LONG TIME ASYMPTOTICS OF SOME WEAKLY INTERACTING PARTICLE SYSTEMS AND HIGHER ORDER ASYMPTOTICS OF GENERALIZED FIDUCIAL DISTRIBUTION

Abhishek Pal Majumder

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2015

Approved by:

Amarjit S. Budhiraja

Jan H. Hannig

Shankar Bhamidi

Vladas Pipiras

Chuanshu Ji

© 2015
Abhishek Pal Majumder
ALL RIGHTS RESERVED

ABSTRACT

ABHISHEK PAL MAJUMDER: LONG TIME ASYMPTOTICS OF SOME WEAKLY INTERACTING PARTICLE SYSTEMS AND HIGHER ORDER ASYMPTOTICS OF GENERALIZED FIDUCIAL DISTRIBUTION
(Under the direction of Amarjit S. Budhiraja and Jan H. Hannig)

In probability and statistics limit theorems are some of the fundamental tools that rigorously justify a proposed approximation procedure. However, typically such results fail to explain how good is the approximation. In order to answer such a question in a precise quantitative way one needs to develop the notion of convergence rates in terms of either higher order asymptotics or non-asymptotic bounds. In this dissertation, two different problems are studied with a focus on quantitative convergence rates.

In first part, we consider a weakly interacting particle system in discrete time, approximating a nonlinear dynamical system. We deduce a uniform in time concentration bound for the Wasserstein-1 distance of the empirical measure of the particles and the law of the corresponding deterministic nonlinear Markov process that is obtained through taking the particle limit. Many authors have looked at similar formulations but under a restrictive compactness assumption on the particle domain. Here we work in a setting where particles take values in a non-compact domain and study several time asymptotics and large particle limit properties of the system. We establish uniform in time propagation of chaos along with a rate of convergence and also uniform in time concentration estimates. We also study another discrete time system that models active chemotaxis of particles which move preferentially

towards higher chemical concentration and themselves release chemicals into the medium dynamically modify the chemical field. Long time behavior of this system is studied.

Second part of the dissertation is focused on higher order asymptotics of Generalized Fiducial inference. It is a relevant inferential procedure in standard parametric inference where no prior information of unknown parameter is available in practice. Traditionally in Bayesian paradigm, people propose posterior distribution based on the non-informative priors but imposition of any prior measure on parameter space is contrary to the "no-information" belief (according to Fisher's philosophy). Generalized Fiducial inference is one such remedy in this context where the proposed distribution on the parameter space is only based on the data generating equation. In this part of dissertation we established a higher order expansion of the asymptotic coverage of one-sided Fiducial quantile. We also studied further and found out the space of desired transformations in certain examples, under which the transformed data generating equation yields first order matching Fiducial distribution.

Baba, Ma,

Chordadu

and all my teachers.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my two advisors: Amarjit Budhiraja and Jan Hannig for their encouragements, guidances and inspirations throughout my graduate life. Apart from being great teachers and providing invaluable inputs to my research, they also helped me grow as a person, enabling me to find a career path for myself. I learned a lot about academic life as much as I did about my impediments. I will remain forever grateful to Amarjit for having patience in me in this process and teaching me the value of discipline, hard work. I am personally grateful to Jan since he introduced me to the topic of Generalized Fiducial Inference and without his constant support and invaluable insights this work would not have been possible. Their teachings will always stay with me for the rest of my career. Furthermore, I take this opportunity to express my gratitude to both of them for financially supporting me from their NSF grants.

I am extremely thankful to Shankar for teaching us measure theory in the very first semester with so much care and later for the course on concentration bounds which helps me enormously to develop a broader intuition for solving any problem through the concepts of probability theory. I am grateful to him for numerous discussions on various other research topics of Probability. I will not forget his countless treats specially the one while Prof. B V Rao was visiting here. I would also like to thank Professor Vladas Pipiras and Chuansu Ji for agreeing to be in my dissertation committee in spite of their extremely busy schedules and for numerous other things that I had learned from them. I thank all my teachers at UNC, Chapel Hill especially Michael Kosorok, Edward Carlstein; for their invaluable advices, teachings which were extremely useful in my later research and for being generous to me with their encouragements. Besides that, I want to thank all other professors and staff in this small wonderful Statistics department, whose doors were always open for me. Many thanks to Alison Kieber for becoming my savior for answering all my "important" questions.

It is imperative that I thank all my fellow graduate students and postdocs who were significant part of my life in last few years. I should specially mention Subhamay, Xuan, Ruyou, Suman and Louis for all interesting conversations. I am thankful to all my batch-mates especially James, Jenny, Michael, Minghui for sharing their valuable thoughts in my research and presentations. Special thanks to all my officemates Chris, Washington, John, Hwo-young for all my good times and enriching my graduate life experiences.

As they say, "*There is nothing on this earth more to be prized than true friendship.*" I feel really blessed to have so many good friends. Special thanks to my roommates Sayan Dasgupta and Ritwik Chaudhuri for always being with me through the highs and lows. Without Ritwik's help I may not end up here. I learned so many things from Sayan that this place will fall short to list them. In one sentence my life would not be easier without him. I should mention Pourab, Sid da, Anji, Pdr, Sujatro, Tommy, Poulami, Suprateek da, Suman, Swarnava, Sayantan for filling my life with your cheerful presence for last five years. Special thanks to Rinku di and Samarpan da for your warmths and making all of us your extended family. We are so lucky to have you in Chapel Hill. I am also thankful to Anirban Basak, Jyotishka da, Shalini di for all great times that we spent together. I will miss all of you in my next venture.

I am grateful to all my teachers at Indian Statistical Institute, Kolkata (especially Prof. S C Bagchi, Alok Goswami, Gopal K Basak, Saurabh Ghosh, Mahuya Dutta and all others) for teaching me statistics and probability as well as building the inherent mathematical foundation. I should specially thank Prof. Debapriya Sengupta for introducing the problem binary matrix simulation constrained by row and column margins which I think changed my notion of research. I would also thank Professor Debapriya Sengupta, Sumitra Purkayastha, Sourabh Bhattacharya, Indranil Mukhopadhyay, Arijit Chakrabarty for their supports during my struggles in JRF days. I would have never reached this far without your helps.

I could not close this section before I thank my family for its unconditional love, support and give special thanks to my first teacher in all of my mathematical training, my father Paritosh Pal Majumder.

TABLE OF CONTENTS

LIST OF TABLES

Chapter1
INTRODUCTION

The dissertation consists of two parts. First part is concerned with some problems in weakly interacting particle systems while second part of the dissertation is focused on higher order asymptotic properties of Generalized Fiducial Inference.

In recent years there has been a significant research activity aimed towards understanding the dynamics of the collective behavior of a group of interacting similarly behaving agents/ particles. Motivation for such problems comes from various examples of self-organizing systems in biological, physical and social sciences. These include, problems of opinion dynamics [46], chemotaxis [15], self organized networks [56], large communication systems [49], swarm robotics [69], etc. For additional references we refer the reader to [70] and references therein. One of the basic challenges is to understand how a large group of autonomous agents with decentralized local interactions gives rise to a coherent behavior.

A commonly studied mathematical formulation of such a system of particles is given in terms of a Markov process whose transition probabilities depend on the empirical process associated with the particle states. Starting from the work of Sznitman [74] there has been an extensive body of work that studies law of large number behavior (Propagation of Chaos), central limit theory (normal fluctuations from the mean) and large deviation principles for such Markovian systems. All of these results concern the behavior of the system over a finite time horizon. In many applications the time asymptotic behavior of the system is of central concern. For example, stability of a communication system, steady state aggregation and self organization in biological and chemical systems, long term consensus formation mechanisms in opinion dynamics modeling, particle based approximation methods for invariant measures, all rely on a careful analysis of the time asymptotic behavior of such systems. Understanding such long time properties is the focus of first part of this dissertation. Specifically the goal of this dissertation is to study several time asymptotic problems for certain families of weakly interacting Markov processes. These problems include, uniform in time law of large number results and uniform in time polynomial and exponential concentration bounds. We consider here a mean field type interaction among the particles. General features of these particle systems include exchangeability under suitable assumptions on initial exchangeability and noise models. Thus the empirical measure of the particle states is a natural summary statistic of the system. There are two types of asymptotics considered here: particle limit ($N \to \infty$) and time limit ($n \to \infty$). Starting point of this work will be to identify the nonlinear dynamical system that is obtained as the formal particle limit ($N \to \infty$) of the occupation measure at every fixed time point. This nonlinear dynamical system can be viewed as the Kolmogorov forward equation for a nonlinear Markov process. We give conditions on the model parameters that ensure stability of the nonlinear system. Next uniform in time convergence of the particle system to the nonlinear system is established along with the the existence of a unique fixed point under some natural integrability assumptions.

One of the main objectives of this dissertation is to quantify the convergence rates in terms of concentration bounds. To see the basic question of interest let L_N be the empirical distribution of N iid samples $\{Y_i\}_{\{i=1,\dots,N\}}$ from distribution μ on a Polish space S. We know from the well known theorem of Varadarajan (1958) [77] that L_N converges to μ weakly, almost surely. An equivalent statement is $\beta(L_N, \mu) \to 0$ almost surely where $\beta(\cdot, \cdot)$ is the bounded Lipschitz metric (metrizing the weak topology) on $\mathcal{P}(S)$ (the space of all probability measures on S). It is natural to seek a non asymptotic bound for $P(\beta(L_N, \mu) > \varepsilon)$ in order to see how fast $\beta(L_N, \mu)$ converges to 0.

One can consider other metrics on the space of probability measures, one such choice being the Wasserstein metric. For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d) := \{\mu : \int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty\}$ the Wasserstein distance metric of order $p \geq 1$ is defined

$$\mathcal{W}_p(\mu, \nu) := \Big[\inf_{X,Y} E|X - Y|^p \Big]^{\frac{1}{p}}$$

where the infimum is taken over all $\mathbb{R}^d$ valued random variables X, Y defined on a common probability space where the marginals of X, Y are respectively μ and ν. In our work we will be interested in giving uniform in time concentration estimates for $\sup_{n\geq 0} P\big[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon\big]$ where μ_n is the law of the nonlinear Markov process obtained as the particle limit of $\mu_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_n^{i,N}}$ which is the occupation measure of the states of N particles $\{X_n^{i,N}\}_{i=1,\dots,N}$ at time instant n.

Finding sharp concentration bounds for $P(\mathcal{W}_p(L_N, \mu) > \varepsilon)$ is a classical problem that is central in probability, statistics, combinatorics and informatics with a number of applications: (see Bolley-Guillin-Villani [14], Boissard [11]). Below we will review some basic results and techniques that have been used to obtain concentration bounds for $\mathcal{W}_P(L_N, \mu)$ in the classical iid setting. One of the key result is due to Boissard which is given in Chapter 2 as Theorem 2.4.1.

Suppose $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, the relative entropy of ν with respect to μ is defined as

$$R(\nu\|\mu) = \begin{cases} \int_{\mathbb{R}^d} \left(\log \frac{d\nu}{d\mu}\right) d\nu & \text{when } \nu \ll \mu, \\ \\ \infty & \text{otherwise.} \end{cases}$$

One major ingredient in proving Boissard's result is the Transportation inequality (in short **TI**).

Definition 1.0.1. *Let $\alpha : [0, \infty) \to \mathbb{R}$ be a convex, increasing left-continuous function such that $\alpha(0) = 0$. We say $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies a $\alpha(\mathcal{T})$ transportation inequality if for all $\tilde{\nu} \in \mathcal{P}(\mathbb{R}^d)$*

$$\alpha(\mathcal{W}_1(\mu, \tilde{\nu})) \leq R(\tilde{\nu}\|\mu). \tag{1.0.1}$$

Applications of transportation inequality include exponential concentration estimate of $\mu\{d(x, A) > \varepsilon\}$ (where $d(x, A)$ is the distance from x to set A) Marton [65],[66]. A seminal contribution of Djellout, Guilin, Wu [29] was to prove the equivalence of the transportation inequalities with the finiteness of exponential moments of various orders. Similar results were also derived in [13] which we use in Chapter 2. In the following section we give a sketch of the proof of concentration bound for $P(\mathcal{W}_1(L_N, \mu) > \varepsilon)$ as done in [11] to see how

TI is used.

1.1 Sketch of the bound for $P(\mathcal{W}_1(L_N, \mu) > \varepsilon)$:

We begin by illustrating a few steps and combine all of them in Step 4.

1. Following proposition can be proved by using the representation formula of $R(\nu\|\mu)$ along with Kantorovich-Rubenstein duality of Wasserstein-1 metric[48] or alternatively using large deviation techniques [47].

   Proposition 1.1.1. *If μ is a measure that satisfies $\alpha(\mathcal{T})$ inequality for some convex $\alpha(\cdot)$ then for any $f \in$ Lip$_1(\mathbb{R}^d)$ (the space of all Lipschitz functions with Lipschitz norm bounded by 1), one has for all $\lambda > 0$*

   $$\int e^{\lambda(f - \int f d\mu)} d\mu \leq e^{\alpha^{\odot}(\lambda)}$$

   where $\alpha^{\odot}(\lambda) := \sup_{x>0}\{\lambda x - \alpha(x)\}$ (the monotone conjugate of $\alpha(\cdot)$).

   Using this proposition:

   $$\begin{aligned} \mu[f - \int f d\mu > \varepsilon] &\leq e^{-\lambda\varepsilon}.Ee^{\lambda(f - \int f d\mu)} \\ &\leq e^{-(\lambda\varepsilon - \alpha^{\odot}(\lambda))}. \end{aligned} \tag{1.1.1}$$

   Now optimizing the right hand side bound in (1.1.1) with respect to λ gives (since

$\alpha^{\odot\odot}(\cdot) = \alpha(\cdot)$):

$$\mu[f - \int f d\mu > \varepsilon] \quad \leq \quad e^{-\alpha(\varepsilon)}. \tag{1.1.2}$$

2. Next we extend this idea to get a concentration bound for a product measure $\mu^{\otimes n}$ instead of μ. For this we need a tensorization idea which will be illustrated in this step. Denote

$$d^{\oplus N}(\{x_i\}_{i=1,\ldots,N}, \{y_i\}_{i=1,\ldots,N}) := \sum_{i=1}^{N} d(x_i, y_i)$$

where $d(x_i, y_i)$ is the component-wise Euclidean distance in $\mathbb{R}^d$. From Kantorovich-Rubenstein duality principle it can be seen that

$$\mathcal{W}_1(L_N, \mu) = \sup_{f \in \mathrm{Lip}_1(\mathbb{R}^d)} \left[\frac{1}{N} \sum_{i=1}^{N} f(Y_i) - \int f d\mu \right]. \tag{1.1.3}$$

Note that the function $(x_1, \ldots, x_N) \longrightarrow \mathcal{W}_1(\frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \mu)$ is a Lipschitz function on $(\mathbb{R}^d)^N$ with Lipschitz constant $\frac{1}{N}$ with respect to the distance metric $d^{\oplus N}$.

3. Now we will state the following tensorization lemma which states a transportation inequality of product measures.

Lemma 1. *If $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^d)$ follow **TI** with functions α_1 and α_2 respectively; then the product measure will also follow **TI** on the space of product measure with function $\alpha_1 \square \alpha_2(t) := \inf\{\alpha_1(t_1) + \alpha_2(t_2) : t_1 + t_2 = t\}$. That is*

$$\alpha_1 \square \alpha_2(\mathcal{W}_1^*(\mu_1 \otimes \mu_2, \nu)) \leq R(\nu \mid \mu_1 \otimes \mu_2) \quad \forall \nu = (\nu_1, \nu_2), \nu_i \in \mathcal{P}(\mathbb{R}^d)$$

where $\mathcal{W}_1^*(\mu_1 \otimes \mu_2, \nu)$ is defined with respect to the distance $d^{\oplus 2}$.

In general, for $N \geq 1$, and functions $\alpha_1(\cdot), \ldots, \alpha_N(\cdot)$ in definition 1.0.1, we define

$$\alpha_1 \square \ldots \square \alpha_N(t) = \inf\{\alpha_1(t_1) + \alpha_2(t_2) + \ldots + \alpha_N(t_2) : t_1 + t_2 + \ldots + t_N = t\}.$$

Since $\{Y_i\}_{i=1,\ldots,N}$ are iid from the measure μ, from convexity of $\alpha(\cdot)$ it follows:

$$\alpha^{\square N}(t) := \alpha \square \alpha \ldots \square \alpha = N\alpha(\frac{t}{N}). \tag{1.1.4}$$

4. $\mu^{\otimes N}$ satisfies TI with function $\alpha^{\square N}(.)$ and with the distance metric $d^{\oplus N}$. Now if f is a Lipschitz functional on $(\mathbb{R}^d)^N$, then from (1.1.2) the concentration under product law $\mu^{\otimes N}$ can be written as

$$\mu^{\otimes N}[f - \int f d\mu > \varepsilon] \leq e^{-\alpha^{\square N}(\frac{\varepsilon}{\|f\|_1})}.$$

The following concentration bound for $\mathcal{W}_1(L_N, \mu)$ now follows from the above observation where $\|f\|_1$ is the Lipschitz norm of f

$$P[\mathcal{W}_1(L_N, \mu) - E\mathcal{W}_1(L_N, \mu) > \varepsilon] \;\; \leq \;\; e^{-N\alpha(\varepsilon)}. \tag{1.1.5}$$

Boissard [11] using Orlicz-norm inequality gives a representation of $E\mathcal{W}_1(L_N, \mu)$ from which the inequality in (1.1.5) can be used to yield an exponential bound for $P(\mathcal{W}_1(L_N, \mu) > t)$. This result will be used in our study of weakly interacting particle systems in chapter 2 (See Theorem 2.4.1).

1.2 Connection to Our Problem and Other Questions:

The sketch given in Section 1.1 essentially uses the independence of $\{Y_i\}_{i=1,\ldots,n}$ which fails to hold for the particle system $\{X_n^{i,N} : i = 1, \ldots, N\}_{n\geq 0}$ considered in our work. However, as will be shown in Chapter 2 one can still apply the result of Boissard (i.e. Theorem 2.4.1), using a coupling technique. In fact we show that under suitable condition; one can give uniform in time n exponential and polynomial concentration bounds for $\sup_{n\geq 0} P\big[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon\big]$.

This uniform in time concentration result implies that the $\mathcal{W}_1$ distance between μ_n^N and the unique fixed point (denoted by μ_∞) of the nonlinear system converges to zero as $n \to \infty$ and $N \to \infty$ in any order. This result is key in developing particle based numerical schemes for approximating the fixed point of the evolution equation. Such a result can also be used for performance evaluation. Given a "cost function" $\psi : \mathbb{R}^d \to \mathbb{R}$ with suitable continuity and integrability properties the above result says that the cost per unit time, per particle measured as $\frac{1}{Nn}\sum_{i=1}^N \sum_{m=1}^n \psi(X_m^{i,N})$ can be well approximated by $\int \psi(x)\mu_\infty(dx)$ for large values of n, N, thus the latter quantity can be taken as a measure of performance for a large system over a long time interval. We also show that for each N, there is unique invariant measure Π_∞^N of the N-particle dynamics with integrable first moment and this sequence of measures is μ_∞-chaotic, namely as $N \to \infty$, the projection of Π_∞^N on the first k-coordinates converges to $\mu_\infty^{\otimes k}$ for every $k \geq 1$. This propagation of chaos property all the way to $n = \infty$ crucially relies on the uniform in time convergence of μ_n^N to μ_∞. Such a result is important since it says that the steady state of a N-dimensional fully coupled Markovian system has a simple approximate description in terms of a product measure when N is large.

As noted earlier, this dissertation is divided into two parts. The first part consists of Chapter 2 and 3 and studies long time properties of large weakly interacting particles. The terms **IP** and **WIPS** are used as abbreviations for "*Intereacting Partcles*" and "*Weakly Interacting Particle Systems*" respectively. An outline of this part is as follows.

1.3 Outline of Part 1

- In Chapter 2, we study a discrete time version of a general Vlasov-McKean process given as a solution of a stochastic differential equation of the form

$$dX(t) = [-BX(t) + f(X(t), \mu(t))]dt + h(X(t), \mu(t))dW(t) \tag{1.3.1}$$

  where $\mu(t) := \mathcal{L}(X(t))$ (i.e. the probability distribution of $X(t)$) and analyze the long time behavior of a N-particle weakly interacting Markov process associated with this system. In a setting where the state space of the particles is compact such questions have been studied in previous works, however for the case of an unbounded state space very few results are available. Under suitable assumptions on the problem data we study several time asymptotic properties of the N-particle system and the associated nonlinear Markov chain. In particular we show that the evolution equation for the law of the nonlinear Markov chain has a unique fixed point and starting from an arbitrary initial condition convergence to the fixed point occurs at an exponential rate. The empirical measure μ_n^N of the N-particles at time n is shown to converge to the law

μ_n of the nonlinear Markov process at time n, in the Wasserstein-1 distance, in L^1, as $N \to \infty$, *uniformly* in n. Several consequences of this uniform convergence are studied, including the interchangeability of the limits $n \to \infty$ and $N \to \infty$ and the propagation of chaos property at $n = \infty$. Rate of convergence of μ_n^N to μ_n is studied by establishing uniform in time polynomial and exponential probability concentration estimates. This work has been accepted for publication in *Stochastic Analysis and Applications.*

- In Chapter 3, a system of N particles in discrete time that models active chemotaxis is introduced. This model is motivated by the following system of equations in continuous time that have been proposed in bio-physics literature to describe cellular transport mechanisms that are governed by a potential field and where the cells themselves dynamically influence the field through their aggregated inputs. $\forall i = 1, \ldots, N$

$$\begin{aligned} dX_i(t) &= [-AX_i(t) + \nabla h(t, X_i(t))]dt + dW_i(t), \quad X_i(0) = x_i \in \mathbb{R}^d \\ \frac{\partial}{\partial x} h(t,x) &= -\alpha h(t,x) + D \triangle h(t,x) + \frac{\beta}{N} \sum_{i=1}^{N} g(X_i(t), x). \end{aligned} \tag{1.3.2}$$

where $h(0, \cdot) = h(\cdot)$ for some function $h(\cdot)$. Here $W_i(t), i = 1, \ldots, N$ are independent Brownian motions that drive the state process X_i of the N interacting particles. The interaction between the particles arises indirectly through the underlying potential field h which changes continuously according to a diffusion equation and through the aggregated input of the N particles. Diffusion of the chemical in the medium is captured by the Laplacian in (1.3.2) and the constant $\alpha > 0$ models the rate of decay or the dissipation of the chemical. Contribution of the agents to the chemical concentration field is given through the last term in the equation. The function g captures the agent response rules and can be used to model a wide range of phenomenon [70]. We introduce a discrete time analogue of the above system and study the N-particle weakly interacting Markov process associated with it. Analogous long time asymptotic properties as for the model in chapter 2, are studied.

We now describe the work in the second part of this dissertation. The philosophy of Generalized Fiducial Inference evolved from extended motivation of Fisher's fiducial argument. Fisher could not accept the Bayes/ Laplace postulate for the non-informative prior. He argued that

"Not knowing the chance of mutually exclusive events and knowing the chance to be equal are two quite different states of knowledge"[73].

He only approved the usage of Bayes' theorem for the case of informative priors since imposing any measure on the parameter space is contrary to "no-information" assumption. But his proposal created some serious controversies once his contemporaries realized that this approach often led to procedures that were not exact in frequentist sense and did not possess other properties claimed by him. In a complete manner [51] gives a list of all references regarding this and subsequent Fiducial approaches.

Much after Fisher, in context of generalized confidence interval Tsui, Weerahandi [76], [80] suggested a new approach for constructing hypothesis testing using the concept of generalized P-values. Hannig, Iyer and Patterson [53] made a direct connection between fiducial intervals and generalized confidence intervals and proved asymptotic frequentist correctness of such intervals. These ideas took a general shape in [51] through applications in various parametric model formulations which is now termed as **Generalized Fiducial Inference** (in short **GFI**). From Fisher [31], [32] one of the goals of Fiducial inference have been to formulate a clear and definite principle that would guide a statistician to a unique fiducial distribution. **GFI** does not have such aim and quite different from that perspective. It treats the techniques as a tool in order to propose a distribution on the parameter space when no prior information is available and uses this distribution to propose useful statistical procedures like uncertainty quantification, or approximate confidence interval, etc.

Suppose $\mathbb{X} = G(\mathbb{U}, \theta)$ is the structural equation through which the data have been generated (under the randomness of $\mathbb{U}$ whose distribution doesn't involve unknown parameter θ). Generalizing Fisher's philosophy, after observing $\mathbb{X} = \mathbf{x}$, the Fiducial distribution $P^{\mathcal{G}}(\cdot)$ of θ is formally defined as the distribution of the solution θ in $\mathbf{x} = G(\mathbb{U}, \theta)$ given it exists. That definition can be further generalized as the distribution of the weak limit (as $\epsilon \downarrow 0$) of the following random quantity

$$\arg\inf_{\theta} \left\|\mathbf{x} - G(U^*, \theta)\right\| \Bigg| \left\{\inf_{\theta} \left\|\mathbf{x} - G(U^*, \theta)\right\| \leq \epsilon\right\} \tag{1.3.3}$$

conditioned on the event $\left\{\inf_{\theta} \left\|\mathbf{x} - G(U^*, \theta)\right\| \leq \epsilon\right\}$ that solution of θ exists in ϵ neighborhood of $\mathbf{x}$. The $\epsilon \to 0$ weak limit in (1.3.3) is considered as the **Generalized Fiducial Distribution** (in short **GFD**) of θ. Now using $\left\|\cdot\right\| = L_\infty$ norm in (1.3.3), applying *increasing precision asymptotics* Hannig[52] showed that the distribution of the unique weak limit has a density of the following form

$$f^G(\theta|\mathbb{X}) \quad = \quad \frac{f_{\mathbb{X}}\left(\mathbf{x}\middle|\theta\right) J_n(\mathbb{X}, \theta)}{\int_{\mathbb{R}} f_{\mathbb{X}}\left(\mathbf{x}\middle|\theta'\right) J_n(\mathbb{X}, \theta') d\theta'}, \tag{1.3.4}$$

where $J_n(\mathbb{X}, \theta)$ is the Jacobian defined in (4.1.7) of Chapter 4 which is unique up to proportional scale. Note that however it might look like a posterior distribution with a data dependent prior proportional to $J_n(\mathbb{X}, \theta)$ technically the derivation doesn't involve Bayes' Theorem.

The Generalized Fiducial Distribution enjoys a number of properties like parametrization invariance, consistency (asymptotically attains the weak limit with a technique similar to Bernstein-Von Misses Theorem) etc. One typical non-uniqueness problem persists due to different choices of the data generating equations and different norms (here we exclusively worked with L_∞ norm). In order to remove the first problem partially we wanted to propose an efficient data generating equation where the fiducial distribution satisfies some desired higher order asymptotic properties.

This part of the dissertation consists of only chapter 4.

1.4 Outline of Chapter 4

In this chapter we study frequentist property of the Fiducial quantile with an exploration of higher order asymptotics. Let $J(\theta_0, \theta)$ be the limit of the Jacobian $J_n(\mathbb{X}, \theta)$ after suitably scaled. Let $\hat{\theta}_n$, $I(\theta)$ be the maximum likelihood of θ and the Fisher information respectively. Denote $\frac{\partial}{\partial\theta} J_n(\mathbb{X}, \theta)\Big|_{\theta=\hat{\theta}_n}$ and $\frac{\partial}{\partial\theta} J(\theta_0, \theta)\Big|_{\theta=\hat{\theta}_n}$ by $J_n^{(1)}(\mathbb{X}, \hat{\theta}_n)$ and $J^{(1)}(\theta_0, \hat{\theta}_n)$ respectively. Under some general conditions, we established a higher order expansion of frequentist coverage of the fiducial quantile. Despite similarities of 1st and 2nd order probability matching terms with Bayesian contexts starting from the third order terms there exist significant differences due to the presence of the following random quantities

$$W_n^{(i)}(\mathbb{X}) := \sqrt{n}\left(\frac{J_n^{(1)}(\mathbb{X},\hat{\theta}_n)}{J_n(\mathbb{X},\hat{\theta}_n)} - \frac{J^{(1)}(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}\right)$$

which appear in the expansion of Fiducial quantile as an additive term. The remaining derivation follows by using Shrinkage method. Under a number of regularity assumptions the main result regarding the frequentist coverage of the fiducial quantile $\theta^{1-\alpha}(\mathcal{G},\mathbb{X},n)$ of order $(1-\alpha)$ follows:

$$P_{\theta_0}\Big[\theta_0 \leq \theta^{1-\alpha}(\mathcal{G},\mathbb{X},n)\Big] - \big(1-\alpha\big) \quad = \quad \frac{c_{1,\alpha}\Delta_1}{\sqrt{n}} + \frac{c_{2,\alpha}\Delta_2}{n} + o\bigg(\frac{1}{n}\bigg), \quad \text{where}$$

$$\begin{aligned}
\Delta_1 &= I_{\theta_0}^{-\frac{1}{2}}\frac{\frac{\partial}{\partial\theta}J(\theta_0,\theta)}{J(\theta_0,\theta_0)}\bigg|_{\theta_0} + \frac{\partial I_\theta^{-\frac{1}{2}}}{\partial\theta}\bigg|_{\theta_0}, \\
\Delta_2 &= \frac{I_{\theta_0}^{-\frac{1}{2}}}{z_\alpha g_2}\bigg[a_1(\theta_0) - \frac{a_0(\theta_0)g_1}{g_2}\bigg] + \frac{1}{6}\frac{\partial}{\partial\theta}\left\{I_\theta^{-2}J(\theta_0,\theta)E_\theta\left[l^{(3)}(\theta\mid X)\right]\right\}\bigg|_{\theta_0} \\
&\quad -\frac{1}{2}\frac{\partial^2}{\partial\theta^2}J(\theta_0,\theta)I_\theta^{-1}\bigg|_{\theta_0}
\end{aligned} \tag{1.4.1}$$

and for some constants $c_{1,\alpha} = \phi(z_\alpha), c_{2,\alpha} = z_\alpha\phi(z_\alpha)$. In (1.4.1) z_α is a $(1-\alpha)$th quantile of normal distribution and $a_1(\theta_0), a_2(\theta_0)$ are expected scaled fluctuations of the Jacobian $J_n(\mathbb{X},\theta_0)$ and its derivative $J_n^{(1)}(\mathbb{X},\theta_0)$ respectively under true θ_0. Eventually we found the transformation $A(\cdot)$ that yields the first order exact (i.e. $\Delta_1 = 0$) fiducial distribution for those pathological cases (Scaled normal family, correlation coefficient in bivariate normal model etc).

Chapter2
LONG TIME ASYMPTOTICS OF SOME WIPS

2.1 Introduction

Stochastic dynamical systems that model the evolution of a large collection of weakly interacting particles have long been studied in statistical mechanics (cf. [55; 74] and references therein). In recent years such models have been considered in many other application areas as well, some examples include, chemical and biological systems(e.g. biological aggregation, chemotactic response dynamics [40; 70; 72]), mathematical finance (e.g. mean field games [17; 50], default clustering in large portfolios [45]), social sciences (e.g. opinion dynamics models [22; 46]), communication systems [1; 44; 49] etc. Most of the existing work considers behavior of systems over a finite time horizon. Here we are interested in the time asymptotic properties. Such behavior for special families of weakly interacting particle systems has been considered by several authors. In [27] the authors give general sufficient conditions for a family of discrete time systems for uniform in time exponential probability concentration estimates to hold. These conditions formulated in terms of Dobrushin's coefficient are not very restrictive when the state space of the particles is compact, however they are hard to verify for settings with an unbounded state space. In [15] a discrete time model with a compact state space for chemotactic cell response dynamics was studied. Several time asymptotic results, including uniform in time law of large numbers, exponential stability of the associated nonlinear Markov process and uniform in time convergence of a particle based simulation scheme are established. For the setting of an unbounded state space and in continuous time, there have been several recent interesting works on granular media equations [14; 19; 58] which establish uniform in time propagation of chaos, time uniform convergence of simulation schemes and uniform in time exponential concentration estimates.

In current work we study a family of interacting particle systems with an unbounded state space in discrete time. Although the form of the nonlinearity can be quite general, we require its contribution to the dynamics to be suitably small. The weakly interacting system and the corresponding nonlinear Markov process we consider evolves in $\mathbb{R}^d$ and is described in terms of a stochastic evolution equation of the following form. Denoting by $X_n^i \equiv X_n^{i,N}$ the state of the i-th particle $(i = 1, \dots, N)$ at time instant n, the evolution is given as

$$X_{n+1}^i = AX_n^i + \delta f(X_n^i, \mu_n^N, \epsilon_{n+1}^i) + g(\epsilon_{n+1}^i), \qquad i = 1, .., N, \quad n \in \mathbb{N}_0 \tag{2.1.1}$$

Here $\mu_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_n^i}$ is the empirical measure of the particle values at time instant n, A is a $d \times d$ matrix, δ is a small parameter, $\{\epsilon_n^i, i = 1, ..., N, \quad n \geq 1\}$ is an i.i.d array of $\mathbb{R}^m$ valued random variables with common probability law θ and $f : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^m \to \mathbb{R}^d$, $g : \mathbb{R}^m \to \mathbb{R}^d$ are measurable functions, where $\mathcal{P}(\mathbb{R}^d)$ denotes the space of probability measures on $\mathbb{R}^d$. Also, $\{X_0^i, i = 1, ..., N\}$ are taken to be exchangeable with common distribution μ_0. As will be seen in Section 2.3, the following nonlinear Markov chain will correspond to the $N \to \infty$ limit of (2.1.1).

$$X_{n+1} = AX_n + \delta f(X_n, \mu_n, \epsilon_{n+1}) + g(\epsilon_{n+1}), \quad \mathcal{L}(X_n) = \mu_n, \; n \in \mathbb{N}_0. \tag{2.1.2}$$

where throughout we denote by $\mathcal{L}(X)$ the probability distribution of a random variable X with values in some Polish space S. Stochastic evolution equations as in (2.1.1) can be used to model many different systems with a large number of dynamic interacting particles. Depending on the application, $\{X_n^i\}$ can represent internal states of a collection of financial firms, physical states of biological entities, opinions in a group of peers, loads on links in a communication network, etc. Stochastic systems in (2.1.1) can also be viewed as discrete time approximations for many stochastic differential equation models for weakly interacting particles. Under conditions on f, g, θ, δ and A we study several long time properties of the N-particle system and the associated nonlinear Markov chain. The stochastic dynamical system (2.1.1) can be regarded as a perturbation of a linear stable stochastic dynamical system with a small interaction term and our results give explicit range of values of the perturbation parameter δ for which the weakly interacting system has many desirable long time properties. We are particularly interested in approximating the distribution of X_n by the empirical measure μ_n^N of the particle system, uniformly in n, with explicit uniform concentration bounds. Such results are particularly useful for developing simulation methods for approximating the steady state distribution of mean field models such as in (2.1.2). We note here that we view the systems (2.1.1)–(2.1.2) in two different ways. One is where N is not too large and the N-particle system is used to obtain a simulation based approximation to the nonlinear system (2.1.2) and the second is when the physical system of interest is (2.1.1) but N is too large to allow for a tractable analysis and one instead uses (2.1.2) as a simplified approximate model. In other words, we use the nonlinear system (2.1.2) as an intermediate model to approximate the properties of the physical system (2.1.1) with a large N by those of a simulated system with a numerically tractable number of particles.

Our starting point is the evolution equation for the law of the nonlinear Markov chain given in (2.2.2). We show in Proposition 2.3.2 that under conditions, that include a Lipschitz property of f with the Wasserstein-1($\mathcal{W}_1$) distance on the space of probability measures (Assumptions 1 and 2), contractivity of A (Assumption 3) and δ being sufficiently small, (2.2.2) has a unique fixed point and starting from an arbitrary initial condition convergence to the fixed point occurs at an exponential rate. Using this result we next argue in Theorem 2.3.1 that under an additional integrability condition (Assumption 4), as $N \to \infty$, the empirical measure μ_n^N of the N-particles at time n converges to the law μ_n of the nonlinear Markov process at time n, in the $\mathcal{W}_1$ distance, in L^1, *uniformly* in n. We next study the rate of this uniform convergence by developing suitable probability concentration estimates. Such results are useful for constructing uniform in time confidence intervals for various quantities of interest. The first result (Theorem 2.3.2), under an assumption of polynomial moments on the initial data and noise sequence (Assumption 4) establishes a corresponding uniform in time polynomial concentration bound. The proof relies on an idea of restricting measures to a compact set and estimates on metric entropy introduced in [14] (see also [79]). The basic idea is to first obtain a concentration bound for the $\mathcal{W}_1$ distance between the truncated law and its corresponding empirical law in a compact ball of radius R along with an estimate on the contribution from the region outside the ball and finally optimize suitably over R. The last two results are concerned with exponential concentration. These impose much stronger integrability conditions on the problem data (Assumption 5). The first considers the setting where the initial random variables form a general exchangeable sequence and gives a concentration bound with an exponential decay rate of $N^{\frac{1}{d+2}}$. The second result uses exponential concentration estimates for empirical measures of i.i.d. sequences based on transportation inequalities from [11; 13] (see also [13; 14; 29; 47; 48]) and considers the setting where the initial data is i.i.d. In this case the concentration bound gives an exponential decay rate of order N. As noted earlier, in continuous time, results analogous to those in this chapter for the specific setting of McKean-Vlasov diffusions in a convex potential have been obtained in many papers, however to the best of our knowledge there is no current work that covers the discrete stochastic dynamical system setting with an unbounded state space of the form considered here.

The following notation will be used in this work. $\mathbb{R}^d$ will denote the d dimensional Euclidean space with the usual Euclidean norm $|.|$. The set of natural numbers (resp. whole numbers) is denoted by $\mathbb{N}$ (resp. $\mathbb{N}_0$). Cardinality of a finite set S is denoted by $|S|$. For a measurable space S, $\mathcal{P}(S)$ denotes the space of all probability measures on S. For $x \in \mathbb{R}^d$, δ_x is the Dirac delta measure on $\mathbb{R}^d$ that puts a unit mass at location x. The space of real valued bounded measurable functions on S is denoted as $BM(S)$. Borel σ field on a metric space will be denoted as $\mathcal{B}(S)$. $\mathcal{C}_b(S)$ denotes the space of all bounded and continuous functions $f : S \to \mathbb{R}$. The supremum norm of a function $f : S \to \mathbb{R}$ is $\|f\|_\infty = \sup_{x\in S} |f(x)|$. When S is a metric space, the Lipschitz seminorm of f is defined by $\|f\|_1 = \sup_{x\neq y} \frac{|f(x)-f(y)|}{d(x,y)}$ where d is the metric on the space S. For a bounded Lipschitz function f on S we define $\|f\|_{BL} := \|f\|_1 + \|f\|_\infty$. $\mathrm{Lip}_1(S)$ (resp. $BL_1(S)$) denotes the class of Lipschitz (resp. bounded Lipschitz) functions $f : S \to \mathbb{R}$ with $\|f\|_1$ (resp. $\|f\|_{BL}$) bounded by 1. Occasionally we will suppress S from the notation and write Lip_1 and BL_1

when clear from the context. For a Polish space S, $\mathcal{P}(S)$ is equipped with the topology of weak convergence. A convenient metric metrizing this topology on $\mathcal{P}(S)$ is given as $\beta(\mu,\gamma) = \sup\{|\int f d\mu - \int f d\gamma| : \|f\|_{BL_1} \le 1\}$ for $\mu, \gamma \in \mathcal{P}(S)$. For a signed measure γ on $\mathbb{R}^d$, we define $\langle f, \gamma\rangle := \int f d\gamma$ whenever the integral makes sense. Let $\mathcal{P}_1(\mathbb{R}^d)$ be the space of $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that

$$\|\mu\|_1 := \int |x| d\mu(x) < \infty.$$

The space $\mathcal{P}_1(\mathbb{R}^d)$ will be equipped with the Wasserstein-1 distance that is defined as follows:

$$\mathcal{W}_1(\mu_0,\gamma_0) := \inf_{X,Y} E|X-Y|$$

where the infimum is taken over all $\mathbb{R}^d$ valued random variables X,Y defined on a common probability space and where the marginals of X, Y are respectively μ_0 and γ_0. From Kantorovich-Rubenstein duality (cf. [79]) one sees the Wasserstein-1 is same as

$$\mathcal{W}_1(\mu_0,\gamma_0) = \sup_{f \in \mathrm{Lip}_1(\mathbb{R}^d)} |\langle f, \mu_0 - \gamma_0\rangle|. \tag{2.1.3}$$

For a signed measure μ on $(S, \mathcal{B}(S))$, the total variation norm of μ is defined as $|\mu|_{TV} := \sup_{\|f\|_\infty \le 1} \langle f, \mu\rangle$. Convergence in distribution of a sequence $\{X_n\}_{n\ge 1}$ of S valued random variable to X will be written as $X_n \Rightarrow X$.

A finite collection $\{Y_1, Y_2, \ldots, Y_N\}$ of S valued random variables is called exchangeable if

$$\mathcal{L}(Y_1, Y_2, \ldots, Y_N) = \mathcal{L}(Y_{\pi(1)}, Y_{\pi(2)}, \ldots, Y_{\pi(N)})$$

for every permutation π on the N symbols $\{1, 2, \ldots, N\}$. Let $\{Y_i^N, i = 1, .., N\}_{N\geq 1}$ be a collection of S valued random variables, such that for every N, $\{Y_1^N, Y_2^N, \ldots, Y_N^N\}$ is exchangeable. Let $\nu_N = \mathcal{L}(Y_1^N, Y_2^N, \ldots, Y_N^N)$. The sequence $\{\nu_N\}_{N\geq 1}$ is called ν -chaotic (cf. [74]) for a $\nu \in \mathcal{P}(\mathcal{S})$, if for any $k \geq 1$, $f_1, f_2, \ldots, f_k \in \mathcal{C}_b(\mathcal{S})$, one has

$$\lim_{N\to\infty} \langle f_1 \otimes f_2 \otimes ... \otimes f_k \otimes 1 ... \otimes 1, \nu_N \rangle = \prod_{i=1}^{k} \langle f_i, \nu \rangle. \tag{2.1.4}$$

Denoting the marginal distribution on first k coordinates of ν_N by ν_N^k, equation (2.1.4) says that, for every $k \geq 1$, $\nu_N^k \to \nu^{\otimes k}$.

2.2 Model Description

Recall the system of N interacting particles in $\mathbb{R}^d$ introduced in (2.1.1). Throughout we will assume that $\{X_0^i, i = 1, ..., N\}$ is exchangeable with common distribution μ_0 where $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$. Assumptions on f, θ, δ and A will be introduced shortly. Note that in the notation we have suppressed the dependence of the sequence $\{X_n^i\}$ on N. Given $\rho \in \mathcal{P}(\mathbb{R}^d)$ define a transition probability kernel $P^\rho : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ as

$$P^\rho(x, C) = \int_{\mathbb{R}^m} 1_{[Ax+\delta f(x,\rho,z)+g(z)\in C]}\theta(dz), \qquad (x, C) \in \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d).$$

With an abuse of notation we will also denote by P^ρ the map from $BM(\mathbb{R}^d)$ to itself, defined as

$$P^\rho \phi(x) = \int_{\mathbb{R}^d} \phi(y) P^\rho(x, dy), \quad \phi \in BM(\mathbb{R}^d), x \in \mathbb{R}^d.$$

For $\mu \in \mathcal{P}(\mathbb{R}^d)$, let $\mu P^\rho \in \mathcal{P}(\mathbb{R}^d)$ be defined as

$$\mu P^\rho(C) = \int_{\mathbb{R}^d} P^\rho(x, C)\mu(dx), \quad C \in \mathcal{B}(\mathbb{R}^d).$$

Note that $\mu P^\rho = \mathcal{L}(AX + \delta f(X, \rho, \epsilon) + g(\epsilon))$ when $\mathcal{L}(X, \epsilon) = \mu \otimes \theta$.

Under Assumptions 1 and 2 introduced in the next section it will follow that, for $\rho, \mu \in \mathcal{P}_1(\mathbb{R}^d)$, $\mu P^\rho \in \mathcal{P}_1(\mathbb{R}^d)$ as well. Under these conditions, one can define $\Psi : \mathcal{P}_1(\mathbb{R}^d) \to \mathcal{P}_1(\mathbb{R}^d)$ as

$$\Psi(\mu) = \mu P^\mu. \tag{2.2.1}$$

Then the evolution of the law of the nonlinear Markov chain given in (2.1.2) is given by the equation

$$\mu_{n+1} = \Psi(\mu_n),\ n \in \mathbb{N}_0. \tag{2.2.2}$$

Using the above notation we see that $((X_n^1, ..., X_n^N), \mu_n^N)$ is a $(\mathbb{R}^d)^N \times \mathcal{P}_1(\mathbb{R}^d)$ valued discrete time Markov chain defined recursively as follows. Let $X_k(N) \equiv (X_k^1, X_k^2, \dots, X_k^N)$ and let $\mathcal{F}_0 = \sigma\{X_0(N)\}$. Then, for $k \geq 1$

$$\begin{cases} P(X_k(N) \in C | \mathcal{F}_{k-1}^N) = \bigotimes_{j=1}^N (\delta_{X_{k-1}^j} P^{\mu_{k-1}^N})(C) \quad \forall C \in \mathcal{B}(\mathbb{R}^d)^N \\ \mu_k^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_k^i} \\ \mathcal{F}_k^N = \sigma\{X_k(N)\} \vee \mathcal{F}_{k-1}^N. \end{cases}$$

2.3 Main Results

Recall that $\{X_0^i, i = 1, ..., N\}$ is assumed to be exchangeable with common distribution μ_0 where $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$. We now introduce our assumptions on the nonlinearity.

Assumption 1. *There is a measurable map $D : \mathbb{R}^m \to \mathbb{R}_+$ such that $\int D(z)\theta(dz) = \sigma < \infty$, and*

$$\sup_{x_1 \neq x_2, \mu_1 \neq \mu_2, x_1, x_2 \in \mathbb{R}^d, \mu_1, \mu_2 \in \mathcal{P}_1(\mathbb{R}^d)} \frac{|f(x_1, \mu_1, z) - f(x_2, \mu_2, z)|}{|x_1 - x_2| + \mathcal{W}_1(\mu_1, \mu_2)} \leq D(z), \text{ for all } z \in \mathbb{R}^m. \tag{2.3.1}$$

Note that the Assumption 1 implies that

$$\sup_{(x,\mu) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)} |f(x, \mu, z)| \leq (|x| + \|\mu\|_1)D(z) + D_1(z), \ z \in \mathbb{R}^m \tag{2.3.2}$$

where $D_1(z) := |f(0, \delta_0, z)|$. We impose the following condition on D_1 and g.

Assumption 2. $\int D_1(z)\theta(dz) = c_0 < \infty$, $\int |g(z)|\theta(dz) = \tilde{c}_0 < \infty$.

Remark 1. *1. One simple example of f that corresponds to the setting of stochastic difference equations is given as:*

$$f(x, \mu, z) = f_1(x, \mu) + f_2(x, \mu)z, \ (x, \mu, z) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^m$$

where f_1 and f_2 are Lipschitz in (x, μ) and $\int |z|\theta(dz) < \infty$.

2. *Suppose* $f_0 : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ *is such that for* $(x_i, y_i, z) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^m, \ i = 1, 2$

$$|f_0(x_1, y_1, z) - f_0(x_2, y_2, z)| \le D(z)(|x_1 - x_2| + |y_1 - y_2|),$$

then $f(x, \mu, z) = \int f_0(x, y, z)\mu(dy)$ *satisfies the Lipschitz property in* (2.3.1). *More generally, suppose* $f_0 : \mathbb{R}^d \times (\mathbb{R}^d)^p \times \mathbb{R}^m \to \mathbb{R}$ *is Lipschitz in the first* $p + 1$ *variables with the Lipschitz parameter* $D_p(\cdot)$ *given as a function of the last variable. Then* $f(x, \mu, z) = \int f(x, y_1, \ldots y_p, z)\mu^{\otimes p}(y_1, \ldots, y_p)$ *satisfies the Lipschitz property* (2.3.1) *with* $D(z) = pD_p(z)$.

We first present a law of large numbers for μ_n^N as $N \to \infty$. The proof is standard but we include it here for completeness. Note that under Assumptions 1 and 2, $\mu_n \in \mathcal{P}_1(\mathbb{R}^d)$ for all $n \in \mathbb{N}_0$. Observing that $\nu \mapsto \mathcal{W}_1(\nu, \mu_n)$ is a continuous map from the Polish space $(\mathcal{P}_1(\mathbb{R}^d), \mathcal{W}_1)$ to $\mathbb{R}_+$ and that μ_n^N is a $\mathcal{P}_1(\mathbb{R}^d)$ valued random variable, we see that

$$\mathcal{W}_1(\mu_n^N, \mu_n) = \sup_{\psi \in \mathrm{Lip}_1} |\langle \psi, \mu_n^N - \mu_n \rangle|$$

is a nonnegative random variable.

Proposition 2.3.1. *Suppose Assumptions 1 and 2 hold and suppose that* $E\mathcal{W}_1(\mu_0^N, \mu_0) \to 0$ *as* $N \to \infty$. *Then, as* $N \to \infty$,

$$E\mathcal{W}_1(\mu_n^N, \mu_n) \to 0 \tag{2.3.3}$$

for all $n \ge 0$.

Remark 2. *Note that Proposition 2.3.1 says that for all $n \geq 0$*

$$\lim_{N\to\infty} E \sup_{\psi \in Lip_1} |\langle \psi, \mu_n^N - \mu_n \rangle| = 0,$$

which in particular implies that $\mu_n^N \to \mu_n$ in probability, in $\mathcal{P}(\mathbb{R}^d)$ (with the topology of weak convergence) as $N \to \infty$.

Next we state a "propagation of chaos" result which is an immediate consequence of Remark 5 and exchangeability of of $\{X_n^i\}_{i=1}^N$.

Corollary 2.3.1. *Suppose Assumptions 1 and 2 hold. Then for any $k \geq 1$ and $n \in \mathbb{N}_0$, $\mathcal{L}(X_n^1, X_n^2, ..., X_n^k) \longrightarrow (\mathcal{L}(X_n))^{\bigotimes k}$ as $N \to \infty$.*

For a $d \times d$ matrix B we denote its norm by $\|B\|$, i.e. $\|B\| = \sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{|Bx|}{|x|}$.

Assumption 3. $\|A\| < 1$.

A measure $\mu^* \in \mathcal{P}_1(\mathbb{R}^d)$ is called a fixed point for the evolution equation in (2.1.2), if $\mu^* = \Psi(\mu^*)$. Let $a_0 = \frac{1-\|A\|}{2\sigma}$.

Proposition 2.3.2. *Suppose Assumptions 1,2 and 3 hold and that $\delta \in (0, a_0)$. Then there exists a unique fixed point μ_∞ of equation (2.2.2). Furthermore, denoting for $\gamma \in \mathcal{P}_1(\mathbb{R}^d)$, $\Psi^n(\gamma) = \underbrace{\Psi \circ \Psi ... \circ \Psi}_{n \text{ times}}(\gamma)$, we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathcal{W}_1(\Psi^n(\gamma), \mu_\infty) < 0,$$

namely $\Psi^n(\gamma)$ converges to μ_∞ as $n \to \infty$, at an exponential rate.

Next, we study uniform in time (i.e n) convergence of μ_n^N to μ_n as the number of particles $N \to \infty$. For this, we will strengthen Assumptions 1 and 2 as follows.

Assumption 4. *For some $\alpha > 0$*

$$E|X_0^i|^{1+\alpha} < \infty, \quad \int D(z)^{1+\alpha}\theta(dz) = \sigma_1(\alpha) < \infty$$
$$\int (D_1(z))^{1+\alpha}\theta(dz) = c_1(\alpha) < \infty, \quad \int |g(z)|^{1+\alpha}\theta(dz) = \tilde{c}_1(\alpha) < \infty.$$

Theorem 2.3.1. *Suppose that Assumptions 3 and 4 hold. Also suppose that $\delta \in (0, a_0)$. Then*

1. *Given $\varepsilon > 0$, there exist $N_0(\varepsilon), n_0(\varepsilon) \in \mathbb{N}$ such that*

$$E\mathcal{W}_1(\mu_n^N, \mu_n) < \varepsilon \qquad \textit{whenever} \quad n \geq n_0(\varepsilon), N \geq N_0(\varepsilon).$$

2. *Suppose $E\mathcal{W}_1(\mu_0^N, \mu_0) \to 0$ as $N \to \infty$. Then $\sup_{n\geq 1} E\mathcal{W}_1(\mu_n^N, \mu_n) \to 0$ as $N \to \infty$.*

Corollary 2.3.2. *Suppose Assumptions 3 and 4 hold and suppose $\delta \in (0, a_0)$. Then*

$$\limsup_{N\to\infty} \limsup_{n\to\infty} E\mathcal{W}_1(\mu_n^N, \mu_\infty) = \limsup_{n\to\infty} \limsup_{N\to\infty} E\mathcal{W}_1(\mu_n^N, \mu_\infty) = 0. \tag{2.3.4}$$

The interchangeability of the limits given in Corollary 4.2.1 allows one to characterize the large N limit of the steady state behavior of the particle system.

Proposition 2.3.3. *Suppose Assumptions 1, 2 and 3 hold and suppose $\delta \in (0, a_0)$. Then for every $N \in \mathbb{N}$, the Markov chain $\{X_n(N)\}_{n\geq 0}$ has a unique invariant measure Π_∞^N that* *satisfies* $\int_{(\mathbb{R}^d)^N} |x| \, \Pi_\infty^N(dx) < \infty$. *Suppose in addition Assumption 4 holds. Then* Π_∞^N *is* μ_∞-*chaotic, where* μ_∞ *is as in Proposition 2.3.2.*

Theorem 2.3.1 gives conditions under which $\mathcal{W}_1(\mu_n^N, \mu_n)$ converges to 0 as $N \to \infty$, in L^1, uniformly in n. The next three theorems show that under additional conditions, one can provide concentration bounds uniformly in n which give estimates on the rate of convergence. Recall the measure μ_0 introduced at the beginning of Section 3.2.

With $\alpha, \sigma_1(\alpha)$ defined in Assumption 4 and ω as in Assumption 3, let

$$a \equiv a(\alpha) := \frac{4^{-\alpha} - \|A\|^{(1+\alpha)}}{2\sigma_1(\alpha)}. \tag{2.3.5}$$

Theorem 2.3.2. *Suppose Assumptions 3 and 4 holds. Fix* $\gamma_0 \in (0, a_0)$ *and suppose that* $\delta \in (0, \min\{a^{\frac{1}{1+\alpha}}, (a_0 - \gamma_0)\})$. *Let* $\vartheta = \frac{1-2\sigma\gamma_0}{\|A\|+2\delta\sigma}$. *Then there exists* $N_0 \in \mathbb{N}_0$ *and* $C_1 \in (0, \infty)$ *such that for all* $\varepsilon > 0$, *and for all* $n \geq 0$,

$$P(\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon) \leq P(\mathcal{W}_1(\mu_0^N, \mu_0) > 2\sigma\gamma_0\vartheta^n\varepsilon) + C_1\varepsilon^{-(1+\alpha)}N^{-\frac{\alpha}{d+2}},$$

for all $N > N_0 \left(\max\left\{1, \log^+ \varepsilon\right\}\right)^{\frac{d+2}{d}}$.

Remark 3. *1. Since* $\delta < a_0 - \gamma_0$, *we have that* $\vartheta > 1$ *and so the above theorem gives the following uniform concentration estimate:*

$$\sup_{n \geq 1} P(\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon) \leq P(\mathcal{W}_1(\mu_0^N, \mu_0) > 2\sigma\gamma_0\varepsilon) + C_1\varepsilon^{-(1+\alpha)}N^{-\frac{\alpha}{d+2}},$$

for all $N > N_0 \left(\max\left\{1, \log^+ \varepsilon\right\}\right)^{\frac{d+2}{d}}$.

2. *Under additional conditions on $\{X_0^{i,N}\}$ one can give concentration bounds for the first term on the right side of the above inequality. For example, when $\{X_0^{i,N}\}_{i=1}^N$ are i.i.d. such concentration bounds can be found in Theorem 2.7 of [14]. Also, although not pursued here, the bound obtained in Theorem 2.3.2 can be used to give an alternative proof of Theorem 2.3.1(2).*

Next we obtain exponential concentration bounds. The bounds depend in particular on our assumptions on the initial condition. Our first result (Theorem 2.3.3) treats the case where the initial random vector has a general exchangeable distribution while the second result (Theorem 2.3.4) considers a more restrictive setting where the initial random vector is i.i.d. In the second case the probabilities will decay exponentially in N whereas in the first case the exponent will be some dimensional dependent power of N.

We start with our main assumption for Theorem 2.3.3.

Assumption 5. *(i) For some $M \in (1, \infty)$, $D(x) \leq M$ for θ a.e. $x \in \mathbb{R}^m$.*

(ii) There exists $\alpha \in (0, \infty)$ such that $\int e^{\alpha|x|}\mu_0(dx) < \infty$ and $\int e^{\alpha(D_1(z)+|g(z)|)}\theta(dz) < \infty$.

Theorem 2.3.3. *Suppose that Assumptions 3 and 5 hold. Fix $\gamma_0 \in (0, a_0)$ and suppose that $\delta \in [0, \min\{a_0 - \gamma_0, \frac{1-\|A\|}{2M}\})$. Then there exists $N_0 \in \mathbb{N}$ and $C_1 \in (0, \infty)$ such that for all $\varepsilon > 0$*

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq P[\mathcal{W}_1(\mu_0^N, \mu_0) > 2\sigma\gamma_0\vartheta^n\varepsilon] + e^{-C_1\varepsilon N^{1/d+2}},$$

for all $n \geq 0$, $N \geq N_0 \max\{(\frac{1}{\varepsilon}\log^+\frac{1}{\varepsilon})^{d+2}, \varepsilon^{(d+2)/(d-1)}\}$, if $d > 1$; and

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq P[\mathcal{W}_1(\mu_0^N, \mu_0) > 2\sigma\gamma_0\vartheta^n\varepsilon] + e^{-C_1(\varepsilon\wedge 1)N^{1/d+2}},$$

for all $n \geq 0$, $N \geq N_0 \max\{(\frac{1}{\varepsilon}\log^+ \frac{1}{\varepsilon})^{d+2}, 1\}$, *if* $d = 1$. *Here* $\vartheta \in (1, \infty)$ *is as in Theorem 2.3.2.*

We note that the bounds in Theorems 2.3.2 and 2.3.3 depend on the dimension parameter d. It will be interesting to see if one can obtain useful dimension independent bounds under the conditions of these theorems. The following result shows that such bounds can be obtained in the case where the initial distribution of the N particles is i.i.d. The proof relies on various estimates from [11; 13].

Theorem 2.3.4. *Suppose that* $\{X_0^{i,N}\}_{i=1,..,N}$ *are i.i.d. with common distribution* μ_0 *for each* N. *Suppose that Assumptions 3 and 5 hold. Fix* $\gamma \in (0, 1 - \|A\|)$. *Suppose that* $\delta \in \left[0, \frac{1-\|A\|-\gamma}{2M}\right)$. *Then there exist* $N_0, a_1, a_2 \in (0, \infty)$ *and a nonincreasing function* $\varsigma_1 : (0, \infty) \to (0, \infty)$ *such that* $\varsigma_1(t) \downarrow 0$ *as* $t \uparrow \infty$ *and for all* $\varepsilon > 0$ *and* $N \geq N_0 \varsigma_1(\varepsilon)$

$$\sup_{n \geq 0} P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq a_1 e^{-N a_2(\varepsilon^2 \wedge \varepsilon)}$$

Remark 4. *1. One can describe the function* ς_1 *in the above theorem explicitly. Define for* $\gamma \in (0, 1)$, $m_\gamma : (0, \infty) \to (0, \infty)$ *as* $m_\gamma(t) = \frac{\gamma t}{\delta M}$, *where* M *is as in Assumption 5. Then*

$$\varsigma_1(t) = \max\left\{1, \frac{\log \mathcal{C}^0_{m_\gamma(t)}}{m_\gamma^2(t)}, \frac{\log \mathcal{C}^0_{\gamma t}}{\gamma^2 t^2}, \frac{1}{t^2}, \frac{1}{t}\right\},$$

where $\mathcal{C}_t^0$ *is defined by the right side of* (2.4.49) *with* ζ *replaced by* ζ_0 *where* ζ_0 *is as in Corollary 11.*

2. *If Assumption 5 is strengthened to* $\int e^{\alpha(D_1(z)+|g(z)|)^2}\theta(dz) < \infty$ *for some* $\alpha > 0$ *then one can strengthen the conclusion of Theorem 2.3.4 as follows: For* δ *sufficiently small there exist* $N_0, a_1, a_2 \in (0,\infty)$ *and a nonincreasing function* $\varsigma_2 : (0,\infty) \to (0,\infty)$ *such that* $\varsigma_2(t) \downarrow 0$ *as* $t \uparrow \infty$ *and for all* $\varepsilon > 0$ *and* $N \geq N_0\varsigma_2(\varepsilon)$

$$\sup_{n\geq 0} P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq a_1 e^{-Na_2\varepsilon^2}.$$

2.4 Proofs

In this section we present the proofs of our main results that were presented in Section 2.3. We begin with some preliminary moment estimates.

2.4.1 Moment Bounds

The following elementary lemma will be used several times in our analysis.

Lemma 2. *Suppose Assumptions 1 and 2 hold. Then, for every* $n \geq 1$*,*

$$M_n := \sup_{N\geq 1}\max_{1\leq i\leq N} E|X_n^i| < \infty.$$

In addition, if Assumption 3 holds and $\delta \in (0, a_0)$ *then* $\sup_{n\geq 1} M_n < \infty$.

Proof. We will only prove the second statement in the lemma. Proof of the first statement is similar. Note that, for $n \geq 1$ and $i = 1, .., N$

$$X_n^i = AX_{n-1}^i + \delta f(X_{n-1}^i, \mu_{n-1}^N, \epsilon_n^i) + g(\epsilon_n^i).$$

Thus

$$|X_n^i| \leq \|A\| \; |X_{n-1}^i| + \delta D(\epsilon_n^i)[|X_{n-1}^i| + \|\mu_{n-1}^N\|_1] + \delta D_1(\epsilon_n^i) + |g(\epsilon_n^i|. \tag{2.4.1}$$

From exchangeability of $\{X_{n-1}^k, k = 1, ..., N\}$ it follows that

$$E\|\mu_{n-1}^N\|_1 = E[\int |x| d\mu_{n-1}^N] = E\frac{1}{N}\sum_{k=1}^N |X_{n-1}^k| = E|X_{n-1}^1|.$$

Taking expectation in (2.4.1) and using independence between ϵ_n^i and $\{X_{n-1}^j\}_{j=1}^N$, we have

$$E|X_n^i| \leq (\|A\| + 2\delta\sigma)E|X_{n-1}^i| + \delta c_0 + \tilde{c}_0. \tag{2.4.2}$$

The assumption on δ implies that $\gamma := \|A\| + 2\delta\sigma \in (0,1)$. A recursive application of (2.4.2) now shows that

$$E|X_n^i| \leq \gamma^n E|X_0^i| + \frac{\delta c_0 + \tilde{c}_0}{1-\gamma}.$$

The result follows. □

Recall the map Ψ defined in (2.2.1). The following lemma is a key ingredient in our truncation arguments.

Lemma 3. *Under Assumptions 1 and 2, for every $\epsilon > 0$ and $n \geq 1$, there exists a compact set $K_{\epsilon,n} \in \mathcal{B}(\mathbb{R}^d)$ such that*

$$\sup_{N\geq 1} E\left\{\int_{K_{\epsilon,n}^c} |x| \left(\mu_n^N(dx) + \Psi(\mu_{n-1}^N)(dx)\right)\right\} < \epsilon.$$

Proof. Note that for any non-negative $\phi : \mathbb{R}^d \to \mathbb{R}$,

$$E \int \phi(x)\mu_n^N(dx) = \frac{1}{N}\sum_{k=1}^{N} E\phi(X_n^k) = E\phi(X_n^1). \tag{2.4.3}$$

Also, letting $f_\delta = \delta f + g$, by a conditioning argument we see that for any non-negative $\phi : \mathbb{R}^d \to \mathbb{R}$

$$\begin{aligned} E \int \phi(x)\Psi(\mu_n^N)(dx) &= \frac{1}{N}\sum_{i=1}^{N} E(E(\langle \phi, \delta_{X_n^i} P^{\mu_n^N}\rangle \mid \mathcal{F}_n)) \\ &= \frac{1}{N}\sum_{i=1}^{N} E\phi\left(AX_n^i + f_\delta(X_n^i, \mu_n^N, \epsilon_{n+1}^i)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} E\phi(X_{n+1}^i) = E\phi(X_{n+1}^1). \end{aligned} \tag{2.4.4}$$

In view of the above observations it now suffices to show that

$$\text{the family } \{X_n^{i,N}, i = 1, ..., N; N \geq 1\} \text{ is uniformly integrable for every } n \geq 0. \tag{2.4.5}$$

We will prove (3.4.18) by induction on n. Once more we suppress N from the super-script. Clearly by our assumptions $\{X_0^i, i = 1, ..., N; N \geq 1\}$ is uniformly integrable. Now suppose that the statement (3.4.18) holds for some n. Note that

$$\begin{aligned} |X_{n+1}^i| &\leq \|A\|\,|X_n^i| + \delta D(\epsilon_{n+1}^i)[|X_n^i| + \|\mu_n^N\|_1] + \delta D_1(\epsilon_{n+1}^i) + |g(\epsilon_{n+1}^i)| \\ &= \|A\|\,|X_n^i| + \delta D(\epsilon_{n+1}^i)[|X_n^i| + \frac{1}{N}\sum_{i=1}^{N}|X_n^i|] + \delta D_1(\epsilon_{n+1}^i) + |g(\epsilon_{n+1}^i)| \end{aligned} \tag{2.4.6}$$

From exchangeability it follows $\frac{1}{N}\sum_{i=1}^{N}|X_n^i| = E[|X_n^i| \mid \mathcal{G}_n^N]$, where $\mathcal{G}_n^N = \sigma\{\frac{1}{N}\sum_{i=1}^{N}\delta_{X_n^i}\}$. Combining this with the induction hypothesis that $\{X_n^i, i = 1, ..., N; N \geq 1\}$ is uniformly integrable, we see that $\{\frac{1}{N}\sum_{i=1}^{N}|X_n^i|, N \geq 1\}$ is uniformly integrable. Here we have used the fact that if $\{Z_\alpha, \alpha \in \Gamma_1\}$ is a uniformly integrable family and $\{\mathcal{H}_\beta, \beta \in \Gamma_2\}$ is a collection of σ-fields where Γ_1, Γ_2 are arbitrary index sets, then $\{E(Z_\alpha \mid \mathcal{H}_\beta), (\alpha, \beta) \in \Gamma_1 \times \Gamma_2\}$ is a uniformly integrable family. Also from Assumptions 1 and 2 the families $\{D(\epsilon_{n+1}^i); i \geq 1\}$, $\{D_1(\epsilon_{n+1}^i); i \geq 1\}$ and $\{g(\epsilon_{n+1}^i); i \geq 1\}$ are uniformly integrable. These observations along with independence between $\{\epsilon_{n+1}^i, i = 1, .., N\}$ and $\{X_n^i : i = 1, ..., N; N \geq 1\}$ yield that the family $\{|X_n^i| : i = 1, ..., N; N \geq 1\}$ is uniformly integrable. The result follows. $\square$

2.4.2 Proof of Proposition 2.3.1

We now proceed to the proof of Proposition 2.3.1. We will argue via induction on $n \geq 0$. By assumption (1) holds for $n = 0$. Assume now that it holds for some $n > 0$. Note that,

$$\begin{aligned}\mathcal{W}_1(\mu_{n+1}^N, \mu_{n+1}) &\leq \mathcal{W}_1(\mu_{n+1}^N, \mu_n^N P^{\mu_n^N}) + \mathcal{W}_1(\mu_n^N P^{\mu_n^N}, \mu_n^N P^{\mu_n}) \\ &+\mathcal{W}_1(\mu_n^N P^{\mu_n}, \mu_{n+1}).\end{aligned} \tag{2.4.7}$$

Consider the last term in (3.4.27). Using Assumption 1 we see that if ϕ is Lipschitz then $P^{\mu_n}\phi$ is Lipschitz and $\|P^{\mu_n}\phi\|_1 \leq (\|A\| + \delta\sigma)\|\phi\|_1$. Thus, almost surely

$$\begin{aligned}\sup_{\phi \in \mathrm{Lip}_1} |\langle \phi, \mu_n^N P^{\mu_n} - \mu_{n+1}\rangle| &= \sup_{\phi \in \mathrm{Lip}_1} |\langle P^{\mu_n}\phi, \mu_n^N - \mu_n\rangle| \\ &\leq (\|A\| + \delta\sigma) \sup_{g \in \mathrm{Lip}_1} |\langle g, \mu_n^N - \mu_n\rangle| \quad .\end{aligned}$$

Taking expectations we obtain,

$$E\mathcal{W}_1(\mu_n^N P^{\mu_n}, \mu_{n+1}) \leq (\|A\| + \delta\sigma) E\mathcal{W}_1(\mu_n^N, \mu_n). \tag{2.4.8}$$

Consider now the second term in (3.4.27). Using Assumption 1 again, we have,

$$\begin{aligned}\sup_{\phi\in\mathrm{Lip}_1} |\langle \phi, \mu_n^N P^{\mu_n^N} - \mu_n^N P^{\mu_n}\rangle| &\leq \frac{1}{N}\sum_{i=1}^N \int \left[|\phi(AX_n^i + f_\delta(X_n^i, \mu_n^N, \xi))\right.\\ &- \left.\phi(AX_n^i + f_\delta(X_n^i, \mu_n, \xi))|\right] \theta(d\xi)\\ &\leq \delta\sigma\mathcal{W}_1(\mu_n^N, \mu_n).\end{aligned}$$

Taking expectations we get

$$\begin{aligned}E\mathcal{W}_1(\mu_n^N P^{\mu_n^N}, \mu_n^N P^{\mu_n}) &= E \sup_{\phi\in\mathrm{Lip}_1(\mathbb{R}^d)} |\langle \phi, \mu_n^N P^{\mu_n^N} - \mu_n^N P^{\mu_n}\rangle|\\ &\leq \delta\sigma E\mathcal{W}_1(\mu_n^N, \mu_n). \end{aligned}\tag{2.4.9}$$

Now we consider the first term of the right hand side of (3.4.27). We will use Lemma 4.7.1.

Fix $\epsilon > 0$ and let K_ϵ be a compact set in $\mathbb{R}^d$ such that

$$\sup_{N\geq 1} E\left\{\int_{K_\epsilon^c} |x|(\mu_{n+1}^N(dx) + \Psi(\mu_n^N)(dx))\right\} < \epsilon.$$

Let $\text{Lip}_1^0(\mathbb{R}^d) := \{\psi \in \text{Lip}_1(\mathbb{R}^d) : \psi(0) = 0\}$. Then,

$$\begin{aligned} E \sup_{\phi \in \text{Lip}_1(\mathbb{R}^d)} |\langle \phi, \mu_{n+1}^N - \mu_n^N P^{\mu_n^N} \rangle| &= E \sup_{\phi \in \text{Lip}_1^0(\mathbb{R}^d)} |\langle \phi, \mu_{n+1}^N - \mu_n^N P^{\mu_n^N} \rangle| \\ &\leq E \sup_{\phi \in \text{Lip}_1^0(\mathbb{R}^d)} |\langle \phi.1_{K_\epsilon}, \mu_{n+1}^N - \mu_n^N P^{\mu_n^N} \rangle| + \epsilon. \end{aligned} \tag{2.4.10}$$

We will now apply Lemma A.1.1 in the Appendix. Note that for any $\phi \in \text{Lip}_1^0(\mathbb{R}^d)$, $\sup_{x \in K_\epsilon} |\phi(x)| \leq diam(K_\epsilon) := m_\epsilon$.

Thus with notation as in Lemma A.1.1

$$\sup_{\phi \in \text{Lip}_1^0(\mathbb{R}^d)} |\langle \phi.1_{K_\epsilon}, \mu_{n+1}^N - \mu_n^N P^{\mu_n^N} \rangle| \;\; \leq \;\; \max_{\phi \in \mathbb{F}^\epsilon_{m_{\epsilon,1}}(K_\epsilon)} |\langle \phi, \mu_{n+1}^N - \mu_n^N P^{\mu_n^N} \rangle| + 2\epsilon. \tag{2.4.11}$$

Here $\mathbb{F}^\epsilon_{m_{\epsilon,1}}(K_\epsilon)$ is as given in Lemma A.1.1 and we have denoted the restrictions of μ_{n+1}^N and $\mu_n^N P^{\mu_n}$ to K_ϵ by the same symbols. Using the above inequality in (3.4.29), we obtain

$$E\mathcal{W}_1(\mu_{n+1}^N, \mu_n^N P^{\mu_n^N}) \;\; \leq \;\; \sum_{\phi \in \mathbb{F}^\epsilon_{m_{\epsilon,1}}(K_\epsilon)} E|\langle \phi, \mu_{n+1}^N - \mu_n^N P^{\mu_n^N} \rangle| + 3\epsilon.$$

Using Lemma A.1.2 we see that the first term on the right hand side can be bounded by $\frac{2m_\epsilon |\mathbb{F}^\epsilon_{m_{\epsilon,1}}(K_\epsilon)|}{\sqrt{N}}$. Combining this estimate with (3.4.27), (3.4.34) and (2.4.9) we now have

$$E\mathcal{W}_1(\mu_{n+1}^N, \mu_{n+1}) \;\; \leq \;\; (\|A\| + 2\delta\sigma) E\mathcal{W}_1(\mu_n^N, \mu_n) + \frac{2m_\epsilon |\mathbb{F}^\epsilon_{m_{\epsilon,1}}(K_\epsilon)|}{\sqrt{N}} + 3\epsilon. \tag{2.4.12}$$

Sending $N \to \infty$ in (3.4.31) and using induction hypothesis, we have

$$\limsup_{N\to\infty} E\mathcal{W}_1(\mu_{n+1}^N, \mu_{n+1}) \le 3\epsilon.$$

Since $\epsilon > 0$ is arbitrary, the result follows. $\square$

2.4.3 Proof of Proposition 2.3.2

In this section we study the asymptotics of the deterministic dynamical system (2.2.2). We begin with the following estimate.

Lemma 4. *Under Assumptions 1,2 and 3*

$$\mathcal{W}_1(\Psi^n(\mu_0), \Psi^n(\gamma_0)) \le (\|A\| + 2\delta\sigma)^n \, \mathcal{W}_1(\mu_0, \gamma_0)$$

for any choice of $\mu_0, \gamma_0 \in \mathcal{P}_1(\mathbb{R}^d)$.

Proof. Given $\mu_0, \gamma_0 \in \mathcal{P}_1(\mathbb{R}^d)$, let $\mathcal{C}(\mu_0, \gamma_0) = \{\mu \in \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d) \mid \mu_0(\cdot) = \mu(\cdot \times \mathbb{R}^d), \gamma_0(\cdot) = \mu(\mathbb{R}^d \times \cdot)\}$. Fix $\mu \in \mathcal{C}(\mu_0, \gamma_0)$ and let (X_0, Y_0) be $\mathbb{R}^d \times \mathbb{R}^d$ valued random variables with distribution μ. Also, let $\{\epsilon_n\}_{n\ge 1}$ be an iid sequence of random variables with common law θ independent of (X_0, Y_0). Define for $n \ge 0$,

$$\begin{aligned} X_{n+1} &= AX_n + \delta f(X_n, \mu_n, \epsilon_{n+1}) + g(\epsilon_{n+1}), \\ Y_{n+1} &= AY_n + \delta f(Y_n, \gamma_n, \epsilon_{n+1}) + g(\epsilon_{n+1}) \end{aligned}$$

where $\mu_n = \mathcal{L}(X_n)$ and $\gamma_n = \mathcal{L}(Y_n)$. Then clearly $\mu_n = \Psi^n(\mu_0), \gamma_n = \Psi^n(\gamma_0)$. For $n \ge 0$, denote $\beta_n = \mathcal{W}_1(\mu_n, \gamma_n), \alpha_n = E|X_n - Y_n|$. Then

$$\begin{aligned}
\beta_{n+1} &= \sup_{\phi \in \mathrm{Lip}_1} \{|\int \phi d\mu_{n+1} - \int \phi d\gamma_{n+1}|\} \\
&= \sup_{\phi \in \mathrm{Lip}_1} \{|E\phi(X_{n+1}) - E\phi(Y_{n+1})|\} \\
&\leq E|X_{n+1} - Y_{n+1}| = \alpha_{n+1}. \qquad (2.4.13)
\end{aligned}$$

Also,

$$\begin{aligned}
\alpha_{n+1} &\leq \|A\|E|X_n - Y_n| + \delta E|f(X_n, \mu_n, \epsilon_{n+1}) - f(Y_n, \gamma_n, \epsilon_{n+1})| \\
&\leq \|A\|E|X_n - Y_n| + \delta\sigma(E|X_n - Y_n| + \mathcal{W}_1(\mu_n, \gamma_n)) \\
&= (\|A\| + \delta\sigma)\alpha_n + \delta\sigma\beta_n, \\
&\leq (\|A\| + 2\delta\sigma)\alpha_n \qquad (2.4.14)
\end{aligned}$$

where the second inequality in the display follows from Assumptions 1 and 3. Combining (2.4.13) and (2.4.14) we have

$$\begin{aligned}
\beta_{n+1} &\leq (\|A\| + 2\delta\sigma)^{n+1} E|X_0 - Y_0| \\
&= (\|A\| + 2\delta\sigma)^{n+1} \int |x - y| \mu(dxdy).
\end{aligned}$$

We now have, on taking infimum on the right hand side of the above display over all $\mu \in \mathcal{C}(\mu_0, \gamma_0)$, that $\beta_{n+1} \leq (\|A\| + 2\delta\sigma)^{n+1}\beta_0$. The result follows. $\square$

We can now complete the proof of Proposition 2.3.2. Observe that under our assumption on δ, $\chi := \|A\| + 2\delta\sigma \in (0, 1)$. The first part of the proposition now follows from Lemma 4 and Banach's fixed point theorem. Furthermore

$$\mathcal{W}_1(\Psi^n(\mu), \mu_\infty) = \mathcal{W}_1(\Psi^n(\mu), \Psi^n(\mu_\infty)) \leq \chi^n \, \mathcal{W}_1(\mu, \mu_\infty). \tag{2.4.15}$$

Second part of the proposition is now immediate. □

2.4.4 Proof of Theorem 2.3.1

In this section we show that the empirical measure μ_n^N is suitably close to the law μ_n of X_n for large n and N. As an immediate consequence we obtain that if μ_0^N converges to μ_0 then μ_n^N conerges to μ_n, as $N \to \infty$, uniformly in n. We start with the following moment bound.

Lemma 5. *Suppose that Assumptions 3 and 4 hold and suppose that* $\delta \in (0, a(\alpha)^{\frac{1}{1+\alpha}})$ *where* $a(\cdot)$ *is as in* (2.3.5). *Then* $\sup_{N\geq 1} \sup_{n \geq 1} E|X_n^{1,N}|^{1+\alpha} < \infty$.

Proof. Using Assumption 3 and (2.3.2)

$$|X_{n+1}^i| \quad \leq \quad \|A\||X_n^i| + \delta|D(\epsilon_{n+1}^N)| \left(|X_n^i| + |\mu_n^N|\right) + \delta D_1(\epsilon_{n+1}^N) + |g(\epsilon_{n+1}^N)|.$$

Taking expectations on both sides and applying Holder's inequality, we have, from Assumption 4

$$\begin{aligned} E|X_{n+1}^i|^{1+\alpha} \quad &\leq \quad 4^\alpha \|A\|^{(1+\alpha)} E|X_n^i|^{1+\alpha} + 4^\alpha \delta^{(1+\alpha)} \sigma_1 [E|X_n^i|^{1+\alpha} + E|\mu_n^N|^{1+\alpha}] + 4^\alpha \bar{c}_1(\alpha) \\ &\leq \quad 4^\alpha \|A\|^{(1+\alpha)} E|X_n^i|^{1+\alpha} + 4^\alpha \delta^{(1+\alpha)} \sigma_1 [2E|X_n^i|^{1+\alpha}] + 4^\alpha \bar{c}_1(\alpha) \qquad (2.4.16) \end{aligned}$$

where the last line in the display follows from Jensen's inequality: $E|\mu_n^N|^{1+\alpha} =$

$E|\int |x|\mu_n^N(dx)|^{1+\alpha} \leq E|\int |x|^{1+\alpha}\mu_n^N(dx)| = E|X_n^i|^{1+\alpha}$ and $\bar{c}_1(\alpha) = 2^\alpha(c_1(\alpha) + \tilde{c}_1(\alpha))$

where $c_1(\alpha)$, $\tilde{c}_1(\alpha)$ are as in Assumption 4.

Note that under our condition on δ

$$\kappa_1 \equiv 4^\alpha[\|A\|^{1+\alpha} + 2\delta^{(1+\alpha)}\sigma_1] < 1.$$

Thus

$$\sup_{n\geq 1} E|X_n^i|^{1+\alpha} \leq \kappa_1 E|X_0^i|^{1+\alpha} + \frac{\kappa_2}{1-\kappa_1}, \tag{2.4.17}$$

where $\kappa_2 = 4^\alpha \bar{c}_1(\alpha)$. The result follows. $\square$

We now complete the proof of Theorem 2.3.1.

Proof of Theorem 2.3.1. We will use the decomposition

$$\mu_n^N - \mu_n = \sum_{k=1}^n [\Psi^{n-k}(\mu_k^N) - \Psi^{n-k}.\Psi(\mu_{k-1}^N)] + [\Psi^n(\mu_0^N) - \Psi^n(\mu_0)].$$

Similar decompositions have been used in previous works on interacting particle systems (cf. [25; 26]). It then follows using Lemma 4 that with $\chi = (\|A\| + 2\delta\sigma)$, almost surely

$$\begin{aligned} \mathcal{W}_1(\mu_n^N, \mu_n) &\leq \sum_{k=1}^n \mathcal{W}_1(\Psi^{n-k}(\mu_k^N), \Psi^{n-k}(\Psi(\mu_{k-1}^N))) + \mathcal{W}_1(\Psi^n(\mu_0^N), \Psi^n(\mu_0)) \\ &\leq \sum_{k=1}^n \chi^{n-k}\, \mathcal{W}_1(\mu_k^N, \Psi(\mu_{k-1}^N)) + \chi^n \mathcal{W}_1(\mu_0^N, \mu_0). \end{aligned} \tag{2.4.18}$$

Taking expectations,

$$E\mathcal{W}_1(\mu_n^N,\mu_n) \;\; \le \;\; \sum_{k=1}^{n} \chi^{n-k}\, E\mathcal{W}_1(\mu_k^N,\Psi(\mu_{k-1}^N)) + \chi^n\, E\mathcal{W}_1(\mu_0^N,\mu_0).$$

Since $a(\alpha_0)^{1/(1+\alpha_0)} \to a_0$ as $\alpha_0 \to 0$ and $\delta \in (0,a_0)$, we can find $\alpha_0 \in (0,\alpha)$ such that $\delta \in (0, a(\alpha_0)^{1/(1+\alpha_0)})$. From Lemma 5 we then have that $\sup_{N\ge 1}\sup_{n\ge 1} E|X_n^1|^{1+\alpha_0} < \infty$ and consequently the family $\{X_n^i, i=1,...,N, N\ge 1, n\ge 1\}$ is uniformly integrable. Similar to the proof of Corollary 4.7.1 (cf. the argument below (2.4.6)) using (3.4.16) and (3.4.17) it follows that, for some compact $K_\epsilon \subseteq \mathbb{R}^d$

$$\sup_{N\ge 1}\sup_{n\ge 1} E\int_{K_\epsilon^c} |x|\;[\mu_n^N(dx) + \Psi(\mu_{n-1}^N)(dx)] < \epsilon. \tag{2.4.19}$$

Now for every $k \ge 1$

$$\begin{aligned} E\mathcal{W}_1(\mu_k^N,\Psi(\mu_{k-1}^N)) \;\; &= \;\; E \sup_{f\in \mathrm{Lip}_1^0(\mathbb{R}^d)} |\langle f, \mu_k^N - \Psi(\mu_{k-1}^N)\rangle| \\ &\le \;\; E \sup_{f\in \mathrm{Lip}_1^0(\mathbb{R}^d)} |\langle f.1_{K_\epsilon}, \mu_k^N - \Psi(\mu_{k-1}^N)\rangle| + \epsilon, \end{aligned}$$

when $\mathrm{Lip}_1^0(\mathbb{R}^d)$ is as introduced above (3.4.29). Applying Lemmas A.1.1 and A.1.2 as in the proof of Proposition 2.3.1 we now see that

$$E\mathcal{W}_1(\mu_k^N,\Psi(\mu_{k-1}^N)) \le |\mathbb{F}^{\epsilon}_{m_\epsilon,1}(K_\epsilon)|\frac{2m_\epsilon}{\sqrt{N}} + 3\epsilon,$$

where $m_\epsilon = diam(K_\epsilon)$. Thus

$$\begin{aligned} E\mathcal{W}_1(\mu_n^N, \mu_n) &\leq \sum_{k=1}^{n} \chi^{n-k} \{|\mathbb{F}^{\epsilon}_{m_\epsilon,1}(K_\epsilon)| \frac{2m_\epsilon}{\sqrt{N}} + 3\epsilon\} + \chi^n E\mathcal{W}_1(\mu_0^N, \mu_0) \\ &\leq \{|\mathbb{F}^{\epsilon}_{m_\epsilon,1}(K_\epsilon)| \frac{2m_\epsilon}{\sqrt{N}} + 2\epsilon\} \frac{1}{1-\chi} + \chi^n \, E\mathcal{W}_1(\mu_0^N, \mu_0). \end{aligned} \tag{2.4.20}$$

Given $\varepsilon > 0$, choose ϵ sufficiently small and N_0 sufficiently large such that $\forall N \geq N_0$

$$\{|\mathbb{F}^{\epsilon}_{m_\epsilon,1}(K_\epsilon)| \frac{2m_\epsilon}{\sqrt{N}} + 2\epsilon\} \frac{1}{1-\chi} \leq \frac{\varepsilon}{2}.$$

Choose n_0 large enough so that $\forall n \geq n_0$, $2\chi^n \, \|\mu_0\|_1 < \frac{\varepsilon}{2}$. Combining the above estimates we have $\forall N \geq N_0$, and $n \geq n_0$ $E\mathcal{W}_1(\mu_n^N, \mu_n) \leq \varepsilon$. This proves the first part of the theorem.

Second part is immediate from the first part and Proposition 2.3.1. $\square$

Corollary 4.2.1 is an immediate consequence of Theorem 2.3.1.

2.4.5 Proof of Corollary 4.2.1

Note that

$$E\mathcal{W}_1(\mu_n^N, \mu_\infty) \leq E\mathcal{W}_1(\mu_n^N, \mu_n) + \mathcal{W}_1(\mu_n, \mu_\infty).$$

Combining this with (2.4.20) we have

$$E\mathcal{W}_1(\mu_n^N, \mu_\infty) \leq \left(|\mathbb{F}^{\epsilon}_{m_\epsilon,1}(K_\epsilon)| \frac{2m_\epsilon}{\sqrt{N}} + 2\epsilon\right) \frac{1}{1-\chi} + \chi^n \, E\mathcal{W}_1(\mu_0^N, \mu_0) + \mathcal{W}_1(\mu_n, \mu_\infty).$$

The result now follows on using Proposition 2.3.2. □

We now consider invariant measures of the Markov chain $\{X_k(N)\}_{k\in\mathbb{N}_0}$.

2.4.6 Proof of Proposition 2.3.3

For $N \geq 1$ and $n \in \mathbb{N}_0$, define $\Pi_n^N \in \mathcal{P}((\mathbb{R}^d)^N)$ as

$$\langle \phi, \Pi_n^N \rangle \;=\; \frac{1}{n}\sum_{j=1}^{n} E\phi(X_j^{1,N}, ..., X_j^{N,N}), \quad \phi \in BM((\mathbb{R}^d)^N) \tag{2.4.21}$$

where $\{X_j^{i,N}, j \in \mathbb{N}_0, i = 1, .., N\}$ are as defined in (2.1.1).

From Lemma 19 it folows that, for each $N \geq 1$, the sequence $\{\Pi_n^N, n \geq 1\}$ is relatively compact and using Assumption 1 it is easy to see that any limit point Π_∞^N of Π_n^N (as $n \to \infty$) is an invariant measure of the Markov chain $\{X_n(N)\}_{n\geq 0}$ and from Lemma 19 it satisfies $\int_{(\mathbb{R}^d)^N} |x|\, \Pi_\infty^N(dx) < \infty$. Uniqueness of an invariant measure can be proved by the following simple coupling argument (see for example [12]): Suppose Π_∞^N, $\tilde{\Pi}_\infty^N$ are two invariant measures that satisfy $\int_{(\mathbb{R}^d)^N} |x|\, \Pi_\infty^N(dx) < \infty$, $\int_{(\mathbb{R}^d)^N} |x|\, \tilde{\Pi}_\infty^N(dx) < \infty$. Let $X_0(N) = (X_0^i)_{i=1}^N$ and $\tilde{X}_0(N) = (\tilde{X}_0^i)_{i=1}^N$ with probability laws Π_∞^N and $\tilde{\Pi}_\infty^N$ respectively be given on a common probability space on which is also given an i.i.d. array of $\mathbb{R}^m$ valued random variables $\{\epsilon_n^i, i = 1, \ldots, N, n \geq 1\}$ that is independent of $(X_0(N), \tilde{X}_0(N))$ with common probability law θ. Let, for $i = 1, \dots n$, $n \in \mathbb{N}_0$

$$X_{n+1}^i = AX_n^i + f_\delta(X_n^i, \mu_n^N, \epsilon_{n+1}^i),\ \mu_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_n^i}$$
$$\tilde{X}_{n+1}^i = A\tilde{X}_n^i + f_\delta(\tilde{X}_n^i, \tilde{\mu}_n^N, \epsilon_{n+1}^i),\ \tilde{\mu}_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\tilde{X}_n^i},$$

where recall $f_\delta = \delta f + g$. Using the independence of the noise sequence and Assumption 1 we have

$$E|X_{n+1}^i - \tilde{X}_{n+1}^i| \le (\|A\| + \delta\sigma)E|X_n^i - \tilde{X}_n^i| + \delta\sigma\frac{1}{N}\sum_{j=1}^N E|X_n^j - \tilde{X}_n^j|.$$

Letting $\|X_{n+1}(N) - \tilde{X}_{n+1}(N)\| = \sum_{i=1}^N |X_{n+1}^i - \tilde{X}_{n+1}^i|$, we have

$$E\|X_{n+1}(N) - \tilde{X}_{n+1}(N)\| \le (\|A\| + 2\delta\sigma)E\|X_{n+1}(N) - \tilde{X}_{n+1}(N)\|.$$

Since $\delta \in (0, a_0)$, $\|A\|+2\delta\sigma = \eta \in (0, 1)$. Also, since Π_∞^N and $\tilde{\Pi}_\infty^N$ are invariant distributions, for every $n \in \mathbb{N}_0$, $X_{n+1}(N) = (X_{n+1}^i)_{i=1}^N$ is distributed as Π_∞^N and $\tilde{X}_{n+1}(N) = (\tilde{X}_{n+1}^i)_{i=1}^N$ is distributed as $\tilde{\Pi}_\infty^N$. Thus $X_{n+1}(N)$ and $\tilde{X}_{n+1}(N)$ define a coupling of random variables with laws Π_∞^N and $\tilde{\Pi}_\infty^N$ respectively. From (2.1.3) we then have

$$\mathcal{W}_1(\Pi_\infty^N, \tilde{\Pi}_\infty^N) \le E\|X_{n+1}(N) - \tilde{X}_{n+1}(N)\| \to 0, \text{ as } n \to \infty.$$

Thus $\Pi_\infty^N = \tilde{\Pi}_\infty^N$ which proves the uniqueness of an invariant measure with an integrable first moment and that, as $n \to \infty$,

$$\Pi_n^N \to \Pi_\infty^N. \tag{2.4.22}$$

This proves the first part of the proposition.

Define $r_N : (\mathbb{R}^d)^N \to \mathcal{P}(\mathbb{R}^d)$ as

$$r_N(x_1, ... x_N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}, (x_1, \cdots x_N) \in (\mathbb{R}^d)^N.$$

Let $\nu_n^N = \Pi_n^N \circ r_N^{-1}$ and $\nu_\infty^N = \Pi_\infty^N \circ r_N^{-1}$. In order to prove that Π_∞^N is μ_∞-chaotic, it suffices to argue that (cf. [74])

$$\nu_\infty^N \to \delta_{\mu_\infty} \text{ in } \mathcal{P}(\mathcal{P}(\mathbb{R}^d)), \text{ as } N \to \infty. \tag{2.4.23}$$

We first argue that as $n \to \infty$

$$\nu_n^N \to \nu_\infty^N \qquad \text{in } \mathcal{P}(\mathcal{P}(\mathbb{R}^d)). \tag{2.4.24}$$

It suffices to show that $\langle F, \nu_n^N \rangle \to \langle F, \nu_\infty^N \rangle$ for any continuous and bounded function $F : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$. But this is immediate on observing that

$$\langle F, \nu_n^N \rangle = \langle F \circ r_N, \Pi_n^N \rangle, \ \langle F, \nu_\infty^N \rangle = \langle F \circ r_N, \Pi_\infty^N \rangle,$$

the continuity of the map r_N and the weak convergence of Π_n^N to Π_∞^N. Next, for any $\psi \in$

$BL_1(\mathcal{P}(\mathbb{R}^d))$

$$|\langle \psi, \nu_n^N \rangle - \langle \psi, \delta_{\mu_\infty} \rangle| = |\frac{1}{n}\sum_{j=1}^{n} E\psi(\mu_j^N) - \psi(\mu_\infty)| \leq \frac{1}{n}\sum_{j=1}^{n} E\mathcal{W}_1(\mu_j^N, \mu_\infty).$$

Fix $\epsilon > 0$. For every $N \in \mathbb{N}$ there exists $n_0(N) \in \mathbb{N}$ such that for all $n \geq n_0(N)$

$$E\mathcal{W}_1(\mu_n^N, \mu_\infty) \leq \limsup_{n\to\infty} E\mathcal{W}_1(\mu_n^N, \mu_\infty) + \epsilon.$$

Thus for all $n, N \in \mathbb{N}$

$$|\langle \psi, \nu_n^N \rangle - \langle \psi, \delta_{\mu_\infty} \rangle| \leq \frac{n_0(N)}{n} \max_{1\leq j\leq n_0(N)} E\mathcal{W}_1(\mu_j^N, \mu_\infty) + \limsup_{n\to\infty} E\mathcal{W}_1(\mu_n^N, \mu_\infty) + \epsilon. \tag{2.4.25}$$

Finally

$$\begin{aligned}
\limsup_{N\to\infty} |\langle \psi, \nu_\infty^N \rangle - \langle \psi, \delta_{\mu_\infty} \rangle| =& \limsup_{N\to\infty} \lim_{n\to\infty} |\langle \psi, \nu_n^N \rangle - \langle \psi, \delta_{\mu_\infty} \rangle| \\
\leq& \limsup_{N\to\infty} \limsup_{n\to\infty} E\mathcal{W}_1(\mu_n^N, \mu_\infty) + \epsilon \\
\leq& \epsilon,
\end{aligned}$$

where the first equality is from (3.4.90), the second uses (3.4.91) and the third is a consequence of Corollary 4.2.1. Since $\epsilon > 0$ is arbitrary, we have (3.4.89) and the result follows. $\square$

2.4.7 Proof of Theorem 2.3.2

We will first develop a concentration bound for $\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N))$ for each fixed n and then combine it with the estimate in (2.4.18) in order to obtain the desired result. The first step is carried out in the lemma below, the proof of which is given in Section 2.4.7.

Lemma 6. *Suppose Assumptions 3 and 4 hold. Then, there exist $a_1, a_2, a_3 \in (0, \infty)$ such that for all $\varepsilon, R > 0$ and $n \in \mathbb{N}$,*

$$P[\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N)) > \varepsilon] \;\; \leq \;\; a_3 \left(e^{-a_2 \frac{N\varepsilon^2}{R^2}} + \frac{R^{-\alpha}}{\varepsilon} \right)$$

for all $N \geq \max\{1, a_1(\frac{R}{\varepsilon})^{d+2}\}$.

We now complete the proof of Theorem 2.3.2 using the lemma.

Proof of Theorem 2.3.2

We will make use of (2.4.18). Recall that $\chi = \|A\| + 2\delta\sigma$ and by our assumption $\chi \in (0, 1)$. Let $\gamma = 2\sigma\gamma_0$. Note that $\gamma < 1 - \|A\|$. Then

$$\begin{aligned} P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \;\; &\leq \;\; P[\cup_{i=1}^n \{\chi^{n-i}\mathcal{W}_1(\mu_i^N, \Psi(\mu_{i-1}^N)) > \gamma(1-\gamma)^{n-i}\varepsilon\} \cup \\ &\quad\;\; \{\chi^n \mathcal{W}_1(\mu_0^N, \mu_0) > \gamma(1-\gamma)^n\varepsilon\}] \\ &\leq \;\; \sum_{i=1}^n P[\mathcal{W}_1(\mu_i^N, \Psi(\mu_{i-1}^N)) > \gamma(\frac{1-\gamma}{\chi})^{n-i}\varepsilon] \\ &+ \;\; P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma(\frac{1-\gamma}{\chi})^n\varepsilon]. \end{aligned} \tag{2.4.26}$$

Let $\beta = \gamma\varepsilon$. Note that $\vartheta = \frac{1-\gamma}{\chi}$ and from our choice of δ, $\vartheta > 1$. Therefore

$$N \geq a_1(\frac{R}{\beta})^{d+2} \vee 1 \text{ implies } N \geq a_1(\frac{R}{\beta\vartheta^n})^{d+2} \vee 1 \text{ for all } n \in \mathbb{N}_0.$$

Thus from Lemma 6, for all $N \geq a_1(\frac{R}{\beta})^{d+2} \vee 1$ and $k = 1, \cdots, n$

$$P[\mathcal{W}_1(\mu_k^N, \Psi(\mu_{k-1}^N)) > \beta\vartheta^{n-k}] \leq a_3 \left(e^{-a_2 \frac{N\beta^2\vartheta^{2(n-k)}}{R^2}} + \frac{R^{-\alpha}}{\beta\vartheta^{n-k}} \right).$$

Using the above estimate in (3.4.99)

$$\begin{aligned} P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq & a_3 \sum_{i=0}^{n-1} \left(e^{-a_2 \frac{N\beta^2\vartheta^{2i}}{R^2}} + \frac{R^{-\alpha}}{\beta\vartheta^i} \right) + P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma\varepsilon] \\ \leq & a_3 \sum_{i=0}^{\infty} e^{-a_2 \frac{N\beta^2\vartheta^{2i}}{R^2}} + \frac{a_3 R^{-\alpha}\vartheta}{\beta(\vartheta - 1)} + P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma\varepsilon]. \end{aligned} \tag{2.4.27}$$

Since $\vartheta > 1$ we can find $m_0 = m_0(\vartheta) \in \mathbb{N}$ such that

$$\vartheta^{2i} \geq i\vartheta^2 \qquad \forall i \geq m_0(\vartheta).$$

Thus

$$\begin{aligned} \sum_{i=0}^{\infty} e^{-a_2 \frac{N\beta^2\vartheta^{2i}}{R^2}} &= \sum_{i=1}^{m_0(\vartheta)} e^{-a_2 \frac{N\beta^2\vartheta^{2i}}{R^2}} + \sum_{m_0(\vartheta)+1}^{\infty} e^{-a_2 \frac{N\beta^2\vartheta^{2i}}{R^2}} \\ &\leq m_0(\vartheta) e^{-a_2 \frac{N\beta^2\vartheta^2}{R^2}} + \sum_{i=m_0(\vartheta)+1}^{\infty} (e^{-a_2 \frac{N\beta^2\vartheta^2}{R^2}})^i \\ &\leq [m_0(\vartheta) + \frac{1}{1 - e^{-a_2 \frac{N\beta^2\vartheta^2}{R^2}}}] e^{-a_2 \frac{N\beta^2\vartheta^2}{R^2}}. \end{aligned} \tag{2.4.28}$$

Now for fixed $N \geq 1$ choose $R = \frac{\gamma\varepsilon N^{1/d+2}}{a_1^{1/d+2}}$. Then (2.4.28) holds for all such N, R. Let $N_0 \geq 1$ be large enough so that for all $N \geq N_0$

$$1 - e^{-a_2 a_1^{\frac{2}{d+2}} \vartheta^2 N^{\frac{d}{d+2}}} > 1/2. \tag{2.4.29}$$

Then letting

$$a_4 = a_3(m_0(\vartheta) + 2), \quad a_5 = a_2 a_1^{\frac{2}{d+2}} \vartheta^2, \quad a_6 = \frac{\vartheta}{\vartheta - 1} a_3 a_1^{\frac{\alpha}{d+2}} \gamma^{-(\alpha+1)},$$

we have for all $N \geq N_0$

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq a_4 e^{-a_5 N^{\frac{d}{d+2}}} + a_6 \varepsilon^{-(\alpha+1)} N^{-\frac{\alpha}{d+2}} + P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma\varepsilon].$$

Choose $N_1 \geq N_0$ such that for all $N \geq N_1$, $N^{\frac{d}{d+2}} \geq \frac{2\alpha}{a_5(d+2)} \log N$. Also let $a_7 = \frac{2(1+\alpha)}{a_5}$. Then for all $N \geq \max(N_1, (a_7 \log^+ \varepsilon)^{(d+2)/d})$

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq (a_4 + a_6) \varepsilon^{-(\alpha+1)} N^{-\frac{\alpha}{d+2}} + P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma\varepsilon].$$

The result follows. $\square$

Proof of Lemma 6.

We now complete the proof of Lemma 6. The proof uses certain truncation ideas from [14]. Fix $\varepsilon > 0$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, $R > 0$ and $\nu_0 \in \mathcal{P}(\mathbb{B}_R(0))$, where $\mathbb{B}_R(0) = \{x \in \mathbb{R}^d :$

$|x| \leq R\}$, define $\mu_R \in \mathcal{P}(\mathbb{B}_R(0))$ as

$$\mu_R(A) = \frac{\mu(A)}{\mu(\mathbb{B}_R(0))} 1_{\{\mu(\mathbb{B}_R(0)) \neq 0\}} + \nu_0(A) 1_{\{\mu(\mathbb{B}_R(0)) = 0\}}, \quad A \in \mathcal{B}(\mathbb{B}_R(0)).$$

For $N, n \in \mathbb{N}$ and $R > 0$, let $\Psi^{(R)}(\mu^N_{n-1}) := \frac{1}{N} \sum_{i=1}^N (\delta_{X^{i,N}_{n-1}} P^{\mu^N_{n-1}})_R$.

Let $\{Y^i_n\}_{i=1}^N$ be $\mathbb{B}_R(0)$ valued random variables which, conditionally on $\mathcal{F}^N_{n-1}$ are mutually independent and also independent of $\{X^{i,N}_n\}_{i=1}^N$, and

$$P(Y^i_n \in A \mid \mathcal{F}^N_{n-1}) = (\delta_{X^{i,N}_{n-1}} P^{\mu^N_{n-1}})_R(A), \quad A \in \mathcal{B}(\mathbb{B}_R(0)).$$

Define

$$Z^i_n = \begin{cases} X^{i,N}_n & \text{when } |X^{i,N}_n| \leq R, \\ \\ Y^i_n & \text{otherwise }. \end{cases}$$

It is easily checked that $P(Y^i_n \in A \mid \mathcal{F}^N_{n-1}) = P(Z^i_n \in A \mid \mathcal{F}^N_{n-1})$ for all A and conditionally on $\mathcal{F}^N_{n-1}$, $\{Z^i_n\}_{i=1}^N$ are mutually independent. Define $\mu^N_{n,R} := \frac{1}{N} \sum_{i=1}^N \delta_{Z^i_n}$. Using triangle inequality we have

$$\begin{aligned} \mathcal{W}_1(\mu^N_n, \Psi(\mu^N_{n-1})) \leq \mathcal{W}_1(\Psi^{(R)}(\mu^N_{n-1}), \Psi(\mu^N_{n-1})) + \mathcal{W}_1(\Psi^{(R)}(\mu^N_{n-1}), \mu^N_{n,R}) \\ + \mathcal{W}_1(\mu^N_n, \mu^N_{n,R}). + \mathcal{W}_1(\mu^N_n, \mu^N_{n,R}). \end{aligned} \tag{2.4.30}$$

Consider first the middle term on the right side of (2.4.30). Recall $\mathrm{Lip}^0_1(\mathbb{B}_R(0)) = \{\psi \in$

$\mathrm{Lip}_1(\mathbb{B}_R(0)) : \psi(0) = 0\}$. Then

$$\begin{aligned}
\mathcal{W}_1(\Psi^{(R)}(\mu_{n-1}^N), \mu_{n,R}^N) &= \sup_{\psi \in \mathrm{Lip}_1^0(\mathbb{B}_R(0))} |\langle \psi, \frac{1}{N}\sum_{i=1}^N (\delta_{X_{n-1}^{i,N}} P^{\mu_{n-1}^N})_R - \frac{1}{N}\sum_{i=1}^N \delta_{Z_n^i}\rangle| \\
&= \sup_{\psi \in \mathrm{Lip}_1^0(\mathbb{B}_R(0))} \left| \frac{1}{N}\sum_{i=1}^N \left(\psi(Z_n^i) - \langle \psi, (\delta_{X_{n-1}^{i,N}} P^{\mu_{n-1}^N})_R \rangle \right) \right| \\
&= \sup_{\psi \in \mathrm{Lip}_1^0(\mathbb{B}_R(0))} |\frac{1}{N}\sum_{i=1}^N Z_{i,n}^\psi|
\end{aligned}$$

where $Z_{i,n}^\psi = \psi(Z_n^i) - \langle \psi, (\delta_{X_{n-1}^{i,N}} P^{\mu_{n-1}^N})_R \rangle$. From Lemma A.1.1(a) there exists a finite subset $\mathbb{F}_{R,1}^{\frac{\varepsilon}{4}}(\mathbb{B}_R(0))$ of $\mathrm{Lip}_1^0(\mathbb{B}_R(0))$ such that

$$\sup_{\psi \in \mathrm{Lip}_1^0(\mathbb{B}_R(0))} |\frac{1}{N}\sum_{i=1}^N Z_{i,n}^\psi| \le \max_{\psi \in \mathbb{F}_{R,1}^{\frac{\varepsilon}{4}}(\mathbb{B}_R(0))} |\frac{1}{N}\sum_{i=1}^N Z_{i,n}^\psi| + \frac{\varepsilon}{2}. \tag{2.4.31}$$

Thus

$$\begin{aligned}
P[\mathcal{W}_1(\Psi^{(R)}(\mu_{n-1}^N), \mu_{n,R}^N) > \varepsilon] &\le E\left[P\left[\max_{\psi \in \mathbb{F}_{R,1}^{\frac{\varepsilon}{4}}(\mathbb{B}_R(0))} |\frac{1}{N}\sum_{i=1}^N Z_{i,n}^\psi| > \frac{\varepsilon}{2} \mid \mathcal{F}_{n-1}^N \right] \right] \\
&\le E \sum_{\psi \in \mathbb{F}_{R,1}^{\frac{\varepsilon}{4}}(\mathbb{B}_R(0))} P\left[|\frac{1}{N}\sum_{i=1}^N Z_{i,n}^\psi| > \frac{\varepsilon}{2} \mid \mathcal{F}_{n-1}^N \right]
\end{aligned} \tag{2.4.32}$$

Since $\psi \in \mathrm{Lip}_1^0(\mathbb{B}_R(0))$, $|Z_{i,n}^\psi| \le 2R$. So by the Azuma - Hoeffding inequality the upperbound of $P[\mathcal{W}_1(\Psi^{(R)}(\mu_{n-1}^N), \mu_{n,R}^N) > \epsilon]$ is

$$\begin{aligned}
&\leq |\mathbb{F}_{R,1}^{\frac{\varepsilon}{4}}(\mathbb{B}_R(0))| \max_{\psi \in \mathbb{F}_{R,1}^{\frac{\varepsilon}{2}}(\mathbb{B}_R(0))} E\left(P[|\frac{1}{N}\sum_{i=1}^N Z_{i,n}^\psi| > \frac{\varepsilon}{2} \,|\, \mathcal{F}_{n-1}^N]\right) \\
&\leq |\mathbb{F}_{R,1}^{\frac{\varepsilon}{4}}(\mathbb{B}_R(0))| 2e^{-\frac{N\varepsilon^2}{32R^2}} \\
&\leq 2\,|\mathbb{F}_{R,1}^{\frac{\varepsilon}{4}}([-R,R]^d)| e^{-\frac{N\varepsilon^2}{32R^2}}. \qquad (2.4.33)
\end{aligned}$$

From Lemma A.1.1(b) we have the following estimate

$$P[\mathcal{W}_1(\Psi^{(R)}(\mu_{n-1}^N), \mu_{n,R}^N) > \varepsilon] \leq \max\left\{2, \frac{16R}{3\varepsilon}(2\sqrt{d}+1)3^{[\frac{8R}{\varepsilon}(\sqrt{d}+1)]^d}\right\} e^{-\frac{N\varepsilon^2}{32R^2}}.$$

Thus there exist $k_1, k_2 \in (0,\infty)$ such that for all $n, N \in \mathbb{N}$, $R > 0, \varepsilon > 0$

$$P[\mathcal{W}_1(\Psi^{(R)}(\mu_{n-1}^N), \mu_{n,R}^N) > \varepsilon] \leq k_2[e^{k_1(R/\varepsilon)^d} \vee 1] e^{-\frac{N\varepsilon^2}{32R^2}}. \qquad (2.4.34)$$

For the first term in the right hand side of (2.4.30) we make use of the observation that if for $i = 1, \cdots N$, U^i, V^i are $\mathbb{R}^d$ valued random variables distributed according to λ_U^i, λ_V^i respectively then

$$\mathcal{W}_1(\frac{1}{N}\sum_{i=1}^N \lambda_U^i, \frac{1}{N}\sum_{i=1}^N \lambda_V^i) \leq \frac{1}{N}\sum_{i=1}^N E|U^i - V^i|.$$

Thus

$$\begin{aligned}\mathcal{W}_1(\Psi^{(R)}(\mu_{n-1}^N), \Psi(\mu_{n-1}^N)) &= \mathcal{W}_1(\frac{1}{N}\sum_{i=1}^N (\delta_{X_{n-1}^{i,N}} P^{\mu_{n-1}^N})_R, \frac{1}{N}\sum_{i=1}^N \delta_{X_{n-1}^{i,N}} P^{\mu_{n-1}^N}) \\ &\leq \frac{1}{N}\sum_{i=1}^N E[|X_n^{i,N} - Z_n^i| \mid \mathcal{F}_{n-1}^N].\end{aligned}$$

Using the definition of $\{Z_n^i\}$ we see

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N E[|X_n^{i,N} - Z_n^i| \mid \mathcal{F}_{n-1}^N] &= \frac{1}{N}\sum_{i=1}^N E[|X_n^{i,N} - Y_n^i| 1_{|X_n^{i,N}|>R} \mid \mathcal{F}_{n-1}^N] \\ &\leq \frac{2}{N}\sum_{i=1}^N E[|X_n^{i,N}| 1_{|X_n^{i,N}|>R} \mid \mathcal{F}_{n-1}^N] \qquad (2.4.35)\end{aligned}$$

From (2.4.17) we have that $B_n^\alpha := E|X_n^{i,N}|^{1+\alpha}$ satisfies

$$B_n^\alpha \leq \kappa_1 E|X_0^{i,N}|^{1+\alpha} + \frac{\kappa_2}{1-\kappa_1} = B(\alpha).$$

Thus

$$\begin{aligned}P[\mathcal{W}_1(\Psi^{(R)}(\mu_{n-1}^N), \Psi(\mu_{n-1}^N)) > \varepsilon] &\leq \frac{1}{\varepsilon} E \frac{1}{N}\sum_{i=1}^N E[|X_n^{i,N} - Y_n^i| 1_{|X_n^{i,N}|>R} \mid \mathcal{F}_{n-1}^N] \\ &\leq \frac{2}{\varepsilon} E\{|X_n^{i,N}| 1_{|X_n^{i,N}|>R}\} \\ &\leq \frac{2R^{-\alpha}}{\varepsilon} B_n^\alpha \leq \frac{2R^{-\alpha}}{\varepsilon} B(\alpha). \qquad (2.4.36)\end{aligned}$$

The third term in (2.4.30) can be treated similarly. Indeed, note that

$$\mathcal{W}_1(\mu_n^N, \mu_{n,R}^N) \leq \frac{1}{N}\sum_{i=1}^N |X_n^{i,N} - Z_n^i| = \frac{1}{N}\sum_{i=1}^N |X_n^{i,N} - Y_i^n| 1_{|X_n^{i,N}|>R}.$$

Thus using the bound for the right side of the first line in (2.4.36) we have that

$$\begin{aligned} P(\mathcal{W}_1(\mu_n^N, \mu_{n,R}^N) > \varepsilon) \leq & \frac{1}{\varepsilon} E \frac{1}{N}\sum_{i=1}^N E[|X_n^{i,N} - Y_i^n| 1_{|X_n^{i,N}|>R} \mid \mathcal{F}_{n-1}^N] \\ \leq & \frac{2R^{-\alpha}}{\varepsilon} B(\alpha). \end{aligned} \tag{2.4.37}$$

Using (2.4.34), (2.4.36) and (2.4.37) in (2.4.30) we have

$$P[\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N)) > \varepsilon] \quad \leq \quad k_2[e^{k_1(3R/\varepsilon)^d} \vee 1] e^{-\frac{N\varepsilon^2}{288R^2}} + \frac{12R^{-\alpha}}{\varepsilon} B(\alpha).$$

Letting $k_3 = 3^d \cdot 576 k_1$, $k_4 = 1/576$ and $k_5 = \max\{k_2, 12B(\alpha)\}$, we have that

$$P[\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N)) > \varepsilon] \leq k_5 \left(e^{-\frac{k_4 N \varepsilon^2}{R^2}} + \frac{R^{-\alpha}}{\varepsilon} \right)$$

for all $N \geq \max\{1, k_3(\frac{R}{\varepsilon})^{d+2}\}$. This completes the proof of the lemma. □

2.4.8 Proof of Theorem 2.3.3

We will proceed as in Section 2.4.7 by first first giving a concentration bound for

$$\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N))$$

for each fixed n and then combining it with (2.4.18) in order to obtain a uniform in n estimate. We begin by observing that from Assumption 5 it follows that there is a $\alpha_0 \in (0, \alpha]$ and $c_2 \in (0, \infty)$ such that for all $\alpha_1 \in [0, \alpha_0]$

$$\mathcal{E}_1(\alpha_1) := \int e^{\alpha_1(D_1(z)+|g_1(z)|)}\theta(dz) \le e^{c_2\alpha_1} \tag{2.4.38}$$

Lemma 7. *Suppose Assumptions 3 and 5 hold. Let γ_0 be as in Theorem 2.3.3. Then for all $\delta \in [0, \min\{a_0 - \gamma_0, \frac{1-\|A\|}{2M}\})$ and $\alpha_1 \in [0, \alpha_0]$*

$$\sup_{n\ge 0}\sup_{N\ge 1} Ee^{\alpha_1|X_n^{1,N}|} < \infty.$$

Proof. Note that for $n \ge 1$

$$|X_n^{i,N}| \le \|A\||X_{n-1}^{i,N}| + \delta M \left(|X_{n-1}^{i,N}| + \|\mu_{n-1}^N\|_1\right) + (\delta D_1(\varepsilon_n^i) + |g_1(\varepsilon_n^i)|).$$

Taking expectations, for all $\alpha_1 \in [0, \alpha_0]$

$$\begin{aligned} Ee^{\alpha_1|X_n^{i,N}|} &\le E\exp\left\{\alpha_1\left(\|A\||X_{n-1}^{i,N}| + \delta M\left(|X_{n-1}^{i,N}| + \|\mu_{n-1}^N\|_1\right) + (\delta D_1(\varepsilon_n^i) + |g_1(\varepsilon_n^i)|)\right)\right\} \\ &\le \mathcal{E}_1(\alpha_1)E\exp\left\{\alpha_1\left(\|A\||X_{n-1}^{i,N}| + \delta M\left(|X_{n-1}^{i,N}| + \|\mu_{n-1}^N\|_1\right)\right)\right\} \end{aligned}$$

Holder's inequality with $p = (\|A\| + 2\delta M)/(\|A\| + \delta M)$ and $q = (\|A\| + 2\delta M)/\delta M$ gives

$$
\begin{aligned}
&E \exp \left\{ \alpha_1 \left(\|A\| |X_{n-1}^{i,N}| + \delta M \left(|X_{n-1}^{i,N}| + \|\mu_{n-1}^N\|_1 \right) \right) \right\} \\
&\quad \leq \left(E \exp \left\{ \alpha_1 (\|A\| + 2\delta M) |X_{n-1}^{i,N}| \right\} \right)^{1/p} \left(E \exp \left\{ \alpha_1 (\|A\| + 2\delta M) \|\mu_{n-1}^N\|_1 \right\} \right)^{1/q}.
\end{aligned}
$$

From Jensen's inequality and exchangeability

$$
\begin{aligned}
E \exp \left\{ \alpha_1 (\|A\| + 2\delta M) \|\mu_{n-1}^N\|_1 \right\} &\leq \frac{1}{N} \sum_{i=1}^{N} E \exp \left\{ \alpha_1 (\|A\| + 2\delta M) |X_{n-1}^{i,N}| \right\} \\
&= E \exp \left\{ \alpha_1 (\|A\| + 2\delta M) |X_{n-1}^{i,N}| \right\}.
\end{aligned}
$$

Using this inequality in the previous display and recalling $1/p + 1/q = 1$ we have

$$
E e^{\alpha_1 |X_n^{i,N}|} \leq \mathcal{E}_1(\alpha_1) E \exp \left\{ \alpha_1 (\|A\| + 2\delta M) |X_{n-1}^{i,N}| \right\}.
$$

Thus for all $\alpha_1 \in [0, \alpha_0]$

$$
f_n(\alpha_1) := E \exp\{\alpha_1 |X_n^{i,N}|\} \leq \mathcal{E}_1(\alpha_1) f_{n-1}(\alpha_1 \kappa_1),
$$

where by our assumption $\kappa_1 = \|A\| + 2\delta M \in (0, 1)$. Iterating the above inequality we have for all $n \geq 1$

$$
f_n(\alpha_1) \leq f_0(\alpha_1) \prod_{j=0}^{n-1} \mathcal{E}_1(\alpha_1 \kappa_1^j) \leq f_0(\alpha_1) e^{c_2 \alpha_1 \sum_{j=0}^{n-1} \kappa_1^j} \leq f_0(\alpha_1) e^{c_2 \alpha_1 / (1 - \kappa_1)}
$$

where the second inequality is a consequence of (2.4.38). The result follows. $\square$

The following lemma is proved in a manner similar to Lemma 6 so only a sketch is provided.

Lemma 8. *There exist* $\tilde{a}_1, \tilde{a}_2, \tilde{a}_3 \in (0, \infty)$ *and, for each* $\alpha_1 \in [0, \alpha_0)$, $\tilde{B}(\alpha_1) \in [0, \infty)$ *such that for all* $\varepsilon, R > 0$ *and* $n \in \mathbb{N}$,

$$P[\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N)) > \varepsilon] \leq \tilde{a}_3 \left(e^{-\tilde{a}_2 \frac{N\varepsilon^2}{R^2}} + \tilde{B}(\alpha_1) \frac{e^{-\alpha_1 R}}{\varepsilon} \right)$$

for all $N \geq \max\{1, \tilde{a}_1 (\frac{R}{\varepsilon})^{d+2}\}$.

Proof. From Lemma 7 we have that for $\alpha_1 \in [0, \alpha_0]$

$$\sup_{n \geq 0} \sup_{N \geq 1} \max_{\{1 \leq i \leq N\}} E e^{\alpha |X_n^{i,N}|} < \infty. \tag{2.4.39}$$

Next, as in the proof of Lemma 6, we will use (2.4.30). For the middle term on the right side of (2.4.30) we use the same bound as in (2.4.34). Now consider the first term in (2.4.30). From (2.4.35) we have that

$$P[\mathcal{W}_1(\Psi(\mu_{n-1}^N), \Psi(\mu_{n-1}^N)_R) > \varepsilon] \leq \frac{2}{\varepsilon} E \left(|X_n^{1,N}| 1_{|X_n^{1,N}| > R} \right). \tag{2.4.40}$$

From (2.4.39) it follows that for every $\alpha_1 \in [0, \alpha_0)$

$$\sup_{n \geq 0} \sup_{N \geq 1} \max_{\{1 \leq i \leq N\}} E \left(|X_n^{i,N}| e^{\alpha_1 |X_n^{i,N}|} \right) = \tilde{B}(\alpha_1) < \infty.$$

Applying Markov's inequality we now have for $\alpha_1 \in [0, \alpha_0)$

$$P[\mathcal{W}_1(\Psi(\mu_{n-1}^N), \Psi(\mu_{n-1}^N)_R) > \varepsilon] \;\; \leq \;\; \frac{2}{\varepsilon} e^{-\alpha_1 R} \tilde{B}(\alpha_1). \tag{2.4.41}$$

The third term in (2.4.30) is bounded similarly. Indeed, as in (2.4.37) we get for $\alpha_1 \in [0, \alpha)$

$$P[\mathcal{W}_1(\mu_n^N, \mu_{n,R}^N) > \varepsilon] \leq \frac{2}{\varepsilon} e^{-\alpha_1 R} \tilde{B}(\alpha_1). \tag{2.4.42}$$

Using (2.4.34), (2.4.41) and (2.4.42) in (2.4.30) we now have for $\alpha_1 \in [0, \alpha_0)$

$$P[\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N)) > \varepsilon] \;\; \leq \;\; k_2[e^{k_1(3R/\varepsilon)^d} \vee 1] e^{-\frac{N\varepsilon^2}{288R^2}} + \frac{12e^{-\alpha_1 R}}{\varepsilon} \tilde{B}(\alpha_1).$$

Thus with k_3, k_4 as in the proof of Lemma 6 and $k_5 = \max\{k_2, 12\}$ we have

$$P[\mathcal{W}_1(\mu_n^N, \Psi(\mu_{n-1}^N)) > \varepsilon] \leq k_5 \left(e^{-\frac{k_4 N\varepsilon^2}{R^2}} + \tilde{B}(\alpha_1) \frac{e^{-\alpha_1 R}}{\varepsilon} \right)$$

for all $N \geq \max\{1, k_3(\frac{R}{\varepsilon})^{d+2}\}$. The result follows. $\square$

We now complete the proof of Theorem 2.3.3.

Proof of Theorem 2.3.3.

Fix $\alpha_1 \in [0, \alpha_0)$. Following the steps in the proof of (2.4.27), with ϑ, β as in Theorem 2.3.2, we have from Lemma 8, for all $N \geq \tilde{a}_1(\frac{R}{\beta})^{d+2} \vee 1$ and $k = 1, \cdots, n$

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq \tilde{a}_3 \sum_{i=0}^{\infty} e^{-\tilde{a}_2 \frac{N\beta^2\vartheta^{2i}}{R^2}} + \frac{\tilde{a}_3 \tilde{B}(\alpha_1) e^{-\alpha_1 R}\vartheta}{\beta(\vartheta - 1)} + P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma\vartheta^n\varepsilon]. \tag{2.4.43}$$

As before for fixed $N \geq 1$ choose $R = \frac{\gamma\varepsilon N^{1/d+2}}{\tilde{a}_1^{1/d+2}}$. Then (2.4.28) holds for all such N, R with a_2 replaced by $\tilde{a}_2$. Let $N_0 \geq 1$ be large enough so that for all $N \geq N_0$, (2.4.29) holds with (a_1, a_2) replaced by $(\tilde{a}_1, \tilde{a}_2)$. Then letting

$$\tilde{a}_4 = \tilde{a}_3(m_0(\vartheta) + 2), \quad \tilde{a}_5 = \tilde{a}_2 \tilde{a}_1^{\frac{2}{d+2}} \vartheta^2, \quad \tilde{a}_6 = \frac{\tilde{a}_3 \vartheta}{\gamma(\vartheta - 1)} \tilde{B}(\alpha_1), \quad \tilde{a}_7 = \frac{\alpha_1 \gamma}{\tilde{a}_1^{1/d+2}},$$

we have for all $N \geq N_0$

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq \tilde{a}_4 e^{-\tilde{a}_5 N^{\frac{d}{d+2}}} + \tilde{a}_6 \varepsilon^{-1} \exp(-\tilde{a}_7 \varepsilon N^{\frac{1}{d+2}}) + P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma\vartheta^n\varepsilon].$$

Note that $\varepsilon^{-1} \exp(-\tilde{a}_7 \varepsilon N^{\frac{1}{d+2}}) < \exp(-\frac{\tilde{a}_7}{2} \varepsilon N^{\frac{1}{d+2}})$ if $N > \left(\frac{2}{\tilde{a}_7}\right)^{d+2} (\frac{1}{\varepsilon} \log^+ \frac{1}{\varepsilon})^{d+2}$.

Consider now the case $d > 1$. Then, taking $L_1 = \max\{(\frac{2}{\tilde{a}_7})^{d+2}, N_0\}$, $L_2 = \tilde{a}_4 + \tilde{a}_6$, $L_3 = \min\{\tilde{a}_5, \tilde{a}_7/2\}$, we have for all $N \geq L_1 \max\{(\frac{1}{\varepsilon} \log^+ \frac{1}{\varepsilon})^{d+2}, \varepsilon^{(d+2)/(d-1)}\}$

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq L_2 e^{-L_3 \varepsilon N^{1/d+2}} + P[\mathcal{W}_1(\mu_0^N, \mu_0) > \gamma\vartheta^n\varepsilon].$$

This proves the theorem for the case $d > 1$. Finally for $d = 1$, with the same choice of L_1, L_2, L_3, we have for all $N \geq L_1 \max\{(\frac{1}{\varepsilon} \log^+ \frac{1}{\varepsilon})^{d+2}, 1\}$

$$P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq L_2 e^{-L_3(\varepsilon \wedge 1) N^{1/d+2}}.$$

The result follows. □

2.4.9 Proof of Theorem 2.3.4.

In order to prove the theorem we will introduce an auxiliary sequence $\{Y_n^{i,N}, i = 1, \cdots N\}_{n \geq 0}$ such that for each n, $\{Y_n^{i,N}\}_{i=1}^N$ are i.i.d. We will then employ results from [11] and [13] in order to give a uniform (in k) concentration bound for $\mathcal{W}_1(\eta_k^N, \mu_k)$, where η_k^N is the empirical measure $\frac{1}{N} \sum_{k=1}^N \delta_{Y_k^{i,N}}$. Finally we will obtain the desired concentration estimate on $\mathcal{W}_1(\mu_k^N, \mu_k)$ by making use of Lemma 9 below. We begin by introducing our auxiliary system.

An Auxiliary System.

Consider the collection of $\mathbb{R}^d$ valued random variables $\{Y_n^{i,N}, i = 1, ..., N\}_{n \geq 0}$ defined as follows.

$$\begin{aligned} Y_{n+1}^{i,N} &= AY_n^{i,N} + f_\delta(Y_n^{i,N}, \mu_n, \epsilon_{n+1}^i), \quad n \geq 0 \\ Y_0^{i,N} &= X_0^{i,N}. \end{aligned} \tag{2.4.44}$$

Note that for each n, $\{Y_n^{i,N}\}_{i=1}^N$ are i.i.d. In fact, since $\mathcal{L}(\{X_0^{i,N}\}_{i=1,...N}) = \mu_0^{\otimes N}$, we have $\mathcal{L}(\{Y_n^{i,N}\}_{i=1,...N}) = \mu_n^{\otimes N}$. Let $\eta_n^N := \frac{1}{N} \sum_{i=1}^N \delta_{Y_n^{i,N}}$. The following lemma will give a useful relation between $\mathcal{W}_1(\eta_n^N, \mu_n)$ and $\mathcal{W}_1(\mu_n^N, \mu_n)$.

Lemma 9. *Suppose Assumptions 3 and 5 hold. Let* $\chi_2 = \|A\| + 2\delta M$. *Then for every* $n \geq 0$ *and* $N \geq 1$

$$\mathcal{W}_1(\mu_{n+1}^N, \mu_{n+1}) \leq \mathcal{W}_1(\eta_{n+1}^N, \mu_{n+1}) + \delta M \sum_{k=0}^{n} \chi_2^{n-k} \mathcal{W}_1(\eta_k^N, \mu_k). \tag{2.4.45}$$

Proof. Since by Assumption 5 $D(\epsilon) \leq M$, we have for each $i = 1, ..., N$

$$\begin{aligned} |X_{n+1}^{i,N} - Y_{n+1}^{i,N}| &\leq \|A\||X_n^{i,N} - Y_n^{i,N}| + \delta M\{|X_n^{i,N} - Y_n^{i,N}| + \mathcal{W}_1(\mu_n^N, \mu_n)\} \\ &= (\|A\| + \delta M)|X_n^{i,N} - Y_n^{i,N}| + \delta M \mathcal{W}_1(\mu_n^N, \mu_n) \end{aligned}$$

Thus

$$|X_{n+1}^{i,N} - Y_{n+1}^{i,N}| \leq \delta M \sum_{k=0}^{n} (\|A\| + \delta M)^{n-k} \mathcal{W}_1(\mu_k^N, \mu_k). \tag{2.4.46}$$

Now note that

$$\mathcal{W}_1(\eta_{n+1}^N, \mu_{n+1}^N) \leq \frac{1}{N} \sum_{i=1}^{N} |X_{n+1}^{i,N} - Y_{n+1}^{i,N}| \leq \delta M \sum_{k=0}^{n} (\|A\| + \delta M)^{n-k} \mathcal{W}_1(\mu_k^N, \mu_k).$$

Using triangle inequality

$$\begin{aligned} \mathcal{W}_1(\eta_{n+1}^N, \mu_{n+1}^N) \leq \delta M \sum_{k=0}^{n} (\|A\| + \delta M)^{n-k} \mathcal{W}_1(\eta_k^N, \mu_k^N) + \\ \delta M \sum_{k=0}^{n} (\|A\| + \delta M)^{n-k} \mathcal{W}_1(\eta_k^N, \mu_k). \end{aligned}$$

Applying Lemma A.1.3 with

$$a_n = \chi_1^{-n}\mathcal{W}_1(\eta_n^N, \mu_n^N),\ b_n = \frac{\delta M}{\chi_1}\sum_{k=0}^{n-1}\chi_1^{-k}\mathcal{W}_1(\eta_k^N, \mu_k),\ c_n = \frac{\delta M}{\chi_1},\ n \geq 0$$

where $\chi_1 := \|A\| + \delta M$, we have

$$\begin{aligned}
\chi_1^{-(n+1)}\mathcal{W}_1(\eta_{n+1}^N, \mu_{n+1}^N) &\leq b_{n+1} + \sum_{k=0}^{n}(\frac{\delta M}{\chi_1})^2\sum_{i=0}^{k-1}\chi_1^{-i}\mathcal{W}_1(\eta_i^N, \mu_i)\left(1 + \frac{\delta M}{\chi_1}\right)^{n-k} \\
&= b_{n+1} + \sum_{i=0}^{n}\sum_{k=i+1}^{n}(\frac{\delta M}{\chi_1})^2(1 + \frac{\delta M}{\chi_1})^{n-k}\chi_1^{-i}\mathcal{W}_1(\eta_i^N, \mu_i) \\
&= b_{n+1} + \sum_{i=0}^{n}(\frac{\delta M}{\chi_1})^2\chi_1^{-i}.\mathcal{W}_1(\eta_i^N, \mu_i)\sum_{m=0}^{n-i-1}(1 + \frac{\delta M}{\chi_1})^m \\
&= b_{n+1} + \sum_{i=0}^{n}(\frac{\delta M}{\chi_1})\chi_1^{-i}\mathcal{W}_1(\eta_i^N, \mu_i)[(1 + \frac{\delta M}{\chi_1})^{n-i} - 1] \qquad (2.4.47)
\end{aligned}$$

Simplifying (3.4.108) one gets

$$\begin{aligned}
\mathcal{W}_1(\eta_{n+1}^N, \mu_{n+1}^N) &\leq \delta M\sum_{k=0}^{n}\chi_1^{n-k}\mathcal{W}_1(\eta_k^N, \mu_k) + \sum_{k=0}^{n}\delta M\chi_1^{n-k}\mathcal{W}_1(\eta_k^N, \mu_k)[(1 + \frac{\delta M}{\chi_1})^{n-k} - 1] \\
&= \delta M\sum_{k=0}^{n}(\chi_1 + \delta M)^{n-k}\mathcal{W}_1(\eta_k^N, \mu_k).
\end{aligned}$$

The result now follows by an application of triangle inequality. $\square$

Proof of Theorem 2.3.4 is based on certain results from [11] and [13] which we summarize in this section. Define $\ell : [0, \infty) \to [0, \infty)$ as $\ell(x) = x\log x - x + 1$. With the definition of "ν satisfies Transportation inequality $\alpha(\mathcal{T})$ " defined in (1.0.1), the following result is established in [11].

Theorem 2.4.1. *(Boissard [11]) Suppose that $\nu \in \mathcal{P}(\mathbb{R}^d)$ satisfies a $\alpha(\mathcal{T})$ inequality and suppose that there is $\zeta > 0$ such that $\int_{\mathbb{R}^d} e^{\zeta|x|}\nu(dx) \leq 2$. Let $L_N = \frac{1}{N}\sum_{i=1}^N \delta_{Z_i}$ where Z_i are i.i.d. with common distribution ν. Then for $t > 0$*

$$P(\mathcal{W}_1(L_N, \nu) \geq t) \leq \exp\left\{-N\alpha(\frac{t}{2} - \Gamma(\mathcal{C}_t, N))\right\},$$

where

$$\Gamma(\mathcal{C}_t, N)) = \inf_{\lambda > 0}\left\{\frac{1}{\lambda}\log \mathcal{C}_t + N\alpha^*(\frac{\lambda}{N})\right\}, \tag{2.4.48}$$

$\alpha^* : \mathbb{R} \to [0, \infty)$ is defined as

$$\alpha^*(s) = \sup_{t \geq 0}\{st - \alpha(t)\}1_{[0,\infty)}(s), \ s \in \mathbb{R},$$

$$\mathcal{C}_t = 2\left(1 + \psi(\frac{32}{\zeta t})\right) 2^{c_d(\psi(\frac{32}{\zeta t}))^d}, \tag{2.4.49}$$

$\psi(x) = x\log(2\ell(x))$, $x \geq 0$ and c_d is a positive scaler depending only on d.

The following result is from [13].

Theorem 2.4.2. *Let $\nu \in \mathcal{P}(\mathbb{R}^d)$. Suppose that $\int_{\mathbb{R}^d} e^{\alpha_0|x|}d\nu(x) < \infty$ for some $\alpha_0 > 0$. Then ν satisfies $\alpha(\mathcal{T})$ inequality with*

$$\alpha(t) = \left(\sqrt{\frac{t}{C} + \frac{1}{4}} - \frac{1}{2}\right)^2, \ t \geq 0 \tag{2.4.50}$$

for any

$$C > 2 \inf_{x_0 \in \mathbb{R}^d, \tilde{\alpha} > 0} \frac{1}{\tilde{\alpha}} \left(\frac{3}{2} + \log \int_{\mathbb{R}^d} e^{\tilde{\alpha}|x - x_0|} d\nu(x) \right).$$

Exponential Integrability.

Transportation inequalities presented in Introduction require exponential integrability of the underlying measure. In this section we show that under Assumption 5 the desired integrability properties hold.

Lemma 10. *Suppose that Assumption 5 holds and that* $\delta \in (0, \frac{1-\|A\|}{2M})$. *Then* $\kappa_1 = (\|A\| + 2\delta M) \in (0,1)$ *and for all* $\alpha_1 \in [0, \alpha_0]$

$$\sup_{n \geq 0} \int_{\mathbb{R}^d} e^{\alpha_1 |x|} \mu_n(dx) \leq (\int e^{\alpha_1 |x|} \mu_0(dx)) \exp\left\{ \frac{c_2 \alpha_1}{1 - \kappa_1} \right\}.$$

Proof. The property that $\kappa_1 \in (0,1)$ is an immediate consequence of assumptions on δ. Let $f_n(\alpha_1) := \int e^{\alpha_1 |x|} \mu_n(dx)$.

From (2.4.1) and the condition $D(\epsilon) \leq M$ we have

$$|X_{n+1}| \leq \|A\| |X_n| + \delta M (|X_n| + \|\mu_n\|_1) + \delta D_1(\epsilon_{n+1}) + |g(\epsilon_{n+1})|. \tag{2.4.51}$$

Using Holder's inequality and taking exponentials we get

$$\begin{aligned} f_{n+1}(\alpha_1) = E e^{\alpha_1 |X_{n+1}|} &\leq E e^{\alpha_1 [\|A\| |X_n| + \delta M (|X_n| + \|\mu_n\|_1) + \delta D_1(\epsilon_{n+1}) + |g(\epsilon_{n+1})|]} \\ &= \mathcal{E}_1(\alpha_1) E e^{\alpha_1 (\|A\| + \delta M) |X_n| + \delta M \|\mu_n\|_1}. \end{aligned} \tag{2.4.52}$$

Applying Jensen's inequality again to the function $x \mapsto \exp\{\alpha_1 \delta M x\}$ and observing $\|\mu_n\|_1 \leq E|X_n|$ we have

$$f_{n+1}(\alpha_1) \quad \leq \quad \mathcal{E}_1(\alpha_1) E[e^{\alpha_1(\|A\|+\delta M)|X_n|}] E[e^{\alpha_1 \delta M |X_n|}]. \tag{2.4.53}$$

Note that for any two non-decreasing, non-negative functions f, g on $\mathbb{R}$ and any $\mu \in \mathcal{P}(\mathbb{R})$,

$$\int f(x)g(x)\mu(dx) \geq \int f(x)\mu(dx) \int g(y)\mu(dy).$$

Using this inequality in the above display yields the following recursion

$$f_{n+1}(\alpha_1) \leq \mathcal{E}_1(\alpha_1) E[e^{\alpha_1(\|A\|+2\delta M)|X_n|}] = \mathcal{E}_1(\alpha_1) f_n(\alpha_1 \kappa_1).$$

Iterating the above inequality we have, for all $n \geq 0$,

$$f_{n+1}(\alpha_1) \leq f_0(\alpha_1) \prod_{j=0}^{n} \mathcal{E}_1(\alpha_1 \kappa_1^j).$$

Thus using (2.4.38) we see

$$f_{n+1}(\alpha_1) \leq f_0(\alpha_1) \prod_{j=0}^{n} \exp\{c_2(\alpha_1 \kappa_1^j)\} \leq f_0(\alpha_1) \exp\{c_2 \alpha_1 \sum_{j=0}^{\infty} \kappa_1^j\}.$$

The result follows. $\square$

2.4.10 Uniform Concentration Bounds for $\{\eta_n^N\}$.

In this section we will give, using results of Theorem 2.4.1 and 2.4.9, uniform concentration bounds for $\{\eta_n^N\}_{n\geq 1}$ as $N \to \infty$.

Lemma 11. *Suppose that Assumption 5 holds and $\delta \in (0, \frac{1-\|A\|}{2M})$. Then the following hold.*

(1) There exists a $\zeta_0 \in (0, \infty)$ such that

$$\sup_{n\in\mathbb{N}_0} \int_{\mathbb{R}^d} e^{\zeta_0|x|}\mu_n(dx) \leq 2 \tag{2.4.54}$$

and for all $n \in \mathbb{N}_0$, μ_n satisfies a $\alpha(\mathcal{T})$ inequality with α as in (2.4.50) *and with*

$$C \geq C_0 = 2\sqrt{2}\frac{1}{\zeta_0}\left(\frac{3}{2} + \log 2\right).$$

(2) For all $t > 0$ and $n \in \mathbb{N}_0$

$$P(\mathcal{W}_1(\eta_n^N, \mu_n) \geq t) \leq \exp\left\{-N\alpha_0(\frac{t}{2} - \Gamma_0(\mathcal{C}_t^0, N))\right\},$$

where α_0 is defined by the right side in (2.4.50) *with C replaced with C_0, Γ_0 is defined by the right side of* (2.4.48) *with α^* replaced by α_0^* and $\mathcal{C}_t^0$ is as in* (2.4.49) *with ζ replaced with ζ_0.*

(3) There exist $N_1 \in \mathbb{N}$ *and* $L_1 \in (0,\infty)$ *such that for all* $t \in [\frac{C_0}{2}, \infty)$*,* $n \in \mathbb{N}_0$ *and* $N \geq N_1$

$$P(\mathcal{W}_1(\eta_n^N, \mu_n) \geq t) \leq \exp(-L_1 N t).$$

(4) There exist $L_2, L_3 \in (0,\infty)$ *such that for all* $t \in (0, \frac{C_0}{2}]$ *and all* $N \geq L_3 \frac{\log \mathcal{C}_t}{t^2}$.

$$P(\mathcal{W}_1(\eta_n^N, \mu_n) \geq t) \leq \exp(-L_2 N t^2).$$

Proof. (1) Suppose that the statement in (2.4.54) fails to hold for any $\zeta_0 > 0$. Then there exist sequences $n_k \uparrow \infty$ and $\zeta_k \downarrow 0$ such that

$$\int_{\mathbb{R}^d} e^{\zeta_k |x|} \mu_{n_k}(dx) > 2. \tag{2.4.55}$$

From Lemma 10 it follows that $\{\mu_{n_k}, k \geq 1\}$ is tight. Suppose along a further subsequence μ_{n_k} converges to some measure μ_0. Then sending $k \to \infty$ along this subsequence in (2.4.55) and using Lemma 10 once again we arrive at a contradiction. This proves the first statement in (1). The second statement in (1) is an immediate consequence of Theorem 2.4.2.

(2) This is immediate from part (1) and Theorem 2.4.1.

(3) It is easy to check that for all $t > 0$, $N \in \mathbb{N}$ (see proof of Corollary 2.5 in [11])

$$\Gamma_0(\mathcal{C}_t^0, N) \leq \frac{C_0}{\left(1 + \frac{N}{\log \mathcal{C}_t^0}\right)^{1/2} - 1}.$$

Thus recalling the expression for $\mathcal{C}_t^0$ in (2.4.49) we see that $\lim_{N\to\infty} \sup_{t \geq C_0/2} \Gamma_0(\mathcal{C}_t^0, N) = 0$. Choose $N_1 \in \mathbb{N}$ such that for all $N \geq N_1$ and $t \geq C_0/2$

$$\Gamma_0(\mathcal{C}_t^0, N) \leq \frac{C_0}{8} \leq \frac{t}{4}.$$

Then for all $N \geq N_1$ and $t \geq C_0/2$

$$\alpha_0 \left(\frac{t}{2} - \Gamma_0(\mathcal{C}_t^0, N)\right) \geq \frac{1}{4}\left((1 + \frac{t}{C_0})^{1/2} - 1\right)^2 \geq \frac{1}{16} \frac{(t/C_0)^2}{(1 + t/C_0)} \geq \frac{t}{48 C_0},$$

where the second inequality follows on using the inequality

$$\sqrt{1+x} - 1 \geq \frac{x}{2\sqrt{1+x}}, \; x \geq 0. \tag{2.4.56}$$

Combining this with (2) completes the proof of (3).

(4) From the proof of Corollary 2.5 of [11] it follows that for $t \leq C_0/2$

$$P(\mathcal{W}_1(\eta_n^N, \mu_n) \geq t) \leq A(N, t) \exp\left(-B_1 N t^2\right) \tag{2.4.57}$$

where $A(N, t) = \exp\left(\frac{N B_2}{\left((1 + N/\log \mathcal{C}_t^0)^{1/2} - 1\right)^2}\right)$, $B_1 = (\sqrt{2} - 1)^2/(2C_1^2)$ and $B_2 = 4(\sqrt{2} - 1)^2$.

From (2.4.56) note that if $N > \log \mathcal{C}_t^0$

$$\left((1+N/\log \mathcal{C}_t^0)^{1/2}-1\right)^2 \geq \frac{N}{8\log \mathcal{C}_t^0}.$$

Thus for all such N, t, $A(N,t) \leq \exp(8B_2 \log \mathcal{C}_t^0)$. Thus if additionally $N \geq \frac{16B_2}{B_1}\frac{\log \mathcal{C}_t^0}{t^2}$, the right side of (2.4.57) is bounded above by $\exp(-B_1 N t^2/2)$. The result follows. $\square$

Proof of Theorem 2.3.4.

In this section we complete the proof of Theorem 2.3.4.

Fix $\gamma \in (0, 1-\|A\|)$. From (3.4.101), for any $\varepsilon > 0$,

$$\begin{aligned} P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] &\leq P[\mathcal{W}_1(\eta_n^N, \mu_n) > \gamma\varepsilon] + \sum_{i=0}^{n-1} P[\mathcal{W}_1(\eta_i^N, \mu_i) \geq \frac{\gamma\varepsilon}{\delta M}(\frac{1-\gamma}{\chi_2})^{n-i}] \\ &= P[\mathcal{W}_1(\eta_n^N, \mu_n) > \gamma\varepsilon] + \sum_{i=1}^{n} P[\mathcal{W}_1(\eta_{n-i}^N, \mu_{n-i}) \geq \frac{\gamma\varepsilon}{\delta M}\vartheta^i] \\ &\equiv T_1 + T_2, \end{aligned} \tag{2.4.58}$$

where $\vartheta = \frac{1-\gamma}{\chi_2}$, which, in view of our assumption on δ, is strictly larger than 1. Let $i^\varepsilon = \max\{i \geq 0 : \frac{\varepsilon\gamma}{\delta M}\vartheta^i < \frac{C_0}{2}\}$. Then

$$T_2 = \sum_{i=1}^{i^\varepsilon} P[\mathcal{W}_1(\eta_{n-i}^N, \mu_{n-i}) \geq \frac{\gamma\varepsilon}{\delta M}\vartheta^i] + \sum_{i=i^\varepsilon+1}^{n} P[\mathcal{W}_1(\eta_{n-i}^N, \mu_{n-i}) \geq \frac{\gamma\varepsilon}{\delta M}\vartheta^i].$$

Note that since $\vartheta > 1$ and $t \mapsto \frac{\mathcal{C}_t^0}{t^2}$ is non-increasing, $N \geq L_3 \frac{\log \mathcal{C}^0_{m_\gamma(\varepsilon)}}{m_\gamma^2(\varepsilon)}$ implies $N \geq L_3 \frac{\log \mathcal{C}^0_{m_\gamma(\varepsilon\vartheta^i)}}{m_\gamma^2(\varepsilon\vartheta^i)}$ for all $i \geq 0$ where m_γ is as introduced in Remark 4. Therefore from Lemma

11(4), for all such N

$$\sum_{i=1}^{i^\varepsilon} P[\mathcal{W}_1(\eta_{n-i}^N, \mu_{n-i}) \geq \frac{\gamma\varepsilon}{\delta M}\vartheta^i] \leq \sum_{i=1}^{i^\varepsilon} \exp\{-L_2 N m_\gamma^2(\varepsilon\vartheta^i)\}.$$

Also, from Lemma 11(3), for all $N \geq N_1$,

$$\sum_{i=i^\varepsilon+1}^{n} P[\mathcal{W}_1(\eta_{n-i}^N, \mu_{n-i}) \geq \frac{\gamma\varepsilon}{\delta M}\vartheta^i] \leq \sum_{i=i^\varepsilon+1}^{n} \exp\{-L_1 N m_\gamma(\varepsilon\vartheta^i)\}.$$

Combining these estimates and letting $N_2 = \max\{N_1, L_3\}$ and $\tilde{\varsigma}_1(t) = \max\{1, \frac{\log \mathcal{C}^0_{m_\gamma(t)}}{m_\gamma^2(t)}\}$, we have for all $N \geq N_2\tilde{\varsigma}_1(\varepsilon)$

$$T_2 \leq 2\sum_{i=1}^{\infty} \exp\left\{-L_4 N(\varepsilon^2 \wedge \varepsilon)\vartheta^i\right\},$$

where $L_4 = \min\{L_2\frac{\gamma^2}{M^2}, L_1\frac{\gamma}{M}\}$. Let $k_0 \in \mathbb{N}$ be such that for all $k \geq k_0$, $\vartheta^k \geq k$. Then

$$T_2 \leq 2k_0 \exp\left\{-L_4 N(\varepsilon^2 \wedge \varepsilon)\right\} + 2\frac{\exp\left\{-L_4 N(\varepsilon^2 \wedge \varepsilon)\right\}}{1 - \exp\left\{-L_4 N(\varepsilon^2 \wedge \varepsilon)\right\}}.$$

Noting that $1 - \exp\{-L_4 N(\varepsilon^2 \wedge \varepsilon)\} \geq 1/2$ whenever $N \geq \frac{\log 2}{L_4}(\frac{1}{\varepsilon^2} \vee \frac{1}{\varepsilon})$, we see that with $\varsigma_1^*(t) = \max\{1, \frac{\log \mathcal{C}^0_{m_\gamma(t)}}{m_\gamma^2(t)}, \frac{1}{t^2}, \frac{1}{t}\}$ and $N_3 = \max\{N_1, L_3, \frac{\log 2}{L_4}\}$

$$T_2 \leq 2(k_0 + 2)\exp\left\{-L_4 N(\varepsilon^2 \wedge \varepsilon)\right\} \quad \text{for all } N \geq N_3\varsigma_1^*(\varepsilon). \tag{2.4.59}$$

Also from Lemma 11, for all $N \geq N_3 \max\{1, \frac{\log \mathcal{C}^0_{\gamma\varepsilon}}{\gamma^2\varepsilon^2}\}$

$$T_1 \leq \exp\left\{-L_5 N(\varepsilon^2 \wedge \varepsilon)\right\}, \tag{2.4.60}$$

where $L_5 = \min\{\gamma^2 L_2, \gamma L_1\}$. Using (2.4.59) and (2.4.60) in (3.4.109) we now get the desired result with $a_1 = 2(k_0+2)+1$, $a_2 = \min\{L_4, L_5\}$, $N_0 = N_3$ and $\varsigma_1(t) = \max\{\varsigma_1^*(t), \frac{\log \mathcal{C}^0_{\gamma t}}{\gamma^2 t^2}\}$.

□

Chapter3
AN IP MODEL FOR ACTIVE CHEMOTAXIS

3.1 Introduction

In contrast to chapter 2 here we will study long time asymptotics of another particle system where interactions among agents with other components of system are more involved. We consider the following model, variations of which have been proposed (see [15],[69] and references therein) for a number of different phenomena in social sciences, biological systems and self organized networks. For each $i = 1, \ldots, N$ $X_i(0) = x_i \in \mathbb{R}^d$ and for a function $h(\cdot,\cdot)$ such that $h(0,\cdot) = h(\cdot)$,

$$\begin{aligned} dX_i(t) &= \Big[-(I-A)X_i(t) + \nabla h(t, X_i(t)) + \frac{1}{n}\sum_{j=1, j\neq i}^{n} K\big(X_i(t), X_j(t)\big)\Big] dt \\ &\quad + dW_i(t), \\ \frac{\partial}{\partial t} h(t,x) &= -\alpha h(t,x) + D \triangle h(t,x) + \frac{\beta}{N}\sum_{i=1}^{N} g(X_i(t), x). \end{aligned} \tag{3.1.1}$$

Here $W_i, i = 1, \ldots, N$ are independent Brownian motions that drive the state process X_i of the N interacting particles. The interaction between the particles arises directly from the evolution equation (3.1.1) and indirectly through the underlying potential field h which changes continuously according to a diffusion equation and through the aggregated input of the N particles. One example of such an interaction is in chemotaxis where cells preferentially move towards a higher chemical concentration and themselves release chemicals into the medium, in response to the local information on the environment, thus modifying the potential field dynamically over time. In this context, $h(t, x)$ represents the concentration of a chemical at time t and location x. Diffusion of the chemical in the medium is captured by the Laplacian in (3.1.1) and the constant $\alpha > 0$ models the rate of decay or dissipation of the chemical. The first equation in (3.1.1) describes the motion of a particle in terms of diffusion process with a drift consisting of three terms. The first term models a restoring force towards the origin where origin represents the natural rest state of the particles. The second term is the gradient of the chemical concentration and captures the fact that particles tend to move particularly towards regions of higher chemical concentration. Finally the third term captures the interaction (e.g attraction or repulsion) between the particles which is represented by the kernel $K(\cdot, \cdot)$. Note that $K = 0$ corresponds to the case (1.3.2)l. Contribution of the agents to the chemical concentration field is given through the last term in the second equation. The function g captures the agent response rules and can be used to model a wide range of phenomenon [70]. Similar models where the particles follow a chemical gradient and themselves actively modify the chemical field, have been proposed for movement of Leukocytes, gliding paths of Myxobacteria (See [41], [16] and references therein) and formation of ant trails etc. However even very basic questions for this continuous time model, such as well posedness, large N limit behavior of fixed $t > 0$, and characterization of the nonlinear dynamical system are not very well understood. A precise mathematical treatment of (3.1.1) presents significant technical obstacles and existing results in literature are limited to simulation and formal asymptotic approximation of the system. In [15] the authors considered a discrete time model which captures some of the key features of the dynamics in (3.1.1) and studied several long time properties of the system. One aspect that greatly simplified the analysis of [15] is that the state space of the particles is taken to be a compact set in $\mathbb{R}^d$. However this requirement is restrictive and may be unnatural for the time scales at which the particle evolution is being modeled. In [69] authors had considered a number of variations of (3.1.1). The theoretical properties obtained in this work on the long time behavior of the particle system can be also applied for such systems with some minor modifications.

We now give a general description of the N- particle system that gives a discrete time approximation of the mechanism outlined above. The space of real valued bounded measurable functions on S is denoted as $BM(S)$. Borel σ field on a metric space will be denoted as $\mathcal{B}(S)$. $\mathcal{C}_b(S)$ denotes the space of all bounded and continuous functions $f : S \to \mathbb{R}$. For a measurable space S, $\mathcal{P}(S)$ denotes the space of all probability measures on S. For $k \in \mathbb{N}$, let $\mathcal{P}_k(\mathbb{R}^d)$ be the space of $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that

$$\|\mu\|_k := \left(\int |x|^k d\mu(x)\right)^{\frac{1}{k}} < \infty.$$

Consider a system of N interacting particles that evolve in $\mathbb{R}^d$ governed by a random dynamic chemical field according to the following discrete time stochastic evolution equation given on some probability space $(\Omega, \mathbb{F}, P)$. Suppose that the chemical field at time instant n is given by a nonnegative C^1 (i.e continuously differentiable) real function on $\mathbb{R}^d$ satisfying $\int_{\mathbb{R}^d} \eta(x)dx = 1$. Then, given that particle state at time instant n is x and the empirical measure of the particle states at time n is μ, the particle state X^+ at time $(n+1)$ is given as

$$X^+ = Ax + \delta f(\nabla\eta(x), \mu, x, \epsilon) + B(\epsilon), \tag{3.1.2}$$

where A is a $d \times d$ matrix, δ is a small parameter, ϵ is a $\mathbb{R}^m$ valued random variable with probability law θ and $f : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^m \longrightarrow \mathbb{R}^d$ is a measurable function. Here we consider a somewhat more general form of dependence of the particle evolution on the concentration profile than the additive form that appears in (3.1.1). Additional assumptions on A, θ, f will be introduced shortly. Nonlinearity (modeled by f and B) of the system can be very general and as described below. Denote by $X_n^i \equiv X_n^{i,N}$ (a $\mathbb{R}^d$ valued random variable) the state of the i-th particle $(i = 1, \ldots, N)$ and by η_n^N the chemical concentration field at time instant n. Let $\mu_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_n^i}$ be the empirical measure of the particle values at time instant n. The stochastic evaluation equation for the N-particle system is given as

$$X_{n+1}^i \quad = \quad AX_n^i + \delta f(\nabla\eta_n^N(X_n^i), \mu_n^N, X_n^i, \epsilon_{n+1}^i) + B(\epsilon_{n+1}^i). \tag{3.1.3}$$

for $i = 1, \ldots, N,\ n \in \mathbb{N}_0$. In (3.1.3) $\{\epsilon_n^i, i = 1, ..., N, \quad n \geq 1\}$ is an i.i.d array of $\mathbb{R}^m$ valued random variables with common probability law θ. Here $\{X_0^i, i = 1, ..., N\}$ are assumed to be exchangeable with common distribution μ_0 where $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$. Note that in the notation we have suppressed the dependence of the sequence $\{X_n^i\}$ on N.

We now describe the evolution of the chemical field approximating the second equation in (3.1.1) and its interaction with the particle system. A transition probability kernel on S is a map $P : S \times \mathcal{B}(S) \to [0,1]$ such that $P(x,\cdot) \in \mathcal{P}(S) \quad \forall x \in S$ and for each $A \in \mathcal{B}(S)$, $P(\cdot, A) \in BM(S)$. Given the concentration profile at time n is a C^1 probability density function η on $\mathbb{R}^d$ and the empirical measure of the state of N-particles at time instant n is μ, the concentration probability density η^+ at time $(n+1)$ is given by the relation

$$\eta^+(y) = \int_{\mathbb{R}^d} \eta(x) R^\alpha_\mu(x,y) l(dx) \tag{3.1.4}$$

where l denotes the Lebesgue measure on $\mathbb{R}^d$, and $R^\alpha_\mu(x,y)$ is the Radon-Nikodym derivative of the transition probability kernel with respect to the Lebesgue measure $l(dy)$ on $\mathbb{R}^d$. The kernel R^α_μ is given as follows. Let P and P' betwo transition probability kernels on $\mathbb{R}^d$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\alpha \in (0,1)$ define the transition probability kernel R^α_μ on $\mathbb{R}^d$ as

$$R^\alpha_\mu(x,C) := (1-\alpha)P(x,C) + \alpha\mu P'(C), \qquad x \in \mathbb{R}^d, C \in \mathcal{B}(\mathbb{R}^d).$$

Here P represents the background diffusion of the chemical concentration while $\delta_x P'$ captures the contribution to the field by a particle with location x. The parameter α gives a convenient way for combining the contribution from the background diffusion and the individual particles. For each $x \in \mathbb{R}^d$, both $P(x,\cdot)$ and $P'(x,\cdot)$ are assumed to be absolutely continuous with respect to Lebesgue measure and throughout this chapter we will denote the corresponding Radon-Nykodim derivatives with the same notations $P(x,\cdot)$ and $P'(x,\cdot)$ respectively. Additional properties of P and P' will be specified shortly. The evolution equation for the chemical field is then given as

$$\eta_{n+1}^N(y) = \int_{\mathbb{R}^d} \eta_n^N(x) R^{\alpha}_{\mu_n^N}(x,y) l(dx). \tag{3.1.5}$$

In contrast to the model studied in Chapter 2 the situation here is somewhat more involved. Note that $\{X_n(N)\}_{n\geq 0} := (X_n^{1,N}, X_n^{2,N}, \ldots, X_n^{N,N})_{n\geq 0}$ is not a Markov process and in order to get a Markovian state descriptor one needs to consider $\{X_n(N), \eta_n^N\}_{n\geq 0}$ which is a discrete time Markov chain with values in $(\mathbb{R}^d)^N \times \mathcal{P}(\mathbb{R}^d)$.

We will show that as $N \to \infty$, $(\mu_n^N, \eta_n^N)_{n\in\mathbb{N}_0}$ converges to a deterministic nonlinear dynamical system $(\mu_n, \eta_n)_{n\in\mathbb{N}_0}$. Under conditions on f, g, θ, δ, A and smoothness parameters of densities of transition kernels P, P' we study several long time asymptotic properties of the N-particle system and the corresponding nonlinear limit dynamical system. *The stochastic dynamical system* (3.1.3) - (3.1.5) *can be regarded as a perturbation of a linear stable stochastic dynamical system with a small interaction term and our results give explicit range of values of the perturbation parameter* δ, α *for which the weakly interacting system has desirable long time properties*. Stochastic systems in (3.1.3)-(3.1.5) can be viewed as discrete time approximations of many stochastic differential equation models for weakly interacting particles.

The gradient of a real differentiable function f on $\mathbb{R}^d$ denoted by ∇f is defined as the d dimensional vector field $\nabla f := (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d})'$. For a function $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$

$$\nabla_x f(x,y) := \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d}\right)'.$$

The function $\nabla_y f(x,y)$ is defined similarly. Absolute continuity of a measure μ with respect to a measure ν will be denoted by $\mu \ll \nu$. We will denote the Radon-Nikodym derivative of μ with respect to ν by $\frac{d\mu}{d\nu}$. For $f \in BM(\mathcal{S})$ and a transition probability kernel P on S, define $Pf \in BM(\mathcal{S})$ as $Pf(\cdot) = \int_S f(y)P(\cdot, dy)$. For any closed subset $B \in S$, and $\mu \in \mathcal{P}(B)$, define $\mu P \in \mathcal{P}(S)$ as $\mu P(A) = \int_B P(x,A)\mu(dx)$. For a matrix B the usual operator norm is denoted by $\|B\|$.

3.2 Description of Nonlinear System:

We now describe the nonlinear dynamical system obtained on taking the limit $N \to \infty$ of (μ_n^N, η_n^N). Given a C^1 density function ρ on $\mathbb{R}^d$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$, define a transition probability kernel $Q^{\rho,\mu}$ on $\mathbb{R}^d$ as

$$Q^{\rho,\mu}(x,C) = \int_{\mathbb{R}^m} 1_{\{Ax+\delta f(\nabla\rho(x),\mu,x,z)+B(z)\in C\}}\theta(dz), \qquad (x,C) \in \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d).$$

With an abuse of notation we will also denote by $Q^{\rho,\mu}$ the map from $BM(\mathbb{R}^d)$ to itself, defined as

$$Q^{\rho,\mu}\phi(x) = \int_{\mathbb{R}^d} \phi(y) Q^{\rho,\mu}(x,dy), \quad \phi \in BM(\mathbb{R}^d), x \in \mathbb{R}^d.$$

For $\mu, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$, let $\mu Q^{\rho,\mu_1} \in \mathcal{P}(\mathbb{R}^d)$ be defined as

$$\mu Q^{\rho,\mu_1}(C) = \int_{\mathbb{R}^d} Q^{\rho,\mu_1}(x,C)\mu(dx), \quad C \in \mathcal{B}(\mathbb{R}^d). \tag{3.2.1}$$

Note that $\mu Q^{\rho,\mu_1} = \mathcal{L}\big(AX + \delta f(\nabla\rho(X), \mu_1, X, \epsilon) + B(\epsilon)\big)$ where $\mathcal{L}(X,\epsilon) = \mu \otimes \theta$.

Define $\mathcal{P}_1^*(\mathbb{R}^d) := \{\mu \in \mathcal{P}_1(\mathbb{R}^d) : \mu \ll l, \frac{d\mu}{dl}$ is continuously differentiable and $\|\nabla \frac{d\mu}{dl}\|_1 <$

$\infty\}$. For notational simplicity we will identify an element in $\mathcal{P}_1^*(\mathbb{R}^d)$ with its density and denote both by the same symbol. Define the map $\Psi : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d) \to \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$ as

$$\Psi(\mu, \eta) \quad = \quad (\mu Q^{\eta,\mu}, \eta R_\mu^\alpha), \qquad (\mu, \eta) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d). \tag{3.2.2}$$

Under suitable assumptions (which will be introduced in Section 3.3) it will follow that for every $(\mu, \eta) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$, η^+ defined by (3.1.4) is in $\mathcal{P}_1^*(\mathbb{R}^d)$ and $\mu Q^{\eta,\mu}$ defined by (3.2.1) is in $\mathcal{P}_1(\mathbb{R}^d)$. Thus (under those assumptions) Ψ is a map from $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$ to itself. Using the above notation we see that $\{(X_n^1, \ldots, X_n^N), \mu_n^N, \eta_n^N\}_{n\geq 0}$ is a $(\mathbb{R}^d)^N \times \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$ valued discrete time Markov chain defined recursively as follows. Let $X_k(N) \equiv (X_k^1, X_k^2, ..., X_k^N)$, and η_0^N be the initial chemical field which is a random element of $\mathcal{P}_1^*(\mathbb{R}^d)$. Let $\mathcal{F}_0 = \sigma\{X_0(N), \eta_0^N\}$. Then, for $k \geq 1$

$$\begin{cases} P(X_k(N) \in C | \mathcal{F}_{k-1}^N) = \bigotimes_{i=1}^N (\delta_{X_{k-1}^j} Q^{\eta_{k-1}^N, \mu_{k-1}^N})(C) \quad \forall C \in \mathcal{B}(\mathbb{R}^{dN}), \\ \mu_k^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_k^i}, \\ \eta_k^N = \eta_{k-1}^N R^\alpha_{\mu_{k-1}^N}, \\ \mathcal{F}_k^N = \sigma\{\eta_k^N, X_k(N)\} \vee \mathcal{F}_{k-1}^N. \end{cases} \tag{3.2.3}$$

We will call this particle system as $\mathbb{IPS}_1$. We next describe a nonlinear dynamical system which is the formal Vlasov-McKean limit of the above system, as $N \to \infty$. Given $(\mu_0, \eta_0) \in$

$\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$ define a sequence $\{(\mu_n, \eta_n)\}_{n\geq 0}$ in $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$ as

$$\mu_{n+1} = \mu_n Q^{\eta_n, \mu_n}, \qquad \eta_{n+1} = \eta_n R^{\alpha}_{\mu_n}, \qquad n \geq 0. \tag{3.2.4}$$

Using (3.2.2) the above evolution can be represented as

$$(\mu_{n+1}, \eta_{n+1}) = \Psi(\mu_n, \eta_n), \qquad n \in \mathbb{N}_0. \tag{3.2.5}$$

As in Chapter 2 the starting point of our investigation on long time asymptotics of the above interacting particle system will be to study the stability properties of (3.2.4). We identify $\eta, \eta' \in \mathcal{P}(\mathbb{R}^d)$ that are equal a.e under the Lebesgue measure on $\mathbb{R}^d$.

From a computational point of view we are primarily interested in approximating (μ_n, η_n) by (μ_n^N, η_n^N) uniformly in time parameter n, with explicit uniform concentration bounds. Such results are particularly useful for developing simulation methods for approximating the steady state distribution of the mean field models such as in (3.2.4). We note here that we view the systems (3.1.3)-(3.1.5) and (3.2.4) in two different ways. One is where N is not too large and the N-particle system is used to obtain a simulation based approximation to the invariant measure of the nonlinear system (3.2.2) and the second is when the physical system of interest is (3.1.3),(3.1.5) but N is too large to allow for a tractable analysis and one instead uses (3.2.4) as a simplified approximate model. In other words, we use the nonlinear system (3.2.4) as an intermediate model to approximate the properties of the physical system (3.1.3) with a large N by those of a simulated system with a numerically tractable number of particles.

The third equation in (3.2.3) makes the simulation of $\mathbb{IPS}_1$ numerically challenging. In section 3.3 we will introduce another particle system referred to as $\mathbb{IPS}_2$ which also gives an asymptotically consistent approximation of (3.2.4) and is computationally more tractable. We show in Theorem 1 that under conditions that include a Lipschitz property of f (Assumptions 6 and 7), smoothness assumptions on the transition kernels of the background diffusion of the chemical medium (Assumption 9) the Wasserstein-1($\mathcal{W}_1$) distance between the occupation measure of the particles along with the chemical medium (μ_n^N, η_n^N) and (μ_n, η_n) converges to 0, for every time instant n. Under an additional condition on the contractivity of A and δ, α being sufficiently small we show that the nonlinear system (3.2.5) has a unique fixed point and starting from an arbitrary initial condition, convergence to the fixed point occurs at a geometric rate. Using these results we next argue in Theorem 3.3.2 that under some integrability conditions (Assumption 12-13), as $N \to \infty$, the empirical occupation measure of the N-particles and density of the chemical medium at time instant n, namely (μ_n^N, η_n^N) converges to (μ_n, η_n) in the $\mathcal{W}_1$ distance, in L^1, *uniformly* in n. This result in particular shows that the $\mathcal{W}_1$ distance between (μ_n^N, η_n^N) and the unique fixed point $(\mu_\infty, \eta_\infty)$ of (3.2.5) converges to zero as $n \to \infty$ and $N \to \infty$ in any order. We next show that for each N, there is unique invariant measure Θ_∞^N of the N-particle dynamics with integrable first moment and this sequence of measures is μ_∞-chaotic, namely as $N \to \infty$, the projection of Θ_∞^N on the first k-coordinates converges to $\mu_\infty^{\otimes k}$ for every $k \geq 1$. This propagation of chaos property all the way to $n = \infty$ crucially relies on the uniform in time convergence of (μ_n^N, η_n^N) to $(\mu_\infty, \eta_\infty)$. Such a result is important since it says that the steady state of a N-dimensional fully coupled Markovian system has a simple approximate description in terms of a product measure when N is large. This result is key in developing particle based numerical schemes for approximating the fixed point of the evolution equation (3.2.5). Next we present some uniform in time concentration bounds of $\mathcal{W}_1(\mu_n^N, \mu_n) + \mathcal{W}_1(\eta_n^N, \eta_n)$. Proof is very similar to that of Theorem 3.8 of Chapter 2 so we only provide a sketch.

3.3 Main Results:

We now introduce our main assumptions on the problem data. Recall that $\{X_0^i, i = 1, \dots N\}$ is assumed to be exchangeable with common distribution μ_0. We assume further $(\mu_0, \eta_0) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$. For a $d \times d$ matrix B we denote its norm by $\|B\|$, i.e. $\|B\| = \sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{|Bx|}{|x|}$.

Assumption 6. *The error distribution θ is such that $\int A_1(z)\theta(dz) := \sigma \in (0, \infty)$ where*

$$A_1(\epsilon) := \sup_{\{x_1,x_2,y_1,y_2 \in \mathbb{R}^d, \mu_1,\mu_2 \in \mathcal{P}_1(\mathbb{R}^d) : (x_1,y_1,\mu_1) \neq (x_2,y_2,\mu_2)\}} \frac{|f(y_1,\mu_1,x_1,\epsilon) - f(y_2,\mu_2,x_2,\epsilon)|}{|x_1 - x_2| + |y_1 - y_2| + \mathcal{W}_1(\mu_1,\mu_2)} \tag{3.3.1}$$

It follows from (3.3.1) that $\forall x, y \in \mathbb{R}^d, \mu \in \mathcal{P}_1(\mathbb{R}^d)$,

$$|f(y, \mu, x, \epsilon)| \leq (|y| + \|\mu\|_1 + |x|)A_1(\epsilon) + A_2(\epsilon) \tag{3.3.2}$$

where $A_2(\epsilon) := f(0, 0, \epsilon)$.

Recall the function $B : \mathbb{R}^m \to \mathbb{R}^d$ introduced in (4.1.10).

Assumption 7. *The error distribution θ is such that*

$$\int_{\mathbb{R}^m} \Big(A_2(z) + |B(z)|\Big)\theta(dz) < \infty.$$

Assumption 8. *η_0^N (the density function) is a Lipschitz function on $\mathbb{R}^d$ and $\eta_0^N \in \mathcal{P}_1^*(\mathbb{R}^d)$.*

Assumptions 9 and 10 on the kernels P and P' hold quite generally. In particular, they are satisfied for Gaussian kernels.

Assumption 9. *There exist $l_P^\nabla \in (0,1]$ and $l_{P'}^\nabla \in (0,\infty)$ such that for all $x, y, x', y' \in \mathbb{R}^d$*

$$\begin{aligned} |\nabla_y P(x,y) - \nabla_y P(x',y')| &\leq l_P^\nabla(|y-y'| + |x-x'|) && (3.3.3) \\ |\nabla_y P'(x,y) - \nabla_y P'(x',y')| &\leq l_{P'}^\nabla(|y-y'| + |x-x'|). && (3.3.4) \end{aligned}$$

Furthermore

$$\sup_{x\in\mathbb{R}^d}\{|\nabla_y P(x,0)| \vee |\nabla_y P'(x,0)|\} < \infty. \tag{3.3.5}$$

Using the Lipschitz property in (3.3.3) and the growth condition (3.3.5) *one has the linear growth property for some $M_P^\nabla \in (0,\infty)$*

$$sup_{x\in\mathbb{R}^d}|\nabla_y P(x,y)| \leq M_P^\nabla(1+|y|). \tag{3.3.6}$$

A similar inequality holds for P' from (3.3.4) with $M_{P'}^\nabla \in (0,\infty)$.

Denote $(1-\alpha)l_P^\nabla + \alpha l_{P'}^\nabla$ by $l_{PP'}^{\nabla,\alpha}$.

Assumption 10. *For every $f \in Lip_1(\mathbb{R}^d)$, Pf and $P'f$ are also Lipschitz and*

$$\sup_{f\in Lip_1(\mathbb{R}^d)} \sup_{x\neq y\in\mathbb{R}^d} \frac{Pf(x) - Pf(y)}{|x-y|} := l(P) < \infty$$

Also $l(P')$ defined as above for P' is finite.

Assumption 11. *Both $P(x,\cdot)$ and $P'(x,\cdot)$ are such that for any compact set $K \subset \mathbb{R}^d$, the families of probability measures $\{P(x,\cdot) : x \in K\}$ and $\{P'(x,\cdot) : x \in K\}$ are both uniformly integrable.*

Let $\max\{l(P), l(P')\} = l_{PP'}$.

Remark 5. *Assumption 10 is satisfied if P, P' are given as follows. For any $f \in \mathcal{C}_b(\mathbb{R}^d)$, let*

$$Pf(\cdot) := Ef(g_1(\cdot,\varepsilon_1)), \qquad P'f(\cdot) := Ef(g_2(\cdot,\varepsilon_2)) \tag{3.3.7}$$

where $\varepsilon_1, \varepsilon_2$ are $\mathbb{R}^m$ valued random variables and $\varepsilon_1, \varepsilon_2$ and $g_1, g_2 : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ are maps with following properties:

$$E(G_1(\varepsilon_1)) \leq l(P) \qquad \textit{and} \qquad E(G_2(\varepsilon_2)) \leq l(P'), \tag{3.3.8}$$

where

$$G_1(y) := \sup_{x_1 \neq x_2} \frac{g_1(x_1,y) - g_1(x_2,y)}{|x_1 - x_2|} \quad \textit{and} \quad G_2(y) := \sup_{x_1 \neq x_2} \frac{g_2(x_1,y) - g_2(x_2,y)}{|x_1 - x_2|}. \tag{3.3.9}$$

Simulation of the system (3.2.3) is numerically intractable due to the step that involves the updating of η^N_{n-1} to η^N_n. This requires computing the integral in (3.1.4) which, since R^α_μ is a mixture of two transition kernels, over time leads to an explosion of terms in the mixture that need to be updated. An approach (proposed in [15]) that addresses this difficulty is, instead of directly updating η^N_{n-1}, to use the empirical distribution of the observations drawn independently from η^N_{n-1}. This leads to the following particle system

Denote $\bar{X}_0(N)$ by $(\bar{X}_0^{1,N}, \ldots, \bar{X}_0^{N,N})$ a sample of size N from μ_0. Let $M \in \mathbb{N}$. The new particle scheme will be described as a family $(\bar{X}_k(N), \bar{\mu}_k^N, \bar{\eta}_k^M)_{k\in\mathbb{N}_0}$ of $(\mathbb{R}^d)^N \times \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}^*(\mathbb{R}^d)$ valued random elements on some probability space defined recursively as follows. Set $\bar{X}_0(N) = (\bar{X}_0^{1,N}, \ldots, \bar{X}_0^{N,N}), \bar{\eta}_0^M = \eta_0, \bar{\mathcal{F}}_0^{M,N} = \sigma(\bar{X}^N(0))$. For $k \geq 1$

$$\begin{cases} \bar{\mu}_k^N = \frac{1}{N}\sum_{i=1}^N \delta_{\bar{X}_k^i}, \\ P(\bar{X}_k(N) \in C | \mathcal{F}_{k-1}^{M,N}) = \bigotimes_{i=1}^N (\delta_{\bar{X}^j_{k-1}} Q^{\bar{\eta}^M_{k-1}, \bar{\mu}^N_{k-1}})(C) \quad \forall C \in \mathcal{B}(\mathbb{R}^d)^N, \\ \bar{\eta}_k^M = (1-\alpha)(S^M(\bar{\eta}_{k-1}^M)P) + \alpha \bar{\mu}_{k-1}^N P', \\ \bar{\mathcal{F}}_k^{M,N} = \sigma\{\bar{\eta}_k^M, \bar{X}_k(N)\} \vee \bar{\mathcal{F}}_{k-1}^{M,N} \end{cases} \tag{3.3.10}$$

where $S^M(\bar{\eta}^M_{k-1})$ is the random measure defined as $\frac{1}{M}\sum_{i=1}^M \delta_{Y^{i,M}_{k-1}}$ where $\{Y^{i,M}_{k-1}\}_{i=1,\ldots,M}$ conditionally on $\bar{\mathcal{F}}^{M,N}_{k-1}$, are M i.i.d distributed according to $\bar{\eta}^M_{k-1}$. We will call this particle system as $\mathbb{IPS}_2$. We remark that our notation is not accurate since both the quantities $\bar{\mu}_k^N, \bar{\eta}_k^M$ depend on M, N. The superscripts only describe the number of particles/samples used in the procedure to combine them. Note that like $\mathbb{IPS}_1$, here $(\bar{X}_k(N), \bar{\eta}_k^M)_{k\geq 0}$ is not a Markov chain on $(\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d)$ anymore. Rather $(\bar{X}^N(k), \bar{\eta}_k^M, S^M(\bar{\eta}_k^M))_{k\geq 0}$ is a discrete time Markov chain on $(\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$.

For any random variable Z we denote $E\big[Z\big|\mathcal{F}_k^{M,N}\big]$ by $E_k^{M,N}\big[Z\big]$. The following result shows that the particle systems in (3.2.3) and (3.3.10) approximate the dynamical system in (3.2.4) as N (respectively $\min\{M,N\}$) becomes large.

Theorem 1. *Suppose Assumptions 6,7,9 and 10 hold.*

1. *Consider the particle system* $\mathbb{IPS}_1$ *in (3.1.3 - 3.1.5). Suppose the sampling of the exchangeable datapoints* $X_0(N) \equiv (X_0^1, X_0^2, \ldots, X_0^N)$ *is exchangeable and* $\{\mathcal{L}(X_0(N))\}_{N\in\mathbb{N}}$ *is* μ_0*- chaotic. Suppose* $E\mathcal{W}_1(\eta_0^N, \eta_0) \to 0$ *as* $N \to \infty$*. Then, as* $N \to \infty$

$$E\left[\mathcal{W}_1(\mu_n^N, \mu_n) + \mathcal{W}_1(\eta_n^N, \eta_n)\right] \to 0 \tag{3.3.11}$$

   for all $n \geq 0$ *where* μ_n, η_n *are as in (3.2.4).*

2. *Consider the second particle system* $\mathbb{IPS}_2$*. Suppose that in addition Assumption 11 holds. Suppose the sampling of the exchangeable datapoints* $\bar{X}_0(N) \equiv (\bar{X}_0^1, \bar{X}_0^2, \ldots, \bar{X}_0^N)$ *is exchangeable and* $\{\mathcal{L}(\bar{X}_0(N))\}_{N\in\mathbb{N}}$ *is* μ_0*- chaotic. Then as* $\min\{N, M\} \to \infty$

$$E\left[\mathcal{W}_1(\bar{\mu}_n^N, \mu_n) + \mathcal{W}_1(\bar{\eta}_n^M, \eta_n)\right] \to 0 \tag{3.3.12}$$

   for all $n \geq 0$*.*

As a consequence of Theorem 1, we have a finite time propagation of chaos result of the following form. Let $\nu_n^N = \mathcal{L}(X_n^{1,N}, X_n^{2,N}, \ldots, X_n^{N,N})$.

Corollary 3.3.1. *Under Assumptions as in Theorem 1 the family* $\{\nu_n^N\}_{N\geq 1}$ *is* μ_n *chaotic for every* $n \geq 1$*.*

As noted in introduction, the primary goal is to study long time properties of (3.1.3) and the non-linear dynamical system (3.2.4). Main contributions of the work are as follows. First we will identify the range of values of modeling parameters that leads to stability of the system. Secondly, we will give non-asymptotic bounds on convergence rates of the particle system to the deterministic nonlinear dynamics that are uniform in time and study their consequences for the steady state behavior.

Theorem 3.3.1. *Suppose Assumptions* (6) - (10) *hold. Then there exist* $\omega_0, \alpha_0, \delta_0 \in (0,1)$ *such that whenever* $\|A\| < \omega_0, \alpha \in (0, \alpha_0)$*, and* $\delta \in (0, \delta_0)$*, the map* Ψ *defined in (3.2.2) has a unique fixed point* $(\mu_\infty, \eta_\infty)$ *in* $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$.

We now provide a result that will strengthen the convergence in (3.3.11) to uniform convergence over all $n \in \mathbb{N}$ and also give rates of convergence. For this result we will need the following additional conditions.

Assumption 12. *For some* $\tau > 0$,

$$\begin{aligned} \mu_0 \in \mathcal{P}_{1+\tau}(\mathbb{R}^d), \qquad & \int A_1(z)^{1+\tau}\theta(dz) := \sigma_1(\tau) < \infty \\ \int \Big(A_2(z) + |B(z)|\Big)^{1+\tau}\theta(dz) \;\; &:= \;\; \sigma_2(\tau) < \infty. \end{aligned} \tag{3.3.13}$$

We need to impose the following condition on P, P' for uniform in time convergence.

Assumption 13. *For some* $\langle |x|^{1+\tau}, \eta_0\rangle < \infty$*. There exist* $m_\tau(P)$ *and* $m_\tau(P')$ *in* $\mathbb{R}^+$ *such that following holds for all* $x \in \mathbb{R}^d$

$$\int_{\mathbb{R}^d} |y|^{1+\tau} P(x, dy) \le m_\tau(P)\left(1 + |x|^{1+\tau}\right), \text{ and } \int_{\mathbb{R}^d} |y|^{1+\tau} P'(x, dy) \le m_\tau(P')\left(1 + |x|^{1+\tau}\right).$$

Now we state a generalization of the Theorem 1, which gives the convergence rate of

$$E\left\{\mathcal{W}_1(\bar{\mu}_n^N, \mu_n) + \mathcal{W}_1(\bar{\eta}_n^M, \eta_n)\right\} \to 0$$

uniformly over all $n \geq 0$ in a nonasymptotic manner.

Recall $l_P^{\nabla}, l_{P'}^{\nabla}$ introduced in Assumption 8. For $\alpha \in (0,1)$, let $l_{PP'}^{\nabla,\alpha} = (1-\alpha)l_P^{\nabla} + \alpha l_{P'}^{\nabla}$. With the notations of Assumption 6 we define

$$a_0 := \frac{1 - \|A\|}{\sigma(2 + l_{PP'}^{\nabla,\alpha})}.$$

For $(\mu_n, \eta_n), (\mu_n', \eta_n') \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$ define the following distance on $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$

$$\mathcal{W}_1((\mu_n, \eta_n), (\mu_n', \eta_n')) := \mathcal{W}_1(\mu_n, \mu_n') + \mathcal{W}_1(\eta_n, \eta_n').$$

Theorem 3.3.2. *Consider the particle system $\mathbb{IPS}_2$. Suppose Assumptions* (6)-(10) *and Assumptions (12),(13) hold for some $\tau > 0$. Let $N_1 := \min\{M, N\}$. Also assume $\delta \in (0, a_0)$, $\quad (1-\alpha)m_\tau(P) < 1$ and*

$$\max\left\{\left(\|A\| + \delta\sigma(2 + l_{PP'}^{\nabla,\alpha}) + \alpha l(P')\right), (1-\alpha)l(P)\right\} + \delta\sigma \max\left\{\alpha l_{P'}^{\nabla}, (1-\alpha)l_P^{\nabla}\right\} < 1, \qquad .$$

Then there exists $\theta < 1$, and $a \in (0, \infty)$ such that for each $n \geq 0$,

$$E\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_n, \eta_n)\right) - a\theta^n E\mathcal{W}_1\left((\bar{\mu}_0^N, \bar{\eta}_0^M), (\mu_0, \eta_0)\right) \leq b(N_1, \tau, d),$$

where

$$
b(N_1,\tau,d) = C \begin{cases} N_1^{-\max\{\frac{1}{2},\frac{\tau}{1+\tau}\}} & \text{if} \quad d=1, \tau \neq 1, \\ N_1^{-\frac{1}{2}} \log N_1 & \text{if} \quad d=1, \tau = 1, \\ N_1^{-\frac{1}{2}} \log N_1 + N_1^{-\frac{\tau}{1+\tau}} & \text{if} \quad d=2, \tau \neq 1, \\ N_1^{-\frac{1}{2}} (\log N_1)^2 & \text{if} \quad d=2, \tau = 1, \\ N_1^{-\max\{\frac{1}{d},\frac{\tau}{1+\tau}\}} & \text{if} \quad d>2, \tau \neq \frac{1}{d-1}, \\ N_1^{-\frac{1}{d}} \log N_1 & \text{if} \quad d>2, \tau = \frac{1}{d-1}, \end{cases} \tag{3.3.14}
$$

and the value of the constant C will vary for each of the cases.

Remark 6. *For the particle system in* (3.2.3) *similar results hold N_1 replaced by N. For $\mathbb{IPS}_2$ if $\mathcal{L}(\bar{X}_0(N))$ is μ_0 -chaotic then one can show $E\mathcal{W}_1(\bar{\mu}_0^N, \mu_0) \to 0$ as $N \to \infty$ and recall that $\bar{\eta}_0^M = \eta_0$. it follows from Theorem 3.3.2*

$$
\sup_{n \geq 0} E\mathcal{W}_1 \left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_n, \eta_n) \right) \to 0
$$

as $\min\{N, M\} \to \infty$. *Similarly for the paticle system in* (3.2.3) , *if $E\mathcal{W}_1(\eta_0^N, \eta_0) \to 0$ as $N \to \infty$, and $\mathcal{L}(X_0(N))$ is μ_0 -chaotic then following holds*

$$
\sup_{n \geq 0} E\mathcal{W}_1 \left((\mu_n^N, \eta_n^N), (\mu_n, \eta_n) \right) \to 0
$$

as $N \to \infty$.

One consequence of Theorems 3.3.1 and 3.3.2 will be the following interchange of limit result.

Corollary 3.3.2. *Under conditions of Theorem 3.3.2*

$$\begin{aligned}
& \limsup_{\min\{N,M\}\to\infty} \limsup_{n\to\infty} E\mathcal{W}_1((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_\infty, \eta_\infty)) \\
= & \limsup_{n\to\infty} \limsup_{\min\{N,M\}\to\infty} E\mathcal{W}_1((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_\infty, \eta_\infty)) \\
= & \ 0. \qquad (3.3.15)
\end{aligned}$$

Suppose Assumptions of Theorem 3.3.2 hold and let $(\mu_\infty, \eta_\infty)$ be the fixed point of the map Ψ of (3.2.5). We are interested in establishing a propagation of chaos result for $n = \infty$. Recall for $\mathbb{IPS}_2$, $S^M(\bar{\eta}_n^M)$ is the random measure defined as $\frac{1}{M}\sum_{i=1}^M \delta_{Y_n^{i,M}}$ where $\{Y_n^{i,M}\}_{i=1,\dots,M}$ conditionally on $\mathcal{F}_n^{M,N}$, are M i.i.d distributed $\mathbb{R}^d$ valued random variables with law $\bar{\eta}_{k-1}^M$. Denote $Y_n(M) := (Y_n^{1,M}, \dots, Y_n^{M,M})$.

Theorem 3.3.3. *Consider the second particle system $\mathbb{IPS}_2$. Suppose Assumptions 6,7,9,10 hold and suppose further*

$$\delta \in (0, a_0), \qquad \textit{and} \qquad \sum_{i=0}^{\infty} (1-\alpha)^i \int_{\mathbb{R}^d} |y| P' P^i(0, dy) < \infty,$$
$$\max\left\{\left(\|A\| + \delta\sigma(2 + l_{PP'}^{\nabla;\alpha}) + \alpha l(P')\right), (1-\alpha) l(P)\right\} + \delta\sigma \max\left\{\alpha l_{P'}^{\nabla}, (1-\alpha) l_P^{\nabla}\right\} < 1.$$

Then for every $N, M \geq 1$, the Markov process $\left(\bar{X}^N(n), \bar{\eta}_n^M, S^M(\bar{\eta}_n^M)\right)_{n\geq 0}$ on $(\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$ has a unique invariant measure $\Theta_\infty^{N,M}$ if following holds Let $\Theta_\infty^{1,N,M}$ be the marginal distribution on $(\mathbb{R}^d)^N$ of the first co-ordinate of $\Theta_\infty^{N,M}$. Suppose additionally

Assumption 8 and Assumption 12,13 hold and for some $\tau > 0$

$$(1-\alpha)m_{\tau}(P) < 1.$$

Then $\Theta_{\infty}^{1,N,M}$ *is* μ_{∞}*- chaotic as* $\min\{M,N\} \to 0$*, where* μ_{∞} *is defined in Theorem 3.3.1.*

Remark 7. *A similar version also holds for the particle system in* (3.2.3)*.*

3.3.1 Concentration Bounds:

In this section we will give concentration bounds of $(\bar{\mu}_n^N, \bar{\eta}_n^N)$ to (μ_n, η_n), in $\mathcal{W}_1$ metric, that are uniform in n. We established two different types of concentration bounds. The first result considers a setting where initiallly the particles have a general exchangeable distribution where as the second considers a more restrictive setting where the initial law of the particles is iid. We will need the following additional assumptions.

Assumption 14. *(i) For some* $K \in (1,\infty)$*,* $A_1(x) \leq K$ *for* θ *a.e.* $x \in \mathbb{R}^m$*.*

(ii) There exists $\alpha \in (0,\infty)$ *such that* $\int e^{\alpha|x|}\mu_0(dx) < \infty$ *and there exists* $\alpha(\delta) \in (0,\alpha)$ *such that*

$$\int_{\mathbb{R}^m} e^{\alpha(\delta)\left(A_2(z)+\frac{|B(z)|}{\delta}\right)}\theta(dz) < \infty.$$

With $\tau, \sigma_1(\tau)$ defined above in Assumption 12 let

$$a(\tau) := \frac{4^{-\tau} - \|A\|^{1+\tau}}{\sigma_1(\tau)\left[1 + (1 + l_{PP'}^{\nabla,\alpha})^{1+\tau}\right]}. \tag{3.3.16}$$

Theorem 3.3.4. *(Polynomial Concentration) Let* $N_1 = \min\{M,N\}$*. Suppose Assumptions*

(6)-(10) *and Assumptions (12),(13) hold for some* $\tau > 0$. *Suppose that* $\delta \in (0, a(\tau)^{\frac{1}{1+\tau}}), (1-\alpha)m_\tau(P) < 1$ *and*

$$\max\left\{\left(\|A\| + \delta\sigma(2 + l^{\nabla,\alpha}_{PP'}) + \alpha l(P')\right), (1-\alpha)l(P)\right\} + \delta\sigma \max\left\{\alpha l^{\nabla}_{P'}, (1-\alpha)l^{\nabla}_{P}\right\} < 1. \tag{3.3.17}$$

Then there exits $\nu > 1, \gamma \in (0,1)$, $N_0 \in \mathbb{N}_0$ *and* $C_1 \in (0,\infty)$ *such that for all* $\varepsilon > 0$, *and for all* $n \geq 0$,

$$P(\mathcal{W}_1((\mu_n^N, \eta_n^M), (\mu_n, \eta_n) > \varepsilon) \leq P(\mathcal{W}_1((\mu_0^N, \eta_0^M), (\mu_0, \eta_0)) > \gamma\nu^n\varepsilon) + C_1\varepsilon^{-(1+\alpha)}N_1^{-\frac{\tau}{d+2}},$$

for all $N_1 > N_0\left(\max\left\{1, \log^+\varepsilon\right\}\right)^{\frac{d+2}{d}}$.

Remark 8. 1. *Similar concentration bounds hold for the particle system* $\mathbb{IPS}_1$.

2. *Note that* $\nu > 1$. *Under the conditions of Theorem 3.3.4(a) the following uniform in time concentration estimate holds.*

$$\sup_{n\geq 1} P(\mathcal{W}_1((\mu_n^N, \eta_n^M), (\mu_n, \eta_n) > \varepsilon) \leq P(\mathcal{W}_1((\mu_0^N, \eta_0^M), (\mu_0, \eta_0)) > \gamma\varepsilon) + C_1\varepsilon^{-(1+\alpha)}N_1^{-\frac{\tau}{d+2}},$$

for all $N_1 > N_0\left(\max\left\{1, \log^+\varepsilon\right\}\right)^{\frac{d+2}{d}}$. *Same thing will also hold for Theorem 3.3.4(b).*

Note that the bounds in Theorems 3.3.4 depend on the state dimension. The following result shows that where initial locations of N particles are i.i.d and under additional conditions on the parameters in a restricted setting one can obtain dimension independent bounds for $d > 2$.

Theorem 3.3.5. *Consider the particle system $\mathbb{IPS}_1$ with initial condition $\eta_0^N \equiv \eta_0$. Suppose that $\{X_0^{i,N}\}_{i=1,\dots,N}$ are i.i.d. with common distribution μ_0 for each N. Let*

$$
\begin{aligned}
C_{PP'}^{\alpha(1)} &:= \frac{\max\{\|A\| + \delta K(1 + l_{PP'}^{\nabla,\alpha}), \alpha l_{P'}^{\nabla}, (1-\alpha)l(P)\}}{\left|\|A\| + \delta K(1 + l_{PP'}^{\nabla,\alpha}) - \max\{\alpha l_{P'}^{\nabla}, (1-\alpha)l(P)\}\right|} \\
C_1 &:= \delta K \max\{1, (1-\alpha)l_P^{\nabla}\alpha l(P')\}C_{PP'}^{\alpha(1)}, \qquad (3.3.18) \\
\chi_1 &:= \delta K \max\{\|A\| + \delta K(1 + l_{PP'}^{\nabla,\alpha}), \alpha l_{P'}^{\nabla}, (1-\alpha)l(P)\} + C_1. \qquad (3.3.19)
\end{aligned}
$$

Suppose that Assumptions 6,9,10 and 14 hold with conditions $\chi_1 \in (0,1)$, $\delta \in \left[0, \frac{1-\|A\|}{(2+l_{PP'}^{\nabla,\frac{\alpha}{\delta}})K}\right)$ and $\alpha_1 < \frac{\alpha(\delta)}{\delta}$. Then there exist $a_1, a_2, a_1', a_2', a_1'', a_2'' \in (0,\infty)$ and N_0, N_1, N_2 for all $\varepsilon > 0$ such that

$$
\sup_{n\geq 0} P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq \begin{cases}
a_1 e^{-Na_2(\varepsilon^2\wedge\varepsilon)} 1_{\{d=1\}} & N \geq N_1 \max\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^2}\}, \\
a_1' e^{-Na_2'\left(\left(\frac{\varepsilon}{\log(2+\frac{1}{\varepsilon})}\right)^2\wedge\varepsilon\right)} 1_{\{d=2\}} & N \geq N_2 \max\{\frac{1}{\varepsilon}, \left(\frac{\log(2+\frac{1}{\varepsilon})}{\varepsilon}\right)^2\}, \\
a_1'' e^{-Na_2''(\varepsilon^d\wedge\varepsilon)} 1_{\{d>2\}} & N \geq N_3 \max\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^d}\}.
\end{cases}
$$

Remark 9. *1. If Assumption 14 is strengthened to $\int e^{\alpha(\delta)\left(A_1^2(z) + \frac{|B(z)|)^2}{\delta^2}\right)}\theta(dz) < \infty$ for some $\alpha(\delta) > 0$ then one can strengthen the conclusion of Theorem 3.3.5 as follows: For δ, α sufficiently small there exist $N_0, a_1, a_2 \in (0,\infty)$ and a nonincreasing function $\varsigma_2 : (0,\infty) \to (0,\infty)$ such that $\varsigma_2(t) \downarrow 0$ as $t \uparrow \infty$ and for all $\varepsilon > 0$ and $N \geq N_0\varsigma_2(\varepsilon)$*

$$
\sup_{n\geq 0} P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq a_1 e^{-Na_2\varepsilon^2}.
$$

2. *Theorem 3.3.5 treats the system for* $\mathbb{IPS}_1$. *For* $\mathbb{IPS}_2$ *we would need to estiate in addition* $\mathcal{W}_1\left(S^M(\bar{\eta}^M_{n-1}), \bar{\eta}^M_{n-1}\right)$. *However obtaining useful concentration bound for this term appears to be a hard problem and is not addressed here.*

3.4 Proofs :

The following two elementary lemmas give a basic moment bound that will be used in the proofs. We denote the function $f(\cdot,\cdot,\cdot,x) + \frac{B(x)}{\delta}$ by $f_\delta(\cdot,\cdot,\cdot,x)$.

Lemma 12. *For an interacting particle system illustrated in (3.1.3) and (3.1.5),*

1. *Suppose Assumptions 6, 7 and 9 hold. Then, for every* $n \geq 1, \quad M_n = \sup_{N\geq 1} E|X^i_n| < \infty$. *Moreover if* $a_0 > 0$ *then for* $\delta \in (0, a_0)$ *we have* $\quad \sup_{n\geq 1} M_n < \infty$.

2. *Suppose in addition Assumption 12 holds for some* $\tau > 0, a(\tau) > 0$, *and suppose* $\delta \in (0, a(\tau)^{\frac{1}{1+\tau}})$. *Then*

$$\sup_{N\geq 1}\sup_{n\geq 1} E|X^i_n|^{1+\tau} < \infty,$$

where in limit $a(\tau)^{\frac{1}{1+\tau}} \to a_0$ *as* $\tau \to 0^+$.

Remark 10. *The same bounds for* $\sup_n \sup_{N,M\geq 1} E|\bar{X}^i_{n+1}|$ *and* $\sup_n \sup_{N,M\geq 1} E|\bar{X}^i_{n+1}|^{1+\tau}$ *also hold for* $\mathbb{IPS}_2$ *under same conditions in part (a) and (b) similarly.*

Proof of Lemma 12

1. We prove the second statement. Proof of the first statement is similar. For each $n \geq 1$ and $i = 1, \ldots, N$, applying Assumption 6 on particle system in (3.1.3) with definitions

of $A_1(\cdot)$ and $A_2(\cdot)$

$$|X^i_{n+1}| \leq \|A\||X^i_n| + \delta A_1(\epsilon^i_{n+1})[|\nabla\eta^N_n(X^i_n)| + \|\mu^N_n\|_1 + |X^i_n|] + \delta A_2(\epsilon^i_{n+1}) \\ +|B(\epsilon^i_{n+1})|.$$

Now by Assumption 9 using DCT one has

$$\nabla\eta_{n+1}(y) = \int_{\mathbb{R}^d} \eta_n(x)[\nabla_y R^\alpha_{\mu_n}(x,y)]dx \tag{3.4.1}$$

for every y since from (3.3.6) $\sup_{x\in\mathbb{R}^d} |\nabla_y R^\alpha_{\mu_n}(x,y)| \leq l^{\nabla,\alpha}_{PP'}|y| + \sup_{x\in\mathbb{R}^d}\big((1-\alpha)|\nabla_y P(x,0)| + \alpha|\nabla_y P'(x,0)|\big)$. Applying the same condition followed by the inequality $|\nabla\eta_{n+1}(y)| \leq \int_{\mathbb{R}^d}\eta_n(x)|\nabla_y R^\alpha_{\mu_n}(x,y)|dx$, one has

$$|\nabla\eta_n(y)| \leq l^{\nabla,\alpha}_{PP'}|y| + c^\alpha_{PP'}. \tag{3.4.2}$$

Also note by exchangeability $E\|\mu^N_n\|_1 = E\int|x|\mu^N_n(dx) = E|X^i_n|$. Taking expectation in (3.4.1) and using (3.4.2) and independence between ϵ^i_{n+1} and $\{X^j_n\}^N_{j=1}$, one has

$$E|X^i_{n+1}| \leq \left(\|A\| + \delta\sigma\left(2 + l^{\nabla,\alpha}_{PP'}\right)\right)E|X^i_n| + \delta[\sigma c^{\nabla,\alpha}_{PP'} + \sigma_2(\delta)]. \tag{3.4.3}$$

The assumption on δ implies that $\gamma := \|A\| + \delta\sigma\left(2 + l^{\nabla,\alpha}_{PP'}\right) \in (0,1)$. A recursion on (3.4.3) will give $M_n \leq \gamma^n E|X^i_0| + \frac{\delta[\sigma c^{\nabla,\alpha}_{PP'}+\sigma_2]}{1-\gamma}$, from which the result follows.

2. By Holder's inequality for any three nonnegative real numbers a_1, a_2, a_3, a_4

$$(a_1 + a_2 + a_3 + a_4)^{1+\tau} \leq 4^\tau (a_1^{1+\tau} + a_2^{1+\tau} + a_3^{1+\tau} + a_4^{1+\tau}). \tag{3.4.4}$$

Starting with (3.4.1), applying (3.4.4), we have

$$|X_{n+1}^i|^{1+\tau} \leq 4^\tau \Bigg[\|A\|^{(1+\tau)} |X_n^i|^{1+\tau} + \Big(\delta A_1(\epsilon_{n+1}^i)[1 + l_{PP'}^{\nabla,\alpha}]|X_n^i|\Big)^{1+\tau} + \big(\delta A_1(\epsilon_{n+1}^i)\|\mu_n^N\|_1\big)^{1+\tau} + \delta^{1+\tau}\big[A_1(\epsilon_{n+1}^i).c_{PP'}^\alpha + A_2(\epsilon_{n+1}^i) + \frac{|B(\epsilon_{n+1}^i)|}{\delta}\big]^{1+\tau}\Bigg].$$

For any convex function $\phi(\cdot)$, applying Jensen's inequality one gets $\phi(\|\mu_n^N\|_1) \leq \int |\phi(x)|\mu_n^N(dx) = \frac{1}{N}\sum_{i=1}^N |\phi(X_n^i)|$. Using $\phi(x) = x^{1+\tau}$, after taking expectation one gets following recursive equation for $E|X_{n+1}^i|^{1+\tau}$,

$$E|X_{n+1}^i|^{1+\tau} \leq 4^\tau \Bigg[\|A\|^{(1+\tau)} + \delta^{1+\tau}\sigma_1(\tau)\big[(1 + l_{PP'}^{\nabla,\alpha})^{1+\tau} + 1\big]\Bigg] E|X_n^i|^{1+\tau} + \delta^{1+\tau} 8^\tau \Bigg[\sigma_1(\tau)c_{PP'}^\tau + \sigma_2(\delta,\tau)\Bigg].$$

Note that for our condition on δ, $\kappa_1 := 4^\tau \Bigg[\|A\|^{(1+\tau)} + \delta^{1+\tau}\sigma_1(\tau)\big[(1 + l_{PP'}^{\nabla,\alpha})^{1+\tau} + 1\big]\Bigg] < 1$. Thus

$$\sup_{n\geq 1} E|X_n^i|^{1+\tau} \leq \kappa_1^n E|X_0^i|^{1+\tau} + \frac{\delta^{1+\tau} 8^\tau \big[\sigma_1(\tau)c_{PP'}^\tau + \sigma_2(\delta,\tau)\big]}{1-\kappa_1}. \tag{3.4.5}$$

□

Lemma 13. *Suppose Assumptions 6,7,9 and 10 hold.*

1. *Consider the particle system $\mathbb{IPS}_1$. Then, for every $n \geq 1$,*

$$\langle |x|, \eta_n \rangle < \infty, \qquad \sup_{N\geq 1} E \left\langle |x|, \eta_n^N \right\rangle < \infty. \tag{3.4.6}$$

   Moreover if $a_0 > 0$ holds, then under conditions

$$\delta \in (0, a_0)\,, \qquad \textit{and} \qquad \sum_{i=0}^{\infty}(1-\alpha)^i \int_{\mathbb{R}^d} |y| P' P^i(0, dy) < \infty, \tag{3.4.7}$$

   one has $\sup_{n\geq 1} \langle |x|, \eta_n \rangle < \infty$.

   Additionally suppose $\sup_{N\geq 1} E \left\langle |x|, \eta_0^N \right\rangle < \infty$. Then

$$\sup_{n\geq 1} \sup_{N\geq 1} E \left\langle |x|, \eta_n^N \right\rangle < \infty.$$

2. *Suppose in addition Assumption 12,13 hold for some $\tau > 0, a(\tau) > 0$ and suppose $\delta \in (0, a(\tau)^{\frac{1}{1+\tau}})$. Then with condition $(1-\alpha) m_\tau(P) < 1$ one has $\sup_{n\geq 1} \langle |x|^{1+\tau}, \eta_n \rangle < \infty$. Additionally suppose $\sup_{N\geq 1} E \left\langle |x|^{1+\tau}, \eta_0^N \right\rangle < \infty$ Then $\sup_{n\geq 1} \sup_{N\geq 1} E \left\langle |x|^{1+\tau}, \eta_n^N \right\rangle < \infty$, where in limit $a(\tau)^{\frac{1}{1+\tau}} \to a_0$ as $\tau \to 0^+$.*

Corollary 3.4.1. *For $\mathbb{IPS}_2$ same conclusion about $\bar{\eta}_n^M$ holds as η_n^N in first particle system specified in Lemma 13 under same set of conditions on δ, α. Note that $\bar{\eta}_0^M = \eta_0$, so we don't need to assume anything about the initial sampling scheme like $\sup_{M\geq 1} E \left\langle |x|, \bar{\eta}_0^M \right\rangle < \infty$ (or $\sup_{M\geq 1} E \left\langle |x|^{1+\tau}, \bar{\eta}_0^M \right\rangle < \infty$) since they automatically hold for $\eta_0 \in \mathcal{P}_1^*(\mathbb{R}^d)$ (or $\eta_0 \in$*

$\mathcal{P}^*_{1+\tau}(\mathbb{R}^d)$) *respectively.*

Proof of Lemma 13

We will start with the second part of part (a) of the lemma. First part will follow similarly. We will show if $\eta_0 \in \mathcal{P}^*_1(\mathbb{R}^d)$ then $\eta_n \in \mathcal{P}_1(\mathbb{R}^d)$ for all $n \geq 1$. Note that

$$\eta_{k+1} = \sum_{i=0}^{k} \left[\alpha(1-\alpha)^i \mu_{k-i} P' P^i\right] + (1-\alpha)^{k+1} \eta_0 P^{k+1}. \tag{3.4.8}$$

From Assumption 10, it is obvious that $P'P^i f$ is $l(P')l(P)^i$ Lipschitz if f is a 1-Lipschitz function. It implies $|P'P^i f(x) - P'P^i f(0)| \leq l(P')l(P)^i|x|$ for any $f \in \mathrm{Lip}_1(\mathbb{R}^d)$. Since $|x|$ is 1-Lipschitz, one has

$$P'P^i|x| \leq l(P')l(P)^i|x| + \int_{\mathbb{R}^d} |y| P'P^i(0, dy).$$

Using this inequality one has from (3.4.8)

$$\begin{aligned}
&\langle |x|, \eta_{k+1}\rangle = \sum_{i=0}^{k} [\alpha(1-\alpha)^i \left\langle |x|, \mu_{k-i} P'P^i \right\rangle] + (1-\alpha)^{k+1} \left\langle |x|, \eta_0 P^{k+1} \right\rangle \\
\leq\ & \sum_{i=0}^{k} [\alpha(1-\alpha)^i \left\langle l(P')l(P)^i|x|, \mu_{k-i} \right\rangle] + \alpha \sum_{i=0}^{\infty} (1-\alpha)^i \int_{\mathbb{R}^d} |y| P'P^i(0, dy) \\
&+ [(1-\alpha)l(P)]^{k+1} \langle |x|, \eta_0 \rangle \\
\leq\ & \alpha l(P') \left\{ \sup_{n \in \mathbb{N}} \langle |x|, \mu_n \rangle \right\} \sum_{i=0}^{k} [(1-\alpha)l(P)]^i + \alpha \sum_{i=0}^{\infty} (1-\alpha)^i \int_{\mathbb{R}^d} |y| P'P^i(0, dy) \\
&+ [(1-\alpha)l(P)]^{k+1} \langle |x|, \eta_0 \rangle .
\end{aligned} \tag{3.4.9}$$

By Assumption 10, $l(P) \leq 1$, implies $(1-\alpha)l(P) < 1$. From similar derivation done in Lemma 12, one has $\sup_{n\in\mathbb{N}} \langle |x|, \mu_n \rangle < \infty$ if $\delta \in (0, a_0)$. The result follows using all the conditions

$$\sup_{k\in\mathbb{N}} \langle |x|, \eta_k \rangle < \infty.$$

For $E\left\langle |x|, \eta_k^N \right\rangle$ note that for any function f,

$$\left\langle f, \eta_{k+1}^N \right\rangle = \sum_{i=0}^{k} \left[\alpha(1-\alpha)^i \left\langle f, \mu_{k-i}^N P' P^i \right\rangle\right] + (1-\alpha)^{k+1} \left\langle f, \eta_0^N P^{k+1} \right\rangle. \tag{3.4.10}$$

From Lemma 12 $\sup_{n\geq 0} \sup_{N\geq 1} E\left\langle |x|, \mu_n^N \right\rangle < \infty$ for $\delta \in (0, a_0)$. Putting $f(x) = |x|$, then expanding $\left\langle |x|, \eta_n^N \right\rangle$ similarly like (3.4.9) after taking expectation one gets a similar bound and finiteness of $\sup_n \sup_{N\geq 1} E\left\langle |x|, \eta_n^N \right\rangle$ follows from that.

□

Proof of Lemma 13(b): From (3.4.8),

$$\left\langle \eta_{k+1}, |x|^{1+\tau} \right\rangle = \sum_{i=0}^{k} \left[\alpha(1-\alpha)^i \left\langle \mu_{k-i} P' P^i, |x|^{1+\tau} \right\rangle\right] + (1-\alpha)^{k+1} \left\langle \eta_0 P^{k+1}, |x|^{1+\tau} \right\rangle \tag{3.4.11}$$

From Assumption 13 we get the following recursion for $a_i := \langle \mu P' P^i, |x|^{1+\tau} \rangle$ for any measure $\mu \in \mathcal{P}_{1+\tau}(\mathbb{R}^d)$

$$a_i = \left\langle \mu P' P^{i-1}, P|x|^{1+\tau} \right\rangle \leq m_\tau(P)(1 + a_{i-1}) \tag{3.4.12}$$

since $P|x|^{1+\tau} \leq m_\tau(P)(1+|x|^{1+\tau})$ from Assumption 13. Using the fact $a_0 := \langle \mu, P'|x|^{1+\tau} \rangle \leq$

$m_\tau(P')(1+\langle \mu, |x|^{1+\tau}\rangle)$, we finally have

$$\begin{aligned}\left\langle \eta_{k+1}, |x|^{1+\tau}\right\rangle &\leq \alpha\sum_{i=0}^{k}(1-\alpha)^i\left[m_\tau(P)\frac{m_\tau^i(P)-1}{m_\tau(P)-1}+m_\tau(P')l_\tau^i(P)\left[1+\left\langle |x|^{1+\tau}, \mu_{k-i}\right\rangle\right]\right] \\ &+(1-\alpha)^{k+1}\left[m_\tau(P)\frac{l_\tau^{k+1}(P)-1}{m_\tau(P)-1}+m_\tau^{k+1}(P)\left\langle \eta_0, |x|^{1+\tau}\right\rangle\right]. \qquad (3.4.13)\end{aligned}$$

Under condition $\delta \in (0, a(\tau)^{\frac{1}{1+\tau}})$ and $(1-\alpha)m_\tau(P) < 1$ one gets $\sup_n \langle \eta_n, |x|^{1+\tau}\rangle < \infty$. Similarly the same bound can be derived for $\sup_n \sup_{N\geq 1} E\left\langle |x|^{1+\tau}, \eta_n^N\right\rangle$ under the same set of conditions.

$\square$

Proof of Corollary 3.4.1

To prove the Corollary about $\bar{\eta}_n^M$, define the random operator $S^M \circ P$ acting on the probability measure μ on $\mathbb{R}^d$: $\mu(S^M \circ P) = (S^M(\mu))P$. Note the following recursive form of $\bar{\eta}_n^M$:

$$\bar{\eta}_{k+1}^M = \sum_{i=0}^{k}\left[\alpha(1-\alpha)^i\bar{\mu}_{k-i}^N P'(S^M \circ P)^i\right] + (1-\alpha)^{k+1}\eta_0(S^M \circ P)^{k+1}. \qquad (3.4.14)$$

Note that for any function f one has

$$E\left\langle \mu(S^M \circ P), f\right\rangle = E\left\langle S^M(\mu), Pf\right\rangle = \langle \mu, Pf\rangle = \langle \mu P, f\rangle.$$

Now by expanding $\mu(S^M \circ P)^k$ one gets,

$$\mu(S^M \circ P)^k = \left[\mu(S^M \circ P)^{k-1}\right](S^M \circ P) = S^M(\mu(S^M \circ P)^{k-1})P.$$

Taking expectation one has

$$\begin{aligned} E\left\langle \mu(S^M \circ P)^k, f\right\rangle &= E\left\langle S^M\left(\mu(S^M \circ P)^{k-1}\right)P, f\right\rangle = E\left\langle S^M\left(\mu(S^M \circ P)^{k-1}\right), Pf\right\rangle \\ &= E\left\langle \mu(S^M \circ P)^{k-1}, Pf\right\rangle = E\left\langle \mu(S^M \circ P)^{k-1}P, f\right\rangle. \end{aligned}$$

Continuing this calculation $k-1$ times one has $E\left\langle \mu(S^M \circ P)^k, f\right\rangle = \left\langle \mu P^k, f\right\rangle$ which leads to the following expression

$$\begin{aligned} E\left\langle \bar{\mu}^N_{k-i}P'(S^M \circ P)^i, f\right\rangle &= EE\left[\left\langle \bar{\mu}^N_{k-i}P'(S^M \circ P)^i, f\right\rangle \middle| \mathcal{F}^{M,N}_{k-i}\right] \\ &= E\left[\left\langle \bar{\mu}^N_{k-i}P'P^i, f\right\rangle\right] = E\left[\left\langle \bar{\mu}^N_{k-i}, P'P^i f\right\rangle\right]. \quad (3.4.15) \end{aligned}$$

The corollary is proved by observing (3.4.15). The same bound holds for both $E\left\langle \bar{\eta}^M_n, f\right\rangle$, $E\left\langle \eta^N_n, f\right\rangle$ because of the similarity of bounds of $E\left\langle f, \mu^N_n\right\rangle$, and $E\left\langle f, \bar{\mu}^N_n\right\rangle$ for $f(x) = |x|, |x|^{1+\tau}, e^{\alpha|x|^p}$ which follows from Remark 10.

□

3.4.1 Proof of Theorem 1

We will prove part (b) of the theorem. Part (a) will follow similarly. We will start with the following lemma.

Lemma 14. *1. Under Assumptions 6,7,9, for every $\epsilon > 0$ and $n \geq 1$, there exists a compact set $K_{\epsilon,n} \in \mathcal{B}(\mathbb{R}^d)$ such that*

$$\sup_{M,N\geq 1} E\left\{\int_{K^c_{\epsilon,n}} |x|\left(\mu^N_n(dx) + \mu^N_{n-1}Q^{\bar{\eta}^M_{n-1},\mu^N_{n-1}}(dx)\right)\right\} < \epsilon.$$

2. *Suppose Assumptions 6,7,9,10,11 hold. Then for every* $\epsilon > 0$ *and* $k \geq 1$*, there exists a compact set* $K_{\epsilon,k} \in \mathcal{B}(\mathbb{R}^d)$ *such that*

$$\sup_{M,N\geq 1} E\left\langle |x|.1_{K_{k,\epsilon}}, S^M(\bar{\eta}_k^M) + \bar{\eta}_k^M \right\rangle < \varepsilon.$$

Proof: Note that for any non-negative $\phi : \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned} E\int \phi(x)\mu_n^N(dx) &= \frac{1}{N}\sum_{k=1}^{N} E\phi(X_n^k) = E\phi(X_n^1), && (3.4.16) \\ E\int \phi(x)\mu_{n-1}^N Q^{\bar{\eta}_{n-1}^M,\mu_{n-1}^N}(dx) &= \frac{1}{N}\sum_{i=1}^{N} E(E(\langle \phi, \delta_{X_n^i} Q^{\bar{\eta}_{n-1}^M,\mu_{n-1}^N}\rangle \mid \mathcal{F}_n)) \\ &= \frac{1}{N}\sum_{i=1}^{N} E\phi\left(AX_n^i + \delta f_\delta(X_n^i, \mu_n^N, \nabla\eta_n^N(X_n^i), \epsilon_{n+1}^i)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} E\phi(X_{n+1}^i) = E\phi(X_{n+1}^1). && (3.4.17) \end{aligned}$$

To get the desired result from above equalities it suffices to show that

the family $\{X_n^{i,N}, i = 1, ..., N; M, N \geq 1\}$ is uniformly integrable for every $n \geq 0$. (3.4.18)

We will prove (3.4.18) by induction on n. Once more we suppress N from the super-script. Clearly by our assumptions $\{X_0^i, i = 1, ..., N; N \geq 1\}$ is uniformly integrable. Now suppose that the Statement (3.4.18) holds for some n. Note that from (3.4.1) and (3.4.2)

$$\begin{aligned}
|X_{n+1}^i| &\leq \|A\||X_n^i| + \delta A_1(\epsilon_{n+1}^i)[|\nabla\eta_n^N(X_n^i)| + \|\mu_n^N\|_1 + |X_n^i|] + \delta A_2(\epsilon_{n+1}^i) + |B(\epsilon_{n+1}^i)|. \\
&\leq \|A\||X_n^i| + \delta A_1(\epsilon_{n+1}^i)[\|\mu_n^N\|_1 + (1 + l_{PP'}^{\nabla,\alpha})|X_n^i|] + \delta A_2(\epsilon_{n+1}^i) + |B(\epsilon_{n+1}^i)| \\
&\quad + \delta c_{PP'}^{\alpha} A_1(\epsilon_{n+1}^i) \\
&\leq \|A\||X_n^i| + \delta A_1(\epsilon_{n+1}^i)[\frac{1}{N}\sum_{i=1}^{N}|X_n^i| + (1 + l_{PP'}^{\nabla,\alpha})|X_n^i|] + \delta A_2(\epsilon_{n+1}^i) + |B(\epsilon_{n+1}^i)| \\
&\quad + \delta c_{PP'}^{\alpha} A_1(\epsilon_{n+1}^i)
\end{aligned}$$

From Assumptions 6 and 7 the families $\{A_1(\epsilon_{n+1}^i); i \geq 1\}$, $\{A_2(\epsilon_{n+1}^i); i \geq 1\}$ $\{B_2(\epsilon_{n+1}^i)$ are uniformly integrable. Now by exchangeability, $\frac{1}{N}\sum_{i=1}^N |X_n^i| = E\Big[|X_n^i|\Big|\sigma\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_n^i}\Big)\Big]$. If $\{X_\alpha : \alpha \in \Gamma_1\}$ is uniformly integrable, and $\{\sigma_\beta, \beta \in \Gamma_2\}$ is a collection of σ- fields where Γ_1, Γ_2 are arbitrary index sets, then $\{E(X_\alpha|\sigma_\beta), (\alpha, \beta) \in \Gamma_1 \times \Gamma_2\}$ is also a uniformly integrable family. It follows that $\{\frac{1}{N}\sum_{i=1}^N |X_N^i|, N \geq 1\}$ is a uniformly integrable family from induction hypothesis. Using (3.4.18) again along with independence between $\{\epsilon_{n+1}^i, i = 1, \ldots, N\}$ and $\{X_n^i : i = 1, \ldots, N; N \geq 1\}$ yield that the family $\{|X_{n+1}^i| : i = 1, \ldots, N; N \geq 1\}$ is uniformly integrable. The result follows. $\square$

Proof of Lemma 14(b): Note that $S^M(\bar\eta_k^M) = \frac{1}{M}\sum_{i=1}^M \delta_{Y_k^{i,M}}$ where $\{Y_k^{i,M}\}_{i=1}^M\Big|\mathcal{F}_k^{M,N}$ are i.i.d from $\bar\eta_k^M$. So for any non-negative function ϕ we have

$$\begin{aligned}
E\langle\phi, S^M(\bar\eta_k^M)\rangle &= E\frac{1}{M}\sum_{i=1}^M \phi(Y_k^{i,M}) = EE\left[\frac{1}{M}\sum_{i=1}^M \phi(Y_k^{i,M})\big|\mathcal{F}_k^{M,N}\right] \\
&= EE\left[\phi(Y_k^{i,M})\big|\mathcal{F}_k^{M,N}\right] = E\phi(Y_k^{i,M}) = E\langle\phi, \bar\eta_k^M\rangle. \qquad (3.4.19)
\end{aligned}$$

We will prove the result if we can show the family

$$\{Y_k^{i,M}, i = 1, \ldots, M; M, N \geq 1\} \quad \text{is uniformly integrable for every} \quad k \geq 0. \tag{3.4.20}$$

We will prove (3.4.20) through induction on k. For $k = 0$, the result follows trivially since $\{Y_0^{i,M}, i = 1, \ldots, M; M \geq 1\}$ are i.i.d from η_0. Suppose it holds for $k = n$. We will show that both,

$$\{S^M(\bar{\eta}_n^M)P : M, N \geq 1\} \quad \text{and} \quad \{\bar{\mu}_n^N P' : N \geq 1\} \quad \text{are uniformly integrable}$$
$$\text{families of probability measures.} \tag{3.4.21}$$

Then from the structure $\bar{\eta}_{n+1}^M = (1-\alpha)S^M(\bar{\eta}_n^M)P + \alpha\bar{\mu}_n^N P$, it is evident that $\{\bar{\eta}_{n+1}^M : M, N \geq 1\}$ is uniform integrable which equivalently implies $\{Y_{n+1}^{i,M} : i = 1, \ldots, M; M, N \geq 1\}$ is UI too. On proving the first assertion in (3.4.21), note that due to the exchangeability of $\{Y_n^{i,M} : i = 1, \ldots, M\}$, one has

$$S^M(\bar{\eta}_n^M)P = E\left[\delta_{Y_n^{1,M}} P \middle| \sigma\left(\frac{1}{M}\sum_{i=1}^{M} \delta_{Y_n^{i,M}}\right)\right]. \tag{3.4.22}$$

We know that if $\{Z_\alpha, \alpha \in \Gamma_1\}$ is a uniformly integrable family and $\{\mathcal{H}_\beta, \beta \in \Gamma_2\}$ is a collection of σ-fields where Γ_1, Γ_2 are arbitrary index sets, then $\{E(Z_\alpha \mid \mathcal{H}_\beta), (\alpha, \beta) \in \Gamma_1 \times \Gamma_2\}$ is a uniformly integrable family. So from (3.4.22) it suffices to prove that $\{\delta_{Y_n^{i,M}} P : i = 1, \ldots, M; M, N \geq 1\}$ is uniformly integrable. Define a function $f_k(.)$ such that, $f_k(x) = 0$, if $|x| \in [0, \frac{k}{2}]$ and $f_k(x) = |x|$, if $|x| \geq k$ and linear in between range. Then by construction

$f_k(.)$ is Lipschitz with coefficient 2 and $x.1_{\{|x|>k\}} \leq f_k(x)$ for all $x \in \mathbb{R}^d$. By Assumption 11 we have that $\{P(z,.) : z \in K\}$ is uniformly integrable. So taking the compact set $K = \{|x| \leq k\}$ assuming $Y_n^{i,M}$ has unconditional law m_i^n for all $i = 1, \ldots, M$, the quantity

$$\begin{aligned} \int_{|z|>L} \int y.1_{\{K^c\}} P(z,dy) m_i^n(dz) &\leq \int_{|z|>L} \left[f_k(y) P(z,dy)\right] m_i^n(dz) \\ &\leq \int_{|z|>L} \left[|Pf_k(0)| + 2l(P)|z|\right] m_i^n(dz) \qquad (3.4.23) \\ \leq Pf_k(0) \int_{|z|>L} m_i^n(dz) &+ 2l(P) \int_{|z|>L} |z| m_i^n(dz). \qquad (3.4.24) \end{aligned}$$

The display in (3.4.23) follows from Assumption 10 and using Lipschitz property of f_k. After taking supremum in the set $\{i = 1, \ldots, M; M, N \geq 1\}$ in both sides of (3.4.24), second part of R.H.S goes to 0, as $L \to \infty$ by induction hypothesis. About the first part $Pf_k(0)$ goes to 0 as $k \to \infty$ by D.C.T since ($\int |y| P(0,dy) < \infty$) and also $\int_{|z|>L} m_i^n(dz)$ converges to 0 (as L goes to ∞) due to the tightness of $\{m_i^n : i = 1, \ldots, M; M, N \geq 1\}$ which also follows from induction hypothesis. The second assertion that $\{\bar{\mu}_n^N P' : N \geq 1\}$ is uniformly integrable follows similarly through induction.

□

We will proceed to the main proof via induction on $n \in \mathbb{N}$ for the quantity

$$E\left[\mathcal{W}_1(\bar{\mu}_n^N, \mu_n) + \mathcal{W}_1(\bar{\eta}_n^N, \eta_n)\right]$$

. For $n = 0$, we will first show that $E\mathcal{W}_1(\bar{\mu}_0^N, \mu_0) \to 0$ as $N \to \infty$. From [74] we have

$$
\begin{aligned}
(\bar{X}_0^1, \bar{X}_0^2, \ldots, \bar{X}_0^N) \text{ is } \mu_0\text{-chaotic} \quad &\Leftrightarrow \quad \bar{\mu}_0^N \text{ converges weakly to } \mu_0 \text{ in probability} \\
&\Leftrightarrow \quad \beta(\bar{\mu}_0^N, \mu_0) \xrightarrow{p} 0. \qquad (3.4.25)
\end{aligned}
$$

From Lemma 14 one can construct $K_{0,\epsilon}$ compact ball containing 0, so that $E\left\langle |x|.1_{K_{0,\epsilon}^c}, \bar{\mu}_0^N \right\rangle < \frac{\varepsilon}{2}$ and $\left\langle |x|.1_{K_{0,\epsilon}^c}, \mu_0 \right\rangle < \frac{\varepsilon}{2}$ hold. So using the fact for any $f \in \mathrm{Lip}_1(\mathbb{R}^d)$ with $f(0) = 0$, one has $|f(x)| \le |x|$.

$$
\begin{aligned}
E\mathcal{W}_1(\bar{\mu}_0^N, \mu_0) &= E \sup_{f \in \mathrm{Lip}_1(\mathbb{R}^d)} |\left\langle f, \bar{\mu}_0^N - \mu_0 \right\rangle| = E \sup_{f \in \mathrm{Lip}_1(\mathbb{R}^d), f(0)=0} |\left\langle f, \bar{\mu}_0^N - \mu_0 \right\rangle| \\
\le \quad & E \sup_{f \in \mathrm{Lip}_1(\mathbb{R}^d), f(0)=0} |\left\langle f 1_{K_{0,\epsilon}}, \bar{\mu}_0^N - \mu_0 \right\rangle| + E\left\langle |x| 1_{K_{0,\epsilon}^c}, \bar{\mu}_0^N \right\rangle + \left\langle |x|.1_{K_{0,\epsilon}^c}, \mu_0 \right\rangle \\
\le \quad & \mathrm{diam}(K_{0,\varepsilon}) E\beta(\bar{\mu}_0^N, \mu_0) + \varepsilon. \qquad (3.4.26)
\end{aligned}
$$

In last display we used the fact that $\sup_{x \in K_{0,\varepsilon}} |f(x)| \le \mathrm{diam}(K_{0,\varepsilon})$. Note that $\beta(\bar{\mu}_0^N, \mu_0)$ is bounded by 2 (so Uniformly Integrable) and $\beta(\bar{\mu}_0^N, \mu_0) \xrightarrow{p} 0$ implies $E\beta(\mu_0^N, \mu_0) \to 0$ as $N \to \infty$ proving the assertion (3.3.12) for $n = 0$. Suppose it holds for $n \le k$. We start with the following triangular inequality

$$
\begin{aligned}
\mathcal{W}_1(\bar{\mu}_{k+1}^N, \mu_{k+1}) \quad \le \quad & \mathcal{W}_1(\bar{\mu}_{k+1}^N, \bar{\mu}_k^N Q^{\bar{\eta}_k^N, \bar{\mu}_k^N}) + \mathcal{W}_1(\bar{\mu}_k^N Q^{\bar{\eta}_k^N, \bar{\mu}_k^N}, \bar{\mu}_k^N Q^{\eta_k, \bar{\mu}_k^N}) \\
& + \mathcal{W}_1(\bar{\mu}_k^N Q^{\eta_k, \bar{\mu}_k^N}, \mu_{k+1}). \qquad (3.4.27)
\end{aligned}
$$

Consider the third term of (3.4.27). From the general calculations follwed by (3.4.45)-

(3.4.47), we have the following estimate,

$$\mathcal{W}_1(\bar{\mu}_k^N Q^{\eta_k,\bar{\mu}_k^N}, \mu_k Q^{\eta_k,\mu_k}) \leq \left(\|A\| + \delta\sigma(2 + l_{PP'}^{\nabla,\alpha})\right) \mathcal{W}_1(\bar{\mu}_k^N, \mu_k). \tag{3.4.28}$$

Now we consider the first term of the right hand side of (3.4.27). We will use Lemma 14(a). Fix $\epsilon > 0$ and let K_ϵ be a compact set in $\mathbb{R}^d$ such that

$$\sup_{N\geq 1} E\left\{\int_{K_\epsilon^c} |x|(\bar{\mu}_{k+1}^N(dx) + \bar{\mu}_k^N Q^{\bar{\eta}_k^N,\bar{\mu}_k^N}(dx))\right\} < \epsilon.$$

Let $\text{Lip}_1^0(\mathbb{R}^d) := \{f \in \text{Lip}_1(\mathbb{R}^d) : f(0) = 0\}$. Then,

$$\begin{aligned} E \sup_{\phi\in \text{Lip}_1(\mathbb{R}^d)} |\langle \phi, \bar{\mu}_{k+1}^N - \bar{\mu}_k^N Q^{\bar{\eta}_k^N,\bar{\mu}_k^N}\rangle| &= E \sup_{\phi\in \text{Lip}_1^0(\mathbb{R}^d)} |\langle \phi, \bar{\mu}_{k+1}^N - \bar{\mu}_k^N Q^{\bar{\eta}_k^N,\bar{\mu}_k^N}\rangle| \\ &\leq E \sup_{\phi\in \text{Lip}_1^0(\mathbb{R}^d)} |\langle \phi.1_{K_\epsilon}, \bar{\mu}_{k+1}^N - \bar{\mu}_k^N Q^{\bar{\eta}_k^N,\bar{\mu}_k^N}\rangle| + \epsilon. \end{aligned} \tag{3.4.29}$$

We will now apply Lemma A.1.1 in the Appendix. Note that for any $\phi \in \text{Lip}_1^0(\mathbb{R}^d)$,

$$\sup_{x\in K_\epsilon} |\phi(x)| \leq diam(K_\epsilon) := m_\epsilon.$$

Thus with notation as in Lemma A.1.1

$$\begin{aligned} \sup_{\phi\in \text{Lip}_1^0(\mathbb{R}^d)} |\langle \phi.1_{K_\epsilon}, \bar{\mu}_{k+1}^N - \bar{\mu}_k^N Q^{\bar{\eta}_k^N,\bar{\mu}_k^N}\rangle| \quad &\leq \quad \max_{\phi\in \mathcal{F}_{m_\epsilon,1}^\epsilon(K_\epsilon)} |\langle \phi, \bar{\mu}_{k+1}^N - \bar{\mu}_k^N Q^{\bar{\eta}_k^N,\bar{\mu}_k^N}\rangle| \\ &\quad +2\epsilon. \end{aligned} \tag{3.4.30}$$

where we have denoted the restrictions of $\bar{\mu}_{k+1}^N$ and $\bar{\mu}_k^N Q^{\bar{\eta}_k^N}$ to K_ϵ by the same symbols. Using the above inequality in (3.4.29), we obtain

$$E\mathcal{W}_1(\bar{\mu}_{k+1}^N, \bar{\mu}_k^N Q^{\bar{\eta}_k^N, \bar{\mu}_k^N}) \le \sum_{\phi \in \mathcal{F}^\epsilon_{m_\epsilon,1}(K_\epsilon)} E|\langle \phi, \bar{\mu}_{k+1}^N - \bar{\mu}_k^N Q^{\bar{\eta}_k^N, \bar{\mu}_k^N} \rangle| + 3\epsilon. \quad (3.4.31)$$

Using Lemma A.1.2 we see that the first term on the right hand side can be bounded by $\frac{2m_\epsilon |\mathcal{F}^\epsilon_{m_\epsilon,1}(K_\epsilon)|}{\sqrt{N}}$.

Consider the second term of R.H.S of (3.4.27). From Assumption 9 applying DCT one has

$$\begin{aligned}
\nabla \bar{\eta}_k^N(y) &= (1-\alpha)\int S^M(\bar{\eta}_{k-1}^N)(dx)\nabla_y P(x,y) + \alpha \int \bar{\mu}_k^N(dx)\nabla_y P'(x,y), & (3.4.32)\\
\nabla \eta_k(y) &= (1-\alpha)\int \eta_{k-1}(dx)\nabla + \alpha\int \mu_k(dx)\nabla_y P'(x,y). & (3.4.33)
\end{aligned}$$

Suppose $\bar{X}_k$ is a random variable conditioned on $\mathcal{F}_k^{M,N}$ is distributed with law $\bar{\mu}_k^N$. Then almost surely $\mathcal{W}_1(\bar{\mu}_k^N Q^{\bar{\eta}_k^N, \bar{\mu}_k^N}, \bar{\mu}_k^N Q^{\eta_k, \bar{\mu}_k^N})$ is

$$\begin{aligned}
&\le \sup_{g \in \mathrm{Lip}_1(\mathbb{R}^d)} E_k^{M,N}\Big[\Big| g(A\bar{X}_k + \delta f_\delta(\nabla\bar{\eta}_k^N(\bar{X}_k), \bar{\mu}_k^N, \bar{X}_k, \epsilon)) \\
&\quad - g(A\bar{X}_k + \delta f_\delta(\nabla\eta_k(\bar{X}_k), \bar{\mu}_k^N, \bar{X}_k, \epsilon))\Big|\Big] \le \delta\sigma E_k^{M,N}\left[\left|\nabla\bar{\eta}_k^N(\bar{X}_k) - \nabla\eta_k(\bar{X}_k)\right|\right] \\
&\le \delta\sigma(1-\alpha)\int \left|\int \left\{S^M(\bar{\eta}_k^M) - \eta_k\right\}(dx).\nabla_y P(x,y)\right| \bar{\mu}_k^N(dy) \\
&\quad + \delta\sigma\alpha \int\left|\int \left\{\bar{\mu}_k^N - \mu_k\right\}(dx).\nabla_y P'(x,y)\right|\bar{\mu}_k^N(dy) \\
&\le \delta\sigma(1-\alpha) l_P^\nabla \mathcal{W}_1(S^M(\bar{\eta}_k^M), \eta_k) + \delta\sigma\alpha l_{P'}^\nabla \mathcal{W}_1(\bar{\mu}_k^N, \mu_k). \qquad (3.4.34)
\end{aligned}$$

(3.4.34) follows by using Assumption 9. About the first term in (3.4.34) note that from triangular inequality,

$$E\mathcal{W}_1(S^M(\bar{\eta}_k^M), \eta_k) \leq E\mathcal{W}_1(S^M(\bar{\eta}_k^M), \bar{\eta}_k^M) + E\mathcal{W}_1(\bar{\eta}_k^M, \eta_k). \tag{3.4.35}$$

The first term in (3.4.35) can be written as

$$\begin{aligned} E\mathcal{W}_1(S^M(\bar{\eta}_k^M), \bar{\eta}_k^M) \leq E \sup_{f \in \mathrm{Lip}_1^0(\mathbb{R}^d)} &| \left\langle f.1_{K_{k,\epsilon}}, S^M(\bar{\eta}_k^M) - \bar{\eta}_k^M \right\rangle | \\ &+ E \left\langle |x|.1_{K_{k,\epsilon}^c}, S^M(\bar{\eta}_k^M) \right\rangle + E \left\langle |x|.1_{K_{k,\epsilon}^c}, \bar{\eta}_k^M \right\rangle. \end{aligned} \tag{3.4.36}$$

By Lemma 14(b), for a specified $\varepsilon > 0$, one can construct a compact set $K_{k,\epsilon}$ containing 0 such that,

$$\sup_{M,N \geq 1} E \left\langle |x|.1_{K_{k,\epsilon}}, S^M(\bar{\eta}_k^M) + \bar{\eta}_k^M \right\rangle < \varepsilon.$$

Denote $m_{k,\epsilon} = \text{diam}(K_{k,\epsilon})$. Using Lemma A.1.1 we have the L.H.S of (3.4.36)

$$\begin{aligned} EE_k^{M,N}\left[\sup_{\phi \in \mathrm{Lip}_1^0(\mathbb{R}^d)} | \left\langle \phi.1_{K_{k,\epsilon}}, S^M(\bar{\eta}_k^M) - \bar{\eta}_k^M \right\rangle | \right] + \varepsilon \leq \\ EE_k^{M,N}\left[\max_{\phi \in \mathcal{F}^{\epsilon}_{m_{k,\epsilon},1}(K_{k,\epsilon})} | \left\langle \phi, S^M(\bar{\eta}_k^M) - \bar{\eta}_k^M \right\rangle | \right] + 2\varepsilon \end{aligned}$$

where (3.4.36) follows from similar arguments used in (3.4.31). Note that the Lemma 14 also suggests the compact set $K_{k,\varepsilon}$ is non-random, which only depends on k and ε only. So from the display above we have

$$EE_k^{M,N}\left[\sum_{\phi\in\mathcal{F}^{\epsilon}_{m_{k,\epsilon,1}}(K_{k,\epsilon})}|\left\langle\phi,S^M(\bar{\eta}_k^M)-\bar{\eta}_k^M\right\rangle|\right]+2\varepsilon \quad \le$$

$$\sum_{\phi\in\mathcal{F}^{\epsilon}_{m_{k,\epsilon,1}}(K_{k,\epsilon})}E|\left\langle\phi,S^M(\bar{\eta}_k^M)-\bar{\eta}_k^M\right\rangle| \quad + \quad 2\varepsilon \tag{3.4.37}$$

Using Lemma A.1.2 we get the final bound of the first term in RHS of (3.4.37) as $\frac{2m_{k,\epsilon}|\mathcal{F}^{\epsilon}_{m_{k,\epsilon,1}}(K_{k,\epsilon})|}{\sqrt{M}}$.

Combining this estimate with (3.4.28),(3.4.31) and (3.4.34) we now have

$$\begin{aligned} E\mathcal{W}_1(\bar{\mu}_{k+1}^N,\mu_{k+1}) \quad &\le \quad \left(\|A\|+\delta\sigma(2+l^{\nabla}_{PP'})+\delta\sigma\alpha l^{\nabla}_{P'}\right)E\mathcal{W}_1(\bar{\mu}_k^N,\mu_k) \\ +\delta\sigma(1-\alpha)l^{\nabla}_P E\mathcal{W}_1(\bar{\eta}_k^M,\eta_k) \quad &+ \quad \frac{2\delta\sigma(1-\alpha)l^{\nabla}_P m_{k,\epsilon}|\mathcal{F}^{\epsilon}_{m_{k,\epsilon,1}}(K_{k,\epsilon})|}{\sqrt{M}}+\frac{2m_{\epsilon}|\mathcal{F}^{\epsilon}_{m_{\epsilon,1}}(K_{\epsilon})|}{\sqrt{N}} \\ &+\left(3+2\delta\sigma(1-\alpha)l^{\nabla}_P\right)\varepsilon. \end{aligned} \tag{3.4.38}$$

For the term $E\mathcal{W}_1(\bar{\eta}_{k+1}^M,\eta_{k+1})$, we start with the following recursive form

$$\begin{aligned} \bar{\eta}_{k+1}^M-\eta_{k+1}=(1-\alpha)\left[S^M(\bar{\eta}_k^M)-\bar{\eta}_k^M\right]P+(1-\alpha)\left[\bar{\eta}_k^M-\eta_k\right]P \\ +\alpha\left[\bar{\mu}_k^N-\mu_k\right]P' \end{aligned} \tag{3.4.39}$$

which leads to the following inequality

$$\begin{aligned} \mathcal{W}_1(\bar{\eta}_{k+1}^M,\eta_{k+1})\le(1-\alpha)l(P)\mathcal{W}_1(S^M(\bar{\eta}_k^M),\bar{\eta}_k^M)+(1-\alpha)l(P)\mathcal{W}_1(\bar{\eta}_k^M,\eta_k) \\ +\alpha l(P')\mathcal{W}_1(\bar{\mu}_k^N,\mu_k). \end{aligned} \tag{3.4.40}$$

Using earlier estimates one has the final estimate for

$$\begin{aligned} E\mathcal{W}_1(\bar{\eta}^M_{k+1}, \eta_{k+1}) \;\; &\leq \;\; 2(1-\alpha)l(P)\frac{m_{k,\epsilon}|\mathcal{F}^{\epsilon}_{m_{k,\epsilon,1}}(K_{k,\epsilon})|}{\sqrt{M}} + (1-\alpha)l(P)\mathcal{W}_1(\bar{\eta}^M_k, \eta_k) \\ &\quad +\alpha l(P')\mathcal{W}_1(\bar{\mu}^N_k, \mu_k) + 2(1-\alpha)l(P)\varepsilon. \end{aligned} \tag{3.4.41}$$

Adding (3.4.38) and (3.4.41), using induction hypothesis and sending $M, N \to \infty$ we have

$$E\mathcal{W}_1(\bar{\mu}^N_{k+1}, \mu_{k+1}) + E\mathcal{W}_1(\bar{\eta}^M_{k+1}, \eta_{k+1}) \leq \left(3 + 2\delta\sigma(1-\alpha)l^{\nabla}_P + 2(1-\alpha)l(P)\right)\varepsilon.$$

Since $\varepsilon > 0$ arbitrary, the result follows.

Part (a) can be proved similarly. The change will come from the structural difference of $\bar{\eta}^N_k$ and η^N_k because of the change in the updating kernel. So the term coming from the quantity $S^M(\bar{\eta}^M_k) - \bar{\eta}^M_k$ will not appear here. Hence we get the following final estimate

$$\begin{aligned} E\left[\mathcal{W}_1(\mu^N_{k+1}, \mu_{k+1}) + \mathcal{W}_1(\eta^N_{k+1}, \eta_{k+1})\right] \leq \Big[&\|A\| + \delta\sigma(2 + l^{\nabla}_{PP'}) + \delta\sigma\alpha l^{\nabla}_{P'} \\ +\alpha l(P')\Big] E\mathcal{W}_1(\mu^N_k, \mu_k) &+ \left[\delta\sigma(1-\alpha)l^{\nabla}_P + (1-\alpha)l(P)\right] E\mathcal{W}_1(\eta^M_k, \eta_k) \\ &+3\varepsilon + \frac{2m_\epsilon|\mathcal{F}^{\epsilon}_{m_{\epsilon,1}}(K_\epsilon)|}{\sqrt{N}} \end{aligned}$$

from which the result follows by induction.

□

3.4.2 Proof of Theorem 3.3.1

We will start with the following lemma and then prove the Theorem 3.3.1 using it.

Lemma 15. *Let* $\mu_0, \mu_0' \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\eta_0, \eta_0' \in \mathcal{P}_1^*(\mathbb{R}^d)$. *Suppose Assumptions 6,7, 9 and 10 hold. Then the transformation* $\Psi : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d) \to \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$ *is well defined if following hold*

$$\delta < a_0 \qquad \textit{and} \qquad \sum_{i=0}^{\infty}(1-\alpha)^i \int_{\mathbb{R}^d} |y| P' P^i(0, dy) < \infty. \tag{3.4.42}$$

Moreover if Assumptions 9,8 and 10 hold along with the following condition:

$$\max\left\{\Big(\|A\| + \delta\sigma(2 + l_{PP'}^{\nabla,\alpha}) + \alpha l(P')\Big), (1-\alpha)l(P)\right\} \\ +\delta\sigma \max\left\{\alpha l_{P'}^{\nabla}, (1-\alpha)l_P^{\nabla}\right\} < 1, \tag{3.4.43}$$

then there exists a $\theta \in (0,1)$ *and a constant* $a_1 \in (0,\infty)$ *such that for any* $n \in \mathbb{N}$,

$$\mathcal{W}_1\big(\Psi^n(\mu_0, \eta_0), \Psi^n(\mu_0', \eta_0')\big) \le a_1\theta^n.$$

Remark 11. *Condition (3.4.43) implies the first condition of (3.4.42).*

Proof of Lemma 19

For fixed $\mu_0, \mu_0' \in \mathcal{P}_1(\mathbb{R}^d)$ and $\eta_0, \eta_0' \in \mathcal{P}_1^*(\mathbb{R}^d)$ define the following quantities for $n \ge 1$

$$(\mu_n, \eta_n) = \Psi^n(\mu_0, \eta_0), \qquad (\mu_n', \eta_n') = \Psi^n(\mu_0', \eta_0') \quad \text{and} \quad \Psi^0 = I.$$

First we will show that under transformation Ψ the $(\mu_n, \nu_n) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$ for $(\mu_0, \nu_0) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1^*(\mathbb{R}^d)$, so that the quantity $\mathcal{W}_1(\mu_n, \mu_n') + \mathcal{W}_1(\nu_n, \nu_n')$ is well defined. Note that , if $\delta \in (0, a_0)$, then $\gamma = \|A\| + \delta\sigma \left(2 + l_{PP'}^{\nabla,\alpha}\right) \in (0,1)$, implying

$$\langle |x|, \mu_n \rangle \leq \gamma^n \langle |x|, \mu_0 \rangle + \frac{\delta[\sigma c_{PP'}^{\nabla,\alpha} + \sigma_2]}{1-\gamma},$$

which follows similarly from the proof of Lemma 12(a). It means if $\delta \in (0, a_0)$ and $\langle |x|, \mu_0 \rangle < \infty$ hold, then $\mu_n \in \mathcal{P}_1(\mathbb{R}^d)$ for all $n \geq 1$. Under conditions in (3.4.42) one also has $\sup_{n>0} \langle |x|, \eta_n \rangle < \infty$ for all $n \in \mathbb{N}$. One has $\nabla \eta_{n+1}(y) = \int_{\mathbb{R}^d} \eta_n(x)[\nabla_y R_{\mu_n}^{\alpha}(x,y)]dx$ by Assumption 9 using DCT. From that condition it follows that for any $n \geq 1$, $\|\nabla \eta_n(\cdot)\|_1 < (1-\alpha) l_P^{\nabla} + \alpha l_{P'}^{\nabla} = l_{PP'}^{\nabla,\alpha} < \infty$ showing $\eta_n \in \mathcal{P}_1^*(\mathbb{R}^d)$ for all $n > 0$ if $\eta_0 \in \mathcal{P}_1^*(\mathbb{R}^d)$.

Now we will go back to the proof of the second part of the lemma regarding the contraction part. Assume $n \geq 2$. The first term of $\mathcal{W}_1((\mu_n, \eta_n), (\mu_n', \eta_n'))$ can be expressed as

$$\begin{aligned} \mathcal{W}_1(\mu_n, \mu_n') = \mathcal{W}_1(\mu_{n-1} Q^{\eta_{n-1}, \mu_{n-1}}, \mu_{n-1}' Q^{\eta_{n-1}', \mu_{n-1}'}) &\leq \mathcal{W}_1(\mu_{n-1} Q^{\eta_{n-1}, \mu_{n-1}}, \mu_{n-1}' Q^{\eta_{n-1}, \mu_{n-1}'}) \\ &+ \mathcal{W}_1(\mu_{n-1}' Q^{\eta_{n-1}, \mu_{n-1}'}, \mu_{n-1}' Q^{\eta_{n-1}', \mu_{n-1}'}) \\ &=: T_1 + T_2. \end{aligned} \tag{3.4.44}$$

$$\begin{aligned} T_1 &= \mathcal{W}_1(\mu_{n-1} Q^{\eta_{n-1}, \mu_{n-1}}, \mu_{n-1}' Q^{\eta_{n-1}, \mu_{n-1}'}) \leq \inf_{\{X,Y: \mathcal{L}(X,Y) = (\mu_{n-1}, \mu_{n-1}'), X, Y \perp \epsilon\}} E\big|A(X-Y) \\ &\quad + \delta[f_\delta(\nabla \eta_{n-1}(X), \mu_{n-1}, X, \epsilon) - f_\delta(\nabla \eta_{n-1}(Y), \mu_{n-1}', Y, \epsilon)]\big| \\ &\leq \inf_{\{X \sim \mu_{n-1}, Y \sim \mu_{n-1}'\}} \left\{ (\|A\| + \delta\sigma) E|X-Y| + \delta\sigma E|\nabla \eta_{n-1}(X) - \nabla \eta_{n-1}(Y)| \right\} \\ &\quad + \delta\sigma \mathcal{W}_1(\mu_{n-1}, \mu_{n-1}') \end{aligned} \tag{3.4.45}$$

The last inequality (3.4.45) follows from Assumption 6. As a consequence of Assumption 9 from (3.4.1) it follows that

$$\begin{aligned}
|\nabla\eta_{n+1}(X)-\nabla\eta_{n+1}(Y)| &\leq \int_{\mathbb{R}^d}\eta_n(x)|\nabla_y R^{\alpha}_{\mu_n}(x,X)-\nabla_y R^{\alpha}_{\mu_n}(x,Y)|dx \\
&\leq (1-\alpha)\int_{\mathbb{R}^d}\eta_n(x)|\nabla_y P(x,X)-\nabla_y P(x,Y)|dx \\
&\quad +\alpha|\nabla_y\mu_n P'(X)-\nabla_y\mu_n P'(Y)| \\
&\leq l^{\nabla,\alpha}_{PP'}|X-Y|. \qquad (3.4.46)
\end{aligned}$$

With that estimate, taking infimum at R.H.S of (3.4.46) with all possible couplings of (X,Y) with marginals respectively μ_{n-1} and μ'_{n-1}, one gets

$$T_1=\mathcal{W}_1\big(\mu_{n-1}Q^{\eta_{n-1},\mu_{n-1}},\mu'_{n-1}Q^{\eta_{n-1},\mu'_{n-1}}\big)\leq\big(\|A\|+\delta\sigma(2+l^{\nabla,\alpha}_{PP'})\big)\mathcal{W}_1(\mu_{n-1},\mu'_{n-1}). \qquad (3.4.47)$$

Let X be a $\mathbb{R}^d$ valued random variable with law μ'_{n-1}. Now about the term T_2,

$$\begin{aligned}
T_2 &= \mathcal{W}_1\big(\mu'_{n-1}Q^{\eta_{n-1},\mu'_{n-1}},\mu'_{n-1}Q^{\eta'_{n-1},\mu'_{n-1}}\big) \\
&\leq \sup_{g\in\mathbf{Lip}_1(\mathbb{R}^d)} E\Big| g(AX+\delta f_\delta(\nabla\eta_{n-1}(X),\mu'_{n-1},X,\epsilon))- \\
&\quad g(AX+\delta f_\delta(\nabla\eta'_{n-1}(X),\mu'_{n-1},X,\epsilon))\Big| \\
&\leq \delta\sigma E\left|\nabla\eta_{n-1}(X)-\nabla\eta'_{n-1}(X)\right|.
\end{aligned}$$

Now expanding $\left|\nabla\eta_{n-1}(X)-\nabla\eta'_{n-1}(X)\right|$ from (3.4.1)

$$
\begin{aligned}
T_2 &\leq \delta\sigma E\left|\int_{\mathbb{R}^d}\eta_{n-2}(x)(\nabla_y R^{\alpha}_{\mu_{n-2}}(x,X))dx-\int_{\mathbb{R}^d}\eta'_{n-2}(x)(\nabla_y R^{\alpha}_{\mu'_{n-2}}(x,X))dx\right| \\
&\leq \alpha\delta\sigma\int_{\mathbb{R}^d}\eta_{n-2}(x)E\left|\nabla_y\mu_{n-2}P'(X)-\nabla_y\mu'_{n-2}P'(X)\right|dx \\
&\quad +(1-\alpha)\delta\sigma E\left|\int_{\mathbb{R}^d}\nabla_y P(x,X)(\eta_{n-2}(x)-\eta'_{n-2}(x))dx\right| \\
&=: T_2^{(1)}+T_2^{(2)} \qquad (3.4.48)
\end{aligned}
$$

Note that

$$
\begin{aligned}
T_2^{(1)} := \alpha\delta\sigma\int_{\mathbb{R}^d}\eta_{n-2}(x)\int_{\mathbb{R}^d}\mu'_{n-1}(dz)\Bigg|\int_{\mathbb{R}^d}\Bigg(&\mu_{n-2}(dy)\nabla_y P'(y,z)-] \\
&\mu'_{n-2}(dy)\nabla_y P'(y,z)\Bigg)\Bigg|dx \qquad (3.4.49)
\end{aligned}
$$

Since from Assumption 9 $\nabla_y P'(\tilde{x},x)$ is a Lipschitz function with coefficient $l^{\nabla}_{P'}$, the first integrand in (3.4.49) will be bounded by $l^{\nabla}_{P'}.\mathcal{W}_1(\mu_{n-2},\mu'_{n-2})$ which gives

$$
T_2^{(1)} \leq \alpha\delta\sigma l^{\nabla}_{P'}\mathcal{W}_1(\mu_{n-2},\mu'_{n-2}). \qquad (3.4.50)
$$

Now using Assumption 8 the second term $T_2^{(2)}$ gives similarly

$$
\begin{aligned}
T_2^{(2)} &= (1-\alpha)\delta\sigma E\left|\int_{\mathbb{R}^d}\nabla_y P(x,X)(\eta_{n-2}(x)-\eta'_{n-2}(x))dx\right| \\
&\leq (1-\alpha)\delta\sigma\int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\nabla_y P(x,y)\{\eta_{n-2}(x)-\eta'_{n-2}(x)\}dx\right|\mu'_{n-1}(dy) \\
&\leq (1-\alpha)\delta\sigma l^{\nabla}_{P}\mathcal{W}_1(\eta_{n-2},\eta'_{n-2}). \qquad (3.4.51)
\end{aligned}
$$

Using the Assumption 10 we have

$$\mathcal{W}_1(\eta_n, \eta'_n) \leq (1-\alpha)l(P)\mathcal{W}_1(\eta_{n-1}, \eta'_{n-1}) + \alpha l(P')\mathcal{W}_1(\mu_{n-1}, \mu'_{n-1}) \tag{3.4.52}$$

Combining (3.4.50),(3.4.51) and (3.4.52) we have the following recursion for $n \geq 2$,

$$\begin{aligned} \mathcal{W}_1(\mu_n, \mu'_n) \quad + \mathcal{W}_1(\eta_n, \eta'_n) \quad &\leq \quad \big(\|A\| + \delta\sigma(2 + l^{\nabla,\alpha}_{PP'})\big)\mathcal{W}_1(\mu_{n-1}, \mu'_{n-1}) \\ +\alpha\delta\sigma l^{\nabla}_{P'}\mathcal{W}_1(\mu_{n-2}, \mu'_{n-2}) \quad &+ \quad \alpha l(P').\mathcal{W}_1(\mu_{n-1}, \mu'_{n-1}) \quad + (1-\alpha)\delta\sigma l^{\nabla}_{P}\mathcal{W}_1(\eta_{n-2}, \eta'_{n-2}) \\ &+ \quad (1-\alpha)l(P)\mathcal{W}_1(\eta_{n-1}, \eta'_{n-1}). \end{aligned} \tag{3.4.53}$$

Define a sequence $a_n := \mathcal{W}_1(\mu_n, \mu'_n) + \mathcal{W}_1(\eta_n, \eta'_n)$, for $n \geq 2$ and and first two terms we set them to be

$$a_0 := \mathcal{W}_1(\mu_0, \mu'_0) + \mathcal{W}_1(\eta_0, \eta'_0), \qquad a_1 := \mathcal{W}_1(\mu_1, \mu'_1) + \mathcal{W}_1(\eta_1, \eta'_1)$$

which are well defined for $\mu_0, \mu'_0 \in \mathcal{P}_1(\mathbb{R}^d)$ and $\eta_0, \eta'_0 \in \mathcal{P}^*_1(\mathbb{R}^d)$. Then from (3.4.53) and denoting $c_1 := \max\Big\{\Big(\big(\|A\| + \delta\sigma(2 + l^{\nabla,\alpha}_{PP'})\big) + \alpha l(P')\Big), (1-\alpha)l(P)\Big\}$, $c_2 := \delta\sigma \max\big\{\alpha l^{\nabla}_{P'}, (1-\alpha)l^{\nabla}_{P}\big\}$ following holds

$$a_n \leq c_1 a_{n-1} + c_2 a_{n-2} \tag{3.4.54}$$

for $n \geq 2$. Given (ω, δ, α) if there exists a $\theta \in (0,1)$ for which the following inequality holds

$$\frac{c_1}{\theta} + \frac{c_2}{\theta^2} \leq 1, \tag{3.4.55}$$

then denoting $\lambda = \frac{c_2}{\theta}$, we have

$$a_n \leq \left[\theta\left(1 - \frac{\lambda}{\theta}\right)\right] a_{n-1} + \theta\lambda a_{n-2} \quad \Leftrightarrow \quad a_n + \lambda a_{n-1} \;\leq\; \theta(a_{n-1} + \lambda a_{n-2}).$$

Existence of a solution $\theta \in (0,1)$ satisfying (3.4.55) is valid under $c_1 + c_2 < 1$ which is equivalent to the condition

$$\begin{aligned} &\max\left\{\Big((\|A\| + \delta\sigma(2 + l^{\nabla,\alpha}_{PP'})) + \alpha l(P')\Big), (1-\alpha)l(P)\right\} \\ &\qquad\qquad + \delta\sigma \max\left\{\alpha l^{\nabla}_{P'}, (1-\alpha)l^{\nabla}_{P}\right\} < 1 \end{aligned} \tag{3.4.56}$$

in (3.4.43) satisfied by $(\delta, \alpha, \|A\|)$. From (3.4.56) it follows

$$a_n \leq a_n + \lambda a_{n-1} \leq \theta^{n-1}[a_1 + \lambda a_0]$$

for $n \geq 2$. Since

$$\begin{aligned} \mathcal{W}_1(\eta_1, \eta_1') &= \mathcal{W}_1(\eta_0 R^{\alpha}_{\mu_0}, \eta_0' R^{\alpha}_{\mu_0'}) \leq (1-\alpha)l(P)\mathcal{W}_1(\eta_0, \eta_0') + \alpha l(P')\mathcal{W}_1(\mu_0, \mu_0'), \\ \mathcal{W}_1(\mu_1, \mu_1') &= \mathcal{W}_1(\mu_0 Q^{\eta_0,\mu_0}, \mu_0' Q^{\eta_0',\mu_0'}) \leq \mathcal{W}_1(\mu_0 Q^{\eta_0,\mu_0}, \mu_0' Q^{\eta_0,\mu_0'}) + \mathcal{W}_1(\mu_0' Q^{\eta_0,\mu_0'}, \mu_0' Q^{\eta_0',\mu_0'}) \\ &\leq \Big(\|A\| + \delta\sigma(2 + l^{\nabla}_{PP'})\Big)\mathcal{W}_1(\mu_0, \mu_0') + \delta\sigma E\left|\nabla\eta_0(X) - \nabla\eta_0'(X)\right| \end{aligned}$$

where $X \sim \mu_0'$. Final estimate for a_n is

$$a_n \leq \theta^{n-1}\Bigg[\Big(\max\Big\{\Big(\|A\| + \delta\sigma(2 + l_{PP'}^{\nabla,\alpha}) + \alpha l(P')\Big), (1-\alpha)l(P)\Big\} + \lambda\Big)a_0 \\ + \delta\sigma E\left|\nabla\eta_0(X) - \nabla\eta_0'(X)\right|\Bigg]. \quad (3.4.57)$$

Since $X \sim \mu_0' \in \mathcal{P}_1(\mathbb{R}^d)$ and $\nabla\eta_0, \nabla\eta_0'$ have linear growth (since $\eta_0, \eta_0' \in \mathcal{P}_1^*(\mathbb{R}^d)$), the second term inside the bracket is finite. A general formula can be observed for a_n

$$\mathcal{W}_1(\Psi^n(\mu_0,\eta_0), \Psi^n(\mu_0',\eta_0')) \leq \theta^n\Big[a\mathcal{W}_1((\mu_0,\eta_0),(\mu_0',\eta_0')) \\ + b\mathcal{W}_1(\mu_0' Q^{\eta_0,\mu_0'}, \mu_0' Q^{\eta_0',\mu_0'})\Big] \quad (3.4.58)$$

where

$$a = \frac{\max\Big\{\Big(\|A\| + \delta\sigma(2 + l_{PP'}^{\nabla,\alpha}) + \alpha l(P')\Big), (1-\alpha)l(P)\Big\} + \lambda}{\theta}, \qquad b = \frac{1}{\theta}.$$

Observe that the quantity inside the bracket of RHS of (3.4.58) is finite for $\mu_0, \mu_0' \in \mathcal{P}_1(\mathbb{R}^d)$ and $\eta_0, \eta_0' \in \mathcal{P}_1^*(\mathbb{R}^d)$. Hence proved the lemma.

$\square$

We now complete the proof of the theorem. Given $l(PP') < 1$ from Assumption (10), one can always find $(\omega_0, \alpha_0, \delta_0) \in (0,1) \times (0,1) \times (0,1)$ for which (3.4.56) holds under

$$\|A\| < \omega_0, \quad \alpha < \alpha_0, \quad \delta < \delta_0.$$

For existence we need to show that under $\mathcal{W}_1\left((\cdot,\cdot),(\cdot,\cdot)\right)$ distance $\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1^*(\mathbb{R}^d)$ is complete. From Lemma 19 one can choose (ω,α,δ) such that (3.4.43) holds. It follows that using the θ from that lemma the sequence $\{\Psi^n(\mu_0,\eta_0)\}_{n\geq 1}^{\infty}$ is a cauchy sequence in $\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d)$ which is a complete metric space under $\mathcal{W}_1\left((\cdot,\cdot),(\cdot,\cdot)\right)$. So there exists a $(\mu_\infty,\eta_\infty)\in\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d)$ such that $\Psi^n(\mu_0,\eta_0)\to(\mu_\infty,\eta_\infty)$ as $n\to\infty$. Our assertion for existence will be proved if we prove $\eta_\infty\in\mathcal{P}_1^*(\mathbb{R}^d)$. Given the initial conditon $\|\nabla\eta_0(x)\|_1<\infty$, we will always have from (3.4.1) $\|\nabla\eta_k(x)\|_1<\infty\quad\forall\quad k>1$. Note that for $\eta_0\in\mathcal{P}_1^*(\mathbb{R}^d)$, one has $\eta_k\in\mathcal{P}_1^*(\mathbb{R}^d)$ for all k. This implies $\eta_\infty\in\mathcal{P}_1^*(\mathbb{R}^d)$. So

$$(\mu_\infty,\eta_\infty)\in\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1^*(\mathbb{R}^d).$$

Observe further for $\theta\in(0,1)$ in (3.4.58) of Lemma 19

$$\begin{aligned}\mathcal{W}_1\big(\Psi^n(\mu_0,\eta_0),(\mu_\infty,\eta_\infty)\big) &= \mathcal{W}_1\big(\Psi^n(\mu_0,\eta_0),\Psi^n(\mu_\infty,\eta_\infty)\big)\\ \leq\theta^n\big[a\mathcal{W}_1\big((\mu_0,\eta_0),(\mu_\infty,\eta_\infty)\big) &+ b\mathcal{W}_1(\mu_\infty Q^{\eta_0,\mu_\infty},\mu_\infty Q^{\eta_\infty,\mu_\infty})\big].\end{aligned} \tag{3.4.59}$$

Uniqueness of fixed points follows immediately from (3.4.59).

□

3.4.3 Proof of Theorem 3.3.2

We will prove part (b) of the theorem. Part (a) will follow similarly. We need to prove the following Lemma first.

Lemma 16. *Consider the second particle system $\mathbb{IPS}_2$. Suppose all the conditions of Theorem*

3.3.2 hold. Denote $N_1 = \min\{N, M\}$. Then there exist a constant $C \in (0, \infty)$ such that the upper-bound $b(\tau, d)$ of the quantity $\sup_{k\geq 1} E\mathcal{W}_1\big((\bar{\mu}_k^N, \bar{\eta}_k^M), \Psi(\bar{\mu}_{k-1}^N, \bar{\eta}_{k-1}^M)\big)$ can be given as $b(N_1, \tau, d)$ as defined in Theorem 3.3.2. The constant C will vary for different cases.

Proof of Lemma 16

We start with the fact that

$$
\begin{aligned}
& E\mathcal{W}_1\big((\bar{\mu}_k^N, \bar{\eta}_k^M), \Psi(\bar{\mu}_{k-1}^N, \bar{\eta}_{k-1}^M)\big) = E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}) + E\mathcal{W}_1(\bar{\eta}_k^M, \bar{\eta}_{k-1}^M R^{\alpha}_{\bar{\mu}_{k-1}^N}) \\
\leq\ & E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}) + (1-\alpha) E\mathcal{W}_1(S^M(\bar{\eta}_{k-1}^M), \bar{\eta}_{k-1}^M) \\
=\ & E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}) + (1-\alpha) E\left[E\mathcal{W}_1(S^M(\bar{\eta}_{k-1}^M), \bar{\eta}_{k-1}^M) \middle| \mathcal{F}_{k-1}^{M,N}\right]. \qquad (3.4.60)
\end{aligned}
$$

In order to bound both terms in (3.4.60) we borrow the following formulation from [33] about the convergence rate of empirical distribution of iid random variables to its common distribution, where the key idea of bounding Wasserstein distance came from the constructive quantization context [24]. A similar idea was also developed in Boissard's work [11]. We will maintain the same notation used in [33]. Let $\mathcal{P}_l$ be the natural partition of $(-1, 1]^d$ into 2^{dl} translations of $(-2^{-l}, 2^{-l}]^d$. Define a sequence of sets $\{B_n\}_{n\geq 0}$ such that $B_0 := (-1, 1]^d$ and, for $n \geq 1$, $B_n := (-2^n, 2^n]^d \setminus (-2^{n-1}, 2^{n-1}]^d$. For a set $F \subset \mathbb{R}^d$ denote the set $2^n F$ as $\{2^n x : x \in F\}$. For any two probability measures μ and ν, combining Lemma 5 and 6 of [33] one has the following inequality for the Wasserstein-1 distance,

$$
\mathcal{W}_1(\mu, \nu) \leq 3C.2^{(1+\frac{d}{2})} \sum_{n\geq 0} 2^n \sum_{l\geq 0} 2^{-l} \sum_{F\in\mathcal{P}_l} \left[\mu(2^n F \cap B_n) - \nu(2^n F \cap B_n)\right], \quad (3.4.61)
$$

where C is a constant depends only on d. We denote $a_k^{i,M,N} := \delta_{\bar{X}_k^i} - \delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}$. It follows that $\bar{\mu}_k^N - \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N} = \frac{1}{N}\sum_{i=1}^N a_k^{i,M,N}$. Note that on conditioned upon $\mathcal{F}_{k-1}^{M,N}$, the family of signed measures $\{a_k^{i,M,N}\}_{i=1,\ldots,M}$ is an independent class of measures while unconditionally they are just identical. Using the fact that for any set $A \in \mathcal{B}(\mathbb{R}^d)$, $\quad \delta_{\bar{X}_k^i}(A)\Big|\mathcal{F}_{k-1}^{M,N} \sim$ Bernoulli$(\delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A))$, we have

$$\begin{aligned} E\left[\left(a_k^{i,M,N}(A)\right)^2 \middle| \mathcal{F}_{k-1}^{M,N}\right] &= \delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A)\left[1 - \delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A)\right] \\ &\leq \delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A) \end{aligned} \tag{3.4.62}$$

which implies the unconditional expectation

$$E\big[\big(a_k^{i,M,N}(A)\big)^2\big] \leq P\big[\bar{X}_{k-1}^i + \delta f_\delta(\nabla\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N, \bar{X}_{k-1}^i, \varepsilon_k^N) \in A\big].$$

Using all these we have

$$\begin{aligned} &E\left|\bar{\mu}_k^N(A) - \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A)\right|^2 = E\left|\frac{1}{N}\sum_{i=1}^N a_k^{i,M,N}(A)\right|^2 \leq \frac{E\left[a_k^{i,M,N}(A)\right]^2}{N} \\ &\leq \frac{P\big[\bar{X}_{k-1}^i + \delta f_\delta(\nabla\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N, \bar{X}_{k-1}^i, \varepsilon_k^N) \in A\big]}{N} = \frac{E\left[\delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A)\right]}{N}. \end{aligned}$$

Using these with Cauchy-Schwarz inequality one gets following bound

$$\begin{aligned} E\left|\bar{\mu}_k^N(A) - \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A)\right| \leq \min\Bigg\{&\sqrt{\frac{E\big[\delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A)\big]}{N}}, \\ &2E\left[\delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}(A)\right]\Bigg\} \end{aligned} \tag{3.4.63}$$

where second term inside the bracket of RHS of (3.4.63) follows trivially. Denoting the whole constant in R.H.S of (3.4.61) as C_d, we have

$$E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M,\bar{\mu}_{k-1}^N}) \leq C_d \sum_{n\geq 0} 2^n \sum_{l\geq 0} 2^{-l} E \sum_{F\in\mathcal{P}_l} \big[\bar{\mu}_k^N(2^n F \cap B_n) - \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M,\bar{\mu}_{k-1}^N}(2^n F\cap B_n)\big] \tag{3.4.64}$$

Note that $\#\mathcal{P}_l = 2^{dl}$. Using Cauchy-Schwarz inequality with (3.4.63) and Jensen's inequality $E\sqrt{X} \leq \sqrt{EX}$ for non-negative random variable X , the last sum $E\sum_{F\in\mathcal{P}_l}\big[\bar{\mu}_k^N(2^n F \cap B_n) - \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M,\bar{\mu}_{k-1}^N}(2^n F\cap B_n)\big]$ in the R.H.S of (3.4.64) can be bounded by

$$\leq \min\left\{2^{\frac{dl}{2}}\left[\frac{E\big[\delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M,\bar{\mu}_{k-1}^N}(B_n)\big]}{N}\right]^{\frac{1}{2}}, 2E\left[\delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M,\bar{\mu}_{k-1}^N}(B_n)\right]\right\}. \tag{3.4.65}$$

Now using Remark 10 along with Lemma 12, if $\delta \in (0, a(\tau))$ the quantity $\sup_{n\geq 0}\sup_{M,N\geq 1} E|\bar{X}_n^i|^{1+\tau} := b(\tau) < \infty$, one has by Chebyshev inequality for $n \geq 1$,

$$\sup_{k\geq 1} E\left[\delta_{\bar{X}_{k-1}^i} Q^{\bar{\eta}_{k-1}^M,\bar{\mu}_{k-1}^N}(B_n)\right] \leq \sup_{k\geq 1} P[|\bar{X}_k^i| > 2^{(n-1)}] \leq \frac{b(\tau)}{2^{(1+\tau)(n-1)}} = b(\tau)2^{-(1+\tau)(n-1)}.$$

Note that $a(\tau)^{\frac{1}{1+\tau}} \to a_0$ as $\tau \to 0$ and $\delta \in (0, a_0)$, we can find $\tau_0 \in (0, a(\tau))$ such that $\delta \in (0, a(\tau_0)^{\frac{1}{1+\tau_0}})$. So the bound in (3.4.64) can be restated as

$$\begin{aligned}\sup_{k\geq 1} E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M,\bar{\mu}_{k-1}^N}) \;&\leq\; C_d \sum_{n\geq 0} 2^n \sum_{l\geq 0} 2^{-l} \min\left\{2^{\frac{dl}{2}}\sqrt{\frac{b(\tau)2^{-(1+\tau)(n-1)}}{N}},\right.\\ \left. 2b(\tau)2^{-(1+\tau)(n-1)}\right\} \;&\leq\; C_d' \sum_{n\geq 0} 2^n \sum_{l\geq 0} 2^{-l}\min\left\{2^{\frac{dl}{2}}\frac{2^{-\frac{(1+\tau)n}{2}}}{\sqrt{N}}, 2^{-(1+\tau)n}\right\}\end{aligned} \tag{3.4.66}$$

where $b(\tau)$ is just a constant and the last display is obtained by accumulating upper bounds of all the constants to C'_d. Now proceeding exactly like step 1 to step 4 of the proof of Theorem 1 (for $p = 1, q = 1+\tau$) in [33] one gets the following bounds

$$\sup_{k\geq 1} E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}) = C \begin{cases} N^{-\max\{\frac{1}{2}, \frac{\tau}{1+\tau}\}} & \text{if} \quad d = 1, \tau \neq 1, \\ N^{-\frac{1}{2}} \log(1+N) + N^{-\frac{\tau}{1+\tau}} & \text{if} \quad d = 2, \tau \neq 1, \\ N^{-\max\{\frac{1}{d}, \frac{\tau}{1+\tau}\}} & \text{if} \quad d > 2, \tau \neq \frac{1}{d-1}. \end{cases}$$

Now we will fill the gaps for each of the three special cases $\tau = 1, \tau = 1$ and $\tau = \frac{1}{d-1}$ of three regimes respectively $d = 1, d = 2$ and $d > 2$. We note that one can generalize the choice of $l_{N,\epsilon}$ done in step 1 of Theorem 1 of [33] where $l_{N,\epsilon}$ could be taken as $\frac{\frac{1}{2}\log(\epsilon N)}{d \log 2} \vee 0$ instead of $\frac{\log(2+\epsilon N)}{d \log 2}$ though it doesn't change the conclusion of the main theorem. After step 1 with $p = 1, q = 1+\tau, \epsilon = 2^{-(1+\tau)n}$ one will get

$$\sum_{l\geq 0} 2^{-l} \min\left\{2^{\frac{dl}{2}} \sqrt{\frac{\epsilon}{N}}, \epsilon\right\} = C \begin{cases} \min\{\epsilon, \left(\frac{\epsilon}{N}\right)^{\frac{1}{2}}\} & \text{if} \quad d = 1, \\ \min\{\epsilon, \left(\frac{\epsilon}{N}\right)^{\frac{1}{2}} [\log(\epsilon N) \vee 0]\} & \text{if} \quad d = 2, \\ \min\{\epsilon, \epsilon\,(\epsilon N)^{-\frac{1}{d}}\} & \text{if} \quad d > 2, \end{cases}$$

where the constant C will vary from case to cases. Suppose $d = 1$. From (3.4.66) for general $\tau > 0$ one has

$$\sup_{k\geq 1} E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}) \quad \leq \quad C'_d \sum_{n\geq 0} 2^n \min\left\{2^{-(1+\tau)n}, \left(\frac{2^{-(1+\tau)n}}{N}\right)^{\frac{1}{2}}\right\} \tag{3.4.67}$$

Note that for $n \geq n_{N,\tau} := \frac{\log N}{(1+\tau)\log 2}$, one has $2^{-(1+\tau)n} \leq \left(\frac{2^{-(1+\tau)n}}{N}\right)^{\frac{1}{2}}$. So for $\tau = 1$,

$$\begin{aligned}\sum_{n\geq 0} 2^n \min\left\{2^{-2n}, \left(\frac{2^{-2n}}{N}\right)^{\frac{1}{2}}\right\} &\leq \sum_{n<n_{N,1}} 2^n \left(\frac{2^{-2n}}{N}\right)^{\frac{1}{2}} + \sum_{n\geq n_{N,1}} 2^{-n} \\ = n_{N,1}N^{-\frac{1}{2}} + C2^{-n_{N,1}} &= N^{-\frac{1}{2}}\frac{\log N}{2\log 2} + CN^{-\frac{1}{2}}. \qquad (3.4.68)\end{aligned}$$

For $d = 2$, from (3.4.66) for general $\tau > 0$ one has

$$\begin{aligned}\sup_{k\geq 1} E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}) &\leq C_d' \sum_{n\geq 0} 2^n \min\left\{2^{-(1+\tau)n}, \left(\frac{2^{-(1+\tau)n}}{N}\right)^{\frac{1}{2}} \times \right. \\ &\qquad \left. \left[\log\left(2^{-(1+\tau)n}N\right) \vee 0\right]\right\}.\end{aligned}$$

For $\tau = 1$, $\epsilon = 2^{-2n}$. Note that if $n < n_N^{(2)} := \log_4 N - \log_2(\log N)$, then one has

$$\epsilon = 2^{-2n} > \left(\frac{2^{-2n}}{N}\right)^{\frac{1}{2}} \left[\log\left(2^{-2n}N\right) \vee 0\right].$$

$$\begin{aligned}&\sum_{n\geq 0} 2^n \min\left\{2^{-2n}, \left(\frac{2^{-2n}}{N}\left[\log\left(2^{-2n}N\right)\vee 0\right]\right)^{\frac{1}{2}}\right\} \\ \leq& \sum_{n<n_N^{(2)}} 2^n \left(\frac{2^{-2n}}{N}\right)^{\frac{1}{2}} \left[\log\left(2^{-2n}N\right)\vee 0\right] + \sum_{n\geq n_N^{(2)}} 2^{-n} \leq n_N^{(2)} \frac{\left[\log(N)\vee 0\right]}{N^{\frac{1}{2}}} + C2^{-n_N^{(2)}} \\ \leq& C_1 N^{-\frac{1}{2}}\left[(\log N)^2 - \log N \log_2(\log N)\right] + C_2\frac{\log N}{\sqrt{N}}. \qquad (3.4.69)\end{aligned}$$

By proceeding similarly, for all non regular cases we will end up getting the following results (the constant C will vary from case to cases):

$$\sup_{k\geq 1} E\mathcal{W}_1(\bar{\mu}_k^N, \bar{\mu}_{k-1}^N Q^{\bar{\eta}_{k-1}^M, \bar{\mu}_{k-1}^N}) = C \begin{cases} N^{-\frac{1}{2}} \log N + N^{-\frac{1}{2}} & \text{if} \quad d=1, \tau=1, \\ \frac{(\log N)^2}{N^{\frac{1}{2}}} & \text{if} \quad d=2, \tau=1, \\ \frac{\log_2 N}{N^{\frac{1}{d}}} + N^{-\frac{1}{d}} & \text{if} \quad d>2, \tau=\frac{1}{d-1}. \end{cases}$$

Now about the second term of (3.4.60) using (3.4.61), the upperbound of $E\mathcal{W}_1(S^M(\bar{\eta}_{k-1}^M)\bar{\eta}_{k-1}^M)$ is

$$3C2^{(1+\frac{d}{2})} \sum_{n\geq 0} 2^n \sum_{l\geq 0} 2^{-l} E \sum_{F\in\mathcal{P}_l} \left[S^M(\bar{\eta}_{k-1}^M)(2^n F\cap B_n) - \bar{\eta}_{k-1}^M(2^n F\cap B_n)\right]. \tag{3.4.70}$$

By Cauchy Schwarz inequality and using Jensen inequality $E\sqrt{X} \leq \sqrt{EX}$ for a nonnegative random variable X, one gets the upperbound of

$$\begin{aligned} & E\left[\sum_{F\in\mathcal{P}_l} \left[S^M(\bar{\eta}_{k-1}^M)(2^n F\cap B_n) - \bar{\eta}_{k-1}^M(2^n F\cap B_n)\right] \middle| \mathcal{F}_{k-1}^{M,N} \right] \\ \leq \; & 2^{\frac{dl}{2}} \left[\sum_{F\in\mathcal{P}_l} E\big[\big(\frac{1}{M}\sum_{i=1}^M \delta_{Y_{k-1}^{i,M}}(2^n F\cap B_n) - \bar{\eta}_{k-1}^M(2^n F\cap B_n)\big)^2 \big| \mathcal{F}_{k-1}^{M,N}\big] \right]^{\frac{1}{2}} \end{aligned} \tag{3.4.71}$$

Using similar argument used in (3.4.62) the R.H.S of (3.4.71) will be less than

$$\begin{aligned} & 2^{\frac{dl}{2}} \left[\frac{\sum_{F\in\mathcal{P}_l} \bar{\eta}_{k-1}^M(2^n F\cap B_n)\left(1-\bar{\eta}_{k-1}^M(2^n F\cap B_n)\right)}{M} \right]^{\frac{1}{2}} \leq 2^{\frac{dl}{2}} \left[\frac{\bar{\eta}_{k-1}^M(B_n)}{M}\right]^{\frac{1}{2}} \\ \leq \; & 2^{\frac{dl}{2}} \left[\frac{\bar{\eta}_{k-1}^M(x:|x|>2^{n-1})}{M}\right]^{\frac{1}{2}} \leq 2^{\frac{dl}{2}} \left[\frac{\left\langle |x|^{1+\tau}, \bar{\eta}_{k-1}^M\right\rangle 2^{-(n-1)(1+\tau)}}{M}\right]^{\frac{1}{2}}. \end{aligned} \tag{3.4.72}$$

Finally using Jensen inequality $E\sqrt{X} \leq \sqrt{EX}$, and from Corollary 3.4.1 followed by Lemma 13(b) denoting $c(\tau) := \sup_{k\geq 1}\sup_{M\geq 1} E\left\langle |x|^{1+\tau}, \bar{\eta}^M_{k-1}\right\rangle$ one gets

$$\begin{aligned}
&\sup_{k\geq 1} E\sum_{F\in\mathcal{P}_l}\left[S^M(\bar{\eta}^M_{k-1})(2^nF\cap B_n) - \bar{\eta}^M_{k-1}(2^nF\cap B_n)\right] \leq \\
&2^{\frac{dl}{2}}\sup_{k\geq 1} E\left[\frac{\left\langle |x|^{1+\tau}, \bar{\eta}^M_{k-1}\right\rangle 2^{-(n-1)(1+\tau)}}{M}\right]^{\frac{1}{2}} \\
\leq\ & 2^{\frac{dl}{2}}\sup_{k\geq 1}\left[\frac{E\left\langle |x|^{1+\tau}, \bar{\eta}^M_{k-1}\right\rangle 2^{-(n-1)(1+\tau)}}{M}\right]^{\frac{1}{2}} \leq 2^{\frac{dl}{2}}\left[\frac{\sup_{k\geq 1} E\left\langle |x|^{1+\tau}, \bar{\eta}^M_{k-1}\right\rangle 2^{-(n-1)(1+\tau)}}{M}\right]^{\frac{1}{2}} \\
\leq\ & 2^{\frac{dl}{2}}\left[\frac{c(\tau)2^{-(n-1)(1+\tau)}}{M}\right]^{\frac{1}{2}}.
\end{aligned}$$

Hence the conclusion about the upper bound of $E\mathcal{W}_1(S^M(\bar{\eta}^M_{k-1}), \bar{\eta}^M_{k-1})$ will be similar to the first term of (3.4.60). It will be a function of the sample size of the concentration gradient M in place of N in the bound of $E\mathcal{W}_1(\bar{\mu}^N_k, \bar{\mu}^N_{k-1}Q^{\bar{\eta}^M_{k-1},\bar{\mu}^N_{k-1}})$. Combining this with the conclusion about the first term of (3.4.60) we can state the bound in terms of $N_1 = \min\{M, N\}$ and the result of Lemma 16 will follow.

□

Now we will complete the theorem. Observe the following identity

$$\begin{aligned}
(\bar{\mu}^N_n, \bar{\eta}^M_n) - (\mu_n, \eta_n) \quad = \quad & \sum_{i=1}^{n}\left[\Psi^{(n-i)}(\bar{\mu}^N_i, \bar{\eta}^M_i) - \Psi^{(n-i)}\circ\Psi(\bar{\mu}^N_{i-1}, \bar{\eta}^M_{i-1})\right] \\
& + \left[\Psi^n(\bar{\mu}^N_0, \bar{\eta}^M_0) - \Psi^n(\mu_0, \eta_0)\right].
\end{aligned}$$

Using Triangular inequality and Lemma 19 following holds

$$
\begin{aligned}
& \mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\bar{\mu}_n, \bar{\eta}_n)\right) \\
\leq \; & \sum_{i=1}^{n} \mathcal{W}_1\left(\Psi^{(n-i)}(\bar{\mu}_i^N, \bar{\eta}_i^M), \Psi^{(n-i)} \circ \Psi(\bar{\mu}_{i-1}^N, \bar{\eta}_{i-1}^M)\right) + \mathcal{W}_1\left(\Psi^n(\bar{\mu}_0^N, \bar{\eta}_0^M), \Psi^n(\mu_0, \eta_0)\right) \\
\leq \; & \sum_{i=1}^{n} \theta^{n-i}\Bigg[a\mathcal{W}_1\left((\bar{\mu}_i^N, \bar{\eta}_i^M), \Psi(\bar{\mu}_{i-1}^N, \bar{\eta}_{i-1}^M)\right) + \\
& b\mathcal{W}_1\left(\bar{\mu}_{M,N}^{(i-1)} Q^{\bar{\eta}_i^M, \bar{\mu}_{M,N}^{(i-1)}}, \bar{\mu}_{M,N}^{(i-1)} Q^{\bar{\eta}_{i-1}^M R^{\alpha}_{\bar{\mu}_{i-1}^N}, \bar{\mu}_{M,N}^{(i-1)}}\right)\Bigg] \\
& +\theta^n\left[a\mathcal{W}_1\left((\bar{\mu}_0^N, \bar{\eta}_0^M), (\mu_0, \eta_0)\right) + b\mathcal{W}_1(\mu_0 Q^{\bar{\eta}_0^M, \mu_0}, \mu_0 Q^{\eta_0, \mu_0})\right] \qquad (3.4.73)
\end{aligned}
$$

where (3.4.73) follows from (3.4.58) with specified constants a and b and $\bar{\mu}_{M,N}^{(i-1)} := \bar{\mu}_{i-1}^N Q^{\bar{\eta}_{i-1}^M, \bar{\mu}_{i-1}^N}$. Let $X_i^{M,N}$ be a random variable, conditioned on $\mathcal{F}_{i-1}^{M,N}$, sampled from $\bar{\mu}_{M,N}^{(i-1)}$. We have

$$
\begin{aligned}
& \mathcal{W}_1\left(\bar{\mu}_{M,N}^{(i-1)} Q^{\bar{\eta}_i^M, \bar{\mu}_{M,N}^{(i-1)}}, \bar{\mu}_{M,N}^{(i-1)} Q^{\bar{\eta}_{i-1}^M R^{\alpha}_{\bar{\mu}_{i-1}^N}, \bar{\mu}_{M,N}^{(i-1)}}\right) \\
\leq \; & \sup_{g \in \mathrm{Lip}_1(\mathbb{R}^d)} E\Big| g(AX_i^{M,N} + \delta f_\delta(\nabla \bar{\eta}_i^M, \bar{\mu}_{M,N}^{(i-1)}, X_i^{M,N}, \epsilon)) - g(AX_i^{M,N} \\
& + \delta f_\delta(\nabla(\bar{\eta}_{i-1}^M R^{\alpha}_{\bar{\mu}_{i-1}^N}), \bar{\mu}_{M,N}^{(i-1)}, X_i^{M,N}, \epsilon))\Big| \\
\leq \; & \delta\sigma E\left[\left|\nabla \bar{\eta}_i^M(X_i^{M,N}) - \nabla \bar{\eta}_{i-1}^M R^{\alpha}_{\bar{\mu}_{i-1}^N}(X_i^{M,N})\right| \big| \mathcal{F}_{i-1}^{M,N}\right] \\
= \; & (1-\alpha)\int \left| \int \left[S^M(\bar{\eta}_{i-1}^M) - \bar{\eta}_{i-1}^M\right](dx) \nabla_y P(x,y)\right| (\bar{\mu}_{i-1}^N Q^{\bar{\eta}_{i-1}^M})(dy) \\
\leq \; & l_P^{\nabla}(1-\alpha)\mathcal{W}_1\left(S^M(\bar{\eta}_{i-1}^M), \bar{\eta}_{i-1}^M\right). \qquad (3.4.74)
\end{aligned}
$$

Last display follows from Assumption 9. Since $\bar{\eta}_0^M = \eta_0$, one has

$$\mathcal{W}_1(\mu_0 Q^{\bar{\eta}_0^M,\mu_0}, \mu_0 Q^{\eta_0,\mu_0}) = 0. \tag{3.4.75}$$

Combining the results (3.4.74),(3.4.75), with (3.4.73) we get for each n,

$$\begin{aligned} E\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_n, \eta_n)\right) &\le \frac{a}{1-\theta} \sup_{k\ge 1} E\mathcal{W}_1\big((\bar{\mu}_k^N, \bar{\eta}_k^M), \Psi(\bar{\mu}_{k-1}^N, \bar{\eta}_{k-1}^M)\big) \\ +\frac{bl_P^{\nabla}(1-\alpha)}{1-\theta} \sup_{k\ge 1} E\mathcal{W}_1\left(S^M(\bar{\eta}_{k-1}^M), \bar{\eta}_{k-1}^M\right) &+ a\theta^n E\mathcal{W}_1\left((\bar{\mu}_0^N, \bar{\eta}_0^M), (\mu_0, \eta_0)\right). \end{aligned} \tag{3.4.76}$$

Using Lemma 16 the result follows.

□

3.4.4 Proof of Corollary 4.7.1:

Using triangular inequality and from (3.4.58) one gets

$$\begin{aligned} & E\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_\infty, \eta_\infty)\right) \le \mathcal{W}_1\big((\mu_n, \eta_n), (\mu_\infty, \eta_\infty)\big) + E\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_n, \eta_n)\right) \\ \le\ & \theta^n \Big[a\mathcal{W}_1((\mu_0, \eta_0), (\mu_\infty, \eta_\infty)) + b\mathcal{W}_1(\mu_0 Q^{\eta_0}, \mu_0 Q^{\eta_\infty}) \Big] \\ & + E\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_n, \eta_n)\right). \end{aligned} \tag{3.4.77}$$

Combining this with (3.4.76) we get

$$\begin{aligned} & E\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_\infty, \eta_\infty)\right) \le \theta^n \Big[a\mathcal{W}_1((\mu_0, \eta_0), (\mu_\infty, \eta_\infty)) + b\mathcal{W}_1(\mu_0 Q^{\eta_0}, \mu_0 Q^{\eta_\infty}) \Big] \\ + \ & \frac{a}{1-\theta} \sup_{k\ge 1} E\mathcal{W}_1\big((\bar{\mu}_k^N, \bar{\eta}_k^M), \Psi(\bar{\mu}_{k-1}^N, \bar{\eta}_{k-1}^M)\big) + \frac{bl_P^{\nabla}(1-\alpha)}{1-\theta} \sup_{k\ge 1} E\mathcal{W}_1\left(S^M(\bar{\eta}_{k-1}^M), \bar{\eta}_{k-1}^M\right). \end{aligned}$$

The result is obvious after using Lemma 16.

□

3.4.5 Proof of Theorem 3.3.3:

Fix N and M. Define $\Theta_n^{N,M} \in \mathcal{P}((\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d))$ as

$$\langle \phi, \Theta_n^{N,M} \rangle \;=\; \frac{1}{n}\sum_{j=1}^{n} E\phi\big(\bar{X}_j(N), \eta_j^M, S^M(\eta_j^M)\big), \tag{3.4.78}$$

for $\phi \in BM\big((\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d\big) \times \mathcal{P}(\mathbb{R}^d)), \quad N \geq 1, M \geq 1$ and $n \in \mathbb{N}_0$ where

$$\{(\bar{X}_j(N), \bar{\eta}_j^M, S^M(\bar{\eta}_j^M)),\ j \in \mathbb{N}_0, i = 1, \dots, N\}$$

are as defined in the context of $\mathbb{IPS}_2$. Note that $(\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$ is a complete separable metric space with metric $d((x,\mu_1,\mu_3),(y,\mu_2,\mu_4)) := \|x-y\| + \frac{1}{2}\mathcal{W}_1(\mu_1,\mu_2) + \frac{1}{2}\mathcal{W}_1\big(\mu_3,\mu_4\big)$ where $\|x\| := \frac{1}{N}\sum_{i=1}^N |x_i|$ for $x = (x_1,\dots,x_N) \in (\mathbb{R}^d)^N$. From Lemma 12 and 13 it follows that, for each $N, M \geq 1$, the sequence $\{\Theta_n^{N,M}, n \geq 1\}$ is relatively compact (By Prohorov's Theorem) and using Assumption 6 it is easy to see that any limit point $\Theta_\infty^{N,M}$ of $\Theta_n^{N,M}$ (as $n \to \infty$) is an invariant measure of the Markov chain $\{X_n(N), \bar{\eta}_n^M, S^M(\bar{\eta}_n^M)\}_{n\geq 0}$ and from Lemma 12 it satisfies $\int_{(\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)} |x|\ \Theta_\infty^{N,M}(dx) < \infty$ (Taking the norm of the product space as $|(x,y,z)| = \|x\| + \frac{1}{2}\|y\|_1 + \frac{1}{2}\|z\|_1$ where $(x,y,z) \in (\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$). Uniqueness of invariant measure can be proved by the following simple coupling argument (see Chapter 2): Suppose $\Theta_\infty^{N,M}$, $\tilde{\Theta}_\infty^{N,M}$ are two invariant measures that satisfy

$$\int_{(\mathbb{R}^d)^N\times\mathcal{P}_1^*(\mathbb{R}^d)\times\mathcal{P}(\mathbb{R}^d)}|x|\;\Theta_\infty^{N,M}(dx)<\infty,\qquad \int_{(\mathbb{R}^d)^N\times\mathcal{P}_1^*(\mathbb{R}^d)\times\mathcal{P}(\mathbb{R}^d)}|x|\tilde{\Theta}_\infty^{N,M}(dx)<\infty$$

.

Let $\big(X_0(N),\eta_0^M,S^M(\eta_0^M)\big)$ and $\big(\tilde{X}_0(N),\tilde{\eta}_0^M,S^M(\tilde{\eta}_0^M)\big)$ with probability laws $\Theta_\infty^{N,M}$ and $\tilde{\Theta}_\infty^{N,M}$ respectively be given on a common probability space under same noise sequence (i.e in which an i.i.d. array of $\mathbb{R}^m$ valued random variables $\{\epsilon_n^i, i=1,\dots,N, n\geq 1\}$ are defined that is independent of $(X_0(N),\eta_0^M,\tilde{X}_0(N),\tilde{\eta}_0^M)$ with common probability law θ) and the evolution equations are following.

$$X_{n+1}^i = AX_n^i+\delta f_\delta(X_n^i,\nabla\eta_n^M(X_n^i),\mu_n^N,\epsilon_{n+1}^i),\;\; \eta_k^M=(1-\alpha)(S^M(\eta_{k-1}^M)P)+\alpha\mu_{k-1}^N P',$$
$$\tilde{X}_{n+1}^i = A\tilde{X}_n^i+\delta f_\delta(\tilde{X}_n^i,\nabla\tilde{\eta}_n^M(\tilde{X}_n^i),\tilde{\mu}_n^N,\epsilon_{n+1}^i),\;\; \tilde{\eta}_k^M=(1-\alpha)(S^M(\tilde{\eta}_{k-1}^M)P)+\alpha\tilde{\mu}_{k-1}^N P',$$

where recall $f_\delta(\cdot,\cdot,\cdot,x)=f(\cdot,\cdot,\cdot,x)+\frac{B(x)}{\delta}$ and $\mu_n^N=\frac{1}{N}\sum_{i=1}^N\delta_{X_n^i},\tilde{\mu}_n^N=\frac{1}{N}\sum_{i=1}^N\delta_{\tilde{X}_n^i}$. Note that

$$\mathcal{W}_1\Big(\frac{1}{N}\sum_{i=1}^N\delta_{X_i},\frac{1}{N}\sum_{i=1}^N\delta_{Y_i}\Big)\leq\frac{1}{N}\sum_{i=1}^N|X_i-Y_i| \tag{3.4.79}$$

for any two arrays $\{X_i\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$. Using the independence of the noise sequence along with (3.4.79) and Assumption 6 we have

$$
\begin{aligned}
E|X_{n+1}^i - \tilde{X}_{n+1}^i| \quad &\leq \quad (\|A\| + \delta\sigma)E|X_n^i - \tilde{X}_n^i| + \delta\sigma\frac{1}{N}\sum_{j=1}^{N} E|X_n^j - \tilde{X}_n^j| \\
&\quad +\delta\sigma E|\nabla\eta_n^M(X_n^i) - \nabla\tilde{\eta}_n^M(\tilde{X}_n^i)|. \qquad (3.4.80)
\end{aligned}
$$

Now applying Assumption 9 (doing similar calculations as in (3.4.48),(3.4.50),(3.4.51)) following inequality holds

$$
\begin{aligned}
&E|\nabla\eta_n^M(X_n^i) - \nabla\tilde{\eta}_n^M(\tilde{X}_n^i)| \leq E|\nabla\eta_n^M(X_n^i) - \nabla\eta_n^M(\tilde{X}_n^i)| \\
&+E|\nabla\eta_n^M(\tilde{X}_n^i) - \nabla\tilde{\eta}_n^M(\tilde{X}_n^i)| \qquad (3.4.81) \\
\leq \quad &l_{pp'}^{\nabla} E|X_n^i - \tilde{X}_n^i| + \alpha l_{P'}^{\nabla} E\mathcal{W}_1\big(\mu_{n-1}^N, \tilde{\mu}_{n-1}^N\big) \\
&+(1-\alpha)l_P^{\nabla} E\mathcal{W}_1\big(S^M(\eta_{n-1}^M), S^M(\tilde{\eta}_{n-1}^M)\big).
\end{aligned}
$$

Note that (3.4.79) implies

$$
E\big[\mathcal{W}_1(S^M(\eta_{k-1}^M), S^M(\tilde{\eta}_{k-1}^M))\big|\mathcal{F}_{k-1}^{M,N}\big] \leq \mathcal{W}_1(\eta_{k-1}^M, \tilde{\eta}_{k-1}^M) \qquad (3.4.82)
$$

from which following holds from (3.4.81)

$$
\begin{aligned}
E|\nabla\eta_n^M(X_n^i) - \nabla\tilde{\eta}_n^M(\tilde{X}_n^i)| \quad &\leq \quad l_{pp'}^{\nabla} E|X_n^i - \tilde{X}_n^i| + \alpha l_{P'}^{\nabla} E|X_{n-1}^i - \tilde{X}_{n-1}^i| \\
&\quad +(1-\alpha)l_P^{\nabla} E\mathcal{W}_1\big(\eta_{n-1}^M, \tilde{\eta}_{n-1}^M\big). \qquad (3.4.83)
\end{aligned}
$$

We also have

$$\mathcal{W}_1\big(\eta^M_{n+1}, \tilde{\eta}^M_{n+1}\big) \le (1-\alpha)l(P)\mathcal{W}_1\big(S^M(\eta^M_n), S^M(\tilde{\eta}^M_n)\big) + \alpha l(P')\mathcal{W}_1\big(\mu^N_n, \tilde{\mu}^N_n\big) \quad (3.4.84)$$

and after taking expectation

$$E\mathcal{W}_1\big(\eta^M_{n+1}, \tilde{\eta}^M_{n+1}\big) \le (1-\alpha)l(P)E\mathcal{W}_1\big(\eta^M_n, \tilde{\eta}^M_n\big) + \alpha l(P')E|X^i_n - \tilde{X}^i_n|. \quad (3.4.85)$$

Letting $A^{(M,N)}_{n+1} := \frac{1}{N}\sum_{i=1}^N |X^i_{n+1} - \tilde{X}^i_{n+1}| + \mathcal{W}_1\big(\eta^M_{n+1}, \tilde{\eta}^M_{n+1}\big)$, we have the following recursion relation for $EA^{(M,N)}_n$ in n combining (3.4.80),(3.4.83) and (3.4.85)

$$\begin{aligned} EA^{(M,N)}_{n+1} &\le \max\Big\{\Big(\|A\| + \delta\sigma(2 + l^{\nabla,\alpha}_{PP'}) + \alpha l(P')\Big), (1-\alpha)l(P)\Big\}EA^{(M,N)}_n \\ &\quad + \delta\sigma\max\{(1-\alpha)l^{\nabla}_P, \alpha l^{\nabla}_{P'}\}EA^{(M,N)}_{n-1} \end{aligned} \quad (3.4.86)$$

which is the same recursion as in (3.4.54). Now for the chosen δ, α satisfying (3.4.56) there exists a $\theta \in (0,1)$ such that

$$EA^{(M,N)}_n \le \theta^{n-1}[EA^{(M,N)}_0 + EA^{(M,N)}_1]. \quad (3.4.87)$$

Also, since $\Theta^{N,M}_\infty$ and $\tilde{\Theta}^{N,M}_\infty$ are invariant distributions, for every $n \in \mathbb{N}_0$, $\big(X_{n+1}(N), \eta^M_{n+1}, S^M(\eta^M_{n+1})\big)$ is distributed as $\Theta^{N,M}_\infty$ and $\big(\tilde{X}_{n+1}(N), \tilde{\eta}^M_{n+1}, S^M(\tilde{\eta}^M_{n+1})\big)$ is distributed as $\tilde{\Theta}^{M,N}_\infty$. Thus $(X_{n+1}(N), \eta^M_{n+1}, S^M(\eta^M_{n+1}))$ and $\big(\tilde{X}_{n+1}(N), \tilde{\eta}^M_{n+1}, S^M(\tilde{\eta}^M_{n+1})\big)$ define a coupling of random variables with laws $\Theta_{\infty}^{N,M}$ and $\tilde{\Theta}_{\infty}^{N,M}$ respectively. From (3.4.87) we then have

$$\mathcal{W}_1(\Theta_{\infty}^{N,M}, \tilde{\Theta}_{\infty}^{M,N}) \leq Ed\big((X_n(N), \eta_n^M, S^M(\eta_n^M)), (\tilde{X}_n(N), \tilde{\eta}_n^M, S^M(\tilde{\eta}_n^M))\big) \leq EA_n^{M,N} \to 0,$$

as $n \to \infty$. So there exists a unique invariant measure $\Theta_{\infty}^{N,M} \in \mathcal{P}_1\big((\mathbb{R}^d)^N \times \mathcal{P}_1^*(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)\big)$ for this Markov chain and, as $n \to \infty$,

$$\Theta_n^{N,M} \to \Theta_{\infty}^{N,M}. \tag{3.4.88}$$

This proves the first part of the theorem. Denote $\Theta_{\infty}^{N,M}\left(\cdot, \mathcal{P}_1^*(\mathbb{R}^d), \mathcal{P}(\mathbb{R}^d)\right)$ by $\Theta_{\infty}^{1,N,M}$ and $\Theta_n^{N,M}\left(\cdot, \mathcal{P}_1^*(\mathbb{R}^d), \mathcal{P}(\mathbb{R}^d)\right)$ by $\Theta_n^{1,N,M}$.

Define $r_N : (\mathbb{R}^d)^N \to \mathcal{P}(\mathbb{R}^d)$ as

$$r_N(x_1, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \qquad (x_1, \ldots x_N) \in (\mathbb{R}^d)^N.$$

Let $\nu_n^{N,M} = \Theta_n^{1,N,M} \circ r_N^{-1}$ and $\nu_{\infty}^{N,M} = \Theta_{\infty}^{1,N,M} \circ r_N^{-1}$. In order to prove that $\Theta_{\infty}^{1,N,M}$ is μ_{∞}-chaotic, it suffices to argue that (cf. [74])

$$\nu_{\infty}^{N,M} \to \delta_{\mu_\infty} \text{ in } \mathcal{P}(\mathcal{P}(\mathbb{R}^d)), \text{ as } N, M \to \infty. \tag{3.4.89}$$

We first argue that as $n \to \infty$

$$\nu_n^{N,M} \to \nu_{\infty}^{N,M} \qquad \text{in } \mathcal{P}(\mathcal{P}(\mathbb{R}^d)). \tag{3.4.90}$$

It suffices to show that $\langle F, \nu_n^{N,M}\rangle \to \langle F, \nu_\infty^{N,M}\rangle$ for any continuous and bounded function $F : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$. But this is immediate on observing that

$$\langle F, \nu_n^{N,M}\rangle = \langle F \circ r_N, \Theta_n^{1,N,M}\rangle, \ \langle F, \nu_\infty^{N,M}\rangle = \langle F \circ r_N, \Theta_\infty^{1,N,M}\rangle,$$

the continuity of the map r_N and the weak convergence of $\Theta_n^{N,M}$ to $\Theta_\infty^{N,M}$. Next, for any $f \in BL_1(\mathcal{P}(\mathbb{R}^d))$

$$\left|\langle f, \nu_n^{N,M}\rangle - \langle f, \delta_{\mu_\infty}\rangle\right| = \Big|\frac{1}{n}\sum_{j=1}^{n} Ef(\bar{\mu}_j^N) - f(\mu_\infty)\Big| \le \frac{1}{n}\sum_{j=1}^{n} E\mathcal{W}_1(\bar{\mu}_j^N, \mu_\infty).$$

Fix $\epsilon > 0$. For every $N, M \in \mathbb{N}$ there exists $n_0(N, M) \in \mathbb{N}$ such that for all $n \ge n_0(N, M)$

$$E\mathcal{W}_1(\bar{\mu}_n^N, \mu_\infty) \le \limsup_{n\to\infty} E\mathcal{W}_1(\bar{\mu}_n^N, \mu_\infty) + \epsilon.$$

Thus for all $n, N, M \in \mathbb{N}$

$$|\langle f, \nu_n^{N,M}\rangle - \langle f, \delta_{\mu_\infty}\rangle| \le \frac{n_0(N, M)}{n} \max_{1\le j\le n_0(N,M)} E\mathcal{W}_1(\bar{\mu}_j^N, \mu_\infty) + \limsup_{n\to\infty} E\mathcal{W}_1(\bar{\mu}_n^N, \mu_\infty) + \epsilon. \tag{3.4.91}$$

Finally

$$
\begin{aligned}
\limsup_{N,M\to\infty} |\langle f, \nu_\infty^{N,M}\rangle - \langle f, \delta_{\mu_\infty}\rangle| = & \limsup_{\min\{N,M\}\to\infty} \lim_{n\to\infty} |\langle f, \nu_n^{N,M}\rangle - \langle f, \delta_{\mu_\infty}\rangle| \\
\leq & \limsup_{\min\{N,M\}\to\infty} \limsup_{n\to\infty} E\mathcal{W}_1(\bar{\mu}_n^N, \mu_\infty) + \epsilon \\
\leq & \epsilon,
\end{aligned}
$$

where the first equality is from (3.4.90), the second uses (3.4.91) and the third is a consequence of Corollary 4.7.1. Since $\epsilon > 0$ is arbitrary, we have (3.4.89) and the result follows.

□

3.4.6 Proof of Concentration bounds:

Proof of Theorem 3.3.4:

We start with the following lemma where we establish a concentration bound for

$$
\mathcal{W}_1\big((\bar{\mu}_n^N, \bar{\eta}_n^M), \Psi(\bar{\mu}_{n-1}^N, \bar{\eta}_{n-1}^M)\big)
$$

for each fixed time $n \in \mathbb{N}$ and then combine it with the estimate in (3.4.73) in order to get the desired result.

Lemma 17. *Let* $N_1 = \min\{M, N\}$. *Suppose Assumptions* (6)-(9) *and Assumptions (12),(13) hold for some* $\tau > 0, a(\tau) > 0$. *Suppose that* $\delta \in (0, a(\tau)^{\frac{1}{1+\tau}})$*, and* $(1-\alpha)m_\tau(P) < 1$. *Then there exist*

$a_1, a_2, a_3, a_1', a_2', a_3' \in (0,\infty)$ *such that for all* $\varepsilon, R > 0, n \in \mathbb{N}$, *and* $N_1 \geq \max\{1, a_1(\frac{R}{\varepsilon})^{d+2}\}$.

$$P[\mathcal{W}_1\big((\bar{\mu}_n^N, \bar{\eta}_n^M), \Psi(\bar{\mu}_{n-1}^N, \bar{\eta}_{n-1}^M)\big) > \varepsilon] \leq a_3\left(e^{-a_2\frac{N_1\varepsilon^2}{R^2}} + \frac{R^{-\tau}}{\varepsilon}\right), \tag{3.4.92}$$

$$P[\mathcal{W}_1\left(S^M(\bar{\eta}_{n-1}^M), \bar{\eta}_{n-1}^M\right) > \varepsilon] \leq a_3'\left(e^{-a_2'\frac{N_1\varepsilon^2}{R^2}} + \frac{R^{-\tau}}{\varepsilon}\right). \tag{3.4.93}$$

Proof of Lemma 17

Second concentration bound will follow by proceeding as Lemma 4.5 of Chapter 2. The proof relies on an idea of restricting measures to a compact set and estimates on metric entropy [14] (see also [79]). The basic idea is to first obtain a concentration bound for the $\mathcal{W}_1$ distance between the truncated law and its corresponding empirical law in a compact ball of radius R and getting a tail estimate from Lemma 13 and Corollary 3.4.1 after conditioning by $\mathcal{F}_{n-1}^{M,N}$. With the notations (for example μ_R is the truncated measure of μ restricted on a ball $B_{\mathrm{R}}(0)$ of R radius) introduced in Lemma 4.5 of [**?**] we sketch the proof of the second bound. With that notation the truncated version of $\bar{\eta}_{n-1}^M$ is denoted by $\bar{\eta}_{n-1,R}^M$. Suppoe $\{Y_{n-1}^{i,M} : i = 1, \ldots, M\}$ are iid from $\bar{\eta}_{n-1}^M$ conditioned on $\mathcal{F}_{n-1}^{M,N}$. where $\{Z_i^{M,R} : i = 1, \ldots, M\}$ are iid from $\bar{\eta}_{n-1,R}^M$ conditioned under $\mathcal{F}_{n-1}^{M,N}$. Define

$$X_{n-1}^{i,M} = \begin{cases} Y_{n-1}^{i,M} & \text{when } |Y_{n-1}^{i,M}| \leq R, \\ \\ Z_{n-1}^{i,M} & \text{otherwise .} \end{cases}$$

Note that $P(X_{n-1}^{i,M} \in A \mid \mathcal{F}_{n-1}^{M,N}) = P(Z_{n-1}^{i,M} \in A \mid \mathcal{F}_{n-1}^{M,N})$. Denote $S^M(\bar{\eta}_{n-1,R}^M) := \frac{1}{M}\sum_{i=1}^M \delta_{X_{n-1}^{i,M}}$. Now denoting $a(1+\tau) := \sup_{n\geq 0}\sup_{M,N} E\left\langle |x|^{1+\tau}, \bar{\eta}_n^M\right\rangle$, from (3.4.79) we have

$$
\begin{aligned}
P\big[\mathcal{W}_1(S^M(\bar\eta^M_{n-1,R}), S^M(\bar\eta^M_{n-1})) > \frac{\varepsilon}{3}\big] &\leq 3\frac{E[\mathcal{W}_1(S^M(\bar\eta^M_{n-1,R}), S^M(\bar\eta^M_{n-1}))]}{\varepsilon} \\
\leq \frac{3}{\varepsilon}EE\big[|X^{i,M}_{n-1} - Y^{i,M}_{n-1}|\big|\mathcal{F}^{M,N}_{n-1}\big] &= \frac{3}{\varepsilon}EE\big[|Z^{i,M}_{n-1} - Y^{i,M}_{n-1}|1_{|Y^{i,M}_{n-1}|>R}\big|\mathcal{F}^{M,N}_{n-1}\big] \\
\leq \frac{6}{\varepsilon}EE\big[|Y^{i,M}_{n-1}|1_{|Y^{i,M}_{n-1}|>R}\big|\mathcal{F}^{M,N}_{n-1}\big] &\leq 6a(1+\tau)\frac{R^{-\tau}}{\varepsilon}. \qquad (3.4.94)
\end{aligned}
$$

Now using Azuma Hoeffding inequality as done in display (4.35) of Lemma 4.5 in Chapter 2 one has

$$
P\big[\mathcal{W}_1(S^M(\bar\eta^M_{n-1,R}), \bar\eta^M_{n-1,R}) > \frac{\varepsilon}{3}\big] \leq \max\left\{2, \frac{16R}{\varepsilon}(2\sqrt{d}+1)3^{[\frac{8R}{\varepsilon}(\sqrt{d}+1)]^d}\right\} e^{-\frac{M\varepsilon^2}{288R^2}} \quad (3.4.95)
$$

From the definition of $\bar\eta^M_{n-1,R}$

$$
P\big[\mathcal{W}_1(\bar\eta^M_{n-1}, \bar\eta^M_{n-1,R}) \geq \frac{\varepsilon}{3}\big] \leq \frac{6}{\varepsilon}E\big[|Y^{i,M}_{n-1}|1_{|Y^{i,M}_{n-1}|>R}\big] \leq 3a(1+\tau)\frac{R^{-\tau}}{\varepsilon}. \qquad (3.4.96)
$$

Using triangular inequality

$$
\begin{aligned}
\mathcal{W}_1(S^M(\bar\eta^M_{n-1}), \bar\eta^M_{n-1}) \leq \mathcal{W}_1(S^M(\bar\eta^M_{n-1,R}), S^M(\bar\eta^M_{n-1})) + \mathcal{W}_1(S^M(\bar\eta^M_{n-1,R}), \bar\eta^M_{n-1,R}) \\
+\mathcal{W}_1(\bar\eta^M_{n-1}, \bar\eta^M_{n-1,R})
\end{aligned}
$$

combining (3.4.94),(3.4.95) and (3.4.96) the result (3.4.93) will follow.

The first one (3.4.92) follows by noting that

$$P\big[\mathcal{W}_1\big((\bar{\mu}_n^N,\bar{\eta}_n^M),\Psi(\bar{\mu}_{n-1}^N,\bar{\eta}_{n-1}^M)\big)>\varepsilon\big]\le P\big[\mathcal{W}_1\big(\bar{\mu}_n^N,\bar{\mu}_{n-1}^N Q^{\bar{\mu}_{n-1}^N,\bar{\eta}_{n-1}^M}\big)>\frac{\epsilon}{2}\big] \quad (3.4.97)$$
$$+P\Big[\mathcal{W}_1\big(S^M(\bar{\eta}_{n-1}^M),\bar{\eta}_{n-1}^M\big)>\frac{\epsilon}{2(1-\alpha)l(P)}\Big].$$

Proceeding like Lemma 4.5 of Chapter 2 the bound for the first term in RHS of (3.4.97) can be established.

□

Proof of Theorem 3.3.4(a)

Combining (3.4.73),(3.4.74) and (3.4.75) it follows that

$$\mathcal{W}_1\left((\bar{\mu}_n^N,\bar{\eta}_n^M),(\mu_n,\eta_n)\right) \le \sum_{i=1}^n \theta^{n-i}\Big[a\mathcal{W}_1\left((\bar{\mu}_i^N,\bar{\eta}_i^M),\Psi(\bar{\mu}_{i-1}^N,\bar{\eta}_{i-1}^M)\right)$$
$$+bl_P^{\nabla}(1-\alpha)\mathcal{W}_1\left(S^M(\bar{\eta}_{i-1}^M),\bar{\eta}_{i-1}^M\right)\Big]+a\theta^n\mathcal{W}_1\left((\bar{\mu}_0^N,\bar{\eta}_0^M),(\mu_0,\eta_0)\right). \quad (3.4.98)$$

Denoting $c_1 := \max\Big\{\Big(\left(\|A\|+\delta\sigma(2+l_{PP'}^{\nabla,\alpha})\right)+\alpha l(P')\Big),(1-\alpha)l(P)\Big\}$, $c_2 := \delta\sigma\max\left\{\alpha l_{P'}^{\nabla},(1-\alpha)l_P^{\nabla}\right\}$ define the function $g_0(\cdot)$ as

$$g_0(\gamma) := c_2+(1-\gamma)c_1-(1-\gamma)^2.$$

Since $g_0(0)=c_2+c_1-1<0$ (from the assumption), $g_0(1)=c_2>0$ and $g(\cdot)$ is continuous.

So there exists a $\gamma > 0$ such that $g_0(\gamma) < 0$ or equivalently

$$\frac{c_1}{1-\gamma} + \frac{c_2}{(1-\gamma)^2} < 1.$$

So there exists a $\theta \in (0, 1-\gamma)$ such that statement of Lemma 19 holds. Now using that γ from (3.4.98) one has

$$P\Big[\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_n, \eta_n)\right) > \varepsilon\Big] \leq P\Big[\bigcup_{i=1}^{n}\big\{a\theta^{n-i}\mathcal{W}_1\left((\bar{\mu}_i^N, \bar{\eta}_i^M), \Psi(\bar{\mu}_{i-1}^N, \bar{\eta}_{i-1}^M)\right) > \frac{\gamma}{2}(1-\gamma)^{n-i}\varepsilon\big\}$$
$$\bigcup_{i=1}^{n}\big\{bl_P^{\nabla}(1-\alpha)\theta^{n-i}\mathcal{W}_1\left(S^M(\bar{\eta}_{i-1}^M), \bar{\eta}_{i-1}^M\right) > \frac{\gamma}{2}(1-\gamma)^{n-i}\varepsilon\big\}\bigcup_{i=1}^{n}\big\{\theta^n\mathcal{W}_1\left((\bar{\mu}_0^N, \bar{\eta}_0^M), (\mu_0, \eta_0)\right)$$
$$> \gamma(1-\gamma)^n\varepsilon\big\}\Big] \quad \leq \quad \sum_{i=1}^{n} P\big[\mathcal{W}_1\left((\bar{\mu}_i^N, \bar{\eta}_i^M), \Psi(\bar{\mu}_{i-1}^N, \bar{\eta}_{i-1}^M)\right) > \frac{\gamma\varepsilon}{2a}\big(\frac{1-\gamma}{\theta}\big)^{n-i}\big] +$$
$$\sum_{i=1}^{n} P\big[\mathcal{W}_1\left(S^M(\bar{\eta}_{i-1}^M), \bar{\eta}_{i-1}^M\right) > \frac{\gamma\varepsilon}{2bl_P^{\nabla}(1-\alpha)}\big(\frac{1-\gamma}{\theta}\big)^{n-i}\big]$$
$$+P\big[\mathcal{W}_1\left((\bar{\mu}_0^N, \bar{\eta}_0^M), (\mu_0, \eta_0)\right) > \gamma\varepsilon\Big(\frac{1-\gamma}{\theta}\Big)^n\big].$$

Let $\beta_1 = \frac{\gamma\varepsilon}{2a}$, $\beta_2 = \frac{\gamma\varepsilon}{2bl_P^{\nabla}(1-\alpha)}$ $\beta_3 = \gamma\varepsilon$. Note that $\nu := \left(\frac{1-\gamma}{\theta}\right) > 1$, from our choice of γ. Therefore denoting $\beta := \min\{\beta_1, \beta_2\}$, $\quad N_1 \geq a_1\Big(\frac{R}{\beta}\Big)^{d+2} \vee 1$ implies $N_1 \geq a_1\Big(\frac{R}{\beta\nu^n}\Big)^{d+2} \vee 1$

for all $n \in \mathbb{N}_0$ and a consequence of Lemma 17 gives

$$
\begin{aligned}
& P\Big[\mathcal{W}_1\left((\bar{\mu}_n^N, \bar{\eta}_n^M), (\mu_n, \eta_n)\right) > \varepsilon\Big] \leq \sum_{i=1}^{n} P\big[\mathcal{W}_1\left((\bar{\mu}_i^N, \bar{\eta}_i^M), \Psi(\bar{\mu}_{i-1}^N, \bar{\eta}_{i-1}^M)\right) > \beta_1 \nu^{n-i}\big] \\
+ & \sum_{i=1}^{n} P\big[\mathcal{W}_1\left(S^M(\bar{\eta}_{i-1}^M), \bar{\eta}_{i-1}^M\right) > \beta_2 \nu^{n-i}\big] + P\big[\mathcal{W}_1\left((\bar{\mu}_0^N, \bar{\eta}_0^M), (\mu_0, \eta_0)\right) > \beta_3 \nu^n\big] \qquad (3.4.99) \\
\leq & \ a_3 \sum_{i=1}^{n} \left(e^{-a_2 \frac{N_1 \beta^2 \nu^{2i}}{R^2}} + \frac{R^{-\tau}}{\beta \nu^i} \right) + a_3' \sum_{i=1}^{n} \left(e^{-a_2' \frac{N_1 \beta^2 \nu^{2i}}{R^2}} + \frac{R^{-\tau}}{\beta \nu^i} \right) \\
& + P\big[\mathcal{W}_1\left((\bar{\mu}_0^N, \bar{\eta}_0^M), (\mu_0, \eta_0)\right) > \beta_3 \nu^n\big].
\end{aligned}
$$

Now proceeding similarly like the proof of Theorem 3.7 of Chapter 2 through optimizing the value of R the conclusion will follow.

3.4.7 Proof of Theorem 3.3.5

We will start by introducing a coupling. Consider a system of $\mathbb{R}^d$ valued auxiliary random variables $\{Y_n^{i,N}, i = 1, \ldots, N\}_{n \geq 0}$ defined as follows.

$$
\begin{aligned}
Y_{n+1}^{i,N} &= AY_n^{i,N} + \delta f(\nabla \eta_n(Y_n^{i,N}), \mu_n, Y_n^{i,N}, \epsilon_{n+1}^i) + B(\epsilon_{n+1}^i), \ i = 1, \ldots, N, \quad n \in \mathbb{N}_0. \\
\eta_{n+1} &= \eta_n R_{\mu_n}^{\alpha}, \\
Y_0^{i,N} &= X_0^{i,N}. \qquad (3.4.100)
\end{aligned}
$$

Now for each $n \in \mathbb{N}$, $\{Y_n^{i,N}, i = 1, \ldots, N\}$ is a set of $\mathbb{R}^d$ valued iid random variables under initial assumption $\mathcal{L}(\{X_0^{i,N}\}_{i=1,\ldots,N}) = \mu_0^{\otimes N}$. Suppose $\zeta_n^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{Y_n^{i,N}}$. The following Lemma will make a connection between ζ_n^N and μ_n^N.

Lemma 18. *(Coupling with the auxiliary system) Suppose Assumptions 6,9,10 and 14 hold. Then for every $n \geq 0$ and $N \geq 1$, with the C_1, and χ_1 defined in* (3.3.18),(3.3.19)

$$\mathcal{W}_1(\mu_{n+1}^N, \mu_{n+1}) \leq \mathcal{W}_1(\zeta_{n+1}^N, \mu_{n+1}) + C_1 \sum_{k=0}^{n} \chi_1^{n-k} \mathcal{W}_1(\zeta_k^N, \mu_k). \quad (3.4.101)$$

Proof. Since by Assumption 6 and $A_1(\epsilon) \leq K$, we have for each $j = 1, \ldots, N$

$$\begin{aligned}|X_{n+1}^j - Y_{n+1}^{j,N}| \leq\ & \|A\||X_n^j - Y_n^{j,N}| + \delta K\big\{|\nabla \eta_n^N(X_n^{j,N}) - \nabla \eta_n(Y_n^{j,N})| + |X_n^{j,N} - Y_n^{j,N}| \\ & + \mathcal{W}_1(\mu_n^N, \mu_n)\big\}\end{aligned}$$

Using the calculations in (3.4.46),(3.4.48),(3.4.49) and (3.4.51)

$$\begin{aligned}& |\nabla \eta_n^N(X_n^{j,N}) - \nabla \eta_n(Y_n^{j,N})| \leq |\nabla \eta_n^N(X_n^{j,N}) - \nabla \eta_n^N(Y_n^{j,N})| \\ & + |\nabla \eta_n^N(Y_n^{j,N}) - \nabla \eta_n(Y_n^{j,N})| \\ \leq\ & l_{PP'}^{\nabla,\alpha}|X_n^j - Y_n^{j,N}| + (1-\alpha) l_P^{\nabla} \mathcal{W}_1(\eta_{n-1}^N, \eta_{n-1}) + \alpha l_{P'}^{\nabla} \mathcal{W}_1(\mu_{n-1}^N, \mu_{n-1})\end{aligned}$$

Thus

$$\begin{aligned}|X_{n+1}^{j,N} - Y_{n+1}^{j,N}| \leq\ & \big[\|A\| + \delta K(1 + l_{PP'}^{\nabla,\alpha})\big]|X_n^j - Y_n^{j,N}| + \delta K\Big[\mathcal{W}_1(\mu_n^N, \mu_n) \\ & + (1-\alpha) l_P^{\nabla} \mathcal{W}_1(\eta_{n-1}^N, \eta_{n-1}) + \alpha l_{P'}^{\nabla} \mathcal{W}_1(\mu_{n-1}^N, \mu_{n-1})\Big] \quad (3.4.102)\end{aligned}$$

Using (3.4.102) as the recursion on $a^j_{n+1} := |X^{j,N}_{n+1} - Y^{j,N}_{n+1}|$ with $a^j_0 = 0$, we get

$$\begin{aligned} a^j_{n+1} &\le \delta K \sum_{k=1}^{n} \left[\|A\| + \delta K(1 + l^{\nabla,\alpha}_{PP'})\right]^{n-k} \Big[\mathcal{W}_1(\mu^N_k, \mu_k) + (1-\alpha) l^{\nabla}_P \mathcal{W}_1(\eta^N_{k-1}, \eta_{k-1}) \\ &\quad + \alpha l^{\nabla}_{P'} \mathcal{W}_1(\mu^N_{k-1}, \mu_{k-1}) \Big]. \end{aligned} \tag{3.4.103}$$

Denote $\|A\| + \delta K(1 + l^{\nabla,\alpha}_{PP'})$ by χ. Observe that

$$\mathcal{W}_1(\eta^N_{n-1}, \eta_{n-1}) = (1-\alpha) l(P) \mathcal{W}_1(\eta^N_{n-2}, \eta_{n-2}) + \alpha l(P') \mathcal{W}_1(\mu^N_{n-2}, \mu_{n-2}). \tag{3.4.104}$$

Denote the quantity in the third bracket of RHS of (3.4.103) by b_k. Using (3.4.104) and $\eta^N_0 = \eta_0$ we have

$$\begin{aligned} b_k &= \mathcal{W}_1(\mu^N_k, \mu_k) + (1-\alpha) l^{\nabla}_P \mathcal{W}_1(\eta^N_{k-1}, \eta_{k-1}) + \alpha l^{\nabla}_{P'} \mathcal{W}_1(\mu^N_{k-1}, \mu_{k-1}) \\ &= \mathcal{W}_1(\mu^N_k, \mu_k) + (1-\alpha) l^{\nabla}_P \alpha l(P') \sum_{i=0}^{k-2} [(1-\alpha) l(P)]^{k-2-i} \mathcal{W}_1(\mu^N_i, \mu_i) \\ &\quad + \alpha l^{\nabla}_{P'} \mathcal{W}_1(\mu^N_{k-1}, \mu_{k-1}) \\ &\le c_4 \sum_{i=0}^{k} c_5^{k-i} \mathcal{W}_1(\mu^N_i, \mu_i). \end{aligned} \tag{3.4.105}$$

where $c_4 := \max\{1, (1-\alpha) l^{\nabla}_P \alpha l(P')\}$ and $c_5 := \max\{\alpha l^{\nabla}_{P'}, (1-\alpha) l(P)\}$. Thus from (3.4.103) we have

$$a^j_{n+1} \le \delta K c_4 \sum_{k=0}^{n} \chi^{n-k} \sum_{i=0}^{k} c_5^{k-i} \mathcal{W}_1(\mu^N_i, \mu_i). \tag{3.4.106}$$

Now applying Lemma A.1.4 we have

$$
\begin{aligned}
a_{n+1}^j &\leq \delta K c_4 \sum_{i=0}^{n} \mathcal{W}_1(\mu_i^N, \mu_i) \left[\frac{\chi^{n+1-i} - c_5^{n+1-i}}{\chi - c_5} \right] \\
&\leq \delta K c_7 \sum_{i=0}^{n} \chi_2^{n+1-i} \mathcal{W}_1(\mu_i^N, \mu_i)
\end{aligned} \tag{3.4.107}
$$

where $\chi_2 := \max\{\chi, c_5\}$ and $c_7 := \frac{c_4}{|\chi - c_5|}$. Note that from (3.4.79) we have for all $n \geq 0$,

$$
\mathcal{W}_1(\zeta_n^N, \mu_n^N) \leq \frac{1}{N} \sum_{j=1}^{N} a_n^j.
$$

Combining the result above and using triangle inequality in (3.4.107)

$$
\mathcal{W}_1(\zeta_{n+1}^N, \mu_{n+1}^N) \leq \delta K c_7 \sum_{k=0}^{n} \chi_2^{n+1-k} \mathcal{W}_1(\zeta_k^N, \mu_k^N) + \delta K c_7 \sum_{k=0}^{n} \chi_2^{n+1-k} \mathcal{W}_1(\zeta_k^N, \mu_k).
$$

Applying Lemma A.1.3 with

$$
a_n = \chi_2^{-n} \mathcal{W}_1(\zeta_n^N, \mu_n^N),\ b_n = \delta K c_7 \sum_{k=0}^{n-1} \chi_2^{-k} \mathcal{W}_1(\eta_k^N, \mu_k),\ p_n = \delta K c_7,\ n \geq 0.
$$

We have

$$
\begin{aligned}
\chi_2^{-(n+1)} \mathcal{W}_1(\zeta_{n+1}^N, \mu_{n+1}^N) &\leq b_{n+1} + \sum_{k=0}^{n} (\delta K c_7)^2 \sum_{i=0}^{k-1} \chi_2^{-i} \mathcal{W}_1(\zeta_i^N, \mu_i)\,(1+\delta K c_7)^{n-k} \\
&= b_{n+1} + \sum_{i=0}^{n} \sum_{k=i+1}^{n} (\delta K c_7)^2 (1+\delta K c_7)^{n-k} \chi_2^{-i} \mathcal{W}_1(\zeta_i^N, \mu_i) \\
&= b_{n+1} + \sum_{i=0}^{n} (\delta K c_7)^2 \chi_2^{-i}.\mathcal{W}_1(\zeta_i^N, \mu_i) \sum_{m=0}^{n-i-1} (1+\delta K c_7)^m \\
&= b_{n+1} + \sum_{i=0}^{n} (\delta K c_7) \chi_2^{-i} \mathcal{W}_1(\zeta_i^N, \mu_i)[(1+\delta K c_7)^{n-i} - 1]. \qquad (3.4.108)
\end{aligned}
$$

Simplifying (3.4.108) one gets

$$
\begin{aligned}
\mathcal{W}_1(\zeta_{n+1}^N, \mu_{n+1}^N) &\leq \delta K c_7 \sum_{k=0}^{n} \chi_2^{n+1-k} \mathcal{W}_1(\zeta_k^N, \mu_k) \\
&\quad + \sum_{k=0}^{n} (\delta K c_7) \chi_2^{n+1-k} \mathcal{W}_1(\zeta_k^N, \mu_k)[(1+\delta K c_7)^{n-k} - 1] \\
&= \sum_{k=0}^{n} (\delta K c_7) \chi_2^{n+1-k} \mathcal{W}_1(\zeta_k^N, \mu_k)(1+\delta K c_7)^{n-k}. \\
&= \delta K c_7 \chi_2 \sum_{k=0}^{n} (\chi_2 + \delta K c_7 \chi_2)^{n-k} \mathcal{W}_1(\zeta_k^N, \mu_k).
\end{aligned}
$$

Note that $\delta K c_7 \chi_2 = C_1$ and $\chi_2 + C_1 = \chi_1$ as defined in (3.3.18) and (3.3.19) respectively.

Thus we have

$$
\mathcal{W}_1(\zeta_{n+1}^N, \mu_{n+1}^N) \leq C_1 \sum_{k=0}^{n} \chi_1^{n-k} \mathcal{W}_1(\zeta_k^N, \mu_k).
$$

The result now follows by an application of triangle inequality. $\square$

Proof of Theorem 3.3.5

Since $\chi_1 < 1$. So we can find $\gamma > 0$ such that $\chi_1 < 1-\gamma$. Taking that γ, we have $\nu_1 := \frac{1-\gamma}{\chi_1} > 1$. For any $\varepsilon > 0$, From Lemma 3.3.5

$$\begin{aligned}
& P[\mathcal{W}_1(\mu_n^N,\mu_n) > \varepsilon] \le P[\mathcal{W}_1(\zeta_n^N,\mu_n) > \gamma\varepsilon] \\
& + \sum_{i=0}^{n-1} P[C_1\chi_1^{n-1-i}\mathcal{W}_1(\zeta_i^N,\mu_i) \ge \gamma\varepsilon(1-\gamma)^{n-i}] \\
= \;& P[\mathcal{W}_1(\zeta_n^N,\mu_n) > \gamma\varepsilon] + \sum_{i=1}^{n} P[\mathcal{W}_1(\zeta_{n-i}^N,\mu_{n-i}) \ge \frac{\gamma\varepsilon\chi_1}{C_1}\nu^i] \qquad (3.4.109) \\
= \;& P[\mathcal{W}_1(\zeta_n^N,\mu_n) > \gamma\varepsilon] + \sum_{i=1}^{i_\varepsilon} P[\mathcal{W}_1(\zeta_{n-i}^N,\mu_{n-i}) \ge \frac{\gamma\chi_1\varepsilon}{C_1}\nu^i] \\
& + \sum_{i=i_\varepsilon+1}^{n} P[\mathcal{W}_1(\zeta_{n-i}^N,\mu_{n-i}) \ge \frac{\gamma\chi_1\varepsilon}{C_1}\nu^i],
\end{aligned}$$

where $i_\varepsilon := \max\{i \ge 0 : \frac{\gamma\chi_1\varepsilon}{C_1}\nu^i < 1\}$. Note that for $\delta \in \Big[0, \frac{1-\|A\|}{(2+l_{PP'}^{\nabla,\alpha})K}\Big)$, and $\alpha_1 \in (0, \frac{\alpha(\delta)}{\delta})$, using similar version of Lemma 10 of Chapter 2, we have $\sup_{n\ge 0}\left\langle e^{\alpha_1|x|,\mu_n}\right\rangle < \infty$. That implies from the statement of Theorem 2 of [33] that for all $N > 0$,

$$P[\mathcal{W}_1(\zeta_n^N,\mu_n) \ge \varepsilon] \le a(N,\varepsilon)1_{\{\varepsilon\le 1\}} + b(N,\varepsilon).$$

where $a(N,\varepsilon) = e^{-cN\varepsilon^2}1_{\{d=1\}} + e^{-cN\left(\frac{\varepsilon}{\log(2+\frac{1}{\varepsilon})}\right)^2}1_{\{d=2\}} + e^{-cN\varepsilon^d}1_{\{d>2\}}$ and $b(N,\varepsilon) = e^{-cN\varepsilon}$. In order to prove (3.3.20) we will prove only for one case $d > 2$. Rest will follow similarly. There exists C_1', C_2', C_3'

$$\sum_{i=i_\varepsilon+1}^{n} P[\mathcal{W}_1(\zeta_{n-i}^N, \mu_{n-i}) \geq \frac{\gamma\chi_1\varepsilon}{C_1}\nu^i] \leq \sum_{i=i_\varepsilon+1}^{n} b(N, \frac{\gamma\chi_1\varepsilon}{C_1}\nu^i) \leq \sum_{i=i_\varepsilon+1}^{n} e^{-C_1'\varepsilon N\nu^i} \tag{3.4.110}$$

$$\sum_{i=1}^{i_\varepsilon} P[\mathcal{W}_1(\zeta_{n-i}^N, \mu_{n-i}) \geq \frac{\gamma\chi_1\varepsilon}{C_1}\nu^i] \leq \sum_{i=1}^{i_\varepsilon} a(N, \frac{\gamma\chi_1\varepsilon}{C_1}\nu^i) \leq \sum_{i=1}^{i_\varepsilon} e^{-C_2' N(\varepsilon\nu^i)^d}$$
$$\leq \sum_{i=1}^{i_\varepsilon} e^{-C_2' N\varepsilon^d\nu^i} \tag{3.4.111}$$

$$P[\mathcal{W}_1(\zeta_n^N, \mu_n) > \gamma\varepsilon] \leq e^{-C_3' N\varepsilon^d\wedge\varepsilon} \tag{3.4.112}$$

Suppose k_0 such that $\nu^i \geq k_0 i$ for all $i \geq 1$. Combining (3.4.110),(3.4.111),(3.4.112) we have for all $N > 1$ and $a_2'' = k_0 \min\{C_1', C_2', C_3'\}$.

$$\sup_n P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq \sum_{i=0}^{\infty} e^{-a_2'' N i\varepsilon^d\wedge\varepsilon} \leq \frac{e^{-a_2'' N\varepsilon^d\wedge\varepsilon}}{1 - e^{-a_2'' N\varepsilon^d\wedge\varepsilon}}. \tag{3.4.113}$$

Now there exists $N_3 := -\frac{1}{a_2''}\log(1 - \frac{1}{a_1''})$ such that $N \geq N_3 \max\{\frac{1}{\varepsilon}, \frac{1}{\varepsilon^d}\}$ we have

$$\sup_n P[\mathcal{W}_1(\mu_n^N, \mu_n) > \varepsilon] \leq a_1'' e^{-a_2'' N\varepsilon^d\wedge\varepsilon}.$$

□

Chapter4
HIGHER ORDER ASYMPTOTICS OF GFI

4.1 Introduction

This chapter renders application of higher order asymptotics to Fiducial based inferential techniques. In last decades there had been a surge of parallel endeavors in modern modifications of fiducial inference. These approaches are well known under a common name: *"Distributional inference"*. Main emphasis for these approaches was defining inferentially meaningful probability statements about subsets of the parameter space without the need for subjective prior information. They include the *"Dempster Shafer theory"* (Dempster [28]; Edlefsen, Liu and Dempster [30]) and its relatively recent extension called inferential models (Martin, Zhang and Liu [64]; Zhang and Liu [83]; Martin and Liu [63], [61], [62]). There is another rigorous framework available called *Objective Bayesian inference* that aims at finding nonsubjective model based priors. An example of a recent breakthrough in this area is the modern development of reference priors (Berger, [4]; Berger and Sun [7]; Berger, Bernardo and Sun [5],[6]; Bayarri et al. [3]). Another related approach is based on higher order likelihood expansions and implied data dependent priors (Fraser, Fraser and Staicu [34]; Fraser [35],[36]; Fraser and Naderi [37]; Fraser et al. [38]; Fraser, Reid and Wong [39]). A different frequentist approach namely confidence distributions looks at the problem of obtaining an inferentially meaningful distribution on the parameter space (Xie and Singh [81]). Recently an interesting work (Taraldsen and Lindqvist [75]) showing how some simple fiducial distributions that are not Bayesian posteriors naturally arises within the decision theoretical framework.

Arguably, Generalized Fiducial Inference has been on the forefront of the modern fiducial revival. The strengths and limitations of the fiducial approach are starting to be better understood; see especially Hannig [51], [52]. In particular, the asymptotic exactness of fiducial confidence sets, under fairly general conditions, was established in Hannig [52]; Hannig, Iyer and Patterson [53]; Sonderegger and Hannig [71].

*Main aim of this chapter is to further study exactness property of the Fiducial quantile in frequentist sense for uni-parameter cases with exploration of higher order asymptotics. From a different point of view it can be seen as a prudent way of selecting a data generating equation (to be defined shortly) so that the non-uniqueness issue of proposing Generalized Fiducial Distribution (in short **GFD**) can be reduced partially.* To start with we address what we mean by **GFD**.

Denote the parameter space by Θ. Let the data $\mathbf{X}$ be a $\mathbf{S}$ valued random variable. It starts by expressing a relationship between the parameter and the data through a deterministic function $G : \mathbf{M} \times \Theta \to \mathbf{S}$ which we call by *data generating equation* (in short **DGE**):

$$\mathbf{X} = \mathbf{G}(\mathbf{U}, \theta) \tag{4.1.1}$$

where $\mathbf{U}$ is a $\mathbf{M}$ valued random variable whose distribution doesn't depend on θ. The distribution of the data $\mathbf{X}$ is determined by $\mathbf{U}$ via the structural equation (4.1.1). That is one can generate $\mathbf{X}$ by generating $\mathbf{U}$ and plugging it into the data generating equation.

For example for one sample of $N(\theta, 1)$ the data generating structural equation is

$$G(U, \theta) = \theta + \Phi^{-1}(U)$$

where $\Phi(.)$ cumulative Normal distribution function and $U \sim U(0, 1)$. One can always find (4.1.1) by following construction. For a realization $\mathbf{x}_0 := (x_1, x_2, \ldots x_n)$ of $\mathbf{X}$ where $\mathbf{X} \sim F_\theta(\cdot)$ for F_θ being a distribution function on $\mathbb{R}^n$ with $\theta \in \Theta$ being the unknown parameter denote the conditonal distributions of first, second and n-th co-ordinate (sequentially given the rest) by $F_{\theta, X_1}(\cdot)$, $F_{\theta, X_2|X_1}(\cdot)$, and $F_{\theta, X_n|(X_1, X_2, \ldots, X_{n-1})}(\cdot)$ respectively. Then (4.1.1) can be written as

$$\begin{aligned} x_1 &= F^{-1}_{\theta, X_1}(U_1) \\ x_2 &= F^{-1}_{\theta, X_2 \big| \{X_1 = x_1\}}(U_2) \\ x_n &= F^{-1}_{\theta, X_n \big| \{(X_1, X_2, \ldots, X_{n-1}) = (x_1, x_2, \ldots x_{n-1})\}}(U_n) \end{aligned} \tag{4.1.2}$$

where $(U_1, U_2, \ldots, U_n)$ iid copies of Uniform $(0, 1)$ random variables. Note that in above illustration changing the order of the variables $(X_1, \ldots, X_n)$ will give different data generating equations unless $X_1, X_2, \ldots, X_n$ are independent.

After observing $\mathbf{x}_0$, given U, define the inverse image $Q_{\mathbf{x}_0}(U)$ in the parameter space from the data generating equation as

$$Q_{\mathbf{x}_0}(U) := \{\theta : G(U, \theta) = \mathbf{x}_0\}.$$

after observing $\mathbf{X} = \mathbf{x}_0$, fiducial approach instructs us to deduce a distribution for θ from the randomness of U via the structural equation (i.e., generate $U^\star$ and invert the structural equation solving for $\theta = Q_{\mathbf{x}_0}(U^\star)$). Now in order to remove the possibility of non-existence of solution for some U^*, we will discard that value, i.e, condition the distribution $Q_{\mathbf{x}_0}(U)$ given the fact the solution always exists (i.e $Q_{\mathbf{x}_0}(U) \neq \emptyset$). As explained, the Fiducial distribution of θ given observed $\mathbf{x}_0$ should be heuristically (hence ill defined) the following conditional distribution

$$Q_{\mathbf{x}_0}(U^*) \Big| \left\{Q_{\mathbf{x}_0}(U^*) \neq \emptyset\right\}. \tag{4.1.3}$$

Immediately three relevant questions arise regarding the non-uniquenesses of Generalized Fiducial distribution (4.1.3):

- **The choice among multiple solutions:** It arises if the inverse image $Q_{\mathbf{x_0}}(U^*)$ has more than one element for U^* and observed $\mathbf{x_0}$. This problems mainly occur in discrete distributions (See [52]) which we did not consider in this article. Moreover in asymptotic regime all those choices will lead to the same distribution limit.
- **Borel Paradox:** Another important problem regarding computing the conditional probability in (4.1.3) arises when the conditioning event $\{Q_{\mathbf{x}_0}(U^*) \neq \emptyset\}$ has measure

0. For example, suppose one observes $\mathbb{X} = \mathbf{x}_0 := (x_1, \ldots, x_n)'$ where $\mathbb{X}_{n\times 1} \sim N_n(\theta.1_{n\times 1}, \mathbf{I}_{n\times n})$ for $\Theta = \mathbb{R}$. Note that, considering the simple data generating equation $\mathbb{X} = \theta + \mathbb{U}_{n\times 1}$ where $\mathbb{U} \sim N_n(0, I_n)$, the inverse image

$$Q_{\mathbf{x}_0}(\mathbf{U}^*) = (x_1 - U_1^*).1_{\{U_2^* - U_1^* = x_2 - x_1, U_3^* - U_1^* = x_3 - x_1, \ldots, U_n^* - U_1^* = x_n - x_1\}}$$

and the set $\{Q_{\mathbf{x}_0}(\mathbf{U}^*) \neq \emptyset\}$ has probability 0 ($n-1$ dimensional manifold in n dimensional space). In that case the conditional probability distribution may not remain unique which in literature is termed as the Borel paradox.

This problem can also be removed by defining the **GFD** as the distribution of the weak limit of the following quantity (in the display) conditioned on the event $\left\{\inf_\theta \left\|\mathbf{x_0} - G(U^*, \theta)\right\| \leq \epsilon\right\}$ as $\epsilon \downarrow 0$.

$$\arg\inf_\theta \left\|\mathbf{x_0} - G(U^*, \theta)\right\| \Bigg| \left\{\inf_\theta \left\|\mathbf{x_0} - G(U^*, \theta)\right\| \leq \epsilon\right\}. \tag{4.1.4}$$

Let's assume that for each fixed $\theta \in \Theta$ the function $G(\cdot, \theta)$ is one-to-one and continuously differentiable denoting the inverse by $G^{-1}(x, \theta)$. If we use L_∞ norm as $\|\cdot\|$ in the definition of 4.1.4, from Theorem 3.1 of [52] it follows that the unique weak limit is a distribution on Θ with density

$$f^{\mathcal{G}}\left(\theta \middle| \mathbf{x_0} = \{X_1, X_2, \ldots, X_n\}\right) = \frac{f_{\mathbb{X}}\left(\mathbf{x_0} \middle| \theta\right) J_n(\mathbf{x_0}, \theta)}{\int_{\mathbb{R}} f_{\mathbb{X}}\left(\mathbf{x_0} \middle| \theta'\right) J_n(\mathbf{x_0}, \theta') d\theta'}, \tag{4.1.5}$$

where the Jacobian $J_n(\mathbf{x},\theta)$ is defined as

$$J_n(\mathbf{x},\theta) :\propto \sum_{i=(i_1,\ldots,i_p)} \left| \det\left(\frac{d}{d\theta}\mathbf{G}(\mathbf{u},\theta)\Big|_{\mathbf{u}=\mathbf{G}^{-1}(\mathbf{x},\theta)} \right)_{(i)} \right| \tag{4.1.6}$$

when θ is p dimensional. $A_{(i)}$ denotes the $p \times p$ submtrix consisting of ith selection of rows of A. Let $\mathbb{X} = (X_1,\ldots,X_n)$. In the uniparameter $(p = 1)$ case, the Jacobian becomes

$$J_n(\mathbb{X},\theta) \propto \sum_{i=1}^{n} \left| \frac{\partial}{\partial\theta} G(U_i,\theta)\Big|_{U_i=G^{-1}(X_i,\theta)} \right|. \tag{4.1.7}$$

In what follows by Fiducial distribution (or density) we will mean the distribution(or density) of θ defined in (4.1.4). Denote the probability distribution on Θ induced by the data generating equation $\mathcal{G}$ in (4.1.1) by $P^{\mathcal{G}}(\cdot)$ whose density is (4.1.5).

As an example consider $\mathbb{X}$ is a sample of n iid observations from **Scaled Normal family** $N(\mu,\mu^q), \mu \in \mathbb{R}^+ := \Theta$. The simplest data generating equation comes from the relation $\mathbb{X} = G(\mathbb{U},\theta) := \mu + \mu^{\frac{q}{2}}\mathbb{U}$, where $\mathbb{U} = (U_1,U_2,\ldots,U_n)$ is an array of n i.i.d $N(0,1)$ random variables. Since $\frac{d}{d\mu}G(U_i,\mu) = 1 + \frac{q}{2}\mu^{q/2-1}U_i$, then plugging in $U_i = \frac{x_i-\mu}{\mu^{q/2}}$, one gets the Jacobian in (4.1.6)

$$J(\mathbf{x},\mu) \propto \sum_{i=1}^{n} \left| \frac{\partial}{\partial\mu} G(U_i,\mu)\Big|_{U_i=G^{-1}(X_i,\mu)} \right| = \sum_{i=1}^{n} \left| 1 + \frac{q(X_i-\mu)}{2\mu} \right| \tag{4.1.8}$$

giving the **GFD** as specified in (4.1.5).

In context of non-informative prior for any one-to-one function $\phi(.)$, inference of θ given

$\mathbb{X}$ and inference of $\phi(\theta)$ given $\mathbb{X}$ should not be different since the *ideal non-informative prior should not impose any extra information on* Θ [73]. Just like Posterior distribution of **Jeffrey's prior**, Generalized Fiducial distribution of θ exhibits this **parametrization invariance** property. For some one-one function ϕ suppose $\eta = \phi(\theta)$. The Fiducial distributions of θ and η denoted respectively as r and $\tilde{r}$ follow

$$\int_A r(\theta)d\theta = \int_{\phi^{-1}(A)} \tilde{r}(\eta)d\eta$$

and it follows trivially from the definition (4.1.4).

Choice of L_∞ norm in (4.1.4) can be intuitively justified from natural discretization point of view. That is for the observed $\mathbf{x}$ one can think a box (an interval for each dimension) around it, in order to get a $\theta \in \left\{\theta : \inf_\theta \left\|\mathbf{x_0} - G(U^*, \theta)\right\| \leq \epsilon\right\}$. Also another reason is independence among the dimensions which helps in the calculation.

4.1.1 The Choice of Structural Equations:

Note that changing norm in (4.1.4) leads to different fiducial distributions [52]. Now we illustrate another issue of non-uniqueness arising from different choices of data generating equation $G(U, \theta)$. Let $\mathbb{X} = (X_1, \ldots, X_n)$ be n iid realizations from a distribution $F(\cdot \mid \theta)$ with density $f(\cdot \middle| \theta)$ parametrized by one dimensional parameter $\theta \in \Theta := \mathbb{R}$ where the true parameter value is θ_0. Suppose $T(.)$ is a smooth, one to one transformation and there exists a weight function $w(x)$ such that $T(x) := \int_{(-\infty,x]} w(y)dy$. Now instead of considering the data generating equation $X = G(U, \theta)$ let's consider $Y := T(X) = T \circ G(U, \theta)$. Then using $\|\cdot\| = L_\infty$ norm in (4.1.4), from (4.1.5) the density of fiducial distribution of θ given

$\mathbb{Y} := (Y_1, Y_2, \ldots, Y_n) = (T(X_1), T(X_2), \ldots, T(X_n)) := B(\mathbb{X})$ will be

$$\begin{aligned} f^{\mathcal{G}}(\theta \big| \mathbb{Y} = \mathbf{y}) &= \frac{f_{\mathbb{Y}}\left(\mathbf{y}\big|\theta\right) J_n(\mathbf{y}, \theta)}{\int_\Theta f_{\mathbb{Y}}\left(\mathbf{y}\big|\theta'\right) J_n(\mathbf{y}, \theta') d\theta'} \\ &= \frac{\prod_{i=1}^n f\left(T^{-1}(\mathbf{y}_i)\big|\theta\right) \left|(T^{-1})'(\mathbf{y}_i)\right| J(\mathbf{y}, \theta)}{\int_\Theta \prod_{i=1}^n f\left(T^{-1}(\mathbf{y}_i)\big|\theta\right) \left|(T^{-1})'(\mathbf{y}_i)\right| J(\mathbf{y}, \theta') d\theta'} \\ &= \frac{\prod_{i=1}^n f\left(\mathbf{x}_i\big|\theta\right) J\left(B(\mathbf{x}), \theta\right)}{\int_\Theta \prod_{i=1}^n f\left(\mathbf{x}_i\big|\theta'\right) J(B(\mathbf{x}), \theta') d\theta'} \end{aligned} \tag{4.1.9}$$

where using (4.1.7)) and the chain rule, resulting Jacobian $J(B(X), \theta)$ will be proportional to

$$\begin{aligned} \sum_{i=1}^n \left| \frac{\partial}{\partial\theta} T \circ G(U_i, \theta) \Big|_{U_i = G^{-1}(X_i, \theta)} \right| &= \sum_{i=1}^n \left| \left[T'(G(U_i, \theta)) \frac{\partial}{\partial\theta} G(U_i, \theta) \right] \Big|_{U_i = G^{-1}(X_i, \theta)} \right| \\ &= \sum_{i=1}^n \left| w(X_i) \frac{\partial}{\partial\theta} G(U_i, \theta) \Big|_{U_i = G^{-1}(X_i, \theta)} \right|. \end{aligned} \tag{4.1.10}$$

the weight function $w(.)$ in the Jacobian. Note that (4.1.9) comes from last expression of $f^{\mathcal{G}}(\theta\big|\mathbb{Y})$ where $(T^{-1})'(\mathbf{y}_i)$ cancels out from both numerator and denominator which is possible since $|(T^{-1})'(Y)|$ is almost surely assumed to be non zero. Intuitively fiducial distribution changes due to the deformation of the $\mathbf{x}$ space through the transformation T along with change in the conditional event $\inf_\theta \left\| T(\mathbf{x}) - T \circ G(U^*, \theta) \right\| \leq \epsilon$ in the definition (4.1.4).

Note that given $F(\cdot, \theta)$ being the distribution function with density $f(\cdot, \theta)$, the following illustration by chain rule yields

$$\frac{\partial}{\partial\theta} G(U, \theta) \Big|_{\{U = G^{-1}(X, \theta)\}} = \frac{\partial}{\partial\theta} X = \frac{1}{\frac{dF(X,\theta)}{dX}} \frac{\partial}{\partial\theta} F(X, \theta) = \frac{\frac{\partial F(X,\theta)}{\partial\theta}}{f(X, \theta)}. \tag{4.1.11}$$

Under the transformation $T(\cdot) = \int_{(-\infty,\cdot]} w(y)dy$ the transformed Jacobian in (4.1.10) can be expressed as

$$J_n(T(X), \theta) \propto \sum_{i=1}^{n} \left| w(X_i) \frac{\frac{\partial F_i(X_i,\theta)}{\partial \theta}}{f(X_i, \theta)} \right|. \tag{4.1.12}$$

So we see, if we consider different data generating equation as a smooth and one-one transformation of the original structural equation it only brings a weight $w(.)$ in the Jacobian part. The question of interest is, *what is an "ideal" transformation $T(.)$ for which the Fiducial distribution enjoys some "desirable" properties.* By **desirability** we mean a data generating structural equation under which the Fiducial distribution is "exact" in a frequentist sense.

Structure of this work follows with initially giving an ideal recipe of Fiducial distribution for a special case when some strong monotonicity conditions of the structural equation are satisfied. In the absence of those conditions, we define a criteria through higher order asymptotics. Throughout this article we considered $|| \cdot || = L_\infty$ norm in the definition of (4.1.4) and the main contention of this article is to get the ideal transformation $T(.)$ under which the data generating equation "$T(\mathbf{X}) = T \circ G(\mathbf{U}, \theta)$" will give first order probability matching Fiducial distribution (to be defined in (4.3.1)). Then we will conclude this article illustrating with some illustrations with their empirical behaviors in small sample situations.

4.2 Why Fisher Might Have Thought Fiducial Distribution Exact?

Fisher developed [31] the Fiducial idea based on the minimal sufficient statistics. Under the same motivation we state the following theorem considering $G_S(\mathbf{U}, \theta)$ as the data generating equation for a one dimensional minimal sufficient statistics S. Denote the inverse image $Q_{\mathbf{s}}(\mathbf{u}) = \{\theta : \mathbf{s} = G_S(\mathbf{u}, \theta)\}$. To study the properties of the generalized fiducial distribution based on (4.2.4) we need to study the conditional distribution of

$$Q_{\mathbf{s}}(\mathbf{U}) \mid \{Q_{\mathbf{s}}(\mathbf{U}) \neq \emptyset\}$$

where as for S having a continuous distribution on $\mathbb{R}$, the conditioning event will be of measure 1 so one can ignore that.

Theorem 2. *Let us assume that for all $\mathbf{U}$, the function $G_S(\mathbf{u}, \theta)$ is non-decreasing in θ and for all $\mathbf{u}$ and θ we have $Q(s, \mathbf{u}) \neq \emptyset$. Then the inverse image $Q(s, \mathbf{u})$ is an interval with bounds $Q_s^-(\mathbf{u}) \leq Q_s^+(\mathbf{u})$. Additionally, for any s_0 and θ_0 the bounds satisfy $P(Q_{s_0}^+(\mathbf{u}) \leq \theta_0) = 1 - \lim_{\epsilon \downarrow 0} F_S(s_0, \theta_0 + \epsilon)$ and $P(Q_{s_0}^- \leq \theta_0) = 1 - \lim_{\epsilon \downarrow 0} F_S(s_0 - \epsilon, \theta_0)$. Finally if for all θ_0 and s_0 the $P_{\theta_0}(S = s_0) = F_S(s_0, \theta_0) - \lim_{\epsilon \downarrow 0} F_S(s_0 - \epsilon, \theta_0) = 0$ then the cdf $F_S(s, \theta)$ is continuous in θ, $Q_{s_0}^+(\mathbf{U}) = Q_{s_0}^-(\mathbf{U})$ with probability one, and*

$$P(Q_{s_0}(\mathbf{U}) \leq \theta_0) = 1 - F_S(s_0, \theta_0). \tag{4.2.1}$$

Proof of Theorem 2: The fact that $Q(s, \mathbf{u})$ is an interval follows by monotonicity. Consider an iid sample $\mathbf{U}_1, \ldots, \mathbf{U}_n$. By SLLN we have

$$\begin{aligned} P(Q_+(s_0, \mathbf{U}) < \theta_0) &= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_{\{Q_+(s_0, \mathbf{U}_i) < \theta_0\}} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_{\{G_S(\mathbf{U}_i, \theta_0) > s_0\}} \\ &= 1 - F_S(s_0, \theta_0), \end{aligned} \tag{4.2.2}$$

where the second equality follows from monotonicity of G_S. The rest of the computation follows. Similarly

$$\begin{aligned} P(Q_-(s_0, \mathbf{U}) > \theta_0) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_{\{Q_-(s_0,\mathbf{U}_i)>\theta_0\}} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I_{\{G_S(\mathbf{U}_i,\theta_0)<s_0\}} \\ &= \lim_{\epsilon\downarrow 0} F_S(s_0 - \epsilon, \theta_0). \end{aligned} \tag{4.2.3}$$

Finally, if $\lim_{\epsilon\downarrow 0} F_S(s_0, \theta_0 + \epsilon) - F_S(s_0, \theta_0) < 0$ then $P_{\theta_0}(S = s_0) > 0$. The rest of the proof follows by simple comparison. □

Remark 12. *If $G_S(\mathbf{u}, \theta)$ is non-increasing in θ then one has $\lim_{\epsilon\downarrow 0} F_S(s_0, \theta_0+\epsilon)-F_S(s_0, \theta_0) \geq 0$.*

Now we will generalize Theorem 2 beyond the existence of 1-dimensional sufficient statistics under the following assumption:

Assumption 15. *Let us consider a data generating equation $\mathbf{X} = \mathbf{G}(\mathbf{U}, \theta)$. Assume that there exists a one-one $\mathcal{C}^1$ transformation $(S(\mathbf{X}), A(\mathbf{X}))$ on n dimensional $\mathbf{X}$, such that $S(\mathbf{X})$ is one dimensional and $A(\mathbf{X})$ is a vector of ancillary statistics (of θ) of $(n-1)$ dimensions.*

From definition of ancillary statistics the derivative $\frac{\partial}{\partial\theta} A(\mathbf{G}(\mathbf{U}, \theta)) = \mathbf{0}$. After the transformation (S, A) on the initial data generating equation $\mathbf{X} = \mathbf{G}(\mathbf{U}, \theta)$, the new one can be written as

$$\mathbf{s} = G_S(\mathbf{U}, \theta) := S \circ G(\mathbf{U}, \theta), \quad \text{and} \quad \mathbf{a} = G_A(\mathbf{U}, \theta) := A \circ G(\mathbf{U}, \theta). \tag{4.2.4}$$

Thus by using the chain rule and some simple calculus the Jacobian based on (4.2.4) is

$$J((S,A)(\mathbf{X}),\theta) = \left| \sum_{i=1}^{n} \frac{dS(\mathbf{X})}{dX_i} \frac{dG_i(\mathbf{U},\theta)}{d\theta} \Big|_{\mathbf{U}=\mathbf{G}^{-1}(\mathbf{X},\theta)} \right| + 0. \tag{4.2.5}$$

This means the weight we can consider is $w(x_i) = \frac{dS(\mathbf{X})}{dX_i}$. Here $\mathbf{U}$ is a random variable or vector with a known distribution independent of the parameter $\theta \in \mathbb{R}$. Denote the CDF of $S(\mathbf{X})$ and the conditional distribution of $S(\mathbf{X}) \mid \{A(\mathbf{X}) = a\}$ by $F_S(\cdot,\theta)$ and $F_{S|a}(\cdot,\theta)$ respectively. Note that the fiducial density of θ given $\{(S,A)(\mathbf{X})\}$ is proportional to

$$J((S,A)(\mathbf{X}),\theta) f(\mathbf{X},\theta) \propto J((S,A)(\mathbf{X}),\theta) f_{S|A}(\mathbf{s},\theta) \tag{4.2.6}$$

since $f_A(\mathbf{a},\theta)$ does not depend on θ. Let $G_{S|\mathbf{a}} := G_S(\mathbf{u},\theta)\Big|_{\{u:\, G_A(u,\theta)=\mathbf{a}\}}$ and $\mathbf{U_a}$ is a realization of the conditional distribution $\mathbf{U} \mid \{A(\mathbf{X}) = \mathbf{a}\}$.

Corollary 4.2.1. *Suppose Assumption 15 and all conditions under which cdf $F_S(s,\theta)$ is continuous in θ hold. Suppose $G_S(\mathbf{u},\theta)$ is non-decreasing in θ and for all $\mathbf{u}$ and θ. Then the generalized fiducial inference based on the data generating equation*

$$S = G_{S|\mathbf{a}}(\mathbf{U_a},\theta)$$

is exact.

Proof. From (4.2.5) and (4.2.6) the Corollary will follow from Theorem 2 if we can show that the **GFD** computed from $(\mathbf{s},\mathbf{a}) = (G_S(\mathbf{U},\theta), G_A(\mathbf{U},\theta))$, is same as $\mathbf{s} = G_{S|\mathbf{a}}(\mathbf{U_a},\theta)$.

Note that

$$\{u : \mathbf{s} = G_S(u,\theta), \mathbf{a} = A \circ G(u,\theta)\} = \{u : \mathbf{s} = G_{S|\mathbf{a}}(u,\theta)\}. \tag{4.2.7}$$

Now we will show that the Jacobian based on RHS of (4.2.7) and is same as $J((S(\mathbf{X}),\mathbf{a})),\theta)$. A consequence of (4.2.7) is

$$\begin{aligned}
\frac{\partial}{\partial\theta} G_{S|\mathbf{a}}(U_{\mathbf{a}},\theta)\Big|_{U_{\mathbf{a}}=G^{-1}_{S|\mathbf{a}}(\mathbf{s},\theta)} &= \frac{\partial}{\partial\theta} G_S(U,\theta)\Big|_{U=G_S^{-1}(\mathbf{s},\theta),\mathbf{a}=A\circ G(U,\theta)} \\
&= S'(\mathbf{X})\frac{\partial}{\partial\theta} G(U,\theta)\Big|_{U=G^{-1}(\mathbf{X},\theta),A(\mathbf{X})=\mathbf{a}} \\
&= J((S(\mathbf{X}),\mathbf{a})),\theta).
\end{aligned}$$

Since $G_S(\mathbf{u},\theta)$ is non-decreasing in θ for all $\mathbf{u}$, so will be $G_{S|\mathbf{a}}(\mathbf{u},\theta)$. As a consequence of the Theorem 2 we notice that the generalized fiducial inference is not affected by the choice of the data generating equation $G_{S|\mathbf{a}}$. *Moreover if the distribution of S is continuous, then the inference based on $S(\mathbf{X})$ conditional on the manifold $\{A(\mathbf{X}) = a\}$ has exact frequentist properties.* This can be seen from (4.2.1) and the fact that if S conditioned on $A(\mathbf{X}) = a$ has been generated using θ_0 then $1 - F_{S|a}(S,\theta_0) \sim U(0,1)$. Consequently all one sided CI will have stated coverage. $\square$ $\square$

4.2.1 Examples:

Here we will judge a few examples of univariate Fiducial distribution and it's exactness.

1. **Location family**: If $\mathbb{X} = (X_i, \ldots, X_n)$ iid from distribution $\mathcal{P}_\theta$ with density function is $f(x,\theta) = f(x-\theta), x \in \mathbb{R}, \Theta = \mathbb{R}$ where $f(\cdot)$ is a probability density function, the

following one-one transformation

$$T : \mathbb{X} \to (\bar{X}_n, (X_1 - \bar{X}_n, \ldots, X_{n-1} - \bar{X}_n)) := (S(\mathbb{X}), A(\mathbb{X}))$$

is one-one and $\mathcal{C}^1$. By the corollary we have the Generalized Fiducial distribution based on $S(\mathbb{X}) \mid A(\mathbb{X})$ is exact. For **Scale family** (with density function is $f(x,\theta) = \frac{1}{\theta}f(\frac{x}{\theta}), x \in \mathrm{R}, \Theta = \mathbb{R}^+$ where $f(.)$ is any probability density function) using the following one-one transformation

$$T_1 : \mathbb{X} \to \left(\tilde{X}_n, \left(\frac{X_1}{\tilde{X}_n}, \ldots, \frac{X_{n-1}}{\tilde{X}_n}\right)\right) := (S_1(\mathbb{X}), A_1(\mathbb{X}))$$

where $\tilde{X}_n$ = Geometric mean of $\mathbb{X}$, one can conclude that the fiducial distribution based on $\tilde{X}_n$ conditional on $A_1(\mathbb{X})$ is exact. For examples, consider $\mathcal{P}_\theta = U(\theta, \theta+1)$, $\mathcal{P}_\theta =$ Cauchy with location parameter μ, or $\mathcal{P}_\theta =$ Cauchy with scale parameter σ etc.

2. **Exponential Family:** Consider the natural parametrization $f_X(\mathbb{X}|\eta) = h(\mathbb{X})e^{\eta S(\mathbb{X}) - A_1(\eta)}$ of exponential family with ne parameter η. Since $S(\mathbb{X})$ is complete sufficient of η, by Basu's Theorem $S(\mathbb{X})$ and $A(\mathbb{X}) := \mathbb{X} \mid \{S(\mathbb{X}) = \mathrm{s}\}$ are independent (since $A(\mathbb{X})$ ancillary for η) implying the Fiducial distribution based on $S(\mathbb{X})$ unconditionally will attain exactness from Corollary 4.2.1. A further generalization is following: For a one parameter exponential family, with parameter η suppose the density

$$f_X(x|\eta) = h(x).e^{\eta S(x) - A_1(\eta)}$$

where $S(x)$ is a smooth sufficient statistic for $\eta \in \Theta$, and the support of the density (denoted as domain(X)) doesn't depend on η. Now in order to impose monotonicity condition of Theorem 2 "$G_S(u, \theta)$ *is monotone in* θ *for all* u" (or equivalently $F(x, \eta)$ in η) we see

$$\frac{\partial F(x,\eta)}{\partial \eta} = \int_{dom(X)\cap(-\infty,x]} (S(y) - A_1'(\eta)) f_X(y|\eta) dy = E_\eta[S(X)1_{\{X\le x\}}] \\ -A_1'(\eta)F(x,\eta)$$

should not alter its sign in Θ for every x. So the weight $w(\cdot) = S'(\cdot)$ will lead to exact fiducial quantile under following conditions:

- if $S(x)$ is smooth, and
- if $E_\eta[S(X)1_{\{X\le x\}}] - A_1'(\eta)F(x,\eta)$ doesn't change its sign.

First we take the instance of Gamma$(\theta, 1), (\theta \in \mathbb{R}^+)$ family with the density function $f_\theta(x) = \frac{e^{-x}x^{\theta-1}}{\Gamma(\theta)}.1_{\{x>0\}}$. As evident from the density structure the minimal sufficient statistics $\log X$ is smooth and one-one in $\mathbb{R}^+$. The quantity $E_\theta\left[\log X.1_{\{X\le x\}}\right] - \Gamma'(\theta).P_\theta\left(X \le x\right)$ doesn't change sign in θ for all $x > 0$. Using the weight $w(x) = \frac{1}{x}$ one will get a fiducial distribution which is exact.

A very similar results on exponential family were also derived in [78]. Following table accounts some other examples under the same umbrella, for which the Generalized Fiducial distribution based on $S(X)$ is exact.

Family: $\mathcal{P}_\theta, \quad \theta \in \Theta := \mathbb{R}^+$	$\frac{\partial F(x,\theta)}{\partial \theta}$ (maintains sign in θ)	$S'(X)$
Exponental (with mean θ)	$-xe^{-\frac{x}{\theta}}$	Constant
Weibull with scale θ (with known k)		
density: $f_\theta(x) = \frac{k}{\theta}(\frac{x}{\theta})^{k-1}e^{-(\frac{x}{\theta})^k}1_{\{x\geq 0\}}$	$-x^k e^{-\frac{x^k}{\theta^k}}$	X^{k-1}
Pareto with $f_\alpha(x) = \frac{\theta x_m^\theta}{x^{\theta+1}}1_{\{x \geq x_m\}}$		
(x_m known)	$-\log\left(\frac{x}{x_m}\right)\left(\frac{x}{x_m}\right)^{-\theta}$	$\frac{1}{X}$

In the remaining part we will present a few examples where Fiducial distribution is not exact.

3. **Scaled normal family:** $N(\mu, \mu^q)$ **(**$\mu \in \mathbb{R}^+ := \Theta$**,** $q > 0$ **known)**. Exactness for case $q = 2$ corresponds to the scale family of Example (A). For $q \neq 2$, since the minimal sufficient statistics $S(\mathbb{X}) = (\sum X_i, \sum X_i^2)$ is 2 dimensional so will be $G_S(U, \theta)$. Here conclusion of Corollary 4.2.1 will not hold, since one cannot have $(n-1)$ dimensional ancillary statistics vector $A(\mathbb{X})$ for which $(S(\mathbb{X}), A(\mathbb{X}))$ is one-one function on $\mathbb{X}$.

   For sample size $n > 1$, we calculated that the fiducial distribution based on the simplest data generating equation in (4.1.8). Two more choices are

$$J_2(\mathbf{x}, \mu) = \left|1 + \frac{q(\bar{x} - \mu)}{2\mu}\right| + \frac{qs_n}{2\mu}, \quad \text{and} \quad J_3(\mathbf{x}, \mu) = 2\bar{x}_n \left|1 + \frac{q(\bar{x} - \mu)}{2\mu}\right| + \frac{q^2 s_n^2}{\mu}$$

which are based on the following data generating equations respectively:

$$\begin{aligned}(\bar{X}_n, S_n) &= \left(\mu + \mu^{q/2} Z, \mu^{q/2} \left(\frac{U}{n-1}\right)^{1/2}\right), \quad \text{and} \\ (\bar{X}_n^2 \text{sgn}(\bar{X}_n), qS_n^2) &= \left((\mu + \mu^{q/2} Z)^2 \text{sgn}(\mu + \mu^{q/2} Z), q\mu^q \frac{U}{n-1}\right) \quad (4.2.8)\end{aligned}$$

where $\mathcal{L}(Z, U) = N(0, \frac{1}{n}) \otimes \chi^2_{n-1}$ and $\text{sgn}(x) = 1_{[x>0]} - 1_{[x<0]}$. Calculation of $J_3(\mathbf{x}, \mu)$ is possible almost surely since measure of non-differentiability of $G_3(u, \mu)$ in μ (due to $\text{sgn}(x)$) is 0.

4. **Correlation coefficient** $\rho \in \Theta := (-1, 1)$ **of a Bivariate normal model:** Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. $\text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$. Goal is to find the Generalized Fiducial distribution of ρ. Here we will propose a few data generating structural equations:

 Simple Proposal: Taking the simplest data generating equation:

$$(X_i, Y_i) = (U, \rho U + \sqrt{1 - \rho^2} V)$$

 where $\mathcal{L}(U, V) = N(0, 1) \otimes N(0, 1)$ or X, Y flipped.

$$J_n^{(1)}((\mathbb{X}, \mathbb{Y}), \rho) = \frac{\sum_{i=1}^n |X_i - \rho Y_i| + |\rho X_i - Y_i|}{n(1 - \rho^2)}.$$

 Other proposals: We will construct the data generating equations based on the minimal sufficient statistics. Denote $V_1 := \frac{1}{2n} \sum_{i=1}^n (X_i + Y_i)^2 = (1 + \rho) U_1, V_2 :=$

$\frac{1}{2n}\sum_{i=1}^{n}(X_i - Y_i)^2$, where $\mathcal{L}(U_1, U_2) = \frac{\chi^2_n}{n} \otimes \frac{\chi^2_n}{n}$. With the data generating equations

$$
\begin{aligned}
(V_1, V_2) &= ((1+\rho)U_1, (1-\rho)U_2) \qquad \text{and} \\
\left(\frac{1}{V_1}, \frac{1}{V_2}\right) &= \left(\frac{1}{(1+\rho)U_1}, \frac{1}{(1-\rho)U_2}\right)
\end{aligned}
\tag{4.2.9}
$$

from (4.1.7) one has

$$
J_n^{(2)}((\mathbb{X}, \mathbb{Y}), \rho) = \frac{V_1}{1+\rho} + \frac{V_2}{1-\rho} \quad \text{and} \quad J_n^{(3)}((\mathbb{X}, \mathbb{Y}), \rho) = \frac{1}{V_1(1+\rho)} + \frac{1}{V_2(1-\rho)}.
$$

In order to choose the better data generating equation among the proposed ones, we will decide the criteria in the next section.

4.3 Probability Matching Data Generating Equation:

Recall for any two sequences $\{a_n\}_{n\geq 1}, \{b_n\}_{n\geq 1}$ we denote $a_n = O(b_n)$ if $\limsup_{n\to\infty} \frac{a_n}{b_n} < \infty$, and $a_n = o(b_n)$ if $\lim_{n\to\infty} \frac{a_n}{b_n} \to 0$. We define $\mathcal{G}_s$ as the **Probability Matching Data Generating Equation** of order $s \in \mathbb{N}$ if

$$
P_{\theta_0}\left[\theta_0 < \theta^{1-\alpha}(\mathbb{X}, \mathcal{G}_s)\right] = P^{\mathcal{G}_s}(\theta < \theta^{1-\alpha}(\mathbb{X}, \mathcal{G}_s) \mid \mathbb{X}) + o(n^{-\frac{s}{2}}) \tag{4.3.1}
$$

where $\theta^{1-\alpha}(\mathbb{X}, \mathcal{G}_s)$ is the upper $(1-\alpha)$-th quantile of the **Generalized Fiducial Distribution** Distribution $P^{\mathcal{G}_s}(\cdot \mid \mathbb{X})$. In other words it characterizes the corresponding data generating equation, so that the frequentist coverage of the $(1-\alpha)$-th Fiducial quantile matches $(1-\alpha)$ upto rate $o(n^{-\frac{s}{2}})$.

We plan to guide our choice of **DGE** based on frequentist coverage. ***Why?*** It gives the information how exactly one sided quantile estimator of **GFD** behaves *asymptotically in frequentist sense*. For illustration, suppose we had generated m batches of n (fixed) i.i.d samples from the original distribution $F(.|\theta_0)$. With each batch of n samples one computes the Fiducial $(1-\alpha)$-th quantile. Then we find how many times out of m, θ_0 is less than the values of those quantiles. Asymptotically as $m \to \infty$, that proportion will converge to $P_{\theta_0}\left[\theta_0 < \theta^{1-\alpha}(\mathbb{X})\right]$ which should be "close" to $(1-\alpha)$. In fact as $n \to \infty$ from [51] following holds

$$\lim_{n\to\infty} P_{\theta_0}\left[\theta_0 < \theta^{1-\alpha}(\mathbb{X})\right] = 1-\alpha$$

which is (4.3.1) for $s = 0$. A stronger result would be finding an asymptotic (as $n \to \infty$) expansion of $P_{\theta_0}\left[\theta_0 < \theta^{1-\alpha}(\mathbb{X})\right]$ and finding conditions under which the coefficients of first few order terms will be 0 (getting (4.3.1) for larger $s > 0$). That is the purpose of this probability matching idea.

Secondly in *Bayesian paradigm*, probability matching prior based approach had been relevant [73] in context of judging superiority of non-informative priors. In fact one criteria for calling a prior noninformative is if $(1-\alpha)$th posterior regions have frequentist coverage equal (approximately or very close) to $(1-\alpha)$. The interpretation is *how least the specified prior influences (giving information about the parameter space) in quantile (or other distribution based estimator) from its "ideal" coverage or how least informative it is for the inference of the parameter*. An ideal non-informative prior should match all order terms at the true parameter value but constructing it non-parametrically is an open challenge. In **GFI**, the challenge is translated into finding the data generating equation for which the fiducial quantile has the exact ideal coverage (i.e least influence on the parameter space). Similar to Bayesian paradigm achieving exactness is immensely hard (though it is possible to get that for univariate cases under certain monotonicity conditions imposed, just seen in section 2) but first and second order terms are relatively easier to analyze which we will describe now.

4.3.1 Regularity Assumptions:

Here we consider one dimensional parameter space Θ containing true θ_0. Define $l(\theta \mid X_i) = \log f(X_i, \theta)$ as the log-likelihood of θ given one sample point X_i. Denote $l^{(m)}(\theta \mid X_1)$ as the m-th derivative of the log likelihood function $l(\theta \mid X_1)$ with respect to θ. Define $L_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} l(\theta|X_i)$ as the likelihood of θ given $\mathbb{X}$ (which is scaled by $\frac{1}{n}$) and $c := -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 l(\theta|X_i)}{\partial\theta^2}$. We denote the quantity $\sqrt{nc}(\theta - \hat{\theta}_n)$ by y. Note that $c = L_n^{(2)}(\hat{\theta}_n)$, and define $a = L_n^{(3)}(\hat{\theta}), \quad a_4 = L_n^{(4)}(\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimate of θ_0 (or any solution of $L_n'(\theta) = 0$). Denote $\sqrt{nc}(\theta - \hat{\theta}_n)$ by y whose Fiducial expansion will be needed for asymptotic analysis. From now on $\phi(x)$ will denote the density of the Gaussian distribution function (i.e $\frac{1}{\sqrt{2\pi e}}e^{-\frac{x^2}{2}}$). For sake of generality from this section onwards by $J_n(\mathbb{X}, \theta)$ we denote *any Jacobian* that appears in the Generalized Fiducial distribution driven by the corresponding data generating equation $\mathcal{G}$. For $m \geq 1$, we denote $\frac{\partial J_n(X,\theta)}{\partial\theta}\big|_{\theta=\hat{\theta}_n}, \frac{\partial^2 J_n(X,\theta)}{\partial\theta^2}\big|_{\theta=\hat{\theta}_n}, \frac{\partial^m J_n(X,\theta)}{\partial\theta^m}\big|_{\theta=\hat{\theta}_n}$ by $J_n'(X, \hat{\theta}_n)$, $J_n''(X, \hat{\theta}_n)$, and $J_n^{(m)}(X, \hat{\theta}_n)$ respectively. We know by virtue of SLLN pointwise for each θ, $\quad m$-th derivative (w.r.t θ) of the simple Jacobian $J_n^{(m)}(X, \theta)$ in (4.1.7) scaled by $\frac{1}{n}$ converges to $J^{(m)}(\theta_0, \theta) := \frac{\partial^m E_{\theta_0}[J(\mathbb{X},\theta)]}{\partial\theta^m}$ almost surely as $n \to \infty$.

Assumption 16. *These Assumptions are essential for proving asymptotic normality of $\hat{\theta}_n$ and here we need a stronger version to ensure a valid higher order likelihood expansion.*

1. *The distribution $F(\cdot|\theta)$ are distinct for $\theta \in \Theta$.*

2. *The set $\{x : f(x|\theta) > 0\}$ is independent of the choice of θ.*

3. *The data $\mathbb{X} = \{X_1, \ldots, X_n\}$ are iid with probability density $f(x|\theta)$.*

4. *There exists $m \geq 1$, such that in a neighborhood $B(\theta_0, \delta)$ of the true value θ_0, all possible $(m+3)$ ordered partial derivatives $\frac{\partial^{m+3} f(x|\theta)}{\partial \theta^{m+3}}$ exist. For all $i = 1, \ldots, m+2$; the quantities $E_{\theta_0} l^{(i)}(\theta_0|X_i)$ are all finite.*

5. *There exists a function $M(x)$ such that*

$$\sup_{\theta \in B(\theta_0,\delta)} \Big| \frac{\partial^{(m+3)}}{\partial \theta^{(m+3)}} \log f(x \mid \theta) \Big| \leq M(x) \quad \textit{and} \quad E_{\theta_0} M(X) < \infty.$$

6. *The information $I(\theta)$ is positive for all $\theta \in B(\theta_0, \delta)$*

Under the Assumption 16 with $m = 2$ one has the following expansion (Consequence of Taylor's theorem). We have for some $\theta' \in (\hat{\theta}_n, \hat{\theta}_n + \frac{y}{\sqrt{nc}})$,

$$\begin{aligned} n[L_n(\hat{\theta}_n + \frac{y}{\sqrt{nc}}) - L_n(\hat{\theta}_n)] &= -\frac{y^2}{2} + \frac{1}{6} \frac{y^3 L_n^{(3)}(\hat{\theta}_n)}{\sqrt{n} c^{\frac{3}{2}}} + \frac{1}{24} \frac{y^4}{nc^2} L_n^{(4)}(\hat{\theta}_n) + \frac{1}{120} \frac{y^5}{n^{3/2} c^2} L_n^{(5)}(\theta') \\ &:= -\frac{y^2}{2} + R_n(\hat{\theta}_n) + \frac{1}{120} \frac{y^5}{n^{3/2} c^2} L_n^{(5)}(\theta'), \end{aligned} \tag{4.3.2}$$

where $R_n(\theta) := \frac{1}{6} \frac{y^3 L_n^{(3)}(\theta)}{\sqrt{n} c^{\frac{3}{2}}} + \frac{1}{24} \frac{y^4}{nc^2} L_n^{(4)}(\theta)$. Following assumptions are needed to control the tail of the numerator of the fiducial distribution.

Assumption 17. *Assume*

1. *for any $\delta > 0$, there exists an $\epsilon > 0$ such that*

$$P_{\theta_0}\left\{\sup_{\theta \in B(\theta_0,\delta)^c}[L_n(\theta) - L_n(\theta_0)] \leq -\epsilon\right\} \to 1 \quad \textit{as} \quad n \to \infty.$$

2. *Let $\mathbf{x} = (x_1, \ldots, x_n)$. There exists $s \in \mathbb{N}$, such that $J_n(\mathbf{x}, \theta) = \sum_{i=1}^n J_i(\mathbf{x}, \theta)$, where $J_i(\mathbf{x}, \theta)$ satisfies*

$$\sup_{i=1,\ldots,n} n^{-s}\int_{\mathbb{R}} J_i(\mathbf{x}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) f(x_i, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) dy \ < \ \infty \qquad \textit{a.s} \quad P_{\theta_0}.$$

3. *The density f satisfies the following property: There exists a constant $c \in [0, 1)$*

$$\inf_{\theta \in B(\theta_0,\delta)^c} \frac{\min_{i=1,\ldots,n} \log f(X_i, \theta)}{n[L_n(\theta) - L_n(\theta_0)]} \xrightarrow{P_{\theta_0}} c. \tag{4.3.3}$$

Remark 13. *1. Assumption 17(b) can be verified with $s = 0$ for the simple Jacobian structure of the form in (4.1.12) by taking $J_i(\mathbf{x}, \theta) := \left|w(x_i)\frac{\frac{\partial F_i(x_i,\theta)}{\partial \theta}}{f(x_i,\theta)}\right|$. Since*

$$\int_{\mathbb{R}} J_i(\mathbb{X}, \theta) f(X_i, \theta) d\theta = \left|w(X_i)\right| [F_i(X_i, \infty) - F_i(X_i, -\infty)] < \infty \quad \textit{a.s} \quad P_{\theta_0}.$$

Note that any polynomial exponent of n can replace the condition "n^s for some $s > 0$."

2. If $X_1, X_2, \ldots, X_n$ are iid realizations from density $f(\cdot|\theta_0)$ then both the numerator and denominator of the left hand side of (4.3.3) converge to $-\infty$ with rate $-C_1 \log n$ and $-C_2 n$ respectively. So in that case Assumption 17(c) is strongly implied by $c = 0$ if $\frac{C_1}{C_2}$ is uniformly bounded for $n \geq 1$.

Next is a very crucial Assumption about the Jacobian $J_n(\mathbb{X}, \theta)$ which gives further conditions on the data generating equation.

Assumption 18. *There exists a function $J(\cdot,\cdot) : \Theta \times \Theta \to \mathbb{R}$ with its i-th derivative with respect to second argument $\frac{\partial^i J(\theta_1,\theta)}{\partial^i \theta}$, denoted by $J^{(i)}(\theta_1, \theta)$ (where $J^{(0)}(\theta_1, \theta) := J(\theta_1, \theta)$); such that following conditions hold.*

1. *There exists $m \geq 1$, such that for each $i = 0, \ldots, m+1$ the Jacobian $J_n^{(i)}(\mathbb{X}, \theta)$, satisfies*

$$\sup_{\theta \in B(\theta_0, \delta)} \left| J_n^{(i)}(\mathbb{X}, \theta) - J^{(i)}(\theta_0, \theta) \right| \to 0 \qquad a.s. \quad P_{\theta_0}. \tag{4.3.4}$$

   Namely a uniform convergence result holds over a neighborhood of true parameter value θ_0 for each $i = 0, \ldots, m+1$ uniformly as $n \to \infty$.

2. *The function $J(\cdot, \theta)$ doesn't vanish in $\theta \in B(\theta_0, \delta)$ for any $\delta > 0$.*

3. *For $i = 0, \ldots, m+1$ the quantities*

$$\sqrt{n} \left[J_n^{(i)}(\mathbb{X}, \hat{\theta}_n) - J^{(i)}(\theta_0, \hat{\theta}_n) \right] = O_{P_{\theta_0}}(1). \tag{4.3.5}$$

Remark 14. *Following comments are on Assumption 18:*

1. *In (4.3.4) a difference from Bayesian paradigm is the extra (Assumption 18 for $i = (m+1)$th order) smoothness condition for data dependent $J_n(\mathbb{X}, \theta)$ which is needed to apply the uniform law of large number in a neighborhood of θ_0.*

2. *Assumption 18(a) holds for the simple Jacobian in (4.1.12) if following are satisfied for each* $i = 0, \ldots, m+1:$

   (a) *For each* x, $J^{(i)}(x, \cdot)$ *is continuous in* $\theta \in [\theta_0 - \delta, \theta_0 + \delta]$.

   (b) *For each* $\theta \in [\theta_0 - \delta, \theta_0 + \delta]$, $J^{(i)}(\cdot, \theta)$ *is a strictly positive measurable function of* x.

   (c) *There exists a* $\delta > 0$ *such that,*

$$E_{\theta_0}\left(\sup_{\theta \in B(\theta_0, \delta)} \left| J^{(i)}(X_1, \theta) \right| \right) < \infty.$$

This follows from Wald's theorem.

In contrast to Probability matching prior framework [23] main difference comes from the fact that here the Jacobian $J_n(\mathbb{X}, \theta)$ has a data dependent $U-$statistics type (functional of empirical distribution) structure than a simple prior function of the parameter. In order to analyze the exapnsion of $J_n(\mathbb{X}, \theta)$, the basic ingredient will be Taylor's theorem:

We have for some $\theta' \in (\hat{\theta}_n, \hat{\theta}_n + \frac{y}{\sqrt{nc}})$,

$$\begin{aligned}
& J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) = J_n(\mathbb{X}, \hat{\theta}_n) + J_n'(\mathbb{X}, \hat{\theta}_n)\frac{y}{\sqrt{nc}} + J_n''(\mathbb{X}, \hat{\theta}_n)\frac{y^2}{2nc} + J_n'''(\mathbb{X}, \theta')\frac{y^3}{6(nc)^{3/2}} \\
= \; & J_n(\mathbb{X}, \hat{\theta}_n)\left[1 + \frac{J_n'(\mathbb{X}, \hat{\theta}_n)}{J_n(\mathbb{X}, \hat{\theta}_n)}\frac{y}{\sqrt{nc}} + \frac{J_n''(\mathbb{X}, \hat{\theta}_n)}{J_n(\mathbb{X}, \hat{\theta}_n)}\frac{y^2}{2nc} + \frac{J_n'''(\mathbb{X}, \theta')}{J_n(\mathbb{X}, \hat{\theta}_n)}\frac{y^3}{6(nc)^{3/2}}\right] \\
= \; & J_n(\mathbb{X}, \hat{\theta}_n)\left[1 + \frac{J'(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)}\frac{y}{\sqrt{nc}} + \frac{1}{n}\left(\mathbb{W}_n^{(1)}(\mathbb{X})\frac{y}{\sqrt{c}} + \frac{J''(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)}\frac{y^2}{2c}\right)\right] \\
& + \frac{J_n(\mathbb{X}, \hat{\theta}_n)}{n^{3/2}}\left(\mathbb{W}_n^{(2)}(\mathbb{X})\frac{y^2}{2c} + \frac{J_n'''(\mathbb{X}, \theta')}{J_n(\mathbb{X}, \hat{\theta}_n)}\frac{y^3}{6(c)^{3/2}}\right) \qquad (4.3.6)
\end{aligned}$$

where $\mathbb{W}_n^{(m)}(\mathbb{X}) := \sqrt{n}\left(\frac{J_n^{(m)}(\mathbb{X},\hat{\theta}_n)}{J_n(\mathbb{X},\hat{\theta}_n)} - \frac{J^{(m)}(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}\right)$ for $m \geq 1$.

In higher order expansion of $P_{\theta_0}[\theta_0 \leq \theta^{(1-\alpha)}(\mathbb{X},\mathcal{G})]$ the main difference of **GFD** from " *Probability matching prior"* framework comes from the presence of the terms $\mathbb{W}_n^{(m)}(\mathbb{X})$ for $m \geq 1$. We gave here an explicit proof of the expansion (4.7.10) which comes by making a product of (4.3.6) and (4.3.2) as it was outlined in Bayesian context [42] Chapter 5, from equations (5.4b),(5.4c) to (5.4f). Following assumptions are necessary for existence of the second order term.

Assumption 19. *There exists $m \geq 0$, for which following hold:*

1. ***Integrability Condition:*** *For $i = 0, \ldots, m$ and any $\delta > 0$, one has for all $\theta \in \Theta$*

$$E_\theta\left[n\left[J_n^{(i)}(\mathbb{X},\hat{\theta}_n) - J^{(i)}(\theta,\hat{\theta}_n)\right]^2\right] = O(1)$$

   where the finite constant may depend on θ.

2. *For $i = 0, \ldots, m$ There exist continuous functions $a_i(.)$ such that*

$$a_i(\theta) := \lim_{n\to\infty} E_\theta\sqrt{n}\left[J_n^{(i)}(\mathbb{X},\hat{\theta}_n) - J^{(i)}(\theta,\hat{\theta}_n)\right]. \tag{4.3.7}$$

3. *For $i = 0, \ldots, m$ the functions $J^{(i)}(\theta_0,\cdot)$ are locally Lipschitz.*

Now we will state the main result for first and second order terms in expansion of $P_{\theta_0}\left[\theta_0 \leq \theta^{1-\alpha}(\mathcal{G},\mathbb{X})\right]$. Define the terms $g_1 := J'(\theta_0,\theta_0), \quad g_2 := J(\theta_0,\theta_0)$ where $J'(\theta_0,\theta_0) = \frac{\partial J(\theta_0,\theta)}{\partial\theta}\Big|_{\theta=\theta_0}$. Let z_α is $(1-\alpha)$-th quantile of Normal distribution.

Theorem 3. *Suppose Assumptions 16,17,18,19 hold with $m = 2$. Following expansion holds for all $\theta_0 \in \Theta$ and for some constants (given α) $c_1 := \phi(z_\alpha), c_2 := z_\alpha\phi(z_\alpha)$ (depends only on θ_0)*

$$P_{\theta_0}\left[\theta_0 \leq \theta^{1-\alpha}(\mathcal{G}, \mathbb{X}, n)\right] = \left(1-\alpha\right) + \frac{c_1\Delta_1(\mathcal{G})}{\sqrt{n}} + \frac{c_2\Delta_2(\mathcal{G})}{n} + o\left(\frac{1}{n}\right), \qquad \text{where}$$

$$\Delta_1(\mathcal{G}) = \left[I_{\theta_0}^{-\frac{1}{2}}\frac{\frac{\partial}{\partial\theta}J(\theta_0,\theta)}{J(\theta_0,\theta_0)} + \frac{\partial I_\theta^{-\frac{1}{2}}}{\partial\theta}\right]\Bigg|_{\theta=\theta_0}, \tag{4.3.8}$$

$$\begin{aligned}\Delta_2(\mathcal{G}) = {} & \frac{I_{\theta_0}^{-\frac{1}{2}}}{z_\alpha g_2}\left[a_1(\theta_0) - \frac{a_0(\theta_0)g_1}{g_2}\right] + \left[\frac{1}{6}\frac{\partial}{\partial\theta}\left\{I_\theta^{-2}J(\theta_0,\theta)E_\theta\left[l^{(3)}(\theta \mid X)\right]\right\}\right. \\ & \left. -\frac{1}{2}\frac{\partial^2}{\partial\theta^2}J(\theta_0,\theta)I_\theta^{-1}\right]\Bigg|_{\theta=\theta_0}.\end{aligned} \tag{4.3.9}$$

Remark 15. *Following are some remarks on Assumption 19 and extention to higher order terms:*

1. *Assumption 19(a) is stronger than Assumption 18(c). We mentioned the latter to emphasis on the fact that it is sufficient for only first order term.*

2. *Note that in all situations where $\mathbb{X}$ is a collection of n random samples, usually we have $\left[\frac{a_1(\theta_0)}{z_\alpha} - \frac{a_2(\theta_0)g_1}{z_\alpha g_2}\right] = 0$ because of the symmetricity reason. Note that $\Delta_1(\mathcal{G})$ and the second term of $\Delta_2(\mathcal{G})$ are both first and second order terms for the asymptotic expansion of $P_{\theta_0}\left[\theta_0 \leq \theta^{1-\alpha}(\pi, \mathbb{X})\right]$ where $\theta^{1-\alpha}(\pi, \mathbb{X})$ is the $(1-\alpha)$th Posterior quantile*

based on the prior :

$$\pi(\cdot) \propto J(\theta_0, \cdot) \qquad \textit{where } \theta_0 \textit{ is the true parameter value.}$$

3. ***Higher order terms:*** *In general we have $J(\theta_0, \theta)$ to be $\lim_{n\to\infty} E_{\theta_0} J_n(\mathbb{X}, \theta)$, implying $a_i(\theta_0)$ will be 0 for $i = 0, 1$. But in Theorem 3 we kept it general since data generating structural equation is not-unique. So conditions for the first two order terms really will not differ from the conditions in* ***probability matching priors****. We did not pursue here explicitly, but under Assumptions 16-19 suppose further following two assumptions hold with $m = 3$:*

   (a) For $i = 0, \ldots, m$ and any $\delta > 0$, one has for all $\theta \in \Theta$

   $$\lim_{n\to\infty} n^{\frac{3}{2}} E_\theta \left[\left[J_n^{(i)}(\mathbb{X}, \hat{\theta}_n) - J^{(i)}(\theta, \hat{\theta}_n) \right]^3 \right] < \infty$$

   where the finite constants may depend on θ.

   (b) Define the following quantities:

   $$\begin{aligned} a_i^{(1)}(\theta_0) &:= \lim_{n\to\infty} n E_{\theta_0} \left[J_n^{(i)}(\mathbb{X}, \hat{\theta}_n) - J^{(i)}(\theta_0, \hat{\theta}_n) \right]^2, \\ a_{0,1}^{()}(\theta_0) &:= \lim_{n\to\infty} n E_{\theta_0} \left[J_n(\mathbb{X}, \hat{\theta}_n) - J(\theta_0, \hat{\theta}_n) \right] \left[J_n^{(1)}(\mathbb{X}, \hat{\theta}_n) - J^{(1)}(\theta_0, \hat{\theta}_n) \right]. \end{aligned}$$

Then analogue to the Theorem 3 a ***third order representation*** *holds:*

$$P_{\theta_0}\Big[\theta_0 \leq \theta^{1-\alpha}(\mathcal{G},\mathbb{X},n)\Big] - \big(1-\alpha\big) = \frac{c_1\Delta_1(\mathcal{G})}{\sqrt{n}} + \frac{c_2\Delta_2(\mathcal{G})}{n} + \frac{c_3\Delta_3(\mathcal{G})}{n^{3/2}},$$

$$+o\bigg(\frac{1}{n^{3/2}}\bigg) \quad \textit{where} \quad \Delta_3(\mathcal{G}) = \bigg[\frac{a_i^{(1)}(\theta_0)g_1}{g_3^2} - \frac{a_{0,1}^{()}(\theta_0)}{g_2}\bigg] + I_{\theta_0}^{-\frac{1}{2}}\bigg[\frac{a_1(\theta_0)}{z} - \frac{a_2(\theta_0)g_1}{zg_2}\bigg]$$
$$+\bigg\{\textit{Third order Probability matching prior term with prior} J(\theta_0,\theta) \textit{ at } \theta = \theta_0\bigg\}.$$

This extra additive quantity $\left(\frac{a_i^{(1)}(\theta_0)g_1}{g_3^2} - \frac{a_{0,1}^{()}(\theta_0)}{g_2}\right)$ *in the display above will come due to the following Taylor's expansion of* $\frac{T_1}{T_2}$ *around* $\frac{g_1}{g_2}$ *where* $T_1 := J_n'(\mathbb{X},\hat{\theta}_n), T_2 := J_n(\mathbb{X},\hat{\theta}_n)$ *and their corresponding limits* $g_1 := J'(\theta_0,\theta_0), g_2 := J(\theta_0,\theta_0)$

$$\frac{T_1}{T_2} = \frac{g_1}{g_2} + (T_1-g_1)\frac{1}{g_2} - (T_2-g_2)\frac{g_1}{g_2^2} + \bigg((T_2-g_2)^2\frac{g_1}{g_2^3} - (T_1-g_1)\times \tag{4.3.10}$$
$$(T_2-g_2)\frac{1}{g_2}\bigg) + O\Bigg(\bigg((T_1-g_1)\frac{\partial}{\partial x_1} + (T_2-g_2)\frac{\partial}{\partial x_2}\bigg)^3\bigg(\frac{x_1}{x_2}\bigg)\bigg|_{x_1\in(T_1,g_1),x_2\in(T_2,g_2)}\Bigg)$$

It implies that after taking expectation

$$E_{\theta_0}\left[W_n^{(1)}(\mathbb{X})\right] = \bigg(\frac{a_1(\theta_0)}{g_2} - \frac{a_0(\theta_0)g_1}{g_2^2}\bigg) + \frac{1}{\sqrt{n}}\bigg[\frac{a_i^{(1)}(\theta_0)g_1}{g_3^2} - \frac{a_{0,1}^{()}(\theta_0)}{g_2}\bigg]$$
$$+O\bigg(\frac{1}{n}\bigg). \tag{4.3.11}$$

keeping an extra order term. Note that for simple data generating equation $J(\theta_0,\theta) = E_{\theta_0}J_n(\mathbb{X},\theta)$*, along with the empirical structure* $J_n(\mathbb{X},\theta) = \frac{1}{n}\sum_{i=1}^n J(X_i,\theta)$ *then one* *gets*

$$\left[\frac{a_i^{(1)}(\theta_0)g_1}{g_3^2} - \frac{a_{0,1}^{()}(\theta_0)}{g_2}\right] = \left[\frac{Var_{\theta_0}(J(X_1,\theta))g_1}{g_3^2} - \frac{Cov_{\theta_0}(J(X_1,\theta), J'(X_1,\theta))}{g_2}\right]\Bigg|_{\theta=\theta_0}$$

which appears as an extra in the third order term. The difference of Fiducial cases will be different from Bayesian paradigm likewise in the further order of terms, starting from 3rd *order due to the presence of* $\{W_n^{(m)}(\mathbb{X}), m \geq 1\}$ *and their respective higher order expansions.*

Corollary 4.3.1. *Under Assumptions 16,17,18 with* $m = 1$, $\mathcal{G}_1$ *will be the* ***first order Probability Matching Generating Equation*** *if* $\quad \Delta_1(\mathcal{G}_1) = 0.$

Under Assumptions 16,17,18,19 with $m = 2$, $\mathcal{G}_2$ *will be the* ***Second order Probability Matching Generating Equation*** *if* $\quad \Delta_1(\mathcal{G}_2) = 0,$ *and* $\Delta_2(\mathcal{G}_2) = 0.$

4.4 Recipe For Perfect DGE For $s = 1, 2$ in (4.3.1)

Main contention of this section is to provide proper guidelines so that for any one parameter family so that the space of transformations can be identified for which the desired matching properties are satisfied. By “perfect” we mean those data generating equations for which one can obtain both $s = 1, 2$ th matching Fiducial distribution. In this section we discuss one such technique that is motivated by a few examples of exponential family.

1. *Find minimal sufficient statistics* $(S_1, S_2, \ldots, S_m)$ *of* θ*. From the point of view of computing* ***GFD*** *the choice* $(S_1, S_2, \ldots, S_m)$ *would be better if they are independent of each other (which are possible for exponential families).*

2. *Find any function $g : \mathbb{R}^m \to \mathbb{R}^m$, that satisfies the conditions of Corollary 4.3.1.*

Now we will apply this recipe on some examples where the dimension of minimal sufficient statistics is strictly greater than dimension of the parameter. We will work on two examples of $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ and $N(\mu, \mu^q)$ for $\mu > 0$. Suppose

$$\mathcal{A} := \{\text{Space of all } C^1 \text{ and one-one transformations from } S \text{ to } S\}.$$

One possible representation of transformation would be

$$A(S_1, S_2) := (A_1(S_1), A_2(S_2)) \tag{4.4.1}$$

where both $A_1(\cdot), A_2(\cdot) \in \mathcal{A}$. Now for computational advantage or because of simplicity instead of (4.4.1) one may use the following transformation

$$A(S_1, S_2) = (A_1(S_1)g_1(S_1), A_2(S_2)g_2(S_2)) \tag{4.4.2}$$

where $g_1(\cdot), g_2(\cdot)$ be two functions on S such that

$$P_{\theta_0}\left[g_1'(S_1) \neq 0, g_2'(S_2) \neq 0\right] = O(e^{-an}) \quad \text{for any constant} \quad a > 0$$

and both of $A_1(\cdot)g_1(\cdot), A_2(\cdot)g_2(\cdot) \in \mathcal{A}$. From Corollary 4.3.1 we define the set of transformations $A(\cdot, \cdot)$ yielding first and second order probability matching data generating equations respectively as:

$$\boldsymbol{\mathcal{A}_{\mathcal{G}}^{(1)} = \{A \in \mathcal{A}_{\mathcal{G}} : \Delta_1(\mathcal{G}) = 0\}, \quad \mathcal{A}_{\mathcal{G}}^{(2)} = \{A \in \mathcal{A}_{\mathcal{G}} : \Delta_1(\mathcal{G}) = 0, \Delta_2(\mathcal{G}) = 0\}.}$$

We will describe the class $\mathcal{A}_{\mathcal{G}}^{(1)}$ in examples of $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ and $N(\mu, \mu^q)$ for $\mu > 0$.

1. Since for $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ the generating equation of minimal-sufficient statistics is

$$(S_1, S_2) = \left(\frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i)^2, \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2\right) = \big((1+\rho)U_1, (1-\rho)U_2\big).$$

For general $A := (A_1, A_2) \in \mathcal{A}_{\mathcal{G}}$, the Jacobian for $(A_1(S_1), A_2(S_2)) = \big(A_1((1+\rho)U_1), A_2((1-\rho)U_2)\big)$ will be (after taking mod) $J_n^A(\mathbb{X}, \rho) = A_1'(S_1)\frac{S_1}{1+\rho} + A_2'(S_2)\frac{S_2}{1-\rho}$ converges to $J^A(\rho_0, \rho) := A_1'(1+\rho_0)\frac{1+\rho_0}{1+\rho} + A_2'(1-\rho_0)\frac{1-\rho_0}{1-\rho}$ as sample size goes to ∞. as sample size (n) goes to ∞. Now before using Theorem 3 to find $\boldsymbol{\mathcal{A}_{\mathcal{G}}^{(1)}}$ we should verify Assumption 18. Note that

$$\begin{aligned}
&J_n^A(\mathbb{X}, \hat{\rho}_n) - J^A(\rho_0, \hat{\rho}_n) = [A_1'(S_1) - A_1'(1+\rho_0)]\frac{S_1}{1+\hat{\rho}_n} \\
&+A_1'(1+\rho_0)\left[\frac{S_1 - (1+\rho_0)}{1+\hat{\rho}_n}\right] + [A_2'(S_2) - A_2'(1-\rho_0)]\frac{S_2}{1-\hat{\rho}_n} \\
&+\frac{A_2'(1-\rho_0)}{1-\hat{\rho}_n}[S_2 - (1-\rho_0)]
\end{aligned}$$

and

$$J_n^{A(1)}(\mathbb{X},\rho) - J^{A(1)}(\rho_0,\rho) = [A_1'(S_1) - A_1'(1+\rho_0)]\frac{S_1}{1+\hat{\rho}_n} + A_1'(1+\rho_0) \times \left[\frac{S_1-(1+\rho_0)}{1+\hat{\rho}_n}\right] + [A_2'(S_2) - A_2'(1-\rho_0)]\frac{S_2}{1-\hat{\rho}_n} + \frac{A_2'(1-\rho_0)}{1-\hat{\rho}_n}[S_2-(1-\rho_0)].$$

Since S_1, S_2 both converge to $(1+\rho_0),(1-\rho_0)$ respectively; using smoothness of A_1, A_2 by applying Delta method and Slutsky's theorem one can show that $\sqrt{n}[J_n^A(\mathbb{X},\hat{\rho}_n) - J^A(\rho_0,\hat{\rho}_n)]$ $\big($similarly for $\sqrt{n}(J_n^{A(1)}(\mathbb{X},\rho) - J^{A(1)}(\rho_0,\rho))\big)$ is $O_{P_{\rho_0}}(1)$.

The first order class follows by equating $\Delta_1(\mathcal{G}) = 0$, which is equivalent of saying

$$\left.\frac{\frac{\partial}{\partial\rho}J^A(\rho_0,\rho)}{J^A(\rho_0,\rho_0)}\right|_{\rho=\rho_0} = \frac{1}{2}\frac{I'_{\rho_0}}{I_{\rho_0}}.$$

Note that $\left.\frac{\frac{\partial}{\partial\rho}J^A(\rho_0,\rho)}{J^A(\rho_0,\rho_0)}\right|_{\rho_0} = \frac{(1+\rho_0)-(1-\rho_0)\left[\frac{A_1'(1+\rho_0)}{A_2'(1-\rho_0)}\right]}{(1-\rho_0^2)\left(1+\left[\frac{A_1'(1+\rho_0)}{A_2'(1-\rho_0)}\right]\right)}$, equating that with $\frac{1}{2}\frac{I'_{\rho_0}}{I_{\rho_0}} = \frac{3\rho_0+\rho_0^3}{(1-\rho_0^2)(1+\rho_0^2)}$, will give the following characterization of the first order matching class:

$$\boldsymbol{\mathcal{A}_{\mathcal{G}}^{(1)}} = \left\{A := \big(A_1(\cdot), A_2(\cdot)\big) \in \mathcal{A}_{\mathcal{G}} : \ \boldsymbol{A_1'(1+\rho) = A_2'(1-\rho).\frac{(1-\rho)^2}{(1+\rho)^2}}, \text{ for } |\boldsymbol{\rho}| < 1\right\}. \quad (4.4.3)$$

The second proposal in (4.2.9) $A = (A_1, A_2)$ such that $A_1(x) = A_2(x) = \frac{1}{x}$, belongs to the class $\boldsymbol{\mathcal{A}_{\mathcal{G}}^{(1)}}$ hence it is first order matching.

Now to judge Assumption 19 in order to ensure second order term we need to find the asymptotic limit of $E_{\rho_0}\sqrt{n}[J_n^{A(i)}(\mathbb{X},\hat{\rho}_n) - J^{A(i)}(\rho_0,\hat{\rho}_n)]$ and we need to check finiteness of $E_{\rho_0}\Big[n[J_n^{A(i)}(\mathbb{X},\hat{\rho}_n) - J^{A(i)}(\rho_0,\hat{\rho}_n)]^2\Big]$. Note that for $i = 0, 1$ each of the four terms of $\sqrt{n}[J_n^{A(i)}(\mathbb{X},\hat{\rho}_n) - J^{A(i)}(\rho_0,\hat{\rho}_n)]$ by Slutsky's theorem converges to normal with mean zero. Since S_1 is chi-square so using its exponential concentration property one can prove uniform integrability of each of those terms. So $E_{\rho_0}\sqrt{n}[J_n^{A(i)}(\mathbb{X},\hat{\rho}_n) - J^{A(i)}(\rho_0,\hat{\rho}_n)]$ asymptotically will converge to the mean of its weak limit which is 0. In order to characterize the second order class $\boldsymbol{\mathcal{A}_{\mathcal{G}}^{(2)}}$ we need to compute the secnd order term from (4.3.9). Note that the first term of $\Delta_2(\mathcal{G})$ is 0 since both of $a_1(\rho_0) = 0, \quad a_2(\rho_0) = 0$. Now denoting $M_\rho = E_\rho[l^{(3)}(\theta|X)]$ and using using $\frac{\frac{\partial}{\partial\theta}J(\theta_0,\theta)}{J(\theta_0,\theta_0)}\Big|_{\theta_0} = \frac{1}{2}\frac{I'_{\theta_0}}{I_{\theta_0}}$ followed by $\Delta_1(\mathcal{G}) = 0$ we have

$$\Delta_2(\mathcal{G}) = \left[\frac{1}{6}\frac{\partial}{\partial\theta}\left\{I_\theta^{-2}J(\theta_0,\theta)M_\theta\right\} - \frac{1}{2}\frac{\partial^2}{\partial\theta^2}J(\theta_0,\theta)I_\theta^{-1}\right]\Bigg|_{\theta=\theta_0} \tag{4.4.4}$$
$$= \frac{J(\theta_0,\theta_0)}{2}\left[\frac{1}{3}\frac{d}{d\theta}\left[I_\theta^{-2}M_\theta\right]\Bigg|_{\theta=\theta_0} + \frac{1}{6}I_{\theta_0}^{-3}M_{\theta_0}I'_{\theta_0} - \frac{1}{J(\theta_0,\theta_0)}\frac{\partial^2}{\partial\theta^2}\left[J(\theta_0,\theta)I_\theta^{-1}\right]\Bigg|_{\theta_0}\right].$$

After simplifying inside the bracket of RHS of (4.4.4) and using $M_\rho = -2I'_\rho$ we will get following

$$- \frac{I_{\theta_0}^{-1}}{J(\theta_0,\theta_0)}\frac{\partial^2}{\partial\theta^2}\left[J(\theta_0,\theta)\right]\Bigg|_{\theta_0} - \frac{1}{2}(I'_{\theta_0})^2 I_{\theta_0}^{-3} + \frac{2}{3}I''_{\theta_0}I_{\theta_0}^{-2} + \frac{1}{3}\frac{d}{d\theta}\left[I_\theta^{-2}M_\theta\right]\Bigg|_{\theta=\theta_0}$$
$$+\frac{1}{6}I_{\theta_0}^{-3}M_{\theta_0}I'_{\theta_0} \tag{4.4.5}$$

Note that $\frac{1}{J^A(\rho_0,\rho_0)}\frac{\partial^2}{\partial\rho^2}\left[J^A(\rho_0,\rho)\right]\Bigg|_{\rho=\rho_0} = \frac{2\left[\frac{A'_1(1+\rho_0)}{(1+\rho_0)^2} + \frac{A'_2(1-\rho_0)}{(1-\rho_0)^2}\right]}{A'_1(1+\rho_0)+A'_2(1-\rho_0)} = \frac{\frac{2}{(1+\rho_0)^2}\left[\frac{A'_1(1+\rho_0)}{A'_2(1-\rho_0)} + \frac{(1+\rho_0)^2}{(1-\rho_0)^2}\right]}{1+\left[\frac{A'_1(1+\rho_0)}{A'_2(1-\rho_0)}\right]}$

which is again a function of $\left[\frac{A_1'(1+\rho_0)}{A_2'(1-\rho_0)}\right]$. It implies that for $A \in \boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{G}}}^{\mathbf{(1)}}$, the value for $\frac{A_1'(1+\rho_0)}{A_2'(1-\rho_0)} = \frac{(1-\rho_0)^2}{(1+\rho_0)^2}$ so the value of $\frac{1}{J^A(\rho_0,\rho_0)}\frac{\partial^2}{\partial\rho^2}\left[J^A(\rho_{{}_0},\rho)\right]\Bigg|_{\rho=\rho_0}$ is automatically fixed. But in this case putting $M_\rho := -2I_\rho'$, So for any $A \in \boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{G}}}^{\mathbf{(1)}}$, we have the second order term is identically

$$\Delta_2(\mathcal{G}) = 0. \tag{4.4.6}$$

So we have

$$P_{\rho_0}\left[\rho_{{}_0} \leq \rho^{1-\alpha}(\mathcal{G}_{\mathcal{A}}, \mathbb{X}, n)\right] - \big(1-\alpha\big) \;\;=\;\; o\bigg(\frac{1}{n}\bigg). \tag{4.4.7}$$

2. Similarly for $N(\mu,\mu^q)$ for $q > 0$(known) and $\mu > 0$, the data generating equation is

$$(S_1, S_2) = (\bar{X}_n, S_n) := \left(\mu + \mu^{q/2}Z, \mu^{q/2}\left(\frac{U}{n-1}\right)^{1/2}\right).$$

If one has an element $A \in \mathcal{A}_{\mathcal{G}}$, the structure of the Jacobian will be $J_n^A(\mathbb{X},\mu) = A_1'(\bar{X}_n)\left|1 + \frac{q(\bar{X}_n-\mu)}{2\mu}\right| + A_2'(S_n)\frac{qS_n}{2\mu}\ J^A(\mu_0,\mu)$ that converges to

$$J^A(\mu_0,\mu) := A_1'(\mu_0)\left|1 + \frac{q(\mu_0-\mu)}{2\mu}\right| + \frac{q}{2}A_2'(\mu_0^{\frac{q}{2}})\frac{\mu_0^{\frac{q}{2}}}{\mu}.$$

By similar techniques used in Bivariate normal contexts one can verify Assumptions 18 and 19 satisfied by the Jacobian $J_n^A(\mathbb{X},\mu)$. Using $J^A(\mu_0,\mu)$ in the equation $\Delta_1 = 0$, one has the following characterization

$$\boldsymbol{\mathcal{A}_{\mathcal{G}}^{(1)}} = \{\boldsymbol{A = (A_1, A_2) \in \mathcal{A}_{\mathcal{G}} : A_2'(\mathrm{x}^{\frac{q}{2}}) = A_1'(\mathrm{x}) q \mathrm{x}^{\frac{q}{2}-1},} \text{ for } \boldsymbol{x > 0}\}$$

which is clearly satisfied by our third choice (4.2.8) $A_1(\mathbf{x}) = \mathbf{x}^2, A_2(\mathbf{y}) = q\mathbf{y}^2$.

So for any $A \in \boldsymbol{\mathcal{A}_{\mathcal{G}}^{(1)}}$, the asymptotic coverage will be

$$P_{\mu_0}\left[\mu_0 \leq \mu^{1-\alpha}(\mathcal{G}_{\mathcal{A}}, \mathbb{X}, n)\right] - \left(1-\alpha\right) \quad = \quad \frac{c_2 \Delta_2(\mathcal{G})}{n} + o\left(\frac{1}{n}\right) \tag{4.4.8}$$

where $\Delta_2(\mathcal{G})$ can be found from (4.3.9).

4.5 Small Sample Simulation:

In this section we will demonstrate some simulation results to understand small sample implication of our work. In particular we chose The Basu's famous Bivariate normal $\mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ model for inference of ρ. We considered three Fiducial distributions in Example 4 respectively based on the simple data generating equation, DGE based on minimal sufficient statistic, and one candidate of the first order probability matching with transformation $A_1(x) = A_2(x) = \frac{1}{x}$ on two dimensional minimal sufficient statistics. In addition we also considered Jefrey based Posterior distribution and the second order probability matching data-dependent prior proposed by [68]. Note that in proposal of [68] one needs a MLE of ρ that cannot be computed in closed form from the likelihood (only tractable through numerical iterations that increases the computation time). So we used $\frac{1}{n}\sum_{i=1}^{n} X_i Y_i$ instead and since $\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \rho \sim N(0, \frac{\rho^2+1}{n})$, so this substitution will not make any difference except for a set with probability $O(e^{-an})$ for some $a > 0$.

For a dataset $\mathbf{x}$, let

$$\rho^{\alpha}(\mathbf{x}, J_1), \quad \rho^{\alpha}(\mathbf{x}, J_2), \quad \rho^{\alpha}(\mathbf{x}, J_3), \quad \rho^{\alpha}(\mathbf{x}, \text{Jeff}), \quad \rho^{\alpha}(\mathbf{x}, B_2)$$

denote α-th quantiles of simple fiducial distribution, Fiducial distribution based on Minimal sufficient statistics, First order probability matching fiducial distribution, Jeffreys' Posterior, second order matching posterior by respectively. We considered following order of quantiles $\alpha = (0.05, 0.15, 0.3, 0.5, 0.7, 0.8, 0.95)$ and sample size $n = 2, 3, 4, 5, 10, 25, 100$.

For each n and α, each cell of Table 4.1 - 4.5 exhibits simulated frequentist coverage of $(-1, \rho^{\alpha}(\mathbf{x}, J_1)], (-1, \rho^{\alpha}(\mathbf{x}, J_2)], (-1, \rho^{\alpha}(\mathbf{x}, J_3)], (-1, \rho^{\alpha}(\mathbf{x}, \text{Jeff})], (-1, \rho^{\alpha}(\mathbf{x}, B_2)]$ respectively from top to bottom each based on 5000 iterations while each table the data are generated from true value

$$\rho_0 = (0.05, 0.3, 0.5, 0.7, 0.9).$$

Observe that we considered only positive ρ_0 since by symmetry coverage of the lower interval for true ρ_0 will be same as the coverage of the upper interval with true $-\rho_0$.

We observe that posterior of second order probability matching prior works uniformly well while the data generating equation of Minimal sufficient statistics is quite poor. Relatively first order matching Fiducial distribution works notably well for the α close to 0.5. Also note z_α is close to 0 if α is close to 0.5 and in such cases the last two quanitities of the second order term will be close 0 as they are multiplied by z_α. From the result it also appears that for small sample size empirical coverage of second order matching Fiducial distribution is smaller than actual order of the quantile (hence liberal) while the second order Bayesian is very sharp. We realize that it is due to the first quantity of the second order term while $a_1(\theta_0), a_2(\theta_0)$ is attained at $n \to \infty$ but for small sample (like $n \leq 5$) that approximation is very crude adding negative values to the empirical coverage.

4.6 Conclusion and Open questions:

The study of non-informative priors and related variants of distributional inference has a long history and perhaps the best conclusion was made by Kass,Wasserman in [54]:

"... research on priors chosen by formal rules are serious and may not be dismissed lightly: When sample sizes are small (relative the number of parameters being estimated), it is dangerous to put faith in any default solution; but when asymptotics take over, Jeffreys rules and their variants remain reasonable choices."

Here we found few other variants of probability distributions (**GFD**) on Θ which are (when they are not exact) as good as the posterior distribution from Jeffrey's prior (satisfying invariance, and at least first order matching) and also they came from the Fiducial framework. We never explored here but there is a possibility of getting a second order matching Fiducial distribution too when there is a set of three minimal suffcient statistics for the unknown parameter and simplifying first and second order term will give us two equations over unknowns $\frac{A_2'(x)}{A_1'(x)}$ and $\frac{A_3'(x)}{A_1'(x)}$. Solving them may give a set of transformations of the form $A(x,y,z) = (A_1(x), A_2(y), A_3(z))$. One possible example would be the family

Table 4.1: Coverage for finite sample simulation for Bivariate normal with ρ_0= 0.05

ρ_0= 0.05	α=0.05	α=0.15	α=0.3	α=0.5	α=0.7	α=0.8	α=0.95
	0.0164	0.0804	0.2326	0.4808	0.7434	0.8498	0.9818
	0.0058	0.0376	0.1604	0.4638	0.8064	0.9018	0.9924
n=2	0.1080	0.2166	0.3550	0.4942	0.6380	0.7126	0.8960
	0.0726	0.1654	0.3126	0.4906	0.6762	0.7642	0.9352
	0.0682	0.1480	0.2960	0.4866	0.6974	0.7866	0.9422
	0.0238	0.1072	0.2616	0.4956	0.7370	0.8314	0.9702
	0.0116	0.0682	0.2174	0.4834	0.7646	0.8616	0.9842
n=3	0.0948	0.2024	0.3504	0.5050	0.6694	0.7398	0.9052
	0.0672	0.1652	0.3202	0.5020	0.6962	0.7770	0.9368
	0.0584	0.1512	0.2994	0.4982	0.7144	0.7960	0.9414
	0.0294	0.1128	0.2722	0.4906	0.7152	0.8256	0.9668
	0.0170	0.0872	0.2392	0.4816	0.7296	0.8456	0.9782
n=4	0.0780	0.1960	0.3428	0.5002	0.6610	0.7480	0.9134
	0.0604	0.1652	0.3170	0.4976	0.6846	0.7780	0.9364
	0.0558	0.1530	0.2998	0.4946	0.7010	0.7972	0.9442
	0.0304	0.1168	0.2720	0.5008	0.7100	0.8178	0.9644
	0.0212	0.0950	0.2498	0.4936	0.7218	0.8324	0.9738
n=5	0.0742	0.1800	0.3254	0.5108	0.6630	0.7648	0.9184
	0.0584	0.1588	0.3052	0.5072	0.6802	0.7876	0.9400
	0.0522	0.1462	0.2924	0.5042	0.6958	0.8014	0.9470
	0.0472	0.1460	0.2888	0.4868	0.7000	0.8004	0.9512
	0.0438	0.1390	0.2814	0.4830	0.7006	0.8026	0.9532
n=10	0.0580	0.1624	0.3022	0.4904	0.6930	0.7888	0.9424
	0.0546	0.1572	0.2994	0.4902	0.6972	0.7924	0.9464
	0.0524	0.1534	0.2952	0.4890	0.6998	0.7986	0.9484
	0.0442	0.1358	0.2788	0.4720	0.6846	0.7870	0.9426
	0.0348	0.1134	0.2488	0.4338	0.6514	0.7648	0.9362
n=25	0.0548	0.1498	0.3038	0.4964	0.7042	0.7986	0.9438
	0.0536	0.1470	0.3010	0.4938	0.7044	0.7988	0.9450
	0.0536	0.1456	0.3000	0.4946	0.7048	0.8010	0.9470
	0.0428	0.1548	0.2926	0.5068	0.6874	0.8026	0.9488
	0.0418	0.1526	0.2892	0.5042	0.6864	0.8024	0.9494
n=100	0.0460	0.1584	0.2978	0.5084	0.6864	0.8004	0.9462
	0.0450	0.1578	0.2970	0.5082	0.6866	0.8012	0.9472
	0.0442	0.1566	0.2956	0.5080	0.6884	0.8030	0.9484

Table 4.2: Coverage for finite sample simulation for Bivariate normal
with ρ_0= 0.3

ρ_0= 0.3	n=2	n=3	n=4	n=5	n=10	n=25	n=100
	0.0088	0.0176	0.0246	0.0246	0.0374	0.0422	0.0482
	0.0014	0.0052	0.0094	0.0112	0.0268	0.0374	0.0466
α=0.05	0.0858	0.0700	0.0682	0.0674	0.0600	0.0530	0.0534
	0.0638	0.0522	0.0568	0.0560	0.0554	0.0514	0.0528
	0.0596	0.0484	0.0550	0.0526	0.0522	0.0494	0.0516
	0.0610	0.0862	0.0984	0.1118	0.1246	0.1336	0.1510
	0.0170	0.0392	0.0608	0.0790	0.0968	0.1184	0.1438
α=0.15	0.2136	0.1946	0.1836	0.1812	0.1656	0.1522	0.1570
	0.1688	0.1578	0.1610	0.1636	0.1538	0.1490	0.1574
	0.1622	0.1504	0. 1500	0.1590	0.1504	0.1464	0.1562
	0.1966	0.2306	0.2350	0.2476	0.2724	0.2818	0.2892
	0.0942	0.1644	0.1802	0.2004	0.2332	0.2616	0.2792
α=0.3	0.3658	0.3492	0.3262	0.3274	0.3176	0.3054	0.2990
	0.3128	0.3126	0.2980	0.3042	0.3054	0.3020	0.2982
	0.2964	0.3002	0.2872	0.2946	0.3014	0.2996	0.2978
	0.4458	0.4476	0.4504	0.4604	0.4604	0.4826	0.4950
	0.3488	0.3930	0.3964	0.4162	0.4300	0.4616	0.4840
α=0.5	0.5168	0.5112	0.4986	0.5104	0.4912	0.4996	0.5056
	0.4992	0.4956	0.4886	0.4988	0.4846	0.4978	0.5050
	0.4918	0.4878	0.4838	0.4912	0.4814	0.4972	0.5054
	0.8320	0.8048	0.7936	0.7932	0.7836	0.7894	0.7904
	0.8552	0.8072	0.7914	0.7860	0.7718	0.7804	0.7812
α=0.8	0.7430	0.7458	0.7596	0.7644	0.7802	0.7934	0.7956
	0.7816	0.7718	0.7774	0.7818	0.7884	0.7964	0.7970
	0.8046	0.7918	0.7920	0.7924	0.7950	0.7992	0.7982
	0.9742	0.9648	0.9532	0.9528	0.9520	0.9482	0.9478
	0.9832	0.9748	0.9610	0.9576	0.9526	0.9466	0.9452
α=0.95	0.8904	0.9072	0.9166	0.9206	0.9382	0.9458	0.9484
	0.9326	0.9350	0.9360	0.9356	0.9466	0.9482	0.9488
	0.9428	0.9446	0.9422	0.9448	0.9518	0.9520	0.9498

$P_\theta := \left\{ N\left(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \theta^2 \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \right) \quad \text{for} \quad \theta \in (-1,1) \right\}$ where three dimensional minimal suffi-

Table 4.3: Coverage for finite sample simulation for Bivariate normal with ρ_0= 0.5

ρ_0= 0.5	α=0.05	α=0.15	α=0.3	α=0.5	α=0.7	α=0.8	α=0.95
	0.0058	0.0452	0.1764	0.3996	0.6668	0.7874	0.9566
	0.0010	0.0060	0.0512	0.2446	0.6198	0.7806	0.9648
n=2	0.0712	0.1936	0.3512	0.5116	0.6710	0.7402	0.8912
	0.0526	0.1578	0.3098	0.4802	0.6780	0.7654	0.9270
	0.0514	0.1526	0.3048	0.4770	0.6780	0.7786	0.9386
	0.0106	0.0734	0.2014	0.4236	0.6464	0.7750	0.9556
	0.0020	0.0230	0.1090	0.3152	0.5944	0.7514	0.9570
n=3	0.0666	0.1786	0.3374	0.5096	0.6703	0.7698	0.9150
	0.0544	0.1558	0.3058	0.4872	0.6686	0.7714	0.9354
	0.0534	0.1526	0.2962	0.4830	0.6724	0.7792	0.9454
	0.0210	0.0896	0.2258	0.4332	0.6566	0.7746	0.9474
	0.0054	0.0370	0.1444	0.3502	0.6072	0.7472	0.9488
n=4	0.0646	0.1664	0.3224	0.5082	0.6806	0.7742	0.9198
	0.0558	0.1476	0.2988	0.4880	0.6792	0.7796	0.9350
	0.0552	0.1454	0.2924	0.4868	0.6808	0.7862	0.9432
	0.0236	0.1016	0.2360	0.4266	0.6686	0.7720	0.9478
	0.0094	0.0534	0.1640	0.3554	0.6254	0.7424	0.9472
n=5	0.0578	0.1706	0.3194	0.4944	0.6940	0.7772	0.9306
	0.0528	0.1576	0.2978	0.4828	0.6914	0.7836	0.9432
	0.0520	0.1530	0.2964	0.4788	0.6944	0.7868	0.9492
	0.0326	0.1180	0.2642	0.4676	0.6654	0.7744	0.9378
	0.0210	0.0850	0.2052	0.4130	0.6266	0.7460	0.9320
n=10	0.0484	0.1486	0.3158	0.5104	0.6868	0.7898	0.9344
	0.0466	0.1430	0.3056	0.5044	0.6858	0.7912	0.9374
	0.0458	0.1428	0.3054	0.5044	0.6864	0.7932	0.9418
	0.0442	0.1358	0.2788	0.4720	0.6846	0.7870	0.9426
	0.0348	0.1134	0.2488	0.4338	0.6514	0.7648	0.9362
n=25	0.0548	0.1498	0.3038	0.4964	0.7042	0.7986	0.9438
	0.0536	0.1470	0.3010	0.4938	0.7044	0.7988	0.9450
	0.0536	0.1456	0.3000	0.4946	0.7048	0.8010	0.9470
	0.0474	0.1486	0.2830	0.4924	0.6954	0.8018	0.9464
	0.0428	0.1366	0.2680	0.4772	0.6784	0.7896	0.9428
n=100	0.0502	0.1556	0.2944	0.5046	0.7024	0.8108	0.9478
	0.0498	0.1552	0.2942	0.5044	0.7024	0.8110	0.9486
	0.0496	0.1552	0.2940	0.5046	0.7032	0.8114	0.9490

Table 4.4: Coverage for finite sample simulation for Bivariate normal with ρ_0= 0.7

ρ_0= 0.7	α=0.05	α=0.15	α=0.3	α=0.5	α=0.7	α=0.8	α=0.95
n=2	0.0046	0.0440	0.1506	0.3824	0.6284	0.7518	0.9502
	0.0000	0.0014	0.0208	0.1578	0.5104	0.6966	0.9456
	0.0568	0.1824	0.3186	0.5164	0.6788	0.7576	0.9068
	0.0490	0.1556	0.2848	0.4884	0.6696	0.7652	0.9316
	0.0474	0.1544	0.2832	0.4884	0.6742	0.7710	0.9406
n=3	0.0154	0.0756	0.1868	0.4152	0.6296	0.7636	0.9372
	0.0008	0.0132	0.0684	0.2620	0.5312	0.6958	0.9258
	0.0528	0.1636	0.3014	0.5086	0.6836	0.7862	0.9188
	0.0474	0.1484	0.2800	0.4894	0.6718	0.7860	0.9322
	0.0476	0.1482	0.2810	0.4934	0.6750	0.7910	0.9376
n=4	0.0186	0.0964	0.2248	0.4178	0.6404	0.7550	0.9392
	0.0034	0.0276	0.1178	0.2936	0.5448	0.6894	0.9290
	0.0558	0.1656	0.3106	0.4978	0.6930	0.7818	0.9320
	0.0496	0.1580	0.2956	0.4826	0.6836	0.7804	0.9386
	0.0504	0.1584	0.2970	0.4852	0.6830	0.7840	0.9436
n=5	0.0256	0.1006	0.2372	0.4366	0.6548	0.7614	0.9436
	0.0066	0.0454	0.1388	0.3176	0.5678	0.7054	0.9298
	0.0480	0.1508	0.3132	0.5034	0.7036	0.7896	0.9408
	0.0458	0.1432	0.3022	0.4948	0.6974	0.7886	0.9440
	0.0456	0.1442	0.3020	0.4952	0.7000	0.7918	0.9486
n=10	0.0390	0.1282	0.2596	0.4598	0.6650	0.7752	0.9384
	0.0208	0.0822	0.2018	0.3860	0.6078	0.7282	0.9218
	0.0554	0.1532	0.2982	0.4990	0.6924	0.7972	0.9428
	0.0546	0.1514	0.2940	0.4960	0.6910	0.7956	0.9440
	0.0544	0.1524	0.2954	0.4984	0.6944	0.7962	0.9464
n=25	0.0414	0.1432	0.2746	0.4854	0.6896	0.7868	0.9526
	0.0326	0.1178	0.2348	0.4382	0.6446	0.7560	0.9414
	0.0470	0.1548	0.2954	0.5074	0.7058	0.8008	0.9560
	0.0470	0.1546	0.2944	0.5068	0.7062	0.8002	0.9562
	0.0472	0.1546	0.2942	0.5068	0.7072	0.8018	0.9570
n=100	0.0466	0.1418	0.2928	0.4848	0.6928	0.7896	0.9532
	0.0402	0.1288	0.2734	0.4624	0.6702	0.7732	0.9472
	0.0498	0.1474	0.2988	0.4932	0.7002	0.7962	0.9546
	0.0498	0.1474	0.2988	0.4934	0.7002	0.7966	0.9548
	0.0498	0.1472	0.2992	0.4940	0.7008	0.7972	0.9552

Table 4.5: Coverage for finite sample simulation for Bivariate normal with ρ_0= 0.9

ρ_0= 0.9	α=0.05	α=0.15	α=0.3	α=0.5	α=0.7	α=0.8	α=0.95
	0.0026	0.044	0.1626	0.37066	0.596	0.732	0.936
	0.00	0.000	0.0046	0.05266	0.3406	0.547	0.920
n=2	0.0466	0.1546	0.3166	0.51066	0.6840	0.780	0.942
	0.0420	0.1373	0.2920	0.48866	0.6653	0.771	0.942
	0.0433	0.1420	0.2986	0.49666	0.6720	0.775	0.941
	0.0253	0.1020	0.2233	0.4206	0.6326	0.7593	0.9313
	0.0006	0.0046	0.0573	0.2120	0.4553	0.6166	0.8846
n=3	0.0600	0.1620	0.3046	0.4960	0.6900	0.8040	0.9433
	0.0586	0.1553	0.2960	0.4880	0.6833	0.7980	0.9440
	0.0593	0.1580	0.3000	0.4926	0.6853	0.8013	0.9480
	0.0273	0.12260	0.23800	0.47330	0.66460	0.75260	0.94460
	0.0040	0.03266	0.08933	0.29400	0.50733	0.64866	0.91466
n=4	0.0533	0.15660	0.29400	0.52200	0.70260	0.78660	0.95200
	0.0526	0.15330	0.29000	0.51800	0.69730	0.78260	0.95130
	0.0533	0.15530	0.29200	0.52066	0.70460	0.78600	0.95260
	0.02600	0.1226	0.2660	0.4726	0.6766	0.7640	0.946
	0.00466	0.0500	0.1446	0.3266	0.5513	0.6573	0.9176
n=5	0.04330	0.1540	0.3020	0.5060	0.7060	0.7913	0.9526
	0.04260	0.1533	0.3007	0.5026	0.7026	0.7886	0.9523
	0.04460	0.1546	0.3026	0.5060	0.7060	0.7926	0.953
	0.0360	0.1440	0.2853	0.4780	0.6713	0.7960	0.9430
	0.0140	0.0880	0.2060	0.3826	0.5830	0.7300	0.9170
n=10	0.0413	0.1513	0.3020	0.4893	0.6853	0.8086	0.9453
	0.0418	0.1515	0.3013	0.4886	0.6846	0.8073	0.9453
	0.0406	0.1526	0.3030	0.4906	0.6860	0.8080	0.9460
	0.0546	0.1260	0.2880	0.4846	0.7080	0.7940	0.9453
	0.0346	0.0913	0.2326	0.4280	0.6650	0.7520	0.9326
n=25	0.0560	0.1273	0.2940	0.4890	0.7126	0.7986	0.9486
	0.0562	0.1270	0.2940	0.4890	0.7126	0.7986	0.9486
	0.0559	0.1270	0.2940	0.4893	0.7130	0.7980	0.9486
	0.0540	0.1506	0.3033	0.5080	0.6840	0.8040	0.9450
	0.0473	0.1306	0.2820	0.4853	0.6620	0.7913	0.9366
n=100	0.0560	0.1513	0.3073	0.5106	0.6866	0.8060	0.9460
	0.0560	0.1513	0.3073	0.5106	0.6860	0.8060	0.9460
	0.0560	0.1506	0.3073	0.5106	0.6860	0.8073	0.9460

cient statistics for θ can be taken as

$$(S_1, S_2, S_3) = \left(\bar{X} + \bar{Y}, \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2, \frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i - \bar{X} - \bar{Y})^2 \right).$$

There is a possibility of getting general results on higher order matching generating equations but they will be case specific and basic ideas are similar. Following are some other open questions related to the extension of the idea of higher order asymptotics as well as the choice of norms:

1. **Different Choices of norm in (1.3.3):** One of the essential ingredients in our work was the derivation of (1.3.4) with L_∞ norm through increasing precision asymptotics. In (1.3.3) choosing L_∞ is also very intuitive and a lot easier to handle to get the distribution of the weak limit (as $\epsilon \downarrow 0$). A natural extension would be to find the structure of Jacobian for the L_2 norm which is proved to be

$$\sqrt{\det\left(\left(\frac{d}{d\theta}\mathbf{G}(\mathbf{u}, \theta)\right)' \left(\frac{d}{d\theta}\mathbf{G}(\mathbf{u}, \theta)\right) \right)}\Bigg|_{\mathbf{u}=G^{-1}(X,\theta)}.$$

   So once we have a good handle on that we will be able to generalize the probability matching criteria for fiducial distributions defined under general norm structures (L_p for $2 < p < \infty$) in (1.3.3).

2. **Non-regular cases:** When the true distribution is supported on $(a(\theta), b(\theta))$ with $|a'(\theta)| \leq |b'(\theta)|$ (for example $U(\theta, \theta^2)$ for $\theta > 1$) then the condition "$\{x : f(x \mid \theta) > 0\}$ doesn't depend on θ" of Theorem 3, gets violated and the expansion of the fiducial distribution

will not converge to Normal distribution anymore. For $U(a(\theta), b(\theta))$ the fiducial distribution of θ on the basis of n iid observations (under assumption both $a(\theta)$ and $b(\theta)$ are increasing and continuous in θ) is

$$f(\theta \mid \mathbb{X}) \propto \frac{a'(\theta) - a(\theta)[\log b(\theta)]' + \bar{X}_n[\log b(\theta)]'}{b(\theta)^n} . 1_{\{a(\theta)-b(\theta)<X_{(1)}, \quad a(\theta)+b(\theta)>X_{(n)}\}}$$

where $(X_{(1)}, X_{(n)}) = (\min \mathbb{X}, \max \mathbb{X}), \bar{X}_n =$ mean($\mathbb{X}$). Even in probability matching prior context under a more restrictive condition, we have seen there is only second order terms present [43] in the frequentist coverage of the posterior quantile of $n(\theta - \hat{\theta}_n)$. We expect a similar result to hold in Fiducial context but with a change that should come from the term $(W_n^{(1)}(\mathbb{X}))^2$.

3. **Multi-parameter context:** Proving analogue version of Theorem 3 in multi parameter cases where there is only one parameter of interest and rest are nuisance is more involved. Generally the Jacobian becomes a *U-Statistics*. Since higher order expansion of fiducial quantile requires convergence of fluctuation of scaled Jacobian (like Assumption 19), deriving concentration properties of U-Statistics (that the Jacobian for multiparameter case resembles) is essential in that context which is very challenging.

4.7 Proof of Theorem 3:

We prove Theorem 3 with a number of steps. First we prove a lemma on the expansion of the fiducial density and then we will give an asymptotic expansion of the Fiducial quantile in Corollary 4.7.1. After that in order to get the frequentist coverage of the quantile with the obtained expression from Corollary 4.7.1, we will proceed with Shrinkage method. Follow the definition of $\mathbb{W}_n^{(m)}(\mathbb{X})$ from (4.3.6). In order to lessen notational burden we denote

$$K(\theta_0, \mathbb{X}, y) := \left[1 + \frac{J'(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)} \frac{y}{\sqrt{nc}} + \frac{1}{n}\left(\mathbb{W}_n^{(1)}(\mathbb{X})\frac{y}{\sqrt{c}} + \frac{J''(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)}\frac{y^2}{2c}\right)\right].$$

Lemma 19. *Suppose Assumptions 16,17,18 hold with $m = 2$. Following quantity is*

$$\begin{aligned} I_{\mathbb{R}} := n \int_{\mathbb{R}} \Bigg| J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) e^{n[L_n(\hat{\theta}_n + \frac{y}{\sqrt{nc}}) - L_n(\hat{\theta}_n)]} - \\ J_n(\mathbb{X}, \hat{\theta}_n) e^{-\frac{y^2}{2}} \left(1 + R_n(\hat{\theta}_n) + \frac{R_n(\hat{\theta}_n)^2}{2}\right) K(\theta_0, \mathbb{X}, y) \Bigg| dy = o_{P_{\theta_0}}(1). \end{aligned} \tag{4.7.1}$$

Proof of Lemma 19: We will proceed traditionally by breaking the integral in three disjoint regions. Denoting $I_{\mathbb{R}}$ as the integral appeared in the left hand side of the Lemma, we have

$$I_{\mathbb{R}} \leq I_{A_1} + I_{A_2} + I_{A_3} \tag{4.7.2}$$

where $A_1 = \{y : |y| < C \log \sqrt{n}\}$, $A_2 = \{y : C \log \sqrt{n} \leq |y| \leq \delta\sqrt{n}\}$, $A_3 = \{y : |y| > \delta\sqrt{n}\}$. The choice of C, δ will be specified later. The third term of (4.7.2) can be written as $I_{A_3} \leq I_{A_3}^1 + I_{A_3}^2$ where

$$\begin{aligned} I_{A_3}^1 &= n \int_{A_3} J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) e^{n[L_n(\hat{\theta}_n + \frac{y}{\sqrt{nc}}) - L_n(\hat{\theta}_n)]} dy, \\ I_{A_3}^2 &= n \int_{A_3} J_n(\mathbb{X}, \hat{\theta}_n) e^{-\frac{y^2}{2}} \left(1 + R_n(\hat{\theta}_n) + \frac{R_n(\hat{\theta}_n)^2}{2}\right) K(\theta_0, \mathbb{X}, y) dy. \end{aligned}$$

Expanding $I^1_{A_3}$, one gets

$$I^1_{A_3} = n\sum_{i=1}^{n}\int_{A_3} J(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})f(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})e^{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]-\log f(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})}dy$$

$$= n\sum_{i=1}^{n}\int_{A_3} J(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})f(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})e^{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]\left[1-\frac{\log f(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})}{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]}\right]}dy$$

$$\le n^2 \sup_{\mathbf{x}\in\mathbb{R}^n}\sup_{i=1,\dots,n}\int_{A_3} J(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})f(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})\times e^{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]\left[1-\frac{\log f(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})}{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]}\right]}dy$$

Notice that from Assumption 17(b) P_{θ_0} almost surely $n^{-s}\int_{\mathbb{R}} J(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})f(X_i,\hat{\theta}_n+\frac{y}{\sqrt{nc}})dy<\infty$. Now by Assumption 17 one has the exponential term to decay as $e^{-n(1-c)\epsilon}$ in probability and that term multiplied with n will also goes to 0 in probability. Rest will follow by dominated convergence theorem. For $I^2_{A_3}$ we have $J_n(\mathbb{X},\hat{\theta}_n)\to^P J(\theta_0,\theta_0)$ from Assumption 18. The multiplicative parts are the integrals $\int_{A_3} y^\alpha e^{-y^2}dy$ for $\alpha=0,1,2$ which under A_3 decays exponentially to 0 resulting the P_{θ_0} limit of the second term 0.

Now consider I_{A_1}. Denote $\frac{1}{120}\frac{y^5}{n^{3/2}c^2}L_n^{(5)}(\theta')$ by M_n. The first integral in region A_1, can be expanded as $I_{A_1}\le I^1_{A_1}+I^2_{A_1}$ where

$$I^1_{A_1} := n\int_{A_1} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}})e^{-\frac{y^2}{2}}\left|e^{R_n(\hat{\theta}_n)+M_n}-1-R_n(\hat{\theta}_n)-\frac{R_n(\hat{\theta}_n)^2}{2}\right|dy \quad \text{and}$$
$$I^2_{A_1} := n\int_{A_1} I(\mathbb{X},n,y)e^{-\frac{y^2}{2}}\left(1+R_n(\hat{\theta}_n)+\frac{R_n(\hat{\theta}_n)^2}{2}\right)dy.$$

given $I(\mathbb{X},n,y) := \left|J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}})-J_n(\mathbb{X},\hat{\theta}_n)K(\theta_0,\mathbb{X},y)\right|$. Note that under A_1, the quantity $n\left(M_n+\frac{M_n^2}{2}+R_n(\hat{\theta}_n)M_n\right) = O_p(\frac{(\log\sqrt{n})^5}{\sqrt{n}})$. Also $L_n^{(5)}(\theta')$ is $O_p(1)$ for $\theta' \in (\theta_0-\delta,\theta_0+\delta)$ along with $L_n^{(3)}(\hat{\theta}_n)$ and $L_n^{(4)}(\hat{\theta}_n)$. Since R_n+M_n is $O_p(\frac{\log^3(\sqrt{n})}{\sqrt{n}})$. Using the inequality

$$e^x-1-x-\frac{x^2}{2}\leq\frac{x^3}{6(1-\frac{x}{4})}\quad\text{for}\quad x\in(0,4)$$

the first term of (4.7.3) can be written as

$$\begin{aligned}
I_{A_1}^1 &\leq n\int_{A_1} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}})e^{-\frac{y^2}{2}}\Big|e^{R_n+M_n}-1-(R_n+M_n)-\frac{(R_n+M_n)^2}{2}\Big|dy\\
&+ n\int_{A_1} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}})e^{-\frac{y^2}{2}}\left(M_n+\frac{M_n^2}{2}+R_n(\hat{\theta}_n)M_n\right)dy\\
&\leq n\int_{A_1} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}})e^{-\frac{y^2}{2}}\Big|\frac{(R_n+M_n)^3}{6(1-\frac{(R_n+M_n)}{4})}\Big|dy\\
&\quad +O_p(\frac{(\log\sqrt{n})^5}{\sqrt{n}})\sup_{y\in A_1} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}}).
\end{aligned}$$

Now $\sup_{y\in A_1} n(R_n+M_n)^3 \leq \sup_{y\in A_1}\frac{y^9}{\sqrt{n}}\max\left[\frac{L_n^{(5)}(\theta')}{c^{\frac{3}{2}}},\frac{L_n^{(3)}(\hat{\theta}_n)}{c^2}\right] \leq \frac{\log(\sqrt{n})^9}{\sqrt{n}}O_p(1)$. So we have

$$I_{A_1}^1 \leq \sup_{y\in A_1} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}})\Big[O_p\left(\frac{\log(\sqrt{n})^9}{\sqrt{n}}\right)\int_{A_1} e^{-\frac{y^2}{2}}dy+O_p(\frac{(\log\sqrt{n})^5}{\sqrt{n}})\Big]. \quad (4.7.3)$$

Since $\hat{\theta}_n \to \theta_0 \quad a.s$, under A_1, we have $(\hat{\theta}_n,\hat{\theta}_n+\frac{y}{\sqrt{nc}}) \subset (\theta_0-\delta,\theta_0+\delta)$ with P_{θ_0} probability 1. We have almost surely $\sup_{y\in A_1} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}}) \leq \sup_{\theta'\in(\theta_0-\delta,\theta_0+\delta)} J_n(\mathbb{X},\theta')$. From Assumption 18 one has almost surely

$$I_{A_1}^1 \leq \sup_{\theta'\in(\theta_0-\delta,\theta_0+\delta)} J_n(\mathbb{X},\theta')\Big[O_p\left(\frac{\log(\sqrt{n})^9}{\sqrt{n}}\right)\int_{A_1} e^{-\frac{y^2}{2}}dy+O_p(\frac{(\log\sqrt{n})^5}{\sqrt{n}})\Big].$$

Now using Wald's theorem one has $\sup_{\theta' \in (\theta_0-\delta,\theta_0+\delta)} |J_n(\mathbb{X},\theta') - J(\theta_0,\theta')| \rightarrow^{a.s} 0$, resulting the following statement almost surely

$$I_{A_1}^1 \leq \sup_{\theta' \in (\theta_0-\delta,\theta_0+\delta)} J(\theta_0,\theta')\Big[O_p\left(\frac{\log(\sqrt{n})^9}{\sqrt{n}}\right)\int_{A_1} e^{-\frac{y^2}{2}}dy + O_p(\frac{(\log\sqrt{n})^5}{\sqrt{n}})\Big].$$

which is $o_{p_{\theta_0}}(1)$. Now from (4.3.6) the second term of I_{A_1}

$$\begin{aligned}
I_{A_1}^2 &:= n\int_{A_1} I(\mathbb{X},n,y)e^{-\frac{y^2}{2}}(1+R_n)dy \\
&\leq \sup_{\theta' \in (\hat{\theta}_n,\hat{\theta}_n+\frac{y}{\sqrt{nc}})} J_n(\mathbb{X},\hat{\theta}_n)\int_{A_1}\frac{1}{n^{1/2}}\left(\mathbb{W}_n^{(2)}(\mathbb{X})\frac{y^2}{2c} + \frac{J_n'''(\mathbb{X},\theta')}{J_n(\mathbb{X},\hat{\theta}_n)}\frac{y^3}{6(c)^{3/2}}\right)\times \\
&\quad e^{-\frac{y^2}{2}}(1+R_n(\hat{\theta}_n))dy.
\end{aligned}$$

Again similarly using almost sure convergence of the event $(\hat{\theta}_n,\hat{\theta}_n+\frac{y}{\sqrt{nc}}) \subset (\theta_0-\delta,\theta_0+\delta)$ and Assumption 18 on $J_n'''(\mathbb{X},\theta')$ we get $I_{A_1}^2$ is of $o_{P_{\theta_0}}(1)$.

Next consider the integral I_{A_2}, that can bounded above by $I_{A_2} \leq I_{A_2}^1 + I_{A_2}^2$ where

$$\begin{aligned}
I_{A_2}^1 &:= n\int_{A_2} J_n(\mathbb{X},\hat{\theta}_n+\frac{y}{\sqrt{nc}})e^{-\frac{y^2}{2}+R_n+M_n}dy, \\
I_{A_2}^2 &:= n\int_{A_2} J_n(\mathbb{X},\hat{\theta}_n)e^{-\frac{y^2}{2}}\left(1+R_n(\hat{\theta}_n)+\frac{R_n(\hat{\theta}_n)^2}{2}\right)K(\theta_0,\mathbb{X},y)dy.
\end{aligned}$$

Consider the term $I_{A_2}^2$. Note that under A_2 for $n \geq e^4$, $(\log\sqrt{n})^2 > \log n$ and also using the fact $\frac{|y|}{\sqrt{n}} < \delta$ the quantity

$$R_n(\hat{\theta}_n) \;\leq\; \frac{1}{6}\delta^3 n \frac{L_n^{(3)}(\hat{\theta}_n)}{c^{3/2}} + \frac{1}{24}\delta^4 n \frac{L_n^{(4)}(\hat{\theta}_n)}{c^2} = O_{P_{\theta_0}}(n)$$

resulting $\left(1 + R_n(\hat{\theta}_n) + \frac{R_n(\hat{\theta}_n)^2}{2}\right)$ is $O_{P_{\theta_0}}(n^2)$. Also from Assumption 18 the remaining term $J_n(\mathbb{X}, \hat{\theta}_n)K(\theta_0, \mathbb{X}, y)$ is $O_{P_{\theta_0}}(1)$. So the upper bound of the second integral $I^2_{A_2}$ is bounded by

$$\begin{aligned} & O_{P_{\theta_0}}(n^3) J_n(\mathbb{X}, \hat{\theta}_n)\Bigg[1 + \frac{J'(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)} \frac{\delta}{\sqrt{c}} \\ + \; & \frac{1}{n}\left(\mathbb{W}_n^{(1)}(\mathbb{X}) \frac{\delta\sqrt{n}}{\sqrt{c}} + \frac{J''(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)} \frac{n\delta^2}{2c}\right)\Bigg] . e^{-\frac{C^2}{2}\log n}\left[\delta\sqrt{n} - C\log\sqrt{n}\right] \\ = \; & O_{P_{\theta_0}}(n^{\frac{7}{2} - \frac{C^2}{2}}) \end{aligned} \tag{4.7.4}$$

which goes to 0 in probability if we choose $C > \sqrt{7}$. This result is due to convergence of $J_n(\mathbb{X}, \hat{\theta}_n), J'_n(\mathbb{X}, \hat{\theta}_n)$ respectively to $J(\theta_0, \theta_0), J'(\theta_0, \theta_0)$ which is validated from Assumption 18. Now considering the first term of the integral (4.7.4) we have $\frac{|y|}{\sqrt{n}} < \delta$, We have under A_2

$$\begin{aligned} |R_n(\hat{\theta}_n)| &\leq \frac{1}{6} \frac{\delta y^2 L_n^{(3)}(\hat{\theta}_n)}{c^{\frac{3}{2}}} + \frac{1}{24}\delta^2 y^2 \frac{L_n^{(4)}(\hat{\theta}_n)}{c^2}, \\ |M_n| &= \frac{1}{120} \frac{y^5}{n^{3/2}c^2} L_n^{(5)}(\theta') \leq \frac{1}{120} \frac{y^2\delta^3}{c^2} L_n^{(5)}(\theta') \end{aligned} \tag{4.7.5}$$

and since under A_2 the quantities $\sup_{\theta' \in (\hat{\theta}_n, \hat{\theta}_n + \frac{y}{\sqrt{nc}})} L_n^{(4)}(\theta'), \frac{L_n^{(4)}(\hat{\theta}_n)}{c^2}$ and $\frac{L_n^{(3)}(\hat{\theta}_n)}{c^{3/2}}$ are $O_p(1)$, given a small $\epsilon > 0$ one can always choose a δ so that we can get

$$P_{\theta_0}\big\{ -\frac{y^2}{2} + R_n(\hat{\theta}_n) + M_n < -\frac{y^2}{4}, \quad \forall y \in A_2 \big\} > 1 - \epsilon \quad \text{for} \quad n > n_0. \tag{4.7.6}$$

So with probability greater than $1 - \epsilon$

$$\begin{aligned} n \int_{A_2} J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) e^{-\frac{y^2}{2} + R_n + M_n} dy \quad &\leq \quad \sup_{y \in A_2} J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) \quad n \int_{A_2} e^{-\frac{y^2}{4}} dy \\ &\to^{a.s} \quad 0 \quad \text{as} \quad n \to \infty. \end{aligned} \tag{4.7.7}$$

The last line follows from the fact under A_2, $\quad (\hat{\theta}_n, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) \subset (\theta_0 - \delta, \theta_0 + \delta)$ almost surely and then by applying Assumption 18, $\sup_{y \in A_2} J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) \leq \sup_{\theta \in (\theta_0 - \delta, \theta_0 + \delta)} J(\theta_0, \theta)$ asymptotically almost surely. The integral will converge to 0 as $n \to \infty$ by choosing a bigger C. Choice of δ will be specified by (4.7.6) given a small $\epsilon > 0$.

$\square$

Now it's obvious to conclude from Lemma 19 that for any $A \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} &\int_A e^{-\frac{y^2}{2}} \left(1 + R_n(\hat{\theta}_n) + \frac{R_n(\hat{\theta}_n)^2}{2}\right) K(\theta_0, \mathbb{X}, y) dy \\ = &\int_A e^{-\frac{y^2}{2}} \left[1 + \frac{1}{\sqrt{n}} \left(A_1 y + A_3 y^3\right) + \frac{1}{n}\left(A_2 y^2 + A_4 y^4 + A_6 y^6 + W_n^{(1)} \frac{y}{\sqrt{c}}\right)\right] dy \\ &+ o_{p_{\theta_0}}\left(\frac{1}{n}\right) \end{aligned} \tag{4.7.8}$$

where

$$A_1 := c^{-\frac{1}{2}} \frac{J'(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)}, A_2 := \frac{1}{2} c^{-1} \frac{J''(\theta_0, \hat{\theta}_n)}{J(\theta_0, \hat{\theta}_n)}, \qquad A_3 := \frac{1}{6} c^{-\frac{3}{2}} a, \quad A_4 := A_1 A_3 + \frac{1}{24} c^{-2} a_4,$$
$$A_6 \;:=\; \frac{1}{2} A_3^2.$$

(4.7.8) follows from the fact that all higher order terms will accumulate in $o_{p_{\theta_0}}(\frac{1}{n})$. For an illustration taking just one cross-product term of second term of $R_n(\hat{\theta}_n)$ and $\frac{A_1 y}{\sqrt{n}}$, one has

$$\frac{1}{24} \frac{y^4}{nc^2} L_n^{(4)}(\hat{\theta}_n) \frac{A_1 y}{\sqrt{n}} \;=\; \frac{1}{24} \frac{A_1 y^5}{n\sqrt{n}c^2} L_n^{(4)}(\hat{\theta}_n) 1_{\{|y| \leq \log n\}} + \frac{1}{24} \frac{A_1 y^5}{n\sqrt{n}c^2} L_n^{(4)}(\hat{\theta}_n) 1_{\{|y| > \log n\}}.$$

Since $R_n(\hat{\theta}_n).A_1$ is $O_{P_{\theta_0}}(1)$

$$\begin{aligned} \int_{\mathbb{R}} e^{-\frac{y^2}{2}} R_n(\hat{\theta}_n) \frac{A_1 y}{\sqrt{n}} dy \;&=\; \frac{L_n^{(4)}(\hat{\theta}_n) A_1}{24 c^2 n^{\frac{3}{2}}} \int_{|y| \leq \log n} y^5 e^{-\frac{y^2}{2}} dy + \frac{L_n^{(4)} A_1}{24 c^2 n^{\frac{3}{2}}} \int_{|y| > \log n} y^5 e^{-\frac{y^2}{2}} dy \\ &\leq\; O_{P_{\theta_0}}\left(\frac{(\log n)^5}{n^{\frac{3}{2}}} \right) + \frac{L_n^{(4)} A_1}{24 c^2 n^{\frac{3}{2}}} \int_{|y| > \log n} y^5 e^{-\frac{y^2}{2}} dy \end{aligned} \qquad (4.7.9)$$

Since Gamma distribution is exponentially tailed, whole R.H.S of (4.7.9) is of $o_{P_{\theta_0}}(\frac{1}{n})$. Now note that the formula for r-th (even) central moment of standard normal distribution $EX^r := (r-1)(r-3)\ldots 1$. Dividing the quantity $J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) e^{n[L_n(\hat{\theta}_n + \frac{y}{\sqrt{nc}}) - L_n(\hat{\theta}_n)]}$ with the expansion of the denominator $\int_{\mathbb{R}} J_n(\mathbb{X}, \hat{\theta}_n + \frac{y}{\sqrt{nc}}) e^{n[L_n(\hat{\theta}_n + \frac{y}{\sqrt{nc}}) - L_n(\hat{\theta}_n)]} dy$, one has the asymptotic expansion of the fiducial density (upto second order in terms of expansion with respect to $\frac{1}{\sqrt{n}}$) of y

$$\begin{aligned} f_{\mathcal{G}}(y \mid \mathbb{X}) &= \phi(y)\Bigg(1 + \frac{1}{\sqrt{n}}\left(A_1 y + A_3 y^3\right) + \frac{1}{n}\Bigg(A_2(y^2-1) + A_4(y^4-3) + A_6(y^6-15) \\ &+ W_n^{(1)}(\mathbb{X})\frac{y}{\sqrt{c}}\Bigg)\Bigg) + o_{p_{\theta_0}}\left(\frac{1}{n}\right) \end{aligned} \tag{4.7.10}$$

where $\phi(.)$ is the density function of the normal distribution. From (4.7.8) we get (4.7.10) using the power series expansion $\frac{1}{1+x} = \sum_{i=1}^{\infty}(-1)^i x^i$ given $|x| < 1$ on first two ordered terms.

Remark 16. *The conclusion (4.7.10) will remain unchanged if $W_n^{(1)}(\mathbb{X})$ is replaced by a random variable $\widehat{W}_n^{(1)}(\mathbb{X})$ that is $\sigma(\mathbb{X})$ measurable with the property*

$$P_{\theta_0}\left[W_n^{(1)}(\mathbb{X}) \neq \widehat{W}_n^{(1)}(\mathbb{X})\right] = e^{-cn} \qquad \textit{for some } c > 0.$$

It is because the quantity $\left(W_n^{(1)}(\mathbb{X}) - \widehat{W}_n^{(1)}(\mathbb{X})\right)$ multiplied with any polynomial term of n will remain $O_{P_{\theta_0}}(e^{-cn})$ so that doesn't hamper in any specific polynomial order terms.

Recall the classical orthogonal **Hermite polynomials** $\{H_n(x)\}_{n\geq 1}$ which is defined as

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}}\left[\frac{d^n}{dx^n}e^{-\frac{x^2}{2}}\right].$$

First few Hermite polynomials are

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3,$$
$$H_5(x) = y^5 - 10y^3 + 15y, \qquad H_6(x) = y^6 - 15y^4 + 45y^2 - 15.$$

Following properties hold where $\phi(x)$ is the density of normal distribution: For all $a \in \mathtt{R}, \forall n \geq 2$

$$\int_{-\infty}^{a} H_1(y)\phi(y)dy = -\phi(a) \quad \text{and} \quad \int_{-\infty}^{a} H_n(y)\phi(y)dy = -H_{n-1}(a)\phi(a) \quad . \quad (4.7.11)$$

Expressing (4.7.10) with each coefficient in terms of Hermite polynomials we get,

$$\begin{aligned} f_{\mathcal{G}}(y) &= \phi(y)\Bigg(1 + \frac{1}{\sqrt{n}}\left(G_1H_1(y) + G_3H_3(y)\right) + \frac{1}{n}\Bigg(G_2H_2(y) + G_4H_4(y) \\ &+ G_6H_6(y) + W_n^{(1)}\frac{H_1(y)}{\sqrt{c}}\Bigg)\Bigg) + o_{p_{\theta_0}}\left(\frac{1}{n}\right) \qquad (4.7.12) \end{aligned}$$

where

$$G_1 := A_1 + 3A_3, \quad G_2 = A_2 + 6A_4 + 45A_6, \quad G_3 := A_3, \quad G_4 = A_4 + 15A_6 \quad G_6 = A_6.$$

Define further

$$\begin{aligned} \beta_1 &:= G_1 + G_3.H_2(z), \\ \beta_2 &:= 2z\beta_1G_3 - \frac{1}{2}\beta_1^2 z + G_2H_1(z) + G_4H_3(z) + G_6H_5(z) + \frac{W_n^{(1)}(\mathbb{X})}{\sqrt{c}} \qquad (4.7.13) \end{aligned}$$

This following illustration is similar with Theorem 2.3.1 of [23] which gives asymptotic expansion of $(1-\alpha)$-th fiducial quantile.

Corollary 4.7.1. *Denote* $\theta^{(1-\alpha)}(\mathbb{X}, \mathcal{G}) := \hat{\theta} + (nc)^{-\frac{1}{2}}(z + n^{-\frac{1}{2}}\beta_1 + n^{-1}\beta_2)$. *Suppose Assumptions* (A_1)-(A_5) *of asymptotic normality of likelihood expansion, along with Assumption*

16,17 18 with $m = 1$ hold. Then we have

$$P^{\mathcal{G}}\left[\theta \leq \theta^{(1-\alpha)}(\mathbb{X}, \mathcal{G})\middle|\mathbb{X}\right] = 1 - \alpha + o_{p_{\theta_0}}(n^{-1}). \tag{4.7.14}$$

Proof of Corollary 4.7.1:

The concerned quantity

$$\begin{aligned}
& P^{\mathcal{G}}\left[\theta \leq \theta^{(1-\alpha)}(\mathbb{X}, \mathcal{G})\middle|\mathbb{X}\right] = P\left[\theta \leq \hat{\theta} + (nc)^{-\frac{1}{2}}\left\{z + n^{-\frac{1}{2}}\beta_1 + n^{-1}\beta_2\right\}\middle|\mathbb{X}\right] \\
= \ & P\left[y \leq z + n^{-\frac{1}{2}}\beta_1 + n^{-1}\beta_2\middle|\mathbb{X}\right] \\
= \ & \int_{-\infty}^{z+n^{-\frac{1}{2}}\beta_1+n^{-1}\beta_2} \phi(y)\left[1 + \frac{1}{\sqrt{n}}\left(G_1 H_1(y) + G_3 H_3(y)\right)\right] dy \\
& + \frac{1}{n}\int_{-\infty}^{z+n^{-\frac{1}{2}}\beta_1+n^{-1}\beta_2} \phi(y)\left[\left(G_2 H_2(y) + G_4 H_4(y) + G_6 H_6(y) + W_n^{(1)}\frac{H_1(y)}{\sqrt{c}}\right)\right] dy \\
& + o_p(n^{-1}).
\end{aligned}$$

Using the properties of Hermite polynomials on (4.7.11) one easily gets

$$\begin{aligned}
& P^{\mathcal{G}}\left[\theta \leq \theta^{(1-\alpha)}(F, \mathbb{X}, n)\middle|\mathbb{X}\right] = \Phi(z + n^{-\frac{1}{2}}\beta_1 + n^{-1}\beta_2) \\
& -n^{-\frac{1}{2}}.\phi(z + n^{-\frac{1}{2}}\beta)\left[G_1 + G_3 H_2(z + n^{-\frac{1}{2}}\beta_1)\right] - n^{-1}\phi(z) \times \\
& \Bigg[G_2 H_1(z) + G_4 H_3(z) + G_6 H_5(z) \\
& + \frac{W_n^{(1)}(\mathbb{X})}{\sqrt{c}}\Bigg] + o_p(n^{-1}).
\end{aligned} \tag{4.7.15}$$

Using Taylor's expansions of $\Phi(x), \phi(x)$ and accumulating the higher order terms into $o_p(n^{-1})$, the RHS of (4.7.15) is simplified to

$$
\begin{aligned}
& P^{\mathcal{G}}\left[\theta \leq \theta^{(1-\alpha)}(\mathbb{X},\mathcal{G})\middle|\mathbb{X}\right] = \Phi(z) + n^{-\frac{1}{2}}\phi(z)\left\{\beta_1 - G_1 - G_3 H_2(z)\right\} \\
+ \quad & n^{-1}\phi(z)\Bigg[\beta_2 - 2z\beta_1 G_3 - \frac{1}{2}\beta_1^2 z + \beta_1 z\Bigg\{G_1 + G_3 H_2(z)\Bigg\} - G_2 H_1(z) \\
& -G_4 H_3(z) - G_6 H_5(z) - \frac{W_n^{(1)}(\mathbb{X})}{\sqrt{c}}\Bigg] + o_p(n^{-1}) = 1 - \alpha + o_p(n^{-1}) \quad (4.7.16)
\end{aligned}
$$

where (4.7.16) follows from the definitions of β_1 and β_2. Corollary 4.7.1 follows from that.

□

Higher order asymptotics in context of Probability matching prior is an old topic and well documented in [23].

The idea of Shrinkage method was essentially originated from [10] in context of establishing higher order asymptotics of Bertlett test statistics. In general it is used to find an expansion of $E_{\theta_0}\left[g(\mathbb{X},\theta)\right]$ for any function $g(\mathbb{X},\theta)$ (in our case $g(\mathbb{X},\theta) := 1_{\{\theta \leq \theta^{(1-\alpha)}(\mathbb{X},\mathcal{G})\}}$). Some relevant works on probability matching data dependent prior were done in [67], [68] but *data dependence is either coming from moments or the maximum likelihood estimator. In comparison to that here the term $W_n^{(1)}(\mathbb{X})$ is much like a ratio estimator where its higher order expansion is interestingly critical for the terms after first order which makes the following calculation relevant.* In order to implement Shrinkage method one formulates an auxiliary prior $\bar{\pi}$ with properties that it is proper, supported on a compact set, having true θ_0 in its interior. It vanishes on the boundary of the support while taking strictly positive values in the interior. It also satisfies all the conditions B_m $(m = 1, 2)$ in [10] ensuring smoothness of $\bar{\pi}$,and $\log \bar{\pi}$ and and its derivatives near the boundary of the support. Basic steps of Shrinkage method (for upto second order) are following:

1. **Step 1:** Start with an auxiliary prior $\bar{\pi}$ with a compact support $\subseteq \mathcal{H}$ containing θ_0 as an interior point. We will find the expansion of $E^{\bar{\pi}}(g(\mathbb{X},\theta)|\mathbb{X})$ upto $o_{P_{\theta_0}}(\frac{1}{n})$.

2. **Step 2:** Under the assumption that the $\mathbb{X} = (X_1, X_2, X_3 \ldots, X_n)$ generated from $F(\cdot \mid \theta)$ we compute : $\lambda(\theta) := E_\theta E^{\bar{\pi}}(g(\mathbb{X},\theta)|\mathbb{X})$ upto $o(\frac{1}{n})$.

3. **Step 3:** Compute $\int \lambda(\theta)\bar{\pi}(d\theta)$ when $\bar{\pi} \rightsquigarrow \delta_{\theta_0}(.)$. The final quantity after taking the weak limit leads to the required expansion of $E_{\theta_0}\left(g(\mathbb{X},\theta_{\theta_0})\right)$ upto $o(\frac{1}{n})$.

Proposition 4.7.1. *Note if one observes $T(\mathbb{X}) := E_{\bar{\pi}}(g(\mathbb{X},\theta) \mid \mathbb{X})$ for an integrable function $T(\mathbb{X})$ (with respect to P_θ for $\theta \in (\theta_0 - \delta, \theta_0 + \delta)$ for some $\delta > 0$) after **Step 1** of Shrinkage method, **Step 2** and **Step 3** virtually compute $E_{\theta_0}T(\mathbb{X})$. Since $\lambda(\theta) = E_\theta T(\mathbb{X})$, through Dominated Convergence Theorem and a consequence of the weak limit gives*

$$\lim_{\bar{\pi} \rightsquigarrow \delta_{\theta_0}(\cdot)} \int E_\theta\left[T(\mathbb{X})\right]\bar{\pi}(d\theta).$$

A good illustration on how Shrinkage method works is given at Chapter 1 of [23].

The conditions B_m in [10] ensures the existence of a set S which contains data $\mathbb{X}$ with probability P_θ, $(1 - o_p(n^{-1}))$ for $\theta \in$ a compact set K. For ensuring second order term we need to just assume B_2 for the auxiliary prior $\bar{\pi}(\cdot)$ containing true θ_0 in interior. All the following calculation of the Shrinkage method is a consequence of those assumptions in B_2. We will complete the remaining by implementing Shrinkage method which is more or less regular:

1. **Step 1:** We will construct a prior $\bar{\pi}$ with aforementioned smoothness properties and with a compact support with θ_0 being an interior point. Now define the following quantities

$$\bar{G}_1 := \bar{A}_1 + 3\bar{A}_3, \quad \bar{G}_2 = \bar{A}_2 + 6\bar{A}_4 + 45\bar{A}_6, \qquad \bar{G}_3 := \bar{A}_3, \quad \bar{G}_4 = \bar{A}_4 + 15\bar{A}_6,$$
$$\bar{G}_6 \;=\; \bar{A}_6$$

where

$$\bar{A}_1 := c^{-\frac{1}{2}}\frac{\bar{\pi}'(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}, \quad \bar{A}_2 := \frac{1}{2}c^{-1}\frac{\bar{\pi}''(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}, \qquad \bar{A}_3 := A_3, \quad \bar{A}_4 := \bar{A}_1\bar{A}_3 + \frac{1}{24}c^{-2}a_4,$$
$$\bar{A}_6 \;:=\; A_6.$$

By proceeding similarly like Lemma 19 or from [42], one gets a similar posterior expansion of $\bar{\pi}(y\big|\mathbb{X}) := \bar{\pi}(\hat{\theta}_n + \frac{y}{\sqrt{nc}})e^{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]}$ like following display,

$$\int_{\mathbb{R}}\left|\bar{\pi}(\hat{\theta}_n + \frac{y}{\sqrt{nc}})e^{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]} \;-\; \bar{\pi}(\hat{\theta}_n)e^{-\frac{y^2}{2}}\left(1 + \frac{\bar{\pi}'(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}\frac{y}{\sqrt{nc}} + \frac{1}{n}\frac{\bar{\pi}''(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}\frac{y^2}{2c}\right)\right.$$
$$\left.\left(1 + R_n(\hat{\theta}_n) + \frac{R_n(\hat{\theta}_n)^2}{2}\right)\right|dy = o_{P_{\theta_0}}\left(\frac{1}{n}\right) \quad .$$

So, $$f^{\bar{\pi}}(\theta \mid \mathbb{X}) = \frac{\bar{\pi}(\theta)e^{n[L_n(\theta)-L_n(\hat{\theta}_n)]}}{\int \bar{\pi}(\theta)e^{n[L_n(\theta)-L_n(\hat{\theta}_n)]}d\theta} = \frac{\bar{\pi}(\hat{\theta}_n + \frac{y}{\sqrt{nc}})e^{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]}}{\int \bar{\pi}(\hat{\theta}_n + \frac{y}{\sqrt{nc}}))e^{n[L_n(\hat{\theta}_n+\frac{y}{\sqrt{nc}})-L_n(\hat{\theta}_n)]}dy}.$$

Using the above expression of "auxiliary" posterior density of $y = \sqrt{nc}(\theta - \hat{\theta}_n)$,

$$f^{\bar{\pi}}(y \mid \mathbb{X}) = \phi(y)\left(1 + \frac{1}{\sqrt{n}}\left(\bar{G}_1 H_1(y) + \bar{G}_3 H_3(y)\right) + \frac{1}{n}\left(\bar{G}_2 H_2(y) + \bar{G}_4 H_4(y) + \bar{G}_6 H_6(y)\right)\right) + o_{p_{\theta_0}}(n^{-1}). \quad (4.7.17)$$

Using the expansion one can write $P^{\bar{\pi}}\left[\theta \le \theta^{(1-\alpha)}(\mathbb{X}, \mathcal{G})\middle|\mathbb{X}\right]$ as

$$\begin{aligned}
& P\left[\theta \le \hat{\theta} + (nc)^{-\frac{1}{2}}\left\{z + n^{-\frac{1}{2}}\beta_1 + n^{-1}\beta_2\right\}\middle|\mathbb{X}\right] = P^{\bar{\pi}}\left[y \le z + n^{-\frac{1}{2}}\beta_1 + n^{-1}\beta_2\middle|\mathbb{X}\right] \\
= & \int_{-\infty}^{z+n^{-\frac{1}{2}}\beta_1+n^{-1}\beta_2} \phi(y)\left[1 + \frac{1}{\sqrt{n}}\left(\bar{G}_1 H_1(y) + \bar{G}_3 H_3(y)\right)\right] dy \\
+ & \frac{1}{n}\int_{-\infty}^{z+n^{-\frac{1}{2}}\beta_1+n^{-1}\beta_2} \phi(y)\left[\left(\bar{G}_2 H_2(y) + \bar{G}_4 H_4(y) + \bar{G}_6 H_6(y)\right)\right] dy + o_p(n^{-1})
\end{aligned}$$

Working similarly like (4.7.15)-(4.7.16) we have

$$\begin{aligned}
& P^{\bar{\pi}}\left[\theta \le \theta^{(1-\alpha)}(F, \mathbb{X}, n)\middle|\mathbb{X}\right] \\
= & \Phi(z) + n^{-\frac{1}{2}}\phi(z)\left\{\beta_1 - \bar{G}_1 - \bar{G}_3 H_2(z)\right\} + n^{-1}\phi(z)\Big[\beta_2 - 2z\beta_1\bar{G}_3 \\
& -\frac{1}{2}\beta_1^2 z + \beta_1 z\left\{\bar{G}_1 + \bar{G}_3 H_2(z)\right\} - \bar{G}_2 H_1(z) - \bar{G}_4 H_3(z) - \bar{G}_6 H_5(z)\Big] + o_p(n^{-1}) \\
= & 1 - \alpha + n^{-\frac{1}{2}}\phi(z)\left\{G_1 - \bar{G}_1\right\} + n^{-1}\phi(z)\Big\{\beta_1 z\left[\bar{G}_1 - G_1\right] + \left[G_2 - \bar{G}_2\right] H_1(z) \\
& + \left[G_4 - \bar{G}_4\right] H_3(z) + \frac{W_n^{(1)}(\mathbb{X})}{\sqrt{c}}\Big\} + o_p(n^{-1}). \quad (4.7.18)
\end{aligned}$$

After inputing the values we have

$$
\begin{aligned}
& P^{\bar{\pi}}\left[\theta \leq \theta^{(1-\alpha)}(F,\mathbb{X},n)\middle|\mathbb{X}\right] \\
= \; & 1-\alpha+n^{-\frac{1}{2}}\phi(z)c^{-\frac{1}{2}}\left\{\frac{J'(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}-\frac{\bar{\pi}'(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}+a_n^{(2)}(\theta,\hat{\theta}_n)\right\} \\
& +n^{-1}\phi(z)z\Bigg\{\left(c^{-1}\frac{J'(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}-\frac{1}{6}c^{-2}l'''(\hat{\theta}_n)\right)\left(\frac{\bar{\pi}'(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}-\frac{J'(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}\right) \\
& +\frac{1}{2}c^{-1}\left[\frac{J''(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}-\frac{\bar{\pi}''(\hat{\theta}_n)}{\bar{\pi}(\hat{\theta}_n)}\right]\Bigg\}+n^{-1}\phi(z)\frac{a_n^1(\mathbb{X},\theta)}{\sqrt{c}}+o_{p_{\theta_0}}(n^{-1}). \quad (4.7.19)
\end{aligned}
$$

where (4.7.19) is obtained after a number of simplifications (putting values of the data dependent constants) and using following decomposition of $\mathbb{W}_n^{(1)}(\mathbb{X})$ in (4.7.18).

$$
\begin{aligned}
\mathbb{W}_n^{(1)}(\mathbb{X}) \;&:=\; \sqrt{n}\left(\frac{J_n^{(1)}(\mathbb{X},\hat{\theta}_n)}{J_n(\mathbb{X},\hat{\theta}_n)}-\frac{J^{(1)}(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}\right) \\
&=\; \sqrt{n}\left(\frac{J_n^{(1)}(\mathbb{X},\hat{\theta}_n)}{J_n(\mathbb{X},\hat{\theta}_n)}-\frac{J^{(1)}(\theta,\hat{\theta}_n)}{J(\theta,\hat{\theta}_n)}\right)+\sqrt{n}\left(\frac{J^{(1)}(\theta,\hat{\theta}_n)}{J(\theta,\hat{\theta}_n)}-\frac{J^{(1)}(\theta_0,\hat{\theta}_n)}{J(\theta_0,\hat{\theta}_n)}\right) \\
&=:\; a_n^1(\mathbb{X},\theta)+\sqrt{n}a_n^{(2)}(\theta,\hat{\theta}_n) \quad \text{(defining the first and second term)} \quad (4.7.20)
\end{aligned}
$$

where each of these terms will be analyzed in next step.

2. **Step 2:** We will now compute the asymptotic value of $\lambda(\theta)=E_\theta P^{\bar{\pi}}(\theta\leq\theta^{(1-\alpha)}(\mathbb{X},\mathcal{G})|X)$. Note that asymptotically $a_n^{(2)}(\theta,\hat{\theta}_n)$ converges to $a^{(2)}(\theta):=\frac{J^{(1)}(\theta,\theta)}{J(\theta,\theta)}-\frac{J^{(1)}(\theta_0,\theta)}{J(\theta_0,\theta)}$, under true θ which becomes $a^{(2)}(\theta_0)=0$ when $\theta=\theta_0$. We will treat $a_n^1(\mathbb{X},\theta)$ by expanding that term. We take the facility of choosing auxiliary $\bar{\pi}(\cdot)$ in a way such that the expression (4.7.19) holds for all data points in a compact set $\bar{S}$ in $\mathbb{R}$ that has P_θ of order $(1-o(n^{-1}))$ uniformly for all $\theta\in K$, where K is the compact domain of $\bar{\pi}$. Under the

assumption that the limits exist and the existence of the set $S \times K$ is ensured by the condition B_m satisfied by $\bar{\pi}$ for $m = 2$ in [10]. From (4.7.19)

$$\begin{aligned}
&\lambda(\theta) := E_\theta\left\{P^{\bar{\pi}}\left[\theta \le \theta^{(1-\alpha)}(\mathbb{X},\mathcal{G})\Big|\mathbb{X}\right]\right\} = 1-\alpha+n^{-\frac{1}{2}}\phi(z)I_\theta^{-\frac{1}{2}}\left\{\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)}\right.\\
-&\left.\frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)}+a^{(2)}(\theta)\right\}+n^{-1}\phi(z)z\left\{\left(I_\theta^{-1}\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)}-\frac{1}{6}I_\theta^{-2}M_\theta\right)\left(\frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)}\right.\right.\\
-&\left.\left.\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)}\right)+\frac{1}{2}I_\theta^{-1}\left[\frac{J''(\theta_0,\theta)}{J(\theta_0,\theta)}-\frac{\bar{\pi}''(\theta)}{\bar{\pi}(\theta)}\right]\right\}+n^{-1}\phi(z)\frac{E_\theta\left[a_n^1(\mathbb{X},\theta)\right]}{\sqrt{I_\theta}}\\
&+o(n^{-1}) \qquad (4.7.21)
\end{aligned}$$

where $M_\theta := El^{(3)}(\theta|\mathbb{X})$. By Assumption $a_i(\cdot)$ is a continuous function, so in a compact domain($\bar{\pi}$) containing θ_0 it will always exist. It is a consequence of the Corollary 4.7.1 but we need to show the integrability of $E_\theta\left[a_n^1(\mathbb{X},\theta)\right]$ in $\theta \in$ domain($\bar{\pi}$). In the following we will give an expansion of $E_\theta\left[a_n^1(\mathbb{X},\theta)\right]$ in terms of $a_i(\theta_0)$, $J(\theta_0,\theta)$, $J'(\theta_0,\theta)$. Now by Assumption 17-19 we have $a_i(\cdot)$ continuous function in $\Theta = \mathbb{R}$. Since also $J(\cdot,\theta)$ won't vanish is $\theta \in (\theta_0-\delta,\theta+\delta)$ one can always find a compact neighborhood of θ_0 where the quantity $n \to \infty \quad E_\theta\left[a_n^1(\mathbb{X},\theta)\right]$ will remain bounded. We will take that compact neighborhood as the domain($\bar{\pi}$).

Now we will prove the higher order expansion of the quantity $E_\theta\left[a_n^1(\mathbb{X},\theta)\right]$. Note that by Taylor's expansion on the function $f(x,y) = \frac{x}{y}$ at the point (T_1,T_2) around (g_1,g_2), we have for some $(g_1^*, g_2^*) \in (T_1, g_1) \times (T_2, g_2)$

$$\frac{T_1}{T_2} = \frac{g_{1,\theta}}{g_{2,\theta}} + (T_1 - g_{1,\theta})\frac{1}{g_{2,\theta}} - (T_2 - g_{2,\theta})\frac{g_{1,\theta}}{g_{2,\theta}^2} + (T_2 - g_{2,\theta})^2\frac{g_{1,\theta}^*}{g_{2,\theta}^{*3}}$$
$$+(T_1 - g_{1,\theta})(T_2 - g_{2,\theta})\frac{1}{g_{*2,\theta}}.$$

That yields $\sqrt{n}[\frac{T_1}{T_2} - \frac{g_{1,\theta}}{g_{2,\theta}}] = \sqrt{n}(T_1 - g_{1,\theta})\frac{1}{g_{2,\theta}} - \sqrt{n}(T_2 - g_{2,\theta})\frac{g_{1,\theta}}{g_{2,\theta}^2} + \sqrt{n}(T_2 - g_{2,\theta})[(T_2 - g_{2,\theta})\frac{g_{1,\theta}^*}{g_{2,\theta}^{*3}} - (T_1 - g_{1,\theta})\frac{1}{g_{*2}}]$. Now choosing $T_1 := J_n'(\mathbb{X}, \hat{\theta}_n), T_2 := J_n(\mathbb{X}, \hat{\theta}_n)$ and their corresponding limits $g_{1,\theta} := J'(\theta, \theta), g_{2,\theta} := J(\theta, \theta)$ one gets $a_n^1(\mathbb{X}, \theta)$ is $O_{P_\theta}(1)$. Last statement is a consequence of Slutsky's theorem and the Assumption 19. Note that $\frac{g_{1,\theta}^*}{g_{2,\theta}^{*3}}, \frac{1}{g_{*2,\theta}}$ are $O_{P_{\theta_0}}(1)$ which follows from the fact $g_1^* \xrightarrow{P_{\theta_0}} g_{1,\theta}$, $g_{2,\theta}^* \xrightarrow{P_\theta} g_{2,\theta}$, and then using continuity theorem one has $\frac{g_{1,\theta}^*}{g_{2,\theta}^{*3}} = \frac{g_{1,\theta}}{g_{2,\theta}^3} + o_{P_\theta}(1)$, $\frac{1}{g_{*2,\theta}} = \frac{1}{g_{2,\theta}} + o_{P_\theta}(1)$ since $g_2 \neq 0$. From Slutsky's theorem the residual term $\sqrt{n}(T_2 - g_2)[(T_2 - g_2)\frac{g_{1,\theta}^*}{g_{2,\theta}^{*3}} - (T_1 - g_{1,\theta})\frac{1}{g_{*2,\theta}}]$ will be $o_{p_\theta}(1)$. Our conclusion that $E_\theta\left[a_n^1(\mathbb{X}, \theta)\right] = \frac{a_1(\theta)}{g_{2,\theta}} - \frac{a_2(\theta)g_1}{g_{2,\theta}^2} + o(1)$, will follow if we provide an additional detail on the expected residual term:

$$E_\theta\left[\sqrt{n}(T_2 - g_2)\left[(T_2 - g_2)\frac{g_1^*}{g_2^{*3}} - (T_1 - g_1)\frac{1}{g_{*2}}\right]\right] \to 0 \tag{4.7.22}$$

for all $\theta \in \text{domain}(\bar{\pi})$. Note that

$$(T_1 - g_1) = (J_n'(\mathbb{X}, \hat{\theta}_n) - J'(\theta, \hat{\theta}_n)) + ((J'(\theta, \hat{\theta}_n) - J'(\theta, \theta)) \tag{4.7.23}$$

$$(T_2 - g_2) = (J_n(\mathbb{X}, \hat{\theta}_n) - J(\theta, \hat{\theta}_n)) + ((J(\theta, \hat{\theta}_n) - J(\theta, \theta)). \tag{4.7.24}$$

Second term after scaling with $\sqrt{n}$, along with the lipschitz property of $J(\theta, \cdot)$ from Assumption 19(c) will give finiteness of the quantity $nE_\theta\left[(J(\theta,\hat{\theta}_n) - J(\theta,\theta)\right]^2 < \infty$ (from the asymptotic expansion of MLE $\hat{\theta}_n$ under true value θ). Along with that and Assumption 19(a) one gets $nE_\theta\left[T_1 - g_1\right]^2 < \infty$ and $nE_\theta\left[T_2 - g_2\right]^2 < \infty$ will follow similarly. Note that these results imply that the set $\{(T_1,T_2) : |T_1 - g_1| < \epsilon, |T_2 - g_2| < \epsilon\}$ denoted by $A_{n,\epsilon}$ has probability P_θ of order $(1 - O_{P_\theta}(\frac{1}{n}))$. Since g_2 is away from 0, fixing $\epsilon \in (0, g_2)$ we can work with $\frac{g_1^*}{g_2^{*3}}.1_{A_{n,\epsilon}}, \frac{1}{g_{*2}}.1_{A_{n,\epsilon}}$ in place of $\frac{g_1^*}{g_2^{*3}}, \frac{1}{g_{*2}}$ in the expansion of $a_n^1(\mathbb{X},\theta)$ in (4.7.19) for $\theta = \theta_0$, since the residual term (that is non zero with probability $o_{P_{\theta_0}}(\frac{1}{\sqrt{n}})$ will be accumulated in the $o_{P_{\theta_0}}(\frac{1}{\sqrt{n}})$ term. Now note $\left|\frac{g_1^*}{g_2^{*3}}.1_{A_{n,\epsilon}}\right| < \frac{g_1+\epsilon}{(g_2-\epsilon)^3}, \left|\frac{1}{g_{*2}}.1_{A_{n,\epsilon}}\right| < \frac{1}{g_2-\epsilon}$ a.s. Using these along with $nE_\theta\left[T_1 - g_1\right]^2 < \infty, \quad nE_\theta\left[T_2 - g_2\right]^2 < \infty$, (4.7.22) will follow by breaking and analyzing each of the two terms.

3. **Step 3:** The last step comes from computing $\int \lambda(\theta)\bar{\pi}(d\theta)$ when $\bar{\pi}(\theta) \to \delta_{\theta_0}(\theta)$. Note that if $\alpha(\theta) := \frac{1}{6}I_\theta^{-2}M_\theta - I_\theta^{-1}\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)}$, it follows from (4.7.21) that

$$\begin{aligned}\lambda(\theta) \quad = \quad & 1 - \alpha + n^{-\frac{1}{2}}\phi(z)I_\theta^{-\frac{1}{2}}\left\{\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)} - \frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)} + a^{(2)}(\theta)\right\} \\ & + n^{-1}\phi(z)z\left\{\alpha(\theta)\left(\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)} - \frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)}\right) + \frac{1}{2}I_\theta^{-1}\left[\frac{J''(\theta_0,\theta)}{J(\theta_0,\theta)} - \frac{\bar{\pi}''(\theta)}{\bar{\pi}(\theta)}\right]\right\} \\ & + n^{-1}\phi(z)\frac{E_\theta\left[a_n^1(\mathbb{X},\theta)\right]}{\sqrt{I_\theta}} + o(n^{-1}). \end{aligned} \tag{4.7.25}$$

From properties of distribution theory one has if $\bar{\pi}(\theta) \to \delta_{\theta_0}(\theta)$, then

$$\int f(\theta).\bar{\pi}(d\theta) \to f(\theta_0), \qquad \int f(\theta).\frac{\bar{\pi}^{(m)}(d\theta)}{\bar{\pi}(\theta)} \to (-1)^m f^{(m)}(\theta_0)$$

where for the second result f is an m-times differentiable at a neighborhood of $\theta = \theta_0$. Note that $a^{(2)}(\theta_0) = 0$. So after taking the weak limit of $\int \lambda(\theta)\bar{\pi}(\theta)$ as $\bar{\pi}(\theta) \to \delta_{\theta_0}(\theta)$, one has

$$P_{\theta_0}\left[\theta_0 \leq \theta^{(1-\alpha)}(\mathbb{X}, \mathcal{G})\right] \;=\; 1 - \alpha + \frac{\phi(z_\alpha)\Delta_1(\theta_0)}{n^{\frac{1}{2}}} + \frac{\phi(z_\alpha)z_\alpha\Delta_2(\theta_0)}{n} + o(n^{-1})$$

where

$$\begin{aligned}
\Delta_1(\theta_0) &= I_{\theta_0}^{-\frac{1}{2}} \frac{\frac{\partial}{\partial\theta}J(\theta_0,\theta)}{J(\theta_0,\theta_0)}\bigg|_{\theta_0} + \frac{\partial I_\theta^{-\frac{1}{2}}}{\partial\theta}\bigg|_{\theta_0}, \\
\Delta_2(\theta_0) &= \left\{\left(\alpha'(\theta) + \alpha(\theta)\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)}\right) + \frac{1}{2}I_\theta^{-1}\frac{J''(\theta_0,\theta)}{J(\theta_0,\theta)} - \frac{d^2}{d\theta^2}\left[\frac{1}{2}I_\theta^{-1}\right]\right\}\bigg|_{\theta_0} \\
&\quad + \frac{E_{\theta_0}\left[a_n^1(\mathbb{X},\theta_0)\right]}{z_\alpha\sqrt{I_{\theta_0}}} \qquad (4.7.26)
\end{aligned}$$

$E_{\theta_0}\left[a_n^1(\mathbb{X},\theta_0)\right] = \frac{a_1(\theta_0)}{g_2} - \frac{a_2(\theta_0)g_1}{g_2^2} + o(1)$ Note that First term of $\Delta_2(\theta_0)$ in (4.7.26)

$$\begin{aligned}
&\left(\alpha'(\theta) + \alpha(\theta)\frac{J'(\theta_0,\theta)}{J(\theta_0,\theta)}\right) + \frac{1}{2}I_\theta^{-1}\frac{J''(\theta_0,\theta)}{J(\theta_0,\theta)} - \frac{d^2}{d\theta^2}\left[\frac{1}{2}I_\theta^{-1}\right] \\
= \;& \frac{1}{J(\theta_0,\theta)}\left[\frac{d}{d\theta}\left[\alpha(\theta)J(\theta_{0,\theta})\right]\right] + \frac{1}{2}J(\theta_0,\theta)^{-1}\frac{d}{d\theta}\left\{I_\theta^{-1}J'(\theta_0,\theta) - J(\theta_0,\theta)\Big(\frac{d}{d\theta}I_\theta^{-1}\Big)\right\} \\
= \;& J(\theta_0,\theta)^{-1}\left[\frac{d}{d\theta}\left\{\alpha(\theta)J(\theta_0,\theta) + \frac{1}{2}I_\theta^{-1}J'(\theta_0,\theta) - \frac{1}{2}J(\theta_0,\theta)\frac{d}{d\theta}[I_\theta^{-1}]\right\}\right] \qquad (4.7.27)
\end{aligned}$$

Using definition of $\alpha(\theta)$ the R.H.S of (4.7.27) becomes

$$\begin{aligned}
& J(\theta_0,\theta)^{-1}\left[\frac{d}{d\theta}\left\{\frac{1}{6}I_\theta^{-2}M_\theta J(\theta_0,\theta)-\frac{1}{2}I_\theta^{-1}J'(\theta_0,\theta)-\frac{1}{2}J(\theta_0,\theta)\frac{d}{d\theta}[I_\theta^{-1}]\right\}\right] \\
= \ & J(\theta_0,\theta)^{-1}\left[\frac{d}{d\theta}\left\{\frac{1}{6}I_\theta^{-2}M_\theta J(\theta_0,\theta)-\frac{d}{d\theta}\left[\frac{1}{2}I_\theta^{-1}J(\theta_0,\theta)\right]\right\}\right] \qquad (4.7.28)
\end{aligned}$$

Combining two estimates from (4.7.26) and (4.7.28) with taking the limit at $\theta = \theta_0$ one gets the second order term and the conclusion of the theorem follows.

□

APPENDIXA

A.1 Auxiliary results

Lemma A.1.1. *(a) For a compact set K in $\mathbb{R}^d$ let $\mathbb{F}_{a,b}(K)$ be the space of functions $\psi : K \to \mathbb{R}$ such that $\sup_{x\in K}|\psi(x)| \le a$ and $|\psi(x)-\psi(y)| \le b|x-y|$ for all $x, y \in K$. Then for any $\epsilon > 0$ there is a finite subset $\mathbb{F}^\epsilon_{a,b}(K)$ of $\mathbb{F}_{a,b}(K)$ such that for any signed measure μ*

$$\sup_{\psi\in\mathbb{F}_{a,b}(K)} |\langle\psi,\mu\rangle| \le \max_{g\in\mathbb{F}^\epsilon_{a,b}(K)} |\langle g,\mu\rangle| + \epsilon|\mu|_{TV}.$$

(b) If $K = [-R,R]^d$ for some $R > 0$, then $\mathbb{F}^\epsilon_{R,1}(K)$ can be chosen such that

$$|\mathbb{F}^\epsilon_{R,1}(K)| \le \max\left\{\frac{2(2\sqrt{d}+1)}{3}\frac{R}{\epsilon}3^{[\frac{2R}{\epsilon}(\sqrt{d}+1)]^d}, 1\right\}.$$

The next lemma is straightforward.

Lemma A.1.2. *Let $P : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a transition probability kernel. Fix $N \ge 1$ and let $y_1, y_2, ..., y_N \in \mathbb{R}^d$. Let $X_1, X_2, ..., X_N$ be independent random variables such that $\mathcal{L}(X_i) = \delta_{y_i}P$. Let $f \in BM(\mathbb{R}^d)$ and let $m^N_0 = \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$, $m^N_1 = \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$. Then*

$$E|\langle\psi, m^N_1 - m^N_0 P\rangle| \le \frac{2\|\psi\|_\infty}{\sqrt{N}}.$$

The following is a discrete version of Gronwall's lemma.

Lemma A.1.3. *Let* $\{a_i\}_{i=0}^{\infty}, \{b_i\}_{i=0}^{\infty}, \{c_i\}_{i=0}^{\infty}$ *be non-negative sequences. Suppose that*

$$a_n \leq b_n + \sum_{k=0}^{n-1} c_k a_k \text{ for all } n \geq 0.$$

Then

$$a_n \leq b_n + \sum_{k=0}^{n-1} \left[c_k b_k \left(\prod_{j=k+1}^{n-1} (1 + c_j) \right) \right] \text{ for all } n \geq 0.$$

Lemma A.1.4. *For any* $a, b > 0$ *and* $\{C_i\}_{i \geq 0}$ *be a nonnegative sequence of elements, then for all* $n \geq 0$

$$\sum_{k=0}^{n} a^{n-k} \sum_{i=0}^{k} b^{k-i} C_i = \sum_{i=0}^{n} C_i \left[\frac{a^{n+1-i} - b^{n+1-i}}{a - b} \right].$$

REFERENCES

[1] N. Antunes, C. Fricker, P. Robert and D. Tibi. Stochastic networks with multiple stable points. *Annals of Probability*, 36 (2008), no. 1, 255–278.

[2] Michele Ballerini, Nicola Cabibbo, Raphael Candelier, Andrea Cavagna, Evaristo Cisbani, Irene Giardina, Alberto Orlandi, Giorgio Parisi, Andrea Procaccini, Massimiliano Viale.

[3] MJ Bayarri, JO Berger, A Forte, G García-Donato, et al. Criteria for bayesian model choice with application to variable selection. *The Annals of statistics*, 40(3):1550–1577, 2012.

[4] James O Berger and José M Bernardo. On the development of reference priors. *Bayesian statistics*, 4(4):35–60, 1992.

[5] James O Berger, José M Bernardo, and Dongchu Sun. The formal definition of reference priors. *Annals of Statistics*, pages 905–938, 2009.

[6] James O Berger, Jose M Bernardo, and Dongchu Sun. Objective priors for discrete parameter spaces. *Journal of the American Statistical Association*, 107(498):636–648, 2012.

[7] James O Berger and Dongchu Sun. Objective priors for the bivariate normal model. *Annals of Statistics*, pages 963–982, 2008.

[8] Dario Benedetto, Emanuele Caglioti, and Mario Pulvirenti. A kinetic equation for granular media. *RAIRO-M2AN Modelisation Math et Analyse Numerique-Mathem Modell Numerical Analysis*, 31(5):615–642, 1997.

[9] Dario Benedetto, Emanuelle Caglioti, José A Carrillo, and Mario Pulvirenti. A non-Maxwellian steady distribution for one-dimensional granular media. *Journal of Statistical Physics*, 91(5-6):979–990, 1998.

[10] Peter J Bickel and JK Ghosh. A decomposition for the likelihood ratio statistic and the bartlett correction–a bayesian argument. *Annals of Statistics*, pages 1070–1090, 1990.

[11] Emmanuel Boissard. Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance. *Electronic Journal of Probability*, 16:2296–2333, 2011.

[12] F. Bolley, A. Guillin and F. Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck Equation , *ESAIM Mathematical Modelling and Numerical Analysis*, 44:867–884, 2010.

[13] François Bolley and Cédric Villani. Weighted Csiszar-Kullback-Pinsker inequalities and applications to transportation inequalities. To appear in *Fac. des Sciences de Toulouse*, 2005.

[14] François Bolley, Arnaud Guillin, and Cédric Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2007.

[15] Amarjit Budhiraja, Pierre Del Moral, Sylvain Rubenthaler. Discrete time Markovian agents interacting through a potential. *ESAIM: Probability and Statistics*, PS 17:614–634, (2013)

[16] Vincent Calvez, Lucilla Corrias, et al. The parabolic-parabolic keller-segel model in $\mathbb{R}^2$. *Communications in Mathematical Sciences*, 6(2):417–447, 2008.

[17] R. Carmona and F. Delarue. Probabilistic Analysis of Mean-Field Games. *math arxiv:1210.5780.*

[18] François Caron, Pierre Del Moral, Arnaud Doucet, Michele Pace. Particle approximations of a class of branching distribution flows arising in multi-target tracking. *SIAM Journal on Control and Optimization*, 49(4):1766–1792, 2011.

[19] José A Carrillo, Robert J McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana*, 19(3):971–1018, 2003.

[20] José A Carrillo, Robert J McCann, and Cédric Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, 179(2):217–263, 2006.

[21] P. Cattiaux, A. Guillin, F. Malrieu. Probabilistic approach for granular media equations in the non uniformly convex case. *Probability Theory and Related Fields*, 140, 19–40, 2008.

[22] Giacomo Como and Fabio Fagnani. Scaling limits for continuous opinion dynamics systems. *Annals of Applied Probability*,21(4): 1537 – 1567, 2011.

[23] Gauri S Datta and Rahul Mukerjee. *Probability matching priors: higher order asymptotics*, volume 178. Springer, 2004.

[24] Steffen Dereich, Michael Scheutzow, Reik Schottstedt, et al. Constructive quantization: approximation by empirical measures. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49(4):1183–1203, 2013.

[25] P. Del Moral and A. Guionnet. On the stability of Measure Valued Processes with

Applications to filtering. *C.R. Acad. Sci. Paris,* t. 329, Serie I, pp. 429-434 (1999).

[26] P. Del Moral and L. Miclo. Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering. *Séminaire de Probabilités XXXIV*, Ed. J. Azéma and M. Emery and M. Ledoux and M. Yor. Lecture Notes in Mathematics, Springer-Verlag Berlin, Vol. 1729 (2000).

[27] Pierre Del Moral and Emmanuel Rio. Concentration inequalities for mean field particle models. *Annals of Applied Probability*, 21(3):1017–1052, 2011.

[28] Arthur P Dempster. The dempster–shafer calculus for statisticians. *International Journal of Approximate Reasoning*, 48(2):365–377, 2008.

[29] Hacene Djellout, Arnaud Guillin, and Liming Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Annals of Probability*, 32(3B):2702–2732, 2004.

[30] Paul T Edlefsen, Chuanhai Liu, and Arthur P Dempster. Estimating limits from poisson counting data using dempster-shafer analysis. *Annals of Applied Statistics*, pages 764–790, 2009.

[31] Ronald A Fisher. The fiducial argument in statistical inference. *Annals of Eugenics*, 6(4):391–398, 1935.

[32] Ronald Aylmer Fisher. The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London. Series A*, 139(838):343–348, 1933.

[33] Nicolas Fournier and Arnaud Guillin. On the Rate of Convergence in Wasserstein Distance of the Empirical Measure. Accepted in *Probability Theory and Related Fields*.

[34] Ailana M Fraser, Donald AS Fraser, Ana-Maria Staicu, et al. Second order ancillary: A differential view from continuity. *Bernoulli*, 16(4):1208–1223, 2010.

[35] DAS Fraser et al. Ancillaries and conditional inference. *Statistical Science*, 19(2):333–369, 2004.

[36] DAS Fraser et al. Is bayes posterior just quick and dirty confidence? *Statistical Science*, 26(3):299–316, 2011.

[37] DAS Fraser and A Naderi. Exponential models: Approximations for probabilities. *Biometrika*, 94(2):1–9, 2008.

[38] DAS Fraser, N Reid, E Marras, and GY Yi. Default priors for bayesian and frequentist inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodol-*

ogy), 72(5):631–654, 2010.

[39] DAS Fraser, N Reid, and A Wong. What a model with data says about theta. *Internat. J. Statist. Sci*, 3:163–178, 2005.

[40] Avner Friedman, Chiu-Yen Kao, and Chih-Wen Shih. Asymptotic phases in a cell differentiation model. *Journal of Differential Equations*, 247(3):736–769, 2009.

[41] Herbert Gajewski, Klaus Zacharias, and Konrad Gröger. Global behaviour of a reaction-diffusion system modelling chemotaxis. *Mathematische Nachrichten*, 195(1):77–114, 1998.

[42] Jayanta K Ghosh. Higher order asymptotics. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, pages i–111. JSTOR, 1994.

[43] Subhashis Ghosal. Probability matching priors for non-regular cases. *Biometrika*, 86(4):956–964, 1999.

[44] R. J. Gibbens, P. J. Hunt and F. P. Kelly. Bistability in communication networks. *Disorder in physical systems*, 113127, Oxford Sci. Publ., Oxford Univ. Press, New York, 1990.

[45] K. Giesecke, K. Spiliopoulos and R. B. Sowers. Default clustering in large portfolios: Typical events. *Ann. App. Prob., to appear*, 2012.

[46] Javier Gómez-Serrano, Carl Graham, and Jean-Yves Le Boudec. The bounded confidence model of opinion dynamics. *Mathematical Models and Methods in Applied Sciences*, 22(02), 2012.

[47] Nathael Gozlan and Christian Léonard. A large deviation approach to some transportation cost inequalities. *Probability Theory and Related Fields*, 139(1-2):235–283, 2007.

[48] Nathael Gozlan and Christian Léonard. Transport inequalities. a survey. *arXiv preprint arXiv:1003.3852*, 2010.

[49] Carl Graham and Philippe Robert. Interacting multi-class transmissions in large stochastic networks. *Annals of Applied Probability*, 19(6):2334–2361, 2009.

[50] O. Guéant, J-M. Lasry, P-L. Lions. Mean field games and applications. *Paris-Princeton Lectures on Mathematical Finance*, 2010, 205266, Lecture Notes in Math., 2003, Springer, Berlin, 2011.

[51] Jan Hannig. On generalized fiducial inference. *Statistica Sinica*, 19(2):491, 2009.

[52] Jan Hannig. Generalized fiducial inference via discretization. *Statistical Sinica*,

23(2):489–514, 2013.

[53] Jan Hannig, Hari Iyer, and Paul Patterson. Fiducial generalized confidence intervals. *Journal of the American Statistical Association*, 101(473):254–269, 2006.

[54] Robert E Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435):1343–1370, 1996.

[55] V. N. Kolokoltsov. *Nonlinear Markov Processes and Kinetic Equations.* Cambridge Universty Press, 2010.

[56] Bibb Latané and Andrzej Nowak. Self-organizing social systems: Necessary and sufficient conditions for the emergence of clustering, consolidation, and continuing diversity. *Progress in communication sciences*, pages 43–74, 1997.

[57] J-M. Lasry and P-L. Lions. *Mean field games*. Jpn. J. Math. 2 (2007), no. 1, 229–260.

[58] Florent Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Annals of Applied Probability*, 13(2):540–560, 2003.

[59] F.Malrieu and D.Talay Concentration Inequalities for Euler Schemes. *Monte Carlo and Quasi-Monte Carlo Methods*, 2004, H. Niederreiter and D. Talay (Eds.), Springer-Verlag (2006), 355–371.

[60] Florient Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE. *Stochastic processes and their applications*, 95(1):109–132, 2001.

[61] Ryan Martin and Chuanhai Liu. Inferential models: A framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association*, 108(501):301–313, 2013.

[62] Ryan Martin and Chuanhai Liu. Marginal inferential models: prior-free probabilistic inference on interest parameters. *arXiv preprint arXiv:1306.3092*, 2013.

[63] Ryan Martin and Chuanhai Liu. Conditional inferential models: combining information for prior-free probabilistic inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2014.

[64] Ryan Martin, Jianchun Zhang, and Chuanhai Liu. Dempster–shafer theory and statistical inference with weak beliefs. *Statistical Science*, pages 72–87, 2010.

[65] K Marton. Bounding d-Distance by Informational Divergence: A Method to Prove Measure Concentration. *Annals of Probability*, 24(2):857–866, 1996.

[66] K Marton. A simple proof of the blowing-up Lemma. *IEEE Transactions on Information Theory*, 32(3):445-446, 1986.

[67] Rahul Mukerjee et al. Data-dependent probability matching priors for empirical and related likelihoods. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, pages 60–70. Institute of Mathematical Statistics, 2008.

[68] SH Ong and Rahul Mukerjee. Data-dependent probability matching priors of the second order. *Statistics: A Journal of Theoretical and Applied Statistics*, 44(3):291–302, 2010.

[69] M. Pace, M. Birattari, and M. Dorigo. The swarm/potential model: Modeling robotics swarms with measure-valued recursions associated to random finite sets. *IEEE Transactions on Robotics*, page submitted, 2013.

[70] Frank Schweitzer. *Brownian agents and active particles: Collective dynamics in the natural and social sciences*. Springer, 2007.

[71] Derek L Sonderegger and Jan Hannig. Fiducial theory for free-knot splines. *Springer Proceedings in Mathematics & Statistics*, page 155, 2014.

[72] A. Stevens. Trail following and aggregation of Myxobacteria. *J. Biol. Syst.*, (3):1059–1068, 1995.

[73] Anne Randi Syversveen. Noninformative bayesian priors. interpretation and problems with construction and applications. *Preprint Statistics*

[74] Alain-Sol Sznitman. Topics in propagation of chaos, École d'Été de Probabilités de Saint-Flour XIX1989, 165251, Lecture Notes in Math., 1464, Springer, Berlin, 1991.

[75] Gunnar Taraldsen, Bo Henry Lindqvist, et al. Fiducial theory and optimal inference. *Annals of Statistics*, 41(1):323–341, 2013.

[76] Kam-Wah Tsui and Samaradasa Weerahandi. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84(406):602–607, 1989.

[77] VS Varadarajan. Weak convergence of measures on separable metric spaces. *Sankhya: The Indian Journal of Statistics*,15–22, 1958

[78] Piero Veronese and Eugenio Melilli. Fiducial and confidence distributions for real exponential families. *Scandinavian Journal of Statistics*, 2014.

[79] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2003.

[80] Samaradasa Weerahandi. Generalized confidence intervals. In *Exact Statistical Meth-*

ods for Data Analysis, pages 143–168. Springer, 1995.

[81] Min-ge Xie and Kesar Singh. Confidence distribution, the frequentist distribution estimator of a parameter: a review. *International Statistical Review*, 81(1):3–39, 2013.

[82] In-Kwon Yeo and Richard A Johnson. A uniform strong law of large numbers for u statistics with application to transforming to near symmetry. *Statistics & probability letters*, 51(1):63–69, 2001.

[83] Jianchun Zhang and Chuanhai Liu. Dempster-shafer inference with weak beliefs. *Statistica Sinica*, 21(2):475, 2011.