceq y} Z_y, \quad i = 1, \dots, d, \tag{4.1}$$

where $(Z_y)_{y \in C_d \setminus \{0\}}$ is a family of $2^d - 1$ independent Poisson random variables with respective intensities $(a_y)_{y \in C_d \setminus \{0\}}$, cf. also Example 4.3 of [4] and Theorem 3 of [6].

To characterize the ordering of Poisson random vectors based on the data of their covariance matrices which contain only $d(d+1)/2$ components, we consider Lévy measures of the form

$$\mu(dx) = \sum_{1 \leqslant k \leqslant l \leqslant d} a_{k,l} \delta_{e_{k,l}}(dx), \tag{4.2}$$

on $\{0,1\}^d$, where $a_{k,l} \in \mathbb{R}_+$, $1 \leqslant k \leqslant l \leqslant d$. In this case, (4.1) rewrites as

$$X_i = \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}}} X_{k,l}, \tag{4.3}$$

where $(X_{k,l})_{1 \leqslant k \leqslant l \leqslant d}$ is a family of independent Poisson random variables whose respective intensities $(a_{i,j})_{1 \leqslant i \leqslant j \leqslant d}$ satisfy $\operatorname{Var}[X_{k,l}] = E[X_{k,l}] = a_{k,l}$, $1 \leqslant k \leqslant l \leqslant d$.

In the remaining of this section we assume that the family $(e_{k,l})_{1 \leqslant k \leqslant l \leqslant d} \subset \{0,1\}^d$ forms a binary tree according to Hypothesis (H). In this case, the Möbius inversion formula (3.1) shows that

$$a_{k,l} = \sum_{\substack{1 \leqslant i \leqslant j \leqslant d \\ e_{k,l} \preceq e_{i,j}}} \mu(e_{i,j}, e_{k,l}) \operatorname{Cov}(X_i, X_j), \quad 1 \leqslant k \leqslant l \leqslant d. \tag{4.4}$$

4.1. Supermodular ordering of Poisson random vectors. Theorem 4.1 below is a direct consequence of the following Lemma 4.1 which yields the decomposition

$$\mu(dx) = \sum_{i=1}^{d} \mathrm{Var}[X_i]\delta_{e_i}(dx) + \sum_{1\leqslant i<j\leqslant d} \mathrm{Cov}(X_i, X_j)(\delta_{e_{i,j}} + \delta_{e_{i,j}\setminus\{i,j\}} - \delta_{e_{i,j}\setminus\{i\}} - \delta_{e_{i,j}\setminus\{j\}})(dx)$$

of a Lévy measure $\mu(dx)$ of the form (4.2) under Hypothesis (H).

LEMMA 4.1. *Let $(X_1, \ldots, X_d)$ be an infinitely divisible Poisson random vector written as in* (4.3) *under Hypothesis* (H), *with Lévy measure $\mu(dx)$ on C_d. Then we have*

$$\int_{\mathbb{R}^d} \phi(x)\mu(dx) = \sum_{i=1}^{d} E[X_i]\phi(e_i) + \sum_{1\leqslant i<j\leqslant d} \mathrm{Cov}(X_i, X_j)\big(\phi(e_{i,j}) + \phi(e_{i,j}\setminus\{i,j\}) - \phi(e_{i,j}\setminus\{i\}) - \phi(e_{i,j}\setminus\{j\})\big) \tag{4.5}$$

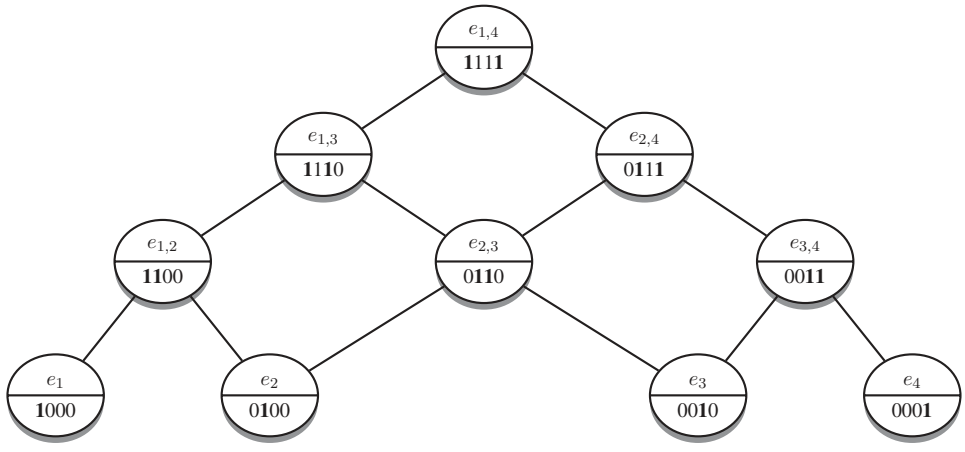
for any function $\phi : \{0,1\}^d \to \mathbb{R}$ such that $\phi(0) = 0$.

P r o o f. By the Möbius inversion formula (3.1) we have

$$\begin{aligned}
\int_{\mathbb{R}^d} \phi(x)\mu(dx) &= \sum_{1\leqslant k\leqslant l\leqslant d} a_{k,l}\phi(e_{k,l}) \\
&= \sum_{1\leqslant k\leqslant l\leqslant d} \phi(e_{k,l}) \sum_{\substack{1\leqslant i\leqslant j\leqslant d \\ e_{k,l}\preceq e_{i,j}}} \mu(e_{i,j}, e_{k,l})\,\mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{d} \mathrm{Cov}(X_i, X_i) \sum_{\substack{1\leqslant k\leqslant d \\ e_k\preceq e_i}} \mu(e_i, e_k)\phi(e_k) \\
&\quad + \sum_{1\leqslant i<j\leqslant d} \mathrm{Cov}(X_i, X_j) \sum_{\substack{1\leqslant k<l\leqslant d \\ e_{k,l}\preceq e_{i,j}}} \mu(e_{i,j}, e_{k,l})\phi(e_{k,l}) \\
&= \sum_{i=1}^{d} E[X_i]\phi(e_i) \\
&\quad + \sum_{1\leqslant i<j\leqslant d} \mathrm{Cov}(X_i, X_j)\big(\phi(e_{i,j}) + \phi(e_{i,j}\setminus\{i,j\}) - \phi(e_{i,j}\setminus\{i\}) - \phi(e_{i,j}\setminus\{j\})\big),
\end{aligned}$$

where we used (3.3a), (3.3b) and the fact that $e_k \preceq e_i$ if and only if $k = i$. ∎

EXAMPLE 4.1. For $d = 4$, the tree structure



is satisfied by the random vector

$$\begin{cases} X_1 = X_{1,1} + X_{1,2} + X_{1,3} + X_{1,4} \\ X_2 = X_{2,2} + X_{1,2} + X_{1,3} + X_{1,4} + X_{2,3} + X_{2,4} \\ X_3 = X_{3,3} + X_{1,3} + X_{1,4} + X_{2,3} + X_{2,4} + X_{3,4} \\ X_4 = X_{4,4} + X_{1,4} + X_{2,4} + X_{3,4}, \end{cases}$$

and relation (4.5) reads

$$\begin{aligned} &\int_{\mathbb{R}^d} \phi(x)\mu(dx) = a_{1,4}\phi(1,1,1,1) + a_{1,3}\phi(1,1,1,0) + a_{2,4}\phi(0,1,1,1) \\ &\quad + a_{1,2}\phi(1,1,0,0) + a_{2,3}\phi(0,1,1,0) + a_{3,4}\phi(0,0,1,1) \\ &\quad + a_{1,1}\phi(1,0,0,0) + a_{2,2}\phi(0,1,0,0) + a_{3,3}\phi(0,0,1,0) + a_{4,4}\phi(0,0,0,1) \\ &= E[X_1]\phi(1,0,0,0) + E[X_2]\phi(0,1,0,0) \\ &\quad + E[X_3]\phi(0,0,1,0) + E[X_4]\phi(0,0,0,1) \\ &\quad + \operatorname{Cov}(X_1,X_2)\big(\phi(1,1,0,0) + \phi(0,0,0,0) - \phi(1,0,0,0) - \phi(0,1,0,0)\big) \\ &\quad + \operatorname{Cov}(X_1,X_3)\big(\phi(1,1,1,0) + \phi(0,1,0,0) - \phi(1,1,0,0) - \phi(0,1,1,0)\big) \\ &\quad + \operatorname{Cov}(X_1,X_4)\big(\phi(1,1,1,1) + \phi(0,1,1,0) - \phi(1,1,1,0) - \phi(0,1,1,1)\big) \\ &\quad + \operatorname{Cov}(X_2,X_3)\big(\phi(0,1,1,0) + \phi(0,0,0,0) - \phi(0,1,0,0) - \phi(0,0,1,0)\big) \\ &\quad + \operatorname{Cov}(X_2,X_4)\big(\phi(0,1,1,1) + \phi(0,0,1,0) - \phi(0,1,1,0) - \phi(0,0,1,1)\big) \\ &\quad + \operatorname{Cov}(X_3,X_4)\big(\phi(0,0,1,1) + \phi(0,0,0,0) - \phi(0,0,1,0) - \phi(0,0,0,1)\big). \end{aligned}$$

Consider now two Poisson random vectors X and Y whose respective Lévy measures μ and ν are represented as in (4.2), i.e.,

$$\mu(dx) = \sum_{1 \leqslant i \leqslant j \leqslant d} a_{i,j}\delta_{e_{i,j}}(dx) \quad \text{and} \quad \nu(dx) = \sum_{1 \leqslant i \leqslant j \leqslant d} b_{i,j}\delta_{e_{i,j}}(dx).$$

If X_i has the same distribution as Y_i for all $i = 1, \dots, d$, then $E[X_i] = E[Y_i]$, $i = 1, \dots, d$, and Lemma 4.1 shows that

$$\begin{aligned}(4.6)\quad &\int_{\mathbb{R}^d} \phi(y)\nu(dy) - \int_{\mathbb{R}^d} \phi(x)\mu(dx) \\ &\quad = \sum_{1\leqslant i<j\leqslant d} \big(\mathrm{Cov}\,(Y_i, Y_j) - \mathrm{Cov}\,(X_i, X_j)\big)\big(\phi(e_{i,j}) + \phi(e_{i,j}\backslash\{i,j\}) \\ &\qquad - \phi(e_{i,j}\backslash\{i\}) - \phi(e_{i,j}\backslash\{j\})\big)\end{aligned}$$

under Hypothesis (H). Relation (4.6) implies in particular that the nonnegativity of the coefficients

$$(4.7)\qquad \mathrm{Cov}\,(Y_i, Y_j) - \mathrm{Cov}\,(X_i, X_j) \geqslant 0, \quad 1 \leqslant i < j \leqslant d,$$

becomes a necessary and sufficient condition for the supermodular ordering of the Lévy measures μ and ν.

The following Theorem 4.1 reformulates (4.7) as a necessary and sufficient condition for supermodular ordering of infinitely divisible Poisson random vectors, based on Theorem 4.5 of [1], which allows one to carry over the notion of supermodularity from the setting of Lévy measures on the discrete cube $C_d = \{0, 1\}^d$ to the setting of Poisson random variables.

THEOREM 4.1. *Consider two Poisson random vectors X and Y both represented as in* (4.3) *under Hypothesis* (H)*. Then the conditions*

$$(4.8)\qquad E[X_i] = E[Y_i], \quad 1 \leqslant i \leqslant d,$$

and

$$(4.9)\qquad \mathrm{Cov}\,(X_i, X_j) \leqslant \mathrm{Cov}\,(Y_i, Y_j), \quad 1 \leqslant i < j \leqslant d,$$

are necessary and sufficient for the supermodular ordering $X \leqslant_{\mathrm{sm}} Y$.

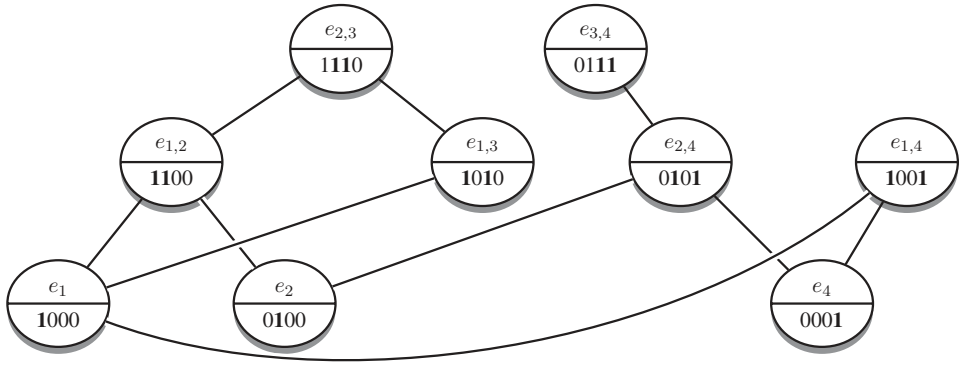
P r o o f. It is well known (cf. e.g. Theorem 3.9.5 of [11]) that for any couple (X, Y) of d-dimensional random vectors, the condition $X \leqslant_{\mathrm{sm}} Y$ implies (4.8) and (4.9), therefore it suffices to show sufficiency. For this, by Theorem 4.5 in [1] it suffices to show that we have

$$(4.10)\qquad \int_{\mathbb{R}^d} \phi(x)\mu(dx) \leqslant \int_{\mathbb{R}^d} \phi(y)\nu(dy)$$

for all supermodular functions $\phi : \mathbb{R}^d \to \mathbb{R}$, where $\mu(dx)$ and $\nu(dy)$ denote the Lévy measures of X and Y, respectively. By Lemma 4.1 we have the identity (4.6) under condition (4.9), which allows us to conclude that (4.10) holds for all supermodular functions ϕ. ∎

Next, we consider a situation where Hypothesis (H) is not satisfied and the equivalence of Theorem 4.1 does not hold.

COUNTEREXAMPLE 4.1. If we take $d = 4$, the tree



does not satisfy Hypothesis (H), and for its corresponding random vector

$$\text{(4.11)}\qquad \begin{cases} X_1 = X_{1,1} + X_{1,2} + X_{1,3} + X_{1,4} + X_{2,3} \\ X_2 = X_{2,2} + X_{1,2} + X_{2,3} + X_{2,4} + X_{3,4} \\ X_3 = X_{1,3} + X_{2,3} + X_{3,4} \\ X_4 = X_{4,4} + X_{1,4} + X_{2,4} + X_{3,4}, \end{cases}$$

relation (4.1) reads

$$\begin{aligned} &\int_{\mathbb{R}^d} \phi(x)\mu(dx) = a_{2,3}\phi(1,1,1,0) + a_{3,4}\phi(0,1,1,1) \\ &\quad + a_{1,2}\phi(1,1,0,0) + a_{1,3}\phi(1,0,1,0) + a_{2,4}\phi(0,1,0,1) + a_{1,4}\phi(1,0,0,1) \\ &\quad + a_{1,1}\phi(1,0,0,0) + a_{2,2}\phi(0,1,0,0) + a_{4,4}\phi(0,0,0,1) \\ &= E[X_1]\phi(1,0,0,0) + E[X_2]\phi(0,1,0,0) \\ &\quad + E[X_3]\phi(0,0,1,0) + E[X_4]\phi(0,0,0,1) \\ &\quad + \operatorname{Cov}(X_1,X_2)\big(\phi(1,1,0,0) + \phi(0,0,0,0) - \phi(1,0,0,0) - \phi(0,1,0,0)\big) \\ &\quad + \operatorname{Cov}(X_1,X_3)\big(\phi(1,0,1,0) + \phi(0,0,0,0) - \phi(1,0,0,0) - \phi(0,0,1,0)\big) \\ &\quad + \operatorname{Cov}(X_1,X_4)\big(\phi(1,0,0,1) + \phi(0,0,0,0) - \phi(1,0,0,0) - \phi(0,0,0,1)\big) \\ &\quad + \operatorname{Cov}(X_2,X_3)\big(\phi(1,1,1,0) + \phi(1,0,0,0) - \phi(1,1,0,0) - \phi(1,0,1,0)\big) \\ &\quad + \operatorname{Cov}(X_2,X_4)\big(\phi(0,1,0,1) + \phi(0,0,0,0) - \phi(0,1,0,0) - \phi(0,0,0,1)\big) \\ &\quad + \operatorname{Cov}(X_3,X_4)\big(\phi(0,1,1,1) + \phi(0,0,0,0) - \phi(0,1,0,1) - \phi(0,0,1,0)\big) \\ &\quad - \operatorname{Cov}(X_3,X_4)\big(\phi(1,1,1,0) + \phi(1,0,0,0) - \phi(1,0,1,0) - \phi(1,1,0,0)\big). \end{aligned}$$

In this case, the conclusion of Theorem 4.1 cannot hold for vectors of the form (4.11) as the sum of the above two terms in factor of $\operatorname{Cov}(X_3, X_4)$ can become negative, e.g. for the supermodular function $\phi(x_1, x_2, x_3, x_4) = x_1x_2x_3$ on the unit cube.

The next proposition replaces the equality of means in (4.8) with an inequality, and is obtained as in Proposition 4.3 of [7] by extending Theorem 4.5 of [1] to nondecreasing supermodular functions ϕ on $\mathbb{R}^d$ satisfying $\phi(0) = 0$, using the same approximation as in Lemma 4.4 therein.

PROPOSITION 4.1. *Consider two Poisson random vectors X and Y both represented as in* (4.3) *under Hypothesis* (H), *and assume that*

$$E[X_i] \leqslant E[Y_i], \quad 1 \leqslant i \leqslant d,$$

and

$$\mathrm{Cov}\,(X_i, X_j) \leqslant \mathrm{Cov}\,(Y_i, Y_j), \quad 1 \leqslant i < j \leqslant d.$$

Then we have

$$E[\Phi(X)] \leqslant E[\Phi(Y)]$$

for all nondecreasing *supermodular functions* $\Phi : \mathbb{R}^d \to \mathbb{R}$.

4.2. Convex ordering. The next result is a remark on the convex ordering of Poisson random vectors represented as in (4.3).

PROPOSITION 4.2. *Consider two Poisson random vectors X and Y both represented as in* (4.3) *under Hypothesis* (H). *Then we have $X \leqslant_{\mathrm{cx}} Y$ if and only if X and Y have the same distribution.*

P r o o f. We assume that $X \leqslant_{\mathrm{cx}} Y$, i.e., we have $E[\Phi(X)] \leqslant E[\Phi(Y)]$ for all convex functions $\Phi : \mathbb{R}^d \to \mathbb{R}$. Clearly, this implies $E[X_k] = E[Y_k]$, $k = 1, \ldots, d$, and by the same argument as in part (b) of the proof of Theorem 4.5 in [1] we also have $\mu \leqslant_{\mathrm{cx}} \nu$. Assume now that $\mathrm{Cov}\,(Y_k, Y_l) > \mathrm{Cov}\,(X_k, X_l)$ for some $1 \leqslant k < l \leqslant d$. The function

$$(x_1, \ldots, x_d) \mapsto \phi_{k,l}(x_1, \ldots, x_d) := \max\big(0, x_l - x_k - \sum_{a \notin e_{k,l}} x_a\big)$$

is convex on $\mathbb{R}^d$ and satisfies $\phi_{k,l}(e_{i,j}) = 1$ when $e_{i,j}$ is a (non-strict) descendant of $e_{k,l}\backslash\{k\}$ that contains l, and $\phi_{k,l}(e_{i,j}) = 0$ in all other cases. This shows that

$$\phi_{k,l}(e_{k,l}) + \phi_{k,l}(e_{k,l}\backslash\{k,l\}) - \phi_{k,l}(e_{k,l}\backslash\{k\}) - \phi_{k,l}(e_{k,l}\backslash\{l\}) = -1,$$

and

$$\phi_{k,l}(e_{i,j}) + \phi_{k,l}(e_{i,j}\backslash\{i,j\}) - \phi_{k,l}(e_{i,j}\backslash\{i\}) - \phi_{k,l}(e_{i,j}\backslash\{j\}) = 0$$

when $(i,j) \neq (k,l)$. Therefore, since $\mathrm{Cov}\,(Y_k, Y_l) > \mathrm{Cov}\,(X_k, X_l)$, Lemma 4.1 shows that

$$\begin{aligned}
&\int_{\mathbb{R}^d} \phi(y)\nu(dy) - \int_{\mathbb{R}^d} \phi(x)\mu(dx) \\
&\quad = \sum_{1\leqslant i<j\leqslant d} \big(\mathrm{Cov}\,(Y_i, Y_j) - \mathrm{Cov}\,(X_i, X_j)\big)\big(\phi(e_{i,j}) + \phi(e_{i,j}\backslash\{i,j\}) \\
&\qquad - \phi(e_{i,j}\backslash\{i\}) - \phi(e_{i,j}\backslash\{j\})\big) \\
&\quad = \big(\mathrm{Cov}\,(Y_k, Y_l) - \mathrm{Cov}\,(X_k, X_l)\big)\big(\phi(e_{k,l}) + \phi(e_{k,l}\backslash\{k,l\}) \\
&\qquad - \phi(e_{k,l}\backslash\{k\}) - \phi(e_{k,l}\backslash\{l\})\big) \\
&\quad < 0,
\end{aligned}$$

which contradicts the fact that $\mu \leqslant_{\mathrm{cx}} \nu$, hence $\mathrm{Cov}\,(Y_k, Y_l) \leqslant \mathrm{Cov}\,(X_k, X_l)$. If $\mathrm{Cov}\,(Y_k, Y_l) < \mathrm{Cov}\,(X_k, X_l)$, we can proceed similarly with the convex function

$$(x_1, \ldots, x_d) \mapsto -\phi_{k,l}(x_1, \ldots, x_d),$$

and conclude that $\mathrm{Cov}\,(Y_k, Y_l) = \mathrm{Cov}\,(X_k, X_l)$ for all $1 \leqslant k \leqslant l \leqslant d$, hence by (4.4) the vectors X and Y have the same distribution. ∎

5. BINOMIAL RANDOM VECTORS

In this section we provide a characterization of the supermodular ordering of binomial random vectors, based on their covariance matrices, cf. Theorem 5.1.

Consider $(Z_1, \ldots, Z_n)$ independent Bernoulli random variables with parameter $p \in [0,1]$ and $\big(A(e_{k,l})\big)_{1\leqslant k\leqslant l\leqslant d}$ a *partition* of $\{1, \ldots, n\}$. Let $(X_{k,l})_{1\leqslant k\leqslant l\leqslant d} = (X_{A(e_{k,l})})_{1\leqslant k\leqslant l\leqslant d}$ denote the family of independent binomial random variables given by

$$X_{k,l} = X_{A(e_{k,l})} := \sum_{i\in A(e_{k,l})} Z_i, \quad 1 \leqslant k \leqslant l \leqslant d,$$

with

$$E[X_{A(e_{k,l})}] = p|A(e_{k,l})|, \quad 1 \leqslant k \leqslant l \leqslant d,$$

where $|A(e_{k,l})|$ denotes the cardinality of $A(e_{k,l})$, and

$$\sigma^2_{k,l} = \mathrm{Var}\,[X_{A(e_{k,l})}] = pq|A(e_{k,l})|, \quad 1 \leqslant k \leqslant l \leqslant d,\ q := 1-p.$$

Let now

$$A_i := \bigcup_{\substack{1\leqslant k\leqslant l\leqslant d \\ e_i \preceq e_{k,l}}} A(e_{k,l}), \quad i = 1, \ldots, d,$$

and consider the vector $(X_1, \ldots, X_d) = (X_{A_1}, \ldots, X_{A_d})$ of binomial random variables defined by

$$X_i = X_{A_i} := \sum_{k\in A_i} Z_k = \sum_{\substack{1\leqslant k\leqslant l\leqslant d \\ e_i \preceq e_{k,l}}} X_{A(e_{k,l})}, \quad i = 1, \ldots, d. \tag{5.1}$$

In general, we have

$$E[X_{A_i}] = p \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}}} |A(e_{k,l})|, \quad i = 1, \dots, d,$$

and

$$\operatorname{Cov}(X_{A_i}, X_{A_j}) = pq \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_{i,j} \preceq e_{k,l}}} |A(e_{k,l})|, \quad 1 \leqslant i \leqslant j \leqslant d.$$

Assuming that the family $(e_{k,l})_{1 \leqslant k \leqslant l \leqslant d} \subset C_d$ forms a binary tree according to Hypothesis (H), the Möbius inversion formula (3.1) shows that we have

$$pq|A(e_{k,l})| = \sum_{\substack{1 \leqslant k \leqslant d \\ e_{k,l} \preceq e_{i,j}}} \mu(e_{i,j}, e_{k,l}) \operatorname{Cov}(X_{A_i}, X_{A_j}), \quad 1 \leqslant k \leqslant l \leqslant d. \tag{5.2}$$

The following is the main result of this section.

THEOREM 5.1. *Consider* $(X_{A_1}, \dots, X_{A_d})$ *and* $(X_{B_1}, \dots, X_{B_d})$ *two binomial random vectors represented as in* (5.1) *under Hypothesis* (H). *Then the conditions*

$$E[X_{A_i}] = E[X_{B_i}], \quad 1 \leqslant i \leqslant d, \tag{5.3}$$

and

$$\operatorname{Cov}(X_{A_i}, X_{A_j}) \leqslant \operatorname{Cov}(X_{B_i}, X_{B_j}), \quad 1 \leqslant i < j \leqslant d, \tag{5.4}$$

are necessary and sufficient for the supermodular ordering

$$(X_{A_1}, \dots, X_{A_d}) \leqslant_{\mathrm{sm}} (X_{B_1}, \dots, X_{B_d}).$$

P r o o f. By Theorem 3.9.5 of [11], it suffices to show sufficiency. Using induction, it is also sufficient to consider the case where

$$\operatorname{Cov}(X_{B_k}, X_{B_l}) = \operatorname{Cov}(X_{A_k}, X_{A_l}) + pq \tag{5.5}$$

for some given $1 \leqslant k < l \leqslant d$, and

$$\operatorname{Cov}(X_{B_i}, X_{B_j}) = \operatorname{Cov}(X_{A_i}, X_{A_j}), \quad 1 \leqslant i \leqslant j \leqslant d, \ (i,j) \neq (k,l). \tag{5.6}$$

By the Möbius inversion formula (5.2), there is a unique way (up to a permutation of $\{1, \dots, n\}$) to choose $\big(A(e_{i,j})\big)_{1 \leqslant i \leqslant j \leqslant d}$ and $\big(B(e_{i,j})\big)_{1 \leqslant i \leqslant j \leqslant d}$ satisfying (5.5) and (5.6), respectively. In this case, (3.1) shows that

$$pq|B(e_{i,j})| = \sum_{\substack{1 \leqslant x \leqslant y \leqslant d \\ e_{i,j} \preceq e_{x,y}}} \mu(e_{x,y}, e_{i,j}) \operatorname{Cov}(X_{B_x}, X_{B_y})$$

$$= pq\mathbf{1}_{\{e_{i,j}\preceq e_{k,l}\}}\mu(e_{k,l}, e_{i,j}) + \sum_{\substack{1\leqslant x\leqslant y\leqslant d\\ e_{i,j}\preceq e_{x,y}}} \mu(e_{x,y}, e_{i,j})\operatorname{Cov}(X_{A_x}, X_{A_y})$$

$$= pq\mathbf{1}_{\{e_{i,j}\preceq e_{k,l}\}}\mu(e_{k,l}, e_{i,j}) + pq|A(e_{i,j})|, \quad 1 \leqslant i \leqslant j \leqslant d,$$

i.e.,

$$|B(e_{i,j})| = \mathbf{1}_{\{e_{i,j}\preceq e_{k,l}\}}\mu(e_{k,l}, e_{i,j}) + |A(e_{i,j})|, \quad 1 \leqslant i \leqslant j \leqslant d. \tag{5.7}$$

Given the children $e_{k,l}\backslash\{k\}$, $e_{k,l}\backslash\{l\} \in \{0,1\}^d$ and grandchild $e_{k,l}\backslash\{k,l\}$ of $e_{k,l} \in \{0,1\}^d$, by (3.3a), (3.3b) and (5.7), we have

$$\begin{cases} |B(e_{k,l})| = |A(e_{k,l})| + 1, \\ |B(e_{k,l}\backslash\{k\})| = |A(e_{k,l}\backslash\{k\})| - 1, \\ |B(e_{k,l}\backslash\{l\})| = |A(e_{k,l}\backslash\{l\})| - 1, \\ |B(e_{k,l}\backslash\{k,l\})| = |A(e_{k,l}\backslash\{k,l\})| + 1, \end{cases} \tag{5.8}$$

with $|B(e_{i,j})| = |A(e_{i,j})|$, since $\mu(e_{k,l}, e_{i,j}) = 0$, in all other cases. We choose to realize the above as

$$\begin{cases} A(e_{k,l}) = B(e_{k,l}) \setminus \{k\}, \\ B(e_{k,l}\backslash\{k\}) = A(e_{k,l}\backslash\{k\}) \setminus \{k\}, \\ B(e_{k,l}\backslash\{l\}) = A(e_{k,l}\backslash\{l\}) \setminus \{l\}, \\ A(e_{k,l}\backslash\{k,l\}) = B(e_{k,l}\backslash\{k,l\}) \setminus \{l\} \end{cases} \tag{5.9}$$

for some given $1 \leqslant k < l \leqslant d$, with $k, l \notin B(e_{i,j}) = A(e_{i,j})$ in all other cases. Noting that

$$l \in B(e_{k,l}\backslash\{k,l\}), \quad k \in A(e_{k,l}\backslash\{k\}), \quad l \in A(e_{k,l}\backslash\{l\}),$$

and

$$B(e_{k,l}\backslash\{k,l\}) \cap B_k = \emptyset, \quad B(e_{k,l}\backslash\{k,l\}) \cap B_l = \emptyset,$$

$$A(e_{k,l}\backslash\{k\}) \cap A_k = \emptyset, \quad A(e_{k,l}\backslash\{l\}) \cap A_l = \emptyset,$$

we find that

$$l \notin B_k, \quad l \notin B_l, \quad k \notin A_k, \quad l \notin A_l.$$

Hence, using the symmetric difference operator $A \setminus B := A \cap B^c$, for $i = 1, \ldots, d$ we have

(5.10)
$$A_i = \begin{cases} \big(B_k \setminus B(e_{k,l}) \setminus B(e_{k,l}\backslash\{k,l\})\big) \cup A(e_{k,l}) \cup \{l\}, & i = k, \\ \big(B_i \setminus B(e_{k,l}) \setminus B(e_{k,l}\backslash\{k,l\})\big) \cup A(e_{k,l}) \cup \{k\} \cup A(e_{k,l}\backslash\{k,l\}) \cup \{l\}, & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i \notin \{k,l\}, & \\ \big(B_l \setminus B(e_{k,l}) \setminus B(e_{k,l}\backslash\{k,l\})\big) \cup A(e_{k,l}) \cup \{k\}, & i = l, \end{cases}$$

and

(5.11)
$$B_i = \begin{cases} \big(B_k \setminus B(e_{k,l}) \setminus B(e_{k,l}\backslash\{k,l\})\big) \cup B(e_{k,l}), & i = k, \\ \big(B_i \setminus B(e_{k,l}) \setminus B(e_{k,l}\backslash\{k,l\})\big) \cup B(e_{k,l}) \cup B(e_{k,l}\backslash\{k,l\}), & i \notin \{k,l\}, \\ \big(B_l \setminus B(e_{k,l}) \setminus B(e_{k,l}\backslash\{k,l\})\big) \cup B(e_{k,l}), & i = l. \end{cases}$$

In other words, by (5.9) we can write

(5.12)
$$\begin{cases} X_{B(e_{k,l})} = X_{A(e_{k,l})} + U, \\ X_{A(e_{k,l}\backslash\{k\})} = X_{B(e_{k,l}\backslash\{k\})} + U, \\ X_{A(e_{k,l}\backslash\{l\})} = X_{B(e_{k,l}\backslash\{l\})} + V, \\ X_{B(e_{k,l}\backslash\{k,l\})} = X_{A(e_{k,l}\backslash\{k,l\})} + V, \end{cases}$$

where $U, V \in \{Z_1, \ldots, Z_n\}$ are two independent Bernoulli random variables, while we have $X_{B(e_{i,j})} = X_{A(e_{i,j})}$ in all other cases, and from (5.10) and (5.11) we get

(5.13)
$$X_{A_i} = \begin{cases} X_{B_k\setminus B(e_{k,l})\setminus B(e_{k,l}\backslash\{k,l\})} + X_{A(e_{k,l})} + V, & i = k, \\ X_{B_i\setminus B(e_{k,l})\setminus B(e_{k,l}\backslash\{k,l\})} + X_{A(e_{k,l})} + U + X_{A(e_{k,l}\backslash\{k,l\})} + V, & i \notin \{k,l\}, \\ X_{B_l\setminus B(e_{k,l})\setminus B(e_{k,l}\backslash\{k,l\})} + X_{A(e_{k,l})} + U, & i = l, \end{cases}$$

and

(5.14)
$$X_{B_i} = \begin{cases} X_{B_k\setminus B(e_{k,l})\setminus B(e_{k,l}\backslash\{k,l\})} + X_{B(e_{k,l})}, & i = k, \\ X_{B_i\setminus B(e_{k,l})\setminus B(e_{k,l}\backslash\{k,l\})} + X_{B(e_{k,l})} + X_{B(e_{k,l}\backslash\{k,l\})}, & i \notin \{k,l\}, \\ X_{B_l\setminus B(e_{k,l})\setminus B(e_{k,l}\backslash\{k,l\})} + X_{B(e_{k,l})}, & i = l. \end{cases}$$

Now, for any supermodular function $\phi : \mathbb{R}^d \to \mathbb{R}$ we have, using formulas (5.14) and (5.12),

$$
\begin{aligned}
&E\big[\phi\big((X_{B_i})_{1\leqslant i\leqslant d}\big)\big]\\
&= E\big[\phi\big((X_{B_i\setminus B(e_{k,l})\setminus B(e_{k,l}\setminus\{k,l\})} + X_{B(e_{k,l})} + X_{B(e_{k,l}\setminus\{k,l\})}\mathbf{1}_{\{i\notin\{k,l\}\}})_{1\leqslant i\leqslant d}\big)\big]\\
&= E\Big[\phi\Big(\big(X_{B_i\setminus B(e_{k,l})\setminus B(e_{k,l}\setminus\{k,l\})} + X_{A(e_{k,l})} + U\\
&\quad + (X_{A(e_{k,l}\setminus\{k,l\})} + V)\mathbf{1}_{\{i\notin\{k,l\}\}}\big)_{1\leqslant i\leqslant d}\Big)\Big]\\
&\geqslant E\big[\phi\big((X_{B_i\setminus B(e_{k,l})\setminus B(e_{k,l}\setminus\{k,l\})} + X_{A(e_{k,l})} + U\mathbf{1}_{\{i\neq k\}}\\
&\quad + X_{A(e_{k,l}\setminus\{k,l\})}\mathbf{1}_{\{i\notin\{k,l\}\}} + V\mathbf{1}_{\{i\neq l\}})_{1\leqslant i\leqslant d}\big)\big]\\
&= E\big[\phi\big((X_{A_i})_{1\leqslant i\leqslant d}\big)\big],
\end{aligned}
$$

where we used (5.13) for the last equality. As for the inequality above, it follows from

$$
\begin{aligned}
&E[\phi(U, U+V, \dots, U+V, U)]\\
&\qquad = p^2\phi\,(1,2,\dots,2,1) + q^2\phi\,(0,0,\dots,0,0) + pq\phi\,(1,1,\dots,1,1)\\
&\qquad\quad + pq\phi\,(0,1,\dots,1,0)\\
&\qquad \geqslant p^2\phi\,(1,2,\dots,2,1) + q^2\phi\,(0,0,\dots,0,0) + pq\phi\,(1,1,\dots,1,0)\\
&\qquad\quad + pq\phi\,(0,1,\dots,1,1)\\
&\qquad = E\,[\phi\,(U, U+V, \dots, U+V, V)]
\end{aligned}
$$

for all supermodular functions $\phi : \mathbb{R}^{|e_{k,l}|} \to \mathbb{R}$, where $|e_{k,l}|$ denotes the cardinality of $e_{k,l}$ whose indices are arranged as $\{k, \dots, l\}$ for convenience of notation, and we did not consider indices $j \notin e_{k,l}$, as U and V do not belong to X_j in this case. ∎

5.1. Multivariate Gaussian vectors. From the central limit theorem, Theorem 5.1 can be used to deal with centered multivariate Gaussian random vectors $(X_1, \dots, X_d)$ and $(Y_1, \dots, Y_d)$ represented as in Example 3.4 as

$$
X = \sum_{1\leqslant k\leqslant l\leqslant d} X_{k,l}e_{k,l}, \qquad Y = \sum_{1\leqslant k\leqslant l\leqslant d} Y_{k,l}e_{k,l}, \tag{5.15}
$$

where $(e_{k,l})_{1\leqslant k\leqslant l\leqslant d} \subset \{0,1\}^d$ satisfies Hypothesis (H). In this case we can apply the Möbius inversion (3.1) to determine the variance coefficients

$$
(\sigma^2_{k,l})_{1\leqslant k\leqslant l\leqslant d} = (\mathrm{Var}\,[X_{k,l}])_{1\leqslant k\leqslant l\leqslant d} \quad \text{and} \quad (\eta^2_{k,l})_{1\leqslant k\leqslant l\leqslant d} = (\mathrm{Var}\,[Y_{k,l}])_{1\leqslant k\leqslant l\leqslant d}
$$

in the decomposition (5.15). Those coefficients can be obtained as the respective limits of normalized variances $(\mathrm{Var}\,[X^n_{k,l}]/n)_{1\leqslant k\leqslant l\leqslant d}$ and $(\mathrm{Var}\,[Y^n_{k,l}]/n)_{1\leqslant k\leqslant l\leqslant d}$ of independent binomial random variables $(X^n_{k,l})_{1\leqslant k\leqslant l\leqslant d}$ and $(Y^n_{k,l})_{1\leqslant k\leqslant l\leqslant d}$. In this case, the sequences $(X^n_1, \dots, X^n_d)_{n\geqslant 1}$ and $(Y^n_1, \dots, Y^n_d)_{n\geqslant 1}$ of independent

random vectors defined by

$$X_i^n := \frac{1}{\sqrt{n}} \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}}} (X_{k,l}^n - E[X_{k,l}^n])$$

and

$$Y_i^n := \frac{1}{\sqrt{n}} \sum_{\substack{1 \leqslant k \leqslant l \leqslant d \\ e_i \preceq e_{k,l}}} (Y_{k,l}^n - E[Y_{k,l}^n]), \quad i = 1, \ldots, d,$$

converge in distribution to the multivariate Gaussian vectors $(X_1, \ldots, X_d)$ and $(Y_1, \ldots, Y_d)$, respectively. The condition $\mathrm{Cov}\,(X_i, X_j) \leqslant \mathrm{Cov}\,(Y_i, Y_j)$ shows that $\mathrm{Cov}\,(X_i^n, X_j^n) \leqslant \mathrm{Cov}\,(Y_i^n, Y_j^n)$ for n sufficiently large, $1 \leqslant i < j \leqslant d$, so by Theorem 5.1 it becomes necessary and sufficient for $(X_1, \ldots, X_d) \leqslant_{\mathrm{sm}} (Y_1, \ldots, Y_d)$ to hold. This is consistent with the general result proved for all multivariate Gaussian random vectors in [10], Theorem 4.2, cf. also Theorem 3.13.5 of [11].

A similar limiting argument can be applied to recover Theorem 4.1 in the Poisson case from Theorem 5.1 and the convergence in distribution of renormalized binomial random variables to Poisson random variables.

5.2. Sums of binomial, Gaussian and Poisson vectors. By Theorem 4.2 of [10] on Gaussian random vectors, Theorems 5.1 and 4.1 above, and the fact that the supermodular ordering is closed under convolution, cf. Theorem 3.9.14-(C) of [11], we deduce that the supermodular ordering of a sum of independent binomial, Gaussian and Poisson vectors is implied by the componentwise ordering of their respective covariances. Proposition 4.1 admits an analog extension to sums of binomial, Gaussian and Poisson random vectors.

REFERENCES

[1] N. Bäuerle, A. Blatter, and A. Müller, *Dependence properties and comparison results for Lévy processes*, Math. Methods Oper. Res. 67 (1) (2008), pp. 161–186.

[2] M. Denuit, E. Frostig, and B. Levikson, *Supermodular comparison of time-to-ruin random vectors*, Methodol. Comput. Appl. Probab. 9 (1) (2007), pp. 41–54.

[3] E. Frostig, *Ordering ruin probabilities for dependent claim streams*, Insurance Math. Econom. 32 (1) (2003), pp. 93–114.

[4] E. Frostig and F. Pellerey, *Supermodular comparison of dependence models and multivariate processes, with applications*, in: *Recent Advances in Probability and Statistics*, Lect. Notes Semin. Interdiscip. Mat., Vol. 12, Potenza, 2015, pp 125–138.

[5] T. Hu, C. Xie, and L. Ruan, *Dependence structures of multivariate Bernoulli random vectors*, J. Multivariate Anal. 94 (1) (2005), pp. 172–195.

[6] K. Kawamura, *The structure of multivariate Poisson distribution*, Kodai Math. J. 2 (3) (1979), pp. 337–345.

[7] B. Kızıldemir and N. Privault, *Supermodular ordering of Poisson arrays*, Statist. Probab. Lett. 98 (2015), pp. 136–143.

[8] M. A. Meyer and B. Strulovici, *The supermodular stochastic ordering*, CEPR Discussion Paper DP9486, May 2013. http://www.cepr.org/pubs/dps/DP9486.

[9] A. Müller, *Stop-loss order for portfolios of dependent risks*, Insurance Math. Econom. 21 (3) (1997), pp. 219–223.
[10] A. Müller and M. Scarsini, *Some remarks on the supermodular order*, J. Multivariate Anal. 73 (1) (2000), pp. 107–119.
[11] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester 2002.
[12] G. Peccati and M. S. Taqqu, *Wiener Chaos: Moments, Cumulants and Diagrams. A Survey with Computer Implementation. Supplementary Material Available Online*, Springer, Milan 2011.
[13] G. Rota, *On the foundations of combinatorial theory: I. Theory of Möbius functions*, Z. Wahrsch. Verw. Gebiete 2 (1964), pp. 340–368.

Bünyamin Kızıldemir
Division of Mathematical Sciences
School of Physical and Mathematical Sciences
Nanyang Technological University
637371 Singapore
E-mail: bunyamin001@e.ntu.edu.sg

Nicolas Privault
Division of Mathematical Sciences
School of Physical and Mathematical Sciences
Nanyang Technological University
637371 Singapore
E-mail: nprivault@ntu.edu.sg

Received on 21.11.2016;
revised version on 5.5.2017

Probability and Mathematical Statistics 38, z. 2, 2018

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 407–428
doi:10.19195/0208-4147.38.2.8

ON THE LONGEST RUNS IN MARKOV CHAINS

BY

ZHENXIA LIU (LINKÖPING) AND XIANGFENG YANG (LINKÖPING)

Abstract. In the first n steps of a two-state (success and failure) Markov chain, the longest success run $L(n)$ has been attracting considerable attention due to its various applications. In this paper, we study $L(n)$ in terms of its two closely connected properties: moment generating function and large deviations. This study generalizes several existing results in the literature, and also finds an application in statistical inference. Our method on the moment generating function is based on a global estimate of the cumulative distribution function of $L(n)$ proposed in this paper, and the proofs of the large deviations include the Gärtner–Ellis theorem and the moment generating function.

2010 AMS Mathematics Subject Classification: Primary: 60F10, 44A1; Secondary: 60J10, 60G70.

Key words and phrases: Longest run, moment generating function, large deviation principle, Markov chain.

1. INTRODUCTION

Let $\{X_k\}_{k\geqslant 1}$ be a time-homogeneous two-state (success and failure) Markov chain. We assume that the initial distribution is $\mathbb{P}(X_1 = 0) = p_0$ and $\mathbb{P}(X_1 = 1) = p_1 = 1 - p_0$, with '1' and '0' denoting the 'success' and 'failure', respectively. The transition matrix of $\{X_k\}_{k\geqslant 1}$ is written as

$$T = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}.$$

To avoid triviality, it is assumed throughout the paper that $0 < p_0 < 1$ and $0 < p_{ij} < 1$ for $i, j = 0, 1$, which indicates that the Markov chain is ergodic. In the first n steps of the Markov chain, the longest success run $L(n)$, namely the longest stretch of consecutive successes, has been attracting considerable attention due to its applications in various fields, such as reliability and statistics (cf. [1]). We refer to [4] and [5] for the first few seminal works in the 1970s, and [8]–[11] for the latest progress.

Among various studies on the longest success run $L(n)$, the probability estimating of $L(n)$ for large n (such as large deviations) is an important topic. Part of the reason is that the exact distribution of $L(n)$ (cf. [7]) is intricate despite known explicit formulas, which gives no information as n approaches infinity. Even in the identically independent case (that is, $\{X_k\}_{k\geqslant 1}$ are independent and identically distributed), there is much complexity of the exact distribution of $L(n)$ which can be seen (for instance cf. [8]) as follows:

$$\mathbb{P}(L(n)<k)=\sum_{r=0}^{[\frac{n+1}{k+1}]}(-1)^r p_1^{rk}p_0^{r-1}\left[\binom{n-rk}{r-1}+p_0\binom{n-rk}{r}\right],$$

where $[\cdot]$ denotes the integer part of a constant. One topic of this paper is to study the large deviations of $L(n)$ in a Markov chain $\{X_k\}_{k\geqslant 1}$ defined above. To appropriately propose such deviations, recall a law of large numbers (cf. e.g. [14]):

$$\frac{L(n)}{\log_{1/p_{11}} n}\to 1 \quad \text{in probability as } n\to\infty.$$

Such a limit in independent trails is a well-known result (cf. [4], [5], [12]). This suggests to study the large deviation probabilities in the form $\mathbb{P}\big(L(n)/\log_{1/p_{11}} n\in A\big)$, where the set A does not include the most probable point 1. Our first result is formulated as follows.

THEOREM 1.1. *For each $x>0$, we have*

$$\lim_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{P}\bigg(\frac{L(n)}{\log_{1/p_{11}} n}\geqslant 1+x\bigg)=-x\cdot\ln(1/p_{11}). \tag{1.1}$$

For each $0<x<1$, we have

$$\lim_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\bigg[-\ln\mathbb{P}\bigg(\frac{L(n)}{\log_{1/p_{11}} n}\leqslant 1-x\bigg)\bigg]=x\cdot\ln(1/p_{11}). \tag{1.2}$$

Theorem 1.1 tells that the probability $\mathbb{P}\big(L(n)/\log_{1/p_{11}} n\geqslant 1+x\big)$ decays in a power rate, while the probability $\mathbb{P}\big(L(n)/\log_{1/p_{11}} n\leqslant 1-x\big)$ decays exponentially fast. If $\{X_k\}_{k\geqslant 1}$ is a sequence of identically independent trails, namely $p_{00}=p_{10}=p_0$ and $p_{01}=p_{11}=p_1$, then the limits (1.1) and (1.2) trivially hold because of a well global estimate (cf. [7] and [9]): for $k=1,\dots,n$,

$$(1-p_1^k)^{n-k+1}\leqslant\mathbb{P}\big(L(n)<k\big)\leqslant(1-p_0p_1^k)^{n-k+1}. \tag{1.3}$$

Due to the lack of satisfactory estimates as above (namely (1.3)) for general Markov chains $\{X_k\}_{k\geqslant 1}$, the proof of Theorem 1.1 will be based on a less precise global estimate proposed below (see Lemma 2.1) in this paper. Here we note that essentially the same large deviation probability as (1.1) was claimed to be proved in [14]

in the form: for all $x > 0$,

(1.4)
$$\lim_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{P}\big(L(n) - \lfloor \log_{1/p_{11}} n \rfloor \geqslant x \cdot \log_{1/p_{11}} n\big) = -x \cdot \ln(1/p_{11}).$$

Unfortunately, the proof of (1.4) therein contains a mistake stemming from the employed (Stein–Chen) method, which seems to be impossible to be corrected in principle. Section 4 includes detailed explanations on this aspect.

A natural generalization of the limit (1.1) (not (1.2)) is a *large deviation principle* for the family of random variables $L(n)/\log_{1/p_{11}} n$. For identically independent trails $\{X_k\}_{k\geqslant 1}$, large deviation principles were recently derived in [9] based on (1.3). There are also related discussions on the large deviations of $L(n)$ in [7] and [11]. The second result of this paper is to establish a large deviation principle for $L(n)$, which includes (1.1) (or (1.4)) as a special case. To this end, we define a function $\Lambda^*(x)$ as

$$\Lambda^*(x) = \begin{cases} +\infty, & x < 1, \\ (x-1)\ln(1/p_{11}), & x \geqslant 1. \end{cases} \tag{1.5}$$

THEOREM 1.2. *The normalized longest success run $L(n)/\log_{1/p_{11}} n$ satisfies a large deviation principle with a good rate function $\Lambda^*(x)$ given by* (1.5) *and a speed $\log_{1/p_{11}} n$. Namely*,

(i) *for any open set $O \subseteq \mathbb{R}$,*

$$\liminf_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{P}\left(\frac{L(n)}{\log_{1/p_{11}} n} \in O\right) \geqslant -\inf_{x\in O} \Lambda^*(x); \tag{1.6}$$

(ii) *for any closed set $F \subseteq \mathbb{R}$,*

$$\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{P}\left(\frac{L(n)}{\log_{1/p_{11}} n} \in F\right) \leqslant -\inf_{x\in F} \Lambda^*(x). \tag{1.7}$$

It is clear that the special case (1.1) (or (1.4)) comes from Theorem 1.2 with an open set $O = (1+x, \infty)$ and a closed set $F = [1+x, \infty)$. The proof of Theorem 1.2 is given in Section 3.2.

The large deviation principle in Theorem 1.2 is *non-trivial* since the rate function $\Lambda^*(x)$ is not always *zero* or *infinity*. Now an interesting question arises: besides the family of random variables $L(n)/\log_{1/p_{11}} n$, are there other families which admit non-trivial large deviation principles? Note that large deviation principles have very close connections with the corresponding Laplace transforms (or the moment generating functions; see the Gärtner–Ellis theorem [3]), thus the above question leads to the third result of this paper: precise logarithmic asymptotics for the moment generating function of $L(n)$ as formulated in the following theorem, based on

which there are (only) two families which admit non-trivial large deviation principles: $\{L(n)/\log_{1/p_{11}} n\}$ and $\{L(n)/n\}$. Throughout the paper, $a(n) \sim b(n)$ as $n \to \infty$ stands for $\lim_{n\to\infty} a(n)/b(n) = 1$.

THEOREM 1.3. *The moment generating function of $L(n)$ has the following logarithmic asymptotics*:

(i) *for* $\lambda < \ln(1/p_{11})$,

$$\ln \mathbb{E}e^{\lambda L(n)} \sim \lambda \log_{1/p_{11}} n;$$

(ii) *for* $\lambda = \ln(1/p_{11})$,

$$\ln \mathbb{E}e^{\lambda L(n)} \sim 2\lambda \log_{1/p_{11}} n;$$

(iii) *for* $\lambda > \ln(1/p_{11})$,

$$\begin{aligned}\lambda - \ln(1/p_{11}) &\leqslant \liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{E}e^{\lambda L(n)} \leqslant \limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{E}e^{\lambda L(n)} \\ &\leqslant \max\left\{\lambda - \ln(1/p_{11}),\ \lambda - \ln \frac{1}{|p_{00} - p_{10}|}\right\},\end{aligned}$$

and, *in particular*, *if* $p_{10} \leqslant p_{00} + p_{11}$, *then*

$$\ln \mathbb{E}e^{\lambda L(n)} \sim \lambda - \ln(1/p_{11}).$$

Similar results for the identically independent case have been recently proved in [9], where the condition $p_{10} \leqslant p_{00} + p_{11}$ is automatically fulfilled. Technically speaking, the condition $p_{10} \leqslant p_{00} + p_{11}$ is due to an extra error term $e(n)$ in Lemma 2.2 below. In terms of the structure of the Markov chain, this condition means that the transition probability p_{10} from the state '1' to the state '0' should not exceed the probability that the chain stays still, which is $p_{00} + p_{11}$. Although we think that such a condition can be removed by using a more precise estimate than the one in Lemma 2.2, the current method in this paper cannot get rid of this condition.

Several new difficulties arise in the proof of Theorem 1.3 due to the lack of satisfactory global estimates of the cumulative distribution function of $L(n)$, and we overcome them using suitable non-global estimates included in Section 2.2. To see how Theorem 1.3 yields non-trivial large deviation principles, we first consider the logarithmic moment generating function of $L(n)/\log_{1/p_{11}} n$ (according to (i) and (ii) of Theorem 1.3) defined as $\Lambda_n(\lambda) = \ln \mathbb{E} \exp\{\lambda \cdot L(n)/\log_{1/p_{11}} n\}$ for $\lambda \in \mathbb{R}$, and the *cumulant* defined as $\Lambda(\lambda) := \lim_{n\to\infty} \Lambda_n(\lambda \cdot \log_{1/p_{11}} n)/\log_{1/p_{11}} n$. Then the Gärtner–Ellis theorem (cf. [3], Section 2.3) suggests that there is a non-trivial large deviation principle for the family $L(n)/\log_{1/p_{11}} n$ with a rate function Λ^* defined via the Fenchel–Legendre transform of Λ: $\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}[\lambda \cdot x - \Lambda(\lambda)]$.

This is verified in detail in Theorem 1.2. Now, according to (iii) of Theorem 1.3, we can also consider the logarithmic moment generating function of $L(n)/n$ as $\widetilde{\Lambda}_n(\lambda) = \ln \mathbb{E} \exp\{\lambda \cdot L(n)/n\}$ for any $\lambda \in \mathbb{R}$, and obtain the cumulant, under the condition $p_{10} \leqslant p_{00} + p_{11}$, in the form

$$\widetilde{\Lambda}(\lambda) := \lim_{n\to\infty} \frac{1}{n}\widetilde{\Lambda}_n(\lambda \cdot n) = \begin{cases} \lambda - \ln(1/p_{11}), & \lambda \geqslant \ln(1/p_{11}), \\ 0, & \lambda < \ln(1/p_{11}). \end{cases}$$

The Gärtner–Ellis theorem again suggests that there is a non-trivial large deviation principle for the family $L(n)/n$ with a rate function $\widetilde{\Lambda}^*(x)$ defined as the Fenchel–Legendre transform of $\widetilde{\Lambda}(\lambda)$:

$$\widetilde{\Lambda}^*(x) = \begin{cases} +\infty, & x < 0, \\ x \ln(1/p_{11}), & 0 \leqslant x \leqslant 1, \\ +\infty, & x > 1. \end{cases} \tag{1.8}$$

This large deviation principle for the family $\{L(n)/n\}$ corresponds to the law of large numbers $L(n)/n \to 0$ which is directly from $L(n)/\log_{1/p_{11}} n \to 1$. We formulate this observation as our last result in the following theorem.

THEOREM 1.4. *If $p_{10} \leqslant p_{00} + p_{11}$, then the normalized longest success run $L(n)/n$ satisfies a large deviation principle with a good rate function $\widetilde{\Lambda}^*(x)$ given by* (1.8) *and a speed n. Namely*,

(i) *for any open set $O \subseteq \mathbb{R}$,*

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left(\frac{L(n)}{n} \in O\right) \geqslant -\inf_{x\in O} \widetilde{\Lambda}^*(x); \tag{1.9}$$

(ii) *for any closed set $F \subseteq \mathbb{R}$,*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left(\frac{L(n)}{n} \in F\right) \leqslant -\inf_{x\in F} \widetilde{\Lambda}^*(x). \tag{1.10}$$

Here we draw the reader's attention that the Gärtner–Ellis theorem will be used to prove the aforementioned two large deviation principles. It should be noted that there are other methods to achieve such large deviation principles, such as the Bryc's Inverse Varadhan Lemma (cf. Section 4.4 in [3]). In [10] the Bryc's Inverse Varadhan Lemma was used to obtain a large deviation principle for $L(n)$ with a general speed in the identically independent case.

The rest of the paper is organized as follows. Section 2 includes global and non-global estimates of the cumulative distribution function of $L(n)$ which will be used throughout the paper. In the first part of Section 3, we give the proof of the main result of the paper: the precise logarithmic asymptotics for the moment

generating function (Theorem 1.3). Then we show that the two large deviation principles (Theorems 1.2 and 1.4) follow from Theorem 1.3 and the Gärtner–Ellis theorem, which is included in the second part of Section 3. The last part of Section 3 contains a very concise proof of Theorem 1.1. The use of the Stein–Chen method in estimating the large deviation probabilities of $L(n)$ is briefly described in Section 4, where a mistake of proving (1.4) in [14] is pointed out. Finally, an application of the derived results to statistical inference is presented in Section 5.

2. ESTIMATES OF THE DISTRIBUTION FUNCTION

In this section, we first propose a global estimate for the cumulative distribution function of $L(n)$ which will be used throughout the paper. Then we present several special non-global estimates which have more explicit forms.

2.1. Global estimate.

LEMMA 2.1. *For all $k = 1, \dots, n$, we have*

(2.1)
$$1 - p_{11}^{k-1}\,[c_1 \cdot (n-k) + c_2] - c(n,k) \leqslant \mathbb{P}\big(L(n) < k\big) \leqslant (1 - c_3 \cdot p_{11}^{k-1})^{n-k+1},$$

where

$$c_1 = \frac{p_{01}p_{10}}{p_{01}+p_{10}} > 0, \quad c_2 = \frac{p_{01}(p_0 p_{01} - p_1 p_{10})}{(p_{01}+p_{10})^2},$$

$$c_3 = \min\left\{p_1,\ \frac{c_1 + (p_{01}+p_{10}) \cdot \min\{0,\ c_2,\ c_2(p_{00}-p_{10}),\ c_2(p_{00}-p_{10})^2\}}{1 + c_1/p_{01} + |c_2|(p_{01}+p_{10})/p_{01}}\right\}$$

$(c_3 > 0)$, *and*

$$c(n,k) = \frac{c_1 p_{11}^{k-1}}{1-p_{11}} - \frac{c_2(p_{01}+p_{10})}{p_{01}}(p_{00}-p_{10})^{n-1} - \frac{c_2(p_{01}+p_{10})}{p_{11}} \cdot \frac{(p_{00}-p_{10})^n - p_{11}^{k-1}(p_{00}-p_{10})^{n-k}}{p_{00}-p_{10}-p_{11}} > 0.$$

P r o o f. We first note that the exact distribution of $L(n)$ has been known (cf. [7]), but it hardly helps to gain useful information on the asymptotics as $n \to \infty$. The proof of Lemma 2.1 is based on a newly built Markov chain $\{\eta_k\}_{1 \leqslant k \leqslant n}$, where η_k is defined as the length of success runs at the end of the k-th step, namely

$$\{\eta_k = i\} \text{ is equivalent to } \{X_k = 1, \dots, X_{k-i+1} = 1, X_{k-i} = 0\}.$$

In this setting, the longest success run $L(n) = \max_{1 \leqslant k \leqslant n} \eta_k$. This enables us to estimate $\mathbb{P}\big(L(n) < k\big)$ a little more explicitly, using the probabilities involving η_k. This idea was introduced in [6], where the derived results are

$$\mathbb{P}\big(L(n) < k\big) \geqslant 1 - p_{01}p_{11}^{k-1} \sum_{i=k}^{n-1} b(i-k) - c(n,k) \tag{2.2}$$

and

(2.3)
$$\mathbb{P}\big(L(n) < k\big) \leqslant (1 - p_1 p_{11}^{k-1}) \prod_{i=k+1}^{n} \left(1 - \frac{p_{01} p_{11}^{k-1} b(i-k)}{b(i-1) + p_{01} \sum\limits_{j=1}^{k-1} p_{11}^{j-1} b(i-j-1)} \right)$$

with
$$b(i) = p_0 (p_{00} - p_{10})^{i-1} + \frac{p_{10}\big(1 - (p_{00} - p_{10})^{i-1}\big)}{1 - p_{00} + p_{10}}.$$

To achieve the upper bound in (2.1) from (2.3), we note that $p_1 \geqslant c_3$, and
$$\frac{p_{01} b(i-k)}{b(i-1) + p_{01} \sum\limits_{j=1}^{k-1} p_{11}^{j-1} b(i-j-1)} \geqslant \frac{p_{01} \min_j b(j)}{\max_j b(j) + 1},$$

since $p_{01} \sum_{j=1}^{k-1} p_{11}^{j-1} b(i-j-1) = \mathbb{P}(\eta_i = 0, 1, \ldots, k-1) \leqslant 1$. To estimate two quantities $\min_j b(j)$ and $\max_j b(j)$, we rewrite $b(j)$ as

$$b(j) = \alpha + \beta \cdot (p_{00} - p_{10})^{j-1}, \quad \text{where } \alpha = \frac{p_{10}}{p_{01} + p_{10}} \text{ and } \beta = \frac{p_0 p_{01} - p_1 p_{10}}{p_{01} + p_{10}}.$$

It then follows that
$$\max_j b(j) \leqslant \alpha + |\beta|,$$

and
$$\min_j b(j) \geqslant \min\{\alpha, \alpha + \beta, \alpha + \beta(p_{00} - p_{10}), \alpha + \beta(p_{00} - p_{10})^2\}.$$

Therefore,
$$\frac{p_{01} \min_j b(j)}{\max_j b(j) + 1} \geqslant c_3,$$

which implies the upper bound in (2.1).

To obtain the lower bound in (2.1) from (2.2), we see that the sum in (2.2) is

$$\begin{aligned} \sum_{i=k}^{n-1} b(i-k) &= (n-k)\alpha + \beta \cdot \frac{1 - (p_{00} - p_{10})^{n-k}}{1 - (p_{00} - p_{10})} \\ &\leqslant (n-k)\alpha + \beta \cdot \frac{1}{p_{01} + p_{10}}, \end{aligned}$$

which gives the lower bound. To see the positivity of $c(n,k)$, we note that

$$c(n,k) = \mathbb{P}(\eta_n \in \{k, k+1, \ldots, n\}) > 0. \quad \blacksquare$$

2.2. Non-global estimates. One might be interested in comparing the global estimate (2.1) in Lemma 2.1 with the i.i.d. case (1.3). They actually look alike under suitable conditions, which will be summarized as follows.

LEMMA 2.2. *If* $n > k := k(n) \geqslant 1 + \log_{1/p_{11}}\left(\frac{n(c_1/(1-p_{11})+|c_2|)}{2}\right)$, *then we have*

$$(1 - c_5 \cdot p_{11}^k)^{n-k+1} - e(n) \leqslant \mathbb{P}\big(L(n) < k\big) \leqslant (1 - c_4 \cdot p_{11}^k)^{n-k+1} \tag{2.4}$$

for large n, *where* c_4 *and* c_5 *are two* (*uniform*) *positive constants*, *and* $e(n)$ *is a term which converges to zero exponentially fast as* $n \to \infty$ (*note that* $e(n) = 0$ *when* $p_{00} = p_{10}$).

P r o o f. In (2.4) the claimed upper bound $\mathbb{P}\big(L(n)<k\big) \leqslant \big(1 - c_4 \cdot p_{11}^k\big)^{n-k+1}$ comes directly from the upper bound of (2.1) by setting $c_4 = c_3/p_{11}$, uniformly in k. To achieve the lower bound of (2.4), we first rewrite the lower bound of (2.1) as follows:

$$\begin{aligned}
&\mathbb{P}\big(L(n) < k\big) \geqslant 1 - p_{11}^{k-1}\left[c_1 \cdot (n-k) + c_2\right] - c(n,k)\\
&= 1 - p_{11}^{k-1}\left[c_1 \cdot (n-k) + c_2 + \frac{c_1}{1-p_{11}}\right] + \frac{c_2(p_{01}+p_{10})}{p_{01}}(p_{00}-p_{10})^{n-1}\\
&\quad + \frac{c_2(p_{01}+p_{10})}{p_{11}} \cdot \frac{(p_{00}-p_{10})^n - p_{11}^{k-1}(p_{00}-p_{10})^{n-k}}{p_{00}-p_{10}-p_{11}}\\
&=: 1 - p_{11}^{k-1}\left[c_1 \cdot (n-k) + c_2 + \frac{c_1}{1-p_{11}}\right] + e(n).
\end{aligned}$$

It is clear that the term $e(n)$ converges to zero exponentially fast for all k, and $e(n) = 0$ if $p_{00} = p_{10}$. If we define $c_* = c_1/(1-p_{11}) + |c_2|$, then (with $k < n$)

$$1 - p_{11}^{k-1}\left[c_1 \cdot (n-k) + c_2 + \frac{c_1}{1-p_{11}}\right] \geqslant 1 - p_{11}^{k-1} \cdot c_* \cdot (n-k).$$

In order to estimate $1 - p_{11}^{k-1} \cdot c_* \cdot (n-k)$, we set $N = n-k+1$, $a = p_{11}^{k-1} \cdot c_*$, and obtain

$$(1-a)^N \leqslant 1 - (N-1)a(1-a)^{N-2}\left[\frac{N}{N-1}(1-a) - Na/2\right].$$

Since $n > k(n) \geqslant 1 + \log_{1/p_{11}}\left(\frac{n(c_1/(1-p_{11})+|c_2|)}{2}\right)$ and n is large, a is small. Therefore,

$$(1-a)^{N-2} = [(1-a)^{1/a}]^{a(N-2)} \geqslant [(1-a)^{1/a}]^{c_* \cdot p_{11}^{c-1}} \geqslant (e/2)^{c_* \cdot p_{11}^{c-1}}$$

with $c = 1 - \log_{1/p_{11}}(2/c_*)$,

$$\frac{N}{N-1}(1-a) \geqslant 1+\delta$$

for some small $\delta > 0$, and

$$Na/2 \leqslant 1.$$

In summary, we have

$$(1-a)^N \leqslant 1 - (N-1)a \cdot \delta(e/2)^{c_* \cdot p_{11}^{c-1}},$$

which gives

$$(1 - c_* \cdot p_{11}^{k-1})^{n-k+1} \leqslant 1 - (n-k) \cdot c_* \cdot p_{11}^{k-1} \cdot \delta(e/2)^{c_* \cdot p_{11}^{c-1}}.$$

Replacing c_* by $c_*/\delta(e/2)^{c_* \cdot p_{11}^{c-1}}$ proves the lower bound of (2.4). ∎

In Lemma 2.2, if k is exactly the size $\alpha \cdot \log_{1/p_{11}} n$ with $\alpha > 1$, then we have the following more explicit estimate.

LEMMA 2.3. *If $x > 0$ and $k(n) = [(1+x)\log_{1/p_{11}} n]$, then*

$$c_6 \cdot n^{-(1+x)}\,(n-k) \leqslant \mathbb{P}\big(L(n) > k\big) \leqslant c_7 \cdot n^{-(1+x)}\,(n-k)$$

for large n, where c_6 and c_7 are two (uniform) positive constants.

P r o o f. To see the lower bound, we infer from Lemma 2.2 that

$$\begin{aligned}
\mathbb{P}\big(L(n) > k\big) &= 1 - \mathbb{P}\big(L(n) \leqslant k\big) \\
&\geqslant 1 - (1 - c_4 \cdot p_{11}^{k+1})^{n-k} \\
&= 1 - [(1 - c_4 \cdot p_{11}^{k+1})^{1/(c_4 \cdot p_{11}^{k+1})}]^{c_4 \cdot p_{11}^{k+1}(n-k)} \\
&= -[(1 - c_4 \cdot p_{11}^{k+1})^{1/(c_4 \cdot p_{11}^{k+1})}]^{\theta_n} \cdot \ln\big((1 - c_4 \cdot p_{11}^{k+1})^{1/(c_4 \cdot p_{11}^{k+1})}\big) \\
&\quad \times c_4 \cdot p_{11}^{k+1}(n-k) \\
&\geqslant \text{const} \cdot p_{11}^{k}(n-k) \geqslant \text{const} \cdot n^{-(1+x)}(n-k),
\end{aligned}$$

where $\theta_n \in [0,\ c_4 \cdot p_{11}^{k+1}(n-k)]$. The upper bound can be similarly handled by noticing that

$$e(n) \sim \text{const} \cdot \exp\left\{-n \cdot \ln\ \frac{1}{|p_{00} - p_{10}|}\right\} \leqslant \text{const} \cdot n^{-(1+x)}\,(n-k)\,. \quad ∎$$

The next estimate is the case when k is of size $\alpha \cdot \log_{1/p_{11}} n$ with $\alpha < 1$.

LEMMA 2.4. *If* $0 < x < 1$ *and* $k(n) = [(1-x)\log_{1/p_{11}} n]$, *then*

$$c_8 \cdot n^x \leqslant \ln \mathbb{P}\big(L(n) < k\big) \leqslant c_9 \cdot n^x$$

for large n, *where* c_8 *and* c_9 *are two* (*uniform*) *negative constants.*

P r o o f. With $k(n) = [(1-x)\log_{1/p_{11}} n]$, it follows from Lemma 2.1 that

$$\begin{aligned}&\mathbb{P}\big(L(n) < k\big) \geqslant 1 - p_{11}^{k-1}\,[c_1 \cdot (n-k) + c_2] - c(n,k)\\ &\geqslant 1 - \text{const}_1 \cdot p_{11}^k (n-k) - \text{const}_2 \cdot |p_{00} - p_{10}|^n - \text{const}_3 \cdot p_{11}^k |p_{00} - p_{10}|^{n-k}.\end{aligned}$$

If we apply the inequality $\ln(1-a) \geqslant -2a$ for $0 < a < 1/2$, then

$$\begin{aligned}&\ln \mathbb{P}\big(L(n) < k\big)\\ &\geqslant -2\,\text{const}_1\, p_{11}^k (n-k) - 2\,\text{const}_2 |p_{00} - p_{10}|^n - 2\,\text{const}_3\, p_{11}^k |p_{00} - p_{10}|^{n-k}\\ &\geqslant \text{const} \cdot n^{-x}.\end{aligned}$$

The upper bound is similarly proved with the help of the arguments in the proof of Lemma 2.3. ∎

3. MOMENT GENERATING FUNCTION AND LARGE DEVIATIONS

In this section, we first give a proof of Theorem 1.3 regarding the precise logarithmic asymptotics for the moment generating function, which is the main result of the paper. Then, using this proved result, we derive two large deviation principles (Theorems 1.2 and 1.4) with the help of the Gärtner–Ellis theorem. At the end, a very concise proof of Theorem 1.1 is included.

3.1. Proof of Theorem 1.3.

S t e p 1. The following estimate holds for all $\lambda \in \mathbb{R}$:

$$\liminf_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \geqslant \lambda.$$

The case when $\lambda = 0$ is trivial. If $\lambda > 0$, then

$$\begin{aligned}&\frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\}\\ &\quad\geqslant \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E}\bigg(\exp\{\lambda \cdot L(n)\}, \bigg\{\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| \leqslant \varepsilon\bigg\}\bigg)\\ &\quad\geqslant \frac{1}{\log_{1/p_{11}} n} \ln \exp\{\lambda \cdot (1-\varepsilon)\log_{1/p_{11}} n\} \cdot \mathbb{P}\bigg(\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| \leqslant \varepsilon\bigg)\\ &\quad= \lambda \cdot (1-\varepsilon) + \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{P}\bigg(\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| \leqslant \varepsilon\bigg).\end{aligned}$$

Since $L(n)/\log_{1/p_{11}} n$ converges to one almost surely, we have

$$\lim_{\varepsilon\to 0^+} \liminf_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \geqslant \lim_{\varepsilon\to 0^+} \lambda\cdot(1-\varepsilon) = \lambda.$$

If $\lambda < 0$, a similar argument as above yields

$$\lim_{\varepsilon\to 0^+} \liminf_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \geqslant \lim_{\varepsilon\to 0^+} \lambda\cdot(1+\varepsilon) = \lambda.$$

S t e p 2. The following estimate holds for $\lambda < \ln(1/p_{11})$:

$$\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \leqslant \lambda.$$

To see this, we first rewrite

$$\begin{aligned}
&\ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \\
&= \ln \mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\}, \bigg\{\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| \leqslant \varepsilon\bigg\} \cup \bigg\{\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| > \varepsilon\bigg\}\bigg).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \\
&= \max\bigg\{ \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\}, \bigg\{\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| \leqslant \varepsilon\bigg\}\bigg), \\
&\qquad \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\}, \bigg\{\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| > \varepsilon\bigg\}\bigg)\bigg\}.
\end{aligned} \tag{3.1}$$

It is clear that the first limit satisfies

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\}, \bigg\{\bigg|\frac{L(n)}{\log_{1/p_{11}} n} - 1\bigg| \leqslant \varepsilon\bigg\}\bigg) \\
&\qquad\qquad \leqslant \begin{cases} \lambda(1+\varepsilon), & \lambda > 0, \\ \lambda(1-\varepsilon), & \lambda < 0. \end{cases}
\end{aligned} \tag{3.2}$$

The second limit is more complicated, and the assumption $\lambda < \ln(1/p_{11})$ is needed. We rewrite

$$\ln \mathbb{E}\bigg(\exp\{\lambda \cdot L(n)\}, \bigg\{ \bigg| \frac{L(n)}{\log_{1/p_{11}} n} - 1 \bigg| > \varepsilon \bigg\} \bigg)$$
$$= \ln \mathbb{E}\bigg(\exp\{\lambda \cdot L(n)\}, \bigg\{ \frac{L(n)}{\log_{1/p_{11}} n} - 1 > \varepsilon \bigg\} \cup \bigg\{ \frac{L(n)}{\log_{1/p_{11}} n} - 1 < -\varepsilon \bigg\} \bigg).$$

On the first part $\left\{ \frac{L(n)}{\log_{1/p_{11}} n} - 1 > \varepsilon \right\}$, if $\lambda < 0$, then similar things can be done as above. But if $\lambda > 0$, then we need to make the following separation:

$$\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E}\bigg(\exp\{\lambda \cdot L(n)\}, \bigg\{ \frac{L(n)}{\log_{1/p_{11}} n} - 1 > \varepsilon \bigg\} \bigg)$$
$$= \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n}$$
$$\times \ln \mathbb{E}\bigg(\exp\{\lambda \cdot L(n)\}, \bigcup_{k=1}^{\infty} \bigg\{ 1 + k\varepsilon < \frac{L(n)}{\log_{1/p_{11}} n} \leqslant 1 + (k+1)\varepsilon \bigg\} \bigg)$$
$$\leqslant \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \bigg(\sum_{k=1}^{\infty} e^{\lambda[1+(1+k)\varepsilon] \log_{1/p_{11}} n} \cdot \mathbb{P}\bigg(1 + k\varepsilon < \frac{L(n)}{\log_{1/p_{11}} n} \bigg) \bigg)$$
$$= \lambda(1+\varepsilon) + \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \bigg(\sum_{k=1}^{\infty} e^{\lambda k\varepsilon \log_{1/p_{11}} n} \cdot \mathbb{P}\bigg(1 + k\varepsilon < \frac{L(n)}{\log_{1/p_{11}} n} \bigg) \bigg).$$

It now follows from Lemma 2.3 that

$$\mathbb{P}\bigg(1 + k\varepsilon < \frac{L(n)}{\log_{1/p_{11}} n} \bigg) = 1 - \mathbb{P}\bigg(\frac{L(n)}{\log_{1/p_{11}} n} \leqslant 1 + k\varepsilon \bigg) \leqslant \text{const} \cdot n^{-k\varepsilon},$$

which gives

$$\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E}\bigg(\exp\{\lambda \cdot L(n)\}, \bigg\{ \frac{L(n)}{\log_{1/p_{11}} n} - 1 > \varepsilon \bigg\} \bigg)$$
$$= \lambda(1+\varepsilon)$$
$$+ \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \bigg(\sum_{k=1}^{\infty} e^{\lambda k\varepsilon \log_{1/p_{11}} n} \cdot \mathbb{P}\bigg(1 + k\varepsilon < \frac{L(n)}{\log_{1/p_{11}} n} \bigg) \bigg)$$
$$\leqslant \lambda(1+\varepsilon) + \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \Big(\sum_{k=1}^{\infty} e^{\lambda k\varepsilon \log_{1/p_{11}} n} \cdot n^{-k\varepsilon} \Big)$$
$$= \lambda(1+\varepsilon) + \limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \Big(\sum_{k=1}^{\infty} n^{-(1-\frac{\lambda}{\ln(1/p_{11})})k\varepsilon} \Big)$$
$$\leqslant \lambda(1+\varepsilon),$$

where the last step follows from the fact that $\lambda < \ln(1/p_{11})$. Namely, we have

proved that

(3.3)
$$\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\},\Big\{\frac{L(n)}{\log_{1/p_{11}} n}-1>\varepsilon\Big\}\bigg)\leqslant\lambda(1+\varepsilon).$$

On the second part $\left\{\frac{L(n)}{\log_{1/p_{11}} n}-1<-\varepsilon\right\}$, the case when $\lambda>0$ can be similarly handled. For the case $\lambda<0$, we can do a similar separation to that in the proof of (3.3), but the argument here is a little different. We have

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\},\Big\{\frac{L(n)}{\log_{1/p_{11}} n}-1<-\varepsilon\Big\}\bigg)\\
&=\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\\
&\quad\times\ln\mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\},\bigcup_{k=1}^{[1/\varepsilon]-1}\Big\{1-(k+1)\varepsilon<\frac{L(n)}{\log_{1/p_{11}} n}\leqslant 1-k\varepsilon\Big\}\bigg)\\
&\leqslant\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\\
&\quad\times\ln\bigg(\sum_{k=1}^{[1/\varepsilon]-1}e^{\lambda[1-(k+1)\varepsilon]\log_{1/p_{11}} n}\cdot\mathbb{P}\Big(1-(k+1)\varepsilon<\frac{L(n)}{\log_{1/p_{11}} n}\leqslant 1-k\varepsilon\Big)\bigg).
\end{aligned}$$

Since there are only finite terms in the summation, we can simplify the above quantity, noticing that it is less than or equal to

$$\begin{aligned}
&\max_{1\leqslant k\leqslant[1/\varepsilon]-1}\bigg\{\lambda[1-(k+1)\varepsilon]+\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{P}\Big(\frac{L(n)}{\log_{1/p_{11}} n}<1-k\varepsilon\Big)\bigg\}\\
&\qquad=\max_{1\leqslant k\leqslant[1/\varepsilon]-1}\{\lambda[1-(k+1)\varepsilon]-\infty\}=-\infty,
\end{aligned}$$

where the '$-\infty$' appears because of Lemma 2.4. Therefore,

(3.4)
$$\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\bigg(\exp\{\lambda\cdot L(n)\},\Big\{\frac{L(n)}{\log_{1/p_{11}} n}-1<-\varepsilon\Big\}\bigg)=-\infty.$$

Now the proof is done by taking the estimates (3.2), (3.3) and (3.4) back into (3.1).

S t e p 3. If $\lambda=\ln(1/p_{11})$, then

$$\lim_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\exp\{\lambda\cdot L(n)\}=2\lambda.$$

On the one hand, it follows from Lemma 2.3 that, for every $\varepsilon>0$,

$$\begin{aligned}
&\frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \\
&\geqslant \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\left\{\lambda \cdot L(n), \left\{\frac{L(n)}{\log_{1/p_{11}} n} > 1+\varepsilon\right\}\right\} \\
&= \frac{1}{\log_{1/p_{11}} n} \\
&\quad \times \ln \mathbb{E} \exp\left\{\lambda \cdot L(n), \bigcup_{k=1}^{q} \left\{1+k\varepsilon < \frac{L(n)}{\log_{1/p_{11}} n} \leqslant 1+(k+1)\varepsilon\right\}\right\} =: K,
\end{aligned}$$

where (and in the sequel) we put

$$q = \left[\frac{1}{\varepsilon}\left(\frac{n}{\log_{1/p_{11}} n} - 1\right)\right].$$

Now we have

$$\begin{aligned}
K &\geqslant \frac{1}{\log_{1/p_{11}} n} \ln \sum_{k=1}^{q} \exp\{(1+k\varepsilon)(\log_{1/p_{11}} n) \cdot \ln(1/p_{11})\} \\
&\quad \times \left(\mathbb{P}\left\{\frac{L(n)}{\log_{1/p_{11}} n} > 1+k\varepsilon\right\} - \mathbb{P}\left\{\frac{L(n)}{\log_{1/p_{11}} n} > 1+(k+1)\varepsilon\right\}\right) \\
&\geqslant \ln(1/p_{11}) + \frac{1}{\log_{1/p_{11}} n} \ln \sum_{k=1}^{q} n^{k\varepsilon}\Big(c_6 \cdot n^{-(1+k\varepsilon)}\big(n - (1+k\varepsilon)\log_{1/p_{11}} n\big) \\
&\qquad\qquad - c_7 \cdot n^{-(1+(k+1)\varepsilon)}\Big(n - \big(1+(k+1)\varepsilon\big)\log_{1/p_{11}} n\Big)\Big) \\
&= \ln(1/p_{11}) + \frac{1}{\log_{1/p_{11}} n} \ln \sum_{k=1}^{q} \left(\frac{c_6}{n}\big(n - (1+k\varepsilon)\log_{1/p_{11}} n\big)\right. \\
&\qquad\qquad \left. - \frac{c_7}{n^{1+\varepsilon}}\Big(n - \big(1+(k+1)\varepsilon\big)\log_{1/p_{11}} n\Big)\right) \\
&\sim \ln(1/p_{11}) + \frac{1}{\log_{1/p_{11}} n} \ln\left[\frac{c_6}{n} \cdot \frac{n^2}{2\varepsilon \log_{1/p_{11}} n} - \frac{c_7}{n^{1+\varepsilon}} \cdot \frac{n^2}{2\varepsilon \log_{1/p_{11}} n}\right] \\
&\sim \ln(1/p_{11}) + \frac{1}{\log_{1/p_{11}} n} \ln\left[\frac{c_6}{n} \cdot \frac{n^2}{2\varepsilon \log_{1/p_{11}} n}\right] \\
&\sim \ln(1/p_{11}) + \ln(1/p_{11}) = 2\ln(1/p_{11}).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\{\lambda \cdot L(n)\} \\
&= \max\left\{\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\left\{\lambda \cdot L(n), \left\{\frac{L(n)}{\log_{1/p_{11}} n} \leqslant 1+\varepsilon\right\}\right\},\right. \\
&\qquad \left.\limsup_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln \mathbb{E} \exp\left\{\lambda \cdot L(n), \left\{\frac{L(n)}{\log_{1/p_{11}} n} > 1+\varepsilon\right\}\right\}\right\}.
\end{aligned}$$

The first limit is estimated as

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\exp\left\{\lambda\cdot L(n),\left\{\frac{L(n)}{\log_{1/p_{11}} n}\leqslant 1+\varepsilon\right\}\right\}\\
&\leqslant\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\exp\{(1+\varepsilon)\log_{1/p_{11}} n\cdot\ln(1/p_{11})\}\mathbb{P}\left\{\frac{L(n)}{\log_{1/p_{11}} n}\leqslant 1+\varepsilon\right\}\\
&=(1+\varepsilon)\ln(1/p_{11}).
\end{aligned}$$

The second limit is estimated as

$$\begin{aligned}
&\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\exp\left\{\lambda\cdot L(n),\left\{\frac{L(n)}{\log_{1/p_{11}} n}>1+\varepsilon\right\}\right\}\\
&=\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\exp\left\{\lambda\cdot L(n),\bigcup_{k=1}^{q}\left\{1+k\varepsilon<\frac{L(n)}{\log_{1/p_{11}} n}\leqslant 1+(k+1)\varepsilon\right\}\right\}\\
&\leqslant\frac{1}{\log_{1/p_{11}} n}\ln\sum_{k=1}^{q}\exp\left\{\big(1+(k+1)\varepsilon\big)\log_{1/p_{11}} n\cdot\ln(1/p_{11})\right\}\\
&\qquad\qquad\times\mathbb{P}\left\{\frac{L(n)}{\log_{1/p_{11}} n}>1+k\varepsilon\right\}\\
&\leqslant(1+\varepsilon)\ln(1/p_{11})\\
&\quad+\frac{1}{\log_{1/p_{11}} n}\ln\sum_{k=1}^{q}n^{k\varepsilon}\cdot c_7\cdot n^{-(1+k\varepsilon)}\big(n-(1+k\varepsilon)\log_{1/p_{11}} n\big)\\
&\sim(1+\varepsilon)\ln(1/p_{11})+\frac{1}{\log_{1/p_{11}} n}\ln\frac{c_7}{n}\cdot\frac{n^2}{\varepsilon\log_{1/p_{11}} n}\\
&\sim(1+\varepsilon)\ln(1/p_{11})+\ln(1/p_{11}).
\end{aligned}$$

Therefore,

$$\limsup_{n\to\infty}\frac{1}{\log_{1/p_{11}} n}\ln\mathbb{E}\exp\{\lambda\cdot L(n)\}\leqslant(1+\varepsilon)\ln(1/p_{11})+\ln(1/p_{11}),$$

which completes the proof.

S t e p 4. In order to study the asymptotic behavior of $\mathbb{E}\exp\{\lambda\cdot L(n)\}$ when $\lambda>\ln(1/p_{11})$, we need to consider a large deviation probability which may be of independent interest.

LEMMA 3.1. *For a fixed* $0<x<1$, *we have*

$$\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left(\frac{L(n)}{n}\geqslant x\right)\geqslant -x\ln(1/p_{11})$$

and

$$\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left(\frac{L(n)}{n}\geqslant x\right)\leqslant\max\left\{-x\ln(1/p_{11}),\ -\ln\frac{1}{|p_{00}-p_{10}|}\right\}.$$

In particular, *if* $p_{10} \leqslant p_{00} + p_{11}$, *then*

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left(\frac{L(n)}{n} \geqslant x\right) = -x \ln(1/p_{11}).$$

Proof of Lemma 3.1. We apply Lemma 2.2 with $k(n) = [nx]$ and obtain the following:

$$1 - (1 - c_4 \cdot p_{11}^k)^{n-k+1} \leqslant \mathbb{P}\left(\frac{L(n)}{n} \geqslant x\right) \leqslant 1 - (1 - c_5 \cdot p_{11}^k)^{n-k+1} + e(n).$$

The lower bound can be handled as

$$\begin{aligned}
&1 - (1 - c_4 \cdot p_{11}^k)^{n-k+1} \\
&= 1 - [(1 - c_4 \cdot p_{11}^k)^{1/(c_4 \cdot p_{11}^k)}]^{c_4 \cdot p_{11}^k (n-k+1)} \\
&= -[(1 - c_4 \cdot p_{11}^k)^{1/(c_4 \cdot p_{11}^k)}]^{\theta_n} \ln\left((1 - c_4 \cdot p_{11}^k)^{1/(c_4 \cdot p_{11}^k)}\right) \cdot c_4 \cdot p_{11}^k (n-k+1),
\end{aligned}$$

where $\theta_n \in [0, c_4 \cdot p_{11}^k (n-k+1)]$. Therefore, for big enough n, the lower bound satisfies

$$1 - (1 - c_4 \cdot p_{11}^k)^{n-k+1} \geqslant c_4 \cdot (1-\delta) p_{11}^k (n-k+1)$$

for some small $\delta > 0$, which proves the lower bound. The upper bound can be handled similarly except for the extra term $e(n)$. In this case,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \ln |e(n)| \\
&\quad \leqslant \limsup_{n\to\infty} \frac{1}{n} \ln[\text{const}_1 \cdot |p_{00} - p_{10}|^n + \text{const}_2 \cdot p_{11}^k |p_{00} - p_{10}|^{n-k}] \\
&\quad \leqslant \max\left\{-\ln \frac{1}{|p_{00} - p_{10}|},\ -x \ln(1/p_{11})\right\},
\end{aligned}$$

from which the upper bound follows. ∎

Step 5. If $\lambda > \ln(1/p_{11})$, then

$$\begin{aligned}
\lambda - \ln(1/p_{11}) &\leqslant \liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{E} e^{\lambda L(n)} \leqslant \limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{E} e^{\lambda L(n)} \\
&\leqslant \max\left\{\lambda - \ln(1/p_{11}),\ \lambda - \ln \frac{1}{|p_{00} - p_{10}|}\right\}.
\end{aligned}$$

It follows from Lemma 3.1 that, for any $0 < x < 1$,

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\exp\{\lambda\cdot L(n)\}&\\ &\geqslant \liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left[\exp\{\lambda\cdot L(n)\},\left\{\frac{L(n)}{n}>x\right\}\right]\\ &\geqslant \lambda x+\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left(\frac{L(n)}{n}>x\right)\\ &= \lambda x - x\ln(1/p_{11}) = \lambda-\ln(1/p_{11}) \quad \text{as } x\to 1.\end{aligned}$$

Furthermore,

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\exp\{\lambda\cdot L(n)\}\\ &\quad=\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left(\exp\{\lambda\cdot L(n)\},\left\{\frac{L(n)}{n}\leqslant\varepsilon\right\}\cup\left\{\frac{L(n)}{n}>\varepsilon\right\}\right)\\ &\quad=\max\left\{\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left(\exp\{\lambda\cdot L(n)\},\left\{\frac{L(n)}{n}\leqslant\varepsilon\right\}\right),\right.\\ &\qquad\qquad\left.\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left(\exp\{\lambda\cdot L(n)\},\left\{\frac{L(n)}{n}>\varepsilon\right\}\right)\right\}.\end{aligned}$$

The first limit is

$$\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left(\exp\{\lambda\cdot L(n)\},\ \left\{\frac{L(n)}{n}\leqslant\varepsilon\right\}\right)\leqslant\lambda\varepsilon.$$

The second limit is handled as follows:

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left(\exp\{\lambda\cdot L(n)\},\ \left\{\frac{L(n)}{n}>\varepsilon\right\}\right)\\ &=\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{E}\left(\exp\{\lambda\cdot L(n)\},\ \bigcup_{k=1}^{[1/\varepsilon]-1}\left\{k\varepsilon<\frac{L(n)}{n}\leqslant(k+1)\varepsilon\right\}\right)\\ &=\max_{1\leqslant k\leqslant[1/\varepsilon]-1}\left\{\lambda(k+1)\varepsilon+\limsup_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left(k\varepsilon<\frac{L(n)}{n}\right)\right\}\\ &\leqslant\max_{1\leqslant k\leqslant[1/\varepsilon]-1}\left\{\lambda(k+1)\varepsilon+\max\left\{-k\varepsilon\ln(1/p_{11}),\ -\ln\frac{1}{|p_{00}-p_{10}|}\right\}\right\}\\ &=\max_{1\leqslant k\leqslant[1/\varepsilon]-1}\left\{\lambda\cdot\varepsilon+k\varepsilon\big(\lambda-\ln(1/p_{11})\big),\ \lambda(k+1)\varepsilon-\ln\frac{1}{|p_{00}-p_{10}|}\right\}\\ &=\max\left\{\lambda-\ln(1/p_{11})+\lambda\cdot\varepsilon,\ \lambda-\ln\frac{1}{|p_{00}-p_{10}|}\right\}.\end{aligned}$$

The condition $\lambda>\ln(1/p_{11})$ is used when the maximum is attained with $k=[1/\varepsilon]-1$. The proof now follows by taking $\varepsilon\to 0^+$.

3.2. Proofs of Theorems 1.2 and 1.4. Using the proved Theorem 1.3, we are now ready to prove Theorems 1.2 and 1.4 with the help of the Gärtner–Ellis theorem. The proofs of Theorems 1.2 and 1.4 are essentially the same, and here we only show the details for the one of Theorem 1.4. Let us define the logarithmic moment generating function of $L(n)/n$ as

$$\widetilde{\Lambda}_n(\lambda) = \ln \mathbb{E} \exp\{\lambda \cdot L(n)/n\}, \quad \lambda \in \mathbb{R},$$

and the *cumulant* as

$$\widetilde{\Lambda}(\lambda) := \lim_{n\to\infty} \frac{1}{n}\widetilde{\Lambda}_n(\lambda \cdot n) = \begin{cases} \lambda - \ln(1/p_{11}), & \lambda \geqslant \ln(1/p_{11}), \\ 0, & \lambda < \ln(1/p_{11}), \end{cases}$$

where the last limit is from Theorem 1.3, under the condition $p_{10} \leqslant p_{00} + p_{11}$. Then the large deviation upper bound (1.10) follows directly from the Gärtner–Ellis theorem (cf. [3], Section 2.3) with the rate function $\widetilde{\Lambda}^*$ in (1.8) defined by the Fenchel–Legendre transform of $\widetilde{\Lambda}$ as $\widetilde{\Lambda}^*(x) = \sup_{\lambda\in\mathbb{R}}[\lambda \cdot x - \widetilde{\Lambda}(\lambda)]$.

For the large deviation lower bound (1.9), it suffices to prove that for a fixed point $0 < y < 1$,

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left(\frac{L(n)}{n} \in B_{y,\delta}\right) \geqslant -y \ln(1/p_{11}), \tag{3.5}$$

where $B_{y,\delta}$ is the open ball centered at y with a radius δ. To achieve (3.5), we write

$$\mathbb{P}\left(\frac{L(n)}{n} \in B_{y,\delta}\right) = \mathbb{P}\left(\frac{L(n)}{n} > y - \delta\right) - \mathbb{P}\left(\frac{L(n)}{n} \geqslant y + \delta\right),$$

and apply an inequality in the form $\ln(a-b) \geqslant \ln(a) - \frac{b}{a-b}$ for $a > b > 0$ to show that

$$\begin{aligned} \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left(\frac{L(n)}{n} \in B_{y,\delta}\right) &\geqslant \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n}\Bigg(\ln\left[\mathbb{P}\left(\frac{L(n)}{n} > y - \delta\right)\right] \\ &\quad - \frac{\mathbb{P}\big(L(n)/n \geqslant y + \delta\big)}{\mathbb{P}\big(L(n)/n > y - \delta\big) - \mathbb{P}\big(L(n)/n \geqslant y + \delta\big)}\Bigg). \end{aligned} \tag{3.6}$$

Lemma 3.1 implies that the first limit is, under the assumption $p_{10} \leqslant p_{00} + p_{11}$,

$$\begin{aligned} \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \ln\left[\mathbb{P}\left(\frac{L(n)}{n} > y - \delta\right)\right] \\ = \lim_{\delta\to 0} -(y-\delta)\ln(1/p_{11}) = -y\ln(1/p_{11}). \end{aligned} \tag{3.7}$$

For the second ratio term, applying Lemma 3.1 twice gives

$$\begin{aligned}(3.8)\quad &\frac{\mathbb{P}\big(L(n)/n \geqslant y+\delta\big)}{\mathbb{P}\big(L(n)/n > y-\delta\big) - \mathbb{P}\big(L(n)/n \geqslant y+\delta\big)} \\ &\qquad = \frac{1}{\mathbb{P}\big(L(n)/n > y-\delta\big)/\mathbb{P}\big(L(n)/n \geqslant y+\delta\big) - 1} \\ &\qquad \leqslant \frac{1}{e^{(2\delta\ln(1/p_{11})-\varepsilon)n} - 1} \to 0,\end{aligned}$$

as $n \to \infty$, for sufficiently small $\varepsilon > 0$ with $2\delta\ln(1/p_{11}) - \varepsilon > 0$. Then (3.5) follows by taking (3.7) and (3.8) back into (3.6).

3.3. Proof of Theorem 1.1. The limit (1.1) comes directly from Lemma 2.3. For the limit (1.2), we apply Lemma 2.4 for each $0 < x < 1$ and obtain

$$\ln\left[-c_9 \cdot n^x\right] \leqslant \ln\left[-\ln\mathbb{P}\left(\frac{L(n)}{\log_{1/p_{11}} n} \leqslant 1-x\right)\right] \leqslant \ln[-c_8 \cdot n^x].$$

Then the proof follows directly by taking the limit $\lim_{n\to\infty} 1/\log_{1/p_{11}} n$.

4. THE STEIN–CHEN METHOD

The aim of this section is to introduce the use of the Stein–Chen method in estimating the large deviation probabilities of $L(n)$ in [14], and point out a mistake in the proof of (1.4). It turns out that the employed Stein–Chen method is insufficient to prove such large deviation probabilities. Let us recall the limit (1.4): for all $x > 0$,

$$\lim_{n\to\infty} \frac{1}{\log_{1/p_{11}} n} \ln\mathbb{P}\big(L(n) - \lfloor\log_{1/p_{11}} n\rfloor \geqslant x \cdot \log_{1/p_{11}} n\big) = -x \cdot \ln(1/p_{11}).$$

The idea used in the proof of (1.4) in [14] is to approximate the large deviation probabilities $\mathbb{P}\big(L(n) - \lfloor\log_{1/p_{11}} n\rfloor \geqslant x \cdot \log_{1/p_{11}} n\big)$ by the ones involving Poisson random variables, and then to control the error term using the Stein–Chen method.

By setting $k = \big\lfloor\lfloor\log_{1/p_{11}} n\rfloor + x \cdot \log_{1/p_{11}} n\big\rfloor + 1$, it was proved on p. 1947 of [14] that

$$\begin{aligned}(4.1)\quad &\big|\mathbb{P}\big(L(n) - \lfloor\log_{1/p_{11}} n\rfloor \geqslant x \cdot \log_{1/p_{11}} n\big) \\ &\qquad - \big(1 - \exp\{-n\pi_1(1-p_{11})p_{11}^{k-1} + o(1)\}\big)\big| \\ &\qquad\qquad \leqslant \mathrm{Error}\Big(W(n), Po\big(\lambda(n)\big)\Big),\end{aligned}$$

where π_1 is a constant, $W(n)$ is a random variable depending on n, defined on p. 1941, and $Po\big(\lambda(n)\big)$ is a Poisson random variable whose intensity $\lambda(n)$, also depending on n, was defined on p. 1942. It was then proved that

$$\big(1-\exp\{-n\pi_1(1-p_{11})p_{11}^{k-1}+o(1)\}\big)=O(1)n^{-x}.$$

The error term was estimated via the Stein–Chen method as

$$\text{Error}\Big(W(n),Po\big(\lambda(n)\big)\Big)=O\bigg(\frac{\ln(n)}{n}\bigg).$$

It is then obvious true that if $0<x<1$, then the limit (1.4) holds since the error term (which is of order $O\big(\frac{\ln(n)}{n}\big)$) is smaller than n^{-x}. But the problem occurs when $x>1$, since in this case the error term is much bigger than the target n^{-x}, and the limit in (1.4) is unclear. Therefore, while employing this method, the limit (1.4) is true only for $0<x<1$. Furthermore, the Stein–Chen method seems to be impossible to remove the restriction $0<x<1$ since it gives an error of power orders, while the target term n^{-x} is also of power order which can be any size depending on x.

5. AN APPLICATION IN CONFIDENCE INTERVALS

Given simulations of the Markov chain $\{X_k\}_{1\leqslant k\leqslant n}$ with the transition matrix

$$\begin{bmatrix}p_{00} & p_{01}\\ p_{10} & p_{11}\end{bmatrix},$$

the aim of this section is to make statistical inferences on the transition probabilities p_{ij}. Since our interest throughout the paper is the longest success run, we will apply Theorem 1.2 to study the confidence intervals of p_{11}.

Theorem 1.2 implies that for each $x\geqslant 1$,

$$\lim_{n\to\infty}\frac{1}{\log_{1/p_{11}}n}\ln\mathbb{P}\bigg(\frac{L(n)}{\log_{1/p_{11}}n}\geqslant x\bigg)=-(x-1)\cdot\ln(1/p_{11}).$$

If $x=1-\ln(\alpha)/\ln(n)$ with a given small $\alpha>0$, then it holds true asymptotically that $\mathbb{P}(p_{11}<e^{-(\ln(n)-\ln(\alpha))/L(n)})=\alpha$. This suggests a $100(1-\alpha)\%$ lower confidence bound of p_{11} as follows:

$$I_{p_{11}}=\bigg(\exp\bigg\{-\frac{\ln(n)-\ln(\alpha)}{\hat{L}(n)}\bigg\},1\bigg),$$

where $\hat{L}(n)$ is a point estimate of $L(n)$. A reasonable point estimate of $L(n)$ is the observed longest success run. We can also obtain a point estimate $\hat{p}_{11}$ of p_{11} using the observed (state '1'$\to$ state '1') proportion. For estimating the transition

probabilities in terms of confidence intervals, there are many existing (more complicated) methods (cf. [2] and [13] for instance), but the advantage of our method is that the lower confidence bound is very simple and neat involving only one observation $\hat{L}(n)$.

Below in Table 1 we have simulations for different transition matrices. Although the point estimate $\hat{p}_{11}$ does not work well, the derived lower confidence bound $I_{p_{11}}$ works really good. We chose the p which is close to 1, since $\hat{p}_{11}$ is only a lower confidence bound. As the other transition probabilities change (see T_2 and T_3), the confidence interval $I_{p_{11}}$ does not change much. This is as expected since the observed longest success run $\hat{L}(n)$ is not supposed to change when the other transition probabilities change. Meanwhile, the point estimates $\hat{p}_{11}$ are quite different due to the fact that the Markov chain with T_3 will have more chance to stay at the state '0' when it is at '0' now.

TABLE 1. $100(1-\alpha)\%$ lower confidence bound of p_{11}.

	$T_1 = \begin{bmatrix} 0.4 & 0.6 \\ 0.05 & 0.95 \end{bmatrix}$	$n = 1000$	$\alpha = 0.05$	
$\hat{p}_{11} = 0.8810$ $\hat{L}(n) = 111$ $I_{p_{11}} = (0.9146, 1)$	$\hat{p}_{11} = 0.8650$ $\hat{L}(n) = 102$ $I_{p_{11}} = (0.9075, 1)$	$\hat{p}_{11} = 0.8780$ $\hat{L}(n) = 190$ $I_{p_{11}} = (0.9492, 1)$	$\hat{p}_{11} = 0.8900$ $\hat{L}(n) = 99$ $I_{p_{11}} = (0.9048, 1)$	$\hat{p}_{11} = 0.8630$ $\hat{L}(n) = 127$ $I_{p_{11}} = (0.9250, 1)$
	$T_2 = \begin{bmatrix} 0.4 & 0.6 \\ 0.02 & 0.98 \end{bmatrix}$	$n = 1000$	$\alpha = 0.05$	
$\hat{p}_{11} = 0.9510$ $\hat{L}(n) = 302$ $I_{p_{11}} = (0.9677, 1)$	$\hat{p}_{11} = 0.9450$ $\hat{L}(n) = 156$ $I_{p_{11}} = (0.9385, 1)$	$\hat{p}_{11} = 0.9530$ $\hat{L}(n) = 259$ $I_{p_{11}} = (0.9625, 1)$	$\hat{p}_{11} = 0.9500$ $\hat{L}(n) = 212$ $I_{p_{11}} = (0.9544, 1)$	$\hat{p}_{11} = 0.9660$ $\hat{L}(n) = 319$ $I_{p_{11}} = (0.9694, 1)$
	$T_3 = \begin{bmatrix} 0.9 & 0.1 \\ 0.02 & 0.98 \end{bmatrix}$	$n = 1000$	$\alpha = 0.05$	
$\hat{p}_{11} = 0.8510$ $\hat{L}(n) = 227$ $I_{p_{11}} = (0.9573, 1)$	$\hat{p}_{11} = 0.9250$ $\hat{L}(n) = 396$ $I_{p_{11}} = (0.9753, 1)$	$\hat{p}_{11} = 0.9140$ $\hat{L}(n) = 203$ $I_{p_{11}} = (0.9524, 1)$	$\hat{p}_{11} = 0.8200$ $\hat{L}(n) = 232$ $I_{p_{11}} = (0.9582, 1)$	$\hat{p}_{11} = 0.7680$ $\hat{L}(n) = 155$ $I_{p_{11}} = (0.9381, 1)$

We remark that the lower confidence bound presented above is very conservative since Theorem 1.2 gives an equivalence up to logarithm. This can be seen from the coverage probabilities. From simulations, the coverage probabilities with the transition matrices $T_i, i = 1, 2, 3$, are all near 100%, which are much higher than the confidence coefficient $100(1 - 0.05)\%$.

It has been seen that Theorem 1.2 yields the lower confidence bound using $x \geqslant 1$. In the same way, Theorem 1.1 can give a two-sided confidence interval of p_{11}. Furthermore, hypothesis testings on p_{11} can be done in a similar way.

Acknowledgments. The authors are grateful to Torkel Erhardsson, Takis Konstantopoulos and Xianyuan Wu for several comments, and to an anonymous referee for the insightful suggestions, especially the one on combining the proofs of Theorems 1.2 and 1.4.

REFERENCES

[1] N. Balakrishnan and M. V. Koutras, *Runs and Scans with Applications*, Wiley, New York 2002.
[2] E. J. Bedrick and J. Aragón, *Approximate confidence intervals for the parameters of a stationary binary Markov chain*, Technometrics 31 (4) (1989), pp. 437–448.
[3] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, corrected reprint of the second (1998) edition, Springer, Berlin 2010.
[4] P. Erdős and A. Rényi, *On a new law of large numbers*, J. Anal. Math. 23 (1970), pp. 103–111.
[5] P. Erdős and P. Révész, *On the length of the longest head-run*, in: *Topics in Information Theory*, I. Csiszár and P. Elias (Eds.), Colloq. Math. Soc. János Bolyai 16 (1975), pp. 219–228.
[6] S. Eryilmaz, *Some results associated with the longest run statistic in a sequence of Markov dependent trials*, Appl. Math. Comput. 175 (1) (2006), pp. 119–130.
[7] J. C. Fu, L. Wang, and W. Y. W. Lou, *On exact and large deviation approximation for the distribution of the longest run in a sequence of two-state Markov dependent trials*, J. Appl. Probab. 40 (2) (2003), pp. 346–360.
[8] L. Holst and T. Konstantopoulos, *Runs in coin tossing: A general approach for deriving distributions for functionals*, J. Appl. Probab. 52 (3) (2015), pp. 752–770.
[9] T. Konstantopoulos, Z. Liu, and X. Yang, *Laplace transform asymptotics and large deviation principles for longest success runs in Bernoulli trials*, J. Appl. Probab. 53 (3) (2016), pp. 747–764.
[10] Z. Liu and X. Yang, *A general large deviation principle for longest runs*, Statist. Probab. Lett. 110 (2016), pp. 128–132.
[11] Y. Mao, F. Wang, and X. Wu, *Large deviation behavior for the longest head run in an IID Bernoulli sequence*, J. Theoret. Probab. 28 (1) (2015), pp. 259–268.
[12] A. Rényi, *Foundations of Probability*, Holden-Day, San Francisco 1970.
[13] R. E. Young and A. L. Sweet, *Confidence intervals for transition probabilities in two-state Markov chains. I*, Comm. Statist. A—Theory Methods 11 (2) (1982), pp. 165–179.
[14] Y. Z. Zhang and X. Y. Wu, *Some results associated with the longest run in a strongly ergodic Markov chain*, Acta Math. Sin. (Engl. Ser.) 29 (10) (2013), pp. 1939–1948.

Zhenxia Liu
Department of Mathematics
Linköping University
581 83 Linköping, Sweden
E-mail: zhenxia.liu@liu.se

Xiangfeng Yang
Department of Mathematics
Linköping University
581 83 Linköping, Sweden
E-mail: xiangfeng.yang@liu.se

Received on 20.8.2015;
revised version on 14.5.2017

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 429–440
doi:10.19195/0208-4147.38.2.9

ENTROPIC UPPER BOUND FOR BAYES RISK IN THE QUANTUM CASE

BY

RAFAŁ WIECZOREK (ŁÓDŹ) AND HANNA PODSĘDKOWSKA (ŁÓDŹ)

Abstract. The entropic upper bound for Bayes risk in a general quantum case is presented. We obtained generalization of the entropic lower bound for probability of detection. Our result indicates upper bound for Bayes risk (in a particular case of loss function – for probability of detection) in a pretty general setting of an arbitrary finite von Neumann algebra. It is also shown under which condition the indicated upper bound is achieved.

2010 AMS Mathematics Subject Classification: Primary: 81P15; Secondary: 81P50.

Key words and phrases: Bayes risk, probability of detection, quantum measurement, entropy.

1. INTRODUCTION

One of the branches of quantum information is the theory of statistics decisions and optimal measurement. It motivates to study the Bayes risk and probability of detection of states of the physical system. Many results in the mentioned field were obtained for the quantum dynamical system represented by algebra of all bounded operators with canonical trace (sometimes even on a finite-dimensional Hilbert space) by [1]–[3], [11], and [12].

In this paper we present more general results received for an arbitrary von Neumann algebra with finite faithful normal trace τ. We employ the definition of the Segal entropy of states from a predual of algebra.

2. BASIC NOTIONS

2.1. Concept of entropy. Let $\mathfrak{M}$ be a semi-finite von Neumann algebra of operators acting on a Hilbert space $\mathcal{H}$ with a normal semi-finite faithful trace τ, identity $\mathbb{1}$, and predual $\mathfrak{M}_*$. By $\mathfrak{M}_*^+$ we shall denote the set of positive functionals in $\mathfrak{M}_*$. These functionals will sometimes be referred to as (non-normalized) states. The set of normalized states, i.e. the elements $\rho \in \mathfrak{M}_*^+$ such that $\rho(\mathbb{1}) = \|\rho\| = 1$, will be denoted by $\mathfrak{S}$.

The algebra of measurable operators $\tilde{\mathfrak{M}}$ is defined as a topological *-algebra of densely defined closed operators on $\mathcal{H}$ affiliated with $\mathfrak{M}$ with strong addition and strong multiplication.

For each $\rho \in \mathfrak{M}_*$ there is a measurable operator h such that

$$\rho(x) = \tau(xh) = \tau(hx), \quad x \in \mathfrak{M}.$$

The space of all such operators is denoted by $L^1(\mathfrak{M}, \tau)$ and the correspondence above is one-to-one and isometric, where the norm of $L^1(\mathfrak{M}, \tau)$, denoted by $\|\cdot\|_1$, is defined as

$$\|h\|_1 = \tau(|h|), \quad h \in L^1(\mathfrak{M}, \tau).$$

Moreover, self-adjoint operators in $L^1(\mathfrak{M}, \tau)$ correspond to Hermitian functionals in $\mathfrak{M}_*$, and positive operators in $L^1(\mathfrak{M}, \tau)$ to the states in $\mathfrak{M}_*$.

For a state ρ the corresponding element in $L^1(\mathfrak{M}, \tau)$ will be denoted by $\hat{\rho}$ and called the *density matrix* of ρ, thus

$$\rho(x) = \tau(x\hat{\rho}) = \tau(\hat{\rho}x), \quad x \in \mathfrak{M}.$$

In particular,

$$\tau(\hat{\rho}) = \rho(\mathbb{1}).$$

Observe that for a finite τ, we have $\mathfrak{M} \subset L^1(\mathfrak{M}, \tau)$.

In the case of the full algebra $\mathbb{B}(\mathcal{H})$, a well-established concept of entropy goes back to J. von Neumann who defined the entropy of a state ρ as

$$S(\rho) = -\operatorname{tr} \hat{\rho} \log \hat{\rho},$$

where $\hat{\rho}$ is a positive trace operator of the trace one.

Unfortunately, when we deal with an arbitrary von Neumann algebra, a satisfactory general definition of entropy is lacking. Thus we employ the Segal entropy (up to the minus sign) of $\rho \in \mathfrak{M}_*$, denoted by $H(\rho)$ and defined as

$$H(\rho) = \tau(\hat{\rho} \log \hat{\rho}),$$

i.e. for the spectral representation of $\hat{\rho}$,

$$\hat{\rho} = \int_0^\infty \lambda e(d\lambda),$$

we have

$$H(\rho) = \int_0^\infty \lambda \log \lambda \tau\big(e(d\lambda)\big).$$

Although for a semi-finite algebra $\mathfrak{M}$ this definition is a straightforward generalization of the von Neumann idea, the reasoning which substantiates Segal entropy properties needs a different setup from the one used in the case of $\mathfrak{M} = \mathbb{B}(\mathcal{H})$.

REMARK 2.1. Despite being a seemingly straightforward generalization of von Neumann entropy, the Segal definition exhibits fundamental differences in many respects from that of von Neumann. For example, while the density operator in the von Neumann definition is a trace-class operator, and thus has a discrete spectrum with the eigenvalues summing up to one, this is not the case in the Segal definition. Furthermore, the von Neumann entropy of a state is nonnegative (which is a consequence of the above property of the density operator), while the Segal entropy of a state need not be such. In addition, there are also some technical problems while dealing with a semi-finite trace. For these reasons, we shall consider the case of a finite von Neumann algebra and adopt a definition of entropy more in the spirit of the classical Boltzmann–Gibbs notion, where for a density function f on a probability space $(\Omega, \mathbb{F}, \mu)$, its entropy is defined as

$$H(f) = \int\limits_{\Omega} f(\log f) d\mu.$$

As will be seen, our definition, which is just that of Segal up to a minus sign, assigns a finite nonnegative entropy to a state, and more generally, for each non-normalized state in $\mathfrak{M}_*^+$ with bounded density, its entropy is finite.

It should be noted that some fundamental investigations concerning entropy and related notions in the above setup were carried out in [9].

REMARK 2.2 (see [6]). *For a finite algebra $\mathfrak{M}$ (this is the case of our interest) with faithful finite normal trace τ, $\tau(\mathbb{1}) = 1$, for each $\rho \in \mathfrak{S}$, $H(\rho) \geqslant 0$, and for $\hat{\rho} \in \mathfrak{M}$, $H(\rho)$ is also bounded from above.*

Indeed, since $\lambda \log \lambda \geqslant \lambda - 1$, we have

$$\begin{aligned} H(\rho) &= \tau\Big(\int\limits_0^\infty \lambda \log \lambda \, e(d\lambda)\Big) = \int\limits_0^\infty \lambda \log \lambda \, \tau\big(e(d\lambda)\big) \geqslant \int\limits_0^\infty (\lambda - 1)\, \tau\big(e(d\lambda)\big) \\ &= \int\limits_0^\infty \lambda \, \tau\big(e(d\lambda)\big) - \int\limits_0^\infty \tau\big(e(d\lambda)\big) = \tau(\hat{\rho}) - \tau(\mathbb{1}) = \rho(\mathbb{1}) - 1, \end{aligned} \tag{2.1}$$

showing that the entropy is bounded from below, and in particular, it is nonnegative for states. Moreover, since $\hat{\rho}$ is bounded, its spectrum is a bounded set; thus, the function $\lambda \mapsto \lambda \log \lambda$ is bounded on the spectrum, which implies that the entropy is bounded from above.

REMARK 2.3. *In the classical quantum case, that is, for $\mathfrak{M} = \mathbb{B}(\mathcal{H})$, the practical Klein's inequality holds* (see [7] and [10]). *The analogue of this inequality is given by the formula*

$$\tau(a \log a - a \log b) \geqslant 0 \quad \text{for } a, b \in \mathfrak{M}_+, \tau(a) = \tau(b) = 1 \text{ and } \operatorname{supp} a \leqslant \operatorname{supp} b, \tag{2.2}$$

and it was proved by Umegaki [9] *in the case of an arbitrary von Neumann algebra $\mathfrak{M}$ with finite faithful normal trace τ. In addition, it was proved* (see [4], Theorem 2.1.2(i)) *that the equality in* (2.2) *holds if and only if $a = b$ for $\mathfrak{M} = \mathbb{B}(\mathbb{C}^d)$.*

REMARK 2.4. *Moreover, apart from the very practical Klein's inequality, which holds for the Segal entropy, we have also the following inequality* (see [6], Proposition 1). *Let* $a, b \in \mathfrak{M}$ *be such that* $0 \leqslant a \leqslant b$. *Then*

$$\tau(a \log b - a \log a) \geqslant 0, \tag{2.3}$$

with equality if and only if $ab = ba = a^2$. *Moreover,* $a \log a$ *and* $a \log b$ *are bounded* (*belong to* $\mathfrak{M}$), *and the numbers* $\tau(a \log b)$ *and* $\tau(a \log a)$ *are finite.*

(This remark is presented and proved with details in [6], but we remind its main idea to make our reasoning clearer.)

P r o o f. Since

$$0 \leqslant a \leqslant b,$$

we have

$$0 \leqslant (\log b)a(\log b) \leqslant (\log b)b(\log b) = b \log^2 b.$$

The operator on the right-hand side of the inequality above is bounded (belongs to $\mathfrak{M}$), hence $(\log b)a(\log b)$ is also bounded (belongs to $\mathfrak{M}$). Moreover,

$$(\log b)a(\log b) = (a^{1/2} \log b)^* a^{1/2} \log b,$$

thus $a^{1/2} \log b$ is bounded (belongs to $\mathfrak{M}$). Consequently, $a^{1/2}(\log b - \log a)$ and $a^{1/2}$ belong to $\mathfrak{M}$, so from the properties of trace we obtain

(2.4)
$$\tau\big(a(\log b - \log a)\big) = \tau\Big(a^{1/2}\big(a^{1/2}(\log b - \log a)\big)\Big) = \tau\big(a^{1/2}(\log b - \log a)a^{1/2}\big).$$

Since the logarithm is an operator monotone function, we have

$$\log b - \log a \geqslant 0,$$

yielding

$$a^{1/2}(\log b - \log a)a^{1/2} \geqslant 0,$$

and finally, by equation (2.4),

$$0 \leqslant \tau\big(a^{1/2}(\log b - \log a)a^{1/2}\big) = \tau\big(a(\log b - \log a)\big).$$

The assumption

$$\tau(a \log b - a \log a) = 0 \tag{2.5}$$

gives

$$a^{1/2}(\log b - \log a)^{1/2} = 0,$$

yielding

$$a(\log a - \log b) = 0,$$

i.e.,

$$a \log a = a \log b.$$

Taking adjoints, we get

$$a \log a = (\log b)a.$$

In particular, $\log b$ commutes with a, leaves the range of a invariant and coincides with $\log a$ on the range of a. Thus, on the range of a we have

$$a|\operatorname{Range} a = e^{\log a}|\operatorname{Range} a = e^{\log b}|\operatorname{Range} a = b|\operatorname{Range} a,$$

which is equivalent to the equalities

$$ab = ba = a^2.$$

Conversely, assume that the equality above holds true. Then, a and b commute, so after taking logarithms of both sides, we get

$$2 \log a = \log a + \log b,$$

that is

$$\log a = \log b,$$

which implies the equality

$$a \log a = a \log b,$$

and thus equation (2.5). ∎

2.2. Bayes risk in the quantum case. We are given a von Neumann algebra $\mathfrak{M}$ describing the (bounded) observables of a physical system. Let $\rho_1, \rho_2, \dots$ be normal states from $\mathfrak{M}_*$. We assume that the physical system can be in state ρ_i with *a priori* probability $\pi_i, i = 1, 2, \dots$, where $\pi = (\pi_1, \pi_2, \dots)$ is a probability distribution. On the system we perform a *measurement* (called also a *strategy*) M by which we mean a sequence $(M_1, M_2, \dots)$ of positive operators from $\mathfrak{M}$ such that

$$\sum_{i=1}^{\infty} M_i = \mathbb{1},$$

where the series is convergent in the weak operator topology on $\mathfrak{M}$.

We want to find, in an optimal way, the state in which the system really is.

If we receive an outcome M_i, we choose the state ρ_i. The probability that the true state is ρ_i when the measurement gives the result M_j is determined by $\rho_i(M_j)$. Thus $\rho_i(M_j)$ is the probability of guessing the state ρ_i correctly. If our guess is ρ_j

while the true one is ρ_i, then we pay a penalty $L(i,j)$. The function $L : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ is called a *loss function*. The risk function is defined by the formula

$$R_M(i) = \sum_{j=1}^{\infty} L(i,j)\rho_i(M_j).$$

The expectation of the risk function is called the *Bayes risk* and denoted by $r(M,\pi)$, i.e.

$$r(M,\pi) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \pi_i L(i,j)\rho_i(M_j).$$

Consider the concrete loss function of the form

$$L(i,j) = 1 - \delta_{ij}.$$

Then we have

$$r(M,\pi) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \pi_i(1-\delta_{ij})\rho_i(M_j) = 1 - \sum_{i=1}^{\infty} \pi_i\rho_i(M_j).$$

In this case, minimizing the Bayes risk is equivalent to maximizing the expression

$$\sum_{i=1}^{\infty} \pi_i\rho_i(M_j)$$

which is the probability of the correct guess while performing the measurement M, and is called the *probability of detection*. We shall denote this probability by $\mathbb{P}_D(M)$.

3. ENTROPIC BOUND

For an arbitrary loss function we have no guarantee of the existence of an optimal measurement, e.g. the one which minimizes the Bayes risk. However. under the assumptions presented in the following theorem, we can consider such an optimal measurement.

THEOREM 3.1 ([5], Theorem 8). *Let L be a loss function that satisfies the following conditions*:

(i) *there are $a_i \geqslant 0$ such that for each i we have $|L(i,j)| \leqslant a_i, j = 1,2,\ldots,$ and $\sum_{i=1}^{\infty} \pi_i a_i < \infty$;*

(ii) *for each i there exists $\lim_{j\to\infty} L(i,j) = b_i$ such that for some j_0 we have $L(i,j_0) \leqslant b_i$ for all $i = 1,2,\ldots$*

Then, there exists an optimal measurement.

In [3] we can find more information on the existence of an optimal measurement. From now on, we will assume that the loss function satisfies the conditions (i) and (ii) of Theorem 3.1. The following theorem will be very helpful for further consideration.

THEOREM 3.2 ([2], Theorem II.2.2). *We have the relation*

$$\min_M r(M,\pi) = \max\{\psi(\mathbb{1}) : \psi \in \mathfrak{M}_*, \psi \leqslant \varphi_j, j = 1,2,\ldots\}, \tag{3.1}$$

where $\varphi_j = \sum_{i=1}^{\infty} \pi_i L(i,j)\rho_i$, $j = 1,2,\ldots$ *Then the following assertions are equivalent*:

(i) *The measurement* $M = (M_i)$ *is optimal for* r, *and* $\psi \in \mathfrak{M}_*$ *maximizes the right-hand side of* (3.1).

(ii) $\psi \leqslant \varphi_j, j = 1,2,\ldots$, *and* $\psi = \sum_{j=1}^{\infty} \varphi_j M_j = \sum_{j=1}^{\infty} M_j \varphi_j$.

Let $c = (c_i)$ be a sequence such that $L(i,j) \leqslant c_i$ for all i,j and the sum $\sum_{i=1}^{\infty} \pi_i c_i$ is convergent. Define the functional

$$r_c(M,\pi) := \sum_{j=1}^{\infty} \varphi_j^c(M_j) = \sum_{i=1}^{\infty} \pi_i c_i - r(M,\pi),$$

where $\varphi_j^c = \sum_{i=1}^{\infty} \pi_i\big(c_i - L(i,j)\big)\rho_i$, $j = 1,2,\ldots$ Minimizing the Bayes risk is equivalent to maximizing the functional r_c with positive functionals φ_j^c.

The next result is a simple consequence of Theorem 3.2.

THEOREM 3.3. *We have the relation*

$$\max_M r_c(M,\pi) = \min\{\varphi(\mathbb{1}) : \varphi \in \mathfrak{M}_*, \varphi \geqslant \varphi_j^c, j = 1,2,\ldots\}. \tag{3.2}$$

Then the following assertions are equivalent:

(i) *The measurement* $M = (M_i)$ *is optimal for* r_c, *and* $\varphi \in \mathfrak{M}_*$ *minimizes the right-hand side of* (3.2).

(ii) $\varphi \geqslant \varphi_j^c, j = 1,2,\ldots$, *and* $\varphi = \sum_{j=1}^{\infty} \varphi_j^c M_j = \sum_{j=1}^{\infty} M_j \varphi_j^c$.

P r o o f. From (3.1) we obtain

$$\min_M r(M,\pi) = \max\big\{\psi(\mathbb{1}) : \psi \leqslant \sum_i \pi_i L(i,j)\rho_i, j = 1,2,\ldots\big\}. \tag{3.3}$$

Denote by φ the functional $\sum_i \pi_i c_i \rho_i - \psi$. Then the above equality takes the form

$$\min_M r(M,\pi) = \sum_i c_i \pi_i - \min\{\varphi(\mathbb{1}) : \varphi_j^c \leqslant \varphi, j = 1,2,\ldots\}.$$

Consequently,

$$\max_M r_c(M,\pi) = \min\{\varphi(\mathbb{1}) : \varphi_j^c \leqslant \varphi, j = 1,2,\ldots\}. \tag{3.4}$$

(i)⇒(ii). Let φ be an optimal functional in (3.2). Then $\varphi \geqslant \varphi_j^c, j = 1, 2, \ldots$ The functional $\psi = \sum_i \pi_i c_i \rho_i - \varphi$ maximizes the right-hand side of (3.1). Then from Theorem 3.2(ii) we have

$$\psi = \sum_j \varphi_j M_j,$$

where $\varphi_j = \sum_i \pi_i L(i,j)\rho_i$ and (M_j) is an optimal measurement. Therefore, the optimal functional φ in (3.4) is of the form $\sum_i \varphi_i^c M_i$.

(ii)⇒(i). Let φ be such that $\varphi \geqslant \varphi_j^c, j = 1, 2, \ldots$, and $\varphi = \sum_{j=1}^\infty \varphi_j^c M_j = \sum_{j=1}^\infty M_j \varphi_j^c$. The functional $\psi = \sum_i \pi_i c_i \rho_i - \varphi$ satisfies the conditions $\psi \leqslant \varphi_j$, $j = 1, 2, \ldots$, and $\psi = \sum_{j=1}^\infty \varphi_j M_j = \sum_{j=1}^\infty M_j \varphi_j$. Then, by Theorem 3.2, the measurement $M = (M_i)$ is optimal for r (also for r_c) and $\psi \in \mathfrak{M}_*$ maximizes the right-hand side of (3.1). Therefore, $\varphi \in \mathfrak{M}_*$ and φ minimizes the right-hand side of (3.2). ∎

In the rest of this article we assume that $\mathfrak{M}$ is a finite von Neumann algebra with faithful finite normal trace τ, $\tau(\mathbb{1}) = 1$, and $\hat{\rho}_1, \hat{\rho}_2, \ldots \in \mathfrak{M}$. Denote by $\|\cdot\|_\infty$ the operator norm in $\mathfrak{M}$.

THEOREM 3.4 (Main theorem). *Let the series* $\sum_{ij} \pi_i\big(c_i - L(i,j)\big)\|\hat{\rho}_i\|_\infty$ *be convergent and let us put* $a_c = \sum_{ij} \pi_i\big(c_i - L(i,j)\big)$. *Then we have the estimate*

$$\min_M r(M,\pi) \leqslant \sum_i \pi_i c_i - 2^{\frac{1}{a_c}(\sum_i H(\varphi_i^c)) - H\left(\frac{1}{a_c}(\sum_i \varphi_i^c)\right)}. \tag{3.5}$$

Proof. Note that the convergence of the series $\sum_{ij} \pi_i\big(c_i - L(i,j)\big)\|\hat{\rho}_i\|_\infty$ implies that $\hat{\varphi}_1^c, \hat{\varphi}_2^c, \ldots \in \mathfrak{M}$, $\sum_i \hat{\varphi}_i^c \in \mathfrak{M}$ and also the convergence of the series $\sum_{ij} \pi_i\big(c_i - L(i,j)\big)$. Let φ' be an optimal functional from the right-hand side of (3.2). Then

$$\hat{\varphi}' = \sum_i \hat{\varphi}_i^c M_i,$$

where $M = (M_i)$ is an optimal measurement. The series $\sum_{ij} \pi_i\big(c_i - L(i,j)\big)\|\hat{\rho}_i\|_\infty$ is convergent, so the series $\sum_i \hat{\varphi}_i^c M_i$ is the Cauchy series for the norm $\|\cdot\|_\infty$. $\hat{\varphi}_i^c M_i \in \mathfrak{M}$, therefore $\hat{\varphi}' \in \mathfrak{M}$. Observe that $a_c = \tau\big(\sum_i \hat{\varphi}_i^c\big)$. By Remark 2.4, the operator $\hat{\varphi}_i^c \log \hat{\varphi}'$ is bounded because $\hat{\varphi}' \geqslant \hat{\varphi}_i^c$. The operator $\big(\sum_i \hat{\varphi}_i^c\big) \log \hat{\varphi}'$ is the pointwise limit of the sequence of operators $\big(\sum_{i=1}^n \hat{\varphi}_i^c\big) \log \hat{\varphi}'$, so it is bounded. On the other hand, using the inequality (2.3), we obtain

$$\tau(\hat{\varphi}_i^c \log \hat{\varphi}') \geqslant \tau(\hat{\varphi}_i^c \log \hat{\varphi}_i^c).$$

In summary, we have the convergence of the series $\sum_i \tau(\hat{\varphi}_i^c \log \hat{\varphi}')$ and

$$\sum_i \tau\big(\hat{\varphi}_i^c (\log \hat{\varphi}' - \log \hat{\varphi}_i^c)\big) \geqslant 0. \tag{3.6}$$

Thus there are two cases. First, the series $\sum_i \tau\big(\hat{\varphi}_i^c(\log\hat{\varphi}' - \log\hat{\varphi}_i^c)\big)$ is divergent, so

$$\sum_i \tau\big(\hat{\varphi}_i^c(\log\hat{\varphi}' - \log\hat{\varphi}_i^c)\big) = \infty.$$

Then $\sum_i \tau(\hat{\varphi}_i^c \log\hat{\varphi}_i^c) = -\infty$ and the inequality (3.5) is true because it takes the form

$$\min_M r(M,\pi) \leqslant \sum_i \pi_i c_i.$$

Second, since the series $\sum_i \tau\big(\hat{\varphi}_i^c(\log\hat{\varphi}' - \log\hat{\varphi}_i^c)\big)$ is convergent, so is the series $\sum_i \tau(\hat{\varphi}_i^c \log\hat{\varphi}_i^c)$. Using the inequality (3.6), we have the estimate

$$\begin{aligned}(3.7)\quad \log\max_M r_c(M,\pi) &\geqslant \log\tau(\hat{\varphi}') - \frac{1}{a_c}\sum_i \tau\big(\hat{\varphi}_i^c(\log\hat{\varphi}' - \log\hat{\varphi}_i^c)\big)\\ &= -\frac{1}{a_c}\sum_i \tau\left(\hat{\varphi}_i^c\left(\log\frac{\hat{\varphi}'}{\tau(\hat{\varphi}')} - \log\hat{\varphi}_i^c\right)\right)\\ &= -\frac{1}{a_c}\tau\left(\Big(\sum_i \hat{\varphi}_i^c\Big)\log\frac{\hat{\varphi}'}{\tau(\hat{\varphi}')}\right) + \frac{1}{a_c}\sum_i \tau(\hat{\varphi}_i^c\log\hat{\varphi}_i^c).\end{aligned}$$

With our notation, it is obvious that the assumptions of Klein's inequality (2.2) hold and

$$(3.8)\qquad \tau\left(\frac{\sum_i \hat{\varphi}_i^c}{a_c}\log\frac{\sum_i \hat{\varphi}_i^c}{a_c}\right) \geqslant \tau\left(\frac{\sum_i \hat{\varphi}_i^c}{a_c}\log\frac{\hat{\varphi}'}{\tau(\hat{\varphi}')}\right),$$

therefore

$$\log\max_M r_c(M,\pi) \geqslant -\tau\left(\frac{\sum_i \hat{\varphi}_i^c}{a_c}\log\frac{\sum_i \hat{\varphi}_i^c}{a_c}\right) + \frac{1}{a_c}\sum_i \tau(\hat{\varphi}_i^c\log\hat{\varphi}_i^c) =: A.$$

Consequently, we obtain the inequality

$$\min_M r(M,\pi) \leqslant \sum_i \pi_i c_i - 2^A. \quad \blacksquare$$

In the proof of Theorem 3.4 we used the idea of the proof of Lemma 2 in [8].

COROLLARY 3.1. *Assume that the series $\sum_{i=1}^\infty \pi_i\|\rho_i\|_\infty$ is convergent. Then for the probability of detection we obtain*

$$(3.9)\qquad \max_M \mathbb{P}_D(M) \geqslant 2^{\sum_i \pi_i\log\pi_i - H(\sum_i \pi_i\rho_i) + \sum_i \pi_i H(\rho_i)}.$$

P r o o f. In Theorem 3.4, let us consider the concrete loss function of the form $L(i,j) = 1 - \delta_{ij}$ and $c = (1,1,\ldots)$. Then we have $r_c(M,\pi) = \sum_i \pi_i\rho_i(M_i)$.

This expression is the probability of detection. From the inequality (3.5) we obtain

$$\max_M \mathbb{P}_D(M) \geqslant 2^{\sum_i H(\pi_i\rho_i) - H(\sum_i \pi_i\rho_i)}.$$

We have

$$\sum_i H(\pi_i\rho_i) = \sum_i \tau(\pi_i\rho_i \log \pi_i\rho_i) = \sum_i [\pi_i \tau(\rho_i \log \rho_i) + \pi_i \log \pi_i].$$

By Remark 2.2, the inequality $\tau(\rho_i \log \rho_i) \geqslant 0$ holds. On the other hand, from the inequality $\log x \leqslant x - 1$ we obtain

$$\tau(\rho_i \log \rho_i) = \int_0^{\|\rho_i\|_\infty} \lambda \log \lambda \tau\big(e_i(d\lambda)\big) \leqslant \log \|\rho_i\|_\infty \leqslant \|\rho_i\|_\infty - 1.$$

Therefore, $0 \leqslant \pi_i \tau(\rho_i \log \rho_i) \leqslant \pi_i \|\rho_i\|_\infty - \pi_i$. By the assumption, the series $\sum_i (\pi_i \|\rho_i\|_\infty - \pi_i)$ is convergent, so the series $\sum_i \pi_i \tau(\rho_i \log \rho_i)$ is also convergent. Consequently,

$$\sum_i [\pi_i \tau(\rho_i \log \rho_i) + \pi_i \log \pi_i] = \sum_i \pi_i \tau(\rho_i \log \rho_i) + \sum_i \pi_i \log \pi_i$$

and

$$\sum_i H(\pi_i\rho_i) - H\big(\sum_i \pi_i\rho_i\big) = \sum_i \pi_i H(\rho_i) + \sum_i \pi_i \log \pi_i - H\big(\sum_i \pi_i\rho_i\big). \quad \blacksquare$$

In the case of a finite-dimensional Hilbert space and a finite number of states, Corollary 3.1 is the main result in [12].

In the next theorem and corollary we assume that $\mathfrak{M} = \mathbb{B}(\mathbb{C}^d)$ and consider a finite number of states $\rho_1, \rho_2, \ldots, \rho_n$.

THEOREM 3.5. *Let $\hat{\varphi}_i^c = \sum_{j=1}^d \lambda_i^j |v_i^j\rangle\langle v_i^j|$ be a spectral decomposition of the operator $\hat{\varphi}_i^c, i = 1, 2, \ldots, n$. Write $E = \{v_i^j : j = 1, 2, \ldots, d, i = 1, 2, \ldots, n\}$. Assume that $\operatorname{Lin} E = \mathbb{C}^d$ and $E \neq A \cup B$, where $A, B \neq \emptyset$ and $\forall_{v\in A} \forall_{w\in B}\, v \perp w$. We have the equality in* (3.5) *if and only if $\hat{\varphi}_i^c = aP_i, i = 1, 2, \ldots, n$, where a is some positive number, P_i is a projection and $\sum_i P_i = \frac{a_c}{a}\mathbb{1}$.*

Proof. Sufficiency. If the equality holds in (3.5), then it must hold also in (3.7). Thus, we have

$$\sum_i \tau\big(\hat{\varphi}_i^c(\log \hat{\varphi}' - \log \hat{\varphi}_i^c)\big) = 0$$

and, by Remark 2.4,

$$\hat{\varphi}' \hat{\varphi}_i^c = \hat{\varphi}_i^c \hat{\varphi}' = (\hat{\varphi}_i^c)^2.$$

The commutation $\hat{\varphi}'\hat{\varphi}_i^c = \hat{\varphi}_i^c\hat{\varphi}'$ means that all eigenvectors from the set E are eigenvectors of $\hat{\varphi}'$. By the assumption on the set E the operator $\hat{\varphi}'$ has only one eigenvalue. Denote it by a. Applying Theorem 3.3, we have the inequalities $\hat{\varphi}' \geqslant \hat{\varphi}_j^c, j = 1, 2, \ldots$, so the operator $\hat{\varphi}'$ is invertible, therefore it is equal to $a\mathbb{1}$. From the equality $\hat{\varphi}_i^c\hat{\varphi}' = (\hat{\varphi}_i^c)^2$ we have $a\hat{\varphi}_i^c = (\hat{\varphi}_i^c)^2$, so all eigenvalues of the operator $\hat{\varphi}_i^c$ are equal to a. Therefore,

$$\hat{\varphi}_i^c = aP_i \quad \text{for some projection } P_i.$$

The equality must also be in formula (3.8), that is,

$$\tau\left(\frac{\sum_i \hat{\varphi}_i^c}{a_c} \log \frac{\sum_i \hat{\varphi}_i^c}{a_c}\right) = \tau\left(\frac{\sum_i \hat{\varphi}_i^c}{a_c} \log \frac{\hat{\varphi}'}{\tau(\hat{\varphi}')}\right).$$

So, by Remark 2.3, $\frac{1}{\tau(\hat{\varphi}')}\hat{\varphi}' = \frac{1}{a_c}\sum_i \hat{\varphi}_i^c$. This gives the condition $\sum_i P_i = \frac{a_c}{a}\mathbb{1}$.

N e c e s s i t y. Let $M = (M_1, M_2, \ldots, M_n)$, $M_i = \frac{a}{a_c}P_i$. We have

$$\sum_i \hat{\varphi}_i^c M_i = \sum_i \frac{a^2}{a_c}P_i = a\mathbb{1} \geqslant aP_i = \hat{\varphi}_i^c,$$

so Theorem 3.3 implies that M is an optimal measurement and $\max_M r_c(M, \pi) = \tau(a\mathbb{1}) = a$. On the other hand,

$$\begin{aligned}\frac{1}{a_c}\Big(\sum_i H(\varphi_i^c)\Big) - H\left(\frac{1}{a_c}\Big(\sum_i \varphi_i^c\Big)\right) &= \frac{1}{a_c}\Big(\sum_i H(aP_i)\Big) - H(\mathbb{1}) \\ &= \frac{a}{a_c}\sum_i \tau\big(P_i \log(aP_i)\big) = \frac{a}{a_c}\sum_i \tau(P_i)\log a = \log a,\end{aligned}$$

therefore

$$\max_M r_c(M, \pi) = 2^{\frac{1}{a_c}(\sum_i H(\varphi_i^c)) - H\left(\frac{1}{a_c}(\sum_i \varphi_i^c)\right)}. \quad \blacksquare$$

For the probability of detection we obtain

COROLLARY 3.2. *Let $\hat{\rho}_i = \sum_{j=1}^d \lambda_i^j |v_i^j\rangle\langle v_i^j|$ be a spectral decomposition of the operator $\hat{\rho}_i, i = 1, 2, \ldots, n$. Write $E = \{v_i^j : j = 1, 2, \ldots, d, i = 1, 2, \ldots, n\}$. Assume that* $\operatorname{Lin} E = \mathbb{C}^d$ *and $E \neq A \cup B$, where $A, B \neq \emptyset$ and $\forall_{v\in A}\forall_{w\in B}\, v \perp w$. We have the equality in* (3.9) *if and only if $\hat{\rho}_i = \frac{1}{m_i}P_i, i = 1, 2, \ldots, n$, where $m_i = \tau(P_i)$, P_i is a projection, $\sum_i P_i = m\mathbb{1}, m = \sum_i m_i$ and $\pi_i = m_i/m$.*

P r o o f. In Theorem 3.5 consider the loss function of the form $L(i, j) = 1 - \delta_{ij}$ and $c = (1, 1, \ldots)$. Then we have $\pi_i\hat{\rho}_i = aP_i$ for some $a > 0$ and the projection P_i. Therefore, $\pi_i = am_i$ and $a = \frac{1}{m}$. Consequently, $\pi_i = m_i/m$ and $\sum_i P_i = m\mathbb{1}$. $\blacksquare$

REFERENCES

[1] A. S. Holevo, *Statistical decision theory for quantum systems*, J. Multivariate Anal. 3 (1973), pp. 337–394.
[2] A. S. Holevo (A. S. Kholevo), *Investigations in the general theory of statistical decisions*, Proc. Steklov Inst. Math. 124 (3) (1978), pp. 1–140.
[3] A. Łuczak, *Maximizing the probability of detection for pure states*, J. Math. Phys. 50 (5) (2009), 053502.
[4] S. Neshveyev and E. Størmer, *Dynamical Entropy in Operator Algebras*, Springer, Berlin 2006.
[5] M. Ozawa, *Optimal measurements for general quantum systems*, Rep. Math. Phys. 18 (1) (1980), pp. 11–28.
[6] H. Podsędkowska, *Entropy of quantum measurement*, Entropy 17 (3) (2015), pp. 1181–1196.
[7] D. Ruelle, *Statistical Mechanics: Rigorous Results*, W. A. Benjamin, Inc., New York–Amsterdam 1969.
[8] M. Tomamichel, R. Colbeck, and R. Renner, *A fully quantum asymptotic equipartition property*, IEEE Trans. Inform. Theory 55 (12) (2009), pp. 5840–5847.
[9] H. Umegaki, *Conditional expectation in an operator algebra. IV. Entropy and information*, Kodai Math. Sem. Rep. 14 (1962), pp. 59–85.
[10] A. Wehrl, *General properties of entropy*, Rev. Modern Phys. 50 (2) (1978), pp. 221–260.
[11] R. Wieczorek, *On the form of the optimal measurement for the probability of detection*, Internat. J. Theoret. Phys. 54 (12) (2015), pp. 4506–4511.
[12] S. Yang, J. Lee, and H. Jeong, *Entropic lower bound for distinguishability of quantum states*, Adv. Math. Phys. (2015), article ID 683658.

Rafał Wieczorek
Faculty of Mathematics and Computer Science
Łódź University
ul. S. Banacha 22, 90-238 Łódź, Poland
E-mail: wieczorek@math.uni.lodz.pl

Hanna Podsędkowska
Faculty of Mathematics and Computer Science
Łódź University
ul. S. Banacha 22, 90-238 Łódź, Poland
E-mail: hpodsedk@math.uni.lodz.pl

Received on 8.2.2017;
revised version on 8.6.2017

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 441–458
doi:10.19195/0208-4147.38.2.10

PRESERVATION PROPERTIES OF STOCHASTIC ORDERS BY TRANSFORMATION TO HARRIS FAMILY

BY

SOMAYEH ABBASI* (ISFAHAN)
AND MOHAMMAD HOSSEIN ALAMATSAZ (ISFAHAN)

Abstract. Stochastic comparisons of lifetime characteristics of reliability systems and their components are of common use in lifetime analysis. In this paper, using Harris family distributions, we compare lifetimes of two series systems with random number of components, with respect to several types of stochastic orders. Our results happen to enfold several previous findings in this connection. We shall also show that several stochastic orders and ageing characteristics, such as IHRA, DHRA, NBU, and NWU, are inherited by transformation to Harris family. Finally, some refinements are made concerning related existing results in the literature.

2010 AMS Mathematics Subject Classification: Primary: 60E15; Secondary: 60E05.

Key words and phrases: Ageing intensity order, Marshall–Olkin distribution, proportional stochastic order, shifted stochastic order, shifted proportional stochastic order, usual stochastic order.

1. INTRODUCTION

Clearly, the lifetime of any reliability system depends on the lifetime of its components. Thus, in practice, to compare stochastically the lifetime of two systems, we need to compare the lifetimes of their components. The Harris family of distributions is a known family for the lifetime of a series system. It was introduced by Aly and Benkherouf [8] as a generalization of the Marshall–Olkin family. The Marshall–Olkin family of distributions is better known as the family with a tilt parameter. It was introduced by Marshall and Olkin [25] and was obtained as the proportional odds family (proportional odds model) by Kirmani and Gupta [23]. However, it was first proposed by Clayton [15].

* The work was supported by the University of Isfahan.

The aim of this paper is to focus on the Harris family and stochastically compare such lifetime systems with each other. We recall that the Harris family is constructed by combining the Harris probability generating function (pgf) introduced by Harris [21] and a baseline distribution function. More precisely, a survival function of the family is defined as

$$\bar{H}(x;\theta,k)=\left(\frac{\theta\bar{F}^k(x)}{1-\bar{\theta}\bar{F}^k(x)}\right)^{1/k}, \tag{1.1}$$
$$-\infty<x<\infty,\ 0<\theta<\infty,\ \bar{\theta}=1-\theta,\ k>0,$$

where $F(x)$ is called the *baseline distribution function* (df) and θ is called the *tilt parameter*. It is easily seen that hazard rates corresponding to $F(x)$ and $H(x;\theta,k)$, namely, $r_F(\cdot)=f(\cdot)/\bar{F}(\cdot)$ and $r_H(\cdot;\theta,k)=h(\cdot;\theta,k)/\bar{H}(\cdot;\theta,k)$, are related by

$$r_H(x;\theta,k)=\frac{r_F(x)}{1-\bar{\theta}\bar{F}^k(x)}, \tag{1.2}$$
$$-\infty<x<\infty,\ 0<\theta<\infty,\ \bar{\theta}=1-\theta,\ k>0.$$

Clearly, $r_H(x;\theta,k)$ is shifted below $(\theta\geqslant 1)$ or above $(0<\theta\leqslant 1)$ $r_F(x)$. When $k=1$, a Harris family distribution reduces to a Marshall–Olkin distribution.

In reliability terms, a random variable (rv) X, with Harris family distribution, can be considered as the lifetime of a series system with independent and identical (iid) component lifetimes $Y_1,Y_2,\ldots,Y_N$, with df's F, when the number of components, N, is itself a Harris rv independent of Y_i's.

Recently, Batsidis and Lemonte [11] discussed another method of constructing the Harris family of distributions. They revealed that the Harris family of distributions is a proportional failure rate model which is obtained from a modified Marshall–Olkin distribution. Then, they provided several results in connection with behavior of the failure rate function for the Harris family and discussed their certain stochastic orders. Al-Jarallah et al. [7] presented a proportional hazard version of the Marshall–Olkin family of distributions as $[\bar{H}(\cdot;\theta,1)]^\gamma$ and investigated likelihood ratio order in this model.

Our aim is to compare a Harris family distribution with its baseline distribution, with respect to several stochastic orders. Stochastic orders are important tools for comparing probability distributions and play a great role in statistical inference and applied probability. Frequently, they are applied in contexts of risk theory, reliability, survival analysis, economic and insurance. For instance, recently, Bartoszewicz and Skolimowska [10], Błażej [14] and Misra et al. [27] studied preservation of stochastic orders under weighting. Benduch-Frąszczak [13] investigated preservation of stochastic orders and the class of life distributions in the proportional odds family. Then, Maiti and Dey [24] applied the result of stochastic orders of [13] to the tilted normal distribution. Nanda and Das [29] studied stochastic orders in the Marshall–Olkin family. Aghababaei and Alamatsaz [3], Aghababaei

et al. [4] and Alamatsaz and Abbasi [6] were concerned with stochastic comparisons of different distributions with their mixtures.

There is no theoretical basis for choosing the baseline distribution and its tilt parameter in a Harris family distribution. Therefore, it is important to see how a Harris family rv responds to the change of the baseline distribution and tilt parameter. This paper mainly investigates how the relations between tilt parameters or baseline distributions affect stochastic orders between two given Harris family distributions. Considering the utility desired, we are able to choose a baseline distribution and the tilt parameter.

Abbasi et al. [1] compared two Harris families with different tilt parameters using stochastic orders. In this paper, we are concerned with four types of stochastic orders: simple stochastic orders, shifted stochastic orders, proportional stochastic orders and shifted proportional stochastic orders. In Section 2, we shall summarize some useful relations among stochastic orders to be used in the sequel. In Section 3, we consider a baseline distribution and compare the two corresponding Harris family distributions, with different tilt parameters, with respect to several stochastic orders. In Section 4, it is observed that certain stochastic orders of the baseline distribution are preserved by transformation to the Harris family with the same tilt parameter and *vice versa*. Finally, in Section 5 we prove that certain ageing characteristics, such as increasing failure rate average (IFRA), decreasing failure rate average (DFRA), new better than used (NBU) and new worse than used (NWU), are preserved by transformation to the Harris family. Thus, our results enfold all findings on stochastic orders of [19], [20], and [23] as special cases. In our investigations, we also reveal that Theorem 2.2 of [20] is valid only if the support of the tilt parameter is corrected. Hence, their result in Theorem 2.3 is not true as it is.

2. STOCHASTIC ORDERS AND CLASSES OF LIFE DISTRIBUTIONS

Let X and Y be rv's with df's F and G, survival functions (sf) $\bar{F}$ and $\bar{G}$, probability density functions (pdf) f and g, hazard rate functions r_F and r_G, reversed hazard rate functions $\tilde{r}_F$ $(= f(\cdot)/F(\cdot))$ and $\tilde{r}_G$ and supports S_X and S_Y, respectively. The lower and upper bounds of supports are denoted by $l_{\cdot}$ and $u_{\cdot}$. In this paper, we consider $F^{-1}(u) = \inf\{x : F(x) \leqslant u\}$, which is called the *quantile function*. Also, throughout the paper, "increasing" is used in place of "non-decreasing" and "decreasing" is used in place of "non-increasing". In what follows, some known stochastic orders and classes of life distributions, used in this article, are recalled and their important properties are stated. For more details, we refer to [28] and [31].

A. *Usual stochastic orders*

(a) X is *statistically* smaller than Y ($X \leqslant_{st} Y$) if $\bar{F}(x) \leqslant \bar{G}(x)$ for all $x \in (-\infty, \infty)$.

(b) X is smaller than Y in the *likelihood ratio order*, denoted by $X \leqslant_{lr} Y$, if $g(x)/f(x)$ increases in x over the $S_X \cup S_Y$.

(c) X is smaller than Y in the *hazard rate order*, denoted by $X \leqslant_{hr} Y$, if $r_F(x) \geqslant r_G(x)$ for all $x \in (-\infty, \infty)$.

(d) X is smaller than Y in the *reversed hazard rate order*, denoted by $X \leqslant_{rh} Y$, if $\tilde{r}_F(x) \leqslant \tilde{r}_G(x)$ for all $x \in (-\infty, \infty)$.

(e) X is smaller than Y in the *expectation order*, denoted by $X \leqslant_E Y$, if $E(X) \leqslant E(Y)$, where expectations are assumed to exist.

(f) The *mean residual life* (mrl) function of X is defined as $m(t) = E(X - t| X > t)$ for $t < t^*$, where $t^* = \sup\{t : \bar{F}(t) > 0\}$. If m and m^* are mrl functions of X and Y, respectively, then X is smaller than Y in the *mrl order*, denoted by $X \leqslant_{mrl} Y$, if $m(t) \leqslant m^*(t)$ for all t or, equivalently, if $\int_t^\infty \bar{F}(u)du / \int_t^\infty \bar{G}(u)du$ decreases in t, when defined.

(g) X is smaller than Y in the *convex order*, denoted by $X \leqslant_{cx} Y$, if for every real-valued convex function $\phi(\cdot)$ defined on the real line, $E\big(\phi(X)\big) \leqslant E\big(\phi(Y)\big)$.

(h) For non-negative rv's, X is smaller than Y in the *Lorenz order*, denoted by $X \leqslant_{Lorenz} Y$, if $L_X(p) \geqslant L_Y(p)$ for all $p \in [0, 1]$, where

$$L_X(p) = \frac{\int_0^p F^{-1}(u)du}{\int_0^1 F^{-1}(u)du}, \quad 0 \leqslant p \leqslant 1,$$

is the Lorenz curve of X.

(i) Zimmer et al. [32] defined the *log-odds function* of an rv X by

$$LO_X(t) = \ln \frac{F_X}{\bar{F}_X}$$

and introduced a new time-to-failure model based on the log-odds ratio (LOR) function. The *LOR function* of an rv X is defined by

$$LOR_X(t) = \frac{d}{dt} LO_X(t) = \frac{f(t)}{F(t)\bar{F}(t)} = \frac{r_X(t)}{F(t)}.$$

We say that X is smaller than Y in the *LOR order*, denoted by $X \leqslant_{LOR} Y$, if $l_X \leqslant l_Y, u_X \leqslant u_Y$ and $LOR_X(t) \geqslant LOR_Y(t)$ for all $t \in (l_Y, u_X)$.

(j) X is smaller than Y in the *dispersive order*, denoted by $X \leqslant_{disp} Y$, if $F^{-1}(\beta) - F^{-1}(\alpha) \leqslant G^{-1}(\beta) - G^{-1}(\alpha)$ whenever $0 < \alpha \leqslant \beta < 1$, or, equivalently, if $G^{-1}F(x) - x$ increases in x.

(k) X is smaller than Y in the *convex transform order*, denoted by $X \leqslant_c Y$, if $G^{-1}F(x)$ is convex in $x \in S_X$.

(l) For non-negative rv's, X is smaller than Y in the *star order*, denoted by $X \leqslant_* Y$, if $G^{-1}F(x)/x$ increases in $x \geqslant 0$.

(m) For non-negative rv's, X is smaller than Y in the *super-additive order*, denoted by $X \leqslant_{su} Y$, if $G^{-1}F(t + u) \geqslant G^{-1}F(t) + G^{-1}F(u)$ for $t \geqslant 0, u \geqslant 0$.

(n) X is smaller than Y in the *ageing intensity order*, denoted by $X \leqslant_{AI} Y$, if for all $x \geqslant 0$,

$$\frac{1}{r_F(x)} \int_0^x r_F(u)du \leqslant \frac{1}{r_G(x)} \int_0^x r_G(u)du.$$

B. *Shifted stochastic orders*

(o) X is smaller than Y in the *up likelihood ratio order*, denoted by $X \leqslant_{lr\uparrow} Y$, if $[X - t \mid X > t] \leqslant_{lr} Y$ for all $t \geqslant 0$ or, equivalently, if $g(x)/f(t+x)$ increases in $x \in [l_Y, u_X - t]$.

(p) X is smaller than Y in the *down likelihood ratio order*, denoted by $X \leqslant_{lr\downarrow} Y$, if $X \leqslant_{lr} [Y - t \mid Y > t]$ for all $x \geqslant 0$ or, equivalently, if $g(t+x)/f(x)$ increases in $x \in [l_X, u_Y - t]$.

(q) X is smaller than Y in the *up hazard rate order* (*up reversed hazard rate order*), denoted by $X \leqslant_{hr\uparrow} (\leqslant_{rh\uparrow}) Y$, if for all $t \geqslant 0$, $[X - t \mid X > t] \leqslant_{hr} (\leqslant_{rh}) Y$ or, equivalently, if $\bar{G}(x)/\bar{F}(t+x)$ $(G(x)/F(t+x))$ increases in $x \in (-\infty, u_Y)$ for all $t \geqslant 0$.

(r) X is smaller than Y in the *down hazard rate order* (*down reversed hazard rate order*), denoted by $X \leqslant_{hr\downarrow} (\leqslant_{rh\downarrow}) Y$, if for all $t \geqslant 0$, $X \leqslant_{hr} (\leqslant_{rh}) [Y - t \mid Y > t]$ or, equivalently, if $\bar{G}(t+x)/\bar{F}(x)$ $(G(t+x)/F(x))$ increases in $x \geqslant 0$ for all $t \geqslant 0$.

C. *Proportional stochastic orders*. Belzunce et al. [12] and Ramos Romero and Sordo Díaz [30] have introduced the proportional likelihood ratio, proportional hazard rate and proportional reversed hazard rate orders as follows. Let X and Y be continuous and non-negative rv's. Then

(s) X is smaller than Y in the *proportional likelihood ratio order* (plr) (*proportional hazard rate order* (phr), *proportional reversed hazard rate order* (prh)), denoted by $X \leqslant_{plr} (\leqslant_{phr}, \leqslant_{prh}) Y$, if for all $0 < \lambda \leqslant 1$, $\lambda X \leqslant_{lr} (\leqslant_{hr}, \leqslant_{rh}) Y$ or, equivalently, if $g(\lambda x)/f(x)$ $(\bar{G}(\lambda x)/\bar{F}(x), G(\lambda x)/F(x))$ increases in x for all $0 < \lambda \leqslant 1$.

D. *Shifted proportional stochastic orders*. Jarrahiferiz et al. [22] have introduced the shifted proportional likelihood ratio order and shifted proportional hazard rate order for continuous and non-negative rv's as follows:

(t) X is smaller than Y in the *up proportional likelihood ratio order*, denoted by $X \leqslant_{plr\uparrow} Y$, if $[X - t \mid X > t] \leqslant_{plr} Y$ or, equivalently, $g(\lambda x)/f(t+x)$ is increasing in $x \in (l_X - t, u_X - t) \cup (l_Y/\lambda, u_Y/\lambda)$ for all $t \geqslant 0$ and $0 < \lambda \leqslant 1$.

(u) X is smaller than Y in the *down proportional likelihood ratio order*, denoted by $X \leqslant_{plr\downarrow} Y$, if $X \leqslant_{plr} [Y - t \mid Y > t]$ or, equivalently, if $g(\lambda x + t)f(x)$ is increasing in $x \geqslant 0$ for all $t \geqslant 0$ and $0 < \lambda \leqslant 1$.

(v) X is smaller than Y in the *up proportional hazard rate order*, denoted by $X \leqslant_{phr\uparrow} Y$, if $[X - t \mid X > t] \leqslant_{phr} Y$ or, equivalently, if $\bar{G}(\lambda x)/\bar{F}(t+x)$ is increasing in $x \in (0, u_Y/\lambda)$ for all $t \geqslant 0$ and $0 < \lambda \leqslant 1$.

(w) X is smaller than Y in the *down proportional hazard rate order*, denoted by $X \leqslant_{phr\downarrow} Y$, if $X \leqslant_{phr} [Y - t \mid Y > t]$ or, equivalently, if $\bar{G}(\lambda x + t)/\bar{F}(x)$ is increasing in $x \geqslant 0$ for all $t \geqslant 0$ and $0 < \lambda \leqslant 1$.

E. ***Classes of life distributions***

(a) X has the *increasing likelihood ratio* (ILR) (*increasing failure rate* (IFR), *increasing reversed failure rate* $(IRFR)$) *property*, denoted by $X \in ILR$ $(IFR, IRFR)$, if

$$X \leqslant_{lr\uparrow} (\leqslant_{hr\uparrow}, \leqslant_{rh\uparrow})\, X$$

or, equivalently, if $f(x)/f(x+t)$ $(\bar{F}(x)/\bar{F}(x+t), F(x)/F(x+t))$ increases in x for any $t \geqslant 0$ and X has the *decreasing likelihood ratio* (DLR) (*decreasing failure rate* (DFR), *decreasing reversed failure rate* $(DRFR)$) *property*, denoted by $X \in DLR$ $(DFR, DRFR)$, if $X \leqslant_{lr\downarrow} (\leqslant_{hr\downarrow}, \leqslant_{rh\downarrow})\, X$ or, equivalently, if $f(x+t)/f(x)$ $(\bar{F}(x+t)/\bar{F}(x), F(x+t)/F(x))$ increases in x for any $t \geqslant 0$.

(b) X has the *increasing proportional likelihood ratio* $(IPLR)$ (*increasing proportional failure rate* $(IPFR)$, *increasing proportional reversed failure rate* $(IPRF)$) *property*, denoted by $X \in IPLR$ $(IPFR, IPRF)$, if $X \leqslant_{plr} (\leqslant_{phr}, \leqslant_{prh})\, X$ or, equivalently, if $f(\lambda x)/f(x)$ $(\bar{F}(\lambda x)/\bar{F}(x), F(\lambda x)/F(x))$ increases in x for all $0 < \lambda \leqslant 1$.

(c) X has the *up increasing proportional likelihood ratio* $(UIPLR)$ (*up increasing proportional failure rate* $(UIPFR)$) *property*, denoted by

$$X \in UIPLR\,(UIPFR),$$

if $X \leqslant_{plr\uparrow} (\leqslant_{phr\uparrow})\, X$ or, equivalently, if $f(\lambda x)/f(x+t)$ $(\bar{F}(\lambda x)/\bar{F}(x+t))$ increases in x for all $0 < \lambda \leqslant 1$ and $t \geqslant 0$ and X has the *down increasing proportional likelihood ratio* $(DIPLR)$ (*down increasing proportional failure rate* $(DIPFR)$) *property*, denoted by $X \in DIPLR$ $(DIPFR)$, if $X \leqslant_{plr\downarrow} (\leqslant_{phr\downarrow})\, X$ or, equivalently, if $f(\lambda x + t)/f(x)$ $(\bar{F}(\lambda x + t)/\bar{F}(x))$ increases in x for all $0 < \lambda \leqslant 1$ and $t \geqslant 0$.

TABLE 1. Some useful relations among various types of stochastic orders.

$$
\begin{array}{ccccccccccccccc}
\leqslant_{lr} & \Rightarrow & \leqslant_{hr} & \Rightarrow & \leqslant_{st} & & \leqslant_{hr\uparrow} & \Rightarrow & \leqslant_{hr} & \Rightarrow & \leqslant_{st} & & & & \\
\Uparrow & & & & & & \Uparrow & & & & & & & & \\
\leqslant_{plr\uparrow} & \Rightarrow & \leqslant_{prh\uparrow} & \Rightarrow & \leqslant_{rh\uparrow} & \Leftarrow & \leqslant_{lr\uparrow} & \Leftarrow & \leqslant_{plr\uparrow} & \Rightarrow & \leqslant_{phr\uparrow} & \Rightarrow & \leqslant_{hr\uparrow} & & \\
\Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \Downarrow & & \\
\leqslant_{plr} & \Rightarrow & \leqslant_{prh} & \Rightarrow & \leqslant_{rh} & \Leftarrow & \leqslant_{lr} & \Leftarrow & \leqslant_{plr} & \Rightarrow & \leqslant_{phr} & \Rightarrow & \leqslant_{hr} & & \\
\Uparrow & & \Uparrow & & \Uparrow & & \Uparrow & & \Uparrow & & \Uparrow & & \Uparrow & & \\
\leqslant_{plr\downarrow} & & \leqslant_{prh\downarrow} & \Rightarrow & \leqslant_{rh\downarrow} & & \leqslant_{lr\downarrow} & \Leftarrow & \leqslant_{plr\downarrow} & \Rightarrow & \leqslant_{phr\downarrow} & \Rightarrow & \leqslant_{hr\downarrow} & & \\
 & & & & \Downarrow & & \Downarrow & & & & & & & & \\
 & & \leqslant_{st} & \Leftarrow & \leqslant_{rh} & & \leqslant_{hr\downarrow} & & & & & & & &
\end{array}
$$

(d) A non-negative rv X has IFRA (DFRA) if $\left(-\frac{1}{t}\right)\ln\bar{F}(t)$ is increasing (decreasing) in $t \geqslant 0$.

(e) A non-negative rv X is NBU (NWU) if $\bar{F}(t+u) \leqslant (\geqslant)\ \bar{F}(t)\bar{F}(u)$ for $t \geqslant 0$ and $u \geqslant 0$.

In Table 1, we summarize some useful relationships among several stochastic orders to be used in the sequel.

3. STOCHASTIC COMPARISON

Assume that the baseline df $F(x)$ in (1.1) is absolutely continuous with pdf $f(x)$. Then, the pdf and df associated with $\bar{H}(x;\theta,k)$ in (1.1) are given by

$$h(x;\theta,k) = \frac{\theta^{1/k} f(x)}{\left(1-\bar{\theta}\bar{F}^k(x)\right)^{1+1/k}}, \tag{3.1}$$
$$-\infty < x < \infty,\ 0 < \theta < \infty,\ \bar{\theta} = 1-\theta,\ k > 0,$$

and

$$H(x;\theta,k) = 1 - \left[\frac{\theta\bar{F}^k(x)}{\left(1-\bar{\theta}\bar{F}^k(x)\right)}\right]^{1/k}, \tag{3.2}$$
$$-\infty < x < \infty,\ 0 < \theta < \infty,\ \bar{\theta} = 1-\theta,\ k > 0,$$

respectively.

Batsidis and Lemonte [11] in their Proposition 2 compared a Harris family distribution with its corresponding baseline distribution with respect to several stochastic and shifted stochastic orders. In the following theorem, we compare two Harris families with respect to their tilt parameter θ.

THEOREM 3.1. *Let X, Y_1 and Y_2 be continuous and non-negative rv's corresponding to survival functions $\bar{F}(\cdot)$, $\bar{H}(\cdot;\theta_1,k_1)$ and $\bar{H}(\cdot;\theta_2,k_2)$, respectively. Moreover, let $\{0 < \theta_1 \leqslant 1, \theta_2 \geqslant 1\}$. Then*:

(i) *If $X \in UIPLR$ $(IPLR, ILR)$, then $Y_1 \leqslant_{plr\uparrow} (\leqslant_{plr}, \leqslant_{lr\uparrow})\ Y_2$.*

(ii) *If $X \in DIPLR$ (DLR), then $Y_1 \leqslant_{plr\downarrow} (\leqslant_{lr\downarrow})\ Y_2$.*

(iii) *If $X \in UIPFR$ $(IPFR, IFR)$, then $Y_1 \leqslant_{phr\uparrow} (\leqslant_{phr}, \leqslant_{hr\uparrow})\ Y_2$.*

(iv) *If $X \in DIPFR$ (DFR), then $Y_1 \leqslant_{phr\downarrow} (\leqslant_{hr\downarrow})\ Y_2$.*

P r o o f. We give the proof for the first part. Proofs of other parts are similar and thus omitted. Let $\{0 < \theta_1 \leqslant 1, \theta_2 \geqslant 1\}$ and $X \in UIPLR$. For $Y_1 \leqslant_{plr\uparrow} Y_2$, it is sufficient to show that

$$\frac{h(\lambda x;\theta_2,k_2)}{h(x+t;\theta_1,k_1)} = \frac{\theta_2^{1/k_2}}{\theta_1^{1/k_1}}\frac{f(\lambda x)}{f(x+t)}\left[\frac{\left(1-\bar{\theta}_1\bar{F}^{k_1}(x+t)\right)^{1/k_1+1}}{\left(1-\bar{\theta}_2\bar{F}^{k_2}(\lambda x)\right)^{1/k_2+1}}\right]$$

is increasing in x for any $0 < \lambda \leqslant 1$, $t \geqslant 0$ and $k_1, k_2 > 0$. Since $X \in UIPLR$, $f(\lambda x)/f(x+t)$ is increasing in x for any $0 < \lambda \leqslant 1$ and $t \geqslant 0$. Also the term in the brackets is increasing in x because

$$\frac{d}{dx}\left[\frac{\big(1-\bar{\theta}_1\bar{F}^{k_1}(x+t)\big)^{1/k_1+1}}{\big(1-\bar{\theta}_2\bar{F}^{k_2}(\lambda x)\big)^{1/k_2+1}}\right] = \left[\frac{\big(1-\bar{\theta}_1\bar{F}^{k_1}(x+t)\big)^{1/k_1+1}}{\big(1-\bar{\theta}_2\bar{F}^{k_2}(\lambda x)\big)^{1/k_2+1}}\right]$$
$$\times\left[\frac{\bar{\theta}_1(k_1+1)f(x+t)\bar{F}^{(k_1-1)}(x+t)}{1-\bar{\theta}_1\bar{F}^{k_1}(x+t)} - \frac{\lambda\bar{\theta}_2(k_2+1)f(\lambda x)\bar{F}^{(k_2-1)}(\lambda x)}{1-\bar{\theta}_2\bar{F}^{k_2}(\lambda x)}\right]$$

is non-negative provided that $\{0 < \theta_1 \leqslant 1, \theta_2 \geqslant 1\}$. Thus, we have the assertion. Our proof above also yields $Y_1 \leqslant_{plr} Y_2$, by putting $t = 0$, and $Y_1 \leqslant_{lr\uparrow} Y_2$, by letting $\lambda = 1$. ∎

THEOREM 3.2. *Let Y_1 and Y_2 be rv's corresponding to the df's $H(\cdot;\theta_1,k_1)$ and $H(\cdot;\theta_2,k_2)$, respectively. If $\{0 < \theta_1 \leqslant 1, \theta_2 \geqslant 1\}$ or $\{0 < \theta_1 \leqslant \theta_2, k_1 = k_2 = k\}$, then $Y_1 \leqslant_{lr} Y_2$.*

Proof. $Y_1 \leqslant_{lr} Y_2$ is equivalent to $h(x;\theta_1,k_1)/h(x;\theta_2,k_2)$ being decreasing in x. But, by equation (3.1), we have

$$\frac{h(x;\theta_1,k_1)}{h(x;\theta_2,k_2)} = \left(\frac{\theta_1^{1/k_1}}{\theta_2^{1/k_2}}\right)\frac{\big[1-\bar{\theta}_2\bar{F}^{k_2}(x)\big]^{1/k_2+1}}{\big[1-\bar{\theta}_1\bar{F}^{k_1}(x)\big]^{1/k_1+1}}.$$

Thus, for any $k_1 > 0$ and $k_2 > 0$ we obtain

$$\frac{d}{dx}\left[\frac{h(x;\theta_1,k_1)}{h(x;\theta_2,k_2)}\right]$$
$$= \frac{h(x;\theta_1,k_1)}{h(x;\theta_2,k_2)}f(x)\left[\frac{(k_2+1)\bar{\theta}_2\bar{F}^{k_2-1}(x)}{1-\bar{\theta}_2\bar{F}^{k_2}(x)} - \frac{(k_1+1)\bar{\theta}_1\bar{F}^{k_1-1}(x)}{1-\bar{\theta}_1\bar{F}^{k_1}(x)}\right]$$

which is non-positive if $\{0 < \theta_1 \leqslant 1, \theta_2 \geqslant 1\}$.

For $k_1 = k_2 = k$, by equation (3.1), we have

$$\frac{h(x;\theta_1,k)}{h(x;\theta_2,k)} = \left(\frac{\theta_1}{\theta_2}\right)^{1/k}\left[\frac{1-\bar{\theta}_2\bar{F}^{k}(x)}{1-\bar{\theta}_1\bar{F}^{k}(x)}\right]^{1+1/k}.$$

Thus, for all $k > 0$ we obtain

$$\frac{d}{dx}\left[\frac{h(x;\theta_1,k)}{h(x;\theta_2,k)}\right] = C(x;k,\theta_1,\theta_2)\left[\frac{1-\bar{\theta}_2\bar{F}^{k}(x)}{1-\bar{\theta}_1\bar{F}^{k}(x)}\right]^{1/k}\frac{\bar{\theta}_2-\bar{\theta}_1}{\big(1-\bar{\theta}_1\bar{F}^{k}(x)\big)^2},$$

where $C(x;k,\theta_1,\theta_2) = (\theta_1/\theta_2)^{1/k}(1+k)f(x)\bar{F}^{k-1}(x) \geqslant 0$ is non-positive if $\theta_1 \leqslant \theta_2$. This completes the proof. ∎

By Theorem 3.2 and Table 1, we immediately obtain

COROLLARY 3.1. *Let Y_1 and Y_2 be rv's corresponding to df's $H(\cdot;\theta_1,k_1)$ and $H(\cdot;\theta_2,k_2)$, respectively. If $\{0<\theta_1\leqslant 1,\theta_2\geqslant 1\}$ or $\{0<\theta_1\leqslant\theta_2,k_1=k_2=k\}$, then $Y_1\leqslant_{hr}(\leqslant_{rh},\leqslant_{st},\leqslant_E)\,Y_2$.*

REMARK 3.1. *It is worth mentioning that, in view of our Theorem* 3.2, *Theorem* 2.3 *of* [20] *concerning the Marshall–Olkin family is not valid unless $\theta_1\geqslant\theta_2$ is replaced by $\theta_2\geqslant\theta_1$.*

REMARK 3.2. *Our results in Theorem* 3.2 *can be viewed as extensions of those of Theorem* 3 *of* [13], *Theorem* 4 *of* [16] *and Proposition* 1 *of* [17], *where they consider the special case of $k=1$, i.e., the Marshall–Olkin family. Furthermore, our result in Corollary* 3.1 *for $k=1$ was proved by Benduch-Frąszczak* [13] *in Corollary* 2.

In the following theorem we study ageing intensity orders between rv's Y_1 and Y_2 corresponding to df's $H(\cdot;\theta_1,k)$ and $H(\cdot;\theta_2,k)$, respectively.

THEOREM 3.3. *Let Y_1 and Y_2 be rv's corresponding to Harris family df's $H(\cdot;\theta_1,k)$ and $H(\cdot;\theta_2,k)$, respectively. Then $Y_1\leqslant_{AI}Y_2$ provided that $\theta_1>\theta_2$.*

P r o o f. $Y_1\leqslant_{AI}Y_2$ if and only if, for all $x>0$, we have

$$\frac{1}{r_H(x;\theta_1,k)}\int_0^x r_H(u;\theta_1,k)du\leqslant\frac{1}{r_H(x;\theta_2,k)}\int_0^x r_H(u;\theta_2,k)du,\quad k>0,$$

or, by equation (1.2),

$$\frac{1-\bar\theta_1\bar F^k(x)}{r_F(x)}\int_0^1\frac{r_F(u)}{1-\bar\theta_1\bar F^k(u)}du\leqslant\frac{1-\bar\theta_2\bar F^k(x)}{r_F(x)}\int_0^x\frac{r_F(u)}{1-\bar\theta_2\bar F^k(u)}du,\quad k>0,$$

which is equivalent to

$$\int_0^x r_F(u)\left[\frac{1-\bar\theta_1\bar F^k(x)}{1-\bar\theta_1\bar F^k(u)}-\frac{1-\bar\theta_2\bar F^k(x)}{1-\bar\theta_2\bar F^k(u)}\right]du\geqslant 0,\quad k>0.$$

But this is true if $\theta_1>\theta_2$ because

$$\frac{d}{d\theta}\left(\frac{1-\bar\theta\bar F^k(x)}{1-\bar\theta\bar F^k(u)}\right)=\frac{\bar F^k(x)-\bar F^k(u)}{\big(1-\bar\theta\bar F^k(u)\big)^2}\leqslant 0,$$

or if $\big(1-\bar\theta\bar F^k(x)\big)/\big(1-\bar\theta\bar F^k(u)\big)$ is decreasing in θ when $0<u<x$. Thus, we have the result. ∎

4. PRESERVATION OF STOCHASTIC ORDERS BY HARRIS FAMILY WITH THE SAME TILT PARAMETERS

Let X_1 and X_2 be two rv's with df's F_1 and F_2 and pdf's f_1 and f_2, respectively. Suppose that Y_1 and Y_2 are their corresponding Harris family rv's, i.e., the df's F_1 and F_2 with baseline, respectively. In this section, we shall study several stochastic order preservations of the baseline distribution by its corresponding Harris family.

Kirmani and Gupta [23] have shown that usual stochastic, hazard rate, convex transform, super-additive and star orders are preserved by transformation to proportional odds ratio (Marshall–Olkin) family. In what follows, their results are generalized to Harris family, i.e., for any $k > 0$ in equation (1.1). In fact, more generally, we have the following necessary and sufficient property.

THEOREM 4.1. *$X_1 \leqslant_{st} X_2$ if and only if $Y_1 \leqslant_{st} Y_2$.*

P r o o f. It is true by Theorem 3.1 of [1] when $\alpha = \beta$. ∎

Since the Harris family of distributions coincides with weighted distributions, with weight $\omega(x) = \theta^{1/k}/\big(1 - \bar{\theta}\bar{F}^k(x)\big)^{1/k+1}$, by Theorem 9(a) of [9] we conclude that the hazard rate order is preserved by transformation to the Harris family. The following theorem also provides a both-sided preservation for different types of hazard rate orders. That is, by comparing lifetimes of two given systems, we can detect which one is made of better quality components. But, in these cases, the range of the tilt parameter values plays a restrictive role.

THEOREM 4.2. (i) *Assume that $\theta \geqslant 1$. If $X_1 \leqslant_{phr\uparrow} (\leqslant_{phr}, \leqslant_{hr\uparrow}, \leqslant_{hr}) X_2$, then $Y_1 \leqslant_{phr\uparrow} (\leqslant_{phr}, \leqslant_{hr\uparrow}, \leqslant_{hr})\, Y_2$.*

(ii) *Assume that $0 < \theta \leqslant 1$. If $Y_1 \leqslant_{phr\uparrow} (\leqslant_{phr}, \leqslant_{hr\uparrow}, \leqslant_{hr})\, Y_2$, then $X_1 \leqslant_{phr\uparrow} (\leqslant_{phr}, \leqslant_{hr\uparrow}, \leqslant_{hr})\, X_2$.*

P r o o f. (i) It is true by Theorem 3.2(i) of [1] when $\alpha = \beta \geqslant 1$.

(ii) For the up proportional hazard rate order, let $Y_1 \leqslant_{phr\uparrow} Y_2$. So, for all x, $t \geqslant 0$ and $0 < \lambda \leqslant 1$ we have $r_{H_1}(x + t; \theta, k) \geqslant \lambda r_{H_2}(\lambda x; \theta, k)$. So, by equation (1.2), we have

$$\frac{r_{F_1}(x+t)}{\lambda r_{F_2}(\lambda x)} \geqslant \frac{1 - \bar{\theta}\bar{F}_1^k(x+t)}{1 - \bar{\theta}\bar{F}_2^k(\lambda x)}. \tag{4.1}$$

Since the hazard rate order is implied by the up proportional hazard rate order (Table 1) and the simple stochastic order is implied by the hazard rate order, for any x and all $k > 0$ we have $\bar{H}_1^k(x) \leqslant \bar{H}_2^k(x)$. Also, by Theorem 4.1, $\bar{F}_1^k(x) \leqslant \bar{F}_2^k(x)$. Further, the survival function is decreasing, so for all $0 < \lambda \leqslant 1$, $t \geqslant 0$, $k > 0$ and x, we get

$$\bar{F}_1^k(x+t) \leqslant \bar{F}_1^k(x) \leqslant \bar{F}_2^k(x) \leqslant \bar{F}_2^k(\lambda x).$$

Thus, when $0 < \theta < 1$, we have

$$-\bar{\theta}\bar{F}_1^k(x+t) \geqslant -\bar{\theta}\bar{F}_2^k(\lambda x) \Longrightarrow 1 - \bar{\theta}\bar{F}_1^k(x+t) \geqslant 1 - \bar{\theta}\bar{F}_2^k(\lambda x).$$

Consequently, the right-hand side of inequality (4.1) is greater than one, which implies $r_{F_1}(x+t) \geqslant \lambda r_{F_2}(\lambda x)$, i.e., $X_1 \leqslant_{phr\uparrow} X_2$, as required.

With proper choices of t or λ, i.e. $t = 0$ or $\lambda = 1$, or both, proofs for the other parts are immediate. ∎

By using the counterexample 3.2 of [1], the following counterexample shows that the up hazard rate order is not preserved by transformation to the Harris family, when $0 < \theta < 1$.

COUNTEREXAMPLE 4.1. Let X_1 and X_2 be two rv's having the Erlang distributions with survival functions $\bar{F}_1(x) = (1+2x)e^{-2x}$, $\bar{F}_2(x) = (x+1)e^{-x}$ and hazard rates $r_{F_1}(x) = 4x/(1+2x)$, $r_{F_2}(x) = x/(x+1)$, for $x > 0$, respectively. So, $X_1 \leqslant_{hr\uparrow} X_2$. However, Figure 1 shows that for some $0 < \theta < 1$, $t > 0$ and some $x > 0$, $r_{H_1}(x+t;\theta,k) \not> r_{H_2}(x;\theta,k)$ or, equivalently, $\bar{H}_2(x;\theta,k)/\bar{H}_1(x+t;\theta,k)$ is not increasing in x, i.e., the up hazard rate order is not preserved by transformation to the Harris family when $0 < \theta < 1$.

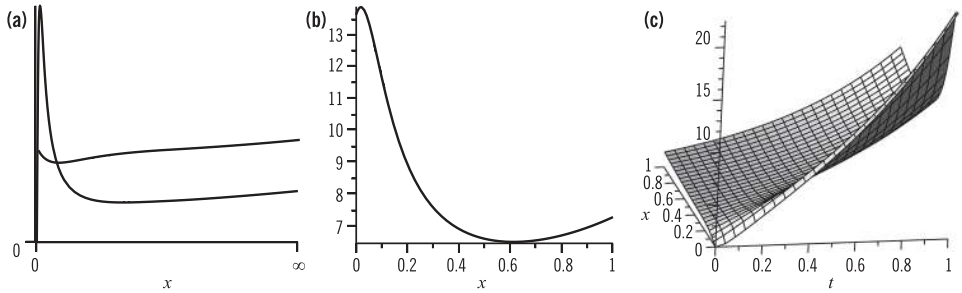


FIGURE 1. (a) showing that $r_{H_1}(x+t;\theta,k) \not\geqslant r_{H_2}(x;\theta,k)$, and (b) and (c) showing that $\bar{H}_2(x;\theta,k)/\bar{H}_1(x+t;\theta,k)$ is not increasing in x.

COROLLARY 4.1. *Let X_1 and X_2 be two rv's with mean residual life (mrl) functions m_1 and m_2 and Harris family rv's Y_1 and Y_2 having mrl functions m_1^* and m_2^*, respectively, such that $m_1(t)/m_2(t)$ increases in t. If $X_1 \leqslant_{mrl} X_2$, then $Y_1 \leqslant_{mrl} Y_2$ provided that $\theta \geqslant 1$. The orders are reversed if $m_1^*(t)/m_2^*(t)$ increases in t and $0 < \theta \leqslant 1$.*

P r o o f. By Theorem 2.A.2 of [31], the assertion holds because if $X_1 \leqslant_{mrl} X_2$ and $m_1(t)/m_2(t)$ increases in t, then $X_1 \leqslant_{hr} X_2$. Thus, by Theorem 4.2(i) we can conclude that $Y_1 \leqslant_{hr} Y_2$. But by the sufficiency of the hazard rate order for mrl order (Theorem 1.D.1 of [31]), this implies that $Y_1 \leqslant_{mrl} Y_2$. ∎

REMARK 4.1. *Note that for the special case when $k = 1$, the log-odds function of an rv X is equal to the log-odds function of the corresponding Harris family rv Y. Consequently, the log-odds ratio order is also preserved by transformation to the Marshall–Olkin family.*

For the ageing intensity order, we have the following

THEOREM 4.3. *Assume that X_1 and X_2 are non-negative rv's. For all $k > 0$, if $X_1 \leqslant_{AI} X_2$ and $X_1 \leqslant_{hr} X_2$, then $Y_1 \leqslant_{AI} Y_2$ provided that $\theta > 1$.*

Proof. Let $k > 0$ and $\theta > 1$. $Y_1 \leqslant_{AI} Y_2$ if and only if

$$\frac{1}{r_{H_1}(x;\theta,k)} \int_0^x r_{H_1}(u;\theta,k)du \leqslant \frac{1}{r_{H_2}(x;\theta,k)} \int_0^x r_{H_2}(u;\theta,k)du$$

or

$$\frac{1-\bar\theta\bar F_1^k(x)}{r_{F_1}(x)} \int_0^x r_{H_1}(u;\theta,k)du \leqslant \frac{1-\bar\theta\bar F_2^k(x)}{r_{F_2}(x)} \int_0^x r_{H_2}(u;\theta,k)du$$

But we have

$$\int_0^x r_H(u;\theta,k)du = -\ln \bar H(x;\theta,k) = -\ln \bar F(x) + \frac{1}{k}\ln\left(\frac{1-\bar\theta\bar F^k(x)}{\theta}\right).$$

So, we should show that

$$\begin{aligned}(4.2)\quad &\big(1-\bar\theta\bar F_1^k(x)\big)\left[\frac{-\ln \bar F_1(x)}{r_{F_1}(x)} + \frac{1}{k}\frac{\ln\big(\big(1-\bar\theta\bar F_1^k(x)\big)/\theta\big)}{r_{F_1}(x)}\right]\\ &\qquad\leqslant \big(1-\bar\theta\bar F_2^k(x)\big)\left[\frac{-\ln \bar F_2(x)}{r_{F_2}(x)} + \frac{1}{k}\frac{\ln\big(\big(1-\bar\theta\bar F_2^k(x)\big)/\theta\big)}{r_{F_2}(x)}\right].\end{aligned}$$

Since $X_1 \leqslant_{AI} X_2$, we also have

$$\frac{1}{r_{F_1}(x)} \int_0^x r_{F_1}(u)du \leqslant \frac{1}{r_{F_2}(x)} \int_0^x r_{F_2}(u)du$$

or

$$\frac{1}{r_{F_1}(x)} \int_0^x \frac{f_1(u)}{\bar F_1(u)}du \leqslant \frac{1}{r_{F_2}(x)} \int_0^x \frac{f_2(u)}{\bar F_2(u)}du.$$

Equivalently, we have

$$(4.3)\qquad \frac{-\ln \bar F_1(x)}{r_{F_1}(x)} \leqslant \frac{-\ln \bar F_2(x)}{r_{F_2}(x)}.$$

On the other hand, if $X_1 \leqslant_{hr} X_2$, for all x we have $1/r_{F_1}(x) \leqslant 1/r_{F_2}(x)$ and also $X_1 \leqslant_{st} X_2$. Thus, $\bar F_1^k(x) \leqslant \bar F_2^k(x)$. So, since $\theta > 1$, we have

$$\frac{1-\bar\theta\bar F_1^k(x)}{\theta} \leqslant \frac{1-\bar\theta\bar F_2^k(x)}{\theta}.$$

Hence, we can conclude that

$$\frac{\ln\big(\big(1-\bar{\theta}\bar{F}_1^k(x)\big)/\theta\big)}{r_{F_1}(x)} \leqslant \frac{\ln\big(\big(1-\bar{\theta}\bar{F}_2^k(x)\big)/\theta\big)}{r_{F_2}(x)}. \tag{4.4}$$

Now, adding up inequalities (4.3) and (4.4) and multiplying the left-hand side by $\big(1-\bar{\theta}\bar{F}_1^k(x)\big)$ and the right-hand side by $\big(1-\bar{\theta}\bar{F}_2^k(x)\big)$, we obtain inequality (4.2). This completes the proof. ∎

In the next lemma we need inverses of the df and survival function of a Harris family distribution. It is easy to verify that equations (1.1) and (3.2) lead to

$$\bar{H}^{-1}(p;\theta,k) = \bar{F}^{-1}\left(\frac{p^k}{\theta+\bar{\theta}p^k}\right)^{1/k}, \quad 0<p<1, \tag{4.5}$$

and

$$H^{-1}(p;\theta,k) = F^{-1}\left(1-\left[\frac{(1-p)^k}{\theta+\bar{\theta}(1-p)^k}\right]^{1/k}\right), \quad 0<p<1. \tag{4.6}$$

Equation (4.6) was observed by Batsidis and Lemonte [11].

LEMMA 4.1. *If $H_1(x) \equiv H_1(x;\theta,k)$ and $H_2(x) \equiv H_2(x;\theta,k)$ are two Harris family df's with baseline df's F_1 and F_2, respectively, then $H_2^{-1}\big(H_1(x)\big) = F_2^{-1}\big(F_1(x)\big)$ for all x.*

P r o o f. This result can be obtained by using the assumed form of H_1 together with H_2^{-1}, which follows from equation (4.6). For any $k>0$ and $\theta>0$, we have

$$H_2^{-1}\big(H_1(x)\big) = F_2^{-1}\left(1-\left[\frac{\big(1-H_1(x)\big)^k}{\theta+\bar{\theta}\big(1-H_1(x)\big)^k}\right]^{1/k}\right). \tag{4.7}$$

Thus, substituting $H_1(x)$ of equation (3.2) into (4.7), we obtain the lemma. ∎

Without any restriction on the tilt parameter θ, we have

THEOREM 4.4. *The following orders are preserved by transformation from a baseline distribution to its corresponding Harris family and vice versa.*

(i) *convex transform order*,

(ii) *star order*,

(iii) *supper-additive order*,

(iv) *dispersive order.*

P r o o f. (i) $X_1 \leqslant_c X_2$ holds if $F_2^{-1}F_1(x)$ is convex in $x \in S_{X_1}$. Thus, by Lemma 4.1, $H_2^{-1}\big(H_1(x)\big)$ is also convex in $x \in S_{Y_1}$. So $Y_1 \leqslant_c Y_2$.

(ii) $X_1 \leqslant_* X_2$ holds if $F_2^{-1}F_1(x)/x$ increases in $x \geqslant 0$. Thus, by Lemma 4.1, $H_2^{-1}\big(H_1(x)\big)/x$ also increases in $x \geqslant 0$. So $Y_1 \leqslant_* Y_2$.

(iii) $X_1 \leqslant_{su} X_2$ if $F_2^{-1}F_1(t+u) \geqslant F_2^{-1}F_1(t) + F_2^{-1}F_1(u)$ for $t \geqslant 0$ and $u \geqslant 0$. Thus, by Lemma 4.1, $H_2^{-1}H_1(t+u) \geqslant H_2^{-1}H_1(t) + H_2^{-1}H_1(u)$ for $t \geqslant 0$ and $u \geqslant 0$. So $Y_1 \leqslant_{su} Y_2$.

(iv)$X_1 \leqslant_{disp} X_2$ if $F_2^{-1}F_1(x) - x$ increases in x. Thus, by Lemma 4.1, it follows that $H_2^{-1}\big(H_1(x)\big) - x$ also increases in x. So $Y_1 \leqslant_{disp} Y_2$.

Proofs of converse transformations are similar. ∎

COROLLARY 4.2. *If $X_1 \leqslant_{Lorenz} X_2$, then $Y_1 \leqslant_{Lorenz} Y_2$ provided that the function $F_2^{-1}(x)/F_1^{-1}(x)$ is increasing for all $x > 0$.*

P r o o f. If $F_2^{-1}(x)/F_1^{-1}(x)$ is increasing for all $x > 0$, then, clearly, it follows that $F_2^{-1}F_1(x)/x$ is also increasing for all $x > 0$. Thus, $X_1 \leqslant_* X_2$ and the Lorenz order is implied by the star order (cf. [9], p. 90), i.e., $X_1 \leqslant_{Lorenz} X_2$. Since, by Theorem 4.4, the star order is preserved by transformation to the Harris family, we have $Y_1 \leqslant_* Y_2$, which yields the Lorenz order, as required. ∎

REMARK 4.2. *The usual stochastic*, *hazard rate*, *convex transform and star orders are preserved by transformation to the frailty family* (*proportional hazard family*, cf. [26], p. 240) *and to Marshall–Olkin family* (cf. [23]). *By combining these facts with Remark* 1 *of* [11], *it follows that such orders are also preserved under transformation to the Harris family.*

5. AGEING PROPERTIES

In the investigations pertaining to ageing concepts, the problem is to examine how a component or system improves or deteriorates with age. In the reliability context, life distributions are classified into different classes based on the monotonic behavior of the failure rate and mean residual life functions. The works [2], [5] and [18] proceed in this direction. Batsidis and Lemonte [11] showed that IFR and DFR properties are preserved by transformation to the Harris family. In what follows, we shall show that the ageing characteristics, i.e., IFRA, DFRA, NBU and NWU, are also preserved by transformation to the Harris family. First, we need to recall some results.

PROPOSITION 5.1 ([26], p. 182). *The following two statements are equivalent*:

(i) *X has IFRA* (*DFRA*),

(ii) *$X \leqslant_* (\geqslant_*) X_1$, where X_1 has an exponential distribution.*

PROPOSITION 5.2 ([26], p. 182). *The following two statements are equivalent*:
(i) X *is NBU* (*NWU*),
(ii) $X \leqslant_{su} (\geqslant_{su}) X_1$, *where* X_1 *has an exponential distribution.*

In the following corollary we shall investigate preservation of the IFRA, DFRA, NBU and NWU characteristics by transformation to the Harris family.

COROLLARY 5.1. *Let* $\theta > 1$ $(0 < \theta < 1)$.
(i) *The IFRA* (*DFRA*) *characteristic is preserved by transformation to the Harris family.*
(ii) *The NBU* (*NWU*) *characteristic is preserved by transformation to the Harris family.*

P r o o f. (i) Assume that an rv X has the IFRA (DFRA) property and that X_1 is an rv with survival function $\bar{F}_1(x) = e^{-x}$ for $x \geqslant 0$. We transform $\bar{F}_1(\cdot)$ to the Harris family as follows:

$$\bar{H}_1(x;\theta,k) = \frac{\theta^{1/k} e^{-x}}{(1-\bar{\theta}e^{-kx})^{1/k}}, \quad x \geqslant 0.$$

Let Y and Y_1 be the corresponding Harris family rv's with survival functions $\bar{H}(\cdot;\theta,k)$ and $\bar{H}_1(\cdot;\theta,k)$, respectively. By Proposition 5.1, we get $X \leqslant_* (\geqslant_*) X_1$. But, by Theorem 4.4(ii), the star order is preserved by transformation to the Harris family, so we have $Y \leqslant_* (\geqslant_*) Y_1$. Thus, by [10], for $\theta > 1$ $(0 < \theta < 1)$, Y_1 has the IFR (DFR) property. Moreover, the IFR (DFR) property implies the IFRA (DFRA) property (cf. [26], p. 181). Thus, Y_1 has IFRA (DFRA), and so, by Proposition 5.1, $Y_1 \leqslant_* (\geqslant_*) X_1$. From the transitivity property of partial order, we obtain $Y \leqslant_* (\geqslant_*) X_1$. Thus, by Proposition 5.1, Y has the IFRA (DFRA) property.

(ii) Let an rv X with survival function $\bar{F}(\cdot)$ be NBU (NWU) and X_1 be an rv with survival function $\bar{F}_1(x) = e^{-x}$ for $x \geqslant 0$. We transform $\bar{F}_1(\cdot)$ to the Harris family as follows:

$$\bar{H}_1(x;\theta,k) = \frac{\theta^{1/k} e^{-x}}{(1-\bar{\theta}e^{-kx})^{1/k}}, \quad x \geqslant 0.$$

Let Y and Y_1 be rv's with survival functions defined in (1.1) and $\bar{H}_1(\cdot;\theta,k)$, respectively. By Proposition 5.2, $X \leqslant_{su} (\geqslant_{su}) X_1$, but by Theorem 4.4(iii), the super-additive order is preserved by transformation to the Harris family. Thus, we have $Y \leqslant_{su} (\geqslant_{su}) Y_1$. For $\theta > 1$ $(0 < \theta < 1)$, it can be easily shown that Y_1 is NBU (NWU). Then, by Proposition 5.2, $Y_1 \leqslant_{su} (\geqslant_{su}) X_1$. Due to the transitivity property of partial order, this implies that $Y \leqslant_{su} (\geqslant_{su}) X_1$. Thus, by Proposition 2, Y is NBU (NWU). ∎

REMARK 5.1. *Since the Harris family of distributions coincides with weighted distributions with weight*

$$\omega(x) = \frac{\theta^{1/k}}{\left(1 - \bar{\theta}\bar{F}^k(x)\right)^{1/k+1}},$$

the above corollary is a consequence of Theorem 3 *of* [10] *and Theorem* 3 *of* [14]. *Note that, by Theorem* 3 *of* [10], *for the IFRA and NBU characteristics we should let* $\omega(x)\bar{F}(x)$ *be increasing in* x, *but in our Corollary* 5.1 *we have a larger class of* θ *values with no restriction on* k *and* x.

6. DISCUSSION AND CONCLUSION

The hazard and lifetime in a series system with variable number of components, model (1.1), are functions of a tilt parameter. So, a proper choice of the range of values of this parameter plays an important role in optimization of the systems lifetime. In Section 3, we indicated how a lower risk (hazard rate order), longer lifetime (usual stochastic order), higher likelihood ratio (likelihood ratio order), etc. can be achieved by a system comparing to its components by a proper choice of the tilt parameter values. In Section 3, we also discussed how one can distinct the optimum case of two systems using their tilt parameters. Section 4 determined when a stochastic order between components is preserved by their corresponding systems and, more interestingly, *vice versa* for the cases in which components are not observable. Finally, in Section 5, we revealed when the ageing properties IFRA, DFRA, NBU and NWU of components are transferred to their corresponding systems.

Acknowledgments. The authors are grateful to two anonymous referees whose valuable comments and suggestions led to the improvement of the manuscript. The first author is also grateful to the office of Graduate Studies of the University of Isfahan for their support.

REFERENCES

[1] S. Abbasi, M. H. Alamatsaz, and E. Cramer, *Preservation properties of stochastic orderings by transformation to Harris family with different tilt parameters*, ALEA Lat. Am. J. Probab. Math. Stat. 13 (1) (2016), pp. 465–479.

[2] A. Abouammoh and E. El-Neweihi, *Closure of the NBUE and DMRL classes under formation of parallel systems*, Statist. Probab. Lett. 4 (5) (1986), pp. 223–225.

[3] M. Aghababaei Jazi and M. H. Alamatsaz, *Ordering comparison of logarithmic series random variables with their mixtures*, Comm. Statist. Theory Methods 39 (18) (2010), pp. 3252–3263.

[4] M. Aghababaei Jazi, M. H. Alamatsaz, and S. Abbasi, *A unified approach to ordering comparison of GPS distributions with their mixtures*, Comm. Statist. Theory Methods 40 (14) (2011), pp. 2591–2604.

[5] I. A. Ahmad and A. R. Mugdadi, *Further moments inequalities of life distributions with hypothesis testing applications: The IFRA, NBUC and DMRL classes*, J. Statist. Plann. Inference 120 (1–2) (2004), pp. 1–12.
[6] M. H. Alamatsaz and S. Abbasi, *Ordering comparison of negative binomial random variables with their mixtures*, Statist. Probab. Lett. 78 (14) (2008), pp. 2234–2239.
[7] R. A. Al-Jarallah, M. E. Ghitany, and R. C. Gupta, *A proportional hazard Marshall–Olkin extended family of distributions and its application to Gompertz distribution*, Comm. Statist. Theory Methods 43 (20) (2014), pp. 4428–4443.
[8] E. A. A. Aly and L. Benkherouf, *A new family of distributions based on probability generating functions*, Sankhya B 73 (1) (2011), pp. 70–80.
[9] B. C. Arnold, *Majorization and the Lorenz Order: A Brief Introduction*, Springer, Berlin 1987.
[10] J. Bartoszewicz and M. Skolimowska, *Preservation of classes of life distributions and stochastic orders under weighting*, Statist. Probab. Lett. 76 (6) (2006), pp. 587–596.
[11] A. Batsidis and A. J. Lemonte, *On the Harris extended family of distributions*, Statistics 49 (6) (2015), pp. 1400–1421.
[12] F. Belzunce, J. M. Ruiz, and M. C. Ruiz, *On preservation of some shifted and proportional orders by systems*, Statist. Probab. Lett. 60 (2) (2002), pp. 141–154.
[13] M. Benduch-Frąszczak, *Some properties of the proportional odds model*, Appl. Math. (Warsaw) 37 (2) (2010), pp. 247–256.
[14] P. Błażej, *Preservation of classes of life distributions under weighting with a general weight function*, Statist. Probab. Lett. 78 (17) (2008), pp. 3056–3061.
[15] D. G. Clayton, *Some odds ratio statistics for the analysis of ordered categorical data*, Biometrika 61 (1974), pp. 525–531.
[16] M. E. Ghitany and S. Kotz, *Reliability properties of extended linear failure-rate distributions*, Probab. Engrg. Inform. Sci. 21 (3) (2007), pp. 441–450.
[17] W. Gui, *Marshall–Olkin extended log-logistic distribution and its application in minification processes*, Appl. Math. Sci. (Ruse) 7 (77-80) (2013), pp. 3947–3961.
[18] R. C. Gupta and S. N. U. A. Kirmani, *The role of weighted distributions in stochastic modeling*, Comm. Statist. Theory Methods 19 (9) (1990), pp. 3147–3162.
[19] R. C. Gupta, S. Lvin, and C. Peng, *Estimating turning points of the failure rate of the extended Weibull distribution*, Comput. Statist. Data Anal. 54 (4) (2010), pp. 924–934.
[20] R. C. Gupta and C. Peng, *Estimating reliability in proportional odds ratio models*, Comput. Statist. Data Anal. 53 (4) (2009), pp. 1495–1510.
[21] T. E. Harris, *Branching processes*, Ann. Math. Statist. 19 (1948), pp. 474–494.
[22] J. Jarrahiferiz, G. R. Mohtashami Borzadaran, and A. H. Rezaei Roknabadi, *Some properties and applications of shifted proportional stochastic orders*, İstatistik 6 (2) (2013), pp. 80–91.
[23] S. N. U. A. Kirmani and R. C. Gupta, *On the proportional odds model in survival analysis*, Ann. Inst. Statist. Math. 53 (2) (2001), pp. 203–216.
[24] S. S. Maiti and M. Dey, *Tilted normal distribution and its survival properties*, J. Data Sci. 10 (2) (2012), pp. 225–240.
[25] A. W. Marshall and I. Olkin, *A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families*, Biometrika 84 (3) (1997), pp. 641–652.
[26] A. W. Marshall and I. Olkin, *Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families*, Springer, New York 2007.
[27] N. Misra, N. Gupta, and I. D. Dhariyal, *Preservation of some aging properties and stochastic orders by weighted distributions*, Comm. Statist. Theory Methods 37 (3–5) (2008), pp. 627–644.
[28] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester 2002.

[29] A. K. Nanda and S. Das, *Stochastic orders of the Marshall–Olkin extended distribution*, Statist. Probab. Lett. 82 (2) (2012), pp. 295–302.
[30] H. M. Ramos Romero and M. A. Sordo Díaz, *The proportional likelihood ratio order and applications*, Qüestiió 25 (2) (2001), pp. 211–223.
[31] M. Shaked and J. G. Shanthikumar, *Stochastic Orders*, Springer, New York 2007.
[32] W. J. Zimmer, Y. Wang, and P. K. Pathak, *Log-odds rate and monotone log-odds rate distributions*, J. Qual. Technol. 30 (1998), pp. 376–385.

Somayeh Abbasi
Department of Statistics
University of Isfahan
Isfahan, Iran
E-mail: s.abbasi@khuisf.ac.ir
s_abbasi_@yahoo.com

Mohammad Hossein Alamatsaz
Department of Statistics
University of Isfahan
Isfahan, Iran
E-mail: alamatho@sci.ui.ac.ir
mh_alamatsaz@yahoo.com

Received on 7.10.2016;
revised version on 15.6.2017

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 38, Fasc. 2 (2018), pp. 459–479
doi:10.19195/0208-4147.38.2.11

BELLMAN FUNCTIONS AND L^p ESTIMATES FOR PARAPRODUCTS*

BY

VJEKOSLAV KOVAČ (ZAGREB) AND KRISTINA ANA ŠKREB (ZAGREB)

Abstract. We give an explicit formula for one possible Bellman function associated with the L^p boundedness of dyadic paraproducts regarded as bilinear operators or trilinear forms. Then we apply the same Bellman function in various other settings, to give self-contained alternative proofs of the estimates for several classical operators. These include the martingale paraproducts of Bañuelos and Bennett and the paraproducts with respect to the heat flows.

2010 AMS Mathematics Subject Classification: Primary: 60G46; Secondary: 42B15.

Key words and phrases: Bellman function, martingale, paraproduct.

1. INTRODUCTION

According to Janson and Peetre [14] the name "paraproduct" denotes an idea rather than a unique object. Various types of paraproducts appear in the literature on analysis or probability and in each case certain boundedness properties (i.e. continuity) are crucial for their applications. An interested reader can find the historical overview and further references in the short expository paper [4]. In this paper we will focus mostly on martingale paraproducts and revisit the L^p estimates, which they are well known to satisfy.

We start with the dyadic paraproduct as a motivation for the forthcoming Bellman function that we construct. For two functions f and g from an appropriate space of real-valued test functions on $\mathbb{R}$ we can define the *dyadic paraproduct* as a bilinear operator in the following way:

$$\Pi_\epsilon(f,g) := \sum_{I\in\mathcal{D}} \epsilon_I |I|^{-2} \langle f, \mathbb{1}_I\rangle \langle g, \mathbb{h}_I\rangle \mathbb{h}_I. \tag{1.1}$$

Here $\mathcal{D}$ denotes the family of dyadic intervals in $\mathbb{R}$, $\mathbb{1}_I$ is the indicator function of an interval I, $\mathbb{h}_I := \mathbb{1}_{I_{\text{left}}} - \mathbb{1}_{I_{\text{right}}}$ is the L^∞-normalized Haar function, while

* This work has been supported by the Croatian Science Foundation under the project 3526.

I_{left} and I_{right} are respectively the left half and the right half of I. Moreover, $\langle \cdot, \cdot \rangle$ denotes the standard inner product with respect to the Lebesgue measure and $\epsilon = (\epsilon_I)_{I \in \mathcal{D}}$ is a collection of real numbers such that $|\epsilon_I| \leqslant 1$ for each $I \in \mathcal{D}$. (If we choose $\epsilon_I \in \{-1, 1\}$, then they simply represent $-$ and $+$ signs.) A convenient choice for the test functions are the so-called *dyadic step functions*, i.e. finite linear combinations of the indicator functions of dyadic intervals.

Typically, such an object is viewed as a linear operator in g with f fixed, when it becomes a particular instance of *Burkholder's martingale transform* [5]. Alternatively, one can fix g and consider it as a linear operator in f, in which case it is known as the linear paraproduct. In this text we prefer to look at Π_ϵ symmetrically and discuss its properties as a bilinear operator. This is partly motivated by the multilinear harmonic analysis, where more singular operators of this type are studied; see the book [24].

Equivalently, we can define the dyadic paraproduct as a trilinear form. We take the third test function h, and dualize (1.1) to get

$$\begin{aligned} \Lambda_\epsilon(f,g,h) &:= \int_{\mathbb{R}} \Pi_\epsilon(f,g)h = \sum_{I \in \mathcal{D}} \epsilon_I |I|^{-2} \langle f, \mathbb{1}_I \rangle \langle g, \mathbb{h}_I \rangle \langle h, \mathbb{h}_I \rangle \\ &= \sum_{I \in \mathcal{D}} \epsilon_I |I| [f]_I \frac{[g]_{I_{\text{left}}} - [g]_{I_{\text{right}}}}{2} \, \frac{[h]_{I_{\text{left}}} - [h]_{I_{\text{right}}}}{2}. \end{aligned} \tag{1.2}$$

Here $[f]_I$ denotes the average of a function f on a dyadic interval I.

It is well known that (1.2) satisfies certain L^p estimates, i.e. there exists a finite constant $\mathcal{C}_{p,q,r} \geqslant 0$ depending only on three exponents p, q, r such that

$$|\Lambda_\epsilon(f,g,h)| \leqslant \mathcal{C}_{p,q,r} \|f\|_{L^p(\mathbb{R})} \|g\|_{L^q(\mathbb{R})} \|h\|_{L^r(\mathbb{R})} \tag{1.3}$$

holds whenever $1 < p, q, r \leqslant \infty$ and $1/p + 1/q + 1/r = 1$. By $\| \cdot \|_{L^p(\mathbb{R})}$ we have denoted the L^p norm on $\mathbb{R}$ with respect to the Lebesgue measure.

The easiest proof of (1.3) when $q, r < \infty$ uses boundedness of the dyadic maximal function and the dyadic square function. We simply apply the Cauchy–Schwarz and Hölder inequalities to get

$$|\Lambda_\epsilon(f,g,h)| \leqslant \int_{\mathbb{R}} (Mf)(Sg)(Sh) \leqslant \|Mf\|_{L^p(\mathbb{R})} \|Sg\|_{L^q(\mathbb{R})} \|Sh\|_{L^r(\mathbb{R})},$$

where

$$Mf := \sup_{I \in \mathcal{D}} |I|^{-1} |\langle f, \mathbb{1}_I \rangle| \mathbb{1}_I \quad \text{and} \quad Sf := \Big(\sum_{I \in \mathcal{D}} |I|^{-2} |\langle f, \mathbb{h}_I \rangle|^2 \mathbb{1}_I \Big)^{1/2}$$

are the *dyadic maximal function* and the *dyadic square function*. Now the well-known L^p estimates for Mf and Sf give us the desired estimate (1.3).

On the side $p = \infty$ of the triangle in Figure 1, without loss of generality we can assume that $f \equiv 1$. The sharp constant in (1.3) was found by Burkholder in [6] and it equals $\mathcal{C}_{\infty,q,r} = \max\{q - 1, r - 1\}$.

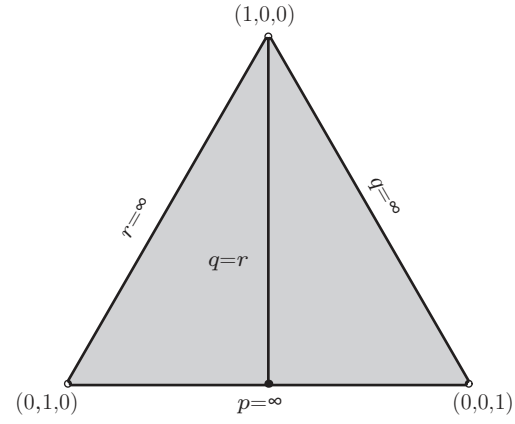


FIGURE 1. The Banach triangle with barycentric coordinates $\left(\frac{1}{p}, \frac{1}{q}, \frac{1}{r}\right)$.

On the other hand, on the sides $q = \infty$ and $r = \infty$, instead of the L^p estimates it is more natural to consider the BMO estimates, which will not be discussed in this paper. On the altitude $q = r$ of the triangle in Figure 1, the L^p estimates for the trilinear form (1.2) reduce to the L^p estimates for the dyadic square function, since

$$\int_{\mathbb{R}} f(Sg)^2 = \Lambda_\epsilon(f, g, g) \quad \text{if } \epsilon_I = 1 \text{ for each } I \in \mathcal{D}.$$

This implies $\|S\|_{L^q(\mathbb{R})\to L^q(\mathbb{R})} \leqslant \sqrt{\mathcal{C}_{p,q,q}}$. Actually, if the constant $\mathcal{C}_{p,q,q}$ is sharp, the last inequality turns into an equality. That sharp constant was found by Davis in [11] and it equals $\mathcal{C}_{p,q,q} = (z_q^*)^{-2}$, where z_q^* is the smallest positive zero of the confluent hypergeometric function (see [1]).

The special cases listed above are well studied and even the appropriate Bellman functions are found. For $p = \infty$ one can find them in the papers by Burkholder [6], Nazarov and Treil [17], Vasyunin and Volberg [25], Bañuelos and Osękowski [3], while for $q = r$ the reader can consult the book by Osękowski [20]. Therefore, because of the symmetry, throughout this paper we restrict our attention to the triples of exponents (p, q, r) satisfying

$$1 < p, q, r < \infty, \quad q > r, \quad \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1, \tag{1.4}$$

which correspond to the right half of the Banach triangle depicted in Figure 1.

Our goal is to give a direct proof of (1.3) using the Bellman function method. Such proofs typically give a better quantitative control and the same Bellman function can often be applied in various other settings.

First, we may assume that f, g, h are non-negative, as otherwise we split them into positive and negative parts. Furthermore, we observe that it turns out to be more practical to apply Young's inequality on the right-hand side of (1.3), but the newly obtained inequality is actually equivalent to the old one, because of

the homogeneity of the left-hand side. Therefore, it is enough to prove a non-homogeneous estimate

$$\sum_{I\in\mathcal{D}} |I|[f]_I \left|\frac{[g]_{I_{\text{left}}}-[g]_{I_{\text{right}}}}{2}\right| \left|\frac{[h]_{I_{\text{left}}}-[h]_{I_{\text{right}}}}{2}\right| \leqslant \mathcal{C}_{p,q,r}\left(\frac{1}{p}\|f\|^p_{L^p(\mathbb{R})}+\frac{1}{q}\|g\|^q_{L^q(\mathbb{R})}+\frac{1}{r}\|h\|^r_{L^r(\mathbb{R})}\right).$$

If we want to recover (1.3), we just have to homogenize the above inequality and use the assumed bound on ϵ_I.

For an arbitrary dyadic interval I we define a scale-invariant expression

$$\Phi_I(f,g,h) := \frac{1}{|I|}\sum_{\substack{J\in\mathcal{D}\\ J\subseteq I}} |J|[f]_J \frac{|[g]_{J_{\text{left}}}-[g]_{J_{\text{right}}}|}{2}\,\frac{|[h]_{J_{\text{left}}}-[h]_{J_{\text{right}}}|}{2},$$

so that we can normalize the desired estimate and rewrite it as

$$\Phi_I(f,g,h) \leqslant \mathcal{C}_{p,q,r}\left(\frac{1}{p}[f^p]_I+\frac{1}{q}[g^q]_I+\frac{1}{r}[h^r]_I\right). \tag{1.5}$$

This is easily seen multiplying (1.5) by $|I|$ and letting I exhaust the positive and the negative half-axis. Splitting $\sum_{J\subseteq I}$ into $\sum_{J\subseteq I_{\text{left}}}$, $\sum_{J\subseteq I_{\text{right}}}$, and $J=I$ gives us the following scaling identity:

$$\begin{aligned}\Phi_I(f,g,h) = {}& \frac{1}{2}\Phi_{I_{\text{left}}}(f,g,h)+\frac{1}{2}\Phi_{I_{\text{right}}}(f,g,h)\\ &+[f]_I\frac{|[g]_{I_{\text{left}}}-[g]_{I_{\text{right}}}|}{2}\,\frac{|[h]_{I_{\text{left}}}-[h]_{I_{\text{right}}}|}{2}.\end{aligned} \tag{1.6}$$

We can define the *abstract Bellman function*

$$\mathbb{B}(u,v,w,U,V,W) := \sup_{f,g,h}\Phi_I(f,g,h),$$

where the supremum is taken over all non-negative functions f,g,h such that $[f]_I=u$, $[g]_I=v$, $[h]_I=w$, $[f^p]_I=U$, $[g^q]_I=V$, $[h^r]_I=W$. Note that the above supremum does not depend on the choice of the "base" interval I.

Now we list some properties of that function.

($\mathcal{B}1$) *Domain*: The function $\mathbb{B}$ is defined on the set

$$\mathbb{D} := \{(u,v,w,U,V,W)\in[0,\infty)^6 : u^p\leqslant U, v^q\leqslant V, w^r\leqslant W\}.$$

The upper bounds simply follow from Jensen's inequality.

($\mathcal{B}2$) *Range*:

$$0 \leqslant \mathbb{B}(u,v,w,U,V,W) \leqslant \mathcal{C}_{p,q,r}\left(\frac{1}{p}U + \frac{1}{q}V + \frac{1}{r}W\right),$$

where on the right-hand side we assume that the estimate (1.5) holds.

($\mathcal{B}3$) *The main inequality*:

$$\mathbb{B}(\boldsymbol{x}) \geqslant \frac{1}{2}\mathbb{B}(\boldsymbol{x}_1) + \frac{1}{2}\mathbb{B}(\boldsymbol{x}_2) + u\frac{|v_1 - v_2|}{2}\frac{|w_1 - w_2|}{2},$$

whenever the six-tuples $\boldsymbol{x} = (u,v,w,U,V,W)$ and $\boldsymbol{x}_i = (u_i, v_i, w_i, U_i, V_i, W_i)$, $i = 1,2$, belong to the domain and satisfy $\boldsymbol{x} = \frac{1}{2}\boldsymbol{x}_1 + \frac{1}{2}\boldsymbol{x}_2$. This can be easily seen by taking the supremum in the scaling identity (1.6) over all non-negative functions f, g, h such that $[f]_{I_{\text{left}}} = u_1$, $[f^p]_{I_{\text{left}}} = U_1$, etc.

Conversely, suppose that we have already found a function $\mathcal{B}$ with properties ($\mathcal{B}1$)–($\mathcal{B}3$). We will show how its existence implies the estimate (1.3). Applying ($\mathcal{B}3$) n times with a fixed choice of the functions $f, g, h \geqslant 0$ and a fixed base interval I gives us

$$\begin{aligned}|I|\,\mathcal{B}\big([f]_I, [g]_I, [h]_I, [f^p]_I, [g^q]_I, [h^r]_I\big) &\geqslant \sum_{\substack{J\subseteq I\\ |J|=2^{-n}|I|}} |J|\mathcal{B}\big([f]_J, [g]_J, [h]_J, [f^p]_J, [g^q]_J, [h^r]_J\big)\\ &\quad + \sum_{\substack{J\subseteq I\\ |J|>2^{-n}|I|}} |J|[f]_J\,\frac{\big|[g]_{J_{\text{left}}} - [g]_{J_{\text{right}}}\big|}{2}\,\frac{\big|[h]_{J_{\text{left}}} - [h]_{J_{\text{right}}}\big|}{2}.\end{aligned}$$

Since, by ($\mathcal{B}2$), the first sum is non-negative and

$$\mathcal{B}\big([f]_I, [g]_I, [h]_I, [f^p]_I, [g^q]_I, [h^r]_I\big) \leqslant \mathcal{C}_{p,q,r}\left(\frac{1}{p}[f^p]_I + \frac{1}{q}[g^q]_I + \frac{1}{r}[h^r]_I\right),$$

letting $n \to \infty$ leads us to the estimate (1.5) and then in turn also to (1.3).

It will be convenient to find a function $\mathcal{B}$ that also satisfies the following condition:

($\mathcal{B}4$) $$\mathcal{B}(\boldsymbol{x}) + (d\mathcal{B})(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}) \geqslant \mathcal{B}(\boldsymbol{x}_1) + \frac{2}{3}u|v_1 - v||w_1 - w|,$$

whenever the six-tuples $\boldsymbol{x} = (u,v,w,U,V,W)$ and $\boldsymbol{x}_1 = (u_1, v_1, w_1, U_1, V_1, W_1)$ belong to the domain ($\mathcal{B}1$). Here $d\mathcal{B}$ denotes the differential of $\mathcal{B}$, which is a linear form, and we consider it at the point $\boldsymbol{x}$ and apply it to the vector $\boldsymbol{x}_1 - \boldsymbol{x}$. Condition ($\mathcal{B}4$) is required by an application considered in Subsection 3.1.

Now we want to find an explicit formula for one possible function $\mathcal{B}$. We define the function $\mathcal{B}\colon \mathbb{D} \to \mathbb{R}$ as

$$\mathcal{B}(u,v,w,U,V,W) := \mathcal{C}_{p,q,r}\left(\frac{1}{p}U + \frac{1}{q}V + \frac{1}{r}W\right) - \mathcal{A}(u,v,w), \tag{1.7}$$

where $\mathcal{A}\colon [0,\infty)^3 \to \mathbb{R}$ is given by

$$\mathcal{A}(u,v,w) :=$$
$$\begin{cases} Au^p + Bv^q + Cw^r, & u^p \leqslant w^r \leqslant v^q,\\ \frac{A(p-1)-C}{p-1}u^p + Bv^q + \frac{Cp}{p-1}uw^{r-r/p}, & w^r \leqslant u^p \leqslant v^q,\\ \frac{A(p-1)-(B+C)}{p-1}u^p + \frac{Bp}{p-1}uv^{q-q/p} + \frac{Cp}{p-1}uw^{r-r/p}, & w^r \leqslant v^q \leqslant u^p,\\ \frac{A(p-1)-(B+C)}{p-1}u^p + \frac{Bq}{2}uv^2w^{1-r/q} + \frac{2Cpr-Bp(q-r)}{2r(p-1)}uw^{r-r/p}, & v^q \leqslant w^r \leqslant u^p,\\ \frac{2Ar(p-1)-B(q+r)}{2r(p-1)}u^p + \frac{Bq^2}{2p(q-2)}u^{p-2p/q}v^2 + \frac{Bq(q-r)}{2r(q-2)}v^2w^{r-2r/q} & \\ \qquad + \frac{2Cr-B(q-r)}{2r}w^r, & v^q \leqslant u^p \leqslant w^r,\\ Au^p + \frac{Bq}{p(q-2)}v^q + \frac{Bq(q-r)}{2r(q-2)}v^2w^{r-2r/q} + \frac{2Cr-B(q-r)}{2r}w^r, & u^p \leqslant v^q \leqslant w^r. \end{cases}$$

The coefficients $A, B, C > 0$ will be appropriately chosen depending only on the exponents p, q, r and then one will be able to take $\mathcal{C}_{p,q,r} = \max\{Ap, Bq, Cr\}$. We see that the function $\mathcal{A}$ has a similar form to the one constructed by Nazarov and Treil [17], which can in our notation be written as

$$\mathcal{NT}(v,w) = A(v^q + w^r) + B\begin{cases} \frac{2}{q}v^q + \left(\frac{2}{r} - 1\right)w^r, & v^q \geqslant w^r,\\ v^2w^{2-r}, & v^q \leqslant w^r. \end{cases}$$

It corresponds to the endpoint case $p = \infty$, $1 < r < 2 < q < \infty$. Instead of one critical curve $v^q = w^r$ for $\mathcal{NT}$, we have three critical surfaces:

$$u^p = v^q, \quad u^p = w^r, \quad v^q = w^r. \tag{1.8}$$

Finally, we are ready to state our main result.

THEOREM 1.1. *For the exponents p, q, r satisfying* (1.4) *it is possible to choose the coefficients A, B, C such that the function $\mathcal{B}$ defined by* (1.7) *is of class* C^1 *on the whole domain $\mathbb{D}$ and satisfies the conditions* ($\mathcal{B}2$) (*with $\mathcal{C}_{p,q,r} = \max\{Ap, Bq, Cr\}$*), ($\mathcal{B}3$), *and* ($\mathcal{B}4$). *One possible choice of the coefficients is*

$$A = \frac{88q^4r}{(p-1)(r-1)(q-r)}, \quad B = 1, \quad \textit{and} \quad C = \frac{11q^3r}{(r-1)(q-r)},$$

which yields

$$\mathcal{C}_{p,q,r} = \frac{88pq^4r}{(p-1)(r-1)(q-r)}.$$

The claim that $\mathcal{B}$ is of class C^1 on $\mathbb{D}$ should be understood in the sense that the function $\mathcal{A}$ is continuous on $[0,\infty)^3$, $\mathcal{A}$ is continuously differentiable on $(0,\infty)^3$, and the partial derivatives of $\mathcal{A}$ can be continuously extended to $[0,\infty)^3$. At a boundary point the differential $d\mathcal{B}$ in ($\mathcal{B}4$) is interpreted as the linear form whose coefficients are the aforementioned continuous extensions of partial derivatives to that point.

The motivation behind finding the explicit Bellman function (instead of just using the abstract one) is that in some contexts the explicit formula could be useful. For example, Carbonaro and Dragičević in [7] and [8] made use of the fact that the explicit Bellman function $\mathcal{NT}$ involves powers. Another source of motivation is that we would also like to find a direct proof (without stopping time arguments) of the estimates for the "twisted" paraproduct considered by one of the authors in [15] or the "twisted" quadrilinear form considered by Durcik in [12] and [13]. This could also extend the range of exponents for a non-adapted stochastic integral considered by the authors in [16] or for the norm-variation of ergodic averages with respect to two commuting transformations [23]. So far we can only say that the Bellman function that has to be constructed for any of the mentioned problems should necessarily encode some structure of the function from Theorem 1.1, as dyadic paraproducts are the simplest and prototypical multilinear multipliers.

The Bellman function that we construct certainly does not give the best possible constants $\mathcal{C}_{p,q,r}$ in (1.3). Indeed, the sharp constant for any triple of exponents from the generic range (1.4) has not yet been determined to the best of our knowledge. Search for the abstract Bellman function $\mathbb{B}$ would lead us to the equations

$$\det\begin{bmatrix} \partial_u^2\mathbb{B} & \partial_u\partial_v\mathbb{B} & \partial_u\partial_w\mathbb{B} & \partial_u\partial_U\mathbb{B} & \partial_u\partial_V\mathbb{B} & \partial_u\partial_W\mathbb{B} \\ \partial_u\partial_v\mathbb{B} & \partial_v^2\mathbb{B} & \partial_v\partial_w\mathbb{B}\pm u & \partial_v\partial_U\mathbb{B} & \partial_v\partial_V\mathbb{B} & \partial_v\partial_W\mathbb{B} \\ \partial_u\partial_w\mathbb{B} & \partial_v\partial_w\mathbb{B}\pm u & \partial_w^2\mathbb{B} & \partial_w\partial_U\mathbb{B} & \partial_w\partial_V\mathbb{B} & \partial_w\partial_W\mathbb{B} \\ \partial_u\partial_U\mathbb{B} & \partial_v\partial_U\mathbb{B} & \partial_w\partial_U\mathbb{B} & \partial_U^2\mathbb{B} & \partial_U\partial_V\mathbb{B} & \partial_U\partial_W\mathbb{B} \\ \partial_u\partial_V\mathbb{B} & \partial_v\partial_V\mathbb{B} & \partial_w\partial_V\mathbb{B} & \partial_U\partial_V\mathbb{B} & \partial_V^2\mathbb{B} & \partial_V\partial_W\mathbb{B} \\ \partial_u\partial_W\mathbb{B} & \partial_v\partial_W\mathbb{B} & \partial_w\partial_W\mathbb{B} & \partial_U\partial_W\mathbb{B} & \partial_V\partial_W\mathbb{B} & \partial_W^2\mathbb{B} \end{bmatrix} = 0. \tag{1.9}$$

One way of simplifying (1.9) is to consider the non-homogeneous function $\mathcal{B}$ of the form (1.7). The function $\mathcal{B}$ is now a supersolution of the equation for the true Bellman function $\mathbb{B}$, but a function of that form can still yield the optimal (unknown) constant. This way (1.9) reduces to

$$\det \mathbb{A}_\pm = 0, \tag{1.10}$$

where $\mathbb{A}_\pm$ are the matrices defined in (2.3) below. Alternatively, one can use the homogeneities of $\mathbb{B}$ to reduce the dimension in (1.9). Equations like (1.10) can sometimes be turned into the Monge–Ampère equation by an appropriate change of variables, which does not seem to be the case here. At the moment, we do not know how to solve (1.10), so we impose slightly weaker conditions on our function $\mathcal{B}$ that result in a constant $\mathcal{C}_{p,q,r}$ which is not optimal. It would be interesting to find

a Bellman function $\mathcal{B}$ that yields the optimal constant, or perhaps even the exact abstract Bellman function $\mathbb{B}$. Let us remark once again that this was achieved by Bañuelos and Osękowski [3] in the endpoint case $p = \infty$, $f \equiv 1$.

We have organized the remainder of the paper as follows. In the next section we present the proof of Theorem 1.1. In Section 3 we apply Theorem 1.1 to reprove the well-known L^p estimates for martingale paraproducts and the heat flow paraproducts.

2. PROOF OF THEOREM 1.1

The continuity of $\mathcal{A}$ on $[0,\infty)^3$ is obvious. Indeed, observe that all exponents appearing in the definition of $\mathcal{A}$ are positive. Thus, $\mathcal{A}$ is clearly well-defined and continuous on each of the six closed regions determined by the inequalities for u, v, w and it is straightforward to verify that the six formulas are compatible on the common boundaries.

To see that $\mathcal{A}$ is continuously differentiable on the open octant $(0,\infty)^3$, we just calculate the first order partial derivatives in the interior of each of the previously mentioned regions. The formula for each of these derivatives inside any of the regions continuously extends to the whole open octant. Moreover, these formulas coincide on the boundaries of each of the two adjacent regions, so we can deduce that $\mathcal{A}$ really is of class C^1 on $(0,\infty)^3$. For instance, both formulas for $\frac{\partial \mathcal{A}}{\partial u}(u,v,w)$ at the common boundary of the two adjacent open regions $v^q < u^p < w^r$ and $u^p < v^q < w^r$, which is a subset of $u^p = v^q$, simplify to Apu^{p-1}. All other cases are treated in the same manner.

Also, it is easy to see that the partial derivatives have limits at each point of the boundary of $[0,\infty)^3$ and hence they can be continuously extended to $[0,\infty)^3$. For example, if $0 < v^q \leqslant w^r \leqslant u^p$, then the partial derivative of $\mathcal{A}$ with respect to w equals

$$\frac{\partial \mathcal{A}}{\partial w}(u,v,w) = \frac{B(q-r)}{2} uv^2 w^{-r/q} + \frac{2Cr - B(q-r)}{2} uw^{r/q}.$$

Obviously, the only problematic points are the ones on the part of the boundary lying on the plane $w = 0$, but since $v^q/w^r \leqslant 1$, the limit as $w \to 0$ still exists and equals zero. The existence of the other limits can be shown in a similar way.

The estimate ($\mathcal{B}2$) follows directly from the definitions of the functions $\mathcal{A}$ and $\mathcal{B}$, since

$$(\mathcal{A}2) \qquad 0 \leqslant \mathcal{A}(u,v,w) \leqslant Au^p + Bv^q + Cw^r$$

as long as $A, B, C \geqslant 0$. This is easily seen by using Young's inequality. The nonnegativity of $\mathcal{B}$ on $\mathbb{D}$ is guaranteed if $\mathcal{C}_{p,q,r} \geqslant Ap, Bq, Cr$.

Observe that ($\mathcal{B}3$) is equivalent to

$$(\mathcal{A}3) \quad \frac{1}{2}\mathcal{A}(u_1,v_1,w_1) + \frac{1}{2}\mathcal{A}(u_2,v_2,w_2) - \mathcal{A}(u,v,w) \geqslant u \frac{|v_1 - v_2|}{2} \frac{|w_1 - w_2|}{2},$$

where (u,v,w), (u_1,v_1,w_1), and (u_2,v_2,w_2) are in $[0,\infty)^3$ and such that

$$(u,v,w)=\frac{1}{2}(u_1,v_1,w_1)+\frac{1}{2}(u_2,v_2,w_2), \tag{2.1}$$

while $(\mathcal{B}4)$ is equivalent to

$$\begin{aligned}(\mathcal{A}4)\qquad \mathcal{A}(u_1,v_1,w_1)\geqslant{}&\mathcal{A}(u,v,w)+(d\mathcal{A})(u,v,w)(u_1-u,v_1-v,w_1-w)\\&+\frac{2}{3}u|v_1-v||w_1-w|,\end{aligned}$$

where (u,v,w) and (u_1,v_1,w_1) are in $[0,\infty)^3$. Instead of proving $(\mathcal{A}3)$ and $(\mathcal{A}4)$ directly, we will reduce them conveniently to an inequality for quadratic forms.

Let $(u,v,w)\in(0,\infty)^3$ be a point that does not lie on any of the three critical surfaces (1.8). This means that $\mathcal{A}$ is of class C^2 on an open ball around that point. If we take (u_1,v_1,w_1), (u_2,v_2,w_2) from that open ball such that (2.1) holds, then substituting $u=(u_1+u_2)/2$, $\Delta u=(u_1-u_2)/2$, etc., and adding Taylor's formulas at (u,v,w) for $\mathcal{A}(u\pm\Delta u,v\pm\Delta v,w\pm\Delta w)$ gives us the infinitesimal version of $(\mathcal{A}3)$:

$$(d^2\mathcal{A})(u,v,w)(\Delta u,\Delta v,\Delta w)\geqslant 2u|\Delta v||\Delta w|. \tag{$\mathcal{A}3'$}$$

Here $d^2\mathcal{A}$ denotes the second differential of $\mathcal{A}$ as a quadratic form, which we consider at the point (u,v,w) and apply to the vector $(\Delta u,\Delta v,\Delta w)$. Notice that $(\mathcal{A}3')$ does not hold on the whole domain of the function $\mathcal{A}$, which is $[0,\infty)^3$, but it does hold on the interior of each of the six regions into which the three surfaces divide $(0,\infty)^3$.

Conversely, $(\mathcal{A}3')$ implies $(\mathcal{A}3)$, i.e. the two inequalities are equivalent for continuously differentiable functions, which is enabled by the convexity of the domain. To show the converse, first take a point $(u,v,w)\in(0,\infty)^3$ and a vector $(\Delta u,\Delta v,\Delta w)\in\mathbb{R}^3$ such that also $(u\pm\Delta u,v\pm\Delta v,w\pm\Delta w)\in(0,\infty)^3$. Now define the function $\alpha\colon[-1,1]\to\mathbb{R}$ as

$$\alpha(t):=\mathcal{A}(u+t\Delta u,v+t\Delta v,w+t\Delta w). \tag{2.2}$$

This function is continuously differentiable on $[-1,1]$ since $\mathcal{A}$ is of class C^1 on $(0,\infty)^3$. Also, α is piecewise C^2 on $[-1,1]$. This follows from the facts that $\mathcal{A}$ is of class C^2 on $(0,\infty)^3$ outside the surfaces (1.8), it has bounded second derivatives away from the coordinate planes $u=0$, $v=0$, and $w=0$, and the segment $\{(u+t\Delta u,v+t\Delta v,w+t\Delta w):t\in[-1,1]\}$ intersects the three critical surfaces at finitely many points. Using the integration by parts and the fundamental theorem of calculus (both in the versions for absolutely continuous functions; see [9]) gives us the equality

$$\frac{1}{2}\alpha(1)+\frac{1}{2}\alpha(-1)-\alpha(0)=\frac{1}{2}\int_{-1}^{1}(1-|t|)\alpha''(t)dt.$$

From the above identity we deduce

$$\frac{1}{2}\mathcal{A}(u+\Delta u, v+\Delta v, w+\Delta w)+\frac{1}{2}\mathcal{A}(u-\Delta u, v-\Delta v, w-\Delta w)-\mathcal{A}(u,v,w)$$
$$=\frac{1}{2}\int_{-1}^{1}(1-|t|)(d^2\mathcal{A})(u+t\Delta u, v+t\Delta v, w+t\Delta w)(\Delta u,\Delta v,\Delta w)dt.$$

Finally, by $(\mathcal{A}3')$ applied at all but finitely many points, the last expression is at least

$$\frac{1}{2}\int_{-1}^{1}(1-|t|)2(u+t\Delta u)|\Delta v||\Delta w|dt = u|\Delta v||\Delta w|,$$

which gives exactly $(\mathcal{A}3)$.

Moreover, $(\mathcal{A}3')$ implies $(\mathcal{A}4)$. To verify this, we also take $(u,v,w)\in(0,\infty)^3$ and $(\Delta u,\Delta v,\Delta w)\in\mathbb{R}^3$ such that $(u+\Delta u, v+\Delta v, w+\Delta w)\in(0,\infty)^3$. We define $\alpha\colon[0,1]\to\mathbb{R}$ again by the formula (2.2). Integration by parts, the fundamental theorem of calculus, and $(\mathcal{A}3')$ this time give

$$\alpha(1)=\alpha(0)+\alpha'(0)+\int_0^1(1-t)\alpha''(t)dt,$$

and therefore,

$$\mathcal{A}(u+\Delta u, v+\Delta v, w+\Delta w)\geqslant\mathcal{A}(u,v,w)+(d\mathcal{A})(u,v,w)(\Delta u,\Delta v,\Delta w)$$
$$+\int_0^1(1-t)2(u+t\Delta u)|\Delta v||\Delta w|dt.$$

Since $u+t\Delta u=(1-t)u+t(u+\Delta u)\geqslant(1-t)u$, the integral in t on the right-hand side is at least $(2/3)u|\Delta v||\Delta w|$, which establishes $(\mathcal{A}4)$.

This way we proved that $(\mathcal{A}3')$ implies $(\mathcal{A}3)$ and $(\mathcal{A}4)$, but only on $(0,\infty)^3$. To see that these two also hold on $[0,\infty)^3$, we just have to extend the obtained inequalities by the continuity of $\mathcal{A}$ and $d\mathcal{A}$. We have commented in the introduction how we interpret $d\mathcal{A}$ at the boundary of the domain.

Now we are left with proving $(\mathcal{A}3')$, which is equivalent to showing that the two matrices

$$\mathbb{A}_\pm=\begin{bmatrix}\partial_u^2\mathcal{A} & \partial_u\partial_v\mathcal{A} & \partial_u\partial_w\mathcal{A}\\ \partial_u\partial_v\mathcal{A} & \partial_v^2\mathcal{A} & \partial_v\partial_w\mathcal{A}\pm u\\ \partial_u\partial_w\mathcal{A} & \partial_v\partial_w\mathcal{A}\pm u & \partial_w^2\mathcal{A}\end{bmatrix} \tag{2.3}$$

are positive semi-definite on each of the six open regions into which the surfaces (1.8) split $(0,\infty)^3$. To do so, we will use Sylvester's criterion and verify that all three principal minors are positive. More precisely, we will prove that the constants A,B,C can be chosen so that this is fulfilled.

We can simplify the calculations a bit by substituting $t = v^q/u^p$, $s = w^r/u^p$ and noting that

$$\mathcal{A}(u,v,w) = u^p\gamma(t,s), \tag{2.4}$$

where $\gamma\colon (0,\infty)^2 \to \mathbb{R}$ is given by

$$\gamma(t,s)=\begin{cases} A+Bt+Cs, & 1\leqslant s\leqslant t,\\ \frac{A(p-1)-C}{p-1}+Bt+\frac{Cp}{p-1}s^{1-1/p}, & s\leqslant 1\leqslant t,\\ \frac{A(p-1)-(B+C)}{p-1}+\frac{Bp}{p-1}t^{1-1/p}+\frac{Cp}{p-1}s^{1-1/p}, & s\leqslant t\leqslant 1,\\ \frac{A(p-1)-(B+C)}{p-1}+\frac{Bq}{2}t^{2/q}s^{1/r-1/q}+\frac{2Cpr-Bp(q-r)}{2r(p-1)}s^{1-1/p}, & t\leqslant s\leqslant 1,\\ \frac{2Ar(p-1)-B(q+r)}{2r(p-1)}+\frac{Bq^2}{2p(q-2)}t^{2/q}+\frac{Bq(q-r)}{2r(q-2)}t^{2/q}s^{1-2/q} & \\ \qquad +\frac{2Cr-B(q-r)}{2r}s, & t\leqslant 1\leqslant s,\\ A+\frac{Bq}{p(q-2)}t+\frac{Bq(q-r)}{2r(q-2)}t^{2/q}s^{1-2/q}+\frac{2Cr-B(q-r)}{2r}s, & 1\leqslant t\leqslant s. \end{cases}$$

After plugging (2.4) into (2.3) and multiplying from both sides with the diagonal matrix $\operatorname{diag}(u^{1-p/2},\, u^{p/q-p/2},\, u^{p/r-p/2})$, we obtain the matrices $M=[m_{ij}]$, where

$$\begin{aligned} m_{11} &= p(p-1)\gamma(t,s) - p(p-1)t\partial_t\gamma(t,s) - p(p-1)s\partial_s\gamma(t,s)\\ &\quad + 2p^2ts\partial_t\partial_s\gamma(t,s) + p^2t^2\partial_t^2\gamma(t,s) + p^2s^2\partial_s^2\gamma(t,s),\\ m_{12} &= m_{21} = -pqt^{1-1/q}s\partial_t\partial_s\gamma(t,s) - pqt^{2-1/q}\partial_t^2\gamma(t,s),\\ m_{13} &= m_{31} = -prts^{1-1/r}\partial_t\partial_s\gamma(t,s) - prs^{2-1/r}\partial_s^2\gamma(t,s),\\ m_{22} &= q(q-1)t^{1-2/q}\partial_t\gamma(t,s) + q^2t^{2-2/q}\partial_t^2\gamma(t,s),\\ m_{23} &= m_{32} = qrt^{1-1/q}s^{1-1/r}\partial_t\partial_s\gamma(t,s) \pm 1,\\ m_{33} &= r(r-1)s^{1-2/r}\partial_s\gamma(t,s) + r^2s^{2-2/r}\partial_s^2\gamma(t,s), \end{aligned}$$

and the problem is reduced to verifying that these matrices are positive definite on the interior of each of the six regions determined by the inequalities for t and s. First, we will calculate the three principal minors of the above matrices for each region, and then we will explain why we can choose the constants A, B, C such that all of them are positive.

The following expressions were calculated using *Mathematica* [26].

Region 1: $1 < s < t$. Minor 1×1: $\boxed{\boxed{Ap(p-1)}}$

Minor 2×2: $\boxed{\boxed{ABp(p-1)q(q-1)t^{1-2/q}}}$

Determinants (with $\pm$):

$$\boxed{\boxed{ABCp(p-1)q(q-1)r(r-1)t^{1-2/q}s^{1-2/r}}} - Ap(p-1)$$

Region 2: $s < 1 < t$. Minor 1×1: $\boxed{\boxed{p\big(A(p-1)-C\big)}}$

Minor 2×2: $\boxed{\boxed{Bp\big(A(p-1)-C\big)q(q-1)t^{1-2/q}}}$

Determinants (with $\pm$):

$$\boxed{\boxed{BCp\big(A(p-1)-C\big)(q-1)r^2t^{1-2/q}s^{1/q-1/r}}} - BC^2q(q-1)r^2t^{1-2/q}s^{2/q} - p\big(A(p-1)-C\big)$$

Region 3: $s < t < 1$. Minor 1×1: $\boxed{\boxed{p\big(A(p-1)-B-C\big)}}$

Minor 2×2: $\boxed{\boxed{\frac{Bp\big(A(p-1)-B-C\big)q^2}{r}t^{1/r-1/q}}} - B^2q^2t^{2/r}$

Determinants (with $\pm$):

$$\boxed{\boxed{BCp\big(Ap(q+r)-qr(B+C)\big)t^{1/r-1/q}s^{1/q-1/r}}} - B^2Cqr^2t^{2/r}s^{1/q-1/r} - BC^2q^2rt^{1/r-1/q}s^{2/q} - p\big(A(p-1)-B-C\big) \pm 2BCqrt^{1/r}s^{1/q}$$

Region 4: $t < s < 1$. Minor 1×1: $\boxed{\boxed{p\big(A(p-1)-B-C\big)}}$

Minor 2×2: $\boxed{\boxed{Bp\big(A(p-1)-B-C\big)qs^{1/r-1/q}}} - B^2q^2t^{2/q}s^{2/r-2/q}$

Determinants (with $\pm$):

$$\boxed{\boxed{\frac{\big(A(p-1)-B-C\big)\big(Bpr\big(2Cr-B(q-r)\big)-2p\big)}{2}}}$$
$$\pm B^2q(q-r)t^{3/q}s^{1/r-2/q}$$
$$\mp 2Bp\big(A(p-1)-B-C\big)(q-r)t^{1/q}s^{-1/q} + \frac{B^3q(q-r)(3q-r)}{4}t^{4/q}s^{1/r-3/q}$$
$$- \frac{B^2p\big(A(p-1)-B-C\big)(q-r)(2q-r)}{2}t^{2/q}s^{-2/q}$$
$$\pm Bq\big(2Cr-B(q-r)\big)t^{1/q}s^{1/r}$$
$$- \frac{Bq\big(2Cr-B(q-r)\big)^2}{4}s^{1/q+1/r}$$
$$+ \frac{B^2q\big(2Cr-B(q-r)\big)(q-2r)}{2}t^{2/q}s^{1/r-1/q}$$

Region 5: $t < 1 < s$. Minor 1×1:

$$\boxed{\boxed{\frac{p\big(2Ar(p-1)-B(q+r)\big)}{2r}}} + \frac{Bp(q-r)}{2r}t^{2/q}$$

Minor 2×2:

$$\boxed{\frac{Bpq(q-r)\big(2Ar(p-1)-B(q+r)\big)}{2r^2(q-2)}s^{1-2/q}} - \frac{B^2q^2(pq+q-2p)}{2p(q-2)}t^{2/q} + \frac{B^2pq(q-r)^2}{2r^2(q-2)}t^{2/q}s^{1-2/q} + \boxed{\frac{Bq^2\big(2Ar(p-1)-B(q+r)\big)}{2r(q-2)}}$$

Determinants (with $\pm$):

$$\boxed{\frac{B(q-r)(p+q)\big(2Ar(p-1)-B(q+r)\big)\big(2Cr-B(q-r)\big)}{4r(q-2)}s^{2/p}}$$
$$+ \boxed{\frac{B\big(2Ar(p-1)-B(q+r)\big)\big(2Cr-B(q-r)\big)q^2(r-1)}{4r(q-2)}s^{1-2/r}}$$
$$- \frac{p\big(2Ar(p-1)-B(q+r)\big)}{2r}$$
$$+ \frac{B^2pq(q-r)^2(r-1)\big(2Cr-B(q-r)\big)}{4r^2(q-2)}t^{2/q}s^{2/p}$$
$$- \frac{B^2q^2(pq-2p+q)(r-1)\big(2Cr-B(q-r)\big)}{4p(q-2)}t^{2/q}s^{1-2/r} - \frac{Bp(q-r)}{2r}t^{2/q}$$
$$\mp \frac{B^2p(q-r)^2}{r}t^{3/q}s^{1/p-1/q}$$
$$+ \frac{B^2q(q-r)(q-p)\big(2Ar(p-1)-B(q+r)\big)}{4p(q-2)}t^{2/q}s^{2/p-1}$$
$$- \frac{B^3p(q-r)^3(qr-2r+q)}{4r^2(q-2)}t^{4/q}s^{2/p-2/q}$$
$$- \frac{B^3qr(pq-2p+q)(q-r)(q-p)}{4p^2(q-2)}t^{4/q}s^{2/p-1}$$
$$\mp \frac{Bp(q-r)\big(2Ar(p-1)-B(q+r)\big)}{r}t^{1/q}s^{1/p-1/q}$$
$$- \frac{B^2(q-r)^2\big(2Ar(p-1)-B(q+r)\big)(2pq-3p-q)}{4r(q-2)}t^{2/q}s^{2/p-2/q}$$

Region 6: $1 < t < s$. Minor 1×1: $\boxed{Ap(p-1)}$

Minor 2×2:

$$\boxed{\frac{ABpq(p-1)(q-r)}{r(q-2)}s^{1-2/q}} + \boxed{\frac{ABq^2(p-1)(q-1)}{q-2}t^{1-2/q}}$$

Determinants (with $\pm$):

$$\boxed{\frac{AB\big(2Cr-B(q-r)\big)(p-1)(q-r)(p+q)}{2(q-2)}s^{2/p}}$$
$$-\frac{AB^2qr(p-1)(q-1)(q-r)(p-q)}{2p(q-2)}ts^{2/p-1}$$
$$+\boxed{\frac{AB\big(2Cr-B(q-r)\big)qr(p-1)(q-1)(p+q)}{2p(q-2)}t^{1-2/q}s^{1-2/r}}-Ap(p-1)$$
$$-\frac{AB^2(p-1)(q-r)^2(2pq-3p-q)}{2(q-2)}t^{2/q}s^{2/p-2/q}$$
$$\mp 2ABp(p-1)(q-r)t^{1/q}s^{1/p-1/q}$$

In each of the expressions there is a unique dominant term (regarding the exponents of t and s) and it is double framed. We choose B arbitrarily (say $B=1$), then take C large enough (depending on p,q,r,B), and finally take A large enough (depending on p,q,r,B,C). While doing so, we take care that the coefficient of the double framed term is greater than the sum of the absolute values of coefficients of the terms that are neither framed nor circled. We can do so because by taking C large enough the expression multiplying A in the coefficient of the dominant term can be made larger than the sum of the absolute values of the corresponding expressions in other non-circled terms that contain A. Consequently, the coefficient of the dominant term grows faster than the sum of the absolute values of the coefficients in the other terms as A tends to infinity. This means that we can take A large enough so that the dominant term actually dominates the sum of all other non-framed and non-circled terms in each expression. Another way of phrasing the argument that sufficiently large A and C make six considered determinantal expressions positive is to observe that each dominant term contains the product AC, as opposed to any other non-circled term.

The only problematic terms that we cannot dominate with the dominant term are the circled ones, because of their uncontrollable growth in A. However, just by taking

$$C \geqslant B(q-r)/(2r) \quad \text{and} \quad A \geqslant B(q+r)/\big(2r(p-1)\big)$$

we make sure that all of them are non-negative, so they only contribute to the positivity of the expressions.

To explain how the values of the coefficients A, B, and C in Theorem 1.1 were obtained, let us consider Region 4 as a representative example. The other regions are treated similarly.

First, notice that the double framed term really is the dominant one, since $t<s<1$ implies

$$t^{\frac{3}{q}}s^{\frac{1}{r}-\frac{2}{q}},\ t^{\frac{1}{q}}s^{-\frac{1}{q}},\ t^{\frac{4}{q}}s^{\frac{1}{r}-\frac{3}{q}}\ t^{\frac{2}{q}}s^{-\frac{2}{q}},\ t^{\frac{1}{q}}s^{\frac{1}{r}},\ s^{\frac{1}{q}+\frac{1}{r}},\ t^{\frac{2}{q}}s^{\frac{1}{r}-\frac{1}{q}} < 1 = t^0s^0.$$

We can choose $B = 1$ and then take C large enough such that

$$r(2Cr - q + r) > \max\{28(q - r) + 2, 7(q - r)(2q - r) + 2\}.$$

Clearly, $C = 11q^3 r/\big((r-1)(q-r)\big)$ satisfies the above condition. This way the expression multiplying A in the coefficient of the dominant term is seven times larger than the expressions multiplying A in the coefficients of the two non-framed terms that contain A. Now we just have to take A large enough such that

$$\big(A(p-1) - C - 1\big)\big(pr(2Cr - q + r) - 2p\big)$$

is at least

$$\max\left\{\frac{7}{2}q(2Cr - q + r)^2,\ 14q(2Cr - q + r),\ 7q(2Cr - q + r)|q - 2r|,\right.$$
$$\left. 14q(q - r),\ \frac{7}{2}q(q - r)(3q - r)\right\}.$$

It is easy to see that $A = 88q^4 r/\big((p-1)(r-1)(q-r)\big)$ is one possible choice. Now the dominant term is more than seven times larger than the absolute value of any other term, which means that the dominant term dominates the sum of all other terms.

This way we accomplish the positivity of each of the expressions, which is exactly what we needed and the proof of $(\mathcal{A}3')$ is completed. This also completes the proof of Theorem 1.1.

In the next section, it will sometimes be more convenient to use the infinitesimal version of $(\mathcal{B}3)$:

$(\mathcal{B}3')$
$$-(d^2\mathcal{B})(u, v, w, U, V, W)(\Delta u, \Delta v, \Delta w, \Delta U, \Delta V, \Delta W) \geqslant 2u|\triangle v||\triangle w|.$$

Again, $(\mathcal{B}3')$ holds only for points (u, v, w, U, V, W) at which the second differential of $\mathcal{B}$ is well defined, i.e. for the points such that (u, v, w) does not lie on any of the three critical surfaces. The equivalence of $(\mathcal{B}3')$ and $(\mathcal{B}3)$ follows from the equivalence of $(\mathcal{A}3')$ and $(\mathcal{A}3)$.

3. APPLICATIONS

Here we present several applications of the existence of the Bellman function from Theorem 1.1. We need to emphasize that the following problems are quite classical and can be solved using more standard tools. We only provide quite straightforward solutions based on Theorem 1.1. Moreover, only the existence of the Bellman function with properties $(\mathcal{B}1)$–$(\mathcal{B}3)$ is needed, even though $(\mathcal{B}4)$ is quite convenient in Subsection 3.1. This existence can also follow if boundedness of the dyadic paraproduct is established in some other way, as commented in

the introduction. However, our goal is to illustrate how several classical problems become methodologically simple once we explicitly construct the function as in Theorem 1.1.

For two non-negative quantities A and B we will write $A \lesssim_P B$ if there exists a finite constant $C_P \geqslant 0$ depending on a set of parameters P such that $A \leqslant C_P B$.

3.1. Discrete-time martingales. Let us consider two martingales $X = (X_n)_{n=0}^{\infty}$ and $Y = (Y_n)_{n=0}^{\infty}$ with respect to the same filtration $(\mathcal{F}_n)_{n=0}^{\infty}$. Their *paraproduct* is a stochastic process $\big((X \cdot Y)_n\big)_{n=0}^{\infty}$ defined as

$$(X \cdot Y)_0 := 0, \quad (X \cdot Y)_n := \sum_{k=1}^{n} X_{k-1}(Y_k - Y_{k-1}) \quad \text{for } n \geqslant 1. \tag{3.1}$$

This process can be regarded as a particular case of Burkholder's martingale transform [5] of the martingale Y with respect to the shifted adapted process X. We have also imposed the martingale property on X, since we want to treat X and Y symmetrically and since this is required by the existence of the L^p estimates in the interior of the Banach triangle in Figure 1. We want to prove that for the exponents p, q, r satisfying (1.4) the estimate

$$\|(X \cdot Y)_n\|_{L^{r'}} \lesssim_{p,q,r} \|X_n\|_{L^p} \|Y_n\|_{L^q} \tag{3.2}$$

holds uniformly in the positive integer n, where r' is the conjugate exponent of r. Instead of proving (3.2) directly, we will rather show the estimate for the dualized form, i.e. that for an arbitrary random variable $Z \in L^r$ the inequality

$$\big|\mathbb{E}\big((X \cdot Y)_n Z\big)\big| \lesssim_{p,q,r} \|X_n\|_{L^p} \|Y_n\|_{L^q} \|Z\|_{L^r} \tag{3.3}$$

holds. This inequality is trivial unless all norms on the right-hand side are finite.

Let us introduce the third martingale $(Z_n)_{n=0}^{\infty}$ with $Z_n := \mathbb{E}(Z|\mathcal{F}_n)$. By splitting $Z = Z_{k-1} + (Z_k - Z_{k-1}) + (Z - Z_k)$ and using the martingale property in the form of $\mathbb{E}(Y_k - Y_{k-1}|\mathcal{F}_{k-1}) = 0$ and $\mathbb{E}(Z - Z_k|\mathcal{F}_k) = 0$, we can write

$$\begin{aligned}\mathbb{E}\big((X \cdot Y)_n Z\big) &= \sum_{k=1}^{n} \mathbb{E}\big(X_{k-1}(Y_k - Y_{k-1})Z\big) \\ &= \sum_{k=1}^{n} \mathbb{E}\big(X_{k-1}(Y_k - Y_{k-1})(Z_k - Z_{k-1})\big).\end{aligned}$$

The estimate (3.3) is now a clear consequence of the Cauchy–Schwarz, Hölder, Doob and Burkholder–Gundy inequalities. Again, we will give a more direct proof using the Bellman function (1.7).

It is enough to consider the times $k = 0, 1, \ldots, n$, but we need to show the estimate that is uniform in n. We can assume that $X_k, Y_k, Z_k \geqslant 0$ for $0 \leqslant k \leqslant n$,

as otherwise we split the variables X_n, Y_n, Z_n into positive and negative parts, and introduce three new martingales (for a fixed n):

$$U_k := \mathbb{E}(X_n^p|\mathcal{F}_k), \quad V_k := \mathbb{E}(Y_n^q|\mathcal{F}_k), \quad W_k := \mathbb{E}(Z_n^r|\mathcal{F}_k).$$

If we write $\boldsymbol{X}_k = (X_k, Y_k, Z_k, U_k, V_k, W_k)$, then property $(\mathcal{B}4)$ of the Bellman function $\mathcal{B}$ gives us

$$\mathcal{B}(\boldsymbol{X}_{k-1}) + (d\mathcal{B})(\boldsymbol{X}_{k-1})(\boldsymbol{X}_k - \boldsymbol{X}_{k-1}) \\ \geqslant \mathcal{B}(\boldsymbol{X}_k) + \frac{2}{3}X_{k-1}|Y_k - Y_{k-1}||Z_k - Z_{k-1}|,$$

from which we deduce

$$\mathcal{B}(\boldsymbol{X}_{k-1}) \geqslant \mathbb{E}\big(\mathcal{B}(\boldsymbol{X}_k)\big|\mathcal{F}_{k-1}\big) + \frac{2}{3}\mathbb{E}\big(X_{k-1}|Y_k - Y_{k-1}||Z_k - Z_{k-1}|\,\big|\mathcal{F}_{k-1}\big),$$

by taking the conditional expectation with respect to $\mathcal{F}_{k-1}$ and using the martingale property. Finally, taking the expectation of the above inequality, summing over $k = 1, \ldots, n$, telescoping, and using $(\mathcal{B}2)$ gives

$$\frac{2}{3}\sum_{k=1}^{n}\mathbb{E}(X_{k-1}|Y_k - Y_{k-1}||Z_k - Z_{k-1}|) \leqslant \mathbb{E}\mathcal{B}(\boldsymbol{X}_0) - \mathbb{E}\mathcal{B}(\boldsymbol{X}_n) \\ \leqslant \mathcal{C}_{p,q,r}\mathbb{E}\left(\frac{1}{p}U_0 + \frac{1}{q}V_0 + \frac{1}{r}W_0\right) = \mathcal{C}_{p,q,r}\left(\frac{1}{p}\|X_n\|_{L^p}^p + \frac{1}{q}\|Y_n\|_{L^q}^q + \frac{1}{r}\|Z_n\|_{L^r}^r\right).$$

Homogenizing the above inequality, we get the desired estimate (3.3) and hence also (3.2).

3.2. Continuous-time martingales. Let $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ be two continuous-time *càdlàg* martingales with respect to the filtration $(\mathcal{F}_t)_{t\geqslant 0}$ that satisfies the "usual hypotheses" [22]. In this case the *martingale paraproduct* is also a stochastic process $\big((X \cdot Y)_t\big)_{t\geqslant 0}$, but now defined via the stochastic integral

$$(X \cdot Y)_t := \int_{0+}^{t} X_{s-}dY_s. \tag{3.4}$$

Since we are allowed to choose dense subspaces on which the initial definition makes sense (and later extend by continuity), we can conveniently assume that X is bounded in L^∞ and Y is bounded in L^2. We want to prove that (3.4) satisfies the same L^p estimates as (3.1). To do so, we take $(\pi_m)_{m=1}^{\infty}$ to be a refining sequence of partitions

$$0 = t_0^{(m)} < t_1^{(m)} < t_2^{(m)} < \ldots < t_{n(m)}^{(m)} = t$$

such that $\lim_{m\to\infty} \operatorname{mesh}(\pi_m) = 0$. We can calculate (3.4) as the limit of the Riemann sums in the following way:

$$\int_{0+}^{t} X_{s-}\,dY_s = \lim_{m\to\infty} \sum_{k=1}^{n(m)} X_{t_{k-1}^{(m)}} (Y_{t_k^{(m)}} - Y_{t_{k-1}^{(m)}}). \tag{3.5}$$

The above limit is interpreted as the convergence in probability; for more details see [22]. Notice that the right-hand side of (3.5) is actually a limit of discrete-time martingale paraproducts (3.1). By passing to an a.s. convergent subsequence, using Fatou's lemma, and applying (3.2), we get the desired estimate for (3.4):

$$\|(X\cdot Y)_t\|_{L^{r'}} \leqslant \sup_m \Big\| \sum_{k=1}^{n(m)} X_{t_{k-1}^{(m)}} (Y_{t_k^{(m)}} - Y_{t_{k-1}^{(m)}}) \Big\|_{L^{r'}} \lesssim_{p,q,r} \|X_t\|_{L^p} \|Y_t\|_{L^q}$$

for the exponents p, q, r satisfying (1.4).

As a special case we can consider martingales with respect to the augmented filtration of the one-dimensional Brownian motion $(B_t)_{t\geqslant 0}$. If we also assume that $Y_0 = 0$, then

$$(X\cdot Y)_t = \int_0^t X_s\,dY_s, \tag{3.6}$$

because $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ now a.s. have continuous paths. We remark that (3.6) are the martingale paraproducts studied by Bañuelos and Bennett in [2] and they established L^p, H^p, and BMO estimates for (3.6). Their proof of the L^p estimates uses Doob's inequality and the Burkholder–Gundy inequality.

Yet another short proof of the L^p estimates in this particular case can be given by applying Itô's formula in combination with $(\mathcal{B}2)$ and $(\mathcal{B}3')$, instead of approximating by discrete-time processes. However, for that purpose our Bellman function should be of class C^2 on the whole domain. This is achieved by shrinking the domain slightly and passing to $\mathcal{B}_\varepsilon$ as in the next section; we omit the details.

3.3. Heat flow paraproducts. In order to be able to use the constructed Bellman function in relationship with the heat equation, we should first "smoothen it up". Let us fix a non-negative even C^∞ function φ supported in $(-1,1)^3$ with integral one. For any $\varepsilon > 0$ we define the function $\mathcal{A}_\varepsilon\colon (\varepsilon,\infty)^3 \to \mathbb{R}$ by the formula

$$\mathcal{A}_\varepsilon(u,v,w) := \int_{(-\varepsilon,\varepsilon)^3} \varepsilon^{-3}\varphi(\varepsilon^{-1}a, \varepsilon^{-1}b, \varepsilon^{-1}c)\mathcal{A}(u-a, v-b, w-c)\,da\,db\,dc.$$

In words, the function $\mathcal{A}_\varepsilon$ is the convolution of $\mathcal{A}$ with the L^1-normalized dilation of φ. The newly obtained function is clearly of class C^∞. We integrate $(\mathcal{A}3)$ translated by (a,b,c) and multiplied by $\varepsilon^{-3}\varphi(\varepsilon^{-1}a, \varepsilon^{-1}b, \varepsilon^{-1}c)$, and then "symmetrize" in (a,b,c) and use the fact that φ is even. That way we conclude that

$\mathcal{A}_\varepsilon$ still satisfies the condition $(\mathcal{A}3)$ and consequently also $(\mathcal{A}3')$ at every point of its domain. By the formula (1.7) with $\mathcal{A}_\varepsilon$ in the place of $\mathcal{A}$ we can define a C^∞ function $\mathcal{B}_\varepsilon$ satisfying property $(\mathcal{B}3')$ for any $u, v, w > \varepsilon$ and $U \geqslant u^p$, $V \geqslant v^q$, $W \geqslant w^r$. Moreover, property $(\mathcal{A}2)$ is retained up to an additional loss by the factor $\max\{2^p, 2^q, 2^r\}$, which in turn guarantees $(\mathcal{B}2)$ for some (sufficiently large) constant $\mathcal{C}_{p,q,r}$ independent of ε.

Now suppose that f, g, h are compactly supported C^∞ functions on $\mathbb{R}$. Also, let $k(x,t) := \frac{1}{\sqrt{2\pi t}} \exp\big(-x^2/(2t)\big)$ be the heat kernel on the real line and u be the heat extension of f:

$$u(x,t) := \int_{\mathbb{R}} f(y) k(x-y,t)\, dy.$$

Note that u is the solution of the heat equation $\partial_t u = \frac{1}{2}\partial_x^2 u$ with the initial condition $\lim_{t\to 0+} u(x,t) = f(x)$. Analogously we define v and w to be the heat extensions of g and h.

We can define the *heat paraproduct*, i.e. the paraproduct with respect to the heat semigroup as a trilinear form

$$\Lambda(f,g,h) := \int_{\mathbb{R}} \int_0^\infty u(x,t)\, \partial_x v(x,t)\, \partial_x w(x,t)\, dt\, dx. \tag{3.7}$$

If we define

$$\varphi_s(x) := k(x, s^2), \qquad \psi_s(x) := -2^{1/2} s\, \partial_x k(x, s^2)$$

and substitute $t = s^2$, we get a more familiar expression:

$$\Lambda(f,g,h) = \int_{\mathbb{R}} \int_0^\infty (f * \varphi_s)(x)\, (g * \psi_s)(x)\, (h * \psi_s)(x)\, \frac{ds}{s}\, dx. \tag{3.8}$$

Smooth paraproducts like (3.8) appear naturally in the proof of the T1 theorem (see [10]), although one usually needs to be more flexible when choosing a bump function φ_s and a mean zero bump function ψ_s.

Again, we want to prove some L^p estimates for (3.7), i.e.

$$|\Lambda(f,g,h)| \lesssim_{p,q,r} \|f\|_{L^p(\mathbb{R})} \|g\|_{L^q(\mathbb{R})} \|h\|_{L^r(\mathbb{R})},$$

where p, q, r are exponents satisfying (1.4). To do so we will imitate the "heating" technique by Nazarov and Volberg [19] or Petermichl and Volberg [21].

Assume that f, g, h are non-negative and that none of them is identically zero. Fix $R > 0$, $\delta > 0$, $T > 2\delta$, and observe that $u(x,t), v(x,t), w(x,t) > \varepsilon$ whenever $x \in [-R, R]$, $t \in [\delta, T-\delta]$ for some sufficiently small $\varepsilon > 0$ depending on R, δ, T, and the functions f, g, h. We introduce U, V, W as the heat extensions of f^p, g^q, h^r respectively and define

$$b(x,t) := \mathcal{B}_\varepsilon\big(u(x,t), v(x,t), w(x,t), U(x,t), V(x,t), W(x,t)\big),$$

where $\mathcal{B}_\varepsilon$ is as above. It is easy to calculate that

$$\begin{aligned}\big(\partial_t - \tfrac{1}{2}\partial_x^2\big)b(x,t) &= (\nabla\mathcal{B}_\varepsilon)(u,v,w,U,V,W)\cdot\big(\partial_t - \tfrac{1}{2}\partial_x^2\big)(u,v,w,U,V,W)\\ &\quad - \tfrac{1}{2}(d^2\mathcal{B}_\varepsilon)(u,v,w,U,V,W)(\partial_x u,\partial_x v,\partial_x w,\partial_x U,\partial_x V,\partial_x W).\end{aligned}$$

(We have omitted writing the variables x,t on the right-hand side.) Since u,v,w, U,V,W all satisfy the heat equation, the first term on the right-hand side is zero and by $(\mathcal{B}3')$ we get

$$\big(\partial_t - \tfrac{1}{2}\partial_x^2\big)b(x,t) \geqslant \pm u(x,t)\,\partial_x v(x,t)\,\partial_x w(x,t).$$

It remains to integrate this inequality over $[-R,R]\times[\delta,T-\delta]$ with an appropriate weight, use Green's formula, and then let $\delta\to 0$, $R,T\to\infty$. We omit the details and refer to [19] and [21].

Let us emphasize once again that the previous trick of "smoothing" the Bellman function was already used in [19] and [21] and no explicit formula is needed for its application.

Acknowledgments. We would like to thank the anonymous referee for several useful comments and suggestions that improved the readability of this paper.

REFERENCES

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, Inc., New York 1992.
[2] R. Bañuelos and A. G. Bennett, *Paraproducts and commutators of martingale transforms*, Proc. Amer. Math. Soc. 103 (4) (1988), pp. 1226–1234.
[3] R. Bañuelos and A. Osękowski, *On the Bellman function of Nazarov, Treil and Volberg*, Math. Z. 278 (2014), pp. 385–399.
[4] Á. Bényi, D. Maldonado, and V. Naibo, *What is ... a paraproduct?*, Notices Amer. Math. Soc. 57 (7) (2010), pp. 858–860.
[5] D. L. Burkholder, *Martingale transforms*, Ann. Math. Statist. 37 (1966), pp. 1494–1504.
[6] D. L. Burkholder, *Boundary value problems and sharp inequalities for martingale transforms*, Ann. Probab. 12 (3) (1984), pp. 647–702.
[7] A. Carbonaro and O. Dragičević, *Bellman function and linear dimension-free estimates in a theorem of Bakry*, J. Funct. Anal. 265 (7) (2013), pp. 1085–1104.
[8] A. Carbonaro and O. Dragičević, *Functional calculus for generators of symmetric contraction semigroups*, Duke Math. J. 166 (5) (2017), pp. 937–974.
[9] D. L. Cohn, *Measure Theory*, second edition, Birkhäuser/Springer, New York 2013.
[10] G. David and J.-L. Journé, *A boundedness criterion for generalized Calderón–Zygmund operators*, Ann. of Math. (2) 120 (2) (1984), pp. 371–397.
[11] B. Davis, *On the L^p norms of stochastic integrals and other martingales*, Duke Math. J. 43 (4) (1976), pp. 697–704.
[12] P. Durcik, *An L^4 estimate for a singular entangled quadrilinear form*, Math. Res. Lett. 22 (5) (2015), pp. 1317–1332.
[13] P. Durcik, *L^p estimates for a singular entangled quadrilinear form*, Trans. Amer. Math. Soc. 369 (10) (2017), pp. 6935–6951.
[14] S. Janson and J. Peetre, *Paracommutators—boundedness and Schatten–von Neumann properties*, Trans. Amer. Math. Soc. 305 (2) (1988), pp. 467–504.

[15] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam. 28 (4) (2012), pp. 1143–1164.
[16] V. Kovač and K. A. Škreb, *One modification of the martingale transform and its applications to paraproducts and stochastic integrals*, J. Math. Anal. Appl. 426 (2) (2015), pp. 1143–1163.
[17] F. L. Nazarov and S. R. Treil, *The hunt for a Bellman function: Applications to estimates of singular integral operators and to other classical problems in harmonic analysis* (in Russian), Algebra i Analiz 8 (1996), pp. 32–162; English transl. in St. Petersburg Math. J. 8 (1997), pp. 721–824.
[18] F. L. Nazarov and A. Volberg, *The Bellman function, the two-weight Hilbert transform, and embeddings of the model spaces K_θ*, J. Anal. Math. 87 (2002), pp. 385–414.
[19] F. L. Nazarov and A. Volberg, *Heating of the Ahlfors–Beurling operator and estimates of its norm*, St. Petersburg Math. J. 15 (2004), pp. 563–573.
[20] A. Osękowski, *Sharp Martingale and Semimartingale Inequalities*, Birkhäuser/Springer, Basel 2012.
[21] S. Petermichl and A. Volberg, *Heating of the Ahlfors–Beurling operator: Weakly quasiregular maps on the plane are quasiregular*, Duke Math. J. 112 (2) (2002), pp. 281–305.
[22] P. E. Protter, *Stochastic Integration and Differential Equations*, second edition, Springer, Berlin 2005.
[23] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Systems 28 (2) (2008), pp. 657–688.
[24] C. Thiele, *Wave Packet Analysis*, American Mathematical Society, Providence, RI, 2006.
[25] V. Vasyunin and A. Volberg, *Bellster and others* (2008), preprint.
[26] Wolfram Research, Inc., *Mathematica, Version 9.0,* Champaign, IL (2012).

Vjekoslav Kovač
Department of Mathematics, Faculty of Science
University of Zagreb
Bijenička cesta 30
10000 Zagreb, Croatia
E-mail: vjekovac@math.hr

Kristina Ana Škreb
Faculty of Civil Engineering
University of Zagreb
Fra Andrije Kačića Miošića 26
10000 Zagreb, Croatia
E-mail: kskreb@grad.hr

Received on 22.9.2016;
revised version on 3.7.2017

Probability and Mathematical Statistics 38, z. 2, 2018

CONTENTS OF VOLUME 38

Probability and Mathematical Statistics 38, z. 2, 2018