université
PARIS-SACLAY

Nonparametric free deconvolution and circular regression estimation

Déconvolution libre non paramétrique et estimation pour la régression circulaire

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 574, Mathématiques Hadamard (EDMH)
Spécialité de doctorat: Mathématiques Appliquées
Unité de recherche: Université Paris-Saclay, CNRS,
Laboratoire de mathématiques d'Orsay (LMO UMR 8628),
91405, Orsay, France
Graduate School: Mathématiques. Référent: Faculté des sciences d'Orsay

Thèse présentée et soutenue à Orsay, le 29 Novembre 2021, par

Tien-Dat NGUYEN

Composition du jury:

Gilles BLANCHARD Professeur des Universités, Université Paris-Saclay	Président du jury
Fabienne COMTE Professeure des Universités, Université de Paris	Rapporteur & Examinatrice
Jeff J.F. YAO Professeur des Universités, University of Hong Kong	Rapporteur & Examinateur
Marc HOFFMANN Professeur des Universités, Université Paris Dauphine - PSL	Examinateur
Florence MERLEVÈDE Professeure des Universités, Université Gustave Eiffel	Examinatrice

Direction de la thèse:

Vincent RIVOIRARD Professeur des Universités, Université Paris Dauphine - PSL	Directeur de thèse
Thanh Mai PHAM NGOC Maîtresse de Conférences, Université Paris-Saclay	Co-Directrice de thèse
Mylène MAÏDA Professeure des Universités, Université de Lille	Co-Encadrante
Viet Chi TRAN Professeur des Universités, Université Gustave Eiffel	Co-Encadrant

Thèse de doctorat

NNT: 2021UPASM041

Acknowledgements

This thesis will never be completed without the support and encouragement from many people to whom I would like to express my deep gratitude.

First of all, I would like to express a deep sense of gratitude to my advisors Prof. Thanh Mai Pham Ngoc, Prof. Vincent Rivoirard, Prof. Viet Chi Tran and Prof. Mylène Maïda not only for introducing to me new and exciting subjects, but also for their constant support, guidance and concern in both academic and life during three years of my PhD study. I began this study with a lack of many skills so everything even little I have learned from my advisors, I am greatly thankful. They taught me to grow up both in doing research and in my life through the past three years. There was actually a very difficult period of time for me during my PhD study so that I would like to thank my advisors for their patience and encouragement. Without their help, advice and guidance, this thesis would never have been able to achieve. I am so fortunate and proud to be their student.

I am very grateful to Prof. Fabienne Comte and Prof. Jeff J.F. Yao for kindly accepting to be the reviewers of this thesis. I highly appreciate their careful reading of the thesis manuscript, the evaluation and the constructive comments. I also would like to thank Prof. Gilles Blanchard, Prof. Marc Hoffmann and Prof. Florence Merlevède for kindly acceptance to be members of my PhD defense committee.

I would like to thank all the members of the Laboratoire de Mathématique d'Orsay (LMO), in particular the Probability and Statistics team, for a warm welcome and providing the facilitatory research environment during these three years of my PhD. In particular, many thanks to Prof. Christophe Giraud, Prof. Stéphane Nonnenmacher for all the encouragement and support for me, as well as Prof. Maxime Février for helpful explanations and discussions. Thanks also administration staffs for the kind support and the IT team at LMO, in particular anh Laurent Dang for helping me kindly a lot to set up the computer in my office and Mme. Mathilde Rousseau for technical help on my PhD defense.

I would like to express my gratitude to the Labex Mathématique Hadamard (LMH) and the Fondation Mathématique Jacques Hadamard (FMJH), for supporting me during my Master-2 and then 3 years (and 2 months extension) for my PhD study at LMO. My PhD study was supported by a public grant as part of the Investissement d'avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH. For formalities and paperwork, I would like to thank particularly Mme. Clotilde d'Epenoux, Mme. Isabelle Jasinowski and Mme. Séverine Simon for their kind help.

My next thank will go to professors as well as some friends of the Faculty of Mathematics and Computer Sciences, University of Sciences, Vietnam National University (VNU) Ho Chi Minh city, particularly Dr. Van Ha Hoang for having helped me a lot when the first time I came to Orsay for Master-2 study, my former PUF internship advisor in Bordeaux Prof. Bernard Bercu as well as Dr. Quang Vu Huynh and Prof. Minh Duc Duong for their helpful

fondation courses in my undergraduate study and especially, my undergraduate advisor Prof. Duc Trong Dang who introduced and advised me to pursue higher studies in Statistics.

A special mention goes to my Vietnamese friends in France. I cannot imagine how dreary the past three years in France would have been without them. Many thanks should go to Duc Thach Son Vu, Van Thanh Nguyen, anh Nhat Thien Pham, Tan Binh Phan, Trung Tin Nguyen, Duc Nam Nguyen, anh chi The Anh-Thanh Tam, anh Anh Thi Dinh, anh Minh Hieu Do, anh Phuong Nguyen, anh Hoang Chinh Lu, anh chi Trung-Van, em Thinh-Hanh, chi Thi Tuyet Trang Chau, anh Phuoc Nhat Dang, anh Khieu Nguyen. In particular, chi Thien Trang Bui and anh Thanh Nhan Le for having taken care of me during my PUF internship in Bordeaux when it was the first time I came to France. In addition, many thanks to former PhD students in office 3D3: Julien Sedro, Minh-Lien Jeanne Nguyen and Solène Thépaut, as well as other PhD students at LMO, particularly, Changzhen Sun, Etienne Lasalle and Guillaume Lachaussee.

Eventually, grateful thanks to my family in Vietnam for unconditional support, having been always encouraging and believing in me. During my study in France, they have been willing to listen and share with me all the joys as well as sorrows and pressure I went through. My family is the place I can lean on and also the motivation for me to overcome challenges and move on in every journey.

Cuối cùng, tôi xin gửi lời biết ơn sâu sắc nhất đến gia đình mình, ba mẹ, anh chị Chuyên-Liên, anh chị Nam-Duyên và các cháu, đã luôn tin tưởng, động viên và yêu thương vô điều kiện. Trong suốt thời gian học xa nhà, mỗi ngày, họ đều là những người sẵn lòng lắng nghe và chia sẻ với tôi mọi niềm vui cũng như áp lực mà tôi trải qua. Gia đình là nơi vững chắc nhất cũng là nơi cuối cùng tôi có thể tựa vào, và còn là động lực tiếp thêm sức mạnh để tôi vượt qua thử thách và bước tiếp trên mọi hành trình.

"Đi khắp thế gian không ai thương con bằng mẹ,
Gánh nặng cuộc đời không ai khổ bằng cha.
Nước biển mênh mông không đong đầy tình mẹ,
Mây trời lồng lộng không phủ kín công cha.
Tần tảo sớm hôm mẹ nuôi con khôn lớn,
Mang cả tấm thân gầy cha che chở đời con. "

(Ca dao-Tục ngữ Việt Nam)

Contents

Chapter 1

Introduction

In this thesis, we study two inverse problems in the statistical framework related to nonparametric density estimation and nonparametric regression estimation, particularly in new contexts:

- The first problem deals with estimation for the density of the random initial condition of a non-linear Fokker-Planck equation from the observations related to time $t > 0$. This is the first time, to our knowledge, that such a problem is examined. For linear partial differential equations (PDEs), this has been considered by Pensky and Sapatinas [48] [49]. An approach to tackle the non-linearity of the Fokker-Planck equation is to work with free probabilistic interpretation of the PDE, involving new techniques of free deconvolution.
- In the second problem, the nonparametric estimation for regression with circular responses is examined. This regression model is inspired by work of Di Marzio et al. [80]. We propose a new warped kernel estimator for the regression function, for which the consistency, rates of convergence and adaptation are carefully studied.

Both problems have specific challenges. It is reasonable to review first the relating standard frameworks of nonparametric estimation in Section 1.1 below. After that, we present the contents of each chapter in this thesis and emphasize the differences with classical literature.

Throughout this manuscript, to assess the optimality of our results, we will consider the minimax framework. More precisely, a minimax rate of convergence ψ_n for estimating a function f over the class $\mathcal{D}$ with loss function ℓ is such that:

1. there is an estimator $\widehat{f}_n$ called optimal in the minimax sense attaining this rate, i.e.
$$\limsup_{n\to+\infty} \sup_{f\in\mathcal{D}} \psi_n^{-1}.\mathbb{E}\left[\ell(\widehat{f}_n, f)\right] \leq C < \infty;$$
2. no other estimator can attain a better rate
$$\liminf_{n\to+\infty} \inf_{\widetilde{f}_n} \sup_{f\in\mathcal{D}} \psi_n^{-1}.\mathbb{E}\left[\ell(\widetilde{f}_n, f)\right] \geq c > 0,$$
for some constants $C > 0, c > 0$ and where the infimum is taken over all possible estimators of f.

1.1 Nonparametric statistical estimation

This chapter starts with a general introduction on two standard statistical problems in nonparametric estimation, namely estimating a density function in the classical convolution model in Section 1.1.1 and estimating a regression function in Section 1.1.2.

1.1.1 Nonparametric density deconvolution estimation

In this subsection, we briefly review some standard results on classical statistical deconvolution for estimating a probability density function. This will remind the reader some useful results which help to make relevant connections with the convergence rates obtained for the free deconvolution problem presented in Chapter 2. Indeed, solving a free deconvolution problem by subordination functions method leads to a convolution (in a classical sense) of the target density with a (centered) Cauchy distribution. In addition, we restrict here to the 1-dimension framework for the sake of simplicity.
In the classical framework (see e.g. [25], [43], [59] and the references therein), we consider the so-called convolution model

$$Y_j = X_j + \varepsilon_j, \quad j = 1, ..., n, \tag{1.1}$$

where X_j are independent identically distributed (i.i.d.) random variables with an unknown probability density f_X with respect to the Lebesgue measure on $\mathbb{R}$, the random noise variables ε_j are i.i.d. with **known** probability density f_ε with respect to the Lebesgue measure on $\mathbb{R}$, and $(\varepsilon_1, ..., \varepsilon_n)$ is independent of $(X_1, ..., X_n)$. The density deconvolution problem consists in estimating the function f_X from observations $Y_1, ..., Y_n$. Let f_Y be the probability density of the variables Y_j, then we have $f_Y = f_X * f_\varepsilon$, where the operator $*$ denotes the classical convolution defined for $y \in \mathbb{R}$ by

$$\big(f_X * f_\varepsilon\big)(y) := \int_{\mathbb{R}} f_X(y - x).f_\varepsilon(x)dx.$$

One of major tools of density deconvolution lies in the field of Fourier analysis (see e.g. [50]). Denote the Fourier transform of a function $g \in \mathbb{L}^1(\mathbb{R})$ by $g^\star$, defined for any $\xi \in \mathbb{R}$ as

$$g^\star(\xi) := \int_{\mathbb{R}} g(x).e^{-ix\xi}dx.$$

Then since $f_Y = f_X * f_\varepsilon$, one has

$$f_Y^\star(\xi) = f_X^\star(\xi).f_\varepsilon^\star(\xi), \quad \xi \in \mathbb{R}. \tag{1.2}$$

Thus, using the Fourier transform is motivated by the fact that it transforms $*$, the convolution operator, into simple multiplication of the characteristic functions. As for the quantity $f_Y^\star(\xi) = \mathbb{E}(e^{-i.\xi.Y})$, it is simply estimated by its empirical version.

Remark. Note that the Fourier transform above is defined for $g \in \mathbb{L}^1(\mathbb{R})$. Actually, thanks to Plancherel Theorem, the Fourier transform can be extended to $g \in \mathbb{L}^2(\mathbb{R})$, see e.g. [50, Chapter 9].

The literature about this problem is actually extensive. We can cite the works of Carroll and Hall [19], Devroye [29], Liu and Taylor [39], Masry [41], Stefanski and Carroll [53], Zhang [65], Hesse [35], Cator [20], Delaigle and Gijbels [28], for mainly kernel methods, Koo [36] for a spline method, Pensky and Vidakovic [47] for wavelet strategies, and Comte et al. [24] for adaptive projection strategies. Last but not least, the book of Meister [43] constitutes a comprehensive review on deconvolution problems in nonparametric statistics.

Assumptions

In density deconvolution problem, two factors determine the accuracy of estimation: the smoothness of the density f_X to be estimated and the smoothness of the noise density f_ε. Accordingly, let us present the usual assumptions for this framework. First, we suppose that the noise is such that for all $x \in \mathbb{R}$, $f_\varepsilon^\star(x) \neq 0$ and satisfies the next hypothesis.

Assumption (N1). There exist constants $s \geq 0$, $b \geq 0$, $\alpha \in \mathbb{R}$ ($\alpha > 0$ if $s = 0$) and $k_0, k_1 > 0$ such that

$$k_0.(x^2+1)^{-\alpha/2}.e^{-b|x|^s} \leq \left|f_\varepsilon^\star(x)\right| \leq k_1.(x^2+1)^{-\alpha/2}.e^{-b|x|^s}.$$

When $b > 0$ and $s > 0$, the Fourier transform of the noise has an exponential decay and its density is infinitely differentiable on $\mathbb{R}$. It is then called a super-smooth noise (SS). For instance, the Cauchy distribution which will be of particular interest in Chapter 2 satisfies Assumption (N1) with $s = 1$. When $s = 0$, the noise density is called ordinary smooth (OS).

As for the density f_X, we will assume that it belongs to the following space

$$\mathcal{S}(\delta, a, r, L) = \left\{ f \text{ a density on } \mathbb{R} : \int |f^\star(x)|^2.(x^2+1)^\delta. \exp(2a|x|^r)dx \leq L \right\}, \tag{1.3}$$

with $r \geq 0, a \geq 0, \delta \geq 0$ and $L > 0$. The same terminology as the noise is used to describe the regularity of the function f_X : supersmooth (SS) if $r > 0$, ordinary smooth (OS) otherwise. Note that $r = 0$ corresponds to a Sobolev class.

Estimators

A standard method to solve the classical deconvolution problem is the kernel method (see e.g. [25],[43]). The classical kernel estimator is defined as:

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - Y_i}{h}\right),$$

where the kernel K is such that its inverse Fourier transform has the following expression

$$K^\star(x) = \frac{\mathbb{1}_{|x| \leq 1}}{f_\varepsilon^\star(-x/h)}.$$

In [16], Butucea and Tsybakov have proved that if $f_X \in \mathcal{S}(\delta, a, r, L)$ and under Assumption (N1), the MISE (Mean Integrated Squared Error) of $\hat{f}_n(x)$ which can be classically decomposed into a bias and a variance term is upper-bounded by:

$$\text{MISE } = \mathbb{E}\left[\|f_X - \hat{f}_n\|^2_{\mathbb{L}^2(\mathbb{R})}\right] = O(h^{2\delta}. \exp(-2a/h^r)) + \frac{h^{s-1-2\alpha}}{n}. \exp(2b/h^s).$$

This last bound of the MISE will be of particular interest when compared to the bound we obtained in (2.108) in Chapter 2. Note that another method developed by Comte et al. [24] is the projection estimator leading to a similar upperbound but achieving adaptation. By adaptation, we mean that the procedure does not need the knowledge of the unknown smoothness of the function to be estimated.

Rates of convergence

We now recall the well-known rates of convergence obtained in the literature for the MISE. As said in the preamble of this section, they heavily depend on both smoothnesses of the noise and the density to be recovered. We can distinguish 4 cases gathered in Table 1.1. As one can see in Table 1.1, the smoother the noise, the harder the deconvolution problem at hand. Except for the case where both f_X and f_ε are supersmooth, these rates are known to be optimal minimax (see [15]).

Let us now dwell on the bottom right case (f_X and f_ε SS) as it will particularly concern us in Chapter 2. In Table 1.1 we have not specified on purpose the corresponding rates as

	f_ε OS	f_ε SS
f_X OS	$n^{\frac{-2\delta}{2\delta+2\alpha+1}}$	$(\log n)^{\frac{-2\delta}{s}}$
f_X SS	$\frac{(\log n)^{\frac{2\alpha+1}{r}}}{n}$	$\frac{(\log n)^{\frac{2\alpha+1}{r}}}{n} < . < (\log n)^{\frac{-2\delta}{s}}$

Table 1.1: rates of convergence for the MISE

they are quite intricate to present briefly. For more details, they have been tackled in [37] or in Corollary 2.7.3 (Chapter 2) for the special case of a Cauchy noise. Nonetheless, at this point of the manuscript, we can stress that considering a supersmooth density to estimate allows to improve greatly the rates of convergence in presence of a supersmooth noise. The rates get faster than the standard logarithmic rates when dealing with f_X in a Sobolev class. Furthermore, from a technical point of view, considering f_X and f_ε SS implies non standard bias variance compromises that require new methods for proving lower bounds. The optimality of these rates have been proved so far in the case $r = s = 1$ (see [58]) or the case $0 < r < s$ with $\delta = 0$ (see [16]).

1.1.2 Nonparametric regression with random design

In this subsection, we introduce briefly the problem of classical nonparametric estimation for regression with random design, and we restrict to one-dimensional setting for the sake of simplicity. Formally, in the classical framework (see e.g. [43],[59]), we consider (X, Y) to be a pair of random variables taking values on $\mathbb{R}$ such that

$$Y = m(X) + \varepsilon, \tag{1.4}$$

where ε is a centered real-valued random variable with finite variance and independent of X. In this model, the observed quantity is denoted by Y which is called *response* variable, X is called *covariate* and ε denotes the regression error. Suppose that we are able to observe an i.i.d. sample $\{(X_j, Y_j)\}_{j=1}^n$ having the same distribution as (X, Y). The set of value $\{X_1, ..., X_n\}$ is called the *design* and here the design is random. Furthermore, we assume that X has a density function f_X (with respect to Lebesgue measure), which is called the design density, and let us denote by $D \subset \mathbb{R}$ the support of f_X.
The function $m(X) = \mathbb{E}(Y|X)$ is called the regression function of Y on X, which is unknown in general. Given $x \in D$, the object is to estimate the function $m(x)$ from the data $\{(X_j, Y_j)\}_{j=1}^n$. The following lemma shows that the regression function m is indeed the best predictor of the response variable Y based on observing the covariate X in the sense of quadratic-risk.

Lemma 1.1.1. *For all measurable functions $g : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}\big[\big(g(X)\big)^2\big] < \infty$,*

$$\mathbb{E}\big[|Y - m(X)|^2\big] \leq \mathbb{E}\big[|Y - g(X)|^2\big].$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\big[|Y - g(X)|^2\big] &= \mathbb{E}\big[|Y - m(X)|^2\big] + \mathbb{E}\big[|m(X) - g(X)|^2\big] + 2.\mathbb{E}\big[\big(Y - m(X)\big).\big(m(X) - g(X)\big)\big] \\
&\geq \mathbb{E}\big[|Y - m(X)|^2\big] + 2.\mathbb{E}\big[\big(\mathbb{E}[Y - m(X)|X]\big).\big(m(X) - g(X)\big)\big] \\
&= \mathbb{E}\big[|Y - m(X)|^2\big] \quad (\text{since } \mathbb{E}[Y - m(X)|X] = \mathbb{E}[Y|X] - m(X) = 0).
\end{aligned}$$

□

Kernel estimators of regression function

A common nonparametric approach to construct an estimator for the regression function m is to apply a kernel method. Consider an integrable kernel function $K : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} K(y)dy = 1$ and we define $K_h(\cdot) := \frac{1}{h}.K(\frac{\cdot}{h})$ for $h > 0$ the bandwidth of K. Among the kernel methods, the popular Nadaraya-Watson (NW) estimator of m is defined for model (1.4) by for $x \in D$:

$$\widehat{m}_h^{NW}(x) := \frac{\sum_{j=1}^n Y_j.K_h(x - X_j)}{\sum_{j=1}^n K_h(x - X_j)}, \quad \text{if } \sum_{j=1}^n K_h(x - X_j) \neq 0,$$

and $\widehat{m}_h^{NW}(x) = 0$ otherwise. If one wants to obtain an adaptive estimation of m (that means the estimators do not require the specification of the regularity m), it requires the automatic selection of the bandwidth h, and then the ratio-form of the NW estimator suggests that two bandwidths should be selected: one for the numerator and one for the denominator. Moreover, when the design X is very irregular (for instance when a "hole" occurs in the data), a ratio-form may lead to instability (see e.g. Pham Ngoc [84]). In addition, for regression analysis with L_2-loss, it is standard to assume that the support D of f_X is a compact subset of $\mathbb{R}$. To our knowledge, Kohler et al. [76] are the first who propose theoretical results with no boundedness assumption on the support of the density design. In turn, it requires a moment assumption on the design X (see Assumption $(A4)$ in [76]).

Warped kernel estimators

The "warped" kernel estimators introduced by Yang (1981, [88]), Stute (1984, [87]) and developed by Kerkyacharian and Picard (2004, [75]) avoid the ratio-form and require very few assumptions on the support of f_X. The warped method is based on the introduction of the auxiliary function $g = m \circ F_X^{<-1>}$, where $F_X : y \in \mathbb{R} \to \mathbb{P}(X \leq y)$ is the cumulative distribution function (c.d.f.) of X and $F_X^{<-1>}$ its inverse. The idea of the estimation procedure is to propose firstly an estimator $\widehat{g}$ for g, and then, the regression function m is estimated using $\widehat{g} \circ \widehat{F}_n$, where $\widehat{F}_n$ is the empirical c.d.f. of X. An advantage of warped kernel methods is that it does not involve a ratio (see Section 3.2.1 for more details). To select the kernel bandwidth, we apply the data-driven selection rule of the unique bandwidth in spirit of Goldenshluger and Lepski [71]. Recently, this strategy is applied in the regression setting using kernel methods to construct adaptive nonparametric estimators in univariate case in Chagny [68] and is extended in multivariate framework in Chagny et al [69].

Presentation of two problems studied in this thesis:

We have presented in Section 1.1.1 a brief introduction on classical framework of nonparametric density estimation and in Section 1.1.2 a brief introduction on classical framework of nonparametric regression estimation. Now, as mentioned at the beginning, this thesis contains two independent parts:

- in the first part, Chapter 2, we study the problem of deconvolution for a free Fokker-Planck equation: to construct a statistical estimate for the random initial condition from the observations of the solution at given fixed time $t > 0$.
- in the second part, Chapter 3, the problem of nonparametric estimation for regression with circular responses is examined. We propose a new warped kernel estimator along with the study of consistency, rates of convergence and the adaptation under a consideration of pointwise quadratic risk.

1.2 Part I - Statistical deconvolution of free Fokker-Planck equation

This section presents a brief introduction for the framework that we consider in Chapter 2. In this work, we are interested in the Fokker-Planck PDE which is a special case of McKean-Vlasov PDE (or so-called the granular media equation) equipped with a logarithmic Coulomb interaction. This equation models the motion of particles with electrostatic repulsion and a probabilistic interpretation that we will adopt has been considered e.g. in [21],[9].

1.2.1 Presentation of Fokker-Plank equation

In general, the granular media equations (see e.g. [18]) are a class of PDEs containing many classical equations raising in physics, that is

$$\frac{\partial \mu_t}{\partial t} = \nabla \cdot \left[\mu_t \nabla \Big(\mathcal{U}'(\mu_t) + \mathcal{V} + \mathcal{W} * \mu_t \Big) \right],$$

where $(\mu_t)_{t\geq 0}$ is a family of probability measures on $\mathbb{R}^d$. In this equation, $*$ denotes the classical convolution, ∇ is the gradient operator and $\nabla\cdot$ is the divergence operator. This equation models the evolution over time of a density of particles subjected to three effects:

- a density of internal energy $\mathcal{U} : \mathbb{R}^+ \to \mathbb{R}$;
- a confinement external potential $\mathcal{V} : \mathbb{R}^d \to \mathbb{R}$;
- an interaction potential $\mathcal{W} : \mathbb{R}^d \to \mathbb{R}$, with some assumptions on the regularity in [18].

The free Fokker-Planck equation (will be introduced in more details in Section 2.1) is a particular case of granular media equations with $d = 1, \mathcal{U}(s) = 0, \mathcal{V}(x) = \frac{1}{2}V(x)$ and $\mathcal{W}(x) = -\ln|x|$, where $V : \mathbb{R} \to \mathbb{R}$ is a given external potential. More precisely, it can be written:

$$\frac{\partial \mu_t}{\partial t} = \frac{\partial}{\partial x} \left[\mu_t \left(\frac{1}{2} V' - H\mu_t \right) \right], \tag{1.5}$$

where $(\mu_t)_{t\geq 0}$ is a family of probability measures on $\mathbb{R}$, with initial condition μ_0 and $H\mu$ denotes the Hilbert transform of a probability measure μ on $\mathbb{R}$, defined for any $x \in \mathbb{R}$ by:

$$H\mu(x) := \lim_{\varepsilon \to 0} \int_{\mathbb{R}\setminus[x-\varepsilon, x+\varepsilon]} \frac{1}{x-y} d\mu(y).$$

In our framework, we consider the free Fokker-Planck equation as follows, without potential function V:

$$\frac{\partial \mu_t}{\partial t} = -\frac{\partial}{\partial x} \mu_t (H\mu_t). \tag{1.6}$$

A solution of equation (1.6) is such that for any test function $g \in C^1(\mathbb{R})$:

$$\int_{\mathbb{R}} g(x) d\mu_t(x) = \int_{\mathbb{R}} g(x) d\mu_0(x) + \int_0^t \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{g'(x)}{x-y} d\mu_s(x) d\mu_s(y) \right) ds, \quad t > 0.$$

Theorem 1.2.1. *For any $t > 0$, the equation (1.6) has a unique solution.*

The existence and uniqueness of the solution of the Fokker-Planck equation (1.6) is proved later by probabilistic techniques in Section 2.3 (see also [1, Section 4.3.2]). In addition, contrarily to the examples considered in [48, 49], 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.

1.2.2 Probabilistic interpretation of the solution with free Brownian motion

We first introduce briefly some useful definitions for explaining measures of an operator and also free Brownian motion in the sense of free probability, (see also [1, Section 5.2]).

Definition of free Brownian motion

We recall some fundamental definitions in operator algebras. An *algebra* is a vector space $\mathcal{A}$ over a field F equipped with a multiplication which is associative, distributive and F-bilinear, that is, for $x, y, z \in \mathcal{A}$ and $\alpha \in F$:

- $x(yz) = (xy)z$,
- $(x + y)z = xz + yz$, and $x(y + z) = xy + xz$,
- $\alpha(xy) = (\alpha x)y = x(\alpha y)$.

An algebra $\mathcal{A}$ is *unital* if there exits a unit element $e \in \mathcal{A}$ such that $ae = ea = a$ for any $a \in \mathcal{A}$. Moreover, a *complex algebra* is an algebra over the complex field $\mathbb{C}$. A *seminorm* on a complex algebra $\mathcal{A}$ is a map from $\mathcal{A}$ into $\mathbb{R}^+$ such that for all $x, y \in \mathcal{A}$ and $\alpha \in \mathbb{C}$,

$$\|\alpha x\| = |\alpha| \, \|x\| \, , \quad \|x + y\| \leq \|x\| + \|y\| \, , \quad \|xy\| \leq \|x\| \, . \|y\| \, ,$$

and, if $\mathcal{A}$ is unital with unit e, also $\|e\| = 1$. A norm on a complex algebra $\mathcal{A}$ is a seminorm satisfying that $\|a\| = 0$ implies $a = 0$ in $\mathcal{A}$.
A normed complex algebra is a complex algebra $\mathcal{A}$ equipped with a norm $\|\cdot\|$. A norm complex algebra $(\mathcal{A}, \|\cdot\|)$ is a Banach algebra if the norm $\|\cdot\|$ induces a complete distance. Furthermore, let $(\mathcal{A}, \|\cdot\|)$ be a Banach algebra,

- an *involution* on $\mathcal{A}$ is a map $* : \mathcal{A} \mapsto \mathcal{A}$ satisfying $(a + b)^* = a^* + b^*$, $(ab)^* = b^*a^*$, $(\lambda a)^* = \overline{\lambda} a^*$ (for $\lambda \in \mathbb{C}$), $(a^*)^* = a$ and $\|a^*\| = \|a\|$.
- $\mathcal{A}$ is a $\mathcal{C}^*$-algebra if it possesses an involution $a \mapsto a^*$ that satisfies $\|a^*a\| = \|a\|^2$.

We introduce in the following a fundamental but important definition of non-commutative probability space

Definition 1.2.2. A *non-commutative probability space* is a pair $\big(\mathcal{A}, \phi\big)$ where $\mathcal{A}$ is a unital algebra over $\mathbb{C}$ and $\phi : \mathcal{A} \mapsto \mathbb{C}$ is a linear functional such that $\phi(1_{\mathcal{A}}) = 1$ with $1_{\mathcal{A}}$ the identity element of $\mathcal{A}$. An element of $\mathcal{A}$ is called *non-commutative random variable*.

We give below some relevant examples of non-commutative probability spaces.

Example 1.2.3.

(i) Classical probability theory: Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, then the algebra $\mathcal{A} = \mathbb{L}^\infty(\Omega, \mathcal{F}, \mu)$ of bounded random variables and ϕ as the expectation $\phi(a) = \int_\Omega a(\omega) d\mu(\omega)$, is a non-commutative probability space.

(ii) Matrices: Let $m \in \mathbb{N}^*$, then $\mathcal{A} = \mathrm{Mat}_m(\mathbb{C})$ the space of $m \times m$ matrices with entries belonging to $\mathbb{C}$ and $\phi_m(a) = \frac{1}{m}\mathrm{Tr}(a)$ for $a \in \mathcal{A}$, is a non-commutative probability space.

(iii) Random matrices: Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. Define $\mathcal{A} = \mathbb{L}^\infty(\Omega, \mu, \mathrm{Mat}_m(\mathbb{C}))$, the space of $m \times m$-dimensional complex random matrices with μ-almost surely uniformly bounded entries. For $a \in \mathcal{A}$, set $\phi_m(a) = \frac{1}{m} \int_\Omega \mathrm{Tr}(a(\omega)) d\mu(\omega)$, then $(\mathcal{A}, \phi_m)$ is a non-commutative probability space.

Now, in this section, consider a subset $J \subset \mathbb{N}$ and denote by $\mathbb{C}\langle X_j | j \in J\rangle$ the set of polynomials in non-commutative indeterminates $\{X_j\}_{j\in J}$, that is,

$$\mathbb{C}\langle X_j | j \in J\rangle = \left\{ \gamma_0 + \sum_{k=1}^{m} \gamma_k X_{j_1^k} ... X_{j_{p_k}^k}, \gamma_k \in \mathbb{C}, m \in \mathbb{N}, j_l^k \in J, p_k \in \mathbb{N} \right\}.$$

Definition 1.2.4. Let $\{a_j\}_{j\in J}$ be a family of elements of a non-commutative probability space $(\mathcal{A}, \phi)$. Then, the *distribution* (or law) of $\{a_j\}_{j\in J}$ is the map $\mu_{\{a_j\}_{j\in J}} : \mathbb{C}\langle X_j | j \in J\rangle \mapsto \mathbb{C}$ such that

$$\mu_{\{a_j\}_{j\in J}}(P) = \phi\big(P(\{a_j\}_{j\in J})\big).$$

Let $\mathbb{C}[X] = \mathbb{C}\langle X\rangle$ denote the set of polynomial functions in one variable.

Example 1.2.5 (Example 1.2.3 continued)**.**

(i) Classical probability theory: If $a \in \mathbb{L}^\infty(\Omega, \mathcal{F}, \mu)$, we get by definition that

$$\mu_a(P) = \int P(a(\omega)) d\mu(\omega),$$

and so μ_a is (the sequence of moments of) the law of a under μ.

(ii) One matrix: Let a be an $m \times m$ Hermitian matrix with eigenvalues $(\lambda_1, ..., \lambda_m)$. Then we have, for all polynomials $P \in \mathbb{C}[X]$,

$$\mu_a(P) = \frac{1}{m}\mathrm{Tr}(P(a)) = \frac{1}{m}\sum_{k=1}^{m} P(\lambda_k).$$

Thus, μ_a is (the sequence of moments of) the spectral measure of a.

(iii) One random matrix: if $a \in \mathbb{L}^\infty(\Omega, \mu, \mathrm{Mat}_m(\mathbb{C}))$ has eigenvalues $(\lambda_1(x), ..., \lambda_m(x))$, we have

$$\phi_m(P(a)) = \frac{1}{m}\int_\Omega \mathrm{Tr}\big(P(a)(\omega)\big)\mu(dx) = \frac{1}{m}\sum_{k=1}^{m}\int P\big(\lambda_k(\omega)\big)d\mu(\omega).$$

Thus, μ_a is (the sequence of moments of) the empirical spectral measure of a.

We introduce in the following the notion of *freeness* in a non-commutative probability space $(\mathcal{A}, \phi)$.

Definition 1.2.6. For $J \subset \mathbb{N}$, a family $\{\mathcal{A}_j\}_{j\in J}$ of sub-algebra of $\mathcal{A}$ is called *free* if for all $n \in \mathbb{N}$, all $j_1, ..., j_n \in J$ such that $j_k \neq j_{k+1}$ and all $(a_1, ..., a_n) \in \mathcal{A}_{j_1} \times ... \times \mathcal{A}_{j_n}$, we have:

$$\phi(a_1) = ... = \phi(a_n) = 0 \quad \Rightarrow \quad \phi(a_1...a_n) = 0.$$

Families of non-commutative variables are said to be free if the generating sub-algebras are free.
Moreover, remind that a C^*-algebra $\mathcal{A}$ is a unital algebra equipped with a norm $\|\cdot\|$ and an involution $*$ so that for $a, b \in \mathcal{A}$:

$$\|ab\| \leq \|a\|\,\|b\|, \quad \|a^*a\| = \|a\|^2,$$

and so that $\mathcal{A}$ is complete under this norm $\|\cdot\|$. An element a of $\mathcal{A}$ is said to be self-adjoint if $a^* = a$.

Definition 1.2.7. A quadruple $(\mathcal{A}, \|\cdot\|, *, \phi)$ is called a $\mathcal{C}^*$-*probability space* if $(\mathcal{A}, \|\cdot\|, *)$ is a $\mathcal{C}^*$-algebra and $\phi : \mathcal{A} \mapsto \mathbb{C}$ is a linear functional satisfying $\phi(1_{\mathcal{A}}) = 1$ and for all $a \in \mathcal{A}$, $\phi(a^*a) \geq 0$.

Consider a $\mathcal{C}^*$-probability space $(\mathcal{A}, *, \|\cdot\|, \phi)$ with a filtration, that is a family $(\mathcal{A}_t)_{t\geq 0}$ of sub-algebras closed for the weak topology-$*$ on $\mathcal{A}$ such that for all $0 \leq s \leq t$, we have $\mathcal{A}_s \subset \mathcal{A}_t$.

Definition 1.2.8. A *free Brownian motion* is a family $\left(W_t\right)_{t\geq 0}$ of non-commutative variables satisfying:

(i) for all $t \geq 0$, W_t is a self-adjoint element of $\mathcal{A}_t$, whose distribution is the semi-circular law σ_t of variance t, defined by

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

(ii) for all $s \leq t$, the element $W_t - W_s$ is free with $\mathcal{A}_s$ and its distribution is the semi-circular law of variance $t - s$.

The probabilistic interpretation

In the classical setting, one can show via the application of Ito's formula that the law of the solution of a well-posed stochastic differential equation (SDE) verifies a partial differential equation. By analogy, Biane and Speicher [9] describe the solution of the free Fokker-Planck equation (1.5) as the family of laws of a process satisfying a free SDE, that is a new kind of SDE, adapted to the non-commutative (free) framework, as shown in the following theorem

Theorem 1.2.9 (see Theorem 3.1 and Section 4.1 in[9])**.** *For a potential $V \in \mathcal{C}^3(\mathbb{R}, \mathbb{R})$, assume that there exist $a, b \in \mathbb{R}$ such that for all $x \in \mathbb{R}$,*

$$-x.V'(x) \leq a.x^2 + b.$$

Then, for any $\mathbf{x}_0$ whose distribution is compactly supported, the free SDE

$$d\mathbf{x}_s = d\mathbf{h}_s - \frac{1}{2}V'(\mathbf{x}_s)ds, \tag{1.7}$$

where $(\mathbf{h}_s)_{s\geq 0}$ is a free Brownian motion, admits a unique solution $(\mathbf{x}_s)_{s\geq 0}$ starting from $\mathbf{x}_0$. Moreover, the family of distributions of $(\mathbf{x}_s)_{s\geq 0}$ satisfies the free Fokker-Planck equation (1.6).

In particular, in case of null external potential, i.e. $V = 0$, the solution of the free SDE (1.7) at time $t > 0$ has a special form: $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{h}_t$. Since $\mathbf{x}_0$ and $\mathbf{h}_t$ are free, we obtain the distribution of $\mathbf{x}_0 + \mathbf{h}_t$ (sum of two non-commutative random variables), as in the following proposition.

Proposition 1.2.10. *(see also [1, Section 5.3.3]) If $\mathbf{x}_0$ admits the spectral measure μ_0, then $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{h}_t$ admits*

$$\mu_t = \mu_0 \boxplus \sigma_t, \tag{1.8}$$

as spectral measure, where the operation $\boxplus$ is free convolution and σ_t is the semi-circular law.

The free convolution operation $\boxplus$ has been introduced by Voiculescu in [63], and a brief definition of this operation is presented later in Section 2.2.1. Moreover, we consequently have the following proposition on the existence of density function of μ_t.

Proposition 1.2.11. *(see [7, Corollary 3]) Assume that the initial condition μ_0 possesses a density $p_0(\cdot) := p(0,\cdot)$ with respect to Lebesgue measure on $\mathbb{R}$, then for any $t > 0$, the solution μ_t of equation (1.6) has a density function $p(t,.)$ such that*

$$\partial_t p(t,x) = -\partial_x\big[p(t,x).\big(Hp(t,x)\big)\big], \quad x \in \mathbb{R}.$$

Thus, we have seen that the solution of this free Fokker-Planck equation (1.6) can be represented as free convolution (this will be studied in details later in Section 2.2) between the initial condition and the semi-circular distribution. As a result, an idea is then to apply free deconvolution techniques (developed recently by Arizmendi et al. [2]) but along with statistical consideration.

1.2.3 Deconvolution procedure

First, the Cauchy transform of a measure μ is defined as

$$G_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(x)}{z-x}, \quad \text{for } z \in \mathbb{C}^+,$$

where $\mathbb{C}^+$ is the set of complex numbers with positive imaginary part, and since G_μ does not vanish on $\mathbb{C}^+$ we defined $F_\mu(z) := 1/G_\mu(z)$, for $z \in \mathbb{C}^+$. For any $z \in \mathbb{C}$, let $\mathrm{Im}(z)$ denote the imaginary part of z, and for $\gamma > 0$ we define a sub-domain $\mathbb{C}_\gamma := \{z \in \mathbb{C}^+ : \mathrm{Im}(z) > \gamma\}$. Now, applying the subordination methods proposed in [2], who work in a general setting, to our case, we prove in Section 2.4 the following result on the subordination functions:

Theorem 1.2.12. *There exist unique subordination functions w_1 and w_{fp} from $\mathbb{C}_{2\sqrt{t}}$ onto $\mathbb{C}^+$ such that:*

(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

$$\lim_{y\to+\infty} \frac{w_1(i.y)}{i.y} = \lim_{y\to+\infty} \frac{w_{fp}(i.y)}{i.y} = 1.$$

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

$$F_{\mu_0}(z) = F_{\sigma_t}(w_1(z)) = F_{\mu_t}(w_{fp}(z)), \quad \text{or} \quad G_{\mu_0}(z) = G_{\sigma_t}(w_1(z)) = G_{\mu_t}(w_{fp}(z)); \quad (1.9)$$

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

$$w_{fp}(z) = z + w_1(z) - F_{\mu_0}(z); \quad (1.10)$$

(iv) Denote $h_{\sigma_t}(w) := w - F_{\sigma_t}(w) = t.G_{\sigma_t}(w)$ (see (2.7)), $\widetilde{h}_{\mu_t}(w) := w + F_{\mu_t}(w)$ on $\mathbb{C}^+$, and define a function $\mathcal{L}_z$ as

$$\begin{aligned} \mathcal{L}_z(w) :&= h_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z \\ &= t.G_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z; \end{aligned} \quad (1.11)$$

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

$$\mathcal{L}_z\big(w_{fp}(z)\big) = w_{fp}(z). \quad (1.12)$$

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

Furthermore, in our framework, the semi-circular measure σ_t is involved in the free convolution (see (1.8)) and one has an explicit formula for the Cauchy transform of σ_t as $G_{\sigma_t}(z) = \frac{z - \sqrt{z^2 - 4t}}{2t}$, for $z \in \mathbb{C}^+$. Thus, we can deduce that the subordination function $w_{fp}(z)$ at time t is related to G_{μ_t} by the functional equation

$$w_{fp}(z) = z + t.G_{\mu_t}(w_{fp}(z)), \quad z \in \mathbb{C}_{2\sqrt{t}}.$$

From this, we can recover G_{μ_0} with the formula $G_{\mu_0}(z) = G_{\mu_t}(w_{fp}(z))$ for $z \in \mathbb{C}_{2\sqrt{t}}$, and thus, recover p_0 (using the Stieltjes inversion formula, see Lemma 2.4.3 and (2.55) for more detail). More precisely, we prove in Section 2.5.1 that for any $\gamma > 2\sqrt{t}$, $f_{\mu_0 * \mathcal{C}_\gamma}$ which is the density of the classical convolution of μ_0 with $\mathcal{C}_\gamma$ satisfies

$$f_{\mu_0 * \mathcal{C}_\gamma}(x) = \frac{1}{\pi t}\left[\gamma - \mathrm{Im} w_{fp}(x + i\gamma)\right], \quad x \in \mathbb{R}, \tag{1.13}$$

where $\mathcal{C}_\gamma$ denotes the Cauchy distribution of parameter γ, defined by its density $f_\gamma(x) := \frac{\gamma}{\pi(x^2 + \gamma^2)}$, $x \in \mathbb{R}$. Then, from (1.13), we can see that to estimate p_0, the density of μ_0, it requires an estimation of the subordination function w_{fp} combined with a classical deconvolution step from a Cauchy distribution.

1.2.4 Dyson's Brownian motion and the convergence to solution of the Fokker-Planck equation

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

$$X_n(t) = X_n(0) + H_n(t), \qquad t \geq 0$$

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 $(d_j^n)_{j \in \{1,\ldots n\}}$ 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 (see Definition 2.3.1 for the definition of $H_n(t)$). 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 the density p_0. The observation consists in the matrix $X_n(t)$ at the fixed time t, from which we can compute the eigenvalues $(\lambda_1^n(t), \cdots, \lambda_n^n(t))$ and then the associated empirical measure. 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 $(\lambda_1^n(t), \cdots, \lambda_n^n(t))$ of $X_n(t)$ solve the following system of stochastic differential equations (SDE):

$$\mathrm{d}\lambda_k^n(t) = \frac{1}{\sqrt{n}}\mathrm{d}\beta_k(t) + \frac{1}{n}\sum_{j \neq k} \frac{\mathrm{d}t}{\lambda_k^n(t) - \lambda_j^n(t)}, \quad 1 \leq k \leq n,$$

where β_k are i.i.d. standard real Brownian motions. Note that this particle system is only introduced for interpretational purpose and will not be used directly in the sequel. Now, if we denote by

$$\mu_t^n := \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j^n(t)}, \tag{1.14}$$

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 the free Fokker-Planck equation, as stated in Proposition 2.3.5 in Section 2.3.

1.2.5 Contributions

This subsection presents the main results for the problem of statistical deconvolution of free Fokker-Planck equation at given fixed time, studied in Chapter 2.
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 (1.14). Then, for $z \in \mathbb{C}^+$, replacing in the procedure $G_{\mu_t}(z)$ by its natural estimator:

$$\widehat{G}_{\mu_t^n}(z) = \int_{\mathbb{R}} \frac{d\mu_t^n(\lambda)}{z-\lambda} = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{z-\lambda_j^n(t)}.$$

First, we provide below a statistical estimator $\widehat{w}_{fp}^n(z)$ for the subordination function w_{fp} presented in Section 1.2.3.

Theorem 1.2.13. *There exists a unique solution to the following functional equation in $w(z)$, for any $z \in \mathbb{C}_{2\sqrt{t}}$:*

$$\frac{1}{t}(w(z)-z) = \widehat{G}_{\mu_t^n}(w(z)). \tag{1.15}$$

This fixed-point is denoted by $\widehat{w}_{fp}^n$. Moreover, $|\widehat{w}_{fp}^n(z)-z| \leq \sqrt{t}$, for all $z \in \mathbb{C}_{2\sqrt{t}}$.

Since the Cauchy transform G_{μ_t} is not invertible on the whole domain $\mathbb{C}^+$, the subordination function $w_{fp}(z)$ will be defined only for $z \in \mathbb{C}_{2\sqrt{t}}$. As explained previously in Section 1.2.3, we estimate p_0 by combining a free deconvolution step via the use of $\widehat{w}_{fp}^n$ with then a classically statistical deconvolution step. Let us consider a bandwidth $h > 0$ and a kernel K. Remind that $\mathrm{Im}(z)$ denotes the imaginary part of $z \in \mathbb{C}$, according to (1.13), with $\gamma > 2\sqrt{t}$, we define our final estimator $\widehat{p}_{0,h}$ via its Fourier transform, denoted by $\widehat{p}_{0,h}^{\star}$, as follows:

$$\widehat{p}_{0,h}^{\star}(\xi) = e^{\gamma|\xi|}.K_h^{\star}(\xi).\frac{1}{\pi t}\left[\gamma - \mathrm{Im}\,\widehat{w}_{fp}^n(\cdot + i.\gamma)^{\star}(\xi)\right], \quad \xi \in \mathbb{R}, \tag{1.16}$$

where we define $K_h(\cdot) := \frac{1}{h}.K(\frac{\cdot}{h})$. Note that, as usual in nonparametric statistics, the last expression depends on $K_h^{\star}$, a regularization term defined through the Fourier transform of a kernel function K_h depending on a bandwidth parameter $h > 0$, see equation (2.64) in Definition 2.5.3 for more details.
From (1.16), one can observe that $\widehat{w}_{fp}^n$ plays a critical role in examining the behavior of $\widehat{p}_{0,h}$. Hence, it is reasonable to study first the consistency of $\widehat{w}_{fp}^n$, given in the following result.

Proposition 1.2.14. *Let $\gamma > 2\sqrt{t}$. Suppose p_0 satisfies the condition*

$$\int_{\mathbb{R}} \log(x^2+1)p_0(x)dx < +\infty. \tag{1.17}$$

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

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

We study theoretical properties of $\widehat{p}_{0,h}$ by deriving asymptotic rates of the mean integrated square error of $\widehat{p}_{0,h}$ in bias-variance decomposition. The study of the variance term is intricate and is based on sharp controls of the difference $\left|\widehat{w}_{fp}^n(z) - w_{fp}(z)\right|$ provided by Proposition 1.2.14.

Theorem 1.2.15. *Let*

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

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

$$\mu_0\big((\kappa, +\infty)\big) \leq \frac{C}{\kappa}. \tag{1.18}$$

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,

$$\mathbb{E}(\Sigma) \leq \frac{\gamma^8}{(\gamma^4 - 4t)^4} . \frac{C_{var}(t) . e^{\frac{2\gamma}{h}}}{n}. \tag{1.19}$$

Theorem 1.2.15 shows that the variance term is of order $\frac{1}{n}.\big(e^{\frac{2\gamma}{h}}\big)$ as desired for deconvolution with the Cauchy distribution with parameter γ, whereas the bias term is driven by the smoothness properties of the function p_0. In particular, when p_0 belongs to a class of super-smooth densities $\mathcal{S}(a, r, L)$ (which is a special case of class $\mathcal{S}(\delta, a, r, L)$ defined in (1.3) with $\delta = 0$), we obtain an upper-bound for the bias (see e.g. [25, Proposition 3]). One reason for considering the class $\mathcal{S}(a, r, L)$ is inspired from [16] in which the optimality of convergent rates have been proved so far for the case $0 < r < s$ with $\delta = 0$ (where s is the smoothness of super-smooth noise, as mentioned in Section 1.1.1). We indeed showed that:

$$\text{MISE} := \mathbb{E}\Big[\left\| \widehat{p}_{0,h} - p_0 \right\|^2 \Big] \leq L.e^{-2ah^{-r}} + \frac{\gamma^8}{(\gamma^2 - 4t)^4} . \frac{C_{\text{var}}(t) . e^{\frac{2\gamma}{h}}}{n}. \tag{1.20}$$

We obtain the following result.

Corollary 1.2.16. *Suppose that μ_0 satisfies Assumption (1.18) and the density p_0 belongs to the space $\mathcal{S}(a, r, L)$ for $a > 0$, $r > 0$ and $L > 0$. Then, for any $\gamma > 2\sqrt{t}$ and by choosing the bandwidth h minimizing the upper bound of (1.20) we have:*

$$\mathbb{E}\Big[\|\widehat{p}_{0,h} - p_0\|^2\Big] = \begin{cases} O\big(n^{-\frac{a}{a+\gamma}}\big) & \text{if } r = 1 \\ O\Big(\exp\Big\{ -\frac{2a}{(2\gamma)^r} \Big[\log n + (r-1)\log\log n + \sum_{i=0}^{k} b_i^* (\log n)^{r+i(r-1)} \Big]^r \Big\} \Big) & \text{if } r < 1 \\ O\Big(\frac{1}{n} \exp\Big\{ \frac{2\gamma}{(2a)^{1/r}} \Big[\log n + \frac{r-1}{r} \log\log n + \sum_{i=0}^{k} d_i^* (\log n)^{\frac{1}{r} - i\frac{r-1}{r}} \Big]^{1/r} \Big\} \Big) & \text{if } r > 1, \end{cases} \tag{1.21}$$

where the integer k is such that

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

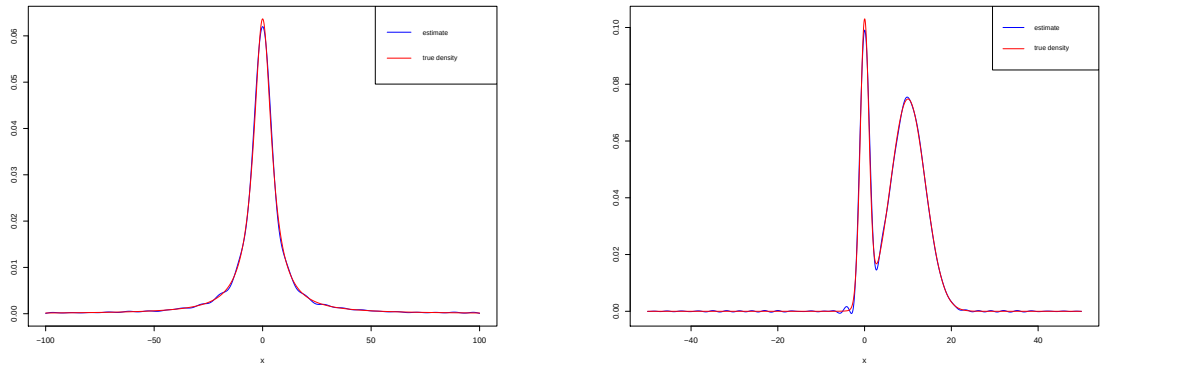
and where the constants b_i^* and d_i^* solve the following triangular system:

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

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

The case of Sobolev-ordinary type regularities leads to logarithmic rates of convergence (see Corollary 2.7.5 for more detail). For further discussions on the rates of convergences see Section 2.7.1.

Now, Figure 1.1 below illustrates two remarkable results of numerical experiment with the number of observations $n = 4000$ at time $t = 1$ for the case where $\mu_0 \sim$ Cauchy(0,5) in Figure 1.1a and for the case $\mu_0 \sim Y \times \mathcal{N}(0,1) + (1-Y) \times \mathcal{N}(10,4)$, with random variable $Y \sim$ Bernoulli(0.25) in Figure 1.1b. In these numerical simulations, optimal bandwidth h is chosen practically by using Cross-Validation. Accordingly, we can see that our estimators perform quite well especially when p_0 belongs to a super-smooth class of densities.



(a) $\mu_0 \sim$ Cauchy(0,5) (b) $\mu_0 \sim Y \times \mathcal{N}(0,1) + (1-Y) \times \mathcal{N}(10,4)$ with $Y \sim$ Bernoulli(0.25)

Figure 1.1: Estimation of p_0

1.3 Part II - Nonparametric estimation for regression with circular responses

This section presents a brief introduction for the framework that we study in Chapter 3.

1.3.1 Circular data

Directional statistics is the branch of statistics which deals with observations that are directions. In this work, we will consider more specifically *circular data* which arises whenever using a periodic scale to measure observations. These data are represented by points lying on the unit circle of $\mathbb{R}^2$ denoted in the sequel by $\mathbb{S}^1$. Note that the term *circular data* is also used to distinguish them from data supported on the real line $\mathbb{R}$ (or some subset of it), which henceforth are referred to as *linear data*. The key difficulty when dealing with such data is the curvature of the sample space since the unit circle is a non-linear manifold. Statistical methods for analyzing circular data have greatly evolved, and now circular counterparts exist for several data analysis techniques. For a comprehensive surveys on statistical methods for circular data, see Lee [77], Ley and Verdebout [78] and recent advances are collected in Pewsey and García-Portugués [83], and the references therein.

Before entering into statistical considerations, we introduce important notations. In the sequel, a point on $\mathbb{S}^1$ will not be represented as a two-dimensional vector $\mathbf{w} = (w_2, w_1)^\top$ with unit Euclidean norm but as an angle $\theta = \text{atan2}(w_1, w_2)$ defined as follows:

Definition 1.3.1. (see [78, Section 1.3]) The function atan2 : $\mathbb{R}^2 \backslash (0,0) \mapsto [-\pi, \pi)$ is defined

for any $(w_1, w_2) \in \mathbb{R}^2 \backslash (0,0)$ by

$$\operatorname{atan2}(w_1, w_2) := \begin{cases} \arctan\left(\frac{w_1}{w_2}\right) & \text{if} \quad w_2 \geq 0, w_1 \neq 0 \\ 0 & \text{if} \quad w_2 > 0, w_1 = 0 \\ \arctan\left(\frac{w_1}{w_2}\right) + \pi & \text{if} \quad w_2 < 0, w_1 > 0 \\ \arctan\left(\frac{w_1}{w_2}\right) - \pi & \text{if} \quad w_2 < 0, w_1 \leq 0, \end{cases}$$

with arctan taking values in $[-\pi/2, \pi/2]$. In particular, $\arctan(+\infty) = \pi/2$ and $\arctan(-\infty) = -\pi/2$.

In this definition, one has arbitrarily fixed the origin of $\mathbb{S}^1$ at $(1,0)^\top$ and uses the anti-clockwise direction as positive. Thus, a circular random variable can be represented as angle over $[-\pi, \pi)$. Observe that $\operatorname{atan2}(0,0)$ is undefined.

1.3.2 Statistical model

In this work, we focus on a nonparametric regression model with circular response and linear predictor. Assume that we have an i.i.d. sample $\{(X_j, \Theta_j)\}_{j=1}^n$ distributed as (X, Θ), where Θ is a circular random variable, and X is a random variable with density f_X supported on $\mathbb{R}$. We look for a function m which contains the dependence structure between the predictors X_j and the observations Θ_j. For $x \in \mathbb{R}$, let $m_1(x) = \mathbb{E}\big(\sin(\Theta)|X = x\big)$ and $m_2(x) = \mathbb{E}\big(\cos(\Theta)|X = x\big)$. This setting is considered in [80] for the first time. We investigate the nonparametric estimation of regression function m at a given point in $\mathbb{R}$.
Due to the difference of circular data from linear one by their periodic nature, a preliminary task is to define an appropriate risk measuring the difference between the two angles Θ and $m(X)$. The circular context leads to a natural approach of using trigonometric functions. For two angles $\theta_1, \theta_2 \in [-\pi, \pi)$, we use the angular distance $d_c(\theta_1, \theta_2) := \big(1 - \cos(\theta_1 - \theta_2)\big)$ to measure the difference. Notice that this angular distance corresponds to the usual Euclidean norm in $\mathbb{R}^2$. Indeed, the angles θ_1 and θ_2 determines the corresponding points $(\cos\theta_1, \sin\theta_1)$ and $(\cos\theta_2, \sin\theta_2)$ respectively on the unit circle $\mathbb{S}^1$. Then, the usual squared Euclidean norm in $\mathbb{R}^2$ reads $\big(\cos\theta_1 - \cos\theta_2\big)^2 + \big(\sin\theta_1 - \sin\theta_2\big)^2 = 2.\big[1 - \cos(\theta_1 - \theta_2)\big] = 2.d_c(\theta_1, \theta_2)$.
Hence, using the angular distance, we look for a function m as the minimizer of the following risk (see also Section 3.1):

$$L(m(X)) := \mathbb{E}\big[1 - \cos(\Theta - m(X))\big|X\big].$$

Actually, given $x \in \mathbb{R}$, the risk $L(m(x))$ is minimized by $m(x) = \operatorname{atan2}(m_1(x), m_2(x))$, see Section 3.1 for more details of computation. In conclusion, the circular nature of the response is taken account by the arctangent of the ratio of the sine and cosine components of the conditional trigonometric moment of Θ given X.
Recall that the statistical regression model is considered for the first time by Di Marzio et al. [80]. In [80], considering that the covariates supported on $[0,1]$, the authors propose a Nadaraya-Watson and a local linear polynomial estimators to estimate $\operatorname{atan2}(m_1(x), m_2(x))$. Under assumptions on the regularity of m_1 and m_2, by using Taylor expansion for the function arctangent, they derive expressions for asymptotic bias and variance, however, the remainder terms in the expansions are not controlled sharply. As for the more recent work of Meilan Vila et al. [82], they studied the multivariate setting $[0,1]^d$ with the same estimators and proofs technics. In both papers, rates of convergence and adaptation still remain as open questions, in addition, kernel bandwidth is selected by cross-validation in practice. By adaptation, we recall that the estimators do not require the specification of the regularity of the regression function which is crucial from a practical point of view. They consider the same regularity for

m_1 and m_2 which can be seen as restricted, whereas it would be more natural to have different regularities for m_1 and m_2. Furthermore, their estimators only depend on one single bandwidth whereas, it would be more natural to consider two bandwidths due to possible different regularities on m_1 and m_2.
In view of this, our motivation is to fill the gap in the literature. In this work, consider predictors supported on $\mathbb{R}$, we first propose new warped kernel estimators $\widehat{m}_{\widehat{h}_1}(x)$ and $\widehat{m}_{\widehat{h}_2}(x)$ for $m_1(x)$ and $m_2(x)$ at a point $x \in \mathbb{R}$ respectively, where bandwidths $\widehat{h}_1$ and $\widehat{h}_2$ are selected adaptively by the Goldenshluger Lepski rule. Then, we set an estimator $\widehat{m}_{\widehat{h}}(x) := \text{atan2}\big(\widehat{m}_{\widehat{h}_1}(x), \widehat{m}_{\widehat{h}_2}(x)\big)$ for the regression function $m(x)$ with $\widehat{h} := (\widehat{h}_1, \widehat{h}_2)$. We obtain an upper-bound for the estimator $\widehat{m}_{\widehat{h}}(x)$ for the pointwise squared risk and over Hölder classes.
A more detailed description of the framework that we examine is given in Section 3.1 of Chapter 3. The summary of the main theoretical and numerical results are presented in Section 1.3.3.

1.3.3 Contributions

This subsection presents the main results for the problem of nonparametric estimation for regression with circular responses, studied in Chapter 3.
Let $K : \mathbb{R} \to \mathbb{R}$ be an integrable function such that K is compactly supported, $K \in \mathbb{L}^\infty(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. We say that K is a kernel if it satisfies $\int_{\mathbb{R}} K(y)dy = 1$. Assume that F_X is invertible, we introduce auxiliary functions $g_1, g_2 : (0,1) \mapsto \mathbb{R}$ defined by

$$g_1 := m_1 \circ F_X^{<-1>}, \quad \text{and} \quad g_2 := m_2 \circ F_X^{<-1>}$$

in such a way that $m_1 = g_1 \circ F_X$ and $m_2 = g_2 \circ F_X$. Suppose that the design distribution is known, for $u \in F_X(\mathbb{R})$, we set estimators :

$$\widehat{g}_{1,h_1}(u) := \frac{1}{n}\sum_{j=1}^{n} \sin(\Theta_j).K_{h_1}(u - F_X(X_j)), \quad \text{and} \quad \widehat{g}_{2,h_2}(u) := \frac{1}{n}\sum_{j=1}^{n} \cos(\Theta_j).K_{h_2}(u - F_X(X_j)), \tag{1.22}$$

to estimate g_1 and g_2 respectively, and $h_1, h_2 > 0$ are bandwidths of kernel $K_{h_1}(\cdot)$ and $K_{h_2}(\cdot)$ respectively. Moreover, as a consequence, for $x \in \mathbb{R}$, the estimators for m_1 and m_2 are

$$\widehat{m}_{1,h_1}(x) = \big(\widehat{g}_{1,h_1} \circ F_X\big)(x), \quad \text{and} \quad \widehat{m}_{2,h_2}(x) = \big(\widehat{g}_{2,h_2} \circ F_X\big)(x). \tag{1.23}$$

Denote $h := (h_1, h_2)$, we then set the estimator for $m(x)$ at $x \in \mathbb{R}$ by

$$\widehat{m}_h(x) := \text{atan2}\big(\widehat{m}_{1,h_1}(x), \widehat{m}_{2,h_2}(x)\big).$$

Now, for the method to get an automatic bandwidth selection, we propose a data-driven selection rule inspired from Goldenshluger and Lepski [71]. This selection procedure is described in Section 3.2.2 of Chapter 3.
Let $\widehat{g}_{1,\widehat{h}_1}$ and $\widehat{g}_{2,\widehat{h}_2}$ be the adaptive estimators of g_1 and g_2 respectively where $\widehat{h}_1$ and $\widehat{h}_2$ are bandwidths selected by our data-driven selection rule (see Section 3.2.2). We establish an oracle-type inequality for $\widehat{g}_{1,\widehat{h}_1}$ and $\widehat{g}_{2,\widehat{h}_2}$ (see Proposition 3.2.2 in Section 3.2.3) which highlights the bias-variance decomposition of the pointwise squared risk.
Then, denote $\widehat{h} := (\widehat{h}_1, \widehat{h}_2)$, we define the adaptive estimator of m at $x \in \mathbb{R}$ by

$$\widehat{m}_{\widehat{h}}(x) := \text{atan2}\big(\widehat{m}_{1,\widehat{h}_1}(x), \widehat{m}_{2,\widehat{h}_2}(x)\big). \tag{1.24}$$

We also consider the following assumption on kernel K:

Assumption 1.3.2. The kernel K is of order $\mathcal{L} \in \mathbb{R}_+$, i.e.

(i) $\int_{\mathbb{R}}(1+|y|^{\mathcal{L}}).|K(y)|dy < \infty$.

(ii) $\forall k \in \{1, ..., \lfloor \mathcal{L} \rfloor\}$, $\int_{\mathbb{R}} y^k.K(y)dy = 0$.

Since the function $\text{atan2}(w_1, w_2)$ is undefined when $w_1 = w_2 = 0$, it is reasonable to consider the following assumption:

Assumption 1.3.3. For given $x \in \mathbb{R}$, assume that

$$|m_1(x)| = \big|g_1(F_X(x))\big| \geq \delta_1 > 0, \quad \text{and} \quad |m_2(x)| = \big|g_2(F_X(x))\big| \geq \delta_2 > 0.$$

Let $\mathcal{H}(\beta, L)$ denote the Hölder class with $\beta, L > 0$ as in Definition 3.2.4. Moreover, let $c_{0,1}$ and $c_{0,2}$ are tuned constants constituted in the selection rule of the adaptive bandwidths $\widehat{h}_1$ and $\widehat{h}_2$ respectively (see Section 3.2.2). The following theorem gives an upper-bound for the mean squared error of the estimator $\widehat{m}_{\widehat{h}}$ defined in (1.24).

Theorem 1.3.4. *Let $\beta_1, \beta_2, L_1, L_2 > 0$. Suppose that g_1 belongs to a Hölder class $\mathcal{H}(\beta_1, L_1)$, g_2 belongs to $\mathcal{H}(\beta_2, L_2)$, the kernel K satisfies Assumption 1.3.2 with an index $\mathcal{L} \in (\mathbb{R}_+)$ such that $\mathcal{L} \geq \max(\beta_1, \beta_2)$. Let $q \geq 1$, and suppose that $\min\{c_{0,1}; c_{0,2}\} \geq 16\big(2+q\big)^2.\big(1+\|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$. Then, under Assumption 1.3.3, for n sufficiently large,*

$$\mathbb{E}\Big[\big|\widehat{m}_{\widehat{h}}(x) - m(x)\big|^2\Big] \leq 32.\pi^2.\Big(\frac{1}{\delta_2^2} + \frac{1}{\delta_1^2}\Big).(C_1^2 + C_2^2).\max\left\{\psi_n(\beta_1)^2, \psi_n(\beta_2)^2\right\},$$

where C_1 is a constant (depending on $\beta_1, L_1, c_{0,1}, K$) and C_2 is a constant (depending on $\beta_2, L_2, c_{0,2}, K$), along with $\psi_n(\beta_1) = \big(\log(n)/n\big)^{\frac{\beta_1}{2\beta_1+1}}$ and $\psi_n(\beta_2) = \big(\log(n)/n\big)^{\frac{\beta_2}{2\beta_2+1}}$.

Observe that if $\beta_1 = \beta_2 = \beta$, then we obtain the rate $\psi_n(\beta) = (\log n/n)^{\beta/(2\beta+1)}$, which is the optimal rate for adaptive univariate regression function estimation and pointwise risk (see e.g. Section 2 in [66]). We conjecture that the rate obtained in Theorem 1.3.4 is optimal even if it remains an open problem.

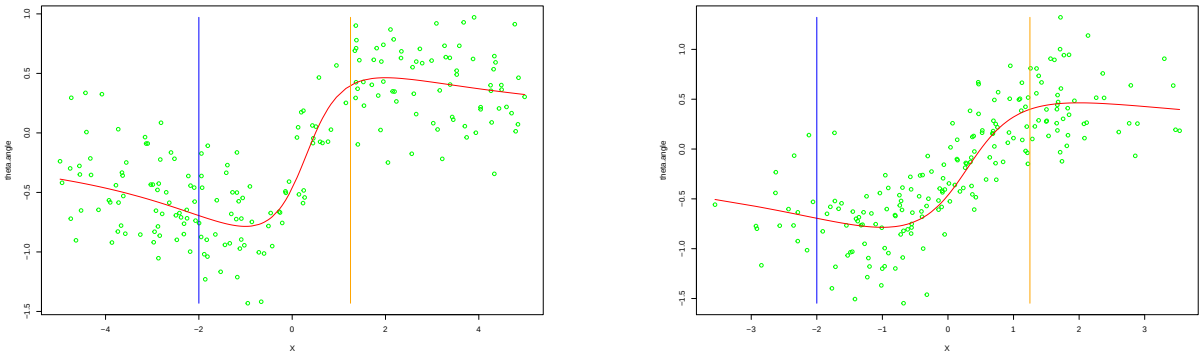
Now, we describe briefly some numerical simulations to illustrate the performance of our estimator $\widehat{m}_{\widehat{h}}$. Consider the following model

$$\text{M1.} \quad \Theta = \text{atan2}\big(2X - 1, X^2 + 2\big) + \zeta \pmod{2\pi}; \tag{1.25}$$

where the circular errors $\{\zeta_j\}_{j=1,n}$ are drawn from a von Mises distribution $vM(0, 10)$, with a location parameter $0 \in [-\pi, \pi)$ and concentration of 10, (see Section 3.3 for more details for definition of von Mises distribution). The implementation is made for the case of uniform distribution $X \sim U([-5, 5])$ and also the case of Gaussian distribution $X \sim \mathcal{N}(0, 1.5)$. In both cases, for the sample size $n = 200$, we aim at estimating $m(x)$ at $x = -2$ and at another point $x = 1.25$. An illustration for simulated data is shown in Figure 1.2. The final estimators are computed using the empirical counterpart $\widehat{F}_n(y) := \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{X_j \leq y}$, with $y \in \mathbb{R}$, of the warping function F_X, and the Epanechnikov kernel $K(v) = \frac{3}{4}.(1 - v^2).\mathbb{1}_{|v|\leq 1}$, $v \in \mathbb{R}$.

First of all, in the bandwidth selection procedure, consider $c_{0,1} = c_{0,2}$, we carry out a preliminary implementation on model M1 to calibrate these two tuning constants.

After that, we compute the adaptive estimator $\widehat{m}_{\widehat{h}}(x)$ (as defined in (1.24)). Moreover, to compare with our adaptive estimator, we compute the Nadaraya-Watson (NW) estimator denoted by $\widehat{m}_h^{NW}$ and local linear (LL) estimator (see [80, Section 4.2]) denoted by $\widehat{m}_h^{LL}$ of m as well.



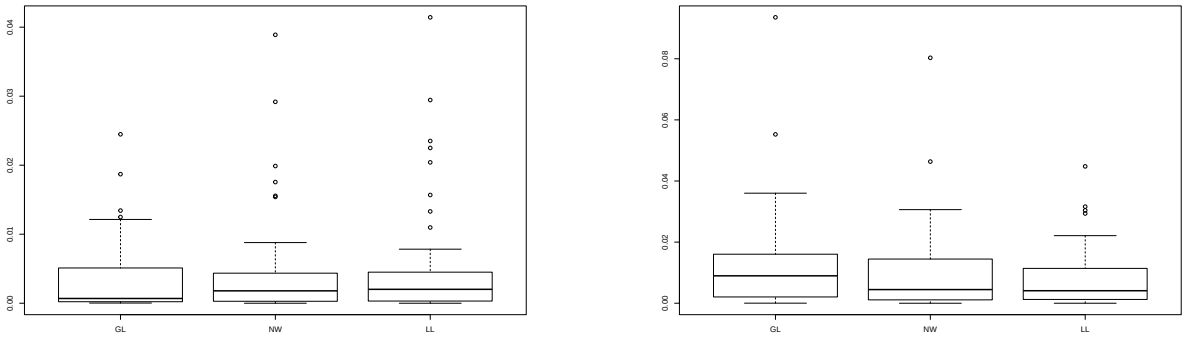
(a) Model M1 : $X_j \sim U([-5,5])$ (b) Model M1 : $X_j \sim \mathcal{N}(0,1.5)$

Figure 1.2: An illustration for simulated data $\{\Theta_j\}_{j=1}^n$ (green points) of model-M1 with $n = 200$. The red curve presents the true regression function m, while the blue vertical line displays point $x = -2$ and the orange vertical line displays point $x = 1.25$ where we aim at estimating $m(x)$.

In addition, Cross-Validation is used to select bandwidth $h > 0$ for $\widehat{m}_h^{NW}$ and $\widehat{m}_h^{LL}$.

In the graphics below, we denote by "GL", "NW" and "LL" the pointwise squared risk (computed with 50 Monte Carlo repetitions) corresponding to $\widehat{m}_{\widehat{h}}(x)$, $\widehat{m}_h^{NW}$ and $\widehat{m}_h^{LL}$ respectively. The comparison between these pointwise squared risks is shown in boxplots in Figures 1.3 and 1.4. These numerical results show that the performances of our estimator are quite satisfying.

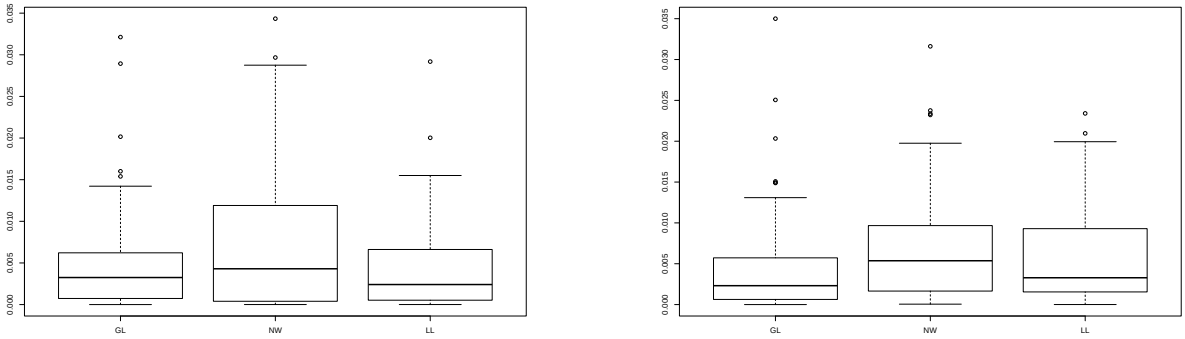
Eventually, we refer to Section 3.3 for more details of numerical simulations.



(a) Model M1: estimation at $x = -2$

(b) Model M1: estimation at $x = 1.25$

Figure 1.3: Boxplot for 50 runs of Monte Carlo resampling with sample size $n = 200$, in case $X \sim U([-5, 5])$, in comparison with NW and LL (using Epanechnikov kernel).



(a) Model M1: estimation at $x = -2$

(b) Model M1: estimation at $x = 1.25$

Figure 1.4: Boxplot for 50 runs of Monte Carlo resampling with sample size $n = 200$, in case $X \sim \mathcal{N}(0, 1.5)$, in comparison with NW and LL (using Epanechnikov kernel).

Introduction (Bref résumé de l'introduction en version française)

Ce chapitre présente une introduction générale de la thèse ainsi que les principales contributions. Il fait aussi office de "résumé en français" requis par l'École Doctorale, d'où la différence dans la langue de rédaction avec les autres chapitres.

Dans cette thèse, nous étudions deux problèmes inverses dans le cadre statistique liés à l'estimation de densité non paramétrique et à l'estimation de régression non paramétrique, cependant dans de nouveaux contextes :

- Le premier problème concerne l'estimation de la densité de la condition initiale aléatoire d'une équation aux dérivées partielles (EDP) de Fokker-Planck non-linéaire à partir des observations liées au temps $t > 0$. C'est la première fois, à notre connaissance, qu'un tel problème est examiné. Pour les EDP linéaires, cela a été considéré par Pensky and Sapatinas [48] [49]. Une approche pour traiter la non-linéarité de l'équation de Fokker-Planck est de travailler avec une interprétation probabiliste libre de l'EDP, impliquant de nouvelles techniques de déconvolution libre.

- Dans le deuxième problème, l'estimation non paramétrique pour la régression avec des réponses circulaires est examinée. Ce modèle de régression est inspiré des travaux de Taylor et al. [80]. Nous proposons un nouvel "warped" estimateur à noyau pour la fonction de régression, dont la consistance, les vitesses de convergence et d'adaptation sont soigneusement étudiés.

Les deux problèmes ont des difficultés spécifiques. Il est raisonnable de passer d'abord en revue les cadres standard correspondants d'estimation non paramétrique dans la section 1.4 ci-dessous. Après cela, nous présentons le contenu de chaque chapitre de cette thèse et soulignons les différences avec la littérature classique.

Tout au long de ce manuscrit, pour évaluer l'optimalité de nos résultats, nous considérerons le cadre minimax. Plus précisément, une vitesse de convergence minimax ψ_n pour estimer une fonction f sur la classe $\mathcal{D}$ avec la fonction de perte ℓ est telle que:

1. il existe un estimateur $\widehat{f}_n$ dit optimal au sens minimax atteignant cette vitesse, c'est-à-dire
$$\limsup_{n\to+\infty} \sup_{f\in\mathcal{D}} \psi_n^{-1}.\mathbb{E}\left[\ell(\widehat{f}_n, f)\right] \leq C < \infty;$$
2. aucun autre estimateur ne peut atteindre une meilleure vitesse
$$\liminf_{n\to+\infty} \inf_{\widetilde{f}_n} \sup_{f\in\mathcal{D}} \psi_n^{-1}.\mathbb{E}\left[\ell(\widetilde{f}_n, f)\right] \geq c > 0,$$
pour certaines constantes $C > 0, c > 0$ et où l'infimum est pris sur tous les estimateurs possibles de f.

1.4 Estimation statistique non paramétrique

Ce chapitre commence par une introduction générale sur deux problèmes statistiques standards en estimation non paramétrique, à savoir l'estimation d'une fonction de densité dans le modèle de convolution classique de la section 1.4.1 et l'estimation d'une fonction de régression dans la section 1.4.2.

1.4.1 Estimation de déconvolution de densité non paramétrique

Dans cette sous-section, nous passons brièvement en revue quelques résultats standard sur la déconvolution statistique classique pour estimer une fonction de densité de probabilité. Cela rappellera au lecteur des résultats utiles pour faciliter les connexions pertinentes avec les vitesses de convergence obtenues pour le problème de déconvolution libre présenté au Chapitre 2. En effet, la résolution d'un problème de déconvolution libre par la méthode des fonctions de subordination conduit à une convolution (au sens classique) de la densité objective avec une distribution de Cauchy (centrée). De plus, nous nous limitons ici au cadre à 1-dimension par souci de simplicité.
Dans le cadre classique (voir par exemple [25], [43], [59] et les références qui s'y trouvent), nous considérons le modèle dit de convolution

$$Y_j = X_j + \varepsilon_j, \quad j = 1, ..., n, \tag{1.26}$$

où X_j sont des variables aléatoires indépendantes identiquement distribuées (i.i.d.) avec une densité de probabilité inconnue f_X par rapport à la mesure de Lebesgue sur $\mathbb{R}$, les variables aléatoires de bruit ε_j sont i.i.d. avec une densité de probabilité **connue** f_ε par rapport à la mesure de Lebesgue sur $\mathbb{R}$, et $(\varepsilon_1, ..., \varepsilon_n)$ est indépendant de $(X_1, ..., X_n)$. Le problème de déconvolution de densité consiste à estimer la fonction f_X à partir d'observations $Y_1, ..., Y_n$. Soit f_Y la densité de probabilité des variables Y_j, alors on a $f_Y = f_X * f_\varepsilon$, où l'opérateur $*$ désigne le convolution classique définie pour $y \in \mathbb{R}$ par

$$\big(f_X * f_\varepsilon\big)(y) := \int_{\mathbb{R}} f_X(y - x).f_\varepsilon(x)dx.$$

L'un des outils majeurs de la déconvolution de la densité réside dans le domaine de l'analyse de Fourier (voir par exemple [50]). En notant la transformée de Fourier d'une fonction $g \in \mathbb{L}^1(\mathbb{R})$ par $g^\star$, définie pour tout $\xi \in \mathbb{R}$ comme

$$g^\star(\xi) := \int_{\mathbb{R}} g(x).e^{-ix\xi}dx.$$

Alors puisque $f_Y = f_X * f_\varepsilon$, on a

$$f_Y^\star(\xi) = f_X^\star(\xi).f_\varepsilon^\star(\xi), \quad \xi \in \mathbb{R}. \tag{1.27}$$

Ainsi, l'utilisation de la transformée de Fourier est motivée par le fait qu'elle transforme $*$, l'opérateur de convolution, en simple multiplication des fonctions caractéristiques. Quant à la quantité $f_Y^\star(\xi) = \mathbb{E}(e^{-i.\xi.Y})$, elle est simplement estimée par sa version empirique.

Remarque. Notez que la transformée de Fourier ci-dessus est définie pour $g \in \mathbb{L}^1(\mathbb{R})$. En fait, grâce au théorème de Plancherel, la transformée de Fourier peut être étendue à $g \in \mathbb{L}^2(\mathbb{R})$, voir par exemple [50, Chapitre 9].

La littérature sur ce problème est en fait extensif. On peut citer les travaux de Carroll et Hall [19], Devroye [29], Liu et Taylor [39], Masry [41], Stefanski et Carroll [53], Zhang [65], Hesse [35], Cator [20], Delaigle et Gijbels [28], pour principalement les méthodes du noyau, Koo [36] pour une méthode spline, Pensky et Vidakovic [47] pour les stratégies d'ondelettes, et de plus, Comte et al. [24] pour les stratégies de projection adaptative. Enfin et surtout, le livre de Meister [43] constitue une revue complète des problèmes de déconvolution en statistique non paramétrique.

Hypothèses

Dans le problème de déconvolution de densité, deux facteurs déterminent la précision de l'estimation: la régularité de la densité f_X à estimer et la régularité de la densité de bruit f_ε. En conséquence, présentons les hypothèses habituelles pour ce cadre. Premièrement, nous supposons que le bruit est tel que pour tout $x \in \mathbb{R}$, $f_\varepsilon^\star(x) \neq 0$ et vérifie l'hypothèse suivante.

Hypothèse (N1)**.** Il existe des constantes $s \geq 0$, $b \geq 0$, $\alpha \in \mathbb{R}$ ($\alpha > 0$ si $s = 0$) et $k_0, k_1 > 0$ telles que

$$k_0.(x^2+1)^{-\alpha/2}.e^{-b|x|^s} \leq \left|f_\varepsilon^\star(x)\right| \leq k_1.(x^2+1)^{-\alpha/2}.e^{-b|x|^s}.$$

Lorsque $b > 0$ et $s > 0$, la transformée de Fourier du bruit a une décroissance exponentielle et sa densité est infiniment dérivable sur $\mathbb{R}$. On parle alors de bruit "super-smooth" (SS). Par exemple, la distribution de Cauchy qui sera particulièrement intéressante dans le chapitre 2 satisfait l'hypothèse (N1) avec $s = 1$. Lorsque $s = 0$, la densité de bruit est appelée "ordinary smooth" (OS).

Quant à la densité f_X, nous supposerons qu'elle appartient à l'espace suivant

$$\mathcal{S}(\delta, a, r, L) = \left\{ f \text{ une densité sur } \mathbb{R} : \int |f^\star(x)|^2.(x^2+1)^\delta.\exp(2a|x|^r)dx \leq L \right\}, \qquad (1.28)$$

avec $r \geq 0, a \geq 0, \delta \geq 0$ et $L > 0$. La même terminologie que le bruit est utilisée pour décrire la régularité de la fonction f_X : super-smooth (SS) si $r > 0$, ordinary smooth (OS) sinon. Notez que $r = 0$ correspond à une classe de Sobolev.

Estimateurs

Une méthode standard pour résoudre le problème de déconvolution classique est la méthode du noyau (voir par exemple [25],[43]). L'estimateur à noyau classique est défini comme:

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - Y_i}{h}\right),$$

où le noyau K est tel que sa transformée de Fourier inverse a l'expression suivante

$$K^\star(x) = \frac{\mathbb{1}_{|x|\leq 1}}{f_\varepsilon^\star(-x/h)}.$$

Dans [16], Butucea and Tsybakov ont prouvé que si $f_X \in \mathcal{S}(\delta, a, r, L)$ et sous l'hypothèse (N1), le MISE (Mean Integrated Squared Error) de $\widehat{f}_n(x)$ qui peut être classiquement décomposé en un biais et un terme de variance majoré par:

$$\text{MISE} \ = \mathbb{E}\left[\|f_X - \hat{f}_n\|^2_{\mathbb{L}^2(\mathbb{R})}\right] = O(h^{2\delta}.\exp(-2a/h^r)) + \frac{h^{s-1-2\alpha}}{n}.\exp(2b/h^s).$$

Cette dernière borne du MISE sera particulièrement intéressante par rapport à la borne que nous avons obtenue dans (2.108) au Chapitre 2. A noter qu'une autre méthode développée par Comte et al. [24] est l'estimateur de projection conduisant à une borne supérieure similaire mais réalisant une adaptation. Par adaptation, nous entendons que la procédure n'a pas besoin de la connaissance de la régularité inconnue de la fonction à estimer.

Les vitesses de convergence

Rappelons maintenant les vitesses de convergence bien connues obtenues dans la littérature pour le MISE. Comme dit dans le préambule de cette section, elles dépendent fortement à la fois des régularités du bruit et de la densité à récupérer. On peut distinguer 4 cas regroupés dans le tableau 1.2. Comme on peut le voir dans le tableau 1.2, plus le bruit est "smooth", plus

	f_ε OS	f_ε SS
f_X OS	$n^{\frac{-2\delta}{2\delta+2\alpha+1}}$	$(\log n)^{\frac{-2\delta}{s}}$
f_X SS	$\frac{(\log n)^{\frac{2\alpha+1}{r}}}{n}$	$\frac{(\log n)^{\frac{2\alpha+1}{r}}}{n} < . < (\log n)^{\frac{-2\delta}{s}}$

Table 1.2: des vitesses de convergence pour le MISE

le problème de déconvolution est difficile. À l'exception du cas où f_X et f_ε sont super-smooth, ces vitesses sont connues pour être minimax optimale (voir [15]).

Attardons-nous maintenant sur le cas en bas à droite (f_X et f_ε SS) car il nous concernera particulièrement au chapitre 2. Dans le tableau 1.2 nous n'avons pas volontairement spécifié les vitesses correspondants car elles sont assez complexes à présenter brièvement. Pour plus de détails, ils ont été abordés dans [37] ou en Corollaire 2.7.3 (Chapitre 2) pour le cas particulier d'un bruit de Cauchy. Néanmoins, à ce stade du manuscrit, nous pouvons souligner que considérer une densité "super-smooth" à estimer permet d'améliorer considérablement les vitesses de convergence en présence d'un bruit "super-smooth". Les vitesses sont plus rapides que les vitesses logarithmiques standard lorsqu'il s'agit de f_X dans une classe de Sobolev. De plus, d'un point de vue technique, considérer f_X et f_ε SS implique des compromis de biais-variance non standard qui nécessitent de nouvelles méthodes pour prouver les bornes inférieures. L'optimalité de ces vitesses a été prouvée jusqu'à présent dans le cas $r = s = 1$ (voir [58]) ou le cas $0 < r < s$ avec $\delta = 0$ (voir [16]).

1.4.2 Régression non paramétrique avec un design aléatoire

Dans cette sous-section, nous introduisons brièvement le problème de l'estimation non paramétrique classique pour la régression avec un design aléatoire, et nous nous limitons à un cadre 1-dimension pour des raisons de simplicité. Formellement, dans le cadre classique (voir par exemple [43], [59]), nous considérons (X, Y) comme un couple de variables aléatoires prenant des valeurs sur $\mathbb{R}$ telles que

$$Y = m(X) + \varepsilon, \tag{1.29}$$

où ε est une variable aléatoire à valeur réelle centrée de variance finie et indépendante de X. Dans ce modèle, la quantité observée est notée Y qui est appelée variable *response*, X est appelée *covariate* et ε dénote l'erreur de régression. Supposons que nous soyons capables d'observer un échantillon i.i.d. $\{(X_j, Y_j)\}_{j=1}^n$ ayant la même distribution que (X, Y). L'ensemble de valeur $\{X_1, ..., X_n\}$ est appelé *design* et ici le design est aléatoire. De plus, nous supposons que X a une fonction de densité f_X (par rapport à la mesure de Lebesgue), qui est

appelée la densité de design, et en notant $D \subset \mathbb{R}$ le support de f_X.
La fonction $m(X) = \mathbb{E}\big(Y|X\big)$ est appelée la fonction de régression de Y sur X, qui est généralement inconnue. Etant donné $x \in D$, le but est d'estimer la fonction $m(x)$ à partir des données $\big\{\big(X_j, Y_j\big)\big\}_{j=1}^n$. Le lemme suivant montre que la fonction de régression m est bien le meilleur prédicteur de la variable de réponse Y basé sur l'observation de la covariate X au sens du risque quadratique.

Lemme 1.4.1. *Pour toutes les fonctions mesurables $g : \mathbb{R} \to \mathbb{R}$ avec $\mathbb{E}\big[\big(g(X)\big)^2\big] < \infty$,*

$$\mathbb{E}\big[|Y - m(X)|^2\big] \leq \mathbb{E}\big[|Y - g(X)|^2\big].$$

Proof. On a

$$\begin{aligned}\mathbb{E}\big[|Y - g(X)|^2\big] &= \mathbb{E}\big[|Y - m(X)|^2\big] + \mathbb{E}\big[|m(X) - g(X)|^2\big] + 2.\mathbb{E}\big[\big(Y - m(X)\big).\big(m(X) - g(X)\big)\big] \\ &\geq \mathbb{E}\big[|Y - m(X)|^2\big] + 2.\mathbb{E}\big[\big(\mathbb{E}[Y - m(X)|X]\big).\big(m(X) - g(X)\big)\big] \\ &= \mathbb{E}\big[|Y - m(X)|^2\big] \quad (\text{car } \mathbb{E}[Y - m(X)|X] = \mathbb{E}[Y|X] - m(X) = 0).\end{aligned}$$

□

Estimateurs à noyau de la fonction de régression

Une approche non paramétrique commune pour construire un estimateur pour la fonction de régression m consiste à appliquer une méthode de noyau. Considérons une fonction noyau intégrable $K : \mathbb{R} \to \mathbb{R}$ telle que $\int_{\mathbb{R}} K(y)dy = 1$ et nous définissons $K_h(\cdot) := \frac{1}{h}.K(\frac{\cdot}{h})$ pour $h > 0$ la fenêtre de K. Parmi les méthodes du noyau, le populaire estimateur Nadaraya-Watson (NW) de m est défini pour le modèle (1.29) à $x \in D$ par:

$$\widehat{m}_h^{NW}(x) := \frac{\sum_{j=1}^n Y_j.K_h(x - X_j)}{\sum_{j=1}^n K_h(x - X_j)}, \quad \text{si } \sum_{j=1}^n K_h(x - X_j) \neq 0,$$

et $\widehat{m}_h^{NW}(x) = 0$ sinon. Si l'on veut obtenir une estimation adaptative de m (c'est-à-dire que les estimateurs ne nécessitent pas la spécification de la régularité de m), il faut la sélection automatique de la fenêtre h, puis la forme ratio de l'estimateur NW suggère de sélectionner deux fenêtres: une pour le numérateur et une pour le dénominateur. De plus, lorsque le design X est très irrégulier (par exemple lorsqu'un "trou" se produit dans les données), une forme de ratio peut conduire à une instabilité (voir par exemple Pham Ngoc [84]). De plus, pour l'analyse de régression avec perte L_2, il est standard de supposer que le support D de f_X est un sous-ensemble compact de $\mathbb{R}$. À notre connaissance, Kohler et al. [76] sont les premiers à proposer des résultats théoriques sans hypothèse de bornage sur le support de la densité du design. À son tour, cela nécessite une hypothèse de moment sur le design X (voir l'hypothèse $(A4)$ dans [76]).

Estimateurs à noyau "warped"

Les estimateurs à noyau "warped" introduits par Yang (1981,[88]), Stute (1984,[87]) et développés par Kerkyacharian et Picard (2004, [75]) évitent la forme-ratio et nécessitent très peu d'hypothèses sur le support de f_X. La méthode "warped" est basée sur l'introduction de la fonction auxiliaire $g = m \circ F_X^{<-1>}$, où $F_X : y \in \mathbb{R} \to \mathbb{P}(X \leq y)$ est la fonction de distribution cumulative (c.d.f.) de X et $F_X^{<-1>}$ son inverse. L'idée de la procédure d'estimation est de proposer d'abord un estimateur $\widehat{g}$ pour g, puis, la fonction de régression m est estimée en utilisant $\widehat{g} \circ \widehat{F}_n$, où $\widehat{F}_n$ est le c.d.f. empirique de X. Un avantage des méthodes de noyau "warped" est qu'elles n'impliquent

pas de ratio (voir Section 3.2.1 pour plus de détails). Pour sélectionner la fenêtre du noyau, nous appliquons la règle de sélection adaptative de la fenêtre unique dans l'esprit de Goldenshluger et Lepski [71]. Récemment, cette stratégie est appliquée dans le cadre de la régression en utilisant des méthodes à noyau pour construire des estimateurs non paramétriques adaptatifs dans le cas univarié dans Chagny [68] et est étendue dans un cadre multivarié en Chagny et al [69].

Présentation de deux problèmes étudiés dans cette thèse:

Nous avons présenté dans la Section 1.4.1 une brève introduction sur le cadre classique de l'estimation de densité non paramétrique et dans la Section 1.4.2 une brève introduction sur le cadre classique de l'estimation par régression non paramétrique. Maintenant, comme mentionné au début, cette thèse contient deux parties indépendantes:

- dans la première partie, Chapitre 2, nous étudions le problème de la déconvolution pour une équation de Fokker-Planck libre: construire une estimation statistique de la condition initiale aléatoire à partir des observations de la solution à temps fixe donné $t > 0$.
- dans la deuxième partie, Chapitre 3, le problème de l'estimation non paramétrique pour la régression avec réponses circulaires est examiné. Nous proposons un nouvel estimateur à noyau "warped" ainsi que l'étude de la consistance, des vitesses de convergence et de l'adaptation sous une considération du risque quadratique ponctuel.

1.5 Partie I - Déconvolution statistique de l'équation Fokker-Planck libre

Cette section présente une brève introduction pour le cadre que nous considérons dans le Chapitre 2. Dans ce travail, nous nous intéressons à l'EDP de Fokker-Planck qui est un cas particulier de l'EDP de McKean-Vlasov (ou dite l'équation des milieux granulaires) équipée d'une interaction logarithmique de Coulomb. Cette équation modélise le mouvement des particules à répulsion électrostatique et une interprétation probabiliste que nous adopterons a été considérée, par exemple dans [21],[9].

1.5.1 Présentation de l'équation de Fokker-Planck

En général, l'équation des milieux granulaires (voir e.g. [18]) est une classe d'EDP contenant de nombreuses équations classiques relevant de la physique, c'est-à-dire

$$\frac{\partial \mu_t}{\partial t} = \nabla \cdot \left[\mu_t \nabla \Big(\mathcal{U}'(\mu_t) + \mathcal{V} + \mathcal{W} * \mu_t \Big) \right],$$

où $(\mu_t)_{t\geq 0}$ est une famille de mesures de probabilité sur $\mathbb{R}^d$. Dans cette équation, $*$ désigne la convolution classique, ∇ est l'opérateur de gradient et $\nabla\cdot$ est l'opérateur de divergence. Cette équation modélise l'évolution en fonction du temps d'une densité de particules soumises à trois effets:

- une densité d'énergie interne $\mathcal{U} : R^+ \to \mathbb{R}$;
- un potentiel externe de confinement $\mathcal{V} : \mathbb{R}^d \to \mathbb{R}$;
- un potentiel d'interaction $\mathcal{W} : \mathbb{R}^d \to \mathbb{R}$, avec quelques hypothèses sur la régularité (voir [18]).

L'équation de Fokker-Planck libre (sera présenté plus en détail dans la Section 2.1) est un cas particulier d'équations de milieux granulaires avec $d = 1$, $\mathcal{U}(s) = 0$, $\mathcal{V}(x) = \frac{1}{2}V(x)$ et $\mathcal{W}(x) = -\ln|x|$, où $V : \mathbb{R} \to \mathbb{R}$ est un potentiel externe donné. Plus précisément, on peut écrire:

$$\frac{\partial \mu_t}{\partial t} = \frac{\partial}{\partial x}\left[\mu_t\left(\frac{1}{2}V' - H\mu_t\right)\right], \tag{1.30}$$

où $(\mu_t)_{t\geq 0}$ est une famille de mesures de probabilité sur $\mathbb{R}$, avec condition initiale μ_0 et $H\mu$ désigne la transformée de Hilbert d'une mesure de probabilité μ sur $\mathbb{R}$, défini pour tout $x \in \mathbb{R}$ par:

$$H\mu(x) := \lim_{\varepsilon \to 0} \int_{\mathbb{R}\setminus[x-\varepsilon, x+\varepsilon]} \frac{1}{x-y} d\mu(y).$$

Dans notre cadre, nous considérons l'équation de Fokker-Planck libre comme suit, sans fonction potentielle V:

$$\frac{\partial \mu_t}{\partial t} = -\frac{\partial}{\partial x}\mu_t(H\mu_t). \tag{1.31}$$

Une solution de l'équation (1.31) est telle que pour toute fonction de test $g \in C^1(\mathbb{R})$:

$$\int_{\mathbb{R}} g(x) d\mu_t(x) = \int_{\mathbb{R}} g(x) d\mu_0(x) + \int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{g'(x)}{x-y} d\mu_s(x) d\mu_s(y)\right) ds, \quad t > 0.$$

Théorème 1.5.1. *Pour tout $t > 0$, l'équation (1.31) a une solution unique.*

L'existence et l'unicité de la solution de l'équation de Fokker-Planck (1.31) sont prouvées plus tard par des techniques probabilistes dans la Section 2.3 (voir aussi [1, Section 4.3.2]). De plus, contrairement aux exemples considérés dans [48, 49], cette EDP est non-linéaire de type McKean-Vlasov avec des interactions logarithmiques. A notre connaissance, il s'agit du premier travail consacré à la déconvolution d'une EDP non-linéaire pour retrouver la condition initiale.

1.5.2 Interprétation probabiliste de la solution avec mouvement brownien libre

Dans le cadre classique, on peut montrer via l'application de la formule d'Ito que la loi de la résolution d'une équation différentielle stochastique (EDS) bien posée vérifie une équation aux dérivées partielles. Par analogie, Biane et Speicher [9] décrivent la solution de l'équation de Fokker-Planck libre (1.30) comme la famille de lois d'un processus satisfaisant une EDS libre, c'est-à-dire un nouveau type de EDS, adapté au cadre non-commutatif (libre), comme le montre le théorème suivant.

Théorème 1.5.2 (voir Théorème 3.1 et Section 4.1 dans [9])**.** *Pour un potentiel $V \in \mathcal{C}^3(\mathbb{R}, \mathbb{R})$, supposons qu'il existe $a, b \in \mathbb{R}$ tel que pour tous $x \in \mathbb{R}$,*

$$-x.V'(x) \leq a.x^2 + b.$$

Ensuite, pour tout $\mathbf{x}_0$ dont la distribution est à support compact, l'EDS libre

$$d\mathbf{x}_s = d\mathbf{h}_s - \frac{1}{2}V'(\mathbf{x}_s)ds, \tag{1.32}$$

où $(\mathbf{h}_s)_{s\geq 0}$ est un mouvement brownien libre, admet une unique solution $(\mathbf{x}_s)_{s\geq 0}$ issue de $\mathbf{x}_0$. De plus, la famille des distributions de $(\mathbf{x}_s)_{s\geq 0}$ satisfait l'équation de Fokker-Planck libre (1.31).

En particulier, en cas de potentiel externe nul, c'est-à-dire $V = 0$, la solution du EDS libre (1.32) au temps $t > 0$ a une forme spéciale: $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{h}_t$. Puisque $\mathbf{x}_0$ et $\mathbf{h}_t$ sont libres, on obtient la distribution de $\mathbf{x}_0 + \mathbf{h}_t$ (somme de deux variables aléatoires non-commutatives), comme dans la proposition suivante.

Proposition 1.5.3. *(voir aussi [1, Section 5.3.3]) Si $\mathbf{x}_0$ admet la mesure spectrale μ_0, alors $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{h}_t$ admet*

$$\mu_t = \mu_0 \boxplus \sigma_t, \tag{1.33}$$

comme mesure spectrale, où l'opération $\boxplus$ est la convolution libre et σ_t est la loi semi-circulaire.

L'opération de convolution libre $\boxplus$ a été introduit par Voiculescu dans [63], et une brève définition de cette opération est présentée plus loin dans la Section 2.2.1. De plus, nous avons par conséquent la proposition suivante sur l'existence de la fonction de densité de μ_t.

Proposition 1.5.4. *(voir [7, Corollaire 3]) Supposons que la condition initiale μ_0 possède une densité $p_0(\cdot) := p(0, \cdot)$ en ce qui concerne la mesure Lebesgue sur $\mathbb{R}$, alors pour tout $t > 0$, la solution μ_t de l'équation (1.31) a une fonction de densité $p(t, .)$ telle que*

$$\partial_t p(t,x) = -\partial_x \big[p(t,x).\big(Hp(t,x)\big)\big], \quad x \in \mathbb{R}.$$

Ainsi, nous avons vu que la solution de cette équation de Fokker-Planck libre (1.31) peut être représentée comme une convolution libre (ceci sera étudié en détail plus loin dans la Section 2.2) entre la condition initiale et la distribution semi-circulaire. En conséquence, une idée est alors d'appliquer des techniques de déconvolution libre (développées récemment par Arizmendi et al. [2]) mais avec des considérations statistiques.

1.5.3 Procédure de déconvolution

Premièrement, la transformée de Cauchy d'une mesure μ est définie comme

$$G_\mu(z) := \int_{\mathbb{R}} \frac{d\mu(x)}{z - x}, \quad \text{pour } z \in \mathbb{C}^+,$$

où $\mathbb{C}^+$ est l'ensemble des nombres complexes de partie imaginaire positive, et comme G_μ ne s'annule pas sur $\mathbb{C}^+$ on définit $F_\mu(z) := 1/G_\mu(z)$, pour $z \in \mathbb{C}^+$. Pour tout $z \in \mathbb{C}$, soit $\mathrm{Im}(z)$ la partie imaginaire de z, et pour $\gamma > 0$ on définit un sous-domaine $\mathbb{C}_\gamma := \{z \in \mathbb{C}^+ : \mathrm{Im}(z) > \gamma\}$. Maintenant, en appliquant les méthodes de subordination proposées dans [2], qui travaillent dans un cadre général, à notre cas, on a prouvé dans la Section 2.4 le résultat suivant sur les fonctions de subordination:

Théorème 1.5.5. *Il existe des fonctions de subordination uniques w_1 et w_{fp} de $\mathbb{C}_{2\sqrt{t}}$ sur $\mathbb{C}^+$ telles que:*

(i) pour $z \in \mathbb{C}_{2\sqrt{t}}$, $Im\big(w_1(z)\big) \geq \frac{1}{2} Im(z)$ et $Im\big(w_{fp}(z)\big) \geq \frac{1}{2} Im(z)$, et aussi

$$\lim_{y \to +\infty} \frac{w_1(i.y)}{i.y} = \lim_{y \to +\infty} \frac{w_{fp}(i.y)}{i.y} = 1.$$

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

$$F_{\mu_0}(z) = F_{\sigma_t}(w_1(z)) = F_{\mu_t}(w_{fp}(z)), \quad ou \quad G_{\mu_0}(z) = G_{\sigma_t}(w_1(z)) = G_{\mu_t}(w_{fp}(z)); \tag{1.34}$$

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

$$w_{fp}(z) = z + w_1(z) - F_{\mu_0}(z); \tag{1.35}$$

(iv) En notant $h_{\sigma_t}(w) := w - F_{\sigma_t}(w) = t.G_{\sigma_t}(w)$ *(voir (2.7)),* $\widetilde{h}_{\mu_t}(w) := w + F_{\mu_t}(w)$ *sur* $\mathbb{C}^+$*, et on définit une fonction* $\mathcal{L}_z$ *comme*

$$\begin{aligned} \mathcal{L}_z(w) :&= h_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z \\ &= t.G_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z; \end{aligned} \tag{1.36}$$

Pour tout $z \in \mathbb{C}_{2\sqrt{t}}$*, on a*

$$\mathcal{L}_z\big(w_{fp}(z)\big) = w_{fp}(z). \tag{1.37}$$

Alors, pour tout w *tel que* $Im(w) > \frac{1}{2}Im(z)$*, la fonction itérée* $\mathcal{L}_z^{\circ m}(w)$ *converge vers* $w_{fp}(z) \in \mathbb{C}^+$ *lorsque* $m \to +\infty$*.*

De plus, dans notre cadre, la mesure semi-circulaire σ_t est impliquée dans la convolution libre (voir (1.33)) et on a une formule explicite pour la transformée de Cauchy de σ_t comme $G_{\sigma_t}(z) = \dfrac{z - \sqrt{z^2 - 4t}}{2t}$, pour $z \in \mathbb{C}^+$. On peut donc en déduire que la fonction de subordination $w_{fp}(z)$ à l'instant t est lié à G_{μ_t} par l'équation fonctionnelle

$$w_{fp}(z) = z + t.G_{\mu_t}(w_{fp}(z)), \quad z \in \mathbb{C}_{2\sqrt{t}}.$$

De cela, on peut récupérer G_{μ_0} avec la formule $G_{\mu_0}(z) = G_{\mu_t}(w_{fp}(z))$ pour $z \in \mathbb{C}_{2\sqrt{t}}$, et ainsi, on peut récupérer p_0 (en utilisant la formule d'inversion de Stieltjes, voir Lemme 2.4.3 et (2.55) pour plus de détail). Plus précisément, on a prouvé dans la Section 2.5.1 que pour tout $\gamma > 2\sqrt{t}$, $f_{\mu_0 * \mathcal{C}_\gamma}$ qui est la densité de la convolution classique de μ_0 avec $\mathcal{C}_\gamma$ satisfait

$$f_{\mu_0 * \mathcal{C}_\gamma}(x) = \frac{1}{\pi t}\left[\gamma - \mathrm{Im} w_{fp}(x + i\gamma)\right], \quad x \in \mathbb{R}, \tag{1.38}$$

où $\mathcal{C}_\gamma$ désigne la distribution de Cauchy du paramètre γ, défini par sa densité $f_\gamma(x) := \dfrac{\gamma}{\pi(x^2 + \gamma^2)}$, $x \in \mathbb{R}$. Alors, d'après (1.38), on voit que pour estimer p_0, la densité de μ_0, il faut une estimation de la fonction de subordination w_{fp} combinée à une étape de déconvolution classique à partir d'une distribution de Cauchy.

1.5.4 Mouvement brownien de Dyson et convergence vers la solution de l'équation de Fokker-Planck

Observations En plus du problème de déconvolution libre, notre observation ne consiste pas en la variable aléatoire à valeur opérateur $\mathbf{x}_t$ mais dans sa contrepartie matricielle. Plus précisément, on observe une matrice $X_n(t)$ pour un $t > 0$ donné, supposée fixe dans la suite, où

$$X_n(t) = X_n(0) + H_n(t), \qquad t \geq 0$$

avec $X_n(0)$ une matrice diagonale dont les entrées sont la statistique ordonnée $\lambda_1^n(0) < \cdots < \lambda_n^n(0)$ d'un vecteur $(d_j^n)_{j\in\{1,\ldots n\}}$ de n indépendant et identiquement distribué (i.i.d.) variables aléatoires distribuées comme $\mu_0(dx) = p_0(x)\,dx$, absolument continue par rapport à la mesure Lebesgue sur $\mathbb{R}$, et $H_n(t)$ un mouvement Brownien Hermitien standard (voir la Définition 2.3.1 pour la définition de $H_n(t)$). Comme la distribution de $H_n(t)$ est invariante par conjugaison, choisir $X_n(0)$ pour être une matrice diagonale n'est pas restrictif.

L'objectif est d'estimer la densité p_0. L'observation consiste en la matrice $X_n(t)$ à le temp fixée t, à partir de laquelle on peut calculer les valeurs propres $(\lambda_1^n(t), \cdots, \lambda_n^n(t))$ puis la mesure empirique associée. On n'observe pas directement la condition initiale $X_n(0)$.

Expliquons maintenant le lien avec l'équation de Fokker-Planck au niveau du système de particules des valeurs propres. On sait que les valeurs propres $(\lambda_1^n(t), \cdots, \lambda_n^n(t))$ de $X_n(t)$ résolvent le système d'équations différentielles stochastiques (SDE) suivant:

$$\mathrm{d}\lambda_k^n(t) = \frac{1}{\sqrt{n}}\mathrm{d}\beta_k(t) + \frac{1}{n}\sum_{j\neq k}\frac{\mathrm{d}t}{\lambda_k^n(t) - \lambda_j^n(t)}, \quad 1 \leq k \leq n,$$

où β_k sont i.i.d. mouvements browniens réels standard. Notez que ce système de particules n'est introduit qu'à des fins d'interprétation et ne sera pas utilisé directement dans la suite. Maintenant, si on note par

$$\mu_t^n := \frac{1}{n}\sum_{j=1}^{n}\delta_{\lambda_j^n(t)}, \tag{1.39}$$

la mesure empirique de ces valeurs propres au temps t, alors le processus $(\mu_t^n)_{t\geq 0}$ converge faiblement presque sûrement lorsque $n \to +\infty$ au processus $(\mu_t)_{t\geq 0}$ avec densité $(p(t,\cdot))_{t\geq 0}$ solution de l'équation de Fokker-Planck libre, comme indiqué dans la Proposition 2.3.5 de la Section 2.3.

1.5.5 Contributions

Cette section présente les principaux résultats pour le problème de déconvolution statistique de l'équation de Fokker-Planck libre à temps fixe donné, étudié au Chapitre 2.
On n'observe pas directement la mesure μ_t. L'observation est la matrice $X_n(t)$ au temps $t > 0$ pour un n donné et donc sa mesure spectrale empirique telle que définie dans (1.39). Alors, pour $z \in \mathbb{C}^+$, on remplace la procédure $G_{\mu_t}(z)$ par son estimateur naturel:

$$\widehat{G}_{\mu_t^n}(z) = \int_{\mathbb{R}}\frac{d\mu_t^n(\lambda)}{z-\lambda} = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{z-\lambda_j^n(t)}.$$

Tout d'abord, nous proposons ci-dessous un estimateur statistique $\widehat{w}_{fp}^n(z)$ pour la fonction de subordination w_{fp} présenté dans la section 1.5.3.

Théorème 1.5.6. *Il existe une solution unique à l'équation fonctionnelle suivante dans $w(z)$, pour tout $z \in \mathbb{C}_{2\sqrt{t}}$:*

$$\frac{1}{t}(w(z) - z) = \widehat{G}_{\mu_t^n}(w(z)). \tag{1.40}$$

Ce point fixe est noté $\widehat{w}_{fp}^n$. De plus, $|\widehat{w}_{fp}^n(z) - z| \leq \sqrt{t}$, pour tout $z \in \mathbb{C}_{2\sqrt{t}}$.

Puisque la transformation de Cauchy G_{μ_t} n'est pas inversible sur tout le domaine $\mathbb{C}^+$, la fonction de subordination $w_{fp}(z)$ ne sera défini que pour $z \in \mathbb{C}_{2\sqrt{t}}$. Comme expliqué précédemment dans la Section 1.2.3, nous estimons p_0 en combinant une étape de déconvolution libre via l'utilisation de $\widehat{w}_{fp}^n$ avec ensuite une étape de déconvolution classiquement statistique. Considérons une fenêtre $h > 0$ et un noyau K. Rappelons que $\mathrm{Im}(z)$ désigne la partie imaginaire de $z \in \mathbb{C}$, selon (1.38), avec $\gamma > 2\sqrt{t}$, nous définissons notre estimateur final $\widehat{p}_{0,h}$ via sa transformée de Fourier, notée $\widehat{p}_{0,h}^{\star}$, comme suit:

$$\widehat{p}_{0,h}^{\star}(\xi) = e^{\gamma|\xi|}.K_h^{\star}(\xi).\frac{1}{\pi t}\left[\gamma - \mathrm{Im}\,\widehat{w}_{fp}^n(\cdot + i.\gamma)^{\star}(\xi)\right], \quad \xi \in \mathbb{R}, \tag{1.41}$$

où nous définissons $K_h(\cdot) := \frac{1}{h}.K(\frac{\cdot}{h})$. Notez que, comme d'habitude dans les statistiques non paramétriques, la dernière expression dépend de $K_h^\star$, un terme de régularisation défini par la transformée de Fourier d'une fonction noyau K_h en fonction d'un paramètre de fenêtre $h > 0$, voir l'équation (2.64) dans la Définition 2.5.3 pour plus de détails.
De (1.41), on peut observer que $\widehat{w}_{fp}^n$ joue un rôle critique dans l'examen du comportement de $\widehat{p}_{0,h}$. Par conséquent, il est raisonnable d'étudier d'abord la consistance de $\widehat{w}_{fp}^n$, donnée dans le résultat suivant.

Proposition 1.5.7. *Soit $\gamma > 2\sqrt{t}$. Supposons que p_0 satisfasse la condition*

$$\int_{\mathbb{R}} \log(x^2+1)p_0(x)dx < +\infty. \tag{1.42}$$

Alors:
(i) Pour tout $z \in \mathbb{C}_{2\sqrt{t}}$, l'estimateur $\widehat{w}_{fp}^n(z)$ converge presque sûrement vers $w_{fp}(z)$ lorsque $n \to \infty$.
(ii) La convergence est uniforme sur $\mathbb{C}_\gamma$.
(iii) On a le vitesse de convergence suivant sur $\mathbb{C}_\gamma$:

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

Nous étudions les propriétés théoriques de $\widehat{p}_{0,h}$ en dérivant les vitesses asymptotiques de "mean integrated squared error" de $\widehat{p}_{0,h}$ dans la décomposition biais-variance. L'étude du terme de variance est complexe et repose sur des contrôles précis de la différence $\left|\widehat{w}_{fp}^n(z) - w_{fp}(z)\right|$ fourni par la Proposition 1.5.7.

Théorème 1.5.8. *Soit*

$$\Sigma := \left\|\widehat{p}_{0,h}^\star - K_h^\star p_0^\star\right\|^2.$$

On suppose qu'il existe une constante $C > 0$ tel que pour suffisamment grand $\kappa > 0$,

$$\mu_0\big((\kappa, +\infty)\big) \leq \frac{C}{\kappa}. \tag{1.43}$$

Alors, pour tout $\gamma > 2\sqrt{t}$, il existe une constante $C_{var}(t)$ dépendant de t seulement tel que pour tout $h > 0$ et n suffisamment grand,

$$\mathbb{E}(\Sigma) \leq \frac{\gamma^8}{(\gamma^4 - 4t)^4}.\frac{C_{var}(t).e^{\frac{2\gamma}{h}}}{n}. \tag{1.44}$$

Le Théorème 1.5.8 montre que le terme de variance est d'ordre $\frac{1}{n}.\big(e^{\frac{2\gamma}{h}}\big)$ comme souhaité pour la déconvolution avec la distribution de Cauchy avec paramètre γ, alors que le terme de biais est déterminé par les propriétés de régularité de la fonction p_0. En particulier, quand p_0 appartient à une classe de densités "super-smooth" $\mathcal{S}(a, r, L)$ (qui est un cas particulier de classe $\mathcal{S}(\delta, a, r, L)$ défini dans (1.28) avec $\delta = 0$), nous obtenons une borne supérieure pour le biais (voir e.g. [25, Proposition 3]). Une raison de considérer la classe $\mathcal{S}(a, r, L)$ s'inspire de [16] dans laquelle l'optimalité des vitesses convergents a été prouvée jusqu'à présent pour le cas $0 < r < s$ avec $\delta = 0$ (où s est la régularité du bruit "super-smooth", comme mentionné dans la Section 1.4.1). Nous avons en effet montré que:

$$\text{MISE} := \mathbb{E}\Big[\left\|\widehat{p}_{0,h} - p_0\right\|^2_{\mathbb{L}^2(\mathbb{R})}\Big] \leq L.e^{-2ah^{-r}} + \frac{\gamma^8}{(\gamma^2 - 4t)^4}.\frac{C_{\text{var}}(t).e^{\frac{2\gamma}{h}}}{n}. \tag{1.45}$$

Nous obtenons les résultats suivants.

Corollaire 1.5.9. *Supposons que μ_0 satisfasse l'hypothèse (1.43) et la densité p_0 appartient à l'espace $\mathcal{S}(a,r,L)$ pour $a>0$, $r>0$ et $L>0$. Alors, pour tout $\gamma > 2\sqrt{t}$ et en choisissant la fenêtre h minimisant la borne supérieure de (1.45) on a:*

$$\mathbb{E}\Big[\|\widehat{p}_{0,h}-p_0\|^2\Big] = \begin{cases} O\big(n^{-\frac{a}{a+\gamma}}\big) & si\ r=1 \\ O\Big(\exp\Big\{-\frac{2a}{(2\gamma)^r}\Big[\log n + (r-1)\log\log n + \sum_{i=0}^{k} b_i^*(\log n)^{r+i(r-1)}\Big]^r\Big\}\Big) & si\ r<1 \\ O\Big(\frac{1}{n}\exp\Big\{\frac{2\gamma}{(2a)^{1/r}}\Big[\log n + \frac{r-1}{r}\log\log n + \sum_{i=0}^{k} d_i^*(\log n)^{\frac{1}{r}-i\frac{r-1}{r}}\Big]^{1/r}\Big\}\Big) & si\ r>1, \end{cases} \tag{1.46}$$

où l'entier k est tel que

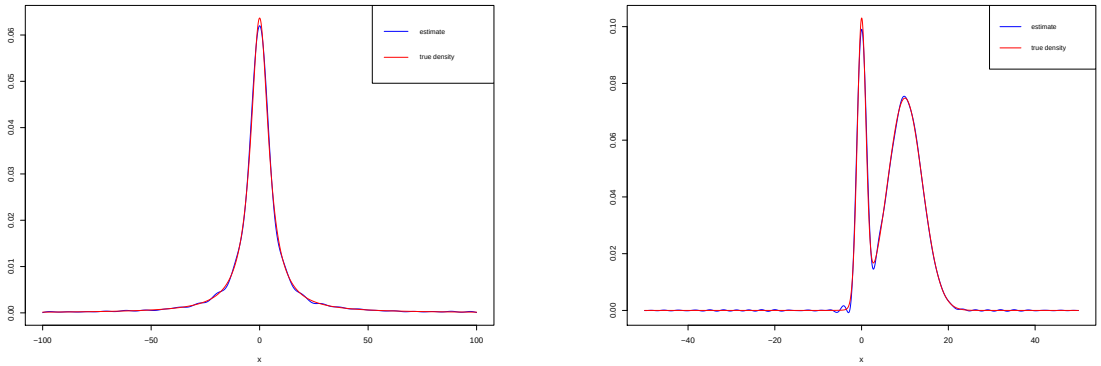
$$\frac{k}{k+1} < \min\big(r,\frac{1}{r}\big) \le \frac{k+1}{k+2},$$

et où les constantes b_i^* et d_i^* résolvent le système triangulaire suivant:

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

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

Le cas des régularités de type "Sobolev-ordinary" conduit à des vitesses de convergence logarithmiques (voir Corollaire 2.7.5 pour plus de détail). Pour plus de détails sur les vitesses de convergence, voir la Section 2.7.1.
Maintenant, la Figure 1.5 ci-dessous illustre deux résultats remarquables d'expérience numérique avec le nombre d'observations $n = 4000$ au temps $t = 1$ pour le cas où $\mu_0 \sim$ Cauchy(0,5) dans la Figure 1.5a et pour le cas $\mu_0 \sim Y \times \mathcal{N}(0,1) + (1-Y)\times\mathcal{N}(10,4)$, avec variable aléatoire $Y \sim$ Bernoulli(0.25) dans la Figure 1.5b. Dans ces simulations numériques, la fenêtre optimale h est choisie pratiquement en utilisant la Validation-Croisée. En conséquence, nous pouvons voir que nos estimateurs fonctionnent assez bien, surtout lorsque p_0 appartient à une classe de densités "super-smooth".



(a) $\mu_0 \sim$ Cauchy(0,5) (b) $\mu_0 \sim Y\times\mathcal{N}(0,1)+(1-Y)\times\mathcal{N}(10,4)$ avec $Y \sim$ Bernoulli(0.25)

Figure 1.5: Estimation de p_0

1.6 Partie II - Estimation non-paramétrique pour la régression avec réponses circulaires

Cette section présente une brève introduction pour le cadre que nous étudions dans le Chapitre 3.

1.6.1 Données circulaires

La statistique directionnelle est la branche de la statistique qui traite des observations qui sont des directions. Dans ce travail, nous considérerons plus spécifiquement les *données circulaires* qui surgissent chaque fois que l'on utilise une échelle périodique pour mesurer des observations. Ces données sont représentées par des points situés sur le cercle unité de $\mathbb{R}^2$ noté dans la suite $\mathbb{S}^1$. Notez que le terme *données circulaires* est également utilisé pour les distinguer des données prises en charge sur la ligne réelle $\mathbb{R}$ (ou un sous-ensemble de celui-ci), que l'on appelle désormais *données linéaires*. La principale difficulté lorsqu'on traite de telles données est la courbure de l'espace échantillon puisque le cercle unité est une variété non-linéaire. Les méthodes statistiques d'analyse des données circulaires ont considérablement évolué et il existe désormais des contreparties circulaires pour plusieurs techniques d'analyse de données. Pour des enquêtes complètes sur les méthodes statistiques pour les données circulaires, voir Lee [77], Ley et Verdebout [78] et les avancées récentes sont collectées dans Pewsey et García-Portugués [83], et les références qui s'y trouvent.

Avant d'entrer dans les considérations statistiques, nous introduisons des notations importantes. Dans la suite, un point sur $\mathbb{S}^1$ ne sera pas représenté comme un vecteur à deux dimensions $\mathbf{w} = (w_2, w_1)^\top$ avec l'unité de norme Euclidienne mais comme un angle $\theta = \text{atan2}(w_1, w_2)$ défini comme:

Définition 1.6.1. (voir [78, Section 1.3]) La function $\text{atan2} : \mathbb{R}^2\backslash(0,0) \mapsto [-\pi, \pi)$ est définie pour tout $(w_1, w_2) \in \mathbb{R}^2\backslash(0,0)$ par

$$\text{atan2}(w_1, w_2) := \begin{cases} \arctan\left(\frac{w_1}{w_2}\right) & \text{si} \quad w_2 \geq 0, w_1 \neq 0 \\ 0 & \text{si} \quad w_2 > 0, w_1 = 0 \\ \arctan\left(\frac{w_1}{w_2}\right) + \pi & \text{si} \quad w_2 < 0, w_1 > 0 \\ \arctan\left(\frac{w_1}{w_2}\right) - \pi & \text{si} \quad w_2 < 0, w_1 \leq 0, \end{cases}$$

avec arctan prenant des valeurs dans $[-\pi/2, \pi/2]$. En particulier, $\arctan(+\infty) = \pi/2$ et $\arctan(-\infty) = -\pi/2$.

Dans cette définition, on a arbitrairement fixé l'origine de $\mathbb{S}^1$ à $(1,0)^\top$ et on utilise le sens anti-horaire comme positif. Ainsi, une variable aléatoire circulaire peut être représentée comme un angle sur $[-\pi, \pi)$. Remarquez que $\text{atan2}(0,0)$ est indéfini.

1.6.2 Modèle statistique

Dans ce travail, nous nous concentrons sur un modèle de régression non paramétrique avec réponse circulaire et prédicteur linéaire. Supposons que nous avons un i.i.d. échantillon $\{(X_j, \Theta_j)\}_{j=1}^n$ distribué comme (X, Θ), où Θ est une variable aléatoire circulaire, et X est une variable aléatoire de densité f_X à supporté sur $\mathbb{R}$. On cherche une fonction m qui contient la structure de dépendance entre les prédicteurs X_j et les observations Θ_j. Pour $x \in \mathbb{R}$, soit $m_1(x) = \mathbb{E}\big(\sin(\Theta)|X = x\big)$ et $m_2(x) = \mathbb{E}\big(\cos(\Theta)|X = x\big)$. Ce contexte est considéré dans [80] pour la première fois. Nous étudions l'estimation non paramétrique de la fonction de régression m en un point donné de $\mathbb{R}$.
En raison de la différence entre les données circulaires et linéaires par leur nature périodique,

une tâche préliminaire consiste à définir un risque approprié mesurant la différence entre les deux angles Θ et $m(X)$. Le contexte circulaire conduit à une approche naturelle de l'utilisation des fonctions trigonométriques. Pour deux angles $\theta_1, \theta_2 \in [-\pi, \pi)$, on utilise la distance angulaire $d_c(\theta_1, \theta_2) := \big(1 - \cos(\theta_1 - \theta_2)\big)$ pour mesurer la différence. Notez que cette distance angulaire correspond à la norme Euclidienne habituelle dans $\mathbb{R}^2$. En effet, les angles θ_1 et θ_2 déterminent les points correspondants $(\cos\theta_1, \sin\theta_1)$ et $(\cos\theta_2, \sin\theta_2)$ respectivement sur le cercle unité $\mathbb{S}^1$. Ensuite, la norme euclidienne au carré habituelle dans $\mathbb{R}^2$ implique que $\big(\cos\theta_1 - \cos\theta_2\big)^2 + \big(\sin\theta_1 - \sin\theta_2\big)^2 = 2.\big[1 - \cos(\theta_1 - \theta_2)\big] = 2.d_c(\theta_1, \theta_2)$.
Par conséquent, en utilisant la distance angulaire, nous cherchons une fonction m comme minimiseur du risque suivant (voir aussi Section 3.1):

$$L(m(X)) := \mathbb{E}\big[1 - \cos(\Theta - m(X))\big|X\big].$$

En fait, étant donné $x \in \mathbb{R}$, le risque $L(m(x))$ est minimisé par $m(x) = \text{atan2}(m_1(x), m_2(x))$, voir la Section 3.1 pour plus de détails sur le calcul. En conclusion, la nature circulaire de la réponse est prise en compte par l'arctangente du rapport des composantes sinus et cosinus du moment trigonométrique conditionnel de Θ étant donné X.
Rappelons que le modèle est considéré pour la première fois par Di Marzio et al. [80]. Dans [80], considérant que les covariables supportées sur $[0, 1]$, les auteurs proposent un estimateur Nadaraya-Watson et un estimateur polynomial linéaire local pour estimer $\text{atan2}(m_1(x), m_2(x))$. Sous des hypothèses sur la régularité de m_1 et m_2, en utilisant l'expansion de Taylor pour la fonction arctangente, ils dérivent des expressions pour le biais asymptotique et la variance, cependant, les termes restants dans les expansions ne sont pas précisément contrôlés. Quant aux travaux plus récents de Meilan Vila et al. [82], ils ont étudié le cadre multivarié $[0, 1]^d$ avec les mêmes estimateurs et techniques de preuves. Dans les deux articles, les vitesses de convergence et d'adaptation restent des questions ouvertes, de plus, la fenêtre du noyau est sélectionnée par validation croisée dans la pratique. Par adaptation, on rappelle que les estimateurs ne nécessitent pas la spécification de la régularité de la fonction de régression ce qui est crucial d'un point de vue pratique. Ils considèrent la même régularité pour m_1 et m_2 qui peut être considérée comme restreinte, alors qu'il serait plus naturel d'avoir des régularités différentes pour m_1 et m_2. De plus, leurs estimateurs ne dépendent que d'une seule fenêtre alors qu'il serait plus naturel de considérer deux fenêtres à cause des régularités différentes possibles sur m_1 et m_2.
Dans cette perspective, notre motivation est de remplir le vide dans la littérature. Dans ce travail, considérons les prédicteurs pris en charge sur $\mathbb{R}$, nous proposons d'abord de nouveaux estimateurs à noyau "warped" $\widehat{m}_{\widehat{h}_1}(x)$ et $\widehat{m}_{\widehat{h}_2}(x)$ pour $m_1(x)$ et $m_2(x)$ en un point $x \in \mathbb{R}$ respectivement, où les fenêtres $\widehat{h}_1$ et $\widehat{h}_2$ sont sélectionnés de manière adaptative par la règle de Goldenshluger Lepski. Ensuite, on pose un estimateur $\widehat{m}_{\widehat{h}}(x) := \text{atan2}\big(\widehat{m}_{\widehat{h}_1}(x), \widehat{m}_{\widehat{h}_2}(x)\big)$ pour la fonction de régression $m(x)$ avec $\widehat{h} := (\widehat{h}_1, \widehat{h}_2)$. On obtient une borne supérieure pour l'estimateur $\widehat{m}_{\widehat{h}}(x)$ pour le risque au carré ponctuel et sur les classes Hölder.
Une description plus détaillée du cadre que nous examinons est donnée dans la Section 3.1 du Chapitre 3. Le résumé des principaux résultats théoriques et numériques sont présentés dans la Section 1.6.3.

1.6.3 Contributions

Cette section présente les principaux résultats pour le problème d'estimation non paramétrique pour la régression avec des réponses circulaires, étudié au Chapitre 3.
Soit $K : \mathbb{R} \to \mathbb{R}$ une fonction intégrable telle que K soit à support compact, $K \in \mathbb{L}^\infty(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. On dit que K est un noyau s'il satisfait $\int_{\mathbb{R}} K(y)dy = 1$. Supposons que F_X

soit inversible, nous introduisons des fonctions auxiliaires $g_1, g_2 : (0,1) \mapsto \mathbb{R}$ défini par

$$g_1 := m_1 \circ F_X^{<-1>}, \quad \text{et} \quad g_2 := m_2 \circ F_X^{<-1>}$$

de telle sorte que $m_1 = g_1 \circ F_X$ et $m_2 = g_2 \circ F_X$. Supposons que la distribution de X soit connue, pour $u \in F_X(\mathbb{R})$, on définit des estimateurs:

$$\widehat{g}_{1,h_1}(u) := \frac{1}{n}\sum_{j=1}^{n} \sin(\Theta_j).K_{h_1}(u - F_X(X_j)), \quad \text{et} \quad \widehat{g}_{2,h_2}(u) := \frac{1}{n}\sum_{j=1}^{n} \cos(\Theta_j).K_{h_2}(u - F_X(X_j)), \tag{1.47}$$

pour estimer g_1 et g_2 respectivement, et $h_1, h_2 > 0$ sont des fenêtres du noyau $K_{h_1}(\cdot)$ et $K_{h_2}(\cdot)$ respectivement. De plus, par conséquent, pour $x \in \mathbb{R}$, les estimateurs pour m_1 et m_2 sont

$$\widehat{m}_{1,h_1}(x) = \big(\widehat{g}_{1,h_1} \circ F_X\big)(x), \quad \text{et} \quad \widehat{m}_{2,h_2}(x) = \big(\widehat{g}_{2,h_2} \circ F_X\big)(x). \tag{1.48}$$

En notant $h := (h_1, h_2)$, on définit alors l'estimateur pour $m(x)$ à $x \in \mathbb{R}$ par

$$\widehat{m}_h(x) := \text{atan2}\big(\widehat{m}_{1,h_1}(x), \widehat{m}_{2,h_2}(x)\big).$$

Maintenant, pour la méthode pour obtenir une sélection automatique de fenêtre, nous proposons une règle de sélection adaptive inspirée de Goldenshluger et Lepski [71]. Cette procédure de sélection est décrite dans la Section 3.2.2 du Chapitre 3.
Soit $\widehat{g}_{1,\widehat{h}_1}$ et $\widehat{g}_{2,\widehat{h}_2}$ être les estimateurs adaptatifs de g_1 et g_2 respectivement où $\widehat{h}_1$ et $\widehat{h}_2$ sont des fenêtres sélectionnées par notre règle "data-driven" de sélection (voir Section 3.2.2). On établit une inégalité de type oracle pour $\widehat{g}_{1,\widehat{h}_1}$ et $\widehat{g}_{2,\widehat{h}_2}$ (voir Proposition 3.2.2 dans la Section 3.2.3) qui illustre la décomposition biais-variance du risque au carré ponctuel.
Ensuite, en notant $\widehat{h} := (\widehat{h}_1, \widehat{h}_2)$, on définit l'estimateur "adaptive" de m à $x \in \mathbb{R}$ par

$$\widehat{m}_{\widehat{h}}(x) := \text{atan2}\big(\widehat{m}_{1,\widehat{h}_1}(x), \widehat{m}_{2,\widehat{h}_2}(x)\big). \tag{1.49}$$

Nous considérons également l'hypothèse suivante sur le noyau K:

Hypothèse 1.6.2. Le noyau K est d'ordre $\mathcal{L} \in \mathbb{R}_+$, c'est-à-dire

(i) $\int_{\mathbb{R}}(1 + |y|^{\mathcal{L}}).|K(y)|dy < \infty$.

(ii) $\forall k \in \{1, ..., \lfloor \mathcal{L} \rfloor\}$, $\int_{\mathbb{R}} y^k.K(y)dy = 0$.

Puisque la fonction $\text{atan2}(w_1, w_2)$ est indéfinie lorsque $w_1 = w_2 = 0$, il est raisonnable de considérer l'hypothèse suivante:

Hypothèse 1.6.3. Pour $x \in \mathbb{R}$ donné, supposons que

$$|m_1(x)| = \big|g_1(F_X(x))\big| \geq \delta_1 > 0, \quad \text{and} \quad |m_2(x)| = \big|g_2(F_X(x))\big| \geq \delta_2 > 0.$$

Soit $\mathcal{H}(\beta, L)$ dénotons la classe Hölder avec $\beta, L > 0$ comme dans la Définition 3.2.4. De plus, soient $c_{0,1}$ et $c_{0,2}$ des constantes de réglage constituées dans la règle de sélection des fenêtres adaptatives $\widehat{h}_1$ et $\widehat{h}_2$ respectivement (voir Section 3.2.2). Le théorème suivant nous donne une borne supérieure pour l'erreur quadratique moyenne des estimateurs $\widehat{m}_{\widehat{h}}$ défini dans (1.49).

Théorème 1.6.4. *Soit $\beta_1, \beta_2, L_1, L_2 > 0$. Supposer que g_1 appartienne à une classe Hölder $\mathcal{H}(\beta_1, L_1)$, g_2 appartienne à $\mathcal{H}(\beta_2, L_2)$, le noyau K satisfait l'hypothèse 1.6.2 avec un indice $\mathcal{L} \in (\mathbb{R}_+)$ tel que $\mathcal{L} \geq \max(\beta_1, \beta_2)$. Soit $q \geq 1$, et supposons que $\min\{c_{0,1}; c_{0,2}\} \geq 16\big(2+q\big)^2.\big(1+\|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$. Alors, sous l'hypothèse 1.6.3, pour n suffisamment grand,*

$$\mathbb{E}\Big[\big|\widehat{m}_{\widehat{h}}(x) - m(x)\big|^2\Big] \leq 32.\pi^2.\Big(\frac{1}{\delta_2^2} + \frac{1}{\delta_1^2}\Big).(C_1^2 + C_2^2).\max\big\{\psi_n(\beta_1)^2, \psi_n(\beta_2)^2\big\},$$

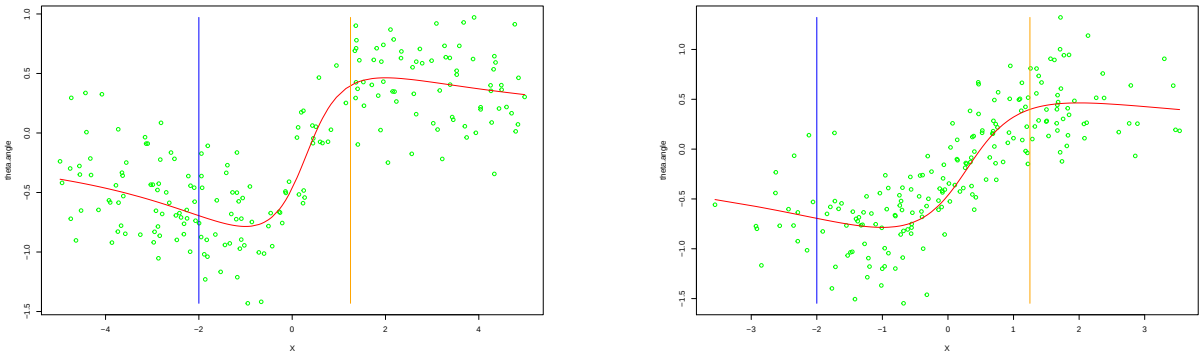
où C_1 est une constante (dépendant de $\beta_1, L_1, c_{0,1}, K$) et C_2 est une constante (dépendant de $\beta_2, L_2, c_{0,2}, K$), avec $\psi_n(\beta_1) = \big(\log(n)/n\big)^{\frac{\beta_1}{2\beta_1+1}}$ et $\psi_n(\beta_2) = \big(\log(n)/n\big)^{\frac{\beta_2}{2\beta_2+1}}$.

Observons que si $\beta_1 = \beta_2 = \beta$, alors on obtient la vitesse $\psi_n(\beta) = (\log n/n)^{\beta/(2\beta+1)}$, qui est la vitesse optimale pour l'estimation de la fonction de régression univariée adaptative et le risque ponctuel (voir par exemple la section 2 dans [66]). Nous conjecturons que la vitesse obtenu dans le Théorème 1.6.4 est optimale même s'il reste un problème ouvert.

Maintenant, nous décrivons brièvement quelques simulations numériques pour illustrer les performances de notre estimateur $\widehat{m}_{\widehat{h}}$. Considérez le modèle suivant

$$\text{M1.} \quad \Theta = \text{atan2}\big(2X - 1, X^2 + 2\big) + \zeta \pmod{2\pi}; \tag{1.50}$$

où les erreurs circulaires ζ_j sont échantillonnées à partir d'une distribution de von Mises $vM(0, 10)$, avec un paramètre de localisation $0 \in [-\pi, \pi)$ et concentration de 10, (voir Section 3.3 pour plus de détails sur la définition de la distribution de von Mises). L'implémentation est faite pour le cas d'une distribution uniforme $X \sim U([-5, 5])$ et aussi le cas de la distribution gaussienne $X \sim \mathcal{N}(0, 1.5)$. Dans les deux cas, pour la taille d'échantillon $n = 200$, nous estimons $m(x)$ à $x = -2$ et à un autre point $x = 1.25$. Une illustration des données simulées est présentée dans Figure 1.6. Les estimateurs finaux sont calculés en utilisant la contrepartie empirique $\widehat{F}_n(y) := \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{X_j \leq y}$, pour $y \in \mathbb{R}$, de la fonction "warping" F_X et le noyau Epanechnikov $K(v) = \frac{3}{4}.(1 - v^2).\mathbb{1}_{|v|\leq 1}$, $v \in \mathbb{R}$. Tout d'abord, dans la procédure de



(a) Model M1 : $X_j \sim U([-5, 5])$ (b) Model M1 : $X_j \sim \mathcal{N}(0, 1.5)$

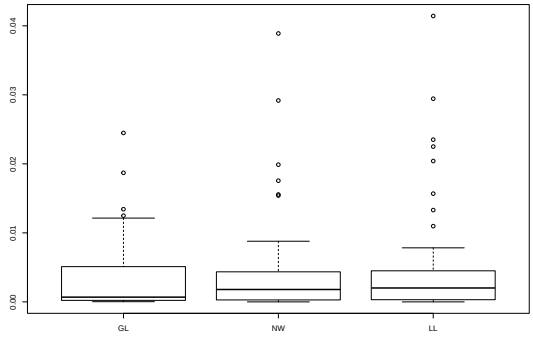
Figure 1.6: Une illustration pour les données simulées $\{\Theta_j\}_{j=1}^n$ (points verts) du model-M1 avec $n = 200$. La courbe rouge présente la vraie fonction de régression m, tandis que la ligne verticale bleue affiche le point $x = -2$ et la ligne verticale orange affiche le point $x = 1.25$ où nous estimons $m(x)$.

sélection de fenêtre, considérons $c_{0,1} = c_{0,2}$, nous effectuons une implémentation préliminaire sur le modèle M1 pour calibrer ces deux constantes de réglage.

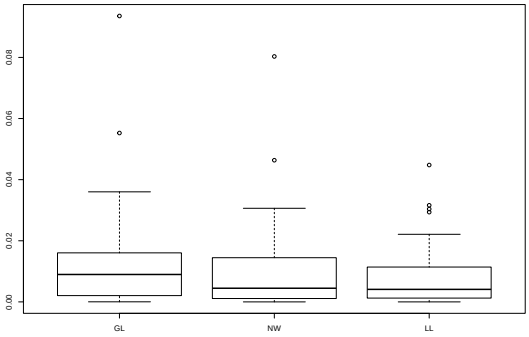
Après cela, nous calculons l'estimateur adaptive $\widehat{m}_{\widehat{h}}(x)$ (tel que défini dans (1.49)). De plus, pour comparer avec notre estimateur adaptatif, nous calculons l'estimateur de Nadaraya-Watson (NW) noté $\widehat{m}_h^{NW}$ et l'estimateur linéaire local (LL) (voir [80, Section 2.2.2]) noté $\widehat{m}_h^{LL}$ de m également. De plus, la validation croisée est utilisée pour sélectionner la fenêtre $h > 0$ pour $\widehat{m}_h^{NW}$ et $\widehat{m}_h^{LL}$.

Dans les graphiques ci-dessous, nous désignons par "GL", "NW" et "LL" le risque ponctuel au carré (calculé avec 50 répétitions Monte Carlo) correspondant à $\widehat{m}_{\widehat{h}}(x)$, $\widehat{m}_h^{NW}$ and $\widehat{m}_h^{LL}$ respectivement. La comparaison entre ces risques ponctuels au carré est montrée dans des "boxplots" dans Figures 1.7 and 1.8. Ces résultats numériques montrent que les performances de notre estimateur sont assez satisfaisantes.

Finalement, nous nous référons à la Section 3.3 pour plus de détails sur les simulations numériques.

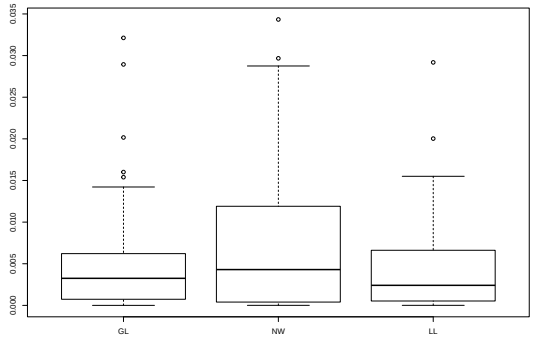


(a) Model M1: estimation à $x = -2$

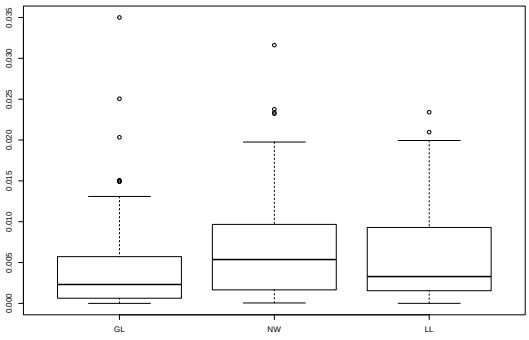


(b) Model M1: estimation à $x = 1.25$

Figure 1.7: Boxplot pour 50 répétitions de rééchantillonnage de Monte Carlo avec une taille d'échantillon $n = 200$, dans le cas $X \sim U([-5, 5])$, en comparaison avec NW et LL (en utilisant le noyau Epanechnikov).



(a) Model M1: estimation à $x = -2$



(b) Model M1: estimation à $x = 1.25$

Figure 1.8: Boxplot pour 50 répétitions de rééchantillonnage de Monte Carlo avec une taille d'échantillon $n = 200$, dans le cas $X \sim \mathcal{N}(0, 1.5)$, en comparaison avec NW et LL (en utilisant le noyau Epanechnikov).

Chapter 2

Statistical deconvolution of free Fokker-Planck equation at fixed time

In this chapter, we consider the problem of statistical deconvolution for Fokker-Planck equation at fixed time $t > 0$. We first give a brief introduction on the free Fokker-Planck equation in Section 2.1, especially with a null potential, in which we can associate the equation with a stochastic differential system (more precisely, a Dyson's Brownian motion) describing the evolution of eigenvalues of a Hermitian matrix. Section 2.2 presents the free deconvolution procedure via subordination method in general setting. Furthermore, in Section 2.3, an introduction of Dyson's Brownian motion is presented and some relevant results are given. After that, we will present the main results of this chapter. Section 2.4 presents a strategy to apply free deconvolution by subordinations method in our framework, and then in Section 2.5, we propose a statistical estimator for the density function of the initial condition p_0. The study of consistency and mean integrated squared error (MISE) for our estimator is examined in Section 2.7. Furthermore, some numerical implements are shown as illustrations in Section 2.8.

This chapter is a version of the paper *Statistical deconvolution of free Fokker-Planck equation at fixed time*([40]) written in a collaboration with Mylène Maïda, Thanh Mai Pham Ngoc, Vincent Rivoirard and Viet Chi Tran, accepted for publication in *Bernoulli*. Furthermore, some supplementary proofs are presented in this chapter. To give a more precise connection, Section 2.3 contains Section 2.1 in [40], Section 2.4 corresponds to Section 2.2 in [40], Section 2.5 associates with Section 2.3 in [40] and Section 2.6 relates to Section 3 in [40]. Finally, Section 2.7 and 2.8 correspond to Section 4 and 5 in [40], respectively.

2.1 The free Fokker-Planck equation

For equations coming from Physics, in particular the Navier-Stokes or the Burger equations, considering random initial data is an interesting way of taking into account the uncertainty due to phenomena such as turbulence (A.S.Monin et al. [46] or M.J.Vishik and A.V.Fursikov [62], see also the survey of M.Vergassolla et al. [61]). Random initial conditions can also traduce a lack of information on the initial state of a system, as in weather forecasting (see A.J.Chorin et al. [22]).

In this work, we are interested in the Fokker-Planck PDE which is a special case of McKean-Vlasov PDE (or so-called the granular media equation) equipped with a logarithmic Coulomb interaction. Moreover, we examine the case where the external potential function is null.

In particular, as mentioned in Section 1.2.1, we consider in our framework the one-dimensional free Fokker-Planck equation as follows:

$$\frac{\partial \mu_t}{\partial t} = -\frac{\partial}{\partial x}\mu_t(H\mu_t). \tag{2.1}$$

where $(\mu_t)_{t\geq 0}$ is a family of probability measures on $\mathbb{R}$, with initial condition μ_0, and $H\mu$ denotes the Hilbert transform of a probability measure μ on $\mathbb{R}$, defined for any $x \in \mathbb{R}$ by:

$$H\mu(x) := \lim_{\varepsilon \to 0} \int_{\mathbb{R}\setminus[x-\varepsilon, x+\varepsilon]} \frac{1}{x-y} d\mu(y).$$

For a test function $g \in \mathcal{C}^1(\mathbb{R})$, we get

$$\int_{\mathbb{R}} g(x) d\mu_t(x) = \int_{\mathbb{R}} g(x) d\mu_0(x) + \int_0^t \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{g'(x)}{x-y} d\mu_s(x) d\mu_s(y) \right) ds, \quad t > 0.$$

Moreover, suppose that μ_0 possesses a density function $p_0(\cdot) = p(0, \cdot)$ such that $d\mu_0(x) = p_0(x)dx$, by Proposition 1.2.11 the measure μ_t has a density function $p(t, \cdot)$ such that $d\mu_t(x) = p(t, x)dx$ and we can rewrite the above equation (2.1) or $t \in \mathbb{R}^+$, $x \in \mathbb{R}$ as:

$$\partial_t p(t, x) = -\partial_x \big[p(t, x).\big(Hp(t, x)\big)\big], \tag{2.2}$$

A natural question is to estimate the random initial condition, given the observation of the PDE solution at given fixed time $t > 0$. For linear PDEs, this inverse problem is solved by deconvolution techniques, and this has been explored for the PDEs such as the heat equation or the wave equation by Pensky and Sapatinas [48, 49]. For 1-d heat equation, it is known that the solution at time t, say $\nu_t(dx)$, is the classical convolution of the initial condition $\nu_0(dx)$ with Green function G_t, which is a Gaussian transition function associated with a 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. The fact is that the Fokker-Planck equation has 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 with a Gaussian distribution is replaced by the *free convolution* with a *semi-circular* law σ_t of variance $t > 0$ characterized by its density with respect to the Lebesgue measure:

$$f_{\sigma_t}(x) = \frac{1}{2\pi t}\sqrt{4t - x^2}\mathbb{1}_{[-2\sqrt{t}, 2\sqrt{t}]}(x). \tag{2.3}$$

More specifically, if $\mathbf{x}_0$ admits the spectral measure μ_0 (note that here the law is considered under the sense of operator), then $\mathbf{x}_t = \mathbf{x}_0 + \mathbf{h}_t$ admits (see also Section 1.2.2)

$$\mu_t = \mu_0 \boxplus \sigma_t,$$

as spectral measure, where the operation $\boxplus$ is the *free convolution* whose definition is introduced later in Section 2.2.1.
Furthermore, it can be proved that the density $p(t, .)$ of μ_t solves equation (2.2). Therefore, in this approach, the problem of estimating the initial condition of equation (2.2) leads to take a *free deconvolution* into account.

Remark. In addition, it is also possible to consider the deconvolution problem without a direct link with PDEs, directly at the level of matrices. The problem can be stated as follows: let A and B be two large matrices, at least one of them, say B, being random. Can you reconstruct accurately the spectrum of A from the observation of its noisy version $A + B$? This has been studied under various assumptions, e.g in J.Bun et al. [11], O.Ledoit and M.Wolf [38] and more recently by P.Tarrago in [56] but there is still a lot of work to do in this direction.

2.2 Free deconvolution by subordination method

Given a fixed time $t > 0$, considering μ_t to be the solution of the Fokker-Planck equation, we have $\mu_t = \mu_0 \boxplus \sigma_t$. The problem of recovering μ_0 from knowledge on μ_t is then a free deconvolution. In this section, we introduce first the definition of *free (additive) convolution* between two probability measures on $\mathbb{R}$ and the corresponding literature in our framewok. We then present a procedure for conducting the free deconvolution in a general setting using subordination methods developed by O.Arizmendi et al in [2].

2.2.1 Cauchy transform and free convolution

To begin with, we introduce some specific functions and notations in complex analysis that will be useful in the sequel.

Definition 2.2.1. Let μ be a probability measure on $\mathbb{R}$, the **Cauchy transform** (or also sometimes called Stieltjes transform) of μ is defined by:

$$G_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(x)}{z - x}, \quad z \in \mathbb{C}\backslash\mathbb{R}. \tag{2.4}$$

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-plane $\mathbb{C}^+ := \{z \in \mathbb{C} | \mathrm{Im}(z) > 0\}$. Moreover, we also get that $G_\mu(\mathbb{C}^+) \subset \mathbb{C}^-$. Indeed, using definition of G_μ, for any $z = u + i.v \in \mathbb{C}^+$ one has:

$$\mathrm{Im}\big(G_\mu(z)\big) = \mathrm{Im}\left(\int_{\mathbb{R}} \frac{u - i.v}{(u-x)^2 + v^2} d\mu(x)\right) = -\int_{\mathbb{R}} \frac{v d\mu(x)}{(u-x)^2 + v^2} < 0.$$

Observe that G_μ does not vanish on $\mathbb{C}\backslash\mathbb{R}$, so we can define the reciprocal Cauchy-transform of μ by:

$$F_\mu(z) = \frac{1}{G_\mu(z)}, \quad z \in \mathbb{C}^+. \tag{2.5}$$

In particular, recall that the **semi-circular distribution** σ_t of variance $t > 0$ is the probability measure with density

$$f_{\sigma_t}(x) = \frac{1}{2\pi t}\sqrt{4t - x^2}dx,$$

on the interval $[-2\sqrt{t}, 2\sqrt{t}]$. Then, its Cauchy transform is (more details of computation are given later in Appendix A.1.2):

$$G_{\sigma_t}(z) = \frac{z - \sqrt{z^2 - 4t}}{2t}, \quad z \in \mathbb{C}^+. \tag{2.6}$$

Moreover, using this explicit formula of G_{σ_t}, for $z \in \mathbb{C}^+$, by computation we obtain

$$z - F_{\sigma_t}(z) = z - \frac{2t}{z - \sqrt{z^2 - 4t}} = z - \frac{2t\big(z + \sqrt{z^2 - 4t}\big)}{z^2 - (z^2 - 4t)} = \frac{z - \sqrt{z^2 - 4t}}{2},$$

thus,

$$z - F_{\sigma_t}(z) = t.G_{\sigma_t}(z). \tag{2.7}$$

Definition 2.2.2. For $\alpha > 0$ and $\beta > 0$, consider sub-domains of $\mathbb{C}^+$:

$$D_\alpha := \{u + i.v \ : \ |u| < \alpha v, v > 0\}, \quad \text{and} \quad D_{\alpha,\beta} := \{z = u + i.v \in \mathbb{C} \ : \ |u| < \alpha v, \quad v > \beta\}.$$

The following lemma shows an asymptotic behavior of G_μ on D_α.

Lemma 2.2.3. *For $\alpha > 0$,*

$$\lim_{|z|\to\infty,\ z\in D_\alpha} zG_\mu(z) = 1.$$

Proof. First, for all $z \in D_\alpha$, for $x \in \mathbb{R}$, one gets $\left|\frac{x}{z-x}\right|^2 \leq \alpha^2 + 1$. Indeed, write $z = u + iv$, we have $|u| < \alpha v$, and $\left|\frac{x}{z-x}\right|^2 = \frac{x^2}{(u-x)^2+v^2} \leq \frac{x^2}{(u-x)^2+\frac{u^2}{\alpha^2}}$. Moreover, we have

$$\frac{x^2}{(u-x)^2+\frac{u^2}{\alpha^2}} \leq \alpha^2+1 \quad \Leftrightarrow \quad x^2 \leq \left(1+\alpha^2\right)\left[(u-x)^2+\frac{u^2}{\alpha^2}\right].$$

The last inequality holds since we have

$$\left(1+\alpha^2\right)\left[(x-u)^2+\frac{u^2}{\alpha^2}\right] \geq \left[(x-u).1+\frac{u}{\alpha}.\alpha\right]^2 = x^2.$$

Now, for $T > 0$ and $z = u + iv \in D_\alpha$, i.e. $v > 0$ and $|u| < \alpha.v$, we consider

$$\begin{aligned}
|zG_\mu(z) - 1| = \left|\int_{\mathbb{R}} \frac{x}{z-x} d\mu(x)\right| &\leq \int_{-T}^{T} \left|\frac{x}{z-x}\right| d\mu(x) + \left(\sqrt{\alpha^2+1}\right).\mu\left(\{x : |x| \geq T\}\right) \\
&\left(\text{but } \frac{|x|}{|z-x|} = \frac{|x|}{\sqrt{(u-x)^2+v^2}} \leq \frac{|x|}{v}\right) \\
&\leq \int_{-T}^{T} \frac{|x|}{v} d\mu(x) + \left(\sqrt{\alpha^2+1}\right).\mu\left(\{x : |x| \geq T\}\right) \\
&\leq \frac{T}{v}.\mu\big(\,(-T,T)\,\big) + \left(\sqrt{\alpha^2+1}\right).\mu\left(\{x : |x| \geq T\}\right),
\end{aligned}$$

and hence for any $T > 0$

$$\limsup_{|z|\to+\infty,\ z\in D_\alpha} |zG_\mu(z) - 1| \leq \left(\sqrt{\alpha^2+1}\right).\mu\left(\{x : |x| \geq T\}\right).$$

Moreover, we also observe that

$$\lim_{T\to+\infty} \left(\sqrt{\alpha^2+1}\right).\mu\left(\{x : |x| \geq T\}\right) = 0,$$

this implies the desired result.

□

Furthermore, for $z = u + i.v \in D_{\alpha,\beta}$, i.e. we have $v > \beta > 0$ and $|u| \leq \alpha.v$, then it is true that the imaginary part of G_μ has a lower bound depending on β:

$$\mathrm{Im}\big(G_\mu(u+i.v)\big) = \int_{\mathbb{R}} \frac{-v}{(u-x)^2+v^2} d\mu(x) \geq \int_{\mathbb{R}} \frac{-1}{v} d\mu(x) > -\frac{1}{\beta};$$

For $\gamma > 0$ and $\delta > 0$, we define a triangle-domain in the lower half-plane $\mathbb{C}^-$ as

$$\Gamma_{\gamma,\delta} = \{\xi = u + i.v \in \mathbb{C} \ :\ |u| < \gamma.(-v), \quad 0 > v > -\delta\},$$

which in fact relates to the image of the Cauchy transform of μ. If we consider two domains $\Gamma_{\gamma_1,\delta_1}$ and $\Gamma_{\gamma_2,\delta_2}$, then the intersection of these two domains is $\Gamma_{\gamma,\delta} := \Gamma_{\gamma_1,\delta_1} \cap \Gamma_{\gamma_2,\delta_2}$, with $\delta = \min\{\delta_1, \delta_2\}$, and $\gamma = \min\{\gamma_1, \gamma_2\}$. Since both δ and γ are strictly positive, the domain $\Gamma_{\gamma,\delta}$ is nonempty.
Now, fix $\alpha > 0$ and $\beta > 0$, actually, the image of $D_{\alpha,\beta}$ by G_μ (i.e the set $G_\mu(D_{\alpha,\beta})$) contains a domain $\Gamma_{\gamma(\mu),\delta(\mu)}$ for some $\gamma(\mu) > 0$ depending also on $\delta(\mu)$ with $0 < \delta(\mu) \leq \frac{1}{\beta}$. Now, let K_μ be the inverse function of G_μ, defined on $\Gamma_{\gamma(\mu),\delta(\mu)}$, and let

$$R_\mu(\xi) = K_\mu(\xi) - \frac{1}{\xi}.$$

The function R_μ is called the R-**transform** of μ. Then given two probability measures μ_1 and μ_2, there exists a unique probability measure μ, such that

$$R_\mu = R_{\mu_1} + R_{\mu_2} \tag{2.8}$$

on a domain $\Gamma_{\gamma(\mu),\delta(\mu)} \subseteq \left(\Gamma_{\gamma(\mu_1),\delta(\mu_1)} \cap \Gamma_{\gamma(\mu_2),\delta(\mu_2)}\right)$ where these three functions are defined. The measure μ is denoted $\mu_1 \boxplus \mu_2$, and is called the **free convolution** of μ_1 and μ_2. The free convolution was first defined by D.Voiculescu for measures with compact support in [63], extended to measures with finite variance by H.Maassen in [42], and then to general probability measures on $\mathbb{R}$ by H.Bercovici and D.Voiculescu in [5].

In particular, as we will see later in Proposition 2.3.5 solution μ_t of the free Fokker-Planck equation can be represented as free convolution between the initial condition μ_0 and the semi-circular law σ_t (defined in (2.3)). In other words, for solution μ_t of the Fokker-Planck equation starting from initial condition μ_0 we have:

$$\mu_t = \mu_0 \boxplus \sigma_t. \tag{2.9}$$

Additionally, the following theorem (see e.g. [1, Theorem 2.4.3]) allows one to recover a probability measure on $\mathbb{R}$ from the associating Cauchy transform.

Theorem 2.2.4 (Stieltjes inversion formula)**.** *Suppose ν is a probability measure on $\mathbb{R}$ and G_ν is its Cauchy transform. For any open interval I on $\mathbb{R}$ we have*

$$\nu(I) = -\lim_{\varepsilon \to 0^+} \frac{1}{\pi} \int_I Im G_\nu(x + i.\varepsilon)\, dx. \tag{2.10}$$

The construction of subordination functions for free deconvolution

In general, it is difficult to obtain explicit descriptions of free convolution, except for some special circumstances. The approach using subordination functions in [4] gives an idea to approximate free convolution through a fixed-point equation concerning the Cauchy transform of measures.
An important property of free (additive) convolution is called *subordination*, which can be stated as follows: let μ_1 and μ_2 be two probability measures on $\mathbb{R}$, there exists an analytic function $v : \mathbb{C}^+ \to \mathbb{C}^+$ such that $G_{\mu_1 \boxplus \mu_2}(z) = G_{\mu_1}\big(v(z)\big)$ $\forall z \in \mathbb{C}^+$, and $\lim_{y \to +\infty} \frac{v(iy)}{iy} = 1$.
To be more precise, the definition of subordination functions (the earliest results in this area) was established due to Voiculescu [64, Proposition 4.4] for compactly supported probability measures, and was proved in full generality by P.Biane [8, Theorem 3.1]. Accordingly we introduce in the following theorem:

Theorem 2.2.5 (Theorem-Definition). *Let μ_1 and μ_2 be probability measures on $\mathbb{R}$. Then, there exist unique subordination functions v_1 and v_2 from $\mathbb{C}^+$ onto $\mathbb{C}^+$ such that:*

(i) for $z \in \mathbb{C}^+$, $Im\big(v_1(z)\big) \geq Im(z)$ and $Im\big(v_2(z)\big) \geq Im(z)$, with

$$\lim_{y\to+\infty} \frac{v_1(iy)}{iy} = \lim_{y\to+\infty} \frac{v_2(iy)}{iy} = 1;$$

(ii) for $z \in \mathbb{C}^+$, $F_{\mu_1\boxplus\mu_2}(z) = F_{\mu_1}\big(v_1(z)\big) = F_{\mu_2}\big(v_2(z)\big)$, and $v_1(z) + v_2(z) = z + F_{\mu_1\boxplus\mu_2}(z)$.

Using this result, Belinschi and Bercovici [4, Section 4] introduce a fixed-point construction of the subordination functions, which Arizmendi et al. [2] adapt for the free deconvolution problem. In our framework, we will apply their result in the special case of the free deconvolution by a semi-circular distribution defined in (2.3).
More precisely, we observe that a crucial property of the reciprocal Cauchy-transform F_μ (defined in (2.5)) is (see also [8, Theorem 3.1]):

$$\mathrm{Im}F_\mu(z) \geq \mathrm{Im}(z), z \in \mathbb{C}^+, \quad \text{and} \quad \lim_{y\to+\infty} \frac{F_\mu(iy)}{iy} = 1. \tag{2.11}$$

Due to [4], this leads to the fact that F_μ is invertible in a region where $|z| = |x+iy|$ is sufficiently large, provided that x/y remains bounded. Using the fact that $F_\mu^{<-1>}(.) = G_\mu^{<-1>}(\frac{1}{.})$, the definition of free convolution in (2.8) (using the R-transform) can be now rewritten as

$$F_{\mu_1}^{<-1>}(\xi) + F_{\mu_2}^{<-1>}(\xi) = \xi + F_{\mu_1\boxplus\mu_2}^{<-1>}(\xi),$$

for ξ on some domain $D_{\alpha',\beta'}$. Then, a natural idea for constructing the subordination functions is to consider the functions $v_1 := F_{\mu_1}^{<-1>} \circ F_{\mu_1\boxplus\mu_2}$ and $v_2 := F_{\mu_2}^{<-1>} \circ F_{\mu_1\boxplus\mu_2}$, defined for z on a domain $\widetilde{D}_\beta := \{z = i.y \in \mathbb{C}^+ : y \geq \beta\}$ with some $\beta > 0$ sufficiently large, and satisfying $\lim_{y\to+\infty} v_j(i.y)/(i.y) = 1$, with $j = 1, 2$, as well as the property (ii)-Theorem 2.2.5 in some open set. In addition, set $h_1(w) = F_{\mu_1}(w) - w$ and $h_2(w) = F_{\mu_2}(w) - w$, for $w \in \mathbb{C}^+$. Since $\mathrm{Im}(F_{\mu_1}(w)) \geq \mathrm{Im}(w)$ and $\mathrm{Im}(F_{\mu_2}(w)) \geq \mathrm{Im}(w)$ for all $w \in \mathbb{C}^+$, the imaginary part of h_1 and h_2 are non-negative. From (ii)-Theorem 2.2.5 above, one can rewrite as:

$$v_1(z) = z + F_{\mu_2}(v_2(z)) - v_2(z) = z + h_2(v_2(z)), \quad z \in \widetilde{D}_\beta,$$

and similarly

$$v_2(z) = z + F_{\mu_1}(v_1(z)) - v_1(z) = z + h_1(v_1(z)), \quad z \in \widetilde{D}_\beta,$$

thus, a fixed-point equation can be established as:

$$v_1(z) = z + h_2\big(z + h_1(v_1(z))\big), \quad z \in \widetilde{D}_\beta. \tag{2.12}$$

One can then consider a function $g : \mathbb{C}^+ \times \mathbb{C}^+ \to \mathbb{C}^+$ defined by:

$$g(z, v) := z + h_2(z + h_1(v)), \quad z, v \in \mathbb{C}^+.$$

For $z, v \in \mathbb{C}^+$, as $\mathrm{Im}(h_1(v)) \geq 0$, one has $\mathrm{Im}(z + h_1(v)) > \mathrm{Im}(z) > 0$, and similarly, $\mathrm{Im}(z) + \mathrm{Im}\big(h_2(z + h_1(v))\big) > 0$. As a result, $\mathrm{Im}(g(z, v)) > 0$ for all $z, v \in \mathbb{C}^+$, and thus, the function g is well-defined on $\mathbb{C}^+ \times \mathbb{C}^+$. The local existence of the function v_j implies that some of the functions $g(z, v)$ have a fixed-point. Therefore, Theorem A.1.1 (in the Appendices) yields a globally defined function v_1 such that $g(z, v_1(z)) = v_1(z)$. Moreover, note that

$$\mathrm{Im}(v_1(z)) = \mathrm{Im}(z) + \mathrm{Im}\big(h_2\big(z + h_1(v_1(z))\big)\big) \geq \mathrm{Im}(z), \quad z \in \mathbb{C}^+.$$

The other function is obtained as $v_2(z) = z + h_1(v_1(z))$. In addition, the uniqueness of the functions v_j is implied from the uniqueness of Denjoy-Wolff points. We do not mention here all the details, since similar arguments will be actually given later, for instance in section 2.4.1.

2.3 A particle system - Dyson's Brownian motion

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 and let us denote that matrix by $X_n(t)$ for given time $t > 0$(assumed to be fixed in the sequel) and the number of observations n. The associating eigenvalues $\left(\lambda_1^n(t), ..., \lambda_n^n(t)\right)$ of matrix $X_n(t)$ solve a particular SDE that is Dyson Brownian motion as stated in Theorem 2.3.4, and moreover, we consider an empirical spectral measure constructed from these eigenvalues. This section presents an introduction of Dyson Brownian motion and a study for the (weak) convergence of empirical spectral measure that we consider in our framework. We refer to the book of Anderson, Guionnet and Zeitouni ([1], section 4.3) as a main reference for this topic.

Let $\mathrm{Mat}_n(\mathbb{C})$ denote the space of $n \times n$ matrices with entries belonging to $\mathbb{C}$. Moreover, we denote by $\mathcal{H}_n(\mathbb{C}) := \{H_n \in \mathrm{Mat}_n(\mathbb{C}) : (H_n)^* = H_n\}$ the subset of Hermitian matrices.

Definition 2.3.1. Let $\left(B_{k,l}, \widetilde{B}_{k,l}, 1 \leq k, l \leq n\right)$ be a collection of independent and identically distributed (i.i.d.) real valued standard Brownian motions, the Hermitian Brownian motion, denoted $H_n \in \mathcal{H}_n(\mathbb{C})$, is the random process with entries $\{(H_n(t))_{k,l}, t \geq 0, 1 \leq k, l \leq n\}$ equal to

$$(H_n)_{k,l} = \begin{cases} \frac{1}{\sqrt{2n}}\left(B_{k,l} + i.\widetilde{B}_{k,l}\right), & \text{if } k < l, \\ \frac{1}{\sqrt{n}} B_{k,k}, & \text{if } k = l. \end{cases} \tag{2.13}$$

We study the process of eigenvalues of time-dependent matrices. Let Δ_n denote the open simplex

$$\Delta_n = \left\{(x_i)_{1 \leq i \leq n} \in \mathbb{R}^n : x_1 < x_2 < ... < x_{n-1} < x_n\right\}, \tag{2.14}$$

with closure $\overline{\Delta_n}$.

Assumption 2.3.2. *Let us detail the initial condition. We assume here that for each $n \geq 1$ and almost surely, $X_n(0)$ is a diagonal random matrix with entries $\left(\lambda_1^n(0), ..., \lambda_n^n(0)\right) \in \overline{\Delta_n}$ that we suppose to be the ordered collection of d_j with random variables d_j i.i.d. with distribution μ_0. Moreover, we assume that these random variables are independent of $\left(H_n(t)\right)_{t \geq 0}$. Additionnally, we suppose that for the measure μ_0 there exists a constant $C > 0$ such that for sufficiently large $\kappa > 0$,*

$$\mu_0\{|x| > \kappa\} \leq \frac{C}{\kappa}. \tag{2.15}$$

Notice that as a consequence of the Assumption (2.3.2), the empirical measure $\mu_0^n = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j^n(0)}$ converges weakly as n goes to infinity towards the measure $\mu_0 \in \mathcal{M}_1(\mathbb{R})$. Moreover, Assumption (2.3.2) also implies that:

$$C_0 := \sup_{n \geq 1} \frac{1}{n} \sum_{j=1}^n \log\left(\lambda_j^n(0)^2 + 1\right) = \sup_{n \geq 1} \int \log(x^2 + 1)\, d\mu_0^n(x) < \infty \quad a.s. \tag{2.16}$$

A more general initial condition could be to consider for $X_n(0)$ a Hermitian random matrix of $\mathcal{H}_n$ with eigenvalues $\left(d_1, \dots d_n\right)$ where d_i are random variables with distribution μ_0. The next lemma (Lemma 2.3.3) explains that we can recover without loss of generality a diagonal matrix with entries $\left(\lambda_1^n(0), ..., \lambda_n^n(0)\right) \in \overline{\Delta_n}$. However, the independence of the eigenvalues d_i is then lost and further hypotheses and work are needed to adapt the proofs of the present manuscript.

Lemma 2.3.3. *As the law of $(H_n(t))_{t\geq 0}$ is invariant under conjugation by a fixed unitary matrix, one can assume without loss of generality that the matrix $X_n(0)$ is diagonal.*

Proof. The law of $\big(H_n(t)\big)_{t\geq 0}$ is invariant under the action of the unitary groups, that is $\big(UH_n(t)U^*\big)_{t\geq 0}$ has the same distribution as $\big(H_n(t)\big)_{t\geq 0}$ if U belongs to unitary groups. Therefore, the law of $\big(\lambda_j^n(t)\big)_{t\geq 0}$ does not depend on the basis of eigenvectors of $X_n(0)$. More precisely, by diagonalizing $X_n(0)$, we can write:

$$X_n(0) + H_n(t) = UD_n(0)U^* + H_n(t) = U\big(D_n(0) + U^*H_n(t)U\big)U^*.$$

The fact is that $D_n(0)+U^*H_n(t)U$ has the same spectral as $X_n(0)+H_n(t)$, and $U^*H_n(t)U \overset{law}{=} H_n(t)$. As a result, we can assume without loss of generality that $X_n(0)$ is diagonal. □

For $t \geq 0$, let $\lambda^n(t) = \big(\lambda_1^n(t), ..., \lambda_n^n(t)\big) \in \overline{\Delta_n}$ denote the ordered collection of eigenvalues of

$$X_n(t) = X_n(0) + H_n(t), \tag{2.17}$$

with H_n is defined as in Definition 2.3.1 and independent with $X_n(0)$.

Theorem 2.3.4 (Dyson)**.** *Let $\Big(X_n(t)\Big)_{t\geq 0}$ be as in (2.17), with eigenvalues $\big(\lambda^n(t)\big)_{t\geq 0}$ and $\big(\lambda^n(t)\big) \in \overline{\Delta_n}$ for all $t \geq 0$. Then, the processes $\big(\lambda^n(t)\big)_{t\geq 0}$ are semi-martingales. Their joint law is the unique distribution on $C\left(\mathbb{R}^+, \mathbb{R}^n\right)$ so that*

$$\mathbb{P}\Big(\forall t > 0, (\lambda_1^n(t), ..., \lambda_n^n(t)) \in \Delta_n\Big) = 1,$$

which is a weak solution to the system

$$d\lambda_j^n(t) = \frac{1}{\sqrt{n}}d\beta_j(t) + \frac{1}{n}\sum_{k\neq j}\frac{dt}{\lambda_j^n(t) - \lambda_k^n(t)}, \quad 1 \leq j \leq n, \tag{2.18}$$

with initial condition $\lambda_j^n(0)$ and where β_j are i.i.d. standard Brownian motion.

According to this theorem, the process $(\lambda^n(t))_{t\geq 0}$ is a vector of semi-martingales, whose evolution is described by a stochastic differential system. We will not mention the proof of this theorem here, since it is beyond the scope of our study, (see e.g. Theorem 4.3.2 in [1], nevertheless, notice that $(\lambda_j^n(0))$ is considered to be random in our case).

Now, for the process of n particles, we consider the empirical spectral measure associating with $\big(\lambda_j^n(t)\big)_{j=1,n}$ at time $t > 0$ as follows

$$\mu_t^n = \frac{1}{n}\sum_{j=1}^{n}\delta_{\lambda_j^n(t)}. \tag{2.19}$$

For $n \geq 1$, we can actually associate the free Fokker-Planck equation (2.2) with the interacting n-particle system $\lambda^n(t) = \big(\lambda_1^n(t), ..., \lambda_n^n(t)\big)_{t\geq 0}$ defined above. A brief heuristic derivation, to reach the equation (2.2) from the SDE (2.18) is given in the following , (see also [9]).

Heuristic of the link between the particle system and free Fokker-Planck equation: To be precise, the generator of the eigenvalue diffusion process defined in (2.18) is thus the sum of the diffusive term $\frac{1}{2\sqrt{n}}\Delta$ (where Δ is the usual Laplace operator), and an entropic term

$$\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{k\neq j}\frac{1}{\lambda_j^n(t)-\lambda_k^n(t)}\right)\frac{\partial}{\partial\lambda_j^n(t)}.$$

We now observe that when n is large, the Brownian term in the equation (2.18) becomes small in front of the other terms, and in finite time, the process behaves like a dynamical system with a small random perturbation. In the large n limit, trajectory of the eigenvalue vector behaves as that of the flow of the deterministic vector field

$$\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{k\neq j}\frac{1}{\lambda_j^n(t)-\lambda_k^n(t)}\right)\frac{\partial}{\partial\lambda_j^n(t)};$$

Assume that the empirical measure $\frac{1}{n}\sum_{j=1}^{n}\delta_{\lambda_j^n(t)}$ converges, as $n\to\infty$, towards a limit distribution $\mu_t(dx)=p(t,x)dx$, then for any smooth test function g one has

$$\frac{\partial}{\partial t}\frac{1}{n}\sum_{j=1}^{n}g(\lambda_j^n(t))\simeq\frac{1}{n^2}\sum_{j=1}^{n}g'(\lambda_j^n(t))\left(\sum_{k\neq j}\frac{1}{\lambda_j^n(t)-\lambda_k^n(t)}\right)$$

and taking the large n limit, one gets

$$\frac{\partial}{\partial t}\int g(x)p(t,x)dx=\int g'(x)\left(Hp(t,x)\right)p(t,x)dx.$$

The flow equation gives then the free Fokker-Planck equation (2.2) for $p(t,\centerdot)$.

Weak convergence of empirical measure μ_t^n

Before studying the (weak) convergence of μ_t^n, we first have some relating results on applying Itô's calculus.

Itô's calculus

For a continuous function f on $\mathbb{R}$, we denote

$$\langle f,\mu_t^n\rangle=\int f(\lambda)\mu_t^n(d\lambda)=\frac{1}{n}\sum_{j=1}^{n}f(\lambda_j^n(t)). \tag{2.20}$$

First, from its differential form, we can rewrite the SDE (2.18) as in the following integral equation:

$$\lambda_j^n(t)=\lambda_j^n(0)+\frac{1}{\sqrt{n}}\int_0^t d\beta_s^j+\frac{1}{n}\int_0^t\sum_{k\neq j}\frac{1}{\lambda_j^n(s)-\lambda_k^n(s)}ds. \tag{2.21}$$

Theorem 2.3.4 shows that $\left(\lambda^n(t)=(\lambda_1(t),...,\lambda_n(t))\right)_{0\leq t\leq T}$ are continuous semi-martingales and then we can apply Itô's calculus (Theorem A.2.1 in the Appendices) to have for any function $f\in\mathcal{C}^2(\mathbb{R})$ and $T>0$:

$$f(\lambda_j^n(t))=f(\lambda_j^n(0))+\frac{1}{n}\int_0^t\sum_{k\neq j}\frac{f'(\lambda_j^n(s))}{\lambda_j^n(s)-\lambda_k^n(s)}ds+\frac{1}{\sqrt{n}}\int_0^t f'(\lambda_j^n(s))d\beta_s^j+\frac{1}{2}\int_0^t f''(\lambda_j^n(s))ds,$$

and hence,

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^n(t)) = \frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^n(0)) + \frac{1}{n^2}\sum_{j=1}^{n}\int_0^t \sum_{k\neq j} \frac{f'(\lambda_j^n(s))}{\lambda_j^n(s)-\lambda_k^n(s)} ds + \frac{1}{n\sqrt{n}}\sum_{j=1}^{n}\int_0^t f'(\lambda_j^n(s)) d\beta_s^j + \frac{1}{2n}\sum_{j=1}^{n}\int_0^t f''(\lambda_j^n(s)) ds.$$

We can rewrite as:

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^n(t)) = \frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^n(0)) + \frac{1}{2n^2}\sum_{j=1}^{n}\int_0^t \sum_{k\neq j} \frac{f'(\lambda_j^n(s)) - f'(\lambda_k^n(s))}{\lambda_j^n(s)-\lambda_k^n(s)} ds + \frac{1}{n\sqrt{n}}\sum_{j=1}^{n}\int_0^t f'(\lambda_j^n(s)) d\beta_s^j + \frac{1}{2n}\sum_{j=1}^{n}\int_0^t f''(\lambda_j^n(s)) ds,$$

or equivalently,

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^n(t)) = \frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^n(0)) + \frac{1}{2n}\sum_{j=1}^{n}\int_0^t \Big[\Big(\frac{1}{n}\sum_{k\neq j} \frac{f'(\lambda_j^n(s)) - f'(\lambda_k^n(s))}{\lambda_j^n(s)-\lambda_k^n(s)}\Big) + f''(\lambda_j^n(s))\Big] ds + \frac{1}{n\sqrt{n}}\sum_{j=1}^{n}\int_0^t f'(\lambda_j^n(s)) d\beta_s^j; \tag{2.22}$$

Finally, we obtain:

$$\int_{\mathbb{R}} f(x)\mu_t^n(dx) = \int_{\mathbb{R}} f(x)\mu_0^n(dx) + \frac{1}{2}\int_0^t \iint \left(\int_0^1 f''\big(\theta x + (1-\theta)y\big) d\theta\right) \mu_s^n(dy)\mu_s^n(dx) ds + M_f^n(t), \tag{2.23}$$

with M_f^n is the martingale given for $0 < t \leq T$ by

$$M_f^n(t) = \frac{1}{n\sqrt{n}}\sum_{j=1}^{n}\int_0^t f'(\lambda_j^n(s)) d\beta_s^j. \tag{2.24}$$

By stochastic calculus, we have a bracket for M_f^n as

$$\begin{aligned}
\langle M_f^n\rangle_t = \sum_{j=1}^{n}\langle \frac{1}{n\sqrt{n}}\int_0^t f'(\lambda_j^n(s)) d\beta_s^j\rangle_t &= \sum_{j=1}^{n}\frac{1}{n^3}\int_0^t \Big(f'(\lambda_j^n(s))\Big)^2 d\langle\beta^j\rangle_s \\
&\text{(the quadratic variation of stochastic integral of } f' \text{ w.r.t. } \beta) \\
&= \frac{1}{n^2}\int_0^t \left(\frac{1}{n}\sum_{j=1}^{n}\Big(f'(\lambda_j^n(s))\Big)^2\right) ds \\
&\text{(since } \langle\beta^j\rangle_s = s, \text{ as } B_s^j \text{ is Brownian motion for all } 1\leq j\leq n \text{)} \\
&= \frac{1}{n^2}\int_0^t \left(\int \big(f'(x)\big)^2 \mu_t^n(dx)\right) ds \quad \text{(by formula (2.20))} \\
&\leq \frac{t.\|f'\|_\infty^2}{n^2};
\end{aligned}$$

Furthermore, by Lemma A.2.2 (in the Appendices) with $X_s = \frac{1}{n\sqrt{n}} \sum_{j=1}^{n} f'(\lambda_j^n(s))$ and an upper bound constant for the bracket $\langle M_f^n \rangle_t$ with $t \in [0,T]$ is $A_T = \frac{T.\|f'\|_\infty^2}{n^2}$, we can obtain for any $L > 0$,

$$\mathbb{P}\left(\sup_{0 \le s \le T} \left| M_f^n(s) \right| \ge L \right) \le 2.e^{-\frac{n^2.L^2}{2T.\|f'\|_\infty^2}} \quad .$$

Now, for fixed $T > 0$, we denote by $\mathcal{C}\left([0,T], \mathcal{M}_1(\mathbb{R})\right)$ the space of continuous processes from $[0,T]$ into $\mathcal{M}_1(\mathbb{R})$ (the space of probability measure on $\mathbb{R}$, equipped with its weak topology). We now prove the convergence of the empirical measure μ_t^n, viewed as an element of $C([0,T], \mathcal{M}_1(\mathbb{R}))$.

Proposition 2.3.5. *Let $\left(X_n(0)\right)$ be a sequence of matrices as described in Assumptions 2.3.2. Let $\lambda^n(t) = \left(\lambda_1^n(t), ..., \lambda_n^n(t)\right)_{t \ge 0}$ be the solution of (2.18) with initial condition $\lambda^n(0)$, and set μ_t^n as in (2.19). Then, for any fixed time $T < \infty$, $\left(\mu_t^n\right)_{t \in [0,T]}$ converges almost surely in $\mathcal{C}\left([0,T], \mathcal{M}_1(\mathbb{R})\right)$. Moreover, its limit is the unique measure-valued process $(\mu_t)_{t \in [0,T]}$ so that $\mu_0 = \mu$ and the function*

$$G_{\mu_t}(z) = \int_{\mathbb{R}} \frac{d\mu_t(x)}{z - x} \tag{2.25}$$

satisfies the equation

$$G_{\mu_t}(z) = G_{\mu_0}(z) - \int_0^t G_{\mu_s}(z) \partial_z G_{\mu_s}(z) ds \tag{2.26}$$

for $z \in \mathbb{C} \backslash \mathbb{R}$. Therefore, for $t > 0$, μ_t coincides with the solution of Fokker-Planck equation (2.1), and furthermore, it can be represented as the free convolution,

$$\mu_t = \mu_0 \boxplus \sigma_t. \tag{2.27}$$

This Proposition 2.3.5 corresponds to Proposition 2.3 in [40], (note that in [40] we do not give a proof of [40, Proposition 2.3] and its proof is referred to be presented in this thesis).

Proof of Proposition 2.3.5. For the sake of completeness, we present here the proof of Proposition 2.3.5, which is an adaptation of the proof of [1, Proposition 4.3.10] to our context, to take into account that the initial condition is random. Note that the proof for deterministic initial conditions also appears in the recent survey [52] and has also been extended to other processes on the same type in [51].
We follow the same structure as the proof in [1]. The proof contains two main parts. First, we show that the sequence $(\mu_t^n)_{t \in [0,T]}$ is almost surely pre-compact in $\mathcal{C}\left([0,T], \mathcal{M}_1(\mathbb{R})\right)$, so it converges to a limit point. Then we prove that there is a unique limit point characterized by (2.26).

The sketch of proof is

- **Step 1:** Construct a family of subsets $\mathcal{K}$ which are compact in $\mathcal{C}\left([0,T], \mathcal{M}_1(\mathbb{R})\right)$.

  Lemma 2.3.6. *Let K be an arbitrary compact subset of $\mathcal{M}_1(\mathbb{R})$, let $(f_i)_{i \ge 0}$ be a sequence of bounded functions dense in $\mathcal{C}_0(\mathbb{R})$, and let C_i be compact subsets of $\mathcal{C}([0,T], \mathbb{R})$. Then, the sets*

  $$\mathcal{K} := \{\forall t \in [0,T], \mu_t \in K\} \cap_{i \ge 0} \{t \to \mu_t(f_i) \in C_i\} \tag{2.28}$$

  are compact subsets of $\mathcal{C}\left([0,T], \mathcal{M}_1(\mathbb{R})\right)$.

- **Step 2:** Show that $(\mu_t^n, t \in [0,T])$ is almost surely contained in $\mathcal{K}$.

  Lemma 2.3.7. *Fix $T \in \mathbb{R}^+$, under the assumptions of Proposition 2.3.5, the sequence $\big(\mu_t^n, t \in [0,T]\big)$ is almost surely pre-compact in $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$.*

- **Step 3:** Show that the limit point of $(\mu_t^n, t \in [0,T])$, coincides with solution of the equation (2.26).

- **Step 4:** The limit point also coincides with the solution of the free Fokker-Planck equation which can be presented in the form of free convolution.

Notation: Let $\mathcal{C}_c(\mathbb{R})$ be the set of continuous functions on $\mathbb{R}$ with compact support and $\mathcal{C}_0(\mathbb{R})$ denotes the uniform closure of $\mathcal{C}_c(\mathbb{R})$.
Now, we will go further into details on each main step of the proof mentioned above.

■ **Step 1:**

Lemma 2.3.6 corresponds to [1, Lemma 4.3.13] and we follow its proof.

Proof of Lemma 2.3.6. The space $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$ is Polish, so it is sufficient to prove that the set $\mathcal{K}$ is *sequentially compact* and *closed*. First, by its definition, $\mathcal{K}$ is an intersection of closed sets, so it is closed.
$\mathcal{K}$ is sequentially compact:
Let $\big(\mu_\cdot^m\big)_{m\geq 0}$ be a sequence in $\mathcal{K}$, then we need to show that $\big(\mu_\cdot^m\big)$ has a subsequence converging to a limit point in $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$.
+ By definition of $\mathcal{K}$, for each $i \in \mathbb{N}$, the functions $t \to \mu_t^m(f_i)$ belong to the sets C_i compact in $C([0,T], \mathbb{R})$, and hence we can find a subsequence $\phi_i(m)$ such that the sequence of bounded continuous functions $t \to \mu_t^{\phi_i(m)}(f_i)$ converges in $C([0,T], \mathbb{R})$.

+ As a result, there is a convergence limit point for each i. By a diagonalization procedure, we can find $\phi(m)$ (a subsequence of $\phi_i(m)$) but independent of i, such that for all $i \in \mathbb{N}$, the functions $t \to \mu_t^{\phi(m)}(f_i)$ converge towards some functions $t \to \mu_t(f_i)$ in $C([0,T], \mathbb{R})$. Note that since f_i have compact support, the limit μ_t may not have a mass 1. This is dealt with by the second condition, i.e. the intersection with K compact in $\mathcal{M}_1(\mathbb{R})$.

+ Indeed, $(f_i)_i$ is a family of bounded continuous function, that is dense in $\mathcal{C}_0(\mathbb{R})$, so $(f_i)_{i\geq 0}$ is convergence determining for $K \subset \mathcal{M}_1(\mathbb{R})$. This implies that

$$\lim_{m\to\infty} \int_{\mathbb{R}} f_i d\nu_m = \int_{\mathbb{R}} f_i d\nu, \forall i \in \mathbb{N} \quad \Rightarrow \quad \nu_m \longrightarrow \nu \text{ in } K, \text{ as } m \to \infty.$$

It follows that one can extract a further subsequence, denote $\phi(m_k)$, such that for a fixed dense countable subset of $[0,T]$, the limit μ_t belong to $\mathcal{M}_1$.

+ The continuity of $t \to \mu_t(f_i)$ then shows that $\mu_t \in \mathcal{M}_1(\mathbb{R})$ for all t. Hence, we have shown that $(\mu_\cdot^n)_{n\geq 0}$ has a subsequence which converges to a limit point in $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$, so $\mathcal{K}$ is sequentially compact.

Eventually, we have proved that $\mathcal{K}$ is compact in $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$.

□

■ **Step 2:**

Lemma 2.3.7 corresponds to [1, Lemma 4.3.14] and we follow its proof. More precisely, we now prove that $(\mu_t^n, t \in [0,T])$ is contained in such a set $\mathcal{K}$ compact in $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$, with the compact set $\mathcal{K}$ introduced in Lemma 2.3.6. A sketch for the proof of Lemma 2.3.7 as follows:

1. Show that for $M > C_0 + T$ (and C_0 is defined in the assumptions of Proposition 2.3.5):
$$\mathbb{P}\left(\sup_{t\in[0,T]} \int \log(x^2+1)\mu_t^n(dx) \geq M\right) \leq 2.e^{-\frac{n^2.(M-C_0-T)^2}{8T}};$$

2. . For any $\delta > 0$, $M > 0$ and denote by $[T/\delta] = \sup\{k \in \mathbb{N} : k \leq T/\delta\}$, for any twice continuously differentiable function f so that $\|f''\|_\infty \leq 2^{-1}.M.\delta^{-\frac{3}{4}}$, then
$$\mathbb{P}\left(\sup_{s,t\in[0,T]:|t-s|\leq\delta} |\langle f, \mu_t^n\rangle - \langle f, \mu_s^n\rangle| \leq M.\delta^{\frac{1}{4}}\right) \leq 2.\left(\left[\frac{T}{\delta}\right]+1\right).e^{-\frac{n^2.M^2}{2^7.\delta^{1/2}.\|f'\|_\infty^2}};$$

3. For $M > 0$, denote $K_M := \big\{\mu \in \mathcal{M}_1(\mathbb{R}) : \int \log(x^2+1)\mu(dx) \leq M\big\}$, and consider the compact subset of $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$ as:
$$C_T(f_k, \varepsilon_k) := \bigcap_{m=1}^{+\infty} \left\{\mu_\cdot \in \mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big) : \sup_{|t-s|\leq m^{-2}} |\mu_t(f) - \mu_s(f)| \leq \frac{1}{\varepsilon_k.m^{1/2}}\right\},$$
where $(\varepsilon_k)_{k\geq 1}$ is a sequence of positive numbers decreasing to 0 as $k \to \infty$, and $(f_k)_{k\geq 1}$ is a family of twice continuously differentiable functions dense in $\mathcal{C}_0(\mathbb{R})$. Then a compact subset $\mathcal{K}(M)$ of the set $\mathcal{C}\big([0,T], \mathcal{M}_1(\mathbb{R})\big)$ of continuous probability measures-valued processes, will be defined as
$$\mathcal{K}(M) := K_M \cap \bigcap_{k\geq 1} C_T(f_k, \varepsilon_k),$$
and moreover, we will show that
$$\mathbb{P}\left(\mu_\cdot^n \in \big(\mathcal{K}(M)\big)^c\right) \leq \widetilde{\gamma}(T).e^{-n^2.2^{-7}},$$
with some constant $\widetilde{\gamma}(T) > 0$, only depending on T. This will imply the result in Lemma 2.3.7.

Proof of Lemma 2.3.7. **1.)** For any $f \in \mathcal{C}^2(\mathbb{R})$, one can write
$$\begin{aligned}\iint \frac{f'(x)}{x-y}\mu_s^n(dy)\mu_s^n(dx) &= \frac{1}{2}\iint \frac{f'(x)-f'(y)}{x-y}\mu_s^n(dy)\mu_s^n(dx)\\ &= \frac{1}{2}\iint\left(\int_0^1 f''\big(\theta x + (1-\theta)y\big)d\theta\right)\mu_s^n(dy)\mu_s^n(dx).\end{aligned}$$

Again, apply the Itô's Theorem as in (2.23) with function $f(x) = \log(x^2+1)$ which is twice continuously differentiable on $\mathbb{R}$, to deduce that for any $t \in [0,T]$
$$\begin{aligned}|\langle f, \mu_t^n\rangle| &\leq |\langle f, \mu_0\rangle| + \int_0^t \frac{1}{2}\iint\left(\int_0^1 \big|f''\big(\theta x + (1-\theta)y\big)\big|d\theta\right)\mu_s^n(dy)\mu_s^n(dx) + \big|M_f^n(t)\big|\\ &\leq |\langle f, \mu_0\rangle| + t.\frac{\|f''\|_\infty}{2} + \big|M_f^n(t)\big|\\ &\leq |\langle f, \mu_0\rangle| + T + \big|M_f^n(t)\big|, \quad \left(\text{since } |f''(s)| = \left|\frac{2(1-s^2)}{(s^2+1)^2}\right| \leq 2, \text{ uniformly bounded}\right).\end{aligned}$$

Thus, we obtain

$$\sup_{0\leq t\leq T}|\langle f,\mu_t^n\rangle| \;\leq\; |\langle f,\mu_0\rangle| + T + \sup_{0\leq t\leq T}\left|M_f^n(t)\right|, \tag{2.29}$$

with the martingale M_f^n $\left(\text{since } f'(x)=\dfrac{2x}{x^2+1}\right)$

$$M_f^n(t)=\frac{1}{n\sqrt{n}}\sum_{j=1}^n\int_0^t f'\big(\lambda_j^n(s)\big)d\beta_s^j=\frac{2}{n\sqrt{n}}\sum_{j=1}^n\int_0^t\frac{\lambda_j^n(s)}{\lambda_j^n(s)^2+1}d\beta_s^j. \tag{2.30}$$

We have a bracket for $M_f^n(t)$ as

$$\langle M_f^n\rangle_t=\frac{4}{n^3}\sum_{j=1}^n\int_0^t\frac{\lambda_j^n(s)^2}{\left(\lambda_j^n(s)^2+1\right)^2}ds\leq\frac{4}{n^3}\sum_{j=1}^n\int_0^t ds=\frac{4t}{n^2}, \tag{2.31}$$

so

$$\mathbb{P}\left(\sup_{0\leq t\leq T}\left|M_f^n(t)\right|\geq L\right)\leq 2.e^{-\frac{n^2.L^2}{8T}}; \tag{2.32}$$

Therefore, for $M>C_0+T$, we have

$$\begin{aligned}\mathbb{P}\left(\sup_{0\leq t\leq T}|\langle f,\mu_t^n\rangle|\geq M\right)&\leq\mathbb{P}\left(|\langle f,\mu_0^n\rangle|+T+\sup_{0\leq t\leq T}\left|M_f^n(t)\right|\geq M\right)\\&\leq\mathbb{P}\left(\sup_{0\leq t\leq T}\left|M_f^n(t)\right|\geq M-C_0-T\right),\end{aligned}$$

and we get

$$\mathbb{P}\left(\sup_{0\leq t\leq T}\left|\int\log(x^2+1)\mu_t^n(dx)\right|\geq M\right)\leq 2.e^{-\frac{n^2.(M-C_0-T)^2}{8T}}; \tag{2.33}$$

2.) We next need an estimate on the Hölder norm of the function $t\to\langle f,\mu_t^n\rangle$, for any $f\in\mathcal{C}_b^2(\mathbb{R})$. We proceed similarly by first cutting $[0,T]$ into intervals of length $\delta>0$.
Let $t_j=j.\delta$, for $0\leq j\leq[T/\delta]+1$, with denote by $[T/\delta]=\sup\{k\in\mathbb{N}:k\leq T/\delta\}$, we have

$$\left\{\sup_{s,t\in[0,T]:|t-s|\leq\delta}|\langle f,\mu_t^n\rangle-\langle f,\mu_s^n\rangle|\geq M\delta^{1/4}\right\}\subset\bigcup_{j=0}^{[T/\delta]}\left\{\sup_{t_j\leq s\leq t_{j+1}}\left|\langle f,\mu_s^n\rangle-\langle f,\mu_{t_j}^n\rangle\right|\geq\frac{M\delta^{1/4}}{2}\right\}.$$

On the other hand, for $s\in[t_j,t_{j+1}]$, using (2.23), one deduces that

$$\begin{aligned}\left|\langle f,\mu_s^n\rangle-\langle f,\mu_{t_j}^n\rangle\right|&\leq\left|\int_{t_j}^s\iint\frac{f'(x)}{x-y}\mu_w^n(dy)\mu_w^n(dx)dw\right|+\left|M_f^n(s)-M_f^n(t_j)\right|\\&\leq|s-t_j|.\frac{1}{2}.\left\|f''\right\|_\infty+\left|M_f^n(s)-M_f^n(t_j)\right|\\&\leq\delta.\frac{\|f''\|_\infty}{2}+\left|M_f^n(s)-M_f^n(t_j)\right|\end{aligned}$$

with bracket

$$\langle M_f^n\rangle_{s-t_j}=\frac{4}{n^3}\sum_{k=1}^n\int_{t_j}^s\left[f'\big(\lambda_k^n(u)\big)\right]^2du\leq\frac{4|s-t_j|.\|f'\|_\infty^2}{n^2}\leq\frac{4\delta\|f'\|_\infty^2}{n^2}.$$

Thus, for any $\gamma > 0$, one gets

$$\mathbb{P}\left(\sup_{t_j \le s \le t_{j+1}} \left|\langle f, \mu_s^n\rangle - \langle f, \mu_{t_j}^n\rangle\right| \le \delta.\frac{\|f''\|_\infty}{2} + \gamma\right) \ge 1 - 2.e^{-\frac{n^2.\gamma^2}{8\delta\|f'\|_\infty^2}}. \tag{2.34}$$

As a conclusion, choose $\gamma = \dfrac{M\delta^{1/4}}{2} - \delta.\dfrac{\|f''\|_\infty}{2} \ge \dfrac{M\delta^{1/4}}{4}$, $\left(\text{since } \|f''\|_\infty \le 2^{-1}.M.\delta^{-3/4}\right)$, we have proved that

$$\begin{aligned}
\mathbb{P}\left(\sup_{s,t\in[0,T]:|t-s|\le\delta} |\langle f, \mu_t^n\rangle - \langle f, \mu_s^n\rangle| \ge M\delta^{1/4}\right) &\le \sum_{j=0}^{[T/\delta]} \mathbb{P}\left(\sup_{t_j \le s \le t_{j+1}} \left|\langle f, \mu_s^n\rangle - \langle f, \mu_{t_j}^n\rangle\right| \ge \frac{M\delta^{1/4}}{2}\right) \\
&\le \sum_{j=0}^{[T/\delta]} 2.e^{-\frac{n^2}{8\delta\|f'\|_\infty^2}.\frac{M^2.\delta^{1/2}}{4^2}} \\
&\le 2.\left(\left[\frac{T}{\delta}\right] + 1\right).e^{-\frac{n^2.M^2}{2^7.\delta^{1/2}.\|f'\|_\infty^2}} \quad ; \qquad (2.35)
\end{aligned}$$

3.) To complete the proof of Lemma 2.3.7, we notice that

- The following set is compact:

$$K_M = \left\{\mu \in \mathcal{M}_1(\mathbb{R}) : \int \log(x^2+1)\mu(dx) \le M\right\} \tag{2.36}$$

  and then by Borel-Cantelli Lemma along with upper bound (2.33), we deduce that

$$\mathbb{P}\left(\{\mu_t^n \in K_M, \forall t \in [0,T]\}^c\right) \le e^{-a(T).M.n^2}, \quad \text{with } a(T) > 0,$$

  so

$$\mathbb{P}\left(\bigcup_{n_0\ge 0}\bigcap_{n\ge n_0} \{\mu_t^n \in K_M, \forall t \in [0,T]\}\right) = 1. \tag{2.37}$$

- Next, we recall that by the Arzela-Ascoli Theorem, sets of the form are compact

$$C = \bigcap_m \left\{g \in \mathcal{C}([0,T],\mathbb{R}) : \sup_{s,t\in[0,T]:|t-s|\le\eta_m} |g(t)-g(s)| \le \varepsilon_m, \sup_{t\in[0,T]} |g(t)| \le M\right\},$$

  where $(\varepsilon_m)_{m\ge0}$ and $(\eta_m)_{m\ge0}$ are sequences of positive numbers decreasing to 0 as $m \to \infty$. Thanks to this remark, now for $f \in \mathcal{C}_b^2(\mathbb{R})$ and $\varepsilon > 0$, we consider the compact subset of $\mathcal{C}\left([0,T],\mathcal{M}_1(\mathbb{R})\right)$ as

$$C_T(f,\varepsilon) := \bigcap_{m=1}^{\infty} \left\{\mu_\cdot \in \mathcal{C}\left([0,T],\mathcal{M}_1(\mathbb{R})\right) : \sup_{|t-s|\le m^{-2}} |\mu_t(f) - \mu_s(f)| \le \frac{1}{\varepsilon.m^{1/2}}\right\}. \tag{2.38}$$

Then, the earlier bound (2.35) implies that

$$\begin{aligned}
\mathbb{P}\left(\mu_\cdot^n \in C_T(f,\varepsilon)^c\right) &\leq \sum_{m=1}^{+\infty} \mathbb{P}\left(\left\{\mu_\cdot^n \in \mathcal{C}\left([0,T],\mathcal{M}_1(\mathbb{R})\right) : \sup_{|t-s|\leq m^{-2}} |\mu_t^n(f)-\mu_s^n(f)| \geq \frac{1}{\varepsilon.m^{1/2}}\right\}\right) \\
&\leq \sum_{m=1}^{+\infty} \left[\frac{T}{\delta}\right].\left(\left[\frac{T}{\delta}\right]+1\right).e^{-\frac{n^2M^2}{2^7.\delta^{1/2}.\|f'\|_\infty^2}} \\
&\quad \text{(using (2.35) upper bound for time-difference)} \\
&\leq \sum_{m=1}^{+\infty} T.m^2.\left(T.m^2+1\right).e^{-\frac{n^2.m}{2^7.\varepsilon^2.\|f'\|_\infty^2}} \quad \text{(as for } M=\varepsilon^{-1}\text{, and } \delta=m^{-2}\text{)} \\
&\leq \gamma(T).e^{-\frac{n^2}{2^7.\varepsilon^2.\|f'\|_\infty^2}} \qquad (2.39)
\end{aligned}$$

with some constant $\gamma(T)>0$, only depending on T.
Choosing a countable family f_k of twice continuously differentiable functions dense in $\mathcal{C}_0(\mathbb{R})$, and set $\varepsilon_k = \dfrac{1}{k\left(\|f_k\|_\infty + \left\|f_k'\right\|_\infty + \left\|f_k''\right\|_\infty\right)^{\frac{1}{2}}} < \dfrac{1}{2}$, we obtain

$$\mathbb{P}\left(\mu_\cdot^n \in \left(\bigcap_{k\geq 1} C_T(f_k,\varepsilon_k)\right)^c\right) \leq \gamma(T).\sum_{k\geq 1} e^{-\frac{n^2}{2^7.\varepsilon_k^2.\left\|f_k'\right\|_\infty^2}} \leq \widetilde{\gamma}(T).e^{-\frac{n^2}{2^7}}$$

with some constant $\widetilde{\gamma}(T)>0$, only depending on T.

- Therefore, we conclude that the compact set

$$\mathcal{K}(M) = K_M \cap \bigcap_{k\geq 1} C_T(f_k,\varepsilon_k) \quad , \qquad (2.40)$$

and $\mathcal{K}(M) \subset \mathcal{C}\left([0,T],\mathcal{M}_1(\mathbb{R})\right)$, is such that

$$\mathbb{P}\left(\mu_\cdot^n \in \mathcal{K}(M)^c\right) \leq \mathbb{P}\left(\mu_\cdot^n \in \left(\bigcap_{k\geq 1} C_T(f_k,\varepsilon_k)\right)^c\right) \leq \widetilde{\gamma}(T).e^{-\frac{n^2}{2^7}}, \qquad (2.41)$$

and thus, by Borel-Cantelli Lemma we get

$$\mathbb{P}\left(\bigcup_{n_0\geq 0}\bigcap_{n\geq n_0}\left\{\mu_\cdot^n \in \mathcal{K}(M)\right\}\right) = 1. \qquad (2.42)$$

This implies that the sequence $(\mu_\cdot^n)$ is almost surely contained in $\mathcal{K}(M)$, and thus is almost surely pre-compact in $\mathcal{C}\left([0,T],\mathcal{M}_1(\mathbb{R})\right)$, which concludes the proof of Lemma 2.3.7.

□

This completes the proof of the first main part in Proposition 2.3.5, which states the convergence of empirical measure $(\mu_t^n, t\in[0,T])$ in an appropriate sense.

■ **Step 3:**

The next step in proof of Proposition 2.3.5 is to show the limit point of $(\mu_t^n, t \in [0,T])$ coincides with the solution of equation (2.26). Indeed, to characterize the limit point of $\mu_\cdot^n$, we rely on the Itô's calculus. Applying (2.23)) we have

$$\begin{aligned}\int_{\mathbb{R}} f(x)\mu_t^n(dx) &= \int_{\mathbb{R}} f(x)\mu_0^n(dx) + \int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{f'(x)}{x-y}\mu_s^n(dy)\mu_s^n(dx)\right) ds + M_f^n(t) \\ &= \int_{\mathbb{R}} f(x)\mu_0^n(dx) + \frac{1}{2}.\int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{f'(x)-f'(y)}{x-y}\mu_s^n(dy)\mu_s^n(dx)\right) ds + M_f^n(t)\end{aligned}$$

with M_f^n is the martingale given for $t \leq T$ by

$$M_f^n(t) = \frac{1}{n\sqrt{n}}\sum_{j=1}^n \int_0^t f'(\lambda_j^n) d\beta_s^j.$$

For a specific test function $f(x) = \dfrac{1}{z-x}$ with $z \in \mathbb{C}^+$, we can rewrite the previous equation as

$$\widehat{G}_{\mu_t^n}(z) = \widehat{G}_{\mu_0^n}(z) + \frac{1}{2}.\int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{\dfrac{1}{(z-x)^2} - \dfrac{1}{(z-y)^2}}{(x-y)}\mu_s^n(dy)\mu_s^n(dx)\right) ds + M_f^n(t)$$

$$\begin{aligned}&= \widehat{G}_{\mu_0^n}(z) + \frac{1}{2}.\int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{(z-y)^2-(z-x)^2}{(x-y).(z-x)^2.(z-y)^2}\mu_s^n(dy)\mu_s^n(dx)\right) ds + M_f^n(t) \\ &= \widehat{G}_{\mu_0^n}(z) + \frac{1}{2}.\int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{(x-y)(2z-x-y)}{(x-y).(z-x)^2.(z-y)^2}\mu_s^n(dy)\mu_s^n(dx)\right) ds + M_f^n(t) \\ &= \widehat{G}_{\mu_0^n}(z) + \int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{(z-x).(z-y)^2}\mu_s^n(dy)\mu_s^n(dx)\right) ds + M_f^n(t) \quad ,\end{aligned}$$

with the martingale M_f^n now becomes

$$M_f^n(t) = \frac{1}{n\sqrt{n}}\sum_{j=1}^n \int_0^t \frac{1}{\left(z-\lambda_j^n(s)\right)^2} d\beta_s^j.$$

Then, we have a bracket

$$\langle M_f^n\rangle_t = \frac{1}{n^3}\sum_{j=1}^n \int_0^t \frac{1}{\left(z-\lambda_j^n(s)\right)^4} ds \leq \frac{t}{n^3}.\frac{1}{|\mathrm{Im}(z)|^4}\sum_{j=1}^n \int_0^t ds \leq \frac{t}{n^2}.\frac{1}{|\mathrm{Im}(z)|^4};$$

Again, by Borel-Cantelli Lemma, the martingale M_f^n goes almost surely to 0. Thus, since $\left|\dfrac{1}{(z-x).(z-y)^2}\right| \leq \dfrac{1}{|\mathrm{Im}(z)|^3}$ for any $x, y \in \mathbb{R}$, and by the almost sure convergence of μ_t^n to μ_t, for $z \in \mathbb{C}\backslash\mathbb{R}$, we can attain from the previous equation that

$$\widehat{G}_{\mu_t^n}(z) = \widehat{G}_{\mu_0^n}(z) + \int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{(z-x).(z-y)^2}\mu_s^n(dy)\mu_s^n(dx)\right) ds + M_f^n(t) \tag{2.43}$$

$$\xrightarrow{n\to\infty} G_{\mu_0}(z) + \int_0^t \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{(z-x).(z-y)^2}\mu_s(dy)\mu_s(dx)\right) ds; \tag{2.44}$$

Moreover, G_{μ_t} is the solution of Burger equation

$$\begin{aligned} G_{\mu_t}(z) &= G_{\mu_0}(z) - \int_0^t G_{\mu_s}(z).\partial_z G_{\mu_s}(z) ds \\ &= G_{\mu_0}(z) + \int_0^t \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{(z-x).(z-y)^2} \mu_s(dy)\mu_s(dx) \right) ds, \quad z \in \mathbb{C}^+. \end{aligned}$$

Therefore, we can conclude that $\widehat{G}_{\mu_t^n}(w)$ converges almost surely to $G_{\mu_t}(w)$ as $n \to \infty$, for any w in $\mathbb{C}^+$.

Finally, the uniqueness of μ_t is guaranteed through its Cauchy transform G_{μ_t}. Since the support of any limit point μ_t lives in $\mathbb{R}$ for all t, this uniqueness implies the uniqueness of the Cauchy Transform of μ_t for all t and hence, by Theorem 2.2.4 (The Stieltjes Inversion Formula), we get the uniqueness of μ_t for all t.

More precisely, the following lemma (see also [1, Lemma 4.3.15]) shows the uniqueness of the solution of equation (2.26).

Lemma 2.3.8. *Let $D_{\alpha,\beta} = \{z = u + i.v \in \mathbb{C}^+ : |u| < \alpha.v, \quad v > \beta\}$, and for $t \geq 0$, set $\Lambda_t := \{z \in \mathbb{C}^+ : z + t.G_{\mu_0}(z) \in \mathbb{C}^+\}$. For all $t \geq 0$, there exist positive constants $\alpha_t, \beta_t, \alpha'_t, \beta'_t$ such that $D_{\alpha_t,\beta_t} \subset \Lambda_t$ and the function $z \in D_{\alpha_t,\beta_t} \to z + t.G_{\mu_0}(z) \in D_{\alpha'_t,\beta'_t}$ is invertible with inverse $F_t : D_{\alpha'_t,\beta'_t} \to D_{\alpha_t,\beta_t}$. Any solution of (2.25), (2.26) is the unique analytic function on $\mathbb{C}^+$ such that for all t and all $z \in D_{\alpha'_t,\beta'_t}$*

$$G_{\mu_t}(z) = G_{\mu_0}\big(F_t(z)\big).$$

Proof of Lemma 2.3.8. Since $|G_{\mu_0}(z)| \leq \frac{1}{|\mathrm{Im}(z)|}$, we have $\mathrm{Im}\big(z + t.G_{\mu_0}(z)\big) \geq \mathrm{Im}(z) - \frac{t}{\mathrm{Im}(z)} > 0$, for $t < (\mathrm{Im}(z))^2$ and $\mathrm{Im}(z) > 0$. Thus, $D_{\alpha_t,\beta_t} \subset \Lambda_t$ for $t < \beta_t^2$. Moreover, $\mathrm{Re}\big(G_{\mu_0}(z)\big) \leq \frac{1}{2|\mathrm{Im}(z)|}$, so we can observe that for all $t \geq 0$, the image of D_{α_t,β_t} by $z + t.G_{\mu_0}(z)$ is contained in some $D_{\alpha'_t,\beta'_t}$, provided β_t is large enough, (see e.g. [7, Lemma 1] and [5, Section 5]).
In addition, we can choose the D_{α_t,β_t} and $D_{\alpha'_t,\beta'_t}$ decreasing in time.
We next use the method of characteristic. Fix $G_{\mu_\cdot}$ a solution of (2.25), (2.26). We associate with $z \in \mathbb{C}^+$ the solution $\{z_t, t \geq 0\}$ of the equation

$$\partial_t z_t = G_{\mu_t}(z_t), \quad z_0 = z; \tag{2.45}$$

We can construct a solution $z_\cdot$ to this equation up to time $\frac{(\mathrm{Im}(z))^2}{4}$ with $\mathrm{Im}(z_t) \geq \frac{\mathrm{Im}(z)}{2}$ as follows. We put for $\varepsilon > 0$,

$$G^\varepsilon_{\mu_t}(z) := \int \frac{\overline{z} - x}{|z-x|^2 + \varepsilon} d\mu_t(x), \quad \partial_t z^\varepsilon_t = G^\varepsilon_{\mu_t}(z^\varepsilon_t), \quad z^\varepsilon_0 = z;$$

We have $z^\varepsilon_\cdot$ exists and is unique since $G^\varepsilon_{\mu_t}$ is uniformly Lipschitz. Moreover,

$$\partial_t \mathrm{Im}(z^\varepsilon_t) = \mathrm{Im}\big(G^\varepsilon_{\mu_t}(z^\varepsilon_t)\big) = \mathrm{Im}\left(\int \frac{\overline{z^\varepsilon_t} - x}{|z^\varepsilon_t - x|^2 + \varepsilon} d\mu_t(x) \right) = -\mathrm{Im}(z^\varepsilon_t). \int \frac{1}{|z^\varepsilon_t - x|^2 + \varepsilon} d\mu_t(x),$$

hence,

$$\frac{\partial_t \mathrm{Im}(z^\varepsilon_t)}{\mathrm{Im}(z^\varepsilon_t)} = -\int \frac{1}{|z^\varepsilon_t - x|^2 + \varepsilon} d\mu_t(x) \in \left[-\frac{1}{|\mathrm{Im}(z^\varepsilon_t)|^2}, 0 \right].$$

Moreover, we also have

$$\partial_t\big((\mathrm{Im}(z_t^\varepsilon))^2\big) = 2.\mathrm{Im}(z_t^\varepsilon).\partial_t\mathrm{Im}(z_t^\varepsilon) = 2.|\mathrm{Im}(z_t^\varepsilon)|^2.\frac{\partial_t\mathrm{Im}(z_t^\varepsilon)}{\mathrm{Im}(z_t^\varepsilon)} \in [-2,0],$$

thus, $|\mathrm{Im}(z_0^\varepsilon)|^2 - 2.t \le |\mathrm{Im}(z_t^\varepsilon)|^2 \le |\mathrm{Im}(z_0^\varepsilon)|^2$. This implies that $|\mathrm{Im}(z_t^\varepsilon)|^2 \in \big[|\mathrm{Im}(z)|^2 - 2t, |\mathrm{Im}(z)|^2\big]$ as $z_0^\varepsilon = z$. On the other hand, for $t \le \dfrac{(\mathrm{Im}(z))^2}{4}$

$$\big|\partial_t\mathrm{Re}(z_t^\varepsilon)\big| = \left|\int \frac{\mathrm{Re}(z_t^\varepsilon) - x}{|z_t^\varepsilon - x|^2 + \varepsilon} d\mu_t(x)\right| \le \frac{1}{\sqrt{|\mathrm{Im}(z_t^\varepsilon)|^2 + \varepsilon}} \le \frac{1}{\sqrt{|\mathrm{Im}(z)|^2 - 2t}},$$

shows that $\mathrm{Re}(z_t^\varepsilon)$ stays uniformly bounded, independently of ε, up to time $\dfrac{(\mathrm{Im}(z))^2}{4}$ as well as its time derivative. Hence, $\left\{z_t^\varepsilon, t \le \dfrac{(\mathrm{Im}(z))^2}{4}\right\}$ is tight by Arzela-Ascoli's Theorem. Any limit point is a solution of the original equation and such that $\mathrm{Im}(z_t) \ge \dfrac{\mathrm{Im}(z)}{2} > 0$. It is unique since G_{μ_t} is uniformly Lipschitz on this domain (the argument showing that G_{μ_t} is a Lipschitz continuous function will be given later in Lemma 2.6.2).
Now, $\partial_t G_{\mu_t}(z_t) = 0$ implies that for $t \le \dfrac{(\mathrm{Im}(z))^2}{4}$,

$$\partial_t z_t = G_{\mu_t}(z_t) = G_{\mu_0}(z_0), \quad \text{so} \quad z_t = t.G_{\mu_0}(z_0) + z_0 = t.G_{\mu_0}(z) + z$$

and thus,

$$G_{\mu_t}\big(z + t.G_{\mu_0}(z)\big) = G_{\mu_t}(z_t) = G_{\mu_0}(z).$$

By the implicit function theorem, $z + t.G_{\mu_0}(z)$ is invertible form D_{α_t,β_t} into $D_{\alpha_t',\beta_t'}$ since $1 + t.G_{\mu_0}'(z) \neq 0$ (note that $\mathrm{Im}(G_{\mu_0}'(z)) \neq 0$) on D_{α_t,β_t}. Its inverse F_t is analytic from $D_{\alpha_t',\beta_t'}$ into D_{α_t,β_t}, and satisfies

$$G_{\mu_t}(z) = G_{\mu_0}\big(F_t(z)\big).$$

□

■ Step 4:

The limit point of $(\mu_t^n, t \in [0,T])$ coincides with the solution of the Fokker-Planck equation. In addition, this solution can be indeed presented as the free convolution of the initial condition μ_0 and the semi-circular law σ_t defined in (2.3). Let ν be a probability measure on $\mathbb{R}$ such that $\nu = \mu_0 \boxplus \sigma_t$. By definition of free convolution and using the fact that $R_{\sigma_t}(w) = t.w$ with w in some domain $\Gamma_{\gamma(\sigma_t),\delta(\sigma_t)}$, we have

$$R_\nu(w) = R_{\mu_0}(w) + R_{\sigma_t}(w) = R_{\mu_0}(w) + t.w, \quad w \in \Gamma_{\gamma(\nu),\delta(\nu)},$$

or equivalently,

$$G_\nu^{<-1>}(w) = G_{\mu_0}^{<-1>}(w) + t.w, \quad w \in \Gamma_{\gamma(\nu),\delta(\nu)};$$

This leads to the fact that for z in some domain D_{α_t,β_t}:

$$G_{\mu_0}(z) = G_\nu\big(z + t.G_{\mu_0}(z)\big), \tag{2.46}$$

and hence, using the inverse function F_t defined in Lemma 2.3.8, we obtain

$$G_{\mu_0}\big(F_t(\xi)\big) = G_\nu(\xi), \quad \xi \in D_{\alpha_t',\beta_t'}; \tag{2.47}$$

Then, the analyticity and the uniqueness of G_{μ_t} along with the uniqueness of F_t implies that $G_v(\xi) = G_{\mu_t}(\xi)$ for all $\xi \in D_{\alpha'_t,\beta'_t}$. Furthermore, since the Stieltjes-Inversion-Formula characterizes a probability measure on $\mathbb{R}$ from its Cauchy transform, we eventually get $\mu_t = \nu$, which is the free convolution measure $\mu_0 \boxplus \sigma_t$.

This concludes the proof of Proposition 2.3.5.

□

2.4 Free deconvolution for a semicircular distribution by subordinations method

Remind that for a fixed time $t > 0$, considering μ_t to be the solution of Fokker-Planck equation, we have $\mu_t = \mu_0 \boxplus \sigma_t$. As a result, the problem of recovering μ_0 from knowledge on μ_t is a free deconvolution problem. Then, the idea is to apply the subornation method introduced (in general setting) in Section 2.2 to our framework where one of the measure involved in free convolution is the semi-circular distribution.
We first introduce a sub-domain in $\mathbb{C}^+$ that is useful in the sequel of this chapter.

Definition 2.4.1. For any $\gamma > 0$, we define a truncated sub-domain $\mathbb{C}_\gamma$ as follows:

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

These domains will appear due to the fact that G_{μ_t} is not invertible on the whole upper half-plane $\mathbb{C}^+$. More specifically, for two probability measures μ_1, μ_2 on $\mathbb{R}$ and $\mu = \mu_1 \boxplus \mu_2$, by definition of free convolution using R-transform (in (2.8)), we have $G_\mu(\mathbb{C}^+) \subset \big(G_{\mu_1}(\mathbb{C}^+) \cap G_{\mu_2}(\mathbb{C}^+)\big)$. This implies the existence of a subordination function $w_2 : \mathbb{C}^+ \to \mathbb{C}^+$ such that $G_\mu = G_{\mu_2} \circ w_2$. In contrast, it prevents us from exploring for a subordination function w defined on the whole $\mathbb{C}^+$ such that $G_{\mu_2} = G_\mu \circ w$. Nevertheless, the following theorem allows us to recover μ_0 (not completely, but there remains just a classical deconvolution with a Cauchy distribution) from the knowledge of σ_t and μ_t.

Theorem 2.4.2. *There exist unique subordination functions w_1 and w_{fp} from $\mathbb{C}_{2\sqrt{t}}$ onto $\mathbb{C}^+$ such that:*

(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

$$\lim_{y \to +\infty} \frac{w_1(iy)}{iy} = \lim_{y \to +\infty} \frac{w_{fp}(iy)}{iy} = 1.$$

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

$$F_{\mu_0}(z) = F_{\sigma_t}(w_1(z)) = F_{\mu_t}(w_{fp}(z)), \quad \text{or} \quad G_{\mu_0}(z) = G_{\sigma_t}(w_1(z)) = G_{\mu_t}(w_{fp}(z)); \tag{2.48}$$

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

$$w_{fp}(z) = z + w_1(z) - F_{\mu_0}(z); \tag{2.49}$$

(iv) Denote $h_{\sigma_t}(w) = w - F_{\sigma_t}(w) = t.G_{\sigma_t}(w)$ (by (2.7)), $\widetilde{h}_{\mu_t}(w) = w + F_{\mu_t}(w)$ on $\mathbb{C}^+$, and define a function $\mathcal{L}_z$ as

$$\begin{aligned} \mathcal{L}_z(w) :&= h_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z \\ &= t.G_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z; \end{aligned} \tag{2.50}$$

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

$$\mathcal{L}_z\big(w_{fp}(z)\big) = w_{fp}(z). \tag{2.51}$$

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

The problem of free deconvolution using subordination methods is studied in [2] in generality. In our framework, we apply for free deconvolution by the semi-circular distribution. On the other hand, a crucial difference between Theorem 2.4.2 and Theorem 2.2.5 lies in the fact that the subordination functions are expressed in terms of $F_{\mu_0 \boxplus \sigma_t}$ and F_{σ_t}, where as in Theorem 2.2.5 it would have been F_{μ_0} and F_{σ_t}. Here the restriction to domain $\mathbb{C}_{2\sqrt{t}}$ comes from the fact that $\mathrm{Im}\big(\widetilde{h}_{\mu_t}(\omega) - z\big)$ appearing in the definition (2.50) of $\mathcal{L}_z$ 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 2.7.3).

Lemma 2.4.3. *For any $z \in \mathbb{C}_{2\sqrt{t}}$,*

$$G_{\mu_0}(z) = G_{\sigma_t}\big(\omega_1(z)\big) = \frac{1}{t}\big(\omega_{fp}(z) - z\big). \tag{2.52}$$

Consequently,

$$|\omega_{fp}(z) - z| \leq \sqrt{t}. \tag{2.53}$$

Proof. Indeed, from (2.49), we rewrite as

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

So, we obtain

$$G_{\sigma_t}\big(\omega_1(z)\big) = \frac{1}{t}\big(\omega_{fp}(z) - z\big).$$

Moreover, one can write

$$\big|\omega_{fp}(z) - z\big| = t.\big|G_{\sigma_t}\big(\omega_1(z)\big)\big| = t.\big|G_{\mu_t}\big(\omega_{fp}(z)\big)\big| \leq \frac{t}{\big|\mathrm{Im}\omega_{fp}(z)\big|} \leq \sqrt{t},$$

in which the last inequality results from Theorem 2.4.2 (i).

□

Notation. For $\gamma > 0$, we denote by $\mathcal{C}_\gamma$ the centered Cauchy distribution with parameter γ, that has a density function

$$f_{\mathcal{C}_\gamma}(x) := \frac{\gamma}{\pi\big(x^2 + \gamma^2\big)}, \quad x \in \mathbb{R}.$$

Actually, we can only recover $G_{\mu_0}\big(s + i.2\sqrt{t}\big)$, which is the Cauchy-transform of classical convolution $\mu_0 * \mathcal{C}_{2\sqrt{t}}$ between μ_0 and a centered Cauchy distribution $\mathcal{C}_{2\sqrt{t}}$. Thus, we can write as

$$G_{\mu_0}\Big(s + i.2\sqrt{t}\Big) = G_{\mu_0 * \mathcal{C}_{2\sqrt{t}}}(s), \quad s \in \mathbb{C}^+; \tag{2.54}$$

- At this point, what can be recovered is $G_{\mu_0 * \mathcal{C}_{2\sqrt{t}}}$;
- Then, eventually there is a deconvolution to recover entirely μ_0.

As a consequence, Stieltjes-inversion-formula allows us to recover the convolution measure $\mu_0 * \mathcal{C}_{2\sqrt{t}}$ from its Cauchy transform. More precisely, for any $x \in \mathbb{R}$ one has the density function

$$\begin{aligned} f_{\mu_0 * \mathcal{C}_{2\sqrt{t}}}(x) = -\lim_{\varepsilon \to 0^+} \frac{1}{\pi} \operatorname{Im} G_{\mu_0 * \mathcal{C}_{2\sqrt{t}}}(x + i.\varepsilon) &= -\lim_{\varepsilon \to 0^+} \frac{1}{\pi} \operatorname{Im} G_{\mu_0}\left(x + i.2\sqrt{t} + i.\varepsilon\right) \\ &= -\lim_{\varepsilon \to (2\sqrt{t})^+} \frac{1}{\pi} \operatorname{Im} G_{\mu_0}(x + i.\varepsilon) \\ \text{(using (2.52))} \quad &= -\lim_{\varepsilon \to (2\sqrt{t})^+} \frac{1}{\pi t} \operatorname{Im}\left[w_{fp}(x + i.\varepsilon) - x - i.\varepsilon\right] \\ &= \frac{1}{\pi t}\left[2\sqrt{t} - \lim_{\varepsilon \to (2\sqrt{t})^+} \operatorname{Im} w_{fp}(x + i.\varepsilon)\right]. \quad (2.55) \end{aligned}$$

2.4.1 Proof of Theorem 2.4.2

The constants of Theorem 2.4.2 are better than the ones of Arizmendi et al. [2] who work in a full generality. We sketch here the main steps of the proof in our context, using the explicit formula for the semi-circular distribution.

In the whole proof, we consider $z \in \mathbb{C}_{2\sqrt{t}}$, i.e. $z \in \mathbb{C}^+$ and $\operatorname{Im}(z) > 2\sqrt{t}$.

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

$$\operatorname{Im}\big(\widetilde{h}_{\mu_t}(w) - z\big) = \operatorname{Im}\big(w + F_{\mu_t}(w) - z\big) \geq 2\operatorname{Im}(w) - \operatorname{Im}(z) > 0, \quad (2.56)$$

where we have used $\operatorname{Im} F_{\mu_t}(w) \geq \operatorname{Im}(w)$ for the first inequality. Indeed, if $w = w_1 + i.w_2$, we have

$$(F_{\mu_t}(w))^{-1} = G_{\mu_t}(w) = \int \frac{d\mu_t(x)}{w_1 + i.w_2 - x} = \int \frac{(w_1 - x)d\mu_t(x)}{(w_1 - x)^2 + w_2^2} - i.w_2 \int \frac{d\mu_t(x)}{(w_1 - x)^2 + w_2^2}$$

and

$$\begin{aligned} \operatorname{Im}(F_{\mu_t}(w)) &= w_2 \times \frac{\int \frac{d\mu_t(x)}{(w_1 - x)^2 + w_2^2}}{\left(\int \frac{(w_1 - x)d\mu_t(x)}{(w_1 - x)^2 + w_2^2}\right)^2 + w_2^2 \left(\int \frac{d\mu_t(x)}{(w_1 - x)^2 + w_2^2}\right)^2} \\ &\geq w_2 \times \frac{\int \frac{d\mu_t(x)}{(w_1 - x)^2 + w_2^2}}{\int \frac{(w_1 - x)^2 d\mu_t(x)}{\left((w_1 - x)^2 + w_2^2\right)^2} + w_2^2 \int \frac{d\mu_t(x)}{\left((w_1 - x)^2 + w_2^2\right)^2}} = w_2. \quad (2.57) \end{aligned}$$

Step 2: We show that $\mathcal{L}_z(\mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)}) \subset \overline{\mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)}}$ and that $\mathcal{L}_z$ is not a conformal automorphism.
+ First, let us show that $\mathcal{L}_z\left(\mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)}\right) \subset \overline{\mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)}}$. Let $w \in \mathbb{C}_{\frac{1}{2}\operatorname{Im}(z)}$, we have:

$$\begin{aligned} \operatorname{Im}\left(\mathcal{L}_z(w)\right) &= \operatorname{Im}\left[t.G_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w) - z\big) + z\right] \\ &= \operatorname{Im}\left(\frac{\widetilde{h}_{\mu_t}(w) - z - \sqrt{\big(\widetilde{h}_{\mu_t}(w) - z\big)^2 - 4t}}{2} + z\right). \quad (2.58) \end{aligned}$$

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

$$\mathrm{Im}\big(\sqrt{v^2-4t}\big) \leq \sqrt{\mathrm{Im}^2(v)+4t}. \tag{2.59}$$

Indeed, write $v = v_1 + i.v_2$ with $v_2 > 0$, and $\sqrt{v^2-4t} = u_1 + i.u_2$, with $u_1, u_2 \in \mathbb{R}$, $u_2 \neq 0$. We have to show that $u_2 \leq \sqrt{v_2^2+4t}$. In case $v_1 = 0$, the inequality is implied directly. For $v_1 \neq 0$, we consider

$$v_1^2 - v_2^2 - 4t + 2v_1v_2.i = u_1^2 - u_2^2 + 2u_1u_2.i,$$

so one gets $u_1.u_2 = v_1.v_2$ and $u_1^2 - u_2^2 = v_1^2 - v_2^2 - 4t$. Thus, using $u_1 = \dfrac{v_1.v_2}{u_2}$, we can write

$$v_1^2.v_2^2 - u_2^4 = v_1^2.u_2^2 - \big(v_2^2+4t\big).u_2^2 \quad \text{, or} \quad \big(v_1^2+u_2^2\big).v_2^2 + 4t.u_2^2 = \big(v_1^2+u_2^2\big).u_2^2,$$

hence, $v_2^2 + 4t = u_2^2 + 4t.\dfrac{v_1^2}{v_1^2+u_2^2} > u_2^2$, this implies that the inequality (2.59) holds.
Therefore, using (2.59), we have:

$$\mathrm{Im}\left(\sqrt{\big(\widetilde{h}_{\mu_t}(w)-z\big)^2 - 4t}\right) \leq \sqrt{\big[\mathrm{Im}\big(\widetilde{h}_{\mu_t}(w)-z\big)\big]^2 + 4t}.$$

Now, (2.58) yields:

$$\mathrm{Im}\big(\mathcal{L}_z(w)\big) \geq \frac{1}{2}\left[\mathrm{Im}\big(\widetilde{h}_{\mu_t}(w)-z\big) - \sqrt{\big[\mathrm{Im}\big(\widetilde{h}_{\mu_t}(w)-z\big)\big]^2+4t}\right] + \mathrm{Im}(z).$$

The function $g(s) = s - \sqrt{s^2+4t}$ is non-decreasing on $\mathbb{R}_+$ and for all $s > 0$, $g(s) \geq -2\sqrt{t}$. This implies that

$$\mathrm{Im}\big(\mathcal{L}_z(w)\big) \geq \mathrm{Im}(z) - \sqrt{t} > \frac{1}{2}\mathrm{Im}(z), \tag{2.60}$$

since $z \in \mathbb{C}_{2\sqrt{t}}$. This guarantees that $\mathcal{L}_z(w) \in \mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$.
+ Let us now prove that $\mathcal{L}_z$ is not an automorphism of $\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$. Consider

$$|\mathcal{L}_z(w) - z| = \Big|F_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w)-z\big) - \big(\widetilde{h}_{\mu_t}(w)-z\big)\Big| = \Big|tG_{\sigma_t}\big(\widetilde{h}_{\mu_t}(w)-z\big)\Big|.$$

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

$$|tG_{\sigma_t}(v)| = t\left|\int_{-2\sqrt{t}}^{2\sqrt{t}} \frac{1}{v-x}d\sigma_t(x)\right| \leq t.\int_{-2\sqrt{t}}^{2\sqrt{t}} \frac{1}{|v-x|}d\sigma_t(x) \leq t.\int_{-2\sqrt{t}}^{2\sqrt{t}} \frac{1}{|v|-|x|}d\sigma_t(x) \leq \sqrt{t}.$$

If $|v| \leq 3\sqrt{t}$,

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

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

$$|\mathcal{L}_z(w) - z| \leq 4\sqrt{t}. \tag{2.61}$$

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

Step 3: Existence and uniqueness of w_{fp}, which is a fixed point of $\mathcal{L}_z$.
By Steps 1 and 2, $\mathcal{L}_z$ satisfies the assumptions of Denjoy-Wolff's fixed-point theorem (see in Appendices and in [4]). The theorem says that for any $w \in \mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$ the iterated sequence $\mathcal{L}_z^{\circ m}(w) = \mathcal{L}_z \circ \mathcal{L}_z^{\circ(m-1)}(w)$ converge to the unique Denjoy-Wolff point of $\mathcal{L}_z$ which we define as $w_{fp}(z)$. The Denjoy-Wolff point is either a fixed-point of $\mathcal{L}_z$ or a point of the boundary of the domain. Let us check that w_{fp} is a fixed point of $\mathcal{L}_z$. For any $z \in \mathbb{C}_{2\sqrt{t}}$, there exists $\gamma > 2$ such that $z \in \mathbb{C}_{\gamma\sqrt{t}}$ and from (2.60), $\mathcal{L}_z(\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}) \subset \mathbb{C}_{(1-\frac{1}{\gamma})\mathrm{Im}(z)}$. Moreover, from (2.61), $\mathcal{L}_z(\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}) \subset B(z, 4\sqrt{t})$. Therefore, $w_{fp}(z) \in \overline{\mathbb{C}_{(1-\frac{1}{\gamma})\mathrm{Im}(z)} \cap B(z, 4\sqrt{t})} \subset \mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$, so that it is necessary a fixed point.

We now define

$$w_1(z) := F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) - z.$$

One can check that

$$\begin{aligned} F_{\sigma_t}(w_1(z)) &= w_1(z) - h_{\sigma_t}(w_1(z)) \\ &= F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) - z - h_{\sigma_t}(F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) - z) \\ &= \tilde{h}_{\mu_t}(w_{fp}(z)) - z - h_{\sigma_t}(\tilde{h}_{\mu_t}(w_{fp}(z)) - z) \\ &= \tilde{h}_{\mu_t}(w_{fp}(z)) - w_{fp}(z) = F_{\mu_t}(w_{fp}(z)). \end{aligned}$$

One can therefore rewrite

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

From (2.60) and the fact that $w_{fp}(z)$ is a fixed point of $\mathcal{L}_z$, one easily gets that

$$\lim_{y \to +\infty} \frac{w_{fp}(i.y)}{i.y} = 1,$$

and

$$\lim_{y \to +\infty} \frac{w_1(i.y)}{i.y} = 1.$$

Now we connect F_{μ_0} to the previous quantities. For z large enough, all the functions we consider are invertible and we have

$$F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) = z + w_1(z) = z + F_{\sigma_t}^{<-1>}(F_{\mu_t}(w_{fp}(z))).$$

On the other hand, for z large enough, using Theorem-definition 2.2.5 for $\mu_1 = \sigma_t$ and $\mu_2 = \mu_0$, we get

$$F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) = v_1(w_{fp}(z)) + v_2(w_{fp}(z)) = F_{\sigma_t}^{<-1>}(F_{\mu_t}(w_{fp}(z))) + F_{\mu_0}^{<-1>}(F_{\mu_t}(w_{fp}(z))).$$

Comparing the two equalities gives

$$F_{\mu_0}^{<-1>}(F_{\mu_t}(w_{fp}(z))) = z,$$

so that, for z large enough,

$$F_{\mu_t}(w_{fp}(z)) = F_{\mu_0}(z).$$

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

Finally, since

$$w_1(z) = F_{\mu_t}(w_{fp}(z)) + w_{fp}(z) - z = F_{\mu_0}(z) + w_{fp}(z) - z,$$

we have, using (2.57) with μ_0 instead of μ_t,

$$\mathrm{Im}(w_1(z)) = \mathrm{Im}(F_{\mu_0}(z)) + \mathrm{Im}(w_{fp}(z)) - \mathrm{Im}(z) \geq \mathrm{Im}(w_{fp}(z)) \geq \frac{1}{2}\mathrm{Im}(z).$$

This ends the proof of Theorem 2.4.2.

2.5 Statistical Estimation

Assume that the initial condition μ_0 possesses a density function p_0, i.e. $\mu_0(dx) = p_0(x)dx$. Then from the Cauchy transform G_{μ_0}, by the Stieltjes inversion formula we can recover p_0. Since we do not know the measure μ_t completely, but the observations, iteration operator of fixed-point equation $\mathcal{L}_z$ is supposed to be estimated by some $\widehat{\mathcal{L}}_z$ constructed from these observations.

2.5.1 Construction of the estimator of p_0

Recall that our observation scheme is matrix $X_n(t)$ at time $t > 0$ for given n (defined by (2.17):

$$X_n(t) = X_n(0) + H_n(t),$$

where $X_n(0)$ is random matrix as in Assumptions 2.3.2, and $H_n(t)$ is a Hermitian Brownian motion. Hence, $X_n(t)$ allows to construct the empirical spectral measure of $\lambda_j^n(t)$ defined in (2.19) as:

$$\mu_t^n = \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j^n(t)}.$$

Then, for $z \in \mathbb{C}^+$ a natural estimator of $G_{\mu_t}(z)$ is established as follows:

$$\widehat{G}_{\mu_t^n}(z) = \int_{\mathbb{R}} \frac{d\mu_t^n(\lambda)}{z-\lambda} = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{z-\lambda_j^n(t)} = \frac{1}{n}\mathrm{tr}\Big(\big(zI_n - X_n(t)\big)^{-1}\Big). \tag{2.62}$$

This provides an estimate of $\mathcal{L}_z$, iterate operator of fixed-point equation. The next step is to estimate subordination function w_{fp}. More specifically, for $z \in \mathbb{C}_{2\sqrt{t}}$, let us define an estimator of $w_{fp}(z)$ as solution of the "empirical" fixed point equation counterpart

$$\widehat{\mathcal{L}}_z(w(z)) =: w(z).$$

In fact, as $\mathcal{L}_z$ is a bit tricky to manipulate, we shall define our estimator via the equivalent empirical fixed-point equation as in the theorem below.

Theorem 2.5.1. *There exists a unique solution to the following functional equation in $\omega(z)$, for any $z \in \mathbb{C}_{2\sqrt{t}}$:*

$$\frac{1}{t}(w(z) - z) = \widehat{G}_{\mu_t^n}(w(z)). \tag{2.63}$$

This fixed-point is denoted by $\widehat{w}_{fp}^n$. Moreover, $|\widehat{w}_{fp}^n(z) - z| \leq \sqrt{t}$.

Notation. In the sequel, remind that we denote the Fourier transform of $g \in L^1(\mathbb{R})$ by $g^\star$.

Lemma 2.5.2. *For $\gamma > 0$, Fourier transform of the Cauchy distribution $\mathcal{C}_\gamma$ with density function $f_{\mathcal{C}_\gamma}$ is $f^\star_{\mathcal{C}_\gamma}(z) = e^{-\gamma|z|}$, $z \in \mathbb{C}$.*

To remind, we have already obtained in (2.55) the density of $\mu_0 * \mathcal{C}_{2\sqrt{t}}$ by the following formula:

$$f_{\mu_0 * \mathcal{C}_{2\sqrt{t}}}(x) = \frac{1}{\pi t}\left[2\sqrt{t} - \lim_{\varepsilon \to (2\sqrt{t})^+} \mathrm{Im}\, w_{fp}(x + i.\varepsilon)\right];$$

Performing the deconvolution from above expression, the Fourier transform of p_0 is the division of the Fourier transform of the right-hand side of (2.55) divided by $f^\star_{\mathcal{C}_\gamma}$ with $\gamma > 2\sqrt{t}$. Then, it is classical to define our ultimate estimator for density function from its Fourier transform, by using Kernel density estimate method.

Definition 2.5.3. *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:*

$$\widehat{p}^{\star}_{0,h}(z) = e^{\gamma.|z|}.K^{\star}_h(z).\frac{1}{\pi t}\left[\gamma - \lim_{\varepsilon\to\gamma^+}\left(Im\widehat{w}^n_{fp}(\cdot + i.\varepsilon)\right)^{\star}(z)\right], \tag{2.64}$$

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

The choice of a kernel with compactly supported in the Fourier domain to ensure the finiteness of estimator. In the sequel, we will choose a specific kernel, for instance the *sinc* kernel $K(x) = sinc(x) = \frac{\sin(x)}{\pi x}$, whose Fourier transform is $K^\star(\xi) = \mathbb{1}_{\{|\xi|\leq 1\}}(\xi)$.

2.5.2 Proof of Theorem 2.5.1

The proof of this theorem follows the steps of the proof of Theorem 2.4.2. First, $\widehat{\mathcal{L}}_z(w) := t\widehat{G}_{\mu^n_t}(w) + z$ is a well-defined and analytic function on $\mathbb{C}^+$. Let us check that $\widehat{\mathcal{L}}_z\left(\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}\right) \subset \mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$ for $z \in \mathbb{C}_{2\sqrt{t}}$. For $w = u + iv \in \mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$,

$$\mathrm{Im}\left(\widehat{G}_{\mu^n_t}(w)\right) = \frac{1}{n}\sum_{j=1}^{n}\mathrm{Im}\left(\frac{u - \lambda^n_j(t) - iv}{(u-\lambda^n_j(t))^2 + v^2}\right) > -\frac{1}{v} = -\frac{1}{\mathrm{Im}(w)}. \tag{2.65}$$

Thus,

$$\mathrm{Im}\left(\widehat{\mathcal{L}}_z(w)\right) = t.\mathrm{Im}\left(\widehat{G}_{\mu^n_t}(w)\right) + \mathrm{Im}(z) > -\frac{t}{\mathrm{Im}(w)} + \mathrm{Im}(z) > -\frac{2t}{\mathrm{Im}(z)} + \mathrm{Im}(z) > \frac{1}{2}\mathrm{Im}(z).$$

The second inequality comes from the choice of $w \in \mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$, and the last inequality is a consequence of $\mathrm{Im}(z) > 2\sqrt{t}$.
Moreover, $\widehat{\mathcal{L}}_z$ is not an automorphism since:

$$\left|\widehat{\mathcal{L}}_z(w) - z\right| = \left|t.\widehat{G}_{\mu^n_t}(w)\right| = \left|\frac{1}{n}\sum_{j=1}^{n}\frac{t}{w - \lambda^n_j(t)}\right| \leq \frac{t}{\mathrm{Im}(w)} \leq \sqrt{t}, \tag{2.66}$$

since $\mathrm{Im}(w) > \frac{1}{2}\mathrm{Im}(z) > \sqrt{t}$. We use again the Denjoy-Wolff fixed-point theorem (see in Appendices and in [4]). Because the inclusion of $\widehat{\mathcal{L}}_z\left(\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}\right)$ into $\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$ is strict, the unique Denjoy-Wolff point of $\widehat{\mathcal{L}}_z$ is necessarily a fixed-point that we denote $\widehat{w}_{fp}(z)$. The last announced estimate is a straightforward consequence of (2.66).

2.6 Estimation of the subordination function

In this section, we not only study the consistency of $\widehat{w}^n_{fp}$, in order to see how this estimator converge to the exact fixed-point w_{fp}, but we also examine the rate of this convergence.

Proposition 2.6.1. *Let $\gamma > 2\sqrt{t}$. Suppose p_0 satisfies the condition*

$$\int_{\mathbb{R}} \log(x^2 + 1)p_0(x)dx < +\infty. \tag{2.67}$$

Then, we have:
(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 $\mathbb{C}_\gamma$*:*

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

With Proposition 2.6.1 at hand, it is possible to recover an estimator of G_{μ_0}

2.6.1 Proof of (i) and (ii) of Proposition 2.6.1

We state a useful lemma.

Lemma 2.6.2. *For any probability measure* μ *on* $\mathbb{R}$ *and* $\gamma>0$*, the Cauchy transform* G_μ *is Lipschitz on* $\mathbb{C}_\gamma$ *with Lipschitz constant* $\frac{1}{\gamma^2}$*, and one has for any* $z\in\mathbb{C}_\gamma$*,* $|G_\mu(z)|\leq\frac{1}{\gamma}$.

Proof. For $z,s\in\mathbb{C}_\gamma$, i.e. we have $\mathrm{Im}(z),\mathrm{Im}(s)\geq\gamma$, consider

$$|G_\mu(z)-G_\mu(s)|=\left|\int_{\mathbb{R}}\frac{d\mu(x)}{z-x}-\int_{\mathbb{R}}\frac{d\mu(y)}{s-y}\right|\leq|z-s|\,.\int_{\mathbb{R}}\frac{d\mu(x)}{|(z-x)(s-x)|}\leq\frac{|z-s|}{\mathrm{Im}(z).\mathrm{Im}(s)}\leq\frac{|z-s|}{\gamma^2},$$

this implies the Lipschitz property of G_μ.
Similarly, we have

$$|G_\mu(z)|=\left|\int_{\mathbb{R}}\frac{d\mu(x)}{z-x}\right|\leq\int_{\mathbb{R}}\frac{d\mu(x)}{|z-x|}\leq\frac{1}{\mathrm{Im}(z)}\leq\frac{1}{\gamma}.$$

□

We now prove the points (i) and (ii) of Proposition 2.6.1.

Proof of Proposition 2.6.1(i)-(ii). Consider $z\in\mathbb{C}_\gamma$ with $\gamma>2\sqrt{t}$, from the empirical fixed-point equation (2.63), we have

$$\begin{aligned}\left|\widehat{w}_{fp}^n(z)-w_{fp}(z)\right|&=t\left|\widehat{G}_{\mu_t^n}(\widehat{w}_{fp}^n(z))-G_{\mu_t}(w_{fp}(z))\right|\\&\leq t\left|\widehat{G}_{\mu_t^n}(\widehat{w}_{fp}^n(z))-\widehat{G}_{\mu_t^n}(w_{fp}(z))\right|+t\left|\widehat{G}_{\mu_t^n}(w_{fp}(z))-G_{\mu_t}(w_{fp}(z))\right|.\quad(2.68)\end{aligned}$$

Since $\widehat{G}_{\mu_t^n}$ is a Lipschitz function on $\mathbb{C}_{\frac{1}{2}\mathrm{Im}(z)}$, with Lipschitz constant $\leq\frac{4}{\gamma^2}$ (by Lemma 2.6.2 along with the fact that $\mathrm{Im}w_{fp}(z)\geq\frac{1}{2}\mathrm{Im}(z)$ by Theorem 2.4.2 and $\mathrm{Im}\widehat{w}_{fp}^n(z)\geq\frac{1}{2}\mathrm{Im}(z)$), we obtain an upper bound for the first term

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

Thus,

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

equivalently,

$$\Leftrightarrow\quad\left(1-\frac{4t}{\gamma^2}\right).\left|\widehat{w}_{fp}^n(z)-w_{fp}(z)\right|\leq t.\left|\widehat{G}_{\mu_t^n}\left(w_{fp}(z)\right)-G_{\mu_t}\left(w_{fp}(z)\right)\right|,$$

implying that

$$\left|\widehat{w}_{fp}^n(z) - w_{fp}(z)\right| \leq \left(\frac{t\gamma^2}{\gamma^2 - 4.t}\right) \times \left|\widehat{G}_{\mu_t^n}\left(w_{fp}(z)\right) - G_{\mu_t}\left(w_{fp}(z)\right)\right|. \tag{2.69}$$

By Proposition 2.3.5, since the function $x \mapsto \frac{1}{z-x}$ is continuous and bounded on $\mathbb{R}$ for any $z \in \mathbb{C}_{2\sqrt{t}}$, $\widehat{G}_{\mu_t^n}(w_{fp}(z)) = \int_{\mathbb{R}} \frac{1}{w_{fp}(z)-x} d\mu_t^n(x)$ converges almost surely to $G_{\mu_t}(w_{fp}(z)) = \int_{\mathbb{R}} \frac{1}{w_{fp}(z)-x} d\mu_t(x)$. 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 $\widehat{G}_{\mu_t^n}(z)$ decay as $1/|z|$ when $|z| \to +\infty$, implying the uniform convergence of the right-hand side of (2.69) on $\mathbb{C}_\gamma$, for $\gamma > 2\sqrt{t}$ and therefore of $\widehat{w}_{fp}^n(z)$ towards $w_{fp}(z)$.

□

2.6.2 Fluctuations of Cauchy transform of empirical measure

In this subsection, we will give a proof of the last point (iii) of Proposition 2.6.1. To do that, we will study the fluctuations of $\mathbb{E}\widehat{G}_{\mu_t^n}(z)$ around the Cauchy transform $G_{\mu_t}(z)$ for $z \in \mathbb{C}_{2\sqrt{t}}$. We first decompose into three terms:

$$\begin{aligned}
&\widehat{G}_{\mu_t^n}(z) - G_{\mu_t}(z) \\
&= \widehat{G}_{\mu_t^n}(z) - \mathbb{E}\left(\widehat{G}_{\mu_t^n}(z)\middle|X_n(0)\right) + \mathbb{E}\left(\widehat{G}_{\mu_t^n}(z)\middle|X_n(0)\right) - G_{\mu_0^n \boxplus \sigma_t}(z) + G_{\mu_0^n \boxplus \sigma_t}(z) - G_{\mu_t}(z) \\
&=: A_{n,1}(z) + A_{n,2}(z) + A_{n,3}(z).
\end{aligned} \tag{2.70}$$

The first term is related to the variance of $\widehat{G}_{\mu_t^n}(z)$ conditionally 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 for empirical measure of initial condition. A similar decomposition for the first two terms is done in Dallaporta and Février [27] and we will adapt their results.

Actually, in their recent paper [27], S.Dallaporta and M.Fevrier have considered a likely similar problem. More precisely, they focus on deformed Wigner matrices, i.e. sums Y_N of a Wigner matrix W_N and a deterministic Hermitian matrix D_N. In fact, the critical difference between [27] and our framework is that we consider a Dyson's Brownian motion

$$X_n(t) = X_n(0) + H_n(t),$$

where the initial condition $X_n(0)$ is a random matrix as in Assumption 2.3.2, while D_N is deterministic in [27]. Nevertheless, it is possible to take advantages from some of their results.

We now remind some properties of the Hermitian Brownian motion $H_n(t)$:

- off-diagonal entries for $1 \leq k < l \leq n$ are: $(H_n(t))_{kl} = \frac{1}{\sqrt{2n}}\left(B_{kl}(t) + i.\widetilde{B}_{kl}(t)\right)$, so

$$\sigma_n^2 := \mathbb{E}\left[\left|(H_n(t))_{kl}\right|^2\right] = \frac{t}{n}, \quad \tau_n := \mathbb{E}\left[(H_n(t))_{kl}^2\right] = 0, \quad m_n := \mathbb{E}\left[\left|(H_n(t))_{kl}\right|^4\right] = \frac{2.t^2}{n^2} \tag{2.71}$$

and thus,

$$\kappa_n := \mathbb{E}\left[\left|(H_n(t))_{kl}\right|^4\right] - 2\left(\mathbb{E}\left[\left|(H_n(t))_{kl}\right|^2\right]\right)^2 - \mathbb{E}\left[(H_n(t))_{kl}^2\right] = 0.$$

- diagonal entries: for $1 \leq k \leq n$ are: $(H_n(t))_{kk} = \frac{1}{\sqrt{n}} B_{kk}(t)$, so

$$\mathbb{E}\left[(H_n(t))_{kk}\right] = 0, \quad \text{and} \quad s_n^2 := \mathbb{E}\left[(H_n(t))_{kk}^2\right] = \frac{t}{n}.$$

These notations $\sigma_n^2, \tau_n, m_n, \kappa_n$ and s_n^2 are introduced initially in [27]. We keep using similar notations to see the correspondence between their computations and the adaptation in our framework.

■ Let us denote the resolvent of $X_n(t)$ by

$$R_{n,t}(z) := \left(zI_n - X_n(t)\right)^{-1}, \tag{2.72}$$

then one can write

$$\widehat{G}_{\mu_t^n}(z) = \frac{1}{n} \operatorname{Tr}\left(R_{n,t}(z)\right).$$

Notations

• Let $R_{n,t}^{(k)}(z)$ be the resolvent of the $(n-1) \times (n-1)$ obtained from $X_n(t)$ by removing the k-th row/column and $C_{k,t}^{(k)}$ be the $(n-1)$-dimensional vector obtained from the k-th column of $H_n(t)$ by removing its k-th component.

• Furthermore, let $\mathbb{E}_k$ denote the expectation with respect to $\left\{(H_n(t))_{jk} : 1 \leq j \leq n\right\}$, and $\mathbb{E}_{\leq k}$ denote the conditional expectation on the sigma-field

$$\sigma\left(((X_n(0))_{ij}, 1 \leq i \leq j \leq n), ((H_n(t))_{ij}, 1 \leq i \leq j \leq k)\right).$$

We begin with some results in linear algebra and martingale theory which are useful in later proofs. These results are given without a proof since being beyond the scope of our work. For more details about these results on inverse matrices and resolvent, one can found in, for example [3, Appendix A.1].

Proposition 2.6.3 (Schur complement). *Let $A \in \mathcal{M}_m(\mathbb{C})$ and $B \in \mathcal{M}_{m-1}(\mathbb{C})$ the submatrix of A obtained by removing its k-th row and k-th column. We denote by c the k-th column of A from which we have removed the diagonal entry A_{kk} and by r the k-th row of A from which we have removed the diagonal entry A_{kk}. Then A is invertible if and only if B is invertible and $A_{kk} - rB^{-1}c \neq 0$, in which case the following formulas hold:*

$$(A^{-1})_{kk} = \frac{1}{A_{kk} - r.B^{-1}.c}, \quad \text{and} \quad \operatorname{Tr}(A^{-1}) - \operatorname{Tr}(B^{-1}) = \frac{1 + rB^{-2}c}{A_{kk} - r.B^{-1}.c} \quad .$$

Lemma 2.6.4 (Resolvent identity). *Let M_1 and M_2 be $m \times m$ Hermitian matrices and denote by R_1 and R_2 their respective resolvents. Then, for all $z_1, z_2 \in \mathbb{C}\backslash\mathbb{R}$,*

$$R_1(z_1) - R_2(z_2) = R_1(z_1).\Big(\left(z_2 - z_1\right) I_m + M_1 - M_2\Big).R_2(z_2) \quad .$$

Consequently, the resolvent R of a $m \times m$ Hermitian matrix M satisfies

$$Im\, \psi\left(R(z)\right) = -Im(z)\psi\left(R(z)^*.R(z)\right), z \in \mathbb{C}\backslash\mathbb{R},$$

where $\psi : \mathcal{M}_m(\mathbb{C}) \to \mathbb{C}$ is any self-adjoint linear functional.

Lemma 2.6.5. *If A is an Hermitian matrix of size n, and $R(z) := (zI_n - A)^{-1}$ its resolvent, then for any $1 \leq k \leq n$ and $z \in \mathbb{C}\backslash\mathbb{R}$,*

$$|R(z)|_{kk} \leq \frac{1}{|Im(z)|}.$$

Lemma 2.6.6. *Let $(M_k)_{k\in\mathbb{N}}$ be a square integrable complex martingale. Then*

$$\mathbb{E}\Big[\Big|\sum_{k=1}^{m}(M_k - M_{k-1})\Big|^2\big|X_n(0)\Big] = \sum_{k=1}^{m}\mathbb{E}\left[\Big|(M_k - M_{k-1})\Big|^2\big|X_n(0)\right].$$

Now, notice that applying the Schur complement to $(zI_n - X_n(t))$ leads to expressions involving random quadratic forms $C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}$. Then it is reasonable to compute the expectation of such quadratic forms. More precisely, since $R_{n,t}^{(k)}(z)$ is deterministic under the expectation $\mathbb{E}_k$, we can write

$$\mathbb{E}\left[C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right] = \mathbb{E}\left[\mathbb{E}_k\left(C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right)\right] = \sigma_n^2.\mathbb{E}\left[\mathrm{Tr}\left(R_{n,t}^{(k)}(z)\right)\right]. \tag{2.73}$$

Moreover, the variance of $C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}$ can be controlled by the following lemma:

Lemma 2.6.7. *For $q \in \{1,2\}$ and $z \in \mathbb{C}\backslash\mathbb{R}$*

$$\mathbb{E}_k\left[\left|C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^q.C_{k,t}^{(k)} - \sigma_n^2.Tr\left(R_{n,t}^{(k)}(z)^q\right)\right|^2\right] = \frac{t^2}{n^2}.Tr\left(R_{n,t}^{(k)}(z)^{q*}.R_{n,t}^{(k)}(z)^q\right).$$

Proof. see Lemma 5, in [27] S.Dallaporta and M.Fevrier 2019, in the case $\tau_n = 0$ and $\kappa_n = 0$. □

The random variable $\mathrm{Tr}\left(R_{n,t}(z)\right) = n.\widehat{G}_{\mu_t^n}(z)$ being bounded since $n.\left|\widehat{G}_{\mu_t^n}(z)\right| \leq \frac{n}{|\mathrm{Im}(z)|}$ as in Lemma 2.6.2, and $\left(\mathbb{E}_{\leq k}\left[\mathrm{Tr}\left(R_{n,t}(z)\right)\right]\right)_{k\in\mathbb{N}}$ is a square integrable complex martingale, hence, by Lemma 2.6.6 in martingale theory, one can write

$$\begin{aligned}\mathrm{Var}\left[\mathrm{Tr}\left(R_{n,t}(z)\right)\big|X_n(0)\right] &= \mathbb{E}\left[\Big|\sum_{k=1}^{n}\left(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\right)\mathrm{Tr}\left(R_{n,t}(z)\right)\Big|^2\big|X_n(0)\right] \\ &= \sum_{k=1}^{n}\mathbb{E}\left[\Big|\left(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\right)\mathrm{Tr}\left(R_{n,t}(z)\right)\Big|^2\big|X_n(0)\right]. \end{aligned}\tag{2.74}$$

We show an upper-bound for the first term $A_{n,1}$ in the main decomposition (2.70).

Proposition 2.6.8. *([40, Proposition 3.2]) For $z \in \mathbb{C}\backslash\mathbb{R}$ and $n \in \mathbb{N}$,*

$$\mathrm{Var}\left(n.A_{n,1}(z)\big|X_n(0)\right) = \mathrm{Var}\big(n.\widehat{G}_{\mu_t^n}(z)\big|X_n(0)\big) = \mathrm{Var}\left[\mathrm{Tr}\left(R_{n,t}(z)\right)\big|X_n(0)\right] \leq \frac{10t}{|Im(z)|^4}.$$

This Proposition 2.6.8 corresponds to Proposition 3.2 in [40], (note that we do not give any proof of [40, Proposition 3.2] and it is referred to a proof presented here in this thesis).

Proof. (see Proposition 3 in [27] S.Dallaporta & M.Fevrier, 2019)
First of all, by Schur complement, one can express diagonal entries and trace of the resolvent of the Hermitian matrix $X_n(t)$. Applying the formula in Proposition 2.6.3 we obtain

$$\mathrm{Tr}\big(R_{n,t}(z)\big) - \mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) = \frac{1 + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}}{z - (H_n(t))_{kk} - (X_n(0))_{kk} - C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}}; \tag{2.75}$$

then it can be bounded by

$$\begin{aligned}
&\left|\frac{1+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}}{z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}}\right| \\
&\leq \frac{\left|1+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}\right|}{\left|\mathrm{Im}\left(z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right)\right|} \\
&\leq \frac{1+\left|C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}\right|}{\left|\mathrm{Im}(z)-\mathrm{Im}\left(C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right)\right|} \quad \text{(due to the fact that } (H_n(t))_{kk},(X_n(0))_{kk}\in\mathbb{R}) \\
&\leq \frac{1+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^*.R^{(k)}(z).C_{k,t}^{(k)}}{\left|\mathrm{Im}(z)+\mathrm{Im}\,(z)\,.C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right|} \quad \text{(by Lemma 2.6.4)} \\
&= \frac{1}{|\mathrm{Im}(z)|}. \qquad (2.76)
\end{aligned}$$

The fact is that in the right-hand side of (2.75), only $(H_n(t))_{kk}$ and $C_{k,t}^{(k)}$ are concerned by the k-th row/column of $H_n(t)$. It is reasonable to decompose the denominator in (2.75) as

$$\begin{aligned}
&\frac{1}{z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}} \\
&= \frac{z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}{\left(z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right).\left(z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right)} \\
&= \frac{\left[z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right]+(H_n(t))_{kk}+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}{\left(z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right).\left(z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right)} \\
&= \frac{1}{z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)} \\
&\qquad + \frac{(H_n(t))_{kk}+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}{\left(z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right).\left(z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right)} \quad ,
\end{aligned}$$

and also for the numerator

$$1+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}=1+\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)+\left[C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)\right] \quad .$$

Thus, we can rewrite (2.75) to get

$$\mathrm{Tr}\big(R_{n,t}(z)\big)=\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)+\frac{1+\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)}{z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}+\Psi_k, \qquad (2.77)$$

where

$$\Psi_k = \frac{\left[1 + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}\right].\left[(H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right]}{\left(z - (H_n(t))_{kk} - (X_n(0))_{kk} - C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right).\left(z - (X_n(0))_{kk} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right)}$$
$$+ \frac{C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)}{z - (X_n(0))_{kk} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}.$$

Since $R_{n,t}^{(k)}(z)$ and $(X_n(0))_{kk}$ do not involve the k-th row/column of $H_n(t)$, we get

$$(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1})\left[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) + \frac{1 + \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)}{z - (X_n(0))_{kk} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}\right] = 0. \tag{2.78}$$

this implies that

$$\begin{aligned}
\mathbb{E}\left[\left|\left(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\right)\left[\mathrm{Tr}\big(R_{n,t}(z)\big)\right]\right|^2 \Big| X_n(0)\right] &= \mathbb{E}\left[\left|\left(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\right)\left[\Psi_k\right]\right|^2 \Big| X_n(0)\right] \\
\text{(since } \mathbb{E}_{\leq k-1} = \mathbb{E}_{\leq k}\mathbb{E}_k\text{)} \quad &= \mathbb{E}\left[\left|\mathbb{E}_{\leq k}\left[\Psi_k - \mathbb{E}_k(\Psi_k)\right]\right|^2 \Big| X_n(0)\right] \\
\text{(Jensen's inequality)} \quad &\leq \mathbb{E}\left[\mathbb{E}_{\leq k}\left(\left|\Psi_k - \mathbb{E}_k(\Psi_k)\right|^2\right) \Big| X_n(0)\right] \\
&\leq \mathbb{E}\left[\left|\Psi_k - \mathbb{E}_k(\Psi_k)\right|^2 \Big| X_n(0)\right] \\
&\leq \mathbb{E}\left[\left(\mathbb{E}_k\left(\left|\Psi_k - \mathbb{E}_k(\Psi_k)\right|^2\right)\right) \Big| X_n(0)\right] \\
&\leq \mathbb{E}\left[\left(\mathbb{E}_k\left(\left|\Psi_k\right|^2\right)\right) \Big| X_n(0)\right].
\end{aligned}$$

We have the following bound as well:

$$\begin{aligned}
\left|\frac{1}{z - (X_n(0))_{kk} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}\right| &\leq \frac{1}{\left|\mathrm{Im}\left[z - (X_n(0))_{kk} - \sigma_n^2.\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right]\right|} \\
&\leq \frac{1}{\left|\mathrm{Im}(z) - \sigma_n^2.\mathrm{ImTr}\big(R_{n,t}^{(k)}(z)\big)\right|} \\
&\text{(since } X_n(0) \text{ is a real random matrix)} \\
&\leq \frac{1}{|\mathrm{Im}(z)|.\left[1 + \sigma_n^2.\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z)\big)\right]} \\
&\text{(since } \mathrm{ImTr}\big(R_{n,t}^{(k)}(z)\big) = -\mathrm{Im}(z).\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z)\big), \text{ by Lemma 2.6.4)} \\
&\leq \min\left(\frac{1}{|\mathrm{Im}(z)|}\ ,\ \frac{1}{|\mathrm{Im}(z)|.\sigma_n^2.\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z)\big)}\right).
\end{aligned}$$

• From Lemma 2.6.7 we can deduce that

$$\begin{aligned}
&\mathbb{E}_k\left[\left|(H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right|^2\right] \\
&= \mathbb{E}_k\left[\left((H_n(t))_{kk}\right)^2\right] + \mathbb{E}_k\left[\left|C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right|^2\right] \\
&\quad \left(\text{ since } \mathbb{E}_k\left((H_n(t))_{kk}\right) = 0, \text{ and } (H_n(t))_{kk} \text{ is independent with } C_{k,t}^{(k)} \text{ under } \mathbb{E}_k\right) \\
&\leq s_n^2 + m_n.\mathrm{Tr}\left(R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z)\right);
\end{aligned}$$

and also

$$\mathbb{E}_k\left[\left|C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)\right|^2\right]\leq m_n.\mathrm{Tr}\left(R_{n,t}^{(k)}(z)^{2*}.R_{n,t}^{(k)}(z)^2\right)$$
$$\leq m_n.\left|\mathrm{Im}(z)\right|^{-2}.\mathrm{Tr}\left(R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z)\right).$$

• Hence, for the first term in Ψ_k we have

$$\mathbb{E}_k\left[\left|\frac{(H_n(t))_{kk}+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}{z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}\right|^2\right]$$
$$=\frac{\mathbb{E}_k\left[\left|(H_n(t))_{kk}+C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right|^2\right]}{\left|z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right|^2}$$
$$\leq\frac{s_n^2+m_n.\mathrm{Tr}\left(R^{(k)}(z)^*.R^{(k)}(z)\right)}{\left|z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right|.\left|\mathbb{E}_k\left[z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right]\right|}$$
$$\Big(\text{since one has }\mathbb{E}_k\left[z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right]$$
$$=z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big)$$
$$\leq\frac{s_n^2+m_n.\mathrm{Tr}\left(R^{(k)}(z)^*.R^{(k)}(z)\right)}{\left|z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right|}.\mathbb{E}_k\left[\frac{1}{\left|z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right|}\right]$$
$$\Big(\text{ by Jensen's inequality applied as }\left|\mathbb{E}_k(X)\right|^{-1}\leq\mathbb{E}_k\big(\left|X\right|^{-1}\big)\Big)$$
$$\leq\left(\frac{s_n^2}{|\mathrm{Im}(z)|}+\frac{m_n.\mathrm{Tr}\left(R^{(k)}(z)^*.R^{(k)}(z)\right)}{|\mathrm{Im}(z)|.\sigma_n^2.\mathrm{Tr}\big(R^{(k)}(z)^*.R^{(k)}(z)\big)}\right).\mathbb{E}_k\left[\left|\big(R_{n,t}(z)\big)_{kk}\right|\right]$$
$$\text{(we have used Proposition 2.6.3 to replace the Schur complement of }\big(R_{n,t}(z)\big)_{kk}\text{)}$$
$$=\left(\frac{s_n^2}{|\mathrm{Im}(z)|}+\frac{m_n}{|\mathrm{Im}(z)|.\sigma_n^2}\right).\mathbb{E}_k\left[\left|\big(R_{n,t}(z)\big)_{kk}\right|\right],$$

then by (2.76), we get

$$\mathbb{E}_k\left[\left|\frac{\left[1+C_{k,t}^{(k)*}.R^{(k)}(z)^2.C_{k,t}^{(k)}\right].\left[(H_n(t))_{kk}+C_{k,t}^{(k)*}.R^{(k)}(z).C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R^{(k)}(z)\big)\right]}{\left(z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R^{(k)}(z).C_{k,t}^{(k)}\right).\left(z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R^{(k)}(z)\big)\right)}\right|^2\right]$$
$$\leq\left(\frac{s_n^2}{|\mathrm{Im}(z)|^3}+\frac{m_n}{|\mathrm{Im}(z)|^3.\sigma_n^2}\right).\mathbb{E}_k\left[\left|\big(R_{n,t}(z)\big)_{kk}\right|\right]\quad;$$

Similarly, for the second term in Ψ_k, one can obtain

$$\mathbb{E}_k\left[\left|\frac{C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)}{z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)}\right|^2\right]$$

$$=\frac{\mathbb{E}_k\left[\left|C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^2\big)\right|^2\right]}{\left|z-(X_n(0))_{kk}-\sigma_n^2\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\right|.\left|\mathbb{E}_k\left[z-(H_n(t))_{kk}-(X_n(0))_{kk}-C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\right]\right|}$$

$$\leq\frac{m_n.\left|\mathrm{Im}(z)\right|^{-2}.\mathrm{Tr}\left(R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z)\right)}{\left|\mathrm{Im}(z)\right|.\sigma_n^2.\mathrm{Tr}\big(R_{n,t}^{(k)}(z)^*.R_{n,t}^{(k)}(z)\big)}.\mathbb{E}_k\left[\left|\big(R_{n,t}(z)\big)_{kk}\right|\right]=\frac{m_n}{\left|\mathrm{Im}(z)\right|^3.\sigma_n^2}.\mathbb{E}_k\left[\left|\big(R_{n,t}(z)\big)_{kk}\right|\right]\quad.$$

Therefore, we have an upper-bound for the residual Ψ_k

$$\mathbb{E}_k\left[\left|\Psi_k\right|^2\right]\leq 2\left(\frac{s_n^2}{\left|\mathrm{Im}(z)\right|^3}+2.\frac{m_n}{\left|\mathrm{Im}(z)\right|^3.\sigma_n^2}\right).\mathbb{E}_k\left[\left|\big(R_{n,t}(z)\big)_{kk}\right|\right].$$

This implies that

$$\begin{aligned}\mathrm{Var}\left[\mathrm{Tr}\left(R_{n,t}(z)\right)\big|X_n(0)\right]&=\sum_{k=1}^n\mathbb{E}\left[\left|\big(\mathbb{E}_{\leq k}-\mathbb{E}_{\leq k-1}\big)\left[\mathrm{Tr}\big(R_{n,t}(z)\big)\right]\right|^2\,\middle|\,X_n(0)\right]\\&\leq\sum_{k=1}^n 2\left(\frac{s_n^2}{\left|\mathrm{Im}(z)\right|^3}+2.\frac{m_n}{\left|\mathrm{Im}(z)\right|^3.\sigma_n^2}\right).\mathbb{E}_k\left[\left|\big(R_{n,t}(z)\big)_{kk}\right|\right]\\&\leq 2\left(\frac{n.s_n^2}{\left|\mathrm{Im}(z)\right|^4}+2.\frac{n.m_n}{\left|\mathrm{Im}(z)\right|^4.\sigma_n^2}\right)=\frac{10t}{\left|\mathrm{Im}(z)\right|^4},\end{aligned}$$

and so we obtain an upper-bound for $\mathrm{Var}\left[\mathrm{Tr}\left(R_{n,t}(z)\right)\big|X_n(0)\right]$, which concludes the proof of Proposition 2.6.8 .

□

Fluctuations of $A_{n,2}(z)$

The bias term is:

$$nA_{n,2}(z)=n\mathbb{E}\big(\widehat{G}_{\mu_t^n}(z)\big|\,X_n(0)\big)-nG_{\mu_0^n\boxplus\sigma_t}(z)=\mathbb{E}\big(\mathrm{Tr}\left(R_{n,t}(z)\right)\,\big|\,X_n(0)\big)-nG_{\mu_0^n\boxplus\sigma_t}(z)\quad(2.79)$$

The control of the fluctuations of $A_{n,2}(z)$ is given by an adaptation of [27, Proposition 4] to the case of a random initial condition.

Proposition 2.6.9. *For $z\in\mathbb{C}^+$ and $n\in\mathbb{N}$,*

$$|nA_{n,2}(z)|\leq\left(1+\frac{4t}{Im^2(z)}\right).\left(\frac{2t}{Im^3(z)}+\frac{12t^2}{Im^5(z)}\right).\quad(2.80)$$

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

$$|n.A_{n,2}(z)|\leq 2n\mathrm{Im}^{-1}(z),\quad(2.81)$$

which is suboptimal due to the factor n.

We follow the ideas of [27] for their 'approximate subordination relations'. Since our initial condition is random, the strategy has to be adapted and we introduce the following analogues of $R_{n,t}(z)$ and $A_{n,2}(z)$, which differ from [27]:

$$\begin{aligned}\widetilde{R}_{n,t}(z) &:= \Big(\big(z - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\left(R_{n,t}(z)\right) \mid X_n(0)\big]\big).I_n - X_n(0)\Big)^{-1} \\ n.\widetilde{A}_{n,2}(z) &:= \mathbb{E}\left(\mathrm{Tr}\left(R_{n,t}(z)\right) \mid X_n(0)\right) - \mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big).\end{aligned} \tag{2.82}$$

We will bound $A_{n,2}(z)$ by using its approximation $\widetilde{A}_{n,2}(z)$.

Step 1: First, we prove an upper bound for $n.\widetilde{A}_{n,2}(z)$:

Lemma 2.6.10. *For $z \in \mathbb{C}^+$,*

$$\big|n.\widetilde{A}_{n,2}(z)\big| \leq \frac{2t}{(Im(z))^3} + \frac{12t^2}{(Im(z))^5}.$$

The proof of this Lemma is postponed and given after the conclusion in proof of Proposition 2.6.9.

Step 2: If $|n.\widetilde{A}_{n,2}(z)| \geq \dfrac{|\mathrm{Im}(z)|}{2\sigma_n^2} = \dfrac{n.\mathrm{Im}(z)}{2t}$ a.s. with respect to $\lambda^n(0)$ then, by (2.81)

$$|n.A_{n,2}(z)| \leq \frac{4t.|n.\widetilde{A}_{n,2}(z)|}{(\mathrm{Im}(z))^2},$$

and along with Lemma 2.6.10 we can conclude the associating bound on $|nA_{n,2}(z)|$.

Step 3: On the other hand, we now consider the case where $|n.\widetilde{A}_{n,2}(z)| < n\mathrm{Im}(z)/(2t)$. We rewrite as:

$$A_{n,2}(z) = \widetilde{A}_{n,2}(z) + \big[A_{n,2}(z) - \widetilde{A}_{n,2}(z)\big] \tag{2.83}$$

The difference $|A_{n,2}(z) - \widetilde{A}_{n,2}(z)|$ will be controlled by $|\widetilde{A}_{n,2}(z)|$ and then conclude with Lemma 2.6.10. By their definitions:

$$n.\big(A_{n,2}(z) - \widetilde{A}_{n,2}(z)\big) = Tr\big(\widetilde{R}_{n,t}(z)\big) - nG_{\mu_0^n \boxplus \sigma_t}(z). \tag{2.84}$$

We follow the trick in [27] which consists in going back to the fluctuations of the subordination functions. Proceeding as in Theorem 2.2.5, with μ_0^n and σ_t, we can define a subordination function $\overline{w}_{fp}(z)$ such that

$$G_{\mu_0^n \boxplus \sigma_t}(z) = G_{\mu_0^n}\big(\overline{w}_{fp}(z)\big). \tag{2.85}$$

Thus, it is natural to express the first term $\mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big)$ of (2.84) similarly. As $\widetilde{R}_{n,t}(z)$ is a diagonal matrix,

$$\mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big) = \sum_{j=1}^{n} \frac{1}{z - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}(z)\big) \mid X_n(0)\big] - \lambda_j^n(0)} = nG_{\mu_0^n}(\widetilde{w}_{fp}(z)), \tag{2.86}$$

where $\widetilde{w}_{fp}(z) := z - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}(z)\big) \mid X_n(0)\big]$ and where $\lambda_j^n(0)$ are the eigenvalues of $X_n(0)$. Thus:

$$A_{n,2}(z) - \widetilde{A}_{n,2}(z) = G_{\mu_0^n}(\widetilde{w}_{fp}(z)) - G_{\mu_0^n}(\overline{w}_{fp}(z)). \tag{2.87}$$

To continue, we first need the following result whose proof is given later.

Lemma 2.6.11. *(i) The function $\overline{w}_{fp}(z)$, defined in (2.85), solves*

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

(ii) The function $\kappa(z) = z + tG_{\mu_0^n}(z)$ is well-defined on $\mathbb{C}^+$ and is the inverse of $\overline{w}_{fp}(z)$ on $\overline{\Omega} = \{z \in \mathbb{C}^+,\ Im(\kappa(z)) > 0\}$. 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 $z \in \mathbb{C}^+$. Indeed, we have

$$\begin{aligned}
\kappa(\widetilde{w}_{fp}(z)) - z &= \widetilde{w}_{fp}(z) + t.G_{\mu_0^n}\big(\widetilde{w}_{fp}(z)\big) - z \\
&= z - \frac{t}{n}\mathbb{E}\big[Tr\big(R_{n,t}(z)\big) \mid X_n(0)\big] + tG_{\mu_0^n}\big(\widetilde{w}_{fp}(z)\big) - z \\
&= -\frac{t}{n}\mathbb{E}\big[Tr\big(R_{n,t}(z)\big) \mid X_n(0)\big] + \frac{t}{n}\ \mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big) \\
&= -t.\widetilde{A}_{n,2}(z), \quad \Big(\text{ by (2.86): } \mathrm{Tr}\big(\widetilde{R}_{n,t}(z)\big) = n.G_{\mu_0^n}\big(\widetilde{w}_{fp}(z)\big)\Big)
\end{aligned} \tag{2.88}$$

Therefore,

$$\big|\mathrm{Im}\big(\kappa(\widetilde{w}_{fp}(z))\big) - \mathrm{Im}(z)\big| \leq |\kappa(\widetilde{w}_{fp}(z)) - z| = t.\big|\widetilde{A}_{n,2}(z)\big| \leq \frac{\mathrm{Im}(z)}{2}. \tag{2.89}$$

Thus, under the condition of Step 3, i.e. $|\widetilde{A}_{n,2}(z)| < \mathrm{Im}(z)/(2t)$, we get $\widetilde{w}_{fp}(z) \in \overline{\Omega}$. Denoting $\widetilde{z} = \overline{w}_{fp}^{<-1>}\big(\widetilde{w}_{fp}(z)\big)$, which is well-defined as $\widetilde{w}_{fp}(z) \in \overline{\Omega}$, we have $\overline{w}_{fp}(\widetilde{z}) = \widetilde{w}_{fp}(z)$. Then plugging this into (2.87),

$$\begin{aligned}
A_{n,2}(z) - \widetilde{A}_{n,2}(z) &= G_{\mu_0^n \boxplus \sigma_t}(\widetilde{z}) - G_{\mu_0^n \boxplus \sigma_t}(z) \\
&= \big(z - \widetilde{z}\big) \int_{\mathbb{R}} \frac{\mu_0^n \boxplus \sigma_t(dx)}{\big(\widetilde{z} - x\big).\big(z - x\big)} \\
&= t\widetilde{A}_{n,2}(z). \int_{\mathbb{R}} \frac{\mu_0^n \boxplus \sigma_t(dx)}{\big(\widetilde{z} - x\big).\big(z - x\big)},
\end{aligned}$$

where we used (2.88) for the last equality.

From there, using (2.89), we get

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

This concludes the proof of Proposition 2.6.9.

□

We present here proofs of Lemma 2.6.10 and Lemma 2.6.11.

Proof of Lemma 2.6.10. Recall that the resolvent $R_{n,t}(z)$ and $\widetilde{R}_{n,t}(z)$ are defined in (2.72) and (2.82) respectively, and that

$$n.\widetilde{A}_{n,2}(z) = \sum_{k=1}^{n} \mathbb{E}\left[\big(R_{n,t}(z)\big)_{kk} \ \big|X_n(0)\right] - \big(\widetilde{R}_{n,t}(z)\big)_{kk}. \tag{2.90}$$

Using Schur's complement (Proposition 2.6.3, see also e.g. [3, Appendix A.1]):

$$\Big(\big(R_{n,t}(z)\big)_{kk}\Big)^{-1} = z - (H_n(t))_{kk} - (X_n(0))_{kk} - C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)},$$

and because $\widetilde{R}_{n,t}(z)$ is a diagonal matrix, one gets:

$$\Big(\big(\widetilde{R}_{n,t}(z)\big)_{kk}\Big)^{-1} = z - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\,(R_{n,t}(z))\,\big|X_n(0)\big] - (X_n(0))_{kk}\,.$$

Then we have

$$\Big(\big(\widetilde{R}_{n,t}(z)\big)_{kk}\Big)^{-1} - \Big(\big(R_{n,t}(z)\big)_{kk}\Big)^{-1} = (H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\,(R_{n,t}(z))\,\big|X_n(0)\big],$$

or equivalently,

$$\begin{aligned}\big(R_{n,t}(z)\big)_{kk} =& \big(\widetilde{R}_{n,t}(z)\big)_{kk}\\ &+\big(\widetilde{R}_{n,t}(z)\big)_{kk}.\big(R_{n,t}(z)\big)_{kk}.\Big((H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\,(R_{n,t}(z))\,\big|X_n(0)\big]\Big).\end{aligned}$$

Furthermore, replacing $\big(R_{n,t}(z)\big)_{kk}$ in the right-hand side of the previous formula, we obtain:

$$\begin{aligned}&(R_{n,t}(z))_{kk} - \big(\widetilde{R}_{n,t}(z)\big)_{kk}\\ &=\big(\widetilde{R}_{n,t}(z)\big)^2_{kk}.\left((H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathbb{E}\left[\mathrm{Tr}\,(R_{n,t}(z))\;\middle|\; X_n(0)\right]\right)\\ &+\big(\widetilde{R}_{n,t}(z)\big)^2_{kk}.\big(R_{n,t}(z)\big)_{kk}.\left((H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathbb{E}\left[\mathrm{Tr}\,(R_{n,t}(z))\;\middle|\; X_n(0)\right]\right)^2.\end{aligned} \tag{2.91}$$

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

$$\begin{aligned}&\mathbb{E}\Big[\Big|\,(H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathbb{E}\left[\mathrm{Tr}\,(R_{n,t}(z))\;\middle|\; X_n(0)\right]\Big|^2\Big|X_n(0)\Big]\\ &= \mathbb{E}\Big[\,\Big|(H_n(t))_{kk} + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) + \frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big]\\ &\qquad +\frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big] - \frac{t}{n}\mathbb{E}\left[\mathrm{Tr}\,(R_{n,t}(z))\;\middle|\; X_n(0)\right]\Big|^2\,\Big|X_n(0)\Big]\\ &= \mathbb{E}\left[(H_n(t))^2_{kk}\right] + \mathbb{E}\Big[\,\Big|C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|^2\,\big|X_n(0)\Big]\\ &\qquad + \frac{t^2}{n^2}\Big(\mathrm{Var}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\big|X_n(0)\big] + \Big|\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) - \mathrm{Tr}\,(R_{n,t}(z))\;\big|\; X_n(0)\big]\Big|^2\Big).\end{aligned} \tag{2.92}$$

in which we have used the fact that

$$\begin{aligned}&\mathbb{E}\Big[\Big(C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} - \frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big).\Big(\frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big]\Big)\Big|X_n(0)\Big]\\ &= \mathbb{E}\Big[C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}.\frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big] - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big].\mathbb{E}\Big[C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}\Big|X_n(0)\Big]\\ &\quad - \Big(\frac{t}{n}\Big)^2.\mathbb{E}\Big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big).\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big] + \Big(\frac{t}{n}\Big)^2.\Big(\mathbb{E}\Big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big]\Big)^2\\ &= \mathbb{E}\Big[\mathbb{E}_k\Big(C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}.\frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big)\Big] - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big].\frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big]\\ &\quad - \Big(\frac{t}{n}\Big)^2.\mathbb{E}\Big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big).\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big] + \Big(\frac{t}{n}\Big)^2.\Big(\mathbb{E}\Big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big]\Big)^2\\ &= \mathbb{E}\Big[\frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big).\frac{t}{n}\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big] - \frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big].\frac{t}{n}\mathbb{E}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\;\big|\; X_n(0)\big]\\ &\quad - \Big(\frac{t}{n}\Big)^2.\mathbb{E}\Big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big).\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big] + \Big(\frac{t}{n}\Big)^2.\Big(\mathbb{E}\Big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big|X_n(0)\Big]\Big)^2\\ &= 0.\end{aligned}$$

We now upper bound each of the term in the right-hand side of (2.92). As in (2.71), the first term equals to t/n.

■ **Step 1:** We upper bound the second term in (2.92). By (2.73) (see also Lemma 5 of [27]),

$$\mathbb{E}\Big[C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} \mid X_n(0)\Big] = \frac{t}{n}\mathbb{E}\Big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) \mid X_n(0)\Big]. \tag{2.93}$$

Thus, the second term in (2.92) equals exactly to $\mathrm{Var}\big(C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} \mid X_n(0)\big)$ and we have:

$$\begin{aligned}\mathrm{Var}\left[C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} \mid X_n(0)\right] =& \frac{t^2}{n^2}\mathbb{E}\left[\mathrm{Tr}\big(R_{n,t}^{(k),*}(z).R_{n,t}^{(k)}(z)\big) \mid X_n(0)\right] \\ \leq& \frac{t^2}{n^2}\mathbb{E}\left[\sum_{j=1}^{n}\frac{1}{|z-\lambda_j^{(k)}|^2} \mid X_n(0)\right]\end{aligned}$$

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

$$\mathrm{Var}\left[C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)} \mid X_n(0)\right] \leq \frac{t^2}{n(\mathrm{Im}(z))^2}. \tag{2.94}$$

■ **Step 2:** We now upper bound the third and fourth terms of (2.92). To reminds, we denote by $\mathbb{E}_k$ the expectation with respect to $\Big\{(H_n(t))_{jk} : 1 \leq j \leq n\Big\}$, and by $\mathbb{E}_{\leq k}$ the conditional expectation on the sigma-field $\sigma\Big(((X_n(0))_{ij}, 1 \leq i \leq j \leq n), ((H_n(t))_{ij}, 1 \leq i \leq j \leq k)\Big)$.
We have:

$$\mathrm{Var}\big[\mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\big|X_n(0)\big] \leq 2\mathrm{Var}\big[\mathrm{Tr}\big(R_{n,t}(z)\big)\big|X_n(0)\big] + 2\mathrm{Var}\big[\mathrm{Tr}\big(R_{n,t}(z)\big) - \mathrm{Tr}(R_{n,t}^{(k)}(z)\big|X_n(0)\big]. \tag{2.95}$$

For the first term,

$$\begin{aligned}\mathrm{Var}\big[\mathrm{Tr}\big(R_{n,t}(z)\big)\big|X_n(0)\big] =& \sum_{k=1}^{n}\mathbb{E}\left[\left|\big(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\big)\mathrm{Tr}\left(R_{n,t}(z)\right)\right|^2 \mid X_n(0)\right] \\ =& \sum_{k=1}^{n}\mathbb{E}\left[\left|\big(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\big)\big(\mathrm{Tr}(R_{n,t}(z)) - \mathrm{Tr}(R_{n,t}^{(k)}(z))\big)\right|^2 \mid X_n(0)\right],\end{aligned} \tag{2.96}$$

as $\big(\mathbb{E}_{\leq k} - \mathbb{E}_{\leq k-1}\big)\mathrm{Tr}\left(R_{n,t}^{(k)}(z)\right) = 0$ since $R_{n,t}^{(k)}(z)$ does not involve the k-th row/column of $H_n(t)$. Again, as in (2.75), the Schur complement formula gives that:

$$\mathrm{Tr}\big(R_{n,t}(z)\big) - \mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big) = \frac{1 + C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z)^2.C_{k,t}^{(k)}}{z - (H_n(t))_{kk} - (X_n(0))_{kk} - C_{k,t}^{(k)*}.R_{n,t}^{(k)}(z).C_{k,t}^{(k)}}.$$

Then, as in (2.76) we get

$$\Big|\mathrm{Tr}\big(R_{n,t}(z)\big) - \mathrm{Tr}\big(R_{n,t}^{(k)}(z)\big)\Big| \leq \frac{1}{\mathrm{Im}(z)}.$$

The second inequality it due to the fact that $(H_n(t))_{kk}, (X_n(0))_{kk} \in \mathbb{R}$ and the third inequality comes from the following equality: With $\Psi : M \in \mathcal{H}_n(\mathbb{C}) \mapsto C^*MC$ with $C \in \mathbb{C}^n$, then, for any $z \in \mathbb{C}$ and any resolvent matrix $R(z)$, we have (see [27, Lemma 1])

$$\mathrm{Im}\big(\Psi(R(z))\big) = -\mathrm{Im}(z)\Psi\big(R(z)^*R(z)\big).$$

The bound (2.76) does not depend on $X_n(0)$. Plugging this bound into (2.96), we obtain:

$$\mathrm{Var}\big[\mathrm{Tr}\big(R_{n,t}(z)\big)\big|X_n(0)\big] \leq \frac{4n}{(\mathrm{Im}(z))^2}.$$

From there, using (2.95),

$$\mathrm{Var}\big[\mathrm{Tr}\big(R^{(k)}_{n,t}(z)\big)\big|X_n(0)\big] \leq \frac{8n+2}{(\mathrm{Im}(z))^2}. \tag{2.97}$$

Similarly, (2.76) also provides an upper bound for the fourth term of (2.92):

$$\Big|\mathbb{E}\big[\mathrm{Tr}\big(R^{(k)}_{n,t}(z)\big) - \mathrm{Tr}\left(R_{n,t}(z)\right) \mid X_n(0)\big]\Big|^2 \leq \frac{1}{\mathrm{Im}^2(z)}. \tag{2.98}$$

■ **Step 3:** In conclusion, combining (2.92), (2.94), (2.97) and (2.98) together, we obtain that:

$$\mathbb{E}\left[\left|(H_n(t))_{kk} + C^{(k)*}_{k,t}.R^{(k)}_{n,t}(z).C^{(k)}_{k,t} - \frac{t}{n}\mathbb{E}\left[\mathrm{Tr}\left(R_{n,t}(z)\right) \mid X_n(0)\right]\right|^2 \Big| X_n(0)\right] \leq \frac{t}{n} + \frac{t^2}{n(\mathrm{Im}(z))^2} + (8n+3)\,.\frac{t^2}{n^2.(\mathrm{Im}(z))^2}.$$

Going back to (2.91) and using (2.93) to upper-bound the first term in the right-hand side:

$$\begin{aligned}
&\Big|\mathbb{E}\left[(R_{n,t}(z))_{kk} - \big(\widetilde{R}_{n,t}(z)\big)_{kk} \mid X_n(0)\right]\Big| \\
&\qquad \leq \frac{t}{n}\big|\big(\widetilde{R}_{n,t}(z)\big)_{kk}\big|^2.\mathbb{E}\left[\big|\mathrm{Tr}\big(R^{(k)}_{n,t}(z)\big) - \mathrm{Tr}\big(R_{n,t}(z)\big)\big| \;\Big|\; X_n(0)\right] \\
&\qquad + \big|\big(\widetilde{R}_{n,t}(z)\big)_{kk}\big|^2.\mathbb{E}\left[\big|\big(R_{n,t}(z)\big)_{kk}\big|.\big|(H_n(t))_{kk} + C^{(k)*}_{k,t}.R^{(k)}_{n,t}(z).C^{(k)}_{k,t}\right. \\
&\qquad \left. - \frac{t}{n}\mathbb{E}\left[\mathrm{Tr}\left(R_{n,t}(z)\right) \mid X_n(0)\right]\big|^2 \;\Big|\; X_n(0)\right] \\
&\qquad \leq \big|\big(\widetilde{R}_{n,t}(z)\big)_{kk}\big|^2.\left(\frac{t}{n.\mathrm{Im}(z)} + \frac{t}{n.\mathrm{Im}(z)} + \frac{t^2}{n.(\mathrm{Im}(z))^3} + \frac{(8n+3)t^2}{n^2(\mathrm{Im}(z))^3}\right) \\
&\qquad \leq \big|\big(\widetilde{R}_{n,t}(z)\big)_{kk}\big|^2.\frac{1}{n}.\left(\frac{2t}{\mathrm{Im}(z)} + \frac{12t^2}{(\mathrm{Im}(z))^3}\right).
\end{aligned}$$

Using this upper bound in (2.90), we obtain by summation the result and using that for any k,

$$\big|\widetilde{R}_{n,t}(z)\big)_{kk}\big|^2 \leq \frac{1}{(\mathrm{Im}(z))^2}.$$

This concludes the proof of Lemma 2.6.10.

□

Proof of Lemma 2.6.11. Using subordination functions, from (2.85) and introducing $\overline{w}_1(z)$ such that:

$$G_{\mu_0^n \boxplus \sigma_t}(z) = G_{\mu_0^n}\big(\overline{w}_{fp}(z)\big) = G_{\sigma_t}(\overline{w}_1(z)).$$

We can derive from Theorem 2.2.5 that $\overline{w}_{fp}(z)$ solves the equation (i) of Lemma 2.6.11 and that:

$$z = \overline{w}_{fp}(z) + tG_{\mu_0^n}(\overline{w}_{fp}(z)),$$

for all $z \in \mathbb{C}^+$. The latter equation justifies (ii) of Lemma 2.6.11. □

Fluctuation of $A_{n,3}(z)$

Finally, the third step is to control $A_{n,3}(z) = G_{\mu_0^n \boxplus \sigma_t}(z) - G_{\mu_t}(z)$, and notice that $\mu_t = \mu_0 \boxplus \sigma_t$.

Proposition 2.6.12. *For any $\gamma > 2\sqrt{t}$ and for any z such that $Im(z) \geq \frac{\gamma}{2}$, we have:*

$$|A_{n,3}(z)| < \frac{\gamma^2}{\gamma^2 - 4t} \left| \int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x} \left[d\mu_0^n(x) - d\mu_0(x) \right] \right| \tag{2.99}$$

and

$$\sup_{n \in \mathbb{N}} \sup_{z \in \mathbb{C}_{\frac{\gamma}{2}}} \mathbb{E}\left[n \left| A_{n,3}(z) \right|^2 \right] < \frac{8\gamma^2}{(\gamma^2 - 4t)^2}. \tag{2.100}$$

Proof. Using again the subordination function $\overline{w}_{fp}(z)$ defined in (2.85) and Lemma 2.6.11-(i), we have

$$G_{\mu_0^n \boxplus \sigma_t}(z) = G_{\mu_0^n}\left(\overline{w}_{fp}(z) \right) = \int_{\mathbb{R}} \frac{d\mu_0^n(x)}{\overline{w}_{fp}(z) - x} = \int_{\mathbb{R}} \frac{d\mu_0^n(x)}{z - tG_{\mu_0^n \boxplus \sigma_t}(z) - x}. \tag{2.101}$$

In this proof, $\mathrm{Im}(z) \geq \gamma/2 \geq \sqrt{t}$. Note that $\mathrm{Im}(\overline{w}_{fp}(z)) \geq \mathrm{Im}(z)$ (by Theorem-Definition 2.2.5) so that

$$|\overline{w}_{fp}(z) - x| = |z - tG_{\mu_0^n \boxplus \sigma_t}(z) - x| \geq \frac{\gamma}{2} \geq \sqrt{t}, \tag{2.102}$$

and the integrand in (2.101) is well-defined and upper-bounded by $1/\sqrt{t}$. Similarly, we can establish that

$$G_{\mu_0 \boxplus \sigma_t}(z) = \int_{\mathbb{R}} \frac{d\mu_0(x)}{z - tG_{\mu_0 \boxplus \sigma_t}(z) - x}. \tag{2.103}$$

Then, we can write

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

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

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

and so

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

Consequently,

$$|A_{n,3}(z)| < \frac{\gamma^2}{\gamma^2 - 4t}\left|\int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x}\left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right]\right|,$$

which gives the first part of the proposition. For the second part (2.100),

$$\mathbb{E}\left[n.\,|A_{n,3}(z)|^2\right] \leq \left(\frac{\gamma^2}{\gamma^2 - 4t}\right)^2 n.\mathbb{E}\left[\left|\int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x}\left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right]\right|^2\right].$$

Now, for any z such that $\mathrm{Im}(z) > \frac{\gamma}{2}$, the function $\varphi_z : x \mapsto \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x}$ is bounded by $2/\gamma$. Remind that $(d_j)_{1\leq j\leq n}$ are random variables i.i.d. from distribution μ_0 and $(\lambda_j^n(0))_{1\leq j\leq n}$ is the ordered collection of $(d_j)_{1\leq j\leq n}$, we then have:

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

for any $z \in \mathbb{C}_{\frac{\gamma}{2}}$.

□

Conclusion: We can now conclude the proof of Proposition 2.6.1-(iii). First, from Proposition 2.6.8 we obtain for $w \in \mathbb{C}_{\frac{\gamma}{2}}$:

$$\mathrm{Var}\left(n.A_{n,1}(w)\middle|X_n(0)\right) = \mathrm{Var}\left(n.\widehat{G}_{\mu_t^n}(w)\middle|X_n(0)\right) \leq \frac{10t}{\left(\mathrm{Im}(w)\right)^4} \leq \frac{10t.2^4}{\gamma^4} < \frac{10.2^2}{\gamma^2};$$

Moreover, from Proposition 2.6.9 we get for $w \in \mathbb{C}_{\frac{\gamma}{2}}$:

$$\begin{aligned}
\left|n.A_{n,2}(w)\right| \leq \left(1 + \frac{4t}{\mathrm{Im}(w)^2}\right).\left(\frac{2t}{\mathrm{Im}(w)^3} + \frac{12t}{\mathrm{Im}(w)^5}\right) &\leq \left(1 + \frac{4t.2^2}{\gamma^2}\right).\left(\frac{2t.2^3}{\gamma^3} + \frac{12t.2^5}{\gamma^5}\right) \\
&= \left(\frac{2.2^2t}{\gamma^2} + \frac{20.2^4t^2}{\gamma^4} + \frac{48.2^6t^3}{\gamma^6}\right).\frac{2}{\gamma} \\
&< \frac{140}{\gamma} \quad \text{(using } \gamma > 2\sqrt{t}\text{)};
\end{aligned}$$

Thus, from (2.70) and combining with the first part of Proposition 2.6.12, we obtain that for $w \in \mathbb{C}_{\frac{\gamma}{2}}$ with $\gamma > 2\sqrt{t}$:

$$\begin{aligned}
\mathbb{E}\left[|\widehat{G}_{\mu_t^n}(w) - G_{\mu_t}(w)|^2 \mid X_n(0)\right] &\leq 3.\left[\mathrm{Var}\left(A_{n,1}(w)\middle|X_n(0)\right) + \left|A_{n,2}(w)\right|^2 + \left|A_{n,3}(w)\right|^2\right] \\
&\leq 3.\left[\frac{40}{\gamma^2}.\frac{1}{n^2} + \frac{140^2}{\gamma^2}.\frac{1}{n^2}\right. \\
&\quad \left. + \left(\frac{\gamma^2}{\gamma^2 - 4t}\right)^2.\left|\int_{\mathbb{R}} \frac{1}{z - t.G_{\mu_0 \boxplus \sigma_t}(z) - x}\left[\mathrm{d}\mu_0^n(x) - \mathrm{d}\mu_0(x)\right]\right|^2\right].
\end{aligned}$$

Hence,

$$\mathbb{E}\big[|\widehat{G}_{\mu_t^n}(w)-G_{\mu_t}(w)|^2 \mid X_n(0)\big] \leq C(\gamma,t).\Big(\frac{1}{n^2}+\Big|\int_{\mathbb{R}} \frac{1}{w-t.G_{\mu_0\boxplus\sigma_t}(w)-x}\big[\mathrm{d}\mu_0^n(x)-\mathrm{d}\mu_0(x)\big]\Big|^2\Big), \tag{2.104}$$

where $C(\gamma,t) := 3.(40+140^2).\frac{1}{\gamma^2}+3.\Big(\frac{\gamma^2}{\gamma^2-4t}\Big)^2$, depends only on γ and t (and converges to $+\infty$ when $\gamma \to 2\sqrt{t}$). We can also see 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:

$$\sup_{n\in\mathbb{N}}\sup_{w\in\mathbb{C}_{\gamma/2}} n.\mathbb{E}\Big[|\widehat{G}_{\mu_t^n}(w)-G_{\mu_t}(w)|^2\Big] = \sup_{n\in\mathbb{N}}\sup_{w\in\mathbb{C}_{\gamma/2}} n.\mathbb{E}\Big[\mathbb{E}\big[|\widehat{G}_{\mu_t^n}(w)-G_{\mu_t}(w)|^2 \mid X_n(0)\big]\Big] \leq \frac{C(t)\gamma^4}{(\gamma^2-4t)^2},$$

by using the second part of Proposition 2.6.12. Equation (2.69) implies that for any $\gamma > 2\sqrt{t}$,

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

since if $z \in \mathbb{C}_\gamma$ then $\mathrm{Im}\big(w_{fp}(z)\big) \geq \frac{1}{2}\mathrm{Im}(z) > \frac{\gamma}{2}$ (by Theorem 2.4.2) so that $w_{fp}(z) \in \mathbb{C}_{\frac{\gamma}{2}}$ and point (iii) of Proposition 2.6.1 is proved.

2.7 Mean Integrated Squared Error - MISE

In Section 2.7.1, we state theoretical results associated with our nonparametric statistical problem. Section 2.7.2 is devoted to the proof of Theorem 2.7.2.

2.7.1 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 $\widehat{p}_{0,h}$. To derive rates of convergence, we rely on the classical bias-variance decomposition of the quadratic risk. Using Parseval's equality we obtain

$$\|\widehat{p}_{0,h}-p_0\|^2 = \frac{1}{2\pi}\left\|\widehat{p}_{0,h}^\star-p_0^\star\right\|^2 \leq \frac{1}{\pi}\left\|\widehat{p}_{0,h}^\star-K_h^\star.p_0^\star\right\|^2 + \frac{1}{\pi}\left\|K_h^\star.p_0^\star-p_0^\star\right\|^2. \tag{2.105}$$

The expectation of the first term is a variance term whereas the second one is a bias term. To derive the order of the bias term, we shall consider two classes of densities, super-smooth densities and densities belonging to Sobolev-ordinary smooth classes. First, we assume that p_0 belongs to the space $\mathcal{S}(a,r,L)$ of supersmooth densities defined for $a>0$, $L>0$ and $r>0$ by:

$$\mathcal{S}(a,r,L) = \left\{g \text{ density such that } \int_{\mathbb{R}} |g^\star(\xi)|^2 e^{2a|\xi|^r} d\xi \leq L\right\}. \tag{2.106}$$

In the literature, this smoothness class of densities has often been considered (see [37], [16], [25]). Most famous examples of supersmooth densities are the Cauchy distribution belonging to $\mathcal{S}(a,r,L)$ with $r=1$ and the Gaussian distribution belonging to $\mathcal{S}(a,r,L)$ with $r=2$. To control the bias, we rely on Proposition 1 in [16] which states that:

Proposition 2.7.1. *For $p_0 \in \mathcal{S}(a,r,L)$, we have*

$$\|K_h^\star.p_0^\star-p_0^\star\|^2 \leq L.e^{-2ah^{-r}}.$$

Whereas the control of the bias term is very classical, the study of the variance term in (2.105) is much more involved. The order of the variance term is provided by the following theorem.

Theorem 2.7.2. *Let*

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

Assume (2.15). *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,*

$$\mathbb{E}(\Sigma) \leq \frac{\gamma^8}{(\gamma^4 - 4t)^4} . \frac{C_{var}(t) . e^{\frac{2\gamma}{h}}}{n} . \tag{2.107}$$

Theorem 2.7.2 is proved in Section 2.7.2.The main point will be to obtain 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 (2.55) and (2.64)). 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 (2.15) is very mild and is satisfied by most classical distributions.

Now, using similar computations to those in [37], we can obtain from Proposition 2.7.1 and Theorem 2.7.2 the rates of convergence of our estimator $\widehat{p}_{0,h}$. We indeed showed that:

$$\text{MISE} := \mathbb{E}\Big[\left\| \widehat{p}_{0,h} - p_0 \right\|^2 \Big] \leq L . e^{-2ah^{-r}} + \frac{\gamma^8}{(\gamma^4 - 4t)^4} . \frac{C_{var}(t) . e^{\frac{2\gamma}{h}}}{n} . \tag{2.108}$$

Minimizing in h the right hand side of (2.108) 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 [37]. 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 2.7.3 is proved later in the end of Section 2.7.2.

Corollary 2.7.3. *Suppose that* μ_0 *satisfies Assumption* (2.15) *and the density* p_0 *belongs to the space* $\mathcal{S}(a, r, L)$ *for* $a > 0$, $r > 0$ *and* $L > 0$. *Then, for any* $\gamma > 2\sqrt{t}$ *and by choosing the bandwidth* h *according to equation* (2.146), *we have:*

$$\mathbb{E}\Big[\left\| \widehat{p}_{0,h} - p_0 \right\|^2 \Big] = \begin{cases} O\big(n^{-\frac{a}{a+\gamma}} \big) & \text{if } r = 1 \\ O\Big(\exp\Big\{ -\frac{2a}{(2\gamma)^r} \Big[\log n + (r-1) \log\log n + \sum_{i=0}^{k} b_i^* (\log n)^{r + i(r-1)} \Big]^r \Big\} \Big) & \text{if } r < 1 \\ O\Big(\frac{1}{n} \exp\Big\{ \frac{2\gamma}{(2a)^{1/r}} \Big[\log n + \frac{r-1}{r} \log\log n + \sum_{i=0}^{k} d_i^* (\log n)^{\frac{1}{r} - i\frac{r-1}{r}} \Big]^{1/r} \Big\} \Big) & \text{if } r > 1, \end{cases} \tag{2.109}$$

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 the following triangular system:*

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

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

Remark. For $r = 1$, the choice $h = 2(a+\gamma)/\log(n)$ yields the rate of convergence. The expressions of the optimal bandwidths for $r > 1$ and $r < 1$ are much more intricate (see (2.147) and (2.149) in the proof of Corollary 2.7.3, and also [37]).

Let us comment the rates of convergence obtained in Corollary 2.7.3. 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 (2.55). 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 2.7.3. 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.

The case of ordinary-smooth regularity assumption on p_0

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

$$\widetilde{\mathcal{S}}(\beta, L) = \Big\{ g \text{ density such that } \int_{\mathbb{R}} |g^\star(\xi)|^2 . (1+\xi^2)^\beta d\xi \leq L \Big\}.$$

We have the following classical estimate for the integrated bias (see e.g. [25, Proposition 3]).

Proposition 2.7.4. *For $p_0 \in \widetilde{\mathcal{S}}(\beta, L)$ we have:*

$$\|K_h^\star . p_0^\star - p_0^\star\|^2 \leq L . h^{2\beta}. \tag{2.110}$$

Proof. Since $p_0 \in \widetilde{\mathcal{S}}(\beta, L)$ with $b > 1/2$ and $L > 0$, then, using the Fourier Transform of sinc kernel as $K_h^\star(z) = \mathbb{1}_{\{|z| \leq 1/h\}}$, we can write

$$\begin{aligned}
\|K_h^\star . p_0^\star - p_0^\star\|^2 = \int_{\mathbb{R}} |K_h^\star(z) . p_0^\star(z) - p_0^\star(z)|^2 \, dz &= \int_{\mathbb{R}} \left| (1 - \mathbb{1}_{\{|z| \leq 1/h\}}) . p_0^\star(z) \right|^2 dz \\
&= \int_{\mathbb{R}} (1+z^2)^\beta \, |p_0^\star(z)|^2 \, . (1+z^2)^{-\beta} . \mathbb{1}_{\{|z| > 1/h\}} dz \\
&\leq h^{2\beta} . \int_{\mathbb{R}} (1+z^2)^\beta \, |p_0^\star(z)|^2 \, dz \leq h^{2\beta} . L.
\end{aligned}$$

□

Then, using Theorem 2.7.2, we obtain the following result.

Corollary 2.7.5. *Suppose that μ_0 satisfies Assumption (2.15) and the density p_0 belongs to the space $\widetilde{\mathcal{S}}(\beta, L)$ 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:*

$$\mathbb{E}\Big[\, \|\widehat{p}_{0,h} - p_0\|^2 \, \Big] = O\Big((\log n)^{-2\beta} \Big). \tag{2.111}$$

Now, let us discuss the optimality of the convergence rates stated in Corollaries 2.7.3 and 2.7.5. 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 [16] for

supersmooth densities or in Fan and Koo [31] for Sobolev-ordinary type regularities. Here, our estimation strategy converts the initial free deconvolution problem into the deconvolution problem (2.55) 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 [16] and [31]. If our observations had been distributed according to the density $f_{\mu_0 \star \mathcal{C}_\gamma}$ as in [16] and [31], for a given γ, the upper bound of the variance term given by Theorem 2.7.2 as well as the bounds for the bias given by Proposition 2.7.1 and Proposition 2.7.4 would have been optimal. Consequently, as part of our strategy, we expect that our rates of convergence cannot be improved for a given γ.

2.7.2 Proof of Theorem 2.7.2

Recall that by Lemma 2.4.3, we have $\mathrm{Im}\left(w_{fp}(z)\right) = t.\mathrm{Im}\big(G_{\mu_t}\big(w_{fp}(z)\big)\big) + \mathrm{Im}(z)$, and similarly by Theorem 2.5.1, $\mathrm{Im}\big(\widehat{w}^n_{fp}(z)\big) = t.\mathrm{Im}\big(\widehat{G}_{\mu^n_t}\big(\widehat{w}^n_{fp}(z)\big)\big) + \mathrm{Im}(z)$ for $z \in \mathbb{C}_{2\sqrt{t}}$. We have

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

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

$$\mathbb{E}(\Sigma) \leq \frac{2C_K^2}{\pi}.e^{\frac{2\gamma}{h}}.\big(I^\kappa + J^\kappa\big) \tag{2.112}$$

where

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

and

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

Before getting into the proof of this lemma, we recall Lemma 4.3.17 of [1], with a null initial condition, which will be useful in the sequel:

Lemma 2.7.6. *Let $(\eta^n_1(t), \ldots, \eta^n_n(t))$ be the eigenvalues of $H_n(t)$. With a large probability, all the eigenvalues $(\eta^n_j(t))$ of $H_n(t)$ belong to a ball of radius $M > 0$ independent of n and t. Introduce*

$$A^{n,t}_M := \left\{\forall 1 \leq j \leq n : \left|\eta^n_j(t)\right| \leq M\right\}. \tag{2.113}$$

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

$$\mathbb{P}\left(\{\eta^n_*(t) > M\}\right) \leq e^{-n.C_{eig}.M}, \tag{2.114}$$

with $\eta^n_*(t) := \max_{j=1,\ldots,n} |\eta^n_j(t)|$.

We will now rely on Lemma 2.7.6 to control the tail distribution of $\mathbb{E}[\mu^n_t]$. We recall that $\lambda^n_1(t) \leq \ldots \leq \lambda^n_n(t)$ are the eigenvalues of $X_n(t) = X_n(0) + H_n(t)$ in non-decreasing order.

Moreover, if we denote by $\eta_*^n(t) := \max_{i=1,\dots,n} |\eta_i^n(t)|$, where $\eta_1^n(t) \leq \dots \leq \eta_n^n(t)$ are the eigenvalues of $H_n(t)$, by Weyl's interlacing inequalities, we have that, for $1 \leq j \leq n$,

$$\lambda_j^n(0) - \eta_*^n(t) \leq \lambda_j^n(t) \leq \lambda_j^n(0) + \eta_*^n(t). \tag{2.115}$$

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

$$\begin{aligned}
\mathbb{E}\left[\mu_t^n\left(\left\{|\lambda| > \frac{\kappa}{2}\right\}\right)\right] = \mathbb{E}\Big[\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\{|\lambda_j^n(t)|>\frac{\kappa}{2}\}}\Big] &\leq \mathbb{E}\Big[\frac{1}{n}\sum_{j=1}^n \left(\mathbb{1}_{\{|\lambda_j^n(0)|>\frac{\kappa}{4}\}} + \mathbb{1}_{\{|\eta_*^n(t)|>\frac{\kappa}{4}\}}\right)\Big] \\
&= \mathbb{E}\left[\mu_0^n\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right)\right] + \mathbb{P}\left(\left\{\eta_*^n(t) > \frac{\kappa}{4}\right\}\right) \\
&\leq \mathbb{E}\left[\mu_0^n\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right)\right] + e^{-\frac{n.C_{eig}.\kappa}{4}}.
\end{aligned}$$

Now if we denote by $d_1^n, \dots, d_n^n$ the diagonal elements of $X_n(0)$ (so that $\lambda_1^n(0) \leq \dots \leq \lambda_n^n(0)$ is the order statistics of $d_1^n, \dots, d_n^n$), as $d_1^n, \dots, d_n^n$ are i.i.d. with law μ_0, we get

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

so that we finally obtain

$$\mathbb{E}\left[\mu_t^n\left(\left\{|\lambda| > \frac{\kappa}{2}\right\}\right)\right] \leq \mu_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right) + e^{-\frac{n.C_{eig}.\kappa}{4}}. \tag{2.116}$$

Upper bound for I^κ

Lemma 2.7.7. *Let us consider $\gamma > 2\sqrt{t}$. There exist constants $C_{I,1}, C_{I,2}$ and $C_{I,3}$ (only depending on M and t) such that:*

$$I^\kappa \leq \frac{\gamma^8}{(\gamma^2 - 4t)^4}.\frac{C_{I,1}}{n} + \frac{\kappa C_{I,2}}{n^2} + C_{I,3}.\kappa.e^{-n.C_{eig}.M}. \tag{2.117}$$

Before proving Lemma 2.7.7, let us establish a result that will be useful in the sequel.

Lemma 2.7.8. *Let us consider $\gamma > 2\sqrt{t}, p > 1$ and $M > 0$. Then, we have*

$$\mathfrak{I}_{p,\gamma,M,t} := \int_0^{+\infty} \int \frac{1}{\left[\left\{\Big||\lambda| - x\Big| - \sqrt{t} - M\right\} \vee \frac{\gamma}{2}\right]^p} d\mu_0(\lambda)dx \leq C(p, M, t), \tag{2.118}$$

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

Proof. The supremum in the denominator equals to $\big||\lambda| - x\big| - \sqrt{t} - M$ when $x < |\lambda| - \sqrt{t} - M - \gamma/2$ (which is possible only if $|\lambda| - \sqrt{t} - M - \gamma/2$ is positive) or when $x > |\lambda| + \sqrt{t} + M + \gamma/2$. Otherwise, the supremum takes $\gamma/2$. Hence,

$$\begin{aligned}
\mathfrak{I}_{p,\gamma,M,t} \leq \int_{\mathbb{R}} \Bigg[&\int_0^{\left(|\lambda| - \sqrt{t} - M - \gamma/2\right) \vee 0} \frac{1}{\left(|\lambda| - x - \sqrt{t} - M\right)^p} dx + \int_{\left(|\lambda| - \sqrt{t} - M - \gamma/2\right) \vee 0}^{|\lambda| + \sqrt{t} + M + \gamma/2} \frac{2^p}{\gamma^p} dx \\
&+ \int_{|\lambda| + \sqrt{t} + M + \gamma/2}^{+\infty} \frac{1}{\left(x - |\lambda| - \sqrt{t} - M\right)^p} dx \Bigg] d\mu_0(\lambda) \\
\leq \int_{\mathbb{R}} \Bigg\{ &\int_{\left(|\lambda| - \sqrt{t} - M\right) \wedge \gamma/2}^{\left(|\lambda| - \sqrt{t} - M\right)} \frac{1}{v^p} dv + \frac{2^p\left(2\sqrt{t} + 2M + \gamma\right)}{\gamma^p} + \int_{\gamma/2}^{+\infty} \frac{1}{v^p} dv \Bigg\} d\mu_0(\lambda) \\
\leq C(p, M, t) < +\infty,
\end{aligned} \tag{2.119}$$

since $\gamma > 2\sqrt{t}$. This concludes the proof of Lemma 2.7.8. □

Proof of Lemma 2.7.7. The idea is to use the examinations for fluctuations of Cauchy transform. We decompose $I^\kappa \leq 3(I_1^\kappa + I_2^\kappa + I_3^\kappa)$ where we define

$$I_1^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(\widehat{w}_{fp}^n(x+i\gamma)\big) - \widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|^2\Big]dx,$$
$$I_2^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big) - \mathbb{E}\Big[\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big]\Big|^2\Big]dx,$$
$$I_3^\kappa := \int_{\{|x|\leq\kappa\}} \Big|\mathbb{E}\Big[\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big] - G_{\mu_t}\big(w_{fp}(x+i\gamma)\big)\Big|^2 dx.$$

■ Let us first give an upper bound for I_1^κ. We have

$$I_1^\kappa = \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\frac{1}{n}\sum_{j=1}^n \frac{w_{fp}(x+i\gamma) - \widehat{w}_{fp}^n(x+i\gamma)}{\big(\widehat{w}_{fp}^n(x+i\gamma) - \lambda_j^n(t)\big).\big(w_{fp}(x+i\gamma) - \lambda_j^n(t)\big)}\Big|^2\Big]dx$$
$$\leq \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\big|\widehat{w}_{fp}^n(x+i\gamma) - w_{fp}(x+i\gamma)\big|^2.\frac{1}{n}\sum_{j=1}^n \frac{1}{\big|\widehat{w}_{fp}^n(x+i\gamma) - \lambda_j^n(t)\big|^2.\big|w_{fp}(x+i\gamma) - \lambda_j^n(t)\big|^2}\Big]dx$$

by Cauchy-Schwarz's inequality. Moreover, $\mathrm{Im}\big(w_{fp}(x+i\gamma)\big) \geq \gamma/2$ by Theorem 2.4.2 (i), and a similar estimate for $\widehat{w}_{fp}^n(x+i\gamma)$ by Theorem 2.5.1, so we can write

$$I_1^\kappa \leq \frac{16}{\gamma^4}.\int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\big|\widehat{w}_{fp}^n(x+i\gamma) - w_{fp}(x+i\gamma)\big|^2\Big]dx$$
$$\leq \frac{16}{n.\gamma^4}.\int_{\{|x|\leq\kappa\}} \sup_{n\in\mathbb{N}}\sup_{z\in\mathbb{C}_\gamma} \mathbb{E}\Big[\Big|\sqrt{n}\big(\widehat{w}_{fp}^n(z) - w_{fp}(z)\big)\Big|^2\Big]dx \leq \frac{32\kappa C_{w_{fp}}}{n.\gamma^4} \leq \frac{2\kappa C_{w_{fp}}}{n.t^2}, \quad (2.120)$$

where the constant $C_{w_{fp}}$ comes from Proposition 2.6.1 -(iii). This gives an upper bound for I_1^κ, but will now 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 2.7.6. Thus, we decompose $I_1^\kappa = I_{11}^\kappa + I_{12}^\kappa$, with

$$I_{11}^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(\widehat{w}_{fp}^n(x+i\gamma)\big) - \widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|^2.\mathbb{1}_{A_M^{n,t}}\Big]dx,$$
$$I_{12}^\kappa := \int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big|\widehat{G}_{\mu_t^n}\big(\widehat{w}_{fp}^n(x+i\gamma)\big) - \widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|^2.\mathbb{1}_{(A_M^{n,t})^c}\Big]dx.$$

For term I_{12}^κ, we have by Theorem 2.4.2-(i) and Lemma 2.7.6 :

$$I_{12}^\kappa \leq 2.\int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big(\Big|\widehat{G}_{\mu_t^n}\big(\widehat{w}_{fp}^n(x+i\gamma)\big)\Big|^2 + \Big|\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|^2\Big).\mathbb{1}_{(A_M^{n,t})^c}\Big]dx$$
$$\leq 2.\int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\Big(\frac{1}{\big|\mathrm{Im}\widehat{w}_{fp}^n(x+i\gamma)\big|^2} + \frac{1}{\big|\mathrm{Im}w_{fp}(x+i\gamma)\big|^2}\Big).\mathbb{1}_{(A_M^{n,t})^c}\Big]dx$$
$$\leq 2.2\kappa.\Big(\frac{4}{\gamma^2} + \frac{4}{\gamma^2}\Big).\mathbb{P}\Big(\big(A_M^{n,t}\big)^c\Big) \leq \frac{32}{\gamma^2}.\kappa.e^{-n.C_{eig}.M}. \quad (2.121)$$

Moreover, for the term I_{11}^κ:

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

by Cauchy-Schwartz's inequality. Using $|w_{fp}(z)-z| \leq \sqrt{t}$ and (2.115) we have

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

We also can write

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

Therefore, using Theorem 2.4.2 as $\big|w_{fp}(x+i\gamma)\big| \geq \mathrm{Im}\big(w_{fp}(x+i\gamma)\big) \geq \gamma/2$, to obtain

$$\big|w_{fp}(x+i\gamma)-\lambda_j^n(t)\big| \geq \Big\{\Big|\big|\lambda_j^n(0)\big|-|x|\Big| - \sqrt{t} - \eta_*^n(t)\Big\} \vee \frac{\gamma}{2}. \tag{2.122}$$

Similarly, we also get

$$\big|\widehat{w}_{fp}^n(x+i\gamma)-\lambda_j^n(t)\big| \geq \Big\{\Big|\big|\lambda_j^n(0)\big|-|x|\Big| - \sqrt{t} - \eta_*^n(t)\Big\} \vee \frac{\gamma}{2}. \tag{2.123}$$

On the event $A_M^{n,t}$ we have $\eta_*^n(t) = \max_{j=1,\ldots,n}|\eta_j^n(t)| \leq M$, and using the constant $C(\gamma,t)$ appearing in (2.104), there exists a constant $C(t)$ only depending on t such that

$$\begin{aligned}I_{11}^\kappa &\leq \int_{\{|x|\leq\kappa\}} \mathbb{E}\left[\big|\widehat{w}_{fp}^n(x+i\gamma)-w_{fp}(x+i\gamma)\big|^2 . \frac{1}{n}\sum_{j=1}^n \frac{\mathbb{1}_{A_M^{n,t}}}{\Big[\Big\{\Big|\big|\lambda_j^n(0)\big|-|x|\Big| - \sqrt{t} - \eta_*^n(t)\Big\} \vee \frac{\gamma}{2}\Big]^4}\right] dx \\ &\leq \frac{1}{n}.\sum_{j=1}^n \int_{\{|x|\leq\kappa\}} \mathbb{E}\left[\frac{\mathbb{1}_{A_M^{n,t}}}{\Big[\Big\{\Big|\big|\lambda_j^n(0)\big|-|x|\Big| - \sqrt{t} - M\Big\} \vee \frac{\gamma}{2}\Big]^4}.\mathbb{E}\Big(\big|\widehat{w}_{fp}^n(x+i\gamma)-w_{fp}(x+i\gamma)\big|^2\Big|X_n(0)\Big)\right] dx \\ &\leq \Big(\frac{t\gamma^2}{\gamma^2-4t}\Big)^2.\frac{C(\gamma,t)}{n}.\sum_{j=1}^n \int_{\{|x|\leq\kappa\}} \mathbb{E}\left[\frac{1}{\Big[\Big\{\Big|\big|\lambda_j^n(0)\big|-|x|\Big| - \sqrt{t} - M\Big\} \vee \frac{\gamma}{2}\Big]^4}.\Big(\frac{1}{n^2}\right. \\ &\qquad \left. + \Big|\int_{\mathbb{R}} \frac{1}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\big[d\mu_0^n(\lambda)-d\mu_0(\lambda)\big]\Big|^2\Big)\right] dx \\ &= \frac{\gamma^8}{(\gamma^2-4t)^4}.\frac{C(t)}{n}.\big(I_{111}^\kappa + I_{112}^\kappa\big),\end{aligned} \tag{2.124}$$

where the third inequality comes from (2.104) combined with (2.69), where the fourth inequality comes from the analysis of the constant $C(\gamma,t)$ led in the study of fluctuations of $A_{n,3}(z)$(see in Section 2.6.2) and where:

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

$$\begin{aligned}I_{112}^\kappa &:= \sum_{j=1}^n \int_{\{|x|\leq\kappa\}} \mathbb{E}\left[\frac{1}{\Big[\Big\{\Big|\big|\lambda_j^n(0)\big|-|x|\Big| - \sqrt{t} - M\Big\} \vee \frac{\gamma}{2}\Big]^4}.\Big|\int_{\mathbb{R}} \frac{\big[d\mu_0^n(\lambda)-d\mu_0(\lambda)\big]}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\Big|^2\right] dx \\ &= \int_{\{|x|\leq\kappa\}} \mathbb{E}\left[\int \frac{d\mu_0^n(\lambda)}{\Big[\Big\{\Big|\big|\lambda\big|-|x|\Big| - \sqrt{t} - M\Big\} \vee \frac{\gamma}{2}\Big]^4}.\Big|\sqrt{n}.\int_{\mathbb{R}} \frac{\big[d\mu_0^n(\lambda)-d\mu_0(\lambda)\big]}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\Big|^2\right] dx.\end{aligned}$$

Now, we intend to find upper bounds for I_{111}^{κ} and I_{112}^{κ}, which depend on κ. For I_{111}^{κ}, we get that

$$\begin{aligned} I_{111}^{\kappa} &= \frac{1}{n}.\int_{\{|x|\leq\kappa\}}\int \frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}d\mu_0(\lambda)dx \\ &\leq \frac{2}{n}.\int_0^{+\infty}\int \frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}d\mu_0(\lambda)dx. \end{aligned} \tag{2.125}$$

This double integral is upper bound by a constant $C(\gamma, M, t)/n$ by Lemma 2.7.8.

For the term I_{112}^{κ}, using Cauchy-Schwarz inequality first for the expectation and then for the integral with respect to x, we have:

$$\begin{aligned} I_{112}^{\kappa} \leq \int_{\{|x|\leq\kappa\}} &\left(\sqrt{\mathbb{E}\left[\int \frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}d\mu_0^n(\lambda)\right]^2}\right. \\ &\left.\times\sqrt{\mathbb{E}\left[\left|\sqrt{n}.\int_{\mathbb{R}}\frac{1}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\left[d\mu_0^n(\lambda)-d\mu_0(\lambda)\right]\right|^4\right]}\right)dx \\ \leq \sqrt{\int_{\{|x|\leq\kappa\}}\mathbb{E}\left[\int \frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}d\mu_0^n(\lambda)\right]^2dx} \\ \times\sqrt{\int_{\{|x|\leq\kappa\}}\mathbb{E}\left[\left|\sqrt{n}.\int_{\mathbb{R}}\frac{1}{x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda}\left[d\mu_0^n(\lambda)-d\mu_0(\lambda)\right]\right|^4\right]dx} \end{aligned} \tag{2.126}$$

The first term can be adapted exactly as I_{111}^{κ} as in the following:

$$\begin{aligned} &\mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\left(\int \frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}d\mu_0^n(\lambda)\right)^2dx\right] \\ &\leq \mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\left(\int\frac{1}{\left(\frac{\gamma}{2}\right)^4}d\mu_0^n(\lambda)\right).\left(\int \frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}d\mu_0^n(\lambda)\right)dx\right] \\ &= \mathbb{E}\left[\int_{\{|x|\leq\kappa\}}\frac{1}{\left(\frac{\gamma}{2}\right)^4}.\left(\int \frac{1}{\left[\left\{\Big||\lambda|-|x|\Big|-\sqrt{t}-M\right\}\vee\frac{\gamma}{2}\right]^4}d\mu_0^n(\lambda)\right)dx\right] = \frac{16n}{\gamma^4}I_{111}^{\kappa}. \end{aligned} \tag{2.127}$$

We now focus on the second term of (2.126). As in the proof of Proposition 2.6.12, if we denote by $\phi_x := \varphi_{x+i\gamma} : \lambda \mapsto (x+i\gamma-t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)-\lambda)^{-1}$, this second term can be written as

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

where we use the notation $d_1^n, ..., d_n^n$ for the non-ordered diagonal elements of $X_n(0)$ (introduced after equation (2.17)). Since the random variables $d_1^n, ..., d_n^n$ are i.i.d. with law μ_0, the random variables $\Big(\phi_x(\lambda_j^n(0)) - \mathbb{E}\Big(\phi_x(\lambda_j^n(0))\Big)\Big)_{1\leq j\leq n}$ are i.i.d. centered with finite moment of order fourth. By Rosenthal and then Cauchy-Schwarz inequality, we have

$$I_{1121}^{\kappa} \leq \frac{C}{n^2}(n+n^2)\int_{\{|x|\leq\kappa\}}\int_{\mathbb{R}}\Big|\phi_x(\lambda) - \int_{\mathbb{R}}\phi_x(\widetilde{\lambda})d\mu_0(\widetilde{\lambda})\Big|^4 d\mu_0(\lambda)dx, \tag{2.128}$$

for a constant C. We can conclude if the above double integral is bounded independently of κ. We would like to use Lemma 2.7.8, but the fact that we have a non-centered moment here. It implies that we should be careful, because a constant integrated with respect to dx on $|x| \leq \kappa$ yields a term proportional to κ that should be pretended.

First let us recall some estimates for the functions ϕ_x. Since $\mathrm{Im}\Big(G_{\mu_0\boxplus\sigma_t}(x+i\gamma)\Big) < 0$, we have

$$\big|x+i\gamma - t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma) - \lambda\big| \geq \mathrm{Im}\big(x+i\gamma - t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma) - \lambda\big) \geq \gamma \geq \gamma/2, \tag{2.129}$$

and so the functions ϕ_x are bounded by $2/\gamma$. Thus, one gets $\big|\int_{\mathbb{R}}\phi_x(\lambda)d\mu_0(\lambda)\big| \leq 2/\gamma$. Moreover, by Lemma 2.4.3, we have $\big|t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma)\big| \leq \sqrt{t}$, so that

$$\big|x+i\gamma - t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma) - \lambda\big| \geq \big(\big|x-\lambda\big| - \sqrt{t}\big) \geq \big(\big||\lambda| - |x|\big| - \sqrt{t}\big). \tag{2.130}$$

As a consequence, we obtain

$$\big|x+i\gamma - t.G_{\mu_0\boxplus\sigma_t}(x+i\gamma) - \lambda\big| \geq \big(\big|x-\lambda\big| - \sqrt{t}\big) \geq \big(\big||\lambda| - |x|\big| - \sqrt{t}\big) \vee \frac{\gamma}{2}. \tag{2.131}$$

Using that d_1^n has distribution μ_0, the double integral in the right hand side of (2.128) can be rewritten as:

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

by using the notation of Lemma 2.7.8 and by neglecting the term $-\big|\mathbb{E}\big[\phi_x(d_1^n)\big]\big|^4 < 0$. The Lemma 2.7.8 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 for the study of I_1^{κ}. This last result, together with (2.128) implies that I_{112}^{κ} is bounded bounded by a constant only depending on t. From (2.124) and (2.125), 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 (2.121), we finally obtain that:

$$I_1^{\kappa} \leq \frac{\gamma^8}{(\gamma^2-4t)^4}.\frac{C_1(M,t)}{n} + \frac{16}{\gamma^2}\kappa.e^{-n.C_{eig}.M}. \tag{2.132}$$

■ For I_2^κ: Using Proposition 2.6.8, we get that

$$I_2^\kappa = \int_{\{|x|\le\kappa\}} \mathbb{E}\Big[\mathrm{Var}\Big(\widehat{G}_{\mu_t^n}\big(w_{fp}(x+i\gamma)\big)\Big|X_n(0)\Big)\Big]dx = \int_{\{|x|\le\kappa\}} \mathbb{E}\Big[\mathrm{Var}\Big(A_{n,1}\big(w_{fp}(x+i\gamma)\big)\Big|X_n(0)\Big)\Big]dx$$
$$\le \int_{\{|x|\le\kappa\}} \frac{10\ t}{n^2 \mathrm{Im}^4\big(w_{fp}(x+i\gamma)\big)}dx \le \frac{10.2^4.t.2\kappa}{n^2\gamma^4} \le \frac{20\kappa}{n^2.t}. \tag{2.133}$$

■ Let us provide finally an upper bound for I_3^κ:

$$I_3^\kappa = \int_{\{|x|\le\kappa\}} \mathbb{E}\Big[\big|A_{n,2}\big(w_{fp}(x+i\gamma)\big) + A_{n,3}\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]dx$$
$$\le 2\int_{\{|x|\le\kappa\}} \mathbb{E}\Big[\big|A_{n,2}\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]dx + 2\int_{\{|x|\le\kappa\}} \mathbb{E}\Big[\big|A_{n,3}\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]dx. \tag{2.134}$$

By using Propositions 2.6.9 together with Theorem 2.4.2-(i) and the fact that $\gamma > 2\sqrt{t}$, we obtain that the first term in the right-hand side is upper-bounded by

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

where c is an absolute constant. Let us now consider the second term in the right-hand side of (2.134). Using the bound of Proposition 2.6.12,

$$2\int_{\{|x|\le\kappa\}} \mathbb{E}\Big[\big|A_{n,3}\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]dx$$
$$\le 2\frac{\gamma^4}{(\gamma^2-4t)^2}\int_{\mathbb{R}} \mathbb{E}\Big[\Big|\int_{\mathbb{R}} \frac{1}{w_{fp}(x+i\gamma) - t.G_{\mu_t}(w_{fp}(x+i\gamma)) - v}\big[d\mu_0^n(v) - d\mu_0(v)\big]\Big|^2\Big] \tag{2.135}$$

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

$$\mathbb{E}\Big[\Big|\int_{\mathbb{R}} \varphi_{w_{fp}(x+i\gamma)}(v)\big[d\mu_0^n(v) - d\mu_0(v)\big]\Big|^2\Big] = \mathrm{Var}\Big[\frac{1}{n}\sum_{j=1}^n \varphi_{w_{fp}(x+i\gamma)}\big(\lambda_j^n(0)\big)\Big]$$
$$\le \frac{1}{n}\mathbb{E}\big[\big|\varphi_{w_{fp}(x+i\gamma)}(d_1^n)\big|^2\big]$$
$$= \frac{1}{n}\int_{\mathbb{R}} \frac{1}{\big|w_{fp}(x+i\gamma) - t.G_{\mu_t}(w_{fp}(x+i\gamma)) - v\big|^2}d\mu_0(v). \tag{2.136}$$

Recall that from Lemma 2.4.3 and Theorem 2.4.2-(i), we have $\big|w_{fp}(x+i\gamma) - t.G_{\mu_t}(w_{fp}(x+i\gamma)) - v\big| \ge \big|\mathrm{Im}\big(w_{fp}(x+i\gamma)\big)\big| \ge \gamma/2$, so that the integrand in the right-hand side of (2.136) is bounded. However, we have to work more to show that it is integrable with respect to x. We have:

$$\big|w_{fp}(x+i\gamma) - t.G_{\mu_t}(w_{fp}(x+i\gamma)) - v\big| \ge \big|\mathrm{Re}\big(w_{fp}(x+i\gamma) - t.G_{\mu_t}(w_{fp}(x+i\gamma))\big) - v\big|$$
$$\ge \big|\mathrm{Re}\big(w_{fp}(x+i\gamma)\big) - v\big| - t.\big|\mathrm{Re}\big(G_{\mu_t}(w_{fp}(x+i\gamma))\big)\big|.$$

By Theorem 2.4.2-(i), we obtain that

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

Also, by using (2.53), we get that $\left|\mathrm{Re}\big(w_{fp}(x+i\gamma)\big)-x\right| \leq \sqrt{t}$. Therefore,

$$\left|w_{fp}(x+i\gamma) - t.G_{\mu_t}(w_{fp}(x+i\gamma)) - v\right| \geq \big||x|-|v|\big| - \sqrt{t} - \frac{2t}{\gamma}. \tag{2.137}$$

From (2.135), (2.136) and (2.137), we have that:

$$\begin{aligned}
&2\int_{\{|x|\leq\kappa\}} \mathbb{E}\Big[\big|A_{n,3}\big(w_{fp}(x+i\gamma)\big)\big|^2\Big]dx \\
&\leq \frac{2\gamma^4}{n(\gamma^2-4t)^2}\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{\Big[\Big\{\big||x|-|v|\big| - \sqrt{t} - \frac{2t}{\gamma}\Big\} \vee \frac{\gamma}{2}\Big]^2} d\mu_0(v)dx \leq \frac{4\gamma^4}{n(\gamma^2-4t)^2}\mathfrak{I}_{2,\gamma,2\sqrt{t},t},
\end{aligned}$$

by Lemma 2.7.8. We conclude as for I_{11}^{κ} and we obtain

$$I_3^{\kappa} \leq \frac{c\kappa}{n^2 t} + \frac{4\gamma^4}{n\big(\gamma^2-4t\big)^2}\mathfrak{I}_{2,\gamma,2\sqrt{t},t}. \tag{2.138}$$

Finally, gathering (2.120), (2.133) and (2.132), we obtain an upper bound for I^{κ} as in Lemma 2.7.7.

□

Upper bound for J^{κ}

Now, we consider the other term, i.e. when $|x|$ is larger than κ. Our goal is to prove the following bound:

Lemma 2.7.9. *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:*

$$J^{\kappa} \leq \frac{C_{J,1}}{\kappa} + C_{J,2}.n.e^{-n.C_{eig}.\kappa/4} + C_{J,3}.\mu_0\left(\Big\{|\lambda| > \frac{\kappa}{4}\Big\}\right). \tag{2.139}$$

Proof of Lemma 2.7.9. We decompose $J^{\kappa} \leq 2(J_1^{\kappa} + J_2^{\kappa})$ where

$$\begin{aligned}
J_1^{\kappa} &:= \int_{\{|x|>\kappa\}} \mathbb{E}\left(\Big|\int_{\mathbb{R}} \frac{d\mu_t^n(\lambda)}{\widehat{w}_{fp}^n(x+i\gamma)-\lambda}\Big|^2\right)dx, \\
J_2^{\kappa} &:= \int_{\{|x|>\kappa\}} \Big|\int_{\mathbb{R}} \frac{d\mu_t(\lambda)}{w_{fp}(x+i\gamma)-\lambda}\Big|^2 dx.
\end{aligned}$$

■ Let us consider the first term J_1^{κ}. Using the estimate of Theorem 2.5.1, we have for all $x \in \mathbb{R}$ that $\left|\mathrm{Re}\big(\widehat{w}_{fp}^n(x+i\gamma)\big)-x\right| \leq \sqrt{t}$, which allows us to prove that

$$(x-\lambda)^2 + \frac{\gamma^2}{4} \leq 3\Big(\big(\mathrm{Re}\big(\widehat{w}_{fp}^n(x+i\gamma)-\lambda\big)\big)^2 + \frac{\gamma^2}{4}\Big), \quad \text{with } \gamma \geq 2\sqrt{t}.$$

Indeed, for the moment set $Re := \mathrm{Re}\widehat{w}^n_{fp}(x+i.\gamma)$, we have for any x and $\lambda \in \mathbb{R}$

$$\begin{aligned}
\frac{(x-\lambda)^2+\frac{1}{4}.\gamma^2}{\big(Re-\lambda\big)^2+\frac{1}{4}.\gamma^2} &= \frac{(x-Re)^2+2.|x-Re|.|Re-\lambda|+(Re-\lambda)^2+\frac{1}{4}.\gamma^2}{\big(Re-\lambda\big)^2+\frac{1}{4}.\gamma^2} \\
&\leq \frac{t}{\big(Re-\lambda\big)^2+\frac{1}{4}.\gamma^2}+2.\sqrt{t}.\frac{|Re-\lambda|}{\big(Re-\lambda\big)^2+\frac{1}{4}.\gamma^2}+1 \quad \Big(\text{using } |Re-x|\leq\sqrt{t}\Big) \\
&\leq \frac{t}{\frac{1}{4}.\gamma^2}+2.\sqrt{t}.\frac{1}{\gamma}+1 \qquad\qquad \Big(\text{by } \big(Re-\lambda\big)^2+\frac{1}{4}.\gamma^2 \geq 2.|Re-\lambda|.\frac{1}{2}.\gamma\Big) \\
&\leq 3. \qquad\qquad \Big(\text{since } \gamma > 2\sqrt{t}\Big)
\end{aligned}$$

This implies that

$$\frac{1}{\big(\mathrm{Re}\widehat{w}^n_{fp}(x+i.\gamma)-\lambda\big)^2+\frac{1}{4}.\gamma^2} \leq 3.\frac{1}{(x-\lambda)^2+\frac{1}{4}.\gamma^2}, \quad \text{for any } x\in\mathbb{R} \text{ and } \lambda\in\mathbb{R}.$$

Thus,

$$\begin{aligned}
J_1^\kappa &\leq \int_{\{|x|>\kappa\}} \mathbb{E}\int_\lambda \frac{d\mu^n_t(\lambda)}{\Big[\mathrm{Re}\big(\widehat{w}^n_{fp}(x+i\gamma)-\lambda\big)\Big]^2+\Big[\mathrm{Im}\big(\widehat{w}^n_{fp}(x+i\gamma)\big)\Big]^2}dx \\
&\leq 3\int_{\{|x|>\kappa\}} \mathbb{E}\bigg[\int_\lambda \frac{d\mu^n_t(\lambda)}{(x-\lambda)^2+\frac{\gamma^2}{4}}\bigg]dx \\
&= 3\,\mathbb{E}\bigg[\int_\lambda\int_{\{|x|>\kappa\}} \frac{1}{(x-\lambda)^2+\frac{\gamma^2}{4}}dx\;\; d\mu^n_t(\lambda)\bigg] \\
&= 3\,\mathbb{E}\bigg[\int_\lambda\bigg(\int_\kappa^{+\infty} \frac{1}{(x-\lambda)^2+\frac{\gamma^2}{4}}dx+\int_{-\infty}^{-\kappa}\frac{1}{(x-\lambda)^2+\frac{\gamma^2}{4}}dx\bigg)\,d\mu^n_t(\lambda)\bigg] \\
&= \frac{6}{\gamma}\,\mathbb{E}\bigg[\int_\lambda\bigg(\pi-\arctan\Big(\frac{2}{\gamma}(\kappa-\lambda)\Big)-\arctan\Big(\frac{2}{\gamma}(\kappa+\lambda)\Big)\bigg)\,d\mu^n_t(\lambda)\bigg] \\
&= \frac{6}{\gamma}\,\mathbb{E}\bigg[\int_\lambda\bigg(\arctan\Big(\frac{4\kappa\gamma}{4\kappa^2-4\lambda^2-\gamma^2}\Big)+\pi\mathbb{1}_{\left\{\lambda^2>\kappa^2-\frac{\gamma^2}{4}\right\}}\bigg)\,d\mu^n_t(\lambda)\bigg],
\end{aligned}$$

where we have used the formula

$$\arctan(a)+\arctan(b)=\arctan\Big(\frac{a+b}{1-ab}\Big)+\pi.\mathbb{1}_{\{ab>1\}}.\left(\mathbb{1}_{\{a,b>0\}}-\mathbb{1}_{\{a,b<0\}}\right).$$

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

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

We therefore get

$$J_1^\kappa \leq \frac{6}{\gamma}\,\mathbb{E}\bigg[\int_\lambda\bigg(\frac{2\gamma}{\kappa}+\frac{\pi}{2}\mathbb{1}_{\left\{\lambda^2>\frac{\kappa^2}{2}-\frac{\gamma^2}{4}\right\}}+\pi\mathbb{1}_{\left\{\lambda^2>\kappa^2-\frac{\gamma^2}{4}\right\}}\bigg)\,d\mu^n_t(\lambda)\bigg].$$

If we assume moreover that $\kappa > \gamma$, then $\left\{\lambda^2 > \kappa^2 - \frac{\gamma^2}{4}\right\} \subset \left\{\lambda^2 > \frac{\kappa^2}{2} - \frac{\gamma^2}{4}\right\} \subset \left\{\lambda^2 > \frac{\kappa^2}{4}\right\}$, and thus this can be simplified as follows:

$$J_1^\kappa \leq 3\left(\frac{4}{\kappa} + \frac{3\pi}{\gamma}\mathbb{E}\left[\mu_t^n\left(\left\{|\lambda| > \frac{\kappa}{2}\right\}\right)\right]\right). \tag{2.140}$$

Hence, by using (2.116), we get that

$$J_1^\kappa \leq 3\left(\frac{4}{\kappa} + \frac{3\pi}{\gamma}.\mu_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right) + \frac{3\pi}{\gamma}.e^{-\frac{n.C_{eig}.\kappa}{4}}\right). \tag{2.141}$$

■ 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 so less details will be given. Using the estimate $|w_{fp}(z) - z| \leq \sqrt{t}$, 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

$$(x-\lambda)^2 + \frac{\gamma^2}{4} \leq 3\Big(\big(\mathrm{Re}\big(w_{fp}(x+i\gamma) - \lambda\big)\big)^2 + \frac{\gamma^2}{4}\Big).$$

Thus,

$$J_2^\kappa \leq 3\int_{\{|x|>\kappa\}}\int_\lambda \frac{d\mu_t(\lambda)}{(x-\lambda)^2 + \frac{\gamma^2}{4}}dx \leq \frac{6}{\gamma}\int_\lambda\left(\frac{2\gamma}{\kappa} + \frac{\pi}{2}\mathbb{1}_{\left\{\lambda^2 > \frac{\kappa^2}{2} - \frac{\gamma^2}{4}\right\}} + \pi\mathbb{1}_{\left\{\lambda^2 > \kappa^2 - \frac{\gamma^2}{4}\right\}}\right)d\mu_t(\lambda).$$

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

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

Moreover, letting n going to infinity in (2.116), by Proposition 2.3.5 along with the 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

$$J_2^\kappa \leq 3\left(\frac{4}{\kappa} + \frac{3\pi}{\gamma}\mu_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right)\right). \tag{2.142}$$

Gathering the upper bounds (2.141) and (2.142), we get that for any $\kappa > \gamma$,

$$J^\kappa \leq 3\left(\frac{8}{\kappa} + \frac{6\pi}{\gamma}.\mu_0\left(\left\{|\lambda| > \frac{\kappa}{4}\right\}\right) + \frac{3\pi}{\gamma}.e^{-\frac{n.C_{eig}.\kappa}{4}}\right). \tag{2.143}$$

This ends the proof.

□

Conclusion

As a result, combining Lemma 2.7.7 and Lemma 2.7.9, we have:

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

We take M the smallest constant only depending on t satisfying conditions of Lemma 2.7.6 and $\kappa = n$ and using Assumption (2.15), we obtain

$$\mu_0\left(\{|\lambda| > n\}\right) \leq Cn^{-1}, \tag{2.144}$$

for some absolute constant C. Then, from (2.112) and previous computations, there exists a constant $C_{var}(t)$ (that depends only on t) such that for n sufficiently large:

$$\mathbb{E}(\Sigma) \leq \frac{\gamma^8}{(\gamma^2 - 4t)^4}.\frac{C_{var}(t).e^{\frac{2\gamma}{h}}}{n}. \tag{2.145}$$

and Theorem 2.7.2 is proved.

Proof of Corollary 2.7.3

Recall that from Proposition 2.7.1 and Theorem 2.7.2, the mean integrated square error is bounded as follows:

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

Minimizing in h amounts to solving the following equation obtained by taking the derivative in the right hand side of (2.108):

$$\psi(h) := \exp\left(\frac{2\gamma}{h} + \frac{2a}{h^r}\right)h^{r-1} = O(n). \tag{2.146}$$

Consequently for the minimizer h_* of (2.146) we get that

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

for some constant $C > 0$. Hence, in view of (2.108), when $r < 1$ the bias dominates the variance and the contrary occurs when $r > 1$. Thus, there are three cases to consider to derive rates of convergence: $r = 1$, $r < 1$ and $r > 1$. To solve the equation (2.146), we follow the steps of Lacour [37].

Case $r = 1$.

The case where $r = 1$ provides a bandwidth $h_* = 2(a + \gamma)/\log n$ and we get

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

Case $r < 1$.

In this case, and in the case $r > 1$, following the ideas in [37], we will look for the bandwidth h expressed as an expansion in $\log(n)$. In this expansion and when $r < 1$, the integer k such that $\frac{k}{k+1} < r \leq \frac{k+1}{k+2}$ will play a role. The optimal bandwidth is of the form:

$$h_* = 2\gamma\Big(\log(n) + (r-1)\log\log(n) + \sum_{i=0}^{k} b_i(\log n)^{r+i(r-1)}\Big)^{-1}, \tag{2.147}$$

where the coefficients b_i's are a sequence of real numbers chosen so that $\psi(h_*) = O(n)$. The heuristic of this expansion is as follows: the first term corresponds to the solution of $e^{2\gamma/h} = n$. The second term is added to compensate the factor h^{r-1} in (2.146) evaluated with the previous bandwidth, and the third term aims at compensating the factor e^{2a/h^r}. Notice that $r-1<0$ and that the definition of k implies that $r > r+(r-1) > \cdots > r+k(r-1) > 0 > r+(k+1)(r-1)$. This explains the range of the index i in the sum of the right hand side of (2.147).

Plugging (2.147) into (2.146),

$$\begin{aligned}
\psi(h_*) =& n\big(\log n\big)^{r-1} \exp\Big(\sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)}\Big) \\
&\times \exp\Big(\frac{2a}{(2\gamma)^r}\big(\log n\big)^r \big(1 + \frac{(r-1)\log\log(n) + \sum_{i=0}^k b_i (\log n)^{r+i(r-1)}}{\log n}\big)^r\Big) \\
&\times (2\gamma)^{r-1}\big(\log n\big)^{-(r-1)} \Big(1 + \frac{(r-1)\log\log(n) + \sum_{i=0}^k b_i (\log n)^{r+i(r-1)}}{\log n}\Big)^{-(r-1)} \\
=& (2\gamma)^{r-1} n (1+v_n)^{1-r} \exp\Big(\sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)}\Big) \\
&\times \exp\Big(\frac{2a}{(2\gamma)^r}\big(\log n\big)^r \Big[1 + \sum_{j=0}^{k} \frac{r(r-1)\cdots(r-j)}{(j+1)!} v_n^{j+1} + o(v_n^{k+1})\Big]\Big)
\end{aligned}$$

where

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

converges to zero when $n \to +\infty$. We note that

$$\begin{aligned}
v_n^{j+1} &= \sum_{i=0}^{k-j-1} \sum_{p_0+\cdots p_j = i} b_{p_0}\cdots b_{p_j} (\log n)^{(i+j+1)(r-1)} + O\Big(\big(\log n\big)^{(k+1)(r-1)}\Big) \\
&= \sum_{\ell=j+1}^{k} \sum_{p_0+\cdots p_j = \ell-j-1} b_{p_0}\cdots b_{p_j} (\log n)^{\ell(r-1)} + O\Big(\big(\log n\big)^{(k+1)(r-1)}\Big).
\end{aligned}$$

So

$$\begin{aligned}
\psi(h_*) =& (2\gamma)^{r-1} n (1+v_n)^{1-r} \exp\Big(\sum_{i=0}^{k} b_i (\log n)^{r+i(r-1)}\Big) \\
&\times \exp\Big\{\frac{2a}{(2\gamma)^r}(\log n)^r + \frac{2a}{(2\gamma)^r}\sum_{\ell=1}^{k}\sum_{j=0}^{\ell-1}\Big[\frac{r(r-1)\cdots(r-j)}{(j+1)!} \sum_{p_0+\cdots p_j=\ell-j-1} b_{p_0}\cdots b_{p_j}\Big](\log n)^{r+\ell(r-1)} \\
&+ O\Big(\big(\log n\big)^{(k+1)(r-1)}\Big)\Big\} \\
=& (2\gamma)^{r-1} n (1+v_n)^{1-r} \exp\Big(\sum_{i=0}^{k} M_i (\log n)^{i(r-1)+r} + o(1)\Big).
\end{aligned}$$

The condition $\psi(h_*) = O(n)$ implies the following choices of constants M_i's:

$$M_0 = b_0 + \frac{2a}{(2\gamma)^r}, \qquad \forall i > 0,\ M_i = b_i + \frac{2a}{(2\gamma)^r}\sum_{j=0}^{i-1} \frac{r(r-1)\cdots(r-j)}{(j+1)!} \sum_{p_0+\cdots p_j = i-j-1} b_{p_0}\cdots b_{p_j}.$$

Since h_* solves (2.146) if all the $M_i = 0$ for $i \in \{0, \cdots k\}$, the above system provides equation by equation the proper coefficients b_i^*.

$$b_0^* = -\frac{2a}{(2\gamma)^r}, \qquad b_i^* = -\frac{2a}{(2\gamma)^r} \sum_{j=0}^{i-1} \frac{r(r-1)\cdots(r-j)}{(j+1)!} \sum_{p_0+\cdots p_j = i-j-1} b_{p_0}^* \cdots b_{p_j}^*. \tag{2.148}$$

Replacing in (2.108), we get:

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

Case $r > 1$.

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

$$h_*^r = 2a \Big(\log n + \frac{r-1}{r} \log\log(n) + \sum_{i=0}^{k} d_i (\log n)^{\frac{1}{r} - i\frac{r-1}{r}} \Big)^{-1}, \tag{2.149}$$

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

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

$$\begin{aligned}
\psi(h_*) =& (2a)^{\frac{r-1}{r}} n (1+v_n)^{-\frac{r-1}{r}} \times \exp\Big(\sum_{i=0}^{k} d_i (\log n)^{\frac{1}{r} - i\frac{r-1}{r}} \Big) \\
& \times \exp\Big(\frac{2\gamma}{(2a)^{1/r}} (\log n)^{1/r} \Big[1 + \\
& \sum_{\ell=1}^{k} \sum_{j=0}^{\ell-1} \sum_{p_0+\cdots p_j = \ell-j-1} \frac{\frac{1}{r}\big(\frac{1}{r}-1\big)\cdots\big(\frac{1}{r}-j\big)}{(j+1)!} d_{p_0}\cdots d_{p_j} (\log n)^{\ell\frac{1-r}{r}} + O\big((\log n)^{k\frac{1-r}{r}}\big) \Big] \Big) \\
=& (2a)^{\frac{r-1}{r}} n (1+v_n)^{-\frac{r-1}{r}} \exp\Big(\sum_{i=0}^{k} M_i (\log n)^{\frac{1}{r} - i\frac{r-1}{r}} + o(1) \Big)
\end{aligned}$$

where here

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

and

$$M_0 = d_0 + \frac{2\gamma}{(2a)^{1/r}}, \qquad \forall i > 0,\ M_i = d_i + \frac{2\gamma}{(2a)^{1/r}} \sum_{j=0}^{i-1} \sum_{p_0+\cdots p_j = i-j-1} \frac{\frac{1}{r}\big(\frac{1}{r}-1\big)\cdots\big(\frac{1}{r}-j\big)}{(j+1)!} d_{p_0}\cdots d_{p_j} \tag{2.150}$$

Solving $M_0 = \cdots = M_k = 0$ provides the coefficients d_i^* so that (2.146) is satisfied.

Plugging the bandwidth h_* with the coefficients d_i^* into (2.108), we obtain:

$$\text{MISE} = O\Big(\frac{1}{n} \exp\Big\{ \frac{2\gamma}{(2a)^{1/r}} \Big[\log n + \frac{r-1}{r} \log\log n + \sum_{i=0}^{k} d_i^* (\log n)^{\frac{1}{r} - i\frac{r-1}{r}} \Big]^{1/r} \Big\} \Big).$$

This concludes the proof of Corollary 2.7.3.

2.8 Numerical simulations

In this section, we conduct a simulation study to assess the performances of our estimator $\widehat{p}_{0,h}$ designed in Definition 2.5.3 based on the n-sample $\lambda^n(t) := \{\lambda_1^n(t), ..., \lambda_n^n(t)\}$ of (non ordered) eigenvalues. We consider the sample size $n = 4000$ and the fixed time value $t = 1$.
Expression (2.64) of Fourier transform $\widehat{p}^{\star}_{0,h}$ is now used with the kernel $K(x) = \mathrm{sinc}(x) = \sin(x)/(\pi x)$, and a 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 $\widehat{p}_{0,h}(x)$ takes complex values by using these approximations. This is the reason why the density p_0 is estimated with $\mathrm{Re}(\widehat{p}_{0,h})$, the real part of $\widehat{p}_{0,h}$.

The theoretical bandwidth h proposed in Section 2.7 cannot be used in practice due to some unknown parameters and thus, we suggest the following data-driven selection rule, inspired from the principle of cross-validation. We decompose the quadratic risk for $\mathrm{Re}(\widehat{p}_{0,h})$ as follows:

$$\|\mathrm{Re}(\widehat{p}_{0,h}) - p_0\|^2 = \int_{\mathbb{R}} |\mathrm{Re}(\widehat{p}_{0,h}(x)) - p_0(x)|^2 \, dx = \|\mathrm{Re}(\widehat{p}_{0,h})\|^2 - 2\int_{\mathbb{R}} \mathrm{Re}(\widehat{p}_{0,h}(x))p_0(x)dx + \|p_0\|^2 .$$

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

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

Nevertheless, one can see that J depends on p_0 through the second term, so we have to investigate a good estimate of this criterion. For this purpose, we divide the sample $\lambda^n(t)$ into two disjoints sets

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

There are $V_{\max} := \binom{n}{n/2}$ possibilities to select the subsets (E, E^c), which becomes huge as n increases. Hence, to reduce computational time, we draw randomly $V = 10$ partitions denoted $(E_j, E_j^c)_{j=1,...,V}$.
In general, we define a discretization of $\mathcal{H}$ as $\mathcal{H} := \{h_0 + \Delta_h.(k-1)\}_{k=1,N_h}$ with step-size Δ_h (i.e., discretizing $\mathcal{H}$ by N_h points). Choosing the grid $\mathcal{H}$ of $N_h = 50$ equispaced points lying between $h_{\min} = 0.25$ and $h_{\max} = 2.7$, our selected bandwidth is

$$\widehat{h} = \underset{h\in\mathcal{H}}{\mathrm{argmin}}\ \mathrm{Crit}(h) \tag{2.151}$$

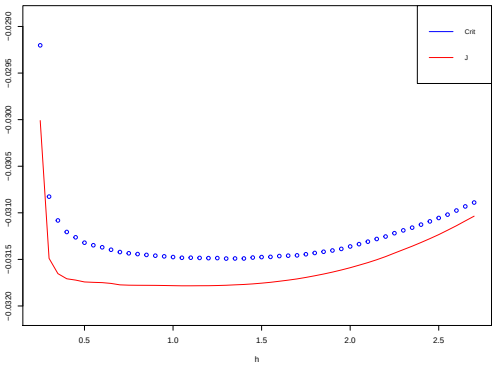
$$\begin{aligned} \text{with} \qquad \mathrm{Crit}(h) &:= \min_{h'\in\mathcal{H}, h'\neq h} \frac{1}{V}\sum_{j=1}^{V} R_j(h,h') \\ &= \min_{h'\in\mathcal{H}, h'\neq h} \frac{1}{V}\sum_{j=1}^{V} \left(\left\| \mathrm{Re}(\widehat{p}_{0,h}^{(E_j)}) \right\|^2 - 2\int_{\mathbb{R}} \mathrm{Re}(\widehat{p}_{0,h}^{(E_j)}(x))\mathrm{Re}(\widehat{p}_{0,h'}^{(E_j^c)}(x))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

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 sub-samples E_j and E_j^c respectively.

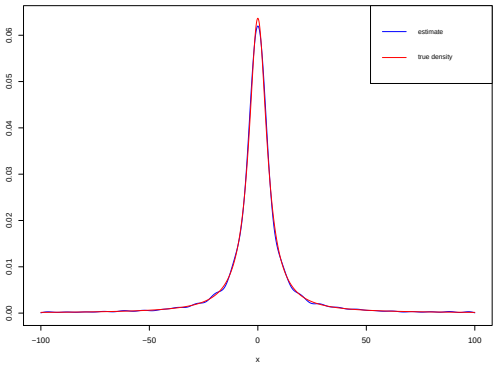
Remark. Note that in the paper [40], we only present the numerical simulations for the case in which μ_0 follows a Cauchy distribution and another case where μ_0 follows a Gaussian mixture distribution. While we also implement here the cases of compact support, for instance, the case of uniform distribution and the case of beta distribution.

$\mu_0 \sim$ **Cauchy(0,d)**

In this first part of numerical experiments, for fixed time $t = 1$ and $\gamma = 2\sqrt{t} + 0.01$, we consider $\mu_0 \sim$ Cauchy(0,sd) with parameter $d = 5$. The exact density function of μ_0 then is $p_0(x) = \frac{1}{\pi}.\frac{d}{(d^2+x^2)}$ with $x \in \mathbb{R}$, and the Fourier Transform of p_0 is given by $p_0^*(z) = e^{-d.|z|}$.
The simulation is implemented for $n = 4000$ observations, on the interval $x \in [-100, 100]$.
To evaluate our approach, Figure 2.1a displays the plot of $h \in \mathcal{H} \mapsto \text{Crit}(h)$ and $h \in \mathcal{H} \mapsto J(h)$ for the Cauchy density p_0. A close inspection of the graphs shows that the first criterion is



(a) Plots of $h \mapsto \text{Crit}(h)$ and $h \mapsto J(h)$ for the Cauchy density p_0



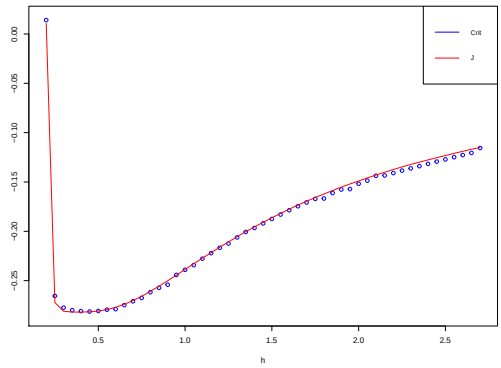
(b) Estimation of p_0

Figure 2.1: Numerical results in case $\mu_0 \sim \text{Cauchy}(0, 5)$.

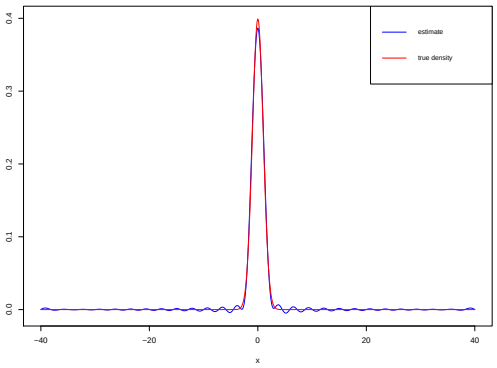
a good estimate of the second one. As expected, for both criterions, we observe a plateau containing minimizers of J and Crit. Outside the plateau, both criterions take large values due to large variance when h is too small and also due to large bias when h is too large.
Furthermore, Figure 2.1b gives the reconstruction provided by $\text{Re}(\widehat{p}_{0,\hat{h}})$ for the Cauchy 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.

$\mu_0 \sim$ **Gaussian(0,1)**

In this second part of numerical experiments, similarly for $t = 1$ and $\gamma = 2\sqrt{t} + 0.01$, we consider $\mu_0 \sim \text{Gaussian}(0, 1)$. The simulation is now implemented for $n = 4000$ observations, on $x \in [-40, 40]$. Actually, in this case, we have to adjust slightly the discretization $\mathcal{H}$ by choosing a smaller beginning value $h_{min} = 0.2$. This adjustment is due to the fact that we would like to observe more evidently that the criterion $J(h)$ and $Crit(h)$ possess minimizers for $h \in \mathcal{H}$.
Figure 2.2a shows that the criterion J works correctly and Figure 2.2b illustrates that the difference between true density p_0 and the estimate $\text{Re}(\widehat{p}_{0,\hat{h}})$ is small in this simulation for the Gaussian case.



(a) Plots of $h \mapsto \text{Crit}(h)$ and $h \mapsto J(h)$ for Gaussian density p_0

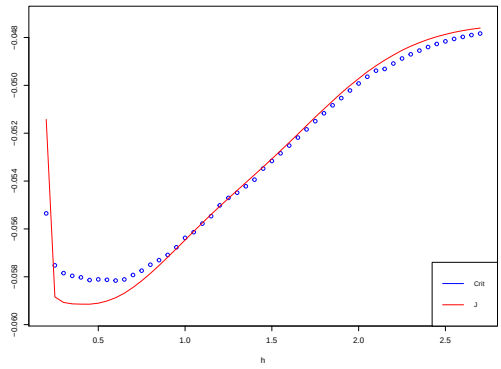


(b) Estimation of p_0

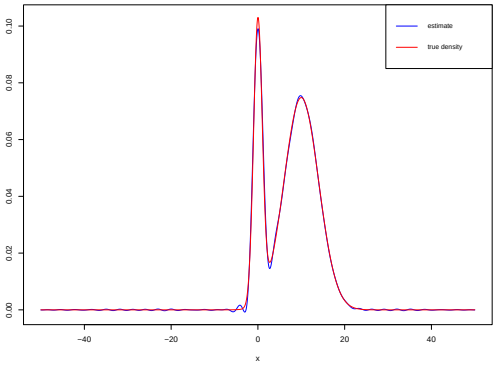
Figure 2.2: Numerical results in case $\mu_0 \sim \text{Gaussian}(0, 1)$.

$\mu_0 \sim$ a Gaussian mixture

We make another numerical experiment with the number of observations $n = 4000$ at time $t = 1$ for the case where $\mu_0 \sim Y \times \mathcal{N}(0, 1) + (1 - Y) \times \mathcal{N}(10, 4)$, with random variable $Y \sim \text{Bernoulli}(0.25)$. Similarly, to evaluate our approach, Figure 2.3a displays the plot of $h \in \mathcal{H} \mapsto \text{Crit}(h)$ and $h \in \mathcal{H} \mapsto J(h)$ for the Gaussian mixture density p_0. It shows that



(a) Plots of $h \mapsto \text{Crit}(h)$ and $h \mapsto J(h)$ for the Gaussian mixture density p_0



(b) Estimation of p_0

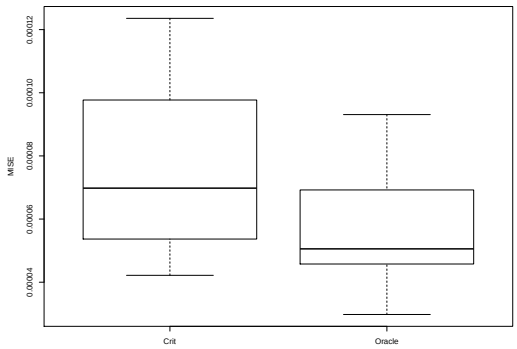
Figure 2.3: Numerical results in case $\mu_0 \sim Y \times \mathcal{N}(0, 1) + (1 - Y) \times \mathcal{N}(10, 4)$ with $Y \sim \text{Bernoulli}(0.25)$.

the first criterion estimates quite precisely the second one. Again, 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. Moreover, Figure 2.3b illustrates the reconstruction provided by $\text{Re}(\widehat{p}_{0,\hat{h}})$ for the Gaussian mixture density p_0.

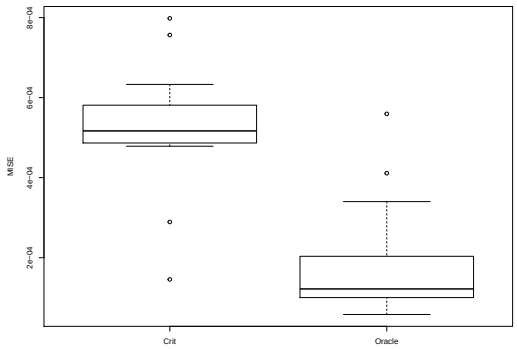
Monte Carlo resampling

Furthermore, we carry out Monte Carlo simulations with 20 runs resampling for both cases of experiment. Figure 2.4a, 2.4b and 2.4c present Boxplots corresponding to Cauchy(0,5), Gaussian(0,1) and Gaussian mixture distribution respectively. The oracle bandwidth corresponds

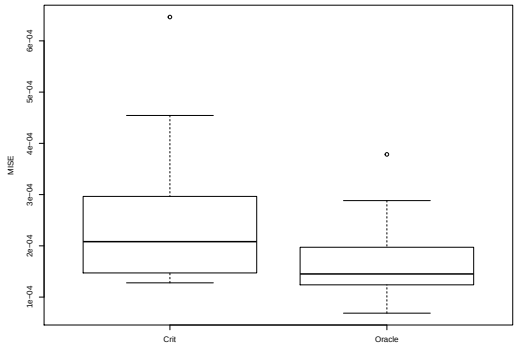
to a minimizer of $h \mapsto \mathbb{E}\big[\, \|\widehat{p}_{0,h} - p_0\|^2 \,\big]$.



(a) $\mu_0 \sim$ Cauchy(0,5)



(b) $\mu_0 \sim$ Gaussian(0,1)



(c) $\mu_0 \sim$ Gaussian mixture

Figure 2.4: Boxplots for 20 runs of Monte Carlo resampling.

Measures with compact support

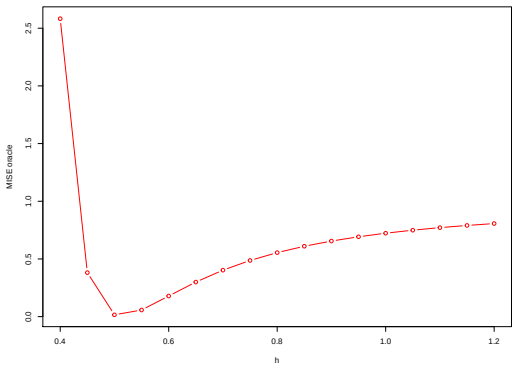
In this subsection, we present illustrations when the initial condition distribution μ_0 has compact support. Such probability densities belong to the Sobolev space $\widetilde{\mathcal{S}}(\beta, L)$, the class of ordinary-smooth densities. In view of comments following Corollary 2.7.5, we expect kernel estimators not to achieve good performances to estimate such densities. As the oracle is expected to perform better than any kernel estimator, we have decided to implement our graphic reconstructions for the oracle.

More specifically, we make implementations first for Uniform([0,1]) distribution (see Figures 2.5b and 2.6b) and then Beta(2,2) (see Figures 2.7b and 2.8b). As previously, we make simulations with $n = 4000$ at fixed time $t = 1$.
We first assume that the support of p_0 is known, so we compute the oracle reconstruction on the interval $[0, 1]$. Then, we compute it on the interval $[-5, 6]$.

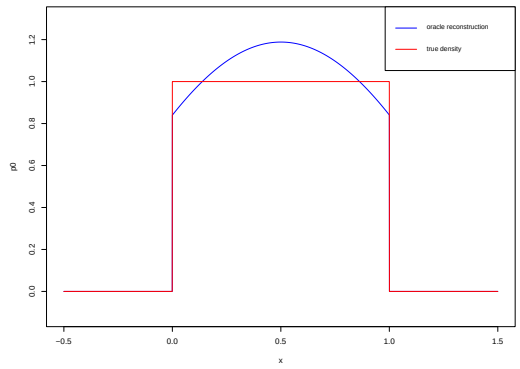
At inspection of Figures 2.5b, 2.6b, 2.7b and 2.8b below, the oracle reconstructions of p_0 are not satisfying as one might have expected.

$\mu_0 \sim$ **Uniform([0,1])**

Figure 2.5b and Figure 2.6b illustrate the plots of reconstructions of $\mathrm{Re}(\widehat{p}_{0,h_*})$ on $\mathrm{supp}(\mu_0)$ and on the interval $[-5, 6]$, respectively. The optimal bandwidth h_* is chosen as minimizer of the oracle MISE (see Figure 2.5a and Figure 2.6a).

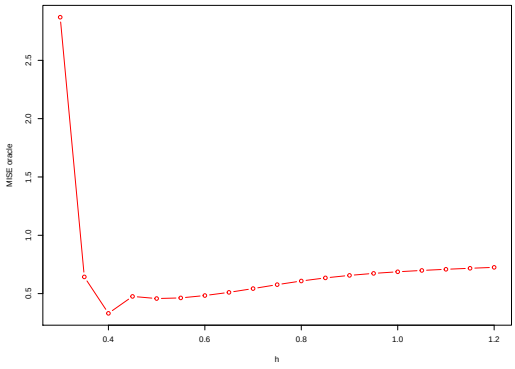


(a) Plots of $h \mapsto \mathrm{MISE}_{\mathrm{oracle}}(h)$ for the Uniform density p_0

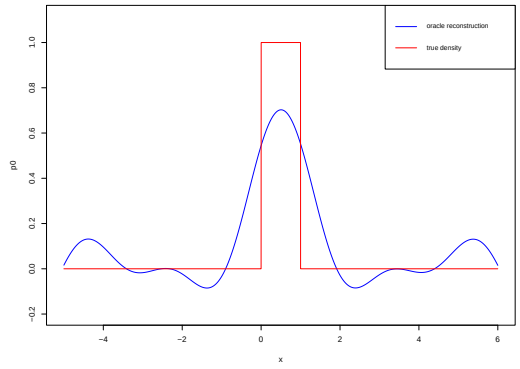


(b) Oracle reconstruction of p_0

Figure 2.5: Numerical results in case $\mu_0 \sim \mathrm{Uniform}([0, 1])$, when considering on $\mathrm{supp}(\mu_0)$.



(a) Plots of $h \mapsto \mathrm{MISE}_{\mathrm{oracle}}(h)$ for the Uniform density p_0

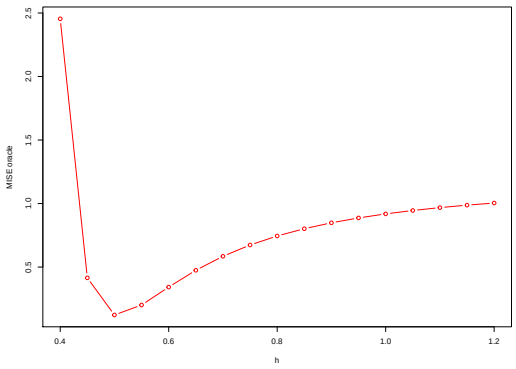


(b) Oracle reconstruction of p_0

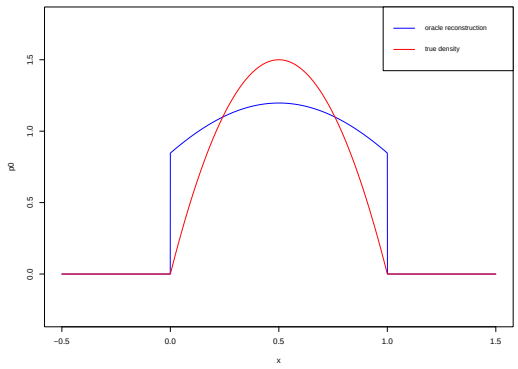
Figure 2.6: Numerical results in case $\mu_0 \sim \mathrm{Uniform}([0, 1])$, when considering on $[-5, 6]$.

$\mu_0 \sim$ **Beta(2,2)**

Considering the initial condition follows a Beta(2,2) distribution, so the density function is $p_0 = x.(1-x).\mathbb{1}_{[0,1]}(x)$. Similarly, Figure 2.7b and Figure 2.8b illustrate the plots of reconstruction of $\mathrm{Re}(\widehat{p}_{0,h_*})$ on $\mathrm{supp}(\mu_0)$ and on the interval $[-5,6]$, respectively. The optimal bandwidth h_* is chosen as minimizer of the oracle MISE (see Figure 2.7a and Figure 2.8a).

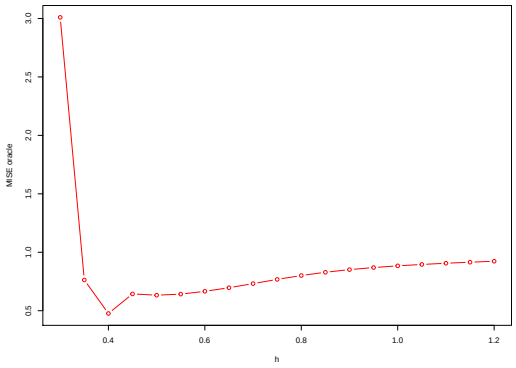


(a) Plots of $h \mapsto \mathrm{MISE}_{\mathrm{oracle}}(h)$ for the Beta(2,2) density p_0

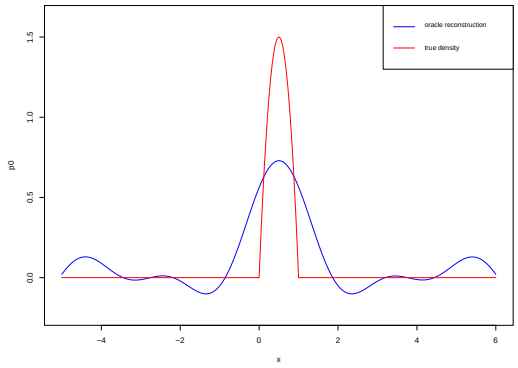


(b) Oracle reconstruction of p_0

Figure 2.7: Numerical results in case $\mu_0 \sim \mathrm{Beta}(2,2)$, when considering on $\mathrm{supp}(\mu_0)$.



(a) Plots of $h \mapsto \mathrm{MISE}_{\mathrm{oracle}}(h)$ for the Beta(2,2) density p_0



(b) Oracle reconstruction of p_0

Figure 2.8: Numerical results in case $\mu_0 \sim \mathrm{Beta}(2,2)$, when considering on $[-5,6]$.

Chapter 3

Adaptive warped kernel estimation of nonparametric regression with circular responses

In this chapter, we study the problem of statistical nonparametric estimation for regression when the response is circular. This chapter is organized as follows. We describe in Section 3.2 the construction of an adaptive warped kernel estimator of $m(x)$ at a given point $x \in \mathbb{R}$ with data-driven bandwidths selected by the Goldenshluger-Lepski method, when the design distribution is supposed to be known and invertible on $\mathbb{R}$. Under mild assumptions, an oracle-type inequality is established in Proposition 3.2.2 and rates of convergence are derived in Theorem 3.2.6. Furthermore, some numerical simulations are conducted in Section 3.3. All the proofs are gathered in Sections 3.4 and 3.5.

3.1 Introduction

Directional statistics is the branch of statistics which deals with observations that are directions. In this paper, we will consider more specifically *circular data* which arises whenever using a periodic scale to measure observations. These data are represented by points lying on the unit circle of $\mathbb{R}^2$ denoted in the sequel by $\mathbb{S}^1$. Circular data are collected in many research fields, for example in ecology (animal movements), earth sciences (wind and ocean current directions), medicine (circadian rhythm), forensics (crime incidence) or social science (clock or calendar effects). Various comprehensive surveys on statistical methods for circular data can be found in Mardia and Jupp [79], Jammalamadaka and SenGupta [73], Ley and Verdebout [78] and recent advances are collected in Pewsey and García-Portugués [83]. Note that the term *circular data* is also used to distinguish them from data supported on the real line $\mathbb{R}$ (or some subset of it), which henceforth are referred to as *linear data*.

Before entering into statistical considerations, it is necessary to equip the reader with important notations. In the sequel, a point on $\mathbb{S}^1$ will not be represented as a two-dimensional vector $\mathbf{w} = (w_2, w_1)^\top$ with unit Euclidean norm but as an angle $\theta = \text{atan2}(w_1, w_2)$ defined as follows:

Definition 3.1.1. The function $\text{atan2} : \mathbb{R}^2 \backslash (0,0) \mapsto [-\pi, \pi)$ is defined for any $(w_1, w_2) \in$

$\mathbb{R}^2\backslash(0,0)$ by

$$\text{atan2}(w_1,w_2) := \begin{cases} \arctan\left(\frac{w_1}{w_2}\right) & \text{if} \quad w_2 \geq 0, w_1 \neq 0 \\ 0 & \text{if} \quad w_2 > 0, w_1 = 0 \\ \arctan\left(\frac{w_1}{w_2}\right) + \pi & \text{if} \quad w_2 < 0, w_1 > 0 \\ \arctan\left(\frac{w_1}{w_2}\right) - \pi & \text{if} \quad w_2 < 0, w_1 \leq 0, \end{cases}$$

with arctan taking values in $[-\pi/2, \pi/2]$. In particular, $\arctan(+\infty) = \pi/2$ and $\arctan(-\infty) = -\pi/2$.

In this definition, one has arbitrarily fixed the origin of $\mathbb{S}^1$ at $(1,0)^\top$ and uses the anti-clockwise direction as positive. Thus, a circular random variable can be represented as angle over $[-\pi, \pi)$. Observe that $\text{atan2}(0,0)$ is undefined.

Figure 3.1: An illustration for the definition of the function $(y,x) \mapsto \text{atan2}(y,x)$.

In this work, we focus on a nonparametric regression model with circular response and linear predictor. Assume that we have an independent identically distributed (i.i.d in the sequel) sample $\{(X_j, \Theta_j)\}_{j=1}^n$ distributed as (X, Θ), where Θ is a circular random variable, and X is a random variable with density f_X supported on $\mathbb{R}$. We look for a function m which would contain the dependence structure between the predictors X_j and the observations Θ_j.
In the directional statistical literature, regression with circular response and linear covariates have been originally and mostly explored from a parametric point of view: main contributions are by Gould [72], Johnson and Wehlry [74], Fisher and Lee [70] or Presnell et al. [86]. Major drawbacks of such parametric models involve implausibility of fitted models, nonidentifiability of parameters, multimodal likelihood, difficulties in computation of parameters estimate or very strong assumptions. In order to get a more flexible approach, nonparametric paradigm has been then considered, in the pioneering work by Di Marzio et al. [80] and more recently in Meilán-Vila et al. [82]. It is quite surprising that to the best of our knowledge it has only been tackled in these single two papers.

We shall now give further precisions on the approach proposed by Di Marzio et al. in [80] to take into account the circular nature of the response. There is no doubt that circular data are fundamentally different from linear one by their periodic nature, and thus need specific tools. A preliminary task is to define how to assess the difference between the two angles Θ and $m(X)$. The circular context leads to a natural approach of using trigonometric functions. For two angles $\theta_1, \theta_2 \in [-\pi, \pi)$, we use the angular distance $d_c(\theta_1, \theta_2) := \big(1 - \cos(\theta_1 - \theta_2)\big)$ to measure the difference. Notice that this angular distance corresponds to the usual Euclidean norm in $\mathbb{R}^2$. Indeed, the angles θ_1 and θ_2 determines the corresponding points $(\cos\theta_1, \sin\theta_1)$ and $(\cos\theta_2, \sin\theta_2)$ respectively on the unit circle $\mathbb{S}^1$. Then, the usual squared Euclidean norm in $\mathbb{R}^2$ reads $\big(\cos\theta_1 - \cos\theta_2\big)^2 + \big(\sin\theta_1 - \sin\theta_2\big)^2 = 2.\big[1 - \cos(\theta_1 - \theta_2)\big] = 2.d_c(\theta_1, \theta_2)$.

Hence, using the angular distance, we look for a function m as the minimizer of the following risk:

$$L(m(X)) := \mathbb{E}\big[1 - \cos(\Theta - m(X))\big|X\big].$$

Minimizing $\mathbb{E}\big[1 - \cos(\Theta - m(X))|X\big]$ with respect to $m(X) \in [-\pi, \pi)$ is equivalent to maximizing $\mathbb{E}\big[\cos(\Theta - m(X))|X\big]$ with respect to $m(X) \in [-\pi, \pi)$. Now for $x \in \mathbb{R}$, let $m_1(x) := \mathbb{E}\big(\sin(\Theta)|X = x\big)$ and $m_2(x) := \mathbb{E}\big(\cos(\Theta)|X = x\big)$. Moreover, write

$$\begin{aligned}\mathbb{E}\big[\cos(\Theta - m(X))|X\big] &= \cos(m(X)).m_2(X) + \sin(m(X)).m_1(X)\\ &= \sqrt{(m_2(X))^2 + (m_1(X))^2}.\cos(m(X) - \gamma),\end{aligned}$$

where $\gamma \in [-\pi, \pi)$ is defined by

$$\cos(\gamma) := \frac{m_2(X)}{\sqrt{(m_2(X))^2 + (m_1(X))^2}}, \quad \text{and} \quad \sin(\gamma) := \frac{m_1(X)}{\sqrt{(m_2(X))^2 + (m_1(X))^2}}.$$

Thus, we have

$$\underset{m(X)\in[-\pi,\pi)}{\operatorname{argmin}} L(m(X)) = \underset{m(X)\in[-\pi,\pi)}{\operatorname{argmax}} \cos(m(X) - \gamma) = \gamma = \operatorname{atan2}\big(m_1(X), m_2(X)\big). \tag{3.1}$$

Hence, the minimizer of $L(m(X))$ is given by $m(x) = \operatorname{atan2}(m_1(x), m_2(x))$. In conclusion, the circular nature of the response is taken into account by the arctangent of the ratio of the sine and cosine components of the conditional trigonometric moment of Θ given X.

In the case of pointwise estimation and covariates supported on $[0, 1]$, Di Marzio et al. [80] investigated the performances of a Nadaraya-Watson and a local linear polynomial estimators to estimate $\operatorname{atan2}(m_1(x), m_2(x))$. Theoretically, for functions m_1 and m_2 being twice continuously differentiable, they obtained expressions for asymptotic bias and variance. Their proofs were based on linearization of the function arctangent by using Taylor expansions. But no sharp controls of the remainder terms in the expansions are obtained. Actually obtaining such controls would be very tedious with an approach based on Taylor expansions. As for the more recent work of Meilan Vila et al. [82], they studied the multivariate setting $[0, 1]^d$ with the same estimators and proofs technics. In both papers, neither rates of convergence nor adaptation are reached and cross-validation is used to select the kernel bandwidth in practice. By adaptation, we mean that the estimators do not require the specification of the regularity of the regression function which is crucial from a practical point of view. Furthermore, their estimators only depend on one single bandwidth whereas, it would be more natural to consider two bandwidths due to possible different regularities on m_1 and m_2. In view of this, we were motivated to fill the gap in the literature. Our goal is twofold: obtaining minimax rates of convergence for predictors supported on $\mathbb{R}$ and adaptation. To achieve this, we propose a new approach based on concentration inequalities along with warping method.

Statistical model

Let $\{(X_j, \Theta_j)\}_{j=1}^n$ be a random sample from (X, Θ), where Θ is a circular random variable taking values on $[-\pi, \pi)$, and X is a random variable with density f_X supported on $\mathbb{R}$ such that the cumulative distribution function of X is invertible on $\mathbb{R}$. It means that f_X is positive on $\mathbb{R}$ and $F_X(\mathbb{R}) = (0, 1)$. Recall that X and Θ are considered to be related through the regression function m. This context is considered in [80] for the first time. Furthermore, we investigate the pointwise nonparametric estimation of regression function m at a given point in $\mathbb{R}$.

Notations We introduce some notations used in the sequel.
Hereafter, $\|\cdot\|_{\mathbb{L}^1(\mathbb{R})}$ and $\|\cdot\|_{\mathbb{L}^2(\mathbb{R})}$ respectively denote the $\mathbb{L}^1$ and $\mathbb{L}^2$ norm on $\mathbb{R}$ with respect to the Lebesgue measure:

$$\|f\|_{\mathbb{L}^1(\mathbb{R})} = \int_{\mathbb{R}} |f(y)|dy, \quad \|f\|_{\mathbb{L}^2(\mathbb{R})} = \Big(\int_{\mathbb{R}} |f(y)|^2 dy \Big)^{1/2}.$$

The $\mathbb{L}^\infty$ norm is defined by $\|f\|_\infty = \sup_{y\in\mathbb{R}} |f(y)|$. Moreover, let $*$ denote the classical convolution defined for functions f, g by $f * g(x) := \int_{\mathbb{R}} f(x-y).g(y)dy$, for $x \in \mathbb{R}$. Finally, for $\alpha \in \mathbb{R}$, we denote by $[\alpha]_+ := \max\{\alpha; 0\}$ and for $\beta > 0$ let $\lfloor \beta \rfloor$ denote the largest integer strictly smaller than β, so $\lfloor \beta \rfloor < \beta$.

3.2 The estimation procedure in case of known design distribution

In this section, we consider the pointwise regression estimation problem with a known design distribution, so the cumulative distribution function of X is consequently considered to be known. First, in Section 3.2.1, at a given point $x \in \mathbb{R}$ which will be fixed along this chapter, we construct an estimator for $m(x)$ using warped kernel method with fixed bandwidth. Then, Section 3.2.2 presents a procedure for bandwidth selection by using the Goldenshluger-Lepski rule. Furthermore, oracle-type inequalities and rates of convergence for our adaptive estimator of $m(x)$ are presented in Section 3.2.3.

3.2.1 Warping strategy

In order to estimate the function $m(x) = \text{atan2}(m_1(x), m_2(x))$, $x \in \mathbb{R}$, we will use warped estimators. Let $F_X : y \in \mathbb{R} \mapsto \mathbb{P}(X \leq y)$ denote the cumulative distribution function (c.d.f) of the design X. Assuming that F_X is invertible, warping methods consists in introducing the auxiliary function $g := m \circ F_X^{<-1>}$, with $F_X^{<-1>}$ the inverse of F_X. The strategy then boils down to first estimating g by say $\hat{g}$ and then estimating the regression function of interest m by $\hat{g} \circ \widehat{F}_n$ with $\widehat{F}_n$ the empirical c.d.f of X. To deal with regression problem with random design, warping strategy has been applied in the past decade. Let us cite papers by Kerkyacharian and Picard [75], Pham Ngoc [84], Chagny [67] and Chagny et al. [69]. Among the advantages of this method, let us cite its stability whatever the design distribution, even with some inhomogeneous design. We first recall the popular Nadaraya-Watson (NW) estimator of m, denoted by $\widehat{m}_h^{NW}$, that is

$$\widehat{m}_h^{NW} : x \mapsto \frac{\frac{1}{n}\sum_{j=1}^n \Theta_j . K_h(x - X_j)}{\frac{1}{n}\sum_{j=1}^n K_h(x - X_j)},$$

with $K : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} K(y)dy = 1$ and $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$, for bandwidth $h > 0$. Unlike the popular Nadaraya-Watson, a warped kernel estimator does not involve a ratio, with a denominator which can be small and thus may lead to some instability. On the other hand, as adaptive estimation requires the data-driven selection of the bandwidth, the ratio form of the NW estimate indicates that we should select two bandwidths: one for the numerator and one for the denominator. Moreover, let $\widehat{m}_{1,h_1}$ and $\widehat{m}_{2,h_2}$ denote a kernel estimator of m_1 and m_2 respectively with bandwidth $h_1, h_2 > 0$. It is natural to consider that h_1 and h_2 are not necessary the same. Consequently, considering NW estimators for m_1 and m_2 may lead to the fact that there will be four bandwidths to be taken into account. This makes the study of these estimators much more complicated. Nevertheless, as we will see below in this Section 3.2.2, by applying the warping device, we not only have to consider two bandwidths in all, but also

avoid the ratio form as in NW estimator of $m(x)$.

We propose to adapt the strategy developed in the linear case by Chagny et al. in [69]. Particularly, in the case of 1-dimension, the warping device is based on the transformation $F_X(X_k)$ of the data X_k, $k = 1, ..., n$. We first define the kernel function considered in our framework as follows.

Definition 3.2.1. Let $K : \mathbb{R} \to \mathbb{R}$ be an integrable function such that K is compactly supported, $K \in \mathbb{L}^\infty(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. We say that K is a kernel if it satisfies $\int_{\mathbb{R}} K(y)dy = 1$.

Now, let $g_1, g_2 : (0, 1) \mapsto \mathbb{R}$ be defined by

$$g_1 := m_1 \circ F_X^{<-1>}, \quad \text{and} \quad g_2 := m_2 \circ F_X^{<-1>},$$

in such a way that $m_1 = g_1 \circ F_X$ and $m_2 = g_2 \circ F_X$. Moreover, for a function $f \in \mathbb{L}^1(\mathbb{R})$, the convolutions of f with g_1 and g_2 are defined respectively for $y \in \mathbb{R}$ by

$$\big(f * g_1\big)(y) := \int_0^1 g_1(v).f(y - v)dv, \quad \text{and} \quad \big(f * g_2\big)(y) := \int_0^1 g_2(v).f(y - v)dv. \tag{3.2}$$

Furthermore, we set for $v \in F_X(\mathbb{R})$,

$$g(v) := \text{atan2}\big(g_1(v), g_2(v)\big).$$

We observe that $m = g \circ F_X$. Now, for $u \in F_X(\mathbb{R})$, we set estimators:

$$\widehat{g}_{1,h_1}(u) := \frac{1}{n}\sum_{j=1}^{n} \sin(\Theta_j).K_{h_1}(u - F_X(X_j)), \quad \text{and} \quad \widehat{g}_{2,h_2}(u) := \frac{1}{n}\sum_{j=1}^{n} \cos(\Theta_j).K_{h_2}(u - F_X(X_j)), \tag{3.3}$$

to estimate g_1 and g_2 respectively, and $h_1, h_2 > 0$ are bandwidths of kernel $K_{h_1}(\cdot)$ and $K_{h_2}(\cdot)$ respectively.
Now, we define an estimator for g:

$$\widehat{g}_h(u) := \text{atan2}\big(\widehat{g}_{1,h_1}(u), \widehat{g}_{2,h_2}(u)\big), \quad \text{for } u \in F_X(\mathbb{R}), \tag{3.4}$$

where we denote $h := (h_1, h_2)$. Moreover, as a consequence, for $x \in \mathbb{R}$, the estimators for m_1 and m_2 are

$$\widehat{m}_{1,h_1}(x) := \widehat{g}_{1,h_1}\big(F_X(x)\big) = \frac{1}{n}\sum_{j=1}^{n} \sin(\Theta_j).K_{h_1}\big(F_X(x) - F_X(X_j)\big), \tag{3.5}$$

and

$$\widehat{m}_{2,h_2}(x) := \widehat{g}_{2,h_2}\big(F_X(x)\big) = \frac{1}{n}\sum_{j=1}^{n} \cos(\Theta_j).K_{h_2}\big(F_X(x) - F_X(X_j)\big). \tag{3.6}$$

We then define the estimator of $m(x)$ at $x \in \mathbb{R}$ by $\widehat{m}_h(x) := \text{atan2}\big(\widehat{m}_{1,h_1}(x), \widehat{m}_{2,h_2}(x)\big)$.

3.2.2 Data-driven estimator

Given any point $x \in \mathbb{R}$, we set $u_x := F_X(x)$. Then the pointwise quadratic-risk of the estimator $\widehat{m}_h(x)$ can be decomposed as

$$\mathbb{E}\Big[\big(\widehat{m}_h(x) - m(x)\big)^2\Big] = \mathbb{E}\Big[\big(\widehat{g}_h(u_x) - g(u_x)\big)^2\Big] = \mathbb{E}\Big[\big(\widehat{g}_h(u_x) - \mathbb{E}\big[\widehat{g}_h(u_x)\big]\big)^2\Big] + \Big[\big(\mathbb{E}\big[\widehat{g}_h(u_x) - g_h(u_x)\big]\big)^2\Big].$$

Through the warping method, we now work on the domain $F_X(\mathbb{R}) \subseteq [0,1]$ and in the sequel, we will focus on the estimator $\widehat{g}_h$ of g.
In the following, we apply the method proposed by Goldenshluger and Lepski in [71] to establish a procedure in order to select an optimal value for bandwidths h_1 and h_2 automatically.
For $h_{\max}$ a constant depending on x (for $A > 0$ satisfying $\mathrm{supp}(K) \subseteq [-A, A]$, we take $h_{\max}$ such that $u_x - A.h_{\max} > 0$ and $u_x + A.h_{\max} < 1$), we consider the following finite collection of bandwidths:

$$\mathcal{H}_n := \left\{ h = k^{-1} : \ k \in \mathbb{N}^*, \ h \leq h_{\max}, \ n.h > \max\left(\frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{\|K\|^2_\infty}; 1 \right). \log(n) \right\}. \tag{3.7}$$

We have $\mathrm{Card}\big(\mathcal{H}_n\big) \approx n/\log n$. First, for $\alpha \in \mathbb{R}$ let $\big[\alpha\big]_+ := \max\{\alpha; 0\}$, then for $h_1 \in \mathcal{H}_n$ and $v \in F_X(\mathbb{R})$ we set

$$A_1(h_1, v) := \sup_{h_1' \in \mathcal{H}_n} \Big[\big| \widehat{g}_{1,h_1,h_1'}(v) - \widehat{g}_{1,h_1'}(v) \big| - \sqrt{\widetilde{V}_1(n, h_1')} \Big]_+, \tag{3.8}$$

with $\widetilde{V}_1(n, h_1') := c_{0,1}. \dfrac{\log(n). \|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_1'}$, $c_{0,1} > 0$ a tuning parameter and

$$\widehat{g}_{1,h_1,h_1'}(v) := \big(K_{h_1'} * \widehat{g}_{1,h_1}\big)(v) = \frac{1}{n} \sum_{k=1}^{n} \sin(\Theta_k). \big(K_{h_1'} * K_{h_1}\big)\big(v - F_X(X_k)\big),$$

so that $\widehat{g}_{1,h_1,h_1'}(v) = \widehat{g}_{1,h_1',h_1}(v)$, where remind that $*$ denotes the classical convolution operator. Then, an adaptive (random) choice of bandwidth h_1 is selected by

$$\widehat{h}_1 = \underset{h_1 \in \mathcal{H}_n}{\mathrm{argmin}} \Big[A_1(h_1, v) + \sqrt{\widetilde{V}_1(n, h_1)} \Big]. \tag{3.9}$$

In a similar way, we define

$$\widehat{h}_2 = \underset{h_2 \in \mathcal{H}_n}{\mathrm{argmin}} \Big[A_2(h_2, v) + \sqrt{\widetilde{V}_2(n, h_2)} \Big], \tag{3.10}$$

where we set

$$A_2(h_2, v) := \sup_{h_2' \in \mathcal{H}_n} \Big[\big| \widehat{g}_{2,h_2,h_2'}(v) - \widehat{g}_{2,h_2'}(v) \big| - \sqrt{\widetilde{V}_2(n, h_2')} \Big]_+, \tag{3.11}$$

with $\widetilde{V}_2(n, h_2') := c_{0,2}. \dfrac{\log(n). \|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_2'}$, $c_{0,2} > 0$ a tuning parameter and

$$\widehat{g}_{2,h_2,h_2'}(v) := \big(K_{h_2'} * \widehat{g}_{2,h_2}\big)(v) = \frac{1}{n} \sum_{k=1}^{n} \cos(\Theta_k). \big(K_{h_2'} * K_{h_2}\big)\big(v - F_X(X_k)\big),$$

so that $\widehat{g}_{2,h_2,h_2'}(v) = \widehat{g}_{2,h_2',h_2}(v)$.
The criteria (3.9) and (3.10) are inspired from [71], in order to mimic the optimal "bias-variance" trade-off using empirical estimations. It is common to use $\widetilde{V}_1(n, h_1')$ and $\widetilde{V}_2(n, h_2')$ to approximate an upper-bound for the corresponding variance terms, whereas the more involved task of the Goldenshluger-Lepski method is to provide an empirical counterpart for the bias term of the upper-bound by comparing pair-by-pair several estimators. In our framework, the bias term corresponds to $\big|\big(K_{h_1} * g_1\big)(v) - g_1(v)\big|$, so it is natural to replace it by $\big|\big(K_{h_1} *$

$\widehat{g}_{1,h_1'}\big)(v) - \widehat{g}_{1,h_1'}(v)\big|$. Thus, the estimator of the bias term is defined as $A_1(h_1, v)$ in (3.8). Actually, the term $\sqrt{\widetilde{V}_1(n, h_1')}$ in the definition of $A_1(h_1, v)$ is used to control the fluctuations of this estimation rule.
Now, we define the kernel estimator of $g(v)$ with data-driven bandwidths as follows:

$$\widehat{g}_{\widehat{h}}(v) := \operatorname{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(v), \widehat{g}_{2,\widehat{h}_2}(v)\big), \tag{3.12}$$

where we denote $\widehat{h} := (\widehat{h}_1, \widehat{h}_2)$, and

$$\widehat{g}_{1,\widehat{h}_1}(v) := \frac{1}{n}\sum_{k=1}^{n} \sin(\Theta_k).K_{\widehat{h}_1}\big(v - F_X(X_k)\big), \quad \text{and} \quad \widehat{g}_{2,\widehat{h}_2}(v) := \frac{1}{n}\sum_{k=1}^{n} \cos(\Theta_k).K_{\widehat{h}_2}\big(v - F_X(X_k)\big). \tag{3.13}$$

Moreover, as a consequence, at $x \in \mathbb{R}$, the adaptive estimators for m_1 and m_2 are

$$\widehat{m}_{1,\widehat{h}_1}(x) := \big(\widehat{g}_{1,\widehat{h}_1} \circ F_X\big)(x) = \frac{1}{n}\sum_{j=1}^{n} \sin(\Theta_j).K_{\widehat{h}_1}\big(F_X(x) - F_X(X_j)\big), \tag{3.14}$$

and

$$\widehat{m}_{2,\widehat{h}_2}(x) := \big(\widehat{g}_{2,\widehat{h}_2} \circ F_X\big)(x) = \frac{1}{n}\sum_{j=1}^{n} \cos(\Theta_j).K_{\widehat{h}_2}\big(F_X(x) - F_X(X_j)\big). \tag{3.15}$$

We finally define the adaptive estimator for $m(x)$ at a given point $x \in \mathbb{R}$ by

$$\widehat{m}_{\widehat{h}}(x) := \operatorname{atan2}\big(\widehat{m}_{1,\widehat{h}_1}(x), \widehat{m}_{2,\widehat{h}_2}(x)\big). \tag{3.16}$$

3.2.3 Oracle-type inequalities and rates of convergence

We first state an oracle-type inequality for $\widehat{g}_{1,\widehat{h}_1}(u_x)$ and $\widehat{g}_{2,\widehat{h}_2}(u_x)$.

Proposition 3.2.2. *Consider the collection of bandwidths $\mathcal{H}_n$ defined in (3.7). Let $q \geq 1$ and assume that $\min\{c_{0,1}; c_{0,2}\} \geq 16\big(2+q\big)^2.\big(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$. Then, for n sufficiently large, with a probability greater than $1 - 4.n^{-q}$,*

$$\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| \leq \inf_{h_1 \in \mathcal{H}_n}\Big[\big(1 + 2.\|K\|_{\mathbb{L}^1(\mathbb{R})}\big).\|g_1 - K_{h_1} * g_1\|_\infty + 3.\sqrt{\widetilde{V}_1(n, h_1)}\Big],$$

and

$$\big|\widehat{g}_{2,\widehat{h}_2}(u_x - g_2(u_x)\big| \leq \inf_{h_2 \in \mathcal{H}_n}\Big[\big(1 + 2.\|K\|_{\mathbb{L}^1(\mathbb{R})}\big).\|g_2 - K_{h_2} * g_2\|_\infty + 3.\sqrt{\widetilde{V}_2(n, h_2)}\Big].$$

The proof of Proposition 3.2.2 is given in Section 3.5.1.
The oracle inequality in Proposition 3.2.2 illustrates a bias-variance decomposition of the pointwise quadratic-risk. The term $\|g_1 - K_{h_1} * g_1\|_\infty$ corresponds to a bias term since $\mathbb{E}\big[\widehat{g}_{1,h_1}(u_x)\big] = \big(K_{h_1} * g_1\big)(u_x)$ (see (3.20)), while remind that $\widetilde{V}_1(n, h_1)$ approximates an upper-bound for the variance term.
The fact is that the pointwise quadratic-risk for the estimator $\widehat{g}_h = \operatorname{atan2}\big(\widehat{g}_{1,h_1}, \widehat{g}_{2,h_2}\big)$ will be derived from results of $\widehat{g}_{1,h_1}$ and $\widehat{g}_{2,h_2}$. Since the function $\operatorname{atan2}(w_1, w_2)$ is undefined when $w_1 = w_2 = 0$, it is reasonable to consider the following assumption:

Assumption 3.2.3. For given $x \in \mathbb{R}$, assume that

$$|m_1(x)| = \big|g_1(F_X(x))\big| \geq \delta_1 > 0, \quad \text{and} \quad |m_2(x)| = \big|g_2(F_X(x))\big| \geq \delta_2 > 0.$$

Moreover, to study upper-bounds for the pointwise quadratic-risk of $\widehat{g}_{1,h_1}$ and $\widehat{g}_{2,h_2}$, we need some further assumptions on the regularity of g_1 and g_2. Thus, we initially introduce the following Hölder classes that are adapted to local estimation.

Definition 3.2.4. Let $\beta > 0$ and $L > 0$. The Hölder class $\mathcal{H}(\beta, L)$ is the set of functions $f : (0,1) \mapsto \mathbb{R}$, such that f admits derivatives up to the order $\lfloor \beta \rfloor$ (recall that $\lfloor \beta \rfloor$ denotes the largest integer strictly smaller than β), and for any $(y, \widetilde{y})$,

$$\left| \frac{d^{\lfloor \beta \rfloor} f}{(dy)^{\lfloor \beta \rfloor}}(\widetilde{y}) - \frac{d^{\lfloor \beta \rfloor} f}{(dy)^{\lfloor \beta \rfloor}}(y) \right| \leq L.|\widetilde{y} - y|^{\beta - \lfloor \beta \rfloor}.$$

We also consider the following assumption on the kernel K:

Assumption 3.2.5. The kernel K is of order $\mathcal{L} \in \mathbb{R}_+$, i.e.

(i) $C_{K,\mathcal{L}} := \int_{\mathbb{R}} (1 + |y|)^{\mathcal{L}}.|K(y)|dy < \infty$;

(ii) $\forall k \in \{1, ..., \lfloor \mathcal{L} \rfloor\}$, $\int_{\mathbb{R}} y^k.K(y)dy = 0$.

Now, we obtain an upper-bound for the pointwise quadratic-risk of our final fully data-driven estimator $\widehat{m}_{\widehat{h}}$ at x defined in (3.16) as in the following theorem:

Theorem 3.2.6. *Let $\beta_1, \beta_2, L_1, L_2 > 0$. Suppose that g_1 belongs to a Hölder class $\mathcal{H}(\beta_1, L_1)$, g_2 belongs to $\mathcal{H}(\beta_2, L_2)$, the kernel K satisfies Assumption 3.2.5 with an index $\mathcal{L} \in \mathbb{R}_+$ such that $\mathcal{L} \geq \max(\beta_1, \beta_2)$. Let $q \geq 1$, and suppose that $\min\{c_{0,1}; c_{0,2}\} \geq 16\big(2+q\big)^2.\big(1 + \|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$. Then, under Assumption 3.2.3, for n sufficiently large,*

$$\mathbb{E}\Big[\big|\widehat{m}_{\widehat{h}}(x) - m(x)\big|^2\Big] \leq 32.\pi^2.\Big(\frac{1}{\delta_2^2} + \frac{1}{\delta_1^2}\Big).(C_1^2 + C_2^2).\max\left\{\psi_n(\beta_1)^2, \psi_n(\beta_2)^2\right\},$$

where C_1 is a constant (depending on $\beta_1, L_1, c_{0,1}, K$) and C_2 is a constant (depending on $\beta_2, L_2, c_{0,2}, K$), along with $\psi_n(\beta_1) = \big(\log(n)/n\big)^{\frac{\beta_1}{2\beta_1+1}}$ and $\psi_n(\beta_2) = \big(\log(n)/n\big)^{\frac{\beta_2}{2\beta_2+1}}$.

A proof of Theorem 3.2.6 is given in Section 3.5.2.
Observe that if $\beta_1 = \beta_2 = \beta$, then we obtain the rate $\psi_n(\beta) = (\log n/n)^{\beta/(2\beta+1)}$, which is the optimal rate for adaptive univariate regression function estimation and pointwise risk (see e.g. Section 2 in [66]). In this case, the obtained rate is much faster than the rate $(\log n/n)^{\beta/(2\beta+2)}$ obtained for bivariate regression function estimation in the isotropic case. This justifies the introduction of specific methodologies for circular regression. We conjecture that the rate obtained in Theorem 3.2.6 is optimal even if it remains an open problem.

Remark. Eventually, to obtain the fully computable estimator, we replace the c.d.f F_X by its empirical counterpart $\widehat{F}_n(y) := \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{X_j \leq y}$. This replacement should not change the final rate of convergence of our nonparametric estimator. The scheme of this switching has been detailed for instance in [68].

3.3 Numerical simulations

In this section, we implement some simulations to study the performance of our estimator. We consider the following regression models:

$$\text{M1.} \quad \Theta = \text{atan2}\big(2X - 1, X^2 + 2\big) + \zeta \pmod{2\pi}; \tag{3.17}$$

$$\text{M2.} \quad \Theta = \text{atan2}\big(-2X + 1, X^2 - 1\big) + \zeta \pmod{2\pi}; \tag{3.18}$$

$$\text{M3.} \quad \Theta = \arccos\big(X^5 - 1\big) + 3.\arcsin\big(X^3 - X + 1\big) + \zeta \pmod{2\pi}, X \in [0,1], \tag{3.19}$$

where the circular errors ζ_k are drawn from a von Mises distribution $vM(0,10)$ and independent with X_k. Recall that the von Mises distribution $vM(\widetilde{\mu},\widetilde{\kappa})$ has density $\theta \in [-\pi,\pi) \mapsto \frac{1}{2\pi.I_0(\widetilde{\kappa})}.\exp\big(\widetilde{\kappa}.\cos(\theta-\widetilde{\mu})\big)$, see e.g [78, Section 2.2.2], with $I_0(\widetilde{\kappa})$ the modified Bessel function of the first kind and of order 0,[1], a location parameter $\widetilde{\mu} \in [-\pi,\pi)$ and concentration $\widetilde{\kappa} > 0$. Moreover, one reason to consider the polynomial X^2+2 in model M1 is due to the fact that by the definition we have $\text{atan2}(w_1,w_2) = \arctan\big(\frac{w_1}{w_2}\big)$, when $w_2 > 0$. By considering first model M1 and then model M2, we intend to cover the whole valued-range $[-\pi,\pi)$ of function atan2. On the other hand, in model M3, the regression function is a sum of inverse of trigonometric functions and so we do not use directly function atan2.
First, we draw a sample $\big\{(\Theta_j, X_j)\big\}_{j=1,n}$. Additionally, we make simulations for the general case of unknown design distribution, i.e. the final estimators are computed using the empirical counterpart $\widehat{F}_n(y) := \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{X_j \le y}$, with $y \in \mathbb{R}$, of the warping function F_X.
In this simulation study, the final adaptive estimator $\widehat{m}_{\widehat{h}}$ is calculated using $\widehat{F}_n(\cdot)$ and a Epanechnikov kernel K defined by $K(y) = \frac{3}{4}.(1-y^2).\mathbb{1}_{|y|\le 1}, y \in \mathbb{R}$. Moreover, we consider a discretization of bandwidth collection $\mathcal{H}_n$ defined as $\mathcal{H}_n := \left\{k^{-1} : k \in \mathbb{N}, 1 \le k \le \frac{n}{\log(n)}\right\}$.

3.3.1 Practical calibration of tuning paremeters

In the bandwidth selection procedure (GL procedure) described in Section 3.2.2, we need to choose two tuning parameters $c_{0,1}$ and $c_{0,2}$ in order to find an optimal value of the quadratic pointwise risk

$$\mathcal{R} := \big|\widehat{m}_{\widehat{h}}(x) - m(x)\big|^2$$

with $\widehat{m}_{\widehat{h}}(x) = \text{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(\widehat{F}_n(x)), \widehat{g}_{2,\widehat{h}_2}(\widehat{F}_n(x))\big)$. To do this, we implement a preliminary simulation to calibrate $c_{0,1}$ and $c_{0,2}$.
For this part of calibration, we consider model M1. In Figure 3.4, a simulated data is illustrated by green points and the true regression function m is presented by the red curve. We intend to estimate $m(x)$ at $x = -2$ and also at $x = 1.25$.

■ In case $c_{0,1} = c_{0,2}$

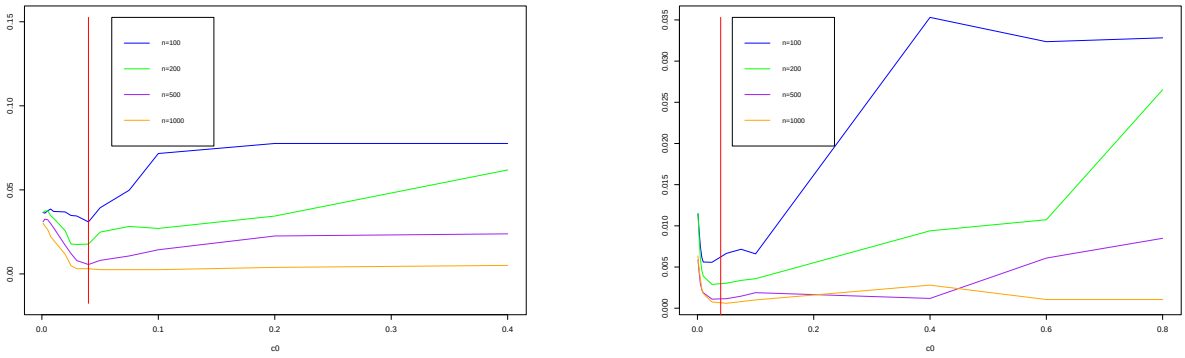
We consider the same values for the two tuning parameters in procedure to find $\widehat{h}_1$ and $\widehat{h}_2$, i.e. $c_{0,1} = c_{0,2} = c_0$. We compute the risk $\mathcal{R}$ defined above as a function of c_0 among the following discretization

$$G_{c_0} := \{0.001;\ 0.0025;\ 0.005;\ 0.0075;\ 0.01;\ 0.025;\ 0.05;\ 0.075;\ 0.1;\ 0.2;\ 0.3;\ 0.4\}.$$

More precisely, for each $c_0 \in G_{c_0}$, we run the GL procedure with $c_{0,1} = c_{0,2} = c_0$ to find $\widehat{h}_1$ and $\widehat{h}_2$. Then, we compute the risk as $\mathcal{R}(c_0) := \big|\widehat{m}_{\widehat{h}}(x) - m(x)\big|^2$. Moreover, we carry out this task for different sample sizes $n \in \{100; 200; 500; 1000\}$ to observe the stability of c_0. The numerical illustrations for this calibration are shown in Figure 3.2 of estimation at $x = -2$ and in Figure 3.3 at $x = 1.25$, respectively. Consequently, we choose $c_{0,1} = c_{0,2} = 0.04$ for the numerical experiments in the sequel.

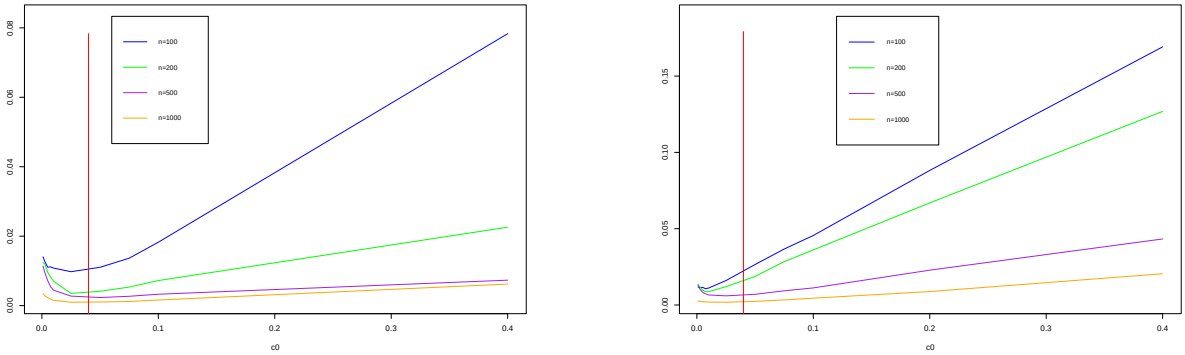
[1] The modified Bessel function of the first kind and of order $\alpha \ge 0$, $I_\alpha(z)$ for $z > 0$, admits the integral representation

$$I_\alpha(z) = \frac{1}{\pi}\int_0^\pi \exp\big(z.\cos(\theta)\big).\cos(\alpha\theta)d\theta - \frac{\sin(\alpha\pi)}{\pi}\int_0^\infty \exp\big(-z.\cosh(y)-\alpha y\big)dy.$$



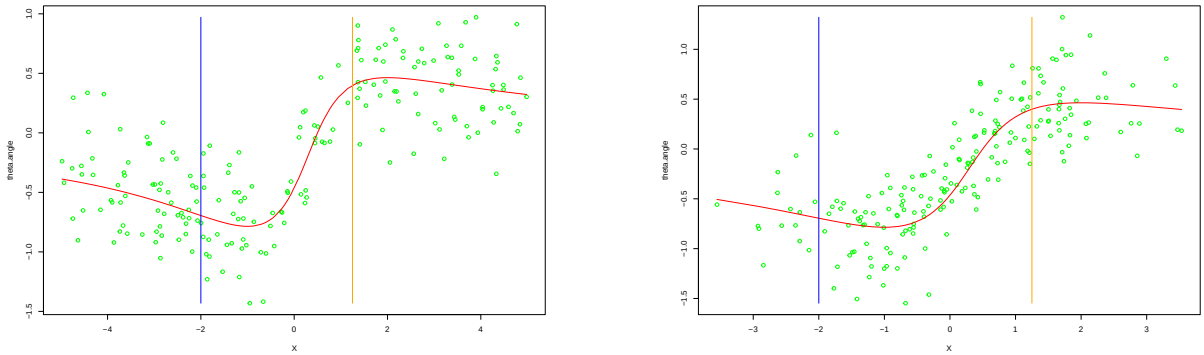
(a) $X_j \sim \mathcal{N}(0, 1.5)$ (b) $X_j \sim U([-5, 5])$

Figure 3.2: Plot of $\mathcal{R}(c_0)$ with $c_0 \in G_{c_0}$ for different sample sizes n of 50 Monte Carlo repetitions of estimation at $x = -2$, (using Epanechnikov kernel).



(a) $X_j \sim \mathcal{N}(0, 1.5)$ (b) $X_j \sim U([-5, 5])$

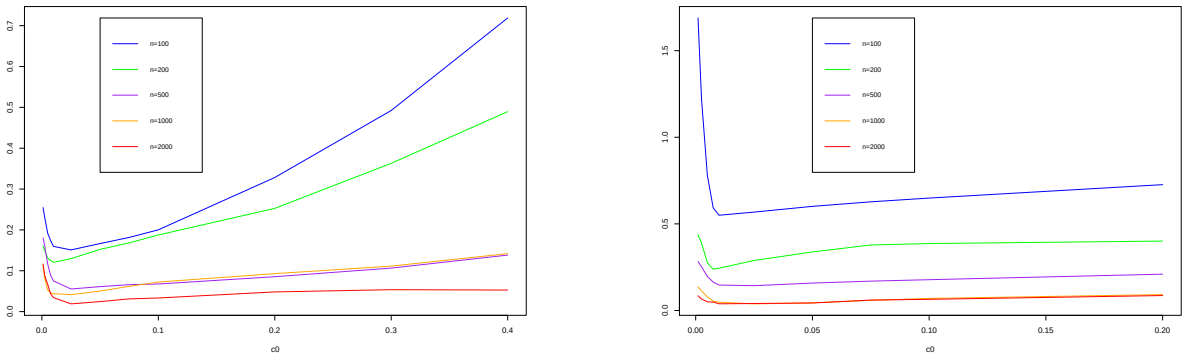
Figure 3.3: Plot of $\mathcal{R}(c_0)$ with $c_0 \in G_{c_0}$ for different sample sizes n of 50 Monte Carlo repetitions of estimation at $x = 1.25$, (using Epanechnikov kernel).



(a) Model M1 : $X_j \sim U([-5, 5])$ (b) Model M1 : $X_j \sim \mathcal{N}(0, 1.5)$

Figure 3.4: An illustration for simulated data $\{\Theta_j\}_{j=1}^n$ (green points) of model-M1 with $n = 200$. The red curve presents the regression function m, while the blue vertical line displays point $x = -2$ and the orange vertical line displays point $x = 1.25$ where we aim at estimating $m(x)$.

Remark. In addition, when using Gaussian kernel defined by $K(y) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{y^2}{2}}, \quad y \in \mathbb{R}$, the numerical illustrations for this calibration are shown in Figure 3.5 for $X \sim U([-5,5])$ and for $X \sim \mathcal{N}(0, 1.5)$. This brief numerical study had shown that the choice of $c_{0,1} = c_{0,2} = 0.04$ is convincing as well for Gaussian kernel.

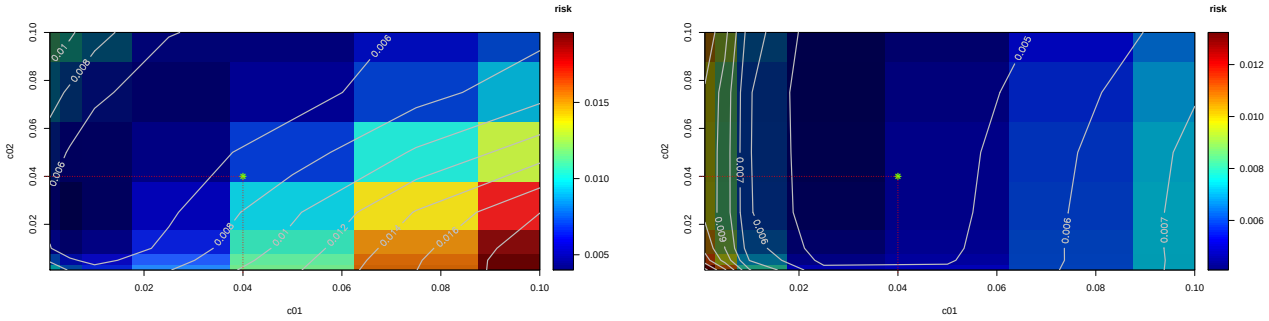


(a) Model M1 : $X \sim U([-5,5])$. (b) Model M1 : $X \sim \mathcal{N}(0, 1.5)$.

Figure 3.5: Plot of $\mathcal{R}(c_0)$ with $c_0 \in G_{c_0}$ for different sample sizes n of 50 Monte Carlo repetitions when estimating at $x = -2$, (using Gaussian kernel).

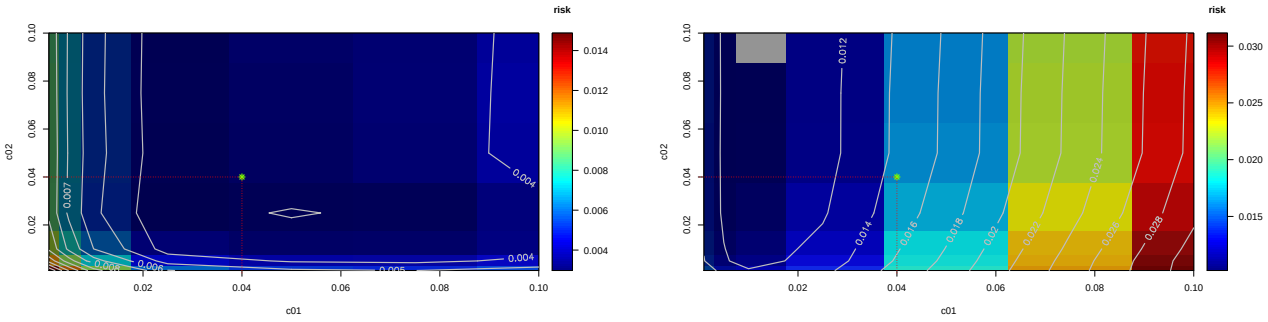
■ In case $c_{0,1}$ may differ from $c_{0,2}$

In this case, with $n = 200$, for $c_{0,1}, c_{0,2}$ among the following discretization $G_{c_0} := \{0.001;\ 0.005;\ 0.01;\ 0.025;\ 0.05;\ 0.075;\ 0.1;\ 0.2;\ 0.3;\ 0.4\}$, we compute the corresponding risk $\mathcal{R}(c_{0,1}, c_{0,2})$. The numerical illustration of the calibration for $c_{0,1}$ and $c_{0,2}$ are presented in Figure 3.6 and Figure 3.7 below for the case $X_j \sim \mathcal{N}(0, 1.5)$ and $X_j \sim U([-5,5])$, respectively. Accordingly, in general, the choice of $c_{0,1} = c_{0,2} = 0.04$ is reasonable for numerical simulations in the sequel.



(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

Figure 3.6: 2D-image of risk $\mathcal{R}(c_{0,1}, c_{0,2})$ corresponding to the grids of $(c_{0,1}, c_{0,2}) \in G_{c_0} \times G_{c_0}$ for Model M1 with $X_j \sim \mathcal{N}(0, 1.5)$, $n = 200$ and 50 repetitions of Monte Carlo, (using Epanechnikov kernel). The specific point displays for $c_{0,1} = c_{0,2} = 0.04$.



(a) estimation at $x = -2$ (b) estimation at $x = 1.25$

Figure 3.7: 2D-image of risk $\mathcal{R}(c_{0,1}, c_{0,2})$ corresponding to the grids of $(c_{0,1}, c_{0,2}) \in G_{c_0} \times G_{c_0}$ for Model M1 with $X_j \sim U([-5, 5])$, $n = 200$ and 50 repetitions of Monte Carlo, (using Epanechnikov kernel). The specific point displays for $c_{0,1} = c_{0,2} = 0.04$.

3.3.2 Numerical results

Remind that in the following numerical experiments, for the GL procedure of selecting bandwidths, we choose $c_{0,1} = c_{0,2} = 0.04$. We first implement the numerical experiments for three models M1, M2, M3 using Epanechnikov kernel, and then a similar scheme is conducted but with Gaussian kernel.

■ Using Epanechnikov kernel

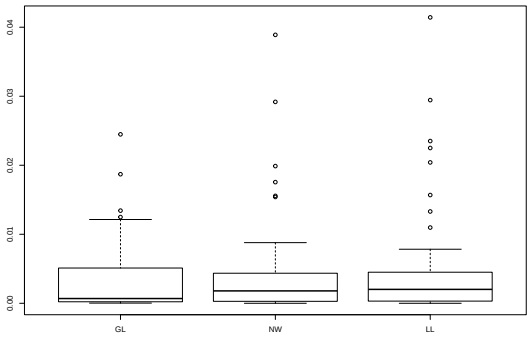
1. For model M1: Boxplots in Figures 3.8 and 3.9 summarize results of our numerical simulations for model M1 in case $X \sim U([-5, 5])$ and the case $X \sim \mathcal{N}(0, 1.5)$, respectively. In both cases, for model M1, we estimate $m(x)$ at $x = -2$ and at another point $x = 1.25$.

2. For model M2: Figure 3.10 shows the numerical simulation for model M2 with $X \sim \mathcal{N}(0, 1.5)$ and we estimate $m(x)$ at $x = 1.05$.

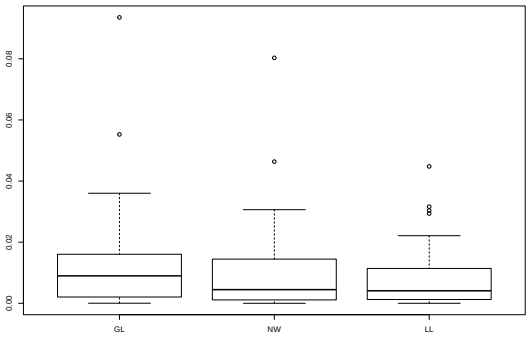
3. For model M3: Figure 3.11 shows the numerical simulation for model M3 with $X \sim U([0, 1])$ and we estimate $m(x)$ at $x = 0.95$.

In the graphics below, we denote by "GL" the pointwise squared risk (computed with 50 Monte Carlo replications) corresponding to $\widehat{m}_{\widehat{h}}(x)$. Moreover, as proposed in [80], we also compute the Nadaraya-Watson (NW) estimator $\widehat{m}_h^{NW}$ (see in Section 3.2.1) and local linear (LL) estimator (see [80, Section 4.2]) denoted by $\widehat{m}_h^{LL}$ of m to make a comparison with our adaptive estimator. In addition, Cross-Validation is used to select bandwidth $h > 0$ for $\widehat{m}_h^{NW}$ and $\widehat{m}_h^{LL}$, then, we compute the corresponding pointwise squared errors (computed with 50 Monte Carlo replications). In the graphics below, we denote by "NW" and "LL" the pointwise squared risk (computed with 50 Monte Carlo replications) corresponding to $\widehat{m}_h^{NW}$ and $\widehat{m}_h^{LL}$, respectively.

Generally, in our numerical experiments, the results presented by boxplots in Figures 3.8, 3.9, 3.10 and 3.11 show that the performances of our estimator are quite satisfying.

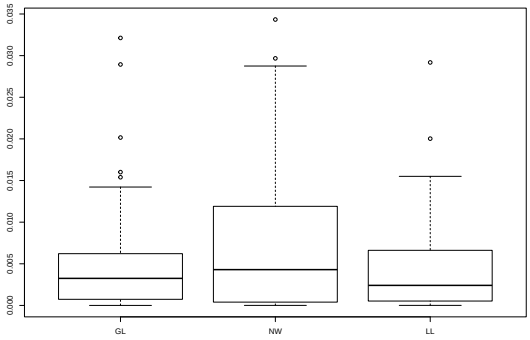


(a) Model M1: estimation at $x = -2$

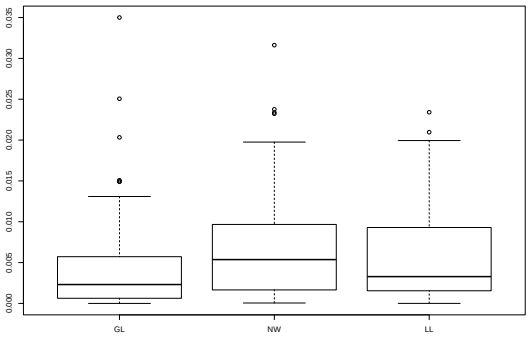


(b) Model M1: estimation at $x = 1.25$

Figure 3.8: Boxplot for 50 runs of Monte Carlo resampling with sample size $n = 200$, in case $X \sim U([-5, 5])$, in comparison with NW and LL (using Epanechnikov kernel).

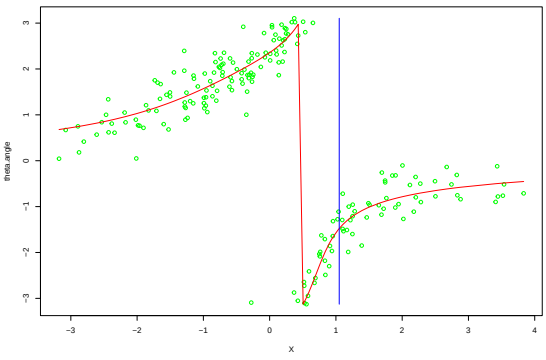


(a) Model M1: estimation at $x = -2$

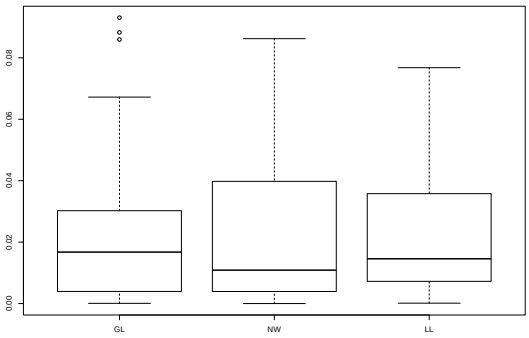


(b) Model M1: estimation at $x = 1.25$

Figure 3.9: Boxplot for 50 runs of Monte Carlo resampling with sample size $n = 200$, in case $X \sim \mathcal{N}(0, 1.5)$, in comparison with NW and LL (using Epanechnikov kernel).

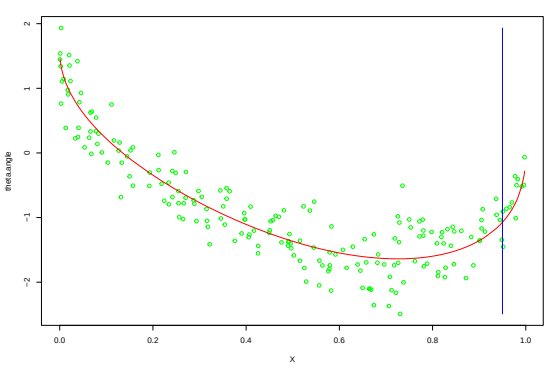


(a) Model M2 with $X \sim \mathcal{N}(0, 1.5)$: simulated data

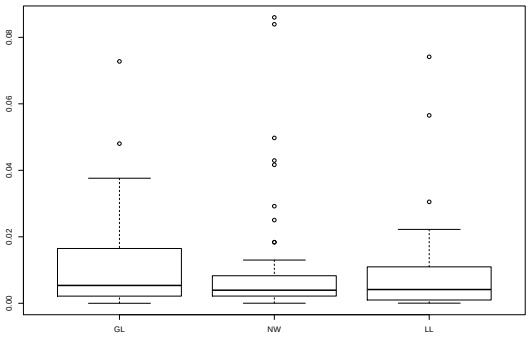


(b) Boxplot of pointwise squared risk in comparison with NW and LL

Figure 3.10: (a): simulated data $\{\Theta_j\}_{j=1}^n$ (green points) of Model-M2 with $n = 200$. The red curve presents the true regression function m, while the blue vertical line displays point $x = 1.05$ where we aim at estimating $m(x)$; (b): boxplot for 50 runs of Monte Carlo, (using Epanechnikov kernel).



(a) Model M3 with $X \sim U([0,1])$: simulated data



(b) Boxplot of pointwise squared risk in comparison with NW and LL

Figure 3.11: (a): simulated data $\{\Theta_j\}_{j=1}^n$ (green points) of Model-M3 with $n = 200$. The red curve presents the true regression function m, while the blue vertical line displays point $x = 0.95$ where we aim at estimating $m(x)$; (b): boxplot for 50 runs of Monte Carlo, (using Epanechnikov kernel).

■ Using Gaussian kernel

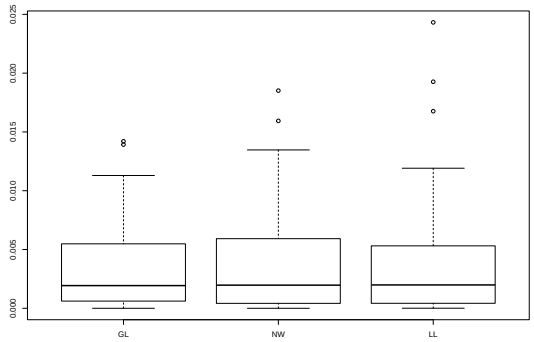
In this subsection, we repeat the numerical experiments studied in the previous part, but now with the use of Gaussian kernel.

1. For model M1: Boxplots in Figures 3.12 and 3.13 summarize the results of our numerical simulations for model M1 in case $X \sim U([-5,5])$ and in the case $X \sim \mathcal{N}(0, 1.5)$, respectively. In both cases, for model M1, we estimate $m(x)$ at $x = -2$ and at another point $x = 1.25$.

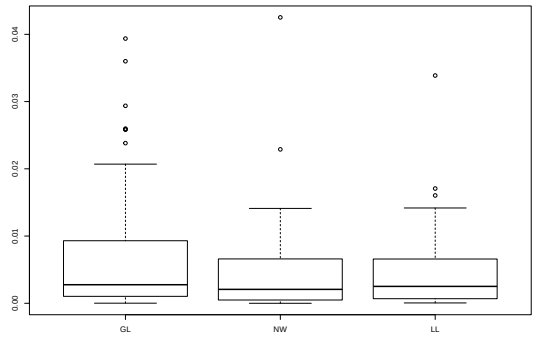
2. For model M2: Figure 3.14a shows the numerical simulation for model M2 with $X \sim \mathcal{N}(0, 1.5)$ and we estimate $m(x)$ at $x = 1.05$.

3. For model M3: Figure 3.14b shows the numerical simulation for model M3 with $X \sim U([0,1])$ and we estimate $m(x)$ at $x = 0.95$.

In this case of using Gaussian kernel, boxplots in Figures 3.12, 3.13, 3.14a and 3.14b show that the performances of our adaptive estimator are quite satisfying as well.

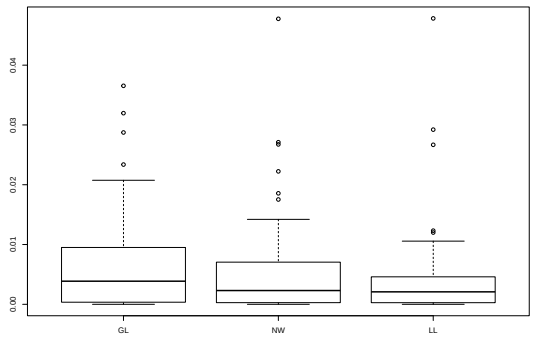


(a) Model M1: estimation at $x = -2$

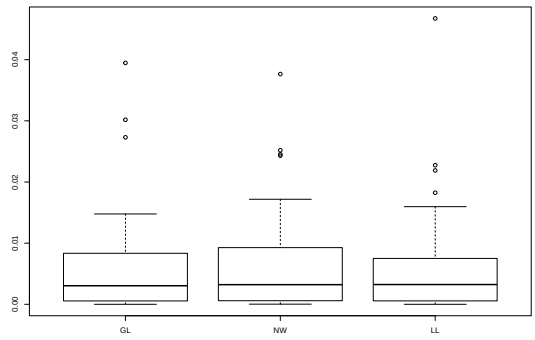


(b) Model M1: estimation at $x = 1.25$

Figure 3.12: Boxplot for 50 runs of Monte Carlo resampling with sample size $n = 200$, in case $X \sim U([-5, 5])$, in comparison with NW and LL, (using Gaussian kernel).

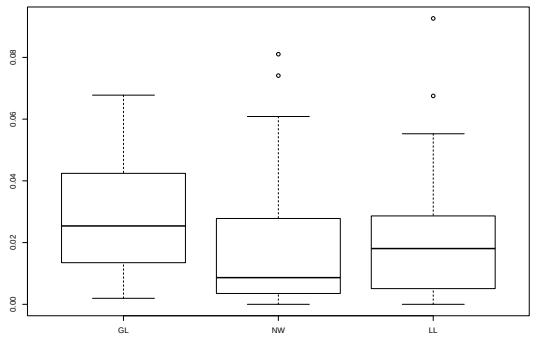


(a) Model M1: estimation at $x = -2$

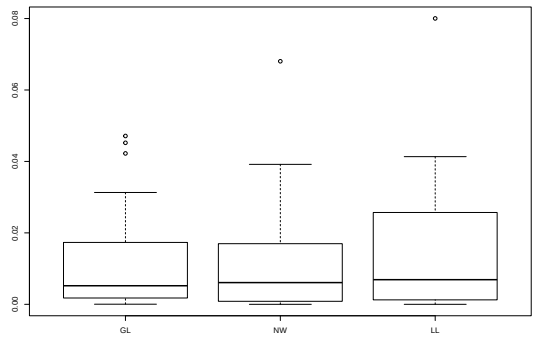


(b) Model M1: estimation at $x = 1.25$

Figure 3.13: Boxplot for 50 runs of Monte Carlo resampling with sample size $n = 200$, in case $X \sim \mathcal{N}(0, 1.5)$, in comparison with NW and LL, (using Gaussian kernel).



(a) Model M2: estimation at $x = 1.05$ with $X \sim \mathcal{N}(0, 1.5)$.



(b) Model M3: estimation at $x = 0.95$ with $X \sim U([0, 1])$.

Figure 3.14: Boxplots for 50 runs of Monte Carlo resampling with sample size $n = 200$, in comparison with NW and LL, (using Gaussian kernel) : (a) Model M2 - (b) Model M3.

3.4 Proofs of preliminary results

In this section, we study several preliminary results for $\widehat{g}_{1,h_1}$ and $\widehat{g}_{2,h_2}$ defined in (3.3). First of all, via the warping method, it is remarkable to notice that for $u_x = F_X(x)$,

$$\begin{aligned}\mathbb{E}\big(\widehat{g}_{1,h_1}(u_x)\big) = \mathbb{E}\big[\frac{1}{n}\sum_{k=1}^{n}\sin(\Theta_k).K_{h_1}\big(u_x - F_X(X_k)\big)\big] &= \mathbb{E}\Big[\mathbb{E}\big[\sin(\Theta)|X\big].K_{h_1}\big(u_x - F_X(X)\big)\Big] \\ &= \int_{\mathbb{R}} \mathbb{E}[\sin(\Theta)|X = y].K_{h_1}\big(u_x - F_X(y)\big).f_X(y)dy \\ &= \int_{\mathbb{R}} m_1(y).K_{h_1}\big(u_x - F_X(y)\big).f_X(y)dy \\ &= \int_{F_X(\mathbb{R})} g_1(w).K_{h_1}\big(u_x - w\big)dw.\end{aligned}$$

Then, recall that following the choice of $h_{\max}$ (see (3.7)), since $u_x = F_X(x) \in (0,1)$ is fixed and $K_{h_1}\big(u_x - w\big) = 0$ for $w \in (0,1)$ such that $\big|u_x - w\big| > A.h_{\max}$ (with $A > 0$ such that the support of K is included into $[-A, A]$), we have (see also (3.2) for the definition of convolution with g_1)

$$\int_{F_X(\mathbb{R})} g_1(w).K_{h_1}\big(u_x - w\big)dw = \int_{u_x - A.h_{\max}}^{u_x + A.h_{\max}} g_1(w).K_{h_1}\big(u_x - w\big)dw = \big(K_{h_1} * g_1\big)(u_x), \quad (3.20)$$

Thus, we obtain

$$\mathbb{E}\big(\widehat{g}_{1,h_2}(u_x)\big) = \big(K_{h_1} * g_1\big)(u_x).$$

and similarly,

$$\mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big) = \big(K_{h_2} * g_2\big)(u_x). \quad (3.21)$$

We obtain upper-bounds for pointwise study of their bias and variance.

Lemma 3.4.1. *Suppose that g_1 belongs to a Hölder class $\mathcal{H}(\beta_1, L_1)$, and g_2 belongs to $\mathcal{H}(\beta_2, L_2)$, with $L_1, L_2, \beta_1, \beta_2 \in \mathbb{R}_+^*$. Assume that the kernel K satisfies Assumption 3.2.5 with an index $\mathcal{L} \in \mathbb{R}_+$ such that $\mathcal{L} \geq \max(\beta_1, \beta_2)$ and a constant $C_{K,\mathcal{L}} > 0$. Then, for $u_x = F_X(x)$, and any $(h_1, h_2) \in \mathcal{H}_n^2$,*

$$\Big|\mathbb{E}\big(\widehat{g}_{1,h_1}(u_x)\big) - g_1(u_x)\Big| \leq L_1.h_1^{\beta_1}.C_{K,\mathcal{L}}, \quad \text{and} \quad Var\big(\widehat{g}_{1,h_1}(u_x)\big) \leq \frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_1},$$

$$\Big|\mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big) - g_2(u_x)\Big| \leq L_2.h_2^{\beta_2}.C_{K,\mathcal{L}}, \quad \text{and} \quad Var\big(\widehat{g}_{2,h_2}(u_x)\big) \leq \frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_2}.$$

A proof of this Lemma 3.4.1 is given in Section 3.4.1.
Now, we introduce below several events on which we will establish some concentration results for $\widehat{g}_{1,h_1}$ and $\widehat{g}_{2,h_2}$.

Definition 3.4.2. For $n \in \mathbb{N}^*$, $p \geq 1$, and $h_1, h_2 > 0$, we define for an arbitrary $v \in F_X(\mathbb{R})$ the following two events

$$\Omega_{1,n}(v, h_1) := \Big\{\big|\widehat{g}_{1,h_1}(v) - \mathbb{E}[\widehat{g}_{1,h_1}(v)]\big| \leq c_1(p).\sqrt{\widetilde{V}_1(n, h_1)}\Big\},$$

and

$$\Omega_{2,n}(v,h_2) := \left\{ \left|\widehat{g}_{2,h_2}(v) - \mathbb{E}[\widehat{g}_{2,h_2}(v)]\right| \leq c_1(p).\sqrt{\widetilde{V}_2(n,h_2)} \right\},$$

where $\widetilde{V}_1(n,h_1) = c_{0,1}.\dfrac{\log(n)}{n.h_1}.\|K\|^2_{\mathbb{L}^2(\mathbb{R})}$, $\widetilde{V}_2(n,h_2) = c_{0,2}.\dfrac{\log(n)}{n.h_2}.\|K\|^2_{\mathbb{L}^2(\mathbb{R})}$ (defined in (3.8) and (3.11) respectively), constants $c_{0,1}, c_{0,2} > 0$ and $c_1(p)$ satisfying $c_1(p).\sqrt{\min\{c_{0,1}; c_{0,2}\}} \geq 4p$. Furthermore, for $L_1, L_2, \beta_1, \beta_2 > 0$, we also introduce

$$E_{1,n}(v,h_1) := \left\{\left|\widehat{g}_{1,h_1}(v) - g_1(v)\right| \leq \Phi_1(n,h_1)\right\}, \text{ and } E_{2,n}(v,h_2) := \left\{\left|\widehat{g}_{2,h_2}(v) - g_2(v)\right| \leq \Phi_2(n,h_2)\right\}$$

where

$$\Phi_1(n,h_1) := c_1(p).\sqrt{\widetilde{V}_1(n,h_1)} + L_1.h_1^{\beta_1}.C_{K,\mathcal{L}},$$

and

$$\Phi_2(n,h_2) := c_1(p).\sqrt{\widetilde{V}_2(n,h_2)} + L_2.h_2^{\beta_2}.C_{K,\mathcal{L}}.$$

Then, the following proposition gives a concentration inequality for $\widehat{g}_{1,h_1}(u_x)$ and $\widehat{g}_{2,h_2}(u_x)$.

Proposition 3.4.3. *For $u_x = F_X(x)$, $p \geq 1$, $n \in \mathbb{N}$ and bandwidths $h_1, h_2 \in \mathcal{H}_n$ (defined in (3.7)), we have:*

$$\mathbb{P}\left(\left(\Omega_{1,n}(u_x,h_1)\right)^c\right) \leq 2.n^{-p}, \quad \text{and} \quad \mathbb{P}\left(\left(\Omega_{2,n}(u_x,h_2)\right)^c\right) \leq 2.n^{-p}.$$

Consequently, suppose further that g_1 belongs to a Hölder class $\mathcal{H}(\beta_1, L_1)$, and g_2 belongs to $\mathcal{H}(\beta_2, L_2)$, with $L_1, L_2, \beta_1, \beta_2 \in \mathbb{R}_+^*$ and the kernel K satisfies Assumption 3.2.5 with an index $\mathcal{L} \in \mathbb{R}_+$ such that $\mathcal{L} \geq \max(\beta_1, \beta_2)$, then we get:

$$\mathbb{P}\left(\left(E_{1,n}(u_x,h_1)\right)^c\right) \leq 2.n^{-p}, \quad \text{and} \quad \mathbb{P}\left(\left(E_{2,n}(u_x,h_2)\right)^c\right) \leq 2.n^{-p}.$$

A proof of Proposition 3.4.3 is given in Section 3.4.2.

3.4.1 Proof of Lemma 3.4.1

Proof. First, for the pointwise bias of $\widehat{g}_{1,h}(u_x)$ at $u_x = F_X(x)$, using (3.20) we can write

$$\begin{aligned}\mathbb{E}\big(\widehat{g}_{1,h_1}(u_x)\big) - g_1(u_x) &= \frac{1}{h_1}.\int_{u_x - A.h_{\max}}^{u_x + A.h_{\max}} K\Big(\frac{u_x - z}{h_1}\Big).g_1(z)dz - g_1(u_x) \\ &= \int_{-A}^{A} K(w).\big(g_1(u_x - h_1.w) - g_1(u_x)\big)dw. \qquad (3.22)\end{aligned}$$

Since g_1 belongs to a Hölder class $\mathcal{H}(\beta_1, L_1)$ with $L_1, \beta_1 \in \mathbb{R}_+^*$, using Taylor expansion for g_1 to get for $w \in [-A, A]$,

$$g_1(u_x - h_1.w) = g_1(u_x) + g_1'(u_x).(-h_1).w + \ldots + \frac{(-h_1.w)^{\lfloor\beta_1\rfloor}}{(\lfloor\beta_1\rfloor)!}.g_1^{(\lfloor\beta_1\rfloor)}(u_x - \tau.h_1.w), \quad 0 \leq \tau \leq 1,$$

then, under Assumption 3.2.5 with an index $\mathcal{L} \in \mathbb{R}_+$ satisfying $\mathcal{L} \geq \max(\beta_1, \beta_2)$, from (3.22) one gets

$$\begin{aligned}&\int_{\text{supp}(K)} K(w).\big(g_1(u_x - h_1.w) - g_1(u_x)\big)dw \\ &= \int_{\text{supp}(K)} K(w).\frac{(-h_1.w)^{\lfloor\beta_1\rfloor}}{(\lfloor\beta_1\rfloor)!}.g_1^{(\lfloor\beta_1\rfloor)}(u_x - \tau.h_1.w)dw \\ &= \int_{\text{supp}(K)} K(w).\frac{(-h_1.w)^{\lfloor\beta_1\rfloor}}{(\lfloor\beta_1\rfloor)!}.\Big(g_1^{(\lfloor\beta_1\rfloor)}(u_x - \tau.h_1.w) - g_1^{(\lfloor\beta_1\rfloor)}(u_x)\Big)dw.\end{aligned}$$

This implies that with $0 \leq \tau \leq 1$, with $C_{K,\mathcal{L}}$ the constant defined in Assumption 3.2.5 ,

$$\begin{aligned}\Big|\mathbb{E}\big(\widehat{g}_{1,h_1}(u_x)\big) - g_1(u_x)\Big| &\leq \int_{\mathrm{supp}(K)} |K(w)|.\frac{|h_1.w|^{\lfloor\beta_1\rfloor}}{(\lfloor\beta_1\rfloor)!}.\Big|g_1^{(\lfloor\beta_1\rfloor)}(u_x - \tau.h_1.w) - g_1^{(\lfloor\beta_1\rfloor)}(u_x)\Big|dw \\ &\leq \int_{\mathrm{supp}(K)} |K(w)|.\frac{|h_1.w|^{\lfloor\beta_1\rfloor}}{(\lfloor\beta_1\rfloor)!}.L_1.\big|\tau.h_1.w\big|^{\beta_1-\lfloor\beta_1\rfloor}dw \quad (\text{since } g_1 \in \mathcal{H}(\beta_1, L_1)) \\ &\leq L_1.h_1^{\beta_1}.\int_{\mathrm{supp}(K)} |K(w)|.|w|^{\beta_1}dw \\ &\leq L_1.h_1^{\beta_1}.\int_{\mathrm{supp}(K)} |K(w)|.(1+|w|)^{\beta_1}dw \leq L_1.h_1^{\beta_1}.C_{K,\mathcal{L}}.\end{aligned}$$

Similarly, we obtain an upper-bound for the pointwise bias of $\widehat{g}_{2,h_2}$ as

$$\Big|\mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big) - g_2(u_x)\Big| \leq L_2.h_2^{\beta_2}.C_{K,\mathcal{L}}.$$

+ For the pointwise variance of $\widehat{g}_{2,h_2}(u_x)$, one gets:

$$\begin{aligned}\mathbb{E}\Big[\Big(\widehat{g}_{2,h_2}(u_x) - \mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big)\Big)^2\Big] = \mathrm{Var}\big(\widehat{g}_{2,h_2}(u_x)\big) &= \mathrm{Var}\left(\frac{1}{n}.\sum_{k=1}^{n}\cos(\Theta_k).K_{h_2}\big(u_x - F_X(X_k)\big)\right) \\ &\leq \frac{1}{n}.\mathbb{E}\Big[\big(\cos(\Theta)\big)^2.\big[K_{h_2}\big(u_x - F_X(X)\big)\big]^2\Big] \\ &\leq \frac{1}{n}.\mathbb{E}\Big(\big[K_{h_2}\big(u_x - F_X(X)\big)\big]^2\Big) \leq \frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_2},\end{aligned}$$

and similarly, for the pointwise variance of $\widehat{g}_{1,h_1}(u_x)$ by

$$\begin{aligned}\mathrm{Var}\big(\widehat{g}_{1,h_1}(u_x)\big) = \mathrm{Var}\left(\frac{1}{n}.\sum_{k=1}^{n}\sin(\Theta_k).K_{h_1}\big(u_x - F_X(X_k)\big)\right) &\leq \frac{1}{n}.\mathbb{E}\Big[\big(\sin(\Theta)\big)^2.\big[K_{h_1}\big(u_x - F_X(X)\big)\big]^2\Big] \\ &\leq \frac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_1};\end{aligned}$$

This concludes the proof of Lemma 3.4.1.

□

3.4.2 A concentration inequality of $\widehat{g}_{1,h_1}(u_x)$ and $\widehat{g}_{2,h_2}(u_x)$: Proof of Proposition 3.4.3

We first state a version of Bernstein inequality (see [25, Lemma 2], which is useful in the sequel, as follows.

Lemma 3.4.4 (Bernstein inequality)**.** *Let $T_1, ..., T_n$ be i.i.d. random variables and $S_n = \sum_{j=1}^{n}\big[T_j - \mathbb{E}(T_j)\big]$. Then, for $\eta > 0$,*

$$\mathbb{P}\Big(\big|S_n\big| \geq n.\eta\Big) \leq 2.\max\left(\exp\Big(-\frac{n.\eta^2}{4.V}\Big), \exp\Big(-\frac{n.\eta}{4.b}\Big)\right),$$

with $Var(T_1) \leq V$ and $|T_1| \leq b$, where V and b are two positive deterministic constants.

Now, we can start to prove Proposition 3.4.3.

Proof of Proposition 3.4.3. We follow the procedure proposed in [25, Proposition 6]. First of all, we define random variables $Z_k(u_x) := \cos(\Theta_k).K_{h_2}\big(u_x - F_X(X_k)\big)$, for $1 \leq k \leq n$, and hence, $\widehat{g}_{2,h_2}(u_x) = \frac{1}{n}\sum_{k=1}^{n} Z_k(u_x)$. Notice that $\mathbb{E}(Z_k(u_x)) = \big(K_{h_2} * g_2\big)(u_x)$ (see (3.21)). Since $\|K\|_\infty < +\infty$, we then have for any $v \in F_X(\mathbb{R})$:

$$|Z_k(v)| = \big|\cos(\Theta_k).K_{h_2}\big(v - F_X(X_k)\big)\big| \leq \frac{\|K\|_\infty}{h_2} =: b(h_2) \quad , \tag{3.23}$$

and $\mathrm{Var}\big(Z_k(v)\big) = n.\mathrm{Var}\big(\widehat{g}_{2,h_2}(v)\big) \leq n.\dfrac{\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_2} =: n.V_0(n,h_2)$. Now, applying Bernstein inequality in Lemma 3.4.4 to the $Z_k(v)$'s, we obtain

$$\begin{aligned}\mathbb{P}\left(\Big|\widehat{g}_{2,h_2}(v) - \mathbb{E}\big(\widehat{g}_{2,h_2}(v)\big)\Big| \geq \eta(h_2)\right) &= \mathbb{P}\left(\Big|\sum_{k=1}^{n} Z_k(v) - \mathbb{E}\Big(Z_k(v)\Big)\Big| \geq n.\eta(h_2)\right) \\ &\leq 2.\max\left\{\exp\left(-\frac{n.\big(\eta(h_2)\big)^2}{4n.V_0(n,h_2)}\right); \exp\left(-\frac{n.\eta(h_2)}{4b(h_2)}\right)\right\}.\end{aligned}$$

For $p \geq 1$, choose $\eta(h_2) = c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}$, with $\widetilde{V}_2(n,h_2) = c_{0,2}.\log(n).V_0(n,h_2)$, a constant $c_{0,2} > 0$, then

$$\begin{aligned}&\mathbb{P}\left(\Big|\widehat{g}_{2,h_2}(v) - \mathbb{E}\big(\widehat{g}_{2,h_2}(v)\big)\Big| \geq c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}\right) \\ &\leq 2.\max\left\{\exp\left(-\frac{n.c_1(p)^2.\widetilde{V}_2(n,h_2)}{4n.V_0(n,h_2)}\right); \exp\left(-\frac{n.c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}}{4b(h_2)}\right)\right\}.\end{aligned} \tag{3.24}$$

By choosing $c_1(p)$, we intend to maximize the right-hand side in the inequality (3.24).
+ First, $c_1(p)$ is chosen such that

$$\frac{n.c_1(p)^2.\widetilde{V}_2(n,h_2)}{4n.V_0(n,h_2)} = \frac{n.c_1(p)^2.c_{0,2}.\log(n).V_0(n,h_2)}{4n.V_0(n,h_2)} \geq p.\log(n), \tag{3.25}$$

that is $c_1(p)$ satisfying $c_1(p)^2.c_{0,2} \geq 4p$.
+ Secondly, we can write

$$\frac{n.c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}}{4b(h_2)} = \frac{c_1(p).\sqrt{c_{0,2}}}{4}.\sqrt{\log(n)}.\frac{n.\sqrt{V_0(n,h_2)}}{b(h_2)}$$

and for $h_2 \in \mathcal{H}_n$,

$$n.\frac{\sqrt{V_0(n,h_2)}}{b(h_2)} = n.\frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}}{\sqrt{n.h_2}}.\frac{h_2}{\|K\|_\infty} = \sqrt{n.h_2}.\frac{\|K\|_{\mathbb{L}^2(\mathbb{R})}}{\|K\|_\infty} > \sqrt{\log(n)},$$

then, we have

$$\frac{n.c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}}{4b(h_2)} = \frac{c_1(p).\sqrt{c_{0,2}}}{4}.\sqrt{\log(n)}.\frac{n.\sqrt{V_0(n,h_2)}}{b(h_2)} \geq \frac{c_1(p).\sqrt{c_{0,2}}}{4}.\log(n) \geq p.\log(n), \tag{3.26}$$

provided that $c_1(p).\sqrt{c_{0,2}} \geq 4p$. Note now that this condition also ensures the constraint $c_1(p)^2.c_{0,2} \geq 4p$.

Now, combining (3.25) and (3.26), we get from (3.24) for any $p \geq 1$:

$$\mathbb{P}\left(\left(\Omega_{2,n}(u_x,h_2)\right)^c\right) := \mathbb{P}\left(\left|\widehat{g}_{2,h_2}(u_x) - \mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big)\right| \geq c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}\right) \leq 2.n^{-p};$$

This implies that with a probability larger than $\big(1-2.n^{-p}\big)$ we have:

$$\left|\widehat{g}_{2,h_2}(u_x) - \mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big)\right| \leq c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}.$$

Then, with a probability larger than $\big(1-2.n^{-p}\big)$, we can obtain

$$\begin{aligned}\left|\widehat{g}_{2,h_2}(u_x) - g_2(u_x)\right| &\leq \left|\widehat{g}_{2,h_2}(u_x) - \mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big)\right| + \left|\mathbb{E}\big(\widehat{g}_{2,h_2}(u_x)\big) - g_2(u_x)\right| \\ &\leq c_1(p).\sqrt{\widetilde{V}_2(n,h_2)} \quad + \quad L_2.h_2^{\beta_2}.C_{K,\mathcal{L}}.\end{aligned}$$

Recall that $\Phi_2(n,h_2) = c_1(p).\sqrt{\widetilde{V}_2(n,h_2)} + L_2.h_2^{\beta_2}.C_{K,\mathcal{L}}$, therefore, we finally attain for $p \geq 1$ that

$$\mathbb{P}\left(\left(E_{2,n}(u_x,h_2)\right)^c\right) = \mathbb{P}\left(\left|\widehat{g}_{2,h_2}(u_x) - g_2(u_x)\right| \geq \Phi_2(n,h_2)\right) \leq 2.n^{-p}. \tag{3.27}$$

By similar arguments we obtain the desired result for $\widehat{g}_{1,h_1}$. This concludes the proof of Proposition 3.4.3.

□

3.5 Proofs of main results

3.5.1 Proof of Proposition 3.2.2

We first have the following concentration result

Corollary 3.5.1. *Under the Assumptions of Proposition 3.2.2, for all $h_1, h_1', h_2, h_2' \in \mathcal{H}_n$, for all $p \geq 1$, for $u_x = F_X(x)$,*

$$\mathbb{P}\left(\left|\widehat{g}_{1,h_1,h_1'}(u_x) - \mathbb{E}\big[\widehat{g}_{1,h_1,h_1'}(u_x)\big]\right| > c_1(p).\left\|K\right\|_{\mathbb{L}^1(\mathbb{R})}.\sqrt{\widetilde{V}_1(n,h_1')}\right) \leq 2.n^{-p},$$

and

$$\mathbb{P}\left(\left|\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\right| > c_1(p).\left\|K\right\|_{\mathbb{L}^1(\mathbb{R})}.\sqrt{\widetilde{V}_2(n,h_2')}\right) \leq 2.n^{-p},$$

with $c_1(p)$ satisfying $c_1(p).\sqrt{\min\{c_{0,1};c_{0,2}\}} \geq 4p$.

Proof of Corollary 3.5.1. We define random variables $\widetilde{Z}_k(u_x) := \cos(\Theta_k).\big(K_{h_2'} * K_{h_2}\big)\big(u_x - F_X(X_k)\big)$, for $1 \leq k \leq n$, and hence, $\widehat{g}_{2,h_2,h_2'}(u_x) = \dfrac{1}{n}\sum_{k=1}^n \widetilde{Z}_k(u_x)$. Since $\|K\|_\infty < +\infty$, we then have:

$$\left|\widetilde{Z}_k(u_x)\right| = \left|\cos(\Theta_k).\big(K_{h_2'} * K_{h_2}\big)\big(u_x - F_X(X_k)\big)\right| \leq \frac{\|K\|_\infty.\|K\|_{\mathbb{L}^1(\mathbb{R})}}{h_2} =: \widetilde{b}(h_2) \quad ,$$

and $\mathrm{Var}\big(\widetilde{Z}_k(u_x)\big) = n.\mathrm{Var}\big(\widehat{g}_{2,h_2,h_2'}(u_x)\big) \leq n.\dfrac{\|K\|^2_{\mathbb{L}^1(\mathbb{R})}.\|K\|^2_{\mathbb{L}^2(\mathbb{R})}}{n.h_2} =: n.\widetilde{V}_0(n,h_2)$.
Using similar arguments in the proof of Proposition 3.4.3, we obtain with a probability greater than $1-2.n^{-p}$ that

$$\left|\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\right| \leq \left\|K\right\|_{\mathbb{L}^1(\mathbb{R})}.c_1(p).\sqrt{\widetilde{V}_2(n,h_2)}.$$

Similarly, we get the desired concentration for $\big|\widehat{g}_{1,h_1,h_1'}(u_x) - \mathbb{E}\big[\widehat{g}_{1,h_1,h_1'}(u_x)\big]\big|$. This concludes the proof of Corollary 3.5.1.

□

Now, we can start to prove the Proposition 3.2.2.

Proof of Proposition 3.2.2. We follow the strategy proposed in [25, Theorem 2]. The target is to find an upper-bound for $\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\big|$. Let $h_2 \in \mathcal{H}_n$ be fixed.
■ We consider the following decomposition:

$$\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\big| \leq \underbrace{\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - \widehat{g}_{2,h_2,\widehat{h}_2}(u_x)\big|}_{=:I_{g_2,1}} + \underbrace{\big|\widehat{g}_{2,h_2,\widehat{h}_2}(u_x) - \widehat{g}_{2,h_2}(u_x)\big|}_{=:I_{g_2,2}} + \big|\widehat{g}_{2,h_2}(u_x) - g_2(u_x)\big|.$$

By the definition of $A_2(h_2, u_x)$, we have

$$\begin{aligned} I_{g_2,1} = \big|\widehat{g}_{2,\widehat{h}_2}(u_x) - \widehat{g}_{2,h_2,\widehat{h}_2}(u_x)\big| &= \big|\widehat{g}_{2,\widehat{h}_2}(u_x) - \widehat{g}_{2,h_2,\widehat{h}_2}(u_x)\big| - \sqrt{\widetilde{V}_2(n,\widehat{h}_2)} + \sqrt{\widetilde{V}_2(n,\widehat{h}_2)} \\ &\leq \sup_{h_2' \in \mathcal{H}_n} \Big[\big|\widehat{g}_{2,h_2'}(u_x) - \widehat{g}_{2,h_2,h_2'}(u_x)\big| - \sqrt{\widetilde{V}_2(n,h_2')}\Big]_+ + \sqrt{\widetilde{V}_2(n,\widehat{h}_2)} \\ &= A_2(h_2,u_x) + \sqrt{\widetilde{V}_2(n,\widehat{h}_2)}; \end{aligned}$$

And similarly, by the definition of $A_2(\widehat{h}_2, u_x)$,

$$\begin{aligned} I_{g_2,2} = \big|\widehat{g}_{2,h_2,\widehat{h}_2}(u_x) - \widehat{g}_{2,h_2}(u_x)\big| &\leq \sup_{h_2' \in \mathcal{H}_n} \Big[\big|\widehat{g}_{2,h_2',\widehat{h}_2}(u_x) - \widehat{g}_{2,h_2'}(u_x)\big| - \sqrt{\widetilde{V}_2(n,h_2')}\Big]_+ + \sqrt{\widetilde{V}_2(n,h_2)} \\ &= A_2(\widehat{h}_2,u_x) + \sqrt{\widetilde{V}_2(n,h_2)}; \end{aligned}$$

Therefore, by using the definition of $\widehat{h}_2$, we get

$$I_{g_2,1} + I_{g_2,2} \leq A_2(h_2,u_x) + \sqrt{\widetilde{V}_2(n,\widehat{h}_2)} + A_2(\widehat{h}_2,u_x) + \sqrt{\widetilde{V}_2(n,h_2)} \leq 2.\Big[A_2(h_2,u_x) + \sqrt{\widetilde{V}_2(n,h_2)}\Big].$$

Hence, we obtain

$$\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\big| \leq 2.A_2(h_2,u_x) + 2.\sqrt{\widetilde{V}_2(n,h_2)} + \big|\widehat{g}_{2,h_2}(u_x) - g_2(u_x)\big|. \tag{3.28}$$

■ Now, to study $A_2(h_2, u_x)$, we can write:

$$\begin{aligned} \widehat{g}_{2,h_2'}(u_x) - \widehat{g}_{2,h_2,h_2'}(u_x) = \widehat{g}_{2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big] - \big(\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\big) \\ + \mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big] - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big], \end{aligned}$$

and, we have $\mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big] = \big(K_{h_2'} * g_2\big)(u_x)$ as well as $\mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big] = \mathbb{E}\big[K_{h_2'} * \widehat{g}_{2,h_2}(u_x)\big] = \big(K_{h_2'} * K_{h_2} * g_2\big)(u_x)$.
Thus,

$$\begin{aligned} \big|\widehat{g}_{2,h_2'}(u_x) - \widehat{g}_{2,h_2,h_2'}(u_x)\big| - \sqrt{\widetilde{V}_2(n,h_2')} \leq \big|\widehat{g}_{2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big]\big| - \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ + \big|\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\big| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ + \big|\mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big] - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\big|; \end{aligned}$$

However, for any $h_2' \in \mathcal{H}_n$,

$$\left|\mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big] - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\right| = \left|K_{h_2'} * \big(g_2 - K_{h_2} * g_2\big)(u_x)\right| \leq \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \|g_2 - K_{h_2} * g_2\|_\infty .$$

Hence, back to the definition of $A_2(h_2, u_x)$, we obtain

$$\begin{aligned} A_2(h_2,u_x) &= \sup_{h_2' \in \mathcal{H}_n} \Big[\left|\widehat{g}_{2,h_2'}(u_x) - \widehat{g}_{2,h_2,h_2'}(u_x)\right| - \sqrt{\widetilde{V}_2(n,h_2')}\Big]_+ \\ &\leq \sup_{h_2' \in \mathcal{H}_n} \Big[\left|\widehat{g}_{2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big]\right| - \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\Big]_+ \qquad (3.29) \\ &+ \sup_{h_2' \in \mathcal{H}_n} \Big[\left|\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\right| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\Big]_+ \qquad (3.30) \\ &\qquad + \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \|g_2 - K_{h_2} * g_2\|_\infty ; \end{aligned}$$

We now need concentration results to control $A_2(h_2, u_x)$ completely. From Corollary 3.5.1, for $h_2, h_2' \in \mathcal{H}_n$,

$$\mathbb{P}\left(\left|\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\right| > c_1(p).\|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \sqrt{\widetilde{V}_2(n,h_2')}\right) \leq 2.n^{-p};$$

It implies that if $c_1(p) = \dfrac{1}{1+\|K\|_{\mathbb{L}^1(\mathbb{R})}}$ and $c_{0,2} \geq 16p^2.\big(1+\|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$, then

$$\mathbb{P}\left(\sup_{h_2' \in \mathcal{H}_n} \Big[\left|\widehat{g}_{2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big]\right| - \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\Big]_+ > 0\right) \leq 2. \sum_{h_2 \in \mathcal{H}_n} n^{-p} \leq 2.n^{1-p},$$

as $\mathrm{Card}(\mathcal{H}_n) \leq n$. In the same way for all $h_2 \in \mathcal{H}_n$,

$$\mathbb{P}\left(\sup_{h_2' \in \mathcal{H}_n} \Big[\left|\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\right| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\Big]_+ > 0\right) \leq 2.n^{1-p};$$

Consequently, the following set

$$\begin{aligned} \widetilde{A}_2 := &\left\{\sup_{h_2' \in \mathcal{H}_n} \Big[\left|\widehat{g}_{2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2'}(u_x)\big]\right| - \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\Big]_+ = 0\right\} \\ &\cap \left\{\forall\, h_2 \in \mathcal{H}_n,\ \sup_{h_2' \in \mathcal{H}_n} \Big[\left|\widehat{g}_{2,h_2,h_2'}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2,h_2'}(u_x)\big]\right| - \|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \frac{\sqrt{\widetilde{V}_2(n,h_2')}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})}\Big]_+ = 0\right\}, \end{aligned}$$

has probability greater than $(1 - 4.n^{2-p})$. Now, choose $p = 2 + q$ and then $c_{0,2} \geq 16\big(2+q\big)^2.\big(1+\|K\|_{\mathbb{L}^1(\mathbb{R})}\big)^2$. Thus, we obtain that $\mathbb{P}\big(\widetilde{A}_2\big) > 1 - 4.n^{-q}$.
Combining inequalities (3.28) and (3.29), we have on $\widetilde{A}_2$:

$$\begin{aligned} \left|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\right| &\leq 2.A_2(h_2,u_x) + 2.\sqrt{\widetilde{V}_2(n,h_2)} + \left|\widehat{g}_{2,h_2}(u_x) - g_2(u_x)\right| \\ &\leq 2.\|K\|_{\mathbb{L}^1(\mathbb{R})} \cdot \|g_2 - K_{h_2} * g_2\|_\infty + 2.\sqrt{\widetilde{V}_2(n,h_2)} + \left|\widehat{g}_{2,h_2}(u_x) - g_2(u_x)\right|, \end{aligned}$$

but still on $\widetilde{A}_2$, one gets $\big|\widehat{g}_{2,h_2}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2}(u_x)\big]\big| - \dfrac{\sqrt{\widetilde{V}_2(n,h_2)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})} \leq 0$, so

$$\begin{aligned}\big|\widehat{g}_{2,h_2}(u_x) - g_2(u_x)\big| &\leq \big|\mathbb{E}\big[\widehat{g}_{2,h_2}(u_x)\big] - g_2(u_x)\big| + \big|\widehat{g}_{2,h_2}(u_x) - \mathbb{E}\big[\widehat{g}_{2,h_2}(u_x)\big]\big| - \frac{\sqrt{\widetilde{V}_2(n,h_2)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &\qquad + \frac{\sqrt{\widetilde{V}_2(n,h_2)}}{(1+\|K\|_{\mathbb{L}^1(\mathbb{R})})} \\ &\leq \|g_2 - K_{h_2} * g_2\|_\infty + \sqrt{\widetilde{V}_2(n,h_2)};\end{aligned}$$

Therefore, on $\widetilde{A}_2$, we finally obtain

$$\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\big| \leq (1 + 2\,\|K\|_{\mathbb{L}^1(\mathbb{R})}).\,\|g_2 - K_{h_2} * g_2\|_\infty + 3.\sqrt{\widetilde{V}_2(n,h_2)}.$$

By similar arguments, we obtain the desired result for $\widehat{g}_{1,\widehat{h}_1}(u_x)$. This concludes the proof of Proposition 3.2.2.

□

3.5.2 Proof of Theorem 3.2.6

First, we establish a concentration result for $\widehat{g}_{1,\widehat{h}_1}(u_x)$ and $\widehat{g}_{2,\widehat{h}_2}(u_x)$ as follows:

Corollary 3.5.2. *For $\beta_1, \beta_2, L_1, L_2 > 0$, suppose that g_1 belongs to a Hölder class $\mathcal{H}(\beta_1, L_1)$, g_2 belongs to $\mathcal{H}(\beta_2, L_2)$ and under the Assumptions in Proposition 3.2.2, for $q \geq 1$,*

$$\begin{aligned}\mathbb{P}\Big(\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c\Big) &:= \mathbb{P}\Big(\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| > C_1.\psi_n(\beta_1)\Big) \leq 4.n^{-q}, \\ \mathbb{P}\Big(\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\Big) &:= \mathbb{P}\Big(\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\big| > C_2.\psi_n(\beta_2)\Big) \leq 4.n^{-q},\end{aligned}$$

where C_1 is a constant (depending on $\beta_1, L_1, c_{0,1}, K$) and C_2 is a constant (depending on $\beta_2, L_2, c_{0,2}, K$), along with $\psi_n(\beta_1) = \big(\log(n)/n\big)^{\frac{\beta_1}{2\beta_1+1}}$ and $\psi_n(\beta_2) = \big(\log(n)/n\big)^{\frac{\beta_2}{2\beta_2+1}}$.

Proof of Corollary 3.5.2. Since g_1 belongs to a Hölder class $\mathcal{H}(\beta_1, L_1)$, from Lemma 3.4.1 we have

$$\|g_1 - K_{h_1} * g_1\|_\infty \leq L_1.h_1^{\beta_1}.C_{K,\mathcal{L}}.$$

From Proposition 3.2.2, this implies that with a probability greater than $1 - 4.n^{-q}$, one gets for any $h_1 \in \mathcal{H}_n$:

$$\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| \leq \big(1 + 2\,\|K\|_{\mathbb{L}^1(\mathbb{R})}\big).L_1.h_1^{\beta_1}.C_{K,\mathcal{L}} + 3.\sqrt{\widetilde{V}_1(n,h_1)}. \tag{3.31}$$

In (3.31), we take h_1 so that h_1^{-1} is an integer and h_1 of order $\Big(\dfrac{\log(n)}{n}\Big)^{\frac{1}{2\beta_1+1}}$. Since $1/(2\beta_1 + 1) < 1$, $h_1 \in \mathcal{H}_n$, for n large enough. As a result, we obtain with probability greater than $1 - 4.n^{-q}$, that

$$\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| \leq C_1.\psi_n(\beta_1),$$

with a constant C_1 (depending on $\beta_1, L_1, c_{0,1}, K$) and $\psi_n(\beta_1) = \big(\log(n)/n\big)^{\frac{\beta_1}{2\beta_1+1}}$.
By similar arguments, we obtain the desired result for $\widehat{g}_{2,\widehat{h}_2}(u_x)$. This concludes the proof of Corollary 3.5.2.

□

Now, we start to prove Theorem 3.2.6.

Proof of Theorem 3.2.6. For given $x \in \mathbb{R}$, at $u_x = F_X(x)$, we have $\mathbb{E}\Big[\big|\widehat{m}_{\widehat{h}}(x) - m(x)\big|^2\Big] = \mathbb{E}\Big[\big|\widehat{g}_{\widehat{h}}(u_x) - g(u_x)\big|^2\Big]$. First, on the event $\widetilde{E}_{2,n}(u_x, \widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x, \widehat{h}_1)$, for n sufficiently large satisfying $C_2.\psi_n(\beta_2) \leq \delta_2/2$ and $C_1.\psi_n(\beta_1) \leq \delta_1/2$, we have $\big|\widehat{g}_{2,\widehat{h}_2}(u_x) - g_2(u_x)\big| \leq C_2.\psi_n(\beta_2) < \dfrac{\big|g_2(u_x)\big|}{2}$ and $\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| \leq C_1.\psi_n(\beta_1) < \dfrac{\big|g_1(u_x)\big|}{2}$. Thus, we get $\widehat{g}_{2,\widehat{h}_2}(u_x).g_2(u_x) > 0$ and $\widehat{g}_{1,\widehat{h}_1}(u_x).g_1(u_x) > 0$. Then, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\big|\widehat{g}_{\widehat{h}}(u_x) - g(u_x)\big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&= \mathbb{E}\Big[\Big|\operatorname{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x), \widehat{g}_{2,\widehat{h}_2}(u_x)\big) - \operatorname{atan2}\big(g_1(u_x), g_2(u_x)\big)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&= \mathbb{E}\Big[\Big|\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) - \arctan\Big(\frac{g_1(u_x)}{g_2(u_x)}\Big)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&\leq 2.\mathbb{E}\Big[\Big|\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) - \arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \qquad (3.32) \\
&\quad + 2.\mathbb{E}\Big[\Big|\arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) - \arctan\Big(\frac{g_1(u_x)}{g_2(u_x)}\Big)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big];
\end{aligned}
$$

• For n sufficiently large, $\big|\widehat{g}_{2,\widehat{h}_2}(u_x)\big| \geq \big|g_2(u_x)\big| - \big|g_2(u_x) - \widehat{g}_{2,h}(u_x)\big| > \delta_2 - C_2.\psi_n(\beta_2) \geq \delta_2/2$ on the event $\widetilde{E}_{2,n}(u_x, \widehat{h}_2)$, and using the 1-Lipschitz continuity of arctan, we get for the first term in (3.32):

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|\arctan\Big(\frac{\widehat{g}_{1,\widehat{h}_1}(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big) - \arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&\leq \frac{4}{\delta_2^2}.\mathbb{E}\Big[\big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&\leq \frac{4}{\delta_2^2}.\mathbb{E}\Big[C_1^2.\psi_n(\beta_1)^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big] \\
&\big(\text{since on } \widetilde{E}_{1,n}(u_x,\widehat{h}_1) \text{ one has } \big|\widehat{g}_{1,\widehat{h}_1}(u_x) - g_1(u_x)\big| \leq C_1.\psi_n(\beta_1)\big) \\
&\leq \frac{4}{\delta_2^2}.C_1^2.\psi_n(\beta_1)^2.\mathbb{P}\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2) \cap \widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big) \\
&\leq \frac{4}{\delta_2^2}.C_1^2.\psi_n(\beta_1)^2.
\end{aligned}
$$

• Moreover, for the second term in (3.32), we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\arctan\Big(\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}\Big)-\arctan\Big(\frac{g_1(u_x)}{g_2(u_x)}\Big)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\\
&=\mathbb{E}\Big[\Big|\arctan\Big(\frac{\widehat{g}_{2,\widehat{h}_2}(u_x)}{g_1(u_x)}\Big)-\arctan\Big(\frac{g_2(u_x)}{g_1(u_x)}\Big)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\\
&\big(\text{since }\frac{g_1(u_x)}{\widehat{g}_{2,\widehat{h}_2}(u_x)}.\frac{g_1(u_x)}{g_2(u_x)}>0\text{ on }\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)\\
&\leq\frac{1}{|g_1(u_x)|^2}.\mathbb{E}\Big[\Big|\widehat{g}_{2,\widehat{h}_2}(u_x)-g_2(u_x)\Big|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\\
&\leq\frac{1}{\delta_1^2}.\mathbb{E}\Big[C_2^2.\psi_n(\beta_2)^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\quad\big(\text{using Corollary 3.5.2}\big)\\
&\leq\frac{1}{\delta_1^2}.C_2^2.\psi_n(\beta_2)^2;
\end{aligned}$$

Therefore, on the event $\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)$, n sufficiently large such that $C_2.\psi_n(\beta_2)\leq\delta_2/2$ and $C_1.\psi_n(\beta_1)\leq\delta_1/2$, we obtain

$$\mathbb{E}\Big[|\widehat{g}_{\widehat{h}}(u_x)-g(u_x)|^2.\mathbb{1}_{\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\cap\widetilde{E}_{1,n}(u_x,\widehat{h}_1)}\Big]\leq 8.\Big(\frac{1}{\delta_2^2}+\frac{1}{\delta_1^2}\Big).(C_1^2+C_2^2).\max\Big(\psi_n(\beta_1)^2,\psi_n(\beta_2)^2\Big);$$

■ On the other hand, on the complementary $\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\cup\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c$, using the fact that $\big|\text{atan2}(w_1,w_2)\big|\leq\pi$, $\forall(w_1,w_2)$, we can simply obtain an upper-bound as follows:

$$\begin{aligned}
&\mathbb{E}\Big[|\widehat{g}_{\widehat{h}}(u_x)-g(u_x)|^2.\mathbb{1}_{\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\cup\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c}\Big]\\
&=\mathbb{E}\Big[\Big|\text{atan2}\big(\widehat{g}_{1,\widehat{h}_1}(u_x),\widehat{g}_{2,\widehat{h}_2}(u_x)\big)-\text{atan2}\big(g_1(u_x),g_2(u_x)\big)\Big|^2.\mathbb{1}_{\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\cup\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c}\Big]\\
&\leq 4\pi^2.\mathbb{P}\Big(\big(\widetilde{E}_{2,n}(u_x,\widehat{h}_2)\big)^c\Big)+4\pi^2.\mathbb{P}\Big(\big(\widetilde{E}_{1,n}(u_x,\widehat{h}_1)\big)^c\Big)\leq 2.4\pi^2.4.n^{-q}\quad\text{(by Corollary 3.5.2)};
\end{aligned}$$

For $q\geq 1$, we get that n^{-q} is negligible in comparison with $C_1^2.\psi_n(\beta_1)^2=C_1^2.\Big(\frac{\log(n)}{n}\Big)^{\frac{2\beta_1}{2\beta_1+1}}$ and $C_2^2.\psi_n(\beta_2)^2=C_2^2.\Big(\frac{\log(n)}{n}\Big)^{\frac{2\beta_2}{2\beta_2+1}}$.
This concludes the proof of Theorem 3.2.6.

□

3.6 Conclusion

For further perspective, first of all, we would like to obtain a lower-bound of the circular regression problem presented in this chapter. Secondly, as a natural extension, it could be very challenging to investigate our regression problem with a response on the sphere $\mathbb{S}^2$ or more generally on the unit hypersphere $\mathbb{S}^{d-1}$. The case of predictors $X\in\mathbb{S}^{d-1}$ and a response $\Theta\in\mathbb{S}^{d-1}$ has been tackled in [81]. The spherical context is of course more complicated than the circular one and the arctangent function approach used here is not easily generalizable in the spherical context. In [81], Di Marzio et al. proposed a local constant estimator by smoothing on each component of the response. Once again no rates of convergence were obtained. Hence, in a future work, a first task would be to obtain convergence rates and then investigate adaptation issue.

Appendix A

Preliminaries

A.1.1 Denjoy-Wolff fixed-point theorem

We use the statements on Denjoy-Wolff fixed-point Theorem as in [4]. Denote by $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ the unit disk in the complex plane, and let $f : \mathbb{D} \longrightarrow \overline{\mathbb{D}}$ be an analytic function. A point $\omega \in \overline{\mathbb{D}}$ is called *Denjoy-Wolff point* for f, if either

(i) $\omega \in \mathbb{D}$ and $f(\omega) = \omega$; or

(ii) $|\omega| = 1$, $\lim_{r\uparrow 1} f(r\omega) = \omega$ and $\lim_{r\uparrow 1} \dfrac{\omega f(r\omega)}{(1-r)\omega} \leq 1$.

Except for the identity map of $\mathbb{D}$, every function f has a unique Denjoy-Wolff point. Moreover, this point is a limit of the iterates of f in most cases.

Theorem A.1.1. *Consider a domain Δ conformally equivalent to $\mathbb{D}$, an open subset $\Omega \subset \mathbb{C}$ and an analytic function $g : \Omega \times \Delta \to \Delta$ such that for some $z \in \Omega$, the map $g(z, v)$ is not a conformal automorphism of Δ and has a fixed point in Δ. Then, there exists an analytic function $\omega : \Omega \to \Delta$ such that $g(z, \omega(z)) = \omega(z), z \in \Omega$. Moreover, for every $u \in \Delta$,*

$$\omega(z) = \lim_{m\to\infty} g^{\circ m}(z, u),$$

uniformly on compact subsets of Δ.

A.1.2 Cauchy transform of semicircular distribution

For $t > 0$, we now calculate G_{σ_t} the Cauchy transform of σ_t (defined in (2.3)) using a generating function (see e.g. [45], [1]). Indeed, σ_t has compact support, so we can expand G_{σ_t} into a power series. To do that, we will compute first, for $k \in \mathbb{N}^*$, the k-th moments of a random variable $Y \sim \sigma_t$ defined by

$$m_k(t) := \mathbb{E}\left(Y^k\right) = \int_{-2\sqrt{t}}^{2\sqrt{t}} x^k f_{\sigma_t}(x)dx,$$

with $m_0(t) = 1$. Then,

- for k odd: since the law σ_t is symmetric about 0 we have $m_k(t) = 0$;
- for k even: $m_k(t) = C_{\frac{k}{2}}.t^{\frac{k}{2}}$, where C_k is so-called the Catalan number $C_k = \dfrac{1}{k+1}\binom{2k}{k}$.

Indeed, we have

$$m_{2k}(t) = \int_{-2\sqrt{t}}^{2\sqrt{t}} x^{2k}.\frac{1}{2\pi t}\sqrt{4t-x^2}dx = (2\sqrt{t})^{2k}.\int_{-1}^{1} u^{2k}.\frac{2\sqrt{t}}{2\pi t}.\sqrt{1-u^2}.2\sqrt{t}du$$
$$= \frac{2.(2\sqrt{t})^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\big(\sin(\theta)\big)^{2k}.\big(\cos(\theta)\big)^2 d\theta$$
$$= \frac{2.(2\sqrt{t})^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\big(\sin(\theta)\big)^{2k}d\theta - \frac{2.(2\sqrt{t})^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\big(\sin(\theta)\big)^{2k+2}d\theta;$$

On the other hand, using integration by parts we get

$$\int_{-\pi/2}^{\pi/2}\big(\sin(\theta)\big)^{2k+1}.\sin(\theta)d\theta = -\Big[\big(\sin(\theta)\big)^{2k+1}.\cos(\theta)\Big]\Big|_{\theta=-\pi/2}^{\pi/2} + \int_{-\pi/2}^{\pi/2}(2k+1).\big(\sin(\theta)\big)^{2k}.\big(\cos(\theta)\big)^2 d\theta$$
$$= (2k+1).\int_{-\pi/2}^{\pi/2}\big(\sin(\theta)\big)^{2k}.\big(\cos(\theta)\big)^2 d\theta;$$

and so,

$$\frac{2.(2\sqrt{t})^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\big(\sin(\theta)\big)^{2k+2}d\theta = (2k+1).m_{2k}(t).$$

Thus, we obtain

$$m_{2k}(t) = (2\sqrt{t})^2.(2k-1).m_{2k-2}(t) - (2k+1).m_{2k}(t) \quad ,\text{or} \quad m_{2k}(t) = \frac{4t(2k-1)}{2k+2}.m_{2k-2}(t),$$

and then using induction as $m_{2k-2}(t) = C_{k-1}.(\sqrt{t})^{2k-2}$ with $m_0(t) = 1$ to get

$$m_{2k}(t) = \frac{4t(2k-1)}{2k+2}.\frac{(\sqrt{t})^{2k-2}.(2k-2)!}{k!(k-1)!} = (\sqrt{t})^{2k}.\frac{2.2k.(2k-1).(2k-2)!}{2(k+1)k.k!(k-1)!}$$
$$= (\sqrt{t})^{2k}.\frac{(2k)!}{(k+1)!k!} = t^k.C_k,$$

this concludes the computation for $m_k(t)$ with k even.
Now, since the support of σ_t is compact, for $z \in \mathbb{C}^+$ (and so $z \notin [-2\sqrt{t}, 2\sqrt{t}]$) such that $|z|^{-1} > 2\sqrt{t}$, we can write:

$$G_{\sigma_t}\left(\frac{1}{z}\right) = \int_{-2\sqrt{t}}^{2\sqrt{t}}\frac{z}{1-zx}.f_{\sigma_t}(x)dx = z.\sum_{j=0}^{+\infty}z^j.\int_{-2\sqrt{t}}^{2\sqrt{t}}x^j.f_{\sigma_t}(x)dx = z.\sum_{j=0}^{+\infty}z^j.m_j(t) = z.\sum_{j=0}^{+\infty}z^{2j}.m_{2j}(t).$$

Let $M(z)$ for $z \in \mathbb{C}^+$ such that $|z|^{-1} > 2\sqrt{t}$, be the power series defined by:

$$M(z) := \frac{G_{\sigma_t}(\frac{1}{z})}{z} = \sum_{j=0}^{+\infty}m_{2j}(t).z^{2j} = m_0(t) + \sum_{j=1}^{\infty}m_{2j}(t).z^{2j} = 1 + C_1.t.z^2 + C_2.t^2.z^4 + ... \quad ,$$

then

$$(M(z))^2 = \sum_{m,j\geq 0}C_m.C_j.t^{m+j}.z^{2(m+j)} = \sum_{k\geq 0}\left(\sum_{m+j=k}C_m.C_j\right).t^k.z^{2k}.$$

Furthermore, we admit a result, namely that $\sum_{m+j=k} C_m.C_j = C_{k+1}$ (which is known as the recurrence property of Catalan numbers, see e.g. [1, section 2.1.1]), thus:

$$(M(z))^2 = \sum_{k\geq 0} C_{k+1}.t^k.z^{2k} = \frac{1}{t.z^2}.\sum_{k\geq 0} C_{k+1}.t^{k+1}.z^{2(k+1)},$$

and therefore, we obtain

$$t.z^2.(M(z))^2 = M(z) - 1, \quad \text{or} \quad t.z^2.(M(z))^2 - M(z) + 1 = 0. \tag{A.1}$$

Consequently, replacing $M(z)$ by $\dfrac{G_{\sigma_t}(\frac{1}{z})}{z}$ into (A.1), i.e. $t.\big(G_{\sigma_t}(\frac{1}{z})\big)^2 - \dfrac{1}{z}.G_{\sigma_t}(\frac{1}{z}) + 1 = 0$. Then we get that G_{σ_t} satisfies the following quadratic equation

$$t.\big(G_{\sigma_t}(z)\big)^2 - z.G_{\sigma_t}(z) + 1 = 0.$$

Solving this quadratic equation, we find the solutions

$$G_{\sigma_t}(z) = \frac{z \pm \sqrt{z^2 - 4t}}{2t}.$$

Now, we need to choose a branch of $\sqrt{z^2-4t}$ for $z \in \mathbb{C}^+$. First, we can write $z^2 - 4t = \big(z - 2\sqrt{t}\big).\big(z + 2\sqrt{t}\big)$, and define each of $\sqrt{z-2\sqrt{t}}$ and $\sqrt{z+2\sqrt{t}}$ on $\mathbb{C}^+$. For $z \in \mathbb{C}^+$, let θ_1 be the angle between the x-axis and the line joining z to $2\sqrt{t}$; while θ_2 be the angle between the x-axis and the line joining z to $-2\sqrt{t}$, (see an illustration in Figure A.1). Then, the corresponding polar representations are $(z-2\sqrt{t}) = |z-2\sqrt{t}|.e^{i\theta_1}$ and $(z+2\sqrt{t}) = |z+2\sqrt{t}|.e^{i\theta_2}$, thus, we define $\sqrt{z^2-4t}$ to be $|z^2-4t|^{1/2}.e^{i(\theta_1+\theta_2)/2}$.

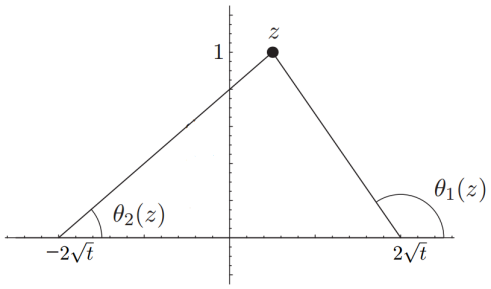


Figure A.1: Choose θ_1, the argument of $z - 2\sqrt{t}$, such that $0 \leq \theta_1 < \pi$. Similarly, choose θ_2, the argument of $z + 2\sqrt{t}$, such that $0 \leq \theta_2 < \pi$.

After obtaining a branch, we also have to choose the sign in front of the square root. By Lemma 2.2.3, we require that $\lim_{|z|\to+\infty,\ z\in D_\alpha} zG_{\sigma_t}(z) = 1$, and equivalently to

$$\lim_{v\to+\infty,\ v>0} ivG_{\sigma_t}(iv) = 1. \tag{A.2}$$

In fact, for $v > 0$, $\sqrt{(iv)^2 - 4t} = i\sqrt{v^2+4t}$, then we have

$$\begin{aligned}
\lim_{v\to+\infty}(iv)\frac{iv - \sqrt{(iv)^2-4t}}{2t} = \lim_{v\to+\infty}(iv)\frac{iv - i\sqrt{v^2+4t}}{2t} &= \lim_{v\to+\infty}\frac{v.\big(\sqrt{v^2+4t} - v\big)}{2t} \\
&= \lim_{v\to+\infty}\frac{v.4t}{2t\left(\sqrt{v^2+4t}+v\right)} \\
&= \lim_{v\to+\infty}\frac{4t}{2t\left(\sqrt{1+\frac{4t}{v^2}}+1\right)} = 1,
\end{aligned}$$

and

$$\lim_{v\to+\infty}(iv)\frac{iv+\sqrt{(iv)^2-4t}}{2t}=\lim_{v\to+\infty}(iv)\frac{iv+i\sqrt{v^2+4t}}{2t}=\lim_{v\to+\infty}\frac{-v.\big(\sqrt{v^2+4t}+v\big)}{2t}=-\infty.$$

Eventually, due to (A.2) , we determine the Cauchy transform of the semicircular distribution σ_t to be:

$$G_{\sigma_t}(z)=\frac{z-\sqrt{z^2-4t}}{2t},\quad z\in\mathbb{C}^+ \text{ such that } |z|>2\sqrt{t},$$

and note that this equality can be extended to the whole domain of analyticity of G_{σ_t}, i.e. for all $z\in\mathbb{C}^+$.

A.1.3 Stieltjes inversion formula

In this section, we give a proof of the Stieltjes inversion formula.

Theorem. *Suppose μ is a probability measure on $\mathbb{R}$ and G_μ is its Cauchy transform on $\mathbb{C}\backslash\mathbb{R}$. Then, for $a<b$, we have the Stieltjes inversion formula*

$$-\frac{2}{\pi}\lim_{v\to0^+}\int_a^b Im\,\Big(G_\mu(u+iv)\Big)du=\mu\Big((a,b)\Big)+\frac{1}{2}\mu\Big(\{a,b\}\Big).$$

Proof. First, one has

$$\text{Im}\,\Big(G_\mu(u+iv)\Big)=\int_{\mathbb{R}}\text{Im}\left(\frac{1}{(u-x)+i.v}\right)d\mu(x)=-\int_{\mathbb{R}}\frac{v}{(u-x)^2+v^2}d\mu(x).$$

Thus, taking the integral, we have

$$\begin{aligned}\int_a^b \text{Im}\,\Big(G_\mu(u+iv)\Big)du&=\int_{\mathbb{R}}\int_a^b\frac{-v}{(u-x)^2+v^2}du\,d\mu(x)\\&=-\int_{\mathbb{R}}\int_{\frac{a-x}{v}}^{\frac{b-x}{v}}\frac{1}{\widetilde{u}^2+1}d\widetilde{u}\,d\mu(x)\\&=-\int_{\mathbb{R}}\left[\tan^{-1}\left(\frac{b-x}{v}\right)-\tan^{-1}\left(\frac{a-x}{v}\right)\right]d\mu(x)\end{aligned}$$

the last equality obtained by changing variable as $\widetilde{u}=\dfrac{u-x}{v}$.

Now, let $f\,(v,x)=\tan^{-1}\left(\dfrac{b-x}{v}\right)-\tan^{-1}\left(\dfrac{a-x}{v}\right)$, and we can observe that

- If $x<a$ or $x>b$:

$$\lim_{v\to0^+}\left[\tan^{-1}\left(\frac{b-x}{v}\right)-\tan^{-1}\left(\frac{a-x}{v}\right)\right]=\frac{\pi}{2}-\frac{\pi}{2}=0;$$

- If $x=a$, then

$$\lim_{v\to0^+}\left[\tan^{-1}\left(\frac{b-x}{v}\right)-\tan^{-1}\left(\frac{a-x}{v}\right)\right]=\lim_{v\to0^+}\left[\tan^{-1}\left(\frac{b-a}{v}\right)-0\right]=\frac{\pi}{2};$$

- Similarly, if $x=b$

$$\lim_{v\to0^+}\left[\tan^{-1}\left(\frac{b-x}{v}\right)-\tan^{-1}\left(\frac{a-x}{v}\right)\right]=\lim_{v\to0^+}\left[0-\tan^{-1}\left(\frac{a-b}{v}\right)\right]=\frac{\pi}{2};$$

- Eventually, if $a < x < b$

$$\lim_{v\to 0^+}\left[\tan^{-1}\left(\frac{b-x}{v}\right)-\tan^{-1}\left(\frac{a-x}{v}\right)\right]=\frac{\pi}{2}+\frac{\pi}{2}=\pi.$$

So, if we consider a function

$$f(x)=\frac{\pi}{2}.\mathbb{I}_{\{a,b\}}(x)+\pi.\mathbb{I}_{(a,b)}(x),$$

where the indicator function $\mathbb{I}_A(x)=1$ if $x\in A$, and otherwise $\mathbb{I}_A(x)=0$ if $x\notin A$. Thus, we can see that $\lim_{v\to 0^+} f(v,x)=f(x)$. Moreover, for all $v>0$ and for all x, we have $|f(v,x)|\leq\pi$. Consequently, by Lebesgue's dominated convergence theorem

$$\begin{aligned}\lim_{v\to 0^+}\int_a^b \operatorname{Im}\Big(G_\mu\left(u+iv\right)\Big)du &= -\lim_{v\to 0^+}\int_{\mathbb{R}} f\left(v,x\right)d\mu(x)\\ &= -\int_{\mathbb{R}} f\left(x\right)d\mu(x)\\ &= -\frac{\pi}{2}\mu\big(\,\{a,b\}\,\big)-\pi\mu\big(\,(a,b)\,\big)\end{aligned}$$

This implies that

$$-\frac{2}{\pi}\lim_{v\to 0^+}\int_a^b \operatorname{Im}\Big(G_\mu(u+iv)\Big)du=\mu\big(\,\{a,b\}\,\big)+2\mu\big(\,(a,b)\,\big)=\mu\big(\,(a,b)\,\big)+\mu\big(\,[a,b]\,\big),$$

so one gets the Stieltjes inversion formula.

□

In particular, we can recover μ from G_μ by using the *Stieltjes inversion formula* as follows:

$$d\mu(x)=-\frac{1}{\pi}\lim_{\varepsilon\to 0^+}\operatorname{Im}\left(G_\mu\left(x+i\varepsilon\right)\right).$$

A.1.4 Implicit functions Theorem

Let E,F,G be three Banach spaces, an open set U of $E\times F$, and a function $f:U\longrightarrow G$; f is a function of two variables $f(x,y)$, where $x\in E$, $y\in F$, the couple (x,y) remains in U.
Give $(a,b)\in U$, and suppose

$$f(a,b)=0.$$

We study the solutions (x,y) of equation

$$f(x,y)=0,$$

in a neighborhood of (a,b). Before doing that, we need the following hypothesis:

Assumption A.1.2. *The partial derivative $f'_y(a,b)$ is an isomorphism of F onto G.*

Now, under the previous assumption, we can state the theorem

Theorem A.1.3 (Implicit Functions Theorem)**.** *There exists in $E\times F$ an open neighborhood V of (a,b), contained in U, there also exists in E an open neighborhood W of a, and there exists a $\mathcal{C}^1$ function*

$$g:W\longrightarrow F$$

that has a following property: the relation

$$(x,y)\in V,\quad \textit{and}\quad f(x,y)=0$$

is equivalent to the relation

$$x\in W,\quad \textit{and } y=g(x).$$

For more details, readers can see for instance in the book "Calcul Différentiel" of Henri Cartan.

A.2 Stochastic calculus and Topological preliminaries

In proof of Proposition 2.3.5, there are several helpful results to rely on, as we mention in the following. We refer to AGZ[1, Appendix B, C and H] for more details on these results.

Theorem A.2.1 (**Itô**(1944), **Kunita-Watanabe** (1967)). *Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^2$ and let $X = \{X_t, \mathcal{F}_t; 0 \leq t < \infty\}$ be a continuous semi-martingale with decomposition*

$$X_t = X_0 + M_t + A_t$$

where M is a local martingale and A the difference of continuous, adapted, non-decreasing processes. Then, almost surely,

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dM_s + \int_0^t f'(X_s)dA_s + \frac{1}{2}\int_0^t f''(X_s)d\langle M\rangle_s, \quad 0 \leq t < \infty.$$

Lemma A.2.2 (See [1], Corollary H.13, page 462). *Let $\{X_t, \mathcal{F}_t, t \geq 0\}$ be an adapted process with values in $\mathbb{R}^d$ such that*

$$\int_0^T \|X_t\|^2 \, dt := \int_0^T \sum_{i=1}^d (X_t^i)^2 dt$$

is uniformly bounded by A_T. Let $\{W_t, \mathcal{F}_t, t \geq 0\}$ be a d-dimensional Brownian motion. Then for any $L > 0$,

$$\mathbb{P}\left(\sup_{0\leq t\leq T}\left|\int_0^t X_s.dW_s\right| \geq L\right) \leq 2e^{-\frac{L^2}{2A_T}}.$$

A.2.1 Some Topological preliminaries

- A set is *pre-compact* (or *relatively compact*) if its closure is compact. A topological space is *locally compact* if every point possesses a neighborhood that is compact.
- Let $(\mathcal{X}, d)$ be a metric space, then a subset $A \subset \mathcal{X}$ is *sequentially compact*, if every sequence contained in A has a subsequence converging to a point in $\mathcal{X}$.
- Any set $A \subset \mathcal{X}$ is *dense* in $\mathcal{X}$ if its closure is $\mathcal{X}$. A topological space $\mathcal{X}$ is *separable* if it contains a countable subset which is dense in $\mathcal{X}$.

Lemma A.2.3. *A subset of a metric space is compact if and only if it is closed and sequentially compact.*

Definition A.2.4. A topological space (E, τ) is called a *Polish* space, if it is separable and if there exists a complete metric that induces the topology τ.

Any set B in a metric space $\mathcal{X}$ is *totally bounded*, if for every $\delta > 0$, it is possible to cover B by a finite number of ball of radius δ centered in B. Moreover, a totally bounded subset of a complete metric space is pre-compact.

A probability measure μ on the metric space Σ is *tight*, if for each $\eta > 0$, there exists a compact set $K_\eta \subset \Sigma$ such that $\mu\big(K_\eta^c\big) < \eta$.

Theorem A.2.5 (Prohorov)**.** *Let Σ be Polish, and let $\Gamma \subset \mathcal{M}_1(\Sigma)$. Then $\overline{\Gamma}$ is compact if and only if Γ is tight.*

We denote $B(\Sigma)$ as the set of bounded measurable functions on Σ. Then, a collection of functions $\mathcal{G} \subset B(\Sigma)$ is called *convergence determining* for $\mathcal{M}_1(\Sigma)$, if

$$\lim_{n\to\infty} \int_\Sigma g d\mu_n = \int_\Sigma g d\mu, \quad \forall g \in \mathcal{G} \quad \Rightarrow \quad \mu_n \overset{n\to+\infty}{\longrightarrow} \mu. \tag{A.3}$$

Lemma A.2.6. *(i) Suppose E is separable metric space. Then, the space of uniformly continuous functions with bounded support is convergence determining for $\mathcal{M}_1(E)$.*

(ii) Suppose E is separable, locally compact metric space. Then, the space of continuous functions with compact support is convergence determining for $\mathcal{M}_1(E)$.

Bibliography

Part I - Statistical deconvolution of free Fokker-Planck equation at fixed time

[1] G.W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[2] O. Arizmendi, P. Tarrago, and C. Vargas. Subordination methods for free deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.*, 56(4):2565–2594, 2020.

[3] Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, volume 37 of *Series in Statistics*. Springer, 2 edition, 2006.

[4] S. Belinschi and H. Bercovici. A new approach to subordination results in free probability. *Journal d'Analyse Mathematique*, 101:357–365, 2007.

[5] H. Bercovici and D. Voiculescu. Free convolution of measures with unbounded support. *Indiana University Mathematics Journal*, 42:733–773, 1993.

[6] J. Bertoin, C. Giraud, and Y. Isozaki. Statistics of a flux in Burgers turbulence with one-sided Brownian initial data. *Comm. Math. Phys.*, 224(2):551–564, 2001.

[7] P. Biane. On the free convolution with a semi-circular distribution. *Indiana University Mathematics Journal*, 46(3):705–718, 1997.

[8] P. Biane. Processes with free increments. *Math. Z.*, 227(1):143–174, 1998.

[9] P. Biane and R. Speicher. Free diffusions, free entropy and free Fisher information. *Ann. Inst. H. Poincaré Probab. Statist.*, 37(5):581–606, 2001.

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

[11] J. Bun, J.-P. Bouchaud, and M. Potters. Cleaning large correlation matrices: tools from random matrix theory. *Phys. Rep.*, 166(1):1–26, 2017.

[12] J.M. Burgers. *The Nonlinear Diffusion Equation*. Springer, 1974.

[13] N. Burq and N. Tzvetkov. Random data cauchy theory for supercritical wave equations i: Local theory. *Inventiones Mathematicae*, 173:449–475, 2008.

[14] N. Burq and N. Tzvetkov. Random data cauchy theory for supercritical wave equations ii: A global result. *Inventiones Mathematicae*, 173:477–496, 2008.

[15] C. Butucea Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics*, Vol. 32, No. 2:181–192, 2004.

[16] C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.*, 52(1):111–128, 2007.

[17] C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.*, 52(2):336–349, 2007.

[18] J.A. Carrillo, R.J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[19] R.J. Carroll and P. Hall. Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83(404):1184–1186, 1988.

[20] E.A. Cator. Deconvolution with arbitrarily smooth kernels. *Statist. Probab. Lett.*, 54(2):205–214, 2001.

[21] E. Cépa and D. Lépingle. Diffusing particles with electrostatic repulsion. *Probability Theory and Related Fields*, 107:429–449, 1997.

[22] A.J. Chorin, O.H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Natl. Acad. Sci. USA*, 97(7):2968–2973, 2000.

[23] F. Comte and V. Genon-Catalot. Penalized projection estimator for volatility density. *Scand. J. Statist.*, 33(4):875–895, 2006.

[24] F. Comte, Y. Rozenholc, and M.-L. Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, 34(3):431–452, 2006b.

[25] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(2):569–609, 2013.

[26] P. Constantin and J. Wu. Statistical solutions of the Navier-Stokes equations on the phase space of vorticity and the inviscid limit. *Journal of Mathematical Physics*, 38(6):3031–3045, 06 1997.

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

[28] A. Delaigle and I. Gijbels. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.*, 56(1):19–47, 2004.

[29] L. Devroye. Consistent deconvolution in density estimation. *Canad. J. Statist.*, 17(2):235–239, 1989.

[30] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272, 1991.

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

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

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

[34] B. Groux. *Grandes déviations de matrices aléatoires et Équation de Fokker-Planck libre.* PhD Thesis, Université Paris-Saclay, 2016.

[35] C.H. Hesse. Data-driven deconvolution. *J. Nonparametr. Statist.*, 10(4):343–373, 1999.

[36] J.-Y. Koo. Logspline deconvolution in Besov space. *Scand. J. Statist.*, 26(1):73–86, 1999.

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

[38] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal*, 88(2):365–411, 2004.

[39] M.C. Liu and R.L. Taylor. A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.*, 17(4):427–438, 1989.

[40] M.Maïda, T.D.Nguyen, T.M.Pham Ngoc, V. Rivoirard and V.C.Tran. Statistical deconvolution of the free Fokker-Planck equation at fixed time. Accepted in *Bernoulli*, 2021.

[41] E. Masry. Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory*, 37(4):1105–1115, 1991.

[42] H. Masseen. Addition of freely independent random variables. *J. Funct. Anal.*, 106(2):409–438, 1992.

[43] A. Meister. *Deconvolution Problems in Nonparametric Statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 2009.

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

[45] J. A. Mingo, and R. Speicher. *Free Probability and Random Matrices*, Schwinger-Verlag, 2017.

[46] A.S. Monin, A.M. Yaglom, and J.L. Lumley. *Statistical Fluid Mechanics*, Vol.2, MIT Press, 1975.

[47] M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053, 1999.

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

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

[50] W. Rudin. *Real and Complex Analysis.* 3rd Ed., McGraw-Hill, Inc., USA, 1978.

[51] J. Song, J. Yao, and W. Yuan. High-dimensional limits of eigenvalue distributions for general Wishart process. *Ann. Appl. Probab.*, 30(4):1642–1668, 2020.

[52] J. Song, J. Yao, and W. Yuan. Recent advances on eigenvalues of matrix-valued stochastic processes. *J. Multivariate Anal.*, 2021.
https://doi.org/10.1016/j.jmva.2021.104847

[53] L. Stefanski and R.J. Carroll. Deconvoluting kernel density estimators. *Statistics*, 21(2):169–184, 1990.

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

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

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

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

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

[59] A. B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer series in Statistics. Springer, New York, 2009.

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

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

[62] M.J. Vishik and A.V. Fursikov. *Mathematical Problems of Statistical Hydromechanics.* Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht, 1988.

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

[64] D. Voiculescu. The analogues of entropy and Fisher's information measure in free probability theory, i. *Commun. Math. Phys.*, 155:71–92, 1993.

[65] C.-H. Zhang. Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, 18(2):806–831, 1990.

Part II - Adaptive warped kernel estimation for nonparametric regression with circular responses

[66] N. Bochkina and T. Sapatinas. Minimax rates of convergence and optimality of Bayes factor wavelet regression estimators under pointwise risks. *Statistica Sinica*, 19(4):1389–1406, 2009.

[67] G. Chagny. Penalization versus Goldenshluger-Lepski strategies in warped bases regression. *ESAIM Probab. Stat.*, 17:328–358, 2013.

[68] G. Chagny. Adaptive warped kernel estimators. *Scandinavian Journal of Statistics*, 42(2):336–360, 2015.

[69] G. Chagny, T. Laloë and R. Servien. Multivariate adaptive warped kernel estimation. *Electron. J. Statist.*, 13(1):1759–1789, 2019.

[70] N.I. Fisher and A.J. Lee. Regression models for angular responses. *Biometrics* **48**, 665–677, 1992.

[71] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality *Ann. Statist* **39** 1608–1632, 2011.

[72] A.L. Gould. A regression technique for angular variates. *Biometrics* **25**, 683–700, 1969.

[73] S.R. Jammalamadaka and A. SenGupta . *Topics in Circular Statistics*. World Scientific, Singapore, 2001.

[74] R.A. Johnson and T.E. Wehlry. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association* **73**, 602–606, 1978.

[75] G. Kerkyacharian and D. Picard. Regression in random design and warped wavelets. *Bernoulli* 10(6):1053–1105, 2004.

[76] M.Kohler, A.Krzyzak, and H.Walk. Optimal global rates of convergence for nonparametric regression with unbounded data. *J. Statist. Plann. Inference* 139(4):1286–1296, 2009.

[77] A. Lee. Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics* 2:477–486, 2010.

[78] C. Ley and T. Verdebout . *Modern Directional Statistics*. (1st ed.). Chapman and Hall/CRC, 2017.

[79] K.V. Mardia and P.E. Jupp. *Directional Statistics*. New York, NY: John Wiley, 2000.

[80] M. Di Marzio, A. Panzera and C.C. Taylor. Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40:238–255, 2013.

[81] M. Di Marzio, A. Panzera and C.C. Taylor. Nonparametric Regression for Spherical Data. *Journal of the American Statistical Association*, 109:748–763, 2014.

[82] A. Meilán-Vila, M. Francisco-Fernández, R.M. Crujeiras and A. Panzera. Nonparametric multiple regression estimation for circular response. *TEST*, (2020).

[83] A. Pewsey and E. García-Portugués. Recent advances in directional statistics. *TEST* **30**, 1–58, (2021).

[84] T.M. Pham Ngoc. Regression in random design and Bayesian warped wavelets estimators. *Electron. J. Stat* 3:1084–1112, 2009.

[85] T.M. Pham Ngoc. Adaptive optimal kernel density estimation for directional data. *J. Multivariate Anal.* 173:248–267, 2019.

[86] B. Presnell, S.P. Morrison and R.C. Littel. Projected multivariate linear models for directional data. *Journal of the American Statistical Association* **93**, 1068–1077, 1998.

[87] W.Stute. Asymptotic normality of nearest neighbor regression function estimates. *Ann. Statist.*, 12(3):917–926, 1984.

[88] S.-S.Yang. Linear combination of concomitants of order statistics with application to testing and estimation. *Annals of the Institute of Statistical Mathematics*, 33(1):463–470, 1981.

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre: Déconvolution libre non paramétrique et estimation pour la régression circulaire.

Mots clés: Estimation non-paramétrique, équation de Fokker-Planck, Déconvolution libre, régression circulaire.

Résumé: Cette thèse contient deux parties indépendantes. Dans la première partie, nous nous intéressons à la reconstruction de la condition initiale d'une équation aux dérivées partielles (EDP) non-linéaire, à savoir l'équation de Fokker-Planck, à partir de l'observation d'un mouvement brownien de Dyson à un instant $t > 0$. La solution de l'équation de Fokker-Planck peut être écrite comme la convolution libre de la condition initiale et de la distribution semi-circulaire. Nous proposons un estimateur non paramétrique de la condition initiale obtenue en effectuant la déconvolution libre via la méthode des fonctions de subordination. Cet estimateur statistique est original car il implique la résolution d'une équation à point fixe, et une déconvolution classique par une distribution de Cauchy. Ceci est dû au fait que, en probabilité libre, l'analogue de la transformée de Fourier est la R-transformée, liée à la transformée de Cauchy. Dans la littérature passée, le focus a été mis sur l'estimation des conditions initiales des EDP linéaires telles que l'équation de la chaleur, mais à notre connaissance, c'est la première fois que le problème est examiné pour une EDP non linéaire. La convergence de l'estimateur est prouvée et le "mean integrated squared error" est calculé, fournissant des vitesses de convergence similaires à celles connues pour les méthodes de déconvolution non paramétriques. Enfin, une étude de simulation illustre les bonnes performances de notre estimateur. Dans la deuxième partie de cette thèse, l'estimation non paramétrique pour la régression avec variables réponses circulaires est étudiée. Nous proposons un nouvel estimateur à noyau "warped" pour la fonction de régression en un point donné. La fenêtre du noyau est sélectionnée par la méthode de Goldenshluger-Lepski. En considérant le risque au carré ponctuel, nous obtenons une inégalité de type oracle et des vitesses de convergence sur la classe Hölder.

Title: Nonparametric free deconvolution and circular regression estimation.

Keywords: Nonparametric estimation, Fokker-Planck equation, Free convolution, circular regression.

Abstract: This thesis contains two independent parts. In the first part, 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 solution of the Fokker-Planck equation can be written as the free convolution of the initial condition and the semi-circular distribution. We propose a nonparametric estimator for the initial condition obtained by performing the free deconvolution via the subordination functions method. This statistical estimator is original as it involves the resolution of a fixed point equation, and a classical deconvolution by a Cauchy distribution. This is due to the fact that, in free probability, the analogue of the Fourier transform is the R-transform, related to the Cauchy transform. In past literature, there has been a focus on the estimation of the initial conditions of linear PDEs such as the heat equation, but to the best of our knowledge, this is the first time that the problem is tackled for a non-linear PDE. The convergence of the estimator is proved and the mean integrated squared 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.
In the second part of this thesis, the nonparametric estimation for regression with circular responses is studied. We propose a new warped kernel estimator for the regression function at a given point. The bandwidth of kernel is selected by the method of Goldenshluger-Lepski. Under the consideration of pointwise squared risk, we obtain an oracle-type inequality and rates of convergence over Hölder class.