ON THE RATE OF CONVERGENCE AND BERRY-ESSEEN TYPE THEOREMS FOR A MULTIVARIATE FREE CENTRAL LIMIT THEOREM

ROLAND SPEICHER (†)

Abstract. We address the question of a Berry Esseen type theorem for the speed of convergence in a multivariate free central limit theorem. For this, we estimate the difference between the operator-valued Cauchy transforms of the normalized partial sums in an operator-valued free central limit theorem and the Cauchy transform of the limiting operator-valued semicircular element.

1. Introduction

The free central limit theorem (due to Voiculescu [12] in the one-dimensional case, and to Speicher [10] in the multivariate case) is one of the basic results in free probability theory. Investigations on the speed of convergence to the limiting semicircular distribution, however, were taken up only recently. In the classical context, the analogous question is answered by the famous Berry-Esseen theorem, which states, in its simplest version, the following: If X_i are i.i.d. random variables, with mean zero and variance 1, then the distance between $S_n := (X_1 + \cdots + X_n)/\sqrt{n}$ and a normal variable γ of mean zero and variance 1 can be estimated in terms of the Kolmogorov distance Δ by

$$\Delta(S_n, \gamma) \leq C \frac{1}{\sqrt{n}} \rho,$$

where C is a constant and ρ is the absolute third moment of the variables x_i.

The question for a free analogue of the Berry-Esseen estimate in the case of one random variable was answered by Chistyakov and Götze [3]:

† Research supported by Discovery and LSI grants from NSERC (Canada) and by a Killam Fellowship from the Canada Council for the Arts.

This project was initiated by disussions with Friedrich Göetze during my visit at the University of Bielefeld in November 2006. I thank the Department of Mathematics and in particular the SFB 701 for its generous hospitality and Friedrich Götze for the invitation and many interesting discussions.

I also thank Uffe Haagerup for pointing out how ideas from [5] can be used to improve the results from an earlier version of this paper.

If x_i are free identically distributed random variables with mean zero and variance 1, then the distance between $S_n := (X_1 + \cdots + X_n)/\sqrt{n}$ and a semicircular variable s of mean zero and variance 1 can, under the assumption of finite fourth moment, be estimated as

$$\Delta(S_n, s) \leq c\frac{|m_3| + \sqrt{m_4}}{\sqrt{n}},$$

where $c > 0$ is an absolute constant, and m_3 and m_4 are the third and fourth moment, respectively, of the x_i. (Independently, the same kind of question was considered, under the more restrictive assumption of compact support for the x_i, by Kargin [8].)

In this paper we want to address the multivariate version of a free Berry-Esseen theorem. In contrast to the classical situation, the multivariate situation is of a quite different nature than the one-dimensional case, because we have to deal with non-commuting operators and all the analytical tools, which are available in the one-dimensional case, break down. However, we are able to deal with this situation by invoking recent ideas of Haagerup and Thorbjornsen [6, 5], in particular, their linearization trick which allows to reduce the multivariate (scalar-valued) to an analogous one-dimensional operator-valued problem. Estimates for the operator-valued Cauchy transform of this operator-valued operator are quite similar to estimates in the scalar-valued case. Actually, on the level of deriving equations for these Cauchy transforms we can follow ideas which are used for dealing with speed of convergence questions for random matrices; here we are inspired in particular by the work of Götze and Tikhomirov [4], but see also [1, 2]. Our main theorem on the speed of convergence in an operator-valued free central limit theorem is the following.

Theorem 1. *Let $1 \in \mathcal{B} \subset \mathcal{A}$, $E : \mathcal{A} \to \mathcal{B}$ be an operator-valued probability space. Consider selfadjoint $X_1, X_2, \cdots \in \mathcal{A}$ which are free with respect to E and have identical $\mathcal{B}$-valued distribution. Assume that the first moments vanish,*

$$E[X_i] = 0$$

and let

$$\eta : \mathcal{B} \to \mathcal{B}, \qquad \eta(b) = E[X_i b X_i]$$

be their covariance. Denote

$$\alpha_2 := \sup_{\substack{b\in\mathcal{B}\\ \|b\|=1}} \|E[X_i b X_i]\| = \|\eta\|$$

and

$$\alpha_4 := \sup_{\substack{b\in\mathcal{B}\\ \|b\|=1}} \|E[X_i b X_i X_i b^* X_i]\|.$$

Consider now the normalized sums

$$S_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

and their $\mathcal{B}$-valued Cauchy transforms

$$G_n(b) := E[\frac{1}{b - S_n}] \qquad (b \in \mathcal{B}_+)$$

on the "upper half plane" $\mathcal{B}_+$ in $\mathcal{B}$,

$$\mathcal{B}_+ := \{b \in \mathcal{B} \mid \operatorname{Im} b \geq 0 \text{ and } \operatorname{Im} b \text{ invertible}\}.$$

By G we denote the operator-valued Cauchy transform of a $\mathcal{B}$-valued semicircular element with covariance η.

Then we have for all $b \in \mathcal{B}_+$ and all $n \in \mathbb{N}$ that

$$\|G_n(b) - G(b)\| \leq 4c_n(b) \left(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\|\right) \cdot \|\frac{1}{\operatorname{Im} b}\|^2, \tag{1}$$

where

$$c_n(b) := \frac{1}{\sqrt{n}}\|\frac{1}{\operatorname{Im} b}\|^3 \sqrt{\alpha_2} \cdot (2\alpha_2 + \sqrt{\alpha_4 + 2\alpha_2^2}) + \frac{1}{n}\|\frac{1}{\operatorname{Im} b}\|^4 \alpha_2^2.$$

In the one-dimensional scalar case one can derive from such estimates corresponding estimates for the Kolmogorov distance between the distribution of S_n and the limiting semicircle s. This relies on the fact that the Kolmogorov metric measures how close the distribution functions of two measures are, and the Stieltjes inversion formula allows to relate the distribution function with Cauchy transforms. (In the proof of the classical Berry-Esseen theorem one follows a similar route, using Fourier transforms instead of Cauchy transforms.) For the multivariate case, say of d variables, where we would like to say something about the speed of convergence of the d-tuple of partial sums $(S_n^{(1)}, \ldots, S_n^{(d)})$ to the limiting semicircular family $(s_1, \ldots, s_d)$, there is no nice replacement for the distribution function, and we also do not know of a canonical metric on joint distributions of several non-commuting variables which relates directly with the above estimates for operator-valued Cauchy transforms.

However, there is a kind of replacement for this; namely, following again [5], estimates for Cauchy transforms of linear combinations with operator-valued coefficients of the variables $(S_n^{(1)}, \ldots, S_n^{(d)})$ should imply corresponding estimates for any non-commutative scalar polynomial in those variables and from those one should be able to estimate, for any selfadjoint non-commutative polynomial p, the Levy distance between $p(S_n^{(1)}, \ldots, S_n^{(d)})$ and $p(s_1, \ldots, s_d)$. However, one has to deal with the following problem in such an approach: as is shown in [5] one can get the Cauchy transform of a polynomial $p(s_1,\dots,s_d)$ as a corner of an operator-valued Cauchy transform of a linear combination P, with matrix-valued coefficients, of $s_1,\dots,s_d$; but, even if p is a selfadjoint polynomial, the corresponding matrix-valued operator P is not selfadjoint, and thus our operator-valued estimates, which were only shown for selfadjoint X, cannot be used directly for P; one would have to reprove most of our statements also for P. It is conceivable that this can be done in a similar manner as in [5]; as this approach is getting quite technical, we will pursue the details in a forthcoming investigation.

Note that for proving such a kind of Berry-Esseen theorem for polynomials $p(s_1,\dots,s_d)$ one also has to face another kind of question: estimates for the difference of Cauchy transforms translate directly only in estimates for the Levy distance between the corresponding measures; in order to get also estimates for the more intuitive Kolmogorov distance one needs to know that the distribution of $p(s_1,\dots,s_d)$ has a continuous density, in particular, has no atoms. We conjecture that this is true for all non-commutative selfadjoint polynomials p in a semicircular family, but this seems to be a non-trivial problem. Note that the question of absence of atoms can be seen as an analogue of the Zero-Divisor Theorem for the free group. We hope to address this question in some future work.

The paper is organized as follows. In the next section we will first relate a multivariate free central limit theorem with a one-dimensional operator-valued free central limit theorem. The proof of Theorem 1 will be given in Section 3.

2. Multivariate free central limit theorem

2.1. **Setting.** Let $\left(x_1^{(k)}\right)_{k=1}^d, \left(x_2^{(k)}\right)_{k=1}^d, \dots$ be free and identically distributed sets of k selfadjoint random variables in some non-commutative probability space $(\mathcal{C},\varphi)$, such that the first moments vanish and the second moments are given by a covariance matrix $\Sigma=(\sigma_{kl})_{k,l=1}^d$. We put

$$S_n^{(k)}=\frac{x_1^{(k)}+\dots+x_n^{(k)}}{\sqrt{n}}.$$

We know [10] that $(S_n^{(1)},\dots,S_n^{(d)})$ converges in distribution for $n\to\infty$ to a semicircular family $(s_1,\dots,s_d)$ of covariance Σ. We want to analyze the rate of this convergence. We would like to get an estimate which involves only small moments of the given variables. As we will see, the second and fourth moments of our variables will show up in the estimates and we will use the upper bound

$$\beta_2 := \max_{k,l} |\sigma_{k,l}| = \max_{k,l} \varphi(x_i^{(k)} x_i^{(l)})$$

for the second and the upper bound

$$\beta_4 := \max_{r,p,k,l} |\varphi(x_i^{(r)} x_i^{(p)} x_i^{(k)} x_i^{(l)})|$$

for the fourth moments.

2.2. **Transition to operator-valued frame.** We will analyze the rate of convergence of the multivariate problem,

$$(S_n^{(1)}, \dots, S_n^{(d)}) \to (s_1, \dots, s_n)$$

by replacing this by an one-dimensional operator-valued problem. The underlying idea for that is the linearization trick [6, 5] that one can understand the joint distribution of several scalar random variables by understanding the distribution of each operator-valued linear combination of those random variables.

Let $\mathcal{B} = M_N(\mathbb{C})$ and put $\mathcal{A} := M_N(\mathbb{C}) \otimes \mathcal{C} = M_N(\mathcal{C})$. Then $\mathcal{B} \cong \mathcal{B} \otimes 1 \subset \mathcal{A}$ is an operator-valued probability space with respect to the conditional expectation

$$E = \mathrm{id} \otimes \varphi : \mathcal{B} \otimes \mathcal{C} \to \mathcal{B}, \qquad b \otimes c \mapsto \varphi(c)b.$$

For some fixed $b_1, \dots, b_k \in M_N(\mathbb{C})$ we put

$$X_i := \sum_{k=1}^{d} b_k \otimes x_i^{(k)}$$

and

$$S_n := \sum_{k=1}^{d} b_k \otimes S_n^{(k)}$$

Note that $X_1, X_2, \cdots$ are free with respect to E and that we have

$$S_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}.$$

The limit of S_n is

$$s := \sum_{k=1}^{d} b_k \otimes s_k,$$

which is an $\mathcal{B} = M_N(\mathbb{C})$-valued semicircular element with covariance mapping $\eta : \mathcal{B} \to \mathcal{B}$ given by

$$\eta(b) = E[sb \otimes 1s] = \sum_{k,l=1}^{d} E[b_k \otimes s_k \cdot b \otimes 1 \cdot b_l \otimes s_l]$$
$$= \sum_{k,l=1}^{d} b_k b b_l \varphi(s_k s_l) = \sum_{k,l=1}^{d} b_k b b_l \sigma_{kl}.$$

We want to determine the rate of convergence for S_n to s. We will do this in the next section in the context of a general operator-valued free central limit theorem.

3. Rate of convergence for operator-valued free central limit theorem

3.1. **Setting.** Let $1 \in \mathcal{B} \subset \mathcal{A}$, $E : \mathcal{A} \to \mathcal{B}$ be an operator-valued probability space. This means that $\mathcal{A}$ is a von Neumann algebra, $\mathcal{B}$ is a sub von Neumann algebra, which contains the identity of $\mathcal{A}$, and E is a conditional expectation from $\mathcal{A}$ onto $\mathcal{B}$, i.e., a linear map which satisfies the property

$$E[b_1 a b_2] = b_1 E[a] b_2$$

for all $a \in \mathcal{A}$ and $b_1, b_2 \in \mathcal{B}$.

Consider selfadjoint $X_1, X_2, \dots \in \mathcal{A}$ which are free with respect to E and have identical $\mathcal{B}$-valued distribution. Assume that the first moments vanish,

$$E[X_i] = 0$$

and let

$$\eta : \mathcal{B} \to \mathcal{B}, \qquad \eta(b) = E[X_i b X_i]$$

be their covariance. We will need

$$\alpha_2 := \sup_{\substack{b \in \mathcal{B} \\ \|b\|=1}} \|E[X_i b X_i]\| = \|\eta\|$$

and

$$\alpha_4 := \sup_{\substack{b \in \mathcal{B} \\ \|b\|=1}} \|E[X_i b X_i X_i b^* X_i]\|.$$

Consider now the normalized sums

$$S_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}}.$$

We know that S_n converges in distribution to an operator-valued semicircular element s with covariance η, see [11]

We want to estimate the rate of this convergence. Let us denote by $\mathcal{B}_+$ the "upper half plane" in $\mathcal{B}$, i.e.,

$$\mathcal{B}_+ := \{b \in \mathcal{B} \mid \operatorname{Im} b \geq 0 \text{ and } \operatorname{Im} b \text{ invertible}\}.$$

We consider, for $b \in \mathcal{B}_+$, the resolvents

$$R_n(b) := \frac{1}{b - S_n}, \qquad R(b) := \frac{1}{b - s}$$

and the Cauchy transforms

$$G_n(b) := E[R_n(b)], \qquad G(b) := E[R(b)].$$

G_n and G are analytic functions in $\mathcal{B}_+$.

3.2. **The main estimates.** We will show that $G_n(b)$ converges to $G(b)$, where we have good control over the difference in terms of n and b. The idea for showing this is the same as in [6]. First we show that G_n satisfies an approximate version of an equation satisfied by G and then we show that this actually implies that G_n and G must be close to each other.

Let us start with deriving the equations for G and G_n.

Since s is an operator-valued semicircular element with covariance η we know [13, 11] that its Cauchy transform satisfies the equation

$$bG(b) - 1 = \eta\left(G(b)\right) \cdot G(b). \tag{2}$$

We want to derive an approximate version of this equation for G_n. For this, we will look at $E[S_n R_n(b)]$.

Let us denote by $S_n^{[i]}$ the version of S_n where the i-th variable X_i is absent, i.e.,

$$S_n^{[i]} := S_n - \frac{1}{\sqrt{n}} X_i,$$

and by $R_n^{[i]}$ and $G_n^{[i]}$ the corresponding resolvent and Cauchy transform, respectively, i.e.,

$$R_n^{[i]}(b) = \frac{1}{b - S_n^{[i]}}$$

and

$$G_n^{[i]}(b) := E[R_n^{[i]}(b)].$$

For each $i = 1, \dots, n$ we have the resolvent identity

$$R_n(b) = R_n^{[i]}(b) + \frac{1}{\sqrt{n}} R_n^{[i]}(b) \cdot X_i \cdot R_n^{[i]}(b) + \frac{1}{n} R_n(b) \cdot X_i \cdot R_n^{[i]}(b) \cdot X_i \cdot R_n^{[i]}(b).$$

Now we can write

$$
\begin{aligned}
E[S_nR_n(b)] &= \sum_{i=1}^n E\big[\frac{X_i}{\sqrt{n}}\cdot R_n(b)\big]\\
&= \sum_{i=1}^n \frac{1}{\sqrt{n}}\Big\{E\big[X_i\cdot R_n^{[i]}(b)\big]\\
&\quad + \frac{1}{\sqrt{n}}E\big[X_i\cdot R_n^{[i]}(b)\cdot X_i\cdot R_n^{[i]}(b)\big]\\
&\quad + \frac{1}{n}E\big[X_i\cdot R_n(b)\cdot X_i\cdot R_n^{[i]}(b)\cdot X_i\cdot R_n^{[i]}(b)\big]\Big\}
\end{aligned}
$$

Now we use our assumption that $X_1, X_2, \dots$ are free with respect to E, which implies that X_i is free from $R_n^{[i]}(b)$ with respect to E. This implies that

$$E[X_i\cdot R_n^{[i]}(b)] = E[X_i]\cdot E[R_n^{[i]}(b)] = 0$$

and

$$
\begin{aligned}
E\left[X_i\cdot R_n^{[i]}(b)\cdot X_i\cdot R_n^{[i]}(b)\right] &= E\left[X_i\cdot E[R_n^{[i]}(b)]\cdot X_i\right]\cdot E\left[R_n^{[i]}(b)\right]\\
&\quad + E[X_i]\cdot E\left[R_n^{[i]}(b)\cdot E[X_i]\cdot R_n^{[i]}(b)\right]\\
&\quad - E[X_i]\cdot E[R_n^{[i]}(b)]\cdot E[X_i]\cdot E[R_n^{[i]}(b)]\\
&= E\left[X_i\cdot E[R_n^{[i]}(b)]\cdot X_i\right]\cdot E[R_n^{[i]}(b)]\\
&= \eta\left(G_n^{[i]}(b)\right)\cdot G_n^{[i]}(b).
\end{aligned}
$$

So we have got finally

$$E[S_nR_n(b)] = \frac{1}{n}\left(\sum_{i=1}^n \eta\left(G_n^{[i]}(b)\right)\cdot G_n^{[i]}(b) + r_1^{[i]}\right), \tag{3}$$

where

$$r_1^{[i]} = \frac{1}{\sqrt{n}}E\left[X_i\cdot R_n(b)\cdot X_i\cdot R_n^{[i]}(b)\cdot X_i\cdot R_n^{[i]}(b)\right]$$

We will now estimate the norm of $r_1^{[i]}$. We could of course just estimate against the operator norm of X_i; however, we prefer, in analogy with the classical case, to do better without invoking the operator norm and use only as small moments of X_i as possible.

Note that for our conditional expectation E we have the Cauchy-Schwarz inequality

$$\|E[AB]\|^2 \le \|E[AA^*]\|\cdot\|E[B^*B]\|,$$

and also

$$E[A]^*E[A] \le E[A^*A] \qquad \text{and} \qquad E[ABB^*A^*] \le \|BB^*\|\cdot E[AA^*]$$

and

$$\|E[A]\| \leq \|A\|$$

for any $A, B \in \mathcal{A}$. Thus, for any $i = 1, \dots, n$, we can estimate

$$\begin{aligned}
&\|E\big[X_i\, R_n(b)\, X_i\, R_n^{[i]}(b)\, X_i\, R_n^{[i]}(b)\big]\|^2 \\
&\qquad \leq \|E\big[X_i\, R_n(b)\, R_n(b)^*\, X_i\big]\| \cdot \\
&\qquad\quad \cdot \Big\|E\big[R_n^{[i]}(b)^*\, X_i\, R_n^{[i]}(b)^*\, X_i\, X_i\, R_n^{[i]}(b)\, X_i\, R_n^{[i]}(b)\big]\Big\|
\end{aligned}$$

We estimate the first factor by

$$\begin{aligned}
\|E\big[X_i\, R_n(b)\, R_n(b)^*\, X_i\big]\| &\leq \|R_n(b)\|^2 \cdot \big\|E\big[X_i X_i\big]\big\| \\
&= \|R_n(b)\|^2 \cdot \|\eta(1)\| \\
&= \alpha_2 \|R_n(b)\|^2
\end{aligned}$$

For the second factor we use again the freeness between X_i and $R_n^{[i]}(b)$. Let us put

$$R := R_n^{[i]}(b)$$

Then X_i and R are $*$-free with respect to E and thus, by also invoking $E[X_i] = 0$, we have

$$\begin{aligned}
E[R^* X_i R^* X_i X_i R X_i R] &= E\Big[R^* \cdot E\big[X_i\, E[R^*]\, X_i\, X_i\, E[R]\, X_i\big] \cdot R\Big] \\
&\quad + E\Big[R^* \cdot \eta\big(E[R^*\, \eta(1)\, R]\big) \cdot R\Big] \\
&\quad - E\Big[R^* \cdot \eta\big(E[R^*]\, \eta(1)\, E[R]\big) \cdot R\Big],
\end{aligned}$$

and thus

$$\begin{aligned}
\big\|E\big[R^* X_i R^* X_i X_i R X_i R\big]\big\| &\leq \Big\|E\Big[R^* \cdot E\big[X_i\, E[R^*]\, X_i\, X_i\, E[R]\, X_i\big] \cdot R\Big]\Big\| \\
&\quad + \Big\|E\Big[R^* \cdot \eta\big(E[R^*\, \eta(1)\, R]\big) \cdot R\Big]\Big\| \\
&\quad + \Big\|E\Big[R^* \cdot \eta\big(E[R^*]\, \eta(1)\, E[R]\big) \cdot R\Big]\Big\|
\end{aligned}$$

We estimate

$$\begin{aligned}
&\Big\|E\Big[R^* \cdot E\big[X_i\, E[R^*]\, X_i\, X_i\, E[R]\, X_i\big] \cdot R\Big]\Big\| \\
&\qquad \leq \|R\| \cdot \|R^*\| \cdot \big\|E\big[X_i\, E[R^*]\, X_i\, X_i\, E[R]\, X_i\big]\big\| \\
&\qquad \leq \|R\|^2 \cdot \alpha_4 \cdot \|E[R]\| \cdot \|E[R^*]\| \\
&\qquad \leq \alpha_4 \cdot \|R\|^4
\end{aligned}$$

$$\Big\|E\Big[R^*\, \eta\big(E[R^*\, \eta(1)\, R]\big)\, R\Big]\Big\| \leq \alpha_2^2 \cdot \|R\|^4,$$

and

$$\Big\| E\Big[R^* \cdot \eta\big(E[R^*]\,\eta(1)\,E[R]\big) \cdot R\Big]\Big\| \le \alpha_2^2 \cdot \|R\|^4$$

Putting this together yields

$$\big\| E\big[R_n^{[i]}(b)^*\, X_i\, R_n^{[i]}(b)^*\, X_i\, X_i\, R_n^{[i]}(b)\, X_i\, R_n^{[i]}(b)\big]\big\| \le (\alpha_4 + 2\alpha_2^2) \cdot \|R_n^{[i]}(b)\|^4,$$

and finally

$$\|r_1^{[i]}\| \le \frac{1}{\sqrt{n}} \cdot \sqrt{\alpha_2(\alpha_4 + 2\alpha_2^2)} \cdot \|R_n(b)\| \cdot \|R_n^{[i]}(b)\|^2.$$

We still need to replace, in (3), $G_n^{[i]}(b) = E[R_n^{[i]}(b)]$ by $G_n(b) = E[R_n(b)]$. By using the resolvent identity

$$R_n(b) = R_n^{[i]}(b) + \frac{1}{\sqrt{n}} R_n^{[i]}(b) \cdot X_i \cdot R_n(b)$$

we have

$$G_n^{[i]}(b) = G_n(b) + r_2^{[i]},$$

where

$$r_2^{[i]} := -\frac{1}{\sqrt{n}} E[R_n^{[i]}(b)\, X_i\, R_n(b)].$$

As before, we estimate

$$\begin{aligned}\|E[R_n^{[i]}(b)\, X_i\, R_n(b)]\|^2 &\le \|E[R_n^{[i]}(b)\, X_i\, X_i\, R_n^{[i]}(b)^*]\| \cdot \|E[R_n(b)^*\, R_n(b)]\| \\ &\le \alpha_2 \cdot \|R_n^{[i]}(b)\|^2 \cdot \|R_n(b)\|^2.\end{aligned}$$

Let us summarize. We have

$$\begin{aligned}E[S_n R_n(b)] &= \frac{1}{n}\sum_{i=1}^n \left(\eta\left(G_n^{[i]}(b)\right) \cdot G_n^{[i]}(b) + r_1^{[i]}\right) \\ &= \frac{1}{n}\sum_{i=1}^n \left(\eta\left(G_n(b) + r_2^{[i]}\right) \cdot \left(G_n(b) + r_2^{[i]}\right) + r_1^{[i]}\right),\end{aligned}$$

and the estimates

$$\|r_1^{[i]}\| \le \frac{1}{\sqrt{n}} \cdot \sqrt{\alpha_2(\alpha_4 + 2\alpha_2^2)} \cdot \|R_n(b)\| \cdot \|R_n^{[i]}(b)\|^2$$

and

$$\|r_2^{[i]}\| \le \frac{1}{\sqrt{n}} \sqrt{\alpha_2} \cdot \|R_n^{[i]}(b)\| \cdot \|R_n(b)\|.$$

It remains to estimate $\|R_n(b)\|$ and $\|R_n^{[i]}(b)\|$. For those we use the usual estimate for Cauchy transforms (where $\operatorname{Im} b := (b - b^*)/(2i)$ denotes the imaginary part of b),

$$\|R_n(b)\| \le \|\frac{1}{\operatorname{Im} b}\|, \qquad \|R_n^{[i]}(b)\| \le \|\frac{1}{\operatorname{Im} b}\|.$$

For a formal proof of this estimate, see, e.g., Lemma 3.1 in [6].

We have now

$$E[S_n R_n(b)] = \eta\left(G_n(b)\right) \cdot G_n(b) + r_3,$$

where

$$r_3 = \frac{1}{n}\sum_{i=1}^{n}\Big(\eta(G_n(b)) \cdot r_2^{[i]} + \eta(r_2^{[i]}) \cdot G_n(b) + \eta(r_2^{[i]}) \cdot r_2^{[i]} + r_1^{[i]}\Big).$$

Hence

$$\|r_3\| \leq \frac{1}{n}\sum_{i=1}^{n}\Big(2\|\eta\| \cdot \|G_n(b)\| \cdot \|r_2^{[i]}\| + \|\eta\| \cdot \|r_2^{[i]}\|^2 + \|r_1^{[i]}\|\Big) \leq c_n,$$

where

$$c_n := c_n(b) := \frac{1}{\sqrt{n}}\Big\|\frac{1}{\operatorname{Im} b}\Big\|^3\sqrt{\alpha_2} \cdot (2\alpha_2 + \sqrt{\alpha_4 + 2\alpha_2^2}) + \frac{1}{n}\Big\|\frac{1}{\operatorname{Im} b}\Big\|^4 \alpha_2^2.$$

Note that $S_n R_n(b) = -1 + bR_n(b)$, hence

$$E[S_n R_n(b)] = bG_n(b) - 1,$$

and so we finally have found

$$\eta(G_n(b)) \cdot G_n(b) - bG_n(b) + 1 = -r_3, \tag{4}$$

or the inequality:

$$\|\eta(G_n(b)) \cdot G_n(b) - bG_n(b) + 1\| \leq c_n. \tag{5}$$

In order to get from this an estimate for the difference between $G_n(b)$ and $G(b)$, we will now follow the ideas in Section 5 of [6], in the improved version from [5].

By (2), we have for all $b \in \mathcal{B}_+$ the equation

$$b = \frac{1}{G(b)} + \eta\big(G(b)\big) \tag{6}$$

for $G(b)$, and, by (4), the corresponding approximate version for $G_n(b)$:

$$\Lambda_n(b) = \frac{1}{G_n(b)} + \eta\big(G_n(b)\big), \tag{7}$$

where

$$\Lambda_n(b) := b - r_3 \cdot G_n(b)^{-1}.$$

A crucial point is now to show that for a sufficiently large set $\tilde{O}_n \subset \mathcal{B}_+$ the quantity $\operatorname{Im} \Lambda_n(b)$ is still positive, so that we can also use equation

(6) for $\Lambda_n(b)$. Let us try

$$\tilde{O}_n := \Big\{ b \in \mathcal{B}_+ \mid c_n(b) < 1/2 \quad \text{and} \\ c_n(b) \cdot \Big(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\| \Big) \cdot \|\frac{1}{\operatorname{Im} b}\| < 1/2 \Big\}.$$

The relevance of the condition $c_n(b) < 1/2$ is the following: Let us denote

$$B_n(b) := b - \eta(G_n(b)),$$

then inequality (5) takes, for $b \in \tilde{O}_n$, the form

$$\|1 - B_n(b)G_n(b)\| \le c_n(b) < 1/2.$$

This, however, implies that $B_n(b)G_n(b)$ is invertible with

$$\|G_n(b)^{-1}B_n(b)^{-1}\| = \|(B_n(b)G_n(b))^{-1}\| \le 2,$$

and thus

$$\begin{aligned} \|G_n(b)^{-1}\| &= \|G_n(b)^{-1}B_n(b)^{-1}B_n(b)\| \\ &\le 2\|B_n(b)\| \\ &= 2\|b - \eta(G_n(b))\| \\ &\le 2\left(\|b\| + \alpha_2 \cdot \|G_n(b)\|\right) \\ &\le 2\left(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\| \right). \end{aligned}$$

But then the other condition in the definition of $\tilde{O}_n$ implies that for $b \in \tilde{O}_n$ we have

$$\begin{aligned} \|r_3 \cdot G_n(b)^{-1}\| &\le \|r_3\| \cdot \|G_n(b)^{-1}\| \\ &\le c_n \cdot 2\left(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\| \right) < \|\frac{1}{\operatorname{Im} b}\|^{-1}. \end{aligned} \tag{8}$$

Since

$$\operatorname{Im} b \ge \|\frac{1}{\operatorname{Im} b}\|^{-1} \cdot 1,$$

it follows that, for $b \in \tilde{O}_n$, $\Lambda_n(b) = b - r_3 \cdot G_n(b)^{-1}$ is still in $\mathcal{B}_+$ and so we can use the equation (6) with $\Lambda_n(b)$ as argument, i.e.,

$$\Lambda_n(b) = \frac{1}{G(\Lambda_n(b))} + \eta\big(G(\Lambda_n(b))\big). \tag{9}$$

The point of having both equation (9) and equation (7) is that this implies that

$$G(\Lambda_n(b)) = G_n(b).$$

In [6, 5] this was shown by analytic continuation arguments. We can simplify that argument by using the fact from [7] that the equation

$$w = \frac{1}{G} + \eta(G) \tag{10}$$

has, for any w with $\operatorname{Im} w > 0$, exactly one solution $G \in \mathcal{B}$ such that $\operatorname{Im} G$ is negative. Since both $G_n(b)$ and $G(\Lambda_n(b))$ have negative imaginary parts (as Cauchy transforms at some arguments) and both satisfy the same equation (10) (for $w = \Lambda_n(b)$), they must agree.

Then we can, still in the case $b \in \tilde{O}_n$, estimate in the usual way, by invoking the resolvent identity:

$$\begin{aligned} \|G_n(b) - G(b)\| &= \|G(\Lambda_n(b)) - G(b)\| \\ &= \|G(\Lambda_n(b)) \cdot (\Lambda_n(b) - b) \cdot G(b)\| \\ &\leq \|(\Lambda_n(b) - b)\| \cdot \|G_n(b)\| \cdot \|G(b)\|. \end{aligned}$$

Both $\|G(b)\|$ and $\|G_n(b)\|$ can be estimated by $\|1/\operatorname{Im} b\|$ and for the first factor we have, by the second inequality in (8), that

$$\|(\Lambda_n(b) - b)\| = \|r_3 G_n(b)^{-1}\| \leq c_n \cdot 2 \left(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\| \right)$$

Thus, for $b \in \tilde{O}_n$, we have shown that

$$\|G_n(b) - G(b)\| \leq c_n \cdot 2 \left(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\| \right) \cdot \|\frac{1}{\operatorname{Im} b}\|^2 \tag{11}$$

For $b \in \mathcal{B}_+ \backslash \tilde{O}_n$, on the other hand, we just use the trivial estimate

$$\|G_n(b) - G(b)\| \leq 2 \cdot \|\frac{1}{\operatorname{Im} b}\|$$

together with

- if we have $c_n(b) \geq 1/2$, then
$$\begin{aligned} \|\frac{1}{\operatorname{Im} b}\| &\leq 2c_n \cdot \|\frac{1}{\operatorname{Im} b}\| \\ &\leq 2c_n \cdot \|\frac{1}{\operatorname{Im} b}\| \cdot \|b\| \cdot \|\frac{1}{\operatorname{Im} b}\| \\ &\leq 2c_n \cdot \|\frac{1}{\operatorname{Im} b}\|^2 \cdot \left(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\| \right) \end{aligned}$$
- if we have $c_n(b) \cdot \left(\|b\| + \alpha_2 \cdot \left\|\frac{1}{\operatorname{Im} b}\right\|\right) \cdot \left\|\frac{1}{\operatorname{Im} b}\right\| \geq 1/2$, then we have again
$$\|\frac{1}{\operatorname{Im} b}\| \leq 2c_n \cdot \left(\|b\| + \alpha_2 \cdot \|\frac{1}{\operatorname{Im} b}\| \right) \cdot \|\frac{1}{\operatorname{Im} b}\|^2$$

Thus we have proved the Theorem.

References

[1] Z.D. Bai: Convergence rate of expected spectral distributions of large random matrices. Part I. Wigner Matrices. Ann. Prob. 21 (1993), 625–648.

[2] Z.D. Bai: Methodologies in spectral analysis of large dimensional random matrices, a review. Statistica Sinica 9 (1999), 611-677.

[3] G.P. Chistyakov, F. Götze: Limit theorems in free probability theory. I. Preprint 2006, math-archive 0602219.

[4] F. Götze, A. Tikhomirov: Limit theorems for spectra of random matrices with martingale structure. Stein's method and applications, 181-193, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., 5, Singapore Univ. Press, Singapore, 2005

[5] U. Haagerup, H. Schultz, S. Thorbjornsen: A random matrix approach to the lack of projections in $C^*_{\mathrm{red}}(\mathbb{F}_2)$. Adv. Math. 204 (2006), 1–83.

[6] U. Haagerup, S. Thorbjornsen: A new application of Random Matrices: $Ext(C^*_{red}(F_2))$ is not a group. Annals of Mathematics 162, 2005.

[7] W. Helton, R. Rashidi Far, R. Speicher: Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints. Preprint, 2007, math.0A/0703510

[8] V. Kargin: Berry-Esseen for free random variables. J. Theor. Probab. 20 (2007), 381–395.

[9] A. Nica, R. Speicher: Lectures on the Combinatorics of Free Probabiltiy. London Mathematical Society Lecture Note Series, no. 335. Cambridge University Press, 2006.

[10] R. Speicher: A New Example of Independence and White Noise. Prob. Th. Rel. Fields 84 (1990), 141–159.

[11] R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," *Mem. Amer. Math. Soc.*, vol. 132, no. 627, pp. x+88, 1998.

[12] D. Voiculescu: Addition of certain non-commuting random variables. J. Funct. Anal. 66 (1986), 323–346.

[13] D. Voiculescu, "Operations on certain non-commutative operator-valued random variables," *Astérisque*, no. 232, pp. 243–275, 1995, recent advances in operator algebras (Orléans, 1992).

Queen's University, Department of Mathematics and Statistics, Jeffery Hall, Kingston, ON, K7L 3N6, Canada

E-mail address: speicher@mast.queensu.ca