Hamiltonian and Lagrangian for the trajectory of the empirical distribution and the empirical measure of Markov processes

Frank Redig\(^{(a)}\), Feijia Wang\(^{(b)}\)

\(^{(a)}\) Delft Institute of Applied Mathematics, Technische Universiteit Delft
Mekelweg 4, 2628 CD Delft, Nederland

\(^{(b)}\) Mathematisch Instituut Universiteit Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

November 12, 2013

Abstract

We compute the Hamiltonian and Lagrangian associated to the large deviations of the trajectory of the empirical distribution for independent Markov processes, and of the empirical measure for translation invariant interacting Markov processes. We treat both the case of jump processes (continuous-time Markov chains and interacting particle systems) as well as diffusion processes. For diffusion processes, the Lagrangian is a quadratic form of the deviation of the trajectory from the Kolmogorov forward equation. In all cases, the Lagrangian can be interpreted as a relative entropy or relative entropy density per unit time.

1 Introduction

The Gibbs formalism (DLR equations, variational principle) plays a crucial role in statistical mechanics of equilibrium systems. Roughly speaking a lattice spin system is called “Gibbs” if it can be described in terms of Boltzmann-Gibbs weights with an interaction such that the total interaction of the spin at the origin with all other spins is finite, uniformly in all configurations. Beyond equilibrium, the appearance of Gibbs measures is less obvious as is illustrated both by the loss of Gibbs property in the course of stochastic dynamics of Glauber type shown in [3] (modeling heating and cooling), as well as by the expected non-Gibbsianness of non-equilibrium stationary states. Outside equilibrium it is natural to think of a Gibbsian description in terms of histories, i.e., using trajectories of the system. In
the context of translation invariant lattice spin systems, one then ends up naturally with a description of the system modulo translations, i.e., on the level of the trajectory of the empirical measure.

In [4] we explained how Gibbs-non-Gibbs transitions in lattice spin systems can be related to a bifurcation phenomenon for the optimal trajectories -in the sense of large deviations- of the empirical measure conditioned to arrive at a given measure at a fixed time $T > 0$. In [12], [14] this idea was developed in the mean-field context, i.e., for the trajectory of the magnetization. The idea is to consider a translation invariant stochastic dynamics, and study the trajectory of the empirical measure. More precisely, one considers the trajectory to arrive at time $T > 0$ at a given empirical measure $\nu$, and at time zero one gives a certain cost $i_0(\mu)$ to each translation invariant starting measure $\mu$. This cost is determined by the choice of the Gibbs measure $\mu_G^0$ from which the dynamics is started, i.e., this cost equals the relative entropy density $s(\mu|\mu_G^0)$ of $\mu$ w.r.t. this initial Gibbs measure. The total cost of a trajectory arriving at time $T$ at $\nu$ is then the sum of $i_0(\mu)$ with the path space cost $\Psi(\mu, \nu, T)$ of the optimal (in the sense of large deviations) trajectory starting from $\mu$ at time zero and arriving at $\nu$ at time $T > 0$. The set of minimizers $\mathcal{O}_T(\nu)$ of this total cost of trajectories arriving at $\nu$ at time $T$ is then the object which should be investigated in order to understand whether or not the initial Gibbs measure, evolved over a time $T$ (denoted by $(\mu_G^0)_T$) is a Gibbs measure.

The precise conjecture is the following. If we have uniqueness for every conditioning of the empirical measure at time $T > 0$, i.e., if $\mathcal{O}_T(\nu)$ is a singleton for every choice of $\nu$, then this should correspond to Gibbsianness of the time-evolved measure $(\mu_G^0)_T$. Conversely, if we have non-uniqueness for a particular conditioning of the empirical measure at time $T$ (a so-called “bad measure”), then we have non-Gibbsianness of the distribution $(\mu_G^0)_T$ at time $T$.

As stated before, the total cost to arrive at time $T > 0$ at a given empirical measure is the sum of the initial cost and a path cost, determined by the Markovian dynamics. The path cost is usually of the form of a Lagrangian action. This means, informally written, that the probability of a trajectory of the empirical measure, where one averages shifts of the point mass of the lattice-spin configuration over the box $[-N, N]^d$, is expected to behave as

$$\mathbb{P}\left((\mathbb{L}_N(\sigma_t))_{0 \leq t \leq T} \approx (\mu_t)_{0 \leq t \leq T}\right) \approx \exp\left(-(2N + 1)^d \int_0^T \Xi(\mu_s, \dot{\mu}_s) \, ds\right)$$

where $\approx$ has to be interpreted in the sense of the large deviation principle on the space of trajectories of translation invariant probability measures.

The Lagrangian $\Xi(\mu_s, \dot{\mu}_s)$ is the object we are after in the present paper. More precisely, we consider two cases in the present paper.
First, in the context of independent Markov processes on a general state space $E$, we study the Lagrangian of the associated to the large deviations of the trajectory of the empirical distribution

$$\mathcal{L}_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i^t}$$

which is a random probability measure on $E$. We compute explicitly the Hamiltonian and provide information on the associated Hamiltonian trajectories for finite state space Markov chains.

For diffusion processes, the Lagrangian is a quadratic form associated to the generator. For Markov chains, the Lagrangian is less explicit (except for two state Markov chains), but can still be characterized as a relative entropy production. The study of the large deviations of the trajectory of the empirical distribution has to be considered as the intermediate step between the magnetization (studied in [12], [14]) and the empirical measure. In particular, for finite state space Markov chains, the empirical distribution is still a finite dimensional object. The corresponding Gibbs-non-Gibbs transitions associated to uniqueness or non-uniqueness of optimal trajectories are then situated in the context of general mean field models, and the notion of Gibbsianness developed there, see e.g. [12].

Second, in the context of translation invariant locally interacting Markov processes, we consider the trajectory of the empirical measure, and compute explicitly the Hamiltonian, both for (interacting) diffusion processes and for jump processes in the class of interacting particle systems [15]. In the context of diffusion processes, the Lagrangian is a quadratic form, while in the context of jump processes (of interacting particle systems type), the Lagrangian is less explicit, but also in that setting a relative entropy production (density) characterization can be given.

This study is a step in the research programme proposed in [4]. Given the Hamiltonians and Lagrangians computed in the present paper, one can then characterize bifurcation phenomena, i.e., non-uniqueness of optimal trajectories for particular choices of initial costs. We leave this problem for future work and focus here on the explicit form of the Hamiltonian and Lagrangian. The full and mathematically complete proof of the validity of the trajectory large deviation principle both for the empirical distribution as well as for the empirical measure will be considered in two future works [10], [11]. This amounts to prove that the Hamiltonian and Lagrangian that we compute here correspond to a unique non-linear semigroup, coinciding with the Nisio semigroup associated to the Lagrangian [6].

Our paper is organized as follows. In section 2 we give a general computation of the Feng-Kurtz Hamiltonian for the trajectory of the empirical distribution. In section 3 we study the case of finite continuous-time Markov chains. In section 4 we consider the case of diffusion processes. In section
we consider the case of interacting Markov processes, both of jump type (interacting particle systems in the spirit of \[15\]) and of diffusion type.

2 The trajectory of the empirical distribution: general case

We consider \(\{X_t : 0 \leq t \leq T\}\) a (Feller) Markov process on a state space \(E\). We assume \(E\) to be a locally compact Polish space. Relevant cases for the present paper are, \(E\) a finite set (finite Markov chains), or \(E = \mathbb{R}^k\) or a compact submanifold of \(\mathbb{R}^k\) (diffusions, diffusions in a domain). The computation of this section is however valid for general \(E\).

We denote by \(Q\) the generator of the process \(\{X_t : 0 \leq t \leq T\}\), i.e.,

\[
Qf(x) = \lim_{t \to 0} \frac{1}{t} (E_x f(X_t) - f(x))
\]

for \(f \in \mathcal{D}(Q)\), and where \(E_x\) denotes expectation in the process starting from \(X_0 = x\). The corresponding semigroup is denoted by \(S_t\), i.e.,

\[
S_tf(x) = E_x f(X_t).
\]

For \(E\) compact \(S_t\) acts on \(\mathcal{C}(E)\), the space of continuous functions, for cases such as \(E = \mathbb{R}^d\), \(S_t\) acts on \(\mathcal{C}_0(E)\), the space of continuous functions vanishing at infinity. We further denote \(\mathcal{C}_b(E)\) the space of bounded continuous functions on \(E\) (of course in the compact case we have \(\mathcal{C}(E) = \mathcal{C}_b(E)\)). For \(\mu\) a finite Borel measure on \(E\) and \(f \in \mathcal{C}(E)\), we denote \(\langle \mu, f \rangle = \int f d\mu\). We denote by \(\mathcal{P}(E)\) the set of probability measures on \(E\).

We now let \(\{X_i^t : 0 \leq t \leq T\}\) be independent copies of the process \(\{X_t : 0 \leq t \leq T\}\) starting at initial points \(X_i^0 = x_i\), and consider the empirical distribution

\[
\mathcal{M}_N(t) = \frac{1}{N} \sum_{i=1}^N \delta_{X_i^t}
\]

This is a random probability measure on \(E\), i.e., a random element of \(\mathcal{P}(E)\), which in the limit \(N \to \infty\) converges to the solution of the Kolmogorov forward equation.

More precisely, if at time zero, \(\mathcal{M}_N(0) \to \mu\) (where \(\mu\) is a probability measure on \(E\)), then at time \(t\), \(\mathcal{M}_N(t) \to \mu_t\), where \(\mu_t\) solves

\[
\frac{d\mu_t}{dt} = Q^* \mu_t
\]

where \(Q^*\) denotes the dual generator defined via

\[
\langle \mu, Qf \rangle = (Q^* \mu, f)
\]

for all \(f\) in the domain of \(Q\).
Indeed, by the law of large numbers, for all \( f \in \mathcal{C}_b(E) \),
\[
\langle \mathcal{M}_N(t), f \rangle = \frac{1}{N} \sum_{i=1}^{N} E_{x_i} f(X_i^t) \to \int f \, d\mu_t
\]
where \( \mu_t = S^*(t)\mu \) denotes the law of \( X_t \) when started initially from \( X_0 \) distributed according to \( \mu \), and where \( \to \) denotes convergence almost surely.

The convergence \( \mathcal{M}_N(t) \to \mu_t \) is a manifestation of the law of large numbers, and therefore it is natural to expect an associated large deviation principle, i.e.,
\[
\mathbb{P}(\{\mathcal{M}_N(t) : 0 \leq t \leq T\} \approx \{\mu_t : 0 \leq t \leq T\}) \approx \exp \left( -N\mathcal{I}(\{\mu_t : 0 \leq t \leq T\}) \right)
\]
(3)

Where \( \approx \) has to be interpreted in the sense of the large deviation principle [2], in a suitable topology on the space of trajectories, i.e., lower bound for open sets \( G \) of trajectories
\[
\liminf_{N \to \infty} \frac{1}{N} \log \mathbb{P}(\{\mathcal{M}_N(t) : 0 \leq t \leq T\} \in G) \geq -\inf_{\gamma \in G} \mathcal{I}(\gamma)
\]
(4)

and for upper bound for closed sets \( F \) of trajectories
\[
\limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P}(\{\mathcal{M}_N(t) : 0 \leq t \leq T\} \in F) \leq -\inf_{\gamma \in F} \mathcal{I}(\gamma)
\]
(5)

By the Markov property, the rate function \( \mathcal{I} \) has the form of a Lagrangian “action”
\[
\mathcal{I}(\{\mu_t : 0 \leq t \leq T\}) = \int_0^T \mathcal{L}(\mu_s, \dot{\mu}_s) \, ds
\]
(6)

where \( \dot{\mu}_s \) denotes the weak derivative of the trajectory at time \( s \), defined via
\[
\langle \dot{\mu}_s, f \rangle = \frac{d}{ds} \langle \mu_s, f \rangle
\]
(7)

Notice that \( \dot{\mu}_s \) is certainly well-defined on functions \( f \) in (7) in the domain of the generator \( Q \), but can possibly not be extended as a finite signed measure on the whole space. We leave the formulation of the precise space on which \( \dot{\mu}_s \) lives to the companion papers [10, 11] where full proofs are given. Our aim here is to compute the Lagrangian \( \mathcal{L} \).

As explained in the introduction, this opens the road to an analysis of bifurcation phenomena related to Gibbs-non-Gibbs transitions, as is done on the level of the magnetization in [12, 13, 14]. The case of the empirical distribution should correspond to Gibbs-non-Gibbs phenomena in the context of mean-field models, where the mean field interaction is a function of possibly several empirical averages (rather than only of the magnetization).
Notice that the expression for the Lagrangian $\mathcal{L}$ is independent of the precise topology (on the space of trajectories of probability measures on $E$) in which the large deviation principle holds. Usually, one then first considers the weakest topology which is the product topology (point-wise convergence at every time), and if one wants to strengthen the topology to e.g. uniform topology, one proves exponential tightness in that topology.

In this paper we focus on the computation of the Lagrangian $\mathcal{L}$ with the scheme of Feng and Kurtz [6], explained e.g. in [14].

In our context this means that we first compute the non-linear generator. To explain this, we need some more notation. First notice that $(X^1_t, X^2_t, \ldots, X^N_t)$ is a Markov process with generator

$$Q_N f(x_1, \ldots, x_n) = \sum_{i=1}^{N} Q_i f$$

where $Q_i$ denotes the generator $Q$ applied to the $i$-th coordinate.

The first computation in the Feng-Kurtz scheme is then the non-linear generator

$$HF(\mu) = \lim_{N \to \infty, \mathcal{M}_N(x_1, \ldots, x_n) \to \mu} \frac{1}{N} \left( e^{-NF(\mathcal{M}_N)} Q_N e^{NF(\mathcal{M}_N)} \right)$$

If $HF$ is of the form $\mathcal{H}(\mu, \nabla F)$, with $\mathcal{H}$ a strictly convex function in the second variable, then we call $\mathcal{H}(\mu, f)$ the Hamiltonian, and the corresponding Lagrangian is then given by the Legendre transform of $\mathcal{H}$:

$$\mathcal{L}(\mu, \alpha) = \sup_{f \in C(E)} \left( \int f \, d\alpha - \mathcal{H}(\mu, f) \right)$$

The interpretation of the “gradient” $\nabla F$ is straightforward when we are in the context of finite state space Markov chains, because the set $\mathcal{P}(E)$ is then finite dimensional. In the context of diffusion processes or more general Markov processes, the gradient will be a (context dependent) functional derivative.

The second variable of the Lagrangian is the velocity variable, which in our context is a signed measure of total mass zero.

The Hamiltonian $\mathcal{H}(\mu, f)$ can be obtained as follows:

$$\mathcal{H}(\mu, f) = \lim_{N \to \infty, \mathcal{M}_N \to \mu} \frac{1}{N} e^{-N(\mathcal{M}_N, f)} Q_N e^{N(\mathcal{M}_N, f)}$$

Notice here that for a given $f \in C(E)$, the function $e^{N(\mathcal{M}_N, f)} = e^{\sum_{i=1}^{N} f(x_i)}$ is a function from $E^N$ to $\mathbb{R}$, on which the generator $Q_N$ can act, i.e., the notation in $Q_N e^{N(\mathcal{M}_N, f)}$ means the generator $Q_N$ acting on that function of $x_1, \ldots, x_N$. 6
The \( \mu \) variable is interpreted as the “position” and the \( f \) variable as the “momentum” (dual to the velocity variable in the Lagrangian formalism).

By the form (8) of the independent generator, we can compute the Hamiltonian \( \mathcal{H}(\mu, f) \):

\[
\mathcal{H}(\mu, f) = \lim_{N \to \infty, \mu_N \to \mu} \frac{1}{N} e^{-N(\mathcal{H}_N,f)} \mathcal{Q}_N e^{N(\mathcal{H}_N,f)}
\]

\[
\mathcal{H}(\mu, f) = \lim_{N \to \infty, \mu_N \to \mu} \frac{1}{N} \sum_{i=1}^{N} e^{-f(x_i)} Q e^{f(x_i)}
\]

\[
\mathcal{H}(\mu, f) = \int e^{-f} Q e^{f} d\mu \quad (12)
\]

Notice that since \( \mathcal{H}(\mu, 0) = 0 \), for the corresponding Lagrangian (10) we have

\[
\mathcal{L}(\mu, \alpha) \geq (\langle \alpha, 0 \rangle - \mathcal{H}(\mu, 0)) = 0
\]

i.e., the Lagrangian is automatically non-negative (as it should be since it is the integrand of the rate function).

3 Finite state space continuous-time Markov chains

3.1 Hamiltonian and Lagrangian

In this case \( E = \{a_1, \ldots, a_k\} \) is a finite set, of which we denote the elements by \( a, b, \ldots \). A function \( f : E \to \mathbb{R} \) is identified with a column of numbers \( f_a, a \in E \), so we will use both notations \( f(a) \), or \( f_a \), idem for probability measures (identified with rows \( \mu_a, a \in E \)).

The continuous-time Markov chain is defined via its transition rates between states \( a, b \in E \), denoted by \( r(a, b) \). The generator is given by

\[
Q f(a) = \sum_{b \in E} r(a, b) (f(b) - f(a)) \quad (13)
\]

For a probability measure \( \mu_a, a \in E \) we then have the Kolmogorov forward equation for the distribution at time \( t \), denoted \( \mu_a(t), a \in E \):

\[
\frac{d\mu_a(t)}{dt} = \sum_b (r(b, a) \mu_b(t) - r(a, b) \mu_a(t)) \quad (14)
\]

with initial condition \( \mu_a(0) = \mu_a, a \in E \).

The Hamiltonian (12) is given by

\[
\mathcal{H}(\mu, f) = \sum_{a, b \in E} \mu_a r(a, b) (e^{f_b - f_a} - 1) \quad (15)
\]
The corresponding Lagrangian is given by
\[
\mathcal{L}(\mu, \alpha) = \sup_f \left( \sum_a f_a \alpha_a - \mathcal{H}(\mu, f) \right)
\]  
(16)

The \( f = f^*(\alpha) \) which realizes the supremum satisfies
\[
\alpha_b = \sum_a \left( \mu_a r(a, b)e^{f^*_b - f^*_a} - \mu_b r(b, a)e^{f^*_a - f^*_b} \right)
\]  
(17)

This leads to
\[
L(\mu, \alpha) = \sum_{a,b} \mu_b r(b, a) \left( f^*_a e^{f^*_b - f^*_a} - f^*_b e^{f^*_a - f^*_b} - (e^{f^*_b - f^*_a} - 1) \right)
\]  
(18)

Defining the “modified” rates
\[
r^*(b, a) = r(b, a)e^{f^*_b - f^*_a}
\]  
the equation (17) reads
\[
\alpha_b = \sum_a \left( \mu_a r^*(a, b) - \mu_b r^*(b, a) \right)
\]  
(19)

which can be interpreted as follows. The modified rates are such that they produce “velocity” (14) equal to \( \alpha \), when started from initial measure \( \mu \). In terms of these modified rates \( r^* \), the Lagrangian reads
\[
L(\mu, \alpha) = \sum_{a,b} \mu_b r^*(b, a) \log \left( \frac{r^*(b, a)}{r(b, a)} \right) - \sum_{a,b} \mu_b (r^*(b, a) - r(b, a))
\]  
(20)

This can be interpreted in terms of relative entropy as follows. The Radon Nikodym derivative of the path space measure of the process with rates \( r^* \) w.r.t. the process with rates \( r \) is given by the Girsanov formula
\[
\frac{d\mathbb{P}^{[0,T]}}{d\mathbb{P}^{[0,T]}} = \exp \left( \sum_{a,b} \left( \log \left( \frac{r^*(b, a)}{r(b, a)} \right) N_{T}^{b,a} - (r^*(b, a) - r(b, a))T \right) \right)
\]  
(21)

where \( N_{T}^{b,a} \) denotes the number of transitions from \( b \) to \( a \) in \([0,T]\). The corresponding relative entropy of these two processes is then given by
\[
s(\mathbb{P}^{[0,T]}_{r^*} | \mathbb{P}^{[0,T]}_{r}) = \int d\mathbb{P}^{[0,T]}_{r^*} \log \left( \frac{d\mathbb{P}^{[0,T]}_{r^*}}{d\mathbb{P}^{[0,T]}_{r}} \right)
\]

Taking the limit \( T \to 0 \) in this expression, starting from initial distribution \( \mu \), we find the connection with the Lagrangian:
\[
\lim_{T \to 0} \frac{1}{T}s(\mathbb{P}^{[0,T]}_{r^*} | \mathbb{P}^{[0,T]}_{r}) = \mathcal{L}(\mu, \alpha)
\]  
(22)
In words this means the following. In order to compute $L(\mu, \alpha)$, we have to consider an auxiliary Markov process with rates that from starting from $\mu$ produce velocity (in the sense of (14)) equal to $\alpha$. The relative entropy of this process w.r.t. the original process in a small interval of time $[0, t]$ is then given by $tL(\mu, \alpha) + O(t^2)$. The Lagrangian $L(\mu, \alpha)$ can thus be viewed as “relative entropy production” needed to force the process to have speed $\alpha$ when started from $\mu$. In particular for the evolution according to the Kolmogorov forward equation: $\alpha = Q^*\mu$, the cost is of course zero, and we indeed have in that case $r^* = r$ and $L(\mu, Q^*\mu) = 0$.

3.2 Hamiltonian trajectories for finite Markov chains

The Hamiltonian (15) has Hamiltonian trajectories given by

$$\dot{f}_a = -\frac{\partial H}{\partial \mu_a} = -\sum_b r(a,b)(e^{f_b} - f_a - 1)$$

$$\dot{\mu}_a = \frac{\partial H}{\partial f_a} = \sum_b \left( \mu_b r(b,a)e^{f_a} - \mu_a r(a,b)e^{f_b} - f_a \right)$$

(23)

The interpretation of the second equation is the following. For a trajectory with “momentum” $f$, the motion of the probability measure is that of a Markov process with rates which are modified according to $f$ via

$$\tilde{r}(a,b) = r(a,b)e^{f_b - f_a}$$

(24)

Indeed, for the modified rates $\tilde{r}$, the second equation of (23) reads simply

$$\dot{\mu}_a = \sum_b \mu_b \tilde{r}(b,a) - \mu_a \tilde{r}(a,b)$$

which is precisely the Kolomogorov forwards equation for the evolution of a probability distribution in a Markov chain with rates $\tilde{r}$.

The equation for the momenta, i.e., the first equation of (23) can be rewritten using the variables $u_a = e^{f_a}, a \in E$:

$$\dot{u}_a = -\sum_b r(a,b)(u_b - u_a) = -(Qu)_a$$

which has the solution

$$u(t) = e^{-tQ}u(0)$$

(25)

The equation for the “position variables” $\mu_a$ is linear and reads

$$\mu(t) = M(u(t))\mu(0)$$

(26)

with $M$ a matrix depending on the solution of the momentum variables, given by

$$M_{a,b}(f) = r(b,a)\frac{u_a}{u_b} - \left( \sum_c r(a,c)\frac{u_c}{u_a} \right) \delta_{a,b}$$

(27)
This matrix has column sums equal to zero, i.e., for all $b \in E$ we have $\sum_a M_{a,b} = 0$, which corresponds to the conservation of mass $\sum_a \mu_a(t) = 1$ in the Hamiltonian evolution. More precisely, the matrix $M_{a,b}$ is precisely the adjoint of the generator corresponding to the modified rates $\tilde{r}$ defined in (24).

We thus conclude that the Hamiltonian trajectories are still Markovian, corresponding with time-dependent rates, steered by the solution of the momentum equation (25).

The solution of (26) is given by

$$\mu(t) = e^{\int_0^t M(u(s))} \mu(0)$$

which means that we have the form of the optimal trajectories, up to the determination of the integration constants given by $u(0)$ and $\mu(0)$. Although the form (28), (29) looks quite explicit, it is not easy in general to find explicit tractable formulas for $\mu(t)$. The action or path-space cost of an optimal trajectory

$$\mathcal{J} = \int_0^T \mathcal{L}(\mu_s, \dot{\mu}_s) \, ds$$

can be rewritten in Hamiltonian formalism as

$$\mathcal{J} (\{\mu(s), f(s) : 0 \leq s \leq T\}) = \left(\sum_a \int_0^T f_a(t) \dot{\mu}_a(t) \, dt\right) - T \mathcal{H}(\mu(0), f(0))$$

This means that in order to find the optimal cost to travel from a starting measure $\mu(0) = \mu$ towards a measure $\mu(T) = \nu$ at time $T$, one has to plug in the solution (28), (26) into the expression (29), and determine the integration constants $\mu(0), f(0)$ by initial and final condition. This leads to a function $\Psi(\mu, \nu, T)$ which is the optimal path cost to travel from $\mu$ to $\nu$ in time $T$. In concrete situations beyond two state Markov chains, in practice, this function is hard to obtain closed formulas for (an issue which we do not want to pursue here).

**Example: Two state symmetric flipping**

To see a concrete example of an explicit solution, we consider the case of two states flipping at rate 1, which corresponds with mean-field independent spin flip dynamics, treated before in [12], [4], [14].

In that case, the state space is given by $E = \{1, 2\}$, the matrix $Q$ is given by

$$Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

and the matrix $M$ of (27) is given by

$$M = \begin{pmatrix} -u_2 & u_1 \\ u_2 & -u_1 \end{pmatrix}$$
where \( u = (u_1, u_2)^T \) satisfies
\[
\dot{u} = -Q u
\] (30)
The equation
\[
\dot{\mu} = M \mu
\]
can be differentiated w.r.t. time once more, which gives
\[
\frac{d^2 \mu}{dt^2} = \left( \frac{dM}{dt} + M^2 \right) \mu
\]
Explicit computation, using (30) then gives
\[
\frac{dM}{dt} + M^2 = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}
\]
which gives the equations
\[
\frac{d^2 \mu_1(t)}{dt^2} = 2\mu_1(t) - 2\mu_2(t) = -2 \frac{d^2 \mu_2(t)}{dt^2}
\]
Putting \( \mu_1 - \mu_2 = x \) we have,
\[
\frac{d^2 x}{dt^2} = 4x
\]
which gives \( x_t = C_1 e^{2t} + C_2 e^{-2t} \) as solutions as found before in [12], or [4].
From this the optimal trajectory starting at \( \mu \) arriving at \( \nu \) and its cost can easily be inferred.

**Remark 3.1.** The fact that \( \frac{dM}{dt} + M^2 \) is a constant matrix is quite exceptional. Even in the two state case, if the rates \( r(1, 2) = \alpha \neq r(2, 1) = \beta \), the matrix \( \frac{dM}{dt} + M^2 \) is not constant and differentiating the equation (26) once more does not lead to further simplification.

### 4 Diffusion processes

#### 4.1 Hamiltonian and Lagrangian

Here we consider the state space \( E = \mathbb{R}^n \) and diffusion processes with generator
\[
Q = \sum_i b_i(x) \partial_i + \sum_{ij} a_{ij}(x) \partial_{ij}^2
\] (31)
where \( \partial_i \) denotes partial derivative w.r.t. \( x_i \). Here \( b_i(x), a_{ij}(x) \) are supposed to be Lipschitz and sufficiently smooth, ensuring uniqueness of the solution of the corresponding stochastic differential equation.

The covariance \( a_{ij}(x) \) is assumed to be a non-degenerate positive definite matrix, i.e., we assume that it is bounded from below by a multiple of the identity.
The Hamiltonian $\mathcal{H}(\mu, f)$ given in (12) can then be computed and this yields:

$$
\mathcal{H}(\mu, f) = e^{-f} Q e^{f} d\mu
= \int \left( Qf + \sum_{ij} \partial_i f(x) \partial_j f(x) a_{ij}(x) \right) d\mu(x) \quad (32)
$$

The measures $\mu$ that we will have to consider are absolutely continuous probability measures w.r.t. Lebesgue measure, $\mu = \mu(x) dx$, where with slight abuse of notation we use the symbol $\mu$ both for the measure and its density.

Although we are in the infinite dimensional context here, because the Hamiltonian is a quadratic form, the corresponding Lagrangian can be obtained more easily than in the previous subsection.

Define the quadratic form

$$
J_{\mu}(f, f) = \int \left( \sum_{ij} \partial_i f(x) \partial_j f(x) a_{ij}(x) \right) d\mu(x) \quad (33)
$$

for $f$ in the domain of $A_{\mu}$.

**Remark 4.1.** Notice that this quadratic form corresponds to the carré du champ operator, i.e.,

$$
J_{\mu}(f, f) = \int \Gamma^{Q}_{2}(f, f) d\mu
$$

where

$$
\Gamma^{Q}_{2}(f, f) = Qf^2 - 2fQf
$$

is the carré du champ operator.

To this quadratic form corresponds a positive self-adjoint operator $A_{\mu}$ (linearly depending on $\mu$) such that

$$
J_{\mu}(f, f) = \frac{1}{2} \langle f, A_{\mu} f \rangle
$$

where $\langle f, g \rangle = \int f(x) g(x) \, dx$ is the usual $L^2$ inner product.

With this notation, the Hamiltonian can be written in the form

$$
\mathcal{H}(\mu, f) = \langle \mu, Qf \rangle + \frac{1}{2} \langle f, A_{\mu} f \rangle = \langle Q^* \mu, f \rangle + \frac{1}{2} \langle f, A_{\mu} f \rangle \quad (34)
$$

Then, the corresponding Lagrangian is computed

$$
\mathcal{L}(\mu, \alpha) = \sup_{f} \left( \langle f, \alpha \rangle - \langle Q^* \mu, f \rangle - \frac{1}{2} \langle f, A_{\mu} f \rangle \right)
= \frac{1}{2} \langle (\alpha - Q^* \mu), A_{\mu}^{-1} (\alpha - Q^* \mu) \rangle \quad (35)
$$
Where \( \langle f, A^{-1} f \rangle \) is to be interpreted in the sense of the spectral theorem, i.e., \( \| A^{-1/2} f \|_2^2 \) for \( f \) in the domain of \( A^{-1/2} \). The Lagrangian is then defined to be infinite when \( \alpha - Q^* \mu \) is not in the domain of \( A^{-1/2} \) (cf. the abstract form of Schilder’s theorem in abstract Wiener spaces see [2]).

We see that the “typical trajectory” which follows the Kolmogorov forward equation has zero cost, since in that case \( \dot{\mu} = \alpha = Q^* \mu \), and hence \( \mathcal{L}(\mu, \alpha) = 0 \), and in general, the Lagrangian is a quadratic expression in the deviation of the trajectory from the Kolmogorov forward equation.

To illustrate this formula, let us consider first the simplest example of the present context, i.e., dimension \( n = 1 \), drift \( b = 0 \), \( a = 1/2 \), corresponding to a one-dimensional Brownian motion. The generator is

\[
Q = \frac{1}{2} \frac{d^2}{dx^2}
\]

\( Q^* = Q \). The quadratic form \( [33] \) reads in this case

\[
J_\mu(f, f) = \frac{1}{2} \int \mu(x)(f')^2 \, dx
\]

and the corresponding operator

\[
A_\mu = \frac{d}{dx} \left( \mu(x) \frac{d}{dx} \right)
\]

which gives

\[
\mathcal{L}(\mu, \alpha) = \frac{1}{2} \left\langle \nabla^{-1} \left( \alpha - \frac{1}{2} \mu'' \right), \frac{1}{\mu} \nabla^{-1} \left( \alpha - \frac{1}{2} \mu'' \right) \right\rangle \quad (36)
\]

The rigorous meaning of the formal expression \( \langle \nabla^{-1} f, \nabla^{-1} g \rangle \) is the inner-product in the space \( H_{-1} \), i.e., \( \langle (\Delta)^{-1/2} f, (\Delta)^{-1/2} g \rangle \), with \( \Delta = \frac{d^2}{dx^2} \), or equivalently the dual space of the Hilbert space \( H_1 \).

**Remark 4.2.** The rate function \( [36] \) has also been obtained in the context of the study of the hydrodynamic limit for independent Brownian particles, in [9]. In general, it is an interesting question to understand the relation between the rate functions which are computed in this paper and the rate functions for deviations of the hydrodynamic limit, see e.g. [8]. For Brownian particles, they coincide because of scale invariance of the Brownian motion.

### 4.2 Relative entropy interpretation

As in the case of finite state space Markov chains, also for diffusion processes, the Lagrangian \( [36] \) can be interpreted in terms of relative entropy. Let us
illustrate this for one-dimensional Brownian motion as the reference process, i.e., \( n = 1 \) with \( b = 0, a = 1/2 \) in (31).

A diffusion process on \( \mathbb{R} \) with drift \( b(x) \) and variance equal to one has the generator
\[
Q_b = b(x) \frac{d}{dx} + \frac{1}{2} \frac{d^2}{dx^2}
\]
if we start this process from a measure \( \mu = \mu(x)dx \), then the infinitesimal change at time zero is given by the adjoint generator working on \( \mu \), i.e.,
\[
\frac{1}{2} \frac{d^2\mu(x)}{dx^2} + \frac{d}{dx}(b(x)\mu(x)) = (Q_b^*\mu)(x)
\] (37)

In particular, for \( \alpha \), a given absolutely continuous signed measure of total mass zero, we can find the drift \( b \) that corresponds to it by solving the equation
\[
\frac{1}{2} \frac{d^2\mu(x)}{dx^2} + \frac{d}{dx}(b(x)\mu(x)) = \alpha(x)
\] (38)

The process with drift \( b \) has a corresponding path space measure on the Wiener space of continuous trajectories denoted by \( \mathbb{P}_b^{[0,T]} \), and we have the Girsanov formula
\[
\frac{d\mathbb{P}_b}{d\mathbb{P}_0} = \exp \left( \int_0^T b(W_s)dW_s - \frac{1}{2} \int_0^T b^2(W_s)ds \right)
\] (39)

relating \( \mathbb{P}_b^{[0,T]} \) with the path space measure of the reference process \( \mathbb{P}_0^{[0,T]} \). The relative entropy of the process with drift \( b \) w.r.t. the zero drift process is thus given by
\[
s(\mathbb{P}_b^{[0,T]}|\mathbb{P}_0^{[0,T]}) = \int d\mathbb{P}_b \log \left( \frac{d\mathbb{P}_b}{d\mathbb{P}_0} \right)
\]
\[
= \mathbb{E}^{(0)} \left( \exp \left( \int_0^T b(W_s)dW_s - \frac{1}{2} \int_0^T b^2(W_s)ds \right) \left( \int_0^T b(W_s)dW_s - \frac{1}{2} \int_0^T b^2(W_s)ds \right) \right)
\] (40)

where the expectation \( \mathbb{E}^{(0)} \) is over the standard Brownian motion, i.e., w.r.t. \( \mathbb{P}_0^{[0,T]} \). Computing then
\[
\lim_{T \to 0} \frac{1}{T} s(\mathbb{P}_b^{[0,T]}|\mathbb{P}_0^{[0,T]})
\]
starting from a distribution \( \mu \) for the reference process, and using
\[
\mathbb{E}^{(0)} \left( \int_0^T b(W_s)dW_s \right)^2 = \mathbb{E}^{(0)} \left( \int_0^T b^2(W_s)ds \right)
\]
gives
\[
\lim_{T \to 0} \frac{1}{T} s(\mathbb{P}_b^{[0,T]}|\mathbb{P}_0^{[0,T]}) = \frac{1}{2} \int b^2(x)\mu(x)dx
\]
which is equal to $\mathcal{L}(\mu, \alpha)$ given in (36), because by (38)

$$
\frac{d}{dx}(b(x)\mu(x)) = \alpha - \frac{1}{2} \mu''(x)
$$

Hence, as in the finite Markov chain case, we see that the Lagrangian can be interpreted as the infinitesimal relative entropy cost to produce a “velocity” $\alpha$ when started from $\mu$. In particular, when $\alpha = Q^*\mu$ this cost is zero, corresponding to the fact that the evolution according to the Kolmogorov forward equation is an optimal trajectory with zero cost.

### 5 Trajectory of the empirical measure

#### 5.1 Context and notation

In the context of translation invariant interacting systems, the empirical distribution is no longer a natural object because of interactions. In particular, the empirical distribution as a function of time is no longer a Markov process. The natural object capturing the essential information about the time evolution, modulo translations is then given by the empirical measure.

In order to describe this setting, we need some more notation. For $N \in \mathbb{N}$ we denote $V_N = \{-N, \ldots, N\}^d$ and denote by $T^d_N$ the $d$-dimensional torus, i.e., $V_N$ endowed with addition modulo $2N + 1$.

We will consider translation invariant systems on this torus which for large $N$ have to be thought of as approximations of an infinite interacting system where the individual components live on the lattice $\mathbb{Z}^d$.

The configuration space is $\Omega_N = E^{T^d_N}$, where $E$, the single-site space, is a locally compact Polish space. Further we denote $\Omega = E^{\mathbb{Z}^d}$ the state space of the infinite volume process. As in the previous sections, we mostly consider $E$ or a finite set (interacting particle systems) or $E = \mathbb{R}^n$ (or a submanifold of $\mathbb{R}^n$) (interacting diffusion processes). Elements of $\Omega_N$ are denoted $\sigma, \eta, \xi, \ldots$, and for $\sigma \in \Omega_N$, $i \in T^d_N$, $\sigma_i$ denotes the value of the configuration at site $i$. On $T^d_N$ we have the addition modulo $2N + 1$, and correspondingly, the shift $\tau_i$ defined on $\Omega_N$ via

$$(\tau_i(\sigma))_j = \sigma_{j+i} \tag{41}$$

on functions $f : \Omega_N \to \mathbb{R}$ via $\tau_i f(\sigma) = f(\tau_i \sigma)$, and on probability measures via $\int f d(\tau_i \mu) = \int \tau_i f d\mu$. If $A$ is a linear operator on functions $f : \Omega_N \to \mathbb{R}$ then we define its shift over $i$ to be $\tau_i A\tau_{-i}$, and an operator is called translation invariant if for all $i$, $\tau_i A\tau_{-i} = Q$. A measure is translation invariant if $\tau_i \mu = \mu$. Natural translation invariant measures on $\Omega_N$ are obtained by periodizing translation invariant measures on $\Omega$, i.e., starting from $\sigma$ distributed according to a translation invariant measure on $\Omega$, we consider $\sigma_i^N = \sigma_i, i \in V_N$, periodically extended to the whole lattice. Conversely,
if we have a probability measure $\mu_N$ on $\Omega_N$ we naturally associate to it a probability measure on the infinite configuration space $\Omega$, namely we consider the periodic extension of a configuration drawn from $\mu_N$ to the whole lattice $\mathbb{Z}^d$. This justifies the fact that with slight abuse of notation we will use sometimes the same symbol $\mu_N$ for a translation invariant measure on $\Omega_N$ as well as for the corresponding translation invariant measure on $\Omega$. We denote by $\mathbb{P}_{\text{inv}}(\Omega)$ the set of translation invariant probability measures on $\Omega$.

A function $f : \Omega \to \mathbb{R}$ is called local if it depends on a finite number of coordinates, i.e., if there exists a (minimal) finite set $D_f$, called the dependence set of $f$ such that for all $\sigma, \eta \in \Omega$: $f(\sigma_{D_f \setminus D_f}) = f(\sigma)$, i.e., the value of the function is not influenced by changing the configuration outside $D_f$. Obviously, a local function $f : \Omega \to \mathbb{R}$ can be thought of as being a function $f : \Omega_N \to \mathbb{R}$ as well, for $N$ large enough such that $V_N \supset D_f$.

The translation $\tau_i f$ of local function is obviously local, with dependence set $D_{\tau_i f} = D_f + i = \{x + i : x \in D_f\}$.

An linear operator (possibly unbounded) $A : \mathcal{D}(A) \subset \mathcal{C}(\Omega) \to \mathcal{C}(\Omega)$ is local if it acts only on $\eta_i$, for $i$ a finite set $D = D_A \subset \mathbb{Z}^d$ of vertices. A local operator acts naturally on functions $f : \Omega_N \to \mathbb{R}$ for $N$ large enough, such that $D_A$ is contained in $V_N$.

### 5.2 Translation invariant sequence of local generators

**Definition 5.1.** A translation invariant sequence of local generators is defined to be a a sequence of generators of the form $L_N = \sum_{i \in \mathbb{T}_N^d} \tau_i Q \tau_{-i}$, with $Q$ a local generator, such that the corresponding infinite volume generator $L = \sum_{i \in \mathbb{Z}^d} \tau_i Q \tau_{-i}$ is well-defined, corresponds to a unique Markov process on $\Omega$, and has a subset of local functions as a core. The generator $Q$ is then called the “source generator”.

As a consequence, the corresponding processes $\{\sigma_{N,t} : t \geq 0\}$ converge weakly in path space to the infinite volume process $\{\sigma_t : 0 \leq t \leq T\}$ with generator $L$. Moreover, for the associated semigroups we have that $S_t^N f \to S_t f$ uniformly as $N \to \infty$, for all local functions $f$.

Let us give some examples in order to make this concept more concrete.

1. **Independent Markov processes.** For $Q$ a generator of a Markov process on $E$, we define

   $$L_N = \sum_{i \in \mathbb{T}_N^d} \tau_i Q_0 \tau_{-i}$$

   Where $Q_0$ is the operator $Q$ working on the variable $\sigma_0$. Under the process with this generator $L_N$ different components evolve independently, as copies of the process with generator $Q$. 
2. **Spin-flip dynamics.** $E$ is finite set (e.g. $E = \{-1, 1\}$ for Ising spins), $\theta : E \to E$ a bijection such that $\theta(a) \neq a$ for all $a \in E$. Furthermore, a local function $r : \Omega \to \mathbb{R}^+$, with dependence set containing the origin, is given. The local generator is then defined $Qf(\sigma) = r(\sigma)(f(\theta_0 \sigma) - f(\sigma))$, where $\theta_0$ means applying $\theta$ to the coordinate $\sigma_0$ and leaving all other coordinates unchanged (similarly we denote $\theta_i$). The corresponding sequence of generators is then given by

$$L_N f(\sigma) = \sum_i ((\tau_i Q\tau_{-i}) f)(\sigma) = \sum_{i \in T^N} r(\tau_i \sigma)(f(\theta_i \sigma) - f(\sigma))$$

3. **Interacting diffusions.** For $E = \mathbb{R}$ and for a finite set $D \subset \mathbb{Z}^d$, we consider the local generator

$$Qf(\sigma) = \left( \sum_{j \in D} \frac{\partial V(\sigma_D)}{\partial \sigma_j} \frac{\partial f}{\partial \sigma_j} \right) + \frac{1}{2} \frac{\partial^2}{\partial \sigma_0^2}$$

and the corresponding

$$L_N = \sum_i \tau_i Q\tau_{-i} f$$

This represents a system of diffusions, interacting via the potential $V$. E.g. for a nearest neighbor potential $V : \mathbb{R} \to \mathbb{R}$ in $d = 1$, the full generator has the form

$$\sum_i V'(|\sigma_i - \sigma_{i-1}|) \left( \frac{\partial}{\partial \sigma_i} - \frac{\partial}{\partial \sigma_{i-1}} \right) + \frac{1}{2} \frac{\partial^2}{\partial \sigma_i^2}$$

corresponding to $D = \{0, 1\}$, $V(\sigma_D) = V(|\sigma_1 - \sigma_0|)$. The core for the generator of the infinite volume process is the set of local smooth ($C^\infty_0$) test functions.

4. **Local interacting particle systems.** $E$ is a finite set. For finite subsets $D_\alpha \subset \mathbb{Z}^d$, a collection of $T_\alpha : E^{D_\alpha} \to E^{D_\alpha}$ $\alpha \in \{1, \ldots, k\}$, and corresponding rates $c(\alpha, \sigma)$ we consider the local generator

$$Qf(\sigma) = \sum_\alpha c(\alpha, \sigma)(f(T_\alpha \sigma_{D_\alpha} \sigma_{D_\alpha}) - f(\sigma))$$

the corresponding local generators then include of course the previous spin-flip case but also translation invariant spin-exchange (Kawasaki) dynamics, combination of spin-flip and spin-exchange, etc.
5. **Local averaging.** For $0 \in D \subset \mathbb{Z}^d$ finite, and $m_D$ a probability measure on $E^D$, consider

$$Qf(\sigma) = r(\sigma) \int (f(\sigma'_{D\sigma'}) - f(\sigma)) m_D(d\sigma'_D)$$

with $r$ a local function. In words, this means that with rate $r$, the configuration inside $D$ is replaced by its average over the measure $m_D$. Examples of this class are the KMP model (a model of heat conduction) [7], or more generally the thermalized BEP process [1].

5.3 **Trajectory of the empirical measure**

For a configuration $\sigma \in \Omega_N$, its corresponding empirical measure is defined by

$$\mathbb{L}_N(\sigma) = \frac{1}{|T_N^d|} \sum_{i \in \mathbb{Z}^d} \delta_{\tau_i \sigma}$$

This is a translation invariant probability measure on $\Omega_N$, capturing all information about $\sigma$, modulo translations.

For a configuration on the full lattice, $\sigma \in \Omega$, with a slight abuse of notation we also denote

$$\mathbb{L}_N(\sigma) = \frac{1}{|T_d^N|} \sum_{i \in \mathbb{Z}_d^N} \delta_{\tau_i \sigma_N}$$

where $\sigma^N$ is the periodized configuration obtained from $\sigma$.

If $\mu$ is a probability measure on $\Omega$, which is ergodic under translations, then, by the Birkhoff ergodic theorem, with $\mu$ probability one

$$\mathbb{L}_N(\sigma) \to \mu$$
as $N \to \infty$, and where “$\to$” means weak convergence.

If $(\mathbb{L}_N)_N$ is a translation invariant sequence of local generators, then we have the associated Markov processes $\sigma_{N,t}$ with semigroups $S^N_t = e^{t \mathbb{L}_N}$. For a probability measure $\mu$ on $\Omega$, let us denote $\mu_t$ to be the distribution at time $t > 0$ in the infinite volume process $\{\sigma_t : t \geq 0\}$, started at initial state distributed according to $\mu$. By locality of the generator $\mathbb{L}$, for $\mu$ ergodic, we have that $\mu_t$ is ergodic as well and hence

$$\mathbb{L}_N(\sigma_t) \to \mu_t$$

weakly, with probability one. Therefore the random trajectory of translation invariant probability measures $\{\mathbb{L}_N(\sigma_t) : 0 \leq t \leq T\}$ converges, as $N \to \infty$ to the deterministic trajectory $\{\mu_t : 0 \leq t \leq T\}$. This convergence of a random $\mathcal{P}_{inv}(\Omega)$-valued trajectory to a deterministic $\mathcal{P}_{inv}(\Omega)$-valued
trajectory can be thought of as a law of large numbers (in an infinite di-
mensional space), and therefore it is natural to ask for an associated large
development principle. For spin-flip dynamics, this was studied in \[4\]. Here
we treat the general case of a translation invariant sequence of local generators. This will naturally lead to a local non-linear operator \(K_Q\) associated to
the local source generator \(Q\), which is in the present context of interacting
systems the analogue of the non-linear operator \(e^{-f_Qe^f}\) in section 2.

More precisely, we want to identify the “path space Lagrangian” (which
is in this section is denoted by \(\Xi\)) such that, in \(\mathcal{P}_{inv}(\Omega)\):

\[
P \left( \{ \mathbb{L}_N(\sigma_t) : 0 \leq t \leq T \} \approx \{ \nu_t : 0 \leq t \leq T \} \right) \approx \exp \left( -\left| T^N_d \right| \int_0^T \Xi(\nu_t, \dot{\nu}_t) \right)
\]

The Lagrangian is now a function of a translation invariant probability mea-
sure and a translation invariant signed measure of total mass zero (or a more
general distribution on the space of functions belonging to the domain of the
generator), and as before, \(\approx\) has to be interpreted in the sense of the large
development principle, in this case, on the space of trajectories with values in
the set \(\mathcal{P}_{inv}(\Omega)\) of translation invariant probability measures on \(\Omega\).

5.4 The Hamiltonian

In this section we compute the Feng-Kurtz Hamiltonian. This Hamiltonian
is now a function from \(\mathcal{C}(\Omega) \times \mathcal{P}_{inv}(\Omega)\) to \(\mathbb{R}\), where the first variable has
to be thought of the “position” variable, whereas the second variable as the
“momentum” variable. The Hamiltonian is then defined as the limit

\[
\mathcal{H}(\mu, f) = \lim_{N \to \infty} \frac{1}{|T^N_d|} \left( e^{-\left| T^N_d \right| \langle \mathbb{L}_N(\sigma) \rangle} \mathcal{L}_N e^{\left| T^N_d \right| \langle \mathbb{L}_N(\sigma) \rangle} \right)
\]

(44)

Note that \(\left| T^N_d \right| \langle \mathbb{L}_N(\sigma) \rangle = \sum_{i \in T^N_d} \tau_i f(\sigma)\).

For the computation of (44), we assume \(f\) to be a local function. Because
the source generator \(Q\) is local we have, that \(Q(\tau_k f) = 0\) for all \(k\) outside
the set \(D(Q, f) = \{ k : D_Q \cap D_f + k \neq \emptyset \} \). Therefore, for \(\Lambda \subset \mathbb{Z}^d\) finite,

\[
Q \left( \prod_{i \in \Lambda} \tau_i e^f \right) = \left( \prod_{i \in \Lambda \setminus D(Q, f)} \tau_i e^f \right) Q \left( \prod_{i \in \Lambda \cap D(Q, f)} \tau_i e^f \right)
\]

(45)
Use (45) to compute
\[
\mathcal{H}(\mu, f) = \lim_{N \to \infty} \mathcal{L}_N(\sigma) - \sum_i \tau_i f \sum_j \tau_j \left( Q e^{\sum_i \tau_i - j f} \right)
\]

We can now introduce the non-linear operator associated to the “source” generator \( Q \), working on local functions \( f \):
\[
\mathcal{K}_Q f = e^{-\sum_{k \in \mathbb{Z}^d} \tau_k f} Q e^{\sum_{k \in \mathbb{Z}^d} \tau_k f}
\]
Using this notation, we obtain from (46)
\[
\mathcal{H}(\mu, f) = \int \mathcal{K}_Q(f) d\mu
\]
This Hamiltonian has to be thought of as the analogue of (12) in the present context.

**Remark 5.1.** Note that we can write, informally,
\[
\mathcal{K}_Q f = e^{-\sum_{k \in \mathbb{Z}^d} \tau_k f} Q e^{\sum_{k \in \mathbb{Z}^d} \tau_k f}
\]
since the terms \( k \notin D(f, Q) \) “cancel”. This is of course not rigorous because the infinite sum \( \sum_{k \in \mathbb{Z}^d} \tau_k f \) does not make sense, but in the “same way” as for a formal infinite volume Hamiltonian, where only local energy differences are well defined. The advantage of this formal representation is that we clearly see that \( \mathcal{H} \) is a translation invariant operator, i.e., \( \mathcal{H}(f) = \mathcal{H}(\tau_i f) \), and as a consequence, the Hamiltonian \( \mathcal{H}(\mu, f) \) is translation invariant as well, both in the measure and in the function, i.e.,
\[
\mathcal{H}(\tau_k \mu, \tau_r f) = \mathcal{H}(\mu, f)
\]
for all \( k, r \in \mathbb{Z}^d \). Another advantage is that one clearly sees the analogy with the corresponding formula for the empirical distribution (12).

The corresponding Lagrangian is then found by Legendre transformation, i.e.,
\[
\Xi(\mu, \dot{\mu}) = \sup_{f \in \mathcal{C}(\Omega)} \left( \int f \, d\dot{\mu} - \mathcal{H}(\mu, f) \right)
\]
where $\dot{\mu}$ denotes a translation invariant signed measure of total mass zero, and $\mu$ a translation invariant probability measure on $\Omega$.

In general, an explicit expression for $\Xi$ cannot be obtained easily. In the examples below we will compute $\Xi$ quite explicitly for diffusion processes and show a relative entropy interpretation of $\Xi$ in the context of interacting particle systems (analogue of finite Markov chains in the previous section) and in the context of interacting diffusions.

5.5 Interacting particle systems: the Lagrangian

We now compute $\mathcal{K}_Q$ for some of the examples discussed before, starting with interacting particle systems. The local generator is of the form

$$Qf = \sum_\alpha r_\alpha (T_\alpha f - f)$$

where $T_\alpha$ are local transformations, which change coordinates only in a finite set $D_\alpha$ containing the origin. This gives

$$\mathcal{K}_Q f = \sum_\alpha r_\alpha \left( e^{D_\alpha(f)} - 1 \right)$$

where the operator $D_\alpha$ is defined by

$$D_\alpha f = \sum_{k \in \mathbb{Z}^d} (T_\alpha(\tau_k f) - \tau_k f)$$

Notice that the sum is in fact a finite sum since $f$ is local, and the transformation $T_\alpha$ is local as well. Let us now first zoom in into two familiar examples.

a) **Independent spin-flip.** For $E = \{-1,1\}$, and for a single transformation $T_\sigma = \sigma^0$ (spin-flip), we get

$$D(f) = \sum_{k \in -D_f} (\tau_k f(\sigma^0) - \tau_k f)$$

for the special functions $f(\sigma) = H_A(\sigma) = \prod_{i \in A} \sigma_i$ we get

$$D(H_A) = \sum_{k \in -A} -2H_{A+k}$$

as we found before in [4].

b) **Symmetric exclusion process.** For $E = \{0,1\}$, $d = 1$ and $T(\sigma) = \sigma^{01}$, where $\sigma^{01}$ denotes exchange of the values at site 0 and 1, i.e.,

$$(\sigma^{01})_j = \sigma_1 \delta_{j,0} + \sigma_0 \delta_{j,1} + \sigma_j (1 - \delta_{j,0} - \delta_{j,1}).$$

We have

$$D(f)(\eta) = \sum_{k:k+D_f \cap \{0,1\} \neq \emptyset} f(\tau_k(\eta^{01})) - f(\eta)$$
Notice that for \( f = \eta_0 \) we find only two terms contributing to \( D(f) \):

\[
D(f) = ((\eta_0^0))_0 - (\eta_0) + ((\tau_1(\eta_0^0)))_0 - (\tau_1(\eta_0))_0
\]

which corresponds to the fact that the density of particles is conserved in this process.

Returning to the general case now, the Lagrangian associated with (50) is

\[
\Xi(\mu, \dot{\mu}) = \sup_{f \in C(\Omega)} \left( \int f d\dot{\mu} - \int \left( \sum_{\alpha} r_{\alpha} \left( e^{D_{\alpha}(f)} - 1 \right) \right) d\mu \right)
\]

This expression is reminiscent of (16) in section 3 (empirical distribution for finite Markov chains). Indeed, a similar relative entropy interpretation of this expression can be given. We will describe this rather informally, as it is quite analogous to the Girsanov formula computation of the section on finite Markov chains. First we note that for a translation invariant measure \( \mu \), its “derivative at time zero” \( L^* \mu \) is formally given by

\[
(L^* \mu)(\sigma) = \sum_i \sum_{\alpha} (r_{\alpha}(\tau_i \sigma)) \mu(\tau_i T_{\alpha} \tau_i - \sigma) - r_{\alpha}(\tau_i \sigma) \mu(\tau_i \sigma)
\]

This object is to be interpreted as working on local functions, i.e., as a distribution.

Suppose now we consider modified rates \( \tilde{r}_{\alpha}(\sigma) = r_{\alpha}(\sigma) e^{f(\sigma) - f(T_{\alpha}(\sigma))} \), and the associated modified local generator \( \tilde{Q} = \sum_{\alpha} \tilde{r}_{\alpha}(T_{\alpha} - I) \), i.e., the same transformations are applied but now with new rates. Then for a given translation invariant signed measure of total mass zero, we look for those modified rates, i.e., choice of \( f \), such that with the starting measure \( \mu \) they produce “derivative at time zero” equal to \( \dot{\mu} \), i.e.,

\[
\dot{\mu}(\sigma) = \sum_i \sum_{\alpha} (\tilde{r}_{\alpha}(\tau_i \sigma)) \mu(\tau_i T_{\alpha} \tau_i - \sigma) - \tilde{r}_{\alpha}(\tau_i \sigma) \mu(\tau_i \sigma)
\]

The Radon Nikodym derivative of the path space measure of the finite-volume process \( (\mathbb{T}_d^N)^{(\eta)} \) with rates \( \tilde{r} \) w.r.t. the process with rates \( r \) is given by the Girsanov formula:

\[
\frac{dP^{[0,T],N_r}}{dP^{[0,T],N_{\tilde{r}}}} = \exp \left( \sum_{i \in \mathbb{T}_d^N} \sum_{\alpha} \left( \int_0^T \log \frac{\tilde{r}_{\alpha}^i(\sigma_s)}{r_{\alpha}^i(\sigma_s)} dN_{\alpha}^i \right) ds - \int_0^T (\tilde{r}_{\alpha}^i(\sigma_s) - r_{\alpha}^i(\sigma_s)) ds \right)
\]

where \( r_{\alpha}^i \), resp. \( \tilde{r}_{\alpha}^i \) denote the rate to flip from \( \sigma \) to \( \tau_i T_{\alpha} \tau_i - \sigma \), i.e., to apply the transformation \( T_{\alpha} \) around the lattice site \( i \), and \( N_{\alpha}^i \) the corresponding counting process counting how many transitions \( \sigma \) to \( \tau_i T_{\alpha} \tau_i - \sigma \) have happened in the time interval \([0,t]\).
We then find, analogously to (22) that the Lagrangian is equal to the limit
\[ \Xi(\mu, \dot{\mu}) = \lim_{T \to 0} \frac{1}{T} \lim_{N \to \infty} \frac{1}{T^N} s([\mathbb{F}_T, \mathbb{F}_N]) \]
i.e., the analogue of (22), replacing relative entropy by relative entropy density.

5.6 Diffusion processes: the Lagrangian.

For diffusion processes, let us start with the simplest case of independent Brownian motions in \( d = 1 \). The general case will be analogous, but the quadratic forms appearing there will be less explicit. The source generator \( Q \) is thus given by
\[ Qf(\sigma) = \frac{1}{2} \partial_0^2 f(\sigma) \]
where we abbreviated \( \partial_0 \) to denote the partial derivative w.r.t. \( \sigma_0 \). As a consequence, for a local function \( f \):
\[ \mathcal{H}_Qf = \sum_k Q(\tau_k f) + \frac{1}{2} \left( \sum_k \partial_0(\tau_k f) \right)^2 \]
and, reminding that the full generator is the sum of shifts of \( Q \), we have
\[ \mathcal{H}(\mu, f) = \int \mathcal{H}_Q f d\mu = \int \mathcal{L} f d\mu + \mathcal{J}_\mu(f, f) \]
\[ = \langle f, \mathcal{L}^* \mu \rangle + \mathcal{J}_\mu(f, f) \tag{52} \]
where
\[ \mathcal{J}_\mu(f, f) = \frac{1}{2} \int \left( \sum_k \partial_0(\tau_k f) \right)^2 d\mu \]
is a \( \mu \) dependent quadratic form. This quadratic form is the analogue of (33). Hence, for the Lagrangian we have
\[ \Xi(\mu, \dot{\mu}) = \sup_f \left( \langle \dot{\mu} - \mathcal{L}^* \mu, f \rangle - \mathcal{J}_\mu(f, f) \right) = \mathcal{J}_{\mu}^*(\dot{\mu} - \mathcal{L}^* \mu, \dot{\mu} - \mathcal{L}^* \mu) \]
where \( \mathcal{J}_{\mu}^* \) is a dual quadratic form defined via
\[ \mathcal{J}_{\mu}^*(\nu, \nu) = \sup_f \langle \nu, f \rangle - \mathcal{J}_\mu(f, f) \tag{53} \]
for \( \nu \) a signed measure of total mass zero. Notice that this indeed defines a quadratic form because for \( \lambda > 0 \) (and similarly for \( \lambda < 0 \))
\[ \mathcal{J}_{\mu}^*(\lambda \nu, \lambda \nu) = \sup_f (\lambda \langle \nu, f \rangle - \mathcal{J}_\mu(f, f)) \]
\[ = \lambda^2 \sup_f (\langle \nu, f/\lambda \rangle - \mathcal{J}_\mu(f/\lambda, f/\lambda)) \]
\[ = \lambda^2 \mathcal{J}_{\mu}^*(\nu, \nu) \]
We see in particular that $\Xi(\mu, \dot{\mu})$ is zero for a solution of the Kolmogorov forward equation, i.e., if $\dot{\mu} = \mathcal{L}^*\mu$, which shows also in the present context that the Markovian evolution of the distribution $\mu$ is an optimal zero cost trajectory.

Finally, let us turn to the general diffusion case. We split $Q$, the source generator, into a first order part and a second order part:

$$Q = Q_1 + Q_2$$

where $Q_2$ contains all second order derivatives (variance part of the diffusion), $Q_1$ all first order derivatives (drift part). To $Q_2$ is then associated the quadratic form

$$\mathcal{J}_\mu^Q(f, f) = \int \left( e^{-\sum_k \tau_k f} Q_2 e^{\sum_k \tau_k f} - Q_2 \left( \sum_k \tau_k f \right) \right) d\mu$$

(54)

Notice that this corresponds to the operator carré du champ associated to $Q_2$, i.e.,

$$\mathcal{J}_\mu^Q(f, f) = \Gamma_{Q_2}^2 \left[ \sum_k \tau_k f \right]$$

The Lagrangian is then given by

$$\Xi(\mu, \dot{\mu}) = (\mathcal{J}_\mu^Q)^*(\dot{\mu} - \mathcal{L}^*\mu, \dot{\mu} - \mathcal{L}^*\mu)$$

(55)

where $(\mathcal{J}_\mu^Q)^*$ is the dual quadratic form of $\mathcal{J}^Q$ (as in (53)).

References

[1] G. Carinci, C. Giardinà, C. Giberti, F. Redig, Duality for Stochastic Models of Transport J. Stat. Phys. 152, 657-697, (2013).

[2] A. Dembo and O. Zeitouni. Large deviations techniques and applications. Second edition, Springer Verlag, (2010).

[3] A.C.D. van Enter, R. Fernández, F. den Hollander and F. Redig, Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures, Comm. Math. Phys. 226, 101-130 (2002).

[4] A.C.D. van Enter, R. Fernández, F. den Hollander and F. Redig, A large-deviation view on dynamical Gibbs-non-Gibbs transitions, Mosc. Math. J. 10, 687-711 (2010).

[5] A.C.D. van Enter, C. Külske, Alex A. Opoku and W.M. Ruszel, Gibbs-non-Gibbs properties for $n$-vector lattice and mean field models, preprint http://arxiv.org/abs/0812.1751 (2008).
[6] J. Feng and T.G. Kurtz, *Large Deviations for Stochastic Processes*, American Mathematical Society, Providence RI, (2006).

[7] C. Kipnis, C. Marchioro and E. Presutti, Heat flow in an exactly solvable model, J. Stat. Phys. 27, 65-74, (1982).

[8] C. Kipnis, C. Landim, *Scaling limits of interacting particle systems*, Springer, Berlin, (1999).

[9] C. Kipnis, S. Olla, Large Deviation from the Hydrodynamic Limit for a System of Independent Brownian Particles, Stochastics and Stochastic Reports 33, 17-25, (1990).

[10] R. Kraaij, Large deviations of the trajectory of empirical distributions of Markov processes, preprint 2013.

[11] R. Kraaij, F. Redig, Large deviations of the trajectory of empirical measure of locally interacting Markov processes, in preparation.

[12] C. Külske and V. Ermolaev, Low-temperature dynamics of the Curie-Weiss model: periodic orbits, multiple histories, and loss of Gibbsianness. J. Stat. Phys. 141, 727-756, (2010).

[13] C. Külske and F. Redig, Loss without recovery of Gibbsianness during diffusion of continuous spins. Prob. Theory Rel. Fields 135, 428-456 (2006).

[14] F. Redig and F. Wang, Gibbs-non-Gibbs transitions via large deviations: computable examples. J. Stat. Phys., 147, 1094-1112, (2012).

[15] T. Liggett, *Interacting Particle Systems*, Springer Berlin, 1985