RELATIVE ENTROPIES AND HYPOCOERCIVITY REVISITED

ANTON ARNOLD, CHRISTIAN SCHMEISER

Abstract. Based on the decay of relative entropies for the probability distributions of Markov processes, different methods for proving decay to equilibrium have been established. The standard situation is coercivity of the entropy dissipation, called Poincaré inequality in analogy to diffusion processes, which is either proved directly or as a consequence of the Bakry-Emery approach via the decay of the entropy dissipation. Even in the absence of the Poincaré inequality decay to equilibrium is still possible, a situation called hypocoercivity.

A short formal derivation of relative entropy dissipation via a local version is given. The connection to the Γ-calculus, to the time reversal of diffusion processes, and several other applications are addressed. Finally, a number of recently developed approaches for proving hypocoercivity are presented for the prototypical model problem, the kinetic Fokker-Planck equation.

This is work in progress.

1. Continuous time stochastic processes

This section is a ridiculously short and incomplete collection of some important notions, which will be used in this course. See, e.g., [8] for much more.

Definition 1. Let \((\Omega, \mathcal{F}, P)\) be a probability space. A continuous time stochastic process (CTSP) in \(\mathbb{R}^d\) is a family of random vectors \(\{X_t\}_{t \in [0, \infty)}\), i.e.

\[ X_t = (X_{1,t}, \ldots, X_{d,t}) : \Omega \to \mathbb{R}^d \quad \text{and} \]

\[ \{\omega \in \Omega : X_{i,t}(\omega) \leq \alpha\} \in \mathcal{F}, \quad i = 1, \ldots, d, \quad t \geq 0, \quad \alpha \in \mathbb{R}. \]

For fixed \(\omega \in \Omega\), the map \(t \mapsto X_t(\omega)\) is called a sample path.

The Borel \(\sigma\)-algebra \(\mathcal{B}_d\) is the smallest \(\sigma\)-algebra on \(\mathbb{R}^d\) containing all sets of the form \((-\infty, \alpha_1] \times \cdots \times (-\infty, \alpha_d]\). A random vector \(X\) induces the probability space \((\mathbb{R}^d, \mathcal{B}_d, P_X)\), where the probability

\[ P_X(A) = P(\{\omega \in \Omega : X(\omega) \in A\}), \quad A \in \mathcal{B}_d. \]

is called the law of \(X\). For CTSPs we shall assume that the law is given in terms of a probability density \(p(x, t)\), \(x \in \mathbb{R}^d, t \geq 0\):

\[ P_X(A) = \int_A p(x, t)dx. \]

Expectation values are then given by

\[ E^{p(t)}(f) := E(f(X_t)) = \int_{\mathbb{R}^d} f(x)p(x, t)dx =: \langle fp(t) \rangle \]

Date: November 7, 2017.
Obviously, probability densities \( p \) have to satisfy \( p \geq 0 \) and \( \langle p \rangle = 1 \). An example for \( d = 1 \) is the Gaussian
\[
p(x) = (2\pi\sigma^2)^{-1/2} \exp \left( \frac{(x - \mu)^2}{2\sigma^2} \right)
\]
with mean \( \mu \in \mathbb{R} \) and variance \( \sigma^2 > 0 \) (i.e., if the Gaussian is the density of \( X \), then \( E(X) = \mu, E((X - \mu)^2) = \sigma^2 \)). For random vectors \( X \) and \( Y \), the probability distributions of \( X \) and of \( Y \) are called marginals of the joint distribution for \( (X,Y) \).

In terms of the probability densities \( p_{(X,Y)} \) for \( (X,Y) \), \( p_X \) for \( X \), and \( p_Y \) for \( Y \),
\[
p_x(x) = \int_{\mathbb{R}^d} p_{(X,Y)}(x,y) dy, \quad p_y(y) = \int_{\mathbb{R}^d} p_{(X,Y)}(x,y) dx
\]
hold. The conditional expectation is given by
\[
E(f(X)|Y = y) = \int_{\mathbb{R}^d} f(x) \frac{p_{(X,Y)}(x,y)}{p_Y(y)} dx.
\]
This formula has to be taken with care. Strictly speaking, it can of course only be used when \( p_Y > 0 \). A general definition of conditional expectation is rather involved.

The random vectors \( X \) and \( Y \) are called independent, if
\[
E(f(X)|Y = y) = E(f(X)) = \int_{\mathbb{R}^d} f(x)p_X(x) dx \quad \forall y \in \mathbb{R}^d.
\]
This holds of course, if \( p_{(X,Y)} \) is factorized: \( p_{(X,Y)}(x,y) = p_X(x)p_Y(y) \).

A CTSP has independent increments, if for every \( n \in \mathbb{N} \) and \( 0 \leq t_1 < \ldots < t_n \), the random vectors \( X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}} \) are mutually independent.

An important class of CTSPs are martingales. They satisfy
\[
E(|X_t|) < \infty, \quad E(X_{t+h}|X_t = x) = x \quad \forall t \geq 0, h > 0, x \in \mathbb{R}^d.
\]
If a CTSP has independent increments and constant mean, i.e. \( E(X_t) = E(X_0) \), \( t \geq 0 \), it is a martingale.

A Markov process is a CTSP without memory: For every \( t \geq 0 \), knowledge of the distribution at time \( t \) is sufficient for predicting the process at later times. It possesses a transition probability with density \( \pi_{t,s}(x,y) \), \( t > s \), such that
\[
E(f(X_t)|X_s = y) = \int_{\mathbb{R}^d} f(x)\pi_{t,s}(x,y) dx.
\]
The transition probability density satisfies the Chapman-Kolmogorov equation
\[
\pi_{t,s}(x,y) = \int_{\mathbb{R}^d} \pi_{t,u}(x,z)\pi_{u,s}(z,y) dz, \quad t > u > s.
\]
Every CTSP with independent increments is a Markov process. A Markov process is called homogeneous, if the transition probability density can be written in the form \( \pi_{t-s}(x,y) \). With \( p(x,0) = p_0(x) \), we then have
\[
p(x,t) = (T(t)p_0)(x) := \int_{\mathbb{R}^d} \pi_t(x,y)p_0(y) dy,
\]
where the propagation operator \( \{T(t)\}_{t \geq 0} \) is a semigroup as a consequence of the Chapman-Kolmogorov equation:
\[
T(t+s) = T(t)T(s), \quad t, s \geq 0.
\]
Its generator will be denoted by

$$L^*p := \lim_{h \to 0^+} \frac{T(h)p - p}{h},$$

for all $p$ such that the limit exists (formally). Accordingly, we shall from now on write $T(t) = e^{Lt}$.

In the following we shall sometimes write $\langle f, g \rangle$ instead of $\langle f, g \rangle$, to emphasize the role of $\langle \cdot, \cdot \rangle$ as a duality pairing. The dual $L$ of $L^*$ ($\langle Lp, q \rangle = \langle p, L^*q \rangle$) is called the generator of the (homogeneous) Markov process, and we will try to explain this terminology in the following.

We start by stating some properties of $\{e^{Lt} : t \geq 0\}$ and the semigroup $\{e^{Lt} : t \geq 0\}$ generated by $L$. $e^{Lt}$ acts on the space of bounded measures on $\mathbb{R}^d$ (typically probability measures), and it preserves mass and positivity, such that, when $p_0$ is a probability measure on $M$, then also $p(t) = e^{Lt}p_0$ is a probability measure for all $t \geq 0$, and $p_0 > 0$ implies $p(t) > 0$, i.e.

$$\langle L^*f \rangle = 0, \quad \langle e^{Lt}f \rangle = \langle f \rangle, \quad L1 = 0.$$ 

If $p_0$ is the probability distribution of $X_0$, then the solution $p(t) = e^{Lt}p_0$ of the forward Kolmogorov equation

$$\partial_t p = L^*p,$$

subject to the initial condition $p(0) = p_0$, is the probability distribution of $X_t$, $t \geq 0$. Note that the transition probability density can be written as $\pi_t(x, y) = (e^{Lt}\delta_x)(y)$, where $\delta_y$ denotes the Delta distribution concentrated at $y$. Expectation values can be computed via the Feynman-Kac formula

$$E(f(X_t)) = E^{p(t)}(f) = \langle f, p(t) \rangle = \langle e^{Lt}f, p_0 \rangle.$$ 

For a deterministic initial position, $p_0 = \delta_x$, this produces the recipe $E(f(X_t)) = u(x, 0)$, where $u(x, s)$ is the solution of the backward Kolmogorov equation

$$\partial_s u + Lu = 0, \quad 0 \leq s \leq t,$$

subject to the terminal condition $u(x, t) = f(x)$. This can be seen as a consequence of $u_t := u(X_t, t)$ being a martingale, when $u$ solves the backward Kolmogorov equation, i.e.

$$\frac{d}{dt}E(u_t) = \frac{d}{dt}\langle up \rangle = 0.$$ 

We shall need a more general version of the Feynman-Kac formula, employing the conditional probability density

$$\pi_{t-s}(\cdot | y) = e^{L^*(t-s)}\delta_y$$

for $(X_t | X_s = y), t \geq s \geq 0$, giving the joint probability density

$$(e^{L^*(t-s)}\delta_y)(x)p(s, y)$$

for $X_t = x$ and $X_s = y$. Therefore

$$E(f(X_t)g(X_s)) = \int_M \int_M f(x)g(y)(e^{L^*(t-s)}\delta_y)(x)p(s, y)dx \, dy$$

(3) $$\quad = \langle f, e^{L^*(t-s)}(gp(s)) \rangle = \langle e^{L(t-s)}f, gp(s) \rangle, \quad t \geq s \geq 0.$$
This can be used in a straightforward computation, producing an explanation for calling \( L \) the generator of the Markov process:

\[
0 = E \left( f(X_t) - f(X_s) - \int_s^t Lf(X_r)dr \right) g(X_s), \quad t \geq s \geq 0.
\]

2. Example 1: diffusion processes

Let \( \{\xi_j : j \in \mathbb{N}\} \) be a sequence of independent identically distributed random variables with \( E(\xi_j) = 0, \ E(\xi_j^2) = 1, \ j \in \mathbb{N}, \) and let

\[
S_k := \sum_{j=1}^k \xi_j, \quad k \geq 1,
\]

implying \( E(S_k) = 0, \ E(S_k^2) = k. \) Now define a CTSP \( B^n_k \) by \( B^n_0 = 0, \)

\[
B^n_k/n := n^{-1/2} S_k, \quad k \in \mathbb{N},
\]

and linear interpolation between these gridpoints, giving \( E(B^n_k) = 0, \ E((B^n_k)^2) \approx t \) as \( n \to \infty. \) These properties are essentially enough to prove that \( B^n_k \) tends to a (weak) limit \( B_t \) as \( n \to \infty. \) The CTSP \( B_t \) is called one-dimensional Brownian motion. The computation

\[
\frac{B^n_k/n - B^n_l/n}{\sqrt{k/n - l/n}} = \frac{S_k - S_l}{\sqrt{k-l}},
\]

suggests Hölder continuity of \( B_t \) with exponent \( 1/2, \) since the right hand side has variance \( 1. \) Actually it can be proven that sample paths of Brownian motion are almost surely (i.e., with probability 1) Hölder continuous with any exponent smaller than \( 1/2 \) and almost surely not Hölder continuous with any exponent bigger than \( 1/2. \) On the other hand the above equation can also be written as

\[
B^n_k/n - B^n_l/n = \frac{1}{\sqrt{n}} \sum_{j=l+1}^k \xi_j,
\]

which, by the fact that all \( \xi_j \) have the same law, implies that \( B^n_{k/n} - B^n_{l/n} \) has the same law as \( B^n_{(k-l)/n}. \) In the limit \( n \to \infty \) this leads to the result that \( B_t \) has stationary increments. One consequence of this, which shall be used below, is

\[
E((B_{t+\Delta t} - B_t)^2) = E(B^2_{\Delta t}) = \Delta t.
\]

The central limit theorem implies that for fixed \( t \) the distribution of \( B_t \) is Gaussian with mean 0 and variance \( t: \)

\[
\pi_t(x,0) = (2\pi t)^{-1/2} e^{-x^2/(2t)}.
\]

The notation indicates that Brownian motion is an homogeneous Markov process with transition probability density

\[
\pi_t(x,y) = (2\pi t)^{-1/2} e^{-(x-y)^2/(2t)}.
\]

This is the fundamental solution of the heat (or diffusion) equation \( \partial_t p = \frac{1}{2} \partial_x^2 p, \) showing that one-dimensional Brownian motion is generated by the second derivative:

\[
L = L^* = \frac{1}{2} \partial_x^2.
\]
Another standard way to derive this starts with choosing \( \xi_j \in \{-1, 1\} \) at the beginning of this section. This implies that only points on the grid \( \{k/\sqrt{n} \colon k \in \mathbb{Z}\} \) can be reached. Then one writes a system of difference equations for the probabilities on this discrete probability space and passes to the limit \( n \to \infty \).

The CTSP \( B_t = (B_{1,t}, \ldots, B_{d,t}) \) with independent one-dimensional Brownian motions \( B_{1,t}, \ldots, B_{d,t} \) is called \( d \)-dimensional Brownian motion. The probability distribution of \( B_t \) is the \( d \)-dimensional Gaussian

\[
\pi_t(x,0) = (2\pi)^{-d/2} e^{-|x|^2/(2t)},
\]

and it is generated by the Laplace operator: \( L = \Delta_x / 2 \).

Solutions of an autonomous stochastic differential equation (SDE) of the form

\[
dX_t = b(X_t)dt + \sigma dB_t
\]

with a vector field \( b \) and a \( d \times d \)-matrix \( \sigma \) are called diffusion processes. We shall assume \( \sigma \) to be invertible. The equation is usually written in this form (not 'divided' by \( dt \)) since \( B_t \) is not differentiable. Its actual meaning is given by the integral equation

\[
X_t = X_0 + \int_0^t b(X_s)ds + \sigma B_t,
\]

(note that \( B_0 = 0 \)). Since \( B_t \) is continuous almost surely, short time existence and uniqueness of continuous sample paths (with \( X_0(\omega) \) and \( B_t(\omega) \)) is guaranteed almost surely for Lipschitz continuous \( b \). As for ODEs, this can be shown by Picard iteration.

Often the volatility matrix \( \sigma \) is allowed to depend on the state, i.e., \( \sigma = \sigma(X_t) \), which makes the interpretation of (6) significantly more complicated. The term \( \sigma B_t \) in the integral formulation then has to be replaced by the stochastic integral \( \int_0^t \sigma(X_s)dB_s \), whose definition and properties are rather different from standard integrals because of the roughness of \( B_t \). In the following we continue with the assumption of constant \( \sigma \) in order to stay away from these difficulties.

In order to find the generator of a diffusion process we use (1), (2). So we assume that \( u_t = u(X_t, t) \) is a martingale and compute for small \( \Delta t > 0 \)

\[
\frac{u_{t+\Delta t} - u_t}{\Delta t} = \frac{u(X_{t+\Delta t}, t + \Delta t) - u(X_t, t)}{\Delta t} \approx \partial_t u(X_t, t) \Delta t + \nabla_x u(X_t, t) \cdot (X_{t+\Delta t} - X_t) + \frac{1}{2} \nabla^2_x u(X_t, t)(X_{t+\Delta t} - X_t).
\]

Noting that

\[
X_{t+\Delta t} - X_t = b(X_t)\Delta t + \sigma(B_{t+\Delta t} - B_t) + O(\Delta t^2),
\]

and, motivated by (5), considering \( B_{t+\Delta t} - B_t \) as an \( O(\sqrt{\Delta t}) \)-term, the approximation in (7) means that terms up to \( O(\Delta t) \) have been kept. The approximation can be simplified further:

\[
\frac{u_{t+\Delta t} - u_t}{\Delta t} = \left( \partial_t u(X_t, t) + b(X_t) \cdot \nabla_x u(X_t, t) \right) \Delta t + \nabla_x u(X_t, t) \cdot (\sigma(B_{t+\Delta t} - B_t)) + \frac{1}{2} \nabla^2_x u(X_t, t)(\sigma(B_{t+\Delta t} - B_t)) + O(\Delta t^{3/2}).
\]

Now we compute the expectation (using the martingale property of \( u_t \), as well as (5) in the second line), divide by \( \Delta t \) and pass to the limit \( \Delta t \to 0 \):

\[
0 = \partial_t u + Lu, \quad \text{with } Lu = b \cdot \nabla_x u + \nabla_x \cdot (D\nabla_x u), \quad D = \frac{1}{2} \sigma^{tr} \sigma.
\]
This is the backward Kolmogorow equation for the diffusion process. Actually it can also be derived by first passing to the limit $\Delta t \to 0$ in (8), giving the result of the Ito Lemma

$$du_t = (\partial_t u + Lu)dt + \nabla_x u \cdot (\sigma dB_t),$$

which, however, requires the rather sophisticated proof of the formal equation $(dB_{i,t})^2 = dt$, which we only have used for the expectation of the left hand side.

The law of the diffusion process solves the forward Kolmogorow equation

$$\partial_t p = L^* p = \nabla_x \cdot (D \nabla_x p - bp).$$

3. Example 2: jump processes

Let $\lambda > 0$, $\mathbb{N} \ni n > \lambda$, and let $\{\xi_j : j \in \mathbb{N}\}$ be a sequence of independent identically distributed random variables with $\xi_j \in \{0, 1\}$, $P(1) = \lambda/n$. We define a CTSP by

$$N^n_t := \sum_{j=1}^{k} \xi_j, \quad \frac{k}{n} \leq t < \frac{k+1}{n},$$

implying that for every $t \geq 0$, $N^n_t$ is a nonnegative integer and, by elementary combinatorics,

$$P(N^n_t = l) = \binom{k}{l} \left(\frac{\lambda}{n}\right)^l \left(1 - \frac{\lambda}{n}\right)^{k-l}, \quad \frac{k}{n} \leq t < \frac{k+1}{n}.$$  

As in the previous section, the uniform (as $n \to \infty$) boundedness of these moments can be used to pass to a weak limit in the CTSP, producing the homogeneous Poisson process $N_t$ with intensity $\lambda$, satisfying

$$P(N_t = l) = \frac{(\lambda t)^l}{l!} e^{-\lambda t},$$

which can be obtained from (9) by using $k \approx nt$ and passing to the limit. The Poisson process is a Markov process with stationary independent increments.

If the (random) time, where $N_t$ jumps from $l$ to $l + 1$ is denoted by $T_l$, then the lengths $T_{l+1} - T_l$ of the time intervals between jumps have independent identical distributions, given by the Exponential$(\lambda)$ law:

$$P(T_{l+1} - T_l > \Delta t) = e^{-\lambda \Delta t}.$$  

Considering the state space $\mathbb{R}$, the probability density of the law of $N_t$ can be written as

$$p(x,t) = e^{-\lambda t} \sum_{l=0}^{\infty} \frac{(\lambda t)^l}{l!} \delta_l(x).$$

Differentiation with respect to time gives

$$\partial_t p(x,t) = \lambda (p(x-1,t) - p(x,t)) = (L^* p(t))(x),$$

and, thus, the generator of the Poisson process is given by

$$(Lu)(x) = \lambda (u(x+1) - u(x)).$$
The Poisson process is the simplest example of a jump process. In the following we shall be interested in more general compound Poisson processes, which can be seen as generalizations in two directions. First, we allow state dependent intensities \( \lambda(x) \geq 0 \), meaning that after a jump taking us to state \( x \in \mathbb{R}^d \), the time to the next jump is Exponential(\( \lambda(x) \)) distributed. At the time of the jump, the new state \( x' \) is also chosen randomly, according to the probability density \( k(x \to x') \geq 0 \), \( \int_{\mathbb{R}^d} k(x \to x') \, dx' = 1 \). This leads to

\[
(L^*p)(x) = \int_{\mathbb{R}^d} \lambda(x') k(x' \to x) p(x') \, dx' - \lambda(x) p(x).
\]

This is often written in terms of the rate \( W(x \to x') := \lambda(x) k(x \to x') \):

\[
(L^*p)(x) = \int_{\mathbb{R}^d} (W(x' \to x) p(x') - W(x \to x') p(x)) \, dx.
\]

For jump processes the forward Kolmogorov equation \( \partial_t p = L^*p \) is often called the master equation. The generator of the jump process is given by

\[
(Lu)(x) = \int_{\mathbb{R}^d} W(x \to x')(u(x') - u(x)) \, dx'.
\]

4. Time reversal – decay of the relative entropy

We are looking for a time reversed version of (4) and compute

\[
E(\langle f(X_t) - f(X_s) \rangle | g(X_t)) = \langle g, f(p(t)) - e^{L(t-s)} f(p(s)) \rangle
\]

\[
= \int_s^t \left( \langle -Le^{L(t-r)} g, f(p(r)) \rangle + \langle e^{L(t-r)} g, f L^* p(r) \rangle \right) \, dr
\]

\[
= -\int_s^t \langle e^{L(t-r)} g, p(r) \Gamma^{p(t)} f \rangle \, dr,
\]

with

\[
\Gamma^{p(t)} f = \frac{1}{p(t)} (L^*(fp(t)) - fL^* p(t)),
\]

implying

\[
0 = E \left( \left[ f(X_t) - f(X_s) + \int_s^t \Gamma^{p(r)} f(X_r) \, dr \right] g(X_t) \right), \quad t \geq s \geq 0.
\]

The operator \( \Gamma^{p(t)} \) can be seen as the generator of the time reversal of the process \( \{X_t : t \geq 0\} \) with probability distribution \( \{p(t) : t \geq 0\} \).

If \( X_t \) is stationary, i.e. \( p(t, x) = p_\infty(x) \) with \( L^* p_\infty = 0 \), and if the detailed balance condition

\[
p_\infty L = L^* (p_\infty)
\]

holds, then \( X_t \) is time reversible, i.e. \( \Gamma^{p_\infty} = L \). The computation

\[
(L^* (p_\infty f), g) = \langle Lf, p_\infty g \rangle = \langle f, L^* (p_\infty g) \rangle
\]

shows that detailed balance is equivalent to the symmetry of the operator \( L^* (p_\infty) \).
Let \( p(t) = e^{t\tau}p_0 \) and \( q(t) = e^{t\tau}q_0 \) denote the probability densities of \( X^p_t \) and, respectively, \( X^q_t \). Then a straightforward computation shows that \( p/q \) solves the backward Kolmogorov equation of the time reversal of \( X^q_t \), i.e.,

\[
\partial_t \left( \frac{p}{q} \right) - \Gamma^q \left( \frac{p}{q} \right) = 0,
\]

implying the backward martingale property: For \( t \geq s \),

\[
\frac{p(t, x)}{q(t, x)} = E \left( \frac{p(s, X^q_s)}{q(s, X^q_s)} \right| X^q_s = x).
\]

The following formal computation can be used instead, without explicit referral to time reversal:

\[
\frac{p(t, x)}{q(t, x)} = \frac{1}{q(t, x)} \left( e^{L^*(t-s)}p(s) \right)(x) = \frac{1}{q(t, x)} \left( e^{L^*(t-s)}p(s), \delta_x \right)
\]

\[
= \frac{p(s)}{q(s)} \frac{q(s)}{q(t, x)} e^{L(t-s)}\delta_x.
\]

The second factor on the right hand side can be interpreted as the probability density for \( X^q_t \) under the condition \( X^q_t = x \). For the following, it is only important that it is a probability density:

\[
\frac{1}{q(t, x)} \left( q(s), e^{L(t-s)}\delta_x \right) = \frac{1}{q(t, x)} \left( e^{L^*(t-s)}q(s), \delta_x \right) = 1.
\]

Thus, for a convex function \( U \), the Jensen inequality implies

\[
U \left( \frac{p(t, x)}{q(t, x)} \right) \leq U \left( \frac{p(s)}{q(s)} \frac{q(s)}{q(t, x)} e^{L(t-s)}\delta_x \right),
\]

leading to the pointwise estimate for the local relative entropy

\[
(12) \ h_U(p(t)|q(t))(x) := q(t, x)U \left( \frac{p(t, x)}{q(t, x)} \right) \leq \left[ e^{L^*(t-s)}h_U(p(s)|q(s)) \right](x).
\]

Integration with respect to \( x \) gives the decay of the entropy of \( p \) relative to \( q \),

\[
H_U(p(t)|q(t)) := \langle h_U(p(t)|q(t)) \rangle \leq \left( e^{L^*(t-s)}h_U(p(s)|q(s)) \right) = H_U(p(s)|q(s)),
\]

by the conservation property of the forward Kolmogorov semigroup. For the typical choices of \( U \), satisfying \( U(1) = U'(1) = 0 \), the local relative entropy \( h_U(p|q) \) is a measure for the distance between \( p \) and \( q \).

Example 1. (Diffusion processes – the Fokker-Planck equation) The generator of the time reversal of the diffusion process \[13\] is, according to \[10\], given by

\[
\Gamma^p f = \nabla_x \cdot (D\nabla_x f) + \nabla_x f \cdot \left( \frac{2D\nabla_x p}{p} - b \right).
\]

The time reversed process \( \overline{X}_\tau = X_{T-\tau} \) can be seen as governed by the SDE system

\[
d\overline{X}_\tau = \left( \frac{2D\nabla_x p}{p} - b \right)(\overline{X}_\tau)d\tau + \sigma dB_\tau.
\]

Remark 1. Some of the uneasiness of PDE people with the idea of time reversal of diffusion is justified, since the coefficient \( \nabla_x p/p \) might blow up at \( \tau = T \) \( (t = 0) \), when the initial density \( p_0 \) is nonsmooth.
Obviously, a stationary diffusion process with distribution $p_{\infty}$ is time reversible, iff $D \nabla_x p_{\infty} = b p_{\infty}$, which is however only possible when $D^{-1} b$ is a gradient field. In this special case, i.e. $b(x) = -D \nabla \Phi(x)$ we also need to assume that $\Phi$ is a confining potential, satisfying $e^{-\Phi} \in L^1(\mathbb{R}^d)$ and shifted such that $(e^{-\Phi}) = 1$. The probability distribution of $X_t$ then satisfies the Fokker-Planck equation
\begin{equation}
\partial_t p = \mathbf{L}^* p = \nabla_x \cdot (D(\nabla_x p + p \nabla_x \Phi)),
\end{equation}
with the unique equilibrium $p_{\infty} = e^{-\Phi}$. The symmetrized form
\begin{equation}
\mathbf{L}^* p = \nabla_x \cdot (D p_{\infty} \nabla_x (p/p_{\infty}))
\end{equation}
of the Fokker-Planck operator immediately implies the detailed balance property
\begin{equation}
\mathbf{L}^* (p_{\infty} f) = \nabla_x \cdot (D p_{\infty} \nabla_x f) = p_{\infty} \mathbf{L} f.
\end{equation}

**Example 2. (Jump processes)** In
\begin{equation}
(\mathbf{L}^* p)(x) = \int_M [W(x' \to x)p' - W(x \to x')p] dx',
\end{equation}
we have adopted the convention of kinetic transport theory to abbreviate $p(x')$ and $p(x)$ by $p'$ and, respectively, $p$. The generator and the time reversed version are given by
\begin{equation}
(\mathbf{L} f)(x) = \int_M W(x \to x')(f' - f) dx'
\end{equation}
and, respectively,
\begin{equation}
(\mathbf{L}^p f)(x) = \int_M \frac{W(x' \to x)p'}{p}(f' - f) dx'
\end{equation}
Time reversibility obviously requires an equilibrium satisfying
\begin{equation}
W(x' \to x)p'_{\infty} = W(x \to x')p_{\infty}, \quad x, x' \in M.
\end{equation}
This is actually the origin of calling this condition detailed balance, since it means that in equilibrium the integrand in the master equation vanishes identically, i.e. each jump $x \to x'$ is balanced by its inverse. In the context of jump processes detailed balance is sometimes also called micro-reversibility.

5. The Entropy Dissipation Rate – Decay to Equilibrium by Poincaré Inequalities

The arguments of the previous section do not provide information about the dissipation rate of the relative entropy. Therefore, we compute the derivative of the right hand side of (12):
\begin{equation}
\partial_s \left( e^{\mathbf{L}^*(t-s)} \left[ q(s) U \left( \frac{p(s)}{q(s)} \right) \right] \right) = -e^{\mathbf{L}^*(t-s)} i_U(p(s)|q(s)),
\end{equation}
with the local Fisher information
\begin{equation}
i_U(p|q) = \mathbf{L}^* \left( q U \left( \frac{p}{q} \right) \right) - U' \left( \frac{p}{q} \right) L^* p - \left( U \left( \frac{p}{q} \right) - \frac{p}{q} U' \left( \frac{p}{q} \right) \right) L^* q \geq 0.
\end{equation}
Integration with respect to $s$ leads to the local (relative) entropy dissipation relation
\begin{equation}
h_U(p(t)|q(t)) - e^{\mathbf{L}^*(t-s)} h_U(p(s)|q(s)) = -\int_s^t e^{\mathbf{L}^*(t-r)} i_U(p(r)|q(r)) dr,
\end{equation}

\textbf{RELATIVE ENTROPIES AND HYPOCOERCIVITY} 9
and a further integration with respect to $x$ to the global result

$$H_U(p(t)|q(t)) - H_U(p(s)|q(s)) = -\int_s^t I_U(p(r)|q(r))dr,$$

with the Fisher information $I_U(p|q) = \langle i_U(p|q) \rangle$.

**Remark 2.** We point out two simple properties of the local Fisher information.

- The local Fisher information is linear in $L^*$, i.e. for a sum of operators the contributions to the local Fisher information can be computed separately.
- For multiplication operators and for differentiation operators, i.e. operators satisfying the Leibniz product rule, the local Fisher information vanishes.

When an equilibrium $p_\infty$ with $L^*p_\infty = 0$ exists, and the Poincaré inequality

$$I_U(p|p_\infty) \geq \lambda H_U(p|p_\infty), \quad \text{with } \lambda > 0 \quad (16)$$

holds, then $p(t) = e^{L^*t}p_0 \rightarrow p_\infty$ as $t \rightarrow \infty$ exponentially in the sense of relative entropy.

**Example 3. (Diffusion processes)** For the diffusion process of Example 1 with $L^*p = \nabla_x \cdot (D\nabla_x p) - p\nabla_x \cdot b - b \cdot \nabla_x p$ only the diffusion term contributes to the local Fisher information, since the second and third terms are multiplication and differentiation operators, respectively. We obtain

$$i_U(p|q) = qU'' \left( \frac{p}{q} \right) \nabla_x \left( \frac{p}{q} \right)^{\text{tr}} D\nabla_x \left( \frac{p}{q} \right) \geq \kappa qU'' \left( \frac{p}{q} \right) \left| \nabla_x \left( \frac{p}{q} \right) \right|^2,$$

with the coercivity constant $\kappa > 0$ of the diffusion matrix. When $D^{-1}b$ is not a gradient field and detailed balance is impossible, showing the existence of a positive equilibrium $p_\infty$ can be difficult (see [1] for some results), and we shall assume it in the following.

For $U(z) = (z-1)^2/2$, the inequality (16) is a consequence of the weighted Poincaré inequality

$$\kappa \left\| \nabla_x u \right\|^2 p_\infty \geq \lambda \langle u^2 p_\infty \rangle \quad \text{for } \langle up_\infty \rangle = 0,$$

which holds, if

$$\liminf_{|x| \rightarrow \infty} \frac{\log(1/p_\infty(x))}{|x|} > 0.$$

In the detailed balance situation $p_\infty = e^{-\Phi}$ this means that the confining potential grows at least linearly.

**Example 4. (Jump processes)** The jump process of Example 2 has the local Fisher information

$$i_U(p|q)(x) = \int_M W(x' \rightarrow x)q' \left[ U \left( \frac{p'}{q'} \right) - U \left( \frac{p}{q} \right) - U' \left( \frac{p}{q} \right) \left( \frac{p'}{q'} - \frac{p}{q} \right) \right] dx'.$$

It is not completely straightforward to prove existence of an equilibrium $p_\infty$. For the case of local micro-reversibility, i.e. $W(x' \rightarrow x)p_\infty' = W(x \rightarrow x')p_\infty$ for all $x, x' \in M$, entropy dissipation has been known (in kinetic transport theory, where $x$ represents velocity) for a long time. Without micro-reversibility it is more recent knowledge in the mathematics community [7] (see, however, [24]).
For $U(z) = (z - 1)^2/2$, the local Fisher information is given by

$$i_U(p|q)(x) = \frac{1}{2} \int_M W(x' \to x) q' \left( \frac{p}{q} - \frac{p'}{q'} \right)^2 dx'.$$

With the equilibrium $p_\infty$ and with the standard assumption $W(x' \to x) \geq \lambda p_\infty(x)$, $\lambda > 0$, we have the local Poincaré inequality

$$i_U(p|p_\infty)(x) \geq \frac{\lambda p_\infty}{2} \left( \left( \frac{p}{p_\infty} - 1 \right)^2 + \int_M \frac{p'^2}{p_\infty} dx' - 1 \right) \geq \lambda h_U(p|p_\infty)(x).$$

Integration gives the Poincaré inequality [16] and therefore exponential decay of the relative entropy but, by [14], actually the stronger local result

$$h_U(p(t)|p_\infty)(x) \leq e^{-\lambda t} h_U(p_0|p_\infty)(x).$$

Under the additional assumption $W(x' \to x) \leq \Lambda p_\infty(x)$ this can be put to use in the following way: In this case it is easy to show that a bound of the form $p_0(x) \leq c p_\infty(x)$ is preserved by the evolution. Since then also $h_U(p_0|p_\infty)$, satisfies such a bound, the same holds for $e^{\Lambda t} h_U(p_0|p_\infty)$. The local decay result then implies

$$|p(x, t) - p_\infty(x)| \leq C e^{-\lambda t/2} p_\infty(x), \quad \text{for all } x,$$

for an appropriate constant $C > 0$.

If $M = \mathbb{R}^d$ and an accelerating force $F \in \mathbb{R}^d$ is added, the resulting forward Kolmogorov operator

$$(L^* p)(x) = - F \cdot \nabla_x p + \int_M [W(x' \to x) p' - W(x \to x') p] dx'$$

produces the same local Fisher information and, thus, the distance between two solutions tends to zero, although it is known that an equilibrium does not exist, if the scattering rate $\lambda(x) = \int_M W(x \to x') dx'$ decays too fast as $|x| \to \infty$ [23]. This effect is called run-away. The result that the distance between solutions shrinks even in this case does not seem to be known in the kinetic theory community, and the entropy decay, in case an equilibrium exists, is again rather recent knowledge (see, e.g. [4]).

Example 5. (Growth and fragmentation of biopolymers) In the previous example, a strictly positive jump rate $W$ has been assumed. The following will show that this is not necessary for the validity of the Poincaré inequality, although the local version fails.

A typical situation for linear actin polymers in biological cells is that they are anchored at one end to a protein complex promoting polymerization. On the other hand there is a severing process, cutting polymers at random positions. The cut-off pieces are then depolymerized completely in a fast process. A simple model for the length distribution of the active polymers is given by [12] [17]

$$\partial_t p + v \partial_x p = \kappa \left( \int_x^\infty p' dx' - xp \right) = \kappa \int_0^\infty [H(x' - x)p' - H(x - x')p] dx',$$

$x > 0$, subject to the boundary condition $p(0, t) = 0$, with the polymerization speed $v > 0$, the severing rate constant $\kappa > 0$, and the Heavyside function $H$. Note that
this is different from most fragmentation models (see, e.g. [16]), since the cut-off pieces are thrown away. The unique equilibrium distribution is given by

\[ p_\infty(x) = \frac{\kappa x}{v} \exp \left( -\frac{\kappa x^2}{2v} \right). \]

With \( U(z) = (z-1)^2/2 \), we can take the local Fisher information from the preceding example, since the transport term does not contribute:

\[ i_U(p)p_\infty(x) = \frac{1}{2} \int_M H(x' - x)p'_\infty \left( \frac{p}{p_\infty} - \frac{p'}{p'_\infty} \right)^2 dx'. \]

This does not permit a local Poincaré inequality, but for the global Fisher information symmetrization implies

\[ I_U(p)p_\infty \geq \frac{1}{2} \sqrt{\frac{ve}{\kappa}} \int_M \int_M p'_\infty p_\infty \left( \frac{p}{p_\infty} - \frac{p'}{p'_\infty} \right)^2 dx'dx = 2 \sqrt{\frac{ve}{\kappa}} H_U(p)p_\infty, \]

where the maximal value \( \sqrt{\kappa/(ve)} \) of \( p_\infty \) appears.

**Example 6.** Fractional diffusion is generated by

\[ L^*p = -P.V. \int d\frac{p - p'}{|x - x'|^{d+\alpha}}dx', \]

with \( 0 < \alpha < 2 \), where \( P.V. \) denotes the principal value. Up to a multiplicative constant, this is, at least when applied to smooth functions, equivalent to the pseudo-differential operator \(-(-\Delta)^{\alpha/2}p\) (defined via the Fourier transform). For \( U(z) = z^2/2 \), a straightforward computation gives

\[ i_U(p|q) = \frac{1}{2} \int_{\mathbb{R}^d} q' \left( \frac{p}{q} - \frac{p'}{q'} \right)^2 dx'. \]

The corresponding Poincaré inequality \( I_U(p|q) \geq \lambda H_U(p|q) \) for \( q(x) \) decaying sufficiently fast as \(|x| \to \infty\) has been proven in [21].

**Example 7.** The concept of generalized relative entropy, developed in [18, 19], also fits into the framework of Markov processes after a transformation. For illustration we consider the finite dimensional situation as in [22], Section 6.3. Consider a matrix \( A = (a_{mn})_{m,n=1,...,d} \) with positive entries \( a_{mn} > 0 \). Then the Perron-Frobenius theorem states that the spectral radius \( \lambda \) of \( A \) is a simple eigenvalue with an eigenvector with positive components. Therefore there are unique normalized left and right eigenvectors \( \phi \) and, respectively, \( R \), such that,

\[ A^* \phi = \lambda \phi, \quad AR = \lambda R, \quad \phi_n R_n > 0, \quad \sum_{n=1}^{d} R_n = \sum_{n=1}^{d} R_n \phi_n = 1. \]

The goal is to analyze the long-time behavior of solutions of

\[ \frac{du}{dt} = Au, \quad u_n(0) \geq 0, \quad \sum_{n=1}^{d} u_n(0) > 0. \]
The new unknown \( p(t) \) is defined by

\[
(17) \quad p_n(t) = \frac{e^{-\lambda t} \phi_n u_n(t)}{\sum_{m=1}^d \phi_m u_m(0)},
\]
satisfying, with the diagonal matrix \( \Phi := \text{diag}(\phi_1, \ldots, \phi_d) \),

\[
\frac{dp}{dt} = L^* p, \quad L^* = \Phi A \Phi^{-1} - \lambda, \quad p_n(0) \geq 0, \quad \sum_{n=1}^d p_n(0) = 1.
\]

Since the rows of \( L^* \) add up to zero, \( p(t) \) is a probability distribution on \( \mathbb{R}^d \), and \( L \) generates a Markov process. The unique equilibrium is given by \( p_\infty = \Phi R \). The local relative entropy can be written as

\[
h_U(p|p_\infty)_n = \phi_n R_n \tilde{U} \left( \frac{r_n}{R_n} \right),
\]
where \( r(t) = e^{-\lambda t} u(t) \) and \( \tilde{U} \) is a rescaled version of \( U \). The right hand side is the local generalized relative entropy of \( r \) relative to \( R \). The local Fisher information according to \((15)\) is given by

\[
i_U(p|p_\infty)_m
= \sum_{n=1}^d \Phi_m a_{mn} R_n \left[ U \left( \frac{p_n}{p_{\infty,n}} \right) - U \left( \frac{p_m}{p_{\infty,m}} \right) - U' \left( \frac{p_m}{p_{\infty,m}} \right) \left( \frac{p_n}{p_{\infty,n}} - \frac{p_m}{p_{\infty,m}} \right) \right].
\]

**Example 8.** Generalized relative entropies have originally been introduced for the analysis of age structured population models like

\[
\partial_t u + \partial_a u = -\nu(a) u, \quad u(0, t) = \int_0^\infty B(a) u(a, t) da,
\]
with \( u(\cdot, t) \) the population density with respect to age \( a \) at time \( t \) (see [22]). The \( a \)-derivative in the differential equation describes aging, the right hand side is the death rate with an age dependent rate constant \( \nu(a) \geq 0 \). The right hand side of the boundary condition is the birth rate, again with an age dependent rate constant \( B(a) \geq 0 \).

The ansatz \( u(a, t) = e^{\lambda t} r_\infty(a) \) leads to the equation

\[
\int_0^\infty B(a) \exp \left( -\lambda a - \int_0^a \nu(a') da' \right) da = 1,
\]
whose unique solution \( \lambda \in \mathbb{R} \) is called the *Malthus parameter* and corresponds to the dominating eigenvalue in the previous example. The result corresponding to the Perron-Frobenius theorem, namely the unique solvability of

\[
\lambda R + \partial_a R = -\nu R, \quad R(0) = \int_0^\infty BR da, \quad \int_0^\infty R da = 1,
\]

\[
\lambda \phi - \partial_a \phi = -\nu \phi + \phi(0) B, \quad \int_0^\infty R \phi da = 1,
\]
and \( R, \phi > 0 \) has been proven in [22].

Analogously to \((17)\), we define the new unknown

\[
p(a, t) = \frac{e^{-\lambda t} \phi(a) u(a, t)}{\int_0^\infty \phi(a') u(a', 0) da'},
\]
satisfying
\[ \partial_t p + \phi \partial_a \left( \frac{p}{\phi} \right) = - (\lambda + \nu) p , \quad p(a = 0) = \phi(0) \int_0^\infty B p \, da . \]

It is easily seen that \( p(\cdot, t) \) is a probability density for \( t \geq 0 \) and, again, the entropy of \( p \) relative to \( p_\infty = \phi R \) is the generalized relative entropy if written in terms of \( r(a, t) = e^{-\lambda t} u(a, t) \).

In this case, the formula (15) for the local Fisher information gives zero since the actual dissipation of relative entropy is derived from the boundary terms in integrations by parts (see gain [22] for details):
\[ \frac{d}{dt} \int_0^\infty \frac{p^2}{2 p_\infty} \, da = \frac{\phi(0)}{2 R(0)} \left( \int_0^\infty B R \frac{p}{p_\infty} \, da \right)^2 - \int_0^\infty B R \, da \int_0^\infty B R \left( \frac{p}{p_\infty} \right)^2 \, da , \]
where the right hand side is nonpositive by the Cauchy-Schwarz inequality and vanishes only for \( p = p_\infty \).

6. \( \Gamma \)-calculus – the Bakry-Emery approach

In this section we assume detailed balance, i.e. the existence of a positive equilibrium distribution \( p_\infty (L^* p_\infty = 0) \), satisfying \( p_\infty L f = L^* (p_\infty f) \) for all \( f \). With the observation that then also \( p_\infty e^{Lt} f = e^{L^* t} (p_\infty f) \) holds, the right hand side of (12) (local relative entropy) with \( q = p_\infty \) can be written as
\[ e^{L^*(t-s)} h_U(p(s)|p_\infty) = p_\infty \Psi(t-s) , \quad \text{with } \Psi(s) = e^{L s} U \left( \frac{p(t-s)}{p_\infty} \right) , \]
where the definition of \( \Psi \) is motivated by the \( \Gamma \)-calculus [4], since
\[ \partial_s \Psi(s) = e^{L s} \Gamma_U (u(t-s)) , \]
with
\[ \Gamma_U(u) = L(U(u)) - \partial_t U(u) = L(U(u)) - U'(u) u , \quad \text{with } u = \frac{p}{p_\infty} . \]

For the choice \( U(z) = z^2 / 2, \Gamma_U \) is the carré du champ operator [4], which can be derived as the quadratic form corresponding to the bilinear expression
\[ \frac{1}{2} \left( L(uv) - uLv - vLu \right) . \]

Note that for arbitrary \( U \) the local Fisher information can be expressed in terms of \( \Gamma_U \):
\[ i_U(p|p_\infty) = p_\infty \Gamma_U \left( \frac{p}{p_\infty} \right) . \]

A relation like the Poincaré inequality [10] can in general not be expected between the local quantities \( \Psi \) and \( \partial_s \Psi \) (or between the local relative entropy and the local Fisher information). The Bakry-Emery approach [4] (see also [22]) is based on another differentiation,
\[ \partial_s^2 \Psi(s) = e^{L s} \Gamma_{U,2} (u(t-s)) , \quad \text{with } \Gamma_{U,2}(u) = L \Gamma_U (u) - \partial_t \Gamma_U (u) , \]
and to relate \( \partial_s \Psi \) and \( \partial_s^2 \Psi \).
**Example 9.** For the diffusion process described by the Fokker-Planck equation \([13]\) we assume \(D = I_{d \times d}\), in order to simplify the otherwise very complicated expression for \(\Gamma_{U,2}\). We obtain

\[
\Gamma_U(u) = U^{(2)}(u) |\nabla_x u|^2
\]

and

\[
\Gamma_{U,2}(u) = 2U^{(2)}(u)\nabla_x u^T \nabla_x^2 \Phi \nabla_x u + 2U^{(2)}(u) |\nabla_x u|^2
+ 4U^{(3)}(u) \nabla_x u^T \nabla_x^2 \nabla_x u \nabla_x u + U^{(4)}(u) |\nabla_x u|^4.
\]

It is easily seen that the sum of the last three terms is nonnegative for admissible entropy generating functions \(U\), which are not only convex, but also satisfy the differential inequality

\[
2(U^{(3)})^2 \leq U^{(2)} U^{(4)}.
\]

A family of admissible choices is given by

\[
U_r(z) = z^r - 1 - r(z - 1), \quad 1 < r \leq 2, \quad U_1(z) = z \log z - z + 1,
\]

where the latter is the limiting case of the former as \(r \to 1\).

Under the Bakry-Emery condition \(\nabla^2_x \Phi \geq \lambda/2 > 0\), \(\Gamma_{U,2}(u) \geq \lambda \Gamma_U(u)\) follows, with the consequence \(\partial_t \psi \succeq \lambda \partial_t \psi\). An application of the Gronwall lemma implies the pointwise estimate

\[
\Gamma_U(u)(t) \leq e^{-\lambda t} e^{Ls} \Gamma_U(u)(t-s) \quad \text{for } s \leq t.
\]

Since the semigroup \(e^{Ls}\) conserves the integral with weight \(p_\infty\), integration and \(s = t\) give

\[
I_U(p|p_\infty) \leq e^{-\lambda t} I_U(p_0|p_\infty),
\]

with

\[
I_U(p|p_\infty) = \int_{\mathbb{R}^d} p_\infty U^{(2)} \left( \frac{p}{p_\infty} \right) \left| \nabla_x \left( \frac{p}{p_\infty} \right) \right|^2 dx.
\]

Since the Fisher information \(I_U(p|p_\infty)\) is the time derivative of the relative entropy, integration of this result leads to exponential decay of the relative entropy (using the fact that the relative entropy decays to zero).

### 7. The kinetic Fokker-Planck equation – hypocoercivity

The evolution of probability densities \(f(t, x, v)\), \(x, v \in \mathbb{R}^d\), corresponding to the stochastic differential system

\[
dX = V dt,
\]

\[
dV = -(V + \nabla_x \Phi(X)) dt + \sqrt{2} dB,
\]

is governed by the kinetic transport equation

\[\partial_t f + \mathbf{T} f = \mathbf{C} f,\]

with the transport operator \(\mathbf{T} f = v \cdot \nabla_x f - \nabla_v \Phi \cdot \nabla_v f\), a confining potential \(\Phi(x)\) as above, and the collision operator \(\mathbf{C} f = \nabla_v \cdot (\nabla_v f + v f)\).

**Torus** We shall also consider a variant, referred to as **Torus** in the following, where instead of using a confining potential, confinement is achieved by restriction of the position variable to the \(d\)-dimensional torus \(\mathbb{T}^d_x\), i.e. \((\mathbf{T} f)(x, v) = v \cdot \nabla_x f(x, v)\), \((x, v) \in \mathbb{T}^d_x \times \mathbb{R}^d_v\).
The solution, subject to the initial condition \( f(t = 0) = f_0 \), is written as \( f(t) = e^{Lt} f_0 \), \( L = C - T \). As \( t \to \infty \), it is expected to converge to a constant multiple of \( F(x, v) = \exp(-|v|^2/2 - \Phi(x)) \in L^1(\mathbb{R}^d) \), where the constant is determined from the initial condition by mass conservation:

\[
\int e^{Lt} f_0 \, dv \, dx = \int f_0 \, dv \, dx, \quad t \geq 0.
\]

In the case (Torus), the equilibrium is position-independent: \( F(x, v) = \exp(-|v|^2/2) \).

It turns out that a convenient functional analytic setting for (18) is \( L^2(d\sigma) \) with \( d\sigma = dv \, dx/F \). In \( L^2(d\sigma) \), the transport operator \( T \) is skew-symmetric and the collision operator \( C \) is symmetric and negative semi-definite, providing the entropy decay relation

\[
\frac{1}{2} \frac{d}{dt} \|f\|^2 = \langle C f, f \rangle \leq 0.
\]

Contrary to the previous section, the entropy dissipation rate

\[
-\langle C f, f \rangle = \int F \left| \nabla_v \left( \frac{f}{F} \right) \right|^2 \, dx \, dv
\]

is not coercive and vanishes already for local equilibria \( f = h(x)e^{-|v|^2/2} \in \mathcal{N}(C) \). Despite of this fact, decay to global equilibrium (with \( h(x) = \text{const} \)) is expected. This property is then called hypocoercivity [25]. The kinetic Fokker-Planck equation is the prototypical example and has witnessed several approaches for proving hypocoercivity. The powerful approach of Desvillettes and Villani [9] has been generalized to prove strong decay to equilibrium for the Boltzmann equation [10], but it only provides algebraic decay for smooth solutions. First results on exponential decay have been based on spectral estimates [15]. Villani [25] has initiated the search for generalized entropies with coercive dissipation rates (see also the earlier work [20]). His method, motivated by hypo-elliptic theory, is based on an \( H^1 \) setting. Comparable results have been derived recently by Baudoin [5] employing the \( \Gamma_2 \) calculus, an extension of the \( \Gamma \) calculus discussed above. Smoothness requirements for the initial conditions have been removed in [11] by a new strategy for finding modified entropies.

8. **Baudoin’s approach**

To keep computations simple, we replace the potential \( \Phi(x) \), confining in position space, by restriction of the dynamics to a torus. The invariant measure then reduces to \( d\mu = e^{-|v|^2/2} dv \, dx \) on \( \mathbb{T}_x \times \mathbb{R}_v \). We also introduce the new unknown

\[
h(t, x, v) := e^{\frac{1}{2}|v|^2/2} f(t, x, v),
\]

leading to the evolution equation

\[
\frac{\partial}{\partial t} h = L h, \quad \text{with } L = \Delta_v - v \cdot \nabla_v - v \cdot \nabla_x.
\]

Mass conservation now reads

\[
\int e^{Lt} h_0 \, d\mu = \int h_0 \, d\mu,
\]

where the right hand side is again assumed to vanish in the following. We shall need the carré du champ

\[
\Gamma(f, g) = \frac{1}{2} \left( (Lf, g) - f(Lg - gL) \right), \quad \Gamma(f) := \Gamma(f, f) = |\nabla_v f|^2.
\]
and its iteration
\[
\Gamma_2(f) = \frac{1}{2} \Gamma(f, f) - \Gamma(f, Lf) = |\nabla^2_v f|^2 + |\nabla_v f|^2 + \nabla_v f \cdot \nabla_x f,
\]
where the computation of the latter is somewhat lengthy, but straightforward. Because of the hypocoercivity, an auxiliary carré du champ and its iteration are needed:
\[
\Gamma^Z(f, g) := Zf \cdot Zg, \quad Z = 2\nabla_x + \nabla_v, \quad \Gamma^Z(f) := \Gamma^Z(f, f),
\]
and
\[
\Gamma^Z_2(f) = \frac{1}{2} \Gamma^Z(f, f) - \Gamma^Z(f, Lf) = |Z \otimes \nabla_v f|^2 + Z f \cdot \nabla_v f + Z f \cdot \nabla_x f.
\]
Another preparation step is the simple proof of
\[
(19) \quad \Gamma_2(f) + \Gamma^Z_2(f) \geq 2|\nabla_x f + \nabla_v f|^2 = \frac{1}{4} \Gamma^Z(f) - \frac{1}{2} \Gamma(f) + \left| \nabla_x f + \frac{3}{2} \nabla_v f \right|^2.
\]
Finally, the Poincaré inequality
\[
\int (\Gamma(f) + \Gamma^Z(f)) d\mu \geq \kappa \int f^2 d\mu
\]
will be needed, which is a consequence of the standard Poincaré inequality on \(\mathbb{T}_x\) and of the weighted Poincaré inequality on \(\mathbb{R}_v\) with weight \(e^{-|v|^2/2}\).

Now we define
\[
\Psi(s) = e^{Ls} \left( \frac{3}{4} h(t-s)^2 + \Gamma(h(t-s)) + \Gamma^Z(h(t-s)) \right)
\]
and compute
\[
\Psi'(s) = e^{Ls} \left( \frac{3}{2} \Gamma(h) + 2\Gamma_2(h) + 2\Gamma^Z(h) \right) (t-s) \geq \frac{1}{2} e^{Ls} (\Gamma(h) + \Gamma^Z(h))(t-s),
\]
using (19) and the maximum principle. Now we integrate and use mass conservation and the Poincaré inequality:
\[
\frac{d}{ds} \int \Psi(s) d\mu \geq \int \left[ \frac{1-\varepsilon}{2} (\Gamma(h) + \Gamma^Z(h)) + \frac{\varepsilon \kappa}{2} f^2 \right] (t-s) d\mu = \lambda \int \Psi(s) d\mu,
\]
with \(\lambda = \frac{2\varepsilon}{3+4\varepsilon}\) (for optimal \(\varepsilon\)). The consequence
\[
\int \Psi(t) d\mu \geq e^{\lambda t} \int \Psi(0) d\mu
\]
is equivalent to
\[
\int \left( \frac{3}{4} h(t)^2 + \Gamma(h(t)) + \Gamma^Z(h(t)) \right) d\mu \leq e^{-\lambda t} \int \left( \frac{3}{4} h_0^2 + \Gamma(h_0) + \Gamma^Z(h_0) \right) d\mu,
\]
which in turn is equivalent to exponential decay of the \(H^1(d\mu)\)-norm of \(h\) (by \(\Gamma(h) + \Gamma^Z(h) \geq (3 - \sqrt{2})(|\nabla_x h|^2 + |\nabla_v h|^2)\)) or the \(H^1(d\sigma)\)-norm of \(f\) from the preceding section, assuming \(H^1\)-regularity of the initial data. Roughly the same result can be obtained by the methods of Villani [25].
9. A Modified Entropy

Here we shall apply the methodology of [11] to

\[ \partial_t f + \mathbf{T} f = \mathbf{C} f, \quad \mathbf{T} f = v \cdot \nabla_x f, \quad \mathbf{C} f = \nabla_v \cdot (\nabla_v f + v f), \]

\[ x \in \mathbb{T}^d, \quad v \in \mathbb{R}^d, \quad t > 0, \] subject to the initial condition \( f(t = 0) = f_0 \) with vanishing total mass: \( \int f_0 dv dx = 0 \). As mentioned above, the transport operator \( \mathbf{T} \) is skew symmetric and the collision operator \( \mathbf{C} \) symmetric negative semidefinite with respect to \( L^2(d\sigma), \quad d\sigma = e^{\|v\|^2/2} dv dx \).

We interpret the equation as an abstract ODE on a Hilbert space \( \mathcal{H} \), and will pose a list of assumptions on the operators \( \mathbf{T} \) and \( \mathbf{C} \), which will be checked later for the Fokker-Planck equation.

The simplest example. is a second order ODE system with

\[ \mathbf{T} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}. \]

Although the entropy dissipation \( -\frac{d}{dt}|f|^2 = 2f_1^2 \) is only semidefinite, the eigenvalues of \( \mathbf{C} - \mathbf{T} \) are given by \( \lambda = (-1 \pm i\sqrt{3})/2 \) and solutions decay exponentially, an example of hypocoercivity. The secret lies in the fact that solutions do in general not remain in \( \mathcal{N}(\mathbf{C}) = \{(f_1,0)\} \) by the rotational motion caused by \( \mathbf{T} \), except when the global equilibrium \( f = 0 \) has been reached (instability of hydrodynamic states in the language of Desvillettes and Villani [10]). For this simple problem, the method of [11] amounts to the introduction of a modified entropy of the form

\[ H[f] := \frac{1}{2} |f|^2 + \epsilon f_1 f_2, \]

with small positive \( \epsilon \), and with the dissipation

\[ -\frac{d}{dt} H[f] = \epsilon f_1^2 + (1 - \epsilon)f_2^2 - \epsilon f_1 f_2. \]

For small enough \( \epsilon \), both \( H \) and its dissipation are obviously coercive, implying exponential decay.

The general approach. With the orthogonal projection \( \Pi : \mathcal{H} \rightarrow \mathcal{N}(\mathbf{C}) \), the main ingredients of the general approach are the assumptions of

**H1.** Diffusive macroscopic limit: \( \Pi \mathbf{T} \Pi = 0 \).

**H2.** Microscopic coercivity: \( -\langle \mathbf{C} f, f \rangle \geq \lambda_m (1 - \Pi) \|f\|^2, \quad \lambda_m > 0 \), and

**H3.** Macroscopic coercivity (or again ‘instability of hydrodynamic states’):

\[ \|\mathbf{T} \Pi f\|^2 \geq \lambda_M \|\Pi f\|^2, \quad \lambda_M > 0. \]

The modified entropy is then given by

\[ H[f] = \frac{1}{2} \|f\|^2 + \epsilon \langle \mathbf{A} f, f \rangle, \quad \mathbf{A} = [1 + (\mathbf{T} \Pi)^* \mathbf{T} \Pi]^{-1}(\mathbf{T} \Pi)^*, \]

and it satisfies

\[ \frac{d}{dt} H[f] = \langle \mathbf{C} f, f \rangle - \epsilon \langle \mathbf{A} \mathbf{T} \Pi f, f \rangle - \epsilon \langle \mathbf{A} (1 - \Pi) f, f \rangle + \epsilon \langle \mathbf{TA} f, f \rangle + \epsilon \langle \mathbf{AC} f, f \rangle. \]

It has been proved in [11] that the operators \( \mathbf{A} \) and \( \mathbf{TA} \) are bounded. Under the additional assumption of

**H4.** Boundedness of the auxiliary operators \( \mathbf{AT} \) and \( \mathbf{AC}, \)
the estimate
\[
\frac{d}{dt} H[f] \leq -\lambda_m ((1 - \Pi)f)^2 - \frac{\varepsilon \lambda_M}{1 + \lambda_M} \|\Pi f\|^2 + \varepsilon C ((1 - \Pi)f \cdot f)
\]
holds, implying (as for the model problem) that both \(H\) and its dissipation are coercive for \(\varepsilon\) small enough.

**The kinetic Fokker-Planck equation.** With the operators \(T\) and \(C\) from (20) and \(H = L_2(d\sigma)\) the orthogonal projection to \(\mathcal{N}(C)\) is given by
\[(\Pi f)(x, v) = \rho_f(x) F(v), \quad F(v) = (2\pi)^{-d/2} e^{-|v|^2/2}, \quad \rho_f(x) = \int_{\mathbb{R}^d} f(x, v) dv.
\]
The marginal density \(\rho_f\) is called the macroscopic or position density. Assumption \(H_1\) is a consequence of the fact that the flux associated to local equilibria of the form \(\rho(x) F(v)\) vanishes: \(\int v F dv = 0\). It is called ’diffusive macroscopic limit’ since it is responsible for the fact that the diffusive rescaling \(t \rightarrow t/\varepsilon^2, x \rightarrow x/\varepsilon\) in (20) and the subsequent macroscopic limit \(\varepsilon \rightarrow 0\) lead to a diffusion equation for the limit of \(\rho_f\).

Assumption \(H_2\) is equivalent to the weighted Poincaré inequality
\[
\int_{\mathbb{R}^d} |\nabla_x h|^2 F dv \geq \lambda_m \int_{\mathbb{R}^d} \left(h - \int h F dv\right)^2 F dv,
\]
which holds for the Gaussian weight \(F\). Similarly the macroscopic coercivity assumption \(H_3\) is equivalent to the Poincaré inequality on the torus,
\[
\int_{\mathbb{T}^d} |\nabla_x \rho|^2 dx \geq \lambda_M \int_{\mathbb{T}^d} \rho^2 dx, \quad \text{for} \int_{\mathbb{T}^d} \rho dx = 0.
\]
The boundedness of \(AT\) is equivalent to an elliptic regularity result, the boundedness of \(\Delta^{-1}_x\) as a map from \(L_2^0(\mathbb{T}^d)\) to \(H_0^2(\mathbb{T}^d)\), where the subscript 0 means the codimension 1 subspaces of functions with zero average. Finally, the boundedness of \(AC\) is a consequence of the preliminary computation \(\rho_{TC} = -\nabla_x \cdot \int v f dv\), showing that \(AC\) does not contain \(v\)-derivatives. Thus, \(H_1-H_4\) holds with the consequence that solutions of (20) with zero total mass converge to zero exponentially in terms of the norm in \(L^2(dx dv/F)\), if the initial datum \(f_0\) is in this space. In particular, no smoothness of \(f_0\) is required as in the results in [5, 25].

The methods of [5, 11, 25] have been carried out for the standard kinetic Fokker-Planck problem on whole space with confining potential under similar assumptions on the potential, except that Baudoin [5] needs the additional assumption of boundedness of its second derivatives.

Finally it should be noted that the method of [11] presented in this section has also been applied to velocity jump processes, where \(C\) is an integral operator and the semigroup generated by \(L = C - T\) does not have any smoothing properties.

**References**


ANTON ARNOLD: INSTITUT FÜR ANALYSE UND SCIENTIFIC COMPUTING, TECHNISCHE UNIVERSITÄT WIEN, WIEDNER HAUPTSTRASSE 8–10, 1040 WIEN, AUSTRIA

E-mail address: Anton.Arnold@tuwien.ac.at

C. SCHMEISER: FAKULTÄT FÜR MATHEMATIK, UNIVERSITÄT WIEN, OSKAR-MORGENSTERN-PLATZ 1, 1090 WIEN, AUSTRIA

E-mail address: Christian.Schmeiser@univie.ac.at