Statistical deconvolution of the free Fokker-Planck equation at fixed time

MYLÈNE MAÏDA 1, TIEN DAT NGUYEN 2,*, THANH MAI PHAM NGOC 2,** VINCENT RIVOIRARD 3 and VIET CHI TRAN 4

1 Univ. Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé, F-59000 Lille, France. E-mail: mylene.maida@univ-lille.fr

2 Laboratoire de Mathématiques d’Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405, Orsay, France. E-mail: *tien-dat.nguyen@math.u-psud.fr; **thanh.pham_ngoc@math.u-psud.fr

3 Ceremade, CNRS, UMR 7534, Université Paris-Dauphine, PSL Research University, 75016 Paris, France. E-mail: Vincent.Rivoirard@dauphine.fr

4 LAMA, Univ Gustave Eiffel, Univ Paris Est Creteil, CNRS, F-77454 Marne-la-Vallée, France. E-mail: chi.tran@univ-eiffel.fr

We are interested in reconstructing the initial condition of a non-linear partial differential equation (PDE), namely the Fokker-Planck equation, from the observation of a Dyson Brownian motion at a given time \( t > 0 \). The Fokker-Planck equation describes the evolution of electrostatic repulsive particle systems, and can be seen as the large particle limit of correctly renormalized Dyson Brownian motions. The solution of the Fokker-Planck equation can be written as the free convolution of the initial condition and the semi-circular distribution. We propose a nonparametric estimator for the initial condition obtained by performing the free deconvolution via the subordination functions method. This statistical estimator is original as it involves the resolution of a fixed point equation, and a classical deconvolution by a Cauchy distribution. This is due to the fact that, in free probability, the analogue of the Fourier transform is the R-transform, related to the Cauchy transform. In past literature, there has been a focus on the estimation of the initial conditions of linear PDEs such as the heat equation, but to the best of our knowledge, this is the first time that the problem is tackled for a non-linear PDE. The convergence of the estimator is proved and the integrated mean square error is computed, providing rates of convergence similar to the ones known for non-parametric deconvolution methods. Finally, a simulation study illustrates the good performances of our estimator.

MSC 2010 subject classifications: Primary 35Q62; 65M32; 62G05; 46L53; 35R30; 60B20; 46L54.

Keywords: PDE with random initial condition, Free deconvolution, Inverse problem, Non-parametric kernel estimation, Fourier transform, Mean Integrated Error, Dyson Brownian motion.

1. Introduction

Letting the initial condition of a partial differential equation (PDE) be random is interesting for considering complex phenomena or for introducing uncertainty and irregularity in the initial state. There is a large literature on the subject, and we can mention that this has been studied for the Navier-Stokes equation, to account for the turbulence arising in fluids with high velocities and low viscosities (see [34, 15]), for the Burgers equation that is used in astrophysics (see [9, 5, 20, 19] or also the survey by...
[33]), for the wave equations, to study the solutions with low-regularity initial data (see [10, 11, 32]) or for the Schrödinger PDE (see [8]). The Burgers PDE or the vortex equation, associated to the Navier-Stokes PDE by considering the curl of the velocity, are of the McKean-Vlasov type as introduced and studied in [26, 22]. Numerical approximations of such PDEs with random initial conditions have been considered in [27, 30]. In this paper, we are interested in the Fokker-Planck PDE which is another case of McKean-Vlasov PDE [13]. This equation models the motion of particles with electrostatic repulsion and a probabilistic interpretation that we will adopt has been considered in [7].

A question naturally raised in this context is to estimate the random initial condition, given the observation of the PDE solution at a given fixed time $t > 0$. For linear PDEs, this inverse problem is solved by deconvolution techniques, and this has been explored for PDEs such as the heat equation or the wave equation by Pensky and Sapatinas [24, 25]. For the 1d-heat equation, it is known that the solution at time $t$, say $\nu_t(dx)$, is the convolution of the initial condition $\nu_0(dx)$ with Green function $G_t$, which is a Gaussian transition function associated with the standard Brownian motion $(B_t)_{t \geq 0}$. The probabilistic interpretation of the heat equation is built on this observation, and $\nu_t$ can be viewed as the distribution of $X_t = X_0 + B_t$ where $X_0$ is distributed as $\nu_0$. Taking the Fourier transforms changes the convolution problem into a multiplication, which paves the way to reconstruct the initial condition. Here, we are interested in estimating the initial condition of a non-linear PDE, namely the Fokker-Planck equation, from the observation of its solution at time $t$. This equation is:

$$\partial_t p(t, x) = -\partial_x \int_{\mathbb{R}^2} H(p(t, x)p(t, y)dy,$$

with

$$H(p(t, x) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}\setminus[x-\varepsilon, x+\varepsilon]} \frac{1}{x-y} p(t, y)dy,$$

and for $t \in \mathbb{R}^+$, $x \in \mathbb{R}$, and initial condition $p_0(x) \in L^1(\mathbb{R})$. Contrarily to the examples considered in [24, 25], this PDE is non-linear of the McKean-Vlasov type with logarithmic interactions. To the best of our knowledge, this is the first work devoted to the deconvolution of a non-linear PDE to recover the initial condition. The choice of this equation is motivated by its strong similarities with the heat equation: the standard Brownian motion of the probabilistic interpretation is replaced here by the free Brownian motion $(h_t)_{t \geq 0}$ (operator-valued), and the usual convolution by a Gaussian distribution is replaced by the free convolution by a semi-circular distribution $\sigma_t$ characterized by its density with respect to the Lebesgue measure:

$$\sigma_t(dx) = \frac{1}{2\pi t} \sqrt{4t - x^2} \mathbb{1}_{[-2\sqrt{t}, 2\sqrt{t}]}(x) dx.$$

If $x_0$ admits the spectral measure $\mu_0$, then $x_t = x_0 + h_t$ admits

$$\mu_t = \mu_0 \boxtimes \sigma_t,$$

as spectral measure, where the operation $\boxtimes$ is the free convolution and has been introduced by Voiculescu in [35]. It can be proved that the density $p(t, \cdot)$ of $\mu_t$ solves (1.1).

For the Fokker-Planck equation, the inverse problem boils down to a free deconvolution, where it was a usual deconvolution for the heat equation. Recently, the problem of free deconvolution has been studied
by Arizmendi, Tarrago and Vargas [2]. To solve (1.3) in a general setting, subordination functions are used. Here, if the Cauchy transform of a measure \( \mu \) is defined as \( G_\mu(z) = \int \frac{1}{x-z} \, d\mu(x) \) for \( z \in \mathbb{C}^+ \), where \( \mathbb{C}^+ \) is the set of complex numbers with positive imaginary part, the subordination function \( w_{fp}(z) \) at time \( t \) is related to \( G_\mu_t \) by the functional equation

\[
  w_{fp}(z) = z + tG_\mu_t(x_{fp}(z)). \tag{1.4}
\]

From this, we can recover \( G_{\mu_0} \) with the formula \( G_{\mu_0}(z) = G_{\mu_t}(w_{fp}(z)) \) and thus \( \mu_0 \) (see Lemma 2.7 and (2.12)). More precisely, we prove in Section 2.3 that for any \( \gamma > 2\sqrt{t} \), \( f_{\mu_0,C_{\gamma}} \) the density of the classical convolution of \( \mu_0 \) with the Cauchy distribution of parameter \( \gamma \), defined by its density \( f_{\gamma}(x) := \gamma/(\pi(x^2 + \gamma^2)) \), satisfies

\[
  f_{\mu_0,C_{\gamma}}(x) = \frac{1}{\pi t} |\gamma - \text{Im}w_{fp}(x + i\gamma)|, \quad x \in \mathbb{R}. \tag{1.5}
\]

Then, estimating \( p_0 \), the density of \( \mu_0 \), requires an estimation of the subordination function \( w_{fp} \) combined with a classical deconvolution step from a Cauchy distribution.

**Observations** Additionally to the free deconvolution problem, our observation does not consist in the operator-valued random variable \( X_t \) but in its matricial counterpart. More precisely, we observe a matrix \( X^n(t) \) for a given \( t > 0 \), assumed to be fixed in the sequel, where

\[
  X^n(t) = X^n(0) + H^n(t), \quad t \geq 0 \tag{1.6}
\]

with \( X^n(0) \) a diagonal matrix whose entries are the ordered statistic \( \lambda^n_i(0) < \cdots < \lambda^n_n(0) \) of a vector \( (d^n_i)_{i \in \{1, \ldots, n\}} \) of \( n \) independent and identically distributed (i.i.d.) random variables distributed as \( \mu_0(dx) = \mu_0(x) \, dx \), absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R} \), and \( H^n(t) \) a standard Hermitian Brownian motion, as defined in Definition 2.1. As the distribution of \( H^n(t) \) is invariant by conjugation, choosing \( X^n(0) \) to be a diagonal matrix is not restrictive.

The purpose is to estimate \( p_0 \). The observation consists in the matrix \( X^n(t) \) at the fixed time \( t \), from which we can compute the eigenvalues \( (\lambda^n_i(t), \ldots, \lambda^n_n(t)) \) and then the associated empirical measure. Of course, we do not observe directly the initial condition \( X^n(0) \).

Let us now explain the link with the Fokker-Planck equation at the level of the particle system of the eigenvalues. It is known that the eigenvalues \( (\lambda^n_1(t), \ldots, \lambda^n_n(t)) \) of \( X^n(t) \) solve the following system of stochastic differential equations (SDE):

\[
  d\lambda^n_i(t) = \frac{1}{\sqrt{n}} d\beta_i(t) + \frac{1}{n} \sum_{j \neq i} \frac{df}{\lambda^n_i(t) - \lambda^n_j(t)}, \quad 1 \leq i \leq n, \tag{1.7}
\]

where \( \beta_i \) are i.i.d. standard real Brownian motions. A rigorous proof can be found e.g. in [1, page 249, Th. 4.3.12], but for the sake of clarity, we will give in Appendix A a short heuristic explanation on how this particle system arises. Now, if we denote by

\[
  \mu^n_{\gamma} = \frac{1}{n} \sum_{i=1}^n \delta_{\gamma^{\lambda^n_i(t)}} \tag{1.8}
\]
the empirical measure of these eigenvalues at time \( t \), then the process \( (\mu_t^n)_{t \geq 0} \) converges weakly almost surely as \( n \) goes to infinity to the process \( (\mu_t)_{t \geq 0} \) with density \( (p(t, \cdot))_{t \geq 0} \) solution of (1.1), as stated in Proposition 2.3 below. For \( n = 1 \), we recover the classical heat equation as the Dyson Brownian motion boils down to a standard Brownian motion.

**Contributions** Relying on the analysis of [2], we provide, in Theorem-Definition 2.8, a statistical estimator \( \hat{w}_{fp}^n(z) \) for the subordination function. As the Cauchy transform \( G_{\mu_t} \) in (1.4) is not invertible on the whole domain \( \mathbb{C}^+ \), the subordination function \( w_{fp}(z) \) will be defined only for \( z \in \mathbb{C}_{2\sqrt{t}} \) where \( \mathbb{C}_\gamma := \{ z \in \mathbb{C}^+, \ \text{Im}(z) > \gamma \} \). We shall prove the following result.

**Proposition 1.1.** Let \( \gamma > 2\sqrt{t} \). Suppose \( p_0 \) satisfies the condition

\[
\int_{\mathbb{R}} \log(x^2 + 1)p_0(x)dx < +\infty. \tag{1.9}
\]

Then:
(i) For any \( z \in \mathbb{C}_{2\sqrt{t}} \), the estimator \( \hat{w}_{fp}^n(z) \) converges almost surely to \( w_{fp}(z) \) as \( n \to \infty \).
(ii) The convergence is uniform on \( \mathbb{C}_\gamma \).
(iii) We have the following convergence rate on \( \mathbb{C}_\gamma \):

\[
\sup_{n \in \mathbb{N}} \sup_{z \in \mathbb{C}_\gamma} \mathbb{E} \left[ n \left| \hat{w}_{fp}^n(z) - w_{fp}(z) \right|^2 \right] < +\infty.
\]

Observe that Condition (1.9) corresponds to the more general assumption that

\[
\sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^{n} \log (\lambda_t^n(0)^2 + 1) < \infty \quad \text{almost surely (a.s.)}
\]

in [1, Proposition 4.3.10] and ensures the convergence of \( p^n \) to the solution of (1.1) by adapting the proof of [1] to our context with a random initial condition. The simplified Condition (1.9) uses here the assumption that the diagonal entries in \( X^n(0) \) are i.i.d.

To obtain uniform convergence and fluctuations (ii) and (iii), we will need to restrict to strict subdomains of \( \mathbb{C}_{2\sqrt{t}} \). The fluctuations (iii) are established in the line of the work of Dallaporta and Février [18].

Proposition 1.1 is the crucial tool to reach the main goal of this paper, namely providing an estimator of \( p_0 \). As explained previously, we estimate \( p_0 \) by combining a free deconvolution step via the use of \( \hat{w}_{fp}^n \) with then a classical deconvolution step. We define our final estimator \( \hat{p}_{0,h} \) via its Fourier transform, denoted \( \hat{p}_{0,h} \); from Equation (1.5), it is natural to define it as follows:

\[
\hat{p}_{0,h}(\xi) = e^{\gamma|\xi|} K_h^*(\xi). \frac{1}{\pi t} \left[ \gamma - \text{Im} \ \hat{w}_{fp}(\cdot + i \gamma)^*(\xi) \right], \quad \xi \in \mathbb{R}.
\]

Note that, as usual in nonparametric statistics, the last expression depends on \( K_h^* \), a regularization term defined through the Fourier transform of a kernel function \( K_h \) depending on a bandwidth parameter \( h \).
We study theoretical properties of $\hat{p}_{0,h}$ by deriving asymptotic rates of the mean integrated square error of $\hat{p}_{0,h}$ decomposed as the sum of bias and variance terms. The study of the variance term is intricate and is based on sharp controls of the difference $\hat{w}_{nfp}(z) - w_{fp}(z)$ provided by Proposition 1.1.

We show in Theorem 4.1 that the variance term is of order $e^{2\gamma h}/n$ as desired for deconvolution with the Cauchy distribution with parameter $\gamma$. The bias term is driven by the smoothness properties of the function $p_0$. In particular, when $p_0$ belongs to a space of supersmooth densities (see (4.5)), we establish convergence rates, after an appropriate (non-adaptive) choice of the bandwidth parameter $h$, see Corollary 4.3. For instance, if $\int_{\mathbb{R}}|p_0(\xi)|^2\exp(2a|\xi|)d\xi \leq L$ for $0 < L < \infty$, then $\mathbb{E}\left[\|\hat{p}_{0,h} - p_0\|^2\right] = O(n^{-\frac{a}{a+\gamma}})$.

The case of Sobolev regularities is tackled in Corollary 4.5 leading to logarithmic rates of convergence. We then discuss the connections of our rates of convergence with those obtained in the classical statistical density deconvolution problem involving a Cauchy distribution of parameter $\gamma$. Note that the exponent in the previous bound reflects the difficulty of our statistical problem: the larger $\gamma$, the slower the rate. Remembering that $\gamma$ is connected to the observational time $t$ through the condition $\gamma > 2\sqrt{t}$, it means that for the previous example, our estimate can achieve the polynomial rate $n^{-\frac{a}{a+2\sqrt{t}+\epsilon}}$ for any $\epsilon > 0$. The question of whether it is possible to consider smaller values for $\gamma$ constitutes a challenging problem. Adaptive choices for $h$ are also a very interesting issue. These problems will be investigated in another work.

Overview of the paper

Important results, namely Theorem 4.1, Corollary 4.3 and Corollary 4.5, which constitute the main contributions of the paper, are contained in Section 4. Before that, in Section 2, we study the free deconvolution and explain the construction of the estimator $\hat{p}_{0,h}$ of $p_0$. Existence results and properties of the subordination functions are precisely stated and proved. Section 3 is devoted to a deeper study of the subordination function and to the proof of Proposition 1.1. Numerical simulations are provided in Section 5.

Notations:

For any $z = u + iv \in \mathbb{C}^+$, we denote $\sqrt{z} := a + ib \in \mathbb{C}$ with $a = \sqrt{(u^2 + v^2 + u)/2}$ and $b = \sqrt{(u^2 + v^2 - u)/2}$. We denote the Fourier transform of a function $g \in L^1(\mathbb{R})$ by $g^* : \xi \in \mathbb{R} \mapsto \int_{\mathbb{R}} g(x)e^{ix\xi}dx$.

2. Free deconvolution of the Fokker-Planck equation

2.1. Dyson Brownian motions

Let us denote by $\mathcal{H}_n(\mathbb{C})$ the space of $n$-dimensional matrices $H_n$ such that $(H_n)^* = H_n$.

Definition 2.1. Let $(B_{i,j}, \tilde{B}_{i,j}, 1 \leq i \leq j \leq n)$ be a collection of i.i.d. real valued standard Brownian motions, the Hermitian Brownian motion, denoted $H^n \in \mathcal{H}_n(\mathbb{C})$, is the random process with entries
\{(H^n(t))_{k,\ell}, t \geq 0, 1 \leq k, \ell \leq n\} equal to
\[(H^n)_{k,\ell} = \begin{cases} 
\frac{1}{\sqrt{2n}} \left( B_{k,\ell} + i \tilde{B}_{k,\ell} \right), & \text{if } k < \ell \\
\frac{1}{\sqrt{n}} B_{k,k}, & \text{if } k = \ell 
\end{cases} \tag{2.1}
\]

Let us now define the initial condition, that we will choose independent of the Hermitian Brownian motion $H^n$. Recall that $\mu_0$ is a probability measure with density $p_0(x)$ with respect to the Lebesgue measure on $\mathbb{R}$. Without loss of generality, we can choose the initial condition $X^n(0)$ to be a diagonal matrix, with entries $(\lambda^n_1(0), \ldots, \lambda^n_n(0))$ the ordered statistics of i.i.d. random variables $(d^n_i)_{1 \leq i \leq n}$ with distribution $\mu_0$.

For $t \geq 0$, let $\lambda^n(t) = (\lambda^n_1(t), \ldots, \lambda^n_n(t))$ denote the ordered collection of eigenvalues of
\[X^n(t) = X^n(0) + H^n(t). \tag{2.2}\]

**Theorem 2.2 (Dyson).** The process $(\lambda^n(t))_{t \geq 0}$ is the unique solution in $C(\mathbb{R}_+, \mathbb{R}^n)$ of the system (1.7) with initial condition $\lambda^n_0(0)$ and where $\beta_i$ are i.i.d. real valued standard Brownian motions. With probability one and for all $t > 0$, $\lambda^n_1(t) < \ldots < \lambda^n_n(t)$.

Moreover, we have for any fixed $T > 0$, the convergence of the process of empirical measures $(\mu^n_t)_{t \geq 0}$ as defined in (1.8), viewed as an element of $C([0,T], M_1(\mathbb{R}))$, the space of continuous processes from $[0,T]$ into the space $M_1(\mathbb{R})$ of probability measure on $\mathbb{R}$, equipped with its weak topology.

**Proposition 2.3.** Under Assumption (1.9), for any fixed time $T < \infty$, $(\mu^n_t)_{t \in [0,T]}$ converges almost surely in $C([0,T], M_1(\mathbb{R}))$. Moreover, its limit is the unique measure-valued process $(\mu_t)_{t \in [0,T]}$ whose densities satisfy (1.1) with initial condition $p_0$.

For deterministic initial conditions, Theorem 2.2 and Proposition 2.3 are classical results and we refer to [1, Section 4.3] for a proof. Both results can be easily extended to random initial conditions, independent of the Hermitian Brownian motion itself. For details, we refer to [23].

### 2.2. Free deconvolution by subordination method

Our starting point is (1.3), for a fixed time $t > 0$. Recovering $\mu_0$ knowing $\mu_t$ is a free deconvolution problem. The generic problem of free deconvolution has been introduced and studied by Arizmendi et al. [2] with the use of the Cauchy transform instead of the Fourier transform. We briefly recall their results, and adapt them to the present setting where one of the measures is the semi-circular distribution. The free convolution with a semi-circular distribution allows notably to exhibit better constants in Theorem 2.6 than the ones of Arizmendi et al. [2] who work in full generality. From a statistical point of view, this is central in improving the convergence rates to estimate $p_0$. Before, we need to introduce a few notations and definitions.
Definition 2.4. Let $\mu$ be a probability measure on $\mathbb{R}$. The Cauchy transform of $\mu$ is defined by:

$$G_\mu(z) = \int_\mathbb{R} \frac{d\mu(x)}{z - x}, \quad z \in \mathbb{C} \setminus \mathbb{R}. \quad (2.3)$$

The fact is that $G_\mu(z) = \overline{G_\mu(\bar{z})}$, so the behavior of the Cauchy transform in the lower half-plane $\mathbb{C}^- = \{z \in \mathbb{C} | \text{Im}(z) < 0\}$ can be determined by its behavior in the upper half-plan $\mathbb{C}^+ = \{z \in \mathbb{C} | \text{Im}(z) > 0\}$. The function $G_\mu$ is a bijection from a neighbourhood of infinity to a neighbourhood of zero (see [6] for example) and we can define the $R$-transform of $\mu$ by:

$$R_\mu(z) = G_\mu^{-1}(z) - \frac{1}{z},$$

where $G_\mu^{-1}(z)$ is the inverse function of $G_\mu$ on a proper neighbourhood of zero. This $R$-transform plays the role of the logarithm of the Fourier transform for the free convolution in the sense that for any probability measures $\mu_1$ and $\mu_2$,

$$R_{\mu_1 \boxplus \mu_2} = R_{\mu_1} + R_{\mu_2}. \quad (2.4)$$

Using this formula for statistical deconvolution requires the computation of two inverse functions, and it is proposed in [2] to use subordination functions which also characterize the free convolution as in (2.4).

Let us recall the definition of subordination functions due to Voiculescu [36]. We first introduce $F_\mu(z) = 1/G_\mu(z)$. As $G_\mu$ does not vanish on $\mathbb{C}^+$, $F_\mu$ is well defined on $\mathbb{C}^+$. Then:

**Theorem-Def 2.5.** There exist unique subordination functions $\alpha_1$ and $\alpha_2$ from $\mathbb{C}^+$ onto $\mathbb{C}^+$ such that:

(i) for $z \in \mathbb{C}^+$, $\text{Im}(\alpha_1(z)) \geq \text{Im}(z)$ and $\text{Im}(\alpha_2(z)) \geq \text{Im}(z)$, and $\lim_{y \to +\infty} \alpha_1(iy)/(iy) = \lim_{y \to +\infty} \alpha_2(iy)/(iy) = 1$.

(ii) for $z \in \mathbb{C}^+$, $F_{\mu_1 \boxplus \mu_2}(z) = F_{\mu_1}(\alpha_1(z)) = F_{\mu_2}(\alpha_2(z))$ and $\alpha_1(z) + \alpha_2(z) = F_{\mu_1 \boxplus \mu_2}(z) + z$.

Using this result, Belinschi and Bercovici [4, Theorem 3.2] introduce a fixed-point construction of the subordination functions, which Arizmendi et al. [2] adapt for the deconvolution problem. We state their result in the special case of the deconvolution by a semi-circular distribution defined in (1.2). In this case, we have an explicit formula for its Cauchy transform $G_{\sigma_t}(z)$ and its reciprocal function $F_{\sigma_t}(z)$:

$$G_{\sigma_t}(z) = \frac{z - \sqrt{z^2 - 4t}}{2t}, \quad \text{and} \quad z - F_{\sigma_t}(z) = t \, G_{\sigma_t}(z). \quad (2.5)$$

Before stating the result, let us recall that, for any $\gamma > 0$,

$$\mathbb{C}_\gamma = \{z \in \mathbb{C}^+ | \text{Im}(z) > \gamma\}.$$ 

These domains will appear since $G_\mu$ is not invertible on the whole plane $\mathbb{C}$. 
For any Lemma 2.7. in the following lemma proved at the end of the section: we will be able to recover the Cauchy transform of the initial condition 2.3. Construction of the estimator of $p_0$ appearing in the definition (2.8) of $F$ had been $F$. Based on Theorem 2.6, we devise the estimation strategy Overview of the estimation strategy. One difference between Theorem 2.6 and Theorem-definition 2.5 lies in the fact that the subordination functions are expressed in terms of $F$, whereas in Theorem-definition 2.5 it would have been $F_{\mu_0}$ and $F_\sigma$. Here the restriction to the domain $\mathbb{C}_{2\sqrt{7}}$ comes from the fact that $\text{Im}(\tilde{h}_\mu(w) - z)$ appearing in the definition (2.8) of $L_z$ has to be positive. It is worth pointing out that the constant $2\sqrt{7}$ in $\mathbb{C}_{2\sqrt{7}}$ that we obtained is better than the one of Arizmendi et al. [2] in a completely general setting, and which in the present case would be $2\sqrt{\pi}$. This improvement has a key impact on the convergence rates through the constant $\gamma$ (see Corollary 4.3).

The proof of Theorem 2.6 will be sketched in Section 2.4 and then proved in detail in Appendix B. We now explain how the subordination functions allow us to construct the estimator of $p_0$.

2.3. Construction of the estimator of $p_0$

Overview of the estimation strategy Based on Theorem 2.6, we devise the estimation strategy of the paper. The theorem allows us to get the subordination function $w_{fp}$ as a fixed point of $L_z$. From there, we will be able to recover the Cauchy transform of the initial condition $\mu_0$ from $w_{fp}$, as stated in the following lemma proved at the end of the section:

Lemma 2.7. For any $z \in \mathbb{C}_{2\sqrt{7}}$,

$$G_{\mu_0}(z) = \frac{1}{t}(w_{fp}(z) - z) = G_{\mu_0}(w_{fp}(z)).$$

Consequently, $|w_{fp}(z) - z| \leq \sqrt{t}$. 

\[\text{imsart-bj ver. 2014/10/16 file: FokkerPlanck_Bernoulli_Revisionv6.tex date: January 24, 2021}\]
Moreover, denoting \( C \) the centered Cauchy distribution with parameter \( \gamma > 0 \), one can check that, for any probability measure \( \mu \) on \( \mathbb{R} \), the density \( f_{\mu+C} \) of the classical convolution of \( \mu \) by \( C \) is given, for \( x \in \mathbb{R} \), by \( f_{\mu+C}(x) = -\text{Im} G_\mu(x + i\gamma)/\pi \). Using the expression of \( G_{\mu_0} \) given by Lemma 2.7 with \( \gamma > 2\sqrt{t} \), we get that for any \( x \in \mathbb{R} \),

\[
f_{\mu_0+C}(x) = \frac{1}{\pi t} \left[ \gamma - \text{Im} w_{fp}(x + i\gamma) \right].
\] (2.12)

From this, we can recover the density \( p_0 \) of \( \mu_0 \) by a classical deconvolution of (2.12) by \( f_\gamma \). The subordination function \( w_{fp} \) in (2.12) is estimated using the second equality of Lemma 2.7. In parallel with our work, Tarrago [29] has used the formula (2.12) to perform spectral deconvolution in a more general setting (including the multiplicative free convolution), but neither the approximation of \( w_{fp} \) by its estimator \( \hat{w}_{fp}^n \) defined in Theorem-Definition 2.8 below nor the (classical) deconvolution of the Cauchy distribution are treated, which are key difficulties encountered in our paper. Tarrago uses a different approach based on concentration inequalities when we use fluctuations in the line of [18]. To get the rates announced in the introduction, we establish very precise estimates of the error terms (see Section 4).

**Proof of Lemma 2.7.** Now, from (2.7), (2.6) and (2.5), we write for \( z \in \mathbb{C}_{2\sqrt{t}} \),

\[
w_{fp}(z) = z + w_1(z) - F_{\mu_0}(z) = z + w_1(z) - F_{\sigma_t}(w_1(z)) = z + t G_{\sigma_t}(w_1(z)).
\]

So, we obtain

\[
G_{\sigma_t}(w_1(z)) = \frac{1}{t} \left( w_{fp}(z) - z \right).
\]

Using again (2.6), we obtain both equalities of (2.10). From there, using Theorem 2.6 (i), we have:

\[
|w_{fp}(z) - z| = t |G_{\sigma_t}(w_1(z))| = t |G_{\mu_t}(w_{fp}(z))| \leq \frac{t}{|\text{Im}(w_{fp}(z))|} \leq \sqrt{t},
\]

\[\square\]

**Estimator of \( p_0 \)** We do not observe directly the measure \( \mu_t \). The observation is the matrix \( X^n(t) \) at time \( t > 0 \) for a given \( n \) and therefore its empirical spectral measure as defined in (1.8). Then, for \( z \in \mathbb{C}^+ \), replacing in the procedure \( G_{\mu_t}(z) \) by its natural estimator:

\[
\hat{G}_{\mu_t}(z) := \int_{\mathbb{R}} \frac{d\mu_t^n(\lambda)}{z - \lambda} = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{z - \lambda_j^n(t)} = \frac{1}{n} \text{tr} \left( (zI_n - X_n(t))^{-1} \right).
\] (2.13)

will lead to the following:

**Theorem-Def 2.8.** There exists a unique fixed point to the following functional equation in \( w(z) \):

\[
\frac{1}{t} (w(z) - z) = \hat{G}_{\mu_t}(w(z)), \quad \text{for } z \in \mathbb{C}_{2\sqrt{t}}
\] (2.14)

This fixed-point is denoted by \( \hat{w}_{fp}^n(z) \). We have \( \text{Im}(\hat{w}_{fp}^n(z)) > \text{Im}(z)/2 \) and \( |\hat{w}_{fp}(z) - z| \leq \sqrt{t} \).
The theorem is proved at the end of this section. We shall prove in Section 3 that \( \hat{w}_{fp}^n(z) \) is a convergent estimator of \( w_{fp}(z) \) and establish a fluctuation result associated with this convergence, which is Proposition 1.1. Let us now explain how the estimator of \( p_0 \) can be obtained from \( \hat{w}_{fp}^n(z) \).

Recall that the Fourier transform of the Cauchy distribution \( C_\alpha \) with \( \alpha > 0 \) is \( f_\alpha^*(\xi) = e^{-\alpha|\xi|} \) for \( \xi \in \mathbb{R} \). Performing the deconvolution from (2.12), the Fourier transform of \( p_0 \) is the division of the Fourier transform of the right-hand side of (2.12) by \( f_\gamma^*(\xi) \) with \( \gamma > 2\sqrt{t} \). It is now classical to define our ultimate estimator for the density function \( p_0 \) from its Fourier transform:

**Definition 2.9.** Let us consider a bandwidth \( h > 0 \) and a regularizing kernel \( K \). We assume that the kernel \( K \) is such that its Fourier transform \( K^* \) is bounded by a positive constant \( C_K < +\infty \) and has a compact support, say \([-1, 1]\). We define the estimator \( \hat{p}_{0,h} \) of \( p_0 \) by its Fourier transform:

\[
\hat{p}_{0,h}(\xi) = e^{\gamma|\xi|} K_h^*(\xi), \quad \frac{1}{\pi t} \left[ \gamma - \text{Im} \hat{w}_{fp}^n(\cdot + i \gamma)^*(\xi) \right],
\]

where we have defined \( K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h}) \).

Note that the assumption on \( K \) ensures finiteness of the estimator. These assumptions are for instance satisfied for the sinc kernel, namely \( K(x) = \text{sinc}(x) = \sin(x)/(\pi x) \) with Fourier transform \( K^*(\xi) = 1_{[-1,1]}(\xi) \) so that \( C_K = 1 \). From now on, \( K \) will denote the sinc kernel.

2.4. Sketch of proof of Theorem 2.6 and Theorem-Definition 2.8

In [2], the authors prove a more general version of Theorem 2.6. Here, as one of the measure involved is the semicircular distribution \( \sigma_\gamma \), one can use the explicit expressions of \( G_{\sigma_\gamma} \) or \( F_{\sigma_\gamma} \) to improve the constants. The detailed proof of the two theorems are postponed to Appendix B.

We start with sketching the proof of Theorem 2.6. Let us define the function \( L_z \) as in Equation (2.8):

\[
L_z(w) := h_{\sigma_\gamma}(\bar{h}_{\mu_\gamma}(w) - z) + z.
\]

One can check that for any \( z \in \mathbb{C}_{2\sqrt{t}} \), \( L_z \) is well defined on \( \mathbb{C}_{\frac{1}{2}\text{Im}z} \) and that it satisfies the assumptions of the Denjoy-Wolff fixed-point theorem, namely that \( L_z(\mathbb{C}_{\frac{1}{2}\text{Im}(z)}) \subset \mathbb{C}_{\frac{1}{2}\text{Im}(z)} \) and \( L_z \) is not a conformal automorphism. Therefore, one can deduce that for any \( z \in \mathbb{C}_{2\sqrt{t}} \), \( L_z \) admits a unique fixed point in \( \mathbb{C}_{\frac{1}{2}\text{Im}(z)} \), denoted by \( w_{fp}(z) \) and point (iv) is proved. We then define \( w_1(z) := F_{\mu_\gamma}(w_{fp}(z)) + w_{fp}(z) - z \), check that \( F_{\sigma_\gamma}(w_1(z)) = F_{\mu_\gamma}(w_{fp}(z)) \) and deduce that \( w_{fp}(z) \) and \( w_1(z) \) satisfy (i), (ii) and (iii).

In the preceding subsection, we have seen how to deduce Lemma 2.7 from Theorem 2.6, getting for \( w_{fp}(z) \) the simple equation \( \frac{1}{2}(w_{fp}(z) - z) = G_{\mu_\gamma}(w_{fp}(z)) \). It is therefore natural to define an estimator for \( w_{fp}(z) \) by replacing \( G_{\mu_\gamma}(z) \) by its estimator \( \hat{G}_{\mu_\gamma}(z) \). Theorem-Definition 2.8 will be obtained along similar arguments as above, applying the Denjoy-Wolff fixed point theorem to \( \hat{L}_z(w) := i\hat{G}_{\mu_\gamma}(w) + z \).

3. Study of the subordination function

This section is devoted to the proof of Proposition 1.1. We show that \( \hat{w}_{fp}^n(z) \) converges uniformly to \( w_{fp}(z) \) on \( \mathbb{C}_\gamma \) with \( \gamma > 2\sqrt{t} \). Next, we establish that its fluctuations are of order \( 1/\sqrt{n} \).
3.1. Proof of (i) and (ii) of Proposition 1.1

We first state a useful lemma.

**Lemma 3.1.** For any probability measure \( \mu \) on \( \mathbb{R} \) and \( \alpha > 0 \), the Cauchy transform \( G_{\mu} \) is Lipschitz on \( \mathbb{C}_\alpha \) with Lipschitz constant \( \frac{1}{\alpha^2} \), and one has for any \( z \in \mathbb{C}_\alpha \), \( |G_{\mu}(z)| \leq \frac{1}{\alpha} \).

**Proof.** For \( z, z' \in \mathbb{C}_\alpha \),

\[
|G_{\mu}(z) - G_{\mu}(z')| = \left| \int_{\mathbb{R}} \frac{d\mu(x)}{z-x} - \int_{\mathbb{R}} \frac{d\mu(y)}{z'-y} \right| \leq |z - z'| \left| \int_{\mathbb{R}} \frac{d\mu(x)}{|(z-x)(z'-x)|} \right| \leq \frac{|z - z'|}{\alpha^2}.
\]

We also have

\[
|G_{\mu}(z)| = \left| \int_{\mathbb{R}} \frac{d\mu(x)}{z-x} \right| \leq \frac{1}{\text{Im}(z)} \leq \frac{1}{\alpha}.
\]

We are now ready to prove the points (i) and (ii) of Proposition 1.1.

**Proof of Proposition 1.1 (i-ii).** Consider \( z \in \mathbb{C}_\gamma \) with \( \gamma > 2\sqrt{t} \). Using the equations (2.10) and (2.14) characterizing \( w_{fp}(z) \) and \( \hat{w}_{fp}^n(z) \), we have

\[
|\hat{w}_{fp}^n(z) - w_{fp}(z)| = t \left| \hat{G}_{\mu^n}(\hat{w}_{fp}^n(z)) - G_{\mu^n}(w_{fp}(z)) \right| \leq t \left| \hat{G}_{\mu^n}(\hat{w}_{fp}^n(z)) - \hat{G}_{\mu^n}(w_{fp}(z)) \right| + t \left| \hat{G}_{\mu^n}(w_{fp}(z)) - G_{\mu^n}(w_{fp}(z)) \right|.
\]

(3.1)

By Theorem 2.6, \( \text{Im}(w_{fp}(z)) \geq \frac{1}{2} \text{Im}(z) \) and since \( \hat{G}_{\mu^n} \) is a Lipschitz function on \( \mathbb{C}_{\frac{1}{2}\text{Im}(z)} \) with Lipschitz constant \( \frac{4}{\text{Im}^2(z)} \leq \frac{4}{\gamma^2} \), by Lemma 3.1, we have an upper bound for the first term

\[
\left| \hat{G}_{\mu^n}(\hat{w}_{fp}^n(z)) - \hat{G}_{\mu^n}(w_{fp}(z)) \right| \leq \frac{4}{\gamma^2} \times |\hat{w}_{fp}^n(z) - w_{fp}(z)|.
\]

Thus,

\[
|\hat{w}_{fp}^n(z) - w_{fp}(z)| \leq \frac{4t}{\gamma^2} |\hat{w}_{fp}^n(z) - w_{fp}(z)| + t \left| \hat{G}_{\mu^n}(w_{fp}(z)) - G_{\mu^n}(w_{fp}(z)) \right|,
\]

implying that

\[
|\hat{w}_{fp}^n(z) - w_{fp}(z)| \leq \left( \frac{t\gamma^2}{\gamma^2 - 4t} \right) \times \left| \hat{G}_{\mu^n}(w_{fp}(z)) - G_{\mu^n}(w_{fp}(z)) \right|.
\]

(3.2)

By Proposition 2.3, since the function \( x \mapsto \frac{1}{z-x} \) is continuous and bounded on \( \mathbb{R} \) for any \( z \in \mathbb{C}_{\sqrt{t}} \), \( \hat{G}_{\mu^n}(w_{fp}(z)) = \int_{\mathbb{R}} \frac{1}{w_{fp}(z)-x} \mu^n(dx) \) converges almost surely to \( G_{\mu^n}(w_{fp}(z)) = \int_{\mathbb{R}} \frac{1}{w_{fp}(z)-x} \mu_t(dx) \). This
concludes the proof of (i). To prove the uniform convergence (ii), we will need Vitali’s convergence theorem, see e.g. [3, Lemma 2.14, p.37-38]: on any bounded compact set of $\mathbb{C}_{2\sqrt{t}}$, the simple convergence is in fact a uniform convergence. Moreover, the functions $G_{\mu_t}(z)$ and $\hat{G}_{\mu_t}(z)$ decay as $1/|z|$ when $|z| \to +\infty$, implying the uniform convergence of the right-hand side of (3.2) on $\mathbb{C}_{\gamma}$, for $\gamma > 2\sqrt{t}$ and of $\hat{w}_{fp}^n(z)$ to $w_{fp}(z)$.

3.2. Fluctuations of the Cauchy transform of the empirical measure

We now prove point (iii) of Proposition 1.1. For this purpose, we first decompose:

$$\hat{G}_{\mu_t}(z) - G_{\mu_t}(z) = \hat{G}_{\mu_t}(z) - E[\hat{G}_{\mu_t}(z)|X^n(0)] + E[\hat{G}_{\mu_t}(z)|X^n(0)] - G_{\mu_0 \boxplus \sigma_t}(z) + G_{\mu_0 \boxplus \sigma_t}(z) - G_{\mu_t}(z) =: A_1^n(z) + A_2^n(z) + A_3^n(z).$$

(3.3)

The first term is related to the variance of $\hat{G}_{\mu_t}(z)$ (conditional on $X^n(0)$). The second term heuristically compares the evolution with the Hermitian Brownian motion to its limit. The third term deals with the fluctuations of the empirical initial condition. A similar decomposition for the first two terms is done in [18] (for a non-random initial condition) and we will adapt their results. In Propositions 3.2 and 3.3 below, we show that the fluctuations of the first two terms are of order $1/n$. The third term, which is associated to a classical central limit theorem, is of order $1/\sqrt{n}$, as proved in Proposition 3.6.

For the term $A_1^n(z)$, the result is a direct consequence of Proposition 3 in [18] and we refer to the detailed computation in [23].

**Proposition 3.2.** For $z \in \mathbb{C}^+$ and $n \in \mathbb{N}$,

$$\text{Var}(nA_1^n(z)|X^n(0)) = \text{Var}(n\hat{G}_{\mu_t}(z)|X^n(0)) \leq \frac{10t}{\text{Im}^3(z)}.$$  

3.2.1. Fluctuations of $A_2^n(z)$

We start with some additional notations. Let us denote the resolvent of $X^n(t)$ by

$$R_{n,t}(z):=(zI_n - X^n(t))^{-1}.$$  

(3.4)

Then one can write

$$\hat{G}_{\mu_t}(z) = \frac{1}{n} \text{Tr}(R_{n,t}(z)).$$

Then, the bias term is:

$$nA_2^n(z) = E[\text{Tr}(R_{n,t}(z)) | X^n(0)] - nG_{\mu_0 \boxplus \sigma_t}(z),$$

and it is given by an adaptation of [18, Proposition 4] to the case of a random initial condition:

**Proposition 3.3.** For $z \in \mathbb{C}^+$ and $n \in \mathbb{N}$,

$$|nA_2^n(z)| \leq \left(1 + \frac{4t}{\text{Im}^2(z)}\right) \left(\frac{2t}{\text{Im}^3(z)} + \frac{12t^2}{\text{Im}^5(z)}\right).$$  

(3.6)
The term $A_n^2(z)$ compares $E[\hat{G}_{\mu^0}(z)|X^n(0)]$ with $G_{\mu^0*\sigma_t}(z)$. Proceeding as in Theorem-Definition 2.5, with $\mu^0_t$ and $\sigma_t$, we can define a subordination function $\overline{w}_{fp}(z)$ such that

$$G_{\mu^0*\sigma_t}(z) = G_{\mu^0_t}(\overline{w}_{fp}(z)).$$

(3.7)

**Proof.** Note that by definition of the resolvent, we have for all $z \in \mathbb{C}^+$,

$$|nA_n^2(z)| \leq 2n\text{Im}^{-1}(z),$$

(3.8)

which is suboptimal due to the factor $n$.

We follow the ideas of ‘approximate subordination relations’ of [18]. As our initial condition is random, the strategy has to be adapted and we introduce the following variants of $R_{n,t}(z)$ and $A_n^2(z)$:

$$\tilde{R}_{n,t}(z) := \left( (z - \frac{t}{n} E[\text{Tr} (R_{n,t}(z)) | X^n(0))].I_n - X^n(0) \right)^{-1}$$

(3.9)

$$n\tilde{A}_n^2(z) := E[\text{Tr} (R_{n,t}(z)) | X^n(0)] - \text{Tr}(\tilde{R}_{n,t}(z)).$$

We will bound $A_n^2(z)$ by using its approximation $\tilde{A}_n^2(z)$.

**Step 1:** First, we prove an upper bound for $\tilde{A}_n^2(z)$, whose proof is postponed to Appendix C:

**Lemma 3.4.** For $z \in \mathbb{C}^+$,

$$|n\tilde{A}_n^2(z)| \leq \frac{2t}{\text{Im}^2(z)} + \frac{12t^2}{\text{Im}(z)}.$$

**Step 2:** If $|\tilde{A}_n^2(z)| \geq \text{Im}(z)/(2t)$ then, by (3.8)

$$|nA_n^2(z)| \leq \frac{4tn|\tilde{A}_n^2(z)|}{\text{Im}^2(z)},$$

and we conclude with Lemma 3.4.

**Step 3:** We now consider the case where $|\tilde{A}_n^2(z)| < \text{Im}(z)/(2t)$. We have:

$$A_n^2(z) = \tilde{A}_n^2(z) + [A_n^2(z) - \tilde{A}_n^2(z)]$$

(3.10)

We will control the difference $|A_n^2(z) - \tilde{A}_n^2(z)|$ by $\tilde{A}_n^2(z)$ and conclude with Lemma 3.4.

By their definitions:

$$n(A_n^2(z) - \tilde{A}_n^2(z)) = \text{Tr}(\tilde{R}_{n,t}(z)) - nG_{\mu^0*\sigma_t}(z).$$

(3.11)
We follow the trick in [18] which consists in going back to the fluctuations of the subordination functions. In view of (3.7), it is natural to express the first term \( \text{Tr}(\bar{R}_{n,t}(z)) \) of (3.11) similarly. As \( \bar{R}_{n,t}(z) \) is a diagonal matrix,

\[
\text{Tr}(\bar{R}_{n,t}(z)) = \sum_{j=1}^{n} \left( z - \frac{t}{n} \mathbb{E} \left[ \text{Tr}(R_{n,t}(z)) \mid X^n(0) \right] \right) - \lambda_j^n(0) = nG_{\mu_0} (\tilde{w}_f(z)),
\]

where

\[
\tilde{w}_f(z) := z - \frac{t}{n} \mathbb{E} \left[ \text{Tr}(R_{n,t}(z)) \mid X^n(0) \right]
\]

and where \( \lambda_j^n(0) \) are the eigenvalues of \( X^n(0) \). Thus:

\[
A_2^n(z) - \bar{A}_2^n(z) = G_{\mu_0} (\tilde{w}_f(z)) - G_{\mu_0} (\bar{w}_f(z)).
\]

To continue, we first need the following result proved in Appendix D.

**Lemma 3.5.** (i) The function \( \bar{w}_f(z) \), defined in (3.7), solves

\[
\bar{w}_f(z) = z - tG_{\mu_0} \mathbb{W}_{\sigma_t}(z).
\]

(ii) The function \( \zeta(z) = z + tG_{\mu_0} \) is well-defined on \( \mathbb{C}^+ \) and is the inverse of \( \bar{w}_f(z) \) on \( \overline{\mathbb{U}} = \{ z \in \mathbb{C}^+ \mid \text{Im}(\zeta(z)) > 0 \} \). For such \( z \in \overline{\mathbb{U}} \), we denote this function \( \bar{w}_f^{z \leftarrow 1} \).

Let us prove that under the condition of Step 3, \( \bar{w}_f(z) \in \overline{\mathbb{U}} \) for all \( z \in \mathbb{C}^+ \).

\[
\zeta(\bar{w}_f(z)) - z = \bar{w}_f(z) + tG_{\mu_0} (\bar{w}_f(z)) - z = z - \frac{t}{n} \mathbb{E} \left[ \text{Tr}(R_{n,t}(z)) \mid X^n(0) \right] + tG_{\mu_0} (\bar{w}_f(z)) - z = -t\bar{A}_2^n(z),
\]

by (3.12). Therefore,

\[
|\text{Im}(\zeta(\bar{w}_f(z))) - \text{Im}(z)| \leq |\zeta(\bar{w}_f(z)) - z| = t|\bar{A}_2^n(z)| \leq \frac{|\text{Im}(z)|}{2}.
\]

Thus, under the condition of Step 3, \( \bar{w}_f(z) \in \overline{\mathbb{U}} \). Denoting \( \bar{z} = \bar{w}_f^{z \leftarrow 1} \), which is well-defined, we have \( \bar{w}_f(z) = \bar{w}_f(\bar{z}) \). Plugging this into (3.14),

\[
A_2^n(z) - \bar{A}_2^n(z) = G_{\mu_0} \mathbb{W}_{\sigma_t}(\bar{z}) - G_{\mu_0} \mathbb{W}_{\sigma_t}(z) = (\bar{z} - z) \int_{\mathbb{R}} \frac{\mu_0^n \boxplus \sigma_t(dx)}{(\bar{z} - x)(z - x)} = t\bar{A}_2^n(z),
\]

where we used (3.15) for the last equality. From there, using (3.10), we get

\[
|A_2^n(z)| \leq \left| 1 + t \int_{\mathbb{R}} \frac{\mu_0^n \boxplus \sigma_t(dx)}{(\bar{z} - x)(z - x)} \right| |\bar{A}_2^n(z)| \leq \left( 1 + \frac{2t}{|\text{Im}^2(z)|} \right) |\bar{A}_2^n(z)|.
\]

This concludes the proof of Proposition 3.3.
3.2.2. Fluctuations of $A^\gamma_3(z)$

Finally, the third step is to control $A^\gamma_3(z) = G_{\mu_0 \boxplus \sigma_t}(z) - G_{\mu_t}(z)$, with $\mu_t = \mu_0 \boxplus \sigma_t$.

**Proposition 3.6.** For any $\gamma > 2\sqrt{t}$ and for any $z$ such that $\text{Im}(z) \geq \frac{\gamma}{t}$, we have:

$$|A^\gamma_3(z)| \leq \frac{\gamma^2}{\gamma^2 - 4t} \left| \int_{\mathbb{R}} \frac{1}{z - tG_{\mu_0 \boxplus \sigma_t}(z) - x} \left[ d\mu^\gamma_t(x) - d\mu_0(x) \right] \right|$$  \hspace{1cm} (3.17)

and

$$\sup_{n \in \mathbb{N}} \sup_{z \in \mathbb{C}} \mathbb{E} \left[ |A^\gamma_3(z)|^2 \right] \leq \frac{8\gamma^2}{(\gamma^2 - 4t)^2}. \hspace{1cm} (3.18)$$

**Proof.** Using again the subordination function $\overline{w}_{fp}(z)$ defined in (3.7) and Lemma 3.5(i), we have

$$G_{\mu_0 \boxplus \sigma_t}(z) = G_{\mu_0} \left( \overline{w}_{fp}(z) \right) = \int_{\mathbb{R}} \frac{d\mu^\gamma_t(x)}{\overline{w}_{fp}(z) - x} = \int_{\mathbb{R}} \frac{d\mu_0(x)}{z - tG_{\mu_0 \boxplus \sigma_t}(z) - x}. \hspace{1cm} (3.19)$$

In this proof, $\text{Im}(z) \geq \gamma/2 \geq \sqrt{t}$. Note that $\text{Im}(\overline{w}_{fp}(z)) \geq \text{Im}(z)$ (Theorem-Definition 2.5) so that

$$|z - tG_{\mu_0 \boxplus \sigma_t}(z) - x| \geq \frac{\gamma}{2} \geq \sqrt{t}, \hspace{1cm} (3.20)$$

and the integrand in (3.19) is well-defined and upper-bounded by $1/\sqrt{t}$. Similarly, we can establish that

$$G_{\mu_0 \boxplus \sigma_t}(z) = \int_{\mathbb{R}} \frac{d\mu_0(x)}{z - tG_{\mu_0 \boxplus \sigma_t}(z) - x}. \hspace{1cm} (3.21)$$

Then, we can write

$$G_{\mu_0 \boxplus \sigma_t}(z) - G_{\mu_0 \boxplus \sigma_t}(z) = t. \int_{\mathbb{R}} \frac{G_{\mu_0 \boxplus \sigma_t}(z) - G_{\mu_0 \boxplus \sigma_t}(z)}{(z - tG_{\mu_0 \boxplus \sigma_t}(z) - x) \cdot (z - tG_{\mu_0 \boxplus \sigma_t}(z) - x)} d\mu^\gamma_t(x)$$

$$+ \int_{\mathbb{R}} \frac{1}{z - tG_{\mu_0 \boxplus \sigma_t}(z) - x} \left[ d\mu^\gamma_t(x) - d\mu_0(x) \right].$$

Thus,

$$\left( G_{\mu_0 \boxplus \sigma_t}(z) - G_{\mu_0 \boxplus \sigma_t}(z) \right) \cdot \left[ 1 - t. \int_{\mathbb{R}} \frac{1}{(z - tG_{\mu_0 \boxplus \sigma_t}(z) - x) \cdot (z - tG_{\mu_0 \boxplus \sigma_t}(z) - x)} d\mu^\gamma_t(x) \right]$$

$$= \int_{\mathbb{R}} \frac{1}{z - tG_{\mu_0 \boxplus \sigma_t}(z) - x} \left[ d\mu^\gamma_t(x) - d\mu_0(x) \right].$$

Similarly to (3.20), we can show that $|z - tG_{\mu_0 \boxplus \sigma_t}(z) - x| \geq \gamma/2$. Thus

$$\left| t. \int_{\mathbb{R}} \frac{1}{(z - tG_{\mu_0 \boxplus \sigma_t}(z) - x) \cdot (z - tG_{\mu_0 \boxplus \sigma_t}(z) - x)} d\mu^\gamma_t(x) \right| \leq \frac{4t}{\gamma^2}.$$

Statistical deconvolution of the free Fokker-Planck equation at fixed time
consequently,
\[
|A_3^n(z)| \leq \frac{\gamma^2}{\gamma^2 - 4t} \left| \int_{\mathbb{R}} \frac{1}{z - tG_{\mu,\beta}(z)} \left[ \mu^n_0(x) - \mu_0(x) \right] dx \right|
\]
which gives the first part of the proposition. For the second part (3.18),
\[
E \left[ n |A_3^n(z)|^2 \right] = \left( \frac{\gamma^2}{\gamma^2 - 4t} \right)^2 nE \left[ \left| \int_{\mathbb{R}} \frac{1}{z - tG_{\mu,\beta}(z)} \left[ \mu^n_0(x) - \mu_0(x) \right] dx \right|^2 \right].
\]
If \( \text{Im}(z) > \frac{\gamma}{2} \), the function \( \varphi(z) = (z - tG_{\mu,\beta}(z) - x)^{-1} \) is bounded by \( 2/\gamma \). Then, for any \( z \in \mathbb{C}_\gamma \):
\[
nE \left[ \left| \int_{\mathbb{R}} \varphi(z) dx - \int_{\mathbb{R}} \varphi(z) dx \right|^2 \right] = nE \left[ \left| \frac{1}{n} \sum_{j=1}^{n} \varphi(z) (\lambda_j^n(0)) - E [\varphi(z) (\lambda_j^n(0))] \right|^2 \right]
\]
\[
= n\text{Var} \left[ \frac{1}{n} \sum_{j=1}^{n} \varphi(z) (\lambda_j^n) \right] = \int_{\mathbb{R}} \left| \varphi(z) dx \right|^2 dx - \left| \int_{\mathbb{R}} \varphi(z) dx \right|^2 \leq \frac{8}{\gamma^2},
\]
\(\Box\)

**Conclusion:** We can now conclude the proof of Proposition 1.1 (iii). From (3.3), Propositions 3.2, 3.3 and the first part of Proposition 3.6, we obtain that for \( z \in \mathbb{C}_{\gamma}/2 \):
\[
E \left[ |\hat{G}_{\mu}(z) - G_{\mu}(z)|^2 \mid X^n(0) \right] \leq C(\gamma,t) \left( \frac{1}{n^2} + \left| \int_{\mathbb{R}} \frac{1}{z - tG_{\mu,\beta}(z)} \left[ \mu^n_0(x) - \mu_0(x) \right] dx \right|^2 \right),
\]
where \( C(\gamma,t) \) depends only on \( \gamma \) and \( t \). The proof also shows that \( \gamma \mapsto C(\gamma,t) \) is bounded when \( \gamma \mapsto +\infty \) and \( \gamma \mapsto (\gamma^2 - 4t)^2 \times C(\gamma,t) \) is bounded when \( \gamma \mapsto 2\sqrt{t} \). Therefore there exists a constant \( C(t) \) only depending on \( t \) such that
\[
\sup_{n \in \mathbb{N}} \sup_{z \in \mathbb{C}_{\gamma}/2} nE \left[ |\hat{G}_{\mu}(z) - G_{\mu}(z)|^2 \right] = \sup_{n \in \mathbb{N}} \sup_{z \in \mathbb{C}_{\gamma}/2} nE \left[ E \left[ |\hat{G}_{\mu}(z) - G_{\mu}(z)|^2 \mid X^n(0) \right] \right] \leq \frac{C(t)\gamma^4}{(\gamma^2 - 4t)^2},
\]
by using the second part of Proposition 3.6. Equation (3.2) implies that for any \( \gamma > 2\sqrt{t} \),
\[
\sup_{n \in \mathbb{N}} \sup_{z \in \mathbb{C}_\gamma} E \left[ n |\hat{G}_{w}(z) - w(z)|^2 \right] \leq \left( \frac{t\gamma^2}{\gamma^2 - 4t} \right)^2 \sup_{n \in \mathbb{N}} \sup_{z \in \mathbb{C}_\gamma} nE \left[ |\hat{G}_{\mu}(z) - G_{\mu}(z)|^2 \right] < +\infty,
\]
since \( z \in \mathbb{C}_\gamma \) implies that \( \text{Im}(w(z)) \geq \frac{1}{2} \text{Im}(z) > \frac{\gamma}{2} \) (Theorem 2.6) so that \( w(z) \in \mathbb{C}_\gamma \) and point (iii) of Proposition 1.1 is proved.

4. Study of the mean integrated squared error

In Section 4.1, we state theoretical results associated with our nonparametric statistical problem. Section 4.2 is devoted to the proof of Theorem 4.1.
4.1. Theoretical results

The goal of this section is to study the rates of convergence of $\mathbb{E}[\|\hat{p}_{0,h} - p_0\|^2]$, the mean integrated squared error of $\hat{p}_{0,h}$, relying on the classical bias-variance decomposition of the quadratic risk. By Parseval’s equality, we obtain:

$$\|\hat{p}_{0,h} - p_0\|^2 = \frac{1}{2\pi} \|\hat{p}_{0,h} - K^* p_0^*\|^2 \leq \frac{1}{\pi} \|\hat{p}_{0,h} - K^* p_0^*\|^2 + \frac{1}{\pi} \|K^* p_0^* - p_0^*\|^2. \tag{4.1}$$

The expectation of the first term is a variance term whereas the second one is a bias term. While the control of the bias term is very classical, the study of the variance term in (4.1) is much more involved. The order of the variance term is provided by the following theorem.

**Theorem 4.1.** Let

$$\Sigma := \|\hat{p}_{0,h} - K^* p_0^*\|^2. \tag{4.2}$$

We assume that there exists a constant $C > 0$ such that for sufficiently large $\kappa > 0$,

$$\mu_0((\kappa, +\infty)) \leq \frac{C}{\kappa}. \tag{4.3}$$

Then, for any $\gamma > 2\sqrt{t}$, there exists a constant $C_{\text{var}}(t)$ only depending on $t$ such that for any $h > 0$ and $n$ large enough,

$$\mathbb{E}(\Sigma) \leq \frac{\gamma^8}{(\gamma^2 - 4t)^4} \frac{C_{\text{var}}(t)e^{\frac{2\pi}{h}}}{n}. \tag{4.4}$$

Theorem 4.1 is proved in Section 4.2. The main point consists in obtaining the optimal $n$ factor appearing at the denominator. The term $e^{\frac{2\pi}{h}}$ appearing at the numerator is classical in our setting and comes from the classical deconvolution by a Cauchy distribution (see (2.12) and (2.15)). The smaller $\gamma$ the better the rate of convergence but if $\gamma \rightarrow 2\sqrt{t}$ the leading constant of the upper bound blows up. Note that Assumption (4.3) is very mild and is satisfied by most classical distributions.

To derive the order of the bias term, we shall consider two classes of densities, supersmooth densities and densities belonging to Sobolev classes. First assume that $p_0$ belongs to the space $S_s(a,r,L)$ of supersmooth densities defined for $a > 0$, $L > 0$ and $r > 0$ by:

$$S_s(a,r,L) = \left\{ p \text{ density such that } \int_{\mathbb{R}} |p^*(\xi)|^2 e^{2ah|\xi|^r} d\xi \leq L \right\}. \tag{4.5}$$

In the literature, this smoothness class of densities has often been considered (see [21], [12], [14]). Most famous examples of supersmooth densities are the Cauchy distribution belonging to $S_s(a,r,L)$ with $r = 1$ and the Gaussian distribution belonging to $S_s(a,r,L)$ with $r = 2$. To control the bias, we rely on Proposition 1 in [12] which states that:

**Proposition 4.2.** For $p_0 \in S_s(a,r,L)$, we have:

$$\|K^* p_0^* - p_0^*\|^2 \leq Le^{-2ah^{-r}}.$$
Now, using similar computations to those in [21], we obtain from Proposition 4.2 and Theorem 4.1 the rates of convergence of our estimator \( \hat{p}_{0,h} \). We indeed showed that:

\[
MISE := E\left[ \| \hat{p}_{0,h} - p_0 \|^2 \right] \leq L e^{-2ah^{-r}} + \frac{\gamma^8}{(\gamma^2 - 4t)^4} \frac{C_{\vartheta}(t) e^{\frac{2\gamma}{n}}}{n}.
\] (4.6)

Minimizing in \( h \) the right hand side of (4.6) provides the convergence rate of the estimator \( \hat{p}_{0,h} \). The rates of convergence are summed up in the following corollary, adapted from the computation of [21]. One can see that there are three cases to consider to derive rates of convergence: \( r = 1 \), \( r < 1 \) and \( r > 1 \), depending on which the bias or variance term dominates the other. For the sake of completeness Corollary 4.3 is proved in Appendix E.

**Corollary 4.3.** Suppose that \( \mu_0 \) satisfies Assumption (4.3) and the density \( p_0 \) belongs to the space \( S_a(a,r,L) \) for \( a > 0, \ r > 0 \) and \( L > 0 \). Then, for any \( \gamma > 2\sqrt{t} \) and by choosing the bandwidth \( h \) according to Equation (E.1), we have:

\[
E\left[ \| \hat{p}_{0,h} - p_0 \|^2 \right] = \begin{cases} O\left( n^{-\frac{r}{r+1}} \right) & \text{if } r = 1 \\ O\left( \exp \left\{ -\frac{2a}{(2\gamma)^r} \left[ \log n + (r - 1) \log \log n + \sum_{i=0}^{k} b_i^* (\log n)^{a+i(r-1)} \right] \right\} \right) & \text{if } r < 1 \\ O\left( \frac{1}{n} \exp \left\{ \frac{2\gamma}{(2a)^{1/r}} \left[ \log n + \frac{r-1}{r} \log \log n + \sum_{i=0}^{k} d_i^* (\log n)^{a+i(r-1)} \right] \right\} \right) & \text{if } r > 1, \end{cases}
\]

where the integer \( k \) is such that

\[
\frac{k}{k+1} < \min\left(r, \frac{1}{r}\right) \leq \frac{k+1}{k+2}.
\]

and where the constants \( b_i^* \) and \( d_i^* \) solve respectively the following triangular systems:

\[
b_0^* = -\frac{2a}{(2\gamma)^r}, \quad \forall i > 0, \quad b_i^* = -\frac{2a}{(2\gamma)^r} \sum_{j=0}^{i-1} \frac{r(r-1)\cdots(r-j)}{(j+1)!} \sum_{p_0+\cdots+p_j=i-j-1} b_{p_0}^* \cdots b_{p_j}^*,
\] (4.8)

\[
d_0^* = -\frac{2\gamma}{(2a)^{1/r}}, \quad \forall i > 0, \quad d_i^* = -\frac{2\gamma}{(2a)^{1/r}} \sum_{j=0}^{i-1} \frac{\frac{r}{r-1}\cdots(\frac{1}{r}-j)}{(j+1)!} \sum_{p_0+\cdots+p_j=i-j-1} d_{p_0}^* \cdots d_{p_j}^*.
\]

**Remark 1.** For \( r = 1 \), the choice \( h = 2(a+\gamma)/\log(n) \) yields the rate of convergence. The optimal bandwidths for \( r > 1 \) and \( r < 1 \) are much more intricate (see (E.2) and (E.4), and also [21]).

Let us comment the rates of convergence obtained in Corollary 4.3. Recall that we have transformed the free deconvolution of the Fokker-Planck equation associated with observation of the matrix \( X^n(t) \) into the deconvolution problem expressed in (2.12). To solve the latter, we have then inverted the convolution operator characterized by the Fourier transform of the Cauchy distribution \( C_\gamma \). The parameter \( \gamma \) represents the difficulty of our deconvolution problem and consequently, the rates of convergence heavily depend on \( \gamma \). The larger \( \gamma \) the harder the problem, as can be observed in rates of convergences of Corollary 4.3. This is not surprising: as \( t \) grows, it becomes naturally harder to reconstruct the initial condition from the observations at time \( t \) and as \( \gamma \) has to be chosen larger than \( 2\sqrt{t} \), \( \gamma \) and therefore
the difficulty of the deconvolution problem grows with $t$ accordingly. It remains an open question if we can take $\gamma$ smaller.

Now, let us consider Sobolev type regularities. Assume that $p_0$ belongs to the Sobolev class $S_b(\beta, L)$ defined for $\beta > 0$ and $L > 0$ as:

$$S_b(\beta, L) = \left\{ p \text{ density such that } \int_R |p^*(\xi)|^2 (1 + \xi^2)\beta d\xi \leq L \right\}.$$

We have the following classical estimate for the integrated bias (see e.g. [14, Proposition 3]).

**Proposition 4.4.** For $p_0 \in S_b(\beta, L)$ we have:

$$\|K_{h}^*p_0^* - p_0^*\|^2 \leq Lh^{2\beta}.$$ (4.9)

Using Theorem 4.1, we obtain the following result.

**Corollary 4.5.** Suppose that $\mu_0$ satisfies Assumption (4.3) and the density $p_0$ belongs to the space $S_b(\beta, L)$ for $\beta > 0$ and $L > 0$. Then, for any $\gamma > 2\sqrt{t}$ and by choosing the bandwidth $h = C \log^{-1}(n)$ with $C > 2\gamma$, we have:

$$E\left[\|\hat{p}_{0,h} - p_0\|^2\right] = O\left((\log n)^{-2\beta}\right).$$ (4.10)

Now, let us discuss the optimality of the convergence rates stated in Corollaries 4.3 and 4.5. To this end, it is relevant to connect them with the minimax rates obtained in the classical statistical density deconvolution problem by Butucea and Tsybakov in [12] for supersmooth densities or in Fan and Koo [17] for Sobolev regularities. Here, our estimation strategy converts the initial free deconvolution problem into the deconvolution problem (2.12) between $\mu_0$ and the Cauchy distribution $C_\gamma$. Thus, our observation scheme is more intricate and involved than the framework of classical density deconvolution tackled in [12] and [17]. If our observations had been distributed according to the density $f_{\mu_0 \ast C_\gamma}$ as in [12] and [17], for a given $\gamma$, the upper bound of the variance term given by Theorem 4.1 as well as the bounds for the bias given by Proposition 4.2 and Proposition 4.4 would have been optimal. Consequently, as part of our strategy, we expect that our rates of convergence cannot be improved for a given $\gamma$.

### 4.2. Proof of Theorem 4.1

Recall the definition of $\Sigma$ in (4.2). By the definition of $\hat{\rho}_{0,h}$:

$$\Sigma = \int_R \frac{1}{\pi^{2}t^2} e^{2\gamma|\xi|} |K_\lambda^*(\xi)|^2 \left[ \left( \text{Im}\left( \hat{w}_{fp}(\cdot + i\gamma) \right) \right)^* - \left( \text{Im}\left( w_{fp}(\cdot + i\gamma) \right) \right)^* \right](\xi)^2 \, d\xi.$$
Recall that by Lemma 2.7, we have \( \text{Im} (w_{fp}(z)) = t \cdot \text{Im}(G_{\mu_1} (w_{fp}(z)) + \text{Im}(z) \), and similarly by Theorem 2.8, \( \text{Im}(\hat{w}_{fp}(z)) = t \cdot \text{Im}(\hat{G}_{\mu_1} (\hat{w}_{fp}(z)) + \text{Im}(z) \) for \( z \in \mathbb{C}_{2\sqrt{T}} \). Since \( K^* (\xi) = K^*(h\xi) \), we have

\[
\Sigma = \int_{\mathbb{R}} e^{2\gamma |\xi|} |K_h^*(\xi)|^2 \cdot \frac{1}{\pi^2} \left| \left( \text{Im} \hat{G}_{\mu_1} (\hat{w}_{fp}(\cdot + i\gamma)) - \text{Im} G_{\mu_1} (w_{fp}(\cdot + i\gamma)) \right)^* (\xi) \right|^2 d\xi
\]

\[
\leq e^{\frac{2\gamma}{\pi}} \cdot \frac{C^2_{\text{eig}}}{\pi^2} \left\| \left( \text{Im} \hat{G}_{\mu_1} (\hat{w}_{fp}(\cdot + i\gamma)) - \text{Im} G_{\mu_1} (w_{fp}(\cdot + i\gamma)) \right)^* \right\|^2
\]

\[
= \frac{2C^2_{\text{eig}}}{\pi} e^{\frac{2\gamma}{\pi}} \left\| \left( \text{Im} \hat{G}_{\mu_1} (\hat{w}_{fp}(\cdot + i\gamma)) - \text{Im} G_{\mu_1} (w_{fp}(\cdot + i\gamma)) \right)^* \right\|^2,
\]

by Parseval’s equality. Taking the expectation, and introducing a constant \( \kappa > 0 \) chosen later (depending on \( n \)), we have

\[
\mathbb{E}(\Sigma) \leq \frac{2C^2_{\text{eig}}}{\pi} e^{\frac{\kappa}{\pi}} (I^\kappa + J^\kappa) \tag{4.11}
\]

where

\[
I^\kappa = \int_{\{x \in \mathbb{R} : |x| \leq \kappa\}} \mathbb{E} \left[ \left| \left( \text{Im} \hat{G}_{\mu_1} (\hat{w}_{fp}(x + i\gamma)) - \text{Im} G_{\mu_1} (w_{fp}(x + i\gamma)) \right)^* \right|^2 \right] dx \tag{4.12}
\]

\[
J^\kappa = \int_{\{x \in \mathbb{R} : |x| > \kappa\}} \mathbb{E} \left[ \left| \left( \text{Im} \hat{G}_{\mu_1} (\hat{w}_{fp}(x + i\gamma)) - \text{Im} G_{\mu_1} (w_{fp}(x + i\gamma)) \right)^* \right|^2 \right] dx. \tag{4.13}
\]

To obtain the announced rates of convergence for the MISE, we need to be very careful in establishing the upper bounds for \( I^\kappa \) and \( J^\kappa \). For this purpose, we recall Lemma 4.3.17 of [1], with a null initial condition, which will be useful in the sequel:

**Lemma 4.6.** Let \( (\eta^n_1(t), \ldots, \eta^n_n(t)) \) be the eigenvalues of \( H^n(t) \). With large probability, all the eigenvalues \( (\eta^n_j(t)) \) of \( H^n(t) \) belong to a ball of radius \( M > 0 \) independent of \( n \). Introduce

\[
A^n_M := \left\{ \forall 1 \leq j \leq n : |\eta^n_j(t)| \leq M \right\}. \tag{4.14}
\]

There exist \( C_{\text{eig}} > 0 \) and \( D_{\text{eig}} > 0 \) depending on \( t \) such that for any \( M > D_{\text{eig}} \) and any \( n \in \mathbb{N}^* \)

\[
\mathbb{P} \left( (A^n_M)^C \right) = \mathbb{P} \left( \{ \eta^n_n(t) > M \} \right) \leq e^{-nC_{\text{eig}} M}, \tag{4.15}
\]

with \( \eta^n_n(t) := \max_{i=1,\ldots,n} |\eta^n_i(t)| \).

Using this lemma, we can control the tail distribution of \( \mathbb{E} [\mu^n] \), which is essential to establish very precise estimates. We recall that \( \lambda^n_j(0) \leq \ldots \leq \lambda^n_n(0) \) are the eigenvalues of \( X^n(t) = X^n(0) + H^n(t) \) in increasing order. By Weyl’s interlacing inequalities, we have that, for \( 1 \leq j \leq n \),

\[
\lambda^n_j(0) - \eta^n_n(t) \leq \lambda^n_j(t) \leq \lambda^n_j(0) + \eta^n_n(t). \tag{4.16}
\]

Therefore, for \( 1 \leq j \leq n \),

\[
\mathbb{E} \left[ \mu^n \left( \left\{ |\lambda| > \frac{K}{2} \right\} \right) \right] \leq \mathbb{E} \left[ \mu^n_0 \left( \left\{ |\lambda| > \frac{K}{4} \right\} \right) \right] + \mathbb{P} \left( \{ \eta^n_n(t) > \frac{K}{4} \} \right) \leq \mathbb{E} \left[ \mu^n_0 \left( \left\{ |\lambda| > \frac{K}{4} \right\} \right) \right] + e^{-\frac{nC_{\text{eig}}}{4}}.
\]
Recall that after (1.6), we introduced the notation \( d_1, \ldots, d_n \) for the i.i.d. random variables with law \( \mu_0 \) and whose order statistic are the diagonal elements of \( X_n(0) \), \( \lambda_1^{(0)} < \ldots < \lambda_n^{(0)} \). We have
\[
E \left[ \mu_0^T \left( \{ |\lambda| > \frac{K}{4} \} \right) \right] = \frac{1}{n} \sum_{i=1}^{n} P \left( |d_i^{(0)}| > \frac{K}{4} \right) = \mu_0 \left( \{ |\lambda| > \frac{K}{4} \} \right),
\]
so that we finally get
\[
E \left[ \mu_0^T \left( \{ |\lambda| > \frac{K}{2} \} \right) \right] \leq \mu_0 \left( \{ |\lambda| > \frac{K}{4} \} \right) + e^{-n C_{eig} \gamma}. \tag{4.17}
\]

Now, we successively study \( I^\kappa \) and \( J^\kappa \).

4.2.1. Upper bound for \( I^\kappa \)

Lemma 4.7. Let us consider \( \gamma > 2\sqrt{t} \). There exist constants \( C_1^2, C_2^2 \) and \( C_3^2 \) only depending on \( M \) and \( t \) such that
\[
I^\kappa \leq \frac{\gamma^{8}}{(\gamma^2 - 4t)^4} \frac{C_1^2}{n^4} + \frac{\kappa C_2^2}{n^2} + C_3^2 e^{-n C_{eig}^2 \gamma}. \tag{4.18}
\]

Before proving Lemma 4.7, let us establish a result that will be useful in the sequel.

Lemma 4.8. Let us consider \( \gamma > 2\sqrt{t} \), \( p > 1 \) and \( M > 0 \). Then, we have
\[
J_{p,\gamma,M,t} := \int_{0}^{+\infty} \int_{\mathbb{R}} \frac{1}{\left( \{ |\lambda| - x - \sqrt{t} - M \} \vee \frac{\gamma}{2} \right)^p} d\mu_0(\lambda) dx \leq C(p,M,t), \tag{4.19}
\]
for \( C(p,M,t) \) a finite constant only depending on \( p, M \) and \( t \).

**Proof.** The supremum in the denominator equals to \( |\lambda| - x - \sqrt{t} - M \) when \( x < |\lambda| - \sqrt{t} - M - \gamma/2 \) (which is possible only if \( |\lambda| - \sqrt{t} - M - \gamma/2 \) is positive) or \( x > |\lambda| + \sqrt{t} + M + \gamma/2 \). Otherwise the supremum is \( \gamma/2 \). Hence
\[
J_{p,\gamma,M,t} \leq \int_{\mathbb{R}} \left\{ \int_{0}^{\left( |\lambda| - \sqrt{t} - M - \frac{\gamma}{2} \right) \vee 0} \frac{1}{(\lambda - x - \sqrt{t} - M)^p} dx + \int_{\left( |\lambda| - \sqrt{t} - M - \frac{\gamma}{2} \right) \vee 0}^{+\infty} \frac{2p}{\gamma^p} dx \right\} \frac{1}{\left( |\lambda| + \sqrt{t} + M + \frac{\gamma}{2} \right)^p} d\mu_0(\lambda)
\]
\[
\leq \int_{\mathbb{R}} \left( \int_{\left( |\lambda| - \sqrt{t} - M \right) \vee \frac{\gamma}{2}}^{+\infty} \frac{1}{\gamma^p} dv + \frac{2p}{\gamma^p} \frac{(2\sqrt{t} + 2M + \gamma)}{\gamma^p} \right) \frac{1}{\gamma^p} d\mu_0(\lambda) \leq C(p,M,t) < +\infty,
\]
since \( \gamma > 2\sqrt{t} \). This concludes the proof of Lemma 4.8. \( \square \)
Proof of Lemma 4.7. We decompose $I^\kappa$ into three parts, $I^\kappa \leq 3(I_1^\kappa + I_2^\kappa + I_3^\kappa)$ where:

$$
I_1^\kappa := \int_{\{|x| \leq \kappa\}} \mathbb{E} \left[ \left| \tilde{G}_\mu^\kappa \left( \tilde{w}_{fp}(x + i\gamma) \right) - \tilde{G}_\mu^\kappa \left( w_{fp}(x + i\gamma) \right) \right|^2 \right] dx,
$$

$$
I_2^\kappa := \int_{\{|x| \leq \kappa\}} \mathbb{E} \left[ \left| \tilde{G}_\mu^\kappa \left( w_{fp}(x + i\gamma) \right) - \mathbb{E} \left[ \tilde{G}_\mu^\kappa \left( w_{fp}(x + i\gamma) \right) | X^n(0) \right] \right|^2 \right] dx,
$$

$$
I_3^\kappa := \int_{\{|x| \leq \kappa\}} \mathbb{E} \left[ \left| \mathbb{E} \left[ \tilde{G}_\mu^\kappa \left( w_{fp}(x + i\gamma) \right) | X^n(0) \right] - G_\mu^\kappa \left( w_{fp}(x + i\gamma) \right) \right|^2 \right] dx.
$$

Step 1: Let us first upper bound $I_1^\kappa$. It is relatively easy to bound $I_1^\kappa$ by an upper bound in $C(\gamma, t)\kappa/n$, but this will not yield in the end the announced convergence rate. To establish more precise upper bounds, we use the event $A_{M}^{n,t}$ defined in Lemma 4.6. We have $I_1^\kappa = I_{11}^\kappa + I_{12}^\kappa$ with

$$
I_{11}^\kappa := \int_{\{|x| \leq \kappa\}} \mathbb{E} \left[ \left| \tilde{G}_\mu^\kappa \left( \tilde{w}_{fp}(x + i\gamma) \right) - \tilde{G}_\mu^\kappa \left( w_{fp}(x + i\gamma) \right) \right|^2 1_{A_{M}^{n,t}} \right] dx,
$$

$$
I_{12}^\kappa := \int_{\{|x| \leq \kappa\}} \mathbb{E} \left[ \left| \tilde{G}_\mu^\kappa \left( \tilde{w}_{fp}(x + i\gamma) \right) - \tilde{G}_\mu^\kappa \left( w_{fp}(x + i\gamma) \right) \right|^2 1_{(A_{M}^{n,t})^c} \right] dx.
$$

For the term $I_{12}^\kappa$, we have by Theorem 2.6(i) and Lemma 4.6:

$$
I_{12}^\kappa \leq \frac{16 n}{\kappa^2} \mathbb{P}(\{A_{M}^{n,t} \cap C\}) \leq \frac{16 n}{\kappa^2} Ke^{-nC\kappa t/M}. \tag{4.20}
$$

Let us now consider the term $I_{11}^\kappa$:

$$
I_{11}^\kappa = \int_{\{|x| \leq \kappa\}} \mathbb{E} \left[ \left| \frac{1}{n} \sum_{j=1}^{n} \left( \frac{w_{fp}(x + i\gamma) - \tilde{w}_{fp}(x + i\gamma)}{\tilde{w}_{fp}(x + i\gamma) - \lambda_j^\kappa(t)} \left( \tilde{w}_{fp}(x + i\gamma) - \lambda_j^\kappa(t) \right) \right)^2 1_{A_{M}^{n,t}} \right] dx
$$

$$
\leq \int_{\{|x| \leq \kappa\}} \mathbb{E} \left[ \left| \tilde{w}_{fp}(x + i\gamma) - w_{fp}(x + i\gamma) \right|^2 \frac{1}{n} \sum_{j=1}^{n} \left| \tilde{w}_{fp}(x + i\gamma) - \lambda_j^\kappa(t) \right|^2 \right] dx
$$

by convexity. Using (2.11) and (4.16), we have

$$
|w_{fp}(x + i\gamma) - \lambda_j^\kappa(t)| \geq |w_{fp}(x + i\gamma)| - |\lambda_j^\kappa(t)|
$$

$$
\geq \left| \text{Re}(w_{fp}(x + i\gamma)) \right| - |\lambda_j^\kappa(0)| - |\eta_j^\kappa(t)|.
$$

Since $\lambda_j^\kappa(t)$ is real, we also have:

$$
|w_{fp}(x + i\gamma) - \lambda_j^\kappa(t)| \geq \left| \text{Re}(w_{fp}(x + i\gamma)) - |\lambda_j^\kappa(t)| \right| \geq |\lambda_j^\kappa(t)| - |\text{Re}(w_{fp}(x + i\gamma))|
$$

$$
\geq |\lambda_j^\kappa(t)| - |x| - \sqrt{t} \geq |\lambda_j^\kappa(0)| - |\eta_j^\kappa(t)| - |x| - \sqrt{t}.
$$

Therefore, using Theorem 2.6,

$$
|w_{fp}(x + i\gamma) - \lambda_j^\kappa(t)| \geq \left| \lambda_j^\kappa(0) - |x| - \sqrt{t} - \eta_j^\kappa(t) \right| \vee \frac{\gamma}{2}. \tag{4.21}
$$
In Theorem-Definition 2.8, it is shown that $\hat{w}^n_{fp}(z)$ satisfies a similar inequality as (2.11). Thus, we obtain with similar computations that:

$$|\hat{w}^n_{fp}(x + i\gamma) - \lambda^n(t)| \geq \left\{ \frac{1}{|\lambda^n(0)| - |x|} \right\} \sqrt{\frac{\gamma}{2}}. \tag{4.22}$$

Then, using the definition of $A_{n,t}^{\kappa}$ and the constant $C(\gamma, t)$ appearing in (3.22), there exists a constant $C(t)$ only depending on $t$ such that

$$I_{11}^n \leq \int_{\{x \leq \kappa\}} E \left[ \frac{1}{n} \sum_{j=1}^{n} \left\{ \frac{1}{\left| \lambda^n(0) - |x| - \sqrt{t} + M \right|} \right\} \right] dx$$

$$\leq \frac{1}{n} \sum_{j=1}^{n} \int_{\{x \leq \kappa\}} E \left[ \left\{ \frac{1}{\left| \lambda^n(0) - |x| - \sqrt{t} + M \right|} \right\} \right] dx$$

$$\leq (\frac{t^2}{\gamma^2 - 4t})^2 \frac{C(\gamma, t)}{n} \left( I_{111}^n + I_{112}^n \right), \tag{4.23}$$

where the third inequality comes from (3.22) combined with (3.2), where the fourth inequality comes from the analysis of the constant $C(\gamma, t)$ in Section 3.2.2, and where:

$$I_{111}^n := \frac{1}{n^2} \sum_{j=1}^{n} \int_{\{x \leq \kappa\}} E \left[ \left\{ \frac{1}{\left| \lambda^n(0) - |x| - \sqrt{t} + M \right|} \right\} \right] dx$$

$$I_{112}^n := \int_{\{x \leq \kappa\}} E \left[ \int_{\{x \leq \kappa\}} \frac{1}{\left| \lambda^n(0) - |x| - \sqrt{t} + M \right|} d\lambda \right] dx$$

$$\left\{ \sqrt{n} \int_{\{x \leq \kappa\}} \frac{1}{\left| \lambda^n(0) - |x| - \sqrt{t} + M \right|} d\lambda \right\}^2 dx. \tag{4.24}$$

Now we wish to upper bound $I_{111}^n$ and $I_{112}^n$ independently of $\kappa$. We first deal with $I_{111}^n$.

$$I_{111}^n = \frac{1}{n} \int_{\{x \leq \kappa\}} E \left[ \int_{\{x \leq \kappa\}} \frac{1}{\left| \lambda^n(0) - |x| - \sqrt{t} + M \right|} d\lambda \right] dx$$

$$\leq \frac{n}{n} \int_{0}^{\infty} \int_{\{x \leq \kappa\}} \frac{1}{\left| \lambda^n(0) - |x| - \sqrt{t} + M \right|} d\lambda \leq \frac{2C(4, M, t)}{n}, \tag{4.24}$$
by Lemma 4.8. Let us now consider \( I_{112}^\kappa \). Using Cauchy-Schwarz inequality, we have:

\[
I_{112}^\kappa \leq \sqrt{\mathbb{E} \left[ \int_{\{ |x| \leq \kappa \}} \left( \int \frac{1}{\{ |\lambda| - |x| - \sqrt{t} - M \} \vee \frac{\gamma}{2}} \right)^2 d\mu_0^\kappa(\lambda) \right] ^2} dx
\]

\[
\leq \mathbb{E} \left[ \int_{\{ |x| \leq \kappa \}} \sqrt{\mathbb{E} \left[ \int \frac{1}{\{ |\lambda| - |x| - \sqrt{t} - M \} \vee \frac{\gamma}{2}} \right]^4 d\mu_0^\kappa(\lambda) \right] dx
\]

The first term can be treated exactly as \( I_{111}^\kappa \) as:

\[
E \left[ \int_{\{ |x| \leq \kappa \}} \left( \int \frac{1}{\{ |\lambda| - |x| - \sqrt{t} - M \} \vee \frac{\gamma}{2}} \right)^2 d\mu_0^\kappa(\lambda) \right] dx
\]

\[
\leq E \left[ \int_{\{ |x| \leq \kappa \}} \left( \int \frac{1}{\{ |\lambda| - |x| - \sqrt{t} - M \} \vee \frac{\gamma}{2}} \right)^4 d\mu_0^\kappa(\lambda) \right] dx = \frac{16n}{\gamma^4} I_{111}^\kappa. \quad (4.26)
\]

We now focus on the second term of (4.25). As in the proof of Proposition 3.6, if we denote by \( \phi_x := \varphi_{x+i\gamma} : \lambda \mapsto (x + i\gamma - tG_{\mu_0 \oplus \sigma_\varepsilon}(x + i\gamma) - \lambda)^{-1} \), the last term can be rewritten as

\[
I_{1121}^\kappa := E \left[ \int_{\{ |x| \leq \kappa \}} \sqrt{n} \int \phi_x(\lambda) \left[ d\mu_0^n(\lambda) - d\mu_0(\lambda) \right] \right] dx = n^2 \int_{\{ |x| \leq \kappa \}} E \left[ \int \frac{\lambda}{n} \sum_{j=1}^n (\phi_x(d_j^n) - E[\phi_x(d_j^n)]) \right] dx
\]

where \( d_1^n, \ldots, d_n^n \) are the non-ordered diagonal elements of \( X_n(0) \) (see after Equation (1.6)). Since the random variables \( d_1^n, \ldots, d_n^n \) are i.i.d. with law \( \mu_0 \), the random variables \( (\phi_x(d_j^n) - E[\phi_x(d_j^n)]) \) are i.i.d. centered with finite fourth moment. By Rosenthal and then Cauchy-Schwarz inequality, we have

\[
I_{1121}^\kappa \leq C n^2 \int_{\{ |x| \leq \kappa \}} E \left[ \left| \phi_x(d_j^n) - E[\phi_x(d_j^n)] \right| \right] dx
\]

for \( C \) a constant. We can conclude if the above double integral is bounded independently of \( \kappa \).

Let us recall now some estimates for the functions \( \phi_x \). As \( \text{Im}(G_{\mu_0 \oplus \sigma_\varepsilon}(x + i\gamma)) < 0 \), we have

\[
|x + i\gamma - tG_{\mu_0 \oplus \sigma_\varepsilon}(x + i\gamma) - \lambda| \geq \text{Im}(x + i\gamma - tG_{\mu_0 \oplus \sigma_\varepsilon}(x + i\gamma) - \lambda) \geq \gamma \geq \frac{\gamma}{2} \quad (4.28)
\]

and the functions \( \phi_x \) are bounded by \( 2/\gamma \). This yields that \( \left| \int \phi_x(\lambda) d\mu_0(\lambda) \right| \leq 2/\gamma \). By Lemma 3.1, \( |tG_{\mu_0 \oplus \sigma_\varepsilon}(x + i\gamma)| \leq \frac{\sqrt{t}}{2} \leq \sqrt{t} \) so that

\[
|x + i\gamma - tG_{\mu_0 \oplus \sigma_\varepsilon}(x + i\gamma) - \lambda| \geq (|x - \lambda| - \sqrt{t}) \geq (|x| - |\lambda|) - \sqrt{t}). \quad (4.29)
\]

As a consequence,

\[
|x + i\gamma - tG_{\mu_0 \oplus \sigma_\varepsilon}(x + i\gamma) - \lambda| \geq (|x| - |\lambda|) - \sqrt{t}) \vee \frac{\gamma}{2}. \quad (4.30)
\]
Using that $d^n_t$ has distribution $\mu_0$, the double integral in the right hand side of (4.27) becomes:

$$I_1 = \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \left| \phi_x(d^n_t) - \mathbb{E}[\phi_x(d^n_t)] \right|^4 \right] \, dx$$

where $\kappa$ is a constant only depending on $\gamma$, $\mu_{\gamma}$ and $\mathbb{E}[\phi_x(d^n_t)]$ are defined in Lemma 4.8 and $\mathbb{E}[\phi_x(d^n_t)]^2 + \mathbb{E}[\phi_x(d^n_t)]^2 \leq |\phi_x(d^n_t)|^4$.

We can now conclude Step 1. The last result, together with (4.27), implies that $I_{12}$ is bounded by a constant only depending on $t$. From (4.23) and (4.24), we have that $I_{11} < \frac{\gamma^8}{(\gamma^2-4t)^4} C_1(M,t) n + 16 \frac{\kappa e^{-n \gamma}}{\gamma^2}$. Gathering this result with (4.20), we finally obtain that:

$$I_1 \leq \frac{\gamma^8}{(\gamma^2-4t)^4} C_1(M,t) n + 16 \frac{\kappa e^{-n \gamma}}{\gamma^2}. \quad (4.31)$$

**Step 2:** Let us consider $I_2$. Using Proposition 3.2, we have:

$$I_2 = \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \text{Var} \left( \hat{\mu}_n \left( w_{fp}(x+i\gamma) \right) \right) \mid X^n(0) \right] \, dx = \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \text{Var} \left( A^n_t \left( w_{fp}(x+i\gamma) \right) \right) \mid X^n(0) \right] \, dx \leq \int_{\{x\leq \kappa\}} \frac{10 t}{n^2 \gamma^4} \, dx \leq \frac{10.2^4 t.2 \kappa}{n^2 \gamma^4} \leq \frac{20 \kappa}{n^2 t}. \quad (4.32)$$

**Step 3:** Let us now provide an upper bound for $I_3$. Recall the definitions of $A_n^2(z)$ and $A_n^3(z)$ in (3.3):

$$I_3 = \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \left| A_n^2 \left( w_{fp}(x+i\gamma) \right) + A_n^3 \left( w_{fp}(x+i\gamma) \right) \right|^2 \right] \, dx\leq 2 \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \left| A_n^2 \left( w_{fp}(x+i\gamma) \right) \right|^2 \right] \, dx + 2 \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \left| A_n^3 \left( w_{fp}(x+i\gamma) \right) \right|^2 \right] \, dx. \quad (4.33)$$

By using Proposition 3.3 together with Theorem 2.6 (i) and the fact that $\gamma > 2\sqrt{t}$, we obtain that the first term in the right hand side is upper-bounded by:

$$2 \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \left| A_n^2 \left( w_{fp}(x+i\gamma) \right) \right|^2 \right] \, dx \leq \frac{c \kappa}{n^2 t},$$

where $c$ is an absolute constant. Let us now consider the second term in the right hand side of (4.33).

**Step 4:** Let us consider $I_4$. Using Proposition 3.3 together with Theorem 2.6 (ii) and the fact that $\gamma > 2\sqrt{t}$, we obtain that the second term in the right hand side is upper-bounded by:

$$2 \int_{\{x\leq \kappa\}} \mathbb{E} \left[ \left| A_n^3 \left( w_{fp}(x+i\gamma) \right) \right|^2 \right] \, dx \leq \frac{c \kappa}{n^2 t},$$

where $c$ is an absolute constant. Let us now consider the third term in the right hand side of (4.33).
Using the bound of Proposition 3.6,

\[ \int_{\{|x| \leq \kappa\}} \mathbb{E}\left[ |A_3^n(w_{fp}(x + i\gamma))|^2 \right] \mathrm{d}x \leq 2 \frac{\gamma^4}{(\gamma^2 - 4t)^2} \int_{\mathbb{R}} \mathbb{E}\left[ \left| \int_{\mathbb{R}} w_{fp}(x + i\gamma) - t.G_{\mu_1}(w_{fp}(x + i\gamma)) - v \left( d\mu_0^n(v) - d\mu_0(v) \right) \right|^2 \right] \mathrm{d}x. \quad (4.34) \]

Recall that \( \mu_0^n \) is the empirical measure of independent random variables \( (d^n_i) \) with distribution \( \mu_0 \) and whose order statistics are the \( (\lambda^n_i(0)) \). Recalling that \( (w_{fp}(x + i\gamma) - t.G_{\mu_1}(w_{fp}(x + i\gamma)) - v)^{-1} = \varphi_{w_{fp}(x+i\gamma)}(v) \), we have that

\[ \mathbb{E}\left[ \left| \int_{\mathbb{R}} \varphi_{w_{fp}(x+i\gamma)}(v) [d\mu^n_0(v) - d\mu_0(v)] \right|^2 \right] = \text{Var}\left[ \frac{1}{n} \sum_{j=1}^n \varphi_{w_{fp}(x+i\gamma)}(\lambda^n_j(0)) \right] \]

\[ \leq \frac{1}{n} \mathbb{E}\left[ |\varphi_{w_{fp}(x+i\gamma)}(d^n_i)|^2 \right] = \frac{1}{n} \int_{\mathbb{R}} |w_{fp}(x + i\gamma) - t.G_{\mu_1}(w_{fp}(x + i\gamma)) - v|^2 d\mu_0(v). \quad (4.35) \]

We have:

\[ |w_{fp}(x + i\gamma) - t.G_{\mu_1}(w_{fp}(x + i\gamma)) - v| \geq |\text{Re}(w_{fp}(x + i\gamma) - t.G_{\mu_1}(w_{fp}(x + i\gamma))) - v| \]

\[ \geq |\text{Re}(w_{fp}(x + i\gamma)) - v| - t |\text{Re}(G_{\mu_1}(w_{fp}(x + i\gamma)))|. \]

By Theorem 2.6 (i), we have that:

\[ |\text{Re}(G_{\mu_1}(w_{fp}(x + i\gamma)))| \leq \left| \int_{\mathbb{R}} \frac{d\mu_1(y)}{w_{fp}(x + i\gamma) - y} \right| \leq \frac{1}{|\text{Im}(w_{fp}(x + i\gamma))|} \leq \frac{2}{\gamma}. \]

Also, by using (2.11), we get that \( |\text{Re}(w_{fp}(x + i\gamma)) - x| \leq \sqrt{t} \). Therefore,

\[ |w_{fp}(x + i\gamma) - t.G_{\mu_1}(w_{fp}(x + i\gamma)) - v| \geq |x| - |v| - \sqrt{t} - \frac{2t}{\gamma}. \quad (4.36) \]

From (4.34), (4.35) and (4.36), we have that:

\[ \int_{\{|x| \leq \kappa\}} \mathbb{E}\left[ |A_3^n(w_{fp}(x + i\gamma))|^2 \right] \mathrm{d}x \leq \frac{2\gamma^4}{n(\gamma^2 - 4t)^2} \int_{\mathbb{R}} \mathbb{E}\left[ \left| \int_{\mathbb{R}} \left\{ |x| - |v| - \sqrt{t} - \frac{2t}{\gamma} \right\} \vee \frac{\gamma}{2} \right|^2 \right] \mathrm{d}\mu_0(v) \mathrm{d}x \leq \frac{4\gamma^4}{n(\gamma^2 - 4t)^2} \int_{\mathbb{R}} \mathbb{E}\left[ \left| \int_{\mathbb{R}} \left\{ |x| - |v| - \sqrt{t} - \frac{2t}{\gamma} \right\} \vee \frac{\gamma}{2} \right|^2 \right] \mathrm{d}\mu_0(v) \mathrm{d}x. \]

by Lemma 4.8. We conclude as for \( I_{11}^n \) and we obtain

\[ I_5^1 \leq \frac{cK}{n^2t} + \frac{4\gamma^4}{n(\gamma^2 - 4t)^2} \int_{\mathbb{R}} \mathbb{E}\left[ \left| \int_{\mathbb{R}} \left\{ |x| - |v| - \sqrt{t} - \frac{2t}{\gamma} \right\} \vee \frac{\gamma}{2} \right|^2 \right] \mathrm{d}\mu_0(v) \mathrm{d}x. \quad (4.37) \]

Gathering (4.31), (4.32) and (4.37) we obtain the result announced in Lemma 4.7. \( \Box \)
4.2.2. Upper bound for $J^r$

Recall the definition of $J^r$ in (4.11). Our goal is to prove the following bound:

**Lemma 4.9.** There exist constants $C_1^\gamma$, $C_2^\gamma$ and $C_3^\gamma$ only depending on $t$ such that, for any $\kappa > \gamma$, we have:

$$J^r \leq \frac{C_1^\gamma}{\kappa} + C_2^\gamma e^{-\frac{n.Ceig.\kappa}{4}} + C_3^\gamma \mu_0 \left( \left\{ |\lambda| > \frac{\kappa}{4} \right\} \right).$$  \hspace{1cm} (4.38)

**Proof.** We decompose $J^r \leq 2(J^r_1 + J^r_2)$ where

$$J^r_1 := \int_{\{|x| > \kappa\}} \mathbb{E} \left( \left| \int_{\mathbb{R}} \overline{w}_{fp}(x + i\gamma) - \lambda \right|^2 \right) dx,$$

and

$$J^r_2 := \int_{\{|x| > \kappa\}} \left| \int_{\mathbb{R}} w_{fp}(x + i\gamma) - \lambda \right|^2 dx.$$

Let us consider the first term $J^r_1$. Using the estimate of Theorem-Definition 2.8, we have for all $x \in \mathbb{R}$ that $|\text{Re} \left( \overline{w}_{fp}(x + i\gamma) - x \right)| \leq \sqrt{t}$ and $\text{Im} \left( \overline{w}_{fp}(x + i\gamma) \right) \geq \gamma/2$. This allows us to prove that there exists a constant $C_4$ only depending on $t$ such that for all $\lambda \in \mathbb{R}$

$$(x - \lambda)^2 + \frac{\gamma^2}{4} \leq C_4 \left( \text{Re}^2 \left( \overline{w}_{fp}(x + i\gamma) - \lambda \right) + \frac{\gamma^2}{4} \right).$$

(4.39)

Thus,

$$J^r_1 \leq \frac{\mathcal{C}_t}{\gamma} \int_{\{|x| > \kappa\}} \mathbb{E} \left[ \int_{\mathbb{R}} \frac{d\mu_n^\gamma(\lambda)}{(x - \lambda)^2 + \frac{\gamma^2}{4}} \right].$$

We now use the simple bounds $|\text{arctan } x| \leq |x|$ and $|\text{arctan } x| \leq \frac{\pi}{2}$ for any $x \in \mathbb{R}$. Moreover, one can easily check that, if $\lambda^2 \leq \frac{\kappa^2}{2} - \frac{\gamma^2}{4}$, then

$$\frac{4\kappa\gamma}{4\kappa^2 - 4\lambda^2 - \gamma^2} \leq \frac{2\gamma}{\kappa}.$$

We therefore get

$$J^r_1 \leq \frac{2\mathcal{C}_t}{\gamma} \int_{\lambda} d\mu_n^\gamma(\lambda) \left( \frac{2\gamma}{\kappa} + \frac{\pi}{2} \left\{ \lambda^2 > \frac{\kappa^2}{2} - \frac{\gamma^2}{4} \right\} + \pi \left\{ \lambda^2 > \frac{\kappa^2}{2} - \frac{\gamma^2}{4} \right\} \right).$$

If we assume moreover that $\kappa > \gamma$, this can be simplified as follows:

$$J^r_1 \leq \mathcal{C}_t \left( \frac{4}{\kappa} + \frac{3\pi}{\gamma} \mathbb{E} \left[ \mu_t^\gamma \left( \left\{ |\lambda| > \frac{\kappa}{2} \right\} \right) \right] \right)$$

$$\leq \mathcal{C}_t \left( \frac{4}{\kappa} + \frac{3\pi}{\gamma} \mu_0 \left( \left\{ |\lambda| > \frac{\kappa}{4} \right\} \right) + \frac{3\pi}{\gamma} e^{-\frac{n.Ceig.\kappa}{4}} \right).$$

(4.40)
by using (4.17).

We now go to the second term $J_2^\kappa$. The strategy will be very similar to what we did for $J_1^\kappa$ and we will give less details. Using the estimate (2.11), we have for all $x \in \mathbb{R}$ that $|\text{Re}(w_{fp}(x + i\gamma)) - x| \leq \sqrt{t}$, which allows us to get that

$$(x - \lambda)^2 + \frac{\gamma^2}{4} \leq C_t \left( \text{Re}^2(w_{fp}(x + i\gamma) - \lambda)^2 + \frac{\gamma^2}{4} \right),$$

with $C_t$ as above. Thus,

$$J_2^\kappa \leq C_t \int_{\{|x| > \kappa\}} \int_\lambda \frac{d\mu_t(\lambda)}{(x - \lambda)^2 + \frac{\gamma^2}{4}} dx \leq \frac{2C_t}{\gamma} \int_\lambda d\mu_t(\lambda) \left( \frac{2\gamma}{\kappa} + \frac{\pi}{2} \{\lambda^2 > \frac{\gamma^2}{4} - \frac{\gamma^2}{4}\} \right).$$

Again, if we assume that $\kappa > \gamma$, this can be simplified as follows:

$$J_2^\kappa \leq C_t \left( \frac{4}{\kappa} + \frac{3\pi}{\gamma} \mu_t \left( \left\{ |\lambda| > \frac{\kappa}{2} \right\} \right) \right).$$

Moreover, letting $n$ going to infinity in (4.17), by Proposition 2.3 and dominated convergence, we get that, for any $\kappa > \gamma$,

$$\mu_t \left( \left\{ |\lambda| > \frac{\kappa}{2} \right\} \right) \leq \mu_0 \left( \left\{ |\lambda| > \frac{\kappa}{4} \right\} \right),$$

so that

$$J_2^\kappa \leq C_t \left( \frac{4}{\kappa} + \frac{3\pi}{\gamma} \mu_0 \left( \left\{ |\lambda| > \frac{\kappa}{4} \right\} \right) \right). \quad (4.41)$$

Gathering the upper bounds (4.40) and (4.41), we get that for any $\kappa > \gamma$,

$$J^\kappa \leq C_t \left( \frac{8}{\kappa} + \frac{6\pi}{\gamma} \mu_0 \left( \left\{ |\lambda| > \frac{\kappa}{4} \right\} \right) + \frac{3\pi}{\gamma} e^{-n.Ceig.M.C_eig.\kappa^4} \right). \quad (4.42)$$

This ends the proof. \hfill \Box

4.2.3. Conclusion

As a result, combining Lemma 4.7 and Lemma 4.9, we have:

$$I^\kappa + J^\kappa \leq \frac{\gamma^8}{(\gamma^2 - 4t)^4} \frac{C_{\text{var}}^2}{n} + \frac{\kappa C_{\text{var}}^2}{n^2} + C_1 C_{\text{var}} C_{\text{eig}} M + \frac{C_1}{\kappa} + C_2 e^{-n.C_{\text{eig}}.n} + C_3 \mu_0 \left( \left\{ |\lambda| > \frac{\kappa}{4} \right\} \right).$$

We take $M$ the smallest constant only depending on $t$ satisfying conditions of Lemma 4.6 and $\kappa = n$. Using Assumption (4.3), we obtain $\mu_0 \left( \left\{ |\lambda| > n \right\} \right) \leq Cn^{-1}$, for some absolute constant $C$. Then, from (4.11) and previous computations, there exists a constant $C_{\text{var}}(t)$ (that depends only on $t$) such that for $n$ sufficiently large:

$$\mathbb{E}(\Sigma) \leq \frac{\gamma^8}{(\gamma^2 - 4t)^4} \frac{C_{\text{var}}(t).e^{\frac{\gamma^2}{4}}}{n} \quad (4.43)$$

and Theorem 4.1 is proved.
5. Numerical simulations

In this section, we conduct a simulation study to assess the performances of our estimator \( \hat{p}_{0,h} \) designed in Definition 2.9 based on the \( n \)-sample \( \lambda^n(t) := \{ \lambda_1^n(t), \ldots, \lambda_n^n(t) \} \) of (non ordered) eigenvalues. We consider the sample size \( n = 4000 \) and the time value \( t = 1 \). We focus on initial conditions following a Cauchy distribution with scale parameter \( s_d = 5 \):

\[
p_0(x) = \frac{1}{\pi} \frac{s_d}{s_d^2 + x^2}, \quad x \in \mathbb{R}.
\]

We also consider the case of a mixture of Gaussian distributions with different variances with \( p_0 \) the density of \( wZ_1 + (1 - w)Z_2 \) where \( w \), \( Z_1 \) and \( Z_2 \) independent and \( w \sim \text{Ber}(0.25), Z_1 \sim \mathcal{N}(0,1) \) and \( Z_2 \sim \mathcal{N}(10,4) \).

Expression (2.15) is used with the kernel \( K(x) = \text{sinc}(x) = \sin(x)/(\pi x) \), and the value \( \gamma = 2\sqrt{t} + 0.01 \) so that the condition \( \gamma > 2\sqrt{t} \) is satisfied. To implement \( \hat{p}_{0,h} \), we approximate integrals involved in Fourier and inverse Fourier transforms by Riemann sums, so it may happen that \( \hat{p}_{0,h}(x) \) is not real. This is the reason why the density \( p_0 \) is estimated with \( \text{Re}(\hat{p}_{0,h}) \), the real part of \( \hat{p}_{0,h} \).

The theoretical bandwidth \( h \) proposed in Section 4 cannot be used in practice and we suggest the following data-driven selection rule, inspired from the principle of cross-validation. We decompose the quadratic risk for \( \text{Re}(\hat{p}_{0,h}) \) as follows:

\[
\| \text{Re}(\hat{p}_{0,h}) - p_0 \|^2 = \int_{\mathbb{R}} |\text{Re}(\hat{p}_{0,h}(x)) - p_0(x)|^2 \, dx = \| \text{Re}(\hat{p}_{0,h}) \|^2 - 2 \int \text{Re}(\hat{p}_{0,h}(x))p_0(x) \, dx + \| p_0 \|^2.
\]

Then, an ideal bandwidth \( h \) would minimize the criterion \( J \) with

\[
J(h) := \| \text{Re}(\hat{p}_{0,h}) \|^2 - 2 \int \text{Re}(\hat{p}_{0,h}(x))p_0(x) \, dx, \quad h \in \mathbb{R}^+.
\]

Since \( J \) depends on \( p_0 \) through the second term, we investigate a good estimate of this criterion. For this purpose, we divide the sample \( \lambda^n(t) \) into two disjoint sets

\[
\lambda^{n,E}(t) := \{ \lambda_i^n(t) \}_{i \in E} \quad \text{and} \quad \lambda^{n,E^c}(t) := \{ \lambda_i^n(t) \}_{i \in E^c}.
\]

There are \( V_{\text{max}} := \binom{n}{n/2} \) possibilities to select the subsets \( (E, E^c) \), which is huge. Hence, to reduce computational time, we draw randomly \( V = 10 \) partitions denoted \( (E_j, E_j^c)_{j=1,...,V} \). Choosing the grid \( \mathcal{H} \) of 50 equispaced points lying between \( h_{\text{min}} = 0.25 \) and \( h_{\text{max}} = 2.7 \), our selected bandwidth is

\[
\hat{h} = \arg\min_{h \in \mathcal{H}} \text{Crit}(h) \tag{5.1}
\]

with

\[
\text{Crit}(h) := \min_{h' \in \mathcal{H}, h' \neq h} \frac{1}{V} \sum_{j=1}^{V} \left( \left\| \text{Re}(\hat{p}_{0,h}(E_j)) \right\|^2 - 2 \int \text{Re}(\hat{p}_{0,h}(x)) \text{Re}(\hat{p}_{0,h'}(x)) \, dx \right)
\]

and our final estimator is then \( \hat{p}_{0,h} \). In the last expression, \( \hat{p}_{0,h}(E_j) \) and \( \hat{p}_{0,h'}(E_j^c) \) are estimates based on the samples \( E_j \) and \( E_j^c \) respectively.
To evaluate our approach, Figure 1 displays the plot of $h \mapsto \text{Crit}(h)$ and $h \mapsto J(h)$ for each density $p_0$. A close inspection of the graphs shows that the minimizer of the first criterion is a good estimate of the minimizer of the second one. As expected, for both criterions, we observe a plateau containing minimizers of $J$ and Crit. Outside the plateau, both criterions take large values due to large variance when $h$ is too small and to large bias when $h$ is too large.

Figure 2 gives the reconstruction provided by $\text{Re}(\hat{p}_{0,h})$ for each density $p_0$. The results are quite satis-

Figure 1: (a): Plots of $h \mapsto \text{Crit}(h)$ and $h \mapsto J(h)$ (a): for the Cauchy density. (b): for the mixture of Gaussian densities.

Figure 2: Estimation of $p_0$ (a): for the Cauchy density (b): for the mixture of Gaussian densities.
fying, meaning that our estimation procedure seems to perform well in practice for estimating initial conditions of the Fokker-Planck equation. For further numerical studies, we refer the reader to [23].

Acknowledgements

The authors thank P. Tarrago for useful discussions. M.M. acknowledges support from the Labex CEMPI (ANR-11-LABX-0007-01). T.D.N. was supported by a public grant as part of the Investissement d’avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH. V.C.T. is partly supported by Labex Bézout (ANR-10-LABX-58) and by the Chair “Modélisation Mathématique et Biodiversité” of Veolia Environnement-Ecole Polytechnique-Museum National d’Histoire Naturelle-Fondation X. The authors would like to thank the Editor and two anonymous referees for valuable comments and suggestions.

References

[1] G.W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
[2] O. Arizmendi, P. Tarrago, and C. Vargas. Subordination methods for free deconvolution. submitted. arxiv:1711.08871, 2020.
[3] Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, volume 37 of *Series in Statistics*. Springer, 2 edition, 2006.
[4] S. Belinschi and H. Bercovici. A new approach to subordination results in free probability. *Journal d’Analyse Mathematique*, 101:357–365, 2007.
[5] J. Bertoin, C. Giraud, and Y. Isozaki. Statistics of a flux in Burgers turbulence with one-sided Brownian initial data. *Comm. Math. Phys.*, 224(2):551–564, 2001.
[6] P. Biane. On the free convolution with a semi-circular distribution. *Indiana University Mathematics Journal*, 46(3):705–718, 1997.
[7] P. Biane and R. Speicher. Free diffusions, free entropy and free Fisher information. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(5):581–606, 2001.
[8] J. Bourgain. Periodic nonlinear Schrödinger equation and invariant measures. *Comm. Math. Phys.*, 166(1):1–26, 1994.
[9] J.M. Burgers. *The Nonlinear Diffusion Equation*. Springer, 1974.
[10] N. Burq and N. Tzvetkov. Random data cauchy theory for supercritical wave equations i: Local theory. *Inventiones Mathematicae*, 173:449–475, 2008.
[11] N. Burq and N. Tzvetkov. Random data cauchy theory for supercritical wave equations ii: A global result. *Inventiones Mathematicae*, 173:477–496, 2008.
[12] C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.*, 52(1):111–128, 2007.
[13] J.A. Carrillo, R.J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.
[14] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(2):569–609, 2013.
[15] P. Constantin and J. Wu. Statistical solutions of the Navier-Stokes equations on the phase space of vorticity and the inviscid limit. *Journal of Mathematical Physics*, 38(6):3031–3045, 06 1997.

[16] L. Erdős and H. T. Yau. A dynamical approach to random matrix theory. *Courant Lecture Notes in Mathematics*, 28, 2017.

[17] J. Fan and J.-Y. Koo. Wavelet deconvolution. *IEEE Trans. Inform. Theory*, 48(3):734-747, 2002.

[18] S. Dallaporta and M. Fevrier. Fluctuations of linear spectral statistics of deformed wigner matrices. submitted. hal-02079313, 2019.

[19] F. Flandoli. Weak vorticity formulation of 2D Euler equations with white noise initial condition. *Comm. Partial Differential Equations*, 43(7):1102–1149, 2018.

[20] C. Giraud. Some properties of burgers turbulence with white noise initial conditions. In *Probabilistic Methods in Fluids*, pages 161–178. World Scientific, 2003.

[21] C. Lacour. Rates of convergence for nonparametric deconvolution. *Comptes rendus de l’Académie des sciences. Série I, Mathématique*, 342(11):877-882, 2006.

[22] S. Méléard. Asymptotic behaviour of some interacting particle systems, McKean-Vlasov and Boltzmann models. In *CIME Lectures*, volume 1627 of *Lecture Notes in Mathematics*, pages 45–95. Springer, 1996.

[23] T.D. Nguyen. *Statistical deconvolution of Fokker-Planck equation*. PhD thesis, Université Paris Saclay, Paris, France, 2021.

[24] M. Pensky and T. Sapatinas. Functional deconvolution in a periodic setting: uniform case. *Ann. Statist.*, 37(1):73–104, 2009.

[25] M. Pensky and T. Sapatinas. On convergence rates equivalency and sampling strategies in functional deconvolution models. *Ann. Statist.*, 38(3):1793–1844, 2010.

[26] A.S. Sznitman. Topics in propagation of chaos. In *Ecole d’Ete de Probabilités de Saint-Flour XIX*, volume 1464 of *Lecture Notes in Mathematics*, pages 165–251, Berlin, 1991. Springer.

[27] D. Talay and O. Vaillant. A stochastic particle method with random weights for the computation of statistical solutions of McKean-Vlasov equations. *The Annals of Applied Probability*, 13(1):140–180, 2003.

[28] T. Tao. Topics in random matrix theory *Graduate Studies in Mathematics*, American Mathematical Society, 132, 2012.

[29] P. Tarrago. Spectral deconvolution of unitary invariant matrix models. arXiv:2006.09356, 2020.

[30] V.C. Tran. A wavelet particle approximation for McKean-Vlasov and Navier-Stokes spatial statistical solutions. *Stochastic Processes and their Applications*, 118(2):284–318, 2008.

[31] Alexandre Tsybakov. On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(9):835–840, 2000.

[32] N. Tzvetkov. Random data wave equations. arXiv:1704.01191, 2017.

[33] M. Vergassola, B. Dubrulle, U. Frisch, and A. Noullez. Burgers’equation, Devil’s staircases and the mass distribution for large-scale structures. *Astronomy and Astrophysics*, 289:325–356, 1994.

[34] M.J. Vishik and A.V. Fursikov. *Mathematical Problems of Statistical Hydromechanics*. Mathematics and its Applications. Kluwer Academic Publishers, 1980.

[35] D. Voiculescu. Addition of certain noncommuting random variables. *J. Funct. Anal.*, 66(3):323–346, 1986.

[36] D. Voiculescu. The analogues of entropy and Fisher’s information measure in free probability theory, i. *Commun. Math. Phys.*, 155:71–92, 1993.
Appendix A: Proof of (1.7)

As mentioned in the introduction, a full proof of (1.7) can be found in [1, Theorem 4.3.2]. The proof therein is involved and proceeds backward, showing that the solutions of (1.7) are the eigenvalues of an Hermitian Brownian motion. In this appendix, we use a more direct approach (following for example [16, 28]) that leads to a non rigorous but more intuitive sketch of proof. Recall that \( X^n(t) = X^n(0) + H^n(t) \) where \( H^n(t) \) is the Hermitian Brownian motion of Definition 2.1. For \( k \leq \ell \) and \( t > 0 \), we denote by \( x_{k\ell}(t) := \text{Re}X^n_{k\ell}(t) \) and \( y_{k\ell}(t) := \text{Im}X^n_{k\ell}(t) \) respectively the real and imaginary parts of the entries of the matrix \( X^n(t) \). The processes \( x_{k\ell} \) and \( y_{k\ell} \) are semi-martingales and we will assume that for any \( m \in \{1, \ldots, n\} \), the \( m \)-th smallest eigenvalue \( \lambda_m^n(t) \) of \( X^n(t) \) is a smooth function of \( (x_{k\ell}, y_{k\ell})_{k \leq \ell} \) so that we can apply Itô’s formula: 

\[
d\lambda_m = \sum_{k < \ell} \frac{\partial \lambda_m}{\partial x_{k\ell}} dx_{k\ell} + \sum_{k < \ell} \frac{\partial \lambda_m}{\partial y_{k\ell}} dy_{k\ell} + \sum_{k=1}^{n} \frac{\partial \lambda_m}{\partial x_{kk}} dx_{kk} + \frac{1}{4n} \sum_{k < \ell} \left( \frac{\partial^2 \lambda_m}{\partial x_{k\ell}^2} + \frac{\partial^2 \lambda_m}{\partial y_{k\ell}^2} \right) dt + \frac{1}{2n} \sum_{k=1}^{n} \frac{\partial^2 \lambda_m}{\partial x_{kk}^2} dt,
\]

where we have used that, in the range of indices we are interested in, \( \langle x_{ij}, y_{k\ell} \rangle = 0 \); if \( i \neq j \), \( d\langle x_{ij}, x_{k\ell} \rangle = d\langle y_{ij}, y_{k\ell} \rangle = \frac{\partial}{\partial m} \delta_{hk} \delta_{\ell\ell} \), and \( d\langle x_{ii}, x_{ii} \rangle = \frac{\partial}{\partial \lambda_m} \). We now have to compute the derivatives. It relies on the so-called Hadamard variation formulae, well-known in perturbation theory.

**Lemma A.1.** Let \( H \) be an Hermitian matrix, with entries \( (h_{k\ell} = x_{k\ell} + iy_{k\ell})_{1 \leq k < \ell \leq n} \). We assume that \( H \) has distinct (real) eigenvalues \( \lambda_1, \ldots, \lambda_n \) and corresponding eigenvectors \( u_1, \ldots, u_n \). Then, denoting by \( u_{km} \) the \( k \)-th component of the vector \( u_m \), we have for all \( m \in \{1, \ldots, n\} \):

\[
\frac{\partial \lambda_m}{\partial x_{k\ell}} = \bar{u}_{km} u_{\ell m} + \bar{u}_{\ell m} u_{km}, \text{ for } k < \ell, \\
\frac{\partial \lambda_m}{\partial y_{k\ell}} = i(\bar{u}_{km} u_{\ell m} - \bar{u}_{\ell m} u_{km}), \text{ for } k < \ell, \\
\frac{\partial \lambda_m}{\partial x_{kk}} = |u_{km}|^2, \\
\frac{\partial^2 \lambda_m}{\partial x^2_{k\ell}} = 2 \sum_{m' \neq m} \frac{1}{\lambda_m - \lambda_{m'}} |\bar{u}_{km'} u_{\ell m} + \bar{u}_{\ell m'} u_{km}|^2, \text{ for } k < \ell \\
\frac{\partial^2 \lambda_m}{\partial y^2_{k\ell}} = 2 \sum_{m' \neq m} \frac{1}{\lambda_m - \lambda_{m'}} |\bar{u}_{km'} u_{\ell m} - \bar{u}_{\ell m'} u_{km}|^2, \text{ for } k < \ell \\
\frac{\partial^2 \lambda_m}{\partial x^2_{kk}} = 2 \sum_{m' \neq m} \frac{1}{\lambda_m - \lambda_{m'}} |u_{km}|^2 |u_{km'}|^2.
\]

**Proof.** Again, we assume here that all the functions that we use hereafter are smooth functions of the real and imaginary parts of the entries of the matrix. For \( k \leq \ell \), let us denote by \( \partial \) the derivative \( \frac{\partial}{\partial x_{k\ell}} \) or \( \frac{\partial}{\partial y_{k\ell}} \). The matrix \( \partial H \) corresponds to the matrix whose entries are \( \partial h_{k\ell} \).

---

1This is far from obvious and the actual rigorous proof does not proceed like that.
For any $m, m' \in \{1, \ldots, n\}$, we have $H.u_m = \lambda_m u_m$, and $u_m^* u_{m'} = \delta_{mm'}$, where in this proof $\delta_{mm'}$ is the Kronecker symbol equal to 1 if and only if $m = m'$ and 0 otherwise, and where $u_m^*$ is the adjoint vector of $u_m$ defined as the row vector with $k$-th component $u_m^{*k} = \text{Re}(u_{km}) - i\text{Im}(u_{km})$. Thus,

$$\partial H.u_m + H.\partial u_m = \partial \lambda_m \times u_m + \lambda_m \partial u_m,$$

(A.2)

and for all $m$ and $m'$ (possibly equal):

$$\partial u_m^* u_{m'} + u_m^* \partial u_{m'} = 0.$$

(A.3)

Multiplying (A.2) by $u_m^*$ on the left, we get the first Hadamard formula:

$$\partial \lambda_m = u_m^* \partial H.u_m.$$

(A.4)

Now multiplying (A.2) by $u_{m'}^*$ on the left, we get, for $m \neq m'$,

$$u_{m'}^* \partial H.u_m = (\lambda_m - \lambda_{m'})u_{m'}^* \partial u_m,$$

so that

$$\partial u_m = \sum_{m' = 1}^{n} (u_{m'}^* \partial u_m) u_{m'} = \sum_{m' \neq m} \frac{u_{m'}^* \partial H.u_m}{\lambda_m - \lambda_{m'}} u_{m'} + (u_m^* \partial u_m) u_m.$$

From there, taking the derivative of the first Hadamard formula (A.4) and using the above equality with (A.3) leads to the second Hadamard equality:

$$\partial^2 \lambda_m = u_m^* \partial^2 H.u_m + 2 \sum_{m' \neq m} \frac{|u_{m'}^* \partial H.u_m|^2}{\lambda_m - \lambda_{m'}}.$$

(A.5)

Now, for $\partial = \frac{\partial}{\partial x}$ or $\partial = \frac{\partial}{\partial y}$, we have that $\partial^2 H = 0$. Moreover, $\frac{\partial H}{\partial x_k}$ is the matrix full of zeros except for the terms $(k, \ell)$ and $(\ell, k)$ that are equal to 1 and $\frac{\partial H}{\partial y_{k\ell}} (k < \ell)$ is the matrix full of zeros except for the terms $(k, \ell)$ equal to i and $(\ell, k)$ that are equal to $-i$. Injecting this information into (A.4) and (A.5) provides the announced derivatives.

Plugging the formulae of Lemma A.1 into the Itô formula (A.1) above, we get

$$\partial \lambda_m = \frac{1}{\sqrt{n}} \sum_{m} \beta_m + \frac{1}{n} \sum_{m' \neq m} \frac{1}{\lambda_m - \lambda_{m'}} dt,$$

with

$$d \beta_m := \frac{1}{\sqrt{2}} \sum_{k < \ell} ((\bar{u}_{km} u_{\ell m} + \bar{u}_{\ell m} u_{km}) dB_{k,\ell} + i (\bar{u}_{km} u_{\ell m} - \bar{u}_{\ell m} u_{km}) d\tilde{B}_{k,\ell}) + \sum_{k = 1}^{n} |u_{km}|^2 dB_{kk}.$$

$\beta_1, \ldots, \beta_n$ are centered semimartingales. Furthermore,

$$d(\beta_m, \beta_{m'}) = \sum_{k, \ell = 1}^{n} \bar{u}_{km} u_{\ell m} \bar{u}_{\ell m'} u_{km'} dt = \delta_{mm'} dt,$$

so that they are independent standard Brownian motions.
Appendix B: Proof of Theorem 2.6 and Theorem-Definition 2.8

B.1. Proof of Theorem 2.6

The constants of Theorem 2.6 are better than the ones of Arizmendi et al. [2] who work in full generality. We develop here the main steps of the proof in our context, using the explicit formula for the semi-circular distribution. In the whole proof, we consider \( z \in \mathbb{C}_{\frac{1}{2} \text{Im}(z)} \).

**Step 1:** We first prove that the function \( L_z(w) = h_{\sigma_t}(\tilde{h}_{\mu_t}(w) - z) + z \) is well-defined and analytic on \( \mathbb{C}_{\frac{1}{2} \text{Im}(z)} \). Since \( h_{\sigma_t} \) is defined on \( \mathbb{C}^+ \), we need to check that \( \tilde{h}_{\mu_t}(w) - z \in \mathbb{C}^+ \) for \( w \in \mathbb{C}_{\frac{1}{2} \text{Im}(z)} \). This is satisfied since for such \( w \),

\[
\text{Im}(\tilde{h}_{\mu_t}(w) - z) = \text{Im}(w + F_{\mu_t}(w) - z) \geq 2\text{Im}(w) - \text{Im}(z) > 0,
\]

where we have used \( \text{Im} F_{\mu_t}(w) \geq \text{Im}(w) \) for the first inequality. Indeed, if \( w = w_1 + iw_2 \), we have

\[
(F_{\mu_t}(w))^{-1} = G_{\mu_t}(w) = \int \frac{d\mu_t(x)}{w_1 + iw_2 - x} = \int \frac{(w_1 - x)d\mu_t(x)}{(w_1 - x)^2 + w_2^2} - iw_2 \int \frac{d\mu_t(x)}{(w_1 - x)^2 + w_2^2}
\]

and

\[
\text{Im}(F_{\mu_t}(w)) = w_2 \times \frac{\int (w_1 - x)d\mu_t(x)}{(w_1 - x)^2 + w_2^2} \geq w_2 \times \frac{\int (w_1 - x)d\mu_t(x)}{(w_1 - x)^2 + w_2^2} = w_2
\]

**Step 2:** We show that \( L_z(\mathbb{C}_{\frac{1}{2} \text{Im}(z)}) \subset \mathbb{C}_{\frac{1}{2} \text{Im}(z)} \) and that \( L_z \) is not a conformal automorphism.

First, let us show that \( L_z(\mathbb{C}_{\frac{1}{2} \text{Im}(z)}) \subset \mathbb{C}_{\frac{1}{2} \text{Im}(z)} \). Let \( w \in \mathbb{C}_{\frac{1}{2} \text{Im}(z)} \), we have:

\[
\text{Im}(L_z(w)) = \text{Im} \left[ tG_{\sigma_t}(\tilde{h}_{\mu_t}(w) - z) + z \right] = \text{Im} \left( \tilde{h}_{\mu_t}(w) - z - \sqrt{\left(\tilde{h}_{\mu_t}(w) - z\right)^2 - 4t} + z \right) \quad \text{(B.3)}
\]

To lower bound the right hand side, note that for all \( v \in \mathbb{C}^+ \), one can check that:

\[
\text{Im} \left( \sqrt{v^2 - 4t} \right) \leq \sqrt{\text{Im}^2(v) + 4t}.
\]

Therefore, we have:

\[
\text{Im} \left( \sqrt{\left(\tilde{h}_{\mu_t}(w) - z\right)^2 - 4t} \right) \leq \sqrt{\text{Im}^2(\tilde{h}_{\mu_t}(w) - z)^2 + 4t}.
\]
This implies that 

\[ L_{\overline{M} \mathrm{a} \ddot{i} d a} \text{ from (B.5), any } z \text{ the unique Denjoy-Wolff point of } H. \]

Hence, for all \( w \)

\[ \text{The function } g(s) = s - \sqrt{s^2 + 4t} \text{ is non-decreasing on } \mathbb{R}_+ \text{ and for all } s > 0, g(s) \geq -2\sqrt{t}. \text{ Thus:} \]

\[ \text{Im}(L_z(w)) \geq \text{Im}(z) - \sqrt{t} > \frac{1}{2} \text{Im}(z), \quad (B.4) \]

since \( z \in \mathbb{C}_{2\sqrt{t}}. \) This guarantees that \( L_z(w) \in \mathbb{C}_{\frac{1}{2}\text{Im}(z)}. \)

Let us now prove that \( L_z \) is not an automorphism of \( \mathbb{C}_{\frac{1}{2}\text{Im}(z)}. \) Consider

\[ |L_z(w) - z| = \left| F_{\sigma_t}(\hat{\mu}_t(w) - z) - (\hat{\mu}_t(w) - z) \right| = |tG_{\sigma_t}(\hat{\mu}_t(w) - z)|. \]

For \( v \in \mathbb{C}^+, \) if \( |v| > 3\sqrt{t}, \) since the support of \( \sigma_t \) is \([-2\sqrt{t}, 2\sqrt{t}], \)

\[ |tG_{\sigma_t}(v)| = \left| \int_{-2\sqrt{t}}^{2\sqrt{t}} \frac{t}{v - x} \, d\sigma_t(x) \right| \leq \sqrt{t}. \]

If \( |v| \leq 3\sqrt{t}, \)

\[ |tG_{\sigma_t}(v)| = \left| \frac{v - \sqrt{v^2 - 4t}}{2} \right| \leq \frac{2|v| + 2\sqrt{t}}{2} \leq 4\sqrt{t}. \]

Hence, for all \( w \in \mathbb{C}_{\frac{1}{2}\text{Im}(z)}, \)

\[ |L_z(w) - z| \leq 4\sqrt{t}. \quad (B.5) \]

This implies that \( L_z \left( \mathbb{C}_{\frac{1}{2}\text{Im}(z)} \right) \) is included in the ball centered at \( z \) with radius \( 4\sqrt{t}. \) As a result, \( L_z \) is not surjective and hence is not an automorphism of \( \mathbb{C}_{\frac{1}{2}\text{Im}(z)}. \)

**Step 3:** Existence and uniqueness of \( w_{fp}, \) which is a fixed point of \( L_z. \)

By Steps 1 and 2, \( L_z \) satisfies the assumptions of Denjoy-Wolff’s fixed-point theorem (see e.g. [4, 2]). The theorem says that for all \( w \in \mathbb{C}_{\frac{1}{2}\text{Im}(z)} \) the iterated sequence \( L_z^n(w) = L_z \circ L_z^{(n-1)}(w) \) converges to the unique Denjoy-Wolff point of \( L_z \) which we define as \( w_{fp}(z). \) The Denjoy-Wolff point is either a fixed-point of \( L_z \) or a point of the boundary of the domain. Let us check that \( w_{fp} \) is a fixed point of \( L_z. \) For any \( z \in \mathbb{C}_{2\sqrt{t}}, \) there exists \( \gamma > 2 \) such that \( z \in \mathbb{C}_{\gamma \sqrt{t}} \) and from (B.4), \( L_z(\mathbb{C}_{\frac{1}{2}\text{Im}(z)}) \subset \mathbb{C}_{(1-\frac{1}{\gamma})\text{Im}(z)}. \) Moreover, from (B.5), \( L_z(\mathbb{C}_{\frac{1}{2}\text{Im}(z)}) \subset B(z, 4\sqrt{t}). \) Therefore, \( w_{fp}(z) \in \mathbb{C}_{(1-\frac{1}{\gamma})\text{Im}(z)} \cap B(z, 4\sqrt{t}) \subset \mathbb{C}_{\frac{1}{2}\text{Im}(z)}, \) so that it is necessarily a fixed point.

We now define

\[ w_1(z) := F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) - z. \]
One can check that
\[ F_{\sigma_t}(w_1(z)) = w_1(z) - h_{\sigma_t}(w_1(z)) \]
\[ = F_{\mu}(w_{fp}(z)) + w_{fp}(z) - z - h_{\sigma_t}(F_{\mu}(w_{fp}(z)) + w_{fp}(z) - z) \]
\[ = \tilde{h}_{\mu_t}(w_{fp}(z)) - z - h_{\sigma_t}(\tilde{h}_{\mu_t}(w_{fp}(z)) - z) \]
\[ = \tilde{h}_{\mu_t}(w_{fp}(z)) - w_{fp}(z) = F_{\mu}(w_{fp}(z)). \]

One can therefore rewrite
\[ w_1(z) = F_{\sigma_t}(w_1(z)) + w_{fp}(z) - z. \]

From (B.5) and the fact that \( w_{fp}(z) \) is a fixed point of \( L_z \), one easily gets that \( \lim_{y \to +\infty} w_{fp}(iy)/(iy) = 1 \), which implies that \( \lim_{y \to +\infty} F_{\mu}(w_{fp}(iy))/(iy) = 1 \), and \( \lim_{y \to +\infty} w_1(iy)/(iy) = 1 \).

Now we connect \( F_{\mu_0} \) to the previous quantities. For \( z \) large enough, all the functions we consider are invertible and we have
\[ F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) = z + w_1(z) = z + F_{\sigma_t}^{-1}(F_{\mu_t}(w_{fp}(z))). \]

On the other hand, for \( z \) large enough, using Theorem-definition 2.5 for \( \mu_1 = \sigma_t \) and \( \mu_2 = \mu_0 \), we get
\[ F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) = \alpha_1(w_{fp}(z)) + \alpha_2(w_{fp}(z)) = F_{\sigma_t}^{-1}(F_{\mu_t}(w_{fp}(z))) + F_{\mu_0}^{-1}(F_{\mu_t}(w_{fp}(z))). \]

Comparing the two equalities gives
\[ F_{\mu_0}^{-1}(F_{\mu_t}(w_{fp}(z))) = z, \]
so that, for \( z \) large enough,
\[ F_{\mu_t}(w_{fp}(z)) = F_{\mu_0}(z). \]

The two functions being analytic on \( \mathbb{C}_{2\sqrt{t}} \), the equality can be extended to any \( z \in \mathbb{C}_{2\sqrt{t}} \).

Finally, since
\[ w_1(z) = F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) - z = F_{\mu_0}(z) + w_{fp}(z) - z, \]
we have, using (B.2) with \( \mu_0 \) instead of \( \mu_t \),
\[ \operatorname{Im}(w_1(z)) = \operatorname{Im}(F_{\mu_0}(z)) + \operatorname{Im}(w_{fp}(z)) - \operatorname{Im}(z) \geq \operatorname{Im}(w_{fp}(z)) \geq \frac{1}{2} \operatorname{Im}(z). \]

This ends the proof of Theorem 2.6.

### B.2. Proof of Theorem-Definition 2.8

The proof of this theorem follows the steps of the proof of Theorem 2.6. First, \( \hat{L}_z(w) := t \hat{\mu}_n^\sigma(w) + z \) is a well-defined and analytic function on \( \mathbb{C}^+ \). Let us check that \( \hat{L}_z(\mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)}) \subset \mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)} \) for \( z \in \mathbb{C}_{2\sqrt{t}} \). For \( w = u + iv \in \mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)} \),
\[
\operatorname{Im}(\hat{\mu}_n^\sigma(w)) = \frac{1}{n} \sum_{j=1}^{n} \operatorname{Im} \left( \frac{u - \lambda_j^n(t) - iv}{(u - \lambda_j^n(t))^2 + v^2} \right) > -\frac{1}{v} = \frac{1}{\operatorname{Im}(w)}. \tag{B.9}
\]
Thus,
\[
\text{Im}(\tilde{L}_z(w)) = t \text{Im}(\tilde{G}_{\mu_1^n}(w)) + \text{Im}(z) > -\frac{t}{\text{Im}(w)} + \text{Im}(z) > -\frac{2t}{\text{Im}(z)} + \text{Im}(z) > \frac{1}{2}\text{Im}(z).
\]
The second and last inequalities come from the choice of \(w \in \mathbb{C}_{\frac{1}{2}\text{Im}(z)}\), and from \(\text{Im}(z) > 2\sqrt{t}\). Moreover, \(\hat{L}_z\) is not an automorphism since:
\[
\left| \hat{L}_z(w) - z \right| = \left| tG_{\mu_1^n}(w) \right| = \left| \frac{1}{n} \sum_{j=1}^{n} \frac{t}{w - \lambda_j^n(t)} \right| \leq \frac{t}{\text{Im}(w)} \leq \sqrt{t}
\] (B.10)

since \(\text{Im}(w) > \frac{1}{2}\text{Im}(z) > \sqrt{t}\). We use again the Denjoy-Wolff fixed-point theorem. Because the inclusion of \(\tilde{L}_z(\mathbb{C}_{\frac{1}{2}\text{Im}(z)})\) into \(\mathbb{C}_{\frac{1}{2}\text{Im}(z)}\) is strict, the unique Denjoy-Wolff point of \(\tilde{L}_z\) is necessarily a fixed point that we denote \(\tilde{w}_{fp}(z)\). From the construction, \(\text{Im}(\tilde{w}_{fp}(z)) > \text{Im}(z)/2\). Finally, the last announced estimate is a straightforward consequence of (B.10).

Appendix C: Proof of Lemma 3.4

Recall that \(R_{n,t}(z)\) and \(\tilde{R}_{n,t}(z)\) are defined in (3.4) and (3.9), and that
\[
nA^2(z) = \sum_{kk} E \left[ \left( R_{n,t}(z) \right)_{kk} \mid X^n(0) \right] - (\tilde{R}_{n,t}(z))_{kk}.
\] (C.1)

Proceeding as in Dallaporta and Février [18], we introduce some notations. Let \(R_{n,t}^{(k)}(z)\) be the resolvent of the \((n-1) \times (n-1)\) obtained from \(X^n(t)\) by removing the \(k\)-th row and column and \(C_{k,t}^{(k)}\) be the \((n-1)\)-dimensional vector obtained from the \(k\)-th column of \(H^n(t)\) by removing its \(k\)-th component. Using Schur’s complement (see e.g. [3, Appendix A.1]):
\[
\left( (R_{n,t}(z))_{kk} \right)^{-1} = z - (H^n(t))_{kk} - (X^n(0))_{kk} - C_{k,t}^{(k)} \cdot R_{n,t}(z) \cdot C_{k,t}^{(k)}.
\]

Because \(\tilde{R}_{n,t}(z)\) is a diagonal matrix, we have easily:
\[
(R_{n,t}(z))_{kk} = (\tilde{R}_{n,t}(z))_{kk}
\]
\[
+ (\tilde{R}_{n,t}(z))_{kk} \cdot (R_{n,t}(z))_{kk} \cdot \left( (H^n(t))_{kk} + C_{k,t}^{(k)} \cdot R_{n,t}(z) \cdot C_{k,t}^{(k)} - \frac{t}{n} E \left[ \text{Tr} \left( R_{n,t}(z) \mid X^n(0) \right) \right] \right).
\]

Replacing \((R_{n,t}(z))_{kk}\) in the right-hand side of the previous formula, we obtain:
\[
(R_{n,t}(z))_{kk} = (\tilde{R}_{n,t}(z))_{kk} \cdot (H^n(t))_{kk} + C_{k,t}^{(k)} \cdot R_{n,t}(z) \cdot C_{k,t}^{(k)} - \frac{t}{n} E \left[ \text{Tr} \left( R_{n,t}(z) \mid X^n(0) \right) \right]
\]
\[
+ (\tilde{R}_{n,t}(z))_{kk} \cdot (R_{n,t}(z))_{kk} \cdot (H^n(t))_{kk} + C_{k,t}^{(k)} \cdot R_{n,t}(z) \cdot C_{k,t}^{(k)} - \frac{t}{n} E \left[ \text{Tr} \left( R_{n,t}(z) \mid X^n(0) \right) \right]^2.
\] (C.2)
Statistical deconvolution of the free Fokker-Planck equation at fixed time

Since $H^n(t)$ and $C_{k,t}^{(k)}$ are independent of $X_n(0)$,

\[
\mathbb{E}\left[ \left( H^n(t) \right)_{kk} + C_{k,t}^{(k)*} R_{n,t}^{(k)}(z) C_{k,t}^{(k)} - \frac{t}{n} \mathbb{E}[\text{Tr}(R_{n,t}(z)) | X^n(0)] \right]^2 | X^n(0) \right] \\
= \mathbb{E}\left[ \left( H^n(t) \right)_{kk} + C_{k,t}^{(k)*} R_{n,t}^{(k)}(z) C_{k,t}^{(k)} - \frac{t}{n} \text{Tr}(R_{n,t}^{(k)}(z)) + \frac{t}{n} \text{Tr}(R_{n,t}^{(k)}(z)) - \frac{t}{n} \mathbb{E}[\text{Tr}(R_{n,t}^{(k)}(z)) | X^n(0)] \right]^2 | X^n(0) \right] \\
+ \frac{t}{n} \mathbb{E}[\text{Tr}(R_{n,t}(z)) | X^n(0)] - \frac{t}{n} \mathbb{E}[\text{Tr}(R_{n,t}(z)) | X^n(0)]^2 | X^n(0) \right] \\
= \mathbb{E}\left[ (H^n(t))_{kk}^{2} \right] + \mathbb{E}\left[ C_{k,t}^{(k)*} R_{n,t}^{(k)}(z) C_{k,t}^{(k)} - \frac{t}{n} \text{Tr}(R_{n,t}^{(k)}(z)) \right]^2 | X^n(0) \right] \\
+ \frac{t^2}{n^2} \left( \mathbb{V} \left[ \text{Tr}(R_{n,t}^{(k)}(z)) | X^n(0) \right] + \mathbb{E}[\text{Tr}(R_{n,t}^{(k)}(z)) - \text{Tr}(R_{n,t}(z)) | X^n(0)] \right)^2 \right) \right). \tag{C.3}
\]

We now upper bound each of the term in the right-hand side of (C.3). The first term equals to $t/n$.

**Step 1:** We upper bound the second term in (C.3). By Lemma 5 of [18],

\[
\mathbb{E}\left[ C_{k,t}^{(k)*} R_{n,t}^{(k)}(z) C_{k,t}^{(k)} | X^n(0) \right] = \frac{t}{n} \mathbb{E}[\text{Tr}(R_{n,t}^{(k)}(z)) | X^n(0)] \tag{C.4}
\]

Thus, the second term in (C.3) equals to $\mathbb{V}(C_{k,t}^{(k)*} R_{n,t}^{(k)}(z) C_{k,t}^{(k)} | X^n(0))$ and we have:

\[
\mathbb{V} \left[ C_{k,t}^{(k)*} R_{n,t}^{(k)}(z) C_{k,t}^{(k)} | X^n(0) \right] = \frac{t^2}{n^2} \mathbb{E} \left[ \text{Tr}(R_{n,t}^{(k)*}(z) R_{n,t}^{(k)}(z)) | X^n(0) \right] \\
\leq \frac{t^2}{n^2} \mathbb{E} \left[ \sum_{j=1}^{n} \frac{1}{|z - \lambda_j^{(k)}|^2} | X^n(0) \right] \]

where the $\lambda_j^{(k)}$'s are the eigenvalues of the matrix with resolvent $R_{n,t}^{(k)}(z)$. Hence,

\[
\mathbb{V} \left[ C_{k,t}^{(k)*} R_{n,t}^{(k)}(z) C_{k,t}^{(k)} | X^n(0) \right] \leq \frac{t^2}{n \text{Im}(z)}. \tag{C.5}
\]

**Step 2:** We now upper bound the third and fourth terms of (C.3). Let us denote in the sequel by $\mathbb{E}_k$ the expectation with respect to $\{(H^n(t))_{jk} : 1 \leq j \leq n\}$, and by $\mathbb{E}_{\leq k}$ the conditional expectation on the sigma-field $\sigma \left( \{ (X^n(0))_{i} : 1 \leq i \leq j \leq n \}, \{ (H^n(t))_{i} : 1 \leq i \leq j \leq k \} \right)$.

We have:

\[
\mathbb{V}[\text{Tr}(R_{n,t}^{(k)}(z)) | X^n(0)] \leq 2 \mathbb{V}[\text{Tr}(R_{n,t}(z)) | X^n(0)] + 2 \mathbb{V}[\text{Tr}(R_{n,t}(z)) - \text{Tr}(R_{n,t}^{(k)}(z)) | X^n(0)]. \tag{C.6}
\]
For the first term,

\[
\text{Var}\left[\text{Tr}(R_{n,t}(z)) \mid X^n(0)\right] = \sum_{k=1}^{n} E \left[\left(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\right)\text{Tr}(R_{n,t}(z))\right]^2 \mid X^n(0)\right]
\]

\[
= \sum_{k=1}^{n} E \left[\left(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\right)(\text{Tr}(R_{n,t}(z)) - \text{Tr}(R_{n,t}(z)))\right]^2 \mid X^n(0)\right].
\]  

(C.7)

as \((\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1})\text{Tr}(R_{n,t}(z)) = 0\). The Schur complement formula (see e.g. [3, Appendix A.1]) gives that:

\[
\text{Tr}(R_{n,t}(z)) - \text{Tr}(R_{n,t}(z)) = \frac{1 + C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot C^{(k)}_k.}{z - (H^n(t))_{kk} - (X^n(0))_{kk} - C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot C^{(k)}_k.}
\]  

(C.8)

Then,

\[
\left|\text{Tr}(R_{n,t}(z)) - \text{Tr}(R_{n,t}(z))\right| \leq \frac{\left|1 + C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot C^{(k)}_k.\right|}{\text{Im}(z) - \text{Im} \left(C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot C^{(k)}_k.\right)}
\]

\[
\leq \frac{1 + C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot C^{(k)}_k.}{\text{Im}(z) - \text{Im} \left(C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot C^{(k)}_k.\right)}
\]

\[
\leq \frac{1 + C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot R_{n,t}(z) \cdot C^{(k)}_k.}{\text{Im}(z) + \text{Im} \left(C^{(k)\ast}_k \cdot R_{n,t}(z) \cdot R_{n,t}(z) \cdot C^{(k)}_k.\right)}
\]

\[
= \frac{1}{\text{Im}(z)}
\]  

(C.9)

The second inequality it due to the fact that \((H^n(t))_{kk}, (X^n(0))_{kk} \in \mathbb{R}\) and the third inequality comes from the following equality: With \(\Psi : M \in \mathcal{H}_n(\mathbb{C}) \rightarrow C^*MC\) with \(C \in \mathbb{C}^n\), then, for any \(z \in \mathbb{C}\) and any resolvent matrix \(R(z)\), we have (see [18, Lemma 1])

\[
\text{Im}(\Psi(R(z))) = -\text{Im}(z)\Psi(R(z)\ast R(z)).
\]

The bound (C.9) does not depend on \(X^n(0)\). Plugging this bound into (C.7), we obtain:

\[
\text{Var}\left[\text{Tr}(R_{n,t}(z)) \mid X^n(0)\right] \leq \frac{4n}{\text{Im}^2(z)}.
\]

From there, using (C.6),

\[
\text{Var}\left[\text{Tr}(R_{n,t}^{(k)}(z)) \mid X^n(0)\right] \leq \frac{8n + 2}{\text{Im}^2(z)}.
\]  

(C.10)

Similarly, (C.9) also provides an upper bound for the fourth term of (C.3):

\[
\left|\mathbb{E}\left[\text{Tr}(R_{n,t}(z)) - \text{Tr}(R_{n,t}(z)) \mid X^n(0)\right]\right| \leq \frac{1}{\text{Im}^2(z)}.
\]  

(C.11)
**Step 3:** In conclusion, using (C.3), (C.5), (C.10) and (C.11), we obtain that:

\[
\mathbb{E}\left[\left|\left(C_{k,t}\right)_{kk} + \mathcal{R}_{n,t}(z) C_{k,t} \frac{t}{n} \mathbb{E}\left[\text{Tr}\left(R_{n,t}(z)\right) \mid X^n(0)\right]\right|^2 \left|X^n(0)\right|\right] \leq \frac{t}{n} + \frac{t^2}{n\text{Im}^2(z)} + \frac{(8n + 3)t^2}{n^2\text{Im}^2(z)}.
\]

Going back to (C.2) and using (C.4) to upper-bound the first term in the right-hand side:

\[
\mathbb{E}\left[\left|\left(R_{n,t}(z)\right)_{kk} - \mathcal{R}_{n,t}(z)\right|_{kk} \mid X^n(0)\right]\right| \leq \frac{t}{n} \left|\mathcal{R}_{n,t}(z)\right|_{kk}^2 \mathbb{E}\left[\left|\text{Tr}\left(R_{n,t}(z)\right) - \text{Tr}\left(R_{n,t}(z)\right)\right| \mid X^n(0)\right]
\]

\[
+ \left|\left(R_{n,t}(z)\right)_{kk}^2 \mathbb{E}\left[\left|\left(R_{n,t}(z)\right)_{kk} \mid (H^n(t))_{kk} + C_{k,t} R_{n,t}(z) C_{k,t}\right] \right| \mid X^n(0)\right|^2 \left|X^n(0)\right]
\]

\[
- \frac{t}{n} \mathbb{E}\left[\text{Tr}\left(R_{n,t}(z)\right) \mid X^n(0)\right]^2 \left|X^n(0)\right]\right| \leq \left|\mathcal{R}_{n,t}(z)\right|_{kk} \cdot \frac{2t}{n\text{Im}(z)} + \frac{12t^2}{\text{Im}^3(z)}.
\]

Using this upper bound in (C.1), we obtain by summation the result and using that for any \(k\),

\[
\left|\mathcal{R}_{n,t}(z)\right|_{kk}^2 \leq \frac{1}{\text{Im}^2(z)}.
\]

**Appendix D: Proof of Lemma 3.5**

From (3.7) and introducing \(\bar{w}_1(z)\) such that:

\[
G_{\mu_0}(\bar{w}_1(z)) = G_{\mu_0}(\bar{w}_1(z)) = G_{\sigma_0}(\bar{w}_1(z)).
\]

We can derive from Theorem-Definition 2.5 that \(\bar{w}_p(z)\) solves the equation (i) of Lemma 3.5 and that:

\[
z = \bar{w}_p(z) + tG_{\mu_0}(\bar{w}_p(z)),
\]

for all \(z \in \mathbb{C}^+\). The latter equation justifies (ii) of Lemma 3.5.

**Appendix E: Proof of Corollary 4.3**

Recall that from Proposition 4.2 and Theorem 4.1, the mean integrated square error is

\[
MISE = \mathbb{E}\left[\left\|\hat{p}_{0,h} - p_0\right\|^2\right] \leq Le^{-2ah} + \frac{\gamma^8}{\left(\gamma^2 - 4t\right)^2} C_{var}.e^{\frac{2\gamma}{n}}.
\]
Minimizing in $h$ amounts to solving the following equation obtained by taking the derivative in the right hand side of (4.6):

$$
\psi(h) := \exp \left( \frac{2\gamma}{h} + \frac{2a}{h^r} \right) h^{r-1} = O(n).
$$

(E.1)

Consequently for the minimizer $h_*$ of (E.1) we get that

$$
\frac{e^{2\gamma}}{n} = Ch_*^{1-r} e^{-2ah_*^r},
$$

for some constant $C > 0$. Hence, in view of (4.6), when $r < 1$ the bias dominates the variance and the contrary occurs when $r > 1$. Thus, there are three cases to consider to derive rates of convergence: $r = 1$, $r < 1$ and $r > 1$. To solve the equation (E.1), we follow the steps of Lacour [21].

**Case $r = 1$.**

The case where $r = 1$ provides a window $h_* = \frac{2(a + \gamma)}{\log n}$ and we get

$$
MISE = O \left( n^{-\frac{a}{a+\gamma}} \right).
$$

**Case $r < 1$.**

In this case, and in the case $r > 1$, following the ideas in [21], we will look for the bandwidth $h$ expressed as an expansion in $\log(n)$. In this expansion and when $r < 1$, the integer $k$ such that $\frac{k}{k+1} < r \leq \frac{k+1}{k+2}$ will play a role. The optimal bandwidth is of the form:

$$
h_* = 2\gamma \left( \log(n) + (r-1) \log \log(n) + \sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)} \right)^{-1},
$$

(E.2)

where the coefficients $b_i$’s are a sequence of real numbers chosen so that $\psi(h_*) = O(n)$. The heuristic of this expansion is as follows: the first term corresponds to the solution of $e^{2\gamma/h} = n$. The second term is added to compensate the factor $h^{r-1}$ in (E.1) evaluated with the previous bandwidth, and the third term aims at compensating the factor $e^{2a/h^r}$. Notice that $r-1 < 0$ and that the definition of $k$ implies that $r > r + (r - 1) > \cdots > r + k(r - 1) > 0 > r + (k + 1)(r - 1)$. This explains the range of the index $i$ in the sum of the right hand side of (E.2).
Plugging (E.2) into (E.1),
\[
\psi(h) = n (\log n)^{r-1} \exp \left( \sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)} \right) \\
\times \exp \left( \frac{2a}{(2\gamma)^r} \left( \log n \right)^r \left( 1 + \frac{(r-1) \log \log n + \sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)}}{\log n} \right)^r \right) \\
\times (2\gamma)^{-1} \left( \log n \right)^{-(r-1)} \left( 1 + \frac{(r-1) \log \log n + \sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)}}{\log n} \right)^{-(r-1)} \\
= (2\gamma)^{-1} n (1 + v_n)^{1-r} \exp \left( \sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)} \right) \\
\times \exp \left( \frac{2a}{(2\gamma)^r} \left( \log n \right)^r \left[ 1 + \sum_{j=0}^{k} \frac{r(r-1) \cdots (r-j)}{(j+1)!} v_n^{j+1} + o(v_n^{k+1}) \right] \right)
\]
where
\[
v_n = \frac{(r-1) \log \log n + \sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)}}{\log n} = (r-1) \log \log n + \sum_{i=0}^{k} b_i (\log n)^{i(r+1)(r-1)}
\]
converges to zero when \( n \to +\infty \). We note that
\[
v_n^{j+1} = \sum_{i=0}^{k} \sum_{p_0 + \cdots + p_j = i} b_{p_0} \cdots b_{p_j} (\log n)^{(i+j+1)(r-1)} + O \left( (\log n)^{(k+1)(r-1)} \right)
\]
\[
= \sum_{i=0}^{k} \sum_{p_0 + \cdots + p_j = i-j-1} b_{p_0} \cdots b_{p_j} (\log n)^{(i+j+1)(r-1)} + O \left( (\log n)^{(k+1)(r-1)} \right).
\]
So
\[
\psi(h) = (2\gamma)^{-1} n (1 + v_n)^{1-r} \exp \left( \sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)} \right) \\
\times \exp \left\{ \frac{2a}{(2\gamma)^r} \left( \log n \right)^r \left[ \sum_{i=1}^{\ell} \left( \frac{r(r-1) \cdots (r-j)}{(j+1)!} \sum_{p_0 + \cdots + p_j = i-j-1} b_{p_0} \cdots b_{p_j} \right) (\log n)^{r+i(r-1)} \right] \right\}
\]
\[
+ O \left( (\log n)^{(k+1)(r-1)} \right) \}
\]
\[
= (2\gamma)^{-1} n (1 + v_n)^{1-r} \exp \left( \sum_{i=0}^{k} M_i (\log n)^{(i+r-1)+r} + o(1) \right).
\]
The condition \( \psi(h) = O(n) \) implies the following choices of constants \( M_i \)'s:
\[
M_0 = b_0 + \frac{2a}{(2\gamma)^r}, \quad \forall i > 0, \quad M_i = b_i + \frac{2a}{(2\gamma)^r} \sum_{j=0}^{i-1} \frac{r(r-1) \cdots (r-j)}{(j+1)!} \sum_{p_0 + \cdots + p_j = i-j-1} b_{p_0} \cdots b_{p_j}.
\]
Since $h_\ast$ solves (E.1) if all the $M_i = 0$ for $i \in \{0, \cdots k\}$, the above system provides equation by equation the proper coefficients $b_i^\ast$.

$$b_0^\ast = -\frac{2a}{(2\gamma)^r}, \quad b_i^\ast = -\frac{2a}{(2\gamma)^r} \sum_{j=0}^{i-1} \frac{r(r-1)\cdots(r-j)}{(j+1)!} \sum_{p_0 + \cdots + p_j = i-j-1} b_{p_0}^\ast \cdots b_{p_j}^\ast.$$  \hspace{1cm} (E.3)

Replacing in (4.6), we get:

$$MISE = O(\exp\left\{ -\frac{2a}{(2\gamma)^r} \left[ \log n + (r-1)\log \log n + \sum_{i=0}^{k} b_i^\ast (\log n)^{r+i(r-1)} \right]^r \right\}).$$

**Case $r > 1$.**

Here, let us denote by $k$ the integer such that $\frac{k}{k+1} < \frac{1}{r} \leq \frac{k+1}{k+2}$. We look here for a bandwidth of the form:

$$h_\ast^k = 2a\left( \log n + \frac{r-1}{r} \log \log(n) + \sum_{i=0}^{k} d_i (\log n)^{\frac{1}{r}-i\frac{1}{r-1}} \right)^{-1}, \hspace{1cm} (E.4)$$

where the coefficients $d_i$'s will be chosen so that $\psi(h_\ast) = O(n)$.

Similar computations as for the case $r < 1$ provide that:

$$\psi(h_\ast) = (2a)^{\frac{-1}{r-1}} n \left( 1 + v_n \right)^{-\frac{1}{r-1}} \times \exp\left( \sum_{i=0}^{k} d_i (\log n)^{\frac{1}{r}-i\frac{1}{r-1}} \right)$$

$$\times \exp\left( \frac{2\gamma}{(2a)^{1/r}} (\log n)^{1/r} \left[ 1 + \sum_{\ell=1}^{k} \sum_{j=0}^{\ell-1} \sum_{p_0 + \cdots + p_j = \ell-j-1} \frac{1}{(j+1)!} (\frac{1}{r} - 1)\cdots(\frac{1}{r} - j) d_{p_0} \cdots d_{p_j} (\log n)^{\frac{1}{r-1}} + O((\log n)^{\frac{1}{r-1}}) \right] \right)$$

$$= (2a)^{\frac{-1}{r-1}} n (1 + v_n)^{-\frac{1}{r-1}} \exp\left( \sum_{i=0}^{k} M_i (\log n)^{\frac{1}{r}-i\frac{1}{r-1}} + o(1) \right)$$

where here

$$v_n = \frac{\frac{-1}{r-1} \log \log(n) + \sum_{i=0}^{k} d_i (\log n)^{\frac{1}{r}-i\frac{1}{r-1}}}{\log n},$$

and

$$M_0 = d_0 + \frac{2\gamma}{(2a)^{1/r}}, \quad \forall i > 0, \quad M_i = d_i + \frac{2\gamma}{(2a)^{1/r}} \sum_{j=0}^{i-1} \sum_{p_0 + \cdots + p_j = i-j-1} \frac{1}{(j+1)!} (\frac{1}{r} - 1)\cdots(\frac{1}{r} - j) d_{p_0} \cdots d_{p_j}.$$  \hspace{1cm} (E.5)

Solving $M_0 = \cdots = M_k = 0$ provides the coefficients $d_i^\ast$ so that (E.1) is satisfied.
Plugging the bandwidth $h_\ast$ with the coefficients $d_\ast^i$ into (4.6), we obtain:

\[
MISE = O\left(\frac{1}{n} \exp\left\{\frac{2\gamma}{(2a)^{1/r}} \left[ \log n + \frac{r - 1}{r} \log \log n + \sum_{i=0}^{k} d_\ast^i (\log n)^\frac{i}{r - i} \right] \right\}\right).
\]

This concludes the proof of Corollary 4.3.