Some Background Material for 18.176
Daniel W. Stroock

Notation

\[ \langle \varphi, \mu \rangle = \int \varphi \, d\mu \]

\( M_1(\mathbb{R}^N) \) is the space of Borel probability measures on \( \mathbb{R}^N \) endowed with the topology of weak convergence. Given \( \mu \in M_1(\mathbb{R}^N) \), its Fourier transform is 
\[ \hat{\mu}(\xi) \equiv \int_{\mathbb{R}^N} e^{i(\xi, y)} \mu(dy). \]

\[ B(a, r) = \{ x \in \mathbb{R}^N : |x - a| < r \} \]
and 
\[ S^{N-1} = \{ x \in \mathbb{R}^N : |x| = 1 \} \]

\( \| \cdot \|_u \) is the uniform norm.

\( C^m(\mathbb{R}^N; \mathbb{R}) \) is the space of \( m \)-times, continuously differentiable functions, 
\( C^m_b(\mathbb{R}^N; \mathbb{R}) \) is the subspace of \( C^m(\mathbb{R}^N; \mathbb{R}) \) whose elements have bounded derivatives of all orders less than or equal to \( m \), 
\( C^m_c(\mathbb{R}^N; \mathbb{R}) \) is the subspace of \( C^m_b(\mathbb{R}^N; \mathbb{R}) \) whose elements have compact support, 
\( C^{1,2}_b([0, \infty) \times \mathbb{R}^N; \mathbb{R}) \) is the space functions \( (t, x) \in [0, \infty) \times \mathbb{R}^N \mapsto u(t, x) \in \mathbb{R} \) which have one bounded, continuous derivative with respect to \( t \) and two bounded, continuous derivatives with respect to \( x \).

\( \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) is the Schwartz space of infinitely differentiable functions all of whose derivatives are rapidly decreasing (i.e., tend to 0 at infinity faster than \( |x|^{-n} \) for every \( n \geq 0 \)).

\[ \|A\|_{\text{op}} = \sup\{ |Ae| : e \in S^{N-1} \} \]

Linear Functionals that Satisfy the Minimum Principle

Suppose that \( t \in (0, \infty) \mapsto \mu_t \in M_1(\mathbb{R}^N) \) is a map with the properties that

\[ A\varphi \equiv \lim_{t \to 0} \frac{\langle \varphi, \mu_t \rangle - \varphi(0)}{t} \]

exists for all \( \varphi \in D \equiv \mathbb{R} \oplus \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) and

\[ \lim_{R \to \infty} \lim_{t \to 0} \frac{\mu_t(B(0, R)^c)}{t} = 0. \]

Then

(i) \( A \) satisfies the minimum principle:

\[ \varphi(0) \leq \varphi \implies A\varphi \geq 0. \]
(ii) $A$ is quasi-local

\begin{equation}
\lim_{R \to \infty} A \varphi_R = 0 \quad \text{where } \varphi_R(x) = \varphi \left( \frac{x}{R} \right)
\end{equation}

if $\varphi$ is constant in a neighborhood of 0.

The first of these is obvious, since

$$\langle \varphi, \mu_t \rangle - \varphi(0) = \langle \varphi - \varphi(0), \mu_t \rangle \geq 0$$

if $\varphi(0) \leq \varphi$. To check the second, choose $\delta > 0$ so that $\varphi(y) = \varphi(0)$ for $|y| < \delta$. Then

$$\left| \langle \varphi_R, \mu_t \rangle - \varphi(0) \right| \leq \int_{B(0, R\delta)} \left| \varphi \left( \frac{y}{R} \right) - \varphi(0) \right| \mu_t(dy) \leq 2\|\varphi\|_{\mu_t(B(0, R\delta))},$$

which, by (1), means that $\lim_{R \to \infty} A \varphi_R = 0$.

The goal here is to show that any linear functional that satisfies these two properties has a very special structure. To begin with, notice that, by applying the minimum principle to both 1 and $-1$, one knows that $A1 = 0$. Before going further, I have to introduce the following partition of unity for $\mathbb{R}^n \setminus \{0\}$. Choose $\psi \in C^\infty(\mathbb{R}^N; [0, 1])$ so that $\psi$ has compact support in $B(0, 2) \setminus B\left(0, \frac{1}{4}\right)$ and $\psi(y) = 1$ when $\frac{1}{2} \leq |y| \leq 1$, and set $\psi_m(y) = \psi(2^m y)$ for $m \in \mathbb{Z}$. Then, if $y \in \mathbb{R}^N$ and $2^{-m-1} \leq |y| \leq 2^{-m}$, $\psi_m(y) = 1$ and $\psi_n(y) = 0$ unless $-m - 2 \leq n \leq -m + 1$. Hence, if $\Psi(y) = \sum_{m \in \mathbb{Z}} \psi_m(y)$ for $y \in \mathbb{R}^N \setminus \{0\}$, then $\Psi$ is a smooth function with values in $[1, 4]$; and therefore, for each $m \in \mathbb{Z}$, the function $\chi_m$ given by $\chi_m(0) = 0$ and $\chi_m(y) = \frac{\psi_m(y)}{\Psi(y)}$ for $y \in \mathbb{R}^N \setminus \{0\}$ is a smooth, $[0, 1]$-valued function that vanishes off of $B(0, 2^{-m+1}) \setminus B(0, 2^{-m-2})$. In addition, for each $y \in \mathbb{R}^N \setminus \{0\}$, $\sum_{m \in \mathbb{Z}} \chi_m(y) = 1$ and $\chi_m(y) = 0$ unless $2^{-m-2} \leq |y| \leq 2^{-m+1}$.

Finally, given $n \in \mathbb{Z}^+$ and $\varphi \in C^n(\mathbb{R}^N; \mathbb{C})$, define $\nabla^n \varphi(x)$ to be the multilinear map on $(\mathbb{R}^N)^n$ into $\mathbb{C}$ by

$$[\nabla^n \varphi(x)](\xi_1, \ldots, \xi_n) = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} \varphi \left( x + \sum_{m=1}^n t_m \xi_m \right) \bigg|_{t_1 = \cdots = t_n = 0}.$$

Obviously, $\nabla \varphi$ and $\nabla^2 \varphi$ can be identified as the gradient of $\varphi$ and Hessian of $\varphi$.

A measure Borel $M$ on $\mathbb{R}^N$ is said to be a **Lévy measure** if

$$M(\{0\}) = 0 \quad \text{and} \quad \int_{B(0, 1)} |y|^2 M(dy) + M(B(0, 1) \setminus \{0\}) < \infty.$$
Theorem 2. Let \( D \) be the space of functions described above. If \( A : D \rightarrow \mathbb{R} \) is a linear functional on \( D \) that satisfies (i) and (ii), then there is a unique Lévy measure \( M \) such that \( A\varphi = \int_{\mathbb{R}^N} \varphi(y) M(dy) \) whenever \( \varphi \) is an element of \( \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) for which \( \varphi(0) = 0, \nabla \varphi(0) = 0, \) and \( \nabla^2 \varphi(0) = 0. \) Next, given \( \eta \in C^\infty_c(\mathbb{R}^N; [0,1]) \) satisfying \( \eta = 1 \) in a neighborhood of \( 0, \) set \( \eta_{\xi}(y) = \eta(y)(\xi,y)_{\mathbb{R}^N} \) for \( \xi \in \mathbb{R}^N, \) and define \( m^\eta \in \mathbb{R}^N \) and \( C \in \text{Hom}(\mathbb{R}^N; \mathbb{R}^N) \) by

\[
(3) \quad (\xi, m^\eta)_{\mathbb{R}^N} = A\eta \quad \text{and} \quad (\xi, C\xi')_{\mathbb{R}^N} = A(\eta\eta'_{\xi}) - \int_{\mathbb{R}^N} (\eta\eta'_{\xi})_{\mathbb{R}^N} M(dy).
\]

Then \( C \) is symmetric, non-negative definite, and independent of the choice of \( \eta. \) Finally, for any \( \varphi \in D, \)

\[
A\varphi = \frac{1}{2} \text{Trace}(C\nabla^2 \varphi(0)) + (m^\eta, \nabla \varphi(0))_{\mathbb{R}^N} \tag{4}
\]

\[
+ \int_{\mathbb{R}^N} \left( \varphi(y) - \varphi(0) - \eta(y)(y, \nabla \varphi(0))_{\mathbb{R}^N} \right) M(dy).
\]

Conversely, if the action of \( A \) is given by (4), then \( A \) satisfies the minimum principle and is quasi-local.

Proof: The concluding converse assertion is easy. Indeed, if \( A \) is given by (4), then it is obvious that \( A1 = 0. \) Next suppose that \( \varphi \in \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) and that \( \varphi(0) \leq \varphi. \) Then \( \nabla \varphi(0) = 0 \) and \( \nabla^2 \varphi(0) \) is non-negative definite. Hence \( (m, \nabla \varphi(0))_{\mathbb{R}^N} = 0, \text{Trace}(C\nabla^2 \varphi(0)) \geq 0, \) and

\[
\varphi(y) - \varphi(0) - 1_{B(a,1)}(y)(y, \nabla \varphi(0))_{\mathbb{R}^N} \geq 0.
\]

To see that \( A \) is quasi local, suppose that \( \varphi \in \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) is constant in a neighborhood of \( 0, \) and choose \( \delta > 0 \) so that \( \varphi = \varphi(0) \) on \( B(0, \delta). \) Obviously, \( \nabla \varphi(0) = 0 \) and \( \nabla^2 \varphi(0) = 0. \) In addition,

\[
\left| \int_{\mathbb{R}^N} \left( \varphi_R(y) - \varphi_R(0) - 1_{B(a,1)}(y, \nabla \varphi_R(0))_{\mathbb{R}^N} \right) M(dy) \right| \\
\leq \int_{B(0, R\delta)\mathbb{R}} |\varphi_R(y)| M(dy) \leq ||\varphi||_a M(B(0, R\delta)\mathbb{R}) \rightarrow 0
\]

as \( R \rightarrow \infty. \)

Referring to the partition of unity described above, define \( \Lambda_m \varphi = A(\chi_m \varphi) \) for \( \varphi \in C^\infty_c(B(0,2^{-m+1}) \setminus B(0,2^{-m-2}); \mathbb{R}), \) where

\[
\chi_m \varphi(y) = \begin{cases} \chi_m(y)\varphi(y) & \text{if } 2^{-m-2} \leq |y| \leq 2^{-m+1} \\ 0 & \text{otherwise.} \end{cases}
\]

Clearly \( \Lambda_m \) is linear. In addition, if \( \varphi \geq 0, \) then \( \chi_m \varphi \geq 0 = \chi_m \varphi(0), \) and so, by (i), \( \Lambda_m \varphi \geq 0. \) Similarly, for any \( \varphi \in C^\infty_c(B(0,2^{-m+1}) \setminus B(0,2^{-m-2}); \mathbb{R}), \)
Clearly, \( \| \varphi \|_{u} \chi_{m} \pm \chi_{m} \varphi \geq 0 = (\| \varphi \|_{u} \chi_{m} \pm \chi_{m} \varphi)(0) \), and therefore \( |\Lambda_{m} \varphi| \leq K_{m} \| \varphi \|_{u} \), where \( K_{m} = A \chi_{m} \). Hence, \( \Lambda_{m} \) admits a unique extension as a continuous linear functional on \( C(\overline{B(0, 2^{-m+1}) \setminus B(0, 2^{-m-2})}; \mathbb{R}) \) that is non-negativity preserving and has norm \( K_{m} \); and so, by the Riesz Representation Theorem, we now know that there is a unique non-negative Borel measure \( M_{m} \) on \( \mathbb{R}^{N} \) such that \( M_{m} \) is supported on \( B(0, 2^{-m+1}) \setminus B(0, 2^{-m-2}) \), \( K_{m} = M_{m}(\mathbb{R}^{N}) \), and \( A(\chi_{m} \varphi) = \int_{\mathbb{R}^{N}} \varphi(y) M_{m}(dy) \) for all \( \varphi \in \mathscr{S}(\mathbb{R}^{N}; \mathbb{R}) \).

Now define the non-negative Borel measure \( M \) on \( \mathbb{R}^{N} \) by \( M = \sum_{m \in \mathbb{Z}} M_{m} \). Clearly, \( M(\{0\}) = 0 \). In addition, if \( \varphi \in C_{c}^{\infty}(\mathbb{R}^{N} \setminus \{0\}; \mathbb{R}) \), then there is an \( n \in \mathbb{Z}^{+} \) such that \( \chi_{m} \varphi \equiv 0 \) unless \( |m| \leq n \). Thus,

\[
A \varphi = \sum_{m=-n}^{n} A(\chi_{m} \varphi) = \sum_{m=-n}^{n} \int_{\mathbb{R}^{N}} \varphi(y) M_{m}(dy)
= \int_{\mathbb{R}^{N}} \left( \sum_{m=-n}^{n} \chi_{m}(y) \varphi(y) \right) M(dy) = \int_{\mathbb{R}^{N}} \varphi(y) M(dy),
\]

and therefore

\[
(5) \quad A \varphi = \int_{\mathbb{R}^{N}} \varphi(y) M(dy)
\]

for \( \varphi \in C_{c}^{\infty}(\mathbb{R}^{N} \setminus \{0\}; \mathbb{R}) \).

Before taking the next step, observe that, as an application of (i), if \( \varphi_{1}, \varphi_{2} \in D \), then

\[
(*) \quad \varphi_{1} \preceq \varphi_{2} \text{ and } \varphi_{1}(0) = \varphi_{2}(0) \implies A \varphi_{1} \preceq A \varphi_{2}.
\]

Indeed, by linearity, this reduces to the observation that, by (i), if \( \varphi \in D \) is non-negative and \( \varphi(0) = 0 \), then \( A \varphi \geq 0 \).

With these preparations, I can show that, for any \( \varphi \in D \),

\[
(**) \quad \varphi \geq 0 = \varphi(0) \implies \int_{\mathbb{R}^{N}} \varphi(y) M(dy) \leq A \varphi.
\]

To check this, apply (*) to \( \varphi_{n} = \sum_{m=-n}^{n} \chi_{m} \varphi \) and \( \varphi \), and use (5) together with the Monotone Convergence Theorem to conclude that

\[
\int_{\mathbb{R}^{N}} \varphi(y) M(dy) = \lim_{n \to \infty} \int_{\mathbb{R}^{N}} \varphi_{n}(y) M(dy) = \lim_{n \to \infty} A \varphi_{n} \leq A \varphi.
\]

Now let \( \eta \) be as in the statement of the lemma, and set \( \eta_{R}(y) = \eta(R^{-1}y) \) for \( R > 0 \). By (**) with \( \varphi(y) = |y|^{2} \eta(y) \) we know that

\[
\int_{\mathbb{R}^{N}} |y|^{2} \eta(y) M(dy) \leq A \varphi < \infty.
\]
At the same time, by (5) and (*),

\[ \int_{\mathbb{R}^N} (1 - \eta(y)) \eta_R(y) M(dy) \leq A(1 - \eta) \]

for all \( R > 0 \), and therefore, by Fatou’s Lemma,

\[ \int_{\mathbb{R}^N} (1 - \eta(y)) M(dy) \leq A(1 - \eta) < \infty. \]

Hence, I have proved that \( M \) is a Lévy measure.

I am now in a position to show that (5) continues to hold for any \( \varphi \in \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) that vanishes along with its first and second order derivatives at 0. To this end, first suppose that \( \varphi \) vanishes in a neighborhood of 0. Then, for each \( R > 0 \), (5) applies to \( \eta_R \varphi \), and so

\[ \int_{\mathbb{R}^N} \eta_R(y) \varphi(y) M(dy) = A(\eta_R \varphi) = A \varphi + A(1 - \eta_R) \varphi. \]

By (*) applied to \( \pm(1 - \eta_R) \varphi \) and \( (1 - \eta_R) \| \varphi \|_u \),

\[ A(1 - \eta_R) \varphi \leq \| \varphi \|_u A(1 - \eta_R) \longrightarrow 0 \quad \text{as} \quad R \to \infty, \]

where I used (ii) to get the limit assertion. Thus,

\[ A \varphi = \lim_{R \to \infty} \int_{\mathbb{R}^N} \eta_R(y) \varphi(y) M(dy) = \int_{\mathbb{R}^N} \varphi(y) M(dy), \]

because, since \( M \) is finite on the support of \( \varphi \) and therefore \( \varphi \) is \( M \)-integrable, Lebesgue’s Dominated Convergence Theorem applies. I still have to replace the assumption that \( \varphi \) vanishes in a neighborhood of 0 by the assumption that it vanishes to second order there. For this purpose, first note that, because \( M \) is a Lévy measure, \( \varphi \) is certainly \( M \)-integrable, and therefore

\[ \int_{\mathbb{R}^N} \varphi(y) M(dy) = \lim_{R \to \infty} A((1 - \eta_R) \varphi) = A \varphi - \lim_{R \to \infty} A(\eta_R \varphi). \]

By our assumptions about \( \varphi \) at 0, we can find a \( C < \infty \) such that \( |\eta_R \varphi(y)| \leq CR|y|^2 \) \( \eta(y) \) for all \( R \in (0,1) \). Hence, by (*), there is a \( C' < \infty \) such that \( A(\eta_R \varphi) \leq C'R \) for small \( R > 0 \), and therefore \( A(\eta_R \varphi) \longrightarrow 0 \) as \( R \to 0 \).

To complete the proof from here, let \( \varphi \in \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) be given, and set

\[ \tilde{\varphi}(x) = \varphi(x) - \varphi(0) - \eta(x)(x, \nabla \varphi(0))_{\mathbb{R}^N} - \frac{1}{2} \eta(x)^2 (x, \nabla^2 \varphi(0) x)_{\mathbb{R}^N}. \]

Then, by the preceding, (5) holds for \( \tilde{\varphi} \) and, after one re-arranges terms, says that (4) holds. Thus, the properties of \( \mathbf{C} \) are all that remain to be proved. That \( \mathbf{C} \) is symmetric requires no comment. In addition, from (*), it is clearly non-negative definite. Finally, to see that it is independent of the \( \eta \) chosen, let \( \eta' \) be a second choice, note that \( \eta'_{\xi} = \eta_{\xi} \) in a neighborhood of 0, and apply (5). \( \square \)
Remark 6. A careful examination of the proof of Theorem 1 reveals a lot. Specifically, it shows why the operation performed by the linear functional $A$ cannot be of order greater than 2. The point is, that, because of the minimum principle, $A$ acts as a bounded, non-negative linear functional on the difference between $\varphi$ and its second order Taylor polynomial, and, because of quasi-locality, this action can be represented by integration against a non-negative measure. The reason why the second order Taylor polynomial suffices is that second order polynomials are, apart from constants, the lowest order polynomials that can have a definite sign.

Another important observation is that there is enormous freedom in the choice of the function $\eta$. Indeed, one can choose any measurable $\eta : \mathbb{R}^N \to [0, 1]$ with the property the

$$ \sup_{y \in B(0,1)} |y|^{-1}(1 - \eta(y)) + \sup_{y \notin B(0,1)} |y|\eta(y) < \infty. $$

In particular, there is no reason not to take $\eta = 1_{B(0,1)}$, and we will usually do so.

Finally, notice that when $A$ is local in the sense that $A\varphi = 0$ if $\varphi$ is constant in a neighborhood of 0, then the corresponding $M = 0$.

My next goal is to show that for each choice of $m \in \mathbb{R}^N$, non-negative definite, symmetric matrix $C$, and Lévy measure $M$, there is a natural map $t \in [0, \infty) \to \mu_t \in M_1(\mathbb{R}^N)$ such that

$$ A\varphi \equiv (m, \nabla \varphi) + \frac{1}{2} \text{Trace}(C \nabla^2 \varphi) $$

$$ + \int_{\mathbb{R}^N} \left( \varphi(y) - 1 - 1_{B(0,1)}(y) (y, \nabla \varphi(0))_{\mathbb{R}^N} \right) M(dy) $$

$$ = \lim_{t \searrow 0} \frac{\langle \varphi, \mu_t \rangle - \varphi(0)}{t} \quad \text{for } \varphi \in D. $$

Lemma 7. Set

$$ \ell(\xi) = i(m, \xi)_{\mathbb{R}^N} - \frac{1}{2} (\xi, C\xi)_{\mathbb{R}^N} + \int_{\mathbb{R}^N} \left( e^{i(\xi, y)} y^N - 1 - i1_{B(0,1)}(y) (\xi, y)_{\mathbb{R}^N} \right) M(dy). $$

Then $\Re(\ell(\xi)) \leq 0$ and $\lim_{|\xi| \to \infty} |\xi|^{-2} |\ell(\xi)| = 0$. Thus there exists a $K < \infty$ such that $|\ell(\xi)| \leq K(1 + |\xi|^2)$. Finally, for each $\varphi \in \mathcal{S}(\mathbb{R}^N; \mathbb{C})$,

$$ \int_{\mathbb{R}^N} \ell(-\xi) \hat{\varphi}(\xi) d\xi = \lim_{t \searrow 0} t^{-1} \int_{\mathbb{R}^N} \hat{\varphi}(\xi) (e^{t\ell(-\xi)} - 1) d\xi = (2\pi)^N A\varphi. $$

Proof: Begin by noting that

$$ \Re(\ell(\xi)) = -\frac{1}{2} (\xi, C\xi)_{\mathbb{R}^N} + \int_{\mathbb{R}^N} (\cos(\xi, y)_{\mathbb{R}^N} - 1) M(dy) \leq 0. $$
Finally, because $\text{rem}$ shows that Fubini's Theorem applies and says that
\[ \int_{\mathbb{R}^N} \left( e^{i(\xi \cdot y)} - 1 - (2 + |\xi|) M(B(0, r)) \right) M(dy) \leq \frac{|\xi|^2}{2} \int_{B(0, r)} |y|^2 M(dy), \]
and therefore
\[ \lim_{|\xi| \to \infty} |\xi|^{-2} \int_{\mathbb{R}^N} \left( e^{i(\xi \cdot y)} - 1 - i1_{B(0,1)}(\xi, y) \right) M(dy) \leq \frac{1}{2} \int_{B(0, r)} |y|^2 M(dy). \]
Since $\int_{B(0, r)} |y|^2 M(dy) \to 0$ as $r \to 0$, this completes the proof of the initial assertions.

To prove the final assertion, note that, by Taylor’s Theorem,
\[ |e^{t\ell(\xi)} - 1 - t\ell(\xi)| \leq \frac{t^2|\ell(\xi)|^2}{2} \leq \frac{K^2 t^2 (1 + |\xi|^2)^2}{2}. \]
Hence, since $\hat{\phi}$ is rapidly decreasing, Lebesgue’s Dominated Convergence Theorem shows that
\[ \lim_{t \to 0} t^{-1} \int_{\mathbb{R}^N} \hat{\phi}(\xi) (e^{t\ell(-\xi)} - 1) d\xi = \int_{\mathbb{R}^N} \hat{\phi}(\xi) \ell(-\xi) d\xi. \]
By the Fourier inversion formula,
\[ (2\pi)^{-N} \int_{\mathbb{R}^N} \left( -i(\mathbf{m}, \xi)_{\mathbb{R}^N} - \frac{1}{2} (\xi, C\xi)_{\mathbb{R}^N} \right) \hat{\phi}(\xi) d\xi = (\mathbf{m}, \nabla \varphi(0))_{\mathbb{R}^N} + \frac{1}{2} \text{Trace}(C \nabla^2 \varphi(0)). \]
Finally, because
\[ \int_{\mathbb{R}^N \times \mathbb{R}^N} |\hat{\phi}(\xi)||e^{i(\xi \cdot y)} - 1 - i1_{B(0,1)}(\xi, y)_{\mathbb{R}^N}| M(dy) d\xi < \infty, \]
Fubini’s Theorem applies and says that
\begin{align*}
\int_{\mathbb{R}^N} \hat{\phi}(\xi) \left( \int_{\mathbb{R}^N} \left( e^{-i(\xi \cdot y)} - 1 - i1_{B(0,1)}(\xi, y)_{\mathbb{R}^N} \right) M(dy) \right) d\xi \\
= \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \left( e^{-i(\xi \cdot y)} - 1 - i1_{B(0,1)}(\xi, y)_{\mathbb{R}^N} \right) \hat{\phi}(\xi) d\xi \right) M(dy) \\
= (2\pi)^N \int_{\mathbb{R}^N} (\varphi(y) - \varphi(0) - 1_{B(0,1)}(y, \nabla \varphi(0))_{\mathbb{R}^N}) M(dy),
\end{align*}
where I again used the Fourier inversion formula in the passage to the last line.

In view of Lemma 6, I will be done once I show that, for each \( t > 0 \), there exists a \( \mu \in \mathbf{M}_1(\mathbb{R}^N) \) such that \( \hat{\mu}_t(\xi) = e^{it\langle \xi \rangle} \). Indeed, by Parseval’s identity,

\[
\langle \varphi, \mu \rangle - \varphi(0) = (2\pi)^{-N} \int_{\mathbb{R}^N} \hat{\varphi}(\xi)(\hat{\mu}_t - 1)(\xi)d\xi
\]

for \( \varphi \in \mathcal{S}(\mathbb{R}^N; \mathbb{C}) \).

Since \( tl(\xi) \) can be represented as the \( \ell(\xi) \) for \( (t\mathbf{m}, t\mathbf{C}, t\mathbf{M}) \), it suffices to take \( t = 1 \). Furthermore, because \( \hat{\mu} * \hat{\nu}(\xi) = \hat{\mu}(\xi)\hat{\nu}(\xi) \), we can treat the terms in \( \ell \) separately. The term corresponding to \( \mathbf{C} \) is the **Gaussian component**, and the corresponding measure \( \gamma_{0, \mathbf{C}} \) is the distribution of \( y \sim \mathbf{C}^2 y \) under the standard Gaussian measure

\[
\gamma_{0, 1}(dy) = (2\pi)^{-N} e^{-\frac{y^2}{2}} dy.
\]

To deal with the term corresponding to \( \mathbf{M} \), initially assume that \( \mathbf{M} \) is finite, and consider the Poisson measure

\[
\Pi_M = e^{-M(\mathbb{R}^N)} \sum_{k=0}^\infty \frac{M^k}{k!},
\]

where \( M^k \equiv \delta_0 \) (\( \delta_0 \) denotes the unit point mass at \( \mathbf{a} \)) and, for \( k \geq 1 \), \( M^k = M * M^{(k-1)} \) is the \( k \)-fold convolution of \( M \) with itself. One then has that

\[
\Pi_M(\xi) = e^{M(\mathbb{R}^N)} \sum_{k=0}^\infty \frac{\hat{M}(\xi)^k}{k!} = \exp \left( \int_{\mathbb{R}^N} (e^{i\langle \xi, y \rangle} - 1) M(dy) \right).
\]

To handle general Lévy measures \( M \), for each \( r \in (0, 1) \), define \( M_r(dy) = 1_{(r, \infty)}(y)M(dy) \) and \( \mathbf{a}_r = \int_{B(0,1)} y M_r(dy) \). Then

\[
\langle \delta_{-\mathbf{a}_r} * \Pi_{M_r}, \gamma \rangle(\xi) = \exp \left( \int_{B(0, r)c} (e^{i\langle \xi, y \rangle} - 1 - i \mathbf{1}_{B(0,1)}(y)\langle \xi, y \rangle_{\mathbb{R}^N}) M(dy) \right) \to \exp \left( \int_{\mathbb{R}^N} (e^{i\langle \xi, y \rangle} - 1 - i \mathbf{1}_{B(0,1)}(y)\langle \xi, y \rangle_{\mathbb{R}^N}) M(dy) \right)
\]

uniformly for \( \xi \) in compact subsets of \( \mathbb{R}^N \). Hence, by Lévy’s Continuity Theorem (which states that if \( \{\mu_n : n \geq 0\} \subseteq \mathbf{M}_1(\mathbb{R}^N) \) and \( \mu_n \) is the characteristic function (i.e., the Fourier transform) of \( \mu_n \), then \( \mu = \lim_{n \to \infty} \mu_n \) exists in \( \mathbf{M}_1(\mathbb{R}^N) \) if and only if \( \hat{\mu}_n(\xi) \) converges for each \( \xi \) and uniformly in a neighborhood of \( 0 \), in which case \( \mu_n \to \mu \) in \( \mathbf{M}_1(\mathbb{R}^N) \) where \( \hat{\mu}(\xi) = \lim_{n \to \infty} \hat{\mu_n}(\xi) \), there is an element \( \mu_M \) of \( \mathbf{M}_1(\mathbb{R}^N) \) whose Fourier transform is

\[
\exp \left( \int_{\mathbb{R}^N} (e^{i\langle \xi, y \rangle} - 1 - i \mathbf{1}_{B(0,1)}(y)\langle \xi, y \rangle_{\mathbb{R}^N}) M(dy) \right).
\]

Hence, if \( \mu = \delta_{\mathbf{m}} * \mu_M * \gamma_{0, \mathbf{C}} \), then \( \hat{\mu}(\xi) = e^{i\langle \xi \rangle} \).
Kolmogorov’s Equations

Notice that \( \mu_{s+t} = \mu_s \ast \mu_t \), and therefore, for each \( n \geq 1 \), \( \mu_1 = \mu_1^n \). A probability measure that admits an \( n \)th convolution root for all \( n \geq 1 \) is said to be infinitely devisable, and a theorem of Lévy and Khinchine says that the only divisible measures \( \mu \) are those for which there exists a \((m, C, M)\) such that \( \hat{\mu}(\xi) = e^{\ell(\xi)} \). Perhaps more important from the perspective here is the evolution equation, known as Kolmogorov’s forward equation, that this property implies \( t \mapsto \mu_t \) satisfies. To describe this equation, define the operator \( L \) on \( D \) by

\[
L \varphi(x) = A \varphi(x + \cdot) = (m, \nabla \varphi(x))_{\mathbb{R}^N} + \frac{1}{2} \text{Trace}(\nabla^2 \varphi(x)) + \int_{\mathbb{R}^N} \left( \varphi(x + y) - \varphi(x) - 1_{B(0,1)}(y) \left( y, \nabla \varphi(x) \right)_{\mathbb{R}^N} \right) M(dy).
\]

What I want to show is that

\[
(8) \quad \frac{d}{dt} \langle \varphi, \mu_t \rangle = \langle L \varphi, \mu_t \rangle.
\]

Since this is trivial when \( \varphi \) is constant, assume that \( \varphi \in \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \). Then, as \( h \downarrow 0 \),

\[
\frac{\langle \varphi, \mu_{t+h} \rangle - \langle \varphi, \mu_t \rangle}{h} = \frac{1}{h} \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} \frac{\langle \varphi(y + \cdot), \mu_h \rangle - \langle \varphi(y) \rangle}{h} \mu_t(dy) \right) \mu_t(dy)
= (2\pi)^{-N} \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} e^{-i(\xi, y)} e^{\ell(t - \xi)} - 1 \hat{\varphi}(\xi) d\xi \right) \mu_t(dy)
\to (2\pi)^{-N} \int_{\mathbb{R}^N} \left( \int_{\mathbb{R}^N} e^{-i(\xi, y)} \ell(t - \xi) \hat{\varphi}(\xi) d\xi \right) \mu_t(dy) = \langle L \varphi, \mu_t \rangle.
\]

I now want to show that (8) continues to hold for all \( \varphi \in C^2_b(\mathbb{R}^N; \mathbb{R}) \). To this end, first note that

\[
|L \varphi(x)| \leq \|C\|_{\text{op}} \|\nabla^2 \varphi(x)\|_{\text{op}} + |m| |\nabla \varphi(x)|
+ \frac{1}{2} \left( \int_{B(0,1)} |y|^2 M(dy) \right) \sup_{y \in B(x,1)} \|\nabla^2 \varphi(y)\|_{\text{op}} + 2M(B(0,1)\mathbb{C}) \|\varphi\|_u.
\]

Now let \( \varphi \in C^2_b(\mathbb{R}^N; \mathbb{R}) \) be given. Then we can choose \( \{\varphi_n : n \geq 1\} \subseteq \mathcal{S}(\mathbb{R}^N; \mathbb{R}) \) to be a sequence of functions that is bounded in \( C^2_b(\mathbb{R}^N; \mathbb{R}) \) and for which \( \varphi_n \to \varphi, \nabla \varphi_n \to \nabla \varphi, \) and \( \nabla^2 \varphi_n \to \nabla^2 \varphi \) uniformly on compacts, in which
case $\sup_{n \geq 1} \|L\varphi_n\|_u < \infty$ and $L\varphi_n(x) \to L\varphi(x)$ uniformly for $x$ in compacts. Hence, from (8), one has that

$$\langle \varphi, \mu_t \rangle - \varphi(0) = \lim_{n \to \infty} \langle \varphi_n, \mu_t \rangle - \varphi_n(0) = \lim_{n \to \infty} \int_0^t \langle L\varphi_n, \mu_t \rangle \, d\tau = \int_0^t \langle L\varphi, \mu_t \rangle \, d\tau,$$

and so (8) continues to hold for $\varphi \in C^2_b(\mathbb{R}^N; \mathbb{R})$. Knowing (8) for $\varphi \in C^2_b(\mathbb{R}^N; \mathbb{R})$ and applying it to $e^{i(\xi, y)}$, one sees that it implies that

$$\frac{d}{dt} \hat{\mu}_t(\xi) = \ell(\xi) \hat{\mu}_t(\xi).$$

Hence (8) together with the initial condition $\mu_0 = \delta_0$ implies that $\hat{\mu}_t(\xi) = e^{t\ell(\xi)}$. That is, (8) plus $\mu_0 = \delta_0$ uniquely determine $\mu_t$.

There is another important equation associated with the measures $\mu_t$. To develop this equation, set $P(t, x, \cdot) = \delta_x^* \mu_t$. Then

$$P(s + t, x, \Gamma) = \int_{\mathbb{R}^N} \mu_t(x + y + \Gamma) \mu_s(dy) = \int_{\mathbb{R}^N} \mu_t(y + \Gamma) P(s, x, dy)$$

$$= \int_{\mathbb{R}^N} P(t, y, \Gamma) P(s, x, dy),$$

and so $P(t, x, \cdot)$ satisfies the Chapman–Kolmogorov equation

$$P(s + t, x, \Gamma) = \int_{\mathbb{R}^N} P(t, y, \Gamma) P(s, x, dy),$$

which means that $(t, x) \mapsto P(t, x, \cdot)$ is a transition probability. Equivalently, if the operators $P_t$ are defined for $t \geq 0$ on bounded, Borel measurable functions $\varphi$ by

$$P_t \varphi(x) = \int_{\mathbb{R}^N} \varphi(y) P(t, x, dy),$$

then $\{P_t : t \geq 0\}$ is a semigroup: $P_{s+t} = P_s P_t$. To compute the generator of this semigroup, first observe that (8) says that

$$\frac{d}{dt} P_t \varphi = P_t L\varphi$$

for $\varphi \in C^2_b(\mathbb{R}^N; \mathbb{R})$, which is also called Kolmogorov's forward equation. Next observe that $P_t$ is a bounded map of $C^2_b(\mathbb{R}^N; \mathbb{R})$ into itself. Therefore, if $\varphi \in C_b(\mathbb{R}^N; \mathbb{R})$, then

$$\frac{P_{t+h} \varphi(x) - P_t \varphi(x)}{h} = \frac{P_h P_t \varphi(x) - P_t \varphi(x)}{h} \to L P_t \varphi(x)$$
as \( h \downarrow 0 \). Equivalently, if \( u(t, x) \equiv P_t \varphi(x) \), then \( u \in C^{1,2}_b([0, \infty) \times \mathbb{R}^N; \mathbb{R}) \) and is a solution to Kolmogorov’s **backward equation**

\[
\partial_t u = L u \quad \text{and} \quad \lim_{t \to 0} u(t, \cdot) = \varphi.
\]

In fact, it is the only such solution. Indeed, if \( u \in C^{1,2}_b([0, \infty) \times \mathbb{R}^N; \mathbb{R}) \) solves (11), then, by (10) and the chain rule,

\[
\frac{d}{d\tau} P_\tau u(t - \tau, \cdot) = P_\tau L u(t - \tau, \cdot) - P_\tau L u(t, \cdot) = 0
\]

for \( \tau \in (0, t) \), and so \( P_t \varphi = u(t, \cdot) \).

**Solving the Forward Equation with Variable Coefficients**

Here I will show how one can go about solving Kolmogorov’s forward equation for \( L \)'s which are local but have variable coefficients. That is,

\[
L \varphi(x) = \frac{1}{2} \sum_{i,j=1}^N a_{ij}(x) \partial_{x_i} \partial_{x_j} \varphi(x) + \sum_{i=1}^N b_i(x) \partial_{x_i} \varphi(x),
\]

where \( a(x) = \left((a_{ij}(x))\right)_{1 \leq i, j \leq N} \) is a non-negative definite, symmetric matrix for each \( x \in \mathbb{R}^N \). In the probability literature, \( a \) is called the **diffusion coefficient** and \( b \) is called the **drift coefficient**. The following is the basic existence result.

**Theorem 13.** Assume that

\[
\Lambda = \sup_{x \in \mathbb{R}^N} \frac{\text{Trace}(a(x)) + 2(x, b(x))_{\mathbb{R}^N}^+}{1 + |x|^2} < \infty.
\]

Then, for each \( \nu \in M_1(\mathbb{R}^N) \), there is a continuous \( t \in [0, \infty) \longmapsto \mu(t) \in M_1(\mathbb{R}^N) \) which satisfies

\[
\langle \varphi, \mu(t) \rangle - \langle \varphi, \nu \rangle = \int_0^t \langle L \varphi, \mu(\tau) \rangle \, d\tau,
\]

for all \( \varphi \in C^2_c(\mathbb{R}^N; \mathbb{C}) \), where \( L \) is the operator in (12). Moreover,

\[
\int (1 + |y|^2) \mu(t, dy) \leq e^{\Lambda t} \int (1 + |x|^2) \nu(dx), \quad t \geq 0.
\]

Before giving the proof, it may be helpful to review the analogous result for ordinary differential equations. Indeed, when applied to the case when \( a = 0 \), our proof is exactly the same as the Euler approximation scheme used there. Namely,
in that case, except for the initial condition, there should be no randomness, and so, when we remove the randomness from the initial condition by taking $\nu = \delta_x$, we expect that $\mu_t = \delta_{X(t)}$, where $t \in [0, \infty) \mapsto X(t) \in \mathbb{R}^N$ satisfies

$$\varphi(X(t)) - \varphi(x) = \int_0^t (b(X(\tau)), \nabla \varphi(X(\tau)))_g \, d\tau.$$  

Equivalently, $t \mapsto X(t)$ is an integral curve of the vector field $b$ starting at $x$. That is,

$$X(t) = x + \int_0^t b(X(\tau)) \, d\tau.$$  

To show that such an integral curve exists, one can use the following approximation scheme. For each $n \geq 0$, define $t \mapsto X_n(t)$ so that $X_n(0) = x$ and

$$X_n(t) = X_n(m2^{-n}) + (t - m2^{-n})b(X(m2^{-n})) \quad \text{for} \quad m2^{-n} < t \leq (m + 1)2^{-n}.$$  

Clearly,

$$X_n(t) = x + \int_0^t b(X_n([\tau]_n)) \, d\tau,$$

where $[\tau]_n = 2^{-n}[2^n \tau]$ is the largest diadic number $m2^{-n}$ dominated by $\tau$. Hence, if we can show that $\{X_n : n \geq 0\}$ is relatively compact in the space $C([0, \infty); \mathbb{R}^N)$, with the topology of uniform convergence on compacts, then we can take $t \mapsto X(t)$ to be any limit of the $X_n$’s.

To simplify matters, assume for the moment that $b$ is bounded. In that case, it is clear that $|X_n(t) - X_n(s)| \leq \|b\|_u|t - s|$, and so the Ascoli–Arzela Theorem guarantees the required compactness. To remove the boundedness assumption, choose a $\psi \in C^\infty_c(B(0, 2); [0, 1])$ so that $\psi = 1$ on $B(0, 1)$ and, for each $k \geq 1$, replace $b$ by $b_k$, where $b_k(x) = \psi(k^{-1}x)b(x)$. Next, let $t \mapsto X_k(t)$ be an integral curve of $b_k$ starting at $x$, and observe that

$$\frac{d}{dt}|X_k(t)|^2 = 2\langle X_k(t), b_k(X_k(t)) \rangle_g \leq \Lambda(1 + |X_k(t)|^2),$$

from which it is an easy step to the conclusion that

$$|X_k(t)| \leq R(T) \equiv (1 + |x|^2)e^{t\Lambda}.$$  

But this means that, for each $T > 0$, $|X_k(t) - X_k(s)| \leq C(T)|t - s|$ for $s, t \in [0, T]$, where $C(T)$ is the maximum value of $|b|$ on the closed ball of radius $R(T)$ centered at the origin, and so we again can invoke the Ascoli–Arzela Theorem to see that

\[1\] I use $[\tau]$ to denote the integer part of a number $\tau \in \mathbb{R}$
\{X_k : k \geq 1\} is relatively compact and therefore has a limit which is an integral curve of $b$.

In view of the preceding, it should be clear that our first task is to find an appropriate replacement for the Ascoli–Arzela Theorem. The one which we will choose is a variant of Lévy’s Continuity Theorem.

In the following, and elsewhere, we say that $\{\varphi_k : k \geq 1\} \subseteq C_b(\mathbb{R}^N; \mathbb{C})$ converges to $\varphi$ in $C_b(\mathbb{R}^N; \mathbb{C})$ and write $\varphi_k \longrightarrow \varphi$ in $C_b(\mathbb{R}^N; \mathbb{C})$ if $\sup_k \|\varphi_k\|_u < \infty$ and $\varphi_k(x) \longrightarrow \varphi(x)$ uniformly for $x$ in compact subsets of $\mathbb{R}^N$. Also, we say that $\{\mu_k : k \geq 1\} \subseteq C(\mathbb{R}^M; M_1(\mathbb{R}^N))$ converges to $\mu$ in $C(\mathbb{R}^M; M_1(\mathbb{R}^N))$ and write $\mu_k \longrightarrow \mu$ in $C(\mathbb{R}^M; M_1(\mathbb{R}^N))$ if, for each $\varphi \in C_b(\mathbb{R}^N; \mathbb{C})$, $\langle \varphi, \mu_k(z) \rangle \longrightarrow \langle \varphi, \mu(z) \rangle$ uniformly for $z$ in compact subsets of $\mathbb{R}^M$.

**Theorem 17.** If $\mu_k \longrightarrow \mu$ in $C(\mathbb{R}^M; M_1(\mathbb{R}^N))$, then

$$
\langle \varphi, \mu_k(z) \rangle \longrightarrow \langle \varphi, \mu(z) \rangle
$$

whenever $z_k \longrightarrow z$ in $\mathbb{R}^M$ and $\varphi_k \longrightarrow \varphi$ in $C_b(\mathbb{R}^N; \mathbb{C})$. Moreover, if $\{\mu_n : n \geq 0\} \subseteq C(\mathbb{R}^M; M_1(\mathbb{R}^N))$ and $f_n(z, \xi) = \mu_n(z)(\xi)$, then $\{\mu_n : n \geq 0\}$ is relatively compact in $C(\mathbb{R}^M; M_1(\mathbb{R}^N))$ if each $(z, \xi) \in \mathbb{R}^M \times \mathbb{R}^N$. In particular, $\{\mu_n : n \geq 0\}$ is relatively compact if, for each $\xi \in \mathbb{R}^N$, $\{f_n(\cdot, \xi) : n \geq 0\}$ is equicontinuous at each $z \in \mathbb{R}^M$ and, for each $r \in (0, \infty)$,

$$
\lim_{R \to \infty} \sup_{n \geq 0} \mu_n(z, \mathbb{R}^N \setminus B(0, R)) = 0.
$$

**Proof:** To prove the first assertion, suppose $\mu_k \longrightarrow \mu$ in $C(\mathbb{R}^M; M_1(\mathbb{R}^N))$, $z_k \longrightarrow z$ in $\mathbb{R}^M$, and $\varphi_k \longrightarrow \varphi$ in $C_b(\mathbb{R}^N; \mathbb{C})$. Then, for every $R > 0$,

$$
\begin{align*}
\limsup_{k \to \infty} \left| \langle \varphi_k, \mu_k(z_k) \rangle - \langle \varphi, \mu(z) \rangle \right| &\leq \limsup_{k \to \infty} \left( |\langle \varphi_k - \varphi, \mu_k(z_k) \rangle| + |\langle \varphi, \mu_k(z_k) \rangle - \langle \varphi, \mu(z) \rangle| \right) \\
&\leq \limsup_{k \to \infty} \left( \sup_{y \in B(0, R)} |\varphi_k(y) - \varphi(y)| + 2 \sup_k \|\varphi_k\|_u \lim_{k \to \infty} \mu_k(z_k, B(0, R))^C \right) \\
&\leq 2 \sup_k \|\varphi_k\|_u \mu(z, B(0, R))^C
\end{align*}
$$

since $\limsup_{k \to \infty} \mu_k(z_k, F) \leq \mu(z, F)$ for any closed $F \subseteq \mathbb{R}^N$. Hence, the required conclusion follows after one lets $R \to \infty$.

Turning to the second assertion, apply the Arzela–Ascoli Theorem to produce an $f \in C_b(\mathbb{R}^M \times \mathbb{R}^N; \mathbb{C})$ and a subsequence $\{n_k : k \geq 0\}$ such that $f_{n_k} \longrightarrow f$ uniformly on compacts. By Lévy’s Continuity Theorem, there is, for each $z \in \mathbb{R}^M$, a $\mu(z) \in M_1(\mathbb{R}^N)$ for which $f(z, \cdot) = \mu(z)$. Moreover, if $z_k \longrightarrow z$ in $\mathbb{R}^M$, then, because $f_{n_k}(z_k, \cdot) \longrightarrow f(z, \cdot)$ uniformly on compact subsets of $\mathbb{R}^N$,
another application of Lévy’s Theorem shows that \( \mu_n(\mathbf{z}_k) \rightarrow \mu(\mathbf{z}) \) in \( \text{M}_1(\mathbb{R}^N) \), and from this it is clear that \( \mu_n \rightarrow \mu \) in \( C(\mathbb{R}^M; \text{M}_1(\mathbb{R}^N)) \).

It remains to show that, under the conditions in the final assertion, \( \{f_n : n \geq 0\} \) is equicontinuous at each \( (\mathbf{z}, \xi) \). But, by assumption, for each \( \xi \in \mathbb{R}^N \), \( \{f_n(\cdot, \xi) : n \geq 0\} \) is equicontinuous at every \( \mathbf{z} \in \mathbb{R}^M \). Thus, it suffices to show that if \( \xi_k \rightarrow \xi \) in \( \mathbb{R}^N \), then, for each \( r > 0 \),

\[
\lim_{k \to \infty} \sup_{n \geq 0} \sup_{|\mathbf{z}| \leq r} |f_n(\mathbf{z}, \xi_k) - f_n(\mathbf{z}, \xi)| = 0.
\]

To this end, note that, for any \( R > 0 \),

\[
|f_n(\mathbf{z}, \xi_k) - f_n(\mathbf{z}, \xi)| \leq R|\xi_k - \xi| + 2\mu_n(\mathbf{z}, B(0, R)\mathcal{C}),
\]

and therefore

\[
\lim_{k \to \infty} \sup_{n \geq 0} \sup_{|\mathbf{z}| \leq r} |f_n(\mathbf{z}, \xi_k) - f_n(\mathbf{z}, \xi)| \leq 2\sup_{n \geq 0} \mu_n(\mathbf{z}, B(0, R)\mathcal{C}) \rightarrow 0
\]

as \( R \to \infty \). \( \square \)

Now that we have a suitable compactness criterion, the next step is to develop an Euler approximation scheme. To do so, we must decide what plays the role in \( \text{M}_1(\mathbb{R}^N) \) that linear translation plays in \( \mathbb{R}^N \). A hint comes from the observation that if \( t \sim X(t, \mathbf{x}) = x + tb \) is a linear translation along the constant vector field \( b \), then \( X(s + t, \mathbf{x}) = X(s, \mathbf{x}) + X(t, 0) \). Equivalently,

\[
\delta_{X(s + t, x)} = \delta_{X(s, x)} * \delta_{X(t, 0)}.
\]

Thus, “linear translation” in \( \text{M}_1(\mathbb{R}^N) \) should be a path \( t \in [0, \infty) \mapsto \mu(t) \in \text{M}_1(\mathbb{R}^N) \) given by \( \mu(t) = \nu * \lambda(t) \), where \( t \sim \lambda(t) \) satisfies \( \lambda(0) = \delta_0 \) and \( \lambda(s + t) = \lambda(s) * \lambda(t) \). That is, \( \mu(t) = \nu * \lambda(t) \), where \( \lambda(t) \) is infinitely divisible. Moreover, because \( L \) is local, the only infinitely divisible laws which can appear here must be Gaussian. With these hints, we now take \( Q(t, \mathbf{x}) = \gamma_{tb(\mathbf{x}), ta(\mathbf{x})} \), the distribution of \( y \sim x + tb(\mathbf{x}) + t\frac{\sigma(\mathbf{x})}{\sigma(\mathbf{x})} y \) under \( \gamma_{0,1} \), where \( \sigma : \mathbb{R}^N \rightarrow \text{Hom}(\mathbb{R}^M; \mathbb{R}^N) \) is a square root of \( a \) in the sense that \( a(\mathbf{x}) = \sigma(\mathbf{x})\sigma(\mathbf{x})^\top \). To check that \( Q(t, \mathbf{x}) \) will play the role that \( x + tb(\mathbf{x}) \) played above, observe that if \( \varphi \in C^2(\mathbb{R}^N; \mathbb{C}) \) and \( \varphi \) together with its derivatives have at most exponential growth, then

\[
\langle \varphi, Q(t, \mathbf{x}) \rangle - \varphi(\mathbf{x}) = \int_0^t \langle L^\tau \varphi, Q(\tau, x) \rangle d\tau,
\]

where \( L^\tau \varphi(\mathbf{y}) = \frac{1}{2} \sum_{i,j} a_{ij}(\mathbf{x}) \partial_{x_i} \partial_{x_j} \varphi(|\mathbf{y}|) + \sum_{i=1}^N b_i(\mathbf{x}) \partial_{y_i} \varphi(\mathbf{y}) \).

\(^2\)At the moment, it makes no difference which choice of square root one chooses. Thus, one might as well assume here that \( \sigma(\mathbf{x}) = a(\mathbf{x})^{\frac{1}{2}} \), the non-negative definite, symmetric square root \( a(\mathbf{x}) \). However, later on it will be useful to have kept our options open.
To verify (18), simply note that
\[
\frac{d}{dt} \langle \varphi, Q(t, x) \rangle = \frac{d}{dt} \int_{\mathbb{R}^N} \varphi(x + \sigma(x)y + tb(x)y) \gamma_{0,1}(dy) = \int_{\mathbb{R}^N} \varphi(x + \sigma(x)y + tb(x)y) \Delta_y \left( \frac{1}{(2\pi t)^{N/2}} e^{-|y|^2/2t} \right) dy,
\]
and integrate twice by parts to move the $\Delta$ over to $\varphi$. As a consequence of either (18) or direct computation, we have
\[
(19) \quad \int |y|^2 Q(t, x, dy) = |x + tb(x)|^2 + t \text{Trace}(a(x)).
\]

Now, for each $n \geq 0$, define the Euler approximation $t \in [0, \infty) \mapsto \mu_n(t) \in M_1(\mathbb{R}^N)$ so that
\[
(20) \quad \mu_n(0) = \nu \quad \text{and} \quad \mu_n(t) = \int Q(t - m2^{-n}, y) \mu_n(m2^{-n}, dy)
\quad \text{for } m2^{-n} < t \leq (m + 1)2^{-n}.
\]

By (19), we know that
\[
(21) \quad \int_{\mathbb{R}^N} |y|^2 \mu_n(t, dy) = \int_{\mathbb{R}^N} \left[ |y + (t - m2^{-n})b(y)|^2 
\quad + (t - m2^{-n}) \text{Trace}(a(y)) \right] \mu_n(m2^{-n}, dy)
\quad \text{for } m2^{-n} \leq t \leq (m + 1)2^{-n}.
\]

**Lemma 22.** Assume that
\[
(23) \quad \lambda = \sup_{x \in \mathbb{R}^N} \frac{\text{Trace}(a(x)) + 2|b(x)|^2}{1 + |x|^2} < \infty.
\]

Then
\[
(24) \quad \sup_{n \geq 0} \int_{\mathbb{R}^N} (1 + |y|^2) \mu_n(t, dy) \leq e^{(1+\lambda)t} \int_{\mathbb{R}^N} (1 + |x|^2) \nu(dx).
\]

In particular, if $\int |x|^2 \nu(dx) < \infty$, then $\{\mu_n : n \geq 0\}$ is a relatively compact subset of $C([0, \infty); M_1(\mathbb{R}^N))$. 
PROOF: Suppose that \( m2^{-n} \leq t \leq (m+1)2^{-n} \), and set \( \tau = t - m2^{-n} \). First note that

\[
|y + \tau b(y)|^2 + \tau \text{Trace}(a(y)) \\
= |y|^2 + 2\tau(y, b(y))_{\mathbb{R}^n} + \tau^2|b(y)|^2 + \tau \text{Trace}(a(y)) \\
\leq |y|^2 + \tau[|y|^2 + 2|b(y)|^2 + \text{Trace}(a(y))] \\
\leq |y|^2 + (1 + \lambda)\tau(1 + |y|^2),
\]

and therefore, by (21),

\[
\int (1 + |y|^2) \mu_n(t, dy) \leq (1 + (1 + \lambda)\tau) \int (1 + |y|^2) \mu_n(m2^{-n}, dy).
\]

Hence,

\[
\int (1 + |y|^2) \mu_n(t, dy) \\
\leq (1 + (1 + \lambda)2^{-n})^m(1 + (1 + \lambda)\tau) \int (1 + |y|^2) \nu(dy) \\
\leq e^{(1+\lambda)t} \int (1 + |x|^2) \nu(dx).
\]

Next, set \( f_n(t, \xi) = [\widehat{\mu_n(t)}](\xi) \). Under the assumption that the second moment \( S \equiv \int |x|^2 \nu(dx) < \infty \), we want to show that \( \{f_n : n \geq 0\} \) is equicontinuous at each \( (t, \xi) \in [0, \infty) \times \mathbb{R}^N \). Since, by (24),

\[
\mu_n(t, B(0, R)\mathcal{C}) \leq S(1 + R^2)^{-1}e^{(1+\lambda)t},
\]

the last part of Theorem 17 says that it suffices to show that, for each \( \xi \in \mathbb{R}^N \), \( \{f_n(\cdot, \xi) : n \geq 0\} \) is equicontinuous at each \( t \in [0, \infty) \). To this end, first observe that, for \( m2^{-n} \leq s < t \leq (m+1)2^{-n} \),

\[
|f_n(t, \xi) - f_n(s, \xi)| \leq \int ||\widehat{Q(t, y)}(\xi) - \widehat{Q(s, y)}(\xi)|| \mu_n(m2^{-n}, dy)
\]

and, by (18),

\[
||\widehat{Q(t, y)}(\xi) - \widehat{Q(s, y)}(\xi)|| = \left| \int_t^s \left( \int L^{\xi}_{y'}(y')Q(\tau, y, dy') \right) d\tau \right| \\
\leq (t-s)\left( \frac{1}{2}(\xi, a(y)\xi)_{\mathbb{R}^n} + ||\xi||b(y)|| \right) \leq \frac{1}{2}(1 + \lambda)(1 + |y|^2)(1 + |\xi|^2)(t-s),
\]

where \( e_\xi(y) \equiv e^{(\xi,y)\mathcal{C}} \). Hence, by (24),

\[
|f_n(t, \xi) - f_n(s, \xi)| \leq \frac{(1 + \lambda)(1 + |\xi|^2)}{2}e^{(1+\lambda)t} \int (1 + |x|^2) \nu(dx)(t-s),
\]
first for \( s < t \) in the same diadic interval and then for all \( s < t \). \( \square \)

With Lemma 22, we can now prove Theorem 13 under the assumptions that \( a \) and \( b \) are bounded and that \( \int |x|^2 \nu(dx) < \infty \). Indeed, because we know then that \( \{\mu_n : n \geq 0\} \) is relatively compact in \( C([0, \infty); M_1(\mathbb{R}^N)) \), all that we have to do is show that every limit satisfies (15). For this purpose, first note that, by (18),

\[
\langle \varphi, \mu_n(t) \rangle - \langle \varphi, \nu \rangle = \int_0^t \left( \int \langle L^\varphi, Q(\tau - [\tau]_n, y) \rangle \mu_n([\tau]_n, dy) \right) d\tau
\]

for any \( \varphi \in C^2_b(\mathbb{R}^N; \mathbb{C}) \). Next, observe that, as \( n \to \infty \),

\[
\langle L^\varphi, Q(\tau - [\tau]_n, y) \rangle \to L^\varphi(y)
\]

boundedly and uniformly for \((\tau, y)\) in compacts. Hence, if

\[
\mu_{nk} \to \mu \text{ in } C([0, \infty); M_1(\mathbb{R}^N)),
\]

then, by Theorem 17,

\[
\langle \varphi, \mu_{nk}(t) \rangle \to \langle \varphi, \mu(t) \rangle \quad \text{and}
\]

\[
\int_0^t \left( \int \langle L^\varphi, Q(\tau - [\tau]_n, y) \rangle \mu_n([\tau]_n, dy) \right) d\tau \to \int_0^t \langle L^\varphi, \mu(\tau) \rangle d\tau.
\]

Before moving on, I want to show that \( \int |x|^2 \nu(dx) < \infty \) implies that (15) continues to hold for \( \varphi \in C^2(\mathbb{R}^N; \mathbb{C}) \) with bounded second order derivatives. Indeed, from (24), we know that

\[
(*) \quad \int (1 + |y|^2) \mu(t, dy) \leq e^{(1+\lambda)t} \int (1 + |y|^2) \nu(dy).
\]

Now choose \( \psi \in C^\infty_c(\mathbb{R}^N; [0, 1]) \) so that \( \psi = 1 \) on \( \overline{B(0, 1)} \) and \( \psi = 0 \) off of \( B(0, 2) \), define \( \psi_R \) by \( \psi_R(y) = \psi(R^{-1} y) \) for \( R \geq 1 \), and set \( \varphi_R = \psi_R \varphi \). Observe that\(^3\)

\[
\frac{\|\varphi(y)\|}{1 + |y|^2} \vee \frac{\|\nabla \varphi(y)\|}{1 + |y|} \vee \|\nabla^2 \varphi(y)\|_{\text{H.S.}}
\]

is bounded independent of \( y \in \mathbb{R}^N \), and therefore so is \( \frac{|L^\varphi(y)|}{1 + |y|^2} \). Thus, by (\( * \)), there is no problem about integrability of the expressions in (15). Moreover, because (15) holds for each \( \varphi_R \), all that we have to do is check that

\[
\langle \varphi, \mu(t) \rangle = \lim_{R \to \infty} \langle \varphi_R, \mu(t) \rangle
\]

\[
\int_0^t \langle L^\varphi, \mu(\tau) \rangle d\tau = \lim_{R \to \infty} \int_0^t \langle L^\varphi_R, \mu(\tau) \rangle d\tau.
\]

\(^3\) We use \( \|\sigma\|_{\text{H.S.}} \) to denote the Hilbert–Schmidt norm \( \sqrt{\sum_{i,j} \sigma_{ij}^2} \) of \( \sigma \).
The first of these is an immediate application of Lebesgue’s Dominated Convergence Theorem. To prove the second, observe that

$$L\varphi_B(y) = \psi_R(y)L\varphi(y) + (\nabla\psi_R(y), a(y)\nabla\varphi)_{\mathbb{R}^N} + \varphi(y)L\psi_R(y).$$

Again the first term on the right causes no problem. To handle the other two terms, note that, because \(\psi_R\) is constant off of \(B(0,2R) \setminus B(0,R)\) and because \(\nabla\psi_R(y) = R^{-1}\nabla\psi(R^{-1}y)\) while \(\nabla^2\psi_R(y) = R^{-2}\nabla^2\psi(R^{-1}y)\), one can easily check that they are dominated by a constant, which is independent of \(R\), times \((1+|y|^2)\mathbf{1}_{[R,2R]}(|y|)\). Hence, again Lebesgue’s Dominated Convergence Theorem gives the desired result.

Knowing that (15) holds for \(\varphi \in \mathcal{C}^2(\mathbb{R}^N; \mathbb{C})\) with bounded second order derivatives, we can prove (16) by taking \(\varphi(y) = 1 + |y|^2\) and thereby obtaining

$$\int (1 + |y|^2) \mu(t, dy)$$

$$= \int (1 + |y|^2) \nu(dy) + \int_0^t \left( \int [\text{Trace}(a(y)) + 2(y, b(y))_{\mathbb{R}^N}] \mu(\tau, dy) \right) d\tau$$

$$\leq \int (1 + |y|^2) \nu(dy) + \Lambda \int_0^t \left( \int (1 + |y|^2) \mu(\tau, dy) \right) d\tau,$$

from which (16) follows by Gronwall’s lemma.

Continuing with the assumption that \(\int |x|^2 \nu(dx) < \infty\), I want to remove the boundedness assumption on \(a\) and \(b\) and replace it by (14). To do this, take \(\psi_R\) as above, set \(a_k = \psi_k a\), \(b_k = \psi_k b\), define \(L_k\) accordingly for \(a_k\) and \(b_k\), and choose \(t \to \mu_k(t)\) so that (16) is satisfied and (15) holds when \(\mu\) and \(L\) are replaced there by \(\mu_k\) and \(L_k\). Because of (16), the argument which was used earlier can be repeated to show that \(\{\mu_k : k \geq 1\}\) is relatively compact in \(C([0, \infty); \mathcal{M}_1(\mathbb{R}^N))\). Moreover, if \(\mu\) is any limit of \(\{\mu_k : k \geq 1\}\), then (16) is satisfied and, just as we did above, one can check (15), first for \(\varphi \in \mathcal{C}^2(\mathbb{R}^N; \mathbb{C})\) and then for all \(\varphi \in \mathcal{C}^2(\mathbb{R}^N; \mathbb{C})\) with bounded second order derivatives.

Finally, to remove the second moment condition on \(\nu\), assume that it fails, and choose \(r_k \to \infty\) so that

$$\alpha_1 \equiv \nu(B(0, r_1)) > 0$$

and \(\alpha_k \equiv \nu(B(0, r_k) \setminus B(0, r_{k-1})) > 0\) for each \(k \geq 2\),

and set \(\nu_1 = \alpha_1^{-1} \nu | B(0, r_1)\) and \(\nu_k = \alpha_k^{-1} \nu | B(0, r_k) \setminus B(0, r_{k-1})\) when \(k \geq 2\).

Finally, choose \(t \to \mu_k(t)\) for \(L\) and \(\nu_k\), and define \(\mu(t) = \sum \alpha_k \mu_k(t)\). It is an easy matter to check that this \(\mu\) satisfies (15) for all \(\varphi \in \mathcal{C}^2(\mathbb{R}^N; \mathbb{C})\).

**Remark 25.** In order to put the result in Theorem 13 into a partial differential equations context, it is best to think of \(t \to \mu(t)\) as a solution to \(\partial_t \mu = L^\dagger \mu(t)\) in the sense of (Schwartz) distributions. Of course, when the coefficients of \(L\) are not smooth, one has to be a little careful about the meaning of \(L^\dagger \mu(t)\).

The reason why this causes no problem here is that, by virtue of the minimum principle (cf. §2.4.1), the only distributions with which we need to deal are probability measures.
Transition Probability Functions

Although we have succeeded in solving Kolmogorov’s forward equation in great generality, we have not yet produced a transition probability function. To be more precise, let $S(x)$ denote the set of maps $t \mapsto \mu(t)$ satisfying (15) with $\nu = \delta_x$. In order to construct a transition probability function, one must make a measurable “selection” $x \in \mathbb{R}^N \mapsto P(\cdot, x) \in S(x)$ in such a way that the Chapman–Kolmogorov equation (9) holds. Thus, the situation is the same as that one encounters in the study of ordinary differential equations when trying to construct a flow on the basis of only an existence result for solutions. In the absence of an accompanying uniqueness result, how does one go about showing that there is a “selection” of solutions which fit together nicely into a flow?

It turns out that, under the hypotheses in Theorem 13, one can always make a selection $x \in \mathbb{R}^N \mapsto P(\cdot, x) \in S(x)$ which forms a transition probability function. The underlying idea is to introduce enough spurious additional conditions to force uniqueness. In doing so, one has to take advantage of the fact that, for each $x \in \mathbb{R}^N$, $S(x)$ is a compact, convex subset of $M_1(\mathbb{R}^N)$ and that, if $y \to x$ in $\mathbb{R}^N$, then $\lim_{y \to x} S(y) \subseteq S(x)$ in the sense of Hausdorff convergence of compact sets. Because I will not be using this result, I will not discuss it further.

Another strategy for the construction of transition probability functions, one that is more satisfactory when it works, is to see whether one can show that the sequence in Lemma 22 is itself convergent. To be more precise, for each $n \geq 0$, let $t \mapsto P_n(t, x)$ be constructed by the prescription in (20) with $\nu = \delta_x$. Then, as we showed, $\{P_n(\cdot, x) : n \geq 0\}$ is relatively compact in $C_b([0, \infty); M_1(\mathbb{R}^N))$ and every limit point is a solution to (15) with $\nu = \delta_x$. Now suppose that, for each $x$, $\{P_n(\cdot, x) : n \geq 0\}$ is convergent, and let $P(\cdot, x)$ be its limit. It would follow that $(t, x) \mapsto P(t, x)$ has to be a continuous transition probability function which, for each $x$, would solve (15). To check the Chapman–Kolmogorov equation, note that, by construction, for any $\phi \in C_b(\mathbb{R}^N; \mathbb{C})$, and $t \in [0, \infty)$,

$$\langle \phi, P_n(s + t, x) \rangle = \int \langle \phi, P_n(t, y) \rangle P_n(s, x, dy)$$

whenever $s = m2^{-n}$ for some $(m, n) \in \mathbb{N}^2$. Hence, after passing to the limit, one has

$$\langle \phi, P(s + t, x) \rangle = \int \langle \phi, P(t, y) \rangle P(s, x, dy)$$

whenever $s = m2^{-n}$ for some $(m, n) \in \mathbb{N}^2$, which, by continuity with respect to $s$, would lead immediately to (9).

The following theorem gives a condition under which one can prove that the required convergence takes place.

**Theorem 26.** Let $a : \mathbb{R}^N \to \text{Hom}(\mathbb{R}^N; \mathbb{R}^N)$ and $b : \mathbb{R}^N \to \mathbb{R}^N$ be given, and define $L$ accordingly, as in (12). Further, assume that there exists a square root
\[ \sigma : \mathbb{R}^N \rightarrow \text{Hom}(\mathbb{R}^M;\mathbb{R}^N) \text{ of a such that} \]
\[ \sup_{x \neq x'} \frac{\|\sigma(x) - \sigma(x')\|_{\text{H.S.}} \vee |b(x) - b(x')|}{|x - x'|} < \infty. \]

Then there exists a continuous transition probability function
\[ (t, x) \in [0, \infty) \times \mathbb{R}^N \mapsto P(t, x) \in \mathcal{M}_1(\mathbb{R}^N) \]
to which the sequence \( \{P_n : n \geq 0\} \) in the preceding discussion converges in \( C([0, \infty) \times \mathbb{R}^N; \mathcal{M}_1(\mathbb{R}^N)) \). In particular, \((t, x) \in [0, \infty) \times \mathbb{R}^N \mapsto P(t, x) \in \mathcal{M}_1(\mathbb{R}^N)\) is a transition probability function with the property that, for each \( x \in \mathbb{R}^N \), \( t \sim P(t, x) \) solves Kolmogorov’s forward equation for any \( \varphi \in C^2_b(\mathbb{R}^N; \mathbb{R}) \).

In order to prove this theorem, we must learn how to estimate the difference between \( P_{n_1}(t, x) \) and \( P_{n_2}(t, x) \) and show that this difference is small when \( n_1 \) and \( n_2 \) are large. The method which we will use to measure these differences is called **coupling**. That is, suppose that \((\mu_1, \mu_2) \in \mathcal{M}_1(\mathbb{R}^N)^2\). Then a coupling of \( \mu_1 \) to \( \mu_2 \) is any \( \tilde{\mu} \in \mathcal{M}_1(\mathbb{R}^N \times \mathbb{R}^N) \) for which \( \mu_1 \) and \( \mu_2 \) are its marginal distributions on \( \mathbb{R}^N \), in the sense that
\[ \mu_1(\Gamma) = \tilde{\mu}(\Gamma \times \mathbb{R}^N) \text{ and } \mu_2(\Gamma) = \tilde{\mu}(\mathbb{R}^N \times \Gamma). \]

Given a coupling of \( \mu_1 \) to \( \mu_2 \), an estimate of their difference is given by
\[ |\langle \varphi, \mu_2 \rangle - \langle \varphi, \mu_1 \rangle| \leq \text{Lip}(\varphi) \left( \int_{\mathbb{R}^N} |y - y'|^2 \tilde{\mu}(dy \times dy') \right)^{\frac{1}{2}}. \]

Equivalently, a coupling of \( \mu_1 \) to \( \mu_2 \) means that one has found a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) on which there are random variables \( X_1 \) and \( X_2 \) for which \( \mu_i \) is the distribution of \( X_i \). Of course, it is only when the choice of \( \tilde{\mu} \) is made judiciously that the method yields any information. For example, taking \( \tilde{\mu} = \mu_1 \times \mu_2 \) yields essentially no information.

To construct our coupling, take \( \mathbb{P} = (\gamma_{0,1})^\mathbb{Z}^+ \) on \((\mathbb{R}^M)^{\mathbb{Z}^+}\), and given \( x \in \mathbb{R}^N \) and \( n \geq 1 \), set \( Z_n(t, x, \omega) = \begin{pmatrix} X_n(t, x, \omega) \\ Y_n(t, x, \omega) \end{pmatrix} \), where \( X_n(0, x, \omega) = Y_n(0, x, \omega) = x \), and, for \( m \geq 0 \) and \( m2^{-n} < t \leq (m + 1)2^{-n} \),
\[ X_n(t, x, \omega) = X_n(m2^{-n}, x, \omega) + (t - m2^{-n})^{\frac{1}{2}} \sigma(X_n(m2^{-n})) \omega_{m+1} \]
\[ + (t - m2^{-n})b(X_n(m2^{-n})) \]
if \( m \) is even,
\[ X_n(t, x, \omega) = X_n(m2^{-n}, x, \omega) + (t - m2^{-n})^{\frac{1}{2}} \sigma(X_n((m - 1)2^{-n})) \omega_{m+1} \]
\[ + (t - m2^{-n})b(X_n((m - 1)2^{-n})) \]
if \( m \) is odd, and
\[
Y_n(t, x, \omega) = Y_n(m2^{-n}, x, \omega) + (t - m2^{-n})\frac{1}{2}\sigma(Y_n(m2^{-n}))\omega_{m+1} + (t - m2^{-n})b(Y_n(m2^{-n})).
\]

Clearly, \( P_n(t, x) \) is the distribution of \( Y_n(t, x) \) under \( \mathbb{P} \). To see that \( P_{n-1}(t, x) \) is the distribution of \( X_n(t, x) \), use the fact that, for odd \( m \) and \( m2^{-n} < t \leq (m+1)2^{-n} \), \( 2^{-n}\omega_{m-1} + (t - m2^{-n})\frac{1}{2}\omega_m \) has the same distribution under \( \mathbb{P} \) as \( (t - (m-1)2^{-n})\omega_m \). Thus the distribution of \( Z_n(t, x) \) under \( \mathbb{P} \) is a coupling of \( P_{n-1}(t, x) \) to \( P_n(t, x) \).

We will begin by showing that for each \( T > 0 \) there is a \( C(T) < \infty \) such that
\[
(28) \sup_{n \geq 0} \mathbb{E}[|Z_n(t, x) - Z_n(s, x)|^2] \leq C(T)(1 + |x|^2)(t-s) \quad \text{for } 0 \leq s < t \leq T.
\]

Set
\[
\Sigma_{m,n}(\omega) = \left( \frac{\sigma(X_n(m2^{-n}, x))}{\sigma(Y_n(m2^{-n}, x))} \right) \quad \text{and} \quad B_{m,n} = \left( \frac{b(X_n(m2^{-n}, x))}{b(Y_n(m2^{-n}, x))} \right)
\]
if \( m \) is even, and
\[
\Sigma_{m,n}(\omega) = \left( \frac{\sigma(X_n((m-1)2^{-n}, x))}{\sigma(Y_n(m2^{-n}, x))} \right) \quad \text{and} \quad B_{m,n} = \left( \frac{b(X_n((m-1)2^{-n}, x))}{b(Y_n(m2^{-n}, x))} \right)
\]
if \( m \) is odd. If \( m2^{-n} < t \leq (m+1)2^{-n} \), then
\[
|Z_n(t, x) - Z_n(m2^{-n}, x)|^2 \leq 2(t - m2^{-n})|\Sigma_{m,n}\omega_{m+1}|^2 + 2(t - m2^{-n})^2|B_{m,n}|^2,
\]
and therefore, since \( \omega_{m+1} \) is independent of \( \Sigma_{m,n} \),
\[
\mathbb{E}[|Z_n(t, x) - Z_n(m2^{-n}, x)|^2] \leq 2(t - m2^{-n})\mathbb{E}[|\Sigma_{m,n}|^2] + 2(t - m2^{-n})^2\mathbb{E}[|B_{m,n}|^2].
\]

Noting that (27) guarantees that (23) holds, and remembering that \( P_{n-1}(\tau, x) \) and \( P_n(\tau, x) \) are, respectively, the distributions of \( X_n(\tau, x) \) and \( Y_n(\tau, x) \), one can apply (16) to see that both the preceding expectation values are bounded by a constant \( C < \infty \) times \( e^{(1+\lambda)t}(1 + |x|^2) \). Hence,
\[
\mathbb{E}[|Z_n(t, x) - Z_n(m2^{-n}, x)|^2] \leq 4C(t - m2^{-n})e^{(1+\lambda)t}(1 + |x|^2).
\]

Next suppose that \( m_1 < m_2 \). Then
\[
|Z_n(m2^{-n}, x) - Z_n(m_12^{-n}, x)|^2 \leq 2^{-n+1} \left| \sum_{m=m_1}^{m_2-1} \Sigma_{m,n}\omega_{m+1} \right|^2 + 2^{-2n+1} \left| \sum_{m=m_1}^{m_2-1} B_{m,n} \right|^2.
\]
By taking advantage of the independence of $\Sigma_{m,n}$ and $\omega_m$ from $\omega_{m'}$ for $m' > m$, one sees that

$$
\mathbb{E}^P \left[ \left| \sum_{m=m_1}^{m_2-1} \Sigma_{m,n} \omega_{m+1} \right|^2 \right] = \sum_{m=m_1}^{m_2-1} \mathbb{E}^P[\|\Sigma_{m,n}\|_{H.S.}^2] \leq C(m_2-m_1)e^{(1+\lambda)m_2^{-n}}(1+|x|^2).
$$

At the same time,

$$
\mathbb{E}^P \left[ \left| \sum_{m=m_1}^{m_2-1} B_{m,n} \right|^2 \right] \leq (m_2-m_1) \sum_{m=m_1}^{m_2-1} \mathbb{E}^P[|B_{m,n}|^2] \leq C(m_2-m_1)^2e^{(1+\lambda)m_2^{-n}}(1+|x|^2).
$$

Hence,

$$
\mathbb{E}^P[|Z_n(m_2^{-n},x) - Z_n(m_1^{-n},x)|^2] \leq 2C((m_2-m_1)2^{-n} + (m_2-m_1)^22^{-2n})e^{(1+\lambda)m_2^{-n}}(1+|x|^2),
$$

and so, after one combines this with our earlier estimate, (28) follows.

Now set $\Delta_{m,n} = Y_n(m_2^{-n},x) - X_n(m_1^{-n},x)$. Then

$$
\Delta_{m+1,n} - \Delta_{m,n} = 2^{-\frac{n}{2}} A_{m,n} \omega_{m+1} + 2^{-n} D_{m,n},
$$

where

$$
A_{m,n} = \sigma(Y_n(m_2^{-n})) - \sigma(X_n(m_2^{-n})) \quad \text{and} \quad D_{m,n} = b(Y_n(m_2^{-n})) - b(X_n(m_2^{-n})).
$$

Hence

$$
\Delta_{m,n} = 2^{-\frac{n}{2}} \sum_{k=0}^{m-1} A_{k,n} \omega_{k+1} + 2^{-n} \sum_{k=0}^{m-1} D_{k,n},
$$

and so

$$
|\Delta_{m,n}|^2 \leq 2^{-n+1} \sum_{k=0}^{m-1} |A_{k,n} \omega_{k+1}|^2 + 2^{-2n+1} \sum_{k=0}^{m-1} |D_{k,n}|^2.
$$

Just as before,

$$
\mathbb{E}^P \left[ \sum_{k=0}^{m-1} A_{k,n} \omega_{k+1} \right]^2 = \sum_{k=0}^{m-1} \mathbb{E}^P[|A_{k,n}|^2_{H.S.}]
$$

and

$$
\mathbb{E}^P \left[ \sum_{k=0}^{m-1} D_{k,n} \right]^2 \leq m \sum_{k=0}^{m-1} \mathbb{E}^P[|D_{k,n}|^2].
$$
Thus,

$$E^P[|\Delta_{m,n}|^2] \leq 2^{-n+1}(1 + m2^{-n}) \sum_{k=0}^{m-1} E^P[\|A_{k,n}\|_{\mathcal{H},S} + |D_{k,n}|^2].$$

Using (27), one can find an $C < \infty$ such that

$$\|A_{k,n}\|_{\mathcal{H},S} + |D_{k,n}|^2 \leq C|\Delta_{k,n}|^2$$

when $k$ is even and

$$\|A_{k,n}\|_{\mathcal{H},S} + |D_{k,n}|^2 \leq C(|\Delta_{k,n}|^2 + |X_n(k2^{-n}, x) - X_n((k-1)2^{-n})|^2)$$

when $k$ is odd. Hence, by (16),

$$E^P[|\Delta_{m,n}|^2] \leq 2^{-n}K(m2^{-n})(1 + |x|^2 + 2^{-n}K(m2^{-n})e^{(1+\lambda)m2^{-n}} \sum_{k=0}^{m-1} E^P[|\Delta_{k,n}|^2],$$

where $K(T) = 2CC(T)(1 + T)$; and, using induction on $M$, one concludes from this that

$$E^P[|\Delta_{m,n}|^2] \leq 2^{-n}K(m2^{-n})(1 + |x|^2)e^{(1+\lambda)m2^{-n}} \left(1 + \frac{K(m2^{-n})}{2^n}\right)^{m-1}$$

$$\leq 2^{-n}K(m2^{-n})(1 + |x|^2)e^{(1+\lambda + K(m2^{-n}))m2^{-n}}.$$ 

Therefore, for any $R > 0$, there exists a $C(R) < \infty$ such that

$$E^P[|Y_n(m2^{-n}, x) - X_n(m2^{-n}, x)|^2]^{\frac{1}{2}} \leq C(R)2^{-\frac{m}{2}}$$

if $m2^{-n} \leq R$ and $|x| \leq R$, and, after combining this with (28), one has

(*) \[\sup_{t \vee |x| \leq R} E^P[|Y_n(t, x) - X_n(t, x)|^2]^{\frac{1}{2}} \leq C(R)2^{-\frac{m}{2}}.\]

for some $C(R) < \infty$. Given (*) and the fact that, for each $x$, $P(\cdot, x)$ is a limit point of $\{P_n(\cdot, x) : n \geq 1\}$ in $C([0, \infty); \mathcal{M}_1(\mathbb{R}^N))$, it follows that

$$\sup_{t \vee |x| \leq R} \left| \langle \varphi, P(t, x) \rangle - \langle \varphi, P_n(t, x) \rangle \right| \leq 2C(R)||\varphi||_{\text{Lip}}2^{-\frac{m}{2}},$$

and from this it is an easy step to the conclusion that

$$P_n \rightarrow P$$

in $C([0, \infty) \times \mathbb{R}^N, \mathcal{M}_1(\mathbb{R}^N)).$

Now that we have a transition probability function for which Kolmogorov’s forward equation holds, one might expect that we can show that his backwards equation holds. However that would require our knowing that $x \mapsto P_t(\varphi)(x)$ is differentiable, and in general it is not.
A Digression on Square Roots: The preceding brings up an interesting question. Namely, knowing \( L \) means that one knows its coefficients \( a \) and \( b \). On the other hand, \( a \) does not uniquely a \( \sigma \) for which \( a = \sigma \sigma^\top \), and the choice of \( \sigma \) can be critical since one would like \( \sigma \) to have the same smoothness as \( a \). From that point of view, when \( L \) is uniformly elliptic (i.e., \( a \geq \epsilon I \) for some \( \epsilon > 0 \)), the following lemma shows that it is hard to do better than take \( \sigma \) to be its positive definite square root \( a^\frac{1}{2} \). In its statement and proof, I use the notation

\[
\| \alpha \| = \sum_{i=1}^{N} \alpha_i \quad \text{and} \quad \partial^\alpha \varphi = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_N}^{\alpha_N} \varphi
\]

for \( \alpha \in \mathbb{N}^N \).

**Lemma 29.** Assume that \( a \geq \epsilon I \). If \( a \) is continuously differentiable in a neighborhood of \( x \), then so is \( a^\frac{1}{2} \) and

\[
\max_{1 \leq i \leq n} \left\| \partial_{x_i} a^\frac{1}{2} (x) \right\|_{op} \leq \frac{\left\| \partial_{x_i} a (x) \right\|_{op}}{2\epsilon^\frac{1}{2}}.
\]

Moreover, if \( n \geq 1 \), then there is a \( C_n < \infty \) such that

\[
\max_{\| \alpha \| = n} \left\| \partial^\alpha a (x) \right\|_{op} \leq C_n \epsilon^\frac{1}{2} \sum_{k=1}^{n-1} \left( \frac{\max_{\| \alpha \| \leq n} \left\| \partial^\alpha a (x) \right\|_{op}}{\epsilon} \right)^k
\]

when \( a \) is \( n \)-times continuously differentiable in a neighborhood of \( x \). Hence, if \( a \in C^n_b(\mathbb{R}^N; \text{Hom}(\mathbb{R}^N; \mathbb{R}^N)) \), then so is \( a^\frac{1}{2} \).

**Proof:** Without loss in generality, assume that \( x = 0 \) and that there is a \( \Lambda < \infty \) such that \( a \leq \Lambda I \) on \( \mathbb{R}^N \).

Set \( d = I - \frac{\Lambda}{\lambda} \). Obviously \( d \) is symmetric, \( 0I \leq d \leq (1 - \frac{\Lambda}{\lambda})I \), and \( a = \Lambda (I - d) \).

Thus, if \( \left( \frac{1}{m} \right) = 1 \) and

\[
\left( \frac{1}{m} \right) = \prod_{k=1}^{m-1} \left( \frac{1}{k} - \frac{1}{m} \right)
\]

are the coefficients in the Taylor expansion of \( x \sim (1 + x)^\frac{1}{2} \) around 0, then

\[
\sum_{m=0}^{\infty} (-1)^m \left( \frac{1}{m} \right) d^m
\]

converges in the operator norm uniformly on \( \mathbb{R}^N \). In addition, if \( \lambda \) is an eigenvalue of \( a(y) \) and \( \xi \) is an associated eigenvector, then \( d(y)\xi = (1 - \frac{\Lambda}{\lambda})\xi \), and so

\[
\left( \Lambda^\frac{1}{2} \sum_{m=0}^{\infty} (-1)^m \left( \frac{1}{m} \right) d^m (y) \right) \xi = \lambda^\frac{1}{2} \xi.
\]
Hence,
\begin{equation}
(*)
a_\frac{1}{2} = \Lambda_\frac{1}{2} \sum_{m=0}^{\infty} (-1)^m \left(\frac{1}{m}\right) d^m.
\end{equation}

Now assume that \(a\) is continuously differentiable in a neighborhood of 0. Then it is easy to see from (*) that
\[
\partial_{\alpha_i} a_\frac{1}{2}(0) = -\Lambda_\frac{1}{2} \left( \sum_{m=1}^{\infty} m(-1)^m \left(\frac{1}{m}\right) d^m(0) \right) \partial_{\alpha_i} a(0),
\]
where again the series converges in the operator norm. Furthermore, because \((-1)^m \left(\frac{1}{m}\right) \geq 0\) for all \(m \geq 0\) and \(d(0) \leq 1 - \frac{\epsilon}{\Lambda}\),
\[
\left\| \sum_{m=1}^{\infty} m(-1)^m \left(\frac{1}{m}\right) d^{m-1}(0) \right\|_{op} < \frac{1}{2} (1 - \|d(0)\|_{op})^{-\frac{1}{2}} = \frac{\Lambda_\frac{1}{2}}{2\epsilon^{\frac{1}{2}}},
\]
and so the first assertion is now proved.

Turning to the second assertion, one again uses (*) to see that if \(\|\alpha\| = n\) then
\[
\partial^\alpha a(0) = \sum_{k=1}^{n} (-1)^k \Lambda_\frac{1}{2}^{k-1} \left( \sum_{m=k}^{\infty} \frac{m!}{(m-k)!} (-1)^m \left(\frac{1}{m}\right) d^{m-k}(0) \right) \times \left( \sum_{\alpha_1 + \ldots + \alpha_k = \alpha} \partial^{\alpha_1} a(0) \cdots \partial^{\alpha_k} a(0) \right).
\]
Proceeding as above, one see that
\[
\left\| \sum_{m=k}^{\infty} \frac{m!}{(m-k)!} (-1)^m \left(\frac{1}{m}\right) d^{m-k}(0) \right\|_{op} \leq \left( \frac{\Lambda}{\epsilon} \right)^{k-\frac{1}{2}},
\]
and from here it is easy to complete the proof. \(\square\)

In view of Lemma 29, what remains to examine are \(a\)'s that can degenerate. In this case, the \(a_\frac{1}{2}\) will often not be the optimal choice of \(\sigma\). For example, if \(N = 1\) and \(a(x) = x^2\), then \(a_\frac{1}{2}(x) = |x|\), which is Lipschitz continuous but not continuously differentiable, and so it is obviously that \(\sigma(x) = x\) is a preferable choice. Another example of the same sort is
\[
a(x) = \begin{pmatrix} 1 & 0 \\ 0 & |x|^2 \end{pmatrix} \text{ for } x \in \mathbb{R}^2.
\]
Again \( a^{\frac{1}{2}} \) is Lipschitz continuous but not differentiable. On the other hand, if 
\[
\sigma(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & x_1 & x_2 \end{pmatrix},
\]
then \( a = \sigma \sigma^\top \) and \( \sigma \) is smooth. However, it can be shown that in general there is no smooth choice of \( \sigma \) even when \( a \) is smooth. The reason why stems from a result of D. Hilbert in classical algebraic geometry. Specifically, he showed that there are non-negative polynomials that cannot be written as the sum of squares of polynomials. By applying his result to the Taylor’s series for \( a \), one can show that it rules out the possibility of always being able to find a smooth \( \sigma \). Nonetheless, the following lemma shows that if all that one wants is Lipschitz continuity, then it suffices to know that \( a \) has two continuous derivatives.

**Lemma 30.** Assume that \( a \) has two continuous derivatives, and let \( K < \infty \) be a bound on the operator norm of its second derivatives. Then 
\[
\|a^{\frac{1}{2}}(y) - a^{\frac{1}{2}}(x)\|_{\text{H.S.}} \leq N\sqrt{2^{-1}K}|y - x|.
\]

**Proof:** The proof turns on a simple fact about functions \( f : \mathbb{R} \rightarrow [0, \infty) \) that have two continuous derivatives and whose second derivative is bounded. Namely,
\[
(*) \quad |f'(0)| \leq \sqrt{2\|f''\|_u f(0)}.
\]
To prove this simply use Taylor’s theorem to write
\[
0 \leq f(h) \leq f(0) + hf'(0) + \frac{h^2}{2}\|f''\|_u
\]
and minimize the right hand side with respect to \( h \).

Turning to the stated result, first observe that it suffices to prove it when \( a \) is uniformly positive definite, since, if this is not already the case, we can replace \( s \) by \( a + \epsilon I \) and then let \( \epsilon \searrow 0 \). Assuming this uniform positivity, we know that \( a^{\frac{1}{2}} \) has two continuous derivatives, and we need to show is that \( |\partial_x a_{ij}^{\frac{1}{2}}| \leq \sqrt{2^{-1}K} \).
For this purpose, let \( x \) be given, and, without loss in generality, assume that \( a(x) \) is diagonal. Then, because \( a = a^{\frac{1}{2}} a^{\frac{1}{2}} \),
\[
\partial_x a_{ij}(x) = \partial_x a_{ij}^{\frac{1}{2}}(x)(\sqrt{a_{ii}(x)} + \sqrt{a_{jj}(x)}) \geq \partial_x a_{ij}^{\frac{1}{2}}(x)\sqrt{a_{ii}(x)} + a_{jj}(x),
\]
and so
\[
|\partial_x a_{ij}^{\frac{1}{2}}(x)| \leq \frac{|\partial_x a_{ij}(x)|}{\sqrt{a_{ii}(x)} + a_{jj}(x)}.
\]
When \( i = j \), apply (*) to \( f(h) = a(x + he_k) \), and conclude that
\[
\left| \partial_{x_k} a_i^1 (x) \right| \leq \sqrt{2} K a_i(x),
\]
which means that
\[
\left| \partial_{x_k} a_i^2 (x) \right| \leq \sqrt{2^{-1}} K a_i(x).
\]
When \( i \neq j \), set
\[
f_{\pm} (h) = a_{ii}(x + he_k) \pm 2a_{ij}(x + he_k) + a_{jj}(x + he_k).
\]
Then, by (*),
\[
\left| \partial_{x_k} a_{ij} (x) \right| \leq \frac{|f'_+ (0)| + |f'_- (0)|}{4} \leq \sqrt{2^{-1}} K (a_{ii}(x) + a_{jj}(x)),
\]
and so
\[
\left| \partial_{x_k} a_{ij}^2 (x) \right| \leq \sqrt{2^{-1}} K. \quad \square
\]

**Euler’s Approach to Itô Integral Equations**

Let \( W \) denote Wiener measure on \( C([0, \infty); \mathbb{R}^M) \) and set \( W_t = \sigma\{ w(\tau) : \tau \in [0, t] \} \) for \( t \geq 0 \).

Given \( \sigma : \mathbb{R}^N \to \text{Hom}(\mathbb{R}^M; \mathbb{R}^N) \) and \( b : \mathbb{R}^N \to \mathbb{R}^N \) satisfying (27), define \( X_n(t, x) \) for \( n \geq 0 \) and \( (t, x) \in [0, \infty) \times \mathbb{R}^N \) so that \( X_n(0, x) = x \) and
\[
X_n(t, x) = X_n([t], x) + \sigma(X_n([t], x))(w(t) - w([t])) + b(X_n([t], x))(t - [t]).
\]
Then, for each \( (t, x) \in [0, \infty) \times \mathbb{R}^N \) and \( n \geq 0 \), the distribution of \( X_n(t, x) \) under \( W \) is the measure \( P_n(t, x) \) described in the discussion preceding Theorem 26. In particular, for each \( T > 0 \),
\[
\sup_{n \geq 0} \sup_{(t, x) \in [0, T] \times \mathbb{R}^N} \frac{\mathbb{E}^W[|X_n(t, x)|^2]}{1 + |x|^2} < \infty
\]
for all \( T > 0 \). In addition, it is clear that \( X_n(t, x) \) is \( W_t \)-measurable.

Set
\[
I_n(t, x) = \sum_{m < 2^n t} \sigma(X_n(m2^{-n}, x))(w((m + 1)2^{-n} \wedge t) - w(m2^{-n})
\]
and
\[
B_n(t, x) = \sum_{m < 2^n t} b(X_n(m2^{-n}, x))((m + 1)2^{-n} \wedge t) - (m2^{-n}).
\]
Note that if $m2^{-n} \leq s < t \leq (m + 1)2^{-n}$, then, because $w(t) - w(s)$ is independent of $W_s$, 

$$E^W[I_n(t, x) - I_n(s, x) \mid W_s] = \sigma(X_n(m2^{-n}, x))E^W[w(t) - w(s) \mid W_s] = 0,$$

and therefore that $(I_n(t, x), W_t, \mathcal{W})$ is a continuous, square integrable martingale. Similarly, if $m2^{-n} \leq s < t \leq (m + 1)2^{-n}$, for any $\xi \in \mathbb{R}^N$, 

$$E^W[(\xi, I_n(t, x))^2_{\mathbb{R}^N} - (\xi, I_n(s, x))^2_{\mathbb{R}^N} \mid W_s]$$

$$= E^W\left(\left(\left(\xi, I_n(t, x)\right)_{\mathbb{R}^N} - (\xi, I_n(x, t))_{\mathbb{R}^N}\right)^2 \mid W_s\right)$$

$$= \sum_{i,j=1}^N (\sigma(X_n(m2^{-n}, x))^\top \xi)_i(\sigma(X_n(m2^{-n}, x))^\top \xi)_j E^W[(w(t)_i - w(s)_i)(w(t)_j - w(s)_j) \mid W_s]$$

$$= (\xi, a(X(m2^{-n}, x))\xi)_{\mathbb{R}^N}(t - s).$$

Thus,

$$\left((\xi, I_n(t, x))^2_{\mathbb{R}^N} - \int_0^t (\xi, a(X_n([\tau], x)))_{\mathbb{R}^N} d\tau, W_t, \mathcal{W}\right)$$

is a martingale, and so

$$E^W[I_n(t, x) - I_n(s, x)]^2 = E^W[I_n(t, x)]^2 - |I_n(s, x)]^2$$

$$= E^W\left[\int_s^t \text{Trace}(a(X_n([\tau], x))) \, d\tau\right].$$

At the same time, $B_n(t, x) = \int_s^t b([\tau], x) \, d\tau$, and so

$$E^W[B_n(t, x) - B_n(s, x)]^2 \leq (t - s) \int_s^t E^W[|b(X_n([\tau], x))|^2] \, d\tau.$$ 

Hence, for each $T > 0$,

$$K(T) \equiv \sup_{n \geq 0} \sup_{0 \leq s \leq t \leq T} \frac{E^W[|X_n(t, x) - X_n(s, x)]^2}{(1 + |x|^2)(t - s)} < \infty.$$ 

Obviously,

$$|X_n(t, x) - X_{n-1}(t, x)]^2 \leq 2|I_n(t, x) - I_{n-1}(t, x)]^2 + 2|B_n(t, x) - B_{n-1}(t, x)]^2.$$ 

Using the same line of reasoning as above, one sees that

$$\left(|I_n(t, x) - I_{n-1}(t, x)]^2 - \int_0^t \|\sigma(X_n([\tau])) - \sigma(X_{n-1}([\tau]))\|^2_{\text{H.S.}} \, d\tau, W_t, \mathcal{W}\right)$$
is a continuous martingate. Hence, since \((I_n(t, x) - I_{n-1}(t, x), W_t, W)\) is a continuous martingale, Doob’s inequality says that

\[
\mathbb{E}^W[|I_n(\cdot, x) - I_{n-1}(\cdot, x)|^2_{[0,T]}] \leq 4\mathbb{E}^W[|I_n(T, x) - I_{n-1}(T, x)|^2]
\]

\[
= 4 \int_0^T \mathbb{E}^W[||\sigma(X_n([T], x)) - \sigma(X_{n-1}([T], x))||_{H.S.}^2] d\tau
\]

\[
\leq 8 \int_0^T \mathbb{E}^W[||\sigma(X_n([T], x)) - \sigma(X_{n-1}([T], x))||_{H.S.}^2] d\tau
\]

\[
+ 8 \int_0^T \mathbb{E}^W[||\sigma(X_n([T], x)) - \sigma(X_{n-1}([T], x))||_{H.S.}^2] d\tau.
\]

At the same time,

\[
|B_n(t, x) - B_{n-1}(t, x)|^2 \leq t \int_0^t |b(X_n([T], x)) - b(X_{n-1}([T], x))|^2 d\tau
\]

\[
\leq 2t \left( \int_0^t |b(X_n([T], x)) - b(X_{n-1}([T], x))|^2 d\tau
\]

\[
+ \int_0^t |b(X_n([T], x)) - b(X_{n-1}([T], x))|^2 d\tau \right).
\]

After combining these with (*) and using the Lipschitz continuity of \(\sigma\) and \(b\), one sees that there is a \(\lambda < \infty\) and, for each \(T > 0\), a \(C(T) < \infty\) such that

\[
\mathbb{E}^W[||X_n(\cdot, x) - X_{n-1}(\cdot, x)||_{[0,T]}^2]
\]

\[
\leq C(T)(1 + |x|^2) 2^{-n} + \lambda \int_0^t \mathbb{E}^W[||X_n(\cdot, x) - X_{n-1}(\cdot, x)||_{[0,T]}^2] d\tau
\]

when \(t \in [0, T]\). Hence, by Gronwall’s lemma,

\[
\mathbb{E}^W[||X_n(\cdot, x) - X_{n-1}(\cdot, x)||_{[0,T]}^2] \leq 2^{-n}C(T)(1 + |x|^2)e^{\lambda T}.
\]

Starting from the preceding, one sees that, for any \(m \geq 0\),

\[
\mathbb{E}^W[\sup_{n > m} ||X_n(\cdot, x) - X_m(\cdot, x)||_{[0,T]}^2] \leq 2^{-m+2}C(T)e^{\lambda T}(1 + |x|^2),
\]

and therefore, for each \(x \in \mathbb{R}^N\), there is a measurable map

\[
w \in C([0, \infty); \mathbb{R}^M) \rightarrow X(\cdot, x) \in C([0, \infty); \mathbb{R}^N)
\]

\footnote{If \(\psi : [0, \infty) \rightarrow \mathbb{R}^N\), then \(\|\psi\|_{[s, t]} = \sup\{|\psi(\tau)| : \tau \in [s, t]\}\).}
and another constant $C(T) < \infty$ such that

$$E^W[\|X(\cdot, x) - X_n(\cdot, x)\|_{[0,T]}^2] \leq 2^{-n}C(T)(1 + |x|^2).$$

**Maringales Everywhere:** Since, $X_n(t,x)$ and $I_n(t,x)$ are $W_t=\sigma(\{w(\tau) : \tau \in [0, t]\})$-measurable,

$$X(t,x), B(t,x) = \int_0^t b(X(\tau,x)) \, d\tau, \text{ and } I(t,x) = X(t,x) - x - B(t,x)$$

are $W_t$-measurable. In addition, because $(I_n(t,x), W_t)$ and, for each $\xi \in \mathbb{R}^N$,

$$\left( (\xi, X_n(t,x))^2 - \int_0^t (\xi, a(X_n(\tau,x))\xi) \, d\tau, W_t, W \right)$$

are martingales for all $n \geq 0$, it follows from (31) that $(I(t,x), W_t)$ is also a continuous, square integrable martingale and, for each $\xi \in \mathbb{R}^N$, that

$$\left( (\xi, I(t,x))^2 - \int_0^t (\xi, a(X(\tau,x))\xi) \, d\tau, W_t, W \right)$$

is a martingale. Using the polarization identity

$$4(\xi, \eta)_{\mathbb{R}^N} = (\xi + \eta, \xi + \eta)_{\mathbb{R}^N} - (\xi - \eta, \xi - \eta)_{\mathbb{R}^N},$$

one sees that

$$\left( (\xi, I(t,x))_{\mathbb{R}^N}(\eta, X(t,\eta))_{\mathbb{R}^N} - \int_0^t (\xi, a(X(\tau,x))\eta) \, d\tau, W_t, W \right)$$

is a martingale for all $\xi, \eta \in \mathbb{R}^N$. Hence we have now proved that

$$\left( I(t,x), W_t, W \right)$$

and

$$\left( I(t,x) \otimes I(t,x) - \int_0^t a(X(\tau,x)) \, d\tau, W_t, W \right)$$

are martingales.

When $\|a\|_u \equiv \|\text{Trace}(a)\|_u$ is finite, there is another important family of martingales associated with $I$. Obviously,

$$|I_n(t,x)| \leq \|a\|_u^2 \sum_{m<2^n t} |w(t \wedge (m+1)2^{-n}) - w(m2^{-n})|,$$

I use $W_t$ here to denote the completion of $W_t$ with respect to $W$. 

---

5 I use $W_t$ here to denote the completion of $W_t$ with respect to $W$. 

---
and so $\mathbb{E}^W[e^{\lambda I_n(t,x)}] < \infty$ for all $\lambda > 0$ if $\|a\|_u < \infty$. Next, given $n \geq 0$ and $\zeta \in \mathbb{C}^N$, set

$$A_n(t,x) = \int_0^t a(X([\tau], x)) \, d\tau$$

and $E_{\zeta,n}(t,x) = \exp \left( (\zeta, I_n(t,x)) \mathbb{R}^N - \frac{1}{2} (\zeta, A_n(t,x) \zeta) \mathbb{R}^N \right)$.

Our first goal is to show that $(E_{\zeta,n}(t,x), W_t, \mathcal{W})$ is a martingale when $\|a\|_u < \infty$, and the argument is very similar to the one given to prove the martingale property for $I_n(t,x)$. Namely, if $m2^{-n} \leq s < t \leq (m+1)2^{-n}$, remember that $w(t) - w(s)$ is independent of $W_s$, and conclude that

$$\mathbb{E}^W[E_{\zeta,n}(t,x) \mid W_s] = E_{\zeta,n}(s,x) \mathbb{E}^W[\exp \left( (\zeta, \sigma(X_n(m2^{-n}, x))(w(t) - w(s)) \right) \mathbb{R}^N$$

$$- \frac{1}{2} (\zeta, a(X(m2^{-n}, x)) \zeta) \mathbb{R}^N \mid W_s]$$

$$= E_{\zeta,n}(s,x),$$

from which the asserted martingale property follows immediately.

As a consequence of the preceding, we know that

$$\mathbb{E}^W[e^{\xi I_n(t,x)}]_{\mathbb{R}^N} \leq 1$$

for all $\xi \in \mathbb{R}^N$, and therefore

$$(1 + t\|a\|_u)^{-\frac{N}{2}} \mathbb{E}^W[e^{\frac{(I_n(t,x))^2}{2(1+t\|a\|_u)^2}}] = \int_{\mathbb{R}^N} \mathbb{E}^W[e^{\xi I_n(t,x)}]_{\mathbb{R}^N} - \frac{t\|a\|_u}{2} |\xi|^2 \gamma_{0,1}(d\xi) \leq 1.$$

Hence

$$\mathbb{E}^W[e^{\frac{(I_n(t,x))^2}{2(1+t\|a\|_u)^2}}] \leq (1 + t\|a\|_u)^{-\frac{N}{2}}.$$

Knowing this, and defining

$$B(t,x) = \int_0^t b(X(\tau,x)) \, d\tau$$

and $I(t,x) = X(t,x) - x - B(t,x)$, it follows from Fatou’s Lemma that

$$(33) \quad \mathbb{E}^W[e^{\frac{(I(t,x))^2}{2(1+t\|a\|_u)^2}}] \leq (1 + t\|a\|_u)^{-\frac{N}{2}}.$$

In addition, if

$$A(t,x) = \int_0^t a(X(\tau,x)) \, d\tau$$
and
\[ E_{\zeta}(t, x) = \exp\left(\left(\zeta, I(t, x)\right)_{RN} - \frac{1}{2} \left(\zeta, A(t, x)\zeta\right)_{RN}\right), \]
then \( \{E_{\zeta,n}(t, x) : n \geq 0\} \) converges to \( E_{\zeta}(t, x) \) in \( L^1(W; C) \), and so
\[ (E_{\zeta}(t, x), \mathcal{W}_t, \mathcal{W}) \]
is a martingale for each \( \zeta \in C^N \).

Continuing under the assumption that \( \|a\|_u < \infty \), we will now show that
\[ (35) \]
is a martingale for all \( \varphi \in C^2(\mathbb{R}^N; C) \) for which there is a \( \lambda \geq 0 \) such that
\[ \sup_{y \in \mathbb{R}^N} e^{-\lambda|y|} \|\nabla^2 \varphi(y)\|_{H.S.} < \infty, \]
and, under the additional that assumption that \( \|b\|_u < \infty \), that
\[ (36) \]
is a martingale for all \( \varphi \in C^2(\mathbb{R}^N; C) \) for which there is a \( \lambda \geq 0 \) such that
\[ \sup_{y \in \mathbb{R}^N} e^{-\lambda|y|} \|\nabla^2 \varphi(y)\|_{H.S.} < \infty, \]
which is the pathspace statement that the distribution of \( X(\cdot, x) \) satisfies Kolmogorov’s forward equation.

To this end, begin by observing that, given (33), standard approximation arguments show that it suffices to handle \( \varphi \in \mathcal{S}(\mathbb{R}^N; C) \). Furthermore, if \( \varphi \in \mathcal{S}(\mathbb{R}^N; C) \), then
\[ (2\pi)^N \left( \varphi(X(t, x)) - \int_0^t L\varphi(X(\tau, x)) \, d\tau, \mathcal{W}_t, \mathcal{W} \right) \]
is a martingale for all \( \varphi \in C^2(\mathbb{R}^N; C) \) for which there is a \( \lambda \geq 0 \) such that
\[ \sup_{y \in \mathbb{R}^N} e^{-\lambda|y|} \|\nabla^2 \varphi(y)\|_{H.S.} < \infty, \]
and so it suffices to prove (36) when \( \varphi(y) = e^{(\zeta, y)_{RN}} \). The same line of reasoning shows that it suffices to prove (35) for \( e^{(\zeta, y)_{RN}} \), and both of these are easy applications of the following lemma.
Lemma 37. Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and \(\{\mathcal{F}_t : t \geq 0\}\) a non-decreasing family of sub-\(\sigma\)-algebras. Suppose that \((M(t), \mathcal{F}_t, \mathbb{P})\) is a continuous \(\mathbb{C}\)-valued martingale and that \(\{V(t) : t \geq 0\}\) is a family of \(\mathbb{C}\)-valued random variables such that \(V(t)\) is \(\mathcal{F}_t\)-measurable for each \(t \geq 0\) and \(t \mapsto V(t)(\omega)\) is a continuous function that has bounded variation \(|V(T)(\omega)|\) on \([0, T]\) for each \(T > 0\). If

\[
\|M(\cdot)\|_{[0,T]}(|V(0)| + |V|(T)) \in L^1(\mathbb{P}; \mathbb{R}) \text{ for all } T > 0,
\]

then

\[
\left( M(t)V(t) - \int_0^t M(\tau)dV(\tau), \mathcal{F}_t, \mathbb{P} \right)
\]

is a martingale.

Proof: Given \(0 \leq s < t\) and \(\Gamma \in \mathcal{F}_s\), set \(\tau_{m,n} = s + \frac{m}{n}(t-s)\) and write

\[
\mathbb{E}^\mathbb{P}[M(t)V(t) - M(s)V(s), \Gamma]
\]

\[
= \sum_{m=0}^{n-1} \mathbb{E}^\mathbb{P}[M(\tau_{m+1,n})V(\tau_{m+1,n}) - M(\tau_{m,n})V(\tau_{m,n}), \Gamma]
\]

\[
= \mathbb{E}^\mathbb{P}\left[ \sum_{m=0}^{n-1} M(\tau_{m+1,n})(V(\tau_{m+1,n}) - V(\tau_{m,n})), \Gamma \right],
\]

where I have used the martingale property of \(M\) in passing to the second line. Now observe that

\[
\sum_{m=0}^{n-1} M(\tau_{m+1,n})(V(\tau_{m+1,n}) - V(\tau_{m,n})) \longrightarrow \int_s^t M(\tau)\,dV(\tau) \text{ in } L^1(\mathbb{P}; \mathbb{C}),
\]

and therefore that

\[
\mathbb{E}^\mathbb{P}[M(t)V(t) - M(s)V(s), \Gamma] = \mathbb{E}^\mathbb{P}\left[ \int_s^t M(\tau)\,dV(\tau), \Gamma \right]. \quad \square
\]

Given Lemma 37 and the remarks preceding it, the proofs of (35) and (36) come down to writing \(e^{(\zeta,t(t,x))_{\mathbb{R}^N}}\) as

\[
E_{\zeta}(t, x) \exp\left( \int_0^t \frac{1}{2}(\zeta, a(X(\tau, x))\zeta)_{\mathbb{R}^N} \,d\tau \right)
\]

and \(e^{(\zeta,X(t,x))_{\mathbb{R}^N}}\) as

\[
e^{(\zeta,x)_{\mathbb{R}^N}} E_{\zeta}(t, x) \exp\left( \int_0^t \left( (\zeta, b(X(t, x))_{\mathbb{R}^N} + \frac{1}{2}(\zeta, a(X(\tau, x))\zeta)_{\mathbb{R}^N} \right) \,d\tau \right).
\]

Some Applications of these Martingales: There are a lot of applications of the preceding results, the first of which is the following localization result.
Lemma 38. Suppose that \( \bar{\sigma} : \mathbb{R}^N \to \text{Hom}(R^M;\mathbb{R}^N) \) and \( \bar{b} : \mathbb{R}^N \to \mathbb{R}^N \) are another pair of Lipschitz continuous function, and define \( \tilde{X}(t, x) \) accordingly. If \( \bar{\sigma} = \sigma \) and \( \bar{b} = b \) on an open set \( G \ni x \) and \( \zeta = \inf\{t \geq 0 : X(t, x) \notin G\} \), then

\[
W(X(t \wedge \zeta, x) = \tilde{X}(t \wedge \zeta, x) \text{ for } t \geq 0) = 1.
\]

Proof: Set

\[
\bar{I}(t, x) = \tilde{X}(t, x) - x - \int_0^t \bar{b}(\tilde{X}(\tau, x)) \, d\tau.
\]

By applying the second part of (32) when \( \sigma \) and \( b \) are replaced by \( \left(\frac{\sigma}{\bar{\sigma}}\right) \) and \( \left(\frac{b}{\bar{b}}\right) \), one sees that

\[
\left( \left| \bar{I}(t, x) - \bar{I}(t, x) \right|^2 - \int_0^t \|\sigma(X(\tau, x)) - \bar{\sigma}(\tilde{X}(\tau, x))\|_{\text{H.S.}}^2 \, d\tau, W_t, W \right)
\]

is a martingale. Thus, by Doob’s stopping time theorem,

\[
\mathbb{E}^W[\left| \bar{I}(t \wedge \zeta) - \bar{I}(t \wedge \zeta, x) \right|^2]
\]

\[
= \mathbb{E}^W\left[ \int_0^{t \wedge \zeta} \|\bar{\sigma}(X(\tau, x)) - \bar{\sigma}(\tilde{X}(\tau, x))\|_{\text{H.S.}}^2 \, d\tau \right]
\]

\[
\leq \int_0^t \mathbb{E}^W[\|\bar{\sigma}(X(\tau \wedge \zeta, x)) - \bar{\sigma}(\tilde{X}(\tau \wedge \zeta, x))\|_{\text{H.S.}}^2 \, d\tau
\]

\[
\leq C \int_0^t \mathbb{E}^W[|X(\tau \wedge \zeta, x) - \tilde{X}(\tau \wedge \zeta, x)|^2] \, d\tau,
\]

where \( C \) is the square of the Lipschitz constant for \( \bar{\sigma} \). At the same time,

\[
|X(t \wedge \zeta, x) - \tilde{X}(t \wedge \zeta, x)|^2 \leq 2|\bar{I}(t \wedge \zeta, x) - \bar{I}(t \wedge \zeta, x)|^2
\]

\[
+ 2 \int_0^{t \wedge \zeta} |\bar{b}(X(\tau, x)) - \bar{b}(\tilde{X}(\tau, x))| \, d\tau,
\]

and

\[
\int_0^{t \wedge \zeta} |\bar{b}(X(\tau, x)) - b(\tilde{X}(\tau, x))| \, d\tau \leq Ct \int_0^t |X(\tau \wedge \zeta) - \tilde{X}(\tau \wedge \zeta)|^2 \, d\tau,
\]

where this time \( C \) is the square of the Lipschitz constant for \( \bar{b} \). Hence, after combining these, we see that there is a \( C < \infty \) such that

\[
\mathbb{E}^W[|X(t \wedge \zeta, x) - \tilde{X}(t \wedge \zeta, x)|^2] \leq C(1 + t) \int_0^t \mathbb{E}^W[|X(t \wedge \zeta, x) - \tilde{X}(t \wedge \zeta, x)|^2] \, d\tau.
\]
Since, by Gronwall’s lemma, this means that
\[ \mathbb{E}^W [ |X(t \wedge \zeta, x) - \tilde{X}(t \wedge \zeta, x)|^2] = 0, \]
it follows that \( X(t \wedge \zeta, x) = \tilde{X}(t \wedge \zeta, x) \) (a.s., \( W \)), first for each and then, by
continuity, simultaneously for all \( t \geq 0 \). □

Lemma 38 will allow us to reduce proofs of many results to the case when \( \sigma \) and \( b \) are bounded.

**Lemma 39.** Given \( p \in [2, \infty) \), set
\[
\lambda_p = \sup_{y \in \mathbb{R}^N} \frac{p(p-1)}{2} \text{Trace}(a(y)) + (y, b(y))_{\mathbb{R}^N}^+, \\
\lambda_0^p = \sup_{y \in \mathbb{R}^N} \frac{p(p-1)}{2} \text{Trace}(a(y))_{\mathbb{R}^N}^+, \\
\text{and} \\
\beta_p = \sup_{y \in \mathbb{R}^N} \frac{|b(y)|^p}{(1 + |y|^p)}. 
\]
Then there exists a constant \( C_p < \infty \), depending only on \( p, \lambda_0^p, \) and \( \beta_p \), such that
\[
\mathbb{E}^W [ \|I(\cdot, x) - I(s, x)\|_{[s, t]}^p]^{\frac{1}{p}} \leq C_p e^{\frac{\lambda_{\mathbb{R}^N}}{p}(1 + |x|)(t - s)^{\frac{1}{2}}} 
\]
and
\[
\mathbb{E}^W [ \|X(\cdot, x) - X(s, x)\|_{[s, t]}^p]^{\frac{1}{p}} \leq C_p e^{\frac{\lambda_{\mathbb{R}^N}}{p}(1 + |x|)(t - s)^{\frac{1}{2}} \vee (t - s)}
\]
for \( 0 \leq s < t \) and \( x \in \mathbb{R}^N \).

**Proof:** For the present, assume that \( a \) and \( b \) are bounded and therefore that (35) and (36) apply.

Begin with the observation that, if \( \varphi_p(y) = |y|^p \)
\[
\nabla \varphi_p(y) = p|y|^{p-2} y \text{ and } H_p(y) \equiv \nabla^2 \varphi_p(y) = |y|^{p-2} \left( p(p-2) \frac{y \otimes y}{|y|^2} + pI \right). 
\]

Hence
\[
L \varphi_p = |y|^{p-2} \left( \frac{p(p-1)(y, a(y)y)_{\mathbb{R}^N}}{2|y|^2} + \frac{p \text{Trace}(a(y))}{2} + (b(y), y)_{\mathbb{R}^N} \right) \leq \lambda_p (1 + |y|^p), 
\]
36

and therefore, by (36),

\[ E^W[1 + |X(t, x)|^p] \leq 1 + |x|^p + \lambda_p \int_0^t E^W[1 + |X(\tau, x)|^p] \, d\tau. \]

Therefore, by Gronwall’s lemma,

(*) \[ E^W[1 + |X(t, x)|^p] \leq (1 + |x|^p)e^{\lambda_p t}. \]

Since

\[ \|B(\cdot, x) - B(s, x)\|_{[s, t]}^p \leq \beta_p \int_s^t (1 + |X(\tau, x)|^p) \, d\tau, \]

it follows that

(**) \[ E^W[\|B(\cdot, x) - B(s, x)\|_{[s, t]}^p] \leq \beta_p (1 + |x|^p)(t - s)^p e^{\lambda_p t}. \]

Next observe that, by (35), for any \( \xi \in \mathbb{R}^N \),

\[ \left( |I(t, x) - \xi|^p - \frac{1}{2} \int_0^t \text{Trace}(a(X(\tau, x))H_p(I(\tau, x) - \xi)) \, d\tau, \overline{W}_t, \mathcal{W} \right) \]

is a martingale, and, as a consequence, for any \( s \geq 0 \),

\[ \left( |I(t \lor s, x) - I(s, x)|^p - \frac{1}{2} \int_s^{t \lor s} \text{Trace}(a(X(\tau, x))H_p(I(\tau, x) - I(s, x))) \, d\tau, \overline{W}_t, \mathcal{W} \right) \]

is a martingale. In particular, for \( 0 \leq s < t \),

\[ E^W[|I(t, x) - I(s, x)|^p] \leq \lambda_p^0 \int_s^t E^W[(1 + |X(\tau, x)|^p)^\frac{p}{2} |I(\tau, x) - I(s, x)|^{p-2}] \, d\tau \]

\[ \leq \lambda_p^0 \left( \int_s^t E^W[(1 + |X(\tau, x)|^p)] \frac{\tau^\frac{p}{2}}{E^W[|I(\tau, x) - I(s, x)|^p]]} \right)^{\frac{2}{p-2}} \]

\[ \leq \lambda_p^0 (t - s)^{\frac{2p}{p-2}} E^W[|I(t, x) - I(s, x)|^p]^{\frac{1}{p-2}}, \]

where I used Hölder’s inequality in the passage to the second line and, in the passage to the third line, (***) and the fact that, because \( I(t \lor s, x) - I(s, x) \) is a martingale, \( E^W[|I(t \lor s, x) - I(s, x)|^p] \) is a non-decreasing function of \( t \). By combining the preceding with

\[ E^W[|I(\cdot, x) - I(s, x)|_{[s, t]}^p]^{\frac{1}{2}} \leq \frac{p}{p - 1} E^W[|I(t, x) - I(s, x)|^p]^{\frac{1}{2}}, \]

the first estimate follows, and, since \( X(t, x) - x = I(t, x) + B(t, x) \), it is clear that the second estimate follows when one combines the first estimate with (**).
In order to remove the assumption that \( a \) and \( b \) are bounded, for each \( R > 0 \), set
\[
y^{(R)} = (y_1 \wedge R) \vee (-R), \ldots, (y_N \wedge R) \vee (-R),
\]
and define \( \sigma_R(y) = \sigma(y^{(R)}) \) and \( b_R(y) = b(y^{(R)}) \). Obviously, the quantities \( \lambda_p, \lambda^0_p, \) and \( \beta_p \) corresponding to \( a_R \equiv \sigma_R \sigma_R^{-1} \) and \( b_R \) are dominated by the original ones, and, by Lemma 38, if \( I_R(\cdot, x) \) and \( X_R(\cdot, x) \) are the stochastic processes determined by \( \sigma_R \) and \( b_R \), then
\[
\|I_R(\cdot, x) - I_R(s, x)\|_{[s,t]} = \|I(\cdot, x) - I(x, x)\|_{[s,t]}
\]
and
\[
\|X_R(\cdot, x) - X_R(s, x)\|_{[s,t]} = \|X(\cdot, x) - X(s, x)\|_{[s,t]}
\]
if \( t \leq \zeta_R \equiv \inf\{\tau : X(\tau, x) \in [-R, R]^N\} \). Hence
\[
\mathbb{E}^W[\|I(\cdot, x) - I(s, x)\|_{[s,t]}^p, \zeta_R \geq t] \leq \mathbb{E}^W[\|I_R(\cdot, x) - I_R(s, x)\|_{[s,t]}^p]
\]
and
\[
\mathbb{E}^W[\|X(\cdot, x) - X(s, x)\|_{[s,t]}^p, \zeta_R \geq t] \leq \mathbb{E}^W[\|X_R(\cdot, x) - X_R(s, x)\|_{[s,t]}^p].
\]
Because \( \zeta_R \not\sim \infty \) as \( R \to \infty \), this completes the proof. \( \square \)

**Theorem 40.** For any \( \varphi \in C^2(\mathbb{R}^N; \mathbb{C}) \) whose second derivatives have at most polynomial growth,
\[
\left( \varphi(I(t, x)) - \frac{1}{2} \int_0^t \text{Trace}(a(X(\tau, x)) \nabla^2 \varphi(I(\tau, x))) \, d\tau, W_t, W \right)
\]
and
\[
\left( \varphi(X(t, x)) - \int_0^t L \varphi(X(\tau, x)) \, d\tau, W_t, W \right)
\]
are martingales.

**Proof:** Determine \( X_R(t, x), I_R(t, x), \) and \( \zeta_R \) as in the last part of the proof of Lemma 39. Using (35), (36), and Doob’s stopping time theorem and Lemma 38, one sees that
\[
\left( \varphi(I(t \wedge \zeta_R, x)) - \frac{1}{2} \int_0^{t \wedge \zeta_R} \text{Trace}(a(X(\tau, x)) \nabla^2 \varphi(I(\tau, x))) \, d\tau, W_t, W \right)
\]
and
\[
\left( \varphi(X(t \wedge \zeta_R, x)) - \int_0^{t \wedge \zeta_R} L \varphi(X(\tau, x)) \, d\tau, W_t, W \right)
\]
are martingales for all $R > 0$. Furthermore, by the estimates in Lemma 39,

$$I(t \land \zeta_R, x) - \frac{1}{2} \int_0^{t \land \zeta_R} \text{Trace}(a(X(\tau, x)) \nabla^2 \varphi(I(\tau, x))) \, d\tau$$

$$\quad \rightarrow I(t, x) - \frac{1}{2} \int_0^t \text{Trace}(a(X(\tau, x)) \nabla^2 \varphi(I(\tau, x))) \, d\tau$$

and

$$\varphi(X(t \land \zeta_R, x)) - \int_0^{t \land \zeta_R} L \varphi(X(\tau, x)) \, d\tau$$

$$\quad \rightarrow \varphi(X(t, x)) - \int_0^t L \varphi(X(\tau, x)) \, d\tau$$

in $L^1(W; \mathbb{C})$ as $R \to \infty$, and so there is nothing more to do. $\square$

**Corollary 41.** For each $p \in [2, \infty)$ and $T > 0$, there exists a $C_p(T) < \infty$, depending only on $p$, $T$, and the Lipschitz norms of $\sigma$ and $b$, such that

$$\mathbb{E}^W[\|X(\cdot, x) - X(\cdot, \bar{x})\|_p]^{\frac{1}{p}} \leq C_p(T) |x - \bar{x}|.$$  

**Proof:** Define $\Phi_p : \mathbb{R}^{2N} = \mathbb{R}^N \times \mathbb{R}^N \mapsto [0, \infty)$ by $\Phi_p(y, \tilde{y}) = |y - \tilde{y}|^p$, and set

$$H_p(y, \tilde{y}) = p(p - 2)|y - \tilde{y}|^{p-2}\frac{(y - \tilde{y}) \otimes (y - \tilde{y})}{|y - \tilde{y}|^2} + p|y - \tilde{y}|^{p-2}I.$$

Then

$$\nabla \Phi_p(y, \tilde{y}) = |y - \tilde{y}|^{p-1}\left(\begin{array}{c} y - \tilde{y} \\ -(y - \tilde{y}) \end{array}\right)$$

and

$$\nabla^2 \Phi_p(y, \tilde{y}) = \left(\begin{array}{cc} H_p(y, \tilde{y}) & -H_p(y, \tilde{y}) \\ -H_p(y, \tilde{y}) & H_p(y, \tilde{y}) \end{array}\right).$$

Next set

$$\Sigma(y, \tilde{y}) = \left(\begin{array}{c} \sigma(y) \\ \sigma(\tilde{y}) \end{array}\right)$$

and

$$B(y, \tilde{y}) = \left(\begin{array}{c} b(y) \\ b(\tilde{y}) \end{array}\right).$$

Then

$$\left(\nabla \Phi_p(z, \tilde{z}), B(y, \tilde{y})\right)_{\mathbb{R}^{2N}} = p|z - \tilde{z}|^{p-2}(z - \tilde{z}, b(y) - b(\tilde{y}))_{\mathbb{R}^N}$$

and

$$\text{Trace}(\Sigma \Sigma^\top (y, \tilde{y})^\top \nabla^2 \Phi_p(z, \tilde{z}))$$

$$\quad = p(p - 2)|z - \tilde{z}|^{p-2} \left(\frac{|(\sigma(y) - \sigma(\tilde{y}))^\top (z - \tilde{z})|^2}{|z - \tilde{z}|^2} + p||\sigma(y) - \sigma(\tilde{y})||_{\text{H.S.}}^2\right).$$
Hence, since, by Theorem 40,
\[
\phi_p(I(t, x), I(t, \bar{x})) - \frac{1}{2} \int_0^t \text{Trace} \left( \sum \sum^\top (X(\tau, x), X(\tau, \bar{x})) \nabla^2 \phi_p(I(\tau, x), I(\tau, \bar{x})) \right) d\tau, W_t, W
\]

and
\[
\phi_p(X(t, x), X(t, \bar{x})) - \int_0^t \left( \left( \nabla \phi_p(X(\tau, x), X(\tau, \bar{x})), B(X(\tau, x), X(\tau, \bar{x})) \right)_{\mathbb{R}^{2N}} + \frac{1}{2} \text{Trace} \left( \sum \sum^\top (X(\tau, x), X(\tau, \bar{x})) \nabla^2 \phi_p(X(\tau, x), X(\tau, \bar{x})) \right) d\tau, W_t, W \right)
\]
are martingales, we see that there is a \( K_p < \infty \), depending only on \( p \) and the Lipschitz norms of \( \sigma \) and \( b \), such that
\[
\mathbb{E}^W[|I(t, x) - I(t, \bar{x})|^p] \leq K_p \int_0^t \mathbb{E}^W[|I(\tau, x) - I(\tau, \bar{x})|^{p-2}|X(\tau, x) - X(\tau, \bar{x})|^2] d\tau
\]
and
\[
\mathbb{E}^W[|X(t, x) - X(t, \bar{x})|^p] \leq |x - \bar{x}|^p + K_p \int_0^t \mathbb{E}^W[|X(\tau, x) - X(\tau, \bar{x})|^p] d\tau.
\]

Starting from these and proceeding as in the proof of Lemma 39, one arrives at the desired result. □

**Corollary 42.** There is a measurable map
\[
w \in C([0, \infty); \mathbb{R}^M) \mapsto Y(\cdot, \cdot, \cdot)(w) \in C([0, \infty) \times \mathbb{R} : \mathbb{R}^N)
\]
such that \( X(\cdot, x) = Y(\cdot, x) \) (a.s., \( W \)) for each \( x \in \mathbb{R}^N \). Moreover, for each \( \alpha \in [0, 1) \), \( p \in [1, \infty) \), and \( R > 0 \)
\[
\mathbb{E}^W \left[ \left( \sup_{x, y \in [-R, R]^N, x \neq y} \frac{|Y(\cdot, y) - Y(\cdot, x)|_{[0, T]}}{|y - x|^{\alpha}} \right)^p \right] < \infty
\]
and
\[ \mathbb{P}^w \left[ \sup_{(s, x), (t, y) \in [-R, R]^{N+1}} \left| \frac{Y(t, y) - Y(s, x)}{|(t, y) - (s, x)|^{\frac{\alpha}{2}}} \right|^p \right] < \infty. \]

**Proof:** Given the estimates in Corollary 41, the existence of the map \( w \rightsquigarrow Y(\cdot, \cdot)(w) \) as well as the estimate on its Hölder continuity as a function of \( x \) are immediate consequences of Kolmogorov’s continuity criterion. As for the estimate on its Hölder continuity as a function of \((t, x)\), one can combine the estimates in Lemma 39 and Corollary 41 to see that, for each \( p \in [1, \infty) \) and \( R > 0 \) there is a \( C(R) < \infty \) such that
\[ \mathbb{E}^w[|X(t, y) - X(s, x)|^p] \leq C(R)|(t, y) - (s, x)|^{\frac{\alpha}{2}} \quad \text{for} \quad (s, x), (t, y) \in [0, R]^{N+1}, \]
and so another application of Kolmogorov’s criterion completes the proof. \( \square \)

**Warning:** Although \( X(\cdot, x) \) and \( Y(\cdot, x) \) are not strictly speaking the same, I will continue to use \( X(\cdot, x) \) instead of \( Y(\cdot, x) \).

**The Markov Property:** In this section we will study the distribution \( \mathbb{P}_x \) on \( C([0, \infty); \mathbb{R}^N) \) of \( X(\cdot, x) \) under \( \mathcal{W} \). In particular, we will show that the family of measures \( \{\mathbb{P}_x : x \in \mathbb{R}^N\} \) satisfies the (strong) Markov property.

Before explaining what this means, observe that, as an application of Corollary 42, we know that \( x \rightsquigarrow \mathbb{E}^x[\Phi] \) is continuous for any bounded, continuous \( \Phi \) on \( C([0, \infty); \mathbb{R}^N) \) and therefore that it is Borel measurable for any Borel measurable \( \Phi \) on \( C([0, \infty); \mathbb{R}^N) \) that is either non-negative or in \( \mathcal{L}^1(\mathbb{P}_x; \mathbb{C}) \). Next let \( \psi \) denote a generic element of \( C([0, \infty); \mathbb{R}^N) \) and set \( \mathcal{B}_t = \sigma(\{\psi(\tau) : \tau \in [0, t]\}) \).

Then a **stopping time** \( \zeta \) relative to \( \{\mathcal{B}_t : t \geq 0\} \) is a measurable map \( \zeta : C([0, \infty); \mathbb{R}^N) \rightarrow [0, \infty] \) with the property that \( \{\zeta \leq t\} \equiv \{\psi : \zeta(\psi) \leq t\} \in \mathcal{B}_t \) for all \( t \geq 0 \). Associated with a stopping time \( \zeta \) is the \( \sigma \)-algebra \( \mathcal{B}_\zeta \) of sets \( \Gamma \subseteq C([0, \infty); \mathbb{R}^N) \) with the property that \( \Gamma \cap \{\zeta \leq t\} \in \mathcal{B}_t \) for all \( t \geq 0 \). One can show that \( \mathcal{B}_\zeta = \sigma(\{\psi(t \wedge \zeta) : t \geq 0\}) \).

Next define the time-shift maps \( \Sigma_s : C([0, \infty); \mathbb{R}^N) \rightarrow C([0, \infty); \mathbb{R}^N) \) so that \( \Sigma_s\psi(t) = \psi(s + t) \). Then the Markov property that we will prove says that for any stopping time \( \zeta \) and any non-negative, Borel measurable function \( \Phi \) on \( C([0, \infty); \mathbb{R}^N)^2 \),
\[ \int_{\{\zeta < \infty\}} \Phi(\psi^\zeta, \Sigma_\zeta\psi) \mathbb{P}_x(d\psi)
= \int \left( \int \Phi(\psi_1, \psi_2) \mathbb{P}_{\psi_1(\zeta)}(d\psi_2) \right) \mathbb{P}_x(d\psi_1), \]
\[ \text{(43)} \]

\(^6\text{See Exercise 7.1.21 in the 2nd edition of my Probability Theory, An Analytic View.}\)
where \( \psi(\zeta) \equiv \psi(\zeta(w)) \), \( \psi^\zeta(t) \equiv \psi(t \land \zeta) \), and \( \Sigma_\zeta \psi \equiv \Sigma_{\psi(\zeta)} \psi \). In particular,

\[
(44) \quad \mathbb{P}^x \left[ \Phi \circ \Sigma_\zeta \right] (\psi) = \mathbb{P}^x_{\psi(\zeta)} [\Gamma] \text{ for } \mathbb{P}_x \text{-a.e. } \psi \in \{ \zeta < \infty \}
\]

for any non-negative, Borel measurable \( \Phi \) on \( C([0, \infty); \mathbb{R}^N) \).

Because \( \mathcal{B}_\zeta = \sigma(\{ \psi^\zeta(t) : t \geq 0 \}) \), (44) follows immediately from (43). Furthermore, (43) holds for general non-negative, Borel measurable \( \Phi \)'s if it holds for bounded, continuous ones. Thus, assume that \( \Phi \) is bounded and continuous, and let \( \zeta \) be a bounded stopping time on \( C([0, \infty); \mathbb{R}^M) \) relative to \( \{ \mathcal{W}_t : t \geq 0 \} \) with values in \( \{ m2^{-\ell} : m \geq 0 \} \). Then, for any \( n \geq \ell \) and \( m \geq 0 \), \( X_n^{m2^{-\ell}}(\cdot, x)(\omega) = X_n(\cdot, \omega)(\omega) \left( w^{m2^{-\ell}} \right) \) and \( \Sigma_{m2^{-\ell}} X_n(\cdot, x) = X_n(\cdot, X_n(m2^{-\ell}, x)(\omega))(\delta_{m2^{-\ell}} w) \), where \( w^\zeta(t) = w(s \land t) \) and \( \delta_s w(t) = w(s + t) - w(s) \). Hence,

\[
\int \Phi(X_n^\zeta(\cdot, x)(w), X_n(\cdot + \zeta, x)(w)) \, dW(w) = \sum_{m=0}^{\infty} \int_{\{ \zeta = m2^{-\ell} \}} \Phi(X_n(\cdot, x)(w^{m2^{-\ell}}), X_n(\cdot, X_n(m2^{-\ell}, x)(\delta_{m2^{-\ell}} w))) \, dW(w).
\]

Because \( (\delta_s w(t), \mathcal{B}_{s+t}, \mathcal{W}) \) is a Brownian motion that is independent of \( \mathcal{B}_s \),

\[
\int_{\{ \zeta = m2^{-\ell} \}} \Phi(X_n(\cdot, x)(w^{m2^{-\ell}}), X_n(\cdot, X_n(m2^{-\ell}, x)(\delta_{m2^{-\ell}} w))) \, dW(w)
\]

and so

\[
\int \Phi(X_n^\zeta(\cdot, x)(w), X_n(\cdot + \zeta, x)(w)) \, dW(w)
\]

\[
= \int \left( \int \phi(X_n^\zeta(\cdot, x)(w_1), X_n(\cdot, X_n(\zeta(w_1), x)(w_2))) \, dW(w_2) \right) \, dW(w_1).
\]

Using (31) and passing to the limit as \( n \to \infty \), one concludes that

\[
\int \Phi(X^\zeta(\cdot, x)(w), X(\cdot + \zeta, x)(w)) \, dW(w)
\]

\[
= \int \left( \int \phi(X^\zeta(\cdot, x)(w_1), X(\zeta(w_1), x)(w_2))) \, dW(w_2) \right) \, dW(w_1)
\]

for any bounded stopping time \( \zeta \) with values in \( \{ m2^{-\ell} : m \geq 0 \} \) for some \( \ell \in \mathbb{N} \). Now suppose that \( \zeta \) is a bounded stopping time, and define \( \zeta_\ell = [\zeta]_\ell + 2^{-\ell} \). Then the preceding holds with \( \zeta_\ell \) replacing \( \zeta \), and therefore, by Corollary 42, it
also holds for $\zeta$ itself. Thus, by replacing $\zeta$ by $\zeta \circ X(\cdot, x)$, we have proved (43) for bounded stopping times $\zeta$ on $C([0, \infty); \mathbb{R}^N)$ relative to $\{B_t : t \geq 0\}$. Finally, given any $\zeta$ and $T > 0$, note that, by the preceding applied to $\zeta \wedge T$,

$$
\int_{\{\zeta < T\}} \Phi(\psi, \Sigma_\zeta \psi) \mathbb{P}_x(d\psi) = \int_{\{\zeta < T\}} \Phi(\zeta \wedge T, \Sigma_{\zeta \wedge T} \psi) \mathbb{P}_x(d\psi)
$$

Thus (43) follows after one lets $T \to \infty$.

**Characterizing the Distribution:** We have constructed the measure $\mathbb{P}_x$ and learned something about its properties, but, as yet, we have not characterized it. That is, starting with coefficients $a$ and $b$, we have given a construction which led to an associated $\mathbb{P}_x$, but there are many other construction methods that we might have adopted, and the question is whether they all would have led to the same place. To answer that question, we first have to decide what is the essential property of the measures on pathspace to which all these constructions lead. Once we have done so, we then need to know whether that property uniquely determines a Borel measure on pathspace.

As we have seen, the measure $\mathbb{P}_x$ to which our construction led has the property that

$$
(\varphi(t)) - \varphi(x) - \int_0^t L\varphi(\psi(\tau)) d\tau, B_t, \mathbb{P}_x
$$

is a martingale with mean value 0 for all $\varphi \in C^2([0, \infty); \mathbb{R})$, and we will say that such a measure solves the **martingale problem for $L$ starting at $x$**. Notice that if $P(t, x)$ is any transition function that satisfies Kolmogorov’s forward equation

$$
\langle \varphi, P(t, x) \rangle = \varphi(x) + \int_0^t \langle L\varphi, P(\tau, x) \rangle d\tau
$$

and if $\mathbb{P}_x$ is a measure on pathspace determined by

$$
\mathbb{P}_x(\psi(0) = x) = 1 \quad \text{and} \quad \mathbb{P}_x(\psi(t) \in \Gamma | B_s) = P(t - s, \psi(s), \Gamma) \quad \text{(a.s., $\mathbb{P}_x$),}
$$

for all $0 \leq s < t$, then

$$
\mathbb{E}^{\mathbb{P}_x}[\varphi(\psi(t)) - \varphi(\psi(s)) | B_s] = \langle \varphi, P(t - s, \psi(s)) \rangle - \psi(s)
$$

$$
= \int_0^{t-s} \langle L\varphi, P(\tau, \psi(s)) \rangle d\tau = \int_0^{t-s} \mathbb{E}^{\mathbb{P}_x}[L\varphi(\psi(s + \tau) | B_s]
$$

$$
= \mathbb{E}^{\mathbb{P}_x}\left[ \int_s^t L\varphi(\psi(\tau)) d\tau \bigg| B_s \right].
$$
and so \( P_x \) is also a solution to (45).

In view of the preceding, knowing that there is only one solution to (45) would afford us the freedom to choose any construction method that led to a solution. Thus it is important to know under what conditions (45) has a unique solution, and, depending on the situation, there are several approaches that one can adopt to find out. For example, consider the case when \( a = 0 \). If \( X(\cdot, x) \) is a solution

\[
\dot{X}(t, x) = b(X(t, x)) \quad \text{with} \quad X(0, x) = x,
\]

then the measure \( P_x = \delta_{X(\cdot, x)} \) will be a solution. Conversely, as the following lemma shows, when \( a = 0 \), any solution \( P_x \) to (45) will be concentrated on path that satisfy (*), and so, when \( a = 0 \), uniqueness of solutions to (45) is equivalent to uniqueness of solutions to (*).

**Lemma 46.** Assume that \( P_x \) satisfies (45). Then, for any \( \varphi \in C^2_b(\mathbb{R}^N; \mathbb{R}) \),

\[
\left( \varphi(\psi(t)) - \varphi(x) - \int_0^t L\varphi(\psi(\tau)) \, d\tau \right)^2 - \int_0^t \left( \nabla\varphi, a\nabla\varphi \right)_{\mathbb{R}^N} (\psi(\tau)) (\psi(\tau)) \, d\tau
\]

is \( P_x \)-martingale relative to \( \{B_t : t \geq 0\} \) with mean value 0. In particular, if \( a = 0 \), then

\[
P_x \left( \psi(t) = x + \int_0^t b(\psi(\tau)) \, d\tau \text{ for } t \geq 0 \right) = 1.
\]

**Proof:** Without loss in generality, we will assume that \( \varphi(x) = 0 \).

Set

\[
V(t) = \int_0^t L\varphi(\psi(\tau)) \, d\tau \quad \text{and} \quad M(t) = \varphi(\psi(t)) - V(t).
\]

Then, by (45) and Lemma 37,

\[
M(t)^2 = \varphi^2(\psi(t)) - 2M(t)V(t) - V(t)^2
\]

\[
\simeq \int_0^t L\varphi^2(\psi(\tau)) \, d\tau - 2 \int_0^t M(\tau) \, dV(\tau) - V(t)^2
\]

\[
= \int_0^t (L\varphi^2 - 2\varphi L\varphi) (\psi(\tau)) \, d\tau + 2 \int_0^t V(\tau) \, dV(\tau) - V(t)^2
\]

\[
= \int_0^t \left( \nabla\varphi, a\nabla\varphi \right)_{\mathbb{R}^N} (\psi(\tau)) \, d\tau,
\]

where "\( \simeq \)" is used here when the difference between the quantities it separates is a \( P_x \)-martingale relative to \( \{B_t : t \geq 0\} \), and, in the second line, I have used the identities \( V(t)^2 = 2 \int_0^t V(\tau) \, dV(\tau) \) and \( L\varphi^2 - 2\varphi L\varphi = \left( \nabla\varphi, a\nabla\varphi \right)_{\mathbb{R}^N} \).
Now assume that $a = 0$. Then the preceding says that
\[
\mathbb{E}^x \left[ \left( \varphi(t) - \varphi(x) - \int_0^t (b, \nabla \varphi)_{\mathbb{R}^N} (\psi(\tau)) \, d\tau \right)^2 \right] = 0
\]
for all $\varphi \in C^2_b(\mathbb{R}^N; \mathbb{R})$ and $t \geq 0$, and from this it is an easy step to a proof of the final assertion. □

When $a \neq 0$, proofs of uniqueness are more technically demanding. One approach relies on existence results for solutions to either Kolmogorov’s backward equation or closely related equations. To understand how this approach works, it is helpful to have the following lemma.

**Lemma 47.** Assume that $P_x$ is a solution to (45). If $v \in C^1_b((0, T) \times \mathbb{R}^N; \mathbb{R}) \cap C^{1,2}((0, T) \times \mathbb{R}^N; \mathbb{R})$ and $f \equiv (\partial + L)v$ is bounded on $(0, T) \times \mathbb{R}^N$, then
\[
\left( v(t \wedge \tau, \psi(t \wedge \tau)) - \int_{t \wedge \tau}^{t \wedge \tau \wedge \zeta_R} f(\tau, \psi(\tau)) \, d\tau, B_t, \mathbb{P}_x \right)
\]
is a martingale. In particular, if $f = 0$, then
\[
\left( v(t \wedge \tau, \psi(t \wedge \tau)), B_t, \mathbb{P}_x \right)
\]
is a martingale.

**Proof:** Assume for the moment that $v \in C^{1,2}((0, T) \times \mathbb{R}^N; \mathbb{R})$. Given $0 \leq s < t < T$, set $\tau_{m,n} = s + \frac{m}{n}(t - s)$. Then
\[
\mathbb{E}^x \left[ v(s, \psi(s)) \right] = \sum_{m=0}^{n-1} \mathbb{E}^x \left[ v(t \wedge \tau_{m+1,n}, \psi(t \wedge \tau_{m+1,n})) - v(t \wedge \tau_{m,n}, \psi(t \wedge \tau_{m,n})) \right] \left| B_s \right]
\]
\[
= \sum_{m=0}^{n-1} \mathbb{E}^x \left[ \int_{\tau_{m,n}}^{\tau_{m+1,n}} \left( \partial \tau \psi(\tau, \psi(t \wedge \tau_{m+1,n})) + Lv(\tau, \psi(t \wedge \tau_{m,n}, \psi(t \wedge \tau_{m,n}))) \right) \, d\tau \right] \left| B_s \right]
\]
\[
\rightarrow \mathbb{E}^x \left[ \int_{s}^{t} \left( \partial \tau - L \right) \psi(\tau, \psi(t \wedge \tau_{m,n}, \psi(t \wedge \tau_{m,n}))) \, d\tau \right] \left| B_s \right]
\]
as $n \to \infty$, and therefore the asserted martingale property holds in this case. To treat the general case, set $\zeta_R = \inf \{ t \geq 0 : \psi(t) \notin B(x, R) \} \wedge T$, and choose $v_R \in C^{1,2}((0, T) \times \mathbb{R}^N; \mathbb{R})$ so that $v = v_R$ on $[0, T] \times \overline{B(0, R)}$. Then, by the preceding applied to $v_R$ and Doob’s stopping time theorem,
\[
\left( v(t \wedge \zeta_R, \psi(t \wedge \zeta_R)) - \int_{0}^{t \wedge \zeta_R} f(\tau, \psi(\tau)) \, d\tau, B_t, \mathbb{P}_x \right)
\]
is a martingale, and so the desired result follows after one lets $R \to \infty$. □

Assume that $P_x$ is a solution to (45), and, for a given $\varphi \in C_b([0, \infty) \times \mathbb{R}^N; \mathbb{R})$ suppose that $u_\varphi \in C_b([0, \infty) \times \mathbb{R}^N; \mathbb{R}) \cap C^{1,2}((0, \infty) \times \mathbb{R}^N; \mathbb{R})$ is a solution to the Kolmogorov backward equation

$$
\partial_t u_\varphi = Lu_\varphi \quad \text{and} \quad u_\varphi(0, \cdot) = \varphi.
$$

Then, by the preceding applied to $v(t, \cdot) = u_\varphi(T - t, \cdot)$, we know that

$$
E^P_x[\varphi(T)|B_s] = u_\varphi(T - s, \psi(s)).
$$

Now assume that there is a solution $u_\varphi$ to (48) for each $\varphi$ in a determining class $\Phi$ of non-negative, bounded, continuous functions. Then, if $P_x$ and $\tilde{P}_x$ are solutions to (45),

$$
E^P_x[\varphi(t)|B_s] = u_\varphi(t, x) = E^\tilde{P}_x[\varphi(t)]
$$

for all $t \geq 0$ and $\varphi \in \Phi$. Next suppose that, for some $n \geq 1$ and every choice of $0 \leq t_1 < \cdots < t_n$, the distribution $\mu_{t_1, \ldots, t_n}$ of $(\psi(t_1), \ldots, \psi(t_n))$ under $P_x$ is the same as that under $\tilde{P}_x$. Then, for any $0 \leq t_1 < \cdots < t_{n+1}$, $\Gamma \in (\mathbb{R}^N)^n$, and $\varphi \in \Phi$,

$$
\begin{align*}
E^P_x[\varphi((t_{n+1}), (\psi(t_1), \ldots, \psi(t_n)) \in \Gamma] &= E^\tilde{P}_x[u_\varphi(t_{n+1} - t_n, \psi(t_n)), (\psi(t_1), \ldots, \psi(t_n)) \in \Gamma] \\
&= \int_{\Gamma} u_\varphi(t_{n+1} - t_n, y_n) \mu_{t_1, \ldots, t_n}(dy_1 \times \cdots \times dy_n),
\end{align*}
$$

and similarly

$$
\begin{align*}
E^\tilde{P}_x[\varphi((t_{n+1}), (\psi(t_1), \ldots, \psi(t_n)) \in \Gamma] &= \int_{\Gamma} u_\varphi(t_{n+1} - t_n, y_n) \mu_{t_1, \ldots, t_n}(dy_1 \times \cdots \times dy_n),
\end{align*}
$$

from which it follows that the distribution of $(\psi(t_1), \ldots, \psi(t_{n+1}))$ is the same under $P_x$ and $\tilde{P}_x$. Hence, by induction on $n$ and the fact that $\sigma(\{\psi(t) : t \geq 0\})$ is the Borel field on $C([0, \infty); \mathbb{R}^N)$, we know that $P_x = \tilde{P}_x$.

Alternately, one can prove uniqueness if one can show that the equation

$$
(\lambda - L)u_{\lambda, \varphi} = \varphi
$$

If $(E, \mathcal{F})$ is a measurable space, then a determining class $\Phi$ is a set of bounded, $\mathcal{F}$-measurable functions $\varphi$ such that, for any pair of finite measures $\mu$ and $\tilde{\mu}$ on $(E, \mathcal{F})$, $\mu = \tilde{\mu}$ if $\langle \varphi, \mu \rangle = \langle \varphi, \tilde{\mu} \rangle$ for all $\varphi \in \Phi$. 

(49)
has a solution in $C^2_b(\mathbb{R}^N; \mathbb{R})$ for $\lambda > 0$ and $\varphi$’s in a determining class $\Phi$ of non-negative, bounded continuous functions $\varphi$. The idea is very much the same. Namely, as the preceding argument shows, all that one needs to show is that, for each $\varphi \in \Phi$ and $0 \leq s < t$,

$$
\mathbb{E}^P_x[\varphi(\psi(t)) | B_s] = \mathbb{E}^{\tilde{P}}_x[\varphi(\psi(t)) | B_s]
$$

whenever $P_x$ and $\tilde{P}_x$ solve (45). Because a bounded, continuous function is determined by its Laplace transform, one will know this if, for each $\lambda > 0$,

$$
\int_0^{\infty} e^{-\lambda \tau} \mathbb{E}^P_x[\varphi(\psi(s + \tau)) | B_s] d\tau = \int_0^{\infty} e^{-\lambda \tau} \mathbb{E}^{\tilde{P}}_x[\varphi(\psi(s + \tau)) | B_s] d\tau,
$$

and, as we about to see, that follows from (49). Namely, by Lemma 47,

$$
e^{-\lambda t} u_{\lambda, \varphi}(\psi(t)) + \int_0^t e^{-\lambda \tau} \varphi(\psi(\tau)) d\tau, B_t, P_x
$$

is a martingale, and therefore

$$
e^{-\lambda t} u_{\lambda, \varphi}(\psi(t)) | B_s = e^{-\lambda s} u_{\lambda, \varphi}(\psi(s)) - \int_s^t e^{-\lambda \tau} \mathbb{E}^P_x[\varphi(\psi(\tau)) | B_s] d\tau.
$$

After letting $t \to \infty$, one concludes that

$$
\int_0^{\infty} e^{-\lambda \tau} \mathbb{E}^P_x[\varphi(\psi(s + \tau)) | B_s] d\tau = u_{\lambda, \varphi}(\psi(s))
$$

for any $P_x$ satisfying (45).

When $a$ and $b$ are uniformly bounded and $a$ is uniformly positive definite, in the sense that $a \geq \epsilon I$ for some $\epsilon > 0$, the theory of partial differential equations provides the needed solutions to (11) and (49). In fact, in that case, all that one needs is for $a$ and $b$ to be Hölder continuous. When $a$ is not uniformly positive definite, the situation is more complicated. As we will see below, when $a = \sigma \sigma^\top$ and both $\sigma$ and $b$ have two bounded, continuous derivatives, one can use the construction that we gave of $X(t, x)$ to show that (11) can be solved for $\varphi \in C^2_b(\mathbb{R}^N; \mathbb{R})$.

**Itô Integration**

The stochastic process $X(t, x)$ has three components: the starting point $x$, the Lebesgue integral $B(t, x) = \int_0^t b(X(\tau, x)) d\tau$, and the somewhat mysterious martingale $I(t, x)$. More generally, when the second derivatives of $\varphi \in C^2(\mathbb{R}^N; \mathbb{R})$ have at most polynomial growth, we saw that

$$
\varphi(X(t, x)) - \int_0^t L\varphi(X(\tau, x)) d\tau
$$

is a martingale for which we gave no explanation. In order to get a better understanding of these martingales, in this chapter I will develop Itô’s theory of stochastic integration and show that all these martingales can be expressed as stochastic integrals.
The Paley–Wiener Integral: Recall that if $f : [0, \infty) \to \mathbb{R}$ is continuous and $g : [0, \infty) \to \mathbb{R}$ has locally of bounded variation, then $f$ and $g$ are Riemann-Stieltjes integrable with respect to each other and

$$
\int_s^t g(\tau) \, df(\tau) = f(t)g(t) - f(s)g(s) - \int_s^t f(\tau) \, dg(\tau),
$$

from which it follows that $\int_s^t g(\tau) \, df(\tau)$ is a continuous function of $t > s$.

Assume that $(B(t), \mathcal{F}_t, \mathbb{P})$ is an $\mathbb{R}^M$-valued Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and, without loss in generality, assume that $B(\cdot)(\omega)$ is continuous for all $\omega \in \Omega$ and that the $\mathcal{F}_t$’s are complete. Given a function $\eta : [0, \infty) \to \mathbb{R}^M$ of locally bounded variation, set

$$I_{\eta}(t) = \int_0^t (\eta(\tau), dB(\tau))_{\mathbb{R}^M},$$

and define $I_{\eta_n}(t) = \eta([t]_n)$. Then $I_{\eta_n}(t) \to I_{\eta}(t)$. In addition, because

$$I_{\eta_n}(t) = \sum_{m < 2^n} (\eta(m2^{-n}), B((m+1)2^{-n} \wedge t) - B(m2^{-n}),$$

for all $0 \leq s < t$, the increment $I_{\eta_n}(t) - I_{\eta_n}(s)$ is an $\mathcal{F}_t$-measurable, centered Gaussian random variable that is independent of $\mathcal{F}_s$ and has variance $\int_s^t |\eta(\tau)|^2 \, d\tau$. Hence $I_{\eta}(t) - I_{\eta}(s)$ is an $\mathcal{F}_t$-measurable, centered Gaussian random variable that is independent of $\mathcal{F}_s$ and has variance $\int_s^t |\eta(\tau)|^2 \, d\tau$. In particular,

$$I_{\eta}(t), \mathcal{F}_t, \mathbb{P}) = \left(I_{\eta}(t)^2 - \int_0^t |\eta(\tau)|^2 \, d\tau, \mathcal{F}_t, \mathbb{P}\right),$$

and, for each $\zeta \in \mathbb{C},$

$$\left(\exp\left(\zeta I_{\eta}(t) - \frac{\zeta^2}{2} \int_0^t |\eta(\tau)|^2 \, d\tau\right), \mathcal{F}_t, \mathbb{P}\right)$$

are all martingales. Thus, by Doob’s inequality,

$$\frac{1}{2} \mathbb{E}^\mathbb{P}\left[\sup_{t \geq 0} I_{\eta}(t)^2\right] \leq \sup_{t \geq 0} \mathbb{E}^\mathbb{P}[I_{\eta}(t)^2] = \int_0^\infty |\eta(\tau)|^2 \, d\tau.$$

Given (52), it is easy to extend the preceding to all square-integrable $\eta : [0, \infty) \to \mathbb{R}^M$. Namely, if $\eta$ is such an function, choose $\{\eta_k : k \geq 1\} \subseteq C^\infty([0, \infty); \mathbb{R}^M)$ so that

$$\int_0^\infty |\eta_k(\tau) - \eta(\tau)|^2 \, d\tau \to 0.$$
Then, by (52),
\[
\mathbb{E}^P \left[ \sup_{t \geq 0} |I_{\eta}(t) - I_{\eta_k}(t)|^2 \right] = \mathbb{E}^P \left[ \sup_{t \geq 0} |I_{\eta} - I_{\eta_k}(t)|^2 \right] \leq 4 \int_0^\infty |\eta(\tau) - \eta_k(\tau)|^2 d\tau,
\]
and so there exists a measurable function $I_{\eta}$, known as the **Paley-Wiener integral** of $\eta$, such that $I_{\eta}(t)$ is $\mathcal{F}_t$ measurable for each $t \geq 0$, $t \mapsto I_{\eta}(t, \omega)$ is continuous for all $\omega \in \Omega$, and
\[
\mathbb{E}^P \left[ \sup_{t \geq 0} |I_{\eta}(t) - I_{\eta_k}(t)|^2 \right] \leq 4 \int_0^\infty |\eta(\tau) - \eta_k(\tau)|^2 d\tau \longrightarrow 0.
\]
Clearly, apart from a $\mathbb{P}$-null set, $I_{\eta}$ does not depend on the choice of the approximating sequence and is a linear function of $\eta$. Moreover, for each $t \geq 0$, $I_{\eta}(t)$ is a centered Gaussian random variable with variance $\int_0^t |\eta(\tau)|^2 d\tau$, the expressions in (50) and (51) are martingales, and (52) is continues to hold. What is no longer true is that $I_{\eta}(t)(\omega)$ is given by a Riemann-Stieltjes integral or that it even defined $\omega$ by $\omega$.

**Itô’s Integral**: Although they resemble one another, the difference between Itô’s integration theory and Paley-Wiener’s is that Itô wanted to allow his integrand to be random. Indeed, keep in mind that his goal was to develop a theory that would enable him to deal with quantities like $I(t, x)$. Of course, as long as the integrand is a random variable with values in the space of functions with locally bounded variation, Riemann-Stieltjes integration can be used. However, in general, such integrals will not satisfy any analog of (52) on which to base an extension to more general integrands. Further remember that he already familiar with the virtues of quantities like $I_{\eta}(t, x)$ and realized that those virtues resulted from using Riemann–Stieltjes approximations in which the integrand is independent of the increments of the Brownian motion. With this in mind, Itô considered integrands $\eta$ that are adapted to the filtration in the sense that $\eta(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$ and for which
\[
\mathbb{E}^P \left[ \int_0^\infty |\eta(\tau)|^2 d\tau \right] < \infty.
\]
If $\eta$ is such an integrand and if, in addition, $\eta(\cdot)(\omega)$ has locally bounded variation, then, for exactly the same reason as it held in the Paley-Wiener setting, (50) continues to hold, and therefore, just as before, one has
\[
\frac{1}{4} \mathbb{E}^P \left[ \sup_{t \geq 0} I_{\eta}(t)^2 \right] \leq \sup_{t \geq 0} \mathbb{E}^P \left[ I_{\eta}(t)^2 \right] = \mathbb{E}^P \left[ \int_0^\infty |\eta(\tau)|^2 d\tau \right]
\]
in place of (52).
Ito’s next step was to use (54) to extend his definition, and for this purpose he chose a $\rho \in C^\infty(\mathbb{R}; [0, \infty))$ that vanishes off $(0, 1)$ and has total integral 1, and set

$$\eta_k(t) = \int_0^\infty \rho_{\frac{t}{k}}(t-\tau)\eta(\tau)\,d\tau,$$

where $\rho_{\epsilon}(t) = \epsilon^{-1}\rho\left(\frac{t}{\epsilon}\right)$. Clearly $\eta_{\frac{t}{k}}$ has locally bounded variation, and, as is well-known,

$$\int_0^\infty |\eta_{\frac{t}{k}}(\tau)|^2\,d\tau \leq \int_0^\infty |\eta(\tau)|^2\,d\tau \quad \text{and} \quad \lim_{k \to \infty} \int_0^\infty |\eta_{\frac{t}{k}}(\tau) - \eta(\tau)|^2\,d\tau = 0.$$

Hence,

$$\lim_{k \to \infty} \mathbb{E}^\mathbb{P}\left[\int_0^\infty |\eta_{\frac{t}{k}}(\tau) - \eta(\tau)|^2\,d\tau \right] = 0.$$

Furthermore, because the construction of $\eta_{\frac{t}{k}}(t)$ involves only $\eta(\tau)$ for $\tau \leq t$, Itô, without further comment, claimed that $\eta_{\frac{t}{k}}$ must be adapted. However, as Doob realized, a rigorous proof of that requires an intricate argument, one that Doob provided when he explained Ito’s ideas in his renowned book *Stochastic Processes*. Fortunately, thanks to P.A. Meyer, there is a way to circumvent this technical hurdle by replacing adapted with the slightly stronger condition of **progressively measurable**. An $\mathbb{R}$-function on $[0, \infty) \times \Omega$ with values in a measurable space is said to be progressively measurable if it’s restriction to $[0, t] \times \Omega$ is $\mathcal{B}_{[0,t]} \times \mathcal{F}_t$-measurable for each $t \geq 0$. The great advantage of this notion is that a function is progressively measurable if and only if it is measurable with respect to the $\sigma$-algebra $\mathcal{P}$ of progressively measurable sets: those subsets of $[0, \infty) \times \Omega$ whose indicator functions are progressively measurable. In particular, a vector valued function will be progressively measurable if and only if each of its components is. Further, and from for us most important, it is elementary to check that $\eta_{\frac{t}{k}}$ is progressively measurable if $\eta$ is. Finally, although every progressively measurable function is adapted, not all adapted functions are progressively measurable. Nonetheless, it is easy to check that an adapted function will be progressively measurable if it is right-continuous with respect to $t$.

For the reason explained in the preceding paragraph, we will now restrict our attention to the class $\mathcal{P}^2(\mathbb{R}^M)$ of integrands $\eta$ which are progressively measurable and satisfy (53), and given $\eta \in \mathcal{P}^2(\mathbb{R}^M)$, one can use (54) to show that $I_{\eta_k}$ converges to an $I_\eta$ which, up to a $\mathbb{P}$-null set, is independent of the choice of approximants and for which (50) and (54) hold. In particular, $\eta \mapsto I_\eta$ is a linear map of $\mathcal{P}^2(\mathbb{R}^M)$ into the space $M^2_c(\mathbb{R})$ of continuous martingales $(M(t), \mathcal{F}_t, \mathbb{P})$ for which

$$\|M\|_{M^2_c(\mathbb{R})} \equiv \sup_{t \geq 0} \|M(t)\|_{L^2(\mathbb{P}; \mathbb{R})} < \infty.$$
then \( \|I_\eta\|_{M^2(\mathbb{R})} = \|\eta\|_{P^2(\mathbb{R}^M)} \), and so \( \eta \to I_\eta \) is an isometry between the metrics determined by the norms \( \| \cdot \|_{P^2(\mathbb{R}^M)} \) and \( \| \cdot \|_{M^2(\mathbb{R})} \).

Because it shares many properties with standard integrals, the quantity \( I_\eta(t) \) is usually denoted by
\[
\int_0^t (\eta(\tau), dB(\tau))_{\mathbb{R}^M}
\]
and is called the Itô stochastic integral, or just the stochastic integral, of \( \eta \) with respect to \( B \).

**Some Properties and Extensions:** Given \( \eta_1, \eta_2 \in P^2(\mathbb{R}^M) \), a simple polarization argument shows that
\[
\left( I_{\eta_1}(t)I_{\eta_2}(t) - \int_0^t (\eta_1(\tau), \eta_2(\tau))_{\mathbb{R}^M} d\tau, \mathcal{F}_t, \mathbb{P} \right)
\]
is a martingale.

Now suppose that \( \eta \in P^2(\mathbb{R}^M) \) and that \( \zeta \) is a stopping time relative to \( \{\mathcal{F}_t : t \geq 0\} \). Then, by Doob’s stopping time theorem,
\[
\mathbb{E}^\mathbb{P} \left[ I_{\eta}(t \wedge \zeta) \int_0^t 1_{[0,\zeta)}(\tau)(\eta(\tau), dB(\tau))_{\mathbb{R}^M} \right] = \mathbb{E}^\mathbb{P} \left[ I_{\eta}(t \wedge \zeta) \int_0^{t \wedge \zeta} 1_{[0,\zeta)}(\tau)(\eta(\tau), dB(\tau))_{\mathbb{R}^M} \right] = \mathbb{E}^\mathbb{P} \left[ \int_0^{t \wedge \zeta} |\eta(\tau)|^2 d\tau \right],
\]
and so
\[
\mathbb{E}^\mathbb{P} \left[ \left| I_{\eta}(t \wedge \zeta) - \int_0^t 1_{[0,\zeta)}(\tau)(\eta(\tau), dB(\tau))_{\mathbb{R}^M} \right|^2 \right] = 0.
\]
Hence
\[
I_{\eta}(t \wedge \zeta) = \int_0^t 1_{[0,\zeta)}(\tau)(\eta(\tau), dB(\tau))_{\mathbb{R}^M}.
\]
In particular, if \( \zeta_1 \) and \( \zeta_2 \) are a pair of stopping times and \( \zeta_1 \leq \zeta_2 \), then
\[
\int_0^{t \wedge \zeta_1} 1_{[0,\zeta_2)}(\tau)(\eta(\tau), dB(\tau))_{\mathbb{R}^M} = \int_0^t 1_{[0,\zeta_1)}(\tau)(\eta(\tau), dB(\tau))_{\mathbb{R}^M}.
\]

The preceding considerations afford us the opportunity to integrate \( \eta \)'s that are not in \( P^2(\mathbb{R}^M) \). Namely, let \( P^2_{\text{loc}}(\mathbb{R}^M) \) be the set of progressively measurable, \( \mathbb{R}^M \)-valued functions with the property that \( \int_0^t |\xi(\tau)|^2 d\tau < \infty \) for all \( t \in [0, \infty) \).

Then, if \( \xi_k = 1_{[0,\zeta_k]} \xi \) where
\[
\zeta_k = \inf \left\{ t \geq 0 : \int_0^t |\xi(\tau)|^2 d\tau \geq k \right\},
\]
\[ I_{\eta_{k+1}}(t \wedge \zeta_k) = I_{\eta_k}(t), \] and so not only does
\[ I_{\eta}(t) = \lim_{k \to \infty} I_{\eta_k}(t) \]
exist, but also \( I(t \wedge \zeta_k) = I_{\eta_k}(t) \) for all \( k \geq 1 \). Of course, in general \( (I_{\eta}(t), \mathcal{F}_t, \mathbb{P}) \) will not be a martingale since \( I_{\eta}(t) \) need not be even \( \mathbb{P} \)-integrable. On the other hand, for each \( k \geq 1 \),
\[ (I_{\eta}(t \wedge \zeta_k), \mathcal{F}_t, \mathbb{P}) \text{ and } \left( I_{\eta}(t)^2 - \int_0^{t \wedge \zeta_k} |\eta(\tau)|^2 \, d\tau, \mathcal{F}_t, \mathbb{P} \right) \]
will be martingales. Such considerations motivate the introduction of continuous local martingales: progressively measurable maps \( M \) that are continuous with respect to time and for which there exists a non-decreasing sequence \( \zeta_k \nearrow \infty \) of stopping times with the property that \( (M(t \wedge \zeta_k), \mathcal{F}_t, \mathbb{P}) \) is a martingale for each \( k \geq 1 \). Observe that if \( \eta \in \mathcal{P}^2_{\text{loc}}(\mathbb{R}^M) \) and \( \zeta \) is a stopping time, then, since
\[ I_{\eta}(t \wedge \zeta) = \lim_{k \to \infty} I_{\eta}(t \wedge \zeta_k) \]
it is still true that
\[ I_{\eta}(t \wedge \zeta) = \int_0^t 1_{[0, \zeta]}(\tau)(\eta(\tau), dB(\tau))_{\mathbb{R}^M}. \]
Thus, if \( \mathbb{E}^\mathbb{P} \left[ \int_0^\zeta |\eta(\tau)|^2 \, d\tau \right] < \infty \), then
\[ (I_{\eta}(t \wedge \zeta), \mathcal{F}_t, \mathbb{P}) \text{ and } \left( I_{\eta}(t \wedge \zeta)^2 - \int_0^{t \wedge \zeta} |\eta(\tau)|^2 \, d\tau, \mathcal{F}_t, \mathbb{P} \right) \]
are martingales. In particular, if \( \mathbb{E}^\mathbb{P} \left[ \int_0^t |\eta(\tau)|^2 \, d\tau \right] < \infty \) for all \( t \geq 0 \), then
\[ (I_{\eta}(t), \mathcal{F}_t, \mathbb{P}) \text{ and } \left( I_{\eta}(t)^2 - \int_0^t |\eta(\tau)|^2 \, d\tau, \mathcal{F}_t, \mathbb{P} \right) \]
are martingales. Finally, if \( \eta \) is an \( \mathbb{R}^M \)-valued, adapted function and \( \eta(\cdot)(\omega) \) is continuous for all \( \omega \in \Omega \), then \( \eta \) is progressively measurable and, by taking
\[ \zeta_k = \inf \left\{ t \geq 0 : \int_0^t |\eta(\tau)|^2 \, d\tau \geq k \right\}, \]
one sees that \( \eta \in \mathcal{P}^2_{\text{loc}}(\mathbb{R}^M) \).
Identify $\mathbb{R}^N \otimes \mathbb{R}^M$ with $\text{Hom}(\mathbb{R}^M; \mathbb{R}^N)$, let $\mathcal{P}^2(\mathbb{R}^N \otimes \mathbb{R}^M)$ be the of progressively measurable, $\mathbb{R}^N \otimes \mathbb{R}^M$-valued functions $\sigma$ with the property that

$$
\|\sigma\|_{\mathcal{P}^2(\mathbb{R}^N \otimes \mathbb{R}^M)} \equiv \mathbb{E}^P \left[ \int_0^\infty \|\sigma(\tau)\|_{H.S.}^2 \, d\tau \right]^{\frac{1}{2}} < \infty,
$$

and define the $\mathbb{R}^N$-valued random variable

$$
I_{\sigma}(t) = \int_0^t \sigma(\tau) \, dB(\tau)
$$

so that

$$(\xi, I_{\sigma}(t))_{\mathbb{R}^N} = \int_0^t (\sigma(\tau)^\top \xi, dB(\tau))_{\mathbb{R}^M} \text{ for each } \xi \in \mathbb{R}^N.$$ 

It is then an easy exercise to check that

$$(55) \quad (I_{\sigma}(t), \mathcal{F}_t, \mathbb{P}) \text{ and } \left( I_{\sigma}(t) \otimes I_{\tilde{\sigma}}(t) - \int_0^t \sigma(\tau)\tilde{\sigma}(\tau)^\top \, d\tau, \mathcal{F}_t, \mathbb{P} \right)$$

are martingales and that

$$
\frac{1}{2} \mathbb{E}^P \left[ \sup_{t \geq 0} |I_{\sigma}(t)|^2 \right]^{\frac{1}{2}} \leq \|I_{\sigma}\|_{M^2(\mathbb{R}^N \otimes \mathbb{R}^M)} \equiv \sup_{t \geq 0} \mathbb{E}^P [ |I_{\sigma}(t)|^2 ]^{\frac{1}{2}}
$$

$$
= \mathbb{E}^P \left[ \int_0^\infty \|\sigma(\tau)\|_{H.S.}^2 \, d\tau \right]^{\frac{1}{2}}.
$$

Further, starting from (55) and using polarization, one sees that if $\tilde{\sigma}$ is a second element of $\mathcal{P}^2(\mathbb{R}^N \otimes \mathbb{R}^M)$, then

$$
\left( I_{\sigma}(t) \otimes I_{\tilde{\sigma}}(t) - \int_0^t \sigma(\tau)\tilde{\sigma}(\tau)^\top \, d\tau, \mathcal{F}_t, \mathbb{P} \right)
$$

is a martingale. Finally, define $\mathcal{P}^2_{\text{loc}}(\mathbb{R}^N \otimes \mathbb{R}^M)$ by analogy to $\mathcal{P}^2_{\text{loc}}(\mathbb{R}^M)$, and define $I_{\sigma}(t)$ for $\sigma \in \mathcal{P}^2_{\text{loc}}(\mathbb{R}^N \otimes \mathbb{R}^M)$ accordingly.

**Stochastic Integral Equations:** We can now identify the quantity $I(t, x)$ in our construction. Indeed, take $\Omega = C([0, \infty); \mathbb{R}^M)$, $\mathcal{F} = \mathcal{B}_\Omega$, $\mathcal{F}_t = \mathcal{B}_t$, and $\mathbb{P} = \mathcal{W}$. Then

$$
I_n(t, x) = \int_0^t \sigma(X_n([\tau]_n, x)) \, dw(\tau).
$$

Recall that, for all $T > 0$,

$$
\lim_{n \to \infty} \mathbb{E}^W [ ||I(\cdot, x) - I_n(\cdot, x)||_{[0,T]}^2 \wedge ||X(\cdot, x) - X_n(\cdot, x)||_{[0,T]}^2 ] = 0
$$
and
\[
\lim_{n \to \infty} \sup_{t \in [0,T]} \mathbb{E}^W[|X(t, x) - X([t]_n, x)|^2] = 0
\]

Hence,
\[
\lim_{n \to \infty} \mathbb{E}^W\left[ \sup_{t \in [0,T]} \left| \int_0^t \sigma(X(\tau, x)) d\tau - I_n(t, x) \right|^2 \right] \leq 4 \lim_{n \to \infty} \mathbb{E}^W\left[ \int_0^T \|\sigma(X(\tau, x)) - \sigma(X_0([\tau]_n, x))\|^2_{H.S.} d\tau \right] = 0,
\]
and so
\[
I(t, x) = \int_0^t \sigma(X(\tau, x)) d\tau.
\]

We now know that \(X(\cdot, x)\) is a solution to the **stochastic integral equation**

(56)  \[
X(t, x) = x + \int_0^t \sigma(X(\tau, x)) d\tau + \int_0^t b(X(\tau, x)) d\tau.
\]

That is, it is a progressively measurable function that satisfies (56). In fact, it is the only such function. To see this, suppose that \(\tilde{X}(\cdot, x) \in \mathcal{P}^2_{loc}(\mathbb{R}^N)\) is a second solution, and set \(\zeta_R = \inf\{t \geq 0 : \tilde{X}(t, x) \geq R\}\).

\[
\mathbb{E}^W[|X(t \wedge \zeta_R, x) - \tilde{X}(t \wedge \zeta_R, x)|^2] \leq 2\mathbb{E}^W\left[ \int_0^{t \wedge \zeta_R} \|\sigma(X(\tau, x)) - \sigma(\tilde{X}(\tau, x))\|^2_{H.S.} d\tau \right]
+ 2t\mathbb{E}^W\left[ \int_0^{t \wedge \zeta_R} |b(X(\tau, x)) - b(\tilde{X}(\tau, x))|^2 d\tau \right]
\leq C(1 + t) \int_0^t \mathbb{E}^W[|X(\tau \wedge \zeta_R, x) - \tilde{X}(\tau \wedge \zeta_R, x)|^2] d\tau,
\]
for some \(C < \infty\), and so, by Gronwall’s lemma, \(X(t, x) = \tilde{X}(t, x)\) for \(t \in [0, \zeta_R]\). Since \(\zeta_R \nearrow \infty\), this proves that \(\tilde{X}(\cdot, x) = X(\cdot, x)\).

Having described \(X(\cdot, x)\) as the solution to (56), it is time for me to admit that the method that I used to construct the solution is not the one chosen by Itô. Instead of using Euler’s approximation scheme, Itô chose to use a Picard iteration scheme. That is, set \(X_0(t, x) = x\) and
\[
\tilde{X}_{n+1}(t, x) = x + \int_0^t \sigma(\tilde{X}_n(\tau, x)) d\tau + \int_0^t b(\tilde{X}_n(\tau, x)) d\tau
\]
for \( n \geq 0 \). Set \( \Delta_n(t) = \| \tilde{X}_{n+1}(\cdot, x) - \tilde{X}_n(\cdot, x) \|_{[0,t]} \). Then

\[
E^W[\Delta_n(t)^2] \leq 8 \int_0^t E^W[\| \sigma(\tilde{X}_n(t, x)) - \sigma(\tilde{X}_{n-1}(t, x)) \|^2_{\text{HS}}] \, d\tau + 2t \int_0^t E^W[|b(\tilde{X}_n(t, x)) - b(\tilde{X}_{n-1}(t, x))|^2] \, d\tau
\]

\[
\leq C(1 + t) \int_0^t E^W[\Delta_{n-1}(\tau)^2] \, d\tau
\]

for some \( C < \infty \) and all \( n \geq 1 \). Working by induction, one concludes that, for \( n \geq 1 \),

\[
E^W[\Delta_n(t)^2] \leq \frac{(C(1 + t)t)^{n-1}}{(n-1)!} E^W[\Delta_0(t)^2],
\]

and from this it is easy to show that there is an \( \tilde{X}(\cdot, x) \) such that

\[
\lim_{n \to \infty} E^W[\| \tilde{X}(\cdot, x) - \tilde{X}_n(\cdot, x) \|^2_{[0,t]}] = 0 \text{ for all } t \geq 0,
\]

and clearly this \( \tilde{X}(\cdot, x) \) will solve (56).

**The Crown Jewel: Itô’s Formula**: Although we interpreted \( X(\cdot, x) \) as a stochastic integral, we have yet to interpret the martingale

\[
(M_{\phi}(t), \mathcal{B}_t, W) \quad \text{where} \quad M_{\phi}(t) = \phi(X(t, x)) - \phi(x) - \int_0^t L\phi(X(\tau, x)) \, d\tau
\]

as one. If Brownian paths, and therefore \( X(\cdot, x) \), had locally bounded variation, we would know that

\[
\phi(X(t, x)) - \phi(x) = \int_0^t \left( \sigma(X(\tau, x))^\top \nabla \phi(X(\tau, x)), dw(\tau) \right)_{\mathbb{R}^N} + \int_0^t b(X(\tau, x)) \, d\tau,
\]

where the first integral is taken in the sense of Riemann-Stieltjes. Hence, it is reasonable to guess that, when we interpret it as an Itô integral, it might be the martingale that we are trying to identify. That is, we are guessing that

\[
\phi(X(t, x)) - \phi(x) = \int_0^t \left( \sigma(X(\tau, x))^\top \nabla \phi(X(\tau, x)), dw(\tau) \right)_{\mathbb{R}^N} + \int_0^t L\phi(X(\tau, x)) \, d\tau
\]

and that the appearance of the term

\[
\frac{1}{2} \int_0^t \text{Trace}(a(X(\tau, x))\nabla^2\phi(X(\tau, x))) \, d\tau
\]
is a progressively measurable Itô’s formula what is now called is a martingale.

It follows that infinitesimal. Further evidence for this conjecture is provided by Lemma 46, from which it follows that

\[
\left( M_{\varphi}(t)^2 - \int_0^t (\nabla \varphi(X(\tau,x)), a(X(\tau,x)) \nabla \varphi(X(\tau,x)))_\mathbb{R}^N \, d\tau, \mathcal{B}, \mathcal{W} \right)
\]

is a martingale.

That the preceding conjecture is correct was proved by Itô when he derived what is now called Itô’s formula, which states that, if \( \sigma \in \mathcal{P}_{loc}^2(\mathbb{R}^{N_2} \otimes \mathbb{R}^M) \), \( V \) is a progressively measurable \( \mathbb{R}^{N_1} \)-valued function for which \( V(\cdot)(\omega) \) is a continuous function of locally bounded variation for all \( \omega \in \Omega \), and \( \varphi \) is an element of \( C^{1,2}(\mathbb{R}^{N_1} \times \mathbb{R}^{N_2}; \mathbb{R}) \), then

\[
\varphi(V(t), I_\sigma(t)) - \varphi(V(0), 0) = \int_0^t \left( (\nabla(\varphi(V(\tau), I_\sigma(\tau)), dV(\tau))_{\mathbb{R}^{N_1}} \\
+ \int_0^\tau (\sigma(\tau)^\top \nabla \varphi(I_\sigma(\tau)), dB(\tau))_{\mathbb{R}^M} \\
+ \frac{1}{2} \int_0^\tau \text{Trace}(\sigma\sigma^\top(\tau)\nabla(\varphi(I_\sigma(\tau)), dB(\tau))) \, d\tau,
\right)
\]

(57)

where \( \nabla(1) \) and \( \nabla(2) \) are the gradient operators for the variables in \( \mathbb{R}^{N_1} \) and \( \mathbb{R}^{N_2} \). Using stopping times and standard approximation methods, one can easily show that it suffices to prove (57) in the case when \( \sigma \) and \( V \) are bounded, \( \sigma(\cdot) \) is continuous, and \( \varphi \in C^\infty_c(\mathbb{R}^{N_1} \times \mathbb{R}^{N_2}; \mathbb{R}) \). Thus I will proceed under those assumptions.

The proof relies on the observation that if \( \xi \) is a bounded, \( \mathcal{F}_s \)-measurable, \( \mathbb{R}^{N_2} \)-valued function, then \( t \to 1_{[s, \infty)}(t)\sigma(t)^\top \xi \) is progressively measurable and

\[
\left( \xi, \int_s^t \sigma(\tau) \, dB(\tau) \right)_{\mathbb{R}^{N_2}} = \int_s^t \left( \sigma(\tau)^\top \xi, dB(\tau) \right)_{\mathbb{R}^M} \quad \text{for } t \geq s.
\]

To check this, simply observe that

\[
\mathbb{E}^P \left[ \left( \xi, \int_s^t \sigma(\tau) \, dB(\tau) \right)_{\mathbb{R}^{N_2}} - \int_s^t \left( \sigma(\tau)^\top \xi, dB(\tau) \right)_{\mathbb{R}^M} \right]^2
\]

\[
= \mathbb{E}^P \left[ \int_s^t |\sigma(\tau)^\top \xi|^2 \, d\tau \right] - 2\mathbb{E}^P \left[ \int_s^t |\sigma(\tau)^\top \xi|^2 \, d\tau \right] + \mathbb{E}^P \left[ \int_s^t |\sigma(\tau)^\top \xi|^2 \, d\tau \right] = 0.
\]

Now define \( \sigma_n(t) = \sigma([t]_n) \). Then

\[
\varphi(V(t), I_\sigma(t)) - \varphi(V(0), 0) = \lim_{n \to \infty} \left( \varphi(V(t), I_{\sigma_n}(t)) - \varphi(V(0), 0) \right)
\]
Next, by Taylor’s theorem,

\[
\varphi(V(t_{m+1,n}), I_{m+1,n}) - \varphi(V(t_{m,n}), I_{m,n}) = \int_{t_{m,n}}^{t_{m+1,n}} (\nabla_1 \varphi(V(\tau), I_{m+1,n}), dV(\tau))_{\mathbb{R}^N},
\]

where \( t_{m,n} = m2^{-n} \land t \) and \( I_{m,n} = I_{m,n}(t_{m,n}) \). Since

\[
\varphi(V(t_{m+1,n}), I_{m+1,n}) - \varphi(V(t_{m,n}, I_{m+1,n})) = \int_{t_{m,n}}^{t_{m+1,n}} (\nabla_1 \varphi(V(\tau), I_{m+1,n}), dV(\tau))_{\mathbb{R}^N},
\]

and obviously

\[
\int_0^t (\nabla_1 \varphi(V(\tau), I_{m,n}(\lfloor \tau \rfloor_n)), dV(\tau))_{\mathbb{R}^N} \longrightarrow \int_0^t (\nabla_1 \varphi(V(\tau), I_{m,n}(\lfloor \tau \rfloor_n)), dV(\tau))_{\mathbb{R}^N}.
\]

Next, by Taylor’s theorem,

\[
\varphi(V(t_{m,n}), I_{m+1,n}) - \varphi(V(t_{m,n}), I_{m,n}) = \left( \nabla_2 \varphi(V(t_{m,n}), I_{m,n}), \Delta_{m,n} \right)_{\mathbb{R}^{N_2}}
\]

\[
+ \frac{1}{2} \text{Trace} \left( \nabla_2^2 \varphi(V(t_{m,n}), I_{m,n}) \Delta_{m,n} \otimes \Delta_{m,n} \right) + E_{m,n},
\]

where \( \Delta_{m,n} = I_{m+1,n} - I_{m,n} = \sigma(t_{m,n})(B(t_{m+1,n}) - B(t_{m,n})) \) and there is a \( C < \infty \) such that \( |E_{m,n}| \leq C|\Delta_{m,n}|^3 \). Because

\[
\mathbb{E}^\mathbb{P}[|B(t_{m+1,n}) - B(t_{m,n})|^3] \leq \mathbb{E}^\mathbb{P}[|B(2^{-n})|^3] = 2^{-3n} \mathbb{E}^\mathbb{P}[|B(1)|^3],
\]

\( \sum_{m < 2^n} E_{m,n} \) tends to 0 \( \mathbb{P} \)-a.s., and, by the preceding observation,

\[
\sum_{m < 2^n} \left( \nabla_2 \varphi(V(t_{m,n}), I_{m,n}), \Delta_{m,n} \right)_{\mathbb{R}^{N_2}}
\]

\[
= \sum_{m < 2^n} \int_{t_{m,n}}^{t_{m+1,n}} \left( \sigma(\lfloor \tau \rfloor_n)^\top \nabla_2 \varphi(V(t_{m,n}, I_{m,n})), dB(\tau) \right)_{\mathbb{R}^M}
\]

\[
= \int_0^t \left( \sigma(\tau)^\top \nabla_2 \varphi(V(\lfloor \tau \rfloor_n), I_{m,n}(\lfloor \tau \rfloor_n)), dB(\tau) \right)_{\mathbb{R}^M}
\]

\[
\longrightarrow \int_0^t \left( \sigma(\tau)^\top \nabla_2 \varphi(V(\tau), I_{m,n}(\tau)), dB(\tau) \right)_{\mathbb{R}^M}
\]
in $L^2(\mathbb{P}; \mathbb{R})$. Finally, write
\[
\text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) \Delta_{m,n} \otimes \Delta_{m,n} \right) \\
= \text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) A_{m,n} \right) \\
+ \text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) (\Delta_{m,n} \otimes \Delta_{m,n} - A_{m,n}) \right),
\]
where $A_{m,n} = \int_{t_{m,n}}^{t_{m+1,n}} \sigma_n \sigma_n^\top (\tau) \, d\tau = (t_{m+1,n} - t_{m,n}) \sigma \sigma^\top (t_{m,n})$. Clearly
\[
\sum_{m<2^n} \text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) A_{m,n} \right) \\
= \int_0^t \text{Trace} \left( \nabla^2_2 \varphi(V(\tau), I_{\sigma_n} \sigma_\tau(\tau)) \sigma \sigma^\top \sigma_\tau(\tau) \right) \, d\tau \\
\rightarrow \int_0^t \text{Trace} \left( \nabla^2_2 \varphi(V(\tau), I_{\sigma} \sigma^\top(\tau)) \right) \, d\tau.
\]
At the same time, because
\[
\mathbb{E}^P \left[ \Delta_{m,n} \otimes \Delta_{m,n} \mid \mathcal{F}_{t_{m,n}} \right] \\
= \mathbb{E}^P \left[ I_{\sigma_n} (t_{m+1,n}) \otimes I_{\sigma_n} (t_{m+1,n}) - I_{\sigma_n} (t_{m,n}) \otimes I_{\sigma_n} (t_{m,n}) \mid \mathcal{F}_{t_{m,n}} \right] \\
= \mathbb{E}^P [A_{m,n} \mid \mathcal{F}_{t_{m,n}}],
\]
the terms
\[
\text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) (\Delta_{m,n} \otimes \Delta_{m,n} - A_{m,n}) \right)
\]
are orthogonal in $L^2(\mathbb{P}; \mathbb{R})$, and so
\[
\mathbb{E}^P \left[ \left( \sum_{m<2^n} \text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) (\Delta_{m,n} \otimes \Delta_{m,n} - A_{m,n}) \right) \right)^2 \right] \\
= \sum_{m<2^n} \mathbb{E}^P \left[ \text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) (\Delta_{m,n} \otimes \Delta_{m,n} - A_{m,n}) \right)^2 \right].
\]
Obviously, $\|A_{m,n}\|_{\text{H.S.}} \leq C2^{-n}$ and $\|\Delta_{m,n} \otimes \Delta_{m,n}\|_{\text{H.S.}} \leq C|B(t_{m+1,n} - B(t_{m,n})|^2$ for some $C < \infty$. Hence, since
\[
\mathbb{E}^P [|B(t_{m+1,n} - B(t_{m,n})|^4] = 2^{-2n} \mathbb{E}^P [|B(1)|^4],
\]
it follows that
\[
\sum_{m<2^n} \text{Trace} \left( \nabla^2_2 \varphi(V(t_{m,n}), I_{m,n}) \Delta_{m,n} \otimes \Delta_{m,n} - A_{m,n} \right) \rightarrow 0
\]
in $L^2(\mathbb{P}; \mathbb{R})$.

It has become customary to write integral equations like the one in (57) in differential form:

$$
\begin{align*}
\quad & d\varphi(V(t), I_\sigma(t)) \\
= & \left( \nabla_1 \varphi(V(t), I_\sigma(t)), dV(t) \right)_{\mathbb{R}^M} + \left( \sigma(\tau)^T \nabla_2 \varphi(V(t), I_\sigma(\tau)), dB(\tau) \right)_{\mathbb{R}^M} \\
& + \frac{1}{2} \text{Trace} \left( \sigma(\tau)^T \nabla_2 \varphi(V(t), I_\sigma(t)) \right) dt.
\end{align*}
$$

However, it is best to remember that such an expression should not be taken too seriously and really makes sense only when one “integrates” it to get the equivalent integral expression.

It should be clear that, by taking $N_1 = N_2 = N$, $V(t) = x + \int_0^t b(X(\tau, x)) d\tau$, and $\sigma(\tau) = \sigma(X(\tau, x))$, the equation

$$
\varphi(X(t, x) = \varphi(x) + \int_0^t \left( \sigma(X(\tau, x)^T \nabla \varphi(X(\tau, x)), dB(\tau) \right)_{\mathbb{R}^M} \\
+ \int_0^t L \varphi(X(\tau, x)) d\tau
$$

becomes a special case of (57) applied to $(x_1, x_2) \mapsto \varphi(x_1 + x_2)$.

**Some Applications:** Let $\sigma \in \mathcal{P}^2_{\text{loc}}(\mathbb{R}^N \otimes \mathbb{R}^M)$, and set $A(t) = \int_0^t \sigma\sigma^T(\tau) d\tau$. My first application will to the development of relations between $I_\sigma(\cdot)$ and $A(\cdot)$.

To begin with, given $\zeta \in \mathbb{C}^N$, apply (57) to the function $(y, v) \in \mathbb{R}^N \times \mathbb{R} \mapsto e^{(\zeta, y)_{\mathbb{R}^N} - \frac{1}{2}(\zeta, A(t)\zeta)_{\mathbb{R}^N}}$ to see that

$$
\exp \left( \left( \zeta, I_\sigma(t) \right)_{\mathbb{R}^N} - \frac{1}{2} \left( \zeta, A(t)\zeta \right)_{\mathbb{R}^N} \right), \mathcal{F}_t, \mathbb{P}
$$

is a local $\mathbb{C}$-valued martingale. Thus, if

$$
\zeta_R = \inf \{ t \geq 0 : |I_\sigma(t)| \vee \text{Trace}(A(t)) \geq R \},
$$

then, for any $\xi \in \mathbb{R}^N$,

$$
\mathbb{E}^\mathbb{P} \left[ \exp \left( \left( \xi, I_\sigma(t \wedge \zeta_R) \right)_{\mathbb{R}^N} - \frac{1}{2} \left( \xi, A(t \wedge \zeta_R)\xi \right)_{\mathbb{R}^N} \right) \right] = 1,
$$

which, by HW3, means that, for $m \geq 1$,

$$
\kappa_2^{-1} \mathbb{E}^\mathbb{P} \left[ \left( \xi, A(t \wedge \zeta_R)\xi \right)_{\mathbb{R}^N}^m \right] \leq \mathbb{E}^\mathbb{P} \left[ \left( \xi, I_\sigma(t \wedge \zeta_R) \right)_{\mathbb{R}^N}^{2m} \right] \leq \kappa_2^{-1} \mathbb{E}^\mathbb{P} \left[ \left( \xi, A(t \wedge \zeta_R)\xi \right)_{\mathbb{R}^N}^m \right].
$$
Therefore, for each \( p \), we have

\[
\mathbb{E}^p[|\xi, I_\sigma(t)|_{\mathbb{R}^N}^{2m}] \leq \kappa_2 m \mathbb{E}^p[|\xi, A(t)\xi|_{\mathbb{R}^N}^{m}].
\]

At the same time, if \( \mathbb{E}^p[|\xi, I_\sigma(t)|_{\mathbb{R}^N}^{2m}] < \infty \), then \( \mathbb{E}^p[|\xi, I_\sigma(\cdot)|_{\mathbb{R}^N}^{2m}] < \infty \), and so, by Lebesgue’s dominated convergence theorem,

\[
\kappa_2^{-1} m \mathbb{E}^p[|\xi, I_\sigma(t)|_{\mathbb{R}^N}^{m}] \leq \mathbb{E}^p[|\xi, I_\sigma(t)|_{\mathbb{R}^N}^{2m}].
\]

Hence, we have now shown that

\[
(58) \quad \kappa_2^{-1} m \mathbb{E}^p[|\xi, A(t)\xi|_{\mathbb{R}^N}^{m}] \leq \mathbb{E}^p[|\xi, I_\sigma(t)|_{\mathbb{R}^N}^{2m}] \leq \kappa_2 m \mathbb{E}^p[|\xi, A(t)\xi|_{\mathbb{R}^N}^{m}],
\]

from which it follows immediately that

\[
(59) \quad \kappa_2^{-1} \kappa_1^{1-m} m \mathbb{E}^p[\text{Trace}(A(t))^{m}] \leq \mathbb{E}^p[|I_\sigma(t)|^{2m}] \leq \kappa_2 m N^{m-1} \mathbb{E}^p[\text{Trace}(A(t))^{m}].
\]

The estimates in (58) and (59) are a very special cases of general inequalities for martingales proved by D. Burkholder, and it should be pointed out that Itô’s formula can be used to prove, with a constant that doesn’t depend on \( N \), the second of inequality in (59). Namely, given \( p \in [2, \infty) \), remember that

\[
\nabla^2 |y|^p = p(p-2)|y|^{p-2} + p|y|^{p-2} I.
\]

Again using stopping times to reduce to the case when \( I_\sigma \) and \( A \) are bounded, apply (57) and Hölder’s inequality to see that

\[
\mathbb{E}^p[|I_\sigma(t)|^p] \leq p(p-1) \mathbb{E}^p \left[ \int_0^t \text{Trace}(\sigma \sigma^\top(\tau)) |I_\sigma(\tau)|^{p-2} d\tau \right]
\]

\[
\leq p(p-1) \mathbb{E}^p \left[ \left( \int_0^t \text{Trace}(\sigma \sigma^\top(\tau)) d\tau \right) \|I_\sigma(\cdot)\|_{[0,t]}^{p-2} \right]
\]

\[
\leq p(p-1) \mathbb{E}^p [A(t)^{\frac{p}{2}}]^{\frac{2}{p}} \mathbb{E}^p \left[ \|I_\sigma(\cdot)\|_{[0,t]}^p \right]^{1-\frac{2}{p}},
\]

and then use Doob’s inequality to conclude that

\[
\mathbb{E}^p \left[ \|I_\sigma(\cdot)\|_{[0,t]}^2 \right]^{\frac{2}{p}} \leq \left( \frac{p^{p+1}}{(p-1)p-1} \right)^{\frac{2}{p}} \mathbb{E}^p [A(t)^{\frac{p}{2}}]^{\frac{2}{p}}.
\]

Therefore, for each \( p \in [2, \infty) \),

\[
(60) \quad \mathbb{E}^p[|I_\sigma(t)|^p] \leq \mathbb{E}^p \left[ \|I_\sigma(\cdot)\|_{[0,t]} \right] \leq K_p \mathbb{E}^p [A(t)^{\frac{p}{2}}].
\]
where \( K_p = \left( \frac{p^{p+1}}{(p-1)^{p-1}} \right)^{\frac{3}{p}} \).

The preceding applications of (57) are mundane by comparison to the one made by H. Tanaka to represent what P. Lévy called local time for an \( \mathbb{R} \)-valued Brownian motion \((B(t), \mathcal{F}_t, \mathbb{P})\). To describe local time, define the occupation time measures \( L(t, \cdot) \) for \( t \geq 0 \) by

\[
L(t, \Gamma) = \int_0^t 1_{\Gamma}(B(\tau)) \, d\tau, \quad \Gamma \in \mathcal{B}_{\mathbb{R}}.
\]

One of Lévy’s many remarkable discoveries is that there is a map \( \omega \in \Omega \mapsto -\ell(\cdot, \cdot) \in C([0, \infty) \times \mathbb{R}) \) such that

\[
\mathbb{P} \left( L(t, \Gamma) = \int_{\Gamma} \ell(t, y) \, dy \text{ for all } t \geq 0 \& \Gamma \in \mathcal{B}_{\mathbb{R}} \right) = 1.
\]

In other words, with probability 1, for all \( t \geq 0 \), \( L(t, \cdot) \) is absolutely continuous with respect to Lebesgue measure \( \lambda_{\mathbb{R}} \) and

\[
\frac{L(t, dy)}{\lambda_{\mathbb{R}}(dy)} = \ell(t, y).
\]

Notice that this results reflects another aspect of non-differentiability of Brownian paths. Indeed, suppose that \( p : [0, \infty) \rightarrow \mathbb{R} \) is a continuously differentiable path, and let \( t \mapsto \mu(t, \cdot) \) be its occupation time measures. If \( \dot{p} = 0 \) on an interval \([a, b]\), then it is clear that, for \( t > a \), \( \mu(t, \{p(a)\}) \geq (t - b + a) \), and therefore \( \mu(t, \cdot) \) can’t be absolutely continuous with respect to \( \lambda_{\mathbb{R}} \). On the other hand, if \( \dot{p} > 0 \) on \([a, b]\), then, for \( t > a \),

\[
\frac{\mu(t, dy) - \mu(a, dy)}{\lambda_{\mathbb{R}}(dy)} = \frac{1_{[a, t]}(y)}{\dot{p}(p(t))} \cdot \frac{1}{\dot{p}(p(t))} = \hat{\mu}(t, y),
\]

and so \( \mu(t, \cdot) - \mu(a, \cdot) \) is absolutely continuous but its Radon-Nikodym derivative cannot be continuous. It is because the graph of a Brownian path is so fuzzy that its occupations time measures can admit a continuous density.

Tanaka’s idea to prove the existence of local time is based on the heuristic representation of \( \ell(t, y) \) as

\[
\int_0^t \delta_y(B(\tau)) \, d\tau,
\]

where \( \delta_y \) is the Dirac delta function at \( y \). Because \( h_y = 1_{[y, \infty)} \) and \( h''_y = \delta_y \), when \( h_y(x) = x \lor y \), a somewhat cavalier application of (57) to \( h_y \) led him to guess that

\[
\frac{1}{2} \ell(t, y) = B(t) \lor y - 0 \lor y - \int_0^t 1_{[y, \infty)}(B(\tau)) \, d\tau.
\]
The key step in the justification of his idea is the proof that there is a continuous map \((t, y) \rightsquigarrow I(t, y)\) such that

\[
\mathbb{P}\left(I(t, y) = \int_0^t 1_{[y, \infty)}(B(\tau)) \, dB(\tau) \text{ for all } t \geq 0\right) = 1 \text{ for each } y \in \mathbb{R}.
\]

To this end, use (60) to see that there is a \(C < \infty\) such that

\[
\mathbb{E}^\mathbb{P}\left[\sup_{t \in [0, T]} \left| \int_0^t 1_{[y, \infty)}(B(\tau)) \, dB(\tau) - \int_0^t 1_{[x, \infty)}(B(\tau)) \, dB(\tau) \right|^4 \right] \\
\leq C \mathbb{E}^\mathbb{P}\left[\left( \int_0^T 1_{[x,y]}(B(\tau)) \, d\tau \right)^2 \right]
\]

for \(T > 0\) and \(x < y\). Next use the Markov property for Brownian motion to write

\[
\mathbb{E}^\mathbb{P}\left[\left( \int_0^T 1_{[x,y]}(B(\tau)) \, d\tau \right)^2 \right] = 2 \mathbb{E}^\mathbb{P}\left[\int_{0 \leq \tau_1 < \tau_2} 1_{[x,y]}(B(\tau_1)) 1_{[x,y]}(B(\tau_2)) \, d\tau_1 d\tau_2 \right]
\]

\[
= \frac{1}{2\pi} \int_{0 \leq \tau_1 < \tau_2} \tau_1^{-\frac{1}{2}} (\tau_2 - \tau_1)^{-\frac{1}{2}} \left( \int_{[x,y]^2} e^{-\frac{\xi_1^2}{2\tau_1}} e^{-\frac{(\xi_2 - \xi_1)^2}{2(\tau_2 - \tau_1)}} \, d\xi_1 d\xi_2 \right) d\tau_1 d\tau_2 \\
\leq \frac{2T(y - x)^2}{\pi}.
\]

Hence the existence of \((t, y) \rightsquigarrow I(t, y)\) follows from Kolmogorov’s continuity criterion.

We now define

\[
\ell(t, y) = 2(B(t) \vee y - B(t) \vee 0 - I(t, y)).
\]

Given \(\varphi \in C_c(\mathbb{R}; \mathbb{R})\), set \(f(x) = \int_{\mathbb{R}} (x \vee y) \varphi(y) \, dy\). Then

\[
f'(x) = \int_{-\infty}^x \varphi(y) \, dy \text{ and } f'' = \varphi,
\]

and so, by (57),

\[
f(x) = f(0) + \int_0^t f'(B(\tau)) \, dB(\tau) + \frac{1}{2} \int_0^t \varphi(B(\tau)) \, d\tau.
\]

At the same time,

\[
\frac{1}{2} \int \varphi(y) \ell(t, y) \, dy = f(B(t)) - f(0) - \int \varphi(y) I(t, y) \, dy,
\]
and because $\varphi$ and $I(t, \cdot)$ are continuous and $\varphi$ has compact support,

$$\int \varphi(y) I(t, y) \, dy = \lim_{n \to \infty} \frac{1}{n} \sum_{m \in \mathbb{N}} \varphi(\frac{m}{n}) I(t, \frac{m}{n}),$$

and

$$\lim_{n \to \infty} \int_0^t \left( \frac{1}{n} \sum_{m \in \mathbb{N}} \varphi(\frac{m}{n}) 1_{(-\infty, B(\tau))}(\frac{m}{n}) \right) \, dB(\tau) = \int_0^t f'(B(\tau)) \, dB(\tau)$$

$\mathbb{P}$-a.s. Hence, we have shown that

$$\int \varphi(y) \ell(t, y) \, dy = \int_0^t \varphi(B(\tau)) \, d\tau \quad (\text{a.s., } \mathbb{P})$$

for each $\varphi \in C_c(\mathbb{R}; \mathbb{R})$. Starting from this it is an easy matter to show that (61) holds, and once one has (61), there is no reason not to replace $\ell(t, y)$ by $\|\ell(\cdot, y)\|_{[0, t]}$ so that it becomes non-decreasing.

The function $\ell(\cdot, y)$ has many strange and interesting properties. Since it is non-decreasing, it determines a Borel measure $\ell(dt, y)$ on $[0, \infty)$ determined by

$$\ell([s, t], y) = \ell(t, y) - \ell(s, y) \text{ for } 0 \leq s < t,$$

and we will now show

(62) $\mathbb{P}\left( \ell(\{t : B(t) \neq y\}, y) = 0 \right) = 1,$

which is the reason why $\ell(\cdot, y)$ is called the local time at $y$. To this end, first observe that if $f$ is a bounded $\mathcal{B}_{[0, \infty)} \times \mathcal{B}_\mathbb{R} \times \mathcal{F}$-measurable $\mathbb{R}$-valued function and $(t, x) \mapsto f(t, x, \omega)$ is continuous for each $\omega \in \Omega$, then, by using Riemann sum approximations, one can see that

(*) \quad \int \left( \int_0^t f(\tau, x, \omega) \ell(d\tau, x)(\omega) \right) \, dx = \int_0^t f(\tau, B(\tau)(\omega), \omega) \, d\tau

for $\mathbb{P}$-almost every $\omega$. Now choose continuous functions $\rho : \mathbb{R} \rightarrow [0, \infty)$ and $\eta : \mathbb{R} \rightarrow [0, 1]$ so that $\rho = 0$ off $(-1, 1)$ and has total integral 1, and $\eta = 0$ on $[-1, 1]$ and 1 off $(-2, 2)$. Next, for $\epsilon, R > 0$, define $\rho_\epsilon(x) = \epsilon^{-1} \rho(\epsilon^{-1} x)$ and $\eta_R(x) = \eta(Rx)$, and set

$$f_{\epsilon, R}(t, x, \omega) = \rho_\epsilon(x - y) \eta_R(B(t)(\omega) - y).$$

If $0 < \epsilon < R^{-1}$, then $f_{\epsilon, R}(\tau, B(\tau)(\omega), \omega) = 0$, and therefore, by (*),

$$\int \rho_\epsilon(x - y) \left( \int \eta_R(B(\tau) - y) \ell(d\tau, x) \right) \, dx = 0$$
\( P \)-a.s. Thus (62) follows when one first lets \( \epsilon \searrow 0 \) and then \( R \nearrow \infty \). As a consequence of (62), we know that, with probability 1, \( t \mapsto \ell(t, y) \) is a singular, continuous, non-decreasing function. Indeed, if

\[
S(\omega) = \{ \tau \in [0, \infty) : B(\tau)(\omega) = y \},
\]

then the expected value of the Lebesgue measure of \( S \) equals

\[
\mathbb{E}^P \left[ \int_0^\infty 1_{\{y\}}(B(\tau)) \ d\tau \right] = \int_0^\infty \gamma_{0,\tau}([y]) \ d\tau = 0.
\]

Here is another application of local time. Because \( x = x \vee 0 + x \wedge 0 \) and \( |x| = x \vee 0 - x \wedge 0 \).

\[
B(t) \wedge 0 = B(t) - \int_0^t 1_{[0,\infty)}(B(\tau)) \ dB(\tau) - \frac{1}{2} \ell(t, 0) = \int_0^t \text{sgn}(B(\tau)) \ dB(\tau) - \frac{1}{2} \ell(t, 0),
\]

and

\[
|B(t)| = \tilde{B}(t) + \ell(t, 0) \quad \text{where} \quad \tilde{B}(t) = \int_0^t \text{sgn}(B(\tau)).
\]

Now observe that, for all \( \xi \in \mathbb{R} \),

\[
(e^{i \xi \tilde{B}(t)} + \frac{\xi^2}{2}, \mathcal{F}_t, \mathbb{P})
\]

is a martingale, and conclude that \((\tilde{B}(t), \mathcal{F}_t, \mathbb{P})\) is a Brownian motion. Hence, since \( \ell(\cdot, 0) \) is \( \mathbb{P} \)-almost surely constant on time intervals during which \( B \) stays away from 0, (63) says that \( t \mapsto |B(t)| \) behaves like a Brownian motion on such intervals and is prevented from becoming negative because \( \ell(\cdot, 0) \) gives it a kick to the right whenever it is at 0. It should also be noted that

\[
\tilde{\mathcal{F}}_t = \sigma(\{B(\tau) : \tau \in [0,t]\}) \subseteq \sigma(\{|B(\tau)| : \tau \in [0,t]\}).
\]

To check this, remember that, \( \mathbb{P} \)-a.s.,

\[
\ell(t, 0) = \lim_{\delta \searrow 0} \frac{1}{2\delta} \int_0^t 1_{[-\delta, \delta]}(B(\tau)) \ d\tau = \lim_{\delta \searrow 0} \frac{1}{2\delta} \int_0^t 1_{[0,\delta]}(|B(\tau)|) \ d\tau,
\]

and therefore both \( \ell(t, 0) \) and \( \tilde{B}(t) \) are \( \sigma(\{|B(\tau)| : \tau \in [0,t]\}) \)-measurable. In particular, this means that \( \text{sgn}(B(t)) \) is not \( \tilde{\mathcal{F}}_t \)-measurable, and so \( B(t) \) isn’t either. Even so, \( B(\cdot) \) is a solution to the stochastic integral equation

\[
(*) \quad X(t) = \int_0^t \text{sgn}(X(\tau)) \ dB(\tau).
\]
To see this, all that one has to show is that
\[
\int_0^t \xi(\tau) \, d\tilde{B}(\tau) = \int_0^t \xi(\tau) \text{sgn}(B(\tau)) \, dB(\tau)
\]
for bounded \( \xi \in P^{2}_{\text{loc}}(\mathbb{R}) \). When \( \xi(\cdot) \) has bounded variation, this is clear from the theory of Riemann-Stieltjes integration, and so it follows in general when one uses the same approximation procedure that allowed us to define stochastic integrals. The interest of this observation is that it shows that solutions to stochastic integral equations need not be measurable with respect to the driving Brownian motion. In addition, it shows that there are stochastic integral equations all of whose solutions have the same distribution even though there is more than one solution. Namely, every solution to (*) must be a Brownian motion, but both \( B(\cdot) \) and \( -B(\cdot) \) are solutions.

We will now examine some properties of \( \ell(t,y) \).

**Lemma 64.** For \( t > 0 \) and \( y \neq 0 \),
\[
\mathbb{P}(\ell(t,y) > s) = \left( \frac{2y^2}{\pi} \right)^{1/2} \int_0^t e^{-\tau^2/4} \mathbb{P}(\ell(t-\tau) > s) \, d\tau.
\]

In addition,
\[
(65) \quad \mathbb{P}(\ell(t,0) > 0) = 1 \text{ for all } t > 0,
\]
and so
\[
\mathbb{P}(\ell(t,y) > 0) = 2 \mathbb{P}(B(t) > y).
\]

**Proof:** First observe that the distribution of \( \ell(\cdot, -y) \) is the same as that of \( \ell(\cdot, y) \). Thus we will assume that \( y > 0 \) throughout.

Next, set \( \zeta_y = \inf\{t \geq 0 : B(t) \geq y\} \). Then \( \zeta_y < t \) implies that
\[
L(t, [y,y+\delta]) = \int_{t \wedge \zeta_y}^t 1_{[y,y+\delta]}(B(\tau)) \, d\tau = \int_0^{t-t \wedge \zeta_y} 1_{[y,\delta]}(B(\tau+t \wedge \zeta_y)-B(t \wedge \zeta_y)) \, d\tau.
\]

Since
\[
(B(\tau + t \wedge \zeta_y) - B(t \wedge \zeta_y), \mathcal{F}_{t \wedge \zeta_y}, \mathbb{P})
\]
is a Brownian motion that is independent of \( \mathcal{F}_{t \wedge \zeta_y} \), it follows that, for any bounded, continuous \( f : [0, \infty) \rightarrow \mathbb{R} \),
\[
\mathbb{E}^{\mathbb{P}}\left[f\left(\frac{1}{2} L(t, [y,y+\delta])\right)\right] = \int_{\zeta_y < t} \left( \int f\left(\frac{1}{2} L(t-t \wedge \zeta_y(\omega_1), [0,\delta])(\omega_2)\right) \, d\omega_2\right) \, d\omega_1.
\]
Hence, after letting $\delta \searrow 0$, we have

$$\mathbb{E}^\pi[f(\ell(t,y))] = \int_{\{\zeta_y < t\}} \left( \int f(\ell(t - \zeta_y(\omega),0) \zeta(\omega)) \mathbb{P}(d\omega) \right) \mathbb{P}(d\omega).$$

Starting from here and using the fact that $\mathbb{P}(\zeta_y < t) = 2\mathbb{P}(B(t) > y)$, it is easy to verify the first assertion.

To prove (65), note that, because $\ell(t,0) = |B(t)| - \tilde{B}(t)$ and $-\tilde{B}(\cdot)$ is a Brownian motion, all that we have to do is note that

$$\mathbb{P}(|\zeta_y| < t) = \lim_{y \searrow 0} \mathbb{P}(\zeta_y < t) = \left( \frac{2}{\pi t} \right)^{\frac{1}{2}} \int_0^\infty e^{-\frac{x^2}{4t}} dx = 1. \quad \square$$

I will close this discussion with another of Lévy’s remarkable insights. What Lévy wanted to do is “count” the number of times that a Brownian motion visits 0 by time $t$. Of course, due to its fuzzy nature, one suspects it visits infinitely often, and so one has to be careful about what one means. With this in mind, Lévy considered the number $N(\cdot)$ of times before $t$ that $B(\cdot)$ returns to 0 after leaving $(-\epsilon, \epsilon)$. That is, $N(\cdot) \geq n$ if and only if there exist $0 < r_1 < \cdots < r_n \leq t$ such that $|B(r_{m+1})| \geq \epsilon$ and $|B(r_m)| = 0$ for $1 \leq m \leq n$. He then showed that

$$\mathbb{P}\left( \lim_{\epsilon \searrow 0} \|\epsilon N(\cdot) - \ell(\cdot, 0)\|_{[0,t]} = 0 \right) = 1 \text{ for all } t > 0. \quad (66)$$

One way to prove (66) is to use (63). Indeed, if $\zeta_0 = 0$ and, for $m \geq 1$, $\zeta_{2m+1} = \inf\{t \geq \zeta_{2m} + 1 : |B(t)| > \epsilon \}$ and $\zeta_{2m} = \inf\{t \geq \zeta_{2m-1} : |B(t)| = 0\}$, then, because $B(\cdot)$ is continuous, $\zeta_m \geq t$ for all but a finite number of $m$’s, and

$$\sum_{m \geq 1} (|B(\zeta_{2m+1}) - B(\zeta_{2m})|) = \epsilon N(\cdot) + (\epsilon - |B(t)|) \sum_{m \geq 1} 1_{\{\zeta_{2m} < \zeta_{2m+1}\}}(t).$$

On the other hand, because $B(\tau) \neq 0$ for $\tau \in (\zeta_{2m-1}, \zeta_{2m})$, $\ell(\zeta_{2m-1}, 0) = \ell(\zeta_{2m}, 0)$, and therefore

$$\sum_{m \geq 1} (|B(\zeta_{2m+1}) - B(\zeta_{2m})|) = \sum_{m \geq 1} (\tilde{B}(\zeta_{2m+1}) - \tilde{B}(\zeta_{2m}))$$

$$= -\tilde{B}(t) + \sum_{m \geq 0} (\tilde{B}(\zeta_{2m+1}) - \tilde{B}(\zeta_{2m})) = -|B(t)| + \ell(t,0) + \int_0^t \xi(\tau) d\tilde{B}(\tau),$$

where

$$\xi(\tau) = \sum_{m \geq 0} \text{sgn}(B(\tau)) 1_{[\zeta_{2m}, \zeta_{2m+1}]}(\tau).$$
After combining these, we see that
\[
\epsilon N_n(t) - \ell(t, 0) = -|B(t)| \sum_{m \geq 0} 1_{[\zeta_{2m}, \zeta_{2m+1}]}(t) - \epsilon \sum_{m \geq 1} 1_{(\zeta_{2m-1}, \zeta_{2m})}(t) + \int_0^t \xi(\tau) dB(\tau).
\]
Since $|B(t)| \leq \epsilon$ for $t \in [\zeta_{2m}, \zeta_{2m+1}]$, the absolute value of the sum the first two terms on the right is no larger than $\epsilon$ and $|\xi(\tau)| \leq 1_{[0, \epsilon]}(|B(\tau)|)$. Hence,
\[
\mathbb{E}^\mathbb{P}[||\epsilon N_n(\cdot) - \ell(\cdot, 0)||_{[0, t]}^2]^{\frac{1}{2}} \leq \epsilon + 2\mathbb{E}^\mathbb{P}\left[\int_0^t 1_{[0, \epsilon]}(B(\tau)) \, d\tau\right]^{\frac{1}{2}},
\]
which means that there is a $C < \infty$ such that
\[
\mathbb{E}^\mathbb{P}[||\epsilon N_n(\cdot) - \ell(\cdot, 0)||_{[0, t]}^2]^{\frac{1}{2}} \leq C(t\epsilon)^{\frac{1}{2}} \text{ for } \epsilon \in (0, 1].
\]

In particular,
\[
\sum_{k \geq 1} \mathbb{E}^\mathbb{P}[||k^{-4}N_{k^{-4}}(\cdot) - \ell(\cdot, 0)||_{[0, t]}^2]^{\frac{1}{2}} < \infty,
\]
and so $||k^{-4}N_{k^{-4}}(\cdot) - \ell(\cdot, 0)||_{[0, t]} \rightarrow 0$ (a.s., $\mathbb{P}$). Finally, note that if $(k+1)^{-4} \leq \epsilon \leq k^{-4}$, then
\[
(k+1)^{-4}N_{k^{-4}}(\cdot) \leq \epsilon N_n(\cdot) \leq k^{-4}N_{k+1^{-4}}(\cdot),
\]
and thereby conclude that (66) holds.

In conjunction with Lemma 64, (66) gives abundant, quantitative evidence that Brownian paths are fuzzy. Indeed, it says that, for any $B$ and thereby conclude that (66) holds.

**Spacial Derivatives of Solutions:** Let $X(\cdot, x)$ be the solution to (56), where $\sigma$ and $b$ be Lipschitz continuous. As we showed, we can assume that $(t, x) \sim X(t, x)$ is continuous. What I want to show now is that, if $\sigma$ and $b$ are continuously differentiable, then $x \sim X(t, x)$ is also. However, in order to carry out that program, I need to make some preparations.

Recall the Euler approximants $X_n(t, x)$, which are the solutions to
\[
X_n(t, x) = x + \int_0^t \sigma(X_n([\tau]_n)) \, dw(\tau) + \int_0^t b(X_n([\tau]_n)) \, d\tau,
\]
and set $\Delta_n(t, x) = X(t, x) - X_n(t, x)$. Using (60) and arguing as we did when $p = 2$, one can show that for each $p \in [1, \infty)$ and $t > 0$ there is a $C_p(t) < \infty$ such that
\[
\mathbb{E}^W[||\Delta_n(\cdot, x)||_{[0, t]}^p]^{\frac{1}{p}} \leq C_p(t)(1 + |x|)2^{-\frac{p}{2}},
\]
\[
\mathbb{E}^W[||X_n(\cdot, x) - X_n(s, x)||_{[s, t]}^p]^{\frac{1}{p}} \leq C_p(t)(1 + |x|)(t-s)^{\frac{1}{2}},
\]
\[
\mathbb{E}^W[||X_n(\cdot, y) - X_n(\cdot, x)||_{[s, t]}^p]^{\frac{1}{p}} \leq C_p(t)|y - x|
\]
for $n \geq 0$, $0 \leq s < t$, and $x, y \in \mathbb{R}^N$. From these it was clear the $\|\Delta_n(\cdot, x)\|_{[0,t]} \to 0$ both (a.s., $\mathbb{P}$) and in $L^p(W; \mathbb{R}^N)$, but I now want to use Kolmogorov’s continuity criterion to show that this convergence is uniform with respect to $x$ in compact subsets. This end, note that

$$
|\Delta_n(t, y) - \Delta_n(t, x)|^p
= |(X(t, y) - X(t, x)) - (X_n(t, y) - X_n(t, x))|^\frac{p}{2} |\Delta_n(t, y) - \Delta_n(t, x)|^\frac{p}{2},
$$

apply Schwarz’s inequality to get

$$
E^W[\|\Delta_n(\cdot, y) - \Delta_n(\cdot, x)\|^p_{[0,t]}]
\leq E^W[\|X(\cdot, y) - X(\cdot, x)\|_{[0,t]}^p]^{\frac{1}{2}} \times E^W[\|\Delta_n(\cdot, y) - \Delta_n(\cdot, x)\|^p_{[0,t]}]^{\frac{1}{2}},
$$

and, after combining these with the first and third estimates in (67), conclude that, for each $R > 0$, there is a $K_p(t, R) < \infty$ such that

$$
E^W[\|\Delta_n(\cdot, y) - \Delta_n(\cdot, x)\|^p_{[0,t]}]^{\frac{1}{2}} \leq K_p(t, R)2^{-\frac{n}{2}} |y - x|^{\frac{p}{2}} \text{ for } x, y \in [0, R]^N.
$$

Hence, by taking $p > \frac{N}{2}$, we can apply Kolmogorov’s continuity criterion to see that

$$
\sup_{n \geq 0} 2^{\frac{n}{2}} \frac{\|\Delta_n\|_{[0,t] \times [-R,R]^N}}{2} < \infty
$$

and therefore that, $\mathbb{P}$-almost surely, $\|X_n(\cdot, x) - X(\cdot, x)\|_{[0,t]} \to 0$ uniformly for $x$ in compact subsets of $\mathbb{R}^N$.

In order to prove the differentiability of $X(t, \cdot)$, I will need the following lemma that is harder to state than it is to prove.

**Lemma 68.** For each $n \geq 0$, let

$$
\tilde{\sigma}_n : [0, \infty) \times \mathbb{R}^N \times \tilde{\mathbb{R}} \times C([0, \infty); \mathbb{R}^M) \to Hom(\mathbb{R}^M; \tilde{\mathbb{R}})
$$

and

$$
\tilde{b}_n : [0, \infty) \times \mathbb{R}^N \times C([0, \infty); \mathbb{R}^M) \to \mathbb{R}^N
$$

be Borel measurable functions with the properties that $\tilde{\sigma}_n(\cdot, x, \tilde{x})$ and $\tilde{b}_n(\cdot, x, \tilde{x})$ are progressively measurable for each $(x, \tilde{x}) \in \mathbb{R}^N \times \tilde{\mathbb{R}}$ and for which there exists a $C < \infty$ such that

$$
\|\tilde{\sigma}_n(t, x, \tilde{x})\|_{H.S.} \vee |\tilde{b}_n(t, x, \tilde{x})| \leq C,
$$

and

$$
\left| \frac{\|\tilde{\sigma}_n(t, x, \tilde{x}) - \tilde{\sigma}_n(t, x, \tilde{y})\|_{H.S.} \vee |\tilde{b}_n(t, x, \tilde{y}) - \tilde{b}_n(t, x, \tilde{x})|}{|\tilde{y} - \tilde{x}|} \right| \leq C.
$$
for all $n \geq 0$, $(t, x, \bar{x}) \in [0, \infty) \times \mathbb{R}^N \times \mathbb{N}$, and $\bar{y} \neq \bar{x}$. In addition, assume that for each $p \in [1, \infty)$ there is a non-decreasing function $t \mapsto C_p(t)$ such that

$$
\mathbb{E}^W\left[ \frac{\|\tilde{\sigma}_{n+1}(t, x, \bar{x}) - \tilde{\sigma}_n(t, x, \bar{x})\|_{H^1} \vee \|\tilde{b}_{n+1}(t, x, \bar{x}) - \tilde{b}_n(t, x, \bar{x})\|_p}{1 + |x| + |\bar{x}|} \right]^{\frac{1}{p}} \leq C_p(t),
$$

and

$$
\mathbb{E}^W\left[ \frac{\|\tilde{\sigma}_n(\cdot, y, \bar{x}) - \tilde{\sigma}_n(\cdot, x, \bar{x})\|_{[0, t]}^p \vee \|\tilde{b}_n(\cdot, y, \bar{y}) - \tilde{b}_n(\cdot, x, \bar{x})\|_{[0, t]}^p}{1 + |\bar{x}|} \right]^{\frac{1}{p}} \leq C_p(t)|y - x|
$$

for all $n \geq 0$, $(t, x, \bar{x}) \in [0, \infty) \times \mathbb{R}^N$, and $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$. Finally, for each $n \geq 0$, let $\tilde{Y}_n : [0, \infty) \times \mathbb{R}^N \times C([0, \infty); \mathbb{R}^M) \rightarrow \mathbb{R}^N$ be a measurable map with the properties that $Y(\cdot, x)$ is progressively measurable of each $x \in \mathbb{R}^N$, and, for each $p \in [2, \infty)$,

$$
\mathbb{E}^W\left[ \frac{\|\tilde{Y}_n(\cdot, x)\|_{[0, t]}^p}{1 + |x|^k} \right]^{\frac{1}{p}} \leq C_p(t),
$$

$$
\mathbb{E}^W\left[ \frac{\|\tilde{Y}_n(\cdot, y) - \tilde{Y}_n(\cdot, x)\|_{[0, t]}^p}{1 + |x|^k + |y|^k} \right]^{\frac{1}{p}} \leq C_p(t)|y - x|,
$$

$$
\mathbb{E}^W\left[ \frac{\|\tilde{Y}_n(\cdot, x) - \tilde{Y}_n(s, x)\|_{[s, t]}^p}{1 + |x|^k} \right]^{\frac{1}{p}} \leq C_p(t)(t - s)^{\frac{1}{2}}
$$

and

$$
\mathbb{E}^W\left[ \frac{\|\tilde{Y}_{n+1}(\cdot, x) - \tilde{Y}_n(\cdot, x)\|_{[0, t]}^p}{1 + |x|^k} \right]^{\frac{1}{p}} \leq C_p(t)2^{-\frac{k}{2}}
$$

for some $k \in \mathbb{N}$. If $\tilde{X}_n(t, x)$ is determined by

$$
\tilde{X}_n(t, x) = \int_0^t \tilde{\sigma}_n(\tau, x, \tilde{X}_n([\tau], x)) \, d\tau + \int_0^t \tilde{b}_n(\tau, x, \tilde{X}_n([\tau], x)) \, d\tau + \tilde{Y}_n(t, x)
$$

and $\tilde{\Delta}_n(t, x) = \tilde{X}_{n+1}(t, x) - \tilde{X}_n(t, x)$, then, for each $p \in [2, \infty)$ there is a non-decreasing $t \mapsto \tilde{C}_p(t)$ such that

$$
\mathbb{E}^W\left[ \|\tilde{X}_n(\cdot, x)\|_{[0, t]}^p \right]^{\frac{1}{p}} \leq \tilde{C}_p(t),
$$

$$
\mathbb{E}^W\left[ \|\tilde{\Delta}_n(\cdot, x)\|_{[0, t]}^p \right]^{\frac{1}{p}} \leq \tilde{C}_p(t)(1 + |x|^k)2^{-\frac{k}{2}},
$$

$$
\mathbb{E}^W\left[ \|\tilde{X}_n(\cdot, x) - \tilde{X}_n(s, x)\|_{[s, t]}^p \right]^{\frac{1}{p}} \leq \tilde{C}_p(t)(1 + |x|^k)(t - s)^{\frac{1}{2}},
$$

$$
\mathbb{E}^W\left[ \|\tilde{X}_n(\cdot, y) - \tilde{X}_n(\cdot, x)\|_{[s, t]}^p \right]^{\frac{1}{p}} \leq \tilde{C}_p(t)|y - x|
$$

for all $n \geq 0$, $0 \leq s < t$, and $x, y \in \mathbb{R}^N$. 
Proof: There are no new ideas required to prove this lemma because each of the conclusions is proved in the same way as the corresponding estimate in (67). For example, to prove the second one, first observe that, by (60),

\[ E^W[\|\tilde{x}_n(\cdot, x) - \tilde{x}_n(s, x)\|^p_{[s,t]}] \]

\[ \leq 3^{p-1}K_pE^W\left[ \left( \int_s^t \|\tilde{\sigma}_n(\tau, x, \tilde{x}_n([\tau]_n, x))\|^2_{H.S.} d\tau \right)^{\frac{p}{2}} \right] \]

\[ + 3^{p-1}E^W\left[ \left( \int_s^t \|\tilde{b}_n(\tau, x, \tilde{x}_n([\tau]_n, x))\| d\tau \right)^p \right] \]

\[ + 3^{p-1}E^W[\|\tilde{y}_n(\cdot, x) - \tilde{y}_n(s, x)\|^p_{[s,t]}]. \]

When \( s = 0 \), this says that

\[ E^W[\|\tilde{x}_n(\cdot, x)\|^p_{[0,t]}] \]

\[ \leq 3^{p-1}(1 + K_p)(t^{\frac{p}{2}} + t^{p-1})C \int_0^t E^W[1 + \|\tilde{x}_n(\cdot, x)\|^p_{[0,\tau]}] d\tau \]

\[ + 6^{p-1}C_p(t^p(1 + |x|^k)^p, \]

and therefore, after an application of Gromwall's inequality, one finds that

\[ \sup_n \sup_{x \in \mathbb{R}^N} \frac{E^W[\|\tilde{x}_n(\cdot, x)\|^p_{[0,t]}]}{(1 + |x|^k)^p} < \infty. \]

Finally, after combining this estimate with the growth assumptions about \( \tilde{\sigma}_n \) and \( \tilde{b}_n \) as functions of \( \tilde{x} \), one arrives quickly at the desired result.

The proofs of the other results are left as an exercise. □

Now assume that \( \sigma \) and \( b \) are twice differentiable and that their first and second derivatives are bounded, and observe that, for each \( n \geq 0 \), \( X_n(t, \cdot) \) is twice continuously differentiable and that its first derivatives satisfy

\[ \frac{\partial X_n(t, x)}{\partial x_j} = \frac{\partial X_n(m2^{-n}, x)}{\partial x_j} \]

\[ + \sum_{\ell=1}^N \sum_{k=1}^M \frac{\partial \sigma(X_n(m2^{-n}, x))_{i,k}}{\partial x_{\ell}} \frac{\partial X_n(m2^{-n}, x)_{\ell}}{\partial x_j} (w(t) - w(m2^{-n}))_k \]

\[ + \sum_{\ell=1}^N \frac{\partial b(X_n(m2^{-n}, x))_{i}}{\partial x_{\ell}} \frac{\partial X_n(m2^{-n}, x)_{\ell}}{\partial x_j} (t - m2^{-n}) \]

for \( m2^{-n} \leq t \leq (m + 1)2^{-n} \). Hence, if we define \( Y_n^{(1)}(t, x) = I \),

\[ \sigma_n^{(1)}: [0, \infty) \times \mathbb{R}^N \times (\mathbb{R}^N \otimes \mathbb{R}^N) \times C([0, \infty); \mathbb{R}^M) \rightarrow \text{Hom}(\mathbb{R}^M, \mathbb{R}^N \otimes \mathbb{R}^N) \]
and

\[ b_n^{(1)} : [0, \infty) \times \mathbb{R}^N \times (\mathbb{R}^N \otimes \mathbb{R}^N) \times C([0, \infty); \mathbb{R}^M) \rightarrow \mathbb{R}^N \otimes \mathbb{R}^N \]

by

\[ \sigma_n^{(1)}(t, x, \bar{x})_{(ij)k} = \sum_{\ell=1}^{N} \frac{\partial \sigma_{ik}}{\partial x_\ell} (X_n([t]_n, x)) x^{(1)}_{\ell j} \]

and

\[ b_n^{(1)}(t, x, \bar{x})_{(ij)} = \sum_{\ell=1}^{N} \frac{\partial b_i}{\partial x_\ell} (X_n([t]_n, x)) x^{(1)}_{\ell j} , \]

then, by (67), \( \sigma_n^{(1)} \), \( b_n^{(1)} \), and \( Y_n^{(1)} \) satisfy the assumptions about \( \tilde{\sigma}_n \), \( \tilde{b}_n \), and \( \tilde{Y}_n \) in Lemma 68 with \( k = 0 \), and

\[ X_n^{(1)}(t, x) = \mathbf{I} + \int_0^t \sigma_n^{(1)}(\tau, x, X_n^{(1)}([\tau]_n, x)) \, dw(\tau) + \int_0^t b_n^{(1)}(\tau, x, X_n^{(1)}([\tau]_n, x)) \, d\tau, \]

where \( X_n^{(1)}(t, x)_{(ij)} = \frac{\partial X_n(t, x)_{(ij)}}{\partial x_k} \). As a consequence, the \( X_n^{(1)} \)'s have the properties proved for the \( \hat{X}_n \)'s in that lemma, and so, by the same argument as we used to prove the corresponding facts for the \( X_n \)'s, we can show that there is a measurable map

\[ X^{(1)} : [0, \infty) \times \mathbb{R}^N \times C([0, \infty); \mathbb{R}^M) \rightarrow \mathbb{R}^N \otimes \mathbb{R}^N \]

such that \( (t, x) \sim X^{(1)}(t, x) \) is continuous, \( X^{(1)}(\cdot, x) \) is a progressively solution to

\[ X^{(1)}(t, x)_{ij} = \mathbf{I} + \sum_{k=1}^{M} \sum_{\ell=1}^{N} \int_0^t \frac{\partial \sigma_{ik}}{\partial x_\ell} (X(\tau, x)) X^{(1)}(\tau, x)_{\ell j} \, dw(\tau)_k \]

\[ + \sum_{\ell=1}^{N} \int_0^t \frac{\partial b_i}{\partial x_\ell} (X(\tau, x)) X^{(1)}(\tau, x)_{\ell j} \, d\tau, \]

and, almost surely, \( X^{(1)} \rightarrow X^{(1)} \) uniformly on compact subsets of \([0, \infty) \times \mathbb{R}^N \). As a result, we now know that \( X(t, \cdot) \) is almost surely continuously differentiable and that \( X^{(1)}(t, x) \) is almost surely equal to its Jacobian matrix. In addition, for each \( p \in [2, 1+\delta] \), there exists a non-decreasing map \( t \in [0, \infty) \mapsto C^{(1)}(t) \in [0, \infty) \) such that

\[ \mathbb{E}^W[\|X_{n+1}^{(1)}(\cdot, x) - X_n^{(1)}(\cdot, x)\|_{[0,t]}^p]^{\frac{1}{p}} \leq C_p^{(1)}(t)^{2^{-\frac{p}{2}}}, \]

\[ \mathbb{E}^W[\|X_n^{(1)}(\cdot, x) - X_n^{(1)}(s, x)\|_{[s,t]}^p]^{\frac{1}{p}} \leq C_p^{(1)}(t-s)^{\frac{1}{2}}, \]

\[ \mathbb{E}^W[\|X_n^{(1)}(\cdot, y) - X_n^{(1)}(\cdot, x)\|_{[s,t]}^p]^{\frac{1}{p}} \leq C_p^{(1)}(t)|y - x|. \]
Next assume that \( \sigma \) and \( b \) have three continuous derivatives, all of which are bounded. Then the \( X_n \) have three continuous spacial derivatives and

\[
X_n^{(2)}(t, x) = \int_0^t \sigma_n^{(2)}(\tau, x, X_n^{(2)}([\tau]_n, x)) \, dw(\tau)
\]

\[
+ \int_0^t b_n^{(2)}(\tau, x, X_n^{(2)}([\tau]_n, x)) \, d\tau + Y_n^{(2)}(t, x),
\]

where \( X_n^{(2)}(t, x)_{ij, k} = \frac{\partial^2 X_n(t, x)}{\partial x_i \partial x_j} \),

\[
\sigma_n^{(2)}(t, x, x^{(2)})_{ij, k} = \sum_{\ell=1}^N \frac{\partial \sigma_{ik}}{\partial x_\ell} (X_n([\tau]_n, x)X_n^{(2)}(\tau, x)_{ij, \ell, k} d\tau)
\]

\[
b_n^{(2)}(t, x, x^{(2)})_{ij, k} = \sum_{\ell=1}^N \frac{\partial b_i}{\partial x_\ell} (X_n([\tau]_n, x)X_n^{(2)}(\tau, x)_{ij, \ell, k} d\tau)
\]

and \( Y_n^{(2)}(t, x)_{ij, k} \) equals

\[
\sum_{k=1}^M \sum_{\ell_1, \ell_2=1}^N \int_0^t \frac{\partial^2 \sigma_{ik}}{\partial x_{\ell_1} \partial x_{\ell_2}} (X_n([\tau]_n, x)) X_n^{(1)}([\tau]_n, x)_{\ell_1, j, k} X_n^{(1)}([\tau]_n, x)_{\ell_2, j, k} d\tau
\]

\[
+ \sum_{\ell_1, \ell_2=1}^N \int_0^t \frac{\partial^2 b_i}{\partial x_{\ell_1} \partial x_{\ell_2}} (X_n([\tau]_n, x)) X_n^{(1)}([\tau]_n, x)_{\ell_1, j, k} X_n^{(1)}([\tau]_n, x)_{\ell_2, j, k} d\tau.
\]

Using (67) and (70), one sees that these quantities again satisfy the conditions in Lemma 68 with \( k = 0 \) and therefore that the analogs of the conclusions just drawn about the \( X_n^{(1)} \)'s hold for the \( X_n^{(2)} \)'s and lead to the existence of an \( X^{(2)} \) with the properties that \( \frac{\partial^2 X_n(t, x)}{\partial x_i \partial x_j} = X^{(2)}(t, x)_{ij, k} \) almost surely and

\[
X^{(2)}(t, x)_{ij, k} = \sum_{k=1}^M \sum_{\ell_1, \ell_2=1}^N \int_0^t \frac{\partial \sigma_{ik}}{\partial x_\ell} (X(\tau, x)) X^{(2)}(\tau, x)_{ij, \ell, k} d\tau
\]

\[
+ \sum_{\ell_1=1}^N \int_0^t \frac{\partial b_i}{\partial x_\ell} (X(\tau, x)) X^{(2)}(\tau, x)_{ij, j, k} d\tau
\]

\[
+ \sum_{k=1}^M \sum_{\ell_1, \ell_2=1}^N \int_0^t \frac{\partial^2 \sigma_{ik}}{\partial x_{\ell_1} \partial x_{\ell_2}} (X(\tau, x)) X^{(1)}(\tau, x)_{\ell_1, j, k} X^{(1)}(\tau, x)_{\ell_2, j, k} d\tau
\]

\[
+ \sum_{\ell_1, \ell_2=1}^N \int_0^t \frac{\partial^2 b_i}{\partial x_{\ell_1} \partial x_{\ell_2}} (X(\tau, x)) X^{(1)}(\tau, x)_{ij, j, k} X^{(1)}(\tau, x)_{ij, j, k} d\tau.
\]
It should be clear that, at the price of introducing increasingly baroque notation but no new ideas, one can prove that, for any $n \geq 1$, $X(t, \cdot)$ has $n$ continuous derivatives if $\sigma$ and $b$ have $n + 1$ bounded continuous ones. In fact, an examination of the argument reveals that we could have afforded the derivatives of order greater than one to have polynomial growth and that the existence of the $(n+1)$st order derivatives could have replaced by a modulus of continuity assumption on the $n$th derivatives.

**An Application to the Martingale Problem:** Recall that to prove that the martingale problem for the operator $L$ have unique solutions it suffices to know that Kolmogorov’s backwards equation (11) has solutions for sufficiently many $\varphi$’s, and we can now show that it does if $\sigma$ and $b$ have three bounded derivatives. Indeed, if $\varphi \in C^2_b(\mathbb{R}^N; \mathbb{R})$ and $u_\varphi(t, x) = \mathbb{E}^W [\varphi(X(t, x))$, then, for any $1 \leq k \leq N$,

$$
u_\varphi(t, x)(t, x + h e_k) - u_\varphi(t, x) = \mathbb{E}^W [\varphi(X(t, x + h e_k)) - \varphi(X(t, x))]$$

where $(e_k)_j = \delta_{k,j}$. Further,

$$
\varphi(X(t, x + h e_k)) - \varphi(X(t, x)) = \sum_{\ell = 1}^h \partial_{x_k} \varphi((t, x + \xi e_k)) X^{(1)}(t, x + \xi e_k)_{\ell j} d\xi,
$$

which, in view of (70), means that $\partial_{x_k} u_\varphi(t, \cdot)$ exists, is continuous, and is given by

$$
\partial_{x_k} u_\varphi(t, x) = \sum_{i=1}^N \mathbb{E}^W [\partial_{x_i} \varphi((t, x)) X^{(1)}(t, x)_{ik}].
$$

Similarly, $\partial_{x_{i_1}} \partial_{x_{i_2}} u_\varphi(t, x)$ exists, is continuous, and is given by

$$
\sum_{i_1, i_2 = 1}^N \mathbb{E}^W [\partial_{x_{i_1}} \partial_{x_{i_2}} \varphi(X(t, x) X^{(1)}(t, x)_{i_1 j_1} X^{(1)}(t, x)_{i_2 j_2}]
+ \sum_{i=1}^N \mathbb{E}^W [\partial_{x_i} \varphi(X(t, x)) X^{(2)}(t, x)_{ij_1 j_2}].
$$

Hence, since, by the Markov property,

$$
u_\varphi(t + h, x) - u(t, x) = \mathbb{E}^W [u_\varphi(t, X(h, x)) - u_\varphi(t, x)] = \int_0^h \mathbb{E}^W [Lu_\varphi(t, X(\xi, x))] d\xi,
$$

it follows that $\partial_t u_\varphi(t, x)$ exist, is continuous, and is equal to $Lu_\varphi(t, x)$. Since two finite Borel measures on $\mathbb{R}^N$ are equal if their integrals of every $\varphi \in C^2_b(\mathbb{R}^N; \mathbb{R})$
agree, we have now proved that the martingale problem for \( L \) has unique solutions.

It turns out that one can do better. In fact, using no probability theory and only clever applications of the minimum principle, O. Olenik proved that, when \( \phi \in C^2_b(\mathbb{R}^N; \mathbb{R}) \), Kolmogorov’s backward equation can be solved if \( a \) and \( b \) are bounded and have two, bounded, continuous derivatives, and therefore the corresponding martingale problem has unique solutions under those assumptions. Her result dramatizes a basic weakness of the Itô theory. Namely, having to find a good \( \sigma \) is a serious drawback.

**Itô’s Exponential**

**A Simple Stochastic Integral Equation:** Let \((B(t), \mathcal{F}_t, \mathbb{P})\) be an \( \mathbb{R}^N \)-valued Brownian motion on \((\Omega, \mathcal{F}, \mathbb{P})\), and, given a \( \xi \in P^2_{\text{loc}}(\mathbb{R}^N) \) and \( \Phi \in P^2_{\text{loc}}(\mathbb{R}) \), consider the stochastic integral equation

\[
X(t) = 1 + \int_0^t X(\tau)(\xi(\tau), dB(\tau))_{\mathbb{R}^N} + \int_0^t X(\tau)\Phi(\tau)\, d\tau,
\]

Using Itô’s formula, it is easy to check that

\[
E_{\xi, \Phi}(t) = \exp \left( \int_0^t (\xi(\tau), dB(\tau))_{\mathbb{R}^N} + \frac{1}{2} \int_0^t (\Phi(\tau) - \frac{1}{2} |\xi(\tau)|^2)\, d\tau \right)
\]

is a solution to (71). In addition, if \( X(\cdot) \) were a second solution, then an application of Itô’s formula to \( \frac{dX(t)}{E_{\xi, \Phi}(t)} \) shows that \( X(\cdot) = E_{\xi, \Phi}(\cdot) \).

There are many applications of this observation, most of which exploit \( E_{\xi, \Phi}(\cdot) \) as an integrating factor for solving various differential equations. Such applications are based on the fact proved in the following lemma.

**Lemma 73.** If \( V \in P^2_{\text{loc}}(\mathbb{R}^N) \), \( \sigma \in P^2_{\text{loc}}(\text{Hom}(\mathbb{R}^M; \mathbb{R}^N)) \) are as in (57), then, for every \( \varphi \in C^{1,2}(\mathbb{R}^{N_1} \times \mathbb{R}^{N_2}; \mathbb{R}) \),

\[
\left( E_{\xi, \Phi}(t)\left( E_{0, \Phi}(t)\varphi(V(t), I_\sigma(t)) - \int_0^t E_{0, \Phi}(\tau)\left( \Phi(\tau)\varphi(V(\tau), I_\sigma(\tau)) + \left( \sigma(\tau)\xi(\tau), \nabla_{(2)}\varphi(V(\tau), I_\sigma(\tau)) \right)_{\mathbb{R}^{N_2}} + \left( \nabla_{(1)}\varphi(V(\tau), I_\sigma(\tau)), dV(\tau) \right)_{\mathbb{R}^{N_1}} \right) \, d\tau \right)_{\mathbb{F}_t, \mathbb{P}} \right)
\]

is a local martingale.

**Proof:** Using stopping times and standard approximation methods, one can reduce to the case in which \( \sigma \) and \( \Phi \) are bounded and \( \varphi \in C^{1,2}_b(\mathbb{R}^{N_1} \times \mathbb{R}^{N_2}; \mathbb{R}) \).
In addition, one can assume that \( V(\cdot) \) has a bounded continuous derivative \( \dot{V} \). Thus I will proceed under these assumptions.

By (57),
\[
M(t) \equiv \varphi(V(t), I_\sigma(t))
- \int_0^t \left( ((\nabla_1) \varphi(V(\tau), I_\sigma(\tau)), \dot{V}(\tau)) \right)_{\mathbb{R}^{N_2}}
+ \frac{1}{2} \text{Trace} \left( \sigma \sigma^\top \nabla^2_2 \varphi(V(\tau), I_\sigma(\tau)) \right) d\tau
= \int_0^t (\sigma(\tau)^\top \nabla_2 \varphi(V(\tau), I_\sigma(\tau), dB(\tau))_{\mathbb{R}^{N_2}},
\]
and therefore, again by (57),
\[
\left( E_{\xi, \Phi}(t) M(t) - \int_0^t E_{\xi, \Phi}(\tau) \left( \Phi(\tau) \varphi(V(\tau), I_\sigma(\tau))
+ \left( \sigma(\tau) \xi(\tau), \nabla_2 \varphi(V(\tau), I_\sigma(\tau)) \right)_{\mathbb{R}^{N_2}} \right) d\tau, F_t, \mathbb{P} \right)
\]
is a martingale. Equivalently, if
\[
H(t) = \Phi(t) \varphi(V(t), I_\sigma(t)) + ((\nabla_1) \varphi(V(t), I_\sigma(t)), \dot{V}(t))_{\mathbb{R}^{N_1}}
+ \left( \sigma(t) \xi(t), \nabla_2 \varphi(V(t), I_\sigma(t)) \right)_{\mathbb{R}^{N_2}}
+ \frac{1}{2} \text{Trace} \left( \sigma(\tau)^\top \nabla^2_2 \varphi(V(\tau), I_\sigma(\tau)) \right),
\]
then
\[
\left( E_{\xi, \Phi}(t) \varphi(V(t), I_\sigma(t)) - \int_0^t E_{\xi, \Phi}(\tau) H(\tau) d\tau, F_t, \mathbb{P} \right)
\]
is a martingale, and therefore
\[
\mathbb{E}^\mathbb{P} \left[ E_{\xi, \Phi}(t) \varphi(V(t), I_\sigma(t)) - E_{\xi, \Phi}(s) \varphi(V(s), I_\sigma(s)) \mid F_s \right]
= \mathbb{E}^\mathbb{P} \left[ \int_s^t E_{\xi, \Phi}(\tau) H(\tau) d\tau \mid F_s \right]
= \mathbb{E}^\mathbb{P} \left[ E_{\xi, 0}(t) \int_s^t E_{0, \Phi}(\tau) H(\tau) d\tau \mid F_s \right]
= \mathbb{E}^\mathbb{P} \left[ E_{\xi, 0}(t) \int_0^t E_{0, \Phi}(\tau) H(\tau) d\tau \mid F_s \right] - \mathbb{E}^\mathbb{P} \left[ E_{\xi, 0}(s) \int_0^s E_{0, \Phi}(\tau) H(\tau) d\tau \right].
\]

Most applications are to cases in which either \( \xi = 0 \) or \( \Phi = 0 \). In the following section I will explain some of its applications when \( \xi = 0 \), and the section after that deals with the case when \( \Phi = 0 \).
The Feynman-Kac Formula: To see how Lemma 73 gets applied when \( \xi = 0 \), suppose that \( X(\cdot, x) \) is the solution to (56) and that \( F: \mathbb{R}^N \to \mathbb{R} \) is a Borel measurable function that is bounded on compact subsets. Then, by taking \( \Phi(t) = \int_0^t F(X(t, x)) \, d\tau \), we see that, for each \( T > 0 \),

\[
\left( e^{\int_0^t F(X(\tau \wedge T, x)) \, d\tau} u(T - t \wedge T, X(t \wedge T, x)), \mathcal{F}_t, \mathbf{W} \right)
\]

is a local martingale if \( u \in C^1_2((0, \infty) \times \mathbb{R}^N; \mathbb{R}) \cap C([0, \infty) \times \mathbb{R}^N; \mathbb{R}) \) satisfies \( \partial_t u = L\varphi + F\varphi \) on \((0, \infty) \times \mathbb{R}^N \). Hence, if \( F \) is bound above and \( u \) is bounded on \([0, t] \times \mathbb{R}^N \), then

\[
(74) \quad u(t, x) = \mathbb{E}^W \left[ \exp \left( \int_0^t F(X(\tau, x)) \, d\tau \right) u(0, X(t, x)) \right].
\]

The equation in (74) is one version of the renowned **Feynman-Kac formula**, whose power is demonstrated in the following computation. One of the few non-trivial evolution equations for which one can write down explicit solutions is the equation

\[
(75) \quad \partial_t u = \frac{1}{2} \left( \partial^2_x u - x^2 u \right) \text{ on } (0, \infty) \times \mathbb{R} \text{ with } \lim_{t \to 0} u(t, \cdot) = \varphi.
\]

Indeed, if

\[ h(t, x, y) = e^{-\frac{1}{2}x^2} g(1 - e^{-2t}, y - e^{-t}x) e^{\frac{y^2}{2}} \]

where \( g(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2}} \),

then an elementary calculation shows that for any \( \varphi \in C(\mathbb{R}; \mathbb{R}) \) with at most exponential growth,

\[ u_\varphi(t, x) = \int_{\mathbb{R}} h(t, x, y) \varphi(y) \, dy \]

is a solution to (75). In particular, by taking \( \varphi = 1 \), one sees that

\[
\mathbb{E}^W \left[ \exp \left( -\frac{1}{2} \int_0^t (x + w(\tau))^2 \, d\tau \right) \right] = \frac{1}{\sqrt{\cosh t}} e^{-\frac{\tanh t}{2} x^2}.
\]

Further, because, for each \( \alpha > 0 \), \( t \sim w(\alpha t) \) has the same distribution as \( t \sim \alpha^{\frac{1}{2}} w(t) \), one knows that

\[
\exp \left( -\frac{\alpha}{2} \int_0^t (x + w(\tau))^2 \, d\tau \right) \text{ and } \exp \left( -\frac{1}{2} \int_0^{\alpha\frac{1}{2} t} (\alpha^{\frac{1}{2}} x + w(\tau))^2 \, d\tau \right)
\]
have the same distribution. Hence,

\[
E^W\left[ \exp\left(-\frac{\alpha}{2} \int_0^t (x + w(\tau))^2 \, d\tau \right) \right] = \left( \cosh(\alpha \frac{1}{2} t) \right)^{-\frac{1}{2}} \exp\left(-\frac{\alpha^2 \tanh(\alpha \frac{1}{2} t)}{2} x^2 \right).
\]

(76)

This computation provides quantitative evidence of the fuzzyness of Brownian paths. Namely, take \( t = 1 \) and integrate both sides of (76) with respect to \( x \) to obtain

\[
E^W\left[ \exp\left(-\frac{\alpha}{2} \left( \int_0^1 w(\tau)^2 \, d\tau - \left( \int_0^1 w(\tau) \, d\tau \right)^2 \right) \right) \right] = \sqrt{\frac{\alpha^2}{\sinh(\alpha \frac{1}{2})}},
\]

and then, again using Brownian scaling,

\[
E^W\left[ \exp\left(-\frac{\alpha}{2} \text{Var}_{[0,1]}(w(\cdot)) \right) \right] = \sqrt{\frac{(\alpha t)^{\frac{1}{2}}}{\sinh(\alpha t)^{\frac{1}{2}}}},
\]

(77)

where \( \text{Var}_{[0,1]}(w(\cdot)) = \frac{1}{t} \int_0^t w(\tau)^2 \, d\tau - \left( \frac{1}{t} \int_0^t w(\tau) \, d\tau \right)^2 \)

is the variance of the path \( w \mid [0,t] \). From this it follows that

\[
W\left( \text{Var}_{[0,1]}(w(\cdot)) \geq \frac{1}{r} \right) \leq e^{-\frac{\alpha^2}{2} t - \frac{2 \sqrt{\alpha t}}{1 - e^{-2 \sqrt{\alpha t}}}}
\]

for all \( \alpha > 0 \), and so, by taking \( \alpha = \frac{t^2}{e} \), we have that

\[
W\left( \text{Var}_{[0,1]}(w(\cdot)) \geq \frac{1}{r} \right) \leq e^{-\frac{rt}{1 - e^{-rt}}}.
\]

(78)

In other words, the probability that the a Brownian path has small variance is exponentially small.

**The Cameron-Martin Formula**: Applications of Lemma 73 when \( \Phi = 0 \) are a bit more complicated because, in order to be useful, one has to know when \( (E_{\xi,0}(t), F_t, P) \) is actually a martingale and not just a local one. For this reason, the following criterion, which was discovered by A. Novikov, can help.

**Lemma 79.** For any \( \xi \in P_{\text{loc}}^2(\mathbb{R}^N) \), \( (E_{\xi,0}(t), F_t, P) \) is a supermartingale, and it is a martingale if and only if \( E^P[|E_{\xi,0}(t)|] = 1 \) for all \( t \geq 0 \). Next, set

\[
A_{\xi}(t) \equiv \int_0^t |\xi(\tau)|^2 \, d\tau.
\]

If

\[
E^P[ e^{\frac{1}{2} A_{\xi}(t)} ] < \infty \quad \text{for all} \quad t \geq 0,
\]

then \( (E_{\xi,0}(t), F_t, P) \) is a martingale.
Proof: Choose stopping times \( \{ \zeta_m : m \geq 1 \} \) for \(( E_{\xi,0}(t), \mathcal{F}_t, \mathbb{P} )\). Then

\[
E_{\xi,0}(t \wedge \zeta_m) \rightarrow E_{\xi,0}(t) \text{ (a.s.,} \mathbb{P} \text{) and } \mathbb{E}^\mathbb{P} \left[ E_{\xi,0}(t \wedge \zeta_m) \right] = 1
\]

for all \( t \geq 0 \). Thus, since \(( E_{\xi,0}(t \wedge \zeta_m), \mathcal{F}_t, \mathbb{P} )\) is a martingale for each \( m \geq 1 \), Fatou’s lemma implies that \(( E_{\xi,0}(t), \mathcal{F}_t, \mathbb{P} )\) is a supermartingale. Moreover, if \( \mathbb{E}^\mathbb{P} \left[ E_{\xi,0}(t) \right] = 1 \), then \( E_{\xi,0}(t \wedge \zeta_m) \rightarrow E_{\xi,0}(t) \) in \( L^1(\mathbb{P} ; \mathbb{R}) \), and so \(( E_{\xi,0}(t), \mathcal{F}_t, \mathbb{P} )\) is a martingale. Since the converse implication in the second one is trivial, we have proved the first and second assertions.

Now make the assumption in the concluding assertion. Because, for any \( m \geq 1 \),

\[
\mathbb{E}^\mathbb{P} \left[ E_{\pm \xi,0}(t \wedge \zeta_m) \right] \leq 1,
\]

and therefore \( \sup_{m \geq 1} \mathbb{E}^\mathbb{P} \left[ e^{\pm \frac{1}{2} \lambda(t \wedge \zeta_m)} \right] < \infty \) is integrable for all \( t \geq 0 \). In particular, this means that \(( I_\xi(t), \mathcal{F}_t, \mathbb{P} )\) is a martingale and therefore that \(( e^{\alpha I_\xi(t)}, \mathcal{F}_t, \mathbb{P} )\) is a submartingale for all \( \alpha \geq 0 \). Next, for \( \lambda \in (0,1) \), determine \( p_\lambda \in (1, \infty) \) by the equation \( 2p_\lambda \lambda(p_\lambda - \lambda) = 1 - \lambda^2 \), and for \( p \in [1, p_\lambda] \), set \( \alpha(p, \lambda) = \frac{p(p_\lambda - \lambda)}{1 - \lambda^2} \). Then \( E_{\lambda,0}(t)^p = (e^{\alpha(p, \lambda) I_\xi(t)})^{1-\lambda^2} E_{p,0}(t)^\lambda I_\xi(t) \), and so, by Hölder’s inequality and the submartingale property of \( e^{\alpha(p, \lambda) I_\xi(t)} \), for any stopping time \( \zeta \),

\[
(* ) \quad \mathbb{E}^\mathbb{P} \left[ E_{\lambda,0}(t \wedge \zeta)^p \right] \leq \mathbb{E}^\mathbb{P} \left[ e^{\alpha(p, \lambda) I_\xi(t)} \right]^{1-\lambda^2} \mathbb{E}^\mathbb{P} \left[ E_{p,0}(t \wedge \zeta) \right]^{\lambda^2}.
\]

Since \( \mathbb{E}^\mathbb{P} \left[ E_{p,0}(t \wedge \zeta) \right] \leq 1 \), by taking \( p = p_\lambda \), we see that

\[
\mathbb{E}^\mathbb{P} \left[ E_{\lambda,0}(t \wedge \zeta)^{p_\lambda} \right] \leq \mathbb{E}^\mathbb{P} \left[ e^{\frac{1}{2} I_\xi(t)} \right],
\]

from which it follows that the local martingale \(( E_{\lambda,0}(t), \mathcal{F}_t, \mathbb{P} )\) is a martingale for each \( \lambda \in (0,1) \). Finally, take \( p = 1 \) and \( \zeta = t \). Then \((*)\) combined with Jensen’s inequality say that

\[
\mathbb{E}^\mathbb{P} \left[ E_{\lambda,0}(t) \right] \leq \mathbb{E}^\mathbb{P} \left[ e^{\frac{1}{2} I_\xi(t)} \right]^{2\lambda(1-\lambda)} \mathbb{E}^\mathbb{P} \left[ E_{\xi,0}(t) \right]^{\lambda^2}
\]

for all \( \lambda \in (0,1) \). Thus, because \( \mathbb{E}^\mathbb{P} \left[ E_{\lambda,0}(t) \right] = 1 \), after letting \( \lambda \uparrow 1 \), one sees that \( \mathbb{E}^\mathbb{P} \left[ E_{\xi,0}(t) \right] = 1 \). \( \square \)

Novikov’s criterion is useful but not definitive. Indeed, consider the case when \( N = 1 \) and \( \xi(t) = B(t) \). Then

\[
\xi(t) = \int_0^t B(\rho) \, dB(\tau) = \frac{B(t)^2 - t}{2} \text{ and } A_\xi(t) = \int_0^t B(\tau)^2 \, d\tau.
\]
As we showed $E^p[e^{\frac{1}{2} B}] \leq E^p[e^{\frac{1}{2} A}] \frac{1}{2}$, and so, since $E^p[e^{\frac{1}{2} B(t)^2}] = \infty$ if $t \geq 2$, $E^p[e^{\frac{1}{2} A}] = \infty$ for $t \geq 2$. Nonetheless, $(E_{\xi,0}(t), F_t, P)$ is a martingale. To see this, choose a sequence $\{\varphi_k : k \geq 1\} \subseteq C_b(\mathbb{R}; [0, \infty))$ such that $\varphi_k(y) \nearrow e^{\frac{y^2}{2}}$. Then, by the application of the Feynman-Kac formula given above (cf. the notation there),

$$E^p[\xi_0(t)] = e^{-\frac{1}{2} \lim_{k \to \infty} E^p[e^{-\frac{1}{2} A_k(t)} \varphi_k(B(t))]} = e^{-\frac{1}{2} \lim_{k \to \infty} \int h(t, 0, y) \varphi_k(y) dy} = e^{-\int g(1 - e^{-2t}, y)e^{y^2} dy} = 1.$$

In case one is thinking that it is the sign of the $B(t)^2$ in the exponential that saves the day, do a similar calculation with $\xi$ replaced by $-\xi$ and arrive at the same conclusion.

The main application to which we will put these considerations requires the use of the following lemma.

**Lemma 80.** Suppose that $\{\mathbb{P}_k : k \geq 1\}$ is a sequence of Borel probability measures on $C([0, \infty); \mathbb{R}^N)$ with the property that $\mathbb{P}_{k+1} \upharpoonright B_k = \mathbb{P}_k \upharpoonright B_k$ for all $k \geq 1$. Then there is a unique Borel probability measure $\mathbb{P}$ on $C([0, \infty); \mathbb{R}^N)$ such that $\mathbb{P} \upharpoonright B_k = \mathbb{P}_k \upharpoonright B_k$ for all $k \geq 1$.

**Proof:** This is a relatively straightforward application of the fact that a family of Borel probability measures on a Polish space $E$ is relatively compact in the topology of weak convergence if and only if it is tight in the sense that for each $\epsilon > 0$ there is a compact subset $K \subseteq E$ such that $K \mathbb{C}$ is assigned measure less that $\epsilon$ by all members of the family. To apply this result here, one also has to know that $K \subseteq C([0, \infty); \mathbb{R}^N)$ is compact if and only if $\{\psi \upharpoonright [0, k] : \psi \in K\}$ is compact in $C([0, k]; \mathbb{R}^N)$ for each $k \geq 1$.

The uniqueness assertion is trivial. To prove the existence, let $\epsilon > 0$ be given, and for each $k \geq 1$ choose a compact subset $K_k$ of $C([0, k]; \mathbb{R}^N)$ such that

$$\max_{1 \leq j \leq k} \mathbb{P}_j(\{\psi : \psi \upharpoonright [0, k] \notin K_k\}) \leq 2^{-k} \epsilon,$$

in which case

$$\sup_{j \geq 1} \mathbb{P}_j(\{\psi : \psi \upharpoonright [0, k] \notin K_k\}) \leq 2^{-k} \epsilon.$$

Finally, set

$$K = \{\psi : \psi \upharpoonright [0, k] \in K_k \text{ for all } k \geq 1\}.$$

Then $K$ is a compact subset of $C([0, \infty); \mathbb{R}^N)$, and

$$\mathbb{P}_j(K) \leq \sum_{k=1}^{\infty} \mathbb{P}_j(\{\psi : \psi \upharpoonright [0, k] \notin K_k\}) \leq \epsilon.$$
Hence, the sequence \( \{ \mathbb{P}_k : k \geq 1 \} \) is relatively compact. Furthermore, any limit \( \mathbb{P} \) will have the property that \( \mathbb{P} \upharpoonright \mathcal{B}_k = \mathbb{P}_k \upharpoonright \mathcal{B}_k \) for all \( k \geq 1 \). Therefore the sequence converges and its limit has the required property. \( \square \)

Let \( X(\cdot, x) \) be the solution to (56) and \( \beta: \mathbb{R}^N \rightarrow \mathbb{R}^M \) a continuous function. Set

\[
E_\beta(t, x) = \exp \left( \int_0^t \left( \beta(X(\tau, x)), dw(\tau) \right)_{\mathbb{R}^M} - \int_0^t |\beta(X(\tau, x))|^2 d\tau \right),
\]

and assume that \( (E_\beta(t, x), \mathcal{W}_t, \mathcal{W}) \) is a martingale. If \( \mathbb{P}_k \) is the Borel probability measure on \( C([0, \infty); \mathbb{R}^M) \) given by

\[
\mathbb{P}_k(\Gamma) = \mathbb{E}^W[E_\beta(k, x), \Gamma],
\]

then \( \mathbb{P}_{k+1} \upharpoonright \mathcal{W}_k = \mathbb{P}_k \upharpoonright \mathcal{W}_k \), and so, by Lemma 80, there is a unique \( \mathcal{W}_{\beta, x} \) with the property that

\[
(81) \quad \mathcal{W}_{\beta, x}(\Gamma) = \mathbb{E}^W[E_\beta(t, x), \Gamma] \quad \text{for all} \quad t \geq 0 \quad \text{and} \quad \Gamma \in \mathcal{W}_t.
\]

Furthermore, if

\[
L_\beta \varphi = L \varphi + (\sigma \beta, \nabla \varphi)_{\mathbb{R}^N},
\]

then, by Lemma 73,

\[
\left( \varphi(X(t, x)) - \int_0^t L_\beta \varphi(X(\tau, x)) d\tau, \mathcal{W}_t, \mathcal{W}_{\beta, x} \right)
\]

is a martingale for all \( \varphi \in C_c^2(\mathbb{R}^N; \mathbb{R}) \). Hence the distribution \( \mathbb{P}_{\beta, x} \) of \( X(\cdot, x) \) under \( \mathcal{W}_{\beta, x} \) is a solution starting at \( x \) to the martingale problem for \( L_\beta \). Now assume that \( \sigma \beta \) is Lipschitz continuous, and let \( X_\beta(\cdot, x) \) be the solution to (56) with \( b + \sigma \beta \) replacing \( b \). Then the distribution of \( X_\beta(\cdot, x) \) is also a solution starting at \( x \) to the martingale problem for \( L_\beta \), and so, if there is only one solution to that martingale problem, then

\[
(82) \quad \mathcal{W}(X_\beta(\cdot, x) \in \Gamma) = \mathbb{E}^W[E_\beta(t, x), X(\cdot, x) \in \Gamma] \quad \text{for all} \quad t \geq 0 \quad \text{and} \quad \Gamma \in \mathcal{B}_t.
\]

The first examples of this type of relationship were discovered by Cameron and Martin, and so I will call the equation in (82) the **Cameron-Martin formula**, even though, because Girsanov introduced a significant generalization of it, it is usually called **Girsanov’s formula**.

There are many ways in which (82) can be used. For instance, it provides an explanation for the conclusion that we reached prior to Lemma 80. To understand how it relates to that discussion, assume that \( \sigma \) is Lipschitz continuous but that \( b = \sigma \beta \) is only locally Lipschitz continuous. Choose an \( \eta \in C^\infty(\mathbb{R}^N, [0, 1]) \).
\[ L_k \varphi = (b_k, \nabla \varphi)_\mathbb{R}^n + \frac{1}{2} \text{Trace}(\sigma \sigma^t \nabla^2 \varphi), \]

and assume that, for each \( k \geq 1 \), the martingale problem for \( L_k \) has only one solution starting at \( x \). Then if \( X(\cdot, x) \) is the solution to (56) when \( b = 0 \), (82) says that

\[ \mathbb{E}_W^W [E_{k,t}(t, x), X(\cdot, x) \in \Gamma] = \mathcal{W}(\tilde{X}_k(\cdot, x) \in \Gamma) \text{ for } t \geq 0 \text{ and } \Gamma \in \mathcal{B}_t, \]

where

\[ E_k(t, x) = \exp \left( \int_0^t \left( \beta_k(X(\tau, x)), dw(\tau) \right)_{\mathbb{R}^M} - \frac{1}{2} \int_0^t \left| \beta_k(\tau) \right|^2 d\tau \right) \]

with \( \beta_k(y) = \eta(k^{-1}y)b(y) \).

In particular, if \( \zeta_k = \inf\{ t \geq 0 : |X(t, x)| \geq k \} \), then

\[ E_k(t, x) = E_{\beta}(t, x) \text{ (a.s.,} \mathcal{W}) \text{ for } t \in [0, \zeta_k), \]

and so

\[ \mathbb{E}_W^W [E_{\beta}(t, x), \zeta_k > t] = \mathcal{W}(\tilde{\zeta}_k > t). \]

Finally, since \( \zeta_k \not\rightarrow \infty \text{ (a.s.,} \mathcal{W}), \) we have shown that

\[ \mathbb{E}_W^W [E_{\beta}(t, x)] = \mathcal{W}(\tilde{\zeta} > t). \]

In other words, the expected value of \( E_{\beta}(t, x) \) is the probability that \( \tilde{X}(\cdot, x) \) hasn’t exploded by time \( t \). Thus, \( \tilde{X}(\cdot, x) \) never explodes \( \text{i.e.,} \mathcal{W}(\tilde{\zeta} = \infty) = 1 \) if and only if \( (E_{\beta}(t, x), W_t) \) is a martingale. Since \( \tilde{X}(\cdot, x) \) will never explode if \( |b(y)| \leq C(1 + |y|) \), this explains our earlier conclusion that \( (E_\xi(t), \mathcal{F}_t, \mathbb{P}) \) is a martingale when \( \xi(t) = B(t) \). Indeed, after translating it into the present setting, one sees that we were dealing with \( \sigma = 1, b(y) = y, \) and \( x = 0 \). When \( \xi(t) = -B(t), \tilde{X}(t, x) \) is the Ornstein-Uhlenbeck process.
The Cameron-Martin-Segal Theorem: One of the most important applications of the ideas discussed in the preceding section has its origins in the early work of Cameron and Martin and reached its final form in the work of I. Segal. To fully appreciate what they did, one should know about the following theorem of Sudakov.

**Theorem 83.** Let $E$ be an infinite dimensional, separable Banach space over $\mathbb{R}$, and, for $a \in E$, define the translation map $T_a : E \to E$ by $T_a = x + a$. Given $\mu$ a Borel probability measure on $E$, there is a dense subset $D$ such that $(T_{-a})_*\mu \perp \mu$ for all $a \in D$.

To prove this theorem, we will use the following lemma.

**Lemma 84.** If $K_1$ and $K_2$ are compact subsets of $E$, then $\text{int}(K_2 - K_1) = \emptyset$.

**Proof:** Since $K_2 - K_1$, as the image of a compact set under a continuous map, is compact if $K_1$ and $K_2$ are compact, it suffices for us to prove that $\text{int}(K) = \emptyset$ for every compact $K$.

Given $K$ and $r > 0$, choose $x_1, \ldots, x_n \in K$ so that $K \subseteq \bigcup_{m=1}^n B_E(x_m, \frac{r}{2})$. Because $E$ is infinite dimensional, the Hahn-Banach theorem guarantees that there is an $a^* \in E^*$ such that $\|a^*\|_{E^*} = 1$ and $\langle x_m, a^* \rangle = 0$ for $1 \leq m \leq n$. Now choose $a \in E$ so that $\|a\|_E = r$ and $\langle a, a^* \rangle \geq \frac{3r}{4}$. If $x$ were in $K \cap (a + K)$, then there would exist $1 \leq m, m' \leq n$ such that $\|x - x_m\|_E < \frac{r}{4}$ and $\|x - a - x_{m'}\|_E < \frac{r}{4}$, which would lead to the contradiction that

$$\langle x, a^* \rangle = \langle x - x_m, a^* \rangle < \frac{r}{4} \quad \text{and} \quad \langle x, a^* \rangle = \langle a, a^* \rangle + \langle x - a - x_{m'}, a^* \rangle \geq \frac{r}{4}.$$

Thus $K \cap (a + K) = \emptyset$, and so $a \notin K - K$, which means that $K - K$ contains no non-empty ball centered at the origin. Finally, if $K \supseteq B_E(b, r)$ for some $b \in E$ and $r > 0$, then $B_E(0, r) \subseteq K - K$, which cannot be. \(\square\)

**Proof of Theorem 83:** Let $\mu$ be a Borel probability measure on $E$. By Ulam’s lemma, for each $m \geq 1$ there exists a compact set $K_m$ such that $\nu(K_m) \geq 1 - \frac{1}{m}$.

Set $A = \bigcup_{m=1}^\infty K_m$ and $B = A - A = \bigcup_{m,n=1}^\infty (K_n - K_m)$. Obviously, $\mu(A) = 1$.

By Lemma 84, $\text{int}(K_n - K_m) = \emptyset$ for all $m, n \geq 1$, and so, by the Baire category theorem, $\text{int}(B) = \emptyset$. Now suppose that $a \notin B$. Then $A \cap (a + A) = \emptyset$, and so $(T_{-a})_*\mu(A) = \mu(a + A) = 0$. Hence, since $a \notin B \iff -a \notin B$, we have shown that $(T_{-a})_*\mu \perp \mu$ for all $a \in D = BC$. \(\square\)

As Theorem 83 makes clear, translation of a Borel probability measure on an infinite dimensional Banach space in most directions will produce a measure that is singular to the original one, and it is not obvious that there are any measures on an infinite dimensional Banach space that are quasi-invariant under translation in a dense set of directions. Thus, it was a significant discovery when Cameron and Martin showed that Wiener measure is quasi-invariant under translation by absolutely continuous paths whose derivatives are square integrable and, later, Segal showed that translates of Wiener by any other paths are singular.
To put these results into the context of Theorem 83, let \( \Omega \) be the space of \( w \in C([0, \infty) ; \mathbb{R}^M) \) with the properties that \( w(0) = 0 \) and \( \lim_{t \to \infty} |w(t)| = 0 \), and observe that \( \mathcal{W}(\Omega) = 1 \). Next define the norm \( \|w\|_\Omega = \sup_{t \geq 0} \frac{|w(t)|}{1+t} \). Then it is not hard to show that \( \Omega \) with this norm is a separable Banach space on which \( \mathcal{W} \) is a Borel probability measure. Finally, denote by \( H^1(\mathbb{R}^M) \) the subset of absolutely continuous \( h \in C([0, \infty) ; \mathbb{R}^M) \) with the properties that \( h(0) = 0 \) and \( h \in L^2((0, \infty) ; \mathbb{R}^*M) \). That is, \( h(t) = \int_0^t \dot{h}(\tau) \, d\tau \) where \( \dot{h} \in L^2((0, \infty) ; \mathbb{R}^M) \). It is easy to check that \( H^1(\mathbb{R}^M) \) becomes a separable Hilbert space if one uses the inner product

\[
(g, h)_{H^1(\mathbb{R}^M)} = (\dot{g}, \dot{h})_{L^2((0, \infty) ; \mathbb{R}^M)}.
\]

In addition, since \( |\dot{h}(t)| \leq t^{\frac{1}{2}} \|\dot{h}\|_{H^1(\mathbb{R}^M)}, \ h \in \Omega \) and \( \|h\|_\Omega \leq \frac{1}{2} \|\dot{h}\|_{H^1(\mathbb{R}^M)} \). Finally, it is obvious that \( H^1(\mathbb{R}^M) \) is a dense subspace of \( \Omega \).

**Theorem 85.** Think of \( \mathcal{W} \) as a Borel measure on \( \Omega \). Given \( h \in H^1(\mathbb{R}^M) \), set

\[
I(h) = \int_0^\infty (\dot{h}(\tau), dw(\tau))_{\mathbb{R}^M} \text{ and } R_h = \exp(I(h) - \frac{1}{2} \|\dot{h}\|_{H^1(\mathbb{R}^M)}^2).
\]

Then \( (T_h)_* \mathcal{W} \ll \mathcal{W} \) and \( \frac{d(T_h)_* \mathcal{W}}{d\mathcal{W}} = R_h \) for all \( h \in H^1(\mathbb{R}^M) \). On the other hand, if \( f \in \Omega \setminus H^1(\mathbb{R}^M) \), then \( (T_f)_* \mathcal{W} \perp \mathcal{W} \).

A proof of the first assertion can be based on the considerations in the preceding section, but there is a more elementary, direct proof. Namely, given \( n \geq 1, 0 \leq t_1 < \cdots < t_n \), and \( \xi_1, \ldots, \xi_n \in \mathbb{R}^M \), it is obvious that

\[
\mathbb{E}[T_h)_* \mathcal{W} \left( e^{i \sum_{m=1}^n (\xi_m, w(t_m))_{\mathbb{R}^M}} \right] = \exp \left( i \sum_{m=1}^n (\xi_m, h(t_m))_{\mathbb{R}^M} - \frac{1}{2} \sum_{m,m'=1}^n (t_m \wedge t_{m'})(\xi_m, \xi_{m'})_{\mathbb{R}^M} \right).
\]

To compute the same integral with respect to \( R_h d\mathcal{W} \), set \( h_m(t) = (t \wedge t_m) \xi_m \) and \( f(t) = \dot{h}(t) + i \sum_{m=1}^n h_m(t) \). Then

\[
R_h(w) e^{i \sum_{m=1}^n (\xi_m, w(t_m))_{\mathbb{R}^M}} = \exp \left( \int_0^\infty (f(\tau), dw(\tau))_{\mathbb{R}^M} - \frac{1}{2} \|\dot{h}\|_{H^1(\mathbb{R}^M)}^2 \right).
\]
and so

\[
\mathbb{E}^W [R_h(w) e^{\frac{1}{2} \sum_{m=1}^n (\xi_m, w(t_m))_{R^M}}] \\
= \exp \left( -\frac{i}{2} \sum_{m=1}^n \int_0^\infty f(\tau)^2 d\tau - \|h\|_{H^1(R^M)}^2 \right) \\
= \exp \left( i \sum_{m=1}^n \langle h_m, \ddot{h} \rangle_{H^1(R^M)} - \frac{1}{2} \sum_{m,m'=1}^n \langle h_m, h_{m'} \rangle_{H^1(R^M)} \right) \\
= \exp \left( i \sum_{m=1}^n \langle \xi_m, \ddot{h}(t_m) \rangle_{R^M} - \frac{1}{2} \sum_{m,m'=1}^n \langle t_m \wedge t_{m'} \rangle (\xi_m, \xi_{m'})_{R^M} \right).
\]

Hence \( d(T_h) W = R_h dW \).

The proof of the second assertion requires some preparations. Let \( L \) denote the subspace of \( H^1(R^M) \) consisting of twice continuously differentiable functions whose first derivatives have compact support. Clearly \( L \) is dense in \( H^1(R^M) \), and so one can find an orthonormal \( \{h_m : m \geq 1\} \subseteq L \) for \( H^1(R^M) \). Now define the linear functional \( \Lambda \) on \( L \) by

\[
\Lambda(h) = -\int_0^\infty \langle f(\tau), \ddot{h}(\tau) \rangle_{R^M} d\tau.
\]

We need to show that \( \sum_{m=1}^\infty \Lambda(h_m)^2 = \infty \). To this end, suppose that \( C = \frac{1}{\sqrt{\sum_{m=1}^\infty \Lambda(h_m)^2}} \) < \( \infty \). Then, if \( h \) is in the span of \( \{h_m : m \geq 1\} \),

\[
|\Lambda(h)| \leq \sum_{m=1}^\infty |\Lambda(h_m)| |\langle h, h_m \rangle_{H^1(R^M)}| \leq C \|h\|_{H^1(R^M)},
\]

and so \( \Lambda \) admits a unique extension as a continuous linear functional on \( H^1(R^M) \). Thus, by the Riesz representation theorem for Hilbert space, there exists an \( h_0 \in H^1(R^M) \) such that \( \Lambda(h) = \langle h, h_0 \rangle_{H^1(R^M)} \), and so

\[
\int_0^\infty \langle f(\tau), \ddot{h}(\tau) \rangle_{R^M} d\tau = \int_0^\infty \langle h_0(\tau), \ddot{h}(\tau) \rangle_{R^M} d\tau \text{ for all } h \in L.
\]

Now given \( t > 0 \), choose \( \rho \in C^\infty(\mathbb{R}; \mathbb{R}) \) so that \( \rho = 0 \) off of \( (0, 1) \) and \( \int \rho(\tau) d\tau = 1 \). For \( 0 < \epsilon < t \), set \( \rho_\epsilon(\tau) = \epsilon^{-1} \rho(\epsilon^{-1} \tau) \) and

\[
\psi_\epsilon(\tau) = \int_0^\tau \int_0^{\tau_1} \rho_\epsilon(t - \tau_2) d\tau_2 d\tau_1 - \tau.
\]
Then \( \psi_\epsilon = 0 \) off \([0, t]\) and \( \tilde{\psi}_\epsilon = \rho_\epsilon \). Hence, \( \psi_\epsilon \in L \) and, for any \( \xi \in \mathbb{R}^M \),

\[
(\xi, f(t))_{\mathbb{R}^M} = \lim_{\epsilon \searrow 0} \int_0^\infty (f(\tau), \rho_\epsilon(t - \tau)\xi)_{\mathbb{R}^M} \, d\tau = -\lim_{\epsilon \searrow 0} \Lambda(\psi, \xi)
\]

\[
= -\lim_{\epsilon \searrow 0} (\psi_\epsilon, h_0)_{H^1(\mathbb{R}^M)} = \lim_{\epsilon \searrow 0} \int_0^\infty (h_0(\tau), \rho_\epsilon(t - \tau)\xi)_{\mathbb{R}^M} \, d\tau = (\xi, h_0(t))_{\mathbb{R}^M},
\]

which leads to the contradiction \( f = h_0 \in H^1(\mathbb{R}^M) \). With this information, we can complete the proof as follows. Define \( F : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}^+} \) so that

\[
F(w)_m = \int_0^\infty (\dot{h}_m(\tau), dw(\tau))_{\mathbb{R}^M} = -\int_0^\infty (w(\tau), \dot{h}_m(\tau))_{\mathbb{R}^M} \, d\tau.
\]

Then \( F_*\mathcal{W} = \gamma_{0,1}^{\mathbb{Z}^+} \), and, because \( F(w + f)_m = F(w)_m + \Lambda(h_m) \),

\[
F_*((T_f)_*\mathcal{W}) = \prod_{m=1}^\infty \gamma_{\Lambda(h_m),1}.
\]

Since \( \gamma_{0,1}^{\mathbb{Z}^+} \perp \prod_{m=1}^\infty \gamma_{a_m,1} \) if \( \sum_{m=1}^\infty a_m^2 = \infty \), it follows that \( F_*((T_f)_*\mathcal{W}) \perp F_*\mathcal{W} \), which means that \((T_f)_*\mathcal{W} \perp \mathcal{W} \).

Among other things, this result allows us to show that \( \mathcal{W} \) gives positive measure to every open subset of \( \Omega \). To see this, first observe that, because \( H^1(\mathbb{R}^M) \) is dense in \( \Omega \), it suffices to show that \( \mathcal{W}(B_\Omega(h, r)) > 0 \) for every \( h \in H^1(\mathbb{R}^M) \).

Second, note that

\[
\mathcal{W}(B_\Omega(0, r)) = \mathbb{E}^\mathcal{W}[\mathcal{R}^{-\frac{1}{2}}_{-h} \mathcal{R}^\frac{1}{2}_{-h}, B_\Omega(0, r)]
\]

\[
\leq \mathbb{E}^\mathcal{W}[\mathcal{R}^{-\frac{1}{2}}_{-h}(T_{-h})_*\mathcal{W}(B_\Omega(0, r))]^{\frac{1}{2}} = e^{\frac{\|h\|^2_{H^1(\mathbb{R}^M)}}{2}} \mathcal{W}(B_\Omega(h, r)),
\]

and therefore it suffices to show that \( \mathcal{W}(B_\Omega(0, r)) > 0 \) for all \( r > 0 \). To this end, let \((B(t), \mathcal{F}_t, \mathbb{P})\) be an \( \mathbb{R} \)-valued Brownian motion, and consider the function

\[
u(t, x) = e^{-\frac{\pi^2}{2r^2}x\sin \frac{\pi}{2}(x + r)}.
\]

Because \( \partial_t u + \frac{1}{2} \partial_x^2 u = 0 \), \((u(t, B(t), \mathcal{F}_t, \mathbb{P}))\) is a martingale. Hence, if \( \zeta_r = \inf\{t \geq 0 : |B(t)| \geq r\} \), then, because \( u(t, \pm r) = 0 \),

\[
1 = \mathbb{E}^\mathbb{P}[u(t \wedge \zeta_r, B(t \wedge \zeta_r))] = e^{\frac{\pi^2}{2r^2} \mathbb{E}^\mathbb{P}\left[\sin \frac{\pi(B(t) + r)}{2} \zeta_r > t\right]},
\]

and so

\[
\mathbb{P}(\zeta_r > t) \geq e^{-\frac{\pi^2}{2r^2}t}.
\]
Because
\[ W(\|w(\cdot)\|_{[0,T]} < r) \geq W\left( \max_{1 \leq j \leq N} \|w(\cdot)_j\|_{[0,T]} < \frac{r}{N^2} \right) = P(\zeta_N - \frac{1}{2} < T)^N, \]
we now know that
\[ W(\|w(\cdot)\|_{[0,T]} < r) \geq e^{-\frac{x^2 r^2}{8}}, \]
and therefore
\[ W(B_{\Omega}(0,r)) \geq W\left( \|w(\cdot)\|_{[0,T]} < \frac{r}{2} \& \sup_{t \geq T} |w(t) - w(T)| \leq \frac{r}{2} \right) \geq e^{-\frac{x^2 r^2}{8}} W\left( \sup_{t \geq 0} \frac{|w(t)|}{1 + t} \leq \frac{r}{2} \right). \]

Finally, because \|w\|_{\Omega} < \infty and \frac{|w(t)|}{1 + t} \to 0 \text{ (a.s.,} W), \text{ we can choose } T \to 0 \text{ so that}
\[ W\left( \sup_{t \in [0,\sqrt{T}]} \frac{|w(t)|}{1 + t + T} \geq \frac{r}{2} \right) \vee W\left( \sup_{t \geq \sqrt{T}} \frac{|w(t)|}{1 + t} \geq \frac{r}{2} \right) \leq \frac{1}{4}, \]
in which case
\[ W\left( \sup_{t \geq \sqrt{T}} \frac{|w(t)|}{1 + t} \geq \frac{r}{2} \right) \leq W\left( \sup_{t \in [0,\sqrt{T}]} \frac{|w(t)|}{1 + t + T} \geq \frac{r}{2} \right) + W\left( \sup_{t \geq \sqrt{T}} \frac{|w(t)|}{1 + t} \geq \frac{r}{2} \right) \leq \frac{1}{2}. \]

Besides what it says about Wiener measure, the preceding result has the following interesting application to partial differential equations. Let \( G \supset 0 \) be a connected, open subset of \( \mathbb{R}^M \) and \( u \) a harmonic function on \( G \) that achieve its minimum value at 0. Then, for bounded, open \( H \) with \( \bar{H} \subseteq G \), \( W(\zeta^H < \infty) = 1 \) and \( (u(w(t \wedge \zeta^H)), W_t, \Omega) \) martingale when \( \zeta^H = \inf\{t \geq 0 : w(t) \notin H\} \). Now suppose that there were a point \( x \in G \setminus \{0\} \) at which \( u(x) > u(0) \), and let \( p : [0,1] \to G \) be a continuous path such that \( p(0) = 0 \) and \( p(1) = x \). Choose \( r > 0 \) so that \( u(y) - u(0) \geq \delta = \frac{u(x) - u(0)}{2} \) for \( y \in B(x,r) \) and \( |p(t) - y| > 2r \) for all \( t \in [0,1] \) and \( y \notin G \). If
\[ \zeta(w) = \inf\{t \geq 0 : |w(t) - p(t)| \geq 2r\} \text{ and } \zeta'(w) = \inf\{t \geq 0 : w(t) \in B(x,r)\}, \]
then, \( \|w(\cdot) - p(\cdot)\|_{[0,1]} < r \implies \zeta'(w) < \zeta(w) \), and so
\[ W(\zeta' < \zeta) \geq \epsilon_r \equiv W(\|w(\cdot) - p(\cdot)\|_{[0,1]} < r) > 0. \]
But this means that
\[u(0) = \mathbb{E}^{W}[u(w(\zeta \land \zeta'))] = \mathbb{E}^{W}[u(w(\zeta'))], \zeta' < \zeta] + \mathbb{E}^{W}[u(w(\zeta)), \zeta > \zeta'] \]
\[\geq u(0) + \delta \epsilon_r,
\]
which is impossible. Therefore no such \(x\) exists, and so we have proved the\textbf{ strong minimum principle}, which is the statement that a harmonic function on a connected open set can achieve a minimum value there only if it is constant. Essentially the same argument proves the parabolic analog of this principle. Namely, suppose that \(G\) is an open subset of \(\mathbb{R} \times \mathbb{R}^M\) which is \textit{upward path connected} in the sense that for each \((s, x), (t, y) \in G\) with \(t > s\) there is a continuous path \(p : [s, t] \rightarrow \mathbb{R}^M\) such that \(p(s) = x, p(t) = y\), and \((\tau, p(\tau)) \in G\) for all \(\tau \in [s, t]\). If \(u \in C^{1,2}(G)\) satisfying \(\partial_t u + \frac{1}{2} \Delta u = 0\) and \((s, x) \in G\), \(u(s, x) \leq u(t, y)\) for all \((t, y) \in G\) with \(t > s\) implies that \(u(t, y) = u(s, x)\) for all such \((t, y)\).

A second application is to the development of a calculus for functions on \(\Omega\) in which Wiener measure plays the role that Lebesgue measure plays in finite dimensions. First observe that
\[\mathbb{E}^{W}[R^p_h] = e^{\frac{\|h\|^2_{H^1(\mathbb{R}^M)}}{2}} \text{ for all } p \in [1, \infty).\]

Hence, by Hölder’s inequality,
\[\|\Phi \circ T_h\|_{L^q(W; \mathbb{R})} \leq e^{\frac{\|h\|^2_{H^1(\mathbb{R}^M)}}{2(p-q)}} \|\Phi\|_{L^p(W; \mathbb{R})}\]
for \(1 < q < p < \infty\) and \(\Phi \in L^p(W; \mathbb{R})\).

Now suppose that \(\Phi \in L^p(W; \mathbb{R})\) for some \(p \in (1, \infty)\) and that there exists a function \(D_h \Phi \in L^p(W; \mathbb{R})\) such that
\[\lim_{\xi \to 0} \frac{\Phi \circ T_{\xi^h} - \Phi}{\xi} \rightarrow D_h \Phi \text{ in } L^1(W; \mathbb{R}).\]

Then, because
\[\mathbb{E}^{W}[\Phi \circ T_{\xi^h} - \Phi] = \mathbb{E}^{W}[(R_{\xi^h} - 1)\Phi]\]
and \(R_{\xi^h}^{-1} \rightarrow I(\dot{h})\) in \(L^p(W; \mathbb{R})\), it follows that
\[\mathbb{E}^{W}[D_h \Phi] = \mathbb{E}^{W}[I(\dot{h})\Phi].\]

Next suppose that \(\Phi_1, \Phi_2 \in L^p(W; \mathbb{R})\) for some \(p \in (2, \infty)\) and that
\[\lim_{\xi \to 0} \frac{\Phi_i \circ T_{\xi^h} - \Phi_i}{\xi} \rightarrow D_h \Phi_i \text{ in } L^p(W; \mathbb{R}) \text{ for } i \in \{1, 2\}.\]

\[\text{Here and elsewhere, } p' = \frac{p}{p-1} \text{ denotes the Hölder conjugate of } p.\]
Then another application of Hölder’s inequality shows that
\[
(Φ_1 ◦ T_ξ h)(Φ_2 ◦ T_ξ h) - Φ_1 Φ_2 \rightarrow Φ_1 D_h Φ_2 + Φ_2 D_h Φ_1 \text{ in } L^1(\mathcal{W}; \mathbb{R}),
\]
and therefore
\[
(89) \quad E^{\mathcal{W}}[Φ_1 D_h Φ_2] = -E^{\mathcal{W}}[Φ_2 D_h Φ_1] + E^{\mathcal{W}}[I(\dot{h})Φ_1 Φ_2].
\]

This formula is the starting point for the Sobolev type calculus on which P. Malliavun based his analysis of functions on Wiener space.

**Wiener’s Spaces of Homogenous Chaos**: Wiener spent a lot of time thinking about noise and how to separate it from signals. One his most profound ideas on the subject was that of decomposing a random variable into components of uniform orders of randomness, or, what, with his usual flare for language, he called components of homogenous chaos. From a mathematical standpoint, what he was doing is write $L^2(\mathcal{W}; \mathbb{R})$ as the direct sum of mutually orthogonal subspaces consisting functions that could be reasonably thought of as having a uniform order of randomness. Wiener’s own treatment of this subject is fraught with difficulties, all of which were resolved by Itô. Thus, we, once again, will be guided by Itô.

We must define what we will mean by *multiple stochastic integrals*. That is, if for $m ≥ 1$ and $t ∈ [0, ∞)$, $□^{(m)}(t) = [0, t]^m$ and $□^{(m)}(∞) = □^{(m)}(∞)$, we want to assign a meaning to expressions like
\[
\tilde{I}^{(m)}_F(t) = \int_{□^{(m)}(t)} (F(\vec{\tau}), d\vec{\tau})_{(\mathbb{R}^M)^m}
\]
when $F ∈ L^2(□^{(m)}; (\mathbb{R}^M)^m)$.

With this goal in mind, when $m = 1$ and $F = f ∈ L^2([0, ∞); \mathbb{R}^M)$, take $I^{(1)}_f(t) = I_f(t)$, where $I_f(t)$ is the Paley–Wiener integral of $f$. When $m ≥ 2$ and $F = f_1 ⊗ ⋯ ⊗ f_m$ for some $f_1, \ldots, f_m ∈ L^2([0, ∞); \mathbb{R}^M)$, we use induction to define $I^{(m)}_F(t)$ so that
\[
(90) \quad I^{(m)}_F(t) = \int_0^t I^{(m-1)}_{f_1 ⊗ ⋯ ⊗ f_{m-1}}(\tau)(f_m(\tau), dw(\tau))_{\mathbb{R}^M},
\]

9 One reason why Wiener is difficult to read is that he insisted on doing all integration theory with respect to Lebesgue measure on the interval $[0, 1]$. He seems to have thought that this decision would make engineers and other non-mathematicians happier.

10 Here we are identifying $f_1 ⊗ ⋯ ⊗ f_m$ with the $(\mathbb{R}^M)^m$-valued function $F$ on $[0, ∞)^m$ such that $(\Xi, F(t_1, \ldots, t_m))_{(\mathbb{R}^M)^m} = (\xi_1, f_1(t_1))_{\mathbb{R}^M} ⋯ (\xi_m, f_m(t_m))_{\mathbb{R}^M}$ for $\Xi = (\xi_1, \ldots, \xi_m) ∈ (\mathbb{R}^M)^m$. 

where now we need Itô’s integral. Of course, in order to do so, we are obliged to check that \( \tau \leadsto f_m(\tau)I_{f_1 \otimes \cdots \otimes f_{m-1}}(\tau) \) is square integrable. But, assuming that \( I_{f_1 \otimes \cdots \otimes f_{m-1}} \) is well defined, we have that

\[
\mathbb{E}^W \left[ \int_0^T |f_m(\tau)I_{f_1 \otimes \cdots \otimes f_{m-1}}(\tau)|^2 d\tau \right] = \int_0^T |f_m(\tau)|^2 \mathbb{E}^W \left[ |I_{f_1 \otimes \cdots \otimes f_{m-1}}(\tau)|^2 \right] d\tau.
\]

Hence, at each step in our induction procedure, we can check that

\[
\mathbb{E}^W \left[ |I_{f_1 \otimes \cdots \otimes f_m}(T)|^2 \right] = \int_{\Delta^{(m)}(T)} |f_1(\tau_1)|^2 \cdots |f_m(\tau_m)|^2 d\tau_1 \cdots d\tau_m,
\]

where \( \Delta^{(m)}(t) \equiv \{(t_1, \ldots, t_m) \in \Box^{(m)} : t_1 < \cdots < t_m < t\} \); and so, after polarization, we arrive at

\[
\mathbb{E}^W \left[ I_{f_1 \otimes \cdots \otimes f_m}(T)I_{f_1' \otimes \cdots \otimes f_m'}(T) \right] = \int_{\Delta^{(m)}(T)} (f_1(\tau_1), f_1'(\tau_1))^M \cdots (f_m(\tau_m), f_m'(\tau_m))^M d\tau_1 \cdots d\tau_m.
\]

I next introduce

\[(91) \quad \tilde{I}_{f_1 \otimes \cdots \otimes f_m}(t) = \sum_{\pi \in \Pi_m} I_{f_{\pi(1)} \otimes \cdots \otimes f_{\pi(m)}}(t),\]

where \( \Pi_m \) is the symmetric group (i.e., the group of permutations) on \( \{1, \ldots, m\} \).

By the preceding, one sees that

\[
\mathbb{E}^W \left[ \tilde{I}_{f_1 \otimes \cdots \otimes f_m}(T)\tilde{I}_{f_1' \otimes \cdots \otimes f_m'}(T) \right] = \sum_{\pi, \pi' \in \Pi_m} \int_{\Delta^{(m)}(T)} \prod_{\ell=1}^m (f_{\pi(\ell)}(\tau_\ell), f_{\pi'(\ell)}(\tau_\ell))^M d\tau_1 \cdots d\tau_m
\]

\[
= \sum_{\pi \in \Pi_m} \int_{\Box^{(m)}(T)} \prod_{\ell=1}^m (f_\pi(\tau_\ell), f_{\pi'(\ell)}(\tau_\ell))^M d\tau_1 \cdots d\tau_m
\]

\[
= \sum_{\pi \in \Pi_m} \prod_{\ell=1}^m (f_\pi, f_{\pi'})_{L^2([0,T];\mathbb{R}^M)}.
\]

In preparation for the next step, let \( \{g_j : j \geq 1\} \) be an orthonormal basis in \( L^2([0,\infty);\mathbb{R}^M) \), and note that \( \{g_{j_1} \otimes \cdots \otimes g_{j_m} : (j_1, \ldots, j_m) \in \mathbb{Z}^+^m\} \) is an
orthonormal basis in $L^2(\Box^m; (\mathbb{R}^M)^m)$. Next, let $\mathcal{A}$ denote the set of $\alpha \in \mathbb{N}^{\mathbb{Z}^+}$ for which $\|\alpha\| = \sum_1^\infty \alpha_j < \infty$. Finally, given $\alpha \in \mathcal{A}$, set

$$S(\alpha) = \{ j \in \mathbb{Z}^+ : \alpha_j \geq 1 \}$$

and

$$G_\alpha = \bigotimes_{j \in S(\alpha)} g^{\alpha_j}.$$

Then, $G_\alpha \in L^2(\Box^m; (\mathbb{R}^M)^m)$, and, for $\alpha, \beta \in \mathcal{A}$ with $\|\alpha\| = \|\beta\| = m$,

$$\left( \tilde{I}_{G_\alpha}, \tilde{I}_{G_\beta} \right)_{L^2(\mathcal{W}; \mathbb{R})} = \delta_{\alpha,\beta} \alpha!,$$

where $\alpha! = \prod_{j \in S(\alpha)} \alpha_j!$.

Now, given $F \in L^2(\Box^m; (\mathbb{R}^M)^m)$, set

$$\tilde{F}(t_1, \ldots, t_m) = \sum_{\pi \in \Pi_m} F(t_{\pi(1)}, \ldots, t_{\pi(m)}),$$

and observe that

$$\|\tilde{F}\|_{L^2(\Box^m; (\mathbb{R}^M)^m)}^2 = \sum_{\|\alpha\| = m} \left( \tilde{F}, g_{j_1} \otimes \cdots \otimes g_{j_m} \right)_{L^2(\Box^m; (\mathbb{R}^M)^m)}^2$$

$$= \sum_{\|\alpha\| = m} \binom{m}{\alpha} \left( \tilde{F}, G_\alpha \right)_{L^2(\Box^m; (\mathbb{R}^M)^m)}^2,$$

where $\binom{m}{\alpha}$ is the multinomial coefficient $\frac{m!}{\alpha! \alpha'}$. Hence, after combining this with calculation in (92), we have that

$$\mathbb{E}^\mathcal{W} \left[ \sum_{\|\alpha\| = m} \frac{\left( \tilde{F}, G_\alpha \right)_{L^2(\Box^m; (\mathbb{R}^M)^m)}}{\alpha!} \tilde{I}_{G_\alpha}(\infty) \right]^2 = \frac{1}{m!} \|\tilde{F}\|_{L^2(\Box^m; (\mathbb{R}^M)^m)}^2.$$

With these considerations, we have proved the following.

**Theorem 94.** There is a unique linear map

$$F \in L^2(\Box^m; (\mathbb{R}^M)^m) \longmapsto \tilde{I}_F^{(m)}(\infty) \in \mathcal{M}^2(\mathcal{W}; \mathbb{R})$$

such that $\tilde{I}_{f_1 \otimes \cdots \otimes f_m}$ is given as in (91) and

$$\mathbb{E}^\mathcal{W} \left[ \tilde{I}_F^{(m)}(\infty) \tilde{I}_{F'}^{(m)}(\infty) \right] = \frac{1}{m!} \left( \tilde{F}, \tilde{F}' \right)_{L^2(\Box^m; (\mathbb{R}^M)^m)}.$$

In fact,

$$\tilde{I}_F^{(m)}(\infty) = \sum_{\|\alpha\| = m} \frac{\left( \tilde{F}, G_\alpha \right)_{L^2(\Box^m; (\mathbb{R}^M)^m)}}{\alpha!} \tilde{I}_{G_\alpha},$$

where the convergence is in $L^2(\mathcal{W}; \mathbb{R})$. 
Although it is somewhat questionable to do so, as indicated at the beginning of this section, I like to think of \( \tilde{I}^{(m)}_F(t) \) as

\[
\int_{\square^{(m)}(t)} (F(\tau), d\tilde{w}(\tau))_{(\mathbb{R}^M)^m}.
\]

The reason why this notation is questionable is that, although it is suggestive, it may suggest the wrong thing. Specifically, in order to avoid stochastic integrals with non-progressively measurable integrands, our definition of \( \tilde{I}^{(m)}_F \) carefully avoids integration across diagonals, whereas the preceding notation gives no hint of that fact.

Take \( Z^{(0)} \) to be the subspace of \( L^2(W; \mathbb{R}) \) consisting of the constant functions, and, for \( m \geq 1 \), set

\[
Z^{(m)} = \{ \tilde{I}^{(m)}_F(\infty) : F \in L^2(\square^{(m)}; (\mathbb{R}^M)^m) \}.
\]

Clearly, each \( Z^{(m)} \) is a subspace of \( L^2(W; \mathbb{R}) \). Furthermore, if \( \{ F_k : k \geq 1 \} \subseteq L^2(\square^{(m)}; (\mathbb{R}^M)^m) \) and \( \{ \tilde{I}^{(m)}_F(\infty) : k \geq 1 \} \) converges in \( L^2(W; \mathbb{R}) \), then \( \{ \tilde{F}_k : k \geq 1 \} \) converges in \( L^2(\square^{(m)}; (\mathbb{R}^M)^m) \) to some symmetric function \( G \). Hence, since \( G = m!G \), we see that \( \tilde{I}^{(m)}_F(\infty) \rightarrow \tilde{I}^{(m)}_F(\infty) \) in \( L^2(W; \mathbb{R}) \) where \( F = \frac{1}{m!}G \).

That is, each \( Z^{(m)} \) is a closed linear subspace of \( L^2(W; \mathbb{R}) \). Finally, \( Z^{(m)} \perp Z^{(m')} \) when \( m' \neq m \). This is completely obvious if either \( m \) or \( m' \) is 0. Thus, suppose that \( 1 \leq m < m' \). Then

\[
\mathbb{E}^W\left[ I^{(m)}_{f_1 \otimes \cdots \otimes f_m}(\infty) I^{(m')}_{f'_1 \otimes \cdots \otimes f'_m}(\infty) \right]
= \int_{\Delta^{(m'-m)}} \prod_{\ell=0}^{m-1} (f_{m-\ell}(\tau_\ell), f'_{m'-\ell}(\tau_\ell))_{\mathbb{R}^M}
\times \mathbb{E}^W\left[ I^{(m'-m)}_{f'_1 \otimes \cdots \otimes f'_{m'-m}}(\tau_{m'-m}) \right] d\tau_1 \cdots d\tau_{m'-m} = 0,
\]

which completes the proof.

The space \( Z^{(m)} \) is the space of \textit{mth order homogeneous chaos}. The reason why elements of \( Z^{(0)} \) are said to be of 0th order chaos is clear: constants are non-random. To understand why \( \tilde{I}^{(m)}_F(\infty) \) is of \textit{m}th order chaos when \( m \geq 1 \), it is helpful to replace \( d\tilde{w}(\tau) \) by the much more ambiguous \( \tilde{w}(\tau) d\tau \) and write

\[
\tilde{I}^{(m)}_F(\infty) = \int_{\square^{(m)}} (F(\tau_1, \ldots, \tau_m), (\tilde{w}(\tau_1), \cdots, \tilde{w}(\tau_m)))_{(\mathbb{R}^M)^m} d\tau_1 \cdots d\tau_m.
\]

In the world of engineering and physics, \( \tau \sim \tilde{w}(\tau) \) is \textit{white noise}.\(^{11}\) Thus, \( Z^{(m)} \) is the space built out of homogeneous \textit{m}th order polynomials in white noises

\(^{11}\)The terminology comes from the observation that, no matter how one interprets \( t \sim \tilde{w}(t) \), it is a stationary, centered Gaussian process whose covariance is the Dirac delta function times the identity. In particular, \( \tilde{w}(t_1) \) is independent of \( \tilde{w}(t_2) \) when \( t_1 \neq t_2 \).
evaluated at different times. In other words, the order of chaos is the order of the white noise polynomial.

The result of Wiener, alluded to at the beginning of this section, now becomes the assertion that

\[(95) \quad L^2(\mathcal{W}; \mathbb{R}) = \bigoplus_{m=0}^{\infty} Z^{(m)}.\]

The key to Itô’s proof of (95) is found in the following.

**Lemma 96.** If \( f \in L^2([0, \infty); \mathbb{R}^M) \), then and \( I^{(0)}_f \equiv 1 \), for each \( \lambda \in \mathbb{C} \),

\[e^{\lambda I_f(\infty)} - \frac{\lambda^2}{2} \| f \|_{L^2([0, \infty); \mathbb{R}^M)}^2 = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \bar{I}^{(m)}_f(\infty) \equiv e^\lambda \lim_{M \to \infty} \sum_{m=0}^{M} \frac{\lambda^m}{m!} \bar{I}^{(m)}_f(\infty) \]

\( W \)-almost surely and in \( L^2(\mathcal{W}; \mathbb{R}) \). In fact, if

\[R^M_f(\infty, \lambda) \equiv e^{\lambda I_f(\infty)} - \frac{\lambda^2}{2} \| f \|_{L^2([0, \infty); \mathbb{R}^M)}^2 - \sum_{m=0}^{M-1} \frac{\lambda^m}{m!} \bar{I}^{(m)}_f(\infty),\]

then

\[\mathbb{E}^W[ |R^M_f(\infty, \lambda)|^2 ] = e^{\lambda^2 \| f \|_{L^2([0, \infty); \mathbb{R}^M)}^2} \sum_{m=0}^{M-1} \frac{\lambda^m}{m!} \bar{I}^{(m)}_f(\infty) \leq \frac{\left( \lambda^2 \| f \|_{L^2([0, \infty); \mathbb{R}^M)}^2 \right)^2}{M!}.

**Proof:** Set \( E(t, \lambda) = e^{\lambda I_f(t)} - \frac{\lambda^2}{2} \| f \|_{L^2([0, t]; \mathbb{R}^M)}^2 \). Then

\[E(t, \lambda) = 1 + \lambda \int_0^t E(\tau, \lambda) (f(\tau), dw(\tau))_{\mathbb{R}^M} \]

Thus, if \( (cf. \ (90)) \ I^{(m)}(t) \equiv \bar{I}^{(m)}_f(t) \equiv \frac{1}{m!} \bar{I}^{(m)}_f \) and

\[R^0(t, \lambda) \equiv E(t, \lambda) \quad \text{and} \quad R^{M+1}(t, \lambda) \equiv \lambda \int_0^t R^M(\tau, \lambda) (f(\tau), dw(\tau))_{\mathbb{R}^M}\]

for \( M \geq 0 \), then, by induction, one sees that

\[E(t, \lambda) = 1 + \sum_{m=1}^{M} \lambda^m I^{(m)}(t) + R^{M+1}(t, \lambda)\]

for all \( M \geq 0 \). Finally, if \( A(t) = \int_0^t |f(\tau)|^2 \, d\tau \), then

\[\mathbb{E}^W[ |R^0(t, \lambda)|^2 ] \leq e^{\lambda^2 A(t)} \mathbb{E}^W[ E(t, 2\mathcal{R} \lambda) ] = e^{\lambda^2 A(t)}, \]

and

\[\mathbb{E}^W[ |R^{M+1}(t, \lambda)|^2 ] = \lambda^2 \int_0^t \mathbb{E}^W[ |R^M(\tau, \lambda)|^2 ] A(\tau) \, d\tau.\]

Hence, the asserted estimate follows by induction on \( M \geq 0 \). \( \square \)

\(^{12}\) Remember that our integrals stay away from the diagonal.
Theorem 97. The span of
\[ \mathbb{R} \oplus \left\{ I_{f_{\infty}}^{(m)}(\infty) : m \geq 1 \& f \in L^2([0, \infty); \mathbb{R}^M) \right\} \]
is dense in \( L^2(\mathcal{W}; \mathbb{R}) \). In particular, (95) holds.

Proof: Let \( \mathbb{H} \) denote smallest closed subspace of \( L^2(\mathcal{W}; \mathbb{R}) \) containing all constants and all the functions \( I_{f_{\infty}}^{(m)}(\infty) \). By the preceding, we know that \( \cos \circ I_{f}(\infty) \) and \( \sin \circ I_{f}(\infty) \) are in \( \mathbb{H} \) for all \( f \in L^2([0, \infty); \mathbb{R}^M) \).

Next, observe that the space of functions \( \Phi : C([0, \infty); \mathbb{R}^M) \rightarrow \mathbb{R} \) which have the form
\[
\Phi = F(I_{f_1}(\infty), \ldots, I_{f_L}(\infty))
\]
for some \( L \geq 1, F \in \mathcal{C}(\mathbb{R}^L; \mathbb{R}) \), and \( f_1, \ldots, f_L \in L^2([0, \infty); \mathbb{R}^M) \) is dense in \( L^2(\mathcal{W}; \mathbb{R}) \). Indeed, this follows immediately from the density in \( L^2(\mathcal{W}; \mathbb{R}) \) of the space of functions of the form
\[
w \mapsto F(w(t_1), \ldots, w(t_L)),
\]
where \( L \geq 1, F \in \mathcal{C}(\mathbb{R}^L; \mathbb{R}) \), and \( 0 \leq t_0 < \cdots < t_L \).

Now suppose that \( F : \mathcal{C}(\mathbb{R}^L; \mathbb{R}) \) is given, and let \( \hat{F} \) denote its Fourier transform. Then, by elementary Fourier analysis,
\[
F_n(x) \equiv (2(4^n + 1)\pi)^{-L} \sum_{\|m\|_\infty \leq 4^n} e^{-\sqrt{1}/(2^{-n}m,x)_{\mathbb{R}^L}} \hat{F}(m 2^{-n}) \rightarrow F(x),
\]
both uniformly and boundedly, where \( m = (m_1, \ldots, m_L) \in \mathbb{Z}^L \) and \( \|m\|_\infty = \max_{1 \leq \ell \leq L} |m_\ell| \). Finally, since \( \hat{F}(\Xi) = \hat{F}(-\Xi) \) for all \( \Xi \in \mathbb{R}^L \), we can write
\[
(2(4^n + 1)\pi)^L F_n(I_{f_1}(\infty), \ldots, I_{f_L}(\infty)) = \hat{F}(0) + 2 \sum_{1 \leq \|m\|_\infty \leq 4^n} \left( \Re(\hat{F}(m 2^{-n})) \cos \circ I_{m,n}(\infty) + \Im(\hat{F}(m 2^{-n})) \sin \circ I_{m,n}(\infty) \right) \in \mathbb{H},
\]
where \( I_{m,n} \equiv 2^{-n}I_{f_m} \) with \( f_m = \sum_{\ell=1}^L m_\ell f_\ell \). \( \square \)

The following corollary is an observation made by Itô after he cleaned up Wiener’s treatment of (95). It is often called Itô’s representation theorem and turns out to play an important role in applications of stochastic analysis to, of all things, models of financial markets.\(^\text{13}\)

\(^\text{13}\)In fact, it shares with Itô’s formula responsibility for the widespread misconception in the financial community that Itô is an economist.
Corollary 98. The map

$$(x, \xi) \in \mathbb{R} \times \mathcal{P}^2(\mathbb{R}^M) \longrightarrow x + I_\xi(\infty) \in L^2(\mathcal{W}; \mathbb{R})$$

is a linear, isometric surjection. Hence, for each $\Phi \in L^2(\mathcal{W}; \mathbb{R})$ there is a $\mathcal{W}$-almost surely unique $\xi \in \mathcal{P}^2(\mathbb{R}^M)$ such that

$$\Phi = \mathbb{E}^\mathcal{W}[\Phi|W] + \int_0^\infty (\xi(\tau), dw(\tau))_{\mathbb{R}^M} \; \mathcal{W}\text{-almost surely}.$$ 

In particular,

$$\mathbb{E}^\mathcal{W}[\Phi|W] = \mathbb{E}^\mathcal{W}[\Phi] + \int_0^t (\xi(\tau), dw(\tau))_{\mathbb{R}^M} \; \mathcal{W}\text{-almost surely for each } t \geq 0.$$ 

Finally, for any $\Phi \in L^1(\mathcal{W}; \mathbb{R})$, there is a version of $(t, w) \sim \mathbb{E}^\mathcal{W}[\Phi|W](w)$ that is is continuous as function of $t \geq 0$.

Proof: Since it is clear that the map is linear and isometric, the first assertion will be proved once we check that the map is onto. But, because it is a linear isometry, we know that its image is a closed subspace, and so we need only show that its image contains a set whose span is dense. However, for each $f \in L^2([0, \infty); \mathbb{R}^M)$ and $m \geq 1$,

$$\tilde{I}_{f \otimes m}(\infty) = m \int_0^\infty (\tilde{I}_{f \otimes (m-1)}(\tau)f(\tau), dw(\tau))_{\mathbb{R}^M},$$

and so, by the first part of Theorem 97, we are done.

Given the first assertion, the second assertion is obvious and shows that $(t, w) \sim \mathbb{E}^\mathcal{W}[\Phi|W](w)$ can be chosen so that it is continuous with respect to $t$ for any $\Phi \in L^2(\mathcal{W}; \mathbb{R})$. Finally, if $\Phi \in L^1(\mathcal{W}; \mathbb{R})$, choose $\{\Phi_k : k \geq 1\} \subseteq L^2(\mathcal{W}; \mathbb{R})$ so that $\Phi_k \longrightarrow \Phi$ in $L^1(\mathcal{W}; \mathbb{R})$, and let $(t, w) \sim X_k(t, w)$ be a version of $(t, w) \sim \mathbb{E}^\mathcal{W}[\Phi|W]$ which is continuous in $t$. Then, by Doob’s inequality, for all $t > 0$,

$$\sup_{k \geq k} \mathcal{W}(\|X_{t}(\cdot) - X_k(\cdot)\|_{[0, t]} \geq \epsilon) \leq \frac{\|X_{t}(t) - X_k(t)\|_{L^1(\mathcal{W}; \mathbb{R})}}{\epsilon} \longrightarrow 0$$

as $k \to \infty$. Hence there exists a progressively measurable $X : [0, \infty) \times \Omega \longrightarrow \mathbb{R}$ to which $\{X_k(\cdot) : k \geq 1\} \mathcal{W}$-almost surely converges uniformly on compacts, and, since, for each $t \geq 0$,

$$\lim_{k \to \infty} \mathcal{W}(X_k(t) - \mathbb{E}^\mathcal{W}[\Phi|W] = 0,$$

it follows that $X(t)$ is a version of $\mathbb{E}^\mathcal{W}[\Phi|W]$. □
Extensions of and Variations on Itô’s Ideas

As is clear from Doob’s presentation of Itô’s theory in his book, Doob understood that one can apply Itô’s ideas to any continuous, square integrable martingale \((M(t), \mathcal{F}_t, \mathbb{P})\) for which one knows that there is a progressively measurable function \(t \mapsto A(t)\) which is continuous and non-decreasing in \(t\) and for which \((M(t)^2 - A(t), \mathcal{F}_t, \mathbb{P})\) is a martingale. At the time, Doob did not know what is now called the Doob-Meyer decomposition theorem, a special case of which guarantees that such an \(A(\cdot)\) always exists. In this chapter, I will first prove this existence result and then, following Kunita and Watanabe, develop an elegant version of the theory that Doob had in mind.

The Doob-Meyer Decomposition Theorem: Doob noticed that if \((X_n, \mathcal{F}_n, \mathbb{P})\) is a discrete parameter, integrable submartingale, then there is a unique \(\{A_n : n \geq 0\}\) such that \(A_0 = 0\), \(A_n\) is \(\mathcal{F}_{n-1}\)-measurable and \(A_{n-1} \leq A_n\) for \(n \geq 1\), and \((X_n - A_n, \mathcal{F}_n, \mathbb{P})\) is a martingale. To see that such \(A_n\’s\) exist, simply set \(A_0 = 0\) and \(A_n = A_{n-1} + \mathbb{E}^\mathbb{P}[X_n - X_{n-1} | \mathcal{F}_{n-1}]\) for \(n \geq 1\), and check that \((X_n - A_n, \mathcal{F}_n, \mathbb{P})\) is a martingale. To prove uniqueness, suppose that \(\{B_n : n \geq 0\}\) is a sequence of random variables such that \(B_0 = 0\), \(B_n\) is \(\mathcal{F}_{n-1}\)-measurable for \(n \geq 1\), and \((X_n - B_n, \mathcal{F}_n, \mathbb{P})\) is a martingale. Then \(B_0 = A_0\) and, since \(B_n - B_{n-1}\) is \(\mathcal{F}_{n-1}\)-measurable,

\[
B_n - B_{n-1} = \mathbb{E}^\mathbb{P}[X_n - X_{n-1} | \mathcal{F}_{n-1}] = A_n - A_{n-1}
\]

for \(n \geq 1\).

Trivial as this observation is, it greatly simplifies proofs of results like his stopping time and convergence theorems. However, even formulating, much less proving, a continuous parameter analog was a non-trivial challenge. Indeed, wholly aside from the a proof of existence, in a continuous parameter context it was not obvious what should replace the condition that \(A_n\) be \(\mathcal{F}_{n-1}\)-measurable. The person who figured out how to carry out this program was P.A. Meyer, who, in the process, launched a program that led to a deep théorie générale of stochastic processes. I know no elementary proof of Meyer’s theorem, and so it is fortunate that we need only the particularly easy, special case covered in the following theorem.\(^{14}\)

**Theorem 99.** If \((M(t), \mathcal{F}_t, \mathbb{P})\) is a continuous, local martingale, then there is an \(\mathbb{P}\)-almost surely unique continuous, non-decreasing, progressively measurable function \(t \mapsto A(t)\) such that \(A(0) = 0\) and \((M(t)^2 - A(t), \mathcal{F}_t, \mathbb{P})\) is a local martingale.

**Proof:**\(^{15}\) Without loss in generality, assume that \(M(0) = 0\). Next, note that if \(M\) is uniformly bounded and \(A(\cdot)\) exists, then \((M(t)^2 - A(t), \mathcal{F}_t, \mathbb{P})\) is a martingale and \(\mathbb{E}^\mathbb{P}[A(t)] = \mathbb{E}^\mathbb{P}[M(t)^2]\). In fact, if \(\{\zeta_m : m \geq 1\}\) are stopping times for \(\Omega\) is continuous if it is continuous as a function of time.

\(^{14}\)In the following and elsewhere, I will say that a progressively measurable function on \([0, \infty) \times \Omega\) is continuous if it is continuous as a function of time.

\(^{15}\)I learned the basic idea for this proof in a conversation with Itô.
the local martingale \((M(t)^2 - A(t), \mathcal{F}_t, \mathbb{P})\), then \(\mathbb{E}^\mathbb{P}[A(t \wedge \zeta_m)] = \mathbb{E}^\mathbb{P}[M(t \wedge \zeta_m)^2]\), \(A(t \wedge \zeta_m) \not\rightarrow A(t)\), and, as a consequence, \((M(t \wedge \zeta_m)^2 - A(t \wedge \zeta_m)) \rightarrow M(t)^2 - A(t)\) in \(L^1(\mathbb{P}; \mathbb{R})\), and therefore \((M(t)^2 - A(t), \mathcal{F}_t, \mathbb{P})\) is a martingale.

We will now show that we can reduce to the case when \(M\) is bounded. To this end, assume that we know both the existence and uniqueness result in this case, and introduce the stopping times \(\zeta_m = \inf\{t \geq 0 : |M(t)| \geq m\}\). Then, for each \(m \geq 1\), there is a unique \(A_m(\cdot)\) for \(t \rightsquigarrow M(t \wedge \zeta_m)\). Further, by uniqueness, we can assume that, for all \(m \geq 1\), \(A_{m+1}(t) = A_m(t)\) when \(t \in [0, \zeta_m)\). Thus, if \(A(t) \equiv A_m(t)\) for \(t \in [0, \zeta_m)\), then \((M(t \wedge \zeta_m)^2 - A(t \wedge \zeta_m), \mathcal{F}_t, \mathbb{P})\) is a martingale for all \(m \geq 1\). In addition, if \(A'(\cdot)\) is a second continuous, non-decreasing, progressively measurable function for which \((M(t)^2 - A'(t), \mathcal{F}_t, \mathbb{P})\) is a local martingale, then

\[
(M(t \wedge \zeta_m)^2 - A'(t \wedge \zeta_m), \mathcal{F}_t, \mathbb{P})
\]

is a martingale for all \(m \geq 1\), and so, by uniqueness, \(A'(t) = A_m(t)\) for \(t \in [0, \zeta_m)\). Having made this observation, from now on we will assume that \(M(0) = 0\) and \(M(\cdot)\) is uniformly bounded.

To prove the uniqueness assertion, we begin by showing that if \((X(t), \mathcal{F}_t, \mathbb{P})\) is a square integrable, continuous martingale with \(X(0) = 0\) and \(t \rightsquigarrow X(t)\) has bounded variation, then \(X(t) = 0\) (a.s., \(\mathbb{P}\)) for all \(t \geq 0\). Thus, suppose that such an \(X(\cdot)\) is given, and let \(|X|(t)\) be the total variation of \(X(\cdot)\) on \([0, t]\). By Riemann-Stieltjes integration theory, \(X(t)^2 = 2\int_0^t X(\tau) dM(\tau)\). Next, for \(R > 0\), set \(\zeta_R = \inf\{t \geq 0 : |X(t)| \geq R\}\) and \(X_R(t) = X(t \wedge \zeta_R)\). By Lemma 37, \(\mathbb{E}^\mathbb{P}[X_R(t)^2] = \mathbb{E}^\mathbb{P}\left(\int_0^t X_R(\tau) dX_R(\tau)\right)\), and so \(\mathbb{E}^\mathbb{P}[X(t \wedge \zeta_R)^2] = 0\).

Since \(\zeta_R \not\rightarrow \infty\) as \(R \rightarrow \infty\), it follows that \(X(t) = 0\) (a.s., \(\mathbb{P}\)). Now suppose that \((M(t) - A(t), \mathcal{F}_t, \mathbb{P})\) and \((M(t) - B(t), \mathcal{F}_t, \mathbb{P})\) are martingales, and take \(\zeta_R = \inf\{t \geq 0 : A(t) \vee B(t) \geq R\}\). Then the preceding result applies to \((B(t \wedge \zeta_R) - A(t \wedge \zeta_R), \mathcal{F}_t, \mathbb{P})\) and says that \(A(t) = B(t)\) for \(t \in [0, \zeta_R)\).

The proof of existence is more involved. Set \(\zeta_{0,n} = 0\) for \(n \geq 0\). Next, assuming that \(\{\zeta_{m,n} : m \geq 0\}\) has been chosen so that \(\zeta_{m,n} \not\rightarrow \infty\) as \(m \rightarrow \infty\), proceed by induction on \(m \geq 1\), and define \(\zeta_{m+1,n+1}\) to be

\[
\zeta_{\ell,n} \wedge \inf\{t \geq \zeta_{m,n+1} : |M(t) - M(\zeta_{m,n})| \geq 2^{-n}\}
\]

where \(\ell\) is the element of \(\mathbb{Z}^+\) for which \(\zeta_{\ell-1,n} < \zeta_{m,n+1} \leq \zeta_{\ell,n}\).

Clearly, for each \(n \geq 0\), \(\{\zeta_{m,n} : n \geq 0\}\) is non-decreasing sequence of bounded stopping times that tend to \(\infty\). Further, these sequences are nested in the sense that \(\{\zeta_{m,n} : m \geq 0\} \subseteq \{\zeta_{m,n+1} : m \geq 0\}\). Now set

\[
M_{m,n} = M(\zeta_{m,n}) \text{ and } \Delta_{m,n}(t) = M(t \wedge \zeta_{m+1,n}) - M(t \wedge \zeta_{m,n}),
\]

and observe that \(M(t)^2 = 2Y_n(t) + A_n(t)\), where

\[
Y_n = \sum_{m=0}^{\infty} M_{m,n} \Delta_{m,n}(t) \text{ and } A_n(t) = \sum_{m=0}^{\infty} \Delta_{m,n}(t)^2.
\]
In addition, \((Y_n(t), \mathcal{F}_t, \mathbb{P})\) is a square integrable martingale, \(A_n(0) = 0, A_n(\cdot)\) is continuous, and \(A_n(t) + 4^{-n} \geq A_n(s)\) for \(0 \leq s < t\). Thus, if, for each \(T > 0\), we show that \(\{A_n(\cdot) \mid [0, T] : n \geq 0\}\) converges in \(L^1(\mathbb{P}; C([0, T]; \mathbb{R}))\) or, equivalently, that \(\{Y_n(\cdot) \mid [0, T] : n \geq 0\}\) does, then we will be done. For that purpose, define
\[
\tilde{M}_{m,n+1} = M_{\ell,n} \text{ if } \zeta_{\ell,n} \leq \zeta_{m,n+1} < \zeta_{\ell+1,n},
\]
and observe that \(|M_{m,n+1} - \tilde{M}_{m,n+1}| \leq 2^{-n}\) and
\[
Y_{n+1}(t) - Y_n(t) = \sum_{m=0}^{\infty} (M_{m,n+1} - \tilde{M}_{m,n+1}) \Delta_{m,n+1}(t).
\]
Since \(M_{m,n+1} - \tilde{M}_{m,n+1}\) is \(\mathcal{F}_{\zeta_{m,n+1}}\)-measurable, the terms in this series are orthogonal in \(L^2(\mathbb{P}; \mathbb{R})\), and therefore, by Doob’s inequality,
\[
\mathbb{E}^\mathbb{P}[\|Y_{n+1}(\cdot) - Y_n(\cdot)\|^2_{[0,T]}] \leq 2^{-n} \sum_{m=0}^{\infty} \mathbb{E}^\mathbb{P}[\Delta_{m,n+1}(t)^2]
= 2^{-n} \sum_{m=0}^{\infty} \left[ M(t \wedge \zeta_{m+1,n+1})^2 - M(t \wedge \zeta_{m,n+1})^2 \right] = 2^{-n} \mathbb{E}^\mathbb{P}[M(t)^2],
\]
and the desired convergence follows from this. \(\square\)

It is important to appreciate how much more subtle than the preceding Meyer’s reasoning was. Perhaps the most subtle aspect of his theory is the one having to do with uniqueness. In the discrete setting, uniqueness relied on Doob’s insistence that \(A_n\) be \(\mathcal{F}_{n-1}\)-measurable, but what is the analog of that requirement in the continuous parameter context? Loosely speaking, Meyer’s answer is that \(A(\cdot)\) should have the property that \(A(t)^2 = 2 \int_0^t A(\tau) \, dA(\tau)\). An example that illustrates this point is the simple Poisson process \(t \leadsto N(t)\), the one that starts at 0, waits a unit exponential holding time before jumping to 1, waits there for a second, independent unit exponential holding time before jumping to 2, etc. Since it is non-decreasing, \(N(t)\) is a submartingale with respect to any filtration \(\{\mathcal{F}_t : t \geq 0\}\) for which \(N(\cdot)\) is progressively measurable. Further, there is no doubt that \(N(t) - A(t)\) is a martingale if one take \(A(t) = N(t)\). But is not the choice that Meyer’s theory makes. Instead, his theory would choose \(A(t) = t\), because, being continuous, since \(N(t) - t\) is also a martingale and this choice satisfies \(A(t) = 2 \int_0^t A(\tau) \, dA(\tau)\). More generally, if a continuous \(A(\cdot)\) exists, his theory would choose it. However, in general, there is no continuous choice of \(A(\cdot)\), and to deal with those cases Meyer had to introduce a raft of new ideas.

From now on, I will use \(\langle M \rangle(\cdot)\) to denote the function \(A(\cdot)\) in Theorem 99. In addition, given a stopping time \(\zeta\), \(M^\zeta\) will be the local martingale \(t \leadsto M(t \wedge \zeta)\). The following lemma contains a few elementary facts that will be used in the next section.
Lemma 100. Let \((M(t), \mathcal{F}_t, \mathbb{P})\) be a continuous local martingale and \(\zeta\) a stopping time. Then \(\langle \langle M^\zeta \rangle \rangle(t) = \langle M \rangle(t \wedge \zeta)\). Furthermore, if \(\mathbb{E}^\mathbb{P}[\langle \langle M \rangle \rangle(\zeta)] < \infty\), then
\[
(\langle M^\zeta \rangle(t), \mathcal{F}_t, \mathbb{P}) \text{ and } (\langle M^\zeta \rangle(t)^2 - \langle \langle M^\zeta \rangle \rangle(t), \mathcal{F}_t, \mathbb{P})
\]
are martingales. Finally, if \(\alpha : \Omega \to \mathbb{R}\) is a bounded, \(\mathcal{F}_\zeta\)-measurable function and \(\tilde{M}(t) = \alpha(M(t) - M^\zeta(t))\), then \((\tilde{M}(t), \mathcal{F}_t, \mathbb{P})\) is a continuous, local martingale and
\[
\langle \langle \tilde{M} \rangle \rangle(t) = \alpha^2(\langle M \rangle(t) - \langle \langle M^\zeta \rangle \rangle(t)).
\]

Proof: Without loss in generality, I will assume that \(M(0) = 0\).

To prove the first assertion, simply observe that, by Doob’s stopping time theorem, \((M^\zeta(t)^2 - \langle \langle M \rangle \rangle(t \wedge \zeta), \mathcal{F}_t)\) is a local martingale, and therefore, by uniqueness, \(\langle \langle M^\zeta \rangle \rangle(t) = \langle \langle M \rangle \rangle(t \wedge \zeta)\).

Now assume that \(\mathbb{E}^\mathbb{P}[\langle \langle M \rangle \rangle(\zeta)] < \infty\). Choose stopping times \(\{\zeta_m : m \geq 1\}\) for the local martingale \((M^\zeta(t), \mathcal{F}_t, \mathbb{P})\). By Doob’s inequality,
\[
\mathbb{E}^\mathbb{P}[\| M^{c_\zeta \wedge \zeta_m} \|^2_{[0,t]}] \leq 4\mathbb{E}^\mathbb{P}[\langle \langle M \rangle \rangle(t \wedge \zeta_m \wedge \zeta)] \leq 4\mathbb{E}^\mathbb{P}[\langle \langle M \rangle \rangle(\zeta)],
\]
and therefore \(\mathbb{E}^\mathbb{P}[\| M^\zeta \|^2_{[0,t]}] < \infty\). In particular, this means that \(M^\zeta(t \wedge \zeta_m) \to M^\zeta(t)\) in \(L^2(\mathbb{P}, \mathbb{R})\), and so, since \(\langle \langle M^\zeta \rangle \rangle(t \wedge \zeta_m) \not\to \langle \langle M^\zeta \rangle \rangle(t)\), there is nothing more to do.

To prove the final assertion, it suffices to prove that
\[
(\tilde{M}(t), \mathcal{F}_t, \mathbb{P}) \text{ and } (\tilde{M}(t)^2 - \alpha^2(\langle \langle M \rangle \rangle(t) - \langle \langle M^\zeta \rangle \rangle(t)), \mathcal{F}_t, \mathbb{P})
\]
are martingales if \((M(t), \mathcal{F}_t, \mathbb{P})\) is a bounded martingale. Thus, assume that \((M(t), \mathcal{F}_t, \mathbb{P})\) is a bounded martingale, and let \(t > s\) and \(\Gamma \in \mathcal{F}_s\) be given. Then
\[
\mathbb{E}^\mathbb{P}[\tilde{M}(t), \Gamma] = \mathbb{E}^\mathbb{P}[\tilde{M}(t), \Gamma \cap \{\zeta \leq s\}] + \mathbb{E}^\mathbb{P}[\tilde{M}(t), \Gamma \cap \{s < \zeta \leq t\}].
\]
Because \(\Gamma \cap \{\zeta \leq s\} \in \mathcal{F}_s\) and \(\mathbf{1}_{\Gamma \cap \{\zeta \leq s\}}\) is \(\mathcal{F}_s\)-measurable,
\[
\mathbb{E}^\mathbb{P}[\tilde{M}(t), \Gamma \cap \{\zeta \leq s\}] = \mathbb{E}^\mathbb{P}[\alpha(M(t) - M(t \wedge \zeta)), \Gamma \cap \{\zeta \leq s\}] = \mathbb{E}^\mathbb{P}[\alpha(M(s) - M(s \wedge \zeta)), \Gamma \cap \{\zeta \leq s\}] = \mathbb{E}^\mathbb{P}[\tilde{M}(t), \Gamma].
\]
At the same time, \(\mathbf{1}_{\Gamma \cap \{s < \zeta \leq t\}}\) is \(\mathcal{F}_{t \wedge \zeta}\)-measurable, and so, by Hunt’s version of the stopping time theorem,
\[
\mathbb{E}^\mathbb{P}[\alpha M(t), \Gamma \cap \{s < \zeta \leq t\}] = \mathbb{E}^\mathbb{P}[\alpha M(t \wedge \zeta), \Gamma \cap \{s < \zeta \leq t\}],
\]
which means that \(\mathbb{E}^\mathbb{P}[\tilde{M}(t), \Gamma \cap \{s < \zeta \leq t\}] = 0\). Thus \((\tilde{M}(t), \mathcal{F}_t, \mathbb{P})\) is a martingale. To show that \((\tilde{M}(t)^2 - \alpha^2(\langle \langle M \rangle \rangle(t) - \langle \langle M^\zeta \rangle \rangle(t)), \mathcal{F}_t, \mathbb{P})\) is martingale, use Hunt’s theorem to see first that
\[
\mathbb{E}^\mathbb{P}[\alpha^2 M(t)M(t \wedge \zeta), \Gamma \cap \{s < \zeta \leq t\}] = \mathbb{E}^\mathbb{P}[\alpha^2 M(t \wedge \zeta)^2, \Gamma \cap \{s < \zeta \leq t\}]
\]
and then that
\[ E^P[\alpha^2(M(t)^2 - M(t \wedge \zeta)^2, \Gamma \cap \{ s < \zeta \leq t \}] = E^P[\alpha^2((M(t) - \langle M \rangle(t)), \Gamma \cap \{ s < \zeta \leq t \}]. \]

Hence,
\[ E^P[\alpha^2(M(t) - M^\zeta(t))^2, \Gamma \cap \{ \zeta > s \}] = E^P[\alpha^2((M(t) - \langle M \rangle(t)), \Gamma \cap \{ \zeta > s \}]. \]

Next, since \( \alpha 1_{\zeta \leq s} \) is \( F_\cdot \)-measurable,
\[ E^P[\alpha^2(M(t) - M^\zeta(t))^2, \Gamma \cap \{ \zeta \leq s \}] \]
\[ = E^P[\alpha^2(M(t)^2 - 2M(t)M^\zeta(s) + M^\zeta(s)^2, \Gamma \cap \{ \zeta \leq s \}] \]
\[ = E^P[\alpha^2(M(s)^2 - 2M(s)M^\zeta(s) + M(s)^2 + \langle M \rangle(t) - \langle M \rangle(s), \Gamma \cap \{ \zeta \leq s \}] \]
\[ = E^P[\alpha^2(M(s) - M^\zeta(s))^2, \Gamma] + E^P[\alpha^2((M(t) - \langle M \rangle(t), \Gamma \cap \{ \zeta \leq s \}]. \]

After combining these, one obtains
\[ E^P[\alpha^2(M(t) - M^\zeta(t))^2 - \alpha^2(M(s) - M^\zeta(s))^2, \Gamma] \]
\[ E^P[\alpha^2((M(t) - \langle M \rangle(t), \Gamma \cap \{ \zeta \leq s \}]. \quad \Box. \]

**Kunita-Watanabe’s Integral:** As I said, Doob already understood that Itô’s integration theory could be extended to martingales other than Brownian motion once one had a result like the one in Theorem 99. However, it was Kunita and Watanabe who not only carried out the program that Doob had in mind but did so in a particularly elegant fashion.

Given \( (\Omega, F, P) \) be a complete probability space and \( \{ F_t : t \geq 0 \} \) a non-decreasing filtration of complete sub \( \sigma \)-algebras, use \( M^2(P; \mathbb{R}) \) and \( M_{loc}(P; \mathbb{R}) \), respectively, to denote the set of all square integral continuous martingales and local martingales under \( P \) relative to \( \{ F_t : t \geq 0 \} \). Clearly, both \( M^2(P; \mathbb{R}) \) and \( M_{loc}(P; \mathbb{R}) \) are vector spaces over \( \mathbb{R} \). Next, given \( M_1, M_2 \in M_{loc}(P; \mathbb{R}) \), set
\[ \langle M_1, M_2 \rangle = \frac{\langle M_1 + M_2 \rangle - \langle M_1 - M_2 \rangle}{4}, \]
and note that
\[ (M_1(t)M_2(t) - \langle M_1, M_2 \rangle(t), F_t, P) \]
is a continuous local martingale. Moreover, by exactly the same argument with which we proved uniqueness in Theorem 99, one sees that \( \langle M_1, M_2 \rangle \) is the only progressively measurable, continuous function \( A(\cdot) \) such that \( A(0) = 0, A(\cdot) \uparrow [0, t] \) has bounded variation for each \( t \geq 0 \), and \( (M_1(t)M_2(t) - A(t), F_t, P) \) is a continuous local martingale.
Lemma 101. The map \((M_1, M_2) \mapsto \langle M_1, M_2 \rangle\) is symmetric and bilinear on \(M_{\text{loc}}(\mathbb{P}; \mathbb{R})^2\) in the sense that \(\langle M_1, M_2 \rangle = \langle M_2, M_1 \rangle\) and, for all \(\alpha_1, \alpha_2 \in \mathbb{R}\),

\[
\langle \alpha_1 M_1 + \alpha_2 M_2, M_3 \rangle = \alpha_1 \langle M_1, M_3 \rangle + \alpha_2 \langle M_2, M_3 \rangle
\]

\(\mathbb{P}\)-almost surely. Furthermore, \(\langle M, M \rangle = \langle\langle M \rangle\rangle \geq 0\) and, for \(0 \leq s < t\),

\[
\left| \langle M_1, M_2 \rangle(t) - \langle M_1, M_2 \rangle(s) \right| \leq \sqrt{\langle\langle M_1 \rangle\rangle(t) - \langle\langle M_1 \rangle\rangle(s)} \sqrt{\langle\langle M_2 \rangle\rangle(t) - \langle\langle M_2 \rangle\rangle(s)}
\]

\(\mathbb{P}\)-almost surely. In particular,

\[
\| \sqrt{\langle\langle M_2 \rangle\rangle} - \sqrt{\langle\langle M_1 \rangle\rangle} \|_{[0, t]} \leq \sqrt{\langle\langle M_2 - M_1 \rangle\rangle(t)} \quad (\text{a.s., } \mathbb{P}),
\]

and for any stopping time \(\zeta\),

\[
\langle M_1^\zeta, M_2 \rangle(t) = \langle M_1, M_2 \rangle(t \wedge \zeta).
\]

Proof: The first assertions are all easy applications of uniqueness. To prove the Schwarz type inequality, one uses the same reasoning as usual. That is,

\[
0 \leq \langle\langle \alpha M_1 \pm \alpha^{-1} M_2 \rangle\rangle = \alpha^2 \langle\langle M_1 \rangle\rangle \pm 2 \langle M_1, M_2 \rangle + \alpha^{-2} \langle\langle M_2 \rangle\rangle
\]

\(\mathbb{P}\)-almost surely, first for each \(\alpha\) and then for all \(\alpha\) simultaneously. Thus, \(\mathbb{P}\)-almost surely,

\[
2 \left| \langle M_1, M_2 \rangle \right| \leq \alpha^2 \langle\langle M_1 \rangle\rangle + \alpha^{-2} \langle\langle M_2 \rangle\rangle,
\]

and so, after minimizing with respect to \(\alpha\), one sees that

\[
\left| \langle M_1, M_2 \rangle \right| \leq \sqrt{\langle\langle M_1 \rangle\rangle} \sqrt{\langle\langle M_2 \rangle\rangle} \quad (\text{a.s., } \mathbb{P}).
\]

Finally, by applying this with \(M_i(t) - M_i(t \wedge s)\) in place of \(M_i(t)\), one arrives at (102).

Once one has (102), the first of the concluding inequalities follows in the same way as the triangle inequality follows from Schwarz’s inequality. To prove the second, note that, by (102),

\[
\left| \langle M_1^\zeta, M_2 \rangle(t) - \langle M_1^\zeta, M_2 \rangle(t \wedge \zeta) \right| \leq \sqrt{\langle\langle M_1^\zeta \rangle\rangle(t) - \langle\langle M_1^\zeta \rangle\rangle(t \wedge \zeta)} \sqrt{\langle\langle M_2 \rangle\rangle(t)} = 0
\]

and

\[
\left| \langle M_1^\zeta, M_2 \rangle(t \wedge \zeta) - \langle M_1, M_2 \rangle(t \wedge \zeta) \right| \leq \sqrt{\langle\langle M_1^\zeta - M_1 \rangle\rangle(t \wedge \zeta)} \sqrt{\langle\langle M_2 \rangle\rangle(t)} = 0,
\]
since
\[ \langle \langle M_1^\xi \rangle \rangle (t, \zeta) = \langle \langle (M_1^\xi - M_1)\zeta \rangle \rangle (t) \]
and \((M_1^\xi - M_1)^\xi = M_1^\xi - M_1^\xi = 0. \square \]

Because \( \langle \langle M \rangle \rangle (t) \) is continuous and non-decreasing, it determines a non-atomic, locally finite Borel measure \( \langle \langle M \rangle \rangle(dt) \) on \([0, \infty)\), and because, \( \langle M_1, M_2 \rangle \) is continuous and of locally bounded variation, it determines a non-atomic, locally finite Borel signed measure \( \langle M_1, M_2 \rangle(dt) \) there. Starting from (102), it is easy to show that, for all \( \alpha > 0, T > 0, \) and \( \varphi \in C([0, T]; \mathbb{R}), \)
\[
2 \int_0^T \varphi(\tau) \langle M_1, M_2 \rangle(d\tau) \leq \alpha^2 \int_0^T \varphi(\tau)^2 \langle \langle M_1 \rangle \rangle(d\tau) + \alpha^{-2} \langle \langle M_2 \rangle \rangle(T),
\]
and therefore that
\[
\left| \int_0^T \varphi(\tau) \langle M_1, M_2 \rangle(d\tau) \right| \leq \sqrt{\int_0^T \varphi(\tau)^2 \langle \langle M_1 \rangle \rangle(d\tau) \sqrt{\langle \langle M_2 \rangle \rangle(T)}}.
\]
Because this inequality holds for all continuous \( \varphi \)'s on \([0, T], \) it also holds for all Borel measurable ones, and therefore we know that, for all Borel measurable \( \varphi : [0, T] \rightarrow \mathbb{R}, \)
\[
(103) \quad \left| \int_0^T \varphi(\tau) \langle M_1, M_2 \rangle(d\tau) \right| \leq \sqrt{\int_0^T \varphi(\tau)^2 \langle \langle M_1 \rangle \rangle(d\tau) \sqrt{\langle \langle M_2 \rangle \rangle(T)}},
\]
where \( \langle \langle M_1, M_2 \rangle \rangle(dt) \) is the variation measure determined by \( \langle M_1, M_2 \rangle(dt) \).

With these preparations, we can say how Kunita and Watanabe defined stochastic integrals with respect to an element \( M \) of \( M_{loc}(\mathbb{P}; \mathbb{R}) \). Use \( P_{loc}^2 * M[\mathbb{R}] \) to denote the space of progressively measurable functions \( \xi \) such that \( \int_0^T |\xi(\tau)|^2 \langle \langle M \rangle \rangle(d\tau) < \infty \) for locally integrable with respect to \( \langle \langle M, M' \rangle \rangle(dt) \) of all \( M' \in M_{loc}(\mathbb{P}; \mathbb{R}) \), and define the stochastic integral \( I_\xi^M \) of \( \xi \) with respect to \( M \) to be the element of \( M_{loc}(\mathbb{P}; \mathbb{R}) \) such that \( I_\xi^M(0) = 0 \) and \( \langle I_\xi^M, M' \rangle(dt) = \xi(t)\langle M, M' \rangle(dt) \) for all \( M' \in M_{loc}(\mathbb{P}; \mathbb{R}) \). It is obvious that this definition uniquely determines \( I_\xi^M \) since, if \( I \) and \( J \) were two such elements of \( M_{loc}(\mathbb{P}; \mathbb{R}), \) then \( \langle I - J, I - J \rangle(dt) = 0. \) Thus, before adopting this definition, all that we have to do is prove that such an element exists.

**Lemma 104.** If \( I_\xi^M \) exists, then \( I_\xi^M \) and \( I_{1[0, \xi]}^M \) exist and both are equal to \((I_\xi^M)^\xi. \)** Next suppose that \( \{\xi_n : n \geq 0\} \cup \{\xi\} \subseteq P_{loc}^2(M; \mathbb{R}) \) and that
\[
\lim_{n \to \infty} \int_0^t (\xi_n(\tau) - \xi(\tau))^2 \langle \langle M \rangle \rangle(d\tau) = 0 \quad (a.s., \mathbb{P})
\]
for all \( t \geq 0. \) If \( I_{1n}^M \) exists for all \( n \geq 0, \) then \( I_{1}^M \) does also.
Proof: The first assertion is an easy application of the last part of Lemma 101. Indeed,

\[
\langle (I_t^M)_{\xi}, M' \rangle (t) = \langle I_t^M, M' \rangle (t \wedge \zeta)
\]

\[
= \int_0^{t \wedge \zeta} \xi(\tau) \langle M, M \rangle (d\tau) = \int_0^t \mathbf{1}_{[0, \zeta]}(\tau) \xi(\tau) \langle M, M' \rangle (d\tau),
\]

\[
\text{and } \langle M^\zeta, M \rangle (d\tau) = \mathbf{1}_{[0, \zeta]}(\tau) \langle M, M' \rangle (d\tau).
\]

Turning to the second assertion, begin by assuming that

\[
\|M\|_{[0, t]} \text{ and } \sup_{n \geq 0} \int_0^t \xi_n(\tau)^2 \langle M \rangle (d\tau)
\]

are uniformly bounded for each \(t \geq 0\), and set \(I_n = I_{\xi_n}^M\). Then

\[
\left( (I_n(t) - I_m(t))^2 - \int_0^t (\xi_n(\tau) - \xi_m(\tau))^2 \langle M \rangle (d\tau), \mathcal{F}_t, \mathbb{P} \right)
\]

is a martingale, and so

\[
\mathbb{E}^\mathbb{P}\left[ (I_n(\cdot) - I_m(\cdot))^2 \right] = \mathbb{E}^\mathbb{P}\left[ \int_0^t (\xi_n(\tau) - \xi_m(\tau))^2 \langle M \rangle (d\tau) \right],
\]

which, by Doob’s inequality, means that

\[
\lim_{m \to \infty} \sup_{n \geq m} \mathbb{E}^\mathbb{P}\left[ \|I_n(\cdot) - I_m(\cdot)\|^2_{[0, t]} \right] = 0.
\]

Hence there exists a continuous, square integrable martingale \(I\) such that

\[
\|I - I_m\|_{[0, t]} \to 0 \text{ in } L^2(\mathbb{P}; \mathbb{R}) \text{ for all } t \geq 0.
\]

In addition, if \(M' \in M_{\text{loc}}(\mathbb{P}; \mathbb{R})\) and \(\langle M' \rangle (t)\) is bounded for all \(t \geq 0\), then, for \(0 \leq s < t\),

\[
\mathbb{E}^\mathbb{P}\left[ I(t) M'(t) \mid \mathcal{F}_s \right] = \lim_{n \to \infty} \mathbb{E}^\mathbb{P}\left[ I_n(t) M'(t) \mid \mathcal{F}_s \right] = \lim_{n \to \infty} \mathbb{E}^\mathbb{P}\left[ \int_0^t \xi_n(\tau) \langle M, M' \rangle (d\tau) \right]
\]

and

\[
\int_0^t |\xi(\tau) - \xi_n(\tau)| |\langle M, M' \rangle| (d\tau) \leq \sqrt{\int_0^t (\xi(\tau) - \xi_n(\tau))^2 \langle M \rangle (d\tau) \sqrt{\langle M' \rangle (t)}} \to 0
\]
\(\mathbb{P}\)-almost surely. Hence, \(\langle I, M' \rangle(dt) = \xi(t)\langle M, M' \rangle(dt)\) (a.s., \(\mathbb{P}\)) when \(\langle M' \rangle(t)\) is bounded for all \(t \geq 0\). To prove that the same equality for general \(M'\), take \(\zeta_k = \inf\{t \geq 0 : \langle M' \rangle(t) \geq k\}\). Then, using Lemma 101, one sees that

\[
\langle I, M' \rangle(t \wedge \zeta_k) = \langle I, (M')^{\zeta_k} \rangle(t) = \int_0^{t \wedge \zeta_k} \xi(\tau)\langle M, M' \rangle(d\tau)
\]

\(\mathbb{P}\)-almost surely for all \(k \geq 1\). Since \(\zeta_k \not\nearrow \infty\) as \(k \to \infty\), it follows that \(I = I^M_\xi\).

To remove the boundedness assumptions on \(M\) and \(\int_0^t \xi_n(\tau)^2 \langle M' \rangle(d\tau)\), define

\[
\zeta_k = \inf\left\{t \geq 0 : |M(t)| \text{ess sup}_{n \geq 0} \int_0^t \xi_n(\tau)^2 \langle M' \rangle(t) \geq k\right\},
\]

and set \(M_k = M^{\zeta_k}\). By the preceding, we know that \(I_k = I^M_{\xi_k}\) exists and is equal to \(I^M_k\), and, by the first part of this lemma, \(I_{k+1}(t \wedge \zeta_k) = I_k(t \wedge \zeta_k)\) (a.s., \(\mathbb{P}\)) for all \(k \geq 1\). Hence, we can choose \(I \in M_{loc}(\mathbb{P}; \mathbb{R})\) so that \(I(t \wedge \zeta_k) = I(t \wedge \zeta_k)\) (a.s., \(\mathbb{P}\)) for all \(k \geq 1\), in which case, it is easy to check that \(\langle I, M' \rangle(dt) = \xi(t)\langle M, M' \rangle(dt)\) for all \(M' \in M_{loc}(\mathbb{P}; \mathbb{R})\) \(\square\)

In view of the preceding, what remains is to find a sufficiently rich set of \(\xi\)'s for which we can show that \(I^M_\xi\) exists, and as in the case when \(M\) was a Brownian motion, a good guess is that \(\xi\)'s of bounded variation should be the place to look. The basic result in this case requires that we know the following elementary fact about Riemann-Stieltjes integrals. Let \(\varphi \in C([0, t]; \mathbb{R})\) with \(\varphi(0) = 0\), and let \(\{\psi_n : n \geq 0\}\) be a sequence of functions on \([0, t]\) such that \(|\psi_n(0)| \vee \text{var}_{[0, t]}(\psi) \leq C < \infty\). If \(\psi_n \to \psi\) pointwise, then \(\text{var}_{[0, t]}(\psi) \leq C\) and

\[
\lim_{n \to \infty} \int_0^t \psi_n(\tau) d\varphi(\tau) = \int_0^t \psi(\tau) d\varphi(\tau).
\]

To prove this result, first note that \(\text{var}_{[0, t]}(\psi) \leq C\) is obvious. Next, choose \(\{\varphi_k : k \geq 1\} \subseteq C^1(\mathbb{R}; \mathbb{R})\) so that \(||\varphi_k||_u \leq ||\varphi||_{[0, t]}\) and \(||\varphi - \varphi_k||_{[0, t]} \leq \frac{1}{k}\). Then

\[
\left| \int_0^t \psi_n(\tau) d\varphi(\tau) - \int_0^t \psi_n(\tau) d\varphi_k(\tau) \right| \\
\leq |\varphi(t) - \varphi_k(t)||\psi(t)| + \left| \int_0^t (\varphi(\tau) - \varphi_k(\tau)) d\psi_n(\tau) \right| \\
\leq 2C||\varphi - \varphi_k||_{[0, t]} \leq \frac{2C}{k},
\]

and similarly

\[
\left| \int_0^t \psi(\tau) d\varphi(\tau) - \int_0^t \psi(\tau) d\varphi_k(\tau) \right| \leq \frac{2C}{k}.
\]
Hence, it suffices to show that
\[
\lim_{n \to \infty} \int_0^t \psi_n(\tau) \, d\varphi_k(\tau) = \int_0^t \psi(\tau) \, d\varphi_k(\tau)
\]
for each \( k \geq 1 \). But, by Lebesgue’s dominated convergence theorem,
\[
\lim_{n \to \infty} \int_0^t \psi_n(\tau) \, d\varphi_k(\tau) = \lim_{n \to \infty} \int_0^t \psi_n(\tau) \varphi_k(\tau) \, d\tau = \int_0^t \psi(\tau) \varphi_k(\tau) \, d\tau.
\]

**Lemma 105.** Suppose \( M \in M_{\text{loc}}(\mathbb{P}; \mathbb{R}) \) and that \( \xi \) is an element of \( \mathcal{P}_{\text{loc}}^2(M; \mathbb{R}) \) for which \( \xi(\cdot) \) has locally bounded variation. Then \( I_\xi^M \) exists and \( I_\xi^M(t) \) is equal to the Riemann-Stieltjes integral \( \int_0^t \xi(\tau) \, dM(\tau) \).

**Proof:** Begin with the assumption that \( M, \langle\langle M\rangle\rangle \), and \( \xi \) are all uniformly bounded. Set \( I(t) = \int_0^t \xi(\tau) \, dM(\tau) \), where the integral is taken in the sense of Riemann-Stieltjes. Then \( I(t) = \lim_{n \to \infty} I_n(t) \), where
\[
I_n(t) \equiv \int_0^t \xi([\tau]_n) \, dM(\tau) = \sum_{m < 2^{-n} t} \xi(m2^{-n})(M(t \wedge (m + 1)2^{-n}) - M(t \wedge m2^{-n})),
\]
and, because \( \langle\langle M\rangle\rangle(dt) \) is non-atomic and \( \xi \) has at most a countable number of discontinuities,
\[
\lim_{n \to \infty} \int_0^t (\xi(\tau) - \xi_n(\tau))^2 \langle\langle M\rangle\rangle(d\tau) = 0,
\]
where \( \xi_n(t) = \xi([t]_n) \). Hence, by Lemma 104, we will know that \( I = I_\xi^M \) once we show that \( I_n = I_{\xi_n}^M \). To this end, note that
\[
I_n(t) = \sum_{m < 2^{-n} t} \Delta_{m2^{-n}}(t \wedge (m + 1)2^{-n}).
\]
where \( \Delta_s(t) \equiv \xi(s)(M(t) - M(t \wedge s)) \). By Lemma 100, \( \Delta_s \in M_{\text{loc}}(\mathbb{P}; \mathbb{R}) \) for each \( s \geq 0 \), and so \( I_n \in M_{\text{loc}}(\mathbb{P}; \mathbb{R}) \). Next, let \( M' \in M_{\text{loc}}(\mathbb{P}; \mathbb{R}) \) be given, and set \( X(t) = M(t)t' - \langle\langle M\rangle\rangle(t) \). Then, again by Lemma 100,
\[
\left( \xi(s)(X(t) - X(t \wedge s)), \mathcal{F}_t, \mathbb{P} \right) \text{ and } \left( \xi(s)M(s)(M'(t) - M'(t \wedge s)), \mathcal{F}_t, \mathbb{P} \right)
\]
are local martingales. Hence, since
\[
\Delta_s(t)M'(t) - \xi(s)(\langle\langle M,M'\rangle\rangle(t) - \langle\langle M,M'\rangle\rangle(t \wedge s)) = \xi(s)(X(t) - X(t \wedge s)) - \xi(s)M(s)(\langle\langle M'\rangle\rangle(t) - \langle\langle M'\rangle\rangle(t \wedge s)),
\]

Lemma 106.  \( I_\xi(t) = \int_0^t \xi(s) \, dM(s) \), and therefore \( I_\xi(M)(t) = \xi_n(t) \langle M, M' \rangle(\langle 0, m \rangle) \), and therefore \( \langle X_n, M' \rangle(dt) = \xi_n(t) \langle M, M' \rangle(dt) \).

To remove the boundedness assumption on \( \xi \), for each \( n \geq 1 \), define \( \xi_n(t) = \xi(t) \) if \( |\xi(t)| \leq n \) and \( \xi_n(t) = \pm n \) if \( |\xi(t)| > n \). Then \( \var{_{[0, t]}(\xi_n)} \leq \var{_{[0, t]}(\xi)} \), \( |\xi_n(t)| \leq n \wedge |\xi(t)| \), and \( \int_0^t (\xi(t) - \xi_n(t))^2 \var{_{(0, t)}}(d\tau) \to 0 \). Hence, by the preceding combined with the above comment and Lemma 104, \( I_\xi^M(t) \to \int_0^t \xi(s) \, dM(s) \), and \( I_{\xi_n}^M(t) \to I_\xi^M(t) \).

Finally, to remove the assumption that \( M \) and \( \langle M \rangle \) are bounded, take \( \zeta_n = \inf\{ t \geq 0 : |\xi(t)| \leq n \} \) and \( \zeta_n = M/\zeta_n \). Then, just as in the proof of Lemma 104, one can define \( I(t) = I_M^\zeta(t) \) for \( t \in [0, \zeta_n] \), in which case \( I = I_\xi^M(t) \) and \( I(t) = \int_0^t \xi(s) \, dM(s) \). \( \square \)

To complete the program, we have to show that if \( \xi \in \mathcal{P}_\text{loc}(M; \mathbb{R}) \) then there exist \( \{ \xi_n : n \geq 1 \} \subseteq \mathcal{P}_\text{loc}(M; \mathbb{R}) \) such that each \( \xi_n \) has locally bounded variation and \( \int_0^t (\xi(t) - \xi_n(t))^2 \var{_{(0, t)}}(d\tau) \to 0 \). The construction of such a sequence is a little trickier than it was in the Brownian case and makes use of the following lemma.

Lemma 106.  Let \( F : [0, \infty) \to \mathbb{R} \) be a continuous, non-decreasing function with \( F(0) = 0 \), and define

\[
F^{-1}(t) = \inf\{ \tau \geq 0 : F(\tau) \geq t \} \quad \text{for } t \geq 0.
\]

Then \( F^{-1} \) is a right-continuous and non-decreasing on \([0, F(\infty))\), \( F \circ F^{-1}(t) = t \), \( F^{-1} \circ F(\tau) \leq \tau \), and

\[
D = \{ \tau : F^{-1} \circ F(\tau) < \tau \}
\]

is either empty or the at most countable union of mutually disjoint intervals of the form \((a, b)\) with the properties that \( 0 \leq a < b \leq \infty \), \( F(a) = F(b) \), \( F(\tau) < F(a) \) for \( \tau < a \), and \( F(\tau) > F(b) \) for \( \tau > b \). Furthermore, if \( F(d\tau) \) is the Borel measure determined by \( F \), then \( F(D) = 0 \) and

\[
\int_0^{F(T)} f \circ F^{-1}(t) \, dt = \int_0^T f(\tau) \, F(d\tau)
\]

for non-negative, Borel measurable \( f \) on \([0, \infty)\). Finally, set \( \rho_\epsilon(t) = \epsilon^{-1} \rho(\epsilon^{-1} t) \) where \( \rho \in C^\infty(\mathbb{R}; [0, \infty]) \) vanishes off of \((0, 1)\) and \( \int \rho(t) \, dt = 1 \), and, given \( f \in L^2(F; [0, \infty)) \), set

\[
f_\epsilon(\tau) = \int_0^{F(\infty)} \rho_\epsilon(F(\tau) - \sigma) f \circ F^{-1}(\sigma) \, d\sigma.
\]

Then \( \|f_\epsilon\|_{L^2(F; \mathbb{R})} \leq \|f\|_{L^2(F; \mathbb{R})} \) and \( \|f - f_\epsilon\|_{L^2(F; \mathbb{R})} \to 0 \) as \( \epsilon \searrow 0 \).
Proof: It is easy to check that $F^{-1}$ is right-continuous, non-decreasing, and satisfies $F^{-1} \circ F(\tau) \leq \tau$. In addition, if $F^{-1} \circ F(\tau) < \tau$, then $\tau \in (a, b]$ where $a = \inf\{\sigma : F(\sigma) = F(\tau)\}$ and $b = \sup\{\sigma : F(\sigma) = F(\tau)\}$. Thus, if $D \neq \emptyset$, then it is the union of mutually disjoint intervals of the described sort and there can be at most a countable number of such intervals. Hence, since $F((a, b)) = F(a) - F(b)$, it follows that $F(D) = 0$.

Next, by the standard change of variables formula for Riemann-Stieltjes integrals,

$$\int_0^T f \circ F(\tau) \, dF(\tau) = \int_0^{F(T)} f(t) \, dt$$

for all non-negative Borel measurable $f$'s. Hence, because $F(D) = 0$,

$$\int_0^{F(T)} f \circ F^{-1}(t) \, dt = \int_0^T f \circ (F^{-1} \circ F)(\tau) \, F(d\tau) = \int_0^T f(\tau) \, F(d\tau).$$

In particular,

$$\int_0^\infty f_\epsilon(\tau)^2 \, F(d\tau) = \int_0^{F(\infty)} \rho_\epsilon \ast (f \circ F^{-1})(t)^2 \, dt$$

$$\leq \int_0^{F(\infty)} f \circ F^{-1}(t)^2 \, dt = \int_0^\infty f(\tau)^2 \, F(d\tau),$$

and similarly

$$\|f - f_\epsilon\|_{L^2(F;\mathbb{R})} = \int_0^{F(\infty)} \left( f \circ F^{-1}(t) - \rho_\epsilon \ast (f \circ F^{-1})(t) \right)^2 \, dt \longrightarrow 0 \quad \Box$$

**Theorem 107.** For each $M \in \mathcal{M}_{\text{loc}}(\mathbb{P}; \mathbb{R})$ and $\xi \in \mathcal{P}^2_{\text{loc}}(M; \mathbb{R})$ there exists a unique $I^M_\xi \in \mathcal{M}_{\text{loc}}(\mathbb{P}; \mathbb{R})$ such that $(I^M_\xi, M')(dt) = \xi(t)(M, M')(dt)$ for all $M' \in \mathcal{M}_{\text{loc}}(\mathbb{P}; \mathbb{R})$. Furthermore, if $\xi(\cdot)$ has locally bounded variation, then $I^M(\cdot)$ equals the Riemann-Stieltjes integral on $[0, t]$ of $\xi$ with respect to $M$.

**Proof:** All that remains is to use the existence assertion when $\xi$ isn’t of locally bounded variation. For this purpose, define $\xi_\epsilon$ from $\xi$ by the prescription in Lemma 106 with $F = \langle M \rangle$ and $f = \xi$. Then $\xi_\epsilon$ has locally bounded variation and

$$\lim_{\epsilon \downarrow 0} \int_0^t (\xi(\tau) - \xi_\epsilon(\tau))^2 \langle M \rangle(d\tau) = 0.$$

Thus, by Lemma 101, we are done. $\Box$
In the future, we will call $I^M_\xi$ the **stochastic integral** of $\xi$ with respect to $M$ and will usually use $\int_0^t \xi(\tau) \, dM(\tau)$ to denote $I^M_\xi(t)$. Notice that if $\xi \in \mathcal{P}^2_{\text{loc}}(M; \mathbb{R})$ and $\eta \in \mathcal{P}^2_{\text{loc}}(I^M_\xi; \mathbb{R})$, then $\xi \eta \in \mathcal{P}^2_{\text{loc}}(M; \mathbb{R})$ and

\[
\int_0^t \eta(\tau) \, dI^M_\xi(\tau) = \int_0^t \xi(\tau) \eta(\tau) \, dM(\tau).
\]

To check this, set $J(t) = \int_0^t \eta(\tau) \, dI^M_\xi(\tau)$. Then, for any $M' \in M_{\text{loc}}(\mathbb{P}; \mathbb{R})$,

\[
\langle J, M' \rangle(dt) = \eta(t) \langle I^M_\xi, M' \rangle(dt) = \xi(t) \eta(t) \langle M, M' \rangle(dt).
\]

**Itô's Formula Again**: Having developed the theory of stochastic integration for general continuous, local martingales, it is only reasonable to see what Itô’s formula look like in that context. What follows is Kunita and Watanabe’s version of his formula.

**Theorem 109**. For each $1 \leq i \leq N_1$, let $V_i$ be a continuous, progressively measurable $\mathbb{R}$-valued function of locally bounded variation, and for each $1 \leq j \leq N_2$ let $(M_j(t), \mathcal{F}_t, \mathbb{P})$ be a continuous local martingale. Set $\tilde{V}(t) = (V_1, \ldots, V_{N_1})$, $\tilde{M}(t) = (M_1(t), \ldots, M_{N_2}(t))$, and

\[
A(t) = \left(\langle M_i, M_j \rangle\right)_{1 \leq i, j \leq N}.
\]

If $\varphi \in C^{1,2}(\mathbb{R}^{N_1} \times \mathbb{R}^{N_2}; \mathbb{C})$, then

\[
\varphi(\tilde{V}(t), \tilde{M}(t)) = \varphi(\tilde{V}(0), \tilde{M}(0)) + \int_0^t \left(\nabla(\varphi(\tilde{V}(\tau), \tilde{M}(\tau)), d\tilde{V}(\tau)\right)_{\mathbb{R}^{N_1}} + \int_0^t \left(\nabla(\varphi(\tilde{V}(\tau), \tilde{M}(\tau)), d\tilde{M}(\tau)\right)_{\mathbb{R}^{N_2}} + \frac{1}{2} \int_0^t Trace\left(\nabla^2(\varphi(\tilde{V}(\tau), \tilde{M}(\tau)))dA(\tau)\right).
\]

**Proof**: Begin by observing that for $0 \leq x < t$, $A(t) - A(s)$ is non-negative definite and symmetric, and therefore $\|A(t) - A(s)\|_{\text{op}} \leq \text{Trace}(A(t) - A(s))$.

It is easy to reduce to the case when $\varphi \in C^\infty_\mathbb{C}(\mathbb{R}^{N_1} \times \mathbb{R}^{N_2}; \mathbb{R})$ and $\text{var}_{[0, \infty]}(V)$, $\|M\|_{[0, \infty]}$, and $\sup_{t \geq 0} \|A(t)\|_{\text{op}}$ are all uniformly bounded. Thus we will proceed under these assumptions. In particular, this means that $(\tilde{M}(t), \mathcal{F}_t, \mathbb{P})$ is a bounded, continuous martingale.

The difference between the proof of this general case and the Brownian one is that we now have control only on the second moment of the $|\tilde{M}(t) - \tilde{M}(s)|$, which makes it more general.
Thus, we must rely more heavily on continuity, and the way to do that is to control the increments by using stopping times. With this in mind, for a given $T > 0$, define $\zeta_{0,n} = 0$ for all $n \geq 0$ and, for $m \geq 1$,

$$\zeta_{m,n} = \inf\{t \geq 0 : |\tilde{M}(t) - \tilde{M}(\zeta_{m-1,n})| \vee \text{Trace}(A(t) - A(\zeta_{m-1,n})) \geq 2^{-n}\} \wedge T.$$ 

By continuity, $\zeta_{m,n} = T$ for all but a finite number of $m$’s, and so

$$\varphi(\tilde{V}(T), \tilde{M}(T)) - \varphi(\tilde{X}(0))$$

$$= \sum_{m=0}^{\infty} \int_{\zeta_{m,n}}^{\zeta_{m+1,n}} \left( \nabla_{(1)} \varphi(\tilde{V}(\tau), M(\zeta_{m+1,n})), d\tilde{V}(\tau) \right)$$

$$+ \sum_{m=0}^{\infty} \left( \varphi(\tilde{V}(\zeta_{m,n}), M(\zeta_{m+1,n})) - \varphi(\tilde{V}(\zeta_{m,n}), M(\zeta_{m,n})) \right).$$

Clearly the first sum on the right tends to $\int_0^T \left( \nabla_{(1)} \varphi(\tilde{V}(\tau), \tilde{M}(\tau), d\tilde{V}(\tau) \right)$ as $n \to \infty$. To handle the second sum, use Taylor’s theorem to write

$$\left( \varphi(\tilde{V}(\zeta_{m,n}), M(\zeta_{m+1,n})) - \varphi(\tilde{V}(\zeta_{m,n}), M(\zeta_{m,n})) \right)$$

$$= \left( \nabla_{(2)} \varphi(\tilde{X}_{m,n}), \Delta_{m,n} \right)_{\mathbb{R}^N} + \frac{1}{2} \text{Trace} \left( \nabla_{(2)}^2 \varphi(\tilde{X}_{m,n}) \Delta_{m,n} \otimes \Delta_{m,n} \right) + E_{m,n},$$

where $\tilde{X}_{m,n} = \left( \tilde{V}(\zeta_{m,n}), \tilde{M}(\zeta_{m,n}) \right), \Delta_{m,n} = \tilde{M}(\zeta_{m+1,n}) - \tilde{M}(\zeta_{m,n})$, and $|E_{m,n}| \leq C|\Delta_{m,n}|^3$ for some $C < \infty$. Using Lemma 100, it is easy to show that

$$\lim_{n \to \infty} \sum_{m=0}^{\infty} \left( \nabla_{(2)} \varphi(\tilde{X}_{m,n}), \Delta_{m,n} \right)_{\mathbb{R}^N} = \int_0^T \left( \nabla_{(2)} \varphi(\tilde{V}(\tau), \tilde{M}(\tau)), d\tilde{M}(\tau) \right)_{\mathbb{R}^N},$$

and, by Hunt’s stopping time theorem,

$$\sum_{m=0}^{\infty} \mathbb{E}^P[|\Delta_{m,n}|^2] = \sum_{m=0}^{\infty} \mathbb{E}^P[|M(\zeta_{m+1,n})|^2 - |M(\zeta_{m,n})|^2] = \mathbb{E}^P[|M(T)|^2],$$

and so

$$\sum_{m=0}^{\infty} \mathbb{E}^P[|E_{m,n}|] \leq 2^{-n} C \mathbb{E}^P[|M(T)|^2].$$

Finally, set $D_{m,n} = \Delta_{m,n} \otimes \Delta_{m,n} - (A(\zeta_{m,n}) - A(\zeta_{m,n}))$. Then

$$\sum_{m=0}^{\infty} \text{Trace} \left( \nabla_{(2)}^2 \varphi(\tilde{X}_{m,n}) \Delta_{m,n} \otimes \Delta_{m,n} \right)$$

$$= \sum_{m=0}^{\infty} \text{Trace} \left( \nabla_{(2)}^2 \varphi(\tilde{X}_{m,n})(A(\zeta_{m+1,n}) - A(\zeta_{m,n})) \right)$$

$$+ \sum_{m=0}^{\infty} \text{Trace} \left( \nabla_{(2)}^2 \varphi(\tilde{X}_{m,n})D_{m,n} \right),$$
and the first sum on the right tends to
\[ \int_0^T \text{Trace} \left( \nabla (\varphi) (V(\tau), \tilde{M}(\tau)) \right) dA(\tau) \]
as \( n \to \infty \). At the same time, by Hunt’s stopping time theorem, the terms in
the second sum are orthogonal in \( L^2(\mathbb{P}; \mathbb{R}) \), and therefore the second moment of
that sum is dominated by a constant times
\[ \sum_{m=0}^{\infty} \mathbb{E}_P \left[ \| D_{m,n} \|_{H.S.}^2 \right] . \]
Since \( \| D_{m,n} \|_{H.S.} \leq C (|\Delta_{m,n}|^4 + \| A(\zeta_{m+1,n}) - A(\zeta_{m,n}) \|_{op}^2) \) for some \( C < \infty \),
the preceding sum is dominated by a constant times
\[ 2^{-n} \mathbb{E}_P \left[ |\tilde{M}(T)|^2 + \text{Trace}(A(T)) \right] . \]

A particularly striking application of this result is Kunita and Watanabe’s
derivation of Lévy’s characterization of Brownian motion.

**Corollary 110.** If \((M(t), \mathcal{F}_t, \mathbb{P})\) is an \( \mathbb{R}^N \)-valued, continuous local martingale,
then it is a Brownian motion if and only if \( M(0) = 0 \) and
\( \langle M_i, M_j \rangle = t \delta_{i,j} \) for \( 1 \leq i, j \leq N \).

**Proof:** The necessity is obvious. To prove the sufficiency, first observe that,
because \( \langle \langle M_i \rangle \rangle(t) \) is bounded for all \( 1 \leq i \leq N, t \geq 0, \ (M(t), \mathcal{F}_t, \mathbb{P}) \) a martingale.
Next, given \( \xi \in \mathbb{R}^N \), apply Theorem 109 to show that
\[ E_{\xi}(t) \equiv e^{i \langle \xi, M(t) \rangle_{\mathbb{R}^N}} \left[ \frac{\| \xi \|_{\mathbb{R}^N}^2}{t} \right] = 1 + \int_0^t e^{i \langle \xi, M(\tau) \rangle_{\mathbb{R}^N}} \left[ \frac{\| \xi \|_{\mathbb{R}^N}^2}{t} \right] dM(\tau) \right]_{\mathbb{R}^N} . \]
Finally, if \( X(t) \) and \( Y(t) \) denote the real and imaginary parts of the preceding
stochastic integral, check that \( \langle X \rangle(t) + \langle Y \rangle(t) \leq e^{\| \xi \|^2 t} - 1 \), and therefore
\((E_{ix}(t), \mathcal{F}_t, \mathbb{P})\) is a martingale. Since this means that
\[ \mathbb{E}_P \left[ e^{i \langle \xi, M(t) - M(s) \rangle_{\mathbb{R}^N}} \mid \mathcal{F}_s \right] = e^{-\| \xi \|^2 (t-s) / 2} \text{ for } 0 \leq s < t , \]
the proof is complete. \( \square \)

Another important consequence of Theorem 109 is what it says about the
way local martingales transform under smooth maps. It is clear that linear
maps preserve the martingale property and equally clear that non-linear ones
do not. Nonetheless, Theorem 109 says that the image of a continuous local
martingale under of twice continuously differentiable map is the sum of a local
martingate and a progressively measurable function of locally bounded variation.
In fact, it shows the such a sum is transformed into another such sum, and so it is reasonable to introduce terminology for such stochastic processes. Thus given an \( \mathbb{R} \)-valued progressively measurable function \( t \mapsto X(t) \), one says that \( (X(t), \mathcal{F}_t, \mathbb{P}) \) is a continuous semi-martingale if \( X = M + V \), where \( (M(t), \mathcal{F}_t, \mathbb{P}) \) is a continuous local martingale with \( M(0) = 0 \) and \( V \) is a progressively measurable function for which \( t \mapsto V(t) \) is a continuous function of locally bounded variation.

By the same argument as we used in Lemma 101 to prove uniqueness, one sees that, up to a \( \mathbb{P} \)-null set, \( M \) and \( V \) are uniquely determined. Thus, we can unambiguously talk about the martingale part \( M \) and the bounded variation part \( V \) of a continuous semi-martingale \( X = M + V \), and so we can define \( \langle X_1, X_2 \rangle = \langle M_1, M_2 \rangle \) if \( X_1 = M_1 + V_1 \) and \( X_2 = M_2 + V_2 \). Notice that \( \langle X_1, X_3 \rangle = 0 \) if either \( M_1 = 0 \) or \( M_2 = 0 \). Finally, if \( X = M + V \) and \( \xi \) is an \( \mathbb{R} \)-valued, continuous, progressively measurable function,

\[
\int_0^t \xi(\tau) \, dX(\tau) = \int_0^t \xi(\tau) \, dM(\tau) + \int_0^t \xi(\tau) \, dV(\tau),
\]

where the first integral on the right is a stochastic integral and the second is a Riemann-Stieltjes one. Obviously, such integrals are again semi-martingales, and, using (108) and the properties of Riemann-Stieltjes integrals, one sees that

\[
\int_0^t \eta(\tau) \left( \int_0^\tau \xi(\sigma) \, dX(\sigma) \right) = \int_0^t \xi(\tau) \eta(\tau) \, dX(\tau)
\]

(111)

if \( \eta \) is a second continuous, progressively measurable function.

The following statement is an immediate consequence of Theorem 109.

**Corollary 112.** Suppose that \( \tilde{X}(t) = (X_1(t), \ldots, X_N(t)) \), where \( (X_j(t), \mathcal{F}_t, \mathbb{P}) \) is a continuous local semi-martingale for each \( 1 \leq j \leq N \). and set

\[
A(t) = \left( \langle X_i(t), X_j(t) \rangle \right)_{1 \leq i, j \leq N}.
\]

If \( \varphi \in C^2(\mathbb{R}^N; \mathbb{C}) \), then

\[
\varphi(\tilde{X}(t)) - \varphi(\tilde{X}(0)) = \int_0^t \left( \nabla \varphi(\tilde{X}(\tau)), d\tilde{X}(\tau) \right)_{\mathbb{R}^N} + \frac{1}{2} \int_0^t \text{Trace} \left( \nabla^2 \varphi(\tilde{X}(\tau)) \, dA(\tau) \right).
\]

**Corollary 113.** If \((X_1(t), \mathcal{F}_t, \mathbb{P})\) and \((X_2(t), \mathcal{F}_t, \mathbb{P})\) are a pair of continuous local semi-martingales, then, for all \( T > 0 \),

\[
\lim_{n \to \infty} \sup_{t \in [0, T]} \left| \sum_{m=0}^\infty \left( X_1(t \wedge (m + 1)2^{-n}) - X_1(t \wedge m2^{-n}) \right) \times \left( X_2(t \wedge (m + 1)2^{-n}) - X_2(t \wedge m2^{-n}) - \langle X_1, X_2 \rangle(t) \right) \right| = 0
\]

in \( \mathbb{P} \)-probability.
Proof: First note that, by polarization, it suffices to treat case when \(X_1 = X = X_2\) where \(X\) and \(\langle X \rangle\) are uniformly bounded.

Observe that, by Corollary 112,

\[
\sum_{m=0}^{\infty} \left( X(t \wedge (m+1)2^{-n}) - X(t \wedge m2^{-n}) \right)^2
\]

\[
= \sum_{m=0}^{\infty} \left( X(t \wedge (m+1)2^{-n})^2 - X(t \wedge m2^{-n})^2 \right.
\]

\[
- 2X(m2^{-n})\left( X(t \wedge (m+1)2^{-n}) - X(t \wedge m2^{-n}) \right)
\]

\[
= \langle X \rangle(t) + \int_0^t (X(\tau) - X([\tau]_n))\,dX(\tau).
\]

Next let \(M\) and \(V\) denote the martingale and bounded variation parts of \(X\).

Clearly

\[
\sup_{t \in [0,T]} \left| \int_0^t (X(\tau) - X([\tau]_n))\,dV(\tau) \right| \leq \int_0^T |X(\tau) - X([\tau]_n)|\,d|V|(\tau) \to 0
\]

as \(n \to \infty\). At the same time,

\[
\mathbb{E}^P \left[ \sup_{t \in [0,T]} \left( \int_0^t (X(\tau) - X([\tau]_n))\,dM(\tau) \right)^2 \right]
\]

\[
\leq 4 \mathbb{E}^P \left[ \int_0^T (X(\tau) - X([\tau]_n))^2 \langle M \rangle(d\tau) \right] \to 0,
\]

and, when combined with the preceding, that completes the proof. \(\square\)

Representing Continuous Local Martingales: Suppose that \((M(t),\mathcal{F}_t,\mathbb{P})\) is an \(\mathbb{R}^N\)-valued continuous local martingale with \(M(0) = 0\), and set

\[
A(t) = \left( \langle M_i, M_j \rangle(t) \right)_{1 \leq i, j \leq N}.
\]

Further, assume that \(t \mapsto \langle M_j \rangle(t)\) is absolutely continuous for each \(1 \leq j \leq N\), and therefore, by Lemma 100, for each \(1 \leq i, j \leq N\), \(t \mapsto \langle M_i, M_j \rangle(t)\) is also absolutely continuous. Then there exists a progressively measurable \(\text{Hom}(\mathbb{R}^N; \mathbb{R}^N)\)-valued function \(a\) such that

\[
a(t) = \lim_{h \searrow 0} \frac{A(t) - A(t-h)}{h}
\]

for Lebesgue almost every \(t > 0\) and

\[
A(t) = \int_0^t a(\tau)\,d\tau.
\]
Because $A(t) - A(s)$ is symmetric and non-negative definite for all $0 \leq s < t$, we can and will assume that $a(t)$ is also symmetric and non-negative definite for all $t \geq 0$. By Lemma 29, $(\epsilon I + a)^{\frac{1}{2}}$ is progressively measurable for all $\epsilon > 0$, and so $\sigma \equiv a^{\frac{1}{2}} = \lim_{\epsilon \searrow 0} (\epsilon I + a)^{\frac{1}{2}}$ is also. Furthermore,

$$\int_0^t \sigma(\tau)^2 d\tau = A(t),$$

and so one can integrate $\sigma$ with respect to a Brownian motion.

Now assume that $a$, and therefore $\sigma$, are strictly positive definite. Then $\sigma^{-1}$ is progressively measurable. In addition,

$$\int_0^t (\sigma^{-1} \xi, dA(\tau)\sigma^{-1} \eta)_{\mathbb{R}^N} = t(\xi, \eta)_{\mathbb{R}^N} \text{ for all } \xi \in \mathbb{R}^N,$$

and so, if

$$B(t) = \int_0^t \sigma^{-1}(\tau) dM(\tau),$$

then $\langle B_i, B_j \rangle(t) = t\delta_{i,j}$, which, by Corollary 110, means that $(B(t), \mathcal{F}_t, \mathbb{P})$ is a Brownian motion. Moreover, if

$$I(t) = \int_0^t \sigma(\tau) dB(\tau),$$

then

$$\langle I_i, I_j \rangle(t) = \int_0^t \sigma^2(\tau)_{i,j} d\tau = A(t)_{i,j}$$

and

$$\langle I_i, M_j \rangle(t) = \sum_{k=1}^N \int_0^t \sigma(\tau)_{i,k} \langle B_k, M_j \rangle(d\tau) = \langle M_i, M_j \rangle(t) = A(t)_{i,j},$$

from which it follows that $\langle (I - M)_i, (I - M)_j \rangle = 0$ and therefore that $I(t) = \int_0^t \sigma(\tau) dB(\tau)$.

When $a$ is not strictly positive definite, the preceding argument breaks down. Indeed, if $a = 0$, the sample space can consist of only one point and there is no way that one can build a Brownian motion from $M$, and in general one will only be able to build part of one. For this reason one has to have a Brownian motion in reserve so that one can insert it to fill the gaps caused by $a$ becoming degenerate. With this in mind, denote by $N(t)$ the null space of $a(t)$ and by $\Pi(t)$ orthogonal projection onto $N(t)$, and let $\sigma^{-1}(t)$ be the linear map for which
\[ N(t) \text{ is the null space and } \sigma^{-1}(t) \upharpoonright N(t) \downarrow \text{ is the inverse of } \sigma(t) \upharpoonright N(t) \downarrow. \] Then \( \sigma \sigma^{-1} = \sigma^{-1} \sigma = \Pi \downarrow = \Pi, \text{ and, since} \]
\[ \Pi(t) = \lim_{\varepsilon \downarrow 0} a(t)(cI + a(t))^{-1} \text{ and } \sigma^{-1}(t) = \lim_{\varepsilon \downarrow 0} \sigma(t)(cI + a(t))^{-1}, \]
both these functions are progressively measurable. In particular, for any \( \xi \in \mathbb{R}^N, \)
\[ \int_0^t (\sigma^{-1}(\tau)\xi, dA(\tau)\sigma^{-1}(\xi))_{\mathbb{R}^N} = \int_0^t |\Pi(\tau)\xi|^2 \, d\tau \leq t|\xi|^2, \]
and so stochastic integrals of \( \sigma^{-1} \) with respect to \( M \) are well defined. Now take \( W \) to be Wiener measure on \( C([0, \infty); \mathbb{R}^N), \mathbb{P} = \mathbb{P} \times W, \text{ and } \tilde{F}_t \text{ to be the completion of } \mathcal{F}_t \times \mathcal{W}_t \text{ with respect to } \mathbb{P}, \text{ and define } \tilde{M}(t)(w, \omega) = M(t)(\omega), \tilde{A}(t)(w, \omega) = A(t)(\omega), \tilde{\Pi}(t)(w, \omega) = \Pi(t)(\omega), \tilde{\sigma}(t)(w, \omega) = \sigma(t)(\omega), \text{ and } \tilde{\sigma}^{-1}(w, \omega) = \sigma^{-1}(\omega). \]
It is then an easy matter to check that \( (\tilde{M}(t), \tilde{F}_t, \tilde{\mathbb{P}}) \text{ is a local martingale and that } (w(t), \tilde{F}_t, \tilde{\mathbb{P}}) \text{ is a Brownian motion. In addition, } (w_1, \tilde{M}_j) = 0. \text{ Thus, we can take} \]
\[ \tilde{B}(t) = \int_0^t \tilde{\sigma}^{-1}(\tau) \, d\tilde{M}(\tau) + \int_0^t \tilde{\Pi}(\tau) \, dw(\tau), \]
and, using the properties discussed above, one sees that \( (\tilde{B}_i, \tilde{B}_j)(t) = t\delta_{i,j} \text{ and therefore that } (\tilde{B}(t), \tilde{F}_t, \tilde{\mathbb{P}}) \text{ is a Brownian motion. Further, if } \tilde{I}(t) = \int_0^t \tilde{\sigma}(\tau) \, d\tilde{B}(\tau), \text{ then} \]
\[ \langle \tilde{I}_i, \tilde{I}_j \rangle(t) = \langle \tilde{I}_i, \tilde{M}_j \rangle(t) = \tilde{A}(t)_{i,j}, \]
and so \( \tilde{M}(t) = \int_0^t \tilde{\sigma}(\tau) \, d\tilde{B}(\tau). \)
We have now proved the following representation theorem.

**Theorem 114.** Let \( (M(t), F_t, \mathbb{P}) \text{ be an } \mathbb{R}^N\text{-valued continuous local martingale on } (\Omega, \mathcal{F}), \text{ assume that } \langle M_i \rangle \text{ is absolutely continuous for each } 1 \leq i \leq N, \text{ and define } \sigma \text{ as in the preceding. Then there is a probability space } (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) \text{ on which there is a Brownian motion } (\tilde{B}(t), \tilde{F}_t, \tilde{\mathbb{P}}) \text{ and measurable map } F \text{ to } \Omega \text{ such that } \mathbb{P} = F_*\mathbb{P} \text{ and} \]
\[ \tilde{M}(t) = \tilde{M}(0) + \int_0^t \tilde{\sigma}(\tau) \, d\tilde{B}(\tau) \]
when \( \tilde{M} = M \circ F \text{ and } \tilde{\sigma} = \sigma \circ F. \)

One of the important applications of this result is to the question of uniqueness of solutions to the martingale problem.

**Theorem 115.** Suppose that \( a : \mathbb{R}^N \rightarrow \text{Hom}(\mathbb{R}^N; \mathbb{R}^N) \text{ is a symmetric, non-definite definite matrix valued function and } b : \mathbb{R}^N \rightarrow \mathbb{R}^N. \text{ If} \]
\[ \sup_{\substack{x, y \in \mathbb{R}^N \\forall \ y \neq x}} \frac{|a^{\frac{1}{2}}(y) - a^{\frac{1}{2}}(x)|_{\text{H.S.}} \vee |b(y) - b(x)|}{|y - x|} < \infty, \]
then, for each \( x \in \mathbb{R}^N, \text{ the martingale problem for the } L \text{ in } (12) \text{ has precisely one solution starting at } x. \)
PROOF: Consider the solution \( X(\cdot, x) \) to (56) when \( \sigma = a^{\frac{1}{2}} \) and \( M = N \). We know that the distribution of \( X(\cdot, x) \) under \( \mathcal{W} \) solves the martingale problem for \( L \) starting at \( x \). In addition, if \((\tilde{B}(t), \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\) is any \( \mathbb{R}^N \)-valued Brownian motion and

\[
(*) \quad \tilde{X}(t, x) = x + \int_0^t \sigma(\tilde{X}(\tau, x)) \, d\tilde{B}(\tau) + \int_0^t b(\tilde{X}(\tau, x)) \, d\tau,
\]

then \( \tilde{X}(\cdot, x) \) has the same distribution under \( \tilde{\mathbb{P}} \) as \( X(\cdot, x) \) has under \( \mathcal{W} \). Indeed, \( (*) \) has only one solution and one can construct that solution from \( \tilde{B}(\cdot) \) by the same approximation procedure as we used to construct \( X(\cdot, x) \) from \( w(\cdot) \), and, since all these approximations have the same distributions whether one uses \( \tilde{B}(\cdot) \) and \( \tilde{\mathbb{P}} \) or \( w(\cdot) \) and \( \mathcal{W} \), so must \( \tilde{X}(\cdot, x) \) and \( X(\cdot, x) \). Thus, all the we have to do is show that if \( \mathbb{P} \) solves the martingale problem for \( L \) starting at \( x \), then there is a Brownian motion \((\tilde{B}(t), \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\) such that \( \tilde{\mathbb{P}} \) is the distribution of the solution to \( (*) \). To that end, set \( M(t) = \psi(t) - x - \int_0^t b(\psi(\tau)) \, d\tau \). Then \((M(t), B_t, \mathbb{P})\) is a continuous martingale, and so Theorem 114 says that there is a \((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\), a Brownian motion \((\tilde{B}(t), \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\), and a measure preserving map \( F : \tilde{\Omega} \rightarrow C([0, \infty); \mathbb{R}^N) \) such that

\[
\tilde{X}(t,x) - x - \int_0^t b(\tilde{X}(\tau, x)) \, d\tau = \int_0^t \sigma(\tilde{X}(\tau, x)) \, d\tilde{B}(\tau)
\]

when \( \tilde{X}(\cdot, x) = X(\cdot, x) \circ F. \) □

Stratonovich Integration: When a probabilist looks at an operator

\[
L = \frac{1}{2} \sum_{i,j} a_{i,j}(x) \partial_{x_i} \partial_{x_j} + \sum_{i=1}^N b_i(x) \partial_{x_i},
\]

he is inclined to think of the matrix \( a \) as governing the diffusive behavior and \( b \) as governing the deterministic behavior of the associated diffusion process, and this is entirely reasonable as long as \( a \) and \( b \) are constant. However, if what one means that the diffusive part of the process is the one whose increments during a time interval \( dt \) are of order \( \sqrt{dt} \) as opposed to the deterministic part whose increments are of order \( dt \), then, as the following example shows, this interpretation of \( a \) and \( b \) is flawed. Take \( N = 2 \) and \( a(x) = \begin{pmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_3 & x_1^2 \end{pmatrix} \) and \( b = 0 \). Then the prediction is that the associated diffusion is purely diffusive, but, that is not true. Indeed, \( a = \sigma \sigma^\top \), where \( \sigma(x) = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \), and so an Itô representation of the associated diffusion is

\[
\begin{pmatrix} X_1(t, x) \\ X_2(t, x) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \int_0^t \begin{pmatrix} -X_2(\tau, x) \\ X_1(\tau, x) \end{pmatrix} \, dw(\tau),
\]
where $w$ is an $\mathbb{R}$-valued Wiener path. The easiest way to solve this equation is to write it in terms of complex variables. That is, set $z = x_1 + ix_2$ and $Z(t, z) = X_1(t, x) + iX_2(t, x)$. Then

$$Z(t, z) = z + i \int_0^t Z(\tau, z) d\tau,$$

and so $Z(t, z) = ze^{i w(t) + \frac{i}{2} t}$, or, equivalently,

$$X(t, x) = e^{\frac{t}{2}} \begin{pmatrix} x_1 \cos w(t) - x_2 \sin t \\ x_2 \cos w(t) + x_1 \sin t \end{pmatrix}.$$

In particular, $|X(t, x)| = |x|e^\frac{t}{2}$, which, if $x \neq 0$, means that the distance of $X(t, x)$ from 0 is deterministic and growing exponentially fast.

What the preceding example reflects is that the interpretation of $a$ and $b$ as the diffusive and deterministic parts of $L$ is naïve because it is too coordinate dependent. Indeed, if one represents the preceding $L$ in terms of polar coordinates, one finds that it is equal to

$$\frac{1}{2} (\partial_\theta^2 + r \partial_r),$$

which makes it clear that, although the angular coordinate of $X(t, x)$ is a Brownian motion, the radial coordinate is deterministic. One reason why this flaw was, for the most part, ignored by probabilists is that it doesn’t cause any problems when $a$ is uniformly elliptic in the sense that $a \geq \epsilon I$, in which case, at least over short time intervals, the associated diffusion is diffusive for every choice of $b$. The Cameron-Martin formula (82) provides a good explanation for this. Namely, when $a \geq \epsilon I$, any $b$ can be represented as $a^{\frac{1}{2}} \beta$, where $\beta = a^{-\frac{1}{2}} b$. Hence, in this case, the distributions on bounded time intervals of the diffusions associated with different $b$’s will be mutually absolutely continuous, and therefore their almost certain behavior over bounded time intervals will look the same in all coordinate systems.

To address the issues raised above, it is desirable to find a way of representing $L$ in a form that looks the same in all coordinate systems, and such a representation was introduced by L. Hörmander. Write $a = \sigma \sigma^T$, where $\sigma \in \mathbb{R}^M, \mathbb{R}^N$, and, for each $1 \leq k \leq M$, let $V_k \in \mathbb{R}^N$ be the $k$th column of $\sigma$ and $\mathcal{L}_{V_k}$ for associated directional derivative operator $\sum_{i=1}^N \sigma_{i,k} \partial_{x_i}$. Assuming that $\sigma$ is continuously differentiable, one then has

\begin{equation}
L = \frac{1}{2} \sum_{j=1}^M \mathcal{L}_{V_j}^2 + \mathcal{L}_{V_0},
\end{equation}

when one takes

$$V_0 = b - \frac{1}{2} \sum_{k=1}^M \mathcal{L}_{V_k} V_k.$$
Thus, in the preceding example, $M = 1$, $V_1 = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$ and $V_0 = \frac{1}{2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. The beauty of the representation in (116) is that it looks the same in all coordinate systems. That is, suppose that $F$ is a diffeomorphism on some open set $G \subseteq \mathbb{R}^N$, and, given a vector field $V$ on $G$, define the vector field $F_*V$ on $F(G)$ so that $\mathcal{L}_{F_*V}\varphi = (\mathcal{L}_V(\varphi \circ F)) \circ F^{-1}$ for $\varphi \in C^1(F(G); \mathbb{R})$. More explicitly,

$$\mathcal{L}_{F_*V}\varphi = \mathcal{L}_{JF,V}\varphi$$

where $J_F = \begin{pmatrix} \partial_{x_j}F_i \end{pmatrix}_{1 \leq i,j \leq N}$.

Then

$$\mathcal{L}_V^2(\varphi \circ F) = \mathcal{L}_V\left(\mathcal{L}_{F_*V}\varphi \circ F\right) = \left(\mathcal{L}_{F_*V,\varphi}^2\right) \circ F,$$

and in the coordinate system on $F(G)$ determined by $F$,

$$L = \frac{1}{2} \sum_{k=1}^M \mathcal{L}_{F_*V_k}^2 + \mathcal{L}_{F_*V_0}.$$

Related to these considerations is the following. In that the Hörmander representation is in terms of vector fields, one suspects that the paths of the diffusion associated with the $L$ in (116) should transform under changes of coordinates the same way of integral curves do. Namely, if $X(\cdot)$ is an integral curve of the vector field $V$ and $F$ is a diffeomorphism, then $F \circ X(\cdot)$ is an integral curve of $F_*V$. Hence, we should expect that if $X(\cdot, x)$ is the solution to (56) when $\sigma = (V_1, \ldots, V_M)$ and $b = V_0 + \frac{1}{2} \sum_{k=0}^M \mathcal{L}_{V_k}V_k$, then $F \circ X(\cdot, x)$ should be solution to (56) with $\sigma$ and $b$ replaced by

$$(F_*V_1, \ldots, F_*V_M) \text{ and } F_*V_0 + \frac{1}{2} \sum_{k=0}^M \mathcal{L}_{F_*V_k}F_*V_k,$$

and, with sufficient patience, one can check that this is true. However, it would be desirable to have a formulism that made such computations easier and brought out the relationship between $X(\cdot, x)$ and integral curves of the $V_k$’s. For that purpose, reconsider the equation

$$\dot{X}(t, x)(w) = \sum_{k=1}^M V_k(X(t, x)(w)) \dot{w}(t)_k + V_0(X(t, x)(w)) \text{ with } X(0, x) = x,$$

where the interpretation now, unlike Itô’s interpretation, is that $X(t, x)(w)$ is constructed by taking the limit of integral curves corresponding to mollifications of the paths $w$. For instance, in our example,

$$F_1(t, x) = \begin{pmatrix} x_1 \cos t - x_2 \sin t \\ x_2 \cos t + x_1 \sin t \end{pmatrix} \text{ and } F_0(t, x) = e^{\frac{t}{2}}x.$$
are the integral curves of \( V_1 \) and \( V_0 \) starting at \( x \), and it is easy to check that, for any continuously differentiable \( w \), \( F_0(t, F_1(w(t), x)) \) is the integral curve of \( \dot{w}(t)\mathcal{L}_{V_1} + \mathcal{L}_{V_0} \) starting at \( x \). Thus, one might guess that \( F_0(t, F_1(w(t), x)) \) a diffusion associated with \( L \), as indeed it is. As we will see below, this example is as simple as it is because \( \mathcal{L}_{V_0} \) commutes with \( \mathcal{L}_{V_1} \), and things are more complicated when dealing with non-commuting vector fields.

With the preceding in mind, one should wonder whether there is a way to incorporate these ideas into a theory of stochastic integration. Such a theory was introduced by the Russian engineer L. Stratonovich and produces what is now called the **Stratonovich integral**. However, Statonovich’s treatment was rather cumbersome, and mathematicians remained skeptical about it until Itô rationalized it. Itô understood that there was no way to define such an integral for all locally square integrable progressively measurable integrands, but he realized that one could do so if the integrand was a semi-martingale. Namely, given a pair of continuous semi-martingales \( (X(t), \mathcal{F}_t, \mathbb{P}) \) and \( (Y(t), \mathcal{F}_t, \mathbb{P}) \), let paths such that \( X_n(m2^{-n}) = X(m2^{-n}) \), \( Y_n(m2^{-n}) = Y(m2^{-n}) \), and \( X_n(\cdot) \) and \( Y_n(\cdot) \) are linear on \([m2^{-n}, (m + 1)2^{-n}]\) for each \( m \geq 0 \). Set \( \Delta_{m,n}^X = X((m + 1)2^{-n}) - X(m2^{-n}) \) and \( \Delta_{m,n}^Y(t) = Y((m + 1)2^{-n}) - Y(m2^{-n}) \). Then, by Corollary 113,

\[
\int_0^t X_n(\tau) \, dY_n(\tau) = \sum_{m < 2^n} X(m2^{-n}) \Delta_{m,n}^Y + \frac{1}{2} \sum_{m < 2^n} 4^n (t \wedge (m + 1)2^{-n} - m2^{-n}) \Delta_{m,n}^X \Delta_{m,n}^Y \\
\rightarrow \int_0^t X(\tau) \, dY(\tau) + \frac{1}{2} (X, Y)(t).
\]

Thus Itô said that the Stratonovich integral of \( Y(\cdot) \) with respect to \( X(\cdot) \) should be

\[
\int_0^t Y(\tau) \, \bullet \, dX(\tau) \equiv \int_0^t Y(\tau) \, dX(\tau) + \frac{1}{2} (X, Y)(t).^{16}
\]

Notice, if one adopts this definition, then the equation in Corollary 112 becomes

\[
(117) \quad \varphi(\bar{X}(t)) = \varphi(\bar{X}(0)) + \sum_{j=1}^N \int_0^t \partial_{x_j} \varphi(\bar{X}(\tau)) \, \bullet \, dX_j(\tau)
\]

for \( \varphi \in \mathcal{C}^3(\mathbb{R}^N; \mathbb{R}) \). Indeed, by Itô’s formula, the martingale part of \( \partial_{x_j} \varphi(\bar{X}(t)) \) is

\[
\sum_{j=1}^N \int_0^t \partial_{x_i} \partial_{x_j} \varphi(\bar{X}(\tau)) \, dM_i(\tau),
\]

\(^{16}\)The standard notation is \( \circ dX(\tau) \) rather than \( \bullet dX(\tau) \), but, because I use \( \circ \) to denote composition, I have chosen to use \( \bullet \) to denote Statonovich integration.
where $M_i(t)$ is the martingale part of $X_i(t)$, and so
\[
\sum_{j=1}^{N} \langle \partial_x \varphi \circ \bar{X}, X_i \rangle (dt) = \sum_{i,j=1}^{N} \partial_x \partial_x \varphi (\bar{X}(t)) \langle X_i, X_j \rangle (dt).
\]

**Remark:** It is a little disappointing that we mollified both $X$ and $Y$ and not just $Y$. Thus, it may be comforting to know that one need only mollify $Y$ if $\langle X, Y \rangle (dt) = c(t) dt$ for a continuous progressively measurable function $c$. To see this, consider
\[
D_n = \int_0^1 (X(\tau) - X_n(\tau)) \, dY_n(\tau)
\]
\[
= \sum_{m<2^n} 2^n \left( \int_{I_{m,n}} \left( (X(\tau) - X_{m,n}) - 2^n (\tau - m2^{-n}) \Delta_{m,n}^X \right) \, d\tau \right) \Delta_{m,n}^Y,
\]
where $I_{m,n} \equiv [m2^{-n}, (m+1)2^{-n}]$ and, for any $F : [0, \infty) \rightarrow \mathbb{R}$, $\Delta_{m,n}^F \equiv F_{m+1,n} - F_{m,n}$ with $F_{m,n} \equiv F(m2^{-n})$. Obviously
\[
4^n \int_{I_{m,n}} (\tau - m2^{-n}) \Delta_{m,n}^X \, d\tau = \frac{1}{2} \Delta_{m,n}^X,
\]
and, using integration by parts,
\[
\int_{I_{m,n}} 2^n \int_{I_{m,n}} (X(\tau) - X_{m,n}) \, d\tau = \Delta_{m,n}^X - 2^n \int_{I_{m,n}} (\tau - m2^{-n}) \, dX(\tau).
\]
Hence, if
\[
Z_n(t) = \int_0^t \left( \frac{1}{2} - 2^n (\tau - \lfloor \tau \rfloor_n) \right) \, dX(\tau),
\]
then (cf. the proof of Corollary 113),
\[
D_n = \sum_{m<2^n} \Delta_{m,n}^Z \Delta_{m,n}^Y
\]
\[
= \langle Z_n, Y \rangle (1) + \int_0^1 (Z_n(\tau) - Z_n(\lfloor \tau \rfloor_n)) \, dY(\tau) + \int_0^1 (Y(\tau) - Y(\lfloor \tau \rfloor_n)) \, dZ_n(\tau).
\]
It is simple to show that the stochastic integrals on the right tend to 0 as $n \rightarrow \infty$. To handle the term $\langle Z_n, Y \rangle (1)$, observe that it equals
\[
\frac{1}{2} \int_0^1 c(\tau) \, d\tau - 2^n \int_0^1 (\tau - \lfloor \tau \rfloor_n) c(\tau) \, d\tau,
\]
and, since
\[
2^n \int_0^1 (\tau - \lfloor \tau \rfloor_n) c(\tau) \, d\tau
\]
\[
= \frac{1}{2} 2^{-2} \sum_{m<2^n} c(m2^{-n}) + 2^n \int_0^1 (\tau - \lfloor \tau \rfloor_n)(c(\tau) - c(\lfloor \tau \rfloor_n)) \, d\tau \rightarrow \frac{1}{2} \int_0^1 c(\tau) \, d\tau,
\]
it follows that $D_n \rightarrow 0$. 

Because the equation in (117) resembles the fundamental theorem of calculus, at first sight one might think that it is an improvement on the formula in Corollary 112. However, after a second look, one realizes that the opposite is true. Namely, in Itô’s formulation, the integrand in a Stratonovich integral has to be a semi-martingale, and so we had to assume that \( \varphi \in C^3(\mathbb{R}^N; \mathbb{R}) \) in order to be sure that \( \partial_x \varphi(\tilde{X}(t)) \) would be one. Thus, we have obtained a pleasing looking formula in which only first order derivatives of \( \varphi \) appear explicitly, but we have done so at price of requiring that \( \varphi \) have three continuous derivatives.

It turns out that there is an ingenious subterfuge that allows one to avoid this objectionable requirement in some cases, but the basic fact remains that the fundamental theorem of calculus and martingales are incompatible companions.

**Stratonovich Integral Equations:** In spite of the somewhat disparaging comments at the end of the previous section, Stratonovich integration has value. However, to appreciate its value, one has to abandon Itô’s clever formulation of it and return to ideas based on its relationship to integral curves. To explain what I have in mind, for each \( 0 \leq k \leq M \), let \( V_k \) be an element of \( C^2(\mathbb{R}^N; \mathbb{R}^N) \) with bounded first and second order derivatives, and consider the stochastic integral equation

\[
X(t, x) = x + \int_0^t V_0(X(\tau, x)) \, d\tau + \sum_{k=1}^M V_k(X(\tau, x)) \cdot dw(\tau)_k.
\]

Of course, this equation is equivalent to (56) when \( \sigma \) is the \( \mathbb{R}^N \times \mathbb{R}^M \)-matrix whose \( k \)th column, for \( 1 \leq k \leq M \) is \( V_k \) and \( b = V_0 + \frac{1}{2} \sum_{k=1}^M \mathcal{L}_{V_k} V_k \). Thus we know that (118) has a solution and the distribution of that solution is the unique solution to the martingale problem for the Hörmander form operator (116). Notice that the using this formalism, it is easy to verify that, if \( F \) is a diffeomorphism, then \( F \circ X(\cdot, x) \) is the solution to (118) with the \( V_k \)'s replaced by \( F_*V_k \)'s. Indeed, by (117),

\[
F(X(t, x)) = F(x) + \int_0^t \mathcal{L}_{V_0} F(X(\tau, x)) \, d\tau + \int_0^t \mathcal{L}_{V_k} F(X(\tau, x)) \cdot dw(\tau)_k,
\]

and \( \mathcal{L}_{V_k} F = (F_*V_k) \circ F \).

Although we know how to solve (118) by converting it to an Itô stochastic integral equation, that is not the way that I want to think about solving it. Instead, for \( \xi = (\xi_0, \ldots, \xi_M) \in \mathbb{R} \times \mathbb{R}^M \), set \( V_\xi = \sum_{k=0}^M \xi_k V_k \), and consider the the ordinary differential equation

\[
\dot{F}_\xi(t, x) = V_\xi(F_\xi(t, x)) \quad \text{with} \quad F_\xi(0, x) = x.
\]

In other words, \( t \mapsto F_\xi(t, x) \) is the integral curve of \( V_\xi \) that passes through \( x \) at time 0. Next define \( E : (\mathbb{R} \times \mathbb{R}^M) \times \mathbb{R}^N \rightarrow \mathbb{R}^N \) by \( E(\xi, x) = F_\xi(1, x) \). As is
easy to check, \( E(t\xi, x) = F_t(t, x) \) for all \( t \in \mathbb{R} \). Furthermore, standard results from the theory of ordinary differential equations show \( E(\xi, \cdot) \) of diffeomorphism from \( \mathbb{R}^N \) onto itself, \( E(-\xi, \cdot) \) is the inverse of \( E(\xi, \cdot) \), and there exist constants \( C < \infty \) and \( \nu \in [0, \infty) \) such that

\[
|\partial_{\xi_1} E(\xi, x)| \vee |\partial_{x_j} E(\xi, x)| \vee |\partial_{\xi_1} \partial_{\xi_l} E(\xi, x)|
\]

\[
|\partial_{\xi_k} \partial_{x_j} E(\xi, x)| \vee |\partial_{x_j} \partial_{x_j} E(\xi, x)| \leq C(1 + |x|)e^{\nu(|\xi| + |x|)}
\]

for all \( 1 \leq i, j \leq M \) and \( 0 \leq k, \ell \leq M \).

For \( V, W \in C^1(\mathbb{R}^N; \mathbb{R}^N) \), define the commutator \([V, W] = \mathcal{L}_V W - \mathcal{L}_W V\).

Equivalently, \( \mathcal{L}_{[V, W]} = \mathcal{L}_V \circ \mathcal{L}_W - \mathcal{L}_W \circ \mathcal{L}_V \).

**Lemma 120.** Assume that \( V, W \in C^1(\mathbb{R}^N; \mathbb{R}^N) \) have bounded first order derivatives, and determine \( t \mapsto F(t, x) \) and \( t \mapsto G(t, x) \) by

\[
\dot{F}(t, x) = V(F(t, x)) \quad \text{and} \quad \dot{G}(t, x) = V(G(t, x)) \quad \text{with} \quad F(0, x) = G(0, x) = x.
\]

If \( [V, W] = 0 \), then \( F(s, G(t, x)) = G(t, F(x, s)) \) for all \( (s, t) \in \mathbb{R}^2 \).

**Proof:** Set \( \Phi(t) = F(s, G(t, x)) \). Then, for \( \varphi \in C^2(\mathbb{R}^N; \mathbb{R}) \),

\[
\partial_t \varphi \circ \Phi(t) = (\mathcal{L}_{(F(t, x), W \varphi)})(G(t, x))
\]

If \( \varphi_s = \varphi \circ F(s, \cdot) \), then, because \( F(s + h, \cdot) = F(h, F(s, \cdot)) \),

\[
\partial_s \left( (\mathcal{L}_{F(s, \cdot), W \varphi})(F(-s, y)) = \partial_h \left( (\mathcal{L}_{F(h, \cdot), W \varphi_s})(F(-s - h, y)) \right) \right)_{h=0}
\]

\[
= (\mathcal{L}_W \circ \mathcal{L}_V - \mathcal{L}_V \circ \mathcal{L}_W) \varphi_s(y) = 0.
\]

Hence \( (\mathcal{L}_{F(s, \cdot), W \varphi})(F(-s, y)) = \mathcal{L}_W \varphi(y) \), which means that \( (\mathcal{L}_{F(s, \cdot), W \varphi})(y) = \mathcal{L}_W \varphi(F(s, y)) \) and therefore that \( \partial_t \varphi \circ \Phi(t) = \mathcal{L}_W \varphi(\Phi(t)) \). Equivalently, \( \Phi(t) = W(\Phi(t)) \), and so, since \( \Phi(0) = F(s, x) \),

\[
F(s, G(t, x)) = \Phi(t) = G(t, F(s, x)) \quad \square
\]

**Theorem 121.** Let \( X(\cdot, x) \) be the solution to (118). If the \( V_k \)'s commute, then \( X(t, x)(w) = E((t, w(t)), x) \).

**Proof:** By (117),

\[
E((t, w(t)), x) = x + \int_0^t \partial_{u_1} E((\tau, w(\tau), x) \, d\tau + \sum_{k=1}^N \partial_{u_k} E((\tau, w(\tau)), x) \cdot dw(\tau)_k,
\]
and so all that remains is to prove that $\partial_\xi E(\xi, y) = V_k(E(\xi, y))$. To this end, apply Lemma 120 to see that $E(\xi, E(\eta, y)) = E(\eta, E(\xi, y))$, and then apply this to check that if $\Phi(t) = E(t\xi, E(t\eta, y))$, then $\dot{\Phi}(t) = V_{\xi + \eta}(\Phi(t))$ with $\Phi(0) = x$. Hence $E(\xi + \eta, y) = E(\eta, E(\xi, y))$. Taking $\eta = t\epsilon_k$, where $(\epsilon_k)_t = \delta_{k, t}$, and then differentiating with respect to $t$ at $t = 0$, one arrives at the desired conclusion. □

Besides confirming the connection between integral curves and Stratonovich integral equations, this theorem shows that, when the $V_k$'s commute, the solution to (118) is a continuous function of $w$. In fact, it has as many continuous Frechet derivates as a function of $w$ as the $V_k$'s have as a functions of $x$. In addition, it indicates that, although we constructed solutions to (56) by perturbing off the constant coefficient case, we should construct solutions to (118) by perturbing off the commuting case. The critical property that was present when the $V_k$'s commute and is absent when they don’t is $\partial_\xi E(\xi, y) = V_k(E(\xi, y))$, which is equivalent to the property $E(\xi + \eta, y) = E(\eta, E(\xi, y))$. A quantitative statement of that equivalence is content of the following lemma.

**Lemma 122.** Set

$$R(\xi, \eta, x) = E(\eta, E(\xi, x)) - E(\xi + \eta, x) - \frac{1}{2}[V_\xi, V_\eta](x).$$

Then $R(\xi, \eta, x)$ is continuously differentiable, $R(\xi, 0, x) = 0$, and there exists a $C < \infty$ such that (cf. (119)) $\|\partial_\eta R(\xi, 0, x)\| \leq C\|\xi\|^2 e^{\nu|\xi|}(1 + |x|)$ for all $0 \leq k \leq M$.

**Proof:** By Taylor’s theorem,

$$E(\eta, E(\xi, x)) = E(\xi, x) + V_\eta(E(\xi, x)) + \frac{1}{2}L_{V_\xi}V_\eta(E(\xi, x)) + R_0(\xi, \eta, x)$$

$$= x + V_\xi(x) + \frac{1}{2}L_{V_\xi}V_\eta(x) + L_{V_\xi}V_\eta(x) + \frac{1}{2}L_{V_\xi}V_\eta(x)$$

$$+ R_0(\xi, \eta, x) + R_1(\xi, x) + R_2(\xi, \eta, x) + R_3(\xi, \eta, x),$$

where

$$R_0(\xi, \eta, x) = \int_0^1 (1 - t)\left(L_{V_\xi}V_\eta(E(t\eta, E(\xi, x))) - L_{V_\xi}V_\eta(E(\xi, x))\right) dt$$

$$R_1(\xi, x) = \int_0^1 (1 - t)\left(L_{V_\xi}V_\eta(E(t\xi, x)) - L_{V_\xi}V_\eta(x)\right) dt$$

$$R_2(\xi, \eta, x) = \int_0^1 \left(L_{V_\xi}V_\eta(E(t\xi, x)) - L_{V_\xi}V_\eta(x)\right) dt$$

$$R_3(x, \eta, x) = \frac{L_{V_\xi}V_\eta(E(\xi, x)) - L_{V_\xi}V_\eta(x)}{2}.$$ 

At the same time,

$$E(\xi + \eta, x) = x + V_{\xi + \eta}(x) + \frac{1}{2}L_{V_{\xi + \eta}}V_{\xi + \eta}(x) + R_1(\xi + \eta, x).$$
Hence, \( E(\eta, E(\xi, x)) - E(\xi + \eta, x) = \frac{1}{2}[V_{\xi}, V_\eta](x) + R(\xi, \eta, x) \), where

\[ R(\xi, \eta, x) = R_0(\xi, \eta, x) + R_1(\xi, x) + R_2(\xi, \eta, x) + R_3(\xi, \eta, x) - R_1(\xi + \eta, x). \]

Finally,

\[
\partial_n R(\xi, 0, x) = \int_0^1 \left( \mathcal{L}_{V_k} V_k \left( E(\xi, x) \right) - \mathcal{L}_{V_k} V_k(x) \right) d\tau \\
+ \int_0^1 (1 - t) \left( (\mathcal{L}_{V_k} V_k + \mathcal{L}_{V_k} V_\xi)(E(t\xi, x)) - (\mathcal{L}_{V_k} V_k + \mathcal{L}_{V_k} V_\xi)(x) + t \mathcal{L}_{V_k} V_\xi \partial_n E(\xi, x) \right) dt,
\]

from which the asserted estimate follows when one uses (119). □

Set \( E_k(\xi, x) = \partial_k E(\xi, \eta) \), \( W_k(\xi, x) = E_k(\xi, x) - V_\xi E(\xi, x) \), and \( \tilde{W}_k(\xi, x) = W_k(\xi, x) - \frac{1}{2}[V_{\xi}, V_\eta](x) \). Since

\[
\left| W_k(\xi, x) \right| \vee |\partial_\xi \tilde{W}_k(\xi, x)| \leq C|\xi|e^{\nu|\xi|}(1 + |\xi|)
\]

for some \( C < \infty \). For \( u \geq 0 \), determine \( X_n(t, \cdot, x) \) by \( X_n(0, x) = x \) and

\[
X_n(t, x) = E(\Delta_n(t), X_n([t]_n, x))
\]

where \( \Delta_n(t)(w) = (t - [t]_n, w(t) - w([t]_n)) \).

Equivalently,

\[
X_n(t, x) = x + \int_0^t E_0(\Delta_n(\tau), X_n([\tau]_n, x)) \, d\tau \\
+ \sum_{k=1}^M E_k(\Delta_n(\tau), X_n([\tau]_n, x)) \cdot dw(\tau)_k
\]

\[
= x + \int_0^t V_0(X_n(\tau, x)) \, d\tau + \mathcal{E}_n(t, x),
\]

where

\[
\mathcal{E}_n(t, x) = \sum_{k=1}^N \int_0^t V_k(X_n(\tau, x)) \cdot dw(\tau)_k
\]

\[
+ \int_0^t W_0(\Delta_n(\tau), X_n([\tau]_n, x)) \, d\tau \\
+ \sum_{k=1}^N \int_0^t W_k(\Delta_n(\tau), X_n([\tau]_n, x)) \cdot dw(\tau)_k.
\]
Observe that, by (123),

\[
\mathbb{E}^W \left[ \sup_{t \in [0,T]} \left| \int_0^t W_0(X_n(\tau, x)) \, d\tau \right|^2 \right] \\
\leq T \int_0^T \mathbb{E}^W \left[ |W_0(\Delta_n(\tau), X_n([\tau]_n, x))|^2 \right] \, d\tau \\
\leq 2C^2T \sum_{m < 2^nT \wedge (m+1)2^{-n} < m2^{-n}} \mathbb{E}^W \left[ |\Delta_n(\tau)|^2 e^{2\nu|\Delta_n(t)|} (1 + |X_n(m2^{-n}, x)|^2) \right] \, d\tau \\
= 2C^2T \sum_{m < 2^nT \wedge (m+1)2^{-n} < m2^{-n}} \mathbb{E}^W \left[ |\Delta_n(\tau)|^2 e^{2\nu|\Delta_n(t)|} \right] \\
\times \mathbb{E}^W \left[ (1 + |X_n(m2^{-n}, x)|^2) \right] \, d\tau,
\]

and there is an \( A < \infty \) such that

\[
\mathbb{E}^W \left[ |\Delta_n(\tau)|^2 e^{2\nu|\Delta_n(t)|} \right] \leq A2^{-n}.
\]

Hence

\[
\mathbb{E}^W \left[ \sup_{t \in [0,T]} \left| \int_0^t W_0(X_n(\tau, x)) \, d\tau \right|^2 \right] \\
\leq 2^{-n+1}C^2AT \int_0^T \mathbb{E}^W \left[ 1 + |X_n(\cdot, x)|^2 \right] \, d\tau.
\]

Estimating the terms involving \( W_k \) for \( 1 \leq k \leq M \) is more challenging because the integrals are Stratonovich and not Itô ones and we don’t have a good estimate on \( \partial_{\xi_k} W_k \). What enables us to handle them is that, since \([V_k, V_k] = 0\), we can write

\[
\int_0^T W_k(\Delta_n(t), X_n([\tau]_n, x)) \, d(\tau)_k \\
= \frac{1}{2} \sum_{0 \leq \ell \leq M, \ell \neq k} \int_0^T \Delta_n(\tau)[V_{\ell}, V_k](X_n([\tau]_n, x)) \, dw(\tau)_k \\
+ \int_0^T \tilde{W}_k(\Delta_n(\tau), X_n([\tau]_n, x)) \, d(\tau)_k,
\]

where the replacement of \( \bullet dw(\tau)_k \) by \( dw(\tau)_k \) in the second line is justified because \( \langle (\Delta_n)_{\ell}, w_k \rangle = 0 \) when \( \ell \neq k \). After applying Doob’s inequality and repeat-
ing the same sort of reasoning as we just used, it is easy to show that

\[
E^W \left[ \sup_{t \in [0,T]} \left| \int_0^T \Delta_n(\tau)[V_\ell, V_k](X_n([\tau]_n, x)) \, dw(\tau)_k \right|^2 \right] 
\leq C 2^{-n} \int_0^T E^W \left[ (1 + \|X_n(\cdot, x)\|_{[0,\tau]}^2) \right]
\]

for some \( C < \infty \). As for the Stratonovich integrals in which the \( \tilde{W}_k \)'s appear, one writes them as

\[
\int_0^t \tilde{W}_k(\Delta_n(\tau), X_n([\tau]_n, x)) \, dw(\tau)_k + \frac{1}{2} \int_0^t \partial_{\xi_k} \tilde{W}_k(\Delta_n(\tau), X_n([\tau]_n, x)) \, d\tau
\]

and then uses the estimates in (123) to obtain

\[
E^W \left[ \sup_{t \in [0,T]} \left| \int_0^t \tilde{W}_k(\Delta_n(\tau), X_n([\tau]_n, x)) \bullet dw(\tau)_k \right| \right] 
\leq C 2^{-n} E^W \left[ 1 + \|X_n(\cdot, x)\|_{[0,\tau]}^2 \right]
\]

for some \( C < \infty \). By combining these, we have that

(124) \[
E^W[\|E_n(\cdot, x)\|_{[0,t]}^2] \leq C 2^{-n} E^W[1 + \|X_n(\cdot, x)\|_{[0,t]}^2].
\]

for some \( C < \infty \). The integrals with \( V_k \) in their integrands are handled in the same way as the ones with integrands \( W_0 \) and \( \tilde{W}_k \), and, in conjunction with the estimate in (124) lead to

\[
E^W[1 + \|X_n(\cdot, x)\|_{[0,t]}^2] \leq 1 + 2|x|^2 + C \int_0^t E^W[\|X_n(\cdot, x)\|_{[0,\tau]}^2] \, d\tau
\]

for some \( C < \infty \). Hence, by Gromwall’s lemma, we know that

(125) \[
\sup_{n \geq 0} E^W[1 + \|X_n(\cdot, x)\|_{[0,t]}^2] \leq K e^{Kt}(1 + |x|^2)
\]

for some \( K < \infty \).

Now let \( X(t, x) \) be the solution to (118) and set \( D_n(t, x) = X_n(t, x) - X(t, x) \).

Then

\[
D_n(t, x) = \int_0^t \left( V_0(X_n(\tau, x)) - V_0(X(\tau, x)) \right) \, d\tau \\
+ \sum_{k=1}^N \int_0^t \left( V_k(X_n(\tau, x)) - V_k(X(\tau, x)) \right) \bullet dw(\tau)_k + E_n(t, x),
\]
which, together with (124) and (125), leads first to
\[
\mathbb{E}[1 + \|D_n(\cdot, x)\|^2_{[0,t]}]
\leq 2^{-n} e^{Kt} KC(1 + |x|^2) + K' \int_0^t \mathbb{E}[1 + \|D_n(\cdot, x)\|^2_{[0,\tau]}] \, d\tau,
\]
and then, by Gromwall’s lemma, to
\[
\mathbb{E}[\|D_n(\cdot, x)\|^2_{[0,t]}] \leq 2^{-n} e^{(K+K')t} KC(1 + |x|^2).
\]

Therefore, we have now shown that, for each \( t > 0 \),

\[
X_n(\cdot, x) - X(\cdot, x) \to 0 \quad \text{(a.s.,} \mathcal{W}) \quad \text{and in} \quad L^2(\mathcal{W}; \mathbb{R}).
\]

As a consequence of the preceding, we have now proved the following version of a result whose origins are in the work of Wong and Zakai.

**Theorem 127.** Define \( S(x) \) to be the closure in \( C([0,\infty); \mathbb{R}^N) \) of the set of continuously differentiable paths \( p : [0,\infty) \to \mathbb{R}^N \) for which \( p(0) = x \) and there exists a smooth function \( \alpha \in C([0,\infty); \mathbb{R}^N) \) such that

\[
\dot{p}(t) - V_0(p(t)) = \sum_{k=1}^M \alpha(t)_k V_k(p(t)).
\]

If \( \mathbb{P}_x \) is the solution to the martingale problem for the \( L \) in (116) starting at \( x \), then \( \mathbb{P}_x(S(x)) = 1 \). In particular, if \( M \subseteq \mathbb{R}^N \) is a closed, differentiable manifold and \( V_k(y) \) is tangent to \( M \) for all \( 0 \leq k \leq M \) and \( y \in M \), then

\[
\mathbb{P}_x(\psi(t) \in M \text{ for all } t \in [0,\infty)) = 1.
\]

**Proof:** Since, if \( x \in M \), every \( p \in S(x) \) stays on \( M \), the last assertion follows from the first. Next observe that for any piecewise constant \( \alpha : [0,\infty) \to \mathbb{R}^N \), the solution to (128) is the limit in \( C([0,\infty); \Theta(\mathbb{R}^N)) \) of solutions to (128) for smooth \( \alpha \)'s. Thus, \( S(x) \) contains all limits in \( C([0,\infty); \mathbb{R}^N) \) of solutions to (128) with piecewise constant \( \alpha \)'s.

For the present assume that the \( V_k \)'s are bounded by some constant \( C < \infty \). Given \( w \in C([0,\infty); \mathbb{R}^N) \), let \( w_n \) be the polygonal path that is equal \( w \) at time \( m2^{-n} \) and is linear on each interval \([m2^{-n}, (m+1)2^{-n}]\). If \( \tilde{X}_n(\cdot, x)(w) \) is the solution to (128) with \( \alpha = w_n \), then

\[
\tilde{X}_n(t, x) = E(t\Delta_n((m+1)2^{-n}), \tilde{X}_n(m2^{-n}, x)) \quad \text{for} \ t \in [m2^{-n}, (m+1)2^{-n}],
\]

where \( \Delta_n(z) = \mathbb{E}[\Delta_n((m+1)2^{-n}) | \mathbb{F}_n(z)] \).

and so $\tilde{X}_n(m2^{-n}, x) = X_n(m2^{-n}, x)$ for all $m \geq 0$. At the same time,

$$|X_n(t, x)(w) - X_n(m2^{-n}, x)(w)| \leq C|w(t) - w(m2^{-n})|$$

$$|\tilde{X}_n(t, x) - \tilde{X}_n(t, x)(w)| \leq C|w((m + 1)2^{-n}) - w(m2^{-n})|,$$

and so $\|X_n(\cdot, x) - \tilde{X}_n(\cdot, x)\|_{[0, t]} \to 0$ for each $t \geq 0$. Hence, by (126), $P_x(S(x)) = W(X(\cdot, x) \in S(x)) = 1$.

To remove the boundedness assumption on the $V_k$'s, for each $R > 0$, choose $\eta_R \in C^\infty_c([R, R+1])$ so that $\eta_R = 1$ on $B(0, R + 1)$, set $V_{k}^{(R)}(y) = \eta_R(y) V_k(y)$, and define $S^{(R)}(x), X^{(R)}(\cdot, x), X_{n}^{(R)}(\cdot, x)$ accordingly. If

$$\zeta_R = \inf\{t \geq 0 : |X(t, x)| \geq R\},$$

then, by Lemma 38, $X(\cdot, x) = X^{(R)}(\cdot, x)$ on $[0, \zeta_R)$. Hence, if $t < \zeta_R$ and $\|X^{(R)}(\cdot, x) - X_{n}^{(R)}(\cdot, x)\|_{[0, t]} < \epsilon < 1$, then $\|X(\cdot, x) - X_n(\cdot, x)\|_{[0, t]} < \epsilon$. Therefore, if $\psi \in S(t, x)$, then

$$\mathbb{P}_x(\psi \in S(t, x)) \geq \mathcal{W}(X^{(R)}(\cdot, x) \in S^{(R)}(x) \& \zeta_R > t) \geq 1 - \mathcal{W}(\zeta_R \leq t)$$

for all $R > 0$. After letting $R \to \infty$ and then $t \to \infty$, one has that $\mathbb{P}_x(S(x)) = 1$. □