CENTRAL LIMIT THEOREM
FOR PRODUCTS OF RANDOM MATRICES

BY

MARC A. BERGER

ABSTRACT. Using the semigroup product formula of P. Chernoff, a central limit theorem is derived for products of random matrices. Applications are presented for representations of solutions to linear systems of stochastic differential equations, and to the corresponding partial differential evolution equations. Included is a discussion of stochastic semigroups, and a stochastic version of the Lie-Trotter product formula.

1. Introduction. During the past thirty years there has been interest concerning the limiting behavior of products of random matrices. Bellman [1] was the first to examine the limiting behavior of $P_n = \prod_{i=1}^{n} D_i$ for an i.i.d. sequence of random $m \times m$ matrices $\{D_i = D_i(\omega)\}$. Under the assumption that the entries of the matrices D_i be strictly positive he showed that

$$(1.1) \qquad \lim_{n \to \infty} \frac{1}{n} \mathbf{E} \log(P_n)_{i,j}$$

exists, where $(P_n)_{i,j}$ denotes the i, j entry of P_n. In Furstenberg and Kesten [14] it was shown that

$$(1.2) \qquad \lim_{n \to \infty} \frac{1}{n} \log(P_n)_{i,j}$$

exists a.s. and is, in fact, independent of i, j. Thus P_n behaves like $e^{kn}J$, where J is the matrix with all entries one. There the sequence $\{D_i\}$ was only assumed to be stationary and metrically transitive. Again, the entries of the D_i were assumed to be strictly positive. (Bellman and Furstenberg and Kesten had other conditions as well.) In addition to the limit (1.2), Furstenberg and Kesten derived a central limit theorem showing that

$$(1.3) \qquad \sqrt{n}\left[\frac{1}{n}\log(P_n)_{i,j} - k\right]$$

tends in distribution to a normally distributed limit (or zero, in a degenerate case). These results involved stationary measures and eventually led to a general analysis of noncommutative "large numbers" phenomena on semisimple Lie groups. Some representative works are Furstenberg [12, 13], Grenander [17], Guivarc'h [18], Guivarc'h et al. [19] and Tutubalin [27], and the bibliographies there.

In this paper we are examining a different type of random matrix product. Let $\{D_{i,n} = D_{i,n}(\omega): 1 \leq i \leq n\}$ be a family of random square-integrable $m \times m$

---

matrices, with the property that $D_{1,n}, D_{2,n}, \ldots, D_{n,n}$ are i.i.d. for each fixed n. The first half of this paper is concerned with establishing the existence of and calculating the distribution of the limiting matrix product

$$\lim_{n\to\infty} \prod_{i=1}^{n} D_{i,n},$$

when $D_{i,n}$ has the form

$$D_{i,n} = I + A_i/\sqrt{n} + B_i/n + o(n^{-1})$$

for large n, with $\mathbf{E} A_i = 0$. In this framework $(A_1, B_1), (A_2, B_2), \ldots$ is an i.i.d. sequence of pairs of square-integrable $m \times m$ matrices. In particular, we obtain "law of large numbers" type results for $\prod_{i=1}^{n}(I + B_i/n)$ and "central limit" type results for $\prod_{i=1}^{n}(I + A_i/\sqrt{n})$, where $\mathbf{E} A_i = 0$. A related result is that of Watkins [28] who takes $D_{i,n} = e^{A_i/\sqrt{n}}$ (albeit in a more general setting than that of $m \times m$ matrices). This corresponds to taking $B_i = \frac{1}{2}A_i^2$ above. The second half of this paper is devoted to applications of this result. To carry this out we introduce the semigroup formula of P. Chernoff [8], as far as it can be applied to central limit type results. Such a connection has already been observed by Goldstein [16].

Unlike the commutative case, where the matrices in the range commute and may be simultaneously diagonalized, there is no connection in the noncommutative case between the asymptotic behavior of large products of terms with a fixed distribution and products of terms close to the identity. The former are multiplicative while the latter are (still) additive in nature.

2. P. Chernoff's product formula. Chernoff's product formula (Chernoff [8, 9]; see also Pazy [23; 24, Chapter 3, Corollary 5.4]) goes as follows.

THEOREM I. *Let $\{T(t): t \geq 0\}$ be a family of linear contraction operators on a Banach space $\mathcal{X}$. Let L be the infinitesimal generator of a C_0 semigroup $\{S_L(t): t \geq 0\}$ on $\mathcal{X}$. If*

$$(2.1) \qquad \lim_{t\downarrow 0} \frac{T(t)x - x}{t} = Lx, \qquad x \in D(L),$$

then

$$(2.2) \qquad \lim_{n\to\infty} T^n\left(\frac{t}{n}\right) x = S_L(t)x, \qquad x \in \mathcal{X}.$$

Chernoff's result can be used to prove the Lie-Trotter and resolvent product formulas. It becomes of fundamental interest in probability theory once we observe, as did Goldstein [16], that it is intimately connected with central limit type results. This connection becomes more marked when we examine stochastic integration, and even more marked when we study (two-parameter) evolution systems, as explained in Berger and Sloan [5]. Here we content ourselves with showing the reader how the multi-dimensional central limit theorem follows at once from Theorem I.

Recall the (simplest) multi-dimensional central limit theorem.

CENTRAL LIMIT THEOREM (MULTI-DIMENSIONAL) (CLT (M-DIM.)). *Let $X_1, X_2, \ldots$ be i.i.d. m-vectors having two moments—mean μ, covariance matrix*

$\Gamma = (\gamma_{ij})$. *Then*

$$(2.3) \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \overset{D}{\to} \mathcal{N}(0, \Gamma).$$

PROOF. Let $\mathbf{R}_m^*$ be the one-point compactification of $\mathbf{R}_m$, and define operators $T(t)$ on the Banach space $\mathcal{X} = C(\mathbf{R}_m^*)$, with the supremum norm, by

$$(2.4) \qquad T(t)f(x) = \mathbf{E}f(x + \sqrt{t}(X - \mu)).$$

These operators are all contractions on $\mathcal{X}$. Take

$$L = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \gamma_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

with $D(L) = \{f \in \mathcal{X} : \text{all first and second order derivatives of } f \text{ exist and are in } \mathcal{X}\}$ (the infinitesimal operator of an m-dimensional Brownian motion). For $f \in D(L)$

$$
\begin{aligned}
T(t)f(x) = \mathbf{E} &\left[f(x) + \sum_{i=1}^{m} \frac{\partial}{\partial x_i} f(x) \sqrt{t}(X - \mu)_i \right. \\
&\left. + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2}{\partial x_i \partial x_j} f(\xi_{x,t}) t(X - \mu)_i (X - \mu)_j \right] \\
= f(x) &+ \frac{1}{2} t \sum_{i=1}^{m} \sum_{j=1}^{m} \gamma_{ij} \frac{\partial^2}{\partial x_i \partial x_j} f(x) \\
&+ \frac{1}{2} t \mathbf{E} \sum_{i=1}^{m} \sum_{j=1}^{m} \left[\frac{\partial^2}{\partial x_i \partial x_j} f(\xi_{x,t}) - \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right] (X - \mu)_i (X - \mu)_j,
\end{aligned}
$$

where $\xi_{x,t}$ is a (random) point on the line (i.e. interval) between x and $x + \sqrt{t}(X - \mu)$.

To establish (2.1) we must show that for each i, j

$$(2.5) \qquad \lim_{t \downarrow 0} \sup_{x \in \mathbf{R}_m} |\mathbf{E}[g(\xi_{x,t}) - g(x)](X - \mu)_i (X - \mu)_j| = 0,$$

where $g = \partial^2 f / \partial x_i \partial x_j$. Indeed, since g is uniformly continuous for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$V_\delta(g) = \sup_{|x-y| \leq \delta} |g(x) - g(y)| \leq \varepsilon.$$

Thus

$$
\begin{aligned}
\mathbf{E}|g(\xi_{x,t}) &- g(x)| \, |(X - \mu)_i (X - \mu)_j| \\
&= \int_{|X-\mu| \leq \delta/\sqrt{t}} |g(\xi_{x,t}) - g(x)| \, |(X - \mu)_i (X - \mu)_j| \\
&\quad + \int_{|X-\mu| > \delta/\sqrt{t}} |g(\xi_{x,t}) - g(x)| \, |(X - \mu)_i (X - \mu)_j| \\
&\leq \varepsilon \sqrt{\gamma_{ii} \gamma_{jj}} + 2\|g\| \int_{|X-\mu| > \delta/\sqrt{t}} |(X - \mu)_i (X - \mu)_j|
\end{aligned}
$$

from which (2.5) follows. Hence it follows from Theorem I that

$$\lim_{n\to\infty} T^n\left(\frac{t}{n}\right) f(x) = S_L(t)f(x), \qquad f \in \mathcal{X}.$$

Since

$$(2.6) \qquad\qquad T^n(t)f(x) = \mathbf{E}f\left(x + \sqrt{t}\sum_{i=1}^n (X_i - \mu)\right)$$

and

$$(2.7) \qquad\qquad S_L(t)f(x) = \mathbf{E}f(x + \mathcal{N}(0,\Gamma t))$$

we have

$$\sqrt{t/n}\sum_{i=1}^n (X_i - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Gamma t),$$

and by setting $t = 1$ we have proved CLT (m-dim.). (Recall that to establish convergence in distribution it suffices to use even compactly supported continuous functions.) □

We want to emphasize here equation (2.6). What makes Chernoff's result so particularly attractive when dealing with central limit results is that the nth power of T corresponds to a sum of i.i.d. random variables $\sum_{i=1}^n (X_i - \mu)$, and the evaluation at time t/n corresponds to the normalization by $1/\sqrt{n}$. Thus $T^n(t/n)$ corresponds exactly to summing n i.i.d. variables and normalizing appropriately by $1/\sqrt{n}$. By taking note of this type of correspondence throughout our discussion the reader will very soon come to the realization that these central limit type results may, in fact, be "the perfect setting" for Chernoff's product formula.

One nice consequence of Theorem I as used in the above proof is that we immediately appreciate the reason why the normal distribution "pops up" here. It simply gives rise to a semigroup. Precisely,

THEOREM II. $\{T(t)\colon t \geq 0\}$ *defined above in* (2.4) *is a semigroup if and only if X is normally distributed.*

PROOF. The sufficiency is immediate. On the other hand, suppose $\{T(t)\colon t \geq 0\}$ is a semigroup. Then, on the one hand, $T^n(t/n) = T(t)$ while, on the other hand,

$$\operatorname*{st-lim}_{n\to\infty} T^n(t/n) = S_L(t).$$

Therefore

$$\mathbf{E}f(x + \sqrt{t}(X - \mu)) = T(t)f(x) = S_L(t)f(x) = \mathbf{E}f(x + \mathcal{N}(0,\Gamma t)). \quad □$$

Thus, in simple terms, since st-$\lim_{n\to\infty} T^n(t/n)$ has to be a semigroup, and since the normal distribution alone generates a semigroup, the limiting distribution must be normal.

3. Auxiliary matrices. In this section, we consider a random $m \times m$ matrix $A = (a_{ij})$ having two moments. Our main result is the existence of a natural number p and auxiliary matrices $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ satisfying

$$(3.1) \qquad\qquad \sum_{n=1}^p \sigma_{ik}^{(n)}\sigma_{jl}^{(n)} = \mathbf{E}a_{ik}a_{jl} \quad \forall i,j,k,l.$$

The number p and the matrices $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ are not unique, but we can make the claim that it is always possible to arrange things so that

$$(3.2) \qquad p \le m^2.$$

To see this simply multiply both sides of (3.1) by $\xi_{ik}\xi_{jl}$ and sum over i, j, k, l. Then it is seen that (3.1) is equivalent to the relation

$$(3.3) \quad \sum_{n=1}^{p} \left(\sum_{i=1}^{m} \sum_{k=1}^{m} \sigma_{ik}^{(n)} \xi_{ik} \right)^2 = \mathbf{E} \left(\sum_{i=1}^{m} \sum_{k=1}^{m} a_{ik} \xi_{ik} \right)^2, \qquad \xi \in \mathbf{R}_m \otimes \mathbf{R}_m \simeq \mathbf{R}_{m^2}.$$

The right-hand side of (3.3) is a nonnegative definite quadratic form in the m^2 variables ξ_{ik} ($i = 1, \ldots, m$; $k = 1, \ldots, m$). Hence it can be written as the sum of m^2 (or fewer) squares of linear forms in these variables.

The relation (3.1) can also be interpreted as follows. Let $X = (X_1, \ldots, X_p)$ be a p-dimensional random variable with mean zero and identity covariance matrix. Then the matrix $\sum_{n=1}^{p} \sigma^{(n)} X_n$ has the same mean and covariance as A.

For reference in the following section we note here two other equivalent forms of (3.1). To begin with, multiply both sides of (3.1) by $x_k x_l$ and sum over k, l. Then it is seen that (3.1) is also equivalent to

$$(3.4) \qquad \sum_{n=1}^{p} (\sigma^{(n)} x)_i (\sigma^{(n)} x)_j = \mathbf{E}(Ax)_i (Ax)_j, \qquad x \in \mathbf{R}_m, \ \forall i, j.$$

Next, multiply both sides of (3.4) by $\xi_i \xi_j$ and sum over i, j. Then it is seen that (3.1) is further equivalent to

$$(3.5) \qquad \sum_{n=1}^{p} \left(\langle \sigma^{(n)} x, \xi \rangle \right)^2 = \mathbf{E} \left(\langle Ax, \xi \rangle \right)^2, \qquad x, \xi \in \mathbf{R}_m.$$

In what follows we will assume that we have chosen p and $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ satisfying (3.1).

4. Notions of convergence. It is important to emphasize now that the results below regarding convergence of sequences of random matrices refer specifically to convergence in distribution in the *operator* sense. More specifically, if M_n is a sequence of random $m \times m$ matrices, we write $M_n \Rightarrow M$ if $M_n x$ converges in distribution to Mx for all $x \in \mathbf{R}_m$. This is weaker than (joint) convergence in distribution of the m^2-tuples of elements of the arrays M_n to the m^2-tuple M. However, it is strong enough for all the applications we introduce. Technically, this mode of convergence is referred to as strong operator convergence in distribution, the adjective strong indicating pointwise (as opposed to uniform). To distinguish between operator convergence in distribution and the stronger m^2-tuple convergence in distribution we adopt the notation "$\Rightarrow$" for the former, and reserve the notation "$\overset{\mathcal{D}}{\rightarrow}$" for the latter. The condition $M_n \Rightarrow M$ is also equivalent to the convergence in distribution of the quadratic forms $\langle M_n x, y \rangle$ to $\langle Mx, y \rangle$. Of course $\langle Mx, y \rangle$ is by no means the most general linear combination of the elements of M, and this is why "$\Rightarrow$" is weaker than "$\overset{\mathcal{D}}{\rightarrow}$". But "$\Rightarrow$" is certainly strong enough to conclude that any given row (column) of M_n converges (jointly) in distribution to the corresponding row (column) of M. This fact is used in Example I in §5.

Another condition equivalent to $M_n \Rightarrow M$ is

$$(4.1) \qquad \lim_{n \to \infty} \sup_{x \in B_R} |\mathbf{E}[f(M_n x) - f(Mx)]| = 0$$

for $f \in C(\mathbf{R}_m^*)$, $R > 0$, where B_R is the m-dimensional sphere $\{x \in \mathbf{R}_m : |x| < R\}$. Motivated by this we say the convergence is uniform, written $M_n \Rightarrow_{\mathrm{unif}} M$, if

$$(4.2) \qquad \lim_{n \to \infty} \sup_{x \in \mathbf{R}_M} |\mathbf{E}[f(M_n x) - f(Mx)]| = 0$$

for $f \in C(\mathbf{R}_m^*)$. Unlike the vector convergence in distribution "$\xrightarrow{\mathcal{D}}$", where uniformity is automatic, the two notions (4.1) and (4.2) are not equivalent. Clearly, if $M_n \to M$ in probability then $M_n \Rightarrow M$. However, uniform convergence in distribution of operators does not necessarily follow from any other type of convergence. Here is a trivial example: Let $M_n = 1/n$ on $\mathbf{R}$. It is not true that $M_n \Rightarrow_{\mathrm{unif}} 0$, although $M_n \xrightarrow{\mathcal{D}} 0$ and $M_n \Rightarrow 0$. Of course the problem here is the singular limit zero, as the following theorem intends to emphasize.

THEOREM III. *Let the random matrices $M_1, M_2, \ldots$ converge in probability to M. Suppose, in addition,*

$$(4.3) \qquad \sup_{H \in \mathrm{GL}_m} \mathbf{P}(|I - MH| < 1) = 1,$$

where GL_m denotes the group of invertible $m \times m$ matrices. Then $M_n \Rightarrow_{\mathrm{unif}} M$.

PROOF. Without loss of generality we may assume that

$$(4.4) \qquad \mathbf{P}(|I - M| < 1) = 1$$

since (i) $M_n H$ converges in probability to MH, and (ii)

$$\sup_{x \in \mathbf{R}_m} |\mathbf{E}[f(M_n x) - f(Mx)]| = \sup_{x \in \mathbf{R}_m} |\mathbf{E}[f(M_n Hx) - f(MHx)]|,$$

for any $H \in \mathrm{GL}_m$. Given $f \in C(\mathbf{R}_m^*)$ and $\varepsilon > 0$ choose $\tilde{R}$ and $\alpha < 1$ such that

$$\sup_{|x| \geq \tilde{R}, |y| \geq \tilde{R}} |f(x) - f(y)| \leq \frac{1}{3}\varepsilon, \quad \mathbf{P}(|I - M| < \alpha) \geq 1 - \frac{1}{6}\frac{\varepsilon}{\|f\|}.$$

Choose $\beta \in (\alpha, 1)$ and then find N large enough so that

$$\mathbf{P}(|M_n - M| \geq \beta - \alpha) \leq \frac{1}{6}\frac{\varepsilon}{\|f\|}$$

whenever $n \geq N$. Set $R = \tilde{R}/(1 - \beta)$. Then for all $x \in \mathbf{R}_m \backslash B_R$, $n \geq N$ we have

$$\mathbf{E}|f(M_n x) - f(Mx)| \leq \int_{|I-M| \geq \alpha} |f(M_n x) - f(Mx)|$$

$$+ \int_{\substack{|I-M| < \alpha \\ |I-M_n| < \beta}} |f(M_n x) - f(Mx)|$$

$$+ \int_{\substack{|I-M| < \alpha \\ |I-M_n| \geq \beta}} |f(M_n x) - f(Mx)|$$

$$\leq \frac{1}{3}\varepsilon + \frac{1}{3}\varepsilon + \int_{|M_n - M| \geq \beta - \alpha} |f(M_n x) - f(Mx)| \leq \varepsilon.$$

This shows that

$$(4.5) \qquad \lim_{\substack{n \to \infty \\ R \to \infty}} \sup_{x \in \mathbf{R}_m \setminus B_R} |\mathbf{E}[f(M_n x) - f(Mx)]| = 0.$$

Observing that $M_n \Rightarrow M$, so that (4.1) holds, suffices to conclude the proof. $\square$

From this follows the observation that if M is deterministic and invertible, then $M_n \Rightarrow_{\text{unif}} M$. (Indeed, (4.3) then holds if and only if M is invertible.)

5. Main result. For $m \times m$ matrices $b, \sigma^{(1)}, \sigma^{(2)}, \dots, \sigma^{(p)}$ consider the solution $\{\Phi(\tau, t) \colon 0 \leq \tau \leq t < \infty\}$ of the linear homogeneous stochastic differential equation

$$(5.1) \qquad d\Phi(\tau, t) = \sum_{n=1}^{p} \sigma^{(n)} \Phi(\tau, t) \, dW_n(t) + b\Phi(\tau, t) \, dt,$$

$$(5.2) \qquad \Phi(\tau, \tau) = I,$$

where $(W_1(t), W_2(t), \dots, W_p(t))$ is a p-dimensional Brownian motion. $\Phi(\tau, t)$ is a random $m \times m$ matrix, and is usually referred to as the fundamental matrix for the linear stochastic differential equation

$$(5.3) \qquad dx(t) = \sum_{n=1}^{p} \sigma^{(n)} x(t) \, dW_n(t) + bx(t) \, dt.$$

We note three basic properties of $\Phi(\tau, t)$:

(i) For fixed τ, there is a version of $\Phi(\tau, t)$ which is continuous in t.

(ii) In case the $p + 1$ matrices $\sigma^{(1)}, \sigma^{(2)}, \dots, \sigma^{(p)}, b$ are mutually commutative

$$(5.4) \qquad \Phi(\tau, t) = \exp\left\{ \sum_{n=1}^{p} \sigma^{(n)} W_n(t) + \left[b - \frac{1}{2} \sum_{n=1}^{p} (\sigma^{(n)})^2 \right] t \right\}.$$

(iii) In general, regardless of commutativity,

$$(5.5) \qquad \det \Phi(\tau, t) = \exp\left\{ \operatorname{tr}\left(\sum_{n=1}^{p} \sigma^{(n)} W_n(t) + \left[b - \frac{1}{2} \sum_{n=1}^{p} (\sigma^{(n)})^2 \right] t \right) \right\}.$$

The first two properties appear in any standard text on the subject of stochastic differential equations. Property (iii) appears in Berger [**2**]. We denote $\Phi(0, t)$ by $\Phi(t)$.

Our main result is

THEOREM IV. *Let $D_{i,n}$, $1 \leq i \leq n$, be square-integrable $m \times m$ matrices such that $D_{1,n}, D_{2,n}, \dots, D_{n,n}$ are i.i.d. for each fixed n. Assume $D_{i,n}$ to be of the form*

$$(5.6) \qquad D_{i,n} = I + \frac{A_i}{\sqrt{n}} + \frac{B_i}{n} + C_i(n),$$

where $\mathbf{E}A_i = 0$ and $\mathbf{E}|C_i(n)|^2 = o(n^{-2})$ for large n. Assume further that

(H) *for any two nonzero vectors $x, y \in \mathbf{R}_m$*

$$(5.7) \qquad \mathbf{P}(\langle Ax, y \rangle = 0) < 1.$$

Then

$$(5.8) \qquad\qquad \prod_{i=1}^{n} D_{i,n} \Rightarrow_{\text{unif}} \Phi(1),$$

where Φ *is defined in* (5.1) *and* (5.2), $b = \mathbf{E}B$ *and* $p, \sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ *satisfy* (3.1).

PROOF. Denote

$$E_i(t) = A_i\sqrt{t} + B_i t + C_i([1/t]).$$

Let $\mathcal{X} = C(\mathbf{R}_m^*)$, with the supremum norm, and define linear operators $T(t)$ by

$$(5.9) \qquad\qquad T(t)f(x) = \int_{|E(t)| \le 1/2} f(x + E(t)x).$$

These form a family of contractions on $\mathcal{X}$. Take

$$L = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left[\sum_{n=1}^{p} (\sigma^{(n)}x)_i (\sigma^{(n)}x)_j \right] \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{m} (bx)_i \frac{\partial}{\partial x_i}$$

with

$$D(L) = \left\{ f \in \mathcal{X} : \text{ all products } x_i \frac{\partial f}{\partial x_j}, \ x_i x_j \frac{\partial^2 f}{\partial x_k \partial x_l} \right.$$
$$\left. \text{exist and are in } \mathcal{X} \text{ and vanish at } x = 0 \text{ and as } |x| \to \infty \right\}$$

(the infinitesimal operator of an m-dimensional diffusion with drift $\mu(x) = bx$ and diffusion $\Lambda x = \sum_{n=1}^{p} \sigma^{(n)}x \otimes \sigma^{(n)}x$).

Notice, by virtue of (H), that the only degeneracy in L is at the origin since the principal symbol $P(x, \xi)$ of L is given by $P(x, \xi) = \mathbf{E}(\langle Ax, \xi \rangle)^2$. In fact, since P is biquadratic it follows directly from (5.7) that

$$(5.10) \qquad\qquad P(x, \xi) \ge \alpha |x|^2 |\xi|^2,$$

where $\alpha > 0$ is the minimum of P on the compact set $|x| = |\xi| = 1$. We compute,

assuming $f \in D(L)$,

$$T(t)f(x) = f(x)\mathbf{P}\left(|E(t)| \leq \frac{1}{2}\right) + tLf(x) - t^{1/2}\sum_{i=1}^{m}\left[\int_{|E(t)|>1/2}(Ax)_i\right]\frac{\partial}{\partial x_i}f(x)$$

$$- t\sum_{i=1}^{m}\left[\int_{|E(t)|>1/2}(Bx)_i\right]\frac{\partial}{\partial x_i}f(x)$$

$$+ \sum_{i=1}^{m}\left[\int_{|E(t)|\leq 1/2}\left(C\left(\left[\frac{1}{t}\right]\right)x\right)_i\right]\frac{\partial}{\partial x_i}f(x)$$

$$- \frac{1}{2}t\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\int_{|E(t)|>1/2}(Ax)_i(Ax)_j\right]\frac{\partial^2}{\partial x_i\partial x_j}f(x)$$

$$+ t^{3/2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\int_{|E(t)|\leq 1/2}(Ax)_i(Bx)_j\right]\frac{\partial^2}{\partial x_i\partial x_j}f(x)$$

$$+ \frac{1}{2}t^2\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\int_{|E(t)|\leq 1/2}(Bx)_i(Bx)_j\right]\frac{\partial^2}{\partial x_i\partial x_j}f(x)$$

$$+ t^{1/2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\int_{|E(t)|\leq 1/2}(Ax)_i\left(C\left(\left[\frac{1}{t}\right]\right)x\right)_j\right]\frac{\partial^2}{\partial x_i\partial x_j}f(x)$$

$$+ t\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\int_{|E(t)|\leq 1/2}(Bx)_i\left(C\left(\left[\frac{1}{t}\right]\right)x\right)_j\right]\frac{\partial^2}{\partial x_i\partial x_j}f(x)$$

$$+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\int_{|E(t)|\leq 1/2}\left(C\left(\left[\frac{1}{t}\right]\right)x\right)_i\left(C\left(\left[\frac{1}{t}\right]\right)x\right)_j\right]\frac{\partial^2}{\partial x_i\partial x_j}f(x)$$

$$+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\int_{|E(t)|\leq 1/2}\left[\frac{\partial^2}{\partial x_i\partial x_j}f(\xi_{x,t}) - \frac{\partial^2}{\partial x_i\partial x_j}f(x)\right](E(t)x)_i(E(t)x)_j$$

$$= f(x) + tLf(x) + o(t)$$

$$+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\int_{|E(t)|\leq 1/2}\left[\frac{\partial^2}{\partial x_i\partial x_j}f(\xi_{x,t}) - \frac{\partial^2}{\partial x_i\partial x_j}f(x)\right](E(t)x)_i(E(t)x)_j,$$

where $\xi_{x,t}$ is a (random) point on the line (i.e. interval) between x and $x + E(t)x$. It is important to note that this interval cannot pass through the origin (where the derivatives of f might be singular) since $|E(t)| \leq \frac{1}{2}$, and thus $|\xi_{x,t}| \geq \frac{1}{2}|x|$. Otherwise we could not use Taylor's theorem. If $x = 0$ then it is valid to set $\xi_{x,t} = 0$ in the above derivation and everything still holds (trivially). (Recall $Lf(0) = 0$ for $f \in D(L)$.) We have also relied above on the fact that $\mathbf{P}(|E(t)| > \frac{1}{2})$ is $o(t)$. This follows from the mean square integrability of A, B, C and the fact that $\mathbf{E}|C(n)|^2 = o(n^{-2})$.

In order to establish (2.1) it suffices now to show that

$$(5.11) \qquad \lim_{t\downarrow 0}\sup_{x\in\mathbf{R}_m}\int_{|E(t)|\leq 1/2}[g(\xi_{x,t}) - g(x)]e(t)x_k x_l = 0,$$

where $g = \partial^2 f / \partial x_i \partial x_j$, $e(t)$ is an integrable stochastic process which converges in mean to $e(0)$ as $t \downarrow 0$ and i, j, k, l are arbitrary. As above, let B_R denote the sphere $\{x \in \mathbf{R}_m : |x| < R\}$. We first show that for any $R > 0$

$$(5.12) \qquad \lim_{t \downarrow 0} \sup_{x \in B_R} |\mathbf{E}[\xi_k \xi_l g(\xi) - x_k x_l g(x)] e(t)| = 0,$$

where ξ is short for $\xi_{x,t}$. For any $\varepsilon > 0$ choose $\delta > 0$ such that $V_\delta(x_k x_l g) \leq \varepsilon$. Then for $x \in B_R$

$$\mathbf{E}|[\xi_k \xi_l g(\xi) - x_k x_l g(x)] e(t)|$$
$$\leq \varepsilon \mathbf{E}|e(t)| + \|x_k x_l g\| \int_{|E(t)x| > \delta} |e(t)|$$
$$\leq \varepsilon \mathbf{E}|e(t)| + \|x_k x_l g\| \int_{|E(t)| > \delta/R} |e(t)|.$$

Next we show that

$$(5.13) \qquad \lim_{\substack{R \uparrow \infty \\ t \downarrow 0}} \sup_{x \in \mathbf{R}_m \setminus B_R} |\mathbf{E}[\xi_k \xi_l g(\xi) - x_k x_l g(x)] e(t)| = 0.$$

For any $\varepsilon > 0$ choose $R > 0$ such that $\sup_{|x|, |y| \geq R/2} |y_k y_l g(y) - x_k x_l g(x)| < \varepsilon$. Then for $x \in \mathbf{R}_m \setminus B_R$

$$\mathbf{E}|[\xi_k \xi_l g(\xi) - x_k x_l g(x)] e(t)| \leq \varepsilon \mathbf{E}|e(t)| + 2\|x_k x_l g\| \int_{|E(t)| > 1/2} |e(t)|.$$

Next we show that

$$(5.14) \qquad \lim_{\substack{r \downarrow 0 \\ t \downarrow 0}} \sup_{x \in B_r} \left| \int_{|E(t)| \leq 1/2} [\xi_k \xi_l - x_k x_l] g(\xi) e(t) \right| = 0.$$

As above, for any $\varepsilon > 0$ choose $r > 0$ such that $\sup_{|x| \leq 3r/2} |x|^2 |g(x)| < \varepsilon$. Then for $x \in B_r$

$$\int_{|E(t)| \leq 1/2} |[\xi_k \xi_l - x_k x_l] g(\xi) e(t)| \leq 5\varepsilon \mathbf{E}|e(t)|,$$

where we have used the fact that

$$|x - y| \leq \tfrac{1}{2}|x| \rightarrow |x_k x_l - y_k y_l| \leq 5|y|^2.$$

Next we show that for any $0 < r < R$

$$(5.15) \qquad \lim_{t \downarrow 0} \sup_{x \in B_R \setminus B_r} \left| \int_{|E(t)| \leq 1/2} [\xi_k \xi_l - x_k x_l] g(\xi) e(t) \right| = 0.$$

Again, for any $\varepsilon > 0$ choose $0 < \delta < r$ such that

$$\sup_{\substack{x,y \in B_{R+\delta} \\ |x-y| \leq \delta}} |x_k x_l - y_k y_l| < \varepsilon.$$

Let $M = \sup_{x \in B_{R+\delta} \setminus B_{r-\delta}} |g(x)|$. Then for $x \in B_R \setminus B_r$

$$\int_{|E(t)| \leq 1/2} |[\xi_k \xi_l - x_k x_l] g(\xi) e(t)|$$
$$\leq \varepsilon M \mathbf{E}|e(t)| + 5\|x_k x_l g\| \int_{|E(t)| > \delta/R} |e(t)|.$$

Finally we show that

$$(5.16) \qquad \lim_{\substack{R\uparrow\infty \\ t\downarrow 0}} \sup_{x\in\mathbf{R}_m\backslash B_R} \left| \int_{|E(t)|\leq 1/2} [\xi_k\xi_l - x_k x_l]g(\xi)e(t) \right| = 0.$$

For any $\varepsilon > 0$ choose $R > 0$ such that $\sup_{x\in R_m\backslash B_{R/2}} |x|^2|g(x)| < \varepsilon$. Then for $x \in B_R$

$$\int_{|E(t)|\leq 1/2} |[\xi_k\xi_l - x_k x_l]g(\xi)e(t)| \leq 5\varepsilon\mathbf{E}|e(t)|.$$

Putting together (5.12)–(5.16) gives us (5.11). Thus, the hypotheses of Theorem I are satisfied.

It thus follows that

$$\lim_{n\to\infty} T^n\left(\frac{t}{n}\right)f(x) = S_L(t)f(x), \qquad f \in \mathcal{X}.$$

Since

$$(5.17) \qquad T^n(t)f(x) = \int_{|E_1(t)|\vee|E_2(t)|\vee\cdots\vee|E_n(t)|\leq 1/2} f\left(\prod_{i=1}^{n}(I + E_i(t))x\right)$$

and

$$(5.18) \qquad S_L(t)f(x) = \mathbf{E}f(\Phi(t)x),$$

it is only necessary to show that

$$(5.19) \qquad \lim_{n\to\infty} \mathbf{P}\left(\left|E_1\left(\frac{t}{n}\right)\right| \vee \left|E_2\left(\frac{t}{n}\right)\right| \vee \cdots \vee \left|E_n\left(\frac{t}{n}\right)\right| > \frac{1}{2}\right) = 0.$$

But this follows at once from the fact that $\mathbf{P}(|E(t/n)| > \frac{1}{2})$ is $o(n^{-1})$, as we indicated above. Hence

$$\lim_{n\to\infty} \mathbf{E}f\left(\prod_{i=1}^{n}\left(I + E_i\left(\frac{t}{n}\right)\right)x\right) = \mathbf{E}f(\Phi(t)x)$$

uniformly in x for all $f \in \mathcal{X}$. Now set $t = 1$ and (5.8) follows at once. $\square$

Hypothesis (H) is a nondegeneracy condition and can be interpreted algebraically as follows. Given general matrices $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ we say they form a *spanning set* provided

$$(5.20) \qquad \mathrm{sp}\{\sigma^{(1)}x, \sigma^{(2)}x, \ldots, \sigma^{(p)}x\} = \mathbf{R}_m \quad \forall x \in \mathbf{R}_m\backslash 0.$$

With these matrices we can associate a bilinear mapping $Q\colon \mathbf{R}^m \times \mathbf{R}^m \to \mathbf{R}^p$ defined by $Q_i(x,y) = \langle\sigma^{(i)}x, y\rangle$; $i = 1, \ldots, p$. We can also associate a biquadratic form P on $\mathbf{R}^m \times \mathbf{R}^m$ by $P(x,y) = \|Q(x,y)\|^2$. In general, a bilinear mapping Q is said to be *definite* if $Q(x,y) = 0 \leftrightarrow x = 0 \vee y = 0$. A positive semidefinite biquadratic form P is said to be *coercive* if $P(x,y) \geq \alpha|x|^2|y|^2$ for some $\alpha > 0$.

We observe that if $p, \sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ satisfy (3.1) then, according to (3.5), the associated biquadratic form is given by $P(x,y) = \mathbf{E}(\langle Ax, y\rangle)^2$. The following result is then immediate.

THEOREM V. *Let A be a random square integrable $m \times m$ matrix and let $p, \sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ satisfy (3.1). The following conditions are equivalent.*

  (i) *$\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ form a spanning set.*

  (ii) *The associated bilinear mapping Q is definite.*

  (iii) *The associated biquadratic form P is coercive.*

  (H) *For any two nonzero vectors $x, y \in \mathbf{R}_m$, $\mathbf{P}(\langle Ax, y \rangle = 0) < 1$.*

Note that for biquadratic forms it follows from scaling that positive definiteness and coercivity are equivalent, as we already indicated above in (5.10).

The coercivity condition (iii) has appeared in the literature of stochastic differential equations, and is known to be a sufficient condition for normalized solutions $x/|x|$ of (5.3) to be ergodic on S_{m-1} (see Khasminskii [21]). Pinsky [26] analyzes the existence of $\lim_{|x| \to 0} f(x)$ for solutions f of $Lf = 0$ in the punctured disc $0 < |x| < \delta$ under the coercivity assumption. Our sole concern is being able to use the semigroup $S_L(t)$ on $C(\mathbf{R}_m^*)$ (rather than on $C(K^*)$ for some cone K, e.g. the positive octant, properly contained in $\mathbf{R}_m$), and for this we need to know that $f = (\lambda - L)^{-1} g$, as well as all the terms appearing in Lf, behave well across the degeneracies of L, for $\lambda > 0$ and $g \in C(\mathbf{R}_m^*)$. More precisely, we need to know that $f = (\lambda - L)^{-1} g$ has a single limit as we approach any x at which L degenerates, and that Lf tends to zero at these points.

A detailed elaboration of the notion of a spanning set of matrices, in the more general case of rectangular matrices, appears in Berger and Friedland [3]. Included there are results concerning the smallest number $\sigma(m, n)$ of $m \times n$ matrices which can possibly form a spanning set.

We concern ourselves next with the degenerate case, where (H) fails.

THEOREM VI. *If hypothesis (H) is dropped in Theorem IV, then $\prod_{i=1}^{n} D_{i,n} \Rightarrow \Phi(1)$ (without the uniformity this time).*

PROOF. Suppose the random matrix A does not satisfy (H). (It is not "random enough".) In particular, in the framework of §3 and Theorem V, let $A \simeq \sum_{n=1}^{p} \sigma^{(n)} X_n$, where $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ do not form a spanning set. Here the symbol $\simeq$ denotes equality as regards the first two (multivariate) moments. (Refer back to the second paragraph of §3 for a minute.) Find matrices $\sigma^{(p+1)}, \sigma^{(p+2)}, \ldots, \sigma^{(\tilde{p})}$ so that the complete set $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(\tilde{p})}$ does form a spanning set. Choose additional (scalar) random variables $X_{p+1}, X_{p+2}, \ldots, X_{\tilde{p}}$ which are independent of all the matrices $D_{i,n}$ and themselves have mean zero and identity covariance matrix. Define i.i.d. matrices $\tilde{A}_i$ all according to $\sum_{n=p+1}^{\tilde{p}} \sigma^{(n)} X_n$, independent of all the matrices $D_{i,n}$. Let $P_n = \prod_{i=1}^{n} D_{i,n}$ and set, for $\varepsilon > 0$,

$$(5.21) \qquad P_n(\varepsilon) = \prod_{i=1}^{n} \left(D_{i,n} + \varepsilon \frac{\tilde{A}_i}{\sqrt{n}} \right).$$

Clearly we can apply Theorem IV to the matrices $P_n(\varepsilon)$. In fact, the principal symbol of the partial differential operator $L(\varepsilon)$, which arises from these matrices, is

$$(5.22) \qquad P_\varepsilon(x, y) = \sum_{n=1}^{p} (\langle \sigma^{(n)} x, y \rangle)^2 + \varepsilon^2 \sum_{n=p+1}^{\tilde{p}} (\langle \sigma^{(n)} x, y \rangle)^2.$$

According to Theorem IV, for each fixed $\varepsilon > 0$

$$(5.23) \qquad P_n(\varepsilon) \Rightarrow_{\text{unif}} \Phi_\varepsilon(1) \quad \text{as } n \to \infty,$$

where

$$(5.24) \qquad \begin{aligned} d\Phi_\varepsilon(\tau, t) &= \sum_{n=1}^{p} \sigma^{(n)} \Phi_\varepsilon(\tau, t)\, dW_n(t) \\ &\quad + \varepsilon \sum_{n=p+1}^{\tilde{p}} \sigma^{(n)} \Phi_\varepsilon(\tau, t)\, dW_n(t) + b\Phi_\varepsilon(\tau, t)\, dt. \end{aligned}$$

It is also clear that $\Phi_\varepsilon(1)$ converges to $\Phi(1)$ in mean square as $\varepsilon \downarrow 0$, and thus $\Phi_\varepsilon(1) \Rightarrow \Phi(1)$. Here is where we actually get around the degeneracy, for irregardless of how degenerate the operator L in the proof of Theorem IV is, the semigroup generated by $L(\varepsilon)$ always converges to that generated by L in the topology of uniform convergence on compact sets. This is simply because the solution of a stochastic differential equation depends continuously on the coefficients (see, for example, Gihman and Skorohod [15, §I.7]).

Finally, we examine $P_n(\varepsilon) - P_n$. We make the following observation. Let $M_1, M_2,$ $\ldots, M_n$ be random matrices with means and covariances, $\mathbf{E}M_i M_j^T$, zero and finite second moments. There exists a constant $\beta > 0$, dependent only on the matrix norm we are using (and the dimension m–but not on the matrices M_i, nor on n), such that

$$(5.25) \qquad \mathbf{E}\left| \sum_{i=1}^{n} M_i \right|^2 \le \beta \sum_{i=1}^{n} \mathbf{E}|M_i|^2.$$

Now by expanding (5.21) we can write

$$(5.26) \qquad P_n(\varepsilon) = P_n + \sum_{k=1}^{n} \varepsilon^k M_{k,n},$$

where $M_{k,n}$ is itself an algebraic sum of $\binom{n}{k}$ means and covariances zero square integrable matrices. For example,

$$\begin{aligned} M_{1,n} = \frac{1}{\sqrt{n}} \Big[&D_{n,n} D_{n-1,n} \cdots D_{2,n} \tilde{A}_1 + D_{n,n} D_{n-1,n} \cdots \tilde{A}_2 D_{1,n} \\ &+ \cdots + D_{n,n} \tilde{A}_{n-1} \cdots D_{2,n} D_{1,n} + \tilde{A}_n D_{n-1,n} \cdots D_{2,n} D_{1,n} \Big]. \end{aligned}$$

By using (5.25) it follows that

$$\mathbf{E}|M_{k,n}|^2 \le \beta \binom{n}{k} \frac{(\mathbf{E}|A|^2)^k (\mathbf{E}|D_{1,n}|^2)^{n-k}}{n^k}.$$

From the form of $D_{1,n}$ it is clear that $\mathbf{E}|D_{1,n}|^2 \le 1 + K/n$ for some $K > 0$, and thus

$$(5.27) \qquad \sup_{n \ge k} \mathbf{E}|M_{k,n}|^2 \le \gamma \frac{\alpha^k}{k!}$$

for $\gamma = \beta e^K$ and $\alpha = \mathbf{E}|A|^2$. From (5.27) follows at once

$$(5.28) \qquad \limsup_{\varepsilon \downarrow 0} \sup_{n} \mathbf{E}|P_n(\varepsilon) - P_n|^2 = 0,$$

and, therefore, for any $f \in C(\mathbf{R}_m^*)$ and $R > 0$

$$(5.29) \qquad \limsup_{\varepsilon \downarrow 0} \sup_n \sup_{x \in B_R} |\mathbf{E}[f(P_n(\varepsilon)x) - f(P_n x)]| = 0.$$

Putting this together with the limits $P_n(\varepsilon) \Rightarrow \Phi_\varepsilon(1)$ as $n \to \infty$, and $\Phi_\varepsilon(1) \Rightarrow \Phi(1)$ as $\varepsilon \downarrow 0$ gives the desired result $P_n \Rightarrow \Phi(1)$. $\square$

There is an interesting observation to make now, based on Theorem III. If $A = 0$ then $\Phi(1)$ is deterministic and, just as with ordinary convergence in distribution, strong operator convergence in distribution to an atom implies convergence in probability. Since $\Phi(1)$ is invertible Theorem III applies (see the remark which follows the proof of this theorem), and we conclude that, in fact,

$$\prod_{i=1}^n D_{i,n} \Rightarrow_{\text{unif}} \Phi(1),$$

even in this most degenerate case.

What happens when $\Phi(1)$ is random, yet, according to (5.5), (everywhere) invertible? This is precisely our setup in Theorem VI. Theorem IV established uniform convergence in distribution by using the nice structure of $D(L)$—guaranteed by (H). In Theorem VI we forfeit this for the weaker conclusion of uniformity only on compact sets. We will shortly see in §6 the consequences of this in terms of approximating solutions for the Cauchy problem $du/dt = Lu$.

We conclude this section with three examples.

EXAMPLE I (CLT, M-DIM.). Let $X_1, X_2, \ldots$ be i.i.d. m-vectors with mean μ and covariance matrix Γ. Write $\Gamma = SS^T$ where the dimensions of S are $m \times p$. Take

$$D_{i,n} = I + \begin{pmatrix} 0 & 0 & \ldots & 0 & \uparrow \\ 0 & 0 & \ldots & 0 & \\ \ldots & \ldots & \ldots & \ldots & \frac{X_i - \mu}{\sqrt{n}} \\ 0 & 0 & \ldots & 0 & \downarrow \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

Then

$$\prod_{i=1}^n D_{i,n} = I + \begin{pmatrix} 0 & 0 & \ldots & 0 & \uparrow \\ 0 & 0 & \ldots & 0 & \\ \ldots & \ldots & \ldots & \ldots & \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \\ 0 & 0 & \ldots & 0 & \downarrow \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

We can take

$$\sigma^{(j)} = \begin{pmatrix} 0 & 0 & \ldots & 0 & s_{1j} \\ 0 & 0 & \ldots & 0 & s_{2j} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & s_{mj} \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix};$$

$j = 1, 2, \ldots, p$. Then

$$\Phi(t) = \exp \begin{pmatrix} 0 & 0 & \ldots & 0 & \uparrow \\ 0 & 0 & \ldots & 0 & \\ \ldots & \ldots & \ldots & \ldots & \mathcal{N}(0, \Gamma t) \\ 0 & 0 & \ldots & 0 & \downarrow \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}$$

$$= I + \begin{pmatrix} 0 & 0 & \ldots & 0 & \uparrow \\ 0 & 0 & \ldots & 0 & \\ \ldots & \ldots & \ldots & \ldots & \mathcal{N}(0, \Gamma t) \\ 0 & 0 & \ldots & 0 & \downarrow \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

Note that we have embedded a nonhomogeneous equation into a homogeneous one by insertion of a dummy variable. The relevant diffusion here is $x(t) = SW(t)$, where $W(t) = (W_1(t), W_2(t), \ldots, W_p(t))$ is a p-dimensional Brownian motion. Let s_j be the jth column of S. Then we convert the nonhomogeneous equation $dx(t) = \sum_{j=1}^{p} s_j\, dW_j(t)$ into a homogeneous equation by insertion of the dummy variable $x_{m+1} \equiv 1$. Then the system becomes

$$d \begin{pmatrix} x_1(t) \\ x_2(t) \\ \ldots \\ x_m(t) \\ x_{m+1}(t) \end{pmatrix} = \sum_{j=1}^{p} \begin{pmatrix} 0 & 0 & \ldots & 0 & \uparrow \\ 0 & 0 & \ldots & 0 & s_j \\ \ldots & \ldots & \ldots & \ldots & \downarrow \\ 0 & 0 & \ldots & 0 & \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \ldots \\ x_m(t) \\ x_{m+1}(t) \end{pmatrix} dW_j(t)$$

with an extra dimension ($m + 1$ instead of m); hence the matrices $\sigma^{(j)}$ in the example. It is important to note that the product of any two of these nilpotent matrices is zero, and we compute Φ directly from (5.4). (Everything commutes and there are no second order terms.)

Clearly any linear system, homogeneous or not, can always be converted into a homogeneous system by insertion of an extraneous variable in this manner. This "nilpotent" change converts additive functionals (cf. the operators $T(t)$ in (2.4)) into multiplicative ones (cf. the operators $T(t)$ in (5.9)). For example,

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a+b \\ 0 & 1 \end{pmatrix}.$$

EXAMPLE II (WEAK LAW OF LARGE NUMBERS). Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean μ. Take $D_{i,n} = 1 + X_i/n$. Then

$$\prod_{i=1}^{n} \left(1 + \frac{X_i}{n}\right) \xrightarrow{D} e^{\mu}.$$

This is equivalent to the Weak Law of Large Numbers. Here we surely need to use Theorem VI rather than Theorem IV (but see the remark following the proof of Theorem VI), and although these theorems call for the existence of two moments, one can start back directly from Theorem I and see that only one is necessary.

Precisely, one can show that

THEOREM VII. *Let $D_{i,n}$, $1 \le i \le n$, be integrable $m \times m$ matrices such that $D_{1,n}, D_{2,n}, \ldots, D_{n,n}$ are i.i.d. for each fixed n. Assume $D_{i,n}$ to be of the form*

$$(5.30) \qquad\qquad D_{i,n} = I + B_i/n + C_i(n),$$

where $\mathbf{E}|C_i(n)| = o(n^{-1})$ for large n. Then

$$(5.31) \qquad\qquad \prod_{i=1}^{n} D_{i,n} \Rightarrow_{\text{unif}} \exp(\mathbf{E}B).$$

Our next example is intended to launch us off into §6.

EXAMPLE III. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with mean zero and variance one. Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

and take $D_{i,n} = e^{AX_i/\sqrt{n}} e^{BX_i/\sqrt{n}}$. Then

$$\prod_{i=1}^{n} D_{i,n} = \begin{pmatrix} e^{1/\sqrt{n}\sum_{i=1}^{n} X_i} & \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e^{1/\sqrt{n}\sum_{j=1}^{i} X_j} \\ 0 & 1 \end{pmatrix}.$$

We can take $p = 1$, $\sigma^{(1)} = A+B$, $b = \frac{1}{2}A+B$. Note that $\sigma^{(1)}$ and b do not commute so we cannot solve for $\Phi(t)$ from (5.14). However, things are so degenerate that we can solve (5.1) directly, and we find that

$$\Phi(t) = \begin{pmatrix} e^{W(t)} & e^{W(t)} \int_0^t e^{-W(s)}\, dW(s) \\ 0 & 1 \end{pmatrix}.$$

In particular, we see here that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i e^{(1/\sqrt{n})\sum_{j=1}^{i} X_j} \xrightarrow{D} e^{W(1)} \int_0^1 e^{-W(t)}\, dW(t).$$

It is interesting to contrast this with

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i e^{(1/\sqrt{n})\sum_{j=1}^{i-1} X_j} \xrightarrow{D} \int_0^1 e^{W(t)}\, dW(t),$$

which can be obtained by interchanging A and B in the above example. In this type of application it is always helpful to make the "convenient choice" $X_i = \sqrt{n}[W(i/n) - W((i-1)/n)]$, and to think in terms of approximating sums for Itô stochastic integrals.

6. Stochastic product formulas. Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with mean zero and variance one, and let L and M be two general (deterministic) $m \times m$ matrices. We can use Theorem VI to determine the limiting distribution of

$$\prod_{i=1}^{n} e^{LX_i/\sqrt{n}} e^{MX_i/\sqrt{n}}.$$

Indeed by using the approximation

$$e^{LX_i/\sqrt{n}} \sim I + L\frac{X_i}{\sqrt{n}} + \frac{1}{2}L^2\frac{X_i^2}{n}$$

we see that we can take $p = 1$, $\sigma^{(1)} = L + M$ and $b = \frac{1}{2}L^2 + LM + \frac{1}{2}M^2$. Since $\sigma^{(1)}$ and b do not necessarily commute we cannot readily compute Φ from (5.4). However, we can readily compute the limiting distribution of

$$\prod_{i=1}^{n} e^{LX_i/\sqrt{n}} e^{MX_i/\sqrt{n}} e^{-(L^2/2+LM+M^2/2)/n}.$$

In this case $p = 1$, $\sigma^{(1)} = L + M$ and $b = 0$. Thus,

(6.1)
$$\prod_{i=1}^{n} e^{LX_i/\sqrt{n}} e^{MX_i/\sqrt{n}} e^{-(L^2/2+LM+M^2/2)/n}$$
$$\Rightarrow e^{(L+M)\mathcal{N}(0,1)-(L+M)^2/2}.$$

Now consider the stochastic system

(6.2)
$$d\Phi(\tau, t) = (L + M)\Phi(\tau, t)\,dW(t),$$

(6.3)
$$\Phi(\tau, \tau) = I.$$

Denote the solution by $\Phi_{L+M}(\tau, t)$. Suppose we solve (6.2), (6.3) using L alone (instead of $L + M$) between times $0, t/n$. Then we solve this sytem using M alone between these same times. Then we multiply these two solutions together and repeat this procedure again between times t/n, $2t/n$ and then between times $2t/n$, $3t/n$, etc. up to time t (the method of "fractional steps"). Based on our intuition from the Lie-Trotter formula and based on (5.4) we might expect that

(6.4)
$$\prod_{i=1}^{n} \Phi_L\left(\frac{i-1}{n}t, \frac{i}{n}t\right)\Phi_M\left(\frac{i-1}{n}t, \frac{i}{n}t\right)e^{-tLM/n} \Rightarrow \Phi_{L+M}(0, t).$$

That this is indeed the case follows at once from (6.1). Simply note that

$$\Phi_L\left(\frac{i-1}{n}t, \frac{i}{n}t\right) = e^{L[W(it/n)-W((i-1)t/n)]-(1/2)L^2t/n},$$

which is distributed according to $e^{L\sqrt{t/n}W_i(1)-(1/2)L^2t/n}$, where $W_i(t)$ is the ith component of an n-dimensional Brownian motion $(W_1(t), W_2(t), \ldots, W_n(t))$.

Consider next the stochastic system

(6.5)
$$d\Phi(\tau, t) = L\Phi(\tau, t)\,dW_1(t) + M\Phi(\tau, t)\,dW_2(t),$$

(6.6)
$$\Phi(\tau, \tau) = I.$$

Again, denote the solution by $\Phi_{L+M}(\tau, t)$, and let $\Phi_L(\tau, t)$ and $\Phi_M(\tau, t)$ denote the solutions to (6.5) when we take $M = 0$ and $L = 0$, respectively. This time, based on (5.4), we do not expect any mixed terms to arise, and simply that

(6.7)
$$\prod_{i=1}^{n} \Phi_L\left(\frac{i-1}{n}t, \frac{i}{n}t\right)\Phi_M\left(\frac{i-1}{n}t, \frac{i}{n}t\right) \Rightarrow \Phi_{L+M}(0, t).$$

This follows directly from Theorem VI. This time let $(X_1, Y_1), (X_2, Y_2), \ldots$ be an i.i.d. sequence of random vectors with mean zero and covariance matrix I. Then

$$(6.8) \qquad \prod_{i=1}^{n} e^{LX_i/\sqrt{n}} e^{MY_i/\sqrt{n}} e^{-(L^2/2 + M^2/2)/n} \Rightarrow \Phi_{L+M}(1).$$

To establish (6.7) from (6.8) make the "convenient choice"

$$X_i = \sqrt{n/t} \left[W_1\left(\frac{i}{n}t\right) - W_1\left(\frac{i-1}{n}t\right) \right],$$
$$Y_i = \sqrt{n/t} \left[W_2\left(\frac{i}{n}t\right) - W_2\left(\frac{i-1}{n}t\right) \right].$$

With this motivation we now state

THEOREM VIII. *Let $(z_1(t), z_2(t), \ldots, z_p(t))$ be p-dimensional white noise. Denote*

$$L\Phi = \sum_{k=1}^{p} \sigma_L^{(k)} \Phi z_k + b_L \Phi,$$

where all matrices $\sigma_L^{(k)}, b_L$ are $m \times m$. Let $\Phi_L(\tau, t)$ denote the solution of the (Itô) stochastic differential equation

$$d\Phi_L(\tau, t) = L\Phi_L(\tau, t)\, dt, \quad \Phi_L(\tau, \tau) = O.$$

Then

$$(6.9) \qquad \begin{aligned} &\prod_{i=1}^{n} \prod_{k=1}^{p} \left\{ \exp\left(\sigma_L^{(k)} \left[W_k\left(\frac{i}{n}t\right) - W_k\left(\frac{i-1}{n}t\right) \right] \right) \right\} \\ &\times \exp\left(\left[b_L - \frac{1}{2} \sum_{k=1}^{p} (\sigma_L^{(k)})^2 \right] \frac{t}{n} \right) \Rightarrow \Phi_L(0, t). \end{aligned}$$

PROOF. This follows at once from repeated application of (6.7) (or, of course, directly from Theorem VI as in (6.8)). □

An immediate consequence of Theorem VIII is

THEOREM IX. *Let $\Phi_L(\tau, t)$ be as in Theorem VIII. Then*

$$(6.10) \qquad \mathbf{E} \left| \Phi_L(0, t) - I - \sum_{k=1}^{p} \sigma_L^{(k)} W_k(t) \right|^2 = O(t^2).$$

From this result we can obtain our generalized Lie-Trotter product formula for linear stochastic systems (the fundamental solutions of which we shall soon "dub" stochastic semigroups).

THEOREM X. *Let $(z_1(t), z_2(t), \ldots, z_p(t))$ be p-dimensional white noise. Denote*

$$L\Phi = \sum_{k=1}^{p} \sigma_L^{(k)} \Phi z_k + b_L \Phi, \quad M\Phi = \sum_{k=1}^{p} \sigma_M^{(k)} \Phi z_k + b_M \Phi,$$

where all matrices $\sigma_L^{(k)}, \sigma_M^{(k)}, b_L, b_M$ are $m \times m$. Let $\Phi_L(\tau, t)$ denote the solution of the (Itô) stochastic differential equation

$$d\Phi_L(\tau, t) = L\Phi_L(\tau, t)\, at, \quad \Phi_L(\tau, t) = I,$$

with similar meaning attached to $\Phi_M(\tau, t), \Phi_{L+M}(\tau, t)$. *Let*

$$(6.11) \qquad \sigma_* = \sum_{k=1}^{p} \sigma_L^{(k)} \sigma_M^{(k)}.$$

Then

$$(6.12) \qquad \prod_{i=1}^{n} \Phi_L\left(\frac{i-1}{n}t, \frac{i}{n}t\right) \Phi_M\left(\frac{i-1}{n}t, \frac{i}{n}t\right) e^{-t\sigma_*/n} \Rightarrow \Phi_{L+M}(0,t).$$

PROOF. By virtue of Theorem VI we need only show that

$$(6.13) \qquad \Phi_L(0,t) = I + \sqrt{t}A_L + tB_L + C_L(t)$$

and similarly for M, $L + M$ where these random matrices A, B, C satisfy

$$\mathbf{E}A_L = \mathbf{E}A_M = \mathbf{E}A_{L+M} = 0,$$

$$(6.14) \qquad A_L + A_M = A_{L+M},$$

$$(6.15) \qquad \mathbf{E}(A_L A_M + B_L + B_M - B_{L+M}) = \sigma_*,$$

and $C_L(t), C_M(t), C_{L+M}(t)$ are $o(t)$ in mean square norm. Since $\mathbf{E}\Phi_L(0,t) = e^{b_L t}$ it follows that $\mathbf{E}B_L = b_L$, and (6.15) can be reduced to

$$(6.16) \qquad \mathbf{E}A_L A_M = \sigma_*.$$

It follows from Theorem IX that

$$(6.17) \qquad A_L = \frac{1}{\sqrt{t}} \sum_{k=1}^{p} \sigma_L^{(k)} W_k(t).$$

Thus (6.14), (6.16) follow at once. $\square$

We remark here that (6.9) and (6.12) apply strictly to Itô integration. If we consider other integration schemes these results must be modified accordingly. For example if we use Stratonovich integration then all of the second order terms disappear, and (6.9), (6.12) become, respectively,

$$(6.18) \qquad \prod_{i=1}^{n} \left\{ \prod_{k=1}^{p} \exp\left(\sigma_L^{(k)}\left[W_k\left(\frac{i}{n}t\right) - W_k\left(\frac{i-1}{n}t\right)\right]\right) \right\} e^{b_L t/n} \Rightarrow \Phi_L(0,t),$$

$$(6.19) \qquad \prod_{i=1}^{n} \Phi_L\left(\frac{i-1}{n}t, \frac{i}{n}t\right) \Phi_M\left(\frac{i-1}{n}t, \frac{i}{n}t\right) \Rightarrow \Phi_{L+M}(0,t).$$

We note further that in the case of Stratonovich integration all of the classical (deterministic) analysis is valid and, in particular, the Lie-Trotter product formula is directly applicable to systems like $d\Phi_L = L\Phi_L \, dt$. In case of other integration schemes, such as those considered in McShane [22], it is readily apparent that results analogous to (6.18), (6.19) are valid. (Most of these stochastic integrals correspond to Itô integrals with appropriately modified integrands. See Berger and Mizel [4].)

There is an important point to be made here. The convergence (in distribution) of Riemann type approximating sums for stochastic integrals, as well as Cauchy-Peano type approximants for solutions of stochastic differential equations, can easily be demonstrated through Chernoff's product formula. We begin with an example. Let us prove that in distribution $(dW(t))^2 = dt$ or, more precisely, that $\int_0^t (dW(\tau))^2 = t$. Let $\mathcal{X} = C(\mathbf{R}^*)$ and take

$$(6.20) \qquad T(t)f(x) = \mathbf{E}f(x + W^2(t)).$$

Corresponding to the equispaced approximating sums

$$\sum_{i=1}^n \left[W\left(\frac{i}{n}t\right) - W\left(\frac{i-1}{n}t\right) \right]^2$$

for $\int_0^t (dW(\tau))^2$ we have

$$(6.21) \qquad T^n\left(\frac{t}{n}\right)f(x) = \mathbf{E}f\left(x + \sum_{i=1}^n \left[W\left(\frac{i}{n}t\right) - W\left(\frac{i-1}{n}t\right) \right]^2\right).$$

Take $L = d/dx$, $D(L) = \{f \in \mathcal{X} : f' \text{ exists, belongs to } \mathcal{X} \text{ and tends to zero as } |x| \to \infty\}$. Clearly $T'(0) = L$ in the strong sense, and since $S_L(t)$ is the shift operator, it follows from (2.2) that

$$(6.22) \qquad \lim_{n\to\infty} T^n\left(\frac{t}{n}\right)f(x) = f(x + t).$$

This is one of those rare examples where we can explicitly calculate $T^n(t/n)$. Indeed since

$$\sum_{i=1}^n \left[W\left(\frac{i}{n}t\right) - W\left(\frac{i-1}{n}t\right) \right]^2$$

is distributed according to $(t/n)\chi_n^2$ we have

$$(6.23) \qquad T^n\left(\frac{t}{n}\right)f(x) = \frac{1}{\Gamma(n/2)} \int_0^\infty f\left(x + \frac{2t}{n}s\right) s^{n/2-1}e^{-s}\,ds.$$

Thus (6.22) can be checked directly. Generally one proves mean square convergence of the approximants, and in some cses (e.g., when the approximants come from successive partition refinements) one even proves a.s. convergence (see Doob [11, p. 395]).

Now we modify (6.20) to

$$(6.24) \qquad T(t)f(x) = \mathbf{E}f(x + \sigma(x)W(t)).$$

Define $\xi_x^{(n)}(\tau)$ by

$$(6.25) \qquad \xi_x^{(n)}(\tau) = \begin{cases} x, & \tau = 0, \\ \xi_x^{(n)}\left(\frac{i-1}{n}t\right) + \sigma\left(\xi_x^{(n)}\left(\frac{i-1}{n}t\right)\right)\left[W\left(\frac{i}{n}t\right) - W\left(\frac{i-1}{n}t\right)\right], \\ \qquad\qquad\qquad\qquad\qquad\qquad \tau \in \left(\frac{i-1}{n}t, \frac{i}{n}t\right]. \end{cases}$$

Then

$$(6.26) \qquad T^n\left(\frac{t}{n}\right)f(x) = \mathbf{E}j\left(\xi_x^{(n)}(t)\right).$$

The $\xi_x^{(n)}(\tau)$ are successive Cauchy-Peano type approximants to the diffusion $\xi_x(\tau)$ satisfying

$$(6.27) \qquad \xi_x(t) = x + \int_0^t \sigma(\xi_x(\tau))\,dW(\tau).$$

Take $L = \frac{1}{2}\sigma^2(x)\,d^2/dx^2$ and then from (2.2) follows

$$(6.28) \qquad \xi_x^{(n)}(t) \overset{D}{\to} \xi_x(t).$$

This type of analysis demonstrates existence (in distribution) of solutions for stochastic differential equations. We note the nonanticipating (i.e., "belated", in McShane's jargon) nature of the approximants in (6.25). The result (6.28) should be contrasted with the Wong-Zakai result concerning approximants stemming from the Brownian bridge

$$(6.29) \quad W^{(n)}(\tau) = W\left(\frac{i-1}{n}t\right) + \left(1 + \frac{n\tau}{t} - i\right)\left[W\left(\frac{i}{n}t\right) - W\left(\frac{i-1}{n}t\right)\right],$$
$$\tau \in ((i-1)/nt, it/n],$$

or a smooth approximation to the Brownian motion (see Wong and Zakai [29]).

We focus now on the implications of Theorem VIII on approximating solutions of the Cauchy problem $du/dt = Lu$ in $\mathbf{R}^m$, where the full symbol of L is given by

$$(6.30) \qquad P(x,y) = \sum_{n=1}^p \left(\langle \sigma^{(n)}x + v^{(n)}, y\rangle\right)^2 + \langle bx + w, y\rangle.$$

Here $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}, b$ are $m \times m$ matrices, and $v^{(1)}, v^{(2)}, \ldots, v^{(p)}, w$ are vectors in $\mathbf{R}_m$. Assume first that $v^{(n)}, w$ are all zero. Then for $f \in C(\mathbf{R}_m^*)$

$$(6.31) \qquad S_L(t)f(x) = \mathbf{E}f(\Phi(t)x),$$

where Φ satisfies (5.1), (5.2). In case $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}, b$ all commute we can use (5.4) to write

$$(6.32) \qquad \mathbf{E}f(\Phi(t)x) = \int_{\mathbf{R}_p} f\left(e^{\sum_{n=1}^p \sigma^{(n)}z_n + \sigma^* t}x\right) t^{-p/2} \prod_{i=1}^p \psi\left(t^{-1/2}z_n\right) dz,$$

where $\psi(r) = e^{-r^2/2}/\sqrt{2\pi}$ and $\sigma^* = b - \frac{1}{2}\sum_{n=1}^p [\sigma^{(n)}]^2$. Suppose now that these matrices do not commute, but that $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ form a spanning set. It follows from Theorem VIII that

$$(6.33) \qquad \prod_{i=1}^k \Psi_{i,k}(t) \Rightarrow_{\text{unif}} \Phi(t),$$

where

$$(6.34) \qquad \Psi_{i,k}(t) = e^{\sum_{n=1}^p \sigma^{(n)}W_{n,i}(t/k) + \sigma^* t/k}.$$

Here $\{W_{n,i}(t): n = 1, 2, \ldots, p;\ i = 1, 2, \ldots\}$ are independent Brownian motions. Hence

$$(6.35) \qquad \begin{aligned} &\int_{\mathbf{R}_k \otimes \mathbf{R}_p} f\left(\prod_{i=1}^k \exp\left(\sum_{n=1}^p \sigma^{(n)}\frac{z_{n,i}}{\sqrt{k}} + \sigma^*\frac{t}{k}\right)x\right) t^{-kp/2} \\ &\times \prod_{i=1}^k \prod_{n=1}^p \psi(t^{-1/2}z_{n,i})\,dz \to \mathbf{E}f(\Phi(t)x) \quad \text{as } k \to \infty, \end{aligned}$$

uniformly for $x \in \mathbf{R}_m$. On the other hand, if L is degenerate, so that $\sigma^{(1)}, \sigma^{(2)},$ $\ldots, \sigma^{(p)}$ do not form a spanning set, then the most we can say is that the convergence in (6.35) is uniform for x in compact sets.

If the vectors $v^{(n)}, w$ are not all zero, we add a dummy variable x_{m+1}. Extend $\sigma^{(n)}$ to an $m \times (m+1)$ matrix by inserting $v^{(n)}$ for the $(m+1)$st column, and then extend it to an $(m+1) \times (m+1)$ matrix by inserting all zeros for the $(m+1)$st row. Similarly, extend b to an $(m+1) \times (m+1)$ matrix using w. Call these extended matrices $\tilde{\sigma}^{(n)}$ and $\tilde{b}$, respectively. For $x \in \mathbf{R}_m$ denote $\tilde{x} = (x_1, x_2, \ldots, x_m, 1)$, and for $f \in C(\mathbf{R}_m^*)$ define $\tilde{f}$ by

$$\tilde{f}(x_1, x_2, \ldots, x_{m+1}) = f(x_1, x_2, \ldots, x_m).$$

Again, if $\tilde{\sigma}^{(1)}, \tilde{\sigma}^{(2)}, \ldots, \tilde{\sigma}^{(p)}, \tilde{b}$ all commute then

$$
\begin{aligned}
S_L(t)f(x) &= \mathbf{E}\tilde{f}(\tilde{\Phi}(t)\tilde{x}) \\
&= \int_{\mathbf{R}_p} \tilde{f}\left(e^{\sum_{n=1}^p \tilde{\sigma}^{(n)} z_n + \tilde{\sigma}^* t}\tilde{x}\right) t^{-p/2} \prod_{n=1}^p \psi(t^{-1/2}z_n)\, dz,
\end{aligned}
$$
(6.36)

where $\tilde{\sigma}^* = \tilde{b} - \frac{1}{2}\sum_{n=1}^p [\tilde{\sigma}^{(n)}]^2$ and $\tilde{\Phi}(t)$ satisfies

$$d\tilde{\Phi}(\tau, t) = \sum_{n=1}^p \tilde{\sigma}^{(n)}\tilde{\Phi}(\tau, t)\, dW_n(t) + \tilde{b}\tilde{\Phi}(\tau, t)\, dt,$$
(6.37)

$$\tilde{\Phi}(\tau, \tau) = I.$$
(6.38)

Otherwise, we have

$$
\begin{aligned}
\int_{\mathbf{R}_k \otimes \mathbf{R}_p} &\tilde{f}\left(\prod_{i=1}^k \exp\left(\sum_{n=1}^p \tilde{\sigma}^{(n)}\frac{z_{n,i}}{\sqrt{k}} + \tilde{\sigma}^*\frac{t}{k}\right)\tilde{x}\right) t^{-kp/2} \\
&\times \prod_{i=1}^k \prod_{n=1}^p \psi(t^{-1/2}z_{n,i})\, dz \to \mathbf{E}\tilde{f}(\tilde{\Phi}(t)\tilde{x}) \quad \text{as } k \to \infty,
\end{aligned}
$$
(6.39)

with uniform convergence for $x \in \mathbf{R}_m$, or else uniform convergence for x in compact sets, depending on whether or not L is strongly coercive, respectively.

7. Stochastic semigroups. A stochastic semigroup on $\mathbf{R}_m$ is a family $\{\Psi(\tau, t) = \Psi(\tau, t, \omega): 0 \le \tau < t\}$ of random square-integrable $m \times m$ matrices, satisfying
   (i) Ψ has independent increments and

$$\Psi(t_2, t_3)\Psi(t_1, t_2) = \Psi(t_1, t_3) \quad \text{a.s., } 0 \le t_1 < t_2 < t_3;$$
(7.1)

   (ii) Ψ has stationary increments;
   (iii) for any $\delta > 0$,

$$\mathbf{P}\{|\Psi(0, t) - I| > \delta\} = o(t).$$
(7.2)

Independent increments means that $\Psi(t_{i-1}, t_i)$ are mutually independent, $i = 1, 2, \ldots, p$, whenever $0 = t_0 < t_1 < t_2 < \cdots < t_p$. Stationary increments means that the distribution of $\Psi(t, t+h)$ depends only on h.

Notice that if $\boldsymbol{\Psi}$ is deterministic then $\boldsymbol{\Psi}(\tau, t) = S(t - \tau)$, where $\{S(t) : t \geq 0\}$ is a (classical) continuous semigroup. However, in general when $\boldsymbol{\Psi}$ is random, the one-parameter family $S(t) = \boldsymbol{\Psi}(0, t)$ does not satisfy the semigroup property.

Although in spirit stochastic semigroups have been around for a long time, as in Pinsky [**25**], the first to consider them through axioms similar to those above was Butsan [**7**]. (Note Butsan's references to his earlier works.) In [**7**] Butsan describes a method for obtaining "generators". Formally his results are of the type $d\boldsymbol{\Psi}(\tau, t) = dX(\tau, t)\boldsymbol{\Psi}(\tau, t)$ where

$$X(\tau, t) = \lim_{n \to \infty} \sum_{i=1}^{n} \left[\boldsymbol{\Psi}\left(\tau + (i-1)\frac{t - \tau}{n}, \tau + i\frac{t - \tau}{n} \right) - I \right]$$

(cf. [**10**, p. 190]). A general exploratory discussion of stochastic semigroups without the stationary independent increment assumptions can be found in Grenander [**17**].

An example of a stochastic semigroup is the fundamental matrix $\boldsymbol{\Phi}(\tau, t)$ of a linear stochastic system (5.1), (5.2). Our first results state that under an additional assumption, as far as distribution is concerned this is the only example.

THEOREM XI. *Let $\{\boldsymbol{\Psi}(\tau, t) : 0 \leq \tau < t\}$ be a stochastic semigroup on $\mathbf{R}_m$. Then $\boldsymbol{\Psi}(\tau, t)$ is a.s. nonsingular (the null set can be taken independent of τ, t) and induces a C_0 semigroup on $C(\mathbf{R}_m^*)$.*

PROOF. Let $d(\tau, t) = \det \boldsymbol{\Psi}(\tau, t)$. Then from (7.1)

(7.3) $d(t_2, t_3)\, d(t_1, t_2) = d(t_1, t_3)$ a.s.

Furthermore, $d(t_1, t_2)$ has the same distribution as $d(0, t_2 - t_1)$. Fix t and let $\alpha = \mathbf{P}(d(0, t) = 0)$. Since

$$d(0, t) = \prod_{i=1}^{n} d\left(\frac{i-1}{n}t, \frac{i}{n}t \right)   \text{a.s.}$$

and since $\mathbf{P}(d((i-1)t/n, it/n) = 0)$ is independent of i, we must have

$$\mathbf{P}(d(0, t/n) = 0) \geq \alpha/n.$$

This in turn implies that $\mathbf{P}(|\boldsymbol{\Psi}(0, t/n) - I| \geq 1) \geq \alpha/n$. This contradicts (7.2) unless $\alpha = 0$.

The claim that the null set can be taken independent of τ, t follows from (7.1). For if $d(\tau, t) = 0$ then $d(\tau', t') = 0$ for some rational numbers τ', t'.

Now define a family $\{T(t) : t \geq 0\}$ of bounded linear operators on $\mathcal{X} = C(\mathbf{R}_m^*)$ by

(7.4) $T(t)f(x) = \mathbf{E}f(\boldsymbol{\Psi}(0, t)x).$

Then

(7.5)
$$T(\tau)T(t)f(x) = \mathbf{E}\mathbf{E}[f(\boldsymbol{\Psi}_1(0, t)\boldsymbol{\Psi}_2(0, \tau)x)|\boldsymbol{\Psi}_2(0, \tau)]$$
$$= \mathbf{E}f(\boldsymbol{\Psi}(\tau, \tau + t)\boldsymbol{\Psi}(0, \tau)x) = T(t + \tau)f(x),$$

where $\boldsymbol{\Psi}_i(0, t)$ indicates an ith independent copy of $\boldsymbol{\Psi}(0, t)$; $i = 1, 2$. Given $f \in \mathcal{X}$ and $\varepsilon > 0$ choose $\delta > 0$ such that $V_\delta(f) \leq \varepsilon$. Then for any $R > 0$

$$\sup_{x \in B_R} |\mathbf{E}[f(\boldsymbol{\Psi}(0, t)x) - f(x)]| \leq \varepsilon + 2\|f\|\mathbf{P}\left(|\boldsymbol{\Psi}(0, t) - I| > \frac{\delta}{R} \right).$$

Similarly, given $\varepsilon > 0$ we can choose $R > 0$ such that

$$\sup_{|x| \geq R/2, |y| \geq R/2} |f(x) - f(y)| \leq \varepsilon.$$

Then

$$\sup_{x \in \mathbf{R}_m \setminus B_R} |\mathbf{E}[f(\Psi(0, t)x) - f(x)]| \leq \varepsilon + 2\|f\|\mathbf{P}\left(|\Psi(0, t) - I| > \tfrac{1}{2}\right).$$

Hence we conclude that $\{T(t): t \geq 0\}$ is a C_0 semigroup on $\mathcal{X}$. □

In (7.4) we need the fact that $\Psi(0, t)$ is nonsingular a.s. Otherwise, it is not clear that $T(t)f$ need belong to $\mathcal{X}$. In (5.9) we avoided any such problem by integrating over the set $\{|E(t)| \leq \frac{1}{2}\}$, which guarantees $I + E(t)$ to be nonsingular. The condition (7.2) guarantees, as in (5.19), that

$$\lim_{n \to \infty} \mathbf{P}\left(\max_{1 \leq i \leq n} \left|\Psi_i\left(0, \frac{t}{n}\right) - I\right| \leq \frac{1}{2}\right) = 1,$$

where the subscript Ψ_i indicates on ith independent copy of Ψ. However, it takes another condition, e.g. (7.1), to be able to conclude that $\Psi(0, t)$ itself is nonsingular a.s.

THEOREM XII. *Let $\{\Psi(\tau, t): 0 \leq \tau < t\}$ be a stochastic semigroup on $\mathbf{R}_m$. Make the additional assumption*

(A) $$\Psi(0, t) = I + A\sqrt{t} + Bt + C(t),$$

where $A, B, C(t)$ are square-integrable, $\mathbf{E}A = 0$ and $\mathbf{E}|C(t)|^2 = o(t^2)$. Then for some natural number p and $m \times m$ matrices $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}, b$ the matrix $\Psi(\tau, t)$ has the same distribution as the fundamental matrix $\Phi(\tau, t)$ of the operator

$$L\Phi = \sum_{n=1}^{p} \sigma^{(n)}\Phi z_n + b\Phi.$$

Here $(z_1(t), z_2(t), \ldots, z_p(t))$ is p-dimensional white noise.

PROOF. $\Psi(\tau, t)$ has the same distribution as $\prod_{i=1}^{n} \Psi_i(0, (t - \tau)/n)$. Let $p, \sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(p)}$ satisfy (3.1) and let $b = \mathbf{E}B$. Then according to Theorem VI, $\Psi(\tau, t)$ must have the same distribution as $\Phi_{\tau,t}(1)$, where

(7.6) $$d\Phi_{\tau,t}(\tau', t') = \sqrt{t - \tau} \sum_{n=1}^{p} \sigma^{(n)}\Phi_{\tau,t}(\tau', t')\, dW_n(t') + (t - \tau)b\Phi_{\tau,t}(\tau', t')\, dt',$$

(7.7) $$\Phi_{\tau,t}(\tau', \tau') = I.$$

By scaling it follows that $\Phi_{\tau,t}(\tau', t')$ has the same distribution as $\Phi((t - \tau)(t' - \tau'))$, where Φ satisfies (5.1), (5.2). □

THEOREM XIII. *In the scalar (1×1) case, where $\{\Psi(\tau, t): 0 \leq \tau < t\}$ is a stochastic semigroup on $\mathbf{R}$, assumption (A) is automatically satisfied. In fact $\Psi(\tau, t)$ is necessarily of the form $\exp(aW(t) + bt)$, where $W(t)$ is a Brownian motion.*

PROOF. First we show that $\Psi(\tau, t) > 0$ a.s. As above, let $\alpha = \mathbf{P}(\Psi(\tau, t) < 0)$. Then $\alpha = \mathbf{P}(\Psi(0, t') < 0)$ for $t' = t - \tau$. Since $\Psi(0, t')$ has the same distribution as $\Psi_1(0, t'/2)\Psi_2(0, t'/2)$ it follows that

$$\mathbf{P}(\Psi(0, t'/2) < 0) \geq \alpha/2$$

and, inductively, that
$$\mathbf{P}(\Psi(0, t'/2^n) < 0) \geq \alpha/2^n.$$
From this follows
$$\mathbf{P}(|\psi(0, t'/2^n) - 1| > 1) \geq \alpha/2^n,$$
which contradicts (7.2) unless $\alpha = 0$.

Now let $\nu(t) = \ln \Psi(0, t)$. Then $\nu(t)$ is a stationary independent increment stochastic process. Furthermore for any $\delta > 0$
$$\mathbf{P}(|\nu(t)| > \delta) \leq \mathbf{P}(|\Psi(0, t) - 1| > 1 - e^{-\delta}) = o(t).$$
Hence (cf. Breiman [6, p. 249]) $\nu(t)$ must be of the form $aW(t) + bt$. $\square$

According to Theorem XI the matrices $\Psi(\tau, t)$ of the stochastic semigroup are in the Lie group $G = \mathrm{GL}_m$. The measures p_t on G induced by $\Psi(0, t)$, $p_t(X) = \mathbf{P}(\Psi(0, t) \in X)$, form what Hunt [20] calls a semigroup of positive measures. In fact, if it were known a priori that $\Psi(0, t) \in G$, then it would suffice to replace (7.2) with the weaker condition

$$(7.8) \qquad\qquad \lim_{t \downarrow 0} \mathbf{P}(|\Psi(0, t) - I| > \delta) = 0$$

in order that these measures p_t form such a semigroup. Hunt then has a complete characterization of the generator of the semigroup $T(t)f(x) = \int_G f(Mx)p_t(dM)$; that is, (7.4). His representation of the generator L is

$$(7.9) \qquad
\begin{aligned}
Lf(x) = & \sum_{i=1}^{m} \sum_{j=1}^{m} a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^{m} b_i(x) \frac{\partial f}{\partial x_i} \\
& + \int_G \left[f(Mx) - f(x) - \sum_{i=1}^{m} \phi_i(M) c_i(x) \frac{\partial f}{\partial x_i} \right] \nu(dM),
\end{aligned}$$

where the a_{ij} are quadratic forms, the b_i and c_i are linear forms, the ϕ_i are generalized coordinates and ν is a measure on $G \backslash \{I\}$ with finite second moment.

Let us briefly indicate the parallel between our setting and that of infinitely divisible distributions. Lévy showed that any infinitely divisible probability μ can always be taken as p_1 in a semigroup of measures p_t. Thus, one constructs a stationary independent increment process $x(t)$ such that $x(1)$ has the distribution μ. When we have a stochastic semigroup $\Psi(\tau, t)$ then $\Psi(0, 1)$ is an infinitely divisible element of G. Perhaps a better name in this case, where the group operation is multiplication, would be infinitely factorable. That is, a random matrix M is said to be infinitely factorable if for every n there are i.i.d. matrices $M_i^{(n)}$ such that $\prod_{i=1}^{n} M_i^{(n)}$ has the same distribution as M. A random matrix M which is the limit in distribution of products $\prod_{i=1}^{n} D_{i,n}$ of i.i.d. matrices $D_{1,n}, D_{2,n}, \ldots, D_{n,n}$ should have such an infinitely factorable distribution. In our Central Limit Theorem IV we have identified the limiting distribution as the distibution p_1 of $\Phi(0, 1)$, for the stochastic semigroup Φ given by (5.1), (5.2).

Hunt briefly discusses this question of whether or not an infinitely factorable distribution μ can be taken as p_1 in a semigroup p_t of measures on G, in his introduction (see also §5.3). Our concern is with limits in distribution

$$M = \lim_{n \to \infty} \prod_{i=1}^{n} D_{i,n},$$

where

$$(7.10) \qquad \max_{1 \le i \le n} |D_{i,n} - I| \xrightarrow{D} 0 \quad \text{as } n \to \infty.$$

It is clear from (5.6) that this condition holds, since A_i, B_i, C_i are square-integrable. It is also clear from (7.2) that

$$(7.11) \qquad \max_{1 \le i \le n} \left| \Psi\left(\frac{i-1}{n}t, \frac{i}{n}t\right) - I \right| \xrightarrow{D} 0 \quad \text{as } n \to \infty.$$

Since, by (7.1),

$$\Psi\left(0, \frac{i}{n}t\right) - \Psi\left(0, \frac{i-1}{n}t\right) = \left[\Psi\left(\frac{i-1}{n}t, \frac{i}{n}t\right) - I\right] \Psi\left(0, \frac{i-1}{n}t\right)$$

we see that (7.11) is a continuity-type condition for Ψ. Hunt remarks, at the end of §5, that if $\Psi(0,t)$ has continuous sample paths, then $\nu = 0$ in (7.9) (i.e. the generator has no "Poisson" component). This would then allow us to conclude, as in Theorem XII, that $\Psi(\tau, t)$ has the same distribution as some fundamental matrix $\Phi(\tau, t)$, without the need to make the a priori assumption (A). This assumption would be automatically satisfied, as in Theorem XIII.

BIBLIOGRAPHY

1. R. Bellman, *Limit theorems for non-commutative operations*. I, Duke Math. J. **21** (1954), 491–500.
2. M. A. Berger, *A remark on stochastic fundamental matrices*, Ann. Probab. **11** (1983), 215–216.
3. M. A. Berger and S. Friedland, *The generalized Radon-Hurwitz numbers* (to appear).
4. M. A. Berger and V. J. Mizel, *Theorems of Fubini type for iterated stochastic integrals*, Trans. Amer. Math. Soc. **252** (1979), 249–274.
5. M. A. Berger and A. D. Sloan, *Semigroup approach to the radical differential calculus*, Lecture Notes, Georgia Institute of Technology, 1980.
6. L. Breiman, *Probability*, Addison-Wesley, Reading, Mass., 1968.
7. G. P. Butsan, *On the mixed product of nonhomogeneous stochastic semigroups*, Theory Probab. Appl. **24** (1979), 168–175.
8. P. R. Chernoff, *Note on product formulas for operator semi-groups*, J. Funct. Anal. **2** (1968), 238–242.
9. ____ , *Product formulas, nonlinear semigroups, and addition of unbounded operators*, Mem. Amer. Math. Soc. No. **140** (1974).
10. J. D. Dollard and C. N. Friedman, *Encyclopedia of mathematics and its applications,* Vol. 10: *Product integration* (G. C. Rota, ed.), Addison-Wesley, Reading, Mass., 1979.
11. J. L. Doob, *Stochastic processes*, Wiley, New York, 1967.
12. H. Furstenberg, *Non-commuting random products*, Trans. Amer. Math. Soc. **108** (1963), 377–428.
13. ____ , *A Poisson formula for semi-simple Lie groups*, Ann. of Math. (2) **77** (1963), 335–386.
14. H. Furstenberg and H. Kesten, *Products of random matrices*, Ann. Math. Statist. **31** (1960), 457–469.
15. I. I. Gihman and A. V. Skorohod, *Stochastic differential equations*, Springer-Verlag, New York, 1972 (translated from Russian).
16. J. A. Goldstein, *Semigroup-theoretic proofs of the central limit theorem and other theorems of analysis*, Semigroup Forum **12** (1976), 189–206. (See also Corrigendum, p. 388.)
17. U. Grenander, *Probabilities on algebraic structures*, Wiley, Stockholm, 1963.
18. Y. Guivarc'h, *Quelques propriétés asymptotiques des produits de matrices aléatoires*, Lecture Notes in Math., vol. 774, Springer-Verlag, Berlin and New York, 1980.
19. Y. Guivarc'h, M. Keane and B. Roynette, *Marches aléatoires sur les groupes de Lie*, Lecture Notes in Math., vol. 624, Springer-Verlag, Berlin and New York, 1977.
20. G. A. Hunt, *Semi-groups of measures on Lie groups*, Trans. Amer. Math. Soc. **81** (1956), 264–293.

21. R. Z. Khasminskii, *Necessary and sufficient conditions for the asymptotic stability of linear stochastic systems*, Theor. Probab. Appl. **12** (1967), 144–147.

22. E. J. McShane, *Stochastic calculus and stochastic models*, Academic Press, New York, 1974.

23. A. Pazy, *A Trotter-Lie formula for compactly generated semigroups of nonlinear operators*, J. Funct. Anal. **23** (1976), 353–361.

24. ____, *Semi-groups of linear operators and applications to partial differential equations*, Springer-Verlag, New York, 1983.

25. M. Pinsky, *Multiplicative operator functionals and their asymptotic properties*, Adv. in Probab. Related Topics, Vol. III (P. E. Ney, ed.), Dekker, New York, 1974.

26. ____, *Isolated singularities of degenerate elliptic equations in* $\mathbf{R}^2$, J. Differential Equations **24** (1977), 274–281.

27. V. N. Tutubalin, *On limit theorems for the product of random matrices*, Theor. Probab. Appl. **10** (1964), 15–27.

28. J. C. Watkins, *A central limit problem in random evolutions*, Ann. Probab. (to appear).

29. E. Wong and M. Zakai, *On the convergence of ordinary integrals to stochastic integrals*, Ann. Math. Statist. **36** (1965), 1560–1564.

SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GEORGIA 30332