Trajectorial Dissipation and Gradient Flow for the Relative Entropy in Markov Chains *

Ioannis Karatzas † Jan Maas ‡ Walter Schachermayer §

May 28, 2020

Abstract

We study the temporal dissipation of variance and relative entropy for ergodic Markov Chains in continuous time, and compute explicitly the corresponding dissipation rates. These are identified, as is well known, in the case of the variance in terms of an appropriate Hilbertian norm; and in the case of the relative entropy, in terms of a Dirichlet form which morphs into a version of the familiar Fisher information under conditions of "detailed balance". Here we obtain trajectorial versions of these results, which are valid along almost every path of the random motion and most transparent in the backwards direction of time. Martingale arguments and time reversal play crucial roles, as in the recent work of Karatzas, Schachermayer and Tschiderer for conservative diffusions. Extensions are developed to general "convex divergences" and to countable state-spaces. The steepest descent and gradient flow properties for the variance, the relative entropy, and appropriate generalizations, are studied along with their respective geometries under conditions of detailed balance.

Keywords and Phrases: Markov Chain; Relative Entropy; Time Reversal; Steepest Descent; Gradient Flow.
AMS 2020 Subject Classifications: 60J27; 60H10; 60G44; 46C05.

1 Introduction and Summary

We present a trajectorial approach to the temporal dissipation of variance and relative entropy in the context of continuous-time ergodic Markov Chains. We follow the methodology of the recent work by Karatzas, Schachermayer & Tschiderer (2019), which is based on stochastic calculus and uses time-reversal in a critical fashion. By "aggregating" the trajectorial results, i.e., by averaging them with respect to the invariant measure, we obtain a very crisp, geometric picture of the steepest descent property for the curve of time-marginals, relative to local perturbations. This is done in terms of an appropriate flat metric on configuration space, defined in terms of a suitable locally weighted Hilbert norm.

We adopt then a more global approach, and establish also the "gradient flow" property—to the effect that the time-evolution for the curve of time-marginals is prescribed by an appropriate *Riemannian metric* on

* We are indebted to Ioannis Kontoyiannis, Abishek Tilva and Lane Yeung for sharing their expertise on the subject, for bringing relevant literature to our attention, and for their many helpful comments. I.K. acknowledges support from the U.S. National Science Foundation under Grant NSF-DMS-20-04997. J.M. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 716117) and from the Austrian Science Fund (FWF) through project F65. W.S. acknowledges support from the Austrian Science Fund (FWF) under grant P28861 and by the Vienna Science and Technology Fund (WWTF) through projects MA14-008 and MA16-021.

† Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027, USA (email: *ik1@columbia.edu*).
‡ Institute of Science and Technology (IST) Austria, Am Campus 1, 3400 Klosterneuburg, Austria (email: *jan.maas@ist.ac.at*).
§ Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
(email: *walter.schachermayer@univie.ac.at*).

the manifold of probability measures on configuration space, and by the differential of the relative entropy functional along this curve; cf. MAAS (2011), MIELKE (2011), ERBAR & MAAS (2012, 2014). Both steepest descent and gradient flow are manifestations of the seminal JORDAN, KINDERLEHRER & OTTO (1998) results and of their outgrowth, the so-called "OTTO (2001) Calculus".

Preview: For a finite state-space, we set up the probabilistic framework in Section 2 and the functional-analytic one in Section 4. The appropriate stochastic-analytic machinery and results appear in Sections 3 and 5. Temporal dissipation and steepest descent are then developed in increasing generality: First in Section 6 for the variance and its associated, globally determined and flat, metric; then in Section 7 for the BOLTZMANN-GIBBS-SHANNON relative entropy; and finally in Section 8 for general entropies induced by convex functions. Gradient flows and their associated Riemannian geometries are taken up in Section 9. Extensions to state-spaces with a countable infinity of elements are developed in Section 10.

2 The Setting

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we consider an irreducible, positive recurrent, discrete-time MARKOV Chain $\mathcal{Z} = (Z_n)_{n \in \mathbb{N}_0}$ with state-space $\mathcal{S}$, transition probability matrix $\Pi = (\pi_{xy})_{(x,y) \in \mathcal{S}^2}$ with entries $\pi_{xy} = \mathbb{P}(Z_{n+1} = y \,|\, Z_n = x)$ for $n \in \mathbb{N}_0$, and initial distribution $P(0) = (p(0,x))_{x \in \mathcal{S}}$ which is a column vector with components $p(0,x) := \mathbb{P}(Z_0 = x) > 0$ for all $x \in \mathcal{S}$. Throughout Sections 2–9, it is assumed that the state-space $\mathcal{S}$ is *finite*. Extensions to countable state-spaces are taken up in Section 10.

It is straightforward to check that the sequence of random variables $\big(M_n^f\big)_{n \in \mathbb{N}_0}$ with $M_0^f := f(Z_0)$,

$$M_n^f \,:=\, f(Z_n) - \sum_{k=0}^{n-1} \big(\Pi f - f\big)(Z_k)\,, \qquad n \in \mathbb{N}\,, \tag{2.1}$$

is a martingale of the filtration generated by the MARKOV Chain $\mathcal{Z}$, for any given function $f : \mathcal{S} \to \mathbb{R}$. Here and in what follows, we denote $(\Pi f)(z) := \sum_{y \in \mathcal{S}} \pi_{zy}\, f(y),\ z \in \mathcal{S}$.

As is well known, such a Chain has a unique invariant distribution: that is, a column vector $Q = \big(q(y)\big)_{y \in \mathcal{S}}$ of *positive* numbers adding up to 1 and satisfying $\Pi^* Q = Q$ or, more explicitly,

$$q(y) \,=\, \sum_{z \in \mathcal{S}} q(z)\, \pi_{zy}\,, \qquad \forall\ y \in \mathcal{S}\,. \tag{2.2}$$

A major result of discrete-time MARKOV Chain theory states that, when $\mathcal{Z}$ is also aperiodic, the $k-$step transition probabilities

$$\pi_{xy}^{(0)} := \mathbf{1}_{x=y}\,, \qquad \pi_{xy}^{(k)} := \mathbb{P}\big(Z_k = y \,\big|\, Z_0 = x\big)\,, \quad k \in \mathbb{N} \tag{2.3}$$

converge as k tends to infinity to $q(y)$, for every pair of states $(x,y) \in \mathcal{S}^2$. We refer to Chapter 1 in NORRIS (1997), in particular Theorems 1.7.7 and 1.8.3, for an excellent account of the relevant theory.

2.1 From Discrete- to Continuous-Time MARKOV Chains, via POISSON

Consider now a POISSON process $\mathcal{N} = \big(N(t)\big)_{0 \le t < \infty}$ with parameter $\lambda = 1$ and *independent* of the discrete-time MARKOV Chain $\mathcal{Z}$. We construct via time-change the continuous-time process

$$X(t) \,:=\, Z_{N(t)}\,, \qquad 0 \le t < \infty\,, \tag{2.4}$$

as well as the filtration $\mathbb{F}^X = \big\{\mathcal{F}^X(t)\big\}_{0 \le t < \infty}$ this process generates via $\mathcal{F}^X(t) := \sigma\big(X(s),\, 0 \le s \le t\big)$.

Straightforward computation shows that this new, continuous-time process $\mathcal{X} = \big(X(t)\big)_{0\le t<\infty}$ has the MARKOV property, and time-homogeneous transition probabilities

$$\varrho_h(x,y) \,:=\, \mathbb{P}\big(X(t+h)=y \,\big|\, X(t)=x\big) = e^{-h} \sum_{k\in\mathbb{N}_0} \frac{h^k}{k!}\, \pi^{(k)}_{xy}\,, \qquad t\ge 0,\ h>0 \tag{2.5}$$

with the notation of (2.3); we set $\varrho_0(x,y) := \mathbf{1}_{x=y}$. The functions $h\mapsto \varrho_h(x,y)$ in (2.5) are uniformly continuous and continuously differentiable; cf. Theorems 2.13, 2.14 in LIGGETT (2010).

More generally, for arbitrary $n\in\mathbb{N}$, $0<\theta_1<\cdots<\theta_n=\theta<t<\infty$, $(x,y_1,\cdots,y_n,z)\in\mathcal{S}^{n+2}$ with $y=y_n$, the finite-dimensional distributions of this process are

$$\begin{aligned}\mathbb{P}\big(X(0)=x, X(\theta_1)=y_1,\cdots,X(\theta_n)=y_n, X(t)=z\big) \,=& \\ =\, p(0,x)\,\varrho_{\theta_1}(x,y_1)\,\varrho_{\theta_2-\theta_1}(y_1,y_2)\cdots\varrho_{\theta_n-\theta_{n-1}}(y_{n-1},y_n)\cdot\varrho_{t-\theta}(y,z)&\end{aligned} \tag{2.6}$$

and we deduce the time-homogeneous MARKOV property

$$\mathbb{P}\big(X(t)=z \,\big|\, \mathcal{F}^X(\theta)\big) = \varrho_{t-\theta}\big(X(\theta),z\big) = \mathbb{P}\big(X(t)=z \,\big|\, X(\theta)\big). \tag{2.7}$$

Finally, from the CHAPMAN-KOLMOGOROV equations $\pi^{(m+n)}_{xy} = \sum_{z\in\mathcal{S}} \pi^{(m)}_{xz}\pi^{(n)}_{zy}$ for the $k-$step transition probabilities of $\mathcal{Z}$ in (2.3), we deduce these same equations

$$\varrho_{t+\theta}(x,y) \,=\, \sum_{z\in\mathcal{S}} \varrho_\theta(x,z)\,\varrho_t(z,y)\,, \qquad (\theta,t)\in[0,\infty)^2,\ \ (x,y)\in\mathcal{S}^2 \tag{2.8}$$

for the quantities in (2.5), the transition probabilities of the continuous-time MARKOV Chain $\mathcal{X}$ in (2.4). Here we think of the temporal argument θ as the “backward variable”, and of t as the “forward variable”.

2.2 Infinitesimal Generators and Martingales

We introduce now the matrix

$$\mathcal{K} := \Pi - \mathrm{I} \,=\, \big\{\kappa(x,y)\big\}_{(x,y)\in\mathcal{S}^2} \qquad \text{with elements} \qquad \kappa(x,y) := \pi_{xy} - \mathbf{1}_{x=y}\ : \tag{2.9}$$

non-negative off the diagonal, adding up to zero across each row. From (2.5) and with the help of time-homogeneity, we obtain for $t\ge 0$, $h>0$ the infinitesimals

$$\mathbb{P}\big(X(t+h)=y \,\big|\, X(t)=x\big) \,=\, h\cdot\kappa(x,y)+o(h)\,, \qquad x\ne y\,, \tag{2.10}$$

$$\mathbb{P}\big(X(t+h)=x \,\big|\, X(t)=x\big) \,=\, 1+h\cdot\kappa(x,x)+o(h) \tag{2.11}$$

with the standard convention $\lim_{h\downarrow 0}\big(o(h)/h\big) = 0$, valid uniformly over $t\in[0,\infty)$. In particular, (2.10) and (2.11) give the infinitesimals $\varrho_h(x,y)-\varrho_0(x,y) = h\cdot\kappa(x,y)+o(h)$ for all $(x,y)\in\mathcal{S}^2$, and thus

$$\partial\varrho_h(x,y)\,\big|_{h=0} \,=\, \kappa(x,y)\,. \tag{2.12}$$

Here and throughout this paper, ∂g denotes partial differentiation of a function g with respect to its temporal argument.

A bit more generally, for any $f:\mathcal{S}\to\mathbb{R}$ we have from (2.10), (2.11) the semigroup computation

$$\big(T_h f\big)(x) \,:=\, \mathbb{E}\big[f\big(X(t+h)\big) \,\big|\, X(t)=x\big] \,=\, f(x)+h\cdot\big(\mathcal{K}f\big)(x)+o(h)\,. \tag{2.13}$$

We deploy, here and in what follows, the *infinitesimal generator* of the Chain, i.e., the linear operator

$$\big(\mathcal{K}f\big)(x) := \big(\Pi f\big)(x) - f(x) = \sum_{y\in\mathcal{S}} \kappa(x,y)\,f(y) = \sum_{y\in\mathcal{S}} \kappa(x,y)\,\big[f(y)-f(x)\big]\,, \quad x\in\mathcal{S}\,. \tag{2.14}$$

Using the computation (2.13), it is shown fairly easily that the exact analogue of the random sequence (2.1) in our present setting, namely, the process

$$f\big(X(t)\big) - \int_0^t \big(\mathcal{K} f\big)\big(X(\theta)\big)\,\mathrm{d}\theta\,, \qquad 0 \le t < \infty\,, \tag{2.15}$$

is an $\mathbb{F}^X-$martingale; cf. Theorem 3.32 in LIGGETT (2010). As a slight generalization, we obtain also the following result (Lemma IV.20.12 in ROGERS & WILLIAMS (1987)).

Proposition 2.1. *Given any function* $g\,:\,[0,\infty)\times\mathcal{S}\to\mathbb{R}$ *whose temporal derivative* $t\mapsto\partial g(t,x)$ *is continuous for every state* $x\in\mathcal{S}$, *the process below is a local* $\mathbb{F}^X-$*martingale:*

$$M^g(t) \,:=\, g\big(t, X(t)\big) - \int_0^t \big(\partial g + \mathcal{K} g\big)\big(\theta, X(\theta)\big)\,\mathrm{d}\theta\,, \qquad 0 \le t < \infty\,. \tag{2.16}$$

Remark 2.1. *The General Case:* Instead of starting with the transition probabilities π_{xy} and defining $\kappa(x,y) = \pi_{xy} - \mathbf{1}_{x=y}$ as in (2.9), one can work instead with *any transition rates* $\kappa(x,y)$ satisfying: (i) $\kappa(x,y)\ge 0$ for $x\neq y$; and (i) $\sum_{y\in\mathcal{S}}\kappa(x,y)=0$ for every $x\in\mathcal{S}$. In this manner, arbitrary irreducible continuous-time MARKOV chains on finite state spaces can be studied with no extra effort.

We have opted here for the somewhat less general, but also very concrete and intuitive, approach of the present Section.

3 Forward and Backward KOLMOGOROV Equations

Let us differentiate both sides of the equations in (2.8) with respect to the backward variable θ, then set $\theta = 0$. We obtain on account of (2.12) the *Backward* KOLMOGOROV *differential equations*

$$\partial\varrho_t(x,y) \,=\, \sum_{z\in\mathcal{S}} \kappa(x,z)\,\varrho_t(z,y). \tag{3.1}$$

We can write this system of equations, for the matrix-valued function $t\mapsto\mathcal{P}_t = \big(\varrho_t(x,y)\big)_{(x,y)\in\mathcal{S}^2}$ of the forward variable $t\in[0,\infty)$, in the form $\partial\,\mathcal{P}_t \,=\, \mathcal{K}\,\mathcal{P}_t\,,\ \mathcal{P}_0 = \mathrm{I}\,.$

In a similar manner, differentiating formally the equations (2.8) with respect to the forward variable t, then evaluating at $t = 0$ and introducing the *transpose*

$$\mathcal{K}' \,:=\, \big(\kappa'(y,z)\big)_{(y,z)\in\mathcal{S}^2}\,, \qquad \kappa'(y,z) \,:=\, \kappa(z,y) \tag{3.2}$$

of the $\mathcal{K}-$matrix, we obtain the *Forward* KOLMOGOROV *equations*

$$\partial\varrho_\theta(x,y) \,=\, \sum_{z\in\mathcal{S}} \varrho_\theta(x,z)\,\kappa(z,y) \,=\, \sum_{z\in\mathcal{S}} \kappa'(y,z)\,\varrho_\theta(x,z)\,, \qquad \text{or} \qquad \partial\,\mathcal{P}_\theta \,=\, \mathcal{K}'\mathcal{P}_\theta\,, \quad \mathcal{P}_0 = \mathrm{I}. \tag{3.3}$$

Here and throughout this paper, prime $'$ denotes transposition of a matrix or vector.

3.1 A Curve of Probability Vectors

For every $t>0$, let us consider the column vector $P(t) = \big(p(t,y)\big)_{y\in\mathcal{S}}$ of probabilities for the $\mathbb{P}-$distribution

$$p(t,y) := \mathbb{P}\big(X(t)=y\big) = e^{-t}\sum_{x\in\mathcal{S}} p(0,x) \sum_{k\in\mathbb{N}_0} \frac{t^k}{k!}\,\pi_{xy}^{(k)} \,>\, 0 \tag{3.4}$$

of the random variable $X(t)$. The forward KOLMOGOROV equations of (3.3), the law of total probability, and the MARKOV property, show that these satisfy their own forward KOLMOGOROV equations, namely

$$\partial p(t,y) \,=\, \sum_{z\in\mathcal{S}} p(t,z)\,\kappa(z,y) \,=\, \sum_{z\in\mathcal{S}} \kappa'(y,z)\,p(t,z) \,=: \big(\mathcal{K}'p\big)(t,y)\,; \tag{3.5}$$

or, more compactly and in matrix form, $\partial P(t) = \mathcal{K}' P(t)\,,\ \ 0 \le t < \infty$ in the notation of (3.2).

We think of $(P(t))_{0\le t<\infty}$ *as a curve on the manifold* $\mathcal{M} = \mathcal{P}_+(\mathcal{S})\,$, *of vectors* $P = (p(x))_{x\in\mathcal{S}}$ *with strictly positive elements and total mass* $\sum_{x\in\mathcal{S}} p(x) = 1$, *thus viewed as probability measures.* In the present context, they are governed by (3.5).

Now suppose that the initial distribution $P(0)$ of the discrete-time MARKOV Chain $\mathcal{Z}$ coincides with the column vector $Q = \big(q(y)\big)_{y\in\mathcal{S}}$ of (2.2) satisfying $\Pi' Q = Q\,$, or equivalently $\mathcal{K}' Q = 0$ on account of (2.9). It follows that $P(t) \equiv Q\,,\ \forall\, t \in [0,\infty)$ provides then the solution of (3.5): the distribution Q is invariant also for the continuous-time MARKOV Chain $\mathcal{X}$ in (2.4).

A bit more generally, Q is the *equilibrium distribution* of $\mathcal{X}$, in the sense that for every initial distribution $P(0) = \big(p(0,x)\big)_{x\in\mathcal{S}}$ and function $f : \mathcal{S} \to \mathbb{R}$ we have

$$\lim_{t\to\infty} p(t,y) \,=\, q(y)\,, \qquad \forall\ y \in \mathcal{S}, \tag{3.6}$$

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f\big(X(t)\big)\,\mathrm{d}t \,=\, \sum_{y\in\mathcal{S}} q(y)\,f(y)\,, \quad \mathbb{P}-\text{a.e.}; \tag{3.7}$$

see Sections 3.6–3.8 in NORRIS (1997) for an account of these results. In the present, continuous-time context, aperiodicity plays no role.

3.2 A Curve of Likelihood Ratios

Let us compare now the components of the probability vector $P(t)$ in (3.4), with those of the invariant probability vector Q in (2.2). One way to do this, very fruitful in the present context, is by considering the likelihood ratio column vector

$$\boldsymbol{\ell}_t \equiv \boldsymbol{\ell}(t) = \big(\ell(t,y)\big)_{y\in\mathcal{S}} \qquad \text{with components} \qquad \ell(t,y) \,:=\, \frac{p(t,y)}{q(y)}\,. \tag{3.8}$$

Plugging the product $p(t,y) = \ell(t,y)\,q(y)$ into the forward KOLMOGOROV equation (3.5), we obtain for the likelihood ratios of (3.8) the *Backward Equation*

$$\partial \ell(t,y) \,=\, \sum_{z\in\mathcal{S}} \widehat{\kappa}(y,z)\,\ell(t,z) \,=\, \sum_{z\in\mathcal{S}} \widehat{\kappa}(y,z)\,\big[\ell(t,z) - \ell(t,y)\big] \,=: \big(\widehat{\mathcal{K}}\,\boldsymbol{\ell}\big)(t,y)\,, \tag{3.9}$$

or equivalently $\partial\boldsymbol{\ell}(t) = \widehat{\mathcal{K}}\,\boldsymbol{\ell}(t)$ in matrix form, with the new transition rates

$$\widehat{\mathcal{K}} \,:=\, \Big(\widehat{\kappa}(y,z)\Big)_{(y,z)\in\mathcal{S}^2}\,, \qquad \widehat{\kappa}(y,z) \,:=\, \frac{q(z)}{q(y)}\,\kappa(z,y)\,. \tag{3.10}$$

The entries of this matrix are non-negative off the diagonal, and add up to zero $\sum_{z\in\mathcal{S}} \widehat{\kappa}(y,z) = 0$ across every row $y \in \mathcal{S}$, on account of (2.2), (2.9).

We think once again of $(\boldsymbol{\ell}(t))_{0\le t<\infty}$ as a curve, now in the space $\mathcal{L} = \mathcal{L}_+(\mathcal{S})$ of vectors $\Lambda = (\lambda(x))_{x\in\mathcal{S}}$ with strictly positive elements and $\sum_{x\in\mathcal{S}} q(x)\,\lambda(x) = 1$, viewed as likelihood ratios with respect to the invariant distribution and evolving in time via (3.9).

Presently, we shall identify $\widehat{\mathcal{K}}$ of (3.10) with the infinitesimal generator of a suitable continuous-time MARKOV Chain, run *backwards* in time. A special case, however, is worth mentioning already.

Definition 3.1. Detailed Balance: The invariant distribution Q in (2.2) is said to satisfy the *detailed-balance* conditions, if

$$q(y)\,\kappa(y,z) \;=\; q(z)\,\kappa(z,y)\,, \quad \forall\,(y,z) \in \mathcal{S}^2. \tag{3.11}$$

This requirement turns out to be equivalent to the identity $q(y)\,\varrho_t(y,z) \;=\; q(z)\,\varrho_t(z,y)$ for all $t \in (0,\infty)$, $(y,z) \in \mathcal{S}^2$; one leg of the equivalence is immediate, courtesy of (2.12). When (3.11) prevails, $\widehat{\mathcal{K}} \equiv \mathcal{K}$ holds in (3.10); and the backward equation (3.9) for the likelihood ratios $(\ell_t(x))_{x\in\mathcal{S}}$ of (3.8), is then exactly the same as the backward equation (3.1) for $(\varrho_t(x,y))_{x\in\mathcal{S}}$. We stress that, whenever the detailed-balance conditions (3.11) are needed in the sequel, they will be invoked explicitly.

Remark 3.1. Clearly, (3.11) holds if, and only if, the operator $\mathcal{K}$ in (2.14) is self-adjoint on $\mathbb{L}^2(\mathcal{S},Q)$.

4 Discrete Gradient and Divergence; DIRICHLET Form, HILBERT Norms

It is apt at this point to introduce some necessary notation. For a given function $f : \mathcal{S} \to \mathbb{R}$ we consider the *discrete gradient* $\nabla f : \mathcal{S}^2 \to \mathbb{R}$ given by

$$\nabla f(x,y) := f(y) - f(x)\,. \tag{4.1}$$

In a similar spirit we consider

$$\big(\nabla \cdot F\big)(x) \;:=\; \frac{1}{2} \sum_{y\in\mathcal{S},\, y\neq x} \kappa(x,y)\,\big[F(x,y) - F(y,x)\big], \tag{4.2}$$

the *discrete divergence* of a function $F : \mathcal{S}\times\mathcal{S} \to \mathbb{R}$, and note the familiar *concatenation* formula

$$\mathcal{K}f = \nabla \cdot \big(\nabla f\big)\,. \tag{4.3}$$

This allows us to think of the operator $\mathcal{K}$ in (2.14) also as a "discrete Laplacian".

We introduce now the set $\mathcal{Z} := \{(x,y) \in \mathcal{S}\times\mathcal{S} \;:\; \kappa(x,y) > 0\}$ consisting of all edges in the incidence graph associated with the MARKOV chain, and the measure C on $\mathcal{Z}$ defined by the "conductances"

$$C\{(x,y)\} \equiv c(x,y) := \frac{1}{2}\,\kappa(x,y)\,q(x)\,, \qquad (x,y) \in \mathcal{Z}, \tag{4.4}$$

and consider the bilinear forms

$$\big\langle f,g\big\rangle_{L^2(\mathcal{S},Q)} \;:=\; \sum_{x\in\mathcal{S}} q(x)\,f(x)\,g(x)\,, \qquad \big\langle F,G\big\rangle_{L^2(\mathcal{Z},C)} \;:=\; \sum_{(x,y)\in\mathcal{Z}} c(x,y)\,F(x,y)\,G(x,y) \tag{4.5}$$

for real-valued functions defined on $\mathcal{S}$ (lowercase $f,\ g$) and on $\mathcal{S}\times\mathcal{S}$ (uppercase $F,\ G$), respectively. These induce the $\mathbb{L}^2-$norms $\big\|f\big\|_{L^2(\mathcal{S},Q)}$ (relative to the probability measure Q) and $\big\|F\big\|_{L^2(\mathcal{Z},C)}$ (relative to the unnormalized measure C on $\mathcal{Z}$ in (4.4)), via

$$\begin{aligned} \big\|f\big\|^2_{\mathbb{L}^2(\mathcal{S},Q)} \;&:= \big\langle f,f\big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = \sum_{x\in\mathcal{S}} q(x)\,f^2(x)\,, \\ \big\|F\big\|^2_{\mathbb{L}^2(\mathcal{Z},C)} \;&:= \big\langle F,F\big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} = \sum_{(x,y)\in\mathcal{Z}} c(x,y)\,F^2(x,y)\,. \end{aligned} \tag{4.6}$$

Finally, we introduce the bilinear DIRICHLET *form* associated with the MARKOV Chain:

$$\mathcal{E}(f,g) \;:=\; -\big\langle f,\mathcal{K}g\big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} \;=\; -\sum_{y\in\mathcal{S}} q(y)\,f(y)\,\big(\mathcal{K}g\big)(y) \;=\; -\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} q(y)\,\kappa(y,x)\,f(y)\,g(x). \tag{4.7}$$

This is not symmetric, in general; but satisfies $\mathcal{E}(f,f) \geq 0$, as follows from Lemma 4.1 below.

Lemma 4.1. *The* DIRICHLET *form (4.7) can be cast equivalently as*

$$\mathcal{E}(f,g) = \frac{1}{2} \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \kappa(y,x)\, q(y) \big(f(y) - g(x)\big)^2 . \tag{4.8}$$

Proof: We have clearly $\sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \kappa(y,x)\, q(y) f^2(y) = 0$ on account of $\sum_{x \in \mathcal{S}} \kappa(y,x) = 0$ for every $y \in \mathcal{S}$; as well as

$$\sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \kappa(y,x)\, q(y)\, g^2(x) \;=\; \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \widehat{\kappa}(x,y)\, q(x)\, g^2(x) \;=\; 0\,,$$

from the adjoint rates of (3.10) and their property $\sum_{y \in \mathcal{S}} \widehat{\kappa}(x,y) = 0$ for every $x \in \mathcal{S}$. It follows from (4.7) that

$$\sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \kappa(y,x)\, q(y) \big(f(y) - g(x)\big)^2 \;=\; -2 \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \kappa(y,x)\, q(y)\, f(y)\, g(x) \;=\; 2\,\mathcal{E}(f,g)\,. \qquad \square$$

4.1 Consequences of Detailed Balance

Under the detailed-balance conditions (3.11) we have, for functions $f : \mathcal{S} \to \mathbb{R}$ and $F : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ and in addition to the concatenation property (4.3), also the *integration-by-parts* formula

$$\big\langle \nabla f, F \big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} \;=\; -\big\langle f, \nabla \cdot F \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)}\,. \tag{4.9}$$

As a result, *the bilinear* DIRICHLET *form of (4.7) is now symmetric,* and induces the HILBERT $\mathbb{H}^1-$*inner product and norm*

$$\big\langle f, g \big\rangle_{\mathbb{H}^1(\mathcal{S},Q)} \;:=\; \mathcal{E}(f,g) \;=\; \Big\langle \nabla f, \nabla g \Big\rangle_{\mathbb{L}^2(\mathcal{Z},C)}\,, \tag{4.10}$$

$$\big\| f \big\|^2_{\mathbb{H}^1(\mathcal{S},Q)} \;:=\; \mathcal{E}(f,f) \;=\; \sum_{(x,y) \in \mathcal{Z}} c(x,y) \big(f(y) - f(x)\big)^2 \;=\; -\big\langle f, \mathcal{K} f \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} \;=\; \big\| \nabla f \big\|^2_{\mathbb{L}^2(\mathcal{Z},C)}\,, \tag{4.11}$$

respectively. We introduce also the dual of this norm, the HILBERT $\mathbb{H}^{-1}-$*norm* $\big|\big| f \big|\big|_{\mathbb{H}^{-1}(\mathcal{S},Q)}$, via

$$\big\| f \big\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \;:=\; \big\| \nabla \big(\mathcal{K}^{-1} f\big) \big\|_{\mathbb{L}^2(\mathcal{Z},C)}\,, \quad \text{if } f \in \text{Range}(\mathcal{K})\,; \qquad \big\| f \big\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \;:=\; +\infty\,, \quad \text{otherwise} \tag{4.12}$$

and note the variational characterizations

$$\big\| f \big\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \;=\; \sup_{g : \mathcal{S} \to \mathbb{R}} \frac{\big\langle f, g \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)}}{\big\| g \big\|_{\mathbb{H}^1(\mathcal{S},Q)}}\,, \tag{4.13}$$

$$\big\| f \big\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \;=\; \inf \big\{ \big\| F \big\|_{\mathbb{L}^2(\mathcal{Z},C)} : f = \nabla \cdot F \big\} \;=\; \inf \big\{ \big\| \nabla g \big\|_{\mathbb{L}^2(\mathcal{Z},C)} : f = \mathcal{K} g \big\}\,. \tag{4.14}$$

HILBERT space theory shows that the infima are attained.

Lemma 4.2. *Under the conditions of (3.11), the expression (4.8) takes the form*

$$\mathcal{E}(f,g) = \frac{1}{2} \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \kappa(y,x)\, q(y) \big[f(y) - f(x)\big] \big[g(y) - g(x)\big] \;=\; \Big\langle \nabla f, \nabla g \Big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} = \big\langle f, g \big\rangle_{\mathbb{H}^1(\mathcal{S},Q)}\,. \tag{4.15}$$

Proof: Let us write the double summation in the above display as

$$\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} \kappa(y,x)q(y)\Big[\, f(y)\,g(y) - f(y)\,g(x) - f(x)\,g(y) + f(x)\,g(x)\Big] =$$

$$= -\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} \kappa(y,x)q(y)\Big[\, f(y)\,g(x) + f(x)\,g(y)\Big] \;=\; -2\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} \kappa(y,x)q(y)\, f(y)\,g(x) \;=\; -2\,\mathcal{E}(f,g)\,.$$

Here, the first equality uses (3.10), as well as the properties $\sum_{x\in\mathcal{S}} \kappa(y,x) = 0$ for every $y \in \mathcal{S}$, and $\sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y) = 0$ for every $x \in \mathcal{S}$; whereas, the second equality uses the detailed balance conditions (3.11), and the third equality is just (4.7).

This proves the first equality in (4.15). The second and third are just restatements of (4.10). □

Remark 4.1. *An Additional Consequence:* It follows from (4.9)–(4.11) that, under the detailed balance conditions (3.11), the mapping

$$\nabla \,:\, \mathbb{H}^1(\mathcal{S},Q) \to \mathbb{L}^2(\mathcal{Z},C)$$

is an isometric embedding. Whereas, the discrete divergence mapping $\nabla\cdot$ in (4.2) is, up to a minus sign, the adjoint of the map $\nabla \,:\, \mathbb{L}^2(\mathcal{S},Q) \to \mathbb{L}^2(\mathcal{Z},C)$.

Remark 4.2. *A Counterexample.* In the absence of detailed balance, the DIRICHLET form $\mathcal{E}(f,g)$ is not an inner product; indeed, Remark 3.1 shows that there exist functions $f:\mathcal{S}\to\mathbb{R}\,,\; g:\mathcal{S}\to\mathbb{R}$ with $\mathcal{E}(f,g) = -\big\langle f, \mathcal{K}g\big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} \neq -\big\langle g, \mathcal{K}f\big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = \mathcal{E}(g,f)$. An explicit example of this situation is provided by the matrix

$$\mathcal{K} \;=\; \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix},$$

whose invariant distribution $Q = (1/3, 1/3, 1/3)$ is uniform on the state space $\mathcal{S} = \{1,2,3\}$ and for which detailed balance fails. Whereas, with $f = \mathfrak{e}_1 = (1,0,0)$ and $g = \mathfrak{e}_2 = (0,1,0)$ the first and second unit row vectors, respectively, and noting $3\,\mathcal{E}(\varphi,\gamma) = \varphi\,\mathcal{K}'\gamma'$ from (4.7), we observe

$$3\,\mathcal{E}(f,g) = \big(1,0,0\big)\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = -1\,, \qquad 3\,\mathcal{E}(g,f) = \big(0,1,0\big)\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 0\,.$$

Nevertheless, $\big\|f\big\|_{\mathbb{H}^1(\mathcal{S},Q)} = \sqrt{\mathcal{E}(f,f)}$ is still a HILBERT norm, whose associated inner product is given by the DIRICHLET form $\mathcal{E}_{\text{sym}}(f,g)$ of the reversible MARKOV Chain with symmetrized rates $\kappa_{\text{sym}}(x,y) := \big(\kappa(x,y) + \widehat{\kappa}(x,y)\big)/2$ in the manner of (4.7), (3.10); namely, $\mathcal{E}_{\text{sym}}(f,f) \equiv \mathcal{E}(f,f)$ and

$$\big\langle f,g\big\rangle_{\mathbb{H}^1(\mathcal{S},Q)} = -\frac{1}{2}\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} \big[q(y)\kappa(y,x) + q(x)\kappa(x,y)\big]\, f(x)g(y) = -\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} q(y)\,\kappa_{\text{sym}}(y,x)\, f(y)\,g(x)\,.$$

5 Time Reversal and Associated Martingales

It is well known that the MARKOV property is invariant under reversal of time (interchanging the roles of "past" and "future", keeping the "present" as is). This means, in particular, that the time-reversed process

$$\widehat{X}(s) := X(T-s)\,, \quad 0 \le s \le T \tag{5.1}$$

is a MARKOV Chain, for any given $T \in (0,\infty)$. *But how about the transition probabilities of this time-reversed process?* These are fairly easy to compute: namely,

$$\mathbb{P}\big(\widehat{X}(s_2) = z \,\big|\, \widehat{\mathcal{G}}(s_1)\big) = \mathbb{P}\big(\widehat{X}(s_2) = z \,\big|\, \widehat{X}(s_1)\big) = \rho^*\big(s_1, \widehat{X}(s_1); s_2, z\big) \tag{5.2}$$

for $0 \le s_1 \le s_2 \le T,\ z \in \mathcal{S}$, where

$$\rho^*\big(s_1, y; s_2, z\big) \,:=\, \frac{p(T-s_2,z)}{p(T-s_1,y)}\; \varrho_{s_2-s_1}\big(z,y\big)\,; \tag{5.3}$$

but need not be time-homogeneous in general.

However: *Let us compute these same transition probabilities when the Chain starts at its invariant distribution* Q. We introduce at this point another probability measure $\mathbb{Q}$ on the underlying measurable space $(\Omega, \mathcal{F})$, under which the MARKOV Chain $\mathcal{X}$ has exactly the same dynamics as before, but its initial distribution is the invariant probability vector $Q = \big(q(y)\big)_{y\in\mathcal{S}}$ in (2.2). Then, in lieu of (2.6), the finite-dimensional distributions of the Chain are

$$\mathbb{Q}\big(X(0)=x, X(\theta_1)=y_1, \cdots, X(\theta_n)=y_n, X(t)=z\big) \,=$$
$$=\, q(x)\,\varrho_{\theta_1}(x,y_1)\,\varrho_{\theta_2-\theta_1}(y_1,y_2)\cdots\varrho_{\theta_n-\theta_{n-1}}(y_{n-1},y_n)\cdot\varrho_{t-\theta}(y,z)\,.$$

On each $\sigma-$algebra $\mathcal{F}^X(t),\ 0 \le t < \infty$, the two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent; in fact, on the smaller $\sigma-$algebra $\sigma(X(t))$, we single out in the notation of (3.8) the so-called *likelihood process*

$$L(t) \,:=\, \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\bigg|_{\sigma(X(t))} \,=\, \ell\big(t, X(t)\big)\,, \qquad 0 \le t < \infty\,. \tag{5.4}$$

Under this dispensation, the transition probabilities are

$$\mathbb{Q}\big(\widehat{X}(s_2)=z\,\big|\,\widehat{\mathcal{G}}(s_1)\big) \,=\, \mathbb{Q}\big(\widehat{X}(s_2)=z\,\big|\,\widehat{X}(s_1)\big) \,=\, \widehat{\varrho}_{s_2-s_1}\big(\widehat{X}(s_1), z\big), \tag{5.5}$$

i.e., *time-homogeneous,* with

$$\widehat{\varrho}_h(y,z) \,:=\, \frac{q(z)}{q(y)}\; \varrho_h\big(z,y\big)\,. \tag{5.6}$$

Invoking (5.6) and (2.12), we see that the $\mathbb{Q}-$infinitesimal-generator of this time-reversed MARKOV Chain $\widehat{X}(s) = X(T-s),\ 0 \le s \le T$ in (5.1), is given *precisely by* $\widehat{\mathcal{K}} = \big(\widehat{\kappa}(y,z)\big)_{(y,z)\in\mathcal{S}^2}$ *as in (3.10).*
(We note parenthetically that, when the "detailed-balance" condition (3.11) holds, the initial distributions and transition probabilities of the continuous-time MARKOV Chain $X(t),\ 0 \le t \le T$, and of its time-reversal (5.1), are *exactly the same* under the probability measure $\mathbb{Q}$.)

As a consequence of these considerations, and by complete analogy with Proposition 2.1, we formulate the following result – then use it to provide an elementary proof for Proposition 5.2 in our context.

Proposition 5.1. *For any given function* $g : [0,T] \times \mathcal{S} \to \mathbb{R}$ *whose temporal derivative* $s \mapsto \partial g(s,x)$ *is continuous for every state* $x \in \mathcal{S},$ *the process right below is a* $\big(\widehat{\mathbb{G}}, \mathbb{Q}\big)-$ *local martingale:*

$$\widehat{M}^{\,g}(s) \,:=\, g\big(s, \widehat{X}(s)\big) - \int_0^s \big(\partial g + \widehat{\mathcal{K}} g\big)\big(u, \widehat{X}(u)\big)\,\mathrm{d}u\,, \qquad 0 \le s \le T\,. \tag{5.7}$$

The following important result is due to FONTBONA & JOURDAIN (2016) in the context of diffusions. Its proof (cf. Theorem 4.2 in KARATZAS, SCHACHERMAYER & TSCHIDERER (2019)) uses only the MARKOV property and the definition of conditional expectation, and carries over verbatim to our present context. An alternative argument, specific to the MARKOV Chain context, uses Proposition 5.1 and is given right below.

Proposition 5.2. Time-Reversed Likelihood Process as Martingale: *Fix* $T \in (0,\infty)$ *and consider the time-reversed Chain (5.1), as well as the filtration* $\widehat{\mathbb{G}} = \big\{\widehat{\mathcal{G}}(s)\big\}_{0\le s\le T}$ *this process generates via* $\widehat{\mathcal{G}}(s) := \sigma\big(\widehat{X}(u),\ 0 \le u \le s\big)$*. Then, the time-reversed likelihood process*

$$L(T-s) \,=\, \ell\big(T-s, \widehat{X}(s)\big)\,, \quad 0 \le s \le T \qquad \textit{is a } \big(\widehat{\mathbb{G}}, \mathbb{Q}\big) - \textit{martingale.} \tag{5.8}$$

Proof: We consider in (5.7) the function $g(s,x) = \ell(T-s,x)\,,\ 0 \le s \le T,\ \ x \in \mathcal{S}$ and note that $\partial g(s,x) = -\partial \ell(T-s,x) = -\big(\widehat{\mathcal{K}}\,\ell\big)(T-s,x)$ holds on account of (3.9). It follows from (5.7) of Proposition 5.1, whose integrand now vanishes, that the time-reversed likelihood ratio process $\ell\big(T-s,\widehat{X}(s)\big)\,,\ 0 \le s \le T$ is a $\mathbb{Q}-$local-martingale of the time-reversed filtration $\widehat{\mathbb{G}}$. But this process is positive, thus also a $\mathbb{Q}-$supermartingale, and its expectation

$$\mathbb{E}^{\mathbb{Q}}\big[\ell\big(T-s,X(T-s)\big)\big] \,=\, \sum_{y\in\mathcal{S}} q(y)\,\frac{p(T-s,y)}{q(y)} \,=\, 1\,, \qquad 0 \le s \le T$$

is constant. Therefore $\ell\big(T-s,\widehat{X}(s)\big)\,,\ 0 \le s \le T$ is a true $\mathbb{Q}-$martingale, exactly as stated in (5.8). $\Box$

6 The Variance Process

For a probability vector $P = (p(y))_{y\in\mathcal{S}}$ with positive components, we introduce its likelihood vector $\boldsymbol{\ell} = (\ell(y))_{y\in\mathcal{S}} \in \mathcal{L}$ with $\ell(y) = p(y)/q(y)$ as in (3.8), relative to the invariant distribution Q of the Chain. We define then in the manner of (4.6) the *Variance* of P relative to Q, also known as χ^2-*divergence,* as

$$V\big(P\,|\,Q\big) \,\equiv\, \mathrm{Var}^{\mathbb{Q}}\big(\boldsymbol{\ell}\big) \,:=\, \sum_{y\in\mathcal{S}} q(y)\,\ell^2(y) - 1 \,=\, \big|\boldsymbol{\ell}\big|^2_{\mathbb{L}^2(\mathcal{S},Q)} - 1\,. \tag{6.1}$$

Let us recall now from (3.4) the curve $\big(P(t)\big)_{0\le t<\infty}$ of time-marginal distributions for our continuous-time MARKOV Chain, and the corresponding curve of likelihoods $\boldsymbol{\ell}_t = (\ell(t,y))_{y\in\mathcal{S}}\,,\ 0 \le t < \infty$ in the space $\mathcal{L}$, with $\ell(t,y) = p(t,y)/q(y)$. We will show in Proposition 6.2 that the variance just defined in (6.1) plays the role of a LYAPUNOV function for this curve.

To see this, we summon the likelihood process $L(t) = \ell(t,X(t)),\ 0 \le t < \infty$ from (5.4) and consider its square $L^2(t),\ 0 \le t < \infty$, the so-called *Variance Process.*

Proposition 6.1. *For any given* $T \in (0,\infty)$, *we have the* DOOB-MEYER *decomposition*

$$\ell^2\big(T-s,\widehat{X}(s)\big) \,=\, \widehat{M}(s) + \int_0^s \sum_{y\neq x} \bigg(\widehat{\kappa}(x,y)\,\Big(\ell(t,y)-\ell(t,x)\Big)^2\bigg)\bigg|_{\substack{t=T-u\\ x=\widehat{X}(u)}} \mathrm{d}u\,, \qquad 0 \le s \le T \tag{6.2}$$

of the time-reversed variance process $\ell^2\big(T-s,\widehat{X}(s)\big),\ 0 \le s \le T$, *where* $\widehat{M}$ *is a* $\big(\widehat{\mathbb{G}},\mathbb{Q}\big)-$*martingale.*

Proof: The first claim follows from Proposition 5.2 and the JENSEN inequality. For the second claim we deploy Proposition 5.1 with $g(s,x) := \ell^2(T-s,x)\,,\ 0 \le s \le T,\ \ x \in \mathcal{S}$, to conclude via the calculation

$$\big(\partial g + \widehat{\mathcal{K}} g\big)(T-s,x) \,=\, \sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y)\Big(\ell(T-s,y) - \ell(T-s,x)\Big)^2 \tag{6.3}$$

that $\widehat{M}$ is a local $\big(\widehat{\mathbb{G}},\mathbb{Q}\big)-$martingale. The uniform continuity of $[0,T] \ni t \mapsto p_t(x,y) \in (0,1)$, and the finiteness of the state-space, imply that this process is actually bounded, thus a true $\mathbb{Q}-$martingale.

Let us now justify the claim (6.3). From the Backwards Equation (3.9), we have

$$\partial g(T-s,x) \,=-2\ell(T-s,x)\,\partial\ell(T-s,x) = -2\ell(T-s,x)\sum_{y\in\mathcal{S}}\widehat{\kappa}(x,y)\ell(T-s,y), \tag{6.4}$$

$$\big(\widehat{\mathcal{K}} g\big)(T-s,x) \,=\, \sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y)\,\ell^2(T-s,y) \,=\, \sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y)\,\Big[\ell^2(T-s,y) + \ell^2(T-s,x)\Big]$$

on account of the property $\sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y) = 0$ for every $x \in \mathcal{S}$; now (6.3) follows readily. $\Box$

Proposition 6.1 deals with the trajectorial behavior of the variance process; and for this, it is crucial to let time run backwards. Now, we want to adopt also an "aggregate" point of view, and take $\mathbb{Q}-$expectations in (6.2). When doing this, it does not matter any more whether time runs forwards or backwards, so we state the following result "forwards in time". Recalling (3.10), we obtain thus the dissipation of the variance along the curve of time-marginals for the Chain, and measure the exact rate of this dissipation.

Proposition 6.2. *Along the curve $\big(P(t)\big)_{0\le t<\infty}$ of time-marginal distributions in (3.4), the variance*

$$V\big(P(t)\,|\,Q\big) = \mathrm{Var}^{\mathbb{Q}}\big(\boldsymbol{\ell}_t\big) = \sum_{y\in\mathcal{S}} q(y)\,\ell^2(t,y) - 1 \;=\; \big|\boldsymbol{\ell}_t\big|^2_{\mathbb{L}^2(\mathcal{S},Q)} - 1\,, \quad 0\le t<\infty \tag{6.5}$$

is decreasing with $\lim_{t\to\infty}\downarrow V\big(P(t)\,|\,Q\big)=0$, *and the rate of its decrease is given by*

$$\partial\,\big|\boldsymbol{\ell}_t\big|^2_{\mathbb{L}^2(\mathcal{S},Q)} \;=\; \partial\, V\big(P(t)\,|\,Q\big) \;=\; -2\,\mathcal{E}\big(\boldsymbol{\ell}_t\,,\boldsymbol{\ell}_t\,\big) \tag{6.6}$$

(thus by $-2\,\big\|\boldsymbol{\ell}_t\big\|^2_{\mathbb{H}^1(\mathcal{S},Q)}$ *under the retailed balance conditions (3.11)). More precisely,*

$$V\big(P(T)\,|\,Q\big) \;=\; V\big(P(0)\,|\,Q\big) - \int_0^T \sum_{(x,y)\in\mathcal{Z}} q(y)\,\kappa(y,x)\,\Big(\ell(t,y)-\ell(t,x)\Big)^2\,\mathrm{d}t \tag{6.7}$$

$$= \int_T^\infty \sum_{(x,y)\in\mathcal{Z}} q(y)\,\kappa(y,x)\,\Big(\ell(t,y)-\ell(t,x)\Big)^2\,\mathrm{d}t\,. \tag{6.8}$$

The decomposition (6.2) is a trajectorial version of this variance dissipation, at the level of the individual particle viewed under the probability measure $\mathbb{Q}$ and under time-reversal. As a consequence of (6.2) and of the BAYES rule, we deduce from (6.2) the DOOB-MEYER decomposition

$$\ell\big(T-s,\widehat{X}(s)\big) \;=\; \widehat{N}(s) + \int_0^s \sum_{y\neq x}\left(\frac{\widehat{\kappa}(x,y)}{\ell\big(t,x\big)}\Big(\ell(t,y)-\ell(t,x)\Big)^2\right)\Bigg|_{\substack{t=T-u\\ x=\widehat{X}(u)}}\mathrm{d}u\,, \qquad 0\le s\le T \tag{6.9}$$

of the time-reversed likelihood process, where $\widehat{N}$ is a $\big(\widehat{\mathbb{G}},\mathbb{P}\big)-$martingale.

6.1 Steepest Descent of the Variance, under Detailed Balance

We shall establish now the following theorem. As pointed out in JORDAN, KINDERLEHRER & OTTO (1998), results of this type go as far back as the paper by COURANT, FRIEDRICHS & LEWY (1928) in the Brownian motion context. We deploy the notation of (3.8) for the likelihood ratios relative to the invariant distribution, as well as the following notion.

Definition 6.1. We say that a smooth curve of probability measures $(P(t))_{t_0\le t<\infty}\subset\mathcal{M}=\mathcal{P}_+(\mathcal{S})$ is a curve of *steepest descent* locally at $t=t_0$, for a given smooth functional $F:\mathcal{M}\to\mathbb{R}$ and relative to a given metric ϱ on $\mathcal{M}$, if it minimizes the infinitesimal rate of change of F as measured on $\mathcal{M}$ in terms of ϱ, i.e.,

$$\lim_{h\downarrow 0}\frac{F\big(\widetilde{P}(t_0+h)\big)-F\big(P(t_0)\big)}{\varrho\big(\widetilde{P}(t_0+h),P(t_0)\big)}\,,$$

among all curves $\big(\widetilde{P}(t)\big)_{t_0\le t<\infty}\subset\mathcal{M}$ satisfying $\widetilde{P}(t_0)=P(t_0)$.

Theorem 6.3. Steepest Descent for the Variance: *Under the conditions (3.11) of detailed balance, the curve $(P(t))_{0\le t<\infty}$ of time-marginal distributions in (3.4) has the property of steepest decent for the variance of (6.5), with respect to the metric distance bequeathed by the norm of (4.12), i.e.,*

$$\varrho\big(P_1,P_2\big) \;:=\; \big\|\,\boldsymbol{\ell}_1-\boldsymbol{\ell}_2\,\big\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \qquad \textit{for}\;\; P_1=\boldsymbol{\ell}_1 Q \;\; \textit{and}\;\; P_2=\boldsymbol{\ell}_2\,Q\,. \tag{6.10}$$

The proof of this result needs Proposition 6.5 below. We pave the way towards it by formulating first a variational version of Propositions 6.1, 6.2. For this purpose we fix an arbitrary time-point $t_0 \in (0,\infty)$, and let $\psi(\cdot) = (\psi(t))_{t_0 \le t < t_0+\varepsilon}$ be a continuous curve of real-valued functions on the state-space $\mathcal{S}$. With these ingredients, we define a new curve $\ell^\psi(\cdot) = (\ell^\psi(t))_{t_0 \le t < t_0+\varepsilon}$ of such functions, for a suitable $\varepsilon > 0$, by specifying in the space $\mathcal{L} = \mathcal{L}_+(\mathcal{S})$ of subsection 3.2 the initial condition $\ell^\psi(t_0) = \boldsymbol{\ell}(t_0) \in \mathcal{L}$ and the dynamics $\partial \ell^\psi(t) = (\widehat{\mathcal{K}}\psi)(t)$ for $t \in [t_0, t_0+\varepsilon)$; in the manner of (3.9) and a bit more explicitly,

$$\partial \ell^\psi(t,x) \,=\, \sum_{y \in \mathcal{S}} \widehat{\kappa}(x,y)\, \psi(t,y)\,, \qquad x \in \mathcal{S}. \tag{6.11}$$

The curve $\ell^\psi(\cdot) = (\ell^\psi(t))_{t_0 \le t < t_0+\varepsilon}$, the "output" of the system (6.11) corresponding to the "input" $\psi(\cdot)$, is only defined on an interval $[t_0, t_0+\varepsilon)$ and lives in the space $\mathcal{L}$, since

$$\partial \sum_{x \in \mathcal{S}} q(x)\, \ell^\psi(t,x) \,=\, \sum_{x \in \mathcal{S}} q(x) \sum_{y \in \mathcal{S}} \widehat{\kappa}(x,y)\, \psi(t,y) \,=\, \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} q(y)\, \kappa(y,x)\, \psi(t,y) \,=\, 0$$

implies $\sum_{x \in \mathcal{S}} q(x)\, \ell^\psi(t,x) = \sum_{x \in \mathcal{S}} q(x)\, \boldsymbol{\ell}(t_0,x) = 1$ for all $t \in [t_0+\varepsilon)$. Thus, the recipe

$$p^\psi(t,x) := q(x)\, \ell^\psi(t,x), \qquad (t,x) \in [t_0, t_0+\varepsilon) \times \mathcal{S} \tag{6.12}$$

procures a curve $(P^\psi(t))_{0 \le t < t_0+\varepsilon}$ on the manifold $\mathcal{M} = \mathcal{P}_+(\mathcal{S})$ in subsection 3.1, of vectors $P = \big(p(x)\big)_{x \in \mathcal{S}}$ with strictly positive elements and total mass $\sum_{x \in \mathcal{S}} p(x) = 1$.

Conversely: By irreducibility, the "input curve" $\psi(\cdot)$ is determined by the "output curve" $\ell^\psi(\cdot)$ up to an additive time-dependent constant. In particular, every smooth curve $\ell^*(\cdot) = (\ell^*(t))_{t_0 \le t < t_0+\varepsilon}$ in $\mathcal{L}$ with $\ell^*(t_0) = \ell(t_0)$ is representable as $\ell^\psi(\cdot)$ for a suitable continuous $\psi(\cdot)$ as above. For instance, $\boldsymbol{\ell}(\cdot) \in \mathcal{L}$ of (3.8) is the "output" that corresponds in this manner to the "input" $\psi(\cdot) \equiv \boldsymbol{\ell}(\cdot)$ in (6.11), via (3.9).

We have the following generalization of Proposition 6.2, to which it reduces when $\psi(\cdot) \equiv \boldsymbol{\ell}(\cdot)$.

Proposition 6.4. *In the above context, we have for* $t \in [t_0+\varepsilon)$ *the properties*

$$\partial\, V\big(P^\psi(t)\,|\,Q\big) = \partial\, \mathbb{E}^{\mathbb{Q}}\Big[\big(\ell^\psi\big)^2\big(t, X(t)\big)\Big] = 2\,\langle \psi_t, \mathcal{K}\ell^\psi_t \rangle_{\mathbb{L}^2(\mathcal{S},Q)} = -2\,\mathcal{E}\big(\psi_t, \ell^\psi_t\big).$$

Whereas, under the detailed balance conditions (3.11), this expression becomes

$$\partial\, V\big(P^\psi(t)\,|\,Q\big) \,=\, -2\,\mathcal{E}\big(\ell^\psi_t, \psi_t\big) \,=\, -2\,\Big\langle \nabla \ell^\psi_t, \nabla \psi_t \Big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} \,=\, -2\,\Big\langle \ell^\psi_t, \psi_t \Big\rangle_{\mathbb{H}^1(\mathcal{S},Q)}\,.$$

Proof: A reasoning similar to that in Propositions 6.1 and 6.2, and carried out once again in the backwards direction of time, can be deployed by applying Proposition 5.1 to $g(s,x) := \big(\ell^\psi\big)^2(T-s,x)\,,\ \ 0 \le s \le T$, $x \in \mathcal{S}$ for arbitrary but fixed $T \in (0, t_0+\varepsilon)$. But here is a simpler argument :

$$\partial\, V\big(P^\psi(t)\,|\,Q\big) \,=\, \partial \big\| \ell^\psi_t \big\|^2_{\mathbb{L}^2(\mathcal{S},Q)} \,=\, 2\,\langle \ell^\psi_t, \widehat{\mathcal{K}}\psi_t \rangle_{\mathbb{L}^2(\mathcal{S},Q)} \,=\, 2\,\langle \psi_t, \mathcal{K}\ell^\psi_t \rangle_{\mathbb{L}^2(\mathcal{S},Q)} = -2\,\mathcal{E}\big(\psi_t, \ell^\psi_t\big)$$

on account of (4.6), (4.7) and (3.10). This proves Proposition 6.2 as well. □

We pass now to the computation of the "infinitesimal cost of moving the curve" $\big(\ell^\psi(t)\big)_{t_0 \le t < t_0+\varepsilon}$.

Proposition 6.5. *Under the conditions (3.11) of detailed balance, we have*

$$\lim_{h \downarrow 0} \frac{1}{h}\, \|\, \boldsymbol{\ell}_{t+h} - \boldsymbol{\ell}_t \,\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \,=\, \|\, \mathcal{K}\, \boldsymbol{\ell}_t \,\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \,=\, \|\, \boldsymbol{\ell}_t \,\|_{\mathbb{H}^1(\mathcal{S},Q)} \tag{6.13}$$

for every $t \in [t_0, t_0+\varepsilon)$; *and a bit more generally, in the notation just developed,*

$$\lim_{h \downarrow 0} \frac{1}{h}\, \|\, \ell^\psi_{t+h} - \ell^\psi_t \,\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \,=\, \|\, \mathcal{K}\, \psi_t \,\|_{\mathbb{H}^{-1}(\mathcal{S},Q)} \,=\, \|\, \psi_t \,\|_{\mathbb{H}^1(\mathcal{S},Q)}\,. \tag{6.14}$$

Proof: From (6.11), (3.11) it follows that

$$\lim_{h\downarrow 0} \frac{1}{h}\Big[\ell^{\psi}_{t+h}(x) - \ell^{\psi}_{t}(x)\Big] \;=\; \big(\mathcal{K}\,\psi_t\big)(x)$$

holds for every $x \in \mathcal{S}$, and the first equality in (6.14) is evident. For the second equality in (6.14), it suffices to observe that $\nabla\psi_t$ is the unique element $F \in \mathbb{L}^2(\mathcal{Z}, C)$ with the property $\nabla \cdot F = \mathcal{K}\,\psi_t$, and to note from Remark 4.1 the isometry $\big\|F\big\|_{\mathbb{L}^2(\mathcal{Z},C)} = \big\|\psi_t\big\|_{\mathbb{H}^1(\mathcal{S},Q)}$ from the space $\mathbb{L}^2(\mathcal{Z}, C)$ to $\mathbb{H}^1(Q)$.

Now, (6.13) is just a special case of (6.14) with $\psi(\cdot) \equiv \boldsymbol{\ell}(\cdot)$, as discussed above. □

6.2 The Proof of Theorem 6.3

We are ready now to tackle the proof of Theorem 6.3. Along *any* smooth curve of the form $(P^{\psi}(t))_{t_0 \le t < t_0+\varepsilon}$ created as in (6.11), (6.12) on the manifold of probability vectors $\mathcal{M} = \mathcal{P}_+(\mathcal{S})$ and with $\ell^{\psi}(t_0) = \boldsymbol{\ell}(t_0) \in \mathcal{L}$, we have from Propositions 6.4, 6.5 the respective rates for the variance and the metric distance

$$\lim_{h\downarrow 0} \frac{V\big(P^{\psi}(t_0+h)\,|\,Q\big) - V\big(P(t_0)\,|\,Q\big)}{h} \;=\; -2\,\Big\langle\, \boldsymbol{\ell}_{t_0}, \psi_{t_0} \Big\rangle_{\mathbb{H}^1(\mathcal{S},Q)}\,,$$

$$\lim_{h\downarrow 0} \frac{\varrho\big(P^{\psi}(t_0+h), P(t_0)\big)}{h} \;=\; \big\|\psi_{t_0}\big\|_{\mathbb{H}^1(\mathcal{S},Q)}\,.$$

Thus, the rate of change for the variance along the *perturbed curve* $(P^{\psi}(t))_{t_0 \le t < t_0+\varepsilon}$, when measured on the manifold $\mathcal{M}$ by the metric distance in (6.10), is

$$\lim_{h\downarrow 0} \frac{V\big(P^{\psi}(t_0+h)\,|\,Q\big) - V\big(P(t_0)\,|\,Q\big)}{\varrho\big(P^{\psi}(t_0+h), P(t_0)\big)} \;=\; -2\,\left\langle\, \boldsymbol{\ell}_{t_0}, \frac{\psi_{t_0}}{\big\|\psi_{t_0}\big\|_{\mathbb{H}^1(\mathcal{S},Q)}} \right\rangle_{\mathbb{H}^1(\mathcal{S},Q)}.$$

On the other hand, along the *original curve* $(P(t))_{0 \le t < \infty}$ of time-marginal distributions for the Chain (that is, with $\psi(\cdot) \equiv \boldsymbol{\ell}(\cdot)$ modulo an affine transformation, as noted above), the rate of variance dissipation measured in terms of the metric distance traveled on the manifold $\mathcal{M}$ is

$$\lim_{h\downarrow 0} \frac{V\big(P(t_0+h)\,|\,Q\big) - V\big(P(t_0)\,|\,Q\big)}{\varrho\big(P(t_0+h), P(t_0)\big)} \;=\; -2\,\big\|\,\boldsymbol{\ell}_{t_0}\big\|_{\mathbb{H}^1(\mathcal{S},Q)} \;<\; 0\,.$$

A simple comparison of the last two displays, via CAUCHY-SCHWARZ, gives the *steepest descent property of the variance with respect to the metric distance in (6.10),* i.e.,

$$\lim_{h\downarrow 0} \frac{V\big(P^{\psi}(t_0+h)\,|\,Q\big) - V\big(P(t_0)\,|\,Q\big)}{\varrho\big(P^{\psi}(t_0+h), P(t_0)\big)} - \lim_{h\downarrow 0} \frac{V\big(P(t_0+h)\,|\,Q\big) - V\big(P(t_0)\,|\,Q\big)}{\varrho\big(P(t_0+h), P(t_0)\big)}$$

$$=\; 2\left(\big\|\boldsymbol{\ell}_{t_0}\big\|_{\mathbb{H}^1(\mathcal{S},Q)} - \left\langle\, \boldsymbol{\ell}_{t_0}, \frac{\psi_{t_0}}{\big\|\psi_{t_0}\big\|_{\mathbb{H}^1(\mathcal{S},Q)}} \right\rangle_{\mathbb{H}^1(\mathcal{S},Q)} \right) \;\ge\; 0\,,$$

along the original curve $(P(t))_{0 \le t < \infty}$ *of time-marginals for the* MARKOV *Chain.* Equality holds here if, and only if, $c + \psi_{t_0}$ is a positive constant multiple of $\boldsymbol{\ell}_{t_0}$ for some $c \in \mathbb{R}$. □

We will revisit this theme in Sections 8 and 9.

7 The Relative Entropy Process

For an arbitrary probability vector $P = (p(x))_{x \in \mathcal{S}}$ with strictly positive elements, let us recall the definition of its *relative entropy*, or KULLBACK–LEIBLER *divergence*,

$$H(P \,|\, Q) \;:=\; \sum_{x \in \mathcal{S}} p(x) \log \Big(\frac{p(x)}{q(x)} \Big) \tag{7.1}$$

with respect to the invariant distribution $Q = (q(x))_{x \in \mathcal{S}}$ of (2.2). In terms of the likelihood function in (3.8), the relative entropy of the probability vector $P(t)$ in (3.4) with respect to Q, is

$$H\big(P(t) \,\big|\, Q\big) \;=\; \mathbb{E}^{\mathbb{P}} \Big[\log \ell\big(t, X(t)\big) \Big], \qquad 0 \le t < \infty \,, \tag{7.2}$$

the $\mathbb{P}-$expectation of the log-likelihood at time t. We shall see presently that this function

$$t \;\longmapsto\; H\big(P(t) \,\big|\, Q\big) \;\text{ is non-negative, and satisfies }\; \lim_{t \to \infty} \downarrow H\big(P(t) \,\big|\, Q\big) = 0\,. \tag{7.3}$$

In other words, the relative entropy functional of (7.1) is also a LYAPUNOV function for the curve $\big(P(t)\big)_{0 \le t < \infty}$ of time-marginal distributions for our continuous-time MARKOV Chain.

We shall compute in Section 7.2 the rate of temporal decrease for this function. Of course, all this is in accordance with general thermodynamic principles governing the approach to equilibrium in physical systems (e.g., Chapter 2 in COVER & THOMAS (1991) in the context of discrete-time MARKOV Chains).

Let us note also, that the relative entropy in (7.2) can be cast equivalently as the $\mathbb{Q}-$expectation

$$H\big(P(t) \,\big|\, Q\big) \;=\; \sum_{y \in \mathcal{S}} q(y)\, \frac{p(t,y)}{q(y)}\, \log \left(\frac{p(t,y)}{q(y)} \right) \;=\; \mathbb{E}^{\mathbb{Q}} \Big[\ell\big(t, X(t)\big) \log \ell\big(t, X(t)\big) \Big] \tag{7.4}$$

of the *relative entropy process* $\ell\big(t, X(t)\big) \cdot \log \ell\big(t, X(t)\big)\,,\; 0 \le t < \infty\,.$ This allows us to justify the first claim in (7.3), regarding non-negativity. Indeed, the convexity of the function $(0,\infty) \ni \ell \mapsto \Phi(\ell) := \ell \,\log \ell$ gives, on the strength of the JENSEN inequality,

$$H\big(P(t) \,\big|\, Q\big) \;=\; \mathbb{E}^{\mathbb{Q}} \big[\Phi\big(\ell\big(t, X(t)\big)\big) \big] \;\ge\; \Phi\Big(\mathbb{E}^{\mathbb{Q}} \big[\ell\big(t, X(t)\big) \big] \Big) \;=\; f(1) \;=\; 0\,. \tag{7.5}$$

(Alternatively, this follows from $H\big(P(t) \,\big|\, Q\big) = \mathbb{E}^{\mathbb{Q}}[\Psi\big(\ell(t, X(t))\big)]$, with $\Psi \ge 0$ as in (7.13) below.)

Proposition 7.1. *In the context of Proposition 5.2, the time-reversed relative entropy process*

$$\ell\big(T-s, \widehat{X}(s)\big) \cdot \log \ell\big(T-s, \widehat{X}(s)\big), \quad 0 \le s \le T \qquad \textit{is a } \big(\widehat{\mathbb{G}}, \mathbb{Q}\big) - \textit{submartingale;} \tag{7.6}$$

the properties in (7.3) hold; and the time-reversed log-likelihood process

$$\log \ell\big(T-s, \widehat{X}(s)\big), \quad 0 \le s \le T \qquad \textit{is a } \big(\widehat{\mathbb{G}}, \mathbb{P}\big) - \textit{submartingale.} \tag{7.7}$$

Proof: The first claim follows from (5.8) and the convexity of the function $\Phi(\ell) = \ell \,\log \ell$ appearing inside the expectation in (7.4), along with the JENSEN inequality. The $\mathbb{Q}-$expectation

$$H\big(P(T-s) \,|\, Q\big) \;=\; \mathbb{E}^{\mathbb{Q}} \big[\Phi\big(\ell\big(T-s, \widehat{X}(s)\big)\big) \big]\,, \qquad 0 \le s \le T \tag{7.8}$$

of the process in (7.6) is thus increasing. This is precisely the monotonicity in (7.3); the remaining claim

$$\lim_{t \to \infty} \downarrow H\big(P(t) \,\big|\, Q\big) = 0 \tag{7.9}$$

there, follows now from (3.6), (7.4), and the finiteness of $\mathcal{S}$. The claim of (7.7) is a consequence of (7.6), (5.4), and the familiar BAYES rule (Lemma 3.5.3 in KARATZAS & SHREVE (1988)). □

7.1 Trajectorial Relative Entropy Dissipation

We read now Proposition 5.1 with $\Phi(\ell) = \ell \, \log \ell$ and the function

$$h(s,x) \,=\, \Phi\big(\ell(T-s,x)\big)\,, \qquad 0 \leq s \leq T,\;\; x \in \mathcal{S}\,. \tag{7.10}$$

As argued in the discussion following Proposition 5.2, and Proposition 7.1, the "time-reversed relative entropy" $H\big(P(T-s)\,|\,Q\big) = \mathbb{E}^{\mathbb{Q}}\big[h\big(s,\widehat{X}(s)\big)\big],\;\; 0 \leq s \leq T$ is increasing; and

$$\widehat{M}^h(s) \,:=\, h\big(s,\widehat{X}(s)\big) - \int_0^s \big(\partial h + \widehat{\mathcal{K}} h\big)\big(u,\widehat{X}(u)\big)\,\mathrm{d}u\,, \qquad 0 \leq s \leq T \tag{7.11}$$

is a $\mathbb{Q}-$local-martingale of the time-reversed filtration $\widehat{\mathbb{G}}$. The integrand in (7.11) is straightforward to compute: from (3.9), (3.10), and with $t = T-s$ for notational convenience, we get

$$\partial h(s,x) = -\big(1+\log \ell(t,x)\big)\,\big(\widehat{\mathcal{K}}\,\ell\big)(t,x) = -\big(1+\log \ell(t,x)\big) \sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y)\,\ell(t,y), \quad \text{thus}$$

$$\big(\partial h + \widehat{\mathcal{K}} h\big)(s,x) \,=\, \sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y)\,\ell(t,y)\Big[\log\frac{\ell(t,y)}{\ell(t,x)} - 1\Big] \,=\, \ell(t,x) \sum_{\substack{y\in\mathcal{S}\\ y\neq x}} \widehat{\kappa}(x,y)\,\Psi\Big(\frac{\ell(t,y)}{\ell(t,x)}\Big) \geq 0\,. \tag{7.12}$$

Here the function

$$\Psi(r) \,:=\, r\log r - r + 1\,, \qquad r > 0 \tag{7.13}$$

is nonnegative, convex, and attains its minimum $\Psi(1) = 0$ at $r = 1$. We have used in the last equality of (7.12) the property $\sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y) = 0$ for every $x \in \mathcal{S}$.

Proposition 7.2. *The submartingales of (7.6), (7.7) admit the* DOOB-MEYER *decompositions of the form*

$$\ell\big(T-s,\widehat{X}(s)\big)\log\big(\ell\big(T-s,\widehat{X}(s)\big)\big) \,=\, \widehat{M}^{\,h}(s) + \int_0^s \lambda^{\mathbb{Q}}(u)\,\mathrm{d}u\,, \qquad 0 \leq s \leq T\,, \tag{7.14}$$

$$\log\big(\ell\big(T-s,\widehat{X}(s)\big)\big) \,=\, \widehat{N}^{\,h}(s) + \int_0^s \lambda^{\mathbb{P}}(u)\,\mathrm{d}u\,, \qquad 0 \leq s \leq T\,, \tag{7.15}$$

in the notation of (7.12), (7.13), with $\lambda^{\mathbb{Q}}(s) = \Lambda^{\mathbb{Q}}\big(T-s,\widehat{X}(s)\big)$ *,* $\lambda^{\mathbb{P}}(s) = \Lambda^{\mathbb{P}}\big(T-s,\widehat{X}(s)\big)$ *and*

$$\Lambda^{\mathbb{Q}}(t,x) \,:=\, \ell(t,x)\,\Lambda^{\mathbb{P}}(t,x) \geq 0\,, \qquad \Lambda^{\mathbb{P}}(t,x) \,:=\, \sum_{y\in\mathcal{S},\,y\neq x} \widehat{\kappa}(x,y)\,\Psi\Big(\frac{\ell(t,y)}{\ell(t,x)}\Big) \geq 0\,. \tag{7.16}$$

Here $\widehat{M}^{\,h}$ *is the process of (7.11) and a* $\big(\widehat{\mathbb{G}},\mathbb{Q}\big)-$*martingale, whereas* $\widehat{N}^{\,h}$ *is a* $\big(\widehat{\mathbb{G}},\mathbb{P}\big)-$*martingale.*

Proof: Let us take a look at the expressions of (7.10)–(7.12). As we have already noted, each function $[0,T] \ni t \mapsto p(t,x) \in (0,1)$ is uniformly continuous. This fact, along with the finiteness of the state space $\mathcal{S}$, implies that the quantities in (7.10), (7.12) are actually uniformly bounded. This implies a similar boundedness for the $\big(\widehat{\mathbb{G}},\mathbb{Q}\big)-$local martingale in (7.11), which is thus seen to be a true $\big(\widehat{\mathbb{G}},\mathbb{Q}\big)-$martingale. The remaining claims follow by means of the BAYES rule. $\square$

The decomposition (7.14) is a trajectorial version of relative entropy dissipation, manifesting itself at the level of the *individual particles* that undergo the MARKOV Chain motion as viewed under the lens of the probability measure $\mathbb{Q}$ and under time-reversal, rather than only at the level of their ensembles.

We note that the quantity of (7.16) provides the exact rate of relative entropy dissipation, in the sense that for every $0 < t < T < \infty$ we have the following convergence, a.e. and in $\mathbb{L}^1(\mathbb{P})$:

$$\lim_{s \uparrow T-t} \frac{1}{T-t-s} \left(\mathbb{E}^{\mathbb{P}} \Big[\log \ell\big(t, X(t)\big) \,\Big|\, \widehat{\mathcal{G}}(s) \Big] - \log \big(\ell\big(T-s, \widehat{X}(s)\big)\big) \right) = \Lambda^{\mathbb{P}}\big(t, X(t)\big). \qquad (7.17)$$

The decomposition (7.15) and the trajectorial rate (7.17) are exact analogues of those in Theorem 3.6 and Proposition 3.12 of KARATZAS, SCHACHERMAYER & TSCHIDERER (2019). They constitute trajectorial versions of relative entropy dissipation, viewed now under the original probability measure $\mathbb{P}$ — and again under time-reversal.

7.2 DE BRUIJN-Type Identities

With this preparation, we are now in a position to recover the precise rate of decay for the relative entropy function in (7.2); cf. DIACONIS & SALOFF-COSTE (1996), Lemma 2.5. All it takes, is to "aggregate" (take $Q-$expectations) in (7.14). This leads to an analogue of equation (3.31) in KARATZAS, SCHACHERMAYER & TSCHIDERER (2019) as we describe now. [1]

Theorem 7.3. DE BRUIJN-type identity for the Dissipation of Relative Entropy: *The relative entropy of (7.2) is a decreasing function of time, and satisfies*

$$H\big(P(T)\,|\,Q\big) = H\big(P(0)\,|\,Q\big) - \int_0^T I(t)\,\mathrm{d}t = \int_T^\infty I(t)\,\mathrm{d}t\,, \qquad I(t) \,=\, \mathcal{E}\big(\boldsymbol{\ell}_t, \log \boldsymbol{\ell}_t\big) \geq 0 \qquad (7.18)$$

for all $T \in [0,\infty)$, in the notation of (4.7), (3.8).

Proof: The first claim is simply a restatement of (7.3); and by taking $\mathbb{Q}-$expectations in (7.14) we obtain in conjunction with (7.10) the first equality of (7.18), with

$$I(t) \,:=\, \mathbb{E}^{\mathbb{Q}}\big[\big(\partial h + \widehat{\mathcal{K}} h\big)\big(T-t, X(t)\big)\big].$$

From (7.12) and (4.7), this quantity coincides with the last expression in (7.18): to wit,

$$I(t) \,=\, \sum_{x \in \mathcal{S}} q(x)\,\big(\partial h + \widehat{\mathcal{K}} h\big)(T-t, x) \,=\, \sum_{(x,y) \in \mathcal{Z}} q(x)\,\widehat{\kappa}(x,y)\,\ell(t,y)\Big[\log \frac{\ell(t,y)}{\ell(t,x)} - 1\Big] \qquad (7.19)$$

$$= \sum_{x \in \mathcal{S}} q(x)\,\ell(t,x) \sum_{\substack{y \in \mathcal{S} \\ y \neq x}} \widehat{\kappa}(x,y)\,\Psi\Big(\frac{\ell(t,y)}{\ell(t,x)}\Big) = \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} \kappa(y,x)\,q(y)\,\ell(t,y)\,\log \ell(t,x) \,=\, \mathcal{E}\big(\boldsymbol{\ell}_t, \log \boldsymbol{\ell}_t\big).$$

It is non-negative on account of the non-negativity of the last expression in (7.12), and uniformly continuous as a function of time. In the display (7.19), the second equality follows from the first equality in (7.12); the third from the last equality in (7.12); the fourth from (3.10) and the property $\sum_{y \in \mathcal{S}} \kappa(y,x) = 0$ for every $y \in \mathcal{S}$; and the fifth from the definition (4.7). We deduce $H\big(P(0)\,|\,Q\big) = \int_0^\infty I(t)\,\mathrm{d}t$ by letting $T \to \infty$ in (7.18) and recalling (7.9); then the second identity in (7.18) follows. □

[1] The seminal paper STAM (1959), from the early days of Information Theory, establishes the first identity of this type, and in a context where the invariant measure Q is standard Gaussian. A.J. STAM gives credit for this result to his teacher, the analyst, number theorist, combinatorialist and logician Nicolaas DE BRUIJN.

Remark 7.1. Whenever there exists a positive real constant α (respectively, β) such that the POINCARÉ (resp., the modified log-SOBOLEV) inequality

$$\alpha \leq \frac{2\,\mathcal{E}(f,f)}{\sum_{y\in\mathcal{S}} q(y) f^2(y) - 1} \qquad \left(\text{resp.,} \quad \beta \leq \frac{\mathcal{E}(f,\log f)}{\sum_{y\in\mathcal{S}} q(y) f(y)\log f(y)}\right) \tag{7.20}$$

holds for every function $f : \mathcal{S} \to (0,\infty)$ with $\sum_{y\in\mathcal{S}} q(y)f(y) = 1$, it is clear from (6.5), (6.6) and (7.4), (7.18) that the variance (resp., the relative entropy) decays exponentially:

$$\mathrm{Var}^{\mathbb{Q}}\big(L(t)\big) \leq \mathrm{Var}^{\mathbb{Q}}\big(L(0)\big)\, e^{-\alpha t} \qquad \Big(\text{resp.,} \quad H\big(P(t)\,|\,Q\big) \leq H\big(P(0)\,|\,Q\big)\, e^{-\beta t}\Big)\,. \tag{7.21}$$

Remark 7.2. To the best of our knowledge, the identities (6.6), (7.18) appear in the MARKOV Chain context first in Lemma 2.5 of DIACONIS & SALOFF-COSTE (1996); see also BOBKOV & TETALI (2006), MONTENEGRO & TETALI (2006), CAPUTO ET AL. (2009), and CONFORTI (2020). These authors use slightly different arguments, based on semigroups. One advantage of the more probabilistic approach we follow here, is that it provides a very sharp picture for the dissipation of relative entropy *along trajectories,* as exemplified in subsection 7.1.

7.3 FISHER Information Under Detailed Balance

The following is now a direct consequence of Lemma 4.2.

Proposition 7.4. *Under the detailed-balance condition (3.11), the rate of relative entropy dissipation in (7.18) can be cast as*

$$\begin{aligned} I(t) = \mathcal{E}\big(\boldsymbol{\ell}_t, \log \boldsymbol{\ell}_t\big) &= \frac{1}{2}\sum_{(x,y)\in\mathcal{Z}} \Big(\log \ell\big(t,y\big) - \log \ell\big(t,x\big)\Big)^2\, \Theta\big(\ell(t,y),\ell(t,x)\big)\, \kappa(y,x)\, q(y) \\ &= \frac{1}{2}\sum_{(x,y)\in\mathcal{Z}} \frac{\big(\ell(t,y)-\ell(t,x)\big)^2}{\Theta\big(\ell(t,y),\ell(t,x)\big)}\, \kappa(y,x)\, q(y) \end{aligned} \tag{7.22}$$

in terms of the "logarithmic mean" function

$$\Theta(q,p) := \frac{q-p}{\log q - \log p} = \int_0^1 q^r\, p^{1-r}\, \mathrm{d}r\,, \qquad (q,p)\in(0,\infty)^2. \tag{7.23}$$

Remark 7.3. The expression in (7.22) is reminiscent of the familiar FISHER Information in Statistics and Information Theory; cf. BOBKOV & TETALI (2006). Always under the detailed-balance condition (3.11), the expression of (7.22) can be cast in terms of a "score function", the discrete logarithmic-gradient of the likelihood ratio, as $\big\langle \nabla \boldsymbol{\ell}_t, \nabla \log \boldsymbol{\ell}_t \big\rangle_{\mathbb{L}^2(\mathcal{Z},C)}$ in the notation of (4.1)-(4.5).

As shown in BOBKOV & TETALI (2006), the inequality $2(a-b)^2 \leq (a^2-b^2)\,\log(a/b)$ for $0 < a, b < \infty$ leads under detailed-balance (3.11) to the DIACONIS AND SALOFF-COSTE (1996) estimate

$$\mathcal{E}\big(e^g, g\big) \geq 4\,\mathcal{E}\big(e^{g/2}, e^{g/2}\big)\,, \qquad \text{and thus to} \qquad I(t) = \mathcal{E}\big(\boldsymbol{\ell}_t, \log \boldsymbol{\ell}_t\big) \geq 4\,\mathcal{E}\big(\sqrt{\boldsymbol{\ell}_t}\,, \sqrt{\boldsymbol{\ell}_t}\,\big).$$

8 The $\Phi-$Relative Entropy Process

In order to reveal the common thread running through the examples of the two preceding Sections, let us consider now a function $\Phi : (0,\infty) \to \mathbb{R}$ which is continuously differentiable and convex, with $\Phi(1) = 0$ and with continuous, strictly positive second derivative. We denote by $\varphi : (0,\infty) \to \mathbb{R}$ its derivative $\Phi' = \varphi$. For each $\eta > 0$, $\xi > 0$ we define the BREGMAN $\Phi-$*divergence*

$$\mathrm{div}^{\Phi}\big(\eta\,|\,\xi\big) := \Phi(\eta) - \Phi(\xi) - (\eta-\xi)\,\varphi(\xi)\,, \tag{8.1}$$

a quantity which is non-negative on account of the convexity of Φ (and has nothing to do with the "discrete divergence" we introduced in (4.2)). For instance, $\mathrm{div}^{\Phi}\big(\eta \,|\, \xi\big) = (\xi - \eta)^2$ for $\Phi(\xi) = \xi^2 - 1$; whereas, for $\Phi(\xi) = \xi \, \log \xi$ and in the notation of (7.13), we have

$$\mathrm{div}^{\Phi}\big(\eta \,|\, \xi\big) \,=\, \mathrm{div}^{\Psi}\big(\eta \,|\, \xi\big) \,=\, \xi \, \Psi(\eta/\xi)\,. \tag{8.2}$$

Let us consider now, for a general convex Φ as above, the $\Phi-$*relative entropy*

$$H^{\Phi}\big(P(t)\big|Q\big) \,:=\, \mathbb{E}^{\mathbb{Q}}\big[\Phi\big(\ell(t, X(t))\big)\big] \,=\, \sum_{y \in \mathcal{S}} q(y)\, \Phi\Big(\frac{p(t,y)}{q(y)}\Big)\,, \qquad 0 \le t < \infty\,; \tag{8.3}$$

see CHAFAÏ (2004) for a general study of such functionals. The convexity of Φ and the JENSEN inequality imply that this function is non-negative, since $\Phi(1) = 0$; and from Proposition 5.2, that the *time-reversed* $\Phi-$*relative entropy process*

$$\Phi\big(\ell(T-s, \widehat{X}(s)\big), \qquad 0 \le s \le T$$

is a $(\widehat{\mathbb{G}}, \mathbb{Q})-$submartingale, for every fixed $T \in (0, \infty)$. As a consequence the function in (8.3) decreases, in fact satisfies $\lim_{t\to\infty} \downarrow H^{\Phi}\big(P(t) \,\big|\, Q\big) = 0$ by virtue of (3.6) and the finiteness of the state space.

8.1 Trajectorial Dissipation of the $\Phi-$Relative Entropy

The DOOB-MEYER decomposition of this submartingale is obtained from Proposition 5.1 as follows: Consider the function $g(s,x) = \Phi\big(\ell(T-s,x)\big)$ and compute, in the manner of (7.12), the quantities

$$\partial g(s,x) = -\varphi\big(\ell(t,x)\big) \sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y)\big[\ell(t,y) - \ell(t,x)\big], \quad \big(\widehat{\mathcal{K}}g\big)(s,x) = \sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y)\big[\Phi\big(\ell(t,y)\big) - \Phi\big(\ell(t,x)\big)\big]$$

with $t = T - s$, on account of (3.9). Putting these expressions together with (8.1), we deduce

$$\big(\partial g + \widehat{\mathcal{K}} g\big)(s,x) \,=\, \sum_{y \in \mathcal{S},\, y \neq x} \widehat{\kappa}(x,y)\; \mathrm{div}^{\Phi}\big(\eta \,|\, \xi\big)\Big|_{\substack{\eta = \ell(t,y)\\ \xi = \ell(t,x)}} \,=:\, \Lambda^{\Phi,\mathbb{Q}}(t,x) \,\ge\, 0\,. \tag{8.4}$$

The following result is now a direct consequence of Proposition 5.1 and the BAYES rule. Once again, the finiteness of the state-space and the continuity of the functions involved, turns local into true martingales.

Proposition 8.1. *For any given $T \in (0, \infty)$, the process below is a $(\widehat{\mathbb{G}}, \mathbb{Q})-$martingale:*

$$\Phi\big(\ell(T-s, \widehat{X}(s))\big) - \int_0^s \Lambda^{\Phi,\mathbb{Q}}\big(T-u, \widehat{X}(u)\big)\,\mathrm{d}u\,, \qquad 0 \le s \le T\,. \tag{8.5}$$

Whereas, with $\Lambda^{\Phi,\mathbb{P}}(t,x) := \Lambda^{\Phi,\mathbb{Q}}(t,x)/\ell(t,x)$, the process below is a $(\widehat{\mathbb{G}}, \mathbb{P})-$ martingale:

$$\frac{\Phi\big(\ell(T-s, \widehat{X}(s))\big)}{\ell(T-s, \widehat{X}(s))} - \int_0^s \Lambda^{\Phi,\mathbb{P}}\big(T-u, \widehat{X}(u)\big)\,\mathrm{d}u\,, \qquad 0 \le s \le T. \tag{8.6}$$

8.2 Generalized DE BRUIJN Identities

In view of these considerations, it is now straightforward to "aggregate" (i.e., take $\mathbb{Q}-$expectations of) the $(\widehat{\mathbb{G}}, \mathbb{Q})-$martingale of (8.5). We obtain in the manner of (7.18) the following result, stated again in the forward direction of time; cf. CHAFAÏ (2004), Proposition 1.1.

Proposition 8.2. Generalized DE BRUIJN-type identity: *The temporal dissipation of the $\Phi-$relative entropy of (8.3) is given for $0 \le T < \infty$ as*

$$H^{\Phi}\big(P(T) \,|\, Q\big) = H^{\Phi}\big(P(0) \,|\, Q\big) - \int_0^T I^{\Phi}(t)\,\mathrm{d}t = \int_T^{\infty} I^{\Phi}(t)\,\mathrm{d}t\,, \qquad I^{\Phi}(t) \,:=\, \mathbb{E}^{\mathbb{Q}}\big[\Lambda^{\Phi,\mathbb{Q}}\big(t, X(t)\big)\big] \ge 0\,. \tag{8.7}$$

On the strength of (8.4), the dissipation rate in (8.7) is given by the $\Phi-$FISHER *Information*

$$I^{\Phi}(t) = \sum_{(x,y)\in\mathcal{Z}} q(x)\,\widehat{\kappa}(x,y)\ \mathrm{div}^{\Phi}\big(\ell(t,y)\big|\ell(t,x)\big) = \sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} q(x)\,\widehat{\kappa}(x,y)\,\mathrm{div}^{\Phi}\big(\ell(t,y)\big|\ell(t,x)\big)$$

$$= -\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} q(y)\,\kappa(y,x)\,\ell(t,y)\,\varphi\big(\ell(t,x)\big) = \mathcal{E}\big(\boldsymbol{\ell}_t, \varphi(\boldsymbol{\ell}_t)\big). \tag{8.8}$$

Proof: The third equality in (8.8) is a consequence of the properties $\sum_{y\in\mathcal{S}} \widehat{\kappa}(x,y) = 0$ for every $x \in \mathcal{S}$, and $\sum_{x\in\mathcal{S}} \kappa(y,x) = 0$ for every $y \in \mathcal{S}$, as well as of (3.10). It underscores the fact that, when passing from the trajectorial to the "aggregate" point of view (that is, when taking $\mathbb{Q}-$expectations), the term $\xi\varphi(\xi) - \Phi(\xi)$ that depends only on the variable $\xi = \ell(t,x)$, as well as the term $\Phi(\eta)$ that depends only on the variable $\eta = \ell(t,y)$, can be ignored in (8.1); only the "mixed term" $-\eta\,\varphi(\xi)$ remains relevant. We note in passing that similar reasoning was deployed in the proof of Lemma 4.1. □

Remark 8.1. *Some Special Cases: (i)* For the convex function $\Phi(\xi) = \xi\log\xi$, and recalling (8.2), (7.13), the quantity $I^{\Phi}(t)$ of (8.8) is seen to coincide with $I(t)$ in (7.19), (7.18).
(i) On the other hand, when $\Phi(\xi) = \xi^2 - 1$ we have $\mathrm{div}^{\Phi}\big(\eta\,|\,\xi\big) = (\eta-\xi)^2$ in (8.1) and

$$H^{\Phi}\big(P(t)\big|Q\big) = \mathbb{E}^{\mathbb{Q}}\big(\ell^2(t,X(t)\big) - 1 = \big|\boldsymbol{\ell}_t\big|^2_{\mathbb{L}^2(\mathcal{S},Q)} - 1 = \mathrm{Var}^{\mathbb{Q}}(L(t)) = V\big(P(t)\,|\,Q\big)\,, \qquad 0 \le t < \infty$$

as in (6.1), and the rate of temporal dissipation for this function is precisely the integrand in (6.8):

$$I^{\Phi}(t) = -2\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} q(y)\,\kappa(y,x)\,\ell(t,x)\,\ell(t,y) = 2\,\mathcal{E}\big(\boldsymbol{\ell}_t, \boldsymbol{\ell}_t\big)\,. \tag{8.9}$$

(ii) A bit more generally, the choice of convex function $\Phi(\xi) = (\xi^m - 1)/(m-1)$ for some $m > 1$, leads to the so-called "RÉNYI relative entropy"

$$H^{\Phi}\big(P(t)\big|Q\big) = \frac{\mathbb{E}^{\mathbb{Q}}\big(\ell^m(t,X(t)\big) - 1}{m-1}\,, \qquad 0 \le t < \infty$$

whose rate of temporal dissipation is a generalized version of (8.9), namely

$$I^{\Phi}(t) = -\frac{m}{m-1}\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} q(y)\,\kappa(y,x)\,\ell(t,y)\,\big(\ell(t,x)\big)^{m-1} = \frac{m}{m-1}\,\mathcal{E}\big(\boldsymbol{\ell}_t, \boldsymbol{\ell}_t^{m-1}\big)\,.$$

The variance $\mathrm{Var}^{\mathbb{Q}}(L(t))$ is thus a special case of the RÉNYI relative entropy, corresponding to $m = 2$; whereas, Proposition 7.3 corresponds to the limit of this quantity as $m \downarrow 1$.

We stress that nowhere in this subsection, or in the one preceding it, have we invoked the detailed-balance conditions of (3.11).

8.3 Locally Steepest Descent for the $\Phi-$Relative Entropy Under Detailed Balance

We formulate now a variational version of Proposition 8.2 *under the conditions (3.11) of detailed balance, which will now be in force throughout this subsection.*

Remark 8.2. First, let us take a look at the expression of (8.8). From the consequence $q(x)\,\widehat{\kappa}(x,y) = q(y)\,\kappa(y,x) = q(y)\,\widehat{\kappa}(y,x)$ of the detailed balance conditions (3.11), as well as from the consequence

$$\mathrm{div}^{\Phi}\big(\eta\,|\,\xi\big) + \mathrm{div}^{\Phi}\big(\xi\,|\,\eta\big) = \big(\eta-\xi\big)\big(\varphi(\eta) - \varphi(\xi)\big)$$

of (8.1), we see that the $\Phi-$FISHER Information of (8.8) can be cast in this case as

$$
\begin{aligned}
I^{\Phi}(t) = \;& \frac{1}{2} \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} q(x)\, \widehat{\kappa}(x,y) \, \Big(\mathrm{div}^{\Phi}\big(\eta \,|\, \xi\big) + \mathrm{div}^{\Phi}\big(\xi \,|\, \eta\big) \Big) \Big|_{\substack{\eta = \ell(t,y) \\ \xi = \ell(t,x)}} \\
= \;& \frac{1}{2} \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} q(y)\, \kappa(y,x) \, \Big(\big(\eta - \xi\big)\big(\varphi(\eta) - \varphi(\xi)\big) \Big) \Big|_{\substack{\eta = \ell(t,y) \\ \xi = \ell(t,x)}} = \mathcal{E}\big(\varphi(\boldsymbol{\ell}_t), \boldsymbol{\ell}_t\big) \qquad (8.10) \\
= \;& \frac{1}{2} \sum_{(x,y) \in \mathcal{Z}} q(x)\, \kappa(x,y)\, \Theta^{\Phi}(\xi, \eta) \big(\varphi(\xi) - \varphi(\eta)\big)^2 \Big|_{\substack{\xi = \ell(t,x) \\ \eta = \ell(t,y)}}
\end{aligned}
$$

in the manner of (7.22); we recall the notation $\varphi = \Phi'$. Here the function

$$
\Theta^{\Phi}(q,p) \;:=\; \frac{q-p}{\varphi(q) - \varphi(p)}\,, \quad 0 < q \neq p < \infty\,; \qquad \Theta^{\Phi}(p,p) \;:=\; \frac{1}{\Phi''(p)}\,, \quad 0 < p < \infty \qquad (8.11)
$$

extends the "logarithmic mean" of (7.23), to which it reduces when $\Phi(\xi) = \xi \log \xi$. With $\Phi(\xi) = \xi^2 - 1$ we get $\Theta^{\Phi} \equiv 1/2,$ and the last expression in (8.10) reduces to $\sum_{(x,y) \in \mathcal{Z}} q(x)\, \kappa(x,y) \cdot \big(\ell(t,x) - \ell(t,y)\big)^2$ as in (6.8).

We set out now to find a metric on the manifold $\mathcal{M} = \mathcal{P}_+(\mathcal{S})$ *of probability vectors on the state-space, relative to which the time-marginals for the* MARKOV *Chain* $(P(t))_{0 \le t < \infty}$ *constitute a curve of steepest descent for the* $\Phi-$*relative entropy.* In other words, we look for a metric on $\mathcal{M}$ that can play — in the current general context — a role similar to that played by the HILBERT norm $\| \cdot \|_{\mathbb{H}^{-1}(Q)}$ in Section 6. This norm defines the metric distance of (6.10) that works for the variance $V(P(t)|Q)$, that is, in the special case $\Phi(\xi) = \xi^2 - 1$. But except for such very special cases, the Riemannian metric on the manifold $\mathcal{M}$ will *not be flat;* i.e., *not* induced by such a simple norm as in Proposition 6.5. For this reason we are forced to consider the machinery of Riemannian geometry, which we take up in the next Section 9. In this section we avoid Riemannian terminology and present the steepest descent property of the curve $(P(t))_{0 \le t < \infty}$, in terms of appropriate Hilbert norms that capture the local behaviour of the Riemannian metric.

We start this effort by recalling from (4.6) the norm $\|F\|_Q^2$ for functions $F : \mathcal{Z} \to \mathbb{R}\,,$ defined on the "off-diagonal Cartesian product" $\mathcal{Z}$ by assigning to its elements $(x,y),\ x \neq y$ the weights $q(x)\, \kappa(x,y)/2$ and taking the usual $\mathbb{L}^2-$norm relative to the positive measure with these weights. For a fixed likelihood ratio ℓ in the space $\mathcal{L} = \mathcal{L}_+(\mathcal{S})$ of subsection 3.2, we consider in place of $c(x,y) \equiv q(x)\, \kappa(x,y)/2$ the new weights

$$
c(x,y) \cdot \vartheta_\ell(x,y)\,, \qquad \text{where} \qquad \vartheta_\ell(x,y) := \Theta^{\Phi}\big(\ell(x), \ell(y)\big) \;=\; \frac{\nabla \ell(x,y)}{\nabla(\varphi \circ \ell)(x,y)} \qquad (8.12)
$$

in the notation of (8.11). The resulting *weighted* inner product and norm, extensions of the respective quantities for real-valued functions on $\mathcal{S} \times \mathcal{S}$ in (4.5), (4.6) (to which they reduce when $\Phi(\xi) = \xi^2/2$), are

$$
\begin{aligned}
\big\langle F, G \big\rangle_{\mathbb{L}^2(\mathcal{Z}, \vartheta_\ell C)} \;&:=\; \sum_{(x,y) \in \mathcal{Z}} c(x,y)\, \vartheta_\ell(x,y) F(x,y)\, G(x,y) \;\;=\; \big\langle\, \vartheta_\ell F, G \,\big\rangle_{\mathbb{L}^2(\mathcal{Z}, C)}\,, \\
\big\| F \big\|^2_{\mathbb{L}^2(\mathcal{Z}, \vartheta_\ell C)} \;&:=\; \big\langle F, F \big\rangle_{\mathbb{L}^2(\mathcal{Z}, \vartheta_\ell C)}\,,
\end{aligned} \qquad (8.13)
$$

respectively. Similarly, we define for functions $f : \mathcal{S} \to \mathbb{R}$ the *Weighted* SOBOLEV *Norm* $\| \cdot \|_{\mathbb{H}^1_\Theta(\mathcal{S}, \ell Q)}$ by replacing on the right-hand sides of (4.10)–(4.12), the norm $\| \cdot \|_{\mathbb{L}^2(\mathcal{Z}, C)}$ by the new norm $\| \cdot \|_{\mathbb{L}^2(\mathcal{Z}, \vartheta_\ell C)}$ in (8.13):

$$
\big\langle f, g \big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S}, \ell Q)} \;:=\; \big\langle \nabla f, \nabla g \big\rangle_{\mathbb{L}^2(\mathcal{Z}, \vartheta_\ell C)}\,, \qquad \big\| f \big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S}, \ell Q)} \;:=\; \big\langle f, f \big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S}, \ell Q)}\,. \qquad (8.14)
$$

Remark 8.3. It is interesting to note at this point, and will become important down the road, that the $\Phi-$FISHER Information of (8.8), (8.10) can be expressed in terms of the square of this new, weighted SOBOLEV norm. Indeed, for any $\ell \in \mathcal{L}_+(\mathcal{S})$ we have

$$\mathcal{E}\big(\ell, \varphi(\ell)\big) = \big\langle \nabla \ell, \nabla \varphi(\ell) \big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} = \big\langle \vartheta_\ell \nabla \varphi(\ell), \nabla \varphi(\ell) \big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} = \big\| \nabla \varphi(\ell) \big\|^2_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} = \big\| \varphi(\ell) \big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)}$$

so that $I^\Phi(t) = \big\| \varphi(\boldsymbol{\ell}_t) \big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S},\boldsymbol{\ell}_t Q)}$

We introduce also, in the manner of (4.13), (4.14), the dual

$$\big\|f\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)} := \sup_{g:\mathcal{S}\to\mathbb{R}} \frac{\big\langle f, g \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)}}{\big\|g\big\|_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)}} \tag{8.15}$$

of this SOBOLEV norm. The following characterization will be crucial in what follows.

Proposition 8.3. *For any function $f : \mathcal{S} \to \mathbb{R}$ we have*

$$\big\|f\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)} = \inf_{G:\mathcal{Z}\to\mathbb{R}} \Big\{ \|G\|_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} \ : \ f + \nabla \cdot \big(\vartheta_\ell G\big) = 0 \Big\} \tag{8.16}$$

Moreover, $\big\|f\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)}$ is finite if, and only if, $\sum_{x\in\mathcal{S}} q(x) f(x) = 0$; and in this case the infimum is attained, uniquely, by the unique discrete gradient that is admissible.

Proof. Consider a function $f : \mathcal{S} \to \mathbb{R}$ such that $\sum_{x\in\mathcal{S}} q(x) f(x) = 0$; if this is not the case, it is straightforward to verify that both sides in (8.16) are infinite. It is easy to see that the set of admissible G on the right-hand side of (8.16) is non-empty (indeed, $G_0 := -\frac{1}{\vartheta_\ell} \nabla \mathcal{K}^{-1} f$ is admissible) and that a minimizer exists. Let $G : \mathcal{Z} \to \mathbb{R}$ be such a minimizer. We show first that G is a discrete gradient, by a projection argument in the HILBERT space $\mathbb{L}^2(\mathcal{Z}, \vartheta_\ell C)$.

To this end, let us denote by ∇h the orthogonal projection of G onto the subspace of discrete gradients in $\mathcal{H}_\ell$. We claim that ∇h is admissible on the right-hand side of (8.16). Indeed, $G - \nabla h$ is orthogonal in $\mathbb{L}^2(\mathcal{Z}, \vartheta_\ell C)$ to ∇g for any $g : \mathcal{S} \to \mathbb{R}$. This implies

$$-\big\langle g, \nabla \cdot \big(\vartheta_\ell (G - \nabla h)\big) \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = \big\langle \nabla g, \vartheta_\ell (G - \nabla h) \big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} = \big\langle \nabla g, G - \nabla h \big\rangle_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} = 0$$

and yields $\nabla \cdot \big(\vartheta_\ell G\big) = \nabla \cdot \big(\vartheta_\ell \nabla h\big)$, proving the claim.

By orthogonality, we have $\|G\|^2_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} = \|\nabla h\|^2_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} + \|G - \nabla h\|^2_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)}$. Since G is a minimizer, we infer $\|G - \nabla h\|_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} = 0$, which implies that $G \equiv \nabla h$. This shows that ∇h is a minimizer, and that the right-hand side of (8.16) is equal to $\big\|h\big\|_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)}$. It is shown in MAAS (2011) that ∇h is actually the *unique* discrete gradient satisfying the constraint in (8.16).

To prove the equality in (8.16), we note for any $g : \mathcal{S} \to \mathbb{R}$ the identities

$$\langle f, g \rangle_{\mathbb{L}^2(\mathcal{S},Q)} = -\big\langle \nabla \cdot \big(\vartheta_\ell \nabla h\big), g \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = \big\langle \vartheta_\ell \nabla h, \nabla g \big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} = \big\langle h, g \big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} .$$

Writing the dual norm as a LEGENDRE transform, we obtain

$$\begin{aligned} \big\|f\big\|^2_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)} &= \sup_{g:\mathcal{S}\to\mathbb{R}} \Big\{ 2\langle f, g \rangle_{\mathbb{L}^2(\mathcal{S},Q)} - \big\|g\big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} \Big\} \\ &= \sup_{g:\mathcal{S}\to\mathbb{R}} \Big\{ 2\big\langle h, g \big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} - \big\|g\big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} \Big\} = \big\|h\big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} , \end{aligned}$$

which establishes the equality in (8.16). □

Let us consider now, as we did in Section 6, for some $\varepsilon > 0$ an arbitrary smooth curve $\ell^\psi(\cdot) = (\ell^\psi(t))_{t_0 \le t < t_0+\varepsilon}$ with initial position $\ell^\psi(t_0) = \boldsymbol{\ell} \equiv \boldsymbol{\ell}(t_0)$ in $\mathcal{L} = \mathcal{L}_+(\mathcal{S})$. In order to compute $\mathbb{H}^{-1}_\Theta(\mathcal{S}, \ell Q)$-norms, it is natural in view of Proposition 8.3 to write the time-evolution in the manner of a "discrete continuity equation" $\partial \ell^\psi_t + \nabla \cdot (\vartheta_{\ell_t} \nabla \psi_t) = 0$ as in subsection 6.1, where $\psi_t : \mathcal{S} \to \mathbb{R}$ is unique up to an additive constant. We regard $\nabla \psi_t$ as an input, or velocity vector field, that yields the infinitesimal change $\partial \ell^\psi_t$ of the likelihood ratio flow. The original backward KOLMOGOROV equation $\partial \boldsymbol{\ell}_t = \nabla \cdot (\nabla \boldsymbol{\ell}_t) = \mathcal{K} \boldsymbol{\ell}_t$ of (3.9) corresponds to $\psi_t = -\varphi(\boldsymbol{\ell}_t)$ in this scheme of things.

We define as in (6.12) the corresponding curve $P^\psi(\cdot) = \big(P^\psi(t)\big)_{t_0 \le t < t_0+\varepsilon}$ on the manifold $\mathcal{M} = \mathcal{P}_+(\mathcal{S})$ of probability vectors on the state-space. We get the following generalization of Proposition 8.2.

Proposition 8.4. *In the above context, we have*

$$\partial H^\Phi\big(P^\psi(t) \,|\, Q\big) = \big\langle\, \varphi(\ell^\psi_t), \psi_t \,\big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S}, \ell_t Q)} \,. \tag{8.17}$$

Proof. Using the discrete continuity equation, a discrete integration by parts, and the definitions of the scalar products, we obtain

$$\begin{aligned}\partial H^\Phi\big(P^\psi(t) \,|\, Q\big) &= \partial\, \mathbb{E}^{\mathbb{Q}}\Big[\Phi\big(\ell^\psi\big(t, X(t)\big)\big)\Big] = -\big\langle\, \varphi(\ell^\psi_t), \nabla \cdot (\vartheta_{\ell_t} \nabla \psi_t) \,\big\rangle_{\mathbb{L}^2(\mathcal{S}, Q)} \\ &= \big\langle\, \nabla\varphi(\ell^\psi_t), \vartheta_{\ell_t} \nabla \psi_t \,\big\rangle_{\mathbb{L}^2(\mathcal{Z}, C)} = \big\langle\, \nabla\varphi(\ell^\psi_t), \nabla \psi_t \,\big\rangle_{\mathbb{L}^2(\mathcal{Z}, \vartheta_{\ell_t} C)} = \big\langle\, \varphi(\ell^\psi_t), \psi_t \,\big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S}, \ell_t Q)} \,,\end{aligned}$$

as desired. □

With the context and notation just established, and always with $\boldsymbol{\ell} \equiv \boldsymbol{\ell}(t_0) = (\ell(t_0, x))_{x \in \mathcal{S}}$, we can formulate the following result, an analogue of Proposition 8.4, which follows from the characterizations of the weighted $\mathbb{H}^{-1}$-norm in (8.15), in view of the identity $\mathcal{K}\boldsymbol{\ell} = \nabla \cdot \big(\vartheta_{\boldsymbol{\ell}} \nabla \varphi(\boldsymbol{\ell})\big)$.

Proposition 8.5. *Under the conditions (3.11) of detailed balance, we have*

$$\lim_{h \downarrow 0} \frac{1}{h} \big\|\, \boldsymbol{\ell}_{t_0+h} - \boldsymbol{\ell}_{t_0} \,\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S}, \boldsymbol{\ell} Q)} \;=\; \big\|\, \mathcal{K}\, \boldsymbol{\ell}_{t_0} \,\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S}, \boldsymbol{\ell} Q)} \;=\; \big\|\, \varphi(\boldsymbol{\ell}_{t_0}) \,\big\|_{\mathbb{H}^1_\Theta(\mathcal{S}, \boldsymbol{\ell} Q)} \,; \tag{8.18}$$

and a bit more generally,

$$\lim_{h \downarrow 0} \frac{1}{h} \big\|\, \ell^\psi_{t_0+h} - \ell^\psi_{t_0} \,\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S}, \boldsymbol{\ell} Q)} \;=\; \big\|\, \nabla \cdot (\vartheta_{t_0} \nabla \psi_{t_0}) \,\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S}, \boldsymbol{\ell} Q)} \;=\; \big\|\psi_{t_0}\big\|_{\mathbb{H}^1_\Theta(\mathcal{S}, \boldsymbol{\ell} Q)} \,. \tag{8.19}$$

We pass now to the principal result of this Section. It is a direct analogue of Theorem 3.4 in KARATZAS, SCHACHERMAYER & TSCHIDERER (2019), where a similar steepest-descent for the dissipation of the relative entropy is established in the context of LANGEVIN-SCHMOLUCHOWSKI diffusions, and with distance on the ambient space measured by the quadratic WASSERSTEIN metric.

Theorem 8.6. Steepest Descent for the $\Phi-$Relative Entropy: *Under the conditions (3.11) of detailed balance, the curve $(P(t))_{t_0 \le t < \infty}$ of time-marginal distributions in (3.4) has the property of steepest descent for the $\Phi-$Relative Entropy of (8.3) locally at $t = t_0$, with respect to the distance induced by the "flat metric"*

$$\varrho_\star\big(P_1, P_2\big) \;:=\; \big\|\, \boldsymbol{\ell}_1 - \boldsymbol{\ell}_2 \,\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S}, \ell Q)} \qquad \text{for} \;\; P_1 = \boldsymbol{\ell}_1 Q \;\; \text{and} \;\; P_2 = \boldsymbol{\ell}_2\, Q \tag{8.20}$$

and with the notation of (8.13)–(8.14), in the sense of Definition 6.1.

Proof: This is proved exactly as in subsection 6.2, with the caveat that the distance-inducing flat metric is now determined "locally", that is, depends on $(t_0, \boldsymbol{\ell}) \equiv (t_0, \boldsymbol{\ell}(t_0))$ in the weighted norms of (8.13)–(8.15). But let us go through the argument again, to highlight the role that these weighted norms play in the present, more general context. From (8.17), and recalling the initial position $\ell^\psi(t_0) = \boldsymbol{\ell}(t_0) \in \mathcal{L}$, we obtain

$$\lim_{h \downarrow 0} \frac{H^\Phi\big(P^\psi(t_0+h) \,|\, Q\big) - H^\Phi\big(P(t_0) \,|\, Q\big)}{h} \;=\; \Big\langle\, \varphi(\boldsymbol{\ell}_{t_0}), \psi_{t_0} \Big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S}, \boldsymbol{\ell} Q)} \,;$$

whereas, (8.19) gives

$$\lim_{h\downarrow 0} \frac{\varrho_\star\big(P^\psi(t_0+h), P(t_0)\big)}{h} \;=\; \big\|\psi_{t_0}\big\|_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)}\,,$$

thus

$$\lim_{h\downarrow 0} \frac{H^\Phi\big(P^\psi(t_0+h)\,|\,Q\big) - H^\Phi\big(P(t_0)\,|\,Q\big)}{\varrho_\star\big(P^\psi(t_0+h), P(t_0)\big)} \;=\; \Big\langle\, \varphi(\boldsymbol{\ell}_{t_0}),\, \frac{\psi_{t_0}}{\big\|\psi_{t_0}\big\|_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)}} \Big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)}\,.$$

This is the rate of change for the $\Phi-$relative entropy along the *perturbed curve* $\big(P^\psi(t)\big)_{t_0\le t<t_0+\varepsilon}$, as measured on the manifold $\mathcal{M}$ with respect to the distance in (8.20).

On the other hand, we have from (8.8), (8.7) and (8.10), the following observation: Along the *original curve* of time-marginal distributions $\big(P(t)\big)_{t_0\le t<\infty}$ for the Chain, which corresponds to taking $\psi(\cdot)\equiv\boldsymbol{\ell}(\cdot)$ above, the rate of $\Phi-$relative entropy dissipation measured in terms of the "flat metric" distance traveled on the manifold $\mathcal{M}$ is given as

$$\lim_{h\downarrow 0} \frac{H^\Phi\big(P(t_0+h)\,|\,Q\big) - H^\Phi\big(P(t_0)\,|\,Q\big)}{\varrho_\star\big(P(t_0+h), P(t_0)\big)} \;=\; -\,\big\|\varphi(\boldsymbol{\ell}_{t_0})\big\|_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} \;<\; 0\,.$$

A simple comparison of the last two displays, via CAUCHY-SCHWARZ, gives the steepest descent property

$$\lim_{h\downarrow 0} \frac{H^\Phi\big(P^\psi(t_0+h)\,|\,Q\big) - H^\Phi\big(P(t_0)\,|\,Q\big)}{\varrho_\star\big(P^\psi(t_0+h), P(t_0)\big)} \;-\; \lim_{h\downarrow 0} \frac{H^\Phi\big(P(t_0+h)\,|\,Q\big) - H^\Phi\big(P(t_0)\,|\,Q\big)}{\varrho_\star\big(P(t_0+h), P(t_0)\big)}$$

$$=\; \big\|\varphi(\boldsymbol{\ell}_{t_0})\big\|_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} + \Big\langle\, \varphi(\boldsymbol{\ell}_{t_0}),\, \frac{\psi_{t_0}}{\big\|\psi_{t_0}\big\|_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)}} \Big\rangle_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} \;\ge\; 0$$

of the $\Phi-$relative entropy with respect to the distance in (6.10), along the original curve of MARKOV Chain time-marginals. Equality holds if, and only if, $\nabla\psi_{t_0}$ is a negative constant multiple of $\nabla\varphi(\boldsymbol{\ell}_{t_0})$. $\square$

8.4 Nonuniqueness of the metric

There exist norms other than $\mathbb{H}^{-1}_\Theta(\mathcal{S},\boldsymbol{\ell}Q)$ of (8.15), for which Theorem 8.6 remains valid; see DIETERT (2015) and Proposition 9.4 below. Here we exhibit an explicit example.

Fix $\ell\in\mathcal{L}_+(\mathcal{S})$ and consider the *"modified weighted* $\mathbb{H}^{-1}-$*norm"* given by

$$\big\|f\big\|^2_{\widetilde{\mathbb{H}}^{-1}_\Theta(\mathcal{S},\ell Q)} \;:=\; \Big\langle \frac{1}{\vartheta_\ell}\nabla\big(\mathcal{K}^{-1}f\big), \nabla\big(\mathcal{K}^{-1}f\big)\Big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} \tag{8.21}$$

for functions $f\,:\,\mathcal{S}\to\mathbb{R}$ with $\sum_{x\in\mathcal{S}} f(x)q(x) \;=\; 0$. This norm is never smaller than the original $\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)-$norm; namely,

$$\big\|f\big\|_{\widetilde{\mathbb{H}}^{-1}_\Theta(\mathcal{S},\ell Q)} \;\ge\; \big\|f\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)}\,. \tag{8.22}$$

And equality holds when $f=\mathcal{K}\ell$; to wit,

$$\big\|\mathcal{K}\ell\big\|_{\widetilde{\mathbb{H}}^{-1}_\Theta(\mathcal{S},\ell Q)} \;=\; \big\|\mathcal{K}\ell\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)}. \tag{8.23}$$

These two facts immediately imply that the curve $\big(P(t)\big)_{t_0\le t<\infty}$ from Theorem 8.6, which corresponds to the original backward equation $\partial\boldsymbol{\ell}_t \;=\; \nabla\cdot(\nabla\boldsymbol{\ell}_t) \;=\; \mathcal{K}\boldsymbol{\ell}_t$ of (3.9), is a curve of steepest descent also with respect to the modified norms $\widetilde{\mathbb{H}}^{-1}_\Theta(\mathcal{S},\ell Q)$ in (8.21).

To prove the inequality (8.22), we use Proposition 8.3 and the identity $\mathcal{K}f = \nabla \cdot (\nabla f)$ to obtain

$$\begin{aligned} \big\|f\big\|^2_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)} &= \inf_{G:\mathcal{Z}\to\mathbb{R}} \Big\{ \big\langle G, \vartheta_\ell G \big\rangle_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} \;:\; f + \nabla \cdot \big(\vartheta_\ell G\big) = 0 \Big\} \\ &\le \Big\langle \frac{1}{\vartheta_\ell} \nabla\big(\mathcal{K}^{-1} f\big),\, \vartheta_\ell \Big(\frac{1}{\vartheta_\ell} \nabla \big(\mathcal{K}^{-1} f\big)\Big) \Big\rangle_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)} = \big\|f\big\|^2_{\widetilde{\mathbb{H}}^{-1}_\Theta(\mathcal{S},\ell Q)} . \end{aligned} \tag{8.24}$$

The equality (8.23) holds since, on the one hand,

$$\begin{aligned} \big\|\mathcal{K}\ell\big\|^2_{\widetilde{\mathbb{H}}^{-1}_\Theta(\mathcal{S},\ell Q)} = \Big\langle \frac{1}{\vartheta_\ell} \nabla \ell, \nabla \ell \Big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} &= \Big\langle \nabla\varphi(\ell), \nabla \ell \Big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} \\ &= \Big\langle \nabla\varphi(\ell), \vartheta_\ell \nabla\varphi(\ell) \Big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} = \big\|\varphi(\ell)\big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} \;; \end{aligned}$$

while, on the other hand, Proposition 8.3 yields

$$\big\|\mathcal{K}\ell\big\|^2_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)} = \big\|\nabla \cdot (\nabla \ell)\big\|^2_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)} = \big\|\nabla \cdot \big(\vartheta_\ell \nabla \varphi(\ell)\big)\big\|^2_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)} = \big\|\varphi(\ell)\big\|^2_{\mathbb{H}^1_\Theta(\mathcal{S},\ell Q)} .$$

Remark 8.4. In general, the norms $\big\|f\big\|_{\widetilde{\mathbb{H}}^{-1}_\Theta(\mathcal{S},\ell Q)}$ and $\big\|f\big\|_{\mathbb{H}^{-1}_\Theta(\mathcal{S},\ell Q)}$ are different. Indeed, it follows from Proposition 8.3 and (8.24) that equality of norms holds if, and only if, $\frac{1}{\vartheta_\ell}\nabla\big(\mathcal{K}^{-1}f\big)$ is a discrete gradient. This is in general false, but it is true in the following very special cases:

- At equilibrium, i.e., with $\ell \equiv 1$, we have $\vartheta_\ell \equiv 1$, so that $\frac{1}{\vartheta_\ell}\nabla\big(\mathcal{K}^{-1}f\big) = \nabla\big(\mathcal{K}^{-1}f\big)$;
- For the multiple $\Phi(\xi) = \frac{1}{2}(\xi^2 - 1)$ of the variance in Section 6, we have $\vartheta_\ell \equiv 1$ for every likelihood ratio ℓ ;
- If the state space $\mathcal{S}$ consists of only two points, $\frac{1}{\vartheta_\ell}\nabla\big(\mathcal{K}^{-1}f\big)$ is a discrete gradient, since this holds for every anti-symmetric function on $\mathcal{S} \times \mathcal{S}$.

9 Gradient Flows

Let us reconsider now, under conditions of detailed balance, the results of Sections 6–9 from a different, "Riemannian" point of view. We shall see here that, *under the conditions (3.11), the curve $(P(t))_{0\le t<\infty}$ of time-marginal distributions for the Chain evolves as a gradient flow of the relative $\Phi-$entropy.* This takes place in a suitable geometry on the space of probability measures, in the spirit of the pioneering work by JORDAN, KINDERLEHRER & OTTO (1998). We refer to ERBAR & MAAS (2012, 2014), MIELKE (2011, 2013) and to the expository paper MAAS (2017), for an in-depth study of such issues in discrete spaces.

We summon from subsection 3.1 the manifold $\mathcal{M} = \mathcal{P}_+(\mathcal{S})$ of probability vectors with strictly positive entries $P = \big(p(x)\big)_{x\in\mathcal{S}}$ on the state-space, and denote by $\mathcal{M}_0(\mathcal{S})$ the collection of vectors $W = \big(w(x)\big)_{x\in\mathcal{S}}$ with total mass $\sum_{x\in\mathcal{S}} w(x) = 0$ (viewed as "signed measures"). We note that $\mathcal{M}$ is a relatively open subset of the $(n-1)-$dimensional affine space $P + \mathcal{M}_0(\mathcal{S}) = \{P + W : W \in \mathcal{M}_0(\mathcal{S})\}$, where $P \in \mathcal{M}$ is arbitrary and n is the cardinality of the state-space. This observation allows us to identify the tangent space at each $P \in \mathcal{M}$ with $\mathcal{M}_0(\mathcal{S})$.

9.1 Gradient Flow for the Variance

As a warmup, let us start as in Section 6 with a derivation for the gradient flow property for the variance functional $\mathcal{M} \ni P \mapsto V(P|Q) \in \mathbb{R}_+$ of (6.1). Following DE GIORGI's approach to *curves of maximal slope* (cf. AMBROSIO, GIGLI & SAVARÉ (2008)), we compute the dissipation of this functional along an arbitrary smooth curve $(\widetilde{P}_t)_{0\le t<\infty}$ in $\mathcal{M}$; or equivalently, along the curve $(\widetilde{\ell}_t)_{0\le t<\infty}$ induced on $\mathcal{L}$ by the likelihood ratios $\widetilde{\ell}_t(y) = \widetilde{p}_t(y)/q(y)$.

As in Section 6, we express the time-evolution of this likelihood ratio curve as $\partial \widetilde{\ell}_t = \mathcal{K} f_t = \nabla \cdot \big(\nabla f_t \big)$ in the manner of (3.9), for a suitable curve $(f_t)_{0 \le t < \infty}$ with $f_t : \mathcal{S} \to \mathbb{R}$. This is uniquely determined up to an additive constant on account of the Chain's irreducibility, and its discrete gradient provides the "momentum vector field" of the motion. Recalling the consequences $\widehat{\mathcal{K}} f = \mathcal{K} f = \nabla \cdot (\nabla f)$ of detailed balance and of (3.11), (4.3), as well as the fact that $\nabla \cdot$ is the adjoint of $-\nabla$ from (4.9), we obtain

$$\begin{aligned} \partial V\big(\widetilde{P}_t | Q\big) &= \partial \big\| \widetilde{\ell}_t \big\|^2_{\mathbb{L}^2(\mathcal{S},Q)} = 2 \big\langle \widetilde{\ell}_t, \partial \widetilde{\ell}_t \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = 2 \big\langle \widetilde{\ell}_t, \mathcal{K} f_t \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = -2\, \mathcal{E}\big(\widetilde{\ell}_t, f_t \big) \\ &= -2 \big\langle \nabla \widetilde{\ell}_t, \nabla f_t \big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} \ge -2 \, \big\| \nabla \widetilde{\ell}_t \big\|_{\mathbb{L}^2(\mathcal{Z},C)} \, \big\| \nabla f_t \big\|_{\mathbb{L}^2(\mathcal{Z},C)} \ge - \big\| \nabla \widetilde{\ell}_t \big\|^2_{\mathbb{L}^2(\mathcal{Z},C)} - \big\| \nabla f_t \big\|^2_{\mathbb{L}^2(\mathcal{Z},C)} . \end{aligned} \tag{9.1}$$

Equality holds in the first (resp., the second) of these inequalities if, and only if, ∇f_t and $\nabla \widetilde{\ell}_t$ are positively collinear (resp., have the same norm). In other words, both these last two inequalities hold as equalities if and only if $\nabla f_t = \nabla \widetilde{\ell}_t$, and this leads to the backwards equation (3.9) on account of detailed balance:

$$\partial \widetilde{\ell}_t = \mathcal{K} f_t = \nabla \cdot \big(\nabla f_t \big) = \nabla \cdot \big(\nabla \widetilde{\ell}_t \big) = \mathcal{K}\, \widetilde{\ell}_t = \widehat{\mathcal{K}}\, \widetilde{\ell}_t \, .$$

But the last two norms in (9.1) are $\| \nabla f_t \|_{\mathbb{L}^2(\mathcal{Z},C)} = \| \nabla (\mathcal{K}^{-1} (\partial \widetilde{\ell}_t)) \|_{\mathbb{L}^2(\mathcal{Z},C)} = \| \partial \widetilde{\ell}_t \|_{\mathbb{H}^{-1}(\mathcal{S},Q)}$ as well as $\| \nabla \widetilde{\ell}_t \|_{\mathbb{L}^2(\mathcal{Z},C)} = \| \widetilde{\ell}_t \|_{\mathbb{H}^1(\mathcal{S},Q)}$.

In this manner we obtain from (9.1) the following classical result. This provides another proof for Theorem 6.3 by identifying the solutions of $\partial P_t = \mathcal{K}' P_t$ in (3.5) as curves in the direction of steepest descent for the variance, relative to the distance induced by the $\mathbb{H}^{-1}(\mathcal{S}, Q)$ norm. But it also strengthens Theorem 6.3, by identifying as well the correct velocity with which the gradient flow moves into this direction.

Theorem 9.1. *For any given probability vector $P \in \mathcal{M}$, along any smooth curve $(\widetilde{P}_t)_{0 \le t < \infty}$ in $\mathcal{M}$ with $\widetilde{P}_0 = P$, and with $\boldsymbol{\ell} \in \mathcal{L}$ the likelihood ratio vector corresponding to P, we have*

$$\bigg(\partial V\big(\widetilde{P}_t | Q\big) + \big\| \partial \widetilde{\ell}_t \big\|^2_{\mathbb{H}^{-1}(\mathcal{S},Q)} \bigg) \bigg|_{t=0} \ge - \big\| \boldsymbol{\ell} \big\|^2_{\mathbb{H}^1(\mathcal{S},Q)} .$$

Equality holds here if, and only if, the curve $(\widetilde{P}_t)_{0 \le t < \infty} \subset \mathcal{M}$ satisfies the forward equation $\partial \widetilde{P}_t = \mathcal{K}' \widetilde{P}_t$ (equivalently, the likelihood ratio curve $(\widetilde{\ell}_t)_{0 \le t < \infty} \subset \mathcal{L}$ satisfies the backward equation $\partial \widetilde{\ell}_t = \mathcal{K} \widetilde{\ell}_t$).

9.2 Gradient Flow for the $\Phi-$Relative Entropy

Let us examine now, how these ideas might work in the context of the generalized relative entropy functional

$$\mathcal{M} \ni P \longmapsto H^{\Phi}\big(P | Q\big) := \sum_{y \in \mathcal{S}} q(y)\, \Phi\Big(\frac{p(y)}{q(y)} \Big) \in [0, \infty) \tag{9.2}$$

corresponding to a convex function Φ, as in Section 8. We fix a smooth curve $(\widetilde{P}_t)_{0 \le t < \infty}$ on $\mathcal{M}$ emanating from a given $\widetilde{P}_0 = P \in \mathcal{M}$; and consider the induced curve $(\widetilde{\ell}_t)_{0 \le t < \infty} \subset \mathcal{L}$ of likelihood ratios $\widetilde{\ell}_t(y) = \widetilde{p}_t(y)/q(y)$, $y \in \mathcal{S}$ emanating from $\boldsymbol{\ell} = \ell_0$.

As in subsection 8.3, we cast the time-evolution of the likelihood ratio curve as a "continuity equation"

$$\partial \widetilde{\ell}_t + \nabla \cdot \big(\widetilde{\vartheta}_t \nabla f_t \big) = 0 \tag{9.3}$$

where the "velocity vector field" is the discrete gradient of a suitable function $f_t : \mathcal{S} \to \mathbb{R}$, and $\widetilde{\vartheta}_t$ is a shorthand for $\vartheta_{\widetilde{\ell}_t}$ from (8.12). In the manner of (9.1), this expresses the time-evolution of the $\Phi-$relative entropy functional in (9.2) along the curve $\big(\widetilde{P}_t\big)_{0 \le t < \infty}$ as

$$
\begin{aligned}
\partial H^{\Phi}\big(\widetilde{P}_t\big|Q\big) &= \big\langle \varphi(\widetilde{\ell}_t), \partial\widetilde{\ell}_t \big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = -\big\langle \varphi(\widetilde{\ell}_t), \nabla\cdot(\widetilde{\vartheta}_t \nabla f_t)\big\rangle_{\mathbb{L}^2(\mathcal{S},Q)} = \big\langle\, \nabla(\varphi(\widetilde{\ell}_t)), \widetilde{\vartheta}_t \nabla f_t \,\big\rangle_{\mathbb{L}^2(\mathcal{Z},C)} \\
&= \big\langle\, \nabla\varphi(\widetilde{\ell}_t), \nabla f_t \,\big\rangle_{\mathbb{L}^2(\mathcal{Z},\widetilde{\vartheta}_t C)} \geq -\big\|\nabla\varphi(\widetilde{\ell}_t)\big\|_{\mathbb{L}^2(\mathcal{Z},\widetilde{\vartheta}_t C)} \big\|\nabla f_t\big\|_{\mathbb{L}^2(\mathcal{Z},\widetilde{\vartheta}_t C)} \\
&\geq -\Big(\big\|\nabla\varphi(\widetilde{\ell}_t)\big\|^2_{\mathbb{L}^2(\mathcal{Z},\widetilde{\vartheta}_t C)} + \big\|\nabla f_t\big\|^2_{\mathbb{L}^2(\mathcal{Z},\widetilde{\vartheta}_t C)}\Big)\Big/ 2\,.
\end{aligned}
\tag{9.4}
$$

Once again, equality holds if and only if $\nabla f_t = \nabla\big(\varphi(\widetilde{\ell}_t)\big)$, and this leads to the backwards equation

$$
\partial\widetilde{\ell}_t = -\nabla\cdot\big(\widetilde{\vartheta}_t\, \nabla f_t\big) = \nabla\cdot\big(\widetilde{\vartheta}_t\, \nabla(\varphi(\widetilde{\ell}_t))\big) = \nabla\cdot\big(\nabla\widetilde{\ell}_t\big) = \mathcal{K}\,\widetilde{\ell}_t = \widehat{\mathcal{K}}\,\widetilde{\ell}_t
$$

of (3.9), by detailed balance. We have used here the elementary but crucial identity $\widetilde{\vartheta}_t\, \nabla(\varphi(\widetilde{\ell}_t)) = \nabla\widetilde{\ell}_t$, a "discrete chain rule" that sheds light on our choice of weight-function Θ^{Φ} in (8.11). But the last two norms displayed in (9.4) are $\|\nabla f_t\|_{\mathbb{L}^2(\mathcal{Z},\widetilde{\vartheta}_t C)} = \|\partial\widetilde{\ell}_t\|_{\mathbb{H}^{-1}_{\Theta}(\mathcal{S},\widetilde{\ell}_t Q)}$ and $\|\nabla\varphi(\widetilde{\ell}_t)\|_{\mathbb{L}^2(\mathcal{Z},\widetilde{\vartheta}_t C)} = \|\varphi(\widetilde{\ell}_t)\|_{\mathbb{H}^{1}_{\Theta}(\mathcal{S},\widetilde{\ell}_t Q)}$.

We summarize the situation in Theorem 9.2 below; this corresponds to Theorem 8.6 in the same manner as Theorem 9.1 corresponds to Theorem 6.3. Again, the DE GIORGI argument (9.4) gives not only the "direction of steepest descent" into which the gradient flow travels, but also the velocity of this flow.

Theorem 9.2. *For any given probability vector $P \in \mathcal{M}$, along any smooth curve $(\widetilde{P}_t)_{0\leq t<\infty}$ in $\mathcal{M}$ with $\widetilde{P}_0 = P$, and with $\boldsymbol{\ell} \in \mathcal{L}$ the likelihood ratio vector corresponding to P, we have*

$$
\Big(2\,\partial H^{\Phi}\big(\widetilde{P}_t\big|Q\big) + \big\|\partial\widetilde{\ell}_t\big\|^2_{\mathbb{H}^{-1}_{\Theta}(\mathcal{S},\boldsymbol{\ell}Q)}\Big)\Big|_{t=0} \;\geq\; -\,\big\|\varphi(\boldsymbol{\ell})\big\|^2_{\mathbb{H}^{1}_{\Theta}(\mathcal{S},\boldsymbol{\ell}Q)}\,.
$$

Equality holds here if, and only if, the curve $(\widetilde{P}_t)_{0\leq t<\infty} \subset \mathcal{M}$ satisfies the forward equation $\partial\widetilde{P}_t = \mathcal{K}'\widetilde{P}_t$ (equivalently, the likelihood ratio curve $(\widetilde{\ell}_t)_{0\leq t<\infty} \subset \mathcal{L}$ satisfies the backward equation $\partial\widetilde{\ell}_t = \mathcal{K}\,\widetilde{\ell}_t$, and the corresponding "driver" in (9.3) is $f_t = -\varphi(\widetilde{\ell}_t)$.)

9.3 A Riemannian Framework

Let us take up these same ideas again, but now in a Riemannian-geometric framework as for instance in MAAS (2011), MIELKE (2011). For any given probability vector $P \in \mathcal{M}$, we define the "likelihood ratio" vector $\boldsymbol{\ell} = \big(\ell(x)\big)_{x\in\mathcal{S}} \in \mathcal{L}$ with strictly positive elements $\ell(x) := p(x)/q(x)$. We then consider the Riemannian metric $(g_{\boldsymbol{\ell}})_{\boldsymbol{\ell}\in\mathcal{L}}$ on $\mathcal{L}$ induced by the scalar products

$$
g_{\boldsymbol{\ell}}(\partial\ell_1, \partial\ell_2) \;:=\; \big\langle \nabla\psi_1, \nabla\psi_2 \big\rangle_{\mathbb{L}^2(\mathcal{Z},\vartheta_\ell C)}\,,
$$

where $\nabla\psi_i$ is the unique discrete gradient satisfying the continuity equation $\partial\ell_i = \nabla\cdot(\vartheta_\ell\nabla\psi_i)$ for $i = 1, 2$. In other words, $g_{\boldsymbol{\ell}}(\partial\ell, \partial\ell) = \|\partial\ell\|^2_{\mathbb{H}^{-1}_{\Theta}(\mathcal{S},\boldsymbol{\ell}Q)}$.

The Riemannian gradient $\operatorname{grad} F$ of a smooth functional $F : \mathcal{L} \to \mathbb{R}$ is then given by

$$
\operatorname{grad} F = -\nabla\cdot\Big(\vartheta_\ell\,\nabla\frac{\delta F}{\delta\ell}\Big)
$$

where $\frac{\delta F}{\delta\ell}$ is the $\mathbb{L}^2(\mathcal{S}, Q)$-derivative defined by $\lim_{\varepsilon\to 0}\varepsilon^{-1}\big(F(\ell + \varepsilon\eta) - F(\ell)\big) = \langle\frac{\delta F}{\delta\ell}, \eta\rangle_{\mathbb{L}^2(\mathcal{S},Q)}$ for $\eta : \mathcal{S} \to \mathbb{R}$ with $\sum_{x\in\mathcal{X}} \eta(x)q(x) = 0$. In particular, the gradient flow equation $\partial\ell = -\operatorname{grad} F(\ell)$ reads

$$
\partial\ell = \nabla\cdot\Big(\vartheta_\ell\,\nabla\frac{\delta F}{\delta\ell}\Big)\,. \tag{9.5}
$$

The Riemannian metric g on $\mathcal{L}$ can of course be turned into a Riemannian metric G on the manifold of probability measures $\mathcal{M}$, via $G_P(\partial P_1, \partial P_2) := g_{\boldsymbol{\ell}}(\partial\ell_1, \partial\ell_2)$, where $P = \boldsymbol{\ell}\,Q$ and $\partial P_i = \partial\ell_i\,Q$ for $i = 1, 2$.

Theorem 9.3. **(MAAS (2011), MIELKE (2011)):** *Under the detailed balance conditions (3.11), and with Θ the function of (8.11), the Forward* KOLMOGOROV *equation $\partial P(t) = \mathcal{K}' P(t)$ in (3.5) is the gradient flow of the $\Phi-$relative entropy in (9.2) with respect to the Riemannian metric induced on the manifold $\mathcal{M}$ as above.*

Proof: Let $(P(t))_{0 \le t < \infty}$ solve the Forward KOLMOGOROV equation $\partial P(t) = \mathcal{K}' P(t)$. By detailed balance, the associated likelihood ratio curve $\big(\boldsymbol{\ell}(t)\big)_{0 \le t < \infty} \subset \mathcal{L}$ satisfies the backward equation $\partial \boldsymbol{\ell}(t) = \mathcal{K}\, \boldsymbol{\ell}(t)$. In view of (9.5), we thus need to verify the identity

$$\mathcal{K}\, \ell = \nabla \cdot \Big(\vartheta_\ell \, \nabla \frac{\delta h^\Phi}{\delta \ell} \Big) \,,$$

where $h^\Phi : \mathcal{L} \to \mathbb{R}$ is defined by $h^\Phi(\ell) = H^\Phi(\ell Q | Q)$.

For $\boldsymbol{\ell} \in \mathcal{L}$ and $\eta : \mathcal{S} \to \mathbb{R}$ with $\sum_{x \in \mathcal{S}} \eta(x) q(x) = 0$, we have the directional derivative computation

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\, h^\Phi\big(\boldsymbol{\ell} + \varepsilon \eta\,\big) \bigg|_{\varepsilon = 0} = \sum_{x \in \mathcal{S}} \eta(x)\, \varphi\big(\ell(x)\big); \qquad \text{thus} \quad \frac{\delta h^\Phi}{\delta \ell} = \varphi(\boldsymbol{\ell}) := \big(\varphi(\ell(x))\big)_{x \in \mathcal{S}}. \tag{9.6}$$

Invoking the "discrete chain rule" $\vartheta_\ell \, \nabla(\varphi(\boldsymbol{\ell})) = \nabla \ell$ we obtain

$$\nabla \cdot \Big(\vartheta_\ell \, \nabla \frac{\delta h^\Phi}{\delta \ell} \Big) = \nabla \cdot \Big(\vartheta_\ell \nabla(\varphi(\boldsymbol{\ell})) \Big) = \nabla \cdot (\nabla \ell) = \mathcal{K} \ell \,,$$

which is the desired identity. □

Theorem 9.3 has a converse, developed in DIETERT (2015) as follows.

Proposition 9.4. *Suppose that there exists a $\mathcal{C}^1$ Riemannian metric on the manifold of probability vectors $\mathcal{M}$, under which the Forward* KOLMOGOROV *equation $\partial P(t) = \mathcal{K}' P(t)$ of (3.5) is the gradient flow for the relative entropy in (7.1). Then the* MARKOV *Chain satisfies the detailed balance conditions (3.11).*

10 Countable State-Space

It is well known that the results of Sections 2 and 3 hold also for countably infinite state-spaces $\mathcal{S}$; see Chapters 2, 3 in NORRIS (1997) and LIGGETT (2010). In particular, the ergodic property (3.7) holds at least for bounded functions $f : \mathcal{S} \to \mathbb{R}$. The crucial Proposition 5.2 also remains valid.

Propositions 6.1, 6.2 carry over to countable state-spaces under the assumption $V\big(P(0)\,|\,Q\big) < \infty$. To see this, we start by observing that we can guarantee now *prima facie* only the local martingale property of the processes $\widehat{M}$ in (6.2). Still, we can localize $\widehat{M}$ by an increasing sequence $\{\sigma_n\}_{n \in \mathbb{N}}$ of $\widehat{\mathbb{G}}-$stopping-times with values in $[0, T]$ and $\lim_{n \to \infty} \uparrow \sigma_n = T$, and create the bounded $(\widehat{\mathbb{G}}, \mathbb{Q})-$martingales $\widehat{M}(s \wedge \sigma_n)$, $0 \le s \le T$. Taking expectations in (6.2) with $s = \sigma_n$, then letting $n \to \infty$ and using monotone convergence, the $\mathbb{Q}-$submartingale property of $\ell^2\big(T - s, \widehat{X}(s)\big)$, $0 \le s \le T$ from Proposition 5.2, and optional sampling, we obtain from (4.8) the inequality

$$\mathbb{E}^{\mathbb{Q}}\big[\ell^2\big(T, X(T)\big)\big] + \int_0^T 2\, \mathcal{E}\big(\ell_t\,, \ell_t\,\big)\, \mathrm{d}t \;=\; \lim_{n \to \infty} \uparrow \mathbb{E}^{\mathbb{Q}}\big[\ell^2\big(T - \sigma_n, \widehat{X}(\sigma_n)\big)\big] \;\le\; \mathbb{E}^{\mathbb{Q}}\big[\ell^2\big(0, X(0)\big)\big].$$

But the reverse of this last inequality also holds, on account of FATOU's Lemma; thus (6.7) follows for countable state-spaces as well, and $\widehat{M}$ is seen to be a true $(\widehat{\mathbb{G}}, \mathbb{Q})-$martingale. Then $\lim_{t \to \infty} V\big(P(t)\,|\,Q\big) = 0$, and with it (6.8), are proved for a countable state-space in the manner of Proposition 10.1 below.

10.1 Relative Entropy Dissipates all the way down to Zero

When the state-spaces $\mathcal{S}$ is countably infinite, the results of Section 7 pertaining to the relative entropy need the additional assumption

$$H\big(P(0)\big|Q\big) = \sum_{y \in \mathcal{S}} p(0,y)\, \log\left(\frac{p(0,y)}{q(y)}\right) < \infty\,. \tag{10.1}$$

Then everything goes through as before, including the non-negativity and decrease claims in (7.3) — except for the argument establishing (7.9), which uses the finiteness of the state-space in a crucial manner.

Here is an argument for this result in the countable case.

Proposition 10.1. *The dissipation of relative entropy all the way down to zero, as in (7.9), holds in the context of a countable state-space under the condition (10.1).*

Proof: Let us recall the likelihood ratio process $L(t) := \ell\big(t, X(t)\big)$, $0 \le t < \infty$ of (5.4), and from (5.8) that its time-reversal $L(T-s),\ 0 \le s \le T$ is a $\big(\widehat{\mathbb{G}}, \mathbb{Q}\big)-$martingale.

Fix $0 \le t_1 < t_2 < \infty$. For any $T \in (t_2, \infty)$, this means $\mathbb{E}^{\mathbb{Q}}\big[L(T-s_1)\,\big|\,\mathcal{G}(T-s_2)\big] = L(T-s_2)$ for $s_1 = T-t_1,\ s_2 = T-t_2$, or equivalently:

$$\mathbb{E}^{\mathbb{Q}}\big[L(t_1)\,\big|\,\sigma\big(X(\theta),\ t_2 \le \theta \le T\big)\big] \;=\; L(t_2)\,.$$

But this last identity holds for any $T \in (t_2, \infty)$, so it leads — on the strength of the P. LÉVY martingale convergence Theorem 9.4.8 in CHUNG (1974) — to

$$\mathbb{E}^{\mathbb{Q}}\big[L(t_1)\,\big|\,\mathcal{H}(t_2)\big] = L(t_2)\,, \qquad \mathcal{H}(t) := \sigma\big(X(\theta),\ t \le \theta < \infty\big). \tag{10.2}$$

To wit, the likelihood ratio process $\big(L(t)\big)_{0 \le t < \infty}$ is a martingale of the backwards filtration $\big(\mathcal{H}(t)\big)_{0 \le t < \infty}$, whose "tail" sigma-algebra is trivial on account of the ergodicity property (3.7) of the MARKOV Chain (BLACKWELL & FREEDMAN (1964)):

$$\mathcal{H}(\infty) \;:=\; \bigcap_{0 \le t < \infty} \mathcal{H}(t) \;=\; \big\{\emptyset, \Omega\big\}\,, \quad \text{mod. } \mathbb{Q}\,.$$

We invoke now the martingale version of the backward submartingale convergence Theorem 9.4.7 in CHUNG (1974). It follows from this result that $\big(L(t)\big)_{0 \le t < \infty}$ is a $\mathbb{Q}-$uniformly integrable family; that the limit $L(\infty) := \lim_{t \to \infty} L(t)$ exists, both a.e. and in $\mathbb{L}^1$ under $\mathbb{Q}$; and that the backward martingale property (10.2) extends all the way to infinity, namely

$$\mathbb{E}^{\mathbb{Q}}\big[L(t_1)\,\big|\,\mathcal{H}(\infty)\big] = L(\infty)\,. \tag{10.3}$$

But the triviality under $\mathbb{Q}$ of the tail sigma-algebra, implies that $L(\infty)$ is $\mathbb{Q}-$a.e. constant. Then the extended martingale property (10.3) identifies this constant as $L(\infty) = \mathbb{E}^{\mathbb{Q}}\big[L(\infty)\big] = \mathbb{E}^{\mathbb{Q}}\big[L(t_1)\big] = 1$.

We recall the relative entropy from (7.5). The convexity of the function $\Phi(\ell) = \ell \log \ell$ shows, in conjunction with (10.2) and the JENSEN inequality, that

$$\Big(\Phi\big(L(t)\big), \mathcal{H}(t)\Big)_{0 \le t < \infty} \qquad \text{is a backward } \mathbb{Q}-\text{submartingale}, \tag{10.4}$$

with decreasing expectation $\mathbb{E}^{\mathbb{Q}}\big[\Phi\big(L(t)\big] = H\big(P(t)\,\big|\,Q\big) \ge 0$. Because this expectation is bounded from below, we can appeal once again to the backward submartingale convergence Theorem 9.4.7 in CHUNG (1974). We deduce that the process in (10.4) is a $\mathbb{Q}-$uniformly integrable family which converges, again both a.e. and in $\mathbb{L}^1$ under $\mathbb{Q}$, to $\lim_{t \to \infty} \Phi\big(L(t)\big) = \Phi\big(L(\infty)\big) = \Phi(1) = 0$.

Furthermore, the aforementioned uniform integrability gives

$$\lim_{t \to \infty} \downarrow H\big(P(t)\,\big|\,Q\big) \;=\; \lim_{t \to \infty} \mathbb{E}^{\mathbb{Q}}\big[\Phi\big(L(t)\big)\big] \;=\; \mathbb{E}^{\mathbb{Q}}\Big(\lim_{t \to \infty} \Phi\big(L(t)\big)\Big) \;=\; 0\,;$$

that is, (7.9) is also valid in this general case with countable state-space. □

10.1.1 Relative Entropy is Continuous at the Origin

We discuss now the validity of the DE BRUIJN identities of (7.18) when the state-space is countable.

Proposition 10.2. *The* DE BRUIJN *identities of (7.18) for the dissipation of relative entropy are valid in the case of a countable state-space, under the condition (10.1).*

To justify this claim, we would like to use the argument already deployed; but there is now no obvious, general way to turn the local martingale $\widehat{M}^h$ of (7.11) into a true $\mathbb{Q}-$martingale. Thus, we localize $\widehat{M}^h$ by an increasing sequence $\{\sigma_n\}_{n\in\mathbb{N}}$ of $\widehat{\mathbb{G}}-$stopping-times with values in $[0,T]$ and $\lim_{n\to\infty}\uparrow\sigma_n = T$. In this manner we create the *bounded* $(\widehat{\mathbb{G}},\mathbb{Q})-$martingales $\widehat{M}^h(s\wedge\sigma_n)\,,\ 0\le s\le T$, which then give

$$\mathbb{E}^{\mathbb{Q}}\int_0^{\sigma_n}\big(\partial h+\widehat{\mathcal{K}}h\big)\big(u,\widehat{X}(u)\big)\,\mathrm{d}u \;=\; \mathbb{E}^{\mathbb{Q}}\big[h\big(\sigma_n,\widehat{X}(\sigma_n)\big)\big]-\mathbb{E}^{\mathbb{Q}}\big[h\big(0,\widehat{X}(0)\big)\big] \tag{10.5}$$

$$= H\big(P(T-\sigma_n)\,\big|\,Q\big)-H\big(P(T)\,\big|\,Q\big)\;\le\;H\big(P(0)\,\big|\,Q\big)-H\big(P(T)\,\big|\,Q\big)\;\le\;H\big(P(0)\,\big|\,Q\big)\;<\;\infty$$

for every $n\in\mathbb{N}$, on account of (10.1); see also the argument straddling (10.8) below. In particular, the sequence of real numbers in (10.5) takes values in the compact interval $[-H(P(0)\,|\,Q),H(P(0)\,|\,Q)]$.

We would like now to let $n\to\infty$ in (10.5), and establish the DE BRUIJN identity (7.18) in this case. The issue once again is continuity of the relative entropy — though now *at the origin* (rather than at infinity, as in (7.9)); and not along fixed times, but rather along an appropriate sequence of stopping times, i.e.,

$$\lim_{n\to\infty}\uparrow H\big(P(T-\sigma_n)\,\big|\,Q\big)\;=\;H\big(P(0)\,\big|\,Q\big)\,. \tag{10.6}$$

Accepting this for a moment, and letting $n\to\infty$ in (10.5), we obtain the DE BRUIJN identity (7.18), i.e.,

$$\int_0^T I(t)\,\mathrm{d}t\;=\;\mathbb{E}^{\mathbb{Q}}\int_0^T\Big(\partial h+\widehat{\mathcal{K}}h\Big)\big(u,\widehat{X}(u)\big)\,\mathrm{d}u\;=\;H\big(P(0)\,\big|\,Q\big)-H\big(P(T)\,\big|\,Q\big) \tag{10.7}$$

by monotone convergence. We let now $T\to\infty$ in (10.7) and arrive at the second identity in (7.18), thanks to the property (7.9) already established in Proposition 10.1.

Proof of (10.6): By analogy with (7.8), and invoking now additionally the optional sampling theorem for the bounded stopping times $\{\sigma_n\}_{n\in\mathbb{N}}$ of $\widehat{\mathbb{G}}$ with values in $[0,T]$, we deduce that the sequence of non-negative real numbers

$$H\big(P(T-\sigma_n)\,|\,Q\big)\;=\;\mathbb{E}^{\mathbb{Q}}\big[\Phi\big(\ell\big(T-\sigma_n,\widehat{X}(\sigma_n)\big)\big)\big]\,,\qquad n\in\mathbb{N} \tag{10.8}$$

is increasing; in particular, $\lim_{n\to\infty}H\big(P(T-\sigma_n)\,\big|\,Q\big)\le H\big(P(0)\,\big|\,Q\big)$. On the other hand, the boundedness-from-below of the function $\Phi(\ell)=\ell\,\log\ell$ gives

$$\lim_{n\to\infty}H\big(P(T-\sigma_n)\,\big|\,Q\big)\;\ge\;\mathbb{E}^{\mathbb{Q}}\Big[\lim_{n\to\infty}\Phi\big(\ell\big(T-\sigma_n,\widehat{X}(\sigma_n)\big)\big)\Big]\;=\;\mathbb{E}^{\mathbb{Q}}\big[\Phi\big(\ell\big(0,X(0)\big)\big)\big]\;=\;H\big(P(0)\,\big|\,Q\big)$$

with the help of FATOU's Lemma, and (10.6) follows. □

Remark 10.1. *The General Case:* Exacly the same methods show that the results of Propositions 8.1 and 8.2, pertaining to a general convex function $\Phi:(0,\infty)\to\mathbb{R}$ with the properties imposed there, continue to hold for the generalized relative entropy functional of (9.2) in the case of a countable state-space $\mathcal{S}$, under the condition $H^{\Phi}\big(P(0)|Q\big)<\infty$.

Once again, it is important to stress that nowhere in the present Section up to this point, have we invoked the detailed-balance conditions of (3.11).

Bibliography

AMBROSIO, L., GIGLI, N. & SAVARÉ, G. (2008) *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Second Edition. *Lectures in Mathematics, ETH Zürich.* Birhäuser Verlag, Basel.

BLACKWELL, D. & FREEDMAN, D. (1964) The tail $\sigma-$field of a Markov chain and a theorem of Orey. *Ann. Math. Statist.* **35**, 1291-1295.

BOBKOV, S.G. & TETALI, P. (2006) Modified logarithmic Sobolev inequalities in discrete settings. *Journal of Theoretical Probability* **19**, 289-336.

CAPUTO, P., DAI PRA, P. & POSTA, G. (2009) Convex entropy decay via the Bochner-Bakry-Émery approach. *Ann. Inst. Henri Poincaré (Sér. B, Probabilités et Statistiques)* **45**, 734-753.

CHAFAÏ, D. (2004) Entropies, convexity and functional inequalities. *J. Math. Kyôto University* **42**, 325-363.

CHUNG, K.L. (1974) *A Course in Probability Theory.* Second Edition, Academic Press, New York.

CONFORTI, G. (2020) A probabilistic approach to convex (ϕ)-entropy decay for Markov chains. Preprint, available at *https://arxiv.org/abs/2004.10850*

COURANT, R., FRIEDRICHS, K. & LEWY, H. (1928) Über die partiellen Differenzellengleichungen der mathematiscen Physik. *Math. Annalen* **100**, 32-74.

COVER, T.M. & THOMAS, J.A. (1991) *Elements of Information Theory.* J. Wiley & Sons, New York.

DIACONIS, P. & SALOFF-COSTE, L. (1996) Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability* **6**, 695-750.

DIETERT, H. (1996) Characterization of gradient flows for finite-state Markov chains. *Electronic Communications in Probability* **20**, no. 29, 1-8.

ERBAR, M. & MAAS, J. (2012) Ricci curvature of finite Markov Chains via convexity of entropy. *Archive for Rational Mechanics and Analysis* **206**, 997-1038.

ERBAR, M. & MAAS, J. (2014) Gradient flow structures for discrete porous medium equations. *Discrete and Continuous Dynamical Systems* **34**, 1355-1374.

FONTBONA, J. & JOURDAIN, B. (2016) A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations. *The Annals of Probability* **44**, 131-170.

JORDAN, R., KINDERLEHRER, D. & OTTO, F. (1998) The variational formulation of the Fokker-Planck equation. *SIAM Journal of Mathematical Analysis* **29**, 1-17.

KARATZAS, I., SCHACHERMAYER, W. & TSCHIDERER, B. (2019) Trajectorial Otto Calculus. Preprint, available at *https://arxiv.org/abs/1811.08686*

KARATZAS, I. & SHREVE, S.E. (1988) *Brownian Motion and Stochastic Calculus.* Volume 113 of series *Graduate Texts in Mathematics.* Springer Verlag, New York.

LIGGETT, T.G. (2010) *Continuous Time Markov Processes: An Introduction.* Volume 113 of the series *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI.

MAAS, J. (2011) Gradient flows of the entropy for finite Markov chains. *Journal of Functional Analysis* **261**, 2250-2292.

MAAS, J. (2017) Entropic Ricci curvature for discrete spaces. *Lecture Notes in Mathematics* **2184**, 159-173. Springer-Verlag, Berlin.

MIELKE, A. (2011) A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity* **24**, 1329-1346.

MIELKE, A. (2013) Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations* **48**, 1-31.

MIELKE, A. (2016) On evolutionary $\Gamma-$convergence for gradient systems. In *Macroscopic and Large Scale Phenomena: Coarse Graining, Mean-Field Limits, and Ergodicity*. Lecture Notes in Applied Mathematics and Mechanics **3**, 187-249. Springer- Verlag.

MONTENEGRO, R. & TETALI, P. (2006) Mathematical aspects of mixing times in Markov chains. *Foundations and Trends in Theoretical Computer Science* **1** (3), 237-354.

NORRIS, J. (1997) *Markov Chains.* Cambridge University Press.

OTTO, F. (2001) The geometry of dissipative evolution equatioins: the porous medium equation. *Communications in Partial Differential Equations* **26**, 101-174.

ROGERS, L.C.G. & WILLIAMS, D. (1987) *Diffusions, Markov Processes and Martingales. Vol. II: Itô Calculus.* J. Wiley & Sons, New York.

STAM, A.J. (1959) Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* **2**, 101-112.