# Swarm dynamics for global optimisation on finite sets

Nhat-Thang Le[†] and Laurent Miclo[‡]

[†]Toulouse School of Economics
Institut de Mathématiques de Toulouse
University of Toulouse

[‡]Toulouse School of Economics
Institut de Mathématiques de Toulouse
CNRS and University of Toulouse

**Abstract**

Consider the global optimisation of a function $U$ defined on a finite set $V$ endowed with an irreducible and reversible Markov generator. By integration, we extend $U$ to the set $\mathcal{P}(V)$ of probability distributions on $V$ and we penalise it with a time-dependent generalised entropy functional. Endowing $\mathcal{P}(V)$ with a Maas' Wasserstein-type Riemannian structure, enables us to consider an associated time-inhomogeneous gradient descent algorithm. There are several ways to interpret this $\mathcal{P}(V)$-valued dynamical system as the time-marginal laws of a time-inhomogeneous non-linear Markov process taking values in $V$, each of them allowing for interacting particle approximations. This procedure extends to the discrete framework the continuous state space swarm algorithm approach of Bolte, Miclo and Villeneuve [4], but here we go further by considering more general generalised entropy functionals for which functional inequalities can be proven. Thus in the full generality of the above finite framework, we give conditions on the underlying time dependence ensuring the convergence of the algorithm toward laws supported by the set of global minima of $U$. Numerical simulations illustrate that one has to be careful about the choice of the time-inhomogeneous non-linear Markov process interpretation.

**Keywords:** Finite global optimisation, swarm algorithms, non-linear finite Markov processes, interacting particle systems, Maas' Wasserstein-like metrics, generalised entropies, gradient flows, functional inequalities.

**MSC2020:** primary: 60J27, secondary: 60E15, 39A12, 37A30, 60J35, 65C05, 65C35.

**Fundings:** This work was supported by the grants ANR-17-EURE-0010 and AFOSR-22IOE016.

# 1 Introduction

The global minimization of a function $U$ given on a set $V$ is in general an important but difficult task. When $V$ is a compact and connected manifold and $U$ is smooth function, a time-inhomogeneous swarm algorithm was proposed in [4] to approach the set of global minimizers. Our purpose here is to deal with discrete optimisation problems and second to go beyond some technical restrictions that have appeared in [4], in particular concerning some functional inequalities.

Let us begin by recalling the swarm algorithm presented in [4]. We start by up-lifting through integrations the function $U$ on $V$ to the functional $\mathcal{U}$ defined on the set $\mathcal{P}(V)$ of probability measures on $V$ via

$$\forall \rho \in \mathcal{P}(V), \quad \mathcal{U}(\rho) \coloneqq \int_V U(x)\rho(dx).$$

Next, we penalise this functional by a $\varphi$-entropy term. Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ ($\mathbb{R}_+ \coloneqq [0, +\infty)$) be a convex function satisfying $\varphi(1) = \varphi'(1) = 0$ and consider the functional

$$\mathcal{H} : \mathcal{P}(V) \ni \rho \mapsto \begin{cases} \int \varphi(\rho(x))\ell(dx) & , \quad \text{when} \quad \rho \ll \ell \\ +\infty & , \quad \text{otherwise} \end{cases} \tag{1.1}$$

where $\ell$ is the Riemannian probability measure on $V$ and where we denoted in the same way a probability measure and its Radon-Nikodym density with respect to the reference measure $\ell$.

For any $\beta \geq 0$, seen as an inverse temperature, consider the functional

$$\mathcal{U}_\beta \coloneqq \beta\mathcal{U} + \mathcal{H}. \tag{1.2}$$

When $\beta$ is large, the global optimisation of $\mathcal{U}_\beta/\beta$ on $\mathcal{P}(V)$ is to some degree equivalent to the global optimisation of $U$ on V. We endow $\mathcal{P}(V)$, or rather its subset $\mathcal{D}_+(V)$ consisting of probability measures admitting a positive and smooth density, with the Wasserstein structure. Then under the additional assumption that $\varphi'(0) = -\infty$ and starting from an initial probability measure $\rho_0 \in \mathcal{D}_+(V)$, we can consider the gradient descent associated to $\mathcal{U}_\beta$ to come close to the unique stationary probability measure, which is almost concentrated on $\mathcal{M}(U)$, the set of global minima of $U$. To really concentrate on $\mathcal{M}(U)$, we have to let $\beta$ depends on time in some way, with in particular $\lim_{t\to+\infty} \beta_t = +\infty$. The resulting evolution $(\rho_t)_{t\geq 0}$ has in the weak sense the non-linear Markov representation

$$\forall t \geq 0, \quad \dot{\rho}_t = \rho_t L_{\beta_t,\rho_t}, \tag{1.3}$$

where for any $\beta \geq 0$ and $\rho \in \mathcal{D}_+(V)$, $L_{\beta,\rho}$ is the diffusion generator on $V$ defined by

$$L_{\beta,\rho}[\cdot] = \alpha(\rho)\Delta \cdot -\beta\langle \nabla, \nabla \cdot\rangle,$$

where $\Delta, \langle\cdot,\cdot\rangle$ and $\nabla$ are the Laplace-Beltrami operator, the Riemannian scalar product and the gradient operator, and with

$$\forall r > 0, \quad \alpha(r) \coloneqq \frac{1}{r}\int_0^r s\varphi''(s)ds,$$

assuming that $\varphi$ is $\mathcal{C}^2$ on $(0, +\infty)$.

In [4], convex functions $\varphi$ of the following forms were considered. For any $m \in \mathbb{R} \setminus \{0, 1\}$ define $\varphi_m$ by

$$\forall r \geq 0, \quad \varphi_m(r) \coloneqq \frac{r^m - 1 - m(r-1)}{m(m-1)}.$$

Observe that $\varphi_m(0) = +\infty$ for $m < 0$ and this situation was not taken into account in [4].

The function $\varphi_0$ and $\varphi_1$ are obtained as limits (respectively for $m \to 0$ and $m \to 1$) and are given by

$$\forall r \geq 0, \quad \begin{cases} \varphi_0(r) \coloneqq -ln(r) + r - 1, \\ \varphi_1(r) \coloneqq r\ln(r) - r + 1, \end{cases} \tag{1.4}$$

(in particular $\varphi_m(0) = +\infty$ iff $m \leq 0$).

We deduce a family of convex functions parametrized by $m_1, m_2 \in \mathbb{R}$ (respectively controlling the behavior at 0 and $+\infty$) via

$$\forall r \geq 0, \quad \varphi_{m_1,m_2} = \begin{cases} \varphi_{m_1}(r) & \text{if} \quad r \in (0, 1], \\ \varphi_{m_2}(r) & \text{if} \quad r \in (1, +\infty), \end{cases}$$

and note that these functions $\varphi_{m_1,m_2}$ are $\mathcal{C}^2$ on $(0, +\infty)$.

It was proven in [4] that if $V$ is the circle, if $\varphi \coloneqq \varphi_{m,2}$ with $m \in (0, 1/2)$ and if the inverse temperature schedule is given by

$$\forall t \geq 0, \quad \beta_t \coloneqq kt^{1/\gamma}, \text{ with } k > 0 \text{ and } \gamma = \frac{3(2-m)}{1-2m} \in [6, +\infty),$$

then the solution of (1.3) concentrates around the set of global minima of $U$ for large times. We expect that a variant of this result holds for any compact connected manifold $V$, but we restricted to the case of the circle to get the underlying functional inequality.

As mentioned previously, here one of our goals is to transpose the above considerations to the situation of a finite set $V$, in particular to get around the difficulty of the underlying functional inequality. As illustrated by the two papers of Holley and Stroock [7] and Holley, Kusuoka and Stroock [8], such inequalities can be easier to obtain in the finite context than in the continuous one.

Let us describe how the previous objects have to be modified. The compact and connected Riemannian manifold is replaced by a finite set $V$ endowed with a Markov generator $L := (L(x,y))_{x,t\in V}$ plays the role of the Beltrami-Laplacian $\Delta$ (which encapsulates the whole Riemannian structure), so we assume that it is irreducible and reversible with a probability distribution still denoted $\ell := (\ell(x))_{x\in V}$ (which necessarily gives a positive weight to all points of $V$). Let $\mathcal{P}(V)$ be the set of probability measures on $V$. To any $\mu := (\mu(x))_{x\in V} \in \mathcal{P}(V)$, we associate its density $\rho$ with respect to $\ell$:

$$\forall x \in V, \quad \rho(x) := \frac{\mu(x)}{\ell(x)}. \tag{1.5}$$

The set of such densities is denoted by $\mathcal{D}(V)$, we will often move back and forth between $\mathcal{P}(V)$ and $\mathcal{D}(V)$, which somewhat respectively corresponds to probabilist and analyst points of view.

Similar to (1.1), the functional $\mathcal{H}$ is given by

$$\forall \mu \in \mathcal{P}(V), \quad \mathcal{H}(\mu) := \sum_{x\in V} \varphi(\rho(x))\ell(x),$$

where $\varphi$ is a convex function as above, except that we furthermore allow $\varphi(0) = +\infty$ (in this case we assume that $\lim_{r\to 0^+} \varphi(r) = +\infty$).

Given a mapping $U : V \to \mathbb{R}$, as above we can then extend it into the functionals $\mathcal{U}$ and $\mathcal{U}_\beta$, for any $\beta \geq 0$, defined on $\mathcal{P}(V)$.

To go further, we have to endow $\mathcal{P}(V)$ with a Riemannian structure (with boundary), an ersatz of the Wasserstein distance, to be able to consider gradient descent for $\mathcal{U}_\beta$. To do so, we follow Erbar and Maas [9]. Choosing a particular metric among those they propose, see the next section for details, and starting from a positive probability $\mu_0$, the gradient descent evolution $(\mu_t)_{t\geq 0}$ satisfies the equation

$$\forall t > 0, \quad \dot{\mu}_t = \mu_t L_{\beta_t,\rho_t}, \tag{1.6}$$

(recall that for $t \geq 0$, $\mu_t$ is the probability admitting $\rho_t$ as density with respect to $\ell$), with the mapping

$$\mathbb{R}_+ \times \mathcal{D}_+(V) \ni (\beta,\rho) \mapsto L_{\beta,\rho} := (L_{\beta,\rho}(x,y))_{x,y\in V} \in \mathcal{G}(V),$$

where $\mathcal{L}(V)$ is the set of Markov generators on $V$,

$$\mathcal{D}_+(V) := \{\rho \in \mathcal{D}(V) : \forall x \in V, \rho(x) > 0\},$$

and where

$$\forall x \neq y, \quad L_{\beta,\rho} := \left( \frac{\rho(y) - \rho(x)}{\rho(x)(\varphi'(\rho(y)) - \varphi'(\rho(x))} \beta(U(y) - U(x)) + \frac{\rho(y)}{\rho(x)} - 1 \right)_- L(x,y),$$

where $(x)_- = \max(0, -x)$ and with the convention that $\forall x, y \in V$ such that $\rho(y) = \rho(x)$,

$$\frac{\rho(y) - \rho(x)}{\varphi'(\rho(y)) - \varphi'(\rho(x))} = \frac{1}{\varphi''(\rho(x))}.$$

Since $L_{\beta,\rho}$ is a Markov generator, we don't need to specify it's diagonal entries, they are given by

$$\forall x \in V, \quad L_{\beta,\rho}(x,x) = -\sum_{y\in V\setminus\{x\}} L_{\beta,\rho}(x,y).$$

Inspired by (1.3), we then consider time-inhomogeneous inverse temperature schemes $(\beta_t)_{t\geq 0}$ and the associated evolution equations

$$\forall t > 0, \quad \dot{\mu}_t = \mu_t L_{\beta_t,\rho_t}, \quad \mu_0 \in \mathcal{P}_+(V), \tag{1.7}$$

where $\mathcal{P}_+(V) := \{\mu \in \mathcal{P}(V) : \rho \in \mathcal{D}_+(V)\}$.

The main result of this paper is then:

**Theorem 1.1.** *For any $m < 0$, consider the function $\varphi = \varphi_{m,2}$ as well as the time-inhomogeneous inverse temperature scheme*

$$\forall t \geq 0, \quad \beta_t = (t_0 + t)^{\kappa(m)} - 1,$$

*where $t_0 \geq 1$ and*

$$\kappa(m) = \frac{-m}{2(1-m)} \in (0, \frac{1}{2}). \tag{1.8}$$

*For the corresponding* (1.7)*, we have*

$$\lim_{t \to +\infty} \mu_t[\mathcal{M}(U)] = 1, \tag{1.9}$$

*where $\mathcal{M}(U)$ is the set of global minimizers of $U$.*

This is a discrete analogue to the corresponding result of [4], with the improvement that there is no more restriction on the "geometry" of the energy landscape $(L, \ell, U)$.

In practice it is often difficult to compute the evolution (1.7), so one traditionally resorts to interacting particle approximations. An numerical illustration is given at the end of the paper.

The paper is constructed according to the following plan. In the following section, we recall the metric constructions of Erbar and Maas on $\mathcal{P}_+(V)$ and our particular choice. In Section 3, we present the details of the adaptation to the finite setting of the program described at the beginning of this introduction, in particular we extend the penalized cost (1.2) for a specific family of convex functions. In Section 4 we consider the convergence to the equilibrium of the time-homogeneous and non-linear Markov evolution (1.6). Since it is a representation of the gradient descent with respect to $\mathcal{U}_\beta$, we use this functional as a Lyapunov function and are led to a functional inequality which is investigated in Section 5. The proof of Theorem 1.1 is given in Section 5 by adapting the same approach. The last section contains the numerical illustration. A first appendix explains why the traditional Metropolis algorithm is not included into our framework based on Maas' formalism [9] and how to extend it. The second appendix recalls some facts relative to linear and non-linear Markov samplings.


### Acknowledgments:

We would particularly like to thank Stéphane Villeneuve for the discussions we had about this paper.


# 2 Riemannian structures on $\mathcal{D}_+(V)$

In this section, we revisit certain Riemannian structures on $\mathcal{D}_+(V)$ and the concept of gradient flow for a smooth functional on $(\mathcal{D}_+(V), \mathcal{W}_\theta)$ as introduced in Maas [9], where $\mathcal{W}_\theta$ denotes a a Riemannian metric. Throughout the paper, we endow the set $V$ with an irreducible Markov generator $L = (L(x,y))_{x,y \in V}$, i.e.,

$$\forall x \neq y, \quad L(x,y) \geq 0 \quad \text{and} \quad \forall x \in V, \quad \sum_{y \in V} L(x,y) = 0.$$

Irreducibility means that for any $x \neq y \in V$, there is a path $x = x_0, x_1, ..., x_n = y$ such that $L(x_i, x_{i+1}) > 0$ for all $i = 0, 1, ..., n-1$. It is well-known that such a generator possesses a unique positive invariant measure $\ell = (\ell(x))_{x \in V}$ by Perron-Frobenius theorem and we assume further that $\ell$ is reversible for $L$, i.e.,

$$\ell(x)L(x,y) = \ell(y)L(y,x) \quad \forall x, y \in V.$$

## 2.1 Geometric notions

**Definition 2.1** (Discrete gradient and divergence)
*For any function $\psi \in \mathbb{R}^V$, the discrete gradient of $\psi$, denoted as $\nabla\psi$, is defined by*

$$\nabla\psi : V \times V \to \mathbb{R}, \qquad \nabla\psi(x,y) := \psi(y) - \psi(x). \tag{2.1}$$

*For any function $\Psi \in \mathbb{R}^{V \times V}$, the discrete divergence of $\Psi$, denoted as $\operatorname{div}\Psi$, is defined by*

$$\operatorname{div}\Psi : S \to \mathbb{R}, \qquad \operatorname{div}\Psi(x) := \frac{1}{2}\sum_{y \in V} L(x,y)(\Psi(x,y) - \Psi(y,x)). \tag{2.2}$$

**Definition 2.2** (Inner products)
*For $\phi, \psi \in \mathbb{R}^V$, the inner product with respect to $\ell$ is defined by*

$$\langle \phi, \psi \rangle_{\mathbb{L}^2(\ell)} := \sum_{x \in V} \ell(x)\phi(x)\psi(x). \tag{2.3}$$

*For $\Phi, \Psi \in \mathbb{R}^{V \times V}$, the inner product with respect to $\ell \ltimes L$ is defined by*

$$\langle \Phi, \Psi \rangle_{\ell \ltimes L} := \frac{1}{2} \sum_{\substack{x,y \in V \\ x \neq y}} \ell(x)L(x,y)\Phi(x,y)\Psi(x,y). \tag{2.4}$$

From the definitions provided above, it can be readily verified that the "integration by parts" formula holds.

$$\langle \nabla \psi, \Phi \rangle_{\ell \ltimes L} = -\langle \psi, \operatorname{div} \Phi \rangle_{\mathbb{L}^2(\ell)}. \tag{2.5}$$

Another crucial notion is the definition of tangent spaces over $\rho \in \mathcal{D}_+(V)$, which serves as a fundamental component for the Riemannian structures on on $\mathcal{D}_+(V)$.

**Definition 2.3** (Tangent space and inner product)
*Let $\rho \in \mathcal{D}_+(V)$, the tangent space over $\rho$ is defined by*

$$T_\rho \coloneqq \{\nabla \psi \in \mathbb{R}^{S \times S} : \psi \in \mathbb{R}^S\}. \tag{2.6}$$

*Note that $T_\rho$ does not depend on $\rho$, but we endow it with a inner product that does:*

$$\forall \nabla \phi, \nabla \psi \in T_\rho, \qquad \langle \nabla \phi, \nabla \psi \rangle_\rho \coloneqq \frac{1}{2} \sum_{x,y \in V} \nabla \phi(x,y) \nabla \psi(x,y) L(x,y) \theta(\rho(x), \rho(y)) \ell(x), \tag{2.7}$$

*where $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a suitable nonnegative function of two variables, chosen carefully in Maas [9] to make $T_\rho$ a Hilbert space. We will give more details on the function $\theta$ in the next section.*

## 2.2 Maas' metric

In [9], Maas introduced a notion of Wasserstein-like metric on $\mathcal{P}(V)$, for which he closely followed Brenier-Benamou's interpretation of the 2-Wasserstein metric on $\mathcal{P}_2(\mathbb{R}^n)$ in Benamou and Brenier [2], the space of probability measures on $\mathbb{R}^n$ with finite second moment. Initially, Maas used a Markov kernel to define the metric, but later in Erbar and Maas [6], they replaced the Markov kernel with a Markov generator, which is the one presented here. The key aspect is that Maas employed a function of two variables $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ satisfying the following collection of assumptions.

**Assumption 2.1.** *The function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ satisfies*

- *(A1): $\theta$ is continuous and is symmetric on $\mathbb{R}_+ \times \mathbb{R}_+$, i.e., $\theta(s,t) = \theta(t,s),\ \forall s,t \geq 0$.*
- *(A2): $\theta$ is $C^\infty$ on $(0,+\infty) \times (0,+\infty)$.*
- *(A3): $\theta(s,t) > 0,\ \forall s,t > 0$, and vanishes at the boundary: $\theta(0,t) = 0, \quad \forall t \geq 0$.*
- *(A4): $\theta(r,t) \leq \theta(s,t)$, for all $0 \leq r \leq s$ and $t \geq 0$.*
- *(A5): For any $T > 0$, there exists a constant $C_T > 0$ such that $\theta(2s,2t) \leq 2C_T\theta(s,t)$, whenever $s,t \leq T$.*

**Definition 2.4** (Maas metric)
*Let $\theta$ be a function as described in Assumption 2.1. For $\bar{\rho}_0, \bar{\rho}_1 \in \mathcal{D}(V)$, we set*

$$\mathcal{W}_\theta^2(\bar{\rho}_0, \bar{\rho}_1) \coloneqq \inf_{\rho,\psi} \left\{ \frac{1}{2} \int_0^1 \sum_{x,y \in V} \ell(x) L(x,y) \theta(\rho_t(x), \rho_t(y)) (\nabla \psi_t(x,y))^2 \, dt \right\} = \inf_{\rho,\psi} \left\{ \int_0^1 \|\nabla \psi_t\|_{\rho_t}^2 dt \right\}, \tag{2.8}$$

*where the infimum runs over all pairs $(\rho, \psi)$ such that $\rho : [0,1] \to \mathcal{D}(V)$ is a piecewise $C^1$ curve in $\mathcal{D}(V)$ and $\psi : [0,1] \to \mathbb{R}^S$ is a measurable function, the pair satisfies, for a.e. $t \in [0,1]$,*

$$\begin{cases} \dot{\rho}_t(x) + \sum_{y \in V} \nabla \psi_t(x,y) L(x,y) \theta(\rho_t(x), \rho_t(y)) = 0, \ \forall x \in V, \\ \rho_0 = \bar{\rho}_0, \ \rho_1 = \bar{\rho}_1. \end{cases} \tag{2.9}$$

We have the following summarized result by Maas, the proof of which can be found in the proofs of Theorems 3.12, 3.19, and Lemma 3.30 in Maas [9].

**Theorem 2.2.** *Suppose that*

$$C_\theta \coloneqq \int_0^1 \frac{1}{\sqrt{\theta(1-r, 1+r)}} dr < +\infty \tag{2.10}$$

*then $\mathcal{W}_\theta$ is a metric on $\mathcal{P}(V)$. Additionally, if $\theta$ is concave, then for any $\bar{\rho}_0, \bar{\rho}_1 \in \mathcal{D}_+(V)$, we can restrict the set in the infimum in (2.8) to curves $\rho = (\rho_t)_{t \in [0,1]} \subset \mathcal{D}_+(V)$. As a consequence, $(\mathcal{D}_+(V), \mathcal{W}_\theta)$ is a Riemannian manifold (i.e., $\mathcal{W}_\theta$ can be intepreted as a Riemannian distance).*

## 2.3 Gradient flows of functionals

**Definition 2.5** (Tangent vector field of a curve)
*Let $\rho = (\rho_t)_{t\geq 0} \subset \mathcal{D}_+(V)$ be a smooth curve. The tangent vector field along $\rho$ is denoted by $D_t\rho \in T_{\rho_t}$. At any time $t \geq 0$, $D_t\rho$ is the unique element $\nabla g_t$ of $T_{\rho_t}$ such that*

$$\dot{\rho}_t + \mathrm{div}(\widehat{\rho}_t \odot \nabla g_t) = 0, \tag{2.11}$$

*where $\widehat{\rho}_t(x,y) \coloneqq \theta(\rho_t(x), \rho_t(y))$ and the notation $\odot$ represents the entrywise product, i.e., if $H, K \in \mathbb{R}^{V\times V}$ then $H \odot K \coloneqq (H(x,y)K(x,y))_{x,y\in V}$.*

In view of Maas [9], we shall consider two special types of functionals:

- For a function $R : \mathcal{S} \to \mathbb{R}$ we consider the *potential energy functional* $\mathcal{V}_1 : \mathcal{D}_+(V) \to \mathbb{R}$ defined by

$$\mathcal{V}_1(\rho) \coloneqq \sum_{x\in V} R(x)\rho(x)\ell(x). \tag{2.12}$$

- For a differentiable function $f : (0, +\infty) \to \mathbb{R}$, we consider the *generalized entropy* $\mathcal{V}_2 : \mathcal{D}_+(V) \to \mathbb{R}$ defined by

$$\mathcal{V}_2(\rho) \coloneqq \sum_{x\in V} f(\rho(x))\ell(x). \tag{2.13}$$

**Definition 2.6** (Gradient of a smooth functional)
*The gradient of a smooth functional $\mathcal{V} : \mathcal{D}_+(V) \to \mathbb{R}$ at $\rho \in \mathcal{D}_+(V)$ with respect to the metric $\mathcal{W}_\theta$, denoted by grad $\mathcal{V}$, is the unique element of $T_\rho$ such that, for any smooth curve $(\rho_t)_{t\in(-\epsilon,\epsilon)} \subset \mathcal{D}_+(V)$ with $\rho_0 = \rho$,*

$$\frac{d}{dt}\mathcal{V}(\rho_t)\Big|_{t=0} = \langle \mathrm{grad}\ \mathcal{V}(\rho), D_t\rho|_{t=0}\rangle_{\rho\ltimes L}.$$

**Theorem 2.3.** *For the functionals $\mathcal{V}_1$ and $\mathcal{V}_2$ introduced in* (2.12), (2.13)*, their gradients at $\rho \in \mathcal{D}_+(V)$ are*

$$\mathrm{grad}\ \mathcal{V}_1(\rho) = \nabla R, \qquad \mathrm{grad}\ \mathcal{V}_2(\rho) = \nabla[f' \circ \rho].$$

*Proof.* See proofs of Propositions 4.1 and 4.2 in Maas [9]. □

**Definition 2.7** (Gradient flow)
*Given a metric $\mathcal{W}_\theta$, a smooth curve $\rho = (\rho_t)_{t\geq 0} \subset \mathcal{D}_+(V)$ is called a gradient flow of a functional $\mathcal{V}$ if*

$$D_t\rho = -\mathrm{grad}\ \mathcal{V}(\rho_t), \quad \forall t \geq 0.$$

In particular, given a functional of the form $\mathcal{V} \coloneqq \beta\mathcal{V}_1 + \mathcal{V}_2$, $\beta \geq 0$, Theorem 2.3 gives

$$\mathrm{grad}\ \mathcal{V}(\rho) = \beta\nabla R + \nabla[f' \circ \rho].$$

Such an example of the functional $\mathcal{V}$ is the penalized cost

$$\mathcal{U}_\beta(\rho) \coloneqq \beta \sum_{x\in V} U(x)\rho(x)\ell(x) + \sum_{x\in V} \varphi(\rho(x))\ell(x), \quad \beta \geq 0,$$

where $\varphi \in C^2(0, +\infty)$ is strictly convex. We will study in detail this functional together with its associated gradient flow in the next section.

# 3 Presentation of the problem

## 3.1 The choice of family of relaxations on $\mathcal{P}(V)$

Let $V$ be the finite set mentioned in the introduction and $U : V \to \mathbb{R}$ be a function on $V$. As previously mentioned in the introduction, our goal is to minimize the function $U$ over $V$. To achieve this, we first up-lift through integration the function $U$ on $V$ to the functional $\mathcal{U}$ defined on the set $\mathcal{P}(V)$ of probability measures on $V$ via

$$\forall \mu \in \mathcal{P}(V), \qquad \mathcal{U}(\mu) \coloneqq \sum_{x\in V} U(x)\mu(x).$$

Recall that in the previous section, we endowed $V$ with an irreducible Markov generator $L$. Its positive invariant probability measure $\ell$ allows us to identify each $\mu \in \mathcal{P}(V)$ with its density with respect to $\ell$:

$$\forall x \in V, \qquad \rho(x) \coloneqq \frac{\mu(x)}{\ell(x)}.$$

Therefore, we will often write $\mathcal{U}(\rho)$ instead of $\mathcal{U}(\mu)$:

$$\forall \rho \in \mathcal{D}(V), \qquad \mathcal{U}(\rho) \coloneqq \sum_{x\in V} U(x)\rho(x)\ell(x). \tag{3.1}$$

We turn to the choice of the $\mathcal{C}^2$, convex function $\varphi$ mentioned in the introduction that we are going to use throughout this paper. Let $m < 0$ be a negative real number, define the function $\varphi : (0, +\infty) \to \mathbb{R}_+$ as follows

$$\forall r > 0, \qquad \varphi(r) \coloneqq \varphi_{m,2}(r) = \begin{cases} \dfrac{r^m - 1 - m(r-1)}{m(m-1)} & , \quad r \in (0,1) \\ \dfrac{(r-1)^2}{2} & , \quad r \in [1, +\infty) \end{cases} \tag{3.2}$$

It can be easily verified that $\varphi$ is $C^2$ with its first derivative given by

$$\forall r > 0, \qquad \varphi'(r) = \begin{cases} \dfrac{r^{m-1} - 1}{m-1} & , \quad r \in (0,1) \\ r - 1 & , \quad r \in [1, +\infty) \end{cases}$$

and its second derivative is given by

$$\forall r > 0, \qquad \varphi''(r) = \begin{cases} r^{m-2} & , \quad r \in (0,1) \\ 1 & , \quad r \in [1, +\infty) \end{cases}$$

so that $\lim_{r\to 0^+} \varphi(r) = +\infty$, $\varphi(1) = \varphi'(1) = 0$. The second derivative $\varphi''$ is decreasing on $(0, +\infty)$ and $\forall r > 0,\ \varphi''(r) \geq 1$, implying that its first derivative $\varphi'$ is strictly increasing and concave. Consequently, $\varphi$ is strictly convex. Additionally, we have $\lim_{r\to 0^+} \varphi'(r) = -\infty$ and $\varphi'(0, +\infty) = \mathbb{R}$, thus $\varphi'$ has an inverse $(\varphi')^{-1} : \mathbb{R} \to (0, +\infty)$ which we denote by $g = (\varphi')^{-1}$. A standard result in real analysis shows that the function $g : \mathbb{R} \to (0, +\infty)$ is strictly positive and increasing, with the first derivative

$$g'(x) = \frac{1}{\varphi''(g(x))} \in (0, 1].$$

We can think of the function $\varphi$ defined in (3.2) as a family of convex functions indexed by $m < 0$. The negativity of $m$ forces $\lim_{r\to 0^+} \varphi(r) = +\infty$, which will be crucial in the proofs of existence, uniqueness and convergence theorems of the gradient flows associated with the penalized cost $\mathcal{U}_\beta$ (given in the next subsection) and its time-dependent version later. We would like to emphasize that from now on, whenever we write $\varphi$, we implicitly refer to the one defined in (3.2) unless explicitly stated otherwise.

## 3.2 Penalized cost functional and stationary measure

Using function $\varphi$ in (3.2), we define the a $\varphi$-entropy term $\mathcal{H}$ by

$$\forall \rho \in \mathcal{D}(V), \qquad \mathcal{H}(\rho) \coloneqq \sum_{x\in V} \varphi(\rho(x))\ell(x),$$

and use this term to penalize $\mathcal{U}(\rho)$. For $\beta \geq 0$, seen as an inverse temperature, consider the penalized-cost functional

$$\begin{aligned} \forall \rho \in \mathcal{D}(V), \qquad \mathcal{U}_\beta(\rho) &\coloneqq \beta\mathcal{U}(\rho) + \mathcal{H}(\rho) \\ &= \beta \sum_{x\in V} U(x)\rho(x)\ell(x) + \sum_{x\in V} \varphi(\rho(x))\ell(x). \end{aligned}$$

From the choice of $\varphi$ in (3.2), we have the following result.

**Theorem 3.1.** *Let $\beta \geq 0$, and $\varphi$ be given in* (3.2)*. The functional $\rho \mapsto \mathcal{U}_\beta(\rho)$ is strictly convex and admits a unique minimizer $\eta_\beta \in \mathcal{D}_+(V)$.*

*Proof.* The functional $\rho \mapsto \mathcal{U}_\beta(\rho)$ is strictly convex because it is a sum of a linear functional and a strictly convex functional. Hence, if such a minimizer $\eta_\beta$ exists, it is necessarily unique. To show the existence of $\eta_\beta \in \mathcal{D}(V)$, we define $\eta_\beta$ to be the (unique) solution of the equation

$$\rho \in \mathcal{D}_+(V), \quad \forall x \in V, \qquad \beta U(x) + \varphi'(\rho(x)) = c(\beta), \tag{3.3}$$

where $c(\beta)$ solves the equation

$$c \in \mathbb{R}, \qquad \sum_{x\in V} \ell(x) g(c - \beta U(x)) = 1, \tag{3.4}$$

with $g = (\varphi')^{-1}$, the inverse function of $\varphi'$. To show $\eta_\beta$ is well-defined, consider

$$f : \mathbb{R} \to (0, +\infty), \qquad c \mapsto f(c) = \sum_{x\in V} \ell(x) g(c - \beta U(x)).$$

The first derivative of $f$ is

$$f'(c) = \sum_{x\in V} \frac{\ell(x)}{\varphi''(g(c - \beta U(x)))} > 0, \qquad (\text{since } \varphi'' > 0)$$

and note that $\lim_{c\to-\infty} f(c) = 0$, $\lim_{c\to+\infty} f(c) = +\infty$, so (3.4) has a unique solution $c(\beta)$. If we let $\forall x \in V$, $\eta_\beta(x) := g(c(\beta) - \beta U(x))$ then $\eta_\beta > 0$ and satisfies $\sum_{x\in V} \ell(x)\eta_\beta(x) = 1$, thus $\eta_\beta \in \mathcal{D}_+(V)$. Also, $\eta_\beta$ satisfies (3.3), i.e., $\beta U(x) + \varphi'(\eta_\beta(x)) = c(\beta)$. Finally, to show $\eta_\beta$ is the unique minimizer, write

$$\begin{aligned}
\mathcal{U}_\beta(\rho) - \mathcal{U}_\beta(\eta_\beta) &= \sum_{x\in V} \beta U(x)(\rho(x) - \eta_\beta(x))\ell(x) + \sum_{x\in V} \Big(\varphi(\rho(x)) - \varphi(\eta_\beta(x))\Big)\ell(x) \\
&= \sum_{x\in V} (-\varphi'(\eta_\beta(x)) + c(\beta))(\rho(x) - \eta_\beta(x))\ell(x) + \sum_{x\in V} \Big(\varphi(\rho(x)) - \varphi(\eta_\beta(x))\Big)\ell(x) \\
&= -\sum_{x\in V} \varphi'(\eta_\beta(x))(\rho(x) - \eta_\beta(x))\ell(x) + \sum_{x\in V} \Big(\varphi(\rho(x)) - \varphi(\eta_\beta(x))\Big)\ell(x) \\
&= \sum_{x\in V} \ell(x)\Big(\varphi(\rho(x)) - \varphi(\eta_\beta(x)) - \varphi'(\eta_\beta(x))(\rho(x) - \eta_\beta(x))\Big) \\
&\geq 0,
\end{aligned}$$

where the last inequality follows from the convexity of $\varphi$. □

The next result shows the limiting behavior of $\eta_\beta$ as $\beta \to +\infty$. Recall that $\mathcal{M}(U)$ is the set of global minimizers of $U$.

**Theorem 3.2.** *Under the hypothesis of Theorem 3.1, the following statements hold*

*(i)* $\forall x, y \in V$ *such that* $U(x) = U(y)$, $\eta_\beta(x) = \eta_\beta(y)$ *(in particular* $x, y \in \mathcal{M}(U)$*).*

*(ii)* $\forall x \notin \mathcal{M}(U)$, $\lim_{\beta\to+\infty} \eta_\beta(x) = 0$. *More precisely, we have*

$$\forall x \notin \mathcal{M}(U), \qquad \lim_{\beta\to+\infty} \beta[\eta_\beta(x)]^{1-m} = \frac{1}{(1-m)(U(x) - \min U)},$$

*where* $m < 0$ *is fixed in* (3.2).

*(iii)* $\forall x \in \mathcal{M}(U)$, $\eta_\beta(x) \geq 1$ *and* $\dfrac{\partial \eta_\beta(x)}{\partial\beta} \geq 0$, *where the equality holds iff* $U$ *is constant on* $V$.

*(iv)* $\forall x \in \mathcal{M}(U)$, $\lim_{\beta\to+\infty} \eta_\beta(x) = \dfrac{1}{\sum_{y\in\mathcal{M}(U)} \ell(y)}$. *In other words, if* $\zeta_\beta$ *is the probability measure with density* $\eta_\beta$ *and* $V \ni x \mapsto 1_{\mathcal{M}(U)}(x)$ *is the function on* $V$ *which equals 1 if* $x \in \mathcal{M}(U)$ *and 0 otherwise, then* $\lim_{\beta\to+\infty} \zeta_\beta(x) = \zeta_\infty(x)$, *where*

$$\forall x \in V, \qquad \zeta_\infty(x) := \frac{\ell(x)}{\sum_{y\in\mathcal{M}(U)} \ell(y)} 1_{\mathcal{M}(U)}(x). \tag{3.5}$$

*Proof.* The first statement $(i)$ is a consequence of (3.3). For $(ii)$, we let $x \notin \mathcal{M}(U)$ and $x_0 \in \mathcal{M}(U)$ then from (3.3) and the fact that $\rho(x_0)\ell(x_0) \leq 1$,

$$\begin{aligned}
\varphi'(\eta_\beta(x)) &= -\beta(U(x) - \min U)) + \varphi'(\eta_\beta(x_0)) \\
&\leq -\beta(U(x) - \min U)) + \varphi'(\ell(x_0)^{-1}) \\
&\longrightarrow -\infty \qquad \text{as} \quad \beta \to +\infty,
\end{aligned}$$

which in turn gives $\eta_\beta(x) \to 0$ as $\beta \to +\infty$ and $\eta_\beta(x_0)$ is bounded away from 0 by a positive constant (in fact the lower bound is 1 by statement $(iii)$, which we will show later). So for $\beta$ big enough, such that $\eta_\beta(x) < 1$, we have

$$-\beta(U(x) - \min U)) + \varphi'(\eta_\beta(x_0)) = \varphi'(\eta_\beta(x)) = \frac{[\eta_\beta(x)]^{m-1} - 1}{m-1},$$

or

$$\begin{aligned}
\beta[\eta_\beta(x)]^{1-m} &= \Big((1-m)(U(x) - \min U) + \frac{(m-1)\varphi'(\eta_\beta(x_0)) + 1}{\beta}\Big)^{-1} \\
&\longrightarrow \frac{1}{(1-m)(U(x) - \min U)} \qquad \text{as} \quad \beta \to +\infty.
\end{aligned}$$

Let us prove $(iii)$. As in the proof of $(ii)$

$$\varphi'(\eta_\beta(x)) - \varphi'(\eta_\beta(x_0)) = -\beta(U(x) - \min U) < 0,$$

so $\eta_\beta(x) < \eta_\beta(x_0)$ by monotonicity of $\varphi'$. Since $x_0 \in \mathcal{M}(U)$ and $x \notin \mathcal{M}(U)$ are arbitrary, from $(i)$,

$$\begin{aligned}\eta_\beta(x_0) &= \eta_\beta(x_0) \sum_{y\in\mathcal{M}(U)} \ell(y) + \eta_\beta(x_0) \sum_{y\notin\mathcal{M}(U)} \ell(y) \\ &= \sum_{y\in\mathcal{M}(U)} \eta_\beta(y)\ell(y) + \eta_\beta(x_0) \sum_{y\notin\mathcal{M}(U)} \ell(y) \\ &= 1 + \sum_{y\notin\mathcal{M}(U)} (\eta_\beta(x_0) - \eta_\beta(y))\ell(y) \\ &\geq 1.\end{aligned}$$

Consider the function of two variables

$$\mathbb{R}_+ \times \mathbb{R} \ni (\beta, c) \mapsto Z(\beta, c) \coloneqq \sum_{x\in V} g(c - \beta U(x))\ell(x), \qquad g = (\varphi')^{-1}.$$

From the proof of Theorem 3.1, for each $\beta \geq 0$ there exists a constant $c(\beta)$ such that $Z(\beta, c(\beta)) = 1$. By the implicit function theorem,

$$\begin{aligned}\frac{\partial c}{\partial \beta} &= -\frac{\partial Z}{\partial \beta}\left(\frac{\partial Z}{\partial c}\right)^{-1} \\ &= \frac{\sum_{x\in V} g'(c - \beta U(x))U(x)\ell(x)}{\sum_{x\in V} g'(c - \beta U(x))\ell(x)} \\ &\geq \min U.\end{aligned}$$

The last inequality follows from the fact that $g' > 0$. Note that $\eta_\beta(x_0) = g(c(\beta) - \beta U(x_0))$, hence

$$\frac{\partial \eta_\beta(x_0)}{\partial \beta} = g'(c(\beta) - \beta U(x_0))\left(\frac{\partial c}{\partial \beta} - \min U\right) \geq 0,$$

where the last inequality becomes equality if and only if $U$ is constant on $V$. Finally, from $(i)$, $(ii)$ and $(iii)$ we deduce

$$\eta_\beta(x_0) = \frac{1 - \sum_{y\notin\mathcal{M}(U)} \eta_\beta(y)\ell(y)}{\sum_{y\in\mathcal{M}(U)} \ell(y)} \overset{\beta\to+\infty}{\longrightarrow} \frac{1}{\sum_{y\in\mathcal{M}(U)} \ell(y)},$$

which establishes $(iv)$ and finishes the proof. □

**Remark 3.3.** *(a) The limit measure* (3.5) *is independent of the choice of $m < 0$ in the definition of $\varphi = \varphi_{m,2}$, it only depends on the invariant measure $\ell = (\ell(x))_{x\in V}$.*

*(b) In the theory of simulated algorithms on finite state spaces, see e.g. Holley and Stroock [7], one also gets a convergence in law toward a measure such as* (3.5)*, where $\ell$ is the reversible measure of the underlying exploration kernel. It appears as the small temperature limit $1/\beta \to 0_+$ of the Gibbs distribution given by*

$$\forall\ x \in V, \qquad \zeta_\beta(x) \coloneqq \frac{\ell(x)e^{-\beta U(x)}}{\sum_{y\in V} \ell(y)e^{-\beta U(y)}}$$

*Note that the latter corresponds to our stationary measure when we take $\varphi = \varphi_1$ in* (1.4)*. But the time-inhomogeneous evolution of probability measures $(\rho_t(x)\,\ell(dx))_{t\geq 0}$ we will construct in Section 5 do not correspond to the evolution of the Simulated Annealing algorithm considered in Holley and Stroock [7]. This a departure with the continuous space situation of of [4], where we recover the Simulated Annealing algorithm of Holley, Kusuoka and Stroock [8] by taking $\varphi = \varphi_1$. For other classical references to the Simulated Annealing algorithms on finite states space, see [12, 1], where $\ell$ is rather taken to be the uniform distribution on $V$.*

# 4 The time-homogeneous situation

## 4.1 Gradient flow of $\mathcal{U}_\beta$

We define the function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ by

$$\theta(s,t) \coloneqq \begin{cases} \frac{s-t}{\varphi'(s)-\varphi'(t)}, & \forall s,t > 0,\ t \neq s \\ \frac{1}{\varphi''(s)}, & s = t > 0 \\ 0, & t = 0 \text{ or } s = 0. \end{cases} \tag{4.1}$$

It will be shown below that $\theta$ satisfies Assumption 2.1 except (A2) because $\theta$ is not differentiable (e.g., at (1,1)). Even without the smoothness of $\theta$, suppose that $C_\theta < \infty$ in (2.10) then $(\mathcal{D}(V), \mathcal{W}_\theta)$ remains a complete metric space by a slight modification of the proof of Theorem 2.2 in Maas [9] (to the best of our understanding, Maas did not rely on the smoothness of $\theta$ to prove $\mathcal{W}_\theta$ is a metric on $\mathcal{D}(V)$). However, $(\mathcal{D}_+(V), \mathcal{W}_\theta)$ ceases to be a Riemannian manifold due to the lack of smoothness in $\theta$. Since we will not investigate curvature notions in this paper, smoothness of the Riemannian metric $\langle\cdot,\cdot\rangle_\rho$ is not needed for our purposes. The continuity of $\theta$ will suffice for a notion of $C^1$ gradient flow dynamic of the penalized cost $\mathcal{U}_\beta$. We give some properties of the function $\theta$ defined in (4.1).

**Lemma 4.1.** *The function $\theta$ in* (4.1) *satisfies Assumption* (2.1) *except (A2). Moreover, we have the following inequality*

$$\forall s,t>0, \qquad (t-s)(\varphi'(t)-\varphi'(s)) \geq \varphi(t)-\varphi(s)-\varphi'(s)(t-s). \tag{4.2}$$

*Proof.* From the definition of $\theta$, (A1) and (A3) follow from the fact that $\varphi \in C^2$, $\lim_{r\to 0^+}\varphi'(r) = -\infty$ and $\lim_{r\to 0^+}\varphi''(r) = +\infty$. To prove (A4), we note that $\varphi'$ is concave because $\varphi''$ is positive and nonincreasing. For fixed $r \leq s$, consider the function $k$ given on $t \in [0,r]\cup[s,+\infty)$ by

$$\forall t \in [0,r]\cup[s,+\infty), \qquad k(t) := (s-t)(\varphi'(r)-\varphi'(t)) - (r-t)(\varphi'(s)-\varphi'(t)), \qquad .$$

Its first derivative is

$$k'(t) = \varphi'(s) - \varphi'(r) - \varphi''(t)(s-r).$$

If $t \in [0,r]$, then $\varphi''(t) \geq \varphi''(r)$ (recall that $\varphi''$ is decreasing) and

$$k'(t) \leq \varphi'(s)-\varphi'(r)-\varphi''(r)(s-r) \leq 0$$

by the concavity of $\varphi'$. Similarly, if $t \geq s$, then $\varphi''(s) \geq \varphi''(t)$ and

$$\begin{aligned} k'(t) &\geq \varphi'(s)-\varphi'(r)-\varphi''(s)(s-r) \\ &= -(\varphi'(r)-\varphi'(s)-\varphi''(s)(r-s)) \\ &\geq 0 \end{aligned}$$

again by the concavity of $\varphi'$. We obtain $t \mapsto k(t)$ is decreasing on $[0,r]$ and increasing on $[s,\infty)$. Hence, $k(t) \geq k(r) = 0$ on $[0,r]$ and $k(t) \geq k(s) = 0$ on $[s,+\infty)$. Thus, for $t \in [0,r)\cup(s,+\infty)$,

$$\theta(s,t)-\theta(r,t) = \frac{(s-t)(\varphi'(r)-\varphi'(t)) - (r-t)(\varphi'(s)-\varphi'(t))}{(\varphi'(s)-\varphi'(t))(\varphi'(r)-\varphi'(t))} \geq 0 \tag{4.3}$$

because the numerator is just $k(t) \geq 0$ and the denominator is always positive due to the fact that $\varphi'$ is strictly increasing on $(0,+\infty)$. If $t \in (r,s)$,

$$\theta(s,t)-\theta(t,t) = \frac{\varphi'(s)-\varphi'(t)-\varphi''(t)(s-t)}{(\varphi'(t)-\varphi'(s))\varphi''(t)} \geq 0$$

since the numerator and the denominator are both negative because $\varphi'$ is concave and strictly increasing. In the same manner, $\theta(t,t) \geq \theta(r,t)$ and hence $\theta(s,t) \geq \theta(r,t)$. Since $r,s,t$ are arbitrary, thus we have established (A4). For the last inequality (4.2), we rewrite it as

$$\varphi(s)-\varphi(t)-\varphi'(t)(s-t) \geq 0,$$

but this is obviously true from the convexity of $\varphi$. □

**Lemma 4.2.** *The function $\theta$ given in* (4.1) *is locally Lipschitz on* $(0,+\infty)\times(0,+\infty)$, *i.e.,* $\forall (s_0,t_0) \in (0,+\infty)\times(0,+\infty)$, *there exist $\delta > 0$ and $K = K(s_0,t_0,\delta)$ such that*

$$\forall (s',t'),(s'',t'') \in B((s_0,t_0),\delta), \qquad |\theta(s',t')-\theta(s'',t'')| \leq K\|(s',t')-(s'',t'')\|_\infty,$$

*where* $\|(s',t')-(s'',t'')\|_\infty := \max\{|s'-s''|,|t'-t''|\}$ *and* $B((s_0,t_0),\delta) := \{(s,t) : \|(s,t)-(s_0,t_0)\|_\infty < \delta\}$.

*Proof.* We first prove that the function $(0,+\infty) \ni r \mapsto \varphi''(r)$ is Lipschitz on any compact intervals. It is clear that $\varphi''$ is continuously differentiable on any compact interval that does not contains 1 and thus Lipschitz continuous there. Suppose $[a,b]$ is an interval that contains 1, and let $s,t \in [a,b]$ be such that $a \leq s \leq 1 \leq t \leq b$ then,

$$\begin{aligned} \left|\varphi''(s)-\varphi''(t)\right| = |s^{m-2}-1| &= \frac{1-s^{2-m}}{s^{2-m}} \\ &\leq \frac{2-m}{a^{2-m}}(1-s) \\ &\leq \frac{2-m}{a^{2-m}}(t-s), \end{aligned}$$

and so $\varphi''$ is Lipschitz continuous on compacts that contain 1 too. Now we fixed a point $(s_0, t_0) \in (0, +\infty) \times (0, +\infty)$ and let $\delta > 0$ be such that $\min\{s_0 - \delta, t_0 - \delta\} > 0$. Consider two points $(s', t'), (s'', t'') \in B((s_0, t_0), \delta)$. By symmetry of $\theta$, we assume without loss of generality $s_0 \le t_0$, $s' < t'$, $s'' < t''$ and $t' \le t''$. Note that $s', t', s'', t'' \in [s_0 - \delta, t_0 + \delta]$. If $t' \le s''$, then by the monotonicity in each variable of $\theta$ in Lemma 4.1, $\theta(s'', t'') \ge \theta(t', t'') \ge \theta(t', s')$. So

$$\begin{aligned}
0 \le \theta(s'', t'') - \theta(s', t') &\le \theta(t'', t'') - \theta(s', s') \\
&= \left| \frac{1}{\varphi''(t'')} - \frac{1}{\varphi''(s')} \right| \\
&\le |\varphi''(t'') - \varphi''(s')| \quad (\text{since } \varphi'' \ge 1) \\
&\le K(s_0, t_0, \delta)(t'' - s') \\
&= K(s_0, t_0, \delta)(t'' - t' + s'' - s' + t' - s'') \\
&\le 2K(s_0, t_0, \delta)\|(s', t') - (s'', t'')\|_\infty,
\end{aligned}$$

where $K(s_0, t_0, \delta)$ is the Lipschitz constant of $\varphi''$ on $[s_0 - \delta, t_0 + \delta]$. If $t' > s''$, let $x \in [s_0 - \delta, t_0 + \delta]$ and consider the function $[x, t_0 + \delta] \ni r \mapsto p_x(r) := \theta(r, x)$. It's derivative is given by

$$\forall r \in (x, t_0 + \delta), \qquad p'_x(r) = \frac{-(\varphi'(x) - \varphi'(r) - \varphi''(r)(x - r))}{(\varphi'(x) - \varphi'(r))^2} \ge 0.$$

We will prove that $p'_x$ is bounded by some constant depending on $s_0, t_0$ and $\delta$. Indeed, by the mean value theorem, there is $r^* \in (x, r)$ such that $\varphi'(x) - \varphi'(r) = (x - r)\varphi''(r^*)$, hence

$$\begin{aligned}
p'_x(r) &= \frac{(r - x)(\varphi''(r^*) - \varphi''(r))}{(r - x)^2 \varphi''(r^*)^2} \\
&\le \frac{\varphi''(r^*) - \varphi''(r)}{r - x} \qquad (\text{since } \varphi'' \ge 1) \\
&\le \frac{\varphi''(x) - \varphi''(r)}{r - x} \qquad (r^* > x,\ \varphi'' \text{ is decreasing}) \\
&\le K(s_0, t_0, \delta) \qquad (x, r \in [s_0 - \delta, t_0 + \delta]).
\end{aligned} \tag{4.4}$$

Similarly, we have the same estimate for the derivative of the function $[s_0 - \delta, y] \ni r \mapsto h_y(r) := \theta(y, r)$, i.e.,

$$h'_y(r) \le K(s_0, t_0, \delta) \tag{4.5}$$

Hence, we have

$$\begin{aligned}
|\theta(s', t') - \theta(s'', t'')| &\le |\theta(s', t') - \theta(s'', t')| + |\theta(s'', t') - \theta(s'', t'')| \\
&\le \left| \int_{s'}^{s''} h'_{t'}(r) dr \right| + \left| \int_{t'}^{t''} p'_{s''}(r) dr \right| \quad (\text{set } x = s'', y = t' \text{ in } (4.4), (4.5)) \\
&\le K(s_0, t_0, \delta)(|s' - s''| + |t' - t''|) \\
&\le 2K(s_0, t_0, \delta)\|(s', t') - (s'', t'')\|_\infty.
\end{aligned}$$

which finishes the proof. □

Following Definition 2.7, but restricting to $C^1$ curves, we define the gradient flow of $\mathcal{U}_\beta(\rho)$ is any $C^1$ curve $(\rho_t)_{t\ge 0} \subset \mathcal{D}_+(V)$ satisfying

$$\forall t > 0, \qquad D_t \rho = -\text{grad}\, \mathcal{U}_\beta(\rho) = -(\beta \nabla U + \nabla[\varphi' \circ \rho]), \tag{4.6}$$

where the second equation follows from Theorem 2.3. We can rewrite (4.6) in terms of each coordinate as

$$\forall x \in V,\ \forall t > 0, \qquad \dot\rho_t(x) = \sum_{y \in V} L(x, y)\theta(\rho_t(x), \rho_t(y))\Big(\beta \nabla U(x, y) + \nabla[\varphi' \circ \rho_t](x, y)\Big), \tag{4.7}$$

with the notation $\nabla[\varphi' \circ \rho_t](x, y) = \varphi'(\rho_t(y)) - \varphi'(\rho_t(x))$. We shall often write the last quantity in (4.7) compactly as $\nabla[\beta U + \varphi' \circ \rho](x, y)$. The existence and uniqueness of (4.7) is guaranteed by the following proposition.

**Proposition 4.3.** *For any initial condition $\rho_0 \in \mathcal{D}_+(V)$, there is a unique solution $(\rho_t)_{t\ge 0} \subset \mathcal{D}_+(V)$, which exists for all time $t \ge 0$ and satisfies* (4.7).

*Proof.* For notational convenience, we define the functional

$$\forall \rho \in \mathcal{D}_+(V), \quad \forall x \in V, \qquad \mathcal{F}(\rho)(x) := \sum_{y \in V} L(x, y)\theta(\rho(x), \rho(y))\Big(\nabla[\beta U + \varphi' \circ \rho](x, y)\Big),$$

then the map $\rho \mapsto \mathcal{F}(\rho)$ is clearly continuous and is also locally Lipschitz. Indeed, fix $\rho^* \in \mathcal{D}_+(V)$, and consider $\delta > 0$ such that $B(\rho^*, \delta) := \{\rho \in \mathcal{D}(V) : \|\rho - \rho^*\|_\infty := \max_{x\in V} |\rho(x) - \rho^*(x)| < \delta\}$ lies in $\mathcal{D}_+(V)$. Let $\rho_1, \rho_2 \in B(\rho^*, \delta)$, and fix one $x \in V$, denote $\|L\|_\infty := \max_{y\in V} |L(y, y)|$, osc $U := \max U - \min U$. Due to the particular choice of $\theta$, for $\rho \in \mathcal{D}_+(V)$,

$$\sum_{y\in V} L(x, y)\theta(\rho(x), \rho(y))\Big(\nabla[\varphi' \circ \rho](x, y)\Big) = \sum_{y\in V} L(x, y)(\rho(y) - \rho(x)) = \sum_{y\in V} L(x, y)\rho(y),$$

hence

$$\begin{aligned}\frac{1}{\|L\|_\infty}|\mathcal{F}(\rho_1)(x) - \mathcal{F}(\rho_2)(x)| &\leq \beta \text{osc } U \sum_{y\in V} \Big|\theta(\rho_1(x), \rho_1(y)) - \theta(\rho_2(x), \rho_2(y))\Big| + \sum_{y\in V} |\rho_1(y) - \rho_2(y)| \\ &\leq K\beta \text{osc } U \sum_{y\in V} \max\{|\rho_1(x) - \rho_2(x)|, |\rho_1(y) - \rho_2(y)|\} + \sum_{y\in V} |\rho_1(y) - \rho_2(y)| \\ &\leq (K\beta \text{osc } U + 1)|V|\|\rho_1 - \rho_2\|_\infty,\end{aligned}$$

where $|V|$ is the cardinality of $V$, $K$ is the local Lipschitz constant of $\theta$ in Lemma 4.2. Therefore, there is a constant $K'$ such that

$$\forall \rho_1, \rho_2 \in B(\rho^*, \delta), \qquad |\mathcal{F}(\rho_1) - \mathcal{F}(\rho_2)|_\infty \leq K'\|\rho_1 - \rho_2\|_\infty.$$

By Cauchy-Picard and extensibility theorems, we have the existence and uniqueness of a solution $(\rho_t)_{t\in[0,T)}$ on some maximal interval $[0, T)$. By differentiating in time the function $t \mapsto \mathcal{U}_\beta(\rho_t)$,

$$\begin{aligned}\frac{d}{dt}\mathcal{U}_\beta(\rho_t) &= \sum_{x\in V} \Big(\beta U(x) + \varphi'(\rho_t(x))\Big)\dot{\rho}_t(x)\ell(x) \\ &= \sum_{x\in V} \Big(\beta U(x) + \varphi'(\rho_t(x))\Big)\ell(x) \sum_{y\in V} L(x, y)\theta(\rho_t(x), \rho_t(y))\Big(\nabla[\beta U + \varphi' \circ \rho_t](x, y)\Big) \\ &= \sum_{x\in V}\sum_{y\in V} \Big(\beta U(x) + \varphi'(\rho_t(x))\Big)\ell(x)L(x, y)\theta(\rho_t(x), \rho_t(y))\Big(\nabla[\beta U + \varphi' \circ \rho_t](x, y)\Big) \\ &= \sum_{y\in V}\sum_{x\in V} \Big(\beta U(y) + \varphi'(\rho_t(y))\Big)\ell(y)L(y, x)\theta(\rho_t(y), \rho_t(x))\Big(\nabla[\beta U + \varphi' \circ \rho_t](y, x)\Big) \\ &= \sum_{y\in V}\sum_{x\in V} \Big(\beta U(y) + \varphi'(\rho_t(y))\Big)\ell(x)L(x, y)\theta(\rho_t(x), \rho_t(y))\Big(\nabla[\beta U + \varphi' \circ \rho_t](y, x)\Big) \\ &= \frac{1}{2}\sum_{y\in V}\sum_{x\in V} \Big(\nabla[\beta U + \varphi' \circ \rho_t](x, y))\Big)\ell(x)L(x, y)\theta(\rho_t(x), \rho_t(y))\Big(\nabla[\beta U + \varphi' \circ \rho_t](y, x))\Big) \\ &= -\frac{1}{2}\sum_{y\in V}\sum_{x\in V} \ell(x)L(x, y)\theta(\rho_t(x), \rho_t(y))\Big(\nabla[\beta U + \varphi' \circ \rho_t](x, y)\Big)^2 \qquad (4.8) \\ &\leq 0,\end{aligned}$$

so $\mathcal{U}_\beta$ is decreasing along $(\rho_t)_{t\in[0,T)}$. If $T < +\infty$ then $\rho_t$ must go to the boundary of $\mathcal{D}_+(V)$, as $t \to T^-$. However, this is not possible since $\rho \mapsto \mathcal{U}_\beta(\rho)$ explodes at the boundary by the fact that $\lim_{r\to 0+} \varphi(r) = +\infty$. Therefore, $T = +\infty$ and the solution exists for all time $t \geq 0$. ☐

## 4.2 A functional inequality for $\beta \geq 0$ fixed

**Proposition 4.4.** *Let $\beta \geq 0$, we consider the functionals*

$$\begin{aligned}\mathcal{I}(\beta, \rho) &:= \mathcal{U}_\beta(\rho) - \mathcal{U}_\beta(\eta_\beta) \\ &= \sum_{x\in V} \ell(x)\Big(\varphi(\rho(x)) - \varphi(\eta_\beta(x)) - \varphi'(\eta_\beta(x))(\rho(x) - \eta_\beta(x))\Big), \qquad (4.9)\end{aligned}$$

$$\mathcal{G}(\beta, \rho) := \frac{1}{2}\sum_{x\in V}\sum_{y\in V} \ell(x)L(x, y)\theta(\rho(x), \rho(y))\Big(\beta\nabla U(x, y) + \nabla[\varphi' \circ \rho](x, y)\Big)^2. \qquad (4.10)$$

*Then*

$$\chi(\beta) := \inf_{\rho\in\mathcal{D}_+(V)\setminus\{\eta_\beta\}} \frac{\mathcal{G}(\beta, \rho)}{\mathcal{I}(\beta, \rho)} > 0.$$

Before giving the proof, we need the following lemmas.

**Lemma 4.5.** *Fix $\beta \geq 0$ For any sequence $(\rho_n)_{n\geq 1} \subset \mathcal{D}_+(V)$ such that $\|\rho_n - \eta_\beta\|_\infty \to 0$, we have*

$$\liminf_{n\to+\infty} \frac{\mathcal{G}(\beta, \rho_n)}{\mathcal{I}(\beta, \rho_n)} > 0.$$

*Proof.* Assume the statement in Lemma 4.5 does not hold, then there exists a subsequence, still denoted by $(\rho_n)_{n\geq 1}$ that

$$\lim_{n\to+\infty}\frac{\mathcal{G}(\beta,\rho_n)}{\mathcal{I}(\beta,\rho_n)}=0.$$

Denote $a_n=\|\rho_n-\eta_\beta\|_\infty$, and $h_n=\frac{\rho_n-\eta_\beta}{a_n}$, so $\|h_n\|_\infty=1$. Recall from (3.3) that $\varphi'(\eta_\beta)+\beta U$ is constant, which gives $\beta\nabla U(x,y)=-\nabla[\varphi'\circ\eta_\beta](x,y)$. We have,

$$\frac{\mathcal{G}(\beta,\rho_n)}{a_n^2}=\frac{1}{2a_n^2}\sum_{x,y\in V}\ell(x)L(x,y)\theta(\rho_n(x),\rho_n(y))\Big(\varphi'(\rho_n(x))-\varphi'(\eta_\beta(x))-[\varphi'(\rho_n(y))-\varphi'(\eta_\beta(y))]\Big)^2$$

and by mean value theorem, for some vector $\lambda_n\in[0,1]^V$, we have

$$\begin{aligned}&\frac{1}{a_n^2}\Big(\varphi'(\rho_n(x))-\varphi'(\eta_\beta(x))-(\varphi'(\rho_n(y))-\varphi'(\eta_\beta(y)))\Big)^2\\&=\frac{1}{a_n^2}\Big(\varphi'(\eta_\beta(x)+a_nh_n(x))-\varphi'(\eta_\beta(x))-(\varphi'(\eta_\beta(y)+a_nh_n(y))-\varphi'(\eta_\beta(y)))\Big)^2\\&=\Big(\varphi''\Big(\eta_\beta(x)+a_n\lambda_n(x)h_n(x)\Big)h_n(x)-\varphi''\Big(\eta_\beta(y)+a_n\lambda_n(y)h_n(y)\Big)h_n(y)\Big)^2.\end{aligned}$$

Also, by Taylor's theorem for second derivative (see e.g. Wolfe [13]), for some vector $\lambda_n'\in[0,1]^V$, we have

$$\mathcal{I}(\beta,\rho_n)=\frac{1}{2}\sum_{x\in V}\ell(x)\varphi''\Big(\eta_\beta(x)+a_n\lambda_n'(x)h_n(x)\Big)a_n^2h_n^2(x),\tag{4.11}$$

or

$$\frac{\mathcal{I}(\beta,\rho_n)}{a_n^2}=\frac{1}{2}\sum_{x\in V}\ell(x)\varphi''\Big(\eta_\beta(x)+a_n\lambda_n'(x)h_n(x)\Big)h_n^2(x).$$

Observe that the set $\mathcal{H}_{0,1}:=\{h:\ell[h]=0,\|h\|_\infty=1\}$ is compact so there is a subsequence $(h_{n_k})_{k\geq1}$ converging to $h_0\in\mathcal{H}_{0,1}$. Furthermore, because $a_{n_k}\to0$, we have

$$\lim_{k\to+\infty}\frac{\mathcal{I}(\beta,\rho_{n_k})}{a_{n_k}^2}=\frac{1}{2}\sum_{x\in V}\ell(x)\varphi''(\eta_\beta(x))h_0^2(x)>0,\tag{4.12}$$

$$\lim_{k\to+\infty}\frac{\mathcal{G}(\beta,\rho_{n_k})}{a_{n_k}^2}=\frac{1}{2}\sum_{x,y}\ell(x)L(x,y)\theta(\eta_\beta(x),\eta_\beta(y))\Big(\varphi''(\eta_\beta(x))h_0(x)-\varphi''(\eta_\beta(y))h_0(y)\Big)^2.\tag{4.13}$$

Since $\lim_{n\to+\infty}\frac{\mathcal{G}(\beta,\rho_n)}{\mathcal{I}(\beta,\rho_n)}=0$, the subsequence $\left(\frac{G(\beta,\rho_{n_k})}{\mathcal{I}(\beta,\rho_{n_k})}\right)_{k\geq1}$ converges to 0 and hence

$$\sum_{x,y\in V}\ell(x)L(x,y)\theta(\eta_\beta(x),\eta_\beta(y))\Big(\varphi''(\eta_\beta(x))h_0(x)-\varphi''(\eta_\beta(y))h_0(y)\Big)^2=0.$$

This, together with irreducibility of $L$ and $\eta_\beta\in\mathcal{D}_+(V)$ imply $\forall x\in V,\ \varphi''(\eta_\beta(x))h_0(x)=C$ for some constant $C$. If $C=0$, then $h_0=0\notin\mathcal{H}_{0,1}$ (because $\varphi''>0$), which is impossible. If $C\neq0$ then $\ell[h_0]\neq0$ since the sign of $h$ is equal to that of $C$, a contradiction. □

**Lemma 4.6.** *Denote*

$$\partial\mathcal{D}_+(V):=\mathcal{D}(V)\setminus\mathcal{D}_+(V)=\{\rho\in\mathcal{D}(V):\rho(x)=0\ \textit{for some}\ x\in V\}.$$

*Let $\rho_*\in\partial\mathcal{D}_+(V)$ and $(\rho_n)_{n\geq0}\subset\mathcal{D}_+(V)$ be a sequence such that $\rho_n\to\rho_*$, i.e. $\|\rho_n-\rho\|_\infty\to0$, then*

$$\lim_{n\to0}\frac{\mathcal{G}(\beta,\rho_n)}{\mathcal{I}(\beta,\rho_n)}=+\infty.$$

*Proof.* For any $f\in\mathbb{R}^V$ we will use the notations $f_\wedge:=\min_{x\in V}f(x)$, $f_\vee:=\max_{x\in V}f(x)$, $\{\rho_*=0\}:=\{x\in V:\rho_*(x)=0\}$ and we call a state $x\in V$ a minimizer of $\rho\in\mathcal{D}(V)$ if $\rho(x)=\rho_\wedge$. Since the set $\{\rho_*=0\}$ is finite and $\rho_n\to\rho_*$, there is $N>0$ such that for all $n\geq N$, the minimizers of $\rho_n$ must lie in $\{\rho_*=0\}$. Up to taking a subsequence, we can assume that for all $n\geq0$, $\rho_n$ admits some $x_0\in\{\rho_*=0\}$ as a minimizer, i.e., $\rho_n(x_0)=\rho_{n,\wedge}$ .

For every $\rho\in\mathcal{D}_+(V)$ set $M_\rho:=\{x\in V:\rho(x)\leq1\}$. We have

$$\begin{aligned}\forall\rho\in\mathcal{D}_+(V),\quad\mathcal{I}(\beta,\rho)&=\sum_{x\in V}\ell(x)\rho(x)\beta U(x)+\sum_{x\in V}\varphi(\rho(x))\ell(x)-\mathcal{U}_\beta(\eta_\beta)\\&\leq\beta\max U-\mathcal{U}_\beta(\eta_\beta)+\ell_\vee\sum_{x\in V}\varphi(\rho(x))\end{aligned}\tag{4.14}$$

$$
\begin{aligned}
&= \beta \max U - \mathcal{U}_\beta(\eta_\beta) + \ell_\vee \Big( \sum_{x \in M_\rho} \varphi(\rho(x)) + \sum_{x \notin M_\rho} \varphi(\rho(x)) \Big) \\
&\leq \beta \max U - \mathcal{U}_\beta(\eta_\beta) + \ell_\vee \Big( \sum_{x \in M_\rho} \varphi(\rho_\wedge) + \sum_{x \notin M_\rho} \varphi(1/\ell_\wedge) \Big) \\
&\leq \beta \max U - \mathcal{U}_\beta(\eta_\beta) + \ell_\vee |V| \Big( \varphi(\rho_\wedge) + \varphi(1/\ell_\wedge) \Big),
\end{aligned}
$$

where $|V|$ is the cardinality of $V$ and we have used the fact that $\ell(x) \leq \ell_\vee$, $\rho(x) \leq 1/\ell(x) \leq 1/\ell_\wedge$, $\varphi$ is decreasing on $(0,1)$ and increasing on $[1,\infty)$. In short, there is a constant $K > 0$ such that $\mathcal{I}(\beta,\rho) \leq K(\varphi(\rho_\wedge) + \frac{1}{m(m-1)})$, $\forall \rho \in \mathcal{D}_+(V)$. Also, from (4.14), we have

$$
\begin{aligned}
\forall \rho \in \mathcal{D}_+(V), \quad \mathcal{I}(\beta,\rho) &\geq \beta \min U - \mathcal{U}_\beta(\eta_\beta) + \ell_\wedge \sum_{x \in V} \varphi(\rho(x)) \\
&\geq \beta \min U - \mathcal{U}_\beta(\eta_\beta) + \ell_\wedge \varphi(\rho_\wedge),
\end{aligned}
$$

thus if $\rho_n \to \rho^* \in \partial \mathcal{D}_+(V)$ then $\mathcal{I}(\beta,\rho_n) \to +\infty$ (recall $\lim_{r\to 0^+} \varphi(r) = +\infty$). Next, we decompose $\mathcal{G}(\beta,\rho) = \mathcal{G}_1(\beta,\rho) + \mathcal{G}_2(\rho)$, where

$$
\begin{aligned}
\mathcal{G}_1(\beta,\rho) &\coloneqq \frac{1}{2} \sum_{x,y \in V} \ell(x) L(x,y) \theta(\rho(x),\rho(y)) \Big( [\beta \nabla U(x,y)]^2 + 2\beta \nabla U(x,y) \nabla[\varphi' \circ \rho](x,y) \Big) \\
\mathcal{G}_2(\rho) &\coloneqq \frac{1}{2} \sum_{x,y \in V} \ell(x) L(x,y) \theta(\rho(x),\rho(y)) \Big( \nabla[\varphi' \circ \rho](x,y) \Big)^2.
\end{aligned}
$$

Observe that $\forall t,s > 0$, $\theta(s,t)(\varphi'(t) - \varphi'(s)) = t - s$ and $\forall x \neq y \in V$, $|\rho(y) - \rho(x)| \leq \frac{2}{\ell_\wedge}$, $|U(y) - U(x)| \leq \text{osc } U \coloneqq \max U - \min U$. We define

$$
\begin{aligned}
\mathcal{G}_1(\beta,\rho) &= \frac{1}{2} \sum_{x \neq y \in V} \ell(x) L(x,y) \theta(\rho(x),\rho(y)) \Big( [\beta \nabla U(x,y)]^2 + 2\beta \nabla U(x,y) \nabla[\varphi' \circ \rho](x,y) \Big) \\
&\leq \frac{1}{2} \ell_\vee \|L\|_\infty \sum_{x \neq y \in V} \theta(1/\ell_\wedge, 1/\ell_\wedge) \beta^2 (\text{osc } U)^2 + \beta \sum_{x \neq y \in V} \ell(x) L(x,y) (\rho(y) - \rho(x))(U(y) - U(x)) \\
&\leq \frac{1}{2} \ell_\vee \|L\|_\infty |V|^2 \beta^2 (\text{osc } U)^2 + \beta \text{osc } U \sum_{x \neq y \in V} \ell_\vee \|L\|_\infty \frac{2}{\ell_\wedge} \\
&\leq \frac{1}{2} \ell_\vee \|L\|_\infty |V|^2 \beta^2 (\text{osc } U)^2 + \beta \text{osc } U |V|^2 \ell_\vee \|L\|_\infty \frac{2}{\ell_\wedge},
\end{aligned}
$$

where $\|L\|_\infty = \max_{x \in V} |L(x,x)|$. Hence, $\mathcal{G}_1$ is bounded above for all $\rho \in \mathcal{D}_+(V)$ and since $\rho_n \to \rho_* \in \partial \mathcal{D}_+(V)$, we have $\dfrac{\mathcal{G}_1(\beta,\rho_n)}{\mathcal{I}(\beta,\rho_n)} \to 0$. Observe that $\rho_{n,\wedge} \leq 1$, so (recall that $m < 0$)

$$
\varphi(\rho_{n,\wedge}) = \varphi(\rho_n(x_0)) = \frac{\rho_n(x_0)^m - 1 - m(\rho_n(x_0) - 1)}{m(m-1)} \leq \frac{\rho_n(x_0)^m - 1}{m(m-1)}
$$

Therefore, the conclusion of the lemma will follow if we can show $\frac{\mathcal{G}_2(\beta,\rho_n)}{\rho_{n,\wedge}^m} \to +\infty$ because

$$
\frac{\mathcal{G}_2(\rho_n)}{\mathcal{I}(\beta,\rho_n)} \geq \frac{\mathcal{G}_2(\rho_n)}{K(\varphi(\rho_{n,\wedge}) + \frac{1}{m(m-1)})} \geq \frac{m(m-1)}{K} \frac{\mathcal{G}_2(\rho_n)}{(\rho_{n,\wedge})^m}.
$$

By the irreducibility of $L$, we can find $y_0 \notin \{\rho_* = 0\}$ and a path $x_0, x_1, ..., x_q = y_0$ in a way such that $x_i \in \{\rho_* = 0\}$ for all $i = 0, 1, ..., q-1$ and $L(x_0,x_1)L(x_1,x_2)...L(x_{q-1},x_q) > 0$. Denote

$$
\Upsilon \coloneqq \min\{\ell(x)L(x,y) : x, y \in V \text{ such that } \ell(x)L(x,y) > 0\},
$$

then

$$
\begin{aligned}
\forall \rho \in \mathcal{D}_+(V), \qquad \mathcal{G}_2(\rho) &= \frac{1}{2} \sum_{x,y \in V} \ell(x) L(x,y) \theta(\rho(x),\rho(y)) \Big( \varphi'(\rho(y)) - \varphi'(\rho(x)) \Big)^2 \\
&= \frac{1}{2} \sum_{x,y \in V} \ell(x) L(x,y) (\rho(y) - \rho(x)) \Big( \varphi'(\rho(y)) - \varphi'(\rho(x)) \Big) \\
&\geq \Upsilon \sum_{i=0}^{q-1} (\rho(x_{i+1}) - \rho(x_i)) \Big( \varphi'(\rho(x_{i+1})) - \varphi'(\rho(x_i)) \Big). \qquad (4.15)
\end{aligned}
$$

Suppose for now that $\rho(x_i) \le 1\ \forall i = 0, 1, ..., q-1$ and denote for $i = 1, ..., q$, $t_i := \frac{\rho(x_i)}{\rho(x_{i-1})}, r_i = t_1...t_i$, $r_0 := 1$ then from (4.15),

$$\frac{\mathcal{G}_2(\rho)}{\rho(x_0)^m} \ge \Upsilon \sum_{i=0}^{q-1} \frac{\rho(x_i)^m}{\rho(x_0)^m} \frac{(\rho(x_{i+1}) - \rho(x_i))}{\rho(x_i)} \frac{\Big(\varphi'(\rho(x_{i+1})) - \varphi'(\rho(x_i))\Big)}{\rho(x_i)^{m-1}}$$

$$= \Upsilon \sum_{i=0}^{q-1} \frac{\rho(x_i)^m}{\rho(x_0)^m} \frac{(\rho(x_{i+1}) - \rho(x_i))}{\rho(x_i)} \frac{\Big(\rho(x_{i+1})^{m-1} - \rho(x_i)^{m-1}\Big)}{(m-1)\rho(x_i)^{m-1}} \tag{4.16}$$

$$= \frac{\Upsilon}{1-m} \sum_{i=0}^{q-1} r_i^m (t_{i+1} - 1)(1 - t_{i+1}^{m-1}). \tag{4.17}$$

Consider the following family of functions indexed by $k \in [0, +\infty)$

$$\forall t > 0, \qquad p(k,t) := (t-1)(1-t^{m-1}) + kt^m.$$

Observe that $p(0,t) \ge 0$, $p(k,t) = p(0,t) + kt^m$ and $\lim_{t\to+\infty} p(0,t) = +\infty$. For $t \in (0, k^{\frac{1}{1-m}})$, we have $p(k,t) \ge kt^m \ge k^{\frac{1}{1-m}}$. Let $k \ge 1$, for $t \ge k^{\frac{1}{1-m}}$ we have $1 - t^{m-1} \ge 1 - 1/k$ and $t - 1 \ge k^{\frac{1}{m-1}} - 1 \ge 0$. Thus, we have

$$w(k) := \min_{t>0} p(k,t) \ge \min\left\{k^{\frac{1}{1-m}}, (k^{\frac{1}{1-m}} - 1)(1 - \frac{1}{k})\right\} \to +\infty, \text{ as } k \to +\infty.$$

We regard $w$ as a function of $k \in [0, +\infty)$ and introduce the notation $w_j := w \circ ... \circ w$ as the convolution of $j$ times the function $w$ for $j \ge 1$. Note that $\lim_{k\to+\infty} w_j(k) = +\infty$ for all $j \ge 1$. We deduce that the sum in (4.17)

$$\sum_{i=0}^{q-1} r_i^m (t_{i+1} - 1)(1 - t_{i+1}^{m-1}) \ge w_{q-1}(p(0, t_q)). \tag{4.18}$$

For example, if $q = 2$ then the sum in the expression (4.17) is

$$\begin{aligned} p(0,t_1) + t_1^m p(0,t_2) + t_1^m t_2^m p(0,t_3) &= p(0,t_1) + t_1^m (p(0,t_2) + t_2^m p(0,t_3)) \\ &= p(0,t_1) + t_1^m p(p(0,t_3), t_2) \\ &\ge p(0,t_1) + t_1^m w(p(0,t_3)) \\ &= p(w(p(0,t_3)), t_1) \\ &\ge w(w(p(0,t_3))) \\ &= w_2(p(0,t_3)). \end{aligned}$$

Now, we consider three cases: $\rho_*(y_0) < 1$ and $\rho_*(y_0) > 1$ and $\rho_*(y_0) = 1$. If $\rho_*(y_0) < 1$, then for $n$ big enough $\rho_n(y_0) < 1$. Using the expression (4.16) and the estimate (4.18) evaluated at $\rho = \rho_n$, we get

$$\frac{\mathcal{G}_2(\rho)}{\rho_n(x_0)^m} \ge \frac{\Upsilon}{1-m} w_{q-1}\Big(p(0, \frac{\rho_n(y_0)}{\rho_n(x_{q-1})})\Big) \to +\infty \tag{4.19}$$

because $\rho_n(x_{q-1}) \to 0$, $\rho_n(y_0) \to \rho_*(y_0) > 0$ (recall that $x_{q-1} \in \{\rho_* = 0\}$, $y_0 \notin \{\rho_* = 0\}$). If $\rho_*(y_0) > 1$, for $n$ big enough $\rho_n(y_0) \ge 1$ then the equality in (4.16) does not hold when evaluated at $\rho = \rho_n$ because $\varphi'(\rho_n(y_0)) = \rho_n(y_0) - 1$. Instead, the last term $i = q-1$ in the sum in (4.16), ignoring the multiplicative term $\left(\frac{\rho_n(x_{q-1})}{\rho_n(x_0)}\right)^m$, is replaced by

$$\left(\frac{\rho_n(x_q)}{\rho_n(x_{q-1})} - 1\right)\left(\rho_n(x_q)\rho_n(x_{q-1})^{1-m} - \frac{1}{m-1} + \rho_n(x_{q-1})^{1-m}(\frac{1}{m-1} - 1)\right) =: k_n,$$

which tends to infinity because the first term $\frac{\rho_n(x_q)}{\rho_n(x_{q-1})} - 1$ tends to infinity while the second term converges to $\frac{1}{1-m} > 0$. Therefore, we can use the inequality (4.19) with $p(0, \frac{\rho_n(y_0)}{\rho_n(x_{q-1})})$ replaced by $k_n \to \infty$. For the last case $\rho_*(y_0) = 1$, if $\rho_n(x_0)$ is either eventually less than 1 ($< 1$) or eventually no less than 1 ($\ge 1$) then we can use the same arguments just above. If this is not the case, $(\rho_n)_{n\ge0}$ can be split into two subsequences $(\rho_n^{(1)})_{n\ge0}, (\rho_n^{(2)})_{n\ge0}$ such that $\forall n \ge 0$, $\rho_n^{(1)}(y_0) \ge 1$ and $\rho_n^{(2)}(y_0) < 1$ then the same arguments apply for these subsequences. We end our proof of this lemma here. ☐

*Proof.* (Of Proposition 4.4) Let $(\rho_n)_{\ge 0}$ be a minimizing sequence. Since $\mathcal{D}(V)$ is compact, there exists a subsequence, still denoted by $(\rho_n)_{n\ge0}$, converging to some $\rho_* \in \mathcal{D}(V)$. If $\rho_* = \eta_\beta$ then from Lemma 4.5,

$$\chi(\beta) = \lim_{n\to+\infty} \frac{\mathcal{G}(\beta,\rho_n)}{\mathcal{I}(\beta,\rho_n)} = \liminf_{n\to+\infty} \frac{\mathcal{G}(\beta,\rho_n)}{\mathcal{I}(\beta,\rho_n)} > 0.$$

If $\rho_* \in \partial\mathcal{D}_+(V)$, then Lemma 4.6 shows

$$\chi(\beta) = \lim_{n\to+\infty} \frac{\mathcal{G}(\beta,\rho_n)}{\mathcal{I}(\beta,\rho_n)} = +\infty,$$

which is impossible. If $\rho_* \in \mathcal{D}_+(V) \setminus \{\eta_\beta\}$ then clearly $\chi(\beta) = \frac{\mathcal{G}(\beta,\rho_*)}{\mathcal{I}(\beta,\rho_*)} > 0$. Hence, $\chi(\beta) > 0$. ☐

**Corollary 4.7.** *Let $\beta \geq 0$ fixed and $(\rho_t)_{t\geq 0}$ be the associated gradient flow of $\mathcal{U}_\beta$ with the initial condition $\rho_0 \in \mathcal{D}_+(V)$, then we have the following inequalities,*

$$\forall t > 0, \qquad \frac{1}{2}\|\rho_t - \eta_\beta\|^2_{\mathbb{L}^2(\ell)} \leq \ \mathcal{I}(\beta, \rho_t) \leq e^{-\chi(\beta)t}\mathcal{I}(\beta, \rho_0). \tag{4.20}$$

*As a consequence,* $\lim_{t\to+\infty} \mathcal{U}_\beta(\rho_t) = \mathcal{U}_\beta(\eta_\beta)$ *and* $\lim_{t\to+\infty} \rho_t = \eta_\beta$*.*

*Proof.* By differentiating with respect to time and using the identities in (4.9), (4.10),

$$\begin{aligned}\forall t > 0, \qquad \partial_t \mathcal{I}(\beta, \rho_t) \coloneqq \frac{\partial}{\partial t}\mathcal{I}(\beta, \rho_t) &= \frac{\partial}{\partial t}\mathcal{U}_\beta(\rho_t) \\ &= -\mathcal{G}(\beta, \rho_t) \\ &\leq -\chi(\beta)\mathcal{I}(\beta, \rho_t),\end{aligned}$$

and by a Gronwall Inequality (see e.g. Pachpatte [11]), we have the second inequality in (4.20). For the first inequality in (4.20), we use the identity (4.9), together with Taylor's theorem (or mean value theorem) for second derivatives,

$$\begin{aligned}\forall t > 0, \qquad \mathcal{I}(\beta, \rho_t) &= \sum_{x\in V} \ell(x)\Big(\varphi(\rho_t(x)) - \varphi(\eta_\beta(x)) - \varphi'(\eta_\beta(x))(\rho_t(x) - \eta_\beta(x))\Big) \\ &= \frac{1}{2}\sum_{x\in V} \ell(x)\varphi''(\lambda_t(x))(\rho_t(x) - \eta_\beta(x))^2 \\ &\geq \frac{1}{2}\sum_{x\in V} \ell(x)(\rho_t(x) - \eta_\beta(x))^2 \\ &= \frac{1}{2}\|\rho_t - \eta_\beta\|^2_{\mathbb{L}^2(\ell)},\end{aligned}$$

where in the second equality, $\lambda_t : V \to \mathbb{R}^+$ is some vector in the Taylor's theorem satisfying $\lambda_t(x) \in (\rho_t(x) \wedge \eta_\beta(x), \rho_t(x) \vee \eta_\beta(x))$, $\forall x \in V$, where $x \wedge y \coloneqq \min\{x, y\}$, $x \vee y \coloneqq \max\{x, y\}$. ☐

**Proposition 4.8** (An upper bound for $\chi(\beta)$)**.** *Define the Markov generator $Q_\beta$ as follows*

$$\forall x \neq y \in V, \qquad Q_\beta(x, y) \coloneqq \varphi''(\eta_\beta(x))L(x, y)\theta(\eta_\beta(x), \eta_\beta(y)).$$

*We can easily check that $Q_\beta$ is irreducible and the probability measure $\ell_\beta = (\ell_\beta(x))_{x\in V}$, defined by*

$$\forall x \in V, \qquad \ell_\beta(x) \coloneqq \frac{\ell(x)[\varphi''(\eta_\beta(x))]^{-1}}{\sum_{y\in V} \ell(y)[\varphi''(\eta_\beta(y))]^{-1}},$$

*is reversible for $Q_\beta$. Then we have*

$$\chi(\beta) \leq \lambda(Q_\beta),$$

*where $\lambda(Q_\beta)$ is the spectral gap of $(Q_\beta, \ell_\beta)$.*

*Proof.* Let $h \in \mathbb{R}^V$ be such that $\ell[h] = 0$. Then from the proof of Lemma 4.5, we have

$$\begin{aligned}\lim_{\epsilon\to 0} \frac{\mathcal{G}(\beta, \eta_\beta + \epsilon h)}{\mathcal{I}(\beta, \eta_\beta + \epsilon h)} &= \frac{\sum_{x,y\in V} \ell(x)L(x, y)\theta(\eta_\beta(x), \eta_\beta(y))\Big(\varphi''(\eta_\beta(x))h(x) - \varphi''(\eta_\beta(y))h(y)\Big)^2}{\sum_{x\in V} \ell(x)\varphi''(\eta_\beta(x))h^2(x)} \\ &= \frac{\sum_{x,y\in V} \ell_\beta(x)Q_\beta(x, y)\Big(\varphi''(\eta_\beta(x))h(x) - \varphi''(\eta_\beta(y))h(y)\Big)^2}{\sum_{x\in V} \ell_\beta(x)[\varphi''(\eta_\beta(x))h(x)]^2}.\end{aligned}$$

Set $f(x) \coloneqq \varphi''(\eta_\beta(x))h(x)$, $\forall x \in V$. Then

$$\begin{aligned}\ell_\beta[f] &= \frac{\sum_{x\in V} \ell(x)[\varphi''(\eta_\beta(x))]^{-1}\varphi''(\eta_\beta(x))h(x)}{\sum_{x\in V} \ell(x)[\varphi''(\eta_\beta(x))]^{-1}} \\ &= \frac{\ell[h]}{\sum_{x\in V} \ell(x)[\varphi''(\eta_\beta(x))]^{-1}} \\ &= 0.\end{aligned}$$

Hence,

$$\lim_{\epsilon\to 0} \frac{\mathcal{G}(\beta, \eta_\beta + \epsilon h)}{\mathcal{I}(\beta, \eta_\beta + \epsilon h)} = \frac{\sum_{x\in V} \ell_\beta(x)Q_\beta(x, y)(f(y) - f(x))^2}{\ell_\beta[f^2]}$$

$$= \frac{-\langle f, Q_\beta[f]\rangle_{\mathbb{L}^2(\ell_\beta)}}{2\mathrm{Var}_{\ell_\beta}[f]} = -\frac{\ell_\beta[fQ_\beta[f]]}{2\mathrm{Var}_{\ell_\beta}[f]}$$
$$\geq \lambda(Q_\beta).$$

As $h$ varies over the set $\{h \in \mathbb{R}^V : \ell[h] = 0\}$, $f$ takes all values in $\{f \in \mathbb{R}^V : \ell_\beta[f] = 0\}$: take $f' \in \{f \in \mathbb{R}^V : \ell_\beta[f] = 0\}$, define $h'(x) \coloneqq f'(x)[\varphi''(\eta_\beta(x)]^{-1}$ then $\ell[h'] = \ell_\beta[f'] = 0$. By the variational characterization of $\lambda(Q_\beta)$,

$$\lambda(Q_\beta) = \inf_{f: f\notin \mathrm{Vect}(1), \ell_\beta[f]=0} -\frac{\ell_\beta[fQ_\beta[f]]}{2\mathrm{Var}_{\ell_\beta}[f]}$$
$$= \inf_{h:\ell[h]=0} \lim_{\epsilon\to 0} \frac{\mathcal{G}(\beta, \eta_\beta + \epsilon h)}{\mathcal{I}(\beta, \eta_\beta + \epsilon h)}$$
$$\geq \chi(\beta)$$

by the definition of $\chi(\beta)$, which shows the announced result. □

## 4.3 Nonlinear Markov representation

For $\beta \geq 0$ fixed, the dynamic (4.7) can be reinterpreted as a nonlinear Markov dynamic satisfying

$$\forall t > 0, \qquad \dot{\mu}_t = \mu_t L_{\beta,\rho_t}, \tag{4.21}$$

where $\mu_t \in \mathcal{P}_+(V)$ is the probability measure on $V$ admitting $\rho_t$ as its density with respect to $\ell$. The generator in (4.21) is given by

$$\forall \rho \in \mathcal{D}_+(V),\ \forall x \neq y, \quad L_{\beta,\rho}(x,y) = L(x,y)\frac{\theta(\rho(x),\rho(y))}{\rho(x)}\Big(\nabla[\beta U + \varphi' \circ \rho](x,y)\Big)_-, \tag{4.22}$$

where $\rho$ is the density with respect to $\ell$ of $\mu$ and $x_- = \max\{0, -x\}$. We can write (4.22) more explicitly as follows

$$\forall \rho \in \mathcal{D}_+(V),\ \forall x \neq y,\ L_{\beta,\rho}(x,y) = L(x,y)\left(\frac{\rho(y)-\rho(x)}{\rho(x)[\varphi'(\rho(y)) - \varphi'(\rho(x))]}\beta(U(y)-U(x)) + \frac{\rho(y)}{\rho(x)} - 1\right)_-. \tag{4.23}$$

Using the formula (4.23), we can therefore simulate a Markov process that has its law satisfying (4.21). Our algorithm is then an approximation of this Markov process through particle systems. The details can be found in the second appendix. We now show that the nonlinear interpretation 4.21 holds.

**Theorem 4.9.** *Let $G$ be an irreducible Markov generator on $V$ with reversible measure $\pi > 0$ and $H = (H(x,y))_{x,y\in V}$ be a function with the property $\forall x \neq y$, $H(x,y) = -H(y,x)$. Define the divergence operator $\mathrm{div}_G : \mathbb{R}^{V\times V} \to \mathbb{R}^V$ by*

$$\forall F \in \mathbb{R}^{V\times V}, \forall x \in V \qquad \mathrm{div}_G[F](x) \coloneqq \frac{1}{2}\sum_{y\in V} G(x,y)(F(x,y) - F(y,x))$$

*Denote $H_-(x,y) \coloneqq -\min\{H(x,y), 0\}$ and similarly $H_+(x,y) \coloneqq \max\{H(x,y), 0\}$, then for any test function $f \in \mathbb{R}^V$ we have*

$$\pi[f\mathrm{div}_G H] = \pi[G_{H_-}[f]], \tag{4.24}$$

*where $G_{H_-}$ is the Markov generator defined by*

$$\forall x \neq y, \qquad G_{H_-}(x,y) \coloneqq G(x,y)H_-(x,y). \tag{4.25}$$

*Proof.* We have

$$\pi[f\mathrm{div}_G H] = \sum_{x\in V} \pi(x)f(x)\mathrm{div}_G H(x)$$
$$= \sum_{x\in V} \pi(x)f(x)\sum_{y\in V}\frac{1}{2}G(x,y)(H(x,y) - H(y,x))$$
$$= \sum_{x\neq y\in V} \pi(x)G(x,y)H(x,y)f(x)$$
$$= \sum_{x\neq y\in V} \pi(x)G(x,y)H_+(x,y)f(x) - \sum_{x\neq y\in V} \pi(x)G(x,y)H_-(x,y)f(x).$$

Recall that $\pi$ is reversible for $G$ and from the assumptions on $H$, it holds that $H_-(y,x) = H_+(x,y)$. We then have

$$\pi[fG_{H_-}[f]] = \sum_{x\in V} \pi(x)G_{H_-}f(x)$$

$$
\begin{aligned}
&= \sum_{x\in V}\sum_{y\in V}\pi(x)G_{H_-}(x,y)f(y)\\
&= \sum_{x\neq y\in V}\pi(x)G_{H_-}(x,y)f(y) + \sum_{x\in V}\pi(x)G_{H_-}(x,x)f(x)\\
&= \sum_{x\neq y\in V}\pi(x)G(x,y)H_-(x,y)f(y) - \sum_{x\in V}\pi(x)f(x)\sum_{y\in V\setminus\{x\}}G(x,y)H_-(x,y)\\
&= \sum_{x\neq y\in V}\pi(y)G(y,x)H_-(y,x)f(x) - \sum_{x\neq y\in V}\pi(x)G(x,y)H_-(x,y)f(x)\\
&= \sum_{x\neq y\in V}\pi(x)G(x,y)H_+(x,y)f(x) - \sum_{x\neq y\in V}\pi(x)G(x,y)H_-(x,y)f(x),
\end{aligned}
$$

and hence $\pi[f\mathrm{div}_G H] = \pi[fG_{H_-}[f]]$. □

Now we can apply Theorem 4.9 to the curve of generator

$$
\forall x\neq y\in V,\ \forall t\geq 0,\qquad G_{\rho_t}(x,y) \coloneqq L(x,y)\frac{\theta(\rho_t(x),\rho_t(y))}{\rho_t(x)}
$$

and the curve

$$
\begin{aligned}
\forall x,y\in V,\ \forall t\geq 0,\qquad H_{\rho_t}(x,y) &\coloneqq \nabla[\beta U + \varphi'\circ\rho_t](x,y)\\
&= \beta(U(y)-U(x)) + \varphi'(\rho_t(y)) - \varphi'(\rho_t(x)).
\end{aligned}
$$

Observe that $\mu_t$ is reversible for $G_{\rho_t}$:

$$
\begin{aligned}
\mu_t(x)G_{\rho_t}(x,y) &= \ell(x)\rho_t(x)\times L(x,y)\frac{\theta(\rho_t(x),\rho_t(y))}{\rho_t(x)}\\
&= \ell(x)L(x,y)\theta(\rho_t(x),\rho_t(y))\\
&= \mu_t(y)G_{\rho_t}(y,x)
\end{aligned}
$$

by the symmetry of $\theta$ and the fact that $\ell$ is reversible for $L$. We have for any test function $f\in\mathbb{R}^V$,

$$
\begin{aligned}
\dot\mu_t[f] &= \sum_{x\in V}\pi(x)\dot\rho_t(x)f(x)\\
&= \sum_{x\in V}\pi(x)f(x)\sum_{y\in V}L(x,y)\theta(\rho_t(x),\rho_t(y))\nabla[\beta U+\varphi'\circ\rho_t](x,y)\\
&= \sum_{x,y\in V}\pi(x)\rho_t(x)f(x)L(x,y)\frac{\theta(\rho_t(x),\rho_t(y))}{\rho_t(x)}\nabla[\beta U+\varphi'\circ\rho_t](x,y)\\
&= \sum_{x,y\in V}\mu_t f(x)G_{\rho_t}(x,y)H_{\rho_t}(x,y)\\
&= \mu_t[f\mathrm{div}_{G_{\rho_t}}H_{\rho_t}]\\
&= \mu_t[L_{\beta,\rho_t}[f]],
\end{aligned}
\tag{4.26}
$$

where we used Theorem 4.9 with the triple $(\pi,G,H) = (\mu_t, G_{\rho_t}, H_{\rho_t})$ and $L_{\beta,\rho_t}$ in (4.22) plays the role of the generator $G_{H_-}$. Hence, we have transformed the dynamic (4.7) to a nonlinear Markov representation as announced in (4.21), which is useful for a simulation.

Observe that the generator given in (4.23) is not irreducible for all $\rho\in\mathcal{D}_+(V)$. For example, it may happen that for a fixed $x$, $\beta\nabla U(x,y) > -\nabla[\varphi'\circ\rho](x,y)$ for all $y\neq x$, and hence $L_{\beta,\rho}(x,y) = 0$. In fact, there is another Markov interpretation such that the nonlinear generator is irreducible for all $\rho\in\mathcal{D}_+(V)$. We can replace the generator $L_{\beta,\rho}$ by $Q_{\beta,\rho}$ defined by

$$
\forall\rho\in\mathcal{D}_+(V),\ \forall x\neq y,\quad Q_{\beta,\rho}(x,y) \coloneqq L(x,y)\Big(1+\frac{\theta(\rho(x),\rho(y))}{\rho(x)}\beta(U(y)-U(x))_-\Big),
\tag{4.27}
$$

or more explicitly

$$
\forall\rho\in\mathcal{D}_+(V),\ \forall x\neq y,\quad Q_{\beta,\rho}(x,y) \coloneqq L(x,y)\Big(1+\frac{\rho(y)-\rho(x)}{\rho(x)(\varphi'(\rho(y))-\varphi'(\rho(x)))}\beta(U(y)-U(x))_-\Big).
\tag{4.28}
$$

Clearly $\forall x\neq y$, $Q_{\beta,\rho}(x,y)\geq L(x,y)$ hence $Q_{\beta,\rho}(x,y)$ is irreducible. The generator $Q_{\beta,\rho}$ is obtained by the following computations. We have

$$
\dot\mu_t[f] = \sum_{x\in V}\ell(x)\dot\rho_t(x)f(x)
$$

$$
\begin{aligned}
&= \sum_{x\in V}\ell(x)f(x)\sum_{y\in V}L(x,y)\theta(\rho_t(x),\rho_t(y))\nabla[\beta U+\varphi'\circ\rho_t](x,y)\\
&= \sum_{x\in V}\ell(x)f(x)\sum_{y\in V}L(x,y)\theta(\rho_t(x),\rho_t(y))\beta\nabla U(x,y)+\sum_{x\in V}\ell(x)f(x)\sum_{y\in V}L(x,y)\theta(\rho_t(x),\rho_t(y))\nabla[\varphi'\circ\rho_t](x,y)\\
&= \sum_{x\in V}\ell(x)f(x)\sum_{y\in V}L(x,y)\theta(\rho_t(x),\rho_t(y))\beta\nabla U(x,y)+\sum_{x\in V}\ell(x)f(x)\sum_{y\in V}L(x,y)(\rho_t(y)-\rho_t(x))\\
&= \sum_{x\in V}\mu_t(x)f(x)\sum_{y\in V}L(x,y)\frac{\theta(\rho_t(x),\rho_t(y))}{\rho_t(x)}\beta\nabla U(x,y)+\sum_{x\in V}\ell(x)f(x)\sum_{y\in V}L(x,y)\rho_t(y)\\
&= \sum_{x\in V}\mu_t(x)f(x)\sum_{y\in V}L(x,y)\frac{\theta(\rho_t(x),\rho_t(y))}{\rho_t(x)}\beta\nabla U(x,y)+\sum_{x\in V}\ell(x)f(x)\sum_{y\in V}L(x,y)\rho_t(y)\\
&= \mu_t[f\mathrm{div}_{G_{\rho_t}}[\beta\nabla U]]+\ell[fL[\rho_t]],
\end{aligned}
$$

where the generator $G_\rho$ is defined by

$$\forall\rho\in\mathcal{D}_+(V),\ \forall x\neq y,\qquad G_\rho(x,y)\coloneqq L(x,y)\frac{\theta(\rho(x),\rho(y))}{\rho(x)}$$

Since $\ell$ is reversible for $L$, the last expression is equal to $\ell[fL[\rho_t]]=\ell[\rho_tL[f]]=\mu_t[L[f]]$. Applying Theorem 4.9, we get

$$\mu_t[f\mathrm{div}_{G_{\rho_t}}[\beta\nabla U]]=\mu_t[G_{\rho_t,\beta\nabla U_-}[f]],$$

where $G_{\rho,\beta\nabla U_-}$ is the Markov generator defined by

$$\forall\rho\in\mathcal{D}_+(V),\ \forall x\neq y,\qquad G_{\rho_t,\beta\nabla U_-}(x,y)\coloneqq L(x,y)\frac{\theta(\rho(x),\rho(y))}{\rho(x)}\beta(U(y)-U(x))_-.$$

Observe that $Q_{\beta,\rho_t}=L+G_{\rho_t,\beta\nabla U_-}$, putting together these computations we get

$$\dot{\mu}_t[f]=\mu_t[(L+G_{\rho_t,\beta\nabla U_-})[f]]=\mu_t[Q_{\beta,\rho_t}[f]],$$

which leads to the non-linear interpretation in (4.27), (4.28).

So far, we have two generators that give rise to two different nonlinear Markov processes, both admitting $(\mu_t)_{t\geq 0}$ as their time-marginal laws. Clearly, we can also replace these generators with any convex combination of these two generators and still get a new Markov process with the same time-marginal laws. For the density $\eta_\beta$ in Theorem 3.1, the probability measure associated with this density, denoted $\zeta_\beta$, is the unique invariant measure of $Q_{\beta,\eta_\beta}$, in the sense that $\zeta_\beta Q_{\beta,\eta_\beta}=0$, while $L_{\beta,\eta_\beta}=0$. Indeed, recall that $\varphi'(\eta_\beta(x))+\beta U(x)=\varphi'(\eta_\beta(y))+\beta U(y)$ for any $x,y\in V$, putting this back in (4.23), we see $L_{\beta,\eta_\beta}(x,y)=0$. Fix one $x\in V$, we have

$$
\begin{aligned}
\zeta_\beta Q_{\beta,\eta_\beta}(x) &= \sum_{y\in V}\zeta_\beta(y)Q_{\beta,\eta_\beta}(y,x)\\
&= \sum_{y\in V\setminus\{x\}}\ell(y)\eta_\beta(y)L(y,x)\Big(1+\frac{\theta(\eta_\beta(x),\eta_\beta(y))}{\eta_\beta(y)}\beta(U(x)-U(y))_-\Big)+\ell(x)\eta_\beta(x)Q_{\beta,\eta_\beta}(x,x)\\
&= \ell(x)\sum_{y\in V\setminus\{x\}}L(x,y)\Big(\eta_\beta(y)+(\eta_\beta(y)-\eta_\beta(x))_-\Big)\\
&\qquad -\ell(x)\eta_\beta(x)\sum_{y\in V\setminus\{x\}}L(x,y)\Big(1+\frac{\theta(\eta_\beta(x),\eta_\beta(y))}{\eta_\beta(x)}\beta(U(y)-U(x))_-\Big)\\
&= \ell(x)\sum_{y\in V\setminus\{x\}}L(x,y)\Big(\eta_\beta(y)-\eta_\beta(x)+(\eta_\beta(y)-\eta_\beta(x))_--(\eta_\beta(x)-\eta_\beta(y))_-\Big)\\
&= 0,
\end{aligned}
$$

where we have used the fact that

$$
\begin{aligned}
\theta(\eta_\beta(x),\eta_\beta(y))\beta(U(x)-U(y))_- &= \Big(\frac{\eta_\beta(y)-\eta_\beta(x)}{\varphi'(\eta_\beta(y))-\varphi'(\eta_\beta(x))}\times(\varphi'(\eta_\beta(y))-\varphi'(\eta_\beta(x)))\Big)_-\\
&= (\eta_\beta(y)-\eta_\beta(x))_-.
\end{aligned}
$$

We will compare the simulations of particle systems using these two different generators (4.22) and (4.27) in Section 6.

# 5 The time-inhomogeneous situation

In this section, we will consider the case when $\beta$ depends on time, i.e., $\beta = (\beta_t)_{t\geq 0}$ is an inverse temperature schedule. Inspired by the gradient flow dynamic of $\mathcal{U}_\beta$ (4.7) and its homogeneous nonlinear Markov representation (4.21), we consider the time-inhomogeneous dynamic

$$\forall x \in V,\ \forall t > 0, \quad \dot{\rho}_t(x) = \sum_{y\in V} L(x,y)\theta(\rho_t(x),\rho_t(y))\Big(\nabla[\beta_t U + \varphi' \circ \rho_t](x,y)\Big), \quad \rho_0 \in \mathcal{D}_+(V). \tag{5.1}$$

As we shall see, for some appropriate choices of the temperature schedule $\beta = (\beta_t)_{t\geq 0}$ such that $t \mapsto \beta_t$ increases slowly enough to infinity, a unique solution to (5.1) will exist. Moreover, $\mu_t$ (the measure with density $\rho_t$) will converge to a measure concentrated on $\mathcal{M}(U)$, the set of global minimizers of $U$. In the meantime, we will always assume that the temperature schedule $\beta = (\beta_t)_{t\geq 0}$ satisfies

**Assumption 5.1.** $\lim_{t\to+\infty}\beta_t = +\infty$ *and* $t \mapsto \beta_t$ *is continuously differentiable with* $\dot{\beta}_t \coloneqq \partial\beta/\partial t > 0$, $\forall t > 0$ *and* $\beta_0 \geq 0$.

The crucial aspect of this section is a functional inequality that provides a lower bound of $\rho_{t,\wedge} \coloneqq \min_{x\in V}\rho_t(x)$ for a solution $(\rho_t)_{t\in[0,T)}$ of (5.1) on some interval $[0,T)$ (the existence and uniqueness on a small interval is guaranteed by Cauchy-Picard theorem) in terms of $\beta_t$. Consequently, it facilitates proving the existence and uniqueness of such a solution on the entire half-real line $[0,+\infty)$ as well as the convergence of $(\mu_t)_{t\geq 0}$ ($\mu_t$ being the measure with density $\rho_t$) to the measure $\mu_\infty$ in (3.5).

## 5.1 A functional inequality

For any $t \geq 0$, we let $\nu_t$ be the global mimimizer of the function $\mathcal{U}_{\beta_t}$, i.e., $\nu_t$ is the density satisfying

$$\forall x \in V, \qquad \varphi'(\nu_t(x)) + \beta_t U(x) = c(\beta_t),$$

where $c(\beta_t)$ is the unique number that solves the equation (recall that $g = (\varphi')^{-1}$ in (3.4))

$$c \in \mathbb{R}, \qquad \sum_{x\in V}\ell(x)g(c - \beta_t U(x)) = 1.$$

In other words, $\nu_t = \eta_{\beta_t}$, where $\beta \mapsto \eta_\beta$ is given in Theorem 3.1. The following properties of $\nu = (\nu_t)_{t\geq 0}$ are consequences of Theorem 3.2.

**Theorem 5.2.** *Let* $\beta = (\beta_t)_{t\geq 0}$ *satisfy Assumption 5.1*

*(i)* $\forall x, y \in V$ *such that* $U(x) = U(y)$, $\nu_t(x) = \nu_t(y)$ *(in particular* $x, y \in \mathcal{M}(U)$*, the set of global minimizers of* $U$ *defined in Theorem 1.1).*

*(ii)* $\forall x \notin \mathcal{M}(U)$, $\lim_{t\to+\infty}\nu_t(x) = 0$. *More precisely, we have*

$$\forall x \notin \mathcal{M}(U), \qquad \lim_{t\to+\infty}\beta_t[\nu_t(x)]^{1-m} = \frac{1}{(1-m)(U(x)-\min U)},$$

*where* $m < 0$ *is fixed in* (3.2).

*(iii)* $\forall x \notin \mathcal{M}(U)$, *we have the inequalities*

$$\forall t \geq 0, \qquad \nu_t(x)^{1-m} \geq \frac{1}{\beta_t(1-m)\mathrm{osc}\ U + 1}, \quad \text{where} \quad \mathrm{osc}\ U = \max U - \min U.$$

*(iv)* $\forall x \in \mathcal{M}(U)$, $\nu_t(x) \geq 1$ *and* $\dot{\nu}_t \coloneqq \dfrac{\partial \nu_t(x)}{\partial t} \geq 0$, *where the equality holds iff* $U$ *is constant on* $V$.

*(v)* $\forall x \in \mathcal{M}(U)$, $\lim_{t\to+\infty}\nu_t(x) = \dfrac{1}{\sum_{y\in\mathcal{M}(U)}\ell(y)}$ *and so* $\lim_{t\to+\infty}\ell(x)\nu_t(x) = \zeta_\infty(x)$, *where* $\zeta_\infty$ *is given in* (3.5).

*Proof.* We prove *(iii)* because the rest are just restatements of Theorem 3.2. Indeed, from the proof of *(ii)* of Theorem 3.2, we have

$$\beta_t[\nu_t(x)]^{1-m} = \frac{1}{(1-m)(U(x)-\min U) + \frac{(m-1)\varphi'(\nu_t(x_0))+1}{\beta_t}}, \tag{5.2}$$

where $x \notin \mathcal{M}(U)$ and $x_0 \in \mathcal{M}(U)$ and observe that $\nu_t(x_0) \geq 0$ so $\varphi'(\nu_t(x_0)) \geq 0$. Therefore the denominator of (5.2) is smaller than $(1-m)(U(x)-\min U) + \beta_t^{-1}$, so after dividing both sides with $\beta_t$, the denominator of the left hand side is $\beta_t(1-m)\mathrm{osc}\ U + 1$, which is the desired result. □

Having defined $\nu = (\nu_t)_{t\geq 0}$, we can interpret its as a curve of "instantaneous invariant density" in the sense that $\forall t \geq 0$, $v_t Q_{\beta_t,\nu_t} = 0$, where $v_t$ is the probability measure admitting $\nu_t$ as its density and $Q_{\beta,\rho}$ is the generator given in (4.27), (4.28). Also, note that $\forall t \geq 0$, $L_{\beta_t,\nu_t} = 0$, where $L_{\beta,\rho}$ is given in (4.22), (4.23), because $\nabla \beta_t U(x,y) = -\nabla \varphi' \circ \nu_t(x,y)$, $\forall x \neq y$. We move on to the definition of the functionals that are concerned in this section. Given $\varphi$ defined in (3.2), consider

$$\forall t \geq 0,\ \forall \rho \in \mathcal{D}_+(V),\ \mathcal{I}(t,\rho) \coloneqq \sum_{x\in V} \ell(x)\Big(\varphi(\rho(x)) - \varphi(\nu_t(x)) - \varphi'(\nu_t(x))(\rho(x) - \nu_t(x))\Big), \tag{5.3}$$

$$\forall t \geq 0,\ \forall \rho \in \mathcal{D}_+(V),\ \mathcal{G}(t,\rho) \coloneqq \frac{1}{2}\sum_{x\in V}\sum_{y\in V} \ell(x)L(x,y)\theta(\rho(x),\rho(y))\Big(\nabla[\beta_t U + \varphi' \circ \rho](x,y)\Big)^2, \tag{5.4}$$

where we recall the notation $\nabla[\beta_t U + \varphi' \circ \rho](x,y) = \beta_t(U(y) - U(x)) + \varphi'(\rho(y)) - \varphi'(\rho(x))$. Observe that $\mathcal{I}(t,\rho)$ and $\mathcal{G}(t,\rho)$ are exactly $\mathcal{I}(\beta_t,\rho)$ and $\mathcal{G}(\beta_t,\rho)$ in (4.9),(4.10) evaluated at $\beta = \beta_t$, respectively. Below is a functional inequality that will be crucial in the proofs of convergence theorems of the dynamic (5.1).

**Theorem 5.3.** *Let $\varphi$ and $\theta$ be given in* (3.2) *and* (4.1), *and fix a $\rho_* \in \mathcal{D}_+(V)$ we consider the following functionals*

$$\forall \rho \in \mathcal{D}_+(V),\quad \mathcal{I}_*(\rho) \coloneqq \sum_{x\in V} \ell(x)\Big(\varphi(\rho(x)) - \varphi(\rho_*(x)) - \varphi'(\rho_*(x))(\rho(x) - \rho_*(x))\Big),$$
$$\mathcal{G}_*(\rho) \coloneqq \frac{1}{2}\sum_{x,y\in V} \ell(x)L(x,y)\theta(\rho(x),\rho(y))\Big(\varphi'(\rho(x)) - \varphi'(\rho_*(x)) - \varphi'(\rho(y)) + \varphi'(\rho_*(y))\Big)^2.$$

*Then it holds that*

$$\mathcal{G}_*(\rho) \geq \Lambda(\rho)\mathcal{I}_*(\rho),$$

*where $\Lambda(\rho)$ is the spectral gap of the generator $K_\rho$ given by*

$$\forall x \neq y, \qquad K_\rho(x,y) \coloneqq L(x,y)\frac{\theta(\rho(x),\rho(y))}{\theta(\rho(x),\rho_*(y))},$$

*which has the reversible invariant measure $\mu_\rho$ defined by*

$$\forall x \in V, \qquad \mu_\rho(x) \coloneqq \ell(x)\theta(\rho(x),\rho_*(x)).$$

*Proof.* We will prove that for all $\rho \in \mathcal{D}_+(V)$, we have

$$\mathcal{G}_*(\rho) \geq \Lambda(\rho)\sum_{x\in V}\Big(\varphi'(\rho(x)) - \varphi'(\rho_*(x))\Big)(\rho(x) - \rho_*(x))\ell(x).$$

Let $f$ defined on $V$ by

$$\forall x \in V, \qquad f(x) \coloneqq \varphi'(\rho(x)) - \varphi'(\rho_*(x)),$$

so that

$$\begin{aligned}\mathcal{G}_*(\rho) &= \frac{1}{2}\sum_{x,y\in V}\mu_\rho(x)K_\rho(x,y)\Big(f(y) - f(x)\Big)^2\\ &= -\mu_\rho[fK_\rho[f]].\end{aligned}$$

Note that

$$\begin{aligned}\mu_\rho[f] &= \sum_{x\in V}\ell(x)\theta(\rho(x),\rho_*(x))\Big(\varphi'(\rho(x)) - \varphi'(\rho_*(x))\Big)\\
&= \sum_{x\in V:\varphi'(\rho(x))\neq\varphi(\rho_*(x))}\ell(x)\theta(\rho(x),\rho_*(x))\Big(\varphi'(\rho(x)) - \varphi'(\rho_*(x))\Big)\\
&= \sum_{x\in V:\varphi'(\rho(x))\neq\varphi(\rho_*(x))}\ell(x)\frac{\rho(x)-\rho_*(x)}{\varphi'(\rho(x)) - \varphi'(\rho_*(x))}\Big(\varphi'(\rho(x)) - \varphi'(\rho_*(x))\Big)\\
&= \sum_{x\in V:\varphi'(\rho(x))\neq\varphi(\rho_*(x))}\ell(x)(\rho(x) - \rho_*(x))\\
&= \sum_{x\in V}\ell(x)(\rho(x) - \rho_*(x))\\
&= 0.\end{aligned}$$

It follows that

$$\begin{aligned}-\mu_\rho[fK_\rho[f]] &\geq \Lambda(\rho)\mu_\rho[f^2]\\ &= \Lambda(\rho)\sum_{x\in V}\ell(x)\theta(\rho(x),\rho_*(x))\Big(\varphi'(\rho(x))-\varphi'(\rho_*(x))\Big)^2\\ &= \Lambda(\rho)\sum_{x\in V}\ell(x)(\rho(x)-\rho_*(x))\Big(\varphi'(\rho(x))-\varphi'(\rho_*(x))\Big),\end{aligned}$$

which is the announced result. Now using Lemma 4.1, we have

$$\begin{aligned}\mathcal{G}_*(\rho) &\geq \Lambda(\rho)\sum_{x\in V}\ell(x)(\rho(x)-\rho_*(x))\Big(\varphi'(\rho(x))-\varphi'(\rho_*(x))\Big),\\ &\geq \Lambda(\rho)\sum_{x\in V}\ell(x)\Big(\varphi(\rho(x))-\varphi(\rho_*(x))-\varphi'(\rho_*(x))(\rho(x)-\rho_*(x))\Big)\\ &= \Lambda(\rho)\mathcal{I}_*(\rho).\end{aligned}$$

□

The dependence of $\Lambda(\rho)$ on $\rho$ can be problematic if we have little infomation on $\rho$. The following observation will be useful in this respect.

**Proposition 5.4.** *With the settings in Theorem 5.3, we have for all $\rho \in \mathcal{D}_+(V)$,*

$$\Lambda(\rho) \geq \lambda\frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_\wedge)},$$

*where $\lambda$ is the spectral gap of $L$, $\rho_\wedge \coloneqq \min_{x\in V}\rho(x)$, $\ell_\wedge \coloneqq \min_{x\in V}\ell(x)$.*

*Proof.* Due to the variational principle and monotonicity in each argument of $\theta$, we have for any $\rho \in \mathcal{D}_+(V)$,

$$\begin{aligned}\Lambda(\rho) &= \frac{1}{2}\inf_{f\in\mathbb{R}^V\setminus\mathrm{Vect}(1)}\frac{\sum_{x,y\in V}\ell(x)L(x,y)\theta(\rho(x),\rho(y))(f(y)-f(x))^2}{\inf_{c\in\mathbb{R}}\sum_{x\in V}\ell(x)\theta(\rho(x),\rho_*(y))(f(x)-c)^2}\\ &\geq \frac{\theta(\rho_\wedge,\rho_\wedge)}{\theta(\rho_\vee,\rho_{*,\vee})}\times\frac{1}{2}\inf_{f\in\mathbb{R}^V\setminus\mathrm{Vect}(1)}\frac{\sum_{x,y\in V}\ell(x)L(x,y))(f(y)-f(x))^2}{\inf_{c\in\mathbb{R}}\sum_{x\in V}\ell(x)(f(x)-c)^2}\\ &= \lambda\frac{\theta(\rho_\wedge,\rho_\wedge)}{\theta(\rho_\vee,\rho_{*,\vee})}\end{aligned}$$

where we denote $\rho_\wedge \coloneqq \min_{x\in V}\rho(x)$ and $\rho_\vee \coloneqq \max_{x\in V}\rho(x)$. We also note that $\rho_\vee \leq 1/\ell_\wedge$, thus $\theta(\rho_\vee,\rho_{*,\vee}) \leq \theta(1/\ell_\wedge, 1/\ell_\wedge)$ and so

$$\Lambda(\rho) \geq \lambda\frac{\theta(\rho_\wedge,\rho_\wedge)}{\theta(1/\ell_\wedge,1/\ell_\wedge)} = \frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_\wedge)}.$$

□

## 5.2 Existence, uniqueness and convergence of $(\rho_t)_{t\geq 0}$

Apply Theorem 5.3 and Proposition 5.4, we have the following estimate

$$\mathcal{G}(t,\rho) \geq \lambda\frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_\wedge)}\mathcal{I}(t,\rho), \tag{5.5}$$

We now come the the main result of this paper

**Theorem 5.5.** *For any $m < 0$, consider the function $\varphi = \varphi_{m,2}$ in (3.2), as well as the time-inhomogeneous inverse temperature scheme*

$$\forall t \geq 0, \qquad \beta_t = (t_0+t)^\alpha - 1,$$

*where $t_0 \geq 1$ and*

$$0 < \alpha \leq \kappa(m) \coloneqq \frac{-m}{2(1-m)} \in \left(0,\frac{1}{2}\right).$$

*Then there exists a unique solution to the equation (5.1), which exists for all time $t \geq 0$ and satisfies*

$$\lim_{t\to+\infty}\mu_t[\mathcal{M}(U)] = 1.$$

The proof is decomposed into several intermediate results.

**Lemma 5.6.** *There exists a unique solution* $(\rho_t)_{t\in[0,T)}$ *on a maximal interval* $[0,T)$ *to the equation* (5.1) *and a constant* $K>0$ *such that*

$$\forall t\in[0,T),\qquad \rho_{t,\wedge}^{-m}\geq \frac{1}{K(\beta_t+1)}.$$

*Proof.* Similarly to Proposition 4.3, the existence and uniqueness on a maximal interval $[0,T)$ of the dynamic (5.1) is guaranteed by Cauchy-Picard Theorem. Indeed, it can be easily checked that the mapping

$$[0,+\infty)\times\mathcal{D}_+(V)\ni(t,\rho)\mapsto\mathcal{F}(t,\rho)(x):=\sum_{y\in V}L(x,y)\theta(\rho(x),\rho(y))\Big(\nabla\beta_tU(x,y)+\nabla[\varphi'\circ\rho](x,y)\Big)$$

satisfies a locally Lipschitz condition in $\rho$, uniformly in $t$ on any compact interval $t\in[a,b]$ like in the proof of Proposition 4.3. Let us differentiate with respect to time (recall that $\varphi'(\nu_t(x))+\beta_tU(x)=c(\beta_t)$)

$$\begin{aligned}\partial_t\mathcal{I}(t,\rho_t)&=\sum_x\dot\rho_t(x)\Big(\beta_tU(x)+\varphi'(\rho_t(x))\Big)\ell(x)-\sum_x\dot\nu_t(x)\Big(\beta_tU(x)+\varphi'(\nu_t(x)\Big)\ell(x)\\&\qquad+\dot\beta_t\sum_{x\in V}U(x)(\rho_t(x)-\nu_t(x))\ell(x)\\&=\sum_x\dot\rho_t(x)\Big(\beta_tU(x)+\varphi'(\rho_t(x))\Big)\ell(x)-c(\beta_t)\sum_x\dot\nu_t(x)\ell(x)+\dot\beta_t\sum_{x\in V}U(x)(\rho_t(x)-\nu_t(x))\ell(x)\\&=-\mathcal{G}(t,\rho_t)+\dot\beta_t\sum_{x\in V}U(x)(\rho_t(x)-\nu_t(x))\ell(x)\qquad(5.6)\\&\leq-\lambda\frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_{t,\wedge})}\mathcal{I}(t,\rho_t)+\dot\beta_t\text{osc }U,\qquad(5.7)\end{aligned}$$

where we have used the computation (4.8) in (5.6) for $\mathcal{G}(t,\rho_t)$ and the fact that $\sum_{x\in V}\dot\nu_t(x)\ell(x)=\frac{\partial}{\partial t}(\sum_{x\in V}\nu_t(x)\ell(x))=\frac{\partial}{\partial t}1=0$. By the time-homogeneous version of Gronwall inequality, we obtain $\forall t\in[0,T)$,

$$\begin{aligned}\mathcal{I}(t,\rho_t)&\leq\text{osc}U\int_0^t\dot\beta_s\exp\left(-\int_s^t\lambda\frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_{u,\wedge})}du\right)ds+\exp\left(-\int_0^t\lambda\frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_{s,\wedge})}ds\right)\mathcal{I}(0,\rho_0)\qquad(5.8)\\&\leq\text{osc }U(\beta_t-\beta_0)+\mathcal{I}(0,\rho_0)\\&\leq C_0(\beta_t+1),\end{aligned}$$

where $C_0=\max\{\text{osc}U;\ \mathcal{I}(0,\rho_0)-\text{osc}U\beta_0\}$. Using this, we express $\mathcal{I}(t,\rho_t)$ as

$$\sum_{x\in V}\beta_tU(x)\Big(\rho_t(x)-\nu_t(x)\Big)\ell(x)+\sum_{x\in V}\Big(\varphi(\rho_t(x))-\varphi(\nu_t(x))\Big)\ell(x)\leq C_0(\beta_t+1),$$

so

$$\begin{aligned}\sum_{x\in V}\varphi(\rho_t(x))\ell(x)&\leq C_0(\beta_t+1)+\sum_{x\in V}\varphi(\nu_t(x))\ell(x)-\sum_{x\in V}\beta_tU(x)\Big(\rho_t(x)-\nu_t(x)\Big)\ell(x)\\&\leq C_1(\beta_t+1)+\sum_{x\in V}\varphi(\nu_t(x))\ell(x),\end{aligned}$$

where $C_1=C_0+\text{osc}U$. Observe that from Theorem 5.2, $\nu_t(x)\geq1$ if $x\in\mathcal{M}(U)$ and recall that $m<0$, we have

$$\begin{aligned}&\sum_{x\in V}\varphi(\nu_t(x))\ell(x)\leq\sum_{x\in V}\Big(\varphi_m(\nu_t(x))+\varphi_2(\nu_t(x))\Big)\ell(x)\\&\quad=\sum_{x\in V}\frac{(\nu_t(x)-1)^2}{2}\ell(x)+\sum_{x\in V}\Big(\frac{\nu_t(x)^m-1-m(\nu_t(x)-1)}{m(m-1)}\Big)\ell(x)\\&\quad=\frac12\sum_{x\in V}\nu_t(x)^2\ell(x)-\frac12+\sum_{x\in\mathcal{M}(U)}\frac{\nu_t(x)^m}{m(m-1)}\ell(x)+\sum_{x\notin\mathcal{M}(U)}\frac{\nu_t(x)^m}{m(m-1)}\ell(x)-\frac{1}{m(m-1)}\\&\quad\leq\frac12\nu_{t,\vee}\sum_{x\in V}\nu_t(x)\ell(x)+\sum_{x\in\mathcal{M}(U)}\frac{\nu_t(x)\ell(x)}{m(m-1)}+\sum_{x\notin\mathcal{M}(U)}\frac{\nu_t(x)\ell(x)}{m(m-1)\nu_t(x)^{1-m}}-\left(\frac12+\frac{1}{m(m-1)}\right)\\&\quad\leq\frac12\nu_{t,\vee}+\sum_{x\in\mathcal{M}(U)}\frac{\nu_t(x)\ell(x)}{m(m-1)}+\sum_{x\notin\mathcal{M}(U)}\frac{\nu_t(x)(\beta_t(1-m)\text{osc }U+1)}{m(m-1)}\ell(x)-\left(\frac12+\frac{1}{m(m-1)}\right)\\&\quad\leq\frac12(\nu_{t,\vee}-1)+\beta_t(1-m)\text{osc }U\sum_{x\notin\mathcal{M}(U)}\frac{\nu_t(x)}{m(m-1)}\ell(x)\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{2}(\frac{1}{\ell_\wedge}-1)+\frac{\beta_t \text{osc } U}{-m} \\
&\leq C_2(\beta_t+1),
\end{aligned}
$$

where $\nu_{t,\vee} := \max_{x\in V}\nu_t(x)$, $\ell_\wedge := \min_{x\in V}\ell(x)$, $C_2 := \max\{\frac{1}{2}(\frac{1}{\ell_\wedge}-1), \frac{\text{osc}U}{-m}\}$. Also write $\rho_{t,\wedge} := \min_{x\in V}\rho_t(x)$, and note that $\rho_{t,\wedge}\leq 1$, we have for all $t\in[0,T)$

$$
\begin{aligned}
\ell_\wedge\varphi(\rho_{t,\wedge}) &\leq \sum_{x\in V}\varphi(\nu_t(x))\ell(x)+C_1(\beta_t+1) \\
&\leq C_3(\beta_t+1),
\end{aligned}
$$

where $C_3 = C_1 + C_2$. Since $\varphi(\rho_{t,\wedge}) = \frac{\rho_{t,\wedge}^m-1-m(\rho_{t,\wedge}-1)}{m(m-1)}$, we get for all $t\in[0,T)$

$$
\begin{aligned}
\rho_{t,\wedge}^m &\leq \frac{1}{\ell_\wedge}m(m-1)C_3(\beta_t+1)+1+m(\rho_{t,\wedge}-1) \\
&\leq \frac{1}{\ell_\wedge}m(m-1)C_3(\beta_t+1)+1-m \qquad (\text{since } m\rho_{t,\wedge}<0) \\
&\leq C_4(\beta_t+1),
\end{aligned}
$$

where $C_4 = \frac{1}{\ell_\wedge}m(m-1)C_3+1-m$, and we arrive at

$$
\rho_{t,\wedge}^{-m} \geq \frac{1}{C_4(\beta_t+1)}.
$$

Hence, choosing $K = C_4$ finishes the proof of the lemma. □

**Lemma 5.7.** *There is a constant $I_0$ such that, for all $t\in[0,T)$, $\mathcal{I}(t,\rho_t)\leq I_0$.*

*Proof.* Using the previous lower bound on $\rho_{t,\wedge}$ in the proof of Lemma 5.6, we deduce for $s\leq t<T$ (recall that $\varphi''(r)=r^{m-2}$ if $r\in(0,1)$)

$$
\begin{aligned}
\exp\left(-\lambda\varphi''(1/\ell_\wedge)\int_s^t\frac{1}{\varphi''(\rho_{u,\wedge})}du\right) &= \exp\left(-\lambda\varphi''(1/\ell_\wedge)\int_s^t\rho_{u,\wedge}^{2-m}du\right) \\
&\leq \exp\left(-\lambda\varphi''(1/\ell_\wedge)\int_s^t[C_4(\beta_u+1)]^{-\frac{2-m}{-m}}du\right) \\
&= \exp\left(-\lambda\varphi''(1/\ell_\wedge)C_4^{\frac{2-m}{m}}\int_s^t(t_0+u)^{\alpha\times\frac{2-m}{m}}du\right) \\
&= \exp\left(-C_5\left[(t_0+t)^{1+\frac{(2-m)\alpha}{m}}-(t_0+s)^{1+\frac{(2-m)\alpha}{m}}\right]\right) \\
&= \frac{\exp(C_5(t_0+s)^{1+\frac{(2-m)\alpha}{m}})}{\exp(C_5(t_0+t)^{1+\frac{(2-m)\alpha}{m}})},
\end{aligned}
$$

where $C_5 := \lambda\varphi''(1/\ell_\wedge)C_4^{\frac{2-m}{m}}$. Putting this back to the inequality (5.8), we have $\forall t\in[0,T)$,

$$
\mathcal{I}(t,\rho_t)\leq \text{osc } U\int_0^t\alpha(t_0+s)^{\alpha-1}\frac{\exp(C_5(t_0+s)^{1+\frac{(2-m)\alpha}{m}})}{\exp(C_5(t_0+t)^{1+\frac{(2-m)\alpha}{m}})}ds+\frac{\exp(C_5t_0^{1+\frac{(2-m)\alpha}{m}})}{\exp(C_5(t_0+t)^{1+\frac{(2-m)\alpha}{m}})}\mathcal{I}(0,\rho_0).
$$

Let

$$
\begin{aligned}
p(t) &:= \frac{1}{C_5\left(1+\frac{(2-m)\alpha}{m}\right)}(t_0+t)^{\alpha-1-\frac{(2-m)\alpha}{m}} = \frac{1}{C_5\left(1+\frac{(2-m)\alpha}{m}\right)}(t_0+t)^{\frac{\alpha}{\kappa(m)}-1}, \\
q(t) &:= \exp(C_5(t_0+t)^{1+\frac{(2-m)\alpha}{m}}), \\
\Longrightarrow\ p'(t) &= \frac{1}{C_5\left(1+\frac{(2-m)\alpha}{m}\right)}\left(\frac{\alpha}{\kappa(m)}-1\right)(t_0+t)^{\frac{\alpha}{\kappa(m)}-2}, \\
q'(t) &= C_5\left(1+\frac{(2-m)\alpha}{m}\right)(t_0+t)^{\frac{(2-m)\alpha}{m}}\exp(C_5(t_0+t)^{1+\frac{(2-m)\alpha}{m}}).
\end{aligned}
$$

We have as $t\to+\infty$

$$
\frac{p'(t)}{p(t)}\frac{q(t)}{q'(t)} = \frac{1}{C_5\left(1+\frac{(2-m)\alpha}{m}\right)}\left(\frac{\alpha}{\kappa(m)}-1\right)(t_0+t)^{-1-\frac{(2-m)\alpha}{m}}\to 0
$$

because $-1 - \frac{(2-m)\alpha}{m} \le -1 + \frac{2-m}{-m} \times \frac{-m}{2(1-m)} = \frac{m}{2(1-m)} < 0$ (recall $m < 0$). Note that $p' < 0$ and

$$\lim_{t\to+\infty} p(t)q(t) = +\infty$$
$$\lim_{t\to+\infty} \int_0^t p(s)q'(s)ds = \lim_{t\to+\infty} \Big(p(t)q(t) - p(0)q(0) + \int_0^t [-p'(s)]q(s)ds\Big) = +\infty.$$

By L'Hopital's rule, we have

$$\lim_{t\to+\infty} \frac{\int_0^t p(s)q'(s)ds}{p(t)q(t)} = \lim_{t\to+\infty} \frac{p(t)q'(t)}{p'(t)q(t)+p(t)q'(t)} = \lim_{t\to+\infty} \frac{1}{1+\frac{p'(t)q(t)}{p(t)q'(t)}} = 1.$$

Hence, there exists a constant $K_{\mathcal{I}}$ such that

$$\frac{\int_0^t (t_0+s)^{\alpha-1}\exp(C_5(t_0+s)^{1+\frac{(2-m)\alpha}{m}})ds}{\exp(C_5(t_0+t)^{1+\frac{(2-m)\alpha}{m}})} = \frac{\int_0^t p(s)q'(s)ds}{q(t)} \le K_{\mathcal{I}}p(t). \tag{5.9}$$

The condition $0 < \alpha \le \kappa(m)$ forces $p(t)$ to be constant or goes to 0. In either case, we have (recall that $(t_0+t)^{\frac{\alpha}{\kappa(m)}-1} \le 1$ since $t_0 \ge 1$)

$$\forall t \in [0,T), \qquad \mathcal{I}(t,\rho_t) \le \frac{K_{\mathcal{I}} \operatorname{osc} U}{C_5(1+\frac{(2-m)\alpha}{m})} + \mathcal{I}(0,\rho_0) =: I_0,$$

which ends the proof of the lemma. □

Now we prove Theorem 5.5.

*Proof.* (of Theorem 5.5.) Suppose $T < +\infty$ then $\rho_t$ is going to the boundary of $\mathcal{D}_+(V)$ as $t \to T^-$, but then $\mathcal{I}(t,\rho_t)$ will go to infinity since

$$\mathcal{I}(t,\rho_t) = \mathcal{U}_{\beta_t}(\rho_t) - \mathcal{U}_{\beta_t}(\nu_t),$$

and

$$[0,T] \ni t \mapsto \mathcal{U}_{\beta_t}(\nu_t) = \sum_{x\in V} \beta_t U(x)\nu_t(x)\ell(x) + \sum_{x\in V} \varphi(\nu_t(x))\ell(x)$$

is continuous and bounded, while $\mathcal{U}_{\beta_t}(\rho_t)$ is going to infinity because $\lim_{r\to0^+}\varphi(r) = +\infty$. Therefore, $T = +\infty$ and we have established the existence and uniqueness of a solution $(\rho_t)_{t\ge0}$ of (5.1). Now using (5.3) together with $\varphi = \varphi_{m,2} \ge \varphi_m$, for large $t > 0$ such that $\nu_t(x) \le 1$ for $x \notin \mathcal{M}(U)$, we have

$$\begin{aligned}\forall x \notin \mathcal{M}(U), \qquad \frac{I_0}{\ell(x)} &\ge \varphi(\rho_t(x)) - \varphi(\nu_t(x)) - \varphi'(\nu_t(x))(\rho_t(x)-\nu_t(x)) \\ &\ge \varphi_m(\rho_t(x)) - \varphi_m(\nu_t(x)) - \varphi_m'(\nu_t(x))(\rho_t(x)-\nu_t(x)) \\ &= \nu_t(x)^m \varphi_m\Big(\frac{\rho_t(x)}{\nu_t(x)}\Big)\end{aligned}$$

because

$$\begin{aligned}\varphi_m(t) - \varphi_m(s) - \varphi_m'(s)(t-s) &= \frac{t^m-1-m(t-1)}{m(m-1)} - \frac{s^m-1-m(s-1)}{m(m-1)} - \frac{s^{m-1}-1}{m-1}(t-s) \\ &= \frac{t^m - s^m - m(t-s) - m(s^{m-1}-1)(t-s)}{m(m-1)} \\ &= \frac{t^m - s^m - ms^{m-1}(t-s)}{m(m-1)} \\ &= \frac{s^m\Big((t/s)^m - 1 - m(t/s-1)\Big)}{m(m-1)} \\ &= s^m\varphi_m(t/s).\end{aligned}$$

Thus $\varphi_m\Big(\frac{\rho_t(x)}{\nu_t(x)}\Big) \le \frac{I_0\nu_t(x)^{-m}}{\ell(x)} \to 0$, which implies $\frac{\rho_t(x)}{\nu_t(x)} \to 1$ for $x \notin \mathcal{M}(U)$ as $t \to +\infty$. Since $\nu_t(x) \to 0$ for $x \notin \mathcal{M}(U)$ it follows that $\mu_t[V \setminus \mathcal{M}(U)] \to 0$ or $\mu_t[\mathcal{M}(U)] \to 1$ as $t \to +\infty$, which finishes the proof of Theorem 5.5. □

**Corollary 5.8.** *Given then settings in Theorem 5.5, let $(\rho_t)_{t\ge0}$ be the unique solution to (5.1), then we have*

$$|\mu_t[\mathcal{M}(U)] - 1| = O(t^{-\frac{\alpha}{1-m}}). \tag{5.10}$$

*Moreover, the choice $m = -1$ and $\alpha = \frac{1}{4}$ give the "best" rate for the upperbound*

$$|\mu_t[\mathcal{M}(U)] - 1| = O(t^{-\frac{1}{8}}). \tag{5.11}$$

*Proof.* From the proof of Theorem 5.5, we know that for any $x \notin \mathcal{M}(U)$ we have $\lim \frac{\rho_t(x)}{\nu_t(x)} = 1$. From Theorem 5.2 (ii), we have $\nu_t(x) = O(\beta_t^{\frac{1}{m-1}})$, so our choice of $\beta_t = (t_0+t)^\alpha - 1 \approx t^\alpha$ leads to $\rho_t(x) = O(t^{\frac{\alpha}{m-1}})$, and therefore

$$|\mu_t[\mathcal{M}(U)] - 1| = \mu_t[V \setminus \mathcal{M}(U)] = \sum_{x\notin\mathcal{M}(U)} \ell(x)\rho_t(x) = O(t^{-\frac{\alpha}{1-m}}),$$

which shows (5.10). Since $\alpha \le \kappa(m) = \frac{-m}{2(1-m)}$, we have $\frac{\alpha}{1-m} \le \frac{-m}{2(1-m)^2} \le \frac{1}{8}$ (use $(1+x)^2 \ge 4x$ for $x = -m > 0$). The bound is obtained with the choice $m = -1$ and $\alpha = \kappa(-1) = \frac{1}{4}$ (corresponds to the choice $\beta_t = (t_0+t)^{1/4} - 1 \approx t^{1/4}$), then we have (5.11). □

**Corollary 5.9.** *Given then settings in Theorem 5.5, let $(\rho_t)_{t\ge0}$ be the unique solution to (5.1). Suppose in addition $\alpha < \kappa(m)$ then it holds as $t \to +\infty$ that*

$$\frac{1}{2}\|\rho_t - \nu_t\|^2_{\mathbb{L}^2(\ell)} \le \mathcal{I}(t,\rho_t) \le \frac{\operatorname{osc} U (t_0+t)^{\frac{\alpha}{\kappa(m)}-1}}{C_5(1+\frac{(2-m)\alpha}{m})} + \frac{\exp(C_5 t_0^{1+\frac{(2-m)\alpha}{m}})}{\exp(C_5(t_0+t)^{1+\frac{(2-m)\alpha}{m}})}\mathcal{I}(0,\rho_0) \longrightarrow 0. \tag{5.12}$$

*From (5.12), we have $\mathcal{I}(t,\rho_t) = O(t^{\frac{\alpha}{\kappa(m)}-1})$ but in fact, we have a better rate, that is $\mathcal{I}(t,\rho_t) = O(t^{\frac{2\alpha}{\kappa(m)}-2})$.*

*Proof.* The first inequality in (5.12) follows from Corollary 4.7, while the second is already shown in (5.9). We prove the last statement. Since the second term on the right handside of the second inequality in (5.12) is going to 0 at exponential rate, the first term governs the convergence speed to 0 of $\mathcal{I}(t,\rho_t)$, which is proportional to $t^{\frac{\alpha}{\kappa(m)}-1}$. Denote $\mathcal{I}_t \coloneqq \sqrt{\mathcal{I}(t,\rho_t)}$. From (5.6), we have

$$\begin{aligned}
\partial_t \mathcal{I}(t,\rho_t) &= -\mathcal{G}(t,\rho_t) + \dot\beta_t \sum_{x\in V} U(x)(\rho_t(x) - \nu_t(x))\ell(x) \\
&= -\mathcal{G}(t,\rho_t) + \dot\beta_t \sum_{x\in V} (U(x) - \ell[U])(\rho_t(x) - \nu_t(x))\ell(x) \\
&\le -\mathcal{G}(t,\rho_t) + \dot\beta_t \sqrt{\sum_{x\in V} \ell(x)(U(x)-\ell[U])^2 \times \sum_{x\in V} \ell(x)(\rho_t(x)-\nu_t(x))^2} \\
&= -\mathcal{G}(t,\rho_t) + \dot\beta_t \sqrt{\operatorname{Var}[U]}\|\rho_t - \nu_t\|_{\mathbb{L}^2(\ell)} \\
&\le -\lambda \frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_{t,\wedge})}\mathcal{I}(t,\rho_t) + \dot\beta_t\sqrt{2\operatorname{Var}[U]\mathcal{I}(t,\rho_t)}.
\end{aligned}$$

Dividing the last quantity by $\sqrt{\mathcal{I}(t,\rho_t)}$, we arrive at

$$2\partial_t \mathcal{I}_t \le -\lambda\frac{\varphi''(1/\ell_\wedge)}{\varphi''(\rho_{t,\wedge})}\mathcal{I}_t + \dot\beta_t\sqrt{2\operatorname{Var}[U]},$$

which looks exactly like (5.7) except $\operatorname{osc} U$ is replaced by $\sqrt{2\operatorname{Var}[U]}$ and note that $\sqrt{2\operatorname{Var}[U]} \le \sqrt{2}\operatorname{osc} U$. Hence, following the same lines of proof in Theorem 5.5, we have that $\mathcal{I}_t = O(t^{\frac{\alpha}{\kappa(m)}-1})$ which implies $\mathcal{I}(t,\rho_t) = O(t^{\frac{2\alpha}{\kappa(m)}-2})$. □

## 5.3 Nonlinear Markov interpretation

Similar to Section 4.3, we can reinterpret the dynamic (5.1) in terms of a time-inhomogeneous nonlinear Markov dynamic satisfying

$$\forall t > 0, \qquad \dot\mu_t = \mu_t L_{t,\rho_t} = \mu_t Q_{t,\rho_t}, \tag{5.13}$$

where the generator $L_{t,\rho}$ is defined by

$$\forall t>0,\ \forall \rho \in \mathcal{D}_+(V),\ \forall x \ne y, \quad L_{t,\rho}(x,y) \coloneqq L(x,y)\frac{\theta(\rho(x),\rho(y))}{\rho(x)}\Big(\nabla[\beta_t U + \varphi'\circ\rho](x,y)\Big)_-, \tag{5.14}$$

or more explicitly

$$\forall t>0,\ \forall \rho \in \mathcal{D}_+(V),\ \forall x \ne y,\ L_{t,\rho}(x,y) = L(x,y)\left(\frac{\rho(y)-\rho(x)}{\rho(x)[\varphi'(\rho(y))-\varphi'(\rho(x))]}\beta_t(U(y)-U(x)) + \frac{\rho(y)}{\rho(x)} - 1\right)_-, \tag{5.15}$$

and the generator $Q_{t,\rho}$ is defined by

$$\forall t>0,\ \forall \rho \in \mathcal{D}_+(V),\ \forall x \ne y, \quad Q_{t,\rho}(x,y) \coloneqq L(x,y)\Big(1 + \frac{\rho(y)-\rho(x)}{\rho(x)[\varphi'(\rho(y))-\varphi'(\rho(x))]}\beta_t(U(y)-U(x))_-\Big) \tag{5.16}$$

Observe that $L_{t,\rho} = L_{\beta_t,\rho}$ and $Q_{t,\rho} = Q_{\beta_t,\rho}$ , where $L_{\beta,\rho}$ and $Q_{\beta,\rho}$ are given in (4.22) and (4.28), and therefore the dependence on $t$ of $L_{t,\rho}$ and $Q_{t,\rho}$ is through $\beta_t$ namely on the choice of the temperature schedule.

With access to these Markov interpretations, we can effectively utilize an interacting particle system to approximate a Markov process denoted as $X = (X_t)_{t\geq 0}$, taking values in $V$ and satisfying $\forall t \geq 0$, $\text{Law}(X_t) \equiv \mu_t$, where $\mu_t$ is the unique solution to (5.13) (recall that the existence and uniqueness of $(\mu_t)_{t\geq 0}$ has been established in Section 5.2). We call it Swarm algorithm since the process $X = (X_t)_{t\geq 0}$ evolves and interacts with its law $\mu_t$ at all times $t \geq 0$ in (5.13) through the generator $L_{t,\rho}$ or $Q_{t,\rho}$. An additional rationale behind the chosen name lies in the essence of our algorithm itself. As an approximation method outlined previously, our algorithm relies on the interaction of particles to approximate the behavior of the Markov process $X$. This interaction goes through the closest neighbors as it is observed in biological systems of birds or fish. This inherent similarity provides further validation for the designation "Swarm algorithm". The details of sampling techniques, including the homogeneous case, can be found in the second appendix.

We now delve into investigating the behavior of the Markov generators $t \mapsto Q_{t,\rho_t}$ and $t \mapsto L_{t,\rho_t}$ as $t$ becomes large. Specifically, let us fix two states $x \neq y$ in $V$ and suppose we are at the state $x$. The following observations will shed some lights on the behavior of the values $Q_{t,\rho_t}(x, y)$ and $L_{t,\rho_t}(x, y)$ for large time $t$. If $\min U < U(y) \leq U(x)$ then $(U(y) - U(x))_- = U(x) - U(y)$ and recall from the end of the proof of Theorem 5.5 that $\lim_{t\to\infty} \rho_t(z)/\nu_t(z) = 1$ if $z \notin \mathcal{M}(U)$. Since for $t$ large enough $\rho_t(x), \rho_t(y) < 1$, we have from Theorem 5.2

$$\lim_{t\to\infty} \frac{\rho_t(y)^{m-1}}{\rho_t(x)^{m-1}} = \lim_{t\to\infty} \frac{\nu_t(x)^{1-m}\beta_t}{\nu_t(y)^{1-m}\beta_t} = \frac{U(y) - \min U}{U(x) - \min U},$$

thus,

$$\begin{aligned}\frac{\rho_t(y) - \rho_t(x)}{\rho_t(x)[\varphi'(\rho_t(y)) - \varphi'(\rho_t(x))]}\beta_t(U(y) - U(x))_- &= (m-1)\frac{\frac{\rho_t(y)}{\rho_t(x)} - 1}{\frac{\rho_t(y)^{m-1}}{\rho_t(x)^{m-1}} - 1}\rho_t(x)^{1-m}\beta_t(U(x) - U(y)) \\ &\longrightarrow \left(\frac{U(y) - \min U}{U(x) - \min U}\right)^{\frac{1}{m-1}} - 1 \end{aligned} \tag{5.17}$$

by Theorem 5.2 too and so $Q_{t,\rho_t}(x, y) \to L(x, y)\left(\frac{U(y)-\min U}{U(x)-\min U}\right)^{\frac{1}{m-1}}$ as $t \to \infty$. The same analysis shows $L_{t,\rho_t}(x, y) \to 0$ as $t \to \infty$. If $U(x) > U(y) = \min U$ then the above limit in (5.17) is $+\infty$ because $\rho_t(y)/\rho_t(x) \to +\infty$ and $(m-1)\rho(x)^{1-m}\beta_t(U(x) - U(y)) \to -1$ while the denominator goes to $-1$. Thus if $L(x, y) > 0$, $Q_{t,\rho_t}(x, y) \to +\infty$, which indicates that for large time $t$, the probability of transitioning from $x$ to $y$ in the next jump approaches 1 when $y$ is the unique state in $\mathcal{M}(U)$ that is a neighbor of $x$. If among the neighbors of $x$ there are more than one state in $\mathcal{M}(U)$ then the process jumps very fast to one of them. However, in this case, we cannot ascertain whether $L_{t,\rho_t}(x, y) \to 0$ or $L_{t,\rho_t}(x, y) \to +\infty$ based solely on the asymptotic behavior of $(\rho_t)_{t\geq 0}$ from the proof of Theorem 5.5. In contrast, if we replace $\rho_t$ by $\nu_t$ then it holds that $L_{t,\nu_t} = 0$ for all times $t \geq 0$. Now if $\min U = U(x) < U(y)$, for large time $t$, we have

$$\begin{aligned}\frac{\rho_t(y) - \rho_t(x)}{\rho_t(x)[\varphi'(\rho_t(y)) - \varphi'(\rho_t(x))]}\beta_t(U(y) - U(x)) &= (1-m)\frac{\left(\frac{\rho_t(y)}{\rho_t(x)} - 1\right)\rho_t(y)^{1-m}\beta_t(U(y) - \min U)}{\rho_t(y)^{1-m} - 1 - (1-m)\rho_t(y)^{1-m}(\rho_t(x) - 1)} \\ &\longrightarrow 1\end{aligned}$$

because $\rho_t(y) \to 0$ (so the denominator goes to 0), $\rho_t(x) \to \frac{\zeta_\infty(x)}{\ell(x)} = (\sum_{z\in\mathcal{M}(U)} \ell(z))^{-1}$ by Theorem 5.2 *i*) and *ii*) and $\lim_{t\to+\infty} \frac{\rho_t(y)}{\nu_t(y)} = 1$ so that

$$\lim_{t\to+\infty} \rho_t(y)^{1-m}\beta_t = \lim_{t\to+\infty} \nu_t(y)^{1-m}\beta_t = \frac{1}{(1-m)(U(y) - \min U)}.$$

Thus, $L_{t,\rho_t}(x, y) \to 0$ while $Q_{t,\rho_t}(x, y) \to L(x, y)$ as $t \to \infty$ in this case. In conclusion, the two generators behave differently for large time $t$. Driven by $Q_{t,\rho_t}$, the process can escape outside $\mathcal{M}(U)$ but have tendency to return immediately to $\mathcal{M}(U)$ while it is uncertain if the same phenomena happens for $L_{t,\rho_t}$. Lastly, if the process under the generator $Q_{t,\rho_t}$ is currently outside $\mathcal{M}(U)$, it can move more freely around than the process under the reducible generator $L_{t,\rho_t}$ because $Q_{t,\rho_t}$ is irreducible.

**Remark 5.10.** *In practice, we can employ the following hybrid generator*

$$\forall t \geq 0,\ \forall \rho \in \mathcal{D}_+(V), \qquad A_{t,\rho} := (1 - a_t)L_{t,\rho} + a_tQ_{t,\rho},$$

*where $a : \mathbb{R}_+ \mapsto (0, 1)$ is a continuous function properly chosen. For instance, $a$ can be a constant in $(0, 1)$ or be such that $a_tQ_{t,\rho_t}(x, y) \to +\infty$ if $U(x) > U(y) = \min U$ as $t \to +\infty$. In this manner, it still holds that $\dot{\mu}_t = \mu_tA_{t,\rho_t}$. The first advantage of using this hybrid generator is that it gives us the option of tuning the parameter $a$ in computer simulations, which could enhance performance. Secondly, it preserves a crucial property of $Q_{t,\rho_t}$: if $x \in \mathcal{M}(U)$, $y \notin \mathcal{M}(U)$ with $L(x, y) > 0$ then $A_{t,\rho_t}(x, y) \to 0$ and $A_{t,\rho_t}(y, x) \to +\infty$. Thus the probability that the process jumps from $y$ to a global minimizer from its neighborhood is converging to 1 just like under $Q_{t,\rho_t}$ for large time $t$. Lastly, as we expect $L_{t,\rho_t} \to 0$, the time spent at $x \in \mathcal{M}(U)$ is longer because $|A_{t,\rho_t}(x, x)| \approx a_t|Q_{t,\rho_t}(x, x)| \leq |Q_{t,\rho_t}(x, x)| = |L(x, x)|$, which is helpful for computer simulations. However, it is uncertain if the irreduciblity of $A_{t,\rho_t}$ would bring about better performance in practice at all.*

# 6 Simulation

In this section we present some simulations of our algorithm in both homogeneous and inhomogeneous cases and compare the generators given in Sections 4.3 and 5.3. Consider the state space $V := \{0, 1, ..., 19\}$ and endow $V$ with the irreducible generator $L$ given by

$$\forall i \in V, \qquad L(i, i+1) = L(i, i-1) = 1, \qquad \text{and} \qquad \forall j \notin V \setminus \{i-1, i, i+1\}, \qquad L(i, j) = 0,$$

with the convention that $19 + 1 \equiv 0$ and $0 - 1 \equiv 19$. $L$ then admits the uniform distribution on $V$ as its unique invariant distribution $\ell$. Let

$$\forall x \in \mathbb{R}, \qquad u(x) \coloneqq \frac{x^2}{10} + 2(\cos(3x) + \sin(7x)),$$

we choose the function $U : V \mapsto \mathbb{R}$ to be minimized as follows:

$$\forall i \in V, \qquad U(i) \coloneqq u(-0.6 + i/5.5).$$

The graph of $U$ is given in Figure 1. Observe that $U$ has four local minima and only one global minimum with $i = 7$ is the unique global minimizer. Finally, in both homogeneous and inhomogeneous cases, we use 50 particles in the system.

The details of the codes can be found in this link: https://github.com/nhatthangle/Swarm-Algorithm.git. The strange picks in the following pictures are possibly explained by the fact that the evaluations are made after random jump times and not deterministic fixed times.

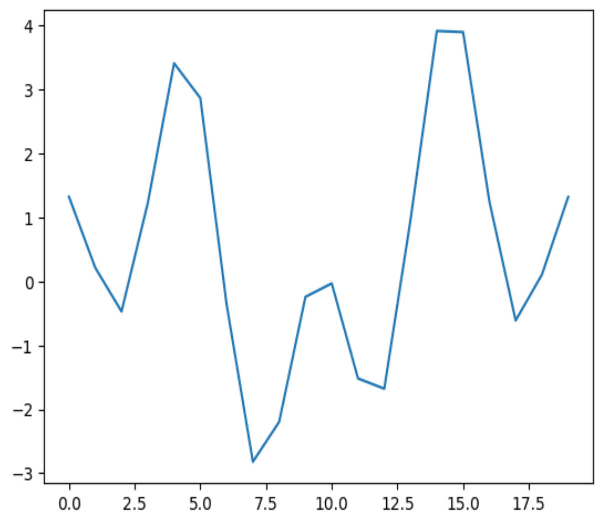


Figure 1: Function $U$

## 6.1 Homogenenous case

We will refer to the generators given in (4.23) and (4.28) as the first generator and the second generator, respectively. Recall that the first generator is not irreducible while the second generator is. To facilitate a comparison of the generators, we select a specific time interval, which we shall choose to be 5 minutes, and allow the particle system to evolve until this interval elapses. Initializing with a uniform distribution, we opt for $\beta = 5$ as our chosen parameter. With this value of $\beta$, the invariant measure $\eta_\beta$ has $\eta_\beta(\{6, 7, 8\}) \approx 0.65$ (recall that $7 = \arg\min_{i \in V} U$).

Figure 2 depicts the graph illustrating the $\mathbb{L}^2(\ell)$-distance between the empirical measure of the particle system and the invariant measure $\eta_\beta$ at each transition (with the x-axis representing the number of transitions), utilizing the first generator. Note that each jump time is random, but we care more about the transitions than time. Meanwhile, Figure 3 presents the analogous graph employing the second generator. Notably, we observe that the distance over time using the second generator exhibits greater fluctuation, attributed to the irreducible nature of this generator.

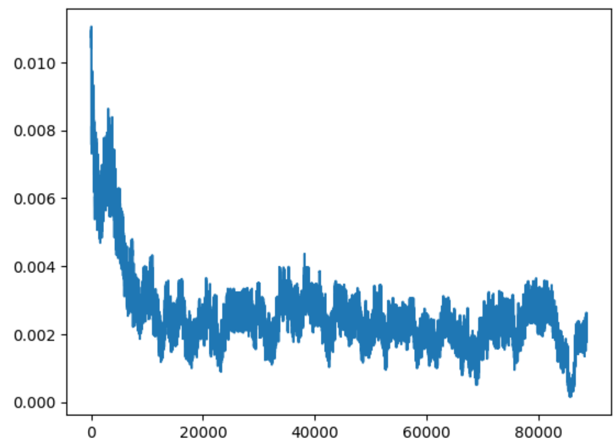


Figure 2: $L^2(\ell)$-distance using first generator

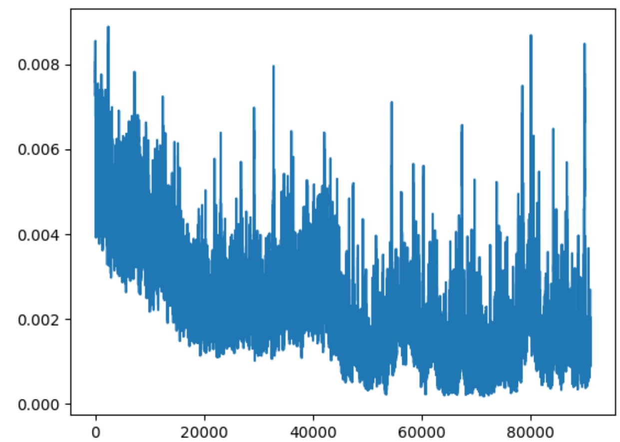


Figure 3: $L^2(\ell)$-distance using second generator

Figures 4, 5 depict the evolution of histograms representing the empirical measures (progressing from left to right), with each picture generated after $1/16$ of the chosen time interval (which amounts to 5 minutes); the initial and final pictures correspond to the start and ending of the simulation, respectively. The two figures look relatively similar, contrary to Figures 2 and 3.

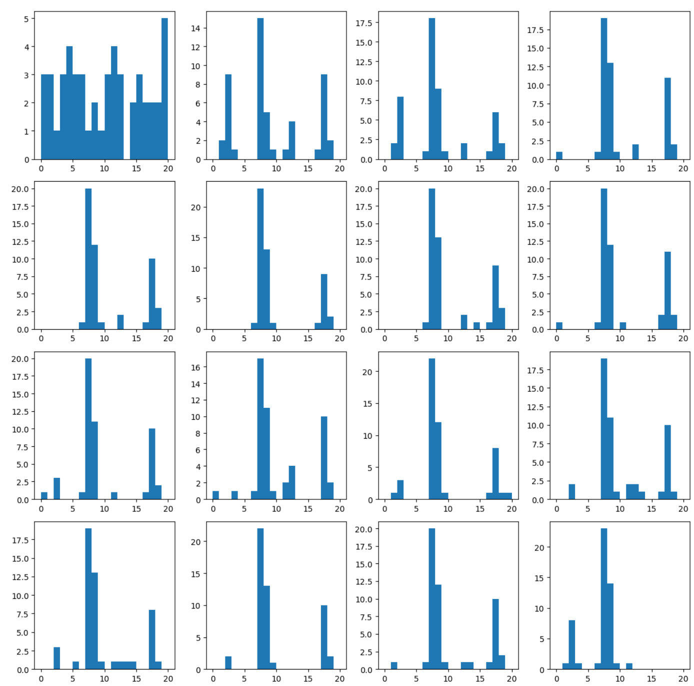

Figure 4: Empirical measure over time (1st generator)

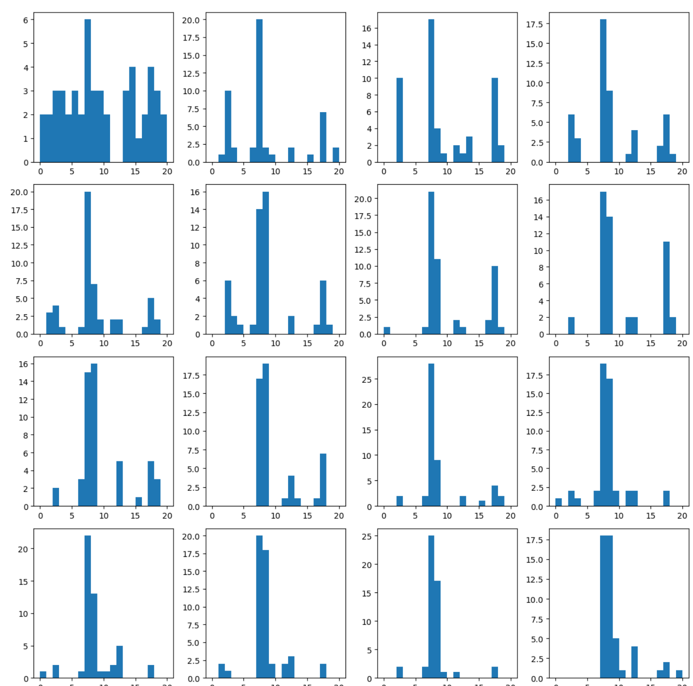

Figure 5: Empirical measure over time (2nd generator)

## 6.2 Inhomogeneous case

In this subsection, we perform the Swarm algorithm introduced in Section 5.3. Once more, we undertake a comparative analysis of the generators (5.15) and (5.16), which we will refer to as the first and the second inhomogeneous families of generators, respectively. We choose $\alpha = 1/4$ and $t_0 = 1$, so that $\beta_t = (1+t)^{\frac{1}{4}} - 1$. We denote the empirical measure by $\hat{\mu}_t$, which is given by

$$\hat{\mu}_t(\cdot) = \sum_{i=1}^{50} \mathbb{1}_{\{X_i(t)\}}(\cdot),$$

where $X(t) = (X_1(t), ..., X_{50}(t))$ is the particle system we are using and for all $i \in V$ and $A \subset V$, $\mathbb{1}_{\{i\}}(A) = 1$ if $i \in A$ and 0 otherwise. We denote $\hat{\rho}_t$ as the empirical density of $\hat{\mu}_t$ with respect to $\ell$. In what follows, we fix a 2-hour period for our simulations.

Figures 6 and 7 illustrate the $\mathbb{L}^2(\ell)$ distance between the empirical density $\hat{\rho}_t$ and the "instantaneous" density $\nu_t$ at each transition (with the x-axis representing the number of transitions) over a 2-hour duration. It is worth noting that within an equivalent time frame, the system governed by the second generator experiences more than eight times the number of transitions compared to that governed by the first generator. This is due to the asymptotic behavior of $L_{t,\rho_t}$ and $Q_{t,\rho_t}$ discussed in Section 5.3. Also, due to the random nature of the particle system, at some point, a particle will move outside $\mathcal{M}(U)$ and causes a few upward spikes as shown in the pictures.

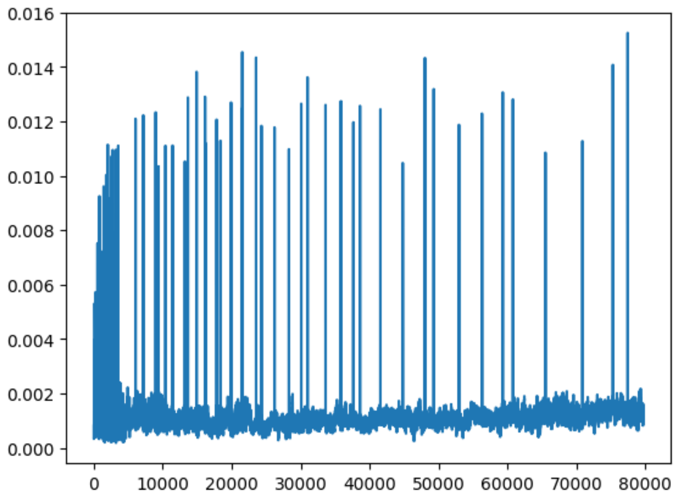


Figure 6: $\|\hat{\rho}_t - \nu_t\|_{\mathbb{L}^2(\ell)}$ (1st generator)

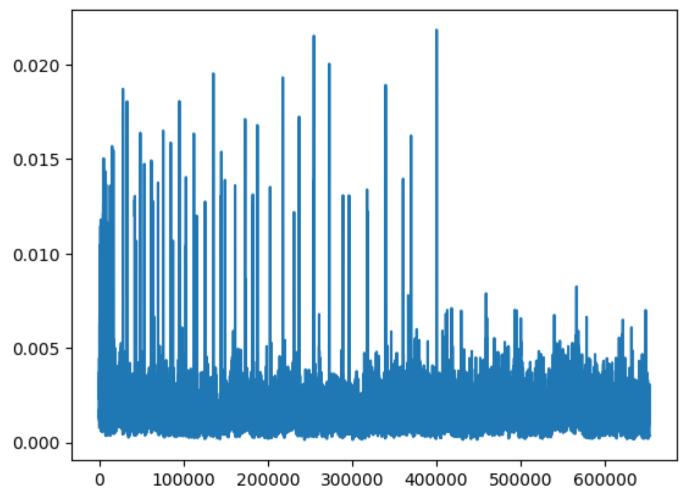


Figure 7: $\|\hat{\rho}_t - \nu_t\|_{\mathbb{L}^2(\ell)}$ (2nd generator)

Figures 8 and 9 show the $\mathbb{L}^2(\ell)$-distance between the empirical density and the density of Dirac measure at state $7 = \arg\min_{i \in V} U$. With a slight abuse of notation, we employ $\mathbb{1}_{\{7\}}$ to represent both the Dirac probability measure and its density with respect to $\ell$. It is evident from these figures that the particle system stemming from the second generator displays more pronounced fluctuations, indicative of its irreducible nature. Conversely, the particle system originating from the first generator appears more stable, with comparatively smaller variance.

Figures 10 and 11 display the histograms illustrating the evolution over time of the empirical distribution $\hat{\mu}_t$. Each histogram, progressing from left to right, represents a snapshot taken after 1/16 of the total time, which equates to 2 hours. The initial frames in both figures portray a sample of 50 particles drawn from the uniform distribution over $V$. The final frames in both figures depict the terminal positions of the particle systems emerging from the two generators.

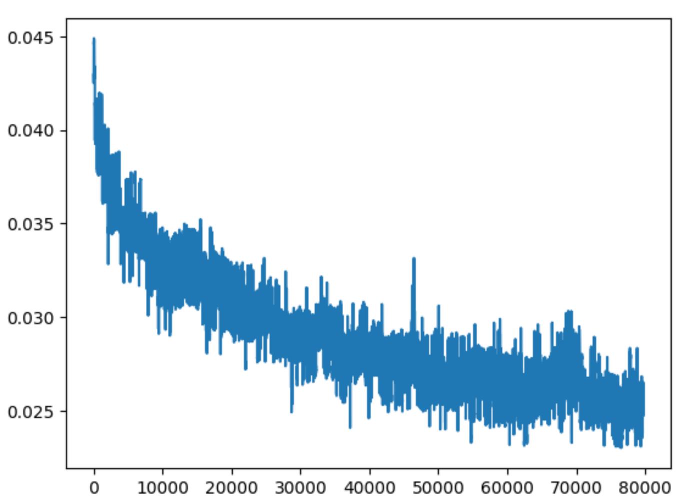


Figure 8: $\|\hat{\rho}_t - \mathbb{1}_{\{7\}}\|_{\mathbb{L}^2(\ell)}$ (1st generator)

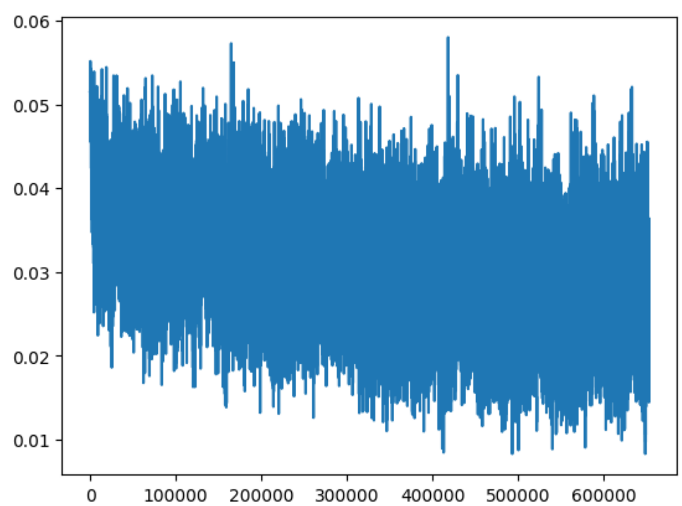


Figure 9: $\|\hat{\rho}_t - \mathbb{1}_{\{7\}}\|_{\mathbb{L}^2(\ell)}$ (2nd generator)

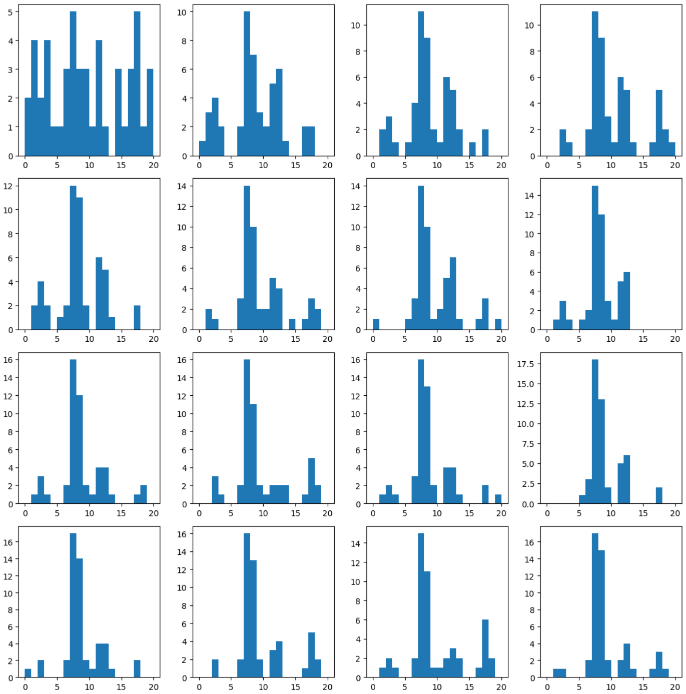

Figure 10: $\|\hat{\rho}_t - \mathbb{1}_{\{7\}}\|_{\mathbb{L}^2(\ell)}$ (1st generator)

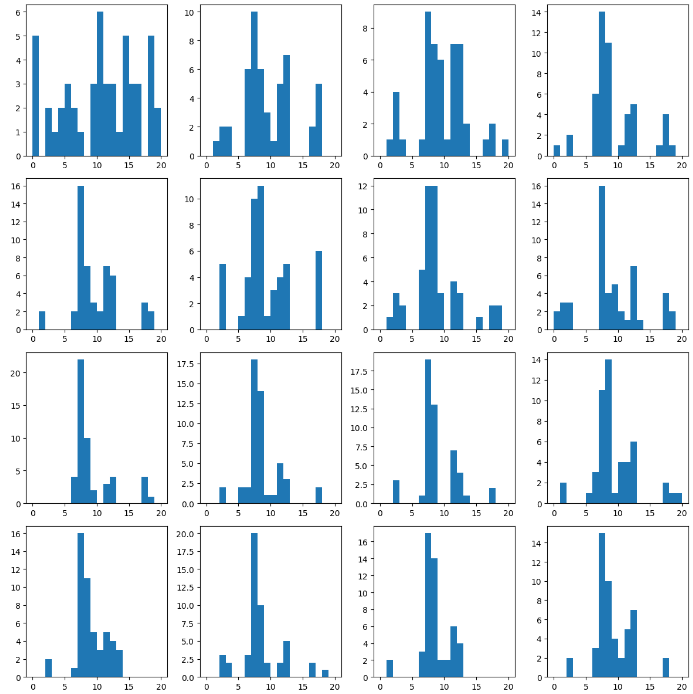

Figure 11: $\|\hat{\rho}_t - \mathbb{1}_{\{7\}}\|_{\mathbb{L}^2(\ell)}$ (2nd generator)

From figures presented above, it appears that the algorithm employing the first generator outperforms the one utilizing the second generator over a fixed time period, particularly concerning variance and convergence characteristics. This is notable despite the fact that the particle system governed by the first generator undergoes fewer transitions. However, it is imperative to note that in practical terms, the number of computations often outweighs the consideration of time alone.

# 7 Conclusion

To globally minimize a function $U$ given on a finite set $V$, we extended it into a penalized functional $\mathcal{U}_\beta$ on the set $\mathcal{P}(V)$ of probability measures on $V$, where $\beta$ is a non-negative parameter. The larger $\beta$ is, the more concentrated on the global minimizers of $U$ is the unique global minimizer $\eta_\beta$ of $\mathcal{U}_\beta$. Another ingredient entering in the functional $\mathcal{U}_\beta$ is a strictly convex function $\varphi : (0, +\infty) \to \mathbb{R}_+$. Following Erbar and Maas [6], we endow $\mathcal{P}(V)$ with a Riemannian structure (except for the regularity), strongly related to $\varphi$. It enables us to consider the gradient descent associated to $\mathcal{U}_\beta$, and as it can be expected, the probability measure-valued dynamical system obtained in this way is converging to $\eta_\beta$ as time goes to infinity. This result leads us to consider a time-inhomogeneous version $(\mu_t)_{t\geq 0}$ of this dynamical system, where the parameter $\beta$ evolves with time, with $\beta_t$ growing to infinity as the time $t$ becomes larger and larger. Our main result gives conditions on the evolution $(\beta_t)_{t\geq 0}$ insuring that for large time $t \geq 0$, $\mu_t$ concentrates on the global minimizers of $U$. The proof is based on a new functional inequality. So while the above considerations are an adaption to the finite setting of the general method described in [4] for the global minimization of Morse functions $U$ on compact Riemannian manifolds $M$, we are able to go much further by relaxing the disappointing geometric restriction imposed in [4] that $M$ should be a circle. Another interesting feature of this approach is that the dynamical system $(\mu_t)_{t\geq 0}$ can be interpreted as the time-marginal distributions of a non-linear and time-inhomogeneous Markov process on $V$, which can thus be approximated by interacting particle systems. The paper ends with an example of such a numerical implementation. We hope to investigate quantitatively the quality of this particle approximation in future works.

# A On Markov-Riemann structures

Our purpose here is to see why the Maas framework [9] based on the introduction of a function $\theta$ is too restrictive to recover the traditional Metropolis algorithm as a gradient descent flow.

Let us extend the Maas framework following [10]. On the finite set $V$, consider $\mathcal{P}_+(V)$ and $\mathcal{L}_\mathrm{i}(V)$, respectively the set of positive probability measures on $V$ and the set of irreducible Markov generators on $V$. Assume we are given a locally Lipschitz (with respect to the total variation) mapping

$$K : \mathcal{P}_+(V) \ni \mu \quad \mapsto \quad K_\mu \in \mathcal{L}_\mathrm{i}(V) \tag{A.1}$$

such that for any $\mu \in \mathcal{P}_+(V)$, $K_\mu$ is reversible with respect to $\mu$.

A form (here we implement Remark 5 of [10], replacing in the terminology "vector fields" by "forms") on $V$ is an anti-symmetric mapping $F : V \times V \ni (x, y) \mapsto F(x, y) \in \mathbb{R}$, i.e. satisfying

$$\forall\ (x,y) \in V \times V, \qquad F(x,y) \quad = \quad -F(y,x)$$

Denote $\mathcal{V}(V)$ the set of forms on $V$. We endow it with the following scalar products, one for each given $\mu \in \mathcal{P}_+(V)$:

$$\forall\ F_1, F_2 \in \mathcal{V}(V), \qquad \langle F_1, F_2 \rangle_{\mu \ltimes K_\mu} \quad \coloneqq \quad \frac{1}{2} \sum_{(x,y)\in V\times V} F_1(x,y) F_2(x,y)\, \mu(x) K_\mu(x,y)$$

(the corresponding Euclidean norm will be denoted $\|\cdot\|_{\mu \ltimes L_\mu}$).

Let $\mathcal{F}(V)$ be the set of real functions defined on $V$. We equally endow it with the family of scalar products corresponding to the $\mathbb{L}^2(\mu)$ space, namely for any $\mu \in \mathcal{P}_+(V)$:

$$\forall\ f_1, f_2 \in \mathcal{F}(V), \qquad \langle f_1, f_2 \rangle_\mu \quad \coloneqq \quad \sum_{x\in V} f_1(x) f_2(x)\, \mu(x)$$

(the corresponding Euclidean norm will be denoted $\|\cdot\|_\mu$).

Consider the mapping d defined by

$$\mathcal{F}(V) \ni f \quad \mapsto \mathrm{d}[f] \coloneqq (f(y) - f(x))_{(x,y)\in V^2} \in \mathcal{V}(V)$$

(it will play the role of the exterior derivative in differential geometry).

The image $\mathrm{d}(\mathcal{F}(V))$ is denoted $\mathcal{E}(V)$ (it corresponds to the set of exact forms in differential geometry and was called the set of gradient fields in [10]).

For any fixed $\mu \in \mathcal{P}_+(V)$, let $\mathrm{d}^*_\mu$ (called the $\mu$-divergence in [10]) be the dual operator to $-\mathrm{d}$ with respect to the Euclidean structures associated to the scalar products $\langle \cdot, \cdot \rangle_\mu$ and $\langle \cdot, \cdot \rangle_{\mu \ltimes K_\mu}$. More explicitly, we compute that

$$\forall\ F \in \mathcal{V}(E), \forall\ x \in V, \qquad \mathrm{d}^*_\mu[F](x) \quad = \quad \sum_{y\in V} K_\mu(x,y) F(x,y)$$

and it appears that

$$\forall\ \mu \in \mathcal{P}_+(V), \qquad K_\mu \quad = \quad \mathrm{d}^*_\mu \circ \mathrm{d}$$

To any $\mu \in \mathcal{P}_+(V)$ and $F \in \mathcal{V}(V)$, we associate the Markov generator $K_{\mu,F}$ given by

$$\forall\ x \neq y \in V, \qquad K_{\mu,F}(x,y) \quad \coloneqq \quad K_\mu(x,y) F_+(x,y)$$

where $F_+(x, y)$ stands for the positive part of $F(x, y)$.

According to (15) in [10], $K_{\mu,F}$ and $\mathrm{d}^*_\mu$ are related through

$$\begin{aligned} \forall\ f \in \mathcal{F}(V), \qquad \mu[K_{\mu,F}[f]] \quad &= \quad -\mu[f \mathrm{d}^*_\mu[F]] \\ &= \quad \langle \mathrm{d}f, F \rangle_{\mu \ltimes K_\mu} \end{aligned} \tag{A.2}$$

To any $F \in \mathcal{V}(V)$ and $\mu \in \mathcal{P}_+(V)$, we associate the semi-flow $(\mathcal{S}_F(\mu, t))_{t\in[0,\tau(F,\mu))}$ solution of

$$\forall\ t \in [0, \tau(F,\mu)), \qquad \dot{\mu}_t \quad = \quad \mu_t K_{\mu_t,F} \tag{A.3}$$

starting with $\mu_0 = \mu$. The time $\tau(F, \mu) > 0$ is assumed to be the explosion time of the above evolution equation, in the sense that $\lim_{t\to\tau(F,\mu)_-} \mu_t$ does not exists in $\mathcal{P}_+(V)$.

**Remark A.1.** *Note that if the mapping defined in* (A.1) *is globally Lipschitz (and in particular bounded), then we get* $\tau(F, \mu) = +\infty$ *for any* $\mu \in \mathcal{P}_+(V)$ *and* $F \in \mathcal{V}(V)$. *It follows that* $\mathcal{S}_F$ *can be seen as a semi-group acting on* $\mathcal{P}_+(V)$, *in the sense that for any* $F \in \mathcal{V}(V)$ *and* $\mu \in \mathcal{P}_+(V)$,

$$\forall\ t, s \geq 0, \qquad \mathcal{S}_F(\mathcal{S}_F(\mu,t), s) \quad = \quad \mathcal{S}_F(\mu, t+s)$$

*Let us explain the interest in our finite setting of* $(\mathcal{S}_F(\delta_x, t))_{t\geq 0}$, *where* $\delta_x$ *is the Dirac mass at* $x \in V$. *Consider* $M$ *a compact Riemannian manifold and let* $\omega$ *be a differential form on* $M$. *The Riemannian structure enables us to transform it into a vector field* $v$. *For any* $x \in M$, *we can consider the flow* $(x(t))_{t\in\mathbb{R}}$ *generated by* $v$, *i.e. the solution of the ordinary differential equation*

$$\begin{cases} x(0) & = \quad x \\ \dot{x}(t) & = \quad v(x(t)), \quad \forall\ t \in \mathbb{R} \end{cases}$$

*Then* $(\mathcal{S}_F(\delta_x, t))_{t\geq 0}$ *for* $x \in V$ *is an analogue of* $(\delta_{x(t)})_{t\geq 0}$ *for* $x \in M$, *when* $\omega$ *is replaced by* $F$. *There are two important differences between these continuous and finite settings. First the flow has to be replaced by a semi-flow, only defined for non-negative times. Secondly, our semi-flow* $(\mathcal{S}_F(\delta_x, t))_{t\geq 0}$ *does not stay in the set of Dirac masses but has to spread, taking values in* $\mathcal{P}_+(V)$ *(the only exception being the case of the zero form).*

In fact we are more interested in the time-inhomogeneous version of (A.3). Let $\mu, \nu \in \mathcal{P}_+(V)$ be given. We denote $\mathcal{D}(\mu,\nu)$ (respectively $\mathcal{D}_{\mathcal{E}(V)}(\mu,\nu)$) the set of continuous paths $F \coloneqq (F(t))_{t\in[0,1]}$ from $[0,1]$ to $\mathcal{V}(V)$ (resp. $\mathcal{E}(V)$) such that the solution of

$$\dot{\mu}_t \quad = \quad \mu_t K_{\mu_t, F(t)} \tag{A.4}$$

starting with $\mu_0 = \mu$ is defined for all $t \in [0,1]$ and satisfies $\mu(1) = \nu$.

**Remark A.2.** *It was shown in [10] that for any $F \coloneqq (F(t))_{t\in[0,1]}$ as above, there exists a unique continuous path $G \coloneqq (G(t))_{t\in[0,1]}$ from $[0,1]$ to $\mathcal{E}(V)$ such that (A.4) is equivalent to*

$$\dot{\mu}_t \quad = \quad \mu_t K_{\mu_t, G(t)}$$

Define

$$D(\mu,\nu) \quad \coloneqq \quad \inf_{F\in\mathcal{D}(\mu,\nu)} \int_0^1 \|F(t)\|_{\mu_t \ltimes K_{\mu_t}} \, dt$$

where $(\mu_t)_{t\in[0,1]}$ is the solution of (A.4) starting from $\mu$. By convention, the above infimum should be $+\infty$ when $\mathcal{D}(\mu,\nu) = \emptyset$. But this does not happen, as it was shown in [10]. Thus, up to a regularity assumption on the mapping $K$ of (A.1), $D$ is a Riemannian metric on $\mathcal{P}_+(V)$. Furthermore from Remark A.2, we have

$$D(\mu,\nu) \quad = \quad \inf_{F\in\mathcal{D}_{\mathcal{E}(V)}(\mu,\nu)} \int_0^1 \|F(t)\|_{\mu_t \ltimes K_{\mu_t}} \, dt$$

This Riemannian structure enables to consider gradient of regular functionals $\mathcal{H}_\varphi$ defined on $\mathcal{P}_+(V)$. Indeed, according to the usual procedure, $\nabla_K \mathcal{H}_\varphi(\mu)$ is defined at any $\mu \in \mathcal{P}_+(V)$ as the unique element from $\mathcal{E}(V)$ such that for any $F \in \mathcal{V}(V)$,

$$\left.\frac{d}{dt}\mathcal{H}_\varphi(\mu_t)\right|_{t=0} \quad = \quad \langle \nabla_K \mathcal{H}_\varphi(\mu), F\rangle_{\mu\ltimes K_\mu} \tag{A.5}$$

where $(\mu_t)_{t\geq 0}$ starts with $\mu_0 = \mu$ and satisfies (A.3). In fact it is sufficient that (A.5) is satisfied for all $F \in \mathcal{E}(V)$, see [10], also for the existence and uniqueness of $\nabla_K \mathcal{H}_\varphi(\mu) \in \mathcal{E}(V)$.

Once this gradient $\nabla_K \mathcal{H}_\varphi$ has been defined from $\mathcal{P}_+(V)$ to $\mathcal{E}(V)$, for any $\mu \in \mathcal{P}_+(V)$, we can consider the gradient descent dynamical system $(\mu_t)_{t\in[0,\tau)}$ starting with $\mu_0 = \mu$ and satisfying

$$\forall\, t \in [0,\tau), \qquad \dot{\mu}(t) \quad = \quad \mu_t K_{\mu_t, -\nabla_K \mathcal{H}_\varphi(\mu_t)} \tag{A.6}$$

where $\tau$ is the explosion time of this $\mathcal{P}_+(V)$-valued flow. The interest of this evolution is that it corresponds to the time-marginal distributions of a non-linear Markov process and thus in principe it can be approximated by interacting particle systems, see e.g. the book of Del Moral [5]. It justifies the consideration of Riemannian structures on $\mathcal{P}_+(V)$ derived from mappings of the form (A.1), called Markov-Riemann structures in [10]. It was checked there that not all Riemannian structures on $\mathcal{P}_+(V)$ are of this form.

Let us give a family of examples that leads to the same traditional Metropolis algorithm. We begin by recalling the latter. The two ingredients are a generator $L \in \mathcal{L}_{\mathrm{i}}(V)$ reversible with respect to a probability $\ell$, as well as a probability $\pi \in \mathcal{P}_+(V)$. The associated Metropolis generator $L_\pi$ is defined by

$$\forall\, x \neq y \in V, \qquad L_\pi(x,y) \quad \coloneqq \quad L(x,y)\left(\frac{\pi(y)\ell(x)}{\pi(x)\ell(y)} \wedge 1\right)$$

Given an initial probability $\mu \in \mathcal{P}_+(V)$, the Metropolis flow $(\mu_t)_{t\in\mathbb{R}_+}$ starting with $\mu_0 = \mu$ is the solution of the linear evolution equation

$$\forall\, t \geq 0, \qquad \dot{\mu}(t) \quad = \quad \mu_t L_\pi \tag{A.7}$$

which is defined for all times and satisfies $\lim_{t\to+\infty} \mu_t = \pi$.

In addition, let us be given a smooth and strictly function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\varphi(1) = 0$ and consider the functional $\mathcal{H}_\varphi$ defined on $\mathcal{P}_+(V)$ via

$$\forall\, \mu \in \mathcal{P}_+(V), \qquad \mathcal{H}_\varphi(\mu) \quad \coloneqq \quad \sum_{x\in V} \varphi\left(\frac{\mu(x)}{\pi(x)}\right)\pi(x)$$

Due to Jensen's inequality and its case of equality, the unique global minimizer of $\mathcal{H}_\varphi$ is $\pi$.

Let us associate to $(L,\pi,\varphi)$ a Markov-Riemann structure on $\mathcal{P}_+(V)$. Consider the function $\theta : \mathbb{R}_+^2 \to \mathbb{R}_+$ defined by

$$\forall\, t,s \in \mathbb{R}_+, \qquad \theta(t,s) \quad \coloneqq \quad \begin{cases} \dfrac{t-s}{\varphi'(t)-\varphi'(s)} & \text{, if } t \neq s \\[2ex] \dfrac{1}{\varphi''(t)} & \text{, if } t = s \end{cases} \tag{A.8}$$

and the mapping (A.1) given by

$$\forall\, \mu \in \mathcal{P}_+(V),\ \forall\, x \neq y \in V, \qquad K_\mu(x,y) \quad \coloneqq \quad \frac{\ell(x)}{\mu(x)} L(x,y) \left(\frac{\pi}{\ell}(x) \wedge \frac{\pi}{\ell}(y)\right) \theta\left(\frac{\mu(x)}{\pi(x)}, \frac{\mu(y)}{\pi(y)}\right) \tag{A.9}$$

Its interest is:

**Proposition A.3.** *Whatever the choice of the convex function $\varphi$, the gradient descent associated to $\mathcal{H}_\varphi$ in the Markov-Riemann structure coming from* (A.9) *is the Metropolis flow* (A.7)*.*

Thus the global minimization of the functional $\mathcal{H}_\varphi$ via a gradient descent in the Riemannian structure coming from (A.9) does not enable us to deduce a new stochastic algorithm. On the other side, it appears that all the above $\mathcal{H}_\varphi$ serve as Liapounov functions for the Metropolis algorithm and their investigations can be performed as part of the general theory of gradient descent and Łojasiewicz' inequalities, see e.g. Blanchet and Bolte [3]. It would be interesting to study more thoroughly the role of the Riemannian metric, for instance what can be said when in (A.8) when one chooses another convex function than $\varphi$? Are there Riemannian structures insuring a faster convergence?

*Proof.* We begin by computing $\nabla_K \mathcal{H}_\varphi(\mu)$ for any given $\mu \in \mathcal{P}_+(V)$. Let $F$ be a form and consider $(\mu_t)_{t\in[0,\tau(F,\mu))}$ the solution of (A.3) starting from $\mu$. We compute

$$\begin{aligned}
\left.\frac{d}{dt}\mathcal{H}_\varphi(\mu_t)\right|_{t=0} &= \left.\sum_{x\in V} \varphi'\left(\frac{\mu_t(x)}{\pi(x)}\right) \frac{d\mu_t}{dt}(x)\right|_{t=0} \\
&= \mu_0\left[K_{\mu_0,F}\left[\varphi'\left(\frac{\mu_0}{\pi}\right)\right]\right] \\
&= \left\langle \mathrm{d}\left[\varphi'\left(\frac{\mu}{\pi}\right)\right], F\right\rangle_{\mu\ltimes K_\mu}
\end{aligned}$$

where we used (A.2), showing that

$$\nabla_K \mathcal{H}_\varphi(\mu) \quad = \quad \mathrm{d}\left[\varphi'\left(\frac{\mu}{\pi}\right)\right]$$

We deduce that for any $\mu \in \mathcal{P}_+(V)$, and any $x, y \in V$,

$$\begin{aligned}
K_{\mu,-\nabla_K\mathcal{H}_\varphi(\mu)}(x,y) &= K_\mu(x,y)\left(-\varphi'\left(\frac{\mu}{\pi}(y)\right) + \varphi'\left(\frac{\mu}{\pi}(x)\right)\right)_+ \\
&= \frac{\ell(x)}{\mu(x)} L(x,y)\left(\frac{\pi}{\ell}(x)\wedge\frac{\pi}{\ell}(y)\right)\theta\left(\frac{\mu(x)}{\pi(x)},\frac{\mu(y)}{\pi(y)}\right)\left(\varphi'\left(\frac{\mu}{\pi}(y)\right) - \varphi'\left(\frac{\mu}{\pi}(x)\right)\right)_- \\
&= \frac{\ell(x)}{\mu(x)} L(x,y)\left(\frac{\pi}{\ell}(x)\wedge\frac{\pi}{\ell}(y)\right)\left(\frac{\mu}{\pi}(y) - \frac{\mu}{\pi}(x)\right)_- \\
&= \frac{\pi(x)}{\mu(x)} L_\pi(x,y)\left(\frac{\mu}{\pi}(y) - \frac{\mu}{\pi}(x)\right)_- \\
&= \frac{\pi(x)}{\mu(x)} L_{\pi,-\mathrm{d}[\mu/\pi]}(x,y)
\end{aligned}$$

It follows that for any test function $f \in \mathcal{F}(V)$,

$$\begin{aligned}
\mu[K_{\mu,-\nabla_K\mathcal{H}_\varphi(\mu)}[f]] &= \mu\left[\frac{\pi}{\mu} L_{\pi,-\mathrm{d}[\mu/\pi]}[f]\right] \\
&= \pi\left[L_{\pi,-\mathrm{d}[\mu/\pi]}[f]\right] \\
&= -\left\langle \mathrm{d}f, \mathrm{d}[\mu/\pi]\right\rangle_{\pi\ltimes L_\pi} \\
&= -\frac{1}{2}\sum_{x,y\in V}(f(y)-f(x))\left(\frac{\mu}{\pi}(y) - \frac{\mu}{\pi}(x)\right)\pi(x)L_\pi(x,y) \\
&= \sum_{x,y\in V}(f(y)-f(x))\frac{\mu}{\pi}(x)\pi(x)L_\pi(x,y) \\
&= \mu[L_\pi[f]]
\end{aligned}$$

where in the second equality we used (A.2), but with the irreducible generator $L_\pi$ reversible with respect to $\pi$ instead of $K_\mu$ and $\mu$.

These computations show that (A.6) reduces to (A.7) as desired. ■

Note that when $\varphi = \varphi_1$ and $\pi$ is the Gibbs distribution given by

$$\forall\, x \in V, \qquad \pi(x) \quad = \quad \frac{\exp(-\beta U(x))}{Z_\beta}\ell(x)$$

where $Z_\beta$ is the normalizing constant, then $\mathcal{H}_\varphi = \mathcal{U}_\beta$, the functional considered in (1.2). Nevertheless in this case we don't recover the non-linear flow investigated in the main text, because the Riemannian structure is different: there the mapping (A.1) is rather given by

$$\forall\, \mu \in \mathcal{P}_+(V),\ \forall\, x \neq y \in V, \qquad K_\mu(x,y) \quad \coloneqq \quad \frac{1}{\mu(x)} L(x,y)\theta\left(\mu(x), \mu(y)\right)$$

# B Sampling finite Markov processes

Our purpose here is to recall how to sample time-homogeneous and time-inhomogeneous Markov processes, as well as sample in an approximate manner time-homogenous and time-inhomogeneous non-linear Markov processes.

## B.1 Time-homogeneous cases

Let $L \coloneqq (L(x,y))_{x,y\in S}$ be a Markov generator on the finite set $S$ and $m_0 \coloneqq (m_0(x))_{x\in S}$ be a probability distribution on $S$. A Markov process $X \coloneqq (X(t))_{t\geq 0}$ with initial law $m_0$ and whose generator is $L$ can be sampled in the following way.

- Sample $X(0)$ according to $m_0$.
- Sample an exponential random variable $E_1$ of parameter 1 and define $\tau_1 \coloneqq E_1/|L(X(0),X(0))|$. For $t \in (0,\tau_1)$, take $X(t) \coloneqq X(0)$. If $L(X(0),X(0)) = 0$, we have $\tau_1 = +\infty$ and the construction stops here. Otherwise we proceed to the next step.
- Sample $X(\tau_1)$ according to the probability $L(X(0),\cdot)/|L(X(0),X(0))|$.
- Sample an exponential random variable $E_2$ of parameter 1 and define $\tau_2 \coloneqq \tau_1 + E_2/|L(X(\tau_1),X(\tau_1))|$. For $t \in (\tau_1,\tau_2)$, take $X(t) \coloneqq X(\tau_1)$. If $L(X(\tau_1),X(\tau_2)) = 0$, we have $\tau_2 = +\infty$ and the construction stop here. Otherwise we proceed to the next step.
- Sample $X(\tau_2)$ according to the probability $L(X(\tau_1),\cdot)/|L(X(\tau_1),X(\tau_1))|$.

In these constructions the samplings are implicitly independent from the previous steps (this will also be so in the following constructions).

The construction proceeds iteratively, to get $\tau_3$, $X(\tau_3)$, $\tau_4$, $X(\tau_4)$, ... If it happens that for some $n \in \mathbb{N}$, $L(X(\tau_n),X(\tau_n)) = 0$, then we get $\tau_{n+1} = +\infty$ and the construction stops there. Otherwise, we obtain an infinite sequence of jump times $(\tau_n)_{n\in\mathbb{N}}$ with

$$\lim_{n\to\infty} \tau_n \quad = \quad +\infty$$

For $t > 0$, let $m_t$ be the law of $X(t)$. It is the solution of the evolution equation (starting from $m_0$)

$$\forall\ t \geq 0, \qquad \partial_t m_t \quad = \quad m_t L$$

(where $m_t$ is seen as a row vector).

## B.2 Time-inhomogeneous cases

The Markov generator $L$ is replaced by a (measurable and locally integrable) family $(L_t)_{t\geq 0}$. A time-inhomogenous Markov process $X \coloneqq (X(t))_{t\geq 0}$ with initial law $m_0$ and whose generators are given by $(L_t)_{t\geq 0}$ can be sampled in the following way.

- Sample $X(0)$ according to $m_0$.
- Sample an exponential random variable $E_1$ of parameter 1 and define $\tau_1$ as

$$\tau_1 \quad \coloneqq \quad \inf\left\{t > 0\ :\ \int_0^t |L_s(X(0),X(0))|\, ds = E_1\right\}$$

  For $t \in (0,\tau_1)$, take $X(t) \coloneqq X(0)$. If $\tau_1 = +\infty$ the construction stops here. Otherwise we proceed to the next step.
- Sample $X(\tau_1)$ according to the probability $L_{\tau_1}(X(0),\cdot)/|L_{\tau_1}((X(0),X(0))|$.
- Sample an exponential random variable $E_2$ of parameter 1 and define $\tau_2$ as

$$\tau_2 \quad \coloneqq \quad \inf\left\{t > 0\ :\ \int_{\tau_1}^{\tau_1+t} |L_s(X(0),X(0))|\, ds = E_2\right\}$$

  For $t \in (\tau_1,\tau_2)$, take $X(t) \coloneqq X(\tau_1)$. If $\tau_2 = +\infty$ the construction stop here. Otherwise we proceed to the next step.
- Sample $X(\tau_2)$ according to the probability $L_{\tau_2}(X(\tau_1),\cdot)/|L_{\tau_2}(X(\tau_1),X(\tau_1))|$.

The construction proceeds iteratively, to get $\tau_3$, $X(\tau_3)$, $\tau_4$, $X(\tau_4)$, ... This procedure may end in a finite number of steps if it happens that for some $n \in \mathbb{N}$ we get $\tau_n = +\infty$. On the contrary when the whole sequence $(\tau_n)_{n\in\mathbb{N}}$ of (finite) jump times is defined, consider

$$\tau_\infty \quad \coloneqq \quad \lim_{n\to\infty} \tau_n$$

definition which is extended to the case where there are only a finite number of jumps by taking $\tau_\infty = +\infty$.

It can be shown that under our local integrability assumption, namely

$$\forall\ t \geq 0, \qquad \int_0^t \max(|L_s(x,x)|\ :\ x \in S)\, ds \quad < \quad +\infty \tag{B.1}$$

we have $\tau_\infty = +\infty$ (a.s.). In particular this is satisfied if the mapping $\mathbb{R}_+ \ni t \mapsto L_t$ is bounded.

If Condition (B.1) is removed (but keeping the mesurability assumption), it may happen that $\tau_\infty < +\infty$, in which case $\tau_\infty$ is called an explosion time. Then $X \coloneqq (X(t))_{t\in[0,\tau_\infty)}$ is only defined on the (random) interval $[0, \tau_\infty)$.

For $t > 0$, let $m_t$ be the law of $X(t)$. It is the solution of the time-inhomogeneous evolution equation (starting from $m_0$)

$$\forall\ t \geq 0, \qquad \partial_t m_t \quad = \quad m_t L_t$$

Note that the above construction coincides with that of Section B.1, in the time-homogeneous cases where $L_t$ does not depend on $t \in \mathbb{R}_+$. In truly time-inhomogeneous cases, one should be able to compute the inverse of the mapping

$$(0, +\infty) \ni t \quad \mapsto \quad \int_s^{s+t} |L_u(x, x)|\, du$$

(where $s \geq 0$ and $x \in S$ are given), which suggests to rather consider simple mappings $\mathbb{R}_+ \ni t \mapsto L_t$.

## B.3 Non-linear cases

Let $\mathcal{P}(S)$ be the set of probability measures on $S$ and $\mathcal{G}(S)$ be the set of Markov generators on $S$. Consider a Lipschitzian mapping

$$\mathcal{P}(S) \ni m \quad \mapsto \quad L_m \in \mathcal{G}(S)$$

Given an initial probability distribution $m_0 \in \mathcal{P}(S)$, we are interested in the solution $(m_t)_{t\geq 0}$ of the non-linear evolution

$$\forall\ t \geq 0, \qquad \partial_t m_t \quad = \quad m_t L_{m_t}$$

It is not easy in general to sample directly a Markov process $X \coloneqq (X(t))_{t\geq 0}$ such that at any time $t \geq 0$, $m_t$ is the law of $X(t)$ and the instantaneous generator is $L_{m_t}$. A probabilistic approximation goes through systems of interacting particles.

Let $N \in \mathbb{N}$ be a number of evolving particles, denoted $X_N \coloneqq (X_{N,l})_{l\in[\![N]\!]} \coloneqq (X_{N,l}(t))_{l\in[\![N]\!],t\geq 0}$. The process $X_N$ is Markovian on $S^N$ and its generator $L_N$ is such that

$$L_N \quad \coloneqq \quad \sum_{l\in[\![N]\!]} L_{N,l}$$

where for any $l \in [\![N]\!]$, $L_{N,l}$ is the Markov generator on $S^N$ given by

$$\forall\ x \coloneqq (x_k)_{k\in[\![N]\!]} \neq y \coloneqq (y_k)_{k\in[\![N]\!]} \in S^N,$$
$$L_{N,l}(x, y) \quad \coloneqq \quad \begin{cases} L_{\eta(x)}(x_l, y_l) & \text{, if } x_k = y_k \text{ for all } k \in [\![N]\!] \setminus \{l\} \\ 0 & \text{, otherwise} \end{cases}$$

where for any $x \coloneqq (x_k)_{k\in[\![N]\!]} \in S^N$, $\eta(x)$ stands for the empirical measure

$$\eta(x) \quad \coloneqq \quad \frac{1}{N} \sum_{l\in[\![N]\!]} \delta_{x_l} \ \in\ \mathcal{P}(S) \tag{B.2}$$

Assume furthermore that the law of $X_N(0)$ is $m_0^{\otimes N}$.

For large $N$, the process $X_{N,1}$ (or any $X_{N,l}$ with $l \in [\![N]\!]$) is an approximation of $X$ and $(\eta(X_N(t)))_{t\geq 0}$ is a random approximation of $(m_t)_{t\geq 0}$.

The Markov process $X_N$ can be sampled as described in Section B.1. Taking into account that if $\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_N$ are $N$ independent exponential random variables of respective parameters $\lambda_1, \lambda_2, ..., \lambda_N \geq 0$, then $\min(\mathcal{E}_l, l \in [\![N]\!])$ is an exponential random variable of parameter $\lambda_1 + \lambda_2 + \cdots + \lambda_N$, we get the following alternative description of the procedure:

- Sample $X_N(0)$ according to $m_0^{\otimes N}$.
- Sample $N$ independent exponential random variables $E_{1,1}, E_{1,2}, ..., E_{1,N}$ of parameter 1, define $\tau_1$ as

$$\tau_1 \quad \coloneqq \quad \min\left(\frac{E_{1,l}}{|L_{\eta(X_N(0))}(X_{N,l}(0), X_{N,l}(0))|}\ :\ l \in [\![N]\!]\right)$$

  and call $I_1$ the index where the minimum is attained (which is a.s. unique if $\tau_1 < +\infty$). For $t \in (0, \tau_1)$, take $X_N(t) \coloneqq X_N(0)$. If $\tau_1 = +\infty$ the construction stops here. Otherwise we proceed to the next step.
- Sample $X_{N,I_1}(\tau_1)$ according to the probability $L_{\eta(X_N(0))}(X_{N,I_1}(0), \cdot)/|L_{\eta(X_N(0))}(X_{N,I_1}(0), X_{N,I_1}(0))|$.
- Keep the other coordinates: for $l \neq I_1$, take $X_{N,l}(\tau_1) \coloneqq X_{N,l}(0)$, this ends the construction of $X_N(\tau_1)$.

- Sample $N$ independent exponential random variables $E_{2,1}, E_{2,2}, ..., E_{2,N}$ of parameter 1, define $\tau_2$ as

$$\tau_2 \quad \coloneqq \quad \tau_1 + \min\left(\frac{E_{1,l}}{|L_{\eta(X_N(\tau_1))}(X_{N,l}(\tau_1), X_{N,l}(\tau_1))|} \,:\, l \in [\![N]\!]\right)$$

  and call $I_2$ the index where the minimum is attained (which is a.s. unique if $\tau_2 < +\infty$). For $t \in (\tau_1, \tau_2)$, take $X_N(t) \coloneqq X_N(\tau_1)$. If $\tau_2 = +\infty$ the construction stops here. Otherwise we proceed to the next step.
- Sample $X_{N,I_2}(\tau_2)$ according to the probability $L_{\eta(X_N(\tau_1))}(X_{N,I_1}(\tau_1), \cdot)/|L_{\eta(X_N(\tau_1))}(X_{N,I_1}(\tau_1), X_{N,I_1}(\tau_1))|$.
- Keep the other coordinates: for $l \neq I_2$, take $X_{N,l}(\tau_2) \coloneqq X_{N,l}(\tau_1)$, this ends the construction of $X_N(\tau_2)$.

The construction proceeds iteratively, to get $\tau_3$, $X_N(\tau_3)$, $\tau_4$, $X_N(\tau_4)$, ... The construction may stop in a finite number of iteration(s), if it happens that $\tau_n = +\infty$ for some $n \in \mathbb{N}$. Otherwise, we obtain an infinite sequence of jump times $(\tau_n)_{n\in\mathbb{N}}$ with

$$\lim_{n\to\infty} \tau_n \quad = \quad +\infty$$

## B.4 Non-linear and time-inhomogeneous cases

Consider a mapping

$$\mathbb{R}_+ \times \mathcal{P}(S) \ni (t, m) \quad \mapsto \quad L_{t,m} \in \mathcal{G}(S)$$

which is locally integrable in the first variable (uniformly with respect to the second) and Lipschitzian in the second variable (locally uniformly with respect to the first variable).

Given an initial probability distribution $m_0 \in \mathcal{P}(S)$, we are interested in the solution $(m_t)_{t\geq 0}$ of the non-linear evolution

$$\forall\, t \geq 0, \qquad \partial_t m_t \quad = \quad m_t L_{t,m_t}$$

It is not easy in general to sample directly a Markov process $X \coloneqq (X(t))_{t\geq 0}$ such that at any time $t \geq 0$, $m_t$ is the law of $X(t)$ and the instantaneous generator is $L_{t,m_t}$. A probabilistic approximation goes through systems of interacting particles.

Let $N \in \mathbb{N}$ be a number of evolving particles, denoted $X_N \coloneqq (X_{N,l})_{l\in[\![N]\!]} \coloneqq (X_{N,l}(t))_{l\in[\![N]\!],t\geq 0}$. The process $X_N$ is Markovian on $S^N$, but time-inhomogeneous, and its instantaneous generator $L_{t,N}$ at time $t \geq 0$ is such that

$$L_{t,N} \quad \coloneqq \quad \sum_{l\in[\![N]\!]} L_{t,N,l}$$

where for any $l \in [\![N]\!]$, $L_{t,N,l}$ is the Markov generator on $S^N$ given by

$$\forall\, x \coloneqq (x_k)_{k\in[\![N]\!]} \neq y \coloneqq (y_k)_{k\in[\![N]\!]} \in S^N, \qquad L_{t,N,l}(x, y) \quad \coloneqq \quad \begin{cases} L_{t,\eta(x)}(x_l, y_l) & \text{, if } x_k = y_k \text{ for all } k \in [\![N]\!] \setminus \{l\} \\ 0 & \text{, otherwise} \end{cases}$$

and where for any $x \coloneqq (x_k)_{k\in[\![N]\!]} \in S^N$, $\eta(x)$ is still given by (B.2).

Assume furthermore that the law of $X_N(0)$ is $m_0^{\otimes N}$.

For large $N$, the process $X_{N,1}$ (or any $X_{N,l}$ with $l \in [\![N]\!]$) is an approximation of $X$ and $(\eta(X_N(t)))_{t\geq 0}$ is a random approximation of $(m_t)_{t\geq 0}$.

The Markov process $X_N$ can be sampled as described in Section B.2. Here is an alternative description of the procedure:

- Sample $X_N(0)$ according to $m_0^{\otimes N}$.
- Sample $N$ independent exponential random variables $E_{1,1}, E_{1,2}, ..., E_{1,N}$ of parameter 1, define $\tau_1$ as

$$\tau_1 \quad \coloneqq \quad \min(\tau_{1,l} \,:\, l \in [\![N]\!]) \tag{B.3}$$

  with

$$\forall\, l \in [\![N]\!], \qquad \tau_{1,l} \quad \coloneqq \quad \inf\left\{t > 0 \,:\, \int_0^t |L_{s,\eta(X_N(0))}(X_{N,l}(0), X_{N,l}(0))|\, ds = E_{1,l}\right\}$$

  and call $I_1$ the index where the minimum is attained in (B.3) (which is a.s. unique if $\tau_1 < +\infty$). For $t \in (0, \tau_1)$, take $X_N(t) \coloneqq X_N(0)$. If $\tau_1 = +\infty$ the construction stops here. Otherwise we proceed to the next step.
- Sample $X_{N,I_1}(\tau_1)$ according to the probability $L_{\tau_1,\eta(X_N(0))}(X_{N,I_1}(0), \cdot)/|L_{\tau_1,\eta(X_N(0))}(X_{N,I_1}(0), X_{N,I_1}(0))|$.
- Keep the other coordinates: for $l \neq I_1$, take $X_{N,l}(\tau_1) \coloneqq X_{N,l}(0)$, this ends the construction of $X_N(\tau_1)$.

- Sample $N$ independent exponential random variables $E_{2,1}, E_{2,2}, ..., E_{2,N}$ of parameter 1, define $\tau_2$ as

$$\tau_2 \quad \coloneqq \quad \min(\tau_{2,l} \,:\, l \in [\![N]\!]) \tag{B.4}$$

with

$$\forall\, l \in [\![N]\!], \qquad \tau_{2,l} \quad \coloneqq \quad \inf\left\{t > 0 \,:\, \int_{\tau_1}^{\tau_1+t} |L_s(X_{N,l}(\tau_1), X_{N,l}(\tau_1))|\, ds = E_{2,l}\right\}$$

and call $I_2$ the index where the minimum is attained in (B.4) (which is a.s. unique if $\tau_2 < +\infty$). For $t \in (\tau_1, \tau_2)$, take $X_N(t) \coloneqq X_N(\tau_1)$. If $\tau_2 = +\infty$ the construction stops here. Otherwise we proceed to the next step.

- Sample $X_{N,I_2}(\tau_2)$ according to the probability $L_{\tau_2,\eta(X_N(\tau_1))}(X_{N,I_2}(\tau_1), \cdot)/|L_{\tau_2,\eta(X_N(\tau_1))}(X_{N,I_2}(\tau_1), X_{N,I_2}(\tau_1))|$.
- Keep the other coordinates: for $l \neq I_2$, take $X_{N,l}(\tau_2) \coloneqq X_{N,l}(\tau_1)$, this ends the construction of $X_N(\tau_2)$.

The construction proceeds iteratively, to get $\tau_3$, $X_N(\tau_3)$, $\tau_4$, $X_N(\tau_4)$, ... The construction may stop in a finite number of iteration(s), if it happens that $\tau_n = +\infty$ for some $n \in \mathbb{N}$. Otherwise, we obtain an infinite sequence of jump times $(\tau_n)_{n\in\mathbb{N}}$ with

$$\lim_{n\to\infty} \tau_n \quad = \quad +\infty$$

# References


[1] E. Aarts and J. Korst, *Simulated annealing and Boltzmann machines. A stochastic approach to combinatorial optimization and neural computing*, Wiley-Intersci. Ser. Discrete Math. Optim., Chichester (UK) etc.: Wiley, 1989.

[2] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[3] A. Blanchet and J. Bolte, *A family of functional inequalities: Łojasiewicz inequalities and displacement convex functions*, J. Funct. Anal., 275 (2018), pp. 1650–1673.

[4] J. Bolte, L. Miclo, and S. Villeneuve, *Swarm gradient dynamics for global optimization: the density case*, Mathematical programming, (2023).

[5] P. Del Moral, *Mean field simulation for Monte Carlo integration*, vol. 126 of Monographs on Statistics and Applied Probability, CRC Press, Boca Raton, FL, 2013.

[6] M. Erbar and J. Maas, *Gradient flow structures for discrete porous medium equations*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 1355–1374.

[7] R. Holley and D. Stroock, *Simulated annealing via Sobolev inequalities*, Commun. Math. Phys., 115 (1988), pp. 553–569.

[8] R. A. Holley, S. Kusuoka, and D. W. Stroock, *Asymptotics of the spectral gap with applications to the theory of simulated annealing*, J. Funct. Anal., 83 (1989), pp. 333–347.

[9] J. Maas, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal., 261 (2011), pp. 2250–2292.

[10] L. Miclo, *On the Helmholtz decomposition for finite Markov processes.* To appear in Séminaire de Probabilités, Springer. Lect. Notes Math., 2024.

[11] B. G. Pachpatte, *Inequalities for differential and integral equations*, vol. 197 of Math. Sci. Eng., San Diego, CA: Academic Press, Inc., 1998.

[12] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated annealing: theory and applications*, vol. 37 of Math. Appl., Dordr., Springer, Dordrecht, 1987.

[13] J. Wolfe, *A proof of Taylor's formula*, The American Mathematical Monthly, 60 (1953), pp. 415–415.



lenhat.thang@tse-fr.eu

miclo@math.cnrs.fr

Toulouse School of Economics,
1, Esplanade de l'université,
31080 Toulouse cedex 6, France.
Institut de Mathématiques de Toulouse,
Université Paul Sabatier, 118, route de Narbonne,
31062 Toulouse cedex 9, France.