ALEA, Lat. Am. J. Probab. Math. Stat. 12 (1), 1–23 (2015)

Donsker-Varadhan asymptotics for degenerate jump Markov processes

Giada Basile and Lorenzo Bertini

Dipartimento di Matematica, Università di Roma 'La Sapienza'
P.le Aldo Moro 5, 00185 Roma, Italy
E-mail address: basile@mat.uniroma1.it
E-mail address: bertini@mat.uniroma1.it

Abstract. We consider a class of continuous time Markov chains on a compact metric space that admit an invariant measure strictly positive on open sets together with absorbing states. We prove the joint large deviation principle for the empirical measure and flow. Due to the lack of uniform ergodicity, the zero level set of the rate function is not a singleton. As corollaries, we obtain the Donsker-Varadhan rate function for the empirical measure and a variational expression of the rate function for the empirical flow.

1. Introduction

The energy transport in insulators can be described within a kinetic approach analogous to the kinetic theory of gases. At low temperatures the lattice vibrations, responsible of energy transport, can be modeled as a gas of interacting particles (phonons) and their time-dependent distribution function solves a Boltzmann type equation. The basic scheme to derive phononic Boltzmann equations from the underlying microscopic dynamics is introduced in Spohn (2006). Following this approach, in Basile et al. (2010) a harmonic chain of oscillators perturbed by a conservative weak stochastic noise is analyzed and the following linear Boltzmann equation is derived

$$\partial_t W(t,r,k) + v(k)\partial_r W(t,r,k) = \int_{\mathbb{T}} dk' R(k,k')[W(t,r,k') - W(t,r,k)]. \qquad (1.1)$$

Here W is the energy density distribution of phonons with wave number $k \in \mathbb{T}$ (the one dimensional torus), r is the space coordinate, t is the time and $v(k)$ is the velocity of a phonon with wave number k and it is given by the gradient of the dispersion relation. The scattering kernel R is positive and symmetric. Referring to

Received by the editors April 1, 2014; accepted January 21, 2015.
2010 *Mathematics Subject Classification.* 60F10, 60J75, 74A25.
Key words and phrases. Large deviations, Donsker-Varadhan theorem, Empirical flow, Phonon Boltzmann equation.

Basile et al. (2010) for explicit expressions of R and v and the analogous equation in higher dimensions, we point out the following features. The velocity is finite for small k while R behaves like k^2 for small k, and like k'^2 for small k'. This means that phonons with small wave numbers travel with finite velocity, but they have low probability to be scattered, therefore their mean free paths have a macroscopic length (ballistic transport).

The equation (1.1) can be interpreted as the Fokker-Planck equation for the Markov process $(K(t), Y(t))$ on $\mathbb{T} \times \mathbb{R}$, where the wave number $K(t)$ is a jump process and the position $Y(t)$ is an additive functional of K, namely $Y(t) = \int_0^t ds\, v(K(s))$. In view of the behavior of the kernel R mentioned above, $k = 0$ is an absorbing state for the process $K(t)$. On the other hand, the skeleton associated to K admits an invariant measure π_0 for which the mean jump time is integrable. As we prove, this condition implies the ergodicity of the process with $K(0)$ different from 0. Nevertheless, in dimension one and two the variance of the mean jump time with respect to π_0 is infinite, so that the standard central limit theorem for the position $Y(t)$ fails. More precisely, in one dimension the position converges to a 3/2 stable Lévy process under the proper scaling Jara et al. (2009); Basile and Bovier (2010), while in two dimensions it converges to a Brownian motion under an anomalous scaling with logarithmic corrections Basile (2014). The purpose of the present paper is to analyze how the degenerate behavior of the kernel R affects the large deviations properties of the process K.

In the general context of continuous time Markov processes, the empirical measure associates to a given trajectory the fraction of time spent on the different states up to a time T. Under ergodicity assumptions, the empirical measure converges to the invariant measure. The corresponding large deviations asymptotic is the content of the Donsker-Varadhan theorem, with a rate function given by a variational formula that can be computed explicitly only in the reversible case. A natural generalization of this framework in the setting of jump processes takes into account, together with the empirical measure, the empirical flow which counts the number of jumps between the different states per unit of time. We remark that a relevant dynamical observable, the empirical current, is directly related to the empirical flow.

The joint large deviation asymptotics for the empirical measure and flow can be derived by contraction from the corresponding result for the empirical process, which yields the information on arbitrary sequences of jumps. The corresponding rate function can be always written in a (simple) closed form. The Gallavotti-Cohen large deviation principle Maes (1999); Lebowitz and Spohn (1999) and the associated fluctuation theorem can be obtained by projection Bertini et al. (2015b). Moreover, by contraction one also derives a dual variational formula for the rate function of the empirical measure.

Alternatively, the joint large deviations for the empirical measure and flow can be directly derived by tilting the underlying Markov chain. Indeed, with this approach it has been firstly derived in Kesidis and Walrand (1993) for a Markov chain with two states. Always in the context of discrete state space, a large deviations principle for flows and currents have been discussed in Baiesi et al. (2009) in relation to statistical mechanics models. The general case of countable state space is analyzed in Bertini et al. (2015a), to which we refer for further references.

With respect to this setting, the phononic chain described above lives on a continuous state space and lacks uniform ergodicity due to the presence of zero as absorbing state. In particular the classical Donsker-Varadhan conditions Donsker and Varadhan (1975, 1976, 1983); Deuschel and Stroock (1989) do not hold. Motivated by this model, we consider a class of continuous time Markov chains which are degenerate in the sense that there exist states with infinite holding time, but the corresponding skeletons admit an invariant measure for which the mean jump time is integrable. For simplicity, we restrict to the case of compact state space. We prove a large deviation principle for the empirical measure and flow, with an initial state different from the absorbing ones.

The rate function is the continuous version of the one derived in Bertini et al. (2015a) for discrete space states. The presence of absorbing states is however reflected in the properties of the rate function. Its zero level set is not a singleton and more precisely it contains convex combinations of the invariant measure with the associated flow and measures supported by the absorbing states with zero flow. Indeed, with sub-exponential probability, the chain may spend almost all the time in a small neighborhood of the absorbing states. Analogous degenerate large deviation asymptotics have been obtained in Lefevere et al. (2011a,b) in the context of renewal processes and in Bodineau et al. (2012); Bodineau and Toninelli (2012) in the context of interacting particle systems.

From the large deviation principle for the empirical measure and flow we deduce by contraction the large deviation principle for the empirical measure. The corresponding rate function can be expressed by the Donsker-Varadhan variational formula, which in this case also admits a not trivial zero level set. Furthermore, we also obtain a variational expression for the rate function describing the large deviation asymptotics of the empirical flow.

The large deviation upper bound for the empirical measure and flow is proven by perturbing the rates of the underlying Markov chain. We remark that this step can be accomplished since the Radon-Nikodym derivative of the corresponding laws can be expressed in terms of the empirical measure and flow. We derive the lower bound by considering first deviations of measures and flows with support bounded away from the absorbing state. For this class we can construct perturbed Markov chains with nice ergodic properties, which have these measures and flows as typical behavior. We then complete the proof by a density argument.

2. Notation and results

Let E be a compact Polish space, i.e. metrizable complete and separable, endowed with its Borel σ-algebra. The spaces of continuous functions on E and $E \times E$, endowed with the uniform norm $\|\cdot\|$, are denoted by $C(E)$ and $C(E \times E)$. We consider a continuous time Markov chain ξ_t, $t \in \mathbb{R}_+$ on the state space E, defined by transitions rates $c(x, dy) = r(x)p(x, dy)$, where $r\colon E \to \mathbb{R}_+$ and p is a transition kernel on E. Throughout all the paper we assume the transition rates satisfy the following conditions which in particular imply that ξ is not explosive and Feller.

Assumption 2.1.

(i) The function $r\colon E \to \mathbb{R}_+$ is continuous. We set $E_0 := \{x \in E : r(x) = 0\}$.

(ii) There exists a probability λ on E, strictly positive on open sets, such that $p(x, dy) = p(x, y)\lambda(dy)$ for some strictly positive density $p \in C(E \times E)$.

(iii) The function $1/r$ is integrable with respect to λ, i.e., $\lambda\big(1/r\big) < +\infty$.

In order to prove the large deviations lower bound we also need the following technical condition.

(iv) For $\delta > 0$, let $A_\delta := \{x \in E : r(x) < \delta\}$ be the (open) level set of r. There exists a sequence $\delta_n \to 0$ such that $\sup_n \lambda(A_{2\delta_n})/\lambda(A_{2\delta_n} \setminus A_{\delta_n}) < +\infty$.

Since r is continuous E_0 is closed. Moreover, in view of condition (iii) $\lambda(E_0) = 0$. Assumption (ii) implies that the kernel p is Feller and satisfies the Doeblin condition. In view of Meyn and Tweedie (2009, Thm. 16.0.2) the discrete time Markov chain with kernel p is uniform ergodic. That is, there exists a probability π_0 on E such that $p^n(x, \cdot)$ converges in total variation to π_0 uniformly with respect to $x \in E$. Moreover, since λ is strictly positive on open sets, π_0 enjoys the same property. In view of items (ii) and (iii), $\pi_0\big(1/r\big) < +\infty$ and therefore

$$\pi(dx) := \frac{1}{\pi_0\big(1/r\big)} \frac{\pi_0(dx)}{r(x)} \tag{2.1}$$

defines a probability on E. As it is simple to check, π is an invariant probability for the continuous time chain ξ.

As discussed in the Introduction, the main novelty of this paper is that we allow the set E_0 to be not empty. If this is the case, the points in E_0 are absorbing states for the chain ξ. In particular any probability supported on a subset of E_0 is also an invariant measure and ξ is not uniformly ergodic. Then the standard conditions for the Donsker-Varadhan theorem, see e.g., Deuschel and Stroock (1989); Donsker and Varadhan (1975, 1976, 1983) do not hold. The phononic chain described by (1.1) (see Basile et al. (2010) for the explicit expression of the rates) meets the requirements in Assumption 2.1 with E_0 the singleton at the point 0.

We next state the ergodic theorem for the chain ξ. For $x \in E$ we denote by $\mathbb{P}_x$ the distribution of the process ξ with initial condition x. Observe that $\mathbb{P}_x$ is a probability on the Skorokhod space $D(\mathbb{R}_+; E)$ whose canonical coordinate will be denoted by X_t, $t \in \mathbb{R}_+$. The expectation with respect to $\mathbb{P}_x$ is denoted by $\mathbb{E}_x$.

Theorem 2.2. *Let $f \in C(E)$ and $x \in E \setminus E_0$. Then*

$$\lim_{T\to+\infty} \frac{1}{T} \int_0^T dt\, f(X_t) = \pi(f) \qquad \text{in } \mathbb{P}_x \text{ probability.}$$

Moreover, the convergence is uniform with respect to x in a compact subset of $E \setminus E_0$.

We denote by $\mathcal{M}_1(E)$ the space of probability measures on E endowed with the topology of weak convergence. Given $T > 0$, the *empirical measure* μ_T is the continuous map from $D(\mathbb{R}_+; E)$ to $\mathcal{M}_1(E)$ defined by

$$\mu_T(f)\,(X) := \frac{1}{T} \int_0^T dt\, f(X_t), \qquad f \in C(E). \tag{2.2}$$

Theorem 2.2 can be then restated as follows. As $T \to +\infty$ the family $\big\{\mathbb{P}_x \circ \mu_T^{-1}\big\}_{T>0}$ converges to δ_π uniformly with respect to x in a compact subset of $E \setminus E_0$.

To describe the large deviation asymptotic of the empirical measure, we follow the approach introduced in Bertini et al. (2015a) for discrete state space. Within this scheme, together with the empirical measure it is also considered the empirical flow which accounts for the number of jumps between two given states. For this purpose, we let $\mathcal{M}_+(E\times E)$ be the space of finite positive measures on $E\times E$ equipped with the *bounded weak** topology. This is defined as follows. Let $\mathcal{M}(E\times E)$ be the set of finite signed measure on $E\times E$. The weak* topology on $\mathcal{M}(E\times E)$ is then defined by identifying it with the dual of $C(E\times E)$. For $Q\in\mathcal{M}(E\times E)$ denote by $\|Q\|_{\mathrm{TV}}$ the total variation of Q and, given $\ell>0$, let $B_\ell := \big\{Q\in\mathcal{M} : \|Q\|_{\mathrm{TV}}\le \ell\big\}$ be the closed ball of radius ℓ in $\mathcal{M}(E\times E)$. The bounded weak* topology on $\mathcal{M}(E\times E)$ is then defined by declaring a set $A\subset\mathcal{M}(E\times E)$ open if and only if $A\cap B_\ell$ is open in the weak* topology of B_ℓ for any $\ell>0$. In particular, the bounded weak* topology is stronger than the weak* topology and, as follows from the Banach-Alaoglu theorem, for each $\ell>0$ the closed ball B_ℓ is compact with respect to the bounded weak* topology. The space $\mathcal{M}(E\times E)$ endowed with the bounded weak* topology is a locally convex, complete, linear topological space, and a completely regular space, i.e., for each closed set $C\subset\mathcal{M}(E\times E)$ and each $Q\in\mathcal{M}(E\times E)\setminus C$ there exists a continuous function $f\colon\mathcal{M}(E\times E)\to[0,1]$ such that $f(Q)=1$ and $f(Q')=0$ for all $Q'\in C$). We refer to Megginson (1998, § 2.7) for the proof of the above statements and for further details. We finally regard $\mathcal{M}_+(E\times E)$ as a (closed) subset of $\mathcal{M}(E\times E)$ and consider it endowed with the relative topology and the associated Borel σ-algebra.

For $T>0$, the *empirical flow* Q_T is the map from $D(\mathbb{R}_+;E)$ to $\mathcal{M}_+(E\times E)$ defined $\mathbb{P}_x$ a.s., $x\in E$, by

$$Q_T(F)\,(X) := \frac{1}{T}\sum_{\substack{t\in[0,T]:\\ X_{t^-}\neq X_t}} F\big(X_{t^-},X_t\big), \qquad F\in C(E\times E).$$

Observe indeed that the right hand side is well defined because $\mathbb{P}_x$ a.s., $x\in E$, the set of discontinuities of X_t is locally finite. In view of Theorem 2.2, a straightforward martingale decomposition, see Proposition 5.2 below, yields the following law of large numbers for empirical flow. Let $Q^\pi(dx,dy) := \pi(dx)c(x,dy)$, then as $T\to+\infty$ the family $\big\{\mathbb{P}_x\circ Q_T^{-1}\big\}_{T>0}$ converges to δ_{Q^π} uniformly with respect to x in a compact subset of $E\setminus E_0$.

We regard the pair (μ_T,Q_T) as a map from $D(\mathbb{R}_+;E)$ to the product space $\mathcal{M} := \mathcal{M}_1(E)\times\mathcal{M}_+(E\times E)$ defined $\mathbb{P}_x$ a.s., $x\in E$. Our main result is the large deviation principle for the family $\big\{\mathbb{P}_x\circ(\mu_T,Q_T)^{-1}\big\}_{T>0}$. We start by defining the rate function. Let $\Psi\colon\mathbb{R}_+\to\mathbb{R}_+$ be the convex function $\Psi(a) := a\log a-(a-1)$, in which we understand that $\Psi(0)=1$. We then define the functional $I\colon\mathcal{M}\to[0,+\infty]$ by

$$I(\mu,Q) := \begin{cases} \displaystyle\iint \mu(dx)c(x,dy)\,\Psi\Big(\frac{Q(dx,dy)}{\mu(dx)c(x,dy)}\Big) & \text{if } Q(\cdot,E)=Q(E,\cdot),\\ +\infty & \text{otherwise.}\end{cases} \tag{2.3}$$

Observe that $I(\mu,Q)<+\infty$ implies that the two marginals of Q are equal and $Q(dx,dy)\ll\mu(dx)c(x,dy)$. Moreover, since the second marginal of $\mu(dx)c(x,dy)$ is absolutely continuous with respect to λ, $I(\mu,Q)<+\infty$ also implies $dQ=q\,d\lambda\times d\lambda$, see Lemma 4.4 below. It is thus possible to express I in terms of the density q as follows. Given $\mu\in\mathcal{M}_1(E)$, decompose it into the absolutely continuous and singular parts with respect to λ. Namely, $d\mu=\varrho\,d\lambda+d\mu_{\mathrm{s}}$ where ϱ is a sub-probability density on E and μ_{s} is singular with respect to λ. Let $\Phi\colon\mathbb{R}_+^2\to[0,+\infty]$

be the convex lower semi-continuous function defined by $\Phi(a,b) := a\log(a/b) - a + b$. If $I(\mu,Q) < +\infty$ then the marginals of Q are equal and

$$I(\mu,Q) = \iint \lambda(dx)\lambda(dy)\,\Phi\big(q(x,y),\varrho(x)r(x)p(x,y)\big) + \mu_{\mathrm{s}}(r). \tag{2.4}$$

In particular, if $\mu = \mu_{\mathrm{s}}$ then $I(\mu,Q) < +\infty$ implies $Q = 0$.

Theorem 2.3. *As $T \to +\infty$ the family $\{\mathbb{P}_x \circ (\mu_T, Q_T)^{-1}\}_{T>0}$ satisfies, uniformly with respect to x bounded away from E_0, a large deviation principle with good convex rate function I. Namely, the functional I has compact level sets and for each compact $E_1 \subset E \setminus E_0$, each closed $C \subset \mathcal{M}$, respectively each open $A \subset \mathcal{M}$,*

$$\varlimsup_{T\to+\infty} \sup_{x\in E_1} \frac{1}{T}\log \mathbb{P}_x\big((\mu_T,Q_T)\in C\big) \le -\inf_{(\mu,Q)\in C} I(\mu,Q),$$
$$\varliminf_{T\to+\infty} \inf_{x\in E_1} \frac{1}{T}\log \mathbb{P}_x\big((\mu_T,Q_T)\in A\big) \ge -\inf_{(\mu,Q)\in A} I(\mu,Q).$$

By the stationarity condition for π, the measure $Q^\pi(dx,dy) = \pi(dx)c(x,dy)$ has equal marginals. We thus deduce, as must be the case, that $I(\pi,Q^\pi) = 0$. On the other hand, if E_0 is not empty, the zero level set of I contains other points and the law of large numbers stated in Theorem 2.2 cannot be deduced from the large deviation result. More precisely, if the measure μ is supported on a subset of E_0, then $I(\mu,0) = 0$ and, by convexity, I vanishes on the segment $\alpha(\pi,Q^\pi) + (1-\alpha)(\mu,0)$, $\alpha\in[0,1]$. The representation (2.4) implies that elements of this form are the only zeros of I.

As a corollary of the previous theorem, we deduce the large deviations asymptotic for the empirical measure. We emphasize that the corresponding rate function is the standard Donsker-Varadhan functional.

Corollary 2.4. *Let $\widehat{I}\colon \mathcal{M}_1(E) \to [0,+\infty]$ be the functional defined by*

$$\widehat{I}(\mu) = \sup_{\phi\in C(E)} \Big\{ -\iint \mu(dx)c(x,dy)\,\big[\exp\{\phi(y)-\phi(x)\} - 1\big]\Big\}.$$

As $T \to +\infty$ the family $\{\mathbb{P}_x\circ\mu_T^{-1}\}_{T>0}$ satisfies, uniformly with respect to x bounded away from E_0, a large deviation principle with convex rate function $\widehat{I}$.

As a further projection of Theorem 2.3, we obtain a variational expression, that appears to be new, of the rate function for the empirical flow.

Corollary 2.5. *Let $\widetilde{I}\colon \mathcal{M}_+(E\times E) \to [0,+\infty]$ be the functional defined by*

$$\widetilde{I}(Q) = \begin{cases} \displaystyle\sup_{\alpha\in(-r_{\mathrm{m}},+\infty)} \Big\{ \iint Q(dx,dy)\,\log\Big[\frac{Q(dx,dy)}{Q(dx,E)c(x,dy)}(r(x)+\alpha)\Big] - \alpha\Big\} \\ \qquad\qquad \text{if } Q(\cdot,E) = Q(E,\cdot) \\ +\infty \qquad\qquad \text{otherwise,} \end{cases}$$

where $r_{\mathrm{m}} := \min r$. As $T \to +\infty$ the family $\{\mathbb{P}_x \circ Q_T^{-1}\}_{T>0}$ satisfies, uniformly with respect to x bounded away from E_0, a large deviation principle with convex rate function $\widetilde{I}$.

3. Law of large numbers

We denote with $\{Z_i\}_{i\geq 0}$ the skeleton of the process ξ, namely the sequence of the visited states, and with $\{\tau_i\}_{i\geq 0}$ the collection of the holding times. The skeleton $\{Z_i\}_{i\geq 0}$ is a Markov chain with transition probability $p(x, dy)$. Conditioned to the skeleton $\{Z_i\}_{i\geq 0}$, $\{\tau_i\}_{i\geq 0}$ are independent, exponentially distributed random variables with parameters $r(Z_i)$. In particular they have the same law as $\{r(Z_i)^{-1}e_i\}_{i\geq 0}$, where $\{e_i\}_{i\geq 0}$ are i.i.d. exponential random variables with parameter 1.

We denote with T_n, $n \geq 0$, the jump times $T_n := \sum_{i=0}^{n-1} \tau_i$ for $n \geq 1$ and $T_0 = 0$. We then define the clock process $\mathcal{T}(t) := T_{\lfloor t\rfloor}$, where $\lfloor\cdot\rfloor$ denotes the integer part. The inverse function $n(t) := \inf\{n : \ T_n \geq t\}$ gives the number of jumps up to time t. By definition, the following inequality holds

$$T_{n(t)-1} < t \leq T_{n(t)}, \tag{3.1}$$

where we take $T_{-1} = 0$.

Proposition 3.1. *Let $\pi_0 \in \mathcal{M}_1(E)$ be the unique invariant measure of the chain $\{Z_i\}$. Then for each $f \in C(E)$ and $Z_0 \in E \setminus x_0$*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(Z_i)\tau_i = \pi_0(f/r) \qquad \textit{in probability.} \tag{3.2}$$

Moreover, the convergence is uniform with respect to Z_0 in a compact subset of $E \setminus E_0$.

Postponing the proof above statement, we first show that it implies the law of large numbers for the continuous time chain ξ.

Proof of Theorem 2.2: Recalling the definition of the empirical measure (2.2),

$$\mu_T(f) = \frac{1}{T}\sum_{i=0}^{n(T)} f(Z_i)\tau_i.$$

For each $f \in C(E)$, $\varepsilon > 0$, and $\sigma > 0$

$$\begin{aligned}\mathbb{P}_x\Big(\big|\mu_T(f) - \pi(f)\big| > \varepsilon\Big) &\leq \mathbb{P}_x\Big(\Big|\frac{1}{T}n(T) - \frac{1}{\pi_0(1/r)}\Big| > \sigma\Big)\\ &\quad + \mathbb{P}\Big(\big|\mu_T(f) - \pi(f)\big| > \varepsilon,\ \Big|\frac{1}{T}n(T) - \frac{1}{\pi_0(1/r)}\Big| \leq \sigma\Big).\end{aligned} \tag{3.3}$$

From (3.2) with $f = 1$, we deduce that the sequence $\{T_n/n\}_{n\geq 1}$ converges in probability to $\pi_0(1/r)$ and therefore, in view of (3.1), the family of random variables $\{n(T)/T\}_{T>0}$ converges in probability to $\big(\pi_0(1/r)\big)^{-1}$. This implies that the first term on the right hand side of (3.3) vanishes as $T \to \infty$.

On the other hand, on the event $\big\{\big|\frac{1}{T}n(T) - \frac{1}{\pi_0(1/r)}\big| \leq \sigma\big\}$,

$$\Big|\mu_T(f) - \frac{1}{T}\sum_{i=0}^{\alpha(T)} f(Z_i)\tau_i\Big| \leq \|f\|\,\frac{1}{T}\sum_{i=\alpha(T)-\lfloor\sigma T\rfloor-1}^{\alpha(T)+\lfloor\sigma T\rfloor+1} \tau_i,$$

where $\alpha(t) = \lfloor t/\pi_0(1/r) \rfloor$. In view of condition (ii) in Assumption 2.1, for each $i \geq 1$ and $Z_0 \in E$

$$\mathbb{E}\big(\tau_i\big) = \int p^{i-1}(Z_0, dx) \int p(x, dy) \frac{1}{r(y)} \leq C \int p^{i-1}(Z_0, dx) \int \lambda(dy) \frac{1}{r(y)},$$

where $C = \max p(x, y)$. By condition (iii) in Assumption 2.1, we thus get

$$\sup_{i \geq 1} \mathbb{E}\big(\tau_i\big) < +\infty. \tag{3.4}$$

Hence, by Chebychev inequality, the second term on the right hand side of (3.3) vanishes as $\sigma \to 0$ uniformly in T. □

Proof of Proposition 3.1: Recalling that, conditionally on $\{Z_i\}$, $\{\tau_i\}$ have the same law as $\{r(Z_i)^{-1} e_i\}$, we set $S_n(f) = \frac{1}{n} \sum_{i=0}^{n-1} f(Z_i) r(Z_i)^{-1} e_i$ and define

$$u^<_{n,i} := \frac{1}{r(Z_i)} 1\!\!\mathrm{I}_{\{r(Z_i)^{-1} \leq n^{1/4}\}}, \qquad u^>_{n,i} := \frac{1}{r(Z_i)} 1\!\!\mathrm{I}_{\{r(Z_i)^{-1} > n^{1/4}\}}.$$

We decompose accordingly

$$S_n(f) = \frac{1}{n} f(Z_0) r(Z_0)^{-1} e_0 + S^<_n(f) + S^>_n(f), \tag{3.5}$$

where

$$S^<_n(f) := \frac{1}{n} \sum_{i=1}^{n-1} f(Z_i) u^<_{n,i} e_i,$$
$$S^>_n(f) := \frac{1}{n} \sum_{i=1}^{n-1} f(Z_i) u^>_{n,i} e_i.$$

Trivially, the first term on the r.h.s. of (3.5) vanishes as $n \to \infty$, uniformly with respect to Z_0 in a compact subset of $E \setminus E_0$. Let us next show that $S^>_n(f)$ converges to zero in L^1. We have

$$\mathbb{E}\big(|S^>_n(f)|\big) \leq \frac{1}{n} \|f\| \sum_{i=1}^{n-1} \mathbb{E}\big(u^>_{n,i}\big),$$

where, by Chapman-Kolmogorov,

$$\begin{aligned} &\frac{1}{n} \|f\| \sum_{i=1}^{n-1} \int p^{i-1}(Z_0, dx) \int p(x, dy) \, \frac{1}{r(y)} 1\!\!\mathrm{I}_{\{r(y)^{-1} > n^{1/4}\}} \\ &\qquad \leq \|f\| \sup_{x \in E} \int p(x, dy) \, \frac{1}{r(y)} 1\!\!\mathrm{I}_{\{r(y)^{-1} > n^{1/4}\}}, \end{aligned}$$

which vanishes as $n \to \infty$ by conditions (ii) and (iii) in Assumption 2.1.

We next show that $S^<_n(f)$ converges to $\pi_0(f/r)$ in L^2. We have

$$\begin{aligned} \mathbb{E}\big(S^<_n(f) - \pi_0(f/r)\big)^2 &\leq \frac{1}{n^2} \sum_{i=1}^{n-1} \mathbb{E}\Big(f(Z_i) u^<_{n,i} e_i - \pi_0(f/r)\Big)^2 \\ &+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n-1} \mathbb{E}\Big(\big[f(Z_i) u^<_{n,i} e_i - \pi_0(f/r)\big]\big[f(Z_j) u^<_{n,j} e_j - \pi_0(f/r)\big]\Big). \end{aligned} \tag{3.6}$$

By definition of $u^<_{n,i}$, $\mathbb{E}\big(f(Z_i)u^<_{n,i}e_i\big)^2 \le \|f\|^2 n^{1/2}$. Therefore the first sum on the right hand side of (3.6) vanishes as $n \to \infty$. In order to estimate the second sum on the r.h.s. of (3.6), we observe that for each $1 \le i < j \le n-1$

$$\begin{aligned}\Big|\mathbb{E}\Big(f(Z_i)\big[u^<_{n,i} - r(Z_i)^{-1}\big]\, f(Z_j)\big[u^<_{n,j} - r(Z_j)^{-1}\big]\Big)\\ \le \|f\|^2 \Big(\sup_{x\in E}\int p(x,dy)\frac{1}{r(y)}\mathbb{1}_{\{r(y)^{-1}>n^{1/4}\}}\Big)^2,\end{aligned}$$

which vanishes as $n \to \infty$ by conditions (ii) and (iii) in Assumption 2.1. Therefore we can replace $u^<_{n,i}$ by $r^{-1}(X_i)$ in the second term of the r.h.s. of (3.6). By the same computations presented above, we get that each term in this modified sum is uniformly bounded, that is

$$\Big|\mathbb{E}\Big(\big[f(Z_i)r(Z_i)^{-1}e_i - \pi_0(f/r)\big]\big[f(Z_j)r(Z_j)^{-1}e_j - \pi_0(f/r)\big]\Big)\Big| \le C^2\,\|f\|^2, \tag{3.7}$$

where $C := \sup_{x\in E}\int p(x,dy)\,1/r(y)$. Given $m < n-1$, we now split the sum

$$\sum_{1\le i<j\le n-1} = \sum_{\substack{1\le i<m\\ i<j\le n-1}} + \sum_{\substack{m\le i<j\le n-1\\ j-i\le m}} + \sum_{\substack{m\le i<j\le n-1\\ j-i>m}}.$$

By (3.7), the first and the second sum give a contribution of order m/n, then it remains to estimate the last sum. Since $\pi_0(\cdot) = \int \pi_0(dx)p(x,\cdot)$, for $\ell \ge 0$ we can write

$$\begin{aligned}\Big|\int p^{\ell+1}(x,dy)\,\frac{f(y)}{r(y)} - \pi_0(f/r)\Big| &= \Big|\int \big(p^\ell(x,dz) - \pi_0(dz)\big)\int p(z,dy)\,\frac{f(y)}{r(y)}\Big|\\ &\le C\,\|f\|\,\sup_{x\in E}\|p^\ell(x,\cdot) - \pi_0(\cdot)\|_{\mathrm{TV}}.\end{aligned} \tag{3.8}$$

Therefore

$$\begin{aligned}\frac{1}{n^2}\sum_{\substack{m\le i<j\le n-1\\ j-i>m}}\Big|\mathbb{E}\Big(\big[f(Z_i)r(Z_i)^{-1}e_i - \pi_0(f/r)\big]\big[f(Z_j)r(Z_j)^{-1}e_j - \pi_0(f/r)\big]\Big)\Big|\\ \le C^2\,\|f\|^2\,\sup_{\ell\ge m}\Big(\sup_{x\in E}\|p^\ell(x,\cdot) - \pi_0(\cdot)\|_{\mathrm{TV}}\Big)^2.\end{aligned}$$

Since condition (ii) in Assumption 2.1 implies the Doeblin condition for p, the uniform ergodicity of the chain, see e.g., Meyn and Tweedie (2009, Thm. 16.0.2), implies that the r.h.s. above vanishes as $m \to \infty$. We then conclude the proof taking the limit $n \to \infty$ and then $m \to \infty$. □

4. Large deviations upper bound

We denote the marginals of $Q \in \mathcal{M}_+(E\times E)$ by $Q^{(1)}$ and $Q^{(2)}$. For $F \in C(E\times E)$ we let $r^F : E \to \mathbb{R}_+$ be the continuous function defined by

$$r^F(x) := \int c(x,dy)\,e^{F(x,y)}; \tag{4.1}$$

observing that for $F = 0$ we get $r^0 = r$. Given $\phi \in C(E)$ and $F \in C(E\times E)$ let $I_{\phi,F} : \mathcal{M} \to \mathbb{R}$ be the continuous affine map defined by

$$I_{\phi,F}(\mu,Q) := Q^{(1)}(\phi) - Q^{(2)}(\phi) + Q(F) - \mu\big(r^F - r\big). \tag{4.2}$$

In this section we first prove, by an exponential tilt of the underlying probability, the large deviation upper bound with rate function $\sup_{\phi,F} I_{\phi,F}$. As in Bertini et al. (2015a), this step can be easily accomplished since we are considering the joint deviations of the empirical measure and flow. We then show that the rate function thus obtained coincides with (2.3). We remark that the upper bound estimate holds uniformly with respect to all initial conditions in E.

Proposition 4.1. *As $T \to +\infty$ the family $\{\mathbb{P}_x \circ (\mu_T, Q_T)^{-1}\}_{T>0}$ satisfies, uniformly with respect to $x \in E$, a large deviation upper bound with lower semi-continuous convex rate function $\sup_{\phi,F} I_{\phi,F}$. Namely, for each closed $C \subset \mathcal{M}$*

$$\varlimsup_{T\to+\infty} \sup_{x\in E} \frac{1}{T} \log \mathbb{P}_x\big((\mu_T, Q_T) \in C\big) \le - \inf_{(\mu,Q)\in C} \sup_{\phi,F} I_{\phi,F}(\mu, Q)$$

where the supremum is carried out over all $(\phi, F) \in C(E) \times C(E \times E)$.

We start by proving the exponential tightness, that is there exists a sequence $\{\mathcal{K}_\ell\}_{\ell\in\mathbb{N}}$ of compacts in $\mathcal{M}$ such that

$$\lim_{\ell\to\infty} \varlimsup_{T\to\infty} \sup_{x\in E} \frac{1}{T} \log \mathbb{P}_x\big((\mu_T, Q_T) \notin \mathcal{K}_\ell\big) = -\infty.$$

Recall $\mathcal{M} = \mathcal{M}_1(E) \times \mathcal{M}_+(E \times E)$. Since $\mathcal{M}_1(E)$ is compact with respect to the topology of weak convergence and $\mathcal{M}_+(E\times E)$ is endowed with the bounded weak* topology, the previous bound follows from the exponential tightness of the sequence of positive random variables $\{Q_T(1)\}_{T>0}$, which count the total number of jumps per unit of time.

Lemma 4.2. *Let $a_\ell \to +\infty$. Then*

$$\lim_{\ell\to\infty} \varlimsup_{T\to\infty} \sup_{x\in E} \frac{1}{T} \log \mathbb{P}_x\big(Q_T(1) > a_\ell\big) = -\infty.$$

Proof: Given $F \in C(E \times E)$, let $\mathbb{M}^F$ be the process defined by

$$\mathbb{M}^F_t = \exp\big\{t\big[Q_t(F) - \mu_t(r^F - r)\big]\big\}, \qquad t \in \mathbb{R}_+. \tag{4.3}$$

By standard Markov chain computations, see e.g., Brémaud (1981, §VI.2), $\mathbb{M}^F$ is a mean one positive $\mathbb{P}_x$ martingale, $x \in E$. By choosing $F(x, y) = \gamma > 0$, $(x, y) \in E \times E$, for $a > 0$, $T > 0$ we then write

$$\begin{aligned}\mathbb{P}_x\big(Q_T(1) > a\big) &= \mathbb{E}_x\big(e^{-T\{\gamma Q_T(1) - \mu_T(r^\gamma - r)\}}\, \mathbb{M}^F_T\, 1\!\!1_{\{Q_T(1)>a\}}\big)\\ &\le e^{-T\gamma a}\, e^{T\|r\|(e^\gamma-1)}\, \mathbb{E}_x\big(\mathbb{M}^F_T\big) = e^{-T\gamma a}\, e^{T\|r\|(e^\gamma-1)}.\end{aligned}$$

The statement follows. □

Lemma 4.3. *For each $(\phi, F) \in C(E) \times C(E \times E)$ and each measurable $\mathcal{B} \subset \mathcal{M}$,*

$$\varlimsup_{T\to\infty} \sup_{x\in E} \frac{1}{T} \log \mathbb{P}_x\Big((\mu_T, Q_T) \in \mathcal{B}\Big) \le - \inf_{(\mu,Q)\in\mathcal{B}} I_{\phi,F}(\mu, Q).$$

Proof: Fix $x \in E$ and observe that the following path-wise continuity equation holds $\mathbb{P}_x$ a.s.,

$$\begin{aligned}0 &= \phi(X_T) - \phi(X_0) - \sum_{t\in[0,T]} \big[\phi(X_t) - \phi(X_t^-)\big]\\ &= \phi(X_T) - \phi(X_0) - T\big[Q^{(2)}_T(\phi) - Q^{(1)}_T(\phi)\big].\end{aligned} \tag{4.4}$$

In view of (4.2) and (4.4), recalling the martingale introduced in (4.3), for each $T > 0$

$$\begin{aligned}
&\mathbb{P}_x\big((\mu_T, Q_T) \in \mathcal{B}\big) \\
&\quad = \mathbb{E}_x\Big(\exp\big\{-T\, I_{\phi,F}(\mu_T, Q_T) - \big[\phi(X_T) - \phi(x)\big]\big\}\, \mathbb{M}_T^F\, 1\!\!1_{\mathcal{B}}(\mu_T, Q_T)\Big) \\
&\quad \le \sup_{(\mu,Q)\in\mathcal{B}} e^{-T\, I_{\phi,F}(\mu,Q)}\, \mathbb{E}_x\Big(\exp\big\{-\big[\phi(X_T) - \phi(x)\big]\big\}\, \mathbb{M}_T^F\, 1\!\!1_{\mathcal{B}}(\mu_T, Q_T)\Big) \\
&\quad \le \sup_{(\mu,Q)\in\mathcal{B}} e^{-T\, I_{\phi,F}(\mu,Q)} e^{2\|\phi\|},
\end{aligned}$$

where in the last step we used $\mathbb{E}_x\big(\mathbb{M}_T^F\big) = 1$. The statement follows. □

Proof of Proposition 4.1: In view of the exponential tightness proven in Lemma 4.2, it is enough to prove the upper bound for compacts. For each compact $\mathcal{K} \subset \mathcal{M}$, by Lemma 4.3 and the min-max lemma in Kipnis and Landim (1999, App. 2, Lemma 3.3)

$$\varlimsup_{T\to+\infty} \sup_{x\in E} \frac{1}{T} \log \mathbb{P}_x\Big((\mu_T, Q_T) \in \mathcal{K}\Big) \le -\inf_{(\mu,Q)\in\mathcal{K}} \sup_{\phi,F} I_{\phi,F}(\mu, Q).$$

Finally, as the map $(\mu, Q) \mapsto I_{\phi,F}(\mu, Q)$ is continuous and affine, the functional $\sup_{\phi,F} I_{\phi,F}$ is lower semi-continuous and convex. □

Recalling that the functional I is defined in (2.3), we show that it coincides with $\sup_{\phi,F} I_{\phi,F}$.

Lemma 4.4. *For each $(\mu, Q) \in \mathcal{M}$,*

$$I(\mu, Q) = \sup_{\phi,F} I_{\phi,F}(\mu, Q). \tag{4.5}$$

In particular, I is lower semicontinuous and convex. Moreover, if $I(\mu, Q) < +\infty$ then $Q \ll \lambda \times \lambda$ and (2.4) holds.

Proof: Clearly, $\sup_\phi \big\{Q^{(1)}(\phi) - Q^{(2)}(\phi)\big\} < +\infty$ if and only if $Q^{(1)} = Q^{(2)}$. For $\mu \in \mathcal{M}_1(E)$ we denote by $Q^\mu \in \mathcal{M}_+(E \times E)$ the measure $Q^\mu(dx, dy) := \mu(dx)c(x, dy)$ and set $\Lambda(\mu, Q) := \sup_F \big\{Q(F) - Q^\mu\big(e^F - 1\big)\big\}$. Recalling (4.1) and (4.2), the proof of (4.5) is achieved once we show that if $Q^{(1)} = Q^{(2)}$ then $\Lambda(\mu, Q) = I(\mu, Q)$.

For Q with equal marginals we next prove that $\Lambda(\mu, Q) \le I(\mu, Q)$. We can assume $I(\mu, Q) < +\infty$ so that $Q \ll Q^\mu$. Then

$$Q(F) - Q^\mu\big(e^F - 1\big) = \iint dQ^\mu \Big\{\frac{dQ}{dQ^\mu} F - \big(e^F - 1\big)\Big\}.$$

Since $\Psi(a) = \sup_{\lambda\in\mathbb{R}} \big\{\lambda a - \big(e^\lambda - 1\big)\big\}$, $a \in \mathbb{R}_+$, we complete this step by taking the supremum over F.

To obtain the converse inequality, we first prove that if $\Lambda(\mu, Q) < +\infty$ then $Q \ll Q^\mu$. Let $\widetilde{B}$ be a Borel set in $E \times E$ such that $Q^\mu(\widetilde{B}) = 0$, we show that also $Q(\widetilde{B}) = 0$. By regularity of the measure Q^μ there exists a sequence of open sets $A_n \supset \widetilde{B}$ in $E\times E$ such that $\lim_n Q^\mu(A_n) = Q^\mu(\widetilde{B}) = 0$. By approximating indicator of open sets with continuous functions we can take as test function $F = \gamma 1\!\!1_{A_n}$, $\gamma > 0$, and deduce

$$\gamma\, Q(B) \le \gamma\, Q(A_n) \le \Lambda(\mu, Q) + \big(e^\gamma - 1\big)\, Q^\mu(A_n).$$

We conclude by taking first the limit as $n \to \infty$ and then $\gamma \to \infty$. To prove $\Lambda(\mu, Q) \ge I(\mu, Q)$ (for Q with equal marginals) we can assume $\Lambda(\mu, Q) < +\infty$ so that $Q \ll Q^\mu$. Pick an array of continuous functions $\{F_{k,n}\}$ equibounded in n such that $\{F_{k,n}\}_{n\ge 0}$ converges to $\log\big[\big(dQ/dQ^\mu \wedge k\big) \vee 1/k\big]$ in $L^1(E\times E, dQ^\mu)$. Then

$$\Lambda(\mu, Q) \ge \iint dQ^\mu \, \frac{dQ}{dQ^\mu} \log \Big[\Big(\frac{dQ}{dQ^\mu} \wedge k\Big) \vee \frac{1}{k}\Big] - \iint dQ^\mu \Big\{\Big[\Big(\frac{dQ}{dQ^\mu} \wedge k\Big) \vee \frac{1}{k}\Big] - 1\Big\}.$$

By monotone convergence, we conclude taking the limit $k \to \infty$.

To prove the last statement of the lemma, we decompose the measure μ into its absolutely continuous and singular parts with respect to λ, i.e. $\mu = \mu_{\mathrm{ac}} + \mu_{\mathrm{s}}$. Accordingly, there exists a Borel set $B \subset E$ such that $\mu_{\mathrm{s}}(B) = \mu_{\mathrm{s}}(E)$ and $\lambda(B) = 0$. Since $Q^\mu(dx, dy) = \mu(dx) r(x) p(x, y) \lambda(dy)$, it holds $Q^\mu(E \times B) = 0$. As $Q \ll Q^\mu$ and $Q^{(1)} = Q^{(2)}$, this implies $Q(E\times B) = Q(B \times E) = 0$. Since the restriction of Q to $(E \setminus B)\times E$ is absolutely continuous with respect to $Q^{\mu_{\mathrm{ac}}} \ll \lambda \times \lambda$, then $Q \ll \lambda \times \lambda$. Straightforward manipulations now yield (2.4). □

The following estimate will be used in the proof of the lower bound.

Lemma 4.5. *Let $(\mu, Q) \in \mathcal{M}$ be such that $I(\mu, Q) < +\infty$. Then*

$$\iint Q(dx, dy) \, \log \frac{1}{r(x)p(x,y)} < +\infty.$$

Proof: For $k > 0$, choose as test function in the variational formula (4.5) the function $(x, y) \mapsto \log\big(k \wedge 1/r(x)p(x, y)\big)$. We deduce

$$\begin{aligned} &\iint Q(dx, dy) \, \log\Big(k \wedge \frac{1}{r(x)p(x,y)}\Big) \\ &\qquad \le I(\mu, Q) + \iint \mu(dx) c(x, dy) \Big(k \wedge \frac{1}{r(x)p(x,y)} - 1\Big) \le I(\mu, Q) + 1 \end{aligned}$$

where we used that $c(x, dy) = r(x)p(x, y)\lambda(dy)$. By taking the limit $k \to \infty$ we conclude the proof. □

5. Large deviations lower bound

We state a general result concerning the large deviation lower bound in which we denote by $\mathrm{Ent}(\widetilde{P}|P)$ the relative entropy of the probability $\widetilde{P}$ with respect to P.

Lemma 5.1. *Let $\{P_n^\alpha, \, \alpha \in A\}_{n\in\mathbb{N}}$ be a sequence of family of probability measures on a completely regular topological space $\mathcal{X}$. Assume that for each $z \in \mathcal{X}$ there exists a sequence of family of probability measures $\{\widetilde{P}_n^\alpha(z)\}$ weakly convergent to δ_z uniformly with respect to $\alpha \in A$ and such that*

$$\varlimsup_{n\to\infty} \sup_{\alpha\in A} \frac{1}{n} \, \mathrm{Ent}\big(\widetilde{P}_n^\alpha(z) \big| P_n^\alpha\big) \le J(z) \tag{5.1}$$

for some $J\colon \mathcal{X} \to [0, +\infty]$. Then the sequence of family $\{P_n^\alpha, \, \alpha \in A\}_{n\in\mathbb{N}}$ satisfies uniformly with respect to $\alpha \in A$ the large deviation lower bound with rate function given by $\mathrm{sc}^- J$, the lower semi-continuous envelope of J, i.e.,

$$\big(\mathrm{sc}^- J\big)(z) := \sup_{U \in \mathcal{N}_z} \; \inf_{w\in U} \; J(w)$$

where $\mathcal{N}_z$ denotes the collection of the open neighborhoods of z.

This lemma is proven in Jensen (2000, Prop. 4.1) in a Polish space setting without the dependence on the parameter α. The proof extends to the present setting. Note that in the our application we shall work only with sequences so that one can avoid the details of the general topological setting.

Our strategy to prove the large deviations lower bound is the following. We first prove the lower bound for a nice subset of $\mathcal{M}$. In view on Lemma 5.1 we then recover the full lower bound by a suitable density argument. More precisely, we let

$$\begin{aligned}\mathcal{M}_0 := \Big\{(\mu, Q) \in \mathcal{M} \,:\, & K := \mathrm{supp}(\mu) \subset E \setminus E_0,\, \mathrm{supp}(Q) = K \times K,\\ & d\mu = \varrho\, d\lambda \text{ with } \varrho \text{ continuous and } \varrho > 0 \text{ on } K, \\ & dQ = q\, d\lambda \times d\lambda \text{ with } q \text{ continuous and } q > 0 \text{ on } K \times K\Big\}.\end{aligned} \tag{5.2}$$

We shall prove the entropy bound (5.1) with J given by the restriction of I, as defined in (2.3), to $\mathcal{M}_0$, that is

$$J(\mu, Q) := \begin{cases} I(\mu, Q) & \text{if } (\mu, Q) \in \mathcal{M}_0, \\ +\infty & \text{otherwise.} \end{cases} \tag{5.3}$$

Then we complete the proof of the lower bound by showing that the lower semi-continuous envelope of J coincides with I.

Proposition 5.2. *Let $(\mu, Q) \in \mathcal{M}_0$ and $K := \mathrm{supp}(\mu)$. Then there exists a Markov family $\widetilde{\mathbb{P}}_x$, $x \in K$, such that $\widetilde{\mathbb{P}}_x \circ (\mu_T, Q_T)^{-1} \to \delta_{(\mu,Q)}$ uniformly with respect to $x \in K$ and*

$$\varlimsup_{T\to\infty} \sup_{x \in K} \frac{1}{T} \,\mathrm{Ent}\big(\widetilde{\mathbb{P}}_{x,\,[0,T]} \big| \mathbb{P}_{x,\,[0,T]}\big) \le I(\mu, Q), \tag{5.4}$$

where $\mathbb{P}_{x,\,[0,T]}$ denotes the restriction of $\mathbb{P}_x$ to $D([0, T], E)$.

Proof: We can assume $I(\mu, Q) < +\infty$, so that Q has equal marginals. Let $d\mu = \varrho\, d\lambda$, $dQ = q\, d\lambda \times d\lambda$, and let $\widetilde{c}$ be the transition rates on K defined by $\widetilde{c}(x, dy) = q(x, y)/\varrho(x)\, \lambda(dy)$. We denote by $\widetilde{\mathbb{P}}_x$ the law of the chain with rates $\widetilde{c}$ starting from x. Since Q has equal marginals, then it is easy to check that $\varrho\, d\lambda$ is an invariant measure of the chain. Moreover, the chain is Feller and satisfies $\widetilde{c}(x, dy) \ge c_0 \lambda(dy)$, $x \in K$, with $c_0 := \min_{K\times K} q(x, y)/\varrho(x) > 0$. Then, by the arguments of Section 3, μ_T converges to $\varrho\, d\lambda$ in $\widetilde{\mathbb{P}}_x$ probability, uniformly with respect to $x \in K$. In order to prove the law of large numbers for the empirical flow Q_T, we use the following semi-martingale decomposition. For each $F \in C(E \times E)$

$$t\, Q_t(F) = \int_0^t ds \int \widetilde{c}(X_s, dy)\, F(X_s, y) + M_t(F),$$

where the $\widetilde{\mathbb{P}}_x$ martingale $M(F)$ has predictable quadratic variation

$$\langle M(F)\rangle_t = \int_0^t ds \int \widetilde{c}(X_s, dy)\, F(X_s, y)^2.$$

Since $\widetilde{c}(x, dy) \le C\lambda(dy)$, then $\langle M(F)\rangle_t \le C\, t\, \lambda(F^2)$. Therefore, as the map $K \ni x \mapsto \int \widetilde{c}(x, dy)\, F(x, y)$ is continuous, the law of large numbers of the empirical measure μ_T implies

$$\lim_{T\to+\infty} Q_T(F) = \iint \mu(dx)\widetilde{c}(x, dy)\, F(x, y) = Q(F), \qquad \text{in } \widetilde{\mathbb{P}}_x \text{ probability,}$$

uniformly with respect to $x \in K$. Since by Lemma 4.2 the family $\{Q_T\}_{T>0}$ is tight, this implies the law of large numbers $\widetilde{\mathbb{P}}_x \circ (\mu_T, Q_T)^{-1} \to \delta_{(\mu,Q)}$ uniformly with respect to $x \in K$. Observe that this argument also shows that for each F the family of random variables $\{Q_T(F)\}_{T>0}$ converges to $Q(F)$ in L^2 with respect to $\widetilde{\mathbb{P}}_x$, uniformly in $x \in K$,

Set $F^*(x,y) := \log[q(x,y)/\varrho(x)r(x)p(x,y)]$, $(x,y) \in K \times K$. Then, by an explicit computation of the Radon-Nikodym derivative, see e.g., Brémaud (1981, §VI.2),

$$\frac{1}{T}\,\mathrm{Ent}\,\big(\widetilde{\mathbb{P}}_{x,\,[0,T]}\big|\mathbb{P}_{x,\,[0,T]}\big) = \frac{1}{T}\,\widetilde{\mathbb{E}}_x\Big(\log\frac{d\widetilde{\mathbb{P}}_{x,\,[0,T]}}{d\mathbb{P}_{x,\,[0,T]}}\Big) = \widetilde{\mathbb{E}}_x\Big(Q_T(F^*) - \mu_T(r^{F^*} - r)\Big).$$

Recalling the representation (2.4) for I, the law of large numbers just proven yields

$$\lim_{T\to\infty}\frac{1}{T}\,\mathrm{Ent}\,\big(\widetilde{\mathbb{P}}_{x,\,[0,T]}\big|\mathbb{P}_{x,\,[0,T]}\big) = Q(F^*) - \mu(r^{F^*} - r) = I(\mu,Q),$$

uniformly with respect to $x \in K$. □

We next show that the lower semi-continuous envelope of J, as defined in (5.3), coincides with I. A set $\mathcal{C} \subset \mathcal{M}$ is called I-dense in $\mathcal{D} \subset \mathcal{M}$ if and only if for each $(\mu,Q) \in \mathcal{D}$ such that $I(\mu,Q) < +\infty$ there exists a net $\{(\mu_\alpha,Q_\alpha)\} \subset \mathcal{C}$ such that $(\mu_\alpha,Q_\alpha) \to (\mu,Q)$ and $\lim_\alpha I(\mu_\alpha,Q_\alpha) = I(\mu,Q)$. We remark that by the lower semi-continuity of I, the second condition is equivalent to $\overline{\lim}_\alpha\, I(\mu_\alpha,Q_\alpha) \le I(\mu,Q)$.

Theorem 5.3. *The set $\mathcal{M}_0$ defined in (5.2) is I-dense in $\mathcal{M}$.*

The proof is split in few lemmata in which we use the following notation. Let A and B be respectively a Borel subset of E of strictly positive λ measure and a Borel subset of $E \times E$ of strictly positive $\lambda \times \lambda$ measure. For a function $f \in L^1(d\lambda)$, respectively a function $F \in L^1(d\lambda \times d\lambda)$ we set

$$⨍_A f := \frac{1}{\lambda(A)}\int_A \lambda(dx)\, f(x), \qquad ⨍\!\!\!\!\int_B F := \frac{1}{(\lambda\times\lambda)(B)}\iint_B \lambda(dx)\lambda(dy)\, F(x,y).$$

Lemma 5.4. *Let*

$$\begin{aligned}\mathcal{M}_1 := \Big\{&(\mu,Q) \in \mathcal{M} \,:\, K := \mathrm{supp}(\mu) \subset E \setminus E_0,\ \mathrm{supp}(Q) = K \times K,\\ &\exists\, D_1,\dots,D_\ell \subset K \textit{ disjoint open sets such that } \lambda\Big(K \setminus \bigcup_i D_i\Big) = 0,\\ &d\mu = \sum_i a_i\, \mathbb{1}_{D_i}\, d\lambda,\ \ a_i > 0,\ i = 1,..,\ell,\\ &dQ = \sum_{i,j} b_{ij}\, \mathbb{1}_{D_i \times D_j}\, d\lambda \times d\lambda,\ \ b_{ij} > 0,\ i,j = 1,..,\ell\Big\}.\end{aligned} \tag{5.5}$$

The set $\mathcal{M}_0$ is I-dense in $\mathcal{M}_1$.

Proof: Let $(\mu,Q) \in \mathcal{M}_1$ with $I(\mu,Q) < +\infty$, so that $Q^{(1)} = Q^{(2)}$. Denoting with $d(\cdot,\cdot)$ the distance in E, by Urysohn lemma, for each D_i, $i = 1,..,\ell$, and $n \in \mathbb{N}$ there exists a continuous function $\phi_i^n \colon K \to [0,1]$ such that

$$\phi_i^n(x) = \begin{cases} 1 & \text{if } x \in \overline{D}_i \\ 0 & \text{if } d(x,D_i) \ge \frac{1}{n},\end{cases}$$

where $\overline{D}_i$ is the closure of D_i. We define the sequence (μ_n, Q_n) by $d\mu_n = \varrho_n d\lambda$, $dQ_n = q_n d\lambda \times d\lambda$, with $\varrho_n = 0$ in $E \setminus K$ and $q_n = 0$ in $(E \times E) \setminus (K \times K)$, and

$$\varrho_n(x) = \sum_i a_i \frac{\phi_i^n(x)}{\int d\lambda\, \phi_i^n} \lambda(D_i),$$

$$q_n(x,y) = \sum_{i,j} b_{ij} \frac{\phi_i^n(x)}{\int d\lambda\, \phi_i^n} \frac{\phi_j^n(y)}{\int d\lambda\, \phi_j^n} \lambda(D_i)\, \lambda(D_j).$$

In particular $\{(\mu_n, Q_n)\} \subset \mathcal{M}_0$, $(\mu_n, Q_n) \to (\mu, Q)$ and, since $Q^{(1)} = Q^{(2)}$, $Q_n^{(1)} = Q_n^{(2)}$. In view of (2.4),

$$I(\mu_n, Q_n) = \iint_{K \times K} \lambda(dx)\, \lambda(dy)\, \Phi\big(q_n(x,y), \varrho_n(x) r(x) p(x,y)\big).$$

Since $r(x)p(x,y) > 0$ in $K \times K$ and $\varrho_n \geq c > 0$ in K, by dominated convergence we conclude that $I(\mu_n, Q_n) \to I(\mu, Q)$. $\square$

Lemma 5.5. *Let*

$$\begin{aligned} \mathcal{M}_2 := \Big\{ &(\mu, Q) \in \mathcal{M} \,:\, K := \mathrm{supp}(\mu) \subset E \setminus E_0,\ \mathrm{supp}(Q) = K \times K, \\ &\mu \ll \lambda,\ Q \ll \lambda \times \lambda \Big\}. \end{aligned} \tag{5.6}$$

The set $\mathcal{M}_1$ is I-dense in $\mathcal{M}_2$.

Proof: Given an integer n, pick a family of disjoint open sets $D_1^n, .., D_n^n \subset K$ such that $\lambda\big(K \setminus \bigcup_i D_i^n\big) = 0$ and the diameter of D_i^n vanishes as $n \to \infty$, $i = 1, \dots, n$. For $(\mu, Q) \in \mathcal{M}_2$ with $Q^{(1)} = Q^{(2)}$, let $d\mu = \varrho\, d\lambda$ and $dQ = q\, d\lambda \times \lambda$. We define $d\mu_n = \varrho_n d\lambda$ and $dQ_n = q_n d\lambda \times d\lambda$, with $\varrho_n = 0$ in $E \setminus K$, $q_n = 0$ in $(E \times E) \setminus (K \times K)$, and

$$\begin{aligned} \varrho_n(x) &:= \fint_{D_i^n} \varrho && \text{if } x \in D_i^n, \\ q_n(x,y) &:= \fint\!\!\fint_{D_i^n \times D_j^n} q && \text{if } (x,y) \in D_i^n \times D_j^n. \end{aligned}$$

In particular, since $Q^{(1)} = Q^{(2)}$, $Q_n^{(1)} = Q_n^{(2)}$. Moreover, $\{(\mu_n, Q_n)\} \subset \mathcal{M}_1$, and $(\mu_n, Q_n) \to (\mu, Q)$. In view of (2.4),

$$\begin{aligned} I(\mu_n, Q_n) &= \iint_{K \times K} \lambda(dx)\lambda(dy)\, \Phi\big(q_n(x,y), \varrho_n(x)\, r(x) p(x,y)\big) \\ &= \sum_{i,j} \lambda(D_i^n)\lambda(D_j^n) \fint\!\!\fint_{D_i^n \times D_j^n} \Big\{ q_n \log \frac{q_n}{\varrho_n} + q_n \log \frac{1}{rp} - \big(q_n - \varrho_n\, rp\big) \Big\}. \end{aligned}$$

By convexity of the function $a \log(a/b)$, using Jensen's inequality we get

$$q_n(x,y) \log \frac{q_n(x,y)}{\varrho_n(x)} \leq \fint\!\!\fint_{D_i^n \times D_j^n} q \log \frac{q}{\varrho}, \qquad (x,y) \in D_i^n \times D_j^n$$

so that

$$\begin{aligned} I(\mu_n, Q_n) \leq I(\mu, Q) &+ \sum_{i,j} \iint_{D_i^n \times D_j^n} d\lambda \times d\lambda\, \big(q - q_n\big) \log \frac{1}{rp} \\ &+ \sum_{i,j} \iint_{D_i^n \times D_j^n} d\lambda \times d\lambda\, \big(\varrho_n - \varrho\big)\, rp. \end{aligned}$$

Since both rp and $\log(1/rp)$ are continuous in $K \times K$, we conclude the proof taking $n \to \infty$. $\square$

The next lemma is the key step and relies on the technical condition (iv) in Assumption 2.1.

Lemma 5.6. *Let*

$$\mathcal{M}_3 := \Big\{(\mu, Q) \in \mathcal{M} \,:\, \mu \ll \lambda,\, Q \ll \lambda \times \lambda\Big\}. \tag{5.7}$$

The set $\mathcal{M}_2$ is I-dense in $\mathcal{M}_3$.

Proof: For $\delta > 0$ let $A_\delta \subset E$ be the open set defined by $A_\delta := \{x \in E : r(x) < \delta\}$. Given $(\mu, Q) \in \mathcal{M}_3$, with $Q^{(1)} = Q^{(2)}$, we write $d\mu = \varrho \, d\lambda$ and $dQ = q \, d\lambda \times d\lambda$. For $\delta > 0$, we set

$$\varrho_\delta(x) := \begin{cases} \varrho(x) & \text{if } \ x \in E \setminus A_{2\delta} \\ \frac{1}{\lambda(A_{2\delta} \setminus A_\delta)} \displaystyle\int_{A_{2\delta}} \lambda(dx') \, \varrho(x') & \text{if } \ x \in A_{2\delta} \setminus A_\delta \\ 0 & \text{if } \ x \in A_\delta, \end{cases}$$

and

$$q_\delta(x,y) := \begin{cases} q(x,y) & \text{if } \ (x,y) \in \big(E \setminus A_{2\delta}\big)^2 \\ \frac{1}{\lambda(A_{2\delta} \setminus A_\delta)} \displaystyle\int_{A_{2\delta}} \lambda(dy') \, q(x,y') & \text{if } \ (x,y) \in E \setminus A_{2\delta} \times A_{2\delta} \setminus A_\delta \\ \frac{1}{\lambda(A_{2\delta} \setminus A_\delta)} \displaystyle\int_{A_{2\delta}} \lambda(dx') \, q(x',y) & \text{if } \ (x,y) \in A_{2\delta} \setminus A_\delta \times E \setminus A_{2\delta} \\ \frac{1}{\big(\lambda(A_{2\delta} \setminus A_\delta)\big)^2} Q\big(A_{2\delta} \times A_{2\delta}\big) & \text{if } \ (x,y) \in \big(A_{2\delta} \setminus A_\delta\big)^2 \\ 0 & \text{if } \ x \in A_\delta \ \text{ or } \ y \in A_\delta. \end{cases}$$

By letting $d\mu_\delta := \varrho_\delta \, d\lambda$ and $dQ_\delta := q_\delta \, d\lambda \times d\lambda$, it follows that $(\mu_\delta,\, Q_\delta) \in \mathcal{M}_2$ and $(\mu_\delta,\, Q_\delta) \to (\mu,\, Q)$. Moreover, since $Q^{(1)} = Q^{(2)}$, $Q_\delta^{(1)} = Q_\delta^{(2)}$. In view of (2.4),

$$I(\mu_\delta,\, Q_\delta) = \iint d\lambda \times d\lambda \, \Phi\big(q_\delta,\, \varrho_\delta \, rp\big). \tag{5.8}$$

Consider first the integral over $(E \setminus A_{2\delta})^2$. By definition of (μ_δ, Q_δ),

$$\lim_{\delta \downarrow 0} \iint_{(E \setminus A_{2\delta})^2} d\lambda \times d\lambda \, \Phi\big(q_\delta, \varrho_\delta \, rp\big) = I(\mu, Q). \tag{5.9}$$

The proof of the lemma will be achieved by showing that the other contributions to the right hand side of (5.8) vanish ad $\delta \downarrow 0$.

Consider the integral over $E \setminus A_{2\delta} \times A_{2\delta} \setminus A_\delta$, namely

$$\begin{aligned} &\iint_{E \setminus A_{2\delta} \times A_{2\delta} \setminus A_\delta} d\lambda \times d\lambda \, \Phi\big(q_\delta,\, \varrho_\delta \, rp\big) \\ &= \frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta} \setminus A_\delta)} \iint_{E \setminus A_{2\delta} \times A_{2\delta} \setminus A_\delta} \lambda(dx) \lambda(dy) \, \Phi\Big(\fint_{A_{2\delta}} q(x,\cdot),\, \fint_{A_{2\delta}} \varrho(x) \, r(x) p(x,\cdot)\Big) + R_\delta \end{aligned}$$

where

$$
\begin{aligned}
R_\delta = \iint_{E\setminus A_{2\delta}\times A_{2\delta}\setminus A_\delta} &\lambda(dx)\lambda(dy) \\
&\bigg\{ \frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)} \fint_{A_{2\delta}} q(x,\cdot)\log \frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)} \frac{\fint_{A_{2\delta}} r(x)p(x,\cdot)}{r(x)p(x,y)} \\
&\quad - \Big(\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)} \fint_{A_{2\delta}} \varrho(x)r(x)p(x,\cdot) - \varrho(x)r(x)p(x,y)\Big)\bigg\}.
\end{aligned}
$$

By choosing $\delta = \delta_n$ as in condition (iv) of Assumption 2.1 and using that p is strictly positive, it follows that $\lim_n R_{\delta_n} = 0$. Moreover, by Jensen inequality,

$$
\begin{aligned}
&\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)} \iint_{E\setminus A_{2\delta}\times A_{2\delta}\setminus A_\delta} \lambda(dx)\lambda(dy)\, \Phi\Big(\fint_{A_{2\delta}} q(x,\cdot), \fint_{A_{2\delta}} \varrho(x)\, r(x)p(x,\cdot)\Big) \\
&\qquad \le \iint_{E\setminus A_{2\delta}\times A_{2\delta}} d\lambda\times d\lambda\, \Phi(q,\varrho\, rp).
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\varlimsup_n \iint_{E\setminus A_{2\delta_n}\times A_{2\delta_n}\setminus A_{\delta_n}} d\lambda\times d\lambda\, \Phi\big(q_{\delta_n}, \varrho_{\delta_n} rp\big) \\
&\le \lim_n \iint_{E\setminus A_{2\delta_n}\times A_{2\delta_n}} d\lambda\times d\lambda\, \Phi(q,\varrho\, rp) = 0
\end{aligned}
\tag{5.10}
$$

We next consider the integral over $A_{2\delta}\setminus A_\delta\times E\setminus A_{2\delta}$, namely

$$
\begin{aligned}
&\iint_{A_{2\delta}\setminus A_\delta\times E\setminus A_{2\delta}} d\lambda\times d\lambda\, \Phi\big(q_\delta, \varrho_\delta\, rp\big) \\
&= \frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)} \iint_{A_{2\delta}\setminus A_\delta\times E\setminus A_{2\delta}} \lambda(dx)\lambda(dy)\, \Phi\Big(\fint_{A_{2\delta}} q(\cdot,y), \fint_{A_{2\delta}} \varrho(\cdot)r(\cdot)p(\cdot,y)\Big) + R_\delta,
\end{aligned}
$$

where

$$
\begin{aligned}
R_\delta = &\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)} \iint_{A_{2\delta}\setminus A_\delta\times E\setminus A_{2\delta}} \lambda(dx)\lambda(dy) \bigg\{ \fint_{A_{2\delta}} q(\cdot,y)\log \frac{\fint_{A_{2\delta}} \varrho(\cdot)\, r(\cdot)p(\cdot,y)}{\fint_{A_{2\delta}} \varrho(\cdot)\, r(x)p(x,y)} \\
&\qquad + \fint_{A_{2\delta}} \varrho(\cdot)\, r(x)p(x,y) - \fint_{A_{2\delta}} \varrho(\cdot)\, r(\cdot)p(\cdot,y)\bigg\}.
\end{aligned}
$$

By condition (ii) in Assumption 2.1, there exists $c>0$ such that $p(x,y)\ge c$. We thus deduce

$$
R_\delta \le \frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)} \iint_{A_{2\delta}\setminus A_\delta\times E\setminus A_{2\delta}} \lambda(dx)\lambda(dy) \fint_{A_{2\delta}} q(\cdot,y)\log \frac{\|rp\|}{cr(x)} + \|rp\|\,\mu(A_{2\delta}).
$$

By definition of the set A_δ, $r(x)>\delta$ for $x\in A_{2\delta}\setminus A_\delta$. Hence

$$
\begin{aligned}
R_\delta &\le \iint_{A_{2\delta}\times E\setminus A_{2\delta}} \lambda(dx)\lambda(dy)\, q(x,y)\log \frac{\|rp\|}{c\delta} + \|rp\|\,\mu(A_{2\delta}) \\
&\le \iint_{A_{2\delta}\times E\setminus A_{2\delta}} \lambda(dx)\lambda(dy)\, q(x,y)\log \frac{2\|rp\|}{cr(x)} + \|rp\|\,\mu(A_{2\delta}),
\end{aligned}
$$

where in the last inequality we used $r(x)<2\delta$ for $x\in A_{2\delta}$. In view of Lemma 4.5 we conclude that $\varlimsup_{\delta\downarrow 0} R_\delta \le 0$. By using Jensen inequality as in the previous step we deduce

$$\lim_{\delta\downarrow 0}\iint_{A_{2\delta}\setminus A_\delta\times E\setminus A_{2\delta}} d\lambda\times d\lambda\;\Phi\big(q_\delta,\varrho_\delta\, rp\big)=0. \tag{5.11}$$

We finally consider the integral over $(A_{2\delta}\setminus A_\delta)^2$, namely

$$\begin{aligned}&\iint_{(A_{2\delta}\setminus A_\delta)^2} d\lambda\times d\lambda\;\Phi\big(q_\delta,\varrho_\delta\, rp\big)\\ &=\Big(\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\Big)^2\iint_{(A_{2\delta}\setminus A_\delta)^2} d\lambda\times d\lambda\,\Phi\Big(⨌_{(A_{2\delta})^2} q, ⨌_{(A_{2\delta})^2}\varrho\, rp\Big)+R_\delta,\end{aligned}$$

where

$$\begin{aligned}R_\delta=\iint_{(A_{2\delta}\setminus A_\delta)^2}&\lambda(dx)\lambda(dy)\\ &\Big\{\Big(\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\Big)^2⨌_{(A_{2\delta})^2} q\log\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\frac{⨌_{(A_{2\delta})^2}\varrho\, rp}{⨍_{A_{2\delta}}\varrho\, r(x)p(x,y)}\\ &\quad+\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}⨍_{A_{2\delta}}\varrho\, r(x)p(x,y)-\Big(\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\Big)^2⨌_{(A_{2\delta})^2}\varrho\, rp\Big\}.\end{aligned}$$

As in the previous step, we now use that $p(x,y)\ge c$ for some $c>0$. We deduce

$$\begin{aligned}R_\delta\le&\Big(\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\Big)^2\iint_{(A_{2\delta}\setminus A_\delta)^2}\lambda(dx)\lambda(dy)⨌_{(A_{2\delta})^2} q\,\log\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\frac{\|rp\|}{cr(x)}\\ &+\|rp\|\,\mu(A_{2\delta})\,\lambda(A_{2\delta}\setminus A_\delta).\end{aligned}$$

Since $r(x)>\delta$ for $x\in A_{2\delta}\setminus A_\delta$,

$$\begin{aligned}R_\delta&\le\iint_{(A_{2\delta})^2}\lambda(dx)\lambda(dy)\,q(x,y)\log\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\frac{\|rp\|}{c\delta}+\|rp\|\,\mu(A_{2\delta})\,\lambda(A_{2\delta}\setminus A_\delta)\\ &\le\iint_{(A_{2\delta})^2}\lambda(dx)\lambda(dy)\,q(x,y)\log\frac{\lambda(A_{2\delta})}{\lambda(A_{2\delta}\setminus A_\delta)}\frac{2\|rp\|}{cr(x)}+C\,\mu(A_{2\delta})\,\lambda(A_{2\delta}\setminus A_\delta),\end{aligned}$$

where in the last inequality we used $r(x)<2\delta$ for $x\in A_{2\delta}$. By choosing $\delta=\delta_n$, where δ_n is the sequence in condition (iv) of Assumption 2.1 and using Lemma 4.5, we deduce that $\varlimsup_n R_{\delta_n}\le 0$. By using Jensen inequality as in the previous steps we conclude

$$\lim_{n\to\infty}\iint_{(A_{2\delta_n}\setminus A_{\delta_n})^2} d\lambda\times d\lambda\;\Phi\big(q_{\delta_n},\varrho_{\delta_n}\, rp\big)=0 \tag{5.12}$$

Since ϱ_δ vanishes on A_δ and q_δ vanishes on $\big(A_\delta\times E\big)\cup\big(E\times A_\delta\big)$, (5.9)-(5.12) yield the statement. □

Lemma 5.7. *Let*

$$\mathcal{M}_4:=\big\{(\mu,Q)\in\mathcal{M}\,:\,\mu\perp\lambda,\,Q=0\big\}, \tag{5.13}$$

The set $\mathcal{M}_3$ is I-dense in $\mathcal{M}_4$.

Proof: Consider a Borel partition $E=\bigcup_i D_i^n$ such that $\lambda(D_i^n)>0$ and the diameter of D_i^n vanishes as $n\to\infty$. Given $\mu\perp\lambda$, we set $d\mu_n=\varrho_n d\lambda$, where

$$\varrho_n(x)=\frac{\mu\big(D_i^n\big)}{\lambda\big(D_i^n\big)},\qquad x\in D_i^n$$

and $Q_n = 0$. In particular $(\mu_n, Q_n) \in \mathcal{M}_3$ and $(\mu_n, Q_n) \to (\mu, 0)$. Moreover

$$\lim_{n\to\infty} I(\mu_n, Q_n) = \lim_{n\to\infty} \iint \lambda(dx)\,\lambda(dy) \varrho_n(x) r(x) p(x,y) = \mu(r) = I(\mu, 0)$$

where we used the representation (2.4) for $I(\mu, Q)$. □

Proof of Theorem 5.3: Let $(\mu, Q) \in \mathcal{M}$ be such that $I(\mu, Q) < +\infty$. We decompose μ into the absolutely continuous and singular parts with respect to λ, i.e. $\mu = \mu_{\mathrm{ac}} + \mu_{\mathrm{s}}$ and we recall that by Lemma 4.4 $Q \ll Q^{\mu_{\mathrm{ac}}}$. In particular, letting $\alpha = \mu_{\mathrm{ac}}(E)$,

$$(\mu, Q) = \alpha \left(\frac{1}{\alpha}\mu_{\mathrm{ac}}, \frac{1}{\alpha}Q\right) + (1-\alpha)\Big(\frac{1}{1-\alpha}\mu_{\mathrm{s}}, 0\Big).$$

In view of Lemmata 5.4–5.6 there exists a sequence $\{(\mu_{1,n}, Q_{1,n})\} \subset \mathcal{M}_0$ such that $(\mu_{1,n}, Q_{1,n}) \to (\alpha^{-1}\mu_{\mathrm{ac}}, \alpha^{-1}Q)$ and $I(\mu_{1,n}, Q_{1,n}) \to I\big(\alpha^{-1}\mu_{\mathrm{ac}}, \alpha^{-1}Q\big)$. Moreover, by Lemmata 5.4–5.7, there exists a sequence $\{(\mu_{2,n}, Q_{2,n})\} \subset \mathcal{M}_0$ such that $(\mu_{2,n}, Q_{2,n}) \to ((1-\alpha)^{-1}\mu_{\mathrm{s}}, 0)$ and $I(\mu_{2,n}, Q_{2,n}) \to I\big((1-\alpha)^{-1}\mu_{\mathrm{s}}, 0\big)$.

The sequence $\{\alpha(\mu_{1,n}, Q_{1,n}) + (1-\alpha)(\mu_{2,n}, Q_{2,n})\}$ is in $\mathcal{M}_0$ and converges to (μ, Q). By the convexity of I,

$$I\big(\alpha(\mu_{1,n}, Q_{1,n}) + (1-\alpha)(\mu_{2,n}, Q_{2,n})\big) \le \alpha I(\mu_{1,n}, Q_{1,n}) + (1-\alpha) I(\mu_{2,n}, Q_{2,n})$$

so that

$$\begin{aligned} &\varlimsup_n I\big(\alpha(\mu_{1,n}, Q_{1,n}) + (1-\alpha)(\mu_{2,n}, Q_{2,n})\big) \\ &\qquad \le \alpha I(\alpha^{-1}\mu_{\mathrm{ac}}, \alpha^{-1}Q\big) + (1-\alpha) I((1-\alpha)^{-1}\mu_{\mathrm{s}}, 0\big) = I(\mu, Q) \end{aligned}$$

where we used the representation (2.4) in the last equality. □

Proof of Theorem 2.3 (conclusion). The upper bound follows from Proposition 4.1 and Lemma 4.4, which also yields the convexity and lower semi-continuity of I. Recalling (5.3), Lemma 5.1 and Proposition 5.2 imply the uniform lower bound with rate function $\mathrm{sc}^- J$. In view of the lower semi-continuity of I and Theorem 5.3 we conclude $\mathrm{sc}^- J = I$. Finally, the goodness of the rate function I follows from the exponential tightness proven in Lemma 4.2 and Dembo and Zeitouni (1998, Lemma 1.2.18). □

6. Projections

Large deviations of the empirical measure. In the context of irreducible finite state Markov chain, the representation of the Donsker-Varadhan functional in terms of I has been obtained in Baldi and Piccioni (1999); Kipnis and Landim (1999). This result has been proven for countable state space in Bertini et al. (2014). The proof presented below relies on the variational representation of Lemma 4.4 and on the Sion's minimax theorem. It takes advantage of the compactness of E.

Proof of Corollary 2.4: Let $I_1 \colon \mathcal{M}_1(E) \to [0, +\infty]$ be the functional

$$I_1(\mu) := \inf_{Q \in \mathcal{M}_+(E\times E)} I(\mu, Q). \tag{6.1}$$

By contraction principle and Theorem 2.3, as $T \to +\infty$ the family $\big\{\mathbb{P}_x \circ \mu_T^{-1}\big\}_{T>0}$ satisfy a large deviation principle with rate function I_1. To complete the proof it is therefore enough to show $\widehat{I} = I_1$.

We first prove the inequality $I_1 \geq \widehat{I}$. In view of (2.3) we can restrict the infimum on the right hand side of (6.1) to elements $Q \in \mathcal{M}_+(E \times E)$ satisfying $Q^{(1)} = Q^{(2)}$. For such elements, by the variational characterization of I proven in Lemma 4.4,

$$I(\mu, Q) = \sup_{F \in C(E \times E)} \big\{ Q(F) - \mu\big(r^F - r\big) \big\}.$$

Fix $f \in C(E)$ and choose $F(x, y) = f(y) - f(x)$, $(x, y) \in E \times E$. Since $Q^{(1)} = Q^{(2)}$,

$$I(\mu, Q) \geq -\iint \mu(dx) c(x, dy) \left[e^{f(y) - f(x)} - 1 \right].$$

As the right hand side does not depend on Q we deduce

$$I_1(\mu) \geq -\iint \mu(dx) c(x, dy) \left[e^{f(y) - f(x)} - 1 \right]$$

and the result follows by optimizing on f.

We next prove the inequality $I_1 \leq \widehat{I}$. Fix $\mu \in \mathcal{M}_1(E)$ and observe that $I_1(\mu) \leq I(\mu, 0) = \mu(r) < +\infty$. By Lemma 4.4,

$$I_1(\mu) = \inf_Q \sup_{\phi, F} \Gamma_\mu(Q, \phi, F)$$

where the infimum is carried out over all $Q \in \mathcal{M}_+(E \times E)$, the supremum over all $(\phi, F) \in C(E) \times C(E \times E)$, and $\Gamma_\mu \colon \mathcal{M}_+(E \times E) \times C(E) \times C(E \times E) \to \mathbb{R}$ is the continuous functional defined by

$$\Gamma_\mu(Q, \phi, F) = Q^{(1)}(\phi) - Q^{(2)}(\phi) + Q(F) - \mu(r^F - r).$$

As follows from a direct application of Hölder inequality, the map $F \mapsto \mu(r^F)$ is convex. Hence, for each Q the map $(\phi, F) \mapsto \Gamma_\mu(Q, \phi, F)$ is concave. Since for each (ϕ, F) the map $Q \mapsto \Gamma_\mu(Q, \phi, F)$ is affine, we would like to apply the Sion's minimax theorem to get

$$I_1(\mu) = \sup_{\phi, F} \inf_Q \Gamma_\mu(Q, \phi, F). \tag{6.2}$$

Since neither $\mathcal{M}_+(E \times E)$ nor $C(E) \times C(E \times E)$ is compact, (6.2) needs however to be justified. Postponing this step, we first conclude the argument. By choosing on the right hand side of (6.2) $Q(dx, dy) = \mu(dx) c(x, dy) e^{F(x,y)}$ we get

$$\begin{aligned} I_1(\mu) &\leq \sup_{\phi, F} \iint \mu(dx) c(x, dy) \big[e^{F(x,y)} \big(F(x, y) + \phi(x) - \phi(y) \big) - \big(e^{F(x,y)} - 1 \big) \big] \\ &\leq \sup_{\phi} \iint \mu(dx) c(x, dy) \sup_{\lambda \in \mathbb{R}} \big[e^{\lambda} \big(\lambda + \phi(x) - \phi(y) \big) - \big(e^{\lambda} - 1 \big) \big] \\ &= \sup_{\phi} \Big\{ -\iint \mu(dx) c(x, dy) \left[e^{\phi(y) - \phi(x)} - 1 \right] \Big\} = \widehat{I}(\mu). \end{aligned}$$

We are left with the proof of (6.2). To this end we apply the generalization of the Sion minimax theorem proven in Ha (1981) that states the following. Under the continuity and convexity/concavity assumptions discussed before, a sufficient condition for the minimax identity

$$\inf_Q \sup_{\phi, F} \Gamma_\mu(Q, \phi, F) = \sup_{\phi, F} \inf_Q \Gamma_\mu(Q, \phi, F)$$

is that there exist a nonempty convex compact $K \subset C(E) \times C(E \times E)$ and a compact $\mathcal{H} \subset \mathcal{M}_+(E \times E)$ such that

$$\inf_Q \sup_{\phi,F} \Gamma_\mu(Q,\phi,F) \le \inf_{Q \notin \mathcal{H}} \sup_{(\phi,F)\in K} \Gamma_\mu(Q,\phi,F). \tag{6.3}$$

We choose $\mathcal{H} = \big\{Q \in \mathcal{M}_+(E\times E) : \|Q\|_{\mathrm{TV}} \le h\big\}$ (here $\|Q\|_{\mathrm{TV}}$ is the total mass of Q) for some $h > 0$ to be fixed later and let K be the singleton $K = \{(0,1)\}$. If $Q \notin \mathcal{H}$ then

$$\Gamma_\mu(Q,0,1) = Q(1) - (e-1)\mu(r) \ge h - (e-1)\mu(r).$$

Since, as already observed,

$$\inf_Q \sup_{\phi,F} \Gamma_\mu(Q,\phi,F) \le I(\mu,0) = \mu(r),$$

by choosing $h \ge e\mu(r)$ the condition (6.3) holds and we have concluded the proof of (6.2). □

Large deviations of the empirical flow.

Proof of Corollary 2.5: Let $I_2 \colon \mathcal{M}_+(E\times E) \to [0,+\infty]$ be the functional

$$I_2(Q) := \inf_{\mu\in\mathcal{M}_1(E)} I(\mu,Q). \tag{6.4}$$

By contraction principle and Theorem 2.3, as $T \to +\infty$ the family $\big\{\mathbb{P}_x \circ Q_T^{-1}\big\}_{T>0}$ satisfy a large deviation principle with rate function I_2. To complete the proof it is therefore enough to show $\widetilde{I} = I_2$.

We first prove the inequality $I_2 \ge \widetilde{I}$. We use the variational characterization (4.5) restricting to Q with equal marginals. Given $\alpha \in (-r_{\mathrm{m}}, +\infty)$, we chose

$$F(x,y) = \log\Big[\frac{Q(dx,dy)}{Q(dx,E)c(x,dy)}(r(x)+\alpha)\Big].$$

By direct computations $Q^\mu(e^F - 1) = \alpha$, so that

$$\begin{aligned} I(\mu,Q) &\ge Q(F) - Q^\mu\big(e^F - 1\big) \\ &= \iint Q(dx,dy)\log\Big[\frac{Q(dx,dy)}{Q(dx,E)c(x,dy)}(r(x)+\alpha)\Big] - \alpha. \end{aligned}$$

The result follows by optimizing over α. We observe the choice of F is not really legal since it could be not continuous, however a truncation procedure similar to the one in Lemma 4.4 leads to the same conclusion.

We next prove $I_2 \le \widetilde{I}$. By definition of $\widetilde{I}$ we can assume that Q has equal marginals. Given Q, if there exists $\alpha > -r_{\mathrm{m}}$ such that $\iint Q(dx,E)/(r(x)+\alpha) = 1$ we chose

$$\mu^Q(dx) = \frac{Q(dx,E)}{r(x)+\alpha},$$

then $I_2(Q) \le I(\mu^Q,Q)$. By a direct computation

$$I(\mu^Q,Q) = \iint Q(dx,dy)\log\Big[\frac{Q(dx,dy)}{Q(dx,E)c(x,dy)}(r(x)+\alpha)\Big] - \alpha \le \widetilde{I}(Q).$$

If such α does not exists, by monotone convergence

$$\int \frac{Q(dx,E)}{r(x)-r_{\mathrm{m}}} \le 1$$

and we can chose

$$\mu^Q(dx) = \frac{Q(dx, E)}{r(x) - r_{\mathrm{m}}} + \Big(1 - \iint \frac{Q(dz, E)}{r(z) - r_{\mathrm{m}}}\Big)\delta_{x_0}(dx),$$

with x_0 such that $r(x_0) = r_{\mathrm{m}}$. From (2.4) by direct computation we obtain

$$I(\mu^Q, Q) = \iint Q(dx, dy) \log\Big[\frac{Q(dx, dy)}{Q(dx, E)c(x, dy)}(r(x) - r_{\mathrm{m}})\Big] + r_{\mathrm{m}} \le \widetilde{I}(Q),$$

where the last inequality follows by monotone convergence. □

Acknowledgements. We thank the referee for suggesting us to formulate result for a general (closed) set E_0.

References

M. Baiesi, C. Maes and K. Netočný. Computation of current cumulants for small nonequilibrium systems. *J. Stat. Phys.* **135** (1), 57–75 (2009). MR2505725.

P. Baldi and M. Piccioni. A representation formula for the large deviation rate function for the empirical law of a continuous time Markov chain. *Statist. Probab. Lett.* **41** (2), 107–115 (1999). MR1665260.

G. Basile. From a kinetic equation to a diffusion under an anomalous scaling. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** (4), 1301–1322 (2014). MR3269995.

G. Basile and A. Bovier. Convergence of a kinetic equation to a fractional diffusion equation. *Markov Process. Related Fields* **16** (1), 15–44 (2010). MR2664334.

G. Basile, S. Olla and H. Spohn. Energy transport in stochastically perturbed lattice dynamics. *Arch. Ration. Mech. Anal.* **195** (1), 171–203 (2010). MR2564472.

L. Bertini, A. Faggionato and D. Gabrielli. From level 2.5 to level 2 large deviations for continuous time Markov chains. *Markov Process. Related Fields* **20** (3), 545–562 (2014). MR3289132.

L. Bertini, A. Faggionato and G. Gabrielli. Asymptotic independence of multiple Wiener-Itô integrals and the resulting limit laws (2015a). To appear in Ann. Inst. H. Poincaré Probab. Statist.

L. Bertini, A. Faggionato and G. Gabrielli. Flows, currents and symmetries for continuous time markov chains: a large deviations approach. (2015b). To appear in Stochastic Processes Appl.

T. Bodineau, V. Lecomte and C. Toninelli. Finite size scaling of the dynamical free-energy in a kinetically constrained model. *J. Stat. Phys.* **147** (1), 1–17 (2012). MR2922756.

T. Bodineau and C. Toninelli. Activity phase transition for constrained dynamics. *Comm. Math. Phys.* **311** (2), 357–396 (2012). MR2902193.

P. Brémaud. *Point processes and queues.* Springer-Verlag, New York-Berlin (1981). ISBN 0-387-90536-7. Martingale dynamics, Springer Series in Statistics. MR636252.

A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition (1998). ISBN 0-387-98406-2. MR1619036.

J.-D. Deuschel and D.W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA (1989). ISBN 0-12-213150-9. MR997938.

M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.* **28**, 1–47; ibid. 28 (1975), 279–301 (1975). MR0386024.

M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** (4), 389–461 (1976). MR0428471.

M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** (2), 183–212 (1983). MR690656.

C.W. Ha. A noncompact minimax theorem. *Pacific J. Math.* **97** (1), 115–117 (1981). MR638178.

M. Jara, T. Komorowski and S. Olla. Limit theorems for additive functionals of a Markov chain. *Ann. Appl. Probab.* **19** (6), 2270–2300 (2009). MR2588245.

L.H. Jensen. *Large deviations of the asymmetric simple exclusion process in one dimension*. ProQuest LLC, Ann Arbor, MI (2000). ISBN 978-0599-75951-0. Thesis (Ph.D.)–New York University. MR2700635.

G. Kesidis and J. Walrand. Relative entropy between Markov transition rate matrices. *IEEE Trans. Inform. Theory* **39** (3), 1056–1057 (1993). MR1237729.

C. Kipnis and C. Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin (1999). ISBN 3-540-64913-1. MR1707314.

J.L. Lebowitz and H. Spohn. A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics. *J. Statist. Phys.* **95** (1-2), 333–365 (1999). MR1705590.

R. Lefevere, M. Mariani and L. Zambotti. Large deviations for renewal processes. *Stochastic Process. Appl.* **121** (10), 2243–2271 (2011a). MR2822776.

R. Lefevere, M. Mariani and L. Zambotti. Large deviations of the current in stochastic collisional dynamics. *J. Math. Phys.* **52** (3), 033302, 22 (2011b). MR2814865.

C. Maes. The fluctuation theorem as a Gibbs property. *J. Statist. Phys.* **95** (1-2), 367–392 (1999). MR1705591.

R.E. Megginson. *An introduction to Banach space theory*, volume 183 of *Graduate Texts in Mathematics*. Springer-Verlag, New York (1998). ISBN 0-387-98431-3. MR1650235.

S. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition (2009). ISBN 978-0-521-73182-9. With a prologue by Peter W. Glynn. MR2509253.

H. Spohn. The phonon Boltzmann equation, properties and link to weakly anharmonic lattice dynamics. *J. Stat. Phys.* **124** (2-4), 1041–1104 (2006). MR2264633.