ENTROPY METHODS

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. Motivation

We consider a system of linear ODEs with constant coefficients:

$$\frac{dp}{dt} = Ap\,, \tag{1}$$

with unknown $p(t) = (p_1(t), \dots, p_N(t)) \in \mathbb{R}^N$, $t \geq 0$, and given constant $A \in \mathbb{R}^{N\times N}$. We assume that the initial datum $p(0)$ is a probability distribution on $\{1, \dots, N\}$, i.e. $p_n(0) \geq 0$, $1 \leq n \leq N$, and $\sum_{n=1}^N p_n(0) = 1$, and we want this property to be conserved by the dynamics given by (1). This leads to the requirements

$$A_{mn} > 0\,, \quad 1 \leq m \neq n \leq N\,, \qquad \sum_{m=1}^N A_{mn} = 0\,, \quad 1 \leq n \leq N\,, \tag{2}$$

where the strict inequality would not be necessary, but is assumed for convenience. It is easily seen that it implies $p_n(t) > 0$, $1 \leq n \leq N$, $t > 0$, abbreviated as $p(t) > 0$ in the following.

Lemma 1. *Let (2) hold and let $p(t), q(t) \in \mathbb{R}^N$ be solutions of (1) with $q(t) > 0$, $t \geq 0$. Then*

$$\frac{dH(p|q)}{dt} = -\sum_{m,n=1}^N A_{nm} q_m \left(\frac{p_n}{q_n} - \frac{p_m}{q_m}\right)^2 \leq 0\,, \qquad \text{with } H(p|q) := \sum_{n=1}^N \frac{(p_n - q_n)^2}{q_n}\,.$$

Proof. First we note that (1) can be written in the form

$$\frac{dp_n}{dt} = \sum_{m\neq n} (A_{nm} p_m - A_{mn} p_n)\,, \qquad 1 \leq n \leq N\,.$$

Date: October 7, 2018.

Using this, a straightforward computation gives

$$\frac{dH(p|q)}{dt} = \sum_{n=1}^{N} \sum_{m\neq n} \left[A_{nm} \left(\frac{2p_n p_m}{q_n} - \frac{p_n^2 q_m}{q_n^2} \right) - A_{mn} \frac{p_n^2}{q_n} \right] .$$

The proof is completed by $m \leftrightarrow n$ in the last term in the bracket. □

This somehow miraculous result can be used for a complete analysis of the long time behavior of solutions of (1).

Lemma 2. *The system (1) has a unique steady state p_∞ such that $\sum_{n=1}^{N} p_{\infty,n} = 1$. It satisfies $p_\infty > 0$.*

Proof. First, it is easily seen that the real parts of the eigenvalues of A are non-positive, because otherwise initial conditions for p and q could be chosen such that $H(p|q)$ would increase exponentially, which is impossible by the previous result. On the other hand, zero is an eigenvalue of A as a consequence of (2). Therefore, for large enough $\mu > 0$, the matrix $A + \mu I$ has only positive entries (by (2)), and it has the eigenvalue μ, which is also the eigenvalue with the largest real part.

At this point we need a result from linear algebra, the Perron-Frobenius theorem (see, e.g., [25]). It says that for positive matrices the spectral radius is a simple eigenvalue with a positive eigenvector (where positive is always meant component-wise). Obviously, this eigenvalue has to be μ for the matrix $A + \mu I$, implying that zero is a simple eigenvalue of A with a positive eigenvector, which can be scaled to become a probability distribution p_∞. □

Lemma 3. *Let $p(t)$, $t \geq 0$, be a solution of (1), where $p(0)$ is a probability distribution. Then, as $t \to \infty$, $p(t)$ converges to p_∞ exponentially.*

Proof. With $\lambda := \min_{m\neq n} A_{nm}/p_{\infty,n} > 0$, we have

$$\frac{dH(p|p_\infty)}{dt} \leq -\lambda \sum_{m,n=1}^{N} \Bigg(p_{\infty,m} \frac{(p_n - p_{\infty,n})^2}{p_{\infty,n}} + p_{\infty,n} \frac{(p_m - p_{\infty,m})^2}{p_{\infty,m}} - 2(p_n - p_{\infty,n})(p_m - p_{\infty,m}) \Bigg) .$$

The term in the second line disappears, since both $p(t)$ and p_∞ are probability distributions, implying

$$\frac{dH(p|p_\infty)}{dt} \leq -2\lambda H(p|p_\infty) .$$

By the Gronwall inequality (`https://en.wikipedia.org/wiki/Grönwall's_inequality`) we get exponential decay of (the weighted ℓ^2-Norm of $p - p_\infty$) $H(p|p_\infty)$ to zero. □

In the following we shall explain general principles leading to results like Lemma 1 in much more general situations, motivated by ideas from stochastic processes. If the existence of an appropriate equilibrium can be verified, as in Lemma 2, then in many cases the result of Lemma 1 can be strengthened to provide convergence to equilibrium, as in Lemma 3.

2. 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.*

$$X_t = (X_{1,t}, \dots, X_{d,t}) : \Omega \to \mathbb{R}^d \qquad \text{and}$$

$$\{\omega \in \Omega : X_{i,t}(\omega) \le \alpha\} \in \mathcal{F}\,, \qquad i = 1, \dots, d\,, \quad t \ge 0\,, \quad \alpha \in \mathbb{R}\,.$$

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

The *Borel σ-algebra* $\mathcal{B}_d$ is the smallest σ-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

$$P_X(A) = P(\{\omega \in \Omega : X(\omega) \in A\})\,, \qquad A \in \mathcal{B}_d\,,$$

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

$$P_{X_t}(A) = \int_A p(x,t)dx\,.$$

Remark 1. *Here we mostly concentrate on describing CTSP in terms of their laws at different times, i.e. vary ω for each fixed t. Alternatively, the pathwise view, i.e. vary t for each fixed ω, sees a CTSP as a function valued random variable. This leads to probability distributions on function spaces, which is much more sophisticated from the analysis point of view. This is necessary when properties of sample paths are of interest, a subject we only touch peripherically in the following.*

Expectation values of functions of random variables are given by

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

Obviously, probability densities p have to satisfy $p \ge 0$ and $\langle p\rangle = 1$. An example is the *Gaussian*

$$p(x) = (2\pi\sigma^2)^{-d/2} \exp\left(-\frac{|x-\mu|^2}{2\sigma^2}\right)$$

with *mean* $\mu \in \mathbb{R}^d$ 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,

$$p_X(x) = \int_{\mathbb{R}^d} p_{(X,Y)}(x,y)dy\,, \qquad p_Y(y) = \int_{\mathbb{R}^d} p_{(X,Y)}(x,y)dx$$

hold.

The conditional expectation. We recall the definition $P(A|B) = P(A\cap B)/P(B)$ of the *conditional probability* of A under the condition B. For fixed B with $P(B) > 0$ this defines a new probability distribution, which allows to compute expectation values, called conditional expectation. What we shall need, is the conditional expectation $E(f(X)|Y = y)$ for two random vectors X and Y. Our formal derivation will be based on a discretization $\mathbb{R}^d = \bigcup_{j\in\mathbb{N}} C_j$ (into disjoint cubes C_j) with $x_j \in C_j$ and $\Delta C := |C_j| \ll 1$, $j \in \mathbb{N}$, which implies

$$P(X \in C_j|Y \in C_k) = \frac{P_{(X,Y)}(C_j \times C_k)}{P_Y(C_k)} \approx \frac{p_{(X,Y)}(x_j, x_k)(\Delta C)^2}{p_Y(x_k)\Delta C}\,,$$

and therefore

$$E(f(X)|Y \in C_k) \approx \sum_{j\in\mathbb{N}} f(x_j)\frac{p_{(X,Y)}(x_j, x_k)}{p_Y(x_k)}\Delta C\,.$$

This allows to formally pass to the limit $\Delta C \to 0$, $x_k \to y$, leading to the desired formula

$$E(f(X)|Y = y) = \int_{\mathbb{R}^d} f(x)\frac{p_{(X,Y)}(x, y)}{p_Y(y)}dx\,.$$

Strictly speaking, this formula can of course only be used when $p_Y > 0$. A general definition of conditional expectation, not necessarily based on probability densities, is rather involved [8].

The random vectors X and Y are called *independent*, if

$$E(f(X)|Y = y) = E(f(X)) = \int_{\mathbb{R}^d} f(x)p_X(x)dx \qquad \forall\, y \in \mathbb{R}^d\,.$$

This holds of course, if $p_{(X,Y)}$ is factorized: $p_{(X,Y)}(x, y) = p_X(x)p_Y(y)$.

Properties of CTSP can often be described in terms of there *finite-dimensional distributions*, i.e. $(X_{t_1}, \ldots, X_{t_n})$ for arbitrary $n \in \mathbb{N}$ and $0 \le t_1 < \ldots < t_n$. For example: A CTSP has *independent increments*, if the random vectors X_{t_1}, $X_{t_2} - X_{t_1}$, $\ldots$, $X_{t_n} - X_{t_{n-1}}$ are mutually independent for all finite-dimensional distributions.

An important class of CTSPs are *martingales*. They satisfy

$$E(|X_t|) < \infty\,, \qquad E(X_{t+h}|X_t = x) = x \qquad \forall\, t \ge 0\,,\, h > 0\,,\, x \in \mathbb{R}^d\,.$$

If a CTSP has independent increments and constant mean, i.e. $E(X_t) = E(X_0)$, $t \ge 0$, it is a martingale. A complete definition of martingales and several of their properties can be found in [8].

A *Markov process* is a CTSP without memory: For every $t \ge 0$, knowledge of the distribution at time t is sufficient for predicting the process at later times. In terms of an arbitrary finite-dimensional distribution this can be formulated as

$$E(f(X_{t_n})|(X_{t_1}, \ldots, X_{t_{n-1}}) = (x_1, \ldots, x_{n-1})) = E(f(X_{t_n})|X_{t_{n-1}} = x_{n-1})\,.$$

In terms of probability densities for three times $s < u < t$, this translates to

$$\frac{p_{(X_t,X_u,X_s)}(x, z, y)}{p_{(X_u,X_s)}(z, y)} = \frac{p_{(X_t,X_u)}(x, z)}{p_{X_u}(z)} =: \pi_{t,u}(x, z)\,,$$

where the right hand side is called *transition probability density,* which satisfies the *Chapman-Kolmogorow equation*

$$\pi_{t,s}(x,y) = \int_{\mathbb{R}^d} \frac{p_{(X_t,X_u,X_s)}(x,z,y)}{p_{(X_u,X_s)}(z,y)} \frac{p_{(X_u,X_s)}(z,y)}{p_{X_s}(y)} dz = \int_{\mathbb{R}^d} \pi_{t,u}(x,z)\pi_{u,s}(z,y)dz\,.$$

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-Kolmogorow equation:

$$T(t+s) = T(t)T(s)\,, \qquad t,s \geq 0\,.$$

Its generator will be denoted by

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

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

In the following we shall sometimes write $\langle f, g\rangle$ instead of $\langle fg\rangle$, to emphasize the role of $\langle\cdot,\cdot\rangle$ as a duality pairing. The dual L of L^* ($\langle \mathsf{L}p, q\rangle = \langle p, \mathsf{L}^* q\rangle$) 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 $\{e^{\mathsf{L}^* t} : t \geq 0\}$ and the semigroup $\{e^{\mathsf{L}t} : t \geq 0\}$, generated by L. $e^{\mathsf{L}^* t}$ acts on the space of bounded measures on $\mathbb{R}^d$ (typically probability measures), and it preserves mass, such that, when p_0 is a probability measure on M, then also $p(t) = e^{\mathsf{L}^* t} p_0$ is a probability measure for all $t \geq 0$, i.e.

$$\langle \mathsf{L}^* f\rangle = 0\,, \qquad \langle e^{\mathsf{L}^* t} f\rangle = \langle f\rangle\,, \qquad \mathsf{L}1 = 0\,.$$

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

$$\partial_t p = \mathsf{L}^* p\,,$$

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 $\pi_t(x,y) = (e^{\mathsf{L}^* t}\delta_y)(x)$, where δ_y denotes the Delta distribution concentrated at y. Expectation values can be computed via the *Feynman-Kac formula*

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

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 Kolmogorow equation*

$$\partial_s u + \mathsf{L}u = 0\,, \qquad 0 \leq s \leq t\,, \tag{3}$$

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.

$$\frac{d}{dt} E(U_t) = \frac{d}{dt}\langle up\rangle = 0\,. \tag{4}$$

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

$$\begin{aligned} E(f(X_t)g(X_s)) &= \int_M \int_M f(x)g(y)(e^{L^*(t-s)}\delta_y)(x)p(s,y)dx\,dy \\ &= \langle f, e^{\mathsf{L}^*(t-s)}(gp(s))\rangle = \langle e^{\mathsf{L}(t-s)}f, gp(s)\rangle\,, \qquad t \geq s \geq 0\,. \end{aligned} \tag{5}$$

This can be used in a straightforward computation, producing an explanation for calling L the generator of the Markov process:

$$0 = E\left(\left[f(X_t) - f(X_s) - \int_s^t \mathsf{L}f(X_r)dr\right] g(X_s)\right)\,, \qquad t \geq s \geq 0\,. \tag{6}$$

3. 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

$$S_k := \sum_{j=1}^k \xi_j\,, \qquad k \geq 1\,,$$

implying $E(S_k) = 0$, $E(S_k^2) = k$. Now define a CTSP B_t^n by $B_0^n = 0$,

$$B_{k/n}^n := n^{-1/2}S_k\,, \qquad k \in \mathbb{N}\,,$$

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

$$\frac{B_{k/n}^n - B_{l/n}^n}{\sqrt{k/n - l/n}} = \frac{S_k - S_l}{\sqrt{k-l}}\,,$$

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

$$B_{k/n}^n - B_{l/n}^n = \frac{S_k - S_l}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{j=l+1}^k \xi_j\,,$$

which, by the fact that all ξ_j have the same law, implies that $B_{k/n}^n - B_{l/n}^n$ has the same law as $B_{(k-l)/n}^n$. 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

$$E((B_{t+\Delta t} - B_t)^2) = E(B_{\Delta t}^2) = \Delta t\,. \tag{7}$$

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:

$$\mathsf{L} = \mathsf{L}^* = \frac{1}{2}\partial_x^2 .$$

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 $\{k/\sqrt{n} : k \in \mathbb{Z}\}$ 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

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

and it is generated by the Laplace operator: $\mathsf{L} = \mathsf{L}^* = \Delta_x/2$.

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

$$dX_t = b(X_t)dt + \sigma dB_t \tag{8}$$

with a vector field b and a $d \times d$-matrix σ are called *diffusion processes*. We shall assume σ 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* σ is allowed to depend on the state, i.e., $\sigma = \sigma(X_t)$, which makes the interpretation of (8) significantly more complicated. The term σB_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 σ in order to stay away from these difficulties.

In order to find the generator of a diffusion process we use (3), (4). So we assume that $U_t = u(X_t, t)$ is a martingale and compute for small $\Delta t > 0$

$$\begin{aligned} U_{t+\Delta t} - U_t &= u(X_{t+\Delta t}, t+\Delta t) - u(X_t, t) \approx \partial_t u(X_t, t)\Delta t \\ &+ \nabla_x u(X_t, t) \cdot (X_{t+\Delta t} - X_t) + \frac{1}{2}(X_{t+\Delta t} - X_t)^{tr} \nabla_x^2 u(X_t, t)(X_{t+\Delta t} - X_t) . \end{aligned} \tag{9}$$

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 (7), considering $B_{t+\Delta t} - B_t$ as an $O(\sqrt{\Delta t})$-term, the approximation in (9) means that terms up to $O(\Delta t)$ have been kept. The approximation can be simplified further:

$$\begin{aligned} U_{t+\Delta t} - U_t &= (\partial_t u(X_t,t) + b(X_t)\cdot\nabla_x u(X_t,t))\,\Delta t + \nabla_x u(X_t,t)\cdot(\sigma(B_{t+\Delta t}-B_t)) \\ &\quad + \frac{1}{2}(B_{t+\Delta t}-B_t)^{tr}\sigma^{tr}\nabla_x^2 u(X_t,t)\sigma(B_{t+\Delta t}-B_t) + O(\Delta t^{3/2})\,. \end{aligned} \tag{10}$$

Now we compute the expectation (using the martingale property of U_t, as well as (7) in the second line), divide by Δt and pass to the limit $\Delta t \to 0$:

$$0 = \partial_t u + \mathsf{L}u\,, \qquad \text{with } \mathsf{L}u = b\cdot\nabla_x u + \nabla_x\cdot(D\nabla_x u)\,, \quad D = \frac{1}{2}\sigma^{tr}\sigma\,.$$

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 (10), giving the result of the *Ito Lemma*

$$du_t = (\partial_t u + \mathsf{L}u)dt + \nabla_x u\cdot(\sigma dB_t)\,,$$

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

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

4. 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

$$N_t^n := \sum_{j=1}^{k}\xi_j\,, \qquad \frac{k}{n} \le t < \frac{k+1}{n}\,,$$

implying that for every $t \ge 0$, N_t^n is a nonnegative integer and, by elementary combinatorics,

$$P\,(N_t^n = l) = \binom{k}{l}\left(\frac{\lambda}{n}\right)^l\left(1-\frac{\lambda}{n}\right)^{k-l}\,, \qquad \frac{k}{n} \le t < \frac{k+1}{n}\,. \tag{11}$$

We also get

$$E(N_t^n) = \frac{\lambda k}{n}\,, \quad E((N_t^n - \lambda k/n)^2) = \frac{\lambda k}{n}\left(1-\frac{\lambda}{n}\right)\,, \qquad \frac{k}{n} \le t < \frac{k+1}{n}\,.$$

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 λ*, satisfying

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

which can be obtained from (11) 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(λ) 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

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

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

$$(\mathsf{L}u)(x) = \lambda(u(x+1) - u(x)) \,.$$

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

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

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

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

For jump processes the forward Kolmogorow equation $\partial_t p = \mathsf{L}^* p$ is often called the *master equation.* The generator of the jump process is given by

$$(\mathsf{L}u)(x) = \int_{\mathbb{R}^d} W(x \to x')(u(x') - u(x))dx' \,.$$

5. Time reversal – decay of the relative entropy

We are looking for a time reversed version of (6) and compute

$$\begin{aligned} E([f(X_t) - f(X_s)]g(X_t)) &= \langle g, f\,p(t)\rangle - \left\langle e^{\mathsf{L}(t-s)} g, f\,p(s)\right\rangle \\ &= \int_s^t \left(\left\langle -\mathsf{L}e^{\mathsf{L}(t-r)} g, f\,p(r)\right\rangle + \left\langle e^{\mathsf{L}(t-r)} g, f\,\mathsf{L}^* p(r)\right\rangle\right) dr \\ &= -\int_s^t \left\langle e^{\mathsf{L}(t-r)} g, p(r)\overline{\mathsf{L}}^{p(r)} f\right\rangle dr \,, \end{aligned}$$

with

$$\overline{\mathsf{L}}^{p(t)} f = \frac{1}{p(t)}(\mathsf{L}^*(fp(t)) - f\mathsf{L}^* p(t)) \,, \tag{12}$$

implying

$$0 = E\left(\left[f(X_t) - f(X_s) + \int_s^t \overline{\mathsf{L}}^{p(r)} f(X_r)dr\right] g(X_t)\right) \,, \qquad t \geq s \geq 0 \,. \tag{13}$$

The operator $\overline{\mathsf{L}}^{p(t)}$ 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 $\mathsf{L}^* p_\infty = 0$, and if the *detailed balance* condition

$$p_\infty \mathsf{L} = \mathsf{L}^*(p_\infty \cdot)$$

holds, then X_t is *time reversible,* i.e. $\overline{\mathsf{L}}^{p_\infty} = \mathsf{L}$. The computation

$$\langle \mathsf{L}^*(p_\infty f), g\rangle = \langle \mathsf{L} f, p_\infty g\rangle = \langle f, \mathsf{L}^*(p_\infty g)\rangle$$

shows that detailed balance is equivalent to the symmetry of the operator $\mathsf{L}^*(p_\infty \cdot)$.

Let $p(t) = e^{\mathsf{L}^* t} p_0$ and $q(t) = e^{\mathsf{L}^* t} q_0$ denote the probability densities of X_t^p and, respectively, X_t^q. Then a straightforward computation shows that p/q solves the backward Kolmogorow equation of the time reversal of X_t^q, i.e.,

$$\partial_t \left(\frac{p}{q}\right) - \overline{\mathsf{L}}^q \left(\frac{p}{q}\right) = 0\,,$$

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

$$\frac{p(t,x)}{q(t,x)} = E\left(\frac{p(s,X_s^q)}{q(s,X_s^q)} \;\Big|\; X_t^q = x\right).$$

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

$$\begin{aligned}\frac{p(t,x)}{q(t,x)} &= \frac{1}{q(t,x)}\left(e^{\mathsf{L}^*(t-s)} p(s)\right)(x) = \frac{1}{q(t,x)}\left\langle e^{\mathsf{L}^*(t-s)} p(s), \delta_x\right\rangle \\ &= \left\langle \frac{p(s)}{q(s)}, \frac{q(s)}{q(t,x)} e^{\mathsf{L}(t-s)}\delta_x\right\rangle.\end{aligned}$$

The second factor on the right hand side can be interpreted as the probability density for X_s^q under the condition $X_t^q = x$. For the following, it is only important that it is a probability density:

$$\frac{1}{q(t,x)}\left\langle q(s), e^{\mathsf{L}(t-s)}\delta_x\right\rangle = \frac{1}{q(t,x)}\left\langle e^{\mathsf{L}^*(t-s)} q(s), \delta_x\right\rangle = 1\,.$$

Thus, for a convex function U, the Jensen inequality implies

$$U\left(\frac{p(t,x)}{q(t,x)}\right) \leq \left\langle U\left(\frac{p(s)}{q(s)}\right), \frac{q(s)}{q(t,x)} e^{\mathsf{L}(t-s)}\delta_x\right\rangle,$$

leading to the pointwise estimate for the *local relative entropy*

$$h_U(p(t)|q(t))(x) := q(t,x) U\left(\frac{p(t,x)}{q(t,x)}\right) \leq \left[e^{\mathsf{L}^*(t-s)} h_U(p(s)|q(s))\right](x)\,. \tag{14}$$

Integration with respect to x gives the decay of the *entropy of p relative to q,*

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

by the conservation property of the forward Kolmogorow 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 [14] is, according to (12), given by

$$\overline{\mathsf{L}}^p f = \nabla_x \cdot (D\nabla_x f) + \nabla_x f \cdot \left(\frac{2D\nabla_x p}{p} - b\right).$$

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 2. *Some of the uneasyness 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_∞ is time reversible, iff $D\nabla_x p_\infty = bp_\infty$, 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 Φ is a *confining potential*, satisfying $e^{-\Phi} \in L^1(\mathbb{R}^d)$ and shifted such that $\langle e^{-\Phi}\rangle = 1$. The probability distribution of X_t then satisfies the *Fokker-Planck equation*

$$\partial_t p = \mathsf{L}^* p = \nabla_x \cdot (D(\nabla_x p + p\nabla_x \Phi))\,, \tag{15}$$

with the unique equilibrium $p_\infty = e^{-\Phi}$. The symmetrized form

$$\mathsf{L}^* p = \nabla_x \cdot (Dp_\infty \nabla_x (p/p_\infty))$$

of the Fokker-Planck operator immediately implies the detailed balance property

$$\mathsf{L}^*(p_\infty f) = \nabla_x \cdot (Dp_\infty \nabla_x f) = p_\infty \mathsf{L} f\,.$$

Example 2. *(Jump processes)* In

$$(\mathsf{L}^* p)(x) = \int_M [W(x' \to x)p' - W(x \to x')p]dx'\,,$$

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

$$(\mathsf{L} f)(x) = \int_M W(x \to x')(f' - f)dx'$$

and, respectively,

$$(\overline{\mathsf{L}}^p f)(x) = \int_M \frac{W(x' \to x)p'}{p}(f' - f)dx'$$

Time reversibility obviously requires an equilibrium satisfying

$$W(x' \to x)p'_\infty = W(x \to x')p_\infty\,, \qquad x, x' \in M\,.$$

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.*

6. 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 (14):

$$\partial_s \left(e^{\mathsf{L}^*(t-s)}\left[q(s)U\left(\frac{p(s)}{q(s)}\right)\right]\right) = -e^{\mathsf{L}^*(t-s)} i_U(p(s)|q(s))\,, \tag{16}$$

with the *local Fisher information*

$$(17) i_U(p|q) = \mathsf{L}^*\left(qU\left(\frac{p}{q}\right)\right) - U'\left(\frac{p}{q}\right)\mathsf{L}^*p - \left(U\left(\frac{p}{q}\right) - \frac{p}{q}U'\left(\frac{p}{q}\right)\right)\mathsf{L}^*q \geq 0\,.$$

Integration with respect to s leads to the local (relative) entropy dissipation relation

$$h_U(p(t)|q(t)) - e^{\mathsf{L}^*(t-s)}h_U(p(s)|q(s)) = -\int_s^t e^{\mathsf{L}^*(t-r)}i_U(p(r)|q(r))dr\,,$$

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 3. *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_∞ with $\mathsf{L}^*p_\infty = 0$ exists, and the *Poincaré inequality*

$$(18) \qquad I_U(p|p_\infty) \geq \lambda H_U(p|p_\infty)\,, \qquad \text{with } \lambda > 0$$

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

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

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

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_∞ 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 (18) is a consequence of the weighted Poincaré inequality

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

which holds, if

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

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

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

It is not completely straightforward to prove existence of an equilibrium p_∞. For the case of *local micro-reversibility,* i.e. $W(x' \to x)p'_\infty = W(x \to x')p_\infty$ 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_∞ 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 (18) and therefore exponential decay of the relative entropy but, by (16), actually the stronger local result

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

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 $p_0(x) \leq cp_\infty(x)$ is preserved by the evolution. Since then also $h_U(p_0|p_\infty)$, satisfies such a bound, the same holds for $e^{\mathsf{L}^* t}h_U(p_0|p_\infty)$. The local decay result then implies

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

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 Kolmogorow operator

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

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. [6]).

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

$$\begin{aligned} I_U(p|p_\infty) &= \frac{1}{4}\int_M\int_M (H(x'-x)p'_\infty + H(x-x')p_\infty)\left(\frac{p}{p_\infty} - \frac{p'}{p'_\infty}\right)^2 dx'dx \\ &\geq \frac{1}{2}\sqrt{\frac{ve}{\kappa}}\int_M\int_M p'_\infty p_\infty \left(\frac{p}{p_\infty} - \frac{p'}{p'_\infty}\right)^2 dx'dx = 2\sqrt{\frac{ve}{\kappa}} H_U(p|p_\infty) , \end{aligned}$$

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

Example 6. *Fractional diffusion* is generated by

$$\mathsf{L}^* p = -P.V. \int_{\mathbb{R}^d} \frac{p - p'}{|x-x'|^{d+\alpha}} dx' ,$$

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

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

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 $A = (a_{mn})_{m,n=1,\dots,d}$ with positive entries $a_{mn} > 0$. Then the Perron-Frobenius theorem states that the spectral radius λ of A is a simple eigenvalue with an eigenvector with positive components. Therefore there are unique normalized left and right eigenvectors ϕ and, repectively, R, such that,

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

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

$$\frac{du}{dt} = Au , \qquad u_n(0) \geq 0 , \quad \sum_{n=1}^d u_n(0) > 0 .$$

The new unknown $p(t)$ is defined by

$$p_n(t) = \frac{e^{-\lambda t}\phi_n u_n(t)}{\sum_{m=1}^d \phi_m u_m(0)}\,, \tag{19}$$

satisfying, with the diagonal matrix $\Phi := \mathrm{diag}(\phi_1, \ldots, \phi_d)$,

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

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

$$h_U(p|p_\infty)_n = \phi_n R_n \widehat{U}\left(\frac{r_n}{R_n}\right)\,,$$

where $r(t) = e^{-\lambda t}u(t)$ and $\widehat{U}$ 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 (17) is given by

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

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

$$\partial_t u + \partial_a u = -\nu(a)u\,, \qquad u(0,t) = \int_0^\infty B(a)u(a,t)da\,,$$

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

$$\begin{aligned} &\lambda R + \partial_a R = -\nu R\,, \qquad R(0) = \int_0^\infty BR\,da\,, \qquad \int_0^\infty R\,da = 1\,, \\ &\lambda\phi - \partial_a \phi = -\nu\phi + \phi(0)B\,, \qquad \int_0^\infty R\phi\,da = 1\,, \end{aligned}$$

and $R, \phi > 0$ has been proven in [22].

Analogously to (19), 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

$$\partial_t p + \phi\,\partial_a\left(\frac{p}{\phi}\right) = -(\lambda+\nu)p\,, \qquad p(a=0) = \phi(0)\int_0^\infty \frac{Bp}{\phi}da\,.$$

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 (17) 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):

$$\frac{d}{dt}\int_0^\infty \frac{p^2}{2p_\infty}da = \frac{\phi(0)}{2R(0)}\left(\left(\int_0^\infty BR\frac{p}{p_\infty}da\right)^2 - \int_0^\infty BR\,da\int_0^\infty BR\left(\frac{p}{p_\infty}\right)^2 da\right),$$

where the right hand side is nonpositive by the Cauchy-Schwarz inequality and vanishes only for $p = p_\infty$.

7. Γ-calculus – the Bakry-Emery approach

In this section we assume detailed balance, i.e. the existence of a positive equilibrium distribution p_∞ ($\mathsf{L}^* p_\infty = 0$), satisfying $p_\infty \mathsf{L} f = \mathsf{L}^*(p_\infty f)$ for all f. With the observation that then also $p_\infty e^{\mathsf{L}t} f = e^{\mathsf{L}^* t}(p_\infty f)$ holds, the right hand side of (14) (local relative entropy) with $q = p_\infty$ can be written as

$$e^{\mathsf{L}^*(t-s)}h_U(p(s)|p_\infty) = p_\infty \Psi(t-s)\,, \qquad \text{with } \Psi(s) = e^{\mathsf{L}s}U\left(\frac{p(t-s)}{p_\infty}\right),$$

where the definition of Ψ is motivated by the Γ-calculus [4], since

$$\partial_s \Psi(s) = e^{\mathsf{L}s}\Gamma_U(u(t-s))\,,$$

with

$$\Gamma_U(u) = \mathsf{L}(U(u)) - \partial_t U(u) = \mathsf{L}(U(u)) - U'(u)\mathsf{L}u\,, \quad u = \frac{p}{p_\infty}\,.$$

For the choice $U(z) = z^2/2$, Γ_U is the *carré du champ* operator [4], which can be derived as the quadratic form corresponding to the bilinear expression

$$\frac{1}{2}\Big(\mathsf{L}(uv) - u\mathsf{L}v - v\mathsf{L}u\Big)\,.$$

Note that for arbitrary U the local Fisher information can be expressed in terms of Γ_U:

$$i_U(p|p_\infty) = p_\infty \Gamma_U\left(\frac{p}{p_\infty}\right).$$

A relation like the Poincaré inequality (18) can in general not be expected between the local quantities Ψ and $\partial_s\Psi$ (or between the local relative entropy and the local Fisher information). The *Bakry-Emery approach* [3] (see also [2]) is based on another differentiation,

$$\partial_s^2\Psi(s) = e^{\mathsf{L}s}\Gamma_{U,2}(u(t-s))\,, \qquad \text{with } \Gamma_{U,2}(u) = L\Gamma_U(u) - \partial_t\Gamma_U(u)\,,$$

and to relate $\partial_s\Psi$ and $\partial_s^2\Psi$.

Example 9. For the diffusion process described by the Fokker-Planck equation (15) 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

$$\begin{aligned}\Gamma_{U,2}(u) &= 2U^{(2)}(u)\nabla_x u^{tr}\nabla_x^2\Phi\nabla_x u + 2U^{(2)}(u)|\nabla_x^2 u|^2 \\ &\quad +4U^{(3)}(u)\nabla_x u^{tr}\nabla_x^2 u\nabla_x u + U^{(4)}(u)|\nabla_x u|^4\,.\end{aligned}$$

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

$$2(U^{(3)})^2 \le U^{(2)}U^{(4)}\,.$$

A family of admissible choices is given by

$$U_r(z) = \frac{z^r - 1 - r(z-1)}{r-1}\,, \quad 1 < r \le 2\,, \qquad U_1(z) = z\log z - z + 1\,,$$

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

Under the *Bakry-Emery condition* $\nabla_x^2\Phi \ge \lambda/2 > 0$, $\Gamma_{U,2}(u) \ge \lambda\Gamma_U(u)$ follows, with the consequence $\partial_s^2\Psi \ge \lambda\partial_s\Psi$. An application of the Gronwall lemma implies the pointwise estimate

$$\Gamma_U(u)(t) \le e^{-\lambda s}e^{\mathsf{L}s}\Gamma_U(u)(t-s) \quad \text{for } s \le t\,.$$

Since the semigroup $e^{\mathsf{L}s}$ conserves the integral with weight p_∞, integration and $s = t$ give

$$I_U(p|p_\infty) \le e^{-\lambda t}I_U(p_0|p_\infty)\,,$$

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).

8. 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

$$\begin{aligned} dX &= V\,dt\,, \\ dV &= -(V + \nabla_x\Phi(X))dt + \sqrt{2}\,dB\,, \end{aligned}$$

is governed by the kinetic transport equation

$$\partial_t f + \mathsf{T}f = \mathsf{C}f\,, \tag{20}$$

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

(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 $\mathbb{T}_x^d$, i.e. $(\mathsf{T}f)(x,v) = v\cdot\nabla_x f(x,v)$, $(x,v) \in \mathbb{T}_x^d \times \mathbb{R}_v^d$.

The solution, subject to the initial condition $f(t=0)=f_0$, is written as $f(t)=e^{\mathsf{L}t}f_0$, $\mathsf{L}=\mathsf{C}-\mathsf{T}$. As $t\to\infty$, it is expected to converge to a constant multiple of $F(x,v)=\exp(-|v|^2/2-\Phi(x))\in L^1(\mathbb{R}^{2d})$, where the constant is determined from the initial condition by mass conservation:

$$\int e^{\mathsf{L}t}f_0\,dv\,dx=\int f_0\,dv\,dx\,,\qquad t\geq 0\,.$$

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 (20) 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

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

Contrary to the previous section, the entropy dissipation rate

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

is not coercive and vanishes already for *local equilibria* $f=h(x)e^{-|v|^2/2}\in\mathcal{N}(\mathsf{C})$. Despite of this fact, decay to global equilibrium (with $h(x)=\text{const}$) is expected. This property is then called *hypocoercivity* [26]. 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 [26] 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 Γ_2 calculus, an extension of the Γ calculus discussed above. Smoothness requirements for the initial conditions have been removed in [11] by a new strategy for finding modified entropies.

9. 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 $\mathbb{T}^d_x\times\mathbb{R}^d_v$. We also introduce the new unknown

$$h(t,x,v):=e^{|v|^2/2}f(t,x,v)\,,$$

leading to the evolution equation

$$\partial_t h=\mathsf{L}h\,,\qquad\text{with }\mathsf{L}=\Delta_v-v\cdot\nabla_v-v\cdot\nabla_x\,.$$

Mass conservation now reads

$$\int e^{\mathsf{L}t}h_0d\mu=\int h_0d\mu\,,$$

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

$$\Gamma(f,g)=\frac{1}{2}(\mathsf{L}(fg)-f\mathsf{L}g-g\mathsf{L}f)\,,\quad\Gamma(f):=\Gamma(f,f)=|\nabla_v f|^2\,,$$

and its iteration

$$\Gamma_2(f) = \frac{1}{2}\mathsf{L}\Gamma(f,f) - \Gamma(f,\mathsf{L}f) = |\nabla_v^2 f|^2 + |\nabla_v f|^2 + \nabla_v f\cdot\nabla_x f\,,$$

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

$$\Gamma_2^Z(f) = \frac{1}{2}\mathsf{L}\Gamma^Z(f,f) - \Gamma^Z(f,\mathsf{L}f) = |Z\otimes\nabla_v f|^2 + Zf\cdot\nabla_v f + Zf\cdot\nabla_x f\,.$$

Another preparation step is the simple proof of

$$\Gamma_2(f) + \Gamma_2^Z(f) \ge 2|\nabla_x f + \nabla_v f|^2 = \frac{1}{4}\Gamma^Z(f) - \frac{1}{2}\Gamma(f) + \left|\nabla_x f + \frac{3}{2}\nabla_v f\right|^2\,. \tag{21}$$

Finally, the Poincaré inequality

$$\int (\Gamma(f) + \Gamma^Z(f))d\mu \ge \kappa\int f^2 d\mu$$

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

$$\Psi(s) = e^{\mathsf{L}s}\left(\frac{3}{4}h(t-s)^2 + \Gamma(h(t-s)) + \Gamma^Z(h(t-s))\right)$$

and compute

$$\Psi'(s) = e^{\mathsf{L}s}\left(\frac{3}{2}\Gamma(h) + 2\Gamma_2(h) + 2\Gamma_2^Z(h)\right)(t-s) \ge \frac{1}{2}e^{\mathsf{L}s}(\Gamma(h) + \Gamma^Z(h))(t-s)\,,$$

using (21) and the maximum principle. Now we integrate and use mass conservation and the Poincaré inequality:

$$\frac{d}{ds}\int\Psi(s)d\mu \ge \int\left[\frac{1-\varepsilon}{2}(\Gamma(h)+\Gamma^Z(h)) + \frac{\varepsilon\kappa}{2}f^2\right](t-s)d\mu = \lambda\int\Psi(s)d\mu\,,$$

with $\lambda = \frac{2\kappa}{3+4\kappa}$ (for optimal ε). The consequence

$$\int\Psi(t)d\mu \ge e^{\lambda t}\int\Psi(0)d\mu$$

is equivalent to

$$\int\left(\frac{3}{4}h(t)^2 + \Gamma(h(t)) + \Gamma^Z(h(t))\right)d\mu \le e^{-\lambda t}\int\left(\frac{3}{4}h_0^2 + \Gamma(h_0) + \Gamma^Z(h_0)\right)d\mu\,,$$

which in turn is equivalent to exponential decay of the $H^1(d\mu)$-norm of h (by $\Gamma(h)+\Gamma^Z(h) \ge (3-\sqrt{2})(|\nabla_x h|^2 + |\nabla_v h|^2)$) 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 [26].

10. A modified entropy

Here we shall apply the methodology of [11] to

$$\partial_t f + \mathsf{T} f = \mathsf{C} f\,, \qquad \mathsf{T} f = v\cdot\nabla_x f\,, \quad \mathsf{C} f = \nabla_v\cdot(\nabla_v f + v f)\,, \tag{22}$$

$x\in\mathbb{T}^d$, $v\in\mathbb{R}^d$, $t>0$, 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 T is skew symmetric and the collision operator C symmetric negative semidefinite with respect to $L^2(d\sigma)$, $d\sigma = e^{|v|^2/2}dx\,dv$.

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

The simplest example. is a second order ODE system with

$$\mathsf{T} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad \mathsf{C} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}.$$

Although the entropy dissipation $-\frac{d}{dt}|f|^2 = 2f_2^2$ is only semidefinite, the eigenvalues of $\mathsf{C}-\mathsf{T}$ 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 $\mathcal{N}(\mathsf{C}) = \{(f_1,0)\}$ by the rotational motion caused by 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

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

with small positive ε, and with the dissipation

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

For small enough ε, both H and its dissipation are obviously coercive, implying exponential decay.

The general approach. With the ortogonal projection $\Pi:\ \mathcal{H}\to\mathcal{N}(\mathsf{C})$, the main ingredients of the general approach are the assumptions of

H1. Diffusive macroscopic limit: $\Pi\mathsf{T}\Pi = 0$,
H2. Microscopic coercivity: $-\langle \mathsf{C}f, f\rangle \ge \lambda_m\|(1-\Pi)f\|^2$, $\lambda_m>0$, and
H3. Macroscopic coercivity (or again 'instability of hydrodynamic states'):
$\|\mathsf{T}\Pi f\|^2 \ge \lambda_M\|\Pi f\|^2$, $\lambda_M>0$.

The modified entropy is then given by

$$H[f] = \frac{1}{2}\|f\|^2 + \varepsilon\langle \mathsf{A}f, f\rangle\,, \qquad \mathsf{A} = [1+(\mathsf{T}\Pi)^*\mathsf{T}\Pi]^{-1}(\mathsf{T}\Pi)^*\,,$$

and it satisfies

$$\frac{d}{dt}H[f] = \langle \mathsf{C}f, f\rangle - \varepsilon\langle \mathsf{A}\mathsf{T}\Pi f, f\rangle - \varepsilon\langle \mathsf{A}\mathsf{T}(1-\Pi)f, f\rangle + \varepsilon\langle \mathsf{T}\mathsf{A}f, f\rangle + \varepsilon\langle \mathsf{A}\mathsf{C}f, f\rangle\,.$$

It has been proved in [11] that the operators A and TA are bounded. Under the additional assumption of

H4. Boundedness of the auxiliary operators AT and AC,

the estimate

$$\frac{d}{dt}H[f] \le -\lambda_m\|(1-\Pi)f\|^2 - \frac{\varepsilon\lambda_M}{1+\lambda_M}\|\Pi f\|^2 + \varepsilon C\|(1-\Pi)f\|\,\|f\|$$

holds, implying (as for the model problem) that both H and its dissipation are coercive for ε small enough.

The kinetic Fokker-Planck equation. With the operators T and C from (22) and $\mathcal{H} = L^2(d\sigma)$ the orthogonal projection to $\mathcal{N}(\mathsf{C})$ is given by

$$(\Pi f)(x,v) = \rho_f(x)F(v)\,, \qquad F(v) = (2\pi)^{-d/2}e^{-|v|^2/2}\,, \quad \rho_f(x) = \int_{\mathbb{R}^d} f(x,v)dv\,.$$

The marginal density ρ_f is called the macroscopic or position density. Assumption H1 is a consequence of the fact that the flux associated to local equilibria of the form $\rho(x)F(v)$ vanishes: $\int vF\,dv = 0$. It is called 'diffusive macroscopic limit' since it is responsible for the fact that the diffusive rescaling $t \to t/\varepsilon^2$, $x \to x/\varepsilon$ in (22) and the subsequent macroscopic limit $\varepsilon \to 0$ lead to a diffusion equation for the limit of ρ_f.

Assumption H2 is equivalent to the weighted Poincaré inequality

$$\int_{\mathbb{R}^d} |\nabla_v h|^2\, F\,dv \ge \lambda_m \int_{\mathbb{R}^d} \left(h - \int hF\,dv\right)^2 F\,dv\,,$$

which holds for the Gaussian weight F. Similarly the macroscopic coercivity assumption H3 is equivalent to the Poincaré inequality on the torus,

$$\int_{\mathbb{T}^d} |\nabla_x \rho|^2 dx \ge \lambda_M \int_{\mathbb{T}^d} \rho^2 dx\,, \qquad \text{for } \int_{\mathbb{T}^d} \rho\,dx = 0\,.$$

The boundedness of AT is equivalent to an elliptic regularity result, the boundedness of Δ_x^{-1} as a map from $L^2_0(\mathbb{T}^d)$ to $H^2_0(\mathbb{T}^d)$, 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 $\rho_{\mathsf{TC}f} = -\nabla_x \cdot \int vf\,dv$, showing that AC does not contain v-derivatives. Thus, H1-H4 holds with the consequence that solutions of (22) 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, 26].

The methods of [5, 11, 26] 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 $\mathsf{L} = \mathsf{C} - \mathsf{T}$ does not have any smoothing properties.

References

[1] F. Achleitner, A. Arnold, D. Stuerzer, Large-time behavior in non-symmetric Fokker-Planck equations, *Riv. Mat. Univ. Parma* (2014).

[2] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations, *Comm. PDE* **26** (2001), pp. 43–100.

[3] D. Bakry, M. Emery, Hypercontractivité de semi-groupes de diffusion, *C.R. Acad. Sc. Paris,* Série I **299**(15) (1984).

[4] D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Springer, 2013.
[5] F. Baudoin, Bakry-Emery meet Villani, *arXiv:* 1308.4938v1, 2013.
[6] N. BenAbdallah, H. Chaker, C. Schmeiser, The high field asymptotics for a fermionic Boltzmann equation: entropy solutions and kinetic shock profiles, *J. of Hyperbolic Diff. Equs.* **4** (2007), pp. 679–704.
[7] P. Degond, T. Goudon, F. Poupaud, Diffusion limit for nonhomogeneous and non-microreversible processes, *Indiana Univ. Math. J.* **49** (2000), pp. 1175–1198.
[8] A. Dembo, *Stochastic Processes*, lecture notes, Stanford Univ., 2013.
[9] L. Desvillettes and C. Villani, On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation, *Comm. Pure Appl. Math.* **54** (2001), pp. 1–42.
[10] L. Desvillettes and C. Villani, On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation, *Invent. Math.* **159** (2005), pp. 245–316.
[11] J. Dolbeault, C. Mouhot, C. Schmeiser, Hypocoercivity for linear kinetic equations conserving mass, *Trans. AMS* **367** (2015), pp. 3807–3828.
[12] L. Edelstein-Keshet and G.B. Ermentrout, Models for the length distribution of actin filaments: I. Simple polymerization and fragmentation, *Bull. Math. Biol.* **60** (1998), 449–475.
[13] J. Fontbona, B. Jourdain, A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations, *Ann. Prob.* **44** (2016), pp. 131–170.
[14] U.G. Haussmann, E. Pardoux, Time reversal of diffusions, *Ann. of Prob.* **14** (1986), pp. 1188–1205.
[15] F. Herau, F. Nier, Isotropic hypoellipticity and trend to the equilibrium for the Fokker-Planck equation with high degree potential, *Archive Rat. Mech. Anal.* **171** (2004), pp. 151–218.
[16] P. Laurencot, B. Perthame, Exponential decay for the growth-fragmentation/cell-division equation, *Comm. Math. Sci.* **7** (2009), pp. 503–510.
[17] A. Manhart, D. Oelz, C. Schmeiser, N. Sfakianakis, An extended Filament Based Lamellipodium Model produces various moving cell shapes in the presence of chemotactic signals, *J. Theor. Biol.* **382** (2015), pp. 244–258.
[18] P. Michel, S. Mischler, B. Perthame, General relative entropy inequality: an illustration on growth models, *J. Math. Pures et Appl.* **84** (2005), pp. 1235-1260.
[19] S. Mischler, B. Perthame, L. Ryzhik, Stability in a Nonlinear Population Maturation Model, *Math. Mod. Meth. Appl. Sci.* **12** (2002), pp. 1751-1772.
[20] C. Mouhot, L. Neumann, Quantitative perturbative study of convergence to equilibrium for collisional kinetic models in the torus, *Nonlinearity*, **19** (2006), pp. 969–998.
[21] C. Mouhot, E. Russ, Y. Sire, Fractional Poincaré inequalities for general measures, *J. Math. Pures Appl.* **95** (2011), pp. 72–84.
[22] B. Perthame, *Transport Equations in Biology,* Birkhäuser, 2007.
[23] F. Poupaud, Runaway phenomena and fluid approximation under high fields in semiconductor kinetic theory, *ZAMM* **72** (1992), pp. 359–372.
[24] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry,* North-Holland, 1992.
[25] R.S. Varga, *Matrix Iterative Analysis,* 2nd ed., Springer-Verlag, 2002.
[26] C. Villani, *Hypocoercivity*, Mem. AMS **202**, 2009.

C. Schmeiser: Fakultät für Mathematik, Universität Wien, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria

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