Discrete Mathematics and Theoretical Computer Science DMTCS vol. **8**, 2006, 31–56

A Survey of Multivariate Aspects of the Contraction Method

Ralph Neininger[1†] and Ludger Rüschendorf[2]

[1] *Department of Mathematics J.W. Goethe University, Frankfurt a.M, Germany*
[2] *Department of Mathematics, University of Freiburg, Freiburg, Germany*

received Jan 19, 2006, *accepted Feb 1, 2006*.

We survey multivariate limit theorems in the framework of the contraction method for recursive sequences as arising in the analysis of algorithms, random trees or branching processes. We compare and improve various general conditions under which limit laws can be obtained, state related open problems and give applications to the analysis of algorithms and branching recurrences.

Keywords: Analysis of algorithms, random trees, branching processes, weak convergence, probability metric, multivariate analysis.

AMS subject classifications. Primary: 68P10, 68Q25, 60F05; secondary: 60J80, 05C05.

1 Introduction

We survey multivariate limit laws for sequences of random vectors which satisfy distributional recursions as they appear under various models of randomness for parameters of trees, characteristics of divide-and-conquer algorithms, or, more generally, for quantities related to recursive structures or branching processes.

While the area of probabilistic analysis of algorithms, since its introduction in the 60s of the last century by Knuth [25, 26, 27] has been dominated by analytic techniques based on generating functions, over the last decade, among other probabilistic techniques, the so called contraction method has been developed. This method was first introduced for the analysis of Quicksort in Rösler [45] and further on developed independently in Rösler [46] and Rachev and Rüschendorf [43], and later on in Rösler [48] and Neininger and Rüschendorf [40, 41], see also the survey article of Rösler and Rüschendorf [49].

In this survey we discuss multivariate aspects of the approach of the contraction method. In particular we study various conditions, under which multivariate limit laws can be established, mention applications to the probabilistic analysis of algorithms and connections to other areas as branching processes, and indicate as to which extend a multivariate point of view may also add flexibility to univariate studies.

†Research supported by an Emmy Noether fellowship of the DFG.

Throughout this note we study sequences of d-dimensional vectors $(Y_n)_{n\geq 0}$, which satisfy the distributional recursion

$$Y_n \stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A_r(n) Y^{(r)}_{I_r^{(n)}} + b_n, \quad n \geq n_0, \tag{1}$$

with $(A_1(n), \ldots, A_K(n), b_n, I^{(n)})$, $(Y_n^{(1)}), \ldots, (Y_n^{(K)})$ independent, $A_1(n), \ldots, A_K(n)$ random $d \times d$-matrices, b_n a random d-dimensional vector, $I^{(n)}$ a vector of random cardinalities $I_r^{(n)} \in \{0, \ldots, n\}$ and $(Y_n^{(1)}), \ldots, (Y_n^{(K)})$ identically distributed as (Y_n). The $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution and we have $n_0 \geq 1$. Note that we do not define the sequence (Y_n) by (1), we only assume that (Y_n) satisfies recurrence (1). The number $K \geq 1$ is, for simplicity of presentation, considered to be fixed in our discussion. However, extensions to random K depending on n have also been studied.

We will indicate below how various problems form the area of analysis of algorithms and other areas fit into this general scheme by taking special choices for the parameters $A_1(n), \ldots, A_K(n), b_n, I^{(n)}$, K, and n_0.

We normalize Y_n by

$$X_n := \Sigma_n^{-1/2}(Y_n - M_n), \quad n \geq 0, \tag{2}$$

where $M_n \in \mathbb{R}^d$ and Σ_n is a positive-definite square matrix. If first or second moments for Y_n are finite the natural choices for M_n and Σ_n are the mean vector $\mathbb{E}\, Y_n$ and the covariance matrix $\mathrm{Cov}(Y_n)$ respectively. The X_n satisfy

$$X_n \stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A_r^{(n)} X^{(r)}_{I_r^{(n)}} + b^{(n)}, \quad n \geq n_0, \tag{3}$$

with

$$A_r^{(n)} := \Sigma_n^{-1/2} A_r(n) \Sigma^{1/2}_{I_r^{(n)}}, \quad b^{(n)} := \Sigma_n^{-1/2}\left(b_n - M_n + \sum_{r=1}^{K} A_r(n) M_{I_r^{(n)}}\right) \tag{4}$$

and independence relations as in (1).

The contraction method provides transfer theorems, which state that, under various conditions, convergence of the coefficients $A_r^{(n)} \to A_r^*$, $b^{(n)} \to b^*$ implies weak convergence of the parameters (X_n) to a limit X. The limit distribution $\mathcal{L}(X)$ satisfies a fixed-point equation obtained from (3) by letting formally $n \to \infty$:

$$X \stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A_r^* X^{(r)} + b^*. \tag{5}$$

Here $(A_1^*, \ldots, A_K^*, b^*), X^{(1)}, \ldots, X^{(K)}$ are independent and $X^{(r)} \sim X$ for $r = 1, \ldots, K$, where $X \sim Y$ denotes equality of the distributions of X, Y.

In the context of the contraction method, the fixed-point equation (5) is used to define a map T from the space $\mathcal{M}^d$ of all Borel measures on $\mathbb{R}^d$ to itself by

$$\begin{aligned} T : \mathcal{M}^d &\to \mathcal{M}^d \\ \mu &\mapsto \mathcal{L}\Big(\sum_{r=1}^{K} A_r^* Z^{(r)} + b^* \Big), \end{aligned} \tag{6}$$

where $(A_1^*, \ldots, A_K^*, b^*), Z^{(1)}, \ldots, Z^{(K)}$ are independent and $Z^{(r)} \sim \mu$ for $r = 1, \ldots, K$. Clearly, a random variable X satisfies (5) if and only if its distribution $\mathcal{L}(X)$ is a fixed-point of the map T.

Usually, maps of type T have multiple fixed-points in $\mathcal{M}^d$, but once restricted to appropriate subspaces of $\mathcal{M}^d$ such fixed-points become unique. The name of the method refers to the fact that such unique fixed-points are obtained by showing that the restriction of T to suitable subspaces of $\mathcal{M}^d$, which are endowed with complete metrics, form a contraction in the sense of Banach's fixed-point theorem and that these fixed-point measures are the distributional limits of the rescaled quantities X_n as given in the basic recurrence (3).

Various probability metrics have been proposed to obtain Lipschitz properties for the maps T. It turned out that different classes of recursive problems of type (3) necessitate different metrics. Two classes of probability metrics are of particular importance in this respect, the minimal L_p metrics and the Zolotarev metrics.

In section 2 we recall these probability metrics together with Lipschitz properties of the map T, then, in section 3, we collect multivariate limit laws, discuss the various conditions needed, give some improvements and state an open problem. In section 4 applications of the general framework are given. First, we discuss some known applications from the area of algorithms and random trees, then we develop asymptotic results for branching processes, that can also be covered by the general framework.

2 Probability metrics

The minimal L_p metric ℓ_p, $p > 0$, is defined for $\mu, \nu \in \mathcal{M}_p^d := \{\sigma \in \mathcal{M}^d : \int \|x\|^p d\sigma(x) < \infty\}$ by

$$\ell_p(\mu, \nu) = \inf\{(\mathbb{E}\,\|X - Y\|^p)^{1\wedge(1/p)} : X \sim \mu, Y \sim \nu\}, \tag{7}$$

where $\|\cdot\|$ denotes Euclidean norm. The metric space $(\mathcal{M}_p^d, \ell_p)$ is complete. The metric ℓ_p has frequently been used in the analysis of algorithms since its introduction in this context by Rösler [45] for the analysis of Quicksort, see, e.g., [31, 39, 37]. An advantage of this metric is that for estimates it is convenient to work with optimal couplings of measures, i.e., with choices of random variables X, Y such that the infimum in (7) becomes a minimum.

Another important class of metrics are the Zolotarev metrics ζ_s, $s > 0$, see [53], defined for $\mu, \nu \in \mathcal{M}^d$, with $X \sim \mu$ and $Y \sim \nu$, by

$$\zeta_s(\mu, \nu) = \sup_{f \in \mathcal{F}_s} |\,\mathbb{E}\,(f(X) - f(Y))|,$$

where for $s = m + \alpha$, $0 < \alpha \leq 1$, $m \in \mathbb{N}_0$, and

$$\mathcal{F}_s := \{f \in C^m(\mathbb{R}^d, \mathbb{R}) : \|f^{(m)}(x) - f^{(m)}(y)\| \leq \|x - y\|^\alpha\},$$

where $C^m(\mathbb{R}^d, \mathbb{R})$ denotes the space of m times differentiable functions and $f^{(m)}$ the mth derivative of a function f. A nontrivial issue is to decide whether $\zeta_s(\mu,\nu)$ is finite or not. Subsequently we will only need that for finiteness of $\zeta_s(\mathcal{L}(X), \mathcal{L}(Y))$ it is sufficient that X and Y have identical mixed moments up to order m and both a finite absolute sth moment. Since ζ_s is of main interest for $s \le 3$, we introduce the following special spaces of measures to ensure finiteness. For $2 < s \le 3$ we have to control the mean and the covariances in order to obtain the finiteness of the ζ_s metric. We define for $0 < s \le 3$, a vector $m \in \mathbb{R}^d$, and a symmetric positive semidefinite $d \times d$ matrix Σ the spaces

$$\begin{aligned}
\mathcal{M}_s^d(m,\Sigma) &:= \{\mu \in \mathcal{M}_s^d : \mathbb{E}\,\mu = m, \mathrm{Cov}(\mu) = \Sigma\}, \quad 2 < s \le 3 \\
\mathcal{M}_s^d(m,\Sigma) &:= \mathcal{M}_s^d(m) := \{\mu \in \mathcal{M}_s^d : \mathbb{E}\,\mu = m\}, \quad 1 < s \le 2, \\
\mathcal{M}_s^d(m,\Sigma) &:= \mathcal{M}_s^d, \quad 0 < s \le 1.
\end{aligned}$$

Then ζ_s is finite on $\mathcal{M}_s^d(m,\Sigma) \times \mathcal{M}_s^d(m,\Sigma)$ for all $0 < s \le 3$, $m \in \mathbb{R}^d$, and symmetric, positive semidefinite Σ. Note that for $s < 2$ the Σ in $\mathcal{M}_s^d(m,\Sigma)$ has no meaning as has the m for $0 < s \le 1$.

The most important property of ζ_s for the contraction method is that it is $(s,+)$ ideal, i.e., we have

$$\zeta_s(X+Z, Y+Z) \le \zeta_s(X,Y), \quad \zeta_s(cX, cY) = |c|^s \zeta_s(X,Y), \tag{8}$$

for all Z independent of X, Y and $c \in \mathbb{R} \setminus \{0\}$, valid, whenever these distances are finite.

Note that both, convergence in ℓ_p for some $p > 0$ or in ζ_s for some $s > 0$, imply weak convergence.

From the perspective of the contraction method it is crucial under which conditions on $(A_1^*, \dots, A_K^*, b^*)$ and on which spaces the map T defined in (6) is a contraction. We have the following estimates on Lipschitz constants for T: For $0 < s \le 3$ restricted to the metric space $(\mathcal{M}_s^d(m,\Sigma), \zeta_s)$ we have

$$\zeta_s(T(\mu), T(\nu)) \le \Big(\sum_{r=1}^K \mathbb{E}\,\|A_r\|_{\mathrm{op}}^s\Big)\zeta_s(\mu,\nu), \quad \mu,\nu \in \mathcal{M}_s^d(m,\Sigma),$$

where, for a matrix A, we denote $\|A\|_{\mathrm{op}} := \sup_{\|x\| \le 1} \|Ax\|$. On the metric space $(\mathcal{M}_p^d, \ell_p)$ for $p \ge 1$ we have

$$\ell_p(T(\mu), T(\nu)) \le \Big(\sum_{r=1}^K \Big\| \|A_r\|_{\mathrm{op}} \Big\|_p\Big)\ell_p(\mu,\nu), \quad \mu,\nu \in \mathcal{M}_p^d,$$

where, for random variates, $\|\cdot\|_p$ denotes the L_p norm. On $(\mathcal{M}_2^d(0), \ell_2)$ we have

$$\ell_2(T(\mu), T(\nu)) \le \left\| \sum_{r=1}^K \mathbb{E}\left[(A_r^*)^t A_r\right] \right\|_{\mathrm{op}}^{1/2} \ell_2(\mu,\nu), \quad \mu,\nu \in \mathcal{M}_2^d(0),$$

where A^t denotes the transposed of a matrix A. See, for references [6, 37, 40].

3 Multivariate limit laws

In this section we state some general limit laws that transfer convergence of the coefficients $A_r^{(n)} \to A_r^*$, $b^{(n)} \to b^*$ to the quantities itself, cf. (3) and (5), and discuss the various conditions needed from the point of view of the probability metric used. We denote by $\xrightarrow{L_s}$ convergence in the L_s norm. The following theorem can, in part, be found in [37, 40]:

Theorem 3.1 *Let (X_n) be s-integrable, $0 < s \leq 3$ and satisfy the recurrence (3), where the X_n are centered if $s > 1$ and have the identity matrix Id_d as covariance matrix if $s > 2$. Assume that, as $n \to \infty$,*

$$\left(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}\right) \xrightarrow{L_s} \left(A_1^*, \dots, A_k^*, b^*\right), \tag{9}$$

$$\mathbb{E} \sum_{r=1}^K \left\|A_r^*\right\|_{\mathrm{op}}^s < 1, \text{ and} \tag{10}$$

$$\mathbb{E}\left[\mathbf{1}_{\{I_r^{(n)} \leq \ell\} \cup \{I_r^{(n)} = n\}} \|A_r^{(n)}\|_{\mathrm{op}}^s\right] \to 0 \tag{11}$$

for all $\ell \in \mathbb{N}$ and $r = 1 \dots, K$. Then we have

$$\zeta_s(X_n, X) \to 0, \quad n \to \infty,$$

where $\mathcal{L}(X)$ is the in $\mathcal{M}_s^d(0, \mathrm{Id}_d)$ unique fixed-point of T.

In the case $s = 2$ and (10) replaced by

$$\sum_{r=1}^K \mathbb{E} \left\|(A_r^*)^t A_r^*\right\|_{\mathrm{op}} < 1, \tag{12}$$

we have

$$\ell_2(X_n, X) \to 0, \quad n \to \infty.$$

Note that the cases $0 < s \leq 1$, $1 < s \leq 2$, and $2 < s \leq 3$ are substantially different from the perspective of applications. For the case $2 < s \leq 3$ the condition $\mathcal{L}(X_n) \in \mathcal{M}_s^d(0, Id_d)$ requires that an original sequence (Y_n) is scaled in (2) by its exact mean M_n and covariance matrix Σ_n. For the verification of the convergence of $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ in (9) one has to draw back to the representations of $A_r^{(n)}$ and $b^{(n)}$ given in (4) that contain M_n and Σ_n. Hence, for the application of Theorem 3.1 with $2 < s \leq 3$ one needs to know $\mathbb{E}\, Y_n$ and $\mathrm{Cov}(Y_n)$ in advance. This is different for $s \leq 2$. For $1 < s \leq 2$ we only need $\mathcal{L}(X_n) \in \mathcal{M}_s^d(0)$, thus by the same argument, $M_n = \mathbb{E}\, Y_n$ needs to be known in advance but $\mathrm{Cov}(Y_n)$ may be unknown. Moreover, in the case $s = 2$, convergence in ζ_2 or ℓ_2 both imply convergence of all second (mixed) moments, i.e., convergence of the covariance matrix $\mathrm{Cov}(X_n)$ to $\mathrm{Cov}(X)$. This fact will be exploited in the applications in sections 4.1 and 4.2. If Theorem 3.1 is applied with $s \leq 1$ there are no conditions on the first two moments of X_n, only $\|X_n\|_s < \infty$ is needed. Similarly, for $s = 1$, convergence in $\zeta_1 = \ell_1$ implies convergence of the expectations, $\mathbb{E}\, X_n \to \mathbb{E}\, X$.

By $\|(A_r^*)^t A_r^*\|_{\mathrm{op}} \leq \|A_r^*\|_{\mathrm{op}}^2$ we obtain that condition (12) is weaker than condition (10) for $s = 2$. However, the map T is a contraction on $(\mathcal{M}_2^d(0), \ell_2)$ under the even weaker condition

$$\left\|\sum_{r=1}^K \mathbb{E}\left[(A_r^*)^t A_r^*\right]\right\|_{\mathrm{op}} < 1, \tag{13}$$

cf. Burton and Rösler [6, Theorem 1], for $K = 1$, and Neininger [37, Lemma 3.1]. Since, intuitively, such an underlying contraction may be sufficient to obtain a convergence result as in Theorem 3.1 we are led to the following open problem:

Problem 3.2 *Weaken condition (12) in Theorem 3.1, so that the assertion $\ell_2(X_n, X) \to 0$ remains true. Can one replace condition (12) by condition (13)?*

Note that weakening (12) towards (13) has the additional advantage that the norm in (13) in applications typically is easy to compute since only the norm of a fixed matrix has to be computed, whereas for the expectation in (12) one has to do an integration over the possibly complicated norm there, see (30) and (31) in section 4.2 for an example.

We will show in Theorem 3.3 below that we can replace (12) by the weaker condition

$$\limsup_{n\to\infty} \sum_{r=1}^{K} \mathbb{E} \left\| \mathbb{E} \left[(A_r^{(n)})^t A_r^{(n)} \middle| \mathcal{A}_r^{(n)} \right] \right\|_{\mathrm{op}} < 1, \tag{14}$$

where $\mathcal{A}_r^{(n)}$ is the σ-algebra generated by $I_r^{(n)}$, $\mathcal{A}_r^{(n)} = \sigma(I_r^{(n)}) \subset \mathcal{A}$, with an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Note that (14) with $\mathcal{A}_r^{(n)} = \mathcal{A}$ for all $n \geq 1$ and $r = 1, \ldots, K$ coincides, under (9), with condition (12), whereas (14) with the trivial σ-algebra $\mathcal{A}_r^{(n)} = \{\emptyset, \Omega\}$ for all $n \geq 1$ and $r = 1, \ldots, K$ is almost condition (13), only the sum being outside the norm. The smaller $\mathcal{A}_r^{(n)}$ is, the weaker is condition (14).

In the special case of diagonal matrices $(A_1^{(n)}, \ldots, A_K^{(n)})$ the assertion of Theorem 3.1 remains true when (12) and (11) are replaced by

$$\limsup_{n\to\infty} \sum_{i=0}^{n} \max_{1\leq k\leq d} \mathbb{E} \sum_{r=1}^{K} \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right) < 1,$$

$$\lim_{n\to\infty} \sum_{i\in\{0,\ldots,\ell\}\cup\{n\}} \max_{1\leq k\leq d} \mathbb{E} \sum_{r=1}^{K} \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right) = 0,$$

see Neininger [37, Corollary 4.2]. Here, the expectation inside the maximum corresponds to the expectation inside the norm in (13). By $(A)_{ij}$ we denote the ijth entry of a matrix A.

In the case of branching recurrences discussed in section 4.3, that is when $I_r^{(n)} = n-1$ for $r = 1, \ldots, K$ and general $(A_1^{(n)}, \ldots, A_K^{(n)}, b^{(n)})$ not depending on n, we are able to replace (12) by (13), see Theorem 4.4 below.

Theorem 3.3 *Let (X_n) be square integrable and satisfy the recurrence (3), where the X_n are centered. Assume that, as $n \to \infty$,*

$$\left(A_1^{(n)}, \ldots, A_K^{(n)}, b^{(n)} \right) \xrightarrow{L_2} \left(A_1^*, \ldots, A_k^*, b^* \right), \tag{15}$$

$$\limsup_{n\to\infty} \sum_{r=1}^{K} \mathbb{E} \left\| \mathbb{E} \left[(A_r^{(n)})^t A_r^{(n)} \middle| I_r^{(n)} \right] \right\|_{\mathrm{op}} < 1, \tag{16}$$

$$\mathbb{E} \left[\mathbf{1}_{\{I_r^{(n)}\leq\ell\}\cup\{I_r^{(n)}=n\}} \left\| (A_r^{(n)})^t A_r^{(n)} \right\|_{\mathrm{op}} \right] \to 0, \tag{17}$$

for all $\ell \in \mathbb{N}$ and $r = 1 \ldots, K$. Then we have

$$\ell_2(X_n, X) \to 0, \quad n \to \infty,$$

where $\mathcal{L}(X)$ is the in $\mathcal{M}_2^d(0)$ unique fixed-point of T.

Proof: By Jensen's inequality (16) implies $\| \sum \mathbb{E}\,[(A_r^*)^t A_r^*]\|_{\mathrm{op}} < 1$. By the definition of $b^{(n)}$ we have $\mathbb{E}\, b^{(n)} = 0$ for all $n \geq n_0$. Thus, the L_2-convergence of $(b^{(n)})$ implies $\mathbb{E}\, b^* = 0$. Therefore, by Lemma 3.1 in Neininger [37], T has a unique fixed-point $\mathcal{L}(X)$ in $\mathcal{M}_2^d(0)$. Let $X_n^{(r)} \sim X_n$, $X^{(r)} \sim X$ so that $(X_n^{(r)}, X^{(r)})$ are optimal couplings of (X_n, X) for all $n \in \mathbb{N}$ and $r = 1, \ldots, K$ and that $(A_1^{(n)}, \ldots, A_K^{(n)}, b^{(n)}, I^{(n)})$, $(X_n^{(1)}, X^{(1)}), \ldots, (X_n^{(K)}, X^{(K)})$ are independent. The first step is to derive an estimate of $\ell_2^2(X_n, X)$ in terms of $\ell_2^2(X_i, X)$ with indices $i \leq n-1$. This reduction inequality, cf. (21), for the sequence $(\ell_2^2(X_n, X))$ will be sufficient to deduce $\ell_2(X_n, X) \to 0$. We use the representations (3) and (5) of X_n and X respectively. For the $X_n^{(r)}$ and $X^{(r)}$ occurring there we use optimal couplings to keep the arising distances small. For $n \geq n_0$,

$$\begin{aligned}
\ell_2^2(X_n, X) &\leq \left\| \sum_{r=1}^K \left(A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right) + b^{(n)} - b^* \right\|_2^2 \\
&= \sum_{r=1}^K \left\| A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right\|_2^2 + \left\| b^{(n)} - b^* \right\|_2^2 \\
&\qquad + \sum_{\substack{r,s=1 \\ r \neq s}}^K \mathbb{E} \left\langle A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)}, A_s^{(n)} X_{I_s^{(n)}}^{(s)} - A_s^* X^{(s)} \right\rangle \\
&\qquad + 2 \sum_{r=1}^K \mathbb{E} \left\langle A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)}, b^{(n)} - b^* \right\rangle.
\end{aligned} \tag{18}$$

The third and fourth summand in (18) are zero by independence and $\mathbb{E}\, X^{(r)} = \mathbb{E}\, X_{I_r^{(n)}}^{(r)} = 0$. By our assumption we have $\|b^{(n)} - b^*\|_2^2 \to 0$ for $n \to \infty$, so we only have to care about the first summand:

$$\begin{aligned}
&\sum_{r=1}^K \left\| A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right\|_2^2 \\
&= \sum_{r=1}^K \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) + \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|_2^2 \\
&= \sum_{r=1}^K \Bigg(\left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2^2 + \left\| \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|_2^2 \\
&\qquad + 2\, \mathbb{E} \left\langle A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right), \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\rangle \Bigg).
\end{aligned} \tag{19}$$

By (15), independence, and $\|X\|_2 < \infty$ we obtain

$$\left\| \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|_2^2 \to 0, \quad n \to \infty,$$

for all $r = 1, \dots, K$. The third summand in (19) can be estimated by

$$\begin{aligned}
&\mathbb{E}\left\langle A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right), \left(A_r^{(n)} - A_r^*\right) X^{(r)}\right\rangle \\
&\quad\le \mathbb{E}\left[\left\|A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\| \left\|\left(A_r^{(n)} - A_r^*\right) X^{(r)}\right\|\right] \\
&\quad\le \left\|A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\|_2 \left\|\left(A_r^{(n)} - A_r^*\right) X^{(r)}\right\|_2 \\
&\quad\le \left\|A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\|_2 o(1) \\
&\quad\le \max\left\{1, \left\|A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\|_2^2\right\} o(1) \\
&\quad\le \left\|A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\|_2^2 o(1) + o(1),
\end{aligned}$$

where the non-trivial factor in the latter display is the same as the first summand in (19). For this we estimate, by conditioning on $I_r^{(n)}$, and using that conditioned on $I_r^{(n)}$ the random variates $(A_r^{(n)})^t A_r^{(n)}$ and $X^{(r)}_{I_r^{(n)}} - X^{(r)}$ are independent,

$$\begin{aligned}
&\left\|A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\|_2^2 \qquad (20)\\
&\quad= \mathbb{E}\left\langle X^{(r)}_{I_r^{(n)}} - X^{(r)}, (A_r^{(n)})^t A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\rangle \\
&\quad= \mathbb{E}\left[\mathbb{E}\left[\left\langle X^{(r)}_{I_r^{(n)}} - X^{(r)}, (A_r^{(n)})^t A_r^{(n)}\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\rangle \middle| I_r^{(n)}\right]\right] \\
&\quad= \mathbb{E}\left[\mathbb{E}\left[\left\langle X^{(r)}_{I_r^{(n)}} - X^{(r)}, \mathbb{E}\left[(A_r^{(n)})^t A_r^{(n)}\middle| I_r^{(n)}\right]\left(X^{(r)}_{I_r^{(n)}} - X^{(r)}\right)\right\rangle \middle| I_r^{(n)}\right]\right] \\
&\quad\le \mathbb{E}\left[\mathbb{E}\left[\left\|\mathbb{E}\left[(A_r^{(n)})^t A_r^{(n)}\middle| I_r^{(n)}\right]\right\|_{\mathrm{op}} \left\|X^{(r)}_{I_r^{(n)}} - X^{(r)}\right\|^2 \middle| I_r^{(n)}\right]\right] \\
&\quad= \mathbb{E}\left[\left\|\mathbb{E}\left[(A_r^{(n)})^t A_r^{(n)}\middle| I_r^{(n)}\right]\right\|_{\mathrm{op}} \mathbb{E}\left[\left\|X^{(r)}_{I_r^{(n)}} - X^{(r)}\right\|^2 \middle| I_r^{(n)}\right]\right] \\
&\quad= \mathbb{E}\left[\left\|\mathbb{E}\left[(A_r^{(n)})^t A_r^{(n)}\middle| I_r^{(n)}\right]\right\|_{\mathrm{op}} a_{I_r^{(n)}}\right],
\end{aligned}$$

where we define the sequence (a_i) by $a_i := \ell_2^2(X_i, X)$ and use that $X_i^{(r)}$ and $X^{(r)}$ are optimal couplings of X_i and X. Subsequently, by $o(1)$ we denote a generic deterministic sequence tending to zero as $n \to \infty$, that may be different at different occurrences. Putting the estimates together and denoting

$$A_r^{((n))} := \left\|\mathbb{E}\left[(A_r^{(n)})^t A_r^{(n)}\middle| I_r^{(n)}\right]\right\|_{\mathrm{op}} (1 + o(1)),$$

we obtain

$$\begin{aligned}
\ell_2^2(X_n, X) = a_n &\leq \sum_{r=1}^{K} \mathbb{E}\left[A_r^{((n))} a_{I_r^{(n)}}\right] + o(1) \\
&= \sum_{r=1}^{K} \mathbb{E}\left[\sum_{i=0}^{n} \mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))} a_i\right] + o(1) \\
&= \sum_{i=0}^{n} \left(\sum_{r=1}^{K} \mathbb{E}\left[\mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))}\right]\right) a_i + o(1).
\end{aligned}$$

With the abbreviations

$$p_n := \sum_{r=1}^{K} \mathbb{E}\left[\mathbf{1}_{\{I_r^{(n)}=n\}} A_r^{((n))}\right], \quad \eta := \limsup_{n\to\infty} \sum_{r=1}^{K} \mathbb{E}\left\|\mathbb{E}\left[(A_r^{(n)})^t A_r^{(n)} \middle| I_r^{(n)}\right]\right\|_{\mathrm{op}}$$

this implies the reduction inequality

$$\begin{aligned}
(1-p_n)a_n &\leq \sum_{r=1}^{K} \mathbb{E}\, A_r^{((n))} \sup_{0\leq i\leq n-1} a_i + o(1) \\
&= (\eta + o(1)) \sup_{0\leq i\leq n-1} a_i + o(1).
\end{aligned} \tag{21}$$

By (17) we have $p_n \to 0$, thus the assumption $\eta < 1$ in (16) implies that (a_n) is a bounded sequence. We define $a := \limsup a_n$. Now, there is a $\eta < \eta^+ < 1$ such that for all $\varepsilon > 0$ there is an $n_1 \in \mathbb{N}$ with $a_n \leq a + \varepsilon$ for all $n \geq n_1$ and such that the pre-factor in (21) satisfies $\sum \mathbb{E}\,[A_r^{((n))}] \leq \eta^+$ for $n \geq n_1$. Then from (21) we deduce

$$\begin{aligned}
a_n &\leq \frac{1}{1-p_n}\Bigg[\sum_{i=0}^{n_1-1}\left(\sum_{r=1}^{K} \mathbb{E}\left[\mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))}\right]\right) a_i \\
&\quad + \sum_{i=n_1}^{n-1}\left(\sum_{r=1}^{K} \mathbb{E}\left[\mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))}\right]\right)(a+\varepsilon) + o(1)\Bigg] \\
&\leq \frac{1}{1-p_n}\left(\eta^+(a+\varepsilon) + o(1)\right),
\end{aligned}$$

where (17) has been used. The $o(1)$ depends on ε. Since $\varepsilon > 0$ is arbitrary we conclude with $n \to \infty$ that $a = 0$. □

4 Applications of the multivariate framework

4.1 Quicksort

In this section as a first application of the multivariate transfer theorems of section 3 the analysis of the median-of-$(2t+1)$ version of Hoare's Quicksort algorithm is given. The problem is to sort an array of n

distinct numbers. The Quicksort algorithm chooses one of the elements (the so-called pivot) and compares all the other elements with the pivot. The elements smaller than the pivot are written in the array to the left of the pivot, the elements larger are written right to the pivot. Then Quicksort is applied recursively to the sub-arrays left and right of the pivot, for details see, e.g., Mahmoud [30]. For measuring the performance of Quicksort algorithms several parameters have been considered, the most important being the number of key comparisons and key exchanges.

We assume that the initial numbers' ranks are given as a random permutation, each permutation being equally likely and that the splitting into the sub-arrays is done while preserving randomness in and independence between the sub-arrays. For the number of key comparisons C_n a huge body of probabilistic results is available even for the median-of-$(2t+1)$ version of Quicksort, a version, where the pivot element is chosen as the median of a sub-sample of $2t+1$ elements taken uniformly at random from the numbers to be sorted. These results include in particular asymptotic expressions for the means and variances, as well as limit laws for the scaled quantities, and large deviation inequalities, see Hennequin [17, 18], Régnier [44], Rösler [45, 48], McDiarmid and Hayward [11], Bruhn [5], and for a detailed survey the book of Mahmoud [30]. For the number of key exchanges B_n executed while creating the sub-arrays (in standard implementations, see Sedgewick [50]), the mean and variance were for general $t \geq 0$ studied in Hennequin [18], Chern and Hwang [7] refined the analysis of the mean, and Hwang and Neininger [20] gave a limit law for the standard case $t = 0$.

Here we sketch a bivariate asymptotic analysis for the joint distribution $Y_n := (C_n, B_n)$ for general $t \geq 0$, as given in Neininger [37]. From a practical point of view linear combinations $C_n + wB_n$ with $w > 0$ are of interest. These model the cost of the algorithm assuming that a key exchange has w times the cost of a comparison. Here, naturally the covariance of C_n and B_n arises that drops automatically out in the bivariate approach below.

The number of key comparisons C_n for median-of-$(2t+1)$ Quicksort satisfies the recursion

$$C_n \stackrel{\mathcal{D}}{=} C^{(1)}_{I_n} + C^{(2)}_{n-1-I_n} + n - 1 + S^c_n, \quad n \geq n_0,$$

where I_n+1 is the order of the pivot element of the first partition stage. Furthermore, $(C^{(1)}_n)$, $(C^{(2)}_n)$, (I_n, S^c_n) are independent, $C^{(1)}_n \sim C^{(2)}_n \sim C_n$, and (S^c_n) is a sequence of uniformly bounded random variables which models the number of key comparisons for the selection of the median in the $2t+1$ sample. No further conditions on S^c_n are required. To initialize the algorithm some (random) bounded costs $C_0, \ldots, C_{n_0-1}$ have to be given with a $n_0 \geq 2t+1$ denoting the maximal size of the sub-arrays, which may be sorted by some other sorting procedure.

For the number of key exchanges we have

$$B_n \stackrel{\mathcal{D}}{=} B^{(1)}_{I_n} + B^{(2)}_{n-1-I_n} + T_n + S^b_n, \quad n \geq n_0, \tag{22}$$

with $(B^{(1)}_n)$, $(B^{(2)}_n)$, (I_n, T_n, S^b_n) being independent, $B^{(1)}_n \sim B^{(2)}_n \sim B_n$, T_n denoting the number of key exchanges during the partitioning step, and (S^b_n) a uniformly bounded sequence counting exchanges for the selection of the pivot element. We also need initial values $B_0, \ldots, B_{n_0-1}$. The T_n depend on the orders $I_n + 1$ of the pivot elements. We have

$$\mathbb{P}(T_n = j \mid I_n = k) = \frac{\binom{k}{j}\binom{n-1-k}{j}}{\binom{n-1}{k}}, \quad 0 \leq j \leq k \leq n-1,$$

see Sedgewick [50].

We emphasize that the relation (22) is only correct due to the assumption that the numbers are permuted uniformly at random and that the randomness and independence between sub-arrays is preserved.

The expectations $\mathbb{E}\,B_n$, $\mathbb{E}\,C_n$ have been studied in Sedgewick [50], Green [15], Hennequin [18], Bruhn [5], Rösler [48], Chern and Hwang [7] and others. What is needed subsequently is that

$$\begin{aligned}
\mathbb{E}\,B_n &= \frac{t+1}{2(2t+3)(H_{2t+2}-H_{t+1})} n\ln(n) + c_t n + o(n), && (23)\\
\mathbb{E}\,C_n &= \frac{1}{H_{2t+2}-H_{t+1}} n\ln(n) + c_t' n + o(n), && (24)
\end{aligned}$$

with constants $c_t, c_t' \in \mathbb{R}$ depending on the indicial conditions and (S_n^c, S_n^b). We abbreviate

$$\mu_c^{(t)} := \frac{1}{H_{2t+2}-H_{t+1}}, \quad \mu_b^{(t)} := \frac{t+1}{2(2t+3)(H_{2t+2}-H_{t+1})}.$$

The vector $Y_n = (C_n, B_n)$ satisfies the recursion

$$Y_n \stackrel{d}{=} Y^{(1)}_{I_1^{(n)}} + Y^{(2)}_{I_2^{(n)}} + b_n, \quad n \geq n_0,$$

with $(Y_n^{(1)}), (Y_n^{(2)}), (I^{(n)}, b_n)$ being independent, $Y_n^{(1)} \sim Y_n^{(2)} \sim Y_n$, $I^{(n)} = (I_n, n-1-I_n)$, $b_n = (n-1+S_n^c, T_n + S_n^b)$, and I_n, T_n as above. We scale using the matrix $\Sigma_n = \mathrm{diag}(n^2, n^2)$, where diag denotes the diagonal matrix of the given entries. With the expansions (23) and (24) we obtain for the scaled quantities $X_n := \Sigma_n^{-1/2}(Y_n - \mathbb{E}\,Y_n)$

$$X_n \stackrel{d}{=} A_1^{(n)} X^{(1)}_{I_1^{(n)}} + A_2^{(n)} X^{(2)}_{I_2^{(n)}} + b^{(n)}, \quad n \geq n_0, \tag{25}$$

with $A_1^{(n)} = \mathrm{diag}(I_n/n, I_n/n)$, $A_2^{(n)} = \mathrm{diag}((n-1-I_n)/n, (n-1-I_n)/n)$,

$$\begin{aligned}
b^{(n)} &= \Bigg(1 + \mu_c^{(t)} \left(\frac{I_1^{(n)}}{n} \ln \frac{I_1^{(n)}}{n} + \frac{I_2^{(n)}}{n} \ln \frac{I_2^{(n)}}{n} \right),\\
&\qquad \frac{T_n}{n} + \mu_b^{(t)} \left(\frac{I_1^{(n)}}{n} \ln \frac{I_1^{(n)}}{n} + \frac{I_2^{(n)}}{n} \ln \frac{I_2^{(n)}}{n} \right)\Bigg) + o(1),
\end{aligned}$$

and independence relations as in the original recursion. The $o(1)$ depends on randomness, but the convergence is uniform. For the L_2 convergence of the coefficients in (25) we use that for all $p > 0$

$$\frac{I_n}{n} \xrightarrow{L_p} V, \quad \frac{T_n}{n} \xrightarrow{L_2} V(1-V), \quad n \to \infty,$$

where V has the beta$(t+1, t+1)$ distribution. We have the L_2-convergences:

$$b^{(n)} \to b^*, \quad A_r^{(n)} \to A_r^*, \quad r = 1, 2,$$

where

$$A_1^* = \begin{pmatrix} V & 0 \\ 0 & V \end{pmatrix}, \quad A_2^* = \begin{pmatrix} 1-V & 0 \\ 0 & 1-V \end{pmatrix}, \tag{26}$$

$$b^* = \left(1 + \mu_c^{(t)} \mathcal{E}(V), V(1-V) + \mu_b^{(t)} \mathcal{E}(V)\right), \tag{27}$$

with $\mathcal{E}(V) := V \ln(V) + (1-V)\ln(1-V)$. From Theorem 3.1 we immediately obtain

Theorem 4.1 *The normalized vector of the number of key comparisons and key exchanges made by a median-of-*$(2t+1)$ *version of Quicksort satisfies*

$$\ell_2\left(\left(\frac{C_n - \mathbb{E}\, C_n}{n}, \frac{B_n - \mathbb{E}\, B_n}{n}\right), X\right) \to 0, \quad n \to \infty,$$

where $\mathcal{L}(X)$ *is the in* $\mathcal{M}_2^2(0)$ *unique distributional fixed-point of* T *defined in (6) with* (A_1^*, A_2^*, b^*) *given by (26), (27) and* V *there being* beta$(t+1, t+1)$ *distributed.*

Since convergence in ℓ_2 implies convergence of all mixed moments up to order two we obtain in particular

Corollary 4.2 *The asymptotic correlation and covariance of the number of key comparisons and key exchanges made by a median-of-*$(2t+1)$ *Quicksort version are given by*

$$\begin{aligned} \mathrm{Cor}(C_n, B_n) &= (1+o(1)) \frac{\mathbb{E}\,[b_1^* b_2^*]}{\sqrt{\mathbb{E}\,[(b_1^*)^2]\,\mathbb{E}\,[(b_2^*)^2]}}, \\ \mathrm{Cov}(C_n, B_n) &= (1+o(1)) \frac{2t+3}{t+1}\,\mathbb{E}\,[b_1^* b_2^*]\, n^2, \end{aligned}$$

where $b^* = (b_1^*, b_2^*)$ *is given in (27).*

The asymptotic correlation from Corollary 4.2 is for, e.g., (median-of-1) Quicksort

$$\frac{\sqrt{5}(39 - 4\pi^2)}{2\sqrt{(21 - 2\pi^2)(99 - 10\pi^2)}} \doteq -0.864.$$

Numerical values for these asymptotic correlations for $t = 0, \ldots, 10$ are listed in Neininger [37, Table 1].

Note that similar bivariate limit laws and correlation coefficients for the number of key comparisons and exchanges can be obtained for various variants and models of the closely related Quickselect algorithm, in particular for the models assumed in Mahmoud, Modarres, and Smythe [31] and Hwang and Tsai [21] and median-of-$(2t+1)$ versions of them.

4.2 Wiener index

The Wiener index of a connected graph is defined as the sum of the distances between all unordered pairs of vertices of the graph, where the distance between two vertices is the minimum number of edges connecting them in the graph. Here, we are discussing the probabilistic behavior of the Wiener index for random binary search trees. A binary search tree is a data structure built up from a set of distinct numbers. The first number becomes the root of the tree. Then the numbers are successively inserted recursively;

each number is compared with the root. If it is smaller than the root, it goes to the left subtree, otherwise to the right subtree. There this procedure is recursively iterated until we reach an empty subtree, where the number is inserted. A *random* binary search tree with n vertices is one built up from an equiprobable permutation of the numbers $1, \dots, n$. For reference see Knuth [27].

Interestingly, this parameter does not fit in our framework due to certain dependencies between the toll cost b_n and the parameter itself (details are given below). However, in a bivariate setup of our framework we surmount these dependencies.

Let I_n and $J_n = n - 1 - I_n$ denote the cardinalities of the left and right subtree of the root of a binary search tree containing n vertices. We denote by (W_{I_n}, P_{I_n}), (W'_{J_n}, P'_{J_n}) the pairs of the Wiener index and the internal path length in the left and right subtree of the root respectively. The internal path length of a rooted tree is defined as the sum of the distances between all vertices and the root. Thus by direct enumeration we obtain the recurrence

$$W_n = W_{I_n} + W'_{J_n} + b_n,$$

where

$$b_n = (P_{I_n} + P'_{J_n} + n - 1) + J_n P_{I_n} + I_n P'_{J_n} + 2 I_n J_n.$$

It is known that the cardinality of the left subtree I_n is uniformly distributed over $\{0, \dots, n-1\}$ and that, conditioned on this cardinality I_n, the left and right subtree have the distributions of random binary search trees of cardinalities I_n and J_n respectively and are (stochastically) independent of each other. This implies that with two sequences (W_n, P_n), (W'_n, P'_n) of pairs of Wiener indices and internal path lengths in random binary search trees such that (W_n, P_n), (W'_n, P'_n) and I_n are independent we obtain the distributional recurrence

$$W_n \stackrel{\mathcal{D}}{=} W_{I_n} + W'_{J_n} + b_n, \quad n \geq 1. \tag{28}$$

Note that we have $W_0 = 0$. The reason why we cannot apply our framework directly to the recurrence (28) is that conditioned on I_n the quantities b_n, W_{I_n}, W'_{J_n} are dependent, where independence is essential in recurrence (1). This dependence is caused by the dependence of the Wiener index and the internal path length in each subtree.

Theorem 4.3 *Let (W_n, P_n) denote the vector of the Wiener index and the internal path length of a random binary search tree with n vertices. Then we have*

$$\begin{aligned} \mathbb{E}\, W_n &= 2n^2 H_n - 6n^2 + 8nH_n - 10n + 6H_n, \qquad (29)\\ \mathrm{Var}(W_n) &= \frac{20 - 2\pi^2}{3} n^4 + o(n^4),\\ \left(\frac{W_n - \mathbb{E}\, W_n}{n^2}, \frac{P_n - \mathbb{E}\, P_n}{n}\right) &\xrightarrow{\mathcal{L}} (W, P), \end{aligned}$$

where $\mathcal{L}(W, P)$ is the in $\mathcal{M}_2^2(0)$ unique fixed-point of the map T given in (6) with

$$A_1^* = \begin{bmatrix} U^2 & U(1-U) \\ 0 & U \end{bmatrix},\ A_2^* = \begin{bmatrix} (1-U)^2 & U(1-U) \\ 0 & 1-U \end{bmatrix},\ b^* = \begin{pmatrix} 6U(1-U) + 2\,\mathcal{E}(U) \\ 1 + 2\,\mathcal{E}(U) \end{pmatrix},$$

and $\mathcal{E}(U)$ defined the line below (27). By H_n the n-th harmonic number $H_n = \sum_{i=1}^n 1/i$ is denoted and by $\xrightarrow{\mathcal{L}}$ convergence in distribution.

Proof: *(Sketch)* For (29) see Hwang and Neininger [20]. Additionally to recurrence (28) we also have

$$P_n \stackrel{\mathcal{D}}{=} P_{I_n} + P'_{J_n} + n - 1, \quad n \geq 1,$$

and obtain as well a distributional recurrence for the bivariate quantities (W_n, P_n):

$$\begin{pmatrix} W_n \\ P_n \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{bmatrix} 1 & n - I_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W_{I_n} \\ P_{I_n} \end{pmatrix} + \begin{bmatrix} 1 & n - J_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} W'_{J_n} \\ P'_{J_n} \end{pmatrix} + \begin{pmatrix} 2I_nJ_n + n - 1 \\ n - 1 \end{pmatrix}.$$

The rescaled quantities $X_0 := 0$ and

$$X_n := \left(\frac{W_n - \mathbb{E}\, W_n}{n^2}, \frac{P_n - \mathbb{E}\, P_n}{n} \right)^t, \quad n \geq 1,$$

and the analogously defined X'_n satisfy the recurrence

$$X_n \stackrel{\mathcal{D}}{=} A_1^{(n)} X_{I_n} + A_2^{(n)} X'_{J_n} + b^{(n)}, \quad n \geq 1,$$

where

$$A_1^{(n)} = \begin{bmatrix} (I_n/n)^2 & I_n(n - I_n)/n^2 \\ 0 & I_n/n \end{bmatrix}, \quad A_2^{(n)} = \begin{bmatrix} (J_n/n)^2 & J_n(n - J_n)/n^2 \\ 0 & J_n/n \end{bmatrix},$$

and $b^{(n)} = (b_1^{(n)}, b_2^{(n)})$ with

$$\begin{aligned} \begin{pmatrix} b_1^{(n)} \\ b_2^{(n)} \end{pmatrix} \quad &= \quad \begin{bmatrix} 1/n^2 & 0 \\ 0 & 1/n \end{bmatrix} \left(\begin{bmatrix} 1 & n - I_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \alpha_{I_n} \\ \gamma_{I_n} \end{pmatrix} + \begin{bmatrix} 1 & n - J_n \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \alpha_{J_n} \\ \gamma_{J_n} \end{pmatrix} \right. \\ &\qquad \left. - \begin{pmatrix} \alpha_n \\ \gamma_n \end{pmatrix} + \begin{pmatrix} 2I_nJ_n + n - 1 \\ n - 1 \end{pmatrix} \right), \end{aligned}$$

where $(\alpha_n, \gamma_n) = \mathbb{E}\,(W_n, P_n)$. We plug in the expansions

$$\begin{aligned} \alpha_n &= 2n^2 \ln(n) + (2\gamma - 6)n^2 + o(n^2), \\ \gamma_n &= 2n \ln(n) + (2\gamma - 4)n + o(n), \end{aligned}$$

with Euler's constant γ. After cancellation we obtain with the convention $x \ln x := 0$ for $x = 0$,

$$\begin{aligned} b_1^{(n)} &= \frac{1}{n^2} \left(2I_n^2 \ln\left(\frac{I_n}{n}\right) + 2J_n^2 \ln\left(\frac{J_n}{n}\right) + 2I_nJ_n \ln\left(\frac{I_n}{n}\right) \right. \\ &\qquad \left. + 2I_nJ_n \ln\left(\frac{J_n}{n}\right) + 6I_nJ_n \right) + o(1), \\ b_2^{(n)} &= \frac{1}{n} \left(2I_n \ln\left(\frac{I_n}{n}\right) + 2J_n \ln\left(\frac{J_n}{n}\right) + n \right) + o(1), \end{aligned}$$

where the $o(1)$s are random but the convergences hold uniformly. We model all quantities on a joint probability space such that $I_n/n \to U$ for a uniform $[0,1]$ distributed random variate, where the convergence holds almost surely and thus in L_2. Then, by dominated convergence, we obtain the following L_2-convergences:

$$A_1^{(n)} \to A_1^*, \quad A_2^{(n)} \to A_2^*, \quad b^{(n)} \to b^*,$$

with (A_1^*, A_2^*, b^*) given in the theorem.

Solving the characteristic equation for $(A_1^*)^t A_1^*$ we obtain that the eigenvalue $\lambda(U)$ of $(A_1^*)^t A_1^*$ being larger in absolute value is given by

$$\lambda(U) = U^2 \left(\frac{1 + U^2 + (1-U)^2}{2} + \sqrt{\frac{(1 + U^2 + (1-U)^2)^2}{4} - U^2} \right). \tag{30}$$

This implies, since $(A_1^*)^t A_1^*$ and $(A_2^*)^t A_2^*$ are identically distributed, that

$$\begin{aligned} \mathbb{E} \left\| (A_1^*)^t A_1^* \right\|_{\mathrm{op}} + \mathbb{E} \left\| (A_2^*)^t A_2^* \right\|_{\mathrm{op}} &= 2\,\mathbb{E}\,\lambda(U) \\ &= \frac{3}{10} + \frac{29}{60}\sqrt{2} + \frac{1}{4} \ln \left(\sqrt{2} - 1 \right) \\ &< 1. \end{aligned} \tag{31}$$

Thus condition (12) is fulfilled and Theorem 3.1 can be applied and covariances and correlations can be extracted from the bivariate fixed-point equation, see Neininger [38] for details. □

For another application of this technique in the analysis of quantities related to phylogenetic tree balance see Blum, François and Janson [2].

4.3 Branching recurrences

Branching recurrences are sequences $(Y_n)_{n \geq 0}$, where Y_0 is a random variable in $\mathbb{R}^d$ and for random $d \times d$ matrices $A_1, \ldots, A_K$ and a random translation b in $\mathbb{R}^d$ we have, for $n \geq 1$,

$$Y_n \stackrel{\mathcal{D}}{=} \sum_{r=1}^{K} A_r Y_{n-1}^{(r)} + b, \quad n \geq 1, \tag{32}$$

where $(A_1, \ldots, A_K, b), Y_{n-1}^{(1)}, \ldots, Y_{n-1}^{(K)}$ are independent and $Y_{n-1}^{(r)} \sim Y_{n-1}$ for $r = 1, \ldots, K$. Hence, these sequences are covered by our general setting (3) choosing $I_r^{(n)} = n - 1$ for $r = 1, \ldots, K$ and with $(A_1(n), \ldots, A_K(n), b_n)$ being independent of n.

The special case $K = 1$ in (32) of an iteration of a random affine map has been studied intensively in the literature with many respects, see, e.g., Kesten [23], Brand [4], Bougerol and Picard [3], Burton and Rösler [6], Goldie and Maller [14], Diaconis and Freedman [10], and the references in these articles. The case $K \geq 2$ leads to branching type recursive sequences. In the one dimensional case without the immigration term b and the A_r being independent and nonnegative this recursion was studied by Mandelbrot [35] for the analysis of a model of turbulence of Yaglom and Kolmogorov. To this case further contributions on nontrivial fixed points of a corresponding operator, the existence of moments of these fixed points and convergence of (Y_n) to the fixed points were made in Kahane and Peyrière [22] and Guivarc'h [16]. The case $b = 0$, $A^{(r)} \geq 0$ for $r = 1, \dots K$ with dependencies was considered in Holley and Liggett [19] and Durrett and Liggett [12] for the purpose of analyzing a problem in infinite particle systems. The case $b = 0$ with deterministic coefficients (and $K = \infty$) was discussed in Rösler [47]. See this paper also for references and an overview on the one-dimensional fixed point equations without immigration term. The general form of the recursion (32) in dimension one was treated in Cramer and Rüschendorf [8]. A two-dimensional version of (32) with $K = 2$ and $b = 0$ has been considered in Cramer and Rüschendorf [9]. For an application in the context of randomized game tree evaluation leading to a two-dimensional version of (32) with $K = 4$ and $b = 0$ see Ali Khan and Neininger [1].

Most of the investigations mentioned considered problems of convergence of the sequence (Y_n) itself. Our general theorems from section 3 apply to cases where the Y_n require a proper scaling in order to allow distributional convergence. We assume $0 < \mathrm{Var}(Y_{n,i}) < \infty$ for all $1 \leq i \leq d$ and $n \geq 1$, where $Y_{n,i}$ denote the coordinates of Y_n. This condition is easy to check and not satisfied only in very special cases. For the normalization of the process (Y_n) define

$$M_n := \mathbb{E}\, Y_n, \quad \Sigma_n := \mathrm{diag}(\mathrm{Var}(Y_{n,1}), \dots, \mathrm{Var}(Y_{n,d})), \quad n \geq 1. \tag{33}$$

Then we obtain X_n by rescaling Y_n as defined in (2), where we have

$$A_r^{(n)} := \Sigma_n^{-1/2} A_r \Sigma_{n-1}^{1/2}, \quad b^{(n)} = \Sigma_n^{-1/2}\Big(b - M_n + \sum_{r=1}^K A_r M_{n-1}\Big), \quad n \geq 2.$$

Thus, we normalize X_n by the components not changing its covariance structure:

$$X_{n,i} = \frac{Y_{n,i} - \mathbb{E}\, Y_{n,i}}{\mathrm{Var}(Y_{n,i})^{1/2}}, \quad 1 \leq i \leq d,$$

$$(A_r^{(n)})_{ij} = \left(\frac{\mathrm{Var}(Y_{n-1,j})}{\mathrm{Var}(Y_{n,i})}\right)^{1/2} (A_r)_{ij}, \quad 1 \leq i, j \leq d.$$

The existence of limits $(A_1^*, \dots, A_K^*, b^*)$ for $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ in this situation reduces to the existence of the following deterministic limits (here $\lim a_n = \infty$ denotes that a sequence (a_n) is definitely divergent):

$$\lim_{n\to\infty} \mathrm{Var}(Y_{n,i}) =: \vartheta_i \in (0, \infty], \tag{34}$$

$$\lim_{n\to\infty} \left(\frac{\mathrm{Var}(Y_{n-1,j})}{\mathrm{Var}(Y_{n,i})}\right)^{1/2} =: c_{ij}, \tag{35}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}\, Y_{n,i}}{(\mathrm{Var}(Y_{n,i}))^{1/2}} =: \gamma_i. \tag{36}$$

Then we have the following relations as $n \to \infty$: With

$$A_\Sigma^* := \sum_{r=1}^K A_r^*, \quad \Sigma_\infty := \lim_{n\to\infty} \Sigma_n = \mathrm{diag}(\vartheta_1, \dots, \vartheta_d), \quad \gamma = (\gamma_1, \dots, \gamma_d),$$

where Σ_∞ may have infinite entries, we have

$$A_r^{(n)} \xrightarrow{L_2} A_r^*, \quad b^{(n)} \xrightarrow{L_2} b^*$$

with

$$(A_r^*)_{ij} = c_{ij}(A_r)_{ij}, \quad b^* = \Sigma_\infty^{-1/2}(b - \mathbb{E}\, b) + (A_\Sigma^* - \mathbb{E}\, A_\Sigma^*)\gamma, \tag{37}$$

for $1 \le i, j \le d,\ 1 \le r \le K$, where we use the convention $x/\infty = 0$ for $x \in \mathbb{R}$. For the second convergence note that

$$\begin{aligned} b^{(n)} &= \Sigma_n^{-1/2}\left(b - M_n + \sum_{r=1}^K A_r M_{n-1}\right) \\ &= \Sigma_n^{-1/2}\left(b - \Big(\sum_{r=1}^K \mathbb{E}\, A_r M_{n-1} + \mathbb{E}\, b\Big) + \sum_{r=1}^K A_r M_{n-1}\right) \\ &= \Sigma_n^{-1/2}\left(b - \mathbb{E}\, b + \Big(\sum_{r=1}^K (A_r - \mathbb{E}\, A_r)\Big) M_{n-1}\right) \\ &= \Sigma_n^{-1/2}(b - \mathbb{E}\, b) + \left[\Sigma_n^{-1/2}\Big(\sum_{r=1}^K (A_r - \mathbb{E}\, A_r)\Big)\Sigma_{n-1}^{1/2}\right]\left[\Sigma_{n-1}^{-1/2} M_{n-1}\right] \\ &\xrightarrow{L_2} \Sigma_\infty^{-1/2}(b - \mathbb{E}\, b) + (A_\Sigma^* - \mathbb{E}\, A_\Sigma^*)\gamma. \end{aligned}$$

We obtain the following asymptotic behavior for (X_n):

Theorem 4.4 *Let (Y_n) be the sequence defined by (32) where the initial distribution, random matrices and the immigration term are square integrable. Assume that conditions (34)–(36) are satisfied and that*

$$\left\| \sum_{r=1}^K \mathbb{E}\, \left[(A_r^*)^t A_r^*\right] \right\|_{\mathrm{op}} < 1. \tag{38}$$

Then the scaled version $X_n = \Sigma_n^{-1/2}(Y_n - M_n)$ with M_n, Σ_n defined in (33) converges to the in $\mathcal{M}_2^d(0)$ unique fixed point $\mathcal{L}(X)$ of the limiting operator T given by (6) with $(A_1^*, \dots, A_K^*, b^*)$ given in (37),

$$\ell_2(X_n, X) \to 0.$$

Note that under the stronger condition

$$\sum_{r=1}^K \mathbb{E}\, \left\|(A_r^*)^t A_r^*\right\|_{\mathrm{op}} < 1$$

the assertion of this theorem is covered by Theorem 3.1. In the special case of Theorem 4.4 the weaker condition (38) coincides with condition (13) of Problem 3.2. Here, this weaker condition can be obtained applying similar techniques as in the proof of Theorem 3.3. The difference that allows to apply sharper estimates here is that the underlying recurrence is somehow more similar to the limiting map T compared to the situation in Theorem 3.3, for details see Neininger [36, Theorem 5.1.3].

Covariance structure

For the application of Theorem 4.4 it is necessary to check conditions (34)–(36). The asymptotic behavior of the mean vector $M_n = \mathbb{E}\,Y_n$ is given by

$$M_n = \mathbb{E}\,Y_n = \mathbb{E}\,A_\Sigma\,\mathbb{E}\,Y_{n-1} + \mathbb{E}\,b = (\mathbb{E}\,A_\Sigma)^n\,\mathbb{E}\,Y_0 + \sum_{k=0}^{n-1}(\mathbb{E}\,A_\Sigma)^k\,\mathbb{E}\,b, \tag{39}$$

where we denote $A_\Sigma := \sum_{r=1}^K A_r$. For the mean, therefore, the computational complexity reduces to the derivation of the powers of the matrix $\mathbb{E}\,A_\Sigma$. For the asymptotic of the variances Var$(Y_{n,1}),\dots,$Var$(Y_{n,d})$ we have to investigate the whole covariance matrix Cov(Y_n). We use the following notation: For an $n\times m$ matrix $A = (a_{ij})$ denote by $(A)_s := (a_{s1},\dots,a_{sm})^t$ the sth row vector of A for $1 \le s \le n$. If X is an random variable in $\mathbb{R}^d$, Cov(X) denotes the matrix $(\mathrm{Cov}(X_i,X_j))_{i,j=1}^d$. For two random variables X,Y in $\mathbb{R}^d$, $\mathbb{R}^{d'}$ respectively, Cov(X,Y) denotes the $d\times d'$ matrix $(\mathrm{Cov}(X_i,Y_j))$, $1\le i\le d, 1\le j\le d'$. A direct computation yields:

Lemma 4.5 *Let (Y_n) satisfy the recursion (32) and $A_\Sigma = \sum_{r=1}^K A_r$. Then for $1 \le s,t \le d$*

$$\begin{aligned}(\mathrm{Cov}(Y_n))_{st} \;&=\; \sum_{r=1}^K \mathbb{E}\,\langle \mathrm{Cov}(Y_{n-1})(A_r)_s, (A_r)_t\rangle \\ &\quad + \langle \mathrm{Cov}((A_\Sigma)_s,(A_\Sigma)_t)\,\mathbb{E}\,Y_{n-1},\,\mathbb{E}\,Y_{n-1}\rangle \\ &\quad + \langle \mathrm{Cov}((A_\Sigma)_s, b_t) + \mathrm{Cov}((A_\Sigma)_t, b_s),\,\mathbb{E}\,Y_{n-1}\rangle \\ &\quad +(\mathrm{Cov}(b))_{st}\end{aligned} \tag{40}$$

For a given $d\times d$ matrix $A=(a_{ij})$ denote by $A^V \in \mathbb{R}^{d^2}$ the vector

$$A^V_{(i-1)d+j} := a_{ij}, \quad 1\le i,j\le d.$$

This means that we write the matrix A row by row into the vector A^V. We also use the notation $(\mathrm{Cov}(\,\cdot\,))^V = \mathrm{Cov}^V(\,\cdot\,)$. Define the $d^2\times d^2$ matrices P,Q and the $d^2\times d$ matrix R by

$$\begin{aligned}P_{(s-1)d+t,(i-1)d+j} \;&:=\; \sum_{r=1}^K \mathbb{E}\,[(A_r)_{si}(A_r)_{tj}]\,, \\ Q_{(s-1)d+t,(i-1)d+j} \;&:=\; \sum_{q,r=1}^K \mathrm{Cov}\left((A_q)_{si},(A_r)_{tj}\right), \\ R_{(s-1)d+t,j} \;&:=\; \sum_{r=1}^K \mathrm{Cov}\left((A_r)_{si}, b_t\right) + \mathrm{Cov}\left((A_r)_{tj}, b_s\right),\end{aligned}$$

for $1\le i,j,s,t\le d$. Furthermore denote by L_n the $d\times d$ matrix

$$L_n := M_n M_n^t.$$

By (39), M_n can be expressed in terms of $A_1, \dots, A_K, b, Y_0$. In this notation the recurrence (40) reads

$$\mathrm{Cov}^V(Y_n) = P\,\mathrm{Cov}^V(Y_{n-1}) + QL^V_{n-1} + RM_{n-1} + \mathrm{Cov}^V(b).$$

Iteration of this formula leads to an explicit representation of the covariance matrix $\mathrm{Cov}(Y_n)$, see Neininger [36, Theorem 5.2.2]

Theorem 4.6 *The covariance matrix Cov (Y_n) of Y_n satisfying recursion (32) has the representation*

$$\mathrm{Cov}^V(Y_n) = P^n \mathrm{Cov}^V(Y_0) + \sum_{k=1}^{n} P^{k-1}\Big(QL^V_{n-k} + RM_{n-k} + \mathrm{Cov}^V(b)\Big). \tag{41}$$

Hence, the computational complexity to determine the asymptotic of $\mathbb{E}\, Y_n$ and Var$(Y_{n,1}), \dots,$Var$(Y_{n,d})$ equals the complexity to determine the powers of the matrices $\mathbb{E}\, A_\Sigma$ and P given by

$$(\mathbb{E}\, A_\Sigma)_{ij} = \sum_{r=1}^{K} \mathbb{E}\left[(A_r)_{ij}\right], \tag{42}$$

$$P_{(s-1)d+t,(i-1)d+j} = \sum_{r=1}^{K} \mathbb{E}\left[(A_r)_{si}(A_r)_{tj}\right] \tag{43}$$

for $1 \le i, j, s, t \le d$. A concrete example in the two-dimensional case for the derivation of these asymptotics leading to the verification of conditions (34)–(36) was elaborated in Cramer and Rüschendorf [9].

Lyapunov exponents

The classical approach to study recursion (32) in the affine case $K = 1$ is based on properties of Lyapunov exponents. For $K = 1$ the process $(Y_n)_{n\ge 0}$ can be defined pointwise by an initial random variate Y_0 and

$$Y_n := A'_n Y_{n-1} + b'_n, \quad n \ge 1, \tag{44}$$

where (A'_n, b'_n) is an independent identically distributed sequence that, for the case $K = 1$ in recurrence (32), corresponds to distributional copies of (A_1, b) there. Iterating (44), Y_n has the representation

$$Y_n = A'_n \cdots A'_1 Y_0 + b'_n + \sum_{j=2}^{n} (A'_n \cdots A'_j) b'_{j-1}.$$

The distributional asymptotics of Y_n are usually analyzed introducing a *change of time* (see, e.g., Vervaat [52]): Let $Y^\star_0 := 0$ and

$$Y^\star_n := b'_1 + \sum_{j=1}^{n-1} (A'_1 \cdots A'_j) b'_{j+1}. \tag{45}$$

Then we obtain the distributional relation

$$Y_n \stackrel{\mathcal{D}}{=} Y^\star_n + A'_1 \cdots A'_n Y_0. \tag{46}$$

Assuming appropriate assumptions involving the Lyapunov exponent for A_1', the summand $A_1' \cdots A_n' Y_0$ becomes asymptotically small and $Y_n^\star$ converges almost surely to

$$Y := b_1' + \sum_{j=1}^{\infty} (A_1' \cdots A_j') b_{j+1}'.$$

As a consequence, $Y_n \to Y$ in distribution can be deduced.

The *(top) Lyapunov exponent* of a random $n \times n$ matrix $A (= A_1')$ satisfying the condition $\mathbb{E} \ln^+ \|A\| < \infty$ is defined by

$$\gamma(A) := \inf_{n \in \mathbb{N}} \frac{1}{n} \mathbb{E} \ln \|A_1' \cdots A_n'\|,$$

where (A_i') are independent with $A_i' \sim A$ for $i \geq 1$. Note that the definition is independent of the norm being used. The analysis of $Y_n^\star$ is based on the fact that

$$\gamma(A) = \lim_{n \to \infty} \frac{1}{n} \ln \|A_1' \cdots A_n'\| \quad \text{almost surely,} \tag{47}$$

which was first proved in Furstenberg and Kesten [13] and is a consequence of the sub-additive ergodic theorem of Kingman [24]. If $\mathbb{E} \ln \|b\| < \infty$ and $\gamma(A_1') < 0$ then convergence of Y_n to Y was shown in Burton and Rösler [6].

For a scaled version of such a result consider

$$X_n := \Sigma_n^{-1/2} Y_n, \quad n \geq 1,$$

with Σ_n given by (33). Define

$$\beta := \frac{1}{2} \min_{1 \leq i \leq d} \liminf_{n \to \infty} \frac{1}{n} \ln \Big(\mathrm{Var}(Y_{n,i}) \Big). \tag{48}$$

Corresponding to Theorem 4.4 we derive, in the case $K = 1$, a limit law for (X_n) based on Lyapunov exponents, see Neininger [36, Theorem 5.3.1].

Theorem 4.7 *Let (Y_n) be given as in (44) with $\mathbb{E} \ln^+ \|b\| < \infty$, $\|Y_0\| < \infty$ almost surely, and $\gamma(A_1') < \beta$, where $\beta > 0$ is given by (48). Then*

$$Z := \lim_{n \to \infty} \Sigma_n^{-1/2} \Big(b_1' + \sum_{j=1}^{n-1} (A_1' \cdots A_j') b_{j+1}' \Big)$$

exists almost surely and, as $n \to \infty$,

$$X_n = \Sigma_n^{-1/2} Y_n \xrightarrow{\mathcal{D}} Z.$$

Proof: By (45) and (46) we have

$$X_n \stackrel{\mathcal{D}}{=} \Sigma_n^{-1/2} Y_n^\star + \Sigma_n^{-1/2} A_1' \cdots A_n' Y_0.$$

Since $\beta > 0$ and $\gamma(A_1') < \beta$ there exists a $0 < \xi < (\beta - \gamma(A_1'))/4$ with $\beta_- := \beta - \xi > 0$ and $\alpha_{++} := \gamma(A_1') + 2\xi \neq 0$. Define $\alpha_+ := \gamma(A_1') + \xi$. By (47) there exists almost surely a random $n_0 \geq 1$ with

$$\left\|A_1' \cdots A_n'\right\|_{\mathrm{op}} \leq \exp(n\alpha_+), \quad n \geq n_0.$$

Therefore, almost surely a random $c_1 > 0$ exists with

$$\left\|A_1' \cdots A_n'\right\|_{\mathrm{op}} \leq c_1 \exp(n\alpha_+), \quad n \geq 1. \tag{49}$$

Analogously, using (48) there is a $c_2 > 0$ with

$$\left\|\Sigma_n^{-1/2}\right\|_{\mathrm{op}} \leq c_2 \exp(-n\beta_-), \quad n \geq 1. \tag{50}$$

Furthermore there is a random $c_3 > 0$ so that almost surely

$$\left\|b_n'\right\| \leq c_3 \exp(n\xi/4), \quad n \geq 1. \tag{51}$$

This can be seen by a standard argument:

$$\begin{aligned} \sum_{n\geq 1} \mathbb{P}\Big(\left\|b_n'\right\| > \exp(n\xi/4)\Big) &= \sum_{n\geq 1} \mathbb{P}\Big((4/\xi)\ln\left\|b_n'\right\| > n\Big) \\ &\leq \frac{4}{\xi}\,\mathbb{E}\,\ln^+\left\|b\right\| < \infty. \end{aligned}$$

Then by the Borel-Cantelli Lemma

$$\mathbb{P}\left(\limsup_{n\to\infty}\left\{\left\|b_n'\right\| > \exp(n\xi/4)\right\}\right) = 0.$$

This means that with probability one $\left\|b_n'\right\| > \exp(n\xi/4)$ occurs only finitely many times. This implies (51).

Now we complete the proof by showing that, as $n \to \infty$, we have almost surely

$$\Sigma_n^{-1/2} A_1' \cdots A_n' Y_0 \longrightarrow 0, \tag{52}$$

$$\Sigma_n^{-1/2} Y_n^\star \longrightarrow Z. \tag{53}$$

For the proof of (52) note that almost surely

$$\begin{aligned} \left\|\Sigma_n^{-1/2} A_1' \cdots A_n' Y_0\right\| &\leq \left\|\Sigma_n^{-1/2}\right\|_{\mathrm{op}} \left\|A_1' \cdots A_n'\right\|_{\mathrm{op}} \left\|Y_0\right\| \\ &\leq c_1 c_2 \exp(-n\beta_-) \exp(n\alpha_+) \left\|Y_0\right\| \\ &\leq c_1 c_2 \left\|Y_0\right\| \exp(-2\xi n) \to 0, \quad \text{as } n \to \infty, \end{aligned} \tag{54}$$

by (49), (50), and $\left\|Y_0\right\| < \infty$ almost surely.

For the proof of (53) we have

$$\Sigma_n^{-1/2} Y_n^\star = \Sigma_n^{-1/2}\left(b_1' + \sum_{j=1}^{n-1} A_1' \cdots A_j' b_{j+1}'\right).$$

For the convergence of $\Sigma_n^{-1/2} Y_n^\star$ we show that $(\Sigma_n^{-1/2} Y_n^\star)$ is almost surely a Cauchy sequence. We prove that almost surely

$$\lim_{n_0\to\infty} \sup_{n_0 \le m \le n} \left\|\Sigma_n^{-1/2} Y_n^\star - \Sigma_m^{-1/2} Y_m^\star\right\| = 0.$$

For $n_0 \le m < n$ we have

$$\begin{aligned}
&\left\|\Sigma_n^{-1/2} Y_n^\star - \Sigma_m^{-1/2} Y_m^\star\right\| \\
&= \left\|\Sigma_n^{-1/2} \sum_{j=m}^{n-1} A_1' \cdots A_j' b_{j+1}' + (\Sigma_n^{-1/2} - \Sigma_m^{-1/2}) \sum_{j=1}^{m-1} A_1' \cdots A_j' b_{j+1}'\right\| \\
&\le \left\|\Sigma_n^{-1/2}\right\|_{\mathrm{op}} \sum_{j=m}^{n-1} \left\|A_1' \cdots A_j'\right\|_{\mathrm{op}} \left\|b_{j+1}'\right\| \\
&\quad + \left\|\Sigma_n^{-1/2} - \Sigma_m^{-1/2}\right\|_{\mathrm{op}} \sum_{j=1}^{m-1} \left\|A_1' \cdots A_j'\right\|_{\mathrm{op}} \left\|b_{j+1}'\right\|.
\end{aligned} \tag{55}$$

The first summand in (55), denoting $c := c_1 c_2 c_3$ and using (49) and (51), is estimated almost surely by

$$\begin{aligned}
&\left\|\Sigma_n^{-1/2}\right\|_{\mathrm{op}} \sum_{j=m}^{n-1} \left\|A_1' \cdots A_j'\right\|_{\mathrm{op}} \left\|b_{j+1}'\right\| \\
&\le c\exp(-n\beta_-) \sum_{j=m}^{n-1} \exp(j\alpha_{++}) \\
&= c\exp(-n\beta_-) \frac{\exp(\alpha_{++})^m - \exp(\alpha_{++})^n}{1-\exp(\alpha_{++})} \\
&= \frac{c}{1-\exp(\alpha_{++})}\Big[\exp(m\alpha_{++} - n\beta_-) - \exp(n(\alpha_{++} - \beta_-))\Big] \\
&\to 0, \quad \text{as } n_0 \to \infty,
\end{aligned}$$

since $\exp(m\alpha_{++} - n\beta_-) \le \exp(-n_0\xi)$ and $\exp(n(\alpha_{++} - \beta_-)) \le \exp(-n_0\xi)$. For the second summand of (55) note that

$$\left\|\Sigma_n^{-1/2} - \Sigma_m^{-1/2}\right\|_{\mathrm{op}} \le c_2 \exp(-m\beta_-),$$

since $m < n$ and $\beta_- > 0$. This implies almost surely

$$\begin{aligned}
&\big\|\Sigma_n^{-1/2} - \Sigma_m^{-1/2}\big\|_{\mathrm{op}} \sum_{j=1}^{m-1} \big\|A_1' \cdots A_j'\big\|_{\mathrm{op}} \big\|b_{j+1}'\big\| \\
\leq\quad & c \exp(-m\beta_-) \sum_{j=1}^{m-1} \exp(j\alpha_{++}) \\
=\quad & c \exp(-m\beta_-) \frac{1 - \exp(\alpha_{++})^m}{1 - \exp(\alpha_{++})} \\
=\quad & \frac{c}{1 - \exp(\alpha_{++})} \Big[\exp(-m\beta_-) - \exp(m(\alpha_{++} - \beta_-)) \Big] \\
\to\quad & 0, \quad \text{as } n_0 \to \infty,
\end{aligned}$$

since $\exp(-m\beta_-) \leq \exp(-n_0\xi)$ and $\exp(n(\alpha_{++} - \beta_-)) \leq \exp(-n_0\xi)$. $\square$

Even in the un-scaled case with $K = 1$ there is not much known about the connection of the conditions arising in the L_2-case formulated in expectations of norms of certain matrices as in (38) with conditions on Lyapunov exponents. For a discussion of this subject see Burton and Rösler [6].

Problem 4.8 *Does* $\|\mathbb{E}\,[(A_1')^t A_1']\|_{\mathrm{op}} < 1$ *imply* $\gamma(A_1') < 0$*? Find a generalization* $\gamma(A_1, \ldots, A_K)$ *for the Lyapunov exponent* $\gamma(A_1)$ *such that an analogue of Theorem 4.4 can be given with (38) replaced by a condition on* $\gamma(A_1^*, \ldots, A_K^*)$.

References

[1] Ali Khan, T. and Neininger, R. (2004) Probabilistic analysis for randomized game tree evaluation. em Mathematics and computer science. III, 163–174, Trends Math., *Birkhäuser, Basel.*

[2] Blum, M. G. B., François, O. and Janson, S. (2005+) The mean, variance and limiting distribution of two statistics sensitive to phylogenetic tree balance. Preprint
Available at `http://www.math.uu.se/ svante/papers/sj180.pdf`

[3] Bougerol, P. and Picard, N. (1992) Strict stationarity of generalized autoregressive processes. *Ann. Probab. 20*, 1714-1730.

[4] Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. in Appl. Probab. 18*, 211-220.

[5] Bruhn, V. (1996). *Eine Methode zur asymptotischen Behandlung einer Klasse von Rekursionsgleichungen mit einer Anwendung in der stochastischen Analyse des Quicksort-Algorithmus*. Ph.D. Thesis, Universität Kiel.

[6] Burton, R. and Rösler, U. (1995). An L_2 convergence theorem for random affine mappings. *J. Appl. Probab. 32*, 183–192.

[7] Chern, H.-H. and Hwang, H.-K. (2001). Transitional behaviors of the average cost of Quicksort with median-of-$(2t+1)$. *Algorithmica 29*, 44–69.

[8] Cramer, M. and Rüschendorf, L. (1996) Convergence of a branching type recursion. *Ann. Inst. H. Poincaré Probab. Statist. 32*, 725-741.

[9] Cramer, M. and Rüschendorf, L. (2001) Convergence of two-dimensional branching recursions. *J. Comput. Appl. Math. 130*, 53-73.

[10] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev. 41*, 45-76.

[11] McDiarmid, C. J. H. and Hayward, R. B. (1996). Large deviations for Quicksort. *J. Algorithms 21*, 476–507.

[12] Durrett, R. and Liggett, T. M. (1983). Fixed points of the smoothing transformation. *Z. Wahrscheinlichkeitstheor. Verw. Geb. 64*, 275-301.

[13] Furstenberg, H. and Kesten, H. (1960). Products of random matrices. *Ann. Math. Stat. 31*, 457-469.

[14] Goldie, C. M. and Maller, R. A. (2000). Stability of perpetuities. *Ann. Probab. 28*, 1195-1218.

[15] Green, D. H. (1983). *Labelled Formal Languages and their Uses*. Ph.D. Thesis, Stanford University.

[16] Guivarc'h, Y. (1996). Sur une extension de la notion de loi semi-stable. *Ann. Inst. H. Poincaré Probab. Statist. 26*, 261-286.

[17] Hennequin, P. (1989). Combinatorial analysis of quicksort algorithm. *RAIRO Inform. Théor. Appl. 23*, 317–333.

[18] Hennequin, P. (1991). *Analyse en moyenne d'algorithme, tri rapide et arbres de recherche*. Ph.D. Thesis, L'École Polytechnique.

[19] Holley, R. and Liggett, T. M. (1981). Generalized potlatch and smoothing processes. *Z. Wahrscheinlichkeitstheor. Verw. Geb. 55*, 165-195.

[20] Hwang, H.-K. and Neininger, R. (2002). Phase change of limit laws in the quicksort recurrence under varying toll functions. *SIAM J. Comput. 31*, 1687-1722.

[21] Hwang, H.-K. and Tsai, T.-H. (2002). Quickselect and Dickman function. *Combinatorics, Probability and Computing 11*, 353-371.

[22] Kahane, J.-P. and Peyrière, J. (1976). Sur certaines martingales de Benoit Mandelbrot. *Adv. Math. 22*, 131-145.

[23] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta. Math. 131*, 207-248.

[24] Kingman, J. F. C. (1973). Subadditive ergodic theory. *Ann. Probab. 1*, 883-909.

[25] Knuth, D. E. (1997) *The art of computer programming. Vol. 1: Fundamental Algorithms. 3rd ed.* Addison-Wesley.

[26] Knuth, D. E. (1997) *The art of computer programming. Vol. 2: Seminumerical Algorithms. 3rd ed.* Addison-Wesley.

[27] Knuth, D. E. (1998) *The art of computer programming. Vol. 3: sorting and searching. 2nd ed.* Addison-Wesley.

[28] Mahmoud, H. M. (1992) *Evolution of random search trees.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York.

[29] Mahmoud, H. M. (1995). The joint distribution of the three types of nodes in uniform binary trees. *Algorithmica 13*, 313–323.

[30] Mahmoud, H. M. (2000). *Sorting. A Distribution Theory*. Wiley-Interscience, New York.

[31] Mahmoud, H. M., Modarres, R. and Smythe, R. T. (1995) Analysis of quickselect: An algorithm for order statistics. *RAIRO Inform. Théor. Appl. 29*, 255-276.

[32] Mahmoud, H. M. and Smythe, R. T. (1991). On the distribution of leaves in rooted subtrees of recursive trees. *Ann. Appl. Probab. 1*, 406–418.

[33] Mahmoud, H. M. and Smythe, R. T. (1992). Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms 3*, 255–266.

[34] Mahmoud, H. M., Smythe, R. T. and Szymański, J. (1993). On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms 4*, 151–176.

[35] Mandelbrot, B. (1974). Multiplications aléatoires itérées et distributions invariantes par moyenne pondérée aléatoire. *C. R. Acad. Sci., Paris, Ser. A 278*, 289-292.

[36] Neininger, R. (1999). *Limit Laws for random recursive structures and algorithms.* Ph.D. Dissertation, Universität Freiburg i. Br., iv, 96 p.

[37] Neininger, R. (2001). On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms 19*, 498-524.

[38] Neininger, R. (2002). The Wiener index of random trees. *Combin. Probab. Comput. 11*, 587-597.

[39] Neininger, R. and Rüschendorf, L. (1999). On the internal path length of d-dimensional quad trees. *Random Structures Algorithms 15*, 25–41.

[40] Neininger, R. and Rüschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab. 14*, 378–418.

[41] Neininger, R. and Rüschendorf, L. (2004). On the contraction method with degenerate limit equation. *Ann. Probab. 32*, 2838–2856.

[42] Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models*. John Wiley.

[43] Rachev, S. T. and Rüschendorf, L. (1995). Probability metrics and recursive algorithms. *Adv. in Appl. Probab. 27*, 770–799.

[44] Régnier, M. (1989). A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl. 23*, 335–343.

[45] Rösler, U. (1991). A limit theorem for “Quicksort”. *RAIRO Inform. Théor. Appl. 25*, 85–100.

[46] Rösler, U. (1992). A fixed point theorem for distributions. *Stochastic Process. Appl. 42*, 195–214.

[47] Rösler, U. (1998). A fixed point equation for distributions. *Berichtsreihe des Mathematischen Seminars Kiel 98-7*.
Available at `http://www.numerik.uni-kiel.de/reports/1998/98-7.ps.gz`

[48] Rösler, U. (2001). On the analysis of stochastic divide and conquer algorithms. *Algorithmica 29*, 238–261.

[49] Rösler, U. and Rüschendorf, L. (2001). The contraction method for recursive algorithms. *Algorithmica 29*, 3–33.

[50] Sedgewick, R. (1975). *Quicksort*. Stanford Computer Science Report STAN-CS-75-492, Ph. D. Thesis, Stanford University.

[51] Sedgewick, R. (1977). The analysis of Quicksort programs. *Acta Informat. 7*, 327–355.

[52] Vervaat, W. (1979) On a stochastic difference equation and a representation of non-negative infinitly divisible. *Adv. in Appl. Probab. 11*, 750-783.

[53] Zolotarev, V. M. (1976). Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theor. Probability Appl. 21*, 721–737.