Convergence Rates in the Probabilistic Analysis of Algorithms

Ralph Neininger

Institute for Mathematics, Goethe University, 60054 Frankfurt a.M., Germany
neininger@math.uni-frankfurt.de

Jasmin Straub

Institute for Mathematics, Goethe University, 60054 Frankfurt a.M., Germany
jstraub@math.uni-frankfurt.de

Abstract

In this extended abstract a general framework is developed to bound rates of convergence for sequences of random variables as they mainly arise in the analysis of random trees and divide and conquer algorithms. The rates of convergence are bounded in the Zolotarev distances. Concrete examples from the analysis of algorithms and data structures are discussed as well as a few examples from other areas. They lead to convergence rates of polynomial and logarithmic order. A crucial role is played by a factor 3 in the exponent of these orders in cases where the normal distribution is the limit distribution.

2012 ACM Subject Classification Theory of computation → Sorting and searching; Theory of computation → Divide and conquer

Keywords and phrases weak convergence, probabilistic analysis of algorithms, random trees, probability metrics

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

1 Introduction and notation

In this extended abstract we consider a general recurrence for (probability) distributions which covers many instances of complexity measures of divide and conquer algorithms and parameters of random search trees. We consider a sequence $(Y_n)_{n\geq 0}$ of d-dimensional random vectors satisfying the distributional recursion

$$Y_n \stackrel{d}{=} \sum_{r=1}^{K} A_r(n) Y^{(r)}_{I^{(n)}_r} + b_n, \qquad n \geq n_0, \tag{1}$$

where $(A_1(n),\dots,A_K(n),b_n,I^{(n)}),(Y^{(1)}_n)_{n\geq 0},\dots,(Y^{(K)}_n)_{n\geq 0}$ are independent, the coefficients $A_1(n),\dots,A_K(n)$ are random $(d\times d)$-matrices, b_n is a d-dimensional random vector, $I^{(n)} = (I^{(n)}_1,\dots,I^{(n)}_K)$ is a random vector in $\{0,\dots,n\}^K$, $n_0 \geq 1$ and $(Y^{(r)}_n)_{n\geq 0} \stackrel{d}{=} (Y_n)_{n\geq 0}$ for $r = 1,\dots,K$. Moreover, $K \geq 1$ is a fixed integer, but extensions to K being random and depending on n are possible.

This is the framework of [14] where some general convergence results are shown for appropriate normalizations of the Y_n. The content of the present extended abstract is to also study the rates of convergence in such limit theorems.

We define the normalized sequence $(X_n)_{n\geq 0}$ by

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

where M_n is a d-dimensional vector and C_n a positive definite $(d\times d)$-matrix. Essentially, we choose M_n as the mean and C_n as the covariance matrix of Y_n if they exist or as the

42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:18
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

leading order terms in expansions of these moments as $n \to \infty$. The normalized quantities satisfy the following modified recursion:

$$X_n \stackrel{d}{=} \sum_{r=1}^{K} A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \qquad n \ge n_0, \tag{2}$$

with

$$A_r^{(n)} := C_n^{-1/2} A_r(n) C_{I_r^{(n)}}^{1/2}, \quad b^{(n)} := C_n^{-1/2}\left(b_n - M_n + \sum_{r=1}^{K} A_r(n) M_{I_r^{(n)}}\right) \tag{3}$$

and independence relations as in (1).

In the context of the contraction method the aim is to establish transfer theorems of the following form: After verifying the assumptions of appropriate convergence of the coefficients $A_r^{(n)} \to A_r^*, b^{(n)} \to b^*$ then convergence in distribution of random vectors (X_n) to a limit X is implied. The limit distribution $\mathcal{L}(X)$ is identified by a fixed-point equation obtained from (2) by considering formally $n \to \infty$:

$$X \stackrel{d}{=} \sum_{r=1}^{K} A_r^* X^{(r)} + b^*. \tag{4}$$

Here $(A_1^*, \ldots, A_K^*, b^*), X^{(1)}, \ldots, X^{(K)}$ are independent and $X^{(r)} \stackrel{d}{=} X$ for $r = 1, \ldots, K$.

The aim of the present extended abstract is to endow such general transfer theorems with bounds on the rates of convergence. As a distance measure between (probability) distributions we use the Zolotarev metric. For various of the applications we discuss, bounds on the rate of convergence have been derived one by one for more popular distance measures such as the Kolmogorov–Smirnov distance. The transfer theorems of the present paper are in terms of the smoother Zolotarev metrics. However, they are easy to apply and cover a broad range of applications at once. A crucial role is played by a factor 3 in the exponent of these orders in cases where the normal distribution is the limit distribution, see Remark 4.

In the rest of this section we fix some notation. Regarding norms of vectors and (random) matrices we denote for $x \in \mathbb{R}^d$ by $\|x\|$ its Euclidean norm and for a random vector X and some $0 < p < \infty$, we set $\|X\|_p := \mathbb{E}[\|X\|^p]^{(1/p)\wedge 1}$. Furthermore, for a $(d \times d)$-matrix A, $\|A\|_{\text{op}} := \sup_{\|x\|=1} \|Ax\|$ denotes the spectral norm of A and for a random such A we define $\|A\|_p := \mathbb{E}[\|A\|_{\text{op}}^p]^{(1/p)\wedge 1}$ for a random square matrix and $0 < p < \infty$. Note that for a symmetric $(d \times d)$-matrix A, we have $\|A\|_{\text{op}} = \max\{|\lambda| : \lambda \text{ eigenvalue of } A\}$. By Id_d the d-dimensional unit matrix is denoted. For multilinear forms the norm is defined similarly.

Furthermore we define by $\mathcal{P}^d$ the space of probability distributions in $\mathbb{R}^d$ (endowed with the Borel σ-field), by $\mathcal{P}_s^d := \{\mathcal{L}(X) \in \mathcal{P}^d : \|X\|_s < \infty\}$ and for a vector $m \in \mathbb{R}^d$, and a symmetric positive semidefinite $d \times d$ matrix C the spaces

$$\mathcal{P}_s^d(m) := \{\mathcal{L}(X) \in \mathcal{P}_s^d : \mathbb{E}[X] = m\}, \quad s > 1, \tag{5}$$
$$\mathcal{P}_s^d(m, C) := \{\mathcal{L}(X) \in \mathcal{P}_s^d : \mathbb{E}[X] = m, \text{Cov}(X) = C\}, \quad s > 2.$$

We use the convention $\mathcal{P}_s^d(m) := \mathcal{P}_s^d$ for $s \le 1$ and $\mathcal{P}_s^d(m, C) := \mathcal{P}_s^d(m)$ for $s \le 2$.

The Zolotarev metrics ζ_s, [19], are defined for probability distributions $\mathcal{L}(X), \mathcal{L}(Y) \in \mathcal{P}^d$ by

$$\zeta_s(X, Y) := \zeta_s(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{f \in \mathcal{F}_s} |E(f(X) - f(Y))| \tag{6}$$

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

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

Note that these distance measures may be infinite. Finite metrics are given by ζ_s on $\mathcal{P}_s^d$ for $0 \leq s \leq 1$, by ζ_s on $\mathcal{P}_s^d(m)$ for $1 < s \leq 2$, and by ζ_s on $\mathcal{P}_s^d(m, C)$ for $2 < s \leq 3$, cf. (5).

2 Results

We return to the situation outlined in the introduction, where we have normalized $(Y_n)_{n \geq 0}$ in the following way:

$$X_n := C_n^{-1/2}(Y_n - M_n), \qquad n \geq 0, \tag{7}$$

where M_n is a d-dimensional random vector and C_n a positive definite $(d \times d)$-matrix. As recalled in Section 1, for $s > 1$, we may fix the mean and covariance matrix of the scaled quantities to guarantee the finiteness of the ζ_s-metric. Therefore, we choose $M_n = \mathbb{E}[Y_n]$ for $n \geq 0$ and $s > 1$. For $s > 2$, we additionally have to control the covariances of X_n. We assume that there exists an $n_1 \geq 0$ such that $\mathrm{Cov}(Y_n)$ is positive definite for $n \geq n_1$ and choose $C_n = \mathrm{Cov}(Y_n)$ for $n \geq n_1$ and $C_n = \mathrm{Id}_d$ for $n < n_1$. For $s \leq 2$, we just assume that C_n is positive definite and set $n_1 = 0$ in this case.

The normalized quantities satisfy the modified recursion

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

with $A_r^{(n)}$ and $b^{(n)}$ given in (3). The following theorem discusses a general framework to bound rates of convergence for the sequence $(X_n)_{n \geq 0}$. For the proof, we need some technical conditions which guarantee that the sizes $I_r^{(n)}$ of the subproblems grow with n. More precisely, we will assume that there exists some monotonically decreasing sequence $R(n) > 0$ with $R(n) \to 0$ such that

$$\left\|\mathbf{1}_{\{I_r^{(n)} < \ell\}} A_r^{(n)}\right\|_s = \mathrm{O}(R(n)), \quad n \to \infty, \tag{8}$$

for all $\ell \in \mathbb{N}$ and $r = 1, \ldots, K$ and that

$$\left\|\mathbf{1}_{\{I_r^{(n)} = n\}} A_r^{(n)}\right\|_s \to 0, \quad n \to \infty, \tag{9}$$

for all $r = 1, \ldots, K$.

2.1 A general transfer theorem for rates of convergence

Our first result is a direct extension of the main Theorem 4.1 in [14], where we essentially only make all the estimates there explicit. The main result of the present extended abstract in contained in the subsequent subsection.

▶ **Theorem 1.** *Let $(X_n)_{n \geq 0}$ be s-integrable, $0 < s \leq 3$, and satisfy recurrence (7) with the choices for M_n and C_n specified there. We assume that there exist s-integrable $A_1^*, \ldots, A_K^*, b^*$ and some monotonically decreasing sequence $R(n) > 0$ with $R(n) \to 0$ such that, as $n \to \infty$,*

$$\left\|b^{(n)} - b^*\right\|_s + \sum_{r=1}^{K} \left\|A_r^{(n)} - A_r^*\right\|_s = \mathrm{O}(R(n)). \tag{10}$$

If conditions (8) and (9) are satisfied and if

$$\limsup_{n\to\infty} \mathbb{E}\sum_{r=1}^{K}\left(\frac{R(I_r^{(n)})}{R(n)}\|A_r^{(n)}\|_{\mathrm{op}}^{s}\right) < 1, \tag{11}$$

then we have, as $n \to \infty$,

$$\zeta_s(X_n, X) = \mathrm{O}(R(n)),$$

where $\mathcal{L}(X)$ is given as the unique fixed point in $\mathcal{P}_s^d(0, \mathrm{Id}_d)$ of the equation

$$X \stackrel{d}{=} \sum_{r=1}^{K} A_r^* X^{(r)} + b^*, \tag{12}$$

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

▶ Remark 2. In applications, the convergence rate of the coefficients (conditions (8) and (10)) is often faster than the convergence rate of the quantities X_n, see, e.g., Section 4.4. In these cases, it is often possible to perform the induction step in the proof of Theorem 1 although condition (11) does not hold. To be more precise, we may assume

$$\left\|\mathbf{1}_{\{I_r^{(n)}<\ell\}}A_r^{(n)}\right\|_s + \left\|b^{(n)} - b^*\right\|_s + \left\|A_r^{(n)} - A_r^*\right\|_s = \mathrm{O}(\widetilde{R}(n))$$

for every $\ell \geq 0$, $r = 1, \dots, K$ and $n \to \infty$. Then, instead of condition (11), it is sufficient to find some $K > 0$ such that

$$\mathbb{E}\left[\sum_{r=1}^{K}\mathbf{1}_{\{n_1\leq I_r^{(n)}<n\}}\frac{R(I_r^{(n)})}{R(n)}\|A_r^{(n)}\|_{\mathrm{op}}^{s}\right] \leq 1 - p_n - \frac{\widetilde{R}(n)}{KR(n)} \tag{13}$$

for all large n with $p_n := \mathbb{E}\left[\sum_{r=1}^{K}\mathbf{1}_{\{I_r^{(n)}=n\}}\|A_r^{(n)}\|_{\mathrm{op}}^{s}\right]$.

2.2 An improved transfer theorem for normal limit distributions

We now consider the special case where the sequence $(X_n)_{n\geq 0}$ is 3-integrable and satisfies recursion (2) with $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L_3} (A_1^*, \dots, A_K^*, b^*)$ for some 3-integrable coefficients $A_1^*, \dots, A_K^*, b^*$ with

$$b^* = 0, \quad \sum_{r=1}^{K} A_r^*(A_r^*)^T = \mathrm{Id}_d$$

almost surely. Corollary 3.4 in [14] implies that, if $\mathbb{E}[\sum_{r=1}^{K}\|A_r^*\|_{\mathrm{op}}^3] < 1$, equation (12) has a unique solution in the space $\mathcal{P}_3^d(0, \mathrm{Id}_d)$. Furthermore, e.g., using characteristic functions, it is easily checked that this unique solution is the standard normal distribution $\mathcal{N}(0, \mathrm{Id}_d)$.

In this special case of normal limit laws, it is possible to derive a refined version of Theorem 1. Instead of the technical condition (8), we now need the weaker condition

$$\left\|\mathbf{1}_{\{I_r^{(n)}<\ell\}}A_r^{(n)}\right\|_3^3 = \mathrm{O}(R(n)), \quad n \to \infty, \tag{14}$$

for all $\ell \in \mathbb{N}$ and $r = 1, \dots, K$. Moreover, condition (10) concerning the convergence rates of the coefficients can be weakened, which is formulated in the following theorem.

▶ **Theorem 3.** *Let $(X_n)_{n\geq 0}$ be given as in (7) and be 3-integrable. We assume that for some $R(n) > 0$ monotonically decreasing with $R(n) \to 0$ as $n \to \infty$ we have*

$$\Big\| \sum_{r=1}^{K} A_r^{(n)} (A_r^{(n)})^T - \mathrm{Id}_d \Big\|_{3/2}^{3/2} + \big\| b^{(n)} \big\|_3^3 = \mathrm{O}(R(n)), \tag{15}$$

and the technical conditions (9) and (14) being satisfied for $s = 3$. If

$$\limsup_{n\to\infty} \mathbb{E} \sum_{r=1}^{K} \left(\frac{R(I_r^{(n)})}{R(n)} \big\| A_r^{(n)} \big\|_{\mathrm{op}}^3 \right) < 1, \tag{16}$$

then we have, as $n \to \infty$,

$$\zeta_3(X_n, \mathcal{N}(0, \mathrm{Id}_d)) = \mathrm{O}(R(n)).$$

Proof. (Sketch) We define an accompanying sequence $(Z_n^*)_{n\geq 0}$ by

$$Z_n^* := \sum_{r=1}^{K} A_r^{(n)} T_{I_r^{(n)}} N^{(r)} + b^{(n)}, \qquad n \geq 0,$$

where $(A_1^{(n)}, \dots, A_K^{(n)}, I^{(n)}, b^{(n)}), N^{(1)}, \dots, N^{(K)}$ are independent, $\mathcal{L}(N^{(r)}) = \mathcal{N}(0, \mathrm{Id}_d)$ for $r = 1, \dots, K$ and $T_n T_n^T = \mathrm{Cov}(X_n)$ for $n \geq 0$. Hence, Z_n^* is L_3-integrable, $\mathbb{E}[Z_n^*] = 0$ and $\mathrm{Cov}(Z_n^*) = \mathrm{Id}_d$ for all $n \geq n_1$. By the triangle inequality, we have

$$\zeta_3(X_n, \mathcal{N}(0, \mathrm{Id}_d)) \leq \zeta_3(X_n, Z_n^*) + \zeta_3(Z_n^*, \mathcal{N}(0, \mathrm{Id}_d)).$$

Then, the assertion follows inductively if one has shown the bound $\zeta_3(Z_n^*, \mathcal{N}(0, \mathrm{Id}_d)) = \mathrm{O}(R(n))$: Using the convolution property of the multidimensional normal distribution, we obtain the representation

$$Z_n^* = \sum_{r=1}^{K} A_r^{(n)} T_{I_r^{(n)}} N^{(r)} + b^{(n)} \stackrel{d}{=} G_n N + b^{(n)}, \tag{17}$$

where $G_n G_n^T = \sum_{r=1}^{K} A_r^{(n)} T_{I_r^{(n)}} T_{I_r^{(n)}}^T (A_r^{(n)})^T$, $\mathcal{L}(N) = \mathcal{N}(0, \mathrm{Id}_d)$ and N is independent of $(G_n, b^{(n)})$. As $\mathrm{Cov}(Z_n^*) = \mathrm{Id}_d$ for all $n \geq n_1$, we have $\mathbb{E}[G_n G_n^T + b^{(n)} (b^{(n)})^T] = \mathrm{Id}_d$ for $n \geq n_1$. Furthermore, we have $\big\| b^{(n)} \big\|_3^3 = \mathrm{O}(R(n))$ and

$$\begin{aligned}
\big\| G_n G_n^T - \mathrm{Id}_d \big\|_{3/2}^{3/2} &= \Big\| \sum_{r=1}^{K} A_r^{(n)} T_{I_r^{(n)}} T_{I_r^{(n)}}^T (A_r^{(n)})^T - \mathrm{Id}_d \Big\|_{3/2}^{3/2} \\
&= \mathrm{O}\Bigg(\Big\| \sum_{r=1}^{K} \mathbf{1}_{\{I_r^{(n)} < n_1\}} A_r^{(n)} (T_{I_r^{(n)}} T_{I_r^{(n)}}^T - \mathrm{Id}_d)(A_r^{(n)})^T \Big\|_{3/2}^{3/2} \\
&\qquad + \Big\| \sum_{r=1}^{K} A_r^{(n)} (A_r^{(n)})^T - \mathrm{Id}_d \Big\|_{3/2}^{3/2} \Bigg) \\
&= \mathrm{O}\Bigg(\sum_{r=1}^{K} \big\| \mathbf{1}_{\{I_r^{(n)} < n_1\}} A_r^{(n)} \big\|_3^3 + \Big\| \sum_{r=1}^{K} A_r^{(n)} (A_r^{(n)})^T - \mathrm{Id}_d \Big\|_{3/2}^{3/2} \Bigg) \\
&= \mathrm{O}(R(n)).
\end{aligned}$$

Thus, the following Lemma 5 implies $\zeta_3(Z_n^*, \mathcal{N}(0, \mathrm{Id}_d)) = \mathrm{O}(R(n))$. Lemma 5 is the main part of the present proof. ◀

▶ **Remark 4.** Theorem 3, when applicable, often improves over Theorem 1 by a factor 3 in the exponent, see Remark 9 for an example. This is caused by the additional exponents in (15) in comparison to (10).

▶ **Lemma 5.** *Let $(Z_n^*)_{n\geq 0}$ be a sequence of d-dimensional random vectors satisfying $Z_n^* \stackrel{d}{=} G_n N + b^{(n)}$, where G_n is a random $(d \times d)$-matrix, $b^{(n)}$ a centered random vector with $\mathbb{E}[G_n G_n^T + b^{(n)}(b^{(n)})^T] = \mathrm{Id}_d$ and $N \sim \mathcal{N}(0, \mathrm{Id}_d)$ independent of $(G_n, b^{(n)})$. Furthermore, we assume that, as $n \to \infty$,*

$$\left\|G_n G_n^T - \mathrm{Id}_d\right\|_{3/2}^{3/2} + \left\|b^{(n)}\right\|_3^3 = \mathrm{O}(R(n))$$

for appropriate $R(n)$. Then, we have, as $n \to \infty$,

$$\zeta_3(Z_n^*, \mathcal{N}(0, \mathrm{Id}_d)) = \mathrm{O}(R(n)).$$

The proof of Lemma 5 builds upon ideas of [15].

3 Expansions of moments

In applications to problems arising in theoretical computer science, where the recurrence (1) is explicitly given one usually has no direct means to identify the orders of the terms $\|b^{(n)} - b^*\|_s$ and $\|A_r^{(n)} - A_r^*\|_s$. This is due to the fact that the mean vector M_n and the covariance matrix C_n, for the cases $1 < s \leq 2$ and $2 < s \leq 3$ respectively, which are used for the normalization (7) are typically not exactly known or too involved to be amenable to explicit calculations. As a substitute one usually has asymptotic expansions of these sequences as $n \to \infty$.

In the present section we assume the dimension to be $d = 1$ and $A_r(n) = 1$ for all $r = 1, \dots, K$ and provide tools to apply the general Theorems 1 and 3 on the basis of expansions of the mean and variance. We assume that

$$\mathbb{E}[X_n] = \mu(n) = f(n) + \mathrm{O}(e(n)), \quad \mathrm{Var}(X_n) = \sigma^2(n) = g(n) + \mathrm{O}(h(n)), \tag{18}$$

with $e(n) = o(f(n))$ and $h(n) = o(g(n))$. To connect Theorems 1 and 3 to recurrences with known expansions we use the following notion.

▶ **Definition 6.** *A sequence $(a(n))_{n\geq 0}$ of non-negative numbers is called* essentially non-decreasing *if there exists a $c > 0$ such that $a(m) \leq c a(n)$ for all $0 \leq m < n$.*

The scaling introduced in (7) with the special choices $A_r(n) = 1$ for all $r = 1, \dots, K$ leads to the scaled recurrence for (X_n) given in (2) with

$$A_r^{(n)} = \frac{\sigma(I_r^{(n)})}{\sigma(n)}, \quad b^{(n)} = \frac{1}{\sigma(n)}\Big(b_n - \mu(n) + \sum_{r=1}^K \mu(I_r^{(n)})\Big). \tag{19}$$

Additionally, we consider the corresponding quantities

$$\overline{A}_r^{(n)} = \frac{g^{1/2}(I_r^{(n)})}{g^{1/2}(n)}, \quad \overline{b}^{(n)} = \frac{1}{g^{1/2}(n)}\Big(b_n - f(n) + \sum_{r=1}^K f(I_r^{(n)})\Big). \tag{20}$$

Then we have:

▶ **Lemma 7.** *With $A_r^{(n)}$, $b^{(n)}$ given in (19), $\overline{A}_r^{(n)}$, $\overline{b}^{(n)}$ given in (20), and the expansions for $\mu(n)$, $\sigma^2(n)$ given in (18) the following holds.*
If the sequence $h/g^{1/2}$ is essentially non-decreasing then

$$\left\|A_r^{(n)} - A_r^*\right\|_s \le \left\|\overline{A}_r^{(n)} - A_r^*\right\|_s + \mathrm{O}\Big(\frac{h(n)}{g(n)}\Big). \tag{21}$$

If the sequence h is essentially non-decreasing then

$$\Big\|\sum_{r=1}^K (A_r^{(n)})^2 - 1\Big\|_s \le \Big\|\sum_{r=1}^K (\overline{A}_r^{(n)})^2 - 1\Big\|_s + \mathrm{O}\Big(\frac{h(n)}{g(n)}\Big). \tag{22}$$

If the sequence e is essentially non-decreasing then

$$\left\|b^{(n)} - b^*\right\|_s \le \left\|\overline{b}^{(n)} - b^*\right\|_s + \mathrm{O}\Big(\frac{h(n)}{g(n)} + \frac{e(n)}{g^{1/2}(n)}\Big). \tag{23}$$

If the sequence g/h is essentially non-decreasing and

$$T(n) := \mathbb{E}\sum_{r=1}^K \frac{g^{s/2-1}(I_r^{(n)})h(I_r^{(n)})R(I_r^{(n)})}{g^{s/2}(n)R(n)}$$

then we have

$$\mathbb{E}\sum_{r=1}^K \frac{\sigma^s(I_r^{(n)})R(I_r^{(n)})}{\sigma^s(n)R(n)} \le \mathbb{E}\sum_{r=1}^K \frac{g^{s/2}(I_r^{(n)})R(I_r^{(n)})}{g^{s/2}(n)R(n)} + \mathrm{O}(T(n)). \tag{24}$$

Proof. We show (21), the other bounds can be shown similarly. Note that $\sigma^2(n) = g(n) + \mathrm{O}(h(n))$ implies $\sigma(n) = g^{1/2}(n) + \mathrm{O}(h(n)/g^{1/2}(n))$ and that for any essentially non-decreasing sequence $(a(n))_{n\ge 0}$ we have $\|a(I_r^{(n)})\|_\infty = \mathrm{O}(a(n))$. Since $h/g^{1/2}$ is essentially non-decreasing we obtain

$$\begin{aligned}
A_r^{(n)} = \frac{\sigma(I_r^{(n)})}{\sigma(n)} &= \frac{g^{1/2}(I_r^{(n)}) + \mathrm{O}(h(I_r^{(n)})/g^{1/2}(I_r^{(n)}))}{\sigma(n)}\\
&= \frac{g^{1/2}(I_r^{(n)}) + \mathrm{O}(h(n)/g^{1/2}(n))}{g^{1/2}(n)} \cdot \frac{g^{1/2}(n)}{\sigma(n)}\\
&= \left(\frac{g^{1/2}(I_r^{(n)})}{g^{1/2}(n)} + \mathrm{O}\left(\frac{h(n)}{g(n)}\right)\right)\left(1 + \mathrm{O}\left(\frac{h(n)}{g(n)}\right)\right)\\
&= \frac{g^{1/2}(I_r^{(n)})}{g^{1/2}(n)} + \mathrm{O}\left(\frac{h(n)}{g(n)}\left(1 + \frac{g^{1/2}(I_r^{(n)})}{g^{1/2}(n)}\right)\right).
\end{aligned}$$

Hence, we obtain

$$\|A_r^{(n)} - A_r^*\|_s \le \|\overline{A}_r^{(n)} - A_r^*\|_s + \mathrm{O}\left(\frac{h(n)}{g(n)}\left(1 + \left\|\overline{A}_r^{(n)}\right\|_s\right)\right).$$

Since $\overline{A}_r^{(n)} \to A_r^*$ in L_s we have $\|\overline{A}_r^{(n)}\|_s = \mathrm{O}(1)$, hence

$$\|A_r^{(n)} - A_r^*\|_s \le \|\overline{A}_r^{(n)} - A_r^*\|_s + \mathrm{O}\left(\frac{h(n)}{g(n)}\right),$$

which is bound (21). ◀

Note that in applications the terms on the right hand side in the estimates (21)–(24) can easily be bound when expansions as in (18) with explicit functions e, f, g, h are available.

4 Applications

We start by deriving a known result to illustrate in detail how to apply our framework of the previous sections.

4.1 Quicksort: Key comparisons

The number of key comparisons Y_n needed by the Quicksort algorithm to sort n randomly permuted (distinct) numbers satisfies the distributional recursion

$$Y_n \stackrel{d}{=} Y_{I_n} + Y'_{n-1-I_n} + n - 1, \quad n \geq 1, \tag{25}$$

where $Y_0 := 0$ and $(Y_k)_{k=0,\ldots,n-1}, (Y'_k)_{k=0,\ldots,n-1}, I_n$ are independent, I_n is uniformly distributed on $\{0,\ldots,n-1\}$, and $Y_k \stackrel{d}{=} Y'_k$, $k \geq 0$. Hence, equation (25) is covered by our general recurrence (1). For the expectation and variance of Y_n exact expressions are known which imply the asymptotic expansions

$$\mathbb{E}Y_n = 2n\log(n) + (2\gamma - 4)n + \mathrm{O}(\log n), \tag{26}$$

$$\mathrm{Var}(Y_n) = \sigma^2 n^2 - 2n\log(n) + \mathrm{O}(n), \tag{27}$$

where γ denotes Euler's constant and $\sigma := \sqrt{7 - 2\pi^2/3} > 0$. We introduce the normalized quantities $X_0 := X_1 := X_2 := 0$ and

$$X_n := \frac{Y_n - \mathbb{E}Y_n}{\sqrt{\mathrm{Var}(Y_n)}}, \quad n \geq 3. \tag{28}$$

To apply Theorem 1 we need to find an $0 < s \leq 3$ and a sequence $(R(n))$ with (10) and (11). Note that the Y_n are bounded, thus L_s-integrable for any $s > 0$. To bound the L_s-norms appearing in (10) we use Lemma 7 and choose

$$f(n) = 2n\log(n) + (2\gamma - 4)n, \quad e(n) = \log n,$$
$$g(n) = \sigma^2 n^2, \quad h(n) = n\log n.$$

With these functions we obtain for the quantities defined in (20) that

$$\overline{A}_1^{(n)} = \frac{I_n}{n}, \quad \overline{A}_2^{(n)} = \frac{n-1-I_n}{n},$$
$$\overline{b}^{(n)} = \frac{1}{\sigma}\left(2\frac{I_n}{n}\log\frac{I_n}{n} + 2\frac{n-1-I_n}{n}\log\frac{n-1-I_n}{n} + \frac{n-1}{n} + \mathrm{O}\left(\frac{\log n}{n}\right)\right)$$

With the embedding $I_n = \lfloor nU \rfloor$ with U uniformly distributed over the unit interval $[0,1]$ we have

$$A_1^* = U, \quad A_2^* = 1 - U, \quad b^* = \frac{1}{\sigma}\left(2U\log(U) + 2(1-U)\log(1-U) + 1\right) =: \frac{1}{\sigma}\varphi(U).$$

The limit theorem $X_n \to X$ has been derived by different methods by Régnier [16] and Rösler [17]. Rösler [17] also found that the scaled limit $Y := \sigma X$ satisfies the distributional fixed-point equation

$$Y \stackrel{d}{=} UY + (1-U)Y' + \varphi(U). \tag{29}$$

Lower and upper bounds for the rate of convergence in $X_n \to X$ have been studied for various metrics in Fill and Janson [6] and Neininger and Rüschendorf [13].

Now, we apply the framework of the present paper: For $r = 1, 2$ and any $s \geq 1$ we find that

$$\|\overline{A}_r^{(n)} - A_r^*\|_s = \mathrm{O}\Big(\frac{1}{n}\Big).$$

Using Proposition 3.2 of Rösler [17] we obtain

$$\|\bar{b}_n - b^*\|_s = \mathrm{O}\Big(\frac{\log n}{n}\Big).$$

Moreover, we have

$$\frac{h(n)}{g(n)} = \mathrm{O}(R(n)) \quad \text{and} \quad \frac{e(n)}{g^{1/2}(n)} = \mathrm{O}(R(n)) \quad \text{with} \quad R(n) := \frac{\log n}{n},$$

thus Lemma 7 implies that condition (10) is satisfied for our choice of the sequence R. To verify condition (11) by use of (24) we obtain that for $T(n)$ given in Lemma 7 we find $T(n) = \mathrm{O}(\log(n)/n) \to 0$ and that

$$\mathbb{E}\sum_{r=1}^{2} \frac{g^{s/2}(I_r^{(n)})R(I_r^{(n)})}{g^{s/2}(n)R(n)} = \mathbb{E}\sum_{r=1}^{2}\left(\frac{I_r^{(n)}}{n}\right)^{s-1}\frac{\log I_r^{(n)}}{\log n}.$$

Note that the latter expression has a limes superior of less than 1 if and only if $s > 2$. Hence, Theorem 1 is applicable for $s > 2$ and yields that

$$\zeta_s(X_n, X) = \mathrm{O}\left(\frac{\log n}{n}\right), \quad \text{for} \quad 2 < s \leq 3. \tag{30}$$

The bound (30) had previously been shown for $s = 3$ in [13], where also the optimality of the order was shown, i.e., that $\zeta_3(X_n, X) = \Theta\left(\log(n)/n\right)$.

In the full paper version we also discuss bounds on rates of convergence for various cost measures of the related Quickselect algorithms under various models for the rank to be selected.

4.2 Size of m-ary search trees

The size of m-ary search trees satisfies the recurrence (1) with $K = m \geq 3$, $A_1(n) = \cdots = A_m(n) = 1$, $n_0 = m$, $b_n = 1$, i.e., we have

$$Y_n \stackrel{d}{=} \sum_{r=1}^{m} Y^{(r)}_{I_r^{(n)}} + 1, \quad n \geq m.$$

For a representation of $I^{(n)}$ we define for independent, identically unif$[0,1]$ distributed random variables $U_1, \ldots, U_{m-1}$ their spacings in $[0,1]$ by $S_1 = U_{(1)}, S_2 = U_{(2)} - U_{(1)}, \ldots, S_m := 1 - U_{(m-1)}$, where $U_{(1)}, \ldots U_{(m-1)}$ denote the order statistics of $U_1, \ldots, U_{m-1}$. Then $I^{(n)}$ has the mixed multinomial distribution:

$$I^{(n)} \stackrel{d}{=} M(n - m + 1, S_1, \ldots, S_m).$$

By this we mean that given $(S_1, \ldots, S_m) = (s_1, \ldots, s_m)$ we have that $I^{(n)}$ is multinomial $M(n - m + 1, s_1, \ldots, s_m)$ distributed. Expectations, variances and limit laws for Y_n have been studied, see[12, 4]. We have

$$\mathbb{E}Y_n = \mu n + \mathrm{O}(1 + n^{\alpha-1}), \quad m \geq 3, \tag{31}$$

$$\mathrm{Var}(Y_n) = \sigma^2 n + \mathrm{O}(1 + n^{2\alpha-2}), \quad 3 \leq m \leq 26, \tag{32}$$

Here, the constants $\mu, \sigma > 0$ depend on m and $\alpha \in \mathbb{R}$ depends on m such that $\alpha < 1$ for $m \leq 13$, $1 \leq \alpha \leq 4/3$ for $14 \leq m \leq 19$, and $4/3 \leq \alpha \leq 3/2$ for $20 \leq m \leq 26$, see, e.g., Mahmoud [12, Table 3.1] for the values $\alpha = \alpha_m$ depending on m. It is known that Y_n standardized by mean and variance satisfies a central limit law for $m \leq 26$, whereas the standardized sequence has no weak limit for $m > 26$ due to dominant periodicities, see Chern and Hwang [4]. The rate of convergence in the central limit law for $m \leq 26$ for the Kolmogorov metric has been identified in Hwang [9]. Our Theorem 3 implies the central limit theorem for Y_n with $m \leq 26$ with the same (up to an ε for $3 \leq m \leq 19$) rate of convergence for the Zolotarev metric ζ_3:

▶ **Theorem 8.** *The size Y_n of a random m-ary search tree with n items inserted satisfies, for $m \leq 26$,*

$$\zeta_3\Big(\frac{Y_n - \mathbb{E}Y_n}{\sqrt{\mathrm{Var}(Y_n)}}, \mathcal{N}(0,1)\Big) = \begin{cases} \mathrm{O}\big(n^{-1/2+\varepsilon}\big), & 3 \leq m \leq 19, \\ \mathrm{O}\big(n^{-3(3/2-\alpha)}\big), & 20 \leq m \leq 26, \end{cases} \tag{33}$$

as $n \to \infty$.

Proof. In order to apply Theorem 3 we have to estimate the orders of $\| \sum_{r=1}^m (A_r^{(n)})^2 - 1\|_{3/2}$ and $\|b^{(n)}\|_3$ with $A_r^{(n)}$ and $b^{(n)}$ defined in (3). For this we apply Lemma 7. From (31) and (32) we obtain that for the quantities appearing in Lemma 7 we can choose $f(n) = \mu n$, $e(n) = 1 \vee n^{\alpha-1}$, $g(n) = \sigma^2 n$, and $h(n) = 1 \vee n^{2(\alpha-1)}$. Hence we obtain

$$\Big\| \sum_{r=1}^m (\overline{A}_r^{(n)})^2 - 1 \Big\|_{3/2} = \Big\| \sum_{r=1}^m \frac{I_r^{(n)}}{n} - 1 \Big\|_{3/2} = \frac{m-1}{n} = \mathrm{O}\big(n^{-1}\big)$$

and $\mathrm{O}(h(n)/g(n)) = \mathrm{O}(n^{-(1\wedge(3-2\alpha))})$. This implies

$$\Big\| \sum_{r=1}^m (A_r^{(n)})^2 - 1 \Big\|_{3/2}^{3/2} = \mathrm{O}\big(n^{-((3/2)\wedge(3(3/2-\alpha)))}\big).$$

Similarly we obtain

$$\big\|\overline{b}^{(n)}\big\|_3 = \frac{1}{\sigma\sqrt{n}} \Big\| 1 - \mu n + \sum_{r=1}^m \mu I_r^{(n)} \Big\|_3 = \frac{1}{\sigma\sqrt{n}} \big\| 1 - \mu(m-1) \big\|_3 = \mathrm{O}\big(n^{-1/2}\big)$$

and $\mathrm{O}(e(n)/g^{1/2}(n)) = \mathrm{O}(n^{-(1\wedge(3/2-\alpha))})$. This implies

$$\big\|b^{(n)}\big\|_3^3 = \mathrm{O}\big(n^{-((3/2)\wedge(3(3/2-\alpha)))}\big).$$

Hence, condition (15) is satisfied with $R(n) = n^{-((3/2)\wedge(3(3/2-\alpha)))}$. ◀

▶ Remark 9. Using Theorem 1 instead of Theorem 3 in the latter proof is also possible but leads to a bound $\mathrm{O}(n^{-(3/2-\alpha)})$ for $20 \leq m \leq 26$, missing the factor 3 appearing in Theorem 8.

In the full paper version we also discuss rates of convergence for the number of leaves of d-dimensional random point quadtrees in the model of [7, 3, 8] where a similar behavior as in Theorem 8 appears. A technically related example is the number of maxima in right triangles in the model of [1, 2], where the order $n^{-1/4}$ appears. Our framework also applies.

4.3 Periodic functions in mean and variance

We now discuss some examples where the asymptotic expansions of the mean and the variance include periodic functions instead of fixed constants. This is the case for several quantities in binomial splitting processes such as tries, PATRICIA tries and digital search trees. Throughout this section, we assume that we have a 3-integrable sequence $(Y_n)_{n\geq 0}$ satisfying the recursion

$$Y_n \stackrel{d}{=} Y^{(1)}_{I_1^{(n)}} + Y^{(2)}_{I_2^{(n)}} + b_n, \quad n \geq n_0, \tag{34}$$

with $(I^{(n)}, b_n), (Y_n^{(1)})_{n\geq 0}$ and $(Y_n^{(2)})_{n\geq 0}$ independent and $(Y_n^{(r)})_{n\geq 0} \stackrel{d}{=} (Y_n)_{n\geq 0}$ for $r = 1, 2$. Furthermore, $I_1^{(n)}$ has the binomial distribution $\mathrm{Bin}(n,\frac{1}{2})$ and $I_2^{(n)} = n - I_1(n)$ or $I_1^{(n)}$ is binomially $\mathrm{Bin}(n-1,\frac{1}{2})$ distributed and $I_2^{(n)} = n-1-I_1(n)$. Mostly, these binomial recurrences are asymptotically normally distributed, see [10, 11, 14, 18] for some examples.

Our first theorem covers the case of linear mean and variance, i.e. we assume that, as $n \to \infty$,

$$\mathbb{E}[Y_n] = nP_1(\log_2 n) + \mathrm{O}(1), \tag{35}$$

$$\mathrm{Var}(Y_n) = nP_2(\log_2 n) + \mathrm{O}(1), \tag{36}$$

for some smooth and 1-periodic functions P_1, P_2 with $P_2 > 0$. Possible applications would start with the analysis of the number of internal nodes of a trie for n strings in the symmetric Bernoulli model and the number of leaves in a random digital search tree, see, e.g., [10].

▶ **Theorem 10.** *Let $(Y_n)_{n\geq 0}$ be 3-integrable and satisfy (34) with $\|b_n\|_3 = \mathrm{O}(1)$, (35) and (36). Then, for any $\varepsilon > 0$ and $n \to \infty$, we have*

$$\zeta_3\Big(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\mathrm{Var}(Y_n)}}, \mathcal{N}(0,1)\Big) = \mathrm{O}(n^{-1/2+\varepsilon}).$$

We now consider the case where our quantities Y_n satisfy recursion (34) with b_n being essentially n. We assume that, as $n \to \infty$, we have

$$\mathbb{E}[Y_n] = n\log_2(n) + nP_1(\log_2 n) + \mathrm{O}(1), \tag{37}$$

$$\mathrm{Var}(Y_n) = nP_2(\log_2 n) + \mathrm{O}(1), \tag{38}$$

for some smooth and 1-periodic functions P_1, P_2 with $P_2 > 0$. This covers, for example, the external path length of random tries and related digital tree structures constructed from n random binary strings under appropriate independence assumptions.

▶ **Theorem 11.** *Let $(Y_n)_{n\geq 0}$ be 3-integrable and satisfy (34) with $\|b_n - n\|_3 = \mathrm{O}(1)$, (37) and (38). Then, for any $\varepsilon > 0$ and $n \to \infty$, we have*

$$\zeta_3\Big(\frac{Y_n - \mathbb{E}[Y_n]}{\sqrt{\mathrm{Var}(Y_n)}}, \mathcal{N}(0,1)\Big) = \mathrm{O}(n^{-1/2+\varepsilon}).$$

4.4 A multivariate application

We consider a random binary search tree with n nodes built from a random permutation of $\{1, \dots, n\}$. For $n \geq 0$, we denote by L_{0n} the number of nodes with no left descendant and by L_{1n} the number of nodes with exactly one left descendant. Defining $Y_n := (L_{0n}, L_{1n})$, we have $Y_0 = (0,0)$ and we obtain the following distributional recurrence:

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

where $(Y_j^{(1)})_{j\geq 0}$ and $(Y_j^{(2)})_{j\geq 0}$ are independent copies of $(Y_j)_{j\geq 0}$, $I_1^{(n)}$ is uniformly distributed on $\{0,\dots,n-1\}$ and independent of $(Y^{(1)})$ and $(Y^{(2)})$, $I_2^{(n)} = n-1-I_1^{(n)}$ and $b_n = (\mathbf{1}_{\{I_1^{(n)}=0\}}, \mathbf{1}_{\{I_1^{(n)}=1\}})$. In Devroye [5] it is shown that, for $n \geq 2$,

$$\mathbb{E}[L_{0n}] = \frac{1}{2}(n+1), \quad \mathbb{E}[L_{1n}] = \frac{1}{6}(n+1),$$

and that the standardized quantities have a limiting normal distribution. Using Devroye's description with local counters one also obtains the covariance structure:

▶ **Lemma 12.** *For $n \geq 4$, we have* $\mathrm{Cov}(Y_n) = (n+1)\,\Gamma$ *with*

$$\Gamma = \frac{1}{360}\begin{pmatrix} 30 & -15 \\ -15 & 28 \end{pmatrix}.$$

For $n \geq 0$, we now set $M_n := \mathbb{E}[Y_n]$, $C_n = \mathrm{Id}_2$ for $n \leq 3$, $C_n := \mathrm{Cov}(Y_n)$ for $n \geq 4$ and define $X_n := C_n^{-1/2}(Y_n - M_n)$ for $n \geq 0$. Note that the matrix Γ in Lemma 12 is symmetric and positive definite, which implies, for $n \geq 4$,

$$C_n^{1/2} = \sqrt{n+1}\,\Gamma^{1/2} \quad \text{and} \quad C_n^{-1/2} = \frac{1}{\sqrt{n+1}}\,\Gamma^{-1/2}.$$

The normalized quantities satisfy $X_0 = (0,0)$ and recursion (2) with $K=2$, $n_0 = 1$,

$$A_r^{(n)} = C_n^{-1/2} C^{1/2}_{I_r^{(n)}} = \mathbf{1}_{\{I_r^{(n)} \geq 4\}} \sqrt{\frac{I_r^{(n)}+1}{n+1}}\,\mathrm{Id}_2 + \mathbf{1}_{\{I_r^{(n)}<4\}} \frac{1}{\sqrt{n+1}}\,\Gamma^{-1/2}$$

for $r = 1,2$ and

$$b^{(n)} = C_n^{-1/2}(b_n - M_n + M_{I_1^{(n)}} + M_{I_2^{(n)}}).$$

Modeling all quantities on a joint probability space such that $I_1^{(n)}/n$ converges almost surely to a uniform random variable U in $[0,1]$, we have the L_3-convergences $A_1^{(n)} \to \sqrt{U}\,\mathrm{Id}_2$, $A_2^{(n)} \to \sqrt{1-U}\,\mathrm{Id}_2$ and $b^{(n)} \to 0$ as $n \to \infty$. Thus, we are in the situation of Section 2.2 and obtain the limiting equation

$$X \stackrel{d}{=} \sqrt{U} X^{(1)} + \sqrt{1-U} X^{(2)},$$

with U uniformly distributed on $[0,1]$ and $X^{(1)}$, $X^{(2)}$ and U independent. We now check the conditions of Theorem 3. Since $A_1^{(n)}(A_1^{(n)})^T + A_2^{(n)}(A_2^{(n)})^T = \mathrm{Id}_2$ on the event $\{I_1^{(n)}, I_2^{(n)} \geq 4\}$, we obtain, as $n \to \infty$,

$$\begin{aligned}\Big\| \sum_{r=1}^{2} A_r^{(n)}(A_r^{(n)})^T - \mathrm{Id}_2 \Big\|_{3/2}^{3/2} &= \mathrm{O}\bigg(\Big\| \mathbf{1}_{\{I_1^{(n)}<4\}} \Big(\frac{1}{n+1}\,\Gamma^{-1} + \frac{I_2^{(n)}+1}{n+1}\,\mathrm{Id}_2 - \mathrm{Id}_2 \Big) \Big\|_{3/2}^{3/2} \bigg) \\ &= \mathrm{O}\bigg(\mathbb{E}\Big[\mathbf{1}_{\{I_1^{(n)}<4\}} \Big\| \frac{1}{n+1}\,\Gamma^{-1} - \frac{I_1^{(n)}+1}{n+1}\,\mathrm{Id}_2 \Big\|_{\mathrm{op}}^{3/2} \Big] \bigg) \\ &= \mathrm{O}\big(n^{-5/2}\big).\end{aligned}$$

Similarly, we obtain

$$\|b^{(n)}\|_3^3 = \mathrm{O}(n^{-5/2}).$$

Since we have $\|\mathbf{1}_{\{I_r^{(n)}<\ell\}} A_r^{(n)}\|_3^3 = \mathrm{O}(n^{-5/2})$ for $\ell \in \mathbb{N}$ and $r = 1, 2$, the technical conditions are satisfied. We now use Theorem 3 with $R(n) = n^{-1/2}$. Note that condition (16) is not satisfied for $R(n) = n^{-1/2}$, but we can use the weakened condition stated in Remark 2 to obtain the following result.

▶ **Theorem 13.** *Denoting by $Y_n := (L_{0n}, L_{1n})$ the vector of the numbers of nodes with no and with exactly one left descendant respectively in a random binary search tree with n nodes we have, for $n \to \infty$, that*

$$\zeta_3\big(\mathrm{Cov}(Y_n)^{-1/2}(Y_n - \mathbb{E}[Y_n]), \mathcal{N}(0, \mathrm{Id}_2)\big) = \mathrm{O}(n^{-1/2}).$$

References

1 Zhi-Dong Bai, Hsien-Kuei Hwang, Wen-Qi Liang, and Tsung-Hsi Tsai. Limit theorems for the number of maxima in random samples from planar regions. *Electron. J. Probab.*, 6:no. 3, 41 pp. (electronic), 2001. URL: `http://dx.doi.org.proxy.ub.uni-frankfurt.de/10.1214/EJP.v6-76`, `doi:10.1214/EJP.v6-76`.

2 Zhi-Dong Bai, Hsien-Kuei Hwang, and Tsung-Hsi Tsai. Berry-Esseen bounds for the number of maxima in planar regions. *Electron. J. Probab.*, 8:no. 9, 26, 2003. URL: `https://doi.org/10.1214/EJP.v8-137`, `doi:10.1214/EJP.v8-137`.

3 Hua-Huai Chern, Michael Fuchs, and Hsien-Kuei Hwang. Phase changes in random point quadtrees. *ACM Trans. Algorithms*, 3(2):Art. 12, 51, 2007. URL: `http://dx.doi.org/10.1145/1240233.1240235`, `doi:10.1145/1240233.1240235`.

4 Hua-Huai Chern and Hsien-Kuei Hwang. Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms*, 19(3-4):316–358, 2001. Analysis of algorithms (Krynica Morska, 2000). URL: `http://dx.doi.org.proxy.ub.uni-frankfurt.de/10.1002/rsa.10005`, `doi:10.1002/rsa.10005`.

5 Luc Devroye. Limit laws for local counters in random binary search trees. *Random Structures Algorithms*, 2(3):303–315, 1991. URL: `http://dx.doi.org.proxy.ub.uni-frankfurt.de/10.1002/rsa.3240020305`, `doi:10.1002/rsa.3240020305`.

6 James Allen Fill and Svante Janson. Quicksort asymptotics. volume 44, pages 4–28. 2002. Analysis of algorithms. URL: `https://doi.org/10.1016/S0196-6774(02)00216-X`, `doi:10.1016/S0196-6774(02)00216-X`.

7 Philippe Flajolet, Gilbert Labelle, Louise Laforest, and Bruno Salvy. Hypergeometrics and the cost structure of quadtrees. *Random Structures Algorithms*, 7(2):117–144, 1995. URL: `https://doi.org/10.1002/rsa.3240070203`, `doi:10.1002/rsa.3240070203`.

8 Michael Fuchs, Noela S. Müller, and Henning Sulzbach. Refined asymptotics for the number of leaves of random point quadtrees. In *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 110 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 23, 16. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018.

9 Hsien-Kuei Hwang. Second phase changes in random m-ary search trees and generalized quicksort: convergence rates. *Ann. Probab.*, 31(2):609–629, 2003. URL: `https://doi.org/10.1214/aop/1048516530`, `doi:10.1214/aop/1048516530`.

10 Hsien-Kuei Hwang, Michael Fuchs, and Vytas Zacharovas. Asymptotic variance of random symmetric digital search trees. *Discrete Math. Theor. Comput. Sci.*, 12(2):103–165, 2010.

11 Philippe Jacquet and Wojciech Szpankowski. Analytical de-Poissonization and its applications. *Theoret. Comput. Sci.*, 201(1-2):1–62, 1998. URL: `https://doi.org/10.1016/S0304-3975(97)00167-9`, `doi:10.1016/S0304-3975(97)00167-9`.

12 Hosam M. Mahmoud. *Evolution of random search trees*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

13 Ralph Neininger and Ludger Rüschendorf. Rates of convergence for Quicksort. volume 44, pages 51–62. 2002. Analysis of algorithms. URL: `https://doi.org/10.1016/S0196-6774(02)00206-7`, `doi:10.1016/S0196-6774(02)00206-7`.

14 Ralph Neininger and Ludger Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14(1):378–418, 2004. URL: `https://doi.org/10.1214/aoap/1075828056`, `doi:10.1214/aoap/1075828056`.

15 Ralph Neininger and Ludger Rüschendorf. On the contraction method with degenerate limit equation. *Ann. Probab.*, 32(3B):2838–2856, 2004. URL: `https://doi-org.proxy.ub.uni-frankfurt.de/10.1214/009117904000000171`, `doi:10.1214/009117904000000171`.

16 Mireille Régnier. A limiting distribution for quicksort. *RAIRO Inform. Théor. Appl.*, 23(3):335–343, 1989. URL: `https://doi.org/10.1051/ita/1989230303351`, `doi:10.1051/ita/1989230303351`.

17 Uwe Rösler. A limit theorem for "Quicksort". *RAIRO Inform. Théor. Appl.*, 25(1):85–100, 1991. URL: `https://doi.org/10.1051/ita/1991250100851`, `doi:10.1051/ita/1991250100851`.

18 Werner Schachinger. Asymptotic normality of recursive algorithms via martingale difference arrays. *Discrete Math. Theor. Comput. Sci.*, 4(2):363–397, 2001.

19 V. M. Zolotarev. Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Teor. Verojatnost. i Primenen.*, 21(4):741–758, 1976.

5 Appendix

Proof. (*Proof of Theorem 1*) Using condition (10), the assumption that R is monotonically decreasing and condition (11), we have

$$\mathbb{E}\left[\sum_{r=1}^{K}\|A_r^*\|_{\mathrm{op}}^s\right] = \lim_{n\to\infty}\mathbb{E}\left[\sum_{r=1}^{K}\|A_r^{(n)}\|_{\mathrm{op}}^s\right] \le \limsup_{n\to\infty}\mathbb{E}\left[\sum_{r=1}^{K}\frac{R(I_r^{(n)})}{R(n)}\|A_r^{(n)}\|_{\mathrm{op}}^s\right] < 1.$$

Furthermore, condition (10) implies $\mathbb{E}[b^*] = \lim_{n\to\infty}\mathbb{E}[b^{(n)}] = 0$ if $s > 1$ and additionally

$$\mathbb{E}\big[b^*(b^*)^T\big] + \mathbb{E}\Big[\sum_{r=1}^{K}A_r^*(A_r^*)^T\Big] = \mathrm{Id}_d$$

if $s > 2$. Thus, Corollary 3.4 in [14] states that equation (12) has a unique fixed-point $\mathcal{L}(X)$ in $\mathcal{P}_s^d(0, \mathrm{Id}_d)$. To establish a rate of convergence to this fixed-point, we introduce the accompanying sequence

$$Z_n^* := \sum_{r=1}^{K}A_r^{(n)}T_{I_r^{(n)}}X^{(r)} + b^{(n)},$$

where $(A_1^{(n)},\dots,A_K^{(n)},I^{(n)},b^{(n)}), X^{(1)},\dots,X^{(K)}$ are independent and $X^{(r)}$ is identically distributed as X for $r = 1,\dots,K$. Here, for $2 < s \le 3$, the sequence $(T_n)_{n\ge0}$ is chosen such that Z_n^* has the same covariance structure as X_n. To be more precise, for $2 < s \le 3$, we choose T_n such that $T_nT_n^T = \mathrm{Cov}(X_n)$ (i.e. $T_n = \mathrm{Id}_d$ for $n \ge n_1$ and $T_nT_n^T = \mathrm{Cov}(Y_n)$ for $n < n_1$). For $s \le 2$, we do not need to control the covariance of Z_n^* and set $T_n := \mathrm{Id}_d$ for $n \ge 0$. Then, Z_n^* is L_s-integrable, we have $\mathbb{E}[Z_n^*] = 0$ for $s > 1$ and in the case $s > 2$

additionally $\mathrm{Cov}(Z_n^*) = \mathrm{Cov}(X_n) = \mathrm{Id}_d$ for $n \geq n_1$. Hence, ζ_s-distances between X_n, Z_n^* and X are finite for $n \geq n_1$. Applying the triangle inequality we have, for $n \geq n_1$,

$$\zeta_s(X_n, X) \leq \zeta_s(X_n, Z_n^*) + \zeta_s(Z_n^*, X). \tag{39}$$

Denoting by Υ_n the joint distribution of $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}, I^{(n)})$, $\alpha = (\alpha_1, \dots, \alpha_K)$, $j = (j_1, \dots, j_K)$ and $\Delta(n) := \zeta_s(X_n, X)$, we obtain by conditioning on Υ_n that, for $n \geq n_1$,

$$\begin{aligned}
\zeta_s(X_n, Z_n^*) &= \zeta_s\left(\sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} X^{(r)} + b^{(n)}\right) \\
&= \sup_{f \in \mathcal{F}_s} \left| \int \mathbb{E}\Big[f\Big(\sum_{r=1}^K \alpha_r X_{j_r}^{(r)} + \beta\Big)\Big] - \mathbb{E}\Big[f\Big(\sum_{r=1}^K \alpha_r T_{j_r} X^{(r)} + \beta\Big)\Big] \mathrm{d}\Upsilon_n(\alpha, \beta, j) \right| \\
&\leq \int \zeta_s\left(\sum_{r=1}^K \alpha_r X_{j_r}^{(r)} + \beta, \sum_{r=1}^K \alpha_r T_{j_r} X^{(r)} + \beta\right) \mathrm{d}\Upsilon_n(\alpha, \beta, j) \\
&\leq \int \sum_{r=1}^K \|\alpha_r\|_{\mathrm{op}}^s \, \zeta_s\left(X_{j_r}^{(r)}, T_{j_r} X^{(r)}\right) \mathrm{d}\Upsilon_n(\alpha, \beta, j) \\
&\leq \left(\mathbb{E} \sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} = n\}} \|A_r^{(n)}\|_{\mathrm{op}}^s\right) \Delta(n) + \mathbb{E}\left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\mathrm{op}}^s \Delta(I_r^{(n)})\right] \\
&\qquad + \mathbb{E}\left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} < n_1\}} \|A_r^{(n)}\|_{\mathrm{op}}^s \sup_{k < n_1} \zeta_s(X_k, T_k X^{(r)})\right].
\end{aligned} \tag{40}$$

Note that the last summand is in $\mathrm{O}(R(n))$ by condition (8). To bound the second summand $\zeta_s(Z_n^*, X)$ in (39), we switch to the Wasserstein metric ℓ_s: By condition (10) and $\|Z_n^*\|_s \leq \sum_{r=1}^K \|A_r^{(n)} T_{I_r^{(n)}}\|_s \|X\|_s + \|b^{(n)}\|_s$, we have $\sup_{n \geq 0} \|Z_n^*\|_s < \infty$. Thus, a standard bound implies that $\zeta_s(Z_n^*, X) \leq C_s \ell_s(Z_n^*, X)$ for some constant $C_s > 0$. Furthermore, we have

$$\begin{aligned}
\ell_s(Z_n^*, X) &\leq \left\| \left(\sum_{r=1}^K A_r^{(n)} T_{I_r^{(n)}} X^{(r)} + b^{(n)}\right) - \left(\sum_{r=1}^K A_r^* X^{(r)} + b^*\right) \right\|_s \\
&\leq \sum_{r=1}^K \left\|A_r^{(n)} T_{I_r^{(n)}} - A_r^*\right\|_s \left\|X^{(r)}\right\|_s + \left\|b^{(n)} - b^*\right\|_s \\
&\leq \sum_{r=1}^K \left(\left\|A_r^{(n)} T_{I_r^{(n)}} - A_r^{(n)}\right\|_s + \left\|A_r^{(n)} - A_r^*\right\|_s\right) \left\|X\right\|_s + \left\|b^{(n)} - b^*\right\|_s \\
&= \sum_{r=1}^K \left(\left\|\mathbf{1}_{\{I_r^{(n)} < n_1\}} A_r^{(n)} (T_{I_r^{(n)}} - \mathrm{Id}_d)\right\|_s + \left\|A_r^{(n)} - A_r^*\right\|_s\right) \left\|X\right\|_s + \left\|b^{(n)} - b^*\right\|_s.
\end{aligned}$$

Using conditions (8) and (10), we obtain $\ell_s(Z_n^*, X) = \mathrm{O}(R(n))$. Hence, putting everything together and introducing the notation $p_n := \mathbb{E}\left[\sum_{r=1}^K \mathbf{1}_{\{I_r^{(n)} = n\}} \|A_r^{(n)}\|_{\mathrm{op}}^s\right]$, we obtain from (39) and (40) that

$$\Delta(n) \leq p_n \, \Delta(n) + \mathbb{E}\left[\sum_{r=1}^K \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\mathrm{op}}^s \Delta(I_r^{(n)})\right] + \mathrm{O}(R(n)). \tag{41}$$

From (11), there exists a $\delta > 0$ such that $\mathbb{E}\left[\sum_{r=1}^K \frac{R(I_r^{(n)})}{R(n)} \left\|A_r^{(n)}\right\|_{\mathrm{op}}^s\right] \leq 1 - \delta$ for all n sufficiently large and from (9) we have $p_n < \delta/2$ for n large. We now choose some $C > 0$

and $n_2 \geq n_1$ sufficiently large such that for $n \geq n_2$ all these inequalities are satisfied and the $\mathrm{O}(R(n))$ term in (41) is bounded by $CR(n)$. By setting

$$L := \frac{2C}{\delta} \vee \max\left\{\frac{\Delta(n)}{R(n)} : n \leq n_2\right\}$$

we now obtain $\Delta(n) \leq LR(n)$ by induction: For $n \leq n_2$, by definition of L, the assertion is true. For $n > n_2$, solving for $\Delta(n)$ in (41), we find

$$\begin{aligned}
\Delta(n) &\leq \frac{1}{1-p_n}\left(\mathbb{E}\left[\sum_{r=1}^{K} \mathbf{1}_{\{n_1 \leq I_r^{(n)} < n\}} \|A_r^{(n)}\|_{\mathrm{op}}^s \Delta(I_r^{(n)})\right] + CR(n)\right)\\
&\leq \frac{1}{1-\delta/2}\left(\mathbb{E}\left[\sum_{r=1}^{K} \|A_r^{(n)}\|_{\mathrm{op}}^s LR(I_r^{(n)})\right] + CR(n)\right)\\
&= \frac{1}{1-\delta/2}\left(L\,\mathbb{E}\left[\sum_{r=1}^{K} \|A_r^{(n)}\|_{\mathrm{op}}^s \frac{R(I_r^{(n)})}{R(n)}\right] R(n) + CR(n)\right)\\
&\leq \frac{1}{1-\delta/2}\big(L(1-\delta) + C\big)R(n)\\
&\leq LR(n).
\end{aligned}$$

◀

Proof. (*Proof of Lemma 5*) As the matrix $G_n G_n^T$ is symmetric and positive-semidefinite, we can decompose it in the following way: Let $\lambda_1 \geq \ldots \geq \lambda_m \geq 1 > \lambda_{m+1} \geq \ldots \geq \lambda_d \geq 0$ be the (random) eigenvalues of $G_n G_n^T$. Then, with a suitable (random) orthogonal matrix O, we have

$$\begin{aligned}
G_n G_n^T &= O\,\mathrm{diag}(\lambda_1, \ldots, \lambda_d)\,O^T\\
&= O\,\mathrm{diag}(1, \ldots, 1, \lambda_{m+1}, \ldots, \lambda_d)\,O^T + O\,\mathrm{diag}(\lambda_1 - 1, \ldots, \lambda_m - 1, 0, \ldots, 0)\,O^T\\
&= B_n B_n^T + C_n C_n^T,
\end{aligned}$$

where we define the random $(d \times d)$-matrices $B_n := O\,\mathrm{diag}(1, \ldots, 1, \sqrt{\lambda_{m+1}}, \ldots, \sqrt{\lambda_d})\,O^T$ and $C_n := O\,\mathrm{diag}(\sqrt{\lambda_1 - 1}, \ldots, \sqrt{\lambda_m - 1}, 0, \ldots, 0)\,O^T$. Hence, we can decompose Z_n^* in the following way:

$$Z_n^* \stackrel{d}{=} G_n N + b^{(n)} \stackrel{d}{=} B_n N + C_n N' + b^{(n)} =: \hat{Z}_n^*,$$

where $(B_n, C_n, b^{(n)})$, N and N' are independent with $\mathcal{L}(N) = \mathcal{L}(N') = \mathcal{N}(0, \mathrm{Id}_d)$. Analogously, we decompose the multivariate normal distribution:

$$N \stackrel{d}{=} B_n N + D_n N' =: \hat{N},$$

where $D_n := O\,\mathrm{diag}(0, \ldots, 0, \sqrt{1 - \lambda_{m+1}}, \ldots, \sqrt{1 - \lambda_d})\,O^T$ is chosen such that $B_n B_n^T + D_n D_n^T = \mathrm{Id}_d$.

By definition of the Zolotarev metric ζ_3 we have

$$\zeta_3(Z_n^*, \mathcal{N}(0, \mathrm{Id}_d)) = \zeta_3(\hat{Z}_n^*, \hat{N}) = \sup_{f \in \mathcal{F}_3} \left|\mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]\right|.$$

For arbitrary $f \in \mathcal{F}_3$ we use Taylor expansion around N and obtain for $x \in \mathbb{R}^d$ that

$$f(x) = f(N) + (x - N)^T\,\nabla f(N) + \frac{1}{2}(x - N)^T H_f(N)(x - N) + R(x, N),$$

where the remainder term satisfies $|R(x,N)| \le \frac{1}{2}\|x-N\|^3$. Thus, we have

$$f(\hat{Z}_n^*) - f(\hat{N}) = (\hat{Z}_n^* - \hat{N})^T \nabla f(N) + \frac{1}{2}(\hat{Z}_n^* - N)^T H_f(N)(\hat{Z}_n^* - N) \\ - \frac{1}{2}(\hat{N} - N)^T H_f(N)(\hat{N} - N) + R(\hat{Z}_n^*, N) - R(\hat{N}, N). \qquad (42)$$

We now study the expectation of these summands: For the first summand, we have

$$\mathbb{E}[(\hat{Z}_n^* - \hat{N})^T \nabla f(N)] = \mathbb{E}[\left((C_n - D_n)N' + b^{(n)}\right)^T \nabla f(N)] \\ = \mathbb{E}[(C_n - D_n)N' + b^{(n)}]^T \, \mathbb{E}[\nabla f(N)] = 0,$$

since N is independent of the other quantities, N' is independent of (C_n, D_n) and $\mathbb{E}[N'] = \mathbb{E}[b^{(n)}] = 0$. For the second summand, we define $F_n := B_n - \mathrm{Id}_d$ and obtain

$$\mathbb{E}[(\hat{Z}_n^* - N)^T H_f(N)(\hat{Z}_n^* - N)] \\ = \mathbb{E}[(F_nN + C_nN' + b^{(n)})^T H_f(N)(F_nN + C_nN' + b^{(n)})] \\ = \mathbb{E}[(F_nN)^T H_f(N)(F_nN)] + \mathbb{E}[(F_nN)^T H_f(N)(C_nN')] + \mathbb{E}[(F_nN)^T H_f(N)b^{(n)}] \\ + \mathbb{E}[(C_nN')^T H_f(N)(F_nN)] + \mathbb{E}[(C_nN')^T H_f(N)(C_nN')] + \mathbb{E}[(C_nN')^T H_f(N)b^{(n)}] \\ + \mathbb{E}[(b^{(n)})^T H_f(N)(F_nN)] + \mathbb{E}[(b^{(n)})^T H_f(N)(C_nN')] + \mathbb{E}[(b^{(n)})^T H_f(N)b^{(n)}].$$

Since N, N' and $(F_n, C_n, b^{(n)})$ are independent with $\mathbb{E}[N'] = 0$, we have

$$\mathbb{E}[(F_nN)^T H_f(N)(C_nN')] = 0.$$

The same argument applies to $\mathbb{E}[(C_nN')^T H_f(N)(F_nN)]$, $\mathbb{E}[(C_nN')^T H_f(N)b^{(n)}]$ and $\mathbb{E}[(b^{(n)})^T H_f(N)(C_nN')]$. Analogously, we obtain for the third summand in (42)

$$\mathbb{E}[(\hat{N} - N)^T H_f(N)(\hat{N} - N)] \\ = \mathbb{E}[(F_nN + D_nN')^T H_f(N)(F_nN + D_nN')] \\ = \mathbb{E}[(F_nN)^T H_f(N)(F_nN)] + \mathbb{E}[(D_nN')^T H_f(N)(D_nN')].$$

This implies together with $\mathbb{E}[(F_nN)^T H_f(N)b^{(n)}] = \mathbb{E}[(b^{(n)})^T H_f(N)(F_nN)]$

$$\mathbb{E}[(\hat{Z}_n^* - N)^T H_f(N)(\hat{Z}_n^* - N) - (\hat{N} - N)^T H_f(N)(\hat{N} - N)] \\ = \mathbb{E}[(C_nN')^T H_f(N)(C_nN')] - \mathbb{E}[(D_nN')^T H_f(N)(D_nN')] + \mathbb{E}[(b^{(n)})^T H_f(N)b^{(n)}] \\ + 2\,\mathbb{E}[(F_nN)^T H_f(N)b^{(n)}]$$

Note that we have $C_nC_n^T - D_nD_n^T = G_nG_n^T - \mathrm{Id}_d$. Furthermore, $\mathbb{E}[G_nG_n^T + b^{(n)}(b^{(n)})^T] = \mathrm{Id}_d$. Thus, with the independence of N, N' and $(C_n, D_n, b^{(n)})$ and $\mathbb{E}[N_i'N_j'] = \mathbf{1}_{\{i=j\}}$ for $i, j = 1, \ldots, d$, we have

$$\mathbb{E}[(C_nN')^T H_f(N)(C_nN')] - \mathbb{E}[(D_nN')^T H_f(N)(D_nN')] + \mathbb{E}[(b^{(n)})^T H_f(N)b^{(n)}] \\ = \sum_{i,j=1}^{d} \mathbb{E}[H_f(N)_{ij}]\mathbb{E}[(C_nN')_i(C_nN')_j - (D_nN')_i(D_nN')_j + b_i^{(n)}b_j^{(n)}] \\ = \sum_{i,j=1}^{d} \mathbb{E}[H_f(N)_{ij}]\mathbb{E}[(C_nC_n^T - D_nD_n^T)_{ij} + (b^{(n)}(b^{(n)})^T)_{ij}] \\ = \sum_{i,j=1}^{d} \mathbb{E}[H_f(N)_{ij}]\mathbb{E}[(G_nG_n^T + b^{(n)}(b^{(n)})^T - \mathrm{Id}_d)_{ij}] \\ = 0.$$

Thus, we have shown that

$$\left|\mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]\right| = \left|\mathbb{E}[(F_nN)^T H_f(N) b^{(n)}] + \mathbb{E}[R(\hat{Z}_n^*, N)] - \mathbb{E}[R(\hat{N}, N)]\right|$$
$$\leq \mathbb{E}[|(F_nN)^T H_f(N) b^{(n)}|] + \mathbb{E}[|R(\hat{Z}_n^*, N)|] + \mathbb{E}[|R(\hat{N}, N)|].$$

We now bound these three terms. For this, without loss of generality, we may assume that $H_f(0) = 0$: If this is not the case, we consider the function $g : \mathbb{R}^d \to \mathbb{R}$ defined by $g(x) := f(x) - \frac{1}{2}x^T H_f(0)x$ for $x \in \mathbb{R}^d$. Then, $H_g(0) = 0$ and $\mathbb{E}[g(\hat{Z}_n^*) - g(\hat{N})] = \mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]$ since $\hat{Z}_n^*$ and $\hat{N}$ have the same mean and covariance structure. The assumption $H_f(0) = 0$ implies, together with the Lipschitz property of the second derivative of f, $\|H_f(N)\|_{\text{op}} \leq \|N\|$. Hence, using the Cauchy-Schwarz inequality, the independence of $(F_n, b^{(n)})$ and N and Hölder's inequality, we have

$$\mathbb{E}[|(F_nN)^T H_f(N) b^{(n)}|] \leq \mathbb{E}[\|F_n\|_{\text{op}} \|N\| \|H_f(N)\|_{\text{op}} \|b^{(n)}\|]$$
$$\leq \mathbb{E}[\|N\|^2]\, \mathbb{E}[\|F_n\|_{\text{op}} \|b^{(n)}\|]$$
$$\leq d\, \|F_n\|_{3/2} \|b^{(n)}\|_3$$
$$\leq d\, \|G_nG_n^T - \text{Id}_d\|_{3/2} \|b^{(n)}\|_3,$$

where the last step follows by $\|G_nG_n^T - \text{Id}_d\|_{\text{op}} = \max\{|\lambda_1 - 1|, |\lambda_d - 1|\}$, $\|F_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_d<1\}} |\sqrt{\lambda_d} - 1|$ and the identity $|\sqrt{a} - 1| \leq |a - 1|$ for $a \geq 0$. The first remainder term is bounded by

$$\mathbb{E}[|R(\hat{Z}_n^*, N)|] \leq \frac{1}{2}\, \mathbb{E}\big[\|\hat{Z}_n^* - N\|^3\big]$$
$$= \frac{1}{2}\, \mathbb{E}\big[\|F_nN + C_nN' + b^{(n)}\|^3\big]$$
$$= \mathrm{O}\big(\,\mathbb{E}\big[\|F_n\|_{\text{op}}^3\big] + \mathbb{E}\big[\|C_n\|_{\text{op}}^3\big] + \mathbb{E}\big[\|b^{(n)}\|^3\big]\big)$$
$$= \mathrm{O}\big(\|G_nG_n^T - \text{Id}_d\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3\big),$$

since $\|C_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_1>1\}} \sqrt{|\lambda_1 - 1|} \leq \|G_nG_n^T - \text{Id}_d\|_{\text{op}}^{1/2}$ and $\|F_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_d<1\}} |\sqrt{\lambda_d} - 1| \leq \|G_nG_n^T - \text{Id}_d\|_{\text{op}}^{1/2}$ (note that we have $|\sqrt{a} - 1| \leq \sqrt{|a - 1|}$ for any $a \geq 0$). With the same arguments, we obtain for the second remainder term

$$\mathbb{E}[|R(\hat{N}, N)|] \leq \frac{1}{2}\, \mathbb{E}\big[\|F_nN + D_nN'\|^3\big] = \mathrm{O}\big(\|F_n\|_3^3 + \|D_n\|_3^3\big)$$
$$= \mathrm{O}\big(\|G_nG_n^T - \text{Id}_d\|_{3/2}^{3/2}\big),$$

as $\|D_n\|_{\text{op}} = \mathbf{1}_{\{\lambda_d<1\}} \sqrt{|\lambda_d - 1|} \leq \|G_nG_n^T - \text{Id}_d\|_{\text{op}}^{1/2}$. This implies

$$\left|\mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]\right| \leq \mathbb{E}[|(F_nN)^T H_f(N) b^{(n)}|] + \mathbb{E}[|R(\hat{Z}_n^*, N)|] + \mathbb{E}[|R(\hat{N}, N)|]$$
$$= \mathrm{O}\big(\|G_nG_n^T - \text{Id}_d\|_{3/2} \|b^{(n)}\|_3 + \|G_nG_n^T - \text{Id}_d\|_{3/2}^{3/2} + \|b^{(n)}\|_3^3\big)$$
$$= \mathrm{O}(R(n)).$$

Note that the constants in the O-notation do not depend on the function f, i.e. we have $\sup_{f \in \mathcal{F}_3} \left|\mathbb{E}[f(\hat{Z}_n^*) - f(\hat{N})]\right| = \mathrm{O}(R(n))$. ◀