On extremals of the entropy production by ‘Langevin–Kramers’ dynamics

Paolo Muratore-Ginanneschi
Department of Mathematics and Statistics, University of Helsinki, PL 68, FI-00014, Finland
E-mail: paolo.muratore-ginanneschi@helsinki.fi

Received 15 January 2014
Accepted for publication 26 March 2014
Published 16 May 2014

Online at stacks.iop.org/JSTAT/2014/P05013
doi:10.1088/1742-5468/2014/05/P05013

Abstract. We refer as ‘Langevin–Kramers’ dynamics to a class of stochastic differential systems exhibiting a degenerate ‘metriplectic’ structure. This means that the drift field can be decomposed into a symplectic and a gradient-like component with respect to a pseudo-metric tensor associated with random fluctuations affecting increments of only a sub-set of the degrees of freedom. Systems in this class are often encountered in applications as elementary models of Hamiltonian dynamics in a heat bath eventually relaxing to a Boltzmann steady state.

Entropy production control in Langevin–Kramers models differs from the now well-understood case of Langevin–Smoluchowski dynamics for two reasons. First, the definition of entropy production stemming from fluctuation theorems specifies a cost functional which does not act coercively on all degrees of freedom of control protocols. Second, the presence of a symplectic structure imposes a non-local constraint on the class of admissible controls. Using Pontryagin control theory and restricting the attention to additive noise, we show that smooth protocols attaining extremal values of the entropy production appear generically in continuous parametric families as a consequence of a trade-off between smoothness of the admissible protocols and non-coercivity of the cost functional. Uniqueness is, however, always recovered in the over-damped limit as extremal equations reduce at leading order to the Monge–Ampère–Kantorovich optimal mass-transport equations.

Keywords: dynamical processes (theory), fluctuations (theory), stochastic processes (theory), diffusion
1. Introduction

The contrivance and development of techniques that can be used to investigate the physics of very small systems is currently attracting great interest [1]. Examples of very small systems are bio-molecular machines consisting of a few molecules or, in some cases,
The structure of the paper is as follows. In section 2 we describe the kinematic properties of Langevin–Kramers diffusion. We also define the class $A$ of admissible Hamiltonians governing the Langevin–Kramers dynamics to which we restrict our attention while considering the optimal control problem for the entropy production. From the mathematical slant, it is obvious that an optimal control problem is well posed if we assign besides the ‘cost functional’ to be minimized, the functional space of admissible controls. From the physics point of view, our aim is to explicitly restrict the attention to ‘macroscopic’ control protocols modeled by smooth Hamiltonians acting on ‘slower’ timescales as opposed to configurational degrees of freedom subject to Brownian forces and fluctuating at the fastest timescales in the model [12] (see also discussion in [11]).

In section 3 we briefly recall the stochastic thermodynamics of Langevin–Kramers diffusion, drawing on [13, 14]. The main result is the expression of the entropy production
On extremals of the entropy production by ‘Langevin–Kramers’ dynamics

$\mathcal{E}_{t_f, t_o}$ over a finite control horizon $[t_o, t_f]$ in terms of the current velocity of the Langevin–Kramers diffusion process. As in the Langevin–Smoluchowski case $[15, 9, 16]$, the current velocity parametrization plays a substantive role in unveiling the properties of the entropy production from both the thermodynamic and the control point of view. In the Langevin–Kramers case, the entropy production turns out to be a non-coercive $[17]$ cost functional. Namely, the decomposition of the current velocity into dissipative and symplectic components evinces that the entropy production is in fact a quadratic functional of the dissipative component alone. The origin of the phenomenon is better understood by revisiting the probabilistic interpretation of the entropy production, which we do in section 4. This section is mostly meant as a service to readers not familiar with $[18, 19, 15, 20]$, as it differs from these references only insofar as we base our discussion on backward-in-time stochastic evolution, as customary in stochastic mechanics $[21, 22, 23]$. Namely, in section 4 we recall that $\mathcal{E}_{t_f, t_o}$ coincides with the Kullback–Leibler divergence $\mathbb{K}(P_\chi || P_{\tilde{\chi}}) [24]$ between the probability measure $P_\chi$ of the primary Langevin–Kramers process $\chi$ and the probability measure $P_{\tilde{\chi}}$ of a process $\tilde{\chi}$ obtained from the former by inverting the sign of the dissipative component of the drift and evolving in the opposite time direction. The Kullback–Leibler divergence is a relative entropy measuring the information loss occasioned when $P_{\tilde{\chi}}$ is used to approximate $P_\chi$. These facts substantiate, on the one hand, the identification of the entropy production as a natural indicator of the irreversibility of a physical process and, as such, as the embodiment of the second law of thermodynamics. On the other hand, they pinpoint that the interpretation of the entropy production as a Kullback–Leibler divergence is possible in the Kramers–Langevin case only by applying in the construction of the auxiliary process $\tilde{\chi}$ a different time-reversal operation from that in the Langevin–Smoluchowski case. Non-coercivity is the result of such a time-reversal operation.

The consequences of non-coercivity of the cost functional on the optimal control of entropy production are, however, tempered by the regularity requirements imposed on the class $\mathcal{A}$ of admissible Hamiltonians. Simple considerations (section 5) based on smoothness of the evolution show that the Langevin–Kramers entropy production must be bounded from below by the entropy production generated by an optimally controlled Langevin–Smoluchowski diffusion connecting in the same horizon the configuration space marginals of the initial and final phase–space probability densities. This result motivates the analysis of section 6, where we avail ourselves of Pontryagin’s maximum principle to directly investigate extremals of the entropy production in a finite-time transition between assigned states. Pontryagin’s maximum principle is formulated in terms of Lagrange multipliers acting as conjugate ‘momentum’ variables (see e.g. $[25, 26, 27]$). We can therefore construe it as an ‘Hamiltonian formulation’ of Bellman’s optimal control theory, which is based upon dynamic programing equations (see e.g. $[17, 28, 29]$). Relying on Pontryagin’s maximum principle we conveniently arrive at the first main finding of the present paper, encapsulated in the extremal equations (48).

On the space $\mathcal{A}$ of admissible control Hamiltonians the entropy production generically attains a highly degenerate minimum value. A distinctive feature of the extremal equations (48) is that the coupling between the dynamic programing equation and the Fokker–Planck equation takes the non-local form of a third auxiliary equation. We attribute the occurrence of a non-local coupling to the divergenceless component of the Langevin–Kramers drift.

doi:10.1088/1742-5468/2014/05/P05013
On extremals of the entropy production by ‘Langevin–Kramers’ dynamics

We illustrate these results in section 7, where we consider the explicitly solvable case of Gaussian statistics. While considering this example, we also investigate the recovery in the over-damped limit of the expression of the minimal entropy production by Langevin–Smoluchowski diffusion, a problem to which we systematically turn in section 8. There we derive our second main result: upon applying a multi-scale (homogenization) asymptotic analysis [30, 31] we show that the cell problem associated with the extremal equations (48) takes the form of a Monge–Ampère–Kantorovich mass-transport problem [32] for configuration space marginals of the phase–space probability densities. The noteworthy aspect of this result is that the degeneracy of the extremals (48) does not appear in the cell problem, therefore consistency with the results obtained for Langevin–Smoluchowski diffusion [7] is guaranteed.

Finally, the section 9 is devoted to a discussion concerning the existence of singular control strategies, which we explicitly ruled out while deriving the extremal equations (48).

2. From kinematics to dynamics

We consider a phase–space dynamics governed by

\[
d\chi_t = (J - G) \cdot \partial_{\chi_t} H \frac{dt}{\tau} + \sqrt{\frac{2}{\beta \tau}} G^{1/2} \cdot d\omega_t,
\]

\[
P(x \leq \chi_{t_0} < x + d\chi) = m_0(x) d^{2d}x.
\]

In (1a) \(\omega = \{\omega_t, t \geq t_0\}\) denotes an \(\mathbb{R}^{2d}\)-valued Wiener process and \(\partial_{\chi_t} H\) is a shortened form of \(\partial_x H(\chi_t, t)\), with \(H\) a time-dependent Hamilton function. \(G\) and \(J\) are the \(2d \times 2d\) constant matrices

\[
G = \begin{bmatrix} 0 & 0 \\ 0 & 1_d \end{bmatrix} \quad \text{and} \quad J = \begin{bmatrix} 0 & 1_d \\ -1_d & 0 \end{bmatrix}
\]

where \(1_d\) stands for the identity in \(d\)-dimensions. We notice that \(G\) has the geometric interpretation of a vertical projector in phase space. Finally, \(\beta\) and \(\tau\) are positive definite constants. We attribute to \(\beta\) the physical interpretation of the inverse of the temperature and to \(\tau\) that of the characteristic timescale of the system. We will measure any other quantity encountered throughout the paper in units of \(\beta\) and \(\tau\).

The kinematics in (1a) satisfies the conditions required by the Hörmander theorem to prove that for any sufficiently regular, bounded from below, and growing sufficiently fast at infinity Hamilton function \(H\), the process \(\chi = \{\chi_t, t \in [t_0, t_f]\}\) admits a smooth transition probability density notwithstanding the degenerate form of the noise (see e.g. [33]). Furthermore, if \(H\) is time independent, it is straightforward to verify that the measure relaxes to a steady state such that

\[
P(x \leq \chi_\infty < x + d\chi) = \beta^d e^{\beta[F - H(x)]} d^{2d}x
\]

\[
F \equiv -\frac{1}{\beta} \ln \int_{\mathbb{R}^{2d}} d^{2d}x \beta^d e^{-\beta H(x)}.
\]
On extremals of the entropy production by 'Langevin–Kramers' dynamics

The choice of the normalization constant in (3a) befits the interpretation of $F$ as equilibrium free energy. For any finite time, we find it expedient to write the probability density of the system in the form

$$P(x \leq \chi_t < x + dx) = m(x, t)d^{2d}x \equiv \beta^d e^{-S(x, t)}d^{2d}x.$$  

(4)

We will refer to the non-dimensional function $S$ as the microscopic entropy of the system, inasmuch as it specifies the amount of information required to describe the state of the system given that the state occurs with probability (4) [34]. The average variation of $S$ with respect to the measure of $\chi$, which we denote by $E(\chi)$,

$$(S_{t_2} - S_{t_1}) = E(\chi)[S(\chi_{t_2}, t_2) - S(\chi_{t_1}, t_1)] \quad \forall t_2 \geq t_1 \geq 0$$  

(5)

specifies the variation of the Shannon–Gibbs entropy of the system. The representation (4) of the probability density establishes an elementary link between the kinematics and the thermodynamics of the system. In order to describe the dynamics, we need to specify the Hamiltonian $H$. Our aim here is to determine $H$ by solving an optimal control problem associated with the minimization of a certain thermodynamic functional, the entropy production, during a transition evolving the initial state (1b) into a final state

$$P(x \leq \chi_0 < x + dx) = m_f(x)d^{2d}x$$  

(6)

in a finite-time horizon $[t_0, t_f]$. From this slant, we need to regard $H$ as the element of a class $A$ of admissible controls satisfying the following requirements.

(i) $H$ must be sufficiently smooth that there exists a unique transition probability density belonging to $C^{(2,1)}(\mathbb{R}^{2d}, [t_0, t_f])$ solving in weak sense (1a). Among the conditions implying this requirement we include that $H$ is itself twice differentiable in space and once in time (see e.g. chapter V of [25] for a precise statement of the set of sufficient conditions). In section 6.1 we use the assumption $A \subseteq C^{(2,1)}(\mathbb{R}^{2d}, [t_0, t_f])$ to derive the equations satisfied by extremal controls of the entropy production.

(ii) The gradients $\partial_x H$, $\partial_x S$ belong for all $t \in [t_0, t_f]$ to $L^2(\mathbb{R}^{2d}, m^{2d}dx)$.1 In other words, they must be square integrable with respect to the probability density of the process (1). As will become clear in section 3, this requirement is a self-consistence condition for the control problem we set out to solve.

2.1. Generator of the process and 'metriplectic' structure

The scalar generator $\mathcal{L}$ of (1) acts on any differentiable phase–space function $f$ as

$$(\mathcal{L}f)(x, t) \equiv \left\{ [(J - G) \cdot \partial_x H(x, t)] \cdot \partial_x + \frac{1}{\beta}G \cdot \partial_x \otimes \partial_x \right\} f(x, t).$$  

(7)

Let us use (7) to explain our conventions for scalar products. The symbol ‘·’ means index contraction by the identity matrix both in $\mathbb{R}^d$ and $\mathbb{R}^{2d}$. We will use the shortened form

$$\|f\|_M^2 \equiv f \cdot M \cdot f$$  

(8)

1 We reserve the simpler notation $L^2(\mathbb{R}^{2d})$ for the space of functions square integrable with respect to the Lebesgue measure.
for quadratic forms of $f$ in $\mathbb{R}^d$ or $\mathbb{R}^{2d}$ with respect to any symmetric matrix $M$ and omitting the subscript if $M$ is the identity. Finally,

$$A : B \equiv \text{Tr} A^\dagger B$$

(9)

stands for for the scalar product between matrices.

The generator (7) admits the decomposition into the sum $\mathcal{L} = \mathcal{L}^- + \mathcal{L}^+$ of a conservative part

$$\mathcal{L}^- f = (\mathbf{J} \cdot \partial_x H) \cdot \partial_x f \equiv (\partial_p H) \cdot \partial_q f - (\partial_q H) \cdot \partial_p f$$

(10)

and a dissipative part

$$\mathcal{L}^+ f = -(G \cdot \partial_x H) \cdot \partial_x f + \frac{1}{\beta} G : \partial_x \otimes \partial_x f.$$  

(11)

Namely, the definition (10) states that $\mathcal{L}^- f$ is a symplectic form specified by the Poisson brackets in a Darboux chart between the Hamilton function $H$ and $f$. The differential operation $\mathcal{L}^+$ describes instead dissipation by a deterministic friction mechanism and by thermal interactions respectively encapsulated in the pseudo-metric form $(G \cdot \partial_x H) \cdot \partial_x f$ and in the second-order differential term. In the analytic mechanics literature it is customary to refer to systems whose generator comprises a symplectic and a (pseudo-)metric form as ‘metriplectic’ see e.g. [35, 36].

The $L^{(2)}(\mathbb{R}^{2d})$ adjoint of $\mathcal{L}$ with respect to the Lebesgue measure governs the evolution of the probability density $\mathbf{m}$ of the system. The anti-symmetry of the Poisson brackets yields

$$\mathcal{L}^\dagger f = -\mathcal{L}^- f + (\mathcal{L}^+)\dagger f$$

(12)

with $(\mathcal{L}^+)\dagger$ the $L^{(2)}(\mathbb{R}^{2d})$-adjoint of (11).

3. Thermodynamic functionals

Following [13], we identify the heat released during individual realizations of $\chi$ with the Stratonovich stochastic integral

$$Q_{t_f,t_o} = -\int_{t_o}^{t_f} \mathbf{d} \chi_t \cdot \frac{1}{2} \partial_{\chi_t} H.$$  

(13)

The $1/2$ betokens Stratonovich’s mid-point convention. If in conjunction with (13) we define the work as

$$W_{t_f,t_o} = \int_{t_o}^{t_f} \mathbf{d} t \partial_t H$$

(14)

we recover the first law of thermodynamics in the form

$$(W - Q)_{t_f,t_o} = \int_{t_o}^{t_f} \left( \mathbf{d} t \partial_t H + \mathbf{d} \chi_t \cdot \frac{1}{2} \partial_{\chi_t} H \right) = H_{t_f} - H_{t_o}.$$  

(15)
As our working hypotheses allow us to perform integrations by parts without generating boundary terms, the definition of the Stratonovich integral yields (see e.g. [21] p 33) the equality

\[
E^{(x)} \int_{t_0}^{t_f} \frac{dt}{\tau} \mathbf{v} \cdot \partial \mathbf{X}_t H = E^{(x)} \int_{t_0}^{t_f} \frac{dt}{\tau} \mathbf{v} \cdot \partial \mathbf{X}_t H
\]

for

\[
\mathbf{v} = J \cdot \partial \mathbf{X}_t H - G \cdot \partial \mathbf{X}_t \left( H - \frac{1}{\beta} S \right)
\]

the current velocity [37] (see also appendix A) and \( S \) the microscopic entropy (4). Upon inserting (17) into (16), after straightforward algebra, we arrive at

\[
\frac{\mathcal{E}_{t_f,t_0}}{\beta} \equiv E^{(x)} \left\{ Q_{t_f,t_0} + \frac{S_{t_f} - S_{t_0}}{\beta} \right\} = E^{(x)} \int_{t_0}^{t_f} \frac{dt}{\tau} \| \partial \mathbf{X}_t A \|^2_G \geq 0.
\]

We interpret the phase–space function

\[
A = H - \frac{1}{\beta} S
\]

as the ‘non-equilibrium’ Helmholtz energy density’ of the system and the non-dimensional quantity \( \mathcal{E} \) as the entropy production during the transition (see e.g. [19, 15, 20]). The interpretation is upheld by observing that the entropy production rate, \( E^{(x)} \| \partial \mathbf{X}_t A \|^2_G \), is a positive definite quantity generically vanishing only at equilibrium. On this basis, we regard the inequality (18) as the embodiment of the second law of thermodynamics. Furthermore, the positive definiteness of the entropy production yields a Jarzynski-type [38] bound for the mean work

\[
E^{(x)} W_{t_f,t_0} = E^{(x)} \left( A_{t_f} - A_{t_0} \right) + \frac{1}{\beta} \mathcal{E}_{t_f,t_0} \geq E^{(x)} \left( A_{t_f} - A_{t_0} \right).
\]

4. Probabilistic interpretation of the thermodynamic functionals

The entropy production (18) admits an intrinsic information theoretic interpretation as a quantifier of the irreversibility of a transition. Namely it coincides with the Kullback–Leibler divergence between the measure of the process (1) and that of the backward-in-time diffusion process \( \tilde{\mathbf{X}} = \{ \tilde{\mathbf{X}}_t; t \in [t_0, t_f] \} \) obtained by reversing the sign of the dissipative component of the drift:

\[
d\tilde{\mathbf{X}}_t = (J + G) \cdot \partial \mathbf{X}_t H \frac{dt}{\tau} + \sqrt{\frac{2}{\beta \tau}} G^{1/2} \cdot d\mathbf{\omega}_t
\]

\[
P(\mathbf{x} \leq \tilde{\mathbf{X}}_{t_f} < \mathbf{x} + d\mathbf{x}) = m(\mathbf{x}, t_f) d^2\mathbf{x}.
\]

In (21b) \( m(\mathbf{x}, t_f) \) is the probability density generated by (1) evaluated at \( t_f \) whilst \( H \) in (21a) is the very same Hamiltonian entering (1). The drift in (21a) must be interpreted
as the mean backward derivative $D^{-}_{\tilde{\chi}_t} \tilde{\chi}_t$ of the process $\tilde{\chi}$ (see appendix A for details). In order to compare $\chi$ with $\tilde{\chi}$ we suppose that the corresponding probability measures $P_\chi$ and $P_{\tilde{\chi}}$ are defined on the same Borel sigma algebra $\mathcal{F}_{[t_0,t_f]}$ and are absolutely continuous with respect to each other. The difference between $\chi$ and $\tilde{\chi}$ consists then in the fact that, for any $t \in [t_0,t_f]$, $\chi_t$ is adapted (i.e. measurable with respect) to the sub-sigma algebra $\mathcal{F}_{[t_0,t]}$ comprising all ‘past’ events at time $t$. The realization $\tilde{\chi}_t$ of $\tilde{\chi}$ is instead adapted to the sub-sigma algebra $\mathcal{F}_{[t,t_f]}$ of $\mathcal{F}_{[t_0,t_f]}$ comprising all ‘future’ events at time $t$ (see e.g. [23] for details). A tangible consequence of this difference is that, for any integrable test vector field $V: \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d}$ and any $t \in [t_0,t_f]$ the Ito pre-point stochastic integral satisfies

$$E^{(\chi)} \int_{t_0}^{t} dt \chi_t \cdot V(\chi_t) = E^{(\chi)} \int_{t_0}^{t} dt (J - G) \cdot \partial_{\chi_t} H(\chi_t, t) \cdot V(\chi_t)$$

(22)

while instead the post-point prescription yields

$$E^{(\tilde{\chi})} \int_{t_0}^{t} dt \tilde{\chi}_t \cdot V(\tilde{\chi}_t) = E^{(\tilde{\chi})} \int_{t_0}^{t} dt (J + G) \cdot \partial_{\tilde{\chi}_t} H(\tilde{\chi}_t, t) \cdot V(\tilde{\chi}_t).$$

(23)

We will now avail us of these observations to prove that

**Proposition 4.1.** If we decompose the current velocity (17) into a ‘dissipative’ component

$$v_+(x, t) = -G \cdot \partial_x A(x, t)$$

(24)

and a divergenceless ‘conservative’ component

$$v_-(x, t) = J \cdot \partial_x H(x, t)$$

(25)

then Kullback–Leibler divergence between $P_{\tilde{\chi}}$ and $P_\chi$ depends only on $v_+$ and is equal to

$$K(P_\chi || P_{\tilde{\chi}}) \equiv E^{(\chi)} \ln \frac{dP_\chi}{dP_{\tilde{\chi}}} = \beta E^{(\chi)} \int_{t_0}^{t_f} dt \| \partial_{\chi_t} A \|^2_G. $$

(26)

**Proof.** The proof proceeds in two steps: first we introduce two auxiliary diffusion processes, one forward and the other backward in time, for which we know the expression of the Radon–Nikodym derivative of the corresponding probability measures; then we apply Cameron–Martin–Girsanov’s formula (see e.g. [39]) to relate the auxiliary processes to $\chi$ and $\tilde{\chi}$.

**First step.** We call $\eta = \{\eta_t; t \in [t_0,t_f]\}$ the diffusion with $\mathcal{F}_{[t_0,t_f]}$-adapted realizations solution of the forward stochastic dynamics

$$d\eta_t = J \cdot \partial_{\eta_t} H \frac{dt}{\tau} + \sqrt{\frac{2}{\beta \tau}} G^{1/2} \cdot d\omega_t$$

(27a)

$$P(x \leq \eta_{t_0} < x + dx) = m_0(x) d^{2d}x.$$ 

(27b)
Similarly, let \( \tilde{\eta} = \{ \tilde{\eta}_t; t \in [t_o, t_f] \} \) be the diffusion governed by the backward dynamics

\[
d\tilde{\eta}_t = J \cdot \partial_t \eta_t \frac{dt}{\tau} + \sqrt{\frac{2}{\beta \tau} G^{1/2}} \cdot d\omega_t
\]

(28a)

\[
P(x \leq \tilde{\chi}_{t_f} < x + dx) = m(x, t_f) \, dx
\]

(28b)

with \( \tilde{\eta}_t \mathcal{F}_{[t_o, t_f]} \)-adapted. The simultaneous occurrence of additive noise and divergenceless drift in (27a), (28a) occasions the identity

\[
p_\eta(x_2, t_2 | x_1, t_1) = p_{\tilde{\eta}}(x_1, t_1 | x_2, t_2)
\]

(29)

satisfied by the transition probability densities of \( \eta \) and \( \tilde{\eta} \) for all \( x_1, x_2 \in \mathbb{R}^d \) and for all \( t_1, t_2 \in [t_o, t_f], t_1 \leq t_2 \). We therefore conclude that

\[
\frac{dP_{\tilde{\eta}}}{dP_\eta} = \frac{m(\eta_{t_f}, t_f)}{m_\circ(\eta_{t_o})}.
\]

(30)

**Second step.** We apply the composition property of the Radon–Nikodym derivative in order to couch (26) in the form

\[
K(P_\chi || P_\hat{\chi}) = -E(\eta) \frac{dP_\hat{\chi}}{dP_\eta} \ln \left( \frac{dP_\hat{\chi}}{dP_\eta} / \frac{dP_\chi}{dP_\eta} \right).
\]

(31)

Cameron–Martin–Girsanov’s formula yields immediately

\[
\frac{dP_\hat{\chi}}{dP_\eta} = \exp \left\{ \beta \int_{t_o}^{t_f} \left[ - (G \cdot \partial_t \eta_t H) \cdot \left( d\eta_t - \frac{dt}{\tau} J \cdot \partial_t \eta_t H \right) - \frac{dt}{\tau} \| \partial_t \eta_t H \|_G^2 \right] \right\}
\]

(32)

which is a martingale at time \( t_f \) by (22). In order to compute \( dP_\hat{\chi}/dP_\eta \) we first use Cameron–Martin–Girsanov’s formula adapted to backward processes [22]

\[
\frac{dP_\chi}{dP_\eta} = \exp \left\{ \beta \int_{t_o}^{t_f} \left[ (G \cdot \partial_t \eta_t H) \cdot \left( d\epsilon_t - \frac{dt}{\tau} J \cdot \partial_t \epsilon_t H \right) - \frac{dt}{\tau} \| \partial_t \epsilon_t H \|_G^2 \right] \right\}.
\]

(33)

Then we apply again the composition property to write

\[
\frac{dP_\hat{\chi}}{dP_\eta} = \frac{dP_\hat{\chi}}{dP_{\tilde{\eta}}} \frac{dP_{\tilde{\eta}}}{dP_\eta} = \frac{m(\eta_{t_f}, t_f)}{m_\circ(\eta_{t_o})} \exp \left\{ \beta \int_{t_o}^{t_f} \left[ (G \cdot \partial_t \eta_t H) \cdot \left( d\eta_t - \frac{dt}{\tau} J \cdot \partial_t \eta_t H \right) - \frac{dt}{\tau} \| \partial_t \eta_t H \|_G^2 \right] \right\}
\]

(34)

since \( \eta_t \) on the right-hand side plays the role of a mute integration variable. Upon inserting (33) and (34) in (31) and expressing the stochastic integrals in the time-reversal invariant Stratonovich mid-point discretization we arrive at
\[
\ln \frac{dP}{dP_{\chi}} = \int_{t_0}^{t_1} \frac{dt}{\tau} \left\{ \left[ (J \cdot \partial_{\chi_1} H) \cdot \partial_{\chi_1} + \tau \partial_t \right] \ln \frac{m}{\beta^d} \right\} \\
+ \beta \int_{t_0}^{t_1} \left[ d\chi_t - \frac{dt}{\tau} (J \cdot \partial_{\chi_1} H) \right]^{1/2} \left( G \cdot \partial_{\chi_1} H + \frac{1}{\beta} \partial_{\chi_1} \ln \frac{m}{\beta^d} \right). 
\]

The first integral vanishes on average since
\[
E(\chi) \left\{ \left[ (J \cdot \partial_{\chi_1} H) \cdot \partial_{\chi_1} + \tau \partial_t \right] \ln \frac{m}{\beta^d} \right\} = \int_{\mathbb{R}^{2d}} d^2 x \left( \partial_x \cdot \mathbf{v} + \tau \partial_t \right) m = 0. 
\]

In virtue of the properties of the Stratonovich integral (see e.g. [21] p 33), the expectation value of the second integral in (35) yields
\[
E(\chi) \int_{t_0}^{t_1} \left[ d\chi_t - \frac{dt}{\tau} (J \cdot \partial_{\chi_1} H) \right]^{1/2} \left( G \cdot \partial_{\chi_1} H + \frac{1}{\beta} \partial_{\chi_1} \ln \frac{m}{\beta^d} \right) = -E(\chi) \int_{t_0}^{t_1} \frac{dt}{\tau} ||\partial_{\chi_1} A||_{\xi}^2 \quad (37)
\]
where the last equality holds because \( G \) is a projector.

Some remarks are in order.

(i) The information theoretic interpretation of the entropy production is a consequence of the fluctuation-relation-type [40, 41, 38, 42, 18, 43, 19, 20] equality (35). Reference [44] discusses in detail the relation between fluctuation relations for Markov processes and exponential martingales. Finally, a recent nice overview of fluctuation theorems can be found in the lectures [10].

(ii) The proof of the identity (26) is based on the comparison between a forward and a backward dynamics in the sense of Nelson [21, 37] and admits a straightforward generalization to all the cases discussed in [20]. In particular, choosing the auxiliary process \( \eta \) to be the stochastic development map (see e.g. [45]) readily yields covariant expressions for the entropy production by diffusion on Riemann manifolds [15, 16].

(iii) The stochastic development map in the Euclidean case with flat metric reduces to the standard Wiener process. An alternative proof of (26) can then be obtained by taking the limit of vanishing noise acting on the position coordinate process.

(iv) The dissipative (24) and conservative (25) components of the current velocity are not \( L^2(\mathbb{R}^{2d}, m^{2d} x) \)-orthogonal. Therefore, it is not natural to regard the dissipative component as an independent control of the entropy production.

5. A general bound for the entropy production from the moments equation

The main consequence of the last remark in the foregoing section is that the Hamiltonian \( H \) is the natural control functional for the entropy production. The entropy production is, however, independent of derivatives of \( H \) with respect to position coordinates. This
fact raises the question whether the uncoerced degrees of freedom can be used to steer a smooth Langevin–Kramers dynamics to accomplish a finite-time transition between assigned states for arbitrarily low values of the entropy production. A simple lower bound provided by the ‘macroscopic’—in the kinetic theory sense (see e.g. [46])—dynamics shows that this cannot be the case. Let

\[ \tilde{m}(q, t) \equiv \int_{\mathbb{R}^d} d^d p \, (p, q, t) \] (38)

be the marginal probability density over the configuration space of (1). Integrating the Fokker–Planck equation governing the evolution of \( m \) over momenta it is readily seen that \( \tilde{m} \) obeys (see also (45), (46) below)

\[ \tau \partial_t \tilde{m} + \partial_q \tilde{m} \tilde{v} = 0. \] (39)

We define the ‘macroscopic drift’ \( \tilde{v} \) as the average

\[ (\tilde{m} \tilde{v})(q, t) \equiv \int_{\mathbb{R}^d} d^d p \, (m(\partial_p A))(p, q, t) \] (40)

over the momentum gradient of the non-equilibrium Helmholtz energy density (19). The probabilistic interpretation of (40) is that of the conditional expectation of \( \partial_p A \) given the value \( q \) of the position process. Since \( G \) is the vertical projector in \( \mathbb{R}^{2d} \), an immediate consequence of (40) is the identity

\[ \mathcal{E}_{t, t_0} = \int_{t_0}^{t} dt \int_{\mathbb{R}^{2d}} d^d x \, m(x, t) \| \partial_x A(x, t) \|_G^2 \]

\[ = \int_{t_0}^{t} dt \int_{\mathbb{R}^d \times \mathbb{R}^d} d^d p \, d^d q \, m(p, q, t) \left[ \| \partial_p A(p, q, t) - \tilde{v}(q, t) \|_G^2 + \| \tilde{v}(q, t) \|_G^2 \right] \] (41)

stating that the conditional expectation value of \( \| \partial_p A \|_G^2 \) over the momentum process can be written as the sum of the variance plus the square of the first moment (40). If we then neglect the variance in this latter expression we get

\[ \mathcal{E}_{t, t_0} \geq \int_{t_0}^{t} dt \int_{\mathbb{R}^d} d^d q \| \tilde{v} \|_G^2 = \tilde{\mathcal{E}}_{t, t_0}. \] (42)

Taking into account that (39) must also hold true, we interpret \( \tilde{v} \) as the current velocity of an effective Langevin–Smoluchowski dynamics. Furthermore, \( \tilde{\mathcal{E}}_{t, t_0} \) attains a minimum if the pair \((\tilde{m}, \tilde{v})\) is determined from the solution of the Monge–Ampère–Kantorovich problem [9, 7]. We will see in section 8 that the bound becomes tight in the presence of a strong separation of scales between position and momentum dynamics.

6. Entropy production extremals via Pontryagin theory

The existence of the general bound (42) indicates that the question of existence of entropy production extremals in the admissible class \( A \) is well-posed. In order to directly pursue the quest, we introduce the Pontryagin functional [25]

\[ \mathcal{A}(m, V, A) = \int_{t_0}^{t} dt \int_{\mathbb{R}^{2d}} d^d x \left\{ m \| \partial_x A \|_G^2 - V (\tau \partial_t - L^d) m \right\} \] (43)
complemented by the boundary conditions
\[ m(x, t_0) = m_o(x) \quad \text{and} \quad m(x, t_f) = m_f(x). \] (44)

The functional (43) specifies a generalized entropy production in which the dynamical constraint on the probability density appears explicitly. The ‘costate’ field \( V \colon \mathbb{R}^{2d} \times [t_o, t_f] \rightarrow \mathbb{R} \) is a Lagrange multiplier forcing the probability density \( m \) to evolve according to the Fokker–Planck specified by equation (1). The sign convention of \( V \) suits the identification of the extremal value of the costate with the ‘value’ or ‘cost-to-go’ function of Bellman’s formulation of optimal control theory [17, 28]. If we exploit the anti-symmetry of the Poisson brackets
\[ (J \cdot \partial_x S) \cdot \partial_x m = -\frac{1}{m} (J \cdot \partial_x m) \cdot \partial_x m = 0 \] (45)
and the definition of the non-equilibrium Helmholtz energy density (19), we can always couch \( \mathcal{L}^\dagger m \) as a first-order differential operation over the probability density
\[ \mathcal{L}^\dagger m = -\partial_x [m(J - G) \cdot \partial_x A]. \] (46)

This fact accounts for regarding (43) as a functional of the non-equilibrium Helmholtz energy density \( A \) and the probability density \( m \). The right-hand side of (46) coincides with the \( L^2(\mathbb{R}^{2d}) \)-dual of the generator of deterministic transport by the vector field
\[ \alpha \equiv (J - G) \cdot \partial_x A \] (47)
effectively describing a ‘coarse graining’ of the underlying stochastic dynamics. Deterministic transport by (47) arises from the fact that the entropy production is a functional of the individual probability density specified by the boundary conditions (44). This is at variance with the stochastic optimal control problems considered in [25, 17], where the cost or pay-off functional is a linear functional of the transition probability density of the process. The entropy production optimal control problem belongs instead to the class encompassed by the ‘weak-sense’ (stochastic) control theory of [28].

6.1. Variations

We determine the extremals of (43) by considering independent variations of \( m, V \) and \( A \) in the class of admissible controls. Such an hypothesis allows us to perform freely all the integrations by parts needed to extricate space-time local stationary conditions. After straightforward algebra (appendix B), the variations of \( m, V \) and \( A \) respectively yield
\[ \tau \partial_t V + (J \cdot \partial_x A) \cdot \partial_x V - (G \cdot \partial_x A) \cdot \partial_x (V - A) = 0 \] (48a)
\[ \tau \partial_t S + (J \cdot \partial_x A) \cdot \partial_x S + G A = 0 \] (48b)
\[ (J \cdot \partial_x S) \cdot \partial_x V + G(V - 2A) = 0. \] (48c)

By \( \mathcal{G} \) we denote in (48a), (48b) the operator
\[ \mathcal{G} f = -(G \cdot \partial_x S) \cdot \partial_x f + G : \partial_x \otimes \partial_x f \] (49)
doi:10.1088/1742-5468/2014/05/P05013 13
negative definite for any $f \in L^2(\mathbb{R}^d, m^d dx)$ (appendix B). The extremal equations (48) are complemented by the boundary conditions:

$$S(x,t_0) = -\ln \frac{m_0(x)}{\beta^d} \quad \text{and} \quad S(x,t_f) = -\ln \frac{m_f(x)}{\beta^d}. \quad (50)$$

The value function (48a) and entropy (48b) equations describe deterministic transport by the ‘coarse-grained’ current velocity (47). This latter vector field vanishes at equilibrium, so that the equations (48) in this case admit the physically natural solution

$$\partial_t V = \partial_t S = A = 0 \quad (51)$$

with

$$H = \frac{1}{\beta} S. \quad (52)$$

The condition (48c) plays for (48) a role analogous to that of pressure in hydrodynamics [47]. It enforces a non-local coupling between the microscopic entropy $S$ and the non-equilibrium Helmholtz energy density $A$. As in the case of hydrodynamics, non-locality arises from the existence of a divergenceless component of the velocity field. In fact, neglecting the Poisson brackets in (48c) would allow us to recover the local extremal condition $V = 2A$, analogous to the one minimizing the entropy production by Langevin–Smoluchowski dynamics [9, 7].

Beside non-locality, a second major difference from Langevin–Smoluchowski is that the extremal equations (48) are degenerate. This is a consequence of the fact that the entropy production does not depend upon $\partial_q A$, which in its turn implies that (48c) does not impose any relation between $\partial_q A$ and the value function $V$. Furthermore, the change of variables $x = p \oplus q \mapsto \tilde{x} = p + \Phi(q,t) \oplus q$ preserves the Poisson brackets. Thus, if the triple $V_*(x,t), S_*(x,t), A_*(x,t)$ solves (48) then also the triple $V_*(\tilde{x},t), S_*(\tilde{x},t), A_*(\tilde{x},t) + U(q,t)$ is solution of (48) for any smooth pair $\Phi(q,t), U(q,t)$ such that $\phi = \partial_q \Phi$ and satisfying the relations

$$\tau \partial_t \Phi(q,t) - U(q,t) = 0 \quad \text{and} \quad \Phi(q,t_o) = \Phi(q,t_f) = 0. \quad (53)$$

The generic consequence of degeneration is that the equations (48) describe a continuous family of controls for which the entropy production attains a local, at least, minimum in $A$. In the coming section 7 we will illustrate the situation with an explicit example.

7. An analytically solvable case

We can explore (48) more explicitly if we assume a Gaussian statistics for the initial and final states of the system. In particular, we restrict our attention to a two-dimensional
phase space and suppose that the microscopic entropy of the initial $i = o$ and final $i = f$
states be at most quadratic in $x = q \oplus p$:

$$S_i(p, q) = \frac{\beta (p - \mu_{pi})^2}{2\sigma_{pi}^2 \cos^2 \theta_i} + \frac{\beta (q - \mu_{qi})^2}{2\sigma_{qi}^2 \cos^2 \theta_i}$$

$$- \beta \tan \theta_i \frac{(p - \mu_{pi})(q - \mu_{qi})}{\sigma_{pi} \sigma_{qi} \cos \theta_i} - \ln \left( \frac{1}{2\pi \sigma_{pi} \sigma_{qi} \cos \theta_i} \right)$$

(54)

corresponding to

$$E\chi_i = \left[ \frac{\mu_{qi}}{\mu_{pi}} \right]$$

(55)

and

$$E(\chi_i - E\chi_i) \otimes (\chi_i - E\chi_i) = \frac{1}{\beta} \begin{bmatrix} \sigma_{qi}^2 & \sigma_{qi} \sigma_{pi} \sin \theta_i & \sigma_{qi} \sigma_{pi} \sin \theta_i \\ \sigma_{qi} \sigma_{pi} \sin \theta_i & \sigma_{pi}^2 & \sigma_{pi}^2 \\ \sigma_{qi} \sigma_{pi} \sin \theta_i & \sigma_{pi}^2 & \sigma_{pi}^2 \end{bmatrix}.$$  

(56)

In particular, we choose $\mu_{pi} = \mu_{qi} = 0$, whilst $0 \leq \theta_i < \pi/2$ parametrizes the degree of correlation between position and momentum variables of the final state. Under these assumptions, we look for a solution of the extremal equations by means of quadratic Ansätze for the microscopic entropy

$$S(p, q, t) = \frac{\beta (p - \mu_{pi})^2}{2\sigma_{pi}^2 \cos^2 \theta_t} + \frac{\beta (q - \mu_{qi})^2}{2\sigma_{qi}^2 \cos^2 \theta_t}$$

$$- \beta \tan \theta_t \frac{(p - \mu_{pi})(q - \mu_{qi})}{\sigma_{pi} \sigma_{qi} \cos \theta_t} - \ln \left( \frac{1}{2\pi \sigma_{pi} \sigma_{qi} \cos \theta_t} \right)$$

(57)

and the non-equilibrium Helmholtz energy

$$A(p, q, t) = \frac{A_{11,t} q^2 + 2A_{12,t} pq + A_{22,t} p^2}{2} + a_{1,t} q + a_{2,t} p$$

(58)

for all $t \in [t_o, t_f]$. The Ansätze imply that the entropy production

$$\mathcal{E}_{t_o, t_f} = \beta \int_{t_o}^{t_f} \frac{dt}{\tau} \left\{ 2a_{2,t} \left( A_{22,t} \mu_{pi} + A_{12,t} \mu_{qi} \right) + a_{2,t}^2 \right\}$$

$$+ \beta \int_{t_o}^{t_f} \frac{dt}{\tau} \left\{ A_{22,t} \left( \mu_{qi}^2 + \sigma_{qi}^2 \cos^2 \theta_t \right) + A_{12,t} \left( \mu_{qi}^2 + \sigma_{qi}^2 \cos^2 \theta_t \right) \right\}$$

$$+ 2\beta \int_{t_o}^{t_f} \frac{dt}{\tau} A_{12,t} A_{22,t} \left( \mu_{pi} \mu_{qi} + \sigma_{pi} \sigma_{qi} \sin \theta_t \right)$$

(59)

does not depend explicitly upon $A_{11,t}$ and $a_{1,t}$.

Using the quadratic Ansätze (57), (58) in (48c) we obtain

$$V(p, q, t) = (q^2 \partial_p \partial_q - 2q \partial_p) \left[ A(p, q, t) - \frac{y_t}{\beta} S(p, q, t) \right] + \frac{2y_t}{\beta} S(p, q, t) + \bar{V}(t)$$

(60)

doi:10.1088/1742-5468/2014/05/P05013
where
\[ y_t \equiv \beta \frac{\partial^2 A}{\partial p^2} \frac{\partial^2 p}{\partial A} \] is a function of the time variable alone, well-defined as long as the probability density of the state is non-degenerate. The explicit value of \( \bar{V}(t) \) does not play any role in the considerations which follow. If we now insert (60) into (48) and (48b), these equations foliate into a closed system of ordinary differential equations for the coefficients of the Ansätze (57) and (58). The calculation is laborious but straightforward. Upon setting
\[ A_{22} \equiv -\tau \dot{y}_t \] we find for the coefficients of second-order monomials in (57) and (58), the set of relations
\[
\begin{align*}
A_{11} &= \frac{y_t}{\beta} \left( \partial_q^2 S - \partial_p \partial_q S - \frac{\dot{y}_t}{y_t} \partial_q \partial_p S \right) - A_{12} \\
A_{12} &= \frac{y_t}{\beta} \partial_p \partial_q S + \tau \frac{\dot{y}_t}{2y_t} \\
\partial_p^2 S &= -\frac{\tau \beta \dot{y}_t}{y_t^2} \\
\partial_q^2 S &= \frac{(\partial_p \partial_q S)^2}{\partial_p^2 S} - \tau \beta \frac{\dddot{y}_t^2 - 2 \ddot{y}_t \dddot{y}_t}{4 \dot{y}_t^3}.
\end{align*}
\] The cross-correlation coefficient \( \partial_p \partial_q S \) of the microscopic entropy enters these equations as a free parameter only subject to the boundary conditions. It turns out that the function \( y_t \) must satisfy the fourth-order nonlinear differential equation
\[
\dddot{y}_t \dot{y}_t^2 - 2 \ddot{y}_t \dddot{y}_t + \dddot{y}_t^3 = 0
\] with solution
\[ y_t = \tau \Omega \left\{ c_0 + c_1 \Omega t + c_1 \left[ \sin (\Omega t + \varphi) - \sin \varphi \right] \right\}. \] The coefficients \( c_0, c_1, \Omega, \varphi \) are fixed by the boundary conditions. Upon imposing the boundary conditions for
\[ t = 0 \] and requiring continuity of solutions for \( S_t \to S_\infty \), we get
\[ c_0 = -\frac{\sigma_{p0} \sigma_{q0}}{2} \] and
\[ c_1 = -\frac{\sigma_{q0}^2}{8 \cos^2 \varphi / 2} \]
while Ω and ϕ satisfy the transcendental equations:

\[
\sigma_{p}^2 = \frac{4\sigma_{p,0} \cos^2 \varphi/2 + \sigma_{q,0} [\Omega t + \sin(\Omega t + \varphi) - \sin \varphi]}{16 \cos^2 \theta_t \cos^2(\varphi/2) \cos^2(\Omega t + \varphi)/2} 
\]

\[
\frac{\sigma_{q}^2}{\sigma_{q,0}^2} = \frac{\cos^2(\Omega t + \varphi)/2}{\cos^2 \varphi/2}.
\]

We verify that the coefficients of the microscopic entropy are positive definite:

\[
\partial_p^2 S = \frac{16 \beta \cos^2 \varphi/2 \cos^2(\Omega t + \varphi)/2}{\{4\sigma_{p,0} \cos^2(\varphi/2) + \sigma_{q,0} [\Omega t + \sin(\Omega t + \varphi) - \sin \varphi]\}^2} \geq 0 \tag{70}
\]

and

\[
\partial_q^2 S = \frac{\beta \cos^2 \varphi/2}{\sigma_{q,0}^2 \cos^2(\Omega t + \varphi)/2} + \left(\frac{\partial_p^2 q S}{\partial_p^2 S}\right) \geq 0.
\]

The equations for the coefficients of the first-degree monomials yield

\[
\mu_{qt} = \mu_{q,t} \frac{t}{t_1}
\]

and

\[
a_{2,t} = \frac{\mu_{q,t} \tau (2 + \Omega t \tan(\Omega t + \varphi)/2)}{2t_1} - \frac{\mu_{p,t} A_{22,t}}{\beta} - \frac{y_{q,t}^2 \Omega^2 t_1}{4 \beta \cos^2 \varphi/2}.
\]

The remaining independent equations determine \(a_{1,t}\) as a functional of \(\partial_p^2 q S\) and \(\mu_{p,t}\) and their time derivatives. We do not need, however, the explicit expression to compute the entropy production, for which we find

\[
\frac{E_{t_1 t_0}}{\beta} = \frac{\mu_{q,t}^2 \tau}{t_1} + \frac{\sigma_{q,0}^2 \Omega^2 t_1}{4 \beta \cos^2 \varphi/2}.
\]

Four properties of the extremal value of the entropy production (74) are worth emphasizing. First (74) is fully specified by the boundary conditions and by the degrees of freedom fixed by the extremal equations (48). This fact is \textit{a posteriori} evidence of the degeneration of the extremal protocols. Second, (74) does not depend upon the expected value of the momentum variable but only upon its variance. The third property is that (74) corresponds to a \textit{constant} entropy production rate over the transition horizon. This phenomenon is reminiscent of the Langevin–Smoluchowski case where the entropy production coincides with the kinetic energy of the current velocity so that the optimal value is attained along free-streaming trajectories. The fourth property pertains to the limit of infinite-time horizon \(t_f \uparrow \infty\). The position variable variance remains finite in such a limit if \(\Omega t_f\) is finite. This lead us to infer generically a \(1/t_f\) decay of the entropy production in such a limit.

The explicit dependence of (74) on the boundary conditions can be obtained in several special cases.

doi:10.1088/1742-5468/2014/05/P05013
7.1. $\sigma_{q_0} = \sigma_{q_f}$

The condition is satisfied for

$$\Omega = y_t = 0$$

in the control horizon. Correspondingly, (69a) yields

$$\sigma_{p,t} = \frac{\sigma_{p,o}}{\cos \theta_t}. \quad (76)$$

If $\theta_t \neq 0$, (76) states that, while enforcing (75), we can use $\partial_p \partial_q S$ to steer the system to a final state with larger momentum variance and non-vanishing correlation between position and momenta. For vanishing $\Omega$ the entropy production is determined by the variation of the position average:

$$E_{t_f,0} = \beta \left[ \frac{\mu_{q,t} \tau_p}{t_f} + \frac{\sigma_{p,o} \tau_p}{2} \frac{d}{dt} \tanh \theta_t - \frac{\tau_q}{t_f} \left( \mu_{q,t} + t_f \mu_{p,t} + \frac{\mu_{q,t} \sigma_{p,o}}{\sigma_{q,o}} \frac{d}{dt} \tanh \theta_t \right) \right] \quad (78a)$$

$$S = \beta \left( p - \mu_{p,t} \right) \left( q - \mu_{q,t} \right) - \ln \frac{1}{2 \pi \sigma_{p,o} \sigma_{q,o}} \quad (78b)$$

with $\tanh \theta_t$, $\sigma_{q,t}$ and $\mu_{p,t}$ arbitrary differentiable functions matching the boundary conditions. If we add the requirement $\theta_t = \theta_t = 0$, (78) shows that a transition changing only the mean value of the position variable requires a quadratic additive Hamiltonian—i.e. of the form $H(p, q, t) = H_p(p, t) + H_q(q, t)$, where $H_p(p, t)$ must, however, include a linear momentum dependence.

7.2. Small one-parameter variation of the statistics

Let us suppose that there exists a non-dimensional parameter $\varepsilon$ such that the elements of the correlation matrix of the final state admit an expansion of the form

$$\sigma_{p,t} = \sigma_{p,o} \left[ 1 + p_1 \varepsilon + \frac{p_2 \varepsilon^2}{2} + \frac{p_3 \varepsilon^3}{6} + O(\varepsilon^4) \right]$$

$$\sigma_{q,t} = \sigma_{q,o} \left[ 1 + q_1 \varepsilon + \frac{q_2 \varepsilon^2}{2} + \frac{q_3 \varepsilon^3}{6} + O(\varepsilon^4) \right]$$

with

$$\cos \theta_t = 1 - \frac{w_3 \varepsilon^2}{2} + O(\varepsilon^4). \quad (80)$$
Under the foregoing hypothesis we find
\[
\Omega = \frac{2(q_1 + p_1)\sigma_{p\alpha}\varepsilon}{t\sigma_{q\alpha}} + \frac{2(p_2 - w_2^2 + q_2 - 2q_1^2)\sigma_{p\alpha}\varepsilon^2}{t\sigma_{q\alpha}} + O(\varepsilon^3)
\]
(81a)
\[
\varphi = -2\arccot \frac{(q_1 + p_1)\sigma_{p\alpha}}{q_1\sigma_{q\alpha}} - \frac{(q_1 + p_1)^3\sigma_{p\alpha}^2 + [q_1(w_2^2 + 2q_1^2) + p_1q_2 - p_2q_1]\sigma_{q\alpha}^2\sigma_{p\alpha}\varepsilon + O(\varepsilon^3)}{(q_1 + p_1)^2\sigma_{p\alpha}^2 + q_1^2\sigma_{q\alpha}^2})
\]
(81b)
which give for the entropy production
\[
\frac{\mathcal{E}_{t,0}}{\beta} = \frac{\mu_{q\alpha}^2\tau}{t\beta t_f} + \frac{\left[(p_1 + q_1)^2\sigma_{p\alpha}^2 + q_1^2\sigma_{q\alpha}^2\right]\tau\varepsilon^2}{\beta t_f} + \frac{\left[(p_1 + q_1)(p_2 + q_2 + (p_1 - q_1)q_1 - w_2^2)\sigma_{p\alpha}^2 + q_2q_1\sigma_{q\alpha}^2\right]\tau\varepsilon^3}{\beta t_f} + O(\varepsilon^4).
\]
(82)
It is interesting to explore the consequence of this formula in three sub-cases. For this purpose we introduce the non-dimensional parameter
\[
\lambda = \frac{\sigma_{p\alpha}}{\sigma_{q\alpha}}
\]
(83)
measuring the scale separation between momentum and position fluctuations.

7.2.1. $q_n = p_n = 0 \forall n > 1$. Both the position and the momentum variances are linear in $\varepsilon$. We can therefore recast the expansion of the entropy production directly in terms of the change of the variances across the control horizon. We obtain
\[
\frac{\mathcal{E}_{t,0}}{\beta} = \frac{\mu_{q\alpha}^2\tau}{t\beta t_f} + \frac{(\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)^2\tau}{\beta t_f} + \frac{[\sigma_{q\alpha}^2 - \sigma_{p\alpha}^2 + \lambda(\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)]^2\tau}{\beta t_f} + \frac{[(\sigma_{q\alpha}^2 - \sigma_{p\alpha}^2)^2 - \lambda^2(\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)](\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)\tau}{\beta t_f} + \frac{2\lambda[\sigma_{q\alpha}^2 - \sigma_{p\alpha}^2 + \lambda(\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)](1 - \cos \theta_t)\tau}{\beta \sigma_{q\alpha}^2 t_f} + \text{h.o.t.}
\]
(84)

7.2.2. $\theta_t = \sigma_{q\alpha} - \sigma_{p\alpha} = 0$ and $\varepsilon = (\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)/\sigma_{q\alpha}$. Under these hypotheses, the marginal momentum distribution in the final state coincides with that of the initial state. As a result, the expansion of the phase $\varphi$ starts from the neighborhood of $\pi/2$. The entropy production reduces to
\[
\frac{\mathcal{E}_{t,0}}{\beta} = \frac{\mu_{q\alpha}^2\tau}{t\beta t_f} + \frac{\tau(1 + \lambda^2)(\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)^2\tau}{\beta t_f} - \frac{\tau\lambda^2(\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)^3}{\beta \sigma_{q\alpha}^2 t_f} + O(\sigma_{q\alpha}^2 - \sigma_{q\alpha}^2)^4.
\]
(85)
In figure 1 we report the behavior of the momentum and position variance for vanishing cross-correlation. It is worth emphasizing that the momentum variance does not remain constant during the control horizon unless
\[
\sigma_{q\alpha} = \sigma_{q\alpha} = \Omega = 0.
\]
(86)
7.2.3. ‘Over-damped’ limit: \( \theta = 0 \) and momentum variance \( \sigma^2_{p,t} = 0 \) and \( \varepsilon = (\sigma_{q,t} - \sigma_{q,0})/\sigma_{q,0} \) for \( \lambda \ll 1 \). At variance with the foregoing we now assume a wide scale separation between the position and momentum variance. We readily see from (81) that \((\varphi, \Omega) = (-\tau + O(\lambda), O(\lambda))\). We can solve the boundary condition equations (69) in the limit of vanishing \( \lambda \) up to all order accuracy in \( \varepsilon \):

\[
\begin{align*}
\Omega &= \frac{2\lambda \varepsilon}{t_f} \left\{ 1 - \varepsilon + \frac{2\varepsilon^2}{3} + O(\varepsilon^3) \right\} \xrightarrow{\lambda \varepsilon \to 0} \frac{6\lambda \varepsilon}{t_f [3 + \varepsilon (3 + \varepsilon)]} + o(\lambda) \\
\varphi &= -\pi + 2\lambda \left\{ 1 - \varepsilon + \frac{2\varepsilon^2}{3} + O(\varepsilon^3) \right\} \xrightarrow{\lambda \varepsilon \to 0} -\pi + \frac{\Omega t_f}{\varepsilon} + o(\lambda).
\end{align*}
\] (87a, 87b)

The corresponding value of the entropy production is

\[
\frac{\mathcal{E}_{t_f,0}}{\beta} = \frac{\mu_{q,t}^2 \lambda}{t_f} + \frac{1}{\beta t_f} + o(\lambda).
\] (88)

We notice that this is exactly the entropy production by a transition governed by Langevin–Smoluchowski dynamics between Gaussian states \([9, 7, 16]\). Indeed, the regime we are considering here corresponds to the ‘over-damped’ asymptotics of the Langevin–Kramers dynamics. Namely, upon inserting (87) into the quadratic Ansätze for the nonequilibrium Helmholtz energy and microscopic entropy densities we get

\[
\begin{align*}
(\partial_q A)(0, q, t) |_{\mu_{q,t}=0} &= -\frac{\mu_{q,t} + q(\sigma_{q,t} - \sigma_{q,0})/\sigma_{q,0}}{1 + t(\sigma_{q,t} - \sigma_{q,0})/(t \sigma_{q,0})} \tau + o(\lambda) \\
(\partial_p A)(0, q, t) |_{\mu_{p,t}=0} &= -(\partial_q A)(0, q, t) |_{\mu_{q,t}=0} + o(\lambda) \\
(\partial_q S)(0, q, t) &= \frac{\beta (q - (\mu_{q,t} t)/t_f)}{\sigma_{q,0}^2 [1 + t(\sigma_{q,t} - \sigma_{q,0})/(t \sigma_{q,0})]^2} + o(\lambda).
\end{align*}
\] (89a, 89b, 89c)

We see that the \( \lambda \)-independent parts of (89a) and (89c) coincide with the values obtained for the same quantities in the Langevin–Smoluchowski case \([9, 7, 16]\). In the

doi:10.1088/1742-5468/2014/05/P05013
forthcoming section, we will show that the equations (48) encapsulate also in general the results obtained for Langevin–Smoluchowski dynamics. In particular, the equality (89b) guarantees that an homogenization theory ‘centering condition’ holds for the Gaussian model so that the Langevin–Smoluchowski dynamics is recovered as the solution of a suitable ‘cell problem’ [31].

8. ‘Over-damped’ asymptotics

In the presence of a wide separation of between the characteristic scales of the momentum and position variables, the Langevin–Smoluchowski or ‘over-damped’ dynamics often provides a good approximation to the Langevin–Kramers dynamics. The entropy production by a smooth Langevin–Smoluchowski dynamics attains a minimum value if the control potential obeys a Monge–Ampère–Kantorovich dynamics [8, 9, 7, 16]. In this section our aim is to investigate in what sense we can recover from (48) the results previously established for Langevin–Smoluchowski dynamics. To address this question, we suppose that the probability densities of the initial and final states take the additive form

$$m_o(p, q) = \left(\frac{\beta}{2\pi \lambda}\right)^d \exp\left\{-\frac{\beta \|p\|^2}{2\lambda^2} - \beta U_o(q)\right\}$$

(90)

and

$$m_f(p, q) = \left(\frac{\beta}{2\pi \lambda}\right)^d \exp\left\{-\frac{\beta \|p\|^2}{2\lambda^2} - \beta U_f(q)\right\}$$

(91)

with $\lambda \ll 1$ a non-dimensional parameter generalizing (83) in order to describe the scale separation between momentum and position dynamics.
Multi-scale perturbation theory (often also referred to as homogenization theory, see e.g. [30, 31]) in powers of $\lambda$ equips us with the tools to extricate the asymptotic expression of solutions of (48) for $\beta \|p\|^2 \ll \lambda \ll 1$ in the form

$$A(x, t) = \sum_{i=0}^{2} \lambda^i A_i(p, q, t, \ldots) + o(\lambda^2) := \tilde{A}(\tilde{p}, q, t, \ldots) \quad (92)$$

and similarly for $S$ and $V$. The ... in (92) stand for the scales which we eventually neglect in the asymptotics. Once we availed us of (92), the extremal equations (48) become

$$\tau \partial_t \tilde{V} + \frac{1}{\lambda} \left[ (\partial_p \tilde{A}) \cdot \partial_q \tilde{V} - (\partial_q \tilde{A}) \cdot \partial_p \tilde{V} \right] - \frac{1}{\lambda^2} (\partial_p \tilde{A}) \cdot \partial_p (\tilde{V} - \tilde{A}) = 0 \quad (93a)$$

$$l \tau \partial_t \tilde{S} + \frac{1}{\lambda} \left[ (\partial_p \tilde{A}) \cdot \partial_q \tilde{S} - (\partial_q \tilde{A}) \cdot \partial_p \tilde{S} \right] + \frac{1}{\lambda^2} \tilde{S} \tilde{A} = 0 \quad (93b)$$

$$\frac{1}{\lambda} \left[ (\partial_p \tilde{S}) \cdot \partial_q \tilde{V} - (\partial_q \tilde{S}) \cdot \partial_p \tilde{V} \right] + \frac{1}{\lambda^2} \tilde{S} (\tilde{V} - 2 \tilde{A}) = 0 \quad (93c)$$

where we used the notation

$$\tilde{S} := -(\partial_p \tilde{S}) \cdot \partial_p + \partial_p \cdot \partial_p. \quad (94)$$

In what follows, we will also write $\tilde{S}^{(0)}$ to denote the replacement in (94) of $\tilde{S}$ with its zeroth-order approximation $\tilde{S}^{(0)}$.

As often occurs in the homogenization of parabolic equation [31], we need to analyze the first three orders of the regular expansion in powers of $\lambda$ in order to fully determine the leading order contributions to $S$ and $A$. This is because the first order is needed to assess the centering condition coupling the widely separated scales which we wish to resolve in the asymptotics. The second order approximation uses the information conveyed by the centering condition to determine the cell problem, a closed equation for the effective dynamics in the limit of vanishing $\lambda$.

### 8.1. Zeroth order

From (93c) we get the condition

$$\tilde{S}^{(0)} (2A_{(0)} - V_{(0)}) = 0 \quad (95)$$

stating that at leading order the value function $V_{(0)}$ may differ from the non-equilibrium Helmholtz energy at most by a function independent of momentum variables:

$$V_{(0)} = 2A_{(0)} + V_{(0;0)} \quad (96)$$

where

$$\partial_p V_{(0;0)} = 0. \quad (97)$$

Once we insert (96) in (93a), (93b) we get into

$$(\partial_p A_{(0)}) \cdot \partial_p A_{(0)} = 0 \quad (98a)$$

doi:10.1088/1742-5468/2014/05/P05013

22
On extremals of the entropy production by 'Langevin–Kramers' dynamics

\( \tilde{S}(0)A(0) = 0. \)  \hfill (98b)

The boundary conditions (90) and (91) translate into

\[ S_o = S_I + o(\lambda^2). \]  \hfill (99)

Hence, we see from (98) that (99) is satisfied upon setting

\[ S(0)(\tilde{p}) = \frac{||\tilde{p}||^2}{2} + S(0)(q, t, \ldots) \]  \hfill (100)

and

\[ \partial_p A(0) = \partial_p V(0) = 0. \]  \hfill (101)

8.2. First order: centering condition

Maxwell momentum distribution is the unique element of the kernel of \( \tilde{S}(0)^\dagger \) in \( L^2(\mathbb{R}^d) \). By Fredholm alternative (see e.g. [31])

\[ (\partial_p S(0)) \cdot \partial_q V(0) - (\partial_q S(0)) \cdot \partial_p V(0) - \tilde{S}(0)(2A(1) - V(1)) = 0 \]  \hfill (102)

admits a unique solution if and only if the solvability condition

\[ 0 = \int_{\mathbb{R}^d} d^dpe^{-S(0)} [(\partial_p e^{-S(0)}) \cdot \partial_q V(0) - (\partial_q e^{-S(0)}) \cdot \partial_p V(0)] \]
\[ = -\int_{\mathbb{R}^d} d^dp [(\partial_p e^{-S(0)}) \cdot \partial_q V(0) - (\partial_q e^{-S(0)}) \cdot \partial_p V(0)] = \partial_q \int_{\mathbb{R}^d} d^dpe^{-S(0)} \partial_p V(0) \]  \hfill (103)

holds true, which is always the case if (101) is verified. Hence we conclude

\[ V(1) = 2(A(1) + \tilde{p} \cdot \partial_q A(0)) + \tilde{p} \cdot \partial_q V(0) + V(1:0) \]  \hfill (104)

with

\[ \partial_p V(1:0) = 0. \]  \hfill (105)

Turning to the value function in equation (93a), we see that

\[ -\sum_{i=0}^{1} (\partial_p V(1-i))^i \cdot \partial_p A(0) + (\partial_p A(0)) \cdot \partial_q V(0) - (\partial_q A(0)) \cdot \partial_p V(0) + 2\partial_p A(1) \cdot \partial_p A(0) = 0 \]  \hfill (106)

is also satisfied by (101). New information comes from the expansion of the microscopic entropy equation

\[ -(\partial_q A(0))^i \cdot \partial_p S(0) + \tilde{S}(0)A(1) = 0 \]  \hfill (107)

which yields the centering condition of the expansion:

\[ A(1) = -\tilde{p} \cdot \partial_q A(0). \]  \hfill (108)

doi:10.1088/1742-5468/2014/05/P05013
On extremals of the entropy production by ‘Langevin–Kramers’ dynamics

The centering condition couples here a linear asymptotic behavior in \( \mathbf{p} \) with a non-trivial cell problem in \((q, t)\), which we will determine by requiring solvability in the sense of Fredholm’s alternative at order \( \mathrm{O}(\lambda^2) \). Contrasting (108) with (104) we infer that

\[
\partial_p V^{(1)} = \partial_q V^{(0:0)}.
\]  

(109)

It is worth emphasizing here the relevant simplification induced by the over-damped limit. The over-damped limit entitled us to neglect the Poisson bracket also in the sub-leading order (102) of the expansion of (48c). The crucial consequence is that the relation between \( V \) and \( A \) remains local within accuracy. Intermediate asymptotics around (98) do not enjoy this property. This is not surprising in light of the example of section 7.2.2 showing that, even in the Gaussian case, the coincidence of the initial and final marginal momentum distribution does not in general imply thermalization.

8.3. Second order: cell problem

The extremal equation (93c) yields now the condition

\[
(\partial_p S^{(0)}) \cdot \partial_q V^{(1)} - (\partial_q S^{(0)}) \cdot \partial_q V^{(0:0)} - \tilde{\mathbf{S}}^{(0)} (2A^{(2)} - V^{(2)}) = 0.
\]

(110)

The solvability condition imposes

\[
\partial_q V^{(0:0)} = 0
\]

(111)

whence

\[
V^{(2)} = 2A^{(2)} + \tilde{\mathbf{p}} \cdot \partial_q V^{(1:0)} + V^{(2:0)}
\]

(112)

with \( \partial_p V^{(2:0)} = 0 \). Hence, combining (96) with (111), (108) and (99), the value function equation reduces to

\[
2\tau \partial_t A^{(0)} - \|\partial_q A^{(0)}\|^2 = 0.
\]

(113)

Finally, the equation for the microscopic entropy is

\[
\partial_t S^{(0)} - (\partial_q A^{(0)}) \cdot \partial_q S^{(0)} + \tilde{\mathbf{p}} \cdot (\partial_q \otimes \partial_q A^{(0)}) \cdot \partial_p S^{(0)} + \tilde{\mathbf{S}}^{(0)} A^{(2)} = 0.
\]

(114)

Regarding this latter as an equation for \( A^{(2)} \) and invoking again Fredholm’s alternative, we see that it admits a unique solution if and only if

\[
\tau \partial_t S^{(0:0)} - (\partial_q A^{(0)}) \cdot \partial_q S^{(0:0)} + \partial_q \partial_q A^{(0)} = 0
\]

(115)

holds true. The system formed by the equalities (100), (108) and the cell problem equations (113) and (115) fully specifies the homogenization asymptotics we set out to derive.

doi:10.1088/1742-5468/2014/05/P05013

24
8.4. Asymptotic expression of the entropy production

The cell problem equations (113) and (115) specify a Monge–Ampère–Kantorovich evolution [32] between an initial configuration space state with density

\[ \tilde{m}_0(q) = \left( \frac{\beta}{2\pi} \right)^{d/2} e^{-\beta U_0(q)} \]

and a final one with density

\[ \tilde{m}_f(q) = \left( \frac{\beta}{2\pi} \right)^{d/2} e^{-\beta U_f(q)} \].

The recovery of the Monge–Ampère–Kantorovich equations unveils the link between the minimum entropy production by the phase–space process (1) and the optimal control of the corresponding thermodynamic quantity, which can be directly defined in the overdamped limit. As a matter of fact, the expansion of \( A \) starts with the \( O(\lambda) \) term specified by the centering condition (108), which is linear in \( \tilde{p} = p/\lambda \). The upshot is that the overdamped expansion of the minimum over \( A \) of the Langevin–Kramers entropy production starts with

\[ E_{t,0} = \beta \int_0^t \frac{dt}{\tau} \int_{\mathbb{R}^d} d^d \tilde{q} \beta^{d/2} e^{-S_{(0,0)}} \| \partial_q A_{(0)} \|^2 + O(\lambda). \]

We therefore proved that the leading order of the expansion coincides with the minimal entropy production by the Langevin–Smoluchowski dynamics.

9. Discussion

Many physical systems are modeled by kinetic-plus-potential Hamiltonians

\[ H(p, q) = \frac{\|p\|^2}{2} + U(q, t). \]

The example of section 7.1 evinces that requiring (119) adds an optimization constraint which is not generically satisfied by the extremal equations (48) over \( A \). Furthermore, the kinetic-plus-potential hypothesis deeply affects the control problem by introducing two new difficulties. First, it restricts to the gradient \( \partial_q U \) of the potential energy the available control degrees of freedom. In this regard, it is worth emphasizing that it is a non-trivial consequence of Hörmander theorem (see e.g. [33] and references therein) that a sufficiently regular (119) is enough to generate a Fokker–Planck evolution of a smooth initial density for a Langevin–Kramers dynamics with degenerate noise acting only on \( d \) out of \( 2d \) degrees of freedom. Physical intuition suggests, however, that the surmise (119) should not create an insurmountable difficulty for controllability, by which we mean the existence of a non-empty set of potentials \( U(q, t) \) able to steer a transition between two probability densities verifying physically plausible assumptions. The second and more substantial difficulty is that inserting (119) into (18) yields an entropy production expression that depends
Upon the control only implicitly through the probability measure. Controls are in such a case only subject to the constraint imposed by the requirement of steering a finite-time transition between smooth probability densities. General considerations [17] lead us to envisage that entropy production may only attain an infimum when evaluated according to a singular control strategy. Such a scenario occurs, for example, when trying to minimize over a fixed time horizon the expected response energy of an harmonic oscillator stirred by a white noise force and controlled by a bounded force not explicitly appearing in the expected response energy [48]. An important difference is, however, that admissible forces in [48] may depend on the momentum process.

A problem more closely related to entropy production control has been instead considered in [49], where the authors set out to derive a lower bound for the work done in the special case of a thermodynamic transition to a state specified by a quadratic final value of the potential energy. Also in the case of entropy production minimization the Gaussian case is very special. Namely, using cumulants and restricting by hypothesis the attention to quadratic potentials, entropy production minimization is amenable to a tractable finite-dimensional problem. We leave, however, the problem of finding entropy production control strategies of the form (119) beyond the scope of the present work.

In conclusion, by insisting that the Hamiltonian belongs to the class of admissible controls \( A \), we focused instead on control strategies which we interpret as ‘macroscopic’ in view the regularity assumptions on the control Hamiltonian. These assumptions are analogous to those adopted in previous studies of the entropy production by Langevin–Smoluchowski dynamics [7] or by Markov jump processes [11]. We therefore gather that the existence of the entropy production minimum (48)—degenerate because of non-coercivity, and which recovers in the over-damped limit the Monge–Ampère–Kantorovich evolution—yields a robust general picture of the ‘optimal’ thermodynamics for a large class of physical processes described by Markovian evolution equations.

Acknowledgments

It is a pleasure to thank Krzysztof Gawędzki and Carlos Mejía–Monasterio for discussions and very useful comments on this manuscript. The author acknowledges support by the Center of Excellence ‘Analysis and Dynamics’ of the Academy of Finland. The results of this paper were first presented during the conference ‘6-th Paladin memorial: Large deviations and rare events in physics and biology’ Rome, September 23–25, 2013. The author wishes to warmly thank the organizers for giving him the opportunity to partake in the event as an invited speaker. He also wishes to dedicate the paper to the memory of Giovanni Paladin.

Appendix A. Mean derivatives and current velocity of a diffusion process

We recall that the drift of an \( \mathbb{R}^d \)-valued diffusion processes \( \zeta \equiv \{ \zeta_t, t \in [t_0, t_f] \} \) with generator

\[
\mathcal{L} = b \cdot \partial_x + \frac{1}{2} K : \partial_x \otimes \partial_x \tag{A.1}
\]

doi:10.1088/1742-5468/2014/05/P05013
can be regarded as the mean forward derivative of the process:
\[ D_x \zeta_t \equiv \lim_{d t \downarrow 0} \mathbb{E}_{\zeta_t=x} \frac{\zeta_{t+dt}-\zeta_t}{dt} = \mathcal{L} x. \] (A.2)

Under standard regularity hypotheses \[37\], it is possible to define the mean backward derivative of the very same process as
\[ D^- x \zeta_t \equiv \lim_{d t \downarrow 0} \mathbb{E}_{\zeta_t=x} \frac{\zeta_t - \zeta_{t-dt}}{dt}. \] (A.3)

**Proposition A.1.** Let \( \zeta \equiv \{ \zeta_t, t \in [t_o, t_f] \} \) be a smooth diffusion with generator (A.1) and density \( m \). The mean forward derivative is
\[ D^- x \zeta_t = -\frac{1}{m(x, t)} \left( \delta^t x - x \delta^t \right) \tau m(x, t). \] (A.4)

**Proof.** By hypothesis, \( \zeta \) is Markovian with density \( m \) in the time interval \([t_o, t_f]\). Given its forward transition probability density \( p \), the backward transition probability \( p^* \) of the same process density must then satisfy
\[ p^*(x_1, t_1 | x_2, t_2) = \frac{1}{m(x_2, t_2)} p(x_2, t_2 | x_1, t_1) m(x_1, t_1) \] (A.5)
for any \( x_1, x_2 \in \mathbb{R}^d, t_1, t_2 \in [t_o, t_f] \) such that \( t_2 \geq t_1 \). By (A.5) it follows immediately that
\[ \mathbb{E}_{\zeta_t=x} \zeta_{t-dt} = \int d^d x_1 p(x, t | x_1, t-dt) m(x, t) \] (A.6)
If we integrate the Fokker–Planck equation and its adjoint equation over a time horizon of order \( O(dt) \) we arrive at
\[ \mathbb{E}_{\zeta_t=x} \zeta_{t-dt} = x + \frac{dt}{\tau} \int d^d x_1 \frac{m(x_1, t) \mathcal{L} \delta^{(2d)}(x_1-x) - \delta^{(2d)}(x_1-x) \mathcal{L}^t m(x_1, t)}{m(x, t)} + O \left( \frac{dt}{\tau} \right) \] (A.7)
which inserted in the definition (A.3) yields the claim.

The mean backward drift governs the Fokker–Planck evolution of the density of the process from \( t_f \) to \( t_o \) \[37\]. By the Hörmander theorem \[33\], the proposition above encompasses the degenerate noise case described by (1). We are therefore entitled to write
\[ \tau D_x X_t = J \cdot \partial_x H - G \cdot \partial_x \left( H + \frac{2}{\beta} \ln \frac{m}{\beta^d} \right). \] (A.8)

The current velocity of a smooth diffusion is defined as
\[ \mathbf{v}(x, t) \equiv \frac{D_x + D^- x}{2} \zeta_t \] (A.9)
whence (17) follows immediately. The advantage of the current velocity representation is that the Fokker–Planck equation for the probability density \( m \) in \([t_o, t_f]\) is mapped by (A.9) into the deterministic mass conservation equation
\[ \tau \partial_t m + \partial_x \cdot \mathbf{v} m = 0. \] (A.10)

do:10.1088/1742-5468/2014/05/P05013
Appendix B. Variations of the Pontryagin functional

We avail us of the identity (46) to treat (43) as a functional of the independent fields $A$ and $m$. The variation of (43) with respect to the costate function being trivial, we restrict the attention here only to those with respect to the probability density $m$ and the non-equilibrium Helmholtz energy density $A$. The boundary terms generated by the variation of $m$ vanish because of the boundary conditions (44):

$$
A'_m(m, V, A) = \int_{t_0}^{t_f} \frac{dt}{\tau} \int_{\mathbb{R}^{2d}} d^2x \ m \left\{ \|\partial_x A\|^2_G + [\tau \partial_t + (\partial_x A) \cdot (J^\dagger - G) \cdot \partial_x ] V \right\}
$$

which readily yields (48a). The variation of $A$ reduces after an integration by parts to

$$
A'_A(m, V, A) = -\int_{t_0}^{t_f} \frac{dt}{\tau} \int_{\mathbb{R}^{2d}} d^2x \ A' \partial_x \cdot m \left\{ 2G \cdot \partial_x A - (J^\dagger - G) \cdot \partial_x V \right\}.
$$

Recalling the definition of the microscopic entropy (4), we see that stationarity of (B.2) admits a geometric interpretation on the De Rahm–Witten complex over $L^2(\mathbb{R}^{2d}, m \text{d}^2x)$ [50] equipped with the exterior derivative

$$
d_S = e^{-S} \text{d} e^S.
$$

Namely it states that the dual $d_S^\ast$ to (B.3) must annihilate the 1-form

$$
\alpha = [2\partial_x A + (J + G) \cdot \partial_x V] \cdot \text{d} x.
$$

In terms of the operator (49) the condition translates into (48c). We also notice that that (49) is a degenerate ‘Witten’ Laplacian [50] on the same complex in consequence of the inequality

$$
\int_{\mathbb{R}^{2d}} d^2x \ m \ S f = -\int_{\mathbb{R}^{2d}} d^2x \ m \|\partial_x f\|^2_G \leq 0
$$

holding for any $f \in L^2(\mathbb{R}^{2d}, m \text{d}^2x)$.

We end this this appendix with a remark. If the nullspace in $L^2(\mathbb{R}^{2d}, m \text{d}^2x)$ of the Witten Laplacian consists only of constant functions then on the De Rahm–Witten complex (B.3) the current velocity (17) admits the Hodge decomposition

$$
\nu = -\partial_x H_0 + h_2
$$

where $H_0$ is a differentiable phase–space function specified by the solution of

$$
\mathcal{G} H_0 = -(J \cdot \partial_x S) \cdot \partial_x A + \mathcal{G} A
$$

and

$$
h_2 \equiv e^S \partial_x e^{-S} H_2 \quad \text{(B.9)}
$$

doi:10.1088/1742-5468/2014/05/P05013
with $H_2$ being a differentiable anti-symmetric rank-two tensor. By construction the elements of the decomposition in (B.7) are orthogonal in $L^2(\mathbb{R}^{2d}, m \, dx)$. 

There are two interesting consequences of (B.7). The first is that mass-transport equation for $m$ depends only upon $H_0$ owing to 

$$\partial_x \cdot \left( m e^{S} \partial_x \cdot e^{-S} H_2 \right) = \beta^d \partial_x \otimes \partial_x \cdot e^{-S} H_2 = 0. \quad (B.10)$$

The second is that identifying the gradient in (B.7) as the dissipative component of the dynamics allows us to define the ‘entropy production’ 

$$\tilde{E}_{t_1,t_0} = \beta \int_{t_0}^{t_1} \frac{dt}{\tau} \int_{\mathbb{R}^{2d}} m \parallel \partial_x H_0 \parallel^2. \quad (B.11)$$

At variance with (18), (B.11) is a coercive functional of $H_0$ the optimal control whereof reduces by (B.10) to that of the Langevin–Smoluchowski case in $\mathbb{R}^{2d}$. It must be stressed, however, that (B.11) carries different physical information from that in (18) since this latter depends also on $h_2$.

References

[1] Ritort F, Nonequilibrium fluctuations in small systems: from physics to biology, 2008 Advances in Chemical Physics vol 137, ed S A Rice (Hoboken, NJ: Wiley) [arXiv:0705.0455]
[2] Haynie D T, 2008 Biological Thermodynamics 2nd edn (Cambridge: Cambridge University Press)
[3] Savir Y and Tlusty T, 2013 Cell 153 471 (www.weizmann.ac.il/complex/tlusty/research.html)
[4] Bennett C H, 1982 Int. J. Theor. Phys. 21 905
[5] Landauer R, 1961 IBM J. Res. Dev. 5 183
[6] Bérut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R and Lutz E, 2012 Nature 483 187
[7] Aurell E, Gawędzki K, Mejía-Monasterio C, Mohayaee R and Muratore-Ginanneschi P, 2012 J. Stat. Phys. 147 487 [arXiv:1201.3207]
[8] Aurell E, Mejía-Monasterio C and Muratore-Ginanneschi P, 2011 Phys. Rev. Lett. 106 250601 [arXiv:1012.2037]
[9] Aurell E, Mejía-Monasterio C and Muratore-Ginanneschi P, 2012 Phys. Rev. E 85 020103(R) [arXiv:1111.2876]
[10] Gawędzki K, 2013 Fluctuation Relations in Stochastic Thermodynamics Lecture Notes [arXiv:1308.1518]
[11] Muratore-Ginanneschi P, Mejía-Monasterio C and Peliti L, 2013 J. Stat. Phys. 150 181 [arXiv:1203.4062]
[12] Alemany A, Ribezi M and Ritort F, 2011 AIP Conf. Proc. 1332 96 [arXiv:1101.3174]
[13] Sekimoto K, 1998 Prog. Theor. Phys. Suppl. 130 17
[14] Sekimoto K, 2010 Stochastic Energetics (Lecture Notes in Physics vol 799) (Berlin: Springer)
[15] Jiang D Q, Qian M and Qian M P, 2004 Mathematical Theory of Nonequilibrium Steady States (Lecture Notes in Mathematics vol 1833) (Berlin: Springer)
[16] Muratore-Ginanneschi P, 2013 J. Phys. A: Math. Gen. 46 275002 [arXiv:1210.1133]
[17] Fleming W H and Soner M H, 2006 Controlled Markov Processes and Viscosity Solutions 2nd edn (Stochastic Modelling and Applied Probability vol 25) (Berlin: Springer)
[18] Lebowitz J L and Spohn H, 1999 J. Stat. Phys. 95 333 [arXiv:cond-mat/9811220]
[19] Maes C, Redig F and Vlieberg A V, 2000 J. Math. Phys. 41 1528
[20] Chetrite R and Gawędzki K, 2008 Commun. Math. Phys. 282 469 [arXiv:0707.2725]
[21] Nelson E, 1985 Quantum Fluctuations (Princeton Series in Physics) (Princeton, NJ: Princeton University Press) https://web.math.princeton.edu/~nelson/books.html
[22] Meyer P A, 1982 Sém. Probab. Strasbourg S16 165
[23] Zambrini J C, 1986 J. Math. Phys. 27 2307
[24] Kulback S and Leibler R, 1951 Ann. Math. Stat. 22 79
[25] Fleming W and Rishel R, 1975 Deterministic and Stochastic Optimal Control (Applications of Mathematics) (Berlin: Springer)
On extremals of the entropy production by ‘Langevin–Kramers’ dynamics

[26] Malliaris A G and Brock W A, 1999 Stochastic Methods in Economics and Finance 8th edn (Advanced Textbooks in Economics vol 17) (Amsterdam: Elsevier Science)

[27] Evans L C, 2010 An Introduction to Mathematical Optimal Control Theory (Lecture Notes) (Berkeley, CA: University of California at Berkeley) (http://math.berkeley.edu/~evans/control.course.pdf)

[28] Guerra F and Morato L M, 1983 Phys. Rev. D 27 1774

[29] van Handel R, 2007 Stochastic Calculus and Stochastic Control (Lecture Notes) California Institute of Technology (http://www.princeton.edu/~rvan/acm217/ACM217.pdf)

[30] Bensoussan A, Lions J L and Papanicolaou G, 1978 Asymptotic Analysis for Periodic Structures (Studies in Mathematics and its Applications vol 5) (Amsterdam: North-Holland)

[31] Pavliotis G A and Stuart A M, Multiscale methods: averaging and homogenization, 2008 Texts in Applied Mathematics vol 53 (Berlin: Springer)

[32] Villani C, 2009 Optimal Transport: Old and New (Grundlehren der Mathematischen Wissenschaften vol 338) (Berlin: Springer) (http://www.umpa.ens-lyon.fr/~cvillani/surveys.html#oldnew)

[33] Rey-Bellet L, Ergodic properties of Markov processes, 2006 Quantum Open Systems II. The Markovian Approach (Lecture Notes in Mathematics vol 1881) (Berlin: Springer) p 1 (http://www.math.umass.edu/~lr7q/ps_files/EMP.pdf)

[34] Shannon C E, 1948 The Bell Syst. Tech. J. 27 379

[35] Morrison P J, 1986 Physica D 18 410

[36] Morrison P J, 2009 J. Phys.: Conf. Ser. 169 012006

[37] Nelson E, 2001 Dynamical Theories of Brownian Motion 2nd edn (Princeton, NJ: Princeton University Press) (https://web.math.princeton.edu/~nelson/books.html)

[38] Jarzynski C, 1997 Phys. Rev. Lett. 78 2690 [arXiv:cond-mat/9610209]

[39] Klebaner F C, 2005 Introduction to Stochastic Calculus With Applications 2nd edn (London: Imperial College Press)

[40] Evans D J and Searles D J, 1994 Phys. Rev. E 50 1645

[41] Gallavotti G and Cohen E G D, 1995 Phys. Rev. Lett. 74 2694 [arXiv:chao-dyn/9410007]

[42] Kurchan J, 1998 J. Phys. A: Math. Gen. 31 3719 [arXiv:cond-mat/9709304]

[43] Crooks G E, 1999 Phys. Rev. E 60 2721 [arXiv:cond-mat/9901352]

[44] Chetrite R and Gupta S, 2011 J. Stat. Phys. 143 543 [arXiv:1009.0707]

[45] Coulibaly-Pasquier K A, 2011 Ann. Inst. Henri Poincare 47 515 [arXiv:hal-00352805]

[46] Struchtrup H, 2005 Macroscopic Transport Equations for Rarefied Gas Flows: Approximation Methods in Kinetic Theory (Interaction of Mechanics and Mathematics vol 258) (Berlin: Springer)

[47] Bloch A M, Crouch P E, Holm D D and Marsden J E, An optimal control formulation for inviscid incompressible ideal fluid flow, 2000 Proc. 39th IEEE Conference on Decision and Control (Piscataway, NJ: IEEE) [arXiv:nlin/0103042]

[48] Dimentberg M F, Iourtchenko D V and Bratus A, 2000 Probab. Eng. Mech. 15 381

[49] Gomez-Marin A, Schmiedt T and Seifert U, 2008 J. Chem. Phys. 129 024114 [arXiv:0803.0269]

[50] Helffer B, 2002 Semiclassical Analysis, Witten Laplacians and Statistical Mechanics (Series on Partial Differential Equations and Applications vol 1) (Singapore: World Scientific)