

# COUPLINGS AND QUANTITATIVE CONTRACTION RATES FOR LANGEVIN DYNAMICS<sup>1</sup>

BY ANDREAS EBERLE\*, ARNAUD GUILLIN<sup>†</sup> AND RAPHAEL ZIMMER\*

*University of Bonn\** and *Université Blaise Pascal<sup>†</sup>*

We introduce a new probabilistic approach to quantify convergence to equilibrium for (kinetic) Langevin processes. In contrast to previous analytic approaches that focus on the associated kinetic Fokker–Planck equation, our approach is based on a specific combination of reflection and synchronous coupling of two solutions of the Langevin equation. It yields contractions in a particular Wasserstein distance, and it provides rather precise bounds for convergence to equilibrium at the borderline between the overdamped and the underdamped regime. In particular, we are able to recover kinetic behaviour in terms of explicit lower bounds for the contraction rate. For example, for a rescaled double-well potential with local minima at distance  $a$ , we obtain a lower bound for the contraction rate of order  $\Omega(a^{-1})$  provided the friction coefficient is of order  $\Theta(a^{-1})$ .

**1. Introduction.** Suppose that  $U$  is a function in  $C^1(\mathbb{R}^d)$  such that  $\nabla U$  is Lipschitz continuous, and let  $u, \gamma \in (0, \infty)$ . We consider a (kinetic) Langevin diffusion  $(X_t, V_t)_{t \geq 0}$  with state space  $\mathbb{R}^{2d}$  that is given by the stochastic differential equation

$$(1.1) \quad \begin{aligned} dX_t &= V_t dt, \\ dV_t &= -\gamma V_t dt - u \nabla U(X_t) dt + \sqrt{2\gamma u} dB_t. \end{aligned}$$

Here,  $(B_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion that is defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Since the coefficients are Lipschitz continuous, a unique strong solution of the Langevin equation exists for any initial condition, and the solution gives rise to a strong Markov process with generator

$$(1.2) \quad \mathcal{L} = u\gamma \Delta_v - \gamma v \cdot \nabla_v - u \nabla U(x) \cdot \nabla_v + v \cdot \nabla_x.$$

The corresponding Kolmogorov forward equation is the *kinetic Fokker–Planck equation*. Under the assumptions on  $U$  imposed below, it can be verified that

Received March 2017; revised June 2018.

<sup>1</sup>Supported by DAAD and the French government through the PROCOPE program, and by the German Science Foundation through the *Hausdorff Center for Mathematics*.

*MSC2010 subject classifications.* 60J60, 60H10, 35Q84, 35B40.

*Key words and phrases.* Langevin diffusion, kinetic Fokker–Planck equation, stochastic Hamiltonian dynamics, reflection coupling, convergence to equilibrium, hypocoercivity, quantitative bounds, Wasserstein distance, Lyapunov functions.

$\exp(-U) \in L^1(\mathbb{R}^d)$ , and that the probability measure

$$(1.3) \quad \mu_*(dx dv) = \mathcal{Z}^{-1} e^{-U(x) - \frac{|v|^2}{2u}} dx dv, \quad \mathcal{Z} = (2\pi u)^{d/2} \int e^{-U(x)} dx,$$

is invariant for the transition semigroup  $(p_t)_{t \geq 0}$ ; see, for example, [36], Proposition 6.1.

In statistical physics, the Langevin equation (1.1) describes the motion of a particle with position  $X_t$  and velocity  $V_t$  in a force field  $b = -\nabla U$  subject to damping and random collisions [18, 28, 35, 38, 43]. In the physical interpretation,  $\gamma$  is the friction coefficient (per unit mass), and  $u$  is the inverse mass. Discretizations of the Langevin equation are relevant for molecular dynamics simulations [29]. Hamiltonian Monte Carlo methods for sampling and integral estimation are based on different types of discrete time analogues to Langevin dynamics [6, 12, 29, 34]. In numerical simulations, often a better performance of these HMC methods compared to traditional MCMC approaches is observed, but the corresponding convergence acceleration is still not well understood theoretically.

For these and other reasons, an important question is how to obtain explicit bounds on the speed of convergence of the law of  $(X_t, V_t)$  toward the invariant probability measure  $\mu_*$ . Since the noise is only acting on the second component, the generator of the Langevin diffusion is degenerate, and thus classical approaches cannot be applied in a straightforward way. Indeed,  $\mathcal{L}$  is a typical example of a hypocoercive operator in the sense of Villani [40, 42]. Several analytic approaches to convergence to equilibrium for kinetic Fokker–Planck equations have been proposed during the last 15 years [3, 7, 10, 11, 17, 19, 20, 25–27, 33, 40, 42]. These are based respectively on Witten Laplacians and functional inequalities, semigroup theory, and in particular on hypocoercivity methods; see also [21] for some explorations around the Gaussian case and the effect of hypoellipticity. There are only few articles which study the ergodic properties of Langevin processes using more probabilistic arguments; cf. [1, 5, 32, 37, 39, 46]. Most of these results ultimately rely on arguments used in Harris-type theorems, that is, they assume a Lyapunov drift condition which implies recurrence of the process w.r.t. a compact set together with a control over the average excursion length. This condition is then combined with an argument showing that for starting points in the recurrent set, the transition probabilities are not singular w.r.t. each other. While the approaches are of a probabilistic nature, the behaviour of the process inside the recurrent set is not very transparent. Correspondingly, these approaches lead to qualitative rather than quantitative convergence results.

An open question asked by Villani ([41], Chapter 2, Bibliographical notes) is how to prove exponential convergence to equilibrium by a direct *coupling approach*. The motivation for this is two-fold: On the one hand, coupling methods often provide a good probabilistic understanding of the dynamics. On the other hand, couplings have been proven useful in establishing precise bounds on the long-time behaviour of nondegenerate diffusion processes [9, 14, 15, 30]. The only

results for Langevin processes in this direction that we are aware of are rather restrictive: Under the assumption that the force field  $\nabla U$  is a small perturbation of a linear function, Bolley, Guillin and Malrieu [5] use a synchronous coupling to show exponential mixing for (1.1) in  $L^2$  Wasserstein distances. Moreover, in [2, 4], couplings for the Kolmogorov diffusion have been considered. This process solves an equation similar to (1.1) without damping and with  $U \equiv 0$ .

Here, we develop a novel coupling approach for Langevin equations that works for a much wider class of force fields. We briefly describe the main ideas behind this approach: A coupling of two solutions of (1.1) is given by stochastic processes  $(X_t, V_t)_{t \geq 0}$  and  $(X'_t, V'_t)_{t \geq 0}$  with state space  $\mathbb{R}^{2d}$  that are defined on a common probability space and satisfy (1.1) and, respectively,

$$(1.4) \quad \begin{aligned} dX'_t &= V'_t dt, \\ dV'_t &= -\gamma V'_t dt - u \nabla U(X'_t) dt + \sqrt{2\gamma u} dB'_t, \end{aligned}$$

where  $(B_t)_{t \geq 0}$  and  $(B'_t)_{t \geq 0}$  are  $d$ -dimensional Brownian motions. The only freedom in constructing a coupling is the way these Brownian motions are related to each other. For a synchronous coupling,  $B_t = B'_t$  for all  $t$ . In this case, the difference process  $(Z_t, W_t) = (X_t - X'_t, V_t - V'_t)$  satisfies a deterministic o.d.e., and contractivity holds if and only if it holds for the equation without noise. This applies for example for overdamped Langevin diffusions in a strictly convex potential or on a positively curved Riemannian manifold, but in general it is a rather restrictive condition that is not satisfied in our case. Nevertheless, one can observe that w.r.t. an appropriately chosen metric on  $\mathbb{R}^{2d}$ , the difference process is contractive without noise as long as it is in a neighbourhood of the hyperplane where  $Q_t := Z_t + \gamma^{-1}W_t = 0$ ; see Section 2.1 below. Therefore, synchronous coupling can be applied in this region.

If the dynamics is not contractive, one has to exploit the random fluctuations to ensure that the two copies approach each other in some sense. A well-known approach is reflection coupling [30] where the noise increments  $dB_t$  and  $dB'_t$  are synchronized in directions orthogonal to the difference of the two copies and reflected in the direction connecting the copies. As a consequence, the difference process is driven by a one-dimensional noise in this direction. It has been shown in [8, 13–15] that this can be exploited to obtain average contractivity with relatively sharp explicit rates in distances that are appropriately chosen concave functions of  $\ell^1$  or  $\ell^2$  metrics. This approach works well for nondegenerate diffusions but it fails for the degenerate case. Therefore it does not apply directly to the Langevin equation. Nevertheless, it can be used in the directions complementary to the contractive hyperplane.

Combining the two types of couplings above suggests that we should apply a coupling that is synchronous whenever  $Q_t$  equals 0 (or is close to 0), and a reflection coupling in the complementary directions otherwise. This means we should

set

$$(1.5) \quad dB'_t = (I_d - 1_{\{Q_t \neq 0\}} 2e_t e_t^T) dB_t,$$

where  $e_t = Q_t/|Q_t|$ . Then the resulting coupling difference process will be driven by noise whenever  $Q_t \neq 0$ , and the noise will be switched off if  $Q_t = 0$ . Lévy's characterization ensures that  $(B'_t)$  is again a Brownian motion, and the resulting coupling process is a diffusion process on  $\mathbb{R}^{4d}$  that is *sticky* [16, 44, 45] on the subspace  $\{(x, v, x', v') \in \mathbb{R}^{4d} : x - x' + \gamma^{-1}(v - v') = 0\}$  where contractivity holds without noise. This means that almost surely, after reaching the subspace due to reflection coupling, the process spends a positive amount of time on this subspace although it does not stay on the subspace for any positive time-interval. Each time it leaves the subspace, it immediately returns due to the random fluctuations that are switched on when  $Q_t \neq 0$ . In total, the set  $\{t \in [0, \infty) : Q_t = 0\}$  of all times where the process visits the subspace has almost surely positive Lebesgue measure although it does not contain any nonempty open interval. The rigorous construction of a corresponding sticky coupling can be carried out by a weak convergence approach that is based on approximating the discontinuous coefficients in (1.5) by Lipschitz continuous functions. This has been done in a slightly different setup in [16]. In general, the stochastic differential equation for the corresponding sticky coupling process does not have a strong solution but the approximation procedure yields a weak solution; see [16] for details.

Since the construction and control of the sticky coupling described above is possible but technically involved, we actually do not consider the sticky coupling itself here. Instead, we use approximations of such a coupling in order to derive bounds for contraction rates; see (3.1) below. The corresponding limit is taken only in the resulting bounds, and the construction of the sticky coupling itself (i.e., the limit of the approximating coupling processes) is not required for our results. The speed of convergence is then measured in Kantorovich distances ( $L^1$  Wasserstein distances) by adapting and optimizing the underlying (semi)metric w.r.t. the given model and the chosen coupling. Here, we basically follow the strategy developed in [15] which extends the results in [13, 14]. The approach taken in [15], which is partially based on ideas from [8, 23, 24], is to build a multiplicative semimetric  $\rho$  out of a concave function of the underlying distance and a Lyapunov function that ensures contractivity at large distances; see Section 2.3 below. In a slight modification of (1.5), we will apply synchronous coupling at large distances, since here a Lyapunov drift condition will ensure contractivity. Both the concave functions and the constants entering the definition of the metric  $\rho$  [see (2.10) and (2.9)] are carefully chosen in order to optimize the order of the resulting contraction rates.

The approach for studying long-time stability properties of diffusion processes by using sticky couplings (or corresponding approximations) seems to be useful in many different contexts; see [16]. For example, it is also related to the application of a similar strategy to infinite-dimensional stochastic differential equations with

possibly degenerate noise in [47]; cf. also [22, 31]. In general, the idea is to identify some submanifold of the state space for the coupling where contraction properties hold for the equation without noise. Then synchronous coupling can be applied on this submanifold whereas outside, random fluctuations introduced by a different coupling ensure that the process reaches the submanifold in finite time.

Besides providing an intuitive understanding for the mechanism of convergence to equilibrium, the coupling approach yields both qualitatively new, and explicit quantitative results in several cases of interest. Before explaining the coupling construction and stating the results in detail, we illustrate this by an example.

**EXAMPLE 1.1 (Double-well potential).** Suppose that  $U \in C^1(\mathbb{R})$  is a Lipschitz continuous double-well potential defined by

$$U(x) = \begin{cases} (|x| - 1)^2/2 & \text{for } |x| \geq 1/2, \\ 1/4 - |x|^2/2 & \text{for } |x| \leq 1/2, \end{cases}$$

and let  $U_a(x) = U(x/a)$  be the rescaled potential with the same height for the potential well, but minima at distance  $2a$ . Then our main result shows that for any  $a, u, \gamma \in (0, \infty)$  there exist a constant  $c \in (0, \infty)$ , a semimetric  $\rho$  on  $\mathbb{R}^{2d}$ , and a corresponding Kantorovich semimetric  $\mathcal{W}_\rho$  such that for all probability measures  $\mu, \nu$  on  $\mathbb{R}^{2d}$ ,

$$\mathcal{W}_\rho(\mu_{p_t}, \nu_{p_t}) \leq e^{-ct} \mathcal{W}_\rho(\mu, \nu) \quad \text{for any } t \geq 0.$$

As a consequence, we also obtain convergence to equilibrium in the standard  $L^2$  Wasserstein distance with the same exponential rate  $c$ . Below, we give explicit lower bounds for the contraction rate. For example, if  $\gamma a \geq \sqrt{30u}$  then

$$c \geq \frac{\sqrt{u}}{107} \min((\gamma a)^{-4} u^2, e^{-8} (\gamma a)^{-2} u, 2^{-3/2} e^{-8}) a^{-1};$$

see Example 2.12 (with parameters  $\mathcal{R} = 4a$ ,  $L = a^{-2}$  and  $\beta = L\mathcal{R}^2/2 = 8$ ). In general, if the value of  $\gamma a$  and  $u$  are fixed (i.e., the friction coefficient  $\gamma$  is adjusted to the potential) then the contraction rate is of order  $\Omega(a^{-1})$ , that is,  $c \geq c_0 \cdot a^{-1}$  for a positive constant  $c_0$ . This clearly reflects the *kinetic behaviour*, and it is in contrast to the rate  $O(a^{-2})$  for convergence to equilibrium of the overdamped limit  $dX_t = -u\nabla U(X_t) dt + \sqrt{2u} dB_t$ .

## 2. Main results.

**2.1. Coupling construction.** We first explain the construction of the coupling briefly; see Section 3 for full details. Suppose that  $((X_t, V_t), (X'_t, V'_t))$  is an arbitrary coupling of two solutions of the Langevin equation (1.1) driven by Brownian

motions  $(B_t)$  and  $(B'_t)$ . Then the difference process  $(Z_t, W_t) = (X_t - X'_t, V_t - V'_t)$  satisfies the stochastic differential equation

$$dZ_t = W_t dt,$$

$$dW_t = -\gamma W_t dt - u(\nabla U(X_t) - \nabla U(X'_t)) dt + \sqrt{2\gamma u} d(B - B')_t.$$

Introducing the new coordinates  $Q_t = Z_t + \gamma^{-1}W_t$ , the system takes the form

$$(2.1) \quad dZ_t = -\gamma Z_t dt + \gamma Q_t dt,$$

$$(2.2) \quad dQ_t = -u\gamma^{-1}(\nabla U(X_t) - \nabla U(X'_t)) dt + \sqrt{2u\gamma^{-1}} d(B - B')_t.$$

Since  $\gamma > 0$ , the first equation is contractive if  $Q_t = 0$ . The key idea is now to apply a synchronous coupling whenever  $Q_t = 0$ , and a reflection coupling if  $Q_t \neq 0$  and  $\alpha|Z_t| + |Q_t| < R_1$  with appropriate constants  $\alpha, R_1 \in (0, \infty)$ ; cf. Figure 1. The synchronous coupling guarantees that the noise coefficient in (2.2) vanishes if  $Q_t = 0$ , that is, the dynamics is not driven away from the “contractive region” by random fluctuations (although it may leave this region by the drift). On the other hand, the reflection coupling for  $Q_t \neq 0$  ensures that the contractive region is recurrent. The resulting coupling process is a diffusion on  $\mathbb{R}^{4d}$  that is sticky on the  $3d$ -dimensional hyperplane  $\{(x, v, x', v') \in \mathbb{R}^{4d} : x - x' + \gamma^{-1}(v - v') = 0\}$  of contractive states, that is, it spends a positive amount of time in this region; cf.

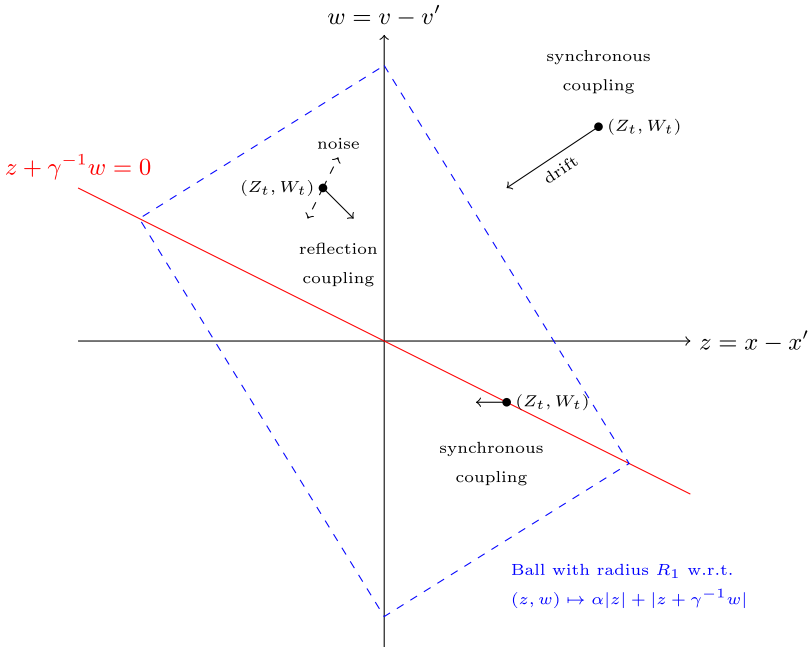


FIG. 1. Sketch of coupling approach.

[16]. Since the construction and control of the sticky couplings described above is possible but technically involved, we actually use approximations of such couplings to derive our results; see (3.1). By designing a special semimetric  $\rho$  on  $\mathbb{R}^{2d}$  that is based on a concave function of the distance and on a Lyapunov function, we can then (similarly as in [15]) make use of the random fluctuations and of a drift condition in order to derive average contractivity for the coupling distance  $\rho((X_t, V_t), (X'_t, V'_t))$ . The construction of a coupling and the proof of contractivity are carried out rigorously in Sections 3 and 4.

2.2. *Drift condition and Lyapunov function.* We now make the following assumption that guarantees, among other things, that the process is nonexplosive.

ASSUMPTION 2.1. *There exist constants  $L, A \in (0, \infty)$  and  $\lambda \in (0, 1/4]$  such that*

$$(2.3) \quad U(x) \geq 0 \quad \text{for all } x \in \mathbb{R}^d,$$

$$(2.4) \quad |\nabla U(x) - \nabla U(y)| \leq L|x - y| \quad \text{for all } x, y \in \mathbb{R}^d \quad \text{and}$$

$$(2.5) \quad x \cdot \nabla U(x)/2 \geq \lambda(U(x) + u^{-1}\gamma^2|x|^2/4) - A \quad \text{for all } x \in \mathbb{R}^d.$$

Notice that the assumption can only be satisfied if

$$(2.6) \quad \lambda \leq 2Lu\gamma^{-2}.$$

Up to the choice of the constants, the drift condition (2.5) is equivalent to the simplified drift condition (2.26) considered further below. It implies the existence of a Lyapunov function for the Langevin process. Indeed, let

$$(2.7) \quad \mathcal{V}(x, v) = U(x) + \frac{1}{4}u^{-1}\gamma^2(|x + \gamma^{-1}v|^2 + |\gamma^{-1}v|^2 - \lambda|x|^2).$$

Note that since  $\lambda \leq 1/4$ ,

$$(2.8) \quad \begin{aligned} \mathcal{V}(x, v) &\geq U(x) + \frac{1}{4}(1 - 2\lambda)u^{-1}\gamma^2(|x + \gamma^{-1}v|^2 + |\gamma^{-1}v|^2) \\ &\geq \frac{1}{8}(1 - 2\lambda)u^{-1}\gamma^2|x|^2. \end{aligned}$$

In particular,  $\mathcal{V}(x, v) \rightarrow \infty$  as  $|(x, v)| \rightarrow \infty$ . Moreover, we have the following.

LEMMA 2.2. *If the drift condition (2.5) holds, then  $\mathcal{L}\mathcal{V} \leq \gamma(d + A - \lambda\mathcal{V})$ .*

The proof of the lemma is included in the [Appendix](#). The choice of the Lyapunov function is motivated by Mattingly, Stuart and Higham [32]; see also [1, 39, 46]. In combination with (2.8), the lemma shows that the process  $\mathcal{V}(X_t, V_t)$  is decreasing on average in regions where

$$|X_t| \geq 8^{1/2}(d + A)^{1/2}u^{1/2}\gamma^{-1}(\lambda - 2\lambda^2)^{-1/2}.$$

2.3. *Choice of metric.* Next, we introduce an appropriate semimetric on  $\mathbb{R}^d$  w.r.t. which the coupling considered below will be contractive on average. Inspired by [23], a similar semimetric has been considered in [15]. For  $(x, v), (x', v') \in \mathbb{R}^{2d}$ , we set

$$(2.9) \quad r((x, v), (x', v')) = \alpha|x - x'| + |x - x' + \gamma^{-1}(v - v')|,$$

$$(2.10) \quad \rho((x, v), (x', v')) = f(r((x, v), (x', v'))) \cdot (1 + \varepsilon\mathcal{V}(x, v) + \varepsilon\mathcal{V}(x', v')),$$

where  $\alpha, \varepsilon \in (0, \infty)$  are appropriately chosen positive constants, and  $f : [0, \infty) \rightarrow [0, \infty)$  is a continuous, nondecreasing concave function such that  $f(0) = 0$ ,  $f$  is  $C^2$  on  $(0, R_1)$  for some constant  $R_1 \in (0, \infty)$  with right-sided derivative  $f'_+(0) = 1$  and left-sided derivative  $f'_-(R_1) > 0$ , and  $f$  is constant on  $[R_1, \infty)$ . The function  $f$  and the constants  $\alpha, \varepsilon$  and  $R_1$  will be chosen explicitly below in order to optimize the resulting contraction rates. For the moment, let us just note that by concavity,

$$(2.11) \quad \min(r, R_1)f'_-(R_1) \leq f(r) \leq \min(r, f(R_1)) \leq \min(r, R_1) \quad \text{for } r \geq 0.$$

For probability measures  $\mu, \nu$  on  $\mathbb{R}^{2d}$ , we define

$$(2.12) \quad \mathcal{W}_\rho(\mu, \nu) = \inf_{\Gamma \in \Pi(\mu, \nu)} \int \rho((x, v), (x', v')) \Gamma(d(x, v) d(x', v')),$$

where the infimum is over all couplings of  $\mu$  and  $\nu$ . We remark that  $\rho$  and the transportation cost  $\mathcal{W}_\rho$  are semimetrics but not necessarily metrics, that is, the triangle inequality may be violated. An important remark is that the distance  $r$  can be controlled by the Lyapunov function. Indeed, let

$$(2.13) \quad R_1 := (16 \cdot (6/5) \cdot (1 + 2\alpha + 2\alpha^2)(d + A)u\gamma^{-2}(\lambda - 2\lambda^2)^{-1})^{1/2}.$$

By (2.8) and since  $U \geq 0$ ,

$$(2.14) \quad \begin{aligned} & r((x, v), (x', v'))^2 \\ & \leq ((1 + \alpha)|x - x' + \gamma^{-1}(v - v')| + \alpha|\gamma^{-1}(v - v')|)^2 \\ & \leq 2((1 + \alpha)^2 + \alpha^2) \\ & \quad \times (|x + \gamma^{-1}v|^2 + |x' + \gamma^{-1}v'|^2 + |\gamma^{-1}v|^2 + |\gamma^{-1}v'|^2) \\ & \leq 8((1 + \alpha)^2 + \alpha^2)(1 - 2\lambda)^{-1}u\gamma^{-2}(\mathcal{V}(x, v) + \mathcal{V}(x', v')) \end{aligned}$$

for any  $(x, v), (x', v') \in \mathbb{R}^{2d}$ . Hence for  $r((x, v), (x', v')) \geq R_1$ ,

$$(2.15) \quad \mathcal{V}(x, v) + \mathcal{V}(x', v') \geq \frac{12}{5}(d + A)/\lambda \quad \text{and thus}$$

$$(2.16) \quad \mathcal{L}\mathcal{V}(x, v) + \mathcal{L}\mathcal{V}(x', v') \leq -\frac{1}{6}\gamma\lambda(\mathcal{V}(x, v) + \mathcal{V}(x', v'))$$



by Lemma 2.2. The bound (2.16) guarantees that for the coupling to be considered below, the process  $\rho_t := \rho((X_t, V_t), (X'_t, V'_t))$  is decreasing on average if  $r_t := r((X_t, V_t), (X'_t, V'_t)) \geq R_1$ . We show that by choosing the coupling and the parameters  $\alpha, \varepsilon$  and  $f$  defining the metric in an adequate way, we can ensure that  $\rho_t$  is also decreasing on average (up to a small error term) for  $r_t < R_1$ . As a consequence, we will obtain our basic contraction result.

2.4. *Main contraction result.* We can now state our main result.

**THEOREM 2.3.** *Suppose that Assumption 2.1 is satisfied. Then there exist constants  $\alpha, \varepsilon \in (0, \infty)$  and a continuous nondecreasing concave function  $f : [0, \infty) \rightarrow [0, \infty)$  with  $f(0) = 0$  such that for all probability measures  $\mu, \nu$  on  $\mathbb{R}^{2d}$ ,*

$$(2.17) \quad \mathcal{W}_\rho(\mu p_t, \nu p_t) \leq e^{-ct} \mathcal{W}_\rho(\mu, \nu) \quad \text{for any } t \geq 0,$$

where the contraction rate  $c$  is given by

$$(2.18) \quad c = \frac{\gamma}{384} \min(\lambda Lu\gamma^{-2}, \Lambda^{1/2} e^{-\Lambda} Lu\gamma^{-2}, \Lambda^{1/2} e^{-\Lambda}) \quad \text{with}$$

$$(2.19) \quad \Lambda := LR_1^2/8 = \frac{12}{5}(1 + 2\alpha + 2\alpha^2)(d + A)Lu\gamma^{-2}\lambda^{-1}(1 - 2\lambda)^{-1}.$$

Explicitly, one can choose the constants  $\alpha, \varepsilon$ , and the function  $f$  determining  $\rho$  in such a way that

$$(2.20) \quad \alpha = (1 + \Lambda^{-1})Lu\gamma^{-2} \leq \frac{11}{6}Lu\gamma^{-2}, \quad \varepsilon = 4\gamma^{-1}c/(d + A),$$

and  $f$  is constant on  $[R_1, \infty)$  and  $C^2$  on  $(0, R_1)$  with

$$(2.21) \quad \frac{1}{2}e^{-2} \exp(-Lr^2/8) \leq f'(r) \leq \exp(-Lr^2/8) \quad \text{for } r \in (0, R_1).$$

More precisely,  $f$  is defined by (4.2), (4.3) and (4.4) below.

**REMARK 2.4.** The constant  $\Lambda$  depends on the parameters  $L, u$  and  $\gamma$  both explicitly and through  $\lambda$  and  $\alpha$ . By (2.6), we always have

$$(2.22) \quad \Lambda \geq 6(d + A)/5 \geq 6/5.$$

Corresponding upper bounds are given in Lemma 2.8 below.

**REMARK 2.5.** We shortly comment on the requirement that  $\nabla U$  is Lipschitz; cf. Assumption 2.1 further above. This condition is not necessary to conclude exponential convergence to equilibrium in Kantorovich distances; cf. [32], Theorem 3.2. We have chosen to limit ourselves here to the Lipschitz case to concentrate on the key techniques rather than on tedious calculations. In the case of

overdamped Langevin equations, the contraction results from [15] are extended in [48] replacing global Lipschitz bounds by local ones. In a similar spirit, it might be possible to extend the results presented here. However, optimizing and keeping track of the constants is more involved in this case.

The proof of Theorem 2.3 is given in Section 5. As a preparation, we introduce the relevant couplings in Section 3, and we apply these to derive a more general contraction result in Section 4. Theorem 2.3 will be obtained from this more general result by choosing the constants  $\alpha$  and  $\varepsilon$  in a specific way.

Theorem 2.3 directly implies convergence of the Langevin process to a unique stationary distribution with exponential rate  $c$ .

**COROLLARY 2.6.** *In the setting of Theorem 2.3, there exists a constant  $C \in (0, \infty)$  such that for all probability measures  $\mu, \nu$  on  $\mathbb{R}^{2d}$ ,*

$$(2.23) \quad \mathcal{W}^2(\mu p_t, \nu p_t)^2 \leq C e^{-ct} \mathcal{W}_\rho(\mu, \nu) \quad \text{for any } t \geq 0.$$

Here,  $\mathcal{W}^2$  denotes the standard  $L^2$  Wasserstein distance w.r.t. the Euclidean metric. In particular,  $\mu_*$  is the unique invariant probability measure for the Langevin process, and  $\mu p_t$  converges towards  $\mu_*$  exponentially fast with rate  $c$  for any initial law  $\mu$  such that  $\mathcal{W}_\rho(\mu, \mu_*) < \infty$ . Here, the constant  $c$  and the semimetric  $\rho$  are given as in Theorem 2.3, and the constant  $C$  can be chosen explicitly as

$$C = 2e^{2+\Lambda} \frac{(1 + \gamma)^2}{\min(1, \alpha)^2} \max\left(1, 4(1 + 2\alpha + 2\alpha^2) \frac{(d + A)u\gamma^{-1}c^{-1}}{\min(1, R_1)}\right).$$

The proof is given in Section 5.

**2.5. Bounds under simplified drift condition.** In order to make the dependence of the bounds on the parameters more explicit, we now replace (2.5) by a simplified drift condition. Instead of Assumption 2.1, we assume the following.

**ASSUMPTION 2.7.** *There exist constants  $L, \mathcal{R}, \beta \in (0, \infty)$  such that*

$$(2.24) \quad U(0) = 0 = \min U,$$

$$(2.25) \quad |\nabla U(x) - \nabla U(y)| \leq L|x - y| \quad \text{for any } x, y \in \mathbb{R}^d \quad \text{and}$$

$$(2.26) \quad x \cdot \nabla U(x) \geq \beta \cdot (|x|/\mathcal{R})^2 \quad \text{for any } x \in \mathbb{R}^d \text{ s.t. } |x| \geq \mathcal{R}.$$

Observe that (2.24) may be assumed w.l.o.g. by subtracting a constant and shifting the coordinate system such that the global minimum of  $U$  [which exists if (2.26) holds] is attained at 0. The Lipschitz condition (2.25) has been assumed before, and up to the values of the constants, the drift condition (2.26) is equivalent

to (2.5). This condition guarantees that the  $x$  marginal of the invariant probability measure (1.3) concentrates on balls of radius  $O(\mathcal{R})$ . Notice that if (2.25) and (2.26) are both satisfied then

$$(2.27) \quad \beta \leq L\mathcal{R}^2.$$

LEMMA 2.8. *Suppose that Assumption 2.7 is satisfied. Then Assumption 2.1 holds with*

$$(2.28) \quad A = (L\mathcal{R}^2 - \beta)/8 \quad \text{and} \quad \lambda = \min\left(\frac{1}{4}, \frac{\beta}{L\mathcal{R}^2} \cdot \frac{2Lu\gamma^{-2}}{1 + 2Lu\gamma^{-2}}\right).$$

Furthermore, if  $Lu\gamma^{-2} \leq 1/8$  then the constant  $\Lambda$  in Theorem 2.3 is bounded by

$$(2.29) \quad \frac{6}{5}(d + A)L\mathcal{R}^2/\beta \leq \Lambda \leq \frac{6}{5}(d + A)(1 + 20Lu\gamma^{-2})L\mathcal{R}^2/\beta.$$

In general, there is an explicit constant  $C_1 \in (0, \infty)$  such that

$$(2.30) \quad \frac{6}{5}(d + A)L\mathcal{R}^2/\beta \leq \Lambda \leq \frac{12}{5}(d + A)(1 + C_1Lu\gamma^{-2})^3L\mathcal{R}^2/\beta.$$

The proof is included in the [Appendix](#). The lemma shows that if Assumption 2.7 holds with fixed constants  $L, \mathcal{R}, \beta \in (0, \infty)$ , then there is a lower bound for the contraction rate in Theorem 2.3 that only depends on the natural parameters  $\gamma, Lu\gamma^{-2}, L\mathcal{R}^2$  and  $\beta$ . The bound is particularly nice if there is sufficient damping.

COROLLARY 2.9. *Let  $\ell \in [1, \infty)$ , and suppose that Assumption 2.7 is satisfied with constants  $L, \mathcal{R}, \beta \in (0, \infty)$  such that  $\beta \geq L\mathcal{R}^2/\ell$ . Suppose further that  $Lu\gamma^{-2} \leq 1/30$ . Then the assertion of Theorem 2.3 holds with a contraction rate*

$$(2.31) \quad \begin{aligned} c &\geq \frac{\gamma}{205} \min\left(\frac{1}{\ell}(Lu\gamma^{-2})^2, \frac{1}{2} \min(d^{1/2}Lu\gamma^{-2}, \Lambda_1^{-1/2})e^{-\Lambda_1}\right) \\ &\geq \frac{\sqrt{\beta u}}{38} \min\left(\frac{1}{\ell}(Lu\gamma^{-2})^2, \frac{1}{2} \min(d^{1/2}Lu\gamma^{-2}, \Lambda_1^{-1/2})e^{-\Lambda_1}\right)\mathcal{R}^{-1}, \end{aligned}$$

where  $\Lambda_1 := (\ell - 1)L\mathcal{R}^2/4 + 2\ell d$ .

The proof of the corollary is given in Section 5. Bounds of the same order as in (2.31) hold if the constant  $1/30$  is replaced by any other strictly positive constant. The specific value  $1/30$  has been chosen in a somehow ad hoc way in order to obtain relatively small constants in the prefactors.

REMARK 2.10 (Kinetic behaviour, Hamiltonian Monte Carlo). Corollary 2.9 shows that by adjusting the friction coefficient  $\gamma$  appropriately, one can obtain a kinetic lower bound for the contraction rate: If  $\gamma$  is chosen such that the value of  $Lu\gamma^{-2}$  is a given constant, and the parameters  $\beta$  and  $\ell$  are fixed as well (i.e.,  $L\mathcal{R}^2$

is within a fixed range), then the lower bound for  $c$  in (2.31) is of order  $\Omega(\mathcal{R}^{-1})$ . This should be relevant for MCMC methods based on discretizations of Langevin equations [6, 29, 34], because it indicates that by adjusting  $\gamma$  appropriately, one can improve on the diffusive order  $O(\mathcal{R}^{-2})$  for the convergence rate to equilibrium.

Before discussing the parameter dependence of the lower bounds for the contraction rate  $c$  that have been stated above, we check the quality of the bounds in the linear case and for drifts that are linear outside a ball.

EXAMPLE 2.11 (Linear drift). Suppose that  $U(x) = L|x|^2/2$ . Then (1.1) reads

$$(2.32) \quad dX_t = V_t dt, \quad dV_t = -\gamma V_t dt - uLX_t dt + \sqrt{2\gamma u} dB_t.$$

Applying Corollary 2.9 with  $\beta = L\mathcal{R}^2$  and  $\ell = 1$  shows that for  $Lu\gamma^{-2} \leq 1/30$ ,

$$(2.33) \quad c \geq \frac{\gamma}{205} \min\left((Lu\gamma^{-2})^2, \frac{1}{2}e^{-2d} \min(d^{1/2}Lu\gamma^{-2}, (2d)^{-1/2})\right).$$

Lower bounds of a similar order can be derived from Theorem 2.3 if  $Lu\gamma^{-2}$  is bounded from above by a fixed constant. On the other hand, the linear Langevin equation (2.32) can be solved explicitly. The solution is a Gaussian process. By [36], Section 6.3, the  $L^2$  spectral gap of the corresponding generator is

$$(2.34) \quad c_{\text{gap}} = (1 - \sqrt{(1 - 4Lu\gamma^{-2})^+})\gamma/2 \quad \text{and, in particular,}$$

$$(2.35) \quad \gamma \min(1/4, Lu\gamma^{-2}) \leq c_{\text{gap}} \leq \gamma \min(1/2, 2Lu\gamma^{-2}).$$

The spectral gap provides an upper bound for the contraction rate  $c$ . For example, for  $d = 1$  and  $Lu\gamma^{-2} = 1/30$ , we obtain the lower bound

$$(2.36) \quad c \geq \frac{\gamma}{184,500} \approx \frac{1}{33,685}(Lu)^{1/2}$$

for the contraction rate, whereas the upper bound given by the spectral gap is

$$c_{\text{gap}} = \frac{1}{2}\gamma(1 - \sqrt{1 - 4/30}) \approx \frac{\gamma}{29} \approx \frac{1}{5.3}(Lu)^{1/2}.$$

EXAMPLE 2.12 (Multi-well potentials, linear drift outside a ball). Assumption 2.7 with  $\beta = L\mathcal{R}^2/2$  is satisfied for the one-dimensional double-well potential

$$U(x) = \begin{cases} L|x|^2/2 & \text{for } x \leq \mathcal{R}/8, \\ -L(x - \mathcal{R}/4)^2/2 + L\mathcal{R}^2/64 & \text{for } \mathcal{R}/8 \leq x \leq 3\mathcal{R}/8, \\ L(x - \mathcal{R}/2)^2/2 & \text{for } x \geq 3\mathcal{R}/8, \end{cases}$$

and also for the triple-well potential  $\tilde{U}(x) = U(|x|)$ . Here, for  $Lu\gamma^{-2} \leq 1/30$ , Corollary 2.9 yields the lower bounds

$$\begin{aligned}
 (2.37) \quad c &\geq \frac{\gamma}{410} \min((Lu\gamma^{-2})^2, \\
 &\quad \min(d^{1/2}Lu\gamma^{-2}, (4d + L\mathcal{R}^2/4)^{-1/2})e^{-4d-L\mathcal{R}^2/4}) \\
 &\geq \frac{\sqrt{\beta u}}{75\mathcal{R}} \min((Lu\gamma^{-2})^2, \\
 &\quad \min(d^{1/2}Lu\gamma^{-2}, (4d + L\mathcal{R}^2/4)^{-1/2})e^{-4d-L\mathcal{R}^2/4})
 \end{aligned}$$

for the contraction rate. Again, lower bounds of similar order hold if  $Lu\gamma^{-2}$  is bounded from above by a fixed constant. More generally, we obtain corresponding bounds if  $U$  is a potential satisfying conditions (2.24) and (2.25), and there exist constants  $\mathcal{R} \in \mathbb{R}_+$  and  $a \in \mathbb{R}^d$  with  $|a| \leq \mathcal{R}/2$  such that  $\nabla U(x) = L(x - a)$  for  $|x| \geq \mathcal{R}$ . The lower bound is of order  $\Theta(\mathcal{R}^{-1})$  if  $Lu\gamma^{-2}$  is fixed and  $L\mathcal{R}^2$  is bounded from above by a fixed constant.

We stress that in low dimensions, we can obtain numerical values for our lower bounds that are in reach for current computer simulations. This is quite remarkable because we have lost some factors during our estimates. For high dimensions, our bounds deteriorate rapidly.

2.6. *Parameter dependence of lower bounds for the contraction rate.* We now discuss the parameter dependence of the bounds derived above, and we compare our results to previously derived bounds on convergence to equilibrium for kinetic Fokker–Planck equations.

Let us first recall that the computation of the spectrum shows that in the linear case, there are two different regimes; cf. [36]. For  $Lu\gamma^{-2} \geq 1/4$  (*underdamped regime*), the spectral gap (2.34) is a linear function of  $\gamma$ . In this case, the friction coefficient  $\gamma$  is so small that the rate of convergence to equilibrium is determined by  $\gamma$ . Conversely, for  $Lu\gamma^{-2} \leq 1/4$ , the spectral gap is a decreasing function of  $\gamma$ . In this regime, the rate of convergence to equilibrium is determined by the transfer of noise from the  $v$ -variable to the  $x$ -variable. If  $\gamma$  increases, then the noise is damped more strongly before it can be transferred to the  $x$ -component, and hence the rate of convergence decreases. In particular, the spectral gap as a function of  $\gamma$  has a sharp maximum for  $Lu\gamma^{-2} = 1/4$ . See Figure 2.

Comparing (2.35) and (2.18), we see that our general lower bound for the contraction rate contains similar terms as the bounds in (2.35). However, these terms are multiplied by constants that again depend on the parameters  $L$ ,  $u$  and  $\gamma$ . We now discuss the parameter dependence of the lower bounds in different regimes:

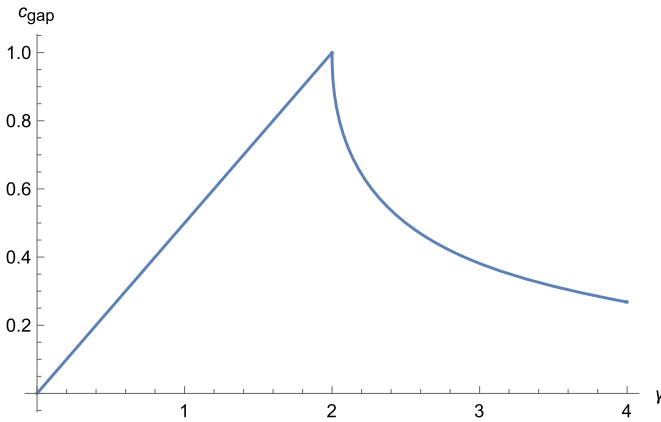


FIG. 2. Spectral gap for  $U(x) = |x|^2/2$  and  $u = 1$ .

$Lu\gamma^{-2} \rightarrow 0$  (overdamped case). As  $Lu\gamma^{-2} \rightarrow 0$ , the lower bound in (2.31) is of order  $\Omega(\gamma(Lu\gamma^{-2})^2)$ . This differs from the order  $\Theta(\gamma Lu\gamma^{-2})$  of the spectral gap in the linear case by a factor  $Lu\gamma^{-2}$ .

$Lu\gamma^{-2}$  fixed (kinetic case). If the friction coefficient  $\gamma$  is chosen such that the value of  $Lu\gamma^{-2}$  is a given constant, and the parameters  $\beta$  and  $L\mathcal{R}^2$  are fixed as well, then the lower bound in (2.31) is of order  $\Omega(\mathcal{R}^{-1})$ .

$Lu\gamma^{-2} \rightarrow \infty$  (underdamped case). For large values of  $Lu\gamma^{-2}$ , our bounds for the contraction rate in Lemma 2.8 are considerably worse than the order  $\Theta(\gamma)$  of the spectral gap in the linear case. This is not surprising, because for small  $\gamma$  it is not clear, how the contractive term  $-\gamma V_t$  in (1.1) can make up for the term  $-u\nabla U(X_t)$  with a potential  $U$  that may be locally very nonconvex.

$L\mathcal{R}^2 \rightarrow \infty$ . For large values of  $L\mathcal{R}^2$ , the lower bound in (2.31) degenerates exponentially in this parameter. This is natural because  $U$  could be a double-well potential with valleys of depth  $\Theta(L\mathcal{R}^2)$ ; see the example above.

$d \rightarrow \infty$ . Our bounds depend exponentially on the dimension. In the general setup considered here, this is unavoidable. An important open question is whether a better dimension dependence can be obtained for a restricted class of models. For overdamped Langevin diffusions, corresponding results have been obtained, for example, in [14, 47, 48].

Let us now set our results in relation to explicit bounds for convergence to equilibrium of kinetic Fokker–Planck equations that have been obtained in [11, 27, 42] by analytic methods. These results are not directly comparable, because they quantify convergence to equilibrium in different distances (e.g., in weighted  $L^2$  or Sobolev norms, or in relative entropy). Moreover, the constants have not always been tracked as precisely as here. Nevertheless, it seems plausible to compare the

orders of the convergence rates. Already in [27], Theorem 0.1, Hérau and Nier have derived a nice explicit bound on the convergence rate in weighted Sobolev spaces under quite general conditions. However, even in the linear case, this bound is far from sharp in the overdamped regime and at the boundary between the overdamped and the underdamped regime. In particular, it seems not to be able to recover kinetic behaviour for properly adjusted friction coefficients. Most of the more recent works are based on extensions of Villani's hypocoercivity approach [42]. In particular, explicit bounds are given in [42], Section 7, in  $H^1$  norms, corresponding bounds in  $L^2$  norms can be deduced from [11], and recently some Wasserstein bounds have been derived in [33]. However, we have not been able to recover a similar behaviour as above for the bounds on the convergence rates in these results.

*2.7. Outline of the proof.* To prove our main result, we proceed in the following way. In Section 3, we precisely define the coupling that we consider. Having introduced both the coupling and the underlying distance function, we study average contraction properties of the coupling distance  $\rho_t$  by standard methods from stochastic analysis. To this end, we compute the semimartingale decomposition of  $e^{ct} \rho_t$  for a given constant  $c > 0$ . Exponential contractivity with rate  $c$  holds if the resulting drift term is negative. In Section 4, we analyse under which conditions on the parameters this holds true in different regions of the state space for the coupling process. By (3.16), contractivity on the hyperplane where synchronous coupling is applied can only be expected provided  $\alpha > Lu\gamma^{-2}$ . This motivates setting  $\alpha = (1 + \eta)Lu\gamma^{-2}$  with  $\eta > 0$  in (3.11). To ensure that reflection coupling yields contractivity outside of this hyperplane, one has to choose  $f$  sufficiently concave. Intuitively, by applying a sufficiently concave function, we can turn the submartingale  $r_t = \alpha|Z_t| + |Q_t|$  into a supermartingale. This approach has been used in several previous works [13–15], and carrying it out in an optimized way leads to the choice of  $f$  given by (4.2), (4.3) and (4.4). Having fixed  $f$ , one can now see that contractivity on the hyperplane holds if  $c$  is sufficiently small depending on  $\eta$ ; cf. (4.8). Moreover, the Lyapunov condition implies contractivity at large distances if (4.11) holds. As a consequence, we obtain a global contraction result with a contraction rate  $c$  depending on the parameters in the definition of the metric; see Theorem 4.1. The final step of the proof of the main result then consists in choosing these parameters in order to maximize the resulting contraction rate approximately. This is carried out in Section 5. Roughly, the constants  $\alpha$  and  $\varepsilon$  in (2.9) and (2.10) are chosen sufficiently small such that the effects of the distortion of the metric do not destroy the contraction properties at small distances, but otherwise as large as possible. This is ensured by (5.3), (5.4) and (5.5). With this choice of constants, the main result then follows from Theorem 4.1.

**3. Coupling and evolution of coupling distance.** We fix a positive constant  $\xi$ . Eventually, we will consider the limit  $\xi \downarrow 0$ . In order to couple two so-

lutions of (1.1), we consider the following SDE on  $\mathbb{R}^{2d} \times \mathbb{R}^{2d}$ :

$$\begin{aligned}
 dX_t &= V_t dt, \\
 dV_t &= -\gamma V_t dt - u \nabla U(X_t) dt + \sqrt{2\gamma u} rc(Z_t, W_t) dB_t^{rc} \\
 &\quad + \sqrt{2\gamma u} sc(Z_t, W_t) dB_t^{sc}, \\
 (3.1) \quad dX'_t &= V'_t dt, \\
 dV'_t &= -\gamma V'_t dt - u \nabla U(X'_t) dt + \sqrt{2\gamma u} rc(Z_t, W_t)(I_d - 2e_t e_t^T) dB_t^{rc} \\
 &\quad + \sqrt{2\gamma u} sc(Z_t, W_t) dB_t^{sc}.
 \end{aligned}$$

Here,  $B^{rc}$  and  $B^{sc}$  are independent Brownian motions,

$$(3.2) \quad Z_t = X_t - X'_t, \quad W_t = V_t - V'_t, \quad Q_t = Z_t + \gamma^{-1} W_t,$$

$$(3.3) \quad e_t = Q_t/|Q_t| \quad \text{if } Q_t \neq 0 \quad \text{and} \quad e_t = 0 \quad \text{if } Q_t = 0.$$

Moreover,  $rc, sc : \mathbb{R}^{2d} \rightarrow [0, 1]$  are Lipschitz continuous functions such that  $rc^2 + sc^2 \equiv 1$ ,

$$(3.4) \quad rc(z, w) = 0 \quad \text{if } z + \gamma^{-1} w = 0 \quad \text{or} \quad \alpha|z| + |z + \gamma^{-1} w| \geq R_1 + \xi,$$

$$(3.5) \quad rc(z, w) = 1 \quad \text{if } |z + \gamma^{-1} w| \geq \xi \quad \text{and} \quad \alpha|z| + |z + \gamma^{-1} w| \leq R_1.$$

The values of the constants  $\alpha, R_1 \in (0, \infty)$  will be fixed later. Note that by (3.4),  $rc(Z_t, W_t)e_t e_t^T$  is a Lipschitz continuous function of  $(X_t, V_t, X'_t, V'_t)$ . Therefore, existence and uniqueness of the coupling process holds by Itô's theorem. Moreover, by Lévy's characterization, for any solution of (3.1),

$$\begin{aligned}
 B_t &= \int_0^t rc(Z_s, W_s) dB_s^{rc} + \int_0^t sc(Z_s, W_s) dB_s^{sc} \quad \text{and} \\
 B'_t &= \int_0^t rc(Z_s, W_s)(I_d - 2e_s e_s^T) dB_s^{rc} + \int_0^t sc(Z_s, W_s) dB_s^{sc}
 \end{aligned}$$

are again Brownian motions. Thus (3.1) defines indeed a coupling of two solutions of (1.1). For  $rc \equiv 1$  and  $sc \equiv 0$ , the coupling is a *reflection coupling*, whereas for  $rc \equiv 0$  and  $sc \equiv 1$  it is a *synchronous coupling*. By (3.4) and (3.5), we choose  $rc$  and  $sc$  such that the synchronous coupling is applied when  $r_t = \alpha|Z_t| + |Q_t|$  is large or  $Q_t$  is close to 0, and the reflection coupling is applied otherwise. Ideally, we would like to choose  $rc(Z_t, W_t) = 1_{\{Q_t \neq 0, r_t < R_1\}}$ . Since the indicator function is not continuous, we consider Lipschitz continuous approximations instead.

The processes  $(Z_t)$ ,  $(W_t)$  and  $(Q_t)$  satisfy the following equations:

$$\begin{aligned}
 (3.6) \quad dZ_t &= W_t dt = \gamma Q_t dt - \gamma Z_t dt, \\
 dW_t &= -\gamma W_t dt - u(\nabla U(X_t) - \nabla U(X'_t)) dt
 \end{aligned}$$



$$\begin{aligned}
 & + \sqrt{8\gamma u rc}(Z_t, W_t) e_t e_t^T dB_t^{rc}, \\
 (3.7) \quad dQ_t & = -u\gamma^{-1}(\nabla U(X_t) - \nabla U(X'_t)) dt \\
 & + \sqrt{8\gamma^{-1}u rc}(Z_t, W_t) e_t e_t^T dB_t^{rc}.
 \end{aligned}$$

Note that in particular,  $Z_t$  is contractive when  $Q_t = 0$ , and  $Q_t$  would be a local martingale if  $U$  would be constant. We set

$$(3.8) \quad r_t := r((X_t, V_t), (X'_t, V'_t)) = \alpha|Z_t| + |Q_t| \quad \text{and}$$

$$(3.9) \quad \rho_t := \rho((X_t, V_t), (X'_t, V'_t)) = f(r_t) \cdot G_t \quad \text{where}$$

$$(3.10) \quad G_t := 1 + \varepsilon\mathcal{V}(X_t, V_t) + \varepsilon\mathcal{V}(X'_t, V'_t).$$

LEMMA 3.1. *Let  $c, \varepsilon \in (0, \infty)$ , and suppose that  $f : [0, \infty) \rightarrow [0, \infty)$  is continuous, nondecreasing, concave and  $C^2$  except for finitely many points. Let*

$$(3.11) \quad \alpha = (1 + \eta)Lu\gamma^{-2}$$

for some constant  $\eta \in (0, \infty)$ . Then

$$(3.12) \quad e^{ct} \rho_t \leq \rho_0 + \gamma \int_0^t e^{cs} K_s ds + M_t \quad \text{for any } t \geq 0,$$

where  $(M_t)$  is a continuous local martingale, and

$$\begin{aligned}
 (3.13) \quad K_t & = 4u\gamma^{-2} rc(Z_t, W_t)^2 f''(r_t)G_t \\
 & + \left( \alpha|Q_t| - \frac{\eta}{1 + \eta} \alpha|Z_t| \right) f'_-(r_t)G_t \\
 & + 4\varepsilon \max(1, (2\alpha)^{-1}) rc(Z_t, W_t)^2 r_t f'_-(r_t) \\
 & + (2(d + A) - \lambda\mathcal{V}(X_t, V_t) - \lambda\mathcal{V}(X'_t, V'_t))\varepsilon f(r_t) \\
 & + \gamma^{-1} cf(r_t)G_t.
 \end{aligned}$$

PROOF. By (3.6), the paths of the process  $(Z_t)$  are almost surely continuously differentiable with derivative  $dZ/dt = -\gamma Z + \gamma Q$ . Therefore,  $t \mapsto |Z_t|$  is almost surely absolutely continuous with

$$\frac{d}{dt}|Z_t| = \frac{Z_t}{|Z_t|} \cdot (-\gamma Z_t + \gamma Q_t) \quad \text{for a.e. } t \text{ such that } Z_t \neq 0 \quad \text{and}$$

$$\frac{d}{dt}|Z_t| \leq \gamma|Q_t| \quad \text{for a.e. } t \text{ such that } Z_t = 0.$$

In particular, we obtain

$$(3.14) \quad \frac{d}{dt}|Z_t| \leq -\gamma|Z_t| + \gamma|Q_t| \quad \text{for a.e. } t \geq 0.$$

The process  $(Q_t)$  satisfies the SDE (3.7). Notice that by (3.4), the noise coefficient vanishes if  $Q_t = 0$ . Therefore, similar to Lemma 4 in [14], we can apply Itô's formula to conclude that almost surely,

$$(3.15) \quad |Q_t| = |Q_0| + A_t^Q + \tilde{M}_t^Q \quad \text{for all } t \geq 0,$$

where  $(A_t^Q)$  and  $(\tilde{M}_t^Q)$  are the absolutely continuous process and the martingale given by

$$A_t^Q = -u\gamma^{-1} \int_0^t e_s^T \cdot (\nabla U(X_s) - \nabla U(X'_s)) ds,$$

$$\tilde{M}_t^Q = \sqrt{8u\gamma^{-1}} \int_0^t rc(Z_s, W_s) e_s^T dB_s^{rc}.$$

Notice that there is no Itô correction, because  $\partial_{q/|q|}^2 |q| = 0$  for  $q \neq 0$  and the noise coefficient vanishes for  $Q_t = 0$ . By (3.14), (3.15) and the Lipschitz condition on  $\nabla U$ , we conclude that  $r_t = \alpha|Z_t| + |Q_t|$  has the semimartingale decomposition

$$r_t = |Q_0| + \alpha|Z_t| + A_t^Q + \tilde{M}_t^Q,$$

where  $t \mapsto \alpha|Z_t| + A_t^Q$  is almost surely absolutely continuous with derivative

$$(3.16) \quad \frac{d}{dt}(\alpha|Z_t| + A_t^Q) \leq ((Lu\gamma^{-2} - \alpha)|Z_t| + \alpha|Q_t|)\gamma \quad \text{for a.e. } t \geq 0.$$

Since by assumption,  $f$  is concave and piecewise  $C^2$ , we can now apply the Itô–Tanaka formula to  $f(r_t)$ . Let  $f'_-$  and  $f''$  denote the left-sided first derivative and the almost everywhere defined second derivative. Notice that the generalized second derivative of  $f$  is a signed measure  $\mu_f$  such that  $\mu_f(dr) \leq f''(r) dr$ . We obtain a semimartingale decomposition

$$(3.17) \quad e^{ct} f(r_t) = f(r_0) + \tilde{A}_t + \tilde{M}_t$$

with the martingale part

$$(3.18) \quad \tilde{M}_t = \sqrt{8u\gamma^{-1}} \int_0^t e^{cs} f'_-(r_s) rc(Z_s, W_s) e_s^T dB_s^{rc}$$

and a continuous finite-variation process  $(\tilde{A}_t)$  satisfying

$$(3.19) \quad d\tilde{A}_t = (cf(r_t) + ((Lu\gamma^{-2} - \alpha)|Z_t| + \alpha|Q_t|)\gamma f'_-(r_t))e^{ct} dt + (4u\gamma^{-1} rc(Z_t, W_t)^2 f''(r_t))e^{ct} dt.$$

Next, we consider the time evolution of the process  $G_t = 1 + \varepsilon\mathcal{V}(X_t, V_t) + \varepsilon\mathcal{V}(X'_t, V'_t)$ . An application of Itô's formula shows that by (3.1),

$$\begin{aligned}
 dG_t &= \varepsilon(\mathcal{L}\mathcal{V})(X_t, V_t) dt + \varepsilon(\mathcal{L}\mathcal{V})(X'_t, V'_t) dt \\
 &+ \varepsilon\sqrt{2u\gamma}(\nabla_v\mathcal{V}(X_t, V_t) - \nabla_v\mathcal{V}(X'_t, V'_t))^T rc(Z_t, W_t)e_t e_t^T dB_t^{rc} \\
 &+ \varepsilon\sqrt{2u\gamma}(\nabla_v\mathcal{V}(X_t, V_t) + \nabla_v\mathcal{V}(X'_t, V'_t))^T rc(Z_t, W_t)(I_d - e_t e_t^T) dB_t^{rc} \\
 &+ \varepsilon\sqrt{2u\gamma}(\nabla_v\mathcal{V}(X_t, V_t) + \nabla_v\mathcal{V}(X'_t, V'_t))^T sc(Z_t, W_t) dB_t^{sc}.
 \end{aligned}
 \tag{3.20}$$

Hence by (3.17), (3.20), the Itô product rule and (3.18), we obtain the semimartingale decomposition

$$e^{ct} \rho_t = e^{ct} f(r_t)G_t = \rho_0 + M_t + A_t,
 \tag{3.21}$$

where  $(M_t)$  is a continuous local martingale, and

$$\begin{aligned}
 dA_t &= G_t d\tilde{A}_t + \varepsilon e^{ct} f(r_t)((\mathcal{L}\mathcal{V})(X_t, V_t) + (\mathcal{L}\mathcal{V})(X'_t, V'_t)) dt \\
 &+ 4\varepsilon u e^{ct} f'_-(r_t) rc(Z_t, W_t)^2 (\nabla_v\mathcal{V}(X_t, V_t) - \nabla_v\mathcal{V}(X'_t, V'_t))^T e_t dt.
 \end{aligned}
 \tag{3.22}$$

Here, we have used that  $B^{rc}$  and  $B^{sc}$  are independent Brownian motions in  $\mathbb{R}^d$ . Now recall that by Lemma 2.2,

$$\mathcal{L}\mathcal{V} \leq (d + A - \lambda\mathcal{V})\gamma.
 \tag{3.23}$$

Furthermore, a simple computation shows that by (2.7),

$$\begin{aligned}
 &|\nabla_v\mathcal{V}(X_t, V_t) - \nabla_v\mathcal{V}(X'_t, V'_t)| \\
 &= \left| \frac{\gamma}{2u} \left( X_t + \frac{2}{\gamma} V_t - X'_t - \frac{2}{\gamma} V'_t \right) \right| \\
 &\leq u^{-1}\gamma(|Q_t| + |Z_t|/2) \\
 &\leq u^{-1}\gamma \max(1, (2\alpha)^{-1})r_t.
 \end{aligned}
 \tag{3.24}$$

By combining (3.22), (3.19), (3.23) and (3.24) we finally obtain  $dA_t \leq \gamma e^{ct} K_t dt$ , where

$$\begin{aligned}
 K_t &= \gamma^{-1} c f(r_t)G_t + ((Lu\gamma^{-2} - \alpha)|Z_t| + \alpha|Q_t|) f'_-(r_t)G_t \\
 &+ 4u\gamma^{-2} rc(Z_t, W_t)^2 f''(r_t)G_t \\
 &+ \varepsilon f(r_t)(2(d + A) - \lambda\mathcal{V}(X_t, V_t) - \lambda\mathcal{V}(X'_t, V'_t)) \\
 &+ 4\varepsilon \max(1, (2\alpha)^{-1})r_t f'_-(r_t) rc(Z_t, W_t)^2.
 \end{aligned}$$

The assertion now follows from (3.21) by setting  $\alpha = (1 + \eta)Lu\gamma^{-2}$ .  $\square$

**4. Contractivity in different regions.** As above, suppose that  $\alpha = (1 + \eta)Lu\gamma^{-2}$  for some  $\eta \in (0, \infty)$ . We will choose  $\varepsilon$  and  $c$  such that

$$(4.1) \quad (d + A)\varepsilon = 4c/\gamma.$$

This choice guarantees in particular that the terms  $2(d + A)\varepsilon$  and  $\gamma^{-1}c$  in (3.13) are of comparable order.

We consider a coupling as introduced in Section 3. In order to make  $\rho_t$  a contraction on average, we have to choose the parameters such that  $K_t \leq 0$ . This will be achieved up to an error term which vanishes as  $\xi \downarrow 0$ . In order to bound  $K_t$ , we distinguish between different regions of the state space:

(i)  $|Q_t| \geq \xi$  and  $r_t \leq R_1$ . Here by (3.5),  $rc(Z_t, W_t) = 1$ . Therefore, by (3.13) and (4.1), and since  $|Q_t| \leq r_t$  and  $G_t \geq 1$ ,

$$K_t \leq 4u\gamma^{-2}f''(r_t)G_t + ((1 + \eta)Lu\gamma^{-2} + 4\varepsilon \max(1, (2\alpha)^{-1}))r_t f'_-(r_t)G_t + 9\gamma^{-1}cf(r_t)G_t.$$

Similar to [14], we can now ensure that  $K_t \leq 0$  by choosing  $f$  appropriately. We set

$$(4.2) \quad f(r) = \int_0^{r \wedge R_1} \varphi(s)g(s) ds \quad \text{where}$$

$$(4.3) \quad \varphi(s) = \exp(-(1 + \eta)Ls^2/8 - \gamma^2u^{-1}\varepsilon \max(1, (2\alpha)^{-1})s^2/2) \quad \text{and}$$

$$(4.4) \quad g(r) = 1 - \frac{9}{4}c\gamma u^{-1} \int_0^r \Phi(s)\varphi(s)^{-1} ds \quad \text{with } \Phi(s) = \int_0^s \varphi(x) dx.$$

Then  $4u\gamma^{-2}\varphi'(r_t) + ((1 + \eta)Lu\gamma^{-2} + 4\varepsilon \max(1, (2\alpha)^{-1}))r_t\varphi(r_t) = 0$ , and hence

$$(4.5) \quad K_t \leq 4u\gamma^{-2}\varphi(r_t)g'(r_t)G_t + 9\gamma^{-1}c\Phi(r_t)G_t \leq 0.$$

In order to ensure  $g(r) \geq 1/2$  for  $r \leq R_1$ , we have to assume

$$(4.6) \quad c \leq \frac{2}{9}u\gamma^{-1} / \int_0^{R_1} \Phi(s)\varphi(s)^{-1} ds.$$

(ii)  $|Q_t| < \xi$  and  $r_t \leq R_1$ . In this region, there is a transition from reflection coupling to synchronous coupling, which is applied for  $Q_t = 0$ . Hence we cannot rely on the additional contraction properties gained by applying reflection coupling and choosing  $f$  sufficiently concave. Instead, however, we can use that

$$(4.7) \quad |Q_t| \leq \xi \quad \text{and} \quad \alpha|Z_t| = r_t - |Q_t| \geq r_t - \xi.$$

By the choice of  $f$  and since  $g \geq 1/2$ , we obtain, similar to Case (i),

$$\begin{aligned} K_t &\leq (4u\gamma^{-2}f''(r_t)G_t + 4\varepsilon \max(1, (2\alpha)^{-1})r_t f'_-(r_t))rc(Z_t, W_t)^2 \\ &\quad - \frac{\eta}{1 + \eta}r_t f'_-(r_t)G_t + (1 + \alpha)\xi f'_-(r_t)G_t + 9\gamma^{-1}cf(r_t)G_t \\ &\leq -\frac{1}{2} \frac{\eta}{1 + \eta}r_t\varphi(r_t)G_t + 9\gamma^{-1}c\Phi(r_t)G_t + (1 + \alpha)\xi G_t. \end{aligned}$$

In order to ensure that the upper bound converges to 0 as  $\xi \downarrow 0$ , we assume

$$(4.8) \quad c \leq \frac{1}{18} \gamma \frac{\eta}{1 + \eta} \inf_{r \in (0, R_1]} \frac{r\varphi(r)}{\Phi(r)}.$$

Under this condition, we obtain

$$(4.9) \quad K_t \leq (1 + \alpha)\xi G_t.$$

(iii)  $r_t > R_1$ . Here  $f'_-(r_t) = 0$ . Hence by (3.13), (3.10), (4.1) and (2.15),

$$(4.10) \quad \begin{aligned} K_t &= [2(d + A) + \gamma^{-1}c\varepsilon^{-1} \\ &\quad - (\lambda - \gamma^{-1}c)(\mathcal{V}(X_t, V_t) + \mathcal{V}(X'_t, V'_t))] \varepsilon f(r_t) \\ &\leq \left[ \frac{9}{4}(d + A) - \frac{15}{16}\lambda(\mathcal{V}(X_t, V_t) + \mathcal{V}(X'_t, V'_t)) \right] \varepsilon f(r_t) \leq 0, \end{aligned}$$

provided we assume

$$(4.11) \quad c \leq \gamma\lambda/16.$$

Notice that the functions  $\varphi$ ,  $g$  and  $f$  in (4.3), (4.4), (4.2) and the constants  $\alpha$ ,  $\varepsilon$  determining  $\rho$  do not depend on the value of  $\xi$ . Therefore, by combining Lemma 3.1, (4.5), (4.9) and (4.10), we obtain the following general result.

**THEOREM 4.1.** *Let  $\alpha = (1 + \eta)Lu\gamma^{-2}$  for some  $\eta \in (0, \infty)$ , and let  $\varepsilon = 4\gamma^{-1}c/(d + A)$  for a positive constant  $c$  such that*

$$(4.12) \quad c \leq \gamma \min\left(\frac{2}{9}u\gamma^{-2} / \int_0^{R_1} \frac{\Phi(s)}{\varphi(s)} ds, \frac{1}{18} \frac{\eta}{1 + \eta} \inf_{s \in (0, R_1]} \frac{s\varphi(s)}{\Phi(s)}, \frac{1}{16}\lambda\right).$$

Moreover, let  $f : [0, \infty) \rightarrow [0, \infty)$  be defined by (4.2), (4.3) and (4.4). Then for any  $t \geq 0$  and for any probability measures  $\mu, \nu$  on  $\mathbb{R}^{2d}$ ,

$$(4.13) \quad \mathcal{W}_\rho(\mu p_t, \nu p_t) \leq e^{-ct} \mathcal{W}_\rho(\mu, \nu).$$

**PROOF.** Let  $\Gamma$  be a coupling of two probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$  such that  $\mathcal{W}_\rho(\mu, \nu) < \infty$ . We consider the coupling process  $((X_t, V_t), (X'_t, V'_t))$  introduced in Section 3 with initial law  $((X_0, V_0), (X'_0, V'_0)) \sim \Gamma$ . By (4.12), the conditions (4.6), (4.8), (4.11) are satisfied. Therefore, in each of the cases (i), (ii) and (iii) considered above, we obtain  $K_t \leq (1 + \alpha)\xi G_t$ . Therefore, we apply Lemma 3.1. By taking expectations in (3.12), evaluated at localizing stopping times  $T_n \uparrow t$  and applying Fatou's lemma as  $n \rightarrow \infty$ , we obtain

$$(4.14) \quad \mathbb{E}[\rho_t] \leq e^{-ct} \mathbb{E}[\rho_0] + \gamma(1 + \alpha)\xi \int_0^t e^{c(s-t)} \mathbb{E}[G_s] ds$$

for any  $\xi > 0$  and  $t \geq 0$ . Note that the coupling process and the coupling distance  $\rho_t$  still depend on the value of  $\xi$ . On the other hand, the expectation of  $G_s$  does not depend on  $\xi$ . Indeed, by (3.10),

$$\begin{aligned} \mathbb{E}[G_s] &= \mathbb{E}[1 + \varepsilon \mathcal{V}(X_s, V_s) + \varepsilon \mathcal{V}(X'_s, V'_s)] \\ (4.15) \quad &= 1 + \varepsilon \int p_s \mathcal{V} d\mu + \varepsilon \int p_s \mathcal{V} d\nu. \end{aligned}$$

Since  $\mathcal{W}_\rho(\mu, \nu) < \infty$ , both  $\int \mathcal{V} d\mu$  and  $\int \mathcal{V} d\nu$  are finite. Therefore, by the Lyapunov condition in Lemma 2.2, the expectation in (4.15) is finite also.

Since  $((X_t, V_t), (X'_t, V'_t))$  is a coupling of  $\mu p_t$  and  $\nu p_t$ , we have  $\mathcal{W}_\rho(\mu p_t, \nu p_t) \leq \mathbb{E}[\rho_t]$  for any  $\xi > 0$ . Moreover,  $\mathbb{E}[\rho_0] = \int \rho d\Gamma$ . Hence by applying (4.14) and taking the limit  $\xi \downarrow 0$ , we obtain

$$(4.16) \quad \mathcal{W}_\rho(\mu p_t, \nu p_t) \leq e^{-ct} \int \rho d\Gamma \quad \text{for any } t \geq 0.$$

The assertion follows since (4.16) holds for an arbitrary coupling of  $\mu$  and  $\nu$ .  $\square$

**5. Choice of the constants.** The function

$$(5.1) \quad \varphi(s) = \exp(-(1 + \eta)Ls^2/8 - u^{-1}\gamma^2\varepsilon \max(1, (2\alpha)^{-1})s^2/2)$$

determining the metric in Theorem 4.1 still depends on the values of the constants  $\eta \in (0, \infty)$ ,  $\alpha = (1 + \eta)Lu\gamma^{-2}$  and  $\varepsilon \in (0, \infty)$ . In order to prove our main result, we now choose explicit values for  $\eta$  and  $\varepsilon$  (and hence for  $\alpha$  and  $c$ ). We first discuss how to choose these constants in order to optimize the resulting bound for the contraction rate in (4.12) approximately, and then we apply Theorem 4.1 with the chosen constants. A first condition we want to be satisfied is that  $\varphi(s)$  is bounded below by the minimal possible value  $\exp(-Ls^2/8)$  up to a multiplicative constants. Therefore, we choose  $\eta$  and  $\varepsilon$  such that

$$(5.2) \quad \varphi(s) \geq e^{-2} \exp(-Ls^2/8) \quad \text{for any } s \in [0, R_1].$$

By (5.1), this bound holds true if

$$(5.3) \quad \eta = 8/(LR_1^2) = \Lambda^{-1} \quad \text{or, equivalently,}$$

$$(5.4) \quad \alpha = (1 + \Lambda^{-1})Lu\gamma^{-2} = (L + 8R_1^{-2})u\gamma^{-2} \quad \text{and}$$

$$(5.5) \quad \varepsilon \leq 2 \min(1, 2\alpha)u\gamma^{-2}R_1^{-2}.$$

We recall from (2.13) that

$$(5.6) \quad R_1 = \sqrt{(1 + \alpha)^2 + \alpha^2}R_0 \quad \text{where } R_0 := u^{1/2}\gamma^{-1} \left( \frac{96(d + A)}{5\lambda(1 - 2\lambda)} \right)^{1/2}.$$

In particular,  $R_1^{-2}$  is a decreasing function of  $\alpha$ , and hence there are unique values  $\alpha, \eta \in (0, \infty)$  such that (5.4) and (5.3) are satisfied. We fix  $\eta$  and  $\alpha$  correspond-

ingly, and then we choose  $\varepsilon > 0$  such that (5.5) is satisfied. By (4.1) and (5.6), the condition (5.5) on  $\varepsilon$  is equivalent to the following additional constraint on the constant  $c$  in Theorem 4.1:

$$(5.7) \quad \begin{aligned} c &= \gamma\varepsilon(d + A)/4 \leq \min(1, 2\alpha)u\gamma^{-1}(d + A)R_1^{-2}/2 \\ &= \frac{5}{192} \frac{\min(1, 2\alpha)}{(1 + \alpha)^2 + \alpha^2} \gamma^\lambda(1 - 2\lambda). \end{aligned}$$

Since the last expression is smaller than  $\gamma\lambda/16$ , we see that Theorem 4.1 applies with parameters  $\eta$  and  $\varepsilon$  satisfying (5.3) and (5.5) whenever  $c$  is smaller than

$$(5.8) \quad \gamma \min\left(\frac{5}{192} \frac{\min(1, 2\alpha)\lambda(1 - 2\lambda)}{(1 + \alpha)^2 + \alpha^2}, \frac{2u}{9\gamma^2} / \int_0^{R_1} \frac{\Phi}{\varphi}, \frac{1}{18} \frac{\eta}{1 + \eta} \inf_{s \in (0, R_1]} \frac{s\varphi(s)}{\Phi(s)}\right).$$

We can now complete the proof of Theorem 2.3 by deriving an explicit lower bound for the right-hand side of (5.8).

**PROOF OF THEOREM 2.3.** We fix  $\eta$  and  $\alpha$  as in (5.3), (5.4) and we choose  $\varepsilon > 0$  such that (5.5) holds. For these parameters, we give explicit lower bounds for the three terms in the minimum in (5.8):

1. By (5.4), (5.6) and (2.6),

$$(5.9) \quad \begin{aligned} Lu\gamma^{-2} \leq \alpha &\leq (L + 8R_0^{-2})u\gamma^{-2} \leq Lu\gamma^{-2} + \frac{5}{12}\lambda/(d + A) \\ &\leq \left(1 + \frac{5}{6}(d + A)^{-1}\right)Lu\gamma^{-2} \leq \frac{11}{6}Lu\gamma^{-2}. \end{aligned}$$

Therefore, and since  $\lambda \leq 1/4$ , we obtain

$$(5.10) \quad \begin{aligned} \frac{\min(1, 2\alpha)}{(1 + \alpha)^2 + \alpha^2} &\geq \min\left(\frac{4}{5}\alpha, \frac{1}{10}\alpha^{-2}\right), \\ \frac{5}{192} \frac{\min(1, 2\alpha)\lambda(1 - 2\lambda)}{(1 + \alpha)^2 + \alpha^2} &\geq \frac{1}{96} \min\left(Lu\gamma^{-2}, \frac{1}{32}(Lu\gamma^{-2})^{-2}\right)\lambda. \end{aligned}$$

2. Recall that  $\Phi(r) = \int_0^r \varphi(s) ds$ . Since  $\varphi(s) \leq \exp(-Ls^2/8)$ , we have

$$(5.11) \quad \Phi(r) \leq \int_0^\infty \exp(-Ls^2/8) ds = \sqrt{2\pi/L} \quad \text{for any } r \geq 0.$$

Furthermore, by (5.2),

$$e^{-2} \int_0^{R_1} 1/\varphi \leq \int_0^{R_1} \exp(Ls^2/8) ds \leq \frac{8}{LR_1} \exp(LR_1^2/8).$$

Here, we have used  $\int_0^x \exp(u^2/2)du \leq 2x^{-1} \exp(x^2/2)$  for  $x > 0$ . We obtain

$$\int_0^{R_1} \frac{\Phi}{\varphi} \leq \int_0^{R_1} \frac{\Phi(R_1)}{\varphi} \leq 8\sqrt{2\pi}e^2L^{-3/2}R_1^{-1}e^{LR_1^2/8} = 4\sqrt{\pi}e^2L^{-1}\Lambda^{-1/2}e^\Lambda,$$

and thus

$$(5.12) \quad \frac{9}{2}u\gamma^{-2} / \int_0^{R_1} (\Phi/\varphi) \geq \frac{1}{18}\pi^{-1/2}e^{-2}Lu\gamma^{-2}\Lambda^{1/2}e^{-\Lambda}.$$

3. By (5.3) and (2.22),  $\eta \leq 5/6$ , and hence

$$(5.13) \quad \frac{\eta}{1+\eta} \geq \frac{6}{11}\eta = \frac{6}{11}\Lambda^{-1}.$$

Moreover, for  $r \leq \min(2/\sqrt{L}, R_1)$ ,

$$r\varphi(r)/\Phi(r) \geq \varphi(r) \geq e^{-1}e^{-(1+\eta)Lr^2/8} \geq e^{-1}e^{-11/12} \geq e^{-2}$$

by (5.1) and the choice of  $\varepsilon$ , and for  $2/\sqrt{L} \leq r \leq R_1$ ,

$$r\varphi(r)/\Phi(r) \geq \sqrt{L/(2\pi)}R_1\varphi(R_1) \geq 2\pi^{-1/2}\sqrt{LR_1^2/8}e^{-LR_1^2/8},$$

where we have used that  $s\varphi(s) = se^{-\beta s^2}$ ,  $\beta \geq L/8$ , is a decreasing function for  $s \geq 2/\sqrt{L}$ . Since  $\Lambda = LR_1^2/8 \geq 1$  by (2.22), we obtain

$$r\varphi(r)/\Phi(r) \geq 2\pi^{-1/2}e^{-2}\Lambda^{1/2}e^{-\Lambda} \quad \text{for all } r \in (0, R_1],$$

and hence by (5.13),

$$(5.14) \quad \frac{1}{18} \frac{\eta}{1+\eta} \inf_{s \in (0, R_1]} \frac{s\varphi(s)}{\Phi(s)} \geq \frac{1}{18}\pi^{-1/2}e^{-2}\Lambda^{-1/2}e^{-\Lambda}.$$

By combining (5.10), (5.12) and (5.14), we see that the right-hand side of (5.8) is lower bounded by the minimum of the expression on the right-hand sides of (5.10), (5.12) and (5.14). As a consequence, we see that Theorem 4.1 applies with constants  $\eta$  given by (5.3) and  $\varepsilon = 4\gamma^{-1}c/(d + A)$  satisfying (5.5), provided  $c \leq c_\star$ , where

$$c_\star = \frac{1}{384}\gamma \min(\lambda Lu\gamma^{-2}, \lambda(Lu\gamma^{-2})^{-2}/8, \Lambda^{1/2}e^{-\Lambda}Lu\gamma^{-2}, \Lambda^{-1/2}e^{-\Lambda}).$$

Here, we have used that  $18\pi^{1/2}e^2 \leq 384$ . By (2.19), and since  $\Lambda \geq 1$  and  $x^{3/2}e^{-x} < 1$  for  $x \geq 1$ , we also have

$$\frac{\lambda}{8(Lu\gamma^{-2})^2} \geq \left(\frac{\lambda}{2Lu\gamma^{-2}}\right)^2 \frac{1}{2\lambda} \geq \Lambda^{-2} \geq \Lambda^{-1/2}e^{-\Lambda}.$$



Therefore, the second term in the minimum defining  $c_*$  can be dropped and the value  $c_*$  is equal to the contraction rate (2.18) in the assertion of Theorem 2.3. Thus we have shown that Theorem 2.3 is indeed a special case of Theorem 4.1.  $\square$

PROOF OF COROLLARY 2.6. By (2.9), for  $(x, v), (x', v') \in \mathbb{R}^{2d}$ ,

$$(5.15) \quad |(x, v) - (x', v')|^2 \leq (1 + \gamma)^2 \max(1, \alpha^{-2}) r((x, v), (x, v'))^2.$$

Moreover, if  $r := r((x, v), (x, v')) \leq \min(1, R_1)$ , then by (2.11) and (2.21),

$$(5.16) \quad r^2 \leq r \leq f(r)/f'_-(R_1) \leq 2e^{2+\Lambda} f(r) \leq 2e^{2+\Lambda} \rho((x, v), (x, v')),$$

and if  $r \geq \min(1, R_1)$ , then by (2.14) and since  $\lambda \leq 1/4$  and  $\varepsilon = 4\gamma^{-1}c/(d + A)$ ,

$$\begin{aligned} r^2 &\leq 16((1 + \alpha)^2 + \alpha^2)u\gamma^{-2}\varepsilon^{-1}(1 + \varepsilon\mathcal{V}(x, v) + \varepsilon\mathcal{V}(x', v')) \\ &\leq 4((1 + \alpha)^2 + \alpha^2)(d + A)u\gamma^{-1}c^{-1}\rho((x, v), (x, v'))/f(\min(1, R_1)) \\ &\leq 8e^{2+\Lambda}(1 + 2\alpha + 2\alpha^2)(d + A)u\gamma^{-1}c^{-1}\rho((x, v), (x, v'))/\min(1, R_1). \end{aligned}$$

Combining the above bounds with Theorem 2.3 implies that

$$\mathcal{W}^2(\mu_{p_t}, \nu_{p_t})^2 \leq C\mathcal{W}_\rho(\mu_{p_t}, \nu_{p_t}) \leq Ce^{-ct}\mathcal{W}_\rho(\mu, \nu) \quad \text{for any } t \geq 0.$$

Uniqueness of the invariant probability measure now follows by standard arguments.  $\square$

PROOF OF COROLLARY 2.9. By (2.28) and since  $Lu\gamma^{-2} \leq 1/30$ ,

$$(5.17) \quad \lambda Lu\gamma^{-2} \geq \frac{15}{8\ell}(Lu\gamma^{-2})^2.$$

Furthermore, by (2.29), (2.27) and (2.28),

$$(5.18) \quad \begin{aligned} d \leq \Lambda \leq 2(d + (LR^2 - \beta)/8)LR^2/\beta \leq \Lambda_1 \quad \text{whence} \\ \Lambda^{1/2}e^{-\Lambda}Lu\gamma^{-2} \geq d^{1/2}e^{-\Lambda_1}Lu\gamma^{-2} \quad \text{and} \quad \Lambda^{-1/2}e^{-\Lambda} \geq \Lambda_1^{-1/2}e^{-\Lambda_1}. \end{aligned}$$

The first inequality in (2.31) now follows from Theorem 2.3, (5.17) and (5.18). Moreover, the second inequality holds since by (2.27),  $\gamma \geq \sqrt{30Lu} \geq \sqrt{30\beta u/R}$ .  $\square$

### APPENDIX: DRIFT CONDITIONS AND LYAPUNOV FUNCTIONS

PROOF OF LEMMA 2.2. By (1.2),  $\mathcal{L}U(x) = v \cdot \nabla U(x)$ ,  $\mathcal{L}|x|^2 = 2x \cdot v$ ,

$$\mathcal{L}\frac{1}{2}|\gamma^{-1}v|^2 = u\gamma^{-1}d - \gamma^{-1}|v|^2 - u\gamma^{-2}v \cdot \nabla U(x),$$

$$\mathcal{L}\frac{1}{2}|x + \gamma^{-1}v|^2 = u\gamma^{-1}d - u\gamma^{-1}(x + \gamma^{-1}v) \cdot \nabla U(x),$$

and hence

$$\begin{aligned} \mathcal{L}\mathcal{V}(x, v) &= \frac{1}{2}\gamma(2d - x \cdot \nabla U(x) - u^{-1}|v|^2 - \lambda u^{-1}\gamma x \cdot v) \\ &\leq \gamma\left(d + A - \lambda U(x) - \frac{1}{4}\lambda u^{-1}\gamma^2(|x|^2 + 2x \cdot \gamma^{-1}v + 2\lambda^{-1}|\gamma^{-1}v|^2)\right) \\ &= \gamma(d + A - \lambda\mathcal{V}(x, v)). \quad \square \end{aligned}$$

**PROOF OF LEMMA 2.8.** Suppose that Assumption 2.7 is satisfied, and let  $A = (L\mathcal{R}^2 - \beta)/8$ . Then by (2.26), (2.25) and (2.27),

$$\begin{aligned} x \cdot \nabla U(x) &= \frac{|x|}{\mathcal{R}} \frac{\mathcal{R}x}{|x|} \nabla U\left(\frac{\mathcal{R}x}{|x|}\right) + x \cdot \left(\nabla U(x) - \nabla U\left(\frac{\mathcal{R}x}{|x|}\right)\right) \\ &\geq \beta\mathcal{R}^{-1}|x| - L(\mathcal{R} - |x|)|x| \geq \beta\mathcal{R}^{-2}|x|^2 - (L - \beta\mathcal{R}^{-2})(\mathcal{R} - |x|)|x| \\ &\geq \beta\mathcal{R}^{-2}|x|^2 - 2A \end{aligned}$$

holds for  $x \in \mathbb{R}^d$  s.t.  $0 < |x| \leq \mathcal{R}$ . Noting that by (2.24) and (2.25),

$$U(x) \leq U(0) + L|x|^2/2 = L|x|^2/2,$$

we obtain

$$\begin{aligned} U(x) + u^{-1}\gamma^2|x|^2/4 &\leq (2L + u^{-1}\gamma^2)|x|^2/4 \\ &\leq \left(\frac{1}{2}x \cdot \nabla U(x) + A\right)(1 + L^{-1}u^{-1}\gamma^2/2)\beta^{-1}L\mathcal{R}^2 \end{aligned}$$

for any  $x \in \mathbb{R}^d$ . Hence (2.5) holds with  $\lambda$  given by (2.28). In particular,  $Lu\gamma^{-2}\lambda^{-1} \geq L\mathcal{R}^2/(2\beta)$ , and hence by (2.19),

$$(A.1) \quad \Lambda \geq \frac{6}{5}(d + A)L\mathcal{R}^2/\beta.$$

Now suppose first that  $Lu\gamma^{-2} \leq 1/8$ . Then since  $\beta \leq L\mathcal{R}^2$ , we have

$$\begin{aligned} \lambda &= \frac{2\beta}{L\mathcal{R}^2} \frac{Lu\gamma^{-2}}{1 + 2Lu\gamma^{-2}}, \\ (1 - 2\lambda)^{-1} &= \frac{1 + 2Lu\gamma^{-2}}{1 + (2 - 4\beta L^{-1}\mathcal{R}^{-2})Lu\gamma^{-2}} \leq 1 + \frac{4Lu\gamma^{-2}}{1 - 2Lu\gamma^{-2}} \frac{\beta}{L\mathcal{R}^2} \\ &\leq 1 + \frac{16}{3} \frac{\beta}{L\mathcal{R}^2} Lu\gamma^{-2}, \end{aligned}$$

and hence by (2.19),

$$(A.2) \quad \Lambda \leq \frac{6}{5}(d + A)\beta^{-1}L\mathcal{R}^2(1 + 2\alpha + 2\alpha^2)(1 + 2Lu\gamma^{-2})\left(1 + \frac{16}{3}Lu\gamma^{-2}\right).$$

By (2.20),  $\alpha \leq 11Lu\gamma^{-2}/6 \leq 11/48$ . Noting that  $Lu\gamma^{-2} \leq 1/8$ , (2.29) follows from (A.1) and (A.2).  $\square$

## REFERENCES

- [1] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. [MR2381160](#)
- [2] BANERJEE, S. and KENDALL, W. S. (2016). Coupling the Kolmogorov diffusion: Maximality and efficiency considerations. *Adv. in Appl. Probab.* **48** 15–35. [MR3539295](#)
- [3] BAUDOIN, F. (2016). Wasserstein contraction properties for hypoelliptic diffusions. Preprint. Available at [arXiv:1602.04177](#).
- [4] BEN AROUS, G., CRANSTON, M. and KENDALL, W. S. (1995). Coupling constructions for hypoelliptic diffusions: Two examples. In *Stochastic Analysis (Ithaca, NY, 1993)*. *Proc. Sympos. Pure Math.* **57** 193–212. Amer. Math. Soc., Providence, RI. [MR1335472](#)
- [5] BOLLEY, F., GUILLIN, A. and MALRIEU, F. (2010). Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov–Fokker–Planck equation. *M2AN Math. Model. Numer. Anal.* **44** 867–884. [MR2731396](#)
- [6] BOU-RABEE, N. and SANZ-SERNA, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. [MR3693523](#)
- [7] CALOGERO, S. (2012). Exponential convergence to equilibrium for kinetic Fokker–Planck equations. *Comm. Partial Differential Equations* **37** 1357–1390. [MR2957543](#)
- [8] CHEN, M.-F. and WANG, F.-Y. (1997). Estimation of spectral gap for elliptic operators. *Trans. Amer. Math. Soc.* **349** 1239–1267. [MR1401516](#)
- [9] CHEN, M. F. and LI, S. F. (1989). Coupling methods for multidimensional diffusion processes. *Ann. Probab.* **17** 151–177. [MR0972776](#)
- [10] DESVILLETES, L. and VILLANI, C. (2001). On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: The linear Fokker–Planck equation. *Comm. Pure Appl. Math.* **54** 1–42. [MR1787105](#)
- [11] DOLBEAULT, J., MOUHOT, C. and SCHMEISER, C. (2015). Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.* **367** 3807–3828. [MR3324910](#)
- [12] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222.
- [13] EBERLE, A. (2011). Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris* **349** 1101–1104. [MR2843007](#)
- [14] EBERLE, A. (2016). Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* **166** 851–886. [MR3568041](#)
- [15] EBERLE, A., GUILLIN, A. and ZIMMER, R. (2019). Quantitative Harris type theorems for diffusions and McKean–Vlasov processes. *Trans. Amer. Math. Soc.* **371** 7135–7173.
- [16] EBERLE, A. and ZIMMER, R. (2016). Sticky couplings of multidimensional diffusions with different drifts. Preprint. Available at [arXiv:1612.06125](#).
- [17] ECKMANN, J.-P. and HAIRER, M. (2003). Spectral properties of hypoelliptic operators. *Comm. Math. Phys.* **235** 233–253. [MR1969727](#)
- [18] EINSTEIN, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **322** 549–560.
- [19] GROTHAUS, M. and STILGENBAUER, P. (2014). Hypocoercivity for Kolmogorov backward evolution equations and applications. *J. Funct. Anal.* **267** 3515–3556. [MR3266239](#)
- [20] GROTHAUS, M. and STILGENBAUER, P. (2016). Hilbert space hypocoercivity for the Langevin dynamics revisited. *Methods Funct. Anal. Topology* **22** 152–168. [MR3522857](#)
- [21] GUILLIN, A. and MONMARCHÉ, P. (2016). Optimal linear drift for the speed of convergence of an hypoelliptic diffusion. *Electron. Commun. Probab.* **21** Paper No. 74, 14. [MR3568348](#)
- [22] HAIRER, M. (2002). Exponential mixing properties of stochastic PDEs through asymptotic coupling. *Probab. Theory Related Fields* **124** 345–380. [MR1939651](#)

- [23] HAIRER, M. and MATTINGLY, J. C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability* **63** 109–117. Birkhäuser/Springer Basel AG, Basel. [MR2857021](#)
- [24] HAIRER, M., MATTINGLY, J. C. and SCHEUTZOW, M. (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149** 223–259. [MR2773030](#)
- [25] HELFFER, B. and NIER, F. (2005). *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians. Lecture Notes in Math.* **1862**. Springer, Berlin. [MR2130405](#)
- [26] HÉRAU, F. (2006). Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation. *Asymptot. Anal.* **46** 349–359. [MR2215889](#)
- [27] HÉRAU, F. and NIER, F. (2004). Isotropic hypoellipticity and trend to equilibrium for the Fokker–Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.* **171** 151–218. [MR2034753](#)
- [28] LANGEVIN, P. (1908). Sur la théorie du mouvement brownien. *C. R. Math. Acad. Sci. Paris* **146** 530–533.
- [29] LELIÈVRE, T., ROUSSET, M. and STOLTZ, G. (2010). *Free Energy Computations: A Mathematical Perspective*. Imperial College Press, London. [MR2681239](#)
- [30] LINDVALL, T. and ROGERS, L. C. G. (1986). Coupling of multidimensional diffusions by reflection. *Ann. Probab.* **14** 860–872. [MR0841588](#)
- [31] MATTINGLY, J. C. (2002). Exponential convergence for the stochastically forced Navier–Stokes equations and other partially dissipative dynamics. *Comm. Math. Phys.* **230** 421–462. [MR1937652](#)
- [32] MATTINGLY, J. C., STUART, A. M. and HIGHAM, D. J. (2002). Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** 185–232. [MR1931266](#)
- [33] MISCHLER, S. and MOUHOT, C. (2016). Exponential stability of slowly decaying solutions to the kinetic-Fokker–Planck equation. *Arch. Ration. Mech. Anal.* **221** 677–723. [MR3488535](#)
- [34] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. 113–162. CRC Press, Boca Raton, FL. [MR2858447](#)
- [35] NELSON, E. (1967). *Dynamical Theories of Brownian Motion, Vol. 2*. Princeton Univ. Press, Princeton, NJ. [MR0214150](#)
- [36] PAVLIOTIS, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations. Texts in Applied Mathematics* **60**. Springer, New York. [MR3288096](#)
- [37] REY-BELLET, L. and THOMAS, L. E. (2002). Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.* **225** 305–329. [MR1889227](#)
- [38] SCHUSS, Z. (2010). *Theory and Applications of Stochastic Processes: An Analytical Approach. Applied Mathematical Sciences* **170**. Springer, New York. [MR2583642](#)
- [39] TALAY, D. (2002). Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields* **8** 163–198. [MR1924934](#)
- [40] VILLANI, C. (2007). Hypocoercive diffusion operators. *Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat.* (8) **10** 257–275. [MR2339441](#)
- [41] VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Springer, Berlin. [MR2459454](#)
- [42] VILLANI, C. (2009). Hypocoercivity. *Mem. Amer. Math. Soc.* **202** iv+141. [MR2562709](#)

- [43] VON SMOLUCHOWSKI, M. (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. Phys.* **326** 756–780.
- [44] WATANABE, S. (1971). On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. *J. Math. Kyoto Univ.* **11** 169–180. [MR0275537](#)
- [45] WATANABE, S. (1971). On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. II. *J. Math. Kyoto Univ.* **11** 545–551. [MR0287612](#)
- [46] WU, L. (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.* **91** 205–238. [MR1807683](#)
- [47] ZIMMER, R. (2017). Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions. *Stoch. Partial Differ. Equ. Anal. Comput.* **5** 368–399. [MR3689216](#)
- [48] ZIMMER, R. (2017). Couplings and Kantorovich contractions with explicit rates for diffusions. Ph.D. thesis, Univ. Bonn.

A. EBERLE  
R. ZIMMER  
INSTITUT FÜR ANGEWANDTE MATHEMATIK  
UNIVERSITÄT BONN  
ENDENICHER ALLEE 60  
53115 BONN  
GERMANY  
E-MAIL: [eberle@uni-bonn.de](mailto:eberle@uni-bonn.de)  
[raphael.zimmer@uni-bonn.de](mailto:raphael.zimmer@uni-bonn.de)  
URL: <http://wt.iam.uni-bonn.de>

A. GUILLIN  
LABORATOIRE DE MATHÉMATIQUES BLAISE PASCAL  
CNRS - UMR 6620  
UNIVERSITÉ CLERMONT-AUVERGNE  
AVENUE DES LANDAIS  
63177 AUBIERE CEDEX  
FRANCE  
E-MAIL: [guillin@math.univ-bpclermont.fr](mailto:guillin@math.univ-bpclermont.fr)  
URL: <http://math.univ-bpclermont.fr/~guillin/>