

On a Multivariate Contraction Method for Random Recursive Structures with Applications to Quicksort

RALPH NEININGER
Institut für Mathematische Stochastik
Universität Freiburg
Eckerstr. 1
79104 Freiburg
Germany

September 10, 2001

Abstract

The contraction method for recursive algorithms is extended to the multivariate analysis of vectors of parameters of recursive structures and algorithms. We prove a general multivariate limit law which also leads to an approach to asymptotic covariances and correlations of the parameters. As an application the asymptotic correlations and a bivariate limit law for the number of key comparisons and exchanges of median-of- $(2t + 1)$ Quicksort is given. Moreover, for the Quicksort programs analyzed by Sedgewick the exact order of the standard deviation and a limit law follow, considering all the parameters counted by Sedgewick.

AMS subject classifications. Primary 60F05, 68Q25; secondary 68P10.

Key words. Contraction method, Quicksort, multivariate limit law, analysis of algorithms, median-of- $(2t + 1)$.

1 Introduction

Over the last ten years limit laws for some parameters of random recursive structures and algorithms, which seemed to resist classical probabilistic techniques, could be derived by the contraction method. This method was introduced by Rösler [43] for the derivation of the limit law of the number of key comparisons needed by Hoare's Quicksort algorithm to sort a list of randomly permuted items. The contraction method was further developed in Rösler [44] and independently in Rachev and Rüschemdorf [40]. A guide for the use of this technique and an overview over the applications up to 1998 is given in the survey article of Rösler and Rüschemdorf [46].

In general, the distribution of a parameter of a recursive structure or algorithm satisfies some recurrence equation on the level of distributions caused by the recursive nature of the structure. In order to derive a limit law for the parameter by the contraction method one proceeds in several steps: First, the right normalization of the parameter has to be found. This is usually done by studying its mean and variance. The original recurrence equation of the parameter induces a modified recursive equation for the normalized quantities, again on the level of distributions. From this a limiting form

has to be determined which gives rise to a transformation on the set of all probability distributions on the real line. Then one chooses a probability metric such that the transformation has contraction properties in this metric. The metric has to be complete on a subspace where the limit distribution is sought. Then Banach's fixed-point theorem yields a unique fixed-point which is the candidate for the limiting distribution. The last step of the method is to establish weak convergence of the scaled parameter to this fixed-point.

In this work we extend this method to the multivariate analysis of vectors of parameters of recursive structures and algorithms and formulate a general theorem in a form to be easily applied. For algorithms usually time and space requirements are of interest, where both quantities may result from various parameters of the algorithm. An accurate asymptotic stochastic description would be a multivariate limit law jointly for all these quantities. For random search trees many parameters such as the depth of insertion of a node, the height, and the internal path length were investigated for its own. It is natural to study these quantities jointly to gain information on the dependence structure beyond the pure marginal distributions of the parameters. For examples of multivariate limit laws in the field of combinatorial structures using various approaches see [1, 29, 18, 30, 31, 26].

This paper is organized as follows: In section two we outline the type of divide-and-conquer structures for which the contraction method is extended to multivariate asymptotic analysis. In the third section contraction properties of transformations from the space of probability distributions to itself are investigated which appear as the limiting operators of the recurrences under consideration. In the fourth section we derive multivariate limit laws for a general recurrence extending a general limit law for one-dimensional stochastic divide-and-conquer algorithms due to Rösler [45]. In contrast to the one-dimensional case it is not clear if the contraction condition for the limiting operator is also sufficient to imply weak convergence of the scaled parameters. Therefore, we have to strengthen the contraction condition in order to get a limit law. This is done in different ways and each of our conditions is tested in the applications. It remains open whether the contraction condition for the *ideal* limiting operator is in general sufficient to imply a limit law or not. This is briefly summarized in section 7.

Section 5 gives the main applications to the analysis of median-of- $(2t+1)$ Quicksort. We consider the vector of the number of key comparisons and key exchanges made by the algorithm. These are the most important parameters of Quicksort since they are of larger order of magnitude than other parameters. The multivariate contraction approach can be applied and results in a bivariate limit law for the joint distribution of these parameters. This leads also to the asymptotic correlation and a first order asymptotic of the covariance of these parameters. As corollaries limit laws and variances of linear combinations of these parameters are obtained. This would also be in the range of a univariate approach but results here without any work. The analysis covers as well the more complex situation of Sedgewick's [50] cost measure for concrete Quicksort implementations. Here the cost of the algorithm is measured as a linear combination of several parameters of the algorithm. Asymptotically only the number of key comparisons and exchanges matter. Therefore we obtain the exact order of the standard deviation of these Quicksort programs as well as a limit law.

In section 6 a family of recurrences is considered in order to test the applicability of the general method and to provide an example, where improvements of the strengthened contraction condition may easily be tested.

The rest of this section is devoted to technical and notational preliminaries. For a random variable X and a probability distribution μ we write $X \sim \mu$ if the law $\mathcal{L}(X)$ of X is μ , similarly $X \sim Y$ for random variables with $\mathcal{L}(X) = \mathcal{L}(Y)$. The law of X is also denoted by \mathbb{P}^X . We will use three different norms. For a vector $x \in \mathbb{R}^d$ by $\|x\|$ the Euclidean norm of x is denoted, $\|X\|_2 := (\mathbb{E} \|X\|^2)^{1/2}$ denotes

the L_2 -norm of a random vector X , and $\|A\|_{\text{op}} := \sup_{\|x\|=1} \|Ax\|$ denotes the spectral radius of a square matrix A . By A^t the transposed of A is denoted. The Wasserstein-metric ℓ_2 is defined on the space of d -dimensional probability distributions with existing second moments by

$$\ell_2(\mu, \nu) := \inf \{ \|X - Y\|_2 : X \sim \mu, Y \sim \nu \}.$$

By $\mathcal{M}_{0,2}^d$ the space of the centered probability measures on \mathbb{R}^d with finite second moment is denoted. The metric space $(\mathcal{M}_{0,2}^d, \ell_2)$ is complete and convergence in ℓ_2 is equivalent to weak convergence plus convergence of the second moments. Random vectors with $X \sim \mu, Y \sim \nu$, and $\ell_2(\mu, \nu) = \|X - Y\|_2$ are called optimal couplings of (μ, ν) . Such optimal couplings exist for all μ, ν with finite second moments. For information on the ℓ_2 metric see [2, 7, 32, 39, 41]. We will also use the notation $\ell_2(X, Y) := \ell_2(\mathcal{L}(X), \mathcal{L}(Y))$. For random variables X, Y with finite second moments we write $\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$ for the covariance of X, Y and $\text{Cor}(X, Y) := \text{Cov}(X, Y)/(\text{Var}(X)^{1/2}\text{Var}(Y)^{1/2})$ for their correlation. By the symbol $\stackrel{\mathcal{D}}{=}$ equality in distribution is denoted even if a random vector and a distribution or two random vectors are compared. The uniform distributions on the unit interval $[0, 1]$ and the unit cube $[0, 1]^d$ are denoted by $\text{unif}[0, 1]$ and $\text{unif}[0, 1]^d$ respectively, $B(n, p)$ denotes the binomial distribution with parameters $n \in \mathbb{N}_0$ and $p \in [0, 1]$, $M(n, p_1, \dots, p_d)$ the corresponding multinomial distribution, and $\text{beta}(a, b)$ stands for the beta distribution with parameters $a, b > 0$.

2 Stochastic divide-and-conquer recurrences

Now, we outline the setting of divide-and-conquer algorithms which are under consideration here. We assume that an algorithm or data structure of (input) size $n \in \mathbb{N}$ is given, where the randomness might come from the input or from the algorithm itself. We consider $d \geq 1$ parameters $Y_n = (Y_{n1}, \dots, Y_{nd})$ which are random variables depending on the input size n . By the recursive nature of the algorithm or structure these parameters can be expressed by the corresponding parameters of the subproblems or substructures into which the original problem or structure is subdivided. We assume that the problem or structure is always subdivided into $K \geq 1$ subproblems of sizes $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$ if the size of the input is n . Here K is a fixed number but $I^{(n)}$ is a random vector. Furthermore, we assume that given the cardinalities $I^{(n)}$ of the subproblems or substructures the vectors of the parameters of these are mutually independent and that the distribution of (Y_{n1}, \dots, Y_{nd}) can be obtained as a random linear combination of all the corresponding $Y_{I_r^{(n)}k}$ for $r = 1, \dots, K, k = 1, \dots, d$ plus a random toll vector b_n . The toll vector measures the cost for subdividing and merging or corresponding effects. More precisely, we assume that the sequence $(Y_n)_{n \in \mathbb{N}_0}$ satisfies the distributional recursion

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

for some $n_0 \geq 1$. Here, the sequences $(Y_n^{(1)}), \dots, (Y_n^{(K)})$ and the vector $(A_1, \dots, A_K, b_n, I^{(n)})$ are independent, A_1, \dots, A_K are random $d \times d$ matrices with some given joint distribution, which might depend on n (suppressed in the notation), b_n is a random vector, $I^{(n)}$ is a vector of random cardinalities $I_r^{(n)} \in \{0, \dots, n\}$ for $r = 1, \dots, K$, and $(Y_n^{(1)}), \dots, (Y_n^{(K)})$ are sequences which are identically distributed as (Y_n) .

Most of the examples given in the survey of Rösler and Rüschemdorf [46] are of the form (1) with dimension $d = 1$. Dependences between $A_1, \dots, A_K, b_n, I^{(n)}$ usually occur in applications to

divide-and-conquer algorithms. We could also allow randomness for K but will drop this in our discussion. An application of the one-dimensional contraction method with random K can be found in Geiger [19].

3 Multivariate Contraction

For the derivation of a limit theorem for the sequence (Y_n) given by recursion (1) we follow the general idea of the contraction method. In this work we restrict ourselves to the use of the Wasserstein metric ℓ_2 . This requires that the random quantities (Y_n) , A_1, \dots, A_K , b_n in (1) are all square-integrable. We scale the Y_{nr} by centering by their mean and by dividing by some $\sigma_r(n) > 0$. Clearly, in an L_2 -setting these $\sigma_r(n)$ should be chosen at the order of the corresponding standard deviations. With the diagonal matrices

$$D_n := \text{diag}(\sigma_1(n), \dots, \sigma_d(n))$$

and the notation $M_n := \mathbb{E} Y_n$ we define the scaled version X_n of Y_n by

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

The original recursion (1) implies the modified recursion for the scaled vectors, $n \geq n_0$,

$$\begin{aligned} X_n &\stackrel{\mathcal{D}}{=} D_n^{-1} \left(\sum_{r=1}^K A_r Y_{I_r^{(n)}}^{(r)} + b_n - M_n \right) \\ &\stackrel{\mathcal{D}}{=} D_n^{-1} \left(\sum_{r=1}^K A_r \left(D_{I_r^{(n)}} X_{I_r^{(n)}}^{(r)} + M_{I_r^{(n)}} \right) + b_n - M_n \right) \\ &= \sum_{r=1}^K \left(D_n^{-1} A_r D_{I_r^{(n)}} \right) X_{I_r^{(n)}}^{(r)} + \sum_{r=1}^K \left(D_n^{-1} A_r M_{I_r^{(n)}} \right) + D_n^{-1} (b_n - M_n) \\ &= \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \end{aligned} \quad (3)$$

where

$$A_r^{(n)} := D_n^{-1} A_r D_{I_r^{(n)}}, \quad b^{(n)} := \sum_{r=1}^K \left(D_n^{-1} A_r M_{I_r^{(n)}} \right) + D_n^{-1} (b_n - M_n), \quad (4)$$

and $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}, I^{(n)})$, $(X_n^{(1)}), \dots, (X_n^{(K)})$ are, corresponding to the original recursion, independent with $(X_n^{(r)}) \sim (X_n)$ for $r = 1, \dots, K$. According to the concept of the contraction method we are looking for a limiting form of equation (3). Therefore we assume L_2 -convergence of $(b^{(n)})$ and $(A_r^{(n)})$ for $r = 1, \dots, K$:

$$\|A_r^{(n)} - A_r^*\|_2^2 \rightarrow 0, \quad \|b^{(n)} - b^*\|_2^2 \rightarrow 0, \quad n \rightarrow \infty, \quad (5)$$

with appropriate $(A_1^*, \dots, A_K^*, b^*)$. Then, one suggests that a limit X of (X_n) satisfies the distributional recursion

$$X \stackrel{\mathcal{D}}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^* \quad (6)$$

with $(A_1^*, \dots, A_K^*, b^*)$, $X^{(1)}, \dots, X^{(K)}$ being independent and $X^{(r)} \sim X$ for $r = 1, \dots, K$.

In the subsequent we have to find conditions which imply that (6) has a unique distributional fixed-point and that in fact convergence of (X_n) to this fixed-point holds.

The following multivariate contraction lemma from the author's dissertation [34] generalizes two special cases. For dimension $d = 1$ this is the well-known contraction lemma (see, e.g., Lemma 1 in Rösler and Rüschendorf [46]). In general dimension but with $K = 1$ our lemma reduces to Theorem 1 in Burton and Rösler [4].

Lemma 3.1 (Multivariate Contraction Lemma) *Let (A_1, \dots, A_K, b) be a square-integrable vector of random $d \times d$ matrices A_1, \dots, A_K and a random d -dimensional vector b with $\mathbb{E}b = 0$, and let the transformation $T : \mathcal{M}_{0,2}^d \rightarrow \mathcal{M}_{0,2}^d$ be defined by*

$$T(\mu) := \mathcal{L} \left(\sum_{r=1}^K A_r Z^{(r)} + b \right), \quad \mu \in \mathcal{M}_{0,2}^d,$$

where (A_1, \dots, A_K, b) , $Z^{(1)}, \dots, Z^{(K)}$ are independent and $Z^{(r)} \sim \mu$ for all $r = 1, \dots, K$. Then T is a contraction with respect to the ℓ_2 -metric if

$$\left\| \sum_{r=1}^K \mathbb{E} [A_r^t A_r] \right\|_{\text{op}} < 1. \quad (7)$$

Proof: Clearly $T(\mu)$ has a second moment and $\mathbb{E}T(\mu) = 0$ for all $\mu \in \mathcal{M}_{0,2}^d$ by the valid independence conditions and $\mathbb{E}b = 0$, so $T : \mathcal{M}_{0,2}^d \rightarrow \mathcal{M}_{0,2}^d$ is a well-defined map. Let $\mu, \nu \in \mathcal{M}_{0,2}^d$ be given and $(W^{(1)}, Z^{(1)}), \dots, (W^{(K)}, Z^{(K)})$ be optimal couplings of (μ, ν) for $r = 1, \dots, K$ so that (A_1, \dots, A_K, b) , $(W^{(1)}, Z^{(1)}), \dots, (W^{(K)}, Z^{(K)})$ are independent. Then

$$\begin{aligned} \ell_2^2(T(\mu), T(\nu)) &\leq \mathbb{E} \left\| \sum_{r=1}^K A_r (W^{(r)} - Z^{(r)}) \right\|^2 \\ &= \mathbb{E} \sum_{r=1}^K \left\langle A_r (W^{(r)} - Z^{(r)}), A_r (W^{(r)} - Z^{(r)}) \right\rangle \\ &\quad + \mathbb{E} \sum_{\substack{r,s=1 \\ r \neq s}}^K \left\langle A_r (W^{(r)} - Z^{(r)}), A_s (W^{(s)} - Z^{(s)}) \right\rangle \end{aligned} \quad (8)$$

$$\begin{aligned} &= \sum_{r=1}^K \mathbb{E} \left\langle W^{(r)} - Z^{(r)}, A_r^t A_r (W^{(r)} - Z^{(r)}) \right\rangle \\ &= \sum_{r=1}^K \mathbb{E} \left\langle W^{(r)} - Z^{(r)}, \mathbb{E} [A_r^t A_r] (W^{(r)} - Z^{(r)}) \right\rangle \end{aligned} \quad (9)$$

$$\begin{aligned} &= \mathbb{E} \left\langle W^{(1)} - Z^{(1)}, \left(\sum_{r=1}^K \mathbb{E} [A_r^t A_r] \right) (W^{(1)} - Z^{(1)}) \right\rangle \\ &\leq \left\| \sum_{r=1}^K \mathbb{E} [A_r^t A_r] \right\|_{\text{op}} \mathbb{E} \|W^{(1)} - Z^{(1)}\|^2 \\ &= \left\| \sum_{r=1}^K \mathbb{E} [A_r^t A_r] \right\|_{\text{op}} \ell_2^2(\mu, \nu). \end{aligned} \quad (10)$$

The sum in (8) is zero by independence and $\mathbb{E}(W^{(r)} - Z^{(r)}) = 0$; the additional expectation in (9) is justified by independence. \blacksquare

The improvement of the contraction condition (7) by the insertion of an additional expectation in (9) is (for the case $K = 1$) discussed in Burton and Rösler [4].

Note that the estimate in (10) is sharp: $S := \sum_{r=1}^K \mathbb{E}[A_r^t A_r]$ is a symmetric and positive semi-definite matrix. Therefore, it is $\|S\|_{\text{op}} = \lambda$ if λ denotes the largest eigenvalue of S . Let $u \in \mathbb{R}^d$ be a corresponding eigenvector and $\mu, \nu \in \mathcal{M}_{0,2}^d$, where ν is the Dirac measure in zero and μ the probability measure with mass 1/2 on u and $-u$. Then $W \sim \mu$ and $Z = 0$ is an optimal coupling of (μ, ν) for all realizations W of μ . It follows

$$\begin{aligned} \mathbb{E} \langle W - Z, S(W - Z) \rangle &= \mathbb{E} \langle W, SW \rangle = \mathbb{E} \langle W, \lambda W \rangle = \lambda \mathbb{E} \|W\|^2 \\ &= \|S\|_{\text{op}} \ell_2^2(\mu, \nu). \end{aligned}$$

This shows the sharpness in (10).

In the case $K = 2$, $d = 2$, and $b = 0$ Cramer and Rüschemdorf [6] were led to the map T in Lemma 3.1 as a limiting operator of a related branching recursion. It is easy to see that their contraction conditions (2.13), (2.14) and Proposition 2.5 coincide with our representation in terms of the spectral radius in (7).

4 Multivariate limit laws

In the following we come back to the situation where we are given a sequence (Y_n) of random vectors satisfying the recurrence (1) and that after scaling the Y_n as in (2) the scaled variates (X_n) satisfy the modified recursion (3). According to the idea of the contraction method we are looking for a theorem saying roughly that convergence of the coefficients as in (5) implies under appropriate conditions convergence of the (X_n) . The following theorem yields such a transfer being an extension of a general one-dimensional limit law for stochastic divide-and-conquer algorithms due to Rösler [45].

Theorem 4.1 *Let (X_n) be a sequence of d -dimensional square-integrable random vectors satisfying the distributional recursion*

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

where $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}, I^{(n)})$, $(X_n^{(1)}), \dots, (X_n^{(K)})$ are independent, $A_1^{(n)}, \dots, A_K^{(n)}$ are square-integrable random $d \times d$ matrices, $b^{(n)}$ is a square-integrable random vector, $X_n^{(r)} \sim X_n$, and $I^{(n)}$ is a vector of random integers with $I_r^{(n)} \in \{0, \dots, n\}$, $r = 1, \dots, K, n \geq 0$. Let the following conditions be satisfied:

$$(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L_2} (A_1^*, \dots, A_K^*, b^*), \quad n \rightarrow \infty, \quad (11)$$

$$\sum_{r=1}^K \mathbb{E} \|(A_r^*)^t A_r^*\|_{\text{op}} < 1, \quad (12)$$

$$\mathbb{E} \left[\mathbf{1}_{\{I_r^{(n)} \leq l\} \cup \{I_r^{(n)} = n\}} \|(A_r^{(n)})^t A_r^{(n)}\|_{\text{op}} \right] \rightarrow 0, \quad n \rightarrow \infty, \quad (13)$$

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

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

where X is the in $\mathcal{M}_{0,2}^d$ unique distributional fixed-point of

$$X \stackrel{\mathcal{D}}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*, \quad (14)$$

with $(A_1^*, \dots, A_K^*, b^*), X^{(1)}, \dots, X^{(K)}$ independent and $X^{(r)} \sim X$ for $r = 1, \dots, K$.

Proof: By Jensen's inequality (12) implies $\|\sum \mathbb{E}[(A_r^*)^t A_r^*]\|_{\text{op}} < 1$. By the definition of $b^{(n)}$ we have $\mathbb{E} b^{(n)} = 0$ for all $n \in \mathbb{N}$. Thus, the L_2 -convergence of $(b^{(n)})$ implies $\mathbb{E} b^* = 0$. Therefore, by Lemma 3.1, the limiting equation (14) has a unique distributional fixed-point X in $\mathcal{M}_{0,2}^d$. 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 $r = 1, \dots, K$ and that $(A_1, \dots, A_K, b_n, I^{(n)}), (X_n^{(1)}, X^{(1)}), \dots, (X_n^{(K)}, X^{(K)})$ are independent. The first step is to derive an estimate of $\ell_2^2(X_n, X)$ in terms of $\ell_2^2(X_i, X)$ only with indices $i \in \{0, \dots, n-1\}$. This reduction inequality for the sequence $(\ell_2^2(X_n, X))$ will be sufficient to deduce $\ell_2(X_n, X) \rightarrow 0$. To derive such a reduction inequality we use the representations (3) and (14) 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. We start for $n \geq n_0$ with the estimate

$$\begin{aligned} \ell_2^2(X_n, X) &\leq \left\| \sum_{r=1}^K \left(A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right) + b^{(n)} - b^* \right\|_2^2 \\ &= \sum_{r=1}^K \left\| A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right\|_2^2 + \left\| b^{(n)} - b^* \right\|_2^2 \\ &\quad + \sum_{\substack{r,s=1 \\ r \neq s}}^K \mathbb{E} \left\langle A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)}, A_s^{(n)} X_{I_s^{(n)}}^{(s)} - A_s^* X^{(s)} \right\rangle \\ &\quad + 2 \sum_{r=1}^K \mathbb{E} \left\langle A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)}, b^{(n)} - b^* \right\rangle. \end{aligned} \quad (15)$$

The third and fourth summand in (15) are zero by independence and $\mathbb{E} X^{(r)} = \mathbb{E} X_{I_r^{(n)}}^{(r)} = 0$. By our assumption we have $\|b^{(n)} - b^*\|_2^2 \rightarrow 0$ for $n \rightarrow \infty$, so we only have to care about the first summand:

$$\begin{aligned} &\sum_{r=1}^K \left\| A_r^{(n)} X_{I_r^{(n)}}^{(r)} - A_r^* X^{(r)} \right\|_2^2 \\ &= \sum_{r=1}^K \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) + \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|_2^2 \\ &= \sum_{r=1}^K \left(\left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2^2 + \left\| \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\|_2^2 \right. \\ &\quad \left. + 2 \mathbb{E} \left\langle A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right), \left(A_r^{(n)} - A_r^* \right) X^{(r)} \right\rangle \right). \end{aligned} \quad (16)$$

By (11), independence, and $\|X\|_2 < \infty$ we obtain

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

and $r = 1, \dots, K$. The first summand in (16) can be estimated by

$$\begin{aligned} & \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2^2 \\ &= \mathbb{E} \left\langle X_{I_r^{(n)}}^{(r)} - X^{(r)}, (A_r^{(n)})^t A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\rangle \\ &\leq \mathbb{E} \left[\left\| (A_r^{(n)})^t A_r^{(n)} \right\|_{\text{op}} \left\| X_{I_r^{(n)}}^{(r)} - X^{(r)} \right\|_2^2 \right]. \end{aligned} \tag{17}$$

Since the operator norm is a Lipschitz continuous map and by the L_2 -convergence of $(A_r^{(n)})$ we deduce

$$\begin{aligned} & \mathbb{E} \left\| A_r^{(n)} - A_r^* \right\|_{\text{op}}^2 \rightarrow 0, \\ & \mathbb{E} \left\| (A_r^{(n)})^t A_r^{(n)} \right\|_{\text{op}} \rightarrow \mathbb{E} \left\| (A_r^*)^t A_r^* \right\|_{\text{op}}, \quad n \rightarrow \infty, \end{aligned}$$

for $r = 1, \dots, K$. In the following the symbol $o(1)$ might denote different sequences tending to zero. The third summand in (16) can be estimated by

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

Putting these estimates together and denoting (with the same $o(1)$)

$$A_r^{((n))} := \left\| (A_r^{(n)})^t A_r^{(n)} \right\|_{\text{op}} (1 + o(1)),$$

we obtain

$$\begin{aligned} \ell_2^2(X_n, X) &\leq \sum_{r=1}^K \mathbb{E} \left[A_r^{((n))} \left\| X_{I_r^{(n)}}^{(r)} - X^{(r)} \right\|_2^2 \right] + o(1) \\ &= \sum_{r=1}^K \mathbb{E} \left[\sum_{i=0}^n \mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))} \left\| X_i^{(r)} - X^{(r)} \right\|_2^2 \right] + o(1) \\ &= \sum_{i=0}^n \left(\sum_{r=1}^K \mathbb{E} \left[\mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))} \right] \right) \ell_2^2(X_i, X) + o(1). \end{aligned}$$

With the abbreviations

$$a_n := \ell_2^2(X_n, X), \quad p_n := \sum_{r=1}^K \mathbb{E} \left[\mathbf{1}_{\{I_r^{(n)}=n\}} A_r^{((n))} \right], \quad (18)$$

this implies

$$\begin{aligned} (1-p_n)a_n &\leq \sum_{r=1}^K \mathbb{E} \left[A_r^{((n))} \right] \sup_{0 \leq i \leq n-1} a_i + o(1) \\ &= \sum_{r=1}^K \mathbb{E} \left[\|(A_r^*)^t A_r^*\|_{\text{op}} + o(1) \right] \sup_{0 \leq i \leq n-1} a_i + o(1). \end{aligned} \quad (19)$$

By (13) we have $p_n \rightarrow 0$, thus the assumption $\sum \mathbb{E} \|(A_r^*)^t A_r^*\|_{\text{op}} < 1$ implies that (a_n) is a bounded sequence. We define $a := \limsup a_n$. Now, it exists a $\xi < 1$ such that for all $\varepsilon > 0$ there exists a $n_1 \in \mathbb{N}$ with $a_n \leq a + \varepsilon$ for all $n \geq n_1$ and such that the prefactor in (19) satisfies $\sum \mathbb{E} [\|(A_r^*)^t A_r^*\|_{\text{op}} + o(1)] \leq \xi$ for $n \geq n_1$. Then from (18) we deduce

$$\begin{aligned} a_n &\leq \frac{1}{1-p_n} \left[\sum_{i=0}^{n_1-1} \left(\sum_{r=1}^K \mathbb{E} \left[\mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))} \right] \right) a_i \right. \\ &\quad \left. + \sum_{i=n_1}^{n-1} \left(\sum_{r=1}^K \mathbb{E} \left[\mathbf{1}_{\{I_r^{(n)}=i\}} A_r^{((n))} \right] \right) (a + \varepsilon) + o(1) \right] \\ &\leq \frac{1}{1-p_n} (\xi(a + \varepsilon) + o(1)), \end{aligned} \quad (20)$$

where (13) has been used. The $o(1)$ depends on ε . Since $\varepsilon > 0$ is arbitrary we conclude with $n \rightarrow \infty$ that $a = 0$. \blacksquare

For the application of this limit law it is necessary to scale the quantities at the right order of magnitude (cf. (2)). With $\sigma_r(n)$ growing too fast we will get $b^* = 0$ for the limiting equation (14). Then the conditions of the limit law might be still satisfied but the unique solution in (14) is the degenerated one $X = 0$. With $\sigma_r(n)$ growing too slow we typically cannot satisfy $b^{(n)} \rightarrow b^*$ as in (11).

Note, that we had to strengthen our *ideal* contraction condition (7) to (12) in order to derive convergence in the ℓ_2 metric. It would be interesting to know whether also (7) in general implies our limit law or not. In the applications in section 6 the fulfillment of the strengthened condition (12) will require some restrictions which would not be necessary to satisfy the ideal contraction condition (7). For the special case of diagonal matrices A_1, \dots, A_K we give an alternative sufficient condition for the limit law. In this condition (22) we try to imitate the expectation inside the spectral radius in (7). The utility of (22) will become clear in the applications of section 6.

Corollary 4.2 *With diagonal matrices A_1, \dots, A_K in the situation of Theorem 4.1, condition (13) replaced by*

$$\sum_{i \in \{0, \dots, l\} \cup \{n\}} \max_{1 \leq k \leq d} \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right) \rightarrow 0 \quad \text{for } n \rightarrow \infty, \quad (21)$$

and (12) replaced by

$$\limsup_{n \rightarrow \infty} \sum_{i=0}^n \max_{1 \leq k \leq d} \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right) < 1 \quad (22)$$

we have

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

where X is the unique distributional fixed-point of (14) in $\mathcal{M}_{0,2}^d$.

Proof: We proceed as in the proof of Theorem 4.1. Note that (22) also implies $\left\| \sum \mathbb{E} [(A_r^*)^t A_r^*] \right\|_{\text{op}} < 1$ in the case of diagonal matrices A_1, \dots, A_K :

$$\begin{aligned} \left\| \sum_{r=1}^K \mathbb{E} [(A_r^*)^t A_r^*] \right\|_{\text{op}} &= \lim_{n \rightarrow \infty} \left\| \sum_{r=1}^K \mathbb{E} \left[\left(A_r^{(n)} \right)^t A_r^{(n)} \right] \right\|_{\text{op}} \\ &= \lim_{n \rightarrow \infty} \max_{1 \leq k \leq d} \sum_{r=1}^K \mathbb{E} \left(A_r^{(n)} \right)_{kk}^2 \\ &= \lim_{n \rightarrow \infty} \max_{1 \leq k \leq d} \sum_{i=0}^n \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right) \\ &\leq \limsup_{n \rightarrow \infty} \sum_{i=0}^n \max_{1 \leq k \leq d} \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right), \end{aligned}$$

thus, by Lemma 3.1, the limiting equation (14) has a unique distributional fixed-point. Now, for the reduction inequality for $\ell_2^2(X_n, X)$ we follow the proof of Theorem 4.1 up to (17). Using $(A_r^{(n)})^t = A_r^{(n)}$ we replace (17) by

$$\begin{aligned} &\sum_{r=1}^K \left\| A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\|_2^2 \\ &= \sum_{r=1}^K \mathbb{E} \left\langle X_{I_r^{(n)}}^{(r)} - X^{(r)}, \left(A_r^{(n)} \right) A_r^{(n)} \left(X_{I_r^{(n)}}^{(r)} - X^{(r)} \right) \right\rangle \\ &= \sum_{r=1}^K \sum_{i=0}^n \mathbb{E} \left[\mathbf{1}_{\{I_r^{(n)}=i\}} \left\langle X_i^{(r)} - X^{(r)}, \left(A_r^{(n)} \right) A_r^{(n)} \left(X_i^{(r)} - X^{(r)} \right) \right\rangle \right] \\ &= \sum_{i=0}^n \mathbb{E} \left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)}=i\}} \sum_{k=1}^d \left(A_r^{(n)} \right)_{kk}^2 \left(X_i^{(r)} - X^{(r)} \right)_k^2 \right] \\ &= \sum_{i=0}^n \sum_{k=1}^d \left(\mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right) \mathbb{E} \left(X_i - X \right)_k^2 \right) \\ &\leq \sum_{i=0}^n \left(\max_{1 \leq k \leq d} \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 \right) \left\| X_i - X \right\|_2^2 \right). \end{aligned}$$

With this estimate and the arguments of the proof of Theorem 4.1 we deduce corresponding to (18)

$$\ell_2^2(X_n, X) \leq \sum_{i=0}^n \left(\left(\max_{1 \leq k \leq d} \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 (1 + o(1)) \right) \right) \ell_2^2(X_i, X) \right) + o(1).$$

Analogously to (18) we define

$$p_n := \max_{1 \leq k \leq d} \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=n\}} \left(A_r^{(n)} \right)_{kk}^2 (1 + o(1)) \right),$$

where the $o(1)$ is the corresponding one from the previous inequality. Then, similarly to (19), we derive

$$(1 - p_n)a_n \leq \sum_{i=0}^{n-1} \left(\max_{1 \leq k \leq d} \mathbb{E} \sum_{r=1}^K \left(\mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 (1 + o(1)) \right) \right) \sup_{0 \leq i \leq n-1} a_i + o(1).$$

By (21) and (22) this again implies that (a_n) is bounded and based on (23) similarly to (20) we deduce $\ell_2^2(X_n, X) \rightarrow 0$ for $n \rightarrow \infty$. ■

5 Applications: Median-of- $(2t + 1)$ Quicksort

In this section we consider the median-of- $(2t + 1)$ version of Hoare's Quicksort algorithm with $t \in \mathbb{N}_0$. For measuring the performance of Quicksort algorithms several parameters have been considered, the most important being the number of key comparisons, key exchanges, partitioning stages, and stack pushes and pops made during the execution of the algorithm. One approach to define a univariate cost measure for Quicksort algorithms may be in taking linear combinations to weight the specific parameters, see Sedgewick [49, 50].

We consider mainly the number of key comparisons C_n and key exchanges B_n , since these parameters are in the mean of the order $n \ln n$, whereas other parameters are of smaller order. The n denotes the number of items to be sorted and the underlying probabilistic model consists of all permutations of the items being equally likely. We assume that the splitting into the subfiles is done while preserving randomness in and independence between the subfiles. For C_n a huge body of probabilistic results is available even for the median-of- $(2t + 1)$ version of Quicksort. These 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 [22, 23], Régnier [42], Rösler [43, 45], McDiarmid and Hayward [11], Bruhn [3], and for a detailed survey the book of Mahmoud [28]. For the number of exchanges B_n the mean and variance were for general $t \in \mathbb{N}_0$ studied in Hennequin [23], Chern and Hwang [5] refined the analysis of the mean, and Hwang and Neininger [25] gave a limit law for the standard case $t = 0$.

Here we will give an asymptotic analysis of the joint distribution $Y_n := (C_n, B_n)$ for general $t \in \mathbb{N}_0$. A bivariate limit law is derived, which covers especially the missing one-dimensional limit laws for B_n with $t \geq 1$. Moreover, asymptotic correlations and covariances for C_n and B_n are derived. Since weak convergence of measures is preserved under continuous transformations the bivariate limit law covers as well continuous functions of the scaled versions of C_n and B_n . The transformations which are of interest from a practical point of view are the linear combinations $C_n + wB_n$ with $w > 0$. This models the cost of the algorithm assuming that a key exchange has w times the cost of a comparison. The subsequent analysis is based on our transfer Theorem 4.1 which gives the

results quite immediate. Due to the special type of the distributional recursion for (C_n, B_n) we could alternatively also combine a purely univariate approach by the contraction method with the Cramér-Wold device. However, for more complex examples (the $A_r^{(n)}$ not being multiples of the identity matrix) the Cramér-Wold device would not be applicable. We will report on such examples in subsequent work [36].

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

$$C_n \stackrel{\mathcal{D}}{=} C_{I_n}^{(1)} + C_{n-1-I_n}^{(2)} + n - 1 + S_n^c, \quad n \geq n_0, \quad (23)$$

where $I_n + 1$ is the order of the pivot element of the first partition stage. Furthermore, $(C_n^{(1)}), (C_n^{(2)}), (I_n, S_n^c)$ are independent, $C_n^{(1)} \sim C_n^{(2)} \sim C_n$, and (S_n^c) 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_n^c are required. To initialize the algorithm some (random) bounded costs C_0, \dots, C_{n_0-1} have to be given with a $n_0 \geq 2t + 1$ denoting the maximal size of the subfiles, which are sorted by some other sorting procedure.

For the number of key exchanges we have

$$B_n \stackrel{\mathcal{D}}{=} B_{I_n}^{(1)} + B_{n-1-I_n}^{(2)} + T_n + S_n^b, \quad n \geq n_0, \quad (24)$$

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

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

see Sedgewick [49].

We emphasize that the relation (24) is only correct due to the assumption that the file is permuted uniformly at random and that the randomness and independence between subfiles is preserved. Note that for the corresponding relation (23) for the key comparisons it would be sufficient to select the pivot element from a uniformly chosen subsample, where the permutation of the file is then irrelevant. This difference is most transparent when looking at a sorted list.

In order to apply our framework we have first to settle some basic facts about T_n and B_n . Sedgewick [49, p. 226] showed the expansion $\mathbb{E}T_n = (t + 1)/(2(2t + 3))n + O(1)$. We rederive this mean since also information on T_n^2 is required.

Lemma 5.1 *The mean of the number of key exchanges T_n during a partitioning stage of $n \geq 2$ elements by the median-of- $(2t + 1)$ quicksort is given by*

$$\mathbb{E}T_n = \frac{t + 1}{2(2t + 3)}n - \frac{1}{2t + 3} + \frac{t}{2t + 3} \frac{1}{n - 1}, \quad n \geq 2t + 1.$$

Proof: Let G_j denote the event that the key with order $1 \leq j \leq n$ is exchanged in the first partition stage. Then conditioned on I_n the number of exchanges T_n in the first partitioning stage is obtained by appropriate counting of these events:

$$\mathbb{E}[T_n \mid I_n = k] = \sum_{j=1}^k \mathbb{P}(G_j \mid I_n = k). \quad (25)$$

Given $I_n = k$, a key with order $1 \leq j \leq k$ is exchanged if it is placed on a position $k + 2, \dots, n$. Since all permutations of the keys are equally likely we have $\mathbb{P}(G_j | I_n = k) = (n - k - 1)/(n - 1)$ for $1 \leq j \leq k \leq n - 1$. Hence we obtain

$$\mathbb{E}[T_n | I_n] = \frac{I_n(n - 1 - I_n)}{n - 1} \quad (26)$$

almost surely. The pivot $I_n + 1$ is given as the median of a random sample of size $2t + 1$ out of $1, \dots, n$, thus we have for $n \geq 2t + 1$

$$\mathbb{P}(I_n + 1 = j) = \frac{\binom{j-1}{t} \binom{n-j}{t}}{\binom{n}{2t+1}}, \quad t + 1 \leq j \leq n - t. \quad (27)$$

By symmetry it is $\mathbb{E} I_n = (n - 1)/2$. Using the combinatorial identity

$$\sum_{j=0}^{n-1} \frac{\binom{j}{t} \binom{n-1-j}{t}}{\binom{n}{2t+1}} (j+1)(j+2) = \frac{t+2}{2(2t+3)} (n+1)(n+2)$$

we obtain

$$\mathbb{E} I_n^2 = \frac{t+2}{2(2t+3)} n^2 - \frac{3(t+1)}{2(2t+3)} n + \frac{1}{2(2t+3)}.$$

The statement follows now by taking expectations in (26). ■

Subsequently we will also need the second moment of T_n given I_n . A representation analogous to (25) and

$$\mathbb{P}(G_i \cap G_j | I_n = k) = \frac{(n - 1 - k)(n - 2 - k)}{(n - 1)(n - 2)}$$

for $1 \leq i < j \leq k \leq n - 1$ imply

$$\mathbb{E}[T_n^2 | I_n] = \frac{I_n(I_n - 1)(n - 1 - I_n)(n - 2 - I_n)}{(n - 1)(n - 2)} + \frac{I_n(n - 1 - I_n)}{n - 1} \quad (28)$$

almost surely. Note, that all the factorial moments of T_n given I_n are implicitly contained in Hwang and Neininger [25].

Lemma 5.2 *The mean of the number of key exchanges B_n for the median-of- $(2t + 1)$ Quicksort applied to a randomly permuted set of items satisfies*

$$\mathbb{E} B_n = \frac{t + 1}{2(2t + 3)(H_{2t+2} - H_{t+1})} n \ln(n) + c_t n + o(n), \quad (29)$$

with a constant $c_t \in \mathbb{R}$ depending on the indicial conditions and (S_n^b) .

Proof: For the transfer from the mean of the toll function $T_n + S_n^b$ to the mean of B_n we apply a general result on median-of- $(2t + 1)$ Quicksort recursions due to Bruhn [3] and Rösler [45], who proved: If B_n satisfies (24) with a toll function T'_n satisfying $\mathbb{E} T'_n = \beta n + O(1)$ for some $\beta > 0$, then it follows $\mathbb{E} B_n = (\beta/(H_{2t+2} - H_{t+1})) n \ln(n) + cn + o(n)$, where the constant $c \in \mathbb{R}$ depends on t, β, n_0 ,

the initial values $\mathbb{E} B_0, \dots, \mathbb{E} B_{n_0-1}$, and the $O(1)$. By Lemma 5.1 our toll function $T'_n := T_n + S_n^b$ satisfies $\mathbb{E} T'_n = (t+1)/(2(2t+3))n + O(1)$. Therefore the transfer theorem of Bruhn and Rösler applies. \blacksquare

The leading term in (29) was given by Hennequin [23, Annexe C, equation (C.4)] with an error estimate of $O(n)$. For the application of the contraction method the refined expansion (29) is required. The mean of the number of comparisons C_n satisfies

$$\mathbb{E} C_n = \frac{1}{H_{2t+2} - H_{t+1}} n \ln(n) + c'_t n + o(n), \quad (30)$$

with a constant $c'_t \in \mathbb{R}$. The derivation of expansions of this type, which are of interest for the application of the contraction method, was the original motivation for the general Bruhn–Rösler transfer theorem. The leading term in (30) was obtained by van Emden [12] and Hurwitz [24]. The constant c'_t depends on the implementation. Contributions to the derivation of explicit representations of c'_t are given in Green [21], Hennequin [22, p. 327], and Chern and Hwang [5, p. 62].

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 $Y_n = (C_n, B_n)^t$ satisfies the recursion

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

with $(Y_n^{(1)}), (Y_n^{(2)}), (I^{(n)}, b_n)$ being independent, $Y_n^{(1)} \sim Y_n^{(2)} \sim Y_n$, $I^{(n)} = (I_n, n-1-I_n)$, $b_n = (n-1 + S_n^c, T_n + S_n^b)^t$, and I_n, T_n as above. We scale using the matrix $D_n := \text{diag}(n, n)$. With the expansions (29) and (30) we obtain for the scaled quantities $X_n := D_n^{-1}(Y_n - \mathbb{E} Y_n)$

$$X_n \stackrel{d}{=} A_1^{(n)} X_{I_1^{(n)}}^{(1)} + A_2^{(n)} X_{I_2^{(n)}}^{(2)} + b^{(n)}, \quad n \geq n_0, \quad (31)$$

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

$$b^{(n)} = \left(1 + \mu_c^{(t)} \left(\frac{I_1^{(n)}}{n} \ln \frac{I_1^{(n)}}{n} + \frac{I_2^{(n)}}{n} \ln \frac{I_2^{(n)}}{n} \right), \right. \\ \left. \frac{T_n}{n} + \mu_b^{(t)} \left(\frac{I_1^{(n)}}{n} \ln \frac{I_1^{(n)}}{n} + \frac{I_2^{(n)}}{n} \ln \frac{I_2^{(n)}}{n} \right) \right)^t + o(1),$$

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 (31) we use that for all $p > 0$

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

where V has the beta($t+1, t+1$) distribution. The convergence of I_n/n is obvious since the median-of- $(2t+1)$ independent unif[0, 1] distributed random variables is beta($t+1, t+1$) distributed and

we are allowed to choose versions of I_n such that the convergence holds in L_p as well. For the convergence of T_n/n we estimate

$$\left\| \frac{T_n}{n} - V(1-V) \right\|_2 \leq \left\| \frac{T_n}{n} - \frac{I_n(n-I_n)}{n^2} \right\|_2 + \left\| \frac{I_n(n-I_n)}{n^2} - V(1-V) \right\|_2.$$

The first summand is seen to tend to zero by taking the square, multiplying out, conditioning by I_n and applying (26) and (28). Thus, we have the L_2 -convergences

$$b^{(n)} \rightarrow b^*, \quad A_r^{(n)} \rightarrow A_r^*, \quad r = 1, 2, \quad (32)$$

with

$$A_1^* = \begin{pmatrix} V & 0 \\ 0 & V \end{pmatrix}, \quad A_2^* = \begin{pmatrix} 1-V & 0 \\ 0 & 1-V \end{pmatrix}, \quad (33)$$

$$b^* = \left(1 + \mu_c^{(t)} \mathcal{E}(V), V(1-V) + \mu_b^{(t)} \mathcal{E}(V) \right)^t, \quad (34)$$

with $\mathcal{E}(V) := V \ln(V) + (1-V) \ln(1-V)$.

Theorem 5.3 *The normalized vector of the number of key comparisons and key exchanges made by a median-of- $(2t+1)$ version of Quicksort satisfies*

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

where X is the unique distributional fixed-point in $\mathcal{M}_{0,2}^2$ of

$$X \stackrel{D}{=} A_1^* X^{(1)} + A_2^* X^{(2)} + b^*,$$

with $X^{(1)}, X^{(2)} \sim X$ being independent and independent of (A_1^*, A_2^*, b^*) , where (A_1^*, A_2^*, b^*) is given by (33), (34) with V there being $\text{beta}(t+1, t+1)$ distributed.

Proof: We apply Theorem 4.1. It is

$$\mathbb{E} [\mathbf{1}_{\{I_r^{(n)} \leq l\}} \|(A_r^{(n)})^t A_r^{(n)}\|_{\text{op}}] \leq \mathbb{P}(I_r^{(n)} \leq l) \rightarrow 0, \quad n \rightarrow \infty,$$

for all $l \in \mathbb{N}$ and $r = 1, 2$, thus (13) is satisfied. The L_2 -convergence of $A_r^{(n)}, b^{(n)}$ was checked in (32). It is $\|A_1^*\|_{\text{op}} = V$, $\|A_2^*\|_{\text{op}} = 1-V$, with $V \sim \text{beta}(t+1, t+1)$, thus

$$\mathbb{E} \|A_1^*\|_{\text{op}}^2 + \mathbb{E} \|A_2^*\|_{\text{op}}^2 = \frac{t+2}{2t+3} < 1.$$

The conditions of Theorem 4.1 are satisfied. ■

Corollary 5.4 *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*

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

where $b^* = (b_1^*, b_2^*)^t$ is given in (34).

Proof: The fixed-point equation for the limit distribution $X = (X_1, X_2)^t$ implies

$$\mathbb{E}[X_1 X_2] = \frac{2t+3}{t+1} \mathbb{E}[b_1^* b_2^*], \quad \text{Var}(X_r) = \frac{2t+3}{t+1} \mathbb{E}[(b_r^*)^2]$$

for $r = 1, 2$. Since convergence in ℓ_2 implies convergence of the correlations the asymptotic correlation follows. Furthermore it is

$$\text{Cov}(C_n, B_n) = \text{Cov}\left(\frac{C_n - \mathbb{E}C_n}{n}, \frac{B_n - \mathbb{E}B_n}{n}\right) n^2 = (1 + o(1)) \mathbb{E}[X_1 X_2] n^2.$$

The second assertion follows. ■

Explicit terms for the variances of C_n and B_n were calculated by Hennequin [23, Proposition IV.8, IV.7]. The asymptotic correlation from Corollary 5.4 is for, e.g., the median-of-3 Quicksort

$$\frac{\sqrt{2}(471 - 48\pi^2)}{\sqrt{(485 - 48\pi^2)(949 - 96\pi^2)}}.$$

Numerical values for these asymptotic correlations for $t = 0, \dots, 10$ are listed in Table 1.

t	$\lim \text{Cor}(C_n, B_n)$
0	-0.86404
1	-0.93766
2	-0.96482
3	-0.97754
4	-0.98446
5	-0.98862
6	-0.99131
7	-0.99316
8	-0.99447
9	-0.99544
10	-0.99618

Table 1 Asymptotic correlations for the number of key comparisons and key exchanges of median-of- $(2t + 1)$ Quicksort.

A univariate cost measure

We consider now the univariate cost measures $W_n := C_n + wB_n$, where $w > 0$ is a weight on the key exchange. By linearity of expectation the mean of W_n follows from (29) and (30). The variance of W_n is obtained from the variances of C_n and B_n and their covariance given in Corollary 5.4. The limit law for W_n follows from the bivariate limit law in Theorem 5.3:

$$\frac{W_n - \mathbb{E}W_n}{n} \xrightarrow{\ell_2} W,$$

where W is the unique distributional solution in $\mathcal{M}_{0,2}^2$ of

$$W \stackrel{d}{=} VW + (1 - V)W^* + 1 + wV(1 - V) + (\mu_c^{(t)} + w\mu_b^{(t)})\mathcal{E}(V), \quad (35)$$

with W, W^*, V independent, $W \sim W^*$, and $V \sim \text{beta}(t+1, t+1)$. This limit law is obtained by applying the map $T_w : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto x + wy$ to the normalized version of (C_n, B_n) in Theorem 5.3.

The asymptotic mean of C_n is decreasing for $t \rightarrow \infty$ whereas the mean of B_n is increasing as $t \rightarrow \infty$. For the identification of the optimal $t \in \mathbb{N}_0$ such that the leading constant $\tau(t, w)$ in $\mathbb{E} W_n = (1 + o(1))\tau(t, w)n \ln(n)$ is minimized for given $w > 0$ see Martínez and Roura [33], who also discuss the more delicate problem of allowing t to be dependent on n .

Sedgewick's Quicksort programs

The analysis of the weighted mixtures W_n can be extended to a first order analysis of the Quicksort versions Sedgewick [50] analyzed; see also [49, program 2.4, program 8.2]. These programs partition up to subfiles of a size $M \in \mathbb{N}$ and sort the remaining subfiles by a final scan of insertion sort. Program 2.4 uses a uniform pivot element ($t = 0$), whereas program 8.2 is the corresponding median-of-3 version. The cost for an assembly language implementation of program 2.4 (for the median-of-3 version a slightly modified mixture being used) is measured by

$$24A_n + 11B_n + 4C_n + 3D_n + 8E_n + 9S_n + 7n, \quad (36)$$

where A_n denotes the number of partitioning stages, B_n the number of key exchanges during the partitioning, C_n the number of comparisons during partitioning, S_n the number of stack pushes, D_n the number of insertions made by the final insertion sort, and E_n the number of keys moved by insertion sort. The computation of the exact mean for $t = 0, 1$ led among other interpretations to the optimal choice of $M = 9$ for both programs. As an indication for the order of the standard deviation the standard deviation for C_n in the case $t = 0$ was derived to be $0.68n$. From the previous discussion we obtain the exact order $\sigma_t n$ of the standard deviation of the costs of the programs as well as the limit law:

Corollary 5.5 *Let \widetilde{W}_n denote the costs of Sedgewick's quicksort programs with median-of- $(2t+1)$ selection of the pivot, $t = 0, 1$. Then we have*

$$\begin{aligned} \text{Var}(\widetilde{W}_n) &= (1 + o(1))\sigma_t^2 n^2, \\ \frac{\widetilde{W}_n - \mathbb{E} \widetilde{W}_n}{n} &\xrightarrow{\ell_2} W, \end{aligned}$$

where W is the fixed-point in (35) with $w = 11/4$ and $V \sim \text{beta}(t+1, t+1)$. We have

$$\begin{aligned} \sigma_0 &= \sqrt{\frac{13771}{960} - \frac{1225}{864}\pi^2} = 0.592814\dots, \\ \sigma_1 &= \sqrt{\frac{917309}{78400} - \frac{2883}{2450}\pi^2} = 0.294047\dots \end{aligned}$$

Proof: A sequence (Y_n) which satisfies a quicksort recurrence as (23), (24) with a (random) toll function of the order $o_{L_2}(n)$ and which has a mean of the form $cn + o(n)$ with $c \geq 0$ is seen by the contraction method to satisfy $(Y_n - \mathbb{E} Y_n)/n \rightarrow 0$ in ℓ_2 . (The order of normalization for a non-degenerate limit being \sqrt{n} .) The quantities A_n, D_n, E_n, S_n in (36) follow this pattern. Therefore, the only asymptotic contributions come from C_n and B_n , and the previous discussion implies the assertions. ■

6 Ideal and strengthened contraction condition

In this section the ideal contraction condition (7) and its strengthened version (12) are compared for a concrete parametric family of recurrences. It will turn out that the strengthened condition (12) is violated for (infinitely many) particular choices of the parameters whereas the ideal condition is always satisfied. This example is worked out in detail in order to provide a bench mark for further improvements of the strengthened condition towards the ideal contraction condition.

Let $d' \geq 2$ be an integer. We use the convention that we identify indices $r \in \{0, \dots, 2^{d'} - 1\}$ with their binary representation $(r_1, \dots, r_{d'}) \in \{0, 1\}^{d'}$, i.e.,

$$r = \sum_{i=1}^{d'} r_i 2^{d'-i}, \quad r_i \in \{0, 1\}.$$

For a vector $u \in [0, 1]^{d'}$ we define for $r \in \{0, \dots, 2^{d'} - 1\}$

$$\langle u \rangle_r := \prod_{\substack{1 \leq i \leq d' \\ r_i=0}} u_i \prod_{\substack{1 \leq i \leq d' \\ r_i=1}} (1 - u_i)$$

and $\langle u \rangle := (\langle u \rangle_0, \dots, \langle u \rangle_{2^{d'}-1})$. Geometrically, $\langle u \rangle$ is the vector of volumes which are generated by subdividing the unit cube $[0, 1]^{d'}$ by the hyperplanes perpendicular to the axis through u . Let $S_1, \dots, S_d \subset \{1, \dots, d'\}$ with $1 \leq |S_k| \leq d' - 1$ for $k = 1, \dots, d$, furthermore U be a unif $[0, 1]^{d'}$ distributed random variable and $\{Q_i^{(k)} : k = 1, \dots, d, i = 1, \dots, d'\}$ be a set of independent unif $[0, 1]$ distributed random variables, which is also independent of U , and

$$\mathbf{1}_r^{S_k}(U, Q^{(k)}) := \prod_{\substack{i \in S_k \\ r_i=0}} \mathbf{1}_{\{Q_i^{(k)} < U_i\}} \prod_{\substack{i \in S_k \\ r_i=1}} \mathbf{1}_{\{Q_i^{(k)} \geq U_i\}}, \quad (37)$$

for $r = 0, \dots, 2^{d'} - 1$ and $k = 1, \dots, d$. Then we define the diagonal matrices

$$A_r := \text{diag} \left(\mathbf{1}_r^{S_1}(U, Q^{(1)}), \dots, \mathbf{1}_r^{S_d}(U, Q^{(d)}) \right)$$

for $r = 0, \dots, 2^{d'} - 1$, and $b := (1, 1, \dots, 1)^t$. With these definitions we setup a recurrence relation as (1) by $Y_0 := 0$ and

$$Y_n \stackrel{\mathcal{D}}{=} \sum_{r=0}^{2^{d'}-1} A_r Y_{I_r^{(n)}} + b, \quad n \geq 1. \quad (38)$$

According to (1) the sequences $(Y_n^{(0)}), \dots, (Y_n^{(2^{d'}-1)})$ are independent copies of (Y_n) being independent of $(A_0, \dots, A_{2^{d'}-1}, I^{(n)})$. The random vector $I^{(n)}$ is defined to be multinomially distributed given U ,

$$\mathbb{P}^{I^{(n)}|U=u} = M(n-1, \langle u \rangle).$$

This implies in particular

$$\frac{I^{(n)}}{n} \xrightarrow{\mathbb{P}} \langle U \rangle = (\langle U \rangle_0, \dots, \langle U \rangle_{2^{d'}-1}). \quad (39)$$

For recurrence (38) we check the conditions of Theorem 4.1 for various choices of d', d, S_1, \dots, S_d . This leads in two ways to examples where the strengthened contraction condition of Theorem 4.1 is violated: firstly, by taking $d = 2$, increasing d' and suitably choosing S_1 and S_2 or secondly by increasing d . Although in some situations the refinement in Corollary 4.2 will save.

Taking means in (38) leads in each component to the recursion for the mean of the number of nodes visited during a partial match query in a d' -dimensional quadtree in the uniform probabilistic model, the set S_k giving the coordinates which are specified in the query. For background on random quadtrees and partial match see [17, 14, 15, 13, 8, 47, 48, 27, 20, 38].

Note that recursion (38) with independent $(Y_n^{(0)}), \dots, (Y_n^{(2^{d'}-1)})$ does model the mean but does not model the distribution of the cost of a partial match in the uniform probabilistic model of Flajolet and Puech [17], since positive correlations of the costs for the subtrees of the root searched simultaneously are to be expected. This effect becomes relevant as soon as variances or limits laws are considered and is present as well in K - d trees, its variants, and in tries under the asymmetric Bernoulli model. Whether the correlation of the cost for searching the subtrees remains small enough, such that the order of variances and limit laws for the idealized model in (38) carry over to the uniform probabilistic model, or not, is open. Such an idealization was in connection with tries also called the model of “adapted queries” and was tacitly assumed in the analysis of K - d trees in [35] and elsewhere. From this perspective, (38) may be considered as an idealized model for d independent partial match queries in a common random quadtree, but we will only look at (38) to test the conditions of Theorem 4.1.

Flajolet, Gonnet, Puech, and Robson [14] proved

$$\mathbb{E} Y_{nk} = (1 + o(1)) \gamma_{s_k, d'} n^{\alpha_k}, \quad (40)$$

with $s_k = |S_k|$, a constant $\gamma_{s_k, d'} > 0$, and α_k given as the in (1, 2) unique solution of

$$\alpha_k^{d' - s_k} (\alpha_k + 1)^{s_k} = 2^{d'}. \quad (41)$$

For $d = 2$ we write $\alpha := \alpha_1$, $\beta := \alpha_2$ and assume that $S_1 = \{1, \dots, q + s\}$, $S_2 = \{1, \dots, q, q + s + 1, \dots, q + s + t\}$ with $0 \leq q, s, t \leq d' - 1$, $1 \leq q + s \leq q + t \leq d' - 1$ and $q + s + t \leq d'$. We define

$$D_n := \text{diag} \left(n^{\alpha-1}, n^{\beta-1} \right)$$

and scale according to (2). Then the quantities $A_r^{(n)}$, $b^{(n)}$ in (4) are given by

$$A_r^{(n)} = \text{diag} \left(\mathbf{1}_r^{S_1}(U, Q^{(1)}) \left(\frac{I_r^{(n)}}{n} \right)^{\alpha-1}, \mathbf{1}_r^{S_2}(U, Q^{(2)}) \left(\frac{I_r^{(n)}}{n} \right)^{\beta-1} \right)$$

for $r = 0, \dots, 2^{d'} - 1$ and

$$b^{(n)} = \left(\left(\sum_{r=0}^{2^{d'}-1} \mathbf{1}_r^{S_1}(U, Q^{(1)}) \left(\frac{I_r^{(n)}}{n} \right)^{\alpha-1} - 1 \right) \gamma_{q+s, d'} + o(1), \right. \\ \left. \left(\sum_{r=0}^{2^{d'}-1} \mathbf{1}_r^{S_2}(U, Q^{(2)}) \left(\frac{I_r^{(n)}}{n} \right)^{\beta-1} - 1 \right) \gamma_{q+t, d'} + o(1) \right)^t,$$

where $\gamma_{q+s,d'}$ and $\gamma_{q+t,d'}$ are the constants in (40). By (39) the L_2 convergence in (11) is satisfied for

$$A_r^* := \text{diag} \left(\mathbf{1}_r^{S_1}(U, Q^{(1)}) \langle U \rangle_r^{\alpha-1}, \mathbf{1}_r^{S_2}(U, Q^{(2)}) \langle U \rangle_r^{\beta-1} \right) \quad (42)$$

for $r = 0, \dots, 2^{d'} - 1$ and

$$b^* = \left(\left(\sum_{r=0}^{2^{d'}-1} \mathbf{1}_r^{S_1}(U, Q^{(1)}) \langle U \rangle_r^{\alpha-1} - 1 \right) \gamma_{q+s,d'}, \right. \\ \left. \left(\sum_{r=0}^{2^{d'}-1} \mathbf{1}_r^{S_2}(U, Q^{(2)}) \langle U \rangle_r^{\beta-1} - 1 \right) \gamma_{q+t,d'} \right)^t. \quad (43)$$

Lemma 6.1 *The conditions of Theorem 4.1 are satisfied if*

$$\left(\frac{1}{\beta} \right)^{q+t} \left(\frac{2}{2\beta-1} \right)^{d'-q-t} + \left(\frac{1}{\alpha} \right)^{q+s} \left(\frac{2}{2\alpha-1} \right)^{d'-q-s} \\ - \left(\frac{2}{2\alpha+1} \right)^q \left(\frac{1}{\alpha} \right)^{s+t} \left(\frac{2}{2\alpha-1} \right)^{d'-q-s-t} < 1. \quad (44)$$

Proof: We apply Theorem 4.1 to our situation. Already checked by our preceding considerations is the L_2 convergence of $(A_0^{(n)}, \dots, A_{2^{d'}-1}^{(n)}, b^{(n)})$ to $(A_0^*, \dots, A_{2^{d'}-1}^*, b^*)$. For the proof of (13) note that $\|(A_k^{(n)})^t A_k^{(n)}\|_{\text{op}}$ remains bounded, that $\sigma := \mathcal{L}(\langle U \rangle_k)$ is Lebesgue-continuous and has in particular no point-mass on zero. Using dominated convergence we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\{I_k^{(n)} \leq l\} \cup \{I_k^{(n)} = n\}) \\ = \int_0^1 \lim_{n \rightarrow \infty} B(n-1, x)(\{0, \dots, l, n\}) d\sigma(x) \\ = 0$$

for all $l \in \mathbb{N}$, i.e., (13) is satisfied. Thus, it remains to show that the condition (12),

$$\sum_{r=0}^{2^{d'}-1} \mathbb{E} \|(A_r^*)^t A_r^*\|_{\text{op}} < 1,$$

is satisfied. First of all, we observe that

$$(A_r^*)^t A_r^* = \text{diag} \left(\mathbf{1}_r^{S_1}(U, Q^{(1)}) \langle U \rangle_r^{2\alpha-2}, \mathbf{1}_r^{S_2}(U, Q^{(2)}) \langle U \rangle_r^{2\beta-2} \right)$$

are identically distributed for $r = 0, \dots, 2^{d'} - 1$. Clearly the spectral radius of these matrices is the maximum of the diagonal entries. Furthermore note, that $q + s \leq q + t$ implies $\beta \leq \alpha$ and

$U_i^{2\alpha-2} \leq U_i^{2\beta-2}$ for the occurring U_i . This implies

$$\begin{aligned}
& \sum_{r=0}^{2^{d'}-1} \mathbb{E} \|(A_r^*)^t A_r^*\|_{\text{op}} \\
&= 2^{d'} \mathbb{E} \max \left\{ \mathbf{1}_0^{S_1}(U, Q^{(1)}) \langle U \rangle_0^{2\alpha-2}, \mathbf{1}_0^{S_2}(U, Q^{(2)}) \langle U \rangle_0^{2\beta-2} \right\} \\
&= 2^{d'} \mathbb{E} \left[\mathbf{1}_0^{S_2}(U, Q^{(2)}) \langle U \rangle_0^{2\beta-2} + (1 - \mathbf{1}_0^{S_2}(U, Q^{(2)})) \mathbf{1}_0^{S_1}(U, Q^{(1)}) \langle U \rangle_0^{2\alpha-2} \right] \\
&= 2^{d'} \left(\left(\frac{1}{2\beta} \right)^{q+t} \left(\frac{1}{2\beta-1} \right)^{d'-q-t} + \left(\frac{1}{2\alpha} \right)^{q+s} \left(\frac{1}{2\alpha-1} \right)^{d'-q-s} \right. \\
&\quad \left. - \left(\frac{1}{2\alpha+1} \right)^q \left(\frac{1}{2\alpha} \right)^{s+t} \left(\frac{1}{2\alpha-1} \right)^{d'-q-s-t} \right) \\
&< 1,
\end{aligned}$$

by (44). ■

It is easy to see that the ideal contraction condition is satisfied for all choices of d', q, s, t . The strengthened conditions (44) is violated for the choices $(d', q, s, t) = (6, 4, 1, 1), (7, 6, 0, 0), (7, 5, 1, 1), (8, 7, 0, 0), (8, 6, 1, 1) \dots$ Up to dimension $d' = 10$ there are 460 qualitatively different cases from which 9 are not covered by the strengthened condition. Interestingly, all these cases are of a special type which can be covered by the alternative contraction condition (22), as it follows from the second part of the next Lemma. However, we have no proof that this holds for arbitrarily large d' .

Now we turn to the case of general $d \geq 2$. Increasing d would easily violated (12) if the sizes $|S_k|$ remain large, e.g., $|S_k| > 0.57d'$ for all k (as seen from the first part of the following proof). Therefore we give two special cases, where we can still for arbitrarily large d get a limit law. For this we scale as in the case $d = 2$ and get the limiting matrices

$$A_r^* := \text{diag} \left(\mathbf{1}_r^{S_1}(U, Q^{(1)}) \langle U \rangle_r^{\alpha_1-1}, \dots, \mathbf{1}_r^{S_d}(U, Q^{(d)}) \langle U \rangle_r^{\alpha_d-1} \right) \quad (45)$$

for $r = 0, \dots, 2^{d'} - 1$, where the α_k are the solutions of the corresponding indicial equations, and the random toll vector

$$b^* = \left(\left(\sum_{r=0}^{2^{d'}-1} \mathbf{1}_r^{S_1}(U, Q^{(1)}) \langle U \rangle_r^{\alpha_1-1} - 1 \right) \gamma_{s_1, d'}, \dots, \right. \quad (46)$$

$$\left. \left(\sum_{r=0}^{2^{d'}-1} \mathbf{1}_r^{S_d}(U, Q^{(d)}) \langle U \rangle_r^{\alpha_d-1} - 1 \right) \gamma_{s_d, d'} \right)^t. \quad (47)$$

Lemma 6.2 *The conditions of Theorem 4.1 are satisfied if*

$$\alpha_k > \frac{3}{2} \quad \left(\text{or equivalently } \frac{s_k}{d'} < \frac{\ln(4/3)}{\ln(5/3)} = 0.56317079 \dots \right) \quad (48)$$

for all $k = 1, \dots, d$ or

$$\alpha_1 = \dots = \alpha_d \quad (\text{or equivalently } s_1 = \dots = s_d). \quad (49)$$

Proof: First we assume that (48) is given. In this case we apply Theorem 4.1 as in the proof of Theorem 6.1. It remains to prove

$$\sum_{r=0}^{2^{d'}-1} \mathbb{E} \left\| (A_r^*)^t A_r^* \right\|_{\text{op}} < 1.$$

Again, $(A_r^*)^t A_r^*$ are identically distributed and with $\alpha := \min_{1 \leq k \leq d} \alpha_k$ we deduce

$$\begin{aligned} \sum_{r=0}^{2^{d'}-1} \mathbb{E} \left\| (A_r^*)^t A_r^* \right\|_{\text{op}} &= 2^{d'} \mathbb{E} \max_{1 \leq k \leq d} \left\{ \mathbf{1}_0^{S_k}(U, Q^{(k)}) \langle U \rangle_0^{2\alpha_k-2} \right\} \\ &\leq 2^{d'} \mathbb{E} \left[\max_{1 \leq k \leq d} \left\{ \mathbf{1}_0^{S_k}(U, Q^{(k)}) \right\} \langle U \rangle_0^{2\alpha-2} \right] \\ &\leq 2^{d'} \mathbb{E} \langle U \rangle_0^{2\alpha-2} \\ &= 2^{d'} \left(\frac{1}{2\alpha-1} \right)^{d'} \\ &< 1, \end{aligned}$$

if $2/(2\alpha-1) < 1$ which means $\alpha > 3/2$. Therefore, $\alpha_k > 3/2$ for $k = 1, \dots, d$ implies (12). This proves the assertion of the theorem under condition (48).

For the second part we assume that $\alpha_k =: \alpha$ for $k = 1, \dots, d$ is given. We apply Corollary 4.2. Note, that condition (21) can be derived similarly to condition (13) in the proof of Theorem 6.1. Hence, it remains to show (22). By conditioning on U we derive

$$\begin{aligned} \mathbb{E} \sum_{r=0}^{2^{d'}-1} \mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{kk}^2 & \tag{50} \\ &= \mathbb{E} \sum_{r=0}^{2^{d'}-1} \mathbf{1}_{\{I_r^{(n)}=i\}} \mathbf{1}_r^{S_k}(U, Q^{(k)}) \left(\frac{i}{n} \right)^{2\alpha-2} \\ &= \left(\frac{i}{n} \right)^{2\alpha-2} \sum_{r=0}^{2^{d'}-1} \int_{[0,1]^{d'}} \binom{n}{i} \langle u \rangle_r^i (1 - \langle u \rangle_r)^{n-i} \prod_{\substack{l \in S_k \\ r_l=0}} u_l \prod_{\substack{l \in S_k \\ r_l=1}} (1 - u_l) du \\ &= \binom{n}{i} \left(\frac{i}{n} \right)^{2\alpha-2} \int_{[0,1]^{d'}} \sum_{r=0}^{2^{d'}-1} \langle u \rangle_r^i (1 - \langle u \rangle_r)^{n-i} \prod_{\substack{l \in S_k \\ r_l=0}} u_l \prod_{\substack{l \in S_k \\ r_l=1}} (1 - u_l) du. \end{aligned}$$

The integral only depends on the cardinality of S_k , thus the expectation in (50) is independent of k . Therefore, in (22) we may remove the maximum and derive

$$\limsup_{n \rightarrow \infty} \sum_{i=0}^n \mathbb{E} \sum_{r=0}^{2^{d'}-1} \mathbf{1}_{\{I_r^{(n)}=i\}} \left(A_r^{(n)} \right)_{11}^2 = \limsup_{n \rightarrow \infty} \mathbb{E} \sum_{r=0}^{2^{d'}-1} \left(A_r^{(n)} \right)_{11}^2 = \mathbb{E} \sum_{r=0}^{2^{d'}-1} (A_r^*)_{11}^2.$$

This is the contraction factor for the one-dimensional problem which was shown to be less than 1 in [38]. ■

7 Conclusion

We proposed an extension of the contraction method with respect to the minimal L_2 -metric ℓ_2 . The contraction condition (7) for the limiting map which determines the limit distribution for the limit law is

$$\left\| \sum_{r=1}^K \mathbb{E} [A_r^t A_r] \right\|_{\text{op}} < 1.$$

However, in order to derive the limit law we had to strengthen this condition to , e.g., (12):

$$\sum_{r=1}^K \mathbb{E} \|A_r^t A_r\|_{\text{op}} < 1.$$

This is contrary to the one-dimensional case $d = 1$ where the contraction condition for the limiting map also implies convergence in the ℓ_2 -metric. Note, that in the one-dimensional case (7) and (12) are equivalent. In our approach it seems to be crucial to find an appropriate estimate for the term (17). Since we have no example where (7) is satisfied and no convergence in ℓ_2 holds it remains open whether (7) is sufficient to imply the limit law or not.

Acknowledgment

I thank my advisor Ludger Rüschemdorf for guiding and encouraging me over years to study multivariate branching type recursions. Thanks also to Luc Devroye and Hsien-Kuei Hwang for helpful discussion on the applications, and a referee for a thorough discussion of the paper.

References

- [1] Bender, E. A., and L. B. Richmond (1983). Central and local limit theorems applied to asymptotic enumeration. II. Multivariate generating functions. *J. Combin. Theory Ser. A* 34, 255–265.
- [2] Bickel, P. J. and P. A. Freedman (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196–1217.
- [3] 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.
- [4] Burton, R. and U. Rösler (1995). An L_2 convergence theorem for random affine mappings. *J. Appl. Probab.* 32, 183–192.
- [5] Chern, H.-H. and H.-K. Hwang (2001). Transitional behaviors of the average cost of Quicksort with median-of- $(2t + 1)$. *Algorithmica* 29, 44–69.
- [6] Cramer, M. and L. Rüschemdorf (2001). Convergence of two-dimensional branching recursions. *J. Comput. Appl. Math.* 130, 53–73.
- [7] Dall’Aglio, G. (1956). Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuola Norm. Sup. Pisa (3)* 10, 35–74.

- [8] Devroye, L. (1998). Branching processes and their applications in the analysis of tree structures and tree algorithms. *Probabilistic Methods for Algorithmic Discrete Mathematics*, 249–314. Springer, Berlin.
- [9] Devroye, L. (1998). Universal limit laws for the depths in random trees. *SIAM J. Comput.* *28*, 409–432.
- [10] Devroye, L. and L. Laforest (1990). An analysis of random d -dimensional quad trees. *SIAM J. Comput.* *19*, 821–832.
- [11] McDiarmid, C. J. H. and R. B. Hayward (1996). Large deviations for Quicksort. *J. Algorithms* *21*, 476–507.
- [12] van Emden, M. H. (1970). Increasing the efficiency of quicksort. *Commun. ACM* *13*, 563–567.
- [13] Finkel, R. and J. Bentley (1974). Quad trees, a data structure for retrieval on composite keys. *Acta Inform.* *4*, 1–9.
- [14] Flajolet, P., G. Gonnet, C. Puech, and J. Robson (1993). Analytic variations on quadtrees. *Algorithmica* *10*, 473–500.
- [15] Flajolet, P., G. Labelle, L. Laforest, and B. Salvy (1995). Hypergeometrics and the cost structure of quadtrees. *Random Structures Algorithms* *7*, 117–144.
- [16] Flajolet, P. and T. Lafforgue (1994). Search costs in quadtrees and singularity perturbation asymptotics. *Discrete Comput. Geom.* *12*, 151–175.
- [17] Flajolet, P. and C. Puech (1986). Partial match retrieval of multidimensional data. *J. ACM* *33*, 371–407.
- [18] Gao, Z. and L. B. Richmond (1992). Central and local limit theorems applied to asymptotic enumeration. IV. Multivariate generating functions. *J. Comput. Appl. Math.* *41*, 177–186.
- [19] Geiger, J. (2000). A new proof of Yaglom’s exponential limit law. *Mathematics and computer science (Versailles, 2000)*, 245–249. Birkhäuser, Basel.
- [20] Gonnet, G. H. and R. Baeza-Yates (1991). *Handbook of Algorithms and Data Structures: in Pascal and C* (2nd ed.). Addison-Wesley, Reading, MA.
- [21] Green, D. H. (1983). *Labelled Formal Languages and their Uses*. Ph.D. Thesis, Stanford University.
- [22] Hennequin, P. (1989). Combinatorial analysis of quicksort algorithm. *RAIRO Inform. Théor. Appl.* *23*, 317–333.
- [23] Hennequin, P. (1991). *Analyse en moyenne d’algorithme, tri rapide et arbres de recherche*. Ph.D. Thesis, L’École Polytechnique.
- [24] Hurwitz, H. Jr. (1971). On the probability distribution of the values of binary trees. *Comm. ACM* *14*, 99–102.
- [25] Hwang, H.-K. and R. Neininger (2001). Phase change of limit laws in the quicksort recurrence under varying toll functions. Preprint.

- [26] Lew, W. and H. M. Mahmoud (1994). The joint distribution of elastic buckets in multiway search trees. *SIAM J. Comput.* 23, 1050–1074.
- [27] Mahmoud, H. M. (1992). *Evolution of Random Search Trees*. John Wiley & Sons Inc., New York.
- [28] Mahmoud, H. M. (2000). *Sorting. A Distribution Theory*. Wiley-Interscience, New York.
- [29] Mahmoud, H. M. and B. Pittel (1988). On the joint distribution of the insertion path length and the number of comparisons in search trees. *Discrete Appl. Math.* 20, 243–251.
- [30] Mahmoud, H. M. and R. T. Smythe (1992). Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms* 3, 255–266.
- [31] Mahmoud, H. M., R. T. Smythe, and J. Szymański (1993). On the structure of random plane-oriented recursive trees and their branches. *Random Structures Algorithms* 4, 151–176.
- [32] Major, P. (1978). On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.* 8, 487–517.
- [33] Martínez, C. and S. Roura (2001). Optimal sampling strategies in Quicksort and Quickselect. *SIAM J. Comput.*, to appear.
- [34] Neininger, R. (1999). *Limit Laws for Random Recursive Structures and Algorithms*. Ph. D. Thesis, Universität Freiburg.
- [35] Neininger, R. (2000). Asymptotic distributions for partial match queries in K - d trees. *Random Structures Algorithms* 17, 403–427.
- [36] Neininger, R. (2001). On the Wiener index in random trees. In preparation.
- [37] Neininger, R. and L. Rüschemdorf (1999). On the internal path length of d -dimensional quad trees. *Random Structures Algorithms* 15, 25–41.
- [38] Neininger, R. and L. Rüschemdorf (2001). Limit laws for partial match queries in quadtrees. *Ann. Appl. Probab.*, to appear.
- [39] Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models*. John Wiley.
- [40] Rachev, S. T. and L. Rüschemdorf (1995). Probability metrics and recursive algorithms. *Adv. in Appl. Probab.* 27, 770–799.
- [41] Rachev, S. T. and L. Rüschemdorf (1998). *Mass Transportation Problems. Vol. I: Theory*. Springer-Verlag, New York.
- [42] Régnier, M. (1989). A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.* 23, 335–343.
- [43] Rösler, U. (1991). A limit theorem for “Quicksort”. *RAIRO Inform. Théor. Appl.* 25, 85–100.
- [44] Rösler, U. (1992). A fixed point theorem for distributions. *Stochastic Process. Appl.* 42, 195–214.
- [45] Rösler, U. (2001). On the analysis of stochastic divide and conquer algorithms. *Algorithmica* 29, 238–261.

- [46] Rösler, U. and L. Rüschemdorf (2001). The contraction method for recursive algorithms. *Algorithmica* 29, 3–33.
- [47] Samet, H. (1990a). *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*. Addison-Wesley, Reading, MA.
- [48] Samet, H. (1990b). *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, Reading, MA.
- [49] Sedgewick, R. (1975). *Quicksort*. Stanford Computer Science Report STAN-CS-75-492, Ph. D. Thesis, Stanford University.
- [50] Sedgewick, R. (1977). The analysis of Quicksort programs. *Acta Informat.* 7, 327–355.
- [51] Tan, K. H. and P. Hadjicostas (1995). Some properties of a limiting distribution in Quicksort. *Statist. Probab. Lett.* 25, 87–94.