The Andersen Thermostat in Molecular Dynamics

WEINAN E
Princeton University

AND

DONG LI
Princeton University

Abstract

We carry out a mathematical study of the Andersen thermostat [1], which is a frequently used tool in molecular dynamics. After reformulating the continuous- and discrete-time Andersen dynamics, we prove that in both cases the Andersen dynamics is uniformly ergodic. A detailed numerical analysis is presented, establishing the rate of convergence of most commonly used numerical algorithms for the Andersen thermostat. Transport properties such as the diffusion constant are also investigated. It is proved for the Lorentz gas model where there is intrinsic diffusion that the diffusion coefficient calculated using the Andersen thermostat converges to the true diffusion coefficient in the limit of vanishing collision frequency in the Andersen thermostat.

1 Introduction

1.1 Continuous-Time Andersen Process

The Andersen thermostat is the first molecular dynamics algorithm for simulating a canonical ensemble at a fixed temperature. The idea is to couple the system to a heat bath by using stochastic impulsive forces that act occasionally on randomly selected particles. After the stochastic collision, the chosen particle "forgets" its old velocity and picks its new velocity from a Maxwell-Boltzmann distribution at the imposed temperature. Between stochastic collisions, the system of N particles evolves according to a Hamiltonian dynamics in $\mathbb{R}^d$ that is given by the following:

$$\begin{cases} \dot{\mathbf{q}}_i = \mathbf{v}_i, \\ \dot{\mathbf{v}}_i = -\nabla_{\mathbf{q}_i}\Phi, \quad i = 1, \dots, N, \end{cases} \tag{1.1}$$

where $\Phi = \Phi(\mathbf{q}_1, \dots, \mathbf{q}_N)$ is the interatomic potential. We shall assume that Φ is infinitely differentiable, although in most places this assumption can be considerably relaxed. Throughout this paper we shall choose units such that the mass is normalized to 1.

We shall define the Andersen process on the phase space $\Gamma = \mathbb{D} \times \mathbb{R}^{dN}$. The configuration space $\mathbb{D}$ is assumed to be a torus in $\mathbb{R}^{dN}$. For simplicity $\mathbb{D} =$

Communications on Pure and Applied Mathematics, Vol. LXI, 0096–0136 (2008)

$\mathbb{R}^{dN}/\mathbb{Z}^{dN}$. For ease of exposition, we introduce the Andersen substitution operator: $\mathcal{S}(i,\mathbf{u}) : \Gamma \to \Gamma$ such that $\mathcal{S}(i,\mathbf{u})\mathbf{y}$ is the vector $\mathbf{y}$ except that the velocity of the i^{th} particle is changed to $\mathbf{u}$. With this notation, the continuous-time Andersen dynamics is defined as a Markov process on Γ.

The d-dimensional Maxwell-Boltzmann distribution with temperature $T = \frac{1}{\beta}$ is given by

$$g_\beta(\mathbf{v})d\mathbf{v} = \left(\frac{\beta}{2\pi}\right)^{d/2} \exp\left(-\frac{\beta|\mathbf{v}|^2}{2}\right) d\mathbf{v}.$$

DEFINITION 1.1 (Continuous-Time Andersen Process) Suppose on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$, $\{T_n\}_1^\infty$ are i.i.d. random variables that are exponentially distributed with mean $\frac{1}{\nu}$ $(\nu > 0)$; $\{Y_n\}_1^\infty$ are i.i.d. random variables such that $P(Y_n = j) = \frac{1}{N}$ for any integer j between 1 and N; and $\{Z_n\}_1^\infty$ are i.i.d. random variables in $\mathbb{R}^d$ that obey a common Maxwell-Boltzmann distribution. Let N_t denote the Poisson counting process generated by $\{T_n\}$; then the continuous-time Andersen dynamics starting at a point $\mathbf{x} \in \Gamma$ is defined as

$$\begin{cases} X_t := \mathcal{H}(t - \sum_1^{N_t} T_n)[\prod_1^{N_t} \mathcal{S}(Y_n, Z_n)\mathcal{H}(T_n)]\mathbf{x}, & t > 0, \\ X_0 = \mathbf{x}, \end{cases} \tag{1.2}$$

where $\mathcal{H}(\cdot)$ is the Hamiltonian flow operator and X_t is related to the starting point $\mathbf{x}$ by successively applying a cascade of operators. Here for a sequence of operators $\mathcal{A}_n$, $\prod_1^N \mathcal{A}_n$ is defined as the backward product, i.e.,

$$\prod_1^N \mathcal{A}_n = \mathcal{A}_N \mathcal{A}_{N-1} \cdots \mathcal{A}_1.$$

It is immediately obvious that the Andersen process has right-continuous sample trajectories, and therefore its invariant distribution can be found by computing the infinitesimal generator associated with the process. To write down an explicit form of the infinitesimal generator, we first introduce the notion of an Andersen collision operator:

DEFINITION 1.2 Let $\mathbf{q} = (\mathbf{q}_1, \ldots, \mathbf{q}_N)$, $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$, and $\mathbf{x} = (\mathbf{q}, \mathbf{v})$, and let $g_\beta(\cdot)$ be the probability density function of the aforementioned Maxwell-Boltzmann distribution at temperature $\frac{1}{\beta}$; then the *Andersen collision operator* $\mathcal{A}$ is defined as

$$(\mathcal{A}f)(\mathbf{x}) := \frac{1}{N}\sum_{i=1}^N \mathcal{A}_i f := \frac{1}{N}\sum_{i=1}^N g_\beta(\mathbf{v}_i) \int\limits_{\mathbb{R}^d} f(\mathcal{S}(i,\mathbf{u})\mathbf{x}) d\mathbf{u}.$$

The *adjoint* $\mathcal{A}^*$ *of the Andersen collision operator* is given by

$$(\mathcal{A}^* f)(\mathbf{x}) := \frac{1}{N}\sum_{i=1}^N \int\limits_{\mathbb{R}^d} f(\mathcal{S}(i,\mathbf{u})\mathbf{x})\ g_\beta(\mathbf{u}) d\mathbf{u}.$$

The infinitesimal generator $\mathcal{G}$ of the Andersen process can be computed and is given by

$$\mathcal{G} := \nu(\mathcal{A}^* - \mathcal{I}) + i\mathcal{L},$$

where $\mathcal{I}$ is the identity operator and the Liouville operator $i\mathcal{L}$ is defined as

$$(i\mathcal{L}f)(\mathbf{q}, \mathbf{v}) := \mathbf{v} \cdot \nabla_{\mathbf{q}} f - \nabla_{\mathbf{q}} \Phi \cdot \nabla_{\mathbf{v}} f.$$

Let us recall the following definition of invariant measure:

DEFINITION 1.3 (Invariant Measure) A Borel measure μ on the phase space Γ is said to be an *invariant* (*probability*) *measure for* T if $\mu(B) = \mu(T^{-1}B)$ for any Borel set B.

Suppose μ is an invariant measure of the Andersen process; then it has to satisfy

$$\mathcal{G}^*\mu = 0$$

or, in more explicit form,

$$[\nu(\mathcal{A} - \mathcal{I}) - i\mathcal{L}]\mu = 0.$$

1.2 Discrete-Time Andersen Process

The discrete-time formulation of the Andersen process is given by the following random dynamical system:

DEFINITION 1.4 (Discrete-Time Andersen Process) Let $\mathcal{H}^{\Delta t} := \mathcal{H}(\Delta t)$, and let $\{\alpha_n\}_{n=1}^{\infty}$ be i.i.d. random variables such that $\mathbb{P}(\alpha_n = 1) = \lambda = \nu\Delta t$ and $\mathbb{P}(\alpha_n = 0) = 1 - \lambda = 1 - \nu\Delta t$; retaining the same notion of Y_n and Z_n as before, we define the *discrete Andersen dynamics* as

$$\mathbf{x}_{n+1} = (1 - \alpha_n)\mathcal{H}^{\Delta t}\mathbf{x}_n + \alpha_n \mathcal{S}(Y_n, Z_n)\mathcal{H}^{\Delta t}\mathbf{x}_n. \tag{1.3}$$

It is straightforward to write down the evolution equation for a probability measure μ under this dynamics. Indeed, let $\mathcal{H}^{-\Delta t}$ be the Markov operator (see Section 1.4 for a definition) associated with $\mathcal{H}^{\Delta t}$; we have

$$\mu_{n+1} = (1 - \lambda)\mathcal{H}^{-\Delta t}\mu_n + \lambda\mathcal{A}\mathcal{H}^{-\Delta t}\mu_n.$$

To see how this is connected to the continuous case, let us note that any invariant measure μ has to satisfy

$$\mu = (1 - \lambda)\mathcal{H}^{-\Delta t}\mu + \lambda\mathcal{A}\mathcal{H}^{-\Delta t}\mu;$$

this is equivalent to

$$\left(\frac{\mathcal{I} - e^{-i\mathcal{L}\Delta t}}{\Delta t}\right)\mu = \nu(\mathcal{A} - \mathcal{I})e^{-i\mathcal{L}\Delta t}\mu.$$

Now formally we can recover the continuous equation by letting Δt go to 0.

The canonical/Gibbsian distribution over the phase space Γ is given by

$$\pi(d\mathbf{x}) = \frac{1}{Z}e^{-\beta H(\mathbf{x})}d\mathbf{x} \tag{1.4}$$

where $H(\mathbf{x}) = \frac{1}{2}|\mathbf{v}|^2 + \Phi(\mathbf{q})$ and $Z = \int_\Gamma e^{-\beta H(\mathbf{x})} d\mathbf{x}$ is the normalization constant. It is straightforward to check that this is an invariant distribution for both the continuous- and discrete-time Andersen processes. In the following, we will prove that this is the unique invariant distribution.

1.3 Outline of the Paper

This paper presents a mathematical analysis of the Andersen thermostat. We take the viewpoint of statistical mechanics. First we prove that the Andersen processes are uniformly ergodic, with the canonical distribution being the unique invariant measure. We then study standard numerical approximations for the equations of molecular dynamics. We choose two representative numerical methods: the forward Euler method and the velocity Verlet scheme. The main difference between these two methods is that the Verlet scheme is symplectic; i.e., it preserves the underlying symplectic structure in phase space. For both schemes, we prove ergodicity, absolute continuity of the invariant distribution, and sharp error estimates for the invariant distribution. One surprising aspect of our results is that the symplectic structure does not seem to play a crucial role in the results.

Finally, we study transport properties. We first construct a simple example showing that, for some systems, the diffusion coefficient computed using molecular dynamics with the Andersen thermostat depends sensitively on the value of the collision rate in the thermostat. We then argue that this is a consequence of the fact that there are no intrinsic time scales of mixing in such systems, and if the underlying system does have intrinsic diffusion, then the diffusion coefficient computed using the Andersen thermostat does converge to the correct value in the limit of zero collision frequency in the thermostat. We prove this rigorously for the example of Lorentz gas.

1.4 Notation and Elementary Facts

The following notation will be used in a number of places:

DEFINITION 1.5 For $\mathbf{q} \in \mathbb{R}^{dN}$, denote by $\{\mathbf{q}\}$ the canonical projection of $\mathbf{q}$ into $\mathbb{D}$, i.e.,

$$\mathbf{q} \equiv \{\mathbf{q}\} \mod \mathbb{Z}^{dN} \quad \text{where } \{\mathbf{q}\} \in \mathbb{D}.$$

Similarly, we define $[\mathbf{q}] := \mathbf{q} - \{\mathbf{q}\} \in \mathbb{Z}^{dN}$.

Denote by M the set of Borel probability measures on Γ. To quantify the distance between probability measures, we introduce the following definition:

DEFINITION 1.6 (Total Variational Norm) Let μ be a finite signed measure on $(\Gamma, \mathcal{B}(\Gamma))$. The *total variational norm* [6] of μ is defined by

$$\|\mu\|_{TV} := \sup_{A \in \mathcal{B}(\Gamma)} |\mu(A)|.$$

If $\mu_1, \mu_2 \in M$ are absolutely continuous with respect to the Lebesgue measure on Γ, then we have

$$\|\mu_1 - \mu_2\|_{TV} = \frac{1}{2} \int_\Gamma |\rho_1(\mathbf{x}) - \rho_2(\mathbf{x})| d\mathbf{x}$$

where $\rho_1 = d\mu_1/d\mathbf{x}$ and $\rho_2 = d\mu_2/d\mathbf{x}$ are the densities.

We define the Markov operator as follows:

DEFINITION 1.7 (Markov Operator) Let $(X, \mathcal{B})$ be a measurable space and $T : X \to X$ a measurable transformation. Denote by M the set of Borel probability measures on X. Then $T^* : M \to M$ is said to be the *Markov operator associated with T* if for any $\mu \in M$ and $B \in \mathcal{B}$, we have

$$(T^*\mu)(B) = \mu(T^{-1}B).$$

Since we will be working with a lot of constants, let us introduce the following notation.

DEFINITION 1.8 For any two real numbers a and b, we denote by $a \lesssim b$ or $b \gtrsim a$ if there is a positive constant c such that $a \le c \cdot b$, where the constant c will depend possibly on the dimension d, the number of particles N, and the potential function Φ, but nothing else.

DEFINITION 1.9 For $\mathbf{x} = (\mathbf{q}, \mathbf{v}) = (\mathbf{q}_1, \dots, \mathbf{q}_N, \mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathbb{R}^{2dN}$, define projection operators $\mathcal{Q}_j$ and $\mathcal{P}_j$, $1 \le j \le N$, by

$$\mathcal{Q}_j \mathbf{x} := \mathbf{q}_j \quad \text{and} \quad \mathcal{P}_j \mathbf{x} := \mathbf{v}_j.$$

Similarly, we define $\mathcal{Q}\mathbf{x} := \mathbf{q}$ and $\mathcal{P}\mathbf{x} := \mathbf{v}$.

DEFINITION 1.10 For ease of notation, in some places we write

$$\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_N) := -\nabla\Phi,$$

where Φ is the potential.

Let us also mention the trivial fact that for any two square matrices A and B, we have

$$\|AB\| \lesssim \|A\|\|B\|.$$

Although all norms in finite-dimensional spaces are equivalent, in some parts of Section 3 and Section 4 we shall choose to work with a fixed matrix norm $\|\cdot\|$ that is submultiplicative, i.e.,

$$\|AB\| \le \|A\|\|B\|.$$

The advantage of doing this is that we can keep track of constants that might blow up if iterated too many times.

The following two coupling inequalities are well-known (see [8] for a proof).

LEMMA 1.11 (Coupling Inequality for Markov Chains) *Let P be the transition kernel of a Markov chain on the state space Γ. Suppose there is a probability measure μ on Γ, a positive constant $\epsilon < 1$, and an integer $n_0 \geq 1$ such that*

$$P^{n_0}(\mathbf{x}, \cdot\,) \geq \epsilon\mu(\cdot) \quad \text{for any } \mathbf{x} \in \Gamma.$$

Then there exists a unique invariant probability measure π such that for any $n \geq 1$ and $\mathbf{x} \in \Gamma$, the following inequality holds:

$$\|P^n(\mathbf{x}, \cdot\,) - \pi(\cdot)\|_{TV} \leq (1-\epsilon)^{\frac{n}{n_0}}.$$

Consequently, there exists a constant $c > 0$ such that

$$\|P^n(\mathbf{x}, \cdot\,) - \pi(\cdot)\|_{TV} \leq c(1-\epsilon)^n.$$

LEMMA 1.12 (Coupling Inequality for Continuous-Time Markov Processes) *Let P^t be the transition semigroup of a continuous-time Markov process on the state space Γ. Assume that P^t has a stationary probability measure π. Also assume that there exists $t_0 > 0$ such that the sampled Markov chain P^{nt_0} satisfies the hypothesis of Lemma* 1.11 *with corresponding (ϵ, μ); then we have for any $t > 0$ and $\mathbf{x} \in \Gamma$,*

$$\|P^t(\mathbf{x}, \cdot\,) - \pi(\cdot)\|_{TV} \leq (1-\epsilon)^{\frac{t}{t_0}}.$$

Consequently, there exist positive constants c and κ such that

$$\|P^t(\mathbf{x}, \cdot\,) - \pi(\cdot)\|_{TV} \leq c\exp(-\kappa t).$$

2 Ergodicity of the Andersen Process

2.1 Continuous-Time Andersen Process, $N = 1$

To build intuition and prepare ourselves for the many-particle case, we shall first prove that the continuous-time Andersen process is uniformly ergodic in the case when $N = 1$. In a later section we shall deal with $N > 1$; there the idea is similar but the proof becomes somewhat involved as we shall see shortly.

LEMMA 2.1 *For each $\mathbf{q}' \in \mathbb{D}$, consider the map: $\mathbf{v}' \to \mathbf{q}^t(\mathbf{q}', \mathbf{v}') = \mathcal{H}_1^t(\mathbf{q}', \mathbf{v}') \in \mathbb{R}^d$. There exists $t_0 > 0$ and positive constants c_1 and c_2, independent of t, $\mathbf{q}'$, and $\mathbf{v}'$, such that for all $0 < t < t_0$,*

$$c_1 t^d \leq \left|\det\left(\frac{\partial \mathbf{q}^t}{\partial \mathbf{v}'}\right)\right| \leq c_2 t^d. \tag{2.1}$$

Furthermore, the map is globally smoothly invertible with its inverse denoted by $\phi^t_{\mathbf{q}'}$, and there exist constants $c_3, c_4, c_5, c_6 > 0$, independent of $(\mathbf{q}, \mathbf{q}', \mathbf{v}, t)$, such that

$$c_3 t^{-d} \leq \left|\det\left(\frac{\partial \phi^t_{\mathbf{q}'}}{\partial \mathbf{q}}\right)\right| \leq c_4 t^{-d} \tag{2.2}$$

and

$$\frac{|\mathbf{q}| - c_5}{t} \leq |\phi^t_{\mathbf{q}'}(\mathbf{q})| \leq \frac{|\mathbf{q}| + c_6}{t}. \tag{2.3}$$

PROOF: By (1.1), we have

$$\mathbf{q}^t = \mathbf{q}' + \mathbf{v}'t + \int_0^t \int_0^s \mathbf{F}(\mathbf{q}^\tau) d\tau\, ds.$$

Differentiating with respect to $\mathbf{v}'$, we get

$$\frac{\partial \mathbf{q}^t}{\partial \mathbf{v}'} = t + \int_0^t \int_0^s \frac{\partial \mathbf{F}}{\partial \mathbf{q}^\tau} . \frac{\partial \mathbf{q}^\tau}{\partial \mathbf{v}'} d\tau\, ds.$$

Now note that $\frac{\partial \mathbf{q}^t}{\partial \mathbf{v}'}\Big|_{t=0} = I$ and $\|\frac{\partial \mathbf{F}}{\partial \mathbf{q}}\|$ is absolutely bounded (by assumption). A simple Gronwall argument immediately gives (2.1). By a result of Chichilnisky [4], we conclude that the map $\mathbf{v}' \longrightarrow \mathbf{q}^t$ is globally smoothly invertible and (2.2) holds. The last inequality (2.3) is obvious from using the fact that $\mathbf{q}' \in \mathbb{D}$. □

LEMMA 2.2 *Let $t_0 > 0$ be the same as in Lemma 2.1, and let $\delta_{\mathbf{x}'}$ be the Dirac measure on Γ concentrated at a point $\mathbf{x}' = (\mathbf{q}', \mathbf{v}')$. Then for any $0 < t < t_0$ and $\mathbf{x}' \in \Gamma$, there exists a positive constant $\eta_t < 1$ independent of $\mathbf{x}'$ and a reference probability measure μ^{ref} independent of $(\mathbf{x}', t)$ such that*

$$\mathcal{A}\mathcal{H}^{-t}\mathcal{A}\delta_{\mathbf{x}'} \geq \eta_t \mu^{\text{ref}}.$$

PROOF: For any $\mathbf{q}' \in \mathbb{D}$, a short computation gives

$$(\mathcal{A}\mathcal{H}^{-t}\mathcal{A}\delta_{\mathbf{x}'})(d\mathbf{q}\, d\mathbf{v}) = g_\beta(\mathbf{v}) \cdot K(\mathbf{q}', \mathbf{q}) d\mathbf{q}\, d\mathbf{v},$$

where the stochastic kernel K is given by

$$K(\mathbf{q}', \mathbf{q}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} g_\beta(\phi^t_{\mathbf{q}'}(\mathbf{q} + \mathbf{k})) \cdot \left| \det\left(\frac{\partial \phi^t_{\mathbf{q}'}(\tilde{\mathbf{q}})}{\partial \tilde{\mathbf{q}}} \bigg|_{\tilde{\mathbf{q}} = \mathbf{q} + \mathbf{k}} \right) \right|.$$

Clearly by Lemma 2.1, for any positive $t < t_0$, there exists a positive constant $\eta_t < 1$ such that

$$K(\mathbf{q}', \mathbf{q}) \geq \eta_t \quad \text{for any } \mathbf{q}', \mathbf{q} \in \mathbb{D}.$$

Now the lemma follows by simply taking $\mu^{\text{ref}} := d\mathbf{q} \otimes g_\beta(\mathbf{v}) d\mathbf{v}$. □

Let us note that from the proof above, the constant η_t satisfies the following relation:

$$\eta_t \geq \frac{\text{const}}{t^d} \exp\left(-\frac{\text{const}}{t^2} \right).$$

This important fact will be used in the proof of the following theorem:

THEOREM 2.3 *Denote by $P^t_{\mathbf{x}'} = P^t(\mathbf{x}', \cdot)$ the Markov semigroup of the Andersen process. Let π be the stationary Gibbsian probability measure for the Andersen process. Then the continuous-time Andersen process is uniformly ergodic: for any $\nu_0 > 0$, there is a constant $\kappa = \kappa(\nu_0) > 0$, independent of $\mathbf{x}'$, such that for any $0 < \nu \leq \nu_0$,*

$$\|P^t_{\mathbf{x}'} - \pi\|_{TV} \leq c \cdot \exp(-\kappa \nu^2 t)$$

where c is a positive constant independent of $(\mathbf{x}', \nu_0, \nu)$.

PROOF: Note that with initial condition $P^0_{\mathbf{x}'} = \delta_{\mathbf{x}'}$, $P^t_{\mathbf{x}'}$ can be viewed as the mild solution to the following Duhamel equation:

$$P^t_{\mathbf{x}'} = e^{-\nu t}\mathcal{H}^{-t}P^0_{\mathbf{x}'} + \nu\int_0^t e^{\nu(s-t)}\mathcal{H}^{-(t-s)}\mathcal{A}P^s_{\mathbf{x}'}\,ds.$$

Let $T^s := e^{-\nu s}\mathcal{H}^{-t}$; two simple iterations produce the inequality

$$P^t_{\mathbf{x}'} \geq \nu^2\int_0^t\int_0^{t_1} T^{t-t_1}\mathcal{A}T^{t_1-t_2}\mathcal{A}T^{t_2}P^0_{\mathbf{x}'}\,dt_2\,dt_1.$$

Let $\varepsilon > 0$ be fixed and chosen sufficiently small. Let $t^* = \min\{t_0/2, 1\}$, where t_0 is defined in Lemma 2.1. Observe that

$$\begin{aligned}P^{t^*}_{\mathbf{x}'} &\geq \nu^2\int_{\varepsilon t^*}^{t^*}\int_0^{t_1-\varepsilon t_1} T^{t^*-t_1}\mathcal{A}T^{t_1-t_2}\mathcal{A}T^{t_2}P^0_{\mathbf{x}'}\,dt_2\,dt_1\\ &\geq \nu^2 e^{-\nu t^*}\int_{\varepsilon t^*}^{t^*}\int_0^{t_1-\varepsilon t_1}\mathcal{H}^{-(t^*-t_1)}\mathcal{A}\mathcal{H}^{-(t_1-t_2)}\mathcal{A}\mathcal{H}^{-t_2}P^0_{\mathbf{x}'}\,dt_2\,dt_1.\end{aligned}$$

Note that for $\varepsilon t^* \leq t_1 \leq t^*$ and $0 \leq t_2 \leq t_1 - \varepsilon t_1$, we have $\varepsilon^2 t^* \leq t_1 - t_2 \leq t^*$. Therefore by Lemma 2.2 and the fact that $\eta_t \geq \frac{\text{const}}{t^d}\exp(-\frac{\text{const}}{t^2})$, we conclude that there exists a constant $\theta(t^*) > 0$ such that for all $\varepsilon t^* \leq t_1 \leq t^*$, $0 \leq t_2 \leq t_1 - \varepsilon t_1$,

$$\mathcal{A}\mathcal{H}^{-(t_1-t_2)}\mathcal{A}\mathcal{H}^{-t_2}P^0_{\mathbf{x}'} \geq \theta(t^*)\mu^{\text{ref}}.$$

Let $\tilde{\mu}^{\text{ref}}$ be the probability measure such that

$$\tilde{\mu}^{\text{ref}} \propto \int_{\varepsilon t^*}^{t^*}\int_0^{t_1-\varepsilon t_1}\mathcal{H}^{-(t^*-t_1)}\mu^{\text{ref}}\,dt_2\,dt_1.$$

Then clearly there exists a constant $\tilde{\theta}(t^*) > 0$ such that the following holds:

$$P^{t^*}_{\mathbf{x}'} \geq (\nu t^*)^2 e^{-\nu t^*}\tilde{\theta}(t^*)\tilde{\mu}^{\text{ref}}.$$

Now by Lemma 1.12, we have for any $t > 0$,

$$\begin{aligned}\|P^t_{\mathbf{x}'} - \pi\|_{TV} &\leq (1 - (\nu t^*)^2 e^{-\nu t^*}\tilde{\theta}(t^*))^{[\frac{t}{t^*}]}\\ &\leq (1 - (\nu t^*)^2 e^{-\nu t^*}\tilde{\theta}(t^*))^{-1}\cdot(1 - (\nu t^*)^2 e^{-\nu t^*}\tilde{\theta}(t^*))^{\frac{t}{t^*}}\\ &\leq c\cdot(1 - (\nu t^*)^2 e^{-\nu t^*}\tilde{\theta}(t^*))^{\frac{t}{t^*}},\end{aligned}$$

where c is a constant depending only on t^*.

Let us consider the function

$$f(\nu) = (1 - (\nu t^*)^2 e^{-\nu t^*}\tilde{\theta}(t^*))^{\frac{1}{\nu^2}}.$$

Clearly there exists a positive constant $d = d(t^*, \nu_0) < 1$ such that for any $\nu \leq \nu_0$,

$$f(\nu) \leq d < 1.$$

Now take $\kappa = -(\log d)/t^*$ and the conclusion of the theorem follows easily. □

2.2 Discrete-Time Andersen Process, $N = 1$

Similar to the continuous case, we shall prove that the discrete-time Andersen process is uniformly ergodic.

LEMMA 2.4 *Let t_0 be the same as in Lemma 2.1. Let $\delta_{\mathbf{x}'}$ be the Dirac measure at $\mathbf{x}' \in \Gamma$. Then for any $0 \le t_1 \le \frac{t_0}{2}$, $\frac{t_0}{6} \le t_2 \le \frac{t_0}{2}$, and $t_3 \ge 0$, there exists a positive constant $c = c(t_0)$ and a reference probability measure $\mu_{t_0}^{\text{ref}}$, both independent of $(t_1, t_2, t_3, \mathbf{x}')$, such that*

$$\mathcal{H}^{-t_1}\mathcal{A}\mathcal{H}^{-t_2}\mathcal{A}\mathcal{H}^{-t_3}\delta_{\mathbf{x}'} \ge c\mu_{t_0}^{\text{ref}}.$$

PROOF: By Lemma 2.2, there exists a positive constant η_{t_0} and a probability measure $d\mu^{\text{ref}} = d\mathbf{q} \otimes g_\beta(\mathbf{v})d\mathbf{v}$ such that

$$\mathcal{A}\mathcal{H}^{-t_2}\mathcal{A}\mathcal{H}^{-t_3}\delta_{\mathbf{x}'} \ge \eta_{t_0}\mu^{\text{ref}}.$$

Now for $t_1 \le \frac{t_0}{2}$ and $\mathbf{x} = (\mathbf{q}, \mathbf{v}) \in \Gamma$, we have

$$|\mathcal{H}_2^{-t_1}\mathbf{x}| \le |\mathbf{v}| + \text{const} \cdot t_0$$

where $\mathcal{H}_2^{-t_1}\mathbf{x}$ is the velocity component of $\mathcal{H}^{-t_1}\mathbf{x}$. This simple estimate immediately gives

$$\begin{aligned}(\mathcal{H}^{-t_1}\mu^{\text{ref}})(d\mathbf{x}) &= g_\beta(\mathcal{H}_2^{-t_1}\mathbf{x})d\mathbf{x} \\ &\ge c \cdot \mu_{t_0}^{\text{ref}}(d\mathbf{x})\end{aligned}$$

where $\mu_{t_0}^{\text{ref}} = d\mathbf{q} \otimes g_{\beta_1}(\mathbf{v})d\mathbf{v}$ and $\beta_1 = (1 + t_0)\beta$. □

With an obvious abuse of notation, let us denote

$$P_{\mathbf{x}'}^{\Delta t} := ((1 - \nu\Delta t)\mathcal{H}^{-\Delta t} + \nu\Delta t\mathcal{A}\mathcal{H}^{-\Delta t})\delta_{\mathbf{x}'}.$$

It is not hard to arrive at the following theorem:

THEOREM 2.5 *Let t_0 be the same as in Lemma 2.1. Let $\nu_0 > 0$ be arbitrary but fixed. Let $\Delta t_0 = t_0/7$. Then for any $0 < \nu < \nu_0$, $0 < \Delta t < \min\{1/2\nu_0, \Delta t_0/4\}$, there exists a constant $\kappa = \kappa(\nu_0)$, independent of $(\Delta t, \nu)$, such that for any $n \ge 1$,*

$$\|(P_{\mathbf{x}'}^{\Delta t})^n - \pi\|_{TV} \le c \cdot \exp(-\kappa\nu^2 n\Delta t)$$

where c is a positive constant independent of $(\nu_0, \Delta t, \nu)$.

PROOF: Let $m = [\Delta t_0/\Delta t]$; clearly $m \ge 4$. Let us also denote

$$S = \{(n_1, n_2, n_3) \in \mathbb{Z}^3 : n_1 \ge 0,\ n_2 \ge m,\ n_3 \ge 0,\ n_1 + n_2 + n_3 = 3m - 2\}.$$

We have

$$\begin{aligned}(P_{\mathbf{x}'}^{\Delta t})^{3m} &= ((1 - \nu\Delta t)\mathcal{H}^{-\Delta t} + \nu\Delta t\mathcal{A}\mathcal{H}^{-\Delta t})^{3m}\delta_{\mathbf{x}'} \\ &\ge \sum_{(n_1,n_2,n_3)\in S} (\nu\Delta t)^2(1 - \nu\Delta t)^{3m-2}\mathcal{H}^{-n_1\Delta t}\mathcal{A}\mathcal{H}^{-(n_2+1)\Delta t}\mathcal{A}\mathcal{H}^{-(n_3+1)\Delta t}\delta_{\mathbf{x}'}.\end{aligned}$$

Observe that for any $(n_1, n_2, n_3) \in S$, we have

$$n_1 \Delta t \le \frac{t_0}{2} \quad \text{and} \quad \frac{t_0}{6} \le (n_2 + 1)\Delta t \le \frac{t_0}{2}.$$

Then by Lemma 2.4, there exist $c_1 = c_1(t_0) > 0$ and $\mu_{t_0}^{\text{ref}}$ such that the following holds for any $(n_1, n_2, n_3) \in S$:

$$\mathcal{H}^{-n_1\Delta t}\mathcal{A}\mathcal{H}^{-(n_2+1)\Delta t}\mathcal{A}\mathcal{H}^{-(n_3+1)\Delta t}\delta_{\mathbf{x}'} \ge c_1 \cdot \mu_{t_0}^{\text{ref}}.$$

Since $m \ge 4$, we have $|S| \ge 2m^2$, so that

$$(P_{\mathbf{x}'}^{\Delta t})^{3m} \ge c_1 \cdot 2m^2(\nu\Delta t)^2(1 - \nu\Delta t)^{3m-2}\mu_{t_0}^{\text{ref}}.$$

Now by Lemma 1.11, we have

$$\begin{aligned} \|(P_{\mathbf{x}'}^{\Delta t})^n - \pi\|_{TV} &\le (1 - c_1 \cdot 2m^2\nu^2\Delta t^2(1 - \nu\Delta t)^{3m-2})^{[\frac{n}{m}]} \\ &\le \left(1 - \frac{c_1}{2}\nu^2\Delta t_0^2(1 - \nu\Delta t)^{3\Delta t_0/\Delta t}\right)^{[\frac{n}{m}]} \\ &\le f(\nu, \Delta t_0)^{-1} \cdot f(\nu, \Delta t_0)^{\frac{n}{m}} \\ &\le c_2 \cdot f(\nu, \Delta t_0)^{n/m}, \end{aligned}$$

where $c_2 > 0$ is independent of $(\nu, \nu_0, \Delta t, n)$ and $f(\nu, \Delta t_0)$ is given by

$$f(\nu, \Delta t_0) = 1 - \frac{c_1}{2}\nu^2\Delta t_0^2 \cdot \exp(-6\nu\Delta t_0).$$

It is easy to show that for any $0 < \nu < \nu_0$, there exists a positive constant $d = d(\Delta t_0, \nu_0) < 1$ such that

$$f(\nu, \Delta t_0) \le d^{\nu^2} < 1,$$

and the desired inequality follows easily. □

2.3 Continuous-Time Andersen Process, $N > 1$

LEMMA 2.6 *There exists a constant $\Delta t_0 > 0$ such that for any $0 < \Delta t < \Delta t_0$, $1 \le i, j \le N$, and $\mathbf{x} \in \Gamma$, the following holds*:

$$\begin{aligned} \left\| \frac{\partial(\mathcal{Q}_j\mathcal{H}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})} - \delta_{ij}I \right\| &\lesssim \Delta t^2, \\ \left\| \frac{\partial(\mathcal{P}_j\mathcal{H}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} - \delta_{ij}I \right\| &\lesssim \Delta t^2, \\ \left\| \frac{\partial(\mathcal{Q}_j\mathcal{H}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} - \Delta t\delta_{ij}I \right\| &\lesssim \Delta t^3, \\ \left\| \frac{\partial(\mathcal{P}_j\mathcal{H}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})} \right\| &\lesssim \Delta t. \end{aligned}$$

PROOF: By (1.1), we have

$$\begin{aligned}\mathcal{Q}\mathcal{H}^t\mathbf{x} &= \mathcal{Q}\mathbf{x} + t\mathcal{P}\mathbf{x} + \int_0^t\int_0^s \mathbf{F}(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})d\tau\, ds,\\ \mathcal{P}\mathcal{H}^t\mathbf{x} &= \mathcal{P}\mathbf{x} + \int_0^t \mathbf{F}(\mathcal{Q}\mathcal{H}^s\mathbf{x})ds.\end{aligned}$$

Differentiating the above equations, we get

$$\begin{aligned}\frac{\partial(\mathcal{Q}\mathcal{H}^t\mathbf{x})}{\partial(\mathcal{Q}\mathbf{x})} &= I + \int_0^t\int_0^s \frac{\partial\mathbf{F}(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}\cdot\frac{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{Q}\mathbf{x})}d\tau\, ds,\\ \frac{\partial(\mathcal{Q}\mathcal{H}^t\mathbf{x})}{\partial(\mathcal{P}\mathbf{x})} &= t + \int_0^t\int_0^s \frac{\partial\mathbf{F}(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}\cdot\frac{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{P}\mathbf{x})}d\tau\, ds,\\ \frac{\partial(\mathcal{P}\mathcal{H}^t\mathbf{x})}{\partial(\mathcal{Q}\mathbf{x})} &= \int_0^t \frac{\partial\mathbf{F}(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}\cdot\frac{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{Q}\mathbf{x})}d\tau,\\ \frac{\partial(\mathcal{P}\mathcal{H}^t\mathbf{x})}{\partial(\mathcal{P}\mathbf{x})} &= I + \int_0^t \frac{\partial\mathbf{F}(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}\cdot\frac{\partial(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})}{\partial(\mathcal{P}\mathbf{x})}d\tau.\end{aligned}$$

Now a simple Gronwall argument gives the result, noting that $\|\frac{\partial\mathbf{F}}{\partial\mathbf{q}}\|$ is bounded (by assumption). We omit the details. □

LEMMA 2.7 *Let Δt_0 be the same as in Lemma 2.6. Let t_i and s_i, $1 \le i \le N$, be a sequence of nonnegative numbers such that $\sum_{i=1}^N (t_i + s_i) = \Delta t \le \Delta t_0$. Let $\mathbf{x}^{(0)} = (\mathbf{q}_1^{(0)}, \dots, \mathbf{q}_N^{(0)}, \mathbf{v}_1^{(0)}, \dots, \mathbf{v}_N^{(0)})$. For $i = 1, \dots, N$, recursively define*

$$\begin{aligned}\mathbf{x}^{(2i-1)} &:= \mathcal{H}^{t_i}\mathcal{S}(i, \mathbf{u}_i)\mathbf{x}^{(2i-2)},\\ \mathbf{x}^{(2i)} &:= \mathcal{H}^{s_i}\mathcal{S}(i, \mathbf{u}_{N+i})\mathbf{x}^{(2i-1)}.\end{aligned}$$

Fix $1 \le i_0 \le N$. Then for any $m \ge 2i_0$ and $1 \le j \le N$, the following holds:

$$\begin{aligned}\left\|\frac{\partial(\mathcal{Q}_j\mathbf{x}^{(m)})}{\partial\mathbf{u}_{i_0}} - t_{i_0}\delta_{ji_0}I\right\| &\lesssim \Delta t^2,\\ \left\|\frac{\partial(\mathcal{P}_j\mathbf{x}^{(m)})}{\partial\mathbf{u}_{i_0}}\right\| &\lesssim \Delta t,\\ \left\|\frac{\partial(\mathcal{Q}_j\mathbf{x}^{(m)})}{\partial\mathbf{u}_{N+i_0}}\right\| &\lesssim \Delta t,\\ \left\|\frac{\partial(\mathcal{P}_j\mathbf{x}^{(m)})}{\partial\mathbf{u}_{N+i_0}} - \delta_{ji_0}I\right\| &\lesssim \Delta t.\end{aligned}$$

PROOF: We shall prove the lemma inductively.

(1) For $m = 2i_0$, define $\mathbf{y} := \mathcal{S}(i_0, \mathbf{u}_{i_0})x^{(2i_0-2)}$, $\mathbf{y}' := \mathcal{S}(i_0, \mathbf{u}_{N+i_0})\mathbf{x}^{(2i_0-1)}$. By Lemma 2.6, we have

$$\begin{aligned}
&\left\| \frac{\partial \mathcal{Q}_j \mathbf{x}^{(2i_0)}}{\partial \mathbf{u}_{i_0}} - t_{i_0}\delta_{ji_0} I \right\| \\
&\quad = \left\| \sum_{k\neq i_0} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{P}_k\mathbf{y}')} \frac{\partial(\mathcal{P}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} + \sum_{k\neq i_0} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{Q}_k\mathbf{y}')} \frac{\partial(\mathcal{Q}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} \right. \\
&\qquad \left. + \left(\frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{Q}_{i_0}\mathbf{y}')} - I\delta_{ji_0} \right) \frac{\partial(\mathcal{Q}_{i_0} \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} + \delta_{ji_0} I \left(\frac{\partial(\mathcal{Q}_{i_0} \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} - t_{i_0} I \right) \right\| \\
&\quad \lesssim \sum_{k\neq i_0} \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{P}_k\mathbf{y}')} \right\| \left\| \frac{\partial(\mathcal{P}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} \right\| + \sum_{k\neq i_0} \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{Q}_k\mathbf{y}')} \right\| \left\| \frac{\partial(\mathcal{Q}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} \right\| \\
&\qquad + \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{Q}_{i_0}\mathbf{y}')} - I\delta_{ji_0} \right\| \left\| \frac{\partial(\mathcal{Q}_{i_0} \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} \right\| + \left\| \delta_{ji_0} I \left(\frac{\partial(\mathcal{Q}_{i_0} \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} - t_{i_0} I \right) \right\| \\
&\quad \lesssim \sum_{k\neq i_0} \Delta t \cdot \Delta t + \sum_{k\neq i_0} \Delta t^3 + \Delta t^2 \cdot \Delta t + \Delta t^3 \\
&\quad \lesssim \Delta t^2
\end{aligned}$$

Now for the remaining inequalities, we have by the previous lemma

$$\begin{aligned}
&\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}^{(2i_0)})}{\partial \mathbf{u}_{i_0}} \right\| \\
&\quad = \left\| \sum_{k=1}^{N} \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{Q}_k\mathbf{y}')} \frac{\partial(\mathcal{Q}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} + \sum_{k\neq i_0} \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{P}_k\mathbf{y}')} \frac{\partial(\mathcal{P}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} \right\| \\
&\quad \lesssim \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{Q}_k\mathbf{y}')} \right\| \left\| \frac{\partial(\mathcal{Q}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} \right\| + \sum_{k\neq i_0} \left\| \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{P}_k\mathbf{y}')} \right\| \left\| \frac{\partial(\mathcal{P}_k \mathcal{H}^{t_{i_0}}\mathbf{y})}{\partial(\mathcal{P}_{i_0}\mathbf{y})} \right\| \\
&\quad \lesssim \sum_{k=1}^{N} \Delta t + \sum_{k\neq i_0} \Delta t \lesssim \Delta t
\end{aligned}$$

and

$$\left\| \frac{\partial(\mathcal{Q}_j \mathbf{x}^{(2i_0)})}{\partial \mathbf{u}_{N+i_0}} \right\| = \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{P}_{i_0}\mathbf{y}')} \right\| \lesssim \Delta t,$$

$$\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}^{(2i_0)})}{\partial \mathbf{u}_{N+i_0}} - \delta_{ji_0} I \right\| = \left\| \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{i_0}}\mathbf{y}')}{\partial(\mathcal{P}_{i_0}\mathbf{y}')} - \delta_{ji_0} I \right\| \lesssim \Delta t.$$

(2) Now assume for $m = l \geq 2i_0$ the lemma is true. We want to show it is true for $m = l + 1$.

First consider the case when l is odd. Denote $n_0 := \frac{l+1}{2}$; then by definition

$$\mathbf{x}^{(l+1)} = \mathcal{H}^{s_{n_0}} \mathcal{S}(n_0, \mathbf{u}_{n_0+N}) \mathbf{x}^{(l)}.$$

Denote $\mathbf{y} := \mathcal{S}(n_0, \mathbf{u}_{N+n_0})\mathbf{x}^{(l)}$. Then by the induction hypothesis and the previous lemma, we have

$$\begin{aligned}
&\left\| \frac{\partial \mathcal{Q}_j \mathbf{x}^{(l+1)}}{\partial \mathbf{u}_{i_0}} - t_{i_0}\delta_{ji_0} I \right\| \\
&\quad = \left\| \sum_{k=1}^{N} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \frac{\partial(\mathcal{Q}_k \mathbf{y})}{\partial \mathbf{u}_{i_0}} + \sum_{k\neq i_0} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \frac{\partial(\mathcal{P}_k \mathbf{y})}{\partial \mathbf{u}_{i_0}} - t_{i_0}\delta_{ji_0} I \right\| \\
&\quad = \left\| \sum_{k\neq i_0} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} + \sum_{k\neq i_0} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} \right. \\
&\qquad \left. + \left(\frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_{i_0} \mathbf{y})} - I\delta_{ji_0} \right) \frac{\partial(\mathcal{Q}_{i_0} \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} + \delta_{ji_0} I \left(\frac{\partial(\mathcal{Q}_{i_0} \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} - t_{i_0} I \right) \right\| \\
&\quad \lesssim \sum_{k\neq i_0} \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \right\| \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} \right\| + \sum_{k\neq i_0} \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \right\| \left\| \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} \right\| \\
&\qquad + \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_{i_0} \mathbf{y})} - I\delta_{ji_0} \right\| \left\| \frac{\partial(\mathcal{Q}_{i_0} \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} \right\| + \left\| \delta_{ji_0} I \left(\frac{\partial(\mathcal{Q}_{i_0} \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} - t_{i_0} I \right) \right\| \\
&\lesssim \sum_{k\neq i_0} \Delta t^2 + \sum_{k\neq i_0} \Delta t \cdot \Delta t + \Delta t^2 \cdot \Delta t + \Delta t^2 \\
&\lesssim \Delta t^2.
\end{aligned}$$

Similarly, we can prove the remaining inequalities:

$$\begin{aligned}
&\left\| \frac{\partial \mathcal{P}_j \mathbf{x}^{(l+1)}}{\partial \mathbf{u}_{i_0}} \right\| \\
&\quad = \left\| \sum_{k=1}^{N} \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} + \sum_{k\neq i_0} \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} \right\| \\
&\quad \lesssim \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \right\| \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} \right\| + \sum_{k\neq i_0} \left\| \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \right\| \left\| \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{i_0}} \right\| \\
&\quad \lesssim \sum_{k=1}^{N} \Delta t \cdot \Delta t + \sum_{k\neq i_0} \Delta t \lesssim \Delta t,
\end{aligned}$$

$$\begin{aligned}
&\left\| \frac{\partial \mathcal{Q}_j \mathbf{x}^{(l+1)}}{\partial \mathbf{u}_{N+i_0}} \right\| \\
&\quad = \left\| \sum_{k=1}^{N} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} + \sum_{k\neq i_0} \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} \right\|
\end{aligned}$$

$$\lesssim \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \right\| \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} \right\| + \sum_{k \neq i_0} \left\| \frac{\partial(\mathcal{Q}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \right\| \left\| \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} \right\|$$

$$\lesssim \sum_{k=1}^{N} \Delta t + \sum_{k \neq i_0} \Delta t \cdot \Delta t \lesssim \Delta t,$$

and

$$\begin{aligned}
&\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}^{(l+1)})}{\partial \mathbf{u}_{N+i_0}} - \delta_{j i_0} I \right\| \\
&\quad = \left\| \sum_{k=1}^{N} \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} + \sum_{k \neq i_0} \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} - \delta_{j i_0} I \right\| \\
&\quad = \left\| \sum_{k=1}^{N} \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} \right. \\
&\qquad \left. + \sum_{k \neq i_0} \left(\frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} - \delta_{jk} I \right) \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} + \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} - \delta_{j i_0} I \right\| \\
&\quad \lesssim \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{Q}_k \mathbf{y})} \right\| \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} \right\| \\
&\qquad + \sum_{k \neq i_0} \left\| \frac{\partial(\mathcal{P}_j \mathcal{H}^{s_{n_0}} \mathbf{y})}{\partial(\mathcal{P}_k \mathbf{y})} - \delta_{jk} I \right\| \left\| \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} \right\| + \left\| \frac{\partial(\mathcal{P}_k \mathbf{x}^{(l)})}{\partial \mathbf{u}_{N+i_0}} - \delta_{j i_0} I \right\| \\
&\quad \lesssim \sum_{k=1}^{N} \Delta t \cdot \Delta t + \sum_{k \neq i_0} \Delta t \cdot \Delta t + \Delta t \lesssim \Delta t.
\end{aligned}$$

The situation when l is even is entirely similar. We omit the details here.

□

LEMMA 2.8 *With the same notation as in the previous lemma, the following holds for any* $1 \le i_0 \le N$ *and* $2i_0 \le m \le 2N$:

$$\begin{aligned}
\Bigg| \mathcal{Q}_{i_0} \mathbf{x}^{(m)} - \left(\mathcal{Q}_{i_0} \mathbf{x}^{(0)} + \sum_{i=1}^{i_0 - 1} (t_i + s_i) \mathcal{P}_{i_0} \mathbf{x}^{(0)} \right) - t_{i_0} \mathbf{u}_{i_0} \\
- \left(\sum_{i=i_0+1}^{[\frac{m+1}{2}]} t_i + \sum_{i=i_0}^{[\frac{m}{2}]} s_i \right) \mathbf{u}_{N+i_0} \Bigg| \lesssim \Delta t^2,
\end{aligned}$$

$$|\mathcal{P}_{i_0} \mathbf{x}^{(m)} - \mathbf{u}_{N+i_0}| \lesssim \Delta t.$$

PROOF: This is quite straightforward. □

COROLLARY 2.9 *We make the same assumptions as in Lemma* 2.7. *In addition, we assume that*

$$\frac{\Delta t}{100N} \le \min_{1\le i\le N} t_i \le \max_{1\le i\le N} t_i \le \frac{\Delta t}{10N}.$$

Then there exists a constant $\Delta t_1 > 0$ *such that for any* $0 < \Delta t \le \Delta t_1$, *the map*

$$\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_{2N}) \in \mathbb{R}^{2dN} \to \mathbf{x}^{(2N)}$$

is a C^∞ *homeomorphism from* $\mathbb{R}^{2dN}$ *to* $\mathbb{R}^{2dN}$. *Also, there exist positive constants* c_1 *and* c_2 *such that the following holds*:

$$c_1 \Delta t^{dN} \le \left|\det\left(\frac{\partial \mathbf{x}^{(2N)}}{\partial \mathbf{u}}\right)\right| \le c_2 \Delta t^{dN}.$$

Consequently, if we denote by $\phi^{\Delta t}$ *the inverse map, then there exist constants* $c_3, c_4 > 0$ *such that*

$$c_3 \Delta t^{-dN} \le \left|\det\left(\frac{\partial \phi^{\Delta t}}{\partial \mathbf{x}}\right)\right| \le c_4 \Delta t^{-dN}.$$

PROOF: Clearly by Lemma 2.7, we have, for any $1 \le i, j \le N$, that there exists a positive constant d_1 such that

$$\left\|\frac{\partial(\mathcal{Q}_j \mathbf{x}^{(2N)})}{\partial \mathbf{u}_i} - t_i \delta_{ji} I\right\| \le d_1 \Delta t^2,$$

$$\left\|\frac{\partial(\mathcal{P}_j \mathbf{x}^{(2N)})}{\partial \mathbf{u}_i}\right\| \le d_1 \Delta t,$$

$$\left\|\frac{\partial(\mathcal{Q}_j \mathbf{x}^{(2N)})}{\partial \mathbf{u}_{N+i}}\right\| \le d_1 \Delta t,$$

$$\left\|\frac{\partial(\mathcal{P}_j \mathbf{x}^{(2N)})}{\partial \mathbf{u}_{N+i}} - \delta_{ji} I\right\| \le d_1 \Delta t.$$

By the assumptions on t_i, we have that there exist constants $\tilde{d}_i$, $i = 1, 2, 3, 4$, such that

$$\tilde{d}_1 \Delta t^{dN} - \tilde{d}_2 \Delta t^{dN+1} \le \det\left(\frac{\partial \mathbf{x}^{(2N)}}{\partial \mathbf{u}}\right) \le \tilde{d}_3 \Delta t^{dN} + \tilde{d}_4 \Delta t^{dN+1}.$$

Clearly by taking $\Delta t_1 > 0$ sufficiently small, we have for any $0 < \Delta t \le \Delta t_1$,

$$c_1 \Delta t^{dN} \le \left|\det\left(\frac{\partial \mathbf{x}^{(2N)}}{\partial \mathbf{u}}\right)\right| \le c_2 \Delta t^{dN}$$

for some constants $c_1 > 0$ and $c_2 > 0$. We omit the smoothness proof since it is quite obvious. □

COROLLARY 2.10 *We make the same assumptions as in Corollary* 2.9. *Then there exists a positive constant* $c < 1$ *and a probability measure* μ^{ref}, *both depending only on* Δt, *such that for any* $\mathbf{x}' \in \Gamma$ *we have*

$$\prod_{j=1}^{N} \big(\mathcal{H}^{-s_j}\mathcal{A}_j\mathcal{H}^{-t_j}\mathcal{A}_j\big)\delta_{\mathbf{x}'} \geq c \cdot \mu^{\mathrm{ref}}.$$

PROOF: For any $\mathbf{x} = (\mathbf{q}, \mathbf{v})$, let us define

$$J(\mathbf{x}) = J(\mathbf{q}, \mathbf{v}) := \left|\frac{\partial \phi^{\Delta t}}{\partial \mathbf{x}}\right|.$$

Then we have

$$\begin{aligned}
&\left(\prod_{j=1}^{N} (\mathcal{H}^{-s_j}\mathcal{A}_j\mathcal{H}^{-t_j}\mathcal{A}_j)\delta_{\mathbf{x}'}\right)(d\mathbf{x})\\
&\quad = \left(\int \delta(\mathbf{q} - \{\mathcal{Q}\mathbf{x}^{(2N)}\})\delta(\mathbf{v} - \mathcal{P}\mathbf{x}^{(2N)})g_\beta(\mathbf{u})d\mathbf{u})\right) \cdot d\mathbf{x}\\
&\quad = \left(\sum_{\mathbf{k}\in\mathbb{Z}^{dN}} \int \delta(\mathbf{q} + \mathbf{k} - \mathcal{Q}\mathbf{x}^{(2N)})\delta(\mathbf{v} - \mathcal{P}\mathbf{x}^{(2N)})g_\beta(\mathbf{u})d\mathbf{u}\right) \cdot d\mathbf{x}\\
&\quad = \left(\sum_{\mathbf{k}\in\mathbb{Z}^{dN}} \big(g_\beta \circ \phi^{\Delta t}\big)(\mathbf{q} + \mathbf{k}, \mathbf{v}) \cdot J(\mathbf{q} + \mathbf{k}, \mathbf{v})\right) \cdot d\mathbf{x}.
\end{aligned}$$

Let $\mathbf{q}'' \in \mathbb{D}$ be such that $\mathcal{Q}_i\mathbf{q}'' = \{\mathcal{Q}_i\mathbf{x}^{(0)} + \sum_{j=1}^{i-1}(t_j + s_j)\mathcal{P}_i\mathbf{x}^{(0)}\}$ for any $1 \leq i \leq N$, and let $\mathbf{k}_0 = [\mathcal{Q}_i\mathbf{x}^{(0)} + \sum_{j=1}^{i-1}(t_j + s_j)\mathcal{P}_i\mathbf{x}^{(0)}]$. Then by Lemma 2.8, the following equality holds for any $\mathbf{u}$ with $\mathbf{x}^{(2N)}(\mathbf{u}) = (\mathbf{q} + \mathbf{k}_0, \mathbf{v})$ and for any $1 \leq i_0 \leq N$:

$$\left|\mathcal{Q}_{i_0}(\mathbf{q} - \mathbf{q}'') - t_{i_0}\mathbf{u}_{i_0} - \left(\sum_{i=i_0+1}^{N} t_i + \sum_{i=i_0}^{N} s_i\right)\mathbf{u}_{N+i_0}\right| \lesssim \Delta t^2,$$

$$|\mathbf{v} - \mathbf{u}_{N+i_0}| \lesssim \Delta t,$$

from which, by using the fact that $\mathbf{q}, \mathbf{q}'' \in \mathbb{D}$, we obtain

$$|\mathbf{u}| \lesssim |\mathbf{v}| + \frac{1}{\Delta t}.$$

Equivalently, there exists a constant $c > 0$ such that

$$|\mathbf{u}| \leq c \cdot \left(|\mathbf{v}| + \frac{1}{\Delta t}\right).$$

With this estimate and using Corollary 2.9, we arrive at

$$\begin{aligned}\sum_{\mathbf{k}\in\mathbb{Z}^{dN}} &(g_\beta \circ \phi^{\Delta t})(\mathbf{q}+\mathbf{k},\mathbf{v})\cdot J(\mathbf{q}+\mathbf{k},\mathbf{v})\\ &\geq (g_\beta \circ \phi^{\Delta t})(\mathbf{q}+\mathbf{k}_0,\mathbf{v})\cdot J(\mathbf{q}+\mathbf{k}_0,\mathbf{v})\\ &\gtrsim \frac{1}{\Delta t^{dN}}(g_\beta \circ \phi^{\Delta t})(\mathbf{q}+\mathbf{k}_0,\mathbf{v})\\ &\gtrsim \frac{1}{\Delta t^{dN}} g_\beta\left(c\cdot\left(|\mathbf{v}|+\frac{1}{\Delta t}\right)\right).\end{aligned}$$

It is now clear that one can take $\mu^{\text{ref}} = d\mathbf{q}\otimes g_{\beta_1}(\mathbf{v})d\mathbf{v}$ with $\beta_1 = (c+\frac{1}{\Delta t})\cdot\beta$. □

A close examination of the proof above gives the following corollary, which will be used later:

COROLLARY 2.11 *Under the same assumptions as in Corollary* 2.9, *let* $0<\alpha\leq 1$ *be fixed and let* $0<\Delta t\leq \Delta t_1$. *Assume* $(t_i)_{1\leq i\leq N}$ *and* $(s_i)_{1\leq i\leq N}$ *are sequences of nonnegative numbers such that*

$$\alpha\Delta t \leq \sum_{i=1}^{N}(t_i+s_i)\leq \Delta t$$

and

$$\frac{\Delta t}{100N}\leq \min_{1\leq i\leq N} t_i \leq \max_{1\leq i\leq N} t_i \leq \frac{\Delta t}{10N}.$$

Then there exist a constant $c=c(\alpha,\Delta t)$ *and a probability measure* μ^{ref} *that depends only on* $(\alpha,\Delta t)$ *such that*

$$\left(\prod_{j=1}^{N}(\mathcal{H}^{-s_j}\mathcal{A}_j\mathcal{H}^{-t_j}\mathcal{A}_j)\right)\mathcal{H}^{\sum_{k=1}^{N}(t_k+s_k)-\Delta t}\delta_{\mathbf{x}'}\geq c\cdot\mu^{\text{ref}}.$$

PROOF: This is quite straightforward, noting that one can always write

$$\delta_{\mathbf{y}} := \mathcal{H}^{\sum_{k=1}^{N}(t_k+s_k)-\Delta t}\delta_{\mathbf{x}'},$$

where $\mathbf{y}=\mathcal{H}^{\Delta t-\sum_{k=1}^{N}(t_k+s_k)}\mathbf{x}'$. □

We are now ready to prove the main theorem of this section.

THEOREM 2.12 *The continous-time Andersen process is uniformly ergodic*: *for any* $\nu_0>0$, *there exists a constant* $\kappa=\kappa(\nu_0)>0$, *independent of* $(\nu,t,\mathbf{x}')$, *such that for any* $0<\nu\leq\nu_0$,

$$\|P^t_{\mathbf{x}'}-\pi\|_{TV}\leq c\cdot\exp(-\kappa\nu^{2N}t),$$

where $c>0$ *is a constant independent of* $(\nu,\nu_0,t,\mathbf{x}')$.

PROOF: We start from the following Duhamel formulation of $P^t_{\mathbf{x}'}$:

$$P^t_{\mathbf{x}'} = e^{-\nu t}\mathcal{H}^{-t}\delta_{\mathbf{x}'} + \nu \int_0^t e^{\nu(s-t)}\mathcal{H}^{-(t-s)}\mathcal{A}P^s_{\mathbf{x}'}\,ds.$$

Iterating the above $2N$ times and making a change of variables, we have

$$\begin{aligned} P^t_{\mathbf{x}'} &\geq \nu^{2N}e^{-\nu t}\int_{S(t)}\left(\prod_{j=1}^{N}(\mathcal{H}^{-s_j}\mathcal{A}\mathcal{H}^{-t_j}\mathcal{A})\right)\mathcal{H}^{\sum_{k=1}^{N}(t_k+s_k)-t}\delta_{\mathbf{x}'}\,d\mathbf{t}\,d\mathbf{s} \\ &\geq \left(\frac{\nu}{2N}\right)^{2N} e^{-\nu t}\int_{S(t)}\left(\prod_{j=1}^{N}(\mathcal{H}^{-s_j}\mathcal{A}_j\mathcal{H}^{-t_j}\mathcal{A}_j)\right)\mathcal{H}^{\sum_{k=1}^{N}(t_k+s_k)-t}\delta_{\mathbf{x}'}\,d\mathbf{t}\,d\mathbf{s}, \end{aligned}$$

where

$$\begin{aligned} S(t) &:= \left\{(t_1,\dots,t_N,s_1,\dots,s_N)\in\mathbb{R}^{2N}_+ : \sum_{k=1}^{N}(t_k+s_k)\leq t\right\}, \\ \mathbb{R}^{2N}_+ &:= \{(t_1,\dots,t_N,s_1,\dots,s_N) : t_k\geq 0,\ s_k\geq 0, \forall\, 1\leq k\leq N\}, \\ d\mathbf{t}\,d\mathbf{s} &:= \prod_{j=1}^{N} dt_j\,ds_j. \end{aligned}$$

Let $0 < t^* < \Delta t_1$ be fixed, where Δt_1 is the same as in Corollary 2.9. By Corollary 2.11 (choose $\alpha = t^*/100N$), there exist a positive constant c and a reference probability measure μ^{ref}, both depending only on t^*, such that

$$\begin{aligned} P^{t^*}_{\mathbf{x}'} &\geq c\cdot\nu^{2N}e^{-\nu t^*}\mu^{\mathrm{ref}}\cdot\int_{O(t^*)} d\mathbf{t}\,d\mathbf{s} \\ &\geq \tilde{c}\cdot(\nu t^*)^{2N}e^{-\nu t^*}\mu^{\mathrm{ref}}, \end{aligned}$$

where in the above $\tilde{c}$ is a positive constant depending only on t^* and

$$O(t^*) = \left\{(t_1,\dots,t_N,s_1,\dots,s_N)\in S(t^*) : \frac{t^*}{100N}\leq \min_{1\leq k\leq N} t_k \leq \max_{1\leq k\leq N} t_k \leq \frac{t^*}{10N}\right\}.$$

The rest of this proof follows by repeating the same argument used at the end of the proof of Theorem 2.3. We omit the details here. □

2.4 Discrete-Time Andersen Process, $N > 1$

We shall prove that the discrete-time Andersen process is uniformly ergodic. Again with a slight abuse of notation let us denote

$$P^{\Delta t}_{\mathbf{x}'} := \left((1-\nu\Delta t)\mathcal{H}^{-\Delta t} + \nu\Delta t\mathcal{A}\mathcal{H}^{-\Delta t}\right)\delta_{\mathbf{x}'},$$

where $\mathcal{A} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{A}_i$ for $N > 1$. Also denote $\Delta t_2 = \Delta t_1/7$, where Δt_1 is the same as in Corollary 2.9.

THEOREM 2.13 *Let $\nu_0 > 0$ be arbitrary but fixed. Then for any $0 < \nu < \nu_0$, $0 < \Delta t < \min\{1/2\nu_0, \Delta t_2/20N\}$, and for any $\mathbf{x}' \in \Gamma$, there exists a constant $\kappa = \kappa(\nu_0)$, independent of $(\nu, \Delta t, \mathbf{x}')$, such that for any $n \geq 1$,*

$$\|(P_{\mathbf{x}'}^{\Delta t})^n - \pi\|_{TV} \leq c \cdot \exp(-\kappa \nu^{2N} n \Delta t),$$

where c is a positive constant independent of $(\nu, \nu_0, n, \mathbf{x}', \Delta t)$.

PROOF: Let $m = [\Delta t_2/N\Delta t]$. Clearly $m \geq 20$ by assumption. Let us also introduce the following notation:

$$\begin{aligned}
&\mathbf{m} := (m_1, \dots, m_N),\\
&\mathbf{n} := (n_1, \dots, n_N),\\
&\alpha_1 = \alpha_1(\nu, \Delta t, m, N) := (\nu\Delta t)^{2N}(1-\nu\Delta t)^{6Nm-2N},\\
&\alpha_2 := \alpha_1/N^{2N},\\
&S_1 := \Bigg\{(k, \mathbf{m}, \mathbf{n}) : k + \sum_{j=1}^{N}(m_j + n_j) = 6Nm - 2N,\\
&\qquad\qquad k \geq 0,\ m_j \geq 0,\ n_j \geq 0,\ \forall\, 1 \leq j \leq N\Bigg\}\\
&S_2 := \Bigg\{(k, \mathbf{m}, \mathbf{n}) : k + \sum_{j=1}^{N}(m_j + n_j) = 6Nm,\ k \geq 1,\ m_j \geq 1,\ n_j \geq 1,\\
&\qquad\qquad \forall\, 1 \leq j \leq N-1,\ m_N \geq 0,\ n_N \geq 1\Bigg\}.
\end{aligned}$$

Then we have

$$\begin{aligned}
(P_{\mathbf{x}'}^{\Delta t})^{6Nm} &= \big((1-\nu\Delta t)\mathcal{H}^{-\Delta t} + \nu\Delta t \mathcal{A}\mathcal{H}^{-\Delta t}\big)^{6Nm}\,\delta_{\mathbf{x}'}\\
&\geq \alpha_1 \cdot \sum_{(k,\mathbf{m},\mathbf{n})\in S_1}\left(\prod_{j=1}^{N}\mathcal{H}^{-m_j\Delta t}\mathcal{A}\mathcal{H}^{-\Delta t}\mathcal{H}^{-n_j\Delta t}\mathcal{A}\mathcal{H}^{-\Delta t}\right)\mathcal{H}^{-k\Delta t}\delta_{\mathbf{x}'}\\
&\geq \alpha_1 \cdot \sum_{(k,\mathbf{m},\mathbf{n})\in S_2}\left(\prod_{j=1}^{N}\mathcal{H}^{-m_j\Delta t}\mathcal{A}\mathcal{H}^{-n_j\Delta t}\mathcal{A}\right)\mathcal{H}^{-k\Delta t}\delta_{\mathbf{x}'}\\
&\geq \alpha_2 \cdot \sum_{(k,\mathbf{m},\mathbf{n})\in S_2}\left(\prod_{j=1}^{N}\mathcal{H}^{-m_j\Delta t}\mathcal{A}_j\mathcal{H}^{-n_j\Delta t}\mathcal{A}_j\right)\mathcal{H}^{-k\Delta t}\delta_{\mathbf{x}'}
\end{aligned}$$

Now note that for $(k, \mathbf{m}, \mathbf{n}) \in S_2$, we have

$$k\Delta t + \sum_{j=1}^{N}(m_j + n_j)\Delta t = 6Nm\Delta t \in (3\Delta t_2, 6\Delta t_2) = \left(\frac{3}{7}\Delta t_1, \frac{6}{7}\Delta t_1\right).$$

If, for any $1 \le j \le N$, we require $\frac{7}{20}m \le n_j \le \frac{7}{10}m$, then after a bit of algebra, one sees that for all such n_j,

$$\frac{\Delta t_1}{100N} \le n_j \Delta t \le \frac{\Delta t_1}{10N}.$$

By Corollary 2.11, we conclude that there exists a constant $c > 0$ and a probability measure μ^{ref}, both depending only on Δt_1 and such that

$$\left(\prod_{j=1}^{N} \mathcal{H}^{-m_j\Delta t}\mathcal{A}_j\mathcal{H}^{-n_j\Delta t}\mathcal{A}_j\right)\mathcal{H}^{-k\Delta t}\delta_{\mathbf{x}'} \ge c\mu^{\text{ref}}$$

where $(k, \mathbf{m}, \mathbf{n}) \in S_3$ and S_3 is given by

$$S_3 := \left\{(k, \mathbf{m}, \mathbf{n}) \in S_2 : \frac{7}{20}m \le n_j \le \frac{7}{10}m,\ \forall\, 1 \le j \le N\right\}.$$

By simple combinatorics, we have $|S_3| \ge (\frac{7}{20}m)^{2N}$. Therefore we have

$$\begin{aligned}
(P_{\mathbf{x}'}^{\Delta t})^{6Nm} &\ge c \cdot \left(\frac{7}{20}m\right)^{2N}\left(\frac{\nu\Delta t}{N}\right)^{2N}(1-\nu\Delta t)^{6Nm-2N}\mu^{\text{ref}} \\
&\gtrsim \nu^{2N} \cdot (\Delta t_2)^{2N}(1-\nu\Delta t)^{\frac{6\Delta t_2}{\Delta t}}\mu^{\text{ref}} \\
&\gtrsim (\nu\Delta t_2)^{2N}e^{-12\nu\Delta t_2}\mu^{\text{ref}}.
\end{aligned}$$

Now the proof is finished by an argument similar to that used in the last part of the proof of Theorem 2.5. We omit the details. □

3 Forward Euler Approximation of the Andersen Process

3.1 Ergodicity of Forward Euler Andersen Process

The forward Euler scheme that approximates the equation of motion (1.1) using a time step Δt is given by the following:

$$\begin{cases} \mathbf{q}^{\Delta t} = \mathbf{q} + \mathbf{v}\Delta t, \\ \mathbf{v}^{\Delta t} = \mathbf{v} + \mathbf{F}(\mathbf{q})\Delta t, \end{cases}$$

where we have denoted $\mathbf{F}(\mathbf{q}) := -\nabla\Phi(\mathbf{q})$. Let us denote by $\mathcal{T}^{\Delta t}$ the map

$$\mathbf{x} = (\mathbf{q}, \mathbf{v}) \longrightarrow \mathbf{x}^{\Delta t} = (\mathbf{q}^{\Delta t}, \mathbf{v}^{\Delta t}).$$

Note that $\mathcal{T}^{\Delta t}$ is not a flow operator in the sense that

$$\mathcal{T}^{\Delta t_1+\Delta t_2} \ne \mathcal{T}^{\Delta t_1} \circ \mathcal{T}^{\Delta t_2}.$$

As we shall see shortly, this complicates some of the proofs below.

DEFINITION 3.1 (Forward Euler Andersen Process) Let $\{\alpha_n\}_{n=1}^{\infty}$ be i.i.d. random variables such that $\mathbb{P}(\alpha_n = 1) = \lambda = \nu\Delta t$ and $\mathbb{P}(\alpha_n = 0) = 1 - \lambda = 1 - \nu\Delta t$. Retain the same notion of $\mathcal{S}$, Y_n, and Z_n as in Definition 1.1. The *discrete-time forward Euler Andersen process* is defined as

$$\mathbf{x}_{n+1} = (1-\alpha_n)\mathcal{T}^{\Delta t}\mathbf{x}_n + \alpha_n \mathcal{S}(Y_n, Z_n)\mathcal{T}^{\Delta t}\mathbf{x}_n.$$

Let $\mathcal{T}^{-\Delta t}$ be the Markov operator associated with $\mathcal{T}^{\Delta t}$. The evolution equation for the probability measures $\mu_n^{\Delta t}$ is given by

$$\mu_{n+1}^{\Delta t} = (1-\lambda)\mathcal{T}^{-\Delta t}\mu_n^{\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t}\mu_n^{\Delta t}.$$

Note that if there is a stationary probability measure $\mu^{\Delta t}$ for the forward Euler Andersen process, then it has to satisfy

$$\mu^{\Delta t} = (1-\lambda)\mathcal{T}^{-\Delta t}\mu^{\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t}\mu^{\Delta t}. \tag{3.1}$$

LEMMA 3.2 *Assume that* $0 < \Delta t_0 \le \varepsilon_0 \|\nabla \mathbf{F}\|$. ε_0 *is sufficiently small such that*

$$\varepsilon_0 \le \min\left\{\frac{1}{\sqrt{8N\|\nabla\mathbf{F}\|^3}}, \frac{1}{\|\nabla\mathbf{F}\|}\right\}.$$

Let $c = 100\|\nabla\mathbf{F}\|\sqrt{d}$. *Then for any* $0 < \Delta t \le \Delta t_0$, $n\Delta t \le \Delta t_0$, *and for any* $1 \le i, j \le N$, *the following inequalities hold*:

$$\left\|\frac{\partial(\mathcal{Q}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})} - \delta_{ij}I\right\| \le c\cdot(n\Delta t)^2,$$

$$\left\|\frac{\partial(\mathcal{P}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} - \delta_{ij}I\right\| \le c\cdot n\Delta t,$$

$$\left\|\frac{\partial(\mathcal{Q}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} - n\Delta t\delta_{ij}I\right\| \le c\cdot(n\Delta t)^2,$$

$$\left\|\frac{\partial(\mathcal{P}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})}\right\| \le c\cdot n\Delta t.$$

PROOF: We shall prove this lemma inductively. Note first, by the definition of $\mathcal{T}^{\Delta t}$, we have

$$\frac{\partial(\mathcal{Q}_j\mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})} = \delta_{ij}I, \qquad \frac{\partial(\mathcal{Q}_j\mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} = \Delta t\delta_{ij}I,$$
$$\frac{\partial(\mathcal{P}_j\mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} = \delta_{ij}I, \qquad \frac{\partial(\mathcal{P}_j\mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})} = \Delta t\frac{\partial\mathbf{F}_j}{\partial(\mathcal{Q}_i\mathbf{x})}.$$

Denote $\mathbf{x}_n := (\mathcal{T}^{\Delta t})^n\mathbf{x}$. Clearly for $l = 0$, the inequalities hold. Suppose that the inequalities hold for $l = n$. Now consider $l = n+1$ and assume $(n+1)\Delta t \le \Delta t_0$. Obviously

$$\frac{\partial(\mathcal{Q}_j\mathbf{x}_{n+1})}{\partial(\mathcal{Q}_i\mathbf{x})} = \frac{\partial(\mathcal{Q}_j\mathbf{x}_n)}{\partial(\mathcal{Q}_i\mathbf{x})} + \Delta t\frac{\partial(\mathcal{P}_j\mathbf{x}_n)}{\partial(\mathcal{Q}_i\mathbf{x})}.$$

By the induction hypothesis

$$\left\| \frac{\partial(\mathcal{Q}_j \mathbf{x}_{n+1})}{\partial(\mathcal{Q}_i \mathbf{x})} - \delta_{ij} I \right\| \le c \cdot (n\Delta t)^2 + c \cdot (n\Delta t)\Delta t \le c \cdot (n+1)^2 \Delta t^2.$$

Similarly, we have

$$\frac{\partial(\mathcal{Q}_j \mathbf{x}_{n+1})}{\partial(\mathcal{P}_i \mathbf{x})} = \frac{\partial(\mathcal{Q}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} + \Delta t \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})}.$$

Hence

$$\left\| \frac{\partial(\mathcal{Q}_j \mathbf{x}_{n+1})}{\partial(\mathcal{P}_i \mathbf{x})} - (n+1)\Delta t \delta_{ij} I \right\| \le c \cdot (n\Delta t)^2 + c \cdot n\Delta t^2 \le c \cdot (n+1)^2 \Delta t^2.$$

The remaining inequalities are proved in a similar fashion:

$$\begin{aligned}\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_{n+1})}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| &= \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} + \sum_{k=1}^{N} \Delta t \frac{\partial \mathbf{F}_j}{\partial(\mathcal{Q}_k \mathbf{x}_n)} \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| \\ &\le \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| + \Delta t \cdot \|\nabla \mathbf{F}\| \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| \\ &\le c \cdot n\Delta t + \Delta t \cdot \|\nabla \mathbf{F}\| (\sqrt{d} + c \cdot N(n\Delta t)^2) \\ &\le c \cdot (n+1)\Delta t.\end{aligned}$$

$$\begin{aligned}\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_{n+1})}{\partial(\mathcal{P}_i \mathbf{x})} - \delta_{ji} I \right\| &= \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} - \delta_{ji} I + \sum_{k=1}^{N} \Delta t \frac{\partial \mathbf{F}_j}{\partial(\mathcal{Q}_k \mathbf{x}_n)} \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} \right\| \\ &\le \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} - \delta_{ji} I \right\| + \Delta t \cdot \|\nabla \mathbf{F}\| \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} \right\| \\ &\le c \cdot n\Delta t + \Delta t \cdot \|\nabla \mathbf{F}\| (n\Delta t\sqrt{d} + c \cdot N(n\Delta t)^2) \\ &\le c \cdot (n+1)\Delta t.\end{aligned}$$

□

LEMMA 3.3 *Let* $0 < \Delta t < \Delta t_1 \le \Delta t_0$ *where* Δt_0 *is the same as in Lemma 3.2. Let* $(m_j)_{1\le j\le N}$ *and* $(n_j)_{1\le j\le N}$ *be sequences of nonnegative integers such that* $\sum_{j=1}^{N} (n_j + m_j)\Delta t \le \Delta t_1$. *Let* $\mathbf{x}^{(0)} = (\mathbf{q}_1^{(0)}, \ldots, \mathbf{q}_N^{(0)}, \mathbf{v}_1^{(0)}, \ldots, \mathbf{v}_N^{(0)})$. *For* $i = 1, \ldots, N$, *recursively define*

$$\begin{aligned}\mathbf{x}^{(2i-1)} &:= (\mathcal{T}^{\Delta t})^{n_i} \mathcal{S}(i, \mathbf{u}_i)\mathbf{x}^{(2i-2)}, \\ \mathbf{x}^{(2i)} &:= (\mathcal{T}^{\Delta t})^{m_i} \mathcal{S}(i, \mathbf{u}_{N+i})\mathbf{x}^{(2i-1)}.\end{aligned}$$

Fix $1 \le i_0 \le N$; *then for any* $m \ge 2i_0$, $1 \le j \le N$, *there exists a constant* $c > 0$ *such that*

$$\left\| \frac{\partial(\mathcal{Q}_j \mathbf{x}^{(m)})}{\partial \mathbf{u}_{i_0}} - n_{i_0} \Delta t \delta_{j i_0} I \right\| \le c \Delta t_1^2,$$

$$\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}^{(m)})}{\partial \mathbf{u}_{i_0}} \right\| \le c \Delta t_1,$$

$$\left\| \frac{\partial(\mathcal{Q}_j \mathbf{x}^{(m)})}{\partial \mathbf{u}_{N+i_0}} \right\| \le c \Delta t_1,$$

$$\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}^{(m)})}{\partial \mathbf{u}_{N+i_0}} - \delta_{j i_0} I \right\| \le c \Delta t_1.$$

PROOF: This follows essentially the same lines as the proof of Lemma 2.7. We omit the details. □

As is similar to what we did before, we have the following:

COROLLARY 3.4 *We make the same assumptions as in Lemma* 3.3 *and*

$$\min_{1 \le i \le N} (n_i \Delta t) \ge \frac{\Delta t_1}{100N}.$$

Then there exists a constant $\Delta t_2 > 0$ *such that for any* $0 < \Delta t_1 \le \Delta t_2$, *the map*

$$\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_{2N}) \in \mathbb{R}^{2dN} \to \mathbf{x}^{(2N)}$$

is a C^∞ *homeomorphism from* $\mathbb{R}^{2dN}$ *to* $\mathbb{R}^{2dN}$ *and there exist positive constants* c_1 *and* c_2 *such that the following holds*:

$$c_1 \Delta t_1^{dN} \le \left| \det\left(\frac{\partial \mathbf{x}^{(2N)}}{\partial \mathbf{u}} \right) \right| \le c_2 \Delta t_1^{dN}$$

COROLLARY 3.5 *Use the same assumptions as in Corollary* 3.4. *There exist a positive constant* $c < 1$ *and a probability measure* μ^{ref}, *both depending only on* Δt_1, *such that for any* $\mathbf{x}' \in \Gamma$, *we have*

$$\left(\prod_{j=1}^{N} (\mathcal{T}^{-\Delta t})^{m_j} \mathcal{A}_j (\mathcal{T}^{-\Delta t})^{n_j} \mathcal{A}_j \right) \delta_{\mathbf{x}'} \ge c \cdot \mu^{\text{ref}}.$$

Let $\Delta t_3 = \Delta t_1 / 7$; then we have the following:

THEOREM 3.6 (Uniform Ergodicity of Forward Euler Andersen Process) *Let* $\nu_0 > 0$ *be arbitrary but fixed. Then for* $0 < \Delta t < \min\{1/2\nu_0, \Delta t_3/20N\}$ *and any* $0 < \nu < \nu_0$, *there exists a unique invariant probability measure that we denote as* $\mu^{\Delta t}$. *Furthermore, for any* $\mathbf{x}' \in \Gamma$, *there exists a constant* $\kappa = \kappa(\nu_0)$, *independent of* $(\Delta t, \mathbf{x}')$, *such that for any* $n \ge 1$,

$$\|(Q_{\mathbf{x}'}^{\Delta t})^n - \mu^{\Delta t}\|_{TV} \le c \cdot \exp(-\kappa \nu^{2N} n \Delta t),$$

where c is a positive constant independent of $(\mathbf{x}', n, \nu_0, \nu, \Delta t)$.

PROOF: This follows by essentially repeating the arguments used in the proof of Theorem 2.13. Therefore we omit the details. □

3.2 Regularity of the Invariant Measure

In this section we assume Δt is sufficiently small such that the forward Euler Andersen process is ergodic. Let $\mathcal{M}$ denote the set of Borel probability measures on Γ.

LEMMA 3.7 *Let $\mathcal{A}_{\text{pr}}^N := \mathcal{A}_N \cdots \mathcal{A}_1$. For any $\mu \in \mathcal{M}$, we have that $\mathcal{A}_{\text{pr}}^N \mathcal{T}^{-\Delta t} \mathcal{A}_{\text{pr}}^N \mu$ is again a probability measure that admits an infinitely differentiable density.*

PROOF: Assume first $\mu = \delta_{\mathbf{x}'}$, where $\mathbf{x}' = (\mathbf{q}', \mathbf{v}') \in \Gamma$ is arbitrary but fixed. For any $\mathbf{x} = (\mathbf{q}, \mathbf{v}) \in \Gamma$, we have

$$
\begin{aligned}
&(\mathcal{A}_{\text{pr}}^N \mathcal{T}^{-\Delta t} \mathcal{A}_{\text{pr}}^N \delta_{\mathbf{x}'})(d\mathbf{x}) \\
&= \left(\int_{\mathbb{R}^{dN}} \delta(\mathbf{q} - \{\mathcal{Q}\mathcal{T}^{\Delta t}(\mathbf{q}', \mathbf{u})\}) g_\beta(\mathbf{u}) d\mathbf{u} \right) g_\beta(\mathbf{v}) \cdot d\mathbf{x} \\
&= \sum_{\mathbf{k} \in \mathbb{Z}^{dN}} \left(\int_{\mathbb{R}^{dN}} \delta(\mathbf{q} + \mathbf{k} - \mathbf{q}' - \mathbf{u}\Delta t) g_\beta(\mathbf{u}) d\mathbf{u} \right) g_\beta(\mathbf{v}) \cdot d\mathbf{x} \\
&= \frac{1}{\Delta t^{dN}} g_\beta(\mathbf{v}) \cdot \sum_{\mathbf{k} \in \mathbb{Z}^{dN}} g_\beta\left(\frac{\mathbf{q} - \mathbf{q}' + \mathbf{k}}{\Delta t} \right) \cdot d\mathbf{x}.
\end{aligned}
$$

Let $\mu \in \Gamma$ be arbitrary. Denote by $\hat{\mu}$ the marginal probability measure on $\mathbb{D}$ induced by μ . The following is immediate:

$$
(\mathcal{A}_{\text{pr}}^N \mathcal{T}^{-\Delta t} \mathcal{A}_{\text{pr}}^N \mu)(d\mathbf{x}) = \frac{1}{\Delta t^{dN}} g_\beta(\mathbf{v}) \cdot h(\mathbf{q}) \cdot d\mathbf{x}
$$

where

$$
h(\mathbf{q}) = \sum_{\mathbf{k} \in \mathbb{Z}^{dN}} \int_{\mathbb{D}} g_\beta \left(\frac{\mathbf{q} - \mathbf{q}' + \mathbf{k}}{\Delta t} \right) \hat{\mu}(d\mathbf{q}').
$$

It is a simple exercise to show that the function $h(\cdot)$ is infinitely differentiable on $\mathbb{D}$. We omit the details. □

Let us recall that $\mu^{\Delta t}$ denotes the unique invariant probability measure of the forward Euler Andersen process. If $N = 1$, the following theorem can be easily proved:

THEOREM 3.8 *Assume $N = 1$. Then $\mu^{\Delta t}$ is absolutely continuous with respect to the Lebesgue measure on Γ, i.e., the Radon-Nikodym derivative $d\mu^{\Delta t}/d\mathbf{x} \in L^1(\Gamma, d\mathbf{x})$.*

PROOF: Since $\mu^{\Delta t}$ is the invariant measure, by (3.1), we have

$$\mu^{\Delta t} = (1-\lambda)\mathcal{T}^{-\Delta t}\mu^{\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t}\mu^{\Delta t}$$

where $\lambda = \nu\Delta t$.
Since $N = 1$, we have $\mathcal{A}^2 = \mathcal{A}$. Hence

$$\mathcal{A}\mu^{\Delta t} = \mathcal{A}\mathcal{T}^{-\Delta t}\mu^{\Delta t}.$$

Let $\tilde{\mathcal{T}} = \frac{1}{1-\lambda}\mathcal{T}^{-\Delta t} + \frac{\lambda}{1-\lambda}\mathcal{A}(\mathcal{T}^{-\Delta t})^2$. Then we have

$$\begin{aligned}\mu^{\Delta t} &= (1-\lambda)\mathcal{T}^{-\Delta t}\mu^{\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t}\left((1-\lambda)\mathcal{T}^{-\Delta t}\mu^{\Delta t} + \lambda\mathcal{A}\mu^{\Delta t}\right)\\ &= (1-\lambda)\mathcal{T}^{-\Delta t}\mu^{\Delta t} + \lambda(1-\lambda)\mathcal{A}(\mathcal{T}^{-\Delta t})^2\mu^{\Delta t} + \lambda^2\mathcal{A}\mathcal{T}^{-\Delta t}\mathcal{A}\mu^{\Delta t}\\ &= (1-\lambda^2)\tilde{\mathcal{T}}\mu^{\Delta t} + \lambda^2\mathcal{A}\mathcal{T}^{-\Delta t}\mathcal{A}\mu^{\Delta t}.\end{aligned}$$

Observe that by definition $\tilde{\mathcal{T}}$ is a Markov operator. By Lemma 3.7, we conclude that $\mathcal{A}\mathcal{T}^{-\Delta t}\mathcal{A}\mu^{\Delta t}$ has a C^∞ density with respect to the Lebesgue measure on Γ. Denote this density as ρ_1; then we have that the following Neumann series converges in $L^1(\Gamma, d\mathbf{x})$:

$$\frac{d\mu^{\Delta t}}{d\mathbf{x}} = \sum_{k=0}^{\infty}\lambda^2(1-\lambda^2)^k\tilde{\mathcal{T}}^k\rho_1,$$

which proves that $d\mu^{\Delta t}/d\mathbf{x} \in L^1(\Gamma, d\mathbf{x})$. □

We need to work a little bit harder to get the same conclusion in the $N > 1$ case. For all $\mathbf{u} \in \mathbb{R}^{2dN}$, let us define

$$\mathbf{x}^{(2N)}(\mathbf{u}; \mathbf{x}') := \left(\prod_{i=1}^{N}\mathcal{T}^{\Delta t}\mathcal{S}(i, \mathbf{u}_{N+i})\mathcal{T}^{\Delta t}\mathcal{S}(i, \mathbf{u}_i)\right)\mathbf{x}'$$

where $\mathbf{x}' = (\mathbf{q}', \mathbf{v}') \in \Gamma$ is the initial point. We shall view $\mathbf{x}^{(2N)}(\cdot\,; \mathbf{x}')$ as a family of maps parametrized by $\mathbf{x}'$. It has been shown before that $\forall\mathbf{x}' \in \Gamma$ the map $\mathbf{u} \to \mathbf{x}^{(2N)}(\mathbf{u}; \mathbf{x}')$ is globally smoothly invertible. Let us denote the inverse maps as $\mathbf{u} = \phi(\mathbf{x}; \mathbf{x}')$; then we have the following important lemma, which is intuitively quite obvious:

LEMMA 3.9 $\forall\mathbf{x}' = (\mathbf{q}', \mathbf{v}') \in \Gamma$, *the probability measure*

$$\left(\prod_{i=1}^{N}\mathcal{T}^{-\Delta t}\mathcal{A}_i\mathcal{T}^{-\Delta t}\mathcal{A}_i\right)\delta_{\mathbf{x}'}$$

is absolutely continuous with respect to the Lebesgue measure on Γ. *Furthermore, let* $\rho(\mathbf{x}; \mathbf{x}') = \rho(\mathbf{q}, \mathbf{v}; \mathbf{q}', \mathbf{v}')$; *then* ρ *is periodic in* $\mathbf{q}$, $\mathbf{q}'$, *and* $\mathbf{v}'$ *with periods* $\mathbb{Z}^{dN}$, $\mathbb{Z}^{dN}$, *and* $\Delta t^{-1}\mathbb{Z}^{dN}$, *respectively, and we have* $\rho \in C^\infty(\mathbb{D} \otimes \mathbb{R}^{dN} \otimes \mathbb{D} \otimes \mathbb{R}^{dN})$.

PROOF: A short computation gives the following representation of ρ:

$$\rho(\mathbf{x}; \mathbf{x}') = \left|\det\left(\frac{\partial\phi(\mathbf{x}; \mathbf{x}')}{\partial\mathbf{x}}\right)\right| \cdot \sum_{\mathbf{k}\in\mathbb{Z}^{dN}} g_\beta(\mathbf{k} + \mathcal{Q}\phi(\mathbf{x}; \mathbf{x}')),$$

where we have used the trivial fact that $|\det(\partial\phi(\mathbf{x};\mathbf{x}')/\partial\mathbf{x})|$ is periodic in $\mathbf{q}$ with period $\mathbb{Z}^{dN}$. Observe that by Corollary 3.4,

$$\left|\det\left(\frac{\partial\phi(\mathbf{x};\mathbf{x}')}{\partial\mathbf{x}}\right)\right| = O(\Delta t^{-dN}),$$

where the implied constant does not depend on $(\mathbf{x},\mathbf{x}')$. Given this fact, the rest of the statements easily follow. □

The next corollary is immediate:

COROLLARY 3.10 *$\forall\mu\in M$, the probability measure $(\prod_{i=1}^{N}\mathcal{T}^{-\Delta t}\mathcal{A}_i\mathcal{T}^{-\Delta t}\mathcal{A}_i)\mu$ is absolutely continuous with respect to the Lebesgue measure on Γ. Furthermore, its density is infinitely differentiable. In other words, the Markov operator*

$$\mathcal{A}_M = \prod_{i=1}^{N}\mathcal{T}^{-\Delta t}\mathcal{A}_i\mathcal{T}^{-\Delta t}\mathcal{A}_i$$

is a smoothing operator.

Remark 3.11. The introduction of the smoothing operator $\mathcal{A}_M$ stems from the intuitive observation that for small Δt, $\mathcal{A}_M \approx \mathcal{A}_{\text{pr}}^{(N)}$.

PROOF OF COROLLARY 3.10: Observe that the following holds true $\forall\mu\in M$ and $A\in\mathbb{B}(\Gamma)$:

$$\left(\prod_{i=1}^{N}\mathcal{T}^{-\Delta t}\mathcal{A}_i\mathcal{T}^{-\Delta t}\mathcal{A}_i\mu\right)(A) = \int_A d\mathbf{x}\int_\Gamma \rho(\mathbf{x};\mathbf{x}')\mu(d\mathbf{x}').$$

Since ρ is C^∞ and periodic in $\mathbf{x}'$ with period $\mathbb{Z}^{dN}\otimes\Delta t^{-1}\mathbb{Z}^{dN}$, the corollary easily follows. □

Now we are ready to prove the regularity of the invariant measure in the many-particle case.

THEOREM 3.12 *The invariant probability measure $\mu^{\Delta t}$ is absolutely continuous with respect to the Lebesgue measure on Γ, i.e., the Radon-Nikodym derivative $\frac{d\mu}{d\mathbf{x}}$ is in $L^1(\Gamma, d\mathbf{x})$.*

PROOF: Since $\mu^{\Delta t}$ is stationary, denoting $\tilde\lambda = (1-\lambda)(\frac{\lambda}{N})^{2N}$, we have

$$\begin{aligned}
\mu^{\Delta t} &= ((1-\lambda)\mathcal{T}^{-\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t})\mu^{\Delta t}\\
&= ((1-\lambda)\mathcal{T}^{-\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t})^{2N+1}\mu^{\Delta t}\\
&= \left((1-\lambda)\mathcal{T}^{-\Delta t} + \frac{\lambda}{N}\sum_{j=1}^{N}\mathcal{A}_j\mathcal{T}^{-\Delta t}\right)^{2N+1}\mu^{\Delta t}
\end{aligned}$$

$$= (1-\tilde{\lambda})\tilde{\mathcal{T}}\mu^{\Delta t} + \tilde{\lambda}\mathcal{T}^{-\Delta t}\prod_{j=1}^{N}(A_j\mathcal{T}^{-\Delta t}\mathcal{A}_j\mathcal{T}^{-\Delta t})\mu^{\Delta t}$$

$$= (1-\tilde{\lambda})\tilde{\mathcal{T}}\mu^{\Delta t} + \tilde{\lambda}\left(\prod_{j=1}^{N}(\mathcal{T}^{-\Delta t}\mathcal{A}_j\mathcal{T}^{-\Delta t}\mathcal{A}_j)\right)\mathcal{T}^{-\Delta t}\mu^{\Delta t}$$

where in the last two steps we have used $\tilde{\mathcal{T}}$ to denote the Markov operator whose definition is given by

$$\tilde{\mathcal{T}} := \frac{((1-\lambda)\mathcal{T}^{-\Delta t} + \frac{\lambda}{N}\sum_{j=1}^{N}\mathcal{A}_j\mathcal{T}^{-\Delta t})^{2N+1} - \tilde{\lambda}(\prod_{j=1}^{N}(\mathcal{T}^{-\Delta t}\mathcal{A}_j\mathcal{T}^{-\Delta t}\mathcal{A}_j))\mathcal{T}^{-\Delta t}}{1-\tilde{\lambda}}.$$

By Corollary 3.10, we denote by ρ_1 the Radon-Nikodym derivative of

$$\left(\prod_{j=1}^{N}(\mathcal{T}^{-\Delta t}\mathcal{A}_j\mathcal{T}^{-\Delta t}\mathcal{A}_j)\right)\mathcal{T}^{-\Delta t}\mu^{\Delta t}$$

with respect to the Lebesgue measure on Γ. Then the following Neumann series obviously converges in $L^1(\Gamma, d\mathbf{x})$:

$$\frac{d\mu^{\Delta t}}{d\mathbf{x}} = \tilde{\lambda}\sum_{k=0}^{\infty}(1-\tilde{\lambda})^k\tilde{\mathcal{T}}^k\rho_1,$$

which proves the theorem. □

3.3 Error Analysis for Forward Euler Approximation

Throughout this section we assume $0 < \nu \leq \nu_0$ and $0 < \Delta t \leq \Delta t^*$ where $(\nu_0, \Delta t^*)$ are fixed and Δt^* is sufficiently small such that the forward Euler Andersen process is uniformly ergodic. To analyze how well the invariant measure of the forward Euler Andersen process approximates the Gibbsian measure, let us denote

$$P^{\Delta t} := (1-\nu\Delta t)\mathcal{H}^{-\Delta t} + \nu\Delta t\mathcal{A}\mathcal{H}^{-\Delta t},$$
$$Q^{\Delta t} := (1-\nu\Delta t)\mathcal{T}^{-\Delta t} + \nu\Delta t\mathcal{A}\mathcal{T}^{-\Delta t}.$$

Since π and $\mu^{\Delta t}$ are stationary for $P^{\Delta t}$ and $Q^{\Delta t}$, respectively, we have

$$\mu^{\Delta t} - \pi = Q^{\Delta t}(\mu^{\Delta t} - \pi) + (Q^{\Delta t} - P^{\Delta t})\pi$$
$$= Q^{\Delta t}(\mu^{\Delta t} - \pi) + (Q^0 + Q^{\Delta t} + \cdots + (Q^{\Delta t})^{(n-1)})(Q^{\Delta t} - P^{\Delta t})\pi.$$

Since the Markov operators $Q^{\Delta t}$ and $\mathcal{A}$ are nonexpansive, we have by the uniform ergodicity of $Q^{\Delta t}$

$$\|\mu^{\Delta t} - \pi\|_{TV} \leq \|(Q^{\Delta t})^n(\mu^{\Delta t} - \pi)\|_{TV} + n\|(Q^{\Delta t} - P^{\Delta t})\pi\|_{TV}$$
$$\leq c\cdot\exp(-\nu^{2N}\kappa n\Delta t) + n\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV}$$

where c and $\kappa = \kappa(\nu_0)$ are positive constants.

To help analyze the error term $\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV}$, we first prove a number of simple lemmas.

LEMMA 3.13 *Denote the Gibbsian density on* Γ *(at temperature* β^{-1}*) as* $\rho(\cdot)$*. Then there exists a constant* $c = c(\beta)$ *such that* $\forall \mathbf{x} \in \Gamma$*, we have*

$$|\nabla_{\mathbf{x}}\rho(\mathbf{x})| \le c \cdot g_\beta(\mathcal{P}\mathbf{x})(1 + |\mathcal{P}\mathbf{x}|).$$

PROOF: We only need to observe that $\nabla_{\mathbf{x}}\rho(\mathbf{x}) = -\beta e^{-H(\mathbf{x})}\nabla_{\mathbf{x}}H(\mathbf{x})$, where $H(\mathbf{x}) = \Phi(\mathcal{Q}\mathbf{x}) + |\mathcal{P}\mathbf{x}|^2/2$. □

LEMMA 3.14 *There exists a constant* $c > 0$*, depending only on the potential* Φ*, such that* $\forall\ 0 < \Delta t < \Delta t^*$ *and* $\mathbf{x} \in \Gamma$*, we have*

$$|\mathcal{H}^{\Delta t}\mathbf{x} - \mathcal{T}^{\Delta t}\mathbf{x}| \le c \cdot (1 + |\mathcal{P}\mathbf{x}|)\Delta t^2.$$

PROOF: From the definition of $\mathcal{H}^{\Delta t}$, we have

$$\mathcal{Q}\mathcal{H}^{\Delta t}\mathbf{x} = \mathcal{Q}\mathbf{x} + \Delta t\mathcal{P}\mathbf{x} + \int_0^{\Delta t}\int_0^s \mathbf{F}(\mathcal{Q}\mathcal{H}^\tau\mathbf{x})d\tau\, ds.$$

Clearly this gives

$$|\mathcal{Q}\mathcal{H}^{\Delta t}\mathbf{x} - \mathcal{Q}\mathcal{T}^{\Delta t}\mathbf{x}| \le \|\mathbf{F}\|_\infty \cdot \Delta t^2.$$

Note that we also get $\forall\ 0 \le s \le \Delta t$,

$$\|\mathcal{Q}\mathcal{H}^{\Delta t}\mathbf{x} - \mathcal{Q}\mathbf{x}\| \le c_1 \cdot (1 + |\mathcal{P}\mathbf{x}|)\Delta t$$

where c_1 only depends on $\|\mathbf{F}\|_\infty$. Since

$$\mathcal{P}\mathcal{H}^{\Delta t}\mathbf{x} = \mathcal{P}\mathbf{x} + \int_0^{\Delta t}\mathbf{F}(\mathcal{Q}\mathcal{H}^s\mathbf{x})ds$$

by the periodicity of $\mathbf{F}$, we have

$$\begin{aligned}|\mathcal{P}\mathcal{H}^{\Delta t}\mathbf{x} - \mathcal{P}\mathbf{x} - \mathbf{F}(\mathcal{Q}\mathbf{x})\Delta t| &\le \int_0^{\Delta t}|\mathbf{F}(\mathcal{Q}\mathcal{H}^s\mathbf{x}) - \mathbf{F}(\mathcal{Q}\mathbf{x})|ds \\ &\le c_2 \cdot \Delta t \cdot \|\nabla\mathbf{F}\|\int_0^{\Delta t}(1 + |\mathcal{P}\mathbf{x}|)ds \\ &\le c \cdot (1 + |\mathcal{P}\mathbf{x}|)\Delta t^2\end{aligned}$$

where in the above c and c_2 are constants depending only on $\mathbf{F}$. □

COROLLARY 3.15 *There exists a constant* $c > 0$ *depending only on the potential* Φ *and* β *such that* $\forall\mathbf{x} \in \Gamma$, $0 < \Delta t \le \Delta t^*$*, we have*

$$|\rho(\mathcal{H}^{\Delta t}\mathbf{x}) - \rho(\mathcal{T}^{\Delta t}\mathbf{x})| \le c \cdot g_{\beta/2}(\mathcal{P}\mathbf{x}) \cdot (1 + |\mathcal{P}\mathbf{x}|^2) \cdot \Delta t^2.$$

PROOF: Denote $\mathbf{x}_1 := \mathcal{H}^{\Delta t}\mathbf{x}$ and $\mathbf{x}_2 := \mathcal{T}^{\Delta t}\mathbf{x}$; then we have

$$\begin{aligned}|\rho(\mathbf{x}_1) - \rho(\mathbf{x}_2)| &= \left|\int_0^1 (\mathbf{x}_2 - \mathbf{x}_1) \cdot \nabla\rho(\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1))dt\right| \\ &\le |\mathbf{x}_2 - \mathbf{x}_1| \max_{0\le t\le 1}|\nabla\rho(\mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1))|.\end{aligned}$$

Now the corollary follows easily by using Lemma 3.13 and 3.14. □

LEMMA 3.16 *There exists a constant $c > 0$ that depends only on the potential Φ such that for $\forall\, 0 < \Delta t \le \Delta t^*$, $\mathbf{x} \in \Gamma$, we have*

$$\left|\det\left(\frac{\partial(\mathcal{T}^{\Delta t}\mathbf{x})}{\partial \mathbf{x}}\right) - 1\right| \le c\Delta t^2,$$
$$\left|\det\left(\frac{\partial(\mathcal{T}^{-\Delta t}\mathbf{x})}{\partial \mathbf{x}}\right) - 1\right| \le c\Delta t^2.$$

PROOF: This is a simple estimate on the determinant. □

With the preceding lemmas and corollaries, we have the following:

LEMMA 3.17 *There exists a constant $c > 0$ independent of Δt such that for $\forall\, 0 < \Delta t \le \Delta t^*$, we have*

$$\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV} \le c\Delta t^2.$$

Remark 3.18. The meaning of the error term $\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV}$ is intuitively clear: it is the error induced by evolving the canonical ensemble by two different operators.

PROOF: By definition, Lemmas 3.13, 3.14, and 3.16, and Corollary 3.15, we have that

$$\begin{aligned}
&\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV} \\
&\quad = \frac{1}{2}\int_\Gamma |\rho(\mathcal{H}^{-\Delta t}\mathbf{x}) - \rho(\mathcal{T}^{-\Delta t}\mathbf{x}) \cdot \left(\left|\det\left(\frac{\partial(\mathcal{T}^{-\Delta t}\mathbf{x})}{\partial \mathbf{x}}\right)\right|\right)| d\mathbf{x} \\
&\quad \le \frac{1}{2}\int_\Gamma |\rho(\mathcal{H}^{-\Delta t}\mathbf{x}) - \rho(\mathcal{T}^{-\Delta t}\mathbf{x})| d\mathbf{x} \\
&\qquad + \frac{1}{2}\int_\Gamma \rho(\mathcal{T}^{-\Delta t}\mathbf{x})\left|1 - \left(\left|\det\left(\frac{\partial(\mathcal{T}^{-\Delta t}\mathbf{x})}{\partial \mathbf{x}}\right)\right|\right)\right| d\mathbf{x} \\
&\quad \le \frac{1}{2}\int_\Gamma |\rho(\mathcal{H}^{-\Delta t}\mathbf{x}) - \rho(\mathcal{T}^{-\Delta t}\mathbf{x})| d\mathbf{x} + \frac{c}{2}\Delta t^2 \int_\Gamma |\rho(\mathcal{T}^{-\Delta t}\mathbf{x})| d\mathbf{x} \\
&\quad = \frac{1}{2}\int_\Gamma |\rho(\mathbf{x}) - \rho(\mathcal{T}^{-\Delta t}\mathbf{x})| d\mathbf{x} + \frac{c}{2}\Delta t^2 \int_\Gamma |\rho(\mathcal{T}^{-\Delta t}\mathbf{x})| d\mathbf{x}
\end{aligned}$$

where the last inequality follows from the fact that the ρ only depends on the energy of the system.

Now denote $\mathbf{y} = \mathcal{T}^{-\Delta t}\mathbf{x}$ and note that

$$\left|\det\left(\frac{\partial(\mathcal{T}^{\Delta t}\mathbf{y})}{\partial \mathbf{y}}\right)\right| \le 1 + c\Delta t^2.$$

Then we have by a change of variables from $\mathbf{x}$ to $\mathbf{y}$,

$$\begin{aligned}
&\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV} \\
&\quad \le \frac{1}{2}\int_\Gamma |\rho(\mathcal{T}^{\Delta t}\mathbf{y}) - \rho(\mathbf{y})|(1 + c\Delta t^2)d\mathbf{y} + \frac{c}{2}\Delta t^2 \int_\Gamma \rho(\mathbf{y})(1 + c\Delta t^2)d\mathbf{y} \\
&\quad = \frac{1}{2}\int_\Gamma |\rho(\mathcal{T}^{\Delta t}\mathbf{y}) - \rho(H^{\Delta t}y)|(1 + c\Delta t^2)d\mathbf{y} + \frac{c}{2}\Delta t^2(1 + c\Delta t^2) \\
&\quad \le \frac{c}{2}\Delta t^2(1 + c\Delta t^2)\cdot\left(1 + \int_\Gamma g_{\beta/2}(\mathcal{P}\mathbf{x})(1 + |\mathcal{P}\mathbf{x}|^2)d\mathbf{x}\right) \\
&\quad = \frac{c}{2}\Delta t^2(1 + c\Delta t^2)\cdot\left(1 + \int_{\mathbb{R}^{dN}} g_{\beta/2}(\mathbf{v})(1 + |\mathbf{v}|^2)d\mathbf{v}\right) \\
&\quad \le c\Delta t^2
\end{aligned}$$

where in the above we have denoted by c the constants, which may vary from line to line but do not depend on Δt. □

Now we are ready to prove our main theorem:

THEOREM 3.19 *There exists a constant* $\Delta t^* > 0$ *such that* $\forall\ 0 < \nu \le \nu_0$*, where* ν_0 *is arbitrary but fixed, and* $\forall\ 0 < \Delta t \le \min\{1/2\nu_0, \Delta t^*\}$*, the following estimate holds*:

$$\|\mu_\nu^{\Delta t} - \pi\|_{TV} \le \frac{c}{\nu^{2N}}\Delta t|\log \Delta t|$$

where $c = c(\nu_0)$ *is a positive constant.*

PROOF: Denote by c the constants that only depend on ν_0 but may vary from line to line. By Lemma 3.17, we have

$$\|\mu_\nu^{\Delta t} - \pi\|_{TV} \le c \cdot \exp(-\nu^{2N}\kappa n\Delta t) + n \cdot c\Delta t^2.$$

Take the integer n to be such that

$$-\log \Delta t \le \nu^{2N}\kappa n\Delta t \le -10\log \Delta t.$$

Clearly such an n exists, and we have

$$\begin{aligned}
\|\mu_\nu^{\Delta t} - \pi\|_{TV} &\le c\Delta t + \frac{1}{\nu^{2N}}c\Delta t|\log \Delta t| \\
&\le \frac{c}{\nu^{2N}}\Delta t|\log \Delta t|.
\end{aligned}$$

□

4 Velocity Verlet Approximation

4.1 Ergodicity of the Velocity Verlet Andersen Process

The velocity Verlet approximation of the equation of motion (1.1) with time step Δt is given by

$$\begin{cases} \mathbf{q}^{\Delta t} = \mathbf{q} + \mathbf{v}\Delta t, \\ \mathbf{v}^{\Delta t} = \mathbf{v} + \frac{\mathbf{F}(\mathbf{q})+\mathbf{F}(\mathbf{q}^{\Delta t})}{2}\Delta t, \end{cases}$$

where we have denoted $\mathbf{F}(\mathbf{q}) := -\nabla\Phi(\mathbf{q})$. Let us denote by $\mathcal{T}^{\Delta t}$ the map

$$\mathbf{x} = (\mathbf{q}, \mathbf{v}) \longrightarrow \mathbf{x}^{\Delta t} = (\mathbf{q}^{\Delta t}, \mathbf{v}^{\Delta t}).$$

Note that $\mathcal{T}^{\Delta t}$ is not a flow operator in the sense that

$$\mathcal{T}^{\Delta t_1 + \Delta t_2} \neq \mathcal{T}^{\Delta t_1} \circ \mathcal{T}^{\Delta t_2}.$$

DEFINITION 4.1 (Velocity Verlet Andersen Process) Let $\{\alpha_n\}_{n=1}^{\infty}$ be i.i.d. random variables such that $\mathbb{P}(\alpha_n = 1) = \lambda = \nu\Delta t$ and $\mathbb{P}(\alpha_n = 0) = 1 - \lambda = 1 - \nu\Delta t$. Let $\mathcal{S}$, Y_n, and Z_n be as in Definition 1.1. The *discrete-time velocity Verlet Andersen process* is defined as

$$\mathbf{x}_{n+1} = (1 - \alpha_n)\mathcal{T}^{\Delta t}\mathbf{x}_n + \alpha_n \mathcal{S}(Y_n, Z_n)\mathcal{T}^{\Delta t}\mathbf{x}_n.$$

Let $\mathcal{T}^{-\Delta t}$ be the Markov operator associated with $\mathcal{T}^{\Delta t}$. The evolution equation for the probability measures $\mu_n^{\Delta t}$ is given by

$$\mu_{n+1}^{\Delta t} = (1 - \lambda)\mathcal{T}^{-\Delta t}\mu_n^{\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t}\mu_n^{\Delta t}.$$

Note that if there is a stationary probability measure $\mu^{\Delta t}$ for the velocity Verlet Andersen process, then it has to satisfy

$$\mu^{\Delta t} = (1 - \lambda)\mathcal{T}^{-\Delta t}\mu^{\Delta t} + \lambda\mathcal{A}\mathcal{T}^{-\Delta t}\mu^{\Delta t}. \tag{4.1}$$

LEMMA 4.2 *Assume* $0 < \Delta t_0 \leq \varepsilon_0\|\nabla\mathbf{F}\|$. ε_0 *is sufficiently small such that*

$$\varepsilon_0 \leq \min\left\{\frac{1}{\sqrt{8N\|\nabla\mathbf{F}\|^3}}, \frac{1}{\|\nabla\mathbf{F}\|}\right\}.$$

Let $c = 100\|\nabla\mathbf{F}\|\sqrt{d}$. *Then for any* $0 < \Delta t \leq \mathrm{q}\Delta t_0$, $n\Delta t \leq \Delta t_0$, *and* $\forall\, 1 \leq i, j \leq N$, *the following inequalities hold*:

$$\left\|\frac{\partial(\mathcal{Q}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})} - \delta_{ij}I\right\| \leq c \cdot (n\Delta t)^2,$$

$$\left\|\frac{\partial(\mathcal{P}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} - \delta_{ij}I\right\| \leq c \cdot n\Delta t,$$

$$\left\|\frac{\partial(\mathcal{Q}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{P}_i\mathbf{x})} - n\Delta t\delta_{ij}I\right\| \leq c \cdot (n\Delta t)^2,$$

$$\left\|\frac{\partial(\mathcal{P}_j(\mathcal{T}^{\Delta t})^n\mathbf{x})}{\partial(\mathcal{Q}_i\mathbf{x})}\right\| \leq c \cdot n\Delta t.$$

PROOF: We shall prove this lemma by induction. Note first by definition of $\mathcal{T}^{\Delta t}$, it is not hard to verify that the following holds $\forall\ 1 \le j, k \le N$:

$$\begin{aligned}
\frac{\partial(\mathcal{Q}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_k \mathbf{x})} &= \delta_{jk} I + \frac{1}{2}\Delta t^2 \frac{\partial(\mathbf{F}_j(\mathcal{Q}\mathbf{x}))}{\partial(\mathcal{Q}_k \mathbf{x})},\\
\frac{\partial(\mathcal{Q}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_k \mathbf{x})} &= \Delta t \delta_{jk} I,\\
\frac{\partial(\mathcal{P}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_k \mathbf{x})} &= \delta_{jk} I + \frac{1}{2}\Delta t^2 \frac{\partial(\mathbf{F}_j(\mathcal{Q}\mathcal{T}^{\Delta t}\mathbf{x}))}{\partial(\mathcal{Q}_k \mathcal{T}^{\Delta t}\mathbf{x})},\\
\frac{\partial(\mathcal{P}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_k \mathbf{x})} 7 &= \frac{1}{2}\Delta t \frac{\partial(\mathbf{F}_j(\mathcal{Q}\mathbf{x}))}{\partial(\mathcal{Q}_k \mathbf{x})}\\
&\quad + \frac{1}{2}\Delta t \sum_{h=1}^{N} \frac{\partial(\mathbf{F}_j(\mathcal{Q}\mathcal{T}^{\Delta t}\mathbf{x}))}{\partial(\mathcal{Q}_h \mathcal{T}^{\Delta t}\mathbf{x})}\left(\delta_{kh} I + \frac{1}{2}\Delta t^2 \frac{\partial(\mathbf{F}_h(\mathcal{Q}\mathbf{x}))}{\partial(\mathcal{Q}_k \mathbf{x})}\right).
\end{aligned}$$

The above can be put into a more compact form:

$$\begin{aligned}
\frac{\partial(\mathcal{Q}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_k \mathbf{x})} &= \delta_{jk} I + A_{jk}(\mathbf{x})\Delta t^2, &\qquad \frac{\partial(\mathcal{P}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_k \mathbf{x})} &= \delta_{jk} I + B_{jk}(\mathbf{x})\Delta t^2,\\
\frac{\partial(\mathcal{Q}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{P}_k \mathbf{x})} &= \Delta t \delta_{jk} I, &\qquad \frac{\partial(\mathcal{P}_j \mathcal{T}^{\Delta t}\mathbf{x})}{\partial(\mathcal{Q}_k \mathbf{x})} &= C_{jk}(\mathbf{x})\Delta t,
\end{aligned}$$

where an easy computation shows that $A_{jk}(\cdot)$, $B_{jk}(\cdot)$, and $C_{jk}(\cdot)$ are absolutely bounded by $2\|\nabla \mathbf{F}\|$. Define $\mathbf{x}_n := (\mathcal{T}^{\Delta t})^n \mathbf{x}$.

For $l = 0$, the inequalities obviously hold. Suppose for $l = n$ the inequalities hold. Now consider $l = n + 1$ and assume $(n + 1)\Delta t \le \Delta t_0$. Obviously

$$\frac{\partial(\mathcal{Q}_j \mathbf{x}_{n+1})}{\partial(\mathcal{Q}_i \mathbf{x})} = \frac{\partial(\mathcal{Q}_j \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} + \Delta t^2 \sum_{k=1}^{N} A_{jk}(\mathbf{x}_n)\frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} + \Delta t \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})}.$$

By the induction hypothesis

$$\begin{aligned}
&\left\| \frac{\partial(\mathcal{Q}_j \mathbf{x}_{n+1})}{\partial(\mathcal{Q}_i \mathbf{x})} - \delta_{ij} I \right\|\\
&\quad \le c \cdot (n\Delta t)^2 + 2\Delta t^2 \cdot \|\nabla \mathbf{F}\|(\sqrt{d} + cN(n\Delta t)^2) + c \cdot (n\Delta t)\Delta t\\
&\quad \le c(n^2 + n)\Delta t^2 + 2\Delta t^2 \cdot \|\nabla \mathbf{F}\|\sqrt{d} + 2\Delta t^2 \cdot \|\nabla \mathbf{F}\| \cdot c \cdot N\Delta t_0^2\\
&\quad \le c(n^2 + n)\Delta t^2 + \frac{1}{2}\Delta t^2 \cdot c + \frac{1}{2}\Delta t^2 \cdot c\\
&\quad \le c \cdot (n + 1)^2 \Delta t^2.
\end{aligned}$$

Similarly, we have

$$\frac{\partial(\mathcal{Q}_j \mathbf{x}_{n+1})}{\partial(\mathcal{P}_i \mathbf{x})} = \frac{\partial(\mathcal{Q}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} + \Delta t \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} + \Delta t^2 \sum_{k=1}^{N} A_{jk}(\mathbf{x}_n) \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})}.$$

Hence

$$\begin{aligned}
&\left\| \frac{\partial(\mathcal{Q}_j \mathbf{x}_{n+1})}{\partial(\mathcal{P}_i \mathbf{x})} - (n+1)\Delta t \delta_{ij} I \right\| \\
&\quad \leq c \cdot (n\Delta t)^2 + c \cdot (n\Delta t)\Delta t + 2\Delta t^2 \cdot \|\nabla \mathbf{F}\| \cdot (\sqrt{d} n \Delta t + cN(n\Delta t)^2) \\
&\quad \leq c \cdot (n^2 + n) \cdot \Delta t^2 + 2\Delta t^2 \|\nabla \mathbf{F}\| \cdot \Delta t_0 \sqrt{d} + 2\Delta t^2 \|\nabla \mathbf{F}\| \cdot cN\Delta t_0^2 \\
&\quad \leq c \cdot (n+1)^2 \Delta t^2.
\end{aligned}$$

The proof of the remaining inequalities is similar:

$$\begin{aligned}
&\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_{n+1})}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| \\
&= \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} + \Delta t \sum_{k=1}^{N} C_{jk}(\mathbf{x}_n) \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} + \Delta t^2 \sum_{k=1}^{N} B_{jk}(\mathbf{x}_n) \frac{\partial(\mathcal{P}_k \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| \\
&\leq \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| + 2\Delta t \|\nabla \mathbf{F}\| \cdot \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| + 2\Delta t^2 \|\nabla \mathbf{F}\| \cdot \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{P}_k \mathbf{x}_n)}{\partial(\mathcal{Q}_i \mathbf{x})} \right\| \\
&\leq c \cdot n\Delta t + 2\Delta t \|\nabla \mathbf{F}\| \cdot (\sqrt{d} + c \cdot N(n\Delta t)^2) + 2\Delta t^2 \|\nabla \mathbf{F}\| \cdot (c \cdot (n\Delta t)) \cdot N \\
&\leq c \cdot (n+1)\Delta t,
\end{aligned}$$

$$\begin{aligned}
&\left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_{n+1})}{\partial(\mathcal{P}_i \mathbf{x})} - \delta_{ij} I \right\| \\
&= \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} - \delta_{ij} I + \Delta t \sum_{k=1}^{N} C_{jk}(\mathbf{x}_n) \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} + \Delta t^2 \sum_{k=1}^{N} B_{jk}(\mathbf{x}_n) \frac{\partial(\mathcal{P}_k \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} \right\| \\
&\leq \left\| \frac{\partial(\mathcal{P}_j \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} - \delta_{ij} I \right\| \\
&\quad + 2\Delta t \|\nabla \mathbf{F}\| \cdot \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{Q}_k \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} \right\| + 2\Delta t^2 \|\nabla \mathbf{F}\| \cdot \sum_{k=1}^{N} \left\| \frac{\partial(\mathcal{P}_k \mathbf{x}_n)}{\partial(\mathcal{P}_i \mathbf{x})} \right\|
\end{aligned}$$

$$\leq c \cdot n\Delta t + 2\Delta t \|\nabla \mathbf{F}\| \cdot (n\Delta t\sqrt{d} + c \cdot N(n\Delta t)^2)$$

$$+ 2\Delta t^2 \|\nabla \mathbf{F}\| \cdot (\sqrt{d} + c \cdot N \cdot (n\Delta t))$$

$$\leq c \cdot (n+1)\Delta t$$

□

Using this lemma we can similarly establish a series of lemmas that are not much different from the ones we gave in previous sections. These give us the following:

THEOREM 4.3 (Ergodicity of the Velocity Verlet Andersen Process) *Let $\nu_0 > 0$ be arbitrary but fixed. There exists $\Delta t^* > 0$ and a constant $\kappa = \kappa(\nu_0)$, independent of $(n, \Delta t, \mathbf{x}')$, such that for any $0 < \nu < \nu_0$, $0 < \Delta t < \min\{1/2\nu_0, \Delta t^*\}$, the velocity Verlet Andersen process has a unique invariant probability measure $\mu^{\Delta t}$. Furthermore, for any $\mathbf{x}' \in \Gamma$ and $n \geq 1$, we have*

$$\|(Q_{\mathbf{x}'}^{\Delta t})^n - \mu^{\Delta t}\|_{TV} \leq c \cdot \exp(-\kappa \nu^{2N} n \Delta t),$$

where c is a positive constant independent of $(\nu, \nu_0, \mathbf{x}', \Delta t)$.

4.2 Regularity of the Invariant Measure

The proof of the following theorem is not much different from the case for the forward Euler Andersen process. Therefore we omit the details here.

THEOREM 4.4 *The invariant probability measure $\mu^{\Delta t}$ for the velocity Verlet Andersen process is absolutely continuous with respect to the Lebesgue measure on Γ; i.e., the Radon-Nikodym derivative $\frac{d\mu}{d\mathbf{x}}$ is in $L^1(\Gamma, d\mathbf{x})$.*

Let us note that as in the forward Euler case, one has an explicit expression for the probability measure $\mathcal{A}_{\text{pr}}^N \mathcal{T}^{-\Delta t} \mathcal{A}_{\text{pr}}^N \delta_{\mathbf{x}'}$:

$$(\mathcal{A}_{\text{pr}}^N \mathcal{T}^{-\Delta t} \mathcal{A}_{\text{pr}}^N \delta_{\mathbf{x}'})(d\mathbf{q}, d\mathbf{v}) = \frac{1}{\Delta t^{dN}} g_\beta(\mathbf{v}) \cdot h(\mathbf{q}; \mathbf{q}') \cdot d\mathbf{q}\, d\mathbf{v}$$

where

$$h(\mathbf{q}) = \sum_{\mathbf{k} \in \mathbb{Z}^{dN}} g_\beta \left(\frac{\mathbf{q} - \mathbf{q}' + \mathbf{k}}{\Delta t} - \frac{\mathbf{F}(\mathbf{q}')}{2} \Delta t \right).$$

4.3 Error Analysis for Velocity Verlet Approximation

The error analysis for the velocity Verlet is not much different from the forward Euler case.

LEMMA 4.5 *Let Δt^* be the same as in Theorem 4.3. There exists a positive constant c that depends only on the potential Φ such that the following holds true $\forall \mathbf{x} \in \Gamma$, $0 < \Delta t \leq \Delta t^*$:*

$$|\mathcal{H}^{\Delta t}\mathbf{x} - \mathcal{T}^{\Delta t}\mathbf{x}| \leq c \cdot (1 + |\mathcal{P}\mathbf{x}|^2)\Delta t^3.$$

PROOF: In this proof we use c to denote constants that depend on the potential Φ but may vary from line to line. By definition of $\mathcal{H}^{\Delta t}$, we have the following:

$$\begin{aligned}\mathcal{Q}\mathcal{H}^{\Delta t}\mathbf{x} &= \mathcal{Q}\mathbf{x} + \Delta t\mathcal{P}\mathbf{x} + \frac{\Delta t^2}{2}\mathbf{F}(\mathcal{Q}\mathbf{x}) \\ &\quad + \int_0^{\Delta t}\int_0^t\int_0^\tau \frac{\partial(\mathbf{F}(\mathcal{Q}\mathcal{H}^s\mathbf{x})}{\partial(\mathcal{Q}\mathcal{H}^s\mathbf{x})}(\mathcal{Q}\mathcal{H}^s\mathbf{x})ds\,d\tau\,dt \\ &= \mathcal{Q}\mathcal{T}^{\Delta t}\mathbf{x} + \int_0^{\Delta t}\int_0^t\int_0^\tau \frac{\partial(\mathbf{F}(\mathcal{Q}\mathcal{H}^s\mathbf{x})}{\partial(\mathcal{Q}\mathcal{H}^s\mathbf{x})}(\mathcal{Q}\mathcal{H}^s\mathbf{x})ds\,d\tau\,dt.\end{aligned}$$

Since $|\mathcal{P}\mathcal{H}^s\mathbf{x} - \mathcal{P}\mathbf{x}| \le s\cdot\|\mathbf{F}\|_\infty$, we have

$$|\mathcal{Q}\mathcal{H}^{\Delta t}\mathbf{x} - \mathcal{Q}\mathcal{T}^{\Delta t}\mathbf{x}| \le c\cdot\|\nabla\mathbf{F}\|(1+|\mathcal{P}\mathbf{x}|)\Delta t^3 \le c(1+|\mathcal{P}\mathbf{x}|)\Delta t^3.$$

It remains for us to show that $|\mathcal{P}\mathcal{H}^{\Delta t}\mathbf{x} - \mathcal{P}\mathcal{T}^{\Delta t}\mathbf{x}| = O(\Delta t^3)$. Indeed, for $0 \le t \le \Delta t$, define $\mathbf{G}(t;\mathbf{x}) := \mathbf{F}(\mathcal{Q}\mathcal{H}^t\mathbf{x})$. It is not hard to verify that $\mathbf{G}$ satisfies the inequality

$$\left|\frac{\partial^2\mathbf{G}(t;\mathbf{x})}{\partial t^2}\right| \le c(1+|\mathcal{P}\mathbf{x}|^2);$$

then we have

$$\begin{aligned}|\mathcal{P}\mathcal{T}^{\Delta t}\mathbf{x} - \mathcal{P}\mathcal{H}^{\Delta t}\mathbf{x}| &= \left|\frac{\mathbf{F}(\mathcal{Q}\mathcal{T}^{\Delta t}\mathbf{x}) + \mathbf{F}(\mathcal{Q}\mathbf{x})}{2}\Delta t - \int_0^{\Delta t}\mathbf{F}(\mathcal{Q}\mathcal{H}^t\mathbf{x})dt\right| \\ &\le \frac{\Delta t}{2}|\mathbf{F}(\mathcal{Q}\mathcal{T}^{\Delta t}\mathbf{x}) - \mathbf{F}(\mathcal{Q}\mathcal{H}^{\Delta t}\mathbf{x})| \\ &\quad + \left|\frac{\mathbf{F}(\mathcal{Q}\mathcal{H}^{\Delta t}\mathbf{x}) + \mathbf{F}(\mathcal{Q}\mathbf{x})}{2}\Delta t - \int_0^{\Delta t}\mathbf{F}(\mathcal{Q}\mathcal{H}^t\mathbf{x})dt\right| \\ &\le c(1+|\mathcal{P}\mathbf{x}|)\Delta t^4 + \frac{\Delta t}{2}\left|\int_0^{\Delta t}\int_0^\tau \frac{\partial^2\mathbf{G}(s;\mathbf{x})}{\partial s^2}ds\,d\tau\right| \\ &\quad + \left|\int_0^{\Delta t}\int_0^t\int_0^\tau \frac{\partial^2\mathbf{G}(s;\mathbf{x})}{\partial s^2}ds\,d\tau\,dt\right| \\ &\le c(1+|\mathcal{P}\mathbf{x}|^2)\Delta t^3.\end{aligned}$$

□

LEMMA 4.6 *Denote by ρ the density of the Gibbsian probability distribution. There exists a constant $c > 0$ depending only on the potential Φ and β such that $\forall\mathbf{x}\in\Gamma$, $0 < \Delta t \le \Delta t^*$, we have*

$$|\rho(\mathcal{H}^{\Delta t}\mathbf{x}) - \rho(\mathcal{T}^{\Delta t}\mathbf{x})| \le c\cdot g_{\beta/2}(\mathcal{P}\mathbf{x})\cdot(1+|\mathcal{P}\mathbf{x}|^3)\cdot\Delta t^3.$$

PROOF: This follows easily from the last lemma. □

LEMMA 4.7 *Let Δt^* be the same as in Theorem 4.3. There exists a constant $c > 0$ independent of Δt such that $\forall\, 0 < \Delta t \le \Delta t^*$, we have*

$$\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV} \le c\Delta t^3.$$

PROOF: Since ρ depends only on the energy of the system, and the velocity Verlet scheme is volume preserving, we have

$$\begin{aligned}\|(\mathcal{H}^{-\Delta t} - \mathcal{T}^{-\Delta t})\pi\|_{TV} &= \frac{1}{2}\int_\Gamma |\rho(\mathcal{H}^{-\Delta t}\mathbf{x}) - \rho(\mathcal{T}^{-\Delta t}\mathbf{x})|d\mathbf{x} \\ &= \frac{1}{2}\int_\Gamma |\rho(\mathbf{x}) - \rho(\mathcal{T}^{-\Delta t}\mathbf{x})|d\mathbf{x} \\ &= \frac{1}{2}\int_\Gamma |\rho(\mathcal{T}^{\Delta t}\mathbf{x}) - \rho(\mathbf{x})|d\mathbf{x} \\ &= \frac{1}{2}\int_\Gamma |\rho(\mathcal{T}^{\Delta t}\mathbf{x}) - \rho(\mathcal{H}^{\Delta t}\mathbf{x})|d\mathbf{x}.\end{aligned}$$

The rest of this proof follows easily from Lemma 4.6. □

Now we are ready to prove the main result of this section:

THEOREM 4.8 *There exists a constant $\Delta t^* > 0$ such that $\forall\ 0 < \nu \le \nu_0$, where ν_0 is arbitrary but fixed, and $\forall\ 0 < \Delta t \le \min\{1/2\nu_0, \Delta t^*\}$, the following holds*:

$$\|\mu^{\Delta t} - \pi\|_{TV} \le \frac{c}{\nu^{2N}}\,\Delta t^2|\log \Delta t|$$

where $c = c(\nu_0)$ is a positive constant.

PROOF: Let us denote by c the positive constant that depends on ν_0 but may vary from line to line; then by Lemma 4.7, we have

$$\|\mu^{\Delta t} - \pi\|_{TV} \le c \cdot \exp(-\nu^{2N}\kappa n\Delta t) + n \cdot c\Delta t^3.$$

Take the integer n to be such that

$$-2\log \Delta t \le \nu^{2N}\kappa n\Delta t \le -10\log \Delta t.$$

Clearly such an n exists, and we have

$$\begin{aligned}\|\mu^{\Delta t} - \pi\|_{TV} &\le c\Delta t^2 + \frac{1}{\nu^{2N}}\, c\Delta t^2|\log \Delta t| \\ &\le \frac{c}{\nu^{2N}}\Delta t^2|\log \Delta t|.\end{aligned}$$

□

5 Transport Properties: Diffusion Matrix for the Lorentz Gas Model

In previous sections we have shown rigorously that the Andersen thermostat correctly generates the canonical ensemble and therefore can be used to calculate equilibrium properties. However, since it uses fictitious dynamics, it is not intuitively clear whether the Andersen thermostat can be used to calculate dynamic quantities such as transport coefficients.

To illustrate what may go wrong, let us consider a counterexample. Imagine that the system consists of free-streaming particles; i.e., the interatomic potential is 0. Since the process is uniformly ergodic, we have by the functional central limit theorem [2] that the diffusion matrix is given by

$$D_\beta = 2 \int \mathbf{v} \otimes h(\mathbf{q}, \mathbf{v}) g_\beta(\mathbf{v}) d\mathbf{q}\, d\mathbf{v}$$

where h is the solution to the problem

$$-\mathbf{v} = \mathcal{G} h.$$

By inspection $h(\mathbf{q}, \mathbf{v}) = \frac{1}{\nu}\mathbf{v}$, so that

$$D_\beta = \frac{2}{\nu} I$$

where I is the identity matrix. Note that D_β depends rather sensitively on ν and therefore is not physical.

The above simple example might seem to suggest that the Andersen thermostat should never be used to calculate the diffusion matrix. However, further consideration shows that this is not always the case. The trouble with the counterexample is that there is only one time scale present: In the absence of the thermostat, the particles simply perform free streaming and there is no diffusion at all. Such is not the case in a generic Hamiltonian system thermostatted by Andersen. In general, there are two competing time scales: One is set by the maximal Lyapunov exponent of the chaotic Hamiltonian dynamics and gives rise to diffusion. The other is set by the collision frequency of the thermostat. In the limit of vanishing collision frequency, the system should spend most of its time doing "deterministic diffusion." Therefore the diffusion constant calculated using the Andersen thermostat should not be too different from the deterministic diffusion constant.

To put the above heuristic discussion into rigorous terms, we must first look for a deterministic system for which there is diffusion. This amounts to proving the convergence of the Green-Kubo formula. At present, the best-known example is the Lorentz gas model [7]. In [3], Bunimovich and Sinai proved the stretched exponential decay of the velocity autocorrelation function and thus rigorously proved the validity of the Green-Kubo formula.

In the Lorentz gas model, the billiard moves freely between its collision with the scatters. The collision is assumed to be elastic and therefore the modulus of the velocity is conserved. The scatters are immobile disks, and they are periodically situated in space and do not overlap each other. As usual, the Lorentz gas is assumed to have a finite horizon; i.e., the length of the free path of the billiard is bounded. We shall study the Lorentz gas model with the Andersen thermostat; i.e., during exponentially distributed time intervals, the particle forgets its old velocity and chooses a new velocity from the Maxwell-Boltzmann distribution at a given temperature $\frac{1}{\beta}$. It is standard to show that this process is well-defined as a Markov process. Note that the stationary distribution is given by

$$\mu_0(d\mathbf{q}\, d\mathbf{v}) \propto d\,q \otimes g_\beta(\mathbf{v})d\mathbf{v}.$$

Let $\mathcal{T}^t$ denote the deterministic flow operator in the phase space; then Sinai's result implies that the following Green-Kubo formula converges absolutely:

$$D_\beta := \int_0^\infty \mathbb{E}^{\mu_0}(\mathbf{v}(X_0) \otimes \mathbf{v}(\mathcal{T}^t X_0))dt.$$

We shall prove that in the limit of vanishing collision frequency, the diffusion matrix of the Andersen thermostatted Lorentz gas converges to the deterministic diffusion matrix D_β:

$$\lim_{\nu \downarrow 0} \int_0^\infty \mathbb{E}^{\mu_0}\Big(\mathbf{v}(X_0) \otimes \mathbf{v}(X_s)\Big)ds = D_\beta. \tag{5.1}$$

Note that in (5.1) we have suppressed the dependence of X_t on ν to avoid cumbersome notation. Let us note that the Lorentz flow operator T^t preserves the magnitude of velocities. As a consequence, the following lemma is trivially true:

LEMMA 5.1 *Assume* $\mathbf{q}_0$ *is an arbitrary but fixed point in the configuration space of the Lorentz gas. Let Z be a centered Gaussian random variable with values in* $\mathbb{R}^2$*. Then for any* $t \geq 0$*, the random variable*

$$\mathbf{v} = \mathbf{v}(T^t(q_0, Z))$$

has the same distribution as Z. In particular, $\mathbb{E}\mathbf{v} = 0$*.*

PROOF: This is obvious. □

Using this lemma, we can prove the following result:

THEOREM 5.2 *Equation* (5.1) *holds*; *i.e., in the limit of vanishing collision frequency, the diffusion matrix of Andersen thermostatted Lorentz gas converges to that of the deterministic Lorentz gas.*

PROOF: Adopt the same notation as in Definition 1.1. For any $t > 0$, let us denote by $\mathcal{F}_t := \mathcal{F}\{T_1, \dots, T_{N_t}\}$ the filtration generated by $T_1, \dots, T_{N_t}$. We have

$$\begin{aligned}
&\mathbb{E}^{\mu_0} \int_0^t \mathbf{v}(X_0) \otimes \mathbf{v}(X_s)ds \\
&\quad = \mathbb{E}^{\mu_0}\left(\sum_{i=1}^{N_t} \int_{\sum_{k=1}^{i-1} T_k}^{\sum_{k=1}^{i} T_k} \mathbf{v}(X_0) \otimes \mathbf{v}(X_s)ds + \int_{\sum_{k=1}^{N_t} T_k}^{t} \mathbf{v}(X_0) \otimes \mathbf{v}(X_s)ds\right) \\
&\quad = \mathbb{E}^{\mu_0}\left(\int_0^{T_1 \wedge t} \mathbf{v}(X_0) \otimes \mathbf{v}(X_s)ds\right) + \mathrm{I} + \mathrm{II}
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{I} &= \mathbb{E}^{\mu_0}\Big(\sum_{2\le i\le N_t}\int_{\sum_{k=1}^{i-1}T_k}^{\sum_{k=1}^{i}T_k}\mathbf{v}(X_0)\otimes\mathbf{v}(X_s)ds\Big)\\
&= \mathbb{E}^{\mu_0}\Big(\sum_{2\le i\le N_t}\int_{\sum_{k=1}^{i-1}T_k}^{\sum_{k=1}^{i}T_k}\mathbb{E}\big(\mathbf{v}(X_0)\otimes\mathbf{v}(X_s)\bigm|\mathcal{F}_t\vee\mathcal{F}\{X_0,Z_1,\dots,Z_{i-2}\}\big)ds\Big)\\
&= \mathbb{E}^{\mu_0}\Big(\sum_{2\le i\le N_t}\int_{\sum_{k=1}^{i-1}T_k}^{\sum_{k=1}^{i}T_k}\mathbf{v}(X_0)\otimes\mathbb{E}\big(\mathbf{v}(X_s)\bigm|\mathcal{F}_t\vee\mathcal{F}\{X_0,Z_1,\dots,Z_{i-2}\}\big)ds\Big).
\end{aligned}$$

Note that in the above the random variables Z_i are defined in Definition 1.1 and $\mathcal{F}\{X_0, Z_1, \dots, Z_{i-2}\}$ is the σ-algebra generated by $X_0, Z_1, \dots, Z_{i-2}$ (if $i = 2$, then it is simply $\mathcal{F}\{X_0\}$). Now by a conditional version of Lemma 5.1, we have

$$\mathbb{E}\big(\mathbf{v}(X_s)\bigm|\mathcal{F}_t\vee\mathcal{F}(X_0,Z_1,\dots,Z_{i-2})\big)=0.$$

This shows that $\mathrm{I} = 0$. Similarly,

$$\mathrm{II}=\mathbb{E}^{\mu_0}\mathbf{1}_{N_t\ge1}\int_{\sum_{k=1}^{N_t}T_k}^{t}\mathbf{v}(X_0)\otimes\mathbf{v}(X_s)ds=0$$

so that

$$\mathbb{E}^{\mu_0}\int_0^t\mathbf{v}(X_0)\otimes\mathbf{v}(X_s)ds=\mathbb{E}^{\mu_0}\int_0^{T_1\wedge t}\mathbf{v}(X_0)\otimes\mathbf{v}(X_s)ds.$$

Now define $h(t) := \int_0^t \mathbb{E}^{\mu_0}(\mathbf{v}(X_0)\otimes\mathbf{v}(\mathcal{T}^s X_0))ds$. Note that by Bunimovich and Sinaĭ's result, we have for any $t \ge 0$, there exists a constant $c_1 > 0$, independent of t, such that $|h(t)| \le c_1$ and $h(t) \to D_\beta$ as $t \to \infty$. Using this nontrivial fact, we have

$$\begin{aligned}
&\lim_{t\to\infty}\mathbb{E}^{\mu_0}\int_0^{T_1\wedge t}\mathbf{v}(X_0)\otimes\mathbf{v}(X_s)ds\\
&\quad=\lim_{t\to\infty}\int_0^\infty \nu e^{-s\nu}h(s\wedge t)ds\\
&\quad=\lim_{t\to\infty}\int_0^\infty e^{-s}h\Big(t\wedge\frac{s}{\nu}\Big)ds\\
&\quad=\int_0^\infty e^{-s}h\Big(\frac{s}{\nu}\Big)ds.
\end{aligned}$$

Now by taking $\nu \downarrow 0$, an application of the dominated convergence theorem yields (5.1). □

6 Concluding Remarks

In this article we have reformulated the continuous-time and discrete-time Andersen process. The main difficulty in proving ergodicity of this type of process is that the noise is degenerate; i.e., only the velocity variables are randomized. Our method of proof follows the intuitive picture that the Hamiltonian flow operator transports the smoothness in the velocity space into the configuration space. In this way the transition kernel has a smooth density (up to exponentially small-in-time terms). Furthermore, it approaches the equilibrium Gibbs measure exponentially fast. We also analyzed the numerical approximations of the Andersen process. Two representative methods are studied: forward Euler and velocity Verlet. We proved existence of the invariant measure and absolute continuity of the invariant measure, and gave sharp error estimates in both cases. For the transport properties, we showed in the low-collision frequency limit that the diffusion matrix of the Lorentz gas system calculated by the Andersen thermostat is close to the true value. This is in accordance with what is observed in molecular dynamics simulations.

Finally, it is interesting to compare the Andersen thermostat to the well-known hybrid Monte Carlo methods (HMC) introduced by Duane et al. [5]. The HMC method also uses fictitious momentum kicks, but only to make trial moves in the configuration space. The main differences between the HMC method and the Andersen thermostat are the following: first, in HMC we are only interested in sampling the configuration space; the fictitious momentum variables are only introduced to advance the system in the configuration space, whereas in the Andersen thermostat the momentum variables are always kept. Therefore, if we are only interested in sampling the equilibrium measure, the HMC method is more efficient. However, if we also want to calculate the transport properties of the system, then HMC cannot be used. The second difference is that in HMC, one always needs to use symplectic integrators to keep detailed balance, while as we have shown here, the symplectic condition is not necessarily needed in the Andersen thermostat.

Bibliography

[1] Andersen, H. C. Molecular dynamics at constant pressure and/or temperature. *J. Chem. Phys.* **72** (1980), 2384–2393.

[2] Bhattacharya, R. N. On the functional central limit theorem and the law of the iterated logrithm for Markov processes. *Z. Wahrsch. Verw. Gebiete.* **60** (1982), no. 2, 185–201.

[3] Bunimovich, L. A.; Sinaĭ, Ya.G. Statistical properties of Lorentz gas with periodic configuration of scatterers. *Comm. Math. Phys.* **78** (1980/81), no. 4, 479–497.

[4] Chichilnisky, G. Topology and invertible maps. *Adv. in Appl. Math.* **21** (1998), no. 1, 113–123.

[5] Duane, S.; Kennedy, A. D.; Pendleton, B. J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **195** (1987), no. 2, 216–222.

[6] Lindvall, T. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, New York, 1992.

[7] Machta, J.; Zwanzig, R. Diffusion in a periodic Lorentz gas. *Phys. Rev. Lett.* **50** (1983), no. 25, 1959–1962.

[8] Nummelin, E. *General irreducible Markov chains and nonnegative operators*. Cambridge Tracts in Mathematics, 83. Cambridge University Press, Cambridge, 1984.

WEINAN E
Princeton University
Department of Mathematics
Fine Hall, Washington Road
Princeton, NJ 08544-1000
E-mail: weinan@math.princeton.edu

DONG LI
Princeton University
PACM
Fine Hall, Washington Road
Princeton, NJ 08544-1000
E:mail: dongli@ias.edu

Received December 2005.