Bernoulli **28**(2), 2022, 771–802
https://doi.org/10.3150/21-BEJ1366

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, 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 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 non-parametric 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.

Keywords: 35Q62; 65M32; 62G05; 46L53; 35R30; 60B20; 46L54

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 [17,37]), for the Burgers equation that is used in astrophysics (see [6,10,21,22] or also the survey by [36]), for the wave equations, to study the solutions with low-regularity initial data (see [11,12,35]) or for the Schrödinger PDE (see [9]). 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 [26,31]. Numerical approximations of such PDEs with random initial conditions have been considered in [32,34]. In this paper, we are interested in the Fokker-Planck PDE which is another case of McKean-Vlasov PDE [14]. This equation models the motion of particles with electrostatic repulsion and a probabilistic interpretation that we will adopt has been considered in [8].

1350-7265

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 [29,30]. 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 ν_t can be viewed as the distribution of $X_t = X_0 + B_t$ where X_0 is distributed as ν_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:

$$\partial_t p(t,x) = -\partial_x \int_{\mathbb{R}^2} Hp(t,x)p(t,x)\mathrm{d}x, \tag{1.1}$$

with

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

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 [29,30], 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 $(\mathbf{h}_t)_{t\geq 0}$ (operator-valued), and the usual convolution by a Gaussian distribution is replaced by the free convolution by a semi-circular distribution σ_t characterized by its density with respect to the Lebesgue measure:

$$\sigma_t(dx) = \frac{1}{2\pi t}\sqrt{4t-x^2}\mathbb{1}_{[-2\sqrt{t},2\sqrt{t}]}(x)\, dx. \tag{1.2}$$

If $\mathbf{x}_0$ admits the spectral measure μ_0, then $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{h}_t$ admits

$$\mu_t = \mu_0 \boxplus \sigma_t, \tag{1.3}$$

as spectral measure, where the operation $\boxplus$ is the free convolution and has been introduced by Voiculescu in [38]. It can be proved that the density $p(t,\cdot)$ of μ_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 μ is defined as $G_\mu(z) = \int_{\mathbb{R}}(z-x)^{-1}\mathrm{d}\mu(x)$ 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_{μ_t} by the functional equation

$$w_{fp}(z) = z + tG_{\mu_t}(w_{fp}(z)). \tag{1.4}$$

From this, we can recover G_{μ_0} with the formula $G_{\mu_0}(z) = G_{\mu_t}(w_{fp}(z))$ and thus p_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 μ_0 with the Cauchy distribution of parameter γ, defined by its density $f_\gamma(x) := \gamma/(\pi(x^2+\gamma^2))$, satisfies

$$f_{\mu_0 * \mathcal{C}_\gamma}(x) = \frac{1}{\pi t}\left[\gamma - \mathrm{Im} w_{fp}(x+i\gamma)\right], \quad x \in \mathbb{R}. \tag{1.5}$$

Then, estimating p_0, the density of μ_0, requires an estimation of the subordination function w_{fp} combined with a classical deconvolution step from a Cauchy distribution. Therefore, some statistical deconvolution tools will be needed and we should mention some recent advances on the topic, for instance [5,15,19] or [28].

Observations. Additionally to the free deconvolution problem, our observation does not consist in the operator-valued random variable $\mathbf{x}_t$ 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

$$X^n(t) = X^n(0) + H^n(t), \qquad t \geq 0 \tag{1.6}$$

with $X^n(0)$ a diagonal matrix whose entries are the ordered statistic $\lambda_1^n(0) < \cdots < \lambda_n^n(0)$ of a vector $(d_i^n)_{i\in\{1,\dots n\}}$ of n independent and identically distributed (i.i.d.) random variables distributed as $\mu_0(dx) = p_0(x)\,dx$, 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 $(\lambda_1^n(t), \dots, \lambda_n^n(t))$ 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 $(\lambda_1^n(t), \dots, \lambda_n^n(t))$ of $X^n(t)$ solve the following system of stochastic differential equations (SDE):

$$\mathrm{d}\lambda_i^n(t) = \frac{1}{\sqrt{n}}\mathrm{d}\beta_i(t) + \frac{1}{n}\sum_{j\neq i}\frac{\mathrm{d}t}{\lambda_i^n(t) - \lambda_j^n(t)}, \quad 1 \leq i \leq n, \tag{1.7}$$

where β_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 of the Supplementary Material [25] a short heuristic explanation on how this particle system arises. Note that the particle system (1.7) is only introduced for interpretational purpose and will not be used directly in the sequel. Now, if we denote by

$$\mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i^n(t)} \tag{1.8}$$

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 $\widehat{w}_{fp}^n(z)$ for the subordination function. As the Cauchy transform G_{μ_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 $\mathbb{C}_\gamma := \{z \in \mathbb{C}^+,\ \mathrm{Im}(z) > \gamma\}$. 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 $\widehat{w}^n_{fp}(z)$ 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 $\mathbb{C}_\gamma$:*

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

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

$$\sup_{n\geq 1} \frac{1}{n}\sum_{i=1}^n \log\left(\lambda_i^n(0)^2+1\right) < \infty \text{ almost surely (a.s.)}$$

in [1], Proposition 4.3.10, and ensures the convergence of μ^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 $\widehat{w}^n_{fp}$ with then a classical deconvolution step. We define our final estimator $\widehat{p}_{0,h}$ via its Fourier transform, denoted $\widehat{p}^\star_{0,h}$; from Equation (1.5), it is natural to define it as follows:

$$\widehat{p}^\star_{0,h}(\xi) = e^{\gamma|\xi|}.K^\star_h(\xi).\frac{1}{\pi t}\left[\gamma - \operatorname{Im}\widehat{w}^n_{fp}(\cdot + i\,\gamma)^\star(\xi)\right], \quad \xi \in \mathbb{R}.$$

Note that, as usual in nonparametric statistics, the last expression depends on $K^\star_h$, a regularization term defined through the Fourier transform of a kernel function K_h depending on a bandwidth parameter h. See Equation (2.15) in Definition 2.9 for more details.

We study theoretical properties of $\widehat{p}_{0,h}$ by deriving asymptotic rates of the mean integrated square error of $\widehat{p}_{0,h}$ 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 $\widehat{w}^n_{fp}(z) - w_{fp}(z)$ provided by Proposition 1.1. We show in Theorem 4.1 that the variance term is of order $e^{\frac{2\gamma}{h}}/n$ as desired for deconvolution with the Cauchy distribution with parameter γ. 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 $\int_{\mathbb{R}} |p_0^\star(\xi)|^2 \exp(2a|\xi|)\mathrm{d}\xi \leq L$ for $0 < L < \infty$, then $\mathbb{E}\Big[\|\widehat{p}_{0,h} - p_0\|^2\Big] = O\big(n^{-\frac{a}{a+\gamma}}\big)$. 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 γ. Note that the exponent in the previous bound reflects the difficulty of our statistical problem: the larger γ, the slower the rate. Remembering that γ 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 γ 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 $\widehat{p}_{0,h}$ 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 $a = \sqrt{(\sqrt{u^2+v^2}+u)/2}$ and $b = \sqrt{(\sqrt{u^2+v^2}-u)/2}$. We denote the Fourier transform of a function $g \in \mathbb{L}^1(\mathbb{R})$ by $g^\star : \xi \in \mathbb{R} \mapsto \int_{\mathbb{R}} g(x)e^{ix\xi}dx$.

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 $\left(B_{i,j}, \tilde{B}_{i,j}, 1 \leq i \leq j \leq n\right)$ 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 $\left\{(H^n(t))_{k,\ell}, t \geq 0, 1 \leq k, \ell \leq n\right\}$ equal to

$$(H^n)_{k,\ell} = \begin{cases} \dfrac{1}{\sqrt{2n}}\left(B_{k,\ell} + i\ \tilde{B}_{k,\ell}\right), & \text{if } k < \ell \\ \dfrac{1}{\sqrt{n}} B_{k,k}, & \text{if } k = \ell \end{cases} \tag{2.1}$$

Let us now define the initial condition, that we will choose independent of the Hermitian Brownian motion H^n. Recall that μ_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 $(\lambda_1^n(0), \dots, \lambda_n^n(0))$ the ordered statistics of i.i.d. random variables $(d_i^n)_{1 \leq i \leq n}$ with distribution μ_0.

For $t \geq 0$, let $\lambda^n(t) = \left(\lambda_1^n(t), \dots, \lambda_n^n(t)\right)$ 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 $\left(\lambda^n(t)\right)_{t \geq 0}$ is the unique solution in $C(\mathbb{R}_+, \mathbb{R}^n)$ of the system (1.7) with initial condition $\lambda_i^n(0)$ and where β_i are i.i.d. real valued standard Brownian motions. With probability one and for all $t > 0$, $\lambda_1^n(t) < \dots < \lambda_n^n(t)$.*

Moreover, we have for any fixed $T > 0$, the convergence of the process of empirical measures $(\mu_t^n)_{t \geq 0}$ as defined in (1.8), viewed as an element of $C\big([0, T], \mathcal{M}_1(\mathbb{R})\big)$, the space of continuous processes from $[0, T]$ into the space $\mathcal{M}_1(\mathbb{R})$ of probability measure on $\mathbb{R}$, equipped with its weak topology.

Proposition 2.3. *Under Assumption* (1.9), *for any fixed time* $T < \infty$, $(\mu_t^n)_{t\in[0,T]}$ *converges almost surely in* $\mathcal{C}([0, T], \mathcal{M}_1(\mathbb{R}))$. *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 [27].

2.2. Free deconvolution by subordination method

Our starting point is (1.3), for a fixed time $t > 0$. Recovering μ_0 knowing μ_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 μ be a probability measure on $\mathbb{R}$. The Cauchy transform of μ is defined by:

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

The fact is that $G_\mu(\overline{z}) = \overline{G_\mu(z)}$, so the behavior of the Cauchy transform in the lower half-plane $\mathbb{C}^- = \{z \in \mathbb{C} | \mathrm{Im}(z) < 0\}$ can be determined by its behavior in the upper half-plan $\mathbb{C}^+ = \{z \in \mathbb{C} | \mathrm{Im}(z) > 0\}$. The function G_μ is a bijection from a neighbourhood of infinity to a neighbourhood of zero (see [7] for example) and we can define the R-transform of μ by:

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

where $G_\mu^{<-1>}(z)$ is the inverse function of G_μ 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 μ_1 and μ_2,

$$R_{\mu_1 \boxplus \mu_2} = R_{\mu_1} + R_{\mu_2}. \tag{2.4}$$

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 [39]. We first introduce $F_\mu(z) = 1/G_\mu(z)$. As G_μ does not vanish on $\mathbb{C}^+$, F_μ is well defined on $\mathbb{C}^+$. Then:

Theorem-Def 2.5. *There exist unique subordination functions α_1 and α_2 from $\mathbb{C}^+$ onto $\mathbb{C}^+$ such that:*
(i) for $z \in \mathbb{C}^+$, $Im\big(\alpha_1(z)\big) \geq Im(z)$ and $Im\big(\alpha_2(z)\big) \geq Im(z)$,
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 $F_{\sigma_t}(z)$:

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

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

$$\mathbb{C}_\gamma = \left\{ z \in \mathbb{C}^+ \middle| \mathrm{Im}(z) > \gamma \right\}.$$

These domains will appear since G_μ is not invertible on the whole plane $\mathbb{C}$.

Theorem 2.6. *There exist unique subordination functions w_1 and w_{fp} from $\mathbb{C}_{2\sqrt{t}}$ onto $\mathbb{C}^+$ such that following properties are satisfied.*

(i) For $z \in \mathbb{C}_{2\sqrt{t}}$, $Im\big(w_1(z)\big) \geq \frac{1}{2} Im(z)$ and $Im\big(w_{fp}(z)\big) \geq \frac{1}{2} Im(z)$, and also $\lim_{y\to+\infty} w_1(iy)/(iy) = \lim_{y\to+\infty} w_{fp}(iy)/(iy) = 1$.

(ii) For $z \in \mathbb{C}_{2\sqrt{t}}$:

$$F_{\mu_0}(z) = F_{\sigma_t}(w_1(z)) = F_{\mu_t}(w_{fp}(z)). \tag{2.6}$$

(iii) For $z \in \mathbb{C}_{2\sqrt{t}}$:

$$w_{fp}(z) = z + w_1(z) - F_{\mu_0}(z). \tag{2.7}$$

(iv) Denote $h_{\sigma_t}(w) = w - F_{\sigma_t}(w) = t\ G_{\sigma_t}(w)$ and $\widetilde{h}_{\mu_t}(w) = w + F_{\mu_t}(w)$ on $\mathbb{C}^+$. We can define the function L_z as

$$L_z(w) := h_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z = t.G_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z. \tag{2.8}$$

For any $z \in \mathbb{C}_{2\sqrt{t}}$, we have

$$L_z\big(w_{fp}(z)\big) = w_{fp}(z), \tag{2.9}$$

and for all w such that $Im(w) > \frac{1}{2} Im(z)$, the iterated function $L_z^{\circ m}(w)$ converges to $w_{fp}(z) \in \mathbb{C}^+$ when $m \to +\infty$.

One difference between Theorem 2.6 and Theorem-definition 2.5 lies in the fact that the subordination functions are expressed in terms of $F_{\mu_0 \boxplus \sigma_t}$ and F_{σ_t} whereas in Theorem-definition 2.5 it would have been F_{μ_0} and F_{σ_t}. Here the restriction to the domain $\mathbb{C}_{2\sqrt{t}}$ comes from the fact that $\mathrm{Im}(\widetilde{h}_{\mu_t}(w) - z)$ appearing in the definition (2.8) of L_z has to be positive. It is worth pointing out that the constant $2\sqrt{t}$ in $\mathbb{C}_{2\sqrt{t}}$ 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 $2\sqrt{2t}$. This improvement has a key impact on the convergence rates through the constant γ (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 of the Supplementary Material [25]. 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 μ_0 from w_{fp}, as stated in the following lemma proved at the end of the section:

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

$$G_{\mu_0}(z) = \frac{1}{t}(w_{fp}(z) - z) = G_{\mu_t}\big(w_{fp}(z)\big). \tag{2.10}$$

Consequently,

$$\big|w_{fp}(z) - z\big| \le \sqrt{t}. \tag{2.11}$$

Moreover, denoting $\mathcal{C}_\gamma$ the centered Cauchy distribution with parameter $\gamma > 0$, one can check that, for any probability measure μ on $\mathbb{R}$, the density $f_{\mu * \mathcal{C}_\gamma}$ of the classical convolution of μ by $\mathcal{C}_\gamma$ is given, for $x \in \mathbb{R}$, by $f_{\mu*\mathcal{C}_\gamma}(x) = -\mathrm{Im} G_\mu(x + i\gamma)/\pi$. Using the expression of G_{μ_0} given by Lemma 2.7 with $\gamma > 2\sqrt{t}$, we get that for any $x \in \mathbb{R}$,

$$f_{\mu_0 * \mathcal{C}_\gamma}(x) = \frac{1}{\pi t}\big[\gamma - \mathrm{Im} w_{fp}(x + i\gamma)\big]. \tag{2.12}$$

From this, we can recover the density p_0 of μ_0 by a classical deconvolution of (2.12) by f_γ (see for instance [13,16] or [24] for Cauchy statistical deconvolution problems or [23] for regularization of Fredholm equations of the first kind). The subordination function w_{fp} in (2.12) is estimated using the second equality of Lemma 2.7. In parallel with our work, Tarrago [33] 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 $\widehat{w}^n_{fp}$ 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 $z \in \mathbb{C}_{2\sqrt{t}}$,

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

So, we obtain

$$G_{\sigma_t}\big(w_1(z)\big) = \frac{1}{t}\big(w_{fp}(z) - z\big).$$

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

$$\big|w_{fp}(z) - z\big| = t.\big|G_{\sigma_t}(w_1(z))\big| = t.\big|G_{\mu_t}(w_{fp}(z))\big| \le \frac{t}{|\mathrm{Im}(w_{fp}(z))|} \le \sqrt{t}, \qquad \square$$

Estimator of p_0. We do not observe directly the measure μ_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:

$$\widehat{G}_{\mu_t^n}(z) := \int_{\mathbb{R}} \frac{d\mu_t^n(\lambda)}{z-\lambda} = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{z-\lambda_j^n(t)} = \frac{1}{n}\mathrm{tr}\Big(\big(zI_n - X_n(t)\big)^{-1}\Big). \tag{2.13}$$

will lead to the following:

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

$$\frac{1}{t}(w(z)-z) = \widehat{G}_{\mu_t^n}(w(z)), \quad \textit{for } z \in \mathbb{C}_{2\sqrt{t}} \tag{2.14}$$

This fixed-point is denoted by $\widehat{w}^n_{fp}(z)$. We have $Im(\widehat{w}^n_{fp}(z)) > Im(z)/2$ and $\left|\widehat{w}^n_{fp}(z) - z\right| \le \sqrt{t}$.

The theorem is proved at the end of this section. We shall prove in Section 3 that $\widehat{w}^n_{fp}(z)$ 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 $\widehat{w}^n_{fp}(z)$.

Recall that the Fourier transform of the Cauchy distribution $\mathcal{C}_\alpha$ with $\alpha > 0$ is $f_\alpha^\star(\xi) = e^{-\alpha|\xi|}$ 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 $f_\gamma^\star(\xi)$ 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 $K^\star$ is bounded by a positive constant $C_K < +\infty$ and has a compact support, say $[-1, 1]$. We define the estimator $\widehat{p}_{0,h}$ of p_0 by its Fourier transform:

$$\widehat{p}^\star_{0,h}(\xi) = e^{\gamma|\xi|}.K_h^\star(\xi).\frac{1}{\pi t}\Big[\gamma - \mathrm{Im}\,\widehat{w}^n_{fp}(\cdot + i\,\gamma)^\star(\xi)\Big], \tag{2.15}$$

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

Note that the assumption on K ensures finiteness of the estimator. These assumptions are for instance satisfied for the sinc kernel, namely $K(x) = \mathrm{sinc}(x) = \sin(x)/(\pi x)$ with Fourier transform $K^\star(\xi) = \mathbb{1}_{[-1,1]}(\xi)$ 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 σ_t, one can use the explicit expressions of G_{σ_t} or F_{σ_t} to improve the constants. The detailed proof of the two theorems are postponed to Appendix B of the Supplementary Material [25].

We start with sketching the proof of Theorem 2.6. Let us define the function L_z as in Equation (2.8): $L_z(w) := h_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z$. One can check that for any $z \in \mathbb{C}_{2\sqrt{t}}$, L_z is well defined on $\mathbb{C}_{\frac{1}{2}\mathrm{Im}z}$ and that it satisfies the assumptions of the Denjoy-Wolff fixed-point theorem, namely that $L_z(\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}) \subset \overline{\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}}$ 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 $\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$, denoted by $w_{fp}(z)$ and point (iv) is proved. We then define $w_1(z) := F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) - z$, check that $F_{\sigma_t}(w_1(z)) = F_{\mu_t}(w_{fp}(z))$ 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 $\frac{1}{t}(w_{fp}(z)-z) = G_{\mu_t}\big(w_{fp}(z)\big)$. It is therefore natural to define an estimator for $w_{fp}(z)$ by replacing $G_{\mu_t}(z)$ by its estimator $\widehat{G}_{\mu_t^n}(z)$. Theorem-Definition 2.8 will be obtained along similar arguments as above, applying the Denjoy-Wolff fixed point theorem to $\widehat{L}_z(w) := t\widehat{G}_{\mu_t^n}(w) + z$.

3. Study of the subordination function

This section is devoted to the proof of Proposition 1.1. We show that $\widehat{w}^n_{fp}(z)$ 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 μ on $\mathbb{R}$ and $\alpha > 0$, the Cauchy transform G_μ is Lipschitz on $\mathbb{C}_\alpha$ with Lipschitz constant $\dfrac{1}{\alpha^2}$, and one has for any $z \in \mathbb{C}_\alpha$, $\big|G_\mu(z)\big| \le \dfrac{1}{\alpha}$.*

Proof. For $z, z' \in \mathbb{C}_\alpha$,

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

We also have

$$\big|G_\mu(z)\big| = \left|\int_{\mathbb{R}} \frac{d\mu(x)}{z-x}\right| \le \frac{1}{\mathrm{Im}(z)} \le \frac{1}{\alpha}. \qquad \square$$

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 $\widehat{w}^n_{fp}(z)$, we have

$$\begin{aligned}\left|\widehat{w}^n_{fp}(z) - w_{fp}(z)\right| &= t\left|\widehat{G}_{\mu_t^n}(\widehat{w}^n_{fp}(z)) - G_{\mu_t}(w_{fp}(z))\right| \\ &\le t\left|\widehat{G}_{\mu_t^n}(\widehat{w}^n_{fp}(z)) - \widehat{G}_{\mu_t^n}(w_{fp}(z))\right| + t\left|\widehat{G}_{\mu_t^n}(w_{fp}(z)) - G_{\mu_t}(w_{fp}(z))\right|. \end{aligned} \tag{3.1}$$

By Theorem 2.6, $\mathrm{Im}(w_{fp}(z)) \ge \dfrac{1}{2}\mathrm{Im}(z)$ and since $\widehat{G}_{\mu_t^n}$ is a Lipschitz function on $\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$ with Lipschitz constant $\dfrac{4}{\mathrm{Im}^2(z)} \le \frac{4}{\gamma^2}$, by Lemma 3.1, we have an upper bound for the first term

$$\left|\widehat{G}_{\mu_t^n}(\widehat{w}^n_{fp}(z)) - \widehat{G}_{\mu_t^n}(w_{f_p}(z))\right| \le \frac{4}{\gamma^2} \times \left|\widehat{w}^n_{fp}(z) - w_{fp}(z)\right|.$$

Thus,

$$\left|\widehat{w}^n_{fp}(z) - w_{fp}(z)\right| \leq \frac{4t}{\gamma^2}\left|\widehat{w}^n_{fp}(z) - w_{fp}(z)\right| + t\left|\widehat{G}_{\mu^n_t}\big(w_{fp}(z)\big) - G_{\mu_t}\big(w_{fp}(z)\big)\right|,$$

implying that

$$\left|\widehat{w}^n_{fp}(z) - w_{fp}(z)\right| \leq \left(\frac{t\gamma^2}{\gamma^2 - 4t}\right) \times \left|\widehat{G}_{\mu^n_t}\big(w_{fp}(z)\big) - G_{\mu_t}\big(w_{fp}(z)\big)\right|. \tag{3.2}$$

By Proposition 2.3, since the function $x \mapsto \frac{1}{z-x}$ is continuous and bounded on $\mathbb{R}$ for any $z \in \mathbb{C}_{\sqrt{t}}$, $\widehat{G}_{\mu^n_t}(w_{fp}(z)) = \int_{\mathbb{R}} \frac{1}{w_{fp}(z)-x}\mu^n_t(\mathrm{d}x)$ converges almost surely to $G_{\mu_t}(w_{fp}(z)) = \int_{\mathbb{R}} \frac{1}{w_{fp}(z)-x}\mu_t(\mathrm{d}x)$. 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 $\widehat{G}_{\mu^n_t}(z)$ 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 $\widehat{w}^n_{fp}(z)$ 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:

$$\begin{aligned}
\widehat{G}_{\mu^n_t}(z) - G_{\mu_t}(z) &= \widehat{G}_{\mu^n_t}(z) - \mathbb{E}\big[\widehat{G}_{\mu^n_t}(z)|X^n(0)\big] + \mathbb{E}\big[\widehat{G}_{\mu^n_t}(z)|X^n(0)\big] - G_{\mu^n_0\boxplus\sigma_t}(z)\\
&\quad + G_{\mu^n_0\boxplus\sigma_t}(z) - G_{\mu_t}(z)\\
&=: A^n_1(z) + A^n_2(z) + A^n_3(z).
\end{aligned} \tag{3.3}$$

The first term is related to the variance of $\widehat{G}_{\mu^n_t}(z)$ (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 $A^n_1(z)$, the result is a direct consequence of Proposition 3 in [18] and we refer to the detailed computation in [27].

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

$$Var\big(nA^n_1(z)|X^n(0)\big) = Var\big(n\widehat{G}_{\mu^n_t}(z)|X^n(0)\big) \leq \frac{10t}{Im^4(z)}.$$

3.2.1. *Fluctuations of $A^n_2(z)$*

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

$$R_{n,t}(z) := \big(zI_n - X^n(t)\big)^{-1}. \tag{3.4}$$

Then one can write

$$\widehat{G}_{\mu^n_t}(z) = \frac{1}{n}\operatorname{Tr}\big(R_{n,t}(z)\big).$$

Then, the bias term is:

$$nA_2^n(z) = \mathbb{E}\left[\mathrm{Tr}\left(R_{n,t}(z)\right) \mid X^n(0)\right] - nG_{\mu_0^n \boxplus \sigma_t}(z), \tag{3.5}$$

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 $n \in \mathbb{N}$,*

$$\left|nA_2^n(z)\right| \leq \left(1 + \frac{4t}{Im^2(z)}\right) \cdot \left(\frac{2t}{Im^3(z)} + \frac{12t^2}{Im^5(z)}\right). \tag{3.6}$$

The term $A_2^n(z)$ compares $\mathbb{E}\left[\widehat{G}_{\mu_t^n}(z)|X^n(0)\right]$ with $G_{\mu_0^n \boxplus \sigma_t}(z)$. Proceeding as in Theorem-Definition 2.5, with μ_0^n and σ_t, we can define a subordination function $\overline{w}_{fp}(z)$ such that

$$G_{\mu_0^n \boxplus \sigma_t}(z) = G_{\mu_0^n}\left(\overline{w}_{fp}(z)\right). \tag{3.7}$$

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

$$|nA_2^n(z)| \leq 2n\mathrm{Im}^{-1}(z), \tag{3.8}$$

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 $A_2^n(z)$:

$$\begin{aligned} \widetilde{R}_{n,t}(z) &:= \left(\left(z - \frac{t}{n}\mathbb{E}\left[\mathrm{Tr}\left(R_{n,t}(z)\right) \mid X^n(0)\right]\right).I_n - X^n(0)\right)^{-1} \\ n\widetilde{A}_2^n(z) &:= \mathbb{E}\left[\mathrm{Tr}\left(R_{n,t}(z)\right) \mid X^n(0)\right] - \mathrm{Tr}\left(\widetilde{R}_{n,t}(z)\right). \end{aligned} \tag{3.9}$$

We will bound $A_2^n(z)$ by using its approximation $\widetilde{A}_2^n(z)$.

Step 1: First, we prove an upper bound for $\widetilde{A}_2^n(z)$, whose proof is postponed to Appendix C of the Supplementary Material [25]:

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

$$\left|n\widetilde{A}_2^n(z)\right| \leq \frac{2t}{Im^3(z)} + \frac{12t^2}{Im^5(z)}.$$

Step 2: If $|\widetilde{A}_2^n(z)| \geq \mathrm{Im}(z)/(2t)$ then, by (3.8)

$$|nA_2^n(z)| \leq \frac{4tn|\widetilde{A}_2^n(z)|}{\mathrm{Im}^2(z)},$$

and we conclude with Lemma 3.4.

Step 3: We now consider the case where $|\widetilde{A}_2^n(z)| < \mathrm{Im}(z)/(2t)$. We have:

$$A_2^n(z) = \widetilde{A}_2^n(z) + \left[A_2^n(z) - \widetilde{A}_2^n(z)\right] \tag{3.10}$$

We will control the difference $|A_2^n(z) - \widetilde{A}_2^n(z)|$ by $\widetilde{A}_2^n(z)$ and conclude with Lemma 3.4.

By their definitions:

$$n\big(A_2^n(z)-\widetilde{A}_2^n(z)\big)=\mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big)-nG_{\mu_0^n\boxplus\sigma_t}(z). \tag{3.11}$$

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 $\mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big)$ of (3.11) similarly. As $\widetilde{R}_{n,t}(z)$ is a diagonal matrix,

$$\mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big)=\sum_{j=1}^{n}\frac{1}{z-\frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}(z)\big)\mid X^n(0)\big]-\lambda_j^n(0)}=nG_{\mu_0^n}(\widetilde{w}_{fp}(z)), \tag{3.12}$$

where

$$\widetilde{w}_{fp}(z):=z-\frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}(z)\big)\mid X^n(0)\big] \tag{3.13}$$

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

$$A_2^n(z)-\widetilde{A}_2^n(z)=G_{\mu_0^n}(\widetilde{w}_{fp}(z))-G_{\mu_0^n}(\overline{w}_{fp}(z)). \tag{3.14}$$

To continue, we first need the following result proved in Appendix D of the Supplementary Material [25].

Lemma 3.5 (i). *The function $\overline{w}_{fp}(z)$, defined in (3.7), solves*

$$\overline{w}_{fp}(z)=z-tG_{\mu_0^n\boxplus\sigma_t}(z).$$

(ii) The function $\zeta(z)=z+tG_{\mu_0^n}(z)$ is well-defined on $\mathbb{C}^+$ and is the inverse of $\overline{w}_{fp}(z)$ on $\overline{\Omega}=\{z\in\mathbb{C}^+,\ Im(\zeta(z))>0\}$. For such $z\in\overline{\Omega}$, we denote this function $\overline{w}_{fp}^{<-1>}(z)$.

Let us prove that under the condition of Step 3, $\widetilde{w}_{fp}(z)\in\overline{\Omega}$ for all $z\in\mathbb{C}^+$.

$$\begin{aligned}\zeta(\widetilde{w}_{fp}(z))-z&=\widetilde{w}_{fp}(z)+tG_{\mu_0^n}\big(\widetilde{w}_{fp}(z)\big)-z\\&=z-\frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}(z)\big)\mid X^n(0)\big]+tG_{\mu_0^n}\big(\widetilde{w}_{fp}(z)\big)-z=-t\widetilde{A}_2^n(z),\end{aligned} \tag{3.15}$$

by (3.12). Therefore,

$$\big|\mathrm{Im}\big(\zeta(\widetilde{w}_{fp}(z))\big)-\mathrm{Im}(z)\big|\le\big|\zeta(\widetilde{w}_{fp}(z))-z\big|=t\big|\widetilde{A}_2^n(z)\big|\le\frac{\mathrm{Im}(z)}{2}. \tag{3.16}$$

Thus, under the condition of Step 3, $\widetilde{w}_{fp}(z)\in\overline{\Omega}$. Denoting $\widetilde{z}=\overline{w}_{fp}^{<-1>}\big(\widetilde{w}_{fp}(z)\big)$, which is well-defined, we have $\overline{w}_{fp}(\widetilde{z})=\widetilde{w}_{fp}(z)$. Plugging this into (3.14),

$$\begin{aligned}A_2^n(z)-\widetilde{A}_2^n(z)&=G_{\mu_0^n\boxplus\sigma_t}(\widetilde{z})-G_{\mu_0^n\boxplus\sigma_t}(z)\\&=(z-\widetilde{z})\int_{\mathbb{R}}\frac{\mu_0^n\boxplus\sigma_t(dx)}{(\widetilde{z}-x).(z-x)}=t\widetilde{A}_2^n(z).\int_{\mathbb{R}}\frac{\mu_0^n\boxplus\sigma_t(dx)}{(\widetilde{z}-x).(z-x)},\end{aligned}$$

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

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

This concludes the proof of Proposition 3.3. □

3.2.2. *Fluctuations of* $A_3^n(z)$

Finally, the third step is to control $A_3^n(z) = G_{\mu_0^n \boxplus \sigma_t}(z) - G_{\mu_t}(z)$, with $\mu_t = \mu_0 \boxplus \sigma_t$.

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

$$\left|A_3^n(z)\right| \leq \frac{\gamma^2}{\gamma^2 - 4t}\left|\int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x}\left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right]\right| \tag{3.17}$$

and

$$\sup_{n\in\mathbb{N}} \sup_{z \in \mathbb{C}_{\frac{\gamma}{2}}} \mathbb{E}\left[n\left|A_3^n(z)\right|^2\right] \leq \frac{8\gamma^2}{(\gamma^2 - 4t)^2}. \tag{3.18}$$

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

$$G_{\mu_0^n \boxplus \sigma_t}(z) = G_{\mu_0^n}\left(\overline{w}_{fp}(z)\right) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0^n(x)}{\overline{w}_{fp}(z) - x} = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_0^n(x)}{z - tG_{\mu_0^n \boxplus \sigma_t}(z) - x}. \tag{3.19}$$

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

$$|z - tG_{\mu_0^n \boxplus \sigma_t}(z) - x| \geq \frac{\gamma}{2} \geq \sqrt{t}, \tag{3.20}$$

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

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

Then, we can write

$$\begin{aligned} G_{\mu_0^n \boxplus \sigma_t}(z) - G_{\mu_0 \boxplus \sigma_t}(z) = {} & t.\int_{\mathbb{R}} \frac{G_{\mu_0^n \boxplus \sigma_t}(z) - G_{\mu_0 \boxplus \sigma_t}(z)}{\left(z - t.G_{\mu_0^n \boxplus \sigma_t}(z) - x\right).\left(z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x\right)} \mathrm{d}\mu_0^n(x) \\ & + \int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x}\left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right]. \end{aligned}$$

Thus,

$$\begin{aligned} \left(G_{\mu_0^n \boxplus \sigma_t}(z) - G_{\mu_0 \boxplus \sigma_t}(z)\right).\left[1 - t.\int_{\mathbb{R}} \frac{1}{\left(z - t.G_{\mu_0^n \boxplus \sigma_t}(z) - x\right).\left(z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x\right)} \mathrm{d}\mu_0^n(x)\right] \\ = \int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x}\left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right]. \end{aligned}$$

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

$$\left| t.\int_{\mathbb{R}} \frac{1}{\left(z - t.G_{\mu_0^n\boxplus\sigma_t}(z) - x\right).\left(z - t.G_{\mu_0\boxplus\sigma_t}(z) - x\right)} \mathrm{d}\mu_0^n(x) \right| \leq \frac{4t}{\gamma^2},$$

consequently,

$$\left|A_3^n(z)\right| \leq \frac{\gamma^2}{\gamma^2 - 4t} \left| \int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0\boxplus\sigma_t}(z) - x} \left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right] \right|,$$

which gives the first part of the proposition. For the second part (3.18),

$$\mathbb{E}\left[n \left|A_3^n(z)\right|^2\right] \leq \left(\frac{\gamma^2}{\gamma^2 - 4t}\right)^2 n\mathbb{E}\left[\left| \int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0\boxplus\sigma_t}(z) - x} \left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right] \right|^2\right].$$

If $\mathrm{Im}(z) > \frac{\gamma}{2}$, the function $\varphi_z(x) = (z - t.G_{\mu_0\boxplus\sigma_t}(z) - x)^{-1}$ is bounded by $2/\gamma$. Then, for any $z \in \mathbb{C}_{\frac{\gamma}{2}}$:

$$\begin{aligned} n\mathbb{E}\left[\left| \int_{\mathbb{R}} \varphi_z(x)\mathrm{d}\mu_0^n(x) - \int_{\mathbb{R}} \varphi_z(x)\mathrm{d}\mu_0(x) \right|^2\right] &= n\mathbb{E}\left[\left| \frac{1}{n}\sum_{j=1}^n \varphi_z\left(\lambda_j^n(0)\right) - \mathbb{E}\left[\varphi_z\left(\lambda_j^n(0)\right)\right] \right|^2\right] \\ &= n\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^n \varphi_z\left(d_j^n\right)\right) = \int_{\mathbb{R}} |\varphi_z(x)|^2 \mathrm{d}\mu_0(x) - \left| \int_{\mathbb{R}} \varphi_z(x)\mathrm{d}\mu_0(x) \right|^2 \leq \frac{8}{\gamma^2}. \qquad \square \end{aligned}$$

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 $z \in \mathbb{C}_{\gamma/2}$:

$$\mathbb{E}\left[|\widehat{G}_{\mu_t^n}(z) - G_{\mu_t}(z)|^2 \mid X^n(0)\right] \leq C(\gamma, t)\left(\frac{1}{n^2} + \left| \int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0\boxplus\sigma_t}(z) - x} \left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right] \right|^2\right), \tag{3.22}$$

where $C(\gamma, t)$ depends only on γ and t. The proof also shows that $\gamma \mapsto C(\gamma, t)$ is bounded when $\gamma \to +\infty$ and $\gamma \mapsto (\gamma^2 - 4t)^2 \times C(\gamma, t)$ is bounded when $\gamma \to 2\sqrt{t}$. Therefore there exists a constant $C(t)$ only depending on t such that

$$\sup_{n\in\mathbb{N}} \sup_{z\in\mathbb{C}_{\gamma/2}} n\mathbb{E}\left[|\widehat{G}_{\mu_t^n}(z) - G_{\mu_t}(z)|^2\right]] = \sup_{n\in\mathbb{N}} \sup_{z\in\mathbb{C}_{\gamma/2}} n\mathbb{E}\left[\mathbb{E}\left[|\widehat{G}_{\mu_t^n}(z) - G_{\mu_t}(z)|^2 \mid X^n(0)\right]\right] \leq \frac{C(t)\gamma^4}{(\gamma^2 - 4t)^2},$$

by using the second part of Proposition 3.6. Equation (3.2) implies that for any $\gamma > 2\sqrt{t}$,

$$\sup_{n\in\mathbb{N}} \sup_{z\in\mathbb{C}_\gamma} \mathbb{E}\left[n\left|\widehat{w}_{fp}^n(z) - w_{fp}(z)\right)\right|^2\right] \leq \left(\frac{t\gamma^2}{\gamma^2 - 4t}\right)^2 \sup_{n\in\mathbb{N}} \sup_{z\in\mathbb{C}_{\frac{\gamma}{2}}} n\mathbb{E}\left[|\widehat{G}_{\mu_t^n}(z) - G_{\mu_t}(z)|^2\right] < +\infty,$$

since $z \in \mathbb{C}_\gamma$ implies that $\mathrm{Im}\left(w_{fp}(z)\right) \geq \dfrac{1}{2}\mathrm{Im}(z) > \frac{\gamma}{2}$ (Theorem 2.6) so that $w_{fp}(z) \in \mathbb{C}_{\frac{\gamma}{2}}$ 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 $\mathbb{E}\big[\|\widehat{p}_{0,h} - p_0\|^2\big]$, the mean integrated squared error of $\widehat{p}_{0,h}$, relying on the classical bias-variance decomposition of the quadratic risk. By Parseval's equality, we obtain:

$$\left\|\widehat{p}_{0,h} - p_0\right\|^2 = \frac{1}{2\pi}\left\|\widehat{p}^{\star}_{0,h} - p^{\star}_0\right\|^2 \le \frac{1}{\pi}\left\|\widehat{p}^{\star}_{0,h} - K^{\star}_h.p^{\star}_0\right\|^2 + \frac{1}{\pi}\left\|K^{\star}_h.p^{\star}_0 - p^{\star}_0\right\|^2. \tag{4.1}$$

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*

$$\Sigma := \left\|\widehat{p}^{\star}_{0,h} - K^{\star}_h p^{\star}_0\right\|^2. \tag{4.2}$$

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

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

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

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

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 $e^{\frac{2\gamma}{h}}$ 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 γ the better the rate of convergence but if $\gamma \to 2\sqrt{t}$ 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 $\mathcal{S}_s(a, r, L)$ of supersmooth densities defined for $a > 0$, $L > 0$ and $r > 0$ by:

$$\mathcal{S}_s(a, r, L) = \left\{p \text{ density such that } \int_{\mathbb{R}} |p^{\star}(\xi)|^2 e^{2a|\xi|^r}\,\mathrm{d}\xi \le L\right\}. \tag{4.5}$$

In the literature, this smoothness class of densities has often been considered (see [13,16,24]). Most famous examples of supersmooth densities are the Cauchy distribution belonging to $\mathcal{S}_s(a, r, L)$ with $r = 1$ and the Gaussian distribution belonging to $\mathcal{S}_s(a, r, L)$ with $r = 2$. To control the bias, we rely on Proposition 1 in [13] which states that:

Proposition 4.2. *For $p_0 \in \mathcal{S}_s(a, r, L)$, we have*:

$$\left\| K_h^{\star}.p_0^{\star} - p_0^{\star} \right\|^2 \leq Le^{-2ah^{-r}}.$$

Now, using similar computations to those in [24], we obtain from Proposition 4.2 and Theorem 4.1 the rates of convergence of our estimator $\widehat{p}_{0,h}$. We indeed showed that:

$$MISE := \mathbb{E}\Big[\left\| \widehat{p}_{0,h} - p_0 \right\|^2 \Big] \leq Le^{-2ah^{-r}} + \frac{\gamma^8}{(\gamma^2 - 4t)^4} \frac{C_{\mathrm{var}}(t).e^{\frac{2\gamma}{h}}}{n}. \tag{4.6}$$

Minimizing in h the right hand side of (4.6) provides the convergence rate of the estimator $\widehat{p}_{0,h}$. The rates of convergence are summed up in the following corollary, adapted from the computation of [24]. 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 of the Supplementary Material [25].

Corollary 4.3. *Suppose that μ_0 satisfies Assumption* (4.3) *and the density p_0 belongs to the space $\mathcal{S}_s(a, r, L)$ 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) *in the Supplementary Material* [25], *we have*:

$$\mathbb{E}\Big[\|\widehat{p}_{0,h} - p_0\|^2 \Big]$$
$$= \begin{cases} O\big(n^{-\frac{a}{a+\gamma}}\big) & \text{if } r = 1 \\ O\Big(\exp\Big\{ -\dfrac{2a}{(2\gamma)^r} \Big[\log n + (r-1)\log\log n + \displaystyle\sum_{i=0}^{k} b_i^*(\log n)^{r+i(r-1)} \Big]^r \Big\} \Big) & \text{if } r < 1 \\ O\Big(\dfrac{1}{n} \exp\Big\{ \dfrac{2\gamma}{(2a)^{1/r}} \Big[\log n + \dfrac{r-1}{r} \log\log n + \displaystyle\sum_{i=0}^{k} d_i^*(\log n)^{\frac{1}{r} - i\frac{r-1}{r}} \Big]^{1/r} \Big\} \Big) & \text{if } r > 1, \end{cases} \tag{4.7}$$

where the integer k is such that

$$\frac{k}{k+1} < \min\Big(r, \frac{1}{r}\Big) \leq \frac{k+1}{k+2},$$

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

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

$$d_0^* = -\tfrac{2\gamma}{(2a)^{1/r}}, \qquad \forall i > 0,\ d_i^* = -\frac{2\gamma}{(2a)^{1/r}} \sum_{j=0}^{i-1} \frac{\frac{1}{r}\big(\frac{1}{r}-1\big)\cdots\big(\frac{1}{r}-j\big)}{(j+1)!} \sum_{p_0+\cdots p_j = i-j-1} d_{p_0}^* \cdots d_{p_j}^*$$

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 equations (E.2) and (E.4) in the Supplementary Material [25], and also [24]).

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 $\mathcal{C}_\gamma$. The parameter γ represents the difficulty of our deconvolution problem and consequently, the rates of convergence heavily depend on γ. The larger γ 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 γ has to be chosen larger than $2\sqrt{t}$, γ and therefore the difficulty of the deconvolution problem grows with t accordingly. It remains an open question if we can take γ smaller.

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

$$\mathcal{S}_b(\beta, L) = \left\{ p \text{ density such that } \int_{\mathbb{R}} |p^\star(\xi)|^2 (1+\xi^2)^\beta d\xi \le L \right\}.$$

We have the following classical estimate for the integrated bias (see e.g. [16], Proposition 3).

Proposition 4.4. *For $p_0 \in \mathcal{S}_b(\beta, L)$ we have*:

$$\left\| K_h^\star . p_0^\star - p_0^\star \right\|^2 \le L h^{2\beta}. \tag{4.8}$$

Using Theorem 4.1, we obtain the following result.

Corollary 4.5. *Suppose that μ_0 satisfies Assumption* (4.3) *and the density p_0 belongs to the space $\mathcal{S}_b(\beta, L)$ 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*:

$$\mathbb{E}\left[\left\| \widehat{p}_{0,h} - p_0 \right\|^2 \right] = O\left((\log n)^{-2\beta} \right). \tag{4.9}$$

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 [13] for supersmooth densities or in Fan and Koo [20] for Sobolev regularities. Here, our estimation strategy converts the initial free deconvolution problem into the deconvolution problem (2.12) between μ_0 and the Cauchy distribution $\mathcal{C}_\gamma$. Thus, our observation scheme is more intricate and involved than the framework of classical density deconvolution tackled in [13] and [20]. If our observations had been distributed according to the density $f_{\mu_0 \star \mathcal{C}_\gamma}$ as in [13] and [20], for a given γ, 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 γ.

4.2. Proof of Theorem 4.1

Recall the definition of Σ in (4.2). By the definition of $\widehat{p}^\star_{0,h}$:

$$\Sigma = \int_{\mathbb{R}} \frac{1}{\pi^2 t^2} e^{2\gamma|\xi|} . \left| K_h^\star(\xi) \right|^2 . \left| \left[\left(\mathrm{Im}\left(\widehat{w}^n_{fp}(\cdot + i\gamma) \right) \right)^\star - \left(\mathrm{Im}\left(w_{fp}(\cdot + i\gamma) \right) \right)^\star \right](\xi) \right|^2 \mathrm{d}\xi.$$

Recall that by Lemma 2.7, we have $\mathrm{Im}\big(w_{fp}(z)\big) = t.\mathrm{Im}\big(G_{\mu_t}\big(w_{fp}(z)\big)\big) + \mathrm{Im}(z)$, and similarly by Theorem-Definition 2.8, $\mathrm{Im}\big(\widehat{w}^n_{fp}(z)\big) = t.\mathrm{Im}\big(\widehat{G}_{\mu^n_t}\big(\widehat{w}^n_{fp}(z)\big)\big) + \mathrm{Im}(z)$ for $z \in \mathbb{C}_{2\sqrt{t}}$. Since $K^\star_h(\xi) = K^\star(h\xi)$, we have

$$\begin{aligned}
\Sigma &= \int_{\mathbb{R}} e^{2\gamma|\xi|}.\left|K^\star_h(\xi)\right|^2.\frac{1}{\pi^2}\left|\Big(\mathrm{Im}\widehat{G}_{\mu^n_t}\big(\widehat{w}^n_{fp}(\centerdot+i\gamma)\big) - \mathrm{Im}G_{\mu_t}\big(w_{fp}(\centerdot+i\gamma)\big)\Big)^\star(\xi)\right|^2 \mathrm{d}\xi \\
&\le e^{\frac{2\gamma}{h}}.\frac{C_K^2}{\pi^2}.\left\|\Big(\mathrm{Im}\widehat{G}_{\mu^n_t}\big(\widehat{w}^n_{fp}(\centerdot+i\gamma)\big) - \mathrm{Im}G_{\mu_t}\big(w_{fp}(\centerdot+i\gamma)\big)\Big)^\star\right\|^2 \\
&= \frac{2C_K^2}{\pi}.e^{\frac{2\gamma}{h}}.\left\|\mathrm{Im}\widehat{G}_{\mu^n_t}\big(\widehat{w}^n_{fp}(\centerdot+i\gamma)\big) - \mathrm{Im}G_{\mu_t}\big(w_{fp}(\centerdot+i\gamma)\big)\right\|^2,
\end{aligned}$$

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

$$\mathbb{E}(\Sigma) \le \frac{2C_K^2}{\pi}.e^{\frac{2\gamma}{h}}.\big(I^\kappa + J^\kappa\big) \tag{4.10}$$

where

$$I^\kappa = \int_{\{x\in\mathbb{R}:|x|\le\kappa\}} \mathbb{E}\Big[\Big|\mathrm{Im}\widehat{G}_{\mu^n_t}\big(\widehat{w}^n_{fp}(x+i\gamma)\big) - \mathrm{Im}G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)\Big|^2\Big]\mathrm{d}x \tag{4.11}$$

$$J^\kappa = \int_{\{x\in\mathbb{R}:|x|>\kappa\}} \mathbb{E}\Big[\Big|\mathrm{Im}\widehat{G}_{\mu^n_t}\big(\widehat{w}^n_{fp}(x+i\gamma)\big) - \mathrm{Im}G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)\Big|^2\Big]\mathrm{d}x. \tag{4.12}$$

To obtain the announced rates of convergence for the MISE, we need to be very careful in establishing the upper bounds for I^κ and J^κ. 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 $(\eta^n_1(t),\dots,\eta^n_n(t))$ 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*

$$A^{n,t}_M := \Big\{\forall 1 \le j \le n : \big|\eta^n_j(t)\big| \le M\Big\}. \tag{4.13}$$

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

$$\mathbb{P}\left((A^{n,t}_M)^c\right) = \mathbb{P}\left(\left\{\eta^n_*(t) > M\right\}\right) \le e^{-n.C_{eig}.M}, \tag{4.14}$$

with $\eta^n_*(t) := \max_{i=1,\dots,n}|\eta^n_i(t)|$.

Using this lemma, we can control the tail distribution of $\mathbb{E}[\mu^n_t]$, which is essential to establish very precise estimates. We recall that $\lambda^n_1(t) \le \cdots \le \lambda^n_n(t)$ 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 $1 \le j \le n$,

$$\lambda^n_j(0) - \eta^n_*(t) \le \lambda^n_j(t) \le \lambda^n_j(0) + \eta^n_*(t). \tag{4.15}$$

Therefore, for $1 \le j \le n$,

$$\mathbb{E}\left[\mu^n_t\left(\left\{|\lambda| > \frac{\kappa}{2}\right\}\right)\right] \le \mathbb{E}\left[\mu^n_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right)\right] + \mathbb{P}\left(\left\{\eta^n_*(t) > \frac{\kappa}{4}\right\}\right) \le \mathbb{E}\left[\mu^n_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right)\right] + e^{-\frac{n.C_{eig}.\kappa}{4}}.$$

Recall that after (1.6), we introduced the notation $d_1^n, \dots, d_n^n$ for the i.i.d. random variables with law μ_0 and whose order statistic are the diagonal elements of $X_n(0)$, $\lambda_1^n(0) < \cdots < \lambda_n^n(0)$. We have

$$\mathbb{E}\left[\mu_0^n\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right)\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{P}\left(|d_i^n| > \frac{\kappa}{4}\right) = \mu_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right),$$

so that we finally get

$$\mathbb{E}\left[\mu_t^n\left(\left\{|\lambda| > \frac{\kappa}{2}\right\}\right)\right] \le \mu_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right) + e^{-\frac{n.C_{eig}.\kappa}{4}}. \tag{4.16}$$

Now, we successively study I^κ and J^κ.

4.2.1. *Upper bound for I^κ*

Lemma 4.7. *Let us consider $\gamma > 2\sqrt{t}$. There exist constants C_I^2, C_I^2 and C_I^3 only depending on M and t such that*

$$I^\kappa \le \frac{\gamma^8}{(\gamma^2 - 4t)^4}\frac{C_I^1}{n} + \frac{\kappa C_I^2}{n^2} + C_I^3 \kappa e^{-n.C_{eig}.M}. \tag{4.17}$$

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*

$$\mathfrak{I}_{p,\gamma,M,t} := \int_0^{+\infty}\int_{\mathbb{R}} \frac{1}{\left[\left\{\left||\lambda| - x\right| - \sqrt{t} - M\right\} \vee \frac{\gamma}{2}\right]^p} \mathrm{d}\mu_0(\lambda)\mathrm{d}x \le C(p, M, t), \tag{4.18}$$

for $C(p, M, t)$ a finite constant only depending on p, M and t.

Proof. The supremum in the denominator equals to $\left||\lambda| - x\right| - \sqrt{t} - M$ 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

$$\begin{aligned}
\mathfrak{I}_{p,\gamma,M,t} \le \int_{\mathbb{R}} \Bigg\{ &\int_0^{\left(|\lambda| - \sqrt{t} - M - \frac{\gamma}{2}\right)\vee 0} \frac{1}{\left(|\lambda| - x - \sqrt{t} - M\right)^p}\mathrm{d}x + \int_{\{|\lambda| - \sqrt{t} - M - \frac{\gamma}{2}\}\vee 0}^{|\lambda| + \sqrt{t} + M + \frac{\gamma}{2}} \frac{2^p}{\gamma^p}\mathrm{d}x \\
&+ \int_{|\lambda| + \sqrt{t} + M + \frac{\gamma}{2}}^{+\infty} \frac{1}{\left[x - |\lambda| - \sqrt{t} - M\right]^p}\mathrm{d}x \Bigg\} \mathrm{d}\mu_0(\lambda) \\
\le &\int_{\mathbb{R}} \left\{ \int_{(|\lambda| - \sqrt{t} - M)\wedge\frac{\gamma}{2}}^{|\lambda| - \sqrt{t} - M} \frac{1}{v^p}\mathrm{d}v + \frac{2^p(2\sqrt{t} + 2M + \gamma)}{\gamma^p} + \int_{\frac{\gamma}{2}}^{+\infty} \frac{1}{v^p}\mathrm{d}v \right\} \mathrm{d}\mu_0(\lambda) \\
\le &C(p, M, t) < +\infty,
\end{aligned}$$

since $\gamma > 2\sqrt{t}$. This concludes the proof of Lemma 4.8. □

Proof of Lemma 4.7. We decompose I^κ into three parts, $I^\kappa \leq 3\big(I_1^\kappa + I_2^\kappa + I_3^\kappa\big)$ where:

$$I_1^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(\widehat{w}_{fp}^n(x+i\gamma)\big) - \widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|^2\Big]\mathrm{d}x,$$

$$I_2^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big) - \mathbb{E}\Big[\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big) \mid X^n(0)\Big]\Big|^2\Big]dx,$$

$$I_3^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\mathbb{E}\Big[\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big) \mid X^n(0)\Big] - G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)\Big|^2\Big]\mathrm{d}x.$$

Step 1: Let us first upper bound I_1^κ. It is relatively easy to bound I_1^κ 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

$$I_{11}^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(\widehat{w}_{fp}^n(x+i\gamma)\big) - \widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|^2 1_{A_M^{n,t}}\Big]\mathrm{d}x,$$

$$I_{12}^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(\widehat{w}_{fp}^n(x+i\gamma)\big) - \widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|^2 1_{(A_M^{n,t})^c}\Big]\mathrm{d}x.$$

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

$$I_{12}^\kappa \leq \frac{16}{\gamma^2}\kappa\mathbb{P}((A_M^{n,t})^c) \leq \frac{16}{\gamma^2}\kappa e^{-n.C_{eig}.M}. \tag{4.19}$$

Let us now consider the term I_{11}^κ:

$$\begin{aligned}
I_{11}^\kappa &= \int_{\{|x|\leq\kappa\}} \mathbb{E}\Bigg[\Bigg|\frac{1}{n}\sum_{j=1}^n \frac{w_{fp}(x+i\gamma) - \widehat{w}_{fp}^n(x+i\gamma)}{\big(\widehat{w}_{fp}^n(x+i\gamma) - \lambda_j^n(t)\big).\big(w_{fp}(x+i\gamma) - \lambda_j^n(t)\big)}\Bigg|^2 1_{A_M^{n,t}}\Bigg]\mathrm{d}x \\
&\leq \int_{\{|x|\leq\kappa\}} \mathbb{E}\Bigg[\Big|\widehat{w}_{fp}^n(x+i\gamma) - w_{fp}(x+i\gamma)\Big|^2 \\
&\qquad .\frac{1}{n}\sum_{j=1}^n \frac{1_{A_M^{n,t}}}{\big|\widehat{w}_{fp}^n(x+i\gamma) - \lambda_j^n(t)\big|^2.\big|w_{fp}(x+i\gamma) - \lambda_j^n(t)\big|^2}\Bigg]\mathrm{d}x
\end{aligned}$$

by convexity. Using (2.11) and (4.15), we have

$$\begin{aligned}
\big|w_{fp}(x+i\gamma) - \lambda_j^n(t)\big| &\geq \big|w_{fp}(x+i\gamma)\big| - \big|\lambda_j^n(t)\big| \\
&\geq \big|\mathrm{Re}(w_{fp}(x+i\gamma))\big| - \big|\lambda_j^n(t)\big| \geq |x| - \sqrt{t} - \big|\lambda_j^n(0)\big| - \eta_*^n(t).
\end{aligned}$$

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

$$\begin{aligned}
\big|w_{fp}(x+i\gamma) - \lambda_j^n(t)\big| &\geq \big|\mathrm{Re}(w_{fp}(x+i\gamma) - \lambda_j^n(t))\big| \geq \big|\lambda_j^n(t)\big| - \big|\mathrm{Re}(w_{fp}(x+i\gamma))\big| \\
&\geq \big|\lambda_j^n(t)\big| - |x| - \sqrt{t} \geq \big|\lambda_j^n(0)\big| - \eta_*^n(t) - |x| - \sqrt{t}.
\end{aligned}$$

Therefore, using Theorem 2.6,

$$\left|w_{fp}(x+i\gamma)-\lambda_j^n(t)\right| \geq \left\{\left|\left|\lambda_j^n(0)\right|-\left|x\right|\right|-\sqrt{t}-\eta_*^n(t)\right\} \vee \frac{\gamma}{2}. \tag{4.20}$$

In Theorem-Definition 2.8, it is shown that $\widehat{w}_{fp}^n(z)$ satisfies a similar inequality as (2.11). Thus, we obtain with similar computations that:

$$\left|\widehat{w}_{fp}^n(x+i\gamma)-\lambda_j^n(t)\right|\Big| \geq \left\{\left|\left|\lambda_j^n(0)\right|-\left|x\right|\right|-\sqrt{t}-\eta_*^n(t)\right\} \vee \frac{\gamma}{2}. \tag{4.21}$$

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

$$\begin{aligned}
I_{11}^\kappa &\leq \int_{\{|x|\leq\kappa\}} \mathbb{E}\Bigg[\left|\widehat{w}_{fp}^n(x+i\gamma)-w_{fp}(x+i\gamma)\right|^2 \\
&\quad \cdot \frac{1}{n}\sum_{j=1}^n \frac{1}{\left[\left\{\left|\left|\lambda_j^n(0)\right|-|x|\right|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4} 1_{A_M^{n,t}}\Bigg]\mathrm{d}x \\
&\leq \frac{1}{n}\sum_{j=1}^n \int_{\{|x|\leq\kappa\}} \mathbb{E}\Bigg[\frac{1}{\left[\left\{\left|\left|\lambda_j^n(0)\right|-|x|\right|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4} \\
&\quad \times \mathbb{E}\left[\left|\widehat{w}_{fp}^n(x+i\gamma)-w_{fp}(x+i\gamma)\right|^2 \middle| X^n(0)\right]\Bigg]\mathrm{d}x \\
&\leq \left(\frac{t\gamma^2}{\gamma^2-4t}\right)^2 \times \frac{C(\gamma,t)}{n}\sum_{j=1}^n \int_{\{|x|\leq\kappa\}} \mathbb{E}\Bigg[\frac{1}{\left[\left\{\left|\left|\lambda_j^n(0)\right|-|x|\right|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\Bigg(\frac{1}{n^2} \\
&\quad + \left|\int_{\mathbb{R}} \frac{1}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\left[\mathrm{d}\mu_0^n(\lambda)-\mathrm{d}\mu_0(\lambda)\right]\right|^2\Bigg)\Bigg]\mathrm{d}x \\
&\leq \frac{\gamma^8}{(\gamma^2-4t)^4}\frac{C(t)}{n}\left(I_{111}^\kappa+I_{112}^\kappa\right),
\end{aligned} \tag{4.22}$$

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

$$I_{111}^\kappa := \frac{1}{n^2}\sum_{j=1}^n \int_{\{|x|\leq\kappa\}} \mathbb{E}\left[\frac{1}{\left[\left\{\left|\left|\lambda_j^n(0)\right|-|x|\right|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\right]\mathrm{d}x$$

$$\begin{aligned}
I_{112}^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Bigg[\int &\frac{1}{\left[\left\{\left|\left|\lambda\right|-|x|\right|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\mathrm{d}\mu_0^n(\lambda) \\
&\left|\sqrt{n}\int_{\mathbb{R}} \frac{1}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\left[\mathrm{d}\mu_0^n(\lambda)-\mathrm{d}\mu_0(\lambda)\right]\right|^2\Bigg]\mathrm{d}x.
\end{aligned}$$

Now we wish to upper bound I^{κ}_{111} and I^{κ}_{112} independently of κ. We first deal with I^{κ}_{111}.

$$\begin{aligned}I^{\kappa}_{111} &= \frac{1}{n}\int_{\{|x|\leq\kappa\}}\int_{\mathbb{R}}\frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\mathrm{d}\mu_0(\lambda)\mathrm{d}x\\ &\leq \frac{2}{n}\int_0^{+\infty}\int_{\mathbb{R}}\frac{1}{\left[\left\{\Big||\lambda|-x\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\mathrm{d}\mu_0(\lambda)\mathrm{d}x \leq \frac{2C(4,M,t)}{n},\end{aligned} \tag{4.23}$$

by Lemma 4.8. Let us now consider I^{κ}_{112}. Using Cauchy-Schwarz inequality, we have:

$$\begin{aligned}I^{\kappa}_{112} \leq &\sqrt{\mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\Big(\int\frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\mathrm{d}\mu_0^n(\lambda)\Big)^2\mathrm{d}x\right]}\\ &\sqrt{\mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\left|\sqrt{n}\int_{\mathbb{R}}\frac{1}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\big[\mathrm{d}\mu_0^n(\lambda)-\mathrm{d}\mu_0(\lambda)\big]\right|^4\mathrm{d}x\right]}\end{aligned} \tag{4.24}$$

The first term can be treated exactly as I^{κ}_{111} as:

$$\begin{aligned}&\mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\Big(\int\frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\mathrm{d}\mu_0^n(\lambda)\Big)^2\mathrm{d}x\right]\\ &\leq \mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\frac{1}{(\frac{\gamma}{2})^4}\Big(\int\frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}\mathrm{d}\mu_0^n(\lambda)\Big)\mathrm{d}x\right] = \frac{16n}{\gamma^4}I^{\kappa}_{111}.\end{aligned} \tag{4.25}$$

We now focus on the second term of (4.24). As in the proof of Proposition 3.6, if we denote by $\phi_x := \varphi_{x+i\gamma} : \lambda \mapsto (x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda)^{-1}$, the last term can be rewritten as

$$\begin{aligned}I^{\kappa}_{1121} :=&\mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\left|\sqrt{n}\int_{\mathbb{R}}\phi_x(\lambda)\big[\mathrm{d}\mu_0^n(\lambda)-\mathrm{d}\mu_0(\lambda)\big]\right|^4\mathrm{d}x\right]\\ =&n^2\int_{\{|x|\leq\kappa\}}\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n\Big(\phi_x(d_j^n)-\mathbb{E}(\phi_x(d_j^n)\Big)\right|^4\right]\mathrm{d}x,\end{aligned}$$

where $d_1^n,\dots,d_n^n$ are the non-ordered diagonal elements of $X_n(0)$ (see after Equation (1.6)). Since the random variables $d_1^n,\dots,d_n^n$ are i.i.d. with law μ_0, the random variables $(\phi_x(d_j^n)-\mathbb{E}(\phi_x(d_j^n)))_{1\leq j\leq n}$ are i.i.d. centered with finite fourth moment. By Rosenthal and then Cauchy-Schwarz inequality, we have

$$I^{\kappa}_{1121} \leq \frac{C}{n^2}n^2\int_{\{|x|\leq\kappa\}}\mathbb{E}\Big(\big|\phi_x(d_1^n)-\mathbb{E}\big[\phi_x(d_1^n)\big]\big|^4\Big)\mathrm{d}x, \tag{4.26}$$

for C a constant. We can conclude if the above double integral is bounded independently of κ.

Let us recall now some estimates for the functions ϕ_x. As $\mathrm{Im}\big(G_{\mu_0\boxplus\sigma_t}(x+i\gamma)\big)<0$, we have

$$|x+i\gamma-tG_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda| \geq \mathrm{Im}(x+i\gamma-tG_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda) \geq \gamma \geq \frac{\gamma}{2} \tag{4.27}$$

and the functions ϕ_x are bounded by $2/\gamma$. This yields that $\left|\int_{\mathbb{R}}\phi_x(\lambda)\mathrm{d}\mu_0(\lambda)\right|\leq 2/\gamma$. By Lemma 3.1, $|tG_{\mu_0\boxplus\sigma_t}(x+i\gamma)|\leq\frac{t}{\gamma}\leq\frac{\sqrt{t}}{2}\leq\sqrt{t}$ so that

$$|x+i\gamma-tG_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda|\geq(|x-\lambda|-\sqrt{t})\geq(\left||x|-|\lambda|\right|-\sqrt{t}). \tag{4.28}$$

As a consequence,

$$|x+i\gamma-tG_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda|\geq(\left||x|-|\lambda|\right|-\sqrt{t})\vee\frac{\gamma}{2}. \tag{4.29}$$

Using that d_1^n has distribution μ_0, the double integral in the right hand side of (4.26) becomes:

$$\begin{aligned}
&\int_{\{|x|\leq\kappa\}}\mathbb{E}\Big(\big|\phi_x(d_1^n)-\mathbb{E}\big[\phi_x(d_1^n)\big]\big|^4\Big)\mathrm{d}x\\
&\quad=\int_{\{|x|\leq\kappa\}}\Big\{\mathbb{E}\big[|\phi_x(d_1^n)|^4\big]-2\mathbb{E}\big[|\phi_x(d_1^n)|^2\phi_x(d_1^n)\big]\mathbb{E}\big[\overline{\phi_x}(d_1^n)\big]+\mathbb{E}\big[\phi_x^2(d_1^n)\big]\mathbb{E}\big[\overline{\phi_x}(d_1^n)\big]^2\\
&\qquad-2\mathbb{E}\big[|\phi_x(d_1^n)|^2\overline{\phi_x}(d_1^n)\big]\mathbb{E}\big[\overline{\phi_x}(d_1^n)\big]+4\mathbb{E}\big[|\phi_x(d_1^n)|^2\big]\big|\mathbb{E}\big[\phi_x(d_1^n)\big]\big|^2\\
&\qquad+\mathbb{E}\big[\overline{\phi_x}(d_1^n)^2\big]\big(\mathbb{E}\big[\phi_x(d_1^n)\big]\big)^2-\big|\mathbb{E}\big[\phi_x(d_1^n)\big]\big|^4\Big\}\\
&\quad\leq\mathfrak{I}_{4,\gamma,0,t}+\frac{8}{\gamma}\mathfrak{I}_{3,\gamma,0,t}+\frac{24}{\gamma^2}\mathfrak{I}_{2,\gamma,0,t}
\end{aligned}$$

by using the notation of Lemma 4.8 and by neglecting the term $-\left|\mathbb{E}\big[\phi_x(d_1^n)\big]\right|^4<0$. The Lemma 4.8 allows us to conclude that I_{1121} is bounded by a constant only depending on t, since $\gamma>2\sqrt{t}$.

We can now conclude Step 1. The last result, together with (4.26), implies that I_{112}^κ is bounded by a constant only depending on t. From (4.22) and (4.23), we have that $I_{11}^\kappa\leq\frac{\gamma^8}{(\gamma^2-4t)^4}\frac{C_1(M,t)}{n}$ for $C_1(M,t)$ a constant only depending on M and t. Gathering this result with (4.19), we finally obtain that:

$$I_1^\kappa\leq\frac{\gamma^8}{(\gamma^2-4t)^4}\frac{C_1(M,t)}{n}+\frac{16}{\gamma^2}\kappa e^{-n.C_{eig}.M}. \tag{4.30}$$

Step 2: Let us consider I_2^κ. Using Proposition 3.2, we have:

$$\begin{aligned}
I_2^\kappa&=\int_{\{|x|\leq\kappa\}}\mathbb{E}\Big[\mathrm{Var}\Big(\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\mid X^n(0)\Big)\Big]\mathrm{d}x\\
&=\int_{\{|x|\leq\kappa\}}\mathbb{E}\Big[\mathrm{Var}\Big(A_1^n\big(w_{fp}(x+i\gamma)\big)\mid X^n(0)\Big)\Big]\mathrm{d}x\\
&\leq\int_{\{|x|\leq\kappa\}}\frac{10\,t}{n^2\mathrm{Im}^4\big(w_{fp}(x+i\gamma)\big)}dx\leq\frac{10.2^4.t.2\kappa}{n^2\gamma^4}\leq\frac{20\kappa}{n^2t}.
\end{aligned} \tag{4.31}$$

Step 3: Let us now provide an upper bound for I_3^κ. Recall the definitions of $A_2^n(z)$ and $A_3^n(z)$ in (3.3):

$$\begin{aligned}
I_3^\kappa&=\int_{\{|x|\leq\kappa\}}\mathbb{E}\Big[\big|A_2^n\big(w_{fp}(x+i\gamma)\big)+A_3^n\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]\mathrm{d}x\\
&\leq2\int_{\{|x|\leq\kappa\}}\mathbb{E}\Big[\big|A_2^n\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]\mathrm{d}x+2\int_{\{|x|\leq\kappa\}}\mathbb{E}\Big[\big|A_3^n\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]\mathrm{d}x.
\end{aligned} \tag{4.32}$$

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

$$2\int_{\{|x|\le\kappa\}}\mathbb{E}\Big[\big|A_2^n\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]\mathrm{d}x \le \frac{c\kappa}{n^2 t},$$

where c is an absolute constant. Let us now consider the second term in the right hand side of (4.32). Using the bound of Proposition 3.6,

$$\begin{aligned}&2\int_{\{|x|\le\kappa\}}\mathbb{E}\Big[\big|A_3^n\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]\mathrm{d}x\\ &\le 2\frac{\gamma^4}{(\gamma^2-4t)^2}\int_{\mathbb{R}}\mathbb{E}\Big[\Big|\int_{\mathbb{R}}\frac{1}{w_{fp}(x+i\gamma)-t.G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)-v}[\mathrm{d}\mu_0^n(v)-\mathrm{d}\mu_0(v)]\Big|^2\Big]\mathrm{d}x. \qquad (4.33)\end{aligned}$$

Recall that μ_0^n is the empirical measure of independent random variables (d_i^n) with distribution μ_0 and whose order statistics are the $(\lambda_i^n(0))$. Recalling that $\big(w_{fp}(x+i\gamma)-t.G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)-v\big)^{-1} = \varphi_{w_{fp}(x+i\gamma)}(v)$, we have that

$$\begin{aligned}&\mathbb{E}\Big[\Big|\int_{\mathbb{R}}\varphi_{w_{fp}(x+i\gamma)}(v)[\mathrm{d}\mu_0^n(v)-\mathrm{d}\mu_0(v)]\Big|^2\Big] = \mathrm{Var}\Big[\frac{1}{n}\sum_{j=1}^n\varphi_{w_{fp}(x+i\gamma)}(\lambda_j^n(0))\Big]\\ &\quad\le \frac{1}{n}\mathbb{E}\big[|\varphi_{w_{fp}(x+i\gamma)}(d_1^n)|^2\big] = \frac{1}{n}\int_{\mathbb{R}}\frac{1}{|w_{fp}(x+i\gamma)-t.G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)-v|^2}\mathrm{d}\mu_0(v). \qquad (4.34)\end{aligned}$$

We have:

$$\begin{aligned}|w_{fp}(x+i\gamma)-t.G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)-v| &\ge \big|\mathrm{Re}\big(w_{fp}(x+i\gamma)-t.G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)\big)-v\big|\\ &\ge \big|\mathrm{Re}\big(w_{fp}(x+i\gamma)\big)-v\big| - t\,\big|\mathrm{Re}\left(G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)\right)\big|.\end{aligned}$$

By Theorem 2.6 (i), we have that:

$$\big|\mathrm{Re}\left(G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)\right)\big| \le \left|\int_{\mathbb{R}}\frac{\mathrm{d}\mu_t(y)}{w_{fp}(x+i\gamma)-y}\right| \le \frac{1}{|\mathrm{Im}(w_{fp}(x+i\gamma))|} \le \frac{2}{\gamma}.$$

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

$$|w_{fp}(x+i\gamma)-t.G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)-v| \ge \big||x|-|v|\big| - \sqrt{t} - \frac{2t}{\gamma}. \qquad (4.35)$$

From (4.33), (4.34) and (4.35), we have that:

$$\begin{aligned}&2\int_{\{|x|\le\kappa\}}\mathbb{E}\Big[\big|A_3^n\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]\mathrm{d}x\\ &\qquad\le \frac{2\gamma^4}{n(\gamma^2-4t)^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{1}{\left(\left\{\big||x|-|v|\big|-\sqrt{t}-\frac{2t}{\gamma}\right\}\vee\frac{\gamma}{2}\right)^2}\mathrm{d}\mu_0(v)\mathrm{d}x \le \frac{4\gamma^4}{n(\gamma^2-4t)^2}\mathfrak{I}_{2,\gamma,2\sqrt{t},t},\end{aligned}$$

by Lemma 4.8. We conclude as for I_{11}^κ and we obtain

$$I_3^\kappa \leq \frac{c\kappa}{n^2 t} + \frac{4\gamma^4}{n(\gamma^2 - 4t)^2} \mathfrak{I}_{2,\gamma,2\sqrt{t},t}. \tag{4.36}$$

Gathering (4.30), (4.31) and (4.36) we obtain the result announced in Lemma 4.7. $\square$

4.2.2. *Upper bound for* J^κ

Recall the definition of J^κ in (4.10). Our goal is to prove the following bound:

Lemma 4.9. *There exist constants C_J^1, C_J^2 and C_J^3 only depending on t such that, for any $\kappa > \gamma$, we have*:

$$J^\kappa \leq \frac{C_J^1}{\kappa} + C_J^2 e^{-\frac{n.C_{eig}.\kappa}{4}} + C_J^3 \mu_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right). \tag{4.37}$$

Proof. We decompose $J^\kappa \leq 2(J_1^\kappa + J_2^\kappa)$ where

$$J_1^\kappa := \int_{\{|x|>\kappa\}} \mathbb{E}\left(\left|\int_{\mathbb{R}} \frac{\mathrm{d}\mu_t^n(\lambda)}{\widehat{w}_{fp}^n(x+i\gamma)-\lambda}\right|^2\right)\mathrm{d}x, \quad \text{and} \quad J_2^\kappa := \int_{\{|x|>\kappa\}} \left|\int_{\mathbb{R}} \frac{\mathrm{d}\mu_t(\lambda)}{w_{fp}(x+i\gamma)-\lambda}\right|^2 \mathrm{d}x.$$

Let us consider the first term J_1^κ. Using the estimate of Theorem-Definition 2.8, we have for all $x \in \mathbb{R}$ that $\left|\mathrm{Re}\left(\widehat{w}_{fp}^n(x+i\gamma)\right) - x\right| \leq \sqrt{t}$ and $\mathrm{Im}\left(\widehat{w}_{fp}^n(x+i\gamma)\right) \geq \gamma/2$. This allows us to prove that there exists a constant C_t only depending on t such that for all $\lambda \in \mathbb{R}$

$$(x-\lambda)^2 + \frac{\gamma^2}{4} \leq C_t\left(\mathrm{Re}^2\left(\widehat{w}_{fp}^n(x+i\gamma)-\lambda\right)^2 + \frac{\gamma^2}{4}\right). \tag{4.38}$$

Thus,

$$\begin{aligned}
J_1^\kappa &\leq \int_{\{|x|>\kappa\}} \mathbb{E}\int_{\mathbb{R}} \frac{\mathrm{d}\mu_t^n(\lambda)}{\mathrm{Re}^2\left(\widehat{w}_{fp}^n(x+i\gamma)-\lambda\right) + \mathrm{Im}^2\left(\widehat{w}_{fp}^n(x+i\gamma)\right)^2}\mathrm{d}x \\
&\leq C_t\,\mathbb{E}\left[\int_{\mathbb{R}} \mathrm{d}\mu_t^n(\lambda)\int_{\{|x|>\kappa\}} \frac{1}{(x-\lambda)^2+\frac{\gamma^2}{4}}\mathrm{d}x\right] \\
&= \frac{2C_t}{\gamma}\mathbb{E}\left[\int_{\mathbb{R}} \mathrm{d}\mu_t^n(\lambda)\left(\pi - \arctan\left(\frac{2}{\gamma}(\kappa-\lambda)\right) - \arctan\left(\frac{2}{\gamma}(\kappa+\lambda)\right)\right)\right] \\
&= \frac{2C_t}{\gamma}\mathbb{E}\left[\int_{\mathbb{R}} \mathrm{d}\mu_t^n(\lambda)\left(\arctan\left(\frac{4\kappa\gamma}{4\kappa^2-4\lambda^2-\gamma^2}\right) + \pi\mathbf{1}_{\left\{\lambda^2>\kappa^2-\frac{\gamma^2}{4}\right\}}\right)\right].
\end{aligned}$$

We now use the simple bounds $|\arctan x| \leq |x|$ and $|\arctan x| \leq \frac{\pi}{2}$ 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

$$J_1^\kappa \leq \frac{2C_t}{\gamma}\mathbb{E}\left[\int_\lambda \mathrm{d}\mu_t^n(\lambda)\left(\frac{2\gamma}{\kappa} + \frac{\pi}{2}1_{\left\{\lambda^2>\frac{\kappa^2}{2}-\frac{\gamma^2}{4}\right\}} + \pi 1_{\left\{\lambda^2>\kappa^2-\frac{\gamma^2}{4}\right\}}\right)\right].$$

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

$$\begin{aligned} J_1^\kappa \leq & C_t \left(\frac{4}{\kappa} + \frac{3\pi}{\gamma} \mathbb{E}\left[\mu_t^n \left(\{|\lambda| > \frac{\kappa}{2}\} \right) \right] \right) \\ \leq & C_t \left(\frac{4}{\kappa} + \frac{3\pi}{\gamma} \mu_0 \left(\left\{ |\lambda| > \frac{\kappa}{4} \right\} \right) + \frac{3\pi}{\gamma} e^{-\frac{n.C_{eig}.\kappa}{4}} \right), \end{aligned} \tag{4.39}$$

by using (4.16).

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

$$(x-\lambda)^2 + \frac{\gamma^2}{4} \leq C_t \Big(\mathrm{Re}^2\big(w_{fp}(x+i\gamma) - \lambda\big)^2 + \frac{\gamma^2}{4} \Big),$$

with C_t as above. Thus,

$$\begin{aligned} J_2^\kappa &\leq C_t \int_{\{|x|>\kappa\}} \int_\lambda \frac{\mathrm{d}\mu_t(\lambda)}{(x-\lambda)^2 + \frac{\gamma^2}{4}} \mathrm{d}x \\ &\leq \frac{2C_t}{\gamma} \int_\lambda \mathrm{d}\mu_t(\lambda) \left(\frac{2\gamma}{\kappa} + \frac{\pi}{2} 1_{\left\{\lambda^2 > \frac{\kappa^2}{2} - \frac{\gamma^2}{4}\right\}} + \pi 1_{\left\{\lambda^2 > \kappa^2 - \frac{\gamma^2}{4}\right\}} \right). \end{aligned}$$

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 (4.16), 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

$$J_2^\kappa \leq C_t \left(\frac{4}{\kappa} + \frac{3\pi}{\gamma} \mu_0 \left(\left\{ |\lambda| > \frac{\kappa}{4} \right\} \right) \right). \tag{4.40}$$

Gathering the upper bounds (4.39) and (4.40), we get that for any $\kappa > \gamma$,

$$J^\kappa \leq C_t \left(\frac{8}{\kappa} + \frac{6\pi}{\gamma} \mu_0 \left(\left\{ |\lambda| > \frac{\kappa}{4} \right\} \right) + \frac{3\pi}{\gamma} e^{-\frac{n.C_{eig}.\kappa}{4}} \right). \tag{4.41}$$

This ends the proof. □

4.2.3. *Conclusion*

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

$$I^\kappa + J^\kappa \leq \frac{\gamma^8}{(\gamma^2 - 4t)^4} \frac{C_I^1}{n} + \frac{\kappa C_I^2}{n^2} + C_I^3 \kappa e^{-n.C_{eig}.M} + \frac{C_J^1}{\kappa} + C_J^2 e^{-nC_{eig}\cdot\frac{\kappa}{4}} + C_J^3 \mu_0 \left(\left\{ |\lambda| > \frac{\kappa}{4} \right\} \right).$$

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 $\mu_0(\{|\lambda| > n\}) \leq Cn^{-1}$, for some absolute constant C. Then, from (4.10) and previous computations, there exists a constant $C_{\text{var}}(t)$ (that depends only on t) such that for n sufficiently large:

$$\mathbb{E}(\Sigma) \leq \frac{\gamma^8}{(\gamma^2-4t)^4} \frac{C_{\text{var}}(t).e^{\frac{2\gamma}{h}}}{n} \tag{4.42}$$

and Theorem 4.1 is proved.

5. Numerical simulations

In this section, we conduct a simulation study to assess the performances of our estimator $\widehat{p}_{0,h}$ 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 $w \sim Ber(0.25)$, $Z_1 \sim \mathcal{N}(0,1)$ 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 $\widehat{p}_{0,h}$, we approximate integrals involved in Fourier and inverse Fourier transforms by Riemann sums, so it may happen that $\widehat{p}_{0,h}(x)$ is not real. This is the reason why the density p_0 is estimated with $\text{Re}(\widehat{p}_{0,h})$, the real part of $\widehat{p}_{0,h}$.

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 $\text{Re}(\widehat{p}_{0,h})$ as follows:

$$\begin{aligned}\left\|\text{Re}(\widehat{p}_{0,h}) - p_0\right\|^2 &= \int_{\mathbb{R}} \left|\text{Re}(\widehat{p}_{0,h}(x)) - p_0(x)\right|^2 dx \\ &= \left\|\text{Re}(\widehat{p}_{0,h})\right\|^2 - 2\int_{\mathbb{R}} \text{Re}(\widehat{p}_{0,h}(x))p_0(x)dx + \|p_0\|^2.\end{aligned}$$

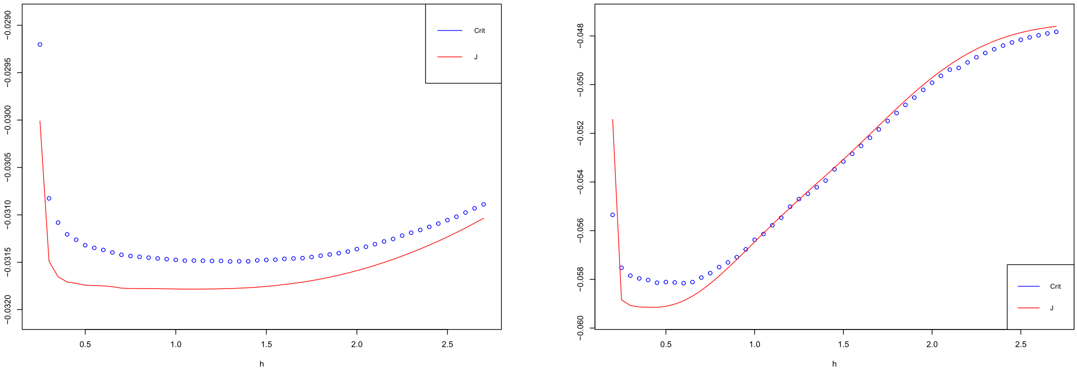
Then, an ideal bandwidth h would minimize the criterion J with

$$J(h) := \left\|\text{Re}(\widehat{p}_{0,h})\right\|^2 - 2\int_{\mathbb{R}} \text{Re}(\widehat{p}_{0,h}(x))p_0(x)dx, \quad h \in \mathbb{R}_+^*.$$

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 disjoints sets

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

There are $V_{\max} := \binom{n}{n/2}$ possibilities to select the subsets (E, E^c), which is huge. Hence, to reduce computational time, we draw randomly $V = 10$ partitions denoted $(E_j, E_j^c)_{j=1,\ldots,V}$. Choosing the grid



(a) (b)

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

$\mathcal{H}$ of 50 equispaced points lying between $h_{\min} = 0.25$ and $h_{\max} = 2.7$, our selected bandwidth is

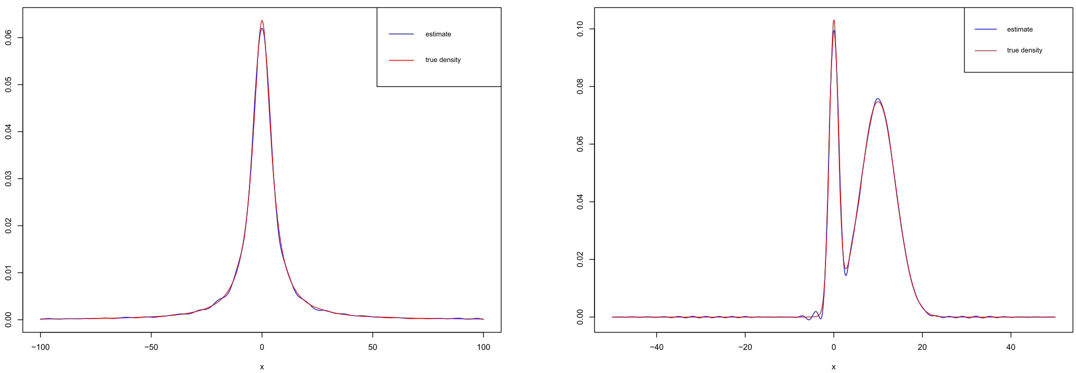
$$\hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmin}}\ \mathrm{Crit}(h) \tag{5.1}$$

with

$$\mathrm{Crit}(h) := \min_{h' \in \mathcal{H}, h' \neq h} \frac{1}{V} \sum_{j=1}^{V} \left(\left\| \mathrm{Re}(\widehat{p}_{0,h}^{(E_j)}) \right\|^2 - 2 \int_{\mathbb{R}} \mathrm{Re}(\widehat{p}_{0,h}^{(E_j)}(x)) \mathrm{Re}(\widehat{p}_{0,h'}^{(E_j^c)}(x)) dx \right)$$

and our final estimator is then $\mathrm{Re}(\widehat{p}_{0,\hat{h}})$. In the last expression, $\widehat{p}_{0,h}^{(E_j)}$ and $\widehat{p}_{0,h'}^{(E_j^c)}$ are estimates based on the samples E_j and E_j^c respectively.

To evaluate our approach, Figure 1 displays the plot of $h \in \mathcal{H} \mapsto \mathrm{Crit}(h)$ and $h \in \mathcal{H} \mapsto J(h)$ for each density p_0. A close inspection of the graphs shows that the minimizer of the first criterion is a good



(a) (b)

Figure 2. Estimation of p_0 (a): for the Cauchy density (b): for the mixture of Gaussian densities.

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 $\mathrm{Re}(\widehat{p}_{0,\hat{h}})$ for each density p_0. The results are quite satisfying, 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 [27].

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.

Supplementary Material

Supplement to "Statistical deconvolution of the free Fokker-Planck equation at fixed time" (DOI: 10.3150/21-BEJ1366SUPP; .pdf). The supplemental material contains all the appendices: Appendix A (Proof of Equation (1.7)), Appendix B (Proof of Theorem 2.6 and Theorem-Definition 2.8), Appendix C (Proof of Lemma 3.4), Appendix D (Proof of Lemma 3.5) and Appendix E (Proof of Corollary 4.3).

References

[1] Anderson, G.W., Guionnet, A. and Zeitouni, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge: Cambridge Univ. Press. MR2760897

[2] Arizmendi, O., Tarrago, P. and Vargas, C. (2020). Subordination methods for free deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2565–2594. MR4164848 https://doi.org/10.1214/20-AIHP1050

[3] Bai, Z. and Silverstein, J.W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

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

[5] Belomestny, D. and Goldenshluger, A. (2021). Density deconvolution under general assumptions on the distribution of measurement errors. *Ann. Statist.* **49** 615–649. MR4255101 https://doi.org/10.1214/20-aos1969

[6] Bertoin, J., Giraud, C. and Isozaki, Y. (2001). Statistics of a flux in Burgers turbulence with one-sided Brownian initial data. *Comm. Math. Phys.* **224** 551–564. MR1870057 https://doi.org/10.1007/PL00005589

[7] Biane, P. (1997). On the free convolution with a semi-circular distribution. *Indiana Univ. Math. J.* **46** 705–718. MR1488333 https://doi.org/10.1512/iumj.1997.46.1467

[8] Biane, P. and Speicher, R. (2001). Free diffusions, free entropy and free Fisher information. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 581–606. MR1851716 https://doi.org/10.1016/S0246-0203(00)01074-8

[9] Bourgain, J. (1994). Periodic nonlinear Schrödinger equation and invariant measures. *Comm. Math. Phys.* **166** 1–26. MR1309539

[10] Burgers, J.M. (1974). *The Nonlinear Diffusion Equation*. Berlin: Springer.

[11] Burq, N. and Tzvetkov, N. (2008). Random data Cauchy theory for supercritical wave equations. I. Local theory. *Invent. Math.* **173** 449–475. MR2425133 https://doi.org/10.1007/s00222-008-0124-z

[12] Burq, N. and Tzvetkov, N. (2008). Random data Cauchy theory for supercritical wave equations. II. A global existence result. *Invent. Math.* **173** 477–496. MR2425134 https://doi.org/10.1007/s00222-008-0123-0

[13] Butucea, C. and Tsybakov, A.B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.* **52** 111–128. MR2354572 https://doi.org/10.1137/S0040585X97982840

[14] Carrillo, J.A., McCann, R.J. and Villani, C. (2003). Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.* **19** 971–1018. MR2053570 https://doi.org/10.4171/RMI/376

[15] Comte, F., Cuenod, C.-A., Pensky, M. and Rozenholc, Y. (2017). Laplace deconvolution on the basis of time domain data and its application to dynamic contrast-enhanced imaging. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 69–94. MR3597965 https://doi.org/10.1111/rssb.12159

[16] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 569–609. MR3088382 https://doi.org/10.1214/11-aihp470

[17] Constantin, P. and Wu, J. (1997). Statistical solutions of the Navier-Stokes equations on the phase space of vorticity and the inviscid limits. *J. Math. Phys.* **38** 3031–3045. MR1449547 https://doi.org/10.1063/1.532032

[18] Dallaporta, S. and Fevrier, M. (2019). Fluctuations of linear spectral statistics of deformed Wigner matrices. Submitted. hal-02079313.

[19] Dattner, I., Reiß, M. and Trabs, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli* **22** 143–192. MR3449779 https://doi.org/10.3150/14-BEJ626

[20] Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *IEEE Trans. Inf. Theory* **48** 734–747. MR1889978 https://doi.org/10.1109/18.986021

[21] Flandoli, F. (2018). Weak vorticity formulation of 2D Euler equations with white noise initial condition. *Comm. Partial Differential Equations* **43** 1102–1149. MR3910197 https://doi.org/10.1080/03605302.2018.1467448

[22] Giraud, C. (2003). Some properties of Burgers turbulence with White noise initial conditions. In *Probabilistic Methods in Fluids* 161–178. River Edge, NJ: World Sci. Publ. MR2083371 https://doi.org/10.1142/9789812703989_0011

[23] Groetsch, C.W. (1984). *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. Research Notes in Mathematics* **105**. Boston, MA: Pitman (Advanced Publishing Program). MR0742928

[24] Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342** 877–882. MR2224640 https://doi.org/10.1016/j.crma.2006.04.006

[25] Maïda, M., Nguyen, T.D., Pham Ngoc, T.M., Rivoirard, V. and Tran, V.C. (2022). Supplement to "Statistical deconvolution of the free Fokker-Planck equation at fixed time." https://doi.org/10.3150/21-BEJ1366SUPP

[26] Méléard, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic Models for Nonlinear Partial Differential Equations* (*Montecatini Terme*, 1995). *Lecture Notes in Math.* **1627** 42–95. Berlin: Springer. MR1431299 https://doi.org/10.1007/BFb0093177

[27] Nguyen, T.D. (2021). Nonparametric free deconvolution and circular regression estimation Ph.D. thesis, Université Paris Saclay, Paris, France.

[28] Pensky, M. (2017). Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity. *Ann. Statist.* **45** 1516–1541. MR3670187 https://doi.org/10.1214/16-AOS1498

[29] Pensky, M. and Sapatinas, T. (2009). Functional deconvolution in a periodic setting: Uniform case. *Ann. Statist.* **37** 73–104. MR2488345 https://doi.org/10.1214/07-AOS552

[30] Pensky, M. and Sapatinas, T. (2010). On convergence rates equivalency and sampling strategies in functional deconvolution models. *Ann. Statist.* **38** 1793–1844. MR2662360 https://doi.org/10.1214/09-AOS767

[31] Sznitman, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math.* **1464** 165–251. Berlin: Springer. MR1108185 https://doi.org/10.1007/BFb0085169

[32] Talay, D. and Vaillant, O. (2003). A stochastic particle method with random weights for the computation of statistical solutions of McKean-Vlasov equations. *Ann. Appl. Probab.* **13** 140–180. MR1951996 https://doi.org/10.1214/aoap/1042765665

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

[34] Tran, V.C. (2008). A wavelet particle approximation for McKean-Vlasov and 2D-Navier-Stokes statistical solutions. *Stochastic Process. Appl.* **118** 284–318. MR2376903 https://doi.org/10.1016/j.spa.2007.04.008

[35] Tzvetkov, N. (2019). Random data wave equations. In *Singular Random Dynamics*. *Lecture Notes in Math*. **2253** 221–313. Cham: Springer. MR3971360

[36] Vergassola, M., Dubrulle, B., Frisch, U. and Noullez, A. (1994). Burgers'equation, Devil's staircases and the mass distribution for large-scale structures. *Astron*. *Astrophys*. **289** 325–356.

[37] Vishik, M.J. and Fursikov, A.V. (1988). *Mathematical Problems of Statistical Hydromechanics*. *Mathematics and Its Applications* (*Soviet Series*) **9**. Dordrecht: Kluwer Academic. MR3444271 https://doi.org/10.1007/978-94-009-1423-0

[38] Voiculescu, D. (1986). Addition of certain noncommuting random variables. *J*. *Funct*. *Anal*. **66** 323–346. MR0839105 https://doi.org/10.1016/0022-1236(86)90062-5

[39] Voiculescu, D. (1993). The analogues of entropy and of Fisher's information measure in free probability theory. I. *Comm*. *Math*. *Phys*. **155** 71–92. MR1228526

Received June 2020 and revised January 2021