Invariant distribution of duplicated diffusions and application to Richardson-Romberg extrapolation

Vincent Lemaire*, Gilles Pagès†, Fabien Panloup‡

October 15, 2013

Abstract

With a view to numerical applications we address the following question: given an ergodic Brownian diffusion with a unique invariant distribution, what are the invariant distributions of the duplicated system consisting of two trajectories? We mainly focus on the interesting case where the two trajectories are driven by the same Brownian path. Under this assumption, we first show that uniqueness is essentially always true in the one-dimensional case. In the multidimensional case, we begin by building some explicit counter-examples where the uniqueness property is not fulfilled. Then, we provide a series of criterions depending on the drift and diffusion coefficients (which can be viewed as the negativity of a *non-infinitesimal Lyapunov exponent* associated with the dynamical system) under which uniqueness of the invariant distribution of the duplicated system (weak confluence) or pathwise confluence hold. As examples, we apply our criterions to some non-trivially confluent settings such as classes of gradient systems with non-convex potentials or diffusions where the confluence is generated by the diffusive component.

As a main application, we investigate the Richardson-Romberg extrapolation for the numerical approximation of the invariant measure of the initial ergodic Brownian diffusion.

Keywords: Invariant measure ; ergodic diffusion ; Lyapunov exponent ; Central Limit Theorem ; Euler scheme ; Gradient System ; Richardson-Romberg extrapolation.

AMS classification (2000): 60G10, 60J60, 65C05, 60F05.

1 Introduction and motivations

When one discretizes a stochastic (or not) differential equation (SDE) by an Euler scheme with step h, a classical method to reduce the discretization error is the so-called Richardson-Romberg (RR) extrapolation introduced in [TT90] for diffusion processes. Roughly speaking, the idea of this method is to introduce a second Euler scheme with step $h/2$ and to choose an appropriate linear combination of the two schemes to cancel the first-order

*Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC, Case 188, 4 pl. Jussieu, F-75252 Paris Cedex 5, France, E-mail: `vincent.lemaire@upmc.fr`

†Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, UPMC, Case 188, 4 pl. Jussieu, F-75252 Paris Cedex 5, France, E-mail: `gilles.pages@upmc.fr`

‡Institut de Mathématiques de Toulouse, Université Paul Sabatier & INSA Toulouse, 135, av. de Rangueil, F-31077 Toulouse Cedex 4, France, E-mail: `fabien.panloup@math.univ-toulouse.fr`

discretization error. Such an idea can be adapted to the long-time setting. More precisely, when one tries to approximate the invariant distribution of a diffusion by empirical measures based on an Euler scheme (with decreasing step) of the diffusion, it is also possible to implement the same strategy by introducing a second Euler scheme with half-step (see [Lem05]). In fact, tackling the rate of convergence of such a procedure involving a couple of Euler schemes of the same SDE leads to studying the long run behaviour of the underlying couple of continuous processes that we will call *duplicated diffusion*. When the two solutions only differ by the starting value and are driven by the same Brownian motion, the resulting coupled process is also known as 2-*point motion* (terminology coming from the more general theory of stochastic flows, see [BS88, Car85, Kun90]). Before being more specific as concerns this motivation, let us now define precisely what we call a duplicated diffusion.
Consider the following Brownian diffusion solution to the stochastic differential equation

$$(SDE) \quad \equiv \quad dX_t = b(X_t)dt + \sigma(X_t)dW_t,\ X_0 = x \in \mathbb{R}^d, \tag{1.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathcal{M}(d, q, \mathbb{R})$ ($d \times q$ matrices with real valued entries) are locally Lipschitz continuous with linear growth and W is a standard q-dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ (satisfying the usual conditions). This stochastic differential equation (SDE) has a unique strong solution denoted $X^x = (X_t^x)_{t\geq 0}$. Let $\rho \in \mathcal{M}(q, q, \mathbb{R})$ be a square matrix with transpose ρ^* such that $I_q - \rho\rho^*$ is non-negative as a symmetric matrix. We consider a filtered probability space, still denoted $(\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ on which is defined a $2q$-dimensional standard $(\mathcal{F}_t)$-Brownian motion denoted $(W, \widetilde{W})$ so that W and $\widetilde{W}$ are two *independent* q-dimensional standard $(\mathcal{F}_t)$-Brownian motions. Then we define $W^{(\rho)}$ a third standard q-dimensional $(\mathcal{F}_t)$-Brownian motions by

$$W^{(\rho)} = \rho^* W + \sqrt{I_q - \rho^*\rho}\,\widetilde{W}, \tag{1.2}$$

which clearly satisfies

$$\langle W^i, W^{(\rho),j}\rangle_t = \rho_{ij}\, t,\ t \geq 0$$

(the square root should be understood in the set of symmetric non-negative matrices). The duplicated diffusion or "duplicated stochastic differential system" $(DSDS)$ is then defined by

$$(DSDS) \quad \equiv \quad \begin{cases} dX_t = b(X_t)dt + \sigma(X_t)\, dW_t, & X_0 = x_1 \in \mathbb{R}^d, \\ dX_t^{(\rho)} = b(X_t^{(\rho)})dt + \sigma(X_t^{(\rho)})\, dW_t^{(\rho)}, & X_0^{(\rho)} = x_2 \in \mathbb{R}^d. \end{cases} \tag{1.3}$$

Under the previous assumptions on b and σ, (1.3) has a unique (strong) solution. Then both $(X_t^x)_{t\geq 0}$ and $(X_t^{x_1}, X_t^{(\rho),x_2})_{t\geq 0}$ are homogeneous Markov processes with transition (Feller) semi-groups, denoted $(P_t(x, dy))_{t\geq 0}$ and $\big(Q_t^{(\rho)}((x_1, x_2), (dy_1, dy_2))\big)_{t\geq 0}$ respectively, and defined on test Borel functions $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, by

$$P_t(f)(x) = \mathbb{E} f(X_t^x) \quad \text{and} \quad Q_t^{(\rho)}(g)(x_1, x_2' = \mathbb{E}\, g(X_t^{x_1}, X_t^{(\rho),x_2}).$$

We will assume throughout the paper that the original diffusion X^x has an unique invariant distribution denoted ν *i.e.* satisfying $\nu P_t = \nu$ for every $t \in \mathbb{R}_+$. The first part of the paper is devoted to determining what are the invariant measures of $(Q_t^{(\rho)})_{t\geq 0}$ (if any) depending on the correlation matrix ρ. Thus, if $\rho = 0$, it is clear that $\nu \otimes \nu$ is invariant for

$Q^{(0)}$ and if $\rho = I_q$ so is $\nu_\Delta = \nu \circ (x \mapsto (x,x))^{-1}$, but are they the only ones? To be more precise, we want to establish easily verifiable criterions on b and σ which ensure that ν_Δ is the *unique* invariant distribution of $(DSDS)$. In the sequel, we will denote by μ a generic invariant measure of $Q^{(\rho)}$. We present the problem in more details (including references to the literature).

$\triangleright$ *Existence of an invariant distribution for* $(Q_t^{(\rho)})_{t\geq 0}$. First, the family of probability measures $(\mu_t^{(\rho)})_{t>0}$ defined on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d)^{\otimes 2})$ by

$$\mu_t^{(\rho)} = \frac{1}{t}\int_0^t \nu^{\otimes 2}(dx_1, dx_2) Q_s^{(\rho)}((x_1,x_2),(dy_1,dy_2))ds$$

is tight since both its marginals on $\mathbb{R}^d$ are equal to ν. Furthermore, the semi-group $(Q_t^{(\rho)})_{t\geq 0}$ being Feller, one easily shows that any of its limiting distributions $\mu^{(\rho)}$ as $t \to \infty$ is an invariant distribution for $(Q_t^{(\rho)})_{t\geq 0}$ such that $\mu^{(\rho)}(dx \times \mathbb{R}^d) = \mu^{(\rho)}(\mathbb{R}^d \times dx) = \nu(dx)$. Also note that, if uniqueness fails and $(P_t)_{t\geq 0}$ has two distinct invariant distributions ν and ν', a straightforward adaptation of the above (sketch of) proof shows that $(Q_t^{(\rho)})_{t\geq 0}$ has (at least) an invariant distribution with marginals (ν, ν') and another with (ν', ν) as marginals.

$\triangleright$ *Uniqueness of the invariant distribution of* $(Q_t^{(\rho)})_{t\geq 0}$. It is clear that in full generality the couple $(X, X^{(\rho)})$ may admit several invariant distributions even if X has only one such distribution. So is the case when $\sigma \equiv 0$ if the flow $\Phi(x,t)$ of the $ODE \equiv \dot{x} = b(x)$ has 0 as a unique repulsive equilibrium and a unique invariant distribution ν on $\mathbb{R}^d \setminus \{0\}$. Then both distributions $\nu^{\otimes 2}$ and ν_Δ (defined as above) on $(\mathbb{R}^d \setminus \{0\})^2$ are invariant and if ν is not reduced to a Dirac mass (think *e.g.* to a 2-dimensional ODE with a limit cycle around 0) $(DSDS)$ has at least two invariant distribution.

In the case ($\sigma \not\equiv 0$) the situation is more involved and depends on the correlation structure ρ between the two Brownian motions W and $W^{(\rho)}$. The diffusion matrix $\Sigma(X_t^{x_1}, X_t^{(\rho),x_2})$ of the couple $(X^{x_1}, X^{(\rho),x_2})$ at time $t > 0$ is given by any continuous solution to the equation

$$\Sigma(\xi_1,\xi_2)\Sigma(\xi_1,\xi_2)^* = \begin{bmatrix} \sigma\sigma^*(\xi_1) & \sigma(\xi_1)\rho\sigma^*(\xi_2) \\ \sigma(\xi_2)\rho^*\sigma^*(\xi_1) & \sigma\sigma^*(\xi_2) \end{bmatrix}$$

(*e.g.* the square root in the symmetric non-negative matrices or the Choleski transform...).

First, note that if $I_q - \rho^*\rho$ is *positive definite* as a symmetric matrix, it is straightforward that ellipticity or uniform ellipticity of $\sigma\sigma^*$(when $q \geq d$) for X^x is transferred to $\Sigma(X_t^{x_1}, X_t^{(\rho),x_2})\Sigma(X_t^{x_1}, X_t^{(\rho),x_2})^*$ for the couple $(X^{x_1}, X^{(\rho),x_2})$. Now, uniform ellipticity, combined with standard regularity and growth/boundedness assumption on the coefficients b, σ and their partial derivatives, classically implies the existence for every $t > 0$ of a (strictly) positive probability density $p_t(x,y)$ for X_t^x. These additional conditions are automatically satisfied by the "duplicated coefficients" of $(DSDS)$. At this stage, it is classical background that any homogeneous Markov process whose transition has a (strictly) positive density for every $t > 0$ has at most one invariant distribution (if any). Consequently, under these standard assumptions on b and σ which ensure uniqueness of the invariant distribution ν for X, we get uniqueness for the "duplicated" diffusion process $(X, X^{(\rho)})$ as well.

The *hypo-elliptic* case also implies the existence of a density for X_t^x and the uniqueness of the invariant distribution under *controllability assumptions* on a companion differential system of the SDE. This property can also be transferred to $(DSDS)$, although the

proof becomes significantly less straightforward than above (see Appendix B for a precise statement and a detailed proof).

We now consider one of the main problems of this paper: the degenerate case $\rho = I_q$. This corresponds to $W^{(\rho)} = W$ so that $X^{(\rho),x_2} = X^{x_2}$, *i.e.* $(DSDS)$ is the equation of the 2-point motion in the sense of [Kun90] section 4.2 and [Har81]. This 2-point motion has been extensively investigated (see [Car85]) from an ergodic viewpoint, especially when the underlying diffusion, or more generally the stochastic flow $\Phi(\omega, x, t)$ lives on a (smooth) compact Riemannian manifold M. When this flow is smooth enough in x, the long run behaviour of such a flow (under its steady regime) can be classified owing to its Lyapunov spectrum. For what we are concerned with, this classification is based on the top Lyapunov exponent defined by

$$\lambda_1 := \limsup_{t \to +\infty} \frac{1}{t} \log \|D_x \Phi(x, t)\|$$

where $\|D_x \Phi(x, t)\|$ denotes the operator norm of the differential (tangent) of the flow. In this compact setting and when the top Lyapunov exponent is positive, the long run behaviour of the two-point on $M^2 \backslash \Delta$ has been deeply investigated in [BS88] (see also [DKK04] for further results in this direction). Such assumption implies that Δ is somewhat repulsive.

Here, we are in fact concerned with the opposite case. Our aim is to identify natural assumptions under which the invariant distribution of the 2-point motion is unique (hence equal to ν_Δ). It seems clear that these conditions should in some sense imply that the paths cluster asymptotically either in a pathwise or in a statistical sense. When $\lambda_1 < 0$, a local form of such a clustering has been obtained in [Car85] (see Proposition 2.3.3) : it is shown that at a given point "asymptotic clustering" holds with an arbitrarily high probability, provided the starting points are close enough. However, this result seems to be not sufficient to imply uniqueness of the invariant distribution for the two-point motion and is still in a compact setting.

In Sections 2 and 3, we provide precise answers under verifiable conditions on the coefficients b and σ of the original $\mathbb{R}^d$-valued diffusion, not assumed to be smooth. More precisely, we show in Section 2 that in the one-dimensional case, uniqueness of ν_Δ is almost always true (as soon as (SDE) has a unique invariant distribution) and that under some slightly more constraining conditions, the diffusion is *pathwise confluent* (*i.e.* pathwise asymptotic clustering holds). This second result slightly extends by a different method a result by Has'minskii in [Has80].

Section 3 is devoted to the multidimensional framework. We first provide a simple counter-example where uniqueness of ν_Δ does not hold. Then, we obtain some sharp criterions for uniqueness. We begin by a general uniqueness result (for ν_Δ) (Theorem 3.2) involving in an Euclidean framework (induced by a positive definite matrix S and its norm $|\,.\,|_S$) a *pseudo-scale* function f_θ designed from a non-negative continuous function $\theta : \mathbb{R}_+ \to \mathbb{R}$. Basically both uniqueness and pathwise confluence follow from conditions involving the coefficients of the diffusion b and σ, S and θ, combined with a requested behavior of the pseudo-scale function at 0^+. The main ingredient of the proof is Birkhoff's ergodic Theorem applied to the one-dimensional Itô process $f_\theta(|X_t^{x_1} - X_t^{x_2}|_S^2)$. Using additional martingale arguments, we also establish that the asymptotic pathwise confluence holds under slightly more stringent conditions.

Then, in Subsection 3.3, we draw a series of corollaries of Theorem 3.2 (illustrated on few examples) which highlight easily verifiable conditions. To this end we introduce a function $\Lambda_S : \mathbb{R}^d \times \mathbb{R}^d \backslash \Delta_{\mathbb{R}^d \times \mathbb{R}^d} \to \mathbb{R}$ called *Non-Infinitesimal S-Lyapunov* (NILS) exponent defined

for every $x,\, y\in \mathbb{R}^d,\ x\neq y$ by

$$\Lambda_S(x,y)=\frac{(b(x)-b(y)|x-y)_S}{|x-y|_S^2}+\frac{1}{2}\frac{\|\sigma(x)-\sigma(y)\|_S^2}{|x-y|_S^2}-\left(\left|\frac{(\sigma^*(x)-\sigma^*(y))S(x-y)}{|x-y|_S^2}\right|^2\right). \quad (1.4)$$

In particular we show (see Corollary 3.2) that if, for every probability measure m on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ such that $m(dx\times\mathbb{R}^d)=m(\mathbb{R}^d\times dy)=\nu$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\Lambda_S(x,y)m(dx,dy)<0$$

then ν_Δ is unique and if furthermore $\Lambda_S\leq -c_0<0$ on a uniform stripe around the diagonal $\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$, then pathwise confluence holds true. Moreover, under a *directional ellipticity* condition on σ, we show that the negativity of Λ_S (at least in an integrated sense) can be localized near the diagonal (see Subsection 3.3 for details). A differential version of the criterion is established when b and σ are smooth (see Corollary 3.3).

Note that these criterions obtained in the case $\rho=I_q$ can be extended to the (last) case $\rho^*\rho=I_q$ using that $W^{(\rho)}=\rho W$ is still a standard B.M. (think to $\rho=-1$ when $d=1$). For the sake of simplicity (and since it is of little interest for the practical implementation of the Richardson-Romberg extrapolation), we will not consider this case in the paper.

We end the first part by some examples and by an application to gradient systems ($b=-\nabla U$ and a constant σ function). In particular, we obtain that our criterions can be applied to some situations where the potential is not convex. More precisely, we prove that for a large class of non-convex potentials, super-quadratic at infinity , the 2-point motion is weakly confluent if the diffusive component σ is sufficiently large. Furthermore, in the particular case $U(x)=(|x|^2-1)^2$, we prove that the result is true for every $\sigma>0$.

The second part of the paper (Section 4) is devoted to a first attempt in a long run ergodic setting to combine the Richardson-Romberg extrapolation with a control of the variance of this procedure (see [Pag07] in a finite horizon framework). To this end we consider two Euler schemes with decreasing steps γ_n and $\tilde\gamma_n$ satisfying $\tilde\gamma_{2n-1}=\tilde\gamma_{2n}=\gamma_n/2$ and ρ-correlated Brownian motion increments. We show that the optimal efficiency of the Richardson-Romberg extrapolation in this framework is obtained when $\rho=0$, at least when the above uniqueness problem for ν_Δ is satisfied. To support this claim we establish a Central Limit Theorem whose variance is analyzed as a function of ρ.

NOTATIONS. • $|x|=\sqrt{xx^*}$ denotes the canonical Euclidean norm of $x\in\mathbb{R}^d$ (x^* transpose of the column vector x).
• $\|A\|=\sqrt{\mathrm{Tr}(AA^*)}$ if $A\in\mathcal{M}(d,q,\mathbb{R})$ and A^* is the transpose of A (which is but the canonical Euclidean norm on $\mathbb{R}^{d^2}$).
• $\Delta_{\mathbb{R}^d\times\mathbb{R}^d}=\{(x,x),\ x\in\mathbb{R}^d\}$ denotes the diagonal of $\mathbb{R}^d\times\mathbb{R}^d$.
• $\mathcal{S}(d,\mathbb{R})=\{S\in\mathcal{M}(d,d,\mathbb{R}),\ S^*=S\}$, $\mathcal{S}^+(d,\mathbb{R})$ the subset of $\mathcal{S}(d,\mathbb{R})$ of non-negative matrices, $\mathcal{S}^{++}(d,\mathbb{R})$ denotes the subset of positive definite such matrices and $\sqrt{S}$ denotes the unique square root of $S\in\mathcal{S}^+(d,\mathbb{R})$ in $\mathcal{S}^{++}(d,\mathbb{R})$ (which commutes with S). $x\otimes y=xy^*=[x_iy_j]\in\mathcal{M}(d,d,\mathbb{R})$, $x,\, y\in\mathbb{R}^d$.
• If $S\in\mathcal{S}^{++}(d,\mathbb{R})$, we denote by $(\,.\,|\,.\,)_S$ and by $|\,.\,|_S$, the induced inner product and norm on $\mathbb{R}^d$, defined by $(x|y)_S=(x|Sy)$ and $|x|_S^2=(x|x)_S$ respectively. Finally, for $A\in\mathcal{M}(d,d,\mathbb{R})$, we set $\|A\|_S^2=\mathrm{Tr}(A^*SA)$.
• $\mu_n\overset{(\mathbb{R}^d)}{\Longrightarrow}\mu$ denotes the weak convergence of the sequence $(\mu_n)_{n\geq1}$ of probability measures defined on $(\mathbb{R}^d,\mathcal{B}or(\mathbb{R}^d))$ toward the probability measure μ. $\mathcal{P}(X,\mathcal{A})$ denotes the set of probability distributions on $(X,\mathcal{A})$.

• For every function $f : \mathbb{R}^d \to \mathbb{R}$, define the Lipschitz coefficient of f by $[f]_{\mathrm{Lip}} = \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|} \leq +\infty$.

2 The one-dimensional case

We first show that, in the one-dimensional case $d = q = 1$, uniqueness of ν implies that ν_Δ, as defined in the introduction, is the unique invariant distribution of the duplicated diffusion. The main theorem of this section is Theorem 2.1 which consists of two claims. The first one establishes this uniqueness claim using some ergodic-type arguments. Note that we do not require that σ never vanishes. The second claim is an asymptotic pathwise confluence property for the diffusion in its own scale, established under some slightly more stringent assumptions involving the scale function p, see below. This second result, under slightly less general assumptions, is originally due to Has'minskii (see [Has80], Appendix to the English edition, Theorem 2.2, p.308). It is revisited here by different techniques, mainly comparison results for one dimensional diffusions and ergodic arguments. Note that uniqueness of ν_Δ can always be retrieved from asymptotic confluence (see Remark 2.1).

Before stating the result, let us recall some definitions. We denote by M the *speed measure* of the diffusion classically defined by $M(d\xi) = \dfrac{d\xi}{(\sigma^2 p')(\xi)}$, where p is the *scale function* defined (up to a constant) by

$$p(x) = \int_{x_0}^{x} d\xi e^{-\int_{x_0}^{\xi} \frac{2b}{\sigma^2}(u)du}, \ x \in \mathbb{R}.$$

Obviously, we will consider p only when it makes sense *as a finite function* (so is the case if b/σ^2 is locally integrable on the real line). We are now in position to state the result.

Theorem 2.1. *Assume that b and σ are continuous functions on $\mathbb{R}$ being such that strong existence, pathwise uniqueness and the Feller Markov property hold for (SDE) from any $x \in \mathbb{R}$. Assume furthermore that there exists $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$, strictly increasing, with $\lambda(0) = 0$ and $\int_{0^+} \lambda(u)^{-2}du = +\infty$ such that for all $x, y \in \mathbb{R}$, $|\sigma(y) - \sigma(x)| \leq \lambda(|x - y|)$. Then, the following claims hold true.*

(a) If $(X_t)_{t\geq 0}$ admits a unique invariant distribution ν, the distribution $\nu_\Delta = \nu \circ (\xi \mapsto (\xi, \xi))^{-1}$ is the unique invariant measure of the duplicated diffusion $(X_t^{x_1}, X_t^{x_2})_{t\geq 0}$.

(b) (Has'minskii) Assume that the scale function p is well-defined as a real function on the real line and that,

$$\lim_{x\to\pm\infty} p(x) = \pm\infty \quad \text{and} \quad M \text{ is finite.}$$

Then, $\nu = M/M(\mathbb{R})$ is the unique invariant distribution of $(X_t)_{t\geq 0}$ and $(p(X_t))_{t\geq 0}$ is pathwise confluent: $\mathbb{P}$-a.s., for every $x_1, x_2 \in \mathbb{R}$, $p(X_t^{x_1}) - p(X_t^{x_2})$ tends to 0 when $t \to +\infty$.

Remark 2.1. ▷ The general assumptions on b and σ are obviously fulfilled whenever these functions are locally Lipschitz with linear growth.

▷ The proofs of both claims are based on (typically one-dimensional) comparison arguments. This also explains the assumption on σ which is a classical sufficient assumption

to ensure comparison of solutions, namely, if $x_1 \leq x_2$, then $X_t^{x_1} \leq X_t^{x_2}$ for every $t \geq 0$ $a.s.$ (see [IW77]).

$\triangleright$ The additional assumptions made in (b) imply the uniqueness of ν (see the proof below). The uniqueness of the invariant distribution ν_Δ for the duplicated diffusion follows by (a). However, it can also be viewed as a direct consequence of the asymptotic pathwise confluence of $p(X_t^{x_i})$, $i = 1, 2$ as $t \to +\infty$. Actually, if for all $x_1, x_2 \in \mathbb{R}^d$, $p(X_t^{x_1}) - p(X_t^{x_2}) \xrightarrow{t\to+\infty} 0$ $a.s$, we deduce that for any invariant distribution μ of (X^{x_1}, X^{x_2}) and every $K > 0$

$$\int_{\mathbb{R}} \Big(|p(x_1) - p(x_2)| \wedge K\Big)\mu(dx_1, dx_2) \leq \limsup_{t\to+\infty} \frac{1}{t} \int_0^t \mathbb{E}_\mu\Big(|p(X_s^{x_1}) - p(X_s^{x_2})| \wedge K\Big)ds = 0.$$

As a consequence, $p(x_1) = p(x_2)$ $\mu(dx_1, dx_2)$-$a.s.$ Since p is an increasing function, it follows that $\mu(\{(x, x), x \in \mathbb{R}\}) = 1$ and thus that $\mu = \nu_\Delta$.

$\triangleright$ As mentioned before, (b) slightly extends a result by Has'minskii obtained in [Has80] with different methods and under the additional assumption that σ never vanishes (whereas we only need the scale function p to be finite which allows e.g. for the existence of integrable singularities of $\frac{b}{\sigma^2}$). Note however that the case of an infinite speed measure M (which corresponds to null recurrent diffusions) is also investigated in [Has80], requiring extra non-periodicity assumptions on σ.

Proof. (a) Throughout the proof we denote by $(X_t^{x_1}, X_t^{x_2})$ the duplicated diffusion at time $t \geq 0$ and by $(Q_t((x_1, x_2), dy_1, dy_2))_{t\geq 0}$ its Feller Markov semi-group. The set $\mathcal{I}_{DSDS}$ of invariant distributions of $(Q_t)_{t\geq 0}$ is clearly nonempty, convex and weakly closed. Since any such distribution μ has ν as marginals (in the sense $\mu(dx_1 \times \mathbb{R}) = \mu(\mathbb{R} \times dx_2) = \nu$), the set $\mathcal{I}_{DSDS}$ is tight and consequently weakly compact in the the topological vector space of signed measures on $(\mathbb{R}^2, \mathcal{B}or(\mathbb{R}^2))$ equipped with the weak topology. As a consequence of the Krein-Millman Theorem, $\mathcal{I}_{DSDS}$ admits extremal distributions and is the convex hull of these extremal distributions.

Let μ be such an extremal distribution and consider the following three subsets of $\mathbb{R}^2$:

$$A^+ = \{(x_1, x_2),\ x_2 > x_1\},\ A^- = \{(x_1, x_2),\ x_1 > x_2\} \text{ and } A_0 = \{(x, x),\ x \in \mathbb{R}\} = \Delta_{\mathbb{R}^2}.$$

We first want to show that if $\mu(A^+) > 0$ then the conditional distribution μ^{A^+} defined by $\mu^{A^+} = \frac{\mu(\,.\,\cap A^+)}{\mu(A^+)}$ is also an invariant distribution for $(Q_t)_{t\geq 0}$. Under the above assumptions on b and σ, one derives from classical comparison theorems and strong pathwise uniqueness arguments for the solutions of (SDE) (see *e.g.* [IW77]) that

$$\forall\, (x_1, x_2) \in {}^cA^+ = \mathbb{R}^2 \setminus A^+, \quad Q_t((x_1, x_2), {}^cA^+) = 1.$$

We deduce that for every $(x_1, x_2) \in \mathbb{R}^2$ and $t \geq 0$,

$$Q_t((x_1, x_2), A^+) = \mathbb{P}\big((X_t^{x_1}, X_t^{x_2}) \in A^+\big) = \mathbf{1}_{A^+}(x_1, x_2)\mathbb{P}(\tau_{x_1,x_2} > t)$$

where $\tau_{x_1,x_2} = \inf\{t \geq 0, X_t^{x_2} \leq X_t^{x_1}\}$. The second equality follows from the pathwise uniqueness since no bifurcation can occur. Now, let $\mu \in \mathcal{I}_{DSDS}$. Integrating the above equality and letting t go to infinity implies

$$\mu(A^+) = \int_{A^+} \mu(dx_1, dx_2)\mathbb{P}(\tau_{x_1,x_2} = +\infty).$$

If $\mu(A^+) > 0$, then $\mu(dx_1, x_2)$-$a.s.$ $\mathbb{P}(\tau_{x_1,x_2} = +\infty) = 1$ on A^+ $i.e.$ $X_t^{x_2} > X_t^{x_1}$ for every $t \geq 0$ $a.s.$. As a consequence, $\mu(dx_1, dx_2)$-$a.s.$, for every $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\mathbf{1}_{(x_1,x_2)\in A^+} Q_t((x_1, x_2), B) = \mathbf{1}_{(x_1,x_2)\in A^+} Q_t((x_1, x_2), B \cap A^+) = Q_t((x_1, x_2), B \cap A^+)$$

where we used again that $Q_t((x_1, x_2), A^+) = 0$ if $x_2 \leq x_1$. Then, since μ is invariant, we deduce from an integration of the above equality that

$$\mu(B \cap A^+) = \int_{\mathbb{R}^2} Q_t((x_1, x_2), B)\mathbf{1}_{(x_1,x_2)\in A^+}\mu(dx_1, dx_2).$$

It follows that if $\mu(A^+) > 0$, μ^{A^+} is invariant.
If $\mu(A^+) < 1$, one shows likewise that $\mu^{{}^cA^+}$ an invariant distribution for $(Q_t)_{t\geq 0}$ as well. Then, if $\mu(A^+) \in (0, 1)$, then μ is a convex combination of elements of $\mathcal{I}_{DSDS}$

$$\mu = \mu(A^+)\mu^{A^+} + \mu({}^cA^+)\mu^{{}^cA^+}$$

so that μ cannot be extremal. Finally $\mu(A^+) = 0$ or 1.

Assume $\mu(A^+) = 1$ so that $\mu = \mu(. \cap A^+)$. This implies that $X_0^1 > X_0^2$ $\mathbb{P}_\mu$-$a.s.$. But μ being invariant, both its marginals are ν $i.e.$ X_0^1 and X_0^2 are ν-distributed. This yields a contradiction. Indeed, let φ be a bounded increasing positive function. For instance, set $\varphi(u) := 1 + \frac{u}{\sqrt{u^2+1}}$, $u \in \mathbb{R}$. Then, $\mathbb{E}[\varphi(X_0^1) - \varphi(X_0^2)] > 0$ since $X_0^1 > X_0^2$ $\mathbb{P}_\mu$-$a.s.$ but we also have $\mathbb{E}[\varphi(X_0^1) - \varphi(X_0^2)] = 0$ since X_0^1 and X_0^2 have the same distribution. This contradiction implies that $\mu(A_+) = 0$.

One shows likewise that $\mu(A^-) = 0$ if μ is an extremal measure. Finally any extremal distribution of $\mathcal{I}_{DSDS}$ is supported by $A_0 = \Delta_{\mathbb{R}^2}$. Given the fact that the marginals of μ are ν this implies that $\mu = \nu_\Delta = \nu \circ (x \mapsto (x, x))^{-1}$ which in turn implies that $\mathcal{I}_{DSDS} = \{\nu_\Delta\}$.

(b) Since the speed measure M is finite and σ never vanishes, the distribution $\nu(d\xi) = M(d\xi)/M(\mathbb{R})$ is the unique invariant measure of the diffusion. Thus, by (a), we also have the uniqueness of the invariant distribution for the duplicated diffusion.Let x_1, $x_2 \in \mathbb{R}$. If $x_1 > x_2$ then $X_t^{x_1} \geq X_t^{x_2}$, still by a comparison argument, and $p(X_t^{x_1}) \geq p(X_t^{x_2})$ since p is increasing. Consequently $M_t^{x_1,x_2} = p(X_t^{x_1}) - p(X_t^{x_2})$, $t \geq 0$, is a non-negative continuous local martingale, hence $\mathbb{P}$-$a.s.$ converging toward a finite random limit $\ell_\infty^{x_1,x_2} \geq 0$. One proceeds likewise when $x_1 < x_2$ (with $\ell_\infty^{x_1,x_2} \leq 0$). When $x_1 = x_2$, $M_t = \ell_\infty^{x_1,x_2} \equiv 0$. The aim is now to show that $\ell_\infty^{x_1,x_2} = 0$ $a.s.$ To this end, we introduce

$$\mu_t(dy_1, dy_2) := \frac{1}{t}\int_0^t Q_s((x_1, x_2), dy_1, dy_2)ds, \qquad (x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$$

and we want to check that for every $(x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$, $(\mu_t(dy_1, dy_2))_{t\geq 1}$ converges weakly to ν_Δ. Owing to the uniqueness of ν_Δ established in (a) and to the fact that any weak limiting distribution of $(\mu_t(dy_1, dy_2))_{t\geq 1}$ is always invariant (by construction), it is enough to prove that $(\mu_t(dy_1, dy_2))_{t\geq 1}$ is tight. Since the tightness of a sequence of probability measures defined on a product space is clearly equivalent to that of its first and second marginals, it is here enough to prove the tightness of $(t^{-1}\int_0^t P_s(x_0, dy)ds)_{t\geq 1}$ for any $x_0 \in \mathbb{R}$.
Let $x_0 \in \mathbb{R}$. Owing to the comparison theorems, we have for all $t \geq 0$ and $M \in \mathbb{R}$,

$P_t(x_0,[M,+\infty)) \le P_t(x,[M,+\infty))$ if $x \ge x_0$ and $P_t(x_0,(-\infty,M]) \le P_t(x,(-\infty,M])$ if $x_0 \ge x$. Since ν is invariant and equivalent to the Lebesgue measure, we deduce that

$$P_t(x_0,[M,+\infty)) \le \frac{\nu([M,+\infty))}{\nu([x_0,+\infty))} \quad \text{and} \quad P_t(x_0,(-\infty,M)) \le \frac{\nu((-\infty,M))}{\nu((-\infty,x_0])}.$$

The tightness of $(P_t(x_0,dy))_{t\ge1}$ follows (from that of ν) and we derive from what preceeds that

$$\forall\,(x_1,x_2)\in\mathbb{R}^d\times\mathbb{R}^d, \quad \frac{1}{t}\int_0^t Q_s((x_1,x_2),dy_1,dy_2)ds \overset{(\mathbb{R}^d)}{\Longrightarrow} \nu_\Delta(dy_1,dy_2).$$

Now, note that for every $L\in\mathbb{N}$, the function $g_L:(y_1,y_2)\mapsto |p(y_1)-p(y_2)|\wedge L$ is continuous and bounded. Hence by Césaro's Theorem, we have that

$$\frac{1}{t}\int_0^t Q_s(g_L)(x_1,x_2)ds = \frac{1}{t}\int_0^t \mathbb{E}\,g_L(X_s^{x_1},X^{x_2})ds \longrightarrow \mathbb{E}\,(|\ell_\infty^{x_1,x_2}|\wedge L)$$

whereas, by the above weak convergence of $(\mu_t(dy_1,dy_2))_{t\ge1}$, we get

$$\frac{1}{t}\int_0^t Q_s(g_L)(x_1,x_2)ds \longrightarrow \int_{\mathbb{R}^d} g_L(y_1,y_2)\nu_\Delta(dy_1,dy_2) = 0 \quad \text{as} \quad t\to+\infty$$

since g_L is identically 0 on $\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$. It follows, by letting L go to infinity, that

$$\mathbb{E}\,|\ell_\infty^{x_1,x_2}| = 0.$$

This implies $\ell_\infty^{x_1,x_2}=0$ $\mathbb{P}$-$a.s.$ which in turn implies that

$$\mathbb{P}\text{-}a.s. \quad p(X_t^{x_1})-p(X_t^{x_2}) \longrightarrow 0 \quad \text{as} \quad t\to+\infty.$$

Finally, it remains to prove that we can exchange the quantifiers, *i.e.* that $\mathbb{P}$-$a.s.$, $p(X_t^{x_1})-p(X_t^{x_2}) \longrightarrow 0$ for every x_1, x_2. Assume that $x_1\ge x_2$. Again by the comparison theorem and the fact that p increases, we have $0\le p(X_t^{x_1})-p(X_t^{x_2})\le p(X_t^{\lfloor x_1\rfloor+1})-p(X_t^{\lfloor x_2\rfloor})$. This means that we can come down to a countable set of starting points. □

In the continuity of the second part of Theorem 2.1(b), it is natural to wonder whether a one-dimensional diffusion is asymptotically confluent, *i.e.* when for all $x_1,x_2\in\mathbb{R}$, $X_t^{x_1}-X_t^{x_2}$ tends to 0 $a.s$ as $t\to+\infty$. In the following corollary, we show that such property holds in a quite general setting.

Corollary 2.1. (a) *Assume the hypothesis of Theorem 2.1(b) hold. If furthermore,*

$$\sigma \text{ never vanishes} \qquad \text{and} \qquad \limsup_{|x|\to+\infty}\int_0^x \frac{b}{\sigma^2}(\xi)d\xi < +\infty$$

then, $\mathbb{P}$-a.s., for every x_1, $x_2\in\mathbb{R}$,

$$X_t^{x_1}-X_t^{x_2} \longrightarrow 0 \quad \text{as} \quad t\to+\infty.$$

(b) *The above condition is in particular satisfied if there exists $M>0$ such that for all*

$$|x|>M \Longrightarrow \operatorname{sign}(x)b(x)\le 0.$$

Proof. (a) Under the assumptions of the theorem, p is continuously differentiable on $\mathbb{R}$ and

$$p'(x) = e^{-\int_{x_0}^{x} \frac{2b}{\sigma^2}(u)du}, \ x \in \mathbb{R}.$$

Then it is clear that $p'_{\inf} = \inf_{x\in\mathbb{R}} p'(x) > 0$ iff $\limsup_{|x|\to+\infty} \int_{x_0}^{x} \frac{2b}{\sigma^2}(\xi)d\xi < +\infty$. By the fundamental theorem of calculus, we know that,

$$|X_t^{x_1} - X_t^{x_2}| \leq \frac{1}{p'_{\min}} |p(X_t^{x_1}) - p(X_t^{x_2})|$$

and the result follows from Theorem 2.1(b).

(b) Since σ never vanishes, p'' is well-defined and for every $x \in \mathbb{R}$, $p''(x) = -\frac{2b(x)p'(x)}{\sigma^2(x)}$. Using that p' is positive, we deduce from the assumptions that

$$\exists M > 0 \quad \text{such that} \quad \begin{cases} p''(x) \geq 0, & x \geq M \\ p''(x) \leq 0, & x \leq -M. \end{cases}$$

Now, p' being continuous, it follows that p' attains a positive minimum $p'_{\min} > 0$. □

Examples. 1. Let U be a positive a twice differentiable function such that $\lim_{|x|\to+\infty} U(x) = +\infty$ and consider the one-dimensional Kolmogorov equation

$$dX_t = -U'(X_t)dt + \sigma dW_t$$

where $\sigma > 0$. Then,

$$\liminf_{|x|\to+\infty} xU'(x) > \frac{\sigma^2}{2} \quad \Longrightarrow \quad X_t^x - X_t^y \xrightarrow{t\to+\infty} 0 \ a.s.$$

Note that in particular, this result holds true even if U has several local minimas.

2. Let $\sigma = \mathbb{R} \to (0, +\infty)$ be a locally Lipschitz continuous function with linear growth so that the SDE

$$dX_t = \sigma(X_t)dW_t$$

defines a (Markov) flow $(X_t^x)_{t\geq 0}$ of local martingales. If $\frac{1}{\sigma} \in L^2(\mathbb{R}, \mathcal{B}or(\mathbb{R}), \lambda)$ then there exists a unique invariant measure $\nu(d\xi) = c_\sigma \frac{d\xi}{\sigma^2(\xi)}$ and $(X_t^{x_i})_{t\geq 0}$, $i = 1, 2$ is pathwise confluent (in the sense of Theorem 2.1(b)) since $p(x) = x$.

Note that the linear growth assumption cannot be significantly relaxed since a stationary process cannot be a true martingale which in turn implies that ν has no (finite) first moment.

3 The multidimensional case

In this section, we begin by an example of a multidimensional Brownian diffusion (X^{x_1}, X^{x_2}) for which ν_Δ (image of ν on the diagonal) is not the only one invariant distribution. Thus, Theorem 2.1 is specific to the case $d = 1$ and we can not hope to get a similar result for the general case $d \geq 2$. It is of course closely related to the classification of two-point motion on smooth compact Riemannian manifolds since the unit circle will turn out to be a uniform attractor of the diffusion.

3.1 Counterexample in 2-dimension

Roughly speaking, saying that ν_Δ is the only one invariant distribution means in a sense that $X_t^x - X_t^y$ has a tendency to converge towards 0 when $t \to +\infty$. Thus, the idea in the counterexample below is to build a "turning" two-dimensional ergodic process where the angular difference between the two coordinates does not depend on t. Such a construction leads to a model where the distance between the two coordinates can not tend to 0 (Note that some proofs are deferred to Appendix B). We consider the 2-dimensional SDE with Lipschitz continuous coefficients defined by, $\forall x \in \mathbb{R}^2$

$$\begin{aligned} b(x) &= \Big(x\mathbf{1}_{\{0\le|x|\le1\}} - \frac{x}{|x|}\mathbf{1}_{\{|x|\ge1\}}\Big)(1-|x|) \\ \sigma(x) &= \vartheta \mathrm{Diag}(b(x)) + \begin{bmatrix} 0 & -cx^2 \\ cx^1 & 0, \end{bmatrix} \end{aligned}$$

where ϑ, $c \in (0,+\infty)$ are fixed parameters.

Switching to polar coordinates $X_t = (r_t \cos\varphi_t, r_t \sin\varphi_t)$, $t \in \mathbb{R}_+$, we obtain that this SDE also reads

$$dr_t = \min(r_t,1)(1-r_t)(dt + \vartheta dW_t^1), \quad r_0 \in \mathbb{R}_+ \tag{3.5}$$

$$d\varphi_t = c\, dW_t^2, \quad \varphi_0 \in [0,2\pi), \tag{3.6}$$

where $x_0 = r_0(\cos\varphi_0, \sin\varphi_0)$ and $W = (W^1, W^2)$ is a standard 2-dimensional Brownian motion.

Standard considerations about Feller classification (see Appendix B for details) show that, if $x_0 \neq 0$ (*i.e.* $r_0 > 0$) and $\vartheta \in (0,\sqrt{2})$ then

$$r_t \longrightarrow 1 \ \text{ as } \ t \to +\infty, \tag{3.7}$$

while it is classical background that

$$\mathbb{P}\text{-}a.s. \quad \forall \varphi_0 \in \mathbb{R}_+, \quad \frac{1}{t}\int_0^t \delta_{e^{i(\varphi_0 + cW_s^2)}} ds \Longrightarrow \lambda_{\mathcal{S}_1} \quad \text{as} \quad t \to +\infty$$

where $\mathcal{S}_1$ denotes the unit circle of $\mathbb{R}^2$. Combining these two results straightforwardly yields

$$\forall\, x \in \mathbb{R}^2 \setminus \{(0,0)\},\ \mathbb{P}\text{-}a.s. \quad \frac{1}{t}\int_0^t \delta_{X_s^x} ds \overset{(\mathbb{R}^2)}{\Longrightarrow} \lambda_{\mathcal{S}_1} \quad \text{as} \quad t \to +\infty.$$

On the other hand, given the form of φ_t, it is clear that if $x = r_0 e^{i\varphi_0}$ and $x' = r_0' e^{i\varphi_0'}$, r_0, $r_0' \neq 0$, φ_0, $\varphi_0' \in [0,2\pi)$, then

$$\lim_{t\to+\infty} |X_t^x - X_t^{x'}| = |e^{i(\varphi_0 - \varphi_0')} - 1|$$

which in turn implies that

$$\lim_{t\to+\infty} \frac{1}{t}\int_0^t |X_s^x - X_s^{x'}| ds = |e^{i(\varphi_0-\varphi_0')} - 1| \qquad \mathbb{P}\text{-}a.s.$$

This limit being different from 0 as soon as $\varphi_0 \neq \varphi_0'$, one derives, as a consequence, that ν_Δ cannot be the only invariant distribution. In fact, a more precise statement can be proved.

Proposition 3.1. *(a) A distribution μ is invariant for the semi-group $(Q_t)_{t\geq 0}$ of the duplicated diffusion if and only if μ has the following form:*

$$\mu = \mathcal{L}(e^{i\Theta}, e^{i(\Theta+V)}) \tag{3.8}$$

where Θ is uniformly distributed over $[0, 2\pi]$ and V is a $[0, 2\pi)$-valued random variable independent of Θ

(b) When $V = 0$ a.s., we retrieve ν_Δ whereas, when V also has uniform distribution on $[0, 2\pi]$, we obtain $\nu \otimes \nu$. Finally, μ is extremal in the convex set of $(Q_t)_{t\geq 0}$ invariant distributions if and only if there exists $\theta_0 \in [0, 2\pi)$ such that $V = \theta_0$ a.s.

The proof is postponed to Appendix B. However, note that the claim about extremal invariant distributions follows from the fact that for every $\theta \in [0, 2\pi)$, $(Q_t)_{t\geq 0}$ leaves the set $\Gamma_\theta := \{(e^{i\varphi}, e^{i\varphi'}) \in \mathcal{S}_1 \times \mathcal{S}_1, \varphi' - \varphi \equiv \theta \,\mathrm{mod.}\, 2\pi\}$ stable.

Remark 3.2. In the above counterexample, the invariant measure of $(r_t)_{t\geq 0}$ is the Dirac measure δ_1. In fact, setting again $x = \varepsilon_0 e^{i\varphi_0}$ and $x' = r'_0 e^{i\varphi'_0}$ and using that $X_t^x - X_t^{x'} = r_t^x \left(e^{i(\varphi_0+W_t^2)} - e^{i(\varphi'_0+W_t^2)}\right) + (r_t^x - r_t^{x'})e^{i(\varphi'_0+W_t^2)}$, an easy adaptation of the above proof shows that it can be generalized to any ergodic non-negative process $(r_t)_{t\geq 0}$ solution to an autonomous SDE and satisfying the following properties:

- Its unique invariant distribution π satisfies $\pi(\mathbb{R}_+^*) = 1$.
- For every $x, y \in (0, +\infty)$, $r_t^x - r_t^y \longrightarrow 0$ *a.s.* as $t \to +\infty$.

For instance, let $(X_t^x)_{t\geq 0}$ be an Ornstein-Uhlenbeck process satisfying the SDE $dX_t = -X_t dt + \sigma dW_t, X_0 = x$. Set $r_t^x = (X_t^x)^2$ (this is a special case of the Cox-Ingersoll-Ross process). The process (r_t^x) clearly satisfies the first two properties. Furthermore, $(X_t^x)_{t\geq 0}$ satisfies *a.s.* for every $x, y \in \mathbb{R}$ and every $t \geq 0$, $|X_t^x - X_t^y| = |x - y|e^{-t}$. Then, since for every $x \in \mathbb{R}$,

$$\frac{X_t^x}{t} = -\frac{1}{t}\int_0^t X_s^x ds + \sigma\frac{W_t}{t} \to 0 \quad a.s. \quad \text{as} \quad t \to +\infty,$$

it follows that $(r_t^x)_{t\geq 0}$ also satisfies for all positive x, y, $r_t^x - r_t^y \longrightarrow 0$ *a.s.* as $t \to +\infty$ (Many other examples can be built using Corollary 2.1).
Finally, note that if $\mu = \mathcal{L}(Re^{i\Theta}, Re^{i(\Theta+V)})$ where R, Θ and V are independent random variables such that the distributions of R and Θ are respectively π and the uniform distribution on $[0, 2\pi]$ and V takes values in $[0, 2\pi)$, then μ is an invariant distribution of the associated duplication system.

In connection with this counterexample we can mention a general result on the Brownian flows of Harris (see [Har81],[Kun90] Theorem 4.3.2). The theorem gives conditions on b and σ under which ν is an invariant measure of the one point motion $(X_t^x)_{t\geq 0}$ and $\nu \otimes \nu$ is an invariant measure of the two point motion $(X_t^{x_1}, X_t^{x_2})_{t\geq 0}$.

3.2 Uniqueness of the invariant measure: (S, θ)-confluence

In the sequel of this section, we propose criterions for the uniqueness of the invariant distribution of the duplicated system in the multidimensional case. The underlying idea of the criterions discussed below is to analyze the coupled diffusion process (X^{x_1}, X^{x_2}) through the squared distance process $r_t = |X_t^{x_1} - X_t^{x_2}|_S^2$ (where we recall that for a given positive definite matrix S, $|\,.\,|_S$ is the Euclidean norm on $\mathbb{R}^d$ induced by the scalar product

$(x|y)_S = (x|Sy)$). It is somewhat similar to that of Has'minskii's test for explosion of diffusions in $\mathbb{R}^d$ or to the one proposed in Chen and Li's work devoted to the coupling of diffusions (see [CL89]). We begin by a general abstract result under an assumption depending on a continuous function $\theta : (0, +\infty) \to \mathbb{R}_+$ to be specified further on. Then, more explicit *pointwise* or *integrated* criterions are derived in the next subsections. In particular, one involves a kind of bi-variate non-infinitesimal Lyapunov exponent. Let us introduce some notations. For a probability measure ν on $\mathbb{R}^d$, we set

$$C_\nu^\star = \Big\{ m \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), m(dx \times \mathbb{R}^d) = m(\mathbb{R}^d \times dy) = \nu, m(\Delta_{\mathbb{R}^d \times \mathbb{R}^d}) = 0 \Big\}.$$

For $S \in \mathcal{S}^{++}(d, \mathbb{R})$, we also set

$$[b]_{S,+} = \sup_{x \neq y} \frac{(b(x) - b(y)|x - y)_S}{|x - y|_S^2}.$$

Note that if $[b]_{S,+} < +\infty$ and if σ is Lipschitz continuous, strong existence, pathwise uniqueness and the Feller Markov property hold for (SDE).

For a continuous function $\theta : (0, +\infty) \to \mathbb{R}_+$, we define the *pseudo-scale* $\mathcal{C}^2$-function f_θ and its companion g_θ by

$$\forall u \in (0, +\infty), \qquad f_\theta(u) = \int_1^u e^{\int_\xi^1 \frac{\theta(w)}{w} dw} d\xi \ \text{ and } \ g_\theta(u) = u f_\theta'(u). \tag{3.9}$$

Finally, for S and θ defined as above, we define the (S, θ)-*confluence* function $\Psi_{\theta,S}$ on ${}^c\Delta_{\mathbb{R}^{2d}}$ by

$$\Psi_{\theta,S}(x, y) = (b(x) - b(y)|x - y)_S + \frac{1}{2}\|\sigma(x) - \sigma(y)\|_S^2 - \theta(|x - y|_S^2)\Big|(\sigma^*(x) - \sigma^*(y))\frac{S(x - y)}{|x - y|_S}\Big|^2.$$

Let us now state the result.

Theorem 3.2. *Let $S \in \mathcal{S}^{++}(d, \mathbb{R})$. Assume that b is a continuous function such that $[b]_{S,+} < +\infty$ and σ is Lipschitz continuous. Assume that the set $\mathcal{I}_{SDE}$ of invariant distributions of SDE is (nonempty, convex and) weakly compact. Furthermore, assume that for every $m \in C_\nu^\star$, the following (S, θ)-confluence condition is satisfied: there exists a continuous function $\theta : (0, +\infty) \to \mathbb{R}_+$ such that*

$$\begin{cases} (i) \limsup_{u \to 0^+} \displaystyle\int_u^1 \frac{\theta(w) - 1}{w} dw < +\infty. \\ (ii) \displaystyle\int_{\mathbb{R}^d \times \mathbb{R}^d} f_\theta'(|x - y|_S^2) \Psi_{\theta,S}(x, y) m(dx, dy) < 0. \end{cases} \tag{3.10}$$

(a) Weak confluence (Uniqueness of both invariant distributions)*: Then, if $\mathcal{I}_{DSDS}$ denotes the set of invariant distributions of the duplicated system $(DSDS)$, one has*

$$\mathcal{I}_{SDE} = \{\nu\} \qquad \text{and} \qquad \mathcal{I}_{DSDS} = \{\nu_\Delta\}$$

keeping in mind that $\nu_\Delta = \nu \circ (x \mapsto (x, x))^{-1}$.

(b) Pathwise confluence: *Let $\theta : (0, +\infty) \to \mathbb{R}_+$ be a continuous function such that $\displaystyle\int_0^1 e^{\int_v^1 \frac{\theta(w)}{w} dw} dv < +\infty$ and such that*

$$\forall\, x,\, y \in \mathbb{R}^d,\ x \neq y, \quad \Psi_{\theta,S}(x, y) < 0.$$

If furthermore, for every $x \in \mathbb{R}^d$, $\Big(\frac{1}{t}\int_0^t P_s(x,dy)ds\Big)_{t\geq 1}$ *is tight, we have a.s. pathwise asymptotic confluence:*

$$\forall\, x_1,\, x_2 \in \mathbb{R}^d, \quad X_t^{x_1} - X_t^{x_2} \longrightarrow 0 \;\; as \;\; t \to +\infty \;\; \mathbb{P}\text{-}a.s. \tag{3.11}$$

Remark 3.3. ▷ Owing to Assumption (i) and to $[b]_{S,+} < +\infty$, $(x,y) \mapsto f'_\theta(|x-y|^2)\Psi_{\theta,S}(x,y)$ is always bounded from above on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ so that the integrals with respect to $m \in C_\nu^\star$ are well-defined. Also note that since f'_θ is positive, Assumption (ii) holds in particular if there exists θ and S such that the (S,θ)-confluence function $\Psi_{\theta,S}$ is negative on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$.

▷ If we also assume in (a), that $\Big(\frac{1}{t}\int_0^t P_s(x,dy)ds\Big)_{t\geq 1}$ is tight then, so is $\Big(\frac{1}{t}\int_0^t Q_s(x,x',dy,dy')ds\Big)_{t\geq 1}$. Since, by construction, the weak limiting distributions of this sequence as $t \to +\infty$ are invariant distributions, it follows that $\frac{1}{t}\int_0^t Q_s(x,x',dy,dy')ds$ weakly converges to ν_Δ as $t \to +\infty$. This motivates the "weak confluence" terminology.

▷ It is natural to wonder if the assumptions for pathwise asymptotic confluence (claim (b)) are more stringent than Assumptions (i) and (ii). The fact that $\Psi_{\theta,S} < 0$ on $\mathbb{R}^d \times \mathbb{R}^d \backslash \Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ implies Assumption (ii) has already been mentioned. One can also checks that $\int_0^1 e^{\int_v^1 \frac{\theta(w)}{w}dw}dv < +\infty$ implies Assumption (i): one first derives from the Cauchy criterion that $\int_0^1 e^{\int_v^1 \frac{\theta(w)}{w}dw}dv < +\infty$ implies that $\int_{\frac{u}{2}}^u e^{\int_v^1 \frac{\theta(w)}{w}dw}dv \to 0$ as $u \to 0^+$. Using that $v \mapsto e^{\int_v^1 \frac{\theta(w)}{w}dw}$ is non-increasing on $(0,1]$, it follows that $ue^{\int_u^1 \frac{\theta(w)}{w}dw} \to 0^+$ as $u \to 0$. Taking the logarithm yields $\int_u^1 \frac{\theta(w)-1}{w}dw \to -\infty$ and thus, Assumption (i).

▷ If b and σ are Lipschitz continuous, Kunita's Stochastic flow theorem (see [Kun90], Section 4.5) ensures in particular that, if $x_1 \neq x_2$, the solutions $X_t^{x_1}$ and $X_t^{x_2}$ *a.s.* never get stuck. Taking advantage of this remark slightly shortens the proof below.

▷ Tightness criterions of $\Big(\frac{1}{t}\int_0^t P_s(x,dy)ds\Big)_{t\geq 1}$ for every $x \in \mathbb{R}^d$ usually rely on the mean-reversion property of the solutions of (SDE) usually established under various assumptions involving the existence of a so-called *Lyapunov function* V going to infinity at infinity and such that $\mathcal{A}V$ is upper-bounded and $\limsup_{|x|\to+\infty} \mathcal{A}V(x) < 0$ where $\mathcal{A}$ denotes the infinitesimal generator of X^x (so-called *Has'minskii's criterion*). Keep in mind that

$$\mathcal{A}V(x) = (b|\nabla V)(x) + \frac{1}{2}\mathrm{Tr}\Big(\sigma\sigma^*(x)D^2V(x)\Big)$$

where $\mathrm{Tr}(A)$ stands for the trace of the matrix A.

On the other hand, a classical criterion for pathwise asymptotic confluence (*a.s.* at exponential rate, see *e.g.* [BB92], [Lem05] and often referred to as *asymptotic flatness*) is

$$\forall\, x,\, y \in \mathbb{R}^d, \quad (b(x)-b(y)|x-y) + \frac{1}{2}\|\sigma(x)-\sigma(y)\|^2 < -c|x-y|^2, \qquad c > 0, \tag{3.12}$$

and, as a straightforward consequence, uniqueness of the invariant distribution ν of (SDE) (and of $(DSDS)$ as well). Moreover, putting $y = 0$ in the above inequality straightforwardly yields real coefficients $\alpha > 0$, $\beta \geq 0$ such that $\mathcal{A}V \leq \beta - \alpha V$ with $V(x) = |x|^2$. Hence Has'minskii criterion is fulfilled, so it is also an existence criterion for the invariant distribution. In fact, both weak and pathwise assumptions in Theorem 3.2 are much

weaker than (3.12) but some of the properties which hold under (3.12) are still preserved. For instance, since the left-hand side of (3.12) corresponds to the $(S,0)$-confluence function, we deduce from the criterions that if the $(S,0)$-confluence function is (only) negative on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$, uniqueness of the invariant distribution ν of (SDE) (and of ν_Δ for $(DSDS)$) holds and, combined with the tightness of the occupation measure of the semi-group, it becomes a criterion for $a.s.$ pathwise asymptotic confluence.

Proof of Theorem 3.2. Step 1: Exactly like in the beginning of the proof of Theorem 2.1(a), one checks that the set $\mathcal{I}_{DSDS}$ of invariant distributions of $(Q_t)_{t\geq 0}$ is a nonempty, convex and weakly compact subset of $\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$. As a a consequence of the Krein-Millman theorem, $\mathcal{I}_{DSDS}$ has extremal distributions (and is their closed convex hull).

On the other hand, it follows from strong uniqueness theorem for SDE's that the semi-group $(Q_t)_{t\geq 0}$ leaves stable the diagonal $\Delta_{\mathbb{R}^d\times\mathbb{R}^d}=\{(x,x),\, x\in\mathbb{R}^d\}$.

Let $x_1,x_2\in\mathbb{R}^d$, $x_1\neq x_2$. We define the stopping time

$$\tau_{x_1,x_2}:=\inf\left\{t\geq 0\,|\,X_t^{x_1}=X_t^{x_2}\right\}.$$

Still by a strong uniqueness argument it is clear that $\{\tau_{x_1,x_2}>t\}=\{X_t^{x_1}\neq X_t^{x_2}\}$ so that

$$Q_t((x_1,x_2),{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d})=\mathbf{1}_{{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}}(x_1,x_2)\mathbb{P}(\tau_{x_1,x_2}>t)$$

and $Q_t((x_1,x_1),{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d})=0$.

Let $\mu\in\mathcal{I}_{DSDS}$ be an extremal invariant measure. We have, for every $t\geq 0$,

$$\mu({}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d})=\int_{{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}}\mu(dx_1,dx_2)\mathbb{P}(\tau_{x_1,x_2}>t).$$

Letting t go to $+\infty$ yields

$$\mu({}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d})=\int_{{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}}\mu(dx_1,dx_2)\mathbb{P}(\tau_{x_1,x_2}=+\infty)$$

so that, on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$, $\mu(dx_1,dx_2)$-$a.s.$, $\mathbb{P}(\tau_{x_1,x_2}=+\infty)=1$ or equivalently the process (X^{x_1},X^{x_2}) lives in ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$. Consequently, if $\mu({}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d})\in(0,1)$, both conditional measures $\mu^{{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}}$ and $\mu^{\Delta_{\mathbb{R}^d\times\mathbb{R}^d}}$ are invariant distributions for $(SDSD)$ as well. If so,

$$\mu=\mu({}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d})\mu^{{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}}+\mu(\Delta_{\mathbb{R}^d\times\mathbb{R}^d})\mu^{\Delta_{\mathbb{R}^d\times\mathbb{R}^d}}$$

cannot be extremal. Consequently $\mu(\Delta_{\mathbb{R}^d\times\mathbb{R}^d})=0$ or 1.

Step 2: Let μ be an extremal distribution in $\mathcal{I}_{DSDS}$ and assume that $\mu({}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d})=1$ so that $\mu\in C_\nu^\star$. We will prove that this yields a contradiction under Assumptions (i) and (ii).

Note that f_θ' and g_θ defined in (3.9) are positive on $(0,+\infty)$, that Assumption (i) reads $\limsup_{u\to 0^+}g_\theta(u)<+\infty$ and that $g_\theta'(u)=f_\theta'(u)(1-\theta(u))$. Moreover, if Assumption (ii) is fulfilled, so is the case for any continuous function $\widetilde{\theta}$ satisfying $\widetilde{\theta}\geq\theta$. As a consequence, we may modify θ on $[1,+\infty)$ so that θ still satisfies (ii) and $\theta\geq 1$ over $[2\varepsilon,+\infty)$. Then the function g_θ is nonincreasing on $[2,+\infty)$. Consequently, without loss of generality, we may assume in the sequel of the proof that

$$\sup_{u>0}g_\theta(u)<+\infty. \tag{3.13}$$

We now define a (Lyapunov) function $\varphi : {}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d} \to \mathbb{R}$ by

$$\varphi(y_1, y_2) := f_\theta(|y_1 - y_2|_S^2).$$

We know from Step 1 that $\mu(dx_1, dx_2)$-$a.s.$, ($X_t^{x_1} \neq X_t^{x_2}$ for every $t \geq 0$) $a.s.$. Then, f_θ being a $\mathcal{C}^2$-function, we derive from Itô's formula applied to $\varphi(X_t^{x_1}, X_t^{x_2})$ that $\mu(dx_1, dx_2)$-$a.s.$,

$$\varphi(X_t^{x_1}, X_t^{x_2}) = \varphi(x_1, x_2) + \int_0^t \mathcal{A}^{(2)}\varphi(X_s^{x_1}, X_s^{x_2})ds + \underbrace{\int_0^t f'_\theta(|X_s^{x_1} - X_s^{x_2}|_S^2)\big((\sigma^*(X_s^{x_1}) - \sigma^*(X_s^{x_2}))S(X_s^{x_1} - X_s^{x_2})|dW_s\big)}_{=:M_t \text{ local martingale}}$$

where, for every $(x_1, x_2) \in (\mathbb{R}^d)^2$,

$$\mathcal{A}^{(2)}\varphi(x_1, x_2) = 2\Big((b(x_1) - b(x_2)|x_1 - x_2)_S + \frac{1}{2}\|\sigma(x_1) - \sigma(x_2)\|_S^2\Big)f'_\theta(|x_1 - x_2|_S^2) + 2f''_\theta(|x_1 - x_2|_S^2)\big|(\sigma^*(x_1) - \sigma^*(x_2))S(x_1 - x_2)\big|^2. \quad (3.14)$$

Using that f_θ is increasing and satisfies the $ODE \equiv \theta(\xi)f'_\theta(\xi) + \xi f''_\theta(\xi) = 0$, $\xi \in (0, +\infty)$, we deduce that

$$\mathcal{A}^{(2)}\varphi(x_1, x_2) = 2f'_\theta(|x_1 - x_2|_S^2)\Psi_{\theta,S}(x_1, x_2)$$

so that $\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathcal{A}^{(2)}\varphi(x_1, x_2)\mu(dx_1, dx_2) < 0$ by Assumption (ii).

On the one hand, since μ is extremal and since $\mathcal{A}^{(2)}\varphi$ is bounded from above (see Remark 3.3), we can apply Birkhoff's theorem and obtain:

$$\mu(dx_1, dx_2)\text{-}a.s., \quad \frac{1}{t}\int_0^t \mathcal{A}^{(2)}\varphi(X_s^{x_1}, X_s^{x_2})ds \xrightarrow{t\to+\infty} \int_{{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}} \mathcal{A}^{(2)}\varphi\, d\mu \in [-\infty, 0) \quad a.s. \quad (3.15)$$

On the other hand, using that g_θ is bounded and σ is Lipschitz continuous, it follows that $(M_t)_{t\geq 0}$ is an L^2-martingale such that

$$\langle M\rangle_t = \int_0^t g_\theta(|X_s^{x_1} - X_s^{x_2}|_S^2)^2\Big|\frac{(\sigma^*(X_s^{x_1}) - \sigma^*(X_s^{x_2}))}{|X_s^{x_1} - X_s^{x_2}|_S}\frac{S(X_s^{x_1} - X_s^{x_2})}{|X_s^{x_1} - X_s^{x_2}|_S}\Big|^2 ds \leq Ct \quad \mathbb{P}_\mu\text{-}a.s. \quad (3.16)$$

where C is a deterministic positive constant so that $\frac{M_t}{t} \to 0$ $\mathbb{P}_\mu$-$a.s.$.

As a consequence, $\mu(dx_1, dx_2)$-$a.s.$,

$$\lim_{t\to+\infty} \frac{\varphi(X_t^{x_1}, X_t^{x_2})}{t} = \int_{{}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}} \mathcal{A}^{(2)}\varphi d\mu < 0 \quad a.s..$$

Hence, $a.s.$, $f(|X_t^{x_1} - X_t^{x_2}|_S^2) = \varphi(X_t^{x_1}, X_t^{x_2}) \xrightarrow{t\to+\infty} -\infty \quad a.s.$
If $\lim_{u\to 0^+} f_\theta(u) > -\infty$, this yields a contradiction since f_θ is increasing on $\mathbb{R}_+^*$. Otherwise $|X_t^{x_1} - X_t^{x_2}|_S^2 \xrightarrow{t\to+\infty} 0$. But applying again Birkhoff's theorem, we obtain $\mu(dx_1, dx_2)$-$a.s.$,

$$\int |y_1 - y_2|_S^2\mu(dy_1, dy_2) = \lim_{t\to+\infty} \frac{1}{t}\int_0^t |X_t^{x_1} - X_t^{x_2}|_S^2 ds = 0 \quad a.s.,$$

which contradicts the assumption $\mu({}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}) = 1$. Consequently, for any extremal invariant distribution μ, we have $\mu(\Delta_{\mathbb{R}^d\times\mathbb{R}^d}) = 1$.

We can now prove Claim (a): by Krein-Millman's Theorem $\mathcal{I}_{SDS}$ is the weak closure of the convex hull of its extremal distributions. Consequently, the diagonal $\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ being a closed subset of $\mathbb{R}^d\times\mathbb{R}^d$, all invariant distributions of the duplicated system are supported by this diagonal. For any such invariant distribution μ, both its marginals are invariant distributions for (SDE). If (SDE) had two distinct invariant distributions ν and ν', we know from the introduction that $\mathcal{I}_{DSDS}$ would contain at least a distribution μ for which the two marginals distributions are $\mu(.\times\mathbb{R}^d) = \nu$ and $\mu(\mathbb{R}^d\times .) = \nu'$ respectively. As a consequence, such a distribution μ could not by supported by the diagonal $\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$. Finally, $\mathcal{I}_{SDE}$ is reduced to a singleton $\{\nu\}$ and $\mathcal{I}_{DSDS} = \{\nu_\Delta\}$.

Step 3 (Claim (b): Proof of (3.11)): Under the additional assumption on θ of (b), we have $\lim_{u\to 0^+} f_\theta(u) > -\infty$ and thus, $\inf_{u>0} f_\theta(u) > -\infty$ since f_θ is increasing. Let x_1, $x_2\in\mathbb{R}^d$. Using again that $\mathcal{A}^{(2)}\varphi < 0$ on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ where $\mathcal{A}^{(2)}\varphi$ is given by (3.14). It follows that $\big(f_\theta(|X_t^{x_1} - X_t^{x_2}|_S^2)\big)_{t\geq 0}$ is a lower-bounded $\mathbb{P}$-supermartingale. Thus, it $a.s.$ converges toward $L_\infty^{x_1,x_2}\in L^1(\mathbb{P})$. Using again that f_θ is increasing, it follows that $|X_t^{x_1} - X_t^{x_2}|_S^2$ $a.s.$ converges toward a finite random variable $\ell_\infty^{x_1,x_2} = f_\theta^{-1}(L_\infty^{x_1,x_2})$.

Now, using that for every $x\in\mathbb{R}^d$, $\Big(\frac{1}{t}\int_0^t P_s(x,dy)ds\Big)_{t\geq 1}$ is tight, we derive that $\Big(\frac{1}{t}\int_0^t Q_s((x_1,x_2),(dy_1,dy_2))ds\Big)_{t\geq 1}$ is tight as well. Then the uniqueness of ν_Δ as an invariant distribution of Q implies that

$$\frac{1}{t}\int_0^t Q_s\big((x_1,x_2),(dy_1,dy_2)\big)ds \overset{(\mathbb{R}^d)}{\Longrightarrow} \nu_\Delta.$$

Now for every bounded continuous function $g:\mathbb{R}^d\to\mathbb{R}$,

$$\frac{1}{t}\int_0^t Q_s(g(|y_1-y_2|_S^2)(x_1,x_2)ds = \frac{1}{t}\int_0^t \mathbb{E}\, g(|X_s^{x_1} - X_s^{x_2}|_S^2)ds \longrightarrow \mathbb{E}\, g(\ell_\infty^{x_1,x_2})$$

so that

$$\mathbb{E}\, g(\ell_\infty^{x_1,x_2}) = \int g(|y_1-y_2|_S^2)\nu_\Delta(dy_1,dy_2) = g(0). \qquad \square$$

In Assumption (ii) of the previous theorem, we see that the function $(x,y)\mapsto|(\sigma^*(x)-\sigma^*(y))S(x-y)|$ plays an important role. In the sequel, we will obtain specific results when this function is not degenerated away from the diagonal. Such type of assumption will be called *strong* or *regular directional S-ellipticity assumption.*

In the following proposition, we first show that when such an assumption is satisfied, claim (b) of the previous theorem still holds without the tightness assumption on $\big(\frac{1}{t}\int_0^t P_s(x,dy)ds\big)_{t\geq 1}$ (although it is not really restrictive in our framework (see the fourth item of Remark 3.3)).

Proposition 3.2. *If the function θ is $(0,1]$-valued and σ satisfies the following strong directional S-ellipticity assumption away from the diagonal*

$$\exists\ \alpha_0 > 0,\ \forall\, x,\ y\in\mathbb{R}^d,\ \big|(\sigma^*(x)-\sigma^*(y))S(x-y)\big| \geq \alpha_0|x-y|^2, \tag{3.17}$$

then the conclusion of Claim (b) in the above proposition remains true without the tightness assumption on $(\frac{1}{t}\int_0^t P_s(x,dy)ds)_{t\geq 1}$.

Proof. First, we recall that under the assumptions of (b), we recall that $(f_\theta(|X_t^{x_1} - X_t^{x_2}|_S^2))_{t\geq 0}$ is a lower-bounded $\mathbb{P}$-supermartingale thus convergent to an integrable random variable and that this implies that $(|X_t^{x_1} - X_t^{x_2}|_S^2)_{t\geq 0}$ is $a.s.$ convergent to a finite random variable $\ell_\infty^{x_1,x_2}$ (since f_θ is increasing). On the other hand, since $-\mathcal{A}^{(2)}\varphi$ is positive and f_θ is lower-bounded, we also have that

$$f_\theta(|X_t^{x_1} - X_t^{x_2}|_S^2) - \int_0^t \mathcal{A}^{(2)}\varphi(X_s^{x_1}, X_s^{x_2})ds = \varphi(x_1, x_2) + M_t$$

is a lower bounded $\mathbb{P}$-(local) martingale starting at a deterministic starting value, hence converging toward an integrable random variable. Owing to the computations of (3.16) (which hold for every starting points x_1, x_2), $(M_t)_{t\geq 0}$ is in fact an L^2- convergent martingale. Thus, $\langle M\rangle_\infty < +\infty$ and taking advantage of the expression of this bracket (see (3.16)) and to Assumption (3.17), we derive that for every $\varepsilon > 0$

$$\int_0^{+\infty} \big(g_\theta\big(|X_s^{x_1} - X_s^{x_2}|_S^2\big)^2 \mathbf{1}_{\{|X_s^{x_1} - X_s^{x_2}|_S^2 \geq \varepsilon\}} ds < +\infty \quad a.s.$$

The function g_θ is positive on $(0, +\infty)$ and non-decreasing since $g_\theta'(u) = f_\theta'(u)(1-\theta(u)) \geq 0$. This implies that, for every $\varepsilon > 0$,

$$\liminf_{t\to+\infty} g_\theta\big(|X_t^{x_1} - X_t^{x_2}|_S^2\big)\mathbf{1}_{\{|X^{x_1} - X^{x_2}|_S^2 \geq \varepsilon\}} = 0 \quad a.s.$$

Combined with the convergence of the squared norm this yields

$$\forall\, \varepsilon > 0, \quad g_\theta\big(\ell_\infty^{x_1,x_2}\big)\mathbf{1}_{\{\ell_\infty^{x_1,x_2} \geq \varepsilon\}} = 0 \quad a.s.$$

which finally implies $\ell_\infty^{x_1,x_2} = 0$ $a.s.$ □

3.3 Global criterions, NILS exponent

In this section and the following, we derive several corollaries of Theorem 3.2 illustrated by different examples.

Proposition 3.3. *Let $S \in \mathcal{S}^{++}(d, \mathbb{R})$. Assume $[b]_{S,+} < +\infty$, σ is Lipschitz continuous and $\mathcal{I}_{SDS}$ is non empty and weakly compact. Then,*

(a) If Assumption (ii) of Theorem 3.2 holds with some continuous functions $\theta : (0, +\infty) \to \mathbb{R}_+$ satisfying: there exists $\varepsilon_0 > 0$ such that $\theta(u) \leq 1$, $u \in (0, \varepsilon_0]$, then (SDE) (1.1) and its duplicated system have ν and ν_Δ as unique invariant distributions respectively.

(b) If for every $x \in \mathbb{R}^d$, $(\frac{1}{t}\int_0^t P_s(x, dy)ds)_{t\geq 1}$ is tight and if $\Psi_{\theta,S} < 0$ on $\Delta^c_{\mathbb{R}^d\times\mathbb{R}^d}$ with a continuous function $\theta : (0, +\infty) \to \mathbb{R}_+$ satisfying: there exists $\kappa > 1$ and $\varepsilon_0 \in (0, e^{-\frac{\kappa}{2}})$ such that

$$\forall\, u \in (0, \varepsilon_0], \quad \theta(u) \leq \Big(1 + \frac{\kappa}{\log u}\Big), \tag{3.18}$$

then the duplicated system of (SDE) is pathwise confluent in the sense of Theorem 3.2(b). This condition is in particular satisfied if there exists $\varepsilon_0 > 0$ and $\theta_0 \in (0, 1)$ such that

$$\forall\, u \in (0, \varepsilon_0], \quad \theta(u) \leq \theta_0.$$

Proof. Claim (a) is obvious. As for (b), one checks that $\int_0^1 e^{\int_v^1 \frac{\theta(w)}{w}dw} dv < +\infty$ as soon as $\liminf_{u\to 0^+} \log(u)\big(\theta(u) - 1\big) > 1$ and the result follows. □

Remark 3.4. The simplest case where the preceding result holds is obtained when $\theta \equiv 0$. In this case, the weak confluence referred to as $(S,0)$-confluence in what follows, condition reads:

$$\forall m \in C_\nu^\star, \quad \int_{\mathbb{R}^d\times\mathbb{R}^d} (b(x)-b(y)|x-y)_S + \frac{1}{2}\|\sigma(x)-\sigma(y)\|_S^2 m(dx,dy) < 0 \tag{3.19}$$

and claim (b) holds as soon as the integrated function is negative on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$.

At this stage, it is important for practical applications to note that *the constant function $\theta \equiv 1$ satisfies the assumption in (a) of the above Proposition.* This leads us to introduce an important quantity of interest for our purpose.

Definition 3.1. The *non-infinitesimal S-Lyapunov (NILS) exponent* is a function on $\mathbb{R}^d \times \mathbb{R}^d \setminus \Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ defined for every $x,\, y \in \mathbb{R}^d,\ x \neq y$, by

$$\Lambda_S(x,y) = \frac{(b(x)-b(y)|x-y)_S}{|x-y|_S^2} + \frac{1}{2}\frac{\|\sigma(x)-\sigma(y)\|_S^2}{|x-y|_S^2} - \left(\left|\frac{(\sigma^*(x)-\sigma^*(y))S(x-y)}{|x-y|_S^2}\right|^2\right).$$

Corollary 3.2. *Assume b and σ are like in Proposition 3.3 and $\mathcal{I}_{SDE}$ is non empty and weakly compact.*

(a) Negative Integrated NILS exponent*: if*

$$\forall m \in C_\nu^\star, \qquad \int \Lambda_S(x,y)m(dx,dy) < 0,$$

then (SDE) and its duplicated system have ν and ν_Δ as unique invariant distributions respectively.

(b) Negative NILS exponent bounded away from 0*: If furthermore $(\frac{1}{t}\int_0^t P_s(x,dy)ds)_{t\geq 1}$ is tight for every $x \in \mathbb{R}^d$ or σ satisfies (3.17) and if there exists $c_0 > 0$ such that*

$$\forall\, x,\, y \in \mathbb{R}^d,\ x \neq y,\ |x-y|_S^2 \leq \varepsilon_0 \Longrightarrow \Lambda_S(x,y) \leq -c_0 \tag{3.20}$$

then the duplicated diffusion is pathwise confluent i.e.

$$\forall\, x_1, x_2 \in \mathbb{R}^d, \quad X_t^{x_1} - X_t^{x_2} \longrightarrow 0 \ a.s. \ \ as\ t \to +\infty.$$

Proof. (a) follows from claim (a) in the above proposition with $\theta \equiv 1$ since $uf_\theta(u) \equiv 1$ on $(0,+\infty)$ so that $\mathcal{A}^{(2)}\varphi(x,y) = 2\Lambda_S(x,y)$ in the proof of Theorem 3.2. (b) follows from claim (b) in the same proposition. □

Remark 3.5. Obviously, the integrated NILS exponent is negative for every $m \in C_\nu^\star$ if the NILS exponent is negative on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$. This second (pointwise) condition is usually the only tractable one for applications. However, we will see in the next subsections that we can build some criterions where Λ_S is not negative everywhere.
Also note that in (b), Condition (3.20) can be replaced by the sharper: for all $x, y \in \mathbb{R}^d$, $\Lambda_S(x,y) < 0$ and there exists $\kappa > 1$ and $\varepsilon_0 \in (0, e^{-\frac{\kappa}{2}})$ such that for all $x, y \in \mathbb{R}^d$ such that $|x-y|_S \leq \varepsilon_0$, $\Lambda_S(x,y) \leq \dfrac{\kappa}{\log(|x-y|_S^2)}$.

When the coefficients are smooth enough, the negativity of Λ_S can be ensured by the following criterion:

Corollary 3.3 (Smooth coefficients)**.** *Assume the functions b and σ are* continuously *differentiable. Let $J_b(x) = \left[\frac{\partial b_i}{\partial x_j}(x)\right]_{1\le i,j\le d}$ denote the Jacobian of b at x and let $\nabla\sigma(x) = \left[\frac{\partial \sigma_{ij}}{\partial x^k}(x)\right]_{i,j,k}$ denote the gradient of σ at x. If both $SJ_b + J_b^*S$ and $\nabla\sigma$ are* Lipschitz continuous *and if $\nabla\sigma$ is bounded then $\Lambda_S(x,y) \le -c_0$ on ${}^c\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ if*

$$\sup_{x\in\mathbb{R}^d}\sup_{|u|_S=1}\left(u^*\left(SJ_b(x) + J_b^*(x)S\right)u + \left\|(\nabla\sigma(x)|u)\right\|_S^2 - 2\left|(\nabla\sigma^*(x)Su|u)\right|^2\right) < 0$$

where, for every $v = (v^1,\dots,v^d)\in\mathbb{R}^d$, $(\nabla\sigma(x)|v) = \left[(\nabla\sigma_{ij}(x)|v)\right]_{1\le i,j\le d}$ and $\nabla\sigma(x)v = \left[\sum_{k=1}^d \frac{\partial\sigma_{ij}}{\partial x_k}(x)v^k\right]_{1\le i,j\le d}$. When $S = I_d$, this may also be written

$$\sup_{x\in\mathbb{R}^d}\sup_{|u|=1}\left(u^*\Big((J_b + J_b^*)(x) + \sum_{i,j}(\nabla\sigma_{ij}(x))^{\otimes 2} - 2\Big[(\nabla\sigma_{ij}(x)|u)\Big]\Big[(\nabla\sigma_{ij}^*(x)|u)\Big]\Big)u\right) < 0.$$

The proof is an easy consequence of the Taylor formula with Laplace (integral) remainder. We leave the computational details to the reader.

3.4 Applications and extensions

3.4.1 Localization around the diagonal

By local we mean that the confluence condition will be effective only in the neighbourhood of the diagonal $\Delta_{\mathbb{R}^d\times\mathbb{R}^d}$. The price to pay is a regular directional ellipticity assumption on $\sigma(x) - \sigma(y)$ in the direction $S(x-y)$ away from the diagonal.

Proposition 3.4. *Assume $[b]_{S,+} < +\infty$, σ is Lipschitz continuous and (SDE) admits at least one invariant distribution ν. If there exists $\varepsilon_0 > 0$ such that*

$$\begin{cases} (i)\ \textit{Directional } S\textit{-ellipticity:}\ \ \eta_0 := \inf\Big\{\big|(\sigma^*(x) - \sigma^*(y))S(x-y)\big|,\ |x-y|_S \ge \varepsilon_0\Big\} > 0, \\ (ii)\ \textit{Locally negative INILS exponent: } \forall\, m\in C_\nu^\star,\ \displaystyle\int_{|x-y|_S\le\varepsilon_0}\Lambda_S(x,y)m(dx,dy) < 0, \end{cases}$$

then (SDE) (1.1) and its duplicated system still have ν and ν_Δ as unique invariant distributions respectively.

Proof. Owing to (i), we have for every $u\in(\varepsilon_0,+\infty)$:

$$\sup_{|x-y|_S=u}\frac{(b(x)-b(y)|x-y)_S + \frac{1}{2}\|\sigma(x)-\sigma(y)\|_S^2}{\left|(\sigma^*(x)-\sigma^*(y))S\frac{x-y}{|x-y|_S}\right|^2} \le \big([b]_+ + \frac{1}{2}[\sigma]_{\mathrm{Lip}}^2\big)\frac{u^4}{\eta_0^2}.$$

For every $\varepsilon_0' > \varepsilon_0$, let $\theta_{\varepsilon_0'} : (0,+\infty)\to\mathbb{R}_+$ denote the continuous function defined by

$$\theta_{\varepsilon_0'}(u) = \begin{cases} \big([b]_{S,+} + \frac{1}{2}[\sigma]_{\mathrm{Lip}}^2\big)\frac{u^4}{\eta_0^2} & \text{if } u\in(\varepsilon_0',\infty), \\ 1 & \text{if } u\in(0,\varepsilon_0], \\ 1 + \theta_{\varepsilon_0'}(\varepsilon_0')\frac{t-\varepsilon_0}{\varepsilon_0'-\varepsilon_0} & \text{if } u\in(\varepsilon_0,\varepsilon_0'). \end{cases}$$

Since $\theta_{\varepsilon_0'}(u) = 1$ in the neighbourhood of 0, Assumption (i) of Theorem 3.2 is satisfied. For Assumption (ii), one first deduces from the construction and to the first assumption that,

$$\forall\, \varepsilon_0' > \varepsilon_0,\ \forall x, y \text{ such that } |x-y| \geq \varepsilon_0', \quad \Psi_{\theta_{\varepsilon_0'},S}(x,y) \leq 0.$$

Using that $\theta_{\varepsilon_0'}(u) = 1$ on $(0, \varepsilon_0]$, it follows that for all $m \in C_\nu^\star$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f_\theta'(|x-y|_S^2)\Psi_{\theta_{\varepsilon_0'},S}(x,y)m(dx,dy) \leq \int_{|x-y|\leq\varepsilon_0} \Lambda_S(x,y)m(dx,dy) + I_2(\varepsilon_0')$$

where

$$I_2(\varepsilon_0') = \int_{\varepsilon_0<|x-y|<\varepsilon_0'} f_\theta'(|x-y|_S^2)\Psi_{\theta_{\varepsilon_0'},S}(x,y)m(dx,dy).$$

Since the integrated function is bounded from above on $\{(x,y), \varepsilon_0 < |x-y| < \varepsilon_0'\}$, we deduce that

$$I_2(\varepsilon_0') \leq C\, m(\varepsilon_0 < |x-y| < \varepsilon_0') \xrightarrow{\varepsilon_0'\to\varepsilon_0} 0.$$

By the second assumption of this proposition, it follows that there exists $\varepsilon_0' > \varepsilon_0$ such that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f_\theta'(|x-y|_S^2)\Psi_{\theta_{\varepsilon_0'},S}(x,y)m(dx,dy) < 0$$

so that Assumption (ii) of Theorem 3.2 holds. This completes the proof.

□

3.4.2 Local criterion on compact sets

As mentioned in Remark 3.4, Theorem 3.2 can be applied under the $(S,0)$-confluence condition (3.19). This condition is in particular satisfied when

$$\forall\, x,\, y\in\mathbb{R}^d,\ x\neq y, \quad (b(x)-b(y)|x-y)_S + \frac{1}{2}\|\sigma(x)-\sigma(y)\|_S^2 < 0.$$

One asset of this more stringent assumption is that it can be localized in two ways: first in the neighbourhood of the diagonal like in the above local criterions, but also on compacts sets of $\mathbb{R}^d\times\mathbb{R}^d$. This naturally leads to a criterion based on the differentials of b and σ when they exist.

Proposition 3.5 (Criterion on compact sets)**.** *(a) Let $S \in \mathcal{S}^{++}(d,\mathbb{R})$ such that for every $R > 0$, there exists $\delta_R > 0$ such that $\forall\, x, y\in B_{|\cdot|_S}(0;R)$,*

$$0 < |x-y|_S \leq \delta_R \implies (b(x)-b(y)|x-y)_S + \frac{1}{2}\|\sigma(x)-\sigma(y)\|_S^2 < 0. \tag{3.21}$$

Then the diffusion is asymptotically $(S,0)$-confluent.

(b) If b and σ are continuously differentiable, then (3.21) holds as soon as

$$(AC)^{diff} \equiv \forall\, x\in\mathbb{R}^d, \quad SJ_b(x) + J_b^*(x)S + \sqrt{S}\sum_{i,j}(\nabla\sigma_{ij}(x))^{\otimes 2}\sqrt{S} < 0 \ \ in\ \ \mathcal{S}(d,\mathbb{R}).$$

Proof. (a) Let $x, y \in \mathbb{R}^d$ such that $x \neq y$. Set $R = \max(|x|_S, |y|_S)$ and

$$x_0 = x,\ x_i = x + \frac{i}{N}(y-x),\ i = 1, \dots, N-1,\ x_N = y$$

where $|y-x|_S < N\delta_R$. Then for every $i \in \{1, \dots, N\}$, $|x_i|_S \leq R$ and $|x_i - x_{i-1}|_S \leq \delta_R$. Then

$$\begin{aligned}
\|\sigma(x) - \sigma(y)\|_S^2 &= \Big\| \sum_{i=1}^N \sigma(x_i) - \sigma(x_{i-1}) \Big\|_S^2 \leq N \sum_{i=1}^N \|\sigma(x_i) - \sigma(x_{i-1})\|_S^2 \\
&< -2N \sum_{i=1}^N (b(x_i) - b(x_{i-1}) | x_i - x_{i-1})_S = -2 \sum_{i=1}^N \big(b(x_i) - b(x_{i-1}) | y - x\big)_S \\
&< -2(b(y) - b(x) | y - x)_S.
\end{aligned}$$

(b) First, we prove the result when $S = I_d$. We note that, for every continuously differentiable function $g : \mathbb{R}^d \to \mathbb{R}$, $g(y) - g(x) = \int_0^1 \big(\nabla g(x + t(y-x)) | y - x\big) dt = \int_0^1 (y-x)^* \nabla g(x + t(y-x)) dt$ so that

$$\big(b(y) - b(x) | y - x\big) = \int_0^1 (y-x)^* J_b(x + t(y-x)) (y-x) dt = \int_0^1 (y-x)^* J_b^*(x + t(y-x)) (y-x) dt$$

and

$$\|\sigma(y) - \sigma(x)\|^2 = \sum_{i,j=1}^d \left(\int_0^1 \big(\nabla \sigma_{ij}(x + t(y-x)) | y - x\big) dt \right)^2.$$

By Schwarz's Inequality and the fact that $(u|v)^2 = u^* v^{\otimes 2} u$, we deduce

$$\begin{aligned}
\big(b(y) - b(x) | y - x\big) + \frac{1}{2}\|\sigma(y) - \sigma(x)\|^2 \leq \frac{1}{2} \int_0^1 \Big(&(y-x)^* (J_b + J_b^*)(x + t(y-x))(y-x) \\
&+ \frac{1}{2} \sum_{ij} (y-x)^* \big(\nabla \sigma_{ij}(x + t(y-x))\big)^{\otimes 2} (y-x) \Big) dt.
\end{aligned}$$

This completes the proof when $S = I_d$. This extends to general matrix $S \in \mathcal{S}^{++}(d, \mathbb{R})$ using that $\|\sigma(y) - \sigma(x)\|_S^2 = \|(\sqrt{S}\sigma)(y) - (\sqrt{S}\sigma)(x)\|^2$ and the fact that $(Au)^{\otimes 2} = Au^{\otimes 2} A^*$ with $A = \sqrt{S}$. □

3.4.3 The case $\Lambda_S \leq 0$

As mentioned before, the main field of applications of Corollary 3.2 seems to be the case $\Lambda_S < 0$ out of the diagonal $\Delta_{\mathbb{R}^d \times \mathbb{R}^d}$. In the two next sections, our objective is to state some results when this condition is not fulfilled. We begin by a simple application of Corollary 3.2 where the NILS exponent is only non-positive and negative outside of a compact set.

Proposition 3.6. *Assume $[b]_{S,+} < +\infty$, σ is Lipschitz continuous and (SDE) has a unique invariant distribution ν whose support is not compact. Then, uniqueness for ν_Δ holds true as soon as*

$$\forall x, y \in \mathbb{R}^d,\ \Lambda_S(x, y) \leq 0 \ \text{ and } \ \exists\, R > 0 \ s.t. \ \max(|x|_S, |y|_S) > R \Longrightarrow \Lambda_S(x, y) < 0. \tag{3.22}$$

Proof. Since the support of ν is not compact, we have for every $m \in C_\nu^\star$: $m(\max(|x|_S, |y|_S) > R) \geq \nu(|x|_S > R) > 0$. It follows from the assumption that

$$\forall m \in C_\nu^\star, \quad \int \Lambda_S(x,y) m(dx,dy) \leq \int \Lambda_S(x,y) 1_{\{\max(|x|_S,|y|_S)>R\}} m(dx,dy) < 0$$

and we deduce the result from Corollary 3.2. □

Remark 3.6. ▷ In the particular case where σ is constant, Condition (3.22) becomes a monotony condition on b (decrease with respect to $(.|.)_S$ at infinity), namely:

$$\begin{aligned} &\forall x, y \in \mathbb{R}^d,\ x \neq y \quad (b(x) - b(y)|x-y)_S \leq 0, \\ \text{and} \quad &\exists\, R > 0 \text{ s.t. } \max(|x|_S, |y|_S) > R \Longrightarrow (b(x) - b(y)|x-y)_S < 0. \end{aligned}$$

This means that b is S-non-increasing on $\mathbb{R}^d$, S-decreasing outside $B_{|.|_S}(0;R)^2$. For instance, if $b = -\nabla U$, the above assumption holds if U is convex and (only) strictly convex outside of a compact set.

▷ Note that when ∇U is only increasing outside $B_{|.|_S}(0;R)$ but possibly with no specific monotony on $B_{|.|_S}(0;R)$, it is still possible to find some diffusion coefficients σ such that the SDE $dX_t = -\nabla U(X_t)dt + \sigma(X_t)dW_t$ remains weakly or pathwise confluent. We refer to the next subsection for models with such *stochastically stabilizing* diffusive components.

▷ Finally, note that the above condition (3.22) can be also localized around the diagonal under the directional S-ellipticity assumption. To be more precise, when ν is unique and its support is not compact, Proposition 3.6 still holds if Assumption (ii) is "localized" into:

$$(ii)_{loc} \equiv \text{ for every } x, y \in \mathbb{R}^d \text{ such that } 0 < |x-y| \leq \varepsilon_0, \Lambda_S(x,y) \leq 0.$$

3.4.4 Λ_S possibly positive in some areas of $\mathbb{R}^d \times \mathbb{R}^d$

In the continuity of the previous section, we try to explore some multidimensional settings where Λ_S can be positive in some parts of the space. More precisely, we focus here on gradient systems with constant noise whose potential U is not convex in all the space (see [Tea08] for other confluence results on this type of model with the "random attractor" viewpoint). For such dynamical systems, we obtain a criterion below that we next apply to super-quadratic non-convex potentials. Then, we will come back to this problem in section 3.5.3 where we focus on the particular example $U(x) = (|x|^2 - 1)^2$, case for which we are able to obtain a sharper result.

Proposition 3.7. *Let $U : \mathbb{R}^d \to \mathbb{R}_+$ be a locally Lipschitz, differentiable function satisfying $\liminf_{|x|\to+\infty} \frac{U(x)}{|x|^\gamma} > 0$ and $\limsup_{|x|\to+\infty} \frac{U(x)}{|x|^\gamma} < +\infty$ for a positive γ. Then, the Brownian diffusion*

$$dX_t^x = -\nabla U(X_t^x)dt + \sigma dW_t,\ X_0^x = x,\ \sigma > 0, \quad \textit{and } W \textit{ standard Brownian motion on } \mathbb{R}^d$$

satisfies a strong existence-uniqueness property with unique invariant distribution $\nu_\sigma(dx) = C_\sigma e^{-\frac{2U(x)}{\sigma^2}} dx$.

Furthermore assume that its NILS exponent satisfies

$$\forall\, x,\, y \in \mathbb{R}^d, \quad \Lambda_{Id}(x,y) \leq \beta - \frac{\alpha}{2}\big(|x|^a + |y|^a\big) \quad \textit{where } \beta \in \mathbb{R},\ \alpha,\ a > 0$$

then there exists $\sigma_c > 0$ such that, for every $\sigma > \sigma_c$, the related $(DSDS)$ system (2-point motion) is weakly confluent.

Proof. The strong existence-uniqueness is classical background. The form of the invariant distribution ν_σ as well. Then by Fatou's Lemma and the asymptotic upper-bound, there exists $A > 0$ such that

$$\liminf_{\sigma \to +\infty} \int_{\mathbb{R}^d} |u|^a e^{-\frac{2U(\sigma^{2/\gamma} u)}{\sigma^2}} du \geq \int_{\mathbb{R}^d} |u|^a e^{-A|u|^\gamma} du > 0.$$

On the other hand, note that

$$\int_{\mathbb{R}^d} |x|^a e^{-\frac{2U(x)}{\sigma^2}} dx = \sigma^{\frac{d}{2}+\frac{2a}{\gamma}} \int_{\mathbb{R}^d} |u|^a e^{-\frac{2U(\sigma^{2/\gamma} u)}{\sigma^2}} du.$$

Owing to the asymptotic lower bound for U at infinity and the (reverse) Fatou's Lemma, there exists a real number $B > 0$ such that

$$\limsup_{\sigma \to +\infty} \int_{\mathbb{R}^d} e^{-\frac{2U(\sigma^{2/\gamma} u)}{\sigma^2}} du \leq \int_{\mathbb{R}^d} e^{-B|u|^\gamma} du < +\infty.$$

As a consequence $\liminf_{\sigma \to +\infty} \nu_\sigma\big(|x|^a\big) = +\infty$. For any distribution $m \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginal ν_σ and assigning no weight to the diagonal, one has

$$\int_{\mathbb{R}^d \times \mathbb{R}^d \setminus \Delta_{\mathbb{R}^d \times \mathbb{R}^d}} \Lambda_{Id}(x,y) m(dx,dy) \leq \beta - \alpha \nu_\sigma\big(|x|^a\big) < 0$$

as soon as σ is large enough to ensure that $\nu_\sigma\big(|x|^a\big) \geq \frac{\beta}{\alpha}$. □

Remark 3.7. We may assume without loss of generality that $\mathrm{argmin}_{\mathbb{R}^d} U = \{U = 0\} \subset \{\nabla U = 0\}$ so that $\nu_\sigma \overset{\mathbb{R}^d}{\Longrightarrow} \nu_0 = \mathrm{Unif}(\{U = 0\})$ as $\sigma \to 0$. Hence, from a practical point of view, the fact that the critical σ_c can be taken as 0 seems a reasonable conjecture if $\beta - \alpha \nu_0\big(|x|^a \mathbf{1}_{\{U(x)=0\}}\big) \leq 0$. Thus, in Section 3.5.3, we prove that such conjecture is valid when $U(x) = (|x|^2 - 1)^2$.

Corollary 3.4. *Assume that $U : \mathbb{R}^d \to \mathbb{R}_+$ is defined by $U(x) = C|x|^{2p} + \varepsilon(x)$ where $p > 1$, $C > 0$ and ε is a $\mathcal{C}^1$-function such that $\nabla\varepsilon$ is Lipschitz continuous. Then, there exists $\sigma_c > 0$ such that, for every $\sigma > \sigma_c$, the $(DSDS)$ system related to $dX_t = -\nabla U(X_t)dt + \sigma dB_t$ is weakly confluent.*

Proof. For every $x \neq y$,

$$\frac{(\nabla(|x|^{2p}) - \nabla(|y|^{2p})|x-y)}{|x-y|^2} \geq 2p \int_0^1 |x + t(y-x)|^{2(p-1)} dt.$$

If $p \geq 2$, we deduce from Jensen's inequality that

$$\begin{aligned}\int_0^1 |x + t(y-x)|^{2(p-1)} dt \geq \left(\int_0^1 |x + t(y-x)|^2 dt\right)^{p-1} &\geq \left(\frac{1}{6}(|x|^2 + |y|^2)\right)^{p-1} \\ &\geq \left(\frac{1}{6}\right)^{p-1} (|x|^{2(p-1)} + |y|^{2(p-1)})\end{aligned}$$

where in the last inequality, we used again that $p-1 \geq 1$. It follows that

$$\Lambda_{Id}(x,y) \leq [\varepsilon]_1 - \alpha(|x|^{2(p-1)} + |y|^{2(p-1)})$$

where $[\varepsilon]_1$ denotes the Lipschitz constant of ε and $\alpha > 0$. The previous result then applies in this case.

When $p \in (1,2)$, we deduce from the elementary inequality $||u|^\rho - |v|^\rho| \leq |u-v|^\rho$ for $\rho < 1$ that

$$\int_0^1 |x + t(y-x)|^{2(p-1)} dt \geq \alpha(|x|^{2(p-1)} + |y|^{2(p-1)})$$

with $\alpha > 0$ and the result again follows. □

3.5 Examples

3.5.1 An example of confluent diffusion with increasing drift

Assume that $\sigma : \mathbb{R}^d \to \mathcal{M}(d,d,\mathbb{R})$ is defined by $\sigma(x) = x \otimes \lambda + \sigma^0$ where $\sigma^0 \in \mathcal{M}(d,d,\mathbb{R})$ and $\lambda : \mathbb{R}^d \to \mathbb{R}^d$ is a bounded Lipschitz function (such that σ is Lipschitz too). If there exists $\rho \in (0, \frac{1}{2})$ such that

$$\limsup_{|x|\to+\infty} \frac{(b(x)|x) - \rho|x|^2|\lambda(x)|^2}{(1+|x|^2)^{\rho+\frac{1}{2}}} = -\infty \tag{3.23}$$

then the diffusion (1.1) has at least one invariant distribution ν. Thus, if

$$[b]_+^0 = \sup_{x\neq 0} \frac{(b(x) - b(0)|x)}{|x|^2} \neq +\infty,$$

the above condition is satisfied as soon as

$$\liminf_{|x|\to+\infty} |\lambda(x)|^2 > 2[b]_+^0.$$

The key is to introduce the Lyapunov function $V(x) = (a+|x|^2)^{\rho+\frac{1}{2}}$. Using that $\|(\sigma - \sigma^0)(x)\|^2 = |(\sigma-\sigma^0)^*(x)\frac{x}{|x|}|^2 = |\lambda(x)|^2|x|^2$, we deduce that

$$\frac{1}{2}\|\sigma(x)\|^2 - (\rho + \frac{1}{2})\Big|\sigma^*(x)\frac{x}{|x|}\Big|^2 = -\rho|\lambda(x)|^2|x|^2 + O(1)$$

and it follows that $\limsup_{|x|\to+\infty} \mathcal{A}V(x) = -\infty$ if (3.23) is fulfilled (where $\mathcal{A}$ denotes the infinitesimal generator of (1.1)).

If the function λ *is constant*, the diffusion is asymptotically pathwise confluent (so that ν is unique for (1.1) and the duplicated system has ν_Δ as unique invariant distribution) as soon as there exists $\varepsilon_0 > 0$ satisfying

$$|x-y| \leq \varepsilon_0 \Longrightarrow (b(x) - b(y)|x-y) - \frac{1}{2}|\lambda|^2|x-y|^2 < 0. \tag{3.24}$$

This is a consequence of Proposition 3.4 applied with $S = I_d$ (the directional ellipticity assumption (i) is clearly true since $|(\sigma^*(x) - \sigma^*(y))(x-y)| = |\lambda|.|x-y|^2$). If b is smooth this condition is satisfied as soon as, for every $x \in \mathbb{R}^d$, $\frac{1}{2}(J_b + J_b^*)(x) < \frac{1}{2}|\lambda|^2 I_d$ in $\mathcal{S}(d,\mathbb{R})$ ($J_b(x)$ denotes the Jacobian matrix of b).

3.5.2 Baxendale's model

Let $\Xi_t = (X_t, Y_t)$ be the unique strong solution to the 2-dimensional SDE

$$\begin{aligned} dX_t &= \big(a - \frac{\sigma^2}{2}\big)X_t dt - (\sigma Y_t - \theta_X)dW_t \\ dY_t &= \big(b - \frac{\sigma^2}{2}\big)Y_t dt + (\sigma X_t + \theta_Y)dW_t \end{aligned}$$

where W is scalar standard Brownian motion, a, b, σ are real numbers satisfying

$$ab < 0,\ a + b < 0,\ \sigma > \sqrt{\frac{2ab}{a+b}}.$$

and θ_X, $\theta_Y \in \mathbb{R}$. When $\theta_X = \theta_Y = 0$, this system is known as Baxendale's system (see *e.g.* [KP92]). Its stochastic stability has been extensively investigated in connection with its Lyapunov exponent. Then set

$$\lambda = \lambda(\sigma) = \frac{b - a + \sqrt{(b-a)^2 + \sigma^4}}{\sigma^2} \in (0,1) \ \text{ and } \ \alpha = \sigma^2 - (a+b) - \sqrt{(a-b)^2 + \sigma^4} > 0.$$

and $|\,.\,|_\lambda = |\,.\,|_S$ with $S = \mathrm{Diag}(1, \lambda)$. Itô's Lemma implies

$$d|\Xi_t|_\lambda^2 = \Big(-\alpha|\Xi_t|_\lambda^2 + \theta_X(\theta_X - 2\sigma Y_t) + \lambda\theta_Y(\theta_Y + 2\sigma X_t)\Big)dt + \Theta(\Xi_t)dW_t$$

where $\Theta(x,y) = 2\big((\lambda - 1)\sigma xy + \lambda\theta_Y x + \theta_X y\big)$. It is clear that there exists $\beta \in \mathbb{R}_+$ such that

$$|\theta_X(\theta_X - 2\sigma y) + \lambda\theta_Y(\theta_Y + 2\sigma x)| \le \beta|(x,y)|_\lambda.$$

Then using that $|\xi|_\lambda \le \frac{1}{2\alpha} + \frac{\alpha}{2}|\xi|_\lambda^2$ and setting $\beta' = \beta + \frac{1}{2\alpha}$, we derive that

$$d|\Xi_t|_\lambda^2 \le \beta' - \frac{\alpha}{2}|\Xi_t|_\lambda^2 dt + \Theta(\Xi_t)dW_t$$

where $\theta(\xi) \le C|\xi|_\lambda$. Hence, the function $V(\xi) = |\xi|_\lambda^2$ is a Lyapunov function for the system since $AV \le \beta' - \frac{\alpha}{2}V$. As a consequence there exists at least one invariant distribution ν for the system and any such distribution satisfies $\nu(V) \le 2\frac{\beta'}{\alpha}$.

At this stage we can compute the non-infinitesimal S-Lyapunov exponent of the duplicated system. Tedious although elementary computations show that, for every $\xi = (x,y)$, $\xi' = (x', y') \in \mathbb{R}^2$,

$$\Lambda_S(\xi, \xi') = -\frac{\alpha}{2} - (\lambda - 1)^2\sigma^2\frac{(x-x')^2(y-y')^2}{|\xi - \xi'|_\lambda^4} < 0.$$

Remark 3.8. Adapting results from [Bax91] obtained for diffusions on compact manifolds, one easily derive another type of criterion for weak confluence. Namely, if the diffusion $(X_t^x)_{t\ge 0}$ (1-point motion) has a unique invariant distribution ν with support $\mathbb{R}^d$ and if $X_t^x \xrightarrow{\mathcal{L}} \nu$ as $t \to +\infty$ for every $x \in \mathbb{R}^d$ and if *no nonempty closed connected subset C of $\mathbb{R}^d \times \mathbb{R}^d \setminus \Delta_{\mathbb{R}^d\times\mathbb{R}^d}$ is left stable by the 2-point motion $(X_t^{x_1}, X_t^{x_2})_{t\ge 0}$* ($(x_1, x_2) \in C$), then the 2-point motion is weakly confluent with invariant distribution $\nu_\Delta = \nu \circ (x \mapsto (x,x))^{-1}$. However, although more intuitive this criterion seems not to be tractable compared to the above criterions based on the $NILS$ exponent.

3.5.3 Example of gradient system with non convex potential

Let $U : \mathbb{R}^d \to \mathbb{R}_+$ be defined by $U(x) = \frac{(|x|^2-1)^2}{4}$, $x \in \mathbb{R}^d$. Applying Corollary 3.4 with $p = 2$ and $\varepsilon(x) = \frac{1}{4}(1 - 2x^2)$, one deduces that there exists $\sigma_c > 0$ such that for every $\sigma > \sigma_c$, the 2-point motion related to $dX_t^x = -\nabla U(X_t^x)dt + \sigma dW_t$ is weakly confluent. In fact, for this function, we obtain the weak confluence for every $\sigma > 0$.

Proposition 3.8. *Let $U : \mathbb{R}^d \to \mathbb{R}_+$ be defined by $U(x) = \frac{(|x|^2-1)^2}{4}$. Then, for every $\sigma > 0$, the (DSDS) related to the Brownian diffusion $dX_t^x = -\nabla U(X_t^x)dt + \sigma dW_t$ is weakly confluent.*

Proof. Elementary computations show that, for every $x, y \in \mathbb{R}^d$,

$$\Lambda_{Id}(x,y) = 1 - \frac{1}{2}\Big(\big(|x|^2 + |y|^2\big) + \frac{(x+y|x-y)^2}{|x-y|^2}\Big) \leq 1 - \frac{1}{2}\big(|x|^2 + |y|^2\big)$$

so that for every $m \in C_\nu^\star$,

$$\int \Lambda_{Id}(x,y)m(dx,dy) < 1 - \frac{1}{2}\left(\int |x|^2\nu_\sigma(dx) + \int |y|^2\nu_\sigma(dy)\right) = \frac{1}{Z_\sigma}\int (1-|x|^2)e^{-\frac{2U(x)}{\sigma^2}}dx$$

with $Z_\sigma = \int_{\mathbb{R}^d} e^{-\frac{2U(x)}{\sigma^2}}dx$. By Corollary 3.2$(a)$, it is now enough to prove that

$$\int_{\mathbb{R}^d} (1-|x|^2)e^{-\frac{2U(x)}{\sigma^2}}(dx) < 0.$$

Thanks to a change of variable,

$$\int_{\mathbb{R}^d} (1-|x|^2)e^{-\frac{2U(x)}{\sigma^2}}dx = \mathrm{Vol}(\mathcal{S}_{d-1})\int_0^{+\infty}(1-r^2)r^{d-1}e^{-\frac{(r^2-1)^2}{2\sigma^2}}dr$$

where $\mathrm{Vol}(\mathcal{S}_{d-1})$ denotes the hyper-volume of the $d-1$-dimensional Euclidean ball. When $d = 2$, it follows that

$$\int_0^{+\infty}(1-r^2)re^{-\frac{(r^2-1)^2}{2\sigma^2}}dr = \frac{\sigma^2}{2}\left[e^{-\frac{(r^2-1)^2}{2\sigma^2}}\right]_0^{+\infty} = \frac{\sigma^2 e^{-\frac{1}{\sigma^2}}}{2} < 0.$$

When $d > 2$, note that $(1-r^2)r^{d-1} \leq (1-r^2)r$ for every $r \in [0,+\infty)$ so that

$$\int_0^{+\infty}(1-r^2)r^{d-1}e^{-\frac{(r^2-1)^2}{2\sigma^2}}dr = \int_0^{+\infty}(1-r^2)re^{-\frac{(r^2-1)^2}{2\sigma^2}}dr < 0.$$

This completes the proof. □

4 Application to the Richardson-Romberg extrapolation for the approximation of invariant distributions

As an application, we investigate in this section the Richardson-Romberg (RR) extrapolation for the approximation of invariant measures. Roughly speaking, the aim of a RR method is generally to improve the order of convergence of an algorithm based on an discretization scheme by cancelling the first order error term induced by the time discretization of the underlying process. However, to be efficient, such a method must be implemented with a control of its variance. We will see that in this context, this control is strongly linked to the uniqueness of the invariant distribution of the duplicated diffusion.

4.1 Setting and Background

4.1.1 Recursive computation of the invariant distribution of a diffusion: the original procedure

Following [LP02] and a series of papers cited in the introduction, we consider here a sequence of empirical measures $(\nu_n^\eta(\omega, dx))_{n\geq 1}$ built as follows: let $(\gamma_n)_{n\geq 1}$ denote a non-increasing sequence of positive *step parameters* satisfying

$$\gamma_n \xrightarrow{n\to+\infty} 0 \quad \text{and} \quad \Gamma_n = \sum_{k=1}^n \gamma_k \xrightarrow{n\to+\infty} +\infty.$$

We denote by $(\bar{X}_n)_{n\geq 0}$ the Euler scheme with step sequence $(\gamma_n)_{n\geq 1}$ defined by $\bar{X}_0 = x \in \mathbb{R}^d$ and

$$\bar{X}_{n+1} = \bar{X}_n + \gamma_{n+1} b(\bar{X}_n) + \sqrt{\gamma_{n+1}}\sigma(\bar{X}_n)U_{n+1}$$

where $(U_n)_{n\geq 1}$ is a sequence of i.i.d. centered $\mathbb{R}^q$-valued random vectors such that $\Sigma_{U_1} = I_q$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The sequence of *weighted empirical measures* $(\nu_n(\omega, dx))_{n\geq 1}$ is then defined for every $n \geq 1$, by

$$\nu_n^\eta(\omega, f) = \frac{1}{H_n}\sum_{k=1}^n \eta_k \delta_{\bar{X}_{k-1}(\omega)}$$

where δ_a denotes the Dirac mass at $a \in \mathbb{R}^d$ and $(\eta_k)_{k\geq 1}$ is a sequence of positive weights such that $H_n = \sum_{k=1}^n \eta_k \xrightarrow{n\to+\infty} +\infty$. For the sake of simplicity, we assume in the sequel that

$$U_1 \sim \mathcal{N}(0, I_q) \quad \text{and} \quad \eta_k = \gamma_k$$

and we refer to [LPP13b] for a more general setting. Note that $\eta_k = \gamma_k$ is a genuine case, since it is the natural discretization of the occupation measure $\frac{1}{t}\int_0^t \delta_{X_s} ds$. For this sequence, we recall in Proposition 4.9 below in a synthesized form the main convergence results (including rates) of the sequence $(\nu_n^\eta(\omega, dx))$ to the invariant distribution ν of (X_t). In this way, we introduce two assumptions:

$(\mathbf{S_a}) : (a > 0)$ There exists a positive $\mathcal{C}^2$-function $V : \mathbb{R}^d \to \mathbb{R}$ with

$$\lim_{|x|\to+\infty} V(x) = +\infty, \quad |\nabla V|^2 \leq CV, \quad \text{and} \quad \sup_{x\in\mathbb{R}^d} \|D^2V(x)\| < +\infty$$

such that there exist some positive constants C_b, β and α such that:

$$(i) \quad |b|^2 \leq C_b V^a, \quad \mathrm{Tr}(\sigma\sigma^*)(x) = o(V^a(x)) \quad \text{as } |x| \to +\infty \quad (ii) \quad (\nabla V|b) \leq \beta - \alpha V^a.$$

This Lyapunov-type assumption is sufficient to ensure the long-time stability of the Euler sheme (in a sense made precise below) as soon as $a \in (0, 1]$. Note that the convergence can be obtained under a less restrictive mean-reverting assumption including the case $a = 0$ (see [Pan06]). The second assumption below is fundamental to establish the rate of convergence of $(\nu_n^\eta(\omega, f))$ to $\nu(f)$ for a fixed smooth enough function $f : \mathbb{R}^d \to \mathbb{R}$: we assume that f has a smooth solution to the Poisson equation (see [PV01] for results on this topic).

$(\mathbf{C(f, k)})$: There exists a $\mathcal{C}^k$-function $g_f : \mathbb{R}^d \to \mathbb{R}$ solution to $f - \nu(f) = \mathcal{A}g_f$ such that f, g_f and its partial derivatives up to k are dominated by V^r $(r \geq 0)$: $|f| \leq CV^r$ and for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ with $|\alpha| := \alpha_1 + \cdots + \alpha_d \in \{0, \ldots, k\}$, $|\partial^{|\alpha|}_{x_{i_1}^{\alpha_1}, \ldots, x_{i_d}^{\alpha_d}} g_f| \leq CV^r$.

Before recalling the results on $(\nu_n(\omega, dx))$, let us introduce the following notations:

$$\forall\, r \in \mathbb{N}, \qquad \Gamma_n^{(r)} = \sum_{k=1}^{n} \gamma_k^r$$

and for a smooth enough function $h : \mathbb{R}^d \to \mathbb{R}$ and an integer $r \geq 2$, we write:

$$D^{(r)}h(x)\, y_1 \otimes \cdots \otimes y_r = \sum_{(i_1,\ldots,i_r)\in\{1,\ldots,d\}^r} \partial^r_{x_{i_1},\ldots,x_{i_r}} h(x) y_1^{i_1} \ldots y_r^{i_r}.$$

Proposition 4.9. *Assume* $(\mathbf{S_a})$ *holds for an* $a \in (0,1]$. *Then,*
(ii) For every $r > 0$, $\sup_{n\geq 1} \nu_n(\omega, V^r) < +\infty$ *a.s. In particular,* $(\nu_n(\omega, dx))_{n\geq 1}$ *is a.s. tight.*
(iii) Every weak limit of $(\nu_n(\omega, dx))_{n\geq 1}$ *is an invariant distribution for* $(X_t)_{t\geq 0}$. *Furthermore, if (SDE) has a unique invariant distribution, say* ν, *then* $\nu_n^\eta(\omega, f) \xrightarrow{n\to+\infty} \nu(f)$ *a.s. for every* ν*-a.s continuous function* f *such that* $|f| \leq CV^r$ *for an* $r > 0$.
(iv) (Rate of convergence): Assume that ν *is unique and let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a function satisfying* $(\mathbf{C(f,5)})$. *Then,*

- *If* $\frac{\Gamma_n^{(2)}}{\sqrt{\Gamma_n}} \xrightarrow{n\to+\infty} 0$,

$$\sqrt{\Gamma_n}\,(\nu_n(\omega, f) - \nu(f)) \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\Big(0; \int |\sigma^* \nabla g_f|^2 d\nu\Big) \quad \text{as } n \to +\infty.$$

- *If* $\frac{\Gamma_n^{(2)}}{\sqrt{\Gamma_n}} \xrightarrow{n\to+\infty} \widetilde{\beta} \in (0, +\infty]$,

$$\triangleright\ \sqrt{\Gamma_n}\Big(\nu_n(\omega, f) - \nu(f)\Big) \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\Big(\widetilde{\beta}\, m_{g_f}^{(1)}; \int_{\mathbb{R}^d} |\sigma^* \nabla g_f|^2 d\nu\Big) \quad \text{as } n \to +\infty \quad \text{if } \widetilde{\beta} \in (0, +\infty),$$

$$\triangleright\ \frac{\Gamma_n}{\Gamma_n^{(2)}}\,(\nu_n(\omega, f) - \nu(f)) \overset{a.s.}{\longrightarrow} m_{g_f}^{(1)} \quad \text{as } n \to +\infty \qquad \text{if } \widetilde{\beta} = +\infty$$

where $m_{g_f}^{(1)} = \int \varphi_1 d\nu$ *with*

$$\varphi_1(x) = \frac{1}{2} D^2 g_f(x) b(x)^{\otimes 2} + \frac{1}{2}\mathbb{E}[D^3 g_f(x) b(x)(\sigma(x)U_1)^{\otimes 2}] + \frac{1}{24}\mathbb{E}[D^4 g_f(x)(\sigma(x)U_1)^{\otimes 4}]. \tag{4.25}$$

The first three claims (i), (ii) and (iii) of the theorem follow from [LP03] whereas (iv) is derived from [LP02] (see Theorem 10).

Polynomial step: Applying (iv) to polynomial steps of the form $\gamma_n = Cn^{-\mu}$, $\mu \in (0,1]$, we observe that the optimal (weak) rate is $n^{-1/3}$ and is attained for $\mu = 1/3$. Then

$$\widetilde{\beta} = \sqrt{6}\, C^{\frac{3}{2}} \ \text{ and } \ \sqrt{\Gamma_n} \sim \sqrt{3C/2}\, n^{\frac{1}{3}}.$$

so that

$$n^{\frac{1}{3}}\Big(\nu_n(\omega, f) - \nu(f)\Big) \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\Big(2C\, m_{g_f}^{(1)}; \frac{2}{3C}\int |\sigma^* \nabla g_f|^2 d\nu\Big).$$

This corresponds to the case where the rate of convergence of the underlying diffusion toward its steady regime ($\sqrt{\Gamma_n}$ corresponding to $\sqrt{t}$ in the continuous time setting, see [Bha82] for the Central Limit Theorem (CLT) for the diffusion itself) and the discretization error are of the same order.

Remark 4.9. From a practical point of view it seems clear that a balance should be made between the asymptotic bias and the asymptotic variance to specify the constant C. Under slightly more stringent assumptions we prove that the L^2–norm of the error $\nu_n(\omega, f) - \nu(f)$ satisfies

$$\|\nu_n(\omega, f) - \nu(f)\|_{L^2} \sim n^{-\frac{1}{3}} \sqrt{4C^2 (m^{(1)}_{g_f})^2 + \frac{2}{3C} \int |\sigma^* \nabla g_f|^2 d\nu}.$$

An optimization with respect to C gives the optimal choice $C = \left(\frac{12 \int_{\mathbb{R}^d} |\sigma^* \nabla g_f|^2 d\nu}{(m^{(1)}_{g_f})^2} \right)^{\frac{1}{3}}$.

4.1.2 The Richardson-Romberg extrapolated algorithm

When the step $(\gamma_n)_{n\geq 1}$ decreases to 0 fast enough so that $\frac{\Gamma_n^{(2)}}{\sqrt{\Gamma_n}} \xrightarrow{n\to+\infty} \widetilde{\beta} \in (0, +\infty)$, $(\nu_n(\omega, f))_{n\geq 1}$ converges toward the invariant distribution ν at the same rate as the empirical measure $\frac{1}{t} \int_0^t f(X_s) ds$ of the underlying diffusion. In both cases the resulting error satisfies a CLT at the rate $\sqrt{\Gamma_n}$ and $\sqrt{t}$ respectively) with the same asymptotic variance when the solution g_f to Poisson equation $Ag_f = -f$ is regular enough.
When γ_n goes to 0 too slowly ($\frac{\Gamma_n^{(2)}}{\sqrt{\Gamma_n}} \xrightarrow{n\to+\infty} +\infty$), the discretization error becomes prominent. This error takes the form of the bias $m^{(1)}_{g_f}$ defined by the above equation (4.25). This bias is the counterpart in the long run framework of the first order term in the expansion of the weak error when dealing with the discretization scheme (Euler) on a finite horizon (see [TT90]). In order to kill this term, we will call upon an *RR* extrapolation which amounts to combining in an appropriate way the occupation measures related to two (Euler) schemes where the step of the second scheme is half of that of the first one. This idea has been first investigated in [Lem05] where two sequences $(\gamma_n)_{n\geq 1}$ and $(\gamma_n/2)_{n\geq 1}$ are considered as step sequences. Both schemes are implemented with the same Gaussian white noise. Doing so does kill the bias and, as an important consequence, extends the range of validity of the CLT at rate $\sqrt{\Gamma_n}$ to larger steps, namely up to steps satisfying $\frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \xrightarrow{n\to+\infty} \widetilde{\beta} \in (0, +\infty)$. This condition reads in a polynomial setting for the steps $\mu \geq \frac{1}{5}$. When $\mu = \frac{1}{5}$, one obtains a biased CLT, this time at rate $n^{\frac{2}{5}}$.
However, as often with *RR* extrapolation, this choice of step did not allow for a control of the asymptotic variance (which is no longer $\int_{\mathbb{R}^d} |\sigma^* \nabla g_f|^2 d\nu$). Thus, the explosion of the variance may annihilate the bias reduction in the evaluation of a quadratic error. To overcome/get rid of this important drawback a *consistent RR* extrapolation can be devised (see [TT90] and, more recently, [Pag07] where this consistent approach is extensively investigated). By consistent, we mean that the Gaussian white noises in both involved discretization schemes are simulated in such a way that they derive form the same underlying standard Brownian motion. Our aim in this section is two fold: first propose a new discretization step to take full advantage of the consistent Brownian increments used in both schemes, then prove that when the duplicated system has ν_Δ as unique invariant distribution, the optimal choice for the brownian increments, among all possible (constant) correlations structures ρ, is definitely the consistent one which corresponds to $\rho = Id$. Then the rates are as above but with *the same variance* as in the non-extrapolated algorithm. These conclusions are contained in Theorem 4.3 and Proposition 4.10 that follows. We do not prove Theorem 4.3 in this paper. This proof will be detailed in a more

general work (in progress) on the RR extrapolation for invariant distribution computation including the case of non-Gaussian increments and with considerations on multistep RR extrapolation. Note however that a proof is also available in an extended version of this paper [LPP13a]).

As mentioned above, the starting idea is to introduce a second Euler scheme with *half step*. This half step sequence $(\widetilde{\gamma}_n)_{n\geq 1}$ is defined in a consistent way by

$$\forall n \geq 1, \quad \widetilde{\gamma}_{2n-1} = \widetilde{\gamma}_{2n} = \frac{\gamma_n}{2}$$

so that in "absolute time" we have $\widetilde{\Gamma}_{2n} = \Gamma_n$. As concerns the white noise of both schemes, our aim is to make them consistent in the above (absolute time) scale and (possibly) correlated (with correlation matrix ρ satisfying $I_q - \rho^*\rho \in \mathcal{S}^+(d,\mathbb{R})$). For the sake of simplicity, we only consider the Gaussian case.

Let $(W, W^{(\rho)})$ be a couple of Brownian Motions defined as in (1.2). We set

$$U_n = \frac{W_{\Gamma_n} - W_{\Gamma_{n-1}}}{\sqrt{\gamma_n}} \quad \text{and} \quad Z_n = \frac{W_{\widetilde{\Gamma}_n} - W_{\widetilde{\Gamma}_{n-1}}}{\sqrt{\widetilde{\gamma}_n}}$$

where $\tilde{\Gamma}_n = \sum_{k=1}^n \tilde{\gamma}_k$ (so that $\tilde{\Gamma}_{2n} = \Gamma_n$). In the sequel, we will then consider the original Euler scheme $(\bar{X}_n)_{n\geq 0}$ built with the above sequence $(U_n)_{n\geq 0}$ and a second one with step $\widetilde{\gamma}_n$ denoted by $(\bar{Y}_n^{(\rho)})_{n\geq 0}$ and defined by:

$$\bar{Y}_{n+1}^{(\rho)} = \bar{Y}_n^{(\rho)} + \widetilde{\gamma}_{n+1} b(\bar{Y}_n^{(\rho)}) + \sqrt{\widetilde{\gamma}_{n+1}}\sigma(\bar{Y}_n^{(\rho)}) Z_{n+1}^{(\rho)}, \; n \geq 1, \; \bar{Y}_0^{(\rho)} = y.$$

The *duplicated Euler scheme* $(\bar{X}_n, \bar{Y}_{2n}^{(\rho)})$ is then an Euler scheme at time Γ_n of the duplicated diffusion $(X_t, X_t^{(\rho)})_{t\geq 0}$.

We denote $(\nu_n^{(\rho)}(\omega, dx))_{n\geq 1}$ the sequence of empirical measures with weight sequence $(\widetilde{\gamma}_n)_{n\geq 1}$ related to $(\bar{Y}_n^{(\rho)}(\omega))_{n\geq 1}$. The empirical measure $(\bar{\nu}_n^{\eta,(\rho)}(\omega, dx))_{n\geq 1}$ associated to the Richardson-Romberg extrapolation is defined by

$$\begin{aligned}
\nu_n^{(\rho)}(\omega, f) &= \frac{1}{\Gamma_n}\sum_{k=1}^n \frac{\gamma_k}{2}\left(f(\bar{Y}_{2(k-1)}^{(\rho)}(\omega)) + f(\bar{Y}_{2k-1}^{(\rho)}(\omega))\right) \\
\bar{\nu}_n^{(\rho)}(\omega, f) &= (2\nu_n^{(\rho)} - \nu_n(\omega, f)) \\
&= \frac{1}{\Gamma_n}\sum_{k=1}^n \gamma_k\left(f(\bar{Y}_{2(k-1)}^{(\rho)}(\omega)) + f(\bar{Y}_{2k-1}^{(\rho)}(\omega)) - f(\bar{X}_k(\omega))\right).
\end{aligned}$$

Under the assumptions of Proposition 4.9, it is clear that $\bar{\nu}_n^{(\rho)}(\omega, dx) \xrightarrow{n\to+\infty} \nu(dx)$ *a.s.*.

Thus, in the next section, we propose to evaluate the effects of the Richardson-Romberg extrapolation on the rate of convergence of the procedure and to explain why the uniqueness of the invariant distribution of the duplicated diffusion plays an important role in this problem.

4.2 Rate of convergence of the extrapolated procedure

We set $(D^3 g_{i,.,.})_{i=1}^d = D^2(\nabla_.)g$ in order that the notation $\mathrm{Tr}(\sigma^* D^2(\nabla_.)g\sigma)$ stands for the *vector* of $\mathbb{R}^d$ defined by $\mathrm{Tr}(\sigma^* D^2(\nabla_.)g\sigma) = (\mathrm{Tr}(\sigma^* D^2(\partial_{x_i} g)\sigma))_{i=1}^d$. For a fixed matrix ρ,

the main result about the RR extrapolation is Theorem 4.3 below. At this stage, we do not discuss the choice of the correlation ρ in this result. This point is tackled in Proposition 4.10 in which we will see that the optimal choice to reduce the asymptotic variance is atteined with $\rho = I_q$ as soon as ν_Δ is the unique invariant distribution of the associated duplicated diffusion. This emphasizes the importance of the question of the uniqueness of the invariant distribution in this pathologic case studied in the previous part of the paper.

Theorem 4.3. *Assume* $(\mathbf{S_a})$ *holds for an* $a \in (0,1]$. *Assume that* $(X_t, X_t^{(\rho)})_{t\geq 0}$ *admits a unique invariant distribution* $\mu^{(\rho)}$ *(with marginals* ν*). Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a function satisfying* $(\mathbf{C(f,7)})$ *and such that* φ_1 *defined by* (4.25) *satisfies* $(\mathbf{C(\varphi_1,5)})$ *with a solution to the Poisson equation denoted by* g_{φ_1}. *Then,*

- *If* $\frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \xrightarrow{n\to+\infty} 0$,

$$\sqrt{\Gamma_n}\Big(\nu_n^{(\rho)}(\omega,f) - \nu(f)\Big) \overset{n\to+\infty}{\Longrightarrow} \mathcal{N}\big(0;\hat\sigma_\rho^2\big)$$

where

$$\hat\sigma_\rho^2 = 5\int_{\mathbb{R}^d} |\sigma^*\nabla g_f|^2 d\nu - 4\int_{\mathbb{R}^d\times\mathbb{R}^d} \big((\sigma^*\nabla g_f)(x)|\rho(\sigma^*\nabla g_f)(y)\big)\mu^{(\rho)}(dx,dy). \tag{4.26}$$

- *If* $\frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \xrightarrow{n\to+\infty} \widetilde\beta \in (0,+\infty]$, *then*

$$\sqrt{\Gamma_n}\big(\nu_n^{(\rho)}(\omega,f) - \nu(f)\big) \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\big(\widetilde\beta\, m_{g_f}^{(2)};\hat\sigma_\rho^2\big) \quad as \;\; n\to+\infty \qquad \text{if } \widetilde\beta\in(0,+\infty),$$

$$\frac{\Gamma_n}{\Gamma_n^{(3)}}(\nu_n^{(\rho)}(\omega,f) - \nu(f)) \overset{\mathbb{P}}{\to} m_{g_f}^{(2)} \quad as \;\; n\to+\infty \qquad \text{if } \widetilde\beta = +\infty,$$

where $m_{g_f}^{(2)} = \dfrac{1}{2}\left(m_{g_{\varphi_1}} + \displaystyle\int_{\mathbb{R}^d}\varphi_2 d\nu\right)$ *with*

$$\varphi_2(x) = \sum_{k=3}^{6} \frac{C_k^{2(k-3)}}{k!}\mathbb{E}\big[D^k g_f(x) b(x)^{\otimes(6-k)}(\sigma(x)U_1)^{\otimes 2(k-3)}\big]. \tag{4.27}$$

Remark 4.10. (Polynomial steps) Let $\gamma_n = Cn^{-\mu}$, $\mu\in(0,1]$. If $\mu > \frac13$, $\Gamma_n^{(3)} \to \Gamma_\infty^{(3)} < +\infty$ so that $\frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \to 0$ as $n\to+\infty$. If $\mu < \frac13$, $\frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \asymp n^{\frac{1-5\mu}{2}}$ (and if $\mu = \frac13$, $\frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \asymp \frac{\log n}{\sqrt{n}}$). Consequently

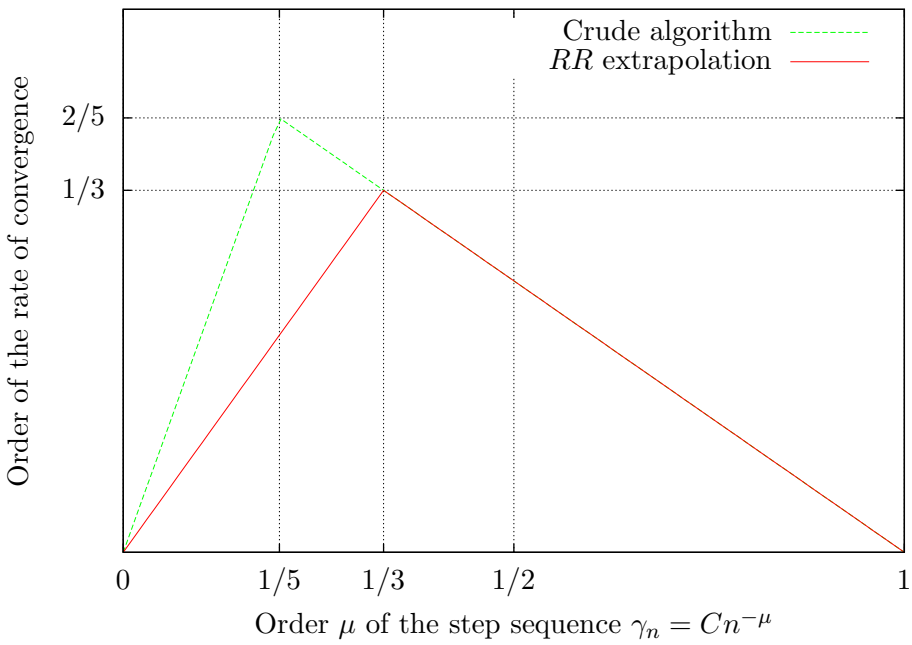
$$\frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \to 0 \iff \mu > \frac15,\; \frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \to +\infty \iff \mu < \frac15 \;\text{ and }\; \frac{\Gamma_n^{(3)}}{\sqrt{\Gamma_n}} \to \widetilde\beta\in(0,+\infty) \iff \mu = \frac15.$$

When $\mu = \frac15$, $\widetilde\beta = C^{\frac52}\sqrt5$ and $\sqrt{\Gamma_n} \sim \frac{\sqrt{5C}}{2} n^{\frac25}$.

As a consequence, if $\gamma_n = \eta_n = Cn^{-\frac15}$,

$$n^{\frac25}\big(\nu_n^{(\rho)}(\omega,f) - \nu(f)\big) \overset{(\mathbb{R})}{\Longrightarrow} \mathcal{N}\Big(2C^2\, m_g^{(2)}; \frac45\frac{\widehat\sigma_\rho^2}{C}\Big).$$

We switch from a weak rate $n^{\frac13}$ to $n^{\frac25}$ *i.e.* a "gain" of $n^{\frac{1}{15}}$ (see figure below). The second noticeable fact is that the bias is now significantly more sensitive to the constant C than in the standard setting. If we minimize the L^2–norm of the error $\nu_n^{(\rho)}(\omega,f) - \nu(f)$ we obtain the optimal choice of C as a function of both bias and standard deviation, precisely $C = \Big(\frac{\widehat\sigma_\rho^2}{20(m_q^{(2)})^2}\Big)^{\frac15}$.



4.2.1 Optimal choice of ρ and uniqueness of $\mu^{(I_d)}$

Proposition 4.10. *Let ρ be an admissible correlation matrix i.e. such that $\rho^*\rho \leq I_q$. Assume that the duplicated diffusion $(X, X^{(\rho)})$ has a unique invariant distribution $\mu^{(\rho)}$ (so that if $\rho = I_q$, $\mu^{(I_q)} = \nu_\Delta$).*

(a) $\widehat{\sigma}^2_\rho \geq \displaystyle\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu$.

(b) If $\rho = 0$ then $\widehat{\sigma}^2_\rho = 5\displaystyle\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu$.

(c) If $\rho = I_q$, $\widehat{\sigma}^2_\rho = \displaystyle\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu$.

Proof. Claims (b) and (c) being obvious thanks to (4.26), we only prove (a). Keeping in mind that both marginals $\mu^{(\rho)}(\mathbb{R}^d \times dy)$ and $\mu^{(\rho)}(dx \times (\mathbb{R}^d)$ are equal to ν, one derives thanks to Schwarz's Inequality (once on $\mathbb{R}^d$ and once on $L^2(\mu)$) from the expression (4.26) of the asymptotic variance $\widehat{\sigma}^2_\rho$ that

$$\begin{aligned}
\widehat{\sigma}^2_\rho &\geq 5\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu - 4\left[\int_{\mathbb{R}^d\times\mathbb{R}^d} |\sigma^*\nabla g|^2(x)\mu^{(\rho)}(dx,dy)\right]^{\frac{1}{2}}\left[\int_{\mathbb{R}^d\times\mathbb{R}^d} |\rho\sigma^*\nabla g|^2(y)\mu^{(\rho)}(dx,dy)\right]^{\frac{1}{2}} \\
&= 5\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu - 4\left[\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2(x)\nu(dx)\right]^{\frac{1}{2}}\left[\int_{\mathbb{R}^d} |\rho\sigma^*\nabla g|^2(y)\nu(dy)\right]^{\frac{1}{2}} \\
&\geq 5\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu - 4\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu = \int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu
\end{aligned}$$

where we used in the last inequality that $|\rho u|^2 \leq |u|^2$.
The previous result says that the structural asymptotic variance of the RR estimator is always greater than that of the standard estimator but can be equal if the Brownian

motions are equal. This condition is in fact almost necessary. Actually, thanks to the Pythagorean identity,

$$\sigma_\rho^2 = 5\int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu + 2\int_{\mathbb{R}^d\times\mathbb{R}^d} |\sigma^*\nabla g(x) - \rho\sigma^*\nabla g(y)|^2 \mu^{(\rho)}(dx,dy)$$
$$- 2\int_{\mathbb{R}^d\times\mathbb{R}^d} |\sigma^*\nabla g(x)|^2 \nu(dx) - 2\int_{\mathbb{R}^d\times\mathbb{R}^d} |\rho\sigma^*\nabla g(y)|^2 \nu(dy).$$

Then, since $\rho^*\rho \leq I_q$, a necessary condition to obtain $\sigma_\rho^2 = \int_{\mathbb{R}^d} |\sigma^*\nabla g|^2 d\nu$ is

$$|\rho\sigma^*\nabla g(y)| = |\sigma^*\nabla g(y)| \quad \nu(dy)\text{-}a.e.$$

When $\rho^*\rho < I_q$, this equality can not hold except if $\sigma^*\nabla g(y) = 0$ $\nu(dy)$-$a.e.$

The above decomposition is built as follows: the second term of (**??**) is the main martingale component of the decomposition whereas $\mathcal{E}_{n,1}$ contains the first order discretization error. Thanks to the Richardson-Romberg extrapolation, $\mathcal{E}_{n,1}$ is in fact negligible when $n \to +\infty$. When the step sequence decreases fast (Theorem4.3(i)), the rate of convergence is ruled by the main martingale component. In Theorem 4.3(ii), the rate is ruled by $\mathcal{E}_{n,1}$ and $\mathcal{E}_{n,2}$. Finally, N_n contains all the negligible martingale terms.

A Hypo-ellipticity of the correlated duplicated system

It is a well-known fact that, for a Markov process, the strong Feller property combined with some irreducibility of the transitions implies uniqueness of the invariant distribution (see *e.g.* [DPZ96], Theorem 4.2.1). For a diffusion process with smooth coefficients, such properties hold if it satisfies the hypoelliptic Hörmander assumption (see [Hör67, Hör85]) and if the deterministic system related to the stochastic differential system (written in the Stratanovich sense) is controllable. In fact, both properties can be transferred from the original SDE to the duplicated system so that its invariant distribution is also unique. The main result of this section is Proposition A.11. Before, we need to introduce some Hörmander-type notations. First, written in a Stratonovich way, X is a solution to

$$dX_t = A_0(X_t)dt + \sum_{j=1}^q A_j(X_t) \circ dW_t^j \tag{A.28}$$

where $A_0, \dots A_q$ are vectors fields on $\mathbb{R}^d$ defined by[1]:

$$A_0(x) = \sum_{i=1}^d \left[b_i(x) - \frac{1}{2}\sum_{l,j} \sigma_{l,j}(x)\partial_{x_j}\sigma_{i,l}(x) \right] \partial_{x_i}$$

and for every $j \in \{1,\dots,q\}$:

$$A_j(x) = \sum_{i=1}^d \sigma_{i,j}(x)\partial_{x_i}.$$

For the sake of simplicity, we assume that b and σ are $\mathcal{C}^\infty$ on $\mathbb{R}^d$ with bounded derivatives. We will also assume the following Hörmander condition at each point: there exists $N \in \mathbb{N}^*$

[1] With a standard abuse of notation, we identify the vectors fields and the associated differential operators.

such that $\forall x \in \mathbb{R}^d$,

$$\dim\left(\text{Span}\left\{A_1(x), A_2(x), \ldots, A_q(x),\ \text{L. B. of length} \leq N \text{ of the } A_j(x)\text{'s }, 0 \leq j \leq q\right\}\right) = d \tag{A.29}$$

where "L.B." stands for Lie Brackets. The above assumptions imply that for every $t > 0$ and $x \in \mathbb{R}^d$, $P_t(x, .)$ admits a density $p_t(x, .)$ *w.r.t.* the Lebesgue measure and that $(x, y) \mapsto p_t(x, y)$ is $\mathcal{C}^\infty$ on $\mathbb{R}^d \times \mathbb{R}^d$ (see *e.g.* [Cat92], Theorem 2.9). In particular, $x \mapsto P_t(x, .)$ is a strong Feller semi-group. Assume also that the control system (associated with (A.28))

$$\dot{x}^{(u)} = A_0(x^{(u)}) + \sum_{j=1}^{q} A_q(x^{(u)}) u_j, \tag{A.30}$$

is *approximatively*-controllable:

$$\text{There exists } T > 0 \text{ such that for every } \varepsilon > 0,\ x_1, x_2 \in \mathbb{R}^d, \text{ there exists } u \in L^2([0, T], \mathbb{R}^d) \text{ such that } (x^{(u)}(t)) \text{ solution to (A.30) satisfies } x(0) = x_1 \text{ and } |x(T) - x_2| \leq \varepsilon. \tag{A.31}$$

Under Assumptions (A.29) and (A.31), the diffusion has a unique invariant distribution ν. Actually, the controllability assumption combined with the Support Theorem implies that for every non-empty open set O, for every $x \in \mathbb{R}^d$, $P_T(x, O) > 0$. The semi-group (P_t) is then irreducible. Owing to the strong Feller property, it follows classically that (P_t) admits a unique invariant distribution (see *e.g.* [DPZ96], Proposition 4.1.1. and Theorem 4.2.1.).
Furthemore, ν is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ and its topological support is $\mathbb{R}^d$ (since for every open set O of $\mathbb{R}^d$, $\nu(O) = \int P_T(x, 0)\nu(dx) > 0$). Let us now consider the duplicated diffusion $(X_t, X_t^{(\rho)})$. Setting $Z_t^{(\rho)} = (X_t, X_t^{(\rho)})$ and using the preceding notations, (1.3) can be written:

$$dZ_t^{(\rho)} = \widetilde{A}_0(Z_t^{(\rho)})dt + \sum_{j=1}^{q} \widetilde{A}_j(Z_t^{(\rho)})dW_t^j + \sum_{j=1}^{q} \widetilde{A}_{d+j}(Z_t^{(\rho)})d\widetilde{W}_t^j$$

where $\widetilde{A}_0(z) = (A_0(x), A_0(y))^T$ (with $A_0(y) = \sum_{i=1}^d \left[b_i(y) - \frac{1}{2}\sum_{l,j} \sigma_{l,j}(y)\partial_{y_j}\sigma_{i,l}(y)\right]\partial_{y_i}$ and $z = (x, y)$), $\widetilde{W}$ is a d-dimensional Brownian Motion independent of W such that $W^{(\rho)} = \rho^* W + (I_q - \rho^*\rho)^{\frac{1}{2}}\widetilde{W}$ and for every $j \in \{1, \ldots, q\}$,

$$\widetilde{A}_j(z) = A_j(x) + A_j^{(\rho)}(y) \quad \text{and,} \quad \widetilde{A}_{q+j}(z) = A_j^{((I_q - \rho^*\rho)^{\frac{1}{2}})}(y) \tag{A.32}$$

where for a for a $q \times q$ matrix B, $A_j^{(B)}(y) = \sum_{i=1}^d (\sigma(y)B)_{i,j}\partial_{y_i}$. Then, the following property holds.

Proposition A.11. *Let $\rho \in \mathbb{M}_{q,q}(\mathbb{R})$ such that $\rho^*\rho < I_q$. Assume that b and σ are $\mathcal{C}^\infty$ on $\mathbb{R}^d$ with bounded derivatives. Assume (A.29) and (A.31). Then, uniqueness holds for the invariant distribution $\nu^{(\rho)}$ of the duplicated diffusion $(X_t, X_t^{(\rho)})$. Furthermore, if $\nu^{(\rho)}$ exists, then $\nu^{(\rho)}$ has a density $p^{(\rho)}$ (w.r.t. λ_{2d}) which is a.s. positive.*

Proof. First, let us check the Hörmander conditions for $(X_t, X_t^{(\rho)})_{t\geq 0}$. Setting $S = (I_q - \rho^*\rho)^{\frac{1}{2}}$, standard computations yield

$$\forall j \in \{1,\dots,q\}, \quad \widetilde{A}_{q+j}(z) = \sum_{l=1}^{q} S_{l,j} A_l(y).$$

Since S is invertible, we deduce that $\{A_l(y), l = 1,\dots,q\}$ belongs to $\mathrm{Span}\{\widetilde{A}_{q+j}(z), j = 1,\dots,q\}$. Similarly, checking that for every $j \in \{1,\dots,q\}$,

$$[\widetilde{A}_0(z), \widetilde{A}_{q+j}(z)] = [A_0(y), A_j^{(S)}(y)] = \sum_{l=1}^{q} S_{l,j}[A_0(y), A_l(y)]$$

one deduces from the invertibility of S that $\{[A_0(y), A_l(y)], l = 1,\dots,q\}$ is included in $\mathrm{Span}\{[\widetilde{A}_0(z), \widetilde{A}_{q+j}(z)], j = 1,\dots,q\}$. Owing to (A.29), it follows that $\mathrm{Span}\{\partial_{y_1},\dots,\partial_{y_d}\}$ is included in

$$V = \mathrm{Span}\left\{\widetilde{A}_1(z), \widetilde{A}_2(z),\dots,\widetilde{A}_q(z), \text{ Lie Brackets of length} \leq N \text{ of the } \widetilde{A}_j(z)\text{'s }, 0 \leq j \leq q\right\}.$$

Now, let us show that $\mathrm{Span}\{\partial_{x_1},\dots,\partial_{x_d}\}$ is included in V. Since $\mathrm{Span}\{\partial_{y_1},\dots,\partial_{y_d}\}$ is included in V, it is clear that for every $x \in \mathbb{R}^d$, $A_j(x) = A_j^{(\rho)}(y) - \widetilde{A}_j(z)$ also belongs to V. Since

$$[\widetilde{A}_0(z), \widetilde{A}_j(z)] = [A_0(x), A_j(x)] + [A_0(y), A_j^{(\rho)}(y)],$$

$[A_0(x), A_j(x)]$ has the same property. Using again (A.29), we deduce that $\mathrm{Span}\{\partial_{x_1},\dots,\partial_{x_d}\}$ is included in V and thus that $\dim(V) = 2d$. As a consequence, for every $z \in \mathbb{R}^d \times \mathbb{R}^d$ and $t > 0$, $Q_t^{(\rho)}(z,.)$ admits a density $q_t(z,.)$ *w.r.t.* λ_{2d} such that $(z,z') \mapsto q_t(z,z')$ is $\mathcal{C}^\infty$ on $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$.
In order to obtain uniqueness for the invariant distribution, it remains to show that there exists $T > 0$ such that for every $z \in \mathbb{R}^d \times \mathbb{R}^d$, for every non-empty open set O of $\mathbb{R}^d \times \mathbb{R}^d$, $Q_T(z,O) > 0$. Owing to (A.31), it is clear that for every $z_1 = (x_1,y_1)$ and $z_2 = (x_2,y_2)$, for every $\varepsilon > 0$, there exist u and $\widetilde{u} \in L^2([0,T],\mathbb{R}^d)$ such that $z(t) = (x^{(u)}(t), x^{(\widetilde{u})}(t))$, where $x^{(u)}$ and $x^{(\widetilde{u})}$ are solutions to (A.30) starting from x_1 and y_1, satisfies $|z(T) - z_2| \leq \varepsilon$. Furthermore, since S is invertible, we can assume that $\widetilde{u} = \rho u + S\omega$ with $\omega \in L^2([0,T],\mathbb{R}^d)$. Then, the support Theorem can be applied to obtain that for every z_1, z_2, ε $Q_T(z_1, B(z_2, \frac{\varepsilon}{2}) > 0$ and thus to conclude that for every $z \in \mathbb{R}^d \times \mathbb{R}^d$ and every non-empty open set O, $Q_T(z,O) > 0$. □

B Additional proofs about the two-dimensional counterexample

Proof of (3.7): For the sake of completeness, we show that $r_t \to 1$ *a.s.* as soon as $r_0 > 0$. First, note that uniqueness holds for the solution of the SDE (3.5) since the coefficients are Lipschitz continuous. In particular, (r_t^1) defined *a.s.* by $r_t^1 = 1$ for every $t \geq 0$ is the unique solution starting from $r_0 = 1$. Owing to the strong Markov property, this implies that if $\tau^1 := \inf\{t \geq 0, r_t = 1\}$, then $r_t = 1$ on $\{\tau \leq t\}$. The same property holds at 0. We deduce that $(r_t)_{t\geq 0}$ lives in $[1,+\infty)$ if $r_0 > 1$ and in $[0,1]$ if $r_0 \in [0,1]$. Moreover, if $r_0 > 1$, we have $d(r_t - 1) = -(r_t - 1)(dt + \vartheta dW_t)$ so that

$$r_t - 1 = e^{-(1+\frac{\vartheta^2}{2})t + \vartheta W_t}.$$

It follows that $\lim_{t\to+\infty} r_t = 1$ since $\lim_{t\to+\infty} \frac{W_t}{t} = 0$ $a.s.$. Now, if $r_0 \in [0,1]$, we have

$$dr_t = r_t(1-r_t)(dt + \vartheta dW_t).$$

Thus, (r_t) is a $[0,1]$-valued submartingale. In particular, r_t converges $a.s.$ to a $[0,1]$-valued random variable r_∞. Since

$$\forall t \geq 0, \qquad \mathbb{E}[r_t] = r_0 + \mathbb{E}\Big(\int_0^t r_s(1-r_s)ds\Big),$$

it follows that $\mathbb{E}[\int_0^{+\infty} r_s(1-r_s)ds]$ which in turn implies that $\int_0^{+\infty} r_s(1-r_s)ds < +\infty$ $a.s.$ As a consequence $\liminf_{t\to+\infty} r_t(1-r_t) = 0$ $a.s.$. The process (r_t) being $a.s.$ convergent to r_∞, it follows that $r_\infty \in \{0,1\}$ $a.s.$. It remains to prove that $\mathbb{P}(r_\infty = 0) = 0$. Denote by p the scale function of (r_t) null at $r = 1/2$. For every $r \in (0,1)$,

$$p(r) = \int_{\frac{1}{2}}^r e^{-\int_{\frac{1}{2}}^{\xi} \frac{2}{\vartheta^2 u(1-u)} du} d\xi = \int_{\frac{1}{2}}^r \left(\frac{1-\xi}{\xi}\right)^{\frac{2}{\vartheta^2}} d\xi.$$

As a consequence, if $\vartheta \in (0, \sqrt{2}]$, $\lim_{r\to+\infty} p(r) = +\infty$. This means that 0 is a repulsive point and that, as a consequence (see *e.g.* [KT81], Lemma 6.1 p. 228),

$$\forall b \in (0,1) \qquad \mathbb{P}(\lim_{a\to 0^+} \tau_a < \tau_b) := \lim_{a\to 0^+} \mathbb{P}(\tau_a < \tau_b) = 0$$

where $\tau_a = \inf\{t \geq 0 \,|\, r_t = a\}$, $y \in [0,1]$. We deduce that $\mathbb{P}(r_\infty = 0) = 0$. This completes the proof. $\square$

Proof of (3.8)**:** We want to prove that μ is invariant for $(X_t^x, X_t^{x'})$ if and only if μ can be represented by (3.8). First, since the unique invariant distribution of (X_t^x) is λ_{S_1}, it is clear that $\mu = \mathcal{L}(e^{i\Theta_0}, e^{i(\Theta_0+V_0)})$ where Θ_0 has uniform distribution on $[0, 2\pi]$ and V_0 is a random variable with values in $[0, 2\pi)$. One can check that if V_0 is independent of Θ_0, μ is invariant. Thus, it remains to prove that it is a necessary condition or equivalently that $K(\theta, dv) := \mathcal{L}(e^{iV_0}|e^{i\Theta_0} = e^{i\theta})$ does not depend on θ. Denote by $(e^{i\Theta_t}, e^{i(\Theta_t+V_t)})$ the (stationary) duplicated diffusion starting from $(e^{i\Theta_0}, e^{i(\Theta_0+V_0)})$. Since μ is invariant, we have for every $t \geq 0$

$$\mathcal{L}(e^{iV_t}|e^{i\Theta_t} = e^{i\theta}) = K(\theta, dv)$$

but thanks to the construction, for every $t \geq 0$, $\Theta_t = \Theta_0 + W_t$ and $V_t = V_0$ (the angular difference between the two coordinates does not change) so that

$$\mathcal{L}(e^{iV_t}|e^{i\Theta_t} = e^{i\theta}) = \int K(\theta', dv)\rho_t(\theta, d\theta')$$

where $\rho_t(\theta, d\theta') = \mathcal{L}(e^{i(\theta+W_t)})$. But $\rho_t(\theta, d\theta')$ converges weakly to λ_{S_1} when $t \to +\infty$. From the two previous equations it follows that $K(\theta, dv)$ does not depend on θ since $\forall \theta \geq 0$, $K(\theta, dv) = \int K(\theta', dv)\lambda_{S_1}(d\theta')$. $\square$

Acknowledgement: We thank an anonymous referee and an associate editor for their suggestions which helped improving the paper.

References

[Bax91] Peter H. Baxendale. Statistical equilibrium and two-point motion for a stochastic flow of diffeomorphisms. In *Spatial stochastic processes*, volume 19 of *Progr. Probab.*, pages 189–218. Birkhäuser Boston, Boston, MA, 1991.

[BB92] G. K. Basak and R. N. Bhattacharya. Stability in distribution for a class of singular diffusions. *Ann. Probab.*, 20(1):312–321, 1992.

[Bha82] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete*, 60(2):185–201, 1982.

[BS88] P. H. Baxendale and D. W. Stroock. Large deviations and stochastic flows of diffeomorphisms. *Probab. Theory Related Fields*, 80(2):169–215, 1988.

[Car85] Andrew Carverhill. Flows of stochastic dynamical systems: ergodic theory. *Stochastics*, 14(4):273–317, 1985.

[Cat92] P. Cattiaux. Stochastic calculus and degenerate boundary value problems. *Ann. Inst. Fourier (Grenoble)*, 42(3):541–624, 1992.

[CL89] M. F. Chen and S. F. Li. Coupling methods for multidimensional diffusion processes. *Ann. Probab.*, 17(1):151–177, 1989.

[DKK04] Dmitry Dolgopyat, Vadim Kaloshin, and Leonid Koralov. Sample path properties of the stochastic flows. *Ann. Probab.*, 32(1A):1–27, 2004.

[DPZ96] G. Da Prato and J. Zabczyk. *Ergodicity for infinite-dimensional systems*, volume 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[Har81] T. E. Harris. Brownian motions on the homeomorphisms of the plane. *Ann. Probab.*, 9(2):232–254, 1981.

[Has80] R. Z. Has′minskiĭ. *Stochastic stability of differential equations*, volume 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn, 1980. Translated from the Russian by D. Louvish.

[Hör67] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[Hör85] L. Hörmander. *The analysis of linear partial differential operators 1-4.*, volume 275 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1985.

[IW77] N. Ikeda and S. Watanabe. A comparison theorem for solutions of stochastic differential equations and its applications. *Osaka J. Math.*, 14(3):619–633, 1977.

[KP92] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[KT81] S. Karlin and H. M. Taylor. *A second course in stochastic processes*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.

[Kun90] H. Kunita. *Stochastic flows and stochastic differential equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[Lem05] V. Lemaire. *Estimation récursive de la mesure invariante d'un processus de diffusion*. Thèse de doctorat, Université de Marne-la-Vallée (France), 2005.

[LP02] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3):367–405, 2002.

[LP03] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. *Stoch. Dyn.*, 3(4):435–451, 2003.

[LPP13a] V. Lemaire, G. Pagès, and F. Panloup. Invariant measure of duplicated diffusions and application to Richardson-Romberg extrapolation. *http://arxiv.org/abs/1302.1651*, 2013.

[LPP13b] V. Lemaire, G. Pagès, and F. Panloup. The multistep Richardson-Romberg extrapolation for the approximation of invariant distributions: Theoretical results and numerical aspects. *in progress*, 2013.

[Pag07] G. Pagès. Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods Appl.*, 13(1):37–70, 2007.

[Pan06] F. Panloup. *Approximation récursive du régime stationnaire d'une Equation Différentielle Stochastique avec sauts.* Thèse de doctorat, Université Pierre et Marie Curie (Paris 6, France), 2006.

[PV01] E. Pardoux and A. Yu. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.

[Tea08] Oliver M. Tearne. Collapse of attractors for ODEs under small random perturbations. *Probab. Theory Related Fields*, 141(1-2):1–18, 2008.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.