Finite Markov Chains Coupled to General Markov Processes and An Application to Metastability I

Thomas G. Kurtz and Jason Swanson

Abstract We consider a diffusion given by a small noise perturbation of a dynamical system driven by a potential function with a finite number of local minima. The classical results of Freidlin and Wentzell show that the time this diffusion spends in the domain of attraction of one of these local minima is approximately exponentially distributed and hence the diffusion should behave approximately like a Markov chain on the local minima. By the work of Bovier and collaborators, the local minima can be associated with the small eigenvalues of the diffusion generator. Applying a Markov mapping theorem, we use the eigenfunctions of the generator to couple this diffusion to a Markov chain whose generator has eigenvalues equal to the eigenvalues of the diffusion generator that are associated with the local minima and establish explicit formulas for conditional probabilities associated with this coupling. The fundamental question then becomes to relate the coupled Markov chain to the approximate Markov chain suggested by the results of Freidlin and Wentzel.

1 Introduction

Fix $\varepsilon > 0$ and consider the stochastic process,

$$X_\varepsilon(t) = X_\varepsilon(0) - \int_0^t \nabla F(X_\varepsilon(s))\, ds + \sqrt{2\varepsilon}\, W(t), \tag{1}$$

where $F \in C^3(\mathbb{R}^d)$ and W is a standard d-dimensional Brownian motion. For the precise assumptions on F, see Section 3.1. Let φ be the solution to the differential

Thomas G. Kurtz
University of Wisconsin-Madison, e-mail: kurtz@math.wisc.edu

Jason Swanson
University of Central Florida, e-mail: Jason.Swanson@ucf.edu

G. Yin, T. Zariphopoulou (eds.), *Stochastic Analysis, Filtering, and Stochastic Optimization*,
https://doi.org/10.1007/978-3-030-98519-6_12

equation $\varphi' = -\nabla F(\varphi)$. We will use φ_x to denote the solution with $\varphi_x(0) = x$. The process X_ε is a small-noise perturbation of the deterministic process φ.

Suppose that $\mathcal{M} = \{x_0, \dots, x_m\}$ is the set of local minima of the potential function F. The points x_j are stable points for the process φ. For X_ε, however, they are not stable. The process X_ε will initially gravitate toward one of the x_j and move about randomly in a small neighborhood of this point. But after an exponential amount of time, a large fluctuation of the noise term will move the process X_ε out of the domain of attraction of x_j and into the domain of attraction of one of the other minima. We say that each point x_j is a point of *metastability* for the process X_ε.

If X is a cadlag process in a complete, separable metric space S adapted to a right continuous filtration (assumptions that are immediately satisfied for all processes considered here) and H is either open or closed, then $\tau_H^X = \inf\{t > 0 : X(t) \text{ or } X(t-) \in H\}$ is a stopping time (see, for example, [8, Proposition 1.5]). If $x \in S$, let $\tau_x^X = \tau_{\{x\}}^X$. We may sometimes also write $\tau^X(H)$, and if the process is understood, we may omit the superscript.

Let

$$D_j = \{x \in \mathbb{R}^d : \lim_{t\to\infty} \varphi_x(t) = x_j\} \tag{2}$$

be the domains of attraction of the local minima. It is well-known (see, for example, [9], [4, Theorem 3.2], [5, Theorems 1.2 and 1.4], and [7]) that as $\varepsilon \to 0$, $\tau^{X_\varepsilon}(D_j^c)$ is asymptotically exponentially distributed under P^{x_j}. It is therefore common to approximate the process X_ε by a continuous time Markov chain on the set $\mathcal{M}$ (or equivalently on $\{0, \dots, m\}$). In fact, metastability can be defined in terms of convergence, in an appropriate sense, to a continuous time Markov chain. (See the survey article [15] for details.) Beltrán and Landim [2, 3] introduced a general method for proving the metastability of a Markov chain. Along similar lines, Rezakhanlou and Seo [19] developed such a method for diffusions. For an alternative approach using intertwining relations, see [1].

In this project, for each $\varepsilon > 0$, we wish to capture this approximate Markov chain behavior by coupling X_ε to a continuous time Markov chain, Y_ε, on $\{0, \dots, m\}$. We will refer to the indexed collection of coupled processes, $\{(X_\varepsilon, Y_\varepsilon) : \varepsilon > 0\}$, as a *coupling sequence*. Our objective is to investigate the possibility of constructing a coupling sequence which satisfies

$$P(X_\varepsilon(t) \in D_j \mid Y_\varepsilon(t) = j) \to 1 \tag{3}$$

as $\varepsilon \to 0$, for all j. We also want the transition rate for Y_ε to go from i to j to be asymptotically equivalent as $\varepsilon \to 0$ to the transition rate for X_ε to go from a neighborhood of x_i to a neighborhood of x_j. That is, we would like

$$E^i[\tau_j^{Y_\varepsilon}] \sim E^{x_i}[\tau_{B_\rho(x_0)}^{X_\varepsilon}] \tag{4}$$

as $\varepsilon \to 0$, for all i and j, where $B_\rho(x)$ is the ball of radius ρ centered at x.

In this paper (Part I), we develop our general coupling construction. The construction goes beyond the specific case of interest here. It is a construction that builds a coupling between a Markov process on a complete and separable metric space and a continuous-time Markov chain where the generators of the two processes have common eigenvalues. The coupling is done in such a way that observations of the chain yield quantifiable conditional probabilities about the process. This coupling construction is built in Section 2 and uses the Markov mapping theorem (Theorem 19). In Section 3, we apply this construction method to reversible diffusions in $\mathbb{R}^d$ driven by a potential function with a finite number of local minima.

With this coupling construction in hand, we can build the coupling sequences described above. In our follow-up work (Part II), we take up the question of the existence and uniqueness of a coupling sequence that satisfies requirements (3) and (4).

2 The general coupling

2.1 Assumptions and definitions

Given a Markov process X with generator A satisfying Assumption 1, we will use the Markov mapping theorem to construct a coupled pair, (X,Y), in such a way that for a specified class of initial distributions, Y is a continuous-time Markov chain on a finite state space. The construction then allows us to explicitly compute the conditional distribution of X given observations of Y.

For explicit definitions of the notation used here and throughout, see the Appendix.

Assumption Let E be a complete and separable metric space.

(i) $A \subset \overline{C}(E) \times \overline{C}(E)$.
(ii) A has a stationary distribution $\varpi \in \mathcal{P}(E)$, which implies $\int_E Af\,d\varpi = 0$ for all $f \in \mathcal{D}(A)$.
(iii) For some m, there exist signed measures $\varpi_1, \ldots, \varpi_m$ on E and positive real numbers $\lambda_1, \ldots, \lambda_m$ such that, for each $k \in \{1, \ldots, m\}$ and $f \in \mathcal{D}(A)$,

$$\int_E Af\,d\varpi_k = -\lambda_k \int_E f\,d\varpi_k, \tag{5}$$

$$\varpi_k(dx) = \eta_k(x)\varpi(dx), \text{ where } \eta_k \in \overline{C}(E), \tag{6}$$

$$\varpi_k(E) = 0. \tag{7}$$

We define $\varpi_0 = \varpi$ and $\eta_0 = 1$. □

Remark If $(1,0) \in A$, then (5) implies (7). □

Remark In what follows, we will make use of the assumption that the functions η_k are continuous. However, this assumption can be relaxed by appealing to the methods in Kurtz and Stockbridge [14]. □

Assumption Let E be a complete and separable metric space. Let $A \subset \overline{C}(E) \times \overline{C}(E)$, $m \in \mathbb{N}$, $Q \in \mathbb{R}^{(m+1)\times(m+1)}$, and $\xi^{(1)}, \ldots, \xi^{(m)} \in \mathbb{R}^{m+1}$.

(i) A and m satisfy Assumption 1.
(ii) Q is the generator of a continuous-time Markov chain with state space $E_0 = \{0, 1, \ldots, m\}$ and eigenvalues $\{0, -\lambda_1, \ldots, -\lambda_m\}$.
(iii) The vectors $\xi^{(1)}, \ldots, \xi^{(m)}$ are right eigenvectors of Q, corresponding to the eigenvalues $-\lambda_1, \ldots, -\lambda_m$.
(iv) For each $i \in \{0, 1, \ldots, m\}$, the function

$$\alpha_i(x) = 1 + \sum_{k=1}^{m} \xi_i^{(k)} \eta_k(x) \tag{8}$$

satisfies $\alpha_i(x) > 0$ for all $x \in E$.

We define $\xi^{(0)} = (1, \ldots, 1)^T$, so that the function $\alpha : E \to \mathbb{R}^{m+1}$ is given by $\alpha = \sum_{k=0}^{m} \xi^{(k)} \eta_k$. □

Remark Given (A, m, Q) satisfying (i) and (ii) of Assumption 4, it is always possible to choose vectors $\xi^{(1)}, \ldots, \xi^{(m)}$ satisfying (iii) and (iv). This follows from the fact that each η_k is a bounded function. □

Definition Suppose $(A, m, Q, \xi^{(0)}, \ldots, \xi^{(m)})$ satisfies Assumption 4. For $0 \le j \ne i \le m$, define

$$q_{ij}(x) = Q_{ij} \frac{\alpha_j(x)}{\alpha_i(x)}. \tag{9}$$

Note that $q_{ij} \in C(E)$. Let $S = E \times E_0$. Define $B \subset \overline{C}(S) \times C(S)$ by

$$Bf(x,i) = Af(x,i) + \sum_{j \ne i} q_{ij}(x)(f(x,j) - f(x,i)), \tag{10}$$

where we take

$$\mathcal{D}(B) = \{f(x,i) = f_1(x) f_2(i) : f_1 \in \mathcal{D}(A), f_2 \in B(E_0)\} \tag{11}$$

In particular, $Af(x,i) = f_2(i) A f_1(x)$.

For each $i \in E_0$, define the measure $\alpha(i, \cdot)$ on E by

$$\alpha(i, \Gamma) = \int_\Gamma \alpha_i(x) \varpi(dx), \tag{12}$$

for all $\Gamma \in \mathcal{B}(E)$. Note that by (8), (7), and (6), these are probability measures. □

2.2 Construction of the coupling

We are now ready to construct our coupled pair, (X,Y), which will have generator B, to prove, for appropriate initial conditions, that the marginal process Y is a Markov chain with generator Q, and to establish our conditional probability formulas. We first require two lemmas.

Lemma 1 *In the setting of Definition 6, let X be a cadlag solution of the martingale problem for A. Then there exists a cadlag process Y such that (X,Y) solves the (local) martingale problem for B. If X is Markov, then (X,Y) is Markov. If the martingale problem for A is well-posed, then the martingale problem for B is well-posed.*

Remark We are not requiring the q_{ij} to be bounded, so for the process we construct,

$$f(X(t),Y(t)) - f(X(0),Y(0)) - \int_0^t Bf(X(s),Y(s))\,ds$$

may only be a local martingale. □

Proof (Proof of Lemma 1) Let $X(t)$ be a cadlag solution to the martingale problem for A. Let $\{N_{ij} : i,j \in E_0, i \neq j\}$ be a family of independent unit rate Poisson processes, which is independent of X. Then the equation

$$Y(t) = k + \sum_{i\neq j}(j-i)N_{ij}\left(\int_0^t 1_{\{i\}}(Y(s))q_{ij}(X(s))\,ds\right) \tag{13}$$

has a unique solution, and as in [12], the process $Z = (X,Y)$ is a solution of the (local) martingale problem for B. If X is Markov, the uniqueness of the solution of (13) ensures that (X,Y) is Markov. Similarly, A well-posed implies B is well posed.□

Lemma 2 *Let A satisfy Assumption 1. Taking $\psi(x,i) = 1 + \sum_{j\neq i} q_{ij}(x) \geq 1$, if A satisfies Condition 17, then B satisfies Condition 17 with E replaced by $S = E \times E_0$.*

Proof Since $\mathcal{D}(A)$ is closed under multiplication, $\mathcal{D}(B)$ defined in (11) is closed under multiplication.

Since we are assuming that $\mathcal{R}(A) \subset \bar{C}(E)$, for each $f \in \mathcal{D}(B)$, there exists $c_f > 0$ such that $|Bf(x,i)| \leq c_f\psi(x)$.

Condition 17(iii) for A and the separability of $B(E_0)$ implies Condition 17(iii) for B_0.

Since A is a pre-generator and B is a perturbation of A by a jump operator, B_0 is a pre-generator. □

Theorem Suppose A satisfies Condition 17 and $(A,m,Q,\xi^{(1)},\ldots,\xi^{(m)})$ satisfies Assumption 4. Let B be given by (10) and for $p_i \geq 0$, $\sum_{i=0}^m p_i = 1$, define

$$\nu(\Gamma \times \{i\}) = p_i\alpha(i,\Gamma), \quad \Gamma \in \mathcal{B}(E), i \in E_0.$$

If $\widetilde{Y}$ is a cadlag E_0-valued Markov chain with generator Q and initial distribution $\{p_i\}$, then there exists a solution (X,Y) of the martingale problem for (B,ν) such that Y and $\widetilde{Y}$ have the same distribution on $D_{E_0}[0,\infty)$, and

$$P(X(t)\in\Gamma \mid \mathcal{F}_t^Y)=\alpha(Y(t),\Gamma), \tag{14}$$

for all $t\geq 0$ and all $\Gamma\in\mathcal{B}(E)$. □

Proof We apply Theorem 19 to the operator $B\subset\overline{C}(S)\times C(S)$.

Let $\gamma : S\to E_0$ be the coordinate projection. Let $\widetilde{\alpha}$ be the transition function from E_0 into S given by the product measure $\widetilde{\alpha}(i,\cdot)=\alpha(i,\cdot)\otimes\delta_i^{E_0}$, where $\alpha(i,\cdot)$ is given by (12). Then $\widetilde{\alpha}(i,\gamma^{-1}(i))=1$ and

$$\widetilde{\psi}(i)\equiv\int_S\psi(z)\widetilde{\alpha}(i,dz)=\int_E\psi(x,i)\alpha_i(x)\varpi(dx)=1+\sum_{j\neq i}Q_{ij}<\infty,$$

for each $i\in E_0$. Define

$$C=\left\{\left(\int_S f(z)\widetilde{\alpha}(\cdot,dz),\int_S Bf(z)\widetilde{\alpha}(\cdot,dz)\right): f\in\mathcal{D}(B)\right\}\subset\mathbb{R}^{m+1}\times\mathbb{R}^{m+1}.$$

The result follow by Theorem 19, if we can show that $Cv=Qv$ for every vector $v\in\mathcal{D}(C)$. Given $f\in\mathcal{D}(B)$, let

$$\overline{f}(i)=\int_S f(z)\widetilde{\alpha}(i,dz)=\int_E f(x,i)\alpha(i,dx)=\int_E f(x,i)\alpha_i(x)\varpi(dx).$$

Note that

$$C\overline{f}(i)=\int_E Bf(x,i)\alpha_i(x)\varpi(dx).$$

Since $\lambda_0=0$, by (5) and the definition of $q_{ij}(x)$,

$$C\overline{f}(i)=-\sum_{k=0}^m\xi_i^{(k)}\lambda_k\int_E f(x,i)\eta_k(dx)+\sum_{j\neq i}Q_{ij}\int_E\alpha_j(x)(f(x,j)-f(x,i))\varpi(dx).$$

By assumption $Q\xi^{(k)}=-\lambda_k\xi^{(k)}$, so $-\xi_i^{(k)}\lambda_k=\sum_{j=0}^m Q_{ij}\xi_j^{(k)}$ and

$$\begin{aligned}-\sum_{k=0}^m\xi_i^{(k)}\lambda_k\int_E f(x,i)\eta_k(dx)&=\sum_{k=0}^m\sum_{j=0}^m Q_{ij}\xi_j^{(k)}\int_E f(x,i)\eta_k(dx)\\&=\sum_{j=0}^m Q_{ij}\sum_{k=0}^m\xi_j^{(k)}\int_E f(x,i)\eta_k(dx)\\&=\sum_{j=0}^m Q_{ij}\int_E f(x,i)\alpha_j(x)\varpi(dx).\end{aligned}$$

This gives

$$C\overline{f}(i) = Q_{ii} \int_E f(x,i)\alpha_i(x)\varpi(dx) + \sum_{j\neq i} Q_{ij} \int_S f(x,j)\alpha_j(x)\varpi(dx)$$
$$= \sum_{j=0}^{m} Q_{ij}\overline{f}(j) = Q\overline{f}(i).$$

It follows that $\widetilde{Y}$ is a solution to the martingale problem for (C,p).

By Theorem 19(a), there exists a solution $Z = (X,Y)$ of the martingale problem for (B,ν) such that $Y = \gamma(Z)$ and $\widetilde{Y}$ have the same distribution on $D_{E_0}[0,\infty)$. Theorem 19(b) implies (14). □

Remark In what follows, we may still write expectations with the notation E^x or E^i, even when we have a coupled process, (X,Y). The meaning will be determined by context, depending on whether the integrand of the expectation involves only X or only Y. □

3 Reversible diffusions

3.1 Assumptions on the potential function

We now consider the special case of our coupling when X is a reversible diffusion on $\mathbb{R}^d$ driven by a potential function F and a small white noise perturbation. We will need to use several results from the literature about the eigenvalues and eigenfunctions of the generator of X. We assume the following on F.

Assumption (i) $F \in C^3(\mathbb{R}^d)$ and $\lim_{|x|\to\infty} F(x) = \infty$.
(ii) F has $m+1 \geq 2$ local minima $\mathcal{M} = \{x_0,\ldots,x_m\}$.
(iii) There exist constants $a_i > 0$ and $c_i > 0$ such that $a_2 < 2a_1 - 2$, and

$$c_1|x|^{a_1} - c_2 \leq |\nabla F(x)|^2 \leq c_3|x|^{a_2} + c_4, \tag{15}$$
$$c_1|x|^{a_1} - c_2 \leq (|\nabla F(x)| - 2\Delta F(x))^2 \leq c_3|x|^{a_2} + c_4. \tag{16}$$

Remark Note that $2 < a_1 \leq a_2$. To see this, observe that (15) implies $a_1 \leq a_2$. Thus, $a_1 \leq a_2 < 2a_1 - 2$, which implies $a_1 > 2$. □

Lemma 3 *Under Assumption 10, there exist constants* $\widetilde{c}_i > 0$ *such that*

$$\widetilde{c}_1|x|^{\widetilde{a}_1} - \widetilde{c}_2 \leq |F(x)| \leq \widetilde{c}_3|x|^{\widetilde{a}_2} + \widetilde{c}_4, \tag{17}$$

where $\widetilde{a}_i = a_i/2 + 1$.

Proof Since

$$F(x) = F(0) + \int_0^1 \nabla F(sx) \cdot x\, ds,$$

it follows from (15) that

$$|F(x)| \le |F(0)| + |x|(c_3|x|^{a_2} + c_4)^{1/2},$$

and the upper bound in (17) follows immediately.

Since $F \to \infty$, there exists $C > 0$ such that $F(x) > -C$ for all $x \in \mathbb{R}^d$, and since $|\nabla F| \to \infty$, there exists $R > 0$ such that $|\nabla F(x)| \ge 1$ whenever $|x| \ge R$.

Recall that φ_x satisfies $\varphi_x' = -\nabla F(\varphi_x)$ and $\varphi_x(0) = x$, and define

$$T_x = \inf\{t \ge 0 : |\varphi_x(t)| < R\}.$$

Suppose there exists x such that $T_x = \infty$. Then, for all $t > 0$,

$$\begin{aligned} -C < F(\varphi_x(t)) &= F(x) + \int_0^t \nabla F(\varphi_x(s)) \cdot \varphi_x'(s)\, ds \\ &= F(x) - \int_0^t |\nabla F(\varphi_x(s))|^2\, ds \\ &\le F(x) - t. \end{aligned}$$

Therefore, $F(x) \ge t - C$ for all t, a contradiction, and we must have $T_x < \infty$ for all $x \in \mathbb{R}^d$.

Let $L = \sup_{|x| \le R} F(x)$. By (15) and the fact that $F \to \infty$, we may choose $R' \ge R$ and $C' > 0$ such that $F(x) > L$ and $|\nabla F(x)| \ge C'|x|^{a_1/2}$ whenever $|x| > R'$.

Fix $x \in \mathbb{R}^d$ with $|x| > 2R'$, so that $F(x) > L$. Since $|\varphi_x(T_x)| = R$, it follows that $F(\varphi_x(T_x)) \le L$. By the continuity of φ_x, we may choose $T' \in (0, T_x]$ such that $F(\varphi_x(T')) = L$. We then have

$$\begin{aligned} L &= F(x) + \int_0^{T'} \nabla F(\varphi_x(t)) \cdot \varphi_x'(t)\, dt \\ &= F(x) - \int_0^{T'} |\nabla F(\varphi_x(t))||\varphi_x'(t)|\, dt. \end{aligned}$$

Let $T'' = \inf\{t \ge 0 : |\varphi_x(t)| < |x|/2\}$. Note that $F(\varphi_x(T')) = L$ implies $|\varphi_x(T')| \le R' < |x|/2$, and therefore $T'' \le T'$. Moreover, for all $t < T''$, we have $|\varphi_x(t)| \ge |x|/2 > R'$, which implies

$$|\nabla F(\varphi_x(t))| \ge C'|\varphi_x(t)|^{a_1/2} \ge C' \left(\frac{|x|}{2}\right)^{a_1/2}.$$

Thus,

$$L \le F(x) - C' \left(\frac{|x|}{2}\right)^{a_1/2} \int_0^{T''} |\varphi_x'(t)|\, dt.$$

But $\int_0^{T''} |\varphi_x'(t)|\, dt$ is the length of φ_x from $t = 0$ to $t = T''$, which is bounded below by

$$|\varphi_x(T'') - \varphi_x(0)| \geq |\varphi_x(0)| - |\varphi_x(T'')| = |x| - \frac{|x|}{2} = \frac{|x|}{2}.$$

Therefore, for all $|x| > 2R'$, we have $F(x) \geq C''|x|^{a_1/2+1} - |L|$, where $C'' = 2^{-a_1/2-1}C'$, and this proves the lower bound in (17). □

3.2 Spectral properties of the generator

Having established our assumptions on F, we now turn our attention to the diffusion process, X_ε, given by (1). To simplify notation, we may sometimes omit the ε. The process X has generator $A = \varepsilon\Delta - \nabla F \cdot \nabla$. To show that A meets the requirements of our coupling from Section 2, we must prove certain results about its eigenvalues and eigenfunctions. For this, we begin with some notation, a lemma, and two results from the literature.

Define $\pi(x) = \pi_\varepsilon(x) = e^{-F(x)/2\varepsilon}$. Let

$$V = V_\varepsilon := \frac{\Delta\pi}{\pi} = \frac{1}{4\varepsilon^2}|\nabla F|^2 - \frac{1}{2\varepsilon}\Delta F. \tag{18}$$

Lemma 4 *Let V_ε be given by* (18)*, where F satisfies Assumption 10. Recall the constants a_i from* (15)-(16)*. For all $\varepsilon \in (0,1)$, there exist constants $c_{i,\varepsilon} > 0$ such that*

$$c_{1,\varepsilon}|x|^{a_1} - c_{2,\varepsilon} \leq V_\varepsilon(x) \leq c_{3,\varepsilon}|x|^{a_2} + c_{4,\varepsilon}.$$

In particular, $V_\varepsilon \to \infty$ for all $\varepsilon \in (0,1)$.

Proof Fix $\varepsilon \in (0,1)$. By (15) and (16), for x sufficiently large,

$$c|x|^{a_1} \leq (|\nabla F(x)| - 2\Delta F)^2 \leq C|x|^{a_2},$$

and

$$c|x|^{a_1} \leq |\nabla F(x)|^2 \leq C|x|^{a_2},$$

for some $0 < c \leq C < \infty$. Note that

$$4V_1 = |\nabla F|^2 - 2\Delta F = (|\nabla F| - 2\Delta F) + (|\nabla F|^2 - |\nabla F|).$$

Hence, for x sufficiently large, $V_1(x) \leq C_1|x|^{a_2}$. Also,

$$V_1(x) \geq \frac{1}{4}(c|x|^{a_1} - C_2|x|^{a_2/2}).$$

Since $a_1 > a_2/2$, it follows that for x sufficiently large, $V_1(x) \geq \widetilde{c}|x|^{a_1}$. Therefore, there exist constants $\widetilde{c_i} > 0$ such that

$$\widetilde{c}_1|x|^{a_1} - \widetilde{c}_2 \le V_1(x) \le \widetilde{c}_3|x|^{a_2} + \widetilde{c}_4,$$

and

$$\widetilde{c}_1|x|^{a_1} - \widetilde{c}_2 \le |\nabla F(x)|^2 \le \widetilde{c}_3|x|^{a_2} + \widetilde{c}_4,$$

for all $x \in \mathbb{R}^d$. Note that

$$V_\varepsilon = \frac{1}{\varepsilon}\left(V_1 + \left(\frac{1-\varepsilon}{4\varepsilon}\right)|\nabla F|^2\right),$$

so that

$$\frac{1}{\varepsilon}V_1 \le V_\varepsilon \le \frac{1}{\varepsilon}V_1 + \frac{1}{\varepsilon^2}|\nabla F|^2.$$

From here, the lemma follows easily. □

The following two theorems are from [6]. Theorem 12 is a consequence of [6, Theorem 4.5.4] and [6, Lemma 4.2.2]. Theorem 13 is part of [6, Theorem 2.1.4].

Theorem Let $H = -\Delta + W$, where W is continuous with $W \to \infty$. Let λ denote the smallest eigenvalue of H, and ψ the corresponding eigenfunction, normalized so that $\|\psi\|_{L^2(\mathbb{R}^d)} = 1$. Define $Uf = \psi f$ and $\widetilde{H} = U^{-1}(H - \lambda)U$. If

$$\widehat{c}_1|x|^{\widehat{a}_1} - \widehat{c}_2 \le |W(x)| \le \widehat{c}_3|x|^{\widehat{a}_2} + \widehat{c}_4,$$

where $\widehat{a}_i > 0$, $\widehat{c}_i > 0$, and $\widehat{a}_2 < 2\widehat{a}_1 - 2$, then $e^{-\widetilde{H}t}$ is an ultracontractive symmetric Markov semigroup on $L^2(\mathbb{R}^d, \psi(x)^2\,dx)$. That is, for each $t \ge 0$, the operator $e^{-\widetilde{H}t}$ is a bounded operator mapping $L^2(\mathbb{R}^d, \psi(x)^2\,dx)$ to $L^\infty(\mathbb{R}^d, \psi(x)^2\,dx)$. □

Theorem Let e^{-Ht} be an ultracontractive symmetric Markov semigroup on $L^2(\Omega, \mu)$, where Ω is a locally compact, second countable Hausdorff space and μ is a Borel measure on Ω. If $\mu(\Omega) < \infty$, then each eigenfunction of H belongs to $L^\infty(\Omega, \mu)$. □

This next proposition establishes the spectral properties of A that are needed to carry out the construction of our coupling.

Proposition Fix $\varepsilon > 0$. The operator $H = -\Delta + V_\varepsilon$ is a self-adjoint operator on $L^2(\mathbb{R}^d)$ with discrete, nonnegative spectrum $\widehat{\lambda}_k \uparrow \infty$ and corresponding orthonormal eigenfunctions ψ_k. Each ψ_k is locally Hölder continuous. Moreover, $\widehat{\lambda}_0 = 0$ is simple and ψ_0 is proportional to π. We define μ by $\mu(dx) = \pi(x)^2\,dx$ and $\varpi = Z^{-1}\mu$, where $Z = \mu(\mathbb{R}^d)$. The operator $\widetilde{H}$ given by $\widetilde{H}f = \pi^{-1}H(\pi f)$ is a self-adjoint operator on $L^2(\varpi)$ with eigenvalues $\widehat{\lambda}_k$ and orthogonal eigenfunctions $\widehat{\eta}_k = \psi_k/\pi$. The functions $\widehat{\eta}_k$ have norm one in $L^2(\mu)$, whereas the functions $\eta_k = Z^{1/2}\widehat{\eta}_k$ have norm one in $L^2(\varpi)$.

For $f \in C_c^\infty(\mathbb{R}^d)$, we have $-\varepsilon\widetilde{H}f = \varepsilon\Delta f - \nabla F \cdot \nabla f$. Hence, if we define A by

$$A = \{(f, -\varepsilon\widetilde{H}f) : f \in C_c^\infty(\mathbb{R}^d)\},$$

then A is the generator for the diffusion process given by (1). For each $x \in \mathbb{R}^d$, (1) has a unique, global solution for all time, so that the process X with $X(0) = x$ is a solution to the martingale problem for (A, δ_x). The operator A is graph separable, and $\mathcal{D}(A)$ is separating and closed under multiplication. The measure ϖ is a stationary distribution for A. Moreover,

$$\int Af\, d\varpi_k = -\lambda_k \int f\, d\varpi_k,$$

where $\varpi_k(dx) = \eta_k(x)\varpi(dx)$ and $\lambda_k = \varepsilon\widehat{\lambda}_k$. The signed measures ϖ_k satisfy $\varpi_k(\mathbb{R}^d) = 0$, and each η_k belongs to $\overline{C}(\mathbb{R}^d)$, the space of bounded, continuous functions on $\mathbb{R}^d$. □

Proof Note that $V \to \infty$ by Lemma 4. Therefore, by [18, Theorem XIII.67], we have that H is a self-adjoint operator on $L^2(\mathbb{R}^d)$ with compact resolvent. It follows (see [6, pp. 108–109, 119–120, and Proposition 1.4.3]) that H has a purely discrete spectrum and there exists a complete, orthonormal set of eigenfunctions $\{\psi_k\}_{k=0}^{\infty}$ with corresponding eigenvalues $\widehat{\lambda}_k \uparrow \infty$. Moreover, $\widehat{\lambda}_0$ is simple and ψ_0 is strictly positive.

Since V is locally bounded, and $(-\Delta + V - \widehat{\lambda}_k)\psi_k = 0$, [10, Theorem 8.22] implies that, for each compact $K \subset \mathbb{R}^d$, ψ_k is Hölder continuous on K with exponent $\gamma(K)$.

Define $U : L^2(\mu) \to L^2(\mathbb{R}^d)$ by $Uf = \pi f$, so that $\widetilde{H} = U^{-1}HU$. Since U is an isometry, $\widetilde{H}$ is self-adjoint on $L^2(\mu)$ and has the same eigenvalues as H. Note that, for any $f \in \mathcal{D}(\widetilde{H})$, it follows from Green's identity that

$$\begin{aligned}\langle f, \widetilde{H}f\rangle_{L^2(\mu)} = \langle \pi f, H(\pi f)\rangle_{L^2(\mathbb{R}^d)} &= \int |\nabla(\pi f)|^2 + \int V(\pi f)^2 \\ &= \int |\nabla(\pi f)|^2 + \int (\Delta\pi)\pi f^2 = \int |\nabla(\pi f)|^2 - \int \nabla\pi \cdot \nabla(\pi f^2).\end{aligned}$$

Using the product rule, $\nabla(gh) = g\nabla h + h\nabla g$, this simplifies to

$$\begin{aligned}\langle f, \widetilde{H}f\rangle_{L^2(\mu)} &= \int (|\nabla\pi|^2 f^2 + 2f\pi(\nabla f \cdot \nabla\pi) + |\nabla f|^2\pi^2 - |\nabla\pi|^2 f^2 - \pi(\nabla(f^2)\cdot\nabla\pi)) \\ &= \int (2f\pi(\nabla f\cdot\nabla\pi) + |\nabla f|^2\pi^2 - \pi(\nabla(f^2)\cdot\nabla\pi)) = \int |\nabla f|^2\pi^2,\end{aligned}$$

showing that $\widetilde{H}$ cannot have a negative eigenvalue. Hence, $\widehat{\lambda}_0 \geq 0$.

By (17), we have $\pi \in L^2(\mathbb{R}^d)$, so that $\pi \in \mathcal{D}(H)$ with $H\pi = 0$. Hence, since $\widehat{\lambda}_0$ is nonnegative and has multiplicity one, it follows that $\widehat{\lambda}_0 = 0$ and ψ_0 is proportional to π.

Observe that, if $f \in C_c^\infty$, then, using the product rule for the Laplacian and the identity $V = \Delta\pi/\pi$, we have

$$-\widetilde{H}f = -\frac{1}{\pi}H(\pi f) = \frac{1}{\pi}(\Delta(\pi f) - V\pi f) = \frac{1}{\pi}(f\Delta\pi + 2\nabla\pi\cdot\nabla f + \pi\Delta f - f\Delta\pi).$$

Since $2\varepsilon\nabla\pi/\pi = -\nabla F$, we have $-\varepsilon\widetilde{H}f = \varepsilon\Delta f - \nabla F\cdot\nabla f$.

Since ∇F is locally Lipschitz, (1) has a unique solution up to an explosion time (see [17, Theorem V.38]). Since $\lim_{|x|\to\infty} F = \infty$ by assumption and $\lim_{|x|\to\infty} AF(x) = \infty$ by Lemma 4.2, it follows that F is a Liapunov function for X_ε proving that X_ε does not explode.

By [13, Remark 2.5], A is graph separable. Clearly $\mathcal{D}(A)$ is closed under multiplication. Since $\mathcal{D}(A)$ separates points and $\mathbb{R}^d$ is complete and separable, $\mathcal{D}(A)$ is separating (see [8, Theorem 3.4.5]).

If $f \in C_c^\infty$, then

$$\int Af\,d\varpi = -\varepsilon\langle 1, \widetilde{H}f\rangle_{L^2(\varpi)} = -\varepsilon\langle \widetilde{H}1, f\rangle_{L^2(\varpi)} = 0,$$

so that ϖ is a stationary distribution for A. For $k \geq 1$, since $\varpi_k(dx) = \eta_k(x)\varpi(dx)$, we have

$$\int Af\,d\varpi_k = -\varepsilon\langle \eta_k, \widetilde{H}f\rangle_{L^2(\varpi)} = -\varepsilon\langle \widetilde{H}\eta_k, f\rangle_{L^2(\varpi)} = -\lambda_k \int f\,d\varpi_k.$$

Also, $\varpi_k(\mathbb{R}^d) = \langle \eta_k, 1\rangle_{L^2(\varpi)} = 0$, since η_k and $\eta_0 = 1$ are orthogonal.

Finally, since $\eta_k = Z^{1/2}\psi_k/\pi$ and ψ_k is locally Hölder continuous, it follows that each η_k belongs to $C(\mathbb{R}^d)$, and the fact that they are bounded follows from Theorems 12 and 13. □

3.3 The coupled process

By Proposition 14, the pair (A, m) satisfies Assumption 1 with $E = \mathbb{R}^d$, so we have the following.

Theorem Let A be the generator for (1) where F satisfies Assumption 10, and let $(-\lambda_0, \eta_0), \ldots, (-\lambda_m, \eta_m)$ be the first $m+1$ eigenvalues and eigenvectors of A. Let $Q \in \mathbb{R}^{(m+1)\times(m+1)}$ be the generator of a continuous-time Markov chain with state space $E_0 = \{0, 1, \ldots, m\}$ and eigenvalues $\{0, -\lambda_1, \ldots, -\lambda_m\}$ and eigenvectors $\xi^{(1)}, \ldots, \xi^{(m)}$ such that α_i defined by (8) is strictly positive. Let B be defined as in Definition 6.

Let $\widetilde{Y}$ be a continuous time Markov chain with generator Q and initial distribution $p = (p_0, \ldots, p_m) \in \mathcal{P}(E_0)$. Then there exists a cadlag Markov process (X, Y) with generator B and initial distribution ν given by

$$\nu(\Gamma \times \{i\}) = p_i \alpha(i, \Gamma), \quad \Gamma \in \mathcal{B}(\mathbb{R}^d), \tag{19}$$

such that Y and $\widetilde{Y}$ have the same distribution on $D_{E_0}[0, \infty)$, and

$$P(X(t) \in \Gamma \mid Y(t) = j) = \int_\Gamma \alpha_j(x)\,\varpi(dx), \tag{20}$$

for all $t \geq 0$, all $0 \leq j \leq m$, and all $\Gamma \in \mathcal{B}(E)$. □

Remark That Q with these properties exists can be seen from [16, Theorem 1]. Remark 5 ensures the existence of the eigenvectors. □

Proof Note that under the assumptions of the theorem, $(A, m, Q, \xi^{(1)}, \ldots, \xi^{(m)})$ satisfies Assumption 4. By Proposition 14, the rest of the hypotheses of Theorem 8 are also satisfied. Consequently, the process (X, Y) exists, and by uniqueness of the martingale problem for B, (X, Y) is Markov. □

We can now construct the coupling sequences described in the introduction. For each $\varepsilon > 0$, choose a matrix Q_ε and eigenvectors $\xi_\varepsilon^{(1)}, \ldots, \xi_\varepsilon^{(m)}$ that satisfy the assumptions of Theorem 15. If $(X_\varepsilon, Y_\varepsilon)$ is the Markov process described in Theorem 15, then the family, $\{(X_\varepsilon, Y_\varepsilon) : \varepsilon > 0\}$, forms a coupling sequence.

The coupling sequence is determined by the collection, $\{Q_\varepsilon, \xi_\varepsilon^{(1)}, \ldots, \xi_\varepsilon^{(m)} : \varepsilon > 0\}$. By making different choices for the matrices and eigenvectors, we can obtain different coupling sequences. In our follow-up paper, we will consider the question of existence and uniqueness of a coupling sequence that satisfies conditions (3) and (4).

Acknowledgements This paper was completed while the first author was visiting the University of California, San Diego with the support of the Charles Lee Powell Foundation. The hospitality of that institution, particularly that of Professor Ruth Williams, was greatly appreciated.

Appendix

Let E be a complete and separable metric space, $\mathcal{B}(E)$ the σ-algebra of Borel subsets of E, and $\mathcal{P}(E)$ the family of Borel probability measures on E. Let $M(E)$ be the collection of all real-valued, Borel measurable functions on E, and $B(E) \subset M(E)$ the Banach space of bounded functions with $\|f\|_\infty = \sup_{x\in E} |f(x)|$. Let $\overline{C}(E) \subset B(E)$ be the subspace of bounded continuous functions, while $C(E)$ denotes the collection of continuous, real-valued functions on E. A collection of functions $D \subset \overline{C}(E)$ is *separating* if $\mu, \nu \in \mathcal{P}(E)$ and $\int f\,d\mu = \int f\,d\nu$ for all $f \in D$ implies $\mu = \nu$.

Condition (i) $B \subset \overline{C}(E) \times C(E)$ and $\mathcal{D}(B)$ is closed under multiplication and separating.

(ii) There exists $\psi \in C(E)$, $\psi \geq 1$, such that for each $f \in \mathcal{D}(B)$, there exists a constant c_f such that

$$|Bf(x)| \leq c_f \psi(x), \quad x \in E.$$

(We write Bf even though we do not exclude the possibility that B is multivalued. In the multivalued case, each element of Bf must satisfy the inequality.)

(iii) There exists a countable subset $B_c \subset B$ such that every solution of the (local) martingale problem for B_c is a solution of the (local) martingale problem for B.

(iv) $B_0 f \equiv \psi^{-1} Bf$ is a pre-generator, that is, B_0 is dissipative and there are sequences of functions $\mu_n : E \to \mathcal{P}(E)$ and $\lambda_n : E \to [0, \infty)$ such that for each

$(f,g) \in B$,

$$g(x) = \lim_{n\to\infty} \lambda_n(x) \int_E (f(y) - f(x))\mu_n(x,dy) \tag{21}$$

for each $x \in E$. □

Remark Condition 17(iii) holds if B_0 is graph-separable, that is, there is a countable subset $B_{0,c}$ of B_0 such that B_0 is a subset of the bounded, pointwise closure of $B_{0,c}$.

An operator is a pre-generator if for each $x \in E$, there exists a solution of the martingale problem for (B, δ_x). □

For a measurable E_0-valued process Y, where E_0 is a complete and separable metric space, let

$$\widehat{\mathcal{F}}_t^Y = \text{completion of } \sigma\left(\int_0^r g(Y(s))\,ds : r \le t, g \in B(E_0)\right) \vee \sigma(Y(0)).$$

Theorem Let (S,d) and (E_0,d_0) be complete, separable metric spaces. Let B satisfy Condition 17. Let $\gamma : S \to E_0$ be measurable, and let $\widetilde{\alpha}$ be a transition function from E_0 into S (that is, $\widetilde{\alpha} : E_0 \times \mathcal{B}(S) \to \mathbb{R}$ satisfies $\widetilde{\alpha}(y,\cdot) \in \mathcal{P}(S)$ for all $y \in E_0$ and $\widetilde{\alpha}(\cdot,\Gamma) \in B(E_0)$ for all $\Gamma \in \mathcal{B}(S)$) satisfying $\int h \circ \gamma(z)\,\widetilde{\alpha}(y,dz) = h(y)$, $y \in E_0$, $h \in B(E_0)$, that is, $\widetilde{\alpha}(y,\gamma^{-1}(y)) = 1$. Assume that $\widetilde{\psi}(y) \equiv \int_S \psi(z)\widetilde{\alpha}(y,dz) < \infty$ for each $y \in E_0$ and define

$$C = \left\{\left(\int_S f(z)\widetilde{\alpha}(\cdot,dz), \int_S Bf(z)\widetilde{\alpha}(\cdot,dz)\right) : f \in \mathcal{D}(B)\right\}.$$

Let $\mu \in \mathcal{P}(E_0)$ and define $\nu = \int \widetilde{\alpha}(y,\cdot)\,\mu(dy)$.

a) If $\widetilde{Y}$ satisfies $\int_0^t E[\widetilde{\psi}(\widetilde{Y}(s))]\,ds < \infty$ a.s. for all $t > 0$ and $\widetilde{Y}$ is a solution of the martingale problem for (C,μ), then there exists a solution Z of the martingale problem for (B,ν) such that $\widetilde{Y}$ has the same distribution on $M_{E_0}[0,\infty)$ as $Y = \gamma \circ Z$. If Y and $\widetilde{Y}$ are cadlag, then Y and $\widetilde{Y}$ have the same distribution on $D_{E_0}[0,\infty)$.
b) Let $\mathbf{T}^Y = \{t : Y(t) \text{ is } \widehat{\mathcal{F}}_t^Y \text{ measurable}\}$ (which holds for Lebesgue-almost every t). Then for $t \in \mathbf{T}^Y$,

$$P(Z(t) \in \Gamma \mid \widehat{\mathcal{F}}_t^Y) = \widetilde{\alpha}(Y(t),\Gamma), \quad \Gamma \in \mathcal{B}(S).$$

c) If, in addition, uniqueness holds for the martingale problem for (B,ν), then uniqueness holds for the $M_{E_0}[0,\infty)$-martingale problem for (C,μ). If $\widetilde{Y}$ has sample paths in $D_{E_0}[0,\infty)$, then uniqueness holds for the $D_{E_0}[0,\infty)$-martingale problem for (C,μ).
d) If uniqueness holds for the martingale problem for (B,ν), then Y restricted to $\mathbf{T}^Y$ is a Markov process. □

Remark If Y is cadlag with no fixed points of discontinuity (that is $Y(t) = Y(t-)$ a.s. for all t), then $\widehat{\mathcal{F}}_t^Y = \mathcal{F}_t^Y$ for all t. □

Remark The main precursor of this Markov mapping theorem is [13, Corollary 3.5]. The result stated here is a special case of Corollary 3.3 of [11]. □

References

1. Luca Avena, Fabienne Castell, Alexandre Gaudillière, and Clothilde Melot. Approximate and exact solutions of intertwining equations through random spanning forests, 2017.
2. J. Beltrán and C. Landim. Tunneling and metastability of continuous time Markov chains. *J. Stat. Phys.*, 140(6):1065–1114, 2010.
3. J. Beltrán and C. Landim. Tunneling and metastability of continuous time Markov chains II, the nonreversible case. *J. Stat. Phys.*, 149(4):598–618, 2012.
4. Anton Bovier, Michael Eckhoff, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4):399–424, 2004.
5. Anton Bovier, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7(1):69–99, 2005.
6. E. B. Davies. *Heat kernels and spectral theory*, volume 92 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1990.
7. Michael Eckhoff. Precise asymptotics of small eigenvalues of reversible diffusions in the metastable regime. *Ann. Probab.*, 33(1):244–299, 2005.
8. Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.
9. M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szücs.
10. David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1983.
11. T. G. Kurtz and G. Nappo. The filtered martingale problem. In *The Oxford handbook of nonlinear filtering*, pages 129–165. Oxford Univ. Press, Oxford, 2011.
12. Thomas G. Kurtz. *Approximation of population processes*, volume 36 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1981.
13. Thomas G. Kurtz. Martingale problems for conditional distributions of Markov processes. *Electron. J. Probab.*, 3:no. 9, 29 pp. (electronic), 1998.
14. Thomas G. Kurtz and Richard H. Stockbridge. Stationary solutions and forward equations for controlled and singular martingale problems. *Electron. J. Probab.*, 6:no. 17, 52 pp. (electronic), 2001.
15. Claudio Landim. Metastable Markov chains. *Probab. Surv.*, 16:143–227, 2019.
16. Hazel Perfect. Methods of constructing certain stochastic matrices. *Duke Math. J.*, 20:395–404, 1953.
17. Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.
18. Michael Reed and Barry Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978.
19. Fraydoun Rezakhanlou and Insuk Seo. Scaling limit of small random perturbation of dynamical systems, 2018.