ESAIM: M2AN 47 (2013) 1287–1314
DOI: 10.1051/m2an/2013068

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

NUMERICAL ANALYSIS OF PARALLEL REPLICA DYNAMICS

Gideon Simpson[1] and Mitchell Luskin[1]

Abstract. Parallel replica dynamics is a method for accelerating the computation of processes characterized by a sequence of infrequent events. In this work, the processes are governed by the overdamped Langevin equation. Such processes spend much of their time about the minima of the underlying potential, occasionally transitioning into different basins of attraction. The essential idea of parallel replica dynamics is that the exit distribution from a given well for a single process can be approximated by the distribution of the first exit of N independent identical processes, each run for only $1/N$-th the amount of time. While promising, this leads to a series of numerical analysis questions about the accuracy of the exit distributions. Building upon the recent work in [C. Le Bris, T. Lelièvre, M. Luskin and D. Perez, *Monte Carlo Methods Appl.* **18** (2012) 119–146], we prove a unified error estimate on the exit distributions of the algorithm against an unaccelerated process. Furthermore, we study a dephasing mechanism, and prove that it will successfully complete.

Mathematics Subject Classification. 60H35, 65C20, 65C30, 70K70, 74S60.

Received April 3, 2012. Revised January 31, 2013.
Published online July 9, 2013.

1. Introduction

Parallel replica dynamics (ParRep) is a numerical tool first introduced by Voter in [26] (see also [22, 27]) for accelerating the simulation of stochastic processes characterized by a sequence of infrequent, but rapid, transitions from one state to another. A standard and important problem in which such a separation of scales is present is the migration of defects through a crystalline lattice; see [22] and references therein for examples.

Roughly, the idea behind parallel replica dynamics is as follows. Suppose a trajectory spends time t in a particular state, before transitioning into another. Furthermore, assume t is large, relative to the scale of the time step discretization. We wish to avoid directly simulating a single realization for time t. We approximate the simulation of a single trajectory for time t with N independent copies, each simulated for time t/N, and follow the particular trajectory that escapes first. This holds out the promise for a linear speedup with the number of independent realizations we are able to simulate.

Of course, this is not exact, and error is introduced. A particular concern is error in the exit distributions of the system as it migrates from one state to another – does ParRep disrupt the state to state dynamics? Inspired by the tools proposed in [4], we prove an error estimate on the exit distributions over a single "cycle" of ParRep (the transition from one state to the next).

Keywords and phrases. Accelerated dynamics, rare events, parallel replica.

[1] School of Mathematics, University of Minnesota, 206 Church St SE, 127 Vincent Hall, Minneapolis MN 55455, USA. gsimpson@umn.edu; luskin@umn.edu

Article published by EDP Sciences

© EDP Sciences, SMAI 2013

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

1.1. The algorithm

We assume the system we wish to accelerate evolves according to the overdamped Langevin equation,

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}B_t, \quad X_t \in \mathbb{R}^n, \tag{1.1}$$

where B_t is a Wiener process and β is proportional to inverse temperature. Though ParRep was originally developed for the Langevin equations, it is readily adapted to this problem.

We next assume that our system is such that V has a denumerable set of local minima, x_j, $j = 1, 2, \ldots$ For each minima, we associate a set $W_j \subset \mathbb{R}^n$, the "well." W_j could be the basin of attraction of x_j; if $y(t)$ solves the ODE

$$\dot{y} = -\nabla V(y), \quad y(0) = y_0 \in \mathbb{R}^n,$$

then

$$W_j = \left\{ y_0 : \lim_{t\to+\infty} y(t) = x_j \right\}.$$

However, this definition is not essential; for the sake of our analysis, W_j need only be a bounded set in $\mathbb{R}^n$ with sufficiently regular boundary.

This motivates defining the well selection function,

$$\mathcal{S} : \mathbb{R}^n \to \mathbb{N}, \tag{1.2}$$

which identifies the basin associated with a given position. Associated with this is the "coarse grained" trajectory,

$$\mathcal{S}_t \equiv \mathcal{S}(X_t) \tag{1.3}$$

which only identifies the present well.

If the wells are "deep" with well-defined minima, then X_t will infrequently transition from one to another. Such a well corresponds to a *metastable* state. Much of the simulation time will be spent waiting for a jump to occur. The goal of ParRep is to reduce this computational expense by providing a satisfactory approximation of the form

$$\mathcal{S}_t \approx \mathcal{S}_t^{\mathrm{ParRep}}. \tag{1.4}$$

In other words, we are willing to sacrifice information about where the trajectory is within each well, for the sake of rapidly computing the sequence of wells the trajectory visits.

We now describe the ParRep algorithm in the following steps: the decorrelation step; the dephasing step; and the parallel step. These steps are diagrammed in Figures 1 and 2. We assume that the reference process X_t^{ref} enters well W_j at time t_{sim}.

A. *Decorrelation Step*: Let X_t^{ref} evolve under (1.1) for $t_{\mathrm{sim}} \le t \le t_{\mathrm{sim}} + t_{\mathrm{corr}}$.
   - If
   $$\mathcal{S}(X_t^{\mathrm{ref}}) = \mathcal{S}(X_{t_{\mathrm{sim}}}^{\mathrm{ref}})$$
   for all $t_{\mathrm{sim}} \le t \le t_{\mathrm{sim}} + t_{\mathrm{corr}}$, then time advances to $t_{\mathrm{sim}} + t_{\mathrm{corr}}$ and proceed.
   - Otherwise, denote the first exit time from the well,
   $$T = \inf\left\{ t \mid \mathcal{S}(X_{t_{\mathrm{sim}}+t}^{\mathrm{ref}}) \neq \mathcal{S}(X_{t_{\mathrm{sim}}}^{\mathrm{ref}}) \right\}$$
   and time advances to $t_{\mathrm{sim}} + T$. Return to the beginning of the decorrelation step in the new well.

B. *Dephasing Step:* In conjunction with the decorrelation step, we launch N replicas with starting positions drawn from distribution μ_{phase}^0. These are run for t_{phase} amount of time, the *dephasing time*. If at any time before t_{phase} a replica leaves the well, it is restarted. A replica has successfully dephased if it remains in the well for all of t_{phase}.

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

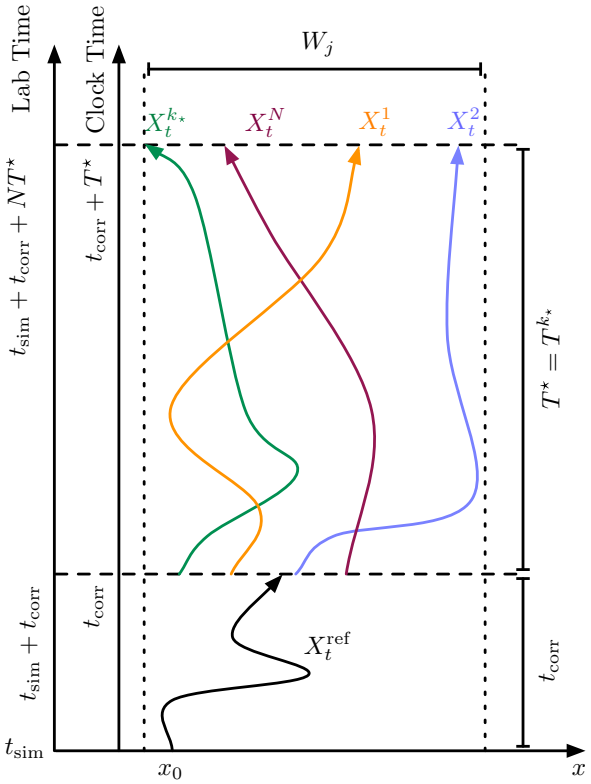


FIGURE 1. An illustration of the decorrelation and parallel steps of the ParRep algorithm in the case that the reference walker never leaves well W_j. $X_t^{k_\star}$ is the first process to exit the well, doing so at the computer time $t_{\rm corr} + T^\star$. This is then translated into the lab, or physical, time $t_{\rm sim} + t_{\rm corr} + NT^\star$. See Figure 2 for an illustration of a dephasing step.

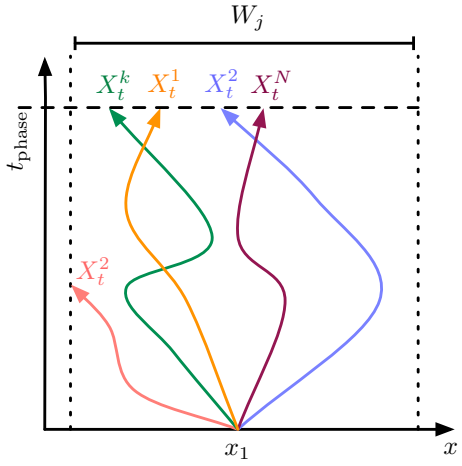


FIGURE 2. An illustration of a dephasing step for the ParRep algorithm. In this implementation, the replicas all start from the same position; $\mu^0_{\rm phase} = \delta_{x_1}$. When X_t^2 leaves before $t_{\rm phase}$, it is relaunched from the same position.

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

At the completion of the decorrelation and dephasing steps, assuming the reference walker has not exited, we have N independent walkers with the same distribution. We discard the reference process. If at any time during the dephasing process the reference walker leaves the well, the dephasing process terminates and the replicas are discarded.

C. *Parallel Step*: We now let the N replicas evolve independently and define

$$k_\star = \operatorname*{argmin}_k T^k, \tag{1.5a}$$

$$X_t^\star = X_t^{k_\star}, \tag{1.5b}$$

$$T^\star = T^{k_\star}. \tag{1.5c}$$

The system advances to the next well:

$$t_{\text{sim}} \mapsto t_{\text{sim}} + t_{\text{corr}} + NT^\star \tag{1.6a}$$

$$X^{\text{ref}}_{t_{\text{sim}}+t_{\text{corr}}+NT^\star} = X^\star_{T^\star}. \tag{1.6b}$$

Finally, we return to the decorrelation step.

This is a different dephasing algorithm than described in [4]. There, after the decorrelation step, the replicas are initiated at the the position of the reference process and run for t_{phase}. The simulation clock is not advanced, and replicas are replaced as need be should they exit the well. Our implementation has the advantage that no processor sits idle.

The reader may wonder why we would want to have a distinguished reference process – why not relaunch the reference process, as we would a replica, should it exit? We retain this feature to allow for realizations where the process is in a well for a very short period, far less than the decorrelation time. These correlated events, such as *recrossings*, appear in serial simulations and should be preserved. One may also ask why we discard the reference process. This is to simplify the analysis, as it permits us to declare that the N replicas are drawn from the same distribution when the parallel step begins.

In addition to the choice of t_{corr} and t_{phase}, there is also the question of what μ^0_{phase} should be. Again, there is significant flexibility. One possibility is to allow the reference process to evolve for some amount of time, and then the replicas could be launched from its position. A method used in practice is to find a local minima associated with the well, and initiate the replicas from that position, [21]. We emphasize that the dephasing mechanism need not depend on any information associated with the reference process.

In principle, ParRep offers a nearly linear speedup with the number of independent replicas, provided t_{corr} is short relative to the typical exit time. With the explosion in the availability of distributed computing clusters, parallel replica dynamics is an attractive tool for studying infrequent event processes.

1.2. Main results

The essential aspects of a process undergoing infrequent transitions are

- How often does it transition from one state to another?
- What state does it transition to?

These properties are captured in $\mathcal{S}_t$. To assess how well $\mathcal{S}_t^{\text{ParRep}}$ approximates it, we are motivated to first consider the exit distribution of a process, and how well it is preserved. In [4], the authors proposed a rigorous framework in which to study ParRep. The purpose of this study is to unify those ideas and assess the total error, over a single cycle of ParRep, as a function of the parameters.

Note. For brevity, we shall now take $t_{\text{sim}} = 0$ and $W_j = W$. Throughout our paper, we shall assume:

- $W \subset \mathbb{R}^n$ is bounded;
- ∂W is sufficiently smooth;

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

- V is sufficiently smooth on $\overline{W}$.

Though W need not correspond to a basin of attraction, we shall continue to call it a well.

To motivate our results, we introduce some important objects. Let μ_t denote the law of X_t, conditioned on having not left the well:

$$\mu_t(A) = \mathbb{P}^{\mu_0}\left[X_t \in A \mid T > t\right] = \frac{\mathbb{P}^{\mu_0}\left[X_t \in A, T > t\right]}{\mathbb{P}^{\mu_0}\left[T > t\right]}. \tag{1.7}$$

The above expression is the probability of finding the processes, X_t, in the set $A \subset W$, at time t, conditioned on the exit time from the well, T, being beyond t, and X_0 being initially distributed by μ_0. Additional details on our notation are given below, in Section 1.4. Under certain assumptions, the limit

$$\lim_{t\to\infty} \mu_t = \nu, \tag{1.8}$$

exists. ν is the *quasistationary distribution* (QSD) and characterizes the long term survivors of (1.1) in well W. The properties of ν are reviewed for the reader in Section 2.

In the following theorems, we shall refer to "admissible distributions." This class is quite broad and includes the Dirac distribution. It is defined and explored in subsequent sections. First, we have the following result on the convergence of the exit distribution of X_t.

Theorem 1.1 (Convergence to the QSD)**.** *Assume μ_0 is admissible. There exist positive constants $\lambda_2 > \lambda_1$, C and $\underline{t}$, such that for all $t \geq \underline{t}$ and bounded and measurable $f(\tau, \xi) : \mathbb{R}^+ \times \partial W \to \mathbb{R}$ we have*

$$\left|\mathbb{E}^{\mu_t}\left[f(T, X_T)\right] - \mathbb{E}^{\nu}\left[f(T, X_T)\right]\right| \leq C\|f\|_{L^\infty} \mathrm{e}^{-(\lambda_2-\lambda_1)t}.$$

The constant C is independent of t and f.

Taking t sufficiently large so as to make this small corresponds to the satisfactory completion of the decorrelation step; this reflects (1.8). We give a more precise statement of this theorem at the beginning of Section 3, after introducing some additional notation in Section 2. This result also plays a role in studying the dephasing step. The constants C and $\underline{t}$ depend on μ_0, V, and the geometry of the well. We will use the notation C_{phase} and C_{corr}, and $\underline{t}_{\mathrm{phase}}$ and $\underline{t}_{\mathrm{corr}}$ to distinguish the constants induced by the dephasing and decorrelation steps.

The next result ensures that the dephasing step terminates successfully:

Theorem 1.2 (Dephasing process)**.** *For an admissible distribution μ^0_{phase} and $t_{\mathrm{phase}} \geq \underline{t}_{\mathrm{phase}}$:*

A. Dephasing produces N independent replicas with distributions μ_{phase};
B. Given any $\epsilon > 0$, by taking $t_{\mathrm{phase}} \geq \underline{t}_{\mathrm{phase}}$,

$$\left|\mathbb{E}^{\mu_{\mathrm{phase}}}\left[f(T^k, X^k_{T^k})\right] - \mathbb{E}^{\nu}\left[f(T, X_T)\right]\right| \leq C_{\mathrm{phase}} \mathrm{e}^{-(\lambda_2-\lambda_1)t_{\mathrm{phase}}} \|f\|_{L^\infty}$$

C. The expected number of times a replica is relaunched is finite.

Next, the error in the parallel step cascading from the dephasing step can be controlled:

Theorem 1.3 (Parallel error)**.** *Given $t_{\mathrm{phase}} \geq \underline{t}_{\mathrm{phase}}$, let*

$$\epsilon_{\mathrm{phase}} \equiv C_{\mathrm{phase}} \mathrm{e}^{-(\lambda_2-\lambda_1)t_{\mathrm{phase}}},$$

and assume the dephasing step has produced N i.i.d. replicas drawn from distribution μ_{phase}.

Then the exit time converges to an exponential law, with parameter $N\lambda_1$,

$$\left|\mathbb{P}^{\mu_{\mathrm{phase}}}\left[T^\star > t\right] - \mathrm{e}^{-N\lambda_1 t}\right| \leq N\epsilon_{\mathrm{phase}}(1+\epsilon_{\mathrm{phase}})^{N-1}\mathrm{e}^{-N\lambda_1 t}.$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

If we additionally assume that $N\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1} < 1$*, then the hitting point distribution is asymptotically independent of the exit time*

$$\left|\mathbb{P}^{\mu_{\text{phase}}}\left[X^{\star}_{T^{\star}} \in A \mid T^{\star} > t\right] - \int_A \mathrm{d}\rho\right| \lesssim \frac{N^2\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}{1 - N\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}},$$

where ρ *is the hitting point density and* $A \subset \partial W$*.*

Thus, for t_{phase} large enough, we achieve the ideal factor of N speedup and we do not disrupt the hitting point distribution too much. The reader may find the N dependence in the error terms to be disconcerting, but it can easily be controlled by taking $t_{\text{phase}} \gtrsim \log N/(\lambda_2 - \lambda_1)$. We will return to this in the discussion. We also note that there is a slight abuse of notation in the above expressions. The superscripts, ν and μ_{phase}, should be interpreted as N-tensor products, with a distinct realization drawn for each replica.

A more detailed statement of this theorem, with explicit constants, is given at the beginning of Section 5. The hitting point density ρ is defined by (2.14).

However, Theorem 1.3 is only a comparison between the parallel step and the QSD. Our final result is a comparison between the ParRep algorithm, including decorrelation, dephasing and parallel steps, with an unaccelerated, serial process:

Theorem 1.4 (ParRep error)**.** *Let* X^{s}_t *denote the unaccelerated (serial) process and* X^{p}_t *denote the ParRep process, and let both the serial process and the reference process be initially distributed under* μ_0*, an admissible distribution. Furthermore, assume the replicas are initialized from* μ^0_{phase}*, also an admissible distribution.*

Given $t_{\text{corr}} \geq \underline{t}_{\text{corr}}$ *and* $t_{\text{phase}} \geq \underline{t}_{\text{phase}}$*, let*

$$\epsilon_{\text{corr}} = C_{\text{corr}}\mathrm{e}^{-(\lambda_2-\lambda_1)t_{\text{corr}}},$$
$$\epsilon_{\text{phase}} = C_{\text{phase}}\mathrm{e}^{-(\lambda_2-\lambda_1)t_{\text{phase}}}.$$

Letting T^{s} *and* T^{p} *denote the physical exit times, we have*

$$\begin{aligned}&\left|\mathbb{P}^{\mu_0}\left[T^{\text{s}} > t\right] - \mathbb{P}^{\mu_0}\left[T^{\text{p}} > t\right]\right| \\ &\quad\lesssim \left[\epsilon_{\text{corr}} + N\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}\right]\mathrm{e}^{-\lambda_1(t-t_{\text{corr}})_+}.\end{aligned}$$

If, in addition, t_{corr} *is sufficiently large such that* $\epsilon_{\text{corr}} < 1$*, then for* $A \subset \partial W$*,*

$$\begin{aligned}&\left|\mathbb{P}^{\mu_0}\left[X^{\text{s}}_{T^{\text{s}}} \in A \mid T^{\text{s}} > t\right] - \mathbb{P}^{\mu_0}\left[X^{\text{p}}_{T^{\text{p}}} \in A \mid T^{\text{p}} > t\right]\right| \\ &\quad\lesssim \frac{\epsilon_{\text{corr}} + N^2\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}{1-\epsilon_{\text{corr}}}\end{aligned}$$

Thus, over a single cycle, the error in ParRep can be approximately decomposed as

$$\text{Error} = \text{Decorrelation error} + \text{Parallel error}(\text{Dephasing error}), \tag{1.9}$$

where we view the parellel error as a function of the dephasing error. The speedup can be seen when T^{p} is given further consideration. When $T^{\text{p}} > t_{\text{corr}}$, $T^{\text{p}} = NT^{\star} + t_{\text{corr}}$ where $T^{\star}$ is the exit time of the particular replica which escapes first. There will be no speedup if the exit is before t_{corr}.

1.3. Outline of the Paper

In Section 2, we review some important results for (1.1). Our main Theorems are proven in Sections 3, 4, and 5. We then discuss our results in Section 6. Some additional calculations appear in the appendix.

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

1.4. Notation

Random variables, such as the position, X_t, and the exit time from the well, T, will appear in capital letters. Deterministic values, such as x, t, t_{corr}, *etc.* will be lower case. We will frequently use indicator functions in our analysis, which we write as 1_A, with A indicating the set on which the value is one.

We are often interested in probabilities and expectations of solutions of X_t solving (1.1), and its exit time T from some region W. When we write

$$\mathbb{E}^x\left[f(T, X_T)\right] \text{ or } \mathbb{P}^x\left[T \geq t\right] = \mathbb{E}^x\left[1_{T \geq t}\right]$$

the superscript x indicates that x is the initial condition of X_t; $X_0 = x$, and the expectation and probability are then taken with respect to the underlying Wiener measure of B_t.

When X_0 is given by some distribution μ_0 over W, we write

$$\mathbb{E}^{\mu_0}\left[f(T, X_T)\right] \equiv \int_W \mathbb{E}^x\left[f(T, X_T)\right] \mathrm{d}\mu_0(x).$$

When we write a conditional expectation with respect to distribution μ_0, we mean

$$\mathbb{E}^{\mu_0}\left[f(T, X_T) \mid T > t\right] \equiv \frac{\mathbb{E}^{\mu_0}\left[f(T, X_T) 1_{T>t}\right]}{\mathbb{P}^{\mu_0}\left[T > t\right]}.$$

For the reader more accustomed to the computational physics literature,

$$\mathbb{E}^{\mu_0}\left[\mathcal{O}(X_t)\right] = \left\langle \mathcal{O}(t) \right\rangle.$$

It is helpful to explicitly include the starting distribution, μ_0 associated with the process X_t, to avoid any ambiguity.

When we write $f \lesssim g$, we mean that there exists a constant $C > 0$ such that $f \leq Cg$, but that the constant is not noteworthy.

2. Preliminary results

Before proceeding to our main results on ParRep, we review some important results on the overdamped Langevin equation. These results are where our regularity assumptions on V, W, and ∂W are needed.

Two essential tools in our study of (1.1) are the Feynman–Kac formula and the quasistationary distribution, which we briefly review here; see [4] for additional details. First, let us recall the Feynman–Kac formula which relates solutions of a parabolic equation with corresponding elliptic operator

$$L \equiv -\nabla V \cdot \nabla + \beta^{-1} \Delta \tag{2.1}$$

to solutions of (1.1).

Proposition 2.1 (Prop. 1 of [4]). *On the parabolic domain $W \times \mathbb{R}^+$, let v solve*

$$\partial_t v = Lv, \tag{2.2a}$$
$$v\,|_{\partial W} = \phi : \partial W \to \mathbb{R}, \tag{2.2b}$$
$$v(t = 0) = v_0 : W \to \mathbb{R}. \tag{2.2c}$$

Then,

$$v(t, x) = \mathbb{E}^x\left[1_{T \leq t} \phi(X_T)\right] + \mathbb{E}^x\left[1_{T > t} v_0(X_t)\right]. \tag{2.3}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

To say a bit more about the elliptic operator L, recall the invariant measure of (1.1):

$$\mathrm{d}\mu \equiv Z^{-1} \exp\left(-\beta V(x)\right) \mathrm{d}x, \tag{2.4}$$

where Z is the appropriate normalization. We introduce the Hilbert space L^2_μ, with inner product

$$\langle f, g \rangle_\mu \equiv \int f g \mathrm{d}\mu. \tag{2.5}$$

An elementary calculation shows that L is self adjoint and negative definite with respect to this inner product when supplemented with homogeneous Dirichlet boundary conditions on ∂W. Standard functional analysis and elliptic theory tell us that L has infinitely many eigenvalue/eigenfunction pairs (λ_k, u_k); the eigenvalues can be ordered

$$0 > -\lambda_1 > -\lambda_2 \geq -\lambda_3 \geq \ldots;$$

and the eigenfunctions form a complete orthonormal basis for $L^2_\mu(W)$. In addition, the ground state, u_1, is unique and positive. For details, see, for example, [13, 14, 16]. The λ_1 and λ_2 appearing in our theorems are precisely the first two eigenvalues.

When solving (2.2) with $\phi = 0$, the solution can be expressed as

$$v(x,t) = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda_k t} \langle v_0, u_k \rangle_\mu u_k. \tag{2.6}$$

Out of this spectral problem, we build the norm

$$\|f\|^2_{H^s_\mu} \equiv \sum_{k=1}^{\infty} \lambda_k^s \left| \langle f, u_k \rangle_\mu \right|^2. \tag{2.7}$$

This generalizes to measures

$$\|\mu_0\|^2_{H^s_\mu} \equiv \sum_{k=1}^{\infty} \lambda_k^s \left| \int u_k \mathrm{d}\mu_0 \right|^2, \tag{2.8}$$

and to sequences, $\mathbf{a} = (a_1, a_2, \ldots)$

$$\|\mathbf{a}\|^2_{H^s_\mu} \equiv \sum_{k=1}^{\infty} \lambda_k^s |a_k|^2. \tag{2.9}$$

If μ_0 has an Radon–Nikodym derivative with respect to μ, (2.7) and (2.8) agree. We then define the function spaces,

$$H^s_\mu = \left\{ v \in \mathscr{S}(W)' \mid \|v\|_{H^s_\mu} < \infty \right\}, \tag{2.10}$$

where $\mathscr{S}$ is the set of smooth functions with support in W, and $\mathscr{S}'$ is its dual. We also define the projection operator, $P_{\mathcal{I}}$, where $\mathcal{I} \subset \mathbb{N}$,

$$P_{\mathcal{I}} f = \sum_{k \in \mathcal{I}} \langle f, u_k \rangle_\mu u_k. \tag{2.11}$$

Having introduced these spaces and norms, we can now clarify what was meant by the term *admissible distribution* used in the introduction. In this work, a distribution will be admissible with respect to W if $\operatorname{supp} \mu_0 \subset W$, and for some $s \geq 0$, $\|\mu_0\|_{H^{-s}_\mu} < \infty$.

The aforementioned quasistationary distribution of (1.1) associated with the set W is closely related to the spectral structure of L. The QSD, ν, is a time independent probability measure satisfying, for all measurable $A \subset W$ and $t > 0$:

$$\nu(A) = \frac{\int_W \mathbb{P}^x \left[X_t \in A,\ t < T\right] \mathrm{d}\nu}{\int_W \mathbb{P}^x \left[t < T\right] \mathrm{d}\nu} = \mathbb{P}^\nu \left[X_t \in A \mid t < T\right]. \tag{2.12}$$

The QSD measure ν exists and

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

Proposition 2.2 (Prop. 2 of [4])**.**

$$\mathrm{d}\nu = \frac{u_1 \mathrm{d}\mu}{\int_W u_1 \mathrm{d}\mu} = \frac{u_1 \mathrm{e}^{-\beta V}\mathrm{d}x}{\int_W u_1 \mathrm{e}^{-\beta V}\mathrm{d}x}. \tag{2.13}$$

We refer the reader to, amongst others, [5–7, 19, 20, 25] for additional details on the QSD. The utility of the QSD stems from the property that if X_0 is distributed according to ν, then:

Proposition 2.3 (Prop. 3 of [4])**.** *Let $\phi : \partial W \to \mathbb{R}$ be smooth. Then for $t > 0$*

$$\mathbb{E}^\nu \left[1_{T<t}\phi(X_T)\right] = \mathbb{P}^\nu \left[T < t\right] E^\nu \left[\phi(X_T)\right] = (1 - \mathrm{e}^{-\lambda_1 t}) \int_{\partial_W} \phi \, \mathrm{d}\rho$$

where the exit density is given by

$$\mathrm{d}\rho = -\frac{1}{\lambda_1 \beta} \nabla \frac{\mathrm{d}\nu}{\mathrm{d}x} \cdot \mathbf{n} \, dS_x = -\frac{\nabla(u_1 \mathrm{e}^{-\beta V}) \cdot \mathbf{n}}{\lambda_1 \beta \int_W u_1 \mathrm{e}^{-\beta V}\mathrm{d}x} \, dS_x, \tag{2.14}$$

with $\mathbf{n}$ *the outward pointing normal and* dS_x *the surface measure.*

In words, T is exponentially distributed with parameter λ_1, and the first hitting point is independent of the first hitting time. Being initially distributed according to ν is, in a sense, ideal. As shown by Proposition 5 of [4], were this the case for the replicas, the parallel step of ParRep would be exact. In practice, X_0 is never distributed by ν, and it is the propagation of this error that we explore.

Many of these quantities can be reformulated in terms of the Fokker–Planck equation for density $p^x(t, y)$, $x \in W$,

$$\begin{aligned} \partial_t p^x = L^* p^x = \nabla_y \cdot \left(p^x \nabla V + \beta^{-1} \nabla p^x\right), \\ p^x|_{\partial W} = 0, \quad p_0^x = \delta_x(y). \end{aligned}$$

Though we will not make use of this, the reader more accustomed to Fokker–Planck may find it helpful to re-express various quantities in terms of p^x. With regard to exit distributions,

$$\begin{aligned} \mathbb{E}^x \left[\phi(X_T) 1_{T<t}\right] &= \int_0^t \int_{\partial W} -\phi(y) \beta^{-1} \nabla p^x \cdot \mathbf{n} \, \mathrm{d}S_y, \\ \mathbb{P}^x \left[t < T\right] &= \int_t^\infty \int_{\partial W} -\beta^{-1} \nabla p^x \cdot \mathbf{n} \mathrm{d}S_y = \int_W p^x(t, y) \, \mathrm{d}y. \end{aligned}$$

These can be integrated against the density of the QSD, $\frac{\mathrm{d}\nu}{\mathrm{d}y}$, which solves $L^* \frac{\mathrm{d}\nu}{\mathrm{d}x} = -\lambda_1 \frac{\mathrm{d}\nu}{\mathrm{d}y}$, to obtain

$$p^\nu(y, t) = \mathrm{e}^{-\lambda_1 t} \frac{\mathrm{d}\nu}{\mathrm{d}y}$$

as a particular solution of the Fokker–Planck equation. This directly shows the independence of exit time and hitting point. Substituting into the above integrals reproduces Proposition 2.3.

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

3. Convergence to the QSD – Proof of Theorem 1.1

In this section we prove Theorem 1.1, which we first restate with more detail:

Theorem 3.1 (Convergence to the QSD)**.** *Given $s \geq 0$, let μ_0 be a distribution with* $\operatorname{supp}\mu_0 \subset W$ *and* $\|\mu_0\|_{H_\mu^{-s}} < \infty$. *There exists*

$$\underline{t} \gtrsim \left\{ \left\| P_{[2,\infty)}\mu_0 \right\|_{H_\mu^{-s}} / \int u_1 \mathrm{d}\mu_0 \right\}^{4/(n+2s)} \tag{3.1}$$

such that for all $t \geq \underline{t}$ and for all bounded and measurable $f(\tau,\xi) : \mathbb{R}^+ \times \partial W \to \mathbb{R}$

$$\begin{aligned} &\left|\mathbb{E}^{\mu_t}\left[f(T,X_T)\right] - \mathbb{E}^{\nu}\left[f(T,X_T)\right]\right| \\ &\lesssim \|f\|_{L^\infty} \left(\int u_1 \mathrm{d}\mu_0\right)^{-1} \underline{t}^{-n/4-s/2} \mathrm{e}^{-(\lambda_2-\lambda_1)(t-\underline{t})} \left\| P_{[2,\infty)}\mu_0 \right\|_{H_\mu^{-s}}. \end{aligned} \tag{3.2}$$

This is a refinement of Proposition 6 from [4], which now admits initial distributions which lack an L^2 Radon–Nikodym derivative. Indeed, for appropriate s, μ_0 can be a Dirac distribution. Though this is a parabolic flow which will instantaneously regularize such rough data, it is essential to an analysis of ParRep as one often wants to use Dirac mass initial conditions.

In addition to this result, we present an extension which is essential to obtaining the results in Section 5 on the parallel step.

3.1. Proof of Theorem 3.1

Proof. We first write

$$\mathbb{E}^{\mu_t}\left[f(T,X_T)\right] = \int_W \mathbb{E}^x\left[f(T,X_T)\right]\mathrm{d}\mu_t = \int_W F(x)\mathrm{d}\mu_t$$

where we have defined $F(x) \equiv \mathbb{E}^x\left[f(T,X_T)\right]$. Thus,

$$\mathbb{E}^{\mu_t}\left[f(T,X_T)\right] = \frac{\int_W \mathbb{E}^x\left[F(X_t)1_{T>t}\right]\mathrm{d}\mu_0}{\int_W \mathbb{E}^x\left[1_{T>t}\right]\mathrm{d}\mu_0}. \tag{3.3}$$

Applying Feynman–Kac, (2.3), to this,

$$\mathbb{E}^{\mu_t}\left[f(T,X_T)\right] = \frac{\int_W v(t,x)\mathrm{d}\mu_0}{\int_W \bar{v}(t,x)\mathrm{d}\mu_0} \tag{3.4}$$

where v solves (2.2) with $v_0 = F$ and $\phi = 0$, while $\bar{v}$ solves it with $v_0 = 1$ and $\phi = 0$. For brevity, let

$$\hat{F}_k = \int F u_k \mathrm{d}\mu, \quad \hat{1}_k = \int u_k \mathrm{d}\mu, \quad \hat{\mu}_{0,k} = \int u_k \mathrm{d}\mu_0. \tag{3.5}$$

Expressing v and $\bar{v}$ as series solutions using (2.6), we have

$$v(t,x) = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda_k t}\hat{F}_k u_k(x), \quad \bar{v}(t,x) = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda_k t}\hat{1}_k u_k(x). \tag{3.6}$$

After a bit of rearrangement, the error can be expressed as

$$\begin{aligned} e(t) &\equiv \left|\mathbb{E}^{\mu_t}\left[f(T,X_T)\right] - \mathbb{E}^{\nu}\left[f(T,X_T)\right]\right| \\ &= \left| \frac{\sum_k \mathrm{e}^{-(\lambda_k-\lambda_1)t}\left(\hat{F}_k - \hat{1}_k \int F\mathrm{d}\nu\right)\hat{\mu}_{0,k}}{\hat{1}_1\hat{\mu}_{0,1} + \sum_k \mathrm{e}^{-(\lambda_k-\lambda_1)t}\hat{1}_k\hat{\mu}_{0,k}} \right| \end{aligned} \tag{3.7}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

where the sums are from $k=2$ to ∞ since $\hat{F}_1 = \hat{1}_1 \int_W F \mathrm{d}\nu$. Noting that

$$\begin{aligned}\left|\hat{F}_k - \hat{1}_k \int F \mathrm{d}\nu\right| &\leq \int |Fu_k| \,\mathrm{d}\mu + \int |F| \,\mathrm{d}\nu \int |u_k| \,\mathrm{d}\mu \\ &\leq 2 \|f\|_{L^\infty} \int |u_k| \,\mathrm{d}\mu \leq 2 \|f\|_{L^\infty} \sqrt{\mu(W)},\end{aligned}$$

we can rewrite the numerator as

$$\begin{aligned}\left|\sum_{k=2}^{\infty} \mathrm{e}^{-(\lambda_k-\lambda_1)t} \left(\hat{F}_k - \hat{1}_k \int F \mathrm{d}\nu\right) \hat{\mu}_{0,k}\right| &\leq 2\sqrt{\mu(W)} \|f\|_{L^\infty} \sum_{k=2}^{\infty} \mathrm{e}^{-(\lambda_k-\lambda_1)t} |\hat{\mu}_{0,k}| \\ &\leq 2\sqrt{\mu(W)} \|f\|_{L^\infty} \mathrm{e}^{-(\lambda_2-\lambda_1)(t-t_1)} \sum_{k=2}^{\infty} \mathrm{e}^{-(\lambda_k-\lambda_1)t_1} |\hat{\mu}_{0,k}| \\ &\leq 2\sqrt{\mu(W)} \|f\|_{L^\infty} \mathrm{e}^{-(\lambda_2-\lambda_1)(t-t_1)} \sum_{k=2}^{\infty} \mathrm{e}^{-\kappa\lambda_k t_1} |\hat{\mu}_{0,k}|\end{aligned}$$

where $\kappa = 1 - \lambda_1/\lambda_2$ and $t \geq t_1 > 0$. Applying Proposition A.2 from the appendix to this, the numerator is bounded by

$$\begin{aligned}&\left|\sum_{k=2}^{\infty} \mathrm{e}^{-(\lambda_k-\lambda_1)t} \left(\hat{F}_k - \hat{1}_k \int F \mathrm{d}\nu\right) \hat{\mu}_{0,k}\right| \\ &\quad\lesssim \left\|P_{[2,\infty)}\mu_0\right\|_{H_\mu^{-s}} \|f\|_{L^\infty} \mathrm{e}^{-(\lambda_2-\lambda_1)(t-t_1)} t_1^{-n/4-s/2}. \end{aligned} \tag{3.8}$$

The constant that has been absorbed into the $\lesssim$ symbol is independent of t, f and μ_0.

To ensure the denominator is uniformly bounded away from zero, we use a similar treatment,

$$\begin{aligned}\sum_{k=2}^{\infty} \mathrm{e}^{-(\lambda_k-\lambda_1)t} \hat{1}_k \hat{\mu}_{0,k} &\leq \sqrt{\mu(W)} \mathrm{e}^{-(\lambda_2-\lambda_1)(t-t_2)} \sum_{k=2}^{\infty} \mathrm{e}^{-\kappa\lambda_k t_2} |\hat{\mu}_{0,k}| \\ &\lesssim \left\|P_{[2,\infty)}\mu_0\right\|_{H_\mu^{-s}} \mathrm{e}^{-(\lambda_2-\lambda_1)t} t_2^{-n/4-s/2}\end{aligned}$$

for $t \geq t_2 > 0$, which may differ from t_1. Therefore,

$$\begin{aligned}&\hat{1}_1\hat{\mu}_{0,1} + \sum_{k=2}^{\infty} \mathrm{e}^{-(\lambda_k-\lambda_1)t} \hat{1}_k \hat{\mu}_{0,k} \\ &\quad\gtrsim \hat{1}_1\hat{\mu}_{0,1} - \mathrm{e}^{-(\lambda_2-\lambda_1)(t-t_2)} t_2^{-n/4-s/2} \left\|P_{[2,\infty)}\mu_0\right\|_{H_\mu^{-s}}.\end{aligned}$$

For a sufficiently large $t \geq \underline{t} \geq t_2 > 0$, the denominator is bounded from below by

$$\hat{1}_1\hat{\mu}_{0,1} + \sum_{k=2}^{\infty} \mathrm{e}^{-(\lambda_k-\lambda_1)t} \hat{1}_k \hat{\mu}_{0,k} \geq \frac{1}{2}\hat{1}_1\hat{\mu}_{0,1} = \frac{1}{2} \int u_1 \mathrm{d}\mu \int u_1 \mathrm{d}\mu_0 > 0. \tag{3.9}$$

Roughly,

$$\underline{t} \gtrsim \left\{\left\|P_{[2,\infty)}\mu_0\right\|_{H_\mu^{-s}} / \int u_1 \mathrm{d}\mu_0\right\}^{4/(n+2s)}. \tag{3.10}$$

Taking $t_1 = t_2 = \underline{t}$ in (3.8) and (3.9) we have that for $t \geq \underline{t}$

$$\begin{aligned}e(t) \lesssim &\left(\int u_1 \mathrm{d}\mu_0\right)^{-1} \mathrm{e}^{-(\lambda_2-\lambda_1)(t-\underline{t})} (\underline{t})^{-n/4-s/2} \\ &\times \|f\|_{L^\infty} \left\|P_{[2,\infty)}\mu_0\right\|_{H_\mu^{-s}}.\end{aligned}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

Finally, for this estimate to hold for general bounded and measurable f, we apply a density argument with respect to the L^∞ norm. □

The inclusion of $\int u_1 \mathrm{d}\mu_0$ in the preceding result is deliberate as μ_0 is, to a degree, a user specified parameter. Moreover, $\int u_1 \mathrm{d}\mu_0$ could be quite small. Indeed, when a X_t first enters W, it is near ∂W and the support of μ_0 is in a neighborhood of ∂W; we may have $\mu_0 = \delta_x$ where x is close to ∂W. As u_1 is continuous and vanishes on ∂W,

$$\int_W u_1 \delta_x = \mathrm{O}\left(\mathrm{dist}(x, \partial W)\right).$$

We also see that as $\mu_0 \to \nu$, $\left\| P_{[2,\infty)}\mu_0 \right\|_{H_\mu^{-s}} \to 0$, and the error vanishes.

It remains to identify distributions and values of s for which $\|\mu_0\|_{H_\mu^{-s}} < \infty$. In the case that μ_0 has an L^2_μ Radon–Nikodym derivative, one readily sees that $\|\mu_0\|_{H_\mu^{-s}} < \infty$ for $s \leq 0$. Indeed, when $s = 0$, this results collapses onto the L^2_μ estimate of [4]. This extends to μ_0 possessing L^p_μ densities for any $p \geq 2$.

For the case $\mu_0 = \delta_x$, a Dirac mass, we shall have that $\mu_0 \in H_\mu^{-s}$ when s is large enough to embed H^s_μ into L^∞. If the ∂W is sufficiently smooth, then by standard elliptic theory, H^s_μ and H^s will be equivalent for $s \geq 0$, and we have the embedding for $s > n/2$, [1, 13, 14]. Refined elliptic estimates may weaken such assumptions on the boundary.

3.2. Exit times

In the case that we are interested in exit times, we have a result closely related to Theorem 3.1.

Theorem 3.2. *Assume μ_0 satisfies the assumptions of Theorem 3.1 and $t_0 \geq \underline{t}$. Then for $t \geq 0$,*

$$\left|\mathbb{P}^{\mu_{t_0}}\left[T > t\right] - \mathrm{e}^{-\lambda_1 t}\right| \leq C \mathrm{e}^{-\lambda_1 t} \mathrm{e}^{-(\lambda_2 - \lambda_1)t_0} \tag{3.11}$$

where C is the pre-exponential factor in (3.2) and is independent of t and t_0.

Proof. As before, we rely on (2.3) and the series expansions (2.6) to write

$$\mathbb{P}^{\mu_{t_1}}\left[T > t\right] = \frac{\mathbb{P}^{\mu_0}[T > t_1 + t]}{\mathbb{P}^{\mu_0}[T > t_1]} = \frac{\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k (t+t_1)} \hat{1}_k \hat{\mu}_{0,k}}{\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k t_1} \hat{1}_k \hat{\mu}_{0,k}}.$$

Comparing against the QSD,

$$\begin{aligned}\left|\mathbb{P}^{\mu_{t_1}}\left[T > t\right] - \mathrm{e}^{-\lambda_1 t}\right| &= \left| \frac{\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k (t+t_1)} \hat{1}_k \hat{\mu}_{0,k}}{\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k t_1} \hat{1}_k \hat{\mu}_{0,k}} - \mathrm{e}^{-\lambda_1 t} \right| \\ &= \left| \frac{\sum_{k=1}^\infty \left(\mathrm{e}^{-\lambda_k (t+t_1)} - \mathrm{e}^{-\lambda_k t_1 - \lambda_1 t}\right) \hat{1}_k \hat{\mu}_{0,k}}{\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k t_1} \hat{1}_k \hat{\mu}_{0,k}} \right|\end{aligned}$$

In the numerator, the $k = 1$ term vanishes, leaving

$$\mathrm{e}^{-\lambda_1 t} \left| \frac{\sum_{k=2}^\infty \left(1 - \mathrm{e}^{-(\lambda_k - \lambda_1)t}\right) \mathrm{e}^{-\lambda_k t_1} \hat{1}_k \hat{\mu}_{0,k}}{\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k t_1} \hat{1}_k \hat{\mu}_{0,k}} \right| \leq \mathrm{e}^{-\lambda_1 t} \frac{\sum_{k=2}^\infty \mathrm{e}^{-\lambda_k t_0} \left|\hat{1}_k \hat{\mu}_{0,k}\right|}{\left|\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k t_0} \hat{1}_k \hat{\mu}_{0,k}\right|}$$

Using the same methods as in the Proof of Theorem 3.1,

$$\frac{\sum_{k=2}^\infty \mathrm{e}^{-\lambda_k t_1} \left|\hat{\mu}_{0,k}\right|}{\left|\sum_{k=1}^\infty \mathrm{e}^{-\lambda_k t_1} \hat{1}_k \hat{\mu}_{0,k}\right|} \lesssim \left(\int u_1 \mathrm{d}\mu_0\right)^{-1} \underline{t}^{-n/4 - s/2} \mathrm{e}^{-(\lambda_2 - \lambda_1)(t_0 - \underline{t})} \left\| P_{[2,\infty)}\mu_0 \right\|_{H_\mu^{-s}}$$

□

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

This estimate plays an important role in our analysis of ParRep. Indeed, we will frequently confront terms of the form $\mathbb{E}^{\mu_{t_0}}\left[f(X,T)1_{T>t}\right]$, and we will want to compare against the corresponding term for the QSD. One could naively apply Theorem 3.1 to estimate such a term, with observable $g_t(\xi,\tau) = f(\xi,\tau)1_{\tau>t}$. However, this is wasteful, as the observable is going to be taken over realizations which not only have not left the well before t_0, but remain in the well for at least an additional t. We thus have the following identity.

Lemma 3.3. *Given $t, t_0 \geq 0$,*

$$\mathbb{E}^{\mu_{t_0}}\left[f(X_T,T)1_{T>t}\right] = \mathbb{E}^{\mu_{t_0+t}}\left[f(X_T,T+t)\right]\mathbb{P}^{\mu_{t_0}}\left[T>t\right]. \tag{3.12}$$

Proof. This reflects the Markovian nature of the process. Writing out the lefthand side,

$$\begin{aligned}\mathbb{E}^{\mu_{t_0}}\left[f(X_T,T)1_{T>t}\right] &= \int_W \mathbb{E}^x\left[f(X_T,T)1_{T>t}\right]\mu_{t_0}(\mathrm{d}x)\\ &= \frac{\int_W \mathbb{E}^x\left[f(X_T,T)1_{T>t}\right]\mathbb{P}^{\mu_0}\left[X_t\in \mathrm{d}x, T>t_0\right]}{\mathbb{P}^{\mu_0}\left[T>t_0\right]}\end{aligned}$$

The numerator is

$$\begin{aligned}\int_W \mathbb{E}^x\left[f(X_T,T)1_{T>t}\right]\mathbb{P}^{\mu_0}\left[X_t\in\mathrm{d}x, T>t_0\right] &= \mathbb{E}^{\mu_0}\left[f(X_T,T-t_0)1_{T-t_0>t}1_{T>t_0}\right]\\ &= \mathbb{E}^{\mu_0}\left[f(X_T,T-t_0)1_{T>t_0+t}\right],\end{aligned}$$

where t_0 is subtracted off to make the observable consistent. The same argument shows

$$\mathbb{E}^{\mu_{t_0+t}}\left[f(X_T,T+t)\right] = \frac{\mathbb{E}^{\mu_0}\left[f(X_T,T-t_0)1_{T>t+t_0}\right]}{\mathbb{P}^{\mu_0}\left[T>t+t_0\right]}.$$

Combining these three expressions completes the proof. □

In principle, we can use this lemma and Theorem 3.2 to obtain refinements on Theorem 3.1 for observables that include $1_{T>t}$ terms.

4. The dephasing step – Proof of Theorem 1.2

We now examine our dephasing step,

Theorem 4.1. *Given $s \geq 0$, assume* $\operatorname{supp}\mu^0_{\mathrm{phase}} \subset W$ *and* $\left\|\mu^0_{\mathrm{phase}}\right\|_{H^{-s}_{\mu}} < \infty$. *Then*

A. The dephasing step produces N independent replicas with distributions μ_{phase},

$$\mu_{\mathrm{phase}}(A) = \mathbb{P}^{\mu^0_{\mathrm{phase}}}\left[X_{t_{\mathrm{phase}}}\in A \mid T>t_{\mathrm{phase}}\right];$$

B. There exists $\underline{t}_{\mathrm{phase}}$ and C_{phase} such that for $t_{\mathrm{phase}} \geq \underline{t}_{\mathrm{phase}}$,

$$\left|\mathbb{E}^{\mu_{\mathrm{phase}}}\left[f(T^k,X^k_{T^k})\right] - \mathbb{E}^{\nu}\left[f(T,X_T)\right]\right| \leq \|f\|_{L^\infty} C_{\mathrm{phase}}\mathrm{e}^{-(\lambda_2-\lambda_1)t_{\mathrm{phase}}};$$

for any bounded measurable $f:\mathbb{R}^+\times\partial W\to\mathbb{R}$ and all $k=1,\ldots,N$.

C. The expected number of times a replica is relaunched is finite.

To prove Theorem 4.1, we must establish:

A. The replicas are independent and have law μ_{phase};
B. The error of μ_{phase} can be made small;
C. The expected number of relaunches is finite.

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

The first property is obvious as each of the replicas is driven by an independent Brownian motion, and we only retain realizations for which $T > t_{\text{phase}}$. The second property follows from Theorem 3.1.

To prove the third property, we must establish that replicas initiated from μ^0_{phase} have a nonzero chance of surviving till t_{phase}:

Lemma 4.2. *Assume that μ^0_{phase} satisfies the hypotheses of Theorem 4.1,*

$$\mathbb{P}^{\mu^0_{\text{phase}}}\left[T^k \geq t_{\text{phase}}\right] \equiv p > 0.$$

Proof. Observe that we have the following monotonicity property for $t_2 > t_1$,

$$0 \leq \mathbb{P}^{\mu^0_{\text{phase}}}\left[T \geq t_2\right] \leq \mathbb{P}^{\mu^0_{\text{phase}}}\left[T \geq t_1\right].$$

We now argue by contradiction. Assume that at some $t_1 > 0$, $\mathbb{P}^{\mu^0_{\text{phase}}}\left[T \geq t_1\right] = 0$. By the above monotonicity, $\mathbb{P}^{\mu^0_{\text{phase}}}\left[T \geq t_2\right] = 0$ for all $t_2 \geq t_1$. Using a similar approach as in the Proof of Theorem 3.1, we write

$$\mathbb{P}^{\mu^0_{\text{phase}}}\left[T \geq t\right] = \bar{v}(x,t) = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda_k t} \int u_k d\mu^0_{\text{phase}} \int u_k \mathrm{d}\mu$$

where $\bar{v}$ solves (2.2) with $v_0 = 1$ and $\phi = 0$. Therefore,

$$\begin{aligned}
\bar{v}(x,t_2) &= \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda_k t_2} \hat{\mu}^0_{\text{phase},k} \int u_k \mathrm{d}\mu = \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda_k t_2} \hat{\mu}^0_{\text{phase},k} \hat{1}_k \\
&\geq \mathrm{e}^{-\lambda_1 t_2} \left\{ \hat{\mu}^0_{\text{phase},1} \hat{1}_1 - \mathrm{e}^{-(\lambda_2-\lambda_1)t_2} \sum_{k=2}^{\infty} \mathrm{e}^{-\kappa\lambda_k t_2} \left|\hat{1}_k\right| \left|\hat{\mu}^0_{\text{phase},k}\right| \right\} \\
&\gtrsim \mathrm{e}^{-\lambda_1 t_2} \left\{ \hat{\mu}^0_{\text{phase},1} \hat{1}_1 - \mathrm{e}^{-(\lambda_2-\lambda_1)t_2} t_2^{-n/4-s/2} \left\| P_{[2,\infty)} \mu^0_{\text{phase}} \right\|_{H_\mu^{-s}} \right\}
\end{aligned}$$

Then taking t_2 sufficiently large,

$$\begin{aligned}
&\hat{\mu}^0_{\text{phase},1} \hat{1}_1 - \mathrm{e}^{-(\lambda_2-\lambda_1)t_2} t_2^{-n/4-s/2} \left\| P_{[2,\infty)} \delta_x \right\|_{H_\mu^{-s}} \\
&\geq \frac{1}{2} \hat{\mu}^0_{\text{phase},1} \hat{1}_1 = \frac{1}{2} \int u_1 \mathrm{d}\mu^0_{\text{phase}} \int u_1 \mathrm{d}\mu > 0,
\end{aligned}$$

since $\int u_1 \mathrm{d}\mu^0_{\text{phase}} > 0$. Thus, we have a contradiction. □

This calculation reveals a role played by the choice of μ^0_{phase}. If concentrated near the well boundary, $\mathbb{P}^{\mu^0_{\text{phase}}}\left[T^k \geq t_{\text{phase}}\right] = p$ could be quite small. This will induce the replicas to relaunch many times, as the next result shows. Thus, for computational efficiency, a distribution concentrated deep in the well's interior is desirable.

Lemma 4.3. *Assume that μ^0_{phase} satisfies the hypotheses of Theorem 4.1, and that $\mathbb{P}^{\mu^0_{\text{phase}}}\left[T^k \geq t_{\text{phase}}\right] = p > 0$. Then*

$$\mathbb{E}^{\mu^0_{\text{phase}}}\left[\textit{Number of relaunches}\right] = (1-p)/p < \infty$$

Proof. The probability of relaunching m times is the probability of exiting m times and surviving on the $m+1$–th time. Interpreting this in terms of T^k and using the assumption, $\mathbb{P}^{\mu^0_{\text{phase}}}[m \text{ relaunches}] = (1-p)^m p$. Thus,

$$\begin{aligned}
\mathbb{E}^{\mu^0_{\text{phase}}}\left[\text{Number of relaunches}\right] &= \sum_{m=0}^{\infty} m \cdot \mathbb{P}^{\mu^0_{\text{phase}}}[m \text{ relaunches}] \\
&= \sum_{m=0}^{\infty} m(1-p)^m p = \frac{1-p}{p} < \infty.
\end{aligned}$$

□

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

5. The parallel step – Proofs of Theorems 1.3 and 1.4

First, we restate Theorem 1.3 with additional detail:

Theorem 5.1 (Parallel error)**.** *Given* $t_{\text{phase}} \geq \underline{t}_{\text{phase}}$*, let*

$$\epsilon_{\text{phase}} \equiv C_{\text{phase}} \mathrm{e}^{-(\lambda_2-\lambda_1)t_{\text{phase}}},$$

and assume the dephasing step has produced N *i.i.d. replicas drawn from distribution* μ_{phase}*. Then the exit time distribution of the parallel step converges to an exponential,*

$$\left|\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star > t\right] - \mathrm{e}^{-N\lambda_1 t}\right| \leq \epsilon_{\text{phase}} N (1+\epsilon_{\text{phase}})^{N-1} \mathrm{e}^{-N\lambda_1 t}. \tag{5.1}$$

If $\phi : \partial W \to \mathbb{R}$ *is bounded and measurable, the exit distribution converges to one that is independent of exit time,*

$$\begin{aligned}\left|\mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^\star>t}\phi(X^\star_{T^\star})\right] - \mathrm{e}^{-N\lambda_1 t}\int_{\partial W}\phi \mathrm{d}\rho\right| \\ \lesssim N^2(1+\epsilon_{\text{phase}})^{N-1}\epsilon_{\text{phase}}\left\|\phi\right\|_{L^\infty}\mathrm{e}^{-N\lambda_1 t}.\end{aligned} \tag{5.2}$$

If, in addition, $N\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1} < 1$*, then*

$$\begin{aligned}\left|\mathbb{E}^{\mu_{\text{phase}}}\left[\phi(X^\star_{T^\star}) \mid T^\star > t\right] - \int_{\partial W}\phi \mathrm{d}\rho\right| \\ \lesssim \frac{N^2\left\|\phi\right\|_{L^\infty}\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}{1-N\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}.\end{aligned} \tag{5.3}$$

Proof. To prove (5.1), we begin by writing,

$$\begin{aligned}\left|\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star > t\right] - \mathrm{e}^{-N\lambda_1 t}\right| &= \left|\Pi_{k=1}^N \mathbb{P}^{\mu_{\text{phase}}}\left[T^k > t\right] - \Pi_{k=1}^N \mathbb{P}^{\nu}\left[T^k > t\right]\right| \\ &= \left|\mathbb{P}^{\mu_{\text{phase}}}\left[T^1 > t\right]^N - \mathrm{e}^{-N\lambda_1 t}\right| \\ &= \left|\mathbb{P}^{\mu_{\text{phase}}}\left[T^1 > t\right] - \mathrm{e}^{-\lambda_1 t}\right| \left|\sum_{k=0}^{N-1}\mathbb{P}^{\mu_{\text{phase}}}\left[T^1 > t\right]^k \mathrm{e}^{-(N-1-k)\lambda_1 t}\right|.\end{aligned}$$

From Theorem 3.2, we know

$$\left|\mathbb{P}^{\mu_{\text{phase}}}\left[T > t\right] - \mathrm{e}^{-\lambda_1 t}\right| \leq \epsilon_{\text{phase}}\mathrm{e}^{-\lambda_1 t}.$$

Therefore,

$$\left|\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star \geq t\right] - \mathrm{e}^{-N\lambda_1 t}\right| \leq \epsilon_{\text{phase}}\mathrm{e}^{-\lambda_1 t} N(1+\epsilon_{\text{phase}})^{N-1}\mathrm{e}^{-(N-1)\lambda_1 t}.$$

To prove (5.2), we begin by writing the expectation as

$$\begin{aligned}\mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^\star>t}\phi(X^\star_{T^\star})\right] &= \mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^{k_\star}>t}\phi(X^{k_\star}_{T^{k_\star}})\right] \\ &= \sum_{k=1}^N \mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^k>t}\phi(X^k_{T^k})1_{k=k_\star}\right] \\ &= \sum_{k=1}^N \mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^k>t}\phi(X^k_{T^k})\Pi_{l\neq k}1_{T^l>T^k}1_{T^l>t}\right].\end{aligned}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

In the above expression, we have used that since $T^\star > t$, $T^l > t$ for each l. Then, using Lemma 3.3 on each of the processes,

$$\begin{aligned}\mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^k>t}\phi(X^k_{T^k})\Pi_{l\neq k}1_{T^l>T^k}1_{T^l>t}\right] &= \mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^k_{T^k})\Pi_{l\neq k}1_{T^l>T^k}\right]\Pi_{l=1}^N\mathbb{P}^{\mu_{\text{phase}}}[T^l>t]\\ &= \mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^k_{T^k})1_{k=k_\star}\right]\mathbb{P}^{\mu_{\text{phase}}}[T>t]^N.\end{aligned}$$

This leads to the expression

$$\mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^\star>t}\phi(X^\star_{T^\star})\right] = \mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^\star_{T^\star})\right]\mathbb{P}^{\mu_{\text{phase}}}[T^\star>t]. \tag{5.4}$$

Comparing against the QSD,

$$\begin{aligned}\left|\mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^\star>t}\phi(X^\star_{T^\star})\right]-\mathbb{E}^{\nu}\left[1_{T^\star>t}\phi(X^\star_{T^\star})\right]\right| &\leq \left|\mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^\star_{T^\star})\right]\right|\left|\mathbb{P}^{\mu_{\text{phase}}}[T^\star>t]-\mathrm{e}^{-N\lambda_1 t}\right|\\ &\quad+\mathrm{e}^{-N\lambda_1 t}\left|\mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^\star_{T^\star})\right]-\mathbb{E}^{\nu}\left[\phi(X^\star_{T^\star})\right]\right|.\end{aligned} \tag{5.5}$$

The first difference can be treated by (5.1), but the second difference requires more care.

Given an arbitrary admissible distribution η for X_0, we define

$$\mathcal{P}^\eta(t) \equiv \mathbb{P}^\eta[T>t] = \mathbb{P}^\eta[T^k>t], \quad k=1\dots N. \tag{5.6}$$

Consequently, $\mathcal{P}^\nu(t) = \mathrm{e}^{-\lambda_1 t}$ and

$$\begin{aligned}\mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^\star_{T^\star})\right] &= \sum_{k=1}^N \mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^k_T)\Pi_{l\neq k}1_{T_l>T_k}\right]\\ &= \sum_{k=1}^N \mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right].\end{aligned} \tag{5.7}$$

An analogous expansion can be made with ν in place of $\mu_{t_{\text{phase}}+t}$. Taking the difference of the two sums, and comparing term by term,

$$\begin{aligned}&\left|\mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]-\mathbb{E}^{\nu}\left[\phi(X^k_T)\mathcal{P}^{\nu}(T^k)^{N-1}\right]\right|\\ &\leq \left|\mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]-\mathbb{E}^{\nu}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]\right|\\ &\quad+\left|\mathbb{E}^{\nu}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]-\mathbb{E}^{\nu}\left[\phi(X^k_T)\mathcal{P}^{\nu}(T^k)^{N-1}\right]\right|.\end{aligned} \tag{5.8}$$

By Theorem 3.1 the first difference in (5.8) is bounded by

$$\begin{aligned}&\left|\mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]-\mathbb{E}^{\nu}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]\right|\\ &\quad\leq \epsilon_{\text{phase}}\mathrm{e}^{-(\lambda_2-\lambda_1)t}\left\|\phi\right\|_{L^\infty},\end{aligned}$$

since $\mathcal{P}\leq 1$.

For the other difference in (5.8), we can replicate the proof of (5.1) to obtain, for any $\tau\geq 0$,

$$\begin{aligned}&\left|\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(\tau)^{N-1}-\mathcal{P}^{\nu}(\tau)^{N-1}\right|\\ &\leq (N-1)\epsilon_{\text{phase}}\mathrm{e}^{-(\lambda_2-\lambda_1)t}(1+\epsilon_{\text{phase}}\mathrm{e}^{-(\lambda_2-\lambda_1)t})^{N-2}\mathrm{e}^{-(N-1)\lambda_1\tau}\\ &\leq (N-1)\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-2}.\end{aligned}$$

Therefore,

$$\begin{aligned}&\left|\mathbb{E}^{\nu}\left[\phi(X^k_T)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]-\mathbb{E}^{\nu}\left[\phi(X^k_T)\mathcal{P}^{\nu}(T^k)^{N-1}\right]\right|\\ &\leq \left\|\phi\right\|_{L^\infty}\epsilon_{\text{phase}}(N-1)(1+\epsilon_{\text{phase}})^{N-2}.\end{aligned}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

So (5.8) can be bounded by

$$\begin{aligned}&\left|\mathbb{E}^{\mu_{t_{\text{phase}}+t}}\left[\phi(X_T^k)\mathcal{P}^{\mu_{t_{\text{phase}}+t}}(T^k)^{N-1}\right]-\mathbb{E}^{\nu}\left[\phi(X_T^k)\mathcal{P}^{\nu}(T^k)^{N-1}\right]\right|\\&\leq\|\phi\|_{L^\infty}\,\epsilon_{\text{phase}}\left[1+(N-1)(1+\epsilon_{\text{phase}})^{N-2}\right].\end{aligned}$$

Returning to (5.5), using (5.1) to treat the first difference and the preceding calculation to treat the second, we have:

$$\begin{aligned}\left|\mathbb{E}^{\mu_{\text{phase}}}\left[1_{T^\star>t}\phi(X^\star_{T^\star})\right]-\mathbb{E}^{\nu}\left[1_{T^\star>t}\phi(X^\star_{T^\star})\right]\right|&\leq N\|\phi\|_{L^\infty}\,\epsilon_{\text{phase}}\mathrm{e}^{-N\lambda_1 t}(1+\epsilon_{\text{phase}})^{N-1}\\&\quad+N\|\phi\|_{L^\infty}\,\epsilon_{\text{phase}}\mathrm{e}^{-N\lambda_1 t}\left[1+(N-1)(1+\epsilon_{\text{phase}})^{N-2}\right]\\&\lesssim N^2\|\phi\|_{L^\infty}\,\epsilon_{\text{phase}}\mathrm{e}^{-N\lambda_1 t}(1+\epsilon_{\text{phase}})^{N-1}.\end{aligned}\tag{5.9}$$

Finally, to prove (5.3),

$$\begin{aligned}&\left|\mathbb{E}^{\mu_{\text{phase}}}\left[\phi(X^\star_{T^\star})\mid T^\star>t\right]-\mathbb{E}^{\nu}\left[\phi(X^\star_{T^\star})\mid T^\star>t\right]\right|\\&=\left|\frac{\mathbb{E}^{\mu_{\text{phase}}}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]}{\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star>t\right]}-\frac{\mathbb{E}^{\nu}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]}{\mathbb{P}^{\nu}\left[T^\star>t\right]}\right|\\&\leq\left|\mathbb{E}^{\mu_{\text{phase}}}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]-\mathbb{E}^{\nu}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]\right|\frac{1}{\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star>t\right]}\\&\quad+\left|\mathbb{E}^{\nu}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]\right|\frac{\left|\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star>t\right]-\mathbb{P}^{\nu}\left[T^\star>t\right]\right|}{\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star>t\right]\mathbb{P}^{\nu}\left[T^\star>t\right]}.\end{aligned}$$

For the first difference,

$$\begin{aligned}&\left|\mathbb{E}^{\mu_{\text{phase}}}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]-\mathbb{E}^{\nu}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]\right|\frac{1}{\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star>t\right]}\\&\lesssim\frac{N^2\|\phi\|_{L^\infty}\,\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}{1-N\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}.\end{aligned}$$

For the second difference,

$$\begin{aligned}&\left|\mathbb{E}^{\nu}\left[\phi(X^\star_{T^\star})1_{T^\star>t}\right]\right|\frac{\left|\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star>t\right]-\mathbb{P}^{\nu}\left[T^\star>t\right]\right|}{\mathbb{P}^{\mu_{\text{phase}}}\left[T^\star>t\right]\mathbb{P}^{\nu}\left[T^\star>t\right]}\\&\leq\frac{N\|\phi\|_{L^\infty}\,\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}{1-N\epsilon_{\text{phase}}(1+\epsilon_{\text{phase}})^{N-1}}.\end{aligned}$$

Combining these estimates, we have our result. □

Lastly, we prove Theorem 1.4, which we first restate with additional detail:

Theorem 5.2 (ParRep error)**.** *Let X_t^{s} denote the unaccelerated (serial) process and X_t^{p} denote the ParRep process, and assume that both X_t^{ref} and X_t^{s} are initially distributed under μ_0, an admissible distribution. Also assume that μ^0_{phase} is admissible.*

Given $t_{\text{corr}}\geq\underline{t}_{\text{corr}}$ and $t_{\text{phase}}\geq\underline{t}_{\text{phase}}$, let

$$\begin{aligned}\epsilon_{\text{corr}}&=C_{\text{corr}}\mathrm{e}^{-(\lambda_2-\lambda_1)t_{\text{corr}}},\\\epsilon_{\text{phase}}&=C_{\text{phase}}\mathrm{e}^{-(\lambda_2-\lambda_1)t_{\text{phase}}}.\end{aligned}$$

Letting T^{s} and T^{p} denote the physical times, we have

$$\begin{aligned}&\left|\mathbb{P}^{\mu_0}\left[T^{\text{s}}>t\right]-\mathbb{P}^{\mu_0}\left[T^{\text{p}}>t\right]\right|\\&\quad\leq\epsilon_{\text{corr}}\mathrm{e}^{-\lambda_1 t}+\epsilon_{\text{phase}}N(1+\epsilon_{\text{phase}})^{N-1}\mathrm{e}^{-\lambda_1(t-t_{\text{corr}})_+},\end{aligned}\tag{5.10}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

$$\begin{aligned}&|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t}\right]-\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t}\right]|\\&\quad\lesssim\left[\epsilon_{\mathrm{corr}}+\epsilon_{\mathrm{phase}}N^2(1+\epsilon_{\mathrm{phase}})^{N-1}\right]\|\phi\|_{L^\infty}\,\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}.\end{aligned}\tag{5.11}$$

If, in addition, $\epsilon_{\mathrm{corr}}<1$*, then*

$$\begin{aligned}&|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})\mid T^{\mathrm{s}}>t\right]-\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})\mid T^{\mathrm{p}}>t\right]|\\&\quad\lesssim\frac{\epsilon_{\mathrm{corr}}+\epsilon_{\mathrm{phase}}N^2(1+\epsilon_{\mathrm{phase}})^{N-1}}{1-\epsilon_{\mathrm{corr}}}\|\phi\|_{L^\infty},\end{aligned}\tag{5.12}$$

Proof. We begin by decomposing

$$\begin{aligned}\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\right]&=\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\mid T^{\mathrm{s}}\le t_{\mathrm{corr}}\right]\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}\le t_{\mathrm{corr}}\right]\\&\quad+\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\mid T^{\mathrm{s}}>t_{\mathrm{corr}}\right]\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t_{\mathrm{corr}}\right].\end{aligned}$$

We analogously decompose $\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}>t\right]$. For $t\le t_{\mathrm{corr}}$, the serial algorithm and the reference process of ParRep have the same law. Hence,

$$\begin{aligned}\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}\le t_{\mathrm{corr}}\right]&=\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}\le t_{\mathrm{corr}}\right],\\\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\mid T^{\mathrm{s}}\le t_{\mathrm{corr}}\right]&=\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}>t\mid T^{\mathrm{p}}\le t_{\mathrm{corr}}\right].\end{aligned}$$

Consequently, error only manifests itself if the parallel step is engaged,

$$\begin{aligned}&|\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\right]-\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}>t\right]|\\&=\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t_{\mathrm{corr}}\right]|\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\mid T^{\mathrm{s}}>t_{\mathrm{corr}}\right]-\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}>t\mid T^{\mathrm{p}}>t_{\mathrm{corr}}\right]|.\end{aligned}$$

Comparing against the QSD,

$$\begin{aligned}&|\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\right]-\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}>t\right]|\\&\le\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t_{\mathrm{corr}}\right]|\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\mid T^{\mathrm{s}}>t_{\mathrm{corr}}\right]-\mathbb{P}^{\nu}\left[T>(t-t_{\mathrm{corr}})_+\right]|\\&\quad+\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t_{\mathrm{corr}}\right]|\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}>t\mid T^{\mathrm{p}}>t_{\mathrm{corr}}\right]-\mathbb{P}^{\nu}\left[T>(t-t_{\mathrm{corr}})_+\right]|.\end{aligned}\tag{5.13}$$

Examining the first term,

$$\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}}>t\mid T^{\mathrm{s}}>t_{\mathrm{corr}}\right]=\mathbb{P}^{\mu_{\mathrm{corr}}}\left[T^{\mathrm{s}}>(t-t_{\mathrm{corr}})_+\right].$$

By assumption and Theorem 3.2

$$|\mathbb{P}^{\mu_{\mathrm{corr}}}\left[T^{\mathrm{s}}>(t-t_{\mathrm{corr}})_+\right]-\mathbb{P}^{\nu}\left[T>(t-t_{\mathrm{corr}})_+\right]|\le\epsilon_{\mathrm{corr}}\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}.\tag{5.14}$$

For the other term, since the exit time is beyond t_{corr} the parallel step engages. The single reference process is replaced by the ensemble of N replicas drawn from μ_{phase}, and $T^{\mathrm{p}}=NT^\star+t_{\mathrm{corr}}$. Hence,

$$\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}}>t\mid T^{\mathrm{p}}>t_{\mathrm{corr}}\right]=\mathbb{P}^{\mu_{\mathrm{phase}}}\left[T^\star>\tfrac{1}{N}(t-t_{\mathrm{corr}})_+\right].$$

Therefore, by Theorem 5.1

$$\begin{aligned}&\left|\mathbb{P}^{\mu_{\mathrm{phase}}}\left[T^\star>\tfrac{1}{N}(t-t_{\mathrm{corr}})_+\right]-\mathbb{P}^{\nu}\left[T>(t-t_{\mathrm{corr}})_+\right]\right|\\&\quad\le\epsilon_{\mathrm{phase}}N(1+\epsilon_{\mathrm{phase}})^{N-1}\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}\end{aligned}\tag{5.15}$$

Substituting (5.14) and (5.15) into (5.13), we obtain (5.10).

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

To obtain (5.11), we again decompose as

$$\begin{aligned}\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t}\right] &= \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} \le t_{\mathrm{corr}}\right]\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} \le t_{\mathrm{corr}}\right] \\ &\quad + \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} > t_{\mathrm{corr}}\right]\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right].\end{aligned}$$

and analogously decompose the ParRep expectation. Again, for $t \le t_{\mathrm{corr}}$, the serial algorithm and the reference process of ParRep have the same law. Thus

$$\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} \le t_{\mathrm{corr}}\right] = \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t} \mid T^{\mathrm{p}} \le t_{\mathrm{corr}}\right].$$

Consequently,

$$\begin{aligned}&\left|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t}\right] - \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t}\right]\right| \\ &= \mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right]\left|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} > t_{\mathrm{corr}}\right] - \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t} \mid T^{\mathrm{p}} > t_{\mathrm{corr}}\right]\right|.\end{aligned}$$

Using the QSD as an intermediary,

$$\begin{aligned}&\left|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} > t_{\mathrm{corr}}\right] - \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t} \mid T^{\mathrm{p}} > t_{\mathrm{corr}}\right]\right| \\ &\le \left|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} > t_{\mathrm{corr}}\right] - \mathbb{E}^{\nu}\left[\phi(X_T)1_{T>(t-t_{\mathrm{corr}})_+}\right]\right| \\ &\quad + \left|\mathbb{E}^{\nu}\left[\phi(X_T)1_{T>(t-t_{\mathrm{corr}})_+}\right] - \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t} \mid T^{\mathrm{p}} > t_{\mathrm{corr}}\right]\right|. \qquad (5.16)\end{aligned}$$

For the first term,

$$\begin{aligned}\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} > t_{\mathrm{corr}}\right] &= \mathbb{E}^{\mu_{\mathrm{corr}}}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>(t-t_{\mathrm{corr}})_+}\right] \\ &= \mathbb{E}^{\mu_{t_{\mathrm{corr}}+(t-t_{\mathrm{corr}})_+}}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})\right]\mathbb{P}^{\mu_{\mathrm{corr}}}\left[T^{\mathrm{s}} > (t-t_{\mathrm{corr}})_+\right].\end{aligned}$$

Hence,

$$\begin{aligned}&\left|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t} \mid T^{\mathrm{s}} > t_{\mathrm{corr}}\right] - \mathbb{E}^{\nu}\left[\phi(X_T)1_{T>(t-t_{\mathrm{corr}})_+}\right]\right| \\ &\lesssim \epsilon_{\mathrm{corr}}\left\|\phi\right\|_{L^\infty}\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}. \qquad (5.17)\end{aligned}$$

For the other term, since the parallel step has engaged,

$$\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t} \mid T^{\mathrm{p}} > t_{\mathrm{corr}}\right] = \mathbb{E}^{\mu_{\mathrm{phase}}}\left[\phi(X^{\star}_{T^{\star}})1_{T^{\star}>\frac{1}{N}(t-t_{\mathrm{corr}})_+}\right]. \qquad (5.18)$$

By Theorem 5.1,

$$\begin{aligned}&\left|\mathbb{E}^{\nu}\left[\phi(X_T)1_{T>(t-t_{\mathrm{corr}})_+}\right] - \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t} \mid T^{\mathrm{p}} > t_{\mathrm{corr}}\right]\right| \\ &= \left|\mathbb{E}^{\nu}\left[\phi(X_T)1_{T>(t-t_{\mathrm{corr}})_+}\right] - \mathbb{E}^{\mu_{\mathrm{phase}}}\left[\phi(X^{\star}_{T^{\star}})1_{T^{\star}>\frac{1}{N}(t-t_{\mathrm{corr}})_+}\right]\right| \\ &\le \epsilon_{\mathrm{phase}}N^2\left\|\phi\right\|_{L^\infty}(1+\epsilon_{\mathrm{phase}})^{N-1}\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}. \qquad (5.19)\end{aligned}$$

Using (5.17) and (5.19) in (5.16) gives (5.11).

(5.12) is proved using the preceding estimates,

$$\begin{aligned}&\left|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}}) \mid T^{\mathrm{s}} > t\right] - \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}}) \mid T^{\mathrm{p}} > t\right]\right| \\ &\le \left|\frac{\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{s}}_{T^{\mathrm{s}}})1_{T^{\mathrm{s}}>t}\right] - \mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t}\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right]}\right| \\ &\quad + \left|\mathbb{E}^{\mu_0}\left[\phi(X^{\mathrm{p}}_{T^{\mathrm{p}}})1_{T^{\mathrm{p}}>t}\right]\right|\left|\frac{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right] - \mathbb{P}^{\mu_0}\left[T^{\mathrm{p}} > t\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right]\mathbb{P}^{\mu_0}\left[T^{\mathrm{p}} > t\right]}\right| \\ &\lesssim \left[\epsilon_{\mathrm{corr}} + \epsilon_{\mathrm{phase}}N^2(1+\epsilon_{\mathrm{phase}})^{N-1}\right]\left\|\phi\right\|_{L^\infty}\frac{\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right]} \\ &\quad + \left[\epsilon_{\mathrm{corr}} + \epsilon_{\mathrm{phase}}N(1+\epsilon_{\mathrm{phase}})^{N-1}\right]\left\|\phi\right\|_{L^\infty}\frac{\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right]} \\ &\lesssim \left[\epsilon_{\mathrm{corr}} + \epsilon_{\mathrm{phase}}N^2(1+\epsilon_{\mathrm{phase}})^{N-1}\right]\left\|\phi\right\|_{L^\infty}\frac{\mathrm{e}^{-\lambda_1(t-t_{\mathrm{corr}})_+}\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right]}. \qquad (5.20)\end{aligned}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

Since $(t - t_{\mathrm{corr}})_+ + t_{\mathrm{corr}} \geq t$,

$$\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > (t - t_{\mathrm{corr}})_+ + t_{\mathrm{corr}}\right] \leq \mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right].$$

Therefore,

$$\frac{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right]} \leq \frac{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > (t - t_{\mathrm{corr}})_+ + t_{\mathrm{corr}}\right]} = \frac{1}{\mathbb{P}^{\mu_{\mathrm{corr}}}\left[T^{\mathrm{s}} > (t - t_{\mathrm{corr}})_+\right]}$$

and

$$\frac{\mathrm{e}^{-\lambda_1 (t - t_{\mathrm{corr}})_+}\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t_{\mathrm{corr}}\right]}{\mathbb{P}^{\mu_0}\left[T^{\mathrm{s}} > t\right]} \leq \frac{1}{1 - \epsilon_{\mathrm{corr}}}.$$

Substituting this estimate into (5.20) yields (5.12). □

6. Discussion

We have proven several theorems on the convergence of the exit distributions of parallel replica dynamics to the underlying unaccelerated problem. We have also demonstrated the effectiveness of a dephasing algorithm done in conjunction with the decorrelation step. However, there remain several problems associated with ParRep, both in fully justifying it as an algorithm, and implementing it in practice.

6.1. Error estimates

As we pointed out in the introduction, the error estimates in Theorem 1.3 and Theorem 1.4 include terms which grow as $N \to \infty$. If we take

$$t_{\mathrm{phase}} \gtrsim k_{\mathrm{phase}} \frac{\log N}{\lambda_2 - \lambda_1}$$

for some multiplier, k_{phase}, then the most egregious term in the estimates is bounded by

$$\begin{aligned}\lim_{N\to\infty} N^2 \epsilon_{\mathrm{phase}}(1 + \epsilon_{\mathrm{phase}})^{N-1} &\leq \lim_{N\to\infty} C_{\mathrm{phase}} \mathrm{e}^{-k_{\mathrm{phase}}/2}\left(1 + \mathrm{e}^{-k_{\mathrm{phase}}} C_{\mathrm{phase}}/N\right)^{N-1} \\ &= \mathrm{e}^{-k_{\mathrm{phase}}/2} \mathrm{e}^{C_{\mathrm{phase}} \mathrm{e}^{-k_{\mathrm{phase}}}}.\end{aligned}$$

Hence, taking k_{phase} large enough, the error can be made arbitrarily small. In contrast, the decorrelation error is independent of N, and reducing the decorrelation error will not correct for the error due to more replicas.

The error estimate on the exit time in Theorem 1.3 is a bit deceiving and merits additional comment. It would appear that when we consider this cumulative distribution function at any $t > 0$, then, sending $N \to \infty$, the error vanishes. This is a reflection on the estimate being an absolute error. Dividing out by $\mathrm{e}^{-N\lambda_1 t}$ lets us evaluate the relative error, which we see is uniformly bounded in t.

We also remark that since

$$\mathbb{E}[T] = \int_0^\infty \mathbb{P}[T > t] \mathrm{d}t,$$

we can obtain error estimates on the expected exit time. Using the estimates in Theorem 1.3, we see that provided $N\epsilon_{\mathrm{phase}}(1 + \epsilon_{\mathrm{phase}})^{N-1} < 1$, we have

$$\left|\mathbb{E}^{\mu_{\mathrm{phase}}}[T^\star] - \frac{1}{N\lambda_1}\right| \leq N\epsilon_{\mathrm{phase}}(1 + \epsilon_{\mathrm{phase}})^{N-1}. \tag{6.1}$$

Similarly, using the estimates in Theorem 1.4,

$$\left|\mathbb{E}^{\mu_0}[T^{\mathrm{s}}] - \mathbb{E}^{\mu_0}[T^{\mathrm{p}}]\right| \lesssim \epsilon_{\mathrm{corr}} + N\epsilon_{\mathrm{phase}}(1 + \epsilon_{\mathrm{phase}})^{N-1}. \tag{6.2}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

It remains to be determined whether our estimates are sharp – is the growth in N real or an artifact of our analysis? While we cannot yet address the sharpness, a simple numerical experiment indicates that there is growth in the error as N increases. Consider the problem

$$\mathrm{d}X_t = -4X_t\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \tag{6.3}$$

for the well $W = [-1, 1]$, and suppose we launch N replicas from the Dirac distribution $X_0 = .1$. By symmetry, we know that if we had perfect dephasing, then during the parallel step

$$\mathbb{P}^\nu[X^\star_{T^\star} = 1] = \mathbb{P}^\nu[X^\star_{T^\star} = -1] = \tfrac{1}{2}.$$

But if we incompletely dephase, then, because of our asymmetric initial condition, we expect a higher probability of escaping at 1 than -1. For this problem, we can compute by spectral methods that $\lambda_1 \approx 0.971972$ and $\lambda_2 \approx 8.98262$.

To test our conjecture, that the error increases with N, we ran 10000 realizations of the dephasing and parallel steps with values of $N = 100, 200, \ldots, 1000$. We employed Euler–Maruyama time stepping with $\Delta t = 10^{-4}$. We then ran this with with $t_{\mathrm{phase}} = .05$, .1 and .2. The results appear in Figure 3.

As we predicted, the errors decrease as t_{phase} increases. For the smallest dephasing time, we also see the error increase with N. At $t_{\mathrm{phase}} = .1$, there is still some increase in the error as N increases, though it is less dramatic. When $t_{\mathrm{phase}} = .2$, the trend appears to have been lost to numerical error and sampling variability.

6.2. Numerical parameters and eigenvalues

An essential question is how to choose of the dephasing and decorrelation time parameters. Based on the arguments in the preceding section, roughly, if we desire the errors from decorrelation and dephasing to be of the same order, then,

$$2\log(N)t_{\mathrm{corr}} \sim t_{\mathrm{phase}}.$$

So, while they should not be the same, if we can estimate one, we can infer the other. There will also be some mismatch due to different starting distributions for the reference process and the dephasing replicas.

t_{corr} must be large enough so as to be representative of the QSD while remaining computationally efficient. Taking too large a value of t_{corr} will just replicate the serial implementation with no acceleration. Theorem 3.1 provides some insight, already discussed in [4]. The error of μ_{corr} is controlled by the following quantities:

- The μ_0 initial distribution,
- $\left\|P_{[2,\infty)}\mu_0\right\|_{H_\mu^{-s}}$, the mismatch between the initial distribution of the reference process and the quasistationary distribution, ν;
- The value of $\underline{t}$;
- $\int u_1 \mathrm{d}\mu_0$;
- $\lambda_2 - \lambda_1$, the spectral gap between the first two eigenvalues.

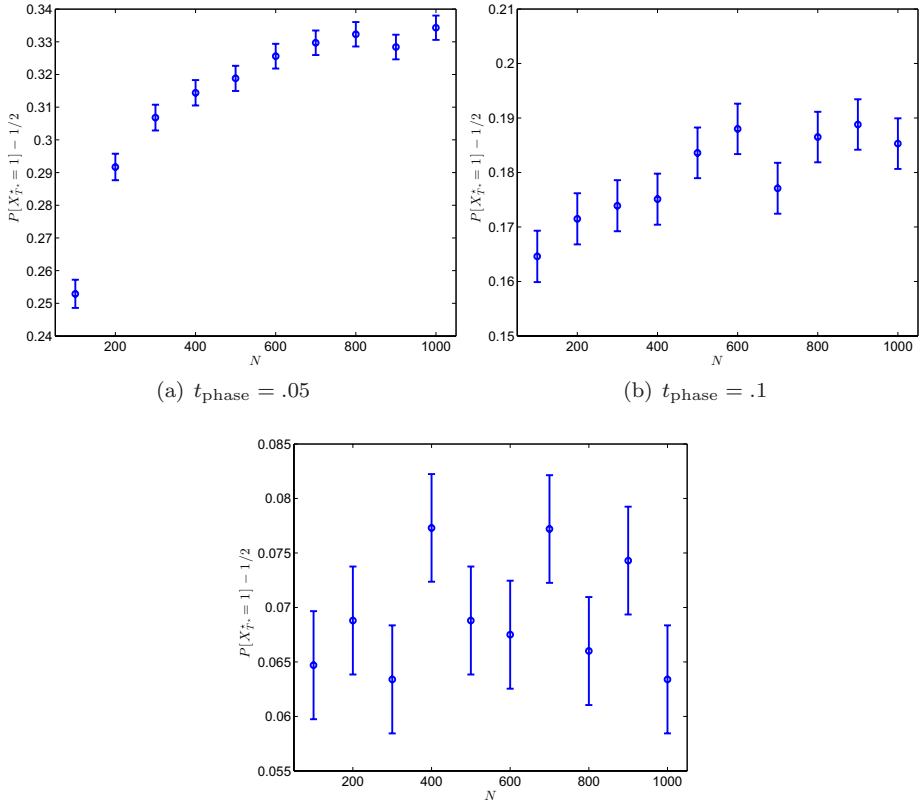
Based on these quantities, and how they relate to t_{corr}, to make the decorrelation error small, we would certainly need

$$t_{\mathrm{corr}} \gtrsim \frac{\ln\left[\left(\int u_1 \mathrm{d}\mu_0\right)^{-1}\left\|P_{[2,\infty)}\mu_0\right\|_{H_\mu^{-s}}\right]}{\lambda_2 - \lambda_1} + \underline{t}. \tag{6.4}$$

The eigenvalues also play an important role in determining which problems would benefit from ParRep is an outstanding problem, which is an outstanding issue. For ParRep to be efficient, we need

$$t_{\mathrm{corr}} \ll \mathbb{E}^{\mu_{\mathrm{corr}}}[T] \sim \mathbb{E}^\nu[T] = \frac{1}{\lambda_1}. \tag{6.5}$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press



(a) $t_{\text{phase}} = .05$ (b) $t_{\text{phase}} = .1$

(c) $t_{\text{phase}} = .2$

FIGURE 3. Three experiments on the impact of imperfect dephasing for (6.3). With perfect dephasing, the probability of exiting at $x = 1$ would be $1/2$, but because the initial condition is .1 and the dephasing time is finite, there is a persistent bias and growth in the error as N increases. 95% confidence intervals are plotted for 10000 realizations of each value of N.

This is desirable because, in the event X_t does not leave the well during the decorrelation step, it is will now take a comparatively long time to exit. In [4], the authors suggested

$$t_{\text{corr}} \leq \mathbb{E}^{\mu_0}[T].$$

However, this can be problematic, depending on μ_0. As previously discussed, if the replicas launch from a position too close to the boundary, $\mathbb{E}^{\mu_0}[T]$ might be rather small. This is mitigated as t_{corr} becomes larger, leading to $\mathbb{E}^{\mu_{\text{corr}}}[T]$ approaching the escape time of the QSD, λ_1^{-1}.

We can see from (6.4) and constraint (6.5) that ParRep will be most effective when

$$\frac{1}{\lambda_2 - \lambda_1} \ll \frac{1}{\lambda_1}, \tag{6.6}$$

or, alternatively, when $\lambda_1 \ll \lambda_2 - \lambda_1$. Under these conditions, μ_{corr} converges to ν much more rapidly than we expect X_t to exit W. (6.6) can also be viewed as a characterization of when W corresponds to a metastable state for (1.1).

Computing λ_1 and λ_2 directly from a discretization of the elliptic operator L is intractable for all but the lowest dimensional systems. Instead, one must use Monte Carlo methods, such as those found in [11,12,17,18,23].

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

However, these studies, some of which use branching particles processes like Fleming–Viot (discussed below), only yield λ_1.

In a forthcoming work, we explore a mechanism for computing $\lambda_2 - \lambda_1$ using observables. The idea stems from calculations in Theorem 1.1, that, for an observable $\mathcal{O}(x)$, as $t \to \infty$,

$$\mathbb{E}^{\mu_0}\left[\mathcal{O}(X_t) \mid T > t\right] = \int_W \mathcal{O}(x)\mathrm{d}\nu(x) + C(\mu_0, \mathcal{O})\mathrm{e}^{-(\lambda_2-\lambda_1)t} + \dots . \tag{6.7}$$

In principle, $\lambda_2 - \lambda_1$ could be extracted from a time series of $\mathbb{E}^{\mu_0}\left[\mathcal{O}(X_t) \mid T > t\right]$. This introduces a variety of questions, such as what observables to use and how to perform such a fitting. Thus, we will have a method for dynamically estimating t_{corr} and t_{phase}.

6.3. Dephasing mechanism

The efficiency of our dephasing algorithm can be improved by the availability of multiple processors. For instance, assume we have N processors available for the replicas and that k replicas have successfully been run until t_{corr}. We are still waiting for $N - k$ replicas to successfully dephase. Rather than let k processors sit idle, they could record the successful replicas, and run independent realizations. As more replicas finish dephasing, more processors can be brought to bear on the outstanding replicas.

In practice, as replicas are deemed to have been successfully dephased, they are promoted to the parallel step, [21]. Thus, there is no bottleneck at the dephasing step from waiting to get N realizations dephased.

There are other approaches to dephasing too, such as Fleming–Viot or Moran branching interacting particle processes, [2,3,9,15]. These merit consideration for ParRep. These approaches, which randomly split a surviving process every time another process exits the well, can provide additional information, such as an estimate of λ_1. Moreover, no processor sits idle at anytime. However, two challenges are introduced. On a practical level, one needs to implement additional communication routines and synchronization across the processors to request and send configurations as trajectories are killed. The second challenge is analytical, as the dephased processes will now be only approximately independent. This complicates the analysis of the how the error in the dephasing step cascades through the parallel step.

6.4. Other challenges

Another task is to assess the cumulative error over many ParRep cycles. The hitting point distribution will be perturbed by the algorithm, meaning that the sequence in which the states are visited would also be perturbed. Quantifying the error across many steps, and showing that it may be made small, would complete the justification of ParRep over the lifetime of a simulation. But to begin such a study, one must decide how to measure

$$\mathrm{dist}(\mathcal{S}_t, \mathcal{S}_t^{\mathrm{ParRep}}).$$

The challenge is that $\mathcal{S}_t$ is not a Markovian process. A particle that sits near the edge of the well is likely to exit much sooner than one which is near the minima of the well. But that information is lost in the coarse graining. Knowing how long X_t has been in the well provides some amount of information; it tells us the proximity to the QSD, from which we can get an exponential exit time.

Despite the challenge of studying the coarse grained flow, we can report that ParRep appears to work as predicted over multiple wells. Consider the flow

$$\mathrm{d}X_t = -2\pi \sin(\pi X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t. \tag{6.8}$$

For this equation, with initial condition $X_0 = 0$, we examined the time it would take to reach the wells centered at $x = \pm 10$. In other words, we sought to compute

$$T_{\pm 10} = \inf\left\{t \mid |X_t| \geq 9\right\}.$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

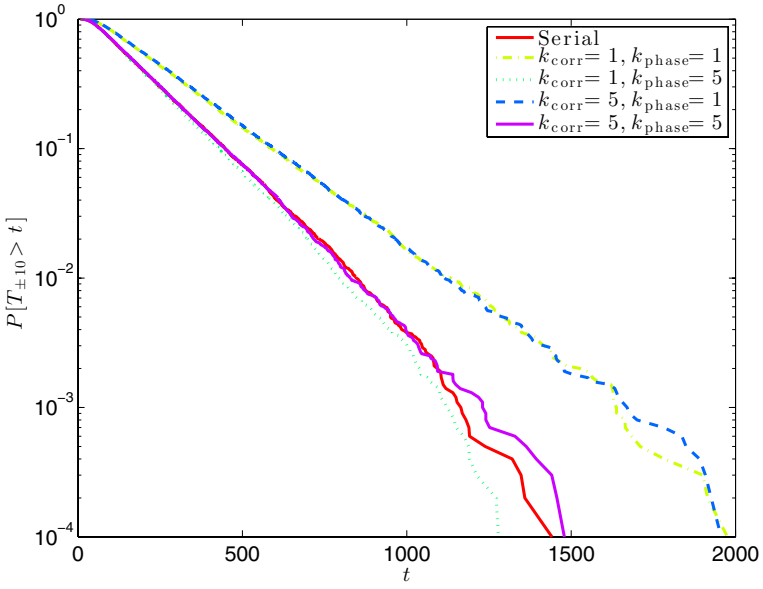


FIGURE 4. The cumulative distribution for the time for it takes trajectory (6.8) to reach the wells centered at ± 10. 10 000 realizations of each case were run with time step $\Delta t = 10^{-4}$. $t_{\rm corr}$ and $t_{\rm phase}$ relate to $k_{\rm corr}$ and $k_{\rm phase}$ via (6.9). As expected, larger values of $t_{\rm corr}$ and $t_{\rm phase}$ give better agreement with an unaccelerated process.

For this problem, we ran the full ParRep algorithm (decorrelation, dephasing and parallel steps) within *each* well. During dephasing, the replicas were initiated from the minima of the present well, $0, \pm 2\pi, \pm 4\pi, \ldots$ We ran 10 000 realizations of this experiment, varying $k_{\rm corr}$ and $k_{\rm phase}$, where

$$t_{\rm corr} = \frac{k_{\rm corr}}{\lambda_2 - \lambda_1}, \quad t_{\rm phase} = \frac{k_{\rm phase}}{\lambda_2 - \lambda_1}. \tag{6.9}$$

Since the wells are periodic, we can use spectral methods to compute $\lambda_1 \approx .202280$ and $\lambda_2 \approx 16.2588$ once, and we then have these values for all the wells. The results, with $\Delta t = 10^{-4}$ and $N = 100$ replicas, appear in Figure 4

As we expect, for sufficiently large values of $t_{\rm corr}$ and $t_{\rm phase}$, the distributions agree with the serial process. Indeed, in the cases $k_{\rm corr} = k_{\rm phase} = 5$ and $k_{\rm corr} = 1$, $k_{\rm phase} = 5$, the exit times agree with the serial realization at 5% significance level under a Kolmogorov–Smirnov test. In addition, this experiment also supports our calculations that, through the dephasing error, the total error should be magnified by N since increasing the dephasing time improves the fit much more than increasing the decorrelation time does.

Finally, we remark that we have only analyzed the continuous in time problem, though we are ultimately interested in the associated discrete in time algorithm. Much of the analysis carries over to the discrete in time case. A discrete in time quasistationary distribution exists, and there are extensive results on using interacting particle algorithms for dephasing, [9, 10]. As in the continuous in time case, there remains the subtlety of how to analyze the parallel step when the dephased ensemble is only approximately independent.

However, the discrete time step introduces other subtleties. Assume one uses Euler–Maruyama time discretization with time step Δt, and define the exit time as

$$T^{\Delta t} = \inf \{ t_n \mid X_{t_n} \notin W \}. \tag{6.10}$$

For a uniform time step, we see that with no acceleration of the dynamics, the exit times are integer multiples of Δt. For ParRep, this remains true for exits that take place during the decorrelation step. But for exit times

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

taking place during the parallel step, the exit times will be determined by multiples of $N\Delta t$. With a large number of processors, this effective time step could be quite large. When comparing against the continuous in time problem, the error of discretization could be magnified in ParRep. In the preceding experiment, $N\Delta t = .01$, which is small relative to the exit time scale ($1/\lambda_1 \approx 4.9$) and the decorrelation time scale ($1/(\lambda_2 - \lambda_1) \approx .062$). Clearly, the discrete in time case warrants a thorough investigation.

Appendix A. Summation bounds

Much of our analysis relies on bounding series solutions, (2.6), of (2.2), to obtain information about X_t through the Feynman–Kac equation, (2.3). The key estimates needed in our work stem from Weyl's Law for L:

Proposition A.1 (Weyl's law for L)**.** *There exist positive constants c_1 and c_2, independent of k, such that the eigenvalues of (2.1) satisfy*

$$c_1 k^{2/n} \leq \lambda_k \leq c_2 k^{2/n}. \tag{A.1}$$

Recall that n denotes the dimension of the underlying problem; $X_t \in \mathbb{R}^n$.

Proof. We will not reproduce the proof here, which is accomplished by rewriting the eigenvalue problem as

$$-\beta^{-1}\nabla \cdot \left(\mathrm{e}^{-\beta V}\nabla u\right) = \lambda \mathrm{e}^{-\beta V} u. \tag{A.2}$$

This is justified because V is smooth and W is bounded; thus $\mathrm{e}^{-\beta V}$ is smooth and nondegenerate. This is now in the form of Theorem 6.3.1 of [8] on Weyl's Law, yielding the result. □

Using Weyl's Law, we have our main summation result,

Proposition A.2. *Given $s \geq 0$, let $\mathbf{a} = (a_1, a_2, \ldots)$ satisfy*

$$\left\{\sum_{k=1}^{\infty} \lambda_k^{-s} |a_k|^2\right\}^{1/2} = \|\mathbf{a}\|_{H_\mu^{-s}} < \infty.$$

Let f be defined as

$$f(\tau) \equiv \sum_{k=1}^{\infty} a_k \lambda_k^{\alpha} \mathrm{e}^{-\tau\lambda_k}. \tag{A.3}$$

For $a > 0$, we have:

A.

$$\sup_{\tau \geq a} |f(\tau)| \lesssim a^{-n/4-\max\{s/2+\alpha,0\}} \|\mathbf{a}\|_{H_\mu^{-s}} < \infty; \tag{A.4}$$

B. The convergence of the series is uniform in $\tau \geq a$;
C. f is continuous.

To prove Proposition A.2, we first have the following lemma.

Lemma A.3. *Let λ_k be the eigenvalues and eigenfunctions of L, (2.1). There exists a constant $C > 0$, independent of τ, such that for all $\tau > 0$,*

$$\sum_{k=1}^{\infty} \lambda_k^{\alpha} \mathrm{e}^{-\tau\lambda_k} \leq C\tau^{-n/2-\max\{\alpha,0\}}. \tag{A.5}$$

The reader should rightfully expect the lefthand side of (A.5) to grow as $\alpha \to \infty$. Indeed, the constant C depends on α and will grow. However, as α is fixed, and we are interested in an estimate in τ, this is suppressed.

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

Proof. For $\alpha \leq 0$,

$$\sum_{k=1}^{\infty} \mathrm{e}^{-\tau\lambda_k}\lambda_k^{\alpha} \leq \sum_{k=1}^{\infty} \mathrm{e}^{-\tau\lambda_k}\lambda_1^{\alpha} \leq \sum_{k=1}^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}\lambda_1^{\alpha}$$
$$\leq \lambda_1^{\alpha}\int_0^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}\mathrm{d}k = \lambda_1^{\alpha}(c_1\tau)^{-n/2}\Gamma\left[1+\frac{n}{2}\right].$$

In the above computation, we approximated the sum as the lower Riemann sum of the integral.

For $\alpha > 0$, we begin by estimating

$$\sum_{k=1}^{\infty} \mathrm{e}^{-\tau\lambda_k}\lambda_k^{\alpha} \leq \sum_{k=1}^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}c_2^{\alpha}k^{2\alpha/n}.$$

For sufficiently large k,

$$k \geq k_1 \equiv \left\lceil \left(\frac{\alpha}{c_1\tau}\right)^{n/2} \right\rceil,$$

the summand is monotonically decreasing, while for $k < k_1$, it is monotonically increasing. Splitting the sum up,

$$\sum_{k=1}^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}k^{2\alpha/n} = \sum_{k=1}^{k_1} \mathrm{e}^{-c_1\tau k^{2/n}}k^{2\alpha/n+1} + \sum_{k=k_1+1}^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}k^{2\alpha/n}$$
$$\leq \mathrm{e}^{-c_1\tau}\sum_{k=1}^{k_1} k^{2\alpha/n} + \sum_{k=k_1+1}^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}k^{2\alpha/n}.$$

Crudely bounding the first sum in terms of a max, and treating the latter sum as a lower Riemann approximations of an integral,

$$\sum_{k=1}^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}k^{2\alpha/n} \leq \mathrm{e}^{-c_1\tau}k_1 \cdot k_1^{2\alpha/n} + \int_{k_1}^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}k^{2\alpha/n}\mathrm{d}k$$
$$\leq \mathrm{e}^{-c_1\tau}\left[\left(\frac{\alpha}{c_1\tau}\right)^{n/2}+1\right]^{2\alpha/n+1} + \int_0^{\infty} \mathrm{e}^{-c_1\tau k^{2/n}}k^{2\alpha/n}\mathrm{d}k$$
$$\leq \left(\frac{c_1\tau}{\alpha}\right)^{-n/2-\alpha}\mathrm{e}^{-c_1\tau}\left[1+\left(\frac{c_1\tau}{\alpha}\right)^{n/2}\right]^{2\alpha/n+1} + \frac{n}{2}(c_1\tau)^{-n/2-\alpha}\Gamma\left[\frac{n}{2}+\alpha\right]$$
$$\lesssim \tau^{-n/2-\alpha}.$$

□

The integrals were computed using Mathematica, with the commands

```
Integrate[Exp[-c*t*k^(2/n)],{k,0,Infinity}]
```

```
Integrate[Exp[-c*t*k^(2/n)]*k^(2*a/n+1),{k,0,Infinity}]
```

Now we prove Proposition A.2.

Proof. We first observe that f is well defined and bounded:

$$|f(\tau)| \leq \sum_{k=1}^{\infty} |a_k|\,\lambda_k^{\alpha}\mathrm{e}^{-\tau\lambda_k} \leq \left\{\sum_{k=1}^{\infty} \lambda^{s+2\alpha}\mathrm{e}^{-2\lambda_k\tau}\right\}^{1/2} \|\mathbf{a}\|_{H_{\mu}^{-s}}.$$

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

Applying Lemma A.3 with $\alpha \mapsto s + 2\alpha$ and $\tau \mapsto 2a$,

$$\sum_{k=1}^{\infty} \lambda^{s+2\alpha} \mathrm{e}^{-2\lambda_k a} \lesssim (2a)^{-n/2-\max\{s+2\alpha,0\}}.$$

To prove uniform convergence, let

$$f_m(\tau) \equiv \sum_{k=1}^{m} a_k \lambda_k^{\alpha} \mathrm{e}^{-\tau\lambda_k}$$

denote the partial sum. Obviously, each partial sum is continuous in τ. Then

$$\begin{aligned} |f(\tau) - f_m(\tau)| &\leq \sum_{k=m+1}^{\infty} |a_k| \, \lambda_k^{\alpha} \mathrm{e}^{-\tau\lambda_k} \\ &\leq \left\{ \sum_{k=m+1}^{\infty} \lambda_k^{s+2\alpha} \mathrm{e}^{-2\lambda_k a} \right\}^{1/2} \left\| P_{[m+1,\infty)} \mathbf{a} \right\|_{H_\mu^{-s}} \\ &\leq \|\mathbf{a}\|_{H_\mu^{-s}} \left\{ \sum_{k=m+1}^{\infty} \lambda_k^{s+2\alpha} \mathrm{e}^{-2\lambda_k a} \right\}^{1/2}. \end{aligned}$$

Examining the sum,

$$\sum_{k=m+1}^{\infty} \lambda_k^{s+2\alpha} \mathrm{e}^{-2\lambda_k a} \lesssim \sum_{k=m+1}^{\infty} k^{2s/n+4\alpha/n} \mathrm{e}^{-2c_1 a k^{2/n}}.$$

Taking m sufficiently large, the summand will be strictly decreasing in k, so we can treat it as a lower Riemann sum for the integral

$$\int_m^{\infty} k^{2s/n+4\alpha/n} \mathrm{e}^{-2c_1 a k^{2/n}} \mathrm{d}k.$$

Changing variables by letting $k^{2/n} = l$,

$$\sum_{k=m+1}^{\infty} \lambda_k^{s+2\alpha} \mathrm{e}^{-2\lambda_k a} \lesssim \int_{m^{2/n}}^{\infty} l^{s+2\alpha+n/2-1} \mathrm{e}^{-2c_1 a l} \mathrm{d}l.$$

If $s + 2\alpha + n/2 - 1 \leq 0$, then

$$\sum_{k=m+1}^{\infty} \lambda_k^{s+2\alpha} \mathrm{e}^{-2\lambda_k a} \lesssim \int_{m^{2/n}}^{\infty} \mathrm{e}^{-2c_1 a l} \mathrm{d}l = \frac{1}{2c_1 a} \mathrm{e}^{-2m^{2/n} c_1 a}.$$

On the other hand, if $s + 2\alpha + n/2 - 1 > 0$, we can trade some of the exponential decay to eliminate the algebraic term,

$$\sum_{k=m+1}^{\infty} \lambda_k^{s+2\alpha} \mathrm{e}^{-2\lambda_k a} \lesssim \int_{m^{2/n}}^{\infty} \mathrm{e}^{-c_1 a l} \mathrm{d}l = \frac{1}{c_1 a} \mathrm{e}^{-m^{2/n} c_1 a}.$$

In either case, we see that for any $a > 0$,

$$\lim_{m\to\infty} \sup_{\tau \geq a} |f(\tau) - f_m(\tau)| = 0.$$

Since the partial sums converge uniformly to f, it is now a classical result to conclude that f is continuous for $\tau \geq a > 0$, [24]. □

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press

Acknowledgements. The authors wish to thank D. Aristoff, K. Leder, S. Mayboroda, A. Shapeev and O. Zeitouni for helpful conversations in developing these ideas. We also thank D. Perez and A.F. Voter for conversations at LANL that motivated important refinements of our estimates. This work was supported by the NSF PIRE grant OISE-0967140 and the DOE grant DE-SC0002085.

References

[1] R.A. Adams and J.J.F. Fournier, *Sobolev spaces*. Academic Press **140** (2003).
[2] M. Bieniek, K. Burdzy and S. Finch, Non-extinction of a Fleming–Viot particle model. *Probab. Theory Relat. Fields* (2011).
[3] M. Bieniek, K. Burdzy and S. Pal, Extinction of Fleming–Viot-type particle systems with strong drift. *Electron. J. Prob.* **17** (2012).
[4] C. Le Bris, T. Lelièvre, M. Luskin and D. Perez, A mathematical formalization of the parallel replica dynamics. *Monte Carlo Methods Appl.* **18** (2012) 119–146.
[5] P. Cattiaux, P. Collet, A. Lambert, S. Martínez, S. Méléard and J. San Martín, Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Prob.* **37** (2009) 1926–1969.
[6] P. Cattiaux and S. Méléard, Competitive or weak cooperative stochastic Lotka–Volterra systems conditioned on non-extinction. *J. Math. Biol.* **60** (2010) 797–829.
[7] P. Collet, S. Martínez and J. San Martín, Asymptotic laws for one-dimensional diffusions conditioned to nonabsorption. *Ann. Prob.* **23** (1995) 1300–1314.
[8] E.B. Davies, *Spectral theory and differential operators*. Cambridge University Press **42** (1996).
[9] P. Del Moral, Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications. *Probability and Its Applications*, Springer (2011).
[10] P. Del Moral and A. Doucet, Particle motions in absorbing medium with hard and soft obstacles. *Stoch. Anal. Appl.* **22** (2004) 1175–1207.
[11] P. Del Moral and L. Miclo, Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman–Kac semigroups. *ESAIM: PS* **7** (2003) 171–208.
[12] M. El Makrini, B. Jourdain and T. Lelievre, Diffusion Monte Carlo method: Numerical analysis in a simple case. *ESAIM: M2AN* **41** (2007) 189–213.
[13] L.C. Evans, Partial Differential Equations. *Amer. Math. Soc.* 2002.
[14] D. Gilbarg and N.S. Trudinger, *Elliptic partial differential equations of second order*. Springer Verlag **224** (2001).
[15] I. Grigorescu and M. Kang, Hydrodynamic limit for a Fleming-Viot type system. *Stoch. Process. Their Appl.* **110** (2004) 111–143.
[16] D. Haroske and H. Triebel, Distributions, Sobolev spaces, elliptic equations. *Europ. Math. Soc.* (2008).
[17] A. Lejay and S. Maire, Computing the principal eigenvalue of the Laplace operator by a stochastic method. *Math. Comput. Simul.* **73** (2007) 351–363.
[18] A. Lejay and S. Maire, Computing the principal eigenelements of some linear operators using a branching Monte Carlo method. *J. Comput. Phys.* **227** (2008) 9794–9806.
[19] S. Martínez and J. San Martín, Quasi–stationary distributions for a Brownian motion with drift and associated limit laws. *J. Appl. Probab.* **31** (1994) 911–920.
[20] S. Martínez and J. San Martín. Classification of killed one-dimensional diffusions. *Ann. Probab.* **32** (2004) 530–552.
[21] D. Perez, Implementation of Parallel Replica Dynamics, Personal Communication (2012).
[22] D. Perez, B.P. Uberuaga, Y. Shim, J.G. Amar and A.F. Voter, Accelerated molecular dynamics methods: introduction and recent developments. *Ann. Reports Comput. Chemistry* **5** (2009) 79–98.
[23] M. Rousset, On the control of an interacting particle estimation of Schrödinger ground states. *SIAM J. Math. Anal.* **38** (2006) 824–844 (electronic).
[24] W. Rudin, *Principles of Mathematical Analysis*. McGraw-Hill (1976).
[25] D. Steinsaltz and S.N. Evans, Quasistationary distributions for one–dimensional diffusions with killing. *Trans. Amer. Math. Soc.* **359** (2007) 1285–1324 (electronic).
[26] A.F. Voter, Parallel replica method for dynamics of infrequent events. *Phys. Rev. B* **57** (1998) 13985–13988.
[27] A.F. Voter, F. Montalenti and T.C. Germann, Extending the time scale in atomistic simulation of materials. *Ann. Rev. Materials Sci.* **32** (2002) 321–346.

https://doi.org/10.1051/m2an/2013068 Published online by Cambridge University Press