Stanford PDE mini-course, Spring 2020

Lenya Ryzhik*

May 11, 2020

1 Examples of relaxation enhancement

In these notes, we will discuss convergence to an equilibrium for solutions of partial differential equations. In this introductory part, we will focus on the situation when the equilibrium is simply $\phi(t,x) \equiv 0$, so the interest is in the decay rate of the solution to zero, and what can make this decay to be faster than naively expected. Generally, we will look at evolution problems of the form

$$\frac{\partial \phi}{\partial t} = L\phi - \Gamma\phi, \quad t \geq 0, \tag{1.1}$$

with an initial condition $\phi(0) = \phi_0 \in H$, where H is some Hilbert space, L is a skew-symmetric operator, and Γ is a symmetric operator on H. Typically, we will take $H = L^2(\mathbb{R}^d)$ but sometimes we will think of more abstract settings. We will usually assume that Γ is strictly dissipative (coercive) there exists $c_0 > 0$ so that

$$\langle \Gamma\phi, \phi\rangle \geq c_0\|\phi\|^2. \tag{1.2}$$

It is immediate to see from (1.2) that the solution to (1.1) satisfies the exponential decay estimate:

$$\frac{1}{2}\frac{d}{dt}(\|\phi(t)\|^2) = -\langle \Gamma\phi, \phi\rangle \leq -c_0\|\phi\|^2, \tag{1.3}$$

so that

$$\|\phi(t)\| \leq \|\phi_0\| \exp(-c_0 t). \tag{1.4}$$

This estimate does not depend on L at all, as long as L is skew-symmetric, and holds, in particular, if $L = 0$.

On the other hand, since L is skew-symmetric, solutions to the Γ-less equation

$$\frac{\partial \psi}{\partial t} = L\psi, \quad \psi(0) = \psi_0, \tag{1.5}$$

do not decay at all:

$$\frac{1}{2}\frac{d}{dt}(\|\psi(t)\|^2) = \langle L\psi, \psi\rangle = 0, \tag{1.6}$$

*Department of Mathematics, Stanford University, Stanford CA, 94305, USA; ryzhik@math.stanford.edu

so that

$$\|\psi(t)\| = \|\psi_0\|. \tag{1.7}$$

The question we are interested in is if the presence of L can make the decay of the solutions in (1.1) much faster than the trivial bound (1.4) even though solutions to the "purely L"-equation (1.6) have no decay whatsoever. This phenomenon is informally known as relaxation enhancement.

The basic mechanism behind relaxation enhancement is very simple. Let us assume that the operator Γ is self-adjoint and has a discrete spectrum with eigenvalues

$$0 < \lambda_0 \le \lambda_1 \le \dots \le \lambda_n \le \dots,$$

with

$$\lambda_n \to +\infty \text{ as } n \to +\infty, \tag{1.8}$$

with the corresponding eigenspaces

$$X_k = \{\phi \in H : \ \Gamma\phi = \lambda_k \phi\}.$$

Then solutions to the "L-less" equation

$$\frac{\partial \phi}{\partial t} = -\Gamma\phi, \tag{1.9}$$

with an initial condition $\phi_0 \in X_k$ will decay as

$$\|\phi(t)\| = \|\phi_0\| \exp(-\lambda_k t). \tag{1.10}$$

Therefore, the decay estimate

$$\|\phi(t)\| \le \|\phi_0\| \exp(-\lambda_0 t) \tag{1.11}$$

is optimal, as ϕ_0 generically has a non-trivial component in the eigenspace X_0.

Consider now what may happen if $L \ne 0$, and to emphasize the effect, let us put a large coefficient in front of L:

$$\frac{\partial \phi}{\partial t} = \frac{1}{\varepsilon} L\phi - \Gamma\phi, \tag{1.12}$$

with $\phi_0 \in X_k$, and with $\varepsilon \ll 1$. Then, for short times, solutions to (1.12) can be well approximated by

$$\frac{\partial \phi_a}{\partial t} = \frac{1}{\varepsilon} L\phi_a, \tag{1.13}$$

with $\phi_a(0) = \phi_0$. This is equivalent to a time-rescaling of

$$\frac{\partial \psi}{\partial t} = L\psi, \quad \psi(0) = \phi_0, \tag{1.14}$$

in the sense that $\phi_a(t) = \psi(t/\varepsilon)$. If L and Γ do not commute, then there is no reason why $\phi_a(t)$ would have a large component in the eigenspaces X_k corresponding to "small" k. We expect that $\phi_a(t)$ will populate all eigenspaces of Γ. Moreover, $\phi_a(t)$ should not concentrate in the low eigenspaces of Γ, and the bulk of the solution will be in the high eigenmodes of Γ. But for such $\phi_a(t)$ the operator Γ may dominate $\varepsilon^{-1}L$ in (1.12) if, say, $\lambda_k \gg \varepsilon^{-1}$ and solutions to the full problem (1.12) should decay rapidly because it has been moved into the high eigenmodes of Γ by the approximate evolution (1.14) that only involves L. This is the basic scenario behind relaxation enhancement. Turning this into a theorem is not always simple but the mechanism is almost always the same.

A newborn toy example

Let us see how this works on the example of 2×2 matrices. We take

$$\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{1.15}$$

with $\lambda \gg 1$, so that $L\Gamma \neq \Gamma L$, and Γ has a small eigenvalue 1 and a large eigenvalue λ. The evolution by Γ alone:

$$\frac{d\phi}{dt} = -\Gamma\phi, \tag{1.16}$$

only has the a priori decay

$$\|\phi(t)\| \leq \|\phi(0)\|e^{-t}, \tag{1.17}$$

unless the initial condition is of the form $\phi(0) = (0, \alpha_0)$, so that the decay for the L-less equation (1.16) is governed by the small eigenvalue 1. On the other hand, the eigenvalues of the matrix

$$nL - \Gamma = \begin{pmatrix} -1 & n \\ -n & -\lambda \end{pmatrix}$$

are the solutions to

$$(\mu + 1)(\mu + \lambda) + n^2 = 0,$$

$$\mu_{1,2} = \frac{-1 - \lambda \pm \sqrt{(1+\lambda)^2 - 4(n^2 + \lambda)}}{2}. \tag{1.18}$$

Then, for $n \gg \lambda$ we have

$$\mathrm{Re}(\mu_{1,2}) = -\frac{1+\lambda}{2}. \tag{1.19}$$

In other words, if n is sufficiently large, then both eigenvalues $\mu_{1,2}$ of the matrix $nL - \Gamma$ have a very large negative real part and the "small" eigenvalue 1 of the matrix Γ disappears. Therefore, solutions to

$$\frac{d\phi}{dt} = (nL - \Gamma)\phi, \quad \phi(0) = \phi_0, \tag{1.20}$$

obey the decay estimate

$$\|\phi(t)\| \leq C_0 e^{-(1+\lambda)t/2}, \tag{1.21}$$

that is much better than (1.17) as $\lambda \gg 1$.

Let us look in a bit more detail at the solution to (1.20). Let $z(t)$ be the solution to

$$\frac{dz}{dt} = Lz, \quad z(0) = \phi_0 = (\phi_{10}, \phi_{20}), \tag{1.22}$$

so that

$$z(t) = \phi_{10} \begin{pmatrix} \cos t \\ -\sin t \end{pmatrix} + \phi_{20} \begin{pmatrix} \sin t \\ \cos t \end{pmatrix}, \tag{1.23}$$

and write

$$\phi(t) = a(t)z(nt) + \frac{1}{n}\psi(t),$$

with a scalar function $a(t)$ and a vector-valued function $\psi(t)$ to be chosen. This gives

$$\frac{da}{dt}z(nt) + \frac{1}{n}\frac{d\psi}{dt} = L\psi(t) - a(t)\Gamma z(nt) - \frac{1}{n}\Gamma\psi(t). \tag{1.24}$$

The function $z(t)$ is periodic with a period 2π and $Lz(t)$ has mean zero because of (1.22). In addition, we have $\|z(t)\| = \|\phi_0\|$ for all $t \geq 0$. Taking the inner product with $z(nt)$, averaging over the period $T = 2\pi$ and recklessly dropping the terms we expect to be small, we arrive at

$$\frac{da}{dt} = -\gamma a(t), \tag{1.25}$$

with

$$\gamma = \frac{1}{T\|\phi_0\|^2}\int_0^T \langle \Gamma z(t), z(t)\rangle dt = \frac{1}{T\|\phi_0\|^2}\int_0^T (z_1^2(t) + \lambda z_2^2(t))dt = \frac{1+\lambda}{2}. \tag{1.26}$$

We used the explicit formula (1.23) above. This explains the precise expression $\gamma = (+1\lambda)/2$ for the limit of the eigenvalues.

Exercise 1.1 The above argument on the behavior of the solutions to (1.20) is a simple case of the general averaging theory for ODE with periodic oscillatory coefficients. Make it rigorous.

Note that one key feature of this very simple dynamics is that the solution to the "pure L" equation (1.22) spends a sufficient time in the eigenspace of the largest eigenvalue of Γ to allow for the decay to kick in. This is very important for the "combined Γ and L "decay mechanism.

The Dirichlet eigenvalues for Laplacian with a drift

We now investigate the same phenomenon for the Laplacian operator with a strong incompressible drift.

The eigenvalues of the Laplacian

Before we explain how relaxation enhancement for the Laplacian operator with a drift comes about, let us first recall some very basic facts about the principal Dirichlet eigenvalues for the Laplacian on a bounded domain [11]. For any smooth bounded domain Ω there exists an eigenvalue λ_1 (called the principal eigenvalue) that corresponds to a positive eigenfunction $\phi_1 > 0$ in Ω:

$$\begin{aligned} -\Delta\phi_1 &= \lambda_1\phi_1, \quad x \in \Omega, \\ \phi_1 &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{1.27}$$

Moreover, λ_1 is the smallest of all eigenvalues of the Dirichlet Laplacian on Ω, λ_1 is a simple eigenvalue and all other eigenfunctions of the Laplacian change sign in Ω. For example, if Ω is an interval $(0,1)$, the eigenvalues of the operator $Lu = -u''$ with the Dirichlet boundary conditions $u(0) = u(1) = 0$ are $\lambda_n = n^2\pi^2$, and the corresponding eigenfunctions are

$$u_n(x) = \sin(n\pi x).$$

In this case, the principal eigenvalue is $\lambda_1 = \pi^2$.

In general, the principal eigenvalue of the Laplacian is given by the variational formula:

$$\lambda_1 = \inf_{\substack{\psi \in H_0^1(\Omega) \\ \|\psi\|_2 = 1}} \int_\Omega |\nabla \psi|^2 dx. \tag{1.28}$$

The principal eigenvalue determines the long time decay of solutions of the parabolic initial value problem in the following way. Consider the initial value problem

$$\begin{aligned} &\psi_t = \Delta \psi, \quad t > 0, x \in \Omega, \\ &\psi(t, x) = 0 \text{ on } \partial\Omega, \\ &\psi(0, x) = \psi_0(x). \end{aligned} \tag{1.29}$$

As $\phi_1(x) > 0$ in Ω, and, as follows from the Hopf lemma, $\partial\phi_1/\partial\nu < 0$ on $\partial\Omega$, we can find a constant $C > 0$ so that $|\psi(t = 1, x)| \leq C\phi_1(x)$ – we can not quite have such estimate at $t = 0$ since the initial condition $\psi_0(x)$ may not satisfy the Dirichlet boundary conditions. The maximum principle implies that

$$\psi(t, x) \leq Ce^{-\lambda_1(t-1)}\phi_1(x), \tag{1.30}$$

for $t > 1$, and, similarly,

$$-\psi(t, x) \leq Ce^{-\lambda_1(t-1)}\phi_1(x), \tag{1.31}$$

so that

$$|\psi(t, x)| \leq Ce^{-\lambda_1(t-1)}\phi_1(x), \quad t \geq 1. \tag{1.32}$$

Therefore, all solutions of the Cauchy problem decay at the exponential rate determined by λ_1 as $t \to +\infty$.

The Dirichlet eigenvalues with a drift

Let us now consider the Dirichlet principal eigenvalue problem in a smooth bounded domain Ω, for a diffusion with a strong incompressible flow:

$$\begin{aligned} &-\Delta\phi + \frac{1}{\varepsilon} u \cdot \nabla\phi = \lambda_1(\varepsilon)\phi, \quad \phi(x) > 0 \text{ in } \Omega, \\ &\phi = 0 \text{ on } \partial\Omega. \end{aligned} \tag{1.33}$$

This is an example of a problem like (1.12), with $\Gamma = -\Delta$, and $L = u \cdot \nabla$. We assume that u is an incompressible flow: $\nabla \cdot u = 0$, and that it does not penetrate the boundary:

$$u \cdot \nu = 0 \text{ on } \partial\Omega. \tag{1.34}$$

This makes the operator $L = u \cdot \nabla$ skew-symmetric:

$$\langle L\psi, \psi \rangle = \int_\Omega (u \cdot \nabla\psi)\psi dx = 0, \tag{1.35}$$

by the divergence theorem, as the boundary term vanishes, due to (1.34).

The operator in (1.33) is not self-adjoint (so that its eigenvalues are not necessarily real), and its eigenvalues do not obey an integral variational principle such as (1.28). Nevertheless, the Krein-Rutman theory for positive operators (see Chapter VIII of [8]) implies that it has a unique eigenvalue $\lambda_1(\varepsilon)$ that corresponds to a positive eigenfunction $\phi_1(x)$. This eigenvalue is real and simple, has the smallest real part of all eigenvalues, and is called the principal eigenvalue. As for the Laplacian, the maximum principle implies that the principal eigenvalue determines the long time decay of the solutions of the corresponding Cauchy problem:

$$\begin{aligned} &\psi_t + \frac{1}{\varepsilon} u \cdot \nabla \psi = \Delta \psi, \quad t > 0, x \in \Omega, \\ &\psi(t,x) = 0 \text{ on } \partial\Omega, \\ &\psi(0,x) = \psi_0(x), \end{aligned} \tag{1.36}$$

that is,

$$\psi(t,x) \sim e^{-\lambda_1(\varepsilon)t} \phi_1(x), \quad \text{as } t \to +\infty. \tag{1.37}$$

Note that when $u = 0$ (or, in our general terminology, $L = 0$) the exponential rate of decay for the solutions of (1.36) is simply the principal eigenvalue of the Laplacian. On the other hand, solutions of the Laplacian-less problem

$$\psi_t + \frac{1}{\varepsilon} u \cdot \nabla \psi = 0 \tag{1.38}$$

do not decay at all – their L^2 norm is preserved, as are all L^p-norms for $p \geq 1$. This is because the flow u is incompressible and parallel to $\partial\Omega$ on the boundary.

Let us now understand whether it is possible that solutions of the "combined" Cauchy problem (1.36) decay much faster in time than when $u = 0$ despite the fact that solutions of (1.38) have no decay whatsoever. To quantify this questions, let us ask if it is possible that

$$\lambda_1(\varepsilon) \to +\infty \text{ as } \varepsilon \to 0. \tag{1.39}$$

The above considerations make it clear that such phenomenon may only come from an interaction of the drift and the Laplacian.

Let us recall the probabilistic interpretation of the solutions of the Cauchy problem (1.36). Consider the stochastic differential equation

$$dX_t = -\frac{1}{\varepsilon} u(X_t) dt + \sqrt{2} dW_t, \quad X_0 = x, \tag{1.40}$$

starting at a point $x \in \Omega$, and let τ be the first time that the process X_t hits the boundary $\partial\Omega$. Then solution of the Cauchy problem (1.36) can be expressed in terms of the diffusion X_t as

$$\psi(t,x) = \mathbb{E}_x[g(X_{\min(t,\tau)})], \tag{1.41}$$

with the convention that

$$g(X_\tau) = 0. \tag{1.42}$$

When would we expect $\psi(t,x)$ to be small as $\varepsilon \to 0$? As one sees from (1.42), this would be true if, with a high probability we have $\tau < t$ – the particle hits the boundary before a given time t. Intuitively, if the trajectories of the incompressible flow are "sufficiently mixing", then, for any starting point x_0 in the interior of Ω, the trajectory of (1.40) that starts at x_0 eventually comes close to the boundary $\partial\Omega$. Therefore, such flow, when sufficiently fast, will force solutions of (1.41) very quickly to pass very close to $\partial\Omega$, and at that time diffusion term in (1.40) will force X_t to exit Ω with a very high probability. Hence, when $\varepsilon > 0$ is sufficiently small, the exit time τ of the solutions of (1.40) should be smaller than a given time $t > 0$ with a high probability. As we have mentioned, this makes $\psi(t, x)$ given by (1.41) very small because of (1.42). Physically, this means that a sufficiently mixing flow, together with diffusion, should dramatically increase the cooling of the interior by the boundary. A natural questions is what "mixing" means in this context, and how one can quantify such property. Usually, the mixing properties of a flow are defined in terms of the dynamic properties of the ODE

$$\dot{X} = u(X),$$

behave. Here, we are asking a PDE question – hence, the first problem is to define what "mixing" means for us. This is quantified by the following beautiful result due to Berestycki, Hamel and Nadirashvili [3]. We denote by $\mathcal{I}_0$ the set of all first integrals of u, solutions of

$$u \cdot \nabla\phi = 0 \text{ a.e. in } \Omega, \tag{1.43}$$

in the space $H_0^1(\Omega)$.

Theorem 1.2 *The principal eigenvalue $\lambda_1(\varepsilon)$ of (1.33) tends to $+\infty$ as $\varepsilon \to 0$ if and only if the flow u has no first integral in $H_0^1(\Omega)$. Moreover, if u has a first integral in $H_0^1(\Omega)$, then*

$$\lambda_1(\varepsilon) \to \bar{\lambda} := \min_{w \in \mathcal{I}_0} \frac{\int_\Omega |\nabla w|^2 dx}{\int_\Omega |w|^2 dx} \text{ as } \varepsilon \to 0, \tag{1.44}$$

and the minimum in the right side is achieved.

A couple of comments are in order. First, notice that the only information about the Laplacian operator in (1.33) that survives in the statement of the theorem is in the condition that the first integral lies in $H_0^1(\Omega)$. This regularity requirement comes exactly from the presence of the Laplacian in (1.33), as irregular first integrals do not prevent strong decay of the solutions of the Cauchy problem. Second, the strong flow essentially forces the eigenfunction to be close to a first integral, and then the variational principle (1.29) for the Laplacian operator is replaced by essentially the same expression (1.44) except that the set of allowed test functions is restricted to the first integrals.

Proof of Theorem 1.2

The proof of this Theorem is nicely short. First, we claim that if u has a non-zero first integral w in $H_0^1(\Omega)$, normalized so that

$$\|w\|_{L^2} = 1,$$

then we have

$$0 \le \lambda_1(\varepsilon) \le \int_\Omega |\nabla w(x)|^2 dx, \tag{1.45}$$

for any $\varepsilon \in \mathbb{R}$. In order to show that (1.45) holds, we take any $w \in \mathcal{I}_0$, and multiply (1.33) by $w^2/(\phi+\delta)$ with $\delta > 0$ fixed:

$$-\int_\Omega \frac{w^2\Delta\phi}{\phi+\delta}dx + \int_\Omega \frac{w^2}{\phi+\delta}(u\cdot\nabla\phi)dx = \lambda_1(\varepsilon)\int_\Omega \frac{w^2\phi}{\phi+\delta}dx. \tag{1.46}$$

Integrating by parts in the first term gives

$$\begin{aligned}-\int_\Omega \frac{w^2\Delta\phi}{\phi+\delta}dx &= \int_\Omega \nabla\phi\cdot\nabla\left(\frac{w^2}{\phi+\delta}\right)dx = \int_\Omega \frac{2w(\phi+\delta)\nabla\phi\cdot\nabla w - w^2|\nabla\phi|^2}{(\phi+\delta)^2}dx\\ &\le \int_\Omega |\nabla w|^2 dx.\end{aligned}$$

The second term in the left side of (1.46) vanishes because $\nabla\cdot u = 0$ and w is a first integral:

$$\int_\Omega \frac{w^2}{\phi+\delta}(u\cdot\nabla\phi)\,dx = \int_\Omega w^2(u\cdot\nabla(\log(\phi+\delta)))\,dx = -\int_\Omega 2w\log(\phi+\delta)(u\cdot\nabla w)dx = 0.$$

The boundary terms above vanish since $w \in H_0^1(\Omega)$ (it vanishes on the boundary). We conclude that

$$\lambda_1(\varepsilon)\int_\Omega \frac{w^2\phi}{\phi+\delta}dx \le \int_\Omega |\nabla w|^2 dx, \tag{1.47}$$

for any $w \in \mathcal{I}_0$. Passing to the limit $\delta \to 0$ in the left side gives (1.45). Thus, existence of a first integral implies that $\lambda_1(\varepsilon)$ are uniformly bounded for all $\varepsilon \in \mathbb{R}$.

On the other hand, if there exists a sequence $\varepsilon_n \to 0$ such that $\lambda_1(\varepsilon_n)$ are bounded, then

$$\int_\Omega |\nabla\phi_n(x)|^2 dx = \lambda_1(\varepsilon_n)\int_\Omega |\phi_n(x)|^2 dx = \lambda_1(\varepsilon_n). \tag{1.48}$$

Here, $\phi_n(x)$ are the associated positive eigenfunctions $\phi_n(x)$ normalized so that $\|\phi_n\|_{L^2(\Omega)} = 1$. As $\lambda_1(\varepsilon_n)$ are uniformly bounded, it follows from (1.48) that there is a subsequence ϕ_{n_k} that converges weakly in $H_0^1(\Omega)$ and strongly in $L^2(\Omega)$ to a function $\bar{w}(x) \in H_0^1(\Omega)$. Next, multiplying (1.33) by ε_{n_k} and passing to the limit $k \to +\infty$ gives

$$\varepsilon_{n_k}\int_\Omega (-\Delta\phi_{n_k})\eta dx = \varepsilon_{n_k}\int_\Omega (\nabla\phi_{n_k})\cdot\nabla\eta)dx \to 0, \tag{1.49}$$

for any test function $\eta \in H_0^1(\Omega)$, because of (1.48). We also have

$$\varepsilon_{n_k}\lambda_{n_k}\int_\Omega \phi_{n_k}\eta dx \to 0, \tag{1.50}$$

because $\|\phi_{n_k}\|_{L^2} = 1$ and $\lambda_1(\varepsilon_{n_k})$ is bounded. It follows that

$$u\cdot\nabla\bar{w} = 0, \quad \text{weakly in } H_0^1(\Omega),$$

which is the same as

$$u\cdot\nabla\bar{w} = 0, \quad \text{a.e. in } \Omega,$$

and

$$\|\bar{w}\|_{L^2(\Omega)} = 1. \tag{1.51}$$

Hence, $\bar{w}$ is a first integral of u in $H_0^1(\Omega)$. Thus, the non-existence of the first integral in $H_0^1(\Omega)$ implies that

$$\lim_{\varepsilon\to 0} \lambda_1(\varepsilon) = +\infty. \tag{1.52}$$

Finally, to show that (1.44) holds if there is a first integral in $H_0^1(\Omega)$, let us assume, once again, that there exists a sequence $\varepsilon_n \to 0$ such that $\lambda_1(\varepsilon_n)$ are bounded. As the convergence of the subsequence ϕ_{n_k} to the first integral $\bar{w}$ is strong in $L^2(\Omega)$ and weak in $H_0^1(\Omega)$, it follows from (1.48), (1.51) and Fatou's lemma that

$$\liminf_{n\to+\infty} \lambda_1(\varepsilon_n) \geq \int_\Omega |\nabla \bar{w}(x)|^2 dx. \tag{1.53}$$

It remains to notice that (1.53) and (1.45) together imply the Rayleigh quotient formula (1.44), and that the minimum is achieved at $\bar{w}(x)$, finishing the proof of Theorem 1.2.

Directions we are not going to take

Let us finish this introductory section mentioning two directions that are important for the interaction of fast flow and diffusion but that we will not discuss. First, there is a large literature on estimating the effective diffusion in random and periodic flows, and its dependence on the fluid flow strength. Second, there is a very beautiful theory by Freidlin and Wentzell on weakly perturbed two-dimensional Hamiltonian flows.

2 Relaxation enhancement in time

As we have discussed, one interpretation of the eigenvalue enhancement estimate in Theorem 1.2 is in terms of the long time decay rate of the solution to the Cauchy problem

$$\begin{aligned} &\psi_t + \frac{1}{\varepsilon} u \cdot \nabla \psi = \Delta \psi, \quad t > 0, x \in \Omega, \\ &\psi(t,x) = 0 \text{ on } \partial\Omega, \\ &\psi(0,x) = g(x), \end{aligned} \tag{2.1}$$

in Ω with the Dirichlet boundary condition. Its solution has the long time asymptotics

$$\psi(t,x) \sim e^{-\lambda_1(\varepsilon)t}\phi(x) \tag{2.2}$$

for $t \gg 1$. Here, $\phi(x)$ is the principal eigenfunction of the operator

$$-\Delta\phi + \frac{1}{\varepsilon} u \cdot \nabla \phi = \lambda_1(\varepsilon)\phi, \tag{2.3}$$

with the Dirichlet boundary conditions. We have seen in Theorem 1.2 that the principal eigenvalue, or the exponential rate of decay in (2.2), satisfies

$$\lambda_1(\varepsilon) \to +\infty \text{ as } \varepsilon \to 0 \tag{2.4}$$

if and only if the flow u has no first integrals in $H_0^1(\Omega)$.

The "eigenvalue approach" to improved mixing by an interaction of a fluid flow and diffusion gets much more complicated if we pose the Neumann boundary conditions on $\partial\Omega$, or if Ω is a manifold without boundary, such as a torus. In that case, the principal eigenfunction is a constant, and the principal eigenvalue vanishes: $\lambda_0 = 0$, regardless of what the flow $u(x)$ is. One may instead study the second eigenvalue but that is not simple since we do not even know a priori that the second eigenvalue is real, or simple, and finding estimates for the real part of a complex eigenvalue that corresponds to an eigenfunction that also need not be real would not be an easy task. Moreover, even if the spectral gap estimate were available, generally it only provides a long time dynamical information, and how fast the long time limit is achieved may depend on ε, since the operator in the left side of (2.3) is neither self-adjoint nor normal: it does not commute with its formal adjoint operator

$$L^*\phi = -\Delta\phi - \frac{1}{\varepsilon}\nabla\cdot(u\phi).$$

This means that the long time behavior may depend not just on the spectrum of L but also on its pseudo-spectrum: the set of λ for which $(L-\lambda I)^{-1}$ exists but is large in an appropriate norm. It is a rather typical situation that the dynamical information is not quite easy to deduce from the spectrum alone.

On the other hand, our general interest is in the speed of convergence of the solution to an equilibrium, the relaxation speed, and there are other ways to measure this, not in terms of the spectrum. Therefore, rather than try to address the spectral behavior, we will reformulate our questions purely in terms of the Cauchy problem. On the other hand, the information we will obtain will not translate into non-trivial quantitative properties of the spectrum in a straightforward way.

Relaxation enhancement in shear flows and hypoellipticity

A reasonable way to approach the relaxation speed for a parabolic equation of the form

$$\psi_t^\varepsilon + \frac{1}{\varepsilon}u\cdot\nabla\psi^\varepsilon = \Delta\psi^\varepsilon, \quad \psi^\varepsilon(0,x) = \psi_0(x), \tag{2.5}$$

posed in an unbounded domain is in terms of the $L^1 - L^\infty$ decay of the solutions. We will always assume that $u(x)$ is incompressible:

$$\nabla\cdot u = 0, \tag{2.6}$$

so that the flow map for the ODE

$$\frac{dX}{dt} = u(X), \quad X(0) = x, \tag{2.7}$$

is measure preserving, and the total mass is preserved by (2.5):

$$\int_{\mathbb{R}^n}\psi^\varepsilon(t,x)dx = \int_{\mathbb{R}^n}\psi_0(x)dx. \tag{2.8}$$

As the total mass of $\psi^\varepsilon(t,x)$ is conserved, we can measure the additional mixing by $u(x)$ in terms of the decay of the L^∞-norm of $\psi^\varepsilon(t,x)$: the smaller $\|\psi^\varepsilon(t,\cdot)\|_{L^\infty}$ is, the more evenly the mass of $\psi^\varepsilon(t,x)$ is spread around. One can rephrase this in terms of the decay of any L^p-norm with $p>1$ but the L^∞-norm gives the most intuitive picture. Let us stress that we always talk about the decay of $\|\psi^\varepsilon(t,\cdot)\|_{L^\infty}$ at a fixed time $t>0$ when $\varepsilon \ll 1$ is sufficiently small, and not as $t\to+\infty$. The reader may think simply of the L^∞-norm of $\psi^\varepsilon(t,x)$ at $t=1$.

One can get a simple estimate on the L^1-L^∞ decay multiplying (2.5) by u and integrating by parts. Using incompressibility of u, gives

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^n}|\psi^\varepsilon|^2dx = -\int_{\mathbb{R}^n}|\nabla\psi^\varepsilon|^2dx. \tag{2.9}$$

Exercise 2.1 Use the Nash inequality, conservation of the total mass, and (2.9) to show that there exists a constant $C>0$ so that for all incompressible $u(x)$ and $\varepsilon>0$ we have

$$|\psi^\varepsilon(t,x)| \le \frac{C}{t^{n/2}}\|\psi_0\|_{L^1}. \tag{2.10}$$

The fact that the constant $C>0$ does not depend on u or $\varepsilon>0$ in (2.9) shows that no incompressible flow can have too little mixing. The next exercise shows that this universal decay also incorporates the incompressibility of $u(x)$ so it does take into account some of the mixing properties.

Exercise 2.2 Show that no such estimate (2.10) may hold with the same constant $C>0$ for all $u(x)$ without the incompressibility assumption (2.6).

However, the estimate (2.10) does not show in any way an improvement of mixing by the flow $u(x)$. In general, this is quite difficult, so let us first look at a simple special case when everything can be done more or less explicitly: the shear flows considered in [4]. These are unidirectional flows of the form $u=(v(y),0)$, with a scalar-valued function $v(y)$. Here, we have introduced the coordinates $x=(x_1,y)$ on $\mathbb{R}^n$, with $x_1\in\mathbb{R}$ and $y\in\mathbb{R}^{n-1}$. Such flows automatically satisfy the incompressibility condition: $\nabla\cdot u=0$. For simplicity, we will consider (2.5) in the cylinder $\Omega=\mathbb{R}\times\mathbb{T}^{n-1}$, so that both $v(y)$ and the solution to (2.5) are 1-periodic in the y_k-variables, $k=1,\dots,n-1$.

Let $\psi(t,x_1,y)$ be 1-periodic in $y\in\mathbb{T}^{n-1}$ and satisfy

$$\psi_t + \frac{1}{\varepsilon}v(y)\frac{\partial\psi}{\partial x_1} = \Delta_{x_1,y}\psi, \tag{2.11}$$

with the initial condition $\psi(0,x_1,y)=\phi_0(x_1,y)$. It is clear that if $v(y)\equiv\bar v$ is a constant flow, then the L^1-L^∞ decay of $\psi(t,x,y)$ is exactly the same as for the equation with $v(y)\equiv 0$, as $\bar v$ simply translates the solution in the x_1-direction. Another clear obstacle to a faster L^1-L^∞ is the existence of a plateau in the profile $v(y)$: if $v(y)\equiv\bar v$ for all $y\in D$, where D is some open set. Indeed, in that case we may bound $\psi(t,x,y)$ from below by the solution to

$$\tilde\psi_t + \frac{1}{\varepsilon}\bar v\frac{\partial\tilde\psi}{\partial x_1} = \Delta_{x_1,y}\tilde\psi, \quad (x_1,y)\in\mathbb{R}\times D, \tag{2.12}$$

with the Dirichlet boundary condition on ∂D. This is a translate of the solution to

$$\phi_t = \Delta_{x_1,y}\phi, \quad (x_1, y) \in \mathbb{R} \times D, \tag{2.13}$$

so that

$$\psi(t, x_1, y) \geq \tilde{\psi}(t, x_1, y) = \phi(t, x_1 - \bar{v}t, y), \quad (x_1, y) \in \mathbb{R} \times D. \tag{2.14}$$

This means that there is no speed-up of the $L^1 - L^\infty$ decay for $\psi^\varepsilon(t, x, y)$ due to the flow $v(y)$ if $v(y)$ has a plateau – the rate of the decay, for a fixed $t > 0$, is constrained by the principal Dirichlet eigenvalue of the domain D.

Let us now assume that $v(y)$ does not have a plateau to see if "no plateau" is a sufficient condition for an improved $L^1 - L^\infty$ decay. A very simple observation is that $\psi(t, x, y)$ can be written as

$$\psi(t, x_1, y) = \int_{-\infty}^{\infty} G(t, x_1 - z)\Psi(t, z, y)dz \tag{2.15}$$

with the function $\Psi(t, x, y)$ satisfying the degenerate parabolic equation

$$\Psi_t + \frac{1}{\varepsilon}v(y)\frac{\partial \Psi}{\partial x_1} = \Delta_y \Psi, \tag{2.16}$$

with the initial condition $\Psi(0, x_1, y) = \psi_0(x_1, y)$ and the one-dimensional heat kernel

$$G(t, x_1) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x_1^2}{4t}\right).$$

Indeed, if $\Psi(t, x_1, y)$ is a solution to (2.16), then $\psi(t, x_1, y)$ defined by (2.15) satisfies

$$\begin{aligned}
\psi_t + \frac{1}{\varepsilon}v(y)\frac{\partial \psi}{\partial x_1} &= \int_{-\infty}^{\infty} [\Delta_{x_1} G(t, x_1 - z)]\Psi(t, z, y)dz \\
&+ \int_{-\infty}^{\infty} G(t, x_1 - z)\Big[\Delta_y \Psi(t, z, y) - \frac{1}{\varepsilon}v(y)\frac{\partial \Psi(t, z, y)}{\partial z}\Big]dz \\
&+ \frac{1}{\varepsilon}\int_{-\infty}^{\infty} v(y)\frac{\partial G(t, x_1 - z)}{\partial x_1}\Psi(t, z, y)dz \\
&= \int_{-\infty}^{\infty} [\Delta_{x_1} G(t, x_1 - z)\Psi(t, z, y) + G(t, x_1 - z)\Delta_y \Psi(t, z, y)]dz = \Delta_{x_1,y}\psi(t, x, y),
\end{aligned} \tag{2.17}$$

so that, indeed, $\psi(t, x_1, y)$ satisfies (2.11).

The operator

$$\mathcal{L}_\varepsilon \Psi = -\Delta_y \Psi + \frac{1}{\varepsilon}v(y)\frac{\partial \Psi}{\partial x_1} \tag{2.18}$$

that appears in (2.16) is not uniformly elliptic: it lacks the Laplacian in the x_1-direction. It is, however, hypoelliptic [19] if there is no point $y \in \mathbb{T}^{n-1}$, where all derivatives of $v(y)$ vanish. We will call this the H-condition. Indeed, the Lie algebra generated by the operators ∇_y and $v(y)\partial_x$ consists of vector fields of the form

$$\nabla_y, v(y)\frac{\partial}{\partial x}, \frac{\partial v(y)}{\partial y_k}\frac{\partial}{\partial x}, \frac{\partial^2 v(y)}{\partial y_k \partial y_m}\frac{\partial}{\partial x}, \ldots, v^{(n)}(y)\frac{\partial}{\partial x}, \ldots \tag{2.19}$$

which span $\mathbb{R}^n$ if $v(y)$ satisfies the H-condition. The study of existence of smooth fundamental solutions for such degenerate operators was initiated by Kolmogorov [24]. Kolmogorov's work with $v(y) = y$ served in part as a motivation for the fundamental result on characterization of hypoelliptic operators of Hörmander [19]. The "no plateau" condition for $v(y)$ is not equivalent to the H-condition but is reasonably close to it.

If $v(y)$ satisfies the H-condition, then the theory of Hörmander [19], and the results of Ichihara and Kunita [18] imply that there exists a smooth transition probability density $p_\varepsilon(t, x_1, y, y')$ such that

$$\Psi(t, x_1, y) = \int\limits_{\mathbb{R}} \int\limits_0^H p_\varepsilon(t, x_1 - x', y, y')\, \psi_0(x', y') dy' dx. \tag{2.20}$$

In particular, the function $p_\varepsilon(t)$ is uniformly bounded from above for any $t > 0$ [18]. Then we have

$$\|\psi(t)\|_{L^\infty_{x_1,y}} \le \|\Psi(t)\|_{L^\infty_{x_1,y}} \le \|p_\varepsilon(t)\|_{L^\infty_{x_1,y}} \|\psi_0\|_{L^1_{x_1,y}}. \tag{2.21}$$

It is straightforward to observe that p_ε has a simple scaling property

$$p_\varepsilon(t, x_1, y, y') = \varepsilon p_0(t, \varepsilon x, y, y') \tag{2.22}$$

with p_0 being the transition probability density for (2.16) with $\varepsilon = 1$. That is, p_0 satisfies

$$\frac{\partial p_0}{\partial t} + v(y) \frac{\partial p_0}{\partial x_1} = \Delta_y p_0, \tag{2.23}$$

with the initial condition $p_0(0, x, y, y') = \delta(x)\delta(y - y')$. Therefore, we obtain

$$\|\psi(t)\|_{L^\infty} \le \varepsilon \|p_0(t)\|_{L^\infty_{x,y}} \|\psi_0\|_{L^1_{x,y}} \le C\varepsilon \|\psi_0\|_{L^1_{x,y}}. \tag{2.24}$$

This is a version of the relaxation enhancement in the whole space that we were looking for: the $L^1 - L^\infty$ decay at a fixed time $t > 0$ is faster as $\varepsilon \to 0$.

As a side remark, we note that this very simple example also shows a connection between relaxation enhancement and hypoellipticity.

Relaxation enhancing flows on a torus

Let us now consider relaxation enhancement on the n-dimensional torus. The discussion below applies verbatim to the case of a smooth compact n-dimensional Riemannian manifold Ω, and generalizations are very straightforwad, so we do not discuss them – see [5] for some of the full cornucopia. We consider solutions to the passive scalar equation

$$\phi^\varepsilon_t + \frac{1}{\varepsilon} u(x) \cdot \nabla \phi^\varepsilon - \Delta \phi^\varepsilon = 0, \quad \phi^\varepsilon(0, x) = \phi_0(x), \tag{2.25}$$

on $\Omega = \mathbb{T}^n$, supplemented by periodic boundary conditions. As always, we assume that u is incompressible: $\nabla \cdot u = 0$. The solution $\phi^\varepsilon(t, x)$ tends to its average,

$$\overline{\phi}(t) \equiv \frac{1}{|\Omega|} \int_\Omega \phi^\varepsilon(t, x)\, d\mu = \frac{1}{|\Omega|} \int_\Omega \phi_0(x)\, dx, \tag{2.26}$$

as $t \to +\infty$. Here $|\Omega|$ is the volume of Ω. To see that, first, integrating (2.25) over M and using incompressibility of $u(x)$ gives

$$\frac{d}{dt}\int_\Omega \phi(t,x)dx = 0,$$

hence $\bar{\phi}(t) = \bar{\phi}(0)$ is preserved in time. Next, multiplying (2.25) by $\phi^\varepsilon(t,x) - \bar{\phi}$, and again using incompressibility of $u(x)$, we have

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |\phi(t,x) - \bar{\phi}|^2 dx = -\int_\Omega |\nabla\phi^\varepsilon(t,x)|^2 dx. \tag{2.27}$$

The Poincaré inequality implies that

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |\phi(t,x) - \bar{\phi}|^2 dx \le -C_p\int_\Omega |\phi^\varepsilon(t,x) - \bar{\phi}|^2 dx, \tag{2.28}$$

whence

$$\|\phi(t,\cdot) - \bar{\phi}\|_{L^2(\Omega)} \le e^{-C_p t}\|\phi_0 - \bar{\phi}\|_{L^2(\Omega)}. \tag{2.29}$$

Exercise 2.3 Strengthen this result to show that

$$\|\phi(t,\cdot) - \bar{\phi}\|_{L^\infty(\Omega)} \to 0 \text{ as } t \to +\infty. \tag{2.30}$$

Note that the decay rate in (2.29) holds for all incompressible flows $u(x)$, with the same constant C_p – this is the analog of the universal $L^1 - L^\infty$ decay estimate (2.10) that holds in the whole space for all incompressible flows $u(x)$. The same is true for (2.30) – the rate of convergence is uniform in incompressible flows.

We would like to understand how the speed of convergence to the average in (2.30) depends on the properties of the flow and determine which flows are particularly efficient in enhancing the relaxation process. We will use the following "fixed time" (no long time limit!) definition as a measure of the flow efficiency in improving the relaxation of the solution to a uniform state.

Definition 2.4 *An incompressible flow u is relaxation enhancing if for all $\tau > 0$ and $\delta > 0$, there exists $\varepsilon(\tau,\delta)$ such that for any $\varepsilon < \varepsilon(\tau,\delta)$ and any $\phi_0 \in L^2(\Omega)$, with $\|\phi_0\|_{L^2(\Omega)} = 1$, we have*

$$\|\phi^\varepsilon(\tau,\cdot) - \overline{\phi}\|_{L^2(\Omega)} < \delta, \tag{2.31}$$

where $\phi^\varepsilon(t,x)$ is the solution of (2.25) and $\overline{\phi}$ the average of ϕ_0.

Exercise 2.5 Show that the choice of the L^2 norm in the definition is not essential and can be replaced by any L^p-norm with $1 \le p \le \infty$, without changing the class of relaxation enhancing flows.

Let us mention that there are various results on Gaussian and other estimates on the heat kernel corresponding to the incompressible drift and diffusion on manifolds such as in the work of Norris [27] and Franke [17], but these estimates lead to upper bounds on the convergence rate to the equilibrium which essentially do not improve as $\varepsilon \to 0$, and thus do not quite address the effect of a strong flow. Such general estimates often deteriorate as the flow gets stronger, which is exactly the opposite of what interests us. Surprisingly, there seems to be no general method to incorporate the "helpful" affects of the advection into the proofs of the heat kernel estimates.

The original motivation for this definition came from the work of Fannjiang, Nonnemacher and Wolowski [12, 13, 14], where relaxation enhancement was studied in the discrete setting (see also [21] for related earlier references). In these papers, a unitary evolution step (a certain measure preserving map on the torus) alternates with a dissipation step, which, for example, acts simply by multiplying the Fourier coefficients by damping factors. The absence of sufficiently regular eigenfunctions appears as a key for the enhanced relaxation in this particular class of dynamical systems. In [12, 13, 14], the authors also provide finer estimates of the dissipation time for particular classes of toral automorphisms – they estimate how many steps are needed to reduce the L^2 norm of the solution by a factor of two if the dissipation strength is ε.

To understand why and when we expect relaxation enhancement, let us first look at the time-splitting approximation for (2.25), in the spirit of [12, 13, 14]. Assume that $\psi(t,x)$ solves the advection equation

$$\psi_t + \frac{2}{\varepsilon} u \cdot \nabla\psi = 0, \quad n\tau \le t \le (n+1/2)\tau, \tag{2.32}$$

followed by the heat equation

$$\psi_t = 2\Delta\psi, \quad (n+1/2)\tau \le n\tau, \tag{2.33}$$

and then again (2.32) followed by (2.33), and so on. As the time step $\tau \to 0$, the solution of this time-splitting scheme converges to the solution of (2.32). However, the smallness of τ that is required to make the error small depends on ε in a way that is very difficult to control efficiently. If we, in a cavalier fashion, instead fix the size of τ that is independent of ε, then solution of the very first step is

$$\psi(\tau/2, x) = \phi_0(X(\tau/\varepsilon, x)), \tag{2.34}$$

where $X(t,x)$ is the trajectory

$$\dot{X}(t) = -u(X), \quad X(0) = x. \tag{2.35}$$

If the flow of (2.35) is sufficiently complex and ε is sufficiently small, the points $X(\tau/\varepsilon, x)$ and $X(\tau/\varepsilon, x')$ may be very far apart, even if x and x' are very close. This would make the difference $\psi(\tau/2,x) - \psi(\tau/2,x')$ large, so that the function $\psi(\tau/2,x)$ given by (2.34) would have a large gradient. This means that the initial condition for the second step in the time-splitting scheme

$$\psi_t = 2\Delta\psi, \quad \tau/2 \le \tau, \tag{2.36}$$

has a very large gradient. On the other hand, the dissipation identity for (2.36)

$$\frac{1}{2}\frac{d}{dt}\int_\Omega |\psi - \bar{\psi}|^2 = -2\int_\Omega |\nabla\psi|^2 dx \tag{2.37}$$

tells us that solutions with a large gradient and zero average decay very fast. Therefore, we would deduce that for "sufficiently mixing" flows $u(x)$ solutions of this time splitting scheme converge to their average very fast if ε is small. The problem with making this argument rigorous is that, as we have mentioned, for the convergence of the time-splitting scheme to the true solution we would need to take τ not fixed but $\tau \ll \varepsilon$, making the interaction of advection and diffusion non-trivial and very difficult to account for carefully. Nevertheless, this intuition is correct. Here is the main result of this section.

Theorem 2.6 *([5]) A Lipschitz continuous incompressible flow $u \in Lip(\Omega)$ is relaxation enhancing if and only if the operator $u \cdot \nabla$ has no eigenfunctions in $H^1(\Omega)$, other than the constant function.*

As in the Berestycki-Hamel-Nadiarshvili theorem, the only information about the Laplacian is in the requirement that the eigenfunction lies in the space $H^1(\Omega)$: rough eigenfunctions do not preclude relaxation enhancement. We will explain below how flows with such rough eigenfunctions can be constructed, more or less explicitly.

The "sufficiently mixing" property of u is encoded in this theorem in the same requirement that it does not have an eigenfunction in $H^1(\Omega)$. The reason for that condition can also be seen from the approximation by the time-splitting scheme (2.34)-(2.36). The operator $u \cdot \nabla$ is skew-symmetric when u is divergence free:

$$\int_\Omega (u \cdot \nabla\eta(x))\eta(x)dx = 0, \tag{2.38}$$

for all $\eta \in H^1(\Omega)$. Therefore, all eigenvalues $\lambda = i\omega$ of $u \cdot \nabla$ are purely imaginary, and if $\phi \in H^1(\Omega)$ is an eigenfunction:

$$u \cdot \nabla\phi = i\omega\phi, \tag{2.39}$$

then solution of (2.34)

$$\psi_t + \frac{2}{\varepsilon} u \cdot \nabla\psi = 0, \quad 0 \le t \le \tau/2 \tag{2.40}$$

with the initial condition $\psi(0, x) = \phi(x)$ satisfies

$$\psi(t, x) = e^{2i\omega t/\varepsilon}\phi(x).$$

Therefore, the H^1-norm of $\psi(t, x)$ does not increase:

$$\|\psi(\tau/2, x)\|_{H^1(\Omega)} = \|\phi\|_{H^1(\Omega)},$$

hence the advection step does not prepare an irregular initial condition for the heat equation in the second step of the time-splitting scheme, and there is no intuitive reason to expect relaxation enhancement when $\varepsilon \to 0$.

The discrepancy between Theorems 1.2 and 2.6 may seem surprising – after all, on the physical level, the conditions for the relaxation enhancement and eigenvalue enhancement need not be very different but the eigenvalue enhancement (with the Dirichlet boundary conditions) requires that the operator $u \cdot \nabla$ does not have first integrals while relaxation enhancement (with the periodic or Neumann boundary conditions) requires that this operator does not have eigenfunctions in $H^1(\Omega)$ with any eigenvalue (the first integral corresponds to a zero eigenvalue). This issue is resolved by the following

Proposition 2.7 *Let $u \in Lip(\Omega)$. If $\phi \in H^1(\Omega)$ is an eigenfunction of the operator $u \cdot \nabla$ corresponding to an eigenvalue $i\omega$, $\omega \in \mathbb{R}$, then $|\phi| \in H^1(\Omega)$ and it is the first integral of u, that is, $u \cdot \nabla|\phi| = 0$.*

Proof. The fact that $|\phi| \in H^1$ follows from the well-known properties of Sobolev functions (see, for example, [11]). If $\phi(x)$ satisfies

$$u \cdot \nabla\phi = i\omega\phi$$

then

$$u \cdot \nabla|\phi|^2 = u \cdot \nabla(\phi\bar{\phi}) = \phi(u \cdot \nabla\bar{\phi}) + \bar{\phi}(u \cdot \nabla\phi) = -i\omega\phi\bar{\phi} + i\omega\phi\bar{\phi} = 0,$$

hence $u \cdot \nabla|\phi| = 0$. □

Therefore, in the case of the Dirichlet boundary conditions, if $\phi \in H_0^1(\Omega)$ is an eigenfunction of the operator $u \cdot \nabla$ then $|\phi|$ is its first integral. Naturally, $|\phi|$ can not be equal identically to a constant since ϕ satisfies the Dirichlet boundary conditions, as it lies in $H_0^1(\Omega)$, and $\phi \not\equiv 0$. Moreover, if $\phi \in H_0^1(\Omega)$ is a first integral: $u \cdot \nabla\phi = 0$ then it is an eigenfunction corresponding to eigenvalue $\lambda = 0$. Hence, for the Dirichlet boundary conditions the requirement that $u \cdot \nabla$ does not have a first integral in $H_0^1(\Omega)$ is equivalent to the condition that it does not have eigenfunctions in $H_0^1(\Omega)$.

On the other hand, existence of mean zero $H^1(\Omega)$ eigenfunctions, without imposing the Dirichlet boundary condition, need not guarantee the existence of a mean zero first integral, as can be seen from the following well-known example. Let $\alpha \in \mathbb{R}^n$ be a constant vector generating an irrational rotation on the n-dimensional torus Ω, in the sense that the components of α are independent over the rationals. The operator $\alpha \cdot \nabla$ has eigenvalues $2\pi i(\alpha \cdot k)$, with any $k \in \mathbb{Z}^n$. The corresponding eigenfunctions are

$$w_k(x) = e^{2\pi i k \cdot x}.$$

Their absolute value is 1, which is a first integral of $\alpha \cdot \nabla$ but there are no non-constant first integrals since α is irrational. Indeed, if there exists a function $\psi \in L^1(\Omega)$ such that

$$\psi(x + \alpha t) = \psi(x), \quad \text{for all } x \in \Omega \text{ and all } t \in \mathbb{R},$$

then the Fourier coefficients of the function ψ, defined by

$$\psi(x) = \sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x}\hat{\psi}_k, \quad \hat{\psi}_k = \int_\Omega e^{-2\pi i k \cdot y}\psi(y)dy,$$

should satisfy

$$\hat{\psi}_k = e^{2\pi i k \cdot \alpha t}\hat{\psi}_k, \quad \text{for all } k \in \mathbb{Z}^n, \text{ and all } t \in \mathbb{R}.$$

Therefore, either all $\hat{\psi}_k = 0$ for $k \neq 0$, or there exists $k \neq 0$ such that

$$k \cdot \alpha = 0.$$

The latter, however, is impossible since α is irrational. Hence, $\hat{\psi}_k = 0$ for all $k \neq 0$, and the only first integrals of $\alpha \cdot \nabla$ for an irrational α are constant functions. Thus, this flow is not relaxation enhancing, since it has eigenfunctions in $H^1(\Omega)$, even though it has no first integrals other than a constant function.

Examples of relaxation enhancing flows

We now present some examples of relaxation enhancing flows on a torus, to assure the reader that this class is not empty. We first describe flows with very rough eigenfunctions, none of which lie in $H^1(\Omega)$, and then flows that have no eigenfunctions – they are weakly mixing. In both cases, the construction is based on a simple modification of a shear flow.

Flows with rough eigenfunctions

Here, we describe a smooth incompressible flow $u(x,y)$, $\nabla \cdot u = 0$, on a torus $\mathbb{T}^2$ that has a purely discrete spectrum but none of the eigenfunctions are in $H^1(\mathbb{T}^2)$. The idea of the construction goes back to Kolmogorov [24]. We present some but not all of the full technical details of the construction [2, 20]. We denote by U^t the flow on $L^2(\mathbb{T}^2)$ generated by u:

$$U^t f(x) = f(X(t;x)),$$

where $X(t,x)$ is the trajectory of

$$\frac{dX}{dt} = -u(X), \quad X(0;x) = x.$$

If $\nabla \cdot u = 0$, so that $u(x)$ is incompressible, then U^t is a map on $C^\infty(\mathbb{T}^2)$ that preserves the L^2-norm. Hence, it can be extended to a unitary map on $L^2(\mathbb{T}^2)$. When we talk of the spectrum of $u \cdot \nabla$, we mean the spectrum of the map U^t: a function $f \in L^2(\Omega)$ is an eigenfunction of the flow u if for any $t \in \mathbb{R}$ there exists $c(t)$ so that

$$U^t f(x) = c(t) f(x). \tag{2.41}$$

This definition is equivalent to the condition that

$$u \cdot \nabla f = \lambda f. \tag{2.42}$$

Indeed, the function $g(t,x) = U^t f(x)$ satisfies the advection equation

$$g_t + u \cdot \nabla g = 0, \quad g(0,x) = f(x). \tag{2.43}$$

therefore, if (2.42) holds then

$$f(X(t,x)) = e^{\lambda t} f(x),$$

so that (2.41) holds with $c(t) = e^{\lambda t}$. On the other hand, if (2.41) holds then the solution to (2.43) has the form

$$g(t,x) = c(t)f(x).$$

Inserting this expression into (2.43) gives

$$\dot{c}(t)f(x) + c(t)u(x)\cdot\nabla f(x) = 0.$$

Separation of variables now implies that there exists $\lambda \in \mathbb{C}$ such that (note that $c(0) = 1$ automatically)

$$c(t) = e^{-\lambda t},$$

and

$$u(x)\cdot\nabla f(x) = \lambda f(x).$$

Moreover, as the map $x \to X(t;x)$ is measure preserving for all $t \in \mathbb{R}$, $|c(t)| = 1$ for all t, whence λ is purely imaginary: $\lambda = i\omega$ with a real number ω.

Exercise 2.8 The above argument made an implicit assumption that $f \in C^1(\mathbb{T}^2)$. Explain what needs to be done to get rid of this assumption.

Here is the key result of this section.

Proposition 2.9 *There exists a smooth incompressible (with respect to the Lebesgue measure) flow $u(x,y)$ on a two-dimensional torus $\mathbb{T}^2$ so that the corresponding unitary evolution U^t has a discrete spectrum on $L^2(\mathbb{T}^2)$ but none of the eigenfunctions of U^t are in $H^1(\mathbb{T}^2)$.*

Proof. The basic idea behind the construction is quite simple: we want to create a unidirectional flow such that the speed with which the particles move along various lines is sufficiently mismatched to create large gradients. If the speed were constant along each straight line trajectory, that would be a shear flow. We will see that it is impossible so we will need the speed to vary along the trajectory. This is incompatible with incompressibility but the flow will preserve another measure that has a non-constant density with respect to the Lebesgue measure. An appropriate mapping of a flow constructed this way will lead to an incomopressible flow. On a slightly more technical level, we will look for a flow that can be mapped to a constant flow $\bar{u} = (\alpha, 1)$ by a measure preserving map S with very low regularity properties. Since the eigenfunctions of the constant flow are explicitly computable, we can compute the eigenfunctions of the original flow. Due to the roughness of S, these will be highly irregular.

As outlined above, we will look at a time change of the constant linear translation flow, of the form

$$\frac{dx}{dt} = \frac{\alpha}{F(x,y)}, \quad \frac{dy}{dt} = \frac{1}{F(x,y)}, \quad x(0) = x_0, \quad y(0) = y_0, \tag{2.44}$$

with an appropriately chosen $\alpha \in \mathbb{R}$ and $F(x,y)$. The trajectories of (2.44) are straight lines:

$$x(t) - \alpha y(t) = x_0 - \alpha y_0, \text{ for all } t \geq 0, \tag{2.45}$$

and the time it takes for the trajectory to go from a point $(x,0)$ at height $y = 0$ to the point $(x + \alpha y, y)$, when the trajectory reaches the height y is

$$T(x,y) = \int_0^y F(x + \alpha z, z)dz. \tag{2.46}$$

Hence, the function $F(x,y)$ is simply the local time change of the flow.

It would be very convenient to take $F(x,y)$ in the form

$$F(x,y) = Q(x-\alpha y), \tag{2.47}$$

so that $F(x,y)$ would be constant on each trajectory of the flow (2.44), and (2.44) would really be a shear flow in the direction $(\alpha, 1)$. However, for $F(x,y)$ as in (2.47) to be 1-periodic both in x and y, the function $Q(x)$ has to be both 1-periodic and α-periodic. If α is irrational, this is impossible unless $Q(x) \equiv \text{const}$.

Thus, instead of trying (2.47), we use cut-offs to modify (2.47), setting

$$F(x,y) = m + \psi(y)(Q(x-\alpha y) - m), \quad 0 \le x, y \le 1. \tag{2.48}$$

Here, $Q(x,y) > 0$ is a 1-periodic function $Q(x) > 0$ such that

$$\int_0^1 Q(\xi)d\xi = 1. \tag{2.49}$$

A smooth cut-off function $\psi(y) \ge 0$ in (2.48) is such that

$$\int_0^1 \psi(y)dy = 1, \tag{2.50}$$

and

$$\psi(y) = 0 \text{ for } 0 \le y \le y_0 \text{ and } y_1 \le y \le 1 \text{ with } y_0 \text{ close to zero and } y_1 \text{ close to one.} \tag{2.51}$$

The constant m in (2.48) is such that $0 < m < \min Q(s)$. The choice of m ensures that the function $F(x,y) > 0$ – this is needed both to interpret $F(x,y)$ as a local time change, and to be able to divide by $F(x,y)$ in (2.44). Note that the function $F(x,y)$ is already 1-periodic in x because $Q(x)$ is 1-periodic. As, in addition, $F(x,y) \equiv m$ near $y = 0$ and $y = 1$, we can extend $F(x,y)$ to the whole plane so that it is also periodic in y. The smoothness of $Q(x)$ and (2.51) imply that the extension is smooth. In addition, because of (2.49) and (2.50), the total mass of $F(x,y)$ is

$$\int_0^1 \int_0^1 F(x,y)dxdy = 1. \tag{2.52}$$

In order to map the flow (2.44) to a constant speed flow $(\alpha, 1)$ moving along the same straight lines, it is natural to attempt to define the transformation $S : (x,y) \to (X,Y)$ as

$$\tilde{X}(x,y) = x + \alpha(\tilde{Y}(x,y) - y), \qquad \tilde{Y}(x,y) = T(x - \alpha y, y), \tag{2.53}$$

with $T(x-\alpha y, y)$ as in (2.46) – the time it takes for the particle starting at $t = 0$ at the point $(x - \alpha y, 0)$ at height $y = 0$ to reach the point (x,y) at the height y. This means that the flow speed in the $\tilde{Y}$-direction would be identically equal to one. In addition, the transformation (2.53) satisfies $x - \alpha y = \tilde{X} - \alpha\tilde{Y}$, thus it preserves the flow trajectories, which are straight lines in the direction $(\alpha, 1)$. Hence, the particle would move with a constant speed along the straight lines in the new variables, the speed in the $\tilde{Y}$-direction would equal to one, and in the $\tilde{X}$-direction it would equal to α. However, the map (2.53) is not well-defined on the torus $\mathbb{T}^2$: it is easy to see that $\tilde{Y}(x,y)$ is not 1-periodic in y, even modulo 1, because

$$\tilde{Y}(x,y+1)T(x-\alpha y-\alpha,y+1) \neq T(x-\alpha y,y) = \tilde{Y}(x,y) \bmod 1. \tag{2.54}$$

However, we do know that the increment

$$\begin{aligned}
&\tilde{Y}(x,y+1)-\tilde{Y}(x,y) = T(x-\alpha y-\alpha,y+1)-T(x-\alpha y,y)\\
&=\int_0^{y+1}F(x-\alpha y-\alpha+\alpha z,z)dz-\int_0^y F(x-\alpha y+\alpha z,z)dz\\
&=\int_0^1 F(x-\alpha y-\alpha+\alpha z,z)dz\\
&+\int_1^{y+1}F(x-\alpha y-\alpha+\alpha z,z)dz-\int_0^y F(x-\alpha y+\alpha z,z)dz\\
&=\int_0^y F(x-\alpha y+\alpha z,z+1)dz-\int_0^y F(x-\alpha y+\alpha z,z)dz\\
&+\int_0^1 F(x-\alpha y-\alpha+\alpha z,z)dz=\int_0^1 F(x-\alpha y-\alpha+\alpha z,z)dz
\end{aligned} \tag{2.55}$$

is constant on each line $x-\alpha y=\text{const}$. Thus, in order to make $Y(x,y)$ be 1-periodic in y, we modify (2.53) as [24, 30]

$$X(x,y)=x+\alpha(Y(x,y)-y), \qquad Y(x,y)=T(x-\alpha y,y)+R(x-\alpha y), \tag{2.56}$$

adding a compensatory shift $R(x-\alpha y)$ that is constant on each trajectory.

We claim that if we choose the 1-periodic function $R(x)$ that satisfies the homology equation [2]

$$R(\xi+\alpha)-R(\xi)=Q(\xi)-1, \quad \xi\in\mathbb{S}^1, \tag{2.57}$$

then the map (2.56) is well-defined on $\mathbb{T}^2$. Note that for (2.57) to have a measurable solution the function $Q(\xi)$ should satisfy the normalization (2.49). For the moment, we will not worry about the existence of $R(x)$ and its properties but will come back to this soon.

Let us now check that, indeed, if R is a solution to the homology equation, then (2.56) defines a mapping of the torus onto itself. The shift in x is simple to understand: the function $T(x,y)$ is clearly 1-periodic in x since $F(x,y)$ is periodic in x, thus

$$Y(x+1,y)=Y(x,y), \tag{2.58}$$

while

$$X(x+1,y)=1+X(x,y)=X(x,y) \bmod 1. \tag{2.59}$$

To verify what happens under the shift $y\to y+1$, we first make some preliminary observations. The normalization (2.50) implies that

$$T(x,1)=\int_0^1 F(x+\alpha z,z)dz=\int_0^1[m+\psi(z)(Q(x)-m)]dz=Q(x). \tag{2.60}$$

Now, it follows that

$$\begin{aligned} T(x,y+1) &= \int_0^{y+1} F(x+\alpha z,z)dz = \int_0^1 F(x+\alpha z,z)dz + \int_1^{y+1} F(x+\alpha z,z)dz \\ &= Q(x) + \int_0^y F(x+\alpha+\alpha z,z+1)dz = Q(x) + \int_0^y F(x+\alpha+\alpha z,z)dz \\ &= Q(x) + T(x+\alpha,y). \end{aligned} \tag{2.61}$$

Using this identity, and the homology equation (2.57) for the function R gives

$$\begin{aligned} Y(x,y+1) &= T(x-\alpha y-\alpha,y+1) + R(x-\alpha y-\alpha) \\ &= T(x-\alpha y,y) + Q(x-\alpha y-\alpha) + R(x-\alpha y) - Q(x-\alpha y-\alpha) + 1 \\ &= T(x-\alpha y,y) + R(x-\alpha y) + 1 = Y(x,y)+1 = Y(x,y) \bmod 1. \end{aligned} \tag{2.62}$$

This computation is the reason why we have chosen $R(x)$ as the solution to the homology equation.

Finally, for $X(x,y)$ we have

$$X(x,y+1) = x+\alpha(Y(x,y+1)-y-1) = x+\alpha(Y(x,y)+1-y-1) = x+\alpha(Y(x,y)-y) = X(x,y). \tag{2.63}$$

We conclude from (2.58), (2.59), (2.62) and (2.63) that S is a well-defined mapping of $\mathbb{T}^2$ to itself.

A key observation is that solutions $R(x)$ of the homology equation (2.57) can be very rough even if the function $Q \in C^\infty(\mathbb{S}^1)$ is smooth. To see that, let us go back to (2.57):

$$R(\xi+\alpha) - R(\xi) = Q(\xi) - 1, \quad \xi \in \mathbb{S}^1. \tag{2.64}$$

Note that it can be solved explicitly using the Fourier transform:

$$R(\xi) = \sum_{n\in\mathbb{Z}} \hat{R}_n e^{2\pi i n\xi}, \tag{2.65}$$

with the Fourier coefficients

$$\hat{R}_n = \frac{\hat{Q}_n}{\exp(2\pi i\alpha n) - 1}. \tag{2.66}$$

The denominators in (2.66) can be dangerously small if αn can be very close to an integer, that is, if α is a Liouvillean irrational number. Recall that an irrational number $\alpha \in \mathbb{R}$ is called β-Diophatine if there exists a constant C such that for each $k \in \mathbb{Z} \setminus \{0\}$ we have

$$\inf_{p\in\mathbb{Z}} |\alpha k + p| \geq \frac{C}{|k|^{\beta+1}}.$$

The vector α is Liouvillean if it is not Diophantine for any $\beta > 0$. The Liouvillean numbers (and vectors) are the ones which can be very well approximated by rationals. The following Proposition is a particular case of Theorem 4.5 of [20].

Proposition 2.10 *Let α be a Liouvillean irrational number. There exists a $C^\infty(\mathbb{S}^1)$ function $Q(\xi)$ so that the homology equation (2.57) has a unique (up to an additive constant) measurable solution $R(\xi) : \mathbb{S}^1 \to \mathbb{R}$ such that for any $\lambda \in \mathbb{R}\backslash\{0\}$, the function $R_\lambda(\xi) = e^{i\lambda R(\xi)}$ is discontinuous everywhere.*

We will not prove this proposition here.

Without loss of generality we may assume that $Q(\xi)$ given by Proposition 2.10 is positive: otherwise, we choose M so that $Q(\xi) + M > 1$ and consider a rescaled function

$$Q_M(\xi) = (M + Q(\xi))/(M + 1).$$

Then, the function

$$R_M(\xi) = \frac{R(\xi)}{M+1}$$

is the solution to (2.57) with Q_M in the right side and, of course, $R_M(\xi)$ has the same set discontinuities as $R(\xi)$.

Let us see what happens to the flow (2.44) under the map (2.56):

$$\frac{dx}{dt} = \frac{\alpha}{F(x,y)}, \quad \frac{dy}{dt} = \frac{1}{F(x,y)}, \quad x(0) = x_0, \quad y(0) = y_0. \tag{2.67}$$

Note that

$$x(t) - y(t) = x_0 - \alpha y_0,$$

hence $Y(t)$ is given by

$$Y(t) = T(x(t) - \alpha y(t), y(t)) + R(x(t) - \alpha y(t)) = T(x_0 - \alpha y_0, y(t)) + R(x_0 - \alpha y_0), \tag{2.68}$$

so that

$$\begin{aligned} \frac{dY}{dt} &= \frac{\partial T(x_0 - \alpha y_0, y(t))}{\partial y}\dot{y}(t) = F(x_0 - \alpha y_0 + \alpha y(t), y(t))\frac{1}{F(x(t), y(t)} \\ &= F(x(t) - \alpha y(t) + \alpha y(t), y(t))\frac{1}{F(x(t), y(t)} = 1. \end{aligned} \tag{2.69}$$

On the other hand, for $X(t)$ we have

$$\frac{dX}{dt} = \dot{x}(t) + \alpha(\dot{Y}(t) - \dot{y}(t)) = \frac{\alpha}{F(x(t), y(t)} + \alpha - \frac{\alpha}{F(x(t), y(t)} = \alpha. \tag{2.70}$$

Therefore, the image of the flow (2.44) under S is simply the uniform flow:

$$\frac{dX}{dt} = \alpha, \quad \frac{dY}{dt} = 1, \tag{2.71}$$

as we desired. We will denote $\bar{u} = (\alpha, 1)$.

Note that the map S is invertible with a measurable inverse. Indeed, we have

$$X - \alpha Y = x - \alpha y, \tag{2.72}$$

so that

$$Y = T(X - \alpha Y, y) + R(X - \alpha Y). \tag{2.73}$$

As the function F is strictly positive, the function $T(x, y)$ is strictly increasing in y, so that (2.73) has a unique solution $y(X, Y)$, and then (2.72) defines $x(X, Y)$ uniquely.

In addition, S is measure preserving in the following sense:

$$\int [S^* f](x, y)F(x, y)dxdy = \int f(S(x, y))F(x, y)dxdy = \int f(X, Y)dXdY \tag{2.74}$$

for any function $f \in C(\mathbb{T}^2)$. In order to see that, let us introduce intermediate changes of variables: $S = S_3 \circ S_2 \circ S_1$, with $S_1 : (x, y) \to (z, y_1)$ with

$$z = x - \alpha y, \quad y_1 = y,$$

followed by $S_2 : (z, y_1) \to (Z, y_2)$

$$Z = z, \quad y_2 = T(z, y_1) + R(z),$$

and finally $S_3 : (Z, y_2) \to (X, Y)$, with

$$X = Z + \alpha y_2, \quad Y = y_2.$$

The corresponding Jacobians are:

$$J_1 = J_3 = 1, \quad J_2 = \frac{\partial T}{\partial y_1}(z, y_1) = F(z + \alpha y_1, y_1) = F(x, y).$$

Therefore, the Jacobian of S is, indeed,

$$J = J_1 J_2 J_3 = F(x, y),$$

hence (2.74) holds and S is measure-preserving.

Hence, S^* may be extended as an operator $L^2(dxdy) \to L^2(d\mu)$ with the preservation of the corresponding norms. It follows that the unitary evolutions U_w^t and U_{unif}^t generated by the flow w given by (2.67) and the uniform flow $\bar{u}$, respectively, are conjugated by means of the unitary transformation

$$S^* : \ L^2(\mathbb{T}^2, dXdY) \to L^2(\mathbb{T}^2, d\mu),$$

that is, we have

$$U_{unif}^t = [S^*]^{-1} U_w^t S^*.$$

Therefore, U_w^t and U_{unif}^t have the same spectrum:

$$\lambda_{nl} = 2\pi i n\alpha + 2\pi i l, \quad l, n \in \mathbb{Z}.$$

It also follows that the eigenfunctions of the operator U_w may be written as

$$\begin{aligned}\psi_{nl}^w(x, y) &= e^{2\pi i n X(x,y) + 2\pi i l Y(x,y)} = e^{2\pi i n(x - \alpha y + \alpha Y(x,y)) + 2\pi i l Y(x,y)} \\ &= e^{2\pi i n(x-\alpha y)} e^{(2\pi i n\alpha + 2\pi i l)(T(x-\alpha y, y) + R(x-\alpha y))} = \zeta(x, y) e^{(2\pi i n\alpha + 2\pi i l)R(x-\alpha y)}\end{aligned} \tag{2.75}$$

with a smooth function $\zeta(x,y) \in C^\infty([0,1]^2)$. Note that the function

$$\zeta(x,y) = e^{2\pi i n(x-\alpha y)} e^{(2\pi i n\alpha + 2\pi i l)T(x-\alpha y, y)}$$

is not periodic in y, even though the function $\psi_{nl}^w(x,y)$ is periodic, but that plays no role. In order to verify that ψ_{nl}^w are not in $H^1(\mathbb{T}^2)$ it suffices to check that the function

$$\Theta_\lambda(x,y) = e^{i\lambda R(x-\alpha y)} = R_\lambda(x-\alpha y)$$

is not in $H^1([0,1]^2)$ for any real $\lambda \neq 0$. Here, $R_\lambda(s)$ is as defined in Proposition 2.10. Since the function $\Theta_\lambda(x,y)$ is constant on the lines

$$x - \alpha y = \text{const},$$

if it were in $H^1([0,1]^2)$, it would force the function $R_\lambda(s)$ to be in $H^1(\mathbb{S}^1)$ and hence continuous. However, R_λ is discontinuous everywhere according to Proposition 2.10. Therefore, the eigenfunctions ψ_{nl}^w cannot be in $H^1(\mathbb{T}^2)$ unless $n = l = 0$.

Finally, to obtain an incompressible flow (with respect to the standard Lebesgue measure) with rough eigenfunctions, we introduce a smooth transformation of the torus

$$\bar{S}: \ (x,y) \to (p,q)$$

by setting

$$p = \int_0^x \bar{F}(s)ds, \quad q = \frac{1}{\bar{F}(x)} \int_0^y F(x,z)dz, \text{ where } \bar{F}(x) = \int_0^1 F(x,z)dz.$$

Note that $\bar{F}(x)$ is periodic, and

$$p(x+1,y) = \int_0^{x+1} \bar{F}(s)ds = p(x,y) + \int_0^1 \bar{F}(s)ds = p(x,y) + 1.$$

We also have $q(x+1,y) = q(x,y)$ and

$$q(x,y+1) = \frac{1}{\bar{F}(x)} \int_0^{y+1} F(x,z)dz = q(x,y) + 1.$$

Therefore, indeed, $\bar{S}$ is a mapping of $\mathbb{T}^2$ to itself. Since $F(x,y)$ is positive, $\bar{S}$ is one-to-one. It is immediate to verify that it maps the measure $d\mu$ onto the Lebesgue measure $dpdq$ – the Jacobian of $\bar{S}$ is $F(x,y)$. Hence, the evolution group generated by the image $u(p,q)$ of the flow $w(x,y)$ will have the same discrete spectrum as U_w. In addition, the eigenfunctions ψ_{nl}^w of U_w are the images of the eigenfunctions ψ_{nl}^u of u under $\bar{S}^*$:

$$\psi_{nl}^w(x,y) = (\bar{S}^* \psi_{nl}^u)(x,y) = \psi_{nl}^u(\bar{S}(x,y)).$$

As the functions ψ_{nl}^w are not in $H^1(\mathbb{T}^2)$ and the map $\bar{S}$ is smooth, it follows that all the eigenfunctions of the incompressible flow $u(p,q)$ are not in $H^1(\mathbb{T}^2)$. This finishes the proof of Proposition 2.9. $\square$

Mixing and weakly mixing flows

An important class of relaxation enhancing flows is given by mixing and weakly mixing flows. Let us recall how they are defined. A flow is mixing if the following condition holds: for any two functions $f, h \in L^2(\Omega)$ we have

$$\lim_{t\to+\infty}\int_\Omega f(X(t;x))h(x)dx = \int_\Omega f(x)dx\int_\Omega h(x)dx. \tag{2.76}$$

The mixing condition (2.76) can be interpreted as follows. Let us start the dynamics

$$\frac{dX}{dt} = -u(X), \quad X(0;x) = x, \tag{2.77}$$

at a random point x, equally distributed over the set Ω. The Lebesgue measure on Ω is invariant under (2.77) since u is incompressible: for any measurable set A we have

$$\mathbb{P}(X(t)\in A) = \int_\Omega \chi_A(X(t;x))dx = \int_\Omega \chi_A(x)dx = |A|.$$

Consider two measurable sets $A \subset \Omega$ and $B \subset \Omega$, and the corresponding characteristic functions $h(x) = \chi_A(x)$ and $f(x) = \chi_B(x)$. Then (2.76) says that

$$P(X(t)\in B \text{ and } X(0)\in A) - |A|\cdot|B| \to 0, \text{ as } t\to+\infty, \tag{2.78}$$

that is, the events $\{X(0) \in A\}$ and $\{X(t) \in B\}$ become asymptotically (as $t \to +\infty$) independent – the fact that you end up in B does not depend on where you start.

We say that a flow u is ergodic if its only first integrals are constant functions. Mixing implies ergodicity: if $U^t f(x) = f(x)$ for all $t \in \mathbb{R}$ then

$$\int_\Omega f(X(t;x))h(x)dx = \int_\Omega f(x)h(x)dx, \text{ for all } t>0, \tag{2.79}$$

for all $h \in L^2(\Omega)$ which is incompatible with mixing unless f is a constant function.

An incompressible flow u is called weakly mixing if the corresponding operators U^t have only continuous spectrum, that is, the only eigenfunctions of U^t are constants. An equivalent definition is that (2.76) holds on average, that is:

$$\lim_{T\to+\infty}\frac{1}{T}\int_0^T\left|\int_\Omega f(X(t;x))h(x)dx - \int_\Omega f(x)dx\int_\Omega h(x)dx\right|dt = 0, \tag{2.80}$$

and the convergence in (2.78) holds for a set of times of density one.

Weakly mixing flows are ergodic: first integrals are eigenfunctions with eigenvalue zero but weakly mixing flows are not necessarily mixing (see, for instance, [6]). On the other hand, mixing flows are weakly mixing: essentially for the same reason that mixing flows are ergodic – if

$$U^t f = c(t)f, \text{ for all } t\in\mathbb{R},$$

then

$$\int_\Omega f(X(t;x))h(x)dx = c(t)\int_\Omega f(x)h(x)dx, \text{ for all } t>0, \tag{2.81}$$

for all $h \in L^2(\Omega)$ which is also incompatible with mixing unless f is a constant function.

A direct consequence of Theorem 2.6 is the following Corollary.

Corollary 2.11 *Any weakly mixing incompressible flow $u \in Lip(\Omega)$ is relaxation enhancing.*

There exist fairly explicit examples of weakly mixing flows [2, 15, 16, 24, 26, 29], some of which we briefly describe below but delving into the detailed constructions would take us too far outside of the PDE realm.

Weakly mixing incompressible flows on a torus

According to Theorem 2.6, a flow $u \in \mathrm{Lip}(\Omega)$ is relaxation enhancing if and only if it has no eigenfunctions in $H^1(\Omega)$. A natural class satisfying this condition is weakly mixing flows – they have no eigenfunctions in $L^2(\Omega)$ at all. Examples of weakly mixing flows on $\mathbb{T}^2$ go back to von Neumann [26] and Kolmogorov [24]. The flow in von Neumann's example is continuous, in the construction suggested by Kolmogorov the flow is smooth. The technical details of Kolmogorov's construction have been carried out in [29], a good review of these results is [20]. More recently, Fayad [15] generalized this example to show that weakly mixing flows are generic. To describe the result of [15] in more detail, let us recall that a vector $\alpha \in \mathbb{R}^n$ is called β-Diophatine if there exists a constant C such that for each $k \in \mathbb{Z}^n \setminus \{0\}$ we have

$$\inf_{p\in\mathbb{Z}} |\langle \alpha, k\rangle + p| \geq \frac{C}{|k|^{n+\beta}}.$$

The vector α is Liouvillean if it is not Diophantine for any $\beta > 0$. The Liouvillean numbers (and vectors) are the ones which can be very well approximated by rationals.

In order to construct a weakly mixing flow on a torus $\mathbb{T}^{n+1}$ we again start with a very simple unidrectional flow that is a local time change of a linear translation flow:

$$\frac{dX}{dt} = \frac{\alpha}{F(X,Y)}, \quad \frac{dY}{dt} = \frac{1}{F(X,Y)}, \quad X(0) = x, \quad Y(0) = y, \tag{2.82}$$

with a smooth positive function $F(x, y)$, $x \in \mathbb{T}^n$, $y \in \mathbb{T}$. We have seen that such flows have a unique invariant measure

$$d\mu = F(x, y)dxdy.$$

Given a smooth function $f(x, y)$, set

$$g(t, x, y) = U^t f(x, y) = f(X(t, x, y), Y(t, x, y)).$$

This function satisfies the first order advection equation

$$\frac{\partial g}{\partial t} - \frac{\alpha}{F(x,y)} \cdot \nabla_x g - \frac{1}{F(x,y)}\frac{\partial g}{\partial y} = 0, \quad g(0, x, y) = f(x, y). \tag{2.83}$$

Let us denote by $C^r_+(\mathbb{T}^d)$ the set of positive C^r functions on the torus. Fayad's result is

Proposition 2.12 *([15]) Assume that the irrational vector $\alpha \in \mathbb{R}^d$ is not β-Diophantine, for some $\beta > 0$. Then, for a dense G_δ set of functions F in $C^{\beta+n}_+(\mathbb{T}^{n+1})$ the flow (2.82) is weakly mixing (for the unique invariant measure $F(x, y)dxdy$).*

The flows given by this proposition have an invariant measure $F(x, y)dxdy$ and not the Lebesgue measure $dxdy$, but as in the Kolmogorov example of a flow with rough eigenfunctions, it is easy to produce a flow that preserves the Lebesgue measure.

An abstract criterion for relaxation enhancement

Theorem 2.6 follows from a rather general abstract criterion, which connects us back to the abstract set-up of (1.1). We start with a self-adjoint, positive, unbounded operator Γ with a discrete spectrum on a separable Hilbert space H. In the setting of Theorem 2.6, Γ corresponds to $-\Delta$, with H the subspace of mean zero functions on $L^2(\Omega)$. We denote by

$$0 < \lambda_1 \leq \lambda_2 \leq \dots$$

the eigenvalues of Γ, and by e_j the corresponding orthonormal eigenvectors forming a basis in H. The (homogenous) Sobolev space $H^m(\Gamma)$ associated with Γ is formed by all vectors

$$\psi = \sum_j c_j e_j \in H,$$

such that

$$\|\psi\|^2_{H^m(\Gamma)} \equiv \sum_j \lambda_j^m |c_j|^2 < \infty.$$

The crucial assumption is that $\lambda_n \to +\infty$ – this makes the set where the dissipation by Γ is not too large a compact subset of H. Recall that part of the definition of the discrete spectrum includes the assumption that each λ_j has a finite multiplicity

We will denote the norm in the Hilbert space H by $\|\cdot\|$, the inner product in H by $\langle \cdot, \cdot \rangle$, the Sobolev spaces $H^m(\Gamma)$ simply by H^m and the norms in these Sobolev spaces by $\|\cdot\|_m$. Note that $H^2(\Gamma)$ is the domain $D(\Gamma)$ of Γ.

The analog of the operator $u \cdot \nabla$ in Theorem 2.6 (or, to be precise, of the self-adjoint operator generating the unitary evolution group U^t which is equal to $iu \cdot \nabla$ on $H^1(\Omega)$) is a self-adjoint operator L such that, for any $\psi \in H^1(\Gamma)$ and $t > 0$ we have

$$\|L\psi\|_H \leq C\|\psi\|_{H^1(\Gamma)} \text{ and } \|e^{iLt}\psi\|_{H^1(\Gamma)} \leq B(t)\|\psi\|_{H^1(\Gamma)} \tag{2.84}$$

with both the constant C and the function $B(t) < \infty$ independent of ψ and $B(t) \in L^2_{\text{loc}}[0, \infty)$. Here, e^{iLt} is the unitary evolution group generated by the self-adjoint operator L. It is easy to see that (2.84) holds when $L = iu(x) \cdot \nabla$ and $\Gamma = -\Delta$ on the n-dimensional torus $\mathbb{T}^n$, as long as $u(x)$ is Lipschitz.

Exercise 2.13 Show that neither of the two conditions in (2.84) implies the other.

Consider a solution $\phi^\varepsilon(t)$ of the rescaled in time version of (1.1):

$$\frac{d}{dt}\phi^\varepsilon(t) = \frac{i}{\varepsilon}L\phi^\varepsilon(t) - \Gamma\phi^\varepsilon(t), \quad \phi^\varepsilon(0) = \phi_0. \tag{2.85}$$

We have the following existence and uniqueness result.

Proposition 2.14 *Assume that for any $\psi \in H^1$, we have*

$$\|L\psi\| \leq C\|\psi\|_1. \tag{2.86}$$

Then for any $T > 0$, there exists a unique solution $\phi(t)$ of the equation

$$\frac{d\phi(t)}{dt} = (iL - \Gamma)\phi(t), \quad \phi(0) = \phi_0 \in H^1.$$

This solution satisfies

$$\phi(t) \in L^2([0,T], H^2) \cap C([0,T], H^1), \;\; \dot{\phi}(t) \in L^2([0,T], H). \tag{2.87}$$

Exercise 2.15 Proposition 2.14 can be proved by standard methods using Galerkin approximations and then establishing uniqueness and regularity. Fill in the details of the argument.

The main result of this section is the following abstract criterion for relaxation enhancement.

Theorem 2.16 *Let Γ be a self-adjoint, positive, unbounded operator with a discrete spectrum, and let a self-adjoint operator L satisfy (2.84). Then the following two statements are equivalent:*

(i) For any $\tau, \delta > 0$ there exists $\varepsilon_0(\tau, \delta)$ such that for any $0 < \varepsilon < \varepsilon_0(\tau, \delta)$ and any $\phi_0 \in H$ with $\|\phi_0\|_H = 1$, the solution $\phi^\varepsilon(t)$ of (2.85) satisfies $\|\phi^\varepsilon(\tau)\|_H^2 < \delta$.

(ii) The operator L has no eigenvectors in $H^1(\Gamma)$.

Theorem 2.16 provides an answer to the general question of when a combination of fast unitary evolution and dissipation produces a significantly stronger dissipative effect than dissipation alone. It can be useful in any model describing a physical situation which involves fast unitary dynamics with dissipation (or, equivalently, unitary dynamics with weak dissipation).

The proof of Theorem 2.16

Eigenvectors in $H^1(\Gamma)$ preclude relaxation enhancement

One direction in the proof of Theorem 2.16 is much easier: existence of $H^1(\Gamma)$ eigenvectors of the operator L ensures existence of $\tau_0, \delta_0 > 0$ and ϕ_0 with $\|\phi_0\| = 1$ such that $\|\phi^\varepsilon(\tau_0)\| > \delta_0$ for all ε – that is, if such eigenvectors exist, then the operator L is not relaxation enhancing.

Assume that the initial condition $\phi_0 \in H^1$ for

$$\frac{d}{dt}\phi^\varepsilon(t) = \frac{i}{\varepsilon}L\phi^\varepsilon(t) - \Gamma\phi^\varepsilon(t), \quad \phi^\varepsilon(0) = \phi_0 \tag{2.88}$$

is an eigenvector of L corresponding to an eigenvalue E, normalized so that $\|\phi_0\| = 1$:

$$L\phi_0 = E\phi_0. \tag{2.89}$$

Take the inner product of (2.88) with ϕ_0. We arrive at

$$\frac{d}{dt}\langle\phi^\varepsilon(t), \phi_0\rangle = \frac{iE}{\varepsilon}\langle\phi^\varepsilon(t), \phi_0\rangle - \langle\Gamma\phi^\varepsilon(t), \phi_0\rangle.$$

This and the assumption $\phi_0 \in H^1$ lead to

$$\left|\frac{d}{dt}\left(e^{-iEt/\varepsilon}\langle\phi^\varepsilon(t),\phi_0\rangle\right)\right| \le \frac{1}{2}\left(\|\phi^\varepsilon(t)\|_1^2 + \|\phi_0\|_1^2\right). \tag{2.90}$$

The H^1-norm in the right side of (2.90) is estimated by the following standard dissipation lemma.

Lemma 2.17 *Assume that* (2.86) *holds, then for any initial condition $\phi_0 \in H$ with $\|\phi_0\| = 1$, the solution $\phi^\varepsilon(t)$ of* (2.88) *satisfies*

$$\int_0^\infty \|\phi^\varepsilon(t)\|_1^2 dt \le \frac{1}{2}. \tag{2.91}$$

Proof. Recall that if $\phi \in H^1(\Gamma)$, then $\Gamma\phi \in H^{-1}(\Gamma)$ and $\langle\Gamma\phi,\phi\rangle = \|\phi\|_1^2$. The fact that L is self-adjoint allows us to compute

$$\frac{d}{dt}\|\phi^\varepsilon\|^2 = \langle\phi^\varepsilon,\phi_t^\varepsilon\rangle + \langle\phi_t^\varepsilon,\phi^\varepsilon\rangle = -2\|\phi^\varepsilon\|_1^2. \tag{2.92}$$

Integrating in time and taking into account the normalization of ϕ_0, we obtain (2.91). □

Going back to (2.90), integrating in time and using (2.91) gives

$$|\langle\phi^\varepsilon(t),\phi_0\rangle| \ge \|\phi_0\|^2 - \frac{t}{2}\|\phi_0\|_1^2 - \frac{1}{2}\int_0^\infty \|\phi^\varepsilon(t)\|_1^2 dt \ge \frac{1}{2} \quad \text{for } 0 \le t \le \tau = \frac{1}{27}\|\phi_0\|_1^2. \tag{2.93}$$

Thus, we have $\|\phi^\varepsilon(\tau)\| \ge 1/2$, uniformly in ε.

Outline of the argument

Even though the proof of the other direction in Theorem 2.16 is slightly technical, the general idea of the proof is completely straightforward and has two intuitive ingredients. As in Lemma 2.17, the dissipation balance for (2.85) is

$$\frac{1}{2}\frac{d}{dt}(\|\phi^\varepsilon\|_H^2) = -\langle\Gamma\phi^\varepsilon,\phi^\varepsilon\rangle = -\|\phi^\varepsilon(t)\|_{H^1(\Gamma)}^2. \tag{2.94}$$

Therefore, if $\|\phi\|_{H^1(\Gamma)}$ is large on a time interval $[t_1,t_2]$ then $\|\phi^\varepsilon(t)\|_H$ will drop significantly over this time interval. On the other hand, we will show that if $\|\phi^\varepsilon(\tau)\|_{H^1(\Gamma)}$ is small at some time τ then, because L does not have $H^1(\Gamma)$-eigenfunctions, the free evolution

$$\frac{d\phi^0}{dt} = \frac{i}{\varepsilon}L\phi^0, \quad \phi^0(\tau) = \phi^\varepsilon(\tau), \quad t \ge \tau, \tag{2.95}$$

will make the $H^1(\Gamma)$-norm of ϕ^0 very large in a time so short that the free evolution is close to the true evolution over this short time interval. This means that the $H^1(\Gamma)$-norm of the solution $\phi^\varepsilon(t)$ to the true problem (2.85) will also be very large. Hence, even if the $H^1(\Gamma)$-norm of ϕ^ε drops, it will go back up again very quickly, forcing the dissipation to be large most of the time, and reducing $\|\phi^\varepsilon(t)\|_H$ very efficiently. Making this argument careful will take us some time, no pun intended. A crucial role is played by the fact that the unit ball in the $H^1(\Gamma)$-norm is compact in H.

It will be more convenient, in terms of notation, to rescale the time back by the factor ε^{-1}, arriving at the equation

$$\frac{d\tilde{\phi}^\varepsilon(t)}{dt} = (iL - \varepsilon\Gamma)\tilde{\phi}^\varepsilon(t), \;\; \tilde{\phi}^\varepsilon(0) = \phi_0. \tag{2.96}$$

The dissipation estimate in Lemma 2.17 translates into

$$\varepsilon \int_0^\infty \|\tilde{\phi}^\varepsilon(t)\|_1^2 dt \le \frac{1}{2}. \tag{2.97}$$

A slight generalization of (2.97) is the following simple estimate, that we state here as a separate lemma for convenience. This quantifies the idea that if the H^1-norm of $\tilde{\phi}^\varepsilon(t)$ stays large on a time interval $[a,b]$, then the H-norm of $\tilde{\phi}^\varepsilon$ drops on that interval.

Lemma 2.18 *Suppose that for all times $t \in (a,b)$ we have $\|\tilde{\phi}^\varepsilon(t)\|_1^2 \ge N\|\tilde{\phi}^\varepsilon(t)\|^2$. Then the following decay estimate holds:*

$$\|\tilde{\phi}^\varepsilon(b)\|^2 \le e^{-2\varepsilon N(b-a)}\|\tilde{\phi}^\varepsilon(a)\|^2.$$

Next we need an estimate on the growth of the difference between solutions corresponding to $\varepsilon > 0$ and $\varepsilon = 0$ in the H-norm. This will be crucial when we show that if the H^1-norm of $\tilde{\phi}^\varepsilon(t_0)$ is small at some time t_0, it will have to become large again quickly,

Lemma 2.19 *Assume, in addition to (2.86), that for any $\psi \in H^1$ and $t > 0$ we have*

$$\|e^{iLt}\psi\|_1 \le B(t)\|\psi\|_1 \tag{2.98}$$

for some $B(t) \in L^2_{\mathrm{loc}}[0,\infty)$. Let $\phi^0(t), \phi^\varepsilon(t)$ be solutions of

$$\frac{d\phi^0(t)}{dt} = iL\phi^0(t), \;\; \frac{d\tilde{\phi}^\varepsilon(t)}{dt} = (iL - \varepsilon\Gamma)\tilde{\phi}^\varepsilon(t),$$

satisfying $\phi^0(0) = \phi^\varepsilon(0) = \phi_0 \in H^1$. Then we have

$$\frac{d}{dt}\|\tilde{\phi}^\varepsilon(t) - \phi^0(t)\|^2 \le \frac{1}{2}\varepsilon\|\phi^0(t)\|_1^2 \le \frac{1}{2}\varepsilon B^2(t)\|\phi_0\|_1^2. \tag{2.99}$$

Proof. Note that $\phi^0(t) = e^{iLt}\phi_0$ by definition. Assumption (2.98) says that this unitary evolution is bounded in the $H^1(\Gamma)$ norm. The regularity guaranteed by conditions (2.86), (2.98) and Proposition 2.14 allows us to multiply the equation

$$\frac{d}{dt}(\tilde{\phi}^\varepsilon - \phi^0) = iL(\tilde{\phi}^\varepsilon - \phi^0) - \varepsilon\Gamma\tilde{\phi}^\varepsilon$$

by $\tilde{\phi}^\varepsilon - \phi^0$. We obtain

$$\frac{d}{dt}\|\tilde{\phi}^\varepsilon - \phi^0\|^2 \le 2\varepsilon(\|\tilde{\phi}^\varepsilon\|_1\|\phi^0\|_1 - \|\tilde{\phi}^\varepsilon\|_1^2) \le \frac{1}{2}\varepsilon\|\phi^0\|_1^2,$$

which is the first inequality in (2.99). The second inequality follows simply from the assumption (2.98). □

The following corollary is immediate.

Corollary 2.20 *Assume that* (2.86) *and* (2.98) *are satisfied, and the initial data* $\phi_0 \in H^1$. *Then the solutions* $\tilde{\phi}^\varepsilon(t)$ *and* $\phi^0(t)$ *defined in Lemma 2.19 satisfy*

$$\|\tilde{\phi}^\varepsilon(t) - \phi^0(t)\|^2 \leq \frac{1}{2}\varepsilon\|\phi_0\|_1^2 \int_0^\tau B^2(s)\, ds$$

for any time $t \leq \tau$.

We will now switch to the equivalent formulation (2.96), and drop the tilde (hoping that this will not cause any confusion).

The RAGE theorem and the time spent in high modes

Our first task is to get good control of the free evolution e^{iLt}, and show that, the H^1 norm of the solution to the free evolution problem can not stay small for too long. The first ingredient that we need to recall is the so-called RAGE theorem, first proved by Ruelle [28] and later generalized by Amrein and Georgescu in [1], and by Enss in [10].

Theorem 2.21 (RAGE) *Let* L *be a self-adjoint operator in a Hilbert space* H. *Let* P_c *be the spectral projection on its continuous spectral subspace, and* $\mathcal{C}$ *be any compact operator. Then for any* $\phi_0 \in H$, *we have*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \|\mathcal{C}e^{iLt}P_c\phi_0\|^2\, dt = 0. \tag{2.100}$$

The proof of the RAGE theorem can be found, for example, in [7]. A very naive but not altogether useless way to think about the RAGE theorem is to take $\mathcal{C}$ as a projection on a finite-dimensional space K. In that case, if (2.100) fails, then the operator L approximately leaves K invariant, as the vector $e^{iLt}\psi_0$, with $\psi_0 = P_c\phi_0$, has a non-trivial component in K for a non-trivial fraction of time. Hence, it is plausible that one could find an eigenvector in K as some sort of a limit of $e^{iLt}\psi_0$, which is not quite compatible with the fact that the initial condition ψ_0 lives in the continuous spectrum part of L.

An analyst perspective is that the RAGE theorem is a generalization of the following classical theorem by Wiener.

Theorem 2.22 *Let* $d\mu$ *be a finite measure on* $\mathbb{R}$ *with the Fourier transform*

$$F(t) = \int_{\mathbb{R}} e^{-ixt} d\mu(x).$$

Then

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T |F(t)|^2 dt = \sum_{x\in\mathbb{R}} |\mu(\{x\})|^2. \tag{2.101}$$

Note that the sum in the right side of (2.101) is finite since μ is a finite measure.

A direct consequence of the RAGE theorem is the following lemma. Recall that we denote by $0 < \lambda_1 \leq \lambda_2 \leq \dots$ the eigenvalues of the operator Γ and by $e_1, e_2, \dots$ the corresponding orthonormal eigenvectors, and that each λ_j has a finite multiplicity. Let us denote by P_N the orthogonal projection on the subspace spanned by the first N eigenvectors $e_1, \dots, e_N$ and by

$$S = \{\phi \in H : \ \|\phi\| = 1\}$$

the unit sphere in H. The following lemma shows that if the initial condition lies in the continuous spectrum of L, then the L-evolution will spend most of time in the higher modes of Γ.

Lemma 2.23 *Let $K \subset S$ be a compact set. For any $N, \sigma > 0$, there exists $T_c(N, \sigma, K)$ such that for all $T \geq T_c(N, \sigma, K)$ and any $\phi \in K$, we have*

$$\frac{1}{T}\int_0^T \|P_N e^{iLt} P_c \phi\|^2 \, dt \leq \sigma. \tag{2.102}$$

The key observation of Lemma 2.23 is that the time $T_c(N, \sigma, K)$ is uniform for all $\phi \in K$.

Proof. Since P_N is compact, RAGE theorem implies that for any vector $\phi \in S$, there exists a time $T_c(N, \sigma, \phi)$ that depends on the function ϕ such that (2.102) holds for $T > T_c(N, \sigma, \phi)$. To prove the uniformity in ϕ, note that the function

$$f(T, \phi) = \frac{1}{T}\int_0^T \|P_N e^{iLt} P_c \phi\|^2 \, dt$$

is uniformly continuous on S for all T (with constants independent of T):

$$|f(T,\phi) - f(T,\psi)| \leq \frac{1}{T}\int_0^T \left| \|P_N e^{iLt} P_c \phi\| - \|P_N e^{iLt} P_c \psi\| \right| \left(\|P_N e^{iLt} P_c \phi\| + \|P_N e^{iLt} P_c \psi\| \right) dt$$

$$\leq (\|\phi\| + \|\psi\|) \frac{1}{T}\int_0^T \|P_N e^{iLt} P_c(\phi - \psi)\| dt \leq 2\|\phi - \psi\|.$$

Now, existence of a uniform $T_c(N, \sigma, K)$ follows from compactness of K by standard arguments. □

The H^1-norm of free solutions with rough eigenfunctions

We also need a lemma which controls from below the growth of the H^1 norm of free solutions corresponding to rough eigenfunctions. We denote by P_p the spectral projection on the pure point spectrum of the operator L.

Lemma 2.24 *Assume that no eigenvectors of the operator L belong to $H^1(\Gamma)$. Let $K \subset S$ be a compact set, and $K_1 = \{\phi \in K : \ \|P_p \phi\| \geq 1/2\}$. Then for any $B > 0$ we can find $N_p(B, K)$ and $T_p(B, K)$ such that for any $N \geq N_p(B, K)$, any $T \geq T_p(B, K)$ and any $\phi \in K_1$, we have*

$$\frac{1}{T}\int_0^T \|P_N e^{iLt} P_p \phi\|_1^2 \, dt \geq B. \tag{2.103}$$

Note that unlike in (2.102), we have the H^1 norm in (2.103).

Proof. The set K_1 may be empty, in which case there is nothing to prove. Otherwise, let us denote by E_j the eigenvalues of L (distinct, without repetitions) and by Q_j the orthogonal projection on the space spanned by the eigenfunctions corresponding to E_j. First, let us show that for any $B > 0$ there is $N(B, K)$ such that for any $\phi \in K_1$ we have

$$\sum_j \|P_N Q_j \phi\|_1^2 \geq 2B \tag{2.104}$$

if $N \geq N(B, K)$. It is clear that for each fixed ϕ with $P_p\phi \neq 0$ we can find $N(B, \phi)$ so that (2.104) holds, since by assumption $Q_j\phi$ does not belong to H^1 whenever $Q_j\phi \neq 0$. Assume that $N(B, \phi)$ cannot be chosen uniformly for $\phi \in K_1$. This means that for any n, there exists $\phi_n \in K_1$ such that

$$\sum_j \|P_n Q_j \phi_n\|_1^2 < 2B.$$

Since K_1 is compact, we can find a subsequence n_l such that ϕ_{n_l} converges to $\bar{\phi} \in K_1$ in H as $n_l \to \infty$. For any N and any $n_{l_1} > N$ we have

$$\sum_j \|P_N Q_j \bar{\phi}\|_1^2 \leq \sum_j \|P_{n_{l_1}} Q_j \bar{\phi}\|_1^2 \leq \liminf_{l\to\infty} \sum_j \|P_{n_{l_1}} Q_j \phi_{n_l}\|_1^2. \tag{2.105}$$

The last inequality follows by Fatou's Lemma from the convergence of ϕ_{n_l} to $\bar{\phi}$ in H and the fact that

$$\|P_{n_{l_1}} Q_j \psi\|_1 = \lambda_{n_{l_1}}^{1/2} \|Q_j \psi\| \leq \lambda_{n_{l_1}}^{1/2} \|\psi\|,$$

for any n_{l_1}. The expression in the right hand side of (2.105) is less than or equal to

$$\liminf_{l\to\infty} \sum_j \|P_{n_l} Q_j \phi_{n_l}\|_1^2 \leq 2B.$$

Thus, we have

$$\sum_j \|P_N Q_j \bar{\phi}\|_1^2 \leq 2B \text{ for any } N,$$

a contradiction since $\bar{\phi} \in K_1$. As a consequence, there exists $N(B, K)$ so that (2.104) holds for all $N \geq N(B, K)$ and all $\phi \in K_1$.

Next, take $\phi \in K_1$ and consider

$$\frac{1}{T} \int_0^T \|P_N e^{iLt} P_p \phi\|_1^2 \, dt = \sum_{j,l} \frac{e^{i(E_j - E_l)T} - 1}{i(E_j - E_l)T} \langle \Gamma P_N Q_j \phi, P_N Q_l \phi \rangle. \tag{2.106}$$

In (2.106), we set

$$\frac{e^{i(E_j - E_l)T} - 1}{i(E_j - E_l)T} \equiv 1 \text{ if } j = l.$$

Notice that the sum in (2.106) converges absolutely. Indeed,

$$P_N Q_j \phi = \sum_{i=1}^{N} \langle Q_j \phi, e_i \rangle e_i,$$

and $\langle \Gamma e_i, e_k\rangle = \lambda_i \delta_{ik}$, therefore

$$\langle \Gamma P_N Q_j \phi, P_N Q_l \phi\rangle = \sum_{i=1}^{N} \lambda_i \langle Q_j\phi, e_i\rangle \overline{\langle Q_l \phi, e_i\rangle}.$$

Hence, the sum in the right side of (2.106) does not exceed

$$\begin{aligned}\sum_{i=1}^{N} \lambda_i \sum_{j,l} |\langle Q_j\phi, e_i\rangle\langle Q_l\phi, e_i\rangle| &\le \lambda_N \sum_{i=1}^{N}\sum_{j,l} \|Q_j\phi\|\|Q_l\phi\||\langle Q_j\phi/\|Q_j\phi\|, e_i\rangle\langle Q_l\phi/\|Q_l\phi\|, e_i\rangle| \\ &\le \lambda_N \sum_{i=1}^{N}\sum_{j,l} \|Q_l\phi\|^2 |\langle Q_j\phi/\|Q_j\phi\|, e_i\rangle|^2 \le \lambda_N N. \qquad (2.107)\end{aligned}$$

The second step above is obtained using the Cauchy-Schwartz inequality, and the third since $\|\phi\| = \|e_i\| = 1$. Then for each fixed N, it follows from the dominated convergence theorem that the expression in (2.106) converges to

$$\sum_j \|\Gamma^{1/2} P_N Q_j \phi\|^2 = \sum_j \|P_N Q_j\phi\|_1^2$$

as $T \to \infty$.

Now assume $N \ge N_p(B, K) \equiv N(B, K)$, so that (2.104) holds. We claim that we can choose $T_p(B, K)$ so that for any $T \ge T_p(B, K)$ we have

$$\left|\frac{1}{T}\int_0^T \|P_N e^{iLt} P_p \phi\|_1^2\, dt - \sum_j \|P_N Q_j \phi\|_1^2\right| = \left|\sum_{l\ne j} \frac{e^{i(E_j - E_l)T} - 1}{i(E_j - E_l)T} \langle \Gamma P_N Q_j\phi, P_N Q_l \phi\rangle\right| \le B \qquad (2.108)$$

for all $\phi \in K_1$. Indeed, this follows from convergence to zero for each individual ϕ as $T \to \infty$, compactness of K_1, and uniform continuity of the expression in the middle of (2.108) in ϕ for each T (with constants independent of T). The latter is proved by estimating the difference of these expressions for $\phi, \psi \in K_1$ and any T by

$$\sum_{l\ne j} |\langle \Gamma P_N Q_j \phi, P_N Q_l(\phi - \psi)\rangle| + |\langle \Gamma P_N Q_j(\phi - \psi), P_N Q_l \psi\rangle|,$$

which is then bounded by $2\lambda_N N\|\phi - \psi\|$ using the trick from (2.107). Combining (2.104) and (2.108) proves the lemma. □

The end of the proof of Theorem 2.16

We can now complete the proof of Theorem 2.16. Recall that given any $\tau, \delta > 0$, we need to show the existence of $\varepsilon_0 > 0$ such that if $\varepsilon < \varepsilon_0$, then solution of the rescaled problem

$$\frac{d\phi^\varepsilon(t)}{dt} = (iL - \varepsilon\Gamma)\phi^\varepsilon(t), \;\; \tilde{\phi}^\varepsilon(0) = \phi_0. \qquad (2.109)$$

satisfies $\|\phi^\varepsilon(\tau/\varepsilon)\| < \delta$ for any initial condition $\phi_0 \in H$, $\|\phi_0\| = 1$. Let us recall again the idea of the proof. Lemma 2.18 tells us that if the H^1 norm of the solution $\phi^\varepsilon(t)$ is large, relaxation is happening quickly. If, on the other hand, $\|\phi^\varepsilon(\tau_0)\|_1^2 \le \lambda_M \|\phi^\varepsilon(\tau_0)\|^2$, where M is to be chosen depending on τ and δ, then the set of all unit vectors satisfying this inequality is compact, and so we can apply Lemma 2.23 and Lemma 2.24. Using these lemmas, we will show that even if the H^1 norm is small at some moment of time τ_0, it will be large on average in some time interval after τ_0. Enhanced relaxation will follow.

We now provide the details. Since Γ is an unbounded positive operator with a discrete spectrum, we know that its eigenvalues $\lambda_n \to \infty$ as $n \to \infty$. Let us choose M large enough, so that

$$e^{-\lambda_M \tau/80} < \delta.$$

Define the sets

$$K = \{\phi \in S : \ \|\phi\|_1^2 \le \lambda_M\} \subset S,$$

and as before,

$$K_1 = \{\phi \in K : \ \|P_p \phi\| \ge 1/2\}.$$

It is easy to see that K is compact. Choose N so that $N \ge M$ and $N \ge N_p(5\lambda_M, K)$ from Lemma 2.24. Define

$$\tau_1 \equiv \max\left\{T_p(5\lambda_M, K), T_c(N, \tfrac{\lambda_M}{20\lambda_N}, K)\right\},$$

with T_p from Lemma 2.24, and T_c from Lemma 2.23. Finally, choose $\varepsilon_0 > 0$ so that $\tau_1 < \tau/2\varepsilon_0$, and

$$\varepsilon_0 \int_0^{\tau_1} B^2(t)\, dt \le \frac{1}{20\lambda_N}, \tag{2.110}$$

where $B(t)$ is the function from condition (2.98).

Take any $\varepsilon < \varepsilon_0$. If we have

$$\|\phi^\varepsilon(s)\|_1^2 \ge \lambda_M \|\phi^\varepsilon(s)\|^2$$

for all $s \in [0, \tau/\varepsilon]$, then Lemma 2.18 implies that

$$\|\phi^\varepsilon(\tau/\varepsilon)\| \le e^{-2\lambda_M \tau} \le \delta,$$

by the choice of M and we are done. Otherwise, let τ_0 be a time in the interval $[0, \tau/\varepsilon]$ such that

$$\|\phi^\varepsilon(\tau_0)\|_1^2 \le \lambda_M \|\phi^\varepsilon(\tau_0)\|^2. \tag{2.111}$$

It may be that $\tau_0 = 0$, of course. Here is a key lemma that reflects the idea that the H^1-norm of the solution can not stay small for too long: even if it is small at a time τ_0 then the dissipation on the the time interval $[\tau_0, \tau_0 + \tau_1]$ will not be small.

Lemma 2.25 *If (2.111) holds at τ_0, then*

$$\|\phi^\varepsilon(\tau_0 + \tau_1)\|^2 \le e^{-\lambda_M \varepsilon \tau_1/20} \|\phi^\varepsilon(\tau_0)\|^2. \tag{2.112}$$

Before we prove Lemma 2.25, we explain how it implies the conclusion of Theorem 2.16 for the case when L has no eigenfunctions in $H^1(\Gamma)$. First, if (2.111) holds then we have (2.112). On the other hand, for any interval $I = [a, b]$ such that

$$\|\phi^\varepsilon(t)\|_1^2 \ge \lambda_M \|\phi^\varepsilon(t)\|^2 \text{ for all } a \le t \le b,$$

we have by Lemma 2.18 that

$$\|\phi^\varepsilon(b)\|^2 \le e^{-2\lambda_M \varepsilon(b-a)}\|\phi^\varepsilon(a)\|^2. \tag{2.113}$$

Combining the decay factors gained from (2.112) and (2.113), and since $\tau_1 < \tau/(2\varepsilon)$, we find that there is $\tau_2 \in [\tau/(2\varepsilon), \tau/\varepsilon]$ such that

$$\|\phi^\varepsilon(\tau_2)\|^2 \le e^{-\lambda_M \tau/40} < \delta^2$$

by our choice of M. Then (2.92) gives $\|\phi^\varepsilon(\tau/\varepsilon)\| \le \|\phi^\varepsilon(\tau_2)\| < \delta$, finishing the proof of Theorem 2.16, except for the proof of Lemma 2.25. □

The proof of Lemma 2.25: tracking the full dynamics with free evolution

For the sake of simplicity, we will denote $\phi^\varepsilon(\tau_0) = \phi_0$. On the interval $[\tau_0, \tau_0 + \tau_1]$, consider the function $\phi^0(t)$ satisfying

$$\frac{d}{dt}\phi^0(t) = iL\phi^0(t), \quad \phi^0(\tau_0) = \phi_0.$$

Note that by the choice of ε_0, (2.110), (2.111), and Corollary 2.20, we have

$$\|\phi^\varepsilon(t) - \phi^0(t)\|^2 \le \frac{\lambda_M}{40\lambda_N}\|\phi_0\|^2 \tag{2.114}$$

for all $t \in [\tau_0, \tau_0 + \tau_1]$. Split

$$\phi^0(t) = \phi_c(t) + \phi_p(t),$$

where ϕ_c and ϕ_p also solve the free equation

$$\frac{d}{dt}\phi_{c,p}(t) = iL\phi_{c,p}(t),$$

but with initial conditions $P_c\phi_0$ and $P_p\phi_0$ at $t = \tau_0$, respectively. We will now consider two cases.

Case I. Assume that

$$\|P_c\phi_0\|^2 \ge \frac{3}{4}\|\phi_0\|^2,$$

or, equivalently, $\|P_p\phi_0\|^2 \le (1/4)\|\phi_0\|^2$. Note that since $\phi_0/\|\phi_0\| \in K$ by the definition of τ_0, we can apply Lemma 2.23. Our choice of τ_1 implies that

$$\frac{1}{\tau_1}\int_{\tau_0}^{\tau_0+\tau_1} \|P_N\phi_c(t)\|^2\,dt \le \frac{\lambda_M}{20\lambda_N}\|\phi_0\|^2. \tag{2.115}$$

By elementary considerations, we obtain

$$\|(I-P_N)\phi^0(t)\|^2 \ge \frac{1}{2}\|(I-P_N)\phi_c(t)\|^2 - \|(I-P_N)\phi_p(t)\|^2 \ge \frac{1}{2}\|\phi_c(t)\|^2 - \frac{1}{2}\|P_N\phi_c(t)\|^2 - \|\phi_p(t)\|^2.$$

Taking into account the fact that the free evolution e^{iLt} is unitary, $\lambda_N \geq \lambda_M$, our assumptions on $\|P_{c,p}\phi_0\|$ and (2.115), we obtain

$$\frac{1}{\tau_1}\int_{\tau_0}^{\tau_0+\tau_1} \|(I-P_N)\phi^0(t)\|^2\,dt \geq \frac{1}{10}\|\phi_0\|^2. \tag{2.116}$$

Using (2.114), we conclude that

$$\frac{1}{\tau_1}\int_{\tau_0}^{\tau_0+\tau_1} \|(I-P_N)\phi^\varepsilon(t)\|^2\,dt \geq \frac{1}{40}\|\phi_0\|^2. \tag{2.117}$$

This estimate implies that

$$\int_{\tau_0}^{\tau_0+\tau_1} \|\phi^\varepsilon(t)\|_1^2\,dt \geq \frac{\lambda_N\tau_1}{40}\|\phi_0\|^2. \tag{2.118}$$

Combining (2.118) with (2.92) yields

$$\|\phi^\varepsilon(\tau_0+\tau_1)\|^2 \leq \left(1-\frac{\lambda_N\varepsilon\tau_1}{20}\right)\|\phi^\varepsilon(\tau_0)\|^2 \leq e^{-\lambda_N\varepsilon\tau_1/20}\|\phi^\varepsilon(\tau_0)\|^2. \tag{2.119}$$

This finishes the proof of (2.112) in the first case since $\lambda_N \geq \lambda_M$.

Case II. Now suppose that $\|P_p\phi_0\|^2 \geq (1/4)\|\phi_0\|^2$. In this case $\phi_0/\|\phi_0\| \in K_1$, and we can apply Lemma 2.24. In particular, by the choice of N and τ_1, we have

$$\frac{1}{\tau_1}\int_{\tau_0}^{\tau_0+\tau_1} \|P_N\phi_p(t)\|_1^2\,dt \geq 5\lambda_M\|\phi_0\|^2. \tag{2.120}$$

Since (2.115) still holds because of our choice of τ_0 and τ_1, it follows that

$$\frac{1}{\tau_1}\int_{\tau_0}^{\tau_0+\tau_1} \|P_N\phi_c(t)\|_1^2\,dt \leq \frac{\lambda_M}{20}\|\phi_0\|^2. \tag{2.121}$$

Note that the H-norm in (2.115) has been replaced in (2.121) by the H^1-norm at the expense of the factor of λ_N. Together, (2.120) and (2.121) imply

$$\frac{1}{\tau_1}\int_{\tau_0}^{\tau_0+\tau_1} \|P_N\phi^0(t)\|_1^2\,dt \geq 2\lambda_M\|\phi_0\|^2. \tag{2.122}$$

Finally, (2.122) and (2.114) give

$$\int_{\tau_0}^{\tau_0+\tau_1} \|P_N\phi^\varepsilon(t)\|_1^2\,dt \geq \frac{\lambda_M\tau_1}{2}\|\phi_0\|^2 \tag{2.123}$$

since $\|P_N\phi^\varepsilon - P_N\phi^0\|_1^2 \leq \lambda_N\|\phi^\varepsilon-\phi^0\|^2$. As before, (2.123) implies

$$\|\phi^\varepsilon(\tau_0+\tau_1)\|^2 \leq e^{-\lambda_M\varepsilon\tau_1}\|\phi^\varepsilon(\tau_0)\|^2, \tag{2.124}$$

finishing the proof of (2.112) in the second case.

3 The Fokker-Planck equations

We now start mostly following Villani's book [31]. The reader should consult this book as the primary source, both for a better presentation and to see more details and more general results. We will also occasionally make stronger assumptions than in [31], to take some small short-cuts. Finally, the reader should be warned that misprints and mistakes have likely creeped in occasionally. These are not Villani's, nor are some of the speculative comments.

The spatial Fokker-Planck equation

The Fokker-Planck equations describe the evolution of a particle in an external potential field. They can be posed either in the physical space, only accounting for the particle position, or in the phase space, keeping track of the particle density both in terms of its position in the physical space and its momentum, or velocity.

The spatial Fokker-Planck equation describes a particle that performs a Brownian motion in addition to following the gradient of an external potential field $V(x)$. Then the forward and backward Kolmogorov equations are, respectively:

$$\phi_t = \Delta\phi - \nabla V(x)\cdot\nabla\phi, \quad \phi(0,x) = \phi_0(x), \tag{3.1}$$

and

$$\rho_t = \Delta\rho + \nabla\cdot(\rho\nabla V(x)), \ \rho(0,x) = \rho_0(x). \tag{3.2}$$

Equation (3.2) conserves the total mass of the particle density $\rho(t,x)$:

$$\int \rho(t,x)dx = \int \rho_0(x)dx. \tag{3.3}$$

To see the corresponding conservation law for $\phi(t,x)$, it is convenient to re-write (3.1) in the form

$$\phi_t = e^{V(x)}\nabla\cdot(e^{-V(x)}\nabla\phi) \tag{3.4}$$

which makes it clear that

$$\int \phi(t,x)e^{-V(x)}dx = \int \phi_0(x)e^{-V(x)}dx. \tag{3.5}$$

The physical reason for (3.5) is that, as it is easy to see, the function

$$\rho_\infty(x) = e^{-V(x)} \tag{3.6}$$

is an equilibrium solution to the density equation (3.2). Thus, (3.5) simply says that the expectation of the observable $\phi_0(X(t))$ with respect to the invariant measure $\rho_\infty(x)$ is preserved by the evolution. Furthermore, if the potential $V(x)$ is sufficiently confining, in the sense that it grows sufficiently rapidly at infinity, and

$$Z = \int e^{-V(x)}dx, \tag{3.7}$$

is finite, then $\rho_\infty(x)$ can be normalized to make it be a probability density:

$$\bar{\rho}_\infty(x) = \frac{e^{-V(x)}}{Z}, \quad Z = \int e^{-V(x)} dx. \tag{3.8}$$

The invariant density encodes the confinement effect of the potential: because of the diffusion, the particle does not just concentrate at the minimum of $V(x)$ but it is more likely to be in the region where $V(x)$ is not too large.

A fundamental question for the dynamics is the convergence, and the rate of convergence, of the solution $\rho(t,x)$ with the initial condition $\rho_0 \in L^1(\mathbb{R}^d)$ to the corresponding multiple of the equilibrium solution $\bar{\rho}_\infty(x)$. This is very much related in spirit to the question of relaxation to the mean we have considered in the previous sections.

Let us also mention another reason behind (3.5): if $\phi(t,x)$ solves (3.1), then a straightforward calculation shows that

$$\rho(t,x) = e^{-V(x)}\phi(t,x) \tag{3.9}$$

is a solution to (3.2), so that (3.3) and (3.5) are equivalent. This connection between the solutions to the forward and backward Kolmogorov equations is a very convenient tool. One should always keep in mind that the function $\exp(V(x))$ blows up at infinity very fast, which limits some of the information that can be inferred from (3.9) but it is still often useful.

The kinetic Fokker-Planck equation

The kinetic version of the Fokker-Planck equation considers evolution of the particles not just in terms of their physical position but in the phase space: particles are described by their position $x \in \mathbb{R}^d$ and velocity $v \in \mathbb{R}^d$. If everything is uniform in the physical space, and there is no external potential, then the forward and backward Kolmogorov equations purely in the space of velocities are, respectively:

$$h_t = \Delta_v h - v \cdot \nabla_v h, \tag{3.10}$$

and

$$f_t = \Delta_v f + \nabla_v \cdot (vf). \tag{3.11}$$

It is easy to see that the Maxwellian

$$M(v) = \frac{1}{(2\pi)^{n/2}} e^{-|v|^2/2} \tag{3.12}$$

is a stationary solution to (3.11) and, as in (3.9), a slightly lengthy but straightforward computation shows that if $h(t,v)$ is a solution to (3.10), then

$$f(t,v) = M(v)h(t,v) \tag{3.13}$$

is a solution to (3.11). This relation is completely analogous to (3.9) that holds for the spatial Fokker-Planck equation, now with the confining potential $V_0(v) = |v|^2$, and the Maxwellian being the corresponding invariant density. This model, with the Maxwellian serving as an equilibrium, is motivated by the standard notion of temperature and kinetic energy in the theory of gases.

In the presence of an external potential $V(x)$, the problem is no longer spatially homogeneous, and the forward and backward Kolmogorov equations become

$$h_t + v \cdot \nabla_x h - \nabla_x V(x) \cdot \nabla_v h = \Delta_v h - v \cdot \nabla_v h, \tag{3.14}$$

and

$$f_t + v \cdot \nabla_x f - \nabla_x V(x) \cdot \nabla_v f = \Delta_v f + \nabla_v \cdot (vf). \tag{3.15}$$

Note that there is no diffusion in the spatial variable, the particle simply has velocity v in the x-variable. Equations (3.14) and (3.15) are no longer a gradient flow as the spatial Fokker-Planck equation: without the right side, their left side alone gives

$$f_t + v \cdot \nabla_x f - \nabla_x V(x) \cdot \nabla_v f = 0, \tag{3.16}$$

which is the Liouville equation corresponding to the classical Hamiltonian

$$E(x, v) = \frac{|v|^2}{2} + V(x). \tag{3.17}$$

In particular, any function of the form

$$f(x, v) = G(E(x, v)), \tag{3.18}$$

with a smooth function $G(E)$ is a steady solution to (3.17). In a sense, the role of the diffusion operator in the right side of (3.15) is precisely to choose one specific invariant measure out of the inifinitely many possibilities.

Our goal will be to understand the long time behavior of the solutions to (3.14) and (3.15), and for that we need to look first at the equilibrium solutions. The steady solutions to (3.14) are simply constants: $h_\infty(t, x) \equiv 1$ and its multiples. The equilibrium solution for (3.15) is no longer the Maxwellian $M(v)$ but the product of the Maxwellian and the equilibrium density that appears in (3.6):

$$f_\infty(x, v) = M(v) \exp\{-V(x)\}, \tag{3.19}$$

or its normalized version

$$\bar{f}_\infty(x, v) = \frac{M(v)}{Z} \exp\{-V(x)\}, \quad Z = \int \exp\{-V(x)\} dx. \tag{3.20}$$

Once again, if $h(t, x)$ is a solution to (3.14), then

$$f(t, x) = f_\infty(x, v) h(t, x) \tag{3.21}$$

is a solution to (3.20), an analog to (3.9). Interestingly, $f_\infty(x, v)$ is in the kernel of both sides of (3.15):

$$\Delta_v f_\infty + \nabla_v \cdot (v f_\infty) = 0, \tag{3.22}$$

and

$$v \cdot \nabla_x f_\infty - \nabla_x V(x) \cdot \nabla_v f_\infty = 0. \tag{3.23}$$

In particular, one can write $\bar{f}_\infty(x, v)$ as

$$\bar{f}_\infty(x, v) = \frac{1}{(2\pi)^{n/2} Z} e^{-E(x,v)}, \tag{3.24}$$

to see that it is of the form (3.18). Another small observation is that the marginal distribution of $\bar f_\infty(x,v)$ in v is $\bar\rho_\infty(x)$, defined in (3.8):

$$\rho_\infty(x)=\frac{1}{Z}e^{-V(x)}=\int\bar f_\infty(x,v)dv. \tag{3.25}$$

This is a direct connection of the gradient flow of the spatial Fokker-Planck equation, and the randomly perturbed Hamiltonian flow described by the kinetic Fokker-Planck equation.

To see the analogs of the self-adjoint form (3.4) and conservation law (3.5) we re-write (3.14) as

$$h_t+\nabla_v E(x,v)\cdot\nabla_x h-\nabla_x E(x,v)\cdot\nabla_v h=e^{E(x,v)}\nabla_v\cdot(e^{-E(x,v)}\nabla_v h), \tag{3.26}$$

or, equivalently, as

$$\partial_t\big(e^{-E(x,v)}h\big)-\nabla_v\cdot\big(e^{-E(x,v)}\nabla_x h\big)+\nabla_x\cdot\big(e^{-E(x,v)}\nabla_v h\big)=\nabla_v\cdot(e^{-E(x,v)}\nabla_v h). \tag{3.27}$$

It follows immediately that we have the conservation law

$$\int h(t,x)d\mu=\int h_0(x)d\mu. \tag{3.28}$$

Here, we have introduced the probability measure

$$d\mu=\bar f_\infty(x,v)dxdv, \tag{3.29}$$

with $\bar f_\infty(x,v)$ as in (3.20). As a consequence of (3.28) and the fact that the steady solutions to (3.14) are constants, it is natural to conjecture that

$$h(t,x)\to\bar h=\int h_0(x,v)d\mu \text{ as } t\to+\infty. \tag{3.30}$$

The goal of this section is to prove this convergence. Before we start, let us note that such result can only come about from the interaction between the diffusive behavior in v in the right side of (3.14), and the transport part in its left side. In particular, the transport part by itself is a first order operator and does not induce any regularization effect that would be necessary to obtain convergence to a constant. This is a hypoelliptic effect: a combination of a diffusion in the velocity variable and transport in the spatial variable leads to regularization in space and long time convergence to the invariant measure.

To state the convergence theorem, we will denote by $H^1(\mu)$ the Sobolev space:

$$\|h\|^2_{H^1(\mu)}=\int(|\nabla_x h(x,v)|^2+|\nabla_v h(x,v)|^2)\mu(dxdv). \tag{3.31}$$

We will assume that the potential $V(x)$ is in $C^2(\mathbb{R}^d)$ and grows faster than linearly at infinity, in the sense that

$$|\nabla V(x)|\to+\infty \text{ as } |x|\to+\infty. \tag{3.32}$$

A growth condition of some kind is needed to make sure that $V(x)$ is sufficiently confining. In addition, we assume that there exists $C>0$ so that

$$|D^2V(x)|\le C(1+|\nabla V(x)|) \text{ for all } x\in\mathbb{R}^d. \tag{3.33}$$

These conditions are not optimal but sufficiently general to make things interesting. Roughly speaking, they mean that $V(x)$ grows super-linearly but not faster than exponentially at infinity. The reader may think of algebraically growing $V(x) \sim |x|^\alpha$ with $\alpha > 1$ as $|x| \to +\infty$. We will show the following result on the long time behavior of the solutions to (3.14).

Theorem 3.1 *There exist $C > 0$ and $\lambda > 0$ so that for all $h_0 \in H^1(d\mu)$, we have*

$$\|h(t,x) - \bar{h}\|_{H^1(\mu)} \le Ce^{-\lambda t}\|h_0\|_{H^1(\mu)}. \tag{3.34}$$

The hypocoercivity structure of the kinetic Fokker-Planck equaiton

As we have mentioned, the main difficulty in the proof of Theorem 3.1, and especially in obtaining an exponential rate of convergence to the equilibrium, is the lack of ellipticity: there is no diffusion in x in (3.14), but only in v. An elegant way to deal with this issue is given by the hypocoercivity tools developed by Villani and collaborators. This approach will utilize a special structure of (3.14):

$$h_t + v \cdot \nabla_x h - \nabla_x V(x) \cdot \nabla_v h = \Delta_v h - v \cdot \nabla_v h. \tag{3.35}$$

We will write (3.35) as

$$\frac{\partial h}{\partial t} + \mathcal{L}h = 0, \tag{3.36}$$

with the operator

$$\mathcal{L}h = -\Delta_v h + v \cdot \nabla_v h + v \cdot \nabla_x h - \nabla_x V(x) \cdot \nabla_v h. \tag{3.37}$$

To see this structure, we will need some preliminary computations. Note that the preceding discussion shows that the natural setting for this problem is in the weighted $L^p(\mu)$ spaces, and not in $L^p(\mathbb{R}^d)$. Accordingly, let us compute the adjoint of the operator $A_k = \partial_{v_k}$ on $L^2(\mu)$:

$$\begin{aligned}
&\langle A_k f, g\rangle_{L^2(\mu)} = \int (\partial_{v_k} f(x,v)) g(x,v) \bar{f}_\infty(x,v)\,dxdv \\
&= -\int f(x,v)(\partial_{v_k} g(x,v))\bar{f}_\infty(x,v)\,dxdv - \int f(x,v)g(x,v)\partial_{v_k}\bar{f}_\infty(x,v)\,dxdv \\
&= -\langle f, A_k g\rangle_{L^2(\mu)} + \langle f, v_k g\rangle_{L^2(\mu)},
\end{aligned} \tag{3.38}$$

so that

$$A_k^* g = -\partial_{v_k} g + v_k g, \tag{3.39}$$

and

$$A_k^* A_k g = (-\partial_{v_k} + v_k)\partial_{v_k} g = -\Delta_v g + v \cdot \nabla_v g. \tag{3.40}$$

Here, and below we use the summation convention – the repeated indices are summed over, unless specified otherwise. Hence, we may write the operator $\mathcal{L}$ as

$$\mathcal{L}g = A_k^* A_k g + Bg, \tag{3.41}$$

with

$$Bg(x) = v \cdot \nabla_x g - \nabla_x V(x) \cdot \nabla_v g. \tag{3.42}$$

As in (3.27) we may re-write B as

$$\begin{aligned}Bg(x) &= v\cdot\nabla_x g - \nabla_x V(x)\cdot\nabla_v g = \nabla_v E(x,v)\cdot\nabla_x g - \nabla_x E(x,v)\cdot\nabla_v g\\ &= e^{E(x,v)}\Big[-\nabla_v\cdot(e^{-E(x,v)}\nabla_x g) + \nabla_x\cdot(e^{-E(x,v)}\nabla_v g)\Big]\\ &= \frac{1}{\bar\rho_\infty(x,v)}\Big[-\nabla_v\cdot(\bar\rho_\infty(x,v)\nabla_x g) + \nabla_x\cdot(\bar\rho_\infty(x,v)\nabla_v g)\Big].\end{aligned}\tag{3.43}$$

As a consequence, we have

$$\begin{aligned}\langle Bg,f\rangle_{L^2(\mu)} &= \int\Big[-\nabla_v\cdot(\bar\rho_\infty(x,v)\nabla_x g) + \nabla_x\cdot(\bar\rho_\infty(x,v)\nabla_v g)\Big]f(x,v)dxdv\\ &= \int\Big[-\nabla_x\cdot(\bar\rho_\infty(x,v)\nabla_v f) + \nabla_v\cdot(\bar\rho_\infty(x,v)\nabla_x f)\Big]g(x,v)dxdv\\ &= -\langle g,Bf\rangle_{L^2(\mu)},\end{aligned}\tag{3.44}$$

so that the operator B is anti-symmetric: $B^*=-B$. Thus, the operator $\mathcal{L}$ has the form (3.41) with an anti-symmetric operator B. This general structure will be our starting point but it will be useful first to get some other algebraic properties.

It is natural to understand the hypoellipticity of the operator $\mathcal{L}$ first, before discussing the long time behavior. For that we need to compute the commutators of A_k and B. First, note that, with $\partial_k=\partial_{v_k}$, we have

$$[A_k,A_m]=[\partial_k,\partial_m]=0,\tag{3.45}$$

and

$$[A_k,A_m^*]g=(A_kA_m^*-A_m^*A_k)g=\partial_k(-\partial_m+v_m)g-(-\partial_m+v_m)\partial_k g=\delta_{mk}g,\tag{3.46}$$

so that

$$[A_k,A_m^*]=\delta_{mk}I.\tag{3.47}$$

Next, we compute, with summation over the repeated indices, the commutator

$$\begin{aligned}C_kg &= [A_k,B]g=(A_kB-BA_k)g\\ &= \partial_{v_k}(v_m\partial_{x_m}g-[\partial_{x_m}V]\partial_{v_m}g)-(v_m\partial_{x_m}-[\partial_{x_m}V]\partial_{v_m})\partial_{v_k}g=\delta_{mk}\partial_{x_m}g=\partial_{x_k}g,\end{aligned}\tag{3.48}$$

hence

$$C_k=[A_k,B]=\partial_{x_k}.\tag{3.49}$$

The commutator C_k is essential here: it brings the differentiation in x that is missing in A_k. This will be absolutely crucial in what follows.

The last pair of commutators we need are

$$[A_k,C_m]=0,\quad [A_k^*,C_m]=0,\tag{3.50}$$

and, also with summation over the repeated indices:

$$\begin{aligned}[B,C_k]g &= (BC_k-C_kB)g=(v_m\partial_{x_m}-[\partial_{x_m}V]\partial_{v_m})\partial_{x_k}g-\partial_{x_k}(v_m\partial_{x_m}-[\partial_{x_m}V]\partial_{v_m})g\\ &= [\partial^2_{x_kx_m}V(x)]\partial_{v_m}g,\end{aligned}\tag{3.51}$$

so that

$$[B, C_k] = [\partial^2_{x_k x_m} V(x)]\partial_{v_m}. \tag{3.52}$$

We will see that assumption (3.33) on the Hessian of $V(x)$ is forced to accommodate this last commutator.

Now, we are in a set-up vaguely reminiscent of the abstract relaxation enhancement situation: equation (3.35) has the form

$$\frac{\partial h}{\partial t} + \Big(A_k^* A_k + B\Big)h = 0, \tag{3.53}$$

with a skew-symmetric operator B. The hypoellipticity of the operator

$$\mathcal{L} = A_k^* A_k + B \tag{3.54}$$

comes from the operator B as it produces the commutator C_k with the differentiation in x. However, the simple-minded dissipation estimate for (3.53):

$$\frac{1}{2}\frac{d}{dt}\Big(\|h(t)\|^2_{L^2(\mu)}\Big) = -\sum_{k=1}^n \|A_k h\|^2_{L^2(\mu)} \tag{3.55}$$

does not see the operator B directly at all. Hence, the relaxation to the equilibrium has to account for the presence of B in a more subtle way. The key property will come from the strict ellipticity of the operator

$$\tilde{\mathcal{L}} = A_k^* A_k + C_k^* C_k. \tag{3.56}$$

This property is essentially automatic for the kinetic Fokker-Planck equation because

$$\langle \tilde{\mathcal{L}} g, g\rangle_{L^2(\mu)} = \sum_{k=1}^n \big[\|A_k g\|^2_{L^2(\mu)} + \|C_k g\|^2_{L^2(\mu)}\big] = \int (|\nabla_x g(x,v)|^2 + |\nabla_v g(x,v)|^2)\bar{f}_\infty(x,v)dxdv, \tag{3.57}$$

so that $\tilde{\mathcal{L}}$ is, indeed, strictly elliptic. The connection between the strict ellipticity of $\tilde{\mathcal{L}}$ and the role of B in the time evolution (3.53) will be the subject of the next section on hypocoercivity.

A hypocoercivity estimate

Motivated by our observations for the kinetic Fokker-Planck equation, we now consider a more general setup where the techniques of hypocoercivity developed by Villani and collaborators can be used. We will consider the operators of the form

$$\mathcal{L} = A_k^* A_k + B, \tag{3.58}$$

with the summation convention, as usually, an anti-symmetric operator B, and unbounded operators A_k, defined on a Hilbert space $\mathcal{H}$. One may think of $\mathcal{H}$ as a weighted L^2-space, and of both A_k and B as differential operators of the first order but that is not necessary. We will denote by $\langle , \rangle$ the inner product on $\mathcal{H}$ and for simplicity will assume that $\mathcal{H}$ is a real Hilbert space, the corresponding results for complex Hilbert spaces can be obtained simply by putting the real part wherever clearly necessary.

We will always assume that $\mathcal{L}$ generates a continuous semigroup $\exp(-t\mathcal{L})$ on $\mathcal{H}$, so that the evolution equation

$$\frac{dh}{dt} + \mathcal{L}h = 0, \quad h(0) = h_0 \tag{3.59}$$

has a unique solution for all $t \geq 0$. This means that existence and uniqueness of the solutions to (3.59) is always assumed.

Obviously, if $h_0 \in \mathcal{K}$ then $h(t) \equiv h_0 \in \mathcal{K}$ for all $t \geq 0$, hence (3.59) leaves $\mathcal{K}$ invariant. Slightly less trivially, while the operators of the form (3.58) are not self-adjoint, they share an important property with self-adjoint operators: the evolution equation (3.59) preserves the orthogonal complement $\mathcal{K}^\perp$ to $\mathcal{K}$. That is, if $h_0 \in \mathcal{K}^\perp$ then $h(t) \in \mathcal{K}^\perp$ for all $t > 0$. The reason is the following simple observation.

Proposition 3.2 *If $h \in \mathcal{K}$, so that*

$$\mathcal{L}h = 0, \tag{3.60}$$

then $A_k h = 0$ for all $1 \leq k \leq n$, and $Bh = 0$.

Proof. Indeed, if $h \in \mathcal{K}$, then, as B is anti-symmetric, we have

$$0 = \langle \mathcal{L}h, h \rangle = \sum_{k=1}^{n} \|A_k h\|^2, \tag{3.61}$$

thus

$$A_k h = 0 \text{ for all } k. \tag{3.62}$$

As $\mathcal{L}h = 0$, it follows that $Bh = 0$ as well. □

It follows that if $g \in \mathcal{K}$, then

$$\mathcal{L}^* g = \sum_{k=1}^{n} A_k^* A_k g - Bg = 0,$$

and thus

$$\frac{d}{dt}\langle h(t), g \rangle = \langle \mathcal{L}h(t), g \rangle = 0.$$

As a consequence, if $h_0 \in \mathcal{K}^\perp$, then $h(t) \in \mathcal{K}^\perp$ for all $t \geq 0$. Therefore, it makes sense to consider the evolution (3.59) restricted to $\mathcal{K}^\perp$, and that is what we will do.

A standard example one may think of is that $\mathcal{L}$ is the Laplacian operator on the torus $\mathbb{T}^n$, the kernel $\mathcal{K}$ of $\mathcal{L}$ in $\mathcal{H} = L^2(\mathbb{T}^n)$, is the one-dimensional subspace that consists of constants, and its orthogonal complement $\mathcal{K}^\perp$ is the sub-space of mean-zero functions on $\mathbb{T}^n$.

Coercivity and hypocoercivity

Let us now recall the usual notion of coercivity, which is essentially strict ellipticity, together with a Poincaré inequality. We say that the operator $\mathcal{L}$ is coercive on $\tilde{\mathcal{H}} = \mathcal{H}/\mathcal{K}$ if there exists $\lambda > 0$ so that for all $h \in \tilde{\mathcal{H}} \cap D(\mathcal{L})$ we have

$$\langle \mathcal{L}h, h \rangle \geq \lambda \|h\|^2. \tag{3.63}$$

If we think of $\mathcal{L}$ as the standard Laplacian, this means that in addition to the strict ellipticity of $\mathcal{L}$, we also assume a version of the Poincare inequality on $\mathcal{H}/\mathcal{K}$, so that the norm in the right side of (3.63) is in L^2 and not in the homogeneous Sobolev space $\dot{H}^1$, that would come up in the definition of strict ellipticity. Villani allows $\tilde{\mathcal{H}}$ to be a dense subspace of $\mathcal{K}$ in [31], with a different norm, but we will simply assume that $\tilde{\mathcal{H}} = \mathcal{H}/\mathcal{K}$.

Exercise 3.3 The coercivity property can be reformulated as follows. Let $h_0 \in \tilde{\mathcal{H}}$ and $h(t)$ be the solution to

$$h_t = -\mathcal{L}h, \quad h(0) = h_0. \tag{3.64}$$

Show that $\mathcal{L}$ is coercive if and only if we have, for all $t \geq 0$ and $h_0 \in \tilde{\mathcal{H}}$ that

$$\|h(t)\| \leq e^{-\lambda t}\|h_0\|. \tag{3.65}$$

The hypocoercivity property ignores (3.63) and generalizes (3.65) instead. In the above setting, we say that the operator $\mathcal{L}$ is hypocoercive on $\tilde{\mathcal{H}}$ if there exists $\lambda > 0$ and $C > 0$ so that for all $h_0 \in \tilde{\mathcal{H}}$ we have

$$\|h(t)\| \leq Ce^{-\lambda t}\|h_0\|. \tag{3.66}$$

Thus, the only difference with the definition of coercivity is that we allow $C > 1$ in (3.66) but not in (3.65) where we must have $C = 1$. This gives us an interesting flexibility: if the Hilbert space $\tilde{\mathcal{H}}$ has two equivalent norms induced by different inner products $\langle f, g\rangle_1$ and $\langle f, g\rangle_2$ and the operator $\mathcal{L}$ is coercive in one of the norms, it may be only hypocoercive in the other. Thus, we may look for an alternative norm that would be equivalent to $\|\cdot\|_{\tilde{\mathcal{H}}}$ in which $\mathcal{L}$ would be actually coercive. That would imply hypocoercivity in the original norm.

A new inner product

Let us now explain which alternative inner product one may look for. Given an operator of the form

$$\mathcal{L} = A_k^* A_k + B, \tag{3.67}$$

with an anti-symmetric operator B, and with the summation convention over the repeated indices, we set

$$C_k = [A_k, B]. \tag{3.68}$$

The key assumption is that the operator

$$\tilde{\mathcal{L}} = A_k^* A_k + C_k^* C_k \tag{3.69}$$

is coercive. This is exactly the set-up we have seen for the kinetic Fokker-Planck equation. Coercivity of $\tilde{\mathcal{L}}$ motivates defining the $\mathcal{H}^1$ space with the norm

$$\|h\|^2_{\mathcal{H}^1} = \langle h, h\rangle + \langle \tilde{\mathcal{L}}h, h\rangle = \|h\|^2 + \sum_{k=1}^n \big(\|A_k h\|^2 + \|C_k h\|^2\big). \tag{3.70}$$

In the kinetic Fokker-Planck case this norm coincides with the standard $H^1(\mu)$-norm, as seen from (3.57). We also define $\dot{\mathcal{H}}^1 = \mathcal{H}^1/\mathcal{K}$, which is simply $\mathcal{K}^\perp$ with the homogeneous $\mathcal{H}^1$-norm

$$\|h\|^2_{\dot{\mathcal{H}}^1} = \sum_{k=1}^n \big(\|A_k h\|^2 + \|C_k h\|^2\big), \tag{3.71}$$

and $\tilde{\mathcal{H}} = \mathcal{H}/\mathcal{K}$. Coercivity of $\tilde{\mathcal{L}}$ means that there exists $\kappa > 0$ so that the Poincaré inequality holds

$$\|h\|_{\mathcal{H}} \le \kappa \|h\|_{\dot{\mathcal{H}}^1} \quad \text{for all } h \in \dot{\mathcal{H}}^1. \tag{3.72}$$

Villani considers in [31] a more general situation when higher order commutators of the form

$$C_m^{(k+1)} = [C_m^{(k)}, B]$$

are also used to obtain a coercive operator, with $C_m^{(0)} = A_m$. We will not consider it here, to focus on the simplest possible setting. This is the setting for Theorem 18 in [31] that is needed for the kinetic Fokker-Planck case, and we leave out further generalizations.

As we have mentioned, an important step is to construct another inner product that would lead to a different but equivalent norm, in which $\mathcal{L}$ itself would be actually coercive. Consider the quadratic form

$$\|h\|_*^2 = (h \cdot h) = \|h\|^2 + a \sum_{k=1}^n \|A_k h\|^2 + 2b \sum_{k=1}^n \langle C_k^* A_k h, h\rangle + c \sum_{k=1}^n \|C_k h\|^2, \tag{3.73}$$

with the constants a, b and c be specified soon, so that

$$0 < c < b < a < 1, \quad b < \sqrt{ac}. \tag{3.74}$$

One should think of these parameters as satisfying

$$0 < c \ll b \ll a \ll 1, \tag{3.75}$$

so that $\|h\|_*$ is a small perturbation of $\|h\|$ that nevertheless incorporates the $\mathcal{H}^1$-norm. We will need further constraints below on a, b and c that would guarantee the coercivity of $\mathcal{L}$ in the $\|\cdot\|_*$ norm. The second condition in (3.74) ensures that $\|\cdot\|_*$ is actually a norm, and there exists $c_0 > 0$ that depends on a, b and c such that the norms $\|\cdot\|_*$ and $\|\cdot\|_{\mathcal{H}^1}$ are equivalent:

$$c_0^{-1} \|h\|_{\mathcal{H}^1} \le \|h\|_* \le c_0 \|h\|_{\mathcal{H}^1}, \text{ for all } h \in \mathcal{H}^1. \tag{3.76}$$

The corresponding inner product on $\mathcal{H}^1$ is defined by the polarization identity:

$$\begin{aligned}(h \cdot g) = \frac{1}{4}\big(\|h+g\|_*^2 - \|h-g\|_*^2\big) &= \langle h, g\rangle + a \sum_{k=1}^n \langle A_k h, A_k g\rangle \\ &+ b \sum_{k=1}^n \langle (C_k^* A_k + A_k^* C_k) h, g\rangle + c \sum_{k=1}^n \langle C_k h, C_k g\rangle.\end{aligned} \tag{3.77}$$

It is sometimes helpful to consider the corresponding homogeneous versions:

$$\|h\|_\bullet^2 = (h \cdot h)_\bullet = a \sum_{k=1}^n \|A_k h\|^2 + 2b \sum_{k=1}^n \langle C_k^* A_k h, h\rangle + c \sum_{k=1}^n \|C_k h\|^2, \tag{3.78}$$

and

$$(h \cdot g)_\bullet = a \sum_{k=1}^n \langle A_k h, A_k g\rangle + b \sum_{k=1}^n \langle (C_k^* A_k + A_k^* C_k) h, g\rangle + c \sum_{k=1}^n \langle C_k h, C_k g\rangle. \tag{3.79}$$

This norm is equivalent to the homogeneous $\dot{\mathcal{H}}^1$-norm:

$$c_0^{-1}\|h\|_{\dot{\mathcal{H}}^1} \le \|h\|_{\bullet} \le c_0\|h\|_{\dot{\mathcal{H}}^1}, \text{ for all } h \in \dot{\mathcal{H}}^1. \tag{3.80}$$

Hence, the Poincaré inequality (3.72) can be written in the form

$$\|h\|_{\mathcal{H}} \le C_1\|h\|_{\dot{\mathcal{H}}^1} \le C_2\|h\|_{\bullet} \le C_2\|h\|_*, \quad \text{for all } h \in \dot{\mathcal{H}}^1. \tag{3.81}$$

The last observation in this section is that the orthogonal complement $\mathcal{K}^\perp$ of $\mathcal{K}$ is the same in all three inner products. Recall that if $h \in \mathcal{K}$, so that

$$\mathcal{L}h = 0, \tag{3.82}$$

then $A_k h = Bh = 0$, so that $C_k h = 0$ for all $1 \le k \le n$ as well. Therefore, if $h \in \mathcal{K}$, then for any $g \in \mathcal{H}^1$ we have

$$\langle h, g\rangle = \langle h, g\rangle_{\mathcal{H}^1} = (h \cdot g), \tag{3.83}$$

so that the orthogonal complement $\mathcal{K}^\perp$, indeed, does not depend on which of the three inner products we use.

The coercivity in the new inner product

The reason we have introduced the new inner product is that $\mathcal{L}$ is actually coercive with respect to it, under mild assumptions on the operators A_k and B. We will assume that all A_k commute with all A_m, and both A_k and A_k^* commute with all C_m:

$$[A_k, A_m] = 0, \quad [A_k, C_m] = 0, \quad [A_k^*, C_m] = 0. \tag{3.84}$$

In the kinetic Fokker-Planck case, these are properties (3.45) and (3.50) that we have verified. We will also need to assume the following, in addition to (3.84) and coercivity of $\tilde{\mathcal{L}}$. First, we assume that, as in (3.47), we have

$$[A_k, A_m^*] = \delta_{km} I. \tag{3.85}$$

This assumption is not all necessary, and is not made in the proof of Theorem 18 of [31], which is the main result of this section. The corresponding assumption on this commutator there is that there exists a constant $\alpha > 0$ so that we have

$$\sum_{k,m=1}^n \|[A_k, A_m^*]f\|^2 \le \alpha^2\|f\|^2 + \alpha^2\sum_{j=1}^n \|A_j f\|^2, \tag{3.86}$$

for all f in the domain of all A_j. However, (3.85) does simplify slightly some algebra in the proof of Proposition 3.4 below, and holds in the kinetic Fokker-Planck case, so we will make this assumption solely for the sake of convenience.

We also assume that there exists a constant $\beta > 0$ so that

$$\begin{aligned}\Big(\sum_{k=1}^n \|[C_k, B]f\|^2\Big)^{1/2} \le \beta\Big(\sum_{k=1}^n \|A_k f\|^2\Big)^{1/2} + \beta\Big(\sum_{k=1}^n \|C_k f\|^2\Big)^{1/2} \\ + \beta\Big(\sum_{k,m=1}^n \|A_k A_m f\|^2\Big)^{1/2} + \beta\Big(\sum_{k,m=1}^n \|C_k A_m f\|^2\Big)^{1/2}.\end{aligned} \tag{3.87}$$

This assumption is made mostly for technical convenience in the proof though some version of this bound seems to be necessary. Here is the key proposition.

Proposition 3.4 *Assume that the commutation relations (3.84) hold, and there exist $\alpha > 0$ and $\beta > 0$ so that the bounds (3.86) and (3.87) hold. Then, there exist a, b, c that satisfy (3.74) and such that there exists $\kappa > 0$ so that for all $h \in \mathcal{H}^1$ we have*

$$(\mathcal{L}h \cdot h) \geq \kappa \sum_{k=1}^{n} \big(\|A_k h\|^2 + \|C_k h\|^2\big). \tag{3.88}$$

Proof. We also remind that repeated indices are summed over throughout the proof. Let us write

$$\begin{aligned}(\mathcal{L}h \cdot h) &= \langle \mathcal{L}h, h\rangle + a\sum_{k=1}^{n}\langle A_k\mathcal{L}h, A_k h\rangle + b\sum_{k=1}^{n}\langle (C_k^* A_k + A_k C_k^*)\mathcal{L}h, h\rangle \\ &\quad + c\sum_{k=1}^{n}\langle C_k\mathcal{L}h, C_k h\rangle = \sum_{k=1}^{n}\|A_k h\|^2 + \sum_{k=1}^{n}(aI_k + bII_k + cIII_k).\end{aligned} \tag{3.89}$$

We will now look at each of the many individual terms separately. The first term can be further decomposed as

$$I = \sum_{k=1}^{n} I_k = \langle A_k\mathcal{L}h, A_k h\rangle = \langle A_k A_j^* A_j h, A_k h\rangle + \langle A_k B h, A_k h\rangle = I^{(A)} + I^{(B)}. \tag{3.90}$$

The last term can be bounded from below, using anti-symmetry of B and the definition of C_k, as

$$I^{(B)} = \langle A_k B h, A_k h\rangle = \langle (A_k B - B A_k)h, A_k h\rangle + \langle B A_k h, A_k h\rangle = \langle C_k h, A_k h\rangle \geq -\|A_k h\|\|C_k h\|. \tag{3.91}$$

For the first term in the right side of (3.90) we write

$$\begin{aligned}I^{(A)} &= \langle A_k A_j^* A_j h, A_k h\rangle = \langle A_j h, A_j A_k^* A_k h\rangle = \langle A_j h, A_k^* A_j A_k h\rangle + \langle A_j h, (A_j A_k^* - A_k^* A_j)A_k h\rangle \\ &= \langle A_k A_j h, A_j A_k h\rangle + \langle A_j h, [A_j, A_k^*]A_k h\rangle = \sum_{k,j=1}^{n}\|A_k A_j h\|^2 + \sum_{k=1}^{n}\|A_k h\|^2.\end{aligned} \tag{3.92}$$

We used in the last step above the fact that A_k and A_j commute, see assumption (3.84), together with our extra simplifying assumption (3.85). Thus, we have

$$I \geq \sum_{k,j=1}^{n}\|A_k A_j h\|^2 + \sum_{k=1}^{n}\|A_k h\|^2 - \sum_{k=1}^{n}\|A_k h\|\|C_k h\|. \tag{3.93}$$

As in (3.90), we split the second term in the right side of (3.89) as

$$\begin{aligned}II = \sum_{k=1}^{n} II_k &= \langle C_k^* A_k\mathcal{L}h, h\rangle + \langle A_k^* C_k\mathcal{L}h, h\rangle = \langle C_k^* A_k A_m^* A_m h, h\rangle + \langle A_k^* C_k A_m^* A_m h, h\rangle \\ &\quad + \langle C_k^* A_k B h, h\rangle + \langle A_k^* C_k B h, h\rangle = II^{(A)} + II^{(B)}.\end{aligned} \tag{3.94}$$

For the first term we write, using the fact that A_m and A_m^* commute with C_k, see (3.84), as well as the extra simplifying assumption (3.85), and, again, that A_k and A_m commute:

$$
\begin{aligned}
II^{(A)} &= \langle C_k^* A_k A_m^* A_m h, h\rangle + \langle A_k^* C_k A_m^* A_m h, h\rangle \\
&= \langle A_k A_m^* A_m h, C_k h\rangle + \langle A_m C_k h, A_m A_k h\rangle \\
&= \langle (A_m^* A_k + \delta_{km} I) A_m h, C_k h\rangle + \langle A_m C_k h, A_m A_k h\rangle \\
&= \langle A_k A_m h, A_m C_k h\rangle + \langle A_k h, C_k h\rangle + \langle A_m A_k h, A_m C_k h\rangle \\
&\geq -2 \sum_{k,m=1}^{n} \|A_k A_m h\| \|A_m C_k h\| - \sum_{k=1}^{n} \|A_k h\| \|C_k h\|.
\end{aligned} \tag{3.95}
$$

For the second term in (3.94), we have, using the commutators in (3.84) and anti-symmetry of B:

$$
\begin{aligned}
II^{(B)} &= \langle C_k^* A_k B h, h\rangle + \langle A_k^* C_k B h, h\rangle = \langle A_k B h, C_k h\rangle + \langle C_k B h, A_k h\rangle \\
&= \langle A_k B h, C_k h\rangle + \langle (B C_k + [C_k, B]) h, A_k h\rangle = \langle A_k B h, C_k h\rangle - \langle C_k h, B A_k h\rangle \\
&+ \langle [C_k, B] h, A_k h\rangle = \langle [A_k, B] h, C_k h\rangle + \langle [C_k, B] h, A_k h\rangle \\
&= \sum_{k=1}^{n} \big(\|C_k h\|^2 - \|A_k h\| \|[C_k, B] h\|\big).
\end{aligned} \tag{3.96}
$$

Thus, for the second term we have a lower bound

$$
II \geq \sum_{k=1}^{n} \big(\|C_k h\|^2 - \|A_k h\| \|[C_k, B] h\| - \|A_k h\| \|C_k h\|\big) - 2 \sum_{k,m=1}^{n} \|A_k A_m h\| \|A_m C_k h\|. \tag{3.97}
$$

For the third term in (3.89) we write

$$
III = \sum_{k=1}^{n} I_k = \langle C_k \mathcal{L} h, C_k h\rangle = \langle C_k A_j^* A_j h, C_k h\rangle + \langle C_k B h, C_k h\rangle = III^{(A)} + III^{(B)}. \tag{3.98}
$$

As C_k commutes both with A_j and A_j^*, the first term is simply

$$
III^{(A)} = \langle C_k A_j^* A_j h, C_k h\rangle = \langle C_k A_j h, C_k A_j h\rangle = \sum_{j,k=1}^{n} \|C_k A_j h\|^2. \tag{3.99}
$$

The second term in (3.98) cab be estimated using the anti-symmetry of B as

$$
III^{(B)} = \langle C_k B h, C_k h\rangle = \langle (B C_k + [C_k, B]) h, C_k h\rangle = \langle [C_k, B] h, C_k h\rangle \geq -\|[C_k, B] h\| \|C_k h\|. \tag{3.100}
$$

Putting all three terms together, we obtain

$$
\begin{aligned}
(\mathcal{L} h \cdot h) \geq& \sum_{k=1}^{n} \|A_k h\|^2 + a \sum_{k,j=1}^{n} \|A_k A_j h\|^2 + a \sum_{k=1}^{n} \|A_k h\|^2 - a \sum_{k=1}^{n} \|A_k h\| \|C_k h\| \\
&+ b \sum_{k=1}^{n} \big(\|C_k h\|^2 - \|A_k h\| \|[C_k, B] h\| - \|A_k h\| \|C_k h\|\big) \\
&- 2b \sum_{k,m=1}^{n} \|A_k A_m h\| \|A_m C_k h\| + c \sum_{j,k=1}^{n} \|C_k A_j h\|^2 - c \sum_{k=1}^{n} \|[C_k, B] h\| \|C_k h\|.
\end{aligned} \tag{3.101}
$$

Let us introduce $\xi_{1,2,3,4} > 0$ by

$$\xi_1^2 = \sum_{k=1}^n \|A_k h\|^2, \quad \xi_2^2 = \sum_{k,j=1}^n \|A_k A_j h\|^2, \quad \xi_3^2 = \sum_{k=1}^n \|C_k h\|^2, \quad \xi_4^2 = \sum_{j,k=1}^n \|C_k A_j h\|^2, \tag{3.102}$$

and rewrite assumption (3.87) as

$$\Big(\sum_{k=1}^n \|[C_k, B]h\|^2\Big)^{1/2} \leq \beta(\xi_1 + \xi_2 + \xi_3 + \xi_4). \tag{3.103}$$

With this notation, and using (3.103), gives

$$\begin{aligned}(\mathcal{L}h \cdot h) \geq (1+a)\xi_1^2 + a\xi_2^2 + b\xi_3^2 + c\xi_4^2 - a\xi_1\xi_3 - b\beta\xi_1(\xi_1+\xi_2+\xi_3+\xi_4) - b\xi_1\xi_3 \\ - 2b\xi_2\xi_4 - c\beta\xi_3(\xi_1+\xi_2+\xi_3+\xi_4) = (Q\xi, \xi),\end{aligned} \tag{3.104}$$

with the matrix

$$Q = \begin{pmatrix} 1+a & -b\beta/2 & -a-(b+c)\beta/2 & -b\beta/2 \\ -b\beta/2 & a & -c\beta/2 & -b \\ -a-(b+c)\beta/2 & -c\beta/2 & b & -c\beta/2 \\ -b\beta/2 & -b & -c\beta/2 & c \end{pmatrix}. \tag{3.105}$$

Taking $M = 1 + \beta/2$, we get

$$(\mathcal{L}h \cdot h) \geq (Q_M \xi, \xi), \tag{3.106}$$

with the matrix

$$Q_M = \begin{pmatrix} 1+a & -Mb & -a-M(b+c) & -bM \\ -bM & a & -cM & -bM \\ -a-(b+c)M & -cM & b & -cM \\ -bM & -bM & -cM & c \end{pmatrix}. \tag{3.107}$$

Assuming that, in addition to (3.74), we have, for example,

$$4c \leq 2b \leq a \leq 1, \tag{3.108}$$

we have the inequality, in the sense of quadratic forms

$$Q_M \geq \bar{Q} = \begin{pmatrix} 1 & -Ma & -2Ma & -Mb \\ -Ma & a & -Mb & -Mb \\ -2Ma & -bM & b & -cM \\ -Mb & -Mb & -Mc & c \end{pmatrix}. \tag{3.109}$$

Note that if for all $i \neq j$ we have

$$|\bar{Q}_{ij}| \leq \frac{1}{4}\sqrt{\bar{Q}_{ii}\bar{Q}_{jj}}, \tag{3.110}$$

then

$$
\begin{aligned}
(\bar{Q}\xi,\xi) &= \sum_{i,j=1}^{4} \bar{Q}_{ij}\xi_i\xi_j = \sum_{i=1}^{4} \bar{Q}_{ii}\xi_i^2 - \sum_{i\neq j} \bar{Q}_{ij}\xi_i\xi_j \geq \sum_{i=1}^{4} \bar{Q}_{ii}\xi_i^2 - \frac{1}{4}\sum_{i\neq j} \sqrt{\bar{Q}_{ii}\bar{Q}_{jj}}\xi_i\xi_j \\
&= \sum_{i=1}^{4} \bar{Q}_{ii}\xi_i^2 - \frac{1}{8}\sum_{i\neq j} (\bar{Q}_{ii}\xi_i^2 + \bar{Q}_{jj}\xi_j^2) = \sum_{i=1}^{4} \bar{Q}_{ii}\xi_i^2 - \frac{3}{4}\sum_{i=1}^{4} \bar{Q}_{ii}\xi_i^2 \\
&= \frac{1}{4}\sum_{i=1}^{4} \bar{Q}_{ii}\xi_i^2 \geq \frac{c}{4}(\xi_1^2+\xi_2^2+\xi_3^2+\xi_4^2).
\end{aligned} \tag{3.111}
$$

To arrange (3.110) for elements in the first row in $\bar{Q}$, we need to have

$$
Ma \leq \frac{1}{4}\sqrt{a},\ 2Ma \leq \frac{1}{4}\sqrt{b},\quad Mb \leq \frac{1}{4}\sqrt{c}, \tag{3.112}
$$

for the second row we need to add the conditions

$$
Mb \leq \frac{1}{4}\sqrt{ab},\ Mb \leq \frac{1}{4}\sqrt{ac}, \tag{3.113}
$$

and for the third row we need the condition

$$
Mc \leq \frac{1}{4}\sqrt{bc}. \tag{3.114}
$$

Hence, we need the following:

$$
a \leq \frac{1}{16M^2},\ 64M^2a^2 \leq b \leq \frac{a}{16M^2},\ 16M^2b^2 \leq c \leq \frac{b}{16M^2}, \tag{3.115}
$$

and also that

$$
b^2 \leq \frac{ac}{16M^2}. \tag{3.116}
$$

To achieve this, we first take $a < 1/(16M^2)$ sufficiently small, and set

$$
b = 64M^2a^2,\ c = \frac{b}{16M^2}. \tag{3.117}
$$

The three inequalities in (3.115) then hold automatically. To see that (3.116) holds if $a > 0$ is sufficiently small, note that $b^2 \sim a^4$, and $ac \sim a^3$.

We conclude that there exists a choice of a, b and c, and a very small constant $\kappa > 0$ so that

$$
(\mathcal{L}h\cdot h) \geq (Q_M\xi,\xi) \geq (\bar{Q}\xi,\xi) \geq \kappa(\xi_1^2+\xi_2^2+\xi_3^2+\xi_4^2) \geq \kappa(\xi_1^2+\xi_3^2) = \kappa\sum_{k=1}^{n}(\|A_kh\|^2+\|C_kh\|^2), \tag{3.118}
$$

finishing the proof of Proposition 3.4. □

The convergence theorem

Let us now look at the convergence to an equilibrium of the solution to the evolution problem

$$\frac{\partial h}{\partial t} + \mathcal{L}h = 0, \quad h(0) = h_0, \tag{3.119}$$

with $h_0 \in \mathcal{K}^\perp$. Recall that then $h(t) \in \mathcal{K}^\perp$ for all $t \geq 0$. Multiplying (3.119) with the $(h \cdot h)$ inner product, we obtain

$$\frac{1}{2}\frac{d}{dt}\|h(t)\|_*^2 = -(\mathcal{L}h \cdot h). \tag{3.120}$$

Using Proposition 3.4 gives then

$$\frac{1}{2}\frac{d}{dt}\|h(t)\|_*^2 \leq -\kappa\|h(t)\|_{\dot{\mathcal{H}}^1}^2. \tag{3.121}$$

So far, we have not used the Poincaré inequality (3.81) – the conclusion of Proposition 3.4 does not assume it. Now we will use it and the fact that $h(t) \in \mathcal{K}^\perp$ to deduce from (3.121) that

$$\frac{1}{2}\frac{d}{dt}\|h(t)\|_*^2 \leq -C\|h(t)\|_*^2, \tag{3.122}$$

so that

$$\|h(t)\|_* \leq \|h_0\|_* e^{-Ct}, \text{ for all } h \in \dot{\mathcal{H}}^1. \tag{3.123}$$

Using (3.81) gives the following convergence theorem.

Theorem 3.5 *Let the operator $\mathcal{L}$ be of the form $\mathcal{L} = A_k^* A_k + B$ with an anti-symmetric operator B. Assume that the operators A_k, B and $C_k = [A_k, B]$ satisfy the following assumptions:*
(i) The operator $\tilde{L} = A_k^* A_k + C_k^* C_k$ is coercive, so that the Poincaré inequality (3.72) holds.
(ii) The commutation relations (3.84) hold.
(iii) There exist $\alpha > 0$ and $\beta > 0$ so that the bounds (3.86) and (3.87) hold.
Then, there exist $C_{1,2} > 0$ and $\lambda > 0$ so that for all $h_0 \in \mathcal{K}^\perp$ the solution to (3.119) satisfies

$$\|h(t)\|_{\mathcal{H}} \leq C_1\|h(t)\|_{\dot{\mathcal{H}}^1} \leq C_2 e^{-\lambda t}\|h_0\|_{\dot{\mathcal{H}}^1}. \tag{3.124}$$

The $L^2 - H^1$ regularization

Theorem 3.5 provides a long time H^1-decay, but does not capture the regularizing effect of the dynamics as it sees no improvement of regularity of $h(t)$ relative to h_0. Regularization is a short time effect, typically instant, and says that if the initial condition $h_0 \in \mathcal{K}^\perp$ then for any $t > 0$ the solution $h(t)$ is in $\mathcal{H}^1$, seemingly a very different phenomenon from the long time convergence in the above theorem. In addition, regularization is a local in space effect and should not depend at all on the validity of the Poincaré inequality, unlike the long time behavior that is global in space. Surprisingly, this question is answered by the same technique.

Let us explain the idea on the very simple example of the heat equation in $\mathbb{R}^n$:

$$h_t = \kappa\Delta h, \quad h(0,x) = h_0(x), \quad x \in \mathbb{R}^n. \tag{3.125}$$

The standard dissipation balance is

$$\frac{1}{2}\frac{d}{dt}\|h\|_{L^2}^2 = -\kappa\|\nabla h\|_{L^2}^2, \tag{3.126}$$

which gives us a time-averaged bound

$$\kappa\int_0^\infty \int |\nabla h(t,x)|^2 dxdt < \int |h_0(x)|^2 dx, \tag{3.127}$$

but not pointwise in time. To get a pointwise in time bound, consider the quadratic form

$$F(t,h) = \|h\|_{L^2}^2 + \varepsilon t\|\nabla h\|_{L^2}^2, \tag{3.128}$$

and compute

$$\begin{aligned}\frac{dF(t,h(t))}{dt} &= -2\kappa\int |\nabla h(t,x)|^2 dx + 2\varepsilon t\int (\nabla h(t,x)\cdot \nabla h_t(t,x))dx + \varepsilon\int |\nabla h(t,x)|^2 dx\\ &= -2(\kappa-\varepsilon)\int |\nabla h(t,x)|^2 dx - 2\varepsilon t\int (\Delta h(t,x))h_t(t,x))dx\\ &= -2(\kappa-\varepsilon)\int |\nabla h(t,x)|^2 dx - 2\varepsilon\kappa t\int |\Delta h(t,x)|^2 dx < 0,\end{aligned} \tag{3.129}$$

if $0<\varepsilon\le\kappa$. It follows that $F(t,h(t))$ is decreasing in time, so that

$$\|h(t)\|_{L^2}^2 + \varepsilon t\|\nabla h(t)\|_{L^2}^2 \le \|h_0\|_{L^2}^2. \tag{3.130}$$

In particular, we see that if $h_0\in L^2(\mathbb{R}^n)$, then $h(t)\in H^1(\mathbb{R}^n)$ for any $t>0$, and, taking $\varepsilon=\kappa$ we get an estimate

$$\|\nabla h(t)\|_{L^2}^2 \le \frac{1}{\kappa t}\|h_0\|_{L^2}^2. \tag{3.131}$$

This very simple argument can be incorporated very nicely into the hypocoercivity strategy. Let us go back to the norm and inner product defined in (3.73) and (3.77):

$$\|h\|_*^2 = \|h\|^2 + a\sum_{k=1}^n \|A_k h\|^2 + 2b\sum_{k=1}^n \langle C_k^* A_k h, h\rangle + c\sum_{k=1}^n \|C_k h\|^2, \tag{3.132}$$

and

$$(h\cdot g) = \langle h,g\rangle + a\sum_{k=1}^n \langle A_k h, A_k g\rangle + b\sum_{k=1}^n \langle (C_k^* A_k + A_k^* C_k)h, g\rangle + c\sum_{k=1}^n \langle C_k h, C_k g\rangle. \tag{3.133}$$

Going through the proof of Proposition 3.4, in particular, looking at (3.111) before the very last inequality, we have actually proved the following: if we choose a, b and c so that (3.115) and (3.116) hold (we denote here $m_1=16M^2$ that appears in (3.115) and (3.116)):

$$a\le\frac{1}{m_1},\ \ 4m_1a^2\le b\le\frac{a}{m_1},\ \ m_1b^2\le c\le\frac{b}{m_1},\quad b^2\le\frac{ac}{m_1}, \tag{3.134}$$

then there exists $\kappa > 0$ that does not depend on a, b and c so that for all $h \in \mathcal{H}^1$ we have

$$(\mathcal{L}h \cdot h) \geq \kappa \sum_{k=1}^n \|A_k h\|^2 + \kappa a \sum_{j,m} \|A_j A_m h\|^2 + \kappa b \sum_{k=1}^n \|C_k h\|^2 + \kappa c \sum_{k,m=1}^n \|C_k A_m h\|^2. \quad (3.135)$$

One choice to ensure that all conditiosn in (3.134) hold is to take $a < 1/m_1$ sufficiently small, and choose b and c as in (3.117). Another possibility is to take $a < 1/m_1$ sufficiently small, set

$$b = 4m_1 a^2, \quad (3.136)$$

and choose

$$c = \alpha a^3 \quad (3.137)$$

with $\alpha > \alpha_0(m_1)$ sufficiently large. Then the two-sided inequality for c in (3.134) holds since $b \sim a^2$ and $c \sim a^3$, and a is sufficiently small, and the last inequality in (3.134) holds because both sides are of the order a^4 and $\alpha > \alpha_0(m_1)$ is large enough. Note that $\alpha_0(m_1)$ depends only on m_1 and not on how small a is.

The idea is then to take a, b and c time-dependent, generalizing (3.128): consider the quadratic form

$$F(h, g; t) = \langle h, g\rangle + \bar{a}t \sum_{k=1}^n \langle A_k h, A_k g\rangle + \bar{b}t^2 \sum_{k=1}^n \langle (C_k^* A_k + A_k^* C_k)h, g\rangle + \bar{c}t^3 \sum_{k=1}^n \langle C_k h, C_k g\rangle, \quad (3.138)$$

and

$$\mathcal{F}(h; t) = \|h\|^2 + \bar{a}t \sum_{k=1}^n \|A_k h\|^2 + 2\bar{b}t^2 \sum_{k=1}^n \langle C_k^* A_k h, h\rangle + \bar{c}t^3 \sum_{k=1}^n \|C_k h\|^2. \quad (3.139)$$

The previous discussion shows that if we choose $\bar{a}$ sufficiently small, and $\alpha > 0$ sufficiently large, and then set $\bar{b} = 4m_1\bar{a}^2$ and $\bar{c} = \alpha\bar{a}^3$, then there exists $t_0 > 0$ so that for $0 < t < t_0$ we have

$$F(\mathcal{L}h, h; t) \geq \kappa \sum_{k=1}^n \|A_k h\|^2 + \kappa\bar{a}t \sum_{j,m} \|A_j A_m h\|^2 + \kappa\bar{b}t^2 \sum_{k=1}^n \|C_k h\|^2 + \kappa\bar{c}t^3 \sum_{k,m=1}^n \|C_k A_m h\|^2. \quad (3.140)$$

Consider now the evolution problem

$$\frac{\partial h}{\partial t} + \mathcal{L}h = 0, \quad h(0) = h_0 \in \mathcal{K}^\perp. \quad (3.141)$$

As in (3.120), we have

$$\frac{1}{2}\frac{d\mathcal{F}(h;t)}{dt} = -F(\mathcal{L}h, h; t) + \bar{a} \sum_{k=1}^n \|A_k h\|^2 + 4\bar{b}t \sum_{k=1}^n \langle C_k^* A_k h, h\rangle + 3\bar{c}t^2 \sum_{k=1}^n \|C_k h\|^2. \quad (3.142)$$

Using (3.140) and dropping the terms with A_jA_m and C_kA_m gives

$$\begin{aligned}\frac{1}{2}\frac{d\mathcal{F}(h,t)}{dt} &\leq -\kappa\sum_{k=1}^{n}\|A_kh\|^2 - \kappa\bar{b}t^2\sum_{k=1}^{n}\|C_kh\|^2 + \bar{a}\sum_{k=1}^{n}\|A_kh\|^2 + 4\bar{b}t\sum_{k=1}^{n}\langle C_k^*A_kh,h\rangle \\ &+ 3\bar{c}t^2\sum_{k=1}^{n}\|C_kh\|^2 \leq (-\kappa+\bar{a}+\frac{2\bar{b}t}{\gamma})\sum_{k=1}^{n}\|A_kh\|^2 \\ &+ \Big((-\kappa\bar{b}+3\bar{c})t^2 + 2\bar{b}t\gamma\Big)\sum_{k=1}^{n}\|C_kh\|^2.\end{aligned} \tag{3.143}$$

We used Young's inequality in the last step. Taking $\gamma = t/(4\kappa)$ to make the coefficient in front of $\|C_kh\|^2$ be negative, and then choosing $\bar{a}$ sufficiently small, so that the coefficient in front of $\|A_k\|^2$ is also negative, shows that with this choice of $\bar{a}$, $\bar{b}$ and $\bar{c}$, we have

$$\frac{d\mathcal{F}(h,t)}{dt} < 0. \tag{3.144}$$

It follows that

$$\mathcal{F}(h,t) = \|h\|^2 + \bar{a}t\sum_{k=1}^{n}\|A_kh\|^2 + 2\bar{b}t^2\sum_{k=1}^{n}\langle C_k^*A_kh,h\rangle + \bar{c}t^3\sum_{k=1}^{n}\|C_kh\|^2 \leq \mathcal{F}(h,0) = \|h_0\|^2. \tag{3.145}$$

We have proved the following regularization theorem.

Theorem 3.6 *Let the operator $\mathcal{L}$ be of the form $\mathcal{L} = A_k^*A_k + B$ with an anti-symmetric operator B. Assume that the operators A_k, B and $C_k = [A_k, B]$ satisfy the following assumptions:*
(i) The commutation relations (3.84) hold.
(ii) There exist $\alpha > 0$ and $\beta > 0$ so that the bounds (3.86) and (3.87) hold.
Then, there exists $t_0 > 0$, and $c > 0$ so that for any $h_0 \in \mathcal{K}^\perp$, the solution to

$$\frac{\partial h}{\partial t} + \mathcal{L}h = 0, \quad h(0) = h_0 \in \mathcal{K}^\perp, \tag{3.146}$$

satisfies the estimates

$$\Big(\sum_{k=1}^{n}\|A_kh\|^2\Big)^{1/2} \leq \frac{c}{t^{1/2}}\|h_0\|, \quad \Big(\sum_{k=1}^{n}\|C_kh\|^2\Big)^{1/2} \leq \frac{c}{t^{3/2}}\|h_0\|. \tag{3.147}$$

As we have mentioned, Proposition 3.4 does not assume the Poincaré inequality, and neither does Theorem 3.6. This is important – an instant regularization is a local effect, and morally can not rely on the Poincaré inequality that is a global result. The regularization effect only relies on the commutator structure of $\mathcal{L}$.

An application to the kinetic Fokker-Planck equation

We now return to the proof of the long time convergence for the kinetic Fokker-Planck equation, Theorem 3.1, that was our motivation for the general hypocoercivity setup. Let us recall the equation itself:

$$h_t + v\cdot\nabla_x h - \nabla_x V(x)\cdot\nabla_v h = \Delta_v h - v\cdot\nabla_v h, \quad h(0,x) = h_0, \tag{3.148}$$

written as

$$\frac{\partial h}{\partial t} + \mathcal{L}h = 0, \tag{3.149}$$

with the operator

$$\mathcal{L}h = -\Delta_v h + v \cdot \nabla_v h + v \cdot \nabla_x h - \nabla_x V(x) \cdot \nabla_v h. \tag{3.150}$$

The problem is posed in $\mathcal{H} = L^2(\mu)$, with

$$d\mu = \frac{1}{Z(2\pi)^{n/2}} \exp\Big\{ -\frac{v^2}{2} - V(x)\Big\} dxdv.$$

Let us recall Theorem 3.1.

Theorem 3.7 *Assume that the potential $V(\cdot) \in C^2(\mathbb{R}^d)$ satisfies*

$$|\nabla V(x)| \to +\infty \ \text{as}\ |x| \to +\infty, \tag{3.151}$$

and

$$|D^2 V(x)| \le C(1 + |\nabla V(x)|) \ \text{for all}\ x \in \mathbb{R}^d. \tag{3.152}$$

There exist $C > 0$ and $\lambda > 0$ so that for all $h_0 \in H^1(d\mu)$, we have

$$\|h(t,x) - \bar{h}\|_{H^1(\mu)} \le Ce^{-\lambda t}\|h_0\|_{H^1(\mu)}, \tag{3.153}$$

with

$$\bar{h} = \int h_0(x,v)d\mu(x,v). \tag{3.154}$$

In addition, we have a regularizing effect: there exists $c > 0$ so that for all $0 < t < 1$ we have

$$\|\nabla_v h\|_{L^2(\mu)} \le \frac{c}{t^{1/2}}\|h_0 - \bar{h}\|_{L^2(\mu)}, \quad \|\nabla_x h\|_{L^2(\mu)} \le \frac{c}{t^{3/2}}\|h_0 - \bar{h}\|_{L^2(\mu)}. \tag{3.155}$$

We will apply Theorems 3.5 and Theorem 3.6, and need to verify that the kinetic Fokker-Planck equation satisfies their assumptions. The operator $\mathcal{L}$ has the form

$$\mathcal{L}g = A_k^* A_k g + Bg, \tag{3.156}$$

with $A_k = \partial_{v_k}$,

$$A_k^* g = -\partial_{v_k} g + v_k g, \tag{3.157}$$

and

$$Bg(x) = v \cdot \nabla_x g - \nabla_x V(x) \cdot \nabla_v g. \tag{3.158}$$

The adjoint A_k^* is with respect to the inner product in $\mathcal{H} = L^2(\mu)$, and B is anti-symmetric in that inner product. We also recall that the first order commutators are

$$C_k = [A_k, B] = \partial_{x_k}. \tag{3.159}$$

This is the general set-up is as in Theorem 3.5, but we need to verify assumptions (i)-(iii) of that theorem. It is immediate to see that the commutation relations (3.84) hold:

$$[A_k, A_m] = 0, \quad [A_k, C_m] = 0, \quad [A_k^*, C_m] = 0. \tag{3.160}$$

The bound (3.86)

$$\sum_{k,m=1}^{n} \|[A_k, A_m^*]f\|^2 \le C\|f\|^2 + C\sum_{j=1}^{n} \|A_j f\|^2, \tag{3.161}$$

also holds trivially, because

$$[A_k, A_m^*] = \delta_{km} I, \tag{3.162}$$

as in (3.47).

To verify the bound (3.87), note that, for the kinetic Fokker-Planck equation, as in (3.52), we have

$$[B, C_k] = [\partial^2_{x_k x_m} V(x)]\partial_{v_m}. \tag{3.163}$$

Hence, to check (3.87) it suffices to prove that

$$\begin{aligned}\sum_{k,m=1}^{n} \int |\partial^2_{x_k x_m} V|^2 |\partial_{v_k} f|^2 \mu(dxdv) \le C \sum_{k=1}^{n} \int |\partial_{v_k} f|^2 \mu(dxdv)\\ + C\sum_{k,m=1}^{n} \int |\partial^2_{v_k x_m} f|^2 \mu(dxdv).\end{aligned} \tag{3.164}$$

This inequality uses only the terms involving A_k and $C_k A_m$ in the right side of (3.87), disregarding those with $A_k A_m$ and C_k as those would not be needed here. Fixing $1 \le k \le n$ and setting $g(x,v) = \partial_{v_k} f$ and using assumption (3.152), we see that (3.164) would follow from the following lemma.

Lemma 3.8 *Let $V(\cdot) \in C^2(\mathbb{R}^d)$ satisfy (3.151) and (3.152), then there exists $C > 0$ so that for all $g \in \mathcal{H}^1(\mu)$ we have*

$$\int |\nabla V|^2 |g(x,v)|^2 \mu(dxdv) \le C \int (|g(x,v)|^2 + |\nabla_x g(x,v)|^2)\mu(dxdv). \tag{3.165}$$

Proof. It suffices to show that for a function $g(x)$ that does not depend on v, we have

$$\int |\nabla V(x)|^2 g^2(x) e^{-V(x)} dx \le C \int (g^2(x) + |\nabla g(x)|^2) e^{-V(x)} dx. \tag{3.166}$$

To get (3.165) for a function $g(x,v)$, we would then simply write (3.166) for each v and integrate in v, with the weight $M(v)dv$. To prove (3.166), let us write

$$\begin{aligned}&\int |\nabla V(x)|^2 g^2(x) e^{-V(x)} dx = -\int g^2(x) \nabla V(x) \cdot \nabla(e^{-V(x)}) dx\\ &= \int \nabla \cdot (g^2(x)\nabla V(x)) e^{-V(x)} dx = \int g^2(x)(\Delta V(x)) e^{-V(x)} dx\\ &+ 2\int g(x)(\nabla g(x) \cdot \nabla V(x)) e^{-V(x)} dx = I + II.\end{aligned} \tag{3.167}$$

We bound the first term above using assumption (3.152) and Young's inequality

$$I \le C \int g^2(x)(1 + |\nabla V(x)|) e^{-V(x)} dx \le C \int g^2(x) e^{-V(x)} dx + \frac{1}{10} \int g^2(x) |\nabla V(x)|^2 e^{-V(x)} dx. \tag{3.168}$$

For the second term in the right side of (3.167) we simply use Young's inequality:

$$II \le C \int |\nabla g(x)|^2 e^{-V(x)} dx + \frac{1}{10} \int g^2(x) |\nabla V(x)|^2 e^{-V(x)} dx. \tag{3.169}$$

Putting (3.167), (3.168) and (3.169) together gives (3.166). □

We conclude that (3.164) holds. The last step is to check assumption (i) in Theorem 3.5: the operator

$$\tilde{\mathcal{L}} = A_k^* A_k + C_k^* C_k$$

is coercive. In our case, this means verifying the following Poincaré inequality:

$$\int (|\nabla_x g(x,v)|^2 + |\nabla_v g(x,v)|^2) d\mu \ge \kappa \int (g(x,v) - \bar{g})^2 d\mu, \tag{3.170}$$

with

$$\bar{g} = \int g(x,v) d\mu. \tag{3.171}$$

Let us denote $y = (x, v)$,

$$E(y) = \frac{|v|^2}{2} + V(x) - \log Z, \tag{3.172}$$

so that

$$\mu(dy) = e^{-E(y)} dy \tag{3.173}$$

is a probability density. Let us note that the Hamiltonian $E(y)$ satisfies the property

$$w(y) = \frac{|\nabla_y E|^2}{2} - \Delta E(y) \to +\infty \quad \text{as } |y| \to +\infty. \tag{3.174}$$

This follows from assumptions (3.151) and (3.152) on $V(x)$.

Lemma 3.9 *Let $E(y) \in C^2(\mathbb{R}^n)$ be such that $\mu(dy)$ given by (3.173) is a probability measure, and (3.174) holds. Then the Poincaré inequality holds for all $g \in H^1(\mu)$:*

$$\int |\nabla g(y)|^2 d\mu(y) \ge \kappa \int (g(y) - \bar{g})^2 d\mu(y), \tag{3.175}$$

with

$$\bar{g} = \int g(y) d\mu(y). \tag{3.176}$$

Proof. Let g be a smooth compactly supported function and write

$$g(y) = h(y) e^{E(y)/2}, \quad \nabla g = \Big[\nabla h + \frac{h}{2} \nabla E\Big] e^{E(y)/2},$$

so that

$$\begin{aligned} \int |\nabla g(y)|^2 d\mu(y) &= \int |\nabla g(y)|^2 e^{-E(y)} dy \ge \frac{1}{4} \int h^2 |\nabla E|^2 dy + \int h(y)(\nabla h(y) \cdot \nabla E(y)) dy \\ &= \frac{1}{2} \int h^2(y) \Big[\frac{|\nabla E(y)|^2}{2} - \Delta E(y)\Big] dy = \frac{1}{2} \int g^2(y) w(y) e^{-E(y)} dy \\ &= \frac{1}{2} \int g^2(y) w(y) d\mu(y), \end{aligned} \tag{3.177}$$

with $w(y)$ defined in (3.174). Let us take $R_0 > 0$ sufficiently large, so that $w(y) > 1$ if $|y| > R_0$, and note that (3.174) implies that

$$\varepsilon(R) = \frac{1}{\min(w(y) : \ |y| \geq R)} \to 0 \text{ as } R \to +\infty. \tag{3.178}$$

Using (3.177) we can write

$$2\int |\nabla g(y)|^2 d\mu(y) \geq \varepsilon_R^{-1} \int_{|y|\geq R} g^2(y) d\mu(y) - M \int_{|y|\leq R} g^2(y) d\mu(y), \tag{3.179}$$

with

$$M = \max_{y\in\mathbb{R}} [w(y)]_-.$$

It follows that

$$\int_{|y|\geq R} g^2(y) d\mu(y) \leq 2\varepsilon_R \int |\nabla g(y)|^2 d\mu(y) + \varepsilon_R M \int g^2(y) d\mu(y). \tag{3.180}$$

Next, let assume, in addition, that

$$\bar{g} = \int g(y) d\mu(y) = 0. \tag{3.181}$$

Let also $\mu_R = (1/\mu(B_R))\mu$ be the the restriction of μ to the ball $B_R = B(0; R)$, normalized to be a probability measure, set

$$\bar{g}_R = \int g(y) d\mu_R(y) = \frac{1}{\mu(B_R)} \int_{|y|\leq R} g(y) d\mu(y),$$

and let p_R be the Poincaré constant for μ_R. Note that p_R is finite because μ_R is compactly supported. We will show that the whole space Poincaré inequality (3.175) holds with a constant κ that is close to p_R if R is large enough.

To see this, note that, as μ_R is a probability measure, the μ_R-Poincaré inequality

$$\int (g(y) - \bar{g}_R)^2 d\mu_R \leq p_R \int |\nabla g(y)|^2 d\mu_R, \tag{3.182}$$

can be written as

$$\int g^2(y) d\mu_R \leq p_R \int |\nabla g(y)|^2 d\mu_R + \bar{g}_R^2. \tag{3.183}$$

Using the definition of μ_R, this is simply

$$\frac{1}{\mu(B_R)} \int_{|y|\leq R} g^2(y) d\mu \leq \frac{p_R}{\mu(B_R)} \int_{|y|\leq R} |\nabla g(y)|^2 d\mu + \frac{1}{\mu(B_R)^2} \Big(\int_{|y|\leq R} g(y) d\mu(y) \Big)^2. \tag{3.184}$$

Taking R sufficiently large, so that $\mu(B_R) > 1/2$, we obtain

$$\int_{|y|\leq R} g^2(y) d\mu \leq p_R \int_{|y|\leq R} |\nabla g(y)|^2 d\mu + 2\Big(\int_{|y|\leq R} g(y) d\mu(y) \Big)^2. \tag{3.185}$$

Recalling (3.181), we see that

$$\Big(\int_{|y|\le R} g(y)d\mu(y)\Big)^2 = \Big(\int_{|y|\ge R} g(y)d\mu(y)\Big)^2 \le \int_{|y|\ge R} g^2(y)d\mu(y). \tag{3.186}$$

Using this in (3.185) gives

$$\int g^2(y)d\mu \le p_R \int_{|y|\le R} |\nabla g(y)|^2 d\mu + 3\int_{|y|\ge R} g^2(y)d\mu(y). \tag{3.187}$$

Inserting this into (3.180), we obtain

$$\int g^2(y)d\mu \le p_R \int_{|y|\le R} |\nabla g(y)|^2 d\mu + 6\varepsilon_R \int |\nabla g(y)|^2 d\mu(y) + \varepsilon_R M \int g^2(y)d\mu(y). \tag{3.188}$$

Taking R sufficiently large so that $\varepsilon_R < 1/(2M)$, we arrive at

$$\int g^2(y)d\mu \le \frac{(p_R + 6\varepsilon_R)}{1-\varepsilon_R M} \int |\nabla g(y)|^2 d\mu, \tag{3.189}$$

finishing the proof. □

We have now verified all assumptions of Theorem 3.5 for the kinetic Fokker-Planck equations, and the proof of Theorem 3.7 is also complete.

The maximum principle and spatial Fokker-Planck transport

We now describe a simple alternative way [22], based on the maximum principle, to see that the solution to the spatial Fokker-Planck equation (3.1)

$$\phi_t = \Delta\phi - \gamma\nabla V(x)\cdot\nabla\phi, \quad \phi(0,x) = \phi_0(x), \tag{3.190}$$

"spreads its mass around", and a corresponding result for the dual equation

$$\rho_t = \Delta\rho + \gamma\nabla\cdot(\rho\nabla V(x)), \quad \rho(0,x) = \rho_0(x), \tag{3.191}$$

Here, the constant γ measures the strength of the potential. This method does not prove convergence of the solution to (3.190) to a constant steady state, or that of (3.191) to the invariant measure, but gives an idea behind the mixing mechanism and how it improves for large $\gamma \gg 1$. We will take a continuously differentiable radially symmetric increasing potential $V(x)$ such that there exists $\delta_0 > 0$ so that

$$V'(r) \ge \frac{c_0}{1+r}, \text{ for all } r > \delta_0, \tag{3.192}$$

so that $V(r)$ has at least logarithmic growth as $r \to +\infty$. This condition is much weaker than our previous assumptions for the kinetic Fokker-Planck equation. Note that if we take $\gamma > 0$ sufficiently large, then we still have

$$\int_{\mathbb{R}^n} e^{-\gamma V(x)}dx < +\infty. \tag{3.193}$$

Our goal will be to show by very simple methods that solutions to (3.190) spread fast if γ is sufficiently large, without concerning ourselves with the much more precise question of the convergence to an equilibrium.

Let us assume for the moment that the initial condition $\phi_0(x)$ is smooth, radially symmetric, and non-increasing in the radial direction, so that $\phi(t,x)$ remains radially symmetric. In the radial coordinates (3.190) takes the form

$$\phi_t = \phi'' + \frac{n-1}{r}\phi' - \gamma V'(r)\phi'. \tag{3.194}$$

Differentiating this equation and using the maximum principle we conclude that $\phi'(t,r) \leq 0$ for all $t > 0$ and $r > 0$ if $\phi_0(r)$ is decreasing in the radial direction.

Recall that (3.190) preserves the weighted mass:

$$\int \phi(t,x)e^{-\gamma V(x)}dx = \int \phi_0(x)e^{-\gamma V(x)}dx. \tag{3.195}$$

In addition, for any $r > 0$ we have

$$\frac{d}{dt}\int_{B_r} \phi(t,x)e^{-\gamma V(x)}dx = \int_{B_r} \nabla \cdot \Big(e^{-\gamma V(x)}\nabla\phi(t,x)\Big)dx = |\partial B_r|e^{-\gamma V(|x|)}\phi'(t,r) < 0, \tag{3.196}$$

so that the weighted mass inside each ball centered at the origin is decreasing in time, which is a very primitive indicator of spreading.

We will prove the following spreading estimate.

Theorem 3.10 *Assume that*

$$\phi_0(x) \geq \mathbb{1}_{B_{r_1}}(x), \tag{3.197}$$

with some $r_1 > 2\delta_0$. Then there exist $c = c(r_1) > 0$ and $\gamma_0(r_1)$ such that for all $\gamma > \gamma_0$ we have

$$\phi(t,x) \geq c\mathbb{1}_{B_{\sqrt{c\gamma t}}}(x) \tag{3.198}$$

for all $t \geq 0$.

Proof. It follows from the maximum principle that it suffices to prove this result for

$$\phi_0(x) = \mathbb{1}_{B_{r_1}}(x), \tag{3.199}$$

and this is what we will assume. Note that then $\phi(t,x)$ is radially symmetric and satisfies (3.194). The first step is the following bound.

Lemma 3.11 *There exists $\gamma_0(r_1)$ so that if $\gamma > \gamma_0(r_1)$ then*

$$\phi(t,x) \geq \frac{1}{2} \text{ for all } |x| = r_1/2 \text{ and } t > 0. \tag{3.200}$$

Proof of Lemma 3.11. We set $r_0 = r_1/2 > \delta_0$. Observe that, as $\phi(t,x)$ is radially symmetric and decreasing in the radial direction, we have, for all x with $|x| = r_0$:

$$\begin{aligned}
&\phi(t,x)\int_{\mathbb{R}^n\setminus B_{r_0}} e^{-\gamma V(y)}dy \geq \int_{\mathbb{R}^n\setminus B_{r_0}} \phi(t,y)e^{-\gamma V(y)}dy \\
&= \int_{\mathbb{R}^n} \phi(t,y)e^{-\gamma V(y)}dy - \int_{B_{r_0}} \phi(t,y)e^{-\gamma V(y)}dy = \int_{\mathbb{R}^n} \phi_0(y)e^{-\gamma V(y)}dy - \int_{B_{r_0}} \phi(t,y)e^{-\gamma V(y)}dy \\
&\geq \int_{\mathbb{R}^n} \phi_0(y)e^{-\gamma V(y)}dy - \int_{B_{d_0}} \phi_0(y)e^{-\gamma V(y)}dy = \int_{\mathbb{R}^n\setminus B_{r_0}} \phi_0(y)e^{-\gamma V(y)}dy \\
&= \int_{\mathbb{R}^n\setminus B_{r_0}} \mathbb{1}_{B_{r_1}}(y)e^{-\gamma V(y)}dy = \int_{B_{r_1}\setminus B_{r_0}} e^{-\gamma V(y)}\,dy.
\end{aligned} \tag{3.201}$$

In the first step we used monotonicity of $\phi(t,x)$ in the radial variable, in the third step the conservation law (3.195), in the fourth step we used (3.196), and in the next to last step we used assumption (3.197) on the initial condition $\phi_0(x)$.

We claim that there exists $\gamma_0(r_0)$ so that

$$\int_{\mathbb{R}^n\setminus B_{r_0}} e^{-\gamma V(x)}\,dx \leq 2\int_{B_{r_1}\setminus B_{r_0}} e^{-\gamma V(x)}\,dx, \tag{3.202}$$

for all $\gamma > \gamma_0(r_0)$. To this end, note from (3.192) that there exists c_1 that depends on r_0 so that

$$|\nabla V(x)| = \partial_r V \geq c_1|x|^{-1}, \quad \text{if } |x| \geq r_0, \tag{3.203}$$

so that for all $z_2 > z_1 \geq r_0$ we have

$$V(z_2) - V(z_1) \geq c_1 \log\Big(\frac{z_2}{z_1}\Big).$$

Multiplying by γ, exponentiating and taking $z_1 = r$, $z_2 = 2r$ gives

$$e^{\gamma(2r)-\gamma V(r)} \geq 2^{c_1\gamma},$$

so that

$$e^{-\gamma V(2r)} \leq \frac{1}{2^{c_1\gamma}}e^{-\gamma V(r)}, \quad \text{for all } r \geq r_0, \tag{3.204}$$

and

$$e^{-\gamma V(2^k r)} \leq \frac{1}{2^{c_1\gamma k}}e^{-\gamma V(r)}, \quad \text{for all } r \geq r_0, \tag{3.205}$$

Then we have

$$\begin{aligned}
\int_{\mathbb{R}^n\setminus B_{r_0}} e^{-\gamma V(x)}dx &= \sum_{k=0}^{\infty}\int_{B_{2^{k+1}r_0}\setminus B_{2^k r_0}} e^{-\gamma V(x)}dx = \sum_{k=0}^{\infty} 2^{kn}\int_{B_{r_1}\setminus B_{r_0}} e^{-\gamma V(2^k x)}dx \\
&\leq \sum_{k=0}^{\infty}\frac{2^{kn}}{2^{c_1\gamma k}}\int_{B_{r_1}\setminus B_{r_0}} e^{-\gamma V(x)}dx \leq 2\int_{B_{r_1}\setminus B_{r_0}} e^{-\gamma V(x)}dx,
\end{aligned} \tag{3.206}$$

so that (3.202) holds, if $\gamma > (n+1)/c_1$, Using (3.201) and (3.202), we conclude that

$$\phi(t,x) \geq 1/2 \text{ for all } |x| = r_0 \text{ and } t > 0, \tag{3.207}$$

and the proof of Lemma 3.11 is complete. □

Next, we construct a sub-solution for $\phi(t,x)$. Take a convex and strictly decreasing function $\omega(r)$ defined for $r > r_0$, such that $\omega(r_0) = 1/2$ and $\omega(r_1) = 0$, and $\omega(r) > 0$ for $r \in (r_0, r_1)$. We look for a sub-solution to $\phi(t,x)$ in the form

$$\xi(t,x) = \omega(r_0 + s(t)(|x| - r_0)), \tag{3.208}$$

with a decreasing function $s(t)$ such that $s(0) = 1$. Note that at $t = 0$ we have

$$\phi(0,x) = \mathbb{1}_{B_{r_1}}(x) \geq \xi(0,x) = \omega(|x|). \tag{3.209}$$

We now compute, with $z = r_0 + s(t)(|x| - r_0)$:

$$\begin{aligned} \xi_t - \Delta\xi + \gamma\nabla V(x)\cdot\nabla\xi &= s'(t)(r-r_0)\omega'(z) - s^2(t)\omega''(z) - \frac{(n-1)s(t)}{r}\omega'(z) \\ + \gamma s(t)V'(r)\omega'(z) &\leq s'(t)(r-r_0)\omega'(z) - \frac{(n-1)s(t)}{r}\omega'(z) + \frac{c_2\gamma s(t)}{r}\omega'(z). \end{aligned} \tag{3.210}$$

In the last step we used the fact that ω is convex and decreasing, and also that (3.192) implies that there exists $c_2(r_0)$ such that for all $r > r_0$ we have

$$V'(r) \geq \frac{c_2}{r}.$$

In order for $\xi(t,x)$ to be a sub-solution for $\phi(t,x)$, it suffices that the right side of (3.210) is negative or, equivalently,

$$s'(t)(r-r_0) - \frac{(n-1)s(t)}{r} + \frac{c_2\gamma s(t)}{r} \geq 0, \tag{3.211}$$

just in the region where $\xi(t,x) > 0$, that is, for $z < r_1$, or

$$r < r_0 + \frac{r_1 - r_0}{s(t)} \leq \frac{r_1}{s(t)}. \tag{3.212}$$

We used the assumption that $s(t) \leq 1$ above. As $s'(t) \leq 0$, it is sufficient to check that (3.211) holds at $r = r_1/s(t)$, which is true if

$$s'(t)\frac{r_1}{s(t)} + \frac{[c_2\gamma + 1 - n]s^2(t)}{r_1} \geq 0, \tag{3.213}$$

or

$$\frac{s'(t)}{s^3(t)} + c_3\gamma \geq 0, \tag{3.214}$$

with c_3 that depends on r_1 and n, provided that γ is sufficiently large. Hence, we may take

$$s(t) = \frac{1}{\sqrt{1 + c_4\gamma t}}, \tag{3.215}$$

with a sufficiently small c_4. Thus, we have shown that $\xi(t,x)$ is a sub-solution for $\phi(t,x)$ and we have

$$\phi(t,x) \geq \omega\Big(r_0 + \frac{|x| - r_0}{\sqrt{1 + c_4\gamma t}}\Big). \tag{3.216}$$

The conclusion of Theorem 3.10 now follows. □

To get the corresponding result for the density equation (3.191), let $\phi(t,x)$ and $\rho(t,x)$ be the solutions to the backward and forward spatial Fokker-Planck equaitons

$$\phi_t = \Delta\phi - \gamma\nabla V(x)\cdot\nabla\phi, \quad \phi(0,x) = \phi_0(x), \tag{3.217}$$

and

$$\rho_t = \Delta\rho + \gamma\nabla\cdot(\rho\nabla V(x)), \quad \rho(0,x) = \rho_0(x). \tag{3.218}$$

By the duality of these two equations, we have

$$\frac{d}{ds}\int_{\mathbb{R}^n} \rho(t-s,x)\phi(s,x)dx = 0, \tag{3.219}$$

so that

$$\int_{\mathbb{R}^n} \rho(t,x)\phi_0(x)dx = \int_{\mathbb{R}^n} \rho_0(x)\phi(t,x)dx. \tag{3.220}$$

The following corollary says that if the initial mass $\rho_0(x)$ is located within distance L from the origin, a non-trivial fraction of the total mass will enter the unit ball by a time of the order $t \sim L^2/\gamma$.

Corollary 3.12 *Let $\rho(t,x)$ satisfy (3.218) with an initial condition ρ_0 that satisfies $\rho_0(x) \geq 0$ and*

$$\int_{1\leq|x|\leq L} \rho_0(x)\,dx = M_0.$$

Then there exist γ_0 and c so that for all $\gamma > \gamma_0$ and $t \geq L^2/(c\gamma)$, we have

$$\int_{B_1} \rho(x,t)\,dx \geq cM_0. \tag{3.221}$$

Proof. Let $\phi(t,x)$ satisfy (3.217) with the initial condition $\phi_0(x) = \mathbb{1}_{B_1}(x)$. Theorem 3.10 and (3.220) imply that

$$\int_{B_1} \rho(t,x)dx = \int_{\mathbb{R}^n} \rho_0(x)\phi(t,x)dx \geq \int_{1\leq|x|\leq L} \rho_0(x)\phi(t,x)dx \geq c\int_{1\leq|x|\leq L} \rho_0(x)dx = cM_0, \tag{3.222}$$

for all t such that $\sqrt{c\gamma t} > L$, if $\gamma > \gamma_0$. Here, c and γ_0 are the constants in Theorem 3.10 corresponding to $r_1 = 1$. □

References

[1] W.O. Amrein, and V. Georgescu, On the characterization of bound states and scattering states in quantum mechanics. Helv. Phys. Acta **46**, 1973/74, 635–658. 32

[2] D.V. Anosov, *The additive functional homology equation that is connected with an ergodic rotation of the circles (Russian),* Izv. Akad. Nauk SSSR Ser. Mat. **37**, 1973, 1259–1274. 18, 21, 27

[3] H. Berestycki, F. Hamel and N. Nadirashvili, Elliptic eigenvalue problems with large drift and applications to nonlinear propagation phenomena. Comm. Math. Phys. **253**, 2005, 451–480. 7

[4] P. Constantin, A. Kiselev and L. Ryzhik, Quenching of flames by fluid advection, Comm. Pure Appl. Math., **54**, 2001, 1320–1342. 11

[5] P. Constantin, A. Kiselev, L. Ryzhik and A. Zlatos, Diffusion and mixing in fluid flow. Ann. of Math., **168**, 2008, 643–674. 13, 16

[6] I.P. Cornfeld, S.V. Fomin and Ya.G. Sinai, *Ergodic Theory,* Springer-Verlag, New York, 1982. 26

[7] H. Cycon, R. Froese, W. Kirsch and B. Simon, *Schrödinger Operators,* Springer-Verlag, 1987. 32

[8] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol. 3: Spectral Theory and Applications, Springer, 2000. 6

[9] D. Dürr, S. Goldstein and J. Lebowitz, Asymptotic motion of a classical particle in a random potential in two dimensions: Landau model, Comm. Math. Phys., **113**, 1987, 209–230.

[10] V. Enss, Asymptotic completeness for quantum mechanical potential scattering. I. Short range potentials. Comm. Math. Phys. **61**, 1978, 285–291. 32

[11] L.C. Evans, Partial Differential Equations, Second Edition, AMS, 2010. 4, 17

[12] A. Fannjiang and L. Wo?owski, Noise induced dissipation in Lebesgue-measure preserving maps on d-dimensional torus. J. Statist. Phys. **113**, 2003, 335–378. 15

[13] A. Fannjiang, S. Nonnenmacher, L. Wołowski, Relaxation time of quantized toral maps. Ann. Henri Poincar **7**, 2006, 161–198. 15

[14] A. Fannjiang, S. Nonnenmacher, L. Wołowski, Dissipation time and decay of correlations. Nonlinearity **17**, 2004, 1481–1508. 15

[15] B. Fayad, *Weak mixing for reparameterized linear flows on the torus,* Ergodic Theory Dynam. Systems **22**, 2002, 187–201. 27

[16] B. Fayad, Fayad, Bassam Smooth mixing flows with purely singular spectra. Duke Math. J. **132**, 2006, 371–391. 27

[17] B. Franke, Integral inequalities for the fundamental solutions of diffusions on manifolds with divergence-free drift. (English summary) Math. Z. **246**, 2004, n 373–403. 15

[18] Ichihara, K. snd Kunita, H., *A classification of the second order degenerate elliptic operators and its probabilistic characterization*, Z. Wahrscheinlichkeitstheorie verw., **30**, 1974, 235-254. 13

[19] Hörmander, L., Hypoelliptic second order differential equations, Acta Math., **119**, 1967, 147-171. 12, 13

[20] A. Katok, *Cocycles, cohomology and combinatorial constructions in ergodic theory,* In collaboration with E. A. Robinson, Jr. Proc. Sympos. Pure Math., 69, Smooth ergodic theory and its applications (Seattle, WA, 1999), 107–173, Amer. Math. Soc., Providence, RI, 2001. 18, 22, 27

[21] Y. Kifer, *Random Perturbations of Dynamical Systems*, Birkhäuser, Boston, 1988. 15

[22] A. Kiselev, F. Nazarov, L. Ryzhik and Y. Yao, Chemotaxis and reactions in biology, Preprint, 2020. 62

[23] A. Kiselev and A. Zlatos, Quenching of combustion by shear flows, Duke Math. J. **132**, 2006, 49–72.

[24] A.N. Kolmogorov, *On dynamical systems with an integral invariant on the torus (Russian),* Doklady Akad. Nauk SSSR (N.S.), **93**, 1953, 763–766. 13, 18, 21, 27

[25] J. Moser, *On the volume elements on a manifold,* Trans. Amer. Math. Soc., **120**, 1965, 286–294.

[26] J. von Neumann, *Zur operatorenmethode in der classichen mechanik,* Annals of Math (2), **33**, 1932, 587–642. 27

[27] J.R. Norris, Long-time behaviour of heat flow: global estimates and exact asymptotics. Arch. Rational Mech. Anal. **140**, 1997, 161–195. 15

[28] D. Ruelle, A remark on bound states in potential-scattering theory. Nuovo Cimento A (10) **61**, 1969, 655–662. 32

[29] M.D. Šklover, *Classical dynamical systems on the torus with continuous spectrum,* Izv. Vuzov, **10**, 1967, 113–124. 27

[30] S. Sternberg, *On differential equations on the torus,* Amer. J. Math., **79**, 1957, 397–402. 21

[31] C. Villani, *Hypocoercivity*, Memoirs of the AMS, **202**, 2009. 39, 47, 48, 49