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.

CONTENTS

1. Motivation

We consider a system of linear ODEs with constant coefficients:

$$(1) \qquad \frac{dp}{dt} = Ap \,,$$

with unknown $p(t) = (p_1(t), \ldots, p_N(t)) \in \mathbb{R}^N$, $t \geq 0$, and given constant $A \in \mathbb{R}^{N \times N}$. We assume that the initial datum $p(0)$ is a probability distribution on $\{1, \ldots, N\}$, 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

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

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

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

Using this, a straightforward computation gives

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

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

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., [41]). 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_∞. $\qquad \square$

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

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

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 square of the weighted ℓ^2-Norm of $p - p_\infty$) $H(p|p_\infty)$ to zero. $\square$

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., [12] 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.*

$$X_t = (X_{1,t}, \ldots, X_{d,t}) : \Omega \to \mathbb{R}^d \qquad and$$
$$\{\omega \in \Omega : X_{i,t}(\omega) \leq \alpha\} \in \mathcal{F}, \qquad i = 1, \ldots, d, \quad t \geq 0, \quad \alpha \in \mathbb{R}.$$

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

The *Borel 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* $p(x,t)$, $x \in \mathbb{R}^d$, $t \geq 0$:

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

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 $p \geq 0$ 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 [12].

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 $0 \leq t_1 < \ldots < t_n$. 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

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

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 complete definition of martingales and several of their properties can be found in [12].

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. 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*

(3) $\partial_s u + \mathsf{L}u = 0 \,, \qquad 0 \le s \le t \,,$

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.

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

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

(5) $p_{(X_t, X_s)}(x, y) = (e^{L^*(t-s)} \delta_y)(x) p(s, y)$

for (X_t, X_s). Therefore

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

(6)

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

(7) $0 = E\left(\left[f(X_t) - f(X_s) - \displaystyle\int_s^t \mathsf{L}f(X_r)dr \right] g(X_s) \right) \,, \qquad t \ge s \ge 0 \,.$

3. 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

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

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*, which means that the law of $B_t - B_s$, $t \geq s$, only depends on $t - s$. One consequence of this, which shall be used below, is

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

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:

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

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

(9) $$dX_t = b(X_t)dt + \sigma dB_t$$

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 (9) 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 $\Delta t > 0$

$$U_{t+\Delta t} - U_t = u(X_{t+\Delta t}, t + \Delta t) - u(X_t, t) \approx \partial_t u(X_t, t)\Delta t$$

$$(10) \quad + \nabla_x u(X_t, t) \cdot (X_{t+\Delta t} - X_t) + \frac{1}{2}(X_{t+\Delta t} - X_t)^{tr}\nabla_x^2 u(X_t, t)(X_{t+\Delta t} - X_t).$$

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 (8), considering $B_{t+\Delta t} - B_t$ as an $O(\sqrt{\Delta t})$-term, the approximation in (10) means that terms up to $O(\Delta t)$ have been kept. The approximation can be simplified further:

$$U_{t+\Delta t} - U_t = (\partial_t u(X_t, t) + b(X_t) \cdot \nabla_x u(X_t, t))\,\Delta t + \nabla_x u(X_t, t) \cdot (\sigma(B_{t+\Delta t} - B_t))$$

$$(11) \qquad + \frac{1}{2}(B_{t+\Delta t} - B_t)^{tr}\sigma^{tr}\nabla_x^2 u(X_t, t)\sigma(B_{t+\Delta t} - B_t) + O(\Delta t^{3/2}).$$

Now we compute the expectation (using the martingale property of U_t, as well as (8) 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 (11), 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. 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,

$$(12) \qquad P(N_t^n = l) = \binom{k}{l}\left(\frac{\lambda}{n}\right)^l \left(1 - \frac{\lambda}{n}\right)^{k-l}, \qquad \frac{k}{n} \le t < \frac{k+1}{n}.$$

We also get

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

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 λ*, satisfying

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

which can be obtained from (12) 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. SPACE HOMOGENEITY – LÉVY PROCESSES

We call a (time) homogeneous Markov process *space homogeneous,* if its transition probabilities $\pi_t(x, y)$ are translation invariant with respect to (x, y), i.e. $\pi_t(x, y) = \pi_t(x - y)$. Recalling for this case the joint probability distribution $p_{(X_t, X_s)}(x, y) = \pi_{t-s}(x - y)p(s, y)$ for (X_t, X_s), $t \geq s$, (see (5)), we can determine the law of increments from

$$E_{X_t - X_s}(f) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x - y)\pi_{t-s}(x - y)p(y, s)dx\,dy = \int_{\mathbb{R}^d} f(z)\pi_{t-s}(z)dz\,.$$

This shows that a time and space homogenous Markov process has stationary increments.

Lévy processes are CTSP X_t with independent, stationary increments, satisfying $X_0 = 0$. Examples are diffusion processes with constant drift, i.e. $X_t = \sigma B_t + tb$, and compound Poisson processes with $X_0 = 0$ and translation invariant jump rate $W(x' \to x) = W(x - x')$.

By $X_t = \sum_{k=1}^n (X_{kt/n} - X_{(k-1)t/n})$, a Lévy process at each time $t \geq 0$ can be written as the sum of arbitrarily many i.i.d. random vectors, a property called *infinite divisibility.* It can be used for proving the space homogeneity of Lévy processes with the aid of the *characteristic function* $\chi(\xi, t) := E(e^{i\xi \cdot X_t})$, i.e. the Fourier transform of $p(t)$. The above splitting leads to

$$\chi(\xi, t) = (\chi(\xi, t/n))^n\,,$$

implying, at least formally, $\chi(\xi, t) = e^{-t\Psi(\xi)}$ with some *characteristic exponent* $\Psi(\xi)$. Since $\chi(\cdot, t)$ is the Fourier transform of $p(t) = e^{L^* t}\delta_0$ ($p_0 = \delta_0$ since $X_0 = 0$), it is obvious that L^* is a pseudo-differential operator with symbol $-\Psi$, i.e. application of L^* corresponds to multiplication by $-\Psi$ in Fourier variables. Therefore L^* is a convolution operator and, thus, translation invariant in space.

Infinite divisibility straightforwardly implies

$$E(X_t) = tE(X_1)\,, \qquad V(X_t) = tV(X_1)\,,$$

which is used in the following computation, where we restrict ourselves to $d = 1$ for simplicity:

$$\begin{aligned}
\Psi(\xi) &= -\log \chi(\xi, 1) = -\partial_t(\chi(\xi, 1)^t)\,|_{t=0} = \lim_{t \to 0} \frac{1 - \chi(\xi, t)}{t} \\
&= \lim_{t \to 0} \frac{1}{t} E(1 - e^{i\xi X_t} + i\xi X_t) - i\xi E(X_1) \\
&= \lim_{t \to 0} \int_{\mathbb{R}} \frac{1 - e^{i\xi y} + i\xi y}{y^2} \mu_t(y)dy - i\xi E(X_1)\,,
\end{aligned}$$

where $\mu_t(y) = \frac{y^2 p(y, t)}{t}$ is bounded in L^1 as $t \to 0$, since

$$\int_{\mathbb{R}} \mu_t(y)dy = \frac{E(X_t^2)}{t} = V(X_1) + tE(X_1)^2\,,$$

and the function $\frac{1 - e^{i\xi y} + i\xi y}{y^2}$ can be continuously extended to $y = 0$ by the value $\xi^2/2$. Denoting by μ_0 a (vage, restricting to a subsequence) limit of μ_t as $t \to 0$, we obtain

$$\Psi(\xi) = \int_{\mathbb{R}} \frac{1 - e^{i\xi y} + i\xi y}{y^2} \mu_0(dy) - i\xi E(X_1)\,.$$

Note that μ_0 is a bounded measure with $\mu_0(\mathbb{R}) = V(X_1)$. With the notation

$$b := E(X_1) - \int_{|y|>1} \frac{\mu_0(dy)}{y}, \qquad D = \frac{\mu_0(\{0\})}{2}, \qquad \nu(dy) = \frac{\mu_0(dy)}{y^2} \quad \text{for } y \neq 0,$$

this becomes the *Lévy-Khinchine representation*

$$\Psi(\xi) = -i\xi b + D\xi^2 + \int_{y\neq 0} (1 - e^{i\xi y} + i\xi y \mathbb{1}_{|y|<1})\nu(dy).$$

An ambiguity in the construction affecting the definition of b is the cutoff for large values of y. Sometimes it might be more convenient to replace $y\mathbb{1}_{|y|<1}$ by something continuous like $\frac{y}{1+y^2}$.

The general result is that also for $d > 1$ and without the assumption of bounded variance the characteristic exponent has a representation

$$(13) \qquad \Psi(\xi) = -i\xi \cdot b + \xi \cdot (D\xi) + \int_{y\neq 0} (1 - e^{i\xi \cdot y} + i\xi \cdot y\mathbb{1}_{|y|<1})\nu(dy),$$

with a vector $b \in \mathbb{R}^d$, a positive semidefinite matrix $D \in \mathbb{R}^{d \times d}$, and the *Lévy measure* ν, satisfying $\int_{y\neq 0} \min\{1, |y|^2\}\nu(dy) < \infty$. Inverse Fourier transformation gives

$$(14) \qquad \begin{aligned} (L^*p)(x) &= \nabla \cdot (D\nabla p(x) - bp(x)) \\ &\quad + \int_{y\neq 0} \big(p(x+y) - p(x) - y \cdot \nabla p(x)\mathbb{1}_{|y|<1}\big)\,\nu(dy). \end{aligned}$$

The first part describes a diffusion process with constant drift. The second needs a little explanation. If the Lévy measure is integrable around $y = 0$, the gradient term in the integral could be included in the definition of the drift vector b. The integral then describes a jump process with $\nu(dy) = W(y)dy$ in the notation of the previous section.

If ν has a stronger singularity at $y = 0$, the gradient term might be necessary for the existence of the integral. The standard example for this situation is *fractional diffusion* with $W(y) = |y|^{-d-\alpha}$, where $0 < \alpha < 2$, and with

$$\Psi(\xi) = \int_{\mathbb{R}^d} e^{i\xi \cdot y}|y|^{-d-\alpha}dy = |\xi|^\alpha \int_{\mathbb{R}^d} \cos(z_1)|z|^{-d-\alpha}dz = c_{\alpha,d}|\xi|^\alpha.$$

By the symmetry of W the gradient term in the integral can be dropped, if the integral is interpreted as principal value.

6. Time reversal – decay of the relative entropy

We are looking for a time reversed version of (7) 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

$$(15) \qquad \overline{\mathsf{L}}^{p(t)}f = \frac{1}{p(t)}(\mathsf{L}^*(fp(t)) - f\mathsf{L}^*p(t)),$$

implying

$$(16) \quad 0 = E\left(\left[f(X_t) - f(X_s) + \int_s^t \overline{\mathsf{L}}^{p(r)} f(X_r)dr\right] g(X_t)\right), \qquad t \geq s \geq 0\,.$$

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 $t \geq s$,

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

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*

$$(17) \quad h_U(p(t)|q(t))(x) := q(t, x) U\left(\frac{p(t, x)}{q(t, x)}\right) \leq \left[e^{\mathsf{L}^*(t-s)} h_U(p(s)|q(s))\right](x)\,.$$

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. *(The heat equation)* For better appreciation of the local entropy inequality (17), we reformulate the proof for a simple example. Note that there is no integration by parts.

Lemma 4. *Let $p(x,t) \geq 0$ and $q(x,t) > 0$ satisfy $\partial_t p = \Delta_x p$, $\partial_t q = \Delta_x q$, $t > 0$, and let $U \in C^2([0,\infty))$ be convex. Let furthermore $u(x,t,s)$, $t \geq s > 0$, satisfy*

$$\partial_t u = \Delta_x u\,, \quad t \geq s\,, \qquad u(t=s) = h_U(p(s)|q(s))\,.$$

Then

$$h_U(p(t)|q(t))(x) \leq u(x,t,s)\,, \qquad 0 < s \leq t\,, \; x \in \mathbb{R}^d\,.$$

Proof: Let $v(x,\tau,s) := u(x, s+\tau, s)$. Then

$$\partial_\tau v = \Delta_x v\,, \qquad v(\tau = 0) = h_U(p(s)|q(s))\,.$$

Therefore, of course,

$$\partial_\tau(\partial_s v) = \Delta_x(\partial_s v)\,, \qquad \partial_s v(\tau = 0) = \partial_s h_U(p(s)|q(s))\,.$$

Since $u(x,t,s) = v(x, t-s, s)$, we have $w := \partial_s u = -\Delta_x u + \partial_s v$, which solves

$$
\begin{aligned}
\partial_t w = \Delta_x w\,, \qquad w(t=s) &= -\Delta_x h(p(s)|q(s)) + \partial_s h_U(p(s)|q(s)) \\
&= -U''\left(\frac{p(s)}{q(s)}\right) q(s) \left|\nabla_x \frac{p(s)}{q(s)}\right|^2 \leq 0\,.
\end{aligned}
$$

The maximum principle implies $w = \partial_s u \leq 0$ which completes the proof. $\square$

Example 2. *(Diffusion processes – the Fokker-Planck equation)* The generator of the time reversal of the diffusion process [22] is, according to (15), 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*

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

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 3. (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.*

Example 4. (Lévy processes) For a general Lévy process with generator (14) and characteristic exponent (13) the existence of a stationary distribution cannot be expected. For diffusion the addition of a drift resulting from a confining potential Φ helped, with the classical example $\Phi(x) = |x|^2/2$ leading to a Gaussian equilibrium. In [20], *Lévy-Fokker-Planck operators* of the form

$$L^*p = L^*_{\text{Lévy}}p + \nabla \cdot (xp)$$

with an operator $L^*_{\text{Lévy}}$ of the form (14), have been studied. Equilibria can be constructed by employing the Fourier transform. The Fourier transform

$$\widehat{p_\infty}(\xi) = \int_{\mathbb{R}^d} e^{i\xi \cdot x} p_\infty(x)dx$$

of an equilibrium distribution has to satisfy

$$\Psi\widehat{p_\infty} + \xi \cdot \nabla\widehat{p_\infty} = 0,$$

with the auxiliary condition $\widehat{p_\infty}(0) = 1$. The ansatz $\widehat{p_\infty}(\xi) = e^{-A(\xi)}$ leads to

$$\xi \cdot \nabla A = \Psi, \qquad A(0) = 0,$$

with the solution

$$A(\xi) = \int_0^1 \frac{\Psi(s\xi)}{s}ds = -i\xi \cdot b + \frac{1}{2}\xi \cdot (D\xi) + \int_{y\neq 0}\int_0^{\xi \cdot y} \frac{1 - e^{iu} + iu\mathbb{1}_{|y|<1}}{u}du\,\nu(dy).$$

For a diffusion process the equilibrium distribution is a Gaussian with mean b. On the other hand for the fractional diffusion case $\Psi(\xi) = |\xi|^\alpha$ we obtain $A(\xi) = |\xi|^\alpha/\alpha$, i.e. p_∞ is the transition probability density of the pure fractional diffusion process for time intervals of length α^{-1}.

The author is not aware of results concerning the time reversibility of the stationary process with probability density p_∞.

7. The entropy dissipation rate – decay to equilibrium by entropy/entropy-dissipation 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 (17):

$$(19) \qquad \partial_s \left(e^{\mathsf{L}^*(t-s)} \left[q(s)U\left(\frac{p(s)}{q(s)}\right) \right] \right) = -e^{\mathsf{L}^*(t-s)} i_U(p(s)|q(s)) \,,$$

with the *local Fisher information*

$$(20) i_U(p|q) = \mathsf{L}^*\left(qU\left(\frac{p}{q}\right) \right) - U'\left(\frac{p}{q}\right) \mathsf{L}^* p - \left(U\left(\frac{p}{q}\right) - \frac{p}{q} U'\left(\frac{p}{q}\right) \right) \mathsf{L}^* q \geq 0 \,.$$

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 *entropy/entropy-dissipation inequality*

$$(21) \qquad\qquad I_U(p|p_\infty) \geq \lambda H_U(p|p_\infty) \,, \qquad \text{with } \lambda > 0$$

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

Example 5. *(Diffusion processes)* For the diffusion process of Example 2 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

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

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 (21) 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 6. *(Jump processes)* The jump process of Example 3 has the local Fisher information

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

The term in the bracket is called the *Bregman distance* (associated to U) between p/q and p'/q'.

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 [11] (see, however, [40]).

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 entropy/entropy-dissipation 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 (21) and therefore exponential decay of the relative entropy but, by (19), 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 $|x| \to \infty$ [38]. 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. [7]).

Example 7. *(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 [17, 30]

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

$x > 0$, 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. [29]), 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:

$$(22) \qquad i_U(p|p_\infty)(x) = \frac{1}{2} \int_M H(x' - x)p'_\infty \left(\frac{p}{p_\infty} - \frac{p'}{p'_\infty} \right)^2 dx' \,.$$

This does not permit a local entropy/entropy-dissipation 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 8. *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 entropy/entropy-dissipation 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 [35].

Example 9. The concept of *generalized relative entropy*, developed in [32, 33], also fits into the framework of Markov processes after a transformation. For illustration we consider the finite dimensional situation as in [37], Section 6.3. Consider a

matrix $A = (a_{mn})_{m,n=1,\ldots,d}$ 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

(23)
$$p_n(t) = \frac{e^{-\lambda t}\phi_n u_n(t)}{\sum_{m=1}^{d} \phi_m u_m(0)}\,,$$

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

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

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 (20) 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 10. 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 [37]). 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 $u(a,t) = e^{\lambda t}R(a)$ 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

$$\lambda R + \partial_a R = -\nu R\,, \qquad R(0) = \int_0^\infty BR\,da\,, \qquad \int_0^\infty R\,da = 1\,,$$

$$\lambda \phi - \partial_a \phi = -\nu \phi + \phi(0)B\,, \qquad \int_0^\infty R\phi\,da = 1\,,$$

and $R, \phi > 0$ has been proven in [37].

Analogously to (23), 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

$$(24) \quad \partial_t p + \partial_a p = -\frac{\phi(0)B}{\phi}p\,, \qquad p(a = 0) = \int_0^\infty \frac{\phi(0)B(a')}{\phi(a')}p(a')da'\,.$$

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 (20) 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 again [37] for details):

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

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

However, there is an alternative way to arrive at the entropy dissipation: The problem for p can be interpreted as containing a jump process, where all jumps go to $a = 0$ with the rate $W(a' \to a) = \delta(a)\phi(0)B(a')/\phi(a')$, leading to an equivalent formulation of (24):

$$\partial_t p + \partial_a p = \delta(a)\int_0^\infty \frac{\phi(0)B(a')}{\phi(a')}p(a')da' - \frac{\phi(0)B}{\phi}p\,.$$

Employing (22), we arrive at an alternative representation of the relative entropy dissipation, corresponding to $U(z) = (z - 1)^2/2$:

$$\begin{aligned} I_U(p|p_\infty) &= \frac{1}{2}\int_0^\infty \frac{\phi(0)B}{\phi}p_\infty\left(\frac{p(0)}{p_\infty(0)} - \frac{p}{p_\infty}\right)^2 da \\ &= \frac{1}{2}\int_0^\infty \frac{\phi(0)B}{\phi}\frac{(p - p_\infty)^2}{p_\infty}da + \frac{3}{2}\frac{(p(0) - p_\infty(0))^2}{p_\infty(0)}\,. \end{aligned}$$

It is now obvious that exponential decay can be shown under the assumption $\phi(0)B/\phi \geq \lambda > 0$, which is also used in [37].

Example 11. The idea of the previous example can be used in various situations. We look at a simple example. Consider the *heat equation* on a bounded domain $\Omega \subset \mathbb{R}^d$ with homogeneous Dirichlet boundary conditions, i.e.

$$\partial_t u = \Delta_x u, \quad \text{on } \Omega, \qquad u = 0, \quad \text{on } \partial\Omega,$$

and positive square integrable initial data: $u(x,0) = u_0(x) > 0$. Let λ be the largest eigenvalue of the Laplacian with these boundary conditions and $\phi(x) > 0$ the corresponding eigenfunction normalized to $\int_\Omega \phi^2 dx = 1$. Then we have

$$\langle \phi u(t) \rangle = e^{\lambda t} \langle \phi u_0 \rangle \,.$$

Therefore

$$p(x,t) := \frac{\phi(x)u(x,t)e^{-\lambda t}}{\langle \phi u_0 \rangle}$$

is for each $t \geq 0$ a probability density on Ω. It satisfies the Fokker-Planck equation

$$\partial_t p = -\lambda p + \phi \Delta_x \left(\frac{p}{\phi} \right) = \nabla_x \cdot \left(\phi^2 \nabla_x \frac{p}{\phi^2} \right) \,.$$

Since the flux can be written as $\phi^2 \nabla_x \frac{p}{\phi^2} = \phi \nabla_x u - u \nabla_x \phi$, zero flux boundary conditions are appropriate as a consequence of the homogeneous Dirichlet conditions for the original problem.

The optimal constant in the Poincaré inequality with weight ϕ^2 is easily seen to be the difference between the first two eigenvalues of the Laplacian, which is the expected result.

8. Γ-CALCULUS – THE BAKRY-ÉMERY 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 (17) (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 [5], 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 [5], which can be derived as the quadratic form corresponding to the bilinear expression

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

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 (21) 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-Émery approach* [4] (see also [3]) 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 12. For the diffusion process described by the Fokker-Planck equation (18) we assume $D = I_{d \times d}$, i.e. $Lu = \Delta_x u - \nabla_x \Phi \cdot \nabla_x u$, 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

$$2(U^{(3)})^2 \leq U^{(2)}U^{(4)} \quad \Leftrightarrow \quad \left(\frac{1}{U^{(2)}}\right)^{(2)} \leq 0.$$

A family of admissible choices is given by

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

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

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

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

Since the semigroup $e^{\mathsf{L}s}$ conserves the integral with weight p_∞, 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).

9. 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

$$(25) \qquad\qquad\qquad \partial_t f + \mathsf{T}f = \mathsf{C}f,$$

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 flat 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 (25) 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

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

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) = $ const) is expected. This property is then called *hypocoercivity* [42]. The kinetic Fokker-Planck equation is the prototypical example and has witnessed several approaches for proving hypocoercivity. The powerful approach of Desvillettes and Villani [13] has been generalized to prove strong decay to equilibrium for the Boltzmann equation [14], but it only provides algebraic decay for smooth solutions. First results on exponential decay have been based on spectral estimates [23]. Villani [42] has initiated the search for generalized entropies with coercive dissipation rates (see also the earlier work [34]). His method, motivated by hypo-elliptic theory, is based on an H^1 setting. Comparable results have been derived recently by Baudoin [6] employing the Γ_2 calculus, an extension of the Γ calculus discussed above. Smoothness requirements for the initial conditions have been removed in [16] by a new strategy for finding modified entropies.

10. The Mouhot-Neumann-Villani approach

To keep computations simple, we replace the potential $\Phi(x)$, confining in position space, by restriction of the dynamics to a torus represented by $[0,1]^d$. However, in this section we consider a different collision term. We consider the problem

$$\partial_t f + v \cdot \nabla_x f = \mathsf{C}f = \rho_f M - f \,, \qquad f(t = 0) = f_0 \,, \quad \int f_0 \, dv \, dx = 0 \,,$$

where

$$M(v) = (2\pi)^{-d/2} e^{-|v|^2/2} \qquad \text{and} \qquad \rho_f(x,t) = \int f(x,v,t) dv$$

are the normalized Gaussian and, respectively, the position density. The integral condition on the initial datum means that f denotes a perturbation of a nontrivial equilibrium. Integrations with respect to x and v are always over $[0,1]^d$ and, respectively, $\mathbb{R}^d$, if not indicated otherwise. We present a (maybe slightly simpler) variant of the procedure of [34].

The relaxation operator $\mathsf{C}f = \rho_f M - f$ satisfies

$$\int f\,\mathsf{C}f\,\frac{dv}{M} = -\int (\mathsf{C}f)^2\,\frac{dv}{M}\,,$$

suggesting to use the Hilbert space $L^2(dx\,dv/M)$ with scalar product $\langle\cdot,\cdot\rangle$ (different from the earlier sections) and norm $\|\cdot\|$ as functional analytic framework. Obviously, the nullspace of C is spanned by M and $\Pi f = \rho_f M$ is the corresponding orthogonal projection. With this notation, the above equation can be written as

$$\langle \mathsf{C}f, f \rangle = -\|(1-\Pi)f\|^2\,.$$

After this preparation we explain the main idea of the approach: The information missing in the entropy dissipation

$$(26) \qquad \frac{1}{2}\frac{d}{dt}\|f\|^2 = -\|(1-\Pi)f\|^2\,,$$

has to come from the transport term $v\cdot\nabla_x f$. A term with a sign can be produced by taking the gradient with respect to v, followed by multiplication with $\nabla_x f$. This observation leads to the computation

$$(27)\frac{d}{dt}\langle\nabla_x f,\nabla_v f\rangle = -\|\nabla_x f\|^2 - 2\langle(1-\Pi)\nabla_x f,\nabla_v f\rangle + \int vf\,\nabla_x\rho_f\,dv\,dx\,.$$

The first term on the right hand side suggests to add this to the entropy dissipation equation. However, this idea produces a number of difficulties. First, the correction to the entropy is not coercive and, second, it is not clear how to control the other two terms on the right hand side. Coercivity of the entropy can be achieved by extending it to a full norm on $H^1(dx\,dv/M)$. For this purpose we compute

$$(28) \qquad \frac{1}{2}\frac{d}{dt}\|\nabla_x f\|^2 = -\|(1-\Pi)\nabla_x f\|^2\,,$$

and

$$(29) \qquad \frac{1}{2}\frac{d}{dt}\|\nabla_v f\|^2 \quad = \quad -\langle\nabla_x f,\nabla_v f\rangle + d\|\Pi f\|^2 - \|\nabla_v f\|^2\,,$$

where the second term on the right hand side follows from the computation

$$\langle\nabla_v(\rho_f M),\nabla_v f\rangle = -\int \rho_f v\cdot\nabla_v f\,dv\,dx = d\int \rho_f^2\,dx = d\|\Pi f\|^2\,.$$

Now we are ready to define a modified entropy functional by

$$H[f] := \frac{1}{2}\left(\|f\|^2 + \alpha\|\nabla_x f\|^2 + \beta\|\nabla_v f\|^2 + 2\gamma\langle\nabla_x f,\nabla_v f\rangle\right)\,,$$

with $\alpha,\beta,\gamma > 0$, satisfying $\gamma^2 < \alpha\beta$, which makes $\sqrt{H(f)}$ a norm on $H^1(dx\,dv/M)$. With the estimates

$$\left|\int vf\,dv\right| \le \|(1-\Pi)f\|_{L^2(dv/M)} \quad\Rightarrow\quad \left|\int vf\,\nabla_x\rho_f\,dv\,dx\right| \le \|\Pi\nabla_x f\|\|(1-\Pi f)\|$$

and with the Poincaré inequality $\|\Pi\nabla_x f\| \geq \|\Pi f\|$ on the torus we obtain

$$
\begin{aligned}
\frac{d}{dt}H[f] \leq \ & -\left(1 - \frac{\gamma c}{2}\right)\|(1-\Pi)f\|^2 - \left(\alpha + \gamma - \frac{\beta}{2a} - \frac{\gamma}{b}\right)\|(1-\Pi)\nabla_x f\|^2 \\
& -A\|\Pi\nabla_x f\|^2 - \left(\beta - \frac{a\beta}{2} - b\gamma\right)\|\nabla_v f\|^2 \\
& -\left(\gamma - \frac{\gamma}{2c} - \frac{\beta}{2a} - \beta d - A\right)\|\Pi f\|^2 \,,
\end{aligned}
$$

with $A, a, b, c > 0$, assuming that all the parantheses can be made nonnegative. Actually, with

$$
a = c = 1\,, \quad b = \frac{1}{4(1+2d)}\,, \quad \alpha = 4(1+2d) + \frac{3}{8(1+2d)}\,,
$$

$$
\beta = \frac{3}{4(1+2d)}\,, \quad \gamma = 1\,, \quad A = \frac{1}{16}\,,
$$

we get $\alpha\beta - \gamma^2 > 2$ and

$$
\frac{d}{dt}H[f] \leq -\frac{1}{8(1+2d)}\|f\|^2_{H^1(dx\,dv/M)} \leq -\lambda H[f]\,,
$$

for some $\lambda > 0$.

11. Hypoellipticity from a hypocoercivity approach

As opposed to the equation treated in the previous section, solutions of the kinetic Fokker-Planck equation are expected to be infinitely smooth although the evolution operator is not elliptic. This property of the equation is called *hypoellipticity*, and it has been known for this equation since the 1930s, when Kolmogorow explicitly computed a fundamental solution [28].

It has been shown by Villani [42] (Appendix A.21) that hypocoercivity estimates can be used for showing hypoellipticity. The presentation in this section has been inspired by [24].

It is a simple observation that for solutions of the heat equation $\partial_t u = \Delta_x u$ on $\mathbb{R}^d$, the functional

$$
\|u(\cdot,t)\|^2_{L^2(\mathbb{R}^d)} + 2t\|\nabla_x u(\cdot,t)\|^2_{L^2(\mathbb{R}^d)}
$$

is nonincreasing in time, implying the estimate

$$
\|\nabla_x u(\cdot,t)\|_{L^2(\mathbb{R}^d)} \leq \|u(\cdot,0)\|_{L^2(\mathbb{R}^d)}(2t)^{-1/2}\,,
$$

i.e. regularity in x for every positive t even for initial data without regularity assumptions.

This idea can be combined with the approach of the previous section to obtain a similar result for the kinetic Fokker-Planck equation

$$
\partial_t f + v \cdot \nabla_x f = \nabla_v \cdot (\nabla_v f + vf)\,,
$$

with $x \in \mathbb{T}^d$, $v \in \mathbb{R}^d$. We introduce the new unknown

$$
h(x,v,t) := \frac{f(x,v,t)}{M(v)}\,,
$$

with the normalized Gaussian M, leading to the evolution equation

$$
\partial_t h + v \cdot \nabla_x h = \mathsf{C}h = \Delta_v h - v \cdot \nabla_v h\,.
$$

Remark 4. Since the approach used here and in the previous section relies on using the gradient with respect to v, it makes a difference to use h instead of f. Actually, in [34] a third option is chosen, namely the unknown $fM^{-1/2} = hM^{1/2}$, which makes the collision operator symmetric with respect to the L^2-norm without weight.

Mass conservation now reads

$$\int h(\cdot,\cdot,t)d\mu = \int h_0 d\mu\,, \qquad \text{with } d\mu := M\, dv\, dx\,,$$

where the right hand side is again assumed to vanish in the following. In this section the notation $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ are used for the scalar product and, respectively, norm in $L^2(d\mu)$. This implies the properties

$$\langle \mathsf{C}h, g\rangle = -\langle \nabla_v h, \nabla_v g\rangle\,, \qquad -\langle \mathsf{C}h, h\rangle \geq \|h\|^2\,,$$

of the collision operator, where the constant 1 in the Poincaré inequality is optimal. We start with computing time derivatives as in the previous section:

$$\frac{1}{2}\frac{d}{dt}\|h\|^2 = -\|\nabla_v h\|^2\,,$$

$$\frac{1}{2}\frac{d}{dt}\|\nabla_x h\|^2 = -\|\nabla_x \otimes \nabla_v h\|^2\,,$$

$$\frac{1}{2}\frac{d}{dt}\|\nabla_v h\|^2 = -\|\nabla_v^2 h\|^2 - \|\nabla_v h\|^2 - \langle \nabla_x h, \nabla_v h\rangle\,,$$

$$\frac{d}{dt}\langle \nabla_x h, \nabla_v h\rangle = -2\langle \nabla_x \otimes \nabla_v h, \nabla_v^2 h\rangle - \|\nabla_x h\|^2 - \langle \nabla_x h, \nabla_v h\rangle\,.$$

Now we define the functional

$$H[h(t),t] := \frac{1}{2}\left(\|h\|^2 + \alpha t^3\|\nabla_x h\|^2 + \beta t\|\nabla_v h\|^2 + 2\gamma t^2\langle \nabla_x h, \nabla_v h\rangle\right)\,,$$

with $\gamma^2 < \alpha\beta$, and estimate its time derivative

$$\frac{d}{dt}H \leq -\left(1 + \beta t - \frac{\beta}{2}\right)\|\nabla_v h\|^2 - t^2\left(\gamma - \frac{3\alpha}{2}\right)\|\nabla_x h\|^2$$
$$+ t(2\gamma - \beta - \gamma t)\langle \nabla_x h, \nabla_v h\rangle\,.$$

With the choice $\beta = 2\gamma$, this can be estimated further:

$$\frac{d}{dt}H \leq -(1-\gamma)\|\nabla_v h\|^2 - t^2\left(\gamma - \frac{3\alpha}{2}\right)\|\nabla_x h\|^2 - \gamma t^2\langle \nabla_x h, \nabla_v h\rangle\,.$$

With the choices $\alpha = 1/2$, $\beta = 7/4$, and $\gamma = 7/8$, the condition $\gamma^2 < \alpha\beta$ is satisfied and the right hand side is nonpositive for $t \leq 2/7$. Therefore we obtain the regularity estimates that for initial data $h_0 \in L^2(d\mu)$,

$$\|\nabla_x h\| = O(t^{-3/2})\,, \qquad \|\nabla_v h\| = O(t^{-1/2})\,, \qquad \text{as } t \to 0\,.$$

Note that the result for the v-derivatives is as for the heat equation, which is not surprising since the operator is elliptic in v.

12. 'Bakry-Émery meets Villani' (copyright F. Baudoin [6])

A subtitle of this section could be *hypocoercivity by the Γ-calculus*. Compared to [6] we keep computations simpler by replacing the potential $\Phi(x)$, confining in position space, by restriction of the dynamics to a torus, considering the same problem as in the preceding section, also using the unknown h, satisfying

$$\partial_t h = \Delta_v h - v \cdot \nabla_v h - v \cdot \nabla_x h =: \mathsf{L}h$$

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

$$(30) \quad \Gamma_2(f) = \frac{1}{2}\mathsf{L}\Gamma(f,f) - \Gamma(f,\mathsf{L}f) = |\nabla_v^2 f|^2 + |\nabla_v f|^2 + \nabla_v f \cdot \nabla_x f,$$

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

$$(31) \quad \Gamma_2(f) + \Gamma_2^Z(f) \geq 2|\nabla_x f + \nabla_v f|^2 = \frac{1}{4}\Gamma^Z(f) - \frac{1}{2}\Gamma(f) + \left|\nabla_x f + \frac{3}{2}\nabla_v f\right|^2.$$

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 $e^{-|v|^2/2}$: Actually we have

$$\Gamma(f) + \Gamma^Z(f) = 4|\nabla_x f|^2 + 2|\nabla_v f|^2 + 4\nabla_x f \cdot \nabla_v f \geq (3 - \sqrt{5})(|\nabla_x f|^2 + |\nabla_v f|^2).$$

On the other hand, with $\Pi f = \int f\, Mdv$,

$$\int \left(|\nabla_x f|^2 + |\nabla_v f|^2\right) d\mu \geq \int |\nabla_x \Pi f|^2 dx + \int f^2 d\mu - \int (\Pi f)^2 dx \geq \int f^2 d\mu,$$

showing that the Poincaré inequality holds with $\kappa = 3 - \sqrt{5}$. 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

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

using (31) and the maximum principle. Now we integrate and use mass conservation and the Poincaré inequality:

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

with $\lambda = \frac{2\kappa}{3+4\kappa} \approx 0.2523$ (for optimal ε). 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. The method and the result are in principle the same as in [34], [42]. However the clean computations lead to a very agreeable estimate of the spectral gap λ.

13. Sharp decay rates

Here we consider the kinetic Fokker-Planck equation on whole space with an harmonic oscillator force field:

$$(32) \qquad \partial_t f + v \cdot \nabla_x f - x \cdot \nabla_v f = \nabla_v \cdot (\nabla_v f + vf),$$

with $x, v \in \mathbb{R}^d$. The collision operator on the right hand side vanishes for Maxwellian velocity distributions, whereas the transport operator on the left hand side vanishes for functions of the total energy $|x|^2 + |v|^2$. Combining these observations leads to the equilibrium

$$f_\infty(x,v) = (2\pi)^{-d} \exp\left(-\frac{|x|^2 + |v|^2}{2}\right),$$

i.e. the normalized Gaussian in $\mathbb{R}^{2d}$. With $z = (x,v)$ the equation can be written in the form

$$(33) \qquad \partial_t f = \nabla_z \cdot (D\nabla_z f + Cz\, f),$$

with

$$D = \begin{pmatrix} 0 & 0 \\ 0 & I_d \end{pmatrix}, \quad C = \begin{pmatrix} 0 & -I_d \\ I_d & I_d \end{pmatrix},$$

a degenerate Fokker-Planck equation because of the degenerate diffusion matrix.

If we start with studying the characteristic ODEs from the first order part of the operator,

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = -C \begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} v \\ -x - v \end{pmatrix},$$

then we observe that there is a lack of coercivity,

$$\frac{1}{2}\frac{d}{dt}(|x|^2 + |v|^2) = -|v|^2,$$

although the eigenvalues

$$\lambda_\pm = \frac{1 \pm i\sqrt{3}}{2}$$

of C have positive real part $1/2$. With a modified norm hypocoercivity can be shown and even the sharp decay rate is obtained:

$$\frac{d}{dt}(|x|^2 + |v|^2 + x \cdot v) = -(|x|^2 + |v|^2 + x \cdot v).$$

Note that

$$|x|^2 + |v|^2 + x \cdot v = z^{tr} P z\,, \qquad \text{with} \quad P = \sum_{j=1}^{2d} p_j \otimes \overline{p_j}\,,$$

where $p_1, \ldots, p_{2d}$ are a complete set of eigenvectors of C^{tr}. This idea can be extended to cases, when C is not diagonalizable:

Lemma 5. *([2]) Let all eigenvalues of the matrix C have positive real parts with minimum μ.*
a) If all eigenvalues with real part μ are nondefective (i.e. algebraic multiplicity = geometric multiplicity), then there exists a positive definite symmetric matrix P, such that

$$C^{tr} P + P C \geq 2\mu P\,.$$

b) In the presence of defective eigenvalues with real part μ: For every $\varepsilon > 0$ there exists a positive definite symmetric matrix $P(\varepsilon)$, such that

$$C^{tr} P + P C \geq 2(\mu - \varepsilon) P\,.$$

This can be used for solutions of $\dot{z} = -Cz$, since

$$\frac{d}{dt}(z^{tr} P z) = -z^{tr}(C^{tr} P + P C)z \leq -2\mu\, z^{tr} P z\,.$$

Remark 5. *Also the result b) of the lemma is sharp in the sense that in this case the best possible decay to equilibrium is like $e^{-\mu t} t^k = O(e^{-(\mu-\varepsilon)t})$.*

Returning to PDEs, we follow [2] and consider general equations of the form (33) with positive semidefinite $D \in \mathbb{R}^{n \times n}$ and with $C \in \mathbb{R}^{n \times n}$ (not necessarily symmetric) having eigenvalues with positive real parts. The Fourier transform $\hat{f}(\xi, t)$ satisfies

$$\partial_t \hat{f} = -\xi^{tr} D\xi\, \hat{f} - \xi^{tr} C\, \nabla_\xi \hat{f}\,.$$

We look for an equilibrium f_∞ which is a generalized Gaussian such that $\hat{f}_\infty(\xi) = \exp(-\xi^{tr} K\xi/2)$ with symmetric K. This leads to the condition

$$\xi^{tr}(-D + CK)\xi = 0 \quad \forall\, \xi\,,$$

meaning that $CK - D$ is antisymmetric, and K has to satisfy the *continuous Lyapunov equation*

$$(34) \qquad\qquad\qquad 2D = CK + KC^{tr}\,,$$

which has a unique, symmetric, positive semidefinite solution [27]. For existence and uniqueness the assumptions on the spectrum of C are sufficient. Positive semi-definiteness is inherited from D. For an integrable equilibrium we need, however, $K > 0$.

Lemma 6. *([2]) Let $\Re(\sigma(C)) > 0$. Then the solution K of (34) is regular, iff there is no eigenvector of C^{tr} in the nullspace of D.*

Proof: Assume K is not regular, i.e. $\exists v \neq 0$ with $Kv = 0$. Then

$$2v^{tr} D v = v^{tr} C K v + v^{tr} K C^{tr} v = 2v^{tr} C K v = 0$$

and, thus, $Dv = 0$, implying $K C^{tr} v = 0$. Since C is regular, $C^{tr} v \neq 0$ is another element of the nullspace of K. Iterating the argument, we get that $\text{span}\{v, C^{tr} v, \ldots, (C^{tr})^{n-1} v\}$

is a C^{tr}-invariant subspace of the nullspace of D. In other words: The nullspace of D contains an eigenvector of C^{tr}.

On the other hand, assume $\exists\, v \neq 0$ with $C^{tr}v = \lambda v$ and $Dv = 0$. Then

$$0 = 2v^{tr}Dv = v^{tr}CKv + v^{tr}KC^{tr}v = 2\,\Re(\lambda)v^{tr}Kv\,,$$

Therefore K is nonregular since $\Re(\lambda) > 0$. $\square$

Remark 6. *The condition that the nullspace of D does not contain eigenvectors of C^{tr} is equivalent to the* Hörmander *condition for hypoellipticity of (33). It is needed for generalizing the computation of Kolmogorow* [28] *of a smooth fundamental solution (see* [2, 25]*).*

With the equilibrium $f_\infty(z) = C_K \exp(-z^{tr}K^{-1}z/2)$ and with $h = f/f_\infty$, we compute the dissipation of the quadratic relative entropy:

$$\frac{d}{dt}\int_{\mathbb{R}^n} \frac{(h-1)^2}{2}f_\infty dz = -\int_{\mathbb{R}^n} \nabla_z h^{tr} D \nabla_z h\, f_\infty\, dz\,.$$

The Bakry-Émery method is not feasible because of the lack of coercivity. The approaches of the last three sections can be seen as the search for a matrix P, such that the functional

$$S(h(t)) := \int_{\mathbb{R}^n} \nabla_z h^{tr} P \nabla_z h\, f_\infty\, dz$$

has good decay properties, or as a modification of the Bakry-Émery approach, where in the entropy dissipation the diffusion matrix D is replaced by P. A long computation [2] gives

$$\frac{d}{dt}S(h) \;=\; -\int_{\mathbb{R}^n} \nabla_z h^{tr}\left(KC^{tr}K^{-1}P + PK^{-1}CK\right)\nabla_z h\, f_\infty\, dz$$
$$-2\int \mathrm{Tr}\left(D\nabla_z^2 h P \nabla_z^2 h\right) f_\infty dz$$

The identity $\mathrm{Tr}(DAPA) = \mathrm{Tr}(\sqrt{D}A\sqrt{P}(\sqrt{D}A\sqrt{P})^{tr})$ (for symmetric A) shows that the last term is nonpositive. It remains to choose P. We use Lemma 5 with C replaced by KCK^{-1} (with the same spectrum as C), and with this choice of P we obtain

$$\frac{d}{dt}S(h) \leq -2\lambda S(h)$$

with $\lambda = \mu$ if all eigenvalues of C with real part μ are nondefective, and $\lambda = \mu - \varepsilon$ with arbitrarily small ε otherwise. Therefore $S(h)$ decays exponentially, implying exponential decay of the entropy dissipation, which is dominated by $S(h)$. The last step, to transfer the decay to the relative entropy, is as in the Bakry-Émery approach leading to

$$\int_{\mathbb{R}^n} \frac{(f - f_\infty)^2}{2f_\infty}dz \leq Ce^{-2\lambda t}\,, \qquad \text{for a } C > 0\,.$$

It is proved in [2] that the constant λ is sharp. This is seen either by the explicit example

$$f(z, 0) = (1 + z_0 \cdot z)f_\infty(z) \qquad \text{with } Cz_0 = \mu z_0\,,$$

with the exact decay rate or by a complete spectral analysis of the problem.

Finally we return to the kinetic Fokker-Planck equation (32). Our computations at the beginning of this section show that in this case K is the identity matrix. Therefore we can use the same P as for the characteristic ODEs and obtain

$$S(h) = \int_{\mathbb{R}^d} (|\nabla_x h|^2 + |\nabla_v h|^2 + \nabla_x h \cdot \nabla_v h) f_\infty \, dx \, dv \,,$$

which decays with the optimal rate e^{-t}.

14. Hypocoercivity without regularity

Here we shall apply the methodology of [16] to

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

$x \in \mathbb{T}^d$, $v \in \mathbb{R}^d$, $t > 0$, 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 [14]). For this simple problem, the method of [16] 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. We consider the abstract ODE

(36) $\partial_t f + \mathsf{T} f = \mathsf{C} f \,,$

in an Hilbert space $\mathcal{H}$ with a symmetric negative semidefinite ('collision') operator C and an antisymmetric ('transport') operator T, and 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:* $-\langle \mathsf{C} f, f \rangle \geq \lambda_m \|(1 - \Pi)f\|^2$, $\lambda_m > 0$, and
   H3. *Macroscopic coercivity:* $\|\mathsf{T}\Pi f\|^2 \geq \lambda_M \|\Pi f\|^2$, $\lambda_M > 0$.

We construct a modification of the relative entropy, i.e. the square of the norm, which is norm equivalent, and with the goal to make use of H3. The modified entropy is given by

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

with $\varepsilon > 0$.

Lemma 7. *([16]) The operators A and $\mathsf{T}\mathsf{A}$ are bounded by $1/2$ and, respectively, 1 in the operator norm.*

Proof: The equation $\mathsf{A}f = g$ is equivalent to $(\mathsf{T}\Pi)^*f = g + [1 + (\mathsf{T}\Pi)^*\mathsf{T}\Pi]g$. The scalar product with g gives

$$\|\mathsf{T}\Pi g\|^2 \leq \|g\|^2 + \|\mathsf{T}\Pi g\|^2 = \langle f, \mathsf{T}\Pi g\rangle \leq \|\mathsf{T}\Pi g\|^2 + \frac{\|f\|^2}{4}\,,$$

implying both results, noting that A maps to $\mathcal{N}(\mathsf{C})$ and, thus, $\mathsf{T}\Pi g = \mathsf{T}\mathsf{A}f$. $\square$

As a consequence, $\sqrt{H[f]}$ is an equivalent norm for $\varepsilon < 1$. The time derivative of the modified entropy is given by

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

The important additional term on the right hand side is the second. The operator $\mathsf{A}\mathsf{T}\Pi$ is the image of the nonnegative operator $(\mathsf{T}\Pi)^*\mathsf{T}\Pi$ under the map $z \mapsto \frac{z}{1+z}$. It therefore has the same spectral decomposition, and assumption H3 can be used to get

$$\langle \mathsf{A}\mathsf{T}\Pi f, f\rangle \geq \frac{\lambda_M}{1+\lambda_M}\|\Pi f\|^2\,.$$

It remains to estimate the remaining three terms. It is an important observation that

$$\langle \mathsf{T}\mathsf{A}f, f\rangle = \langle \mathsf{T}\mathsf{A}(1-\Pi)f, (1-\Pi)f\rangle\,, \qquad \langle \mathsf{A}\mathsf{C}f, f\rangle = \langle \mathsf{A}\mathsf{C}(1-\Pi)f, f\rangle\,,$$

where the second is obvious and the first a consequence of H1. Under the additional assumption of

H4. Boundedness of auxiliary operators: $\|\mathsf{A}\mathsf{T}(1-\Pi)f\| + \|\mathsf{A}\mathsf{C}f\| \leq C_M\|(1-\Pi)f\|$,

and with Lemma 7 the estimate

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

holds, implying (as for the model problem) that the dissipation of $H[f]$ is coercive for ε small enough. This implies the existence of constants $C, \lambda > 0$ such that

$$\|e^{(\mathsf{C}-\mathsf{T})t}\| \leq Ce^{-\lambda t}\,.$$

Explicit formulas for C and λ in terms of λ_m, λ_M, and C_M can be found in [9].

The kinetic Fokker-Planck equation. With the operators T and C from (35) and $\mathcal{H} = L^2(d\sigma)$ the orthogonal projection to $\mathcal{N}(\mathsf{C})$ is given by

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

The marginal density ρ_f is called the macroscopic or position density. We consider solutions with $\int_{\mathbb{T}^d} \rho_f(x,t)dx = 0$.

With $\mathsf{T}\Pi f = vM \cdot \nabla_x\rho_f$, Assumption H1 is a consequence of the fact that the flux associated to local equilibria of the form $\rho(x)M(v)$ vanishes: $\int vM\,dv = 0$. 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 (35) 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

$$\int_{\mathbb{R}^d} |\nabla_v h|^2\, M\, dv \geq \lambda_m \int_{\mathbb{R}^d} \left(h - \int hM\, dv\right)^2 M\, dv\,,$$

($f = hM$) which holds for the Gaussian weight M. Similarly the macroscopic coercivity assumption H3 is equivalent to the Poincaré inequality on the torus,

$$\|\mathsf{T}\Pi f\|^2 = \int_{\mathbb{T}^d} |\nabla_x \rho_f|^2 dx \geq \lambda_M \int_{\mathbb{T}^d} \rho_f^2\, dx\,, \qquad \text{for } \int_{\mathbb{T}^d} \rho_f\, dx = 0\,.$$

The action of the operator A can be described as follows: $\mathsf{A}f = \rho_{\mathsf{A}f}M$, where $\rho_{\mathsf{A}f}$ is the unique solution of

$$\Delta_x \rho_{\mathsf{A}f} - \rho_{\mathsf{A}f} = \nabla_x \cdot \int_{\mathbb{R}^d} vf\, dv\,.$$

Instead of directly proving the boundedness of AT, we consider the adjoint

$$(\mathsf{AT})^* = -\mathsf{T}^2\Pi(1 + (\mathsf{T}\Pi)^*\mathsf{T}\Pi)^{-1}\,.$$

Similarly to above we have that $(\mathsf{AT})^*f = -\mathsf{T}^2(\rho M)$ with ρ satisfying

$$\Delta_x \rho - \rho = -\rho_f\,,$$

and $\|\mathsf{T}^2(\rho M)\|^2 = 3\|\nabla_x^2 \rho\|_{L^2(dx)}^2$. So the boundedness of $(\mathsf{AT})^*$, and thus of AT, is a consequence of the $(L^2 \to H^2)$-regularization of the elliptic equation on the torus: $\|\rho\|_{H^2(dx)} \leq c\|\rho_f\|_{L^2(dx)}$.

Finally, the boundedness of AC is a consequence of the preliminary computation $\rho_{\mathsf{TC}f} = -\nabla_x \cdot \int vf\, dv$, showing that $\mathsf{AC} = -\mathsf{A}$. Thus, H1-H4 holds with the consequence that solutions of (35) 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 the previous sections with the methods of [6, 34, 42].

Finally it should be noted that the method of [16] 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.

15. Hypocoercivity and fast-reaction limit for a kinetic system with chemical reactions

The hypocoercivity approaches presented in the preceding sections are rather robust. In this section we consider a model for a simple chemical reaction network where the transport of the species is described by a kinetic model, and we follow

[36]. This is the first nonlinear problem considered here, and it can be seen as a transition to the following sections.

The chemical process is simple pair generation and recombination. We consider a background of particles with vanishing average velocity. These particles can split into pairs of particles of two different species A and B, and the reverse reaction is also possible, where one particle of species A and one particle of species B recombine and become a background particle. We assume the concentrations of A and B to be much smaller than the background concentration, which is therefore assumed given and stationary. When a pair of particles is created, their velocities are randomly sampled from a Gaussian velocity distribution with mean zero and fixed (background) temperature. The recombination rate is assumed independent of the velocities of the recombining particles. This leads to the model

$$
\begin{aligned}
\partial_t f + v \cdot \nabla_x f &= M(v) - \rho_g f \,, \\
\partial_t g + v \cdot \nabla_x g &= M(v) - \rho_f g \,,
\end{aligned}
\tag{37}
$$

where f, g are the phase space densities of the species A, B, and the definitions of the normalized Maxwellian M and of the position densities ρ_f, ρ_g are as above. This is based on a nondimensionalization, where the reference velocity is the thermal velocity corresponding to the background temperature. The first terms on the right hand sides describe the generation of particle pairs, and the last terms their recombination.

Like in previous examples we choose the domains $x \in \mathsf{T}^d$, $v \in \mathbb{R}^d$, and initial conditions

$$
f(x, v, 0) = f_I(x, v) \,, \qquad g(x, v, 0) = g_I(x, v) \,.
\tag{38}
$$

There is one conserved quantity, the difference of the numbers of particles of species A and B:

$$
\int_{\mathsf{T}^d \times \mathbb{R}^d} (f(x, v, t) - g(x, v, t)) dx \, dv = \int_{\mathsf{T}^d \times \mathbb{R}^d} (f_I(x, v) - g_I(x, v)) dx \, dv \,, \qquad t \geq 0 \,.
$$

As $t \to \infty$, we expect convergence of (f, g) to $(f_\infty, g_\infty) = (\rho_\infty M, \frac{1}{\rho_\infty} M)$ where ρ_∞ is the unique positive solution of

$$
|\mathsf{T}^d| \left(\rho_\infty - \frac{1}{\rho_\infty} \right) = \int_{\mathsf{T}^d \times \mathbb{R}^d} (f_I(x, v) - g_I(x, v)) dx \, dv \,.
$$

The relative entropy functional

$$
\mathsf{H}(f, g) = \int_{\mathbb{T}^3} \int_{\mathbb{R}^3} \left[f \ln \frac{f}{f_\infty} - f + f_\infty + g \ln \frac{g}{g_\infty} - g + g_\infty \right] dv \, dx \,,
\tag{39}
$$

decreases as long as (f, g) is different from $(\rho(x) M(v), \frac{1}{\rho(x)} M(v))$ for some $\rho(x)$:

$$
\frac{d}{dt} \mathsf{H}(f, g) = \int_{\mathbb{T}^3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} (MM' - fg') \ln \left(\frac{fg'}{MM'} \right) dv' \, dv \, dx \leq 0 \,,
\tag{40}
$$

showing that we are again confronted with the problem of a lack of coercivity of the entropy dissipation. The question of finding suitable entropy functionals will be the subject of the following sections.

Solvability of the problem under suitable assumptions on the initial data is a consequence of a maximum principle:

Theorem 8. *([36]) Assume there exist positive constants $\gamma_1 < \rho_\infty$ and γ_2 such that*

$$(\rho_\infty - \gamma_1)M \le f_I \le (\rho_\infty + \gamma_2)M \quad and \quad \frac{1}{\rho_\infty + \gamma_2}M \le g_I \le \frac{1}{\rho_\infty - \gamma_1}M\,.$$

Then the initial value problem (37), (38) *has a unique global mild solution* $(f, g) \in C([0, \infty), L^\infty(dx\,dv/M))^2$ *satisfying, for all* $(x, v, t) \in \mathsf{T}^d \times \mathbb{R}^d \times [0, \infty)$,

$$(\rho_\infty - \gamma_1)M(v) \le f(x, v, t) \le (\rho_\infty + \gamma_2)M(v)\,, \,,$$

and

$$\frac{1}{\rho_\infty + \gamma_2}M(v) \le g(x, v, t) \le \frac{1}{\rho_\infty - \gamma_1}M(v)\,.$$

Note that this not only an existence and uniqueness result, but also shows neutral stability of the equilibrium state in $L^\infty(dx\,dv/M)$.

Hypocoercivity. We start by showing linearized asymptotic stability, i.e. we prove convergence to zero of solutions of the linearized system

(41)
$$\partial_t f + v \cdot \nabla_x f = -\rho_\infty M \rho_g - \frac{1}{\rho_\infty}f\,,$$

$$\partial_t g + v \cdot \nabla_x g = -\frac{1}{\rho_\infty}M\rho_f - \rho_\infty g\,,$$

satisfying

(42)
$$\int_{\mathsf{T}^d \times \mathbb{R}^d} (f(x, v, t) - g(x, v, t))dx\,dv = 0\,, \qquad t \ge 0\,.$$

Note that after linearization we are not interested in positive solutions anymore, and (41) does not describe a Markov process. However, introducing $-g$ as a new variable, nonnegativity of f and $-g$ and their total mass would be preserved. Anyway, a suitable functional analytic setting can be deduced from 'quadratization' (i.e. quadratic approximation close to the minimizer) of (39):

$$\mathsf{H}(f_\infty + f, g_\infty + g) \approx \|f\|^2_{L^2(dx\,dv/f_\infty)} + \|g\|^2_{L^2(dx\,dv/g_\infty)} =: \|(f, g)\|^2\,.$$

This defines the Hilbert space $\mathcal{H}$ in the abstract formulation of the problem with

$$\mathsf{T}(f, g) = \begin{pmatrix} v \cdot \nabla_x f \\ v \cdot \nabla_x g \end{pmatrix}, \qquad \mathsf{C}(f, g) = \begin{pmatrix} -\rho_\infty M \rho_g - \frac{1}{\rho_\infty}f \\ -\frac{1}{\rho_\infty}M\rho_f - \rho_\infty g \end{pmatrix}.$$

The orthogonal projection to $\mathcal{N}(\mathsf{C})$ is given by

$$\Pi(f, g) = \frac{\rho_f - \rho_g}{\rho_\infty^2 + 1}\begin{pmatrix} \rho_\infty^2 \\ -1 \end{pmatrix}M$$

Now we compute

$$-\langle \mathsf{C}(f, g), (f, g)\rangle = \frac{1}{\rho_\infty}\|f - \rho_f M\|^2_{L^2(dx\,dv/f_\infty)} + \rho_\infty\|g - \rho_g M\|^2_{L^2(dx\,dv/g_\infty)}$$

$$+ \frac{1}{\rho_\infty^2}\int_{\mathsf{T}^d}(\rho_f + \rho_\infty^2 \rho_g)^2 dx$$

and

$$\|(1 - \Pi)(f, g)\|^2 = \|f - \rho_f M\|^2_{L^2(dx\,dv/f_\infty)} + \|g - \rho_g M\|^2_{L^2(dx\,dv/g_\infty)}$$

$$+ \frac{1}{\rho_\infty(1 + \rho_\infty^2)}\int_{\mathsf{T}^d}(\rho_f + \rho_\infty^2 \rho_g)^2 dx$$

proving *microscopic coercivity*,

$$\langle \mathsf{C}(f,g),(f,g) \rangle \le -\lambda_m \|(1-\Pi)(f,g)\|^2, \qquad \text{with} \quad \lambda_m = \min\left\{\rho_\infty, \frac{1}{\rho_\infty}\right\}.$$

For proving *macroscopic coercivity*, we need

$$\|\Pi(f,g)\|^2 = \frac{\rho_\infty}{1+\rho_\infty^2}\|\rho_f - \rho_g\|^2_{L^2(dx)}, \quad \|\mathsf{T}\Pi(f,g)\|^2 = \frac{\rho_\infty}{1+\rho_\infty^2}\|\nabla_x(\rho_f - \rho_g)\|^2_{L^2(dx)},$$

showing macroscopic coercivity with $\lambda_M = 1$, using $\int_{\mathsf{T}^d}(\rho_f - \rho_g)dx = 0$ and the optimal constant 1 in the Poincaré inequality on the torus. The conditions $\Pi\mathsf{T}\Pi = 0$ and the boundedness of AT are shown as in the previous section for the kinetic Fokker-Planck equation. Finally, the boundedness of AC is a consequence of the obvious boundedness of C. Thus Assumptions H1–H4 of Section 14 are satisfied, showing that solutions of the linearized system (41) satisfying (42) converge to zero exponentially in $\mathcal{H}$.

In view of the maximum principle estimates from the existence theorem, this result can be extended to a local asymptotic stability result for the nonlinear problem.

Theorem 9. *([36]) With the assumptions of Theorem 8 with γ_1, γ_2 small enough, the solution (f,g) of (37), (38) converges to the steady state (f_∞, g_∞) as $t \to \infty$, exponentially in $\mathcal{H}$ (defined above).*

For the proof, the modified entropy for the linear problem is used. The difference between the nonlinear reaction operator and its linearization can be bound by $\gamma\|(f - f_\infty, g - g_\infty)\|$, where γ is small if γ_1, γ_2 are. Therefore the nonlinear perturbation term can be dominated by the entropy dissipation.

In [36], also a second proof of Theorem 9 is given, based on the methods presented in Section 10. The result is analogous except H^1-regularity assumptions on the initial data. Since decay in Sobolev spaces with higher differentiability order can also be proven, the maximum principle estimates can be replaced by Sobolev imbedding for the control of the nonlinearity.

Fast-reaction limit. Entropy estimates are THE basic tool for a rigorous derivation of macroscopic limits of kinetic models, with many examples in the literature, culminating in the derivation of the incompressible Navier-Stokes equations from the Boltzmann equation [21].

In the system (37) we introduce a *macroscopic length scale*, which is by a large factor $\frac{1}{\varepsilon}$ bigger than the *mean free path* between reaction events. The dimensionless parameter ε is called the *Knudsen number*. Since the equilibrium velocity distribution M is unbiased, i.e. it has mean velocity zero, we expect a diffusive macroscopic limit and choose a parabolic rescaling of time by $\frac{1}{\varepsilon^2}$. This leads to the rescaled system

$$(43) \qquad \begin{aligned} \varepsilon^2 \partial_t f + \varepsilon v \cdot \nabla_x f &= M(v) - \rho_g f, \\ \varepsilon^2 \partial_t g + \varepsilon v \cdot \nabla_x g &= M(v) - \rho_f g. \end{aligned}$$

The first step is to use the entropy dissipation (40) and to integrate with respect to time

$$(44) \qquad \varepsilon^2 \mathsf{H}(f(T), g(T)) + \int_0^T \int_{\mathbb{T}^3 \times \mathbb{R}^6} (fg' - MM') \ln \left(\frac{fg'}{MM'} \right) dv' \, dv \, dx \, dt$$
$$= \varepsilon^2 \mathsf{H}(f_I, g_I) \, .$$

Now some useful information has to be extracted from this. Heuristically, smallness of the entropy dissipation means that its quadratic approximation $-\langle \mathsf{C}(f,g), (f,g) \rangle$ controlling $\|(1-\Pi)(f,g)\|^2$ can be used. Rigorously, one has to work a little harder, using the inequality $(\sqrt{a} - 1)^2 \le \frac{1}{4}(a-1) \ln a$.

Lemma 10. *([36]) Let $\mathsf{H}(f_I, g_I) < \infty$. Then the entropy dissipation relation (44) implies $(1 - \Pi)(f,g) = O(\varepsilon)$ in $L^2(\mathbb{R}_+, \mathcal{H})$.*

As a consequence there exists $\rho(x, t)$, still depending on ε but uniformly bounded by the maximum principle estimates, such that

$$f^\perp := \frac{f - \rho M}{\varepsilon} \quad \text{and} \quad g^\perp := \frac{g - M/\rho}{\varepsilon}$$

are uniformly bounded in $L^2(\mathbb{R}_+, \mathcal{H})$. Therefore, after division by ε, the system (43) reads

$$(45) \qquad \begin{aligned} \varepsilon \partial_t f + v \cdot \nabla_x f &= \frac{M(v) - \rho_g f}{\varepsilon} = -\rho_{g^\perp} \rho \, M - \frac{1}{\rho} f^\perp + O(\varepsilon) \, , \\ \varepsilon \partial_t g + v \cdot \nabla_x g &= \frac{M(v) - \rho_f g}{\varepsilon} = -\rho_{f^\perp} \frac{1}{\rho} M - \rho \, g^\perp + O(\varepsilon) \, . \end{aligned}$$

The uniform boundedness of the right hand sides permits the application of an averaging lemma to obtain some smoothness in x of the position densities of f and g. With the help of an interpolation result similar to the Aubin-Lions lemma eventually strong convergence of ρ can be shown [36]. The final limiting equation will be the conservation law

$$\partial_t(\rho_f - \rho_g) + \nabla_x \cdot \int_{\mathbb{R}^d} v(f^\perp - g^\perp) dv = 0 \, .$$

The flux can be computed from the limit of (45):

$$\int_{\mathbb{R}^d} v(f^\perp - g^\perp) dv = -\rho \nabla_x \rho + \frac{1}{\rho} \nabla_x \left(\frac{1}{\rho} \right) = -\left(\rho + \frac{1}{\rho^3} \right) \nabla_x \rho \, .$$

The limiting density ρ satisfies

$$\partial_t \left(\rho - \frac{1}{\rho} \right) = \nabla_x \cdot \left(\left(\rho + \frac{1}{\rho^3} \right) \nabla_x \rho \right) \, ,$$

which can be written as a nonlinear diffusion equation for $u = \rho - 1/\rho$. A rigorous convergence result has been proven in [36].

16. Entropies for nonlinear problems – 1. Inverting the equilibrium distribution

Consider initial value problems for nonlinear abstract ODEs,

$$(46) \qquad \qquad \partial_t p = F(p) \, , \qquad p(0) = p_0 \, ,$$

where $p_0 \in L^1_+(M, d\mu)$ for some measure space $(M, d\mu)$. We assume that the evolution preserves total mass and nonnegativity, such that $p(t) \in L^1_+(M, d\mu)$ and

$$\langle p(t) \rangle := \int p(t) d\mu = m := \langle p_0 \rangle, \qquad t \geq 0.$$

It will be useful to permit arbitrary positive m (which is of course not necessarily mass in applications), and not only probability distributions. We shall assume the existence of a one-parameter family $\{p_\infty(m) : m \geq 0\} \subset L^1_+(M, d\mu)$ of equilibria with the total mass as parameter:

$$F(p_\infty(m)) = 0, \quad \langle p_\infty(m) \rangle = m, \qquad m \geq 0.$$

Furthermore we assume that $p_\infty(m)$ is a pointwise strictly increasing function of m:

(47) $$m \mapsto p_\infty(m, x) \quad \text{strictly increasing } \forall\, x \in M.$$

In the linear case, where $F = L^*$ generates the evolution of the law of an homogeneous Markov process, we have

$$p_\infty(m, x) = m\, p_\infty(1, x),$$

and the condition (47) is satisfied, whenever the probability density $p_\infty(1)$ is strictly positive, as has been assumed in the previous sections. The dissipation of relative entropy (Section 7) relies on the inequality

$$\frac{d}{dt} H_U(p|p_\infty(1)) = \left\langle U'\left(\frac{p}{p_\infty(1)}\right) L^* p \right\rangle \leq 0$$

for arbitrary convex U. We attempt to generalize this situation by interpreting the ratio $\frac{p(x)}{p_\infty(1,x)}$ as the unique solution m of the equation $p(x) = p_\infty(m, x)$. Therefore the following recipe is suggested: First, find $\phi(p, x)$ by inversion of the equilibrium distribution with respect to m:

$$m = \phi(p, x) \quad \Longleftrightarrow \quad p = p_\infty(m, x), \qquad p > 0,\, x \in M.$$

Second, check if

$$\langle \chi(\phi(p))\, F(p) \rangle \leq 0.$$

for any increasing χ. If yes, the relative entropy

$$H_U(p|\{p_\infty\}) := \langle \Phi(p) \rangle, \quad \text{with } \Phi(p, x) = \int_{p_\infty(m,x)}^{p} \chi(\phi(q, x)) dx$$

is nonincreasing along solutions of (46). The notation for the relative entropy should be a reminder that its definition requires the whole family of equilibria.

The question is of course: Does this ever work? The answer is yes for large groups of examples, but not always.

Example 13. *An ODE system:* We start with $M = \{1, 2\}$ and the system

$$\dot{p}_1 = p_2 - p_1^2, \qquad \dot{p}_2 = p_1^2 - p_2,$$

with $p_1, p_2 \geq 0$ and $p_1 + p_2 = m$ conserved. We compute

$$p_{\infty,1}(m) = \frac{1}{2}\left(-1 + \sqrt{1 + 4m}\right), \qquad p_{\infty,2}(m) = m + \frac{1}{2}\left(1 - \sqrt{1 + 4m}\right),$$

and

$$\phi_1 = p_1 + p_1^2, \qquad \phi_2 = \sqrt{p_2} + p_2,$$

and we obtain

$$\langle \chi\left(\phi(p)\right) F(p)\rangle = -\left(\chi(p_1 + p_1^2) - \chi(\sqrt{p_2} + p_2)\right)(p_1^2 - p_2) \leq 0\,,$$

for any monotone χ. With $\chi = $ id we obtain the entropy

$$\frac{p_1^2}{2} + \frac{p_1^3}{3} + \frac{2p_2^{3/2}}{3} + \frac{p_2^2}{2}\,.$$

Example 14. *Nonlinear parabolic conservation laws,*

$$\partial_t p + \nabla_x \cdot f(p) = \nabla_x \cdot (D(p)\nabla_x p)\,,$$

with $M = \mathbb{R}^d$, given nonlinear flux function $f(p)$, and positive semidefinite diffusivity matrix $D(p)$, have constant solutions. This suggests to use $\phi = p$ and to compute

$$\langle \chi(p)F(p)\rangle = -\langle \chi'(p)\nabla_x p^{tr} D(p)\nabla_x p\rangle \leq 0\,,$$

for $\chi' \geq 0$. For degenerate D, smooth solutions exist in general only for finite time, and the continuation by weak solutions is nonunique. In this situation, entropy decay can be used as part of the definition of solutions, reestablishing uniqueness (see, e.g. [10]).

Example 15. *The Pauli exclusion principle in kinetic models:* We consider a model for an ensemble of fermions, e.g. electrons, moving in a stationary background with constant temperature. The Pauli exclusion principle for fermions means for a semiclassical kinetic description that there is an upper bound, here 1, for the phase space density (see, e.g. [31]). An appropriate kinetic model has the form

$$\partial_t f + v \cdot \nabla_x f = Q(f) := \int_{\mathbb{R}^3} (f'(1-f)M - f(1-f')M')dv'\,,$$

with the normalized Maxwellian velocity distribution $M(v)$. A family of equilibria is given by the *Fermi-Dirac distributions*

$$(48) \qquad\qquad f_\infty(\mu, v) = \frac{1}{1 + e^{|v|^2/2 - \mu}}\,,$$

with the *Fermi energy* $\mu \in \mathbb{R}$. Expressing e^μ from this formula suggests trying

$$\begin{aligned}
\left\langle \chi\left(\frac{f}{M(1-f)}\right) F(f)\right\rangle &= -\int MM'(1-f)(1-f')\left(\frac{f}{M(1-f)} - \frac{f'}{M'(1-f')}\right) \\
&\times \left(\chi\left(\frac{f}{M(1-f)}\right) - \chi\left(\frac{f'}{M'(1-f')}\right)\right) dv'\, dv\, dx \leq 0\,.
\end{aligned}$$

Again any increasing χ is possible. In [39] the macroscopic limit has been carried out, and $\chi(z) = \frac{-1}{1+z}$ was convenient. Another interesting choice is $\chi = \ln$, leading to the entropy

$$\int \left(\frac{|v|^2}{2}f + f(\ln f - 1) + (1-f)(\ln(1-f) - 1)\right) dv\, dx\,,$$

which would be called the *free energy* in thermodynamics.

In [7] a constant electric field described by the vector $E \in \mathbb{R}^3$ has been added:

$$\partial_t f + v \cdot \nabla_x f + E \cdot \nabla_v f = Q(f)\,.$$

The main idea still works. However, serious work is required to show the existence of a family of equilibria $f_\infty(m)$, satisfying $E \cdot \nabla_v f_\infty = Q(f_\infty)$ and having the required strict monotonicity property with respect to m. The entropy estimate

is instrumental for carrying out the macroscopic limit, in this case a nonlinear hyperbolic conservation law.

Example 16. The *reaction-kinetic system* from Section 15: We recall that for this problem we had

$$f_\infty(x,v) = \rho_\infty M(v), \qquad g_\infty(x,v) = \frac{1}{\rho_\infty} M(v),$$

meaning we can take

$$\phi(f,g,v) = \left(\frac{f}{M(v)}, \frac{M(v)}{g} \right).$$

The fact that we do not have total mass conservation but conservation of the difference of the masses shows in the fact that the second component is a decreasing function of g. If we multiply the right hand sides of (37) by $\chi(f/M)$ and, respectively, $\chi(M/g)$, take the difference and integrate, we obtain

$$\int_{\mathsf{T}^d \times \mathbb{R}^{2d}} \left((MM' - fg')\chi\left(\frac{f}{M}\right) - (MM' - f'g)\chi\left(\frac{M}{g}\right) \right) dv'\, dv\, dx$$

$$= \int_{\mathsf{T}^d \times \mathbb{R}^{2d}} MM' \left(1 - \frac{fg'}{MM'} \right) \left(\chi\left(\frac{f}{M}\right) - \chi\left(\frac{M'}{g'}\right) \right) dv'\, dv\, dx \leq 0.$$

With $\chi = \mathrm{id}$ and with $\chi = \log$, we obtain the entropy densities

$$h(f,g) = \frac{f^2}{2M} - M \log g,$$

and, respectively,

$$h(f,g) = f \log(f/M) - f + g \log(g/M) - g,$$

the second being the choice in Section 15.

Example 17. *BGK models:* In [15] the macroscopic limit of models of the form

$$\partial_t f + v \cdot \nabla_x f - \nabla_x V(x) \cdot \nabla_v f = \gamma\left(\frac{|v|^2}{2} - \mu(\rho_f) \right) - f,$$

has been carried out for various choices of the equilibrium profile γ. The Fermi energy $\mu(\rho)$ is determined by the assumption of mass conservation:

$$\int_{\mathbb{R}^d} \gamma\left(\frac{|v|^2}{2} - \mu(\rho) \right) dv = \rho.$$

Under the assumption that γ is decreasing with sufficient decay, $\mu(\rho)$ is uniquely determined and strictly increasing. As an entropy density, the free energy

$$\int^f \phi(g,v)dg = \int^f \left(\frac{|v|^2}{2} - \gamma^{-1}(g) \right) dg = \frac{|v|^2}{2} f - \int^f \gamma^{-1}(g)dg,$$

has been used in [15].

17. Entropies for nonlinear problems – 2. The logarithmic entropy for mass action kinetics

Chemical reaction networks with mass action kinetics always dissipate the logarithmic relative entropy. This result will be proven in this section, following the fundamental work [26] on this subject.

Chemical reaction networks are described by lists with entries of the form

$$\mathcal{C}_j := \sum_{l=1}^{m} y_j^l \mathcal{A}^l \quad \xrightarrow{k_{ij}} \quad \sum_{l=1}^{m} y_i^l \mathcal{A}^l =: \mathcal{C}_i \,,$$
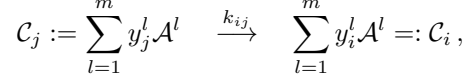
where $\mathcal{A}^1, \ldots, \mathcal{A}^m$ are *chemical species*, types of molecules, and $y_j^l \in \mathbb{N}_0$ are *stoichiometric coefficients*, combined into *complex vectors* $y_j = (y_j^1, \ldots, y_j^m)$, i.e. m-dimensional multi-indices, describing the *complexes* $\mathcal{C}_1, \ldots, \mathcal{C}_n$. The stoichiometric coefficients encode the information, how many molecules of species $\mathcal{A}^l$ are consumed (y_j^l) and produced (y_i^l) in the reaction $\mathcal{C}_j \to \mathcal{C}_i$.

The goal is to derive a mathematical model for the evolution of the *species concentrations* $c(t) = (c^1(t), \ldots, c^m(t))$ in a well stirred reactor. It will be based on the assumption of *mass action kinetics*, which means that the *reaction rate* r_{ij}, i.e. the number of reaction events $\mathcal{C}_j \to \mathcal{C}_i$ per time is given by the model

$$r_{ij}(c) = k_{ij} c^{y_j} := k_{ij} \prod_{l=1}^{m} (c^l)^{y_j^l} \,, \qquad i, j = 1, \ldots, n \,.$$

with the standard notation for vector$^{\text{multi-index}}$. The constants $k_{ij} \geq 0$ are called *rate constants*, and we use the convention $k_{ij} = 0$, if $\mathcal{C}_j \to \mathcal{C}_i$ does not appear in the list of reactions, as well as $k_{ii} = 0$. We also introduce the rate of formation of complex $\mathcal{C}_i$ by

$$g_i(c) = \sum_{j=1}^{n} (r_{ij}(c) - r_{ji}(c)) \,,$$

and the *species formation vector*

$$f(c) = \sum_{i,j=1}^{n} r_{ij}(c)(y_i - y_j) \,.$$

The latter can also be written as $f = Yg$ with the *stoichiometric matrix* $Y = (y_1, \ldots, y_n) \in \mathbb{R}^{m \times n}$ and the *complex formation vector* $g = (g_1, \ldots, g_n)$. The desired model for the dynamics is the ODE system

(49) $\dot{c} = f(c) \,,$

which is completely determined by the stoichiometric matrix Y and by the rate constant matrix $K := (k_{ij}) \in \mathbb{R}^{n \times n}$, whence we call $\langle Y, K \rangle$ a *mass action system*. The mass action kinetics implies that the admissible set $\overline{\mathbb{R}_+^m}$ is invariant under the flow. The *order of the reaction* $\mathcal{C}_j \to \mathcal{C}_i$ is given by $|y_j|$, and it is obvious that for a system of only first order reactions the complexes can be identified with the species and (49) is linear.

We introduce the *stoichiometric space*

$$S := \text{span}\{y_i - y_j \, : \, k_{ij} > 0\} \subset \mathbb{R}^m \,.$$

It is obvious that $f(c) \in S$ and that for the initial conditions $c(0) = c_0 \in \overline{\mathbb{R}^m_+}$ we have that $c(t)$, $t \geq 0$, lies in the *reaction simplex* $(c_0 + S) \cap \overline{\mathbb{R}^m_+}$. If $S \neq \mathbb{R}^m$ then vectors $0 \neq \phi \in S^\perp$ provide *conservation laws*

$$\frac{d}{dt} \phi \cdot c = 0 \,.$$

In particular, we say that the system is *conservative*, if there exists such a ϕ with all entries positive, i.e. if $S^\perp \cap \mathbb{R}^m_+ \neq \{\}$. In this case all reaction simplices are bounded. An example of this situation occurs, when all reactions conserve mass, i.e.

$$M \cdot y_i = M \cdot y_j \,, \quad \forall k_{ij} > 0 \,,$$

where $M = (M^1, \ldots, M^m) \in \mathbb{R}^m_+$ is the vector of molecular masses of the species. Then the reaction simplex is a subset of $\{c \in \overline{\mathbb{R}^m_+} : M \cdot c = M \cdot c_0\}$.

The form

$$f(c) = Y g(c) = Y(R(c) - R(c)^{tr})1$$

of the species formation vector (with $R = (r_{ij}) \in \mathbb{R}^{n \times n}$ and $1 = (1, \ldots, 1) \in \mathbb{R}^n$) suggests to distinguish between different kinds of *equilibria*, and we shall only be interested in equilibria in $\mathbb{R}^m_+$.

Therefore the *equilibrium set* is defined by

$$E := \{c \in \mathbb{R}^m_+ : f(c) = 0\} \,.$$

Obviously $c \in E$ iff $g(c) \in \ker Y$. The subset of *complex balanced equilibria* is given by

$$C := \{c \in \mathbb{R}^m_+ : g(c) = 0\} \,,$$

and the even smaller subset of *detailed balanced equilibria* by

$$D := \{c \in \mathbb{R}^m_+ : R(c) = R(c)^{tr}\} \,.$$

There are 4 different situations for the equilibrium set $E = \mathbb{R}^m_+ \cap g^{-1}(\ker Y)$, depicted in Fig. 1:
a) $g(\mathbb{R}^m_+) \cap \ker Y = \{\}$, which implies $E = \{\}$.
b) $g(\mathbb{R}^m_+) \cap \ker Y = \{0\}$, which implies $C = E \neq \{\}$.
c) $0 \notin g(\mathbb{R}^m_+) \cap \ker Y \neq \{\}$, which implies $E \neq \{\}$, but $C = \{\}$.
d) $g(\mathbb{R}^m_+) \cap \ker Y$ contains 0 and at least one other point, which implies $\{\} \neq C \neq E$.

Examples:
a) $\mathcal{A}_1 \to \mathcal{A}_2$,
b) $\mathcal{A}_1 \to \mathcal{A}_2 \to \mathcal{A}_3 \to \mathcal{A}_1$,
c) $\mathcal{A}_1 \to \mathcal{A}_2$, $2\mathcal{A}_2 \to 2\mathcal{A}_1$.

It will be shown below that case d) cannot occur for mass action kinetics. In other words: If a mass action kinetic network has one complex balanced equilibrium, then we are in case b) and all equilibria are complex balanced.

Definition 2. *A J-tuple of different complexes $\mathcal{C}_{j_1}, \ldots, \mathcal{C}_{j_P}$ is called a <u>reaction cycle of length J</u>, iff $k_{j_1 j_2}, \ldots, k_{j_{P-1} j_P}, k_{j_P j_1} > 0$.*

Definition 3. *A mass action system $\langle Y, K \rangle$ is called <u>cyclic</u>, if it only consists of one reaction cycle, i.e. $k_{j,j+1} > 0$ for $j = 1, \ldots, n$ with indices understood modulo n, and all other $k_{ij} = 0$.*
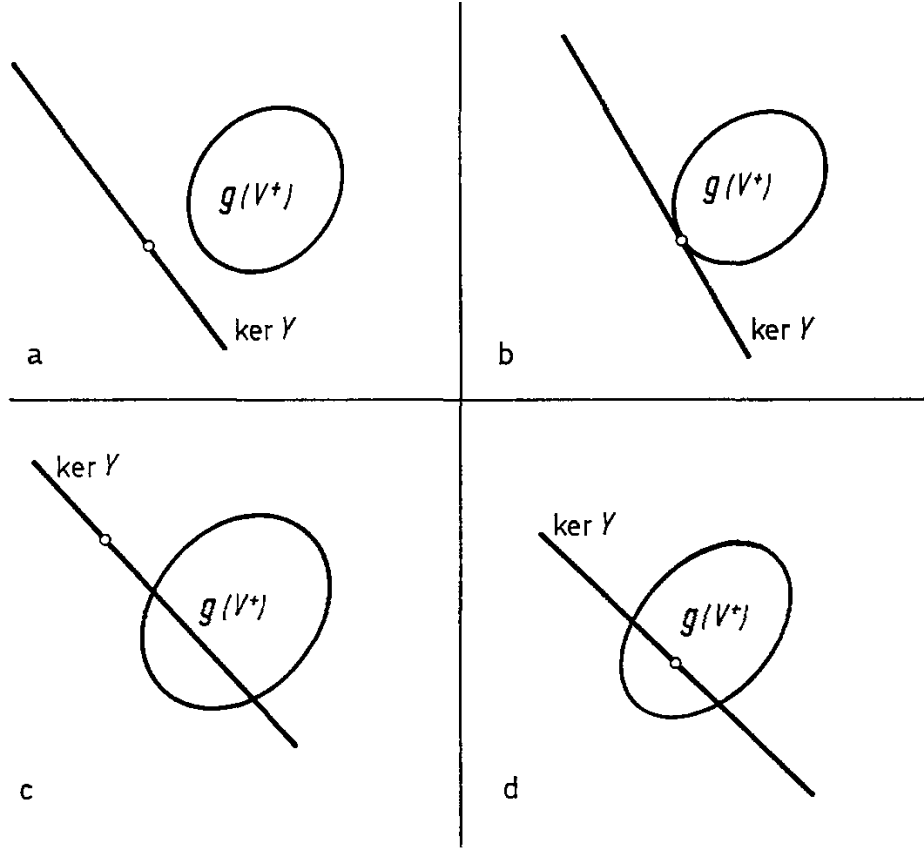
FIGURE 1. From [26]. Possible relations between $\ker Y$ and $g\left(\mathbb{R}^m_+\right)$ (with $V^+ = \mathbb{R}^m_+$). The $\circ$ on $\ker Y$ indicates the origin of $\mathbb{R}^n$.

For a cyclic mass action system we have

$$f(c) = \sum_{j=1}^{n} k_{j+1,j} \, c^{y_j} \left(y_{j+1} - y_j\right), \qquad g(c)_j = k_{j,j-1} c^{y_{j-1}} - k_{j+1,j} c^{y_j} \,.$$

If there exists a complex balanced equilibrium a, then

(50) $$\kappa = k_{j+1,j} a^{y_j} > 0\,, \qquad j = 1, \dots, n\,.$$

Therefore

$$f(c) = \kappa \sum_{j=1}^{n} \left(\frac{c}{a}\right)^{y_j} \left(y_{j+1} - y_j\right),$$

where the division c/a of vectors is meant componentwise, just as $\log c$ used below. We compute

$$\log\left(\frac{c}{a}\right) \cdot f(c) = \kappa \sum_{j=1}^{n} \left(\frac{c}{a}\right)^{y_j} \left(\log\left(\frac{c}{a}\right)^{y_{j+1}} - \log\left(\frac{c}{a}\right)^{y_j}\right)$$

Lemma 11. *Let $\phi : D(\phi) \subset \mathbb{R} \to \mathbb{R}$ strictly increasing, let $\xi_1, \ldots, \xi_n \in D(\phi)$, $n \geq 2$, and set $\xi_{n+1} := \xi_1$. Then*

$$\Phi_n(\xi_1, \ldots, \xi_n) := \sum_{j=1}^{n} \xi_j(\phi(\xi_{j+1}) - \phi(\xi_j)) \leq 0 \,,$$

with equality, iff $\xi_1 = \cdots = \xi_n$.

Proof. By induction with respect to n: For $n = 2$, obviously

$$\Phi_2(\xi_1, \xi_2) = (\xi_1 - \xi_2)(\phi(\xi_2) - \phi(\xi_1)) \leq 0 \,,$$

with equality, iff $\xi_1 = \xi_2$.

Now let the statement hold for some n and let $\xi_1, \ldots, \xi_{n+1} \in D(\phi)$ with (w.l.o.g.) $\xi_{n+1} = \max\{\xi_1, \ldots, \xi_{n+1}\}$. Then

$$\Phi_{n+1}(\xi_1, \ldots, \xi_{n+1}) = \Phi_n(\xi_1, \ldots, \xi_n) + (\xi_{n+1} - \xi_n)(\phi(\xi_1) - \phi(\xi_{n+1})) \,.$$

By the induction hypothesis and by $\xi_{n+1} \geq \xi_1, \xi_n$, this is the sum of two nonpositive terms, which completes the proof. $\square$

Corollary 12. *If a cyclic mass action system has a complex balanced equilibrium a, then the relative entropy (called the pseudo-Helmholtz function in [26])*

$$H[c|a] := \sum_{j=1}^{n} \left(c_j \log\left(\frac{c_j}{a_j}\right) - c_j + a_j \right) \,,$$

generated by $U(z) = z \log z - z + 1$, is a Lyapunov functional, i.e. for solutions $c(t)$ of (49) we have

$$\frac{d}{dt} H[c|a] = \log\left(\frac{c}{a}\right) \cdot f(c) \leq 0 \,,$$

with equality iff c is a complex balanced equilibrium, which is true for every equilibrium.

Proof. By Lemma 11, $\frac{d}{dt} H[c|a] = 0$ iff

$$\left(\frac{c}{a}\right)^{y_j} = \text{const} \,,$$

holding, by (50), iff

$$k_{j+1,j} c^{y_j} = \text{const} \,,$$

which is equivalent to c being a complex balanced equilibrium. Since obviously $\frac{d}{dt} H[c|a] = 0$ for every equilibrium, every equilibrium is complex balanced. $\square$

This result will be used by showing that every mass action system can be decomposed into cycles.

Lemma 13. *Every mass action system $\langle Y, K \rangle$ with $K \neq 0$ and with a complex balanced equilibrium contains a reaction cycle.*

Proof. Let a be a complex balanced equilibrium. Then

$$a^{y_i} \sum_{j=1}^{n} k_{ji} = \sum_{j=1}^{n} k_{ij} a^{y_j} \,, \qquad \forall i = 1, \ldots, n \,.$$

Since there exists at least one $k_{i_1,j} > 0$, the right hand side of this equation for $i = i_1$ is positive. Now the equation implies the existence of $k_{i_2,i_1} > 0$. Repeating

the argument with $i = i_2$ leads to $k_{i_2,i_3} > 0$ and, eventually to an arbitrarily long chain. Since there is only a finite number of complexes, it has to contain a reaction cycle. $\square$

Theorem 14. *Let $\langle Y, K \rangle$ be a mass action system with $K \neq 0$ and with a complex balanced equilibrium a. Then*

$$K = \sum_{r=1}^{R} K^r \,,$$

where $\langle Y, K^r \rangle$ is cyclic with complex balanced equilibrium a, $1 \leq r \leq R$.

Proof. By Lemma 13, $\langle Y, K \rangle$ contains a cycle $\mathcal{C}_{i_1}, \ldots, \mathcal{C}_{i_P}$. Define

$$\kappa^1 := \min_{p=1,\ldots,P} \{ k_{i_{p+1},i_p} a^{y_{i_p}} \}$$

and

$$k^1_{i_{p+1},i_p} := \frac{\kappa^1}{a^{y_{i_p}}} \leq k_{i_{p+1},i_p} \,, \qquad p = 1, \ldots, P \,,$$
$$k^1_{ij} = 0 \,, \qquad \text{for all other } (i,j) \,.$$

Then $\langle Y, K^1 \rangle$ is cyclic with complex balanced equilibrium a, $K^1 \leq K$, $K - K^1$ has at least one positive entry less than K, and $\langle Y, K - K_1 \rangle$ still has the complex balanced equilibrium a. Iterating this process leads to

$$K - \sum_{r=1}^{R} K^r = 0 \,,$$

after $R \leq n^2$ steps. $\square$

Corollary 15. *The result of Corollary 12 holds for every mass action system.*

The result is of course also true for detailed balanced equilibria a, where

$$R(a) = R(a)^{tr} \quad \Leftrightarrow \quad \kappa_{ij} := k_{ij} a^{y_j} = k_{ji} a^{y_i} \,, \qquad \forall \, (i,j) \,,$$

but there is a simpler proof:

$$
\begin{aligned}
\log\left(\frac{c}{a}\right) \cdot f(c) &= \sum_{l=1}^{m} \sum_{i,j=1}^{n} \log\left(\frac{c^l}{a^l}\right) y_i^l (k_{ij} c^{y_j} - k_{ji} c^{y_i}) \\
&= \sum_{l=1}^{m} \sum_{i,j=1}^{n} \kappa_{ij} \left(\left(\frac{c}{a}\right)^{y_j} - \left(\frac{c}{a}\right)^{y_i} \right) \log\left(\frac{c}{a}\right)^{y_i} \\
(51) \qquad &= -\frac{1}{2} \sum_{l=1}^{m} \sum_{i,j=1}^{n} \kappa_{ij} \left(\left(\frac{c}{a}\right)^{y_j} - \left(\frac{c}{a}\right)^{y_i} \right) \left(\log\left(\frac{c}{a}\right)^{y_j} - \log\left(\frac{c}{a}\right)^{y_i} \right) \leq 0 \,.
\end{aligned}
$$

At this point we stop the presentation of the theory of mass action kinetics, but not without mentioning that there are results on the existence of complex balanced equilibria and on their uniqueness within a reaction simplex. This requires additional conditions on the reaction network (see, e.g., [18]).

Example 18. *Extension to continuous species spaces:* As in Example 15 we consider fermions. However we do not model their interaction with a background, but we

consider binary collisions. The corresponding semiclassical model is the *quantum Boltzmann* or *Uehling-Uhlenbeck equation* (see, e.g., [8])

$$(52) \qquad \partial_t f + v \cdot \nabla_x f$$

$$= \int_{\mathbb{R}^3 \times \mathbb{S}^2} B(\omega, v - v_*)(f' f'_*(1-f)(1-f_*) - f f_*(1-f')(1-f'_*)) d\omega \, dv_*$$

where $f(x, v, t)$ is the phase space distribution function. On the right hand side f, f_*, f', f'_* mean evaluation at v, v_*, v', and, respectively, v'_*, where the post-collisional velocities v', v'_* are given in terms of the pre-collisional velocities v, v_* by

$$v' = v - ((v - v_*) \cdot \omega)\omega , \qquad v'_* = v_* + ((v - v_*) \cdot \omega)\omega ,$$

which can be seen as a parametrization of the set of all quadruples (v, v_*, v', v'_*) satisfying momentum and energy conservation:

$$v + v_* = v' + v'_* , \qquad |v|^2 + |v_*|^2 = |v'|^2 + |v'_*|^2 .$$

On the other hand, in the setting of colliding spheres, the interpretation of ω is the direction between the centers of the spheres. More generally, the *collision cross section B* describes the physical nature of the interaction. It has the pre-post-collision invariance property $B(\omega, v' - v'_*) = B(\omega, v - v_*)$.

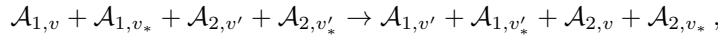It is easily seen that Fermi-Dirac distributions (48) with constant Fermi energy are equilibria of (52).

We can write (52) as a system of two equations for f and $g = 1 - f$:

$$\partial_t f + v \cdot \nabla_x f = \int_{\mathbb{R}^3 \times \mathbb{S}^2} B(\omega, v - v_*)(f' f'_* g g_* - f f_* g' g'_*) d\omega \, dv_* ,$$

$$\partial_t g + v \cdot \nabla_x g = \int_{\mathbb{R}^3 \times \mathbb{S}^2} B(\omega, v - v_*)(f f_* g' g'_* - f' f'_* g g_*) d\omega \, dv_* .$$

Note that the system propagates the property $f + g = 1$, and that for each value of $\mu \in \mathbb{R}$ it has the positive equilibrium

$$f_\infty = \frac{1}{1 + e^{|v|^2/2 - \mu}} , \qquad g_\infty = \frac{e^{|v|^2/2 - \mu}}{1 + e^{|v|^2/2 - \mu}} .$$

The terms on the right hand sides of the system for f and g can be seen as the model for a fourth order chemical reaction

$$\mathcal{A}_{1,v} + \mathcal{A}_{1,v_*} + \mathcal{A}_{2,v'} + \mathcal{A}_{2,v'_*} \to \mathcal{A}_{1,v'} + \mathcal{A}_{1,v'_*} + \mathcal{A}_{2,v} + \mathcal{A}_{2,v_*} ,$$

where species $\mathcal{A}_{1,v}$ corresponds to a particle with velocity v, and $\mathcal{A}_{2,v}$ to available space at velocity v. Since not only the integrals, but the integrands vanish at (f_∞, g_∞), the equilibrium is detailed balanced. The transport terms on the left hand side are generators of a Markov process and, thus, we expect the logarithmic relative entropy

$$\int \left(f \log \left(\frac{f}{f_\infty} \right) - f + f_\infty + g \log \left(\frac{g}{g_\infty} \right) - g + g_\infty \right) dv \, dx$$

$$= \int \left(f \log \left(\frac{f}{f_\infty} \right) + (1 - f) \log \left(\frac{1 - f}{1 - f_\infty} \right) \right) dv \, dx ,$$

to be nonincreasing along solutions, which can be easily checked as in (51).

References

[1] F. Achleitner, A. Arnold, D. Stuerzer, Large-time behavior in non-symmetric Fokker-Planck equations, *Riv. Mat. Univ. Parma* **6** (2015), pp. 1–68.

[2] A. Arnold, J. Erb, Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift, arXiv:1409.5425.

[3] 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.

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

[5] D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Springer, 2013.

[6] F. Baudoin, Bakry-Emery meet Villani, *arXiv:* 1308.4938v1, 2013.

[7] 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.

[8] D. Benedetto, F. Castella, R. Esposito, M. Pulvirenti, A short review on the derivation of the nonlinear quantum Boltzmann equations, *Comm. in Math. Sci.* **5** (2007), pp. 55–71.

[9] E. Bouin, J. Dolbeault, S. Mischler, C. Mouhot, C. Schmeiser, Hypocoercivity without confinement, arXiv:1708.06180.

[10] C. Dafermos, *Hyperbolic Conservation Laws in Continuum Physics*, Springer-Verlag, 1999.

[11] P. Degond, T. Goudon, F. Poupaud, Diffusion limit for nonhomogeneous and non-micro-reversible processes, *Indiana Univ. Math. J.* **49** (2000), pp. 1175–1198.

[12] A. Dembo, *Stochastic Processes*, lecture notes, Stanford Univ., 2013.

[13] 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.

[14] 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.

[15] J. Dolbeault, P. Markowich, D. Oelz, C. Schmeiser, Nonlinear diffusion as limit of kinetic equations with relaxation collision kernels, *Archive Rat. Mech. Anal.* **186** (2007), pp. 133–158.

[16] J. Dolbeault, C. Mouhot, C. Schmeiser, Hypocoercivity for linear kinetic equations conserving mass, *Trans. AMS* **367** (2015), pp. 3807–3828.

[17] 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.

[18] M. Feinberg, *Lectures on Chemical Reaction Networks,* MRC, Madison, 1980, https://crnt.osu.edu/LecturesOnReactionNetworks.

[19] 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.

[20] I. Gentil, C. Imbert, The Lévy-Fokker-Planck equation: Φ-entropies and convergence to equilibrium, *Asymptotic Anal.* **59** (2008), pp. 125–138.

[21] F. Golse, L. Saint-Raymond, The incompressible Navier-Stokes limit of the Boltzmann equation for hard cutoff potentials, *J. Math. Pures Appl.* **91** (2009), pp. 508–552.

[22] U.G. Haussmann, E. Pardoux, Time reversal of diffusions, *Ann. of Prob.* **14** (1986), pp. 1188–1205.

[23] 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.

[24] M. Herda, L.M. Rodrigues, Large-time behavior of solutions to Vlasov-Poisson-Fokker-Planck equations: from evanescent collisions to diffusive limit, *J. Stat. Phys.* **170** (2018), pp. 895–931.

[25] L. Hörmander, Hypoelliptic second order differential equations, *Acta Math.* **119** (1969), pp. 147–171.

[26] F. Horn, R. Jackson, General mass action kinetics, *Archive Rat. Mech. Anal.* **47** (1972), pp. 81–116.

[27] R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis,* Cambridge University Press (1991).

[28] A. Kolmogoroff, Zufällige Bewegungen (zur Theorie der Brownschen Bewegung). *Ann. of Math.* **35** (1934), 116–117.

[29] P. Laurencot, B. Perthame, Exponential decay for the growth-fragmentation/cell-division equation, *Comm. Math. Sci.* **7** (2009), pp. 503–510.

[30] 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.

[31] P.A. Markowich, C.A. Ringhofer, C. Schmeiser, *Semiconductor Equations,* Springer-Verlag, Vienna, 1990.

[32] 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.

[33] S. Mischler, B. Perthame, L. Ryzhik, Stability in a Nonlinear Population Maturation Model, *Math. Mod. Meth. Appl. Sci.* **12** (2002), pp. 1751–1772.

[34] C. Mouhot, L. Neumann, Quantitative perturbative study of convergence to equilibrium for collisional kinetic models in the torus, *Nonlinearity*, **19** (2006), pp. 969–998.

[35] C. Mouhot, E. Russ, Y. Sire, Fractional Poincaré inequalities for general measures, *J. Math. Pures Appl.* **95** (2011), pp. 72–84.

[36] L. Neumann, C. Schmeiser, A kinetic reaction model: decay to equilibrium and macroscopic limit, *Kinetic and Related Models* **9** (2016), pp. 571–585.

[37] B. Perthame, *Transport Equations in Biology,* Birkhäuser, 2007.

[38] F. Poupaud, Runaway phenomena and fluid approximation under high fields in semiconductor kinetic theory, *ZAMM* **72** (1992), pp. 359–372.

[39] F. Poupaud, C. Schmeiser, Charge transport in semiconductors with degeneracy effects, *Math. Meth. in the Appl. Sci.* **14** (1991), pp. 301–318.

[40] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry,* North-Holland, 1992.

[41] R.S. Varga, *Matrix Iterative Analysis,* 2nd ed., Springer-Verlag, 2002.

[42] 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