The Annals of Applied Probability
1993, Vol. 3, No. 3, 897–913

ACCELERATING GAUSSIAN DIFFUSIONS[1]

BY CHII-RUEY HWANG,[2] SHU-YIN HWANG-MA AND SHUENN-JYI SHEU

Academia Sinica, Soochow University and Academia Sinica

Let $\pi(x)$ be a given probability density proportional to $\exp(-U(x))$ in a high-dimensional Euclidean space $\mathbb{R}^m$. The diffusion $dX(t) = -\nabla U(X(t))\,dt + \sqrt{2}\,dW(t)$ is often used to sample from π. Instead of $-\nabla U(x)$, we consider diffusions with smooth drift $b(x)$ and having equilibrium $\pi(x)$. First we study some general properties and then concentrate on the Gaussian case, namely, $-\nabla U(x) = Dx$ with a strictly negative-definite real matrix D and $b(x) = Bx$ with a stability matrix B; that is, the real parts of the eigenvalues of B are strictly negative. Using the rate of convergence of the covariance of $X(t)$ [or together with $EX(t)$] as the criterion, we prove that, among all such $b(x)$, the drift Dx is the worst choice and that improvement can be made if and only if the eigenvalues of D are not identical. In fact, the convergence rate of the covariance is $\exp(2\lambda_M(B)t)$, where $\lambda_M(B)$ is the maximum of the real parts of the eigenvalues of B and the infimum of $\lambda_M(B)$ over all such B is $1/m \operatorname{tr} D$. If, for example, a "circulant" drift

$$\left(\frac{\partial U}{\partial x_m} - \frac{\partial U}{\partial x_2}, \frac{\partial U}{\partial x_1} - \frac{\partial U}{\partial x_3}, \dots, \frac{\partial U}{\partial x_{m-1}} - \frac{\partial U}{\partial x_1}\right)$$

is added to Dx, then for essentially all D, the diffusion with this modified drift has a better convergence rate.

1. Introduction. Probability distributions in high-dimensional Euclidean spaces appear frequently in many applications, for example, in image analysis and image synthesis or in mathematical physics. Direct sampling from these distributions is not feasible in practice: one has to resort to approximations. Diffusions with the underlying probability as their equilibrium distribution could be used to do the approximation. This leads us to investigate the theoretical issue of how to achieve rapid convergence of these diffusions. We first study their structure and then concentrate on the Gaussian case. In the latter, our results allow us to devise a better drift and, in fact, the commonly used gradient drift is the worst one.

Using diffusions to sample from the prior distribution for image synthesis at the model building stage and to sample from the posterior distribution for image reconstruction can be found in the works of Grenander (1984), Roysam,

Received May 1992; revised January 1993.
[1]Partially supported by NSC Grant 81-0208-M-001-50.
[2]Partially supported by Office of Naval Research Grant N00014-91-J-1021 and NSF Grant DMS-88-13699 while visiting the Division of Applied Mathematics, Brown University.
AMS 1991 *subject classifications*. Primary 60J60, 62E25; secondary 65C05, 82B31, 68U10.
Key words and phrases. Diffusions, convergence rate, stochastic relaxation, Monte Carlo method, stability matrix, Ornstein–Uhlenbeck process, reversible process, image analysis, covariance matrix.

Miller, Bhattacharjya and Turner (1992), Grenander and Miller (1992) and references therein.

Geman and Hwang (1986), Chiang, Hwang and Sheu (1987), Hwang and Sheu (1990) studied diffusions for global optimization, using a version of simulated annealing. This also motivates us to search for drifts other than the gradient one that might have a better convergence rate.

Related works can be found in Barone and Frigessi (1990) for improving stochastic relaxation for Gaussian random fields, in Amit and Grenander (1991) for comparison of sweep strategies for Gaussian distribution, in Amit (1991) for rates of convergence of stochastic relaxation, in Frigessi, Hwang and Younes (1992) for an optimal Monte Carlo method with respect to certain criteria and in Frigessi, Hwang, Sheu and Di Stefano (1993) for comparisons of convergence rates of some updating dynamics. Diaconis and Stroock (1991), Fill (1991) and Chiang and Chow (1992) also contain interesting results. Goodman and Sokal (1989) gave a very extensive treatment of Monte Carlo simulation for Gaussian distributions and we shall come back to this at the end of this section.

Let π be a fixed probability on $\mathbb{R}^m$ with density

$$\frac{1}{Z}\exp - U(x), \tag{1.1}$$

where Z is the norming constant. Usually U is given and has nice properties. The diffusion

$$dX(t) = -\nabla U(X(t))\,dt + \sqrt{2}\,dW(t), \qquad t > 0,\ X(0) = x_0, \tag{1.2}$$

is used to approximate π. Here W is the standard Brownian motion in $\mathbb{R}^m$.

Note that we try to approximate $\pi(x)$ and only have $U(x)$ at our disposal. The diffusion (1.2) seems to be a reasonable choice. But we approach the problem from another angle by first studying diffusions with $\pi(x)$ as their equilibrium, and then construct new drifts from $U(x)$ such that the corresponding diffusions would have better approximations. Hence, we consider a diffusion with smooth drift $b(x)$,

$$dX(t) = b(X(t))\,dt + \sqrt{2}\,dW(t), \qquad t > 0,\ X(0) = x_0, \tag{1.3}$$

such that π is its equilibrium distribution.

If the diffusion (1.3) is reversible, then $b(x)$ has to be $-\nabla U(x)$ (see Section 2). This leads us to consider the nonreversible ones. Write $b(x) = -\nabla U(x) + c(x)$, where $c(x)$ is chosen to ensure that $\pi(x)$ remains the equilibrium of $X(t)$ in (1.3). Later we shall characterize $c(x)$. The "circulant" drift

$$\left(\frac{\partial U}{\partial x_m} - \frac{\partial U}{\partial x_2}, \frac{\partial U}{\partial x_1} - \frac{\partial U}{\partial x_3}, \dots, \frac{\partial U}{\partial x_{m-1}} - \frac{\partial U}{\partial x_1}\right)$$

and the "shift" drift

$$\left(-\frac{\partial U}{\partial x_2}, \frac{\partial U}{\partial x_1} - \frac{\partial U}{\partial x_3}, \dots, \frac{\partial U}{\partial x_{m-2}} - \frac{\partial U}{\partial x_m}, \frac{\partial U}{\partial x_{m-1}}\right)$$

are examples of equilibrium-preserving vectors, $c(x)$. Let L and L_b denote the infinitesimal generators of (1.2) and (1.3), respectively. For nicely chosen $c(x)$, the spectral gaps of L and L_b and, hence, the rates of convergence of the corresponding Markov processes, are compared. These are studied in Section 2.

In Section 3 we consider the Gaussian case, that is, U is a quadratic form and $-\nabla U(x)$ can be expressed as Dx with a strictly negative-definite real matrix D. For the diffusion (1.3), we consider the Ornstein–Uhlenbeck process with linear drift $b(x) = Bx$ with a stability matrix B. Denote by $\lambda_M(B)$ the maximum (but negative) of the real parts of the eigenvalues of B. The rate of convergence of the covariance matrix of $X(t)$, which is $\exp(2\lambda_M(B)t)$ [or together with the convergence of $EX(t)$], is used as the comparison criterion. We prove that, among all drifts with equilibrium π, $-\nabla U(x)$ is the worst case and improvement can be made if and only if the eigenvalues of D are not identical. If the "circulant" drift $c(x)$ (previously defined) is added to $-\nabla U(x)$, then for essentially all the quadratic forms U, one has faster convergence.

Although the convergence rate of the mean is improved along with that of the covariance, the overall convergence of the mean is slower if $x_0 \neq 0$. Of course one can always start at the origin to eliminate the bias.

In Section 4 we prove that the infimum of $\lambda_M(B)$ over all equilibrium-preserving B is $(1/m)\operatorname{tr} D$ by studying the asymptotic behavior of $\lambda_M(D + \alpha C)$, as $\alpha \to \infty$, for a particular C. The construction of an appropriate C depends on a natural orthogonal decomposition of $\mathbb{R}^m$ w.r.t. a weighted inner product. The attainability of the optimum is still not known (added in proof: the optimum is attained). As an example, the three-dimensional case is studied in detail. Two numerical simulations of $\lambda_M(D + \alpha C)$ are included at the end of Section 4.

As an example, let us consider the optimal rate $(1/m)\operatorname{tr} D$ for the two-dimensional discretized Laplacian Δ_L on an $L \times L$ square lattice, that is, $-\Delta_L$ is the inverse covariance matrix of the underlying Gaussian distribution and the dimensionality m equals L^2. The diagonal entries of Δ_L are identically equal to -4. Notice that the optimal rate -4 is independent of the dimensionality m. Hence, theoretically the "critical slowing down" seems to be avoided by using proper diffusion processes.

The consideration along this line is related to the following work. Goodman and Sokal (1989) show for a large class of discrete time algorithms, including multigrid methods with appropriate structures, that the Monte Carlo simulation procedure for the Gaussian distribution with mean $-D^{-1}y$ and inverse covariance matrix $-D$ has the corresponding deterministic procedure for the solution of the linear equation $-Dx = y$. These two procedures have the same convergence rate. Note that for nonzero y, similar to what was mentioned before, the rate is determined by the convergence rate of the mean, which is slower than that of the covariance, and the multigrid methods may avoid the critical slowing down as the dimension increases.

Whether our result corresponds to some accelerated procedure for solving the linear equation needs further investigation. Nevertheless, the following points in this direction are observed.

Let us discretize (1.3) with $b(x) = Bx$ and time step δ:

$$X_{n+1} - X_n = \delta B X_n + \sqrt{2\delta}\,\xi_n,$$

where $X_n = X(n\delta)$ and ξ_n are i.i.d. standard Gaussian random variables in $\mathbb{R}^m$. This is a discretized Langevin equation. Let $M = I + \delta B$. The corresponding scheme for solving the linear equation $-Dx = y$ is

$$\varphi_{n+1} = M\varphi_n + Ny,$$

where $N = \delta(I + S)$. Note that $B = D + SD$ for a skew symmetric matrix S (see Theorem 3.1). For a proper chosen δ, say δ^{-1} greater than or equal to the spectral radius of D (see Theorem 3.3), the real parts of the eigenvalues of M are nonnegative and strictly less than 1. If the imaginary parts can be controlled, this should be a good scheme.

The Gaussian diffusions considered here have drifts Dx plus linear perturbations. Do nonlinear perturbations have faster convergence (compared with linear ones)? It seems quite complicated even for a quadratic perturbation in the two-dimensional case. Additionally, what is the optimal rate for the general case? These questions are still under investigation.

2. Preliminary results for the general case. Let $\langle x, y\rangle$ denote the usual inner product in $\mathbb{R}^m$ and $\mathbb{C}^m$. For a reversible diffusion (1.3) with a smooth drift $b(x)$, we have the following proposition.

PROPOSITION 2.1. *For the diffusion* (1.3) *with a smooth* $b(x)$, *if the diffusion is reversible, then* $b(x) = -\nabla U(x)$.

PROOF. See, for example, Kolmogorov (1937) or Nagasawa (1961). □

Now we will concentrate on the nonreversible case. Let $L_b f = \Delta f + \langle b(x), \nabla f\rangle$ and let $L_b^* f = \Delta f + \operatorname{div}(fb)$ denote the adjoint operator of L_b, where div is the abbreviation for divergence. It is known [Varadhan (1980)] that if $L_b^* \exp(-U(x)) = 0$ (and there is no explosion), then $\pi(x)$ is the equilibrium distribution of (1.3). Conversely, if π is the equilibrium distribution and the coefficients are smooth enough, then $L_b^* \exp(-U(x)) = 0$.

PROPOSITION 2.2. $L_b^* \exp(-U(x)) = 0$ *if and only if the drift* $b(x)$ *can be written as*

$$b(x) = -\nabla U(x) + (\exp U(x))g(x), \tag{2.1}$$

where $g = (g_1, \ldots, g_m)$, $\operatorname{div} g = 0$ *and there exist smooth functions* f_{ij}, $1 \le i < j \le m$, *such that*

$$g_i = (-1)^{i-1}\left(\sum_{i<j} \frac{\partial f_{ij}}{\partial x_j}(-1)^{j-2} + \sum_{j<i} \frac{\partial f_{ji}}{\partial x_j}(-1)^{j-1}\right).$$

PROOF. Direct computation and the Poincaré lemma [Sternberg (1964)] yield the results. □

Again, direct calculation yields the following useful corollary.

COROLLARY 2.1. *Let* $c(x)$ *be chosen such that* $\langle c(x), \nabla U(x)\rangle = 0$ *for* $x \in \mathbb{R}^m$ *and* $\operatorname{div} c = 0$. *Then* $b(x) = -\nabla U(x) + c(x)$ *satisfies* $L_b^* \exp(-U(x)) = 0$.

REMARK. If $c(x) = S(\nabla U(x))$, where S is a skew symmetric matrix, then $c(x)$ satisfies the conditions of the corollary. Hence, with $b(x) = -\nabla U(x) + c(x)$ in (1.3), $X(t)$ has equilibrium π.

EXAMPLES. *Circulant drift*. Define

$$c(x) = \left(\frac{\partial U(x)}{\partial x_m} - \frac{\partial U(x)}{\partial x_2}, \frac{\partial U(x)}{\partial x_1} - \frac{\partial U(x)}{\partial x_3}, \ldots, \frac{\partial U(x)}{\partial x_{m-1}} - \frac{\partial U(x)}{\partial x_1}\right),$$

that is, $c(x) = F(\nabla U(x))$, where F is a circulant matrix

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ a_m & a_1 & \cdots & a_{m-1} \\ \vdots & & & \vdots \\ a_2 & a_3 & \cdots & a_1 \end{pmatrix}$$

with $a_2 = -1$, $a_m = 1$ and $a_k = 0$ for other k's. Clearly F is skew symmetric.

Shift drift. Define

$$c(x) = \left(-\frac{\partial U(x)}{\partial x_2}, \frac{\partial U(x)}{\partial x_1} - \frac{\partial U(x)}{\partial x_3}, \ldots, \frac{\partial U(x)}{\partial x_{m-2}} - \frac{\partial U(x)}{\partial x_m}, \frac{\partial U(x)}{\partial x_{m-1}}\right),$$

that is, $c(x) = (R - L)(\nabla U(x))$, where R is the right shift matrix and L is the left shift matrix. Again, $R - L$ is skew symmetric.

The following simple calculation reveals a preliminary result for the improvement of convergence rate for some general diffusions. In the Gaussian case, the results are more definite and complete (the investigation in the next two sections is devoted to this case):

$$\int (L_b f) f\pi = \int (Lf) f\pi + \int \langle c, \nabla f\rangle f\pi,$$

where $c(x) = (\exp U(x))g(x)$ in (2.1),

$$\begin{aligned}\int\langle c, \nabla f\rangle f\pi &= \frac{1}{2}\int\langle c, \nabla f^2\rangle\pi \\ &= \frac{1}{2Z}\int\langle ce^{-U}, \nabla f^2\rangle \\ &= -\frac{1}{2Z}\int(\operatorname{div} g)f^2 \\ &= 0\end{aligned}$$

if $c(x)$ and $f(x)$ are good enough; for example, $c(x)$ is smooth with compact support and f is smooth with some integrability conditions. So we have $\int(L_b f)f\pi = \int(Lf)f\pi$.

Suppose that U has a nice growth condition. Then L has a discrete spectrum [see, e.g., Reed and Simon (1978)]. We also assume that the perturbation $c(x)$ is of compact support and small. Let $f + ig$ be a normalized (w.r.t. π) eigenfunction of L_b with eigenvalue $\lambda + i\mu$, where λ, μ are real numbers and f, g are real functions. Then

$$L_b f = \lambda f - \mu g,$$
$$L_b g = \lambda g + \mu f,$$

$$\begin{aligned}\int(Lf)f\pi + \int(Lg)g\pi &= \int(L_b f)f\pi + \int(L_b g)g\pi \\ &= \lambda\left(\int f^2\,d\pi + \int g^2\,d\pi\right) \\ &= \lambda.\end{aligned}$$

We may restrict ourselves to the subspace of $L^2(\pi)$ such that $\int h\pi = 0$. Let e_k denote a complete set of orthonormalized eigenfunctions of L with eigenvalues λ_k, ordered by $0 > \lambda_1 \geq \lambda_2 \geq \cdots$:

$$\lambda = \int(Lf)f\pi + \int(Lg)g\pi = \sum\lambda_k\left(\left(\int fe_k\pi\right)^2 + \left(\int ge_k\pi\right)^2\right) \leq \lambda_1.$$

Equality holds if and only if f and g are in the eigenspace corresponding to λ_1. This means that the convergence rate of $\int(e^{tL_b}f(x) - \int f\pi)^2\,d\pi$ is better than or equal to that of $\int(e^{tL}f(x) - \int f\pi)^2\,d\pi$.

3. The Gaussian case. Now we consider $U(x) = \frac{1}{2}\langle -Dx, x\rangle$, where D is a strictly negative-definite real matrix. Consider the Ornstein–Uhlenbeck process with π as its equilibrium,

$$dX(t) = BX(t)\,dt + \sqrt{2}\,dW(t), \qquad t > 0, X(0) = x_0, \tag{3.1}$$

where B is a stability matrix, that is, the real parts of the eigenvalues of B are negative. In other words, Bx plays the role of $b(x)$ in (1.3) and $D(x)$ plays the role $-\nabla U(x)$ in (1.2).

As mentioned before, we only have $U(x)$ available, that is, D is known. The relationship between B and D is stated in the following theorem.

THEOREM 3.1. *The Ornstein–Uhlenbeck process* (3.1) *has a Gaussian equilibrium with density* $(1/Z)\exp\frac{1}{2}\langle Dx, x\rangle$ *if and only if* $B = C + D$, *where the matrix* C *satisfies* $\langle Cx, Dx\rangle = 0$ *for any* x *in* $\mathbb{R}^m$. *In fact*, $C = SD$ *with a skew symmetric real matrix* S.

PROOF. Basically this is Proposition 2.2 for the Gaussian case:

$$\begin{aligned} \operatorname{div}((Cx)\exp - U(x)) &= 0 \quad \text{for all } x \text{ in } \mathbb{R}^m \\ \Rightarrow (\operatorname{tr} C + \langle Cx, Dx\rangle) &= 0 \quad \text{for all } x \text{ in } \mathbb{R}^m. \end{aligned}$$

Because $\operatorname{tr} C$ is a constant and $\langle Cx, Dx\rangle$ is quadratic, $\langle Cx, Dx\rangle = 0$ for all x in $\mathbb{R}^m$.

Conversely, if $\langle Cx, Dx\rangle = 0$ for all x in $\mathbb{R}^m$, then $\operatorname{tr} C = 0$, which is equivalent to $\operatorname{div} Cx = 0$. By Corollary 2.3, we establish the assertion. □

REMARK 3.1. Stability matrices have been studied extensively in the literature; see, for example, Bellman (1970). Our study here may be classified as a dual problem [Taussky (1967)], because D is fixed and we are interested in different B's and how to obtain B from D. However, our concern here is different from that in the study of stability matrices.

The solution $X(t)$ of (3.1) can be written as

$$\begin{aligned} X(t) &= e^{Bt}\left(x_0 + \int_0^t e^{-Bs}\sqrt{2}\, dW(s)\right), \\ EX(t) &= e^{Bt}x_0, \\ \operatorname{Cov}(X(t)) &= 2\int_0^t e^{Bs}e^{B's}\, ds \end{aligned}$$

and

$$\int_0^t e^{Bs}e^{B's}\, ds \to \int_0^\infty e^{Bs}e^{B's}\, ds = \frac{-D^{-1}}{2}.$$

Note that $\int_t^\infty e^{Bs}e^{B's}\, ds$ is strictly positive-definite and decreasing in the sense of positive-definiteness.

For a matrix A, let $\lambda_M(A)$ and $\lambda(A)$ denote the maximum and minimum of the real parts of the eigenvalues of A, respectively.

THEOREM 3.2.

$$\frac{1}{t}\ln\left\|\int_t^\infty e^{Bs}e^{B's}\, ds\right\| \to 2\lambda_M(B) < 0. \tag{3.2}$$

PROOF. Because all the norms are equivalent and for a positive-definite matrix A, $(1/m)\operatorname{tr} A \le \lambda_M(A) \le \operatorname{tr} A$, we may replace the norm here by tr. To establish (3.2), it suffices to show that for any $\varepsilon > 0$,

$$\lim_{t\to\infty} \frac{\operatorname{tr}\int_t^\infty e^{Bs}e^{B's}\,ds}{\exp(2t(\lambda_M(B)+\varepsilon))} = 0, \tag{3.3}$$

$$\lim_{t\to\infty} \frac{\operatorname{tr}\int_t^\infty e^{Bs}e^{B's}\,ds}{\exp(2t(\lambda_M(B)-\varepsilon))} = \infty. \tag{3.4}$$

Because tr is a linear function, by applying l'Hôpital's rule, (3.3) equals

$$\begin{aligned}
&\lim_{t\to\infty} \frac{-\operatorname{tr} e^{tB}e^{tB'}}{2(\lambda_M(B)+\varepsilon)\exp(2t(\lambda_M(B)+\varepsilon))}\\
&\quad = \lim_{t\to\infty} \frac{-1}{2(\lambda_M(B)+\varepsilon)}\left[\frac{(\operatorname{tr} e^{tB}e^{tB'})^{1/(2t)}}{e^{\lambda_M(B)}e^{\varepsilon}}\right]^{2t}\\
&\quad = 0.
\end{aligned}$$

In fact, by using the Euclidean norm for e^{tB}, one rewrites $(\operatorname{tr} e^{tB}e^{tB'})^{1/(2t)} = (\|e^{tB}\|^2)^{1/(2t)}$ and $\|e^{tB}\|^{1/t}$ converges to the spectral radius of e^B, which is $e^{\lambda_M(B)}$. Note that the latter holds for integers t [Halmos (1958)], but the same result remains true in our case.

(3.4) can be proved similarly. □

The rate of convergence of the covariance is $\exp(2\lambda_M(B)t)$ and that of the mean is $\exp(\lambda_M(B)t)x_0$, so the mean has slower convergence if x_0 is not zero. This coincides with the result in Barone and Frigessi (1990). Of course, we would take the starting point at 0 to speed up the convergence.

Now the rate is determined by $\lambda_M(B)$. We have the following comparison theorem.

THEOREM 3.3. *Let $B = C + D$ with $\langle Cx, Dx\rangle = 0$ for x in $\mathbb{R}^m$. Then*

$$\lambda(D) \le \lambda(B) \le \lambda_M(B) \le \lambda_M(D) \tag{3.5}$$

and $\lambda_M(B) = \lambda_M(D)$ if and only if there exist nonzero u, v in $\mathbb{R}^m$ and ρ in $\mathbb{R}$ such that $u + iv$ is an eigenvector of B with eigenvalue $\lambda_M(D) + i\rho$ and u, v are eigenvectors of D with eigenvalue $\lambda_M(D)$. Moreover, $u + iv$ is an eigenvector of C with eigenvalue $i\rho$ and $Cu = -\rho v$, $Cv = \rho u$. If ρ is nonzero, then $u \perp v$.

PROOF. First notice that $\langle Cx, Dx\rangle = 0$ for any x in $\mathbb{R}^m$ is equivalent to $\langle C(x+iy), D(x+iy)\rangle$ being purely imaginary for any x, y in $\mathbb{R}^m$:

$$\begin{aligned}\langle D(x+iy), (D+C)(x+iy)\rangle &= \langle Dx, Dx\rangle + \langle Dy, Dy\rangle\\ &\quad + \text{a purely imaginary part.}\end{aligned}$$

If $B(x+iy) = (\lambda + i\eta)(x+iy)$, λ, η in $\mathbb{R}$, then

$$\langle D(x+iy), (D+C)(x+iy)\rangle = \lambda(\langle Dx, x\rangle + \langle Dy, y\rangle) + \text{a purely imaginary part}.$$

Hence, for an eigenvector $x+iy$ of B with x, y in $\mathbb{R}^m$ and the real part of the eigenvalue λ,

$$\lambda(\langle Dx, x\rangle + \langle Dy, y\rangle) = \langle Dx, Dx\rangle + \langle Dy, Dy\rangle. \tag{3.6}$$

Let $\{e_k\}$ be a complete orthonormal set in $\mathbb{R}^m$ with $D(e_k) = \lambda_k e_k$. Then (3.6) can be rewritten as

$$\lambda \sum_k \lambda_k\big(\langle x, e_k\rangle^2 + \langle y, e_k\rangle^2\big) = \sum_k \lambda_k^2\big(\langle x, e_k\rangle^2 + \langle y, e_k\rangle^2\big).$$

Therefore, $\lambda(D) \le \lambda \le \lambda_M(D)$ and $\lambda = \lambda_M(D)$ if and only if $\langle x, e_k\rangle^2 + \langle y, e_k\rangle^2 = 0$ for all $\lambda_k \neq \lambda_M(D)$. The last statement is equivalent to $Dx = \lambda_M(D)x$ and $Dy = \lambda_M(D)y$.

Now take $u + iv$ as in the statement of the theorem,

$$B(u+iv) = D(u+iv) + C(u+iv) = \lambda_M(D)(u+iv) + C(u+iv),$$

but $B(u+iv) = (\lambda_M(D) + i\rho)(u+iv) = \lambda_M(D)(u+iv) + i\rho(u+iv)$. Hence, $C(u+iv) = i\rho(u+iv)$, $Cu = -\rho v$, $Cv = \rho u$. If $\rho \neq 0$, $0 = \langle Du, Cu\rangle = -\rho\lambda_M(D)\langle u, v\rangle$. □

REMARK 3.2. The inequality (3.5) is known [see, e.g., Barnett and Storey (1967)]. Our proof here also gives the condition when equality will hold.

REMARK 3.3. If C has $\langle Cx, Dx\rangle = 0$ for $x \in \mathbb{R}^m$, then so does αC for any α in $\mathbb{R}$. Now consider $\lambda_M(\alpha C + D)$ as a function of α. We have the maximum at $\alpha = 0$. It is interesting to know the behavior of this function. For a given D, what is the optimal value $\inf_{C \perp D} \lambda_M(D + C)$ and is it attainable? This will be studied in Section 4.

The following theorem theoretically shows that by adding a drift orthogonal to $-\nabla U(x)$, we can make an improvement as long as the eigenvalues of D are not identical.

THEOREM 3.4. *There exist a real matrix C such that* $\langle Cx, Dx\rangle = 0$ *for any* x *in* $\mathbb{R}^m$ *and* $\lambda_M(C+D) < \lambda_M(D)$ *if and only if the eigenvalues of D are not identical.*

PROOF. If all the eigenvalues of D are the same,

$$\lambda_M(C+D) \ge \lambda(D) = \frac{1}{m}\operatorname{tr}(D) = \lambda_M(D)$$

by Theorem 3.3.

Now we prove the only if part. Let P be a real unitary transformation such that PDP^{-1} is diagonal and $\langle PCP^{-1}x, PDP^{-1}x\rangle = \langle CP^{-1}x, DP^{-1}x\rangle = 0$ for x in $\mathbb{R}^m$. So it suffices to show the result for the diagonal case:

$$U(x) = \frac{-1}{2}\sum_1^m \lambda_k x_k^2, \qquad (Dx)_k = \lambda_k x_k, \qquad 1 \le k \le m.$$

We use the shift drift mentioned in Section 2 for the added term. The corresponding matrix C can be defined by

$$(3.7)\qquad \begin{aligned}&(Cx)_1 = -\lambda_2 x_2, \qquad (Cx)_m = \lambda_{m-1}x_{m-1},\\ &(Cx)_k = \lambda_{k-1}x_{k-1} - \lambda_{k+1}x_{k+1}, \quad \text{for } 1 < k < m.\end{aligned}$$

By using the result in Theorem 3.3, we will prove the assertion by contradiction. Assume that $\lambda_1 = \lambda_M(D)$, the dimensionality of the eigenspace corresponding to λ_1 is n, which is strictly less than m, and $(C + D)(u + iv) = (\lambda_1 + i\rho)(u + iv)$. By Theorem 3.3, $C(u + iv) = i\rho(u + iv)$, $Cu = -\rho v$, $Cv = \rho u$ and $u_k = 0 = v_k$ for $k > n$.

$$(Cu)_{n+1} = \lambda_1 u_n = -\rho v_{n+1} = 0; \qquad \text{hence } u_n = 0.$$

Similarly, $(Cv)_{n+1} = \lambda_1 v_n = \rho u_{n+1} = 0$; hence $v_n = 0$. So we have $u_k = 0 = v_k$ for $k \ge n - 1$. If we iterate the same procedure, we will arrive at $u = 0 = v$, which is a contradiction. □

Using the previous theorem to construct C for the nondiagonal case, one needs to know the eigenvectors of D. This is not practical. In the following we will study the circulant drift.

THEOREM 3.5. *Let C be the circulant drift corresponding to D. That is, $C = FD$ and F is a circulant matrix such that the entries on its first row are zero except the second term is -1 and the last term is $+1$. Let*

$$e_k = \left(\cos\frac{2\pi k(m-1)}{m}, \cos\frac{2\pi k(m-2)}{m}, \dots, 1\right),$$

$$f_k = \left(\sin\frac{2\pi k(m-1)}{m}, \sin\frac{2\pi k(m-2)}{m}, \dots, 0\right).$$

Then $\lambda_M(C + D) < \lambda_M(D)$ if: (i) *for odd m, D does not have e_k and f_k for some k as its eigenvectors with eigenvalues $\lambda_M(D)$; or* (ii) *for even m, D does not have $ae_k + be_{m/2-k}$ and $af_k + bf_{m/2-k}$ for some $1 \le k < m/2$ or $ae_k + be_{3m/2-k}$ and $af_k + bf_{3m/2-k}$ for some $m/2 \le k \le m$, for some a, b as its eigenvectors with eigenvalue $\lambda_M(D)$.*

PROOF. If D is such that $\lambda_M(D + C) = \lambda_M(C)$, then by Theorem 3.3, there exists $u + iv$ such that

$$F^2 u = (D^{-1}C)^2 u = \frac{-\rho^2}{\lambda_M^2(D)}u \quad \text{and} \quad F^2 v = -\frac{\rho^2}{\lambda_M^2(D)}v.$$

The eigenvectors of the circulant matrix F are $e_k + if_k$ with corresponding eigenvalues $\lambda_k = -i2\sin(2\pi k/m)$.

If m is odd, then all the eigenvalues λ_k are distinct. Hence, as long as D does not have e_k and f_k for some k as its eigenvectors with eigenvalue $\lambda_M(D)$, then $D + C$ has a better $\lambda_M(D + C)$. Note that for $k = m$, there is only one eigenvector $e_m = (1, 1, 1, \dots, 1)$.

For the even case,

$$\lambda_k = \begin{cases} \lambda_{m/2-k}, & \text{if } 1 \le k < m/2, \\ \lambda_{3m/2-k}, & \text{if } m/2 \le k \le m. \end{cases}$$

Similar reasoning as in the odd case is applicable here. □

If we have D, then the circulant drift C is easily constructed and, except for very specific D as mentioned in Theorem 3.5, we will have a better $\lambda_M(C + D)$. Hence, one would just use the drift $Cx + Dx$ to run the diffusion without checking the conditions (i) and (ii).

Concerning the relationship to other types of convergence, results are presented in the following propositions. The proofs are a little bit tedious but straightforward; hence, they are omitted.

PROPOSITION 3.6. *Let f_t denote the density of the solution $X(t)$ of* (3.1) *with $x_0 = 0$. Then $\exists\ a > 0$ such that for large t,*

$$a\int|f_t - \pi| \le \|\mathrm{cov}(X(t)) + D^{-1}\| \le \frac{1}{a}\int|f_t - \pi|.$$

In particular,

$$\frac{1}{t}\ln\int|f_t - \pi| \to 2\lambda_M(B).$$

PROPOSITION 3.7. *There exists a constant $a > 0$ such that for large t,*

$$\int\left(E_x h(X_t) - \int h\pi\right)^2 \pi(x)\,dx \le a\|e^{tB}\|^2\int h^2\pi.$$

4. The optimal rate. $\lambda_M(D + C)$ represents the convergence rate. It is interesting to find out the infimum of $\lambda_M(D + C)$. We will prove that $\inf\lambda_m(D + C)$ is $(1/m)\,\mathrm{tr}\,D$.

We construct a specific $\hat{C}$ that relies on a natural orthogonal decomposition of $\mathbb{R}^m$ with respect to the weighted inner product $\langle x, -Dy\rangle$. Then we establish that $\inf_\alpha \lambda_M(D + \alpha\hat{C})$ is indeed $(1/m)\,\mathrm{tr}\,D$.

Our approach does not answer the attainability of the infimum. For $m = 2$ or 3, the answer is positive. We will discuss the three-dimensional case and show some simulation results at the end of this section.

LEMMA 4.1. *If A is self-adjoint w.r.t. an inner product $[\ ,\ ]$ in $\mathbb{R}^m$, then there exists an orthogonal decomposition w.r.t. $[\ ,\ ]$,*

$$\mathbb{R}^m = H_0 \oplus H_1 \oplus \cdots \oplus H_{[m/2]} \tag{4.1}$$

such that:

(i) *for* $1 \le k \le [m/2]$, H_k *is two-dimensional and* $\frac{1}{2}\,tr\,P_k A = (1/m)\,\mathrm{tr}\,A$, *where* P_k *is the orthogonal projection onto* H_k;

(ii) *for odd* m, $tr(P_0 A) = (1/m)\,\mathrm{tr}\,A$, *where* P_0 *is the projection onto the one-dimensional* H_0.

PROOF. The assertion will be proved by induction.

It is obvious for $m = 2$. For $m = 3$, let e_k, $1 \le k \le 3$, be normalized eigenvectors with corresponding eigenvalues λ_k and let $f_1 = (e_1 + e_2 + e_3)/\sqrt{3}$. Define $H_0 = \mathrm{span}\{f_1\}$ and let H_1 be the orthogonal complement of H_0. Then

$$[P_0 A f_1, f_1] = [A f_1, f_1] = \tfrac{1}{3}[\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3, e_1 + e_2 + e_3] = \tfrac{1}{3}\,\mathrm{tr}\,A.$$

Let f_2, f_3 be an orthonormal base for H_1,

$$\mathrm{tr}\,A = [Af_1, f_1] + [Af_2, f_2] + [Af_3, f_3] = \tfrac{1}{3}\,\mathrm{tr}\,A + \mathrm{tr}(P_1 A).$$

Hence $\frac{1}{3}\,\mathrm{tr}\,A = \frac{1}{2}\,\mathrm{tr}\,P_1 A$.

Suppose that the assertion holds for dimensionality less than m. We are going to establish that it is true for m. Let $e_1, \dots, e_m$ be orthonormalized eigenvectors of A. For $m = 2n$, let $H^1 = \mathrm{span}\{e_1, \dots, e_n\}$ and $H^2 = \mathrm{span}\{e_{n+1}, \dots, e_m\}$. In H^l, we use the same [,] with corresponding $P^l A$, where P^l is the projection onto H^l. By induction, we have the decompositions $H^l = H_0^l \oplus \cdots \oplus H_{[n/2]}^l$, $l = 1, 2$. Now choose orthogonal bases u_k^l, v_k^l for H_k^l, $1 \le k \le [n/2]$, $l = 1, 2$. If n is odd, define $H_0 = \{0\}$ and $H_1 = H_0^1 \oplus H_0^2$, and for $1 \le k \le [n/2]$,

$$\begin{aligned} H_{2k} &= \mathrm{span}\{u_k^1 - u_k^2, v_k^1 - v_k^2\}, \\ H_{2k+1} &= \mathrm{span}\{u_k^1 + u_k^2, v_k^1 + v_k^2\}. \end{aligned} \tag{4.2}$$

If n is even, define $H_0 = \{0\}$ and for $1 \le k \le [n/2]$,

$$\begin{aligned} H_{2k-1} &= \mathrm{span}\{u_k^1 + u_k^2, v_k^1 + v_k^2\}, \\ H_{2k} &= \mathrm{span}\{u_k^1 - u_k^2, v_k^1 - v_k^2\}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} 2\,\mathrm{tr}(P_{2k} A) &= \left[A(u_k^1 - u_k^2), u_k^1 - u_k^2\right] + \left[A(v_k^1 - v_k^2), v_k^1 - v_k^2\right] \\ &= \left[Au_k^1, u_k^1\right] + \left[Av_k^1, v_k^1\right] + \left[Au_k^2, u_k^2\right] + \left[Av_k^2, v_k^2\right] \\ &= \frac{2}{n}\mathrm{tr}(P^1 A) + \frac{2}{n}\mathrm{tr}(P^2 A) \\ &= \frac{2}{n}\mathrm{tr}\,A. \end{aligned}$$

Therefore, $\frac{1}{2}\,\mathrm{tr}\,P_{2k} A = (1/2n)\,\mathrm{tr}\,A = (1/m)\,\mathrm{tr}\,A$.

Similar proofs hold for H_{2k+1} in (4.2), H_{2k-1} in (4.3) and for H_0^1 and H_0^2 when they are one-dimensional.

Now for the odd case, $m = 2n + 1$. Let $f_0 = (e_0 + \cdots + e_m)/\sqrt{m}$, $H_0 = \mathrm{span}\{f_0\}$ and $\tilde{H} = H_0^{\perp}$ with the corresponding projection $\tilde{P}$. Again we use the same inner product in $\tilde{H}$. The transformation $\tilde{P}A = \tilde{P}A\tilde{P}$ from $\tilde{H}$ to $\tilde{H}$ is self-adjoint. Choose an orthonormal base $f_1, f_2, \dots, f_{2n}$ in $\tilde{H}$:

$$\begin{aligned}\operatorname{tr} A &= [Af_0, f_0] + \sum_{k\geq 1} [Af_k, f_k]\\ &= [Af_0, f_0] + \sum_{k\geq 1} \left[\tilde{P}Af_k, f_k\right]\\ &= \frac{1}{m}\operatorname{tr} A + \operatorname{tr}(\tilde{P}A).\end{aligned}$$

Hence

$$\frac{1}{m}\operatorname{tr} A = \frac{1}{2}\operatorname{tr} \tilde{P}A. \tag{4.4}$$

Using induction on $\tilde{H}$ with $\tilde{P}A$ and the fact $\tilde{P}_k\tilde{P} = P_k$, where $\tilde{P}_k, P_k$ are defined in the obvious way, for $k \geq 1$, $\frac{1}{2}\operatorname{tr} P_k A = \frac{1}{2}\operatorname{tr}(\tilde{P}_k\tilde{P}A) = (1/2n)\operatorname{tr}\tilde{P}A = (1/m)\operatorname{tr} A$. For H_0, $[P_0Af_0, f_0] = (1/m)\operatorname{tr} A$.

We have established the assertion for m. □

The preceding lemma is applied in our setup as follows:

$[x, y]$ is defined by $\langle x, -Dy\rangle$ and D is the self-adjoint matrix, $\mathbb{R}^m = H_0 \oplus \cdots \oplus H_{[m/2]}$ such that $\operatorname{tr} P_0 D = (1/m)\operatorname{tr} D = \frac{1}{2}\operatorname{tr} P_k D$ for $1 \leq k \leq [m/2]$.

We are going to define $\hat{C}$ and consider the limiting behavior of $D + \alpha\hat{C}$. Let u_k, v_k be an orthonormal base in H_k and let d_k's be distinct positive numbers, for $k \geq 1$. A linear transform $\hat{C}$ from $\mathbb{R}^m$ to $\mathbb{R}^m$ is determined by $\hat{C}(u) = 0$ if u is in H_0 and $\hat{C}(u_k) = -d_k v_k, \hat{C}(v_k) = d_k u_k$ for $1 \leq k \leq [m/2]$.

THEOREM 4.1. $\inf_{\alpha>0} \lambda_M(D + \alpha\hat{C}) = (1/m)\operatorname{tr} D$.

PROOF. For a normalized eigenvector $u^\alpha + iv^\alpha$ of $D + \alpha\hat{C}$ with eigenvalue $\lambda^\alpha + i\mu^\alpha$,

$$\begin{aligned} Du^\alpha + \alpha\hat{C}u^\alpha &= \lambda^\alpha u^\alpha - \mu^\alpha v^\alpha,\\ Dv^\alpha + \alpha\hat{C}v^\alpha &= \mu^\alpha u^\alpha + \lambda^\alpha v^\alpha.\end{aligned} \tag{4.5}$$

$$\hat{C}(u^\alpha + iv^\alpha) = \left(\frac{\lambda^\alpha}{\alpha} + i\frac{\mu^\alpha}{\alpha}\right)(u^\alpha + iv^\alpha) - \frac{D}{\alpha}(u^\alpha + iv^\alpha). \tag{4.6}$$

Suppose that u, v are limits of u_α, v_α along a subsequence of α, still denoted by $\alpha \to \infty$. Then

$$\hat{C}(u + iv) = \lim_{\alpha\to\infty}\left(\frac{\lambda^\alpha}{\alpha} + i\frac{\mu^\alpha}{\alpha}\right)(u + iv). \tag{4.7}$$

Because the eigenvalue of $\hat{C}$ is purely imaginary, $\lambda^\alpha/\alpha \to 0$ and μ^α/α has a limit γ. We will discuss $\gamma = 0$ and $\gamma \neq 0$ separately. In both cases, we will prove $\lambda^\alpha \to (1/m)\operatorname{tr} D$.

If $\gamma \neq 0$, then it must be d_k or $-d_k$. It suffices to prove the positive case. The corresponding u, v are in H_k. By (4.5),

$$[Du^\alpha, u^\alpha] = \lambda^\alpha[u^\alpha, u^\alpha] - \mu^\alpha[v^\alpha, u^\alpha],$$
$$[Dv^\alpha, v^\alpha] = \lambda^\alpha[v^\alpha, v^\alpha] + \mu^\alpha[u^\alpha, v^\alpha];$$
$$\lambda^\alpha = \frac{[Du^\alpha, u^\alpha] + D[v^\alpha, v^\alpha]}{[u^\alpha, u^\alpha] + [v^\alpha, v^\alpha]} \to \frac{[Du, u] + [Dv, v]}{[u, u] + [v, v]} := \lambda; \tag{4.8}$$
$$[Du^\alpha, v^\alpha] + \alpha[\hat{C}u^\alpha, v^\alpha] = \lambda^\alpha[u^\alpha, v^\alpha] - \mu^\alpha[v^\alpha, v^\alpha],$$
$$[Dv^\alpha, u^\alpha] + \alpha[\hat{C}v^\alpha, u^\alpha] = \lambda^\alpha[v^\alpha, u^\alpha] + \mu^\alpha[u^\alpha, u^\alpha].$$

Note that $[\hat{C}x, y] + [\hat{C}y, x] = 0$, $\mu^\alpha/\alpha([u^\alpha, u^\alpha] - [v^\alpha, v^\alpha]) = 2/\alpha([Du^\alpha, v^\alpha] - \lambda^\alpha[u^\alpha, v^\alpha])$. Let $\alpha \to \infty$, $\gamma([u, u] - [v, v]) = 0$. This implies $[u, u] = [v, v]$ and $u \perp v$ in H_k. Then by (4.8) $\lambda = \frac{1}{2} \operatorname{tr} P_k D = (1/m) \operatorname{tr} D$.

Now consider the case $\gamma = 0$. By (4.7), u and v are in H_0. Because $[u, u]^2 + [v, v]^2 = 1$, this can only happen in the odd dimensional case. Moreover, H_0 is one-dimensional, and in view of (4.8), λ is an eigenvalue of $P_0 D$ restricted to H_0 which is $(1/m) \operatorname{tr} D$. This completes the proof of the theorem.

□

Intuitively one would expect that $\lambda_M(D + \alpha C)$ is decreasing in α as $\alpha \to \infty$. Unfortunately this is not true. The situation is more complicated as we will see in the following study of the three-dimensional case. Let

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & a & b \\ -\dfrac{\lambda_1}{\lambda_2} a & 0 & c \\ -\dfrac{\lambda_1}{\lambda_3} b & -\dfrac{\lambda_2}{\lambda_3} c & 0 \end{bmatrix},$$

$0 > \lambda_1 \geq \lambda_2 \geq \lambda_3$, $\rho_k = \lambda_k - (\lambda_1 + \lambda_2 + \lambda_3)/3$, $1 \leq k \leq 3$. We will show that the optimal rate is attainable by studying the characteristic polynomial of $D + \alpha C$ directly. This corresponds to proving that there are two nonzero solutions, η and $-\eta$, of the following set of equations:

$$\rho_1 \rho_2 \rho_3 + \alpha^2 \left(\frac{\lambda_1}{\lambda_3} b^2 \rho_2 + \frac{\lambda_2}{\lambda_3} c^2 \rho_1 + \frac{\lambda_1}{\lambda_2} a^2 \rho_3 \right) = 0, \tag{4.9}$$

$$\eta^2 = (\rho_1 \rho_2 + \rho_2 \rho_3 + \rho_3 \rho_1) + \alpha^2 \left(\frac{\lambda_1}{\lambda_3} b^2 + \frac{\lambda_2}{\lambda_3} c^2 + \frac{\lambda_1}{\lambda_2} a^2 \right). \tag{4.10}$$

Note that $\rho_3 \leq 0 \leq \rho_1$ and one equality holds iff both equalities hold. Hence, it is clear that (4.9) has solutions α, a, b, c with fixed α, b, but a, c can go to infinity. So if we choose a, c large enough and keep the same α, b, then (4.10) has nonzero solutions η and $-\eta$.

If for some C such that

$$\liminf_{\alpha\to\infty} \lambda_M(D + \alpha C) = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3},$$

then from (4.9),

$$\frac{\lambda_1}{\lambda_3} b^2 \rho_2 + \frac{\lambda_2}{\lambda_3} c^2 \rho_1 + \frac{\lambda_1}{\lambda_2} a^2 \rho_3 = 0.$$

If $\rho_2 \neq 0$, then the optimum is not attained for finite α. On the other hand, if

$$\lambda_M(D + \tilde{C}) = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

and $\rho_2 \neq 0$, then

$$\liminf_{\alpha\to\infty} \lambda_M(D + \alpha C) > \tfrac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3).$$

To conclude, we plotted two examples of $\lambda_M(D + 0.1\beta C)$ as a function of β from 1 to 160. For both cases,

$$D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{bmatrix}.$$

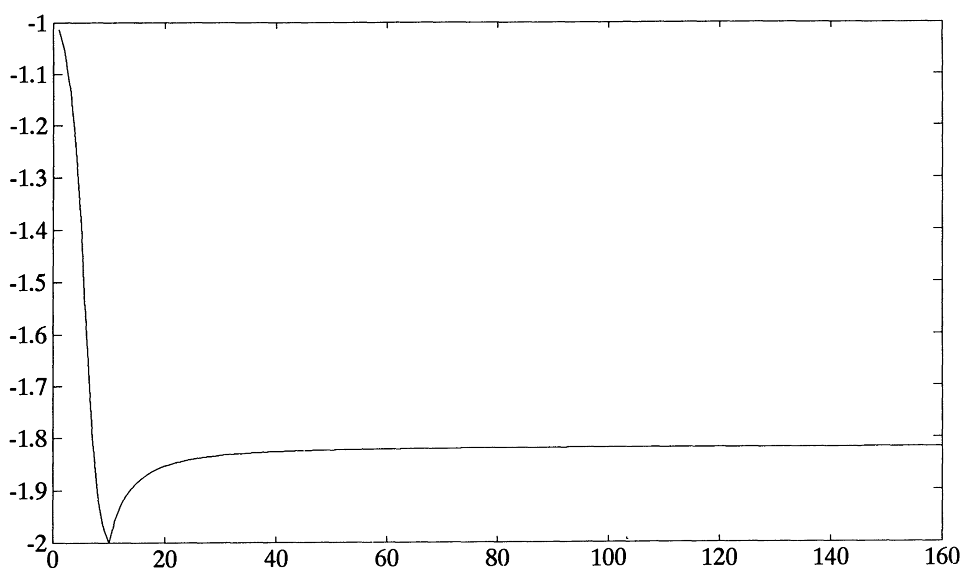

FIG. 1. *Rate of convergence of* $\lambda_M(D + 0.1\beta C)$ *as a function of* β:

$$D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & \sqrt{3} & 4 \\ -\sqrt{3} & 0 & 4 \\ -1 & -1 & 0 \end{bmatrix}.$$

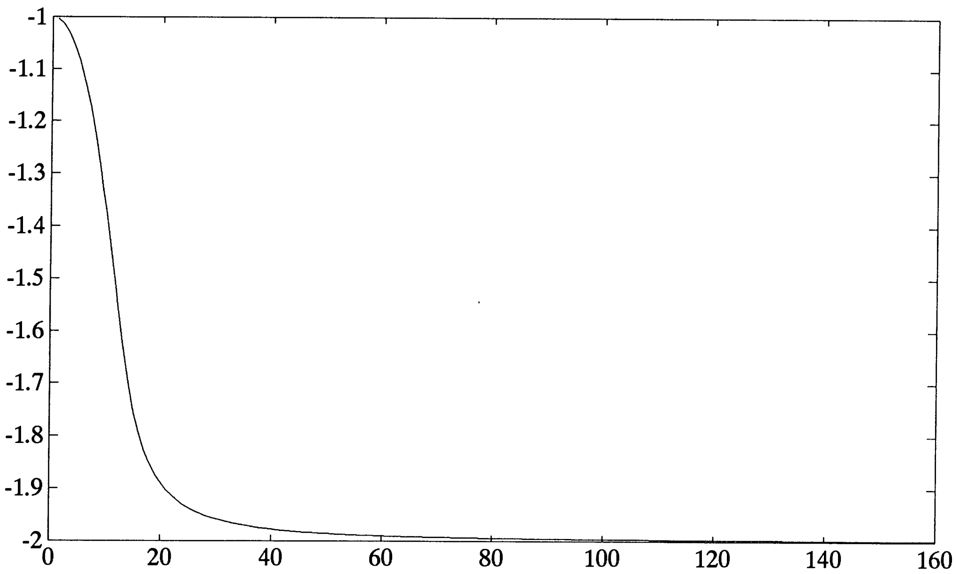

FIG. 2. *Rate of convergence of* $\lambda_M(D + 0.1\beta C)$ *as a function of* β:

$$D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -4 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 2 \\ -\frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix}.$$

C in Figure 1 is

$$\begin{bmatrix} 0 & \sqrt{3} & 4 \\ -\sqrt{3} & 0 & 4 \\ -1 & -1 & 0 \end{bmatrix},$$

which is chosen such that $\lambda_M(D + C)$ is the optimum -2. The plot shows $\lambda_M(D + \alpha C)$ is increasing for $\alpha = 0.1\beta > 1$. C in Figure 2 is

$$\begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 2 \\ -\frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix},$$

which is chosen such that $\liminf_{\alpha \to \infty} \lambda_M(D + \alpha C) = -2$. The plot shows that $\lambda_M(D + \alpha C)$ decreases to -2.

Acknowledgments. Part of this work was done while C.-R. Hwang was visiting the Division of Applied Mathematics, Brown University. He would like to thank Professor Ulf Grenander, Professor Stuart Geman and Professor Donald McClure for many inspiring discussions. S.-J. Sheu would like to thank the Center for Mathematical Sciences Research, Rutgers University, for their hospitality during his stay. The authors would like to express their gratitude to the referees for their very helpful comments and suggestions.

REFERENCES

AMIT, Y. (1991). On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. *J. Multivariate Anal.* **38** 82–99.

AMIT, Y. and GRENANDER, U. (1991). Comparing sweeping stretegies for stochastic relaxation. *J. Multivariate Anal.* **37** 197–222.

BARNETT, S. and STOREY, C. (1967). Analysis and synthesis of stability matrices. *J. Differential Equations* **2** 414–422.

BARONE, P. and FRIGESSI, A. (1990). Improving stochastic relaxation for Gaussian random fields. *Probab. Engrg. Inform. Sci.* **4** 369–389.

BELLMAN, R. (1970). *Methods of Nonlinear Analysis. Math. Sci. Engrg.* **61-I**. Academic, New York.

CHIANG, T.-S. and CHOW, Y. (1992). *A Comparison of Simulated Annealing of Gibbs Sampler and Metropolis Algorithms. Lecture Notes in Statist.* **74** 117–124. Springer, New York.

CHIANG, T.-S., HWANG, C.-R. and SHEU, S.-J. (1987). Diffusion for global optimization in $\mathbb{R}^n$. *SIAM J. Control Optim.* **25** 735–753.

DIACONIS, P. and STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.

FILL, J. A. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.

FRIGESSI, A., HWANG, C.-R. and YOUNES, L. (1992). Optimal spectral structures of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields. *Ann. Appl. Probab.* **2** 610–628.

FRIGESSI, A., HWANG, C.-R., SHEU, S.-J. and DI STEFANO, P. (1993). Convergence rates of the Gibbs sampler, the Metropolis algorithm, and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B* **55** 205–219.

GEMAN, S. and HWANG, C.-R. (1986). Diffusions for global optimization. *SIAM J. Control Optim.* **24** 1031–1043.

GOODMAN, J. and SOKAL, A. D. (1989). Multigrid Monte Carlo method. Conceptual foundations. *Phys. Rev. D* **40** 2035–2071.

GRENANDER, U. (1984). Tutorial in pattern theory. Div. Appl. Math., Brown Univ.

GRENANDER, U. and MILLER, M. I. (1992). Representations of knowledge in complex systems. Div. Appl. Math., Brown Univ.

HALMOS, P. R. (1958). *Finite-Dimensional Vector Spaces*, 2nd ed. Van Nostrand, New York.

HWANG, C.-R. and SHEU, S.-J. (1990). Large-time behavior of perturbed diffusion Markov processes with applications to the second eigenvalue problem for Fokker–Planck operators and simulated annealing. *Acta Appl. Math.* **19** 253–295.

KOLMOGOROV, A. (1937). Zür Umkehrbarkeit der statisticischen Naturgesetze. *Math. Ann.* **113** 766–772.

NAGASAWA, M. (1961). The adjoint process of a diffusion with reflecting barrier. *Kodai Math. Sem. Rep.* **13** 235–248.

REED, M. and SIMON, B. (1978). *Methods of Modern Mathematical Physics* **4**. Academic, New York.

ROYSAM, B., MILLER, M. I., BHATTACHARJYA, A. K. and TURNER, J. N. (1992). Combining stochastic and syntactic processing with analog computing methods. *Digital Signal Processing*. To appear.

STERNBERG, S. (1964). *Lectures on Differential Geometry*. Prentice-Hall, Englewood Cliffs, NJ.

TAUSSKY, O. (1967). Positive-definite matrices, "Inequalities." In *Proceedings of a Symposium Held at Wright-Patterson Air Force Base*, 309–319. Academic, New York.

VARADHAN, S. R. S. (1980). *Lectures on Diffusion Problems and Partial Differential Equations*. Springer, New York.

CHII-RUEY HWANG
SHUENN-JYI SHEU
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI, TAIWAN 11529

SHU-YIN HWANG-MA
DEPARTMENT OF BUSINESS MATHEMATICS
SOOCHOW UNIVERSITY
TAIPEI, TAIWAN 10001