CONTRACTION IN THE WASSERSTEIN METRIC FOR THE KINETIC FOKKER-PLANCK EQUATION ON THE TORUS

HELGE DIETERT, JOSEPHINE EVANS AND THOMAS HOLDING
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

(Communicated by Anton Arnold)

Abstract. We study contraction for the kinetic Fokker-Planck operator on the torus. Solving the stochastic differential equation, we show contraction and therefore exponential convergence in the Monge-Kantorovich-Wasserstein $W_2$ distance. Finally, we investigate if such a coupling can be obtained by a co-adapted coupling, and show that then the bound must depend on the square root of the initial distance.

1. Introduction. The kinetic Fokker-Planck equation, also known as the Kramers equation, is a basic model for the spreading of a solute due to interaction with the fluid background. It is derived from Langevin dynamics, where the time scale of observation is much larger than the correlation time of the solute-fluid interactions (see e.g. [17]).

We prove contraction properties of the spatially periodic kinetic Fokker-Planck equation in the Wasserstein metric, and show to what extent the probabilistic technique of coupling can be used in such situations. This is of interest, both intrinsically, and in the broader context of analytic and probabilistic methods of proving convergence to equilibrium and contraction properties of Fokker-Planck equations which we summarise in the paragraphs below. The Monge-Kantorovich-Wasserstein (MKW) distance comes from optimal transport and is defined as

$$W_2(\mu, \nu) = \inf_{\pi \in \Pi_{\mu, \nu}} \left( \int |x - y|^2 \, d\pi(x, y) \right)^{1/2},$$

where $\Pi_{\mu, \nu}$ is the set of all couplings between $\mu$ and $\nu$.

A common analytic technique to show contraction or convergence to equilibrium of Fokker-Planck equations is to work in a $L^2$ space weighted by the reciprocal of the equilibrium measure. Here, in the spatially homogeneous setting, contractivity is established by showing that the generator of the Fokker-Planck semi-group is coercive on this $L^2$ space, which implies that the generator has a spectral gap. In the spatially inhomogeneous setting, which is common in kinetic theory, the generator is, however, not coercive in this space and this method fails.

2010 Mathematics Subject Classification. 60J60; 35Q84, 60H30, 82C31.

Key words and phrases. Convergence to equilibrium, hypocoercivity, Monge-Kantorovich-Wasserstein $W_2$ distance, Fokker-Planck equation, torus, co-adapted couplings.

The authors were supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/H023348/1 for the University of Cambridge Centre for Doctoral Training, the Cambridge Centre for Analysis.
This lead to the development of the celebrated theory of hypocoercivity for which an excellent reference is [15], where a spectral gap and contraction of the semi-group are shown, roughly speaking, by constructing equivalent ‘skew’ $L^2$ or Sobolev norms on which the generator is coercive. This theory is well developed and in Proposition 5 in the first section of [15], Villani shows that the equation on the law for a large class of SDEs can be put in the $A^*A + B$ form he uses in his theorems. Hypocoercive SDEs are also studied in [14] which is one of the early works in the development of hypocoercivity. The techniques of hypocoercivity apply also to collisional models [6]. The kinetic Fokker-Planck equation in particular has received much attention [7, 5, 13] both in the case of a spatial confining potential and in, the analytically simpler, case of spatial periodicity. The paper [5] considers exactly our equation and finds explicitly the optimal rates of convergence in weighted $L^2$ space. The motivation is similar to that of this paper, which is to study a simple toy model on which more explicit calculations can be performed in order to explore alternative methods for proving hypocoercivity. These works, however, do not address the question of contraction in the Wasserstein metric $W_2$, as this distance is inaccessible from these analytic tools; the closest result to this being [11] where $W_1$ results are obtained by duality. Using interpolation estimates and convergence results in other spaces, one can conclude exponential decay in the Wasserstein $W_w$ distance. However, then the control in terms of the initial data only holds for a power strictly less than one.

Another viewpoint, strongly related to the first, comes from the theory of gradient flows [8], in which the Fokker-Planck equation is identified with the steepest descent flow of an entropy functional in the Wasserstein space $W_2$. However, the theory does not cover the considered model due to the kinetic structure. Dissipation in the Wasserstein distance can also be shown for non-gradient drifts in the homogeneous setting using analytic methods [1].

A common probabilistic technique to show contraction or convergence is to construct a coupling between two copies of the stochastic process that realises the desired bound on the metric between the laws. In the spatially homogeneous Fokker-Planck equation, the synchronisation coupling, where the infinitesimal motions of the noise are coupled together, gives contraction in Wasserstein metrics when the velocity potential is strongly convex. In the spatially inhomogeneous case with a confining potential, such a straightforward coupling only establishes contraction if the confining potentials are quadratic (or a small perturbation thereof) see for example [2]. Establishing contraction in the Wasserstein metric for more general confining potentials is an open problem. In the spatially periodic case results are even more limited. In this case the synchronisation coupling does not cause the spatial distance on the torus to decay. Thus the spatially periodic case is more difficult in the probabilistic case. This is in contrast to the analytic setting, where having the spatial variable on the torus means hypocoercivity can be shown by a very similar, and in fact slightly simpler, computation to that in part 1 section 7 of [15] will show hypocoercivity.

In this work we study the contraction properties in the Wasserstein metric of the kinetic Fokker-Planck equation with spatial variable on the torus and a quadratic velocity potential. Despite the simplicity of this equation, to the authors’ knowledge this question has not been answered in the literature, and a second goal of this manuscript it to understand what difficulties might explain this.
This kinetic Fokker-Planck equation describes the law of a particle moving in the phase space $\mathbb{T} \times \mathbb{R}$ whose location in the phase space is $(X_t, V_t)$ and evolves as
\[
\begin{align*}
    dX_t &= V_t \, dt, \\
    dV_t &= -\lambda V_t \, dt + dW_t,
\end{align*}
\]
where $W_t$ is a Brownian motion and the spatial variable is in the torus $\mathbb{T} = \mathbb{R}/(2\pi L \mathbb{Z})$ of length $2\pi L$.

The corresponding law $\mu_t$ on $\mathbb{T} \times \mathbb{R}$ evolves as
\[
\partial_t \mu_t + v \partial_x \mu_t = \partial_v \left[ \lambda v \mu_t + \frac{1}{2} \partial_v \mu_t \right],
\]
where this equation is considered in the weak sense. The equilibrium state for this equation is
\[
\frac{1}{2\pi L} \text{Leb} \otimes \sqrt{\frac{\lambda}{\pi}} \exp \left( -\frac{1}{4\lambda} v^2 \right) \text{Leb}.
\]

Solving the stochastic evolution, we show exponential decay of the distance between two solutions.

**Theorem 1.1.** If $\mu_t$ and $\nu_t$ are two solutions to the kinetic Fokker-Planck equation (2), then we have
\[
W_2(\mu_t, \nu_t) \leq (e^{-\lambda t} + ce^{-t/2\lambda^2 L^2}) W_2(\mu_0, \nu_0)
\]
for a constant $c$ only depending on $L$.

**Remark 1.** We are not aware of any paper showing optimal rate of convergence for this process in $W_2$. The paper [5] shows that for large times this is the optimal rate of convergence in a weighted $L^2$ space. Also, we show later that we can split the process in components which are broadly an Ornstein-Uhlenbeck process with rate $\lambda$ and a Brownian motion with diffusivity $1/\lambda$ on the torus. One would expect the optimal rate of convergence for an O-U process in any reasonable distance to be $\lambda$ and the optimal rate of convergence for the diffusion process to be $1/2\lambda^2 L^2$. Therefore it seems likely that our rates are optimal.

The key idea is that, after conditioning on the final velocity, the spatial variable has enough randomness left to allow such a coupling. This approach is not based on a functional inequality which is integrated over time.

In fact the evolution is not a contraction semigroup in the considered distance which we can show directly in a straightforward way using the explicit solution to the SDE. Precisely,

**Proposition 1.** The kinetic Fokker-Planck operator is not coercive in the MKW distance. The inequality
\[
W_2(\mu_t, \nu_t) \leq e^{-\gamma t} W_2(\mu_0, \nu_0), \quad \forall \mu_0, \nu_0
\]
cannot hold for any $\gamma > 0$.

In order to construct a coupling showing convergence in the MKW distance, random variables $(X^i_t, V^i_t)$ are constructed for $t \in \mathbb{R}^+$ and $i = 1, 2$ such that $(X^1_t, V^1_t)$ has law $\mu_t$ and $(X^2_t, V^2_t)$ has law $\nu_t$. Then for $t \in \mathbb{R}^+$ the coupling $((X^1_t, V^1_t), (X^2_t, V^2_t))$ gives an upper bound of the MKW distance $W_2(\mu_t, \nu_t)$.

The fact that (1) is an evolution equation means that it could be considered more natural from a probabilistic viewpoint to consider couplings that evolve along the flow of the equation. This motivates us to look at couplings where $(X^1_t, V^2_t)$ are
Theorem 1.4. Suppose there exists a function \( \alpha : \mathbb{R}^+ \to \mathbb{R}^+ \) and a constant \( \gamma > 0 \) such that for all initial distributions \( \mu_0 \) and \( \nu_0 \) there exists a co-adapted coupling \( ((X^1_t, V^1_t), (X^2_t, V^2_t)) \) such that
\[
\left( \mathbb{E} \left( |X^1_t - X^2_t|^2 + (V^1_t - V^2_t)^2 \right) \right)^{1/2} \leq \alpha(t) \left( \mathbb{W}_2(\mu_0, \nu_0) e^{-\gamma t} \right).
\]
Then there exists a constant \( C \) such that for \( z \in (0, \pi L] \) we have the following lower bound on the dependence on the initial distance
\[
\alpha(z) \geq C \sqrt{z}.
\]

The idea is to focus on a drift-corrected position on the torus, which evolves as a Brownian motion. By stopping the Brownian motion at a large distance we can then prove the claimed lower bound.

Definition 1.2 (co-adapted coupling). The coupling \( ((X^1_t, V^1_t), (X^2_t, V^2_t)) \) is co-adapted if, for \( i = 1, 2 \), under the filtration \( \mathcal{F} \) that is generated by the coupling \( ((X^1_t, V^1_t), (X^2_t, V^2_t)) \), the process \( (X^i_t, V^i_t) \) is a continuous Markov process whose transition semigroup is determined by (1).

Remark 2. Kendall [9] calls an equi-filtration coupling a Brownian motion. By stopping the Brownian motion at a large distance we can then prove the claimed lower bound.
This shows that a simple hypocoercivity argument on a Markovian coupling cannot work. Precisely, there cannot exist a semigroup \( P_t \) on the probability measures over \((\mathbb{T} \times \mathbb{R})^2\), whose marginals behave like the solution of (1) and which satisfies
\[ H(P_t(\pi)) \leq cH(\pi)e^{-\gamma t} \quad \text{for} \quad H^2(\pi) = \int[(X_1^2 + X_2^2) + (V_1^2 + V_2^2)]d\pi(X_1, V_1, X_2, V_2). \]
Otherwise, the Markov process associated to \( P \) would be a coupling contradicting Theorem 1.4.

2. Set up. The stochastic differential equation (1) has an explicit solution, when posed in \( \mathbb{R}^2 \). For clarity, when we are considering \( X \) to be in \( \mathbb{R} \) rather than the torus we will denote it \( \hat{X} \). The explicit solution is
\[
\hat{X}_t = \hat{X}_0 + \frac{1}{\lambda}(1 - e^{-\lambda t})V_0 + \int_0^t \frac{1}{\lambda}(1 - e^{-\lambda(t-s)})dW_s, \\
V_t = e^{-\lambda t}V_0 + \int_0^t e^{-\lambda(t-s)}dW_s,
\]
where \( W_t \) is the common Brownian motion. In this we separate the stochastic driving as \((A_t, B_t)\) given by the stochastic integrals
\[
A_t = \int_0^t \frac{1}{\lambda}(1 - e^{-\lambda(t-s)})dW_s, \\
B_t = \int_0^t e^{-\lambda(t-s)}dW_s,
\]
which evolve as a vector in \( \mathbb{R}^2 \) with the common Brownian motion \( W_t \). By Itô’s isometry \((A_t, B_t)\) is a Gaussian random variable with covariance matrix \( \Sigma(t) \) given by
\[
\Sigma_{AA}(t) = \frac{1}{\lambda^2} \left[ t - \frac{2}{\lambda}(1 - e^{-\lambda t}) + \frac{1}{2\lambda}(1 - e^{-2\lambda t}) \right], \\
\Sigma_{AB}(t) = \frac{1}{\lambda^2} \left[ (1 - e^{-\lambda t}) - \frac{1}{2}(1 - e^{-2\lambda t}) \right], \\
\Sigma_{BB}(t) = \frac{1}{2\lambda}(1 - e^{-2\lambda t}).
\]
From this we calculate that the conditional distribution of \( A_t \) given \( B_t \) is a Gaussian with variance \( \Sigma_{AA}(t) - \Sigma_{AB}^2(t, b) \Sigma_{BB}^{-1}(t) b \) and mean given by
\[
\mu_{A|B}(t, b) = \Sigma_{AB}(t) \Sigma_{BB}^{-1}(t)b.
\]
We write \( g_{A|B} \) for the conditional density of \( A \) given \( B \) and \( g_B \) for the marginal density of \( B \). Hence
\[
g(t, a, b) = g_{A|B}(t, a, b)g_B(t, b)
\]
is the joint density of \( A \) and \( B \).

The last part of the set up is the change of variables we will need for the Markovian coupling. We define new coordinates \((Y, V)\) in \( \mathbb{T} \times \mathbb{R} \) by taking the drift away
\[
\begin{cases}
Y = X + \frac{1}{\lambda}V, \\
V = V.
\end{cases}
\]
The motivation for this change is the explicit formulas found in (3) from which we see that \( Y \) is the limit as \( t \to \infty \) of \( X_t \) without additional noise. In the new
variables, (1) becomes
\[
\begin{aligned}
    dY_t &= \frac{1}{\lambda} dW_t, \\
    dV_t &= -\lambda V_t dt + dW_t,
\end{aligned}
\]
for the common Brownian motion \(W_t\). Note that the motion of \(Y_t\) does not depend explicitly upon \(V_t\) and is a Brownian motion on the torus.

It remains to show that these new coordinates define an equivalent norm on \(T \times \mathbb{R}\). This follows from the triangle inequality and we have
\[
|X^1 - X^2|_T + |V^1 - V^2| \leq |Y^1 - Y^2|_T + \left(1 + \frac{1}{\lambda}\right) |V^1 - V^2|
\]
and the other direction is similar. Thus, the two norms are equivalent up to a constant factor that depends only on \(\lambda\).

3. **Non-Markovian coupling.** We wish to estimate how much the spatial variable will spread out over time. We will then use this to construct a coupling at a fixed time \(t\) which exploits the fact that a proportion of the spatial density is distributed uniformly. In order to do this we give a lemma on the spreading of a Gaussian density wrapped on the torus.

**Lemma 3.1.** For \(\sigma^2 > 2L^2 \log(3)\) consider the Gaussian density \(h\) on \(\mathbb{R}\) given by
\[
h(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}
\]
and wrap it onto the torus \(T\), i.e. define the density \(Qh\) on \(T\) by
\[
(Qh)(x) = \sum_{n \in \mathbb{Z}} h(x + 2\pi Ln).
\]
We have the following estimate on the spatial spreading
\[
Qh(x) \geq \frac{\beta}{2\pi L}
\]
where
\[
1 - \beta = \frac{2e^{-\sigma^2/2L^2}}{1 - e^{-\sigma^2/2L^2}} \in (0, 1).
\]

**Proof.** We define the Fourier transform of a function on \(T\) to be
\[
(Fg)(k) = \int_T e^{ikx/L} g(x) dx,
\]
where
\[
\int_T g(x) dx = \int_0^{2\pi L} g(x) dx.
\]
By the definition of \(Q\), the Fourier transform of \(Qh\) is given by
\[
(FQh)(k) = \int_T \sum_{n \in \mathbb{Z}} h(x + 2\pi Ln) e^{ikx/L} dx = \int_\mathbb{R} h(x) e^{ikx/L} dx = \exp\left(-k^2\sigma^2/2L^2\right)
\]
where we have used the well-known Fourier transformation of a Gaussian.

Writing \(Qh\) in terms of its Fourier series and subtracting the \(k = 0\) term, we have for any \(x \in \mathbb{T}\)
\[
Qh(x) - \frac{\beta}{2\pi L} = \frac{1}{2\pi L} \sum_{|k| \geq 1} e^{-k^2\sigma^2/2L^2 - ikx/L} + \frac{1 - \beta}{2\pi L}.
\]
We want this to be positive. Therefore it is sufficient to show that
\[ \left| \sum_{|k| \geq 1} e^{-k^2 \sigma^2 / 2L^2} - ikx / L \right| \leq 1 - \beta. \]

We estimate the left hand side by
\[ \left| \sum_{|k| \geq 1} e^{-k^2 \sigma^2 / 2L^2} - ikx / L \right| \leq 2 \sum_{k \geq 1} e^{-k \sigma^2 / 2L^2} = 1 - \beta \]
where the final equality follows from summing the geometric series.

We can now use this to construct a coupling at time \( t \). We will use this coupling to prove exponential decrease in the Wasserstein distance.

**Lemma 3.2.** Let \( t \geq 0 \), be large enough so the variance of \( g_A|B \) is greater than \( 2L \log(3) \), and \( \beta \) be such that
\[ (Qg_A|B)(t,a,b) \geq \frac{\beta}{2\pi L}, \]
where \( g_A|B \) is defined by (7) above. Let \( \mu_t \text{ resp. } \nu_t \) be the distribution of the solution to the Fokker-Planck equation (2) with deterministic initial data \( \mu_0 = \delta_{x_1^0,v_1^0} \text{ and } \nu_0 = \delta_{x_2^0,v_2^0} \) respectively, at time \( t \). Then there exists a coupling \( ((X_1^t, V_1^t), (X_2^t, V_2^t)) \) between \( \mu_t \) and \( \nu_t \) satisfying
\[ \mathbb{E} \left[ (V_1^t - V_2^t)^2 \right] = e^{-2\lambda t} \left[ (v_0^1 - v_0^2)^2 \right] \]
and
\[ \mathbb{E} \left[ |X_1^t - X_2^t|^2 \right] \leq 2(1 - \beta) \left[ |x_1^0 - x_2^0|^2 + \frac{1}{\lambda \sigma^2} (v_0^1 - v_0^2)^2 \right]. \]

**Proof.** Let us construct such a coupling. Since we have seen that \( g_A|B \) is Gaussian density with variance \( \sigma^2 = \Sigma_{AA}(t) - \Sigma_{AB}(t)\Sigma_{BB}^{-1}(t) \), we can use Lemma 3.1 to split the distribution \( Qg_A|B \) as
\[ Qg_A|B(t,a,b) = \frac{\beta}{2\pi L} + (1 - \beta)s(t,a,b). \]
Then by assumption \( s \) is again a probability density for the variable \( a \) on the torus \( T \). We now consider the torus as a subset of \( \mathbb{R} \) and then \( Qg_A|B \) and \( 1/2\pi L \) are probability density functions. Therefore, \( s \) is also probability density functions supported on \( [0, 2\pi L] \). Let \( B \) be an independent random variable with density \( g_B(t,b) \), let \( Z \) be an independent uniform random variable over \( [0,1] \) and let \( U \) be an independent uniform random variable over the torus. Finally let \( S \) be a random variable on \( \mathbb{R} \) with density \( s(t, \cdot, B) \), viewed as a density function on \( \mathbb{R} \), only depending on \( B \).

With this define the random parts \( A^1, A^2 \) of \( X_1^t, X_2^t \) as
\[ A^1 = 1_{Z \leq \beta} \left[ U - x_0^1 - \frac{1}{\lambda}(1 - e^{-\lambda t})v_0^1 \right] + 1_{\beta > Z} S, \]
\[ A^2 = 1_{Z \leq \beta} \left[ U - x_0^2 - \frac{1}{\lambda}(1 - e^{-\lambda t})v_0^2 \right] + 1_{\beta > Z} S. \]
We then construct \((X^1_t, V^1_t)\) defined by
\[
X^1_t = x^1_0 + \frac{1}{\lambda}(1 - e^{-\lambda t})v^1_0 + A^1,
\]
\[
V^1_t = e^{-\lambda t}v^1_0 + B,
\]
and \((\hat{X}^2_t, \hat{V}^2_t)\) defined by
\[
\hat{X}^2_t = x^2_0 + \frac{1}{\lambda}(1 - e^{-\lambda t})v^2_0 + A^2,
\]
\[
\hat{V}^2_t = e^{-\lambda t}v^2_0 + B.
\]
We then construct \(X^i_t\) by wrapping \(\hat{X}^i_t\) onto the torus (i.e., \(X^i_t \in [0, 2\pi L]\) and \(X^i_t \equiv X^i_t \mod 2\pi L\). By construction the pairs \((X^i, V^i)\) have the right laws so they form a valid coupling.

We find
\[
E \left[ (V^1_t - V^2_t)^2 \right] = e^{-2\lambda t} \left[ (v^1_0 - v^2_0)^2 \right]
\]
and
\[
E \left[ |X^1_t - X^2_t|^2 \right] = (1 - \beta) \left[ |x^1_0 - x^2_0 + \frac{1}{\lambda}(1 - e^{-\lambda t})(v^1_0 - v^2_0)|^2 \right]
\]
and we can use Young’s inequality to find the claimed control.

We now put these two lemmas together to prove Theorem 1.1, which states exponential convergence in the MKW \(W_2\) distance.

**Proof of Theorem 1.1.** We first show that we can reduce to working with deterministic initial conditions. We denote \(\mu^x_v\) to be the law of the solution to the SDE initialized at \((x, v)\). Suppose we know that
\[
\mathcal{W}_2(\mu^x_1, \nu_1) \leq \omega(t) \int d((x_1, v_1), (x_2, v_2)).
\]

Since, \(\mu_t, \nu_t\) are the laws of Markov processes we know that,
\[
\mu_t(\phi) = \int \int \phi(y, u) \mu_t^{x,v}(y, u) d\mu_0(x, v).
\]
Hence given, \(\pi\) a coupling of \(\mu_0, \nu_0\) we can construct a coupling of \(\mu_t, \nu_t\) by
\[
\pi_t(\psi) = \int \int \psi((y_1, u_1), (y_2, u_2)) \mu_t^{x_1,v_1}(y_1, u_1) d\pi(x, v) d\mu_0(x, v).
\]
The couplings of this form are a subset of all the couplings of \(\mu_t, \nu_t\) therefore we can take the infimum over these couplings in order to bound the Wasserstein distance. Then given any coupling \(\pi\) of initial measures \(\mu_0, \nu_0\) we have
\[
\mathcal{W}_2(\mu_t, \nu_t)^2 \leq \int_{(\mathbb{R} \times \mathbb{R})^2} \mathcal{W}_2(\mu_t^{x,v_1}, \mu_t^{x,v_2})^2 d\pi((x_1, v_1), (x_2, v_2))
\]
\[
\leq \omega(t)^2 \int_{(\mathbb{R} \times \mathbb{R})^2} d((x_1, v_1), (x_2, v_2))^2 d\pi((x_1, v_1), (x_2, v_2)).
\]
Then taking an infimum over \(\pi\) shows that this implies
\[
\mathcal{W}_2(\mu_t, \nu_t) \leq \omega(t) \mathcal{W}_2(\mu_0, \nu_0).
\]
Given any initial points \(((x^1_0, v^1_0), (x^2_0, v^2_0))\), we can use Lemma 3.2 to construct a coupling \(((X^1_t, V^1_t), (X^2_t, V^2_t))\) of \(\mu_t\) and \(\nu_t\). From explicitly calculating the variance of the distribution of \(A|B\) using (4), (5), (6), we see that the variance grows.
asymptotically as $t/\lambda^2$. Hence by Lemma 3.1 we can choose $\beta$ so that $1 - \beta \to 0$ exponentially fast with rate $1/2\lambda^2 L^2$. This, combined with the control from the second lemma, shows that
\[
\mathbb{E} \left[ (V_t^1 - V_t^2)^2 + |X_t^1 - X_t^2|^2 \right] \leq \left( e^{-2\lambda t} + c e^{-t/2\lambda^2 L^2} \right) \left( (V_0^1 - V_0^2)^2 + |x_0^1 - x_0^2|^2 \right).
\]

The explicit solution also allows to prove that the evolution is not a contraction semigroup.

\textit{Proof of Proposition 1.} We will prove the theorem by contradiction. Suppose $\gamma > 0$ and let $a \neq b$ be two distinct points on the torus. Consider the initial measures
\[
\mu_0 = \delta_{x=a}\delta_{v=0}
\]
and
\[
\nu_0 = \delta_{x=b}\delta_{v=0}.
\]
Then the distance is $W_2(\mu_0, \nu_0) = |a - b|_T$.

At time $t$ the spatial distribution of $\mu_t$ and $\nu_t$, interpreted in $\mathbb{R}$, is a Gaussian with variance $\Sigma_{AA}$ which by the explicit formula (4) can be bounded as
\[
\Sigma_{AA}(t) \leq C_A t^2
\]
for a constant $C_A$ and $t \leq 1$.

Hence for $d > 0$ and $t \leq 1$ the spatial spreading is controlled as
\[
\mu_t((T \setminus [a - d, a + d]) \times \mathbb{R}) \leq \frac{2\Sigma_{AA}(t)}{d} \exp \left( \frac{-d^2}{2\Sigma_{AA}(t)} \right) \leq C_1 \frac{t^2}{d} \exp \left( -C_2 \frac{d^2}{t^4} \right)
\]
for positive constants $C_1$ and $C_2$, where we have used the standard tail bound for the Gaussian distribution (see e.g. [12, Lemma 12.9]).

For any $d > 0$ small enough that $a \pm d$ and $b \pm d$ do not wrap around the torus, any coupling between $\mu_t$ and $\nu_t$ must transfer at least the mass
\[
1 - \mu_t((T \setminus [a - d, a + d]) \times \mathbb{R}) - \nu_t((T \setminus [b - d, b + d]) \times \mathbb{R})
\]
between $[a - d, a + d]$ and $[b - d, b + d]$.

Hence the Wasserstein distance is bounded by
\[
W_2^2(\mu_t, \nu_t) \geq (|a - b|_T - 2d)^2 \left( 1 - 2C_1 \frac{t^2}{d} \exp \left( -C_2 \frac{d^2}{t^4} \right) \right).
\]

Taking $d = |a - b|_T t^{3/2}$ for $t$ sufficiently small, this shows that
\[
W_2^2(\mu_t, \nu_t) \geq |a - b|_T^2 (1 - 2t^{3/2})^2 \left( 1 - \frac{2C_1}{|a - b|_T} \sqrt{t} \exp \left( -C_2 \frac{|a - b|_T^2}{t} \right) \right).
\]

However, for all small enough positive $t$, we have
\[
(1 - 2t^{3/2})^2 > e^{-\gamma t/2}
\]
and
\[
\left( 1 - \frac{2C_1}{|a - b|_T} \sqrt{t} \exp \left( -C_2 \frac{|a - b|_T^2}{t} \right) \right) > e^{-\gamma t/2}
\]
contradicting the assumed contraction. For the second estimate we use $\exp(-c/t) \leq (1 + c/t)^{-1} = t/(c + t)$. \qed
4. Co-adapted couplings.

4.1. Existence. For Theorem 1.3 we construct a reflection/synchronisation coupling using the drift-corrected positions $Y^1_t$. As the positions are on the torus we can use a reflection coupling until $Y^1_t$ and $Y^2_t$ agree. Afterwards, we use a synchronisation coupling which keeps $Y^1_t = Y^2_t$ and reduces the velocity distance.

For a formal definition let $((X^1_0,V^1_0), (X^2_0,V^2_0))$ be a coupling between $\mu$ and $\nu$ obtaining the MKW distance (the existence of such a coupling is a standard result, see e.g. [16, Theorem 4.1.]).

We then define the evolution of this coupling in two stages. First, define $(X^1_t,V^1_t)$ and $(X^2_t,V^2_t)$ to be strong solutions to (1) with initial conditions $((X^1_0,V^1_0), (X^2_0,V^2_0))$ respectively and driving Brownian motion $W^1_t$. Then we recall the definition of $Y^i$ from (8), and define the stopping time $T := \inf\{t \geq 0 : Y^1_t = Y^2_t\}$. Then we define a new process $W^2_t$ by

$$W^2_t = \begin{cases} -W^1_t & t \leq T, \\ W^1_t - 2W^2_t & t > T. \end{cases}$$

By the reflection principle, $W^2$ is a Brownian motion. We use this to define a new solution $(X^2_t, V^2_t)$ to be the strong solution to (1) with driving Brownian motion $W^2$ and initial condition $(X^2_0,V^2_0)$. Note now that $T = \inf\{t \geq 0 : Y^1_t = Y^2_t\}$.

For the analysis we introduce the notation

$$M_t = Y^1_t - Y^2_t,$$
$$Z_t = V^1_t - V^2_t.$$  

Then by the construction the evolution is given by

$$dM_t = \frac{2}{\lambda} 1_{t \leq T} dW^1_t, $$  
$$dZ_t = -\lambda Z_t dt + 2 \cdot 1_{t \leq T} dW^1_t,$$  

where $M_t$ evolves on the torus $T$.

As a first step we introduce a bound for $T$.

**Lemma 4.1.** *The stopping time $T$ satisfies*

$$\mathbb{P}(T > t | M_0) = 4 \pi \sum_{k=0}^{\infty} \frac{1}{2k+1} 2k+1 \exp\left(-\frac{(2k+1)^2}{2\lambda^2 L^2} t\right) \sin\left(\frac{(2k+1)\pi |M_0|^2}{2L}\right).$$  

(12)

**Proof.** As $M_t$ evolves on the torus, $T$ is the first exit time of a Brownian motion starting at $M_0$ from the interval $(0,2\pi L)$. See [12, (7.14-7.15)], from which the claim follow after rescaling to incorporate the $2/\lambda$ factor.

**Remark 3.** The second expression in (12) is obtained by solving the heat equation on $[0,2\pi L]$ with Dirichlet boundary conditions and initial condition $\delta_{M_0}$.

**Lemma 4.2.** *There exists a constant $C$ such that for any $t > 0$ the following holds*

$$\mathbb{P}(T > t | M_0) \leq C |M_0|^2 (1 + t^{-1/2}) e^{-t/(2\lambda^2 L^2)}.$$  

(13)

**Proof.** Using (12) and the inequality $\sin(x) \leq x$ for $x \geq 0$, we have

$$\mathbb{P}(T > t | M_0) \leq \frac{4}{\pi} e^{-t/(2\lambda^2 L^2)} \sum_{k=0}^{\infty} \frac{|M_0|^2}{2L} \frac{2k+1}{2k+1} e^{-4k^2 t/(2\lambda^2 L^2)}$$  

(12)
\[ \leq \frac{2}{\pi L} |M_0|_T e^{-t/(2\lambda^2 L^2)} \left( 1 + \int_0^\infty e^{-4u^2t/(2\lambda^2 L^2)} du \right) \]
\[ = \frac{2}{\pi L} |M_0|_T e^{-t/(2\lambda^2 L^2)} \left( 1 + \frac{\pi}{\sqrt{8t/(\lambda^2 L^2)}} \right) \]
\[ \leq C |M_0|_T (1 + t^{-1/2}) e^{-t/(2\lambda^2 L^2)} \]

where on the second line we have bounded the sum by an integral. \( \square \)

Using these simple estimates, we now study the convergence rate of the coupling.

**Lemma 4.3.** There exists a constants \( C \) such that for any \( t \geq 0 \) we have the bound

\[ E \left[ |M_t|^2 + |Z_t|^2 \left| (Z_0, M_0) \right. \right] \leq |Z_0|^2 e^{-2\lambda t} + \begin{cases} C |M_0|_T e^{-2\lambda t} & 2\lambda < 1/(2\lambda^2 L^2) \\
C |M_0|_T (1 + t) e^{-2\lambda t} & 2\lambda = 1/(2\lambda^2 L^2) \\
C |M_0|_T e^{-t/(2\lambda^2 L^2)} & 2\lambda > 1/(2\lambda^2 L^2). \end{cases} \]

**Proof.** Without loss of generality we may assume that \( Z_0 \) and \( M_0 \) are deterministic in order to avoid writing the conditional expectation.

Applying Itô’s lemma, we find from (11) that

\[ d|Z_t|^2 = -2\lambda |Z_t|^2 dt + 4 \cdot 1_{t \leq T} Z_t dW_t^1 + 2 \cdot 1_{t \leq T} dt. \]

After taking expectations we see that

\[ \frac{d}{dt} E|Z_t|^2 = -2\lambda E|Z_t|^2 + 2E(t \leq T). \] \hspace{1cm} (14)

By explicitly solving (14) and using Lemma 4.2, we obtain

\[ E|Z_t|^2 = |Z_0|^2 e^{-2\lambda t} + 2e^{-2\lambda t} \int_0^t e^{2\lambda s} \mathbb{P}(s \leq T) ds \]
\[ \leq |Z_0|^2 e^{-2\lambda t} + C |M_0|_T e^{-2\lambda t} \int_0^t e^{(2\lambda - 1/(2\lambda^2 L^2))s} (1 + s^{-1/2}) ds. \]

Let us bound \( I_t \). As the integrand is locally integrable, we have for a constant \( C \)

\[ I_t \leq C \left( 1 + \int_0^t e^{(2\lambda - 1/(2\lambda^2 L^2))s} ds \right). \]

Here the \( s^{-1/2} \) term can be bounded by 1 for \( s > 1 \) and for \( s \leq 1 \) the additional contribution can be absorbed into the constant. To bound the remaining integral we consider three cases:

- \( 2\lambda < 1/(2\lambda^2 L^2) \): The integral (and \( I_t \)) are uniformly bounded, \( I_t \leq C \).
- \( 2\lambda = 1/(2\lambda^2 L^2) \): The integrand is equal to 1 and \( I_t \leq C(1 + t) \).
- \( 2\lambda > 1/(2\lambda^2 L^2) \): The integrand grows and \( I_t \leq C(1 + e^{(2\lambda - 1/(2\lambda^2 L^2))t}) \).

In each case we multiply \( I_t \) by \( e^{-2\lambda t} \) to obtain the decay rate. In the first two cases this gives the dominant term with \( |M_0|_T \) (as opposed to \( |Z_0| \)) dependence, while in the last case it is lower order than the \( e^{-t/(2\lambda^2 L^2)} \) decay we obtain from \( E|M_t|^2_T \) below.

Next let us consider \( E|M_t|^2_T \). Using the finite diameter of the torus we have the simple estimate

\[ E|M_t|^2_T \leq \pi^2 L^2 \mathbb{P}(T > t). \]
For $t \geq 1$ (say), we can use Lemma 4.2, to obtain

$$\mathbb{E}|M_t|_T^2 \leq C|M_0|_T e^{-t/(2\lambda^2 L^2)} \quad \text{for } t \geq 1.$$ 

This leaves the case when $t \leq 1$ where (13) blows up. We instead use the martingale property of $M_t$. Without loss of generality we may assume that $M_0 \in [0, \pi L]$. Then as $M_t$ is stopped at $T$ we know that $M_t \in [0, 2\pi L]$ for all $t \geq 0$. Hence, for any $t \geq 0$,

$$\mathbb{E}|M_t|_T^2 \leq \mathbb{E}|M_t|^2 \leq 2\pi LEM_t = 2\pi LM_0 = 2\pi L|M_0|_T$$

by the martingale property. Combining the $t \leq 1$ and $t \geq 1$ estimates we have

$$\mathbb{E}|M_t|_T^2 \leq C|M_0|_T e^{-t/(2\lambda^2 L^2)} \quad \text{for } t \geq 0.$$ 

This together with the bound for $\mathbb{E}|Z_t|^2$ provides the claimed bounds of the lemma and completes its proof. \hfill \Box

**Proof of Theorem 1.3.** By the equivalence of the norms from $(X, V)$ and $(Y, V)$, we see that

$$\mathbb{E}\left(|X_1^1 - X_1^2|^2 + |V_1^1 - V_1^2|^2\right) \leq 2 \left(1 + \frac{1}{\lambda}\right) \mathbb{E}\left(|M|_T^2 + |Z|^2\right)$$

$$\leq C\zeta(t) \mathbb{E}(|M_0|_T + |Z|^2)$$

$$\leq C\zeta(t) \mathbb{E}\left(|X_0^1 - X_0^2|^2 + |V_0^1 - V_0^2|^2\right)^{1/2} + \left(|X_1^1 - X_0^2|^2 + |V_1^1 - V_0^2|^2\right).$$

Here we used Lemma 4.3 to go between the first and second line, and to find the exponentially decreasing term $\zeta$. The constants $C$ and $C'$ come from the constants in equivalence of norms. \hfill \Box

4.2. **Optimality.** In order to show Theorem 1.4, we focus on the drift-corrected positions $Y_1^1$ and $Y_2^1$ which behave like time-rescaled Brownian motion on the torus. For their quadratic distance we prove the following decay bound.

**Proposition 2.** Suppose there exist functions $\alpha : (0, \pi L] \mapsto \mathbb{R}^+$ and $\zeta : [0, \infty) \mapsto \mathbb{R}^+$ with $\zeta \in L^1([0, \infty))$, such that, for any $z \in (0, \pi L]$ there exist two standard Brownian motions $W_t^1$ and $W_t^2$ taking values on the torus $T = \mathbb{R}/(2\pi L \mathbb{Z})$ and both adapted to a common filtration such that $|W_t^1 - W_t^2| = z$, and for $t \in \mathbb{R}^+$ it holds that

$$\mathbb{E}|W_t^1 - W_t^2|_T^2 \leq (\alpha(z))^2 \zeta(t).$$

Then with a constant $c$ only depending on $L$, the function $\alpha$ satisfies the bound

$$\alpha(z) \geq c\zeta_{L, \alpha}^{-1/2} \zeta(z).$$

From this Theorem 1.4 follows easily.

**Proof of Theorem 1.4.** Fix $z \in (0, \pi L]$ and consider the initial distributions $\mu = \delta_{X_0=0} \delta_{V_0=0}$ and $\nu = \delta_{X_0=z} \delta_{V_0=0}$. Between $\mu$ and $\nu$, there is only one coupling and $W_2(\mu, \nu) = z$.

If there exists a co-adapted coupling $((X_1^1, V_1^1), (X_2^2, V_2^2))$ satisfying the bound, then $Y_{t/\lambda_2}^1$ and $Y_{t/\lambda_2}^2$ are Brownian motions on the torus with a common filtration. Moreover,

$$\mathbb{E}|Y_t^1 - Y_t^2|_T^2 \leq C \mathbb{E}|X_t^1 - X_t^2|_T^2 + |V_1^1 - V_2^2|^2$$

for a constant $C$ only depending on $\lambda$. Hence we can apply Proposition 2 to find the claimed lower bound for $\alpha$. \hfill \Box

For the proof of Proposition 2, we first prove the following lemma.
Lemma 4.4. Given two Brownian motions \(W^1_t\) and \(W^2_t\) on the torus adapted to the same filtration, then there exists a numerical constant \(c\) such that
\[
\mathbb{E}[|W^1_t - W^2_t|^2] \geq c e^{-2t/L^2} \mathbb{E}[|W^1_0 - W^2_0|^2].
\]

Proof. The natural (squared) metric \(|x - y|^2\) on the torus is not a global smooth function of \(x, y \in \mathbb{R}\) as it takes \(x, y \mod 2\pi L\). Therefore we introduce the equivalent metric
\[
d^2_W(x, y) = L^2 \sin^2 \left( \frac{x - y}{2L} \right),
\]
which is a smooth function of \(x, y \in \mathbb{R}\). Moreover, the constants of equivalence are independent of \(L\), i.e. there exist numerical constants \(c_1\) and \(c_2\) such that
\[
c_1|x - y|^2 \leq d^2_W(x, y) \leq c_2|x - y|^2.
\]

Now consider \(H_t\) defined by
\[
H_t = L \sin \left( \frac{W^1_t - W^2_t}{2L} \right) \exp \left( \frac{[W^1_t - W^2_t]}{4L^2} \right).
\]
As \(W^1_t\) and \(W^2_t\) are Brownian motions, their quadratic variation is controlled as
\[
[W^1 - W^2]_t \leq 4t.
\]
By Itô’s lemma
\[
dH_t = \frac{1}{2} \cos \left( \frac{W^1_t - W^2_t}{2L} \right) \exp \left( \frac{[W^1_t - W^2_t]}{4L^2} \right) d(W^1_t - W^2)_t.
\]
Therefore we may bound the quadratic variation of \(H\) by
\[
[H]_t = \int_0^t \frac{1}{4} \cos^2 \left( \frac{W^1_t - W^2_t}{2L} \right) \exp \left( \frac{[W^1_t - W^2_t]}{2L^2} \right) d([W^1 - W^2]_t)
\leq \int_0^t \exp \left( \frac{2t}{L^2} \right) dt
< \infty.
\]

Therefore, as also \(|H_0| \leq L\), the local martingale \(H_t\) is a true martingale and by Jensen’s inequality
\[
\mathbb{E}[|H_t|^2] \geq \mathbb{E}[|H_0|^2].
\]
Using the equivalence of two metrics, we thus find the required bound
\[
\mathbb{E}[|W^1_t - W^2_t|^2] \geq c_2^{-1} \mathbb{E}[|H_t|^2 \exp \left( -\frac{[W^1_t - W^2_t]}{2L^2} \right)]
\geq c_2^{-1} \mathbb{E}[|H_0|^2 \exp \left( -\frac{2t}{L^2} \right)]
\geq c_1 c_2^{-1} \mathbb{E}[|W^1_0 - W^2_0|^2] \exp \left( -\frac{2t}{L^2} \right). \quad \Box
\]

With this we approach the final proof.

Proof of Proposition 2. Fix \(a \in (0, 1)\), let \(z \in (0, \pi L)\) be given, and by symmetry assume without loss of generality that \(W^1_0 - W^2_0 = |W^1_0 - W^2_0| = z\). Then define the stopping time
\[
T = \inf\{t \geq 0 : W^1_t - W^2_t \notin (az, \pi L)\}.
\]
The distance can be directly bounded as
\[
\mathbb{E}[|W^1_t - W^2_t|^2] \geq \mathbb{P}[T \geq t](az)^2.
\]
As $\zeta$ is integrable, it must decay along a subsequence of times and thus $T$ must be almost surely finite.

As $W^1_t$ and $W^2_t$, considered on $\mathbb{R}$, are continuous martingales, their difference is also a continuous martingale. By the construction of the stopping time, the stopped martingale $(W^1_t - W^2_t)_{t \wedge T}$ is bounded by $\pi L$ and the optional stopping theorem implies

$$\mathbb{P}[W^1_T - W^2_T = \pi L] = \frac{z - az}{\pi L - az}.$$ 

Since Brownian motions satisfy the strong Markov property, we find together with Lemma 4.4

$$\mathbb{E} \int_0^\infty |W^1_t - W^2_t|^2 dt \geq \mathbb{E} \int_T^\infty |W^1_t - W^2_t|^2 dt 
\geq \mathbb{P}[W^1_T - W^2_T = \pi L] \mathbb{E} \int_T^\infty |W^1_t - W^2_t|^2 dt \bigg| W^1_T - W^2_T = \pi 
\geq \mathbb{E} [W^1_T - W^2_T = \pi L] c(\pi L)^2 \int_0^\infty e^{-2t/L^2} dt 
\geq \frac{z - az}{\pi L - az} c(\pi L)^2 \frac{L^2}{2} 
\geq C_a z$$

for a constant $C_a$ only depending on $a$ and $L$, where the strong Markov property and then the lemma are applied on the second line.

On the other hand, integrating the assumed bound gives

$$\mathbb{E} \int_0^\infty |W^1_t - W^2_t|^2 dt \leq (\alpha(z))^2 \int_0^\infty \zeta(t) dt \leq (\alpha(z))^2 \|\zeta\|_{L^1([0,\infty))}.$$ 

Hence

$$C_a z \leq (\alpha(z))^2 \|\zeta\|_{L^1([0,\infty))}$$

which is the claimed result. \qed

**Acknowledgments.** The authors would like to thank Clément Mouhot for the initial discussion to look into the problem, and José A. Carrillo for discussion relating to the (lack of) a gradient flow representation of the kinetic Fokker-Planck equation.

**REFERENCES**

[1] F. Bolley, I. Gentil and A. Guillin, Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations, *J. Funct. Anal.*, **263** (2012), 2430–2457.

[2] F. Bolley, A. Guillin and F. Malrieu, Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation, *M2AN Math. Model. Numer. Anal.*, **44** (2010), 867–884.

[3] K. Burdzy and W. S. Kendall, Efficient Markovian couplings: Examples and counterexamples, *Ann. Appl. Probab.*, **10** (2000), 362–409.

[4] M. Chen, Optimal Markovian couplings and applications, *Acta Math. Sinica (N.S.)*, **10** (1994), 260–275; A Chinese summary appears in *Acta Math. Sinica*, **38** (1995), p575.

[5] S. Gadat and L. Miclo, Spectral decompositions and $L^2$-operator norms of toy hypocoercive semi-groups, *Kinet. Relat. Models*, **6** (2013), 317–372.

[6] F. Hérau, Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation, *Asymptot. Anal.*, **46** (2006), 349–359.

[7] F. Hérau and F. Nier, Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential, *Arch. Ration. Mech. Anal.*, **171** (2004), 151–218.

[8] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.*, **29** (1998), 1–17.
[9] W. S. Kendall, Coupling, local times, immersions, *Bernoulli*, 21 (2015), 1014–1046.
[10] K. Kuwada, Characterization of maximal Markovian couplings for diffusion processes, *Electron. J. Probab.*, 14 (2009), 633–662.
[11] S. Mischler and C. Mouhot, Exponential stability of slowly decaying solutions to the kinetic Fokker-Planck equation, *Arch. Ration. Mech. Anal.*, 221 (2016), 677–723.
[12] P. Mörters and Y. Peres, *Brownian Motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2010, URL http://books.google.co.uk/books?id=e-TbA-dSrzYC.
[13] C. Mouhot and L. Neumann, Quantitative perturbative study of convergence to equilibrium for collisional kinetic models in the torus, *Nonlinearity*, 19 (2006), 969–998.
[14] D. Talay, Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme, *Markov Process. Related Fields*, 8 (2002), 163–198, Inhomogeneous random systems (Cergy-Pontoise, 2001).
[15] C. Villani, Hypocoercivity, *Mem. Amer. Math. Soc.*, 202 (2009), iv+141pp.
[16] C. Villani, *Optimal Transport. Old and New*, vol. 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009.
[17] R. Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, USA, 2001, URL https://books.google.co.uk/books?id=4cI5136OdoMC.

Received April 2016; revised July 2017.

E-mail address: helge.dietert@imj-prg.fr
E-mail address: jahe2@cam.ac.uk
E-mail address: t.holding@imperial.ac.uk