Barycentric Markov processes and stability of stochastic integrators

Kennerberg, Philip

2020

Document Version:
Version created as part of publication process; publisher's layout; not normally made publicly available

Citation for published version (APA):
Kennerberg, P. (2020). *Barycentric Markov processes and stability of stochastic integrators*. [Doctoral Thesis (compilation), Centre for Mathematical Sciences]. Media-Tryck, Lund University, Sweden.

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Barycentric Markov processes and stability of stochastic integrators

av

Philip Kennerberg

Lund University

Avhandling som för avläggande av filosofie doktorsexamen vid
naturvetenskapliga fakulteten
vid Lunds universitet kommer att offentligen försvaras
a Matematikcentrum, Sölvegatan 18, sal MH:Hörmander-salen (MH:C),
fredagen den 18 december 2020 kl. 9^{15}.

Organization
LUND UNIVERSITY

Centre for Mathematical Sciences
Mathematical Statistics
Faculty of Science

Document name
DOCTORAL DISSERTATION

Date of issue
2020-11-09

ISRN:
LUNFMS-1028-2020

Sponsoring organization

Author
Philip Kennerberg

Title and subtitle
Barycentric Markov processes and stability of stochastic integrators

This thesis consists of four papers that broadly concerns two different topics. The first topic is so called barycentric Markov processes. By a barycentric Markov process we mean a process that consists of a point/particle system evolving in (discrete) time, whose evolution depends in some way on mean value of the current points in the system. In common for the three first papers which are on this topic is that we study how all the points of the so called core (a certain subset of points in the system) of the system converges to the same point.

The first article concerns how an N-point system behaves when we reject the K<N/2 points that minimize the sample variance of the remaining N-K points (the core). We then replace the rejected points with K new points which follow some fixed distribution and which are all independent from the past points. When K=1 this is equivalent to rejecting the point which is furthest from the center of mass. We prove that under rather weak assumptions on the sampling distribution, the points of the core converge to the same point as well as that regardless of any assumptions on our sampling distribution, the sampling variance of the core converges to zero or the core "drifts off to infinity".

The second article concerns a similar problem as the first one. We once again consider an N-point system but at each time step we reject the point furthest from the center of mass multiplied by a positive number p and replace it with a point from a fixed distribution with full support on [0,1], which is independent from all past points. If p=1 we obtain a special case of the previous article. If p\not=1 it turns out that this process behaves very differently from the process in the first article, the stationary distribution to which the core points converge turns out to be a Bernoulli distribution.

The third article studies yet another N-point system but now on a discrete circle. During each time step we compute the distances for each point from the mean of its two neighbors and reject the one with largest such distance (thereby obtaining our core) and replace it with a new point independent from past points. Two different cases are considered, the first is with uniformly distributed points in [0,1] and the other is with a discrete uniform distribution (i.e. uniformly distributed on an equally spaced grid).

The fourth and last article is on the topic of stochastic calculus. The main objective is to study "stability" of integrators for stochastic integrals. We examine how converging sequences of processes in the role of integrators retain their convergence properties for their corresponding integrals when the integrators are transformed under certain classes of functions. The convergence is on one hand in the uniform (over compact time intervals) in L^p-sense and on the other hand in the UCP-sense (uniform convergence in probability on compact time intervals). We examine processes with quadratic variations (along some refining sequence) transformed by absolutely continuous functions as well as Dirichlet processes transformed by C^1 functions.

Key words
Dirichlet processes, Markov processes, Geometric probability, Stochastic calculus.

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISBN
978-91-7895-702-6

Recipient's notes

Number of pages
188

Price

Security classification

Distribution by (name and adress)

signature ____________________ Date 2020-11-09

Barycentric Markov processes and stability of stochastic integrators

Philip Kennerberg

LUND UNIVERSITY

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118
SE-221 00 Lund
Sweden
http://www.maths.lth.se/

Doctoral Theses in Mathematical Sciences 2020:4
ISSN 1404-0034

ISBN 978-91-7895-702-6 (print)
ISBN 978-91-7895-703-3 (pdf)
LUNFMS-1028-2020

Printed in Sweden by Media Tryck, Lund 2020

Contents

Acknowledgement

First I would like to express my gratitude towards my supervisor Professor Stanislav Volkov, who gave me this opportunity, for this I will always be grateful. My deepest thanks for your time and guidance throughout this process. I would also like to thank my co-supervisor Associate Professor Magnus Wiktorsson, for all of your help and time. Thanks also to my second co-supervisor Prof. Tatyana Turova for inspiration and fruitful discussions. Additionally, thank you to my friend Maria Juhlin with whom I got to discover the jaw dropping city of Tokyo, and with whom I have had many fun chats with, about all sorts of nonsense. My old childhood friend Terry, whom I have known since the age of eight, I have always felt your support. My sincere thank you to the always helpful and jovial James, Susann and Denise. I want to thank Carl, Rachele, Unn and Vladimir for all the good times. Thank you Dmitrii Silvestrov for encouraging me to pursue a PhD and for recommending me.

Last but not least, I would like to thank my family for their support throughout all my life.

List of papers

A Jante's law process

Kennerberg, P., Volkov, S. (2018). Advances in Applied Probability, 50(2), 414-439. doi:10.1017/apr.2018.20

B Convergence in the p-Contest

Kennerberg, P., Volkov, S. Convergence in the p-Contest. J Stat Phys 178, 1096–1125 (2020).

C A local barycentric version of the Bak-Sneppen model

Submitted

D Some stability results for stochastic integrators

Working paper

Abstract

This thesis consists of four papers that broadly concerns two different topics. The first topic is so-called barycentric Markov processes. By a barycentric Markov process we mean a process that consists of a point/particle system evolving in (discrete) time, whose evolution depends in some way on the mean value of the current points in the system. In common for the three first papers which are on this topic is that we study how all the points of the so-called core (a certain subset of points in the system) of the system converge to the same point.

The first article concerns how an N-point system behaves when we reject the $K < N/2$ points that minimize the sample variance of the remaining $N - K$ points (the core). We then replace the rejected points with K new points which follow some fixed distribution and which are all independent from the past points. When $K = 1$ this is equivalent to rejecting the point which is furthest from the center of mass. We prove that under rather weak assumptions on the sampling distribution, the points of the core converge to the same point as well as that regardless of any assumptions on our sampling distribution, the sampling variance of the core converges to zero or the core "drifts off to infinity".

The second article concerns a similar problem as the first one. We once again consider an N-point system but at each time step we reject the point furthest from the center of mass multiplied by a positive number p and replace it with a point from a fixed distribution with full support on $[0, 1]$, which is independent from all past points. If $p = 1$ we obtain a special case of the previous article. If $p \neq 1$ it turns out that this process behaves very differently from the process in the first article, the stationary distribution to which the core points converge turns out to be a Bernoulli distribution.

The third article studies yet another N-point system but now on a discrete circle. During each time step we compute the distances for each point from the mean of it's two neighbours and reject the one with largest such distance (thereby obtaining our core) and replace it with a new point independent from past points. Two different cases are considered, the first is with uniformly distributed points in $[0, 1]$ and the other is with a discrete uniform distribution (i.e. uniformly distributed on an equally spaced grid).

The fourth and last article is on the topic of stochastic calculus. The main objective is to study "stability" of integrators for stochastic integrals. We examine how converging

sequences of processes in the role of integrators retain their convergence properties for their corresponding integrals when the integrators are transformed under certain classes of functions. The convergence is on one hand in the uniform (over compact time intervals) in L^p-sense and on the other hand in the UCP-sense (uniform convergence in probability on compact time intervals). We examine processes with quadratic variations (along some refining sequence) transformed by absolutely continuous functions as well as Dirichlet processes transformed by C^1 functions.

Populärvetenskaplig sammanfattning

Denna avhandling behandlar i huvudsak två olika ämnen. Det första ämnet är, i en bred mening, så kallade barycentriska Markovprocesser. Med barycentriska Markovprocesser åsyftas här, processer som består av en punkt/partikeluppsättning som utvecklas under (diskret) tid och vars utveckling på något sätt styrs av dem ingående partiklarnas medelpunkt. Gemensamt för alla tre artiklar som avhandlar detta ämne är att vi studerar hur punkterna i systemets så kallade kärna (en viss sorts delmängd av systemet punkter) alla konvergerar till en och samma punkt.

Den första artikeln avhandlar hur ett N-punktssystem beter sig när vi förkastar $K < N/2$ punkter som minimerar sampelvariansen av de $N - K$ återstående punkterna (kärnan). Sedan ersätter vi de förkastade punkterna med K nya punkter som följer någon fix fördelning och som är oberoende av tidigare punkter, detta är vad som sker vid ett tidssteg. När $K = 1$ så svarar detta mot att förkasta den punkt som ligger längst ifrån systemets medelpunkt. Vi bevisar att under ganska så svaga antaganden på fördelningen som de nya punkterna slumpas utifrån så konvergerar systemets punkter mot en och samma punkt, när antalet tidssteg går mot oändligheten samt att utan några fördelningsantaganden alls så måste sampelvariansen av kärnan konvergera till noll eller så drar kärnan ”iväg mot oändligheten”.

Den andra artikeln avhandlar ett snarlikt problem som den första artikeln. Vi betraktar återigen ett N-punktssystem men vid varje tidssteg så förkastar vi den här gången den punkt som ligger längst ifrån medelpunkten multiplicerad med en positiv konstant p och ersätter den med en oberoende punkt vars fördelning har fullt stöd på intervallet $[0, 1]$. Om $p = 1$ så har vi ett specialfall av processen som studerades i den första artikeln. Om $p \neq 1$ så visar det sig att denna process beter sig väldigt annorlunda från processen i den föregående artikeln, den stationära fördelningen mot vilken kärnans punkter närmar sig visar sig alltid vara Bernoullifördelad.

Den tredje artikeln studerar återigen ett N-punktssystem men nu på en diskret cirkel. Under ett tidssteg så beräknar vi avstånden till medelpunkten mellan varje punkts två grannar och förkastar den punkt med högst sådant avstånd (detta är vår kärna) och ersätter sedan denna med en ny punkt som är obereonde av tidigare punkter. Vi låter antalet tidssteg gå mot oändligheten och studerar konvergens av kärnans punkter. Två fall behandlas i denna artikel, dels likformigt fördelade punkter på intervallet $[0, 1]$, samt diskret likformigt fördelade punkter (dvs likformigt fördelade över en grid där punkterna är ekvidistanta).

Den fjärde och sista artikeln är inom ämnet stokastisk kalkyl. Huvudsyftet med denna artikel är att studera ”stabilitet” av integratorer för stokastiska integraler. Vi undersöker hur konvergerande följder av processer i egenskap av integratorer bibehåller sin konvergensegenskap för tillhörande integraler när integratorerna transformeras av olika klasser av funktioner. Konvergensen är dels i likformig (över kompakta tidsintervall) L^p-mening och dels i UCP-mening (likformig konvergens i sannolikhet över kompakta tidsintervall). Vi undersöker dels processer med kvadratisk variation (längs någon given förfiningssekvens) transformerade under absolutkontinuerliga funktioner samt Dirichletprocesser transformerade under C^1 funktioner.

Chapter 1

Introduction

1 Semimartingales

In this section we give a *very* minimal introduction to the concept of semimartingales. No proofs of the results are given here, they can all be found in Chapter two of [3]. Semimartingales can be defined in two equivalent manners. The most straightforward definition is to define it as a sum of a local martingale and a processes of finite variation. The second definition, which has more of a functional analysis flare is the following. Let $\mathbf{S}$ denote the space of so called simple predictable processes which are of the form

$$H_t = H_0 + \sum_{k=1}^{n} H_k I_{(T_k, T_{k+1}]}(t), \tag{1.1}$$

for some finite sequence of stopping times $T_1 \leq ... \leq T_{n+1} < \infty$ and where $H_k \in \mathcal{F}_{T_k}$. Given a continuous-time stochastic processes X and a simple predictable process $H \in \mathbf{S}$, we define the linear operator $I_X : \mathbf{S} \to \mathbf{L}^0$ by

$$I_X(H) = H_0 X_0 + \sum_{k=1}^{n} H_k \left(X_{T_{k+1}} - X_{T_k} \right),$$

when H has the form of (1.1).

Definition 1. *An adapted càdlàg process X is called a semimartingale if for each $t \in \mathbb{R}^+$, the map $I_{X^t} : \mathbf{S} \to \mathbf{L}^0$ is continuous in the sense that $\sup_{t,\omega} |H^n_\omega(t) - H_\omega(t)| \to 0$ implies $I_{X^t}(H^n) \xrightarrow{\mathbb{P}} I_{X^t}(H)$.*

The fact that these two definitions are equivalent is known as the Bichteler-Dellacherie Theorem.

Definition 2. *A sequence of processes* $\{X^n\}_n$ *is said to converge in the topology of uniform convergence in probability (ucp) to a* X *if* $(X^n - X)^*_t = \sup_{s \leq t} |X^n_s - X_S| \xrightarrow{\mathbb{P}} 0$.

Theorem 1. *The space* $\mathbf{S}$ *is dense in* $\mathbb{L}$ *(the space of caglad processes) under the ucp topology.*

Definition 3. *Let* $H \in \mathbf{S}$ *and* X *be a cadlag process, we define the (linear) map* $J_X : \mathbf{S} \to \mathbb{D}$*, called the stochastic integral of* H *with respect to* X *by*

$$J_X(H) = H_0 X_0 + \sum_{i=1}^{n} H_i \left(X^{T_{i+1}} - X^{T_i} \right)$$

with H *of the form in* (1.1).

Theorem 2. *The map* $J_X : \mathbf{S}_{ucp} \to \mathbb{D}_{ucp}$ *is continuous (here* $mathbf S_{ucp}$ *and* $\mathbb{D}_{ucp}$ *denotes* $mathbf S$ *and* $\mathbb{D}$ *equipped with the ucp-topology respectively.*

From Theorem 1 and Theorem 2 it follows that $J_X : \mathbb{L}_{ucp} \to \mathbb{D}_{ucp}$ is continuous.

2 Random measures

This section gives a brief introduction to the concept of random measures and their corresponding integrals. All results and definition are taken either from chapter 2 of [12] or chapter 3 of [7], all proofs are omitted but can be found in these textbooks.

Definition 4. *A random measure on* $\mathbb{R}^+ \times \mathbb{R}^d$ *is a family* $\{\mu_\omega(.,.), \omega \in \Omega\}$ *of non-negative measures on* $(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{B}(\mathbb{R}^+) \times \mathbb{B}(\mathbb{R}^d))$ *with the property that* $\mu_\omega(\{0\}, A) = 0$ *for every* $A \in \mathbb{B}(\mathbb{R}^d)$ *(i.e. no point mass at time zero).*

Let $\Omega' := \Omega \times \mathbb{R}^+ \times \mathbb{R}^d$, $\mathcal{O}' := \mathcal{O} \times \mathbb{B}(\mathbb{R}^d)$ and $\mathcal{P}' := \mathcal{P} \times \mathbb{B}(\mathbb{R}^d)$, where $\mathcal{O}$ denotes the optional sigma algebra (the sigma algebra generated by mappings of the type $(\omega, t) \to f(\omega, t)$ where $f(\omega, .) \in \mathbb{D}$, the space of cadlag functions) and $\mathcal{P}$ denotes the predictable sigma algebra (the sigma algebra generated by mappings of the type $(\omega, t) \to f(\omega, t)$ where $f(\omega, .)$ is a caglad function). We say a function on Ω' is optional if it is $\mathcal{O}'$-measurable and say it is predictable if it is $\mathcal{P}'$-measurable. Let $W(\omega, t, x) : \Omega' \to \mathbb{R}$

be an optional function on Ω', since $W(\omega,.,.)$ is $\mathbb{B}(\mathbb{R}^+)\times\mathbb{B}(\mathbb{R}^d))$-measurable for every $\omega\in\Omega$, we may define the integral process $(W\cdot\mu)_t$ as $\int_{[0,t]\times\mathbb{B}(\mathbb{R}^d)} W(\omega,s,x)\mu_\omega(ds,dx)$ when $\int_{[0,t]\times\mathbb{B}(\mathbb{R}^d)} |W(\omega,s,x)|\mu_\omega(ds,dx)<\infty$ and as $+\infty$ otherwise. We say that μ is optional if $W\cdot\mu$ is an optional process for every optional function W. Similarly we say that μ is predictable if $W\cdot\mu$ is predictable for every predictable W.

Definition 5. *An optional measure μ is called $\mathcal{P}'$-σ-finite if there exists a strictly positive predictable function V on Ω' such that* $\mathbb{E}\left[\int_{\mathbb{R}^+\times\mathbb{R}^d} V(\omega,s,x)\mu_\omega(ds,dx)\right]<\infty$

We shall say that nonnegative process is *integrable* if it has an a.s. limit as $t\to\infty$ and this limit has a finite expected value. Also we will use the term *transition kernel* (as is done in [12]) of a measurable space $(A,\mathcal{A})$ into another measurable space $(B,\mathcal{B})$ we will mean a family $\{\alpha(a,.) : a\in A\}$ of non-negative measures on $(B,\mathcal{B})$ such that $\alpha(.,C)$ is $\mathcal{A}$-measurable for each $C\in\mathcal{B}$. Recall (See) the following property, if $(G,\mathcal{G})$ is any measurable space and m is any finite nonnegative measure on $(\mathbb{R}^d\times G,\mathbb{B}(\mathbb{R}^d)\times\mathcal{G})$ with G-marginal $\hat{m}(A)=m(A\times\mathbb{R}^d)$ then there exists a transition kernel α from $(G,\mathcal{G})$ into $(\mathbb{R}^d,\mathbb{B}(\mathbb{R}^d))$ such that $m(B)=\int\int 1_B(g,x)\alpha(g,dx)\hat{m}(dg)$ for all $B\in\mathcal{G}\times\mathbb{B}(\mathbb{R}^d)$.

Theorem 3. *If μ is an optional $\mathcal{P}'$-σ-finite measure then there exists a predictable random measure ν, called the compensator of μ, which is unique up to a $\mathbb{P}$-null set satisfying either one of the following two equivalent conditions*
(i) $\mathbb{E}\left[\int_{\mathbb{R}^+\times\mathbb{R}^d} W(\omega,s,x,)\nu_\omega(ds,dx)\right]=\mathbb{E}\left[\int_{\mathbb{R}^+\times\mathbb{R}^d} W(\omega,s,x,)\mu_\omega(ds,dx)\right]$ *for every non-negative $\mathcal{P}'$-measurable random function W on Ω'.*
(ii) For every $\mathcal{P}'$-measurable function W on Ω' such that $\int_{[0,t]\times\mathbb{R}^d} |W(\omega,s,x,)|\mu_\omega(ds,dx)$ is an integrable process then $\int_{[0,t]\times\mathbb{R}^d} |W(\omega,s,x,)|\nu_\omega(ds,dx)$ is also integrable and $\int_{[0,t]\times\mathbb{R}^d} |W(\omega,s,x,)|\nu_\omega(ds,dx)$ is the compensator process of $\int_{[0,t]\times\mathbb{R}^d} |W(\omega,s,x,)|\mu_\omega(ds,dx)$.
Moreover there exists a (predictable) increasing and integrable process A and a transition kernel $K(\omega,t,dx)$ from $(\Omega\times\mathbb{R}^+,\mathcal{P})$ into $(\mathbb{R}^d,\mathbb{B}(\mathbb{R}^d))$ such that

$$\nu_\omega(dt,dx)=dA_t(\omega)K(\omega,t,dx)$$

Definition 6. *We say that μ is an integer valued random measure if it satisfies:*
1) $\mu_\omega(\{t\}\times\mathbb{R}^d)\leq 1$ a.s. for all $t\in\mathbb{R}^+$,
2) for each $A\in\mathbb{B}(\mathbb{R}^+)\times\mathbb{B}(\mathbb{R}^d)$, $\mu_(A)\in\bar{\mathbb{N}}$
3) μ is optional and $\mathcal{P}'$-σ-finite.

Let ν denote the compensator of an integer valued random measure μ. Denote

$$a_t(\omega)=\nu(\omega,\{t\}\times\mathbb{R}^d)$$

Lemma 1. *There exists a version of the compensator ν of μ such that*

$$a_t(\omega) \leq 1.$$

We say that a random set D is thin if $D = \bigcup_{n\geq 1}[[T_n]]$ where $\{T_n\}_n$ are stopping times and $[[T_n]] = \{(\omega, t) : t \in \mathbb{R}^+, T(\omega) = t\}$

Proposition 1. *If μ is an integer-valued random measure, there exists a thin random set D and an $\mathbb{R}^d$-valued optional process β such that*

$$\mu_\omega(A, B) = \sum_{s\in A} 1_D(\omega, s)\delta_{(s,\beta(s))}(A, B),$$

for $A \in \mathbb{B}(\mathbb{R}^+)$ and $B \in \mathbb{B}(\mathbb{R}^d)$.

Proposition 2. *Let X be an $\mathbb{R}^d$-valued càdlàg process then*

$$\mu(A, B) = \sum_s 1_{\Delta X_s \neq 0}\delta_{(s,\Delta X_s)}(A, B)$$

with $A \in \mathbb{B}(\mathbb{R}^+)$ and $B \in \mathbb{B}(\mathbb{R}^d)$ defines an integer-valued random measure with $D = \{\Delta X_s \neq 0\}$ being the thin set and $\beta(s) = \Delta X_s$ the optional process in the previous proposition.

Let μ be an integer valued random measure, ν it's compensator and such that $|U| \cdot \nu$ is a locally integrable process then so is $|U| \cdot \nu$

Let μ be an integer valued random measure, define the integer valued measure p by

$$p(\omega, \Gamma) = \sum_{s\in\Gamma} 1_{a_s(\omega)>0}(1 - \mu(\omega, \{s\} \times \mathbb{R}^d)),$$

for every $\Gamma \in \mathbb{B}(\mathbb{R}^+)$. The compensator of p is the ,measure q given by (see section 5 of chapter 3 in [7])

$$q(\omega, \Gamma) = \sum_{s\in\Gamma} 1_{a_s(\omega)>0}(1 - a_s(\omega)).$$

Let U be a $\mathcal{P}'$-measurable function such that for each stopping time T

$$\int_{\mathbb{R}^d} |U(\omega, T, x)\nu(\omega, \{T\}, dx)1_{T<\infty} < \infty \text{ a.s.}$$

and define for such U

$$\hat{U}(\omega, t) = \int_{\mathbb{R}^d} U(\omega, t, x)\nu(\omega, \{t\}, dx).$$

Define now the process

$$G(U) = \frac{(U - \hat{U})^2}{1 + |U - \hat{U}|} \cdot p + \frac{\hat{U}^2}{1 + \hat{U}} \cdot q.$$

If $G(U)$ is locally integrable then the integral $U \cdot (\mu - \nu)$ is a well defined local martingale (again, see section 5 of chapter 3 in [7] for a proof).

3 Dirichlet processes

Originally studied by Föllmer in the paper REF for the purpose of developing a pathwise Ito calculus. Today there are several definitions of Dirichlet processes (which are not all equivalent). We will present the original definition proposed by Föllmer (in [5]), but first we introduce the notion of a refining seqeunce. Given some $t > 0$ we say that $\{D_k\}_k$ is a refining sequence of each D_k is a partition of $[0, t]$, $D_k \subseteq D_{k+1}$ and the mesh of D_k tends to zero as $k \to \infty$.

Definition 7. *X is said to be a Dirichlet process (in Föllmer sense) if for any $t > 0$, and for some refining sequence $\{D_k\}_k$ of $[0, t]$,*

$$\sup_{l \geq k} \sum_{t_i \in D_k} \mathbb{E}\left[\left(\sum_{t_i \leq s_j \leq t_{i+1},\, s_j \in D_l} \mathbb{E}\left[X_{s_{j+1}} - X_{s_j} \,|\, \mathcal{F}_{s_j}\right]\right)^2\right],$$

a.s. converges to zero as $k \to \infty$.

Föllmer also showed in [6] that this definition is equivalent to saying that X can be (uniquely) decomposed into a square integrable martingale plus an adapted continuous process starting in 0 with zero quadratic variation along $\{D_k\}_k$. We will work with a definition not equivalent Föllmers, which is much weaker.

Definition 8. *A cadlag process X is called a semimartingale if $X = Z + C$ where Z is a semimartingale and C is an adapted continuous process of zero quadratic variation along some refining sequence $\{D_k\}_{k \geq 1}$.*

By the above definition any semimartingale is a Dirichlet process, something which is obviously not true by Föllmer's definition. By transforming a semimartingale by a C^2 function we know that by Ito's formula we get yet another semimartingale. If we however transform a semimartingale by a C^1 function then in general we do not get a semimartingale but a Dirichlet process. In fact, it was shown in [9] that if we transform a Dirichlet process by a C^1 function we retain another Dirichlet process. In the same article the following representation formula was also proven

Theorem 4. *Let $X = Z + C$ where Z is a semimartingale and C has zero quadratic variation and f be a C^1-function. We have $f(X_s) = Y_s + \Gamma_s$ where Y is a semimartingale, Γ is continuous and $[\Gamma]_t = 0$ for all $t > 0$. The expression for Y is given by*

$$\begin{aligned} Y_t &= f(X_0) + \sum_{s \le t} \left(f(X_s) - f(X_{s-}) - \Delta X_s f(X_{s-})\right) I_{|\Delta X_s|>1} + \int_0^t f'(X_{s-}) dZ_s \\ &+ \int_0^t \int_{|x|\le 1} \left(f(X_{s-}+x) - f(X_{s-}) - x f'(X_{s-})\right)(\mu - \nu)(ds, dx) \\ &+ \sum_{s\le t} \int_{|x| \le a} \left(f(X_{s-}+x) - f(X_{s-}) - x f'(X_{s-})\right) \nu(\{s\}, dx). \end{aligned} \tag{3.2}$$

4 Background for papers A, B and C

Papers A and B (and to a lesser extent paper C) are directly related to the paper by [11], in this paper a number of open problems were posed. Among these problems was one referred to as "a repeated Keynesian beauty contest". Fix a parameter $p > 0$. Start with a uniform array of N elements on $[0, 1]$. At each step, compute the mean μ of the N elements, and replace by a $U[0, 1]$ random variable the element that is farthest (amongst all the N points) from $p\mu$. Thus at each step, either the minimum or maximum is replaced, depending on the current configuration. This is related to the "p-beauty contest" [27, p. 72] in which N players choose a number between 0 and 100, the winner being the player whose choice is closest to p times the average of all the N choices. The stochastic process described above is a repeated, randomized version of this game (without any learning, and with random player behaviour) in which the worst performer is replaced by a new player. In the paper "Convergence in a multidimensional randomized Keynesian beauty contest" in [2] the case $p = 1$ was studied and also generalized to the multivariate case, i.e. considering points in $\mathbb{R}^n$ and with $U[0, 1]^n$ distributed replacement points.

5 Summary of paper A

In this paper we generalize the results of those in [2]. We also redefine the process so that we allow for $K < N/2$ points to be replaced at each step. The criteria for which points that are to be replaced is that we replace those K points that minimize the sample variance of the remaining $N - K$ points (this minimization is the reason for the given name, Jante's law process). If $K = 1$ then this is actually the same as to replace the point furthest from the center of mass, so it coincides with the model in [2]. We also allow for much more general distributions.

5.1 Main outline of paper

In section one we define the model and introduce some auxiliary tools. Section two is devoted to studying the "dissipation" of the system (or to put it more succinctly, convergence of the sample variance to zero). Theorem 1 tells us that if $K \leq N - 2$ (this result is thus more general than our other results which require that $K < N/2$) regardless of the sampling distribution (on $\mathbb{R}^d$), the system can only have two long term behaviours; either the sample variance converges to zero or the system "runs of to infinity" (i.e. the absolute value of at least one of the core points goes to infinity). As a corollary to this theorem we show that when $d = 1$ and the sampling distribution is singular, the core converges to a single point.

In Section 2.1 we prove in Theorem 2 that in the real-valued case, when we remove only one point, under a rather weak assumption on the distribution of the tail, the sample variance converges almost surely to zero. It is possible to construct counterexamples to this assumption by hand, but most common continuous distributions will in fact fulfil the proposed assumption. In addition we also prove that if the core eventually stays in the tail region then the process converges to a single point almost surely; if the tail condition is valid on the whole line then the process converges to a single point (almost surely). An additional result has been added to this theorem that was not present in the published article namely, we prove that if the sampling distribution has finite first moment with the tail condition being valid through all of the support then limit of the expected value of the order statistics (up to $N - 1$) of the core all exist and coincide.

In Section three we first introduce a "local" regularity assumption for distributions on $\mathbb{R}^d$, this is a very weak assumption that we have yet to find any counterexamples to.

In Theorem 3, we prove that if $K < N/2$, either the core drifts off to infinity or it almost surely converges to a single point. In particular, if the sampling distribution has compact support then the core converges almost surely. We also introduce an even weaker assumption (the "matryoshka" condition) when $d = 1$, and show that in the real-valued case we have convergence of the core under this weaker condition. Finally we show that if one combines the "matryoshka" condition in some bounded region and assumes the tail condition used for Theorem 2 then we also have almost sure convergence of the core.

6 Summary of paper B

We study a discrete-time Markovian system of $N \geq 3$ number of points, $\{X_1^0, ..., X_N^0\}$ taking values in $[0, 1]$. Given a fixed parameter $p \in \mathbb{R}^+$ we start with our N points at time $t = 0$, compute their average $\mu(X_1^0, ..., X_N^0)$, remove the point furthest from $p\mu(X_1^0, ..., X_N^0)$ (the remaining $N-1$ points are called "the core" similarly to the Jante's law process) and then replace it with a point independent from all past points, having some fixed distribution ζ, with full support in $[0, 1]$. This procedure is then repeated indefinitely and we study convergence of the core. Study of this process (with $\zeta \in U([0, 1])$) was posed as an open problem in [11].

6.1 Main outline of paper B

The problem is divided into two different cases, $p < 1$ and $p > 1$ (The case $p = 1$ was dealt with in [2] and generalized in [8]). In Section 1 we formalize the model and introduce some notation. In section 2 we tackle the case $p < 1$ when the sampling distribution is uniform on $[0, 1]$. In Theorem 1 we prove that the core almost surely converges to zero. In part, this is based on some very tedious but elementary inequalities that have been deferred to the Appendix. In section three we study the case $p > 1$. We consider all sampling distributions with full support on $[0, 1]$ and establish that the core must converge almost surely either to zero or one, and we provide examples when both outcomes are possible, i.e. when the stationary distribution is a true Bernoulli distribution. In Section four we study the case when $N = 3$, ζ has a nondecreasing density in some neighbourhood of zero and show that in this case $\mathcal{X}'(t) \to 1$ a.s.. The second section in the Appendix contains the original proof (the published one) of

Theorem 2.

7 Summary of paper C

We study the behaviour of an interacting particle system, related to the Bak-Sneppen model and Jante's law process defined in [8]. Let $N \geq 3$ vertices be placed on a circle, such that each vertex has exactly two neighbours. To each vertex we assign a real number, called *fitness*. We pick the vertex which fitness deviates most from the average of the fitnesses of its two immediate neighbours (in case of a tie, draw uniformly among such vertices), and replace it by a random value drawn independently according to some distribution ζ. We show that when ζ is finitely supported on a uniform grid or has a continuous uniform distribution, all the fitnesses, except one, converge to the same value. The model we study in this paper is a "marriage" between Jante's law process (defined in [8]) and the Bak-Sneppen (BS) model. In the BS model, N species are located around a circle, and each of them is associated with a so-called "fitness", which is a real number. The algorithm consists in choosing the least fit individual, and then replacing it *and both of its two closest neighbours* by a new species, with a new random and independent fitness.

7.1 Main outline of paper C

In Section one we formally define the model and introduce necessary notation. In Section two we study the discrete case and show that the process is a finite state space Markov chain which almost surely gets absorbed, regardless of initial state. In Section three we deal with the case when the sampling distribution is $U[0,1]$ and show that the core converges almost surely to a single point.

8 Summary of paper Paper D

We consider limits for sequences of the type $\int Y_- df_n(X^n)$, for semimartingale integrands, where both the functions $\{f_n\}_n$ and the processes $\{X^n\}_n$ tend to some limits, f and X respectively. An important ingredient is then to study the limit of $[f_n(X^n) - f(X)]$ which is an interesting problem in its own right. We provide an important application which is jump removal. We consider processes $\{X^n\}_n$ admitting to quadratic variations and absolutely continuous functions $\{f_n\}_n$ which are dominated by some locally integrable function and study convergence in the UCP setting. We also consider the case when $\{X^n\}_n$ are Dirichlet processes and $\{f_n\}_n$ are C^1 functions whose derivatives converge uniformly on compacts. We provide important examples of how to apply this theory for sequential jump removal.

8.1 Main outline of paper D

In the first section we introduce the notion of quadratic- and covariation along a refining sequence. In the second section we go through some notation, recall some results from the theory of Dirichlet processes and prove lemma's that will be used in the main results. Of more general interest we prove that processes admitting to quadratic variations are closed under absolutely continuous transforms and that these processes are well-defined integrators for semimartingale integrands. In the third section we state our main results and prove the results concerning stability of integrators under either C^1- (for Dirichlet processes) or absolutely continuous transforms (for processes admitting to quadratic variations), in either in uniform L^p convergence setting (for Dirichlet processes) or in the UCP setting (for processes admitting to quadratic variations). In section four we provide examples for the results of the previous section in terms of so-called jump truncations. Of particular interest are processes with jumps of finite variation where we can simply remove jumps on a by-modulus basis for smaller and smaller truncation levels. We also show certain commutation properties of such jump truncation with regards to our earlier stability results. In the Appendix we give proofs of results deferred to this section.

9 Bibliography

[1] Sandemose, Aksel (1936). A fugitive crosses his tracks. translated by Eugene Gay-Tifft. New York: A. A. Knopf.

[2] Grinfeld, Michael; Volkov, Stanislav; Wade, Andrew R. Convergence in a multidimensional randomized Keynesian beauty contest. Adv. in Appl. Probab. 47 (2015), no. 1, 57–82.

[3] *Stochastic Integration and Differential Equations.* Springer-Verlag, Berlin, Heidelberg, Second edition.

[4] Parthasarathy, K. R. (2005). Probability measures on metric spaces. AMS Chelsea Publishing, Providence, RI.

[5] Föllmer, H. (1980) Calcul d'Ito sans probabilités. Séminaire de Probabilités, pp. 144-150. Lecture Notes Math., 850. Berlin: Springer-Verlag.

[6] Föllmer, H. (1981). Dirichlet processes. Stochastic Integrals, pp. 476-478. Lecture Notes Math., 851. Berlin: Springer-Verlag

[7] Liptser, R. and Shiryaev, A. (1986) *Theory of Martingales.* Springer-Verlag.

[8] Jante's law process. *Adv. in Appl. Probab.* **50**, 414–439.

[9] Coquet, F., Mémin, J. and Slomiński, L. (2003). On Non-Continuous Dirichlet Processes. *Journal of Theoretical Probability*, **16**, 197.

[10] On the discrete Bak-Sneppen model of self-organized criticality. Proceedings Of The Twelth Annual ACM-SIAM Symposium On Discrete Algorithms (SODA), Washington DC.

[11] Rank-driven Markov processes, *J. Stat. Phys.* **146** (2012), 378–407.

[12] Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes.* Springer-Verlag.

A

Chapter 2

Paper A

Jante's law process

Philip Kennerberg[1,2] and Stanislav Volkov[*,3]

29th November 2020

[1]Center for Mathematical Sciences, Lund University, Sölvegatan 18A, Box 118, 22100 Lund, Sweden.
[2]Email address: philipk@maths.lth.se
[3]Email address: s.volkov@maths.lth.se

Abstract

Fix some integers $d \geq 1$, $N \geq 3$, $1 \leq K < N$ and a $d-$dimensional random variable ζ. Define *an energy* of configuration of m points as the sum of all pairwise distances squared[1]. The process starts with initially N distinct points on $\mathbb{R}^d$. Next, of the total N points keep those $N-K$ which minimize the energy amongst all the subsets of size $N-K$, and replace thrown out points by K i.i.d. points sampled according to ζ, and of the total $N+K$ points keep those N which minimize the energy amongst all the subsets of size N. Repeat this process ad infinitum. We obtain various quite non-restrictive conditions under which the set of points converges to a certain limit. Observe that this is a very substantial generalization of the "Keynesian beauty contest process" introduced in [3] where $K = 1$ and the distribution ζ was uniform on the unit cube.

Keywords: Keynesian beauty contest, rank driven processes, interacting particle systems.

Subject classification: 60J05, 60D05, 60K35.

1 Introduction and main result

We study a generalization of the model presented in Grinfeld et al. [3]. Fix an integer $N \geq 3$ and some d-dimensional random variable ζ. Now arbitrary choose N distinct points on $\mathbb{R}^d$, $d \geq 1$. The process in [3], called there "Keynesian beauty contest process", is a discrete-time process with the following dynamics: given the configuration of N points we compute its center of mass μ and throw away the most distant from μ point; if there is more than one, we choose each one with equal probability. Then this point is replaced with a new point drawn independently each time from the distribution of ζ. In [3] it was shown that when ζ has a uniform distribution on a unit cube, then the configuration converges to some random point on $\mathbb{R}^d$, with the exception of the most distant point.

The aim of this paper is to remove the assumption on uniformity of ζ and obtain

[1] Please note that in physics this often corresponds to the moment of inertia; however, it can be viewed as "the energy" from the perspective of potential theory. For simplicity, we will use this term in the current paper.

some general sufficient conditions under which the similar convergence takes place. Additionally, it turns out we can naturally generalize the process by removing not just one but $K \geq 2$ points at the same time, and then replacing them with new K i.i.d. points sampled from ζ. We also give the process we introduce a different name, which we believe describes its essence much better. The "Law of Jante" is the concept that there is a pattern of group behaviour towards individuals within Scandinavian countries that criticises individual success and achievement as unworthy and inappropriate, in other words, it is better to be "like everyone else". The concept was created by Aksel Sandemose in [1], identified the Law of Jante as ten rules, and has been a very popular concept in Nordic countries since then.

We will use mostly the same notations as in [3]. Namely, let $\mathcal{X}_n = (x_1, x_2, \dots, x_n)$ for a vector of n points $x_i \in \mathbb{R}^d$; let $\mu_n(\mathcal{X}_n) := n^{-1} \sum_{i=1}^n x_i$ be the barycentre of $\mathcal{X}_n$. Denote by $\operatorname{ord}(\mathcal{X}_n) = (x_{(1)}, x_{(2)}, \dots, x_{(n)})$ the barycentric order statistics of $x_1, \dots, x_n$, so that

$$\|x_{(1)} - \mu_n(\mathcal{X}_n)\| \leq \|x_{(2)} - \mu_n(\mathcal{X}_n)\| \leq \dots \leq \|x_{(n)} - \mu_n(\mathcal{X}_n)\|.$$

Here and throughout the paper $\|x\|$ denotes the Euclidean norm in $\mathbb{R}^d$, $x \cdot y$ is a dot product of two vectors $x, y \in \mathbb{R}^d$, and $B_r(x) = \{y \in \mathbb{R}^d : \|y - x\| < r\}$ is an open ball of radius r centred at x. As in [3] let us also define for $\mathcal{X}_n = (x_1, x_2, \dots, x_n) \in \mathbb{R}^{dn}$

$$G_n(\mathcal{X}_n) := G_n(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{i-1} \|x_i - x_j\|^2 = \sum_{i=1}^n \|x_i - \mu_n(\mathcal{X}_n)\|^2 = \inf_{y \in \mathbb{R}^d} \sum_{i=1}^n \|x_i - y\|^2.$$

We can think of $G_n(\mathcal{X}_n)$ as of a measure of "diversity" among individuals with properties $x_1, \dots, x_n$.

In [3] where $K = 1$, the authors called $x_{(n)}$ the *extreme* point of $\mathcal{X}_n$, that is, a point of $x_1, \dots, x_n$ farthest from the barycentre, and the defined *core* of $\mathcal{X}_n$ as $\mathcal{X}_n' := (x_{(1)}, \dots, x_{(n-1)})$, the vector of $x_1, \dots, x_n$ with (one of) the extreme point removed. They also defined $F_n(\mathcal{X}_n) := G_{n-1}(\mathcal{X}_n')$ and $F(t) = F_N(\mathcal{X}(t))$.

In our paper, when $K \geq 1$, we re-define the core as the subset of $x_1, \dots, x_N$ containing $N - K$ elements which minimizes the diversity of the remaining individuals, that is the subset which minimizes

$$\min_{\{y_1, \dots, y_{N-K}\} \subset \{x_1, \dots, x_N\}} G_{N-K}(y_1, \dots, y_{N-K}).$$

We will show below that, in fact, when $K = 1$ both definitions coincide.

The process runs as follows: Let $\mathcal{X}(t) = \{X_1(t), \dots, X_N(t)\}$ be distinct points in $\mathbb{R}^d$. Given $\mathcal{X}(t)$, let $\mathcal{X}'(t)$ be the core of $\mathcal{X}(t)$ and replace $\mathcal{X}(t) \setminus \mathcal{X}'(t)$ by K i.i.d. ζ-distributed random variables so that

$$\mathcal{X}(t+1) = \mathcal{X}'(t) \cup \{\zeta_{t+1;1}, \dots, \zeta_{t+1;K}\},$$

where $\zeta_{t;j}$, $t = 1, 2, \dots$, $j = 1, 2, \dots, K$, are i.i.d. random variables with a common distribution ζ. In case there is more than one element in the core, that is, a few configurations which minimize diversity, we chose any of it with equal probability, precisely as it was done in [3]. Now let $F(t) = G_{n-K}(\mathcal{X}'(t))$.

Finally, to finish specification of the process, we allow the initial configuration $\mathcal{X}(0)$ be arbitrary or random, with the only requirement that all the points of $\mathcal{X}(0)$ must lie in the support of ζ.

The following statement links the case $K = 1$ with the general $K \geq 1$.

Lemma 1. *If $K = 1$ then the only point not in the core is the one which is the furthermost from the center of mass of $\mathcal{X}$.*

Proof. Let $\mathcal{X} = (x_1, \dots, x_N)$. W.l.o.g. assume $\sum_{i=1}^{N} x_i = \mathbf{0} \in \mathbb{R}^d$ and thus the center of mass of $\mathcal{X}$ is located at $\mathbf{0}$. Here L consists of all subsets of $\{1, \dots, N\}$ containing just one element. If we throw away the l-th point, denoting $\mu_l = \frac{1}{N-1} \sum_{i \neq l} x_i = -\frac{x_l}{N-1}$ we get

$$\begin{aligned} G(l, \mathcal{X}) &= \sum_{i=1}^{N} \|x_i - \mu_l\|^2 - \|x_l - \mu_l\|^2 = \sum_{i=1}^{N} \|x_i\|^2 + N\|\mu_l\|^2 - 2\mu_l \cdot \sum_{i=1}^{N} x_i - \|x_l - \mu_l\|^2 \\ &= \sum_{i=1}^{N} \|x_i\|^2 + N\frac{\|x_l\|^2}{(N-1)^2} - \frac{\|x_l N\|^2}{(N-1)^2} = -\|x_l\|^2 \frac{N}{(N-1)^2} + \sum_{i=1}^{N} \|x_i\|^2. \end{aligned}$$

Therefore, the minimum of $G(l, \mathcal{X})$ is achieved by choosing an x_l with the largest $\|x_l\|$, that is, the furthermost from the centre of mass. □

Corollary 1. *For $K = 1$ Jante's law process coincides with the process studied in [3].*

The following statement is a trivial consequence of the definition of F.

Lemma 2. *For any $1 \leq K \leq N - 2$ and any distribution of ζ we have $F(t+1) \leq F(t)$.*

In case $K = 1$ the above statement coincides with Corollary 2.1 in [3].

Remark 1. *It is worth noting that throwing away $\mathcal{X}^*$ in general does not mean necessary throwing the K furthest points from the centre of mass of $\mathcal{X}$, unlike the case $K = 1$.*

Here's an example with $d = 1$, $N = 5$ and $K = 3$: set $\mathcal{X} = (-24, -19, -14, 28, 29)$. Then the centre of mass is at $\mu = 0$ and thus 28 and 29 have the largest and the second largest distance from μ, while it is clear that the energy is minimized by keeping exactly these two points in the core and throwing away the rest.

Finally, define *the range* of the configuration: for $n \geq 2$ and $x_1, \ldots, x_n \in \mathbb{R}^d$, write

$$D_n(x_1, \ldots, x_n) := \max_{1 \leq i,j \leq n} \|x_i - x_j\|.$$

The following statement is taken from [3] (Lemma 2.2.).

Lemma 3. *Let $n \geq 2$ and $x_1, \ldots, x_n \in \mathbb{R}^d$. Then*

$$\frac{1}{2} D_n(x_1, \ldots, x_n)^2 \leq G_n(x_1, \ldots, x_n) \leq \frac{1}{2}(n-1) D_n(x_1, \ldots, x_n)^2.$$

Let $D(t) = D_{N-K}(\mathcal{X}'(t))$. Then we have from Lemma 3

$$\sqrt{\frac{2}{N-K-1} \cdot F(t)} \leq D(t) \leq \sqrt{2F(t)}. \tag{1.1}$$

From Lemmas 2 and 3 it also follows immediately that

$$D(t+1) \leq \sqrt{2F(t)} \leq D(t)\sqrt{N-K-1}. \tag{1.2}$$

Let also $\mu'(t) = \mu_{N-K}(\mathcal{X}'(t))$ be the centre of mass of the core.

Assumption 1. $2K < N$.

Observe that if Assumption 1 is not fulfilled, then all the points of the points of the core can migrate large distances and that $F = 0$ does not necessarily imply that the configuration stops moving. For example, one can take $N = 4$, $K = 2$, and $\zeta \sim Bernoulli(p)$ to see that the core jumps from 0 to 1 and back infinitely often a.s.

In the other case, the new core must contain at least one point of the old core, and the following statement shows that if newly sampled points are far from the core, they immediately get rejected.

Lemma 4. *Under Assumption 1, if all the distances between K newly sampled points and the points of the core are more than $C := D\sqrt{N-K-1}$ then $\mathcal{X}'(t+1) = \mathcal{X}'(t)$.*

Proof. Since $N - 2K \geq 1$ the new core $\mathcal{X}'(t+1)$ must contain at least one point of the old core $\mathcal{X}'(t)$. By (1.2) $D(t+1) \leq D(t)\sqrt{N-K-1}$ and therefore if one of the new points is in the new core, it should be no further than $D(t)\sqrt{N-K-1}$ from one of the points of the old core. □

Finally, we will use the following notations. For any two sets $A, B \subset \mathbb{R}^d$ let

$$\operatorname{dist}(A, B) = \inf_{x \in A, y \in B} \|x - y\|.$$

If $d = 1$ then write $\mathcal{X}'(t) \to +\infty$ if $\lim_{t\to\infty} \min\{x,\ x \in \mathcal{X}'(t)\} = \infty$ and similarly $\mathcal{X}'(t) \to -\infty$ if $\lim_{t\to\infty} \max\{x,\ x \in \mathcal{X}'(t)\} = -\infty$. If $d \geq 2$ we will write $\mathcal{X}'(t) \to \infty$ if $\min\{\|x\|,\ x \in \mathcal{X}'(t)\} = \operatorname{dist}(\mathcal{X}'(t), 0) \to \infty$, otherwise we will write $\mathcal{X}'(t) \not\to \infty$. We will also write $\mathcal{X}'(t) \to \phi \in \mathbb{R}^d$ if all the coordinates of $\mathcal{X}'(t)$ converge to ϕ as $t \to \infty$.

2 Shrinking

Let ζ be *any* random variable on $\mathbb{R}^d$. As usual, define the support of this random variable as

$$\operatorname{supp}\zeta = \overline{\{A \in \mathbb{R}^d : \ \mathbb{P}(\zeta \in A) > 0\}} = \{x \in \mathbb{R}^d : \ \forall \varepsilon > 0 \ \mathbb{P}(\zeta \in B_\epsilon(x)) > 0\},$$

where the overline denotes set closure (see e.g. [5]). We also say that $\operatorname{supp}\zeta$ is bounded in $\mathbb{R}^d$ if there is an $M > 0$ such that $\mathbb{P}(\|\zeta\| < M) = 1$.

It turns out that the following statement, which is probably known, is true.

Proposition 1. *$\operatorname{supp}\zeta$ is bounded if and only if there exists some function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that for any $x \in \operatorname{supp}\zeta$*

$$\mathbb{P}(\zeta \in B_\delta(x)) \geq f(\delta)$$

for all $\delta > 0$.

Proof. Suppose such a function exists, but the support of ζ is not bounded. Fix any $\Delta > 0$. Then there must exist a infinite sequence of points $\{x_n\}_{n=1}^\infty \subseteq \operatorname{supp}\zeta$, such that $\|x_i - x_j\| > 2\Delta$, whenever $i \neq j$. Since the sets $\{B_\Delta(x_n)\}$ are disjoint, this would imply that

$$\mathbb{P}\left(\zeta \in \mathbb{R}^d\right) \geq \mathbb{P}\left(\bigcup_{n=1}^\infty \{\zeta \in B_\Delta(x_n)\}\right) \geq \sum_{n=1}^\infty f(\Delta) = \infty$$

which is impossible.

Conversely, assume that $\operatorname{supp}\zeta$ is bounded. For all $\delta > 0$ define

$$f(\delta) = \inf_{x \in \operatorname{supp}\zeta} \mathbb{P}(\|\zeta - x\| \le \delta).$$

We will show that $f(\delta) > 0$. Indeed, if not, there exists a sequence $\{x_n\}$ such that $\mathbb{P}(\|\zeta - x_n\| \le \delta) \to 0$ as $n \to \infty$. Since the support of ζ is compact, $\{x_n\}$ must have a convergent subsequence; w.l.o.g. we can assume that it is $\{x_n\}$ itself and thus there is an x such that $x_n \to x$ and there exists N such that $\|x_n - x\| < \delta/2$ for all $n \ge N$. On the other hand, for these n

$$\mathbb{P}(\|\zeta - x\| \le \delta/2) \le \mathbb{P}(\|\zeta - x_n\| \le \delta)$$

by the triangle inequality. Since the RHS converges to zero, this implies $\mathbb{P}(\|\zeta - x\| \le \delta/2) = 0$ so $x \notin \operatorname{supp}\zeta$ which contradicts the fact that $x = \lim_{n\to\infty} x_n \in \operatorname{supp}\zeta$ by the definition of the support. $\square$

Theorem 1. *Given any distribution ζ on $\mathbb{R}^d$, for any $N \ge 3$ and $1 \le K \le N-2$ we have*

$$\mathbb{P}\left(\{F(t) \to 0\} \bigcup \{\mathcal{X}'(t) \to \infty\}\right) = 1.$$

In particular if ζ has compact support, then $F(t) \to 0$ a.s.

Note that $F(t) \to 0$ is equivalent to $D(t) \to 0$.

Proof. We will first make use of the following lemma.

Lemma 5. *Suppose we are given a bounded set $S \in \mathbb{R}^d$ such that $\mathbb{P}(\zeta \in S) > 0$ and $N - K$ points $x_1, ..., x_{N-K}$ in $\operatorname{supp}(\zeta) \bigcap S$ satisfying $F(\{x_1, ..., x_{N-K}\}) > \varepsilon_1$. Let $\varepsilon_2 = \frac{\varepsilon_1}{2(N-K)^2}$. Then there exists a positive constant σ, only depending on ε_1, S, K and N, such that*

$$\mathbb{P}\left(F\left(\{\zeta_1, \dots, \zeta_K, x_1, \dots, x_{N-K}\}'\right) < F(\{x_1, \dots, x_{N-K}\}) - \varepsilon_2\right) \ge \sigma.$$

Proof. We start with the case $K = 1$. Denote $D = \max_{1 \le i,j \le N-K} \|x_i - x_j\|$, and $S_* = \{x : \operatorname{dist}(x, S) < D\sqrt{N-K-1}\}$, then the set $\overline{S_*}$ is a compact set such that $\{\zeta, x_1, \dots, x_{N-1}\}' \in \overline{S_*}$ regardless of where the point ζ is sampled, by Lemma 4. Since $\overline{S_*}$ is compact it follows from Proposition 1 applied to $\zeta \cdot 1_{\{\zeta \in S\}}$ that there is an $f : \mathbb{R}^+ \to \mathbb{R}^+$, such that for any $x \in \operatorname{supp}\zeta \bigcap \overline{S_*}$, we have $\mathbb{P}(\zeta \in B_\delta(x)) \ge f(\delta)$. Assume that the core centre of mass $\mu' = 0$, and that (without loss of generality) $\|x_1\| \geq \|x_l\|, \forall 1 \leq l \leq N-1$. Let $\mu' = \frac{y+x_2+\ldots+x_{N-1}}{N-1}$ and consider the function

$$h(y) = \sum_{i=2}^{N-1} \|x_i - \mu'\|^2 + \|y - \mu'\|^2,$$

continuous in y. Pick a point x_j from $\{x_2, ..., x_{N-1}\}$ such that $\|x_1 - x_j\| \geq \frac{D}{2}$ – otherwise $\|x_i - x_j\| \leq \|x_1 - x_j\| + \|x_1 - x_i\| < D$, for all indices i, j, contradicting the definition of D.

Consider the configuration $\{x_j, x_2, ..., x_{N-1}\}$, where we have removed the point x_1 and replaced it with x_j. This configuration has centre of mass $\mu' = \frac{x_2+\ldots+x_{N-1}+x_j}{N-1} = \frac{x_j - x_1}{N-1}$. The Lyapunov function evaluated for this configuration is precisely $h(x_j)$. Denote $F_{\text{old}} = F(\{x_1, ..., x_{N-1}\})$, then

$$\begin{aligned}
h(x_j) &= \sum_{i=2}^{N-1} \|x_i - \mu'\|^2 + \|x_j - \mu'\|^2 = \sum_{i=1}^{N-1} \|x_i - \mu'\|^2 + \|x_j - \mu'\|^2 - \|x_1 - \mu'\|^2 \\
&= \sum_{i=1}^{N-1} \left(\|x_i\|^2 + \|\mu'\|^2 - 2x_i \cdot \mu'\right) + \|x_j\|^2 + \|\mu'\|^2 - 2x_j \cdot \mu' - \|x_1\|^2 - \|\mu'\|^2 + 2x_1 \cdot \mu' \\
&= \sum_{i=1}^{N-1} \|x_i\|^2 + (N-1)\|\mu'\|^2 + \|x_j\|^2 - \|x_1\|^2 - 2\left(x_j - x_1\right) \cdot \left(\frac{x_j - x_1}{N-1}\right) \\
&\leq F_{\text{old}} + \frac{\|x_j - x_1\|^2}{N-1} - 2\frac{\|x_j - x_1\|^2}{N-1} \leq F_{\text{old}} - \frac{D^2}{4(N-1)} \leq \left(1 - \frac{1}{2\left(N-1\right)^2}\right) F_{\text{old}},
\end{aligned}$$

where the last inequality follows from (1.1). Hence for some $\delta > 0$ if $\|y - x_j\| \leq \delta$ then $h(y) < \left(1 - \frac{1}{4(N-1)^2}\right) F_{\text{old}}$. So if ζ is sampled in $B_\delta(x_j)$ then we have a substantial decrease and this is with probability bounded below by $f(\delta)$, the result is thus proved for the case $K = 1$ with $\sigma = f(\delta)$.

The general case can be reduced to the case $K = 1$ as follows. Set $N' := N - K + 1$ and replace all N by N' in the arguments above. The decrease of F in this case will be at least by $\varepsilon_2(N')$. Indeed, since if at least one particle falls in the ball $\{y : \|y - x_j\| \leq \delta\}$ we could choose the sub-configuration where exactly one point falls in this ball while x_1 is removed, and since we are taking the minimum over all available configurations, the decrease has to be greater or equal than for this specific choice. ☐

Assume that $\mathbb{P}(\mathcal{X}'(t) \to \infty) < 1$, otherwise the theorem follows immediately. Recall that $B_r(0)$ is a ball of radius r centred at the origin and note that

$$\{\mathcal{X}'(t) \not\to \infty\} = \bigcup_{r=1}^{\infty} \{\mathcal{X}'(t) \in B_r(0) \ i.o.\} = \bigcup_{r=1}^{\infty} G_r \tag{2.3}$$

where $G_r = \bigcap_{k\geq 0}\{\tau_{k,r} < \infty\}$, $\tau_{k,r} = \inf\{t : t > \tau_{k-1,r}, \mathcal{X}'(t) \in B_r(0)\}$, $\quad k = 1, 2, \dots,$

with the convention that $\tau_{0,r} = 0$, $\inf\emptyset = +\infty$ and that if $\tau_{k,r} = +\infty$ then $\tau_{k',r} = +\infty$ for all $k' \geq k$.

By the monotonicity of F we have $F(t) \downarrow F_\infty \geq 0$. We will show that in fact

$$\mathbb{P}\left(\{\mathcal{X}'(t) \not\to \infty\}\bigcap\{F_\infty > 0\}\right) = 0 \tag{2.4}$$

which is equivalent to the statement of the Theorem.

Let n_0 be some integer larger than $4(N-K)^2$, this quantity being related to ε_2 from Lemma 5. Since

$$\{F_\infty > 0\} = \bigcup_{n=n_0}^{\infty}\left\{F_\infty > \frac{1}{n}\right\} = \bigcup_{n=n_0}^{\infty}\bigcup_{m=0}^{\infty}\{F_\infty \in I_{n,m}\}, \quad \text{where} \quad I_{n,m} = \left[\frac{1}{n} + \frac{m}{n^2}, \frac{1}{n} + \frac{m+1}{n^2}\right)$$

are disjoint sets for each fixed n. Consequently, taking into account (2.3), to establish (2.4) it suffices to show for each fixed n and m and r we have

$$\mathbb{P}\left(G_r \bigcap \{F_\infty \in I_{n,m}\}\right) = 0.$$

Let $A_k = \{F(\tau_{k,r} + 1) \in I_{n,m}\}\bigcap\{\tau_{k,r} < \infty\}$ then obviously

$$G_r\bigcap\{F_\infty \in I_{n,m}\} \subset \bigcup_{k_0\geq 0}\bigcap_{k\geq k_0} A_k. \tag{2.5}$$

We will show now that for all k_0 we have $\mathbb{P}\left(\bigcap_{k\geq k_0} A_k\right) = 0$. which will imply that the probability of the RHS and hence that of the LHS of (2.5) is 0. Indeed, for any positive integer L

$$\mathbb{P}\left(\bigcap_{k\geq k_0} A_k\right) \leq \mathbb{P}\left(\bigcap_{k=k_0}^{k_0+L} A_k\right) = \mathbb{P}(A_{k_0})\prod_{k=k_0+1}^{t_0+L}\mathbb{P}\left(A_k \mid \bigcap_{s=k_0}^{k-1} A_s\right).$$

We now proceed to calculate the conditional probabilities, $\mathbb{P}\left(A_k \mid \bigcap_{s=k_0}^{k-1} A_s\right)$. Setting $\varepsilon_1 = \frac{1}{n}$ and letting S be the ball of radius $\sqrt{2(1/n + (m+1)/n^2)}\left(1 + \sqrt{N-K-1}\right)$ centred at 0 in Lemma 5 and using the bound (1.1), we obtain

$$\varepsilon_2 = \frac{\varepsilon_1}{4(N-K)^2} = \frac{1}{4n(N-K)^2} > \frac{1}{n^2}$$

and thus with probability at least σ, given by Lemma 5, F will exit $I_{n,m}$, that is,

$$\mathbb{P}\left(F(\tau_{k,r} + 1) \in I_{n,m} \mid F(\tau_{k_0,r} + 1), F(\tau_{k_0+1,r} + 1), \dots, F(\tau_{k-1,r} + 1) \in I_{n,m}, \tau_{k,k} < \infty\right) \leq 1-\sigma,$$

since $\zeta_{\tau_{k,r}+1;j}$ are all independent from $\mathcal{F}_{\tau_{k,r}}$ for $1 \le j \le K$.

From this we can conclude that, $\mathbb{P}\left(A_k \mid \bigcap_{s=k_0}^{k-1} A_s\right) \le 1-\sigma$ yielding $\mathbb{P}\left(\bigcap_{k\ge k_0} A_k\right) \le (1-\sigma)^L$ for all $L \ge 1$. Letting $L \to \infty$ shows that $\mathbb{P}\left(\bigcap_{k\ge k_0} A_k\right) = 0$, which in turn proves (2.4). □

Corollary 2. *Suppose Assumption 1 holds, $d = 1$, and ζ has a singular distribution. Then*

$$\mathbb{P}\left(\{\exists \phi :\ \mathcal{X}'(t) \to \phi\} \bigcup \{\mathcal{X}'(t) \to \infty\}\right) = 1.$$

Proof. Assume that $\mathcal{X}'(t) \not\to \infty$ occurs and for $a < b$ define

$$E_{a,b} = \{\liminf_{t\to\infty} x_{(k)}(t) < a\} \cap \{\limsup_{t\to\infty} x_{(k)}(t) > b\},$$

where $k \in \{1, 2, \dots, N-K\}$ and $x_{(k)}$ is the $k-$th point of the core. By Theorem 1 $F(t) \to 0$ implying, in turn, that $D(t) \to 0$, and hence by Lemma 4

$$\mathsf{dist}(\mathcal{X}'(t), \mathcal{X}'(t+1)) := \max_{1\le i,j\le N-K} |x_{(i)}(t) - x_{(j)}(t+1)| \to 0 \tag{2.6}$$

as $t \to \infty$.

Since ζ is singular $\exists x \in (a, b)$ and $\epsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \subseteq \text{supp}(\zeta)^c$. However, then

$$E_{a,b} \subseteq \text{dist}(\mathcal{X}'(t), \mathcal{X}'(t+1)) > 2\varepsilon \ \text{ i.o.}\}$$

implying from (2.6) that $\mathbb{P}(E_{a,b}) = 0$. Since this is true for all a and b, $\mathcal{X}'(t)$ must converge. □

2.1 Case $K = d = 1$.

In the case where $K = 1$ and the space is $\mathbb{R}^1$ we can obtain some more detailed results, given some further assumptions.

Assumption 2 (at most exponential oscillations in the tail)**.** *Suppose that there exist some $R_+, R_- \in \mathbb{R}$, a constant $C \ge 0$ such that given for any $a \ge R_+$ and $u > 0$ we have*

$$\mathbb{P}\left(a+u < \zeta \le a+2u\right) \le C\,\mathbb{P}\left(a < \zeta \le a+u\right).$$

Similarly for all $a \le R_-$ and $u < 0$ we have

$$\mathbb{P}\left(a+2u < \zeta \le a+u\right) \le C\,\mathbb{P}\left(a+u < \zeta \le a\right).$$

Remark 2. *Observe that nearly all common continuous distributions satisfy this assumption (exponential, normal, Pareto, etc.). An example of distribution for which the assumption is not fulfilled is e.g. one with the density*

$$f(x) = \begin{cases} \frac{1}{2}e^{-|x|}, & \lfloor x \rfloor \textit{ is even,} \\ e^{-2|x|}, & \textit{otherwise} \end{cases}$$

which has support on the whole $\mathbb{R}$.

By iterating the property in Assumption 2 for $a \geq R_+$ one attains that for $k = 1, 2, \ldots$

$$\mathbb{P}\left(\zeta \in (a + (k-1)u, a + ku]\right) \leq C^{k-1}\, \mathbb{P}\left(\zeta \in (a, a+u]\right).$$

It also follows that if we take $R_+ < a < b < c$ then

$$\mathbb{P}\left(\zeta \in (b, c]\right) \leq \mathbb{P}\left(\zeta \in \bigcup_{k=1}^{\lceil \frac{c-a}{b-a} \rceil} \left(a + (k-1)(b-a), k(b-a)\right]\right) \leq \sum_{k=1}^{\lceil \frac{c-a}{b-a} \rceil} C^{k-1}\, \mathbb{P}\left(\zeta \in (a, b]\right). \tag{2.7}$$

Using (2.7) one can compare the probabilities of selecting a new point in the intervals of different length and/or that are not consecutive; we see that in this case the upper bound we get is a polynomial in C.

Remark 3. *The assumption is somewhat related to the concept of O-regular variation (see [2], page 65) in the following sense: if we let* $g(x) := \mathbb{P}(R_+ < \zeta \leq R_+ + x)$ *for* $x > 0$ *then we see from* (2.7) *that* $\limsup_{x\to\infty} \frac{g(tx)}{g(x)} \leq \sum_{k=1}^{\lceil t \rceil} C^{k-1}$ *for* $t \geq 1$. *Therefore,* g *is an O-regularly varying function; moreover, if the support of* ζ *is* $\mathbb{R}^+$ *and* $R_+ = 0$ *then the distribution function of* ζ *itself is an O-regularly varying function.*

Assumption 2 immediately implies that the tail region is free of isolated atoms; moreover, it turns out that the tail region is free of atoms altogether.

Claim 1. *Suppose that Assumption 2 holds. Then* $\mathbb{P}(\zeta = x) = 0$ *for every* $x \in (-\infty, R_-) \cup (R_+, \infty)$.

Proof. Assume to the contrary that $\exists x \in (-\infty, R_-) \cup (R_+, \infty)$ such that $\mathbb{P}(\zeta = x) > 0$. Since $\mathbb{P}(\zeta = x) = \mathbb{P}\left(\bigcap_{n=1}^{\infty}\{\zeta \in (x - \frac{1}{n}, x]\}\right)$, by continuity of probability it follows that $\exists N$ such that $\mathbb{P}(\zeta \in (x - \frac{1}{N}, x]) \leq \left(\frac{1}{2C} + 1\right)\mathbb{P}(\zeta = x)$ which implies that $\mathbb{P}(\zeta \in (x - \frac{1}{N}, x)) \leq \frac{1}{2C}\,\mathbb{P}(\zeta = x)$. Therefore we have

$$\mathbb{P}\left(\zeta \in \left(x - \frac{1}{2N}, x - \frac{1}{N}\right]\right) \leq \mathbb{P}(\zeta \in (x - \frac{1}{N}, x)) \leq \frac{1}{2C}\,\mathbb{P}(\zeta = x) \leq \frac{1}{2C}\,\mathbb{P}\left(\zeta \in (x - \frac{1}{2N}, x]\right),$$

which contradicts Assumption 2. □

Theorem 2. *Suppose $K = 1$ and ζ satisfies Assumption 2 for some R_+ and R_-. Then*

(a) $\mathcal{X}' \not\to \infty$ a.s. and consequently by Theorem 1 we have $F(t) \to 0$ a.s..

(b)

$$\left\{\liminf_{t\to\infty} x_{(1)}(t) > R_+\right\} \bigcup \left\{\limsup_{t\to\infty} x_{(N-1)}(t) < R_-\right\} \subseteq \{\exists \phi : \ \mathcal{X}'(t) \to \phi\}$$

except perhaps a set of measure 0.

(c) Assuming $\mathbb{E}\,|\zeta| < \infty$, then if $\operatorname{supp}\zeta = [R_+, \infty)$ or if $\operatorname{supp}\zeta = (-\infty, R_-]$ then the limits $\lim_{t\to\infty} \mathbb{E}\, x_{(k)}(t)$ and $\lim_{t\to\infty} \mathbb{E}\, \mu'(t)$ are both well defined for $1 \le k \le N-1$ ($x_{(k)}(t)$ denotes the k:th barycentric order statistic) and moreover

$$\lim_{t\to\infty} \mathbb{E}\, x_{(k)}(t) = \lim_{t\to\infty} \mathbb{E}\, \mu'(t).$$

(d) If $R_- > R_+$ then $\mathbb{P}\left(\exists \phi : \ \mathcal{X}'(t) \to \phi\right) = 1$.

Remark 4. *The last part of the theorem above applies to many distributions for which $\operatorname{supp}\zeta = \mathbb{R}$, e.g. to normal, Laplace or Cauchy distribution (one can take $R_+ = -1$ and $R_- = +1$).*

Proof. We begin with the proof of (a). Given some $L \ge 1$, from now on assume that $A_L = \left\{\sqrt{2F(0)} < \frac{L}{2}, |\zeta_{0;k}| < L, k = 1 \dots N\right\}$ occurs, this will imply that $D(t) \le \frac{L}{2}$ for all t. Notice that since the distance between any two points in the core at time t is bounded by $D(t)$ it follows that if one core point diverges to $+\infty$ so must all the other points, similarly if one of the points diverges to $-\infty$ so must all of the rest. Therefore it is enough to show that $\mathbb{P}\left(\{\mathcal{X}'(t) \to +\infty\} \bigcup \{\mathcal{X}'(t) \to -\infty\}\right) = 0$. We shall prove now that $\mathcal{X}'(t) \not\to +\infty$ a.s.; the proof that $\mathcal{X}'(t) \not\to -\infty$ a.s. is completely analogous.

Let $\pi_a = \inf\{t : \sqrt{2F(t)} < \frac{a}{2}\}$, $\eta_{1,a} = \tau_{1,a} = \pi_a$ and for $k > 1$ let

$$\begin{aligned}
\tau_{k,a} &= \inf\left\{t > \eta_{k-1,a} : \ x_{(1)}(t) > R_+ + a\right\}, \\
\eta_{k,a} &= \inf\left\{t > \tau_{k,a} : \quad x_{(1)}(t) < R_+ + a\right\}, \\
\gamma_{k,t,a} &= \min\left(\eta_{k,a}, \tau_{k,a} + t\right),
\end{aligned}$$

where $x_{(1)}(t)$ denotes the left-most point of the core at time t. If $\tau_{k,a} = \infty$ for some k then we set $\eta_{m,a} = \tau_{m,a} = \infty$ for all $m \ge k$. It is obvious that on A_L, $\pi_L = 0$.

Furthermore

$$\{\tau_{k,L} = \infty\} \cap \{\eta_{k-1,L} < \infty\} \subseteq \{\limsup_{t\to\infty} x_{(1)}(t) \le R_+ + a\} \subseteq \{\mathcal{X}'(t) \not\to +\infty\}.$$

Let $C_k = \{\eta_{k,L} < \infty\}$ and note

$$\left(\bigcap_{k=2}^{\infty} C_k\right) \subseteq \{\mathcal{X}'(t) \subseteq B_{R_+ + 2L}(0)\ i.o.\} \subseteq \{\mathcal{X}'(t) \not\to +\infty\}.$$

Since $(\bigcap_{k=1}^{\infty} C_k) \subseteq \{\mathcal{X}'(t) \not\to +\infty\}$, if we could also show that

$$\mathbb{P}\left(\left(\bigcap_{k=2}^{\infty} C_k\right)^c \setminus \{\mathcal{X}'(t) \not\to +\infty\}\right) = \mathbb{P}\left(\left(\bigcup_{k=2}^{\infty} \{\eta_{k,L} = \infty\}\right) \bigcap \{\mathcal{X}'(t) \to +\infty\}\right) = 0, \tag{2.8}$$

then it would follow that $\mathbb{P}\left(A_L \bigcap \{\mathcal{X}'(t) \not\to +\infty\}\right) = \mathbb{P}(A_L)$ and since $\mathbb{P}\left(\bigcup_{L=1}^{\infty} A_L\right) = 1$ it would then follow from continuity of probability that $\mathbb{P}\left(\mathcal{X}'(t) \to +\infty\right) = 0$.

Now we will show that $\mathbb{P}\left(\{\eta_{k,L} = \infty\} \bigcap \{\mathcal{X}'(t) \to +\infty\}\right) = 0$ for every $k > 1$ which will establish (2.8). For this purpose (and for the purpose of showing the other statements of the theorem) we will need the following lemma

Lemma 6. *For some fixed $k > 1$ and $a > 0$ let*

$$h_c(s) = \left(\sqrt{F(s)} + c\left[\mu'(s) + \max(0, -R_+)\right]\right) I_{A_L}.$$

Then there exists $c > 0$ such that $\lim_{t\to\infty} h_c(\gamma_{k,t,a})$ exists a.s. on $\tau_{k,a} < \infty$.

Proof of Lemma 6. We will show that $h_c(\gamma_{k,t,a})$ is a non-negative supermartingale with respect to $\{\mathcal{F}_{\gamma_{k,t,a}}\}_{t\ge 0}$, and then the result will follow from the supermartingale convergence theorem. In order to make notations less cluttered from now on we set $\gamma_t := \gamma_{k,t,a}$ throughout the proof of this lemma. First, observe that the positivity of $h_c(\gamma_t)$ is ensured by the term $c\max(0, -R_+)$, and by the definition of γ_t and π_a. Therefore, from now on we can assume that $R_+ \ge 0$ without loss of generality. We have

$$\mathbb{E}\,|h_c(s)| \le \mathbb{E}\left[\left(\sqrt{F(s)} + c|\mu'(s)|\right) I_{A_L}\right] \le$$
$$\mathbb{E}\left[\left(\sqrt{F(0)} + c\left(|\mu'(0)| + \sum_{l=1}^{s} |\mu'(l) - \mu'(l-1)|\right)\right) I_{A_L}\right]$$
$$\le \mathbb{E}\left[\left(\frac{L}{2\sqrt{2}} + c\left(|\mu'(0)| + \sum_{l=1}^{s} D(l)\right)\right) I_{A_L}\right] \le \mathbb{E}\left[\left(\frac{L}{2\sqrt{2}} + c\left(L + \sum_{l=1}^{s} \sqrt{2F(l)}\right)\right) I_{A_L}\right]$$

$$\leq \mathbb{E}\left[\left(L + c\left(L + s\sqrt{2F(0)}\right)\right) I_{A_L}\right] \leq L\left(1 + c\left(1 + s/2\right)\right) < \infty,$$

where we used Lemma 3, the fact that $|\mu'(0)| \leq \max_{x \in \mathcal{X}'(0)} |x| \leq L$, $|\mu'(s+1) - \mu'(s)| \leq D(s+1), s \geq 0$, and that F is non-increasing. Hence $\mathbb{E}\,|h_c(s)| < \infty$.

Since $\{\gamma_t < \eta_{k,a}\} \in \mathcal{F}_{\gamma_t}$ we have

$$\begin{aligned}
&\mathbb{E}\left[h_c(\gamma_{t+1}) - h_c(\gamma_t) \,|\, \mathcal{F}_{\gamma_t}\right] = \mathbb{E}\left[(h_c(\gamma_{t+1}) - h_c(\gamma_t))\left(I_{\gamma_t = \eta_{k,a}} + I_{\gamma_t < \eta_{k,a}}\right) \,|\, \mathcal{F}_{\gamma_t}\right] \\
&= \mathbb{E}\left[(h_c(\gamma_t + 1) - h_c(\gamma_t))\, I_{\gamma_t < \eta_{k,a}} \,|\, \mathcal{F}_{\gamma_t}\right] = \mathbb{E}\left[h_c(\gamma_t + 1) - h_c(\gamma_t) \,|\, \mathcal{F}_{\gamma_t}\right] I_{\gamma_t < \eta_{k,a}} \\
&\leq \max\left(0, \mathbb{E}\left[(h_c(\gamma_t + 1) - h_c(\gamma_t)) \,|\, \mathcal{F}_{\gamma_t}\right]\right) I_{\gamma_t < \eta_{k,a}} \\
&\leq \max\left(0, \mathbb{E}\left[(h_c(\gamma_t + 1) - h_c(\gamma_t)) \,|\, \mathcal{F}_{\gamma_t}\right]\right).
\end{aligned}$$

It will suffice now to show that $\mathbb{E}(h(\gamma_t + 1) - h(\gamma_t) \,|\, \mathcal{F}_{\gamma_t}) \leq 0$ a.s. Since $\gamma_t \leq \eta_{k,a}$ we can deduce

$$x_{(1)}(\gamma_t) \geq x_{(1)}(\eta_{k,a}) \geq x_{(1)}(\eta_{k,a} - 1) - D(\eta_{k,a} - 1) > R_+ + a - \sqrt{2F(\pi_a)} > R_+ + \frac{a}{2}. \tag{2.9}$$

The above inequalities show that all the core points lie to the right of R_+ at time γ_t, since this region is free of atoms we can conclude that $D(\gamma_t) > 0$ a.s.. Recall that the points of the core at time γ_t are ordered as $x_{(1)}(\gamma_t) \leq ... \leq x_{(N-1)}(\gamma_t)$, and let $\zeta = \zeta_{\gamma_t+1}$.

Let us introduce some new variables where we drop the time indices for the sake of brevity:

$$\begin{aligned}
D &= D(\gamma_t), & \mathcal{F} &= \mathcal{F}_{\gamma_t}, \\
y_k &= \frac{x_{(k)}(\gamma_t) - x_{(1)}(\gamma_t)}{D}, & \zeta' &= \frac{\zeta - x_{(1)}(\gamma_t)}{D}, \\
F_o &= \sqrt{F\left(\{y_1, \cdots, y_{N-1}\}\right)}, & F_n &= \sqrt{F\left(\{y_1, \cdots, y_{N-1}, \zeta'\}'\right)}, \\
\mu_o' &= \mu\left(\{y_1, \cdots, y_{N-1}\}\right), & \mu_n' &= \mu\left(\{y_1, \cdots, y_{N-1}, \zeta'\}'\right).
\end{aligned}$$

At time γ_t the transformed core consists of the new points $(y_1, \ldots, y_k)$ such that $0 = y_1 \leq \cdots \leq y_{N-1} = 1$. Notice that we will always reject ζ' if $\zeta' < -1$ but this is equivalent to $\zeta < x_{(1)}(\gamma_t) - D$ which is bounded below by $x_{(1)}(\gamma_t) - \frac{a}{2}$, by (2.9) this is strictly larger than R_+ so we can conclude that ζ is accepted into the core only if it lies to the right of R_+. Furthermore if $a > -1$ then since ζ is independent of $\mathcal{F}$ it follows that

$$\mathbb{P}\left(\zeta' \in (a + u, a + 2u]\right) = \mathbb{P}\left(\zeta \in \left((a + u)D + x_{(1)}(\gamma_t), (a + 2u)D + x_{(1)}(\gamma_t)\right]\right)$$

$$\le C\,\mathbb{P}\left(\zeta \in \left(aD + x_{(1)}(\gamma_t), (a+u)D + x_{(1)}(\gamma_t)\right]\right) = C\,\mathbb{P}\left(\zeta' \in (a, a+u]\right), \qquad (2.10)$$

hence Assumption 2 translates to ζ'. If we combine (2.10) with the same type of argument as in (2.7) we see that if $-1 < a < b < c$, then

$$\mathbb{P}\left(\zeta' \in (b, c]\right) \le \sum_{k=1}^{\lceil \frac{c-a}{b-a} \rceil} C^{k-1}\,\mathbb{P}\left(\zeta' \in (a, b]\right). \qquad (2.11)$$

Due to the translation invariance of $\sqrt{F}$ and μ we have

$$\mu'(\gamma_t + 1) - \mu'(\gamma_t) = D\left(\mu_n' - \mu_o'\right),$$
$$F\left(\gamma_t + 1\right) - F\left(\gamma_t\right) = D\left(\sqrt{F_n} - \sqrt{F_o}\right),$$

implying

$$\frac{1}{D}\left(h(\gamma_t + 1) - h(\gamma_t)\right) = \sqrt{F_n} - \sqrt{F_o} + c\left(\mu_n' - \mu_o'\right).$$

Denote $\Delta h = \sqrt{F_n} - \sqrt{F_o} + c\left(\mu_n' - \mu_o'\right)$; since $D > 0$ a.s. it follows that

$$\mathbb{E}\left[\left(h(\gamma_{t+1}) - h(\gamma_t)\right) \mid \mathcal{F}\right] \le 0 \quad \iff \quad \mathbb{E}\left[\Delta h \mid \mathcal{F}\right] \le 0.$$

When the new point ζ is sampled then either 0,1 or ζ' is eliminated in the next step. There are 4 different cases, either $\zeta' < 0$, $\zeta' \in (0, 1)$, $\zeta' > 1$ (recall that ζ has no atoms under Assumption 2). The new centre of mass for the whole configuration is thus

$$\mu_n = \frac{\zeta' + M\mu_o'}{M+1}, \qquad \text{where } M := N - 1.$$

If the point 0 is eliminated then centre of mass of the new core is $\mu_n' = \frac{\zeta'}{M} + u_o'$, and if the point 1 is eliminated then $\mu_n' = \frac{\zeta'-1}{M} + \mu_o'$. Note that by Claim 1 our probability measure is non-atomic to the right of R_+ and therefore the probability of a tie between which point should be eliminated is zero; consequently, we can disregard these events.

- In the case $\zeta' < 0$, only ζ' or 1 can be eliminated. The point 1 is eliminated if and only if $\mu_n - \zeta' < 1 - \mu_n$. This is equivalent to $\zeta' > \frac{M(2\mu_o'-1)-1}{M-1}$. So in this case the point 1 is eliminated if and only if $\zeta' \in \left(\frac{M(2\mu_o'-1)-1}{M-1}, 0\right)$. Denote this event by
$$L_1 = \left\{\min\left(\frac{M(2\mu_o' - 1) - 1}{M-1}, 0\right) < \zeta' < 0\right\}.$$

- In the case $\zeta' \in (0,1)$, ζ' is never eliminated, but one of the points 0 or 1 must be. The point 0 is eliminated iff $\mu_n > 1-\mu_n$, which is equivalent to $\zeta' > \frac{M+1}{2} - M\mu_o'$, hence $\zeta' \in \left(\min\left(\frac{M+1}{2} - M\mu_o', 1\right), 1\right)$. Let
$$B_0 = \left\{\min\left(\frac{M+1}{2} - M\mu_o', 1\right) < \zeta' < 1\right\}.$$
The point 1 is eliminated otherwise, in other words if
$\zeta' \in (0,1) \setminus \left[\min\left(\frac{M+1}{2} - M\mu_o', 1\right), 1\right]$. Let
$$B_1 = \left\{0 < \zeta' < \min\left(\frac{M+1}{2} - M\mu_o', 1\right)\right\}.$$

- In the case $\zeta' > 1$ only ζ' or 0 can be eliminated. The point 0 will be eliminated if $\zeta' - \mu_n < \mu_n \iff \zeta' < \frac{2M\mu_o'}{M-1}$, that is if $\zeta' \in \left(1, \max\left(\frac{2M\mu_o'}{M-1}, 1\right)\right)$. Let
$$R_0 = \left\{1 < \zeta' < \max\left(\frac{2M\mu_o'}{M-1}, 1\right)\right\}.$$

We begin with the case $M = 2$. We have $\mu_o' = \frac{1}{2}$, $F_o = \frac{1}{2}$, $L_1 = \{-1 < \zeta' < 0\}$, $B_1 = \{0 < \zeta' < 1/2]\}$, $B_0 = \{1/2 < \zeta' < 1\}$, $R_0 = \{1 < \zeta' < 2\}$. When 1 is eliminated then $F_n = \frac{\zeta'^2}{2}$, moreover notice that in this case $\mu_o' - \mu_n'$ is non-positive. When 0 is eliminated then $\mu_n' = \frac{1+\zeta'}{2}$. We have

$$\begin{aligned}
\mathbb{E}(\Delta h \,|\, \mathcal{F}) &= \mathbb{E}\left[(\mu_n' - \mu_o') + c\,(F_n - F_o) \,|\, \mathcal{F}\right] \le c\,\mathbb{E}\left[(F_n - F_o)\, I_{L_1 \cup B_1} \,|\, \mathcal{F}\right] \\
&\quad + \mathbb{E}\left[(\mu_n' - \mu_o')\, I_{R_0 \cup B_0} \,|\, \mathcal{F}\right] \le \frac{c}{2}\,\mathbb{E}\left[\left(\zeta'^2 - 1\right) I_{B_1} \,|\, \mathcal{F}\right] + \frac{1}{2}\,\mathbb{E}\left[\zeta' I_{R_0 \cup B_0} \,|\, \mathcal{F}\right] \\
&\le \frac{c}{2}\left(\frac{1}{4} - 1\right) \mathbb{P}\left(0 < \zeta' < 1/2\right) + \frac{2}{2}\,\mathbb{P}\left(1/2 < \zeta' < 2\right) \\
&\le -\frac{3}{8} c\,\mathbb{P}\left(0 < \zeta' < 1/2\right) + \left(1 + C + C^2 + C^3\right) \mathbb{P}\left(0 < \zeta' < 1/2\right),
\end{aligned}$$

where we used (2.11) in the last inequality. It is obvious that the last expression can be made negative for large enough $c > 0$, as required.

Let us now consider the case $M \ge 3$. First we note that $\mu_o' \in \left(\frac{1}{M}, \frac{M-1}{M}\right)$ a.s., where the lower bound is approached as $y_2, ..., y_{M-1}$ all go to 0 while the upper bound is approached as $y_2, ..., y_{M-1}$ all go to 1. If we now denote by K_0 the event that 0 is eliminated, and K_1 the event that 1 is eliminated, then we have $K_0 = R_0 \cup B_0$ and $K_1 = L_1 \cup B_1$. Furthermore,

$$\mu_n' - \mu_o' = \frac{\zeta'}{M} I_{K_0} + \frac{\zeta' - 1}{M} I_{K_1}.$$

We also have

$$F_n = \left(F_o + \frac{M-1}{M}\zeta'^2 - 2\mu'_o\zeta'\right) I_{K_0}$$
$$+ \left(F_o + \frac{M-1}{M}\zeta'^2 - \frac{2(M\mu'_o - 1)}{M}\zeta' + \frac{2(M\mu'_o - 1)}{M} - \frac{M-1}{M}\right) I_{K_1}.$$

Observe that $\Delta h = h_0 I_{K_0} + h_1 I_{K_1}$, where

$$h_i = \sqrt{F_o + \Delta_i(\zeta', \mu'_o)} + c\frac{\zeta'}{M} - \sqrt{F_o} \quad i = 0, 1;$$

$$\Delta_i(x, y) = \frac{1}{M} \cdot \begin{cases} (M-1)x^2 - 2Mxy, & i = 0; \\ (M-1)(x^2-1) + 2(1-x)(My-1) & i = 1. \end{cases}$$

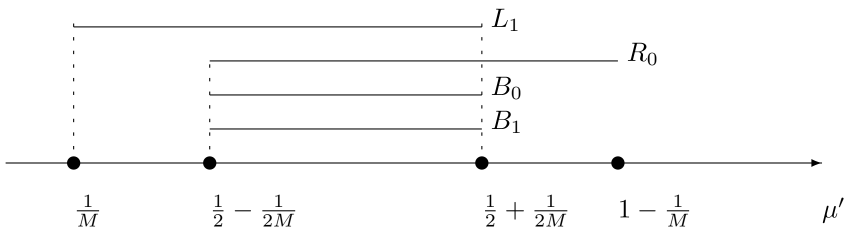
Using these notations we obtain

$$\begin{aligned} \mathbb{E}[\Delta h \,|\, \mathcal{F}] &= \mathbb{E}[h_0 I_{K_0} \,|\, \mathcal{F}] + \mathbb{E}[h_1 I_{K_1} \,|\, \mathcal{F}] \\ &= \mathbb{E}[h_0 I_{R_0} \,|\, \mathcal{F}] + \mathbb{E}[h_0 I_{B_0} \,|\, \mathcal{F}] + \mathbb{E}[h_1 I_{L_1} \,|\, \mathcal{F}] + \mathbb{E}[h_1 I_{B_1} \,|\, \mathcal{F}] \\ &= (I) + (II) + (III), \end{aligned}$$

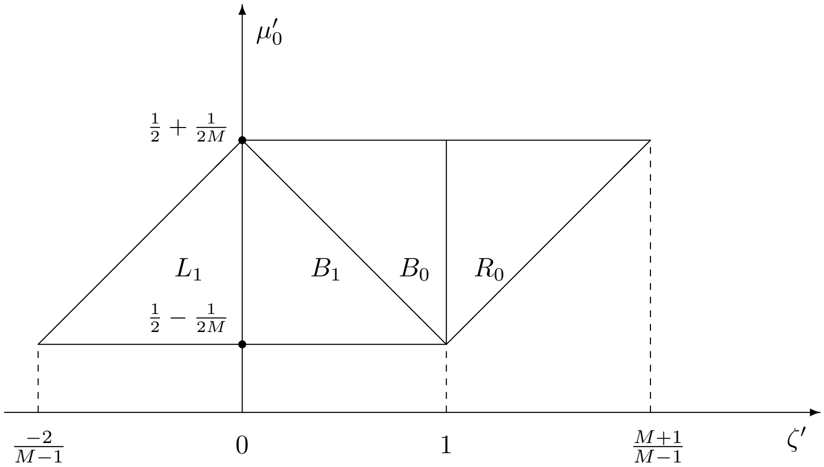
where

$$\begin{aligned} (I) &= (\mathbb{E}[h_1 I_{L_1} \,|\, \mathcal{F}])\, I_{\mu'_o \in \left(\frac{1}{M}, \frac{M-1}{2M}\right)}, \\ (II) &= (\mathbb{E}[h_1 I_{L_1} \,|\, \mathcal{F}] + \mathbb{E}[h_1 I_{B_1} \,|\, \mathcal{F}] + \mathbb{E}[h_0 I_{R_0} \,|\, \mathcal{F}] + \mathbb{E}[h_0 I_{B_0} \,|\, \mathcal{F}])\, I_{\mu'_o \in \left(\frac{M-1}{2M}, \frac{M+1}{2M}\right)}, \\ (III) &= (\mathbb{E}[h_0 I_{R_0} \,|\, \mathcal{F}])\, I_{\mu'_o \in \left(\frac{M+1}{2M}, \frac{M-1}{M}.\right)} \end{aligned}$$

(Please see also the following diagram showing locations of ζ' for the events L_1, B_1, B_0 and R_0.)



It will suffice to show that all the three terms in the expression for $\mathbb{E}[\Delta h \,|\, \mathcal{F}]$ are non-positive. The fact that $(I) \leq 0$ is obvious, since if 1 is eliminated then the core centre of mass must move leftwards while F is always non-increasing. The second term (II) is a little more complicated and requires more careful study. We illustrate the possible combinations of ζ' and μ'_o on the following diagram.



We know present the following elementary statement.

Claim 2. *Let* $\Delta < 0$. *Then*

$$\sqrt{F_o + \Delta} - \sqrt{F_o} \le -\frac{\Delta}{2M}.$$

Proof of Claim 2. The inequality follows from the fact that $\sqrt{F_0} \le \sqrt{M/2} \le M$ and the trivial inequality $\sqrt{x+y} - \sqrt{x} \le \frac{y}{2\sqrt{x}}$ valid for all $x > 0$ and $x + y \ge 0$. □

Next, we find an upper bound for $\Delta_1(x, y)$ on the rectangle

$$A_1 = \left\{ (x,y) : \frac{M-1}{2M} \le y \le \frac{1}{2}, 0 \le x \le \frac{1}{2} \right\}.$$

Combining these estimates with Claim 2 we get that for $\frac{M-1}{2M} \le \mu_o' \le \frac{1}{2}$ and $0 \le \zeta' \le \frac{1}{2}$ (which is a subset of $B_1 \cap \{\frac{M-1}{2M} \le \mu_o' \le \frac{1}{2}\}$)

$$\sqrt{F_o + \Delta_1(\zeta', \mu_o')} - \sqrt{F_o} \le -\frac{1}{2M^2}. \tag{2.12}$$

On the other hand, if $\mu_o' \ge 1/2$ and $0 \le \zeta' \le 1$ then $\Delta_0(\zeta', \mu_o') \le ((M-1)/M - 2\mu_o')\,\zeta' \le -\zeta'/M$ and therefore by Claim 2

$$\sqrt{F_o + \Delta_0(\zeta', \mu_o')} - \sqrt{F_o} \le -\frac{\zeta'}{2M^2}. \tag{2.13}$$

Our next task is to find an upper bound for $\Delta_0(x, y)$ on the rectangle

$$A_2 := \left\{ (x,y) : \frac{1}{2} \le y \le \frac{M+1}{2M}, 1 \le x \le \frac{2M-1}{2M-2} \right\}.$$

As a result, we conclude that $\Delta_0 \leq -\frac{1}{4M}$ on A_2. Combining this with Claim 2 we get that when $\frac{1}{2} \leq \mu_o' \leq \frac{M+1}{2M}$ and $1 \leq \zeta' \leq \frac{2M-1}{2(M-1)}$ (this is a subset of $R_0 \cap \{\frac{1}{2} \leq \mu_o' \leq \frac{M+1}{2M}\}$)

$$\sqrt{F_o + \Delta_0(\zeta', \mu_o')} - \sqrt{F_o} \leq -\frac{1}{8M^2}. \tag{2.14}$$

We will also again make use of the fact that by definition $h_1 I_{L_1} \leq 0$ and $h_1 I_{B_1} \leq 0$ so therefore,

$$\left(\mathbb{E}\left[h_1 I_{L_1} \,|\, \mathcal{F}\right] + \mathbb{E}\left[h_1 I_{B_1} \,|\, \mathcal{F}\right]\right) I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{M+1}{2M}\right)} \leq \mathbb{E}\left[h_1 I_{B_1} \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)}.$$

Now we make the following estimates:

$$\begin{aligned}
(II) \leq\ & \mathbb{E}\left[h_1 I_{B_1} \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)} + \left(\mathbb{E}\left[h_0 I_{R_0} \,|\, \mathcal{F}\right] + \mathbb{E}\left[h_0 I_{B_0} \,|\, \mathcal{F}\right]\right) I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{M+1}{2M}\right)} \\
\leq & \,\mathbb{E}\left[\left(\sqrt{F_o + \Delta_1(\zeta', \mu_o')} - \sqrt{F_o}\right) I_{B_1} \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)} \\
+ & \ \mathbb{E}\left[\left(\sqrt{F_o + \Delta_0(\zeta', \mu_o')} - \sqrt{F_o}\right) \left(I_{B_0} + I_{R_0}\right) \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)} \\
+ & \ \frac{c}{M} \left(\mathbb{E}\left[\zeta' I_{B_0} \,|\, \mathcal{F}\right] + \mathbb{E}\left[\zeta' I_{R_0} \,|\, \mathcal{F}\right]\right) I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{M+1}{2M}\right)} \\
\leq & \ \mathbb{E}\left[\left(\sqrt{F_o + \Delta_1(\zeta', \mu_o')} - \sqrt{F_o}\right) I_{0 \leq \zeta' \leq \frac{1}{2}} \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)} \\
+ & \ \mathbb{E}\left[\left(\sqrt{F_o + \Delta_0(\zeta', \mu_o')} - \sqrt{F_o}\right) \left(I_{B_0} + I_{1 \leq \zeta' \leq \frac{2M-1}{2(M-1)}}\right) \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)} \\
+ & \ \frac{c}{M} \left(\mathbb{E}\left[\zeta' I_{B_0} \,|\, \mathcal{F}\right] + \mathbb{E}\left[\zeta' I_{R_0} \,|\, \mathcal{F}\right]\right) I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{M+1}{2M}\right)},
\end{aligned} \tag{2.15}$$

where we used the fact that $\{0 < \zeta' < 1/2\} \cap \{\frac{M-1}{2M} < \mu_o' < 1/2\} \subseteq \{\frac{M-1}{2M} < \mu_o' < 1/2\} \cap B_1$, that $\{1 \leq \zeta' \leq \frac{2M-1}{2(M-1)}\} \cap \{\frac{1}{2} < \mu_o' < \frac{M+1}{2M}\} \subseteq \{\frac{1}{2} < \mu_o' < \frac{M+1}{2M}\} \cap R_0$, and that on B_1 we have that $h_1 \leq \sqrt{F_o + \Delta_1(\zeta', \mu_o')} - \sqrt{F_o}$. Let us now study the terms in (2.15). Notice that the term in the last line of (2.15) (a.s.) equals

$$\frac{c}{M} \left(\mathbb{E}\left[\zeta' I_{B_0} \,|\, \mathcal{F}\right] + \mathbb{E}\left[\zeta' I_{R_0} \,|\, \mathcal{F}\right]\right) \left(I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)} + I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)}\right),$$

while it follows from (2.13) and (2.14) that

$$\begin{aligned}
& \mathbb{E}\left[\left(\sqrt{F_o + \Delta_0(\zeta', \mu_o')} - \sqrt{F_o}\right) \left(I_{B_0} + I_{1 \leq \zeta' \leq \frac{2M-1}{2(M-1)}}\right) \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)} \\
& \leq \left(\mathbb{E}\left[-\frac{\zeta'}{2M^2} I_{B_0} \,|\, \mathcal{F}\right] - \frac{1}{8M^2} \mathbb{P}\left(1 \leq \zeta' \leq \frac{2M-1}{2(M-1)}\right)\right) I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)}.
\end{aligned}$$

From (2.12) it also follows that

$$\mathbb{E}\left[\left(\sqrt{F_o + \Delta_1(\zeta', \mu_o')} - \sqrt{F_o}\right) I_{0 < \zeta' < \frac{1}{2}} \,|\, \mathcal{F}\right] I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)} \leq -\frac{1}{2M^2} \mathbb{P}\left(0 < \zeta' < \frac{1}{2}\right) I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)}.$$

Furthermore we note that $\mathbb{E}[\zeta' I_{B_0} \,|\, \mathcal{F}] \le \mathbb{P}(B_0)$ and $\mathbb{E}[\zeta' I_{R_0} \,|\, \mathcal{F}] \le \frac{M}{M-1}\,\mathbb{P}(R_0)$ for $\frac{M-1}{2M} < \mu_o' < \frac{1}{2}$ while $\mathbb{E}[\zeta' I_{R_0} \,|\, \mathcal{F}] \le \frac{M+1}{M-1}\,\mathbb{P}(R_0)$ when $\mu_o' < \frac{M+1}{2M}$. We can now conclude

$$
\begin{aligned}
(II) \le & \left[-\frac{1}{2M^2}\,\mathbb{P}\left(0 < \zeta' < 1/2\right) + \frac{c}{M}\left(\mathbb{P}(B_0) + \frac{M}{M-1}\,\mathbb{P}(R_0) \right) \right] I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2}\right)} \\
& + \mathbb{E}\left[\left(\frac{c\zeta'}{M} - \frac{\zeta'}{2M^2} \right) I_{B_0} \,|\, \mathcal{F} \right] I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)} \\
& + \left(-\frac{1}{8M^2}\,\mathbb{P}\left(1 < \zeta' < \frac{2M-1}{2(M-1)} \right) + \frac{c}{M}\frac{M+1}{M-1}\,\mathbb{P}(R_0) \right) I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)} \\
\le & \left[\frac{c}{M}\left(C_1 + \frac{M}{M-1} C_2 \right) - \frac{1}{2M^2} \right] \mathbb{P}\left(0 < \zeta' < 1/2\right) I_{\mu_o' \in \left(\frac{M-1}{2M}, \frac{1}{2},\right)} \\
& + \left[C_3 \frac{c}{M}\frac{M+1}{M-1} - \frac{1}{8M^2} \right] \mathbb{P}\left(1 < \zeta' < \frac{2M-1}{2(M-1)} \right) I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)} \\
& + \mathbb{E}\left[\zeta' \left(\frac{c}{M} - \frac{1}{2M^2} \right) I_{B_0} \,|\, \mathcal{F} \right] I_{\mu_o' \in \left(\frac{1}{2}, \frac{M+1}{2M}\right)},
\end{aligned}
$$

where

$$
\begin{aligned}
C_1 &= \frac{\mathbb{P}\big(\zeta' \in (0,1)\big)}{\mathbb{P}(0<\zeta'<1/2)} \ge \frac{\mathbb{P}(B_0)}{\mathbb{P}(0<\zeta'<1/2)}, \qquad C_2 = \frac{\mathbb{P}\big(\zeta' \in (1,2)\big)}{\mathbb{P}(0<\zeta'<1/2)} \ge \frac{\mathbb{P}(R_0)}{\mathbb{P}(0<\zeta'<1/2)}, \\
C_3 &= \frac{\mathbb{P}\big(\zeta' \in (1,2)\big)}{\mathbb{P}\big(1<\zeta'<\frac{2M-1}{2(M-1)}\big)} \ge \frac{\mathbb{P}(R_0)}{\mathbb{P}\big(1<\zeta'<\frac{2M-1}{2(M-1)}\big)}.
\end{aligned}
$$

It follows from (2.11) that these constants are all bounded above by some polynomial in C whose power depends only on M; also note that $\zeta' \ge 0$ on $B_0 \cap \{\frac{1}{2} \le \mu_0' \le \frac{M+1}{2M}\}$. Therefore it is obvious that we can pick c small enough to make the first two terms in the last displayed inequality above non-positive, the last term is trivially non-positive since $\zeta' \ge 0$ on B_0.

Now we will show that $(III) \le 0$. We begin by finding an upper bound for $\Delta_0(x,y)$ on the rectangle

$$
A_3 = \left\{ (x,y) : \frac{M+1}{2M} \le y \le \frac{M-1}{M}, 1 \le x \le \frac{M}{M-1} \right\}.
$$

Hence $\Delta_0 \le -\frac{1}{M}$ on A_3, and combining this with Claim 2 we obtain that if $\frac{M+1}{2M} \le \mu_o' \le \frac{M-1}{M}$ then

$$
\begin{aligned}
(III) = \mathbb{E}[h_0 I_{R_0} \,|\, \mathcal{F}] &\le \mathbb{E}\left[\left(\sqrt{F_o + \Delta_0(\zeta', \mu_o')} - \sqrt{F_o} \right) I_{1 \le \zeta' \le \frac{M}{M-1}} \,|\, \mathcal{F} \right] + \frac{c}{M}\,\mathbb{E}[\zeta' I_{R_0} \,|\, \mathcal{F}] \\
&\le \mathbb{E}\left[\left(\sqrt{F_o - \frac{1}{M-1}} - \sqrt{F_o} \right) I_{1 \le \zeta' \le \frac{M}{M-1}} \,|\, \mathcal{F} \right] + \frac{c}{M}\,\mathbb{E}[2 I_{R_0}], \qquad (2.16)
\end{aligned}
$$

where we used the fact that $\{\frac{M+1}{2M} < \mu_o' < \frac{M-1}{M}\} \cap \{1 \leq \zeta' \leq \frac{M}{M-1}\} \subseteq \{\frac{M+1}{2M} < \mu_o' < \frac{M-1}{M}\} \cap R_0$ for the first term and that $\zeta' < 2$ on R_0 (since $\mu_0' < \frac{M-1}{M}$) for the second term. If we apply Claim 2 to the first term in (2.16) and again apply the fact that $\zeta' < 2$ on R_0 for the second term then we see that it is less or equal to

$$\begin{aligned}&\leq \left(\sqrt{F_o - \frac{1}{M-1}} - \sqrt{F_o}\right) \mathbb{P}\left(\zeta' \in \left(1, \frac{M}{M-1}\right)\right) + 2\frac{c}{M}\, \mathbb{P}\left(\zeta' \in (1,2)\right) \\ &\leq \left(-\frac{1}{2M(M-1)} + 2\frac{c}{M} C_4\right) \mathbb{P}\left(\zeta' \in \left(1, \frac{M}{M-1}\right)\right),\end{aligned}$$

where $C_4 = \frac{\mathbb{P}(1<\zeta'<2)}{\mathbb{P}\left(1<\zeta'<\frac{M}{M-1}\right)}$, which again is by bounded above by some polynomial in C according to (2.11). For this reason it is clear that we can again pick c small enough to make also this term non-positive, which proves that that $\mathbb{E}[\Delta h \,|\, \mathcal{F}] \leq 0$ and hence h_k is a non-negative supermartingale. □

Now we continue with the proof of (a) of Theorem 2. Fix k and $a := L$, and let c be defined by Lemma 6. If we denote by h_∞ the a.s. limit of $h_c(\gamma_{k,t,L})$ as $t \to \infty$ on $\{\tau_{k,L} < \infty\} \cap \{\eta_{k,L} = \infty\}$ then

$$h_\infty = \lim_{t\to\infty} \left(\sqrt{F(\tau_{k,L}+t)} + c\mu'(\tau_{k,L}+t)\right) I_{A_L} = \left(\sqrt{F_\infty} + \lim_{t\to\infty} c\mu'(t)\right) I_{A_L},$$

that is $\exists \lim_{t\to\infty} \mu'(t) \in \mathbb{R}$ on A_L, implying $\mathcal{X}'(t) \not\to +\infty$.

We will now prove (b). Notice that we have just proved that $F(t) \to 0$ a.s., and hence $\pi_{1/n} < \infty$ a.s., $\forall n > 0$. First, we will show that

$$\mathbb{P}\left(\left\{\liminf_{t\to\infty} x_{(1)}(t) > R_+\right\} \setminus \{\exists \phi :\ \mathcal{X}'(t) \to \phi\}\right) = 0. \tag{2.17}$$

Indeed, let $E_n = \left\{\liminf_{t\to\infty} x_{(1)}(t) \geq R_+ + \frac{1}{n}\right\}$, then $\left\{\liminf_{t\to\infty} x_{(1)}(t) > R_+\right\} = \bigcup_{n=1}^\infty E_n$ and it suffices to prove that $\mathbb{P}\left(E_n \setminus \{\exists \phi :\ \mathcal{X}'(t) \to \phi\}\right) = 0$. Notice that $E_n \subseteq \bigcup_{k=1}^\infty \left(\{\eta_{k,1/n} = \infty\} \cap \{\tau_{k,1/n} < \infty\}\right) \subseteq \bigcup_{k=1}^\infty \{\lim_t \gamma_{k,t,1/n} = \infty\}$. By Lemma 6 $h_c(\gamma_{k,t,1/n})$ has an a.s. limit for some $c > 0$ on $\{\eta_{k,1/n} = \infty\} \cap \{\tau_{k,1/n} < \infty\} \cap A_L$, thus

$$\mathbb{P}\left(A_L \cap \left(\{\eta_{k,1/n} = \infty\} \cap \{\tau_{k,1/n} < \infty\}\right) \setminus \{\exists \lim_{t\to\infty} \mu'(t)\}\right) = 0.$$

Using continuity of probability again, applied to the sets A_L, $L \to \infty$, we can get rid of the term A_L in the expression above. Since $F(t) \to 0$ a.s. from the first part of the theorem, we have $\{\exists \lim_{t\to\infty} \mu'(t)\} = \{\exists \phi :\ \mathcal{X}'(t) \to \phi\}$ except perhaps a set of measure zero, therefore

$$\mathbb{P}\left(E_n \setminus \{\exists \lim_{t\to\infty} \mathcal{X}'(t)\}\right) = \mathbb{P}\left(E_n \setminus \{\exists \lim_{t\to\infty} \mu'(t)\}\right) \leq$$
$$\mathbb{P}\left((\{\eta_{k,1/n} = \infty\} \cap \{\tau_{k,1/n} < \infty\}) \setminus \{\exists \lim_{t\to\infty} \mu'(t)\}\right) = 0.$$

Noting that $E_n \subseteq E_{n+1}$, (2.17) follows from continuity of probability; the proof of the respective statement for lim sup is completely analogous, and they together are equivalent to the second statement of the theorem.

We will now prove (c). Assume that $R_+ \geq 0$ and $\operatorname{supp}\zeta \subseteq [R_+, \infty)$, the case $R_- \leq 0$, $\operatorname{supp}\zeta = (-\infty, R_-]$ is analogous. If $R_+ \geq 0$ and $\operatorname{supp}\zeta \subseteq [R_+, \infty)$ then consider

$$h(t) = c\mu'(t) + \sqrt{F(t)} \geq 0,$$

for all $t \geq 0$ and some $c > 0$ to be chosen later. Notice that compared to h in Lemma 6 the restrictions to the sets A_L will no longer be necessary, neither will we need the stopping time construction introduced in the beginning of the proof of this theorem since our configuration will always stay in an area where Assumption 2 is valid. We can now make the following estimate,

$$\mathbb{E}\, h(t) \leq \mathbb{E}\left[\sqrt{F(0)} + c\left(|\mu'(0)| + \sum_{l=1}^{t} |\mu'(s) - \mu'(s-1)|\right)\right]$$
$$\leq c\,\mathbb{E}\,|\mu'(0)| + (1 + tc\sqrt{2})\,\mathbb{E}\left[\sqrt{F(0)}\right] \leq \left(c + 1 + tc\sqrt{2}(N-1)\right)\mathbb{E}\,|\zeta| < +\infty,$$

where we skipped a few steps which are analogous to those in the beginning of the proof of Lemma 6. Since the left most point of the core always lies to the right of R_+, calculations analogous to those in the proof of Lemma 6 will show that $\mathbb{E}[h(t+1)\,|\,\mathcal{F}_t] \leq h(t)$ a.s. for some $c > 0$ and we will assume that c is chosen in this way from now on. We conclude that for $t \geq 0$, then $\mathbb{E}\, h(t+1) \leq \mathbb{E}\, h(t)$ which leads to,

$$\exists \lim_{t\to\infty} \mathbb{E}\left[c\mu'(t) + \sqrt{F(t)}\right].$$

Since $\mathbb{E}\sqrt{F(0)} \leq 2(N-1)\,\mathbb{E}\,|\zeta|$ is finite by assumption on ζ and since $F(t) \to 0$ a.s. (by the previous part of this theorem), the dominated convergence theorem (since $F(t) \leq F(0)$, for $t \geq 0$) implies that $\lim_{t\to\infty} \mathbb{E}\sqrt{F(t)} = 0$. This implies that $\exists \lim_{t\to\infty} \mathbb{E}\,\mu'(t)$ (although this limit might be $+\infty$). For $1 \leq k \leq N-1$ we have $|x_{(k)}(t) - \mu'(t)| \leq D(t) \leq \sqrt{2F(t)}$ and therefore,

$$\liminf_{t\to\infty} \mathbb{E}\, x_{(k)}(t) \geq \liminf_{t\to\infty} \mathbb{E}\, \mu'(t) - \limsup_{t\to\infty} \mathbb{E}\, |x_{(k)}(t) - \mu'(t)|$$
$$\geq \lim_{t\to\infty} \mathbb{E}\, \mu'(t) - \lim_{t\to\infty} \sqrt{2F(t)} = \lim_{t\to\infty} \mathbb{E}\, \mu'(t).$$

Similarly

$$\limsup_{t\to\infty} \mathbb{E}\, x_{(k)}(t) \leq \limsup_{t\to\infty} \mathbb{E}\, \mu'(t) + \limsup_{t\to\infty} \mathbb{E}\, |x_{(k)}(t) - \mu'(t)| = \lim_{t\to\infty} \mathbb{E}\, \mu'(t),$$

and so $\lim_{t\to\infty} \mathbb{E}\, x_{(k)}(t) = \lim_{t\to\infty} \mathbb{E}\, \mu'(t)$.

We now prove (d). Assume that $R_+ < R_-$ in Assumption 2. Let $u = \liminf_{t\to\infty} x_{(1)}(t)$, $v = \limsup_{t\to\infty} x_{(N-1)}(t)$. Consider the event $A_{a,b} = \{u < a\} \cap \{v > b\}$ for some $a < b$. If $b \leq R_-$ or $a \geq R_+$ we have already showed that we have convergence, so suppose that $b > R_-$ and $a < R_+$. We now make the observation that the interval $[R_+, R_-]$ is regular with parameters $\delta = \frac{2}{3}$, $r = \frac{1}{2C}$ (see Definition 2 in the next Section) and in the event of $A_{a,b}$ we cross the interval $\left(a + \frac{b-a}{2}, b - \frac{b-a}{2}\right)$ i.o., however since this interval also inherits the regularity property, this would contradict Proposition 2 which states that a regular interval cannot be visited i.o. a.s. and so $P(A_{a,b}) = 0$. From this we can conclude that

$$\mathbb{P}\left(\{\exists \phi,\ s.t.\ \mathcal{X}'(t) \to \phi\}^c\right) = \mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q}, a<R_+, b>R_-} A_{a,b}\right) = 0,$$

i.e. the core converges to a point a.s. □

3 Convergence of the core

Definition 1. *A subset $A \subseteq \mathrm{supp}(\zeta)$ is regular with parameters $\delta_A \in (0,1), \sigma_A > 0, r_A > 0$ if*

$$\mathbb{P}(\zeta \in B_{r\delta_A}(x) \,|\, \zeta \in B_r(x)) \geq \sigma_A \tag{3.18}$$

for any $x \in A$ and $r \leq r_A$.

Assumption 3. *For any $x \in supp(\zeta)$ there exists some $\gamma = \gamma(x)$ such that the set $B_\gamma(x) \cap \mathrm{supp}(\zeta)$ is regular.*

Remark 5. *We can iterate the inequality (3.18) to establish that*

$$\mathbb{P}(\zeta \in B_{r\delta_A^k}(x) \,|\, \zeta \in B_r(x)) \geq \sigma_A^k, \quad k \geq 2.$$

Hence it is not hard to check that if Definition 1 holds for some $\delta_A \in (0,1)$ it holds for all $\delta \in (0,1)$, albeit possibly with a smaller σ_A.

Lemma 7. *Under Assumption 3, for any compact subset $A \subset \operatorname{supp}(\zeta)$ and $\delta \in (0,1)$ there exists r_A and σ_A such that A is regular with parameters δ, σ_A, r_A.*

Proof. The union $\bigcup_{x\in A} B_{\gamma(x)}(x)$ is an open covering of A, where $B_{\gamma_x}(x)$ is the regular ball centred in x given to us by Assumption 3. Since A is compact it follows that there is a finite subcover of A. In other words there exist sequences

$$\{x_k\}_{k=1}^M \subseteq A, \qquad \{\sigma_k\}_{k=1}^M, \{r_k\}_{k=1}^M, \{\delta_k\}_{k=1}^M, \{\gamma_k\}_{k=1}^M \subseteq \mathbb{R}^+$$

such that $A \subseteq \bigcup_{k=1}^M B_{\gamma_k}(x_k)$ and $\mathbb{P}(\zeta \in B_{r\delta_k}(x) \,|\, \zeta \in B_r(x)) \geq \sigma_k$ for $x \in B_{\gamma_k}(x_k)$ and $r \leq r_k$. Now let $\sigma' = \min_{1\leq k\leq M} \sigma_k$, $\delta' = \max_{1\leq k\leq M} \delta_k$,$r' = \min_{1\leq k\leq M} r_k$. It follows that for any $x \in A$

$$\mathbb{P}(\zeta \in B_{r\delta'}(x) \,|\, \zeta \in B_r(x)) \geq \sigma',$$

when $r \leq r'$. We conclude by noting that by Remark 5 there exists σ_A such that for each $x \in A$

$$\mathbb{P}(\zeta \in B_{r\delta}(x) \,|\, \zeta \in B_r(x)) \geq \sigma_A.$$

□

Theorem 3. *Under Assumptions 1 and 3*

$$\mathbb{P}\left(\{\exists \phi \in \mathbb{R}^d: \ \mathcal{X}'(t) \to \phi\} \cup \{\mathcal{X}'(t) \to \infty\}\right) = 1.$$

Proof. Firstly, $\mathbb{P}(\{\exists \lim_t \mu'(t)\}\Delta\{\exists \phi, \ s.t. \ \mathcal{X}'(t) \to \phi\}) = 0$, since if $\mu'(t)$ converges then $\mathcal{X}'(t) \not\to \infty$ which implies $D(t) \to 0$ by Theorem 1, yielding convergence of the core to the same point.

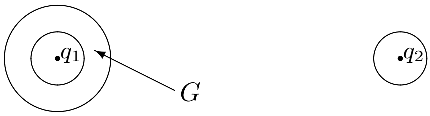
From an elementary calculus it follows that if neither the centre of mass converges to a finite point nor the configurations goes to infinity, then there must exist two arbitrarily small non-overlapping balls (w.l.o.g. centred at rational points) which are visited by μ' infinitely often, that is

$$\{\nexists \lim_t \mu'(t)\} \cap \{\mathcal{X}'(t) \not\to \infty\} = \bigcup_{n=1}^{\infty} \bigcup_{\substack{q_1,q_2\in\mathbb{Q}^d, \\ \|q_1-q_2\|\geq 6/n}} E_{q_1,q_2,n}, \tag{3.19}$$

$$\text{where } E_{q_1,q_2,n} = \left\{\mu'(t) \in B_{\frac{2}{n}}(q_1) \text{ i.o.}\right\} \bigcap \left\{\mu'(t) \in B_{\frac{2}{n}}(q_2) \text{ i.o.}\right\}.$$

To show (3.19), note that $\{\nexists \lim_t \mu'(t)\} \cap \{\mathcal{X}'(t) \not\to \infty\}$ is equivalent to existence of at least two distinct points in the set of accumulation points of $\{\mu'(t)\}_{t=1}^{\infty}$, say x_1 and x_2. Now take $q_1, q_2 \in \mathbb{Q}^d$ such that $\|q_j - x_j\| \le \frac{1}{n}$, $j = 1, 2$, then $\mu' \in B_{\frac{1}{n}}(x_j) \subseteq B_{\frac{2}{n}}(q_j)$, $j = 1, 2$, infinitely often; moreover $\|q_1 - q_2\| \ge \frac{8}{n} - \frac{1}{n} - \frac{1}{n} = \frac{6}{n}$ as required. Thus it suffices to prove that $\mathbb{P}(E_{q_1,q_2,n}) = 0$ for all $n \in \mathbb{N}$ and $q_1, q_2 \in \mathbb{Q}^d$ such that $\|q_1 - q_2\| \ge \frac{6}{n}$ to show that the LHS of (3.19) has measure zero, and then the Theorem will follow.

For simplicity w.l.o.g. assume that $q_1 = 0$ and denote $E := E_{0,q_2,n}$, $R = 2/n$, and $G = \text{supp}(\zeta) \cap (B_{2R}(0) \setminus B_R(0))$. Since every path from $B_{\frac{2}{n}}(0)$ to $B_{\frac{2}{n}}(q_2)$ must cross G, on E the shell G must be crossed infinitely often (by this we mean that $\|\mu'(t)\| > 2R$ i.o. and $\|\mu'(t)\| < R$ i.o.) – please see the illustration.



Since $\mathcal{X}'(t) \not\to \infty$ on E it follows from Theorem 1 that $F(t) \to 0$ a.s. on E and therefore additionally $\mathcal{X}'(t) \subset G$ i.o. (the core points cannot jump over the set G once the spread is sufficiently small); moreover the set G is regular by Lemma 7. We have also the following result.

Lemma 8. *Under Assumption 3, given $N - K$ points $x_1, \cdots, x_{N-K}$ in G, there are constants $a, \sigma \in (0, 1)$ depending on N, K and σ_G only, such that*

$$\mathbb{P}\left(\{F\left(\{\zeta_1, \ldots, \zeta_K, x_1, \ldots, x_{N-K}\}'\right) \le aF(\{x_1, \ldots, x_{N-K}\})\}\right) \ge \sigma.$$

(Remark the similarity of this statement with Lemma 5; the difference here, however, comes from the fact that the probability of decay σ, does not depend on the value of F, thanks to Assumption 3.)

Proof. We start with the case $K = 1$. Due to the translation invariance of F we can assume w.l.o.g. that $\sum_{i=1}^{N-1} x_i = 0$. Let $D = \max_{i,j \in \{1,\cdots,N-1\}} \|x_i - x_j\|$ and assume furthermore that $\|x_1\| \ge \|x_k\|, \forall k$ and take x_j such that $\|x_1 - x_j\| \ge \frac{D}{2}$. Let $\mu' = \frac{x_2 + \cdots + x_{N-1} + \zeta}{N-1} = \frac{\zeta - x_1}{N-1}$ and $F_{old} = F(\{x_1, \cdots, x_{N-1}\})$. If we take $\zeta \in B_{\frac{1}{8}\sqrt{\frac{F_{old}}{N}}}(x_1)$ then

$$\|\zeta - x_1\| \ge \|x_1 - x_j\| - \|\zeta - x_j\| \ge \frac{D}{2} - \frac{1}{8}\sqrt{\frac{F_{old}}{N}}.$$

From this we can deduce that $\|\zeta - x_1\|^2 \geq \frac{D^2}{8} \geq \frac{F_{old}}{4(N-1)}$. for some fixed $a \in (0,1)$ (which is only a function of N and K). By Lemma 4 the event $\left\{\zeta \notin B_{H\sqrt{2F_{old}}}(x_j)\right\}$, where $H = \sqrt{N-K-1}$, implies that $\{\zeta_1, x_1, \cdots, x_{N-1}\}' = \{x_1, \cdots, x_{N-1}\}$ (i.e. ζ is eliminated) and by Lemma 7 we can assume that δ and σ are chosen such that

$$\mathbb{P}\left(\zeta \in B_{\frac{1}{8}\sqrt{\frac{F_{old}}{N}}}(x_j) \,|\, \zeta \in B_{H\sqrt{2F_{old}}}(x_j)\right) \geq \sigma.$$

Skipping the first few steps that are identical to those in Lemma 5, we obtain the following estimate

$$F\left(\{\zeta, x_2, \cdots, x_{N-K}\}\right) = \sum_{i=2}^{N-1} \|x_i - \mu'\|^2 + \|\zeta - \mu'\|^2 \leq \left(1 - \frac{1}{4(N-1)^2}\right) F_{old}.$$

Since $F\left(\{\zeta, x_2, \cdots, x_{N-K}\}\right) < F_{old}$ one of the points $x_1, \cdots, x_{N-1}$ must be discarded. So in the case $K = 1$ we can pick $a = 1 - \frac{1}{4(N-1)^2}$. For general K one can make an argument analogue to the one made at the end of the proof of Lemma 5. □

Define for $t \geq 0$,

$$\eta(t) = \inf\{s \geq t+1 : \ \mathcal{X}'(s) \neq \mathcal{X}'(s-1) \ \text{ or } F(s) = 0\}.$$

(Notice that by definition if $F(\eta(t)) = 0$, i.e. all the points of the core have converged to a single point, then $\eta(t+1) = \eta(t) + 1$. So from now we assume that this is not the case.) Fix some large $M \geq 5$ such that

$$a^{\sigma M} \leq \frac{1}{16},$$

and define $\tau_0 = \tau_0^{(M)}$ such that

$$\mathcal{X}'(\tau_0) \subseteq B_{\frac{7}{4}R}(0) \setminus B_{\frac{5}{4}R}(0), \quad F(\tau_0) \leq \frac{R^2}{M^2\, 4^M}$$

and set also $\tau_i = \eta(\tau_{i-1}), i = 1, 2, \ldots$ (that is, the next time the core changes). Since $F(t) \to 0$ on E and we cross G infinitely often, we must visit the region $B_{\frac{7}{4}R}(0) \setminus B_{\frac{5}{4}R}(0)$ infinitely often as well, therefore $E \subseteq A_M := \{\tau_0^{(M)} < \infty\}$ for all $M \in \mathbb{N}$.

For $m \geq 0$ define

$$\begin{aligned} A'_m &= A'_{m,M} = \left\{F(\tau_{(m+M)^2}) \leq \frac{R^2}{M^2 4^{2m+M}}\right\}, \\ A''_m &= A''_{m,M} = \left\{\mathcal{X}'(\tau_{(m+M)^2}) \subseteq B_{[2-2^{-m-2}]R}(0) \setminus B_{[1+2^{-m-2}]R}(0)\right\}, \qquad (3.20) \\ A_m &= A_{m,M} = A_{m-1} \cap \left(A'_m \cap A''_m\right). \end{aligned}$$

Note that the definition is even consistent for $m = 0$ if we define $A_{-1} := \{\tau_0 < \infty\}$ and that in principle A_m, A'_m and A''_m also depend on M, but we omit the second index where this does not create a confusion.

Lemma 9. $\mathbb{P}\left(A_{m+1} \,|\, A_m\right) \geq 1 - e^{-\sigma^2(m+M)}$*,* $m = 0, 1, 2, \ldots$*.*

Proof. First, note that $A_m \subseteq A''_{m+1}$. Indeed, since $2K < N$, in the core of the new configuration we must have at least one point from the previous core (this is not true in general if $2K \geq N$), so

$$\min_{x \in \mathcal{X}'(t+1)} \|x\| \geq \min_{x \in \mathcal{X}'(t)} \|x\| - D(t+1)$$

and as a result on A_m we have

$$\begin{aligned}
\operatorname{dist}\left(\mathcal{X}'(\tau_{(m+M+1)^2}), B_R(0)\right) &= \min_{x \in \mathcal{X}'(\tau_{(m+M+1)^2})} \|x\| - R \\
&\geq \min_{x \in \mathcal{X}'(\tau_{(m+M)^2})} \|x\| - R - \sum_{t=\tau_{(m+M)^2+1}}^{\tau_{(m+M+1)^2}} D(t) \\
&\geq \min_{x \in \mathcal{X}'(\tau_{(m+M)^2})} \|x\| - R - [2(m+M)+1]\sqrt{2F(\tau_{(m+M)^2})} \\
&\geq \left(1 + \frac{1}{2^{m+2}} - 1 - \frac{2(m+M)+1}{\sqrt{M^2 4^{2m+M}}}\right) R \\
&\geq \left(\frac{1}{2^{m+2}} - \frac{1}{2^{m+3}} \frac{2(m+M)+1}{M\, 2^{M+m-3}}\right) R \geq \frac{R}{2^{m+3}}
\end{aligned}$$

since for all $j \geq 0$ we have $D(t+j) \leq \sqrt{2F(t)}$ by Lemmas 2 and 3, and $\frac{2(m+M)+1}{M\, 2^{M+m-3}} < 1$ for all $m \geq 0$ as long as $M \geq 5$. By a similar argument

$$\operatorname{dist}\left(\mathcal{X}'(\tau_{(m+M+1)^2}), (B_{2R}(0))^c\right) = 2R - \max_{x \in \mathcal{X}'(\tau_{(m+M+1)^2})} \|x\| \geq \frac{R}{2^{m+3}},$$

and hence A''_{m+1} occurs.

Consequently, when A_m occurs then $\mathcal{X}'(t) \subseteq G$ for all $t \in \left(\tau_{(m+M)^2}, \tau_{(m+1+M)^2}\right)$. At the same time the core undergoes $N = 2(m+M)+1$ changes between the times $\tau_{(m+M)^2}$ and $\tau_{(m+M+1)^2}$. During each of these changes the function F does not increase, and with probability at least σ decreases by a factor at least $a < 1$ regardless of the past, by Lemma 8 . Hence

$$\mathbb{P}\left(F(\tau_{(m+M+1)^2}) > a^{\sigma N/2} F(\tau_{(m+M)^2})\right) \leq \mathbb{P}(Y_1 + \cdots + Y_N < \sigma N/2),$$

where Y_i are i.i.d. Bernoulli(σ) random variables. It suffices now to get a bound on the RHS since $a^{\sigma N/2} \leq a^{\sigma(m+M)} \leq a^{\sigma M} \leq \frac{1}{16}$. However, the bound for the sum of Y_i follows from the Hoeffding inequality [4]:

$$\mathbb{P}(Y_1 + \cdots + Y_N < \sigma N/2) \leq \exp(-\sigma^2 N/2) \leq \exp(-\sigma^2(m+M)).$$

Consequently, A'_{m+1} and hence A_{m+1} also occur, with probability at least $\exp(-\sigma^2(m+M))$. □

Note that for a fixed M, $A_{m,M}$ is a decreasing sequence of events. Let $\bar{A}_M = \bigcap_{m=0}^{\infty} A_{m,M}$. Lemma 9 implies by induction on m that

$$\begin{aligned}\mathbb{P}\left(\bar{A}_M\right) &= \mathbb{P}\left(A_{0,M}\right) \prod_{m=1}^{\infty} \mathbb{P}\left(A_{m,M} \mid A_{m-1,M}\right) \geq \mathbb{P}\left(A_{0,M}\right) \prod_{m=1}^{\infty} \left(1 - e^{-\sigma^2(M+m)}\right) \\ &\geq \mathbb{P}(A_{0,M}) \left[1 - \sum_{m=1}^{\infty} e^{-\sigma^2(M+m)}\right] \geq \mathbb{P}(A_{0,M}) \left[1 - \sigma^{-2} e^{-\sigma^2 M}\right].\end{aligned}$$

It is easy to see that on $\bar{A}_M$ the points of the core $\mathcal{X}'(t)$ do not ever leave the set G after time τ_0, hence $\sup_{t>\tau_0} \|\mu'(t)\| < \frac{3R}{4}$ on $\bar{A}_M$. At the same time on E we must visit $B_{2/n}(q_2)$ which lies outside of the convex hull of G, thus $\sup_{t>\tau_0} \|\mu'(t)\| > \frac{3R}{4}$, therefore $E \cap \bar{A}_M = \emptyset$. Since $E \subseteq A_{0,M}$ and $\bar{A}_M \subseteq A_{0,M}$ we have

$$\mathbb{P}(E) = \mathbb{P}(E \setminus \bar{A}_M) \leq \mathbb{P}\left(A_{0,M} \setminus \bar{A}_M\right) = \mathbb{P}(A_{0,M}) - \mathbb{P}(\bar{A}_M) \leq \sigma^{-2} e^{-\sigma^2 M}\, \mathbb{P}\left(A_{0,M}\right) \leq \sigma^{-2} e^{-\sigma^2 M}$$

for any $M \geq 0$. Since M can be arbitrarily large we see that $\mathbb{P}(E) = 0$, finishing the proof. □

3.1 Convergence in $\mathbb{R}^1$

In case $d = 1$ we can obtain stronger results than for the general case $\zeta \in \mathbb{R}^d$, $d \geq 1$. For any interval $(a, b) \subset \mathbb{R}$ and any $\delta \in (0, 1)$ let us define a δ-*truncation* of (a, b) as

$$(a, b)_\delta = \left(a + \frac{\delta}{2}(b-a), b - \frac{\delta}{2}(b-a)\right).$$

Definition 2. *The interval* (a_1, b_1) *is called* regular, *if there are* $\delta, r \in (0, 1)$ *such that for any* $(a_2, b_2) \subseteq (a_1, b_1)$ *we have*

$$\mathbb{P}(\zeta \in (a_2, b_2)_\delta \mid \zeta \in (a_2, b_2)) \geq r. \tag{3.21}$$

Remark 6. *We can iterate the inequality (3.21) to establish that*

$$\mathbb{P}(\zeta \in \underbrace{(\dots(a_2, b_2)\,_\delta)\dots)_\delta}_{k\ times} \,|\, \zeta \in (a_2, b_2)) \geq r^k, \quad k \geq 2$$

and the iteration of δ-truncation eventually shrinks an interval to a point while r^k is still $\in (0,1)$. Hence it is not hard to check that if Definition 2 holds for some $\delta \in (0,1)$ it holds for all δ in this interval.

Assumption 4 ("matryoshka" property)**.** *Suppose that any interval (a,b) such that $\mathbb{P}(\zeta \in (a,b)) > 0$ contains a regular interval.*

Remark 7. *The property above seems to hold for all common distribution; we were not able, in fact, to construct even a single counterexample, nor, unfortunately, to show that none exists.*

Theorem 4. *Under Assumptions 1 and 4, $\mathcal{X}'(t) \to \phi \in [-\infty, +\infty]$ a.s.*

The proof of this theorem immediately follows from the next proposition, since if $\{\mathcal{X}'(t) \not\to \pm\infty\} = \{\mu'(t) \not\to \pm\infty\}$ occurs then $\mu'(t)$ either converges to a finite number or crosses some interval infinitely often. However, every interval contains some regular interval by Assumption 4 and by Theorem 1 $D(t) \to 0$ a.s. if $\mu'(t) \not\to \pm\infty$, so the core must converge in this case.

Proposition 2. *For any a, b such that $a < b$, with probability one $\mu'(t)$ cannot cross the interval (a,b) infinitely many times.*

Proof. Suppose the contrary. From Assumption 4 it follows that (a,b) contains some regular interval, say (a_1, b_1) which also must be crossed infinitely often. Now the rest of the proof is almost the same as that of Theorem 3 so we will only highlight the differences, which lie in how Assumption 4 is used (in place of the stronger Assumption 3) when we define our "absorbing" region G. Here we let $G = (a_1, b_1)$ and assume w.l.o.g. that $a_1 = 0, b_1 = R$. Let $\zeta(t)$ and M satisfy the conditions of Theorem 3 and define $\tau_0 = \tau_0^{(M)}$ such that

$$\mathcal{X}'(\tau_0) \subseteq \left[\frac{1}{4}R, \frac{3}{4}R\right], \quad F(\tau_0) \leq \frac{R^2}{M^2\,4^M}.$$

Let us define the events A'_m, A''_m, A_m for $m = 1, 2, \dots$ as in (3.20) with the only change that

$$A''_m = A''_{m,M} = \left\{\mathcal{X}'(\tau_{(m+M)^2}) \subseteq \left(2^{-(m+2)}R, \left[1 - 2^{-(m+2)}\right]R\right)\right\}.$$

We note that since G is regular so Lemma 8 can still be applied. The rest of the proof is identical to that of Theorem 3. □

Corollary 3. *Suppose that* $\operatorname{supp}\zeta$ *is bounded. Then under Assumptions 1 and 4 we have* $\mathcal{X}'(t) \to \phi \in \mathbb{R}$ *a.s.*

Corollary 4. *Suppose that* $K = 1$ *and that Assumption 4 is valid in some interval* $[a, b]$ *and that in addition Assumption 2 is valid for some* $R_- \geq a$ *and* $R_+ \leq b$*. Then* $\mathcal{X}'(t) \to \phi \in \mathbb{R}$ *a.s.*

Proof. Let $u = \liminf_{t\to\infty} x_{(1)}(t)$, $v = \limsup_{t\to\infty} x_{(N-1)}(t)$. Consider the event

$$A_{c,d} = \{u \leq c\} \cap \{v \geq d\}$$

for some $c < d$. If $d < R_-$ or $c > R_+$ we already know from Theorem 2 that we have convergence, so suppose that both $c, d \in [a, b]$. In this case the interval $\left(c + \frac{d-c}{2}, d - \frac{d-c}{2}\right) \subset [c, d]$ is visited i.o. but since this interval inherits the property of Assumption 4 it follows from Proposition 2 that $\mathbb{P}(A_{c,d}) = 0$. From this it follows that

$$\mathbb{P}\left(\nexists\phi : \mathcal{X}'(t) \to \phi\right) = \mathbb{P}\left(\bigcup_{c,d\in\mathbb{Q},d<b,c>a} A_{c,d}\right) = 0,$$

i.e. the core converges to a point a.s. □

4 Acknowledgement.

We would like to thank the Associate Editor and the referees for many useful suggestions and recommendations on how to improve the present paper. Research of P.K. and partly of S.V. is supported by the Swedish Research Council grant VR2014–5157.

5 Bibliography

[1] Sandemose, Aksel (1936). A fugitive crosses his tracks. translated by Eugene Gay-Tifft. New York: A. A. Knopf.

[2] Bingham, N. H.; Goldie, C. M.; Teugels, J. L. Regular variation. Encyclopedia of Mathematics and its Applications, 27. Cambridge University Press, Cambridge, (1987).

[3] Grinfeld, Michael; Volkov, Stanislav; Wade, Andrew R. Convergence in a multidimensional randomized Keynesian beauty contest. Adv. in Appl. Probab. 47 (2015), no. 1, 57–82.

[4] Hoeffding, Wassily. Probability Inequalities for Sums of Bounded Random Variables. Journal of the American Statistical Association **58**, (1963), 13–30.

[5] Parthasarathy, K. R. (2005). Probability measures on metric spaces. AMS Chelsea Publishing, Providence, RI.

B

Chapter 3

Paper B

Convergence in the p-contest

Philip Kennerberg* and Stanislav Volkov[1]

1st January 2021

[1]Centre for Mathematical Sciences, Lund University, Box 118 SE-22100, Lund, Sweden

Abstract

We study the asymptotic properties of a Markov system of $N \geq 3$ points in $[0, 1]$ in which, at each step in discrete time, the point farthest from the current centre of mass times $p > 0$ is removed and replaced by an independent ζ-distributed point; the problem was posed in [4] when $\zeta \sim U[0, 1]$. In the present paper we obtain various criteria for the convergences of the system, both for $p < 1$ and $p > 1$.

In particular, when $p < 1$ and $\zeta \sim U[0, 1]$, we show that the limiting configuration converges to zero. When $p > 1$ (except a finite set of values of p depending on N), we show that the configuration must converge to either zero or one, and we present an example where both outcomes are possible. Finally, when $p > 1$, $N = 3$ and $\zeta \sim U[0, 1]$, we prove that the configuration can only converge to one a.s.

Our paper extends the results of [3, 5] where it was assumed that $p = 1$. It turns out that one can no longer use the Lyapunov function based just on the radius of gyration; when $0 < p < 1$ one has to find a much finer tuned function which turns out to be a supermartingale; the proof of this fact constitutes a large portion of the present paper.

Keywords: Keynesian beauty contest; Jante's law, rank-driven process.

AMS 2010 Subject Classifications: 60J05 (Primary) 60D05, 60F15, 60K35, 82C22, 91A15 (Secondary)

1 Introduction

This paper extends the results of [3] and [5] on the so-called *Keynesian beauty contest*, or, as it was called in [5], *Jante's law process*. Following [3], we recall that in the Keynesian beauty contest, we have N players guessing a number, and the person who guesses closest to the mean of all the N guesses wins; see [6, Ch. 12, §V]. The formal version, suggested by Moulin [8, p. 72], assumes that this game is played by choosing numbers on the interval $[0, 1]$, the "p-beauty contest", in which the target is the mean value, multiplied by a constant $p > 0$. For the applications of the p-contest in the game theory, we refer the reader to e.g. [1]; see also [2] and [3] and references therein for further applications and other relevant papers.

The version of the p-contest with $p \equiv 1$ was studied in [3, 5]. In [3] it was shown that in the model where at each unit of time the point farthest from the center of mass is replaced by a point chosen uniformly on $[0, 1]$, then eventually all (but one) points converge almost surely to some random limit the support of which is the whole interval $[0, 1]$; many of the results were extended for the version of the model on $\mathbb{R}^d$, $d \geq 2$. The results of [3] were further generalized in [5], by removing the assumption that a new point is chosen uniformly on $[0, 1]$, as well as by removing more than one point at once, these points being chosen in such a way that the moment of inertia of the resulting configuration is minimized. However, the case $p \neq 1$ was not addressed in either of these two papers.

Let us now formally define the model; the notation will be similar to those in [3, 5]. Let $\mathcal{X} = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^N$ be an unordered N-tuple of points in $\mathbb{R}$, and $(x_{(1)}, x_{(2)}, \dots, x_{(N)})$ be these points put in non-decreasing order, that is, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$. As in [3, 5] let us define the barycentre of the configuration as

$$\mu_N(x_1, \dots, x_N) := N^{-1} \sum_{i=1}^{N} x_i.$$

Fix some $p > 0$ and also define the $p-$*centre of mass* as $p\mu_N(x_1, \dots, x_N)$.

The point, farthest from the $p-$centre of mass, is called the *extreme* point of $\mathcal{X}$, and it can be either $x_{(1)}$ or $x_{(N)}$ (with possibility of a tie), and *the core* of $\mathcal{X}$, denoted by $\mathcal{X}'$, is constructed from $\mathcal{X}$ by removing the extreme point; in case of a tie between the left-most and the right-most point, we choose either of them with equal probability (same as in [3, 5]). Throughout the rest of the paper, $x_{(1)}(t), \dots, x_{(N-1)}(t)$ shall denote the points of the core[1] $\mathcal{X}'(t)$ put into non-decreasing order.

Our process runs as follows. Let $\mathcal{X}(t) = \{X_1(t), \dots, X_N(t)\}$ be an unordered N-tuple of points in $\mathbb{R}$ at time $t = 0, 1, 2, \dots$. Given $\mathcal{X}(t)$, let $\mathcal{X}'(t)$ be the core of $\mathcal{X}(t)$ and replace $\mathcal{X}(t) \setminus \mathcal{X}'(t)$ by a ζ-distributed random variable so that

$$\mathcal{X}(t+1) = \mathcal{X}'(t) \cup \{\zeta_{t+1}\},$$

where ζ_t, $t = 1, 2, \dots$, are i.i.d. random variables with a common distribution ζ.

Finally, to finish the specification of our process, we allow the initial configuration $\mathcal{X}(0)$ to be arbitrary or random, with the only requirement being that all the points of $\mathcal{X}(0)$ must lie in the support of ζ.

[1] rather than of $\mathcal{X}(t)$

Throughout the paper we will use the notation $A \underset{\text{a.s.}}{\Longrightarrow} B$ for two events A and B, whenever $\mathbb{P}(A \cap B^c) = 0$, that is, when $A \subseteq B$ up to a set of measure 0. We will also write, with some abuse of notations, that $\lim_{t\to\infty} \mathcal{X}'(t) = a \in \mathbb{R}$ or equivalently $\mathcal{X}'(t) \to a$ as $t \to \infty$ if $\mathcal{X}'(t) \to (a, a, \dots, a) \in \mathbb{R}^{N-1}$, i.e. $\lim_{t\to\infty} x_{(i)}(t) = a$ for all $i = 1, 2, \dots, N-1$. Similarly, for an interval (a, b) we will write $\mathcal{X}'(t) \in (a, b)$ whenever *all* $x_{(1)}(t), \dots, x_{(N-1)}(t) \in (a, b)$. Finally, we will assume that $\inf \emptyset = +\infty$, and use the notation $y^+ = \max(y, 0)$ for $y \in \mathbb{R}$.

Also we require that ζ has a *full support* on $[0, 1]$, that is, $\mathbb{P}(\zeta \in (a, b)) > 0$ for all a, b such that $0 \le a < b \le 1$.

2 The case $p < 1$

Throughout this Section we assume that $0 < p < 1$ and that $\operatorname{supp} \zeta = [0, 1]$. Because of the scaling invariance, our results may be trivially extended to the case when $\operatorname{supp} \zeta = [0, A]$, $A \in (0, \infty)$; some of them are even true when $A = \infty$; however, to simplify the presentation from now on we will deal only with the case $A = 1$.

First, we present some general statements; more precise results will follow in case where $\zeta \sim U[0, 1]$.

Proposition 1. *We have*

(a) $\liminf_{t\to\infty} x_{(N-1)}(t) = 0$*;*

(b) $\mathbb{P}\left(\exists \lim_{t\to\infty} \mathcal{X}'(t) \in (0, 1]\right) = 0$*;*

(c) if $p < \frac{1}{2} + \frac{1}{2(N-1)}$ *then* $\mathbb{P}\left(\lim_{t\to\infty} \mathcal{X}'(t) = 0\right) = 1$*;*

(d) if $p < \frac{1}{2} + \frac{1}{N-2}$ *then* $\{x_{(1)}(t) \to 0\} \underset{a.s.}{\Longrightarrow} \{\lim_{t\to\infty} \mathcal{X}'(t) = 0\}$.

Proof. (a) Since ζ has full support on $[0, 1]$ it follows that (see [5], Proposition 1) there exists a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\mathbb{P}(\zeta \in (a, b)) \ge f(b - a) > 0 \qquad \text{for all } 0 \le a < b \le 1. \tag{2.1}$$

Also, to simplify notations, we write $\mu = \mu_N(\mathcal{X}(t))$ throughout the proof.

Fix a small positive ε such that $p+2\varepsilon < 1$. Suppose that for some t we have $x_{(N-1)}(t) \leq b \leq 1$. We will show that $x_{(N-1)}(t+N) \leq b(1-\varepsilon)$ with a strictly positive probability which only depends on p, b, ϵ and N. Assume that we have $\zeta_{t+1}, \dots, \zeta_{t+N-1} \in (pb, (p+\varepsilon)b) \subset (pb, b)$; this happens with probability no less than $[f(p\varepsilon b)]^{N-1}$. We claim that by the time $t+N$ we have $x_{(N-1)}(t+N-1) < (p+\varepsilon)b$. Indeed, $p\mu \leq pb$ always lies to the left of the newly sampled points, therefore either there are no more points to the right of $(p+\varepsilon)b$ at some time $s \in [t, t+N-1]$ (which implies that there will be no points there at time $t+N$ due to the sampling range of the new points), or one of the older points, i.e. present at time t, gets removed (it can be the one to the left of pb). Since we eventually have to replace all the $N-1$ old points, then $x_{(N-1)}(t+N) \leq b(1-\varepsilon)$.

Fix a $\delta > 0$ and find M so large that $(1-\varepsilon)^M < \delta$. Let the event $C(s) = \{x_{(N-1)}(s) < \delta\}$. By iterating the above argument, we get that $\mathbb{P}(C(t+NM)|\mathcal{F}_t) \geq \prod_{i=1}^{M} \left[f(p\varepsilon(1-\varepsilon)^{i-1})\right]^{N-1}$, since at time t we can set $b = 1$. Therefore, $\sum_m \mathbb{P}(C(NM(m+1))|\mathcal{F}_{NMm}) = \infty$ and by Lévy's extension of the Borel-Cantelli lemma (see e.g. [7]) infinitely many $C(s)$ occur. Since $\delta > 0$ is arbitrary, we get $\liminf_{t\to\infty} x_{(N-1)}(t) = 0$.

(b) Let $r = \frac{1+p^{-1}}{2} > 1$. Suppose that the core converges to some point $x \in (0,1]$; then there exist a rational $q \in (0,1]$ and a $T > 0$ such that $\mathcal{X}'(t) \in (q, rq)$ for all $t \geq T$, formally

$$\{\exists \lim \mathcal{X}'(t) \in (0,1]\} \subseteq \bigcup_{q\in Q\cap(0,1]} \bigcup_{T>0} \bigcap_{t\geq T} A_{q,t} \tag{2.2}$$

where $A_{q,t} = \{\mathcal{X}'(t) \in (q, rq)\}$. We will show that

$$\mathbb{P}(A_{q,t+1}|\mathcal{F}_t, A_{q,t}) < 1 - \nu_q \quad \text{for all } t$$

for some $\nu_q > 0$. This will imply, in turn, that

$$\mathbb{P}\left(\bigcap_{t\geq T} A_{q,t}\right) = 0$$

and hence the RHS (and thus the LHS as well) of (2.2) has the probability 0.

Suppose $A_{q,t}$ has occurred and the newly sampled point $\zeta \in (pq, q)$. Then

$$p\mu_N(\mathcal{X}'(\tau_k) \cup \{\zeta\}) < prq = \frac{pq+q}{2} < \frac{\zeta + x_{(N-1)}}{2}$$

Consequently, $x_{(N-1)}$ lies further from the $p-$center of mass, and hence it should be removed. The new configuration will, however, contain the point $\zeta \notin (q, rq)$ and hence $A_{q,t+1}$ does not occur. Thus

$$\mathbb{P}\left(A_{q,t+1}|\mathcal{F}_t, A_{q,t}\right) \leq 1 - \mathbb{P}\left(\zeta \in (q, rq)\right) \leq 1 - f(pq - q) =: 1 - \nu_q$$

as required.

(c) First, we will show that it is the right-most point of the configuration which should be always removed; note that it suffices to check this only when $x_{(N)} > 0$. Indeed, by the assumption on p we have

$$\mu \leq \frac{(N-1)x_{(1)} + (N-1)x_{(N)}}{N} = \frac{2p(N-1)}{N} \cdot \frac{x_{(1)} + x_{(N)}}{2p} < \frac{x_{(1)} + x_{(N)}}{2p}$$

implying that

$$x_{(N)} - p\mu > p\mu - x_{(1)} \Longleftrightarrow x_{(N)} - p\mu > |p\mu - x_{(1)}|$$

Therefore, $x_{(N)}$ is the farthest point from the $p-$centre of mass. This implies that $x_{(N-1)}(t)$ is non-increasing and therefore result now easily follows from part (a) since $x_{(N-1)}(t)$ is an upper bound for all the core points.

(d) Apply Corollary 3 in the appendix with $k = 1$; this is possible because of Remark 6.

□

We are ready to present the main result of this Section.

Theorem 1. *Suppose that $\zeta \sim U[0, 1]$. Then $\mathcal{X}'(t) \to 0$ a.s.*

Proof. Proposition 1 (c) implies that we now only need to consider the case $p \geq \frac{N}{2(N-1)}$, which we will assume from now on.

Let us introduce a modification of this process on $[0, +\infty)$ which we will call the *borderless p-contest*; it is essentially the same process as the one in Section 3.4 of [3]. In order to do this, we need the following statement.

Lemma 1. *Suppose that $x_1, \ldots, x_{N-1} > 0$. Then there exists an $R = R(x_{(N-1)}) \geq 0$ such that x is the farthest point from $p\mu = \frac{p}{N}(x_1 + \cdots + x_{N-1} + x)$ whenever $x > R$.*

Proof of Lemma 1. Set $R = 6x_{(N-1)}$. Then $x > x_{(1)}$ is farther from the centre of mass than $x_{(1)}$ if and only if

$$x - p\mu > |p\mu - x_{(1)}| \Longleftrightarrow x - p\mu > p\mu - x_{(1)} \Longleftrightarrow x\left(1 - \frac{2p}{N}\right) > 2p\frac{x_1 + \cdots + x_{N-1}}{N} - x_{(1)}$$

This is true, due to the fact that $x > R$ and

$$x\left(1 - \frac{2p}{N}\right) > \frac{x}{3} > 2x_{(N-1)} > 2px_{(N-1)} > 2p\frac{x_1 + \cdots + x_{N-1}}{N}$$

since $p < 1$ and $N \geq 3$. $\square$

The borderless process is constructed as follows. Our core configuration starts as before in $[0, 1]$, and we use the same rejection/acceptance criteria for new points. However, we will now allow points to be generated to the right of 1 as well. Let $R_t = R(x_{(N-1)}(t))$ where R is taken from Lemma 1. Then a new point is sampled uniformly and independently of the past on the interval $[0, R_t]$; formally, it is given by $R_t U_t$ where U_t are i.i.d. uniform $[0, 1]$ random variables independent of everything. Observe that if we consider the embedded process only at the times when the core configuration changes, then the exact form of the function $R(\cdot)$ is irrelevant, due to the fact that the uniform distribution conditioned on a subinterval is also uniform on that subinterval.

Next, for $y = \{y_1, \ldots, y_{N-1}\}$ define the function

$$h(y) = F(y) + k\mu(y)^2, \tag{2.3}$$

where

$$F(y) = \sum_{i=1}^{N-1} (y_i - \mu(y))^2, \quad \mu(y) = \frac{1}{N-1}\sum_{i=1}^{N-1} y_i, \quad k = \frac{(N-1)^2(1-p)}{N-2}.$$

We continue with the following

Lemma 2. *For the borderless p-contest the sequence of random variables $h(\mathcal{X}'(t)) \geq 0$, $t = 1, 2, \ldots,$ is a supermartingale.*

Remark 1. *Note that the function $F(\cdot)$ defined above is a Lyapunov function for the process in [3]; this is no longer the case as long as $p \neq 1$; that is why we have to use a carefully chosen "correction" factor which involves the barycentre of the configuration.*

Proof of Lemma 2. Assume that $x_{(N-1)}(t) > 0$ (otherwise the process has already stopped, and the result is trivial). The inequality, which we want to obtain is

$$\mathbb{E}[h(\mathcal{X}'(t+1)) - h(\mathcal{X}'(t))|\mathcal{F}_t]|_{x(t)=y} \le 0$$

for all $y = (y_1, \dots, y_{N-1})$ with $y_i \in [0,1]$. Note that the function $h(y)$ is homogeneous of degree 2 in y, therefore w.l.o.g. we can assume that $\max y \equiv 1$.

For simplicity let $M = N - 1 \ge 2$, and let

$$z = 6U_t \text{ (the newly sampled point)}, \quad a = \min y < 1 \text{ (the leftmost point)}$$

Note also that

$$p \ge \frac{N}{2(N-1)} = \frac{M+1}{2M} = \frac{1}{2} + \frac{1}{2M}. \tag{2.4}$$

Define

$$\begin{aligned} F_{old} &= F(y), & F_{new} &= F\left((y \cup \{z\})'\right) \\ \mu'_{old} &= \mu(y), & \mu'_{new} &= \mu\left((y \cup \{z\})'\right), \\ h_{old} &= F_{old} + k\left(\mu'_{old}\right)^2, & h_{new} &= F_{new} + k\left(\mu'_{new}\right)^2 \end{aligned}$$

Thus we need to establish

$$\mathbb{E}[h_{new} - h_{old}|\mathcal{F}_t] \le 0. \tag{2.5}$$

First of all, observe that if $\tilde{y} = (y \setminus \{y_i\}) \cup \{z\}$, that is, $\tilde{y}$ is obtained from y by replacing y_i with y_0, then

$$\begin{aligned} F(\tilde{y}) - F(y) &= \frac{z - y_i}{M}\left[(M-1)z + (M+1)y_i - 2M\mu(y)\right] \\ \mu(\tilde{y})^2 - \mu(y)^2 &= \frac{z - y_i}{M^2}\left[z - y_i + 2M\mu(y)\right] \end{aligned}$$

In particular, if we replace point a by the new point z, then

$$\Delta_a(z) := h_{new} - h_{old} = \frac{z-a}{M}\left[(M-1)z + (M+1)a - 2M\mu(y) + \frac{k}{M}(z - a + 2M\mu(y))\right]$$

and if we replace point 1, then

$$\Delta_1(z) := h_{new} - h_{old} = \frac{z-1}{M}\left[(M-1)z + (M+1) - 2M\mu(y) + \frac{k}{M}(z - 1 + 2M\mu(y))\right]$$

Note that both Δ_a and Δ_1 depend only on four variables (a, z, μ, M) but not the whole configuration. Let us also define

$$m(z) = p \cdot \frac{y_1 + \dots + y_M + z}{M+1} = p \cdot \frac{M\mu + z}{M+1},$$

the $p-$centre of mass of the old core and the newly sampled point.

There are three different cases that can occur: either (a) the point a is removed, (b) 1, the rightmost point of the previous core, is removed, or (c) the newly sampled point z is removed. In the third case the core remains unchanged, and the change in the value of the function h is trivially zero. The point a can only be removed if $z > a$; the point 1 can only be removed if $z < 1$; the point z can be possibly removed only if $z \in (0, a)$ or $z \in (1, \infty)$. Let us compute the critical values for z, for which there is a tie between the farthest points.

Which point to remove?

(*i*) Suppose $\boxed{z < a}$. Then there is a tie between z and 1 if and only if $m(z) = \frac{z+1}{2}$, that is if

$$z = t_{z1} := \frac{M(2p\mu - 1) - 1}{M + 1 - 2p} \in \begin{cases} (-\infty, 0) & \text{if } p < p_1 := \frac{M+1}{2M\mu} \\ (0, a) & \text{if } p_1 < p < p_2 := \frac{(M+1)(a+1)}{2M\mu+2a} \\ (a, +\infty) & \text{if } p > p_2. \end{cases}$$

Thus, we have:

- when $p < p_1$, point 1 is removed;
- when $p_1 < p < p_2$, if $z < t_{z1}$ then z is removed; if $z > t_{z1}$ point 1 is removed;
- when $p > p_2$, point z is removed.

(*ii*) Suppose $\boxed{a < z < 1}$. There is a tie between a and 1 if and only if $m(z) = \frac{a+1}{2}$, that is if

$$z = t_{a1} := \frac{(M + 1)(a + 1) - 2M\mu p}{2p} \in \begin{cases} (1, +\infty) & \text{if } p < p_3 := \frac{(M+1)(a+1)}{2M\mu+2}, \\ (a, 1) & \text{if } p_3 < p < p_2, \\ (-\infty, a) & \text{if } p > p_2. \end{cases}$$

Thus, we have:

- when $p < p_3$, point 1 is removed;
- when $p_3 < p < p_2$, if $z < t_{a1}$ then 1 is removed; if $z > t_{a1}$ then point a is removed;

- when $p > p_2$, point a is removed.

(iii) Suppose $\boxed{z > 1}$. There is a tie between z and a if and only if $m(z) = \frac{z+a}{2}$, that is if

$$z = t_{za} := \frac{2M\mu p - (M+1)a}{M+1-2p} \in \begin{cases} (-\infty, 1) & \text{if } p < p_3, \\ (1, +\infty) & \text{if } p > p_3. \end{cases}$$

Thus, we have:

- when $p < p_3$, point z is removed;
- when $p > p_3$, if $z < t_{za}$ then a is removed; if $z > t_{za}$ then point z is removed.

We always have $p_1 < p_2$, $p_3 < p_2$ since

$$p_2 - p_1 = \frac{a(M+1)(M\mu - 1)}{2M\mu(M\mu + a)} = \frac{a(M+1)(a + (M-2)f)}{2M\mu(M\mu + a)} > 0,$$
$$p_2 - p_3 = \frac{(1-a)^2(M+1)}{2(M\mu + 1)(M\mu + a)} > 0,$$

while

$$p_1 < p_3 \iff Ma\mu > 1 \iff f > \frac{1 - a - a^2(M-1)}{a(M-2)(1-a)} \text{ (when } M > 2)$$

The final observation is that $t_{za} < 6$, so there is indeed no need to sample the new point outside of the range $(0, 6)$; this holds since $M \geq 2$ and

$$6 - t_{za} = \frac{-2p(M\mu + 6) + Ma + 6M + a + 6}{M+1-2p} > \frac{-2M\mu + Ma + 6M + a - 6}{M+1-2p}$$
$$> \frac{-2M\mu + 6M - 6}{M+1-2p} = \frac{2M(1-\mu) + 4M - 6}{M+1-2p} > \frac{2}{M+1-2p} > 0.$$

The five cases for the removal:

- $p < \min\{p_1, p_3\}$:
  - when $z < 1$, point 1 is removed
  - when $z > 1$, point z is removed
- $p > p_2$:
  - when $z < a$ or $z > t_{za} \in (1, \infty)$ point z is removed

- when $a < z < t_{za}$, point a is removed

- $\max\{p_1, p_3\} < p < p_2$
  - when $z < t_{z1} \in (0, a)$ or $t > t_{za} \in (1, +\infty)$, point z is removed
  - when $t_{z1} < z < t_{a1} \in (a, 1)$, point 1 is removed
  - when $t_{a1} < z < t_{za}$, point a is removed

- $p_1 < p < p_3$ $(< p_2)$:
  - when $z < t_{z1} \in (0, a)$ or $z > 1$, point z is removed
  - when $t_{z1} < z < 1$, point 1 is removed

- $p_3 < p < p_1$ $(< p_2)$:
  - when $z < t_{a1} \in (a, 1)$, point 1 is removed
  - when $t_{a1} < z < t_{za} \in (1, +\infty)$, point a is removed
  - when $z > t_{za}$, point z is removed

Let

$$\begin{aligned} X_1 &= p - p_1 = \frac{M(2\mu p - 1) - 1}{2M\mu}, \\ X_2 &= p - p_2 = \frac{2ap - a - 1 + (2\mu p - a - 1)M}{2(M\mu + a)}, \\ X_3 &= p - p_3 = \frac{2p - a - 1 + (2\mu p - a - 1)M}{2(M\mu + 1)}. \end{aligned}$$

Define

$$\begin{aligned} \tilde{\mathbf{I}}_1 &= \mathbb{E}(h(\mathcal{X}'(t+1)) - h(\mathcal{X}'(t))|\mathcal{F}_t)|_{x(t)=y} \cdot 1_{X_1<0} \cdot 1_{X_3<0}, \\ \tilde{\mathbf{I}}_2 &= \mathbb{E}(h(\mathcal{X}'(t+1)) - h(\mathcal{X}'(t))|\mathcal{F}_t)|_{x(t)=y} \cdot 1_{X_2>0}, \\ \tilde{\mathbf{I}}_3 &= \mathbb{E}(h(\mathcal{X}'(t+1)) - h(\mathcal{X}'(t))|\mathcal{F}_t)|_{x(t)=y} \cdot 1_{X_2<0} \cdot 1_{X_1>0} \cdot 1_{X_3>0}, \\ \tilde{\mathbf{I}}_4 &= \mathbb{E}(h(\mathcal{X}'(t+1)) - h(\mathcal{X}'(t))|\mathcal{F}_t)|_{x(t)=y} \cdot 1_{X_1>0} \cdot 1_{X_3<0}, \\ \tilde{\mathbf{I}}_5 &= \mathbb{E}(h(\mathcal{X}'(t+1)) - h(\mathcal{X}'(t))|\mathcal{F}_t)|_{x(t)=y} \cdot 1_{X_1<0} \cdot 1_{X_3>0}. \end{aligned}$$

Since $\max y = 1$, because of the comment on the restriction of the uniform distribution on a subinterval, we have $\tilde{\mathbf{I}}_j = c_j \mathbf{I}_j$, $j = 1, 2, 3, 4, 5$, where c_j's are some positive constants and

$$\begin{aligned}
\mathbf{I}_1 &= \mathbf{A}_1 \cdot 1_{X_1<0} \cdot 1_{X_3<0}, & \mathbf{A}_1 &= \int_0^1 \Delta_1 dz,\\
\mathbf{I}_2 &= \mathbf{A}_2 \cdot 1_{X_2>0}, & \mathbf{A}_2 &= \int_a^{t_{za}} \Delta_a dz,\\
\mathbf{I}_3 &= \mathbf{A}_3 \cdot 1_{X_2<0} \cdot 1_{X_1>0} \cdot 1_{X_3>0}, & \mathbf{A}_3 &= \int_{t_{z1}}^{t_{a1}} \Delta_1 dz + \int_{t_{a1}}^{t_{za}} \Delta_a dz,\\
\mathbf{I}_4 &= \mathbf{A}_4 \cdot 1_{X_1>0} \cdot 1_{X_3<0}, & \mathbf{A}_4 &= \int_{t_{z1}}^{1} \Delta_1 dz,\\
\mathbf{I}_5 &= \mathbf{A}_5 \cdot 1_{X_1<0} \cdot 1_{X_3>0}, & \mathbf{A}_5 &= \int_0^{t_{a1}} \Delta_1 dz + \int_{t_{a1}}^{t_{za}} \Delta_a dz.
\end{aligned}$$

Thus to establish (2.5), it suffices to show that $\mathbf{I}_j \le 0$ for each $j = 1, 2, 3, 4, 5$. This is done by very extensive and tedious calculations, which can be found in the appendix. □

We now return to our original p-contest process $\mathcal{X}(t)$. For $L \ge 2$ define

$$\tau_L = \inf\{t > 0 : \ x_{(N-1)}(t) < 1/L\};$$
$$\eta_L = \inf\{t > \tau_L : \ x_{(N-1)}(t) \ge 1/2\},$$

note that τ_L is a.s. finite for every L by Proposition 1. Let $W(s) = \{w_1(s), \dots, w_N(s)\}$ be a borderless p-contest with $W(0) = \mathcal{X}(\tau_L)$; let $W'(s)$ be its core. By Lemma 2 the quantity $\xi_t = h(W'(t \wedge \eta_L))$ is a supermartingale, that converges to some ξ_∞. Since ξ_t is bounded,

$$\mathbb{E}\xi_0 \ge \mathbb{E}\xi_\infty = \mathbb{E}[\xi_\infty \cdot 1_{\eta_L<\infty}] + \mathbb{E}[\xi_\infty \cdot 1_{\eta_L=\infty}] \ge \mathbb{E}[\xi_\infty \cdot 1_{\eta_L<\infty}] \ge \frac{k}{(2(N-1))^2}\mathbb{P}(\eta_L < \infty)$$

since on $\{\eta_L < \infty\}$ we have $\xi_\infty = W'(\eta_L)$ and the largest coordinate of $W'(\eta_L)$ is larger than $1/2$, implying that $\mu(W'(\eta_L)) \ge \frac{1}{2(N-1)}$ and thus $h(W'(\eta_L)) = F(W'(\eta_L)) + k\mu(W'(\eta_L))^2 \ge \frac{k}{(2(N-1))^2}$. We also have

$$\xi_0 = h(\mathcal{X}'(\tau_L)) = F(\mathcal{X}'(\tau_L)) + k\mu(\mathcal{X}'(\tau_L))^2 \le \frac{N-1}{L^2} + \frac{k}{L^2} \quad \Longrightarrow \quad \mathbb{E}\xi_0 \le \frac{N+k-1}{L^2}$$

since $\mathcal{X}'(\tau_L) \subset [0, 1/L]$ and so $\mu(\mathcal{X}'(\tau_L)) \in [0, 1/L]$.

Combining the above inequalities, we conclude that $\mathbb{P}(\eta_L < \infty) \to 0$ as $L \to \infty$. However, on $\eta_L = \infty$ the core of the regular p-contest process can be trivially coupled with the core of the borderless process $W'(s)$ which converges to zero, so $\mathcal{X}'(t) \to 0$ as well. Since $\mathbb{P}(\eta_L = \infty)$ can be made arbitrarily close to 1 by choosing a large L, we conclude that $\mathcal{X}'(t) \to 0$ a.s. □

3 The case $p > 1$

Throughout this section we suppose that ζ has a full support on $[0,1]$, and, unless explicitly stated otherwise, that $p > 1$.

Theorem 2. *(a)* $\mathbb{P}\left(\{\mathcal{X}'(t) \to 0\} \cup \{\mathcal{X}'(t) \to 1\}\right) = 1$;

(b) if $x_{(1)}(0) \geq 1/p$ *then* $\mathbb{P}(\mathcal{X}'(t) \to 1) = 1$*;*

(c) if $x_{(k)}(0) > 0$, *where* k *satisfies*

$$\{2p(N-k) > N - 2p\} \Longleftrightarrow \left\{k < N - \frac{N}{2p} + 1\right\}, \tag{3.6}$$

then $\mathbb{P}\left(\mathcal{X}'(t) \to 1\right) > 0$.

Remark 2. *In general, both convergences can have a positive probability. Let* $N = 3$, $p \in (1, 3/2)$, *and*

$$\zeta = \begin{cases} U, & \text{with probability } 1/3; \\ 0, & \text{with probability } 1/3; \\ 1, & \text{with probability } 1/3, \end{cases}$$

where $U \in U[0,1]$ *(so* ζ *has full support). Suppose we sample the points of* $\mathcal{X}(0)$ *from* ζ*. If they all start off in* 0*, then* $p\mu \leq p/3 < 1/2$*, so they cannot escape from* 0*. On the other hand, there is a positive probability they all start in* $(1/p, 1]$*, and then Theorem 2(b) says that they converge to* 1*.*

The key idea behind the proof of Theorem 2 is that one can actually find the "ruling" order statistic of the core; namely, there exists some non-random $k = k(N,p) \in \{1, 2, \dots, N-1\}$ such that $x_{(k)}(t) \to 0$ implies $\mathcal{X}'(t) \xrightarrow{\text{a.s.}} 0$, while $x_{(k)}(t) \not\to 0$ implies that $\mathcal{X}'(t) \xrightarrow{\text{a.s.}} 1$.

Proof. We start with the following two results, which tells us that there is *an absorbing area* $[\frac{1}{p}, 1]$ for the process, such that, once the core enters this area, it will never leave it, and moreover the core will keep moving to the right.

Claim 1. *Suppose that* $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_N \leq 1$ *and* $x_2 \geq p^{-1}$*. Then* $\{x_1, \cdots, x_N\}' = \{x_2, \cdots, x_N\}$

Proof. Let $\mu = \frac{x_1 + \cdots x_N}{N}$. If $p\mu \geq x_N$ then the claim follows immediately; assume instead that $p\mu < x_N$. We need to check if $p\mu - x_1 > x_N - p\mu$, that is, if

$$2p(x_2 + \cdots + x_{N-1}) > (N - 2p)(x_1 + x_N) \tag{3.7}$$

However, since $x_i \geq x_2$ for $i = 3, \ldots, N-1$ we have

$$2p(x_2 + \cdots + x_{N-1}) \geq 2px_2(N-2) \geq 2(N-2)$$

while $(N - 2p)(x_1 + x_N) \leq 2(N - 2p) < 2(N - 2)$. Hence (3.7) follows. □

Lemma 3. *If $x_{(1)}(t_0) \geq 1/p$ for some t_0, then $\mathcal{X}'(t) \to 1$ a.s.*

Proof. If $x_{(1)}(t_0) \geq 1/p$, then any point that lands in $[0, 1/p)$ is extreme, so $x_{(2)}(t) \geq 1/p$ for all $t \geq t_0$. Choose any positive $\varepsilon < 1 - \frac{1}{p}$, and let $A_t = \{\zeta_{t+1}, \ldots, \zeta_{t+N-1} \in (1-\varepsilon, 1]\}$. Then if A_t happens for $s > t_0$, any point in $[0, 1-\varepsilon]$ is removed in preference to any of the new points coming in, so $x_{(2)}(s + N - 1) > 1 - \varepsilon$. As a result, by Claim 1 we get that $\mathcal{X}'(t) \in [0, 1-\varepsilon]$ for all $t \geq s$.

On the other hand, $\mathbb{P}(A_t) \geq [f(\varepsilon)]^{N-1} > 0$ (see (2.1)) for any t, and the events $A_t, A_{t+N}, A_{t+2N}, \ldots$ are independent. Hence, eventually with probability 1, one of the A_t's must happen for some $t > t_0$, so a.s. $\mathcal{X}'(t) \in [0, 1-\varepsilon]$ for all large t. Since ε can be chosen arbitrary small, we get the result. □

The next two results show that if the is some $\varepsilon > 0$ such that infinitely often the core does not have any points in $[0, \varepsilon)$, then it must, in fact, converge to 1.

Lemma 4. *If $x_{(1)}(t_0) \geq \varepsilon$ for some t_0 and $\varepsilon > 0$, then $\mathbb{P}(x_{(1)}(t_0 + \ell) \geq p^{-1} | \mathcal{F}_t) \geq \delta$ for some $\ell = \ell(\varepsilon)$ and $\delta = \delta(\varepsilon) > 0$.*

Proof. Suppose that for some t we have $x_{(1)}(t) \geq \varepsilon$. We claim that it is possible to move $x_{(1)}$ to the right of $\frac{1+p}{2}\varepsilon$ in at most $N-1$ steps with positive probability, depending only on p and ε. Indeed, if $x_{(1)}(t) > \frac{1+p}{2}\varepsilon$ then we are already done. Otherwise, if the new point ζ_{t+1} is sampled in $\left(\frac{1+p}{2}\varepsilon, p\varepsilon\right] \subset [0,1]$ it cannot be rejected. If at this stage $x_{(1)}(t+1) > \frac{1+p}{2}\varepsilon$, then we are done. If not, we proceed again by sampling $\zeta_{t+2} \in \left(\frac{1+p}{2}\varepsilon, p\varepsilon\right]$, etc. After at most $N-1$ steps of sampling new points in $\left(\frac{1+p}{2}\varepsilon, p\varepsilon\right]$, the leftmost point $x_{(1)}$ will have moved to the right of $\frac{1+p}{2}\varepsilon$.

Thus, in no more than $N-1$ steps, with probability no less than $\left[f\left(\frac{p-1}{2}\varepsilon\right)\right]^{N-1} > 0$, $x_{(1)}$ is to the right of $\frac{1+p}{2}\varepsilon$. By iterating this argument at most m times, where $m \in \mathbb{N}$

is chosen such that $\left[\frac{1+p}{2}\right]^m \varepsilon > 1/p$, we achieve that $x_{(1)}$ is to the right of $1/p$ (for definiteness, one can chose $\ell = (N-1)m$ and $\delta = \left[f\left(\frac{p-1}{2}\varepsilon\right)\right]^{(N-1)m}$.) $\square$

Lemma 5. *Let $\varepsilon \in (0,1)$, and define $B(\varepsilon) := \{x_{(1)}(t) \geq \varepsilon \text{ i.o.}\}$ Then $B(\varepsilon) \underset{a.s.}{\Longrightarrow} \{\mathcal{X}'(t) \to 1\}$.*

Corollary 1. *We have $\left\{\liminf_{t\to\infty} x_{(1)}(t) > 0\right\} \underset{a.s.}{\Longrightarrow} \{\mathcal{X}'(t) \to 1\}$.*

Proof of Lemma 5. Assume that $\varepsilon < \frac{1}{p}$ (otherwise the result immediately follows from Lemma 3). Also suppose that $\mathbb{P}(B(\varepsilon)) > 0$, since otherwise the result is trivial. Let ℓ and δ be the quantities from Lemma 4.

Define

$$\tau_0 = \inf\{t > 0 : \ x_{(1)}(t) > \varepsilon\},$$
$$\tau_k = \inf\{t > \tau_{k-1} + \ell : \ x_{(1)}(t) > \varepsilon\}, \quad k \geq 1,$$

with the convention that if $\tau_k = \infty$ then $\tau_m = \infty$ for all $m > k$. Notice that $B(\varepsilon) = \bigcap_{k=0}^{\infty}\{\tau_k < \infty\}$. On $B(\varepsilon)$ we can also define $D_{\tau_k} = \{x_{(1)}(\tau_k + \ell) \geq 1/p\}$. Since $\tau_k - \tau_{k-1} > \ell$ whenever both are finite, we have from Lemma 4 we have $\mathbb{P}(D_{\tau_{k+1}} | \mathcal{F}_{\tau_k}) \geq \delta$. Therefore,

$$B(\varepsilon) \underset{\text{a.s.}}{\Longrightarrow} \left\{\sum_{k\geq 0} \mathbb{P}(D_{\tau_{k+1}} | \mathcal{F}_{\tau_k}) = \infty\right\}$$

hence by Lévy's extension of the Borel-Cantelli lemma it follows that a.s. on $B(\varepsilon)$ infinitely many (and hence at least one) of D_{τ_k} occur, that is, $x_{(1)}(\tau_k + \ell) \geq 1/p$. Now the result follows from Lemma 3. $\square$

Assume for now that $p < \frac{N}{2}$; in this case $N - \frac{N}{2p} + 1 < N$ (see (3.6)). The case $p \geq \frac{N}{2}$ will be dealt with separately.

Claim 2. *Suppose $0 \leq x_1 \leq \cdots \leq x_N$ and k is such that*

$$k \in \{2, \ldots, N-1\}, \qquad N > 2p(N-k). \tag{3.8}$$

Let $\beta = \frac{2p(k-1)}{N-2p(N-k)} = 1 + \frac{(p-1)N + p(N-2)}{N-2p(N-k)} > 1$. If $x_N > \beta x_k$ then $\{x_1, \ldots, x_N\}' = \{x_1, \ldots, x_{N-1}\}$.

Proof. $x_N > \beta x_k$ implies

$$0 < [N - 2p(N-k)]x_N - 2p(k-1)x_k = Nx_N - 2p[(k-1)x_k + (N-k)x_N]$$

$$\leq Nx_N - 2p[x_2 + \cdots + x_N] \leq Nx_N + Nx_1 - 2p[x_1 + \cdots + x_N] = 2N \cdot \left[\frac{x_1 + x_N}{2} - p\mu\right],$$

since $N - 2p > 0$, hence $|x_N - p\mu| \geq x_N - p\mu > p\mu - x_1 = p|\mu - x_1|$ and thus x_N is the furthermost point from the p-centre of mass. □

Lemma 6. *Let k satisfy the conditions* (3.8) *and β be defined as in Claim 2. Then* $\{x_{(k)}(t) \to 0\} \underset{a.s.}{\Longrightarrow} \{\mathcal{X}'(t) \to 0\}$.

Proof. Let

$$H_s^\delta = \bigcap_{t \geq s} \{x_{(k)}(t) \leq \delta, x_{(N-1)}(t) > \beta\delta\}$$

and

$$V_s^\delta = \bigcap_{t \geq s} \left\{ \zeta_t \notin [\delta, \frac{\delta(1+\beta)}{2}] \right\}.$$

By Claim 2 it follows that $(V_s^\delta)^c \cap H_s^\delta = \emptyset$ for all $s \in \mathbb{N}$ and $\delta > 0$. Meanwhile, by the assumption of full support, $\mathbb{P}(V_s^\delta) = 0$ for for all $s \in \mathbb{N}$ and $\delta > 0$. Therefore

$$\mathbb{P}(H_s^\delta) = \mathbb{P}(H_s^\delta \cap (V_s^\delta)^c) + \mathbb{P}(H_s^\delta \cap V_s^\delta) = 0,$$

all $s \in \mathbb{N}$ and $\delta > 0$. But

$$\{x_{(k)}(t) \to 0\} \cap \{\mathcal{X}'(t) \not\to 0\} \subset \bigcup_{n \in \mathbb{N}} \bigcup_{s \in \mathbb{N}} H_s^{\frac{1}{n}},$$

hence $\mathbb{P}\left(\{x_{(k)}(t) \to 0\} \cap \{\mathcal{X}'(t) \not\to 0\}\right) = 0$, which is equivalent to the statement of the lemma. □

The following statement shows that if all the points to the right of $x_{(k)}$ lie very near each other, while the left-most one lies near zero, then it is to be removed.

Claim 3. *Let $a \in (0, 1]$ and suppose that $k \in \{2, \ldots, N-1\}$ satisfies* (3.6)*. Then there exist small $\delta, \Delta > 0$, depending on N, k, p, a such that if*

$$\begin{aligned} 0 \leq x_1 \leq \delta; \\ x_1 \leq x_i \leq x_N \quad \textit{for } i = 2, \ldots, N-1; \\ x_k, x_{k+1}, \ldots, x_N \in [a(1-\Delta), a) \end{aligned}$$

then $\{x_1, \ldots, x_N\}' = \{x_2, \ldots, x_N\}$.

Proof. The condition to remove the leftmost point is $p\mu - \frac{x_1+x_N}{2} > 0$ where $\mu = (x_1 + \cdots + x_N)/N$. However,

$$\begin{aligned}2N\left(p\mu - \frac{x_1+x_N}{2}\right) &= 2p(x_2 + \cdots + x_{N-1}) - (N-2p)x_1 - (N-2p)x_N \\ &\geq 2p(x_k + \cdots + x_{N-1}) - (N-2p)\delta - (N-2p)a \\ &\geq 2p(N-k)a(1-\Delta) - (N-2p)\delta - (N-2p)a \\ &= a\left[2p(N-k)(1-\Delta) - (N-2p)\right] - (N-2p)\delta\end{aligned}$$

The RHS is linear in δ and Δ, and when $\delta = \Delta = 0$ it is strictly positive by the assumption on k; hence it can also be made positive, by allowing $\delta > 0$ and $\Delta > 0$ to be sufficiently small. □

Corollary 2. *Suppose that $\mathcal{X}(t) = \{x_1, \ldots, x_N\}$ satisfies the conditions of Claim 3 for some a and k. Let δ be the quantity from this claim. Then*

$$\mathbb{P}(x_{(1)}(t+j) > \delta \text{ for some } 1 \leq j \leq k|\mathcal{F}_t) \geq c = c_{a\Delta} > 0.$$

Proof. The probability to sample a new point $\zeta \in (a(1-\Delta), a]$ is bounded below by $f(a\Delta)$ where f is the same function as in (2.1). On the other hand, if the new point is sampled in $(a(1-\Delta), a]$ then $\mathcal{X}(t+1)$ continues to satisfy the conditions of Claim 3 as long as the leftmost point is in $[0, \delta]$. By repeating this argument at most k times and using the induction, we get the result with $c = [f(a\Delta)]^k > 0$. □

Lemma 7. *Let $k \in \mathbb{N}$ satisfy* (3.6). *Then*

$$\left\{x_{(k)}(t) \not\to 0\right\} \underset{a.s.}{\Longrightarrow} \left\{\mathcal{X}'(t) \to 1\right\}.$$

Proof. Note that by Lemma 5, it suffices to show that $\left\{x_{(k)}(t) \not\to 0\right\} \underset{\text{a.s.}}{\Longrightarrow} \left\{x_{(1)}(t) \not\to 0\right\}$.

If $x_{(k)}(t) \not\to 0$, there exists an $a > 0$ such that $x_{(k)}(t) \geq a$ for infinitely many t's. Let s be such a time. Now suppose that $\zeta_{s+i} \in I := (a(1-\Delta), a]$ for $i = 0, 1, \ldots, N-1$ where Δ is defined in Claim 3; the probability of this event is strictly positive and depends only on a and δ (see (2.1)). As long as there are points of $\mathcal{X}(s+i)$ on *both* sides of the interval I, none of the points inside I can be removed; hence, for some $u \in \{s, s+1, \ldots, s+N-1\}$ we have that either $\min \mathcal{X}(u) > a(1-\Delta)$ or $\max \mathcal{X}(u) \leq a$. In the first case, $x_{(1)}(u) > a(1-\Delta)$.

In the latter case, both $x_{(N)}(u) \in I$ and $x_{(k)}(u) \in I$, since every time we replaced a point, the number of points to the left of I did not increase (and there were initially at most $k-1$ of them). As a result

$$a(1-\Delta) \leq x_{(k)}(u) \leq x_{(k+1)}(u) \leq \cdots \leq x_{(N)}(u) \leq a.$$

Together with Corollary 2, this yields

$$\{x_{(k)}(t) \geq a \text{ i.o.}\} \underset{\text{a.s.}}{\Longrightarrow} \{x_{(1)}(t) \geq \min\{a(1-\Delta), \delta\} \text{ i.o.}\} \underset{\text{a.s.}}{\Longrightarrow} \{x_{(1)}(t) \not\to 0\}$$

which proves Lemma 7. □

So far we have shown that if $k > N(1-\frac{1}{2p})$ then $\{x_{(k)}(t) \to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{\mathcal{X}'(t) \to 0\}$ and if $k < N(1-\frac{1}{2p})+1$ then $\{x_{(k)}(t) \not\to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{\mathcal{X}'(t) \to 1\}$. From this we may conclude that if $N(1-\frac{1}{2p}) < k < N(1-\frac{1}{2p})+1$ (this is possible if $\frac{N}{2p}$ is not an integer) then $\mathcal{X}'(t)$ must converge either to 0 or 1. It remains to consider the case when $\frac{N}{2p}$ *is* an integer and we then choose $k = N(1-\frac{1}{2p})$ and proceed to show that $\{x_{(k)}(t) \to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{\mathcal{X}'(t) \to 0\}$, while $\{x_{(k)}(t) \not\to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{\mathcal{X}'(t) \to 1\}$. We start by showing $\{x_{(k)}(t) \to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{\mathcal{X}'(t) \to 0\}$.
For this purpose let $a \in (0,1)$ and define $I_a^0 = \left[0, \frac{a}{2N^N}\right]$, $I_a^1 = \left(\frac{a}{2N^N}, \frac{a}{N^N}\right)$, $I_a = I_a^0 \cup I_a^1$ and

$$D_a(s) := \{\sup_{t \geq s} x_{(k)}(t) < \frac{a}{2N^N}\},$$

$$C_a(s) = \{\zeta_{s+1}, \ldots, \zeta_{s+(N-k)} \in I_a^1\},$$

$$E_a := \{x_{(k)}(t) \to 0, \quad x_{(k+1)}(t) \geq a \quad i.o.\}$$

Lemma 8. *Suppose that $k = N(1-\frac{1}{2p})$ then $\mathbb{P}(E_a) = 0$ for any $a > 0$.*

Proof. Fix $a > 0$. Clearly $E_a \subseteq \{x_{(k)}(t) \to 0\} \subseteq \bigcup_{s=1}^{\infty} D_a(s)$. Let

$$\tau_0 = \inf\{t \geq 0 : x_k(t) < \frac{a}{2N^N}, \quad x_{k+1}(t) \geq a\},$$

while for $l \geq 1$

$$\tau_l = \inf\{t > \tau_{l-1} + N - 1 : x_k(t) < \frac{a}{2N^N}, \quad x_{k+1}(t) \geq a\}$$

and note that $\tau_l < \infty$ a.s. on E_a, for each $l \geq 1$. If at time $\tau_l + 1$, ζ_{τ_l+1} is sampled in I_a^1 then, using $\frac{k+2}{2N^{2N}} < 1$ (since $N \geq 2$) and by also plugging in $p = \frac{N}{2(N-k)}$ we find that,

$$p\mu(\tau_l+1) = p\frac{(N-1)\mu'(\tau_l) + \zeta_{\tau_l+1}}{N} \leq p\frac{\left(k\frac{a}{2N^N} + (N-k-1)x_{(N-1)}(\tau_l)\right) + \frac{a}{N^N}}{N}$$

$$\leq p\frac{a\left(\frac{k+2}{2N^N}-1\right)}{N}+p\frac{N-k}{N}x_{(N-1)}(\tau_l)=\frac{a\left(\frac{k+2}{2N^N}-1\right)}{2(N-k)}+\frac{x_{(N-1)}(\tau_l)}{2}<\frac{x_{(N-1)}(\tau_l)}{2},$$

which means that the right-most point gets rejected. Similarly if ζ_{τ_l+j} is sampled in I_a^1 for $j=1...N-k$ (i.e. if $C_a(\tau_l)$ occurs) then

$$p\mu(\tau_l+j)\leq p\frac{a\left(\frac{k+2j}{2N^N}-j\right)+(N-k)x_{(N-1)}(\tau_l+j-1)}{N}<\frac{x_{(N-1)}(\tau_l+j-1)}{2},$$

for $j=1,\dots,N-k$, implying that all the $N-k$ new points get accepted (and thereby rejecting all points to the right of a). If $C_a(\tau_l)\cap D_a(s(l))$ occurs then for $t\geq s(l):=\tau_l+N-k-1$, the number of points in I_a^0 must be at least k. Consider the interval $I=I_a^0\cup I_a^1$ and note that at time $s(l)$ all points of the core will lie in I. We will establish through the next claim that if at some time we have $N-1$ points in I, it will be impossible for $x_{(N-1)}(t)$ to ever reach above a while still keeping at least k points in I.

Claim 4. *Fix some $l\in\mathbb{N}$. On the event $C_a(\tau_l)\cap D_a(s(l))$ we have that if there are $0\leq j\leq N-k-1$ points of the core in I^c at any $t\geq s(l)$ then $x_{(N-1)}(t)\leq N^j\frac{a}{N^N}=\frac{a}{N^{N-j}}$. In particular $x_{(N-1)}(t)<a$ on $C_a(\tau_l)\cap D_a(s(l))$ for all $t\geq s(l)$.*

Proof. We prove this by induction. When $j=0$ this is true (on $C_a(\tau_l)\cap D_a(s(l))$) since at time $s(l)$ we have all core points in I and if we do not move any points into I^c then the right most point is to the left of $\frac{a}{N^n}$. Assume the claim is true for $j=J$. Since at time $s(l)$ there are no points in in I^c then if there are ever to be $J+1$ points in I^c there exists a time $t>s(l)$ when we go from J points in I^c to $J+1$ points in I^c at time $t+1$.

At time $t+1$, ζ_{t+1} will be rejected if $\zeta_{t+1}>\max(p\mu(t+1),x_{(N-1)}(t))$. Since $\zeta_{t+1}>p\mu(t+1)$ if and only if (by plugging in p) $\zeta_{t+1}>\frac{N-1}{N-k-1}\mu'(t)$, but $\frac{N-1}{N-k-1}\mu'(t)\leq Nx_{(N-1)}(t)$, so we may conclude that ζ_{t+1} will be rejected if $\zeta_{t+1}>Nx_{(N-1)}(t)$ regardless of how the $N-1-L$ points in I and the remaining J points in I^c are distributed. We obtain that

$$x_{(N-1)}(t+1)\leq Nx_{(N-1)}(t)\leq N^{J+1}\frac{a}{N^N}=\frac{a}{N^{N-(J+1)}},$$

by the induction hypothesis and this proves the claim. ☐

It follows directly from Claim 4 that

$$E_a\cap D_a(s(l))\subseteq C_a(\tau_{l+k})^c,\quad \forall l,k\in\mathbb{N}_0,$$

as a consequence, since $\mathbb{P}\left(C_a(\tau_l)|\mathcal{F}_{\tau_l}\right) \geq f\left(|I_a^1|\right)^{N-l}$ on $\{\tau_l < \infty\}$ and since $\{C_a(\tau_{l+k})^c\}_k$ are all independent for every l by the construction of τ_l

$$\mathbb{P}\left(E_a \cap D_a(s(l))\right) \leq \mathbb{P}\left(\bigcap_{k=0}^{\infty} C_a(\tau_{l+k})^c\right) \leq \prod_{k=0}^{\infty}\left(1 - f(|I_a^1|)\right) = 0.$$

Using $E_a \subseteq \bigcup_{l=0}^{\infty} D_a(s(l))$, we conclude

$$\mathbb{P}\left(E_a\right) = \mathbb{P}\left(\bigcup_{l=1}^{\infty}\left(E_a \cap D_a(s(l))\right)\right) \leq \sum_{l=0}^{\infty} \mathbb{P}\left(E_a \cap D_a(s(l))\right) = 0.$$

□

Lemma 8 implies that when $k = N(1 - \frac{1}{2p})$ then

$$\mathbb{P}\left(x_{(k)}(t) \to 0, x_{(k+1)}(t) \not\to 0\right) = \mathbb{P}\left(\bigcup_{n \geq 1} E_{1/n}\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(E_{1/n}) = 0,$$

i.e. $\{x_{(k)}(t) \to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{x_{(k+1)}(t) \to 0\}$, but $k+1 > N(1-\frac{1}{2p})$ and so by Lemma 6, $\{x_{(k)}(t) \to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{\mathcal{X}'(t) \to 0\}$. Note that since $k = N(1-\frac{1}{2p})$ then k obviously satisfies (3.6) so Lemma 7 implies that $\{x_{(k)}(t) \not\to 0\} \underset{\text{a.s.}}{\Longrightarrow} \{\mathcal{X}'(t) \to 1\}$ which completes the proof when $p < \frac{N}{2}$. For the case $p \geq \frac{N}{2}$ we have

Lemma 9. *If $p \geq \frac{N}{2}$ then $\mathcal{X}'(t) \to 1$ a.s.*

Proof. The case $p > \frac{N}{2}$ is easy: unless $x_{(N)} = 0$ we have

$$p \cdot \frac{x_{(1)} + \dots x_{(N)}}{N} > \frac{x_{(1)} + \cdots + x_{(N)}}{2} \geq \frac{x_{(1)} + x_{(N)}}{2}$$

hence it is the left-most point which is always removed. For the case $p = \frac{N}{2}$ we notice that at each moment of time we either have a tie (between the left-most and right-most point) or remove the left-most point. At time t we can only have a tie if $x_{(1)}(t) = ... = x_{(N-1)}(t) = 0$ and if this is true then eventually the right-most point will be kept and the process becomes monotone after this (the left-most point will always be rejected). □

To prove part (c), note that unless $x_{(1)}(0) > 0$ already, by repeating the arguments from the beginning of the proof of Lemma 7, with a positive probability we can "drag" the whole configuration in at most $N-1$ steps to the right of zero, that is, there is

$0 \leq t_0 \leq N-1$ such that $\mathbb{P}(\min \mathcal{X}'(t_0) > 0) > 0$. Now we can apply Lemma 4 and then Lemma 3.

This concludes the proof $\square$

Remark 3. *For an alternative proof see section two of the Appendix.*

4 Non-convergence to zero for $p > 1$ and $N = 3$

In this section we prove the following

Theorem 3. *Suppose that $N = 3$, $p > 1$ and ζ, restricted to some neighbourhood of zero, is a continuous random variable with a non-decreasing density (e.g. uniformly distributed). Assume also that the initial points are all i.i.d. ζ-distributed. Then $\mathcal{X}'(t) \to 1$ as $t \to \infty$ a.s.*

Remark 4.

- *In case $p \geq 3/2$ we already know that $\mathcal{X}'(t) \to 1$ for any initial configuration and any distribution (see Lemma 14), so we have to prove the theorem only for $p \in (1, 3/2)$.*
- *Simulations suggest that the statement of Theorem 3 holds, in fact, for a much more general class of distributions ζ.*

Proof of Theorem 3. Let $\varepsilon \in (0, 1/2)$ be such that ζ conditioned[2] on $\{\zeta \leq \varepsilon\}$ has a non-decreasing density; according to the statement of the Theorem 3 such an ε must exist. Furthermore we can also assume that the denisty of ζ is bounded on $[0, 2\epsilon]$ by choosing ϵ small enough, indeed since the density is non-decreasing and is integrable it follows that it can have at most a single integrable singularity which then must be located at it's right-most point of definition. Let us fix this ε from now on. As before, denote by $x_1, \ldots, x_N$ N distinct points on $[0, 1]$, and let $x_{(1)}, \ldots, x_{(N)}$ be this unordered N-tuple sorted in the increasing order. Let

$$\{y_1, \ldots, y_{N-1}\} = \{x_1, \ldots, x_N\}'_p$$

be the unordered N-tuple $\{x_1, \ldots, x_N\}$ with the farthest point from p-centre of mass removed; w.l.o.g. assume that y_i are already in the increasing order.

[2] note that the full support assumption ensures that the probability of this event is positive

Lemma 10. *The operation $\{\dots\}'_p$ is monotone in p, that is, if $\hat{p} \ge \tilde{p}$ and*

$$\{\hat{y}_1, \dots, \hat{y}_{N-1}\} = \{x_1, \dots, x_N\}'_{\hat{p}},$$
$$\{\tilde{y}_1, \dots, \tilde{y}_{N-1}\} = \{x_1, \dots, x_N\}'_{\tilde{p}}$$

then $\hat{y}_i \ge \tilde{y}_i$, $i = 1, \dots, N-1$.

Proof. Assume w.l.o.g. $x_1 \le ... \le x_N$, and let $\mu = \mu(\{x_1, \dots, x_N\})$. Notice that, regardless of the value of p, the only points which can possibly be removed are x_1 or x_N (since they are the two extreme points). Therefore, it suffices to show that $\{x_1, \dots, x_N\}'_{\tilde{p}} = \{x_2, \dots, x_N\}$ implies $\{x_1, \dots, x_N\}'_{\hat{p}} = \{x_2, \dots, x_N\}$. Note also that $|x_1 - p\mu| = p\mu - x_1$ for all $p \ge 1$.

If $\tilde{p}\mu - x_1 > |\tilde{p}\mu - x_N|$ and $\tilde{p}\mu - x_N > 0$, that is, the $p-$centre of mass lies to the right of x_N, then $\hat{p}\mu > \tilde{p}\mu > x_N$ as well, and hence x_1 is discarded.

On the other hand, if $\tilde{p}\mu - x_1 > |\tilde{p}\mu - x_N|$ and $\tilde{p}\mu < x_N$ then either $\hat{p}\mu < x_N$, or $\hat{p}\mu \ge x_N$. In the first case,

$$\hat{p}\mu - x_1 > \tilde{p}\mu - x_1 > |\tilde{p}\mu - x_N| = x_N - \tilde{p}\mu > x_N - \hat{p}\mu = |x_N - \hat{p}\mu|$$

so x_1 is discarded. In the second case, $p-$centre of mass lies to the right of x_N and so x_1 is also discarded. □

Lemma 11. *Let h be a real-valued function on the sets of N real numbers. Suppose that h is non-increasing in each of its arguments, namely*

$$h(x_1, x_2, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_N) \le h(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_N)$$

whenever $x'_i \ge x_i$. Let $\mathcal{E}_t$ be some $\mathcal{F}_t$-measurable event, and suppose that

$$\mathbb{E}\left(h(\mathcal{X}'(t+1))|\mathcal{F}_t\right) \le h(\mathcal{X}'(t)) \text{ on } \mathcal{E}_t \tag{4.9}$$

for $p = 1$. Then (4.9) *holds for $p > 1$ as well.*

Proof. Let

$$G_p(\mathcal{X}'(t), \zeta_{t+1}) = \{x_{(1)}(t), x_{(2)}(t), \dots, x_{(N-1)}(t), \zeta_{t+1}\}'_p$$

be the new core after the new point ζ_{t+1} is sampled and the farthest point from the $p-$centre of mass is removed; note that $\mathcal{X}'(t+1) = G_p(\mathcal{X}'(t), \zeta_{t+1})$. Then on $\mathcal{E}_t$

$$\mathbb{E}(h(\mathcal{X}'(t+1))|\mathcal{F}_t) = \mathbb{E}(h(G_p(\mathcal{X}'(t), \zeta_{t+1}))|\mathcal{F}_t) \le \mathbb{E}(h(G_1(\mathcal{X}'(t), \zeta_{t+1}))|\mathcal{F}_t) \le h(\mathcal{X}'(t))$$

since the operation $\{\dots\}'_p$ is monotone in p by Lemma 10 and h is decreasing in each argument. $\square$

From now on assume $N = 3$ and $p = 1$. Denote $x_{(1)}(t) = a$, $x_{(2)}(t) = b$ and consider the events

$$\begin{aligned} L_b &= \{\zeta_{t+1} \in ((2a-b)^+, a)\}, & R_a &= \{\zeta_{t+1} \in (b, 2b-a)\}, \\ B_b &= \left\{\zeta_{t+1} \in \left(a, \tfrac{a+b}{2}\right)\right\}, & B_a &= \left\{\zeta_{t+1} \in \left(\tfrac{a+b}{2}, b\right)\right\} \end{aligned}$$

(we assume that b is smaller than $1/2$, yielding $2b - a < 1$.) If $x_{(2)}(t) \le \varepsilon$ then $\mathcal{X}'(t+1) \ne \mathcal{X}'(t)$ implies that one of the events L_b, B_b, B_a or R_a occurs (i.e. all points sampled outside of $((2a-b)^+, 2b-a)$ are rejected at time $t+1$). Let us study the core $\mathcal{X}'(t+1) = \{\zeta_{t+1}, a, b\}'$ on these events: on L_b and B_b we have $\mathcal{X}'(t+1) = \{x, a\}$, while on B_a and R_a we have $\mathcal{X}'(t+1) = \{x, b\}$.

We have, assuming $x_{(1)}(t) = a$ and $x_{(2)}(t) = b$,

$$\mathbb{E}(h(\mathcal{X}'(t+1)) - h(\mathcal{X}'(t))|\mathcal{F}_t) = \mathbb{E}(h(\{\zeta, a, b\}') - h(a, b)).$$

When $0 \le a \le b \le \varepsilon$ we have $2b - a \le 2\varepsilon$. Define

$$g(x) = h(\{x, a, b\}') - h(a, b) = \begin{cases} h(x, a) - h(a, b), & \text{if } x \in ((2a-b)^+, a); \\ h(a, x) - h(a, b), & \text{if } x \in (a, (a+b)/2); \\ h(x, b) - h(a, b), & \text{if } x \in ((a+b)/2, b); \\ h(b, x) - h(a, b), & \text{if } x \in (b, 2b-a) \\ 0, & \text{otherwise,} \end{cases}$$

which is positive in the first two cases, and negative in the next two. Let $\varphi(x)$ be the density of ζ conditioned on $\{\zeta \in [0, 2\varepsilon]\}$. By the monotonicity of φ and the positivity (negativity resp.) of g on the first (second resp.) interval,

$$\begin{aligned} \Delta(a, b) := \mathbb{E}\left[g(\zeta) 1_{\zeta \in [0, 2\varepsilon]}\right] &= \int_{(2a-b)^+}^{\frac{a+b}{2}} g(x)\varphi(x)\mathrm{d}x + \int_{\frac{a+b}{2}}^{2b-a} g(x)\varphi(x)\mathrm{d}x \\ &\le \varphi\left(\frac{a+b}{2}\right) \int_{(2a-b)^+}^{\frac{a+b}{2}} g(x)\mathrm{d}x + \varphi\left(\frac{a+b}{2}\right) \int_{\frac{a+b}{2}}^{2b-a} g(x)\mathrm{d}x = \varphi\left(\frac{a+b}{2}\right) \cdot \Lambda, \end{aligned} \tag{4.10}$$

where

$$\Lambda = \Lambda(a, b) = \int_{(2a-b)^+}^{a} (h(x, a) - h(a, b))\mathrm{d}x + \int_{a}^{\frac{a+b}{2}} (h(a, x) - h(a, b))\mathrm{d}x$$

$$+\int_{\frac{a+b}{2}}^{b} (h(x,b) - h(a,b))\mathrm{d}x + \int_{b}^{2b-a} (h(b,x) - h(a,b))\mathrm{d}x.$$

So if we can establish that $\Lambda \le 0$ for a suitable function h, then indeed $\Delta(a,b) \le 0$, and the supermartingale property follows.

Remark 5. *Notice that the method of proof, presented here, could possibly work for $N > 3$ as well; that is, if one can find a function $h(x_1, \dots, x_{N-1})$, which is positive and decreasing in each of its arguments, and $h(\mathcal{X}'(t))$ is a supermartingale provided* $\max \mathcal{X}'(t) < \varepsilon$ *for some $\varepsilon > 0$. Unfortunately, however, we were not able to find such a function.*

Set

$$h(x,y) = -2\log\left(\max\left\{x, \frac{y}{2}\right\}\right) \ge 0; \tag{4.11}$$

it is easy to check h is indeed monotone in each of its arguments as long as $x, y \in (0,1]$. Let us now compute the integrals in the expression for Λ. We have

$$\Lambda = \begin{cases} 3(a-b)\ln 2 - 3a + 2b, & \text{if } a \le \frac{b}{3}; \\ (a+b)\ln(a+b) - (a+b)\ln a + (a-5b)\ln 2 + b, & \text{if } \frac{b}{3} < a \le \frac{b}{2}; \\ (a+b)\ln(a+b) + (2a-4b)\ln b + 3(b-a)\ln a + (b-5a)\ln 2 + b, & \text{if } \frac{b}{2} < a \le \frac{2b}{3}; \\ (a+b)\ln(a+b) + (2a-4b)\ln b + (5b-7a)\ln a - (a+b)\ln 2 & \\ \quad +3(b-a) + (4a-2b)\ln(2a-b), & \text{if } \frac{2b}{3} < a \le b. \end{cases}$$

It turns out that $h(\mathcal{X}'(t))$ indeed has a non-positive drift, provided $0 < a \le b \le \varepsilon$, as is shown by the following

Lemma 12. $\Lambda \le 0$ *for* $0 < a \le b \le 1$.

Proof. Substitute $a = b\nu$ in the expression for Λ. Then for $\nu \le 1/3$ we easily obtain

$$\Lambda = -b\left[3\nu(1-\ln 2) + \ln 8 - 2\right] \le 0.$$

For $1/3 < \nu \le 1/2$ we have $2\Lambda = -bC_1(\nu) \le 0$ where

$$C_1(\nu) = (1+\nu)\ln\frac{\nu}{1+\nu} + (5-\nu)\ln 2 - 1 > 0,$$

since $\frac{\partial^2 C_1(\nu)}{\partial^2 \nu} = -\frac{1}{\nu^2(1+\nu)} < 0$ and hence $\min_{1/3 \le \nu \le 1/2} C_1(\nu)$ is achieved at one of the endpoints $\nu = 1/3$ or $\nu = 1/2$; the values there are $C_1(1/3) = \ln(4) - 1 > 0$ and $C(1/2) = \frac{1}{2}\ln\left(\frac{512}{27}\right) - 1 > 0$ respectively.

For $1/2 < \nu \le 2/3$ we have $\Lambda = -bC_2(\nu) \le 0$ where

$$C_2(\nu) = -(1+\nu)\ln(1+\nu) + (3\nu - 3)\ln \nu - 1 + (5\nu - 1)\ln 2 > 0,$$

since $\frac{\partial^2 C_2(\nu)}{\partial^2 \nu} = \frac{2\nu^2+6\nu+3}{\nu^2(1+\nu)} > 0$ and $\left.\frac{\partial C_2(\nu)}{\partial \nu}\right|_{\nu=2/3} = \ln\left(\frac{256}{45}\right) - \frac{5}{2} < 0$ implies that $\frac{\partial C_2(\nu)}{\partial \nu} < 0$ for all $\nu \in [1/2, 2/3]$ and hence $\min_{1/2 \le \nu \le 2/3} C_2(\nu) = C_2(2/3) = \frac{1}{3}\ln\left(\frac{104976}{3125}\right) - 1 > 0$.

Finally, for $2/3 < \nu \le 1$ we have $\Lambda = -bC_3(\nu) \le 0$, where

$$C_3(\nu) = \nu \log \frac{2\nu^7}{(2\nu-1)^4(\nu+1)} + \log \frac{2(2\nu-1)^2}{\nu^5(\nu+1)} + 3(\nu - 1) > 0$$

since

$$\frac{\mathrm{d}^2 C_3(\nu)}{\mathrm{d}\nu^2} = \frac{(2\nu+5)(2\nu^2-1)}{(2\nu-1)\nu^2(\nu+1))}$$

changes its sign from $-$ to $+$ at $1/\sqrt{2} \in (2/3, 1)$ and therefore $\frac{\partial C_3(\nu)}{\partial \nu}$ achieves its maximum at the endpoints of the interval; thus

$$\max_{2/3 \le \nu \le 1} \frac{\partial C_3(\nu)}{\partial \nu} = \max_{\nu = 2/3, 1} \frac{\partial C_3(\nu)}{\partial \nu} = \max\left\{-\frac{5}{2} + \ln\left(\frac{256}{45}\right), 0\right\} = 0$$

Therefore, $C_3(\nu)$ is decreasing and hence $\min_{2/3 \le \nu \le 1} C_3(\nu) = C_3(1) = 0$. □

Choose $\tau_0 = 0$, and for $k = 1, 2, \dots$, define the sequence of stopping times

$$\begin{aligned}
\eta_k &= \inf\left\{t > \tau_{k-1} : \quad x_{(2)}(t) < \epsilon\right\}, \\
\tau_k &= \inf\left\{t > \eta_k : \ x_{(2)}(t) > \epsilon\right\}, \\
\gamma_{k,t} &= \min\left(\eta_k + t, \tau_k\right),
\end{aligned}$$

so that $\tau_0 < \eta_1 \le \gamma_{1,t} \le \tau_1 < \eta_2 \le \gamma_{2,t} \le \tau_2 < \dots$ for all $t \ge 0$, with the usual conventions that $\inf \emptyset = +\infty$ and that if one of the stopping times is $+\infty$ then the subsequent ones are also $+\infty$. Note that $\{\mathcal{X}'(t) \to 0\} \subseteq \bigcup_{k=1}^{\infty}(\{\tau_k = \infty\} \cap \{\eta_k < \infty\})$ so it suffices to show that $\mathbb{P}\left(\{\tau_k = \infty\} \cap \{\eta_k < \infty\} \cap \{\mathcal{X}'(t) \to 0\}\right) = 0$ for all $k \ge 1$.

We will show that with h as in (4.11), $\lim_{t\to\infty} h(\mathcal{X}'(\gamma_{k,t}))$ exists and is finite a.s. on $\{\tau_k = \infty\} \cap \{\eta_k < \infty\}$. Since $\lim_{b \downarrow 0} h(a,b) = +\infty$ this implies $\liminf_{t\to\infty} x_{(1)}(\gamma_{k,t}) > 0$ a.s. (since otherwise $\mathbb{P}\left(\lim_{t\to\infty} h(\mathcal{X}'(\gamma_{k,t})) = +\infty\right) > 0$). If we can show that

$h(\mathcal{X}'(\gamma_{k,t}))$ is a positive supermartingale on $\{\tau_k = \infty\} \cap \{\eta_k < \infty\}$ then we are done. From now on fix $k \geq 1$ so that we may denote $\gamma_t = \gamma_{k,t}$ without loss of generality.

The positivity of $h(\mathcal{X}'(\gamma_t))$ follows by the definitions of h. Letting $\xi_t = h(\mathcal{X}'(t))$ (which is always non-negative) then by (4.10) and Lemma 12 it follows that

$$\mathbb{E}\left[(\xi_{t+1} - \xi_t)1_{\zeta_{t+1}<2\epsilon}|\mathcal{F}_t\right] \leq 0. \tag{4.12}$$

Note that if $\zeta_{t+1} \geq 2\epsilon$ and $x_{(1)}(t) < 2\epsilon$ then if ζ_{t+1} is accepted into the core, $\mathcal{X}'(t+1) = \{x_{(2)}(t), \zeta_{t+1}\}$ and if it is not accepted into the core then $\xi_{t+1} = \xi_t$. Since

$$\max\left(x_{(1)}(t), x_{(2)}(t)/2\right) \leq \max\left((x_{(2)}(t) \wedge \zeta_{t+1}, (x_{(2)}(t) \vee \zeta_{t+1})/2\right),$$

it follows that

$$\mathbb{E}\left[(\xi_{t+1} - \xi_t)1_{\zeta_{t+1}\in[2\epsilon,1]}1_{x_{(1)}(t)<2\epsilon}|\mathcal{F}_t\right] \leq 0. \tag{4.13}$$

If $\zeta_{t+1} \geq 2\epsilon$ and $x_{(1)}(t) \geq 2\epsilon$ then $\xi_t \leq -2\log(\epsilon)$ and $\xi_{t+1} \leq -2\log(\epsilon)$ which implies

$$\mathbb{E}\left[(\xi_{t+1} - \xi_t)1_{\zeta_{t+1}\in[2\epsilon,1]}1_{x_{(1)}(t)\geq 2\epsilon}|\mathcal{F}_t\right] \leq -4\log(\epsilon). \tag{4.14}$$

We can conclude

$$\mathbb{E}\left[(\xi_{t+1} - \xi_t)1_{\zeta_{t+1}\in[2\epsilon,1]}|\mathcal{F}_t\right] \leq -4\log(\epsilon) := c. \tag{4.15}$$

Combining (4.12) and (4.15) gives us $\mathbb{E}\left[(\xi_{t+1} - \xi_t)|\mathcal{F}_t\right] \leq c$ taking expectations and iterating over t, we find that $\mathbb{E}\left[\xi_t\right] \leq ct + \mathbb{E}[\xi_0]$ so $\mathbb{E}\left[\xi_t\right] < \infty$ if $\mathbb{E}[\xi_0] < \infty$. Now

$$\begin{aligned}\mathbb{E}[\xi_0] &\leq -2\mathbb{E}[\log(\max(\zeta_1/2, \zeta_2/2))] \leq -2\mathbb{E}[\log(\zeta/2)] \leq -2\mathbb{E}\left[\log(\zeta/2)1_{\zeta<\epsilon}\right] - 2\log(\epsilon/2)\\ &\leq -2\left(\varphi(\epsilon)\int_0^\epsilon \log(x)dx + \log(\epsilon)\right) = -2\left(\varphi(\epsilon)\epsilon\left(\log(\epsilon) - 1\right) + \log(\epsilon/2)\right),\end{aligned}$$

which is finite. Now it suffices to show $\mathbb{E}\left[h(\mathcal{X}'(\gamma_t + 1))|\mathcal{F}_{\gamma_t}\right] \leq h(\mathcal{X}'(\gamma_t))$ since this will in fact imply
$\mathbb{E}\left[h(\mathcal{X}'(\gamma_{t+1}))|\mathcal{F}_{\gamma_t}\right] \leq h(\mathcal{X}'(\gamma_t))$ (see the proof of Theorem 2 in [5] for details). Note that $x_{(2)}(\gamma_t + 1) \leq px_{(2)}(\gamma_t) \leq p \cdot \epsilon < 2\epsilon$, therefore it follows from Lemma 12 that $\mathbb{E}\left[h(\mathcal{X}'(\gamma_t + 1))|\mathcal{F}_{\gamma_t}\right] \leq h(\mathcal{X}'(\gamma_t))$. We have thus showed that $h(\mathcal{X}'(\gamma_t))$ is a positive supermartingale and this concludes the proof. □

5 Appendix: The calculations for the proof of Lemma 2.

Observe that all expressions for $\mathbf{A}_j$ are fractions of the polynomials in (a, f, p, M); moreover, their denominators

$$\begin{array}{ll} 3M(M-1) & (\text{for } \mathbf{A}_1), \\ 3M(M-1)(M+1-2p)^3 & (\text{for } \mathbf{A}_2 \text{ and } \mathbf{A}_4), \\ 12M(M-1)(M+1-2p)^3p^3 & (\text{for } \mathbf{A}_3 \text{ and } \mathbf{A}_5) \end{array}$$

are always positive. Throughout the rest of the proof let $\mathsf{n}(w)$ denote the numerator of such a fraction w.

Case 1: $\mathbf{I}_1 \leq 0$

Observe that

$$\mathsf{n}(\mathbf{A}_1) = -2M^2 - 3M\mu + 2M + 1 + [3M\mu - 1]Mp$$

and the term in the square brackets is positive as $M\mu \geq 1$, so the maximum of $\mathsf{n}(\mathbf{A}_1)$ is achieved at the highest possible value of p. However, in this case we have $p \leq p_1$, hence

$$\mathsf{n}(\mathbf{A}_1)1_{X_1 \leq 0} \leq \mathsf{n}(\mathbf{A}_1)|_{p=p_1} = -\frac{s_1}{2\mu}$$

where

$$s_1 = (M^2-2)\mu + (1-6\mu)(1-\mu)M + 1 = \begin{cases} 3(2\mu-1)^2, \text{ if } M = 2; \\ 4\mu^2 + 1/2 + 14(\mu - 1/2)^2, \text{ if } M = 3; \\ (M-3)[(M-4)\mu + 6\mu^2 + 1] + s_1|_{M=3}, \text{ if } M \geq 4 \end{cases}$$

Hence $s_1 \geq 0$ for $M = 2, 3, \ldots$ and thus $\mathbf{I}_1 \leq 0$.

Case 2: $\mathbf{I}_2 \leq 0$

Here

$$\mathsf{n}(\mathbf{A}_2) = -4\left[a(M-p+1) - M\mu p\right]^2 s_2$$

where

$$s_2 = M^3\mu p - 4M^2\mu p^2 - M^3 a + 2M^2 ap + 5M^2\mu p + 2Map^2 - 3M^2\mu - 6Ma * p + 4M\mu p \\ + 3Ma - 3M\mu - 2ap + 2a,$$

and we need to show that $s_2 \geq 0$.

Assume first $M = 2$. Then (using the fact that $\mu = (1+a)/2$)

$$X_2 \geq 0 \Longleftrightarrow p \geq \frac{3a+3}{4a+2} \geq 1$$

which is impossible; so from now on $M \geq 3$.

To establish $\mathbf{I}_2 \leq 0$, it will suffice to demonstrate that

$$s_3 := 2Ms_2 - 2M^3(M\mu + a)X_2 \geq 0$$

as $\mathbf{I}_2$ has a factor $1_{X_2 \geq 0}$, and $s_2 1_{X_2\geq 0} \geq \frac{s_3}{2M} 1_{X_2 \geq 0}$. Substituting

$$p = \left[\frac{1}{2} + \frac{1}{2M}\right] + \left[\frac{1}{2} - \frac{1}{2M}\right] w$$

where $w \in [0,1)$ corresponding to the condition (2.4), we get

$$\begin{aligned} s_3 &= M\left(-2M^2\mu w^2 + M^2\mu w + 4M\mu w^2 + M^3 - 3M^2\mu - M\mu w - 2\mu w^2 + M^2 - M\mu + 2\mu\right) \\ &\quad - a\cdot(M-1)\left[M\left((M-1)^2 - (w-2)^2\right) + (1-w)\left(M^2 - w - 1\right)\right] \end{aligned}$$

The expression in the square brakets is non-negative for $M \geq 3$, so the minimum of s_3 is achieved when $a = 1$; i.e.

$$\begin{aligned} s_3 \geq s_3|_{a=1} &= -2M^3\mu w^2 + M^3\mu w + 4M^2\mu w^2 - 3M^3\mu + M^3 w - M^2\mu w + M^2 w^2 \\ &\quad - 2M\mu w^2 + 3M^3 - M^2\mu - 5M^2 w - 2Mw^2 + 2M^2 + 2M\mu + 4Mw + w^2 - 2M - 1 =: s_4 \end{aligned}$$

But

$$\frac{\partial s_4}{\partial \mu} = -M\left((3-w)M^2 + (1+w)M - 2 + 2(M-1)^2 w^2\right) < 0$$

so

$$s_4 \geq s_4|_{\mu=1} = (1-w)(M-1)(wM(2M-3) + M + w + 1) \geq 0.$$

Case 3: $\mathbf{I}_3 \leq 0$

Here

$$\mathsf{n}(\mathbf{A}_3) = -(M+1)(1-a)s_5$$

and it suffices to show that $s_5 \geq 0$. If $M = 2$, then $\mu = (a+1)/2$ and $p \geq 3/4$, so

$$\begin{aligned} s_5 = 3(3-2p)\big[&(1-a)^2(8p-5) + (32(1-a)^2 + 144a)(1-p)^4 \\ &+12(1-p)^2(4p + a(4ap + 10p - 3))\big] \geq 0. \end{aligned}$$

For $M \geq 3$, let $M = 3 + \delta$, $\delta = 0, 1, \ldots$. Then $s_5 = \sum_{i=0}^{5} e_{i+1}\delta^i$ where we will show that all $e_i \geq 0$. Indeed, we have

$$\begin{aligned}
e_1 = &-432a\mu p^5 - 1296\mu^2 p^5 + 288a^2p^4 + 2736a\mu p^4 - 144ap^5 + 432p^4\mu^2 - 432\mu p^5 - 1632p^3a^2 \\
&- 2880a\mu p^3 + 1200ap^4 + 4320\mu^2 p^3 + 2736\mu p^4 + 2624p^2a^2 - 1152a\mu p^2 - 3744ap^3 - 1728\mu^2 p^2 \\
&- 2880\mu p^3 + 288p^4 - 1536pa^2 + 4928p^2a - 1152p^2\mu - 1632p^3 + 768a^2 - 3072ap + 2624p^2 \\
&+ 1536a - 1536p + 768 \\
e_2 = &-288a\mu p^5 - 1296\mu^2 p^5 + 168a^2p^4 + 2208a\mu p^4 - 48ap^5 - 360p^4\mu^2 - 288\mu p^5 - 1160p^3a^2 \\
&- 1392a\mu p^3 + 600ap^4 + 6984\mu^2 p^3 + 2208\mu p^4 + 1760p^2a^2 - 3840a\mu p^2 - 2264ap^3 - 2016\mu^2 p^2 \\
&- 1392\mu p^3 + 168p^4 - 576pa^2 + 2720p^2a - 3840p^2\mu - 1160p^3 + 768a^2 - 1152ap + 1760p^2 \\
&+ 1536a - 576p + 768 \\
e_3 = &-48a\mu p^5 - 432\mu^2 p^5 + 24a^2p^4 + 576a\mu p^4 - 600p^4\mu^2 - 48\mu p^5 - 268p^3a^2 + 216a\mu p^3 + 72ap^4 \\
&+ 4404\mu^2 p^3 + 576\mu p^4 + 324p^2a^2 - 3240a\mu p^2 - 412ap^3 - 876\mu^2 p^2 + 216\mu p^3 + 24p^4 + 336pa^2 \\
&+ 180p^2a - 3240p^2\mu - 268p^3 + 288a^2 + 672ap + 324p^2 + 576a + 336p + 288 \\
e_4 = &-48\mu^2 p^5 + 48a\mu p^4 - 216p^4\mu^2 - 20p^3a^2 + 192a\mu p^3 + 1356\mu^2 p^3 + 48\mu p^4 - 4p^2a^2 - 1164a\mu p^2 \\
&- 20ap^3 - 168\mu^2 p^2 + 192\mu p^3 + 228pa^2 - 112p^2a - 1164p^2\mu - 20p^3 + 48a^2 + 456ap - 4p^2 \\
&+ 96a + 228p + 48
\end{aligned}$$

$$\begin{aligned}
e_5 = &-24p^4\mu^2 + 24a\mu p^3 + 204\mu^2 p^3 - 4p^2a^2 - 192a\mu p^2 - 12\mu^2 p^2 + 24\mu p^3 + 45pa^2 - 16p^2a \\
&- 192p^2\mu + 3a^2 + 90ap - 4p^2 + 6a + 45p + 3 \\
e_6 = &\ 360p(a + 1 - 2\mu p)^2 \geq 0.
\end{aligned}$$

The fact that $e_6 \geq 0$ is trivial; we will prove separately that $e_1, \ldots, e_5 \geq 0$ below. In what follows, we substitute $p = \frac{1+\nu}{2}$, where $\nu \in (0, 1)$.

Proof that $e_1 \geq 0$

We have

$$\frac{\partial^2 e_1}{\partial a^2} = 4[9\nu^4 - 66\nu^3 + 76\nu^2 + 2\nu + 235] > 0,$$

hence e_1 achieves its minimum at

$$a_{cr} = \frac{9\nu^5 - 105\nu^4 + 426\nu^3 - 46nu^2 + 397\nu - 1669 + 9\mu(1+\nu)^2(3\nu^3 - 29\nu^2 + 13\nu + 109)}{8[9\nu^4 - 66\nu^3 + 76\nu^2 + 2\nu + 235]}$$

which solves $\frac{\partial e_1}{\partial a} = 0$. Note that it is possible that $a_{cr} \notin [0, 1]$. However, in any case,

$$e_1 \geq e_1|_{a=a_{cr}} = \frac{1}{32} \cdot \frac{3(1+\nu)^2 c_1}{9\nu^4 - 66\nu^3 + 76\nu^2 + 2\nu + 235}$$

so it will suffice to show that

$$\begin{aligned}
c_1 &= 16(1-\nu)^2 c_{1a} + 3(1-\mu)c_{1b}, \quad \text{where} \\
c_{1a} &= -27\nu^6 + 144\nu^5 - 102\nu^4 + 1620\nu^3 - 9883\nu^2 + 12484\nu + 1732 \\
c_{1b} &= 81\mu\nu^8 - 108\mu\nu^7 + 135\nu^8 - 1260\mu\nu^6 - 828\nu^7 - 12276\mu\nu^5 + 276\nu^6 + 84774\mu\nu^4 \\
&\quad - 4404\nu^5 - 157140\mu\nu^3 + 69170\nu^4 + 152628\mu\nu^2 - 198372\nu^3 - 156108\mu\nu + 182084\nu^2 \\
&\quad + 27969\mu - 60588\nu + 73967
\end{aligned}$$

is positive. We have

$$c_{1a} = 3\nu^3(540 - 9\nu^3 + 48\nu^2 - 34\nu) + \nu(12484 - 9883\nu) + 1732 > 0.$$

Similarly,

$$c_{1b} = 61440(1-\mu) + (1-\nu)[c_{1b1} + c_{1b2}\mu]$$

where

$$\begin{aligned}
c_{1b1} &= (-135\nu^7 + 693\nu^6 + 417\nu^5 + 4821\nu^4) - 64349\nu^3 + 134023\nu^2 - 48061\nu + 12527 \\
&\geq -64349\nu^3 + 134023\nu^2 - 48061\nu + 12527 \geq 1000(-67\nu^3 + 134\nu^2 - 67\nu + 12) \\
&= \frac{1000}{27}\left[56 + 67(4-3\nu)(1-3\nu)^2\right] > 0
\end{aligned}$$

and

$$\begin{aligned}
c_{1b2} &= (-81\nu^7 + 27\nu^6 + 1287\nu^5 + 13563\nu^4) - 71211\nu^3 + 85929\nu^2 - 66699\nu + 894 \\
&\geq -71211\nu^3 + 85929\nu^2 - 66699\nu + 89409 > 80000(-\nu^3 + \nu^2 - \nu + 1) \geq 0.
\end{aligned}$$

So, $c_{1b1}, c_{1b2} > 0 \implies c_{1b} > 0$ and since $c_{1a} > 0$ we have $c_1 \geq 0$ and thus $e_1 \geq 0$.

Proof that $e_2 \geq 0$

We have

$$\frac{\partial^2 e_2}{\partial a^2} = 21\nu^4 - 206\nu^3 + 136\nu^2 + 398\nu + 1571 > 0$$

so, similarly to the previous case,

$$e_2 \geq e_2|_{a=a_{cr}} = \frac{3(1+\nu^2)[582912(1-\mu)^2 + (1-\nu)c_2]}{8[21\nu^4 - 206\nu^3 + 136\nu^2 + 398\nu + 1571]}$$

where

$$a_{cr} = \frac{3\nu^5 - 60\nu^4 + 296\nu^3 - 82\nu^2 - 155\nu - 2786 + \left(18\nu^3 - 222\nu^2 - 150\nu + 2010\right)(1+\nu)^2\mu}{2[21\nu^4 - 206\nu^3 + 136\nu^2 + 398\nu + 1571]}$$

solves $\frac{\partial e_2}{\partial a} = 0$ and

$$\begin{aligned} c_2 &= 3\nu^7 - 123\nu^6 + 1330\nu^5 - 1918\nu^4 - 28897\nu^3 + 65177\nu^2 + 93100\nu + 120544 \\ &+ \left(36\nu^7 - 624\nu^6 + 348\nu^5 + 25616\nu^4 - 7332\nu^3 - 272368\nu^2 - 134556\nu + 688784\right) \\ &+ \left(108\nu^7 - 72\nu^6 - 4848\nu^5 - 35916\nu^4 + 247548\nu^3 - 252720\nu^2 + 144456\nu - 647676\right)\mu^2 \end{aligned}$$

Now,

$$\frac{\partial^2 c_2}{\partial \mu^2} = \nu^4(216\nu^3 - 144\nu^2 - 9696\nu - 71832) + \nu^3(495096\nu - 505440) + (288912\nu - 1295352) < 0$$

hence the minimum of c_2 w.r.t. $\mu \in [0, 1]$ can be achieved either at $\mu = 0$ or at $\mu = 1$. At the same time

$$\begin{aligned} c_2|_{\mu=0} &= 3\nu^7 + 1330\nu^5 + 65177\nu^2 + 93100\nu + (120544 - 123\nu^6 - 1918\nu^4 - 28897\nu^3) > 0, \\ c_2|_{\mu=1} &= (1-\nu)(161652 - 147\nu^6 + 672\nu^5 + 3842\nu^4 + 16060\nu^3 + (264652 - 195259\nu)\nu) \geq 0, \end{aligned}$$

so $c_2 \geq 0$ and hence $e_2 \geq 0$.

Proof that $e_3 \geq 0$

We have

$$\frac{\partial^2 e_3}{\partial a^2} = 3\nu^4 - 55\nu^3 - 21\nu^2 + 471\nu + 1010 > 0$$

so, similarly to the previous case,

$$e_3 \geq e_3|_{a=a_{cr}} = \frac{3(1+\nu)^2\left[(1-\nu)^2 c_{3a} + (1-\mu)c_{3b}\right]}{8(3\nu^4 - 55\nu^3 - 21\nu^2 + 471\nu + 1010)}$$

where

$$a_{cr} = \frac{-9\nu^4 + 67\nu^3 + 165\nu^2 - 579\nu - 1820 + 3(1+\nu)^2\mu(\nu^3 - 21\nu^2 - 63\nu + 499)}{2(3\nu^4 - 55\nu^3 - 21\nu^2 + 471\nu + 1010)}$$

solves $\frac{\partial e_3}{\partial a} = 0$ and

$$\begin{aligned} c_{3a} &= -3\nu^6 + 12\nu^5 + 632\nu^4 + 1794\nu^3 - 37624\nu^2 + 65244\nu + 64877 > 0 \\ c_{3b} &= 2(1-\nu)(-3\nu^7 + 12\nu^6 + 652\nu^5 + 2417\nu^4 - 42561\nu^3 + 73864\nu^2 + 41336\nu + 91323) \\ &+ (1-\mu)(3\nu^6(220 - \nu^2 + 4\nu) + 2\nu(1490\nu^4 - 22993\nu^3 + 39898\nu^2 + 890\nu + 109262) + 8477) \geq 0 \end{aligned}$$

Hence $e_3 \geq 0$.

Proof that $e_4 \geq 0$

We have

$$\frac{\partial^2 e_4}{\partial a^2} = 209\nu + 317 - 5\nu^3 - 17\nu^2 > 0$$

so, similarly to the previous case,

$$e_4 \geq e_4|_{a=a_{cr}} = \frac{3(1+\nu)^2[(1-\nu)^2 c_{4a} + 4(1-\mu)c_{4b}]}{8(209\nu + 317 - 5\nu^3 - 17\nu^2)}$$

where

$$a_{cr} = \frac{5\nu^3 + 71\nu^2 - 329\nu - 587 + 6\mu(1+\nu)^2(88 - 10\nu - \nu^2)}{2(209\nu + 317 - 5\nu^3 - 17\nu^2)}$$

solves $\frac{\partial e_4}{\partial a} = 0$ and

$$\begin{aligned} c_{4a} &= 8\nu^4 + 40\nu^3 - 1395\nu^2 + 4354\nu + 4757 > 0 \\ c_{4b} &= 4(1-\nu)(4\nu^5 + 21\nu^4 - 712\nu^3 + 2011\nu^2 + 3102\nu + 3050) \\ &\quad + (1-\mu)(2\nu^6 + 11\nu^5 - 360\nu^4 + 912\nu^3 + 1705\nu^2 + 3655\nu + 543) \geq 0. \end{aligned}$$

Hence $e_4 \geq 0$.

Proof that $e_5 \geq 0$

We have

$$\frac{\partial^2 e_5}{\partial a^2} = 49 + 41\nu - 2\nu^2 > 0$$

so, similarly to the previous case,

$$e_5 \geq e_5|_{a=a_{cr}} = \frac{3(1+\nu)^2[(1-\nu)^2 c_{5a} + (1-\mu)c_{5b}]}{2(49 + 41\nu - 2\nu^2)}$$

where

$$a_{cr} = \frac{4\nu^2 - 37\nu - 47 + 3\mu(15-\nu)(1+\nu)^2}{49 + 41\nu - 2\nu^2}$$

solves $\frac{\partial e_5}{\partial a} = 0$ and

$$\begin{aligned} c_{5a} &= 15 - \nu^2 + 15\nu > 0 \\ c_{5b} &= 2(1-\nu)(14\nu^2 + 28\nu + 19 - \nu^3) + (1-\mu)(13\nu^3 + 40\nu^2 + 49\nu + 11 - \nu^4) \geq 0. \end{aligned}$$

Hence $e_5 \geq 0$.

As a result, $s_5 \geq 0$ and thus $\mathbf{I}_3 \leq 0$.

Case 4: $\mathbf{I}_4 \leq 0$

Here

$$\mathsf{n}(\mathbf{A}_4) = -4(M\mu p - M + p - 1)^2 s_6,$$
$$s_6 = 2p - 2 + (3\mu + 6p - 3 - 4\mu p - 2p^2)M + (4\mu p^2 - 5\mu p + 3\mu - 2p)M^2 + (1 - \mu p)M^3$$

Then, substituting $M = 2 + \delta$,

$$\frac{\partial s_6}{\partial \delta} = [5(1-\mu) + 2(1-p)(p + 2 + 10\mu - 8\mu p)] + [8(1-\mu) + 2(1-p)(2 + 7\mu - 4\mu p)]\delta \geq 0$$

and as a result for $\delta \geq 0$ we have

$$s_6 \geq s_6|_{\delta=0} = 2(3-2p)[p(1-\mu) + \mu(1-3p)] \geq 0.$$

Case 5: $\mathbf{I}_5 \leq 0$

Here

$$\mathsf{n}(\mathbf{A}_5) = -s_7.$$

We need to show that $s_7 \geq 0$ when $X_1 \leq 0$ and $X_3 \geq 0$.

Since $X_1 \leq 0$, we have $2Mp\mu \leq M + 1$. Together with $X_3 \geq 0$ this implies

$$0 \leq \mathsf{n}(X_3) = 2Mp\mu - (M+1) - a(M+1) + 2p \leq -a(M+1) + 2p$$

whence

$$a \leq \frac{2p}{M+1}.$$

Let us show that for this a we have $s_7 \geq 0$; substitute $a = b \cdot \frac{2p}{M+1}$, where $b \in [0, 1]$.

First, let $\boxed{\text{M=2}}$, then $\mu = \frac{1+a}{2}$, $p \in [3/4, 1)$, and $s_7 = \frac{3-2p}{27} s_8$ where

$$s_8 = 512b^3p^8 - 2688b^3p^7 + 5760b^3p^6 + 3456b^2p^7 - 6912b^3p^5 - 12672b^2p^6 + 5184b^3p^4$$
$$+ 16416b^2p^5 + 5184bp^6 - 1944b^3p^3 - 11664b^2p^4 - 10368bp^5 + 7776b^2p^3 + 1728p^5 - 2916b^2p^2$$
$$+ 11664bp^3 - 11664bp^2 - 7776p^3 + 4374bp + 17496p^2 - 17496p + 6561$$

Note that we can write $s_8 = e_1 + e_2(1-p) + e_3(1-p)^2$, where

$$128e_1 = (9 - \nu^2 - 6\nu)(81 - \nu^3 - 9\nu^2 - 63\nu)(\nu^3 + 15\nu^2 + 81 - 9\nu) > 0$$
$$128e_2 = 3(9 - \nu^2)(\nu^6 + 21\nu^5 + 168\nu^4 + 666\nu^3 + 81\nu^2 + 81\nu + 486) > 0$$
$$64(e_1 + e_3) = [2\nu^8 + 33\nu^7 + 234\nu^6 + 783\nu^5] + [-648\nu^4 - 6561\nu^3 + 30618\nu^2 - 28431\nu + 13122]$$
$$\geq -648\nu^4 - 6561\nu^3 + 30618\nu^2 - 28431\nu + 13122$$
$$\geq -1000\nu^4 - 7000\nu^3 + 24000\nu^2 - 29000\nu + 13000$$

$$= 1000(1-\nu)(5+8(1-\nu)^2+\nu^3) \geq 0.$$

with $p = \frac{3+\nu}{4}$, $\nu \in [0,1]$. Consequently, since $(1-p)^2 < 1$ and $e_1 > 0$,

$$s_8 = e_1 + e_2(1-p) + e_3(1-p)^2 \geq e_2(1-p) + (e_1+e_3)(1-p)^2 \geq 0$$

and thus $s_7 \geq 0$ as required.

For $\boxed{M \geq 3}$, set $M = 3+\delta$, $\delta \geq 0$. Then

$$s_7 = \sum_{i=0}^{9} e_{i+1}\delta^i$$

where

$$\begin{aligned}
e_1 &= 196608 + (98304b - 393216)p + (-49152b^2 - 442368\mu^2 - 196608b - 737280\mu + 589824)p^2 \\
&\quad + (-24576b^3 + 221184b\mu^2 + 98304b^2 + 1990656\mu^2 + 294912b + 663552\mu - 540672)p^3 \\
&\quad + (49152b^3 + 73728b^2\mu - 552960b\mu^2 - 331776\mu^3 - 147456b^2 - 2322432\mu^2 - 270336b \\
&\qquad + 110592\mu + 233472)p^4 \\
&\quad + (-83968b^3 + 184320b^2\mu - 55296b\mu^2 + 774144\mu^3 + 135168b^2 + 1050624\mu^2 + 116736b \\
&\qquad - 239616\mu - 36864)p^5 \\
&\quad + (52224b^3 - 175104b^2\mu + 165888b\mu^2 - 331776\mu^3 - 58368b^2 - 165888\mu^2 - 18432b + 55296\mu)p^6 \\
&\quad + (-9216b^3 + 27648b^2\mu + 9216b^2)p^7
\end{aligned}$$

$$\begin{aligned}
e_2 &= 393216 + (172032b - 540672)p + (-73728b^2 - 958464\mu^2 - 221184b - 2088960\mu + 835584)p^2 \\
&\quad + (-30720b^3 + 423936b\mu^2 + 86016b^2 + 4589568\mu^2 + 344064b + 1898496\mu - 823296)p^3 \\
&\quad + (30720b^3 + 282624b^2\mu - 1308672b\mu^2 - 663552\mu^3 - 135168b^2 - 5031936\mu^2 - 344064b \\
&\qquad - 18432\mu + 344064)p^4 \\
&\quad + (-77312b^3 + 181248b^2\mu + 4608b\mu^2 + 1658880\mu^3 + 138240b^2 + 2068992\mu^2 + 142848b \\
&\qquad - 360960\mu - 49152)p^5 \\
&\quad + (50176b^3 - 228864b^2\mu + 290304b\mu^2 - 691200\mu^3 - 56832b^2 - 290304\mu^2 - 19968b + 78336\mu)p^6 \\
&\quad + (-7680b^3 + 32256b^2\mu + 7680b^2)p^7
\end{aligned}$$

$$\begin{aligned}
e_3 &= 344064 + (129024b - 208896)p + (-46080b^2 - 906240\mu^2 - 49152b - 2558976\mu + 430080)p^2 \\
&\quad + (-15360b^3 + 347136b\mu^2 + 3072b^2 + 4718592\mu^2 + 129024b + 2217984\mu - 522240)p^3 \\
&\quad + (-6144b^3 + 334848b^2\mu - 1337856b\mu^2 - 566784\mu^3 - 30720b^2 - 4778496\mu^2 - 175104b \\
&\qquad - 198144\mu + 210432)p^4
\end{aligned}$$

$$+ (-24448b^3 + 42240b^2\mu + 100992b\mu^2 + 1543680\mu^3 + 52992b^2 + 1744512\mu^2 + 69504b - 223872\mu - 26112)p^5$$
$$+ (17856b^3 - 118464b^2\mu + 210816b\mu^2 - 615168\mu^3 - 20544b^2 - 210816\mu^2 - 8064b + 44160\mu)p^6$$
$$+ (-2112b^3 + 14016b^2\mu + 2112b^2)p^7$$

$$e_4 = 172032 + (53760b + 64512)p + (-15360b^2 - 488448\mu^2 + 44544b - 1790976\mu + 64512)p^2$$
$$+ (-3840b^3 + 157440b\mu^2 - 23040b^2 + 2843904\mu^2 + 1416960\mu - 176640)p^3$$
$$+ (-9984b^3 + 193536b^2\mu - 772608b\mu^2 - 268032\mu^3 + 7680b^2 - 2598912\mu^2 - 44544b - 170496\mu + 68352)p^4$$
$$+ (93024b\mu^2 - 13632b^2\mu + 32\mu(25448\mu^2 + 25515\mu - 2283) - 32b(77b^2 - 282b - 525) - 6912)p^5$$
$$+ (2784b^3 - 30336b^2\mu + 81312b\mu^2 - 303168\mu^3 - 3264b^2 - 81312\mu^2 - 1440b + 12384\mu)p^6$$
$$+ (-192b^3 + 2688b^2\mu + 192b^2)p^7$$

$$e_5 = 53760 + (13440b + 91392)p + (-2880b^2 - 164160\mu^2 + 34560b - 792768\mu - 26880)p^2$$
$$+ (-480b^3 + 42720b\mu^2 - 11520b^2 + 1109088\mu^2 - 13440b + 548640\mu - 33600)p^3$$
$$+ (62496b^2\mu - 275952b\mu^2 - 885744\mu^2 - 67536\mu - 75792\mu^3 - 96b(34b^2 - 50b + 59) + 12432)p^4$$
$$+ (160b^3 - 8544b^2\mu + 38928b\mu^2 + 266192\mu^3 + 576b^2 + 229104\mu^2 + 2016b - 13200\mu - 912)p^5$$
$$+ (160b^3 - 3840b^2\mu + 17568b\mu^2 - 89344\mu^3 - 192b^2 - 17568\mu^2 - 96b + 1728\mu)p^6 + 192b^2\mu p^7$$

$$e_6 = 10752 + (2016b + 38976)p + (-288b^2 - 35232\mu^2 + 10848b - 230880\mu - 14784)p^2$$
$$+ (-24b^3 + 6936b\mu^2 - 2544b^2 + 290664\mu^2 - 4032b + 132888\mu - 3408)p^3$$
$$+ (11568b^2\mu - 456b^3 - 62472b\mu^2 - 12816\mu^3 + 816b^2 - 193752\mu^2 - 288b - 14328\mu + 1200)p^4$$
$$+ (32b^3 - 1536b^2\mu + 8688b\mu^2 + 55168\mu^3 + 38544\mu^2 + 96b - 1248\mu - 48)p^5$$
$$+ (-192b^2\mu + 2016b\mu^2 - 15744\mu^3 - 2016\mu^2 + 96\mu)p^6$$

$$e_7 = 1344 + (168b + 9072)p + (-12b^2 - 4716\mu^2 + 1824b - 44340\mu - 3024)p^2$$
$$+ (624b\mu^2 - 276b^2 + 51252\mu^2 - 504b + 19764\mu - 144)p^3$$
$$+ (-24b^3 + 1152b^2\mu - 8760b\mu^2 - 1200\mu^3 + 48b^2 - 26568\mu^2 - 1584\mu + 48)p^4$$
$$+ (-96b^2\mu + 1008b\mu^2 + 7072\mu^3 + 3600\mu^2 - 48\mu)p^5 + (96b\mu^2 - 1536\mu^3 - 96\mu^2)p^6$$

$$e_8 = 96 + (6b + 1236)p + (-360\mu^2 + 162b - 5424\mu - 300)p^2$$
$$+ (24b\mu^2 - 12b^2 + 5868\mu^2 - 24b + 1656\mu)p^3 + (48b^2\mu - 696b\mu^2 - 48\mu^3 - 2088\mu^2 - 72\mu)p^4$$
$$+ (48b\mu^2 + 512\mu^3 + 144\mu^2)p^5 - 64\mu^3p^6$$

and the expressions for e_9 and e_{10} are given a little bit further.

First, we will show that $e_i \geq 0$, $i = 1, \ldots, 8$.

Proof that $e_1, \dots, e_8 > 0$

It turns out that it is easiest is to use a computer-assisted proof in this case; to this end we developed the method which we call a *Box method*; it may have been described by other authors, but since we do not have the reference to the right source, we give its description below.

First of all, we substitute

$$p = \frac{1+x_1}{2},\ b = x_2, \mu = x_3; \quad x_i \in [0,1],\ i = 1,2,3.$$

Let $\mathrm{m} = \min_{a_i \le x_i \le b_i, i=1,2,3} f(x_1, x_2, x_3)$ where

$$f(x_1, x_2, x_3) = f_+(x_1, x_2, x_3) - f_-(x_1, x_2, x_3)$$

and f_+ and f_- are polynomials with non-negative coefficients. We want to show that $\mathrm{m} > 0$.

Let

$$G_{f;M} = \min_{i_1, i_2, i_3 = 0, \dots, M-1} \left[f_+\left(\frac{i_1}{M}, \frac{i_2}{M}, \frac{i_3}{M}\right) - f_-\left(\frac{i_1+1}{M}, \frac{i_2+1}{M}, \frac{i_3+1}{M}\right)\right].$$

Since

$$\mathrm{m} \ge G_{f;M} \to \mathrm{m}$$

as $M \to \infty$, we conclude that $\mathrm{m} > 0$ if and only if $G_{f,M} \ge 0$ for some $M \ge 1$. Checking that $G_{f,M} \ge 0$ can be quite tedious and time-consuming for large M, however, this could be easily accomplished with the help of a computer; please note, that the results are still *completely rigorous*, unlike e.g. simulations.

The results of application of this method to $e_1, \dots, e_8$ are presented in the following table:

$$G_{e_1,2000} > 825, \quad G_{e_2,500} > 25, \quad G_{e_3,400} > 1860, \quad G_{e_4,300} > 2397,$$
$$G_{e_5,200} > 672, \quad G_{e_6,200} > 148, \quad G_{e_7,200} > 5, \quad G_{e_8,400} > 3.$$

Consequently, $e_j > 0$ for all $j = 1, \dots, 8$.

Proof that $e_9 \ge 0$ and $e_{10} \ge 0$

The Box method of the previous section would not work for e_9 and e_{10}, since these functions do touch zero in the required area, and hence the minimum is, in fact, 0. Therefore, we have to handle these two cases analytically.

We have

$$e_9 = 4p^2\mu(4\mu^2p^3 - 18\mu p^2 + 99\mu p - 3 + 15p - 96) - 12p^2 + 93p + 3 + [6p^2(1-2\mu p)(2\mu p + 1)]b,$$

hence, the minimum is achieved either at $b = 0$ or $b = 1$.

For $\mu < 1/(2p)$ we have $e_9 \geq e_{9a}$, where

$$e_{9a} = e_9|_{b=0} = 2s^3p^2 - 18p^2s^2 + 30sp^2 + 99s^2p - 12p^2 - 192ps - 3s^2 + 93p + 3$$
$$= 2p^2 + (1-s)[6(1-p) + (1-s)(99p + 2p^2s - 14p^2 - 3)] \geq 0$$

where $s = 2p\mu \in [0, 1]$.

In case $\mu \geq 1/(2p)$ we have $e_9 \geq e_{9b}$, where

$$e_{9b} = e_9|_{b=1} = 16p^5s^3 - 24p^4s^3 - 72p^4s^2 + 12p^3s^3 + 468p^3s^2 - 2s^3p^2 - 24p^3s - 426p^2s^2$$
$$+ 24sp^2 + 111s^2p + 2p^2 - 18ps - 3s^2 + 6s$$

where $\mu = \frac{1}{2p} + s\left(1 - \frac{1}{2p}\right)$, $s \in [0, 1]$. Now,

$$\frac{\partial^2}{\partial s^2} e_{9b} = 6(2p-1)^2(14 + (2p-1)(2p^2s - 3p + 15)) \geq 0$$

so the minimum of e_{9b} w.r.t. s is achieved where $\frac{\partial}{\partial s} e_{9b} = 0$, i.e.

$$s_{cr} = \frac{6p^2 - 33p + 1 + R}{2p^2(2p-1)}, \quad \text{where} \quad R = \sqrt{44p^4 - 400p^3 + 1105p^2 - 66p + 1}$$

and equals

$$\frac{3996p^5 - 284p^6 - 19956p^4 + 37329p^3 - 3291p^2 + 99p - 1 + (400p^3 - 44p^4 - 1105p^2 + 66p - 1)R}{2p^4}$$

$$\geq 22120.5 - 1576\sqrt{197} = 0.285896\dot{>}0$$

for $p \geq 1/2$.

Finally, trivially, we have $e_{10} = 3p(2\mu p - 1)^2 \geq 0$. Consequently, $s_7 \geq 0$ and $\mathbf{I}_5 \leq 0$.

Combining this with the previously established inequalities $\mathbf{I}_j \leq 0$, $j = 1, 2, 3, 4$, we complete the proof Lemma 2. }

6 Alternative proof of Theorem 2

Proof. Assume for now that $p < \frac{N}{2}$; in this case $N - \frac{N}{2p} + 1 < N$ (see (3.6)). The case $p \geq \frac{N}{2}$ will be dealt with separately.

Claim 5. *Let* $A_i := \left\{x_{(i)}(t) \to 0\right\}$ *and suppose that for some* $1 \le k \le N-2$ *we have*

$$\{2p(N-k-1) < N\} \Longleftrightarrow \left\{k > N - \frac{N}{2p} - 1\right\}. \tag{6.16}$$

Then $A_k \subseteq \{\exists \lim_{t\to\infty} x_{(k+1)}(t)\}$.

Proof. Fix any $a > 0$. Let $\delta > 0$ be so small that

$$2pN\delta < [N - 2p(N-k-1)]a. \tag{6.17}$$

In the event A_k there exists a finite $\tau = \tau_\delta(\omega)$ such that

$$\left\{\sup_{t\ge\tau} x_{(k)}(t) \le \delta\right\} \Longleftrightarrow \{\text{card } (\mathcal{X}'(t) \cap [0,\delta]) \ge k \text{ for all } t \ge \tau.\}$$

From now on assume that $t \ge \tau$. We will show below that $x_{(k+1)}(t+1) \le \max\{x_{(k+1)}(t), a\}$.

To begin, let us prove that $x_{(k+1)}(t+1) \le x_{(k+1)}(t)$ as long as $x_{(k+1)}(t) > \delta$. Indeed, if the new point ζ is sampled to the left of $x_{(k+1)}(t)$, then regardless of which point is to be removed, $x_{(k+1)}(t+1) \le x_{(k+1)}(t)$. If the new point ζ is sampled to the right, then the farthest point from the $p-$centre of mass must be the rightmost one (and hence $x_{(k+1)}(t+1) = x_{(k+1)}(t)$) since there are exactly k points in $[0,\delta]$ and none of these can be removed by the definition of τ.

On the other hand, if $x_{(k+1)}(t) \le \delta$ then either $x_{(k+2)}(t) \le a$ or $x_{(k+2)}(t) > a$. In the first case, $x_{(k+1)}(t+1) \le x_{(k+2)}(t) \le a$ even if $x_{(1)}$ is removed. In the other case, when $x_{(k+2)}(t) > a$, we have $x_{(N-1)} > a$ as well, and

$$\begin{aligned} p\mu(\mathcal{X}(t+1)) &\le p\frac{(k+1)\delta + (N-k-1)x_{(N)}}{N} < \frac{2pN\delta - [N-2p(N-k-1)]x_{(N)} + Nx_N}{2N} \\ &\le \frac{Nx_N - \{[N-2p(N-k-1)]a - 2pN\delta\}}{2N} < \frac{x_{(N)}}{2} \end{aligned}$$

by (6.17), so $x_{(N)} = x_{(N)}(t)$ must be removed and thus $x_{(k+1)}(t+1) \le x_{(k+1)}(t)$.

Consequently, we obtained

$$\begin{aligned} A_k &\subseteq \bigcap_{t\ge\tau} \{x_{(k+1)}(t+1) \le \max\{x_{(k+1)}(t), a\}\} \\ &\subseteq \left(\bigcup_{t\ge0} \{x_{(k+1)}(s) \le a \text{ for all } s \ge t\}\right) \cup \left(\bigcup_{t\ge0} \{x_{(k+1)}(s) \le x_{(k+1)}(s+1) \text{ for all } s \ge t\}\right) \\ &\subseteq \left\{\limsup_{t\to\infty} x_{(k+1)}(t) \le a\right\} \cup \left\{\exists \lim_{t\to\infty} x_{(k+1)}(t)\right\} \end{aligned}$$

since $a > 0$ is arbitrary, we get

$$A_k \subseteq \left\{\limsup_{t\to\infty} x_{(k+1)}(t) \le 0\right\} \cup \left\{\exists \lim_{t\to\infty} x_{(k+1)}(t)\right\} = \left\{\exists \lim_{t\to\infty} x_{(k+1)}(t) \ge 0\right\}$$

□

Lemma 13. *Suppose that* (6.16) *holds for some* $1 \le k \le N-2$*. Then* $A_k \underset{a.s.}{\Longrightarrow} A_{k+1}$*.*

Proof. Let $\tilde{A}^{\ge a}_{k+1} := \left\{\lim_{t\to\infty} x_{(k+1)}(t) \ge a\right\}$ (the existence of this limit on A_k follows from Claim 5). It suffices to show that $\mathbb{P}\left(A_k \cap \tilde{A}^{\ge a}_{k+1}\right) = 0$ for all $a > 0$; then from the continuity of probability we get that $\mathbb{P}\left(A_k \cap \{\lim_{t\to\infty} x_{(k+1)}(t) > 0\}\right) = 0$ and hence $A_k \underset{\text{a.s.}}{\Longrightarrow} A_{k+1}$.

Fix an $a > 0$. Let

$$C_t = \left\{x_{(k)}(t) < \frac{a}{3} \text{ and } x_{(k+1)}(t) > \frac{2a}{3}\right\}, \qquad \bar{C}_T = \bigcap_{t\ge T} C_t,$$

then

$$A_k \cap \tilde{A}^{\ge a}_{k+1} \subseteq \bigcup_{T\ge 0} \bar{C}_T = \left\{\exists T > 0 :\ x_{(k)}(t) < \frac{a}{3} \text{ and } x_{(k+1)}(t) > \frac{2a}{3} \text{ for all } t \ge T\right\}.$$

If the probability of the LHS is positive, then, using the continuity of probability and the fact that $\bar{C}_T$ is an increasing sequence of events, we obtain that $\lim_{T\to\infty} \mathbb{P}(\bar{C}_T) > 0$. Consequently, there exists a *non-random* T_0 such that $\mathbb{P}(\bar{C}_{T_0}) > 0$.

This is, however, impossible, as at each time point t, with probability at least $f(a/3)$ (see (2.1)) the new point ζ_t is sampled in $B := \left(\frac{a}{3}, \frac{2a}{3}\right)$ and then either $x_{(k)}(t+1) \in B$ or $x_{(k+1)}(t+1) \in B$. Formally, this means that

$$\mathbb{P}(C_{t+1}|C_t, \mathcal{F}_t) \le 1 - f(a/3) \quad \text{for all } t \ge 0.$$

By induction, for all $k \ge 1$,

$$\mathbb{P}(\bar{C}_{T_0}|\mathcal{F}_{T_0}) \le \mathbb{P}\left(\bigcap_{T=T_0}^{T_0+k} C_t|\mathcal{F}_{T_0}\right) \le \left[1 - f(a/3)\right]^k.$$

Since k is arbitrary, and $f(a/3) > 0$, by taking the expectation, we conclude that $\mathbb{P}(\bar{C}_{T_0}) = 0$ yielding a contradiction.

Hence the probability of the event $A_k \cap \tilde{A}^{\ge a}_{k+1}$ is zero. □

Corollary 3. *Suppose that* (6.16) *holds for some* $1 \le k \le N-2$. *Then*

$$\left\{x_{(k)}(t) \to 0\right\} \underset{a.s.}{\Longrightarrow} \left\{\mathcal{X}'(t) \to 0\right\}.$$

Proof. Observe that if k satisfies (6.16) then $k+1$ satisfies (6.16) as well. Thus by iterating Lemma 13 we obtain that $A_k \underset{\text{a.s.}}{\Longrightarrow} A_{k+1} \underset{\text{a.s.}}{\Longrightarrow} A_{k+2} \underset{\text{a.s.}}{\Longrightarrow} \dots \underset{\text{a.s.}}{\Longrightarrow} A_{N-1}$, i.e. $x_{(N-1)}(t) \to 0$, which is equivalent to the statement of Corollary. □

Remark 6. *Note that the condition* (6.16) *does not assume* $p > 1$*; hence the conclusion of Corollary 3 holds for the case* $0 < p \le 1$ *as well.*

For the case $p \ge \frac{N}{2}$ we have

Lemma 14. *If* $p \ge \frac{N}{2}$ *then* $\mathcal{X}'(t) \to 1$ *a.s.*

Proof. The case $p > \frac{N}{2}$ is easy: with a positive probability the newly sampled point $\zeta > 0$ and then

$$p\,\frac{x_{(1)} + \dots + x_{(N-1)} + \zeta}{N} > \frac{x_{(1)} + \dots + x_{(N-1)} + \zeta}{2} \ge \frac{x_{(1)} + \zeta}{2}$$

hence it is the left-most point which is always removed, implying $\liminf_{t\to\infty} x_{(1)}(t) > 0$. Hence by Corollary 1, $\mathcal{X}'(t) \to 1$ a.s.

For the case $p = \frac{N}{2}$ we notice that at each moment of time we either have a tie (between the left-most and right-most point) or remove the left-most point. However, we can only have a tie if $x_{(1)}(t) = ... = x_{(N-1)}(t) = 0$; in this case, eventually the right-most point will be kept and the left-most removed. After this moment of time, there will be more ties, and the left-most point will always be removed, leading to the same conclusion as in the case $p > N/2$. □

Part (b) follows from Lemma 3.

To prove part (c), note that unless $x_{(1)}(0) > 0$ already, by repeating the arguments from the beginning of the proof of Lemma 7, with a positive probability we can "drag" the whole configuration in at most $N-1$ steps to the right of zero, that is, there is $0 \le t_0 \le N-1$ such that $\mathbb{P}(\min \mathcal{X}'(t_0) > 0) > 0$. Now we can apply Lemma 4 and then Lemma 3.

Let us now prove part (a). First, assume $p < \frac{N}{2}$. It is always possible to find an integer k which satisfies both (3.6) and (6.16), so let k be such that

$$N - \frac{N}{2p} - 1 < k < N - \frac{N}{2p} + 1$$

(if $N/(2p) \in \mathbb{N}$ this k will be unique). Now the statement of the theorem follows from Corollary 3 and Lemma 7.

Finally, in case $p \geq \frac{N}{2}$ the theorem follows from Lemma 14. $\square$

Acknowledgement

We would like to thank the anonymous referees for very careful reading our manuscript and for giving many useful suggestions and corrections.

7 Bibliography

[1] Seel, Christian; Tsakas, Elias. Rationalizability and Nash equilibria in guessing games. *Games Econom. Behav.* **106** (2017), 75–88.

[2] E. De Giorgi and S. Reimann, The α-beauty contest: Choosing numbers, thinking intervals, *Games Econom. Behav.* **64** (2008) 470–486.

[3] Grinfeld, Michael; Volkov, Stanislav; Wade, Andrew R. Convergence in a multidimensional randomized Keynesian beauty contest. Adv. in Appl. Probab. **47** (2015), 57–82.

[4] M. Grinfeld, P.A. Knight, and A.R. Wade, Rank-driven Markov processes, *J. Stat. Phys.* **146** (2012), 378–407.

[5] Kennerberg, Philip; Volkov, Stanislav. Jante's law process. Adv. in Appl. Probab. **50** (2018), 414–439.

[6] J.M. Keynes, *The General Theory of Employment, Interest and Money*, Macmillan, London, 1936.

[7] Williams, D. Probability with martingales, Cambridge University Press, 1991.

[8] H. Moulin, *Game Theory for the Social Sciences*, 2nd ed., New York University Press, New York, 1986.

C

Chapter 4

Paper C

A local barycentric version of the Bak-Sneppen model

Philip Kennerberg* and Stanislav Volkov[1]

1st January 2021

[1]Centre for Mathematical Sciences, Lund University, Box 118 SE-22100, Lund, Sweden

Abstract

We study the behaviour of an interacting particle system, related to the Bak-Sneppen model and Jante's law process defined in [8]. Let $N \geq 3$ vertices be placed on a circle, such that each vertex has exactly two neighbours. To each vertex assign a real number, called *fitness*[1]. Now find the vertex which fitness deviates most from the average of the fitnesses of its two immediate neighbours (in case of a tie, draw uniformly among such vertices), and replace it by a random value drawn independently according to some distribution ζ. We show that in case where ζ is a finitely supported or continuous uniform distribution, all the fitnesses except one converge to the same value.

Keywords: Bak-Sneppen model, Jante's law process, interacting particle systems.

Subject classification: 60J05, 60K35, 91D10.

1 Introduction

The model we study in the current paper is a "marriage" between Jante's law process and the Bak-Sneppen model.

Jante's law process refers to the interacting particle model studied in [6] under the name "Keynesian beauty contest process", and generalized in [8]. This model runs as follows. Fix an integer $N \geq 3$, $d \geq 1$, and some d-dimensional random variable ζ. Let the initial configuration consist of N arbitrary points in $\mathbb{R}^d$. The process runs in discrete time according to the following algorithm: first, compute the centre of mass μ of the given configuration of N points; then replace the point which is the most distant from μ by a new $\zeta-$distributed point drawn independently each time. It was shown in [6] that if ζ has a uniform distribution on the unit cube, then all but one points converge to some random point in $\mathbb{R}^d$. This result was further generalized in [8], by allowing ζ to have an arbitrary distribution, and additionally removing not just 1, but $K \geq 1$ points chosen to minimize a certain functional. The term "Jante's law process" was also coined in [8], to reflect that this process is reminiscent of the "Law of Jante" principle, which describes patterns of group behaviour towards individuals within Scandinavian countries that criticises individual success and achievement as unworthy and inappropriate; in other words, it is better to be "like everyone else". The origin of this "law" dates back to

[1] we use this term, as it is quite standard for Bak-Sneppen models

Aksel Sandemose [12]. Another modification of this model in one dimension, called the p-contest, was introduced in [6, 7] and later studied e.g. in [9]. This model runs as follows: fix some constant $p \in (0,1) \cup (1,\infty)$, and replace the point which is the farthest from $p\mu$ (rather than μ).

Finally, we want to mention that the phenomenon of conformity is observed in many large social networks, see, for example, [4, 10, 13] and references therein.

Pieter Trapman (2018, personal communications) suggested to study Jante's law model with *local interactions*, thus making it somewhat similar to the famous Bak-Sneppen (BS) model see e.g. [1]. In the BS model, N species are located around a circle, and each of them is associated with a so-called "fitness", which is a real number. The algorithm consists in choosing the least fit individual, and then replacing it *and both of its two closest neighbours* by a new species, with a new random and independent fitness. After a long time, there will be a minimum fitness, below which species do not survive. The model proceeds through certain events, called "avalanches", until it reaches a state of relative stability where all fitnesses are above a certain threshold level. There is a version of the model where fitnesses take only values 0 and 1 (see [2] and [15]), but even this simplified version turns out to be notoriously difficult to analyse, see e.g. [11]. Some more recent results can be found in [3, 14].

The barycentric Bak-Sneppen model, or, equivalently, Jante's law process with local interactions, is defined as follows. Unlike the classical Bak-Sneppen model, our model is based on some *local phenomena*, which makes it much more tractable mathematically, and hence we are able to obtain substantial rigorous results.

Fix an integer $N \geq 3$, and let $S = \{1, 2, \dots, N\}$ be the set of nodes uniformly spaced on a circle. At time t, each node $i \in S$ has a certain "fitness" $X_i(t) \in \mathbb{R}$; let $X(t) = (X_1(t), \dots, X_N(t))$. Next, for the vector $x = (x_1, \dots, x_N)$, define

$$d_i(x) = \left| x_i - \frac{x_{i+1} + x_{i-1}}{2} \right|,$$

as the measure of local "non-conformity" of the fitness at node i (here and further we will use the convention that $N+1 \equiv 1$, $N+2 \equiv 2$, and $1-1 \equiv N$ for indices on x). Let also $d(x) = \max_{i \in S} d_i(x)$.

The process runs as follows. Let ζ be some fixed one-dimensional random variable. At time t, $t = 0, 1, 2, \dots$, we chose the "least conformist node"[2] i, i.e. the one maximizing

[2]The intuition for choosing the deviance as the criteria for removal is the follows. In many Scand-

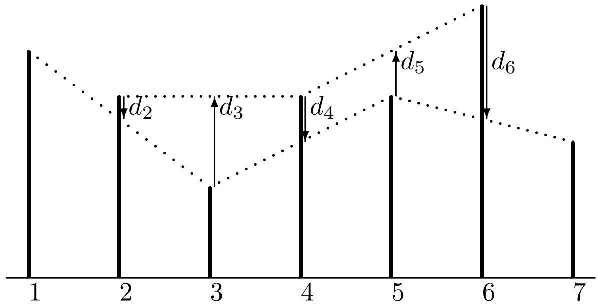


Figure 1: Illustration of the distances from the average of the two neighbours; $\jmath = 6$.

$d_i(X(t))$, and replace it by a ζ-distributed random variable. By $\jmath(x)$ we denote the index of such a node in the configuration $x = (x_1, \ldots, x_N)$, that is

$$d_{\jmath(x)}(x) = d(x)$$

(see Figure 1). If there is more than one such node, we choose any of them with equal probability, thus $\jmath(x)$ is, in general, a random variable. Also assume that all the coordinates of the initial configuration $X(0)$ lie in the support of ζ. We are interested in the long-term dynamics of this process.

We start with a somewhat easier version of the problem, where ζ takes finitely many distinct values (Section 2), and then extend this result to the case where $\zeta \sim U[0, 1]$ (Section 3). We will show that all the fitnesses (except the one which has just been updated) converge to the same (random) value. This will hold for each of the two models.

Remark 1. *One can naturally extend this model to any finite connected non-oriented graph G with vertex set V, as follows. For any two vertices $v, u \in V$ that are connected by an edge we write $u \sim v$. To each vertex v assign a fitness $x_v \in \mathbb{R}$, and define the*

---

inavian countries, non-conformity is considered as a very bad treat, and as a result, individuals which divert from the average, tend to be less successful in these societies. This phenomenon is called "The Jante's Law". We understand that the word "fitness" is thus somewhat misleading here, but would like to use it to keep in line with the standard Bak-Sneppen model.

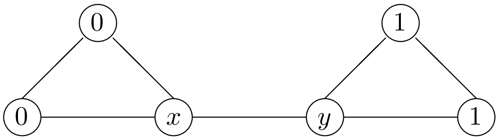


Figure 2: On this graph with $N = 6$ vertices, only values x and $y \in \{0, 1\}$ are updated all the time; infinitely often half of the fitnesses equal 0, while the other half equals 1.

measure of non-conformity of this vertex as

$$d_v(x) = \left| x_v - \frac{\sum_{u:\ u \sim v} x_u}{N_v} \right|,$$

where $N_v = |u \in V : u \sim v|$ *denotes the number of neighbours of* v*, and the replacement algorithm runs exactly as it is described earlier.*

In particular, if G is a cycle graph, we obtain the model studied in the current paper. On the other hand, if G is a complete graph, we obtain the model equivalent to that studied in [6, 8].

Remark 2. *Unfortunately, our results cannot be extended to a general model, described in Remark 1. Indeed, assume that* $\operatorname{supp} \zeta = \{0, 1\}$*. It is not hard to show that if for some* v *we have* $N_v = 1$*, then the statement of Theorem 1 does not have to hold.*

Moreover, it turns out that even when all the vertices have at least two neighbours (i.e., $N_v \geq 2$ *for all* $v \in V$*), then there are still counterexamples: please see Figure 2.*

The rest of the paper is organized as follows. In Section 2 we study the easier, discrete, case. We show the convergence by explicitly finding all the absorbing classes for the finite-state Markov chain.

Section 3 contains the main result of our paper, Theorem 2, which shows that all but one fitness converge to the same (random) limit, similarly to the main result of [6].

2 Discrete case

In this Section we study the case when fitnesses take finitely many values, equally spaced between each other. Due to the shift- and scale-invariance of the model, without loss of generality we may assume that $\operatorname{supp}\zeta = \{1, 2, \ldots, M\} =: \mathcal{M}$, and that $p = \min_{j\in\mathcal{M}} \mathbb{P}(\zeta = j) > 0$. In this case $X(t)$ becomes a finite state-space Markov chain on $\mathcal{M}^N$.

Note that if $N - 1$ fitnesses coincide and are equal to some $a \in \mathcal{M}$, then it is the fitness that differs from a that will keep being replaced, until it finally coincides with the others. When this happens, we will have to choose randomly one among all the vertices, and replace its fitness. The replaced fitness may or may not differ from a, and then this procedure will repeat over and over again. Hence, to simplify the rest of the argument, we can (and will) safely modify the process as follows:

$$X(t+1) \equiv X(t) \text{ as soon as } d(X(t)) = 0 \text{ i.e. all } X_i(t) = a \text{ for some } a \in \mathcal{M}.$$

We will say that the process that the process is *absorbed* at value a.

Remark 3. *The fact that the values of ζ are equally spaced is, surprisingly, crucial. Let* $\operatorname{supp}\zeta = \{0, 1, 5, 6\} =: \mathcal{M}$ *and $N = 8$. Then the set of configurations*

$$[0, 1, x, 5, 6, 5, y, 1], \quad x, y \in \mathcal{M}$$

is stable; the maximum distance from the average of the fitnesses of the neighbours is always at nodes 3 *or* 7*, and it equals* 2 *or* 3*, while the other distances are at most* 1.5 *or* 2 *respectively.*

Theorem 1. *The process $X(t)$ gets absorbed at some value $a \in \mathcal{M}$, regardless of its starting configuration $X(0) \in \mathcal{M}^N$.*

First, observe that since $X(t)$, $t = 0, 1, 2, \ldots$ is a finite-state Markov chain on $\mathcal{M}^N$ with the set of absorbing states

$$\mathrm{O} = (1, 1, \ldots, 1) \cup (2, 2, \ldots, 2) \cup \ldots (M, M, \ldots, M) \subset \mathcal{M}^N$$

it suffices to show that O is accessible (can be reached with a positive probability in some number of steps) from any starting configuration $X(0)$.

First, for $x = (x_1, x_2, \ldots, x_N) \in \mathcal{M}^N$, define

$$\mathrm{Max}(x) = \max_{1\le i\le N} x_i,$$

$$S(x) = \{j \in \{1, 2, \ldots, N\} : \; x_j = \text{Max}(x)\} .$$

that is, the maximum of x, and the indices of x where this maximum is achieved[3]. Let us also define

$$f(x) = \sum_{i=1}^{N} (x_i - x_{i+1})^2$$

with the convention $x_{N+1} \equiv x_1$, which we will use as some sort of Lyapunov function. The following two algebraic statements are not difficult to prove.

Claim 1. *$f(x) = 0$ if and only if $d(x) = 0$.*

Proof. Let $x = (x_1, \ldots, x_N)$. One direction is trivial: if $f(x) = 0$, then $x_i \equiv x_1$ for all $i \in S$ and hence $d_i(x) = 0$ for all $i \in S \iff d(x) = 0$.

On the other hand, suppose that $d_i(x) = 0$ for all i. If not all x_i's are equal, there must be an index j for which $x_j = \max_{i \in S} x_i$, and either $x_{j-1} < x_j$ or $x_{j+1} < x_j$. This, in turn, implies that $2d_j(x) = |(x_j - x_{j-1}) + (x_j - x_{j+1})| = (x_j - x_{j-1}) + (x_j - x_{j+1}) > 0$ yielding a contradiction. □

Claim 2. *Let $x = (x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_N)$ and $x' = (x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_N)$ where $a = \left\lfloor \frac{x_{i-1} + x_{i+1}}{2} \right\rfloor$. Then*

(a) $f(x') \le f(x)$;

(b) if additionally $d_i(x) \ge 1$ then $f(x') \le f(x) - 1$.

Remark 4. *One may expect that there are simpler Lyapunov functions; while we cannot rule this out, let us illustrate two natural candidates that, unfortunately, fail. First, consider $d(x)$; however this function does not work as the next example shows. Let $x = [1, 3, 9, 18, 24, 27, 27, 24, 18, 9, 3, 1]$. Then $d_i(x)$ is the largest at $i = 2$ and $i = 11$; thus $d(x) = d_2(x) = 2$. If we replace a "3" by "4"$= (1 + 9)/2$, then $x' = [1, 4, 9, 18, 24, 27, 27, 24, 18, 9, 3, 1]$ so $d(x') = d_3(x) = 2.5 > d(x)$.*

Another possible candidate, $\tilde{f}(x) = \sum_i d_i(x)^2$ does not work either: let $x = [1, 6, 9, 6, 1]$, then $x' = [1, 6, 6, 6, 1]$ and $\tilde{f}(x') > \tilde{f}(x)$, so it is not a Lyapunov function either.

Proof of Claim 2. . From simple algebra it follows that

$$\frac{f(x') - f(x)}{2} = (a - x_i)(a + x_i - x_{i-1} - x_{i+1})$$

[3] for example, if $x = (1, 4, 2, 4, 4, 2)$ then $\text{Max}(x) = 4$, $S(x) = \{2.4.5\}$.

$$= \left(a - \frac{x_{i-1}+x_{i+1}}{2}\right)^2 - \left(x_i - \frac{x_{i-1}+x_{i+1}}{2}\right)^2 = d_i(x')^2 - d_i(x)^2 =: (*).$$

Note that if $d_i(x) = 0$ or $d_i(x) = 1/2$, then $d_i(x') = d_i(x)$ and thus $(*) = 0$. On the other hand, if $d_i(x) \geq 1$, since $d_i(x') \leq 1/2$, we get $(*) \leq -1/2$. □

To simplify notations, denote

$$\jmath_t = \jmath(X(t)), \qquad \Delta_t = d(X(t)), \qquad f_t = f(X(t)).$$

Now we are going to construct an explicit path through which $X(t)$ can reach O starting from any initial state. Let

$$A_t = \left\{ X_{\jmath_t}(t) \text{ is replaced by } X_{\jmath_t}(t+1) = \left\lfloor \frac{X_{\jmath_t - 1}(t) + X_{\jmath_t + 1}(t)}{2} \right\rfloor, \right.$$
$$\left. \text{and } \jmath_t \in S(X(t)) \text{ if possible} \right\}.$$

Note that the second condition is always possible to satisfy when $\Delta_t = 1/2$. Indeed, if $\Delta_t = 1/2$ for $X(t) = x$, then there must be a j such that $x_j = \mathrm{Max}(x)$ but $x_{j+1} \leq \mathrm{Max}(x) - 1$. As a result, $d_j(x) \geq 1/2$ and hence x_j is one of the points which can be potentially replaced.

Now the statement of Theorem 1 will follow from the following Lemma.

Lemma 1. *For any $X(0)$ there is a $T \geq 0$ such that on the event*

$$A_0 \cap A_1 \cap ... \cap A_T$$

we have $X(T) \in$ O.

This Lemma, in turn, immediately follows from the next statement and the observation that $0 \leq f(x) \leq M^2 N$, as well as the fact that $f(X_T) = 0 \Longleftrightarrow \Delta_T = 0 \Longleftrightarrow X_T \in$ O (see Claim 1).

Claim 3. *If $f_s > 0$ then $f_{s+N-2} \leq f_s - 1$ on $A_s \cap A_{s+1} \cap \cdots \cap A_{s+N-2}$.*

Proof. Note that Δ_t can take only values $\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots\}$. W.l.o.g. we assume that $s = 0$.

First, if $\Delta_t = 0$ for some $0 \leq t \leq N-2$, then $f_t = 0$ by Claim 1 and by Claim 2(a) and the fact that $f_0 \geq 1$, we have $f_{N-2} \leq 0 = f_t \leq f_0 - 1$. From now on suppose that $\min_{0 \leq t \leq N-2} \Delta_t \geq 1/2$.

We will show that it is impossible to have $\Delta_t = \frac{1}{2}$ simultaneously for all $t = 0, 1, 2, \ldots, N-3$ (observe that the case $\Delta_t = 1/2$ contains, quite counter-intuitively, a very rich set of states, see Figure 3). Indeed, the set $S(X(t))$ of indices of the maximum fitnesses must

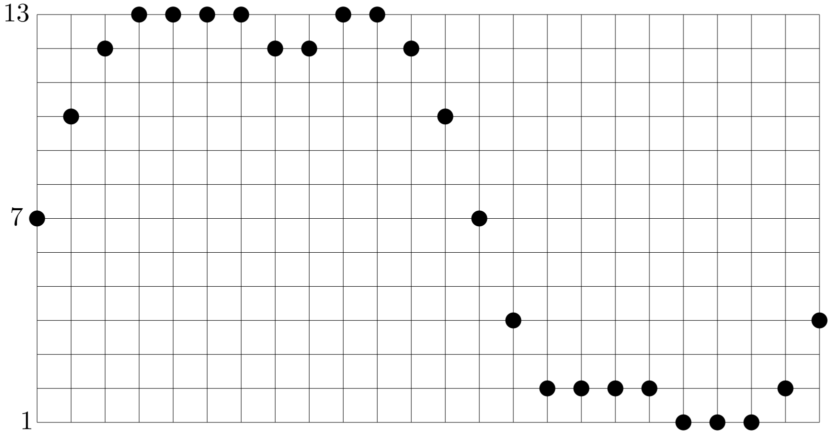


Figure 3: *A configuration with $\Delta_t = 1/2$ (note the periodic boundary conditions), $\mathcal{M} = \{1, 2, \ldots, 13\}$ and $N = 24$. Observe that if $\Delta_t = 1/2$ then there will be a number of "plateaus" each containing at least two maximal fitnesses; moreover, any two such plateaus will be separated by at least two non-maximal fitnesses.*

contain between 2 and $N-2$ elements[4]. However, on A_t we have $S(X(t+1)) \subset S(X(t))$ and $|S(X(t+1))| = |S(X(t))| - 1$ by construction. Since $S(X(0)) \leq N-2$, the value Δ_t cannot stay equal to $1/2$ for $N-2$ consecutive steps, and thus this case is impossible.

As a result, we conclude that $\Delta_t \geq 1$ for some $t \in \{0, 1, \ldots, N-3\}$. Then $f_{t+1} \leq f_t - 1$ by Claim 2(b). As a result, $f_{N-2} \leq f_{t+1} \leq f_t - 1 \leq f_0$ by Claim 2(a). □

Remark 5. *We have actually shown that T in Lemma 1 can be chosen no larger than $M^2N \times (N-2)$, i.e. $\mathbb{P}(X(M^2N(N-2)) \in \mathrm{O} \,|\, X(0) = x) > 0$ for any $x \in \mathcal{M}^N$.*

Remark 6. *It would be interesting to find the distribution of the limiting absorbing configuration, i.e. $\xi := \lim_{t\to\infty} X_i(t)$; clearly it will depend on $X(0)$. This is quite hard problem, and we can present only results based on simulations. Figure 4 shows the histograms of the distribution of ξ for different values of M and N, starting from a random initial condition, i.e. $X_i(0)$ are i.i.d. random variable uniformly distributed on $\mathcal{M}$.*

[4] a single maximum would imply $\Delta_t \geq 1$, the same holds if there are $N-1$ coinciding maxima; finally, $|S(X(t))| = N$ would imply that $\Delta_t = 0$.

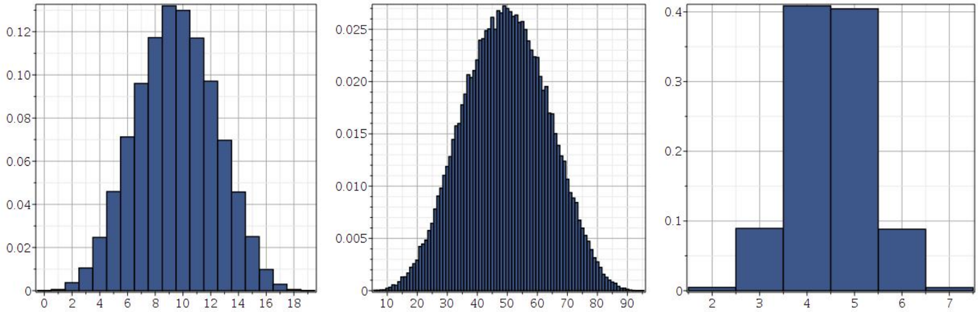


Figure 4: Distribution of ξ based on simulations, for $(N, M) = (20, 20)$, $(20, 100)$, and $(200, 10)$ respectively. Uniform random initial conditions.

3 Continuous case

Throughout this section, we assume that $\zeta \sim U[0,1]$, and $X_i(t) \in [0,1]$ for all $i \in S$ and $t = 0, 1, 2, \dots$. We also assume that $X(0)$ is such that $\jmath(X(0))$ is non-random.

Theorem 2. *There exists a.s. a random variable $\bar{X} \in [0,1]$ such that as $t \to \infty$*

$$(X_1(t), X_2(t), \dots, X_{\jmath(X(t))-1}(t), X_{\jmath(X(t))+1}(t), \dots, X_N(t)) \to (\bar{X}, \bar{X}, \dots, \bar{X}) \in [0,1]^{N-1} \qquad a.s.$$

The proof of this theorem will consists of two parts. Firstly (see Lemma 8), we will show that the properly defined "spread" between the values $X_1(t), \dots, X_N(t)$ converges to zero. This does not, however, imply the the desired result, as hypothetically we can have the situation best described by the "Dance of the Little Swans" from Tchaikovsky's "Swan Lake": while the mutual distances between the X_i's decrease or even some stay 0, their common location changes with time, and thus does not converge to a single point in $[0,1]$. This can happen, for example, if the diameter of the configuration converges to zero too slowly.

The second part of the proof will show that not only the distances between the X_i's decrease, but they all (but the most recently changed one) converge to the same random limit. Please note that the similar strategy was used in [6], however, in our case both steps require much more work.

It turns out that it is much easier to work with the embedded process, for which either the non-conformity of the node at which the value is replaced, is smaller than the initial non-conformity, or at least the location of the "worst" node (i.e. the one where d_i is the largest) has changed, whichever comes first. Formally, let $\nu_0 = 0$ and recursively define for $k = 0, 1, 2, \ldots$

$$\nu_{k+1} = \inf \left\{ t > \nu_k : \ \jmath(X(t)) \neq \jmath(X(\nu_k)) \text{ or } d(X(t)) < d(X(\nu_k)) \right\}.$$

Note that due to the continuity of ζ each $\jmath(X(t))$ is uniquely defined a.s., and that all ν_k are finite a.s..

Examples:

(a) $x = (\ldots 0.5, \underline{0.6}, 0.5, 0.3, \ldots)$. The "worst" node is the second one (with the fitness of 0.6) and $d = d_2(x) = 0.1$; it is replaced, say, by 0.32. Now the configuration becomes

$$x' = (\ldots, 0.5, 0.32, \underline{0.5}, 0.3, \ldots)$$

and the worst node is the third one with $d(x') = d_3(x') = 0.19 > 0.1 = d(x)$;

(b) x is the same as in (a), but x_2 is replaced by 0.58. Now the configuration becomes

$$x = (\ldots, 0.5, \underline{0.58}, 0.5, 0.3, \ldots)$$

and the worst node is still the second one with $d(x') = d_2(x') = 0.08 < 0.1 = d(x)$.

Now let $\tilde{X}(s) = X(\nu_s)$ and $\tilde{\mathcal{F}}_s = \sigma\left(\tilde{X}(1), \ldots, \tilde{X}(s)\right)$ be the filtrations associated with this embedded process. Since throughout time $[\nu_k, \nu_{k+1})$ the value $\jmath$ remains constant at $\jmath_{\nu_k}$ and only $X_{\jmath_{\nu_k}}$ is updated, we have

$$X_i(t) = X_i(\nu_k) \quad \text{for all } i \neq \jmath(X(t))$$

for $t \in [\nu_k, \nu_{k+1})$. Moreover, the process $\tilde{X}$ evolves as a Markov process but with the "update" distribution restricted from the full range, since a uniform distribution conditioned to be in some subinterval is still uniform (this will be used later in Lemma 2). Hence Theorem 2 follows immediately from

Theorem 3. *There exists a.s. a random variable $\bar{X} \in [0, 1]$ such that as $s \to \infty$*

$$(\tilde{X}_1(s), \tilde{X}_2(s), \ldots, \tilde{X}_N(s)) \to (\bar{X}, \bar{X}, \ldots, \bar{X}) \in [0, 1]^N \qquad a.s.$$

(Moreover, this convergence happens exponentially fast: there is an $s_0 = s_0(\omega) < \infty$ and a non-random $\gamma \in (0, 1)$ such that $\left|\tilde{X}_i(s) - \bar{X}\right| \leq \gamma^s$ for all $i \in S$ and $s \geq s_0$.)

Remark 7. *In what follows, we assume that $N \geq 5$. The cases $N = 3$ and $N = 4$ can be studied somewhat easier, and we leave this as an exercise.*

We will use the Lyapunov functions method, with a clever choice of the function. For $x = (x_1, x_2, \dots, x_N)$ define

$$h(x) = 2 \cdot \sum_{i \in S} (x_i - x_{i+1})^2 + \sum_{i \in S} (x_i - x_{i+2})^2 = 2 \sum_{i \in S} \left(3x_i^2 - 2x_i x_{i+1} - x_i x_{i+2}\right).$$

We start by showing that $h(\tilde{X}(s))$ is a non-negative supermartingale (Lemma 2), hence it must converge a.s. Then we show that this limit is actually 0 (Lemma 8). Combined with the fact that $h(\tilde{X}(s))$, as a metric, is equivalent to $\max_{i,j} |\tilde{X}_i(t) - \tilde{X}_j(t)|$, (see Lemma 3) this ensures that eventually all $\tilde{X}_i$ become very close to each other, thus establishing the first necessary ingredient of the proof of the main theorem.

Lemma 2. *$\xi(s) = h\left(\tilde{X}(s)\right)$ is a non-negative supermartingale.*

Proof. The non-negativity of $\xi(s)$ is obvious. To show that it is a supermartingale, assume that $\tilde{X}(s) = (x_1, x_2, x_3, x_4, x_5, \dots)$ and w.l.o.g. that $\jmath(\tilde{X}(s)) = 3$. Suppose that the allowed range (i.e., for which either d decreases or the location of the minimum changes) for the newly sampled point is $[a, b] \subseteq [0, 1]$. Assuming the newly sampled point is uniformly distributed on $[a, b]$ (since a restriction of the uniform distribution to a subinterval is also uniform), we get

$$\begin{aligned} \Delta := \mathbb{E}(\xi(s+1) - \xi(s) | \tilde{\mathcal{F}}_s) = \int_a^b & \left\{2(x_2 - u)^2 + 2(u - x_4)^2 + (x_1 - u)^2 + (u - x_5)^2 \right. \\ & \left. - \left[2(x_2 - x_3)^2 + 2(x_3 - x_4)^2 + (x_1 - x_3)^2 + (x_3 - x_5)^2\right]\right\} \frac{\mathrm{d}u}{b - a} \qquad (3.1) \\ = 2(a^2 + b^2 + ab) + (2x_3 - a - b)&(x_1 + 2x_2 + 2x_4 + x_5) - 6x_3^2. \end{aligned}$$

Now we need to compute the appropriate a and b, and then show that $\Delta \le 0$.

W.l.o.g. we can assume that $x_3 > \frac{x_2 + x_4}{2}$, the case $x_3 < \frac{x_2 + x_4}{2}$ is equivalent to $(1 - x_3) > \frac{(1-x_2)+(1-x_4)}{2}$. Now setting $\tilde{x}_i = 1 - x_i$ for all i yields identical calculations.

Suppose that the fitness at node 3 is replaced by some value $X(\nu_s + 1) =: u$, let the new value of the non-conformity at node 3 be $d_3' = d_3(x_1, x_2, u, x_4, x_5, \dots) = d_3(X(\nu_s + 1))$.

- If x_3 is replaced by $u > x_3$, then this value will be "rejected", in the sense that d has only increased while the $\arg\max_{i \in S} d_i$ is still at the same node (i.e., 3). Indeed, when x_3 increases by some $\delta > 0$, so does d_3, while d_2 and d_4 can potentially increase only by $\delta/2$ and thus cannot overtake d_3.

- When $u \in \left(\frac{x_2+x_4}{2}, x_3\right)$, d_3' is definitely smaller than the original d_3.

  Assume from now on that $u \in \left(0, \frac{x_2+x_4}{2}\right)$. When x_3 is replaced by u, it might happen that while the new d_3 is larger than the original one, the value of d_2 or d_4 overtakes d_3.

- When $u \in \left(0, \frac{x_2+x_4}{2}\right)$ the condition that $d_3' < d_3$ is equivalent to

$$\frac{x_2+x_4}{2} - u < x_3 - \frac{x_2+x_4}{2} \Longleftrightarrow u > x_2 + x_4 - x_3 =: Q_0.$$

- For d_2 to overtake d_3, we need

$$\left|x_2 - \frac{x_1+u}{2}\right| > \frac{x_2+x_4}{2} - u \quad \Longleftrightarrow \quad \begin{cases} u > x_1 - x_2 + x_4 =: Q_1 \\ \text{or} \\ u > \frac{-x_1+3x_2+x_4}{3} =: Q_2 \end{cases}$$

- For d_4 to overtake d_3, we need

$$\left|x_4 - \frac{u+x_5}{2}\right| > \frac{x_2+x_4}{2} - u \quad \Longleftrightarrow \quad \begin{cases} u > x_2 - x_4 + x_5 =: Q_3 \\ \text{or} \\ u > \frac{x_2+3x_4-x_5}{3} =: Q_4 \end{cases}$$

  As a result, the condition for d_3 to be overtaken by some other node, or $d_3' < d_3$ is

$$u > \min_{j=0,1,2,3,4} Q_j.$$

Consequently, we must set

$$\begin{aligned} a &= \max\{0, \min\{Q_0, Q_1, Q_2, Q_3, Q_4\}\} \\ &= \max\left\{0, \min\left\{x_2 + x_4 - x_3, x_1 - x_2 + x_4, \frac{-x_1+3x_2+x_4}{3}, x_2 - x_4 + x_5, \frac{x_2+3x_4-x_5}{3}\right\}\right\}, \\ b &= x_3. \end{aligned}$$

Note that we are guaranteed that $a \leq b$. This is trivial when $a = 0$; on the other hand, when $a > 0$ we have

$$a \leq x_2 + x_4 - x_3 = \frac{x_2+x_4}{2} - \left[x_3 - \frac{x_2+x_4}{2}\right] < \frac{x_2+x_4}{2} < x_3 = b$$

since $x_3 > \frac{x_2+x_4}{2}$.

By substituting $b = x_3$ into the expression for the drift (3.1), we get

$$\Delta = (x_3 - a)(x_1 + 2x_2 - 4x_3 + 2x_4 + x_5 - 2a)$$

and to establish $\Delta \leq 0$ it suffices to show

$$x_1 + 2x_2 - 4x_3 + 2x_4 + x_5 \leq 2a = 2\max\{0, \min\{Q_0, Q_1, Q_2, Q_3, Q_4\}\} \tag{3.2}$$

under the assumption that

$$x_3 - \frac{x_2 + x_4}{2} > \max\left\{\left|x_2 - \frac{x_1 + x_3}{2}\right|, \left|x_4 - \frac{x_3 + x_5}{2}\right|\right\}$$

that is, equivalently,

$$x_3 > \max\{Q_1, Q_2, Q_3, Q_4\}. \tag{3.3}$$

In order to show (3.2) we consider a number of cases. First, assume that $x_2 + x_4 < x_3$. Then $Q_0 < 0$ and $a = 0$. From (3.3) we get that $2x_3 > Q_1 + Q_3 = x_1 + x_5$, thus

$$x_1 + 2x_2 - 4x_3 + 2x_4 + x_5 = (x_1 + x_5 - 2x_3) + 2(x_2 + x_4 - x_3) < 0 = a$$

and (3.2) is fulfilled.

The next case is when $\frac{x_2+x_4}{2} < x_3 < x_2 + x_4$. We need to verify if all of the following holds:

$$x_1 + 2x_2 - 4x_3 + 2x_4 + x_5 - 2Q_j \leq 0 \quad \text{subject to}$$
$$Q_0 \geq 0,\ x_3 \geq Q_1 \geq 0,\ x_3 \geq Q_2 \geq 0,\ x_3 \geq Q_3 \geq 0,\ x_3 \geq Q_4 \geq 0$$

and

$$x_1 + 2x_2 - 4x_3 + 2x_4 + x_5 \leq 0 \quad \text{subject to}$$
$$Q_j \leq 0,\ x_3 \geq Q_1,\ x_3 \geq Q_2,\ x_3 \geq Q_3,\ x_3 \geq Q_4$$

for $j = 0, 1, 2, 3, 4$. This can be done using Linear Programming method. Thus $\Delta \leq 0$. □

The next statement shows that the metrics provided by $h(x)$, $d(x)$, and $\max_{i \in S} |x_i - x_{i-1}|$, where $x \in \mathbb{R}^N$ are, in fact, equivalent.

Lemma 3. *Let $x = (x_1, \ldots, x_N)$ and $\Delta_i(x) := x_i - x_{i-1}$, $i \in S$. Then*

$$\begin{array}{rcccl} d(x) & \le & \max_{i\in S} |\Delta_i| & \le & N d(x), \\ 2\, d(x)^2 & \le & h(x) & \le & 6\, N^3\, d(x)^2. \end{array}$$

Proof. Note that $\Delta_1 + \cdots + \Delta_N = 0$ and

$$h(x) = \sum_{i\in S} \left[2\Delta_i^2 + (\Delta_i + \Delta_{i+1})^2\right],$$
$$d(x) = \frac{1}{2} \max_{i\in S} |\Delta_{i+1} - \Delta_i|.$$

Let j be such that $d_j(x) = d(x)$, then by the triangle inequality

$$|\Delta_{j+1}| + |\Delta_j| \ge |\Delta_{j+1} - \Delta_j| = 2d(x)$$

so at least one of the two terms on the LHS $\ge d(x)$, hence $\max_{i\in S} |\Delta_i| \ge d(x)$.

Now we will show that $\max_{i\in S} |\Delta_i| \le N d(x)$. Indeed, suppose that this is not the case, and w.l.o.g. $\Delta_1 > N d(x)$. For all i we have $|\Delta_{i+1} - \Delta_i| \le 2d(x)$, hence by induction and the triangle inequality we get

$$\begin{aligned} &\Delta_2 > (N-2)\, d(x), \\ &\Delta_3 > (N-4)\, d(x), \\ &\ldots, \\ &\Delta_{N-1} > (N - 2(N-2))\, d(x), \\ &\Delta_N > (N - 2(N-1))\, d(x). \end{aligned}$$

As a result, $\Delta_1 + \Delta_2 + \cdots + \Delta_N > \left[N^2 - 2(1 + 2 + \cdots + (N-1))\right] d(x) = N d(x) \ge 0$, which yields a contradiction, since the LHS is identically equal to 0.

Thus $|\Delta_i| \le N d(x)$, and so $|\Delta_i + \Delta_{i+1}| \le 2N d(x)$ for all $i \in S$. Consequently, $h(x) \le 2N(N d(x))^2 + N(2N d(x))^2 = 6N^3 d(x)^2$. On the other hand, $h(x) \ge \max_{i\in S} 2\Delta_i^2 \ge 2d(x)^2$. □

The following four statements (Lemmas 4 and 5 and Corollaries 1 and 2) show that $\xi(t)$ can actually decrease by a non-trivial factor with a positive (and bounded from below) probability.

Lemma 4. *Suppose that* $X(t) = x = (x_1, x_2, x_3, x_4, x_5, \dots)$, *and* $d_3(x) \geq \max\{d_2(x), d_4(x)\}$. *Let* $\mu = \frac{x_2+x_4}{2}$ *and* $\delta = |x_3 - \mu| = d_3(x)$. *If* x_3 *is replaced by some* $u \in [\mu - \delta/6, \mu + \delta/6]$ *then* $\Delta_h := h(X(t+1)) - h(X(t)) \leq -\frac{5}{6}\delta^2$. *(Note that the Lebesgue measure of* $[\mu - \delta/6, \mu + \delta/6] \bigcap [0,1]$ *is always at least* $\delta/6$*; also after this replacement* d_3 *must decrease.)*

Proof. Note that the change in h equals

$$\Delta_h = -2(x_3 - u)(3u + A), \qquad \text{where } A = 3x_3 - x_1 - 2x_2 - 2x_4 - x_5.$$

W.l.o.g. assume $x_3 > \mu$. Then

$$x_3 - u \geq \mu + \delta - \left(\mu + \frac{\delta}{6}\right) = \frac{5}{6}\,\delta.$$

At the same time, recalling that $d_3(x) \geq \max\{d_2(x), d_4(x)\}$, we obtain that

$$\min_{x_1,\dots,x_5 \geq 0} A \qquad \text{subject to} \qquad x_3 - \mu > \max\left\{\left|x_2 - \frac{x_1 + x_3}{2}\right|, \left|x_4 - \frac{x_3 + x_5}{2}\right|\right\}$$

equals $-3\mu + \delta$. Hence

$$3u + A \geq 3\left(\mu - \frac{\delta}{6}\right) - 3\mu + \delta = \frac{\delta}{2}$$

and thus $\Delta_h \leq -2\,\frac{5\delta}{6} \cdot \frac{\delta}{2}$. □

Lemma 5. *Suppose that* $X(t) = x = (x_1, x_2, x_3, x_4, x_5, \dots)$, *and* $d_3(x) = d(x)$. *Let* $\mu = \frac{x_2+x_4}{2}$ *and* $\delta = |x_3 - \mu| = d_3(x)$. *Given that* $x_3 > \mu$, *if* x_3 *is replaced by some* $u \notin [\mu - 3\delta, x_3]$ *then* $d_3(x') > d_3(x)$ *and* $d_3(x')$ *is still the largest of* $d_i(x')$, *where* $x' = (x_1, x_2, u, x_4, x_5, \dots)$. *The same conclusion holds if* $x_3 < \mu$ *and* x_3 *is replaced by some* $u \notin [x_3, \mu + 3\delta]$.

Before presenting the proof of Lemma 5, we state the obvious

Corollary 1. *Let* $\delta = d(\tilde{X}(s))$. *If* $i = \jmath(\tilde{X}(s))$ *then*

$$\tilde{X}_i(s+1) \in [\tilde{X}_i(s) - 4\delta, \tilde{X}_i(s) + 4\delta]$$

(and if $i \neq \jmath(\tilde{X}(s))$ *then trivially* $X_i(s+1) = X_i(s)$*). Hence we always have*

$$\max_{i \in S} \left|\tilde{X}_i(s+1) - \tilde{X}_i(s)\right| \leq 4\delta.$$

(Note that in Corollary 1 we have 4δ for the following reason: the newly accepted point can deviate from μ by at most 3δ by Lemma 5, while $|\tilde{X}_i(s) - \mu| = \delta$.)

The next implication of Lemma 5 requires a bit of work.

Corollary 2. *Let* $\rho = 1 - \frac{5}{36\, N^3} < 1$*. Then*

$$\mathbb{P}\left(\xi(s+1) \le \rho\xi(s) \,|\, \tilde{\mathcal{F}}_s\right) \ge \frac{1}{48}.$$

Proof of Corollary 2. From Corollary 1 we know that given $x = \tilde{X}(s)$, the allowed range for the newly sampled point to be in $\tilde{X}(s+1)$ is at most 8δ where $\delta = d(x)$. At the same time if the newly sampled point falls into the interval $[\mu - \delta/6, \mu + \delta/6]$ (see Lemma 5), at least half of which lies in $[0,1]$, then $\xi(s+1) - \xi(s) \le -\frac{5}{6}\delta^2$; the probability of this event is no less than $\frac{\delta/6}{8\delta} = \frac{1}{48}$. Since $\xi(s) = h(x)$ and by Lemma 3 we have $d(x)^2 \ge \frac{h(x)}{6N^3}$, the inequality $\xi(s+1) - \xi(s) \le -\frac{5}{6}\delta^2$ implies $\xi(s+1) - \xi(s) \le -\frac{5}{36N^3}\xi(s)$. □

Proof of Lemma 5. By symmetry, it suffices to show just the first part of the statement. First, observe that

$$\begin{aligned}
&d_j(x') = d_j(x) \le d_3(x) \text{ for } j \in S \setminus \{2,3,4\};\\
&d_2(x') = \left|\left(\frac{x_1 + x_3}{2} - x_2\right) + \frac{u - x_3}{2}\right| \le d_2(x) + \left|\frac{u - x_3}{2}\right| \le d_3(x) + \left|\frac{u - x_3}{2}\right|. \qquad (3.4)
\end{aligned}$$

If $u > x_3 > \mu$, then from (3.4)

$$\begin{aligned}
&d_3(x') = u - \frac{x_2 + x_4}{2} > x_3 - \frac{x_2 + x_4}{2} = d_3(x);\\
&d_2(x') \le d_3(x) + \left|\frac{u - x_3}{2}\right| = d_3(x') - (u - x_3) + \left|\frac{u - x_3}{2}\right| = d_3(x') - \left|\frac{u - x_3}{2}\right| < d_3(x');\\
&d_4(x') < d_3(x') \qquad \text{(by the same argument as } d_2\text{)}
\end{aligned}$$

so indeed $d_3(x) < d_3(x') = \max_{i \in S} d_i(x')$.

On the other hand, if $u < \mu - 3\delta < x_3 = \mu + \delta$, then d_j for $j \in S \setminus \{2,3,4\}$ still remain unchanged, but

$$\begin{aligned}
d_3(x') &= \mu - u > 3\delta > d_3(x);\\
d_2(x') &\le d_3(x) + \left|\frac{u - x_3}{2}\right| = \delta + \frac{x_3 - u}{2} = \delta + \frac{x_3 - \mu}{2} + \frac{\mu - u}{2} = \frac{3\delta}{2} + \frac{\mu - u}{2}\\
&< \frac{\mu - u}{2} + \frac{\mu - u}{2} = d_3(x');
\end{aligned}$$

$$d_4(x') < d_3(x') \quad \text{(by the same argument as } d_2\text{)}$$

hence $d_3(x) < d_3(x') = \max_{i \in S} d_i(x')$ in this case as well. □

At the same time, it turns out that $\xi(t)$ cannot increase too much in one step, as follows from

Lemma 6. *There is a non-random $r > 0$ such that for all s we have $\xi(s+1) \le r\xi(s)$.*

Proof. By Corollary 1 it follows that the worst outlier (w.l.o.g. x_3) can be replaced only by a point at most at the distance 4δ from x_3 at time ν_{s+1}. Let the new value of the fitness at node 3 be $x_3 + v$, $|v| \le 4\delta$. The change in the Lyapunov function is given by

$$\begin{aligned}
\xi(s+1) - \xi(s) &= \left[2((x_3+v)-x_2)^2 + 2((x_3+v)-x_4)^2 + ((x_3+v)-x_1)^2 + ((x_3+v)-x_5)^2\right] \\
&\quad - \left[2(x_3-x_2)^2 + 2(x_3-x_4)^2 + (x_3-x_1)^2 + (x_3-x_5)^2\right] \\
&= (12x_3 - 2x_2 - 2x_4 - 4x_1 - 4x_5)\, v + 6\, v^2 \qquad (3.5)
\end{aligned}$$

Since

$$\begin{aligned}
|12x_3 - 2x_2 - 2x_4 - 4x_1 - 4x_5| &= \left|8\left(x_2 - \frac{x_1+x_3}{2}\right) + 8\left(x_4 - \frac{x_5+x_3}{2}\right) + 20\left(x_3 - \frac{x_2+x_4}{2}\right)\right| \\
&\le 8\delta + 8\delta + 20\delta = 36\delta
\end{aligned}$$

from (3.5) and the fact that $\delta = d(\tilde{X}(s)) \le \sqrt{\frac{\xi(s)}{2}}$ by Lemma 3

$$|\xi(s+1) - \xi(s)| \le 36\delta \times 4\delta + 6\,(4\delta)^2 = 240\delta^2 \le 120\xi(s),$$

so we can take $r = 121$. □

Finally, we want to show that, roughly speaking, one does not have to wait for too long before $\xi(t)$ increases or decreases by a *substantial* amount.

Lemma 7. *Fix some $k > 1$ and $s_0 > 0$. Let $\tau_1 = \inf\{s > 0 :\ \xi(s_0+s) \le \xi(s_0)/k\}$ and $\tau_2 = \inf\{s > 0 :\ \xi(s_0+s) \ge k\xi(s_0)\}$. Then $\tau = \min(\tau_1, \tau_2)$, given $\tilde{\mathcal{F}}_{s_0}$, is stochastically smaller than some random variable with a finite mean, the distribution of which does not depend on anything except N and k.*

Proof. Fix a positive integer L. For each $t \geq s_0$ define

$$B_t = \left\{ \xi(t+L) \leq \frac{\xi(t)}{k^2} \right\}.$$

It suffices to show that $\mathbb{P}(B_t|\tilde{\mathcal{F}}_t) \geq p$ for some $p > 0$ uniformly in t, since for $j = 0, 1, 2, \dots$

$$\begin{aligned} B_{s_0+jL} &\subseteq \{\xi(s_0+jL) < k\xi(s_0) \text{ and } \xi(s_0+(j+1)L) < \xi(s_0)/k\} \cup \{\xi(s_0+jL) \geq k\xi(s_0)\} \\ &\subseteq \{\tau_1 \leq (j+1)L\} \cup \{\tau_2 \leq jL\} \subseteq \{\tau \leq (j+1)L\}. \end{aligned}$$

which, in turn, would imply that τ is stochastically smaller than L multiplied by a geometric random variable with parameter $p = p(N, k)$.

To show that $\mathbb{P}(B_t \mid \tilde{\mathcal{F}}_t) \geq p$, note that by Corollary 2,

$$\mathbb{P}(B_m^* \,|\, \tilde{\mathcal{F}}_{m-1}) \geq \frac{1}{48}, \quad \text{where } B_m^* = \{\xi(m) < \rho\xi(m-1)\}, \quad \rho = 1 - \frac{5}{36N^3}.$$

Let L be so large that $\rho^L < 1/k^2$. Then, on one hand,

$$\bigcap_{m=1}^{L} B_{t+m}^* \subseteq B_t \text{ whence } \mathbb{P}\left(B_t \,|\, \tilde{\mathcal{F}}_t\right) \geq \mathbb{P}\left(\bigcap_{m=1}^{L} B_{t+m}^* \,|\, \tilde{\mathcal{F}}_t\right),$$

while on the other hand

$$\mathbb{P}\left(\bigcap_{m=1}^{L} B_{t+m}^* \,|\, \tilde{\mathcal{F}}_t\right) \geq \frac{1}{48^L} =: p$$

which depends on N and k only. $\square$

The proof of the next statement, which completes the first part of the proof of the main theorem, requires a bit more work than that of Lemma 2.4 in [6]. In fact, we will prove a stronger statement (Corollary 3) later, however, it is still useful to see a fairly quick proof of the following

Lemma 8. *$\xi(s) \to 0$ a.s. as $s \to \infty$ (and as a result $\Delta_i(\tilde{X}(s)) \to 0$ a.s. and $d(\tilde{X}(s)) \to 0$ a.s. as $s \to \infty$).*

Proof. From Lemma 2 it follows that $\xi(s)$ converges a.s. to a non-negative limit, say ξ_∞. Let us show that $\xi_\infty = 0$. From Corollary 2 we have

$$\mathbb{P}\left(\xi(s+1) \leq \rho\xi(s) \,|\, \mathcal{F}_s\right) \geq \frac{1}{48}. \tag{3.6}$$

Fix an $\varepsilon > 0$ and a $T \in \mathbb{N}$. Let $\sigma_{\varepsilon,T} = \inf\{s \geq T : \ \xi(s) \leq \varepsilon\}$. Then (3.6) implies

$$\mathbb{P}(A_{s+1} \,|\, \mathcal{F}_s) \geq \frac{1_{s<\sigma_{\varepsilon,T}}}{48}, \quad \text{where } A_{s+1} = \{\xi(s+1) \leq \xi(s) - (1-\rho)\varepsilon\}$$

(Compare this with the inequality (2.18) in [6]). From the non-negativity of $\xi(s)$, we know that only finitely many of A_s can occur. By the Levy's extension to the Borel-Cantelli lemma, we get that $\sum_{s=T}^{\infty} \mathbb{P}(A_{s+1} \,|\, \mathcal{F}_s) < \infty$ a.s., and hence $\sum_{s=T}^{\infty} 1_{s<\sigma_{\varepsilon,T}} < \infty$. This, in turn, implies that $\sigma_{\varepsilon,T} < \infty$ a.s. Consequently, since T is arbitrary,

$$\liminf_{s\to\infty} \xi(s) \leq \varepsilon \quad \text{a.s.}$$

Since $\varepsilon > 0$ is also arbitrary and $\xi(s)$ converges, $\lim_{s\to\infty} \xi(s) = \liminf_{s\to\infty} \xi(s) = 0$ a.s. □

The next general statement may be known, but since we could not find it in the literature, we present its fairly short proof. We need it in order to show that $\xi(t)$ converges to zero quickly.

Proposition 1. *Suppose that $\xi(s)$ is a positive bounded supermartingale with respect to a filtration $\tilde{\mathcal{F}}_s$. Suppose there is a constant $r > 1$ such that $\xi(s+1) \leq r\xi(s)$ a.s. and that for all k large enough the stopping times*

$$\tau_s = \inf\{t > s : \ \xi(t) > k\xi(s) \ \textit{or} \ \xi(t) < k^{-1}\,\xi(s)\}$$

are stochastically bounded above by some finite–mean random variable $\bar{\tau} > 0$, which depends on k only (and, in particular, independent of $\tilde{\mathcal{F}}_s$). Let $\mu = \mathbb{E}\bar{\tau} < \infty$. Then

$$\limsup_{s\to\infty} \frac{\ln \xi(s)}{s} \leq -\frac{1}{4\mu} < 0 \qquad \textit{a.s.}$$

Proof. First, observe that by the Optional Stopping Theorem

$$\mathbb{E}(\xi(\tau_s) \,|\, \tilde{\mathcal{F}}_s) \leq \xi(s) \tag{3.7}$$

(where $\tau_s < \infty$ a.s. by the stochastic dominance condition) while, on the other hand,

$$\begin{aligned} \mathbb{E}(\xi(\tau_s) \,|\, \tilde{\mathcal{F}}_s) &= \mathbb{E}(\xi(\tau_s), \xi(\tau_s) > k\xi(s) \,|\, \tilde{\mathcal{F}}_s) + \mathbb{E}(\xi(\tau_s), \xi(\tau_s) < k^{-1}\,\xi(s) \,|\, \tilde{\mathcal{F}}_s) \\ &\geq \mathbb{E}(\xi(\tau_s), \xi(\tau_s) > k\xi(s) \,|\, \tilde{\mathcal{F}}_s) \geq k\xi(s) \cdot \mathbb{P}(\xi(\tau_s) > k\xi(s) \,|\, \tilde{\mathcal{F}}_s). \end{aligned} \tag{3.8}$$

From (3.7) and (3.8) we conclude

$$p := \mathbb{P}(\xi(\tau_s) > k\xi(s) \,|\, \tilde{\mathcal{F}}_s) < \frac{1}{k}. \tag{3.9}$$

Now let us define a sequence of stopping times as follows: $\eta_0 = 0$ and for $n = 1, 2, \ldots,$

$$\eta_n = \inf \left\{ s > \eta_{n-1} : \ \xi(s) > k\xi(\eta_{n-1}) \text{ or } \xi(s) < k^{-1}\, \xi(\eta_{n-1}) \right\}$$

and let

$$N_s = \max\{n : \ \eta_n \le s\}.$$

From the definition of the stopping times η, it follows

$$\xi(s) \le k\xi(\eta_{N_s}), \qquad \xi(\eta_{n+1}) \le rk\xi(\eta_n). \tag{3.10}$$

Consider now the sequence of random variables $\xi(\eta_n)$. From (3.9) and (3.10) we obtain that $\log_k \frac{\xi(\eta_n)}{\xi(\eta_{n-1})}$ is stochastically bounded above by a random variable $X_n \in \{-1, 1 + \log_k r\}$ such that

$$1 - \mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1 + \log_k r) = \frac{1}{k}$$

yielding

$$\mathbb{E}X_n = \frac{2 + \frac{\ln r}{\ln k}}{k} - 1 =: g(r,k);$$

we can also assume that X_n are i.i.d. One can choose $k > 1$ so large[5] that $g(r,k) < -\frac{1}{2}$. Then, by the Strong Law applied to $\sum_{i=1}^{n} X_i$, we get

$$\limsup_{n\to\infty} \frac{\log_k \xi(\eta_n)}{n} \le \limsup_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} < -\frac{1}{2} \qquad \text{a.s.}$$

From the condition of the proposition we know that the differences $\eta_n - \eta_{n-1}$, $n = 1, 2, \ldots,$ are stochastically bounded by independent random variables with the distribution of $\bar{\tau}$ with $\mathbb{E}\bar{\tau} =: \mu < \infty$. Then by the Strong Law for renewal processes (see e.g. [5], Theorem I.7.3) applied to the sum of independent copies of $\bar{\tau}$, we get

$$\liminf_{s\to\infty} \frac{N_s}{s} \ge \frac{1}{\mu} \quad \text{a.s.} \quad \Longrightarrow \quad s \le 2\mu N_s \text{ for all large enough } s. \tag{3.11}$$

Combining (3.10) and (3.11), we get

$$\limsup_{s\to\infty} \frac{\log_k \xi(s)}{s} \le \limsup_{s\to\infty} \frac{\log_k (k\xi(\eta_{N_s}))}{s} = \limsup_{s\to\infty} \frac{\log_k \xi(\eta_{N_s})}{s}$$

[5] if $r > 4.1$, then $k = \ln(r)$ will be sufficient.

$$\leq \limsup_{s\to\infty} \frac{\log_k \xi(\eta_{N_s})}{2\mu N_s} = \frac{1}{2\mu} \limsup_{n\to\infty} \frac{\log_k \xi(\eta_n)}{n} \leq -\frac{1}{4\mu} \qquad \text{a.s.}$$

since $N_s \to \infty$ when $s \to \infty$ a.s. □

The next statement strengthens Lemma 8.

Corollary 3. *$\xi(s) \to 0$ exponentially fast as $s \to \infty$.*

Proof. The statement follows immediately from Proposition 1: the bound for r we have by Lemma 6; the other condition follows from Lemma 7. □

Now we are ready to finish the proof of the main statement.

Proof of Theorem 3. According to Corollary 3 there exist $a, b > 0$ which are a.s. finite and such that $\xi(t) \leq ae^{-bt}$. If we take s_0 such that $ae^{-bs} \leq \epsilon$ for all $s \geq s_0$ then if $s_0 \leq s < t$,

$$\begin{aligned} |\tilde{X}_i(t) - \tilde{X}_i(s)| &\leq \sum_{k=s+1}^{t} 4\, d(\tilde{X}(k)) \leq \sum_{k=s+1}^{t} \sqrt{8\xi(k)} \\ &\leq \sqrt{8\epsilon} \sum_{k=s+1}^{t} e^{-bk/2} \leq \frac{\sqrt{8\epsilon}}{1 - e^{-b/2}}, \end{aligned} \tag{3.12}$$

where we used Corollary 1 in the first inequality and Lemma 3 in the second inequality. We can thus conclude that $\{\bar{X}_i(t)\}_t$ is a Cauchy sequence in the a.s. sense; therefore the limit $\bar{X}_i(\infty) = \lim_{t\to\infty} \tilde{X}_i(t)$ exists a.s. Moreover, by letting $t \to \infty$ in (3.12), we get that $|\tilde{X}_i(s) - \tilde{X}_i(\infty)| \leq Ce^{-bs/2}$ for some $C > 0$.

Furthermore, assuming w.l.o.g. that $i < j$,

$$|\bar{X}_i(\infty) - \bar{X}_j(\infty)| = \lim_{t\to\infty} |\tilde{X}_i(t) - \tilde{X}_j(t)| \leq \lim_{t\to\infty} \sum_{k=i+1}^{j} \left|\Delta_k(\tilde{X}(t))\right| = 0$$

by Lemma 8, which completes the proof. □

4 Discussion and open problems

One may be interested in the speed of convergence, established in Theorem 3. In Lemma 6 we can take $r = 121$ and from the proof of Proposition 1, $k = \ln r = \ln(121) = 2\ln(11)$ will be sufficient. Then, for Lemma 7, find L such that

$$\left(1 - \frac{5}{36N^3}\right)^L < \frac{1}{23} < \frac{1}{k^2}$$

We can take, e.g.,

$$L \approx 7.2N^3 \cdot \ln(23) \approx 22.6N^3$$

This, in turn, will provide a bound on $\mu = \mathbb{E}\bar{\tau} \le \frac{L}{p} = L \cdot 48^L$ for Proposition 1, and hence the speed of the convergence for large s:

$$2\,[d(\tilde{X}(s))]^2 \le h(\tilde{X}(s)) = \xi(s) \le k^{-\frac{s}{4\mu}} \le \exp\left\{-\frac{s}{8\,L\,48^L\,\ln(11)}\right\} \approx \exp\left\{-\frac{s}{433 \cdot 10^{38N^3}}\right\}$$

This bound is, however, far from the optimal one. The simulations seem to indicate that, depending on N,

$$\xi(s) \sim e^{-\rho_N s},$$

where e.g. $\rho_5 \in (0.47, 0.77)$, $\rho_{10} \in (0.14, 0.23)$, $\rho_{20} \in (0.02, 0.03)$, $\rho_{40} \in (0.003, 0.006)$, suggesting that (a) ρ_N can be, in fact, random, and (b) the average value of ρ_N decays roughly like $5/N^2$. We leave the study of the properties of ρ_N for further research.

We believe that the convergence, described by Theorems 2 and 3 holds for a much more general class of replacement distributions ζ, not just uniform; for example, for the continuous distributions with the property that their density is uniformly bounded away from zero. Unfortunately, our proof is based on the construction of the Lyapunov function which cannot be easily transferred to other cases (obviously, it will work for any $\zeta \sim U[a, b]$, where $a < b$).

One can also attempt to generalize the theorems for more general graphs as described in Remark 1; this should be done, however, with care, as it will not work for all the distributions (see Remark 2).

5 Bibliography

[1] Bak, P., and Sneppen, K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters* **71**, 4083–4086.

[2] Barbay, J., and Kenyon C. (2001). On the discrete Bak-Sneppen model of self-organized criticality. *Proceedings Of The Twelth Annual ACM-SIAM Symposium On Discrete Algorithms (SODA)*, Washington DC.

[3] Ben-Ari, I., Silva, R. (2018). On a local version of the Bak-Sneppen model. *J. Stat. Phys.* **173**, 362–380.

[4] Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy* **102**, 841–877,

[5] Durrett, R. (2010). Probability: Theory and Examples. Cambridge Series in Statistical and Probabilistic Mathematics.

[6] Grinfeld, M., Volkov, S., Wade, A. R. (2015). Convergence in a multidimensional randomized Keynesian beauty contest. *Adv. in Appl. Probab.* **47**, 57–82.

[7] Teck-Hua Ho, Colin Camerer and Keith Weigelt. (1998). Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests" *The American Economic Review* **88**, 947–969.

[8] Kennerberg, P., and Volkov, S. (2018). Jante's law process. *Adv. in Appl. Probab.* **50**, 414–439.

[9] Kennerberg, P. and Volkov, S. (2020). Convergence in the p-contest. *Journal of Statistical Physics* **178**, 1096–1125.

[10] Morrison, C., Naumov, P. (2020). Group Conformity in Social Networks. *Journal of Logic, Language and Information* **29**, 3–19.

[11] Meester, R., and Znamenski, D. (2003). Limit behavior of the Bak-Sneppen evolution model. *Ann. Probab.* **31**, 1986–2002.

[12] Sandemose, A. (1936). A fugitive crosses his tracks. translated by Eugene Gay-Tifft. New York: A. A. Knopf.

[13] Tang, J., Wu, S., and Sun. J. (2013). Confluence: Conformity Influence in Large Social Networks. *KDD'13*, August 11–14, 2013, Chicago, Illinois, USA.

[14] Veerman, J., Prieto, F. (2014). On rank driven dynamical systems. *J. Stat. Phys.* **156** 455–472.

[15] Volkov, S. (2020). Rigorous upper bound for the discrete Bak-Sneppen model. https://arxiv.org/abs/2003.00222, preprint.

D

Chapter 5

Paper D

Some stability results for stochastic integrators

Philip Kennerberg* and Magnus Wiktorsson[1]

9th February 2021

[1]Centre for Mathematical Sciences, Lund University, Box 118 SE-22100, Lund, Sweden

Abstract

We consider limits for sequences of the type $\int Y_- df_n(X^n)$ and $[f_n(X^n) - f(X)]$ for semimartingale integrands, where $\{X^n\}_n$ either are Dirichlet processes or more generally processes admitting to quadratic variations. We here assume that the functions $\{f_n\}_n$ are either C^1 or absolutely continuous. We also provide important examples of how to apply this theory for sequential jump removal.

1 Introduction

We will study stability of integrators and quadratic variations under sequences of C^1 as well as absolutely continuous transformations. What what we mean by this is that assuming $\{f_n\}_n$ and f are C^1 functions, $\{X^n\}_n$ and X are cadlag processes such that $X^n \to X$ (in a sense specified below) then $\int Y_- df_n(X^n) \to \int Y_- df(X)$ and $[f_n(X^n) - f(X)] \to 0$ in a corresponding sense, for a semimartingale integrand Y. The types of processes we shall consider are both the more general, processes with quadratic variation, as well as the more specific class of Dirichlet processes. For processes with quadratic variations we will study convergence in the ucp setting, while for the Dirichlet case we will study convergence in L^p uniformly over time. We will assume that our processes live on some compact time interval say $[0, t]$ for some $t \in \mathbb{R}^+$. The term refining sequence will refer to a sequence of partitions of $[0, t]$, $\{D_k\}_k$ such that $D_k \subseteq D_{k+1}$ and $\lim_{k\to\infty} \max_{t_i \in D_k} |t_{i+1} - t_i| = 0$. We say that a cadlag X process admits to a quadratic variation if there exists an increasing process $[X]$ such that

$$[X]_s = [X]^c_s + \sum_{u \leq s} (\Delta X_u)^2, \tag{1.1}$$

for every $0 < s \leq t$, there exists at least one refining sequence $\{D_k\}_k$ such that if we let

$$(S_n(X))_s := \sum_{t_i \in D_n, t_i \leq s} \left(X_{t_{i+1}} - X_{t_i}\right)^2,$$

for every $0 < s \leq t$ (where we use the convention that $t_{i+1} = s$ if $t_i = s$) then

$$(S_n(X))_s \xrightarrow{\mathbb{P}} [X]_s \ \textit{as } n \to \infty. \tag{1.2}$$

We say that two cadlag processes admit to a covariation $[X,Y]$ along $\{D_n\}_n$ if $[X,Y]$ is a finite variation process such that

$$[X,Y]_s = [X,Y]^c_s + \sum_{u\leq s} \Delta X_u \Delta Y_u,$$

for every $0 < s \leq t$ and if we let

$$S_n(X,Y)_s := \sum_{t_i\in D_n, t_i\leq s} \left(X_{t_{i+1}} - X_{t_i}\right)\left(Y_{t_{i+1}} - Y_{t_i}\right),$$

for every $0 < s \leq t$ then $S_n(X,Y)_s \xrightarrow{\mathbb{P}} [X,Y]_s$ as $n \to \infty$.
We recall that a Dirichlet process is a sum of a semimartingale and an adapted continuous process with zero quadratic variation (there are other characterizations, but this is the one that suits our purposes), they where originally described in [2]. By this definition we see that Dirichlet processes are a subclass of the processes admitting to quadratic variation.

2 Preliminaries

We assume that all our processes are defined on a common filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ and that all the defined processes are adapted to the same filtration $\{\mathcal{F}_t\}_{t\geq 0}$ fulfilling the usual hypothesis.

Given a cadlag process X_t with $t \geq 0$ and a stopping time T we defined X stopped at T as $X^T_t = X_{t\wedge T}$, we also define the supremum process of X as $X^*_t = \sup_{s\leq t} |X_s|$.

Definition 2.1. *A property of a stochastic process is said to hold locally (pre-locally) if there exist a sequence of stopping times T_k increasing to infinity such that the property holds for the stopped process X^{T_k} (X^{T_k-}) for each k.*

A basic property of processes that admits to quadratic variations is that they are closed under absolutely continuous maps. Such processes also have a vector-space structure *per refining sequence*, i.e. all processes that have quadratic variations along the same refining sequence form a vector-space.

Lemma 2.2. *Suppose that g_1, g_2 are absolutely continuous functions, that X and Y both have quadratic variations along the refining sequence $\{D_k\}_k$ then so does $g_1(X)+g_2(Y)$.*

Proof. See Appendix ☐

As one would expect, if two processes have quadratic variations along the same refining sequence this implies that they have a covariation along this sequence as well.

Lemma 2.3. *Suppose that X and Y both have quadratic variations along the refining sequence $\{D_k\}_k$ then the covariation process $[X, Y]$ also exists along this sequence.*

Proof. See Appendix ☐

The following two lemma's are just some very elementary bounds that we shall make frequent use of.

Lemma 2.4. *Suppose $X^1, ..., X^n$ are processes with quadratic variations along the same refining sequence then*

$$\left[\sum_{k=1}^{n} X^k\right]_t \leq \left(\sum_{k=1}^{n} [X^k]_t^{\frac{1}{2}}\right)^2$$

Proof. First of all we note that $X^1, ..., X^n$ having quadratic variation along the same refining sequence implies that $\sum_{k=1}^{m} X^k$ has a quadratic variation along this sequence as well. We prove the stated inequality by induction, the case $n = 1$ is trivial. Assume the statement is true for $n = m$ then for $n = m + 1$,

$$\left[\sum_{k=1}^{m+1} X^k\right]_t = \left[\sum_{k=1}^{m} X^k\right]_t + 2\left[\sum_{k=1}^{m} X^k, X^{m+1}\right]_t + [X^{m+1}]_t \leq \left[\sum_{k=1}^{m} X^k\right]_t +$$
$$2\left[\sum_{k=1}^{m} X^k\right]_t^{\frac{1}{2}} [X^{m+1}]_t^{\frac{1}{2}} + [X^{m+1}]_t = \left(\left[\sum_{k=1}^{m} X^k\right]_t^{\frac{1}{2}} + [X^{m+1}]_t^{\frac{1}{2}}\right)^2 \leq \left(\sum_{k=1}^{m} [X_k]_t^{\frac{1}{2}} + [X^{m+1}]_t^{\frac{1}{2}}\right)^2$$
$$= \left(\sum_{k=1}^{m+1} [X_k]_t^{\frac{1}{2}}\right)^2,$$

where we used the Kunita-Watanabe inequality in the second step. ☐

Lemma 2.5. *Suppose $X^1, ..., X^n$ are processes with quadratic variations along the same refining sequence then*

$$\left[\sum_{k=1}^{n} X^k\right]_t \leq 2^{n-1} \sum_{k=1}^{n} [X^k]_t$$

Proof. We again prove the stated inequality by induction, the case $n = 1$ is trivial and the case $n = 2$ follows from Lemma 2.4 combined with the inequality $(a+b)^2 \leq 2a^2+2b^2$. Assume the statement is true for $n = m$ then for $n = m + 1$,

$$\left[\sum_{k=1}^{m+1} X^k\right]_t \leq 2\left[\sum_{k=1}^{m} X^k\right]_t + 2\left[X^{m+1}\right]_t \leq 2^{(m-1)+1}\sum_{k=1}^{m}[X^k]_t + 2\left[X^{m+1}\right]_t \leq 2^m\sum_{k=1}^{m+1}[X^k]_t,$$

where we used the case $n = 2$ in the first step and the induction hypothesis in the second step. □

The following definition is taken from [1]

Definition 2.6. *Given two cadlag processes X, Y we say that the integral $U_\cdot = \int_0^\cdot Y_{s-}dX_s$ is well defined if there exists a refining sequence $\{D_k\}_k$ and U is a cadlag process such that for every $s \leq t$, $\Delta U_s = Y_{s-}\Delta X_s$ such that $I^k(X,Y) \to U$, as $k \to \infty$ in probability under the $J1$ topology, where $I^k(X,Y)_t = \sum_{t_i \in D_k, t_i \leq t} Y_{t_i}(X_{t_{i+1}} - X_{t_i})$.*

The next lemma shows that all processes that admit to quadratic variations are admissible integrators in the sense of Defnition 2.6 for semimartingale integrands.

Lemma 2.7. *Let Y be a semimartingale and X a cadlag process with quadratic variation along $\{D_k\}_k$ then the integral $\int_0^\cdot Y_{s-}dX_s$ exists in the sense of Definition 2.6. Moreover this integral satisfies the integration by parts formula (along $\{D_k\}_k$).*

Proof. Since Y is a semimartingale it has quadratic variation along any refining sequence and $\{D_k\}_k$ in particular which implies that $X + Y$ also has quadratic variation along $\{D_k\}_k$ by Lemma 2.2. For any k and $s \leq t$

$$I^k(X,Y) = Y_\cdot X_\cdot - Y_0X_0 - I^k(Y,X) - \sum_{t_i \in D_k, t_i \leq \cdot}(Y_{t_{i+1}} - Y_{t_i})(X_{t_{i+1}} - X_{t_i}). \tag{2.3}$$

Since Y is a semimartingale $I^k(Y,X)$ converges in ucp (and therefore in probability under the $J1$ topology). We now expand the last term,

$$\sum_{t_i \in D_k, t_i \leq \cdot}(Y_{t_{i+1}} - Y_{t_i})(X_{t_{i+1}} - X_{t_i}) =$$

$$\frac{1}{2}\left(\sum_{t_i \in D_k, t_i \leq \cdot}(Y_{t_{i+1}} + X_{t_{i+1}} - Y_{t_i} - X_{t_i})^2 - \sum_{t_i \in D_k, t_i \leq \cdot}(Y_{t_{i+1}} - Y_{t_i})^2 - \sum_{t_i \in D_k, t_i \leq \cdot}(X_{t_{i+1}} - X_{t_i})^2\right), \tag{2.4}$$

all three terms on the right side of (2.4) converge in the $J1$ topology due to Lemma 1.3 in [1] and therefore the final term in (2.3) converges in the $J1$ topology. By Lemma 2.3 we have that $[X,Y]$ exists, is the limit of the left-hand side of (2.4) in probability at every $s \le t$ and we have that for any $s \le t$ that

$$(\Delta[X,Y])_s = \frac{1}{2}\left(\Delta[X+Y] - \Delta[X] - \Delta[Y]\right)_s = \Delta X_s \Delta Y_s.$$

Taking limits on the right-hand side of (2.3) and evaluating the jump at time u gives us

$$\Delta\left(\int_0^{\cdot} Y_{s-} dX_s\right)_u = \Delta(XY)_u - \Delta\left(\int_0^{\cdot} X_{s-} dY_s\right)_u - \Delta X_u \Delta Y_u =$$
$$\Delta(XY)_u - (X_{u-} + \Delta X_u)\,\Delta Y_u = Y_{u-}\Delta X_u$$

and hence the requirements of Definition 2.6 are met. Evaluating (2.3) at time $s \le t$ and taking limits in probability gives us the integration by parts formula. □

Combining Lemma 2.2 with Lemma 2.7 gives us the following.

Corollary 2.8. *Suppose X has a quadratic variation along the refining sequence $\{D_k\}_k$, f is an absolutely continuous function and Y a semimartingale then the integral $\int Y_- df(X)$ is well defined in the sense of Definition 2.6*

For general a semimartingale X we define [see e.g. 5, p. 245.] the $\underline{\underline{H}}^p$-norm, we here however work on a finite interval $[0,t]$.

Definition 2.9. *For $1 \le p \le \infty$ define*

$$j_p(M,A) = \|\ |X_0| + [M]_t^{1/2} + \int_0^t |dA_s|\ \|_{L^p}$$

and then

$$\|X\|_{\underline{\underline{H}}^p} = \inf_{X = X_0 + M + A} j_p(M,A)$$

for all possible decompositions $X = X_0 + M + A$ where M is a local martingale with $M_0 = 0$ and A is a process of finite variation with $A_0 = 0$.

Recall that $\underline{\underline{H}}^p$ is the Banach space of all semi martingales with finite $\underline{\underline{H}}^p$-norm and note that all semimartingales with $X_0 \in L^p$ are pre-locally in $\underline{\underline{H}}^p$ [see e.g. 5, p. 247.].

Next is one of the main tools for dealing with the quadratic variations of Dirichlet processes, it is just a very minor modification of a result in [1].

Theorem 2.10. *Let $X = Z + C$ where Z is a semimartingale and C has zero quadratic variation and f be a C^1-function. For any $a \in \mathbb{R}^+$ we have $f(X_s) = Y_s^a + \Gamma_s^a$ where Y^a is a semimartingale, Γ^a is continuous and $[\Gamma^a]_t = 0$ for all $t > 0$. The expression for Y^a is given by*

$$\begin{aligned} Y_t^a = {} & f(X_0) + \sum_{s \le t} \left(f(X_s) - f(X_{s-}) - \Delta X_s f(X_{s-}) \right) I_{|\Delta X_s| > a} + \int_0^t f'(X_{s-}) dZ_s \\ & + \int_0^t \int_{|x| \le a} \left(f(X_{s-} + x) - f(X_{s-}) - x f'(X_{s-}) \right) (\mu - \nu)(ds, dx) \\ & + \sum_{s \le t} \int_{|x| \le a} \left(f(X_{s-} + x) - f(X_{s-}) - x f'(X_{s-}) \right) \nu(\{s\}, dx). \end{aligned} \tag{2.5}$$

Proof. According to Theorem 2.1. in [1] (2.5) is true with $a = 1$ but the proof works just as well with any other $a \in \mathbb{R}^+$. $\square$

3 Main results and lemmas

We will work with the following assumptions with regards to convergence of functions

Assumption 3.1. *$\{f_n\}_n$ and f are assumed to be absolutely continuous, $f_n(x_0) \to f(x_0)$ for some fixed point x_0, $f_n' \to f'$ a.e. and there exists a locally integrable function $h(x)$ such that $|f_n'(x)| \le |h(x)|, a.e. x \in \mathbb{R}$ for all n.*

A stronger condition than Assumption 3.1 is that we replace the condition of a dominating function h with that of uniform convergence of f_n' to f' on compacts. To see that this condition indeed is stronger, given some compact set I, fix $\epsilon > 0$ and take N so that $\sup_{x \in I} |f_n'(x) - f'(x)| < \epsilon$ for $n \ge N$ and let $M_N = \max_{1 \le k < N} \sup_{x \in I} |f_n'(x)|$ and $M = \sup_{x \in I} |f'(x)|$ then for $n < N$, $\sup_{x \in I} |f_n'(x)| \le M_N$ and for $n \ge N$,

$$\sup_{x \in I} |f_n'(x)| \le \sup_{x \in I} |f_n'(x) - f'(x)| + \sup_{x \in I} |f'(x)| < \epsilon + M$$

and hence $\sup_n \sup_{x \in I} |f_n'(x)| \le \max(M_N, M + \epsilon)$, i.e. the constant function $\max(M_N, M + \epsilon)$ is a dominating locally integrable function on A.

Assumption 3.2. *$\{f_n\}_n$ and f are assumed to be C^1, $f_n(x_0) \to f(x_0)$ for some fixed point x_0 and $f_n' \to f'$ uniformly on compacts.*

Remark 3.3. *By the fundamental theorem of calculus Assumption 3.2 is obviously equivalent to requiring that both* $f_n \to f$ *as well as* $f_n' \to f'$ *uniformly on compacts.*

Lemma 3.4. *Assumption 3.1 implies that* $f_n \to f$ *uniformly on compacts.*

Proof. Assume Assumption 3.1. Fix any $R > |x_0|$ and consider any $x \in [-R, R]$. Assume $x \geq x_0$ ($x < x_0$ is handled analogously) then for any $n \in \mathbb{N}$ we have $f_n(x) = \int_{x_0}^x f_n'(y)dy + f_n(x_0)$ and similarly $f(x) = \int_{x_0}^x f'(y)dy + f(x_0)$ and therefore

$$|f_n(x) - f(x)| \leq |f_n(x_0) - f(x_0)| + \int_{x_0}^x |f_n'(y) - f'(y)|dy \leq$$
$$|f_n(x_0) - f(x_0)| + \int_{[-R,R]} |f_n'(y) - f'(y)|dy,$$

since $|f_n'(y) - f'(y)| \leq 2|h(y)|$ it follows from dominated convergence theorem that the right-most side, which does not depend on x (other than we assume $x \in [-R, R]$), converges to zero and hence $f_n \to f$ uniformly on compacts. □

We present the first Theorem concerning the stability of quadratic variations. We consider processes admitting to quadratic variations and absolutely continuous maps, so in particular this result covers Dirichlet processes and C^1 maps.

Theorem 3.5. *Given* $0 \leq u \leq t$, *suppose that* $(\{f_n\}_n, f)$ *fulfil Assumption 3.1, for each* n X *and* X^n *have quadratic variations along the same refining sequence* $\{D_k^n\}_k$, *that* $[X^n - X]_u \xrightarrow{\mathbb{P}} 0$ *and* $(X^n - X)_u^* \xrightarrow{\mathbb{P}} 0$ *then*

$$[f_n(X^n) - f(X)]_u \xrightarrow{\mathbb{P}} 0.$$

Proof. See Appendix A.1.5. □

We will now present a nice application of Theorem 3.5, which extends the kind of maps under which Dirichlet processes are stable. The following variant of the Itô formula due to Föllmer will be used.

Theorem 3.6. *Suppose* X *is a process admitting to a quadratic variation,* f *a* C^2 *function then*

$$f(X_t) = f(X_0) + \sum_{s \leq t} \left(\Delta f(X_s) - \Delta X_s f'(X_{s_-})\right) + \int_0^t f'(X_{s_-})dX_s + \frac{1}{2}\int_0^t f''(X_{s_-})d[X, X]_s^c.$$

Let f be a function that has either a left- or right-derivative at every point. By a version of f' we will mean a function that is equal to the derivative of f at all of its differentiable points and at each non-differentiable point we define it to be one of it's directional derivatives at that point.

Theorem 3.7. *Suppose X is a Dirichlet process with fixed time jumps of finite variation, that f is an absolutely continuous function which has either a left- or right-derivative at every point and such that every version of f' is locally bounded (i.e. bounded on compacts). Then $f(X)$ is a Dirichlet process and $f(X) = Y^a + \Gamma^a$ where*

$$
\begin{aligned}
Y_t^a = f(X_0) &+ \sum_{s\leq t} \left(f(X_s) - f(X_{s-}) - \Delta X_s f'(X_{s-})\right) I_{|\Delta X_s|>a} + \int_0^t f'(X_{s-})dZ_s \\
&+ \int_0^t \int_{|x|\leq a} \left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)(\mu-\nu)(ds,dx) \\
&+ \sum_{s\leq t} \int_{|x|\leq a} \left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)\nu(\{s\},dx),
\end{aligned}
$$

here Γ^a is an adapted continuous process with zero quadratic variation and f' is any fixed version of f'.

Remark 3.8. *Note that Γ^a will depend on the version of f' that is chosen.*

Proof. Begin by defining $\Gamma = f(X) - Y^a$. Note that the jumps of $f(X)$ and Y^a coincide. Indeed, the $\Delta X_s f'(X_{s-})$ jumps in the first sum is cancelled by the jumps exceeding a in the dZ integral, the smaller jumps in the dZ integral are cancelled from the $(\mu-\nu)$ integral. Therefore Γ is continuous and adapted so the result is true if we can show that $[\Gamma]_t = 0$ and that Y^a is a semimartingale. If we let $T_m := \{\Delta X_t^* \vee X_t^* \vee Z_t^* \leq m\}$ then $\lim_{m\to\infty} \mathbb{P}(T_m > t) = 1$. So for any given $\epsilon > 0$ we may chose m such that $\mathbb{P}(T_m > t) > 1-\epsilon$ we can therefore restrict our attention to X^{T_m-} and for our purpose we can without loss of generality assume that $X = X^{T_m-}$ for some large m. Since f' is locally integrable we may approximate it below, pointwise (in the absolute value sense) by a sequence of step functions, $\{g_n\}_n$. Given such a step function $g_n = \sum_{k=1}^{m_n} c_k 1_{[a_k,a_{k+1})}$, for some increasing seuqnce of real numbers $\{a_k\}_{k=1}^{m_n}$, we now construct h_n by linear interpolating g_n on the intervals $[a_{k+1} - (a_{k+1}-a_k)/n, a_{k+1}]$. Then h_n is continuous so we may take a polynom p_n such that $|p_n(x) - h_n(x)| < 1/n$ for every $x \in [-m,m]$. It follows that $p_n(x) \to f'(x)$ pointwise and moreover $\sup_{x\in[-m,m]} |p_n(x)| \leq \sup_{x\in[-m,m]} |f'(x)| + 1 < \infty$, for all n. Moreover note that since $X_- \in \mathbb{L}$ then $p_n(X_-) \in \mathbb{L}$ and so $f'(X_-) = \lim_{n\to\infty} p_n(X_-) \in \mathcal{P}$ ($f'(X_-)$ is predictable). We now verify that Y^a is a semimartingale for each $a > 0$. This follows from arguments rather analogous to those made in the proof of Thoerem 2.1 in [1], but we include this for the sake of completeness. First note that by assumption $\sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx) < \infty$ a.s. so we can consider the localizing sequence $U_0 = 0$, $U_m = \inf\{t > U_{m-1} : \sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx) < m\}$ for $m \ge 1$. Since $\mathbb{E}\left[\int_{|x|\le a}|x|1_{s\le U_m}\mu(\{s\},dx)\right] = \mathbb{E}\left[\int_{|x|\le a}|x|1_{s\le U_m}\mu(\{s\},dx)\right]$, it then follows from monotone convergence that

$$\mathbb{E}\left[\sum_{s\le t}\int_{|x|\le a}|x|\nu(\{s\},dx)1_{U_m\le t}\right] \le \mathbb{E}\left[\left(\sum_{s\le \cdot}\int_{|x|\le a}|x|\nu(\{s\},dx)\right)_t^{U_m}\right]$$
$$= \mathbb{E}\left[\sum_{s\le t}\int_{|x|\le a}|x|1_{s\le U_m}\nu(\{s\},dx)\right] = \mathbb{E}\left[\sum_{s\le t}\int_{|x|\le a}|x|1_{s\le U_m}\mu(\{s\},dx)\right] \le m + a$$

so $\sum_{s\le t}\int_{|x|\le a}|x|\nu(\{s\},dx) < \infty$ on $\{U_m \le t\}$ and since $\mathbb{P}\left(\bigcup_{m\ge1}\{U_m \le t\}\right) = 1$ we conclude that
$\sum_{s\le t}\int_{|x|\le a}|x|\nu(\{s\},dx) < \infty$ a.s.. Because of this we may also deduce that

$$\left|\sum_{s\le t}\int_{|x|\le a}\left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)\nu(\{s\},dx)\right| \le$$
$$\sum_{s\le t}\int_{|x|\le a}|x|\left|\int_0^1 f'(X_{s-}+\theta x)d\theta - f'(X_{s-})\right|\nu(\{s\},dx) \le 2S_m\sum_{s\le t}\int_{|x|\le a}|x|\nu(\{s\},dx) < \infty,$$

a.s.. We may now also expand

$$\int_0^t\int_{|x|\le a}\left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)(\mu-\nu)(ds,dx) =$$
$$\int_0^t\int_{|x|\le a}\left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)(\tilde{\mu}-\nu_c)(ds,dx) +$$
$$\sum_{s\le t}\int_{|x|\le a}\left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)(\mu-\nu)(\{s\},dx), \tag{3.6}$$

where $\tilde{\mu}$ denotes the jump measure μ with all fixed time jumps removed.

- The first sum contains a finite number of jumps for each path and is trivially of finite variation
- It suffices to show that $(\int_0^\cdot f'(X_{s-})dZ_s)^{T_m}$ is a semimartingale for each m (so that it is a total semimartingale). Note that $f(X_{s-})$ is predictable and for $s \le T_m$ it is both bounded and predictable, so clearly the integral is a semimartingale.

- The integrand in the $(\mu-\nu)$- integral may be re-written as

$$f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})=x\left(\int_0^1 f'(X_{s-}+\theta x)d\theta-f'(X_{s-})\right)$$

this leads to

$$\left(\int_0^{\cdot}\int_{|x|\le a}\left(f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})\right)^2\nu(ds,dx)\right)^{T_m}\le\int_0^{T_m}\int_{|x|\le a}x^2 4S_m^2\nu(ds,dx)$$

which is a process with jumps smaller than a^2, so after further localization we see that this is an increasing predictable process which is locally square integrable. This implies that

$\int_0^t\int_{|x|\le a}\left(f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})\right)(\mu-\nu)(ds,dx)$ is a well-defined local martingale.

- For the final term we have the following estimate

$$\sum_{s\le t}\int_{|x|\le a}|f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})|\,\nu(\{s\},dx)\le\sum_{s\le t}\int_{|x|\le a}|x|2S_m\nu(\{s\},dx),$$

which shows that the series is summable, so it is a pure jump semimartingale.

We will from now on denote $S_m=\sup_{x\in[-m,m]}|f'(x)|$. Define $f_n(x)=f(-m)+\int_{-m}^x p_n(u)du$ for $x\in[-m,m]$ so that both $f_n(x)\to f(x)$ and $p_n(x)\to f'(x)$ on $[-m,m]$. By Theorem 3.6 it follows that

$$f_n(X_t)=f_n(X_0)+\sum_{s\le t}\left(\Delta f_n(X_s)-\Delta X_s p_n(X_{s_-})\right)+\int_0^t p_n(X_{s_-})dX_s+\frac{1}{2}\int_0^t p_n'(X_{s_-})d[X,X]_s^c.$$

We now let

$$\begin{aligned}Y_t^{n,a}&=f_n(X_0)+\sum_{s\le t}\left(f_n(X_s)-f_n(X_{s-})-\Delta X_s p_n(X_{s-})\right)I_{|\Delta X_s|>a}+\int_0^t p_n(X_{s-})dZ_s\\&+\int_0^t\int_{|x|\le a}\left(f_n(X_{s-}+x)-f_n(X_{s-})-xp_n(X_{s-})\right)(\tilde\mu-\nu_c)(ds,dx)\\&+\sum_{s\le t}\int_{|x|\le a}\left(f_n(X_{s-}+x)-f_n(X_{s-})-xp_n(X_{s-})\right)\mu(\{s\},dx),\end{aligned}\tag{3.7}$$

where $\tilde\mu$ denotes μ with all fixed time jumps removed and ν_c is ν with all fixed time jumps removed. Define $\Gamma_t^n=\int_0^t p_n(X_{s_-})dC_s+\int_0^t p_n'(X_{s_-})d[X,X]_s^c$, so that $f_n(X)=Y^{n,a}+\Gamma_n$. By part (2) of the proof of Lemma 2.5 in [1], Γ^n is a process with zero quadratic variation. We want to show $[\Gamma]_t=0$.

$$[\Gamma]_t^{\frac12}\le[\Gamma-\Gamma^n]_t^{\frac12}+[\Gamma^n]_t^{\frac12}=[f(X)-Y-(f_n(X)-Y^{n,a})]_t^{\frac12}\le[f(X)-f_n(X)]_t^{\frac12}+[Y^a-Y^{n,a}]_t^{\frac12}.$$

The first term converges to zero in probability as $n \to \infty$ by Theorem 3.5. As for the second term we make the following estimates,

$$[Y^a - Y^{n,a}]_t^{\frac{1}{2}} \leq \left[\sum_{s \leq t} (f_n(X_s) - f(X_s) - (f_n(X_{s-}) - f(X_{s-})) - \Delta X_s (p_n(X_{s-}) - f'(X_{s-}))) I_{|\Delta X_s|>$$

$$+ \left[\int_0^{\cdot} (p_n(X_{s-}) - f'(X_{s-})) dZ \right]_t^{\frac{1}{2}}$$

$$+ \left[\int_0^{\cdot} \int_{|x| \leq a} (f_n(X_{s-} + x) - f(X_{s-} + x) - (f_n(X_{s-}) - f(X_{s-})) - x(f'(X_{s-}) - p_n(X_{s-}))) (\tilde{\mu} - \nu_c$$

$$\leq \sqrt{\sum_{s \leq t} \left((f_n(X_s) - f(X_s)) I_{|\Delta X_s|>a} \right)^2} + \sqrt{\sum_{s \leq t} \left((f_n(X_{s_-}) - f(X_{s_-})) I_{|\Delta X_s|>a} \right)^2} + \sqrt{\int_0^t (p_n(X_{s-})}$$

$$+ \left(\int_0^t \int_{|x| \leq a} (f_n(X_{s-} + x) - f(X_{s-} + x) - (f_n(X_{s-}) - f(X_{s-})) - x(f'(X_{s-}) - p_n(X_{s-})))^2 \tilde{\mu}(ds,$$

$$+ \left(\sum_{s \leq t} \left(\int_{|x| \leq a} (f_n(X_{s-} + x) - f(X_{s-} + x) - (f_n(X_{s-}) - f(X_{s-})) - x(f'(X_{s-}) - p_n(X_{s-})) \mu(\{s\}$$

Since the first two sums contain only finitely many jumps the pointwise convergence ensures pathwise convergence to zero. For the third we may use the fact that $(p_n(X_{s-}) - f'(X_{s-}))^2$ is bounded by some constant so we may employ the stochastic version of the dominated convergence theorem to conclude that this term vanishes. Now we consider the fifth term (the fourth term is dealt with afterwards). We have the following dominating bound for the integrand,

$$(f_n(X_{s-} + x) - f(X_{s-} + x) - (f_n(X_{s-}) - f(X_{s-})) - x(f'(X_{s-}) - p_n(X_{s-})))^2 =$$
$$\left(x \int_0^1 (p_n(X_{s_-} + \theta x) - f'(X_{s_-} + \theta x)) d\theta - x(f'(X_{s-}) - p_n(X_{s-})) \right)^2 \leq$$
$$x^2 \left(\int_0^1 (|p_n(X_{s_-} + \theta x)| + S_m) d\theta + (S_m + |p_n(X_{s-})|) \right)^2 \leq 4x^2 (2S_m + 1)^2, \qquad (3.8)$$

which is $\tilde{\mu}(ds, dx)$-integrable, so we may apply the dominated convergence theorem (pathwise). For the fourth term we can analogously dominate the integrand by $2|x|(2S+1)$ and use dominated convergence again. □

The next Theorem instead concerns (uniform) convergence in the L^p sense, since this is a much stronger form of convergence we will obviously need moment conditions of some kind as well as some kind of tail growth conditions on $\{f_n\}_n$. The proof is rather tedious and has been deferred to the Appendix.

Theorem 3.9. *Let $(\{f_n\}_n, f)$ fulfill Assumption 3.2, $\{X^n\}_n$ and X be Dirichlet processes such that for each n, X^n and X have quadratic variations along the same refining sequence and such that $X^n \xrightarrow{ucp} X$. Assume also that X and X^n have fixed time jumps of finite variation. Letting $\mathcal{A} = \{s \le t : \mathbb{P}(\Delta X_s \neq 0)\}$ and $\mathcal{A}_n = \{s \le t : \mathbb{P}(\Delta X^n_s \neq 0)\}$ we also assume $\sum_{s\le t, s\in\mathcal{A}_n} |\Delta X^n_s| \xrightarrow{\mathbb{P}} \sum_{s\le t, s\in\mathcal{A}} |\Delta X_s|$ (i.e. the total variation of the fixed time jumps of X^n converge in probability to the total variation of the fixed time jumps of X).*
a) Suppose $|f_n'| \le U$ for some $U \in \mathbb{R}^+$, $[X]^p_t \in L^1$, $\mathbb{E}\left[[X^n - X]^p_t\right] \to 0$ with $p \in \{\frac{1}{2}, 1\}$ and
$\left\{\left(\sum_{s\le t, s\in\mathcal{A}_n} |\Delta X^n_s|\right)^{2p}\right\}_n$ *is u.i. then ,*
$[f_n(X^n) - f(X)]^p_t \xrightarrow{L^1} 0$ *as $n \to \infty$.*
b) Suppose $[X]^q_t \in L^1$, $\mathbb{E}\left[[X^n - X]^q_t\right] \to 0$ with $q \in \{1, 2\}$, $\{((X^n)^*_t)^{2q}\}_n$ is u.i., $\lim_{|x|\to\infty} \sup_n \frac{|f_n'(x)|}{|x|} \le C$, for some $C \in \mathbb{R}^+$ and $\left\{\left(\sum_{s\le t, s\in\mathcal{A}_n} |\Delta X^n_s|\right)^{2q}\right\}_n$ is u.i. then
$[f_n(X^n) - f(X)]^{\frac{q}{2}} \xrightarrow{L^1} 0$ *as $n \to \infty$.*

Proof. See Appendix A.1.6. □

A slightly stronger condition than assuming $X^n \xrightarrow{ucp} X$ and $\{((X^n)^*_t)^{2q}\}_n$ being u.i. is to instead assume $(X^n - X)^*_t \xrightarrow{L^{2q}} 0$. An example where we have this is if we consider a sequence of Dirichlet processes $X^n = Z^n + C^n$ such that $Z^n \xrightarrow{\underline{\underline{H}}^q([0,t])} Z$, $((C^n - C)^*_t)^{2q} \xrightarrow{L^{2q}} 0$ and Z^n is without fixed time jumps. We turn now to stability of integrators in the UCP-topology.

Theorem 3.10. *Suppose that $(\{f_n\}_n, f)$ fulfil Assumption 3.1, for each n X and X^n have quadratic variations along the same refining sequence $\{D^n_k\}_k$, that $[X^n - X]_t \xrightarrow{\mathbb{P}} 0$ and $(X^n - X)^*_t \xrightarrow{\mathbb{P}} 0$ then for any semimartingale Y*

$$\int_0^\cdot Y_{s-} df_n(X^n_s) \xrightarrow{ucp} \int_0^\cdot Y_{s-} df(X_s).$$

Remark 3.11. *An immediate consequence of this is that*

$$\lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\left(\left(\int_0^\cdot Y_{s-} df_n(X^m_s) - \int_0^\cdot Y_{s-} df(X_s)\right)^*_t \ge \epsilon\right)$$

$$= \lim_{m\to\infty}\lim_{n\to\infty}\mathbb{P}\left(\left(\int_0^\cdot Y_{s-}df_n(X_s^m) - \int_0^\cdot Y_{s-}df(X_s)\right)_t^* \geq \epsilon\right)$$
$$= \lim_{n\to\infty}\mathbb{P}\left(\left(\int_0^\cdot Y_{s-}df_n(X_s^n) - \int_0^\cdot Y_{s-}df(X_s)\right)_t^* \geq \epsilon\right) = 0,$$

for all $\varepsilon > 0$.

Proof. Fix $t \in \mathbb{R}^+$ and $\epsilon, c > 0$ arbitrary. By Corollary 2.8 the integrals $\int_0^\cdot Y_{s-}df_n(X_s^n)$ and $\int_0^\cdot Y_{s-}df(X_s)$ both exist and we also have

$$\int_0^\cdot Y_{s-}df_n(X_s^n) = f_n(X_s^n)Y_s - \int_0^\cdot f_n(X_{s-}^n)dY_s - [f_n(X^n), Y]$$

as well as

$$\int_0^\cdot Y_{s-}df(X_s) = f(X_s)Y_s - \int_0^\cdot f(X_{s-})dY_s - [f(X), Y],$$

so

$$\left(\int_0^\cdot Y_{s-}df_n(X_s^n) - \int_0^\cdot Y_{s-}df(X_s)\right)_t^* \leq$$
$$\left(f_n(X^n) - f(X)\right)_t^* Y_t^* + \left(\int_0^\cdot f_n(X_{s-}^n)dY_s - \int_0^\cdot f(X_{s-})dY_{s-}\right)_t^* + \left([f_n(X^n) - f(X), Y]\right)_t^* \tag{3.9}$$

and it therefore suffices to show that each term in the right hand side of (3.9) converges to zero in probability. By Lemma 3.4, $f_n \to f$ uniformly on compacts and since $\{f_n\}_n$ and f are continuous this implies that $\{f_n\}_n$ are equicontinuous. Now take take R so large that $\mathbb{P}\left(\sup_{n\geq 1}(X^n)_t^* \geq R\right) < \epsilon$ and $\delta > 0$ so small that $|f_n(x) - f_n(y)| < c/2$ when $|x - y| < \delta$, $x, y \in [-R, R]$ and all n. Pick n_1 so large that $n \geq n_1$ implies $|f_n(x) - f(x)| < c$ for $x \in [-R, R]$ and $\mathbb{P}\left((X^n - X)_t^* \geq c/2\right) < \epsilon$ then for $n \geq n_1$

$$\left(f_n(X^n) - f(X)\right)_t^* \leq \left(f_n(X^n) - f_n(X)\right)_t^* + \left(f_n(X) - f(X)\right)_t^* \leq c$$

on the set $\{(X^n - X)_t^* < c\} \cap \{\sup_{n\geq 1}(X^n)_t^* \geq R\}$ which has measure less than 2ϵ and this shows that $f_n(X^n) \xrightarrow{ucp} f(X)$ so by the continuity of the (regular) stochastic integral, the second term converges to zero in probability. Moreover for any $L \in \mathbb{R}^+$

$$\mathbb{P}\left(\left(f_n(X^n) - f(X)\right)_t^* Y_t^* \geq c\right) \leq \mathbb{P}\left(\left(f_n(X^n) - f(X)\right)_t^* Y_t^* \geq c, Y_t^* \leq L\right) + \mathbb{P}\left(Y_t^* > L\right) \leq$$
$$\mathbb{P}\left(\left(f_n(X^n) - f(X)\right)_t^* \geq \frac{c}{L}\right) + \mathbb{P}\left(Y_t^* > L\right),$$

by first letting $n \to \infty$ the first term vanishes and then by letting $L \to \infty$ the second term vanishes by continuity of probability. For the last term in (3.9) we note that by the Kunita-Watanabe inequality

$$([f_n(X^n)-f(X),Y])_t^* \le \left([f_n(X^n)-f(X)]^{\frac{1}{2}}[Y]^{\frac{1}{2}}\right)_t^* = [f_n(X^n)-f(X)]_t^{\frac{1}{2}}[Y]_t^{\frac{1}{2}}$$

and therefore, for any $L\in\mathbb{R}^+$

$$\begin{aligned}&\mathbb{P}\left(([f_n(X^n)-f(X),Y])_t^*\ge c\right)\le \mathbb{P}\left([f_n(X^n)-f(X)]_t^{\frac{1}{2}}[Y]_t^{\frac{1}{2}}\ge c\right)\le\\&\mathbb{P}\left([f_n(X^n)-f(X)]_t^{\frac{1}{2}}[Y]_t^{\frac{1}{2}}\ge c,[Y]_t^{\frac{1}{2}}\le L\right)+\mathbb{P}\left([Y]_t^{\frac{1}{2}}>L\right)\le\\&\mathbb{P}\left([f_n(X^n)-f(X)]_t^{\frac{1}{2}}\ge\frac{c}{L}\right)+\mathbb{P}\left([Y]_t^{\frac{1}{2}}>L\right),\end{aligned}$$

where the first term vanishes as $n\to\infty$ by Theorem 3.5 and the second term vanishes as $L\to\infty$, hence all four terms in (3.9) vanishes as $n\to\infty$ which completes the proof. □

Now we consider integrator stability in the L^p setting. This will naturally require us to make some type of moment assumptions on the integrand Y.

Theorem 3.12. *Assume $Y\in\underline{\underline{H}}^2([0,t])$, as well as either*
1) Hypothesis a) of Theorem 3.9 and $\{((X^n)_t^*)^2\}_n$ is u.i.
or
2) Hypothesis b) of Theorem 3.9 and $Y_t^*\in L^\infty$,
or
3) Hypothesis b) of Theorem 3.9 $\{((X^n)_t^*)^4\}_n$ is u.i., then

$$\mathbb{E}\left[\left(\int_0^\cdot Y_{s-}df_n(X_s^n)-\int_0^\cdot Y_{s-}df(X_s)\right)_t^*\right]\to 0,$$

as $n\to\infty$.

Remark 3.13. *It is in fact possible to prove Theorem 3.9 with a weaker, albeit perhaps slightly less appetizing condition than requiring that X^n and X have quadratic variations along the same refining sequence, namely that instead $f_n(X^n)$ and $f(X)$ have quadratic variations along the same refining sequence. This weaker requirement is insufficient for Theorem 3.5.*

Proof. Let $L=\|Y_t^*\|_\infty$. It suffices to show that all for terms on the right-hand side of (3.9) converges to zero in expectation. Assuming 1) we have that

$$|f_n(x)|\le|f_n(x_0)|+\left|\int_{x_0}^x|f_n'(y)|dy\right|\le|f_n(x_0)|+U(|x_0|+|x|). \tag{3.10}$$

Let $Y = M + A$ be a decomposition of Y such that $j_2(M,A) < 2\|Y\|_{\underline{\underline{H}}^2([0,t])}$. Note that

$$\mathbb{E}\left[(Y_t^*)^2\right] \le \mathbb{E}\left[2(M_t^*)^2 + 2(A_t^*)^2\right] \le 2D\mathbb{E}\left[[M]_t\right] + 2\mathbb{E}\left[\left(\int_0^t |dA_s|\right)^2\right] \le 4(D+1)\|Y\|_{\underline{\underline{H}}^2([0,t])},$$

for some constant D by the Burkholder Davis-Gundy. Therefore

$$\begin{aligned}
&\mathbb{E}\left[(f_n(X^n) - f(X))_t^* Y_t^* 1_E\right] \le \mathbb{E}\left[(|f_n(x_0)| + U|x_0| + U(X_t^* + (X^n)_t^*))\, Y_t^* 1_E\right] \le \\
&(|f_n(x_0)| + U|x_0|)\,\mathbb{E}\left[Y_t^* 1_E\right] + U\mathbb{E}\left[(X_t^* + (X^n)_t^*))Y_t^* 1_E\right] \le \\
&(|f_n(x_0)| + U|x_0|)\,\mathbb{E}\left[Y_t^* 1_E\right] + U\sqrt{E\left[(2(X_t^*)^2 + 2((X^n)_t^*)^2)1_E\right]}\sqrt{\mathbb{E}\left[(Y_t^*)^2 1_E\right]},
\end{aligned}$$

for any measurable set E, which shows that $(f_n(X^n) - f(X))_t^* Y_t^*$ is u.i. and since this term converges to zero in probability as established by the proof of Theorem 3.10 it follows that it converges to zero in expectation as well. Since $\int_0^\cdot (f_n(X^n) - f(X))\, dM_s$ is a local martingale (by martingale preservation property),

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^\cdot (f_n(X^n) - f(X))\, dY_s\right)_t^*\right] \le \mathbb{E}\left[\left(\int_0^\cdot (f_n(X^n) - f(X))\, dM_s\right)_t^*\right] + \\
&\mathbb{E}\left[\left(\int_0^\cdot (f_n(X^n) - f(X))\, dA_s\right)_t^*\right] \le D\mathbb{E}\left[\left[\int_0^t (f_n(X^n) - f(X))\, dM_s\right]^{\frac{1}{2}}\right] + \\
&\mathbb{E}\left[\int_0^t \left|f_n(X_{s-}^n) - f(X_{s-})\right| |dA_s|\right] = D\mathbb{E}\left[\sqrt{\int_0^t (f_n(X^n) - f(X))^2\, d[M]_s}\right] + \\
&\mathbb{E}\left[\int_0^t \left|f_n(X_{s-}^n) - f(X_{s-})\right| |dA_s|\right] \le D\mathbb{E}\left[(f_n(X^n) - f(X))_t^* \sqrt{[M]_t}\right] + \\
&\mathbb{E}\left[(f_n(X^n) - f(X))_t^* \int_0^t |dA_s|\right] \le D\sqrt{\mathbb{E}\left[((f_n(X^n) - f(X))_t^*)^2\right]}\sqrt{\mathbb{E}\left[[M]_t\right]} + \\
&\sqrt{\mathbb{E}\left[((f_n(X^n) - f(X))_t^*)^2\right]}\sqrt{\mathbb{E}\left[\left(\int_0^t |dA_s|\right)^2\right]} \le \\
&(D+1)\sqrt{\mathbb{E}\left[((f_n(X^n) - f(X))_t^*)^2\right]}\, 2\|Y\|_{\underline{\underline{H}}^2([0,t])}
\end{aligned}$$

where we applied both the Burkholder Davis-Gundy and the Cauchy-Schwarz inequality. From (3.10) it follows that $((f_n(X^n) - f(X))_t^*)^2 \le A + B((X_t^*)^2 + ((X^n)_t^*)^2)$ for some positive constants A, B and since $\{((X^n)_t^*)^2\}_n$ is u.i. so is $(((f_n(X^n) - f(X))_t^*)^2$ which also converges to zero in probability which then implies $\mathbb{E}\left[((f_n(X^n) - f(X))_t^*)^2\right] \to 0$ and hence
$\mathbb{E}\left[\left(\int_0^\cdot (f_n(X^n) - f(X))\, dY_s\right)_t^*\right] \to 0$. For the final term we proceed as in the proof Theorem 3.10 and note that

$([f_n(X^n) - f(X), Y])_t^* \leq [f_n(X^n) - f(X)]_t^{\frac{1}{2}} [Y]_t^{\frac{1}{2}}$. By the Lemma 2.5, the Cauchy-Schwarz and Kunita-Watanabe inequalities

$$\mathbb{E}\left[([f_n(X^n) - f(X), Y])_t^*\right] \leq \sqrt{\mathbb{E}\left[[f_n(X^n) - f(X)]_t\right]}\sqrt{\mathbb{E}\left[[Y]_t\right]} \leq$$

$$\sqrt{\mathbb{E}\left[[f_n(X^n) - f(X)]_t\right]}\sqrt{\mathbb{E}\left[2[M]_t + 2[A]_t\right]} \leq \sqrt{\mathbb{E}\left[[f_n(X^n) - f(X)]_t\right]}\sqrt{2}\sqrt{\mathbb{E}\left[[M]_t + \left(\int_0^t |dA_s|\right)^2\right]}$$

$$\leq \sqrt{\mathbb{E}\left[[f_n(X^n) - f(X)]_t\right]} 2^{\frac{3}{2}} \|Y\|_{\underline{\underline{H}}^2([0,t])},$$

which converges to zero by Theorem 3.9 a) and this concludes the proof when assuming 1).

Assuming either 2) or 3) then we may choose $R > 0$ such that $|f_n'(x)| \leq 2C|x|$ for $|x| \geq R$ and by letting
$M = \sup_n \sup_{y \in [-R,R]} |f_n'(y)|$ we have that for $|x| \leq R$

$$|f_n(x)| \leq |f_n(x_0)| + M(|x_0| + |x|),$$

while for $|x| \geq R$

$$|f_n(x)| \leq |f_n(x_0)| + 2C\left|\int_{x_0}^x |y| dy\right| \leq |f_n(x_0)| + C(x_0^2 + x^2),$$

which implies

$$|f_n(x)| \leq |f_n(x_0)| + C(x_0^2 + x^2) + M(|x_0| + |x|). \tag{3.11}$$

Assuming 2) then for the first term on the right-hand side of (3.9)

$$\mathbb{E}\left[(f_n(X^n) - f(X))_t^* Y_t^* 1_E\right] \leq$$
$$(|f_n(x_0)| + M|x_0|) L\mathbb{P}(E) + ML\mathbb{E}\left[(X_t^* + (X^n)_t^*)1_E\right] + CL\mathbb{E}\left[((X_t^*)^2 + (X^n)_t^*)^2)1_E\right],$$

which shows that $(f_n(X^n) - f(X))_t^* Y_t^*$ is also u.i., so the first term converges to zero in expectation. For the second term we proceed as in the case when assuming 1) but use the bound (3.11) instead of (3.10) and use the L^∞ bound on Y_t^* instead of the Cauchy-Schwarz inequality. For the final term,

$$\mathbb{E}\left[([f_n(X^n) - f(X), Y])_t^*\right] \leq \sqrt{L}\mathbb{E}\left[[f_n(X^n) - f(X)]^{\frac{1}{2}}\right]$$

which converges to zero by Theorem 3.9 b), this concludes the proof for the case 2). Assuming 3) then for the first term on the right-hand side of (3.9) we apply the bound in (3.11) and find that

$$\mathbb{E}\left[(f_n(X^n) - f(X))_t^* Y_t^* 1_E\right] \leq$$

$$\mathbb{E}\left[\left(|f_n(x_0)| + M(X_t^* + (X^n)_t^* + |x_0|) + C((X_t^*)^2 + (X^n)_t^*)^2)\right) Y_t^* 1_E\right] \leq$$
$$(|f_n(x_0)| + M|x_0|)\,\mathbb{E}\left[Y_t^* 1_E\right] + M\mathbb{E}\left[(X_t^* + (X^n)_t^*)Y_t^* 1_E\right] + C\mathbb{E}\left[((X_t^*)^2 + (X^n)_t^*)^2)Y_t^* 1_E\right]$$
$$\leq (|f_n(x_0)| + M|x_0|)\,\mathbb{E}\left[Y_t^* 1_E\right] + M\sqrt{\mathbb{E}\left[(2(X_t^*)^2 + 2((X^n)_t^*)^2 1_E\right]}\sqrt{\mathbb{E}\left[(Y_t^*)^2 1_E\right]} +$$
$$\sqrt{E\left[(2(X_t^*)^4 + 2((X^n)_t^*)^4)1_E)\right]}\sqrt{\mathbb{E}\left[(Y_t^*)^2 1_E\right]},$$

for any measurable set E, which shows that $(f_n(X^n) - f(X))_t^* \, Y_t^*$ is u.i.. For the second term we proceed as in the case when assuming 1) with the only difference that we use the bound (3.11) instead of (3.10) together with the assumption that $\{((X^n)_t^*)^4\}_n$ is u.i..

The third term is dealt with analogously to the case when assuming 2) but we instead invoke Theorem 3.9 b). □

We present an application of Theorem 3.10 for continuous semimartingales in the following corollary.

Corollary 3.14. *Suppose that $\{f_n\}_n$ and f fulfil Assumption 3.1. Assume that $\{X^n\}_n$ are continuous semimartingales such that $X^n \xrightarrow{ucp} X$, where X is some semimartingale. Let M and M_n denote the (unique) local martingale parts of X and X^n respectively. Suppose also that $M^n \xrightarrow{ucp} M$ and $|M^n| \leq |Z|$ for every n, for some semimartingale Z with $Z_0 \in L^1$, then $\int Y_- df_n(X^n) \xrightarrow{ucp} \int Y_- df(X)$, for any semimartingale Y.*

Proof. Let $Z' = Z - Z_0$, by Theorem V.4 in [5] Z' is pre-locally in $\underline{\underline{H}}^1$ (See Definitions 2.1 and 2.9). So let $\{T(m)\}_m$ be stopping times such that $T(m) \xrightarrow{a.s.} \infty$ and $Z^{T(m)-} \in \underline{\underline{H}}^1$. Take $\epsilon > 0$, $c > 0$ arbitrary and pick m so large that $\mathbb{P}(T(m) \leq t) < \epsilon$. Let $X^n = M^n + A^n$, $X = M + A$ be the canonical decompositions of X^n and X respectively. Since $\{X^n\}_n$ are continuous so are X, A, M, $\{M^n\}_n$ and $\{A^n\}_n$. By the continuity of $M^n - M$,

$$|M^n - M|^{T(m)} = |M^n - M|^{T(m)-} \leq 2|Z|^{T(m)-}.$$

Since $A^n - A$ is both FV and continuous it follows that

$$[M^n - M, A^n - A]_t = [A^n - A]_t = 0$$

and therefore $[X^n - X]_t = [M^n - M]_t$. Due to the Burkholder Davis-Gundy inequality

$$\|[(X^n - X)^{T(m)}]_t^{\frac{1}{2}}\|_{L^1} = \|[(M^n - M)^{T(m)}]_t^{\frac{1}{2}}\|_{L^1} \leq C_1 \|((M^n - M)^{T(m)})_t^*\|_{L^1},$$

For some positive constant C_1. Meanwhile $((M^n - M)^{T(m)})^*_t \leq 2(Z^{T(m)-})^*_t$, however by Theorem V.2 in [5]

$$\|(Z^{T(m)-})^*_t\|_{L^1} \leq \left(\|(Z'^{T(m)-})^*_t\|_{L^1} + \|Z_0\|_{L^1}\right) \leq C_2\|Z'^{T(m)-}\|_{\underline{\underline{H}}^1} + \|Z_0\|_{L^1} < \infty,$$

for some positive constant C_2. Since $(M^n - M)^{T(m)} \xrightarrow{ucp} 0$ and $|M^n - M|^{T(m)} \leq 2|Z|^{T(m)-} \in L^1$, it follows from a variant of the dominated convergence theorem that $\lim_n \|((M^n - M)^{T(m)})^*_t\|_{L^1} = 0$ and therefore $[(X^n - X)^{T(m)})]^{\frac{1}{2}}_t \xrightarrow{L^1} 0$. By the Markov inequality

$$\lim_{n\to\infty} \mathbb{P}\left([(X^n - X)^{T(m)})]_t \geq c\right) = \lim_{n\to\infty} \mathbb{P}\left([(X^n - X)^{T(m)})]^{\frac{1}{2}}_t \geq \sqrt{c}\right) \leq$$
$$\lim_{n\to\infty} \frac{1}{\sqrt{c}}\|[(X^n - X)^{T(m)})]^{\frac{1}{2}}_t\|_{L^1} = 0$$

for any $c > 0$ so in other words $[(X^n - X)^{T(m)})]_t \xrightarrow{\mathbb{P}} 0$. Hence,

$$\mathbb{P}([(X^n - X)]_t \geq c) \leq \mathbb{P}([(X^n - X)^{T(m)})]_t \geq c) + \mathbb{P}(T(m) \leq t) \leq$$
$$\mathbb{P}([(X^n - X)^{T(m)})]_t \geq c) + \epsilon,$$

which converges to ϵ which was arbitrary and therefore the hypothesis in Theorem 3.10 is fulfilled and the result follows. □

4 Jump truncation

By a jump truncation we mean a modification of some given process X where jumps below some level are not present. Jump truncations serve as important applications for our results. For processes with jumps of finite variation one may simply discard all jumps of X with modulus less than say $a > 0$ and let $a \to 0^+$ and in the limit we retain our original process. Given a cadlag process X with jumps of finite variation we define

$$X(a)_t = X^c_t + \sum_{s\leq t} \Delta X_s I_{|\Delta X_s| \geq a}.$$

In the general case however, $X(a)$ will not have a well-defined limit as $a \to 0^+$ since $X^d = \sum_{s\leq .} \Delta X_s$ is not well-defined. We will denote by μ the jump measure corresponding to a given cadlag process X and we will denote by ν the compensator measure of μ w.r.t. the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ (see [3] ch. II for details on this). Suppose now that we are given a cadlag process $X = Z + C$ which is the sum of a semimartingale Z and a continuous adapted process C. This means that all jumps of X belong to the semimartingale part. We may represent X by

$$X = M^c + b + \sum_{s\leq \cdot} \Delta X_s I_{|\Delta X_s|>1} + \int_0^{\cdot} \int_{|x|\leq 1} x d(\mu - \nu)(s,x) + C$$

where M^c is the continuous part of the local martingale part of Z and b is a process of finite variation. Given a process $X = Z + C$ of the form above we define for $a < 1$,

$$\widehat{X}(a)_t = M^c_t + b(a)_t + \sum_{s\leq t} \Delta X_s I_{|\Delta X_s|>1} + \int_0^t \int_{a\leq|x|\leq 1} x d(\mu - \nu)(s,x) + C_t.$$

Notice that $\widehat{X}(a)$ only has jumps greater or equal to a.

Lemma 4.1. *Let X be a Dirichlet process whose jumps at fixed times have finite variation, then $\widehat{X}(a) \xrightarrow{ucp} X$ as well as $[\widehat{X}(a) - X]_t \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$.*

Proof. See Appendix 4.1 □

Definition 4.2. *Given a cadlag process X we define the following set of reals*

$$\mathbb{A}_X := \left\{a \in \mathbb{R}^+ : \mathbb{P}\left(\exists s \leq t : |\Delta X_s| = a\right) > 0\right\}.$$

Remark 4.3. *Since X is cadlag it follows that $\mathbb{A}_X$ is countable*

4.1 The case with finite variation of jumps

We now take a closer look at jump truncation for processes with jumps of finite variation. Assuming that Z has jumps of finite variation then $(X - X(a))^*_t \leq \sum_{s\leq t} |\Delta X_s| I_{|\Delta X_s|\leq a}$ so it follows that $(X - X(a))^*_t \to 0$ a.s. when $a \to 0$ and therefore, obviously, $X(a) \xrightarrow{ucp} X$ as $a \to 0$. Also $[X - X(a)]_t = \sum_{s\leq t} (\Delta X_s)^2 I_{|\Delta X_s|\leq a}$ which also converges a.s. to zero since the l^1 norm dominates the l^2 norm. Moreover this kind of truncation actually has some slightly peculiar properties, one may wonder whether $X^n \xrightarrow{ucp} X$ implies $X^n(a) \xrightarrow{ucp} X(a)$ for all $a > 0$ and if $[X^n - X]_t \xrightarrow{\mathbb{P}} 0$ implies $[X^n(a) - X(a)]_t \xrightarrow{\mathbb{P}} 0$. In general neither is true, as the following example shows.

Example 4.4. *Let X be a Poisson process with intensity 1 and let $X^n = (1 - \frac{1}{n})X$ then $X^n \xrightarrow{ucp} X$ as well as $[X^n - X]_t = \frac{1}{n^2}[X]_t \to 0$ as $n \to \infty$ but $X^n(1) = 0$*

for all n so $X^n(1) \xrightarrow{ucp} 0 \neq X(1) = X$ and $[X^n(1) - X(1)]_t \overset{\mathbb{P}}{\not\to} 0$. Notice that in this example however, that there is only a single jump size which is troublesome but we may generalize this example to have countably many troublesome points as follows. Let $X_t = \sum_{k=1}^{\infty} \frac{1}{k^2} N_k(t)$, where $\{N_k(t)\}_k$ are i.i.d. Poisson processes with intensity 1. Setting $X^n = (1 - \frac{1}{n})X$ just as before shows despite $X^n \xrightarrow{ucp} X$ as well as $[X^n - X]_t = \frac{1}{n^2}[X]_t \to 0$ as $n \to \infty$ we get $X^n(\frac{1}{k^2}) \overset{ucp}{\not\longrightarrow} X(\frac{1}{k^2})$ and $[X^n(\frac{1}{k^2}) - X(\frac{1}{k^2})]_t \overset{\mathbb{P}}{\not\to} 0$ for any $k \in \mathbb{N}$.

The above example illustrates the difficulty of truncation when the distribution of the jump size of X is non-continuous. There are however some reasonable assumptions that can help us work around this problem.

Lemma 4.5. *Suppose $\{X^n\}_n$ and X are cadlag processes withs jumps of finite variation. If $X^n \xrightarrow{ucp} X$ then the following three statements are equivalent,*
1) $(X^n)^c \xrightarrow{ucp} X^c$
2) $\sum_{s \le t} \Delta X_s^n \xrightarrow{ucp} \sum_{s \le t} \Delta X_s$
3) $\lim_{a \to 0^+, a \notin \mathbb{A}_X} \lim_{n \to \infty} \mathbb{P}\left(\left(\sum_{s \le \cdot} \Delta X_s^n I_{|\Delta X_s^n| \le a}\right)_t^* \ge c\right) = 0$ for all $c > 0$ and $t \in \mathbb{R}^+$.
Moreover if either one of the three above conditions hold then $X^n(a) \xrightarrow{ucp} X(a)$ for any $a \notin \mathbb{A}_X$. Finally, if assume we the slightly stronger condition,
4) $\lim_{a \to 0^+} \lim_{n \to \infty} \mathbb{P}\left(\sum_{s \le t} |\Delta X_s^n| I_{|\Delta X_s^n| \le a} \ge c\right) = 0$ for all $c > 0$ and $t \in \mathbb{R}^+$ then

$$\lim_{a \to 0^+} \lim_{n \to \infty} \mathbb{P}\left((X^n(a) - X(a))_t^* \ge c\right) = 0,$$

for all $c > 0$.

Remark 4.6. *Notice that if $X^n \xrightarrow{ucp} X$ and we assume one of 1)-3) then (obviously)*

$$\lim_{a \to 0^+, a \notin \mathbb{A}_X} \lim_{n \to \infty} \mathbb{P}\left((X^n(a) - X(a))_t^* \ge c\right) = 0,$$

in comparison to the final statement of the Lemma.

Proof. See Appendix A.1.4. □

Definition 4.7. *For $X(a)$ we define the following properties,*
(V1): X and $\{X^n\}_n$ have jump processes of finite variation and either one of 1)-3) in Lemma 4.5 holds true.
(V2): X and $\{X^n\}_n$ have jump processes of finite variation and 4) in Lemma 4.5 holds true.

Lemma 4.8. *Suppose $a \notin \mathbb{A}_X$, $[X^n - X]_t \xrightarrow{\mathbb{P}} 0$ and $X^n \xrightarrow{ucp} X$ then $[X^n(a) - X(a)]_t \xrightarrow{\mathbb{P}} 0$ and if we assume assumption (V2) then*

$$\lim_{a \to 0^+} \lim_{n \to \infty} \mathbb{P}\left([X^n(a) - X(a)]_t \geq c\right) = 0,$$

for all $c > 0$.

Proof. Since

$$[(X^n)^c - X^c, J^n - J]_t = \sum_{s \leq t} \Delta((X^n)^c - X^c)_s \Delta(J^n - J)_s = 0,$$

it follows that $[X^n - X.]_t = [(X^n)^c - X^c]_t + [J^n - J]_t$, since the left hand side converges to zero and both terms on the right are non-negative we conclude that $[(X^n)^c - X^c]_t \xrightarrow{\mathbb{P}} 0$. Similarly $[X^n(a) - X(a)]_t = [(X^n)^c - X^c]_t + [J^n(a) - J(a)]_t$ so it suffices to show that $[J^n(a) - J(a)]_t \xrightarrow{\mathbb{P}} 0$. A similar approach to the proof of the implication 3) $\to$ 2) in Lemma 4.5 shows that also $[J^n(a) - J(a)]_t \xrightarrow{\mathbb{P}} 0$ if $a \notin \mathbb{A}_X$ (indeed as in that proof on the set $B_{(k,n)}$ we have $[J^n(a) - J(a)]_t = \sum_{l=1}^{k}(\Delta(X^n - X))^2_{T_l})$. Let us now assume assumption 4) of Lemma 4.5 then, since the l^1-norm dominates the l^2-norm, $\lim_{a \to 0^+} \lim_{n \to \infty} \mathbb{P}\left(\sum_{s \leq t} |\Delta X^n_s|^2 I_{|\Delta X^n_s| \leq a} \geq c\right) = 0$ and therefore we may take $a' \notin \mathbb{A}_X$ so small and n_1 so large that $a \leq a'$ and $n \geq n_1$ implies $\mathbb{P}\left(\sum_{s \leq t} |\Delta X^n_s|^2 I_{|\Delta X^n_s| \leq a} \geq c\right) < \epsilon$ as well as $\mathbb{P}\left(\sum_{s \leq t} |\Delta X_s|^2 I_{|\Delta X_s| \leq a} \geq c\right) < \epsilon$. Letting

$$B_n = \left\{\sum_{s \leq t} |\Delta X^n_s|^2 I_{|\Delta X^n_s| \leq a} < c/9\right\} \cap \left\{\sum_{s \leq t} |\Delta X_s|^2 I_{|\Delta X_s| \leq a} \geq c/9\right\}$$
$$\cap \left\{[J^n(a') - J(a')]_t < c/9]\right\}$$

so on B_n,

$$\begin{aligned}[J^n(a) - J(a)]_t &\leq 3\left([J^n(a') - J^n(a)]_t + [J^n(a') - J(a')]_t + [J(a') - J(a)]_t\right) \\ &< \sum_{s \leq t} |\Delta X^n_s|^2 I_{|\Delta X^n_s| \leq a} + c/3 + \sum_{s \leq t} |\Delta X_s|^2 I_{|\Delta X_s| \leq a} < 3(c/9 + c/9 + c/9) = c,\end{aligned}$$

where we used Lemma 2.4 together with the inequality $(x + y + z)^2 \leq 3(x^2 + y^2 + z^2)$ in the first step. Hence,

$$\mathbb{P}\left([J^n(a) - J(a)]_t \geq c\right) \leq \mathbb{P}\left(\{[J^n(a) - J(a)]_t \geq c\} \cap B_n\right) + \mathbb{P}(B_n^c) = \mathbb{P}(B_n^c) < 3\epsilon,$$

when $a \le a'$ and $n \ge n_1$, which shows the final statement.

$$[J^n(a) - J(a)]_t \le \sum_{s \le t} \Delta(X^n - X)_s^2 I_{\Delta|X^n - X|_s \le a},$$

converges to zero a.s. for each $t \in \mathbb{R}^+$ it follows that
$\lim_{a \to 0^+} \lim_{n \to \infty} \mathbb{P}([J^n(a) - J(a)]_t \ge c) = 0$. If we again apply Lemma 4.5,

$$[X^n(a) - X(a)]_t = [(X^n)^c - X^c + J^n(a) - J(a)]_t \le \sqrt{[(X^n)^c - X^c]_t} + \sqrt{[J^n(a) - J(a)]_t}$$

and therefore $[X^n(a) - X(a)]_t \xrightarrow{\mathbb{P}} 0$ for $a \notin \mathbb{A}_X$ and
$\lim_{a \to 0^+} \lim_{n \to \infty} \mathbb{P}([X^n(a) - X(a)]_t \ge c) = 0$. □

The following example serves as a reminder that $X^n \xrightarrow{ucp} X$ does in general not imply $[X^n - X]_t \xrightarrow{\mathbb{P}} 0$ or vice versa.

Example 4.9. *Take $X \equiv 0$ and let*

$$X^n = 2^{-(n+1)/2} \sum_{k=1}^{2^{n-1}} I_{[\frac{2(k-1)}{2^n}, \frac{2k-1}{2^n})}(\cdot) - I_{[\frac{2k-1}{2^n}, \frac{2k}{2^n})}(\cdot).$$

We then have that $(X^n)^*_\infty = 2^{-(n+1)/2} \to 0$ but $[X^n]_1 = (2^{-(n+1)/2})^2(2(2^n - 1) + 1) = 1 - 2^{-(n+1)} \to 1$. Which shows that $[X^n - X]_t \not\xrightarrow{\mathbb{P}} 0$ so that

$$X^n \xrightarrow{ucp} X \not\Rightarrow [X^n - X]_t \xrightarrow{\mathbb{P}} 0.$$

On the other hand take any semimartingale Z plus a sequence $\{V^n\}_n$ of continuous processes of zero quadratic variation not converging to zero, then if we let $X^n := V^n + Z$ then $[X^n - Z]_t = 0$ for all $t \in \mathbb{R}^+$.

We now show that when we are dealing with jumps finite variation then under $(V1)$ or $(V2)$ conditions, taking limits over n in Theorem 3.10 commutes in a sense with taking limits over the truncation level a.

Theorem 4.10. *Let $(X, \{X^n\}_n)$, $(f, \{f_n\}_n)$ fulfil hypothesis of Theorem 3.5.*
1) If $\{X^n\}_n$ has the (V1) property from Definition 4.7 then

$$\lim_{a \to 0^+, a \notin \mathbb{A}_X} \lim_{n \to \infty} \mathbb{P}\left(\left(\int_0^\cdot Y_{s-} df_n(X^n(a)_s) - \int_0^\cdot Y_{s-} df(X_s)\right)_t^* \ge \epsilon\right) =$$
$$\lim_{n \to \infty} \lim_{a \to 0^+, a \notin \mathbb{A}_X} \mathbb{P}\left(\left(\int_0^\cdot Y_{s-} df_n(X^n(a)_s) - \int_0^\cdot Y_{s-} df(X_s)\right)_t^* \ge \epsilon\right) = 0,$$

for every $\varepsilon > 0$.
2) If $\{X^n\}_n$ has the (V2) property from Definition 4.7 then

$$\lim_{a\to 0^+}\lim_{n\to\infty}\mathbb{P}\left(\left(\int_0^{\cdot} Y_{s-}df_n(X^n(a)_s) - \int_0^{\cdot} Y_{s-}df(X_s)\right)_t^* \geq \epsilon\right) =$$
$$\lim_{n\to\infty}\lim_{a\to 0^+}\mathbb{P}\left(\left(\int_0^{\cdot} Y_{s-}df_n(X^n(a)_s) - \int_0^{\cdot} Y_{s-}df(X_s)\right)_t^* \geq \epsilon\right) = 0,$$

for every $\varepsilon > 0$.

Proof. We start with 1). Assume that $a \notin \mathbb{A}_X$. From the integration by parts formula we have

$$\int_0^{\cdot} Y_{s-}df_n(X^n(a)_s) = f_n(X^n(a))Y - f_n(X_0)Y_0 - \int_0^{\cdot} f_n(X^n(a)_s)dY_s - [f_n(X^n(a)), Y], \tag{4.12}$$

so to prove the first claim it will suffice to show that the right-hand side above converges in ucp to $\int Y_- df(X(a))$. Since $X^n(a) \xrightarrow{ucp} X(a)$ and since $f_n \to f$ uniformly by Lemma 3.4 we have that $f_n(X^n(a)) \xrightarrow{ucp} f_n(X^n(a))$ we see that the first (by continuity of probability) and second (by continuity of the stochastic integral) term in (4.12) converges in ucp to $f(X)Y$ and $\int f_n(X^n(a))_- dY$ respectively.
We now study the last term in 4.12, $[f_n(X^n(a)), Y]$. Since $[X^n(a) - X(a)]_t \xrightarrow{\mathbb{P}} 0$ by Lemma 4.8 and $X^n(a) \xrightarrow{ucp} X(a)$ Theorem it follows from Theorem 3.9 that $[f_n(X(a)^n) - f(X(a))]_t \xrightarrow{\mathbb{P}} 0$ and therefore we may argue as in the proof of Theorem 3.5 to show that $[f_n(X^n(a)), Y] \xrightarrow{ucp} [f(X(a)), Y]$. So we have shown that

$$\int_0^{\cdot} Y_{s-}df_n(X^n(a)_s) \xrightarrow{ucp} f(X(a))Y - f(X_0)Y_0 - \int_0^{\cdot} f(X(a)_s)dY_s - [f(X(a)), Y]. \tag{4.13}$$

Since f is continuous and $X(a) \xrightarrow{ucp} X$ it follows that $f(X(a)) \xrightarrow{ucp} f(X)$ so the first and second term converges in ucp to $f(X)Y$ and $\int f(X)_- dY$ respectively. For the last term in (4.13) we have $[f(X(a)), Y]_t \leq [f(X(a))]_t^{\frac{1}{2}}[Y]_t^{\frac{1}{2}}$, but by Theorem 3.5 (the fact that we are are working with a continuous parameter will pose no problem, as is evident by the proof) $[f(X(a)) - f(X)]_t \xrightarrow{\mathbb{P}} 0$ which implies $[f(X(a))]_t \xrightarrow{\mathbb{P}} [f(X)]_t$ so $[f(X(a)), Y] \xrightarrow{ucp} [f(X), Y]$ as $a \to 0^+$, so the right-hand side of 4.13 converges in ucp to

$$f(X)Y - f(X_0)Y_0 - \int_0^{\cdot} f(X_s)dY_s - [f(X), Y] = \int_0^{\cdot} Y_{s-}df(X_s)$$

as $a \to 0^+$, this shows

$$\lim_{a\to 0^+, a\notin \mathbb{A}_X} \lim_{n\to\infty} \mathbb{P}\left(\left(\int_0^\cdot Y_{s-} df_n(X^n(a)_s) - \int_0^\cdot Y_{s-} df(X_s)\right)_t^* \geq \epsilon\right) = 0.$$

To show the second equality of 1) note that by Lemma 2.4,

$$[f_n(X^n(a)) - f(X)]_t^{\frac{1}{2}} \leq [f_n(X^n(a)) - f_n(X^n)]_t^{\frac{1}{2}} + [f_n(X^n) - f(X)]_t^{\frac{1}{2}},$$

by first letting $a \to 0^+$ the first term on the right-hand side above converges to zero in probability. Letting $n \to \infty$ makes the second term converge to zero, i.e. $[f_n(X^n(a)) - f(X)]_t^{\frac{1}{2}} \xrightarrow{\mathbb{P}} 0$ which implies $[f_n(X^n(a)) - f(X), Y]_t \xrightarrow{ucp} 0$. By the fact that f_n is C^1, $f_n(X^n(a)) \xrightarrow{ucp} f_n(X^n)$ as $a \to 0^+$ and using the integration by parts formula concludes the proof of 1). Showing 2) is completely analogous. □

A.1 Additional proofs

A.1.1 Proof of Lemma 2.2

Proof. If we can prove that

1) if X and Y have quadratic variations along $\{D_k\}_k$ then so does $X + Y$,

as well as

2) if g is a absolutely continuous function and X has quadratic variation along $\{D_k\}_k$ then so does $g(X)$,

then it follows by 2) that $g_1(X)$ and $g_2(Y)$ have quadratic variations along $\{D_k\}_k$ and by 1) that $g_1(X) + g_2(Y)$ does as well. We start by showing 1). Since X and Y have a quadratic variations along $\{D_k\}_k$ this implies that

$$\lim_{k,l\to\infty} \mathbb{P}\left(\left|\sum_{t_i\in D_k, t_i\leq s} \left(X_{t_i} - X_{t_{i-1}}\right)^2 - \sum_{t_i\in D_l, t_i\leq s} \left(X_{t_i} - X_{t_{i-1}}\right)^2\right| \geq c\right) =$$

$$\lim_{k,l\to\infty} \mathbb{P}\left(\sum_{t_i \ or \ t_{i-1}\in D_l^t\setminus D_k^s} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \geq c\right) = 0$$

and similarly for Y,

$$\lim_{k,l\to\infty} \mathbb{P}\left(\left|\sum_{t_i\in D_k, t_i\leq s} \left(Y_{t_i} - Y_{t_{i-1}}\right)^2 - \sum_{t_i\in D_l, t_i\leq s} \left(Y_{t_i} - Y_{t_{i-1}}\right)^2\right| \geq c\right) = 0,$$

for all $c > 0$, $0 < s \leq t$ and $n \in \mathbb{N}$, where we used the notation $D_k^s = \{t_i \in D_k : t_i \leq s\}$. Now by the Cauchy-Schwarz inequality,

$$\left| \sum_{t_i \in D_k, t_i \leq s} \left(X_{t_i} - X_{t_{i-1}}\right)\left(Y_{t_i} - Y_{t_{i-1}}\right) - \sum_{t_i \in D_l, t_i \leq s} \left(X_{t_i} - X_{t_{i-1}}\right)\left(Y_{t_i} - Y_{t_{i-1}}\right) \right| =$$

$$\left| \sum_{t_i \ or \ t_{i-1} \in D_l^s \setminus D_k^s} \left(X_{t_i} - X_{t_{i-1}}\right)\left(Y_{t_i} - Y_{t_{i-1}}\right) \right| \leq$$

$$\left(\sum_{t_i \ or \ t_{i-1} \in D_l^s \setminus D_k^s} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \right)^{\frac{1}{2}} \left(\sum_{t_i \ or \ t_{i-1} \in D_l^t \setminus D_k^t} \left(Y_{t_i} - Y_{t_{i-1}}\right)^2 \right)^{\frac{1}{2}}.$$

Now

$$\sum_{t_i \ or \ t_{i-1} \in D_l^s \setminus D_k^s} \left(X_{t_i} + Y_{t_i} - X_{t_{i-1}} - Y_{t_{i-1}}\right)^2 \leq$$

$$\sum_{t_i \ or \ t_{i-1} \in D_l^s \setminus D_k^s} \left(X_{t_i} - X_{t_{i-1}}\right)^2 + \sum_{t_i \ or \ t_{i-1} \in D_l^s \setminus D_k^s} \left(Y_{t_i} - Y_{t_{i-1}}\right)^2 +$$

$$\left(\sum_{t_i \ or \ t_{i-1} \in D_l^s \setminus D_k^s} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \right)^{\frac{1}{2}} \left(\sum_{t_i \ or \ t_{i-1} \in D_l^s \setminus D_k^s} \left(Y_{t_i} - Y_{t_{i-1}}\right)^2 \right)^{\frac{1}{2}},$$

and all three terms above converge to zero in probability as $l, k \to \infty$ which shows that $\{S_n(X+Y)_s\}_n$ is Cauchy in probability and therefore there exists a finite random variable $S(X+Y)_s$, if we can show that $S(X+Y)_s$ fulfils (1.1) then this shows 1). First note that since $S(X+Y)_s$ is finite and increasing in s it is a finite variation process so we may decompose $S(X+Y)_s$ as $S(X+Y)_s = S_s^c + \sum_{u \leq s} (\Delta S(X+Y))_u$ where S^c is a continuous increasing process. It now suffices to show that $(\Delta S(X+Y))_u = (\Delta(X+Y))_u^2$. To this end, we may take a subsequence of $\{S_n(X+Y)_s\}_n$, $\{S_{n_k}(X+Y)_s\}_k$ such that $S_{n_k}(X+Y)_s \xrightarrow{a.s.} S(X+Y)_s$. Using this subsequence we see that

$$(\Delta S(X+Y))_u = \lim_{k \to \infty} \sum_{t_i \in D_k, \ t_i \leq u} \left(X_{t_i} + Y_{t_i} - X_{t_{i-1}} - Y_{t_{i-1}}\right)^2 -$$

$$\lim_{k \to \infty} \sum_{t_i \in D_k, \ t_i < u} \left(X_{t_i} + Y_{t_i} - X_{t_{i-1}} - Y_{t_{i-1}}\right)^2 = \lim_{k \to \infty} \left(X_{t_k'} + Y_{t_k'} - X_{t_k''} - Y_{t_k''}\right)^2,$$

where $t_k' = \min\{t_i \in D_k : t_i \geq u\}$ and $t_k'' = \max\{t_i \in D_k : t_i < u\}$. Since $D_k \subseteq D_{k+1}$ it follows that $t_k' \downarrow t$, $t_k'' \uparrow t$ and since X and Y are cadlag it follows that

$$\lim_{k \to \infty} X_{t_k'} + Y_{t_k'} - X_{t_k''} - Y_{t_k''} = \Delta X_t + \Delta Y_t = \Delta(X+Y)_u,$$

hence $(\Delta S(X+Y))_u = (\Delta(X+Y)_u)^2$ and this finishes the proof of 1). We now show 2). Since g is absolutely continuous it is easy to verify that $h(\theta) = g(X_{t_i} + \theta(X_{t_i} - X_{t_{i-1}}))$ is (pathwise) as well and therefore $h'(\theta)$ exists almost everywhere (in the Lebesgue sense). For any point θ, where $h'(\theta)$ exists we have by the chain rule that $h'(\theta) = (X_{t_i} - X_{t_{i-1}})g'(X_{t_i} + \theta(X_{t_i} - X_{t_{i-1}}))$. Noting that $h(1) - h(0) = g(X_{t_i}) - g(X_{t_{i-1}})$ and meanwhile, by Theorem 10 of section 5, chapter 6 in [6], $h(1) - h(0) = \int_0^1 h'(\theta)d\theta$ from which we derive the following representation

$$g(X_{t_i}) - g(X_{t_{i-1}}) = \left(X_{t_i} - X_{t_{i-1}}\right) \int_0^1 g'\left(X_{t_{i-1}} + \theta(X_{t_i} - X_{t_{i-1}})\right) d\theta.$$

Let $A_R = \{X_t^* \leq R\}$, $M_R = \int_{-R}^R |g'(x)|dx$ and note that on A_R

$$\left|\int_0^1 g'\left(X_{t_{i-1}} + \theta(X_{t_i} - X_{t_{i-1}})\right) d\theta\right| \leq \int_0^1 \left|g'\left(X_{t_{i-1}} + \theta(X_{t_i} - X_{t_{i-1}})\right)\right| d\theta \leq M_R.$$

Now

$$\lim_{k,l\to\infty} \mathbb{P}\left(\left|\sum_{t_i\in D_k, t_i\leq s} \left(g(X_{t_i}) - g(X_{t_{i-1}})\right)^2 - \sum_{t_i\in D_l, t_i\leq s} \left(g(X_{t_i}) - g(X_{t_{i-1}})\right)^2\right| \geq c\right) =$$

$$\lim_{k,l\to\infty} \mathbb{P}\left(\sum_{t_i \;\; or \;\; t_{i-1}\in D_l^s\setminus D_k^s} \left(g(X_{t_i}) - g(X_{t_{i-1}})\right)^2 \geq c\right) =$$

$$\lim_{k,l\to\infty} \mathbb{P}\left(\sum_{t_i \;\; or \;\; t_{i-1}\in D_l^s\setminus D_k^s} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \left(\int_0^1 g'\left(X_{t_{i-1}} + \theta(X_{t_i} - X_{t_{i-1}})\right) d\theta\right)^2 \geq c\right) \leq$$

$$\lim_{k,l\to\infty} \mathbb{P}\left(\sum_{t_i \;\; or \;\; t_{i-1}\in D_l^s\setminus D_k^s} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \geq \frac{c}{M_R^2}\right) + \mathbb{P}(A_R^c) = \mathbb{P}(A_R^c)$$

letting $R \to \infty$ makes the right-most side above vanish by continuity of probability and this shows that $S_n(g(X))_u \xrightarrow{\mathbb{P}} S(g(X))_u$ for $u \leq t$ where $S(g(X))$ is some increasing finite process. As in 1) it now suffices to show that $(\Delta S(g(X)))_u = (\Delta g(X))_u^2$ for $u \leq t$. We again choose a subsequence $\{S_{n_k}(g(X))_u\}_{n_k}$ such that $S_{n_k}(g(X))_u \xrightarrow{a.s.} S(g(X))_u$. Now

$$(\Delta S(g(X)))_u = \lim_{k\to\infty} \left(g(X_{t_k'}) - g(X_{t_k''})\right)^2 = (\Delta g(X))_u^2,$$

with t_k' and t_k'' as defined before using the same kind of argument as in 1). □

A.1.2 Proof of Lemma 2.3

Proof. Expanding $S_n(X,Y)$ at time $s \leq t$ gives us

$$S_k(X,Y)_s = \frac{1}{2}\left(\sum_{t_i \in D_k, t_i \leq s} (Y_{t_{i+1}} + X_{t_{i+1}} - Y_{t_i} - X_{t_i})^2 \right.$$
$$\left. - \sum_{t_i \in D_k, t_i \leq s} (Y_{t_{i+1}} - Y_{t_i})^2 - \sum_{t_i \in D_k, t_i \leq s} (X_{t_{i+1}} - X_{t_i})^2 \right),$$

by assumption the right-hand side converges in probability to $\frac{1}{2}\left([X+Y]_s - [Y]_s - [X]_s\right)$ which can be decomposed as

$$\frac{1}{2}\left([X+Y]_s - [Y]_s - [X]_s\right) = \frac{1}{2}\left([X+Y]^c_s - [Y]^c_s - [X]^c_s\right) +$$
$$\frac{1}{2}\left(\sum_{u \leq t}(\Delta(X+Y))^2_u - \sum_{u \leq t}(\Delta X))^2_u - \sum_{u \leq t}(\Delta Y)^2_u\right) =$$
$$\frac{1}{2}\left([X+Y]^c_s - [Y]^c_s - [X]^c_s\right) + \sum_{u \leq t}(\Delta X)_u(\Delta Y)_u$$

with $\frac{1}{2}\left([X+Y]^c_s - [Y]^c_s - [X]^c_s\right)$ as $[X,Y]^c_s$ the result follows. □

A.1.3 Proof of Lemma 4.1

Proof. Let $\tilde{X}(a) := \widehat{X}(a)_t - X_t = \int_0^t \int_{0<|x|<a} x d(\mu - \nu)(s,x) + b - b(a)$. Since b is of finite variation (and therefore has jumps of finite variation) we trivially have that $[b - b(a)]_t \xrightarrow{a.s.} 0$ as well as $b - b(a) \xrightarrow{ucp} 0$. Now we rewrite

$$\int_0^t \int_{0<|x|<a} x(\mu - \nu)(ds,dx) = \int_0^t \int_{0<|x|<a} x(\tilde{\mu} - \nu_c)(ds,dx) + \sum_{s \leq t} \int_{0<|x|<a} x\mu(\{s\},dx),$$

where $\tilde{\mu}$ is the measure μ with all fixed time jumps removed and ν_c is the compensator of $\tilde{\mu}$ (which is the time-continuous part of the compensator of μ). By Lemma 2.4,

$$[\tilde{X}(a)]_t^{\frac{1}{2}} \leq [b - b(a)]_t^{\frac{1}{2}} + \left[\int_0^{\cdot} \int_{|x|<a} x^2(\tilde{\mu} - \nu_c)(ds,dx)\right]_t + \left[\sum_{s \leq \cdot} \int_{|x|<a} |x|\mu(\{s\},dx)\right]_t =$$
$$[b - b(a)]_t^{\frac{1}{2}} + \int_0^t \int_{0<|x|<a} x^2 \tilde{\mu}(ds,dx) + \left(\sum_{s \leq t} \int_{|x|<a} |x|\mu(\{s\},dx)\right)^2$$

since $\int_0^t \int_{|x|\leq 1} x^2\mu(ds,dx) \leq [X]_t < \infty$ a.s. and $\sum_{s\leq t}\int_{|x|\leq 1}|x|\mu(\{s\},dx) < \infty$ a.s., it follows that $[\widehat{X}(a)-X]_t$ converges to zero a.s. (and hence in probability) as $a \to 0$. If we let $N(a)_t = \int_0^t \int_{|x|<a} xd(\tilde{\mu}-\nu_c)(ds,dx)$ then $N(a)$ is a local martingale with $N(a)_0 = 0$ for every $a > 0$ by Theorem 2.21 in [3] (since ν_c is the compensator measure of $\tilde{\mu}$). Let $\{S_k\}_{k\geq 0}$ be a sequence of stopping times such that $N(a)^{S_k}$ is a martingale for every k. Let $T_0 = 0$ and for $k \geq 1$, $T_k := \inf\{t > T_{k-1} : (b-b(a))_t^* \vee |N(a)_t| \geq k\}$. Let $U_k = T_k \wedge S_k$ then $N(a)^{U_k}$ is a martingale for every k with $|N(a)^{U_k}| \leq k+a$. For a given t and $\epsilon > 0$ we may take k such that $\mathbb{P}(U_k \leq t) < \epsilon$. By the Burkholder Davis-Gundy inequality

$$\mathbb{E}\left[(\tilde{X}(a)^{U_k})_t^*\right] \leq \mathbb{E}\left[(b-b(a))_{U_k\wedge t}^*\right] + \mathbb{E}\left[(N(a)^{U_k})_t^*\right] \leq k+a+C_1\mathbb{E}\left[\sqrt{[N(a)^{U_k}]_t}\right]$$
$$\leq k+a+C_2\mathbb{E}\left[(N(a))_{t\wedge U_k}^*\right] \leq k+a+C_2\left((N(a))_{(t\wedge U_k)-}^* + a\right) \leq (k+a)(C_2+1),$$

for some $C_1, C_2 \in \mathbb{R}^+$. Since $[\tilde{X}(a)^{T_k}]_t \to 0$ a.s. by the first part of the Lemma that we already proved, it now follows from the dominated convergence theorem that $\lim_{a\to 0^+}\mathbb{E}\left[(\tilde{X}(a)^{T_k})_t^*\right] = 0$ and then if we apply the Markov inequality we also see that for any $c > 0$, $\lim_{a\to 0^+}\mathbb{P}\left((\widehat{X}(a)-X)_{t\wedge T_k}^* \geq c\right) = 0$. We conclude that for any $c > 0$,

$$\lim_{a\to 0^+}\mathbb{P}\left((\widehat{X}(a)-X)_t^* \geq c\right) \leq \lim_{a\to 0^+}\mathbb{P}\left((\widehat{X}(a)-X)_{t\wedge T_k}^* \geq c\right) + \mathbb{P}(T_k \leq t)$$
$$= \mathbb{P}(T_k \leq t) < \epsilon,$$

hence $\widehat{X}(a) \xrightarrow{ucp} X$. □

A.1.4 Proof of Lemma 4.5

Proof. The equivalence of 1) and 2) follows trivially from the hypothesis and the linearity of ucp convergence. We will now establish that 2) and 3) are equivalent. Let $\epsilon > 0$, $c > 0$ be arbitrary. Let N denote the number of jumps of X with modulus larger or equal to a on $[0,t]$ for X and denote $A_k = \{N = k\}$ for $k \geq 0$ by the cadlag property of X and continuity of probability, $1 = \mathbb{P}(\bigcup_{k=0}^{\infty} A_k) = \lim_{K\to\infty}\mathbb{P}\left(\bigcup_{k=0}^{K} A_k\right)$. We may now choose K_0 such that $\mathbb{P}\left(\bigcup_{k=0}^{K_0} A_k\right) > 1-\epsilon$. Moreover if we let

$$A_t(a,\delta) = \{\exists s \leq t : |\Delta X_s| \in [a, a+\delta)\}$$

then $\bigcap_{n\geq 1} A_t(a,\frac{1}{n}) = \{\exists s \leq t : |\Delta X_s| = a\}$ this event has probability zero by definition and so by continuity of probability we see that we may choose $\delta > 0$ such that for a given $a \notin \mathbb{A}_X$, $\mathbb{P}(A_t(a,\delta)) < \epsilon/K_0$. Take such a δ and assume we are given $a \notin \mathbb{A}_X$. Define analogously to J_u, $J^n_u = \sum_{s\leq u} \Delta X^n_s$. We will now show that $X^n \xrightarrow{ucp} X$ implies $J^n(a) \xrightarrow{ucp} J(a)$ when $a \notin \mathbb{A}_X$. Let N_n denote the number of jumps with modulus larger or equal to a for X^n and denote $\{T^n_l\}_{l=1}^{N_n}$ the jump times of $X^n(a)$, let $\{T_l\}_{l=1}^k$ denote the jump times of $X(a)$ on A_k. $X^n \xrightarrow{ucp} X$ implies that there exists n_1 such that if $n \geq n_1$ then $\mathbb{P}\left(A_k \setminus \left\{(X^n - X)^*_t \geq \min\left(\frac{c}{2K_0}, \delta\right)\right\}\right) < \epsilon/K_0$. Let

$$B_{(k,n)} = A_k \cap \left\{(X^n - X)^*_t < \min\left(\frac{c}{2K_0}, \delta\right)\right\} \cap \{|\Delta X_s| \notin [a, a+\delta] \quad \forall s \leq t\},$$

then $\mathbb{P}\left(A_k \setminus B_{(k,n)}\right) < \frac{2\epsilon}{K_0}$. It follows that if X^n has a jump of size greater or equal to a in $[0,t]$ on $B_{(k,n)}$ then so must X and vice versa since otherwise $|\Delta(X^n - X)| \geq \delta$ contradicting the definition of $B_{(k,n)}$. i.e. the jump times of $J^n(a)$ and $J(a)$ coincide on $B_{(k,n)}$. Therefore we may write

$$J(a)_t - J^n(a)_t = \sum_{l=1}^{k} \Delta(X^n - X)_{T_l},$$

on $B_{(k,n)}$. We have for $1 \leq k \leq K_0$ on $B_{(k,n)}$

$$\begin{aligned}(J(a) - J^n(a))^*_t &\leq \sum_{l=1}^{k} |\Delta(X^n - X)|_{T_i} \leq \sum_{l=1}^{k} \left((X^n - X)^*_{T_i} + (X^n - X)^*_{T_i-}\right) \\ &\leq 2K_0 (X^n - X)^*_t < c,\end{aligned}$$

and obviously for $k = 0$, $J(a) - J^n(a) = 0$. So

$$\mathbb{P}\left(A_k \cap \{(J(a) - J^n(a))^*_t \geq c\}\right) \leq \mathbb{P}\left(B^c_{(k,n)} \cap A_k \cap \{(J(a) - J^n(a))^*_t \geq c\}\right) < \frac{\epsilon}{K_0}.$$

This leads to

$$\begin{aligned}\mathbb{P}\left((J(a) - J^n(a))^*_t \geq c\right) &\leq \sum_{k=1}^{K_0} \mathbb{P}\left(A_k \cap \{(J(a) - J^n(a))^*_t \geq c\}\right) + \mathbb{P}\left(\bigcup_{k>K_0} A_k\right) \\ &\leq K_0 \frac{\epsilon}{K_0} + \epsilon = 2\epsilon.\end{aligned}$$

This shows that if $a \notin \mathbb{A}_X$ then $J^n(a) \xrightarrow{ucp} J(a)$ and if we combine this with 2) (which is equivalent to 1)) we see that this implies $X^n(a) \xrightarrow{ucp} X(a)$ and it also trivially implies $\lim_{a\to 0^+} \lim_{n\to\infty} \mathbb{P}\left((J^n(a) - J(a))^*_t \geq c\right) = 0$ for all $c > 0$. Assume that 3) is true. Take any $\epsilon > 0$ and take $a > 0$ so small that $\lim_{n\to\infty} \mathbb{P}\left(\left(\sum_{s\leq \cdot} \Delta X^n_s I_{|\Delta X^n_s| \leq a}\right)^*_t \geq c\right) < \epsilon$ as well as $\mathbb{P}\left(\left(\sum_{s\leq \cdot} \Delta X_s I_{|\Delta X_s| \leq a}\right)^*_t \geq c\right) < \epsilon$, which we may since J is of finite variation. Therefore

$$\lim_{n\to\infty} \mathbb{P}\left((J - J^n)^*_t \geq c\right) \leq \lim_{n\to\infty} \mathbb{P}\left((J(a) - J^n(a))^*_t \geq c\right) +$$

$$\lim_{n\to\infty}\mathbb{P}\left(\left(\sum_{s\le\cdot}\Delta X_s I_{|\Delta X_s|\le a}\right)_t^*\ge c\right)\lim_{n\to\infty}\mathbb{P}\left(\left(\sum_{s\le\cdot}\Delta X_s^n I_{|\Delta X_s^n|\le a}\right)_t^*\ge c\right)<2\epsilon,$$

which is zero if $a\notin\mathbb{A}_X$ and this shows that 2) holds true. If we instead assume 2) is true then we have that

$$\lim_{n\to\infty}\mathbb{P}\left(((J^n-J^n(a))-(J-J(a)))_t^*\ge c\right)\le\lim_{n\to\infty}\mathbb{P}\left((J^n-J)_t^*\ge c\right)+$$
$$\lim_{n\to\infty}\mathbb{P}\left((J^n(a)-J(a))_t^*\ge c\right)=\lim_{n\to\infty}\mathbb{P}\left((J^n(a)-J(a))_t^*\ge c\right).$$

Letting $a\to 0^+$ eliminates the remaining term i.e.

$$\lim_{a\to 0^+}\lim_{n\to\infty}\mathbb{P}\left(\left(\sum_{s\le\cdot}\Delta X_s^n I_{|\Delta X_s^n|\le a}-\sum_{s\le\cdot}\Delta X_s I_{|\Delta X_s|\le a}\right)_t^*\ge c\right)=0,$$

which implies

$$\lim_{a\to 0^+}\lim_{n\to\infty}\mathbb{P}\left(\left(\sum_{s\le\cdot}\Delta X_s^n I_{|\Delta X_s^n|\le a}\right)_t^*\ge c\right)\le\lim_{a\to 0^+}\lim_{n\to\infty}\mathbb{P}\left(\left(\sum_{s\le\cdot}\Delta X_s I_{|\Delta X_s|\le a}\right)_t^*\ge c\right)$$
$$+\mathbb{P}\left(\left(\sum_{s\le\cdot}\Delta X_s^n I_{|\Delta X_s^n|\le a}-\sum_{s\le\cdot}\Delta X_s I_{|\Delta X_s|\le a}\right)_t^*\ge c\right)=0,$$

since X has jumps of finite variation. This shows 3). Suppose now that 4) is true. Take $a'\in\mathbb{A}_X$ so small that $a\le a'$ implies $\lim_{n\to\infty}\mathbb{P}\left(\sum_{s\le t}|\Delta X_s^n|I_{|\Delta X_s^n|\le a}\ge c\right)<\epsilon$ as well as $\lim_{n\to\infty}\mathbb{P}\left(\sum_{s\le t}|\Delta X_s|I_{|\Delta X_s|\le a}\ge c\right)<\epsilon$ then

$$\mathbb{P}\left((J^n(a)-J(a))_t^*\ge c\right)\le\mathbb{P}\left((J^n(a)-J^n(a'))_t^*\ge c\right)+\mathbb{P}\left((J^n(a')-J(a'))_t^*\ge c\right)+$$
$$\mathbb{P}\left((J(a')-J(a))_t^*\ge c\right)\le\mathbb{P}\left(\sum_{s\le t}|\Delta X_s^n|I_{|\Delta X_s^n|\le a'}\ge c\right)+\mathbb{P}\left((J^n(a')-J(a'))_t^*\ge c\right)+$$
$$\mathbb{P}\left(\sum_{s\le t}|\Delta X_s|I_{|\Delta X_s|\le a'}\ge c\right)\le 2\epsilon+\mathbb{P}\left((J^n(a')-J(a'))_t^*\ge c\right),$$

since $a'\notin\mathbb{A}_X$ the second term goes to zero as $n\to\infty$. This shows that $\lim_{a\to\infty}\lim_{n\to\infty}\mathbb{P}\left((J^n(a)-J(a))_t^*\ge c\right)=0$ but since 4) is stronger than 3) it implies 2) and combined with this we have the final statement. □

A.1.5 Proof of Theorem 3.5

Proof. By Lemma 2.2 $f_n(X^n) - f(X), f_n(X) - f(X)$ and $f_n(X^n) - f_n(X)$ all have quadratic variations along $\{D_k^n\}_k$, so by Lemma 2.5 we have that

$$[f_n(X^n) - f(X)]_t \leq 2[f_n(X) - f(X)]_t + 2[f_n(X^n) - f_n(X)]_t,$$

so therefore it will suffice to show,
1) $[f_n(X) - f(X)]_u \xrightarrow{\mathbb{P}} 0$,
and then
2) $[f_n(X^n) - f_n(X)]_u \xrightarrow{\mathbb{P}} 0$. We proceed with 1). Let $h(x) \in L_{1,loc}([-R, R])$ be such that $|f_n'| \leq |h|$ a.e. for all n (and therefore $|f'| \leq |h|$ a.e.) as guaranteed by Assumption 3.1. Let $t \geq 0$, $c > 0$ and $\epsilon > 0$ be arbitrary. Since X is cadlag we may choose $R > 0$ such that $\mathbb{P}(\max(X_u^*, [X]_u) > R) < \epsilon$. Since $|f_n'(x) - f'(x)| \leq 2|h(x)|$ a.e. it follows from the dominated convergence theorem that there exists some n' such that if $n \geq n'$ then $\int_{-R}^{R} |f_n'(x) - f'(x)| dx < \sqrt{\frac{c}{2(R+1)}}$.

Finally, letting $\{D_k\}$ be a refining sequence along which X has a quadratic variation we take k so large that

$$\mathbb{P}\left(\left([X] - \sum_{t_i \leq \cdot,\ t_i \in D_k} \left(X_{t_{i+1}} - X_{t_i}\right)^2\right)_u^* > 1\right) < \epsilon,$$

$$\mathbb{P}\left(\left([f(X)] - \sum_{t_i \leq \cdot,\ t_i \in D_k} \left(f(X)_{t_{i+1}} - f(X)_{t_i}\right)^2\right)_u^* > \frac{c}{2}\right) < \epsilon$$

and

$$\mathbb{P}\left(\left([f_n(X)] - \sum_{t_i \leq \cdot,\ t_i \in D_k} \left(f_n(X)_{t_{i+1}} - f_n(X)_{t_i}\right)_s^2\right)_u^* > \frac{c}{2}\right) < \epsilon,$$

Next we define

$$A = \{\max(X_t^*, [X]_t) < R\} \cap \left\{\left([X] - \sum_{t_i \leq \cdot,\ t_i \in D_k} \left(X_{t_{i+1}} - X_{t_i}\right)^2\right)_u^* < 1\right\},$$

then $\mathbb{P}(A^c) < 2\epsilon$. We have

$\mathbb{P}([f_n(X) - f(X)]_u > c) \leq$

$$\mathbb{P}\left(\left([f_n(X) - f(X)] - \sum_{t_i \leq \cdot,\ t_i \in D_k} \left(\left(f_n(X)_{t_{i+1}} - f(X)_{t_{i+1}}\right) - \left(f_n(X)_{t_i} - f(X)_{t_i}\right)\right)^2\right)_u^* > \frac{c}{2}\right)$$

$$+\mathbb{P}\left(\left(\sum_{t_i\le \cdot,\ t_i\in D_k}\left(\left(f_n(X)_{t_{i+1}}-f(X)_{t_{i+1}}\right)-\left(f_n(X)_{t_i}-f(X)_{t_i}\right)\right)^2\right)^*_u>\frac{c}{2}\right)\le$$

$$\mathbb{P}\left(\left(\sum_{t_i\le \cdot,\ t_i\in D_k}\left(\left(f_n(X)_{t_{i+1}}-f(X)_{t_{i+1}}\right)-\left(f_n(X)_{t_i}-f(X)_{t_i}\right)\right)^2\right)^*_u>\frac{c}{2}\right)+\epsilon\le$$

$$\mathbb{P}\left(\left\{\left(\sum_{t_i\le \cdot,\ t_i\in D_k}\left(\left(f_n(X)_{t_{i+1}}-f(X)_{t_{i+1}}\right)-\left(f_n(X)_{t_i}-f(X)_{t_i}\right)\right)^2\right)^*_u>\frac{c}{2}\right\}\cap A\right)$$

$$+\mathbb{P}(A^c)+\epsilon.$$

We shall see that

$$B:=\left\{\left(\sum_{t_i\le \cdot,\ t_i\in D_k}\left(\left(f_n(X)_{t_{i+1}}-f(X)_{t_{i+1}}\right)-\left(f_n(X)_{t_i}-f(X)_{t_i}\right)\right)^2\right)^*_u>\frac{c}{2}\right\}$$

is disjoint from A. By applying the same type of representation as used in the proof of Lemma 2.2, we find that on A when $s\le u$,

$$\sum_{t_i\le s,\ t_i\in D_k}\left(\left(f_n(X)_{t_{i+1}}-f(X)_{t_{i+1}}\right)-\left(f_n(X)_{t_i}-f(X)_{t_i}\right)\right)^2_s=$$

$$\sum_{t_i\le s,\ t_i\in D_k}\left(\Delta_i\int_0^1\left(f'(X_{s\wedge t_i}+\theta\Delta_i)-f_n'(X_{s\wedge t_i}+\theta\Delta_i)\right)d\theta\right)^2\le$$

$$\sum_{t_i\le s,\ t_i\in D_k}\Delta_i^2\left(\int_0^1\left|f'(X_{s\wedge t_i}+\theta\Delta_i)-f_n'(X_{s\wedge t_i}+\theta\Delta_i)\right|d\theta\right)^2\le$$

$$\left(\int_{-R}^{R}\left|f'(\theta)-f_n'(\theta)\right|d\theta\right)^2\sum_{t_i\le s,\ t_i\in D_k}\Delta_i^2\le\frac{c}{2(R+1)}\left([X]_s+1\right)\le\frac{c}{2(R+1)}\left([X]_u+1\right)<\frac{c}{2},$$

which establishes that A and B are disjoint (hence $\mathbb{P}(A\cap B)=0$) so we conclude that $\mathbb{P}\left([f_n(X)-f(X)]_t>c\right)\le 6\epsilon$, since $\epsilon>0$ and $c>0$ are arbitrary it follows that $[f_n(X)-f(X)]_u\xrightarrow{\mathbb{P}}0$.
We now prove 2). Since $[X^n-X]_u\xrightarrow{\mathbb{P}}0$ implies that $[X^n]_u\xrightarrow{\mathbb{P}}[X]_u$ and since $X^n\xrightarrow{ucp}X$ there exists R' such that

$$\mathbb{P}\left(\max\left([X^n]_u,[X]_u,(X^n)^*_u,X^*_u\right)\ge R'\right)<\epsilon,$$

for all n. Now we take n_1 so large that if $n\ge n_1$ then $\int_{[-R,R]}|f_n'(x)-f'(x)|dx<\sqrt{\frac{c}{24(R'+1)}}$. And take $\delta>0$ so small that if $x,y\in[-R-a,R+a]$ and $|x-y|\le\delta$ then $|f'(x)-f'(y)|<\sqrt{\frac{c}{24(R'+1)}}$.

Now we may take n_2 so large that if $n\ge n_2$ then $\mathbb{P}\left([X^n-X]_u\ge\frac{c}{24\left(\int_{[-R,R]}|h(x)|dx\right)^2}\right)<\epsilon$. Given any $n\ge\max(n_1,n_2)$ we may, since X, X^n-X and $f_n(X^n)-f_n(X)$ all have quadratic variations along $\{D^n_k\}_k$, choose a partition $D^n_k = \{t^n_i\}_{i\geq 1}$ (which depends on n) such that

$$\mathbb{P}\left(\left(\sum_{t_i\leq \cdot,\ t_i\in D^n_k}\left(\left(X^n_{t^n_{i+1}} - X_{t^n_{i+1}}\right) - \left(X^n_{t^n_i} - X_{t^n_i}\right)\right)^2 - [X^n - X]\right)^*_u < \frac{c}{24\left(\int_{[-R,R]}|h(x)|dx\right)^2}\right)$$
$$< \epsilon,$$

$$\mathbb{P}\left(\left(\sum_{t_i\leq \cdot,\ t_i\in D^n_k}\left(X_{t^n_{i+1}} - X_{t^n_i}\right)^2 - [X]\right)^*_u < 1\right) < \epsilon, \mathbb{P}\left((X^n - X)^*_u < \delta/3\right) < \epsilon$$

and

$$\mathbb{P}\left(\left([f_n(X^n) - f_n(X)] - \sum_{t_i\leq \cdot,\ t_i\in D^n_k}\left(\left(f_n(X^n)_{t^n_{i+1}} - f_n(X)_{t^n_{i+1}}\right) - \left(f_n(X^n)_{t^n_i} - f_n(X)_{t^n_i}\right)\right)^2\right)^*_u\right.$$
$$\left. > \frac{c}{2}\right) < \epsilon.$$

Let

$$A(n) = \left\{\max\left([X^n]_u, [X]_u, (X^n)^*_u, X^*_u\right) < R'\right\} \cap \left\{[X^n - X]_u < \frac{c}{24\left(\int_{[-R,R]}|h(x)|dx\right)^2}\right\} \cap$$
$$\left\{\left(\sum_{t_i\leq \cdot,\ t_i\in D^n_k}\left(\left(X^n_{t^n_{i+1}} - X_{t^n_{i+1}}\right) - \left(X^n_{t^n_i} - X_{t^n_i}\right)\right)^2 - [X^n - X]\right)^*_u < \frac{c}{24\left(\int_{[-R,R]}|h(x)|dx\right)^2}\right\}$$
$$\cap \left\{\left([X] - \sum_{t_i\leq \cdot,\ t_i\in D^n_k}\left(X_{t^n_{i+1}} - X_{t^n_i}\right)^2\right)^*_u < 1\right\} \cap \{(X^n - X)^*_u < \delta/3\},$$

We now let $N = \max(n_1, n_2, n_2)$ then $\mathbb{P}(A(n)^c) < 7\epsilon$ for $n \geq N$.

So, for $n \geq N$,

$$\mathbb{P}\left([f_n(X^n) - f_n(X)]_u > c\right) \leq$$
$$\mathbb{P}\left(\left([f_n(X^n) - f_n(X)] - \sum_{t_i\leq \cdot,\ t_i\in D^n_k}\left(\left(f_n(X^n)_{t^n_{i+1}} - f_n(X)_{t^n_{i+1}}\right) - \left(f_n(X^n)_{t^n_i} - f_n(X)_{t^n_i}\right)\right)^2\right)^*_u\right.$$
$$\left. > \frac{c}{2}\right) + \mathbb{P}\left(\left(\sum_{t_i\leq \cdot,\ t_i\in D^n_k}\left(\left(f_n(X^n)_{t^n_{i+1}} - f_n(X)_{t^n_{i+1}}\right) - \left(f_n(X^n)_{t^n_i} - f_n(X)_{t^n_i}\right)\right)^2\right)^*_u > \frac{c}{2}\right)$$

$$\leq \epsilon + \mathbb{P}\left(\left(\sum_{t_i \leq \cdot,\ t_i \in D_k^n} \left(\left(f_n(X^n)_{t_{i+1}^n} - f_n(X)_{t_{i+1}^n}\right) - \left(f_n(X^n)_{t_i^n} - f_n(X)_{t_i^n}\right)\right)^2\right)_u^* > \frac{c}{2}\right) \leq$$

$$\mathbb{P}\left(\left\{\left(\sum_{t_i \leq \cdot,\ t_i \in D_k^n} \left(\left(f_n(X^n)_{t_{i+1}^n} - f_n(X^n)_{t_i^n}\right) - \left(f_n(X)_{t_{i+1}^n} - f_n(X)_{t_i^n}\right)\right)^2\right)_u^* > \frac{c}{2}\right\} \cap A(n)\right)$$

$$+ \mathbb{P}(A(n)^c) + \epsilon.$$

We shall see that

$$B' := \left\{\left(\sum_{t_i \leq \cdot,\ t_i \in D_k^n} \left(\left(f_n(X^n)_{t_{i+1}^n} - f_n(X^n)_{t_i^n}\right) - \left(f_n(X)_{t_{i+1}^n} - f_n(X)_{t_i^n}\right)\right)^2\right)_u^* > \frac{c}{2}\right\}$$

is disjoint from $A(n)$. Using the same trick as in 1) we find that,
$(f_n(X^n)_{t_{i+1}^n} - f_n(X^n)_{t_i^n})_s = \Delta_i^n \int_0^1 f_n'(X_{t_i^n}^n + \theta\Delta_i^n)d\theta$ and similarly
$(f_n(X)_{t_{i+1}^n} - f_n(X)_{t_i^n})_s = \Delta_i \int_0^1 f_n'(X_{t_i^n} + \theta\Delta_i)d\theta$ where $\Delta_i^n = X_{t_{i+1}^n}^n - X_{t_i^n}^n$ and $\Delta_i = X_{t_{i+1}^n} - X_{t_i^n}$. This allows us to rewrite

$$\begin{aligned}
&\left(f_n(X^n)_{t_{i+1}^n} - f_n(X^n)_{t_i^n}\right) - \left(f_n(X)_{t_{i+1}^n} - f_n(X)_{t_i^n}\right) = \\
&\Delta_i^n \int_0^1 f_n'(X_{t_i^n}^n + \theta\Delta_i^n)d\theta - \Delta_i \int_0^1 f_n'(X_{t_i^n} + \theta\Delta_i)d\theta \\
&= \int_0^1 f_n'(X_{t_i^n}^n + \theta\Delta_i^n)d\theta\,(\Delta_i^n - \Delta_i) + \int_0^1 \left(f_n'(X_{t_i^n}^n + \theta\Delta_i^n) - f_n'(X_{t_i^n} + \theta\Delta_i)\right) d\theta\Delta_i
\end{aligned}$$

and furthermore on $A(n)$,

$$\begin{aligned}
&\left|\int_0^1 \left(f_n'(X_{t_i^n}^n + \theta\Delta_i^n) - f_n'(X_{t_i^n} + \theta\Delta_i)\right) d\theta\Delta_i\right| \leq \int_0^1 \left|f_n'(X_{t_i^n}^n + \theta\Delta_i^n) - f'(X_{t_i^n}^n + \theta\Delta_i^n)\right| d\theta|\Delta_i| \\
&+ \int_0^1 \left|f'(X_{t_i^n} + \theta\Delta_i) - f_n'(X_{t_i^n} + \theta\Delta_i)\right| d\theta|\Delta_i| + \int_0^1 \left|f'(X_{t_i^n}^n + \theta\Delta_i^n) - f'(X_{t_i^n} + \theta\Delta_i)\right| d\theta|\Delta_i| \\
&\leq \left(2\int_{-R}^R |f_n'(x) - f'(x)|\, dx|\Delta_i| + \sqrt{\frac{c}{24(R'+1)}}\right)|\Delta_i|
\end{aligned}$$

where we used the fact that

$$|X_{t_i^n}^n + \theta\Delta_i^n - X_{t_i^n} + \theta\Delta_i| \leq (X^n - X)_t^* + \theta|\Delta_i^n - \Delta_i| \leq \delta + 2\theta(X^n - X)_t^* \leq 3\delta$$

Plugging in the above representations expanding the squares and then using the Cauchy-Schwartz inequality (for sums) on the cross terms we see that on $A(n)$ for $s \leq u$,

$$\sum_{t_i \leq s,\ t_i \in D_k^n} \left(\left(f_n(X^n)_{t_{i+1}^n} - f_n(X^n)_{t_i^n}\right) - \left(f_n(X)_{t_{i+1}^n} - f_n(X)_{t_i^n}\right)\right)^2$$

$$
\begin{aligned}
&\le \sum_{t_i\le s,\ t_i\in D_k^n} \left(\int_0^1 f_n'(X^n_{t_i^n}+\theta\Delta_i^n)d\theta\right)^2 (\Delta_i^n-\Delta_i)^2\\
&+\sum_{t_i\le s,\ t_i\in D_k^n} \left(\int_0^1 \left(f_n'(X^n_{t_i^n}+\theta\Delta_i^n)-f_n'(X_{t_i^n}+\theta\Delta_i)\right)d\theta)\right)^2 (\Delta_i)^2\\
&+2\left(\sum_{t_i\le s,\ t_i\in D_k^n} \left(\int_0^1 f_n'(X^n_{t_i^n}+\theta\Delta_i^n)d\theta\right)^2 (\Delta_i^n-\Delta_i)^2\right)^{\frac{1}{2}}\\
&\times\left(\sum_{t_i\le s,\ t_i\in D_k^n} \left(\int_0^1 \left(f_n'(X^n_{t_i^n}+\theta\Delta_i^n)-f_n'(X_{t_i^n}+\theta\Delta_i)\right)d\theta)\right)^2 (\Delta_i)^2\right)^{\frac{1}{2}}\le\\
&3\sum_{t_i\le s,\ t_i\in D_k^n} \left(\int_0^1 f_n'(X^n_{t_i^n}+\theta\Delta_i^n)d\theta\right)^2 (\Delta_i^n-\Delta_i)^2+\\
&3\sum_{t_i\le s,\ t_i\in D_k^n} \left(\int_0^1 \left(f_n'(X^n_{t_i^n}+\theta\Delta_i^n)-f_n'(X_{t_i^n}+\theta\Delta_i)\right)d\theta)\right)^2 (\Delta_i)^2\\
&\le 3\left(\int_{[-R,R]}|h(x)|dx\right)^2 \sum_{t_i\le s,\ t_i\in D_k^n}\left(\left(X^n_{t^n_{i+1}}-X_{t^n_{i+1}}\right)-\left(X^n_{t_i^n}-X_{t_i^n}\right)\right)^2\\
&+3\left(2\int_{-R}^{R}\left|f_n'(x)-f'(x)\right|dx|+\sqrt{\frac{c}{24(R'+1)}}\right)^2\sum_{t_i,\ t_i\in D_k^n}\left(X_{t^n_{i+1}}-X_{t_i^n}\right)^2\\
&\le 3\left(\int_{[-R,R]}|h(x)|dx\right)^2\left([X^n-X]_s+\frac{c}{24\left(\int_{[-R,R]}|h(x)|dx\right)^2}\right)+3\frac{c}{12(R'+1)}([X]_s+1)\\
&<3\frac{2c}{24}+3\frac{c}{12(R'+1)}(R'+1)=\frac{c}{2},
\end{aligned}
$$

where we applied the AM-GM inequality to the cross term after the second inequality. We therefore conclude that $A(n)$ and B' are disjoint and therefore we conclude that $[f_n(X^n)-f_n(X)]_u \xrightarrow{\mathbb{P}} 0$. □

A.1.6 Proof of Theorem 3.9

Proof. We will first show the Theorem for a) with $p=\frac{1}{2}$ and b) with $q=1$, at the end of this proof we outline how to carry out the other cases which are very much analogous. Let $\epsilon>0$ be arbitrary. Let $\epsilon>0$ be arbitrary. Let $a>1$ be some (large) constant to be chosen later. Let $X=Z+C$ where Z is a semimartingale and C has zero quadratic variation. According to Theorem 2.10 we have that $f(X_s)=Y_s^a+\Gamma_s^a$ where Y^a is a semimartingale, Γ^a is continuous and $[\Gamma^a]_t=0$ for all $t>0$. The expression for Y^a is given by

$$Y_t = f(X_0) + \sum_{s\le t} \left(f(X_s) - f(X_{s-}) - \Delta X_s f(X_{s-})\right) I_{|\Delta X_s|>a} + \int_0^t f'(X_{s-})dZ_s$$
$$+ \int_0^t \int_{|x|\le a} \left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)(\mu-\nu)(ds,dx)$$
$$+ \sum_{s\le t} \int_{|x|\le a} \left(f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})\right)\nu(\{s\},dx). \qquad \text{(A.1.14)}$$

We have by the mean-value theorem that for some $\theta(s,\omega)\in[0,1]$

$$\sum_{s\le t} \int_{|x|\le a} |f(X_{s-}+x) - f(X_{s-}) - xf'(X_{s-})|\,\nu(\{s\},dx) =$$
$$\sum_{s\le t} \int_{|x|\le a} |xf'(X_{s-}+\theta(s,\omega)x) - xf'(X_{s-})|\,\nu(\{s\},dx) \le$$
$$\left(\sup_{u\le t}|f'(X_{u-})| + \sup_{u\le t, x\in[-a,a]} |f'(X_{u-}+x)|\right) \sum_{s\le t}\int_{|x|\le a} |x|\nu(\{s\},dx), \qquad \text{(A.1.15)}$$

the factor in the parenthesis is a.s. finite since $f\in C^1$ and X is cadlag, so the whole expression on the left-most side of (A.1.15) is a.s. finite if $\sum_{s\le t}\int_{|x|\le a}|x|\nu(\{s\},dx)$ is. If we assume be) then since $\sum_{s\le t}\int_{|x|\le a}|x|\mu_n(\{s\},dx) \xrightarrow{\mathbb{P}} \sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx)$ and $\left\{\left(\sum_{s\le t}\int_{|x|\le a}|x|\mu_n(\{s\},dx)\right)^2\right\}_n$ is u.i. it follows that

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\left(\sum_{s\le t}\int_{|x|\le a}|x|\mu_n(\{s\},dx)\right)^2 - \left(\sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx)\right)^2\right|\right] = 0,$$

so $\left(\sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx)\right)^2 \in L^1$ and this implies (by Jensen's inequality) that $\sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx)\in L^1$. Similarly if we instead assume a) then $\sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx)\in L^1$. We conclude our argument by noting that $\mathbb{E}\left[\int_{|x|\le a}|x|\nu(\{s\},dx)\right] = \mathbb{E}\left[\int_{|x|\le a}|x|\mu(\{s\},dx)\right]$ and by monotone convergence

$$\mathbb{E}\left[\sum_{s\le t}\int_{|x|\le a}|x|\nu(\{s\},dx)\right] = \sum_{s\le t}\mathbb{E}\left[\int_{|x|\le a}|x|\nu(\{s\},dx)\right] =$$
$$\sum_{s\le t}\mathbb{E}\left[\int_{|x|\le a}|x|\mu(\{s\},dx)\right] = \mathbb{E}\left[\sum_{s\le t}\int_{|x|\le a}|x|\mu(\{s\},dx)\right] < \infty,$$

so indeed, $\sum_{s\leq t}\int_{|x|\leq a}|x|\nu(\{s\},dx)<\infty$ a.s..

This allows us to to expand

$$\int_0^t\int_{|x|\leq a}\left(f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})\right)(\mu-\nu)(ds,dx)=$$
$$\int_0^t\int_{|x|\leq a}\left(f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})\right)(\tilde{\mu}-\nu_c)(ds,dx)+$$
$$\sum_{s\leq t}\int_{|x|\leq a}\left(f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})\right)(\mu-\nu)(\{s\},dx),$$

where $\tilde{\mu}$ denotes the jump measure μ with all fixed time jumps removed. We may now rewrite (A.1.14)

$$\begin{aligned}&=f(X_0)+\sum_{s\leq t}\left(f(X_s)-f(X_{s-})-\Delta X_s f(X_{s-})\right)I_{|\Delta X_s|>a}+\int_0^t f'(X_{s-})dZ_s\\&+\int_0^t\int_{|x|\leq a}\left(f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})\right)(\tilde{\mu}-\nu_c)(ds,dx)\\&+\sum_{s\leq t}\int_{|x|\leq a}\left(f(X_{s-}+x)-f(X_{s-})-xf'(X_{s-})\right)\mu(\{s\},dx).\end{aligned}$$

Similarly let $X^n=Z^n+C^n$ where Z^n is a semimartingale and C^n has zero quadratic variation. We again apply Theorem 2.10, we have that $f_n(X^n_s)=(Y^n)^a_s+(\Gamma^n)^a_s$ where $(Y^n)^a$ is a semimartingale, $(\Gamma^n)^a$ is continuous and $[(\Gamma^n)^a]_t=0$ for all $t>0$. Arguing as above we see that the expression for $(Y^n)^a$ will be given by

$$\begin{aligned}Y^n_t&=f_n(X^n_0)+\sum_{s\leq t}\left(f_n(X^n_s)-f_n(X^n_{s-})-\Delta X^n_s f(X^n_{s-})\right)I_{|\Delta X^n_s|>a}\\&+\int_0^t\int_{|x|\leq a}\left(f_n(X^n_{s-}+x)-f_n(X^n_{s-})-xf_n'(X^n_{s-})\right)(\tilde{\mu_n}-(\nu_n)_c)(ds,dx)\\&+\sum_{s\leq t}\int_{|x|\leq a}\left(f_n(X^n_{s-}+x)-f_n(X^n_{s-})-xf_n'(X^n_{s-})\right)\mu_n(\{s\},dx).\end{aligned}$$

Since

$$\begin{aligned}&f_n(X^n_{s-}+x)-f(X_{s-}+x)-f_n(X^n_{s-}))+f(X_{s-})-x\left(f_n'(X^n_{s-})-f'(X_{s-})\right)\\&=x\left(\int_0^1\left(f_n'(X^n_{s-}+\theta x)-f'(X_{s-}+\theta x)\right)d\theta-f_n'(X^n_{s-})+f'(X_{s-})\right)\end{aligned}$$

and the term in the parenthesis is clearly locally bounded we can conclude that

$$\int_0^{\cdot}\int_{|x|\leq a}\left(f_n(X^n_{s-}+x)-f(X_{s-}+x)-f_n(X^n_{s-}))+f(X_{s-})\right.$$

$$-x\left(f_n'(X_{s-}^n)-f'(X_{s-})\right)\right)(\tilde{\mu}-\nu_c)(ds,dx)$$

is well defined (as a local martingale). Furthermore, since X^n and X have quadratic variations along the same refining sequence then so does $f_n(X^n)$ and $f(X)$ by Lemma 2.2 which implies that $(\Gamma^n)^a-\Gamma^a$ has a quadratic variation along this sequence which is zero (the quadratic variation of the semimartingale parts of $f_n(X^n)$ and $f(X)$ do not depend on the refining sequence). With this in mind we make the following estimate,

$$[f_n(X^n)-f(X)]_t^{\frac{1}{2}}=[(Y^n)^a-Y^a]_t^{\frac{1}{2}}\leq\left[\int f'(X_-)d\left(Z-Z^n\right)\right]_t^{\frac{1}{2}}+\left[\int\left(f_n'(X_-^n)-f'(X_-)\right)dZ^n\right]_t^{\frac{1}{2}}$$
$$+\left[\int_0^{\cdot}\int_{|x|\leq a}\left(f_n(X_{s-}^n+x)-f(X_{s-}+x)-f_n(X_{s-}^n))+f(X_{s-})\right)-x\left(f_n'(X_{s-}^n)-f'(X_{s-})\right)\right)$$
$$(\tilde{\mu}-\nu_c)(ds,dx)\Bigg]_t^{\frac{1}{2}}+\left[\int_0^{\cdot}\int_{|x|\leq a}\left(\left(f_n(X_{s-}^n+x)-f_n(X_{s-}^n)\right)\right)-xf_n'(X_{s-}^n)\right)(\tilde{\mu}-\nu_c)(ds,dx)$$
$$-\int_0^{\cdot}\int_{|x|\leq a}\left(\left(f_n(X_{s-}^n+x)-f_n(X_{s-}^n)\right)\right)-xf_n'(X_{s-}^n)\right)(\tilde{\mu_n}-(\nu_n)_c)(ds,dx)\Bigg]_t^{\frac{1}{2}}+$$
$$\left[\sum_{s\leq t}\int_{|x|\leq a}\left(f_n(X_{s-}^n+x)-f(X_{s-}+x)-f_n(X_{s-}^n)+f(X_{s-})-x(f_n'(X_{s-}^n)-f'(X_{s-}))\right)\mu(\{s\},dx)\right]_t^{\frac{1}{2}}$$
$$+\left[\sum_{s\leq t}\int_{|x|\leq a}\left(f_n(X_{s-}^n+x)-f_n(X_{s-}^n)-xf_n'(X_{s-}^n)\right)(\mu-\mu_n)(\{s\},dx)\right]_t^{\frac{1}{2}}$$
$$+\Bigg[\sum_{s\leq t}\left(f_n(X_s^n)-f_n(X_{s-}^n)-\Delta X_s^n f_n'(X_{s-}^n)\right)I_{|\Delta X_s^n|>a}$$
$$-\sum_{s\leq t}\left(f(X_s)-f(X_{s-})-\Delta X_s f'(X_{s-})\right)I_{|\Delta X_s|>a}\Bigg]_t^{\frac{1}{2}} \tag{A.1.16}$$

Where we substituted the expressions for Y, Y^a, did a bit of rearrangement and then used Lemma 2.4.

Since $\tilde{\mu}$ is void of any fixed time jumps it follows from Theorem 1 of chapter 3, section 5 in [4], that if $g(s,x,\omega)$ is locally integrable then

$$\left[\int_{[0,.]}\int_{|x|\leq a}g(s,x,\omega)(\tilde{\mu}-\nu_c)(ds,dx)\right]_t=\int_{[0,t]}\int_{|x|\leq a}g(s,x,\omega)^2\tilde{\mu}(ds,dx),$$

(and similarly for $\tilde{\mu}_n$). We now expand (A.1.16) as

$$=\left(\int_{0+}^t f'(X_-)^2d\left[Z-Z^n\right]_s\right)^{\frac{1}{2}}+\left(\int_{0+}^t\left(f_n'(X_{s\_}^n)-f'(X_{s\_})\right)^2d[Z^n]_s\right)^{\frac{1}{2}}+$$
$$\left(\int_0^t\int_{|x|\leq a}\left((f_n(X_{s-}^n+x)-f(X_{s-}+x)+f_n(X_{s-}^n))-f(X_{s-})\right)-x(f_n'(X_{s-}^n)-f'(X_{s-}))\right)^2\tilde{\mu}(ds,dx)\right)^{\frac{1}{2}}$$

$$+\left[\int_0^{\cdot}\int_{|x|\le a}\left((f_n(X^n_{s-}+x)-f_n(X^n_{s-}))\right)-xf'_n(X^n_{s-})\right)((\tilde{\mu}-\nu_c)(ds,dx)$$

$$-\int_0^{\cdot}\int_{|x|\le a}\left((f_n(X^n_{s-}+x)-f_n(X^n_{s-}))\right)-xf'_n(X^n_{s-})\right)(\tilde{\mu}_n-(\nu_n)_c)(ds,dx)\Bigg]_t^{\frac{1}{2}}$$

$$+\left(\sum_{s\le t}\left(\int_{|x|\le a}\left(f_n(X^n_{s-}+x)-f(X_{s-}+x)-f_n(X^n_{s-})+f(X_{s-})-x(f'_n(X^n_{s-})-f'(X_{s-}))\right)\right.\right.$$

$$\left.\left.\mu(\{s\},dx)\right)^2\right)^{\frac{1}{2}}+\left(\sum_{s\le t}\left(\int_{|x|\le a}\left(f_n(X^n_{s-}+x)-f_n(X^n_{s-})-xf'_n(X_{s-})\right)(\mu-\mu_n)(\{s\},dx)\right)^2\right)^{\frac{1}{2}}$$

$$+\left[\sum_{s\le t}\left(f_n(X^n_s)-f_n(X^n_{s-})-\Delta X^n_s f'_n(X^n_{s-})\right)I_{|\Delta X^n_s|>a}\right.$$

$$\left.-\sum_{s\le t}\left(f(X_s)-f(X_{s-})-\Delta X_s f'(X_{s-})\right)I_{|\Delta X_s|>a}\right]_t^{\frac{1}{2}}, \tag{A.1.17}$$

here we used the fact that for term five and six of (A.1.16) (corresponding to term five and six of (A.1.17)), the expressions inside the quadratic variations are quadratic pure jump semimartingales so their contributions are just the square sums of their jumps.

We now collect some preliminary facts that will be used to tackle the terms of (A.1.17). Note that assuming a), $\mathbb{E}\left[\left(\int_0^t\int_{\mathbb{R}}x^2\mu(ds,dx)\right)^{\frac{1}{2}}\right]\le\mathbb{E}[[X]_t^{\frac{1}{2}}]<\infty$ and

$$\int_0^t\int_{|x|\le r}x^2\mu(ds,dx)=\sum_{s\le t}(\Delta X_s)^2 1_{|\Delta X_s|\le r}\xrightarrow{a.s.}0,$$

as $r\to 0^+$. So by dominated convergence it follows that $\lim_{r\to 0^+}\mathbb{E}\left[\left(\int_0^t\int_{|x|\le r}x^2\mu(ds,dx)\right)^{\frac{1}{2}}\right]=0$, while if we instead assume b) then $\lim_{r\to 0^+}\mathbb{E}\left[\left(\int_0^t\int_{|x|\le r}x^2\mu(ds,dx)\right)^{\frac{1}{2}}\right]=0$.

Since $[X^n]_t\le 2[X^n-X]_t+2[X]_t$ it follows that $|[X^n]_t-[X]_t|\le 2[X^n-X]_t+3[X]_t$ and since $[X^n]_t\xrightarrow{\mathbb{P}}[X]_t$ we may conclude by Pratt's lemma that if we assume b) then $[X^n]_t\xrightarrow{L^1}[X]_t$ and therefore both $\{[X^n]_t\}_n$ and $\{[X^n-X]_t\}_n$ are u.i., under this assumption. Similarly we have that $[X^n]_t^{\frac{1}{2}}\le[X^n-X]_t^{\frac{1}{2}}+[X]_t^{\frac{1}{2}}$ implying $|[X^n]_t^{\frac{1}{2}}-[X]_t^{\frac{1}{2}}|\le[X^n-X]_t^{\frac{1}{2}}+3[X]_t^{\frac{1}{2}}$ and so if we assume a), both $\{[X^n]_t^{\frac{1}{2}}\}_n$ and $\{[X^n-X]_t^{\frac{1}{2}}\}_n$ are u.i. in this case. By assumption a) $\left(\sum_{s\le t}\int_{\mathbb{R}}|x|\mu_n(\{s\},dx)\right)^{\frac{1}{2}}$ is u.i. and $\sum_{s\le t}\int_{\mathbb{R}}|x|\mu_n(\{s\},dx)\xrightarrow{\mathbb{P}}\sum_{s\le t}\int_{\mathbb{R}}|x|\mu(\{s\},dx)$ as $n\to\infty$, this implies $\left(\sum_{s\le t}\int_{\mathbb{R}}|x|\mu_n(\{s\},dx)\right)^{\frac{1}{2}}\xrightarrow{L^1}\left(\sum_{s\le t}\int_{\mathbb{R}}|x|\mu(\{s\},dx)\right)^{\frac{1}{2}}$. Similarly, $\sum_{s\le t}\int_{\mathbb{R}}|x|\mu_n(\{s\},dx)\xrightarrow{L^1}\sum_{s\le t}\int_{\mathbb{R}}|x|\mu(\{s\},dx)$ if we instead assume b).

Furthermore,

$$\lim_{r\to 0^+}\limsup_n \mathbb{P}\left(\sum_{s\le t}\int_{|x|\le r}|x|\mu_n(\{s\},dx)\ge c\right)\le$$

$$\lim_{r\to 0^+}\limsup_n\left(\mathbb{P}\left(\left|\sum_{s\le t}\int_{|x|\le r}|x|\mu_n(\{s\},dx)-\sum_{s\le t}\int_{|x|\le r}|x|\mu(\{s\},dx)\right|\ge c\right)\right.$$

$$\left.+\mathbb{P}\left(\sum_{s\le t}\int_{|x|\le r}|x|\mu(\{s\},dx)\ge c\right)\right)=\lim_{r\to 0^+}\mathbb{P}\left(\sum_{s\le t}\int_{|x|\le r}|x|\mu(\{s\},dx)\ge c\right)=0,$$

for every $c>0$.

It remains to show that all seven terms on the right-most side of (A.1.17) converge to zero in expectation under the assumed hypotheses. Our strategy will be the following. We will begin by showing that if we assume a), term seven of (A.1.17) can be made arbitrarily small in L^1 (or in L^2 if we instead assume b)) uniformly over n by choosing a large enough. Once we have shown that term seven converges to zero in L^1 (or L^2) as $a\to\infty$ we fix a so large that this term is smaller than say ϵ. After this we show that all other terms converge to zero as $n\to\infty$ which means that the right-hand side of (A.1.17) is less than ϵ when $n\to\infty$ and then we can finally let ϵ, which is arbitrary, go to zero. So for term seven, first note that $\mathbb{P}\left((\Delta X^n)^*_t>a\right)\le 2\mathbb{P}\left((X^n)^*_t>a/2\right)$, which can be made arbitrarily small for large enough a and therefore term seven converges to zero in probability as $a\to\infty$ for all n. so letting $a\to\infty$ shows that the above expression tends to zero in probability for all n as $a\to\infty$. Assuming a) we have by the mean value theorem that for some $\theta_1(s,\omega)$ and $\theta_2(s,\omega)$,

$$\left[\sum_{s\le t}\left(f_n(X^n_s)-f_n(X^n_{s-})-\Delta X^n_s f_n'(X^n_{s-})\right)I_{|\Delta X^n_s|>a}-\sum_{s\le t}\left(f(X_s)-f(X_{s-})-\Delta X_s f'(X_{s-})\right)I_{|\Delta X_s|>a}\right]^{\frac{1}{2}}$$

$$=\left[\sum_{s\le t}\Delta X^n_s\left(f_n'(X^n_{s-}+\theta_1(s)\Delta X^n_s)-f_n'(X^n_{s-})\right)I_{|\Delta X^n_s|>a}\right.$$

$$\left.-\sum_{s\le t}\Delta X_s\left(f'(X_{s-}+\theta_2(s)\Delta X_s)-f'(X_{s-})\right)I_{|\Delta X_s|>a}\right]^{\frac{1}{2}}\le$$

$$\left[\sum_{s\le t}\Delta X^n_s\left(f_n'(X^n_{s-}+\theta_1(s)\Delta X^n_s)-f_n'(X^n_{s-})\right)I_{|\Delta X^n_s|>a}\right]^{\frac{1}{2}}+$$

$$\left[\sum_{s\le t}\Delta X_s\left(f'(X_{s-}+\theta_2(s)\Delta X_s)-f'(X_{s-})\right)I_{|\Delta X_s|>a}\right]^{\frac{1}{2}}=$$

$$\left(\sum_{s\le t}(\Delta X^n_s)^2\left(f_n'(X^n_{s-}+\theta_1(s)\Delta X^n_s)-f_n'(X^n_{s-})\right)^2 I_{|\Delta X^n_s|>a}\right)^{\frac{1}{2}}+$$

$$\left(\sum_{s\le t}(\Delta X_s)^2\left(f'(X_{s-}+\theta_2(s)\Delta X_s)-f'(X_{s-})\right)^2 I_{|\Delta X_s|>a}\right)^{\frac{1}{2}}\le\left(4U^2\sum_{s\le t}(\Delta X^n_s)^2 I_{|\Delta X^n_s|>a}\right)^{\frac{1}{2}}+$$

$$\left(4U^2\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|>a}\right)^{\frac{1}{2}} \le 2U\left([X^n]_t^{\frac{1}{2}}1_{H_a(n)} + [X]_t^{\frac{1}{2}}1_{H_a}\right),$$

where $H_a(n) = \{\exists s \le t : |\Delta X_s^n| > a\}$ and $H_a = \{\exists s \le t : |\Delta X_s| > a\}$. We know that $\mathbb{P}(H_a(n))$ tends to zero uniformly over n as $a \to \infty$, as well as $\mathbb{P}(H_a) \to 0$ as $a \to \infty$. This implies (due to the u.i. property of $\{[X^n]_t^{\frac{1}{2}}\}_n$) that $\lim_{a\to\infty}\sup_n \mathbb{E}\left[[X^n]_t^{\frac{1}{2}}1_{H_a(n)}\right] = 0$, as well as $\lim_{a\to\infty}\mathbb{E}\left[[X]_t^{\frac{1}{2}}1_{H_a}\right] = 0$ which shows the L^1 convergence when assuming a). If we instead assume b) then by using the bounds $|X_{s-}^n + \theta_1(s)\Delta X_s^n| \le 3(X^n)_t^*$ and $|X_{s-} + \theta_2(s)\Delta X_s| \le 3X_t^*$ we find that,

$$\mathbb{E}\left[\left[\sum_{s\le t}\left(f_n(X_s^n) - f_n(X_{s-}^n) - \Delta X_s^n f_n'(X_{s-}^n)\right)I_{|\Delta X_s^n|>a}\right.\right.$$

$$\left.\left.-\sum_{s\le t}\left(f(X_s) - f(X_{s-}) - \Delta X_s f'(X_{s-})\right)I_{|\Delta X_s|>a}\right]^{\frac{1}{2}}\right] \le$$

$$\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X_s^n)^2\left(f_n'(X_{s-}^n + \theta_1(s)\Delta X_s^n) - f_n'(X_{s-}^n)\right)^2 I_{|\Delta X_s^n|>a}\right)^{\frac{1}{2}}\right] +$$

$$\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X_s)^2\left(f'(X_{s-} + \theta_2(s)\Delta X_s) - f'(X_{s-})\right)^2 I_{|\Delta X_s|>a}\right)^{\frac{1}{2}}\right] \le$$

$$\mathbb{E}\left[2(M\vee 2C(X^n)_t^*)\left(\sum_{s\le t}(\Delta X_s^n)^2 I_{|\Delta X_s^n|>a}\right)^{\frac{1}{2}}\right] + \mathbb{E}\left[2(M\vee 2CX_t^*)\left(\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|>a}\right)^{\frac{1}{2}}\right]$$

$$\le 2\mathbb{E}\left[[X^n]_t 1_{H_a(n)}\right]^{\frac{1}{2}}\mathbb{E}\left[(M\vee 2C(X^n)_t^*)^2 1_{H_a(n)}\right]^{\frac{1}{2}} + 2\mathbb{E}\left[[X]_t 1_{H_a}\right]^{\frac{1}{2}}\mathbb{E}\left[(M\vee 2CX_t^*)^2 1_{H_a}\right]^{\frac{1}{2}},$$

where we used the Cauchy-Schwarz inequality in the last step. By assumption b) and the same kind of argument we made when assuming a) this shows uniform convergence to zero as $a \to \infty$ uniformly over n.

Since $X^n \xrightarrow{ucp} X$ it follows that $\lim_{R\to\infty}\sup_n \mathbb{P}\left((X^n)_t^* \ge R\right) = 0$ hence we may choose R so large that $\mathbb{P}\left(A_R(n)\right) < \epsilon$ for all n while also $|f_n'(x)| \le 2C|x|$ for all n if $|x| \ge R$ (under assumption b)), where $A_R(n) = \{\max\left((X^n)_t^*, X_t^*\right) \le R\}$. Let us also define $M = \sup_n \sup_{x\in[-R-a,R+a]}|f_n'(x)|$. For the first term in (A.1.17), under assumption a)

$$\left(\int_{0^+}^t f'(X_-)^2 d\left[Z - Z^n\right]_s\right)^{\frac{1}{2}} \le U[Z^n - Z]_t^{\frac{1}{2}} = U[X^n - X]_t^{\frac{1}{2}},$$

which converges to zero in L^1 by assumption.

If we instead assume b)

$$\mathbb{E}\left[\sqrt{\int_{0+}^{t} f'(X_{-})^2 d\left[Z-Z^n\right]_s}\right] \leq \mathbb{E}\left[M \vee 2CX_t^*\sqrt{\int_{0+}^{t} d\left[X-X^n\right]_s}\right]$$
$$\leq \mathbb{E}\left[[X^n-X]_t\right]^{\frac{1}{2}} \mathbb{E}\left[2C(X_t^*)^2+M\right]^{\frac{1}{2}},$$

by the Cauchy-Schwartz inequality and this converges to zero by assumption.

For the second term in (A.1.17), fix $c>0$. Let M_1 be so large that $\mathbb{P}\left(X_t^* \geq M_1\right) < \epsilon$. Since $[X-X^n]_t \xrightarrow{\mathbb{P}} 0$ as $n\to\infty$ and $[X^n]_t \leq \left([X^n-X]_t^{\frac{1}{2}}+[X]_t^{\frac{1}{2}}\right)^2$ it follows that we may take M_2 so large that $\mathbb{P}\left([X^n]_t \geq M_2\right) < \varepsilon$ for $n \geq n_2$. Let $\delta>0$ be so small that if $x,y \in [-M_1-a, M_1+a]$ and $|x-y|<\delta$ then $|f(x)-f(y)| < \frac{\sqrt{c}}{\sqrt{M_2}}$. Take $n_1>0$ so large that $\mathbb{P}\left((X^n-X)_t^* \geq \delta\right) < \epsilon$ for $n \geq n_1$. Define $E_2(n) = \{(X^n-X)_t^* \leq \delta\} \cap \{[X^n]_t \leq M_2\}$ then $(f'(X^n)-f'(X))_t^* < \frac{\epsilon}{\sqrt{M_2}}$ on $E_2(n)$ and $\mathbb{P}(E_2(n)^c) < 2\varepsilon$ if $n \geq n' := \max(n_1,n_2)$ and on $E_2(n)$,

$$\int_{0+}^{t}\left(f_n'(X_{s_-}^n)-f'(X_{s_-})\right)^2 d[Z^n]_s < \frac{c}{M_2}[Z^n]_t = \frac{c}{M_2}[X^n]_t \leq c.$$

So for $n \geq n'$

$$\mathbb{P}\left(\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[Z^n]_s \geq c\right) \leq$$
$$\mathbb{P}\left(\left\{\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[Z^n]_s \geq c\right\}\cap E_2(n)\right)+\mathbb{P}(E_2(n)^c) = \mathbb{P}(E_2(n)^c) \leq 2\varepsilon,$$

thus $\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[Z^n]_s \xrightarrow{\mathbb{P}} 0$. To show convergence in expectation we notice similarly to the first term that under assumption a) we have

$$\left(\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[Z^n]_s\right)^{\frac{1}{2}} \leq 2U[Z^n]_t^{\frac{1}{2}} = 2U[X^n]_t^{\frac{1}{2}} \in L^1$$

and by dominated convergence this implies $\mathbb{E}\left[\left(\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[Z^n]_s\right)^{\frac{1}{2}}\right] \to 0$. If we instead assume b) then note that, by letting $A_R(n)$ and M be as above

$$\mathbb{E}\left[\sqrt{\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[Z^n]_s}\right] =$$
$$\mathbb{E}\left[\sqrt{\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[X^n]_s 1_{A_R(n)}}\right]+\mathbb{E}\left[\sqrt{\int_{0+}^{t}\left(f_n(X_{s_-}^n)-f(X_{s_-})\right)^2 d[X^n]_s 1_{A_R(n)^c}}\right]$$

$$\leq M\mathbb{E}\left[[X^n]_t^{\frac{1}{2}}\right] + 2C\mathbb{E}\left[(X_t^* + (X^n)_t^*)\sqrt{[X^n - X]_t}\right]$$

$$\leq M\mathbb{E}\left[[X^n - X]_t^{\frac{1}{2}}\right] + 2C\sqrt{\mathbb{E}\left[[X^n - X]_t\right]}\sqrt{\mathbb{E}\left[2(X_t^*)^2 + 2((X^n)_t^*)^2\right]}$$

which converges to zero since $\mathbb{E}\left[2(X_t^*)^2 + 2((X^n)_t^*)^2\right]$ is uniformly bounded in n by the u.i. assumption (note that $((X^n)_t^*)^2 \xrightarrow{ucp} (X_t^*)^2$ implies $\mathbb{E}\left[(X_t^*)^2\right] < \infty$ by the u.i. property as well).

The third term of (A.1.17) will be split into two terms, we will show that the first one will vanish and the second terms is completely analogous. Given any $\epsilon > 0$ and $c > 0$, take $r > 0$ so small that $\mathbb{P}\left(\left\{\int_0^t \int_{|x|\leq r} |x|^2 \tilde{\mu}(ds,dx) \geq \frac{c^2}{48M}\right\} \cap A_R(n)\right) < \epsilon$ and take K so large that $\mathbb{P}\left(\tilde{\mu}([0,t]\times[-a,a]\setminus[-r,r]) \geq K\right) < \epsilon$. Let $\delta > 0$ be so small that if $x, y \in [-R-a, R+a]$ and $|x-y| < \delta$ then $|f(x) - f(y)| < \frac{c^2}{24K}$. Take $n_1 > 0$ so large that $\mathbb{P}\left((X^n - X)_t^* \geq \delta\right) < \epsilon$ for $n \geq n_1$. Since $\tilde{\mu}([0,t]\times[-a,a]\setminus[-r,r]) < \infty$ and since $f_n' \to f'$ uniformly on compacts there exists n_2 such that if $n \geq n_2$ then

$$\int_0^t \int_{r<|x|\leq a} \left|f_n'(X_{s-}^n) - f'(X_{s-})\right|^2 \tilde{\mu}(ds,dx) < \frac{c}{10},$$

on $A_R(n)$. We define

$$E_3(n) = A_R(n) \cap \{(X^n - X)_t^* \leq \delta\} \cap \left\{\int_0^t \int_{|x|\leq r} x^2 \tilde{\mu}(ds,dx) < \frac{c^2}{64M^2}\right\} \cap$$
$$\cap \{\tilde{\mu}([0,t]\times[-a,a]\setminus[-r,r]) < K\}$$

Now let $n \geq \max(n_1, n_2)$. If we now restrict our attention to the set $E_3(n)$ and apply the mean-value theorem then for some $\theta_1(s,\omega), \theta_2(s,\omega) \in [0,1]$,

$$\int_0^t \int_{|x|\leq r} \left(f_n(X_{s-}^n + x) - f(X_{s-} + x) - f_n(X_{s-}^n) + f(X_{s-}) - x(f_n'(X_{s-}^n) - f'(X_{s-}))\right)^2 \tilde{\mu}(ds,dx)$$
$$\leq \int_0^t \int_{|x|\leq r} \left(\left|f'(X_{s-} + \theta_1 x)\right||x| + \left|f_n'(X_{s-}^n + \theta_2 x)\right||x| + |x|\left|f'(X_{s-})\right| + |x|\left|f_n'(X_{s-}^n)\right|\right)^2 \tilde{\mu}(ds,dx)$$
$$\leq \int_0^t \int_{|x|\leq r} 8x^2 \left(\left|f_n'(X_{s-}^n + \theta_2 x)\right|^2 + \left|f'(X_{s-} + \theta_1 x)\right|^2 + |f'(X_{s-})|^2 + |f_n'(X_{s-}^n)|^2\right) \tilde{\mu}(ds,dx)$$
$$\leq 32M^2 \int_0^t \int_{|x|\leq r} x^2 \tilde{\mu}(ds,dx) < 32M^2 \cdot \frac{c^2}{64M^2} = c^2/2,$$

while

$$\int_0^t \int_{r<|x|\leq a} \left((f_n(X_{s-}^n + x) - f(X_{s-} + x)) - f_n(X_{s-}^n) + f(X_{s-})) - x\left(f_n'(X_{s-}^n) - f'(X_{s-})\right)\right)^2 \tilde{\mu}(ds,dx)$$
$$\leq \int_0^t \int_{r<|x|\leq a} \left(\left|f_n(X_{s-}^n + x) - f(X_{s-} + x)\right| + \left|f(X_{s-}) - f_n(X_{s-}^n)\right| + \left|f_n'(X_{s-}^n) - f'(X_{s-})\right|\right)^2 \tilde{\mu}(ds,dx)$$
$$\leq 3\frac{c^2}{6K}\tilde{\mu}([0,t]\times[-a,a]\setminus[-r,r]) < \frac{c^2}{2},$$

so therefore

$$\left(\int_0^t \int_{|x|\le a} \left(\left(f_n(X^n_{s-}+x) - f_n(X^n_{s-})\right) + f(X_{s-}+x) - f(X_{s-})\right)\right.$$
$$\left. -x\left(f_n'(X^n_{s-}) - f'(X_{s-})\right)\right)^2 \tilde{\mu}(ds,dx)\Bigg)^{\frac{1}{2}} < \left(\frac{c^2}{2}+\frac{c^2}{2}\right)^{\frac{1}{2}} = c.$$

Hence

$$\mathbb{P}\left(\left(\int_0^t \int_{|x|\le a} \left((f(X_{s-}+x) - f(X_{s-})) - xf'(X_{s-})\right)^2 \tilde{\mu}(ds,dx)\right)^{\frac{1}{2}} \ge c\right) \le \mathbb{P}(E_3(n)^c) < 4\epsilon.$$

To show convergence in L^1 under assumption a) we notice that by applying the mean value theorem

$$\left(\int_0^t \int_{|x|\le a} (f_n(X^n_{s-}+x) - f(X_{s-}+x) - f_n(X^n_{s-}) + f(X_{s-})\right.$$
$$\left. -x(f_n'(X^n_{s-}) - f'(X_{s-})))^2 \tilde{\mu}(ds,dx)\right)^{\frac{1}{2}} \le \sqrt{32}U\left(\int_0^t \int_{|x|\le a} x^2 \tilde{\mu}(ds,dx)\right)^{\frac{1}{2}} \le \sqrt{32}U[X]_t^{\frac{1}{2}},$$

so we have a L^1 bound for this term. Assuming b) then

$$\left(\int_0^t \int_{|x|\le a} \left(f_n(X^n_{s-}+x) - f(X_{s-}+x) - f_n(X^n_{s-}) + f(X_{s-}) - x(f_n'(X^n_{s-}) - f'(X_{s-}))\right)^2 \tilde{\mu}(ds,dx)\right)^{\frac{1}{2}}$$
$$= \left(\int_0^t \int_{|x|\le a} x^2\left(f_n'(X^n_{s-}+\theta_1 x) - f'(X_{s-}+\theta_2 x) - f_n'(X^n_{s-}) + f'(X_{s-})\right)^2 \tilde{\mu}(ds,dx)\right)^{\frac{1}{2}} \le$$
$$\left(\int_0^t \int_{|x|\le a} x^2\left(4M \vee 2C(|X^n_{s-}+\theta_1 x| + |X_{s-}+\theta_2 x| + |X^n_{s-}| + |X_{s-}|)\right)^2 \tilde{\mu}(ds,dx)\right)^{\frac{1}{2}} \le$$
$$\left(\int_0^t \int_{|x|\le a} x^2 \left(4M \vee 2C(2|x| + (X^n)^*_t + X^*_t)\right)^2 \tilde{\mu}(ds,dx)\right)^{\frac{1}{2}} \le (4M + 2C(2|a| + (X^n)^*_t + X^*_t))[X]_t^{\frac{1}{2}},$$

and since

$$\mathbb{E}\left[(4\tilde{M} + 2C(2|a| + (X^n)^*_t + X^*_t))[X]_t^{\frac{1}{2}} 1_E\right] \le C_1\mathbb{P}(E) + C_2\sqrt{\mathbb{E}\left[((X^n)^*_t)^2 + (X^*_t))^2)1_E\right]}\sqrt{\mathbb{E}\left[[X]_t\right]},$$

for any measurable E, some constants C_1, C_2 that depend only on $\tilde{M}, a$ and C, we see that term three is u.i. but we already established convergence in probability to zero therefore this implies convergence in L^1 to zero of term three.

For term four of (A.1.17) let $r \in \mathbb{R}^+$ and $\tilde{N}$ denote the number of jumps of X larger than $r/2$ and take K so large that $\mathbb{P}(\tilde{N} \ge K) < \epsilon$. We may now take $L(r) \in [r/2, r]$ such that $\mathbb{P}\left(\exists s \le t : |\Delta X_s| \in [L(r)-\gamma, L(r)+\gamma]\right) < \epsilon/K$ for some $\gamma > 0$ (this is possible since there are only a finite number of jumps exceeding $r/2$ on $[0,t]$ for X) and we may assume without loss of generality that $\gamma < r/2$. Now let N and N_n denote the number of jumps larger than $L(r)$ of X and X^n in $[0,t]$ respectively, then clearly $N \le \tilde{N}$

since $L(r) \geq r/2$. Let $T_1, ..., T_k$ and $T_1^n, ..., T_{N_n}^n$ denote the jumps of X and X^n larger than $L(r)$ in $[0,t]$. Let $A_k := \{N = k\} \cap \left(\bigcap_{l=1}^N \{|\Delta X_{T_l}| \notin [L(r) - \gamma, L(r) + \gamma)\}\right)$ then $\mathbb{P}\left(\left(\bigcup_{k=1}^K A_k\right)^c\right) < 2\epsilon$. Now on A_k we know that if $||\Delta X_{T_l}| \geq L(r)$ then $|\Delta X_{T_l}| \geq L(r) + \gamma$ and if $|\Delta X_{T_l}^n| < L(r)$ then $|\Delta(X^n - X)_{T_l}| \geq \gamma$ for any $l \in \{1, ..., k\}$, similarly if $|\Delta X_{T_l^n}| < L(r)$ then in fact $|\Delta X_{T_l^n}| < L(r) - \gamma$ implying $|\Delta(X^n - X)_{T_l^n}| \geq \gamma$, so

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\{\{T_1^n, ..., T_{N_n}^n\} \neq \{T_1, ..., T_k\}\} \cap A_k\right) \leq \lim_{n\to\infty} \mathbb{P}\left(\{(\Delta(X^n - X))_t^* \geq \gamma\} \cap A_k\right) \\
&\leq \lim_{n\to\infty} \mathbb{P}\left((\Delta(X^n - X))_t^* \geq \gamma\right) \leq \lim_{n\to\infty} \left(\mathbb{P}\left(((X^n - X)_t^* \geq \gamma/2\right) + \mathbb{P}\left(((X^n - X)_{t-}^* \geq \gamma/2\right)\right) \\
&\leq 2 \lim_{n\to\infty} \mathbb{P}\left(((X^n - X)_t^* \geq \gamma/2\right) = 0,
\end{aligned}$$

for every k. Therefore

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\left(\{T_1^n, ..., T_{N_n}^n\} \neq \{T_1, ..., T_N\}\right) \leq \lim_{n\to\infty} \mathbb{P}\left(\{\{T_1^n, ..., T_{N_n}^n\} \neq \{T_1, ..., T_k\}\} \cap \left(\bigcup_{k=1}^K A_k\right)\right) \\
&+ \mathbb{P}\left(\left(\bigcup_{k=1}^K A_k\right)^c\right) \leq \sum_{k=1}^K \lim_{n\to\infty} \mathbb{P}\left(\{\{T_1^n, ..., T_{N_n}^n\} \neq \{T_1, ..., T_k\}\} \cap A_k\right) + \mathbb{P}\left(\left(\bigcup_{k=1}^K A_k\right)^c\right) \\
&= \mathbb{P}\left(\left(\bigcup_{k=1}^K A_k\right)^c\right) < 2\epsilon
\end{aligned}$$

letting $\epsilon \to 0^+$ shows $\lim_{n\to\infty} \mathbb{P}\left(\{T_1^n, ..., T_{N_n}^n\} \neq \{T_1, ..., T_N\}\right) = 0$, i.e. with probability tending to one the jump times corresponding to jumps of modulus greater or equal to $L(r)$ for X and and X^n on $[0,t]$ coincide, this implies

$$\sum_{s\leq t} \left|1_{|\Delta X_s^n| \geq L(r)} - 1_{|\Delta X_s| \geq L(r)}\right| \xrightarrow{\mathbb{P}} 0, \tag{A.1.18}$$

as $n \to \infty$ and since this sum cannot assume any values between 0 and 1 this means that for large n the sum is exactly zero with large probability. Let $B(n,r) = \left\{\sum_{s\leq t} \left|1_{|\Delta X_s^n| \geq L(r)} - 1_{|\Delta X_s| \geq L(r)}\right| = 0\right\}$ then $\mathbb{P}(B(n,r)) \to 1$ as $n \to \infty$, for each $r > 0$. On $B(n,r)$, by applying the AM-GM inequality

$$\begin{aligned}
&\sum_{s\leq t} (\Delta X_s^n)^2 I_{|\Delta X_s^n| \leq L(r)} = \sum_{s\leq t} (\Delta(X^n - X)_s + \Delta X_s^n)^2 I_{|\Delta X_s^n| \leq L(r)} \\
&\leq 2\sum_{s\leq t} (\Delta(X^n - X)_s)^2 + 2\sum_{s\leq t} (\Delta X_s)^2 I_{|\Delta X_s^n| \leq L(r)} \\
&\leq 2\sum_{s\leq t} (\Delta(X^n - X)_s)^2 + 2\sum_{s\leq t} (\Delta X_s)^2 I_{|\Delta X_s| \leq L(r)} \leq 2[X^n - X]_t + 2\sum_{s\leq t} (\Delta X_s)^2 I_{|\Delta X_s| \leq L(r)}
\end{aligned} \tag{A.1.19}$$

the last inequality follows from the fact that $|\Delta X_s^n| > L(r)$ if and only if $|\Delta X_s| > L(r)$ which is the same as to say that $|\Delta X_s^n| \leq L(r)$ if and only if $|\Delta X_s| \leq L(r)$ so the last sum contains at most all the jumps less or equal to $L(r)$ of X $((\Delta X_s)^2 I_{|\Delta X_s|\le L(r)} > 0$ only if $(\Delta X_s) \le L(r)$ on $B(n,r))$.

Applying Lemma 2.4 and doing a bit of rearrangement of term four of (A.1.16) gives us

$$\begin{aligned}
&\Bigg[\int_0^{\cdot}\int_{|x|\le a}\big(\big(f_n(X^n_{s-}+x)-f_n(X^n_{s-})\big)\big)-xf_n'(X^n_{s-})\big)\,((\tilde{\mu}-(\nu)_c)(ds,dx)\\
&-\int_0^{\cdot}\int_{|x|\le a}\big(\big(f_n(X^n_{s-}+x)-f_n(X^n_{s-})\big)\big)-xf_n'(X^n_{s-})\big)\,(\tilde{\mu}_n-(\nu_n)_c)(ds,dx)\Bigg]_t^{\frac{1}{2}}\\
&\le\left[\int_0^{\cdot}\int_{|x|\le L(r)}\big(\big(f_n(X^n_{s-}+x)-f_n(X^n_{s-})\big)\big)-xf_n'(X^n_{s-})\big)\,((\tilde{\mu}-\nu_c)(ds,dx)\right]_t^{\frac{1}{2}}\\
&+\left[\int_0^{\cdot}\int_{|x|\le L(r)}\big(\big(f_n(X^n_{s-}+x)-f_n(X^n_{s-})\big)\big)-xf_n'(X^n_{s-})\big)\,(\tilde{\mu}_n-(\nu_n)_c)(ds,dx)\right]_t^{\frac{1}{2}}\\
&+\left[\int_0^{\cdot}\int_{L(r)<|x|\le a}\big(\big(f_n(X^n_{s-}+x)-f_n(X^n_{s-})\big)\big)-xf_n'(X^n_{s-})\big)\,(\tilde{\mu}-\tilde{\mu}_n)(ds,dx)\right]_t^{\frac{1}{2}}\\
&+\left[\int_0^{\cdot}\int_{L(r)<|x|\le a}\big(\big(f_n(X^n_{s-}+x)-f_n(X^n_{s-})\big)\big)-xf_n'(X^n_{s-})\big)\,(\nu-\nu_n)_c(ds,dx)\right]_t^{\frac{1}{2}}.
\end{aligned}\tag{A.1.20}$$

We will now show convergence of each term in the right-hand side of (A.1.20). Let $\tilde{\epsilon} > 0$ be arbitrary. We will show that if we assume a), the L^1-limit (L^2-limit if we assume b)) in n, for some small r, of the right-hand side of (A.1.20) can be bounded by some constant times $\tilde{\epsilon}$ and since $\tilde{\epsilon}$ is arbitrary this will show that (A.1.20) converges to zero in L^1 if we assume a) (in L^2 if we assume b)). Assuming a), then by taking r so small that $\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|\le L(r)}\right)^{\frac{1}{2}}\right] < \tilde{\varepsilon}$ and using the fact that $\lim_{n\to\infty}\mathbb{E}\left[[X^n]_t^{\frac{1}{2}}1_{B(n,r)^c}\right] = 0$ (this is true since $\{[X^n]_t^{\frac{1}{2}}\}_n$ is u.i. and $\lim_{n\to\infty}\mathbb{P}\left(B(n,r)\right) = 0$ for every fixed r) we see that

$$\lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X^n_s)^2 I_{|\Delta X^n_s|\le L(r)}\right)^{\frac{1}{2}}\right]\le\lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X^n_s)^2 I_{|\Delta X^n_s|\le L(r)}\right)^{\frac{1}{2}}1_{B(n,r)^c}\right]+$$
$$\sqrt{2}\left(\lim_{n\to\infty}\mathbb{E}\left[[X^n-X]_t^{\frac{1}{2}}1_{B(n,r)}\right]+\lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|\le L(r)}\right)^{\frac{1}{2}}1_{B(n,r)}\right]\right)\le$$

$$\lim_{n\to\infty}\mathbb{E}\left[[X^n]_t^{\frac{1}{2}}1_{B(n,r)^c}\right]+\sqrt{2}\lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|\le L(r)}\right)^{\frac{1}{2}}\right]<\sqrt{2}\tilde{\epsilon},$$

where we used (A.1.19) and the triangle inequality at the first step. In the same fashion if we take r so small that $\mathbb{E}\left[\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|\le L(r)}\right]<\varepsilon$ we see that if we assume b) then $\lim_{n\to\infty}\mathbb{E}\left[\sum_{s\le t}(\Delta X_s^n)^2 I_{|\Delta X_s^n|\le L(r)}\right]<2\tilde{\epsilon}$. For the first term on the right-hand side of (A.1.20) we have assuming a)

$$\mathbb{E}\left[\left[\int_0^{\cdot}\int_{|x|\le L(r)}\left((f_n(X^n_{s-}+x)-f_n(X^n_{s-})))-xf_n'(X^n_{s-})\right)((\tilde{\mu}-\nu_c)(ds,dx)\right]_t^{\frac{1}{2}}\right]=$$

$$\mathbb{E}\left[\left(\int_0^{t}\int_{|x|\le L(r)}\left((f_n(X^n_{s-}+x)-f_n(X^n_{s-})))-xf_n'(X^n_{s-})\right)^2\tilde{\mu}(ds,dx)\right)^{\frac{1}{2}}\right]\le$$

$$\mathbb{E}\left[\left(\int_0^{t}\int_{|x|\le L(r)}x^2\left(f_n'(X^n_{s-}+\theta x)-f_n'(X^n_{s-})\right)^2\tilde{\mu}(ds,dx)\right)^{\frac{1}{2}}\right]\le$$

$$2U\mathbb{E}\left[\left(\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|\le L(r)}\right)^{\frac{1}{2}}\right]\le 2U\tilde{\epsilon},$$

Assuming b) then

$$\mathbb{E}\left[\left[\int_0^{\cdot}\int_{|x|\le L(r)}\left((f_n(X^n_{s-}+x)-f_n(X^n_{s-})))-xf_n'(X^n_{s-})\right)((\tilde{\mu}-\nu_c)(ds,dx)\right]_t^{\frac{1}{2}}\right]\le$$

$$\mathbb{E}\left[\left(\int_0^{t}\int_{|x|\le L(r)}x^2\left(2M\vee 2C(2(X^n)^*_t+|x|)\right)^2\tilde{\mu}(ds,dx)\right)^{\frac{1}{2}}1_E\right]\le$$

$$\mathbb{E}\left[(2M+2C(2(X^n)^*_t+|a|))\left(\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|\le L(r)}\right)^{\frac{1}{2}}\right]\le$$

$$\sqrt{\mathbb{E}\left[(2M+2C(2(X^n)^*_t+|a|))^2\right]}\sqrt{\mathbb{E}\left[\sum_{s\le t}(\Delta X_s)^2 I_{|\Delta X_s|\le L(r)}\right]}\le\sqrt{\mathbb{E}\left[(2M+2C(2(X^n)^*_t+|a|))^2\right]}2\tilde{\epsilon},$$

since $\{((X^n)^2)^*_t\}_n$ is u.i.the factor in front of $\tilde{\epsilon}$ can be made uniformly bounded over all n.

For the second term of (A.1.20) the first steps are analogous to those of the first term. Assuming a) and skipping ahead,

$$\mathbb{E}\left[\left(\left[\int_0^{\cdot}\int_{|x|\le L(r)}(((f_n(X^n_{s-}+x)-f_n(X^n_{s-})))-xf_n'(X^n_{s-}))(\tilde{\mu}_n-(\nu_n)_c)(ds,dx)\right]_t\right)^{\frac{1}{2}}\right]$$

$$\le\mathbb{E}\left[\left(\int_0^{\cdot}\int_{|x|\le L(r)}x^2\left(f_n'(X^n_{s-}+\theta x)-f_n'(X^n_{s-})\right)^2\tilde{\mu}_n(ds,dx)\right)^{\frac{1}{2}}\right]$$

$$\leq 2U\mathbb{E}\left[\left(\sum_{s\leq t}(\Delta X_s^n)^2 1_{|\Delta X_s^n|\leq L(r)}\right)^{\frac{1}{2}}\right] \leq 2U\sqrt{2}\tilde{\epsilon}.$$

Assuming b) also gives the analogous bound that we had for term one of (A.1.20).

Next, for the third term in the right-hand side of (A.1.20) fix some arbitrary $\epsilon, c > 0$. Note that since $f_n \to f$ uniformly on $[-R-a, R+a]$ there exists $\delta > 0$ such that if $|x-y| < \delta$ then $|f_n(x) - f_n(y)| < \frac{\sqrt{c}}{2K}$ for $x, y \in [-R-a, R+a]$. Let

$$E_4(n) := \left(\bigcup_{k=1}^{K} A_k\right) \cap \left\{\sum_{s\leq t}\left|1_{|\Delta X_s^n|\geq L(r)} - 1_{|\Delta X_s|\geq L(r)}\right| = 0\right\} \cap A_R(n)$$
$$\cap \left\{(\Delta(X^n - X))_t^* < \min\left(\delta, \frac{\sqrt{c}}{2MK}\right)\right\},$$

then $\mathbb{P}(E_4(n)^c) < 6\epsilon$ for large enough n. We have on $E_4(n)$, since $1_{L(r)<|\Delta X_s|\leq a} = 1_{L(r)<|\Delta X_s^n|\leq a}$ for every $s \leq t$,

$$\left[\int_0^{\cdot}\int_{L(r)<|x|\leq a}\left(\left(f_n(X_{s-}^n + x) - f_n(X_{s-}^n)\right)\right) - x f_n'(X_{s-}^n)\right)(\tilde{\mu} - \tilde{\mu}_n)(ds, dx)\right]_t$$
$$= \left[\sum_{s\leq \cdot}\left(\left(f_n(X_{s-}^n + \Delta X_s) - f_n(X_{s-}^n)\right)\right) - \Delta X_s f_n'(X_{s-}^n)\right) 1_{L(r)<|\Delta X_s|\leq a}\right.$$
$$\left. - \sum_{s\leq \cdot}\left(\left(f_n(X_{s-}^n + \Delta X_s^n) - f_n(X_{s-}^n)\right)\right) - \Delta X_s^n f_n'(X_{s-}^n)\right) 1_{L(r)<|\Delta X_s^n|\leq a}\right]_t$$
$$= \left[\sum_{s\leq \cdot}\left(f_n(X_{s-}^n + \Delta X_s) - f_n(X_{s-}^n + \Delta X_s^n) - \Delta(X^n - X)_s f_n'(X_{s-}^n)\right) 1_{L(r)<|\Delta X_s|\leq a}\right]_t$$
$$\leq \sum_{s\leq t}\left(f_n(X_{s-}^n + \Delta X_s) - f_n(X_{s-}^n + \Delta X_s^n) - \Delta(X^n - X)_s f_n'(X_{s-}^n)\right)^2 1_{L(r)<|\Delta X_s|\leq a}$$
$$\leq K\left(\frac{\sqrt{c}}{4\sqrt{K}} + M\frac{\sqrt{c}}{4\sqrt{K}M}\right)^2 < c.$$

Assuming a) we have for some $\theta_1(s,\omega), \theta_1(s,\omega)$,

$$\left[\int_0^{\cdot}\int_{L(r)<|x|\leq a}\left(\left(f_n(X_{s-}^n + x) - f_n(X_{s-}^n)\right)\right) - x f_n'(X_{s-}^n)\right)(\tilde{\mu} - \tilde{\mu}_n)(ds, dx)\right]_t^{\frac{1}{2}} \leq$$
$$\left(\sum_{s\leq t}\left(f_n'(X_{s-}^n + \theta_1(s)\Delta X_s) - f_n'(X_{s-}^n)\right)^2 (\Delta X_s)^2 1_{L(r)<|\Delta X_s|\leq a} + \right.$$

$$\sum_{s\le t}\left(f_n'(X^n_{s-}+\theta_2(s)\Delta X^n_s)-f_n'(X^n_{s-})\right)^2(\Delta X^n_s)^2 1_{L(r)<|\Delta X^n_s|\le a}\Bigg)^{\frac{1}{2}} \le$$

$$\left(\sum_{s\le t}4U^2(\Delta X_s)^2 1_{L(r)<|\Delta X_s|\le a}+4U^2\sum_{s\le t}(\Delta X^n_s)^2 1_{L(r)<|\Delta X^n_s|\le a}\right)^{\frac{1}{2}}$$

$$\le 2U\left([X]_t+[X^n]_t\right)^{\frac{1}{2}}\le 2U\left(\sqrt{[X]_t}+\sqrt{[X^n]_t}\right),$$

the right-most side is u.i. since $\sqrt{[X^n]_t}$ is. Assuming b)

$$\mathbb{E}\left[\left[\int_0^{\cdot}\int_{L(r)<|x|\le a}\left((f_n(X^n_{s-}+x)-f_n(X^n_{s-})))-xf_n'(X^n_{s-})\right)(\tilde\mu-\tilde\mu_n)(ds,dx)\right]_t^{\frac{1}{2}}1_E\right]\le$$

$$\mathbb{E}\left[\tilde C(M\vee((X^n)^*_t+X^*_t))\left(\sum_{s\le t}(\Delta X_s)^2 1_{L(r)<|\Delta X_s|\le a}+\sum_{s\le t}(\Delta X^n_s)^2 1_{L(r)<|\Delta X^n_s|\le a}\right)^{\frac{1}{2}}1_E\right]\le$$

$$\sqrt{\mathbb{E}\left[\tilde C^2(M\vee((X^n)^*_t+X^*_t))^2 1_E\right]}\sqrt{\mathbb{E}\left[([X]_t+[X^n]_t)\,1_E\right]},$$

so this term is u.i. as well.
Since the process $\int_0^{\cdot}\int_{L(r)<|x|\le a}((f_n(X^n_{s-}+x)-f_n(X^n_{s-})))-xf_n'(X^n_{s-}))\,(\nu-\nu_n)_c(ds,dx)$ is continuous and of finite variation it has zero quadratic variation so the final term gives no contribution. We have thus shown that when assuming a) the right-hand side of (A.1.20) has an L^1 limit less or equal $D'\tilde\epsilon$ for some $D'\in\mathbb{R}^+$. Similarly the right-hand side of (A.1.20) has an L^2 limit less or equal $D''\tilde\epsilon$ for some $D''\in\mathbb{R}^+$ when assuming b). This takes care of term four of (A.1.17).

For term five of (A.1.17) we note that

$$\Bigg(\sum_{s\le t}\Bigg(\int_{a<|x|\le a}\big(f_n(X^n_{s-}+x)-f(X_{s-}+x)-f_n(X^n_{s-})+f(X_{s-})$$

$$-x(f_n'(X^n_{s-})-f'(X_{s-}))\big)\,\mu(\{s\},dx)\Bigg)^2\Bigg)^{\frac{1}{2}}\le$$

$$\sum_{s\le t}\left|\int_{a<|x|\le a}\left(f_n(X^n_{s-}+x)-f(X_{s-}+x)-f_n(X^n_{s-})+f(X_{s-})-x(f_n'(X^n_{s-})-f'(X_{s-}))\right)\mu(\{s\},dx)\right|$$

$$\le\sum_{s\le t}\int_{|x|\le a}\left|f_n(X^n_{s-}+x)-f(X_{s-}+x)-f_n(X^n_{s-})+f(X_{s-})-x(f_n'(X^n_{s-})-f'(X_{s-}))\right|\mu(\{s\},dx). \tag{A.1.21}$$

Take $W\in\mathbb{R}^+$ so large that $\mathbb{P}\left(\mu([0,t]\times[-a,a]\setminus[-r,r])\ge W\right)<\epsilon$. Also let n' be so large and $\delta'>0$ be so small that $|x-y|<\delta'$ and $n\ge n'$ implies that

$\max\left(|f_n(x) - f(y)|, |f_n'(x) - f'(y)|\right) < \frac{c}{6W}$. Define

$$E_5(n) = A_R(n) \cap \{(X^n - X)_t^* \leq \delta\} \cap \left\{ \sum_{s \leq t} \int_{|x| \leq r} |x| \mu(\{s\}, dx) < \frac{c}{8M} \right\} \cap$$

$$\cap \left\{ \mu([0,t] \times [-a,a] \setminus [-r,r]) \geq W \right\},$$

then for large n we have $\mathbb{P}\left(E_5(n)^c\right) < 4\epsilon$. Let us now restrict our attention to this subset. Similarly to term three of (A.1.17) we may split the right-hand side of (A.1.21)

$$\sum_{s \leq t} \int_{|x| \leq r} \left| f_n(X_{s-}^n + x) - f(X_{s-} + x) - f_n(X_{s-}^n) + f(X_{s-}) - x(f_n'(X_{s-}^n) - f'(X_{s-})) \right| \mu(\{s\}, dx)$$

$$\leq \sum_{s \leq t} \int_{|x| \leq r} |x| \left(\left| f'(X_{s-} + \theta_1 x) \right| + \left| f_n'(X_{s-}^n + \theta_2 x) \right| + \left| f'(X_{s-}) \right| + \left| f_n'(X_{s-}^n) \right| \right) \mu(\{s\}, dx)$$

$$\leq \sum_{s \leq t} \int_{|x| \leq r} 4M |x| \mu(\{s\}, dx) < 4M \frac{c}{8M} = \frac{c}{2}$$

while

$$\sum_{s \leq t} \int_{r < |x| \leq a} \left| \left(f_n(X_{s-}^n + x) - f(X_{s-} + x) \right) - f_n(X_{s-}^n) + f(X_{s-}) \right) - x \left(f_n'(X_{s-}^n) - f'(X_{s-}) \right) \right| \mu(\{s\}, dx)$$

$$\leq \sum_{s \leq t} \int_{r < |x| \leq a} \left(\left| f_n(X_{s-}^n + x) - f(X_{s-} + x) \right| + \left| f(X_{s-}) - f_n(X_{s-}^n) \right| + \left| f_n'(X_{s-}^n) - f'(X_{s-}) \right| \right) \mu(\{s\}, dx)$$

$$\leq 3 \frac{c}{10W} \mu([0,t] \times [-a,a] \setminus [-r,r]) < 3 \frac{c}{6W} \cdot W = \frac{c}{2},$$

and therefore

$$\sum_{s \leq t} \int_{|x| \leq a} \left| \left(f_n(X_{s-}^n + x) - f_n(X_{s-}^n) \right) + f(X_{s-} + x) - f(X_{s-}) \right) - x \left(f_n'(X_{s-}^n) - f'(X_{s-}) \right) \right| \mu(\{s\}, dx)$$

$$< c,$$

on $E_5(n)$ so term five converges to zero in probability. Assuming a) we can bound the right-most side of (A.1.21) by $3U \sum_{s \leq t} \int_{|x| \leq a} |x| \mu(\{s\}, dx) \in L^1$, and use dominated convergence. Assuming b) then

$$\mathbb{E}\left[\sum_{s \leq t} \int_{|x| \leq a} \left| \left(f_n(X_{s-}^n + x) - f_n(X_{s-}^n) \right) + f(X_{s-} + x) - f(X_{s-}) \right) \right.$$

$$\left. - x \left(f_n'(X_{s-}^n) - f'(X_{s-}) \right) \right| \mu(\{s\}, dx) 1_E \right] \leq$$

$$\mathbb{E}\left[\sum_{s \leq t} \int_{|x| \leq a} |x| \left(\left| f'(X_{s-} + \theta_1 x) \right| + \left| f_n'(X_{s-}^n + \theta_2 x) \right| + \left| f'(X_{s-}) \right| + \left| f_n'(X_{s-}^n) \right| \right) \mu(\{s\}, dx) 1_E \right]$$

$$\leq \mathbb{E}\left[2C \left(2a + 2(X^n)_t^* + 2X_t^* \right) \sum_{s \leq t} \int_{|x| \leq a} |x| \mu(\{s\}, dx) 1_E \right]$$

$$\leq 2C\sqrt{\mathbb{E}\left[(2a+2(X^n)^*_t+2X^*_t)^2\, 1_E\right]}\sqrt{\mathbb{E}\left[\left(\sum_{s\leq t}\int_{|x|\leq a}|x|\mu(\{s\},dx)\right)^2 1_E\right]},$$

which shows u.i.

For term six of (A.1.17), take r so small that

$$\mathbb{P}\left(\left\{\sup_n\sum_{s\leq t}\int_{|x|\leq r}|x|\tilde{\mu}_n(\{s\},dx)<\frac{c}{4M}\right\}\cap\left\{\sum_{s\leq t}\int_{|x|\leq r}|x|\tilde{\mu}(\{s\},dx)<\frac{c}{4M}\right\}\right)<\epsilon$$

let $L(r)$, A_k and K be as defined earlier (for term four). Since

$$\sum_{s\leq t}|\Delta(X^n-X)_s|\, 1_{L(r)<|\Delta X^n_s|\leq a}$$

only contains a finite number of terms and since $(\Delta(X^n-X))^*_t\xrightarrow{\mathbb{P}}0$ it follows that this term converges to zero in probability. Let

$$E_6(n):=\left\{\sup_n\sum_{s\leq t}\int_{|x|\leq L(r)}|x|\tilde{\mu}_n(\{s\},dx)<\frac{c}{4M}\right\}\cap\left\{\sum_{s\leq t}\int_{|x|\leq L(r)}|x|\tilde{\mu}(\{s\},dx)<\frac{c}{4M}\right\}$$

$$\cap\left\{\sum_{s\leq t}\left|1_{|\Delta X^n_s|\geq L(r)}-1_{|\Delta X_s|\geq L(r)}\right|=0\right\}\cap A_R(n)\cap\left\{(\Delta(X^n-X))^*_t<\frac{c}{4MK}\right\}\cap\left(\bigcup_{k=1}^K A_k\right),$$

then for large n we can make $E_6(n)$ smaller than say 6ϵ. Now for some $\theta(s,\omega)\in[0,1]$ we have on $E_6(n)$

$$\left(\sum_{s\leq t}\left(\int_{|x|\leq a}\left(f_n(X^n_{s-}+x)-f_n(X^n_{s-})-xf'_n(X_{s-})\right)(\mu-\mu_n)(\{s\},dx)\right)^2\right)^{\frac{1}{2}}\leq$$

$$\sum_{s\leq t}\left|\int_{|x|\leq a}\left(f_n(X^n_{s-}+x)-f_n(X^n_{s-})-xf'_n(X_{s-})\right)(\mu-\mu_n)(\{s\},dx)\right|\leq$$

$$\sum_{s\leq t}\left|\int_{|x|\leq L(r)}\left(f_n(X^n_{s-}+x)-f_n(X^n_{s-})-xf'_n(X_{s-})\right)(\mu-\mu_n)(\{s\},dx)\right|+$$

$$\sum_{s\leq t}\left|\int_{L(r)<|x|\leq a}\left(f_n(X^n_{s-}+x)-f_n(X^n_{s-})-xf'_n(X_{s-})\right)(\mu-\mu_n)(\{s\},dx)\right|\leq$$

$$2M\sum_{s\leq t}\int_{|x|\leq L(r)}|x|(\mu+\mu_n)(\{s\},dx)+$$

$$\sum_{s\leq t,}\left|\left(\left(f_n(X^n_{s-}+\Delta X_s)-f_n(X^n_{s-})\right)\right)-\Delta X_s f'_n(X^n_{s-})\right)1_{L(r)<|\Delta X_s|\leq a,s\in\mathcal{A}}$$

$$
- \left(\left(f_n(X^n_{s-} + \Delta X^n_s) - f_n(X^n_{s-})) \right) - \Delta X^n_s f_n'(X^n_{s-}) \right) 1_{L(r)<|\Delta X^n_s|\le a, s\in\mathcal{A}_n} \Big| \le
$$

$$
\frac{c}{2} + \sum_{s\le t} \left| f_n(X^n_{s-} + \Delta X_s) - f_n(X^n_{s-} + \Delta X^n_s) - \Delta(X^n - X)_s f_n'(X^n_{s-}) \right| 1_{L(r)<|\Delta X^n_s|\le a, s\in\mathcal{A}_n}
$$

$$
+ \sum_{s\le t} \left| \Delta f_n(X^n_{s-}) - \Delta X^n_s f_n'(X^n_{s-}) \right| \left| 1_{L(r)<|\Delta X^n_s|\le a, s\in\mathcal{A}_n} - 1_{L(r)<|\Delta X_s|\le a, s\in\mathcal{A}} \right| \le
$$

$$
\frac{c}{2} + \sum_{s\le t} \left| \left(f_n'(X^n_{s-} + \Delta X_s + \theta\Delta(X^n - X)_s) - f_n'(X^n_{s-}) \right) \Delta(X^n - X)_s \right| 1_{L(r)<|\Delta X^n_s|\le a} \le
$$

$$
\frac{c}{2} + 2M \sum_{s\le t} |\Delta(X^n - X)_s| \, 1_{L(r)<|\Delta X^n_s|\le a} < \frac{c}{2} + 2MK\frac{c}{4MK} = c
$$

where we used the fact that $\sum_{s\le t} \left| 1_{L(r)<|\Delta X^n_s|\le a} - 1_{L(r)<|\Delta X_s|\le a} \right| = 0$ on $E_6(n)$. We have now established that term six does go to zero in probability. Assuming a)

$$
\left(\sum_{s\le t} \left(\int_{|x|\le a} \left(f_n(X^n_{s-} + x) - f_n(X^n_{s-}) - x f_n'(X_{s-}) \right) (\mu - \mu_n)(\{s\}, dx) \right)^2 \right)^{\frac{1}{2}}
$$

$$
\le 2U \sum_{s\le t} \int_{|x|\le L(r)} |x| (\mu + \mu_n)(\{s\}, dx)
$$

which is u.i. since $\sum_{s\le t} \int_{|x|\le L(r)} |x| \mu_n(\{s\}, dx) \xrightarrow{L^1} \sum_{s\le t} \int_{|x|\le L(r)} |x| \mu(\{s\}, dx)$. Assuming b)

$$
\mathbb{E}\left[\left(\sum_{s\le t} \left(\int_{|x|\le a} \left(f_n(X^n_{s-} + x) - f_n(X^n_{s-}) - x f_n'(X_{s-}) \right) (\mu - \mu_n)(\{s\}, dx) \right)^2 \right)^{\frac{1}{2}} 1_E \right] \le
$$

$$
\mathbb{E}\left[2C \left(2|x| + 2(X^n)^*_t + 2X^*_t \right) \sum_{s\le t} \int_{|x|\le L(r)} |x| (\mu + \mu_n)(\{s\}, dx) 1_E \right] \le
$$

$$
2C \sqrt{\mathbb{E}\left[\left(2|x| + 2(X^n)^*_t + 2X^*_t \right)^2 1_E \right]} \sqrt{\mathbb{E}\left[\left(\sum_{s\le t} \int_{|x|\le L(r)} |x| (\mu + \mu_n)(\{s\}, dx) \right)^2 1_E \right]}
$$

We now outline the proof in the case a) with $p = 1$ and b) with $q = 2$. In place of the inequality (A.1.16) we instead estimate $[f_n(X^n) - f(X)]_t$ by using Lemma 2.5. Note that by doing so we get the constant 2^6 in front of each term (but this is obviously will not affect the L^1 convergence). Now that we do not take the square root of each term this has the only effect that it doubles the moment requirements and this will be the only difference in the proof for this case. □

Bibliography

[1] Coquet, F., Mémin, J. and Słomiński, L. (2003). On Non-Continuous Dirichlet Processes. *Journal of Theoretical Probability*, **16**, 197.

[2] Föllmer, H. (1981). Dirichlet processes. In *Lecture Notes in Maths.*, Vol. **851**, , pp. 476-–478, Springer-Verlag, Berlin/Heidelberg/New York.

[3] Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes.* Springer-Verlag.

[4] Liptser, R. and Shiryaev, A. (1986) *Theory of Martingales.* Springer-Verlag.

[5] Protter, P. (1992). *Stochastic Integration and Differential Equations.* Springer-Verlag, Berlin, Heidelberg, Second edition.

[6] Royden, H.L., Fitzpatrick P.M. (2010). *Real analysis.* Prentice Hall, Fourth edition.