QUANTITATIVE CONVERGENCE GUARANTEES FOR THE MEAN-FIELD DISPERSION PROCESS

Fei Cao[✉1] and Jincheng Yang[✉*2]

[1]Department of Mathematics and Statistics, University of Massachusetts Amherst, USA

[2]Department of Applied Mathematics and Statistics, Johns Hopkins University, USA

(Communicated by José Antonio Carrillo)

ABSTRACT. We study the discrete Fokker–Planck equation associated with the mean-field dynamics of a particle system called the dispersion process. For different regimes of the average number of particles per site (denoted by $\mu > 0$), we establish various quantitative long-time convergence guarantees toward the global equilibrium (depending on the sign of $\mu - 1$), which is also confirmed by numerical simulations. The main novelty/contribution of this manuscript lies in the careful and tricky analysis of a nonlinear Volterra-type integral equation satisfied by a key auxiliary function.

1. **Introduction.** In this paper, we study the following one-dimensional discrete Fokker–Planck equation which models the evolution of the particle distribution $\mathbf{p}(t) = \{p_n(t)\}_{n=0}^{\infty}$ associated with the mean-field dispersion process for $t > 0$ (see Section 1.1 for more details on the origin of the following dynamical system):

$$\mathbf{p}'(t) = \mathcal{D}_+[\{n\mathbf{1}_{\{n\geq 2\}}\}_n \mathbf{p}(t)] - \mathcal{D}_-[a(t)\mathbf{p}(t)], \qquad a(t) = \sum_{n=2}^{\infty} np_n(t). \qquad (1)$$

Here, $\mathbf{p} : [0, \infty) \to \mathcal{S}$ represents the probability mass function of the number of particles in a site, where

$$\mathcal{S} = \left\{ \mathbf{q} \in [0,1]^{\mathbb{N}} : \mathbf{q} = \{q_n\}_{n=0}^{\infty} \text{ with } \sum_{n=0}^{\infty} q_n = 1 \right\}$$

is the set of probability distributions on $\mathbb{N}$. $\mathcal{D}_+/\mathcal{D}_-$ are the forward/backward difference operators, and $a : [0, \infty) \to \mathbb{R}$ represents the number of active particles per site. This is an infinite-dimensional dynamical system $\mathbf{p}'(t) = \mathcal{F}[\mathbf{p}(t)]$ where the generator $\mathcal{F} : \mathcal{S} \to \mathbb{R}^{\mathbb{N}}$ can be written down in components as follows:

$$\mathcal{F}[\mathbf{q}]_n = \begin{cases} -\left(\sum_{k\geq 2} kq_k \right) q_0 & n = 0, \\ 2q_2 - \left(\sum_{k\geq 2} kq_k \right)(q_1 - q_0) & n = 1, \\ (n+1)q_{n+1} - nq_n - \left(\sum_{k\geq 2} kq_k \right)(q_n - q_{n-1}) & n \geq 2. \end{cases} \qquad (2)$$

Our main result is summarized in the following theorem. Below, W_0 denotes the principal branch of the Lambert W function [35] (defined to be the unique solution $y = W_0(x)$ to the relation $y \mathrm{e}^{-y} = x$ whenever $x \geq -\mathrm{e}^{-1}$), and we employ the Japanese bracket notation $\langle t \rangle := \sqrt{1 + t^2}$.

Theorem 1.1. *Let* $\mathbf{p}^0 = \{p_n^0\}_{n=0}^{\infty} \in \mathcal{S}$, *and denote its mean by* $\mu = \sum_{n=0}^{\infty} n p_n^0$. *If* $\mu \leq 1$, *we assume in addition that its second moment is finite. Then, there exists a positive constant* C *depending only on the second moment of* $\mathbf{p}^0$ *when* $\mu \leq 1$, *or on the first moment* μ *when* $\mu > 1$, *such that the solution* $\mathbf{p}(t)$ *to* (1) *with initial value* $\mathbf{p}(0) = \mathbf{p}^0$ *converges strongly as* $t \to +\infty$ *at the following rate:*

  i) *If* $0 < \mu < 1$, *then*

$$\|\mathbf{p}(t) - \mathbf{p}^*\|_{\ell^1} \leq C \, \mathrm{e}^{-2(1-\mu)t} \,.$$

 ii) *If* $\mu = 1$, *then*

$$\|\mathbf{p}(t) - \mathbf{p}^*\|_{\ell^1} \leq C \, t^{-1} \,.$$

  For $\mu > 1$, *denote* $\nu := \mu + W_0(-\mu \, \mathrm{e}^{-\mu})$. *Then:*

iii) *If* $1 < \mu < \frac{\mathrm{e}}{\mathrm{e}-1}$, *then*

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1} \leq C \langle t \rangle^{\frac{1}{2}} \, \mathrm{e}^{-\nu t} \,.$$

iv) *If* $\mu \geq \frac{\mathrm{e}}{\mathrm{e}-1}$, *then there exists* $K > 0$ *depending only on* μ *such that*

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1} \leq C \langle t \rangle^{K+\frac{1}{2}} \, \mathrm{e}^{-t} \,.$$

  The equilibrium is given by the following:

$$\mathbf{p}^* = \{1 - \mu, \mu, 0, 0, \dots\}, \tag{3}$$

$$\bar{\mathbf{p}} = \{0, \bar{p}_1, \bar{p}_2, \bar{p}_3, \dots\} \qquad with \qquad \bar{p}_n = \frac{\nu^n}{n!} \cdot \frac{1}{\mathrm{e}^{\nu} - 1}, \qquad n \geq 1. \tag{4}$$

The proof of this theorem is divided into Corollary 3.3 in Section 3 for $\mu \leq 1$ and Corollary 4.14 in Section 4 for $\mu > 1$. We also emphasize that searching for sharp or optimal decay rates in the aforementioned results is out of the scope of this manuscript.

1.1. **Derivation of the model.** Equation (1) can be derived by resorting to a similar approach as taken in [15] as the mean-field limit of the dispersion process, which describes the evolution of a stochastic particle system on a graph. The dispersion process is a continuous-time Markov process where M indistinguishable particles are distributed and moving across the vertices of a complete graph with N vertices known as "sites". At any given time, each particle that shares a common site with at least one other particle is called "active", and it will move to a neighbor site, uniformly randomly chosen, at a prescribed fixed rate (e.g. by Poisson signal). Particles that reside alone on a site are called "inactive", and they will stay there until activated by an incoming particle. This scheme is illustrated in Figure 1. The process will terminate when there are no remaining active particles. This model was first introduced in [21] and further studied in [23, 28, 44], where the random time of termination is analyzed using probabilistic tools.

The evolution of the particle distribution on the graph depends heavily on the ratio $\mu = M/N$, which represents the average number of particles per site. In the underpopulated scenario $\mu \leq 1$, the process will terminate almost surely when all
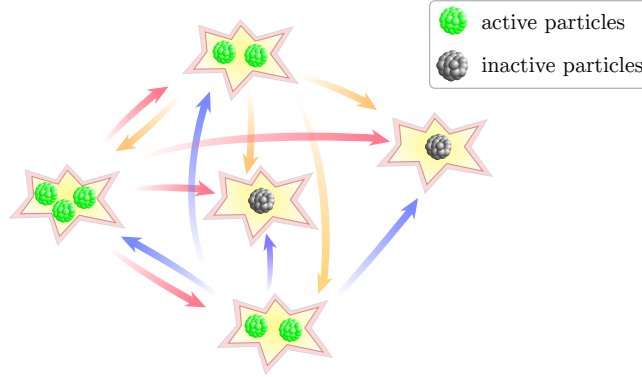
FIGURE 1. Illustration of the dispersion dynamics on a complete graph with $N = 5$ nodes/sites and $M = 9$ particles. Particles which share a common site will be "active" and move across sites.

particles become inactive, but in the overpopulated case $\mu > 1$ the process will continue indefinitely.

Let us now analyze this process from the sites' perspective. Denote the number of particles on site i at time t by $X_i(t)$. Particle dynamics can be represented by the following update rule: For each pair (i, j) of sites, if $X_i > 1$, then one particle will move from site i to site j at a rate proportional to X_i:

$$(X_i, X_j) \xrightarrow{\text{rate } X_i} (X_i - 1, X_j + 1) \qquad \text{if } X_i > 1. \tag{5}$$

The rate at which site i loses a particle is proportional to X_i, because every particle here is active. As a consequence, the rate at which any site gains a particle is proportional to the total number of active particles per site $\frac{1}{N} \sum_{i=1}^{N} X_i \mathbf{1}_{\{X_i \geq 2\}}$. By taking the large population limit as $N, M \to +\infty$, we formally obtain the Fokker–Planck equation (1). We will keep the scaling law $M/N \to \mu \in (0, +\infty)$ so that the law of X_i will converge.

The continuous-time dispersion model we have described is a standard interacting particle system and is amenable to mean-field type analysis under the large population limit $N \to \infty$, which is detailed in a recent work [15] on a related model. The rigorous justification of this transition from the stochastic interacting agents systems (5) into the associated mean-field ODE system (1) requires the proof of the *propagation of chaos* property [45], which is beyond the scope of the present manuscript.

1.2. **Review on related models.** Here, we detail several related models in the literature. One closely related model is the *sticky dispersion process*, which is a variant of the dispersion process where each inactive particle stays permanently inactive. This model was studied in [15], where the mean-field limit was derived and the convergence to the equilibrium was established. The update rule is as follows:

$$(X_i, X_j) \xrightarrow{\text{rate } X_i - 1} (X_i - 1, X_j + 1) \qquad \text{if } X_i > 1. \tag{6}$$

In the mean-field limit, the long-time equilibrium distribution is also the Bernoulli distribution for $\mu \leq 1$, and for $\mu > 1$ the equilibrium is the modified Poisson distribution [15]. The analysis is slightly simpler than the standard dispersion process

because the number of active particles $a(t)$ can be explicitly solved. Although the dispersion process investigated in this work resembles (at the microscopic level) its sticky version studied in the companion work [15] and the final asymptotic behavior might appear similar in one sense or another, the large time analysis of the mean-field ODE system (1) is significantly harder in the region where $\mu > 1$ compared to its sticky counterpart, largely due to a much tricker nonlinearity caused by $a(t) = \sum_{n \geq 2} n \, p_n$ in equation (1) (this nonlinearity is more benign in the mean-field sticky dispersion model). For these reasons, Section 4 constitutes the most technical part of this manuscript and requires significant novel ideas from the theory of PDEs.

A related particle system is the Becker–Döring model [2, 3, 4, 31, 39, 40], which describes the evolution of the size distribution of clusters in a nucleation process. Each cluster may absorb or emit one particle at a rate depending on its size. Let X_i be the number of particles in the i-th cluster. Then, the update rule is

$$(X_i, X_j) \xrightarrow{\text{rate } a_{X_i}} (X_i + 1, X_j - 1) \qquad \text{if } X_i > 1, X_j = 1$$

$$(X_i, X_j) \xrightarrow{\text{rate } b_{X_i}} (X_i - 1, X_j + 1) \qquad \text{if } X_i > 1, X_j = 0.$$

Here, (i, j) should be sampled from the set of all pairs of non-empty clusters and the first empty cluster, and the cluster-dependent jump rates $a_n, b_n \geq 0$ depend on the specific set-up of the model, which we omit to describe at this moment. We will elaborate on the Becker–Döring model in Section 2.

Another example is the poor-biased exchange model [8, 17] studied in econophysics. It can be viewed as a variant of the dispersion process where every particle is always active, disregarding the number of particles on a site. The update rule is as follows:

$$(X_i, X_j) \xrightarrow{\text{rate } X_i} (X_i - 1, X_j + 1) \qquad \text{if } X_i > 0. \tag{7}$$

In this and many other related econophysics models [5, 6, 7, 16, 18, 20, 30, 36, 37, 38], particles are currencies which are exchanged by agents represented by nodes. Each agent passes a currency to another agent at a rate proportional to the amount of currency they have. Our model can be viewed as a variant of the poor-biased exchange model with a wealth-flooring policy, where agents with less than 2 currencies are not allowed to exchange.

We refer interested readers to [19, 24, 25, 34, 42, 43] for more information on econophysics. The aforementioned propogation of chaos property was proved for many econophysics models [8, 9, 10, 13, 17, 22].

1.3. **Numerical simulations.** To verify the result, we perform two agent-based simulations on $N = 1000$ sites, with 800 and 2000 particles respectively. We used two different initial values. For $M = 800$, all particles are initially placed at one single site; for $M = 2000$, they are evenly distributed on the graph at the beginning. We record the number of particles in each site X_i over time. The distributions of X_i in the first 10 units of time are plotted in Figure 2. We observe that for the underpopulated scenario $\mu = 0.8$, the distribution of the number of particles in a site converges to a Bernoulli distribution $\mathbf{p}^*$ defined in (3) with mean 0.8 after approximately 6.5 seconds (Figure 2-left). For the overpopulated case, it stabilizes near a zero-truncated Poisson distribution $\bar{\mathbf{p}}$ with mean 2 defined in (4) (Figure 2-right). To further confirm the convergence, we continue the simulation for 2000 time units, and the bar plot of X_i's is presented in Figure 3.
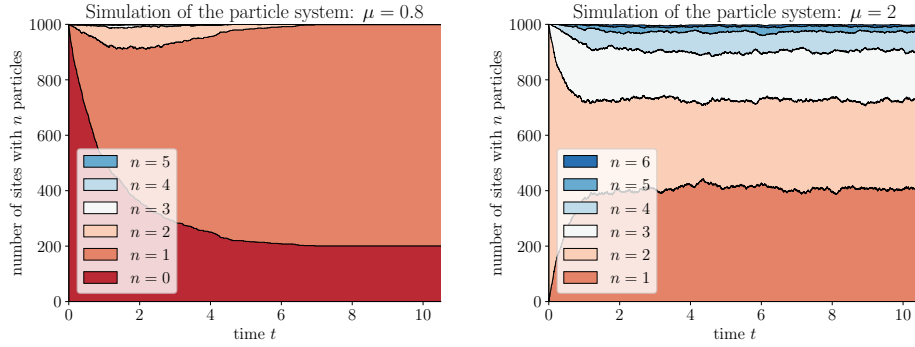
FIGURE 2. Stackplots of the agent-based simulation with $\mu = 0.8$ and $\mu = 2$ during the first 10 units of time. At a given time t, the width of n-th layer represents the number of sites hosting n particles, making a total of $N = 1000$ sites. For $\mu = 0.8$, we initially put all particles at one single site; the distribution of particles converges to a Bernoulli distribution and no site hosts two or more particles (left). For $\mu = 2$, we put two particles in each site initially; each site hosts at least one particle and the distribution of particles stabilizes around a zero-truncated Poisson distribution after a few units of time (right).
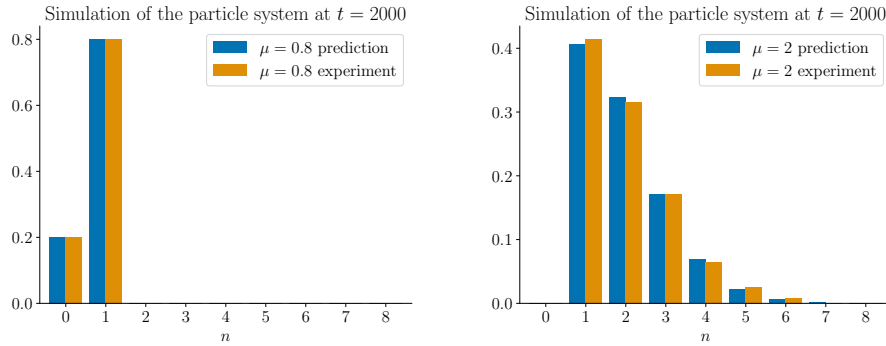

FIGURE 3. Distribution of particles for the dispersion model with $N = 1000$ agents after 2000 units of time, using two different values of μ. For $\mu = 0.8$, the final distribution coincides exactly with the Bernoulli distribution with mean 0.8, where we put all the particles into a single site initially. For $\mu = 2$, the terminal distribution is well-approximated by a zero-truncated Poisson distribution (4), where we put $X_i(0) = \mu$ for all $1 \leq i \leq N$ initially.

Next, we numerically solve the infinite system (2) by truncating the first 101 components (i.e. $(p_0(t), \ldots, p_{100}(t))$). We again choose the parameters $\mu = 2$ and $\mu = 0.8$. For $\mu = 0.8$, we use $p_{100}(0) = \mu/100$, $p_0(0) = 1 - \mu/100$, and $p_i(0) = 0$ on other components for the initial condition. For $\mu = 2$, we use $p_2(0) = 1$ and $p_i(0) = 0$ on other components. We apply the standard Runge–Kutta fourth-order scheme,

and the time step is set to $\Delta t = 0.01$. We plot the value of $p_n(t)$ as a stackplot in Figure 4, and we observe in both cases convergence to equilibrium distributions. Corresponding results for agent-based simulation in Figure 2 are plotted here in yellow for comparison. We find that the numerical solution to the ODE system (1) is in good agreement with the agent-based simulation results, which suggests (1) is a valid mean-field limit.
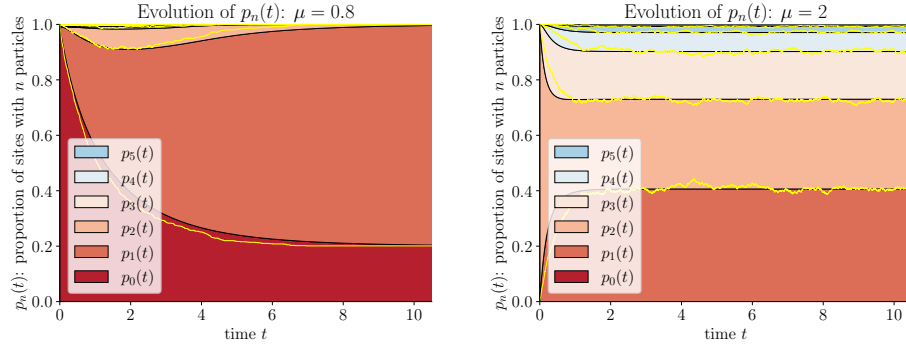

FIGURE 4. Stackplot of the numerical solution to the truncated ODE system $\{p_n(t)\}_{n=0}^{100}$ with $\mu = 0.8$ and $\mu = 2$ during the first 10 units of time. At a given time t, the width of n-th layer represents $p_n(t)$ which sum up to 1. For $\mu = 0.8$ the distibution of particles converges to a Bernoulli distribution with mean μ (left). For $\mu = 2$, the distribution converges to the zero-truncated Poisson distribution with mean μ (right). The overlayed yellow lines represent the corresponding agent-based simulation results.

1.4. **Organization of the paper.** In Section 2, we show some basic properties of system (1) and locate the equilibrium distribution. The form of the equilibrium depends on the parameter $\mu \in (0, \infty)$ which represents the average amount of particles per site initially. For the underpopulated case $\mu \in (0, 1]$, we prove the solution converges to the Bernoulli distribution $\mathbf{p}^*$ using a Lyapunov functional in Section 3. In Section 4, we analyze the overpopulated case $\mu > 1$ and show that the solution relaxes to the zero-truncated Poisson distribution $\bar{\mathbf{p}}$ at an exponential rate. The proof is based on the analysis of the probability generating function (PGF) of the solution, which satisfies a transport equation. Finally, we draw a conclusion in Section 5.

2. **Elementary properties of the ODE system.** After we achieved the transition from the interacting agents system (5) to the deterministic nonlinear ODE system (1), our main goal is to show convergence of solution of (1) to its (unique) equilibrium solution. We aim to describe some elementary properties of solutions of (1) in this section. As we have indicated in the introduction, the large time behavior of solutions to (1) depends critically on the range to which the parameter μ belongs. Before we dive into the detailed analysis of the system of nonlinear ODEs, we first establish some preliminary observations regarding solutions of (1).

As we want $a(t)$ to be finite, we equip $\mathcal{S}$ with the following weighted ℓ^1 norm:

$$\|\mathbf{q}\|_{\mathcal{S}} := \sum_{n=0}^{\infty} (1+n) |q_n|, \qquad \mathbf{q} \in \mathbb{R}^{\mathbb{N}}.$$

Clearly, $\|\mathbf{q}\|_{\ell^{\infty}} \leq \|\mathbf{q}\|_{\mathcal{S}}$. It can be seen from (2) that $\|\mathcal{F}[\mathbf{q}]\|_{\ell^{\infty}} \leq C \|\mathbf{q}\|_{\mathcal{S}}$. Therefore, we say $\mathbf{p} : [0, \infty) \to \mathcal{S}$ is a **(classical) solution** to system (1) if

$$\mathbf{p} \in C([0, \infty); \mathcal{S}) \cap C^1([0, \infty); \ell^{\infty})$$

which verifies $\mathbf{p}'(t) = \mathcal{F}[\mathbf{p}(t)]$.

Lemma 2.1. *If $\mathbf{p}(t)$ is a solution to system (1), then*

$$\sum_{n=0}^{\infty} \mathcal{F}[\mathbf{p}(t)]_n = 0 \quad and \quad \sum_{n=0}^{\infty} n\mathcal{F}[\mathbf{p}(t)]_n = 0. \tag{8}$$

In particular, the total probability mass and the average amount of particles per site are conserved.

The proof of Lemma 2.1 is based on straightforward computations and will be skipped. Thanks to these conservation relations, the solution $\mathbf{p}(t)$ lives in the space of probability distributions on $\mathbb{N}$ with the prescribed mean value μ, defined by

$$\mathcal{S}_{\mu} := \left\{ \mathbf{q} \in [0, 1]^{\mathbb{N}} : \mathbf{q} = \{q_n\}_{n=0}^{\infty} \text{ with } \sum_{n=0}^{\infty} q_n = 1, \sum_{n=0}^{\infty} nq_n = \mu \right\}. \tag{9}$$

We can use the Galerkin method (see [1, section 8.6]) to show the existence of a classical solution. The idea is to show the solutions to the truncated finite-dimensional ODEs are uniformly bounded in $C([0, T); \mathcal{S})$, so a weak limit (say in $L^2(0, T; \ell^2)$) can be extracted. One then verifies the weak limit is indeed a classical solution. As for uniqueness, we provide a proof in Appendix A for interested readers. More importantly, if the initial value has mean μ, then the number of active particles per site $a(t)$ is given by

$$a(t) = \sum_{n=2}^{\infty} np_n(t) = \mu - p_1(t).$$

System (1) is equivalent to the following system of nonlinear ODEs:

$$\mathbf{p}'(t) = \mathcal{D}_+[\{n\mathbf{1}_{\{n\geq 2\}}\}_n \mathbf{p}(t)] - \mathcal{D}_-[(\mu - p_1(t))\mathbf{p}(t)]. \tag{10}$$

Writing in components, we have $\mathbf{p}'(t) = \mathcal{F}[\mathbf{p}(t)]$, where the generator $\mathcal{F} : \mathcal{S}_{\mu} \to \mathbb{R}^{\mathbb{N}}$ is given by

$$\mathcal{F}[\mathbf{q}]_n = \begin{cases} -(\mu - q_1)q_0 & n = 0, \\ 2q_2 - (\mu - q_1)(q_1 - q_0) & n = 1, \\ (n+1)q_{n+1} - nq_n - (\mu - q_1)(q_n - q_{n-1}) & n \geq 2. \end{cases} \tag{11}$$

The Fokker–Planck type equation (10) admits a heuristic interpretation as a jump process with loss and gain, and we illustrate this perspective via Figure 5 below. Recall that $a(t) = \mu - p_1(t)$ is the number of active particles per site.

Remark 2.2. System (10) also resembles another system of nonlinear ODEs known as the Becker–Döring cluster equations. Following the convention, if p_n represents
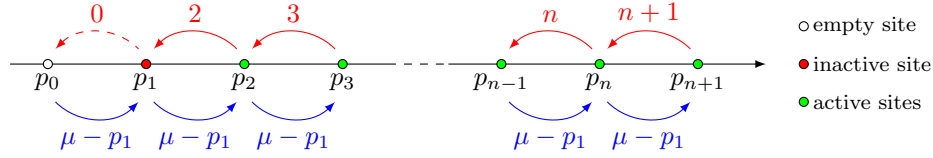
FIGURE 5. Schematic illustration of the Fokker–Planck type system of nonlinear ODEs (10) as a jump process with loss and gain.

the number of clusters with n particles, and μ is the total number of particles, then $\mathbf{p} = \{p_n\}_{n=1}^{\infty}$ satisfies $\mathbf{p}'(t) = \mathcal{BD}(\mathbf{p}(t))$ where the generator $\mathcal{BD} : \mathcal{S}_{\mu} \to \mathbb{R}^{\mathbb{N}^*}$ is

$$\mathcal{BD}[\mathbf{q}]_n = \begin{cases} (b_{n+1}q_{n+1} - b_n q_n) - q_1(a_n q_n - a_{n-1}q_{n-1}) & n \geq 2, \\ b_2 q_2 - a_1 q_1^2 + \sum_{n=1}^{\infty}(b_{n+1}q_{n+1} - q_1 a_n q_n) & n = 1. \end{cases}$$

Here, $\mathbf{a} = \{a_n\}_{n=1}^{\infty}$, $\mathbf{b} = \{b_n\}_{n=1}^{\infty}$ are given positive constants, with $b_1 = 0$. We can also write the Becker–Döring equations as a discrete Fokker–Planck type equation:

$$\mathbf{p}'(t) = \mathcal{D}_+[\mathbf{bp}(t)] - \mathcal{D}_-[p_1(t)\mathbf{ap}(t)] + \mathbf{f}(t),$$
$$\mathbf{f}(t) = ((\mathbf{b} - p_1(t)\mathbf{a}) \cdot \mathbf{p}, 0, 0, \dots).$$

Here, $\mathbf{bp}$ is the pointwise product, and "$\cdot$" denotes the inner product. Compared with (10), the generator of Becker–Döring equations $\mathcal{BD}$ is also a second-order difference operator; upward drift $\mathbf{b}$ is constant, whereas downward drift $p_1\mathbf{a}$ depends only on p_1. However, the total mass here is not conserved due to a quadratic reaction term $\mathbf{f}$, which makes Becker–Döring more difficult to analyze. Nevertheless, both $\mathcal{F}$ and $\mathcal{BD}$ are linear in $(q_2, q_3, \dots)$. We refer interested readers to [2, 4, 3, 31, 39, 40] and references therein.

Next, we identify the unique equilibrium solution associated with the system (10).

Proposition 2.3. *The unique equilibrium solution of* (10) *in the space* $\mathcal{S}_{\mu}$, *for* $\mu \in (0, 1]$, *is given by* $\mathbf{p}^*$ *defined in* (3). *The unique equilibrium solution of* (10) *in the space* $\mathcal{S}_{\mu}$, *when* $\mu \in (1, \infty)$ *is provided by* $\bar{\mathbf{p}}$, *defined in* (4).

Proof. To find equilibrium point of system (10), we need to solve $\mathcal{F}(\mathbf{q}) = 0$, which reduces to the following by induction:

$$(\mu - q_1)q_0 = 0, \tag{12}$$
$$(\mu - q_1)q_n = (n + 1)q_{n+1} \qquad \text{for } n \geq 1. \tag{13}$$

On the one hand, if $\mu = q_1 \leq 1$, then by (13), $q_n = 0$ for all $n \geq 2$, and we deduce that the unique equilibrium solution is $\mathbf{q} = \mathbf{p}^*$ with

$$p_0^* = 1 - \mu, \qquad p_1^* = \mu, \qquad p_n^* = 0 \qquad \text{for } n \geq 2.$$

Note that in the case $\mu = 1$, the Bernoulli distribution (3) degenerates to the Dirac delta distribution centered at 1. On the other hand, for $\mu > 1 \geq p_1$, we deduce from (12) that $p_0 = 0$, and the unique equilibrium distribution is $\mathbf{q} = \bar{\mathbf{p}}$ with

$$\bar{p}_0 = 0, \qquad \bar{p}_n = \frac{(\mu - \bar{p}_1)^{n-1}}{n!}\bar{p}_1 \qquad \text{for } n \geq 1$$

where $\bar{p}_1 > 0$ is chosen such that $\bar{\mathbf{p}} \in \mathcal{S}_\mu$. Since $\sum_{n=0}^{\infty} \bar{p}_n = 1$, we deduce that $(\mu - \bar{p}_1)/\bar{p}_1 = \sum_{n=1}^{\infty} (\mu - \bar{p}_1)^n/n! = \mathrm{e}^{\mu - \bar{p}_1} - 1$, and thus $\bar{p}_1 \, \mathrm{e}^{-\bar{p}_1} = \mu \, \mathrm{e}^{-\mu}$, from which we deduce $\bar{p}_1 = -W_0(-\mu \, \mathrm{e}^{-\mu})$ because $\bar{p}_1 < \mu$. We finish the proof by introducing a new constant $\nu = \mu - \bar{p}_1$. $\hfill\square$

Remark 2.4. The zero-truncated Poisson distribution $\bar{\mathbf{p}}$ defined in (4) admits a simple interpretation in terms of random variables. Indeed, if $X \sim \mathrm{Poisson}(\nu)$, then the distribution of X conditioned on $X \geq 1$ obeys the zero-truncated Poisson distribution, whose law is given by $\bar{\mathbf{p}}$.

3. **Underpopulated case: Convergence to the Bernoulli distribution.** To justify the large time convergence of solutions of system (10) when $\mu \in (0, 1]$, we employ a suitable Lyapunov functional associated to the dynamics (10). For this purpose, we define the following energy functional:

$$\mathcal{E}[\mathbf{q}] \coloneqq \sum_{n \geq 0} n^2 q_n - \mu \qquad \mathbf{q} \in \mathcal{S}_\mu, \tag{14}$$

which is just a shifted version of the second (raw) moment of the distribution $\mathbf{q}$. We first start with an elementary variational characterization of the Bernoulli distribution $\mathbf{p}^*$.

Lemma 3.1. *For each $\mu \in (0, 1]$, the Bernoulli distribution $\mathbf{p}^*$ with parameter μ satisfies*

$$\mathbf{p}^* = \operatorname*{argmin}_{\mathbf{q} \in \mathcal{S}_\mu} \sum_{n \geq 0} n^2 q_n. \tag{15}$$

Consequently, $\mathcal{E}[\mathbf{q}] \geq 0$ for all $\mathbf{q} \in \mathcal{S}_\mu$, and the equality holds if and only if $\mathbf{q} = \mathbf{p}^*$.

Proof. For $\mathbf{q} \in \mathcal{S}_\mu$, we have $\sum_{n \geq 0} n q_n = \mu$, and thus $\sum_{n \geq 1} n^2 q_n \geq \sum_{n \geq 1} n q_n = \mu$, in which the inequality will become an equality if and only if $q_n = 0$ for all $n \geq 2$. This finishes the proof of Lemma 3.1. $\hfill\square$

We now prove the following quantitative convergence result for the dissipation of $\mathcal{E}[\mathbf{p}(t)]$ along solutions to the system of nonlinear ODEs (10) when $\mu \in (0, 1]$.

Theorem 3.2. *Assume that $\mathbf{p}(t)$ is the classical solution to system (10) with $\mathbf{p}(0) \in \mathcal{S}_\mu$ and $\mu \in (0, 1]$. Then, for all $t \geq 0$, we have*

$$\mathcal{E}[\mathbf{p}(t)] \leq \mathcal{E}[\mathbf{p}(0)] \, \mathrm{e}^{-2(1-\mu)t} \qquad \text{when } \mu < 1, \tag{16}$$

and

$$\mathcal{E}[\mathbf{p}(t)] \leq \mathcal{E}[\mathbf{p}(0)] \, \mathrm{e}^{-2t} + \frac{4}{t + 2/p_0(0)} + 2p_0(0) \, \mathrm{e}^{-t} \qquad \text{when } \mu = 1. \tag{17}$$

Proof. Assume $\mathcal{E}[\mathbf{p}(0)] < +\infty$, otherwise the conclusion holds trivially. A straightforward computation gives us

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}[\mathbf{p}] &= 2p_2 + (\mu - p_1)(p_0 - p_1) \\
&\quad + \sum_{n \geq 2} n^2 \left[(n+1)p_{n+1} - np_n - (\mu - p_1)(p_n - p_{n-1})\right] \\
&= 2p_2 + (\mu - p_1)(p_0 - p_1) \\
&\quad + \sum_{n \geq 2}\left[n^2(n+1)p_{n+1} - n^3 p_n\right] - (\mu - p_1)\sum_{n \geq 2} n^2(p_n - p_{n-1}) \\
&= 2p_2 + (\mu - p_1)(p_0 - p_1) \\
&\quad + \left(p_1 - 2p_2 + \mu - 2\sum_{n \geq 0} n^2 p_n\right) - (\mu - p_1)(p_0 - p_1 - 1 - 2\mu) \\
&= p_1 + \mu - 2\sum_{n \geq 0} n^2 p_n + (\mu - p_1)(1 + 2\mu) \\
&= -2\mathcal{E}[\mathbf{p}] + 2\mu(\mu - p_1).
\end{aligned}
\tag{18}
$$

We observe that $0 \leq \mu - p_1 = \sum_{n \geq 2} np_n \leq \sum_{n \geq 2}(n^2 - n)p_n = \mathcal{E}[\mathbf{p}]$ for all $\mathbf{p} \in \mathcal{S}_\mu$, thus for $\mu \in (0, 1)$, we deduce from (18) that

$$
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}[\mathbf{p}] \leq -(2 - 2\mu)\mathcal{E}[\mathbf{p}],
$$

from which the exponential decay of $\mathcal{E}[\mathbf{p}(t)]$ (16) follows immediately. On the other hand, for $\mu = 1$, we have $2p_0 + p_1 \geq 1$ since $\mathbf{p} \in \mathcal{S}_1$. The differential equation satisfied by p_0 implies that $p_0' = -(1 - p_1)p_0 \leq -p_0^2$, whence

$$
p_0(t) \leq \frac{1}{t + 1/p_0(0)} \quad \text{and} \quad p_1(t) \geq 1 - \frac{2}{t + 1/p_0(0)}
$$

for all $t \geq 0$. Consequently, we derive from (18) the following differential inequality:

$$
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}[\mathbf{p}] \leq -2\mathcal{E}[\mathbf{p}] + \frac{4}{t + 1/p_0(0)},
$$

whence

$$
\mathcal{E}[\mathbf{p}(t)] \leq \mathcal{E}[\mathbf{p}(0)]\,\mathrm{e}^{-2t} + 4\,\mathrm{e}^{-2t}\int_0^t \frac{\mathrm{e}^{2s}}{s + 1/p_0(0)}\,\mathrm{d}s.
\tag{19}
$$

To conclude the proof and reach the advertised upper bound (17), it suffices to notice that

$$
\begin{aligned}
\int_0^t \frac{\mathrm{e}^{2s}}{s + 1/p_0(0)}\,\mathrm{d}s &= \int_0^{\frac{t}{2}} \frac{\mathrm{e}^{2s}}{s + 1/p_0(0)}\,\mathrm{d}s + \int_{\frac{t}{2}}^t \frac{\mathrm{e}^{2s}}{s + 1/p_0(0)}\,\mathrm{d}s \\
&\leq p_0(0)\int_0^{\frac{t}{2}} \mathrm{e}^{2s}\,\mathrm{d}s + \frac{1}{\frac{t}{2} + \frac{1}{p_0(0)}}\int_{\frac{t}{2}}^t \mathrm{e}^{2s}\,\mathrm{d}s \\
&\leq \frac{1}{2}p_0(0)\,\mathrm{e}^t + \frac{\mathrm{e}^{2t}}{t + 2/p_0(0)}
\end{aligned}
$$

for all $t \geq 0$. Thus, the proof of Theorem 3.2 is completed. $\qquad\square$

To illustrate the decay of the energy $\mathcal{E}$ numerically, we use the same set-up as in the previous experiment shown in Figure 6 for two different values of $\mu \in (0, 1]$ using the semi-log scale.
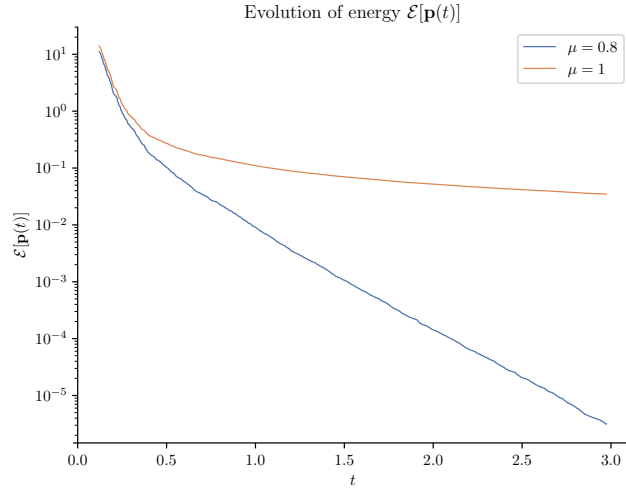

FIGURE 6. Evolution of the energy $\mathcal{E}[\mathbf{p}(t)]$ over $0 \le t \le 3$ with $\mu = 0.8$ and $\mu = 1$. It can be seen from the picture that the energy decays exponentially for $\mu = 0.8$ with rate $C\,\mathrm{e}^{-0.4t}$. For $\mu = 1$, the decay is slower.

As an immediate corollary, we can readily deduce the following strong convergence in ℓ^1.

Corollary 3.3. *Under the settings of Theorem 3.2, if $\mathbf{p}(0)$ has a finite variance, then there exists some constant $C > 0$ depending only on μ and the initial datum $\mathbf{p}(0)$ such that for all $t > 0$, it holds that*

$$\|\mathbf{p}(t) - \mathbf{p}^*\|_{\ell^1} \le C\,\mathrm{e}^{-2(1-\mu)t} \qquad when \ \mu < 1, \tag{20}$$

and

$$\|\mathbf{p}(t) - \mathbf{p}^*\|_{\ell^1} \le \frac{C}{t} \qquad when \ \mu = 1. \tag{21}$$

Proof. We only prove the bound (21) when $\mu = 1$ as the other bound (20) which is valid for $\mu \in (0, 1)$ can be handled in a pretty similar way. Notice that

$$\mathcal{E}[\mathbf{p}(t)] = \sum_{n \ge 0}(n^2 - n)p_n(t) = \sum_{n \ge 2}(n^2 - n)p_n(t)$$

and $n^2 \le 2(n^2 - n)$ for all $n \ge 2$, hence $\sum_{n \ge 2} n^2 p_n(t) \le 2\mathcal{E}[\mathbf{p}(t)]$. Therefore, we deduce that

$$\|\mathbf{p}(t) - \mathbf{p}^*\|_{\ell^1} = |p_0(t)| + |1 - p_1(t)| + \sum_{n \ge 2}|p_n(t)|$$

$$= 1 - 2p_1(t) + \sum_{n=0}^{\infty}p_n(t) = 2(1 - p_1(t)).$$

Moreover, from the definition of $\mathcal{E}$ we know $\mathcal{E}[\mathbf{p}] = p_1 + \sum_{n \geq 2} n^2 p_n - \mu$, so

$$1 - p_1(t) = \sum_{n \geq 2} n^2 p_n(t) - \mathcal{E}[\mathbf{p}(t)] \leq \mathcal{E}[\mathbf{p}(t)] \leq \frac{C}{t}$$

for some constant $C > 0$ depending on $\mathbf{p}(0)$ and μ. $\qquad \square$

We comment here that the proof of Corollary 3.3 implies an inequality of the form $\|\mathbf{p}(t) - \mathbf{p}^*\|_{\ell^1} \leq C\,\mathcal{E}[\mathbf{p}]$ for some generic constant $C > 0$, which is reminiscent of a Csiszár–Kullback–Pinsker type inequality in the study of certain convex entropy-like Lyapunov functionals (see for instance [33]).

Remark 3.4. It is worth mentioning that the Lyapunov functional (or "energy") $\mathcal{E}[\mathbf{q}]$ employed in the large time analysis of system (10) when $\mu \in (0, 1]$ differs drastically from the Lyapunov functional constructed in [15] for the large time analysis of the mean-field sticky dispersion model when $\mu \in (0, 1]$. In fact, the Lyapunov functional discovered in the study of the mean-field sticky dispersion model [15] in the region where $\mu \in (0, 1]$ is the so-called Gini index, which can be utilized to prove convergence in terms of the Wasserstein distance W_1 [12] (which is certainly weaker than the ℓ^1 convergence reported here). To some extent, it demonstrates the necessity to significantly modify the appropriate "energy", and even the underlying mean-field dynamical systems appear to be similar from a structural perspective.

4. **Overpopulated case: Relaxation to zero-truncated Poisson distribution.** We use a different approach to study system (10) when $\mu > 1$. Treating the active particles per site $a(t)$ as a known function, we first show that the probability generating function solves a first-order partial differential equation (PDE), which turns out to be an explicitly solvable transport equation. By writing out the solution, we show that an auxiliary function

$$v(t) := \exp\left(\int_0^t \mathrm{e}^{-t+s}\, a(s)\, \mathrm{d}s \right) \tag{22}$$

must satisfy a nonlinear Volterra-type integral equation [29]. We then study this integral equation to extract the convergence and the convergence rate, which further sheds light on the convergence of the distribution $\mathbf{p}(t)$ to the zero-truncated Poisson distribution $\bar{\mathbf{p}}$ (4).

4.1. **Probability generating function.** Define the probability generating function $G : [0, +\infty) \times [-1, 1]$ of the solution $\mathbf{p}(t)$ to (1) by

$$G(t, z) = \sum_{n=0}^{\infty} p_n(t) z^n.$$

Since $p_n(t) \geq 0$ and $\sum_{n=0}^{\infty} p_n(t) = 1$, we know the above series is absolutely summable. Moreover, because $\sum_{n=0}^{\infty} n p_n(t) = \mu$, we know that

$$\frac{\partial G}{\partial z}(t, z) = \sum_{n=1}^{\infty} n p_n(t) z^{n-1}$$

is absolutely summable. The ODE system (1) can thus be written as the following PDE for G:

$$\partial_t G = (1 - z)[\partial_z G - p_1(t) - a(t)G] = (1 - z)[\partial_z G - \mu + a(t) - a(t)G]. \tag{23}$$

Recall that $a(t) = \mu - p_1(t)$. We also recall that the probability generating function can recover the following statistics:

$$\frac{\partial^k}{\partial z^k} G(t, 0) = k! p_k(t)$$

$$\frac{\partial^k}{\partial z^k} G(t, 1) = \sum_{n=0}^{\infty} n(n-1)\cdots(n-k+1) p_n(t).$$

Moreover, since $p_k(t) \geq 0$ for all k, we have monotonicity in all derivatives:

$$\frac{\partial^k}{\partial z^k} G(t, 0) < \frac{\partial^k}{\partial z^k} G(t, z) < \frac{\partial^k}{\partial z^k} G(t, 1), \qquad (t, z) \in [0, \infty) \times (0, 1).$$

We first solve (23) in terms of $p_1(t)$ using the method of characteristics [27].

Lemma 4.1. *The probability generating function G can be expressed using the following explicit formula: For $z \in \mathbb{C}$ with $|1 - z| \leq 1$, $t \geq 0$, we have*

$$G(t, 1 - z) = 1 + \left(G(0, 1 - z\,\mathrm{e}^{-t}) - 1 - \mu z \int_0^t [v(s)]^{z\,\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s \right) [v(t)]^{-z},$$

where $v : [0, \infty) \to \mathbb{R}$ is defined by (22).

Proof. We only prove for $z \in \mathbb{R}$ with $0 \leq z \leq 2$. This will be sufficient because both sides are analytic in z and the equality follows by the identity theorem.

Define $\gamma(s) = 1 - z\,\mathrm{e}^{-t+s}$. Then,

$$\gamma(0) = 1 - z\,\mathrm{e}^{-t}, \qquad \gamma(t) = 1 - z, \qquad \gamma'(s) = -z\,\mathrm{e}^{-t+s} = -(1 - \gamma(s)).$$

Now we let $g(s) = G(s, \gamma(s)) - 1$, and thus the evolution of g satisfies

$$\begin{aligned}
g'(s) &= \partial_t G(s, \gamma(s)) + \gamma'(s)\partial_z G(s, \gamma(s)) \\
&= [\partial_t - (1 - \gamma(s))\partial_z] G(s, \gamma(s)) \\
&= -(1 - \gamma(s))[a(s)G(s, \gamma(s)) - a(t) + \mu] \\
&= -z\,\mathrm{e}^{-t+s}[a(s)g(s) + \mu],
\end{aligned}$$

with

$$g(0) = G(0, \gamma(0)) - 1 = G(0, 1 - z\,\mathrm{e}^{-t}) - 1.$$

So, g satisfies the following first-order linear ODE with the above initial condition:

$$g'(s) + z\,\mathrm{e}^{-t+s}\,a(s)g(s) = -\mu z\,\mathrm{e}^{-t+s}.$$

Setting

$$H(t) := \int_0^t \mathrm{e}^s\,a(s)\,\mathrm{d}s, \tag{24}$$

we have $v(t) = \mathrm{e}^{\mathrm{e}^{-t} H(t)}$ and

$$\begin{aligned}
g'(s) + z\,\mathrm{e}^{-t}\,H'(s)g(s) &= -\mu z\,\mathrm{e}^{-t+s}, \\
\frac{\mathrm{d}}{\mathrm{d}s}\left(g(s)\,\mathrm{e}^{z\,\mathrm{e}^{-t} H(s)} \right) &= -\mu z\,\mathrm{e}^{-t+s}\,\mathrm{e}^{z\,\mathrm{e}^{-t} H(s)}, \\
\frac{\mathrm{d}}{\mathrm{d}s}\left(g(s)[v(s)]^{z\,\mathrm{e}^{-t+s}} \right) &= -\mu z[v(s)]^{z\,\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}.
\end{aligned}$$

Integrating from $s = 0$ to $s = t$ and using $v(0) = 1$, we deduce that

$$g(t)[v(t)]^z = g(0) - \mu z \int_0^t [v(s)]^{z\,\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s.$$

Finally, noticing that $g(t) = G(t, \gamma(t)) - 1 = G(t, 1 - z) - 1$ and $g(0) = G(0, 1 - z\,\mathrm{e}^{-t}) - 1$, we conclude the proof of Lemma 4.1. $\square$

Remark 4.2. When $z = 0$, we have $G(t, 1) = G(0, 1) = 1$ for all $t \geq 0$. This can be seen from (23), which shows $\partial_t G(t, 1) = 0$, so the total mass is conserved. We can similarly derive conservation of the first moment by $\partial_t \partial_z G(t, 1) = -\partial_z G(t, 1) + \mu$ from (23).

4.2. **Convergence of the auxiliary function.** In this subsection, we first show that the auxiliary v function satisfies an integral equation, and then prove it converges to a limit which depends on the value of μ.

Lemma 4.3. *For $t \geq 0$, $v(t)$ satisfies*

$$v(t) = 1 - f_0(t) + f_0(0)\,\mathrm{e}^{-\int_0^t \log v(s)\,\mathrm{d}s} + \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s,$$

where $f_0(t) := G(0, 1 - \mathrm{e}^{-t})$.

Proof. We recall that $G(t, 0) = p_0(t)$. Applying Lemma 4.1 with $z = 1$, we obtain

$$G(t, 0) = 1 + \left(G(0, 1 - \mathrm{e}^{-t}) - 1 - \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s \right) [v(t)]^{-1}$$

$$\implies v(t) = v(t)p_0(t) + 1 - f_0(t) + \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s. \qquad (25)$$

It remains to solve for $p_0(t)$ in terms of v. Using the differential equation satisfied by $p_0(t)$, we know that

$$p_0(t) \exp\left(\int_0^t \mu - p_1(s)\,\mathrm{d}s \right) = p_0(0) = G(0, 0) = f_0(0). \qquad (26)$$

In view of (24), we can solve $p_0(t)$ in terms of v by

$$a(s) = \mathrm{e}^{-s}\,H'(s)$$

$$\implies \int_0^t a(s)\,\mathrm{d}s = \mathrm{e}^{-t}\,H(t) + \int_0^t H(s)\,\mathrm{e}^{-s}\,\mathrm{d}s = \log v(t) + \int_0^t \log v(s)\,\mathrm{d}s$$

$$\implies p_0(t) = f_0(0) \exp\left(-\int_0^t a(s)\,\mathrm{d}s \right) = f_0(0)[v(t)]^{-1}\,\mathrm{e}^{-\int_0^t \log v(s)\,\mathrm{d}s}.$$

Combining this with (25), we conclude the proof. $\square$

We now investigate the limiting behavior of $v(t)$ as $t \to \infty$. First, since $0 \leq p_1(s) \leq 1$ for all $s \in [0, t]$, we can bound v using its definition (22):

$$(\mu - 1) \int_0^t \mathrm{e}^{-t+s}\,\mathrm{d}s \leq \log v(t) \leq \mu \int_0^t \mathrm{e}^{-t+s}\,\mathrm{d}s.$$

By direct computations, we obtain the following bound:

$$(\mu - 1)(1 - \mathrm{e}^{-t}) \leq \log v(t) \leq \mu(1 - \mathrm{e}^{-t}). \qquad (27)$$

In particular, $1 \leq v(t) \leq \mathrm{e}^{\mu}$.

Next, we control the nonlinear integral term using the following lemma.

Lemma 4.4. *Let $t_0 \geq 0$. If $1 \leq m \leq v(t) \leq M$ for all $t \geq t_0$, then*

$$\varphi(m) - e_1(t) \leq \mu \int_0^t [v(s)]^{e^{-t+s}} e^{-t+s} \, ds \leq \varphi(M) + e_1(t), \qquad \forall t \geq t_0,$$

where $\varphi : (0, +\infty) \to \mathbb{R}_+$ is a strictly increasing continuous function defined by

$$\varphi(x) := \mu \int_0^\infty x^{e^{-t}} e^{-t} \, dt = \begin{cases} \mu \cdot \dfrac{x-1}{\log x}, & x > 0, x \neq 1 \\ \mu, & x = 1 \end{cases}$$

and the remainder term is $e_1(t) := \mu \, e^{\mu + t_0} \, e^{-t}$.

Proof. First, we separate our target integral as follows:

$$\mu \int_0^t [v(s)]^{e^{-t+s}} e^{-t+s} \, ds = \mu \int_0^{t_0} [v(s)]^{e^{-t+s}} e^{-t+s} \, ds + \mu \int_{t_0}^t [v(s)]^{e^{-t+s}} e^{-t+s} \, ds.$$

The first term has an exponential decay since

$$\mu \int_0^{t_0} [v(s)]^{e^{-t+s}} e^{-t+s} \, ds \leq \mu \int_0^{t_0} e^{\mu \, e^{-t+s}} e^{-t+s} \, ds \leq \mu \, e^\mu \, e^{-t+t_0}.$$

The second term is bounded from above by

$$\mu \int_{t_0}^t [v(s)]^{e^{-t+s}} e^{-t+s} \, ds \leq \mu \int_{t_0}^t M^{e^{-t+s}} e^{-t+s} \, ds = \mu \int_0^{t-t_0} M^{e^{-s}} e^{-s} \, ds \leq \varphi(M)$$

and from below by

$$\mu \int_{t_0}^t [v(s)]^{e^{-t+s}} e^{-t+s} \, ds \geq \mu \int_{t_0}^t m^{e^{-t+s}} e^{-t+s} \, ds = \mu \int_0^{t-t_0} m^{e^{-s}} e^{-s} \, ds$$

$$= \varphi(m) - \mu \int_{t-t_0}^\infty m^{e^{-s}} e^{-s} \, ds \geq \varphi(m) - \mu m \, e^{-t+t_0}.$$

The proof is thus completed because $m \leq v(t) \leq e^\mu$ and $\mu m \, e^{-t+t_0} \leq e_1(t)$. $\quad\square$

We are now ready to prove the convergence of $v(t)$.

Lemma 4.5. *Let $\nu = \mu + W_0(-\mu \, e^{-\mu})$. Then,*

$$\lim_{t \to +\infty} v(t) = e^\nu.$$

Proof. Denote

$$e_2(t) = 1 - f_0(t) + f_0(0) \exp\left(-\int_0^t \log v(s) \, ds\right). \tag{28}$$

Since

$$0 \leq 1 - f_0(t) = 1 - G(0, 1 - e^{-t}) \leq \partial_z G(0,1) \, e^{-t} = \mu \, e^{-t},$$

$$\int_0^t \log v(s) \, ds \geq \int_0^t (\mu - 1)(1 - e^{-s}) \, ds = (\mu - 1)(t + e^{-t} - 1) \geq (\mu - 1)(t - 1),$$

we deduce that

$$e_2(t) \leq \mu \, e^{-t} + f_0(0) \, e^{-(\mu-1)(t-1)} \to 0 \qquad \text{as } t \to +\infty. \tag{29}$$

With the above error function, we can write the integral equation in Lemma (4.3) as

$$v(t) = e_2(t) + \mu \int_0^t [v(s)]^{e^{-t+s}} e^{-t+s} \, ds. \tag{30}$$

Since $v(t) \in [1, e^\mu]$ is bounded (uniformly in time), $e_2(t) \to 0$ as $t \to +\infty$, the liminf and limsup of v exist and are bounded between

$$m := \liminf_{t \to +\infty} v(t) \geq \lim_{t \to +\infty} \exp\left((\mu - 1)(1 - e^{-t})\right) = e^{\mu - 1},$$

$$M := \limsup_{t \to +\infty} v(t) \leq \lim_{t \to +\infty} \exp\left(\mu(1 - e^{-t})\right) = e^\mu.$$

We claim that $\varphi(m) \leq m \leq M \leq \varphi(M)$. To prove this, we first fix $\varepsilon > 0$ with $\varepsilon < e^{\mu - 1} - 1$. Then, there exists $t_0 > 0$ such that

$$1 < m - \varepsilon \leq v(t) \leq M + \varepsilon, \qquad \forall t \geq t_0.$$

Invoking Lemma 4.4 together with the monotonicity of the map φ, we conclude for all $t \geq t_0$ that

$$\varphi(m - \varepsilon) - e_1(t) \leq \mu \int_0^t [v(s)]^{e^{-t+s}} e^{-t+s} \, ds \leq \varphi(M + \varepsilon) + e_1(t), \qquad \forall t \geq t_0.$$

Therefore, the limsup and liminf of v are bounded by

$$M = \limsup_{t \to +\infty} v(t) \leq \lim_{t \to +\infty} e_2(t) + \varphi(M + \varepsilon) + \lim_{t \to +\infty} e_1(t) = \varphi(M + \varepsilon),$$

$$m = \liminf_{t \to +\infty} v(t) \geq \lim_{t \to +\infty} e_2(t) + \varphi(m - \varepsilon) - \lim_{t \to +\infty} e_1(t) = \varphi(m - \varepsilon).$$

This is true for any sufficiently small ε, so the advertised claim $\varphi(m) \leq m \leq M \leq \varphi(M)$ is justified.

Within the interval $[e^{\mu - 1}, e^\mu]$, we demonstrate that the function φ is a contraction mapping. Indeed, note that

$$\varphi(e^{\mu - 1}) = \mu \cdot \frac{e^{\mu - 1} - 1}{\mu - 1} = e^{\mu - 1} + \frac{e^{\mu - 1} - \mu}{\mu - 1} > e^{\mu - 1},$$

$$\varphi(e^\mu) = \mu \cdot \frac{e^\mu - 1}{\mu} = e^\mu - 1 < e^\mu.$$

Moreover, the derivatives of φ are

$$\varphi'(e^x) = \mu x^{-2} e^{-x}(1 - e^x + x e^x) > 0,$$

$$\varphi''(e^x) = \mu x^{-3} e^{-x}(2 - 2 e^{-x} - x e^{-x} - x) < 0.$$

So, for $x \in [\mu - 1, \mu]$, we know $\varphi'(e^x)$ is bounded by

$$0 < \varphi'(e^\mu) \leq \varphi'(e^x) \leq \varphi'(e^{\mu - 1}) = \frac{\mu^2 - 2\mu + \mu e^{1-\mu}}{\mu^2 - 2\mu + 1} := L_\mu < 1.$$

Since $\varphi : [e^{\mu - 1}, e^\mu] \to [e^{\mu - 1}, e^\mu]$ is a contraction mapping and $[m, M]$ is contained in this interval, we conclude from $\varphi(m) \leq m \leq M \leq \varphi(M)$ that $m = M = e^\nu$, where e^ν is the unique fixed point of φ in $[e^{\mu - 1}, e^\mu]$. It satisfies

$$e^\nu = \mu \frac{e^\nu - 1}{\nu} \iff \nu e^\nu = \mu e^\nu - \mu \iff (\nu - \mu) e^{\nu - \mu} = -\mu e^{-\mu}.$$

We observe that $\nu - \mu$ and $-\mu$ are two real roots to the equation $z e^z = -\mu e^{-\mu}$, hence we can use the Lambert W function to select the principal branch and arrive at the relation $\nu - \mu = W_0(-\mu e^{-\mu})$. This finishes the proof of Lemma 4.5. $\qquad\square$

Remark 4.6. The convergence of v implies the pointwise convergence of the probability generating function. Indeed, sending $t \to +\infty$ in Lemma 4.1 and applying the result of Lemma 4.4 (with v being replaced by v^z), we obtain

$$\lim_{t \to +\infty} G(t, 1 - z) = 1 + (G(0, 1) - 1 - z\varphi(e^{\nu z})) e^{-\nu z}$$

$$= 1 - \mu \cdot \frac{1 - e^{-\nu z}}{\nu} = \frac{\mu - \nu}{\nu} \left(e^{\nu(1-z)} - 1 \right) = G_{\bar{\mathbf{p}}}(1-z),$$

where $G_{\bar{\mathbf{p}}}$ denotes the PGF of the zero-truncated Poisson distribution $\bar{\mathbf{p}}$ (4). According to a classical result [26], if the PGF exists and the above convergence holds in a neighborhood of $z = 0$, then $\mathbf{p}(t)$ converges to $\bar{\mathbf{p}}$ in distribution. However, in the following we will strengthen the results obtained so far and prove ℓ^p convergence for $p = 1, 2$, which are stronger convergence guarantees.

4.3. **Quantitative convergence of the auxiliary function.** Next, we study the quantitative rate of convergence of $v(t) \xrightarrow{t \to \infty} e^\nu$. We do it in two steps. In the first step, we prove $\mathcal{O}(e^{-c\sqrt{t}})$ decay, and in the second step we refine the previous estimate to reach $\mathcal{O}(e^{-ct})$ decay. First, we show an improved lower bound.

Lemma 4.7. *There exists $T_0 = T_0(\mu) > 0$ such that $v(t) \geq e^{\mu-1}$ for all $t \geq T_0$.*

Proof. For some $t_0 > 0$ to be determined, since $v(t) \geq e^{(\mu-1)(1-e^{-t})} \geq e^{(\mu-1)(1-e^{-t_0})}$ for all $t \geq t_0$, by Lemma 4.4 we know that

$$v(t) \geq e_2(t) + \varphi\left(e^{(\mu-1)(1-e^{-t_0})} \right) - e_1(t) \geq \varphi\left(e^{(\mu-1)(1-e^{-t_0})} \right) - \mu e^{\mu+t_0} e^{-t}$$

holds for all $t \geq t_0$. In particular, for all $t \geq 2t_0$, we obtain

$$v(t) \geq \varphi\left(e^{(\mu-1)(1-e^{-t_0})} \right) - \mu e^{\mu-t_0} \to \varphi(e^{\mu-1}) \qquad \text{as } t_0 \to +\infty.$$

Since $\varphi(e^{\mu-1}) > e^{\mu-1}$ (recall the proof of Lemma 4.5), we can find $T_0 := 2t_0$ for some sufficiently large t_0 depending on μ such that $v(t) \geq e^{\mu-1}$ for all $t \geq T_0$. $\qquad\square$

Lemma 4.8. *There exist constants $C, \delta > 0$ depending only on μ such that*

$$|v(t) - e^\nu| \leq C e^{-\sqrt{\delta t}}, \qquad \forall t > 0.$$

Proof. For $t \geq 0$, we set

$$r(t) := \sup_{s \geq t} |v(s) - e^\nu|,$$

which is a nonnegative and decreasing function.

Let $t_k \geq T_0$ be a sequence of increasing times to be determined later. Denote $m_k = e^\nu - r(t_k)$ and $M_k = e^\nu + r(t_k)$. Then, $m_k \leq v(t) \leq M_k$ for all $t \geq t_k$. Using Lemma 4.7, we deduce that $e^{\mu-1} \leq m_k \leq M_k \leq e^\mu$. By Lemma 4.4, we know that at any time $s \geq t_k$, it holds that

$$\varphi(m_k) - \mu e^{\mu+t_k-s} \leq v(s) - e_2(s) \leq \varphi(M_k) + \mu e^{\mu+t_k-s}.$$

In particular,

$$|v(s) - e^\nu| \leq \max\{e^\nu - \varphi(m_k), \varphi(M_k) - e^\nu\} + e_2(s) + \mu e^{\mu+t_k-s}.$$

Since φ is L_μ-Lipschitz in $[e^{\mu-1}, e^\mu]$, we have

$$\max\{e^\nu - \varphi(m_k), \varphi(M_k) - e^\nu\} \leq L_\mu r(t_k).$$

Now we take the supremum over $s \geq t_{k+1}$ to obtain the recursive inequality

$$r(t_{k+1}) \leq L_\mu r(t_k) + \sup_{s \geq t_{k+1}} e_2(s) + \mu e^{\mu+t_k-t_{k+1}}$$

$$\leq L_\mu r(t_k) + C e^{t_k-t_{k+1}} + e^{-(\mu-1)(t_{k+1}-1)},$$

in which $C > 0$ depends only on μ, and we employed the bound (29) for $e_2(s)$.

Denote $\delta = -\log L_\mu > 0$. For $k \geq 0$, we take $t_k := T_0 + k^2\delta$. Then,

$$\mathrm{e}^{t_k - t_{k+1}} = \mathrm{e}^{-(2k+1)\delta} = L_\mu^{2k+1},$$

$$\mathrm{e}^{-(\mu-1)(t_{k+1}-1)} \leq \mathrm{e}^{-(\mu-1)[(k+1)^2\delta - 1]} \leq C\,\mathrm{e}^{-(2k+1)\delta} = CL_\mu^{2k+1},$$

where $C > 0$ is chosen such that $\log C + (\mu - 1)[(k+1)^2\delta - 1] \geq (2k+1)\delta$ for all $k \geq 0$. The recursive inequality now becomes

$$r(t_{k+1}) \leq L_\mu r(t_k) + CL_\mu^{2k+1},$$

$$L_\mu^{-k-1} r(t_{k+1}) \leq L_\mu^{-k} r(t_k) + CL_\mu^k.$$

Taking the summation from 0 to $k-1$, we have

$$L_\mu^{-k} r(t_k) \leq r(t_0) + \frac{C}{1 - L_\mu} = C,$$

$$r(t_k) \leq CL_\mu^k = C\,\mathrm{e}^{-\sqrt{\delta(t_k - T_0)}} = C\,\mathrm{e}^{-\sqrt{\delta t_k}}.$$

Thus, we conclude by monotonicity that

$$r(t) \leq C\,\mathrm{e}^{-\sqrt{\delta t}}, \qquad t \geq T_0.$$

Finally, the restriction $t \geq T_0$ can be easily removed by taking C to be sufficiently large. $\qquad\square$

We can see that both the convergence and the above estimate are based on comparison. To obtain a sharper estimate, we take the difference

$$v(t) - \mathrm{e}^\nu = e_2(t) + \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s - \mu \int_0^\infty \mathrm{e}^{\nu\,\mathrm{e}^{-s}}\,\mathrm{e}^{-s}\,\mathrm{d}s$$

$$= e_2(t) + \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s$$

$$\quad - \mu \int_0^t \mathrm{e}^{\nu\,\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s - \mu \int_t^\infty \mathrm{e}^{\nu\,\mathrm{e}^{-s}}\,\mathrm{e}^{-s}\,\mathrm{d}s$$

$$= e_2(t) - \mathrm{e}^{-t}\,\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-t}}) + \mu \int_0^t \left([v(s)]^{\mathrm{e}^{-t+s}} - \mathrm{e}^{\nu\,\mathrm{e}^{-t+s}} \right) \mathrm{e}^{-t+s}\,\mathrm{d}s. \quad (31)$$

Therefore, we can control the difference by

$$|v(t) - \mathrm{e}^\nu| \leq \left| e_2(t) - \mathrm{e}^{-t}\,\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-t}}) \right| + \mu \int_0^t \left| [v(s)]^{\mathrm{e}^{-t+s}} - \mathrm{e}^{\nu\,\mathrm{e}^{-t+s}} \right| \mathrm{e}^{-t+s}\,\mathrm{d}s \quad (32)$$

$$\leq \left| e_2(t) - \mathrm{e}^{-t}\,\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-t}}) \right| + \mu \int_0^t |v(s) - \mathrm{e}^\nu|\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s. \quad (33)$$

Here, we used the fact that the power function $x \mapsto x^\alpha$ is α-Lipschitz on $[1, \infty)$ for $0 < \alpha < 1$. Define this integral quantity on the right by

$$y_2(t) := \int_0^t |v(s) - \mathrm{e}^\nu|\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s.$$

Then, $y_2'(t) = |v(t) - \mathrm{e}^\nu| - 2y_2(t)$, so y_2 satisfies the following differential inequality:

$$y_2'(t) + (2 - \mu)y_2(t) \leq \left| e_2(t) - \mathrm{e}^{-t}\,\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-t}}) \right|.$$

If $\mu < 2$, then y_2 will decay exponentially. However, this estimate is useless when $\mu > 2$. We need to estimate the difference in the following integral sense.

Lemma 4.9. *Denote $\bar{c} = \nu \wedge 1$. For any $c < \bar{c}$, define*

$$y_c(t) = \int_t^\infty |v(s) - e^\nu| \, e^{-ct+cs} \, ds.$$

If $y_c(0) < \infty$ converges, then for $c' = c + \frac{e^{-\nu}}{2}(1-c)$, it holds for all $t \geq 0$ that

$$y_c(t) + y_2(t) \leq C y_c(0) \, e^{-c't} + \begin{cases} \frac{C}{(\bar{c}-c)(\bar{c}-c')} \, e^{-c't}, & c' < \bar{c}, \\ \frac{C}{\bar{c}-c} t \, e^{-\bar{c}t}, & c' = \bar{c}, \\ \frac{C}{(\bar{c}-c)(c'-\bar{c})} \, e^{-\bar{c}t}, & c' > \bar{c}, \end{cases} \tag{34}$$

in which $C > 0$ is a constant depending only on μ.

Proof. Notice that

$$y_c'(t) = -|v(t) - e^\nu| - c y_c(t), \qquad y_2'(t) = |v(t) - e^\nu| - 2 y_2(t).$$

Combined, they satisfy the following differential equality:

$$(y_c + y_2)'(t) = -c y_c(t) - 2 y_2(t) = -c(y_c(t) + y_2(t)) - (2-c) y_2(t). \tag{35}$$

As $c < 2$, if we drop the last term then we directly get a decay at rate

$$y_c(t) + y_2(t) \leq y_c(0) \, e^{-ct}, \qquad \forall t \geq 0. \tag{36}$$

In the following, we would like to improve the decay rate from c to c'.
By (32), we can estimate y_c by

$$y_c(t) \leq E_c(t) + \mu \int_t^\infty \int_0^s \left| [v(r)]^{e^{-s+r}} \, e^{-s+r} - e^{\nu \, e^{-s+r}} \, e^{-s+r} \right| e^{-ct+cs} \, dr \, ds,$$

where

$$E_c(t) := \int_t^\infty \left| e_2(s) - e^{-s} \, \varphi(e^{\nu \, e^{-s}}) \right| e^{-ct+cs} \, ds.$$

Note that now we have an improved estimate of e_2. Indeed, using the crude estimate established in Lemma 4.8, we have

$$\int_0^t |\log v(s) - \nu| \, ds \leq \int_0^t |v(s) - e^\nu| \, ds \leq C.$$

Therefore, we can improve estimate (29) to

$$e_2(t) \leq \mu \, e^{-t} + C \, e^{-\nu t} \leq C \, e^{-\bar{c}t}. \tag{37}$$

As $\varphi(e^{\nu \, e^{-s}}) \leq \varphi(e^\nu)$ is bounded, we have

$$E_c(t) \leq C \int_t^\infty e^{-\bar{c}s} \, e^{-ct+cs} \, ds = \frac{C}{\bar{c}-c} \, e^{-\bar{c}t}, \qquad \forall t \geq 0.$$

We now exchange the order of the double integral:

$$\mu \int_t^\infty \int_0^s \left| [v(r)]^{e^{-s+r}} \, e^{-s+r} - e^{\nu \, e^{-s+r}} \, e^{-s+r} \right| e^{-ct+cs} \, dr \, ds$$

$$= \mu \int_0^\infty \int_{r \vee t}^\infty \left| [v(r)]^{e^{-s+r}} \, e^{-s+r} - e^{\nu \, e^{-s+r}} \, e^{-s+r} \right| e^{-cr+cs} \, ds \, e^{-ct+cr} \, dr$$

$$= \mu \int_0^\infty \left| \int_{r \vee t}^\infty [v(r)]^{e^{-s+r}} \, e^{-(1-c)(s-r)} - e^{\nu \, e^{-s+r}} \, e^{-(1-c)(s-r)} \, ds \right| e^{-ct+cr} \, dr$$

$$= \mu \int_0^\infty \left| \int_{(t-r)_+}^\infty [v(r)]^{e^{-s}} \, e^{-(1-c)s} - e^{\nu \, e^{-s}} \, e^{-(1-c)s} \, ds \right| e^{-ct+cr} \, dr.$$

The absolute value symbol is extracted because the (inner) integrand has a fixed sign for each r and t. To compute the inner integral, we define

$$\varphi_c(x) := \mu \int_0^\infty x^{\mathrm{e}^{-s}} \, \mathrm{e}^{-(1-c)s} \, \mathrm{d}s.$$

Then, the inner integral can be expressed as

$$\mu \int_{(t-r)_+}^\infty [v(r)]^{\mathrm{e}^{-s}} \, \mathrm{e}^{-(1-c)s} - \mathrm{e}^{\nu \, \mathrm{e}^{-s}} \, \mathrm{e}^{-(1-c)s} \, \mathrm{d}s$$

$$= \begin{cases} \mathrm{e}^{-(1-c)(t-r)} \left(\varphi_c([v(r)]^{\mathrm{e}^{-t+r}}) - \varphi_c(\mathrm{e}^{\nu \, \mathrm{e}^{-t+r}}) \right), & r < t, \\ \varphi_c(v(r)) - \varphi_c(\mathrm{e}^\nu), & r \geq t. \end{cases}$$

We will leave the detailed proof of the Lipschitzness of φ_c in Lemma 4.10. Denote $\epsilon_c = \frac{\mu \, \mathrm{e}^{-\nu}}{2(2-c)}(1-c) \in (0,1)$. Then, we show in Lemma 4.10 that φ_c is $(\mu-\epsilon_c)$-Lipschitz on $[1,\infty)$, and $(1-\epsilon_c)$-Lipschitz on $[\mathrm{e}^\nu - \epsilon_c/\mu, +\infty)$. By Lemma 4.8, for $t > T_c$ with

$$T_c = \frac{1}{\delta} \log^2 \left(\frac{C\mu}{\epsilon_c} \right) \leq C \log^2 \left(\frac{2}{1-c} \right),$$

we have $v(t) \in [\mathrm{e}^\nu - \epsilon_c/\mu, \mathrm{e}^\nu + \epsilon_c/\mu]$, so

$$\mu \left| \int_{(t-r)_+}^\infty [v(r)]^{\mathrm{e}^{-s}} \, \mathrm{e}^{-(1-c)s} - \mathrm{e}^{\nu \, \mathrm{e}^{-s}} \, \mathrm{e}^{-(1-c)s} \, \mathrm{d}s \right|$$

$$\leq \begin{cases} (\mu - \epsilon_c) \, |v(r) - \mathrm{e}^\nu| \, \mathrm{e}^{-(2-c)(t-r)}, & r < t, \\ (1 - \epsilon_c) \, |v(r) - \mathrm{e}^\nu|, & r \geq t. \end{cases}$$

Thus,

$$y_c(t) \leq E_c(t) + (\mu - \epsilon_c) \int_0^t |v(r) - \mathrm{e}^\nu| \, \mathrm{e}^{-2t+2r} \, \mathrm{d}r$$

$$+ (1 - \epsilon_c) \int_t^\infty |v(r) - \mathrm{e}^\nu| \, \mathrm{e}^{-ct+cr} \, \mathrm{d}r$$

$$= E_c(t) + (\mu - \epsilon_c)y_2(t) + (1 - \epsilon_c)y_c(t). \tag{38}$$

Rearranging (38) leads us to

$$\epsilon_c(y_c + y_2)(t) \leq E_c(t) + \mu y_2(t)$$

$$\implies \frac{\epsilon_c(2-c)}{\mu}(y_c + y_2)(t) \leq \frac{2-c}{\mu} E_c(t) + (2-c)y_2(t).$$

Combining it with (35), we obtain

$$(y_c + y_2)'(t) + \left(c + \frac{\mathrm{e}^{-\nu}}{2}(1-c) \right)(y_c + y_2)(t) \leq 2E_c(t).$$

Recall that $c' = c + \frac{1}{2}(1-c) \, \mathrm{e}^{-\nu}$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[(y_c + y_2)(t) \, \mathrm{e}^{c't} \right] \leq 2E_c(t) \, \mathrm{e}^{c't}.$$

Upon integration of this inequality from T_c to t, for all $t \geq T_c$, we get that

$$(y_c + y_2)(t) \leq (y_c + y_2)(T_c) \, \mathrm{e}^{-c'(t-T_c)} + 2 \int_{T_c}^t E_c(s) \, \mathrm{e}^{-c'(t-s)} \, \mathrm{d}s.$$

Note that (36) yields

$$(y_c + y_2)(T_c) \, e^{-c'(t-T_c)} \leq y_c(0) \, e^{-cT_c} \, e^{-c'(t-T_c)} \leq e^{(c'-c)T_c} \, y_c(0) \, e^{-c't}.$$

The right-hand side also dominates $(y_c + y_2)(t)$ for $t \leq T_c$. Indeed, from (36), we have that

$$(y_c + y_2)(t) \leq y_c(0) \, e^{ct} = y_c(0) \, e^{-cT_c} \, e^{c(T_c-t)} \leq y_c(0) \, e^{-cT_c} \, e^{c'(T_c-t)}$$
$$= y_c(0) \, e^{-c't} \, e^{(c'-c)T_c}.$$

Next, notice that for $c \in (0,1)$, $(c'-c)T_c$ is uniformly bounded by

$$(c'-c)T_c \leq C \log^2 \left(\frac{2}{1-c} \right) \cdot \frac{e^{-\nu}}{2}(1-c) \leq C.$$

As for the error term, we have

$$\int_{T_c}^t E_c(s) \, e^{-c'(t-s)} \, \mathrm{d}s \leq \int_{T_c}^t \frac{C}{\bar{c}-c} \, e^{-\bar{c}s} \, e^{-c't+c's} \, \mathrm{d}s$$

$$\leq \begin{cases} \frac{C}{(\bar{c}-c)(\bar{c}-c')} \, e^{-c't-(\bar{c}-c')T_c}, & c' < \bar{c}, \\ \frac{C}{\bar{c}-c}(t-T_c) \, e^{-\bar{c}t}, & c' = \bar{c}, \\ \frac{C}{(\bar{c}-c)(c'-\bar{c})} \, e^{-\bar{c}t}, & c' > \bar{c}, \end{cases}$$

from which the advertised estimate (34) follows. □

Lemma 4.10 (Lipschitzness of the function φ_c)**.** *Denote $\epsilon_c = \frac{\mu \, e^{-\nu}}{2(2-c)}(1-c) \in (0,1)$. Then, φ_c is $(\mu - \epsilon_c)$-Lipschitz on $[1, \infty)$, and $(1 - \epsilon_c)$-Lipschitz on $[e^\nu - \epsilon_c/\mu, +\infty)$.*

Proof. We remark that φ_c can be computed and expressed using Euler's Gamma function, but we will not need the exact form. Its derivative

$$\varphi_c'(x) = \mu \int_0^\infty e^{-s} \, x^{e^{-s}-1} \, e^{-(1-c)s} \, \mathrm{d}s = \mu \int_0^\infty x^{e^{-s}-1} \, e^{-(2-c)s} \, \mathrm{d}s$$

is positive and decreasing in $[1, \infty)$. Thus, for all $x \geq 1$,

$$\varphi_c'(x) \leq \varphi_c'(1) = \mu \int_0^\infty e^{-(2-c)s} \, \mathrm{d}s = \frac{\mu}{2-c} \leq \mu - \epsilon_c.$$

Moreover, the partial derivative of φ_c' with respect to c reads as

$$\partial_c \varphi_c'(x) = \mu \int_0^\infty x^{e^{-s}-1} \, e^{-(2-c)s} \, s \, \mathrm{d}s \geq \frac{\mu}{x} \int_0^\infty s \, e^{-(2-c)s} \, \mathrm{d}s = \frac{1}{(2-c)^2} \frac{\mu}{x},$$

hence, for any $c \leq 1$, it holds that

$$\varphi_c'(x) \leq \varphi_1'(x) - \frac{\mu}{x} \int_c^1 \frac{1}{(2-t)^2} \, \mathrm{d}t = \frac{\varphi(x)}{x} - \frac{1-c}{2-c} \cdot \frac{\mu}{x}.$$

In particular,

$$\varphi_c'(e^\nu) \leq 1 - \frac{1-c}{2-c} \cdot \frac{\mu}{e^\nu} = 1 - 2\epsilon_c < 1.$$

On the other hand, for $x \geq 1$, we have

$$\varphi_c''(x) = -\mu \int_0^\infty e^{-s}(1 - e^{-s}) x^{e^{-s}-2} \, e^{-(1-c)s} \, \mathrm{d}s \geq -\mu \int_0^\infty e^{-s}(1 - e^{-s}) \, e^{-(1-c)s} \, \mathrm{d}s$$

$$= -\frac{\mu}{(2-c)(3-c)} \geq -\mu.$$

Therefore, if $x \in [\mathrm{e}^\nu - \epsilon_c/\mu, \mathrm{e}^\nu]$, then

$$\varphi_c'(x) \le \varphi_c'(\mathrm{e}^\nu) + \mu(\mathrm{e}^\nu - x) \le 1 - 2\epsilon_c + \epsilon_c = 1 - \epsilon_c.$$

If $x \ge \mathrm{e}^\nu$, then $\varphi_c'(x) \le \varphi_c'(\mathrm{e}^\nu) \le 1 - 2\epsilon_c$. Thus, we conclude that the function φ_c is $(1 - \epsilon_c)$-Lipschitz on $[\mathrm{e}^\nu - \epsilon_c/\mu, +\infty)$. $\qquad\square$

Lemma 4.9 provides an iteration scheme to improve the decay rate. In the following proposition, we will use Lemma 4.9 to bootstrap the decay rate from Lemma 4.8 to a sharper exponential decay rate.

Proposition 4.11. *If $\nu < 1$, then there exists a constant $C > 0$ depending on μ such that*

$$y_2(t) + |v(t) - \mathrm{e}^\nu| + |v'(t)| \le C \, \mathrm{e}^{-\nu t}.$$

If $\nu \ge 1$, then there exist constants $C, K > 0$ depending on μ such that

$$y_2(t) + |v(t) - \mathrm{e}^\nu| + |v'(t)| \le C \langle t \rangle^K \, \mathrm{e}^{-t},$$

where $\langle t \rangle := \sqrt{1 + t^2}$ denotes the usual Japanese bracket shorthand.

Proof. We first claim that for any $\alpha \le \nu$ with $\alpha < 1$, it holds that

$$|v(s) - \mathrm{e}^\nu| \le C_\alpha \, \mathrm{e}^{-\alpha t}.$$

Indeed, given any $\tilde{c} < \min\{c', \bar{c}\}$, we have

$$y_{\tilde{c}}(0) = \int_0^\infty |v(s) - \mathrm{e}^\nu| \, \mathrm{e}^{\tilde{c}s} \, \mathrm{d}s = \int_0^\infty (-y_c'(s) - cy_c(s)) \, \mathrm{e}^{\tilde{c}s} \, \mathrm{d}s$$

$$= y_c(0) + \int_0^\infty (\tilde{c} - c)y_c(s) \, \mathrm{e}^{\tilde{c}s} \, \mathrm{d}s$$

$$\le y_c(0) + Cy_c(0)\frac{\tilde{c} - c}{c' - \tilde{c}} + \begin{cases} \frac{C(\tilde{c}-c)}{(\bar{c}-c)(\bar{c}-c')(c'-\bar{c})}, & c' < \bar{c} \\ \frac{C(\tilde{c}-c)}{(\bar{c}-c)(\bar{c}-\tilde{c})(c'-\bar{c})}, & c' = \bar{c} \\ \frac{C(\tilde{c}-c)}{(\bar{c}-c)(c'-\bar{c})(\bar{c}-\tilde{c})}, & c' > \bar{c}, \end{cases}$$

which is convergent. We now iterate this using Lemma 4.9. Clearly, $y_0(0) < \infty$ in view of Lemma 4.8. After finitely many steps, $c' > \alpha$, and we have

$$y_{c'}(t) + y_2(t) \le C \, \mathrm{e}^{-\alpha t}.$$

From this claim, we proved $y_2(t) \le C \, \mathrm{e}^{-\nu t}$ for $\nu < 1$. For the case $\nu \ge 1$, we know that $y_c(0) < \infty$ for all $c < 1$. To use Lemma 4.9, we need to quantify the size of $y_c(0)$, which is actually the Laplace transform of $|v - \mathrm{e}^\nu|$ evaluated at $-c$. Although it can be estimated from the above iteration, we use the following strategy instead. Taking the derivative of $y_c(0)$ with respect to c yields

$$\frac{\mathrm{d}}{\mathrm{d}c} y_c(0) = \frac{\mathrm{d}}{\mathrm{d}c} \int_0^\infty |v(t) - \mathrm{e}^\nu| \, \mathrm{e}^{ct} \, \mathrm{d}t = \int_0^\infty |v(t) - \mathrm{e}^\nu| \, t \, \mathrm{e}^{ct} \, \mathrm{d}t$$

$$= \int_0^\infty (-y_c'(t) - cy_c(t)) \, t \, \mathrm{e}^{ct} \, \mathrm{d}t = \int_0^\infty y_c(t)(1 + ct - ct) \, \mathrm{e}^{ct} \, \mathrm{d}t$$

$$= \int_0^\infty y_c(t) \, \mathrm{e}^{ct} \, \mathrm{d}t.$$

Using Lemma 4.9 and bearing in mind that $\nu \ge 1$ implies that $\bar{c} = 1 > c'$, we can control it by

$$\frac{\mathrm{d}}{\mathrm{d}c} y_c(0) \le \frac{C}{c' - c} y_c(0) + \frac{C}{(1-c)(1-c')(c'-c)}$$

$$\leq \frac{K}{1-c}y_c(0) + \frac{C}{(1-c)^3},$$

whence $y_c(0) \leq C(1-c)^{-K}$ by taking $K > 2$. Thus, we also have

$$y_2(t) \leq Cy_c(0)\,\mathrm{e}^{-c't} + \frac{C}{(\bar{c}-c)(\bar{c}-c')}\,\mathrm{e}^{-c't} \leq C(1-c)^{-K}\,\mathrm{e}^{-ct}.$$

This is true for all $c < 1$, so for $t > 1$ we can take $c = 1 - \frac{1}{t}$ and obtain

$$y_2(t) \leq Ct^K\,\mathrm{e}^{-t}.$$

In summary, we find y_2 has the desired decay rate.

To get the desired estimate for the convergence rate of $v(t)$, we employ (33):

$$\left|v(t) - \mathrm{e}^{\nu}\right| \leq \left|e_2(t) - \mathrm{e}^{-t}\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-t}})\right| + \mu y_2(t) \leq \begin{cases} C\,\mathrm{e}^{-\nu t}, & \nu < 1, \\ C\langle t\rangle^K\,\mathrm{e}^{-t}, & \nu \geq 1. \end{cases}$$

For the derivative estimate, we differentiate v using (30) and get

$$v'(t) = e_2'(t) + \mu v(t) - \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\,\mathrm{e}^{-t+s}\,\mathrm{d}s$$

$$- \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\log(v(s))\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s$$

$$= e_2'(t) + (\mu-1)v(t) + \mu e_2(t) - \mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\log(v(s))\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s,$$

in which

$$e_2'(t) = -f_0'(t) - f_0(0)\exp\left(-\int_0^t \log v(s)\,\mathrm{d}s\right)\log v(t) = \mathcal{O}(\mathrm{e}^{-(1\wedge\nu)t}).$$

Note that $x \mapsto x^\alpha \log x$ is Lipschitz on $[1, \mathrm{e}^\mu]$ uniformly for $\alpha \in [0,1]$, hence

$$\left|\mu \int_0^t [v(s)]^{\mathrm{e}^{-t+s}}\log(v(s))\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s - \mu \int_0^t \nu\,\mathrm{e}^{\nu\,\mathrm{e}^{-t+s}}\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s\right|$$

$$\leq C\mu \int_0^t |v(s) - \mathrm{e}^\nu|\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s = Cy_2(t).$$

Therefore,

$$|v'(t)| \leq |e_2'(t)| + (\mu-1)|v(t) - \mathrm{e}^\nu| + \mu e_2(t) + Cy_2(t)$$

$$+ (\mu-1)\,\mathrm{e}^\nu - \mu\nu \int_0^t \mathrm{e}^{\nu\,\mathrm{e}^{-s}}\,\mathrm{e}^{-2s}\,\mathrm{d}s$$

$$\leq |e_2'(t)| + (\mu-1)|v(t) - \mathrm{e}^\nu| + \mu e_2(t) + Cy_2(t) + \mu\nu \int_t^\infty \mathrm{e}^{\nu\,\mathrm{e}^{-s}}\,\mathrm{e}^{-2s}\,\mathrm{d}s$$

$$\leq \begin{cases} C\,\mathrm{e}^{-\nu t}, & \nu < 1, \\ C\langle t\rangle^K\,\mathrm{e}^{-t}, & \nu \geq 1. \end{cases}$$

This completes the proof. $\qquad\square$

As a corollary, we deduce the following convergence rate of $p_1(t) \to \mu - \nu$ as $t \to \infty$.

Corollary 4.12. *For $t \geq 0$, we have*

$$|p_1(t) - \mu + \nu| \leq \begin{cases} C\,\mathrm{e}^{-\nu t}, & \nu < 1, \\ C\langle t \rangle^K \mathrm{e}^{-t}, & \nu \geq 1. \end{cases}$$

Proof. It suffices to notice that $\mu - p_1(t)$ can be recovered from v via

$$\mu - p_1(t) = \frac{v'(t)}{v(t)} + \log v(t).$$

Since both $v'(t) \xrightarrow{t \to \infty} 0$ and $v(t) \xrightarrow{t \to \infty} \mathrm{e}^\nu$ occur at this rate, the result follows. $\square$

4.4. **Strong convergence of the ODE system.** In this subsection, we are ready to prove the various strong convergence results regarding the solution $\mathbf{p}(t)$ of system (10) toward the zero-truncated Poisson distribution $\bar{\mathbf{p}}$ (4) as $t \to +\infty$. First, we show convergence in ℓ^2. Recall that

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^2}^2 = \sum_{n=0}^{\infty} \left(p_n(t) - \bar{p}_n\right)^2.$$

Theorem 4.13. *Let $\mu > 1$. There exist constants C, K depending on μ such that for all $t \geq 0$, it holds that*

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^2} \leq \begin{cases} C\,\mathrm{e}^{-\nu t}, & \nu < 1, \\ C\langle t \rangle^K \mathrm{e}^{-t}, & \nu \geq 1. \end{cases}$$

Proof. We first recall the classical Parseval's identity:

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^2}^2 = \frac{1}{2\pi} \int_0^{2\pi} \left| G(t, \mathrm{e}^{i\theta}) - G_{\bar{\mathbf{p}}}(\mathrm{e}^{i\theta}) \right|^2 \mathrm{d}\theta.$$

By Lemma 4.1, we have

$$(G(t, 1-z) - 1)[v(t)]^z = G(0, 1 - z\,\mathrm{e}^{-t}) - 1 - \mu z \int_0^t [v(s)]^{z\,\mathrm{e}^{-t+s}} \mathrm{e}^{-t+s}\,\mathrm{d}s$$

for all $z \in \mathbb{C}$ with $|z - 1| \leq 1$. Notice that

$$|G(0, 1 - z\,\mathrm{e}^{-t}) - 1| \leq \partial_z G(0,1)|z|\,\mathrm{e}^{-t} \leq C\,\mathrm{e}^{-t},$$

$$\left| \int_0^t [v(s)]^{z\,\mathrm{e}^{-t+s}} \mathrm{e}^{-t+s}\,\mathrm{d}s - \int_0^t \mathrm{e}^{\nu z\,\mathrm{e}^{-t+s}} \mathrm{e}^{-t+s}\,\mathrm{d}s \right| \leq C|z| \int_0^t |v(s) - \mathrm{e}^\nu|\,\mathrm{e}^{-2t+2s}\,\mathrm{d}s$$
$$\leq C y_2(t),$$

$$\left| \mu \int_0^t \mathrm{e}^{\nu z\,\mathrm{e}^{-t+s}} \mathrm{e}^{-t+s}\,\mathrm{d}s - \varphi(\mathrm{e}^{\nu z}) \right| \leq \mathrm{e}^{-t} \left| \varphi(\mathrm{e}^{\nu z\,\mathrm{e}^{-t}}) \right| \leq C\,\mathrm{e}^{-t}.$$

On the other hand, we know for $z \in \mathbb{C}$ with $|z - 1| \leq 1$ that

$$\left| (G(t, 1-z) - 1)[v(t)]^z - (G(t, 1-z) - 1)\,\mathrm{e}^{\nu z} \right| \leq C|v(t) - \mathrm{e}^\nu|.$$

Assembling these estimates, we have proved for $z \in \mathbb{C}$ with $|z - 1| \leq 1$ that

$$|(G(t, 1-z) - 1)\,\mathrm{e}^{\nu z} + z\varphi(\mathrm{e}^{\nu z})| \leq \begin{cases} C\,\mathrm{e}^{-\nu t}, & \nu < 1, \\ C\langle t \rangle^K \mathrm{e}^{-t}, & \nu \geq 1. \end{cases}$$

Since

$$G_{\bar{\mathbf{p}}}(1-z) = \frac{1}{e^{\nu}-1} \sum_{n=1}^{\infty} \frac{\nu^n}{n!}(1-z)^n = \frac{e^{\nu(1-z)}-1}{e^{\nu}-1} = 1 - e^{-\nu z} z\varphi(e^{\nu z}),$$

the above implies uniform convergence of $G(t, 1-z)$ to $G_{\bar{\mathbf{p}}}(1-z)$ for all $z \in \mathbb{C}$ with $|z-1| \leq 1$, which shows that $\mathbf{p}(t)$ converges to $\bar{\mathbf{p}}$ in ℓ^2 by Parseval's identity. $\square$

Utilizing the tail estimate for the (zero-truncated) Poisson distribution, we can also establish the following convergence result in ℓ^1.

Corollary 4.14. *Let $\mu > 1$. There exist constants C, K depending on μ such that for all $t \geq 0$, it holds that*

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1} \leq \begin{cases} C\langle t\rangle^{\frac{1}{2}} e^{-\nu t}, & \nu < 1, \\ C\langle t\rangle^{K+\frac{1}{2}} e^{-t}, & \nu \geq 1. \end{cases}$$

Proof. For $x \in \mathbb{N}$ to be specified later, we have

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1} \leq \|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1([0,x])} + \|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1([x+1,\infty))}$$
$$\leq 2\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1([0,x])} + 2\|\bar{\mathbf{p}}\|_{\ell^1([x+1,\infty))}$$

The first term is easily controlled by the ℓ^2 norm:

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1([0,x])} \leq \sqrt{x}\,\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^2}.$$

The second term is amenable to explicit computations, leading us to

$$\|\bar{\mathbf{p}}\|_{\ell^1([x+1,\infty))} = \sum_{n=x+1}^{\infty} \bar{p}_n = \frac{1}{e^{\nu}-1} \sum_{n=x+1}^{\infty} \frac{\nu^n}{n!}.$$

Thanks to the Chernoff bound for the Poisson distribution, for $x \geq \nu$ it holds that

$$\sum_{n=x}^{\infty} \frac{\nu^n e^{-\nu}}{n!} \leq \frac{(e\,\nu)^x e^{-\nu}}{x^x}.$$

We know for our zero-truncated Poisson distribution that

$$\|\bar{\mathbf{p}}\|_{\ell^1([x+1,\infty))} \leq \frac{1}{e^{\nu}-1}\left(\frac{e\,\nu}{x}\right)^x.$$

Finally, setting $x = \lceil t \vee \nu\, e^2 \rceil$ allows us to deduce that $\|\bar{\mathbf{p}}\|_{\ell^1([x+1,\infty))} \leq C\, e^{-t}$, whence

$$\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1} \leq \begin{cases} C\langle t\rangle^{\frac{1}{2}} e^{-\nu t}, & \nu < 1 \\ C\langle t\rangle^{K+\frac{1}{2}} e^{-t}, & \nu \geq 1. \end{cases}$$

This completes the proof. $\square$

To illustrate the quantitative convergence guarantee reported in Corollary 4.14, we plot the evolution of the ℓ^1 error $\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1}$ over time (see Figure 7) with $\nu \in \{0.5, 1, 1.5, 2, 2.5, 3\}$ under the same set-up as used for Figure 4. We observe the exponential decay of $\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1}$ as predicted by Theorem 4.13, although our analytical rate might be sub-optimal for $\nu \geq 1$. Meanwhile, we also emphasize that Figure 7 contains a part where the machine precision is reached, and hence the numerical error can no longer decay further. Lastly, we observe that the numerical experiments displayed above for "large" ν seem to demonstrate two regimes in the convergence behavior: $e^{-(2\wedge\nu)t}$ in the first period, and $e^{-(1\wedge\nu)t}$ afterwards.
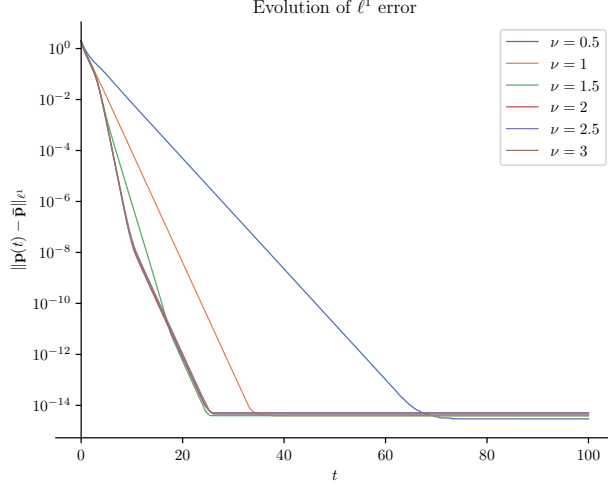
FIGURE 7. Evolution of the ℓ^1 error $\|\mathbf{p}(t) - \bar{\mathbf{p}}\|_{\ell^1}$ over time for different values of ν. It can be seen that larger values of ν (or μ) leads to faster convergence, although such improvement in terms of the convergence rate saturates when ν becomes large enough.

To understand the reason for the two-scale behavior, we examine the convergence for $\nu > 1$ in more detail and provide a non-rigorous explanation here. Instead of the convergence rate of v, we examine the convergence rate of a directly. We do so by linearizing the Volterra-type integral equation. Let $u(t) = \log v(t) - \nu$, and denote the negative exponential function as $\mathrm{e}^-(t) = \mathrm{e}^{-t}$. Then, (22) can be rewritten as the one-sided convolution

$$u = \mathrm{e}^- * a - \nu = \mathrm{e}^- * (a - \nu) - \nu \, \mathrm{e}^- . \tag{39}$$

Moreover, since $v = \mathrm{e}^\nu \, \mathrm{e}^u$, we rewrite (31) with u as

$$\mathrm{e}^\nu(\mathrm{e}^{u(t)} - 1) = e_2(t) - \mathrm{e}^{-t} \, \varphi(\mathrm{e}^{\nu \, \mathrm{e}^{-t}}) + \mu \int_0^t \left([v(s)]^{\mathrm{e}^{-t+s}} - \mathrm{e}^{\nu \, \mathrm{e}^{-t+s}} \right) \mathrm{e}^{-t+s} \, \mathrm{d}s$$

$$= e_2(t) - \mathrm{e}^{-t} \, \varphi(\mathrm{e}^{\nu \, \mathrm{e}^{-t}}) + \mu \int_0^t \left(\mathrm{e}^{u(s) \, \mathrm{e}^{-t+s}} - 1 \right) \mathrm{e}^{\nu \, \mathrm{e}^{-t+s}} \, \mathrm{e}^{-t+s} \, \mathrm{d}s .$$

Denote $k(t) = \mu \, \mathrm{e}^{\nu \, \mathrm{e}^{-t}} \, \mathrm{e}^{-2t}$. We can write the above equation in convolution form as

$$\mathrm{e}^\nu \, u = e_3 + k * u \tag{40}$$

with remainder

$$e_3(t) := e_2(t) - \mathrm{e}^{-t} \, \varphi(\mathrm{e}^{\nu \, \mathrm{e}^{-t}}) - \mathrm{e}^\nu(\mathrm{e}^{u(t)} - 1 - u(t))$$

$$+ \mu \int_0^t \left(\mathrm{e}^{u(s) \, \mathrm{e}^{-t+s}} - 1 - u(s) \, \mathrm{e}^{-t+s} \right) \mathrm{e}^{\nu \, \mathrm{e}^{-t+s}} \, \mathrm{e}^{-t+s} \, \mathrm{d}s .$$

From Proposition 4.11, we know $|u(t)| = \mathcal{O}(\langle t \rangle^K \, \mathrm{e}^{-t})$. Ignoring the algebraic factor $\langle t \rangle^K$ for now and using $\varphi(\mathrm{e}^{\nu \, \mathrm{e}^{-t}}) = \varphi(1) + \mathcal{O}(\mathrm{e}^{\nu \, \mathrm{e}^{-t}} - 1) = \mu + \mathcal{O}(\mathrm{e}^{-t})$, we

know

$$e_3(t) = e_2(t) - \mu\,\mathrm{e}^{-t} + \mathcal{O}(\mathrm{e}^{-2t}).$$

Combining (39) and (40), we have

$$\mathrm{e}^{\nu}\,\mathrm{e}^{-} * (a - \nu) - \nu\,\mathrm{e}^{\nu}\,\mathrm{e}^{-} = e_3 + k * \mathrm{e}^{-} * (a - \nu) - k * \nu\,\mathrm{e}^{-}$$

$$\implies (\mathrm{e}^{\nu}\,\mathrm{e}^{-} - k * \mathrm{e}^{-}) * (a - \nu) = e_3 + \nu\,\mathrm{e}^{\nu}\,\mathrm{e}^{-} - k * \nu\,\mathrm{e}^{-}. \tag{41}$$

We can compute one of the convolutions explicitly:

$$k * \mathrm{e}^{-}(t) = \mu \int_0^t \mathrm{e}^{\nu\,\mathrm{e}^{-s}}\,\mathrm{e}^{-2s}\,\mathrm{e}^{-t+s}\,\mathrm{d}s$$

$$= \mathrm{e}^{-t}\left(\mu \int_0^\infty \mathrm{e}^{\nu\,\mathrm{e}^{-s}}\,\mathrm{e}^{-s}\,\mathrm{d}s - \mu \int_t^\infty \mathrm{e}^{\nu\,\mathrm{e}^{-s}}\,\mathrm{e}^{-s}\,\mathrm{d}s\right)$$

$$= \mathrm{e}^{-t}\left(\mathrm{e}^{\nu} - \mathrm{e}^{-t}\,\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-t}})\right).$$

Thus, (41) becomes

$$(\mathrm{e}^{-})^2\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-}}) * (a - \nu) = e_3 + \nu(\mathrm{e}^{-})^2\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-}})$$

$$= e_2 - \mu\,\mathrm{e}^{-t} + \mathcal{O}((\mathrm{e}^{-})^2). \tag{42}$$

Recall that

$$e_2(t) = 1 - G(0, 1 - \mathrm{e}^{-t}) + \mathcal{O}(\mathrm{e}^{-\nu t}),$$

where

$$G(0, 1 - \mathrm{e}^{-t}) = G(0, 1) - \mathrm{e}^{-t}\,\partial_z G(0, 1) + \mathcal{O}(\mathrm{e}^{-2t}) = 1 - \mu\,\mathrm{e}^{-t} + \mathcal{O}(\mathrm{e}^{-2t}) \tag{43}$$

assuming finite second moment. Thus, we have a cancellation on the right-hand side of (42), and we are left with

$$(\mathrm{e}^{-})^2\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-}}) * (a - \nu)(t) = \mathcal{O}(\mathrm{e}^{-\nu t}) + \mathcal{O}(\mathrm{e}^{-2t}). \tag{44}$$

The convolution kernel has decay rate $\mathrm{e}^{-2t}\,\varphi(\mathrm{e}^{\nu\,\mathrm{e}^{-t}}) = \mathcal{O}(\mathrm{e}^{-2t})$. Therefore, provided that $a - \nu$ has an asymptotic leading term $C\,\mathrm{e}^{-\alpha t}$ for some α, then $\alpha = 2 \wedge \nu$ (unless there exists a further cancellation). Thus, we conjure that the sharp decay rate should be $\mathrm{e}^{-(2 \wedge \nu)t}$ for $\mu > 1$, which is supported by the experiments in Figure 7.

Suppose that, due to simulation machine error, the first moment of the initial condition $\partial_z G(0, 1)$ is not μ but slightly off. Then, in (43) the e^{-t} term cannot perfectly cancel $\mu\,\mathrm{e}^{-t}$ in (42), and the right-hand side of (44) would have a decay rate of e^{-t} instead of $\mathrm{e}^{-(2 \wedge \nu)t}$. We believe this is the reason why we observe a change in decay rate for $\nu > 1$ in Figure 7 beneath a certain scale.

5. **Conclusion.** In this work, we studied the Fokker–Planck equation associated with the mean-field limit of the so-called dispersion process (on a complete graph with N vertices) introduced and studied in a number of recent works [21, 23, 28, 44]. We prove quantitative global convergence results for the solution of the Fokker–Planck equation toward the equilibrium. The equilibrium and the obtained convergence rate differ in the underpopulated regime ($\mu < 1$) and the overpopulated regime ($\mu > 1$).

This work also leaves some important follow-up problems which deserve separate treatments and attention on their own. For instance, can we design a natural Lyapunov functional associated to the solution of the infinite-dimensional nonlinear

ODE system (1) when $\mu > 1$? Is it possible to improve the various decay rates reported in Theorem 1.1? We believe that answers to these questions will enhance our understanding on the role played by the model parameter μ and contribute to the ever-growling literature on the large time analysis of mean-field dynamical systems arising from social-economic, life, and natural sciences [11, 14, 32, 41].

Appendix A. **Uniqueness of solution to the mean-field ODE system.** We provide a proof for the uniqueness of solution to the mean-field dynamical system (1) under the framework of Wasserstein distances. The proof is inspired by the coupling method (see [46, section 7.5]). To illustrate the main ideas, let us recast the equation into the most general form. Let $\mathbf{p}, \tilde{\mathbf{p}} : [0, T] \to \mathbb{R}^{\mathbb{N}}$ be two solutions to the linear Fokker–Planck equations with different drifts

$$\mathbf{p}'(t) = \mathcal{D}_+[\mathbf{b}(t)\mathbf{p}(t)] - \mathcal{D}_-[\mathbf{a}(t)\mathbf{p}(t)], \tag{45}$$

$$\tilde{\mathbf{p}}'(t) = \mathcal{D}_+[\tilde{\mathbf{b}}(t)\tilde{\mathbf{p}}(t)] - \mathcal{D}_-[\tilde{\mathbf{a}}(t)\tilde{\mathbf{p}}(t)]. \tag{46}$$

where $\mathbf{a}, \mathbf{b}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}} : [0, T] \to \mathbb{R}^{\mathbb{N}}$ are upward/downward drift, and the product is in the pointwise sense. A transport plan $\boldsymbol{\pi} = \{\pi_{m,n}\}_{m,n=0}^{\infty}$ between $\mathbf{q}$ and $\tilde{\mathbf{q}}$ is a probability measure on $\mathbb{N} \times \mathbb{N}$ satisfying

$$\pi_{m,n} \geq 0, \qquad \sum_{n=0}^{\infty} \pi_{m,n} = q_m, \qquad \sum_{m=0}^{\infty} \pi_{m,n} = \tilde{q}_n, \qquad \sum_{m,n=0}^{\infty} \pi_{m,n} = 1.$$

The last condition is redundant because $\mathbf{q}$ and $\tilde{\mathbf{q}}$ are both probability measures. The set of transport plans between $\mathbf{q}$ and $\tilde{\mathbf{q}}$ is denoted as $\Pi(\mathbf{q}, \tilde{\mathbf{q}})$. The 1-Wasserstein distance is defined as

$$\mathcal{W}(\mathbf{q}, \tilde{\mathbf{q}}) = \inf_{\boldsymbol{\pi} \in \Pi(\mathbf{q}, \tilde{\mathbf{q}})} \sum_{m,n=0}^{\infty} |m - n| \pi_{m,n}.$$

Let $\boldsymbol{\gamma}_0 \in \Pi(\mathbf{p}(0), \tilde{\mathbf{p}}(0))$ be a transport plan between the initial conditions. We want to design a transport plan $\boldsymbol{\gamma}(t) \in \Pi(\mathbf{p}(t), \tilde{\mathbf{p}}(t))$. To do this, we first need to generalize the forward/backward difference operator to $\mathbb{N}^2$:

$$\mathcal{D}_{++}[\boldsymbol{\pi}]_{m,n} = \pi_{m+1,n+1} - \pi_{m,n}, \qquad \mathcal{D}_{--}[\boldsymbol{\pi}]_{m,n} = \pi_{m,n} - \pi_{m-1,n-1},$$
$$\mathcal{D}_{+\circ}[\boldsymbol{\pi}]_{m,n} = \pi_{m+1,n} - \pi_{m,n}, \qquad \mathcal{D}_{-\circ}[\boldsymbol{\pi}]_{m,n} = \pi_{m,n} - \pi_{m-1,n},$$
$$\mathcal{D}_{\circ+}[\boldsymbol{\pi}]_{m,n} = \pi_{m,n+1} - \pi_{m,n}, \qquad \mathcal{D}_{\circ-}[\boldsymbol{\pi}]_{m,n} = \pi_{m,n} - \pi_{m,n-1}.$$

We define the drifts $\mathbf{A}^{--}, \mathbf{A}^{-\circ}, \mathbf{A}^{\circ-}$ and $\mathbf{B}^{++}, \mathbf{B}^{+\circ}, \mathbf{B}^{\circ+} : [0, T] \to \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ by

$$A_{m,n}^{--}(t) = a_m(t) \wedge a_n(t), \qquad B_{m,n}^{++}(t) = b_m(t) \wedge b_n(t),$$
$$A_{m,n}^{-\circ}(t) = [a_m(t) - a_n(t)]_+, \qquad B_{m,n}^{+\circ}(t) = [b_m(t) - b_n(t)]_+,$$
$$A_{m,n}^{\circ-}(t) = [a_n(t) - a_m(t)]_+, \qquad B_{m,n}^{\circ+}(t) = [b_n(t) - b_m(t)]_+.$$

Here, $\mathbf{a} = \{a_m\}_{m=0}^{\infty}$, $\mathbf{b} = \{b_m\}_{m=0}^{\infty}$, $\mathbf{A}^{--} = \{A_{m,n}^{--}\}_{m,n=0}^{\infty}$, etc., and $(x)_+ = \max\{x, 0\}$ denotes the positive part. Define $\boldsymbol{\gamma} : [0, T] \to \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ to be the solution to the following Fokker–Planck equation in $\mathbb{N}^2$ with initial condition $\boldsymbol{\gamma}(0) = \boldsymbol{\gamma}_0$:

$$\boldsymbol{\gamma}'(t) = \mathcal{D}_{++}[\mathbf{B}^{++}(t)\boldsymbol{\gamma}(t)] + \mathcal{D}_{+\circ}[\mathbf{B}^{+\circ}(t)\boldsymbol{\gamma}(t)] + \mathcal{D}_{\circ+}[\mathbf{B}^{\circ+}(t)\boldsymbol{\gamma}(t)]$$
$$- \mathcal{D}_{--}[\mathbf{A}^{--}(t)\boldsymbol{\gamma}(t)] - \mathcal{D}_{-\circ}[\mathbf{A}^{-\circ}(t)\boldsymbol{\gamma}(t)] - \mathcal{D}_{\circ-}[\mathbf{A}^{\circ-}(t)\boldsymbol{\gamma}(t)].$$

Same as (45) and (46), the drifts are passive rather than active, so it is indeed a well-posed linear Fokker–Planck equation. To see that the solution is a transport

plan, we first examine the forward drift of the first marginal. From the definition of $\mathcal{D}_{\circ-}$ and $\mathcal{D}_{-\circ}$, we know that for any $\mathbf{C} \in \mathbb{R}^{\mathbb{N} \times \mathbb{N}}$ the following identities hold:

$$\sum_{n=0}^{\infty} \mathcal{D}_{\circ-}[\mathbf{C}]_{m,n} = \sum_{n=0}^{\infty} C_{m,n} - C_{m,n-1} = 0,$$

$$\sum_{n=0}^{\infty} \mathcal{D}_{-\circ}[\mathbf{C}]_{m,n} = \sum_{n=0}^{\infty} C_{m,n} - C_{m-1,n} = \sum_{n=0}^{\infty} C_{m,n} - C_{m-1,n-1} = \sum_{n=0}^{\infty} \mathcal{D}_{--}[\mathbf{C}]_{m,n}.$$

Therefore,

$$\sum_{n=0}^{\infty} \mathcal{D}_{--}[\mathbf{A}^{--}(t)\boldsymbol{\pi}]_{m,n} + \mathcal{D}_{-\circ}[\mathbf{A}^{-\circ}(t)\boldsymbol{\pi}]_{m,n} + \mathcal{D}_{\circ-}[\mathbf{A}^{\circ-}(t)\boldsymbol{\gamma}(t)]_{m,n}$$

$$= \sum_{n=0}^{\infty} \mathcal{D}_{--}[\mathbf{A}^{--}(t)\boldsymbol{\pi}]_{m,n} + \mathcal{D}_{--}[\mathbf{A}^{-\circ}(t)\boldsymbol{\pi}]_{m,n} + 0$$

$$= \sum_{n=0}^{\infty} \mathcal{D}_{--}[(\mathbf{A}^{--}(t) + \mathbf{A}^{-\circ}(t))\boldsymbol{\pi}]_{m,n}$$

$$= \sum_{n=0}^{\infty} a_m(t)\pi_{m,n} - a_{m-1}(t)\pi_{m-1,n-1}$$

$$= \mathcal{D}_{-}[\mathbf{a}(t)\boldsymbol{\pi}_1]_m.$$

Here, $\boldsymbol{\pi}_1$ is the first marginal of $\boldsymbol{\pi}$. By a similar computation to the backward drift, we conclude the first marginal $\boldsymbol{\gamma}_1$ of $\boldsymbol{\gamma}$ solves a linear Fokker–Planck equation

$$\boldsymbol{\gamma}_1'(t) = \mathcal{D}_{+}[\mathbf{b}(t)\boldsymbol{\gamma}_1(t)] - \mathcal{D}_{-}[\mathbf{a}(t)\boldsymbol{\gamma}_1(t)].$$

From (45), we know $\mathbf{p}$ and $\boldsymbol{\gamma}_1$ solve the same linear equation (45). They share the same initial condition, so by the uniqueness of linear Fokker–Planck equation, we know $\boldsymbol{\gamma}_1(t) = \mathbf{p}(t)$ for all $t \in [0, T]$. Similarly the second marginal $\boldsymbol{\gamma}_2(t) = \tilde{\mathbf{p}}(t)$, and hence $\boldsymbol{\gamma}(t) \in \Pi(\mathbf{p}(t), \tilde{\mathbf{p}}(t))$.

With this transport plan, we can estimate the Wasserstein distance

$$\mathcal{W}(\mathbf{p}(t), \tilde{\mathbf{p}}(t)) \leq D(t) := \sum_{m,n=0}^{\infty} |m - n|\gamma_{m,n}(t).$$

Taking the derivative, we know

$$D'(t) = \sum_{m,n=0}^{\infty} |m - n|\Bigg(\mathcal{D}_{++}[\mathbf{B}^{++}(t)\boldsymbol{\gamma}(t)] + \mathcal{D}_{+\circ}[\mathbf{B}^{+\circ}(t)\boldsymbol{\gamma}(t)] + \mathcal{D}_{\circ+}[\mathbf{B}^{\circ+}(t)\boldsymbol{\gamma}(t)]$$

$$- \mathcal{D}_{--}[\mathbf{A}^{--}(t)\boldsymbol{\gamma}(t)] - \mathcal{D}_{-\circ}[\mathbf{A}^{-\circ}(t)\boldsymbol{\gamma}(t)] - \mathcal{D}_{\circ-}[\mathbf{A}^{\circ-}(t)\boldsymbol{\gamma}(t)] \Bigg).$$

From the definition of $\mathcal{D}_{++}$ and $\mathcal{D}_{--}$, immediately we have

$$\sum_{m,n=0}^{\infty} |m - n|\mathcal{D}_{++}[\mathbf{B}^{++}(t)\boldsymbol{\gamma}(t)] = \sum_{m,n=0}^{\infty} |m - n|\mathcal{D}_{--}[\mathbf{A}^{--}(t)\boldsymbol{\gamma}(t)] = 0.$$

Moreover, using the discrete integration by parts, we have

$$\sum_{m,n=0}^{\infty} |m - n|\mathcal{D}_{-\circ}[\mathbf{A}^{-\circ}(t)\boldsymbol{\gamma}(t)]_{m,n} = \sum_{m,n=0}^{\infty} -\mathcal{D}_{+\circ}[|m - n|]A_{m,n}^{-\circ}(t)\gamma_{m,n}(t)$$

$$\leq \sum_{m,n=0}^{\infty} A_{m,n}^{-\circ}(t)\gamma_{m,n}(t).$$

Similar computations yield

$$D'(t) \leq \sum_{m,n=0}^{\infty} \left(A_{m,n}^{-\circ}(t) + A_{m,n}^{\circ-}(t) + B_{m,n}^{+\circ}(t) + B_{m,n}^{\circ+}(t)\right) \gamma_{m,n}(t)$$

$$= \sum_{m,n=0}^{\infty} \left(|a_m(t) - a_n(t)| + |b_m(t) - b_n(t)|\right) \gamma_{m,n}(t). \tag{47}$$

If the right-hand side can be controlled by $D(t)$, then we would have a differential inequality.

Now, let $\mathbf{p}$ and $\tilde{\mathbf{p}}$ be two solutions to (1). Then, they solve (45) and (46) with the drifts

$$a_m(t) = \sum_{n=2}^{\infty} np_n(t) = \mu - p_1(t), \quad \tilde{a}_m(t) = \mu - \tilde{p}_1(t), \quad b_n(t) = \tilde{b}_n(t) = n\mathbf{1}_{\{n\geq 2\}}.$$

The backward drift part in (47) can be controlled by

$$\sum_{m,n=0}^{\infty} |b_m(t) - b_n(t)|\gamma_{m,n}(t) \leq \sum_{m,n=0}^{\infty} |m - n|\gamma_{m,n}(t) = D(t).$$

As for the forward drift in (47), it is independent of m, n, so

$$\sum_{m,n=0}^{\infty} |a_m(t) - a_n(t)|\gamma_{m,n}(t) = \sum_{m,n=0}^{\infty} |\mu - p_1(t) - \mu + \tilde{p}_1(t)|\gamma_{m,n}(t)$$

$$= |p_1(t) - \tilde{p}_1(t)|.$$

It remains to show that $|p_1(t) - \tilde{p}_1(t)| \leq D(t)$, which is elementary:

$$D(t) \geq \sum_{m\neq n} \gamma_{m,n}(t) \geq \sum_{m=0}^{\infty} \gamma_{m,1}(t) + \sum_{n=0}^{\infty} \gamma_{1,n}(t) - 2\gamma_{1,1}(t)$$

$$= p_1(t) + \tilde{p}_1(t) - 2\gamma_{1,1}(t).$$

Because $\gamma_{1,1}(t) \leq p_1(t) \wedge \tilde{p}_1(t)$, we have $|p_1(t) - \tilde{p}_1(t)| \leq D(t)$.

To summarize, we have shown that

$$D'(t) \leq 2D(t).$$

Gronwall's inequality provides an upper bound on the Wasserstein distance between two solutions:

$$\mathcal{W}(\mathbf{p}(t), \tilde{\mathbf{p}}(t)) \leq D(t) \leq D(0)\,\mathrm{e}^{2t}.$$

Taking the infimum of all transport plans at time zero, we have

$$\mathcal{W}(\mathbf{p}(t), \tilde{\mathbf{p}}(t)) \leq \mathcal{W}(\mathbf{p}(0), \tilde{\mathbf{p}}(0))\,\mathrm{e}^{2t}.$$

If $\mathbf{p}(0) = \tilde{\mathbf{p}}(0)$, then $\mathcal{W}(\mathbf{p}(t), \tilde{\mathbf{p}}(t)) = 0$, which implies $\mathbf{p}(t) = \tilde{\mathbf{p}}(t)$ for all $t > 0$.

REFERENCES

[1] T. Arbogast and J. L. Bona, *Functional Analysis for the Applied Mathematician*, 1st edition, Chapman and Hall/CRC, New York, 2025.

[2] J. M. Ball, J. Carr and O. Penrose, The Becker-Döring cluster equations: Basic properties and asymptotic behaviour of solutions, *Communications in Mathematical Physics*, **104** (1986), 657-692.

[3] J. M. Ball and J. Carr, The discrete coagulation-fragmentation equations: Existence, uniqueness, and density conservation, *Journal of Statistical Physics*, **61** (1990), 203-234.

[4] J. M. Ball and J. Carr, Asymptotic behaviour of solutions to the Becker-Döring equations for arbitrary initial data, *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, **108** (1988), 109-116.

[5] B. M. Boghosian, M. Johnson and J. A. Marcq, An *H* theorem for Boltzmann's equation for the Yard-Sale model of asset exchange: The Gini coefficient as an *H* functional, *Journal of Statistical Physics*, **161** (2015), 1339-1350.

[6] F. Cao and S. Reed, A biased dollar exchange model involving bank and debt with discontinuous equilibrium, *Mathematical Modelling of Natural Phenomena*, **20** (2025), 5.

[7] F. Cao and N. F. Marshall, From the binomial reshuffling model to Poisson distribution of money, *Networks and Heterogeneous Media*, **19** (2024), 24-43.

[8] F. Cao and S. Motsch, Derivation of wealth distributions from biased exchange of money, *Kinetic & Related Models*, **16** (2023), 764-794.

[9] F. Cao, P.-E. Jabin and S. Motsch, Entropy dissipation and propagation of chaos for the uniform reshuffling model, *Mathematical Models and Methods in Applied Sciences*, **33** (2023), 829-875.

[10] F. Cao, Explicit decay rate for the Gini index in the repeated averaging model, *Mathematical Methods in the Applied Sciences*, **46** (2023), 3583-3596.

[11] F. Cao and R. Cortez, Fractal opinions among interacting agents, *SIAM Journal on Applied Dynamical Systems*, **24** (2025), 1529-1552.

[12] F. Cao, From Gini index as a Lyapunov functional to convergence in Wasserstein distance, preprint, (2024), `arXiv:2409.15225`.

[13] F. Cao and P.-E. Jabin, From interacting agents to Boltzmann-Gibbs distribution of money, *Nonlinearity*, **37** (2024), 125020.

[14] F Cao and S. Reed, The iterative persuasion-polarization opinion dynamics and its mean-field analysis, *SIAM Journal on Applied Mathematics*, **85** (2025), 1596-1620.

[15] F. Cao and S. Motsch, Sticky dispersion on the complete graph: A kinetic approach, *SIAM Journal on Mathematical Analysis*, **57** (2025), 3953-3980.

[16] F. Cao and S. Motsch, Uncovering a two-phase dynamics from a dollar exchange model with bank and debt, *SIAM Journal on Applied Mathematics*, **83** (2023), 1872-1891.

[17] F. Cao and R. Cortez, Uniform propagation of chaos for a dollar exchange econophysics model, *European Journal of Applied Mathematics*, **36** (2025), 27-39.

[18] A. Chakraborti and B. K. Chakrabarti, Statistical mechanics of money: How saving propensity affects its distribution, *The European Physical Journal B-Condensed Matter and Complex Systems*, **17** (2000), 167-170.

[19] A. Chakraborti, I. M. Toke, M. Patriarca and F. Abergel, Econophysics review: I. Empirical facts, *Quantitative Finance*, **11** (2011), 991-1012.

[20] A. Chatterjee, B. K. Chakrabarti and S. S. Manna, Pareto law in a kinetic model of market with random saving propensity, *Physica A: Statistical Mechanics and its Applications*, **335** (2004), 155-163.

[21] C. Cooper, A. McDowell, T. Radzik, N. Rivera and T. Shiraga, Dispersion processes, *Random Structures & Algorithms*, **53** (2018), 561-585.

[22] R. Cortez and J. Fontbona, Quantitative propagation of chaos for generalized Kac particle systems, *The Annals of Applied Probability*, **26** (2016), 892-916.

[23] U. De Ambroggio, T. Makai and K. Panagiotou, Dispersion on the complete graph, *Random Structures & Algorithms*, **66** (2025), e21266.

[24] A. Dragulescu and V. M. Yakovenko, Statistical mechanics of money, *The European Physical Journal B-Condensed Matter and Complex Systems*, **17** (2000), 723-729.

[25] B. Düring, D. Matthes and G. Toscani, Kinetic equations modelling wealth redistribution: A comparison of approaches, *Physical Review E*, **78** (2008), 056103.

[26] M. L. Esquivel, Probability generating functions for discrete real-valued random variables, *Theory of Probability & Its Applications*, **52** (2008), 40-57.

[27] L. C. Evans, *Partial Differential Equations*, raduate Studies in Mathematics, Vol. 19. American Mathematical Society, Providence, 2022.

[28] A. Frieze and W. Pegden, A note on dispersing particles on a line, *Random Structures & Algorithms*, **53** (2018), 586-591.

[29] G. Gripenberg, S.-O. Londén and O. Staffans, *Volterra Integral and Functional Equations*, 1st edition, Cambridge University Press, Cambridge, 1990.

[30] E. Heinsalu and P. Marco, Kinetic models of immediate exchange, *The European Physical Journal B*, **87** (2014), 1-10.

[31] P.-E. Jabin and B. Niethammer, On the rate of convergence to equilibrium in the Becker–Döring equations, *Journal of Differential Equations*, **191** (2003), 518-543.

[32] P.-E. Jabin, A review of the mean field limits for Vlasov equations, *Kinetic & Related Models*, **7** (2014), 661-711.

[33] A. Jüngel, *Entropy Methods for Diffusive Partial Differential Equations*, 1st edition, Springer, Vol. 804, Cham, 2016.

[34] R. Kutner, M. Ausloos, D. Grech, T. Di Matteo, C. Schinckus and H. E. Stanley, Econophysics and sociophysics: Their milestones & challenges, *Physica A: Statistical Mechanics and its Applications*, **516** (2019), 240-253.

[35] J.-H. Lambert, Observationes variae in mathesin puram, *Acta Helvetica*, **3** (1758), 128-168.

[36] N. Lanchier, Rigorous proof of the Boltzmann–Gibbs distribution of money on connected graphs, *Journal of Statistical Physics*, **167** (2017), 160-172.

[37] N. Lanchier and S. Reed, Rigorous results for the distribution of money on connected graphs (models with debts), *Journal of Statistical Physics*, **176** (2019), 1115-1137.

[38] D. Matthes and G. Toscani, On steady distributions of kinetic models of conservative economies. *Journal of Statistical Physics*, **130** (2008), 1087-1117.

[39] B. Niethammer, On the evolution of large clusters in the Becker-Döring model, *Journal of Nonlinear Science*, **13** (2003), 115-122.

[40] B. Niethammer, A scaling limit of the Becker-Döring equations in the regime of small excess density, *Journal of Nonlinear Science*, **14** (2004), 453-468.

[41] L. Pareschi and G. Toscani, *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods*, 1st edition, Oxford University Press, Oxford, 2014.

[42] E. J. de A. L. Pereira, M. F. da Silva and H. de B. Pereira, Econophysics: Past and present, *Physica A: Statistical Mechanics and its Applications*, **473** (2017), 251-261.

[43] G. Savoiu, *Econophysics: Background and Applications in Economics, Finance, and Sociophysics*, 1st edition, Academic Press, Oxford, 2013.

[44] Y. Shang, Longest distance of a non-uniform dispersion process on the infinite line, *Information Processing Letters*, **164** (2020), 106008.

[45] A.-S. Sznitman, Topics in propagation of chaos, *Ecole d'été de Probabilités de Saint-Flour XIX—1989*, (1991), 165-251.

[46] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, Vol. 58. American Mathematical Society, Providence, 2003.