Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 4, 2565–2594
https://doi.org/10.1214/20-AIHP1050

ANNALES DE L'INSTITUT HENRI POINCARÉ PROBABILITÉS ET STATISTIQUES

Subordination methods for free deconvolution

Octavio Arizmendi[a], Pierre Tarrago[b] and Carlos Vargas[c]

[a]*Department of Probability and Statistics, CIMAT, Guanajuato, Mexico. E-mail: octavius@cimat.mx*
[b]*Faculté des Sciences et Ingénierie, PARIS cedex 05, Paris, France. E-mail: pierre.tarrago@upmc.fr*
[c]*Catedras CONACYT-CIMAT, Guanajuato, Mexico. E-mail: carlosv@cimat.mx*

Received 30 July 2018; revised 8 December 2019; accepted 19 February 2020

Abstract. We derive subordination functions for free additive and free multiplicative deconvolutions under mild moment conditions. Our results include an algorithm to calculate these subordination functions, and thus the associated Cauchy transforms, for complex numbers with imaginary part greater than a parameter depending on the measure to deconvolve. The existence of these subordination functions on such domains reduces the problem of free deconvolutions to the problem of the classical additive deconvolution with a Cauchy distribution. Thus, our results, combined with known methods for the deconvolution with a Cauchy distribution, allow us to solve the free deconvolution problem. We also present extensions of these results to the case of operator-valued deconvolutions.

Résumé. Nous dérivons des fonctions de subordination pour la déconvolution libre additive et multiplicative sous des conditions de moment faibles. Nos résultats incluent un algorithme pour calculer ces fonctions de subordination, et donc les transformées de Cauchy associées, pour les nombres complexes ayant une partie imaginaire supérieure à un paramètre dépendant de la mesure à déconvoler. L'existence des fonctions de subordination sur de tels domaines réduit le problème de la déconvolution libre au problème de la déconvolution additive classique par une distribution de Cauchy. Ainsi, nos résultats, combinés à des méthodes connues de déconvolution classique par une distribution de Cauchy, nous permettent de résoudre le problème de déconvolution libre. Nous présentons également des extensions de ces résultats au cas des déconvolutions à valeur opérateur.

MSC2020 subject classifications: Primary 46L54; secondary 60B20; 30D05

Keywords: Deconvolution; Free probability; Random matrices; Subodination

1. Introduction

Voiculescu introduced free independence in [46] as a new special kind of non-commutative relation between collections of operators, similar to the usual stochastic notion of independence, but inspired by free products, rather than by tensor products. A few years later, in [49], he observed that free independence occurs naturally as a fundamental conceptual relation describing the collective behavior of large random matrices appearing in pioneering (and modern) works on asymptotic random matrix theory.

The free additive convolution $\mu_1 \boxplus \mu_2$ and the free multiplicative convolution $\mu_1 \boxtimes \mu_2$ are binary operations of probability measures. They correspond to the distributions of the sum and the product of *free* non-commutative random variables with distributions μ_1 and μ_2.

The approach for computing free convolutions using *analytic subordination functions* [5,13,50,52], has shown to be very effective for concrete calculations.

In this work, we are concerned with the inverse problem known as free additive (resp. multiplicative) *deconvolution*, which is just recovering μ_2 from the knowledge of μ_1 and $\mu_1 \boxplus \mu_2$ (resp. $\mu_1 \boxtimes \mu_2$).

Our main contribution is to solve the problem of computing *free deconvolutions* of distributions by means of analytic subordination [5]. We include an algorithm to compute free deconvolutions numerically. Our methods cover free deconvolutions in the broader context of operator-valued free independence.

1.1. *Basic framework and motivation*

We recall here basic models in asymptotic random matrix theory. For each $N \geq 1$, let X_N be a self-adjoint $N \times N$ Wigner matrix.[1] In addition, let $D_1^{(N)}, D_2^{(N)}$ be self-adjoint $N \times N$ deterministic matrices, such that their uniform spectral probability distributions converge to fixed probability measures μ_1, μ_2.

Due to Wigner's semicircle law [53], the eigenvalue distribution of X_N converges to the semicircle distribution. As further examples, consider the following three random matrix models:

$$P_N = P\big(X_N, D_1^{(N)}\big) := X_N + D_1^{(N)}, \tag{1}$$

$$Q_N = Q\big(X_N, D_1^{(N)}\big) := X_N D_1^{(N)} X_N, \tag{2}$$

$$R_N = R\big(X_N, D_1^{(N)}, D_2^{(N)}\big) = X_N D_1^{(N)} X_N + D_2^{(N)}. \tag{3}$$

Marčenko and Pastur described in [32] the asymptotic eigenvalue distributions of such combinations of matrices. For example, they observed in particular, by studying the Stieltjes transform $z \mapsto \frac{1}{N}\mathrm{Tr} \circ \mathbb{E}((R_N - zI_N)^{-1})$, that in the limit (as the matrix size grows), the limiting transform (and thus the limiting spectral distribution) depends on the deterministic matrices $D_1^{(N)}, D_2^{(N)}$ only through their limiting distributions μ_1, μ_2.

The relation between free probability and large random matrix theory begins with Voiculescu's seminal paper [49], showing that the *asymptotic collective* behavior of the involved random matrices $X_N, D_1^{(N)}, D_2^{(N)}$, is exactly described by free independence, which he defined a few years before [46].

More precisely, the limiting distributions of P_N, Q_N, R_N as $N \to \infty$, are those of the abstract operators

$$P_\infty = x + d_1, \qquad Q_\infty = x d_1 x, \qquad R_\infty = x d_1 x + d_2,$$

where x, d_1 and d_2 are *free random variables* in a non-commutative probability space $(\mathcal{A}, \tau)$, x has semicircular distribution, and the distributions of $(d_i)_i$ are the (given) limiting distributions $(\mu_i)_i$, $i \in \{1, 2\}$.

Thus, in terms of distributions

$$\mu_{P_\infty} = \mu_s \boxplus \mu_1, \qquad \mu_{Q_\infty} = \pi_1 \boxtimes \mu_1, \qquad \mu_{R_\infty} = (\pi_1 \boxtimes \mu_1) \boxplus \mu_2,$$

where μ_s is the semicircle distribution and π_λ is the Marčenko–Pastur distribution of parameter λ.

In early works Voiculescu also derived analytic transforms to compute these free additive [47] and multiplicative [48] convolutions, based on the Cauchy–Stieltjes[2] transform $G_\mu : \mathbb{C}^+ \to \mathbb{C}^-$ and its reciprocal, the F-transfom, $F_\mu : \mathbb{C}^+ \to \mathbb{C}^+$:

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-t}\, d\mu(t), \qquad F_\mu(z) = \frac{1}{G_\mu(z)}, \quad z \in \mathbb{C}^+$$

where $\mathbb{C}^+$ (and $\mathbb{C}^-$) denote the upper (resp. lower) complex half-plane. A few years later, alternative combinatorial methods [36,43] were derived to calculate free convolutions.

However, in order to obtain exact formulas for free convolutions, combinatorial methods require us to recognize distributions from their moment sequences, and analytic methods require solving equations involving compositional inverses of analytic maps. Thus, outside special situations, explicit descriptions of free convolutions are rare.

The subordination approach of [5] (which is based on the works [13,50,52]) has been quite successful. The idea is to approximate free convolutions numerically, by deriving fixed point equations for analytic transforms of the convolutions, from which the Cauchy–Stieltjes transform can be recovered with high precision. Thus, the associated probability measure can be efficiently approximated using the Stieltjes inversion.

1.2. *Statement of the results*

We provide a similar subordination approach to the problem of computing free deconvolutions. Our method is mainly based on the work of Belinschi and Bercovici [5]. To illustrate it, let us recall their main theorem for the additive case.

---

[1] $X_N = 2^{-1/2}(Z_N + Z_N^*)$, where $Z_N = \frac{1}{\sqrt{N}}(z_{ij})$ is a Ginibre matrix (that is, the entries z_{ij} are centered i.i.d. random variables).

[2] The Stieltjes transform is just the negative of the Cauchy transform. The method of Stieltjes inversion to recover distributions may of course be performed using the Cauchy transform, by simply adding a sign on the Stieltjes-inversion formula (see Section 2). The analytic theory of non-commutative probability (free, boolean and monotone) heavily relies on the fact that the reciprocal of the Cauchy transform $F_\mu(z)$ is an analytic self-map on the upper complex half-plane. We use both names since we work directly with Cauchy transforms and reciprocals but rely at the end on Stieltjes inversions.

Theorem 1.1 ([5, Theorem 3.2]). *Given probability measures μ_1, μ_2 on $\mathbb{R}$, there exist unique functions $\omega_1, \omega_2 : \mathbb{C}^+ \to \mathbb{C}^+$ such that*

(1) $\Im\omega_j(z) \geq \Im z$ *for* $z \in \mathbb{C}^+$ *and*

$$\lim_{y\to\infty} \frac{\omega_j(iy)}{iy} = 1, \quad j = 1, 2.$$

(2) $F_{\mu_1 \boxplus \mu_2}(z) = F_{\mu_1}(\omega_1(z)) = F_{\mu_2}(\omega_2(z))$.
(3) $\omega_1(z) + \omega_2(z) = z + F_{\mu_1 \boxplus \mu_2}(z)$ *for all* $z \in \mathbb{C}^+$.
(4) *Denote by* $h_1(w) = w - F_{\mu_1}(w)$, $h_2(w) = w - F_{\mu_2}(w)$ *and* $T_z(w) = z - h_1(z - h_2(w))$.

Then for any $w \in \mathbb{C}^+$, *the iterated function* $T_z^{\circ n}(w)$ *converges to* $w_2(z)$.

The functions ω_1 and ω_2 are known as subordination functions. It is worth mentioning that because of their analytical properties ω_1 and ω_2 correspond to F-transforms (reciprocal Cauchy–Stieltjes transforms) of certain measures, sometimes denoted by $\mu_1 \boxminus\!\!\!\!\boxminus \mu_2$ and $\mu_2 \boxminus\!\!\!\!\boxminus \mu_1$.

Apart from its theoretical importance, the above theorem allows to estimate the density of free convolutions: We may implement numerical approximations for ω_1 (similarly for ω_2) by application of (4), and then use (2) to calculate the F-transform of $\mu_1 \boxplus \mu_2$.

In this paper, we consider the inverse problem of recovering one of the factors of the free convolution, known as free deconvolution ($\boxminus$). A first motivation is given by the following problem in random matrix theory. Suppose that we have a large random matrix B_N, perturbed by some additive noise which one knows statistically, say X_N, and one is given the information of the matrix

$$A_N = B_N + X_N.$$

In view of Voiculescu's results on asymptotic freeness, if we want to recover the eigenvalue distribution of B_n in terms of the eigenvalue distributions of A_N and of X_N, we may replace the triplet (A_N, B_N, X_N) by the system of operators (a, b, x) in an abstract non-commutative probability space $(\mathcal{A}, \tau)$, where $a = b + x$ and b, x are free.

The distribution of μ_b of b in terms of the distribution of a, μ_a, and the distribution of x, μ_x, is known as *deconvolving* x from a, and the distribution of b is called the free additive deconvolution [40,41]. The probability measure μ_b can be used as an approximation to the desired empirical distribution μ_{B_N}.

Our results deal only with the limiting distributions as the dimension of the matrices tends to infinity. However, the convergence of spectral distributions (empirical or averaged) of random matrices to their limit is very strong, see [23]. For example, models involving Gaussian or Unitary matrices are considered large enough already for $N \geq 10$, in the sense that the distribution of the desired operator is well-approximated by the distribution of the limit operator.

A combinatorial approach to free deconvolution has been considered in [11] and amounts to calculate the moments of μ_a and μ_x up to a certain order, then calculating their free cumulants and substracting them. One finally chooses a (non-unique) distribution with these free cumulants as a candidate for an approximation of b. The method has obvious limitations such as moment conditions or non-uniqueness. In [17, ch. 17], the authors propose an analytic approach to free deconvolution. Their method, albeit very efficient in certain situations, relies on a specific functional equation which has strong practical limitations and holds only when μ_x is a Marčenko–Pastur distribution.

Our main results give general solutions to free additive and free multiplicative deconvolutions following the lines of Theorem 1.1. For deconvolutions, we cannot get subordination functions in the whole upper half-plane $\mathbb{C}^+$. Indeed, if $\mu_1 \boxplus \mu_2 = \mu_3$, then $G_{\mu_3}(\mathbb{C}^+) \subset G_{\mu_1}(\mathbb{C}^+) \cap G_{\mu_2}(\mathbb{C}^+)$, which yields directly (at least at a set-theoretical level) the existence of a subordination function $w_2 : \mathbb{C}^+ \to \mathbb{C}^+$ such that $G_{\mu_3} = G_{\mu_2} \circ w_2$. However, the same inequality $G_{\mu_3}(\mathbb{C}^+) \subset G_{\mu_2}(\mathbb{C}^+)$ prevents us from finding a subordination function w_3 defined on $\mathbb{C}^+$ such that $G_{\mu_2} = G_{\mu_3} \circ w_3$. Therefore, the purpose of our result is to build subordination functions in controlled sub-domains, wide enough to ultimately allow the recovery of μ_2 by classical deconvolution (see the discussion after Theorem 1.2).

For $\alpha > 0$, let $\mathbb{C}_\alpha = \{z : \Im z > \alpha\}$. Our main result for free additive deconvolutions reads as follows:

Theorem 1.2. *Let μ_1 and μ_3 be probability measures on $\mathbb{R}$ such that μ_1 has finite variance σ_1^2. There exist unique functions $\omega_1, \omega_3 : \mathbb{C}_{2\sqrt{2}\sigma_1} \to \mathbb{C}^+$, such that*

(1) $\Im\omega_j(z) \geq \frac{1}{2}\Im z$ *for* $z \in \mathbb{C}_{2\sqrt{2}\sigma_1}$ *and*

$$\lim_{y\to\infty} \frac{\omega_j(iy)}{iy} = 1, \quad j = 1, 3.$$

(2) *If* μ_2 *is such that* $\mu_1 \boxplus \mu_2 = \mu_3$, *then*

$$F_{\mu_2}(z) = F_{\mu_3}\big[w_3(z)\big] = F_{\mu_1}\big[w_1(z)\big]$$

for $z \in \mathbb{C}_{2\sqrt{2}\sigma_1}$.

(3) $\omega_1(z) - \omega_3(z) = F_{\mu_3}[w_3(z)] - z$ *for all* $z \in \mathbb{C}_{2\sqrt{2}\sigma_1}$.

(4) *Denote by* $h_1(w) = w - F_{\mu_1}(w)$, $\tilde{h}_3(w) = F_{\mu_3}(w) + w$ *and* $T_z(w) = h_1(\tilde{h}_3(w) - z) + z$.

Then for any w *with* $\Im w > (3\Im z)/4$, *the iterated function* $T_z^{\circ n}(w)$ *converges to* $w_3(z) \in \mathbb{C}^+$ *independent of* w.

Remark that the parameter $2\sqrt{2}\sigma_1$ only depends on the intensity of the noise, and this dependence is linear in the standard deviation σ_1. This parameter is not optimal for many cases, but there is also no hope to give subordination functions for much lower imaginary parts (see Remark 3.2 more details). Thus, if asked to recover μ_2 from the knowledge of μ_1 and μ_3, Theorem 1.2 only recovers $F_{\mu_2}(z + i2\sqrt{2}\sigma_1)$, which is actually the F-transform of the *classical* convolution $\mu_2 * \mathcal{C}$ of our desired distribution μ_2 and a centered Cauchy distribution $\mathcal{C}$ with parameter $2\sqrt{2}\sigma_1$.

Hence, the problem of calculating free additive deconvolutions is reduced to the one of *classically* deconvolving a Cauchy distribution, which amounts to solve a Fredholm equation of the first kind (see [21]) in our case. This can be achieved using a regularization technique with convex optimization, as explained in Section 4. The simulations provided in that section also show the efficiency of the method.

On the other hand, we notice that in Theorem 1.2 we only assumed the fact that $\mu_1 \boxplus \mu_2 = \mu_3$ in part (2), but (3) is satisfied as long as we are $2\sqrt{2}\sigma_1$ above the real line. This has a nontrivial consequence in the arithmetic of free probability; adding a large enough Cauchy distribution to *any* measure with finite variance automatically ensures the existence of a free deconvolution. More precisely, we have the following result.

Theorem 1.3. *The function* $\tilde{F}_2(z) = F_{\mu_3} \circ w_3(z + 2\sqrt{2}\sigma_1 i)$ *is analytic on* $\mathbb{C}^+$ *and there exists a probability measure* $\tilde{\mu_2} \in \mathcal{P}(\mathbb{R})$ *such that* $\tilde{F}_2 = F_{\tilde{\mu}_2}$. *Moreover,* $\tilde{\mu}_2$ *satisfies that*

$$\mu_1 \boxplus \tilde{\mu_2} = \mu_3 \boxplus \mathcal{C}_{2\sqrt{2}\sigma_1},$$

where $\mathcal{C}_{2\sqrt{2}\sigma_1}$ *denotes the Cauchy distribution with parameter* $2\sqrt{2}\sigma_1$.

We derive a similar theorem for free multiplicative deconvolutions, motivated by the analog problem in random matrix theory, of reconstructing the distribution of B_N from the distributions of $A_N B_N$ (or $B_N^{1/2} A_N B_N^{1/2}$) and A_N. In the limit, this operation is exactly the *free multiplicative deconvolution* ($\boxbslash$).

Our approach to the free multiplicative deconvolution follows the same ideas as in the additive case. At the end of the process, we must perform a classical additive deconvolution with a Cauchy distribution. For the first step of the deconvolution, instead of using Theorem 1.2, we use the following result (which also follows ideas of the multiplicative case in [5, Theorem 3.3]).

Theorem 1.4. *Let* $\mu_1, \mu_3 \in \mathcal{P}(\mathbb{R})$ *be such that* μ_1 *has non-negative support and admits moments of order* 4, *and such that* μ_3 *admits moments of order* 2 *with non-zero first moment. Without loss of generality, suppose that the first moments of* μ_1 *and* μ_3 *are equal to one. Then, there exists* $K > 0$ *and unique functions* $\omega_1, \omega_3 : \mathbb{C}_K \to \mathbb{C}^+$ *such that*:

(1) *The constant* K *depends only on the respective variances* σ_1^2 *and* σ_3^2 *of* μ_1 *and* μ_3 *and on the Jacobi coefficients* β_1, γ_1 *of* μ_1, *and we can choose*

$$K \leq \Big[R + \sqrt{5/4R^2 + 5R\sigma_3^2}\Big],$$

with $R = (2\sqrt{\gamma_1} \vee \beta_1 \vee \frac{20\sigma_1^2}{\sqrt{3}})$

(2) *If* μ_2 *is such that* $\mu_1 \boxtimes \mu_2 = \mu_3$, *then*

$$F_{\mu_2}(z) = F_{\mu_3}\big[w_3(z)\big] z w_3(z)^{-1}.$$

(3) *Denote by* $h_1(w) = w - F_{\mu_1}(w)$, $\tilde{h}_3(w) = w^{-2}[w - F_{\mu_3}(w)]$ *and* $T_z(w) = zh_1(\tilde{h}_3(w)^{-1}z^{-1})$. *Then for* $z \in \mathbb{C}_K$ *and any* w *in* $D(z, \frac{\Im z}{5})$, *the iterated function* $T_z^{\circ n}(w)$ *converges to* ω_3.

(4) $\omega_1(z) = \frac{1}{z\tilde{h}_3(w_3(z))}$ *for all* $z \in \mathbb{C}_K$.

The parameter K in the latter theorem can be numerically computed by solving a system of two polynomial equations. In the multiplicative case, this parameter is not optimal, and it would be very interesting to have further improvements of its value. Indeed, the lower it is, the better is the precision of the recovery of the desired distribution.

Using our results for obtaining subordination functions, we implement in Section 4 an algorithm to compute free deconvolutions, including the last step, which requires a classical deconvolution with a Cauchy distribution. We test our method for both discrete and continuous distributions and compare it with random matrix simulations.

1.3. *Applications and related works*

Recovering the spectral distribution of a matrix from a noisy version is of central importance in the estimation of the covariance matrix of a large random vector when the sampling is large. In general, one wants to estimate a positive semi-definite matrix $\Sigma \in \mathbb{R}^{p\times p}$ from the observation of $M = X\Sigma X^*$, where $X \in \mathbb{R}^{n\times p}$ is a random matrix with i.i.d. entries. When n and p go to infinity with $n \approx p$, the estimation of Σ turns into a difficult problem. The shrinkage is a way to solve this question by constructing an estimator $\hat{\Sigma}$ by keeping the eigenvectors of M unchanged while changing the corresponding eigenvalues (see [15] for an overview of the method). In its simplest form this procedure is a linear shift of all the eigenvalues by a constant (see [27] for a study of this estimator). In [26], Ledoit and Péché provided an optimal shrinkage based on the knowledge of the spectral distribution of the original covariance matrix Σ. The main lacking step is thus the estimation of this spectral distribution.

Several methods have been proposed to recover the spectral distribution of the covariance matrix Σ when n and p are large. They are either based on a moment approach [3,24,39], the use of the Marčenko–Pastur equation [20,33] or a mix of both [30]. Ledoit and Wolf [28] successfully used these estimations to implement the shrinkage procedure of Ledoit and Péché. Although the goal of the present paper is to provide a general formalism for the spectral deconvolution, it would be interesting to apply our method to the shrinkage estimation of covariance matrices.

Our analytic deconvolution may also be used for finding outliers of large matrices from randomly perturbated versions. In random matrix theory, an outlier is a large eigenvalue which is outside of the bulk of the spectral distribution. They are of special importance in high-dimensional data analysis since they capture the typical dominant behavior of a linear system: for example they are used in PCA of large random vectors. It is therefore important to estimate the exact value of the outliers of a large matrix from the data of the matrix perturbated by a matricial noise. The first important result in this direction is the seminal paper of Baik, Ben Arous and Péché [4] which gives the distribution of an outlier in M from the value of the corresponding outlier in Σ. The law of large numbers of [4] has been generalized to arbitrary models in [8] by showing that one can determine the positions of outliers in free convolutions of spiked models in terms of the subordination functions. This result together with our deconvolution procedure easily yields an estimator for the outlier, as we show in Section 4. Our simulations show agreement between the outlier and our estimator.

The problem of efficiently computing free deconvolutions is also relevant in view of practical problems in random matrix theory and wireless communications, as in [41]. In a more indirect way, free deconvolutions are also important whenever free additive or multiplicative convolutions (or the relevant transforms) are main objects of study. For example, the free multiplicative convolution and the S-transforms play important roles in applied works on quantum information theory [2] and neural networks [37,45]. In those situations we are interested in certain features of the distribution of the convolutions (e.g. weight of a certain atom, symmetry, positivity, concentration around a certain point, etc.). Consequently, a better understanding of how free deconvolutions map classes of probability measures may help in those situations.

1.4. *Generalization to operator-valued case*

Voiculescu's operator-valued free probability theory (or B-valued free probability) has greatly extended the applicability of free probability. In particular, operator-valued independence (or B-independence) is a much broader relation which may be observed more frequently between models in random matrices or operator theory. Thus, it has become relevant to derive new tools to compute B-free convolutions.

In the B-valued case, explicit expressions for convolutions are hard to obtain (even harder than in the scalar situation), and thus the approach using *analytic subordination functions* [6,7] is very important. It has led to a robust toolbox for computing asymptotic distributions of random matrices [2,7,10,16,42], including a remarkable algorithm for computing distributions of *arbitrary, self-adjoint, non-commutative polynomials* evaluated in free self-adjoint random variables [6].

Therefore, in the last section, we include extensions of our analytic subordinations methods to compute B-free deconvolutions (for the case of *bounded operators*) on certain regions of the B-valued upper half-plane.

1.5. *Organization of the article*

Apart from this introduction, the paper is organized in four more sections.

Section 2 includes preliminaries on transforms, free convolutions and fixed-point theorems required for proofs of our main results. In Section 3 we deal with the case of scalar-valued free deconvolution (that is, we prove Theorems 1.2, 1.3 and 1.4). Section 4 gives concrete examples of the deconvolution procedure. We explain how to implement the algorithms of Theorem 1.2, and 1.4, we show two applications to random matrices: first by considering the problem of recovering a random matrix from its deformed version by adding noise, and second, we see how to approximate the outlier of a matrix from the spike of the deformed model. The simulations show the efficiency of our algorithms. Finally, in Section 5 we treat the operator valued case. First we give basic elements for operator-valued free probability, including some technical lemmas used afterwards and finally we prove our theorems for computing B-valued Cauchy transforms of B-valued free deconvolutions.

2. Preliminaries

2.1. *Transforms*

We denote by $\mathcal{P}_2(\mathbb{R})$ the set of probability measures on $\mathbb{R}$ having a finite second moment ($\int t^2\,d\mu(t) < \infty$) and by $\mathcal{P}_\infty$ the set of probability measures with bounded support. For $\sigma \in \mathbb{R}$, denote by $\mathbb{C}_\sigma$ the upper half-plane $\mathbb{C}_\sigma := \{z \in \mathbb{C}, \Im z > \sigma\}$.

For $\mu \in \mathcal{P}_2(\mathbb{R})$, let $G_\mu : \mathbb{C}^+ \to \mathbb{C}^-$ denote its Cauchy transform, defined by

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-t}\,d\mu(t).$$

The Stieltjes inversion formula recovers a measure from its Cauchy transform as follows:

$$\mu\big(]a,b[\big) = -\frac{1}{\pi}\lim_{y\downarrow 0}\int_a^b \Im\big[G_\mu(x+iy)\big]\,dx. \tag{4}$$

The reciprocal Cauchy transform $F_\mu : \mathbb{C}^+ \to \mathbb{C}^+$ is defined by $F_\mu(z) = \frac{1}{G_\mu(z)}$. It satisfies the important relation

$$\Im\big(F_\mu(z)\big) \geq \Im z, \quad z \in \mathbb{C}^+. \tag{5}$$

Let μ be a probability measure with $2n+2$-moments, that is $\int_{\mathbb{R}} x^{2n+2}\mu(dx) < \infty$. Then the Cauchy transform can be expressed in the form

$$G_\mu(z) = \cfrac{1}{z-\beta_0-\cfrac{\gamma_0}{z-\beta_1-\cfrac{\gamma_1}{\ddots\,\cfrac{}{z-\beta_n-\gamma_n G_\nu(z)}}}}, \tag{6}$$

where ν is a probability measure. The sequences $\beta_m = \beta_m(\mu) \in \mathbb{R}$, $\gamma_m = \gamma_m(\mu) \geq 0$ are respectively called the Jacobi parameters of μ of first and second order. If $\mu \in \mathcal{P}_2(\mathbb{R})$, then β_0 and γ_0 are respectively the expected value and the variance of μ. Notice that (6) at $n=0$ gives $F_\mu(z) = z - \beta_0 - \gamma_0 G_\nu$ for some probability measure ν, so that applying (5) to G_ν yields

$$\big|h_\mu(z) - \beta_0\big| \leq \gamma_0/\Im z, \tag{7}$$

where $h_\mu(z) = z - F_\mu(z)$. The latter inequality plays an important role in our proof. In particular, if $\beta_0 = 0$ then $|h_\mu(z)| \leq \gamma_0/\Im z$.

Moreover, as a consequence of the analyticity of G_μ outside of the support of μ, Hasebe [22, Lemma 4.1] proved that if μ has a positive support and admits enough moments to get the expansion (6), then ν has also a positive support and each coefficient β_m is non-negative.

2.2. *Free convolutions*

Free additive convolution was defined by Voiculescu in [46] for probability measures with compact support and later generalized by Maassen [31] for measures in $\mathcal{P}_2(\mathbb{R})$ and in [12] for general probability measures. Here we will use the analytic definition from [31] via Voiculescu's transform ϕ_μ. For this we need the following lemma.

Lemma 2.1 ([31, Lemma 2.4]). *Let μ be a probability measure on $\mathbb{R}$ with mean* 0, *variance σ^2, and reciprocal Cauchy transform F. Then the restriction of F to $\mathbb{C}_\sigma$ takes every value in $\mathbb{C}_{2\sigma}$, precisely once. The inverse function $F^{\langle -1\rangle} : \mathbb{C}_{2\sigma} \to \mathbb{C}_\sigma$ thus defined satisfies*

$$\left|F^{\langle -1\rangle}(u) - u\right| < \frac{2\sigma^2}{\Im u}.$$

The Voiculescu's transfom, $\phi_\mu : \mathbb{C}_{2\sigma} \to \mathbb{C}_\sigma$, is defined by the formula $\phi_\mu(z) = F_\mu^{\langle -1\rangle}(z) - z$. The *free additive convolution* of two probability measures $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R})$ with variance σ_1^2 and σ_2^2 is the unique probability measure $\mu_3 = \mu_1 \boxplus \mu_2$ on $\mathbb{R}$ such that $\phi_{\mu_3} = \phi_{\mu_1} + \phi_{\mu_2}$ on $\mathbb{C}_{2\sigma_3}$ with $\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2}$ and it is denoted by $\mu_1 \boxplus \mu_2$.

For the multiplicative version of free convolution, let $\eta_\mu : \mathbb{C}^+ \to \mathbb{C}^+$ denotes the η-transform of a distribution μ, which is defined by the formula $\eta_\mu(w) = [1 - wF_\mu(w^{-1})]$. The inverse of η_μ is well defined in a neighborhood of 0 as long as the first moment of μ does not vanish. Define $\Sigma_\mu : \Omega_\mu^+ \to \mathbb{C}$ by $\Sigma_\mu(z) = \eta_\mu^{\langle -1\rangle}(z)/z$, where Ω_μ^+ is a neighborhood of 0 in $\mathbb{C}^+$. Then, for $\mu_1, \mu_2 \in \mathbb{P}_2(\mathbb{R})$ such that μ_1 is supported on the positive real line and μ_2 has non-zero first moment, the *free multiplicative convolution* of μ_1 and μ_2 is the unique probability measure μ_3 such that $\Sigma_{\mu_1}\Sigma_{\mu_2} = \Sigma_{\mu_3}$ on $\Omega_{\mu_1}^+ \cap \Omega_{\mu_2}^+$. In this case we write μ_3 as $\mu_1 \boxtimes \mu_2$.

2.3. *Fixed point theorems*

In the proof of the main theorems we will use the following two theorems on convergence to fixed points of a function. The first one, proved independently by Denjoy [18] and Wolff [55], considers holomorphic maps from the unit disc, $\mathbb{D} = \{z : |z| < 1\}$, to itself. Let $f : \mathbb{D} \to \mathbb{D}$ be an analytic function. A point $w \in \mathbb{D}$ is called a Denjoy–Wolff point for f if either

(1) $w \in \mathbb{D}$ and $f(\omega) = \omega$; or
(2) $|w| = 1$, $\lim_{r\uparrow 1} f(r\omega) = \omega$ and $\lim_{r\uparrow 1} \frac{\omega - f(r\omega)}{(1-r)\omega} \le 1$.

Except for the identity map of $\mathbb{D}$ every function f has a unique Denjoy–Wolff point. The theorem of Denjoy and Wolff shows that for generic maps this point is the limit of the iterates of f.

Theorem 2.2 ([18,55]). *Assume that $f : \mathbb{D} \to \mathbb{D}$ is not a conformal automorphism of $\mathbb{D}$ and denote by ω its Denjoy–Wolff point. Let $f^{\circ n}$ donote the n-fold composition of f. Then, for any $z \in \mathbb{D}$ the sequence $(f^{\circ n}(z))_{n=0}^\infty$ converges to ω.*

The above theorem is obviously still valid for any open set comformally equivalent to the unit disc. For the operator valued case we use a similar result for Banach spaces due to Earl and Hamilton [19]. In this case we need that f maps D strictly inside D.

Theorem 2.3 ([19]). *Let D be a connected open subset of a complex Banach space X and let f be a holomorphic mapping of D into itself such that*:

(1) *the image $f(D)$ is bounded in norm*;
(2) *the distance between points $f(D)$ and points in the exterior of D is bounded from below by a positive constant*.

Then the mapping f has a unique fixed point w in D and for any point $z \in D$, the sequence $(f^{\circ n}(z))_{n=0}^\infty$ converges to w.

3. Free deconvolutions

In this section we prove the main theorems, by first considering the free additive deconvolution and then the free multiplicative deconvolution. Our aim is to find suitable sets and suitable transforms so that we obtain a fixed point equation. For this we will need to give some estimates of the image of these sets under the different transforms in order to be able to use the above fixed point theorem.

3.1. *Additive deconvolution*

Let $\mu_1, \mu_3 \in \mathcal{P}_2(\mathbb{R})$, and suppose without loss of generality that μ_1 is centered. We are looking to solve the equation $\mu_1 \boxplus \mu_2 = \mu_3$. For this we will find subordinations function ω_1, ω_3 such that

$$F_{\mu_2}(z) = F_{\mu_3}\big[w_3(z)\big] = F_{\mu_1}\big[w_1(z)\big].$$

As described above we will use an iterative procedure. So, let us recall two particular functions used in in the statement of Theorem 1.2:

- The h-transform of μ_1 is the function $h_1 : \mathbb{C}^+ \to \mathbb{C}^-$ defined by

$$h_1(w) = w - F_{\mu_1}(w).$$

- The $\tilde{h}$-transform of μ_3 is the function $\tilde{h}_3 : \mathbb{C}^+ \to \mathbb{C}^+$ defined by

$$\tilde{h}_3(w) = F_{\mu_3}(w) + w.$$

We denote by σ_1^2 the variance of μ_1. For $z \in \mathbb{C}_{2\sqrt{2}\sigma_1}$, set $\alpha(z) = \frac{3\Im(z)}{4}$.

Proposition 3.1. *For $z \in \mathbb{C}_{2\sqrt{2}\sigma_1}$, the function $T_z(w) = h_1(\tilde{h}_3(w) - z) + z$ is well defined and analytic on $\mathbb{C}_{\alpha(z)}$.*
For any $w \in \mathbb{C}_{\alpha(z)}$, the iterated function $T_z^{\circ n}(w)$ converges to $w_3(z) \in \mathbb{C}_{\alpha(z)}$ which is the unique fixed point of T_z.

Proof. Let $z \in \mathbb{C}_{2\sqrt{2}\sigma_1}$ and simply write α instead of $\alpha(z)$. Let us prove first that T_z is well defined on $\mathbb{C}_\alpha$. Since h_1 is defined on $\mathbb{C}^+$, we just have to check that $\tilde{h}_3(w) - z \in \mathbb{C}^+$ for $w \in \mathbb{C}_\alpha$. Let $w \in \mathbb{C}_\alpha$. By the definition of α, $\Im(w) > \frac{3\Im(z)}{4}$ and thus

$$\Im\big(\tilde{h}_3(w) - z\big) = \Im\big(F_{\mu_3}(w) + w - z\big) \geq 2\frac{3\Im(z)}{4} - \Im(z) > \frac{\Im(z)}{2}, \tag{8}$$

where we have used (5) in the second inequality.

In view of applying Denjoy–Wolff theorem, we prove now that $T_z(\mathbb{C}_\alpha) \subset \overline{\mathbb{C}_\alpha}$. Let $w \in \mathbb{C}_\alpha$. Then, since F_{μ_1} is the F-transform of a centered probability measure having variance σ_1^2, applying (7) yields

$$\big|F_{\mu_1}(x) - x\big| \leq \frac{\sigma_1^2}{\Im(x)},$$

for $x \in \mathbb{C}^+$, which implies

$$\Im\big[F_{\mu_1}(x)\big] \leq \Im(x) + \frac{\sigma_1^2}{\Im(x)}. \tag{9}$$

Applying (8) and (9) to $x = \tilde{h}_3(w) - z$ we obtain

$$\Im\big[F_{\mu_1}\big(\tilde{h}_3(w) - z\big)\big] \leq \Im\big[\tilde{h}_3(w) - z\big] + \frac{\sigma_1^2}{\Im[\tilde{h}_3(w) - z]} < \Im\big[\tilde{h}_3(w)\big] - \Im(z) + \frac{2\sigma_1^2}{\Im(z)}.$$

Hence, for $z \in \mathbb{C}_{2\sqrt{2}\sigma_1}$,

$$\begin{aligned}
\Im\big[T_z(w)\big] &= \Im\big[h_1\big(\tilde{h}_3(w) - z\big) + z\big] \\
&= \Im\big[\tilde{h}_3(w) - F_{\mu_1}\big(\tilde{h}_3(w) - z\big)\big] \\
&> \Im(z) - \frac{2\sigma_1^2}{\Im(z)} \geq \frac{3\Im(z)}{4},
\end{aligned}$$

where we used the inequality $t - \frac{2\sigma_1^2}{t} \geq \frac{3t}{4}$, valid for $t \geq 2\sqrt{2}\sigma_1$. Thus we have proved that $T_z(w) \in \mathbb{C}_\alpha$, as desired.

Since $T_z(\mathbb{C}_\alpha) \subset \mathbb{C}_\alpha$, we just have to prove that T_z is not an automorphism of $\mathbb{C}_\alpha$ in order to apply Denjoy–Wolff Theorem. But, if $w \in \mathbb{C}_\alpha$,

$$\left|T_z(w) - z\right| = \left|h_1\left(\tilde{h}_3(w) - z\right) + z - z\right| = \left|F_{\mu_1}\left(\tilde{h}_3(w) - z\right) - \left(\tilde{h}_3(w) - z\right)\right| \le \frac{\sigma_1^2}{\Im(\tilde{h}_3(w) - z)}.$$

Hence, by (8), $|T_z(w) - z| < \frac{2\sigma_1^2}{\Im(z)}$ and

$$T_z(\mathbb{C}_\alpha) \subset D\left(z, \frac{2\sigma_1^2}{\Im(z)}\right), \tag{10}$$

where the latter is the disk with center z and radius $\frac{2\sigma_1^2}{\Im(z)}$. Therefore, T_z is not surjective and hence is not an automorphism of $\mathbb{C}_\alpha$. By Denjoy–Wolff Theorem, there exists $w_3(z) \in \overline{\mathbb{C}_\alpha} \cup \{\infty\}$ such that $T_z^{\circ n}(w)$ converges to $w_3(z)$ for all $w \in \mathbb{C}_\alpha$. By (10), $w_3(z) \in \overline{D(z, \frac{2\sigma_1^2}{\Im(z)})} \subset \mathbb{C}_\alpha$ and thus $w_3(z)$ is a fixed point of T_z. □

Remark 3.2.

(1) Without any additional property on μ_1 and μ_3, the constant $2\sqrt{2}\sigma$ is sharp. Indeed, if we only assume the inequality

$$\left|F_\mu(z) - z\right| \le \frac{\sigma^2}{\Im(z)}$$

for distribution μ with finite variance σ, a computation yields that the stability condition $T_z(\mathbb{C}_\alpha) \subset \overline{\mathbb{C}_\alpha}$ implies that α satisifies the inequality

$$\left(2\alpha - \Im(z)\right)\left(\Im(z) - \alpha\right) - \sigma_1^2 > 0,$$

which is possible if and only if $\Im(z)^2 > 8\sigma_1^2$.

(2) Notice that if we consider $z \in \mathbb{C}_\beta$, for some $\beta = c2\sqrt{2}\sigma_1$ and $c > 1$, then the function T_z, satisfies

$$T\left(D\left(z, \frac{1}{4}\Im(\zeta)\right)\right) \subset T(\mathbb{C}_{\alpha(z)}) \subset D\left(z, \frac{2\sigma_1^2}{\Im(\zeta)}\right) \subset 1/c^2 T\left(D\left(z, \frac{1}{4}\Im(\zeta)\right)\right), \tag{11}$$

Thus, for $z \in \mathbb{C}_\beta$, $T_z : D(z, 1/4\Im(\zeta)) \to \mathbb{C}$ is a contraction with Lipschitz constant smaller than $1/c^2$. In particular, if we $\beta = 3$, then $c^2 = 9/8$.

Proposition 3.3. *The function w_3 is analytic on $\mathbb{C}_{2\sqrt{2}\sigma_1}$ and $\lim_{n\to\infty} \frac{w_3(iy)}{iy} = 1$. Moreover, we have*

$$\phi_{\mu_3}\left[F_{\mu_3}\left(w_3(z)\right)\right] - \phi_{\mu_1}\left[F_{\mu_3}\left(w_3(z)\right)\right] = z - F_{\mu_3}\left(w_3(z)\right)$$

for z large enough.

Proof. The analiticity of w_3 follows from Theorem 2.3 in [5]. Now, for $z \in \mathbb{C}_{2\sqrt{2}\sigma_1^2}$, the fact that $w_3(z)$ is a fixed point of T_z implies that it is in $\mathbb{C}_{\alpha(z)}$ which yields that $\Im[w_3(z)] > 3/4\Im(z)$. Therefore, (8) yields that

$$\Im\left[\tilde{h}_3\left(w_3(yi)\right) - yi\right] \ge 3y/4$$

for $y > 0$. Hence, since $w_3(yi)) = T_{yi}(w_3(yi))$,

$$\left|w_3(yi) - yi\right| = \left|T_{yi}\left(w_3(yi)\right) - yi\right| = \left|h_1\left(\tilde{h}_3\left(w_3(yi)\right) - yi\right)\right| \le \frac{\sigma_1^2}{3y/4},$$

and

$$\lim_{n\to\infty} \frac{w_3(iy)}{iy} = 1.$$

By Section 2.2, ϕ_{μ_1}, ϕ_{μ_2} and ϕ_{μ_3} are well-defined on $\mathbb{C}_{2\sigma_3}$. Let $z \in \mathbb{C}_{4\sigma_3}$, so that $\Im[w_3(z)] > 2\sigma_3$. Since $\Im[F_{\mu_3}(w)] \geq \Im(w)$ for $w \in \mathbb{C}^+$, we thus also have

$$\Im\big[F_{\mu_3}\big(w_3(z)\big)\big] \geq w_3(z) > 2\sigma_3$$

for $z \in \mathbb{C}_{4\sigma_3}$, so that ϕ_{μ_1} and ϕ_{μ_3} are well-defined on $F_{\mu_3}(w_3(z))$ for $z \in \mathbb{C}_{4\sigma_3}$. For $z \in \mathbb{C}_{4\sigma_3}$, set

$$w_1(z) = \tilde{h}_3\big(w_3(z)\big) - z = F_{\mu_3}\big(w_3(z)\big) + w_3(z) - z. \tag{12}$$

Since $w_3(z)$ is a fixed point of T_z, we have

$$\begin{aligned} F_{\mu_1}\big(w_1(z)\big) &= -h_1\big(w_1(z)\big) + w_1(z) \\ &= -h_1\big(\tilde{h}_3\big(w_3(z)\big) - z\big) + \tilde{h}_3\big(w_3(z)\big) - z \\ &= -T_z\big(w_3(z)\big) + z + \tilde{h}_3\big(w_3(z)\big) - z \\ &= -w_3(z) + F_{\mu_3}\big(w_3(z)\big) + w_3(z) = F_{\mu_3}\big(w_3(z)\big), \end{aligned}$$

so that

$$\begin{aligned} \big[w_1(z) - F_{\mu_1}\big(w_1(z)\big)\big] + \big[z - F_{\mu_3}\big(w_3(z)\big)\big] &= \tilde{h}_3\big(w_3(z)\big) - z + z - 2F_{\mu_3}\big(w_3(z + i\sigma_1)\big) \\ &= w_3(z) - F_{\mu_3}\big(w_3(z)\big). \end{aligned} \tag{13}$$

Hence $w_3(z) \in \mathbb{C}_{2\sigma_3}$, and by [34, Lemma 24] $F_{\mu_3}^{\langle -1\rangle}[F_{\mu_3}(w_3(z))] = w_3(z)$. Therefore,

$$\begin{aligned} w_3(z) - F_{\mu_3}\big(w_3(z)\big) &= F_{\mu_3}^{\langle -1\rangle}\big[F_{\mu_3}\big(w_3(z)\big)\big] - F_{\mu_3}\big(w_3(z)\big) \\ &= \phi_{\mu_3}\big[F_{\mu_3}\big(w_3(z)\big)\big]. \end{aligned}$$

Likewise, since $w_1(z) \in \mathbb{C}_{2\sigma_3} \subset \mathbb{C}_{2\sigma_1}$,

$$\begin{aligned} w_1(z) - F_{\mu_1}\big(w_1(z)\big) &= F_{\mu_1}^{\langle -1\rangle}\big[F_{\mu_1}\big(w_1(z)\big)\big] - F_{\mu_1}\big(w_1(z)\big) \\ &= \phi_{\mu_1}\big[F_{\mu_1}\big(w_1(z)\big)\big] = \phi_{\mu_1}\big[F_{\mu_3}\big(w_3(z)\big)\big]. \end{aligned}$$

Therefore,

$$\phi_{\mu_3}\big[F_{\mu_3}\big(w_3(z)\big)\big] - \phi_{\mu_1}\big[F_{\mu_3}\big(w_3(z)\big)\big] = z - F_{\mu_3}\big(w_3(z)\big). \qquad \square$$

We can now turn to the proof of Theorem 1.2 and Theorem 1.3.

Proof of Theorem 1.2. By Proposition 3.1, $\Im\omega_3(z) \geq 3\Im z/4$. By (12), $\omega_1(z)$ is defined as

$$\omega_1(z) = F_{\mu_3}\big(\omega_3(z)\big) + \omega_3(z) - z.$$

Hence, the fact that $\Im F_{\mu_3}(w) \geq \Im w$ for $w \in \mathbb{C}^+$ yields that $\Im\omega_1(z) \geq \Im z/2$. The last part of statement (1) is given by Proposition 3.3 for ω_3, and is deduced by (12) and the fact that $\lim_{y\to\infty} \frac{F_{\mu_3}(yi)}{yi} = 1$ for ω_1.

For the second statement, suppose that there exists $\mu_2 \in \mathcal{P}_2(\mathbb{R})$ such that $\mu_1 \boxplus \mu_2 = \mu_3$. Then, by the first statement, $F_{\mu_3}(w_3(z))$ goes to infinity when z goes to infinity along $i\mathbb{R}_{\geq 0}$. Hence, $\phi_{\mu_2}(F_{\mu_3}(w_3(z))$ is well-defined for $z \in i\mathbb{R}_{\geq 0}$ large enough. Moreover, by the equality $\mu_1 \boxplus \mu_2 = \mu_3$ and by Proposition 3.3,

$$\phi_{\mu_2}\big[F_{\mu_3}\big(w_3(z)\big)\big] = \phi_{\mu_3}\big[F_{\mu_3}\big(w_3(z)\big)\big] - \phi_{\mu_1}\big[F_{\mu_3}\big(w_3(z)\big)\big] = z - F_{\mu_3}\big(w_3(z)\big).$$

Since $\phi_{\mu_2}(w) = F_{\mu_2}^{\langle -1\rangle}(w) - w$ on its domain of definition, the above equality yields

$$F_{\mu_2}^{\langle -1\rangle}\big(F_{\mu_3}\big(w_3(z)\big)\big) = z,$$

and thus $F_{\mu_2}(z) = F_{\mu_3}(w_3(z))$ for $z \in i\mathbb{R}_{\geq 0}$ large enough. By Proposition 3.3, $F_{\mu_3} \circ w_3$ is analytic on its domain of definition. Since F_{μ_2} is also analytic and coincides with $F_{\mu_3} \circ w_3$ in a set which is not discrete, the two functions are equal on the intersection of their domains of definition, which is $\mathbb{C}_{2\sqrt{2}\sigma_1}$. Statement (3) is the definition of ω_1 in (12), and statement (4) is the content of Proposition 3.1. □

Proof of Theorem 1.3. Set $\tilde{F}_2(z) = F_{\mu_3}(w_3(z + 2\sqrt{2}\sigma_1 i))$. Since w_3 is defined on $\mathbb{C}_{2\sqrt{2}\sigma_1}$, $\tilde{F}_2$ is a well-defined function from $\mathbb{C}^+$ to $\mathbb{C}^+$. Moreover, by Proposition 3.3,

$$\lim_{n\to\infty} \frac{w_3(iy)}{iy} = 1,$$

which implies

$$\lim_{n\to\infty} \frac{w_3(iy + 2\sqrt{2}\sigma_1 i)}{iy} = 1.$$

Since F_{μ_3} satisfies also the asymptotic behavior $\lim_{y\to\infty} \frac{F_{\mu_3}(yi)}{yi} = 1$, we finally get

$$\lim_{n\to\infty} \frac{\tilde{F}_2(iy)}{iy} = 1,$$

so that by Nevanlinna representation theorem, there exists a probability measure $\tilde{\mu} \in \mathcal{P}(\mathbb{R})$ such that $\tilde{F}_2 = F_{\tilde{\mu}}$. By definition of $\tilde{F}$, for z large enough,

$$F_{\tilde{\mu}}^{\langle -1\rangle}\big(F_{\mu_3}\big(w_3(z)\big)\big) = z - 2\sqrt{2}\sigma_1 i.$$

Hence, by Proposition 3.3, for z large enough we have

$$\phi_{\mu_3}\big(F_{\mu_3}\big(w_3(z)\big)\big) - \phi_{\mu_1}\big(F_{\mu_3}\big(w_3(z)\big)\big) = z - F_{\mu_3}\big(w_3(z)\big) = \phi_{\tilde{\mu}}\big(F_{\mu_3}\big(w_3(z)\big)\big) + 2\sqrt{2}\sigma_1 i.$$

Since $F_{\mathcal{C}_{2\sqrt{2}\sigma_1}}(z) = z + 2\sqrt{2}\sigma_1$, we have $\phi_{\mathcal{C}_{2\sqrt{2}\sigma_1}} = -2\sqrt{2}\sigma_1 i$, so that

$$\phi_{\mu_3}\big(F_{\mu_3}\big(w_3(z)\big)\big) + \phi_{\mathcal{C}_{2\sqrt{2}\sigma_1}}\big(F_{\mu_3}\big(w_3(z)\big)\big) = \phi_{\mu_1}\big(F_{\mu_3}\big(w_3(z)\big)\big) + \phi_{\tilde{\mu}}\big(F_{\mu_3}\big(w_3(z)\big)\big)$$

for z large enough. We deduce that

$$\mu_1 \boxplus \tilde{\mu} = \mu_3 + \mathcal{C}_{2\sqrt{2}\sigma_1}.$$ □

3.2. *Multiplicative deconvolution*

Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_2(\mathbb{R})$ be such that μ_1 admits moments of order four and has support on $[0, +\infty[$ (with $\mu_1 \neq \delta_0$), and such that μ_3 admits moments of order two and has non-zero first moment.

This subsection is dedicated to the free multiplicative convolution

$$\mu_1 \boxtimes \mu_2 = \mu_3, \tag{14}$$

and the objective is to recover the Cauchy transform of μ_2 from the ones of μ_1 and μ_3. Up to a rescaling of μ_1 and μ_3, we can assume that the first moment of μ_1 and μ_3 are equal to 1. Following (6), we denote by β_1, γ_1 the second Jacobi parameters of μ_1 of respectively first and second order, and we set $R = (2\sqrt{\gamma_1} \vee \beta_1 \vee \frac{20\sigma_1^2}{\sqrt{3}})$. Set

$$K = \big[R + \sqrt{5/4R^2 + 5R\sigma_3^2}\big].$$

For $z \in \mathbb{C}_K$, set $r_z = \frac{\Im(z)}{5|z|}$, and define the function

$$T_z(w) = h_1\big(z(1+w)^2 h_3\big[z(1+w)\big]^{-1}\big) - 1$$

on $\Delta_z := D(0, r_z)$.

Proposition 3.4. *The map T_z is well-defined on Δ_z, and for all $w \in \Delta_z$ we have*

$$\lim_{n\to\infty} T_z^{\circ n}(w) = \tilde{w}_3(z)$$

for some $\tilde{w}_3(z)$ independent of the original choice of w and such that $T_z(\tilde{w}_3(z)) = \tilde{w}_3(z)$. Moreover, $\tilde{w}_3(z)$ goes to zero as $\Im(z)$ goes to infinity.

Proof. Let us write $r_z = \frac{t\Im(z)}{|z|}$ with $t < 1$ varying for now; we will show below that the result holds for $t = \frac{1}{5}$.

Let us first prove that T_z is well-defined. Set $I = \Im(z)$. Since the support of μ_1 is included in $[0, \infty[$, h_1 can be analytically extended to $\mathbb{C} \setminus [0, \infty[$. On the one hand, for $w \in \Delta_z$

$$\Im\big(\big[z(1+w)\big]\big) = I + \Im(zw) > I(1-t),$$

where we have used the fact that $|w| \le \frac{tI}{|z|}$ in the last inequality. By the definition of h_3 and (7), the latter inequality with $t < 1$ yields that $h_3[z(1+w)] \in \mathbb{C}^-$ and $|h_3[z(1+w)] - 1| \le \frac{\sigma_3^2}{(1-t)I}$. Hence, we have

$$h_3\big[z(1+w)\big]^{-1} = \frac{1}{1+u} \quad \text{with } u \in \mathbb{C}^-, |u| \le \frac{\sigma_3^2}{(1-t)I}. \tag{15}$$

On the other hand, for $w \in \Delta_z$, we have $z(1+w)^2 = z + \tilde{u}$ with

$$|\tilde{u}| \le 2z|w| + |z| \cdot |w|^2 \le 2tI + \frac{t^2 I^2}{|z|} \le \big(2t + t^2\big)I.$$

Hence, we have

$$\big|z(1+w)^2\big| \ge \big(1 - 2t - t^2\big)|z|$$

and

$$\Im\big(z(1+w)^2\big) \ge \big(1 - 2t - t^2\big)I.$$

In particular, $z(1+w)^2 \in \mathbb{C}^+$ for t small enough (smaller than $1/3$ for example). Since $h_3[z(1+w)]^{-1} \in \mathbb{C}^+$ by (15), we finally get that

$$z(1+w)^2 h_3\big[z(1+w)\big]^{-1} \in \mathbb{C} \setminus [0, \infty[,$$

and T_z is well defined on Δ_z.

Set $\delta = z(1+w)^2 h_3[z(1+w)]^{-1}$. If $\Re(\delta) \ge 0$, $d(\delta, [0, +\infty[) = |\Im(\delta)|$. Since $h_3[z(1+w)]^{-1} \in \mathbb{C}^+$, $\arg(\delta) \ge \arg(z(1+w)^2))$ and by (15),

$$\big|\Im(\delta)\big| \ge \big|h_3\big[z(1+w)\big]^{-1}\big|\Im\big(z(1+w)^2\big) \ge \frac{(1-2t-t^2)I}{1 + \sigma_3^2/[(1-t)I]},$$

which yields

$$d\big(\delta, [0, +\infty[\big) \ge \frac{I^2(1-t)(1-2t-t^2)}{(1-t)I + \sigma_3^2} := F(t, I).$$

If $\Re(\delta) \le 0$, $d(\delta, \infty) = |\delta|$. Moreover, using again (15) yields

$$|\delta| \ge \frac{|z|I(1-t)(1-2t-t^2)}{(1-t)I + \sigma_3^2} = \frac{|z|}{I} F(t, I) \tag{16}$$

and, since $|z| \ge I$, we get also $d(\delta, [0, +\infty[) \ge F(t, I)$. Remark that (16) is also valid when $\Re(\delta) \ge 0$. We suppose now that t, I are such that

$$F(t, I) \ge \left(2\sqrt{\gamma_1} \vee \beta_1 \vee \frac{4\sigma_1^2}{\sqrt{3}rt}\right) \tag{17}$$

for some $0 < r < 1$. By Section 2.1,

$$\left|h_1(\delta) - 1\right| = \left|\frac{\sigma_1^2}{\delta - \beta_1 - \gamma_1 G_\nu(\delta)}\right| \leq \left|\frac{\sigma_1^2}{|\delta - \beta_1| - |\gamma_1 G_\nu(\delta)|}\right|,$$

with ν a probability measure supported on $[0, \infty[$. On the one hand, since $\beta_1 \geq 0$ by [22, Lemma 4.1], $|\delta - \beta_1| \geq d(\delta, [0, +\infty[)$. On the other hand, since ν is supported on $[0, \infty[$ (see Section 2.1), $|\gamma_1 G_\nu(\delta)| \leq \frac{\gamma_1}{d(\delta, [0, +\infty[)}$. By (17), $d(\delta, [0, +\infty[) \geq 2\sqrt{\gamma_1}$, and thus

$$\left|\gamma_1 G_\nu(\delta)\right| \leq \frac{d(\delta, [0, +\infty[)}{4} \leq \frac{|\delta - \beta_1|}{4}.$$

Hence,

$$\left|h_1(\delta) - 1\right| \leq \frac{4\sigma_1^2}{3|\delta - \beta_1|} = \frac{4\sigma_1^2}{3|z|}\frac{|z|}{|\delta|}\frac{|\delta|}{|\delta - \beta_1|}.$$

Since $d(\delta, [0, +\infty[) \geq \beta_1$ by the first inequality of (17), a geometric argument yields that $\frac{|\delta|}{|\delta - \beta_1|} < \sqrt{3}$. Hence, the second inequality of (17) yields

$$\left|h_1(\delta) - 1\right| < \frac{4\sigma_1^2}{\sqrt{3}|z|}\frac{|z|}{|\delta|} \leq \frac{rtF(t, I)}{|z|}\frac{|z|}{|\delta|} \leq r\frac{tI}{|z|},$$

so that $T_z(w) \in r\Delta_z$ for some $0 < r < 1$. Hence, conditioned on the fact that t, I satisfy (17), T_z is an analytic map which is a strict contraction of Δ_z, and Denjoy–Wolff theorem yields that for all $w \in \Delta_z$, $T_z^{\circ n}(z)$ converges to the unique fixed point of T_z in Δ_z. Let $t = \frac{1}{5}$. Then, I satisfies (17) if

$$I^2\left(\frac{4}{5}\cdot\frac{14}{25}\right) - \left(\frac{4}{5}I + \sigma_3\right)R \geq 0$$

with $R = (2\sqrt{\gamma_1} \vee \beta_1 \vee \frac{20\sigma_1^2}{\sqrt{3}})$. The two roots of the above second degree polynomials are

$$x_\pm = \frac{25}{28}R \pm \frac{125}{112}\sqrt{(4/5)R^2 + 4\frac{112}{125}R\sigma_3^2}.$$

Since $K \geq [R + \sqrt{5/4R^2 + 5R\sigma_3^2}]$, for $I \geq K$ we have $I \geq x_+$ and the inequality of (17) is satisfied.

Finally, for any $0 < t < 1$ fixed, for I large enough (t, I) satisfies (17). Hence, for all small $0 < t < 1$ and $\Im(z)$ large enough,

$$\left|\tilde{w}_3(z)\right| \leq \frac{t\Im(z)}{|z|} \leq t,$$

and $\tilde{w}_3(z)$ goes to zero as $\Im(z)$ goes to infinity. □

Remark 3.5. The choice the constant K could be certainly improved, depending on the value of σ_1, σ_2, β_1 and γ_1. One of the way to improve K is to find the set $\mathcal{K}$ of values I in the above proof such that the inequalities in (17) is satisfied for some $0 < t < \sqrt{2} - 1$ (the restriction on t is given by the condition $1 - 2t - t^2 \geq 0$). This involves a polynomial in $\mathbb{R}[t, I]$ of degree 4 in t and 2 in I, and it can be easily seen that $\mathcal{K}$ is an interval $[K_0, \infty[$. The constant K_0, which can be obtained numerically, is a better constant than K. We chose to give the above explicit constant K, since our simulations showed that K does not differ much from K_0.

Set $w_3(z) = (1 + \tilde{w}_3(z))z$, and for $z \in \mathbb{C}_K$, set

$$F(z) = F_{\mu_3}\big(w_3(z)\big)zw_3(z)^{-1}.$$

Proposition 3.6. *The function F is analytic on $\mathbb{C}_K$ and coincides with F_{μ_2} on its domain of definition.*

In the following proof, recall that the η-transform of a distribution μ is defined by $\eta_\mu(w) = w[w^{-1} - F_\mu(w^{-1})]$. We have in particular $\eta_\mu(w) = wh_\mu(w^{-1})$ on $\mathbb{C}^+$.

Proof. By Denjoy–Wolff Theorem, $|T_z[\tilde{w}_3(z)]| < 1$, thus the implicit function theorem applied to the analytic function $g(w,z) = T_z(w) - w$ on $\{(w,z)|z \in \mathbb{C}_K, w \in \Delta_z\}$ yields the analyticity of $\tilde{w}_3$ and w_3. For all $z \in \mathbb{C}_K$, $\tilde{w}_3(z) \leq \frac{\Im(z)}{5|z|}$, thus $w_3(z) = z(1+\tilde{w}_3(z)) \in \mathbb{C}^+$, and F is well-defined and analytic on $\mathbb{C}_K$.

Set $w_1(z) = \frac{w_3(z)^2 z^{-1}}{h_3(w_3(z))}$. Since $\tilde{w}_3(z) = T_z(\tilde{w}_3(z)) = h_1(z(1+\tilde{w}_3(z))^2 h_3[z(1+\tilde{w}_3(z))]^{-1}) - 1$,

$$w_3(z) = z\big(1+\tilde{w}_3(z)\big) = zh_1\big(w_3(z)^2 z^{-1} h_3\big[w_3(z)\big]^{-1}\big) = zh_1\big(w_1(z)\big).$$

Hence,

$$\eta_{\mu_3}\big(w_3(z)^{-1}\big) = w_3(z)^{-1} h_3\big(w_3(z)\big) = w_1(z)^{-1} w_3(z) z^{-1} = w_1(z)^{-1} h_1\big(w_1(z)\big) = \eta_{\mu_1}\big(w_1(z)^{-1}\big).$$

Set $\eta_2(w) = 1 - wF(w^{-1})$ for $z \in \mathbb{C}_K$. Then, for w such that $w^{-1} \in \mathbb{C}_K$,

$$\eta_2(w) = 1 - wF\big(w^{-1}\big) = 1 - w_3\big(w^{-1}\big)^{-1} F_{\mu_3}\big(w_3\big(w^{-1}\big)\big) = \eta_{\mu_3}\big(w_3\big(w^{-1}\big)^{-1}\big) = \eta_{\mu_1}\big(w_1\big(w^{-1}\big)^{-1}\big).$$

Since $w_3(z) = z(1+\tilde{w}_3(z))$ with $|\tilde{w}_3(z)| \leq \frac{\Im(z)}{5|z|}$, $\Im(w_3(z)) \geq 4/5\Im(z)$ and $\Im[w_3(z)]$ goes to infinity when $\Im(z)$ goes to infinity. Hence, by (7), $h_3(w_3(z))$ converges to 1 as $\Im(z)$ goes to infinity, so that $|w_1(z)| = |\frac{w_3(z)^2 z^{-1}}{h_3(w_3(z))}|$ goes to infinity when $\Im(z)$ goes to infinity. For $i \in \{1,3\}$, $\eta_i(z) \sim z$ for z going to zero; hence, for $\Im(z)$ large enough, $w_3(z)^{-1}$, $w_1(z)^{-1}$ are respectively in the image of $\eta_{\mu_3}^{\langle -1\rangle}$, $\eta_{\mu_1}^{\langle -1\rangle}$, and $\eta_2(z^{-1}) = \eta_{\mu_3}(w_3(z)^{-1})$ is in the domain of $\eta_{\mu_2}^{\langle -1\rangle}$. This implies in particular that

$$\eta_{\mu_3}^{\langle -1\rangle}\big(\eta_{\mu_3}\big(w_3(z)^{-1}\big)\big) = w_3(z)^{-1}, \qquad \eta_{\mu_1}^{\langle -1\rangle}\big(\eta_{\mu_1}\big(w_1(z)^{-1}\big)\big) = w_1(z)^{-1}.$$

Therefore, since $\eta_{\mu_1}(w_1(z)^{-1}) = \eta_{\mu_3}(w_3(z)^{-1}) = \eta_2(z^{-1})$, for $\Im(z)$ large enough we have

$$\begin{aligned}\frac{\Sigma_3(\eta_2(z^{-1}))}{\Sigma_1(\eta_2(z^{-1}))} &= \frac{\eta_{\mu_3}^{\langle -1\rangle}(\eta_{\mu_3}(w_3(z)^{-1})\eta_{\mu_1}(w_1(z)^{-1})}{\eta_{\mu_3}(w_3(z)^{-1})\eta_{\mu_1}^{\langle -1\rangle}(\eta_{\mu_1}(w_1(z)^{-1}))} \\ &= \frac{w_3(z)^{-1}\eta_{\mu_1}(w_1(z)^{-1})}{\eta_{\mu_3}(w_3(z)^{-1})w_1(z)^{-1}} = \frac{w_1(z)}{w_3(z)} \\ &= \frac{w_3(z)}{zh_3(w_3(z))} = \frac{z^{-1}}{\eta_2(z^{-1})}.\end{aligned}$$

On the other hand, by the relation $\mu_1 \boxtimes \mu_2 = \mu_3$, for $\Im(z)$ large enough we have

$$\frac{\Sigma_3(\eta_2(z^{-1}))}{\Sigma_1(\eta_2(z^{-1}))} = \Sigma_2\big(\eta_2\big(z^{-1}\big)\big) = \frac{\eta_{\mu_2}^{\langle -1\rangle}(\eta_2(z^{-1}))}{\eta_2(z^{-1})}.$$

Hence, $\eta_{\mu_2}^{\langle -1\rangle}(\eta_2(z^{-1})) = z^{-1}$, which yields, after applying η_{μ_2} on both sides,

$$\eta_{\mu_2}\big(z^{-1}\big) = \eta_2\big(z^{-1}\big).$$

Therefore, η_2 and η_{μ_2} coincide in a neighborhood of zero. Since both maps are analytic, $\eta_{\mu_2}(z^{-1}) = \eta_2(z^{-1})$ for $z \in \mathbb{C}_K$, which yields

$$F = F_{\mu_2}$$

on $\mathbb{C}_K$. $\square$

The proof of Theorem 1.4 is given by Proposition 3.4 and Proposition 3.6.

4. Implementation of free deconvolution

As explained in the introduction, the subordination techniques developed in Theorem 1.2 and Theorem 1.4 provide a first step towards recovering the unknown distribution μ_2, by obtaining the distribution of $\mu_2 * \mathcal{C}_\lambda$, where $\mathcal{C}_\lambda$ is a Cauchy distribution with a parameter λ depending on the first moments of μ_1 and μ_3. Thus, we need to solve the classical deconvolution by the Cauchy distribution in order to complete the algorithm for free deconvolution. In this section we describe how to implement both steps.

4.1. *Free subordination functions*

One very useful consequence about Theorem 1.2 and Theorem 1.4 is that they provide a very direct method to calculate the subordination functions. We describe briefly this method for the additive convolution; the multiplicative case is identical, by choosing the correct function T_z to iterate.

First we choose a small $\epsilon > 0$ which will be our level of approximation. Given G_{μ_1} and G_{μ_3}, we can easily calculate the functions T_z from part (4) of Theorem 1.2.

Let $z \in \mathbb{C}_K$ for K given by Theorem 1.2. We start with an arbitrary point $w_0(z)$ in some proper domain $D = D(z, \frac{2\sigma_1^2}{\Im(z)})$ (for example take $w_0(z) = z$) and define $w^{(n+1)}(z) = T_z(w^{(n)}(z))$. Theorem 1.2 ensures the existence of $N > 0$ such that $w^{(N+1)}(z) - w^{(N)}(z) < \epsilon$ and we call $w^{(N+1)}(z) = w_\infty(z)$. Our approximation for $F_{\mu_2}(z)$ is given by $F_3(w_\infty(z))$. Here we note that (10) implies that for $D = D(z, \frac{2\sigma_1^2}{\Im(z)})$, $T_z : D \to D$ has a fixed point inside D. Moreover, the speed of convergence to the fixed point is *exponential* because T_z is a contractive map with respect to the Schwartz distance in D. For specific Lipschitz constats see Remark 3.2.

Let us choose a discretization $(x(i))_{1\le i\le n}$ of an interval $I \subset \mathbb{R}$ large enough. We are given the functions F_1 and F_3, and we start with a vector $[z(1), z(2), \dots, z(n)] \in \mathbb{C}^n_{2\sqrt{2}\sigma_1}$ where $z(i) = x(i) + 2\sqrt{2}\sigma_1$. We obtain a vector $[F_2(z(1)), \dots, F_2(z(n))]$ as follows.

I. 1. Set an approximation threshold $\epsilon > 0$.
   2. Define the functions $h_1(w) := w - F_1(w)$, and $h_3(w) := w + F_3(w)$.

II. For ($i = 1$ to n)
   1. Set $w_{\text{now}} := z(i)$.
   2. Set $w_{\text{past}} := w_{\text{now}}$.
   3. Set $w_{\text{now}} := h_1(h_3(w_{\text{past}} - z)) + z$.
   4. If $(|w_{\text{past}}(z(i)) - w_{\text{now}}| > \epsilon)$, go to 2.
      else set $w_\infty := w_{\text{now}}$.
   5. Set $F_2(z(i)) := F_3(w_\infty)$.

Applying the latter procedure and then Stieltjes inversion formula yields an approximation $V \in \mathbb{R}^n$ of the density $\tilde{f}$ of $\mu_2 * \mathcal{C}_{2\sqrt{2}\sigma_1}$ on I.

4.2. *Classical deconvolution with the Cauchy distribution*

In order to recover μ_2, one needs to perform afterwards the classical deconvolution of $\tilde{f}$ by the Cauchy distribution of paramter $2\sqrt{2}\sigma_1$. Deconvolving with a Cauchy kernel amounts to solve the Fredholm equation of the first kind (see [21] for more details on this class of equations)

$$\int_{\mathbb{R}} K(x, y)\, d_2\mu(y) = \tilde{f}(x), \quad x \in \mathbb{R},$$

with $K(x, y) = \frac{1}{\pi}\frac{\lambda}{(x-y)^2+\lambda^2}$, $\tilde{f}$ given by the previous step and μ_2 unknown. The latter is known to be a severely ill-posed problem and thus requires regularization. The natural procedure is given by a Tychonov regularization using jointly quadratic programming, which we now explain briefly.

After having discretized the problem and done the first step of the deconvolution, we end up with the linear equation

$$KU = V, \tag{18}$$

with $K \in M_{n\times m}(\mathbb{R})$, $V \in \mathbb{R}^n$ and $U \in \mathbb{R}^n$ are respectively a discrete version of the Cauchy kernel, the discrete approximation of $\tilde{f}$ we obtained in the first step, and a discrete version of the unknown density $d\mu_2$. The ill-possedness of the problem comes from the fact that K is singular (or has very small non-zero eigenvalues), which makes the solution

U unstable with respect to small perturbations of V. The goal of Tychonov regularization is to replace the negligible eigenvalues of K by small ones in order to make the linear problem stable. Namely, instead of solving (18), we will look for a solution which minimizes the convex function $\|KU - V\|^2 + \alpha^2\|U\|^2$, where $\alpha > 0$ is a parameter to be chosen. Moreover, we want to ensure that the solution is a probability distribution, which results in the minimization problem

$$U = \underset{\substack{U_i \geq 0 \\ \sum U_i * \delta = 1}}{\operatorname{argmin}} \left(\|KU - V\|^2 + \alpha^2\|V\|^2\right), \tag{19}$$

where δ is the step of the discretization. In general, the choice of the parameter α is crucial in the success of the Tychonov regularization, and we refer to [21, Section 3.3] for a possible strategy for the choice of such a parameter. In our case of study, we noticed that in all case we get good approximation of μ in Levy distance by simply setting $\alpha = 0$ in (19). This means that regularization could be avoided when we are only interested in approximation of μ_2 in the Levy distance: this important simplification should be the subject of further investigation. In order to achieve the minimization of (19), we used the quadratic programming package CVXOPT [1] with *Python* (see also [14] for theoretical background on the subject). For all examples listed below, the result is obtained in few seconds.

4.3. *Application*: *Recovering spikes in deformed model*

As we mentioned in the introduction, a possible interesting application of free deconvolutions is the recovery of outliers from a deformed matrix model. Namely, assume that $A \in M_n(\mathbb{R})$ is a Hermitian matrix with an outlier λ, and suppose that we know the deformed matrix $M = A + X$ or $M = X^{1/2}AX^{1/2}$, where X is a noise matrix whose spectral distribution μ_1 is known. We denote as usual by μ_2 (resp. μ_3) the spectral distribution of A (resp. M). Let us assume that λ is the unique outlier of A, and that this outlier yields an outlier λ_M on the deformed matrix M. The first hypothesis is only given to simplify the results, and the same results hold for several outliers. The main result of [8] relates the value of λ to the one of λ_M as follows:

- Additive case: Let $\tilde{w}_2$ be the subordination function from the *additive convolution* of μ_1 with μ_2, then

$$\lambda = \tilde{w}_2(\lambda_M). \tag{20}$$

- Multiplicative case: Let $\tilde{w}_2$ be the subordination function from the *multiplicative convolution* of μ_1 with μ_2, then

$$\lambda = \tilde{w}_2\left(\lambda_M^{-1}\right)^{-1}. \tag{21}$$

These results together with our method for free deconvolution provide a way to recover spikes of deformed models, as follows:

(1) Compute the distribution μ_2 using the given subordination methods.
(2) Compute the subordination $\tilde{w}_2$ of the additive (resp. multiplicative) convolution of μ_1 and μ_2.
(3) Apply the relation (20) (resp. (21)) to recover the original spike λ.

Two examples showing the efficiency of such procedure are displayed in Example 4.5 and Example 4.6.

4.4. *Simulations*

We include simulations for additive and multiplicative deconvolutions of two situations: one example where the unknown distribution is atomic and one example where the unknown distribution has a density.

We compare our results with actual simulations of large (however finite dimensional and thus only *approximately free*) random matrices. The distributions obtained by our free deconvolution method are close to the true distributions. All simulations are done with *Python*.

Additive case
Let us consider a (possibly random) matrix $A \in M_n(\mathbb{R})$ with limiting spectral distribution μ_A and a Wigner matrix $W \in M_n(\mathbb{R})$, whose spectral distribution is known to converge to a semicircular distribution $\mathbf{s}$.

We simulate $A + W$ and want to recover an approximation for μ_A. Free probability theory states that the distribution of $A + W$ should be close to the free convolution $\mathbf{s} \boxplus \mu_A$.

Example 4.1 (Discrete distribution). A is a diagonal matrix of size 1200 with eigenvalues -1, 0 and 1 with respective weights $1/2$, $1/6$ and $1/3$. Figure 1 shows the results of our method.

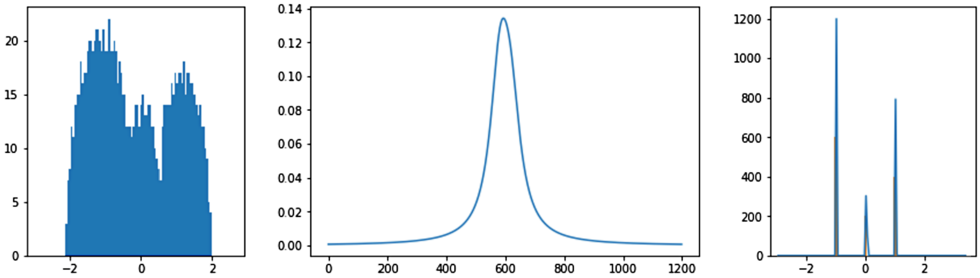

Fig. 1. Histogram of the spectral distribution of $A + W$ (left), result after first step of the deconvolution (center), and the final result compared with the original atomic distribution in orange (right). We did not use Tychonov regularization in this example.

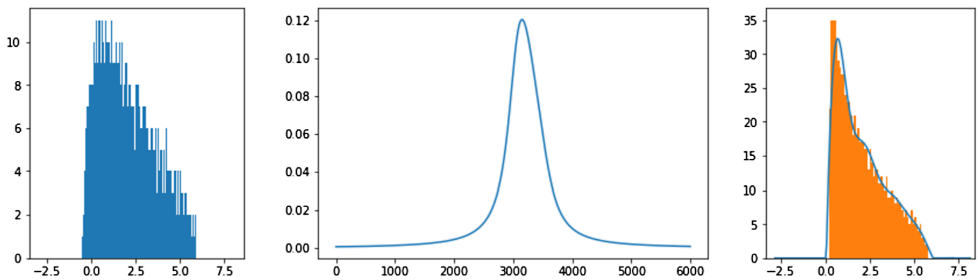

Fig. 2. Histogram of the spectral distribution of $A + W$ (left), result of the first step of the deconvolution (center), and result after Tychonov regularization compared with the histogram of eigenvalues of the original Wishart matrix.

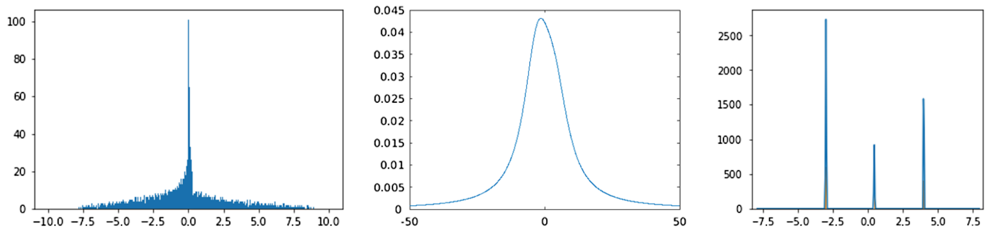

Fig. 3. Histogram of the spectral distribution of WAW^* of size $n = 1200$ (left), result of the first step of the deconvolution (center) and result after the second step compared with the original distribution in orange (right). We did not use Tychonov regularization in this example.

Example 4.2 (Marčenko–Pastur). $A = XX^*$, where X is a random rectangular matrix of size 800×1600 with independent Gaussian entries of variance $1/n$ (with $n = 800$). Figure 2 shows the results of our method.

Multiplicative case

Consider now a matrix $A \in M_n(\mathbb{R})$ and a Ginibre matrix $W \in M_n(\mathbb{R})$. Let us recover the spectral distribution A from the distribution of WAW^*, as follows.

Since WW^* is a Wishart matrix whose spectral distribution approximates the Marčenko–Pastur distribution $\mathbf{m}_1$ (or free Poisson) of parameter 1, the spectral distribution of WAW^* is approximately the free multiplicative convolution $\mathbf{m}_1 \boxtimes \mu_A$.

To approximate the original spectral distribution of the matrix A, we must calculate the multiplicative free deconvolution of the spectral distribution of WAW^* with the Marčenko–Pastur distribution $\mathbf{m}_1$. In the first step of the deconvolution, we found in these examples that we were able to use a lower parameter K than the one theoretically given by our theorem. This improved the precision of the realization of the second step.

Example 4.3 (Discrete distribution). A is a diagonal matrix with eigenvalues -3, $1/2$, 4 and 1 with respective weights $1/2$, $1/6$ and $1/3$. Figure 3 shows the results of our method.

Example 4.4 (Modification of Marčenko–Pastur). $A = 1/2(X^2 + (X^*)^2)$, where X is a random square matrix of size $n = 800$ with independent Gaussian entries with variance $1/n$. Figure 4 shows the results of our method.

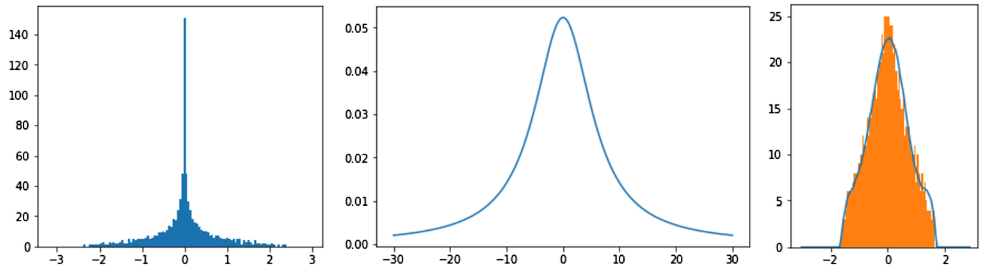

Fig. 4. Histogram of the spectral distribution of WAW^* (left), result of the first step of the deconvolution (center), result after Tychonov regularization and quadratic programming (with $n = 800$) and comparison with the histogram of eigenvalues of the original random matrix (right).

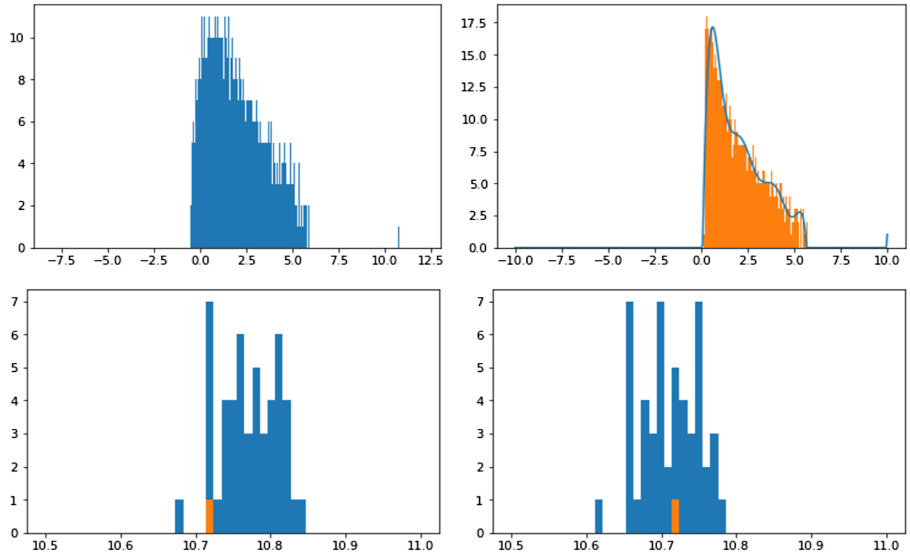

Fig. 5. Spectral distribution of $A + W$, computation of the spectral distribution μ_A by the subordination technique, histogram of the measured value of the outlier λ_M of $A + W$ and of the value after deconvolution (50 trials). The true value λ is in orange.

Recovery of spikes

We consider again Example 4.2 and Example 4.4, but we add this time a shift $\delta > 0$ to the largest eigenvalue of A, in order to get an outlier λ. The shift δ must be large enough to ensure that the outlier still exists after adding the matricial noise, resulting in an outlier λ_M (see [4] for more details on this phenomenon). The minimal value of the shift can be computed from the spectral distribution of A and the one of the matricial noise. We then apply the procedure given in Section 4.3 to recover λ from λ_M and the estimated spectral distribution μ_A of A.

Example 4.5. In the additive case, $A = (XX^*)$, where X is a random square matrix of size 800×1600 with independent Gaussian entries with variance $1/800$; then we added 5 to the largest eigenvalue (outlier at 10.75). The noise is a Wigner additive noise as in Example 4.2. Figure 5 shows the results of our method.

Example 4.6. In the multiplicative case, $A = (XX^*)^2$, X is random square matrix of size $n = 800$ with independent Gaussian entries with variance $1/n$; then we added 4 to the largest eigenvalue (outlier at 6.9). The noise is a Marčenko–Pastur distributed, as in Example 4.4. Figure 6 shows the results of our method.

4.4.1. *Comparison with previous methods*

We now compare our results with the methods of Ledoit and Wolf [27–29]. For this, we consider the matrix ZZ^* with $Z = YT$, where Y is a Wishart matrix of size $(2p, p)$ and T is a diagonal matrix that we want to recover. Following the example of the QuEST method (see [29] for details) the spectral distribution of T is given by discretization of the distribution function

$$H_4(x) = \begin{cases} \frac{1}{2}(1 - [1 - (2x)^3]^{1/3}), & x \in [0, 1/2], \\ \frac{1}{2}(1 + [1 - (2 - 2x)^3]^{1/3}), & x \in [1/2, 1]. \end{cases}$$

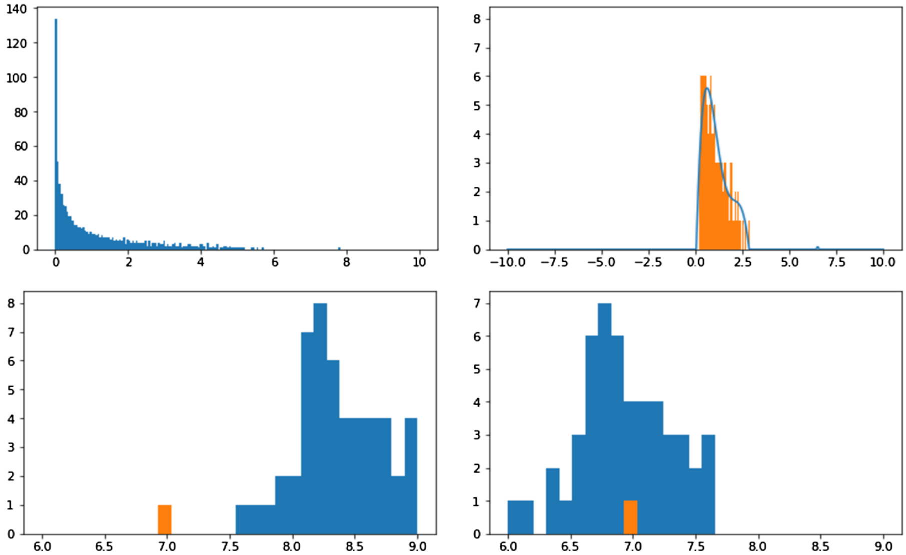

Fig. 6. Spectral distribution of WAW^*, computation of the spectral distribution μ_A by the subordination technique, histogram of the measured value of the outlier λ_M of WAW^* and of the value after deconvolution (50 trials of the noise W). The true value λ is in orange.

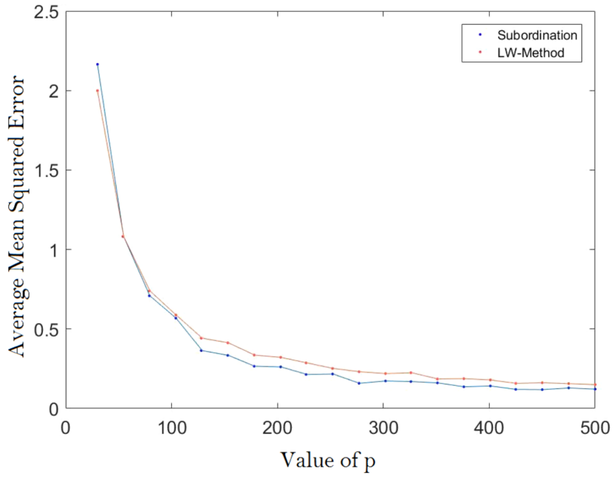

Fig. 7. Accuracy in terms of mean squared error in the subordination method compared to Ledoit–Wolf method.

For each p between 30 and 500 we made 100 simulations of the deconvolution both with the proposed method and with the method of Ledoit and Wolf, and then calculated the mean square error $\frac{1}{p}\sum_i((\hat{\lambda}_i - \lambda_i)^2))$. Since we were interested in the population of eigenvalues rather than the density of the spectral distribution, we chose to modify the second step of our algorithm described in Section 4.2: instead of minimizing the regularized distance (19) over all possible probability distributions, we are only minimizing it on the set of probability distributions with p atoms of mass $1/p$, and we set the regularizing parameter α to zero. Then we took the average over the 100 simulations for each value of p. The result is shown in Figure 7.

Similary, for each p as above between 30 and 500 we calculated the running time for calculating the deconvolution and took the average over the 100 simulations for each value of p. The result is shown in Figure 8. Both methods seem to have running times of similar order.

We see that both methods seem to provide the same accuracy; this should not be a surprise, since both methods rely on the relation between the spectral distribution of the average matrix $\mathbb{E}(ZZ^*)$ and the one of T (although concrete implementions differ from one method to the other). Our method should therefore be seen as a generalization of Ledoit–Wolf's viewpoint to the case of arbitrary multiplicative noise.

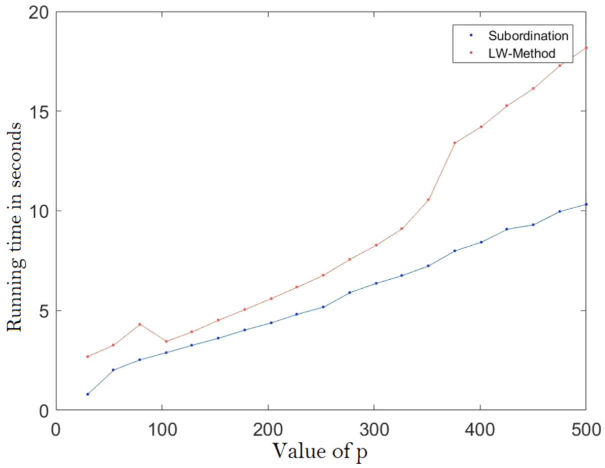

Fig. 8. Speed of the subordination method compared to the speed in the Ledoit–Wolf method.

5. Operator-valued free deconvolutions

In [51], Voiculescu provided generalizations of most concepts and tools from free probability theory to a broader theory of operator-valued (or B-valued) free probability. The fundamental concept of a non-commutative probability space (A, τ), consisting of a $*$-algebra with unit and a state $\tau : A \to \mathbb{C}$, is replaced by a triple (A, B, E), where $E : A \to B$ is a conditional expectation onto a smaller algebra $B \subseteq A$ (to be thought as the algebra of constants).

These theoretical generalizations found immediate applications to the description of more general models in random matrix theory. For example Shlyaktenko [42] used this framework to study band random matrices and block-random matrices, and provided at the same time a general pipeline for applications of B-free probability. In particular, these works showed the need to study how B-distributions of operators behave as we consider different choices of algebra B [35].

Indeed, in order to compute or approximate a desired B_0-valued distribution of a certain operator x, it has been often useful to rephrase the problem in terms of a B_1-distribution of an auxiliary operator y, where B_1 is an auxiliary algebra usually larger than B_0, in such a way that y is built-up by B_1-free pieces that we understand. As the algebra B_1 becomes larger, the notion of B_1-freeness becomes less meaningful or practical, up-to the extreme situation where B contains the relevant operators, and their B-freeness thus follows tautologically, as the corresponding conditional expectation restricted to the relevant operator algebras is the identity.

For non-trivial scenarios where B_1 is minimal, the machinery of B_1-free probability theory is used to compute the B_1-distribution or B_1-Cauchy transform of y, from which x is then extracted, typically by simple means. For example, if $B_0 \subseteq B_1$ are compatible expectations (i.e. the corresponding conditional expectations E_0, E_1 satisfy $E_0 \circ E_1 = E_0$), then the B_0-Cauchy transform is just the projected, restricted B_1-transform,

$$G_x^{B_0}(b_0) = E_0\big(G_x^{B_1}(b_0)\big), \quad b_0 \in B_0.$$

Along with the theoretical developments in B-valued free probability theory, the notion of B-free independence has been more frequently observed in applied models. Thus, the problems of computing B-free additive and multiplicative convolutions gained more interest, and the methods via analytic subordination have been particularly useful and effective.

In this section, we find the B-Cauchy–Stieltjes transform G_x^B of the B-free deconvolutions through subordination functions in a certain region of the B-upper half-plane. For simplicity, we only consider the case of B-independent *bounded* operators.

Unlike the scalar-valued case, in the B-valued case is not obvious what should be used in the second step of the algorithm for replacement of the Cauchy distribution, which allowed us to transfer the analytic distributions (with some small error), from a region away from the real line to a close neighborhoods of the real line in the upper half-plane, from which we obtain the deconvolved distribution (with a small error). Thus, for the moment, our algorithm deals only with the first part of the deconvolution process, which computes the B-Cauchy transform in a region away from the self-adjoint space B^{sa}.

We should warn the reader that, for our method to be useful for practical situations, we should find operator models where the desired B_0 distribution and the auxiliary B_1 distributions either coincide or are not too distant. For example, in the context of the block-modified random matrices studied in [2], the authors give a general numerical method, using a certain auxiliary algebra B_1 and a more restrictive but more explicit method, using a simpler algebra B_1'.

Before stating our results on B-free deconvolutions, let us recall some basic elements of B-valued probability.

5.1. *Elements of operator-valued free probability*

We refer to [44] for a basic introduction to operator-valued non-commutative spaces. In this section, we consider unital inclusions $B \subset A$ of C^*-algebras, and we denote by $E : A \to B$ a unit-preserving conditional expectation. Moreover, we denote by $B(\mathcal{X})$ the $*$-algebra of non-commutative polynomials in a self-adjoint variable $\mathcal{X}$ with coefficients in B. Following [38], we define a B-valued non-commutative distribution as a unital B-module map $\mu : B(\mathcal{X}) \to B$ such that

$$\left[\mu\left(f_i(\mathcal{X})^* f_j(\mathcal{X})\right)\right]_{1\le i,j\le n} \ge 0 \quad \text{in } M_n(B)$$

for all subsets $\{f_i(x)\}_{1\le i\le n}$ of $B(\mathcal{X})$. The distribution μ is said bounded by $M > 0$ if

$$\mu(\mathcal{X} b_1 \mathcal{X} \cdots \mathcal{X} b_n \mathcal{X}) < M^{n+1} \|b_1\| \cdots \|b_n\|$$

for $b_1, \dots, b_n \in B$.

Note that for $a \in A$ self-adjoint, the map $\phi_a : B(\mathcal{X}) \to B$ defined by $\phi_a(P) = E(P(a))$ is a non-commutative distribution. For any non-commutative distribution μ, there exists a unital inclusion of C^*-algebras $B \subset A$, a conditional expectation $E : A \to B$ and an element $a \in A$ such that $\mu = \phi_a$ (see [38, Proposition 1.2] and [54, Theorem 2.8]).

In this section, every non-commutative distribution is assumed to be B-valued.

5.2. *Statement of new results*

Let us introduce first several operator-valued versions of the transforms considered in Section 2.1. Let us denote by B^+ the subset of B consisting of elements with positive imaginary part. Namely, $b \in B^+$ if b is written $b = b_1 + ib_2$ with b_1 self-adjoint and $b_2 > 0$. Likewise, we define B^- as the set of elements of B with negative imaginary part. Given a bounded non-commutative distribution μ, we introduce the following maps:

- $G_\mu : B^+ \to B^-$ its Cauchy transform, defined by

  $$G_\mu(b) = \mu\left[(b - \mathcal{X})^{-1}\right].$$

  In the case that $\mu = \phi_a$, the Cauchy transform of μ can also be written as $G_\mu = E[(b-a)^{-1}]$.
- $F_\mu : B^+ \to B^+$ its reciprocal Cauchy transform $F_\mu = G_\mu^{-1}$.
- $\eta_\mu : B^+ \to B$ its η-transform defined by $\eta_\mu(b) = b[b^{-1} - F_\mu(b^{-1})]$.
- $\phi_\mu : B^+ \to B^+$, the operator-valued Voiculescu trasform $\phi_\mu(b) = F^{\langle -1\rangle}(b) - b$.
- $\Sigma_\mu(b) = b^{-1}\eta_\mu^{\langle -1\rangle}(b)$, defined on B^+ in a neighborhood of 0.

As in the scalar case, additive and multiplicative B-free convolutions may be defined at the level of the transforms on suitable domains: $\phi_{\mu_1 \boxplus \mu_2}(b) = \phi_{\mu_1}(b) + \phi_{\mu_2}(b)$ and $\Sigma_{\mu_1 \boxtimes \mu_2}(b) = \Sigma_{\mu_1}(b)\Sigma_{\mu_2}((\Sigma_{\mu_1}(b))^{-1} b \Sigma_{\mu_1}(b))$.

Although these definitions imply considering non-commutative series in $\mathcal{X}$ instead of polynomials, we can show that all these maps are well-defined and analytic by a limit argument (see [51] for a rigorous proof). Finally, we denote by $\sigma^2 := \|\mu(\mathcal{X}^2) - \mu(\mathcal{X})^2\|$ the norm of the variance of μ, and as in the scalar case we set $h_\mu(b) = b - F_\mu(b)$ for $b \in B^+$.

Our method for computing additive B-free deconvolutions reads as follows:

Theorem 5.1. *Suppose that $\mu_1 \boxplus \mu_2 = \mu_3$, with μ_1, μ_2 and μ_3 bounded B-valued distributions, and let $\sigma_1^2 = \|E(\mathcal{X}_1^2) - E(\mathcal{X}_1)^2\|$ be the variance of μ_1. For $b \in B$ such that $\Im b > 4\sqrt{2}\sigma_1$, set $\Delta_b = \{r \in B^+, \Im r > 3\Im b/4\}$ and define $T_b : \Delta_b \to B$ to be*

$$T_b(w) = h_{\mu_1}^B\left(h_{\mu_3}^B(w) + 2w - b\right) + b,$$

Then, T_b is well-defined and for any $w \in \Delta_b$, the sequence $T_b^{\circ n}(w)$ converges to an element $w_3(b)$ independent of the initial choice of w. Moreover,

$$F_{\mu_2}^B(b) = F_{\mu_3}^B\left(w_3(b)\right).$$

For the multiplicative case, let us first introduce some notations. Given μ_1, μ_3 two bounded B-valued distributions,

- R_i is the bound of the distribution μ_i,
- $\alpha_i := \|E(\mathcal{X}_i)\|$ is the norm of the first moment of μ_i, and
- $\alpha_i^* := \inf \operatorname{Spec} E(\mathcal{X}_i)$ is the minimum of the spectrum of $E(\mathcal{X}_i)$.

Then the result is the following.

Theorem 5.2. *Suppose that $\mu_1 \boxtimes \mu_2 = \mu_3$, with $\mu_1 \geq 0$. Set*

- $K := \frac{2}{\alpha_{\mu_1}^*} \max(\frac{2}{\alpha_{\mu_1}^*}(\sigma_3 + \alpha_{\mu_3})(\|\mu_1\| + 2\frac{\sigma_{\mu_1}^2}{\alpha_{\mu_1}^*}), \|\mu_3\| + \sigma_{\mu_3})$,
- *for b invertible such that $\|b\| \leq K^{-1}$, set $\Delta_b = bD(0, \frac{2}{\alpha_{\mu_1}^*})$ and define $T_b : \Delta_b \to B$ by*

$$T_b(w) = bH_{\mu_1}\big(b^{-1}wH_{\mu_3}(w)w\big)^{-1},$$

where $H_\mu(b) = h_\mu(b^{-1})$.

Then, for any b such that $\|b^{-1}\| \leq K$, $T_{b^{-1}}$ is well-defined, and for any $w \in \Delta_{b^{-1}}$ the sequence $T_{b^{-1}}^{\circ n}(w)$ converges to an element $w_3(b) \in B$ independent of the initial choice of w. Moreover,

$$F_{\mu_2}(b) = bw_3(b)F_{\mu_3}\big(w_3(b)^{-1}\big).$$

5.3. *Auxiliary lemmas*

We give three lemmas of independent interest that will be used in the proofs of Theorems 5.1 and 5.2.

Let us first slightly improve a bound of [38, Proposition 1.2] for later purposes.

Lemma 5.3. *Let $P \in B\langle\mathcal{X}\rangle$. Then,*

$$\mu\big(P^*b^*bP\big) \leq \big\|b^*b\big\|\mu\big(P^*P\big) \quad \text{and} \quad \mu\big(P^*\mathcal{X}^2P\big) \leq M^2\mu\big(P^*P\big),$$

where M is any constant bounding μ.

Proof. The first inequality is already proven in the proof of [38, Proposition 1.2]. In the same paragraph, the authors have also proven that

$$\mu\big(P^*\mathcal{X}^2P\big) \leq 4M^2\mu\big(P^*P\big).$$

We will adapt their proof to give our result: define for each monomial $f = b_0\mathcal{X}b_1 \cdots \mathcal{X}b_n \in B\langle\mathcal{X}\rangle$ the quantity $\mathfrak{p}(f) = M^n\|b_0\| \cdots \|b_n\|$, and denote by $\hat{B}\langle\mathcal{X}\rangle$ the $*$-algebra

$$\hat{B}\langle\mathcal{X}\rangle = \left\{\sum_{n=0}^{\infty} f_n \,\middle|\, f_n \text{ monomial in } B\langle\mathcal{X}\rangle \text{ such that } \sum_{n=0}^{\infty}\mathfrak{p}(f_n) < \infty\right\}.$$

Let $\tilde{\mu}$ be the positive B-valued linear map extending μ from $B\langle\mathcal{X}\rangle$ to $\hat{B}\langle\mathcal{X}\rangle$ with the formula

$$\tilde{\mu}\left(\sum_{n=0}^{\infty} f_n\right) = \sum_{n=0}^{\infty}\mu(f_n).$$

For $T > M$ and $n \geq 0$, let $g_{n,T} = (2n)![(1-2n)(n!)^2T^{2n}4^n]^{-1}\mathcal{X}^{2n}$. Then, $g_{n,T} = g_{n,T}^*$ and $\mathfrak{p}(g_{n,T}) \leq (M/T)^n$. Thus, $g_T = \sum_{n=0}^{\infty} g_{n,T} \in \hat{B}\langle\mathcal{X}\rangle$. Since $g_T^2 = 1 - [\mathcal{X}/T]^2$, we have

$$0 \leq \tilde{\mu}\big(P^*g_T^2P\big) = \tilde{\mu}\big(P^*\big(1 - [\mathcal{X}/T]^2\big)P\big) = \mu\big(P^*P\big) - T^{-2}\mu\big(P^*\mathcal{X}^2P\big).$$

Hence, $\mu(P^*\mathcal{X}^2P) \leq T^2\mu(P^*P)$. Since this holds for all $T > M$, we finally get

$$\mu\big(P^*\mathcal{X}^2P\big) \leq M^2\mu\big(P^*P\big). \qquad \square$$

We give then in the operator valued context an estimate of h_μ similar to (7).

Lemma 5.4. *Denote by $\sigma_{\inf}(v)$ the minimum of the spectrum of a self-adjoint operator $v \in B$. For $b \in B^+$, we have*

$$\|h_\mu(b) - \mu(\mathcal{X})\| \leq \frac{4\|\mu(\mathcal{X}^2) - \mu(\mathcal{X})^2\|}{\sigma_{\inf}\Im(b)}.$$

Proof. The proof follows the method of [9, Remark 2.5] and [6, Lemma 2.3]. Let $b = u + iv$, with $v > 0$ and let $\phi \in B^*$ be a positive functional. Set

$$f_\phi(z) = \phi\big(h_\mu(u + zv) - \mu(\mathcal{X})\big)$$

for $z \in \mathbb{C}^+$. By [9, Remark 2.5], $f_\phi : \mathbb{C}^+ \to \mathbb{C}^+$, and by [6, Lemma 2.3] we have asymptotically

$$\lim_{z\to\infty} f_\phi(z) = 0, \qquad \lim_{z\to\infty} zf_\phi(z) = \phi\big(\mu(\mathcal{X})v^{-1}\mu(\mathcal{X}) - \mu\big(\mathcal{X}v^{-1}\mathcal{X}\big)\big).$$

Thus, by the Nevanlinna representation, there exists a probability measure ρ on $\mathbb{R}$ such that

$$f_\phi(z) = \phi\big(\mu\big(\mathcal{X}v^{-1}\mathcal{X}\big) - \mu(\mathcal{X})v^{-1}\mu(\mathcal{X})\big)\int_{\mathbb{R}} \frac{1}{t - z}\, d\rho(t),$$

and then, by (5), $|f_\phi(z)| \leq \phi(\mu(\mathcal{X}v^{-1}\mathcal{X}) - \mu(\mathcal{X})v^{-1}\mu(\mathcal{X}))/\Im z$.

Now, note that $\Phi(b) := \mu(\mathcal{X}b\mathcal{X}) - \mu(\mathcal{X})b\mu(\mathcal{X}) = \mu([\mathcal{X} - \mu(\mathcal{X})]b[\mathcal{X} - \mu(\mathcal{X})])$ is a positive map, so that

$$\Phi\big(v^{-1}\big) \leq \Phi\big(\|v^{-1}\|\big) = \|v^{-1}\|\big(\mu\big(\mathcal{X}^2\big) - \mu(\mathcal{X})^2\big).$$

Therefore,

$$\phi\big(\mu\big(\mathcal{X}v^{-1}\mathcal{X}\big) - \mu(\mathcal{X})v^{-1}\mu(\mathcal{X})\big) \leq \|v^{-1}\|\phi\big(\mu\big(\mathcal{X}^2\big) - \mu(\mathcal{X})^2\big) \leq \|v^{-1}\|\,\|\mu\big(\mathcal{X}^2\big) - \mu(\mathcal{X})^2\|.$$

In particular,

$$\phi\big(h_\mu(b) - \mu(\mathcal{X})\big) = f_\phi(i) \leq \|v^{-1}\|\,\|\mu\big(\mathcal{X}^2\big) - \mu(\mathcal{X})^2\|.$$

Hence, since any functional on B is the sum of four positive functionals and since B is isometrically embedded in its bidual,

$$\|h_\mu(b) - \mu(\mathcal{X})\| \leq 4\|v^{-1}\|\,\|\mu\big(\mathcal{X}^2\big) - \mu(\mathcal{X})^2\| = 4\frac{\|\mu(\mathcal{X}^2) - \mu(\mathcal{X})^2\|}{\sigma_{\inf}(v)}. \qquad \square$$

We give now a strengthened inequality when μ is bounded. Let μ be a realizable non-commutative distribution, and define $H_\mu : B^+ \to B$ by

$$H_\mu(b) = h_\mu\big(b^{-1}\big) = b^{-1} - F\big(b^{-1}\big).$$

Lemma 5.5. *If μ is bounded by M, then the map H_μ can be extended to an analytic function on the open disk $D_{M^{-1}} := \{b \in B, \|b\| < (M)^{-1}\}$. Moreover, H_μ satisfies the inequality*

$$\|H_\mu(b) - \mu(\mathcal{X})\| \leq \|\mu\big(\mathcal{X}^2\big) - \mu(\mathcal{X})^2\|\frac{1}{\|b\|^{-1} - M}.$$

Proof. By [38], the Boolean cumulant transform of μ is defined by $B_\mu(b) = 1 - F_\mu(b^{-1})b$ for $b \in B^+$. Therefore, $H_\mu(b) = B_\mu(b)b^{-1}$ for $b \in B^+$. The series expansion of B_μ holds for every b in $D_{M^{-1}}$, and we have

$$B_\mu(b) = \sum_{n\geq 1} B_{\mu,n}(b, \dots, b),$$

where $B_{\mu,n} : B^n \to B$ are the non-commutative Boolean cumulants of μ, which satisfy the right B-module property $B_\mu(b_1, \dots, b_n) = B_\mu(b_1, \dots, b_{n-1}, 1)b$. Since $B_{\mu,1}(b) = \mu(\mathcal{X})b$, we have

$$H_\mu(b) = \bigg(\sum_{n\geq 1} B_{\mu,n}(b, \dots, b)\bigg)b^{-1} = \bigg(\sum_{n\geq 1} B_{\mu,n}(b, \dots, 1)b\bigg)b^{-1} = \mu(\mathcal{X}) + \sum_{n\geq 2} B_{\mu,n}(b, \dots, b, 1),$$

on $B^+ \cap D_{M^{-1}}$, and by analytic continuation this equality holds on $D_{M^{-1}}$. Following [38, Lemma 2.9], we introduce on $B(\mathcal{X})$ the B-valued sesquilinear inner-product $\langle P, Q\rangle = \mu(Q^* P)$. Note that $\langle\cdot,\cdot\rangle$ satisfies the B-module condition $\langle Pb, Q\rangle = \langle P, Q\rangle b$ for $b \in B$. We equip $B(\mathcal{X})$ with the semi-norm $\|\cdot\|$ coming from this B-valued inner product and from the norm of B: namely,

$$\|P\| = \big\|\langle P, P\rangle\big\|_B^{1/2}.$$

We recall the B-valued Cauchy–Schwartz inequality [25, p. 3] for B-valued sesquilinear inner-product,

$$\langle x, y\rangle\langle y, x\rangle \leq \big\|\langle y, y\rangle\big\|_B \langle x, x\rangle,$$

which yields the norm inequality

$$\big\|\langle x, y\rangle\big\|_B^2 \leq \big\|\langle y, y\rangle\big\|_B \big\|\langle x, x\rangle\big\|_B.$$

Denote by $B(\mathcal{X})_0$ the complement of 1 in $B(\mathcal{X})$: remark that when $P \in B(\mathcal{X})_0$, then $\langle 1, P\rangle = 0$ and by the B-module structure of the inner product, $\langle b, P\rangle = 0$ for all $b \in B$. By [38, Proof of Theorem 2.5], for $n \geq 2$ we have

$$B_{\mu,n}(b, \dots, b, 1) = \big\langle b(Tb)^{n-2}\xi, \xi\big\rangle,$$

where $\xi = \mathcal{X} - \mu(\mathcal{X})$ and $T : B(\mathcal{X}) \to B(\mathcal{X})$ is defined by $T(b) = 0$ for $b \in B$ and $T(P) = \mathcal{X}P - \mu(\mathcal{X}P)$ for $P \in B(\mathcal{X})_0$. By Lemma 5.3, we have

$$\langle bP, bP\rangle \leq \|b\|^2\langle P, P\rangle \quad \text{and} \quad \langle \mathcal{X}P, \mathcal{X}P\rangle \leq M^2\langle P, P\rangle, \tag{22}$$

for all $b \in B$ and $P \in B(\mathcal{X})$, and where the inequality is understood in the lattice of selfadjoint elements of B. Let $P \in B(\mathcal{X})$.

Hence, the left multiplication by b is a bounded linear map with bound $\|b\|$. Likewise, by the second inequality of (22), the left multiplication by $\mathcal{X}$ is a bounded linear map on $B(\mathcal{X})$ with bound M. Let $P \in B(\mathcal{X})$ and write $P = b + P'$ with $b \in B$ and $P' \in B(\mathcal{X})_0$. Then,

$$\langle TP, TP\rangle = \big\langle \mathcal{X}P' - \mu(\mathcal{X}P'), \mathcal{X}P' - \mu(\mathcal{X}P')\big\rangle = \big\langle \mathcal{X}P', \mathcal{X}P'\big\rangle - \mu(\mathcal{X}P')^*\mu(\mathcal{X}P') \leq \big\langle \mathcal{X}P', \mathcal{X}P'\big\rangle.$$

Thus,

$$\big\|\langle TP, TP\rangle\big\|_B \leq \big\|\langle \mathcal{X}P', \mathcal{X}P'\rangle\big\|_B \leq M^2\big\|\langle P', P'\rangle\big\|_B \leq M^2\big\|\langle P, P\rangle\big\|_B.$$

Therefore, T is also bounded by M. By the B-valued Cauchy–Schwartz inequality and by the above bounds,

$$\big\|B_{\mu,n}(b, \dots, b, 1)\big\| = \big\|\big\langle b(Tb)^{n-2}\xi, \xi\big\rangle\big\| \leq \big\|b(Tb)^{n-2}\xi\big\|\|\xi\| \leq \|b\|^{n-1}(M)^{n-2}\|\xi\|^2.$$

Since $\|\langle\xi, \xi\rangle\| = \|\mu(\mathcal{X}^2) - \mu(\mathcal{X})^2\|$, we conclude that for $\|b\| < M^{-1}$,

$$\big\|H_\mu(b) - \mu(\mathcal{X})\big\| \leq \sum_{n\geq 2}\|b\|^{n-1}M^{n-2}\big\|\mu(\mathcal{X}^2) - \mu(\mathcal{X})^2\big\| \leq \frac{\|b\|}{1 - M\|b\|}\big\|\mu(\mathcal{X}^2) - \mu(\mathcal{X})^2\big\|.$$

□

5.4. *Proof of the additive case*

In this subsection, we are given three B-valued distributions μ_1, μ_2 and μ_3 such that

$$\mu_1 \boxplus \mu_2 = \mu_3,$$

and we want to recover the distribution of μ_2. We suppose without loss of generality that $\mu_1(\mathcal{X}) = 0$, and that all distributions are bounded. Note that the latter condition could be weakened to unbounded distributions admitting moments of order 2 without changing the proof. Since we did not want to introduce affiliated operators, we only are considering the bounded case.

This section is very similar to the scalar case, only the constant K differs. We set $K = 4\sqrt{2}\sigma_1$, and we define

$$B_K := \{b \in B \mid \Im b > K\}.$$

Define moreover the function $\tilde{h}_3(b) = F_{\mu_3}(b) + b$ on B, which is the operator valued version of $\tilde{h}_3$. Recall that for $a, b \in B$ self-adjoint, we write $b > a$ when $b - a > 0$.

Proposition 5.6. *For $b \in B_K$, the function $T_b(w) = h_1(\tilde{h}_3(r) - b) + b$ is well defined and analytic on $\Delta_b = \{r \in B^+, \Im r > 3\Im b/4\}$.*

For any $r \in \Delta_b$, the iterated function $T_b^{\circ n}(r)$ converges to the unique fixed point $w_3(b)$ of Δ_b.

Proof. Let $b \in B_K$. Let $r \in \Delta_b$. Then, $\Im(r) > \frac{3\Im(b)}{4}$, which yields

$$\Im\big(\tilde{h}_3(r) - b\big) = \Im\big(F_{\mu_3}(r) + r - b\big) > 2\frac{3\Im b}{4} - \Im b > \frac{\Im(b)}{2}, \tag{23}$$

where we have used in the first inequality that $\Im[F_{\mu_3}(r)] \geq \Im(r)$ for $r \in B^+$ (see [9]). Since h_1 is defined on B^+, T_b is in particular well-defined.

Since $\mu_1(\mathcal{X}) = 0$ by hypothesis, Lemma 5.4 together with 23 yield

$$\big\|h_{\mu_1}\big(\tilde{h}_3(r) - b\big)\big\| \leq \frac{4\sigma_1^2}{\sigma_{\inf}\Im(\tilde{h}_3(r) - b)} \leq \frac{8\sigma_1^2}{\sigma_{\inf}\Im(b)} \tag{24}$$

for $r \in \Delta_b$. Hence,

$$\begin{aligned}\Im\big[T_b(r)\big] &= \Im\big[h_{\mu_1}\big(\tilde{h}_3(r) - b\big) + b\big] \\ &\geq \Im b - \frac{8\sigma_1^2}{\sigma_{\inf}\Im(b)}.\end{aligned}$$

Since $\sigma_{\inf}\Im(b) > 4\sqrt{2}\sigma_1$,

$$\Im\big[T_b(r)\big] - 3\Im(b)/4 \geq \Im b/4 - \frac{8\sigma_1^2}{\sigma_{\inf}\Im(b)} \geq \sigma_{\inf}\Im(b)/4 - \frac{8\sigma_1^2}{\sigma_{\inf}\Im(b)} > \epsilon$$

for some constant $\epsilon > 0$. Hence, $T_b(\Delta_b) \subset \Delta_b$. Moreover, if $s \notin \Delta_b$, then $\Im s \not> 3\Im b/4$. Hence, there exists a positive functional ϕ with $\|\phi\| = 1$ such that $\phi(\Im s) \leq 3\phi(\Im b)/4$, which yields

$$\phi\big(\Im\big[T_b(r)\big]\big) - \phi(\Im s) > \epsilon$$

for $r \in \Delta_b$, and

$$\big|\phi\big(T_b(r) - s\big)\big| \geq \big|\Im\phi\big(T_b(r) - s\big)\big| = \big|\phi\big(\Im T_b(r)\big) - \phi(\Im s)\big| > \epsilon.$$

Hence, by the isometric embedding of B in the bidual B^*,

$$\|T_b(r)] - s\| = \sup_{\substack{\phi \in B' \\ \|\phi\|_{B'} = 1}} \big|\phi\big(T_b(r) - s\big)\big| > \epsilon.$$

Therefore, $d(\partial\Delta_b, T_b(\Delta_b)) > 0$, and we can apply Earl-Hamilton theorem to the map $T_b : \Delta_b \to \Delta_b$. This implies that for all $r \in \Delta_b$, $T_b^{\circ n}(r)$ converges to the unique fixed point $w_3(b)$ of T_b in Δ_b. □

Proposition 5.7. *The function w_3 is Gateaux analytic on B_K and we have*

$$F_{\mu_2}(b) = F_{\mu_3}\big(w_3(b)\big)$$

for $z \in B_K$.

Proof. Let $a \in B_k$ and $b \in B$. Since B_K is open, there exists a bounded open set $U \subset B$ such that $0 \in U$ and $a + rb \in B_K$ for $r \in U$. We denote by M the bound on U. For $\phi \in B'$, define the function $f(r) = \phi(w_3(a + rb))$ for $r \in U$. By Proposition 5.7, f is the pointwise limit of $f_n(r) = \phi(T_{a+rb}^{\circ n}(a + rb))$. By the definition of T_b, $T_b(b)$ is analytic, which yields that f_n is analytic on U. Moreover, by (24), $\|T_b(w) - b\| \subset \frac{8\sigma_1^2}{K}$ for $b \in B_k$, $w \in \Delta_b$, which implies that

$$\big\|T_{a+rb}^{\circ n}(a + rb)\big\| \leq \|a\| + M\|b\| + \frac{8\sigma_1^2}{K}.$$

Hence, $(f_n)_{n\geq 1}$ a family of uniformly bounded analytic functions which converges pointwise to f, and Montel's theorem implies that $f=\phi\circ w_3$ is analytic. Since this holds for all $\phi\in B'$, w_3 is Gateaux analytic.

Since (24) implies that

$$\|w_3(b)-b\|=\|T_b(w_3(b))-b\|\leq \frac{8\sigma_1^2}{K},$$

for b large enough, $F_{\mu_3}(w_3(b))$ is in the domain of definition of ϕ_{μ_1} and ϕ_{μ_3}. The same reasoning as in the proof of Proposition 3.1 yields that for b large enough,

$$\phi_{\mu_3}\big(F_{\mu_3}(w_3(b))\big)-\phi_{\mu_1}\big(F_{\mu_3}(w_3(b))\big)=b-F_{\mu_3}(w_3(b)).$$

On the other hand, since $\mu_1\boxplus\mu_2=\mu_3$, $\phi_{\mu_1}+\phi_{\mu_2}=\phi_{\mu_3}$ on the intersection of their domain of definition. Therefore, for b large enough,

$$\phi_{\mu_2}\big(F_{\mu_3}(w_3(b))\big)=b-F_{\mu_3}(w_3(b)),$$

which yields

$$F_{\mu_2}(b)=F_{\mu_3}(w_3(b)).\qquad\square$$

5.5. *Proof of the multiplicative case*

Given two realizable bounded non-commutative distributions μ_1 and μ_3 we are interested in finding a realizable distribution μ_2 such that

$$\mu_1\boxtimes\mu_2=\mu_3. \tag{25}$$

We first recall some notations of Theorem 5.2:

- R_i is the bound of the distribution μ_i,
- $\alpha_i:=\|\mu_i(\mathcal{X})\|$ is the norm of the first moment of μ_i, and
- $\alpha_i^*:=\inf\operatorname{Spec}\mu_i(\mathcal{X})$ is the minimum of the spectrum of $\mu_i(\mathcal{X})$.
- $\sigma_i^2:=\|\mu(\mathcal{X}^2)-\mu(\mathcal{X})^2\|$ is the variance of μ_i.

Since we assumed $\mu_1(\mathcal{X})>0$, we have $\alpha_1^*>0$. We introduce the constants

- $K_1:=(R_1+2\frac{\sigma_1^2}{\alpha_1^*})$,
- $K_3:=\sup(\frac{2}{\alpha_1^*}(\sigma_3+\alpha_3)K_1, R_3+\sigma_3)$, and
- $K:=\frac{2}{\alpha_1^*}K_3$.

Lemma 5.8. *Let $\kappa<1$. For all $w\in D_{\kappa K_1^{-1}}$, $H_1(w)$ is well-defined, invertible and*

$$\|H_1(w)^{-1}\|\leq\frac{2}{(2-\kappa)\alpha_1^*}.$$

Proof. Since $\kappa K_1^{-1}\leq R_1^{-1}$, H_1 is well-defined on $D_{\kappa K_1^{-1}}$ by Lemma 5.5. Let $w\in D_{\kappa K_1^{-1}}$. Then, Lemma 5.5 yields that

$$\|H_1(w)-\mu(\mathcal{X})\|\leq\frac{\sigma_1^2}{\|w\|^{-1}-R_1}.$$

Since $\|w\|^{-1}\geq\kappa^{-1}K_1$ and $K_1=R_1+2\frac{\sigma_1^2}{\alpha_1^*}$,

$$\|H_1(w)-\mu(\mathcal{X})\|\leq\frac{\sigma_1^2}{2\kappa^{-1}\sigma_1^2/\alpha_1^*}\leq\frac{\kappa\alpha_1^*}{2}.$$

Thus, there exists $d \in B$ such that $\|d\| \leq \frac{\kappa\alpha_1^*}{2}$ and $H_1(w) = \mu(\mathcal{X}) + d = \mu(\mathcal{X})(1 + \mu(\mathcal{X})^{-1}d)$. By definition of α_1^*, we have $\|\mu(\mathcal{X})^{-1}\| = (\alpha_1^*)^{-1}$, and thus $\|d\mu(\mathcal{X})^{-1}\| \leq \|d\|(\alpha_1^*)^{-1} \leq \kappa/2$. Hence, $(1 + \mu(\mathcal{X})^{-1}d)$ is invertible and

$$\left\|\left(1 + d\mu(\mathcal{X})^{-1}\right)^{-1}\right\| \leq \frac{1}{1 - \kappa/2} \leq 2/(2 - \kappa).$$

Therefore, $H_1(w)$ is also invertible and

$$\left\|H_1(w)^{-1}\right\| \leq \left\|\mu(\mathcal{X})^{-1}\right\| \left\|\left(1 + d\mu(\mathcal{X})^{-1}\right)^{-1}\right\| \leq \frac{2}{(2 - \kappa)\alpha_1^*}.$$

□

We denote by Ω the open set $\{b \in D_{K^{-1}}, b \text{ invertible}\}$. For $b \in \Omega$, we denote by Δ_b the open set $bD_{2(\alpha_1^*)^{-1}}$. Remark that Δ_b always contains the point $b\mu_1(\mathcal{X})^{-1}$, because $\|\mu_1(\mathcal{X})^{-1}\| = (\alpha_1^*)^{-1} < 2(\alpha_1^*)^{-1}$.

For $b \in \Omega$, let $T_b : \Delta_b \to B$ be the function

$$T_b(w) = bH_1\left(b^{-1}\tilde{H}_3(w)\right)^{-1},$$

where we recall that $\tilde{H}_3(w) = wH_3(w)w$ for $\|z\| \leq R_3^{-1}$.

Lemma 5.9. *The map T_b is well-defined on Δ_b, and there is a unique fixed point $w_3(b)$ of T_b in Δ_b. Moreover, for all $w \in \Delta_b$, $T_b^{\circ n}(w)$ converges to $w_3(b)$ as n goes to infinity.*

Proof. Let $b \in \Omega$, so that there exists $\kappa < 1$ such that $\|b\| = \kappa K^{-1}$. Let $w \in \Delta_b$. Then, $w = bw'$ with $w' \in D_{2(\alpha_1^*)^{-1}}$, and thus

$$\|w\| \leq \|b\| \left\|w'\right\| < K^{-1}2\left(\alpha_1^*\right)^{-1} \leq \kappa K_3^{-1}.$$

Since $K_3 = \sup(\frac{2}{\alpha_1^*}(\sigma_3 + \alpha_3)K_1, R_3 + \sigma_3) > R_3$, $H_3(w)$ is well-defined and by Lemma 5.5,

$$\left\|H_3(w) - \mu_3(\mathcal{X})\right\| \leq \frac{\sigma_3^2}{K_3 - R_3} \leq \sigma_3.$$

Hence, $\|H_3(w)\| \leq \alpha_3 + \sigma_3$ and thus

$$\left\|b^{-1}wH_3(w)w\right\| \leq \left\|w'\right\| \left\|H_3(w)\right\| \|w\| \leq \frac{2}{\alpha_1^*}(\alpha_3 + \sigma_3)\kappa K_3^{-1}.$$

Since $K_3 \geq \frac{2}{\alpha_1^*}(\sigma_3 + \alpha_3)K_1$,

$$\left\|b^{-1}wH_3(w)w\right\| \leq \kappa K_1^{-1},$$

Hence, by Lemma 5.8, $H_1(b^{-1}wH_3(w)w)$ is invertible and $\|H_1(b^{-1}wH_3(w)w)^{-1}\| \leq \frac{2}{(2-\kappa)\alpha_1^*}$, which implies that $T_b(w) \in \tilde{\Delta}_b := bD_{2((2-\kappa)\alpha_1^*)^{-1}}$. Remark that $\tilde{\Delta}_b \subset \Delta_b$. In order to apply Earle–Hamilton's theorem it remains to show that $d(\tilde{\Delta}_b, \partial\Delta_b) > 0$. Let $u \notin \Delta_b$ and $v \in \tilde{\Delta}_b$, and set $u' = b^{-1}u$ and $v' = b^{-1}v$. Then, $\|u'\| \geq \frac{2}{\alpha_1^*}$ because $bu' \notin bD_{2(\alpha_1^*)^{-1}}$ and $\|v'\| < 2((2 - \kappa)\alpha_1^*)^{-1}$. Thus,

$$\left\|u' - v'\right\| \geq \left|\left\|u'\right\| - \left\|v'\right\|\right| \geq \frac{2}{\alpha_1^*} - \frac{2}{(2 - \kappa)\alpha_1^*} = \frac{2(1 - \kappa)}{(2 - \kappa)\alpha_1^*}.$$

Since

$$\left\|u' - v'\right\| = \left\|b^{-1}(u - v)\right\| \leq \left\|b^{-1}\right\| \|u - v\|,$$

we deduce that

$$\|u - v\| \geq \frac{2(1 - \kappa)}{\alpha_1^*\|b^{-1}\|},$$

which yields

$$d(\tilde{\Delta}_b, \partial\Delta_b) \geq \frac{2(2-1-\kappa)}{\alpha_1^* \|b^{-1}\|} > 0.$$

Hence, $d(T_b(\Delta_b), \Delta_b^c) > 0$ and T_b satisfies the hypothesis of Earl-Hamilton theorem. There exists thus a unique fixed point $w_3(b)$ of T_b in Δ_b, and for all $w \in \Delta_b$, $K^{\circ n}(w)$ converges to $w_3(b)$ when n goes to infinity. □

We can now turn to the actual computation of the Cauchy transform of F_{μ_2}.

Proposition 5.10. *If* (25) *has a solution, then* F_{μ_2} *is defined by*

$$F_{\mu_2}(b) = bw_3(b^{-1})F_{\mu_3}(w_3(b^{-1})^{-1}),$$

for $b \in B$ *such that* $\inf \operatorname{Spec} b > K$.

Proof. Let us show first that $w_3 : \Omega \to B$ is Gateaux holomorphic and invertible. Let $\phi \in B^*$ and let $a \in \Omega$ and $c \in B$. Since Ω is open, there exist $U \subset \mathbb{C}$ such that for $z \in U$, $a + zc \in \Omega$. Define $f_n : U \to \mathbb{C}$ by $f_n(z) = \phi(T_{a+zc}^{\circ n}(0))$. Since $0 \in \Delta_b$ for all $b \in \Omega$, f_n is well-defined on U. Moreover, H_1 and $\tilde{H}_3$ are analytic, thus $b \mapsto T_b^{\circ n}(0)$ is analytic on Ω for all $n \geq 1$. Therefore, each map f_n is analytic on U. Since $T_b^n(w) \in \Delta_b$ for all $n \geq 1$,

$$\left\| T_b^{\circ n}(w) \right\| \leq 2\|b\|/\alpha_1^* \leq \frac{2K^{-1}}{\alpha_1^*},$$

for all $b \in \Omega$, $w \in \Delta_b$ and $n \geq 1$. Hence, the family $(f_n)_{n\geq 1}$ is uniformly bounded and converges pointwise, which yields by Montel's theorem that $(f_n)_{n\geq 1}$ converges uniformly to a holomorphic function f. By Lemma 5.9, we already now that $f(z) = \phi(w_3(a + zb))$ which yields the Gateaux holomorphicity of w_3.

For b small enough, set $w(b) = \eta_{\mu_3}^{\langle -1 \rangle} \eta_{\mu_2}(b) = w_2^{\langle -1 \rangle}(b)$, where $w_2(b)$ is the function introduced in [7, Theorem 2.2]. By Lemma 5.5 and the definition of η_μ, we have $\eta_\mu(b) \sim b\mu(\mathcal{X})$ as b goes to zero. Therefore, $w(b) \sim b\mu_2(\mathcal{X})\mu_3(\mathcal{X})^{-1}$ as b goes to zero. Moreover, by [7, Theorem 2.2 (3)],

$$w_2(b) = bH_1(H_{\mu_2}(w_2(b))b), \tag{26}$$

and by definition of w_2, $\eta_{\mu_3}(b) = \eta_{\mu_2}(w_2(b))$. Hence, since we have also $\eta_{\mu_2}(b) = \eta_{\mu_3}(w(b))$, evaluating (26) on $w(b)$ yields

$$b = w(b)H_1(H_{\mu_2}(b)w(b)).$$

By Lemma 5.5, $H_{\mu_2}(b)$ converges to $\mu_2(\mathcal{X})$ as b goes to zero; hence, by Lemma 5.8, for b small enough $H_1(H_{\mu_2}(b)w(b))$ is invertible with $\|H_1(H_{\mu_2}(b)w(b))^{-1}\| < \alpha_1^*/2$, which yields

$$w(b) = bH_1(H_{\mu_2}(b)w(b))^{-1} \in \Delta_b.$$

Since $H_{\mu_2}(b) = b^{-1}\eta_{\mu_2}(b) = b^{-1}\eta_{\mu_3}(w(b))$, the latter equation yields

$$w(b) = bH_1(b^{-1}\eta_{\mu_3}(w(b))w(b))^{-1} = bH_1(b^{-1}\tilde{H}_3(w(b)))^{-1} = T_b(w(b)). \tag{27}$$

Therefore, $w(b)$ is a fixed point of T_b. Since $w(b) \in \Delta_b$, we must have $w(b) = w_3(b)$ by Lemma 5.9. Since $\eta_{\mu_2}(b) = \eta_{\mu_3}(w(b))$, this yields $\eta_{\mu_2}(b) = \eta_{\mu_3}(w_3(b))$ for b small enough. The functions η_{μ_2} and $\eta_{\mu_3} \circ w_3$ are two Gateaux holomorphic maps defined on the connected domain Ω and they coincide on an open subset of Ω, thus they are equal on Ω, and we have

$$\eta_{\mu_2}(b) = \eta_{\mu_3}(w_3(b))$$

for $b \in \Omega$. Let $b \in B$ be such that $\inf \operatorname{Spec} b > K$. Then, b is invertible and $b^{-1} \in \Omega$. Therefore,

$$\begin{aligned} F_{\mu_2}(b) &= b(1 - \eta_{\mu_2}(b^{-1})) \\ &= b(1 - \eta_{\mu_3}(w_3(b^{-1})) \\ &= bw_3(b^{-1})F_{\mu_3}(w_3(b^{-1})^{-1}). \end{aligned}$$

□

References

[1] M. Andersen, J. Dahl and L. Vandenberghe. CVXOPT: A Python package for convex optimization, 2013. Available at abel.ee.ucla.edu/cvxopt.

[2] O. Arizmendi, I. Nechita and C. Vargas. On the asymptotic distribution of block-modified random matrices. *J. Math. Phys.* **57** (2016) 015216. MR3432744 https://doi.org/10.1063/1.4936925

[3] Z. Bai, J. Chen and J. Yao. On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52** (4) (2010) 423–437. MR2791528 https://doi.org/10.1111/j.1467-842X.2010.00590.x

[4] J. Baik, G. Ben Arous and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** (5) (2005) 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

[5] S. Belinschi and H. Bercovici. A new approach to subordination results in free probability. *J. Anal. Math.* **101** (2007) 357–365. MR2346550 https://doi.org/10.1007/s11854-007-0013-1

[6] S. Belinschi, T. Mai and R. Speicher. Analytic subordination theory of operator-valued free additive convolution and the solution of a general random matrix problem. *J. Reine Angew. Math.* **732** (2017) 21–53. MR3717087 https://doi.org/10.1515/crelle-2014-0138

[7] S. Belinschi, R. Speicher, J. Treilhard and C. Vargas. Operator-valued free multiplicative convolution: Analytic subordination theory and applications to random matrix theory. *Int. Math. Res. Not.* **14** (2015) 5933–5958. MR3384463 https://doi.org/10.1093/imrn/rnu114

[8] S. T. Belinschi, H. Bercovici, M. Capitaine and M. Février. Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.* **45** (6A) (2017) 3571–3625. MR3729610 https://doi.org/10.1214/16-AOP1144

[9] S. T. Belinschi, M. Popa and V. Vinnikov. Infinite divisibility and a noncommutative Boolean-to-free Bercovici–Pata bijection. *J. Funct. Anal.* **262** (1) (2012) 94–123. MR2852257 https://doi.org/10.1016/j.jfa.2011.09.006

[10] F. Benaych-Georges. Rectangular random matrices, related convolution. *Probab. Theory Related Fields* **144** (3–4) (2009) 471–515. MR2496440 https://doi.org/10.1007/s00440-008-0152-z

[11] F. Benaych-Georges and M. Debbah. Free deconvolution: From theory to practice. In *Paradigms for Biologically-Inspired Autonomic Networks and Services*, 2010.

[12] H. Bercovici and D. Voiculescu. Free convolution of measures with unbounded support. *Indiana Univ. Math. J.* **42** (3) (1993) 733–773. MR1254116 https://doi.org/10.1512/iumj.1993.42.42033

[13] P. Biane. Processes with free increments. *Math. Z.* **227** (1998) 143–174. MR1605393 https://doi.org/10.1007/PL00004363

[14] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004. MR2061575 https://doi.org/10.1017/CBO9780511804441

[15] J. Bun, J. P. Bouchaud and M. Potters. Cleaning large correlation matrices: Tools from random matrix theory. *Phys. Rep.* **666** (2017) 1–109. MR3590056 https://doi.org/10.1016/j.physrep.2016.10.005

[16] M. Capitaine and C. Donati-Martin. Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56** (2) (2007) 767–803. MR2317545 https://doi.org/10.1512/iumj.2007.56.2886

[17] R. Couillet and M. Debbah. *Random Matrix Methods for Wireless Communications*. Cambridge University Press, Cambridge, 2011. MR2884783 https://doi.org/10.1017/CBO9780511994746

[18] A. Denjoy. Sur l'itération des fonctions analytiques. *C. R. Acad. Sci.* **182** (1926) 255–257.

[19] C. J. Earle and R. S. Hamilton. A fixed point theorem for holomorphic mappings. In *Proc. Sympos. Pure Math.* **XVI** 61–65. American Mathematical Society, Providence, 1970. MR0266009

[20] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** (6) (2008) 2757–2790. MR2485012 https://doi.org/10.1214/07-AOS581

[21] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations*. Boston Pitman Publication, 1984. MR0742928

[22] T. Hasebe. Monotone convolution and monotone infinite divisibility from complex analytic viewpoint. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **13** (1) (2010) 111–131. MR2646794 https://doi.org/10.1142/S0219025710003973

[23] V. Kargin. A concentration inequality and a local law for the sum of two random matrices. *Probab. Theory Related Fields* **154** (3–4) (2012) 677–702. MR3000559 https://doi.org/10.1007/s00440-011-0381-4

[24] W. Kong and G. Valiant. Spectrum estimation from samples. *Ann. Statist.* **45** (5) (2017) 2218–2247. MR3718167 https://doi.org/10.1214/16-AOS1525

[25] E. Lance. *Hilbert C^*-Modules: A Toolkit for Operator Algebraists*. *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1995. MR1325694 https://doi.org/10.1017/CBO9780511526206

[26] O. Ledoit and S. Péché. Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** (1–2) (2011) 233–264. MR2834718 https://doi.org/10.1007/s00440-010-0298-3

[27] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariancematrices. *J. Multivariate Anal.* **88** (2004) 365–411. MR2026339 https://doi.org/10.1016/S0047-259X(03)00096-4

[28] O. Ledoit and M. Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** (2) (2012) 1024–1060. MR2985942 https://doi.org/10.1214/12-AOS989

[29] O. Ledoit and M. Wolf. Numerical implementation of the QuEST function. *Comput. Statist. Data Anal.* **115** (2017) 199–223. MR3683138 https://doi.org/10.1016/j.csda.2017.06.004

[30] W. Li and J. Yao. A local moment estimator of the spectrum of a large dimensional covariance matrix. *Statist. Sinica* **24** (2014) 919–936. MR3235405

[31] H. Maassen. Addition of freely independent random variables. *J. Funct. Anal.* **106** (1992) 409–438. MR1165862 https://doi.org/10.1016/0022-1236(92)90055-N

[32] V. Marchenko and L. Pastur. Distribution of eigenvalues of some sets of random matrices. *Math. USSR, Sb.* **1** (1967) 457–486.

[33] X. Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inf. Theory* **54** (11) (2008) 5113–5129. MR2589886 https://doi.org/10.1109/TIT.2008.929938

[34] J. Mingo and R. Speicher. *Free Probability and Random Matrices*. *Fields Institute Monographs*. Amer. Math. Soc., Providence, RI, 2017. MR3585560 https://doi.org/10.1007/978-1-4939-6942-5

[35] A. Nica, D. Shlyaktenko and R. Speicher. Operator-valued distributions I: Characterizations of freeness. *Int. Math. Res. Not.* **29** (2002) 1509–1538. MR1907203 https://doi.org/10.1155/S1073792802201038

[36] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. *LMS Lecture Note Series* **335**. Cambridge University Press, Cambridge, 2006. MR2266879 https://doi.org/10.1017/CBO9780511735127
[37] J. Pennington, S. Schoenholz and S. Ganguli. The emergence of spectral universality in deep networks. In *Proceedings of Machine Learning Research, 84 International Conference on Artificial Intelligence and Statistics*, 2018.
[38] M. Popa and V. Vinnikov. Non-commutative functions and the non-commutative free Lévy–Hinčin formula. *Adv. Math.* **236** (2013) 131–157. MR3019719 https://doi.org/10.1016/j.aim.2012.12.013
[39] N. R. Rao, J. A. Mingo, R. Speicher and A. Edelman. Statistical eigen-inference from large Wishart matrices. *Ann. Statist.* **36** (6) (2008) 2850–2885. MR2485015 https://doi.org/10.1214/07-AOS583
[40] Ø. Ryan and M. Debbah. Free deconvolution for signal processing applications. In *Proceedings of IEEE International Symposium of Information Theory (ISIT' 07)* 1846–1850. *Nice, France*, 2007.
[41] Ø. Ryan and M. Debbah. Multiplicative free convolution and information-plus-Noise type matrices. Preprint. Available at arXiv:math/0702342.
[42] D. Shlyaktenko. Random Gaussian band matrices and freeness with amalgamation. *Int. Math. Res. Not.* **20** (1996) 1013–1025. MR1422374 https://doi.org/10.1155/S1073792896000633
[43] R. Speicher. Multiplicative functions on the lattice of non-crossing partitions and free convolution. *Math. Ann.* **298** (1994) 611–628. MR1268597 https://doi.org/10.1007/BF01459754
[44] R. Speicher. Combinatorial theory of the free product with amalgamation and operator-valued free probability theory. *Mem. Amer. Math. Soc.* **132** (627) (1998) x+88 pp. MR1407898 https://doi.org/10.1090/memo/0627
[45] W. Tarnowski, P. Warchol, S. Jastrzebski, J. Tabor and M. Nowak. Dynamical isometry is achieved in residual networks in a universal way for any activation function. In *Proceedings of Machine Learning Research, 89 the 22nd International Conference on Artificial Intelligence and Statistics*, 2019.
[46] D. Voiculescu. Symmetries of some reduced free product C^*-algebras. In *Operator Algebras and Their Connections with Topology and Ergodic Theory (Busteni, 1983)* 556–588. *Lecture Notes in Math.* **1132**. Springer, Berlin, 1985. MR0799593 https://doi.org/10.1007/BFb0074909
[47] D. Voiculescu. Addition of certain non-commutative random variables. *J. Funct. Anal.* **66** (1986) 323–346. MR0839105 https://doi.org/10.1016/0022-1236(86)90062-5
[48] D. Voiculescu. Multiplication of certain non-commutative random variables. *J. Oper. Theory* **18** (1987) 223–235. MR0915507
[49] D. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.* **104** (1991) 201–220. MR1094052 https://doi.org/10.1007/BF01245072
[50] D. Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory. I. *Comm. Math. Phys.* **155** (1) (1993) 71–92. MR1228526
[51] D. Voiculescu. Operations on certain non-commutative operator-valued random variables. *Astérisque* **232** (1995) 243–275. MR1372537
[52] D. Voiculescu. The coalgebra of the free difference quotient and free probability. *Int. Math. Res. Not.* **2** (2000) 79–106. MR1744647 https://doi.org/10.1155/S1073792800000064
[53] E. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* **67** (1958) 325–327. MR0095527 https://doi.org/10.2307/1970008
[54] J. D. Williams. Analytic function theory for operator-valued free probability. *J. Reine Angew. Math.* **729** (2017) 119–149. MR3680372 https://doi.org/10.1515/crelle-2014-0106
[55] J. Wolff. Sur l'itération des fonctions holomorphes dans une région, et dont les valeurs appartiennent a cette région. *C. R. Acad. Sci.* **182** (1926) 42–43.