Alea 6, 199–238 (2009)

On characterizations of Metropolis type algorithms in continuous time

Persi Diaconis and Laurent Miclo

Departments of Statistics and Mathematics, Stanford University, USA

Laboratoire d'Analyse, Topologie, Probabilités, UMR 6632, Université de Provence and CNRS, France

Abstract. In the continuous time framework, a new definition is proposed for the Metropolis algorithm $(\widetilde{X}_t)_{t\geq 0}$ associated to an a priori given exploratory Markov process $(X_t)_{t\geq 0}$ and to a tarjet probability distribution π. It should be the minimizer for the relative entropy of the trajectorial law of $(\widetilde{X}_t)_{t\in[0,T]}$ with respect to the law of $(X_t)_{t\in[0,T]}$, when both processes start with π as initial law and when π is assumed to be reversible for $(\widetilde{X}_t)_{t\geq 0}$. This definition doesn't depend on the time horizon $T > 0$ and the corresponding minimizing process is not difficult to describe. Even if this procedure can be made general, the details were only worked out in situation of finite jump processes and of compact manifold-valued diffusion processes (a sketch is also given for Markov processes admitting both a diffusive part and a jump part). The proofs rely on an alternative approach to general Girsanov transformations in the spirit of Kunita. The case of φ-relative entropies is also investigated, in particular to make a link with a previous work of Billera and Diaconis on the traditional Metropolis algorithm in the discrete time setting.

1. Introduction

The Metropolis algorithm is a very popular Monte Carlo procedure to sample approximatively according to a given law π (cf. for instance Metropolis et al., 1953, Hammersley and Handscomb, 1965, Fishman, 1996 or Liu, 2008). In discrete time and finite state space, it starts with a Markov kernel K and transform it into a kernel K_π which is reversible with respect to π. Under mild assumptions, K_π is ergodic so that π can be approached by simulating a Markov chain whose transitions are didacted by K_π. Billera and Diaconis (2001) took a geometrical point of view on the construction of K_π by introducing a distance d on the set of Markov kernels

Received by the editors March 2, 2009; accepted June 16, 2009.

2000 *Mathematics Subject Classification.* 60J22, 60J25, 60J27, 60J57, 94A17, 58J65.

Key words and phrases. Metropolis algorithms, finite jump Markov processes, compact manifold-valued diffusion processes, relative entropy minimizations, φ-relative entropies, discrepancies on sets of Markov generators, martingale problems, general Girsanov transformations.

and by showing that K_π is a minimizer for the distance from K to the set of Markov kernels which are reversible with respect to π. The goal of this paper is extend this point of view to continuous time, in particular via new entropy-type discrepancies among generators.

We begin by recalling more precisely the main result of Billera and Diaconis (2001). Let S be a finite set and denote by $\mathcal{K}$ the set of $S \times S$ Markov matrices. We assume that we are given π a positive probability measure on S and $K \in \mathcal{K}$. The associated Metropolis kernel K_π is defined by

$$\forall\; x, y \in S, \qquad K_\pi(x,y) \quad := \quad \begin{cases} \min\left\{K(x,y), \frac{\pi(y)}{\pi(x)}K(y,x)\right\} & , \text{ if } x \neq y, \\ 1 - \sum_{z\in S\setminus\{x\}} K_\pi(x,z) & , \text{ if } x = y. \end{cases}$$

Consider the distance d on $\mathcal{K}$ defined by

$$\forall\; K, K' \in \mathcal{K}, \qquad d(K', K) \quad := \quad \sum_{x\in S} \pi(x) \sum_{y \in S\setminus\{x\}} |K'(x,y) - K(x,y)|$$

and $\mathcal{K}(\pi)$ the subset of $\mathcal{K}$ consisting of Markov matrices M which are reversible with respect to π, namely satisfying

$$\forall\; x, y \in S, \qquad \pi(x)M(x,y) \quad = \quad \pi(y)M(y,x).$$

Let us also denote $\mathcal{K}(\pi, K)$ the subset of $\mathcal{K}(\pi)$ consisting of Markov matrices M whose off-diagonal entries are less or equal than those of K.
Billera and Diaconis (2001) proved the following result:

Theorem 1.1. *With respect to d, the Metropolis kernel K_π minimizes the distance from K to $\mathcal{K}(\pi)$ and it is the unique minimizer of the distance from K to $\mathcal{K}(\pi, K)$.*

The arguments of Billera and Diaconis (2001) enable to extend immediately this result to the corresponding continuous time setting. Let $\mathcal{L}$ be the set of (Markov) generators on S, i.e the $S \times S$ matrices whose off-diagonal entries are nonnegative and such that the sums along the lines are zero. Let π be fixed as above and $L \in \mathcal{L}$ be given. Define the distance d on $\mathcal{L}$ by the same formula as before and similarly consider $\mathcal{L}(\pi)$ (respectively $\mathcal{L}(\pi, L)$) the subset of $\mathcal{L}$ consisting of generators which are reversible with respect to π (resp. and whose off-diagonal entries are dominated by those of L). Finally the associated Metropolis generator L_π is defined by

$$\forall\; x, y \in S, \qquad L_\pi(x,y) \quad := \quad \begin{cases} \min\left\{L(x,y), \frac{\pi(y)}{\pi(x)}L(y,x)\right\} & , \text{ if } x \neq y, \\ -\sum_{z\in S\setminus\{x\}} L_\pi(x,z) & , \text{ if } x = y. \end{cases} \tag{1.1}$$

Then we have

Proposition 1.2. *With respect to d, the Metropolis generator L_π minimizes the distance from L to $\mathcal{L}(\pi)$ and it is the unique minimizer of the distance from L to $\mathcal{L}(\pi, L)$.*

But this result is not satisfying, because it does not admit an obvious extension to more general generators, for instance to diffusion generators, whereas there is often an analogue to the Metropolis construction in such a setting. For example in the Euclidean space $\mathbb{R}^d$, let π be a probability measure with a smooth and positive density (still denoted π) with respect to the Lebesgue measure and consider the generator $L := \triangle/2$ of the Brownian motion. Then it is well-known that the corresponding Metropolis generator should be the Langevin operator $L_\pi\cdot :=$ $(\triangle\cdot + \langle\nabla\ln(\pi), \nabla\cdot\rangle)/2$. One of our objectives is to propose a geometric justification of the latter kind of assertion, in the same spirit as Proposition 1.2, by providing abstract definitions of Metropolis type algorithms associated to a given probability measure and to a given Markov process. This will lead us to replace the distance d by other discrepancy quantities. To keep the development of this subject to a reasonable size and to avoid technicalities, we postpone the treatment of general Markov processes to a future work. Here we will only consider finite state space jump processes and regular diffusions taking values in a compact manifold, but we will adopt a general formalism each time it is possible without too much digression. Nevertheless, for the remaining part of this introduction, we will keep working in the finite state space setting and the analogous results concerning compact Riemannian diffusions will be presented in Section 4.

As above Proposition 1.2, L stands for an a priori given generator and we denote by M another generic generator. For any probability measure μ on S, we consider $(X^{(\mu)}(t))_{t\geq 0}$ (respectively $(Y^{(\mu)}(t))_{t\geq 0}$) a Markov process on S whose initial distribution is μ and whose generator is L (resp. M). Without real loss of generality, their trajectories are assumed to be càdlàg, namely admitting left limits and being right continuous and for any $T\in\mathbb{R}_+$, we denote by $\mathcal{L}(X^{(\mu)}([0,T]))$ the law of $X^{(\mu)}([0,T]) := (X^{(\mu)}(t))_{t\in[0,T]}$ on the set of càdlàg trajectories from $[0,T]$ to S. Recall that the relative entropy of two probability measures μ and ν given on the same probability space is defined by

$$\mathrm{Ent}(\mu|\nu) \quad := \quad \begin{cases} \int \frac{d\mu}{d\nu}\ln\left(\frac{d\mu}{d\nu}\right) d\nu \ \leq\ +\infty & \text{, if } \mu\ll\nu, \\ +\infty & \text{otherwise.} \end{cases}$$

where $d\mu/d\nu$ stands for the Radon-Nikodym derivative of μ with respect to ν. This notion is convenient to introduce naturally a new discrepancy (by this term we mean a $\mathbb{R}_+\sqcup\{+\infty\}$-valued quantity measuring in some sense if two probability distributions are close, but without satisfying the axioms of a distance) on $\mathcal{L}$. As above, π is a given positive probability measure on S, π is said to be invariant for the generator M if we have

$$\forall\ y\in S, \qquad \sum_{x\in S}\pi(x)M(x,y) \quad = \quad 0.$$

Proposition 1.3. *If π is invariant for M, for any $T\geq 0$, we have*

$$\mathrm{Ent}(\mathcal{L}(Y^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) \quad = \quad T\widetilde{d}(M,L) \ \leq\ +\infty,$$

where $\widetilde{d}$ is the discrepancy defined by

$$\forall\ M,M'\in\mathcal{L},$$

$$\widetilde{d}(M',M) \quad := \quad \sum_{x\in S}\pi(x)\sum_{y\in S\setminus\{x\}} M'(x,y)\ln\left(\frac{M'(x,y)}{M(x,y)}\right) - M'(x,y) + M(x,y)$$

To get a better justification of the introduction of the discrepancy $\widetilde{d}$, we are lacking a "computational complexity" interpretation of the relative entropy of two probability measures μ and ν, something saying heuristically that "$\mathrm{Ent}(\mu|\nu)$ is a measurement of the difficulty (or maybe of the necessary quantity of additional randomness) to simulate according to μ when we know how to simulate from ν". Sanov's theorem (see for instance the book of Dembo and Zeitouni, 1998) goes in this direction, we would like a more simulation oriented result. The above proposition may still seems disturbing, since we are using chains starting from the distribution π we want to approximate! But the discrepancy $\widetilde{d}$ should be seen as an asymptotic object: we will show that if π is furthermore assumed to be attractive for M, then we have for any initial distribution μ,

$$\lim_{T\to+\infty} \mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T]))|\mathcal{L}(X^{(\mu)}([0,T])))/T \quad = \quad \widetilde{d}(M,L). \tag{1.2}$$

We can use the discrepancy $\widetilde{d}$ to define another Metropolis type generator:

Theorem 1.4. *The mapping $\mathcal{L}(\pi) \ni M \mapsto \widetilde{d}(M,L)$ admits a unique minimizer $\widetilde{L}_\pi$ which is given by*

$$\forall\; x,y\in S,\qquad \widetilde{L}_\pi(x,y) \quad := \quad \begin{cases} \sqrt{\frac{\pi(y)}{\pi(x)}}\sqrt{L(x,y)L(y,x)} & ,\ \text{if } x\neq y \\ -\sum_{z\in S\setminus\{x\}} \widetilde{L}_\pi(x,z) & ,\ \text{if } x=y \end{cases}$$

In particular, we don't find the usual Metropolis generator L_π as in Proposition 1.2. To make a link between them, we need to consider more general projection procedures.

Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a convex function satisfying $\varphi(1)=0$, $\varphi'(1)=0$ (so φ is assumed to admits a derivative at 1) and whose growth is at most of polynomial order. We define the corresponding φ-relative entropy of two probability measures μ and ν given on the same probability space, by

$$\mathrm{Ent}_\varphi(\mu|\nu) \quad := \quad \begin{cases} \int \varphi\left(\frac{d\mu}{d\nu}\right)\, d\nu & ,\ \text{if } \mu \ll \nu, \\ +\infty & \text{otherwise.} \end{cases}$$

(the previous relative entropy corresponds to the function $\varphi : \mathbb{R}_+ \ni r \mapsto r\ln(r) - r + 1$). Contrary to the usual relative entropy case, when π is invariant with respect to M, the quantity $\mathrm{Ent}_\varphi(\mathcal{L}(Y^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))$ is no longer linear with respect to the time $T\in\mathbb{R}_+$, but we can define a discrepancy d_φ in the following way:

Proposition 1.5. *Without any assumption on M, we have*

$$\lim_{T\to 0_+} \mathrm{Ent}_\varphi(\mathcal{L}(Y^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))/T \quad = \quad d_\varphi(M,L),$$

where the discrepancy d_φ is given by

$$\forall\; M,M'\in\mathcal{L},$$

$$d_\varphi(M',M) \quad := \quad \begin{cases} \sum_{x\in S}\pi(x)\sum_{y\in S\setminus\{x\}} M(x,y)\varphi\left(\frac{M'(x,y)}{M(x,y)}\right) & ,\ \text{if } M'\ll M, \\ +\infty & \text{otherwise.} \end{cases}$$

where $M'\ll M$ means that

$$\forall\; x,y\in S,\qquad M(x,y)=0 \quad\Longrightarrow\quad M'(x,y)=0.$$

If furthermore we assume that φ is strictly convex then the mapping

$$\mathcal{L}(\pi)\ni M\mapsto d_\varphi(M,L)$$

admits a unique minimizer $L_{\varphi,\pi}$.

To recover the usual Metropolis generator L_π defined in (1.1), we consider for $\epsilon \in (0, 1/2]$, the convex function φ_ϵ satisfying $\varphi_\epsilon(x) = (x-1)^2$ for any $x \in [1-\epsilon, 1+\epsilon]$, $\varphi'_\epsilon = -1$ on $[0, 1-\epsilon)$ and $\varphi'_\epsilon = 1+\epsilon$ on $(1+\epsilon, +\infty)$ (this is a kind of Huber loss function). Despite the fact that φ_ϵ is not strictly convex, the mapping $\mathcal{L}(\pi) \ni M \mapsto d_{\varphi_\epsilon}(M, L)$ admits nevertheless a unique minimizer $L_{\varphi_\epsilon,\pi}$ and we have

$$\lim_{\epsilon \to 0_+} L_{\varphi_\epsilon,\pi} \quad = \quad L_\pi. \tag{1.3}$$

The advantage of this approach is that it can be extended to more general situations than the finite state space case. For instance in the Euclidean diffusion setting, we will see that the discrepancies d_{φ_ϵ} do not depend on $\epsilon \in (0, 1/2]$ and are also equal to $2\widetilde{d}$ (this quantity will be symmetrical in this situation, but not a distance, because it can still take the value $+\infty$). This provides a natural definition for d in this framework and "projecting" accordingly, we will recover the notion of Metropolis algorithm mentioned previously in the Brownian motion case.

The paper has the following plan. The above assertions about relative entropy and their generalizations to φ-relative entropies will be developed respectively in Sections 2 and 3. In Section 4 we will extend these considerations to the compact Riemannian diffusion situation. Finally in two appendices, we will respectively present the analogue of Girsanov formula in the finite state space setting, which enables to compute Radon-Nikodym densities on trajectory spaces, and the usual Girsanov formula in the compact Riemannian diffusion framework.

2. Entropy minimization

This section contains all the considerations concerning relative entropy mentioned in the introduction. In particular we prove Proposition 1.3 and Theorem 1.4.

In order to deal with Proposition 1.3 and (1.2) by the same computation, we first consider an arbitrary initial condition μ on S. From now on, we adopt the notations and conventions made in Appendix 1, in particular, $(X^{(\mu)}(t))_{t\geq 0}$ (respectively $(Y^{(\mu)}(t))_{t\geq 0}$) is a Markov process on S whose initial distribution is μ and whose generator is L (resp. M). For $t \geq 0$, we denote by μ_t the law of $Y^{(\mu)}(t)$. We also introduce the extended function F (F does not necessary belong to $\mathcal{F}(S)$, because it can take $+\infty$ as value) defined by

$$\forall\, x \in S, \qquad F(x) \quad := \quad \sum_{y \in S \setminus \{x\}} M(x,y) \ln\left(\frac{M(x,y)}{L(x,y)}\right) - M(x,y) + L(x,y)$$

(in particular, $F(x) = +\infty$ is equivalent to the fact that there exist $y \neq x$ such that $L(x,y) = 0$ and $M(x,y) > 0$).
These notations enable to compute the relative entropy of interest:

Lemma 2.1. *For any $T \geq 0$, we have*

$$\mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T]))|\mathcal{L}(X^{(\mu)}([0,T]))) \quad = \quad \int_0^T \mu_t[F]\, dt.$$

Proof: Assume first that $M \ll L$. Then according to Theorem 4.6, the Radon-Nikodym density of $\mathcal{L}(Y^{(\mu)}([0,T]))$ with respect to $\mathcal{L}(X^{(\mu)}([0,T]))$ is given by

$$\frac{d\mathcal{L}(Y^{(\mu)}([0,T]))}{d\mathcal{L}(X^{(\mu)}([0,T]))}(X([0,T])) = \exp\left(\sum_{(x,y)\in S^{(2)}} \ln(A(x,y))N_T^{(x,y)} + \int_0^T H(X(s))\,ds\right)$$

(recall that $(X(t))_{t\geq 0}$ is a generical trajectory of $\mathbb{D}$), with

$$\begin{array}{rrcl} \forall\ x\in S, & H(x) & := & M(x,x)-L(x,x),\\ \forall\ (x,y)\in S^{(2)}, & A(x,y) & := & \dfrac{M(x,y)}{L(x,y)}. \end{array}$$

Thus the entropy $\mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T]))|\mathcal{L}(X^{(\mu)}([0,T])))$ is equal to the expectation with respect to $\mathcal{L}(Y^{(\mu)}([0,T]))$ of $\sum_{(x,y)\in S^{(2)}}\ln(A(x,y))N_T^{(x,y)} + \int_0^T H(X(s))\,ds$ (with the convention that this expectation is $+\infty$ if the latter expression is not integrable). Taking into account that for any $(x,y)\in S^{(2)}$,

$$(N_t^{(x,y)} - \int_0^t M(x,y)\mathbb{1}_x(X(s))\,ds)_{t\in[0,T]}$$

is a martingale under $\mathcal{L}(Y^{(\mu)}([0,T]))$, which is $(\mathcal{L}(Y^{(\mu)}([0,T]))$-a.s. identically null if $A(x,y)=0$, we can write

$$\begin{aligned} &\mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T]))|\mathcal{L}(X^{(\mu)}([0,T])))\\ &= \mathbb{E}\left[\sum_{(x,y)\in S^{(2)}}\int_0^T M(x,y)\ln(A(x,y))\mathbb{1}_x(Y^{(\mu)}(t))\,dt + \int_0^T H(Y^{(\mu)}(t))\,dt\right]\\ &= \mathbb{E}\left[\int_0^T F(Y^{(\mu)}(t))\,dt\right]\\ &= \int_0^T \mathbb{E}\left[F(Y^{(\mu)}(t))\right]\,dt\\ &= \int_0^T \mu_t[F]\,dt, \end{aligned}$$

so we get the announced equality.

We now come to the general case. Let $\widetilde{S}_\mu$ be the set of points attainable by $Y^{(\mu)}$ (defined before Lemma 4.4, with $\widetilde{L}$ replaced by M). According to Remark 4.11, the above arguments are still valid if we have

$$\forall\ x\in\widetilde{S}_\mu,\ \forall\ y\in S,\qquad L(x,y)=0 \implies M(x,y)=0.$$

But if the latter condition is not satisfied, then on one hand for $T>0$, $\mathcal{L}(Y^{(\mu)}([0,T]))$ is not absolutely continuous with respect to $\mathcal{L}(X^{(\mu)}([0,T]))$ and thus $\mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T]))|\mathcal{L}(X^{(\mu)}([0,T])))=+\infty$. On the other hand, there exists $x\in\widetilde{S}_\mu$ such that $F(x)=+\infty$. But from the probabilistic description of $Y^{(\mu)}$, it appears that for any $t>0$, μ_t gives positive weight to any point of $\widetilde{S}_\mu$ and in particular $\mu_t[F]=+\infty$. It follows that in this situation the equality of the above lemma is also true for $T > 0$, since both terms are equal to $+\infty$. For $T = 0$, the equality holds because both terms are clearly null. $\square$

Proposition 1.3 follows immediately: if μ is invariant for M we have $\mu_t = \mu$ for all $t \geq 0$, so for any $T \geq 0$,

$$\mathrm{Ent}(\mathcal{L}(Y^{(\mu)}([0,T]))|\mathcal{L}(X^{(\mu)}([0,T]))) \quad = \quad T\mu[F]$$

and one would have noticed that $\pi[F] = \widetilde{d}(M, L)$.
Concerning (1.2), the assumption that π is attractive for M means that for any initial condition μ, we have

$$\lim_{t \to +\infty} \mu_t \quad = \quad \pi.$$

Thus Cesaro's lemma enables to conclude to (1.2), at least if F is a true function. But if F takes the value $+\infty$ at some point of S, the hypothesis that π is positive on S implies that $\widetilde{d}(M, L) = +\infty$. On the other hand the attractive assumption of π with respect to M is well-known to be equivalent to the fact that M is irreducible. This implies that for any initial distribution μ, $\widetilde{S}_\mu = S$ and by consequence, for any $T > 0$, the law $\mathcal{L}(Y^{(\mu)}([0,T]))$ cannot be absolutely continuous with respect to $\mathcal{L}(X^{(\mu)}([0,T]))$ (see Remark 4.11). It follows that the l.h.s. of (1.2) is also equal to $+\infty$.

One advantage of the discrepancy $\widetilde{d}$ over the distance d introduced by Billera and Diaconis (2001) is that we don't need an extra condition to define the projection on $\mathcal{L}(\pi)$ with respect to $\widetilde{d}$, as it will be clear in the

Proof of Theorem 1.4: Put an arbitrary total ordering on S, whose strict inequality is denoted $\prec$. The fact that $M \in \mathcal{L}(\pi)$ is equivalent to the fact that

$$\forall\, x \prec y, \qquad M(y, x) \quad = \quad \frac{\pi(x)}{\pi(y)} M(x, y)$$

and in particular M is determined by the free choice of the quantities $M(x, y) \geq 0$ for $x \prec y$. This leads us to write

$$\begin{aligned}
&\widetilde{d}(M, L)\\
&= \sum_{x \prec y} \pi(x)M(x, y)\ln\left(\frac{M(x, y)}{L(x, y)}\right) - \pi(x)M(x, y) + \pi(x)L(x, y)\\
&\qquad +\pi(y)M(y, x)\ln\left(\frac{M(y, x)}{L(y, x)}\right) - \pi(y)M(y, x) + \pi(y)L(y, x)\\
&= \sum_{x \prec y} \pi(x)M(x, y)\ln\left(\frac{\pi(x)M(x, y)}{\pi(x)L(x, y)}\right) - 2\pi(x)M(x, y) + \pi(x)L(x, y)\\
&\qquad +\pi(x)M(x, y)\ln\left(\frac{\pi(x)M(x, y)}{\pi(y)L(y, x)}\right) + \pi(y)L(y, x)\\
&= \sum_{x \prec y} \pi(x)M(x, y)\ln\left(\frac{(\pi(x)M(x, y))^2}{\pi(x)L(x, y)\pi(y)L(y, x)}\right)\\
&\qquad -2\pi(x)M(x, y) + \pi(x)L(x, y) + \pi(y)L(y, x).
\end{aligned}$$

It appears that if we want to find a minimizer $M \in \mathcal{L}(\pi)$ of $\widetilde{d}(M, L)$, we can minimize each summand of the above sum separately.

So let $x \prec y$ be fixed and denote

$$\begin{aligned}\alpha &:= \pi(x)M(x,y),\\ \beta &:= \pi(x)L(x,y),\\ \beta' &:= \pi(y)L(y,x).\end{aligned}$$

We are led to minimize as a function of $\alpha \geq 0$ the expression

$$2(\alpha\ln(\alpha)-\alpha)-\alpha\ln(\beta\beta')+\beta+\beta'.$$

By differentiating this strictly convex function, we see that the unique minimizer is $\alpha_* := \sqrt{\beta\beta'}$.
The announced results follow at once. □

Due to the fact that

$$\forall\ a,b\in\mathbb{R}_+,\qquad \min(a,b)\quad\leq\quad\sqrt{ab},$$

we get that

$$\forall\ x\neq y,\qquad L_\pi(x,y)\quad\leq\quad\widetilde{L}_\pi(x,y)$$

and it follows that the Markov process associated to $\widetilde{L}_\pi$ goes faster to equilibrium than that associated to L_π, if we measure this property by the spectral gap or by the logarithmic Sobolev constant (see for instance the lectures given by Saloff-Coste, 1997). Nevertheless the speed of convergence should not be the unique criterion for choosing a Metropolis algorithm (otherwise one would just multiply $\widetilde{L}_\pi$ by a big constant) as the numbers of operations by unit time should also be taken into account. Furthermore, the initial generator L should also play a role, maybe in the spirit of the discussion below Proposition 1.3.

Remark 2.2. When $L\in\mathcal{L}$ admits an invariant probability measure μ which is positive, the construction of $\widetilde{L}_\pi$ can be seen as a two steps procedure:
- First we symmetrize L into $\widetilde{L}$ defined by

$$\forall\ x\neq y\in S,\qquad \widetilde{L}(x,y)\quad:=\quad\sqrt{L(x,y)L^*(x,y)},$$

where L^* corresponds to the dual generator of L in $\mathbb{L}^2(\mu)$:

$$\forall\ x\neq y\in S,\qquad L^*(x,y)\quad:=\quad\frac{\mu(y)}{\mu(x)}L(y,x)$$

- Next we "reweight" $\widetilde{L}$ into $\widetilde{L}_\pi$ using $f=d\pi/d\mu$ the Radon-Nikodym derivative of π with respect to μ:

$$\forall\ x\neq y\in S,\qquad \widetilde{L}_\pi(x,y)\quad:=\quad\sqrt{\frac{f(y)}{f(x)}}\widetilde{L}(x,y).$$

3. Other Metropolis type projections

In order to better understand the difference between L_π and $\widetilde{L}_\pi$, we generalize here the projection procedure of the previous section.

We work in the finite state space setting as above: for any probability measure μ on S, we consider $(X^{(\mu)}(t))_{t\geq 0}$ a Markov process on S whose initial distribution is μ and whose generator is L and we will denote by $\mathbb{P}_\mu$ the probability measure on the underlying probability space Ω (only the image of $\mathbb{P}_\mu$ by $(X^{(\mu)}(t))_{t\geq 0}$ will be important, so we could restrict ourselves to the canonical situation where $\Omega = \mathbb{D}$ and $\mathbb{P}_\mu$ is the solution of the martingale problem associated to μ and L). For any $(x,y) \in S^{(2)}$ and $t \geq 0$, $N_t^{(x,y)} := \sum_{s\in(0,t]} \mathbb{1}_{\{X^{(\mu)}(s-)=x,\, X^{(\mu)}(s)=y\}}$ is the number of jumps from x to y in the time interval $(0,t]$ and

$$\forall\ t \geq 0, \qquad N_t \ := \sum_{(x,y)\in S^{(2)}} N_t^{(x,y)}$$

is the total number of jumps before time t. From the probabilistic description of $(X^{(\mu)}(t))_{t\geq 0}$, it appears that the process $(N_t)_{t\geq 0}$ is stochastically dominated by $(\bar{N}_t)_{t\geq 0}$, a Poisson process of parameter $l := \max_{x\in S} |L(x,x)|$. The following preliminary result is well-known.

Lemma 3.1. *For any $t \geq 0$ and $(x,y) \in S^{(2)}$, consider the events*

$$\begin{aligned}
\Omega_{t,x,x} &:= \{N_t = 0\},\\
\Omega_{t,x,y} &:= \{N_t^{(x,y)} = N_t = 1\},\\
\widetilde{\Omega}_{t,x} &:= \Omega \setminus \sqcup_{y\in S}\Omega_{t,x,y}.
\end{aligned}$$

Then we have

$$\begin{aligned}
\lim_{t\to 0_+} \frac{1-\mathbb{P}_x[\Omega_{t,x,x}]}{t} &= -L(x,x),\\
\lim_{t\to 0_+} \frac{\mathbb{P}_x[\Omega_{t,x,y}]}{t} &= L(x,y),\\
\lim_{t\to 0_+} \frac{\mathbb{P}_x[\widetilde{\Omega}_{t,x}]}{t^2} &< +\infty.
\end{aligned}$$

Proof: Let $(x,y) \in S^{(2)}$ be given, using the martingale problem for the indicator function of y, we get that

$$\lim_{t\to 0_+} \frac{\mathbb{P}_x[X^{(x)}(t) = y]}{t} = L(x,y).$$

But for any $t \geq 0$, we have

$$\{X^{(x)}(t) = y\} \setminus \Omega_{t,x,y} \quad \subset \quad \{N_t \geq 2\}.$$

Taking into account the stochastic domination of $(N_t)_{t\geq 0}$ by $(\bar{N}_t)_{t\geq 0}$, it appears that $\mathbb{P}_x[\{X^{(x)}(t) = y\} \setminus \Omega_{t,x,y}]$ is at most of order t^2 for small $t > 0$. The second convergence announced in the lemma follows. The first one is obtained in the same way, since we also have

$$\lim_{t\to 0_+} \frac{1-\mathbb{P}_x[X^{(x)}(t) = x]}{t} = -L(x,x)$$

The last assertion of the lemma is due to the fact that $\widetilde{\Omega}_{t,x}$ is included into $\{N_t \geq 2\}$. $\square$

For the next result, let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be a locally bounded function vanishing at 1 and whose growth is at most of polynomial order at $+\infty$. Let also be given a family $(a(x,y))_{(x,y)\in S^{(2)}}$ of elements from $\mathbb{R} \sqcup \{-\infty\}$ (in fact only the quantities $a(x,y)$ with $L(x,y) > 0$ will play a role).

Lemma 3.2. *We have for any initial distribution μ,*

$$\lim_{t\to 0_+} t^{-1}\mathbb{E}_\mu\left[\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right] = \sum_{(x,y)\in S^{(2)}} \mu(x)L(x,y)\varphi(\exp(a(x,y))).$$

Proof: We begin by remarking that for any $b\geq 0$, we can find a constant $c(b)\in\mathbb{R}_+$ such that

$$\forall\ t\geq 0,\qquad \mathbb{E}_\mu\left[\exp\left(b\sum_{y\in S\setminus\{x\}} a_+(x,y)N_t^{(x,y)}\right)\right] \quad\leq\quad \exp(tc(b))$$

(where for any $r\in\mathbb{R}\sqcup\{-\infty\}$, $r_+ := \max(r,0)$). Indeed, define the function $H_b\in\mathcal{F}(S)$ by

$$\forall\ x\in S,\qquad H_b(x)\quad :=\quad -\sum_{y\in S\setminus\{x\}} L(x,y)(\exp(ba_+(x,y))-1),$$

then according to Theorem 4.6, for any $t\geq 0$, the functional

$$\exp\left(\sum_{(x,y)\in S^{(2)}} ba_+(x,y)N_t^{(x,y)} + \int_0^t H_b(X^{(\mu)}(s))\,ds\right)$$

is a density under $\mathbb{P}_\mu$, so it is sufficient to take

$$c(b)\quad :=\quad -\min_{x\in S} H_b(x).$$

In particular the expectations considered in the lemma are well defined and we can write for any $t>0$ and $x'\in S$,

$$\begin{aligned}
&\mathbb{E}_{x'}\left[\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right]\\
&\quad=\quad \mathbb{E}_{x'}\left[\mathbb{1}_{\Omega_{t,x',x'}}\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right]\\
&\qquad+\sum_{y'\in S\setminus\{x'\}}\mathbb{E}_{x'}\left[\mathbb{1}_{\Omega_{t,x',y'}}\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right]\\
&\qquad+\mathbb{E}_{x'}\left[\mathbb{1}_{\widetilde{\Omega}_{t,x'}}\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right]
\end{aligned}$$

The first term of the r.h.s. is null, since $\varphi(1)=0$. The second term is equal to

$$\sum_{y'\in S\setminus\{x'\}}\varphi(\exp(a(x',y')))\mathbb{P}_{x'}\left[\Omega_{t,x',y'}\right]\quad\sim\quad t\sum_{y'\in S\setminus\{x'\}}\varphi(\exp(a(x',y')))L(x',y')$$

as t goes to 0_+.

Concerning the third term, we bound it using Hölder inequality with conjugate exponents $1 < p, q < +\infty$,

$$\begin{aligned}
&\mathbb{E}_{x'}\left[\mathbb{1}_{\widetilde{\Omega}_{t,x'}}\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right]\\
&\quad\leq \left(\mathbb{P}_{x'}[\widetilde{\Omega}_{t,x'}]\right)^{\frac{1}{p}}\left(\mathbb{E}_{x'}\left[|\varphi|^q\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right]\right)^{\frac{1}{q}}\\
&\quad\leq \left(\mathbb{P}_{x'}[\widetilde{\Omega}_{t,x'}]\right)^{\frac{1}{p}}\left(\mathbb{E}_{x'}\left[K^q\exp\left(qr\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right]\right)^{\frac{1}{q}}\\
&\quad\leq K\left(\mathbb{P}_{x'}[\widetilde{\Omega}_{t,x'}]\right)^{\frac{1}{p}}\left(\mathbb{E}_{x'}\left[\exp\left(qr\sum_{(x,y)\in S^{(2)}} a_+(x,y)N_t^{(x,y)}\right)\right]\right)^{\frac{1}{q}}\\
&\quad\leq \left(\mathbb{P}_{x'}[\widetilde{\Omega}_{t,x'}]\right)^{\frac{1}{p}} K\exp(tc(rq)/q),
\end{aligned}$$

where $K, r \geq 0$ are such that we have

$$\forall\ s\in\mathbb{R}_+,\qquad |\varphi(s)| \quad\leq\quad Ks^r$$

So if we choose $p \in (1,2)$, Lemma 3.1 shows that

$$\lim_{t\to 0_+} t^{-1}\mathbb{E}_{x'}\left[\mathbb{1}_{\widetilde{\Omega}_{t,x'}}\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}\right)\right)\right] \quad=\quad 0$$

and the announced result follows. □

Now let us assume furthermore that φ is convex and differentiable at 1 with $\varphi'(1) = 0$. Let also be given $h \in \mathcal{F}(S)$. Then the addition of the additive functional associated to h in the exponential does not change the previous result:

Lemma 3.3. *We have for any initial distribution* μ*,*

$$\begin{aligned}
\lim_{t\to 0_+} t^{-1}\mathbb{E}_\mu&\left[\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)} + \int_0^t h(X^{(\mu)}(s))\,ds\right)\right)\right]\\
&= \sum_{(x,y)\in S^{(2)}} \mu(x)L(x,y)\varphi(\exp(a(x,y))).
\end{aligned}$$

Proof: To simplify the notations, let us write for any $t \geq 0$,

$$N_t^{(a)} := \sum_{(x,y)\in S^{(2)}} a(x,y)N_t^{(x,y)}.$$

In view of Lemma 3.2, it is sufficient to prove that

$$\begin{aligned}
\lim_{t\to 0_+} t^{-1}\Bigg|\mathbb{E}_\mu&\left[\varphi\left(\exp\left(N_t^{(a)} + \int_0^t h(X^{(\mu)}(s))\,ds\right)\right)\right]\\
&-\mathbb{E}_\mu\left[\varphi\left(\exp\left(N_t^{(a)}\right)\right)\right]\Bigg| = 0. \qquad (3.1)
\end{aligned}$$

We denote by φ'_+ the right derivative of φ (recall that it exists since φ is convex) and we begin by assuming furthermore that $\varphi'_+(0) > -\infty$. Writing that for any $s \in \mathbb{R}_+$ we have

$$\varphi'_+(0) \;\leq\; \varphi'_+(s) \;\leq\; \frac{\varphi(2s)-\varphi(s)}{s} \;\leq\; \frac{\varphi(2s)}{s}$$

(φ is nonnegative since $\varphi'(1) = 0$), it appears that φ'_+ is locally bounded and that its growth is at most of polynomial order at $+\infty$. Thus it exist $K', r' \geq 0$ such that

$$\forall\, s \in \mathbb{R}_+, \qquad \left|\varphi'_+(s)\right| \;\leq\; K' s^{r'}.$$

This leads to the following bound, with the function $\psi : \mathbb{R}_+ \ni s \mapsto s\varphi'_+(s)$,

$$\begin{aligned}
&\left|\varphi\left(\exp\left(N_t^{(a)} + \int_0^t h(X^{(\mu)}(s))\,ds\right)\right) - \varphi\left(\exp\left(N_t^{(a)}\right)\right)\right| \\
&\quad = \quad \left|\int_0^t h(X^{(\mu)}(s))\,ds \int_0^1 \psi\left(\exp\left(N_t^{(a)} + u\int_0^t h(X^{(\mu)}(s))\,ds\right)\right)\,du\right| \\
&\quad \leq \quad K' \left\|h\right\|_\infty t \exp\left((r'+1)[N_t^{(a_+)} + \left\|h\right\|_\infty t]\right).
\end{aligned}$$

We deduce that the l.h.s. of (3.1) is bounded above by

$$K' \left\|h\right\|_\infty \exp((1+r')\left\|h\right\|_\infty t)\mathbb{E}_\mu\left[\exp\left((r'+1)N_t^{(a_+)}\right)\right].$$

According to Lemma 3.2 applied with a_+, the latter expectation is of order t for small $t > 0$, in particular the above expression vanishes as t goes to zero, so (3.1) is proven. It remains to deal with the case where $\varphi'_+(0) = -\infty$. Let $\eta = \min(1/2, \min_{(x,y)\in S^{(2)}} \exp(a(x,y)))$ and consider $\widetilde{\varphi}, \widehat{\varphi} : \mathbb{R}_+ \to \mathbb{R}_+$ two convex functions which coincide with φ on $[\eta, +\infty)$, which are affine on $[0,\eta]$ and which satisfy $\widetilde{\varphi} \leq \varphi \leq \widehat{\varphi}$ on $\mathbb{R}_+$. By the above proof, we have with $\phi = \widetilde{\varphi}$ or $\phi = \widehat{\varphi}$,

$$\begin{aligned}
&\lim_{t\to 0_+} t^{-1}\mathbb{E}_\mu\left[\phi\left(\exp\left(N_t^{(a)} + \int_0^t h(X^{(\mu)}(s))\,ds\right)\right)\right] \\
&\qquad\qquad = \sum_{(x,y)\in S^{(2)}} \mu(x)L(x,y)\phi(\exp(a(x,y))) \\
&\qquad\qquad = \sum_{(x,y)\in S^{(2)}} \mu(x)L(x,y)\varphi(\exp(a(x,y))).
\end{aligned}$$

The announced result then follows by comparison. $\square$

We can now come to the

Proof: of Proposition 1.5 So let π be a positive probability measure on S and besides $(X^{(\pi)}(t))_{t\geq 0}$, we are given $(Y^{(\pi)}(t))_{t\geq 0}$ a Markov process on S whose initial distribution is π and whose generator is M.
We first assume that $M \ll L$. Then according to Theorem 4.6, we have for any $T \geq 0$,

$$\begin{aligned}
&\mathrm{Ent}_\varphi(\mathcal{L}(Y^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) \\
&\quad = \quad \mathbb{E}_\pi\left[\varphi\left(\exp\left(\sum_{(x,y)\in S^{(2)}} a(x,y)N_T^{(x,y)} + \int_0^T H(X^{(\mu)}(s))\,ds\right)\right)\right],
\end{aligned}$$

with

$$\begin{array}{rrcl}\forall\ x\in S, & H(x) & := & M(x,x)-L(x,x),\\ \forall\ (x,y)\in S^{(2)}, & a(x,y) & := & \ln\left(\dfrac{M(x,y)}{L(x,y)}\right)\ \in\ \mathbb{R}\sqcup\{-\infty\}.\end{array}$$

The convergence announced in Proposition 1.5 is then an immediate consequence of Lemma 3.3. If the property $M \ll L$ is not satisfied, we have, because π is positive and according to Lemma 4.5, that for any $T>0$, $\mathcal{L}(Y^{(\pi)}([0,T]))$ is not absolutely continuous with respect to $\mathcal{L}(X^{(\pi)}([0,T]))$. So due to our conventions, the first part of Proposition 1.5 is also valid.

The proof of the second part is similar to that of Theorem 1.4, except we cannot give an explicit formula for the minimizer $L_{\varphi,\pi}$.
Again we consider an arbitrary total ordering on S, whose strict inequality is denoted $\prec$. If π is reversible for M and $M \ll L$, we can write

$$\begin{aligned} d_\varphi(M,L) &= \sum_{x\prec y}\pi(x)L(x,y)\varphi\left(\frac{M(x,y)}{L(x,y)}\right)+\pi(y)L(y,x)\varphi\left(\frac{M(y,x)}{L(y,x)}\right)\\ &= \sum_{x\prec y}\Phi_{\pi(x)L(x,y),\pi(y)L(y,x)}(\pi(x)M(x,y)),\end{aligned}$$

where Φ is defined by

$$\forall\ \alpha,\beta,\beta'\geq 0,\qquad \Phi_{\beta,\beta'}(\alpha)\ :=\ \beta\varphi\left(\frac{\alpha}{\beta}\right)+\beta'\varphi\left(\frac{\alpha}{\beta'}\right)\tag{3.2}$$

(the convention that zero multiplied by anything is still zero is always enforced). In the above sum, if $x\prec y$ are such that $L(x,y)=0$ or $L(y,x)=0$, then we must have $M(x,y)=0$, since $M\ll L$. So to find a minimizer $M\in\mathcal{L}(\pi)$ of $d_\varphi(M,L)$ is equivalent to find a minimizer $M\in\mathcal{L}(\pi)$ with $M\ll L$ of the expression

$$\sum_{x\lhd y}\Phi_{\pi(x)L(x,y),\pi(y)L(y,x)}(\pi(x)M(x,y)),$$

where $x\lhd y$ means that $x\prec y$ with $L(x,y)>0$ and $L(y,x)>0$. Since each summand of the above sum can be minimized separately, we are led to show that for any given $(\beta,\beta')\in(\mathbb{R}_+\setminus\{0\})^2$, the mapping $\mathbb{R}_+\ni\alpha\mapsto\Phi_{\beta,\beta'}(\alpha)$ admits a unique minimizer $\alpha_{\beta,\beta'}\in\mathbb{R}_+$ (note that by symmetry we will have $\alpha_{\beta,\beta'}=\alpha_{\beta',\beta}$). The unique minimizer $L_{\varphi,\pi}$ of $\mathcal{L}(\pi)\ni M\mapsto d_\varphi(M,L)$ will then be given by

$$\forall\ (x,y)\in S^{(2)},\ \widetilde{L}_{\varphi,\pi}(x,y):=\begin{cases}\frac{1}{\pi(x)}\alpha_{\pi(x)L(x,y),\pi(y)L(y,x)} & \text{, if } x\lhd y \text{ or if } y\lhd x,\\ 0 & \text{, otherwise.}\end{cases}$$

So let us come back to the problem of minimizing $\phi:=\Phi_{\beta,\beta'}$, where $(\beta,\beta')\in(\mathbb{R}_+\setminus\{0\})^2$ are assumed to be fixed. The function ϕ is convex and let us denote by ϕ'_+ its right derivative. To see that Φ admits α_* as unique minimizer is equivalent to show that

$$\forall\ \alpha\in\mathbb{R}_+,\qquad \phi'_+(\alpha)\qquad \begin{cases}<0 & \text{, if } \alpha<\alpha_*\\ >0 & \text{, if } \alpha>\alpha_*\end{cases}\tag{3.3}$$

We compute that

$$\forall\ \alpha\in\mathbb{R}_+,\qquad \phi'_+(\alpha)\ :=\ \varphi'_+\left(\frac{\alpha}{\beta}\right)+\varphi'_+\left(\frac{\alpha}{\beta'}\right).$$

By the fact that $\varphi'(1) = 0$ and the strict convexity of φ, it appears that for $\alpha > 0$ sufficiently small $\phi'_+(\alpha) < 0$, that for $\alpha > 0$ sufficiently large $\phi'_+(\alpha) > 0$ and that ϕ'_+ is increasing, thus there exists a unique $\alpha_* > 0$ such that (3.3) is satisfied. □

In fact the strict convexity condition of Proposition 1.5 to get the uniqueness of the minimizer can be relaxed. To go in this direction, let us introduce

$$\begin{aligned} D_+ &:= \{t \in \mathbb{R}_+ : \lambda((\varphi'_+)^{-1}(\{t\})) > 0\} \\ D_- &:= \{t \in \mathbb{R}_+ : \lambda((\varphi'_+)^{-1}(\{-t\})) > 0\} \end{aligned}$$

(where λ denotes the Lebesgue measure), which correspond to the nonnegative slopes of φ where φ is locally affine and respectively to the opposites of the non-positive slopes of φ where φ is locally affine. The strict convexity of φ is equivalent to the fact that $D_- \cup D_+ = \emptyset$. The last part of Proposition 1.5 can be improved as follows:

Proposition 3.4. *Assume that $D_- \cap D_+ = \emptyset$, then the mapping $\mathcal{L}(\pi) \ni M \mapsto d_\varphi(M, L)$ admits a unique minimizer $L_{\varphi,\pi}$.*

Proof: As in the proof of Proposition 1.5, we consider the mapping $\phi := \Phi_{\beta,\beta'}$, where $(\beta, \beta') \in (\mathbb{R}_+ \setminus \{0\})^2$ are fixed and where $\Phi_{\beta,\beta'}$ is defined in (3.2). To get the wanted uniqueness, it is sufficient to check that (3.3) is satisfied for some $\alpha_* > 0$. Assume that this is not true, then there would exist $0 < \alpha_- < \alpha_+$ such that $\phi'_+ = 0$ on (α_-, α_+), which means that

$$\forall\ \alpha \in (\alpha_-, \alpha_+), \qquad \varphi'_+\left(\frac{\alpha}{\beta}\right) \;=\; -\varphi'_+\left(\frac{\alpha}{\beta'}\right).$$

But the l.h.s. is nondecreasing as a function of α while the r.h.s. is nonincreasing, so we deduce that both $\varphi'_+\left(\frac{\cdot}{\beta}\right)$ and $\varphi'_+\left(\frac{\cdot}{\beta'}\right)$ should be constant on the interval (α_-, α_+). Denoting l the absolute value of these constants (one being the opposite of the other), we would get that $l \in D_- \cap D_+$, which is in contradiction with our assumption. □

In fact the condition given in the above proposition is optimal:

Remark 3.5. Assume that the convex function φ (still satisfying $\varphi(1) = \varphi'(1) = 0$) is such that for any $L \in \mathcal{L}$, the mapping $\mathcal{L}(\pi) \ni M \mapsto d_\varphi(M, L)$ admits a unique minimizer, then we have $D_- \cap D_+ = \emptyset$. Indeed, let $l \in D_- \cap D_+$, then we can find $0 < \beta < \beta'$ and $0 < \alpha_- < \alpha_+$ such that

$$\forall\ \alpha \in (\alpha_-, \alpha_+), \qquad \varphi'_+\left(\frac{\alpha}{\beta}\right) \;=\; -\varphi'_+\left(\frac{\alpha}{\beta'}\right) \;=\; l.$$

It follows that $\Phi_{\beta,\beta'}$ admits the whole interval $[\alpha_-, \alpha_+]$ as minimizers. So according to the proof of Proposition 1.5, if $L \in \mathcal{L}$ is such that for some $x \neq y$ we have $\pi(x)L(x, y) = \beta$ and $\pi(y)L(y, x) = \beta'$, then we can freely choose the value of $\pi(x)M(x, y) = \pi(y)M(y, x)$ inside $[\alpha_-, \alpha_+]$ for a minimizer M of $d_\varphi(\cdot, L)$ on $\mathcal{L}(\pi)$.

We can now justify the first assertion made after Proposition 1.5. So let $\epsilon \in (0, 1/2]$ be given and consider the function φ_ϵ, for which $D_- = \{1\}$ and $D_+ = \{1 + \epsilon\}$. Thus the condition of Proposition 3.4 is fulfilled and the mapping $\mathcal{L}(\pi) \ni M \mapsto d_{\varphi_\epsilon}(M, L)$ admits a unique minimizer $L_{\varphi_\epsilon,\pi}$. Before proving the convergence

(1.3), let us give a more precise version of Proposition 1.2. From now on, endow S with the binary relation $\prec$ defined by

$$\forall\ x, y \in S, \qquad x \prec y \quad \Leftrightarrow \quad \pi(x)L(x,y) \leq \pi(y)L(y,x) \text{ and } x \neq y.$$

Then we have

Proposition 3.6. *A generator $M \in \mathcal{L}(\pi)$ is a minimizer of $d(\cdot, L)$ on $\mathcal{L}(\pi)$ if and only if for any $x \prec y$, we have*

$$\pi(x)M(x,y) = \pi(y)M(y,x) \quad \in \quad [\pi(x)L(x,y), \pi(y)L(y,x)].$$

Proof: Consider the convex function $\psi\ :\ \mathbb{R}_+ \ni u \mapsto |u-1|$. If we replace the distance d by the discrepancy d_ψ, which coincides with d on couples $(M', M) \in \mathcal{L}^2$ with $M' \ll M$ and is equal to $+\infty$ otherwise, the proof of the second part of Proposition 1.5 shows that a generator $M \in \mathcal{L}(\pi)$ is a minimizer of $d_\psi(\cdot, L)$ on $\mathcal{L}(\pi)$ if and only if for any $x \prec y$, with $L(x,y)L(y,x) > 0$, we have that $\pi(x)M(x,y)$ is a minimizer of $\Psi_{\pi(x)L(x,y),\pi(y)L(y,x)}$, where

$$\forall\ \alpha, \beta, \beta' \geq 0, \qquad \Phi_{\beta,\beta'}(\alpha) \quad := \quad |\alpha - \beta| + |\alpha - \beta'|\,.$$

This function is easy to minimize and we find the condition

$$\pi(x)M(x,y) \quad \in \quad [\pi(x)L(x,y), \pi(y)L(y,x)].$$

Looking closely at these arguments, it appears that the situation with d instead of d_ψ is similar, except we don't have to take into account the absolutely continuity constraint and we end up with the condition given in the proposition. $\square$

In particular, it appears that if there exists $(x,y) \in S^{(2)}$ with $L(x,y) = 0$ and $L(y,x) > 0$, then some of the minimizers M of $d(\cdot, L)$ on $\mathcal{L}(\pi)$ do not satisfy $M \ll L$.

We can now come to the

Proof of the convergence (1.3): If $(x,y) \in S^{(2)}$ is such that $L(x,y)L(y,x) = 0$, then we have for any $\epsilon \in (0, 1/2]$, $L_{\varphi_\epsilon,\pi}(x,y) = 0 = L_\pi(x,y)$ and the convergence (1.3) is trivial. So consider $(x,y) \in S^{(2)}$ with $0 < \beta < \beta'$ where $\beta := \pi(x)L(x,y)$ and $\beta' := \pi(y)L(y,x)$. For $\epsilon \in (0, 1/2]$, let $\alpha_*(\epsilon)$ be defined as in (3.3) but with φ replaced by φ_ϵ. According to the proofs of Propositions 1.5 and 3.6, we have $L_{\varphi_\epsilon,\pi}(x,y) = \alpha_*(\epsilon)/\pi(x)$ and the convergence $\lim_{\epsilon\to 0_+} L_{\varphi_\epsilon,\pi}(x,y) = L_\pi(x,y)$ is equivalent to the fact that

$$\lim_{\epsilon \to 0_+} \alpha_*(\epsilon) \quad = \quad \beta.$$

But this convergence is a consequence of the observation that for $\epsilon > 0$ small enough,

$$\begin{aligned} \forall\ s \in [0, (1-\epsilon)\beta), \qquad (\phi_\epsilon)'(s) \quad &= \quad -2, \\ \forall\ s \in ((1+\epsilon)\beta, (1-\epsilon)\beta'), \qquad (\phi_\epsilon)'(s) \quad &= \quad \epsilon, \end{aligned}$$

which implies that $\alpha_*(\epsilon) \in [(1-\epsilon)\beta, (1+\epsilon)\beta]$. $\square$

But the above result should not lead one to think that if $(\varphi_\epsilon)_{\epsilon\in(0,\epsilon_0]}$, for some $\epsilon_0 > 0$, is another family of functions satisfying the assumption of Proposition 3.4 and converging to $\psi\ :\ \mathbb{R}_+ \ni u \mapsto |u-1|$ for small $\epsilon > 0$, then (1.3) is fulfilled. Indeed, consider for $\epsilon \in (0, 1/2]$, the convex function φ_ϵ defined by $\varphi_\epsilon(x) = (1-x)^2$

for any $x \in [1-\epsilon, 1+\epsilon]$, $\varphi'_\epsilon = -1-\epsilon$ on $[0, 1-\epsilon)$ and $\varphi'_\epsilon = 1$ on $(1+\epsilon, +\infty)$. Then computations similar to the previous ones show that

$$\lim_{\epsilon \to 0_+} L_{\varphi_\epsilon, \pi} = \widehat{L}_\pi,$$

with

$$\forall\ (x,y) \in S^{(2)},$$
$$\widehat{L}_\pi(x,y) := \begin{cases} \max\left\{L(x,y), \frac{\pi(y)}{\pi(x)}L(y,x)\right\} & \text{, if } L(x,y)L(y,x) > 0, \\ 0 & \text{, otherwise.} \end{cases}$$

Conversely, one should neither think that all the minimizers M described in Proposition 3.6 could be approached by minimizers of discrepancies as in Proposition 3.4. Indeed, this is clear for minimizers M not satisfying $M \ll L$, but even some minimizers satisfying this condition cannot be approached. Consider for instance the case where there exist $(x,y) \in S^{(2)}$ and $(x',y') \in S^{(2)}$ with $\{x,y\} \neq \{x',y'\}$ and $\pi(x)L(x,y) = \pi(x')L(x',y') < \pi(y)L(y,x) = \pi(y')L(y',x')$. Then there are some minimizers M such that $\pi(x)M(x,y) \neq \pi(x')M(x',y')$, while for any convex function φ satisfying the assumption of Proposition 3.4, we have $\pi(x)L_{\varphi,\pi}(x,y) = \pi(x')L_{\varphi,\pi}(x',y')$.

Remark 3.7. The considerations of Proposition 3.6 can be extended to other discrepancies on $\mathcal{L}$. For instance for $\epsilon > 0$, consider the (non symmetrical) discrepancy

$$\forall\ M, L \in \mathcal{L},$$
$$d_\epsilon(M,L) := \sum_{x \in S} \pi(x) \sum_{y \in S \setminus \{x\}} (|M(x,y) - L(x,y)| + \epsilon(M(x,y) - L(x,y))_+$$

(where $(\cdot)_+$ stands for the positive part). Then it appears that the mapping $\mathcal{L}(\pi) \ni M \mapsto d_\epsilon(M,L)$ admits the usual Metropolis generator L_π as unique minimizer (if one replace the positive part by the negative part in the above discrepancy, one gets the previously defined generator $\widehat{L}_\pi$). But Lemma 3.3 is not clear for the corresponding convex function $\mathbb{R}_+ \ni u \mapsto |u-1| + \epsilon(u-1)_+$ and this alternative definition L_π does not extend either to the diffusion situation.

4. On the compact Riemannian diffusion situation

We now leave the finite state space setting to see how our previous considerations can be extended to a compact Riemannian diffusion framework. As our main purpose is to illustrate how the approach of the previous sections can be generalized, we will adopt strong regularity assumptions. That is also why we have chosen a compact setting, to avoid the investigation of boundedness conditions. For computations of trajectory entropies in Euclidean spaces, see for instance the papers of Dawson and Gärtner (1987) or of Cattiaux and Léonard (1994) and the references given therein.

So let S be a smooth compact manifold of dimension $n \in \mathbb{N} \setminus \{0\}$. We are interested into Markov generators L of diffusion type, which can be written in a chart C as

$$\forall\ f \in \mathcal{C}^\infty,\ \forall\ x \in S,$$
$$L[f](x) \quad := \quad \frac{1}{2} \sum_{i,j \in [\![1,n]\!]} a_{i,j}(x)\partial_{i,j} f(x) + \sum_{i \in [\![1,n]\!]} b_i(x)\partial_i f(x), \quad (4.1)$$

where $\mathcal{C}^\infty$ is the space of smooth functions on S, where ∂_i is the differentiation operator with respect to the i^{th} coordinate of the underlying chart C, $\partial_{i,j} = \partial_i \partial_j$ and where the diffusive symmetric matrix field $C \ni x \mapsto a(x) := (a_{i,j}(x))_{i,j \in [\![1,n]\!]}$ and the drift vector field $C \ni x \mapsto b(x) := (b_i(x))_{i \in [\![1,n]\!]}$ are smooth. Furthermore $a(x)$ is assumed to be invertible for all $x \in C$.

If we are given a probability measure μ on S, it is well-known (see for instance the books of Émery, 1989 and of Ikeda and Watanabe, 1989) that we can associate to μ, seen as the initial distribution, and to the generator L a Markov process $(X^{(\mu)}(t))_{t \geq 0}$ with continuous trajectories in S (a Markov process with continuous trajectories is traditionally called a diffusion). The law $\mathbb{P}_\mu$ of $(X^{(\mu)}(t))_{t \geq 0}$ on the space of continuous mappings from $\mathbb{R}_+$ to S is characterized as the unique solution of the martingale problem associated to μ and L:

- The law of $X^{(\mu)}(0)$ is μ.
- For any function $f \in \mathcal{C}^\infty$, the process $\mathcal{M}^{(f)}$ defined by

$$\forall\ t \in \mathbb{R}_+, \qquad \mathcal{M}_t^{(f)} \quad := \quad f(X^{(\mu)}(t)) - f(X^{(\mu)}(0)) - \int_0^t L[f](X^{(\mu)}(s))\, ds$$

is a martingale.

For more informations on martingale problems associated to diffusion operators, we refer to the books of Stroock and Varadhan (2006) and Ethier and Kurtz (1986).

The generator L and the Markov process $(X^{(\mu)}(t))_{t \geq 0}$ should be seen as an a priori given Markov framework and they will play exactly the same role they had in the previous sections. The other data will be a probability measure π with a smooth and positive density. But first let us consider another diffusion generator $\widetilde{L}$ (whose diffusion symmetric matrix field and drift vector field will be denoted $(\widetilde{a}_{i,j}(x))_{i,j \in [\![1,n]\!]}$ and $(\widetilde{b}_i(x))_{i \in [\![1,n]\!]}$ for x belonging to a chart C) satisfying the same conditions as L. A corresponding diffusion process is denoted $(\widetilde{X}^{(\mu)}(t))_{t \geq 0}$ (instead of $(Y^{(\mu)}(t))_{t \geq 0}$ in the previous sections). In proposition 4.13 of Appendix 2 we recall why a necessary condition for absolute continuity on finite time horizons is that $\widetilde{a} = a$. Since in the end we will mainly be interested into diffusions $(\widetilde{X}_t^{(\mu)})_{t \geq 0}$ which are absolutely continuous with respect to $(X_t^{(\mu)})_{t \geq 0}$ on finite time horizons, the diffusive matrix field of $\widetilde{L}$ will thus have to coincide with that of L. This leads us to consider the corresponding Riemannian structure.

For $x \in S$, we introduce a scalar product $\langle \cdot, \cdot \rangle(x)$ on the tangent space $\mathrm{T}_x S$ of S at x: if $v, w \in \mathrm{T}_x S$ and if C is a chart with $x \in C$, we define

$$\langle v, w \rangle(x) \quad := \quad \sum_{i,j \in [\![1,n]\!]} (a^{-1}(x))_{i,j} v_i w_j,$$

where $(v_i)_{i \in [\![1,n]\!]}$ and $(w_i)_{i \in [\![1,n]\!]}$ are the coordinates of v and w in the chart C. It can be checked that $(\langle \cdot, \cdot \rangle(x))_{x \in S}$ satisfies the rules of 2-covariant tensor and since it is symmetrical and positive definite, it gives rise to a Riemannian structure on S, see for instance the book of Ikeda and Watanabe (1989). From now on, we denote by $\langle\cdot,\cdot\rangle$, $|\cdot|$, ∇, div, $\triangle$ and λ, respectively the scalar product, the norm, the gradient, the divergence, the Laplacian and the probability associated to this Riemannian structure. The operator L can then be rewritten under a more intrinsic way:

$$L\cdot \quad = \quad \triangle/2\cdot + \langle b, \nabla\cdot\rangle, \tag{4.2}$$

where b is a vector field on S (but whose coordinates $(b_i)_{i\in[\![1,n]\!]}$ in a chart don't necessary coincide with those given in the formulation (4.1)). Assuming that $\widetilde{a} = a$, the other generator $\widetilde{L}$ can be written under the same form, $\widetilde{L}\cdot = \triangle/2\cdot + \langle\widetilde{b}, \nabla\cdot\rangle$, with another vector field $\widetilde{b}$.

The usual Girsanov theorem recalled in Proposition 4.14 of Appendix 2, as well as Remark 4.15 following it, lead us to introduce the notation $\widetilde{L} \sim L$, where L and $\widetilde{L}$ are two generators as in the beginning of this section, if $\widetilde{a} = a$ (by analogy with the previous section, we could also have adopted the notation $\widetilde{L} \ll L$, but we prefered to write $\widetilde{L} \sim L$ to indicate that in the present continuous framework, this is an equivalence relation). Let π be a probability measure on S. We define an associated discrepancy d on the set $\mathcal{L}$ of generators as above by

$$\forall\ L, \widetilde{L} \in \mathcal{L}, \qquad d(\widetilde{L}, L) \quad := \quad \begin{cases} \int \left|\widetilde{b} - b\right|^2 (x)\,\pi(dx) & \text{, if } \widetilde{L} \sim L, \\ +\infty & \text{, otherwise.} \end{cases}$$

Note that the discrepancy d is symmetrical: for any $L, \widetilde{L} \in \mathcal{L}$, $d(\widetilde{L}, L) = d(L, \widetilde{L})$.

To justify the introduction of d (and to try to smooth the abuse of notation consisting in denoting it in the same way as the distance Billera and Diaconis (2001) in the discrete time and space setting), let us consider $\varphi : \mathbb{R}_+ \to \mathbb{R}$ a convex function differentiable in a neighborhood of 1, admitting a second derivative $\varphi''(1) > 0$, satisfying $\varphi(1) = \varphi'(1) = 0$ and whose growth is at most of polynomial order. Then we have a result analogous to Proposition 1.5:

Proposition 4.1. *For any $\widetilde{L}, L \in \mathcal{L}$, we have*

$$\lim_{T\to 0_+} \mathrm{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))/T \quad = \quad \frac{\varphi''(1)}{2} d(\widetilde{L}, L).$$

As it will appear in the following proof, this convergence is also true if $\varphi''(1) = 0$, but we have to replace $\varphi''(1)d(\widetilde{L}, L)$ by 0 if $\widetilde{L} \sim L$ and $+\infty$ otherwise.

Proof: If the condition $\widetilde{L} \ll L$ is not satisfied, then according to Remark 4.15 we have

$$\forall\ T > 0, \qquad \mathrm{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) \quad = \quad +\infty.$$

Thus it is sufficient to treat the situation where $\widetilde{L} \sim L$. According to Proposition 4.14, we have for any $T \geq 0$,

$$\mathrm{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) \quad = \quad \mathbb{E}\left[\varphi(\mathcal{E}_T[\widetilde{b} - b])\right].$$

As in the finite state space, this expectation is finite because for any $T \geq 0$ and $p \geq 1$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{E}_T^p[\widetilde{b}-b]\right] &= \mathbb{E}\left[\exp\left(p\mathcal{M}_T^{(\widetilde{b}-b)} - \frac{p}{2}\int_0^T \left|\widetilde{b}-b\right|^2 (X^{(\mu)}(t))\,dt\right)\right] \qquad (4.3)\\
&= \mathbb{E}\left[\mathcal{E}_T[p(\widetilde{b}-b)]\exp\left(\frac{p^2-p}{2}\int_0^T \left|\widetilde{b}-b\right|^2 (X^{(\mu)}(t))\,dt\right)\right]\\
&\leq \exp\left(\frac{p^2-p}{2}T\left\|\widetilde{b}-b\right\|_\infty^2\right)\mathbb{E}\left[\mathcal{E}_T[p(\widetilde{b}-b)]\right]\\
&= \exp\left(\frac{p^2-p}{2}T\left\|\widetilde{b}-b\right\|_\infty^2\right).
\end{aligned}$$

We begin by assuming furthermore that φ is smooth on $\mathbb{R}_+$ and that the growth of φ'' is at most of polynomial order. In this case we can use directly It's formula to get that for any $T \geq 0$,

$$\begin{aligned}
\mathbb{E}\left[\varphi(\mathcal{E}_T[\widetilde{b}-b])\right] &= \mathbb{E}\left[\int_0^T \varphi'(\mathcal{E}_t[\widetilde{b}-b])\mathcal{E}_t[\widetilde{b}-b]\,d\mathcal{M}_t^{(\widetilde{b}-b)}\right.\\
&\qquad\left. + \frac{1}{2}\int_0^T \varphi''(\mathcal{E}_t[\widetilde{b}-b])\mathcal{E}_t^2[\widetilde{b}-b]\,d\langle\mathcal{M}^{(\widetilde{b}-b)}\rangle_t\right]\\
&= \frac{1}{2}\mathbb{E}\left[\int_0^T \varphi''(\mathcal{E}_t[\widetilde{b}-b])\mathcal{E}_t^2[\widetilde{b}-b]\left|\widetilde{b}-b\right|^2(X^{(\pi)}(t))\,dt\right]\\
&= \frac{1}{2}\int_0^T \mathbb{E}\left[\varphi''(\mathcal{E}_t[\widetilde{b}-b])\mathcal{E}_t^2[\widetilde{b}-b]\left|\widetilde{b}-b\right|^2(X^{(\pi)}(t))\right]dt.
\end{aligned}$$

So if we divide by T and let T go to zero, we get in the limit

$$\frac{1}{2}\mathbb{E}\left[\varphi''(\mathcal{E}_0[\widetilde{b}-b])\mathcal{E}_0^2[\widetilde{b}-b]\left|\widetilde{b}-b\right|^2(X^{(\pi)}(0))\right] = \frac{\varphi''(1)}{2}d(\widetilde{L},L),$$

where to justify the convergence, we have taken into account the fact that the growth of φ'' is at most of polynomial order and the bound (4.3). Note that the convexity of φ was not necessery to deduce this convergence.

We now come to the general case. By our assumptions on φ, for any $\epsilon > 0$, we can find two smooth functions $\varphi_{\epsilon,-}$ and $\varphi_{\epsilon,+}$ on $\mathbb{R}_+$ satisfying the following conditions:

$$\varphi_{\epsilon,-}(1) = \varphi'_{\epsilon,-}(1) = 0 = \varphi_{\epsilon,+}(1) = \varphi'_{\epsilon,+}(1),$$

$$\varphi''_{\epsilon,-}(1) = \varphi''(1)-\epsilon \quad \text{and} \quad \varphi''_{\epsilon,+}(1) = \varphi''(1)+\epsilon,$$

$$\forall\, s \in \mathbb{R}_+, \qquad \varphi_{-,\epsilon}(s) \leq \varphi(s) \leq \varphi_{+,\epsilon}(s),$$

$\varphi''_{-,\epsilon}$ and $\varphi''_{+,\epsilon}$ grow at most polynomially.

Then using the previous convergence with the functions $\varphi_{\epsilon,-}$ and $\varphi_{\epsilon,+}$ and the following bound valid for any $T \geq 0$,

$$\begin{aligned}\mathbb{E}\left[\varphi_{-,\epsilon}(\mathcal{E}_T[\widetilde{b}-b])\right] &\leq \operatorname{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))\\ &\leq \mathbb{E}\left[\varphi_{+,\epsilon}(\mathcal{E}_T[\widetilde{b}-b])\right],\end{aligned}$$

it appears that

$$\begin{aligned}\limsup_{T\to 0_+}\frac{\operatorname{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))}{T} &\leq \frac{d(\widetilde{L},L)}{2}(\varphi''(1)+\epsilon),\\ \liminf_{T\to 0_+}\frac{\operatorname{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))}{T} &\geq \frac{d(\widetilde{L},L)}{2}(\varphi''(1)-\epsilon).\end{aligned}$$

It remains to let ϵ go to zero to be convinced of the announced convergence. □

In the case of the usual entropy, corresponding to the function $\mathbb{R}_+ \ni u \mapsto u\ln(u)-u+1$, we recover the analogous result to Proposition 1.3: if π is invariant for $\widetilde{L}$, then for any $T\geq 0$, we have

$$\operatorname{Ent}(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) = \frac{T}{2}d(\widetilde{L},L) \leq +\infty.$$

Indeed, the only nontrivial case corresponds to $T>0$ and $\widetilde{L}\ll L$, for which we compute that

$$\begin{aligned}&\forall\, T>0,\\ &\operatorname{Ent}(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) = \widetilde{\mathbb{E}}_\pi\left[\ln\left(\mathcal{E}_T[\widetilde{b}-b]\right)\right]\\ &\qquad = \widetilde{\mathbb{E}}_\pi\left[\mathcal{M}_T^{(\widetilde{b}-b)} - \frac{1}{2}\int_0^T \left|\widetilde{b}-b\right|^2(\widetilde{X}^{(\pi)}(t))\,dt\right],\end{aligned}$$

where $\widetilde{\mathbb{E}}_\pi[\cdot]$ stands for the expectation with respect to $\widetilde{X}^{(\pi)}$. To go further, note that $\mathcal{M}^{(\widetilde{b}-b)}$ is no longer a martingale under this law and to transform it into a martingale we have to subtract the process $\langle\mathcal{M}^{(\widetilde{b}-b)},\mathcal{M}^{(\widetilde{b}-b)}\rangle$ (this comes from arguments similar to those used in the proof of Proposition 4.14, see also the book of Revuz and Yor, 1999). Thus it follows that

$$\begin{aligned}\operatorname{Ent}(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T]))) &= \widetilde{\mathbb{E}}_\pi\left[\frac{1}{2}\int_0^T\left|\widetilde{b}-b\right|^2(\widetilde{X}^{(\pi)}(t))\,dt\right]\\ &= \frac{1}{2}\int_0^T\widetilde{\mathbb{E}}_\pi\left[\left|\widetilde{b}-b\right|^2(\widetilde{X}^{(\pi)}(t))\right]dt\\ &= \frac{T}{2}\widetilde{\mathbb{E}}_\pi\left[\left|\widetilde{b}-b\right|^2(\widetilde{X}^{(\pi)}(0))\right],\end{aligned}$$

because π is invariant for $\widetilde{L}$ and we recognize the last term as $Td(\widetilde{L},L)/2$.

Thus as it was announced, the discrepancies $2\widetilde{d}$ and d_{φ_ϵ} for $\epsilon\in(0,1/2]$ all coincides with d in our diffusion framework. It is then quite natural to propose as the Metropolis generator associated to π and L, the minimizer L_d of the mapping

$\mathcal{L}(\pi) \ni \widetilde{L} \mapsto d(\widetilde{L}, L)$, at least if it exists. Here $\mathcal{L}(\pi)$ also stands for the set of generators $\widetilde{L} \in \mathcal{L}$ for which π is reversible. This means that

$$\forall\ f, g \in \mathcal{C}^\infty, \qquad \pi[f\widetilde{L}[g]] \quad = \quad \pi[g\widetilde{L}[f]].$$

Under our smoothness assumptions, for a given $\widetilde{L} \in \mathcal{L}$, if such a reversible probability measure π exists, it admits a positive and smooth density with respect to λ, because π must also be invariant for $\widetilde{L}$ and so it is a weak solution to the elliptic equation $\widetilde{L}^*[\pi] = 0$, where $\widetilde{L}^*$ is dual operator of $\widetilde{L}$ in $\mathbb{L}^2(\lambda)$. Conversely, if π is a probability measure admitting a positive and smooth density with respect to λ, then $\mathcal{L}(\pi)$ is not empty, since it contains at least the operator $(\triangle\cdot + \langle \nabla \ln(\pi), \nabla\cdot\rangle)/2$. In fact let us check that this is the only element $\widetilde{L}$ of $\mathcal{L}(\pi)$ satisfying $\widetilde{L} \sim L$, where L is given in (4.2). The condition $\widetilde{L} \sim L$ implies that there exists a vector field $\widetilde{b}$ such that we can write $\widetilde{L}\cdot = \frac{1}{2}\triangle\cdot + \left\langle \widetilde{b}, \nabla\cdot \right\rangle$. The dual operator $\widetilde{L}^*$ of $\widetilde{L}$ in $\mathbb{L}^2(\lambda)$ is then given by $\widetilde{L}^*\cdot = \frac{1}{2}\triangle \cdot - \mathrm{div}(\cdot\widetilde{b})$. The reversibility of π with respect to $\widetilde{L}$ is equivalent to the fact that $\widetilde{L}$ is equal to its dual operator in $\mathbb{L}^2(\pi)$, condition which can be written

$$\widetilde{L}\cdot \quad = \quad \frac{1}{\pi}\widetilde{L}^*[\pi\cdot] \tag{4.4}$$

(where we denote in the same way the probability measure π and its density with respect to λ). By expanding this relation and by taking into account the equation $\widetilde{L}^*[\pi] = 0$, we get that the vector field $\frac{\nabla\pi}{\pi} - 2\widetilde{b}$ must vanish everywhere, which amounts to the above assertion.

So the minimization of $\mathcal{L}(\pi) \ni \widetilde{L} \mapsto d(\widetilde{L}, L)$ is very simple, since $L_d := (\triangle\cdot + \langle \nabla \ln(\pi), \nabla\cdot\rangle)/2$ is the unique element of $\mathcal{L}(\pi)$ with $d(L_d, L) < +\infty$.

The above approach can be extended to the situation where we add jumps to the above diffusion generators. But as already mentioned in the discussion after Proposition 4.14, we don't want to develop here the underlying theory of martingales associated to functions of two variables. So let us close this paper by sketching the line of reasoning, which is a mix of the arguments given in the finite and diffusion case, without entering into the details of the corresponding computations:

Remark 4.2. Consider generators of the type $H = L + Q$, where the diffusion part L is given as in (4.1), with the same assumptions on the coefficients a and b, and where the jump part can be written as

$$\forall\ f \in \mathcal{C}^\infty, \forall\ x \in S, \qquad Q[f](x) \quad := \quad \int (f(y) - f(x))\, q(x,y)\lambda(dy),$$

where the intensity of jumps q is a positive and smooth function on S^2. It is easy to see that the decomposition of H under the form $L+Q$ is unique and more precisely that the action of H on $\mathcal{C}^\infty$ determine the coefficients a, b (in any chart) and q. The set of such generators H will be denoted $\mathcal{H}$.
If μ is a probability measure on S, we denote $(X^{(\mu)})_{t\geq 0}$ any càdlàg Markov process admitting μ as initial distribution and H as generator. Its law on the set of càdlàg trajectories is the unique solution of the martingale problem associated to μ and H.
Let $\widetilde{H} \in \mathcal{H}$ be another generator of the same kind, all notions relative to $\widetilde{H}$ will receive a tilde at their top. As recalled at the beginning of this section, for any $T > 0$, we have $\mathcal{L}(\widetilde{X}^{(\mu)}([0,T]) \ll \mathcal{L}(X^{(\mu)}([0,T])$ if and only if $\widetilde{a} = a$, property

which is still denoted by $\widetilde{H} \sim H$. Again we endow S with the Riemannian structure induced by a^{-1}, in order to adopt intrinsic notations: $H = \triangle/2 \cdot + \langle b, \nabla \cdot \rangle + Q[\cdot]$, where b is a vector field on S (and idem for $\widetilde{H}$ if $\widetilde{H} \sim H$). Let us introduce the functions V and v defined by

$$\begin{aligned}
\forall\ (x,y) \in S^2, \qquad V(x,y) \quad &:= \quad \begin{cases} \ln\left(\frac{\widetilde{q}(x,y)}{q(x,y)}\right) & , \text{if } x \neq y, \\ 0 & , \text{if } x = y. \end{cases} \\
\forall\ x \in S, \qquad v(x) \quad &:= \quad \int (V(x,y) - \exp(V(x,y)) + 1) q(x,y)\, \lambda(dy)
\end{aligned}$$

Then it is possible to associate naturally to V a martingale $\mathcal{M}^{(V)}$, such that if $\widetilde{H} \sim H$, we have for any $T \geq 0$,

$$\frac{d\mathcal{L}(\widetilde{X}^{(\mu)}([0,T]))}{d\mathcal{L}(X^{(\mu)}([0,T]))} \quad = \quad \mathcal{E}_T[\widetilde{b} - b]\mathcal{E}_T[V],$$

where the first factor was defined in (4.11) and the second factor is given by

$$\mathcal{E}_T[V] \quad := \quad \exp\left(\mathcal{M}_T^{(V)} + \int_0^T v(X^{(\mu)}(t))\, dt\right).$$

The fact that these densities so neatly decompose into a diffusion factor and a jump factor is related to the fact that the martingale $\mathcal{M}^{(V)}$ is totally discontinuous (namely its co-bracket with any continuous martingale is null, see for instance the book of Dellacherie and Meyer, 1980). So in this situation, we see that in the completion of the space of simple functions of two variables with respect to the norm (4.10), two kinds of orthogonal objects appear: germs of functions on the diagonal (corresponding to vector fields) and functions of two variables which vanish on the diagonal (corresponding to jump rate densities). The parameterization of Girsanov transformations by these objects (the couple $(\widetilde{b} - b, V)$ in the above example) is in fact very general, as it was alluded to in Remark 4.12. These formulas enable to handle φ-entropies of trajectorial laws and to define corresponding discrepancies. Indeed, under the assumptions on φ preceding Proposition 4.1, we get for any probability measure π on S,

$$\lim_{T \to 0_+} \mathrm{Ent}_\varphi(\mathcal{L}(\widetilde{X}^{(\pi)}([0,T]))|\mathcal{L}(X^{(\pi)}([0,T])))/T \quad = \quad d_\varphi(\widetilde{L}, L),$$

where the discrepancy in the r.h.s. is given by

$$d_\varphi(\widetilde{L}, L) :=$$
$$:= \begin{cases} \frac{\varphi''(1)}{2} \int \left|\widetilde{b} - b\right|^2 (x)\, \pi(dx) + \int \varphi\left(\frac{\widetilde{q}(x,y)}{q(x,y)}\right) \pi(dx) q(x,y) \lambda(dy) & , \text{if } \widetilde{H} \sim H, \\ +\infty & , \text{otherwise.} \end{cases}$$

Assume now that π has a smooth and positive density with respect to λ (which will still be denoted π) and consider $\mathcal{H}(\pi)$ the subset of $\mathcal{H}$ consisting of generators from $\mathcal{H}$ admitting π as a reversible measure. As it can be guessed, our purpose is to find the minimizer H_φ of the mapping $\mathcal{H}(\pi) \ni \widetilde{H} \mapsto d_\varphi(\widetilde{H}, H)$, under the additional condition on φ given in Proposition 3.4. In order to go in this direction the following result is important.

Proposition 4.3. *The generator $\widetilde{H}$ is reversible with respect to π if and only if its diffusion part $\widetilde{L}$ and its jump part $\widetilde{Q}$ are both reversible with respect to π.*

Proof: As in the case (4.4) of the diffusion generators, the reversibility of π is equivalent to the equation $\widetilde{H}\cdot = \frac{1}{\pi}\widetilde{H}^*[\pi\cdot]$, where $\widetilde{H}^*$ is the dual operator of $\widetilde{H}$ in $\mathbb{L}^2(\lambda)$. Expanding the r.h.s. we get that for any $f \in \mathcal{C}^\infty$ and $x \in S$,

$$\begin{aligned} \frac{1}{\pi(x)}\widetilde{H}^*[\pi f](x) &= \frac{1}{2}\triangle f(x) + \left\langle \nabla \ln(\pi) - \widetilde{b}, \nabla f \right\rangle(x) \\ &+ \int (f(y) - f(x)) \frac{\pi(y)}{\pi(x)}\widetilde{q}(y,x)\lambda(dy) + \frac{\widetilde{H}^*[\pi]}{\pi(x)} f(x). \end{aligned}$$

So the previous equality is equivalent to the fact that

$$\begin{aligned} \widetilde{H}^*[\pi] &= 0, \\ \nabla \ln(\pi) - \widetilde{b} &= \widetilde{b}, \\ \forall\, x \neq y \in S, \qquad \frac{\pi(y)}{\pi(x)}\widetilde{q}(y,x) &= q(x,y). \end{aligned}$$

The second equation is equivalent to the reversibility of π with respect to the diffusion part $\widetilde{L}$ and the third equation is equivalent to the reversibility of π with respect to the jump part $\widetilde{Q}$. The first equation, which is the formulation of the stationarity of π, is in fact implied by the two other equations. □

Indeed, the last proposition is also true without assumption of regularity on the probability measure π, but it appears a posteriori that π has to admit a smooth and positive density with respect to λ, since it must also be reversible for the elliptic operator $\widetilde{L}$.
Thus H_φ is a minimizer of the mapping $\mathcal{H}(\pi) \ni \widetilde{H} \mapsto d_\varphi(\widetilde{H}, H)$ if and only if its diffusion part L_φ is a minimizer of the mapping $\mathcal{L}(\pi) \ni \widetilde{L} \mapsto \frac{\varphi''(1)}{2} d(\widetilde{L}, L)$ and its jump part Q_φ is a minimizer of the mapping $\widetilde{q} \mapsto \int \varphi\left(\frac{\widetilde{q}(x,y)}{q(x,y)}\right) \pi(dx)q(x,y)\lambda(dy)$. For the diffusion part, this leads to $L_\varphi = (\triangle \cdot + \langle \nabla \ln(\pi), \nabla \cdot\rangle)/2$. For the jump part, the computations of the proof of Proposition 1.5 (taking into account Proposition 3.4) lead to

$$\forall\, x \neq y, \qquad q_\varphi(x,y) = \frac{1}{\pi(x)}\alpha_{\pi(x)q(x,y),\pi(y)q(y,x)},$$

where the function $(\mathbb{R}_+^*)^2 \ni (\beta, \beta') \mapsto \alpha_{\beta,\beta'}$ was defined in term of φ in the proof of Proposition 1.5.
In particular in the case of the classical entropy, we find the rate density $\widetilde{q}$ given by

$$\forall\, x \neq y, \qquad \widetilde{q}(x,y) = \sqrt{\frac{\pi(y)}{\pi(x)}}\sqrt{q(x,y)q(y,x)}.$$

Appendix 1: finite state space Girsanov formula

This long appendix presents in the finite state space setting the changes of law which are analogous to those considered by Girsanov (1960) for Euclidean diffusions. They corresponds to all the transformations of the underlying probability measure $\mathbb{P}$ of a given Markov process which remain absolutely continuous with respect to $\mathbb{P}$ on finite time intervals and which preserve the (time-homogeneous) Markov property.

We begin by recalling the martingale problem approach to Markov processes, since it is the natural framework to deal with this subject. We need some notations: let $\mathbb{D}$ be the set of càdlàg trajectories from $\mathbb{R}_+$ to the finite set S. We denote by $(X(t))_{t\geq 0}$ the process of the canonical coordinates on $\mathbb{D}$ and we endow $\mathbb{D}$ with the σ-field $\mathcal{D}$ generated by the $X(t)$, for $t \in \mathbb{R}_+$. For any $T \in \mathbb{R}_+$, we also denote $\mathcal{D}([0,T])$ the σ-field generated by the $X(t)$, for $t \in [0,T]$. We will always implicitly assume that the set $\mathbb{D}$ is endowed with the filtration $(\mathcal{D}([0,t]))_{t\geq 0}$. Let a probability measure μ on S and $L \in \mathcal{L}$ be given. We say that a probability measure $\mathbb{P}$ on $\mathbb{D}$ is a solution to the martingale problem associated to μ and L if:

- The law of $X(0)$ (called the initial law) under $\mathbb{P}$ is μ.
- For any function f defined on S, the process $\mathcal{M}^{(f)}$ defined by

$$\forall\, t \in \mathbb{R}_+, \qquad \mathcal{M}_t^{(f)} \quad := \quad f(X(t)) - f(X(0)) - \int_0^t L[f](X(s))\, ds \tag{4.5}$$

is a martingale. In the above expression, L was interpreted as an operator acting on functions (seen as column vector, indexed by S) through the formula

$$\forall\, x \in S, \qquad L[f](x) \quad := \quad \sum_{y\in S} L(x,y)f(y).$$

It is well-known, see for instance the book of Ethier and Kurtz (1986), that such a solution exists and is unique, we denote it $\mathbb{P}_\mu$ (or simply $\mathbb{P}_x$ when μ is equal to the Dirac mass δ_x for some $x \in S$). The probabilistic description of the evolution of $(X(t))_{t\geq 0}$ is the following: for $s > 0$, let $X(s-) := \lim_{u\to s, u<s} X(u)$ and consider the first jump time $\tau_1 := \inf\{s > 0 \,:\, X(s) \neq X(s-)\}$. Then conditionally to $X(0)$, τ_1 is distributed as an exponential law of parameter $|L(X(0),X(0))|$ (this is the Dirac measure at $+\infty$ if $L(X(0),X(0)) = 0$). Conditionally to $X(0)$ and $\tau_1 < +\infty$, the first jump position $X(\tau_1)$ is distributed according to the law $(L(X(0),y)/|L(X(0),X(0))|)_{y\neq X(0)}$ on $S \setminus \{X(0)\}$. Next conditionally to $X(0)$, τ_1 and $X(\tau_1)$, the waiting time for the second jump is distributed as an exponential law of parameter $|L(X(\tau_1),X(\tau_1))|$ etc.
If one doesn't want to make reference to a particular initial distribution μ (and this is usually the point of view adopted in Markov process theory, compare for instance Lemmas 4.4 and 4.5 below to see how convenient it can be), one considers simultaneously all the martingale problems associated to δ_x and L, for $x \in S$, and one says that the solution to the martingale problems relative to L is the family $(\mathbb{P}_x)_{x\in S}$. Note that all the $\mathbb{P}_\mu$ can be recovered from this family, since we have $\mathbb{P}_\mu = \sum_{x\in S} \mu(x)\mathbb{P}_x$. Furthermore, the Markov property of $(X(t))_{t\geq 0}$ under $\mathbb{P}_\mu$ is deduced without difficulty from the uniqueness of the family $(\mathbb{P}_x)_{x\in S}$ as solution to the above martingale problems.

Let $\widetilde{L} \in \mathcal{L}$ be another generator and $\widetilde{\mathbb{P}}_\mu$ be the solution of the martingale problem associated to μ and $\widetilde{L}$. More generally all notions relative to $\widetilde{L}$ will be covered by a tilde sign when they have to be distinguished from those relative to L. For $T \geq 0$, consider $\mathbb{P}_{\mu,[0,T]}$ the restriction of $\mathbb{P}_\mu$ to the measurable space $(\mathbb{D}, \mathcal{D}([0,T]))$. We are interested in the generators $\widetilde{L}$ which are such that for any $x \in S$ and any $T \geq 0$, $\widetilde{\mathbb{P}}_{x,[0,T]}$ is absolutely continuous with respect to $\mathbb{P}_{x,[0,T]}$ (using the Markov property, it is clear that is equivalent to the existence of $T > 0$ such that for any $x \in S$, $\widetilde{\mathbb{P}}_{x,[0,T]}$ is absolutely continuous with respect to $\mathbb{P}_{x,[0,T]}$). A necessary condition is easy to obtain. For $x \in S$, let S_x be the set of points which are attainable by X

under $\mathbb{P}_x$: a point $y \in S$ belongs to S_x if and only if there exists a finite sequence $x_0 = x, x_1, ..., x_n = y$ in S with $n \in \mathbb{N}$ and $L(x_i, x_{i+1}) > 0$ for all $i \in [\![0, n-1]\!]$. More generally, let $S_\mu := \cup_{x \in \mathrm{supp}(\mu)} S_x$, where $\mathrm{supp}(\mu)$ is the support of μ.

Lemma 4.4. *If there exists $T > 0$ such that $\widetilde{\mathbb{P}}_{\mu,[0,T]}$ is absolutely continuous with respect to $\mathbb{P}_{\mu,[0,T]}$, then $\widetilde{S}_\mu \subseteq S_\mu$ and we have*

$$\forall\, x \in \widetilde{S}_\mu, \forall\, y \in S, \qquad L(x,y) = 0 \quad \Longrightarrow \quad \widetilde{L}(x,y) = 0$$

Proof: For $x, y \in S$ and $T > 0$, consider the event

$$A_{x,y,T} \quad := \quad \{\exists\, s \in (0,T] \,:\, X(s-) = x,\, X(s) = y\}.$$

From the probabilistic description of $\mathbb{P}_\mu$, we see that $\mathbb{P}_\mu[A_{x,y,T}] > 0$ if and only if $x \in S_\mu$ and $L(x,y) > 0$ and let denote by R_μ the set of such couples (x,y).
Now let $T > 0$ be as in the above lemma. From the absolute continuity assumption, we get that $\widetilde{R}_\mu \subseteq R_\mu$ and this amounts to the conclusions given in the lemma. □

The necessary condition given in Lemma 4.4 is in fact sufficient, see remark 4.11 below. But to avoid embarrassing notations, it is better to work simultaneously with the whole family $(\mathbb{P}_x)_{x \in S}$. As an immediate consequence of the above result, we get:

Lemma 4.5. *If there exists $T > 0$ such that for any $x \in S$, $\widetilde{\mathbb{P}}_{x,[0,T]}$ is absolutely continuous with respect to $\mathbb{P}_{x,[0,T]}$, then we have $\widetilde{L} \ll L$ (with the notation of Proposition 1.5), which just means that the transitions forbidden by L are also forbidden by $\widetilde{L}$.*

Our main goal in this appendix is to show the reciprocal result and to exhibit the density $d\widetilde{\mathbb{P}}_{x,[0,T]}/d\mathbb{P}_{x,[0,T]}$. To do so, again we need more notations.

For $x \neq y \in S$ and $t \geq 0$, consider

$$N_t^{(x,y)} \quad := \quad \sum_{s \in (0,t]} \mathbb{1}_{X(s-)=x, X(s)=y}$$

the number of jumps from x to y which have occurred before time t and

$$\mathcal{M}_t^{(x,y)} \quad := \quad N_t^{(x,y)} - \int_0^t L(x,y) \mathbb{1}_x(X(s))\, ds.$$

Let us check that $(\mathcal{M}_t^{(x,y)})_{t \geq 0}$ is a martingale under $\mathbb{P}_\mu$ for any initial condition μ. This comes from the fact that we can represent it as a stochastic integral:

$$\forall\, t \geq 0, \qquad \mathcal{M}_t^{(x,y)} \quad = \quad \int_0^t \mathbb{1}_x(X(s-))\, d\mathcal{M}_s^{(y)} \tag{4.6}$$

where $(\mathcal{M}_t^{(y)})_{t\geq 0}$ is the martingale associated to the indicator function $\mathbb{1}_y$ via formula (4.5). Indeed, let us compute the r.h.s.:

$$\begin{aligned}
\int_0^t \mathbb{1}_x(X(s-))\,d\mathcal{M}_s^{(y)} &= \int_0^t \mathbb{1}_x(X(s-))\,d\bigg(\mathbb{1}_y(X(s)) - \mathbb{1}_y(X(0)) \\
&\qquad\qquad - \int_0^s L[\mathbb{1}_y](X(u))\,du\bigg) \\
&= \int_0^t \mathbb{1}_x(X(s-))\,d\mathbb{1}_y(X(s)) \\
&\quad - \int_0^t \mathbb{1}_x(X(s-))L[\mathbb{1}_y](X(s))\,ds \\
&= \sum_{0<s\leq t} \mathbb{1}_x(X(s-))(\mathbb{1}_y(X(s)) - \mathbb{1}_y(X(s-))) \\
&\quad - \int_0^t \mathbb{1}_x(X(s-))L[\mathbb{1}_y](x)\,ds \\
&= N_t^{(x,y)} - \int_0^t L(x,y)\mathbb{1}_x(X(s))\,ds.
\end{aligned}$$

Now consider another generator $\widetilde{L}$ satisfying the conclusion of Lemma 4.5. We define

$$\forall\, x \neq y, \qquad A(x,y) \;:=\; \frac{\widetilde{L}(x,y)}{L(x,y)}$$

(taking into account the usual convention that $0\cdot\infty = 0$, we have $A(x,y)=0$ if $L(x,y)=0$, since $\widetilde{L} \ll L$). Consider also the functions G and H defined by

$$\forall\, x\in S,\ \left\{\begin{array}{lcl} G(x) &:=& \sum_{y\neq x} L(x,y)(\ln(A(x,y)) - A(x,y) + 1), \\ H(x) &:=& -\sum_{y\neq x} L(x,y)(A(x,y)-1) \;=\; \widetilde{L}(x,x) - L(x,x). \end{array}\right.$$

Then we have

Theorem 4.6. *Under the assumption that $\widetilde{L} \ll L$, for any initial condition μ and any finite time horizon $T\geq 0$, we have $\widetilde{\mathbb{P}}_{\mu,[0,T]} \ll \mathbb{P}_{\mu,[0,T]}$ and the corresponding Radon-Nikodym derivative is given by*

$$\begin{aligned}
\frac{d\widetilde{\mathbb{P}}_{\mu,[0,T]}}{d\mathbb{P}_{\mu,[0,T]}} &= \exp\left(\sum_{x\neq y\in S} \ln(A(x,y))N_T^{(x,y)} + \int_0^T H(X(s))\,ds\right) \\
&= \exp\left(\sum_{x\neq y\in S} \ln(A(x,y))\mathcal{M}_T^{(x,y)} + \int_0^T G(X(s))\,ds\right)
\end{aligned}$$

(note that these quantities vanish if $N_T^{(x,y)} > 0$ for some $x\neq y$ satisfying $\widetilde{L}(x,y)=0$ and $L(x,y)>0$).

This result could be proven using Doléans-Dade stochastic exponential (cf. for instance the book of Dellacherie and Meyer, 1980) and stochastic calculus for martingales with jumps. Nevertheless this approach does not seem to us the most appropriate to deal with the present situation. Indeed, we just want to use immediate consequences of the martingale problem formulation and elementary stochastic calculus, enabling to compute the product of a absolutely continuous adapted process with a martingale. The proof of Theorem 4.6 will take several steps. The first one consist in introducing some exponential martingales (which are not easily described as Doléans-Dade stochastic exponential processes).

To any function f on S, we associate a new function $I[f]$ via the formula

$$\forall\ x \in S, \qquad I[f](x) \quad := \quad \exp(-f(x))L[\exp(f)](x) - L[f](x).$$

Remark 4.7. The mapping $f \mapsto I[f]$ should be seen as a modified "carré du champ". Recall that the usual carré du champ of f (see for instance the lecture of Bakry, 1994) is defined by

$$\forall\ x \in S, \qquad \Gamma[f](x) \quad := \quad L[f^2](x) - 2f(x)L[f](x).$$

It is easily checked for instance that we have

$$\exp(-\mathrm{osc}(f))\Gamma[f] \ \leq \ 2I[f] \ \leq \ \exp(\mathrm{osc}(f))\Gamma[f],$$

where $\mathrm{osc}(f) = \max_S f - \min_S f$ is the oscillation of f. Furthermore, if L is a diffusion generator (namely if the processes admitting L as generator have continuous paths, as it is the case in Section 4), we have $I[f] = \Gamma[f]/2$. The interest of the carré du champ is that it enables to compute the bracket of $\mathcal{M}^{(f)}$ (i.e. the previsible process $(\langle \mathcal{M}^{(f)} \rangle_t)_{t \geq 0}$ starting from 0 such that $((\mathcal{M}^{(f)})^2 - \langle \mathcal{M}^{(f)} \rangle_t)_{t \geq 0}$ is a martingale):

$$\forall\ t \geq 0, \qquad \langle \mathcal{M}^{(f)} \rangle_t \quad = \quad \int_0^t \Gamma[f](X(s))\, ds.$$

The modified carré du champ will play a similar role of integrand for a compensator, but with an exponential mapping rather than a square mapping.

We associate to any function f on S the process $(\mathcal{E}_t[f])_{t \geq 0}$ given by

$$\forall\ t \geq 0, \qquad \mathcal{E}_t[f] \quad := \quad \exp\left(\mathcal{M}_t^{(f)} - \int_0^t I[f](X(s))\, ds\right). \tag{4.7}$$

Then we have another characterization of the solution to the martingale problem associated to μ and L:

Proposition 4.8. *Under $\mathbb{P}_\mu$ the process $(\mathcal{E}_t[f])_{t \geq 0}$ is a martingale. Conversely if $\mathbb{P}$ is a probability measure on $\mathbb{D}$ such that μ is the law of $X(0)$ and such that for any function f, the process $(\mathcal{E}_t[f])_{t \geq 0}$ is a martingale, then $\mathbb{P} = \mathbb{P}_\mu$.*

Proof: For the first point, we write that for any $t \geq 0$,

$$\mathcal{E}_t[f] \quad = \quad \exp\left(-\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\, ds\right) \exp(f(X(t)) - f(X(0))),$$

so its stochastic differential is

$$\begin{aligned}
d\mathcal{E}_t[f] &= \exp\left(-\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)\exp(-f(X(0)))\\
&\quad (L[\exp(f)](X(t))dt + d\mathcal{M}_t^{(\exp(f))})\\
&\quad -\exp\left(-\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)\\
&\quad \exp(f(X(t))-f(X(0)))\exp(-f(X(t)))L[\exp(f)](X(t))dt\\
&= \exp\left(-\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)\\
&\quad \exp(-f(X(0)))d\mathcal{M}_t^{(\exp(f))}.
\end{aligned}$$

It follows that $(\mathcal{E}_t[f])_{t\geq 0}$ is a martingale as a stochastic integral (of a previsible integrand) with respect to the martingale $(\mathcal{M}_t^{(\exp(f))})_{t\geq 0}$.
Conversely, we write that for any $t \geq 0$,

$$\begin{aligned}
\exp(f(X(t))) &= \exp(f(X(0)))\\
&\quad \exp\left(\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)\mathcal{E}_t[f],
\end{aligned}$$

thus differentiating we get

$$\begin{aligned}
&d\exp(f(X(t)))\\
&= \exp(f(X(0)))\exp\left(\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)\mathcal{E}_t[f]\\
&\quad \exp(-f(X(t)))L[\exp(f)](X(t))dt\\
&\quad +\exp(f(X(0)))\exp\left(\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)d\mathcal{E}_t[f]\\
&= L[\exp(f)](X(t))dt\\
&\quad +\exp(f(X(0)))\exp\left(\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)d\mathcal{E}_t[f].
\end{aligned}$$

It follows that

$$d\mathcal{M}_t^{(\exp(f))} = \exp(f(X(0)))\exp\left(\int_0^t \exp(-f(X(s)))L[\exp(f)](X(s))\,ds\right)d\mathcal{E}_t[f]$$

and thus $(\mathcal{M}_t^{(\exp(f))})_{t\geq 0}$ is a martingale if $(\mathcal{E}_t[f])_{t\geq 0}$ is a martingale. The second announced result is a consequence from the fact that any function f on S can be written as $f = \exp(g) - c\exp(0)$, with $c > -\min_S f$) and the function $g := \ln(c+f)$, so that $\mathcal{M}^{(f)} = \mathcal{M}^{(\exp(g))} - c\mathcal{M}^{(\exp(0))} = \mathcal{M}^{(\exp(g))}$ is a martingale. □

We need more exponential martingales than those of the form $(\mathcal{E}_t[f])_{t\geq 0}$ where $f \in \mathcal{F}(S)$, the space of functions defined on S. Let us note that for any $f \in \mathcal{F}(S)$, we can write

$$\forall\ t \geq 0, \qquad \mathcal{M}_t^{(f)} = \sum_{x\in S} f(x)\mathcal{M}_t^{(x)}.$$

This suggests to consider the new martingales

$$\forall\, t \geq 0, \qquad \mathcal{M}_t^{(f)} \;=\; \sum_{(x,y)\in S^{(2)}} f(x,y)\mathcal{M}_t^{(x,y)},$$

where $S^{(2)} := S^2 \setminus \{(x,x)\,:\, x \in S\}$ and f is a function defined on $S^{(2)}$. We denote $\mathcal{F}(S^{(2)})$ the space of such functions and we will also see it as a subset of $\mathcal{F}(S^2)$ by extending functions $f \in \mathcal{F}(S^{(2)})$ on S^2 with the convention that they vanish on the diagonal $\{(x,x)\,:\, x \in S\}$. This enables to extend the modified carré du champ I as a non-linear operator from $\mathcal{F}(S^{(2)})$ to $\mathcal{F}(S)$ by action on the second variable. More explicitly, we define

$$\begin{aligned}\forall\, f \in \mathcal{F}(S^{(2)}), \forall\, x \in S,\; I[f](x) \;&:=\; I[f(x,\cdot)](x)\\ &=\; \sum_{y\in S\setminus\{x\}} L(x,y)(\exp(f(x,y)) - 1 - f(x,y)).\end{aligned}$$

Next we associate to any $f \in \mathcal{F}(S^{(2)})$ the process $(\mathcal{E}_t[f])_{t\geq 0}$ via (4.7). We hope that the fact that the same symbols $\mathcal{M}^{(f)}$, $I[f]$ and $\mathcal{E}[f]$ are used for related but different meanings according to $f \in \mathcal{F}(S)$ or $f \in \mathcal{F}(S^{(2)})$ is not leading to confusion.

Proposition 4.9. *For any initial condition μ, the process $(\mathcal{E}_t[f])_{t\geq 0}$ is a martingale under $\mathbb{P}_\mu$.*

Proof: We remark that $(\mathcal{E}_t[f])_{t\geq 0}$ is a multiplicative process, in the sense that

$$\forall\, t,s \geq 0, \qquad \mathcal{E}_{t+s}[f] \;=\; \mathcal{E}_t[f]\mathcal{E}_s[f]\circ\theta_t,$$

where $(\theta_t)_{t\geq 0}$ is the family of the natural shift mappings from $\mathbb{D}$ to itself:

$$\forall\, t,s \geq 0, \forall\, \omega \in \mathbb{D}, \qquad X_s(\theta_t(\omega)) \;=\; X_{t+s}(\omega).$$

Thus taking into account the fact that $(\mathcal{E}_t[f])_{t\geq 0}$ is adapted and the Markov property, we see that to prove that $(\mathcal{E}_t[f])_{t\geq 0}$ is a martingale under $\mathbb{P}_\mu$, it is sufficient to show that

$$\forall\, t \geq 0, \forall\, x \in S, \qquad \mathbb{E}_x[\mathcal{E}_t[f]] \;=\; 1$$

To do so, we will use Proposition 4.8 and an approximation procedure. Let $\epsilon > 0$ be given. We consider the $\mathcal{F}(S)$-valued previsible process $(F_\epsilon(t))_{t\geq 0}$ given by

$$\forall\, t \geq 0, \forall\, x \in S, \qquad F_\epsilon(t)(x) \;:=\; f(X(\epsilon\lfloor t/\epsilon\rfloor), x),$$

where $\lfloor\cdot\rfloor$ stands for a modified integer part: for any $r \in \mathbb{R}_+$, $\lfloor r\rfloor$ is the largest integer number strictly less than r, with the exception that $\lfloor 0\rfloor = 0$ (this is to insure that $(F_\epsilon(t))_{t\geq 0}$ is previsible, in fact this is also true if $\lfloor\cdot\rfloor$ is the usual integer part, but to be convinced of this assertion, one needs to know that the jump times of the underlying Markov processes are totally imprevisible and we don't want to enter in such technicalities, cf. for instance the book of Dellacherie and Meyer, 1980). Next we define

$$\forall\, t \geq 0, \qquad \mathcal{M}_t^{(F_\epsilon)} \;:=\; \sum_{x\in S}\int_0^t F_\epsilon(s)(x)\, d\mathcal{M}_s^{(x)} \tag{4.8}$$

and quite naturally

$$\forall\, t \geq 0, \qquad \mathcal{E}_t(\epsilon, f) \;:=\; \exp\left(\mathcal{M}_t^{(F_\epsilon)} - \int_0^t I[F_\epsilon(s)](X(s))\, ds\right).$$

Since for $t \in [0, \epsilon]$, $F_\epsilon(t)(x)$ does not depend on t and depends on the underlying trajectory $(X(s))_{s\geq 0}$ only through $X(0)$, the first part of Proposition 4.8 shows that for any initial point $x \in S$,

$$\forall\, t \in [0, \epsilon], \qquad \mathbb{E}_x[\mathcal{E}_t(\epsilon, f)] \;=\; 1.$$

More generally, conditioning successively with respect to $\mathcal{D}([0, i\epsilon])$, $\mathcal{D}([0, (i-1)\epsilon])$, ..., $\mathcal{D}([0, \epsilon])$, with $i := \lfloor t/\epsilon \rfloor$, the same argument shows that

$$\forall\, t \geq 0,\, \forall\, x \in S, \qquad \mathbb{E}_x[\mathcal{E}_t(\epsilon, f)] \;=\; 1.$$

Thus to deduce the announced result, it is sufficient to prove that for any fixed $x \in S$, $t \geq 0$ and $f \in \mathcal{F}(S^{(2)})$, $\mathcal{E}_t(\epsilon, f)$ converges to $\mathcal{E}_t[f]$ in $\mathbb{L}^1(\mathbb{P}_x)$ as ϵ goes to 0_+. One way to deduce this convergence is to show on one hand that $\sup_{\epsilon \in (0,1/2]} \mathbb{E}_x[\mathcal{E}_t^2(\epsilon, f)] < +\infty$ and on the other hand that pointwise ($\mathbb{P}_x$-a.s. or even only in probability would be enough), $\mathcal{E}_t(\epsilon, f)$ converges to $\mathcal{E}_t[f]$ as ϵ goes to 0_+. For the first point, we compute that

$$\begin{aligned}
\mathbb{E}_x[\mathcal{E}_t^2(\epsilon, f)] &= \mathbb{E}_x\left[\exp\left(\mathcal{M}_t^{(2F_\epsilon)} - 2\int_0^t I[F_\epsilon](X(s))\, ds\right)\right] \\
&= \mathbb{E}_x\left[\mathcal{E}_t(\epsilon, 2f) \exp\left(\int_0^t (I[2F_\epsilon] - 2I[F_\epsilon])(X(s))\, ds\right)\right] \\
&\leq \exp\left(t \sup_{y,z \in S} I[2f(y,\cdot)](z) - 2I[f(y,\cdot)](z)\right),
\end{aligned}$$

which is a finite upper bound independent from $\epsilon \in (0, 1/2]$.

Concerning the pointwise convergence, we note that for any $s \geq 0$ and $x \in S$,

$$\lim_{\epsilon \to 0_+} F_\epsilon(s)(x) \;=\; f(X(s-), x).$$

Thus, passing in the limit in (4.8) (taking into account that in the finite setting the stochastic differentials $d\mathcal{M}_s^{(x)}$ can be seen as simple Stieljes differentials), we get that

$$\begin{aligned}
\lim_{\epsilon \to 0_+} \mathcal{M}_t^{(F_\epsilon)} &= \sum_{x \in S} \int_0^t f(X(s-), x)\, d\mathcal{M}_s^{(x)} \\
&= \sum_{x \in S} \int_0^t \sum_{y \neq x} f(y, x) \mathbb{1}_{\{y\}}(X(s-))\, d\mathcal{M}_s^{(x)} \\
&= \sum_{(y,x) \in S^{(2)}} f(y, x) \int_0^t \mathbb{1}_{\{y\}}(X(s-)) d\mathcal{M}_s^{(x)} \\
&= \sum_{(x,y) \in S^{(2)}} f(x, y) \mathcal{M}_t^{(x,y)} \\
&= \mathcal{M}_t^{(f)},
\end{aligned}$$

where the last but one equality comes from (4.6).
We also deduce that

$$\begin{aligned}\lim_{\epsilon\to 0_+}\int_0^t I[F_\epsilon(s)](X(s))\,ds &= \int_0^t I[f(X(s-),\cdot)](X(s))\,ds\\ &= \int_0^t I[f(X(s),\cdot)](X(s))\,ds\\ &= \int_0^t I[f](X(s))\,ds\end{aligned}$$

(where in the second equality we have used that $X(s-)=X(s)$ almost everywhere with respect to the Lebesgue measure ds), result leading to the wanted pointwise convergence. □

Let the probability μ on S and the function $f\in\mathcal{F}(S^{(2)})$ be fixed. For any $t\geq 0$, consider $\widetilde{\mathbb{P}}_{\mu,[0,t]}$ the probability on the measurable space $(\mathbb{D},\mathcal{D}([0,t]))$ which admits $\mathcal{E}_t[f]$ as density with respect to $\mathbb{P}_{\mu,[0,t]}$. Since the process $(\mathcal{E}_s[f])_{s\geq 0}$ is a nonnegative martingale starting from 1, it is easy to deduce that the family $(\widetilde{\mathbb{P}}_{\mu,[0,t]})_{t\geq 0}$ satisfies the compatibility relations of Kolmogorov: for $0\leq s\leq t$, the restriction of $\widetilde{\mathbb{P}}_{\mu,[0,t]}$ to $(\mathbb{D},\mathcal{D}([0,s]))$ is equal to $\widetilde{\mathbb{P}}_{\mu,[0,s]}$. By consequence there exists a unique probability measure $\widetilde{\mathbb{P}}_\mu$ on $(\mathbb{D},\mathcal{D})$ such that for any $t\geq 0$, the restriction of $\widetilde{\mathbb{P}}_\mu$ to $(\mathbb{D},\mathcal{D}([0,t]))$ coincides with $\widetilde{\mathbb{P}}_{\mu,[0,t]}$. Next result is the key to Theorem 4.6.

Theorem 4.10. *The probability $\widetilde{\mathbb{P}}_\mu$ is solution to the martingale problem associated to μ and to the generator $\widetilde{L}$ defined by*

$$\forall\ x,y\in S,\qquad \widetilde{L}(x,y)\ :=\ \begin{cases} L(x,y)\exp(f(x,y)) & ,\ \text{if } x\neq y,\\ -\sum_{z\in S\setminus\{x\}}\widetilde{L}(x,z) & ,\ \text{if } x=y.\end{cases}$$

Proof: We use the criterion presented in Proposition 4.8. The law of $X(0)$ under $\widetilde{\mathbb{P}}_\mu$ is μ since $\widetilde{\mathbb{P}}_\mu$ and $\mathbb{P}_\mu$ coincides on the σ-field $\mathcal{D}([0,0])=\sigma(X(0))$. So let $g\in\mathcal{F}(S)$ be a test function. We want to show that $(\widetilde{\mathcal{E}}_t[g])_{t\geq 0}$ is a martingale under $\widetilde{\mathbb{P}}_\mu$, where

$$\forall\ t\geq 0,\qquad \widetilde{\mathcal{E}}_t[g]\ :=\ \exp\left(\widetilde{\mathcal{M}}_t^{(g)}-\int_0^t\widetilde{I}[g](X(s))\,ds\right),$$

with of course,

$$\begin{aligned}\forall\ t\geq 0,\qquad \widetilde{\mathcal{M}}_t^{(g)} &:= g(X(t))-g(X(0))-\int_0^t\widetilde{L}[g](X(s))\,ds,\\ &= \mathcal{M}_t^{(g)}-\int_0^t(\widetilde{L}-L)[g](X(s))\,ds\\ \forall\ x\in S,\qquad \widetilde{I}[f](x) &:= \exp(-f(x))\widetilde{L}[\exp(f)](x)-\widetilde{L}[f](x).\end{aligned}$$

Due to the definition of $\widetilde{\mathbb{P}}_\mu$, $(\widetilde{\mathcal{E}}_t[g])_{t\geq 0}$ is a martingale under $\widetilde{\mathbb{P}}_\mu$ if and only if $(\widetilde{\mathcal{E}}_t[g]\mathcal{E}_t[f])_{t\geq 0}$ is a martingale under $\mathbb{P}_\mu$. But for any $t\geq 0$, we compute that

$$\widetilde{\mathcal{E}}_t[g]\mathcal{E}_t[f]\ =\ \exp\left(\mathcal{M}_t^{(g)}+\mathcal{M}_t^{(f)}-\int_0^t F(X(s))\,ds\right),$$

where $F\in\mathcal{F}(S)$ is the function defined by, for any $x\in S$,

$$
\begin{aligned}
F(x) &:= (\widetilde{L}-L)[g](x)+\exp(-g(x))\widetilde{L}[\exp(g)](x)-\widetilde{L}[g](x)\\
&\quad +L[\exp(f(x,\cdot)](x)-L[f(x,\cdot)](x)\\
&= \exp(-g(x))\widetilde{L}[\exp(g)](x)+L[\exp(f(x,\cdot)](x)-L[g+f(x,\cdot)](x)\\
&= \sum_{y\neq x}\Big((\exp(g(y)-g(x))-1)\widetilde{L}(x,y)+(\exp(f(x,y))-1)L(x,y)\Big)\\
&\quad -L[g+f(x,\cdot)](x)\\
&= \sum_{y\neq x}\big([\exp(g(y)-g(x))-1]\exp(f(x,y))+\exp(f(x,y))-1\big)L(x,y)\\
&\quad -L[g+f(x,\cdot)](x)\\
&= \sum_{y\neq x}\left(\exp[g(y)-g(x)+f(x,y)]-1\right)L(x,y)-L[g+f(x,\cdot)](x)\\
&= I[h](x),
\end{aligned}
$$

with $h\in\mathcal{F}(S^{(2)})$ given by

$$\forall\ (x,y)\in S^{(2)},\qquad h(x,y) \quad:=\quad g(y)-g(x)+f(x,y)$$

Let us check that

$$\forall\ t\geq 0,\qquad \mathcal{M}_t^{(g)}+\mathcal{M}_t^{(f)} \quad=\quad \mathcal{M}_t^{(h)}.$$

It amounts to verify that $\mathcal{M}^{(g)}=\mathcal{M}^{(\bar{g})}$ where $\bar{g}\in\mathcal{F}(S^{(2)})$ is defined by

$$\forall\ (x,y)\in S^{(2)},\qquad \bar{g}(x,y) \quad:=\quad g(y)-g(x).$$

Indeed, we have for any $t\geq 0$,

$$
\begin{aligned}
\mathcal{M}_t^{(\bar{g})} &= \sum_{(x,y)\in S^{(2)}}(g(y)-g(x))\mathcal{M}_t^{(x,y)}\\
&= \sum_{(x,y)\in S^{(2)}}(g(y)-g(x))\int_0^t \mathbb{1}_{\{x\}}(X(s-))\,d\mathcal{M}_s^{(y)}\\
&= \sum_{(x,y)\in S^{2}}(g(y)-g(x))\int_0^t \mathbb{1}_{\{x\}}(X(s-))\,d\mathcal{M}_s^{(y)}\\
&= \sum_{y\in S}g(y)\int_0^t\sum_{x\in S}\mathbb{1}_{\{x\}}(X(s-))\,d\mathcal{M}_s^{(y)}\\
&\quad -\sum_{x\in S}g(x)\int_0^t\mathbb{1}_{\{x\}}(X(s-))\,d\left(\sum_{y\in S}\mathcal{M}_s^{(y)}\right)\\
&= \sum_{y\in S}g(y)\int_0^t d\mathcal{M}_s^{(y)}-\sum_{x\in S}g(x)\int_0^t\mathbb{1}_{\{x\}}(X(s-))\,d\mathcal{M}_s^{(\mathbb{1})})\\
&= \sum_{y\in S}g(y)\mathcal{M}_t^{(y)}\\
&= \mathcal{M}_t^{(g)},
\end{aligned}
$$

where for the last but one equality we have used that the martingale associated to the function identically equal to one, $\mathcal{M}^{(\mathbb{1})}$, is null.

So finally we get that $(\widetilde{\mathcal{E}}_t[g]\mathcal{E}_t[f])_{t\geq 0}$=$(\mathcal{E}_t[h])_{t\geq 0}$ is a martingale under $\mathbb{P}_\mu$ according to Proposition 4.9. □

We can now come to the

Proof: of Theorem 4.6 So let be given a generator $\widetilde{L}$ satisfying $\widetilde{L} \ll L$. We begin by assuming that $\widetilde{L}$ satisfies furthermore $\widetilde{L} \gg L$. Then we consider the previous construction with the function $f \in \mathcal{F}(S^{(2)})$ defined by

$$\forall\ (x,y) \in S^{(2)}, \qquad f(x,y) \quad := \quad \ln\left(\frac{\widetilde{L}(x,y)}{L(x,y)}\right).$$

According to Theorem 4.10, we have

$$\forall\ T \geq 0, \qquad \frac{d\widetilde{\mathbb{P}}_{\mu,[0,T]}}{d\mathbb{P}_{\mu,[0,T]}} \quad = \quad \mathcal{E}_T[f]$$

and this can be rewritten under the form given in Theorem 4.6.
In the general case, we introduce for $\epsilon > 0$ the new generator $\widetilde{L}^{(\epsilon)} := \widetilde{L} + \epsilon L$, to which we apply the above result. Then with obvious notations, we get that for any $T > 0$,

$$\begin{aligned}
&\lim_{\epsilon\to 0_+} \exp\left(\sum_{x\neq y\in S} \ln(A^{(\epsilon)}(x,y))N_T^{(x,y)} + \int_0^T H^{(\epsilon)}(X(s))\,ds\right)\\
&\quad = \quad \exp\left(\sum_{x\neq y\in S} \ln(A(x,y))N_T^{(x,y)} + \int_0^T H(X(s))\,ds\right).
\end{aligned}$$

Furthermore this pointwise convergence also takes place in $\mathbb{L}^p(\mathbb{P}_{\mu,[0,T]})$, for any $p \geq 1$, because it is not difficult to find three constants $K_1, K_2, K_3 \geq 0$ (depending only on T, L and $\widetilde{L}$) such that for any $\epsilon \in (0,1]$, the random variable $\sum_{x\neq y\in S} \ln(A^{(\epsilon)}(x,y))N_T^{(x,y)} + \int_0^T H^{(\epsilon)}(X(s))\,ds$ is stochastically bounded above by $K_1N + K_2$, where N is a Poisson variable of parameter K_3, and the latter random variable admits exponential moments of all order. We deduce that as $\epsilon \to 0_+$, the probability $\mathbb{P}^{(\epsilon)}_{\mu,[0,T]}$ converges in total variation toward the probability $\widehat{\mathbb{P}}_{\mu,[0,T]}$ on $(\mathbb{D}, \mathcal{D}([0,T]))$ which has $\exp\left(\sum_{x\neq y\in S} \ln(A(x,y))N_T^{(x,y)} + \int_0^T H(X(s))\,ds\right)$ as density with respect to $\mathbb{P}_{\mu,[0,T]}$. It follows that the family $(\widehat{\mathbb{P}}_{\mu,[0,T]})_{T\geq 0}$ satisfies the Kolmogorov compatibility conditions, so that its elements are the restrictions to the σ-fields $(\mathcal{D}([0,T]))_{T\geq 0}$ of a probability $\widehat{\mathbb{P}}_\mu$ on $(\mathbb{D}, \mathcal{D})$. Furthermore the above convergence in total variation and the convergence of $\widetilde{L}^{(\epsilon)}$ to $\widetilde{L}$ imply that $\widehat{\mathbb{P}}_\mu$ is the solution of the martingale problem associated to μ and $\widetilde{L}$, i.e. $\widehat{\mathbb{P}}_\mu = \widetilde{\mathbb{P}}_\mu$. Thus the conclusions of Theorem 4.6 also hold in the general case. □

We end up this appendix with two remarks.

Remark 4.11. Let us justify the already announced fact that the necessary condition given in Lemma 4.4 is also sufficient. So let $\widetilde{L}$ a generator satisfying the conclusions of this lemma. Let $\widehat{L}$ be another generator such that

$$\forall\, x \in \widetilde{S}_\mu,\ \forall\, y \in S, \qquad \widehat{L}(x,y) \;=\; \widetilde{L}(x,y).$$

Then $\widetilde{\mathbb{P}}_\mu$ is also solution to the martingale problem associated to μ and $\widehat{L}$, because for any $f \in \mathcal{F}(S)$ and any $t \geq 0$ we have $\int_0^t \widetilde{L}[f](X(s))\,ds = \int_0^t \widehat{L}[f](X(s))\,ds$, $\widetilde{\mathbb{P}}_\mu$-a.s. In particular consider $\widehat{L}$ defined by

$$\forall\, x,y \in S, \qquad \widehat{L}(x,y) \;:=\; \begin{cases} \widetilde{L}(x,y) & \text{, if } x \in \widetilde{S}_\mu, \\ 0 & \text{, otherwise.} \end{cases}$$

Since $\widehat{L} \ll L$, we can apply Theorem 4.6 to get that for any $T \geq 0$, $\widetilde{\mathbb{P}}_{\mu,[0,T]} \ll \mathbb{P}_{\mu,[0,T]}$ (and the expressions for the density given in Theorem 4.6 are also valid, with $\widetilde{L}$ and $\widehat{L}$).

The second remark goes in the direction of an abstract Girsanov theory.

Remark 4.12. The carré du champ alluded to in Remark 4.7 can be polarized into a bilinear map from $\mathcal{F}(S)^2$ to $\mathcal{F}(S)$ via the formula

$$\forall\, f,g \in \mathcal{F}(S),\ \forall\, x \in S, \qquad \Gamma[f,g](x) \;:=\; L[fg](x) - f(x)L[g](x) - g(x)L[f](x).$$

Next, similarly to what we have done before Proposition 4.9, we can extend Γ to $\mathcal{F}(S^2)$ by action on the second variable:

$$\forall\, f,g \in \mathcal{F}(S^2),\ \forall\, x \in S, \qquad \Gamma[f,g](x) \;:=\; \Gamma[f(x,\cdot),g(x,\cdot)](x)$$

Now let a function $f \in \mathcal{F}(S^{(2)})$ be fixed and consider the generator $\widetilde{L}$ constructed in Theorem 4.10. Then its action on functions can be expressed under the form

$$\forall\, g \in \mathcal{F}(S),\ \forall\, x \in S, \qquad \widetilde{L}[g](x) \;=\; L[g](x) + \Gamma[\exp(f),g](x)$$

(where g in the carré du champ is considered as a function of two variables via $g(x,y) = g(y)$ for any $x,y \in S$).

This procedure associating to a bivariate function $f \in \mathcal{F}(S^{(2)})$ an exponential martingale $(\mathcal{E}_t[f])_{t\geq 0}$ which leads to the construction of the solution to the martingale problem corresponding to the operator $\widetilde{L}$ given in the above form is in fact very general Indeed, under regularity assumptions, we can recover in this way all the transformations of the underlying probability of a Markov process which preserve the Markov property and which are absolutely continuous over finite time horizons. These generalized Girsanov transformations were first exihibited by Kunita (1969) which used another formalism. Heuristically, these changes amounts to modifying the intensity of jumps and adding drifts in the directions permitted by the diffusion coefficients. According to the above formulation, they can be parameterized by functions belonging to subspaces of $\mathcal{F}(S^{(2)})$ (satisfying some regularity or boundedness assumptions in general) quotiented by the subspace of bivariate functions whose carré du champ vanishes everywhere. For instance starting with the Laplacian generator over a Riemannian compact manifold, we recover all the vector fields as parameters of Girsanov transformations in this setting, see also the discussion after Proposition 4.14 in Appendix 2.

Appendix 2: the usual Girsanov theorem

We present here this famous result in the compact diffusion framework of Section 4.

First we recall a preliminary condition for absolute continuity on finite time horizons, relatively to generators given in the form (4.1), so no Riemannian structure is necessary.

Proposition 4.13. *Let μ and $\widetilde{\mu}$ be two initial distributions, if for some $T>0$ we have $\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([0,T])) \ll \mathcal{L}(X^{(\mu)}([0,T]))$ then on any chart C, we have*

$$\forall\ x \in C, \qquad a(x) \quad = \quad \widetilde{a}(x).$$

Proof: First we assume that $\mu = \widetilde{\mu} = \delta_x$, the Dirac mass at some fixed point $x \in S$ and we consider C an open chart containing x. Let us show that in this chart, we have $a(x) = \widetilde{a}(x)$.
For $i \in [\![1,n]\!]$, we denote by x_i the i^{th} coordinate mapping in C centered in x (namely $x_i(x) = 0$ for all $i \in [\![1,n]\!]$). Let $(v_i)_{i\in[\![1,n]\!]} \in \mathbb{R}^n$ be a given vector and consider C' an open set of S satisfying $x \in C'$ and $\bar{C}' \subset C$. It is possible to extend the mapping $C' \ni x \mapsto \sum_{i\in[\![1,n]\!]} v_i x_i$ into a function $f \in \mathcal{C}^\infty$. Since the martingale $\mathcal{M}^{(f)}$ associated to f is continuous, the iterated logarithm law (see for instance the book of Revuz and Yor, 1999) shows that a.s.

$$\limsup_{t\to 0_+} \frac{\mathcal{M}_t^{(f)}}{\sqrt{\langle \mathcal{M}^{(f)}\rangle_t \ln(\ln(1/\langle \mathcal{M}^{(f)}\rangle_t))}} \quad = \quad 1,$$

where according to Remark 4.7, the bracket $\langle \mathcal{M}^{(f)}\rangle$ of the martingale $\mathcal{M}^{(f)}$ is given by

$$\forall\ t \geq 0, \qquad \langle \mathcal{M}^{(f)}\rangle_t \quad = \quad \int_0^t \Gamma[f](X^{(\delta_x)}(s))\, ds.$$

We compute that on any chart C'', the carré du champ $\Gamma[f]$ can be written down as

$$\forall\ y \in C'', \qquad \Gamma[f](y) \quad = \quad \sum_{i,j\in[\![1,n]\!]} a_{i,j}(y)\partial_i f(y) \partial_j f(y).$$

In particular, since a.s. $\lim_{t\to 0_+} X^{(\delta_x)}(t) = x$, it follows that as t goes to zero,

$$\begin{aligned}\langle \mathcal{M}^{(f)}\rangle_t \quad &\sim \quad t \sum_{i,j\in[\![1,n]\!]} a_{i,j}(x) v_i v_j \\ &= \quad t v^{\mathrm{t}} a(x) v\end{aligned}$$

and we get that a.s.,

$$\begin{aligned}\limsup_{t\to 0_+} \frac{f(X^{(\delta_x)}(t))}{\sqrt{t\ln(\ln(1/t))}} \quad &= \quad \limsup_{t\to 0_+} \frac{\mathcal{M}_t^{(f)}}{\sqrt{t\ln(\ln(1/t))}} \\ &= \quad \sqrt{v^{\mathrm{t}} a(x) v}.\end{aligned}$$

Similarly we have a.s.

$$\limsup_{t\to 0_+} \frac{f(\widetilde{X}^{(\delta_x)}(t))}{\sqrt{t\ln(\ln(1/t))}} \quad = \quad \sqrt{v^{\mathrm{t}} \widetilde{a}(x) v},$$

so we deduce that for any vector $v \in \mathbb{R}^n$,

$$v^{\mathrm{t}}\widetilde{a}(x)v \quad = \quad v^{\mathrm{t}}a(x)v.$$

Since $a(x)$ and $\widetilde{a}(x)$ are symmetrical, this relation implies that $\widetilde{a}(x) = a(x)$.

More generally, for any initial distributions μ and $\widetilde{\mu}$, the above arguments show that for any function $f \in \mathcal{C}^\infty$, we have a.s.,

$$\begin{aligned}
\limsup_{t\to 0_+} \frac{f(X^{(\mu)}(t))}{\sqrt{t\ln(\ln(1/t))}} &= \sqrt{\Gamma[f](X^{(\mu)}(0))},\\
\limsup_{t\to 0_+} \frac{f(\widetilde{X}^{(\widetilde{\mu})}(t))}{\sqrt{t\ln(\ln(1/t))}} &= \sqrt{\widetilde{\Gamma}[f](X^{(\widetilde{\mu})}(0))}.
\end{aligned}$$

Assume now that the support of $\widetilde{\mu}$ is the whole state space S and that the conclusion of the above proposition is not satisfied. Then we can find a chart C, a point $x_0 \in C$ and a vector $v \in \mathbb{R}^n$ such that $\eta = \beta - \widetilde{\beta} \neq 0$, where $\beta := v^{\mathrm{t}}a(x_0)v$ and $\widetilde{\beta} := v^{\mathrm{t}}\widetilde{a}(x_0)v$. Let $f \in \mathcal{C}^\infty$ be a function as above which coincides with $x \mapsto \sum_{i\in[\![1,n]\!]} v_i x_i$ on a neighborhood $C' \subset C$ of x_0. Then we can find another neighborhood $C'' \subset C'$ such that for any $x \in C''$,

$$\begin{aligned}
|\Gamma[f](x) - \beta| &\leq \eta/3,\\
\left|\widetilde{\Gamma}[f](x) - \widetilde{\beta}\right| &\leq \eta/3.
\end{aligned}$$

Assume for instance that $\beta > \widetilde{\beta}$ and consider in the space of continuous trajectories from $[0, T]$ to S the event

$$E \quad := \quad \left\{X(0) \in C'',\ \limsup_{t\to 0_+} \frac{f(X(t))}{\sqrt{t\ln(\ln(1/t))}} \leq \frac{\beta + \widetilde{\beta}}{2}\right\},$$

where $(X(t))_{t\in[0,T]}$ is the canonical coordinate process. Then under $\mathcal{L}(X^{(\mu)}([0, T]))$, E is negligible while it has probability $\widetilde{\mu}(C'') > 0$ under $\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([0, T]))$ and this is contradictory with our absolute continuity assumption. So the conclusion of the above proposition also holds if the support of $\widetilde{\mu}$ is S.

Finally we consider the general case for the initial distributions μ and $\widetilde{\mu}$. Due to the compactness of S and our assumption on the diffusive matrix fields, the operator $\widetilde{L}$ is uniformly elliptic, so for $T > 0$ the law $\widetilde{\mu}_{T/2}$ of $X^{(\mu)}_{T/2}$ admits a smooth and positive density and in particular its support is S. Thus applying the above arguments to $\mathcal{L}(X^{(\mu)}([T/2, T]))$ and $\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([T/2, T]))$ we get the announced conclusion, at least if we know that $\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([T/2, T])) \ll \mathcal{L}(X^{(\mu)}([T/2, T]))$. But this is an immediate consequence of the assumption $\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([0, T])) \ll \mathcal{L}(X^{(\mu)}([0, T]))$. $\square$

Now we assume that the compact manifold S is endowed with a Riemannian structure and that the generators are written under the form (4.2). The usual Girsanov formula can then be stated as:

Proposition 4.14. *For any initial distribution μ and any finite time horizon $T \geq 0$, the law $\mathcal{L}(\widetilde{X}^{(\mu)}([0, T]))$ is absolutely continuous with respect to $\mathcal{L}(X^{(\mu)}([0, T]))$ and the corresponding Radon-Nikodym density is equal to*

$$\frac{d\mathcal{L}(\widetilde{X}^{(\mu)}([0, T]))}{d\mathcal{L}(X^{(\mu)}([0, T]))} \quad = \quad \exp\left(\mathcal{M}_T^{(\widetilde{b}-b)} - \frac{1}{2}\int_0^T \left|\widetilde{b} - b\right|^2 (X(t))\, dt\right),$$

where $(\mathcal{M}_t^{(\widetilde{b}-b)})_{t\geq 0}$ *is a martingale whose bracket is given by*

$$\forall\ t\geq 0,\qquad \langle \mathcal{M}^{(\widetilde{b}-b)}\rangle_t \quad=\quad \int_0^t \left|\widetilde{b}-b\right|^2 (X(s))\,ds$$

(in these formulas $(X(t))_{t\in[0,T]}$ *stands for a generic trajectory, in the proof below it will coincides with* $(X^{(\mu)}(t))_{t\in[0,T]}$*, which is a more natural notation when working under* $\mathcal{L}(X^{(\mu)}([0,T]))$*).*

This result can be proven by an approach formally similar to that of Appendix 1. Indeed, if F is a smooth function on S^2 such that

$$\forall\ x\in S,\qquad \nabla_y F(x,y)_{|y=x} \quad=\quad \widetilde{b}(x)-b(x),$$

then with the notations of Appendix 1, $\mathcal{M}^{(\widetilde{b}-b)}$ corresponds to $\mathcal{M}^{(F)}$ (see also Remark 4.12). The definition of the martingales associated to functions of two variables goes through an approximation procedure, starting with simple functions of the kind $F=f\otimes g$, with $f,g\in\mathcal{C}^\infty$, for which we take

$$\forall\ t\geq 0,\qquad \mathcal{M}_t^{(F)} \quad:=\quad \int_0^t f(X^{(\mu)}(s-))\,d\mathcal{M}_s^{(g)}. \tag{4.9}$$

The completion of the vector space generated by functions of the form $F=f\otimes g$, with $f,g\in\mathcal{C}^\infty$, is done with respect to the semi-norm $\|\cdot\|$ on $\mathcal{C}^\infty(S^2)$ defined by

$$\forall\ F\in\mathcal{C}^\infty(S^2),\qquad \|F\| \quad:=\quad \sup_{x\in S}\Gamma[F(x,\cdot)](x), \tag{4.10}$$

where Γ is the carré du champ associated to L (acting on the second variable in the above formula).

In the diffusion framework the computation of the Radon-Nikodym densities are even easier, for instance the martingale

$$(\mathcal{E}_t[\widetilde{b}-b])_{t\geq 0} \quad:=\quad \left(\exp\left(\mathcal{M}_t^{(\widetilde{b}-b)}-\frac{1}{2}\int_0^t\left|\widetilde{b}-b\right|^2(X^{(\mu)}(s)\,ds\right)\right)_{t\geq 0} \tag{4.11}$$

is the Doléans-Dade exponential of the martingale $\mathcal{M}^{(\widetilde{b}-b)}$, namely the solution of the s.d.e.

$$\begin{aligned} \mathcal{E}_0[\widetilde{b}-b] &= 1,\\ \forall\ s\geq 0,\qquad d\mathcal{E}_s[\widetilde{b}-b] &= \mathcal{E}_{s-}[\widetilde{b}-b]\,d\mathcal{M}_s^{(\widetilde{b}-b)} \end{aligned}$$

(in the diffusion setting, the $s-$ (indicating left limit) in the above formula and in (4.9) can be replaced by s).
But for our restricted purposes here, it is not necessary to develop this approach and we resort to a more traditional point of view:

Proof: of Proposition 4.14 To construct directly the martingale $\mathcal{M}^{(\widetilde{b}-b)}$, we imbed S isometrically into $\mathbb{R}^N$ via the Nash's theorem (see for instance the book of Han and Hong, 2006). Let $(e_i)_{i\in[\![1,N]\!]}$ be the canonical basis of $\mathbb{R}^N$. For each $i\in[\![1,N]\!]$ and $x\in S$, we denote $v_i(x)\in \mathrm{T}_xS$ the orthogonal projection of e_i on the vector space T_xS (seen as a vector subspace of the Euclidean space $\mathbb{R}^N$). If $X^{(\mu)}(0)$ is given and distributed according to μ on S and if $((W_i(t))_{t\geq 0})_{i\in[\![1,N]\!]}$ is a family of independent standard one-dimensional Brownian motions (furthermore independent from $X^{(\mu)}(0)$), it is well-known that the solution $(X^{(\mu)}(t))_{t\geq 0}$ of s.d.e.

$$\forall\ t\geq 0,\qquad dX^{(\mu)}(t) \;=\; \sum_{i\in[\![1,N]\!]} v_i(X^{(\mu)}(t))\,dW_i(t)+b(X^{(\mu)}(t))\,dt$$

stays in S and that its law is solution to the martingale problem associated to μ and L. Indeed, using It's formula we get that for any $f\in\mathcal{C}^\infty$,

$$f(X^{(\mu)}(t)) \;=\; f(X^{(\mu)}(0))+\int_0^t L[f](X^{(\mu)}(s))\,ds+\mathcal{M}_t^{(f)},$$

where $\mathcal{M}^{(f)}$ is the martingale defined by

$$\forall\ t\geq 0,\qquad \mathcal{M}_t^{(f)} \;:=\; \sum_{i\in[\![1,N]\!]}\int_0^t \langle v_i,\nabla f\rangle(X^{(\mu)}(s))\,dW_i(s).$$

Now we define the martingale $\mathcal{M}^{(\widetilde{b}-b)}$ by

$$\forall\ t\geq 0,\qquad \mathcal{M}_t^{(\widetilde{b}-b)} \;:=\; \sum_{i\in[\![1,N]\!]}\int_0^t \langle\widetilde{b}-b,v_i\rangle(X^{(\mu)}(s))\,dW_i(s)$$

and we compute that its bracket is given by

$$\begin{aligned}\forall\ t\geq 0,\qquad \langle\mathcal{M}^{(\widetilde{b}-b)}\rangle_t \;&:=\; \sum_{i\in[\![1,N]\!]}\int_0^t \langle\widetilde{b}-b,v_i\rangle^2(X^{(\mu)}(s))\,ds\\ &=\; \sum_{i\in[\![1,N]\!]}\int_0^t \langle(\widetilde{b}-b)(X^{(\mu)}(s)),e_i\rangle^2\,ds\\ &=\; \int_0^t \left|\widetilde{b}-b\right|^2(X^{(\mu)}(s))\,ds.\end{aligned}$$

Since the martingale $\mathcal{M}^{(\widetilde{b}-b)}$ is continuous, we deduce that the process $\mathcal{E}[\widetilde{b}-b]$ defined in (4.11) is a positive martingale starting from 1. Thus via Kolmogorov's theorem, we can use it to construct a probability measure $\widetilde{\mathbb{P}}_\mu$ on the σ-algebra generated by $X^{(\mu)}$ by imposing that on $\sigma(X_t^{(\mu)}\,:\,t\in[0,T])$, $d\widetilde{\mathbb{P}}_\mu/d\mathbb{P}_\mu=\mathcal{E}_T[\widetilde{b}-b]$, for any $T\geq 0$.

Next for any function $f\in\mathcal{C}^\infty$, we get the co-bracket of $\mathcal{M}^{(\widetilde{b}-b)}$ and $\mathcal{M}^{(f)}$ by a similar computation:

$$\begin{aligned}\forall\ t\geq 0,\ \langle\mathcal{M}^{(\widetilde{b}-b)},\mathcal{M}^{(f)}\rangle_t \;&:=\; \sum_{i\in[\![1,N]\!]}\int_0^t \langle\widetilde{b}-b,v_i\rangle(X^{(\mu)}(s))\langle\nabla f,v_i\rangle(X^{(\mu)}(s))\,ds\\ &=\; \int_0^t \langle\widetilde{b}-b,\nabla f\rangle(X^{(\mu)}(s))\,ds.\end{aligned}$$

It follows that the process

$$\left(\left(\mathcal{M}_t^{(f)}-\int_0^t\langle\widetilde{b}-b,\nabla f\rangle(X^{(\mu)}(s))\,ds\right)\mathcal{E}_t[\widetilde{b}-b]\right)_{t\geq 0}$$

is a martingale under $\mathbb{P}_\mu$, fact which is immediately translated into the assertion that the process $\widetilde{\mathcal{M}}^{(f)}$ defined by

$$\begin{aligned}\forall\, t\in\mathbb{R}_+,\qquad \widetilde{\mathcal{M}}_t^{(f)} &:= \mathcal{M}_t^{(f)}-\int_0^t\langle\widetilde{b}-b,\nabla f\rangle(X^{(\mu)}(s))\,ds\\ &= f(X^{(\mu)}(t))-f(X^{(\mu)}(0))-\int_0^t\widetilde{L}[f](X^{(\mu)}(s))\,ds\end{aligned}$$

is a martingale under $\widetilde{\mathbb{P}}_\mu$. Thus the image of $\widetilde{\mathbb{P}}_\mu$ under the process $X^{(\mu)}$ is the solution to the martingale problem associated to μ and $\widetilde{L}$ and the validity of the above proposition follows. $\square$

This absolute continuity property enables to obtain a converse result to Proposition 4.13:

Remark 4.15. By conditioning with respect to $X^{(\mu)}(0)$ and applying Proposition 4.14 with μ a Dirac mass, we get the equivalence between the following assertions:

- We have $\widetilde{\mu}\ll\mu$ and $\widetilde{a}=a$.
- There exists $T>0$ such that $\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([0,T]))\ll\mathcal{L}(X^{(\mu)}([0,T]))$.
- For any $T\geq 0$, we have $\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([0,T]))\ll\mathcal{L}(X^{(\mu)}([0,T]))$.

Furthermore in this situation, the Radon-Nikodym density is given, for any $T\geq 0$, by

$$\frac{d\mathcal{L}(\widetilde{X}^{(\widetilde{\mu})}([0,T]))}{d\mathcal{L}(X^{(\mu)}([0,T]))} = \frac{d\widetilde{\mu}}{d\mu}(X^{(\mu)}(0))\exp\left(\mathcal{M}_T^{(\widetilde{b}-b)}-\frac{1}{2}\int_0^T\left|\widetilde{b}-b\right|^2(X^{(\mu)}(t))\,dt\right).$$

Acknowledgement: We are grateful to the ANR's Chaire d'Excellence program for providing financial support, as well as to the Institut de Mathématiques (UMR 5219) in Toulouse, where parts of this work were done, for its hospitality. We are also indebted to the referee for its careful reading which helped to improve the readability of the paper.

References

Dominique Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on probability theory (Saint-Flour, 1992)*, volume 1581 of *Lecture Notes in Math.*, pages 1–114. Springer, Berlin (1994). MR1307413.

Louis J. Billera and Persi Diaconis. A geometric interpretation of the Metropolis-Hastings algorithm. *Statist. Sci.* **16** (4), 335–339 (2001). ISSN 0883-4237. MR1888448.

Patrick Cattiaux and Christian Léonard. Minimization of the Kullback information of diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **30** (1), 83–132 (1994). ISSN 0246-0203. MR1262893.

Donald A. Dawson and Jürgen Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics* **20** (4), 247–308 (1987). ISSN 0090-9491. MR885876.

Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel. Chapitres V à VIII*, volume 1385 of *Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics]*. Hermann, Paris, revised edition (1980). ISBN 2-7056-1385-4. Théorie des martingales. [Martingale theory]; MR566768.

Amir Dembo and Ofer 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.

Michel Émery. *Stochastic calculus in manifolds*. Universitext. Springer-Verlag, Berlin (1989). ISBN 3-540-51664-6. With an appendix by P.-A. Meyer; MR1030543.

Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York (1986). ISBN 0-471-08186-8. Characterization and convergence; MR838085.

George S. Fishman. *Monte Carlo*. Springer Series in Operations Research. Springer-Verlag, New York (1996). ISBN 0-387-94527-X. Concepts, algorithms, and applications; MR1392474.

I. V. Girsanov. On transforming a class of stochastic processes by absolutely continuous substitution of measures. *Teor. Verojatnost. i Primenen.* **5**, 314–330 (1960). ISSN 0040-361x. MR0133152.

J. M. Hammersley and D. C. Handscomb. *Monte Carlo methods*. Methuen & Co. Ltd., London (1965). MR0223065.

Qing Han and Jia-Xing Hong. *Isometric embedding of Riemannian manifolds in Euclidean spaces*, volume 130 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI (2006). ISBN 978-0-8218-4071-9; 0-8218-4071-1. MR2261749.

Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition (1989). ISBN 0-444-87378-3. MR1011252.

Hiroshi Kunita. Absolute continuity of Markov processes and generators. *Nagoya Math. J.* **36**, 1–26 (1969). ISSN 0027-7630. MR0250387.

Jun S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York (2008). ISBN 978-0-387-76369-9; 0-387-95230-6. MR2401592.

N. Metropolis, A. Rosenbluth, R. Rosenbluth, A. Teller and E. Teller. *Equation of state calculations by fast computing machines*, volume 21. J. Chem. Phys. (1953).

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition (1999). ISBN 3-540-64325-7. MR1725357.

Laurent Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin (1997). MR1490046.

Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin (2006). ISBN 978-3-540-28998-2; 3-540-28998-4. Reprint of the 1997 edition; MR2190038.