Electronic Journal of Probability

Vol. 1 (1996) Paper no. 9, pages 1–21.

Journal URL: http://www.math.washington.edu/˜ejpecp/

Paper URL: http://www.math.washington.edu/˜ejpecp/EjpVol1/paper9.abs.html

Quantitative bounds for convergence rates of continuous time Markov processes

by

Gareth O. Roberts* and Jeffrey S. Rosenthal**

Abstract. We develop quantitative bounds on rates of convergence for continuous-time Markov processes on general state spaces. Our methods involve coupling and shift-coupling, and make use of minorization and drift conditions. In particular, we use auxiliary coupling to establish the existence of small (or pseudo-small) sets. We apply our method to some diffusion examples. We are motivated by interest in the use of Langevin diffusions for Monte Carlo simulation.

Keywords. Markov process, rates of convergence, coupling, shift-coupling, minorization condition, drift condition.

AMS Subject Classification. 60J25.

Submitted to EJP on January 29, 1996. Final version accepted on May 28, 1996.

* Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K. Internet: G.O.Roberts@statslab.cam.ac.uk.

** Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: jeff@utstat.toronto.edu.

1. Introduction.

This paper concerns quantitative bounds on the time to stationarity of continuous time Markov processes, in particular diffusion processes.

Quantitative bounds for discrete time Markov chains have recently been studied by several authors, using drift conditions and minorization conditions (i.e. small sets) to establish couplings or shift-couplings for the chains. Bounding the distance of $\mathcal{L}(X_t)$ to stationarity for a fixed t has been considered by Meyn and Tweedie (1994), Rosenthal (1995), and Baxendale (1994). In Roberts and Rosenthal (1994), periodicity issues are avoided by instead bounding the distance of ergodic average laws $\sum_{k=1}^{t} \mathcal{L}(X_k)$ to stationarity. In each of these cases, quantitative bounds are obtained in terms of drift and minorization conditions for the chain.

In this paper we extend these results to honest (i.e. stochastic) continuous time Markov processes. We derive general bounds, using coupling and shift-coupling, which are similar to the discrete case (Section 2). Drift conditions are defined in terms of infinitesimal generators, and are fairly straightforward. However, the task of establishing minorization conditions is rather less clear, and this is the heart of our paper. We approach this problem for both one-dimensional (Section 3) and multi-dimensional (Section 4) diffusions, by producing an auxiliary coupling of certain processes started at different points of the proposed small set, which has probability $\epsilon > 0$ of being successful by some fixed time t_0. This implies (Theorem 7) the existence of a corresponding minorization condition. Our construction relies on the use of "medium sets" on which the drifts of the diffusions remain bounded. It makes use of the Bachelier-Lévy formula (Lévy, 1965; Lerche, 1986) to lower-bound the probability ϵ of coupling.

For one-dimensional diffusions, we are able to simultaneously couple an entire collection of processes, started from each point of the proposed small set. However, for multi-dimensional diffusions, this is not possible. Instead, we couple the processes pairwise, thus establishing the existence of a *pseudo-small set* (Section 4) rather than an actual small set. We show that we can still use such a construction to obtain a coupling, even though there is no longer a regenerative structure. This suggests a certain advantage to studying minorization conditions through the use of coupling (e.g. Nummelin, 1992, Section III.10;

Rosenthal, 1995) rather than the use of regeneration times as is often done (e.g. Athreya and Ney, 1978; Nummelin, 1984; Asmussen, 1987).

We apply our results to some examples of Langevin diffusions in Section 5.

In discrete time, studies of quantitative convergence rates have been motivated largely by Markov chain Monte Carlo algorithms (see Gelfand and Smith, 1990; Smith and Roberts, 1993). The current study is motivated partially by recent work (Grenander and Miller, 1994; Philips and Smith, 1994; Roberts and Tweedie, 1995; Roberts and Rosenthal, 1995) considering the use of Langevin diffusions for Monte Carlo simulations.

2. Bounds involving minorization and drift conditions.

We begin with some general results related to convergence rates of positive recurrent, continuous time Markov processes to stationarity.

Let $P^t(x,\cdot)$ be the transition probabilities for a Markov process on a general state space $\mathcal{X}$. We say that a subset $C \subseteq \mathcal{X}$ is (t,ϵ)-*small*, for a positive time t and $\epsilon > 0$, if there exists a probability measure $Q(\cdot)$ on $\mathcal{X}$, satisfying the *minorization condition*

$$P^t(x,\cdot) \;\geq\; \epsilon\, Q(\cdot) \qquad x \in C\,. \tag{1}$$

The subset C is *small* if it is (t,ϵ)-small for some positive t and ϵ. (For background on small sets and the related notion of Harris chains, see Nummelin, 1978; Meyn and Tweedie, 1993b; Asmussen, 1987, pp. 150–158; Lindvall, 1992, pp. 91–92.)

The advantage of small sets, for our purposes, is that they can be used (together with information about the return times to C) to establish couplings and shift-couplings of Markov processes, leading to bounds on convergence rates of processes to their stationary distributions. This is a well-studied idea; for extensive background, see Nummelin (1992, pp. 91–98).

For simplicity, we begin with general results related to shift-coupling, which provides for bounds on the ergodic averages of distances to stationary distributions. (Corresponding results for ordinary coupling are considered in Theorem 3 and Corollary 4.) These results are analogous to the discrete time results of Roberts and Rosenthal (1994). They bound the total variation distance between various probability measures, defined by

$$\|\mu - \nu\| \;\equiv\; \sup_{S \subseteq \mathcal{X}} |\mu(S) - \nu(S)| \,.$$

The proofs are deferred until after the statements of all of the initial results.

Theorem 1. *Given a Markov process with transition probabilities $P^t(x,\cdot)$ and stationary distribution $\pi(\cdot)$, suppose $C \subseteq \mathcal{X}$ is (t^*, ϵ)-small, for some positive time t^*, and $\epsilon > 0$. Suppose further that there is $\delta > 0$ and a non-negative function $V : \mathcal{X} \to \mathbf{R}$ with $V(x) \geq 1$ for all $x \in \mathcal{X}$, such that*

$$\mathbf{E}_x\left(e^{\delta \tau_C}\right) \;\leq\; V(x)\,, \qquad x \notin C \tag{2}$$

where τ_C is the first hitting time of C. Set $A = \sup_{x \in C} \mathbf{E}_x(V(X_{t^*}))$, and assume that $A < \infty$. Then for $t > 0$, and for any $0 < r < 1/t^*$ with $Ae^{\delta t^* - \delta/r} < 1$,

$$\| \int_0^t \mathbf{P}(X_s \in \cdot)ds \;-\; \pi(\cdot)\| \;\leq\; \frac{1}{t}\left(\frac{2}{r\epsilon} + A^{-1}(\mathbf{E}(V(X_0) + \mathbf{E}_\pi(V))\frac{e^{\delta/r}}{r(1 - Ae^{\delta t^* - \delta/r})}\right)\,.$$

It can be difficult to directly establish bounds on return times such as (2). It is often easier to establish *drift conditions*. Thus, we now consider using drift conditions and generators to imply (2). We require some notation. We let $\mathcal{A}$ be the weak generator of our Markov process, and let $\mathcal{D}$ be the "one-sided extended domain" consisting of all functions $U : \mathcal{X} \to \mathbf{R}$ which satisfy the one-sided Dynkin's formula

$$\mathbf{E}_x\left(U(X_t)\right) \;\leq\; \mathbf{E}_x\left(\int_0^t \mathcal{A}U(X_s)\,ds\right) \;+\; U(x)\,.$$

This formula holds with equality if U is in the domain of the strong generator; furthermore, it holds with inequality if U is in the domains of certain stopped versions of the diffusion. For background and discussion see, e.g., Meyn and Tweedie (1993a). For smooth functions in concrete examples, Dynkin's formula can be verified directly using Itô's formula and the dominated convergence theorem, as we do in Section 5.

Corollary 2. *Suppose* (1) *holds, and in addition there is a function* $U \in \mathcal{D}$ *with* $U(x) \geq 1$ *for all* $x \in \mathcal{X}$*, and* $\delta > 0$ *and* $\Lambda < \infty$*, such that*

$$\mathcal{A}U(x)) \leq -\delta U(x) + \Lambda \mathbf{1}_C(x)\,, \qquad x \in \mathcal{X}\,.$$

Then (2) *holds for this* δ*, with* $V = U$*, and hence the conclusion of Theorem 1 holds with* $V = U$*. Furthermore, we have* $A \leq \frac{\Lambda}{\delta} + e^{-\delta t^*} \sup\limits_{x \in C} U(x)$.

Remark. Under the hypothesis of Corollary 2 (or similarly of Corollary 4 below), we can also derive a bound on $\mathbf{E}_\pi(U)$. Indeed, we have that

$$\mathbf{E}_x U(X_t) - U(x) \;\leq\; \mathbf{E}_x \int_0^t \mathcal{A}U(X_s)ds \;\leq\; \mathbf{E}_x \int_0^t (-\delta\, U(X_s) + \Lambda)ds\,.$$

Integrating both sides with respect to $\pi(dx)$ and using the stationarity of π, we obtain that $0 \leq -\delta \mathbf{E}_\pi(U) + \Lambda$, so that

$$\mathbf{E}_\pi(U) \;\leq\; \Lambda/\delta\,.$$

This may be of help in cases where the stationary distribution $\pi(\cdot)$ is too complicated for direct computation, as is often the case in the Markov chain Monte Carlo context.

We now consider results which directly bound $\|\mathcal{L}(X_t) - \pi(\cdot)\|$, rather than bounding ergodic averages. The statements are similar, except that the return-time conditions are more involved since we now need two copies of the process to *simultaneously* return to C. These results are analogous to the discrete-time results of Rosenthal (1995).

Theorem 3. *Given a Markov process with transition probabilities* $P^t(x, \cdot)$ *and stationary distribution* $\pi(\cdot)$*, suppose* $C \subseteq \mathcal{X}$ *is* (t^*, ϵ)*-small, for some positive time* t^**, and* $\epsilon > 0$*. Suppose further that there is* $\delta > 0$ *and a non-negative function* $h : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ *with* $h(x, y) \geq 1$ *for all* $x \in \mathcal{X}$*, such that*

$$\mathbf{E}_{x,y}\left(e^{\delta \tau_{C\times C}}\right) \;\leq\; h(x, y)\,, \qquad (x, y) \notin C \times C \tag{3}$$

where $\tau_{C\times C}$ *is the first hitting time of* $C \times C$*. Set* $A = \sup_{(x,y)\in C\times C} \mathbf{E}_{x,y}(h(X_{t^*}, Y_{t^*}))$*, where* $\{X_t\}$ *and* $\{Y_t\}$ *are defined jointly as described in the proof, and assume that* $A < \infty$*.*

Then for $t > 0$, and for any $0 < r < 1/t^*$,

$$\|\mathcal{L}(X_t) \; - \; \pi(\cdot)\| \; \leq \; (1-\epsilon)^{[rt]} + e^{-\delta(t-t^*)} A^{[rt]-1} \mathbf{E}(h(X_0, Y_0)) \, .$$

As before, it can be hard to establish (3) directly, and it is thus desirable to relate this bound to a drift condition.

Corollary 4. *Suppose (1) holds, and in addition there is a function $U \in \mathcal{D}$ with $U(x) \geq 1$ for all $x \in \mathcal{X}$, and $\lambda > 0$ and $\Lambda < \infty$, such that*

$$\mathcal{A}U(x)) \leq -\lambda U(x) + \Lambda \mathbf{1}_C(x) \, , \qquad x \in \mathcal{X} \, .$$

Then, setting $B = \inf_{x \notin C} U(x)$, we have that (3) holds with $h = \frac{1}{2}(U(x) + U(y))$ and with $\delta = \lambda - \frac{\Lambda}{B}$. Hence the conclusion of Theorem 3 holds with these values. Furthermore, we have $A \leq \frac{\Lambda}{\delta} + e^{-\delta t^*} \sup_{x \in C} U(x)$.

We now proceed to the proofs of these results. Recall (see e.g. Lindvall, 1992) that T is a *coupling time* for $\{X_t\}$ and $\{Y_t\}$ if the two processes can be defined jointly so that $X_{T+t} = Y_{T+t}$ for $t \geq 0$, in which case the *coupling inequality* states that $\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\| \leq \mathbf{P}(T > t)$.

Similarly, following Aldous and Thorisson (1993), we define T and T' to be *shift-coupling epochs* for $\{X_t\}$ and $\{Y_t\}$ if $X_{T+t} = Y_{T'+t}$ for $t \geq 0$. To make use of shift-coupling, we use the following elementary result. (For a proof, and additional background on shift-coupling, see Thorisson, 1992, equation 10.2; Thorisson, 1993; Thorisson, 1994; Roberts and Rosenthal, 1994.)

Proposition 5. *Let $\{X_t\}$, $\{Y_t\}$ be continuous-time Markov processes, each with transition probabilities $P(x, \cdot)$, and let T and T' be shift-coupling epochs for $\{X_t\}$ and $\{Y_t\}$. Then the total variation distance between ergodic averages of $\mathcal{L}(\{X_t\})$ and $\mathcal{L}(\{Y_t\})$ satisfies*

$$\left\| \frac{1}{t} \int_0^t P(X_s \in \cdot) \, ds \; - \; \frac{1}{t} \int_0^t P(Y_s \in \cdot) \, ds \right\| \leq \; \frac{1}{t} \mathbf{E} \left(\min(t, \max(T, T')) \right)$$

$$\leq \; \frac{1}{t} (\mathbf{E}(T) + \mathbf{E}(T')) \; = \; \frac{1}{t} \left(\int_0^\infty \mathbf{P}(T > s) \, ds \; + \; \int_0^\infty \mathbf{P}(T' > s) \, ds \right) .$$

The key computation for the proofs of Theorems 1 and 3 is contained in the following lemma, which is a straightforward consequence of regeneration theory (cf. Nummelin, 1992, p. 92; see also Athreya and Ney, 1978; Asmussen, 1987; Meyn and Tweedie, 1994; Rosenthal, 1995).

Lemma 6. *With notation and assumptions as in Theorem 1, there is a random stopping time T with $\mathcal{L}(X_T) = Q(\cdot)$, such that for any $0 < r < 1/t^*$ and $s > 0$,*

$$\mathbf{P}(T > s) \;\leq\; (1-\epsilon)^{[rs]} + e^{-\delta(s-t^*)} A^{[rs]-1} \mathbf{E}(V(X_0)) .$$

Proof. We construct $\{X_t\}$ as follows. We begin by choosing $q \sim Q(\cdot)$, and letting $Z_1, Z_2, \ldots$ be a sequence of i.i.d. random variables with $\mathbf{P}(Z_i = 1) = 1 - \mathbf{P}(Z_i = 0) = \epsilon$. Let τ_1 be the first time the process $\{X_t\}$ is in the set C. If $Z_1 = 1$, we set $X_{\tau_1 + t^*} = q$. If $Z_1 = 0$, we choose $X_{\tau_1 + t^*} \sim \frac{1}{1-\epsilon}(P(X_{\tau_1}, \cdot) - \epsilon Q(\cdot))$. In either case, we then fill in X_t for $\tau_1 < t < \tau_1 + t^*$ from the appropriate conditional distributions. Similarly, for $i \geq 2$, we let τ_i be the first time $\geq \tau_{i-1} + t^*$ at which the process $\{X_t\}$ is in the set C, and again if $Z_i = 1$ we set $X_{\tau_i + t^*} = q$, otherwise choose $X_{\tau_i + t^*} \sim \frac{1}{1-\epsilon}(P(X_{\tau_i}, \cdot) - \epsilon Q(\cdot))$. This defines the process $\{X_t\}$ for all times $t \geq 0$, such that $\{X_t\}$ follows the transition probabilities $P^t(x, \cdot)$.

To proceed, we define $N_t = \max\{i; \tau_i \leq t\}$. Now, for any $0 < r < \frac{1}{t^*}$,

$$\mathbf{P}(T > s) \;\leq\; (1-\epsilon)^{[rs]} + \mathbf{P}(N_{s-t^*} < [rs]) .$$

However, setting $D_1 = \tau_1$ and $D_i = \tau_i - \tau_{i-1}$ for $i \geq 2$, for any positive integer j, we have from Markov's inequality that

$$\mathbf{P}(N_s < j) = \mathbf{P}\left(\sum_{i=1}^{j} D_i > s\right) \leq e^{-\delta s} \mathbf{E}\left(\prod_{i=1}^{j} e^{\delta D_i}\right) .$$

Moreover, from (2), as in Rosenthal (1995), we have

$$\mathbf{E}\left(\prod_{i=1}^{j} e^{\delta D_i}\right) \;\leq\; \prod_{i=1}^{j} \mathbf{E}\left(e^{\delta D_i} \,|\, \mathcal{F}_{i-1}\right) \;\leq\; A^{j-1} \mathbf{E}(V(X_0)) ,$$

where $\mathcal{F}_i = \sigma\{X_t; 0 \leq t \leq \tau_i\}$.

Hence,

$$\mathbf{P}(T > s) \ \le \ (1-\epsilon)^{[rs]} + e^{-\delta(s-t^*)} A^{[rs]-1} \mathbf{E}(V(X_0)),$$

as required. ∎

Proof of Theorem 1. We augment the process $\{X_t\}$ with a second process $\{Y_t\}$, also following the transition probabilities $P^t(x,\cdot)$, but with $\mathcal{L}(Y_0) = \pi(\cdot)$, so that $P(Y_s \in \cdot) = \pi(\cdot)$ for all times $s \ge 0$. From the lemma, there are times T and T' with $\mathcal{L}(X_T) = \mathcal{L}(Y_{T'}) = Q(\cdot)$. We define the two processes jointly in the obvious way so that $X_T = Y_{T'}$. We complete the construction by *re-defining* Y_s for $s > T'$, by the formula $Y_{T'+t} = X_{T+t}$ for $t > 0$. It is easily seen that $\{Y_s\}$ still follows the transition probabilities $P^t(x,\cdot)$. The times T and T' are then shift-coupling epochs as defined above. The lemma gives upper bounds on $P(T > s)$ and $P(T' > s)$. Hence, integrating,

$$\int_0^\infty \mathbf{P}(T > s)\,ds \ \le \ \frac{1}{r\epsilon} + A^{-1}(\mathbf{E}(V(X_0)) \frac{e^{\delta/r}}{r(1 - Ae^{\delta t^* - \delta/r})},$$

with a similar formula for $\int P(T' > s)ds$. The result then follows from Proposition 5. ∎

Proof of Corollary 2. The statement about (2) follows directly by a standard martingale argument, as in Rosenthal (1995). Specifically, $e^{\delta(t\wedge\tau_C)}U(X_{t\wedge\tau_C})$ is a non-negative local supermartingale, so that

$$\mathbf{E}U(X_0) \ge \mathbf{E}\left(e^{\delta\tau_C}U(X_{\tau_C})\right) \ge \mathbf{E}\left(e^{\delta\tau_C}\right).$$

For the statement about A, let $E(t,x) = \mathbf{E}_x(U(X_t))$. Then

$$\begin{aligned} E(t,x) &\le \mathbf{E}_x\left(\int_0^t \mathcal{A}U(X_s)\,ds\right) + U(x) \\ &\le -\delta \int_0^t E(s,x)\,ds + \Lambda t + U(x). \end{aligned}$$

Therefore, $E(t,x) \le W(t,x)$ for all t and x, where

$$W(t,x) = -\delta \int_0^t W(s,x)\,ds + \Lambda t + U(x),$$

with $W(0,x) = E(0,x)$ for all $x \in \mathcal{X}$. Hence, solving, we have

$$E(t,x) \le W(t,x) = \frac{\Lambda}{\delta} + U(x)e^{-\delta t}\,.$$

The result follows by taking $t = t^*$ and taking supremum over $x \in \mathcal{X}$. ■

Proof of Theorem 3. We construct the processes $\{X_t\}$ and $\{Y_t\}$ jointly as follows. We begin by choosing $q \sim Q(\cdot)$. We further set $t_0 = \inf\{t \ge 0\,;\,(X_t, Y_t) \in C \times C\}$, and $t_n = \inf\{t \ge t_{n-1} + t^*\,;\,(X_t, Y_t) \in C \times C\}$ for $n \ge 1$. Then, for each time t_i, if we have not yet coupled, then we proceed as follows. With probability ϵ we set $X_{t_i+t^*} = Y_{t_i+t^*} = q$, set $T = t_i + t^*$, and declare the processes to have coupled. Otherwise (with probability $1-\epsilon$, we choose $\{X_{t+t^*}\} \sim \frac{1}{1-\epsilon}\left(P^{t^*}(X_t,\cdot) - \epsilon\, Q(\cdot)\right)$ and $\{Y_{t+t^*}\} \sim \frac{1}{1-\epsilon}\left(P^{t^*}(Y_t,\cdot) - \epsilon\, Q(\cdot)\right)$, conditionally independently. In either case, we then fill in the values X_s and Y_s for $t_i < s < t_i + t^*$ conditionally independently, using the correct conditional distributions given $X_{t_i}, Y_{t_i}, X_{t_i+t^*}, Y_{t_i+t^*}$.

Finally, if $\{X_t\}$ and $\{Y_t\}$ have already coupled, then we let them proceed conditionally independently.

It is easily seen that $\{X_t\}$ and $\{Y_t\}$ each marginally follow the transition probabilities $P^t(\cdot,\cdot)$. Furthermore, T is a coupling time. We bound $P(T > s)$ as in Lemma 6 above. The result then follows from the coupling inequality. ■

Proof of Corollary 4. Setting $h(x,y) = \frac{1}{2}(U(x) + U(y))$, we compute that

$$\begin{aligned}
\mathcal{A}(h(x,y)) &\le -\frac{1}{2}\lambda(U(x) + U(y)) + \frac{\Lambda}{2}(\mathbf{1}_C(x) + \mathbf{1}_C(y)) \\
&\le -\lambda h(x,y) + \frac{\Lambda}{2} + \frac{\Lambda}{2}\mathbf{1}_{C\times C}(x,y) \\
&\le -(\lambda - \frac{\Lambda}{B})h(x,y) + \Lambda \mathbf{1}_{C\times C}(x,y) \\
&= -\delta\, h(x,y) + \Lambda \mathbf{1}_{C\times C}(x,y)\,.
\end{aligned}$$

Statement (3), and also the statement about A, then follow just as in Corollary 2. ■

3. Computing minorizations for one-dimensional diffusions.

To use Theorem 1 or Theorem 3, it is necessary to establish drift and minorization conditions. Consideration of the action of the generator $\mathcal{A}$ on test functions U provides a method of establishing drift conditions. However, finding a minorization condition appears to be more difficult. In this section, we present a method for doing so.

We shall have considerable occasion to estimate first hitting time distributions of one-dimensional diffusions. A key quantity is provided by the *Bachelier-Lévy formula* (Lévy, 1965; Lerche, 1986): Suppose $\{Z_t\}$ is a Brownian motion with drift μ and diffusion coefficient σ, defined by $dZ_t = \mu dt + \sigma dB_t$, started at $Z_0 = 0$. Then if T_x is the first time that $\{Z_t\}$ hits a point $x > 0$, then

$$\mathbf{P}(T_x < t) = \Phi\left(\frac{-x + t\mu}{\sqrt{\sigma^2 t}}\right) + e^{2x\mu/\sigma^2}\,\Phi\left(\frac{-x - t\mu}{\sqrt{\sigma^2 t}}\right),$$

where $\Phi(s) = \int\limits_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$.

Now consider a one-dimensional diffusion process $\{X_t\}$ defined by

$$dX_t = \mu(X_t)dt + dB_t\,.$$

Assume that $\mu(\cdot)$ is C^1, and that there exist positive constants a, b, N such that $\operatorname{sgn}(x)\,\mu(x) \leq a|x| + b$ whenever $|x| \geq N$. This implies (see for example Roberts and Tweedie, 1995, Theorem 2.1) that the diffusion is non-explosive. Furthermore, it is straightforward to check that this diffusion has a unique strong solution.

For $c, d \in \mathbf{R}$, call a set $S \subseteq \mathcal{X}$ a "$[c,d]$-medium set" if $c \leq \mu(x) \leq d$ for all $x \in S$. The set S is medium if it is $[c,d]$-medium for some $c < d$. Note that if $\mu(\cdot)$ is a continuous function, then all compact sets are medium. On medium sets we have some control over the drift of the process; this will allow us to get bounds to help establish minorization conditions.

Theorem 7. *Let $\{X_t\}$ be a one-dimensional diffusion process defined by $dX_t = \mu(X_t)dt + dB_t$. Suppose $C = [\alpha, \beta]$ is a finite interval. Suppose further that $S = [a, b]$ is an interval containing C, which is $[c,d]$-medium. Then for any $t > 0$, there exists an $\epsilon > 0$ such that*

C is (t,ϵ)-small. Moreover, given any $t_0>0$, for all $t\geq t_0$, we have that C is (t,ϵ)-small where

$$\begin{aligned}\epsilon = \Phi\left(\frac{-(\beta-\alpha)-t_0(d-c)}{\sqrt{4t_0}}\right) + e^{-(\beta-\alpha)(d-c)/2}\Phi\left(\frac{t_0(d-c)-(\beta-\alpha)}{\sqrt{4t_0}}\right)\\ -\Phi\left(\frac{-(\alpha-a)-t_0c}{\sqrt{t_0}}\right) - e^{-2(\alpha-a)c}\Phi\left(\frac{t_0c-(\alpha-a)}{\sqrt{t_0}}\right)\\ -\Phi\left(\frac{-(b-\beta)+t_0d}{\sqrt{t_0}}\right) - e^{2(b-\beta)d}\Phi\left(\frac{-t_0d-(b-\beta)}{\sqrt{t_0}}\right).\end{aligned}$$

Proof. Given a standard Brownian motion $\{B_t\}$, we simultaneously construct, for each $x\in C$, a version $\{X_t^x\}$ of X_t started at x, and defined by

$$dX_t^x = \mu(X_t^x)dt + (1-2\mathbf{1}_{\tau_\beta^x\geq t})dB_t\,,$$

where $\tau_\beta^x = \inf\{t\geq 0\,;\,X_t^x = X_t^\beta\}$ is the first time the process $\{X_t^x\}$ hits the process $\{X_t^\beta\}$ (in particular, $\tau_\beta^\beta = 0$). Thus, for $t\geq\tau_\beta^x$, we have $X_t^x = X_t^\beta$. Note also that, if $x\leq y$, then $X_t^x\leq X_t^y$ for all $t\geq 0$. Hence, for $t\geq\tau_\beta^\alpha$, the collection of processes $\{\{X_t^x\}\,;\,\alpha\leq x\leq\beta\}$ are all coincident.

Set $r=\mathbf{P}(\tau_\beta^\alpha\leq t_0)$, and for $t\geq t_0$ let

$$Q_t(\cdot) = \mathbf{P}(X_t^\alpha\in\cdot\,|\,\tau_\beta^\alpha\leq t)\,.$$

Then it follows by construction that

$$P^t(x,\cdot)\ \geq\ r\,Q_t(\cdot)\,,\qquad x\in C\,,$$

as desired.

It remains to compute a lower bound for r. We let $Z_t = X_t^\beta - X_t^\alpha$, and let U_t satisfy the S.D.E.

$$dU_t = (d-c)dt - 2\,dB_t\,,$$

for the same Brownian motion $\{B_t\}$ as before, with $U_0=\beta-\alpha$. We let T_0 be the first hitting time of $\{U_t\}$ to 0. We have that $Z_{t_0}\leq U_{t_0}$ on the event $E_1^C\cap E_2^C$, where

$$E_1 = \{\exists t\leq t_0; X_t^\alpha\leq a\}$$

and

$$E_2 = \{\exists t \le t_0; X_t^\beta \ge b\}\,.$$

Furthermore, defining processes $\{V_t^1\}$ and $\{V_t^2\}$ by

$$dV_t^1 = c\,dt + (1 - 2\mathbf{1}_{\tau_\beta^\alpha \le t})dB_t\,;$$

$$dV_t^2 = d\,dt - dB_t\,;$$

with $V_0^1 = \alpha$ and $V_0^2 = \beta$, it is clear that $E_1 \cap E_2 \subseteq \tilde{E}_1 \cap \tilde{E}_2$ where

$$\tilde{E}_1 = \{\exists t \le t_0; V_t^1 \le a\}$$

and

$$\tilde{E}_2 = \{\exists t \le t_0; V_t^2 \ge b\}\,.$$

Hence,

$$\begin{aligned} r = \mathbf{P}(\tau_\beta^\alpha \le t_0) &\ge \mathbf{P}(T_0 \le t_0) - \mathbf{P}(E_1 \cup E_2) \\ &\ge \mathbf{P}(T_0 \le t_0) - \mathbf{P}(\tilde{E}_1) - \mathbf{P}(\tilde{E}_2)\,. \end{aligned}$$

The result now follows from applying the Bachelier-Lévy formula to each of these three probabilities. ■

4. Multi-dimensional diffusions and pseudo-small sets.

In this section we consider k-dimensional diffusions $\{\mathbf{X}_t\}$ defined by $d\mathbf{X}_t = \mu(\mathbf{X}_t)dt + d\mathbf{B}_t$. Here $\mathbf{B}$ is standard k-dimensional Brownian motion, and $\mathbf{X}_t, \mu(\cdot) \in \mathbf{R}^k$. As before, we assume that $\mu(\cdot)$ is C^1, and that there exist positive constants a, b, N such that $\mathbf{n_x} \cdot \mu(x) \le a\|x\|_2 + b$ whenever $\|x\|_2 \ge N$. This again implies (see for example Roberts and Tweedie, 1995, Theorem 2.1) that the diffusion is non-explosive; and again the diffusion has a unique strong solution.

It may appear that the method of the previous section is quite specific to one-dimension. There, diffusions were jointly constructed so that they had probability $\epsilon > 0$ of coupling by time t^*, and this produced the required minorization condition. However, in multi-dimensions, it may be possible for the two diffusions to "miss" each other, so that bounding the probability of their coupling would not be easy.

It is possible to get around this difficulty by considering what might be called "rotating axes". Specifically, we shall break up the two processes into components parallel and perpendicular to the direction of their difference. We shall define them so that they proceed identically in the perpendicular direction, but are perfectly negatively correlated in the parallel direction. Thus, they will have a useful positive probability of coupling in finite time t^*. (For other approaches to coupling multi-dimensional processes in various contexts, see Lindvall and Rogers, 1986; Davies, 1986; Chen and Li, 1989; Lindvall, 1992, Chapter VI.)

However, it is no longer possible to use such a construction to simultaneously couple *all* of the processes started from different points of a proposed small set. This is because the parallel and perpendicular axes are different for different pairs of processes. Instead, we shall couple the processes pairwise only. This does not establish the existence of a small set, however it does establish the existence of a pseudo-small set as we define now.

Definition. A subset $S \subseteq \mathcal{X}$ is (t, ϵ)-*pseudo-small* for a Markov process $P^t(\cdot, \cdot)$ on a state space $\mathcal{X}$ if, for all $x, y \in S$, there exists a probability measure $Q_{xy}(\cdot)$ on $\mathcal{X}$, such that

$$P^t(x, \cdot) \geq \epsilon\, Q_{xy}(\cdot) \quad \text{and} \quad P^t(y, \cdot) \geq \epsilon\, Q_{xy}(\cdot)\,.$$

That pseudo-small sets are useful for our purposes is given by the following theorem, whose statement and proof are very similar to Theorem 3, but which substitutes pseudo-smallness for smallness.

Theorem 8. *Given a Markov process with transition probabilities $P^t(x, \cdot)$ and stationary distribution $\pi(\cdot)$, suppose $C \subseteq \mathcal{X}$ is (t^*, ϵ)-pseudo-small, for some positive time t^*, and $\epsilon > 0$. Suppose further that there is $\delta > 0$ and a non-negative function $h : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ with $h(x, y) \geq 1$ for all $x \in \mathcal{X}$, such that*

$$\mathbf{E}_{x,y}\left(e^{\delta \tau_{C \times C}}\right) \;\leq\; h(x, y)\,, \qquad (x, y) \notin C \times C$$

where $\tau_{C\times C}$ is the first hitting time of $C \times C$. Set $A = \sup_{(x,y)\in C\times C} \mathbf{E}_{x,y}(h(X_{t^*}, Y_{t^*}))$, where $\{X_t\}$ and $\{Y_t\}$ are defined jointly as described in the proof, and assume that $A < \infty$. Then for $s > 0$, and for any $0 < r < 1/t^*$,

$$\|\mathcal{L}(X_s) \ - \ \pi(\cdot)\| \ \leq \ (1-\epsilon)^{[rs]} + e^{-\delta(s-t^*)} A^{[rs]-1} \mathbf{E}(h(X_0, Y_0)) .$$

Proof. We construct the processes $\{X_t\}$ and $\{Y_t\}$ jointly as follows. We first set $T = \infty$. Then, each time $(X_t, Y_t) \in C \times C$ and $T = \infty$, with probability ϵ we choose $X_{t+t^*} = Y_{t+t^*} \sim Q_{X_t Y_t}(\cdot)$ and set $T = t + t^*$, while with probability $1-\epsilon$ we choose $X_{t+t^*} \sim \frac{1}{1-\epsilon}\left(P^{t^*}(X_t, \cdot) - \epsilon Q_{X_t Y_t}(\cdot)\right)$ and $Y_{t+t^*} \sim \frac{1}{1-\epsilon}\left(P^{t^*}(Y_t, \cdot) - \epsilon Q_{X_t Y_t}(\cdot)\right)$, conditionally independently. If $(X_t, Y_t) \notin C \times C$ or $T < \infty$, we simply choose $X_{t+t^*} \sim P^{t^*}(X_t, \cdot)$ and $Y_{t+t^*} \sim P^{t^*}(Y_t, \cdot)$, conditionally independently.

Having chosen X_{t+t^*} and Y_{t+t^*}, we then fill in the values of X_s and Y_s for $t < s < t+t^*$ conditionally independently, with the correct conditional distributions given X_t, Y_t, X_{t+t^*}, and Y_{t+t^*}.

It is easily seen that $\{X_t\}$ and $\{Y_t\}$ each marginally follow the transition probabilities $P^t(\cdot, \cdot)$. Furthermore, T is a coupling time. We bound $P(T > s)$ as in Lemma 6 above. The result then follows from the coupling inequality. ∎

We now turn our attention to the multi-dimensional diffusions. As in the one-dimensional case, it is necessary to restrict attention to events where the processes remain in some medium set on which their drifts are bounded. Here, for $\mathbf{c}, \mathbf{d} \in \mathbf{R}^k$, we say a set $S \subseteq \mathbf{R}^k$ a "$[\mathbf{c}, \mathbf{d}]$-medium set" if $c_i \leq \mu_i(x) \leq d_i$ for all $x \in S$, for $1 \leq i \leq k$.

Theorem 9. *Let $\{\mathbf{X}_s\}$ be a multi-dimensional diffusion process defined by $d\mathbf{X}_s = \mu(\mathbf{X}_s)ds + d\mathbf{B}_s$. Suppose C is contained in $\prod_{i=1}^{k} [\alpha_i, \beta_i]$, and let $D = \sup_{x,y\in C} \|x-y\|_2$ be the L^2 diameter of C. Let $S = \prod_{i=1}^{k} [a_i, b_i]$, where $a_i < \alpha_i < \beta_i < b_i$ for each i, and suppose S is a $[\mathbf{c}, \mathbf{d}]$-medium set. Set $L = \|\mathbf{d} - \mathbf{c}\|_2 \equiv \left(\sum_{i=1}^{k} (d_i - c_i)^2\right)^{1/2}$. Then for any $t > 0$, there exists an $\epsilon > 0$ such that C is (t, ϵ)-pseudo-small. Moreover, given any $t_0 > 0$, for all*

$t \geq t_0$, *we have that* C *is* (t, ϵ)*-pseudo-small where*

$$\begin{aligned}
\epsilon = {} & \Phi\left(\frac{-D - t_0 L}{\sqrt{4t_0}}\right) + e^{-DL/2}\Phi\left(\frac{t_0 L - D}{\sqrt{4t_0}}\right) \\
& - 2\sum_{i=1}^{k}\Phi\left(\frac{-(\alpha_i - a_i) - t_0 c_i}{\sqrt{t_0}}\right) - 2\sum_{i=1}^{k} e^{-2(\alpha_i - a_i)c_i}\Phi\left(\frac{t_0 c_i - (\alpha_i - a_i)}{\sqrt{t_0}}\right) \\
& - 2\sum_{i=1}^{k}\Phi\left(\frac{-(b_i - \beta_i) + t_0 d_i}{\sqrt{t_0}}\right) - 2\sum_{i=1}^{k} e^{2(b_i - \beta_i)d_i}\Phi\left(\frac{-t_0 d_i - (b_i - \beta_i)}{\sqrt{t_0}}\right) .
\end{aligned}$$

The above estimates are clearly based again upon the Bachelier-Lévy formula; in fact many aspects of this proof are similar to those of Theorem 7. In some special cases the estimates can be improved upon considerably; the method of proof will indicate where improvements might be possible.

Proof. Given a standard k-dimensional Brownian motion $\{\mathbf{B}_s\}$, and a pair of points $\mathbf{x}, \mathbf{y} \in C$, we define diffusions $\mathbf{X}^{\mathbf{x}}$ and $\mathbf{X}^{\mathbf{y}}$ simultaneously by $\mathbf{X}^{\mathbf{x}}_0 = \mathbf{x}$, $\mathbf{X}^{\mathbf{y}}_0 = \mathbf{y}$,

$$d\mathbf{X}^{\mathbf{y}}_s = \mu(\mathbf{X}^{\mathbf{y}}_s)ds + d\mathbf{B}_s ,$$

and

$$d\mathbf{X}^{\mathbf{x}}_s \;=\; \mu(\mathbf{X}^{\mathbf{x}}_s)ds \;+\; d\mathbf{B}_s - 2(\mathbf{n}_s \cdot d\mathbf{B}_s)\mathbf{1}_{\tau \geq s}\mathbf{n}_s ,$$

where τ denotes the first time that $\mathbf{X}^{\mathbf{x}}$ and $\mathbf{X}^{\mathbf{y}}$ coincide, and $\mathbf{n}_s$ denotes the unit vector in the direction from $\mathbf{X}^{\mathbf{x}}_s$ to $\mathbf{X}^{\mathbf{y}}_s$.

This has the effect that $\mathbf{X}^{\mathbf{x}}$ and $\mathbf{X}^{\mathbf{y}}$ proceed identically tangentially to the axis joining them, but are perfectly negatively correlated in the direction parallel to this axis. The processes $\mathbf{X}^{\mathbf{x}}$ and $\mathbf{X}^{\mathbf{y}}$ coalesce on coincidence. (Note that, unlike in the one-dimensional case, this intricate construction is necessary to ensure that coupling is achieved with positive probability.) Hence, if $r = \mathbf{P}(\tau \leq t_0)$, then for $t \geq t_0$ we have $P^t(x, \cdot) \geq r\, Q_t(\cdot)$ and $P^t(y, \cdot) \geq r\, Q_t(\cdot)$, exactly as in Theorem 7. We now proceed to lower-bound r.

If we set $\mathbf{Z}_s = \mathbf{X}^{\mathbf{y}}_s - \mathbf{X}^{\mathbf{x}}_s$, then by a simple application of Itô's formula, and at least up until the first time one of the $\mathbf{X}^{\mathbf{x}}$ processes exits S, we have that

$$\|\mathbf{Z}_s\|_2 \leq |U_s|$$

where U_s is the one-dimensional diffusion defined by $U_0 = D$ and

$$dU_s = 2\,d\tilde{B}_s + \mathrm{sgn}(U_s)\,L\,ds\,,$$

where $\tilde{B}_s = \mathbf{B}_s \cdot \mathbf{n}_s$ is a standard one-dimensional Brownian motion.

Now note that, if $A = \{\mathbf{X}_s^x \in S \text{ and } \mathbf{X}_s^y \in S, \text{ for } 0 \le s \le \tau\}$, then

$$\begin{aligned} r &\ge \mathbf{P}(\{\tau \le t_0\} \cap A) \\ &\ge \mathbf{P}(\{U \text{ hits zero before time } t_0\} \cap A) \\ &\ge \mathbf{P}(U \text{ hits zero before time } t_0) - \mathbf{P}(A^c)\,. \end{aligned}$$

We compute the first of these two probabilities using the Bachelier-Lévy formula, exactly as in Theorem 7. We bound the second by writing $P(A^c) \le \sum_{i=1}^{k} P(A_i^c)$, where $A_i = \{\mathbf{X}_{s,i}^x \in S \text{ and } \mathbf{X}_{s,i}^y \in S, \text{ for } 0 \le s \le \tau\}$ (here $\mathbf{X}_{s,i}^x$ is the i^{th} component of $\mathbf{X}_s^x$), and computing each $P(A_i)$ similarly to Theorem 7 (after first redefining $\mu(\mathbf{x}) = \mathbf{c}$ for $\mathbf{x} \notin S$, so that we can assume $c_i \le \mu_i(x) \le d_i$ for all $x \in \mathcal{X}$). (The extra factor of 2 is required because either of $\mathbf{X}^{\mathbf{x}}$ and $\mathbf{X}^{\mathbf{y}}$ can now escape from either side; orderings like $\mathbf{X}_{s,i}^{\mathbf{x}} \le \mathbf{X}_{s,i}^{\mathbf{y}}$ need not be preserved.) The result follows. ∎

5. Examples: Langevin diffusions.

In this section, we consider applying our results to some simple examples of Langevin diffusions. A Langevin diffusion requires a probability distribution $\pi(\cdot)$ on $\mathbf{R}^k$, having C^1 density $f(\cdot)$ with respect to Lebesgue measure. It is defined by the S.D.E.

$$d\mathbf{X}_t \;=\; d\mathbf{B}_t \;+\; \mu(\mathbf{X}_t)\,dt \;\equiv\; d\mathbf{B}_t \;+\; \frac{1}{2}\nabla \log f(\mathbf{X}_t)\,dt\,.$$

Such a diffusion is reversible with respect to $\pi(\cdot)$, which is thus a stationary distribution (intuitively, the diffusion is more likely to proceed in the direction where f is increasing). This fact has been exploited to use Langevin diffusions for Monte Carlo simulation (see Grenander and Miller, 1994; Philips and Smith, 1994; Roberts and Tweedie, 1995; Roberts and Rosenthal, 1995).

5.1. Standard normal distribution; the Ornstein-Uhlenbeck process.

Here $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$; hence, the Langevin diffusion has drift $\mu(x) = \frac{1}{2} f'(x)/f(x) = -x/2$ (and is thus an Ornstein-Uhlenbeck process). We recall that this process has generator

$$\mathcal{A} = \frac{1}{2}\frac{d^2}{dx^2} - \frac{x}{2}\frac{d}{dx}.$$

We choose $U(x) = 1 + x^2$, and compute that $\mathcal{A}U(x) = 1 - x^2$. Setting

$$M(t) \;=\; U(X_t) - \int_0^t \mathcal{A}U(X_s)\,ds\,,$$

Itô's lemma (see, e.g., Bhattacharya and Waymire, Chapter VII) implies that $M(t)$ is a local martingale, satisfying the S.D.E.

$$dM(t) \;=\; X_t\,dB_t\,.$$

This is L^2-bounded on compact time intervals, so by the dominated convergence theorem, $M(t)$ is in fact a martingale. Thus $U \in \mathcal{D}$.

We choose $C = [-\beta, \beta]$, and choose $S = [-b, b]$ (where $1 < \beta < b$ may be chosen as desired to optimize the results). To make $\mathcal{A}U(x) \le -\lambda U(x) + \Lambda \mathbf{1}_C(x)$, we choose $\lambda = \frac{\beta^2-1}{\beta^2+1}$ and $\Lambda = \lambda + 1$.

We now apply Theorem 7 and Corollary 4. Choosing $\beta = 1.6$, $b = 5$, and $t_0 = t^* = 0.32$, we compute that $\epsilon = 0.00003176$ and $\delta = 0.03421$, to obtain that for any $0 < r < 1/t^*$,

$$\|\mathcal{L}(X_t) - \pi(\cdot)\| \;\le\; (0.9999683)^{[rt]} + (0.96637 \cdot 45.6361^r)^t(\frac{3}{2} + E)\,,$$

where $E = \mathbf{E}_{\mu_0}(X_0)^2$ is found from the initial distribution of our chain. Choosing $r > 0$ sufficiently small, we can ensure that $0.96637 \cdot 45.6361^r < 1$, thus giving an exponentially-decreasing quantitative upper bound as a function of t.

The main point is that Theorem 7 and Corollary 4 provide us with concrete, quantitative bounds on the distance of $\mathcal{L}(X_t)$ to the stationary distribution. ■

Remark. *Multivariate normal distributions.* In $\mathbf{R}^k$, if $f(\cdot)$ is the density for a multivariate normal distribution, then we can get similar bounds without any additional work. Indeed, an appropriate linear transformation reduces this case to that of a standard multivariate normal. From there, we inductively use the inequality $\|\mu_1 \times \mu_2 - \nu_1 \times \nu_2\| \leq \|\mu_1 - \nu_1\| + \|\mu_2 - \nu_2\|$, to reduce this to the one-dimensional case. Thus, we obtain bounds identical to the above, except multiplied by a global factor of k.

5.2. A two-dimensional diffusion.

Following Roberts and Tweedie (1995), we consider the density on $\mathbf{R}^2$ given by

$$f(x, y) \;=\; C\, e^{-x^2 - y^2 - x^2 y^2}\,,$$

where $C > 0$ is the appropriate normalizing constant. The associated Langevin diffusion has drift

$$\mu(x, y) \;=\; -\left(x(1+y^2), y(1+x^2)\right).$$

This diffusion has generator

$$\mathcal{A} = \frac{1}{2}\left(\frac{\partial^2}{(\partial x)^2} + \frac{\partial^2}{(\partial y)^2}\right) - x(1+y^2)\frac{\partial}{\partial x} - y(1+x^2)\frac{\partial}{\partial y}\,.$$

We choose the function $U(x, y) = 1 + x^2 + y^2 + x^2 y^2$, and compute that $\mathcal{A}U(x, y) = 2 + x^2 + y^2 - 2x^2(1+y^2)^2 - 2y^2(1+x^2)^2$. As in the previous example, we set

$$M(t) \;=\; U(X_t) - \int_0^t \mathcal{A}U(X_s)\, ds\,.$$

Again, by Itô's lemma, $M(t)$ is a local martingale. Furthermore, by comparison to two independent Ornstein-Uhlenbeck processes, we see that $M(t)$ is again L^2-bounded on compact time intervals, so that by dominated convergence $M(t)$ is again a martingale. Hence again $U \in \mathcal{D}$.

We let C be the square with corners $(\pm\beta, \pm\beta)$ and let S be the square with corners $(\pm b, \pm b)$. For fixed $\lambda > 0$ (to be chosen later to optimize the results), we choose $\beta = \sqrt{\frac{2+\lambda}{1-\lambda}}$

and $\Lambda = 2 + \lambda$, so that the drift condition $\mathcal{A}U(x,y) \leq -\lambda U(x,y) + \Lambda \mathbf{1}_C(x,y)$ is satisfied. We note that we have $d_1 = d_2 = -c_1 = -c_2 = b(1+b^2)$.

We apply Theorem 9 and Corollary 4. We choose $\lambda = 0.5$ so that $\beta = 2.236$. We further choose $b = 6.836$ and $t_0 = t^* = 0.3$, to get that $\delta = 0.0833$ and $\epsilon = 4.146 \cdot 10^{-9}$. We thus obtain that for any $0 < r < 1/t^*$,

$$\|\mathcal{L}(X_t) - \pi(\cdot)\| \;\leq\; (1 - 4.146 \cdot 10^{-9})^{[rt]} + \frac{1}{2}(0.9201 \cdot 66.915^r)^t \Big(\mathbf{E}_{\mu_0} U(X_0) + \mathbf{E}_\pi U(Y_0)\Big)\,.$$

As before, choosing $r > 0$ sufficiently small, we can ensure that $0.9201 \cdot 66.915^r < 1$, thus giving an exponentially-decreasing quantitative upper bound as a function of t. ∎

Acknowledgements. We thank Tom Salisbury and Richard Tweedie for helpful comments.

REFERENCES

D.J. Aldous and H. Thorisson (1993), Shift-coupling. Stoch. Proc. Appl. **44**, 1-14.

S. Asmussen (1987), Applied Probability and Queues. John Wiley & Sons, New York.

K.B. Athreya and P. Ney (1978), A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. **245**, 493-501.

P.H. Baxendale (1994), Uniform estimates for geometric ergodicity of recurrent Markov chains. Tech. Rep., Dept. of Mathematics, University of Southern California.

R.N. Bhattacharya and E.C. Waymire (1990), Stochastic processes with applications. Wiley & Sons, New York.

M.F. Chen and S.F. Li (1989), Coupling methods for multidimensional diffusion processes. Ann. Prob. **17**, 151-177.

P.L. Davies (1986), Rates of convergence to the stationary distribution for k-dimensional diffusion processes. J. Appl. Prob. **23**, 370-384.

A.E. Gelfand and A.F.M. Smith (1990), Sampling based approaches to calculating marginal densities. J. Amer. Stat. Assoc. **85**, 398-409.

U. Grenander and M.I. Miller (1994), Representations of knowledge in complex systems (with discussion). J. Roy. Stat. Soc. B **56**, 549-604.

H.R. Lerche (1986), Boundary crossings of Brownian motion. Springer-Verlag, London.

P. Lévy (1965), Processus stochastiques et mouvement Brownien. Gauthier-Villars, Paris.

T. Lindvall (1992), Lectures on the coupling method. Wiley & Sons, New York.

T. Lindvall and L.C.G. Rogers, Coupling of multidimensional diffusions by reflection. Ann. Prob. **14**, 860-872.

S.P. Meyn and R.L. Tweedie (1993a), Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. Adv. Appl. Prob. **25**, 518-548.

S.P. Meyn and R.L. Tweedie (1993b), Markov chains and stochastic stability. Springer-Verlag, London.

S.P. Meyn and R.L. Tweedie (1994), Computable bounds for convergence rates of Markov chains. Ann. Appl. Prob. **4**, 981-1011.

E. Nummelin (1984), General irreducible Markov chains and non-negative operators. Cambridge University Press.

D.B. Phillips and A.F.M. Smith (1994), Bayesian model comparison via jump diffusions. Tech. Rep. **94-20**, Department of Mathematics, Imperial College, London.

G.O. Roberts and J.S. Rosenthal (1994), Shift-coupling and convergence rates of ergodic averages. Preprint.

G.O. Roberts and J.S. Rosenthal (1995), Optimal scaling of discrete approximations to Langevin diffusions. Preprint.

G.O. Roberts and R.L. Tweedie (1995), Exponential Convergence of Langevin Diffusions and their discrete approximations. Preprint.

J.S. Rosenthal (1995), Minorization conditions and convergence rates for Markov chain Monte Carlo. J. Amer. Stat. Assoc. **90**, 558-566.

A.F.M. Smith and G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J. Roy. Stat. Soc. Ser. B **55**, 3-24.

H. Thorisson (1992), Coupling methods in probability theory. Tech. Rep. RH-18-92, Science Institute, University of Iceland.

H. Thorisson (1993), Coupling and shift-coupling random sequences. Contemp. Math., Volume **149**.

H. Thorisson (1994), Shift-coupling in continuous time. Prob. Th. Rel. Fields **99**, 477-483.