Some Background Material for 18.176

Daniel W. Stroock

Notation

$\langle \varphi, \mu \rangle = \int \varphi \, d\mu$

$\mathbf{M}_1(\mathbb{R}^N)$ is the space of Borel probability measures on $\mathbb{R}^N$ endowed with the topology of weak convergence. Given $\mu \in \mathbf{M}_1(\mathbb{R}^N)$, its Fourier transform is $\hat{\mu}(\boldsymbol{\xi}) \equiv \int_{\mathbb{R}^N} e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}} \, \mu(d\mathbf{y})$.

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

$\| \cdot \|_{\mathrm{u}}$ is the uniform norm.

$C^m(\mathbb{R}^N; \mathbb{R})$ is the space of m-times, continuously differentiable funtions, $C^m_{\mathrm{b}}(\mathbb{R}^N; \mathbb{R})$ is the subspace of $C^m(\mathbb{R}^N; \mathbb{R})$ whose elements have bounded derivatives of all orders less than of equal to m, $C^m_{\mathrm{c}}(\mathbb{R}^N; \mathbb{R})$ is the subspace of $C^m_{\mathrm{b}}(\mathbb{R}^N; \mathbb{R})$ whose elements have compact support, $C^{1,2}_{\mathrm{b}}\big([0,\infty) \times \mathbb{R}^N; \mathbb{R}\big)$ is the space functions $(t, \mathbf{x}) \in [0,\infty) \times \mathbb{R}^N \longmapsto u(t, \mathbf{x}) \in \mathbb{R}$ which have one bounded, continuous derivative with respect to t and two bounded, continuous derivatives with respect to $\mathbf{x}$.

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

$\|A\|_{\mathrm{op}} = \sup\{|A\mathbf{e}| : \mathbf{e} \in \mathbb{S}^{N-1}\}$

Linear Functionals that Satisfy the Minimum Principle

Suppose that $t \in (0,\infty) \longmapsto \mu_t \in \mathbf{M}_1(\mathbb{R}^N)$ is a map with the properties that

$$A\varphi \equiv \lim_{t \searrow 0} \frac{\langle \varphi, \mu_t \rangle - \varphi(\mathbf{0})}{t}$$

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

$$\varlimsup_{R \to \infty} \varlimsup_{t \searrow 0} \frac{\mu_t\big(B(\mathbf{0}, R)\complement\big)}{t} = 0.$$

Then

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

$$\varphi(\mathbf{0}) \leq \varphi \implies A\varphi \geq 0.$$

(ii) A is **quasi-local**

(1) $$\lim_{R\to\infty} A\varphi_R = 0 \quad \text{where } \varphi_R(\mathbf{x}) = \varphi\left(\frac{\mathbf{x}}{R}\right)$$

if φ is constant in a neighborhood of $\mathbf{0}$.

The first of these is obvious, since

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

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

$$\begin{aligned}\big|\langle \varphi_R, \mu_t\rangle - \varphi(\mathbf{0})\big| &\leq \int_{B(\mathbf{0},R\delta)\complement} \big|\varphi\big(\tfrac{\mathbf{y}}{R}\big) - \varphi(\mathbf{0})\big|\, \mu_t(d\mathbf{y}) \\ &\leq 2\|\varphi\|_{\mathrm{u}}\mu_t\big(B(\mathbf{0}, R\delta)\big),\end{aligned}$$

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 $\mathbf{1}$ and $-\mathbf{1}$, one knows that $A\mathbf{1} = 0$. Before going further, I have to introduce the following partition of unity for $\mathbb{R}^N \setminus \{\mathbf{0}\}$. Choose $\psi \in C^\infty\big(\mathbb{R}^N; [0,1]\big)$ so that ψ has compact support in $B(\mathbf{0}, 2) \setminus \overline{B\big(\mathbf{0}, \frac{1}{4}\big)}$ and $\psi(\mathbf{y}) = 1$ when $\frac{1}{2} \leq |\mathbf{y}| \leq 1$, and set $\psi_m(\mathbf{y}) = \psi(2^m\mathbf{y})$ for $m \in \mathbb{Z}$. Then, if $\mathbf{y} \in \mathbb{R}^N$ and $2^{-m-1} \leq |\mathbf{y}| \leq 2^{-m}$, $\psi_m(\mathbf{y}) = 1$ and $\psi_n(\mathbf{y}) = 0$ unless $-m-2 \leq n \leq -m+1$. Hence, if $\Psi(\mathbf{y}) = \sum_{m\in\mathbb{Z}} \psi_m(\mathbf{y})$ for $\mathbf{y} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$, then Ψ is a smooth function with values in $[1, 4]$; and therefore, for each $m \in \mathbb{Z}$, the function χ_m given by $\chi_m(\mathbf{0}) = 0$ and $\chi_m(\mathbf{y}) = \frac{\psi_m(\mathbf{y})}{\Psi(\mathbf{y})}$ for $\mathbf{y} \in \mathbb{R}^N \setminus \{\mathbf{0}\}$ is a smooth, $[0,1]$-valued function that vanishes off of $B(\mathbf{0}, 2^{-m+1}) \setminus \overline{B(\mathbf{0}, 2^{-m-2})}$. In addition, for each $\mathbf{y} \in \mathbb{R}^N \setminus \{0\}$, $\sum_{m\in\mathbb{Z}} \chi_m(\mathbf{y}) = 1$ and $\chi_m(\mathbf{y}) = 0$ unless $2^{-m-2} \leq |\mathbf{y}| \leq 2^{-m+1}$.

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

$$\big[\nabla^n\varphi(x)\big](\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) = \frac{\partial^n}{\partial_{t_1}\cdots\partial t_n}\varphi\left(\mathbf{x} + \sum_{m=1}^n t_m\boldsymbol{\xi}_m\right)\Bigg|_{t_1=\cdots=t_n=0}.$$

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

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

$$M(\{\mathbf{0}\}) = 0 \quad \text{and} \quad \int_{B(\mathbf{0},1)} |\mathbf{y}|^2\, M(\mathbf{y}) + M\big(B(\mathbf{0}, 1)\complement\big) < \infty.$$

THEOREM 2. *Let* $\mathbf{D}$ *be the space of functions described above. If* $A : \mathbf{D} \longrightarrow \mathbb{R}$ *is a linear functional on* $\mathbf{D}$ *that satisfies* (i) *and* (ii), *then there is a unique Lévy measure* M *such that* $A\varphi = \int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M(d\mathbf{y})$ *whenever* φ *is an element of* $\mathscr{S}(\mathbb{R}^N;\mathbb{R})$ *for which* $\varphi(\mathbf{0}) = 0$, $\nabla\varphi(\mathbf{0}) = \mathbf{0}$, *and* $\nabla^2\varphi(\mathbf{0}) = \mathbf{0}$. *Next, given* $\eta \in C_c^\infty\big(\mathbb{R}^N;[0,1]\big)$ *satisfying* $\eta = 1$ *in a neighborhood of* $\mathbf{0}$, *set* $\eta_{\boldsymbol{\xi}}(\mathbf{y}) = \eta(\mathbf{y})(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}$ *for* $\boldsymbol{\xi} \in \mathbb{R}^N$, *and define* $\mathbf{m}^\eta \in \mathbb{R}^N$ *and* $\mathbf{C} \in \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$ *by*

$$(3) \qquad (\boldsymbol{\xi},\mathbf{m}^\eta)_{\mathbb{R}^N} = A\eta_{\boldsymbol{\xi}} \quad \text{and} \quad \big(\boldsymbol{\xi},\mathbf{C}\boldsymbol{\xi}'\big)_{\mathbb{R}^N} = A\big(\eta_{\boldsymbol{\xi}}\eta_{\boldsymbol{\xi}'}\big) - \int_{\mathbb{R}^N} (\eta_{\boldsymbol{\xi}}\eta_{\boldsymbol{\xi}'})(\mathbf{y})\, M(d\mathbf{y}).$$

Then $\mathbf{C}$ *is symmetric, non-negative definite, and independent of the choice of* η. *Finally, for any* $\varphi \in \mathbf{D}$,

$$(4) \qquad \begin{aligned} A\varphi &= \tfrac{1}{2}\mathrm{Trace}\big(\mathbf{C}\nabla^2\varphi(\mathbf{0})\big) + \big(\mathbf{m}^\eta, \nabla\varphi(\mathbf{0})\big)_{\mathbb{R}^N} \\ &\quad + \int_{\mathbb{R}^N} \Big(\varphi(\mathbf{y}) - \varphi(\mathbf{0}) - \eta(\mathbf{y})\big(\mathbf{y},\nabla\varphi(\mathbf{0})\big)_{\mathbb{R}^N}\Big)\, M(d\mathbf{y}). \end{aligned}$$

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 $A\mathbf{1} = 0$. Next suppose that $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$ and that $\varphi(\mathbf{0}) \le \varphi$. Then $\nabla\varphi(\mathbf{0}) = \mathbf{0}$ and $\nabla^2\varphi(\mathbf{0})$ is non-negative definite. Hence $\big(\mathbf{m},\nabla\varphi(\mathbf{0})\big)_{\mathbb{R}^N} = 0$, $\mathrm{Trace}\big(\mathbf{C}\nabla^2\varphi(\mathbf{0})\big) \ge 0$, and

$$\varphi(\mathbf{y}) - \varphi(\mathbf{0}) - \mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\mathbf{y},\nabla\varphi(\mathbf{0})\big)_{\mathbb{R}^N} \ge 0.$$

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

$$\begin{aligned} &\left|\int_{\mathbb{R}^N} \Big(\varphi_R(\mathbf{y}) - \varphi_R(\mathbf{0}) - \mathbf{1}_{B(\mathbf{0},1)}\big(\mathbf{y},\nabla\varphi_R(\mathbf{0})\big)_{\mathbb{R}^N}\Big) M(d\mathbf{y})\right| \\ &\qquad \le \int_{B(\mathbf{0},R\delta)\complement} |\varphi_R(\mathbf{y})|\, M(d\mathbf{y}) \le \|\varphi\|_{\mathrm{u}} M\big(B(\mathbf{0},R\delta)\complement\big) \longrightarrow 0 \end{aligned}$$

as $R \to \infty$.

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

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

Clearly Λ_m is linear. In addition, if $\varphi \ge 0$, then $\chi_m\varphi \ge 0 = \chi_m\varphi(\mathbf{0})$, and so, by (i), $\Lambda_m\varphi \ge 0$. Similarly, for any $\varphi \in C^\infty\big(\overline{B(\mathbf{0},2^{-m+1})} \setminus B(\mathbf{0},2^{-m-2});\mathbb{R}\big)$,

$\|\varphi\|_{\mathrm{u}}\chi_m \pm \chi_m\varphi \ge 0 = \big(\|\varphi\|_{\mathrm{u}}\chi_m \pm \chi_m\varphi\big)(\mathbf{0})$, and therefore $|\Lambda_m\varphi| \le K_m\|\varphi\|_{\mathrm{u}}$, where $K_m = A\chi_m$. Hence, Λ_m admits a unique extension as a continuous linear functional on $C\big(\overline{B(\mathbf{0},2^{-m+1})} \setminus B(\mathbf{0},2^{-m-2});\mathbb{R}\big)$ 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 $\overline{B(\mathbf{0},2^{-m+1})} \setminus B(\mathbf{0},2^{-m-2})$, $K_m = M_m(\mathbb{R}^N)$, and $A(\chi_m\varphi) = \int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M_m(d\mathbf{y})$ 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, $M(\{\mathbf{0}\}) = 0$. In addition, if $\varphi \in C_{\mathrm{c}}^{\infty}\big(\mathbb{R}^N \setminus \{\mathbf{0}\};\mathbb{R}\big)$, then there is an $n \in \mathbb{Z}^+$ such that $\chi_m\varphi \equiv 0$ unless $|m| \le n$. Thus,

$$\begin{aligned} A\varphi &= \sum_{m=-n}^{n} A(\chi_m\varphi) = \sum_{m=-n}^{n} \int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M_m(d\mathbf{y}) \\ &= \int_{\mathbb{R}^N} \left(\sum_{m=-n}^{n} \chi_m(\mathbf{y})\varphi(\mathbf{y}) \right) M(d\mathbf{y}) = \int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M(d\mathbf{y}), \end{aligned}$$

and therefore

$$A\varphi = \int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M(d\mathbf{y}) \tag{5}$$

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

Before taking the next step, observe that, as an application of (i), if $\varphi_1, \varphi_2 \in \mathbf{D}$, then

$$\varphi_1 \le \varphi_2 \text{ and } \varphi_1(\mathbf{0}) = \varphi_2(\mathbf{0}) \implies A\varphi_1 \le A\varphi_2. \tag{*}$$

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

With these preparations, I can show that, for any $\varphi \in \mathbf{D}$,

$$\varphi \ge 0 = \varphi(\mathbf{0}) \implies \int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M(d\mathbf{y}) \le A\varphi. \tag{**}$$

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

$$\int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M(d\mathbf{y}) = \lim_{n\to\infty} \int_{\mathbb{R}^N} \varphi_n(\mathbf{y})\, M(d\mathbf{y}) = \lim_{n\to\infty} A\varphi_n \le A\varphi.$$

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

$$\int_{\mathbb{R}^N} |\mathbf{y}|^2\eta(\mathbf{y})\, M(d\mathbf{y}) \le A\varphi < \infty.$$

At the same time, by (5) and (*),

$$\int_{\mathbb{R}^N} \big(1-\eta(\mathbf{y})\big)\eta_R(\mathbf{y})\, M(d\mathbf{y}) \le A(\mathbf{1}-\eta)$$

for all $R>0$, and therefore, by Fatou's Lemma,

$$\int_{\mathbb{R}^N} \big(1-\eta(\mathbf{y})\big)\, M(d\mathbf{y}) \le A(\mathbf{1}-\eta) < \infty.$$

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 $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{R})$ that vanishes along with its first and second order derivatives at $\mathbf{0}$. To this end, first suppose that φ vanishes in a neighborhood of $\mathbf{0}$. Then, for each $R>0$, (5) applies to $\eta_R\varphi$, and so

$$\int_{\mathbb{R}^N} \eta_R(\mathbf{y})\varphi(\mathbf{y})\, M(d\mathbf{y}) = A(\eta_R\varphi) = A\varphi + A\big((\mathbf{1}-\eta_R)\varphi\big).$$

By (*) applied to $\pm(\mathbf{1}-\eta_R)\varphi$ and $(\mathbf{1}-\eta_R)\|\varphi\|_{\mathrm{u}}$,

$$\big|A\big((\mathbf{1}-\eta_R)\varphi\big)\big| \le \|\varphi\|_{\mathrm{u}} A(\mathbf{1}-\eta_R) \longrightarrow 0 \quad \text{as } R\to\infty,$$

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

$$A\varphi = \lim_{R\to\infty}\int_{\mathbb{R}^N} \eta_R(\mathbf{y})\varphi(\mathbf{y})\, M(d\mathbf{y}) = \int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M(d\mathbf{y}),$$

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

$$\int_{\mathbb{R}^N} \varphi(\mathbf{y})\, M(d\mathbf{y}) = \lim_{R\searrow 0} A\big((\mathbf{1}-\eta_R)\varphi\big) = A\varphi - \lim_{R\searrow 0} A(\eta_R\varphi).$$

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

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

$$\tilde\varphi(\mathbf{x}) = \varphi(\mathbf{x}) - \varphi(\mathbf{0}) - \eta(\mathbf{x})\big(\mathbf{x},\nabla\varphi(\mathbf{0})\big)_{\mathbb{R}^N} - \tfrac12\eta(\mathbf{x})^2\big(\mathbf{x},\nabla^2\varphi(\mathbf{0})\mathbf{x}\big)_{\mathbb{R}^N}.$$

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 η chosen, let η' be a second choice, note that $\eta'_{\boldsymbol{\xi}} = \eta_{\boldsymbol{\xi}}$ in a neighborhood of $\mathbf{0}$, 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 φ 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 η. Indeed, one can choose any measurable $\eta : \mathbb{R}^N \longrightarrow [0,1]$ with the property the

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

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

Finally, notice that when A is **local** in the sanse that $A\varphi = 0$ if φ is constant in a neighborhood of $\mathbf{0}$, then the corresponding $M = 0$.

My next goal is to show that for each choice of $\mathbf{m} \in \mathbb{R}^N$, non-negative definite, symmetric matix $\mathbf{C}$, and Lévy measure M, there is a natural map $t \in [0,\infty) \longrightarrow \mu_t \in \mathbf{M}_1(\mathbb{R}^N)$ such that

$$\begin{aligned} A\varphi &\equiv (\mathbf{m}, \nabla\varphi) + \tfrac{1}{2}\mathrm{Trace}(\mathbf{C}\nabla^2\varphi) \\ &\qquad + \int_{\mathbb{R}^N} \Big(\varphi(\mathbf{y}) - 1 - \mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\mathbf{y}, \nabla\varphi(0)\big)_{\mathbb{R}^N}\Big) M(d\mathbf{y}) \\ &= \lim_{t\searrow 0} \frac{\langle \varphi, \mu_t\rangle - \varphi(\mathbf{0})}{t} \quad \text{for } \varphi \in \mathbf{D}. \end{aligned}$$

LEMMA 7. *Set*

$$\ell(\boldsymbol{\xi}) = i(\mathbf{m}, \boldsymbol{\xi})_{\mathbb{R}^N} - \tfrac{1}{2}\big(\boldsymbol{\xi}, \mathbf{C}\boldsymbol{\xi}\big)_{\mathbb{R}^N} + \int_{\mathbb{R}^N} \Big(e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}} - 1 - i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\boldsymbol{\xi}, \mathbf{y}\big)_{\mathbb{R}^N}\Big) M(d\mathbf{y}).$$

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

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

PROOF: Begin by noting that

$$\mathfrak{Re}\big(\ell(\boldsymbol{\xi})\big) = -\tfrac{1}{2}\big(\boldsymbol{\xi}, \mathbf{C}\boldsymbol{\xi}\big)_{\mathbb{R}^N} + \int_{\mathbb{R}^N} \big(\cos(\boldsymbol{\xi}, \mathbf{y})_{\mathbb{R}^N} - 1\big) M(d\mathbf{y}) \le 0.$$

Next, given $r \in (0,1]$, observe that

$$\begin{aligned}&\left|\int_{\mathbb{R}^N}\Big(e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}}-1-i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\boldsymbol{\xi},\mathbf{y}\big)_{\mathbb{R}^N}\Big)M(d\mathbf{y})\right|\\&\qquad\le\frac{|\boldsymbol{\xi}|^2}{2}\int_{B(\mathbf{0},r)}|\mathbf{y}|^2\,M(d\mathbf{y})+(2+|\boldsymbol{\xi}|)M\big(B(\mathbf{0},r)\complement\big),\end{aligned}$$

and therefore

$$\begin{aligned}&\lim_{|\boldsymbol{\xi}|\to\infty}|\boldsymbol{\xi}|^{-2}\left|\int_{\mathbb{R}^N}\Big(e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}}-1-i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\boldsymbol{\xi},\mathbf{y}\big)_{\mathbb{R}^N}\Big)M(d\mathbf{y})\right|\\&\qquad\le\frac{1}{2}\int_{B(\mathbf{0},r)}|\mathbf{y}|^2\,M(d\mathbf{y}).\end{aligned}$$

Since $\int_{B(\mathbf{0},r)}|\mathbf{y}|^2\,M(d\mathbf{y})\longrightarrow 0$ as $r\searrow 0$, this completes the proof of the intitial assertions.

To prove the final assertion, note that, by Taylor's Theorem,

$$\big|e^{t\ell(\boldsymbol{\xi})}-1-t\ell(\boldsymbol{\xi})\big|\le\frac{t^2|\ell(\boldsymbol{\xi})|^2}{2}\le\frac{K^2t^2(1+|\boldsymbol{\xi}|^2)^2}{2}.$$

Hence, since $\hat{\varphi}$ is rapidly decreasing, Lebesgue's Dominated Convergence Theorem shows that

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

By the Fourier inversion formula,

$$(2\pi)^{-N}\int_{\mathbb{R}^N}\Big(-i(\mathbf{m},\boldsymbol{\xi})_{\mathbb{R}^N}-\frac{1}{2}\big(\boldsymbol{\xi},\mathbf{C}\boldsymbol{\xi}\big)_{\mathbb{R}^N}\Big)\hat{\varphi}(\boldsymbol{\xi})\,d\boldsymbol{\xi}=\big(\mathbf{m},\nabla\varphi(\mathbf{0})\big)_{\mathbb{R}^N}+\tfrac{1}{2}\mathrm{Trace}\big(\mathbf{C}\nabla^2\varphi(\mathbf{0})\big).$$

Finally, because

$$\iint_{\mathbb{R}^N\times\mathbb{R}^N}|\hat{\varphi}(\boldsymbol{\xi})|\Big|e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}}-1-i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\boldsymbol{\xi},\mathbf{y}\big)_{\mathbb{R}^N}\Big|\,M(d\mathbf{y})d\boldsymbol{\xi}<\infty,$$

Fubini's Theorem applies and says that

$$\begin{aligned}&\int_{\mathbb{R}^N}\hat{\varphi}(\xi)\left(\int_{\mathbb{R}^N}\Big(e^{-i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}}-1-i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\boldsymbol{\xi},\mathbf{y}\big)_{\mathbb{R}^N}\Big)\,M(d\mathbf{y})\right)d\boldsymbol{\xi}\\&\quad=\int_{\mathbb{R}^N}\left(\int_{\mathbb{R}^N}\Big(e^{-i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}}-1-i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\boldsymbol{\xi},\mathbf{y}\big)_{\mathbb{R}^N}\Big)\hat{\varphi}(\boldsymbol{\xi})\,d\boldsymbol{\xi}\right)M(d\mathbf{y})\\&\quad=(2\pi)^N\int_{\mathbb{R}^N}\Big(\varphi(\mathbf{y})-\varphi(\mathbf{0})-\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\mathbf{y},\nabla\varphi(\mathbf{0})\big)_{\mathbb{R}^N}\Big)\,M(d\mathbf{y}),\end{aligned}$$

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 $\mu_t \in \mathbf{M}_1(\mathbb{R}^N)$ such that $\widehat{\mu_t}(\boldsymbol{\xi}) = e^{t\ell(\boldsymbol{\xi})}$. Indeed, by Parseval's identity,

$$\langle \varphi, \mu_t \rangle - \varphi(\mathbf{0}) = (2\pi)^{-N} \int_{\mathbb{R}^N} \hat{\varphi}(\boldsymbol{\xi})\big(\widehat{\mu_t} - 1\big)(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

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

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

$$\gamma_{\mathbf{0},\mathbf{I}}(d\mathbf{y}) = (2\pi)^{-\frac{N}{2}} e^{-\frac{|\mathbf{y}|^2}{2}}\, d\mathbf{y}.$$

To deal with the term corresponding to M, initially assume that M is finite, and consider the Poisson measure

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

where $M^{*0} \equiv \delta_{\mathbf{0}}$ ($\delta_{\mathbf{a}}$ 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

$$\widehat{\Pi_M}(\boldsymbol{\xi}) = e^{M(\mathbb{R}^N)} \sum_{k=0}^{\infty} \frac{\hat{M}(\boldsymbol{\xi})^k}{k!} = \exp\left(\int_{\mathbb{R}^N} \left(e^{i(\boldsymbol{\xi},\mathbf{y})} - 1\right) M(d\mathbf{y}) \right).$$

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

$$\begin{aligned}(\delta_{-\mathbf{a}_r} * \Pi_{M_r})\hat{\ }(\boldsymbol{\xi}) &= \exp\left(\int_{B(\mathbf{0},r)\complement} \left(e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}} - 1 - i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N} \right) M(d\mathbf{y}) \right) \\ &\longrightarrow \exp\left(\int_{\mathbb{R}^N} \left(e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}} - 1 - i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N} \right) M(d\mathbf{y}) \right)\end{aligned}$$

uniformly for $\boldsymbol{\xi}$ 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 $\hat{\mu}_n$ is the characteristic function (i.e., the Fourier transform) of μ_n, then $\mu = \lim_{n\to\infty} \mu_n$ exists in $\mathbf{M}_1(\mathbb{R}^N)$ if and only if $\hat{\mu}_n(\boldsymbol{\xi})$ converges for each $\boldsymbol{\xi}$ and uniformly in a neighborhood of 0, in which case $\mu_n \longrightarrow \mu$ in $\mathbf{M}_1(\mathbb{R}^N)$ where $\hat{\mu}(\boldsymbol{\xi}) = \lim_{n\to\infty} \hat{\mu}_n(\boldsymbol{\xi})$), there is an element μ_M of $\mathbf{M}_1(\mathbb{R}^N)$ whose Fourier transform is

$$\exp\left(\int_{\mathbb{R}^N} \left(e^{i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}} - 1 - i\mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N} \right) M(d\mathbf{y}) \right).$$

Hence, if $\mu = \delta_{\mathbf{m}} * \mu_M * \gamma_{\mathbf{0},\mathbf{C}}$, then $\hat{\mu}(\boldsymbol{\xi}) = e^{\ell(\boldsymbol{\xi})}$.

Kolmogorov's Equations

Notice that $\mu_{s+t} = \mu_s * \mu_t$, and therefore, for each $n \geq 1$, $\mu_1 = \mu_{\frac{1}{n}}^{*n}$. A probability measure that admits an nth 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 μ are those for which there exists a $(\mathbf{m}, \mathbf{C}, M)$ such that $\hat{\mu}(\boldsymbol{\xi}) = e^{\ell(\boldsymbol{\xi})}$. Perhaps more important from the perspective here is the evolution equation, known as Kolmogorov's **forward equaltion**, that this property implies $t \rightsquigarrow \mu_t$ satisfies. To describe this equation, define the operator L on $\mathbf{D}$ by

$$\begin{aligned} L\varphi(\mathbf{x}) &= A\varphi(\mathbf{x} + \,\cdot\,) \\ &= \big(\mathbf{m}, \nabla\varphi(\mathbf{x})\big)_{\mathbb{R}^N} + \tfrac{1}{2}\mathrm{Trace}\big(\mathbf{C}\nabla^2\varphi(\mathbf{x})\big) \\ &\qquad + \int_{\mathbb{R}^N} \Big(\varphi(\mathbf{x}+\mathbf{y}) - \varphi(\mathbf{x}) - \mathbf{1}_{B(\mathbf{0},1)}(\mathbf{y})\big(\mathbf{y}, \nabla\varphi(\mathbf{x})\big)_{\mathbb{R}^N}\Big) M(d\mathbf{y}). \end{aligned}$$

What I want to show is that

$$\frac{d}{dt}\langle \varphi, \mu_t\rangle = \langle L\varphi, \mu_t\rangle. \tag{8}$$

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

$$\begin{aligned} \frac{\langle \varphi, \mu_{t+h}\rangle - \langle \varphi, \mu_t\rangle}{h} &= \frac{1}{h}\int_{\mathbb{R}^N}\left(\int_{\mathbb{R}^N} \frac{\langle \varphi(\mathbf{y} + \,\cdot\,), \mu_h\rangle - \varphi(\mathbf{y})}{h}\right)\mu_t(d\mathbf{y}) \\ = (2\pi)^{-N}&\int_{\mathbb{R}^N}\left(\int_{\mathbb{R}^N} e^{-i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}} \frac{e^{t\ell(-\boldsymbol{\xi})} - 1}{h}\hat{\varphi}(\boldsymbol{\xi})\, d\boldsymbol{\xi}\right)\mu_t(d\mathbf{y}) \\ \longrightarrow (2\pi)^{-N}&\int_{\mathbb{R}^N}\left(\int_{\mathbb{R}^N} e^{-i(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}} \ell(-\boldsymbol{\xi})\hat{\varphi}(\boldsymbol{\xi})\, d\boldsymbol{\xi}\right)\mu_t(d\mathbf{y}) = \langle L\varphi, \mu_t\rangle. \end{aligned}$$

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

$$\begin{aligned} |L\varphi(\mathbf{x})| \leq& \|\mathbf{C}\|_{\mathrm{op}}\|\nabla^2\varphi(\mathbf{x})\|_{\mathrm{op}} + |\mathbf{m}||\nabla\varphi(\mathbf{x})| \\ &+ \frac{1}{2}\left(\int_{B(\mathbf{0},1)} |\mathbf{y}|^2\, M(d\mathbf{y})\right) \sup_{\mathbf{y}\in B(\mathbf{x},1)} \|\nabla^2\varphi(\mathbf{y})\|_{\mathrm{op}} + 2M\big(B(\mathbf{0},1)\complement\big)\|\varphi\|_{\mathrm{u}}. \end{aligned}$$

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

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

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

$$\frac{d}{dt}\widehat{\mu_t}(\boldsymbol{\xi})=\ell(\boldsymbol{\xi})\widehat{\mu_t}(\boldsymbol{\xi}).$$

Hence (8) together with the initial condition $\mu_0=\delta_0$ implies that $\widehat{\mu_t}(\boldsymbol{\xi})=e^{t\ell(\boldsymbol{\xi})}$. That is, (8) plus $\mu_0=\delta_{\mathbf{0}}$ uniquely determine μ_t.

There is another important equation associated with the measures μ_t. To develop this equation, set $P(t,\mathbf{x},\,\cdot\,)=\delta_{\mathbf{x}}*\mu_t$. Then

$$\begin{aligned}P(s+t,\mathbf{x},\Gamma)&=\int_{\mathbb{R}^N}\mu_t(\mathbf{x}+\mathbf{y}+\Gamma)\,\mu_s(d\mathbf{y})=\int_{\mathbb{R}^N}\mu_t(\mathbf{y}+\Gamma)\,P(s,\mathbf{x},d\mathbf{y})\\&=\int_{\mathbb{R}^N}P(t,\mathbf{y},\Gamma)\,P(s,\mathbf{x},d\mathbf{y}),\end{aligned}$$

and so $P(t,\mathbf{x},\,\cdot\,)$ satisfies the **Chapman–Kolomogorov equation**

$$P(s+t,\mathbf{x},\Gamma)=\int_{\mathbb{R}^N}P(t,\mathbf{y},\Gamma)\,P(s,\mathbf{x},d\mathbf{y}),\tag{9}$$

which means that $(t,\mathbf{x})\rightsquigarrow P(t,\mathbf{x},\,\cdot\,)$ is a **transition probability**. Equivalently, if the operators $\mathbf{P}_t$ are defined for $t\geq 0$ on bounded, Borel measurable functions φ by

$$\mathbf{P}_t\varphi(\mathbf{x})=\int_{\mathbb{R}^N}\varphi(\mathbf{y})\,P(t.\mathbf{x},d\mathbf{y}),$$

then $\{\mathbf{P}_t:\,t\geq 0\}$ is a **semigroup**: $\mathbf{P}_{s+t}=\mathbf{P}_s\mathbf{P}_t$. To compute the generator of this semigroup, first observe that (8) says that

$$\frac{d}{dt}\mathbf{P}_t\varphi=\mathbf{P}_tL\varphi\tag{10}$$

for $\varphi\in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$, which is also called Kolmogorov's **forward equation**. Next observe that $\mathbf{P}_t$ is a bounded map of $C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$ into itself. Therefore, if $\varphi\in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$, then

$$\frac{\mathbf{P}_{t+h}\varphi(\mathbf{x})-\mathbf{P}_t\varphi(\mathbf{x})}{h}=\frac{\mathbf{P}_h\mathbf{P}_t\varphi(\mathbf{x})-\mathbf{P}_t\varphi(\mathbf{x})}{h}\longrightarrow L\mathbf{P}_t\varphi(\mathbf{x})$$

as $h \searrow 0$. Equivalently, if $u(t,\mathbf{x}) \equiv \mathbf{P}_t\varphi(\mathbf{x})$, then $u \in C^{1,2}_{\mathrm{b}}\big([0,\infty)\times\mathbb{R}^N;\mathbb{R}\big)$ and is a solution to Kolmogorov's **backward equation**

$$\partial_t u = Lu \quad \text{and} \quad \lim_{t\searrow 0} u(t,\,\cdot\,) = \varphi. \tag{11}$$

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

$$\frac{d}{d\tau}\mathbf{P}_\tau u(t-\tau,\,\cdot\,) = \mathbf{P}_\tau Lu(t-\tau,\,\cdot\,) - \mathbf{P}_\tau Lu(t-\tau,\,\cdot\,) = 0$$

for $\tau \in (0,t)$, and so $\mathbf{P}_t\varphi = u(t,\,\cdot\,)$.

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,

$$L\varphi(\mathbf{x}) = \frac{1}{2}\sum_{i,j=1}^{N} a_{ij}(\mathbf{x})\partial_{x_i}\partial_{x_j}\varphi(\mathbf{x}) + \sum_{i=1}^{N} b_i(\mathbf{x})\partial_{x_i}\varphi(\mathbf{x}), \tag{12}$$

where $a(\mathbf{x}) = \big(\big(a_{ij}(\mathbf{x})\big)\big)_{1\le i,j\le N}$ 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*

$$\Lambda \equiv \sup_{x\in\mathbb{R}^N} \frac{\mathrm{Trace}\big(a(\mathbf{x})\big) + 2\big(\mathbf{x}, b(\mathbf{x})\big)^+_{\mathbb{R}^N}}{1+|\mathbf{x}|^2} < \infty. \tag{14}$$

Then, for each $\nu \in \mathbf{M}_1(\mathbb{R}^N)$, there is a continuous $t \in [0,\infty) \longmapsto \mu(t) \in \mathbf{M}_1(\mathbb{R}^N)$ which satisfies

$$\langle \varphi, \mu(t)\rangle - \langle \varphi, \nu\rangle = \int_0^t \langle L\varphi, \mu(\tau)\rangle\, d\tau, \tag{15}$$

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

$$\int (1+|\mathbf{y}|^2)\,\mu(t,d\mathbf{y}) \le e^{\Lambda t}\int (1+|\mathbf{x}|^2)\,\nu(d\mathbf{x}), \quad t \ge 0. \tag{16}$$

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 $\nu = \delta_{\mathbf{x}}$, we expect that $\mu_t = \delta_{X(t)}$, where $t \in [0,\infty) \longmapsto X(t) \in \mathbb{R}^N$ satisfies

$$\varphi\big(X(t)\big) - \varphi(\mathbf{x}) = \int_0^t \big(b(X(\tau)), \nabla\varphi(X(\tau)\big)_{\mathbb{R}^N}\, d\tau.$$

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

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

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

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

Clearly,

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

where[1] $[\tau]_n = 2^{-n}[2^n\tau]$ is the largest diadic number $m2^{-n}$ dominated by τ. Hence, if we can show that $\{X_n : n \geq 0\}$ is relatively compact in the space $C\big([0,\infty);\mathbb{R}^N\big)$, with the topology of uniform convergence on compacts, then we can take $t \rightsquigarrow X(t)$ 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 $|X_n(t) - X_n(s)| \leq \|b\|_{\mathrm{u}}|t-s|$, and so the Ascoli–Arzela Theorem guarantees the required compactness. To remove the boundedness assumption, choose a $\psi \in C_{\mathrm{c}}^\infty\big(B(0,2);[0,1]\big)$ so that $\psi = 1$ on $\overline{B(0,1)}$ and, for each $k \geq 1$, replace b by b_k, where $b_k(\mathbf{x}) = \psi(k^{-1}\mathbf{x})b(\mathbf{x})$. Next, let $t \rightsquigarrow X_k(t)$ be an integral curve of b_k starting at x, and observe that

$$\frac{d}{dt}|X_k(t)|^2 = 2\big(X_k(t), b_k(X_k(t))\big)_{\mathbb{R}^N} \leq \Lambda\big(1 + |X_k(t)|^2\big),$$

from which it is an easy step to the conclusion that

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

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 $\{\varphi_k : k \geq 1\} \subseteq C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$ *converges to* φ *in* $C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$ and write $\varphi_k \longrightarrow \varphi$ in $C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$ if $\sup_k \|\varphi_k\|_{\mathrm{u}} < \infty$ and $\varphi_k(\mathbf{x}) \longrightarrow \varphi(\mathbf{x})$ uniformly for $\mathbf{x}$ in compact subsets of $\mathbb{R}^N$. Also, we say that $\{\mu_k : k \geq 1\} \subseteq C\big(\mathbb{R}^M;\mathbf{M}_1(\mathbb{R}^N)\big)$ *converges to* μ *in* $C\big(\mathbb{R}^M;\mathbf{M}_1(\mathbb{R}^N)\big)$ and write $\mu_k \longrightarrow \mu$ in $C\big(\mathbb{R}^M;\mathbf{M}_1(\mathbb{R}^N)\big)$ if, for each $\varphi \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$, $\langle \varphi, \mu_k(\mathbf{z})\rangle \longrightarrow \langle \varphi, \mu(\mathbf{z})\rangle$ uniformly for $\mathbf{z}$ in compact subsets of $\mathbb{R}^M$.

Theorem 17. *If* $\mu_k \longrightarrow \mu$ *in* $C\big(\mathbb{R}^M;\mathbf{M}_1(\mathbb{R}^N)\big)$, *then*

$$\langle \varphi_k, \mu_k(\mathbf{z}_k)\rangle \longrightarrow \langle \varphi, \mu(\mathbf{z})\rangle$$

whenever $\mathbf{z}_k \longrightarrow \mathbf{z}$ *in* $\mathbb{R}^M$ *and* $\varphi_k \longrightarrow \varphi$ *in* $C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$. *Moreover, if* $\{\mu_n : n \geq 0\} \subseteq C\big(\mathbb{R}^M;\mathbf{M}_1(\mathbb{R}^N)\big)$ *and* $f_n(\mathbf{z},\boldsymbol{\xi}) = \widehat{\mu_n(\mathbf{z})}(\boldsymbol{\xi})$, *then* $\{\mu_n : n \geq 0\}$ *is relatively compact in* $C\big(\mathbb{R}^M;\mathbf{M}_1(\mathbb{R}^N)\big)$ *if* $\{f_n : n \geq 0\}$ *is equicontinuous at each* $(\mathbf{z},\boldsymbol{\xi}) \in \mathbb{R}^M \times \mathbb{R}^N$. *In particular,* $\{\mu_n : n \geq 0\}$ *is relatively compact if, for each* $\boldsymbol{\xi} \in \mathbb{R}^N$, $\{f_n(\,\cdot\,,\boldsymbol{\xi}) : n \geq 0\}$ *is equicontinuous at each* $\mathbf{z} \in \mathbb{R}^M$ *and, for each* $r \in (0,\infty)$,

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

Proof: To prove the first assertion, suppose $\mu_k \longrightarrow \mu$ in $C\big(\mathbb{R}^M;\mathbf{M}_1(\mathbb{R}^N)\big)$, $\mathbf{z}_k \longrightarrow \mathbf{z}$ in $\mathbb{R}^M$, and $\varphi_k \longrightarrow \varphi$ in $C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$. Then, for every $R > 0$,

$$\begin{aligned}
&\varlimsup_{k\to\infty} \big|\langle \varphi_k, \mu_k(\mathbf{z}_k)\rangle - \langle \varphi, \mu(\mathbf{z})\rangle\big| \\
&\leq \varlimsup_{k\to\infty} \Big(\big|\langle \varphi_k - \varphi, \mu_k(\mathbf{z}_k)\rangle\big| + \big|\langle \varphi, \mu_k(\mathbf{z}_k)\rangle - \langle \varphi, \mu(\mathbf{z})\rangle\big|\Big) \\
&\leq \varlimsup_{k\to\infty} \sup_{y\in B(0,R)} |\varphi_k(y) - \varphi(y)| + 2\sup_k \|\varphi_k\|_{\mathrm{u}} \varlimsup_{k\to\infty} \mu_k\big(\mathbf{z}_k, B(0,R)\complement\big) \\
&\leq 2\sup_k \|\varphi_k\|_{\mathrm{u}} \mu\big(\mathbf{z}, B(0,R)\complement\big)
\end{aligned}$$

since $\varlimsup_{k\to\infty} \mu_k(\mathbf{z}_k, F) \leq \mu(\mathbf{z}, F)$ 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 $f \in C_{\mathrm{b}}(\mathbb{R}^M \times \mathbb{R}^N;\mathbb{C})$ 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 $\mathbf{z} \in \mathbb{R}^M$, a $\mu(\mathbf{z}) \in \mathbf{M}_1(\mathbb{R}^N)$ for which $f(\mathbf{z},\,\cdot\,) = \widehat{\mu(\mathbf{z})}$. Moreover, if $\mathbf{z}_k \longrightarrow \mathbf{z}$ in $\mathbb{R}^M$, then, because $f_{n_k}(\mathbf{z}_k,\,\cdot\,) \longrightarrow f(\mathbf{z},\,\cdot\,)$ uniformly on compact subsets of $\mathbb{R}^N$, another application of Lévy's Theorem shows that $\mu_{n_k}(\mathbf{z}_k) \longrightarrow \mu(\mathbf{z})$ in $\mathbf{M}_1(\mathbb{R}^N)$, and from this it is clear that $\mu_{n_k} \longrightarrow \mu$ in $C\big(\mathbb{R}^M; \mathbf{M}_1(\mathbb{R}^N)\big)$.

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

$$\lim_{k\to\infty} \sup_{n\geq 0} \sup_{|\mathbf{z}|\leq r} \big|f_n(\mathbf{z}, \boldsymbol{\xi}_k) - f_n(\mathbf{z}, \boldsymbol{\xi})\big| = 0.$$

To this end, note that, for any $R > 0$,

$$\big|f_n(\mathbf{z}, \boldsymbol{\xi}_k) - f_n(\mathbf{z}, \boldsymbol{\xi})\big| \leq R|\boldsymbol{\xi}_k - \boldsymbol{\xi}| + 2\mu_n\big(\mathbf{z}, B(0,R)\complement\big),$$

and therefore

$$\varlimsup_{k\to\infty} \sup_{n\geq 0} \sup_{|\mathbf{z}|\leq r} \big|f_n(\mathbf{z}, \boldsymbol{\xi}_k) - f_n(\mathbf{z}, \boldsymbol{\xi})\big| \leq 2 \sup_{n\geq 0} \sup_{|\mathbf{z}|\leq r} \mu_n\big(\mathbf{z}, B(0,R)\complement\big) \longrightarrow 0$$

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 $\mathbf{M}_1(\mathbb{R}^N)$ that linear translation plays in $\mathbb{R}^N$. A hint comes from the observation that if $t \rightsquigarrow X(t, \mathbf{x}) = x + tb$ is a linear translation along the constant vector field b, then $X(s+t, \mathbf{x}) = X(s, \mathbf{x}) + X(t, \mathbf{0})$. Equivalently,

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

Thus, "linear translation" in $\mathbf{M}_1(\mathbb{R}^N)$ should be a path $t \in [0,\infty) \longmapsto \mu(t) \in \mathbf{M}_1(\mathbb{R}^N)$ given by $\mu(t) = \nu * \lambda(t)$, where $t \rightsquigarrow \lambda(t)$ satisfies $\lambda(0) = \delta_{\mathbf{0}}$ 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 $\mathbf{y} \rightsquigarrow x + tb(\mathbf{x}) + t^{\frac{1}{2}}\sigma(\mathbf{x})\mathbf{y}$ under $\gamma_{\mathbf{0},\mathbf{I}}$, where $\sigma : \mathbb{R}^N \longrightarrow \mathrm{Hom}(\mathbb{R}^M; \mathbb{R}^N)$ is a square root[2] 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 φ together with its derivatives have at most exponential growth, then

$$\begin{aligned} &\langle \varphi, Q(t,\mathbf{x})\rangle - \varphi(\mathbf{x}) = \int_0^t \langle L^{\mathbf{x}}\varphi, Q(\tau, x)\rangle \, d\tau, \\ &\text{where } L^{\mathbf{x}}\varphi(\mathbf{y}) = \frac{1}{2}\sum_{i,j}^N a(\mathbf{x})\partial_{\mathbf{y}_i}\partial_{\mathbf{y}_j}\varphi(\mathbf{y}) + \sum_{i=1}^N b_i(\mathbf{x})\partial_{\mathbf{y}_i}\varphi(\mathbf{y}). \end{aligned} \tag{18}$$

[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

$$\begin{aligned}\frac{d}{dt}\langle\varphi,Q(t,\mathbf{x})\rangle &= \frac{d}{dt}\int_{\mathbb{R}^N}\varphi\big(\mathbf{x}+\sigma(\mathbf{x})\mathbf{y}+tb(\mathbf{x})\mathbf{y}\big)\,\gamma_{\mathbf{0},t\mathbf{I}}(d\mathbf{y})\\ &= \int_{\mathbb{R}^N}\varphi\big(\mathbf{x}+\sigma(\mathbf{x})\mathbf{y}+tb(\mathbf{x})\big)\Delta_{\mathbf{y}}\left(\frac{1}{(2\pi t)^{\frac{N}{2}}}e^{-\frac{|\mathbf{y}|^2}{2t}}\right)\,d\mathbf{y},\end{aligned}$$

and integrate twice by parts to move the Δ over to φ. As a consequence of either (18) or direct computation, we have

$$\int |y|^2\,Q(t,\mathbf{x},d\mathbf{y}) = \big|\mathbf{x}+tb(\mathbf{x})\big|^2 + t\mathrm{Trace}\big(a(\mathbf{x})\big). \tag{19}$$

Now, for each $n\geq 0$, define the Euler approximation $t\in[0,\infty)\longmapsto\mu_n(t)\in\mathbf{M}_1(\mathbb{R}^N)$ so that

$$\begin{gathered}\mu_n(0)=\nu \quad\text{and}\quad \mu_n(t)=\int Q\big(t-m2^{-n},\mathbf{y}\big)\,\mu_n(m2^{-n},d\mathbf{y})\\ \text{for } m2^{-n}<t\leq(m+1)2^{-n}.\end{gathered} \tag{20}$$

By (19), we know that

$$\begin{aligned}\int_{\mathbb{R}^N}|\mathbf{y}|^2\,\mu_n(t,d\mathbf{y}) = \int_{\mathbb{R}^N}\Big[\big|y+(t-m2^{-n})b(\mathbf{y})\big|^2&\\ +(t-m2^{-n})\mathrm{Trace}\big(a(\mathbf{y})\big)\Big]\,\mu_n\big(m2^{-n},d\mathbf{y}\big)&\end{aligned} \tag{21}$$

for $m2^{-n}\leq t\leq(m+1)2^{-n}$.

LEMMA 22. *Assume that*

$$\lambda\equiv\sup_{x\in\mathbb{R}^N}\frac{\mathrm{Trace}\big(a(\mathbf{x})\big)+2|b(\mathbf{x})|^2}{1+|\mathbf{x}|^2}<\infty. \tag{23}$$

Then

$$\sup_{n\geq 0}\int_{\mathbb{R}^N}(1+|\mathbf{y}|^2)\,\mu_n(t,d\mathbf{y})\leq e^{(1+\lambda)t}\int_{\mathbb{R}^N}(1+|\mathbf{x}|^2)\,\nu(d\mathbf{x}). \tag{24}$$

In particular, if $\int|\mathbf{x}|^2\,\nu(d\mathbf{x})<\infty$, *then* $\{\mu_n : n\geq 0\}$ *is a relatively compact subset of* $C\big([0,\infty);\mathbf{M}_1(\mathbb{R}^N)\big)$.

PROOF: Suppose that $m2^{-n} \le t \le (m+1)2^{-n}$, and set $\tau = t - m2^{-n}$. First note that

$$
\begin{aligned}
&\big|\mathbf{y} + \tau b(\mathbf{y})\big|^2 + \tau \mathrm{Trace}\big(a(\mathbf{y})\big) \\
&= |\mathbf{y}|^2 + 2\tau\big(\mathbf{y}, b(\mathbf{y})\big)_{\mathbb{R}^N} + \tau^2 |b(\mathbf{y})|^2 + \tau \mathrm{Trace}\big(a(\mathbf{y})\big) \\
&\le |\mathbf{y}|^2 + \tau\big[|\mathbf{y}|^2 + 2|b(\mathbf{y})|^2 + \mathrm{Trace}\big(a(\mathbf{y})\big)\big] \le |\mathbf{y}|^2 + (1+\lambda)\tau(1+|\mathbf{y}|^2),
\end{aligned}
$$

and therefore, by (21),

$$
\int (1+|\mathbf{y}|^2)\, \mu_n(t, d\mathbf{y}) \le \big(1 + (1+\lambda)\tau\big) \int (1+|\mathbf{y}|^2)\, \mu_n\big(m2^{-n}, d\mathbf{y}\big).
$$

Hence,

$$
\begin{aligned}
&\int (1+|\mathbf{y}|^2)\, \mu_n(t, d\mathbf{y}) \\
&\quad \le \big(1 + (1+\lambda)2^{-n}\big)^m \big(1 + (1+\lambda)\tau\big) \int (1+|\mathbf{y}|^2)\, \nu(d\mathbf{y}) \\
&\quad \le e^{(1+\lambda)t} \int (1+|\mathbf{x}|^2)\, \nu(d\mathbf{x}).
\end{aligned}
$$

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

$$
\mu_n\big(t, \overline{B(0,R)}\complement\big) \le S(1+R^2)^{-1} e^{(1+\lambda)t},
$$

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

$$
\big|f_n(t, \boldsymbol{\xi}) - f_n(s, \boldsymbol{\xi})\big| \le \int \big|[\widehat{Q(t,y)}](\boldsymbol{\xi}) - [\widehat{Q(s,y)}](\boldsymbol{\xi})\big|\, \mu_n\big(m2^{-n}, d\mathbf{y}\big)
$$

and, by (18),

$$
\begin{aligned}
&\big|[\widehat{Q(t,y)}](\boldsymbol{\xi}) - [\widehat{Q(s,y)}](\boldsymbol{\xi})\big| = \left| \int_s^t \left(\int L^{\mathbf{y}} e_{\boldsymbol{\xi}}(\mathbf{y}')\, Q(\tau, y, d\mathbf{y}') \right) d\tau \right| \\
&\le (t-s)\Big(\tfrac{1}{2}\big(\boldsymbol{\xi}, a(\mathbf{y})\boldsymbol{\xi}\big)_{\mathbb{R}^N} + |\boldsymbol{\xi}||b(\mathbf{y})|\Big) \le \tfrac{1}{2}(1+\lambda)(1+|\mathbf{y}|^2)(1+|\boldsymbol{\xi}|^2)(t-s),
\end{aligned}
$$

where $e_{\boldsymbol{\xi}}(\mathbf{y}) \equiv e^{i(\boldsymbol{\xi}, \mathbf{y})_{\mathbb{R}^N}}$. Hence, by (24),

$$
\big|f_n(t, \boldsymbol{\xi}) - f_n(s, \boldsymbol{\xi})\big| \le \frac{(1+\lambda)(1+|\boldsymbol{\xi}|^2)}{2} e^{(1+\lambda)t} \int (1+|\mathbf{x}|^2)\, \nu(d\mathbf{x})(t-s),
$$

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 $\int |\mathbf{x}|^2 \, \nu(d\mathbf{x}) < \infty$. Indeed, because we know then that $\{\mu_n : n \geq 0\}$ is relatively compact in $C\big([0,\infty); \mathbf{M}_1(\mathbb{R}^N)\big)$, all that we have to do is show that every limit satisfies (15). For this purpose, first note that, by (18),

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

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

$$\langle L^{\mathbf{y}}\varphi, Q(\tau - [\tau]_n, \mathbf{y})\rangle \longrightarrow L\varphi(\mathbf{y})$$

boundedly and uniformly for $(\tau, \mathbf{y})$ in compacts. Hence, if

$$\mu_{n_k} \longrightarrow \mu \text{ in } C\big([0,\infty); \mathbf{M}_1(\mathbb{R}^N)\big),$$

then, by Theorem 17,

$$\langle \varphi, \mu_{n_k}(t)\rangle \longrightarrow \langle \varphi, \mu(t)\rangle \quad \text{and}$$
$$\int_0^t \left(\int \langle L^{\mathbf{y}}\varphi, Q(\tau - [\tau]_n, \mathbf{y})\rangle \mu_n([\tau]_n, d\mathbf{y}) \right) d\tau \longrightarrow \int_0^t \langle L\varphi, \mu(\tau)\rangle \, d\tau.$$

Before moving on, I want to show that $\int |\mathbf{x}|^2 \, \nu(d\mathbf{x}) < \infty$ 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

$$(*) \qquad \int (1 + |\mathbf{y}|^2) \, \mu(t, d\mathbf{y}) \leq e^{(1+\lambda)t} \int (1 + |\mathbf{y}|^2) \, \nu(d\mathbf{y}).$$

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

$$\frac{|\varphi(\mathbf{y})|}{1 + |\mathbf{y}|^2} \vee \frac{|\nabla\varphi(\mathbf{y})|}{1 + |\mathbf{y}|} \vee \|\nabla^2\varphi(\mathbf{y})\|_{\mathrm{H.S.}}$$

is bounded independent of $\mathbf{y} \in \mathbb{R}^N$, and therefore so is $\frac{|L\varphi(\mathbf{y})|}{1+|\mathbf{y}|^2}$. Thus, by (*), there is no problem about integrability of the expressions in (15). Moreover, because (15) holds for each φ_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 $\|\sigma\|_{\mathrm{H.S.}}$ to denote the Hilbert–Schmidt norm $\sqrt{\sum_{ij} \sigma_{ij}^2}$ of σ.

The first of these is an immediate application of Lebesgue's Dominated Convergence Theorem. To prove the second, observe that

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

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

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

$$\begin{aligned}
&\int (1+|\mathbf{y}|^2)\,\mu(t,d\mathbf{y}) \\
&\quad = \int (1+|\mathbf{y}|^2)\,\nu(d\mathbf{y}) + \int_0^t \left(\int \Big[\mathrm{Trace}\big(a(\mathbf{y})\big) + 2\big(\mathbf{y}, b(\mathbf{y})\big)_{\mathbb{R}^N}\Big]\,\mu(\tau,d\mathbf{y})\right) d\tau \\
&\quad \le \int (1+|\mathbf{y}|^2)\,\nu(d\mathbf{y}) + \Lambda \int_0^t \left(\int (1+|\mathbf{y}|^2)\,\mu(\tau,d\mathbf{y})\right) d\tau,
\end{aligned}$$

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

Continuing with the assumption that $\int |\mathbf{x}|^2\,\nu(d\mathbf{x}) < \infty$, I want to remove the boundedness assumption on a and b and replace it by (14). To do this, take ψ_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 $t \rightsquigarrow \mu_k(t)$ so that (16) is satisfied and (15) holds when μ and L are replaced there by μ_k and L_k. Because of (16), the argument which was used earlier can be repeated to show that $\{\mu_k : k \ge 1\}$ is relatively compact in $C\big([0,\infty);\mathbf{M}_1(\mathbb{R}^N)\big)$. Moreover, if μ is any limit of $\{\mu_k : k \ge 1\}$, then (16) is satisfied and, just as we did above, one can check (15), first for $\varphi \in C^2_{\mathrm{c}}(\mathbb{R}^N;\mathbb{C})$ and then for all $\varphi \in C^2(\mathbb{R}^N;\mathbb{C})$ with bounded second order derivatives.

Finally, to remove the second moment condition on ν, assume that it fails, and choose $r_k \nearrow \infty$ so that

$$\alpha_1 \equiv \nu\big(B(0,r_1)\big) > 0 \text{ and } \alpha_k \equiv \nu\big(B(0,r_k) \setminus \overline{B(0,r_{k-1})}\big) > 0 \text{ for each } k \ge 2,$$

and set $\nu_1 = \alpha_1^{-1}\nu \restriction B(0,r_1)$ and $\nu_k = \alpha_k^{-1}\nu \restriction B(0,r_k) \setminus \overline{B(0,r_{k-1})}$ when $k \ge 2$. Finally, choose $t \rightsquigarrow \mu_k(t)$ for L and ν_k, and define $\mu(t) = \sum_{k=1}^{\infty} \alpha_k \mu_k(t)$. It is an easy matter to check that this μ satisfies (15) for all $\varphi \in C^2_{\mathrm{c}}(\mathbb{R}^N;\mathbb{C})$.

Remark 25. In order to put the result in Theorem 13 into a partial differential equations context, it is best to think of $t \rightsquigarrow \mu(t)$ as a solution to $\partial_t\mu = L^\top\mu(t)$ 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 $L^\top\mu(t)$. 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 $S(\mathbf{x})$ denote the set of maps $t \rightsquigarrow \mu(t)$ satisfying (15) with $\nu = \delta_x$. In order to construct a transition probability function, one must make a measurable "selection" $x \in \mathbb{R}^N \longmapsto P(\,\cdot\,,\mathbf{x}) \in S(\mathbf{x})$ 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 $\mathbf{x} \in \mathbb{R}^N \longmapsto P(\,\cdot\,,\mathbf{x}) \in S(\mathbf{x})$ 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 $\mathbf{x} \in \mathbb{R}^N$, $S(\mathbf{x})$ is a compact, convex subset of $\mathbf{M}_1(\mathbb{R}^N)$ and that, if $\mathbf{y} \to \mathbf{x}$ in $\mathbb{R}^N$, then $\overline{\lim}_{\mathbf{y}\to\mathbf{x}} S(\mathbf{y}) \subseteq S(\mathbf{x})$ 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 $t \rightsquigarrow P_n(t,\mathbf{x})$ be constructed by the prescription in (20) with $\nu = \delta_x$. Then, as we showed, $\{P_n(\,\cdot\,,\mathbf{x}) : n \geq 0\}$ is relatively compact in $C_{\mathrm{b}}\big([0,\infty);\mathbf{M}_1(\mathbb{R}^N)\big)$ and every limit point is a solution to (15) with $\nu = \delta_{\mathbf{x}}$. Now suppose that, for each $\mathbf{x}$, $\{P_n(\,\cdot\,,\mathbf{x}) : n \geq 0\}$ is convergent, and let $P(\,\cdot\,,\mathbf{x})$ be its limit. It would follow that $(t,\mathbf{x}) \rightsquigarrow P(t,\mathbf{x})$ has to be a continuous transition probability function which, for each $\mathbf{x}$, would solve (15). To check the Chapman–Kolmogorov equation, note that, by construction, for any $n \geq 0$, $\varphi \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$, and $t \in [0,\infty)$,

$$\langle \varphi, P_n(s+t,\mathbf{x})\rangle = \int \langle \varphi, P_n(t,\mathbf{y})\rangle\, P_n(s,\mathbf{x},d\mathbf{y})$$

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

$$\langle \varphi, P(s+t,\mathbf{x})\rangle = \int \langle \varphi, P(t,\mathbf{y})\rangle\, P(s,\mathbf{x},d\mathbf{y})$$

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 $a : \mathbb{R}^N \longrightarrow \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$ and $b : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ be given, and define L accordingly, as in (12). Further, assume that there exists a square root*

$\sigma : \mathbb{R}^N \longrightarrow \mathrm{Hom}(\mathbb{R}^M;\mathbb{R}^N)$ of a such that

$$\sup_{x\neq x'} \frac{\|\sigma(\mathbf{x})-\sigma(x')\|_{\mathrm{H.S.}} \vee |b(\mathbf{x})-b(x')|}{|x-x'|} < \infty. \tag{27}$$

Then there exists a continuous transition probability function

$$(t,\mathbf{x}) \in [0,\infty)\times\mathbb{R}^N \longmapsto P(t,\mathbf{x}) \in \mathbf{M}_1(\mathbb{R}^N)$$

to which the sequence $\{P_n : n \geq 0\}$ in the preceding discussion converges in $C\big([0,\infty)\times\mathbb{R}^N;\mathbf{M}_1(\mathbb{R}^N)\big)$. In particular, $(t,\mathbf{x}) \in [0,\infty)\times\mathbb{R}^N \longmapsto P(t,\mathbf{x}) \in \mathbf{M}_1(\mathbb{R}^N)$ is a transition probability function with the property that, for each $\mathbf{x} \in \mathbb{R}^N$, $t \rightsquigarrow P(t,\mathbf{x})$ solves Kolmogorov's forward equation for any $\varphi \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$.

In order to prove this theorem, we must learn how to estimate the difference between $P_{n_1}(t,\mathbf{x})$ and $P_{n_2}(t,\mathbf{x})$ 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 $(\mu_1,\mu_2) \in \mathbf{M}_1(\mathbb{R}^N)^2$. Then a coupling of μ_1 to μ_2 is any $\tilde{\mu} \in \mathbf{M}_1(\mathbb{R}^N\times\mathbb{R}^N)$ for which μ_1 and μ_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 μ_1 to μ_2, an estimate of their difference is given by

$$\big|\langle\varphi,\mu_2\rangle - \langle\varphi,\mu_1\rangle\big| \leq \mathrm{Lip}(\varphi)\left(\int_{\mathbb{R}^N} |\mathbf{y}-\mathbf{y}'|^2\,\tilde{\mu}(d\mathbf{y}\times d\mathbf{y}')\right)^{\frac{1}{2}}.$$

Equivalently, a coupling of μ_1 to μ_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 μ_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 $\mathbb{P} = (\gamma_{\mathbf{0},\mathbf{I}})^{\mathbb{Z}^+}$ on $(\mathbb{R}^M)^{\mathbb{Z}^+}$, and given $\mathbf{x} \in \mathbb{R}^N$ and $n \geq 1$, set $Z_n(t,\mathbf{x},\boldsymbol{\omega}) = \begin{pmatrix} X_n(t,\mathbf{x},\boldsymbol{\omega}) \\ Y_n(t,\mathbf{x},\boldsymbol{\omega}) \end{pmatrix}$, where $X_n(0,\mathbf{x},\boldsymbol{\omega}) = Y_n(0,\mathbf{x},\boldsymbol{\omega}) = \mathbf{x}$, and, for $m \geq 0$ and $m2^{-n} < t \leq (m+1)2^{-n}$,

$$\begin{aligned} X_n(t,\mathbf{x},\boldsymbol{\omega}) = X_n(m2^{-n},\mathbf{x},\boldsymbol{\omega}) &+ (t-m2^{-n})^{\frac{1}{2}}\sigma\big(X_n(m2^{-n})\big)\omega_{m+1} \\ &+ (t-m2^{-n})b\big(X_n(m2^{-n})\big) \end{aligned}$$

if m is even,

$$\begin{aligned} X_n(t,\mathbf{x},\boldsymbol{\omega}) = X_n(m2^{-n},\mathbf{x},\boldsymbol{\omega}) &+ (t-m2^{-n})^{\frac{1}{2}}\sigma\big(X_n((m-1)2^{-n})\big)\omega_{m+1} \\ &+ (t-m2^{-n})b\big(X_n((m-1)2^{-n})\big) \end{aligned}$$

if m is odd, and

$$Y_n(t,\mathbf{x},\boldsymbol{\omega}) = Y_n(m2^{-n},\mathbf{x},\boldsymbol{\omega}) + (t-m2^{-n})^{\frac{1}{2}}\sigma\big(Y_n(m2^{-n})\big)\omega_{m+1} + (t-m2^{-n})b\big(Y_n(m2^{-n})\big).$$

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

We will begin by showing that for each $T>0$ there is a $C(T)<\infty$ such that

$$\sup_{n\ge 0}\mathbb{E}^{\mathbb{P}}\big[|Z_n(t,\mathbf{x})-Z_n(s,\mathbf{x})|^2\big] \le C(T)(1+|\mathbf{x}|^2)(t-s) \quad \text{for } 0\le s<t\le T. \tag{28}$$

Set

$$\Sigma_{m,n}(\boldsymbol{\omega}) = \begin{pmatrix}\sigma\big(X_n(m2^{-n},\mathbf{x})\big)\\ \sigma\big(Y_n(m2^{-n},\mathbf{x})\big)\end{pmatrix} \text{ and } B_{m,n} = \begin{pmatrix}b\big(X_n(m2^{-n},\mathbf{x})\big)\\ b\big(Y_n(m2^{-n},\mathbf{x})\big)\end{pmatrix}$$

if m is even, and

$$\Sigma_{m,n}(\boldsymbol{\omega}) = \begin{pmatrix}\sigma\big(X_n((m-1)2^{-n},\mathbf{x})\big)\\ \sigma\big(Y_n(m2^{-n},\mathbf{x})\big)\end{pmatrix} \text{ and } B_{m,n} = \begin{pmatrix}b\big(X_n((m-1)2^{-n},\mathbf{x})\big)\\ b\big(Y_n(m2^{-n},\mathbf{x})\big)\end{pmatrix}$$

if m is odd. If $m2^{-n} < t \le (m+1)2^{-n}$, then

$$|Z_n(t,\mathbf{x})-Z_n(m2^{-n},\mathbf{x})|^2 \le 2(t-m2^{-n})|\Sigma_{m,n}\omega_{m+1}|^2 + 2(t-m2^{-n})^2|B_{m,n}|^2,$$

and therefore, since ω_{m+1} is independent of $\Sigma_{m,n}$,

$$\mathbb{E}^{\mathbb{P}}\big[|Z_n(t,\mathbf{x})-Z_n(m2^{-n},\mathbf{x})|^2\big] \le 2(t-m2^{-n})\mathbb{E}^{\mathbb{P}}\big[\|\Sigma_{m,n}\|^2_{\mathrm{H.S.}}\big] + 2(t-m2^{-n})^2\mathbb{E}^{\mathbb{P}}\big[|B_{m,n}|^2\big].$$

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

$$\mathbb{E}^{\mathbb{P}}\big[|Z_n(t,\mathbf{x})-Z_n(m2^{-n},\mathbf{x})|^2\big] \le 4C(t-m2^{-n})e^{(1+\lambda)t}(1+|\mathbf{x}|^2).$$

Next suppose that $m_1 < m_2$. Then

$$|Z_n(m_2 2^{-n},\mathbf{x})-Z_n(m_1 2^{-n},\mathbf{x})|^2 \le 2^{-n+1}\left|\sum_{m=m_1}^{m_2-1}\Sigma_{m,n}\omega_{m+1}\right|^2 + 2^{-2n+1}\left|\sum_{m=m_1}^{m_2-1}B_{m,n}\right|^2.$$

By taking advantage of the independence of $\Sigma_{m,n}$ and ω_m from $\omega_{m'}$ for $m' > m$, one sees that

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

At the same time,

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

Hence,

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\big[|Z_n(m_2 2^{-n},\mathbf{x}) - Z_n(m_1 2^{-n},\mathbf{x})|^2\big]\\ &\qquad\qquad \le 2C\big((m_2-m_1)2^{-n} + (m_2-m_1)^2 2^{-2n}\big)e^{(1+\lambda)m_2 2^{-n}}(1+|\mathbf{x}|^2),\end{aligned}$$

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

Now set $\Delta_{m,n} = Y_n(m2^{-n},\mathbf{x}) - X_n(m2^{-n},\mathbf{x})$. Then

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

where

$$A_{m,n} = \sigma\big(Y_n(m2^{-n})\big) - \sigma\big(X_n(m2^{-n})\big) \text{ and } D_{m,n} = b\big(Y_n(m2^{-n})\big) - b\big(X_n(m2^{-n})\big).$$

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

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

Just as before,

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

and

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

Thus,

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

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

$$\|A_{k,n}\|_{\mathrm{H.S.}}^2+|D_{k,n}|^2 \le C|\Delta_{k,n}|^2$$

when k is even and

$$\|A_{k,n}\|_{\mathrm{H.S.}}^2+|D_{k,n}|^2 \le C\big(|\Delta_{k,n}|^2+|X_n(k2^{-n},\mathbf{x})-X_n\big((k-1)2^{-n}\big)|^2\big)$$

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

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

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

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[|\Delta_{m,n}|^2\big] &\le 2^{-n}K(m2^{-n})(1+|\mathbf{x}|^2)e^{(1+\lambda)m2^{-n}}\left(1+\frac{K(m2^{-n})}{2^n}\right)^{m-1}\\ &\le 2^{-n}K(m2^{-n})(1+|\mathbf{x}|^2)e^{(1+\lambda+K(m2^{-n}))m2^{-n}}.\end{aligned}$$

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

$$\mathbb{E}^{\mathbb{P}}\big[|Y_n(m2^{-n},\mathbf{x})-X_n(m2^{-n},\mathbf{x})|^2\big]^{\frac12} \le C(R)2^{-\frac{n}{2}}$$

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

$$(*)\qquad \sup_{t\vee|\mathbf{x}|\le R}\mathbb{E}^{\mathbb{P}}\big[|Y_n(t,\mathbf{x})-X_n(t,\mathbf{x})|^2\big]^{\frac12} \le C(R)2^{-\frac{n}{2}}.$$

for some $C(R)<\infty$. Given (*) and the fact that, for each $\mathbf{x}$, $P(\,\cdot\,,\mathbf{x})$ is a limit point of $\{P_n(\,\cdot\,,\mathbf{x}):\,n\ge1\}$ in $C\big([0,\infty);\mathbf{M}_1(\mathbb{R}^N)\big)$, it follows that

$$\sup_{t\vee|\mathbf{x}|\le R}\big|\langle\varphi,P(t,\mathbf{x})\rangle-\langle\varphi,P_n(t,\mathbf{x})\rangle\big| \le 2C(R)\|\varphi\|_{\mathrm{Lip}}2^{-\frac{n}{2}},$$

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

$$P_n \longrightarrow P \text{ in } C\big([0,\infty)\times\mathbb{R}^N,\mathbf{M}_1(\mathbb{R}^N)\big).$$

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 reguire our knowing that $\mathbf{x}\rightsquigarrow\mathbf{P}_t\varphi(\mathbf{x})$ 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 σ for which $a=\sigma\sigma^\top$, and the choice of σ can be critical since one would like σ to have the same smoothness as a. From that point of view, when L is **uniformly elliptic** (i.e., $a\ge\epsilon\mathbf{I}$ for some $\epsilon>0$), the following lemma shows that it is hard to do better than take σ to be its positive definite square root $a^{\frac12}$. In its statement and proof, I use the natation

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

for $\boldsymbol{\alpha}\in\mathbb{N}^N$.

Lemma 29. *Assume that $a\ge\epsilon\mathbf{I}$. If a is continuously differentiable in a neighborhood of $\mathbf{x}$, then so is $a^{\frac12}$ and*

$$\max_{1\le i\le n}\left\|\partial_{x_i}a^{\frac12}(\mathbf{x})\right\|_{\mathrm{op}}\le\frac{\|\partial_{x_i}a(\mathbf{x})\|_{\mathrm{op}}}{2\epsilon^{\frac12}}.$$

Moreover, if $n\ge1$, then there is a $C_n<\infty$ such that

$$\max_{\|\boldsymbol{\alpha}\|=n}\left\|\partial^{\boldsymbol{\alpha}}a(\mathbf{x})\right\|_{\mathrm{op}}\le C_n\epsilon^{\frac12}\sum_{k=1}^{n-1}\left(\frac{\max_{\|\boldsymbol{\alpha}\|\le n}\left\|\partial^{\boldsymbol{\alpha}}a(\mathbf{x})\right\|_{\mathrm{op}}}{\epsilon}\right)^k$$

when a is n-times continuously differentiable in a neighborhood of $\mathbf{x}$. Hence, if $a\in C^n_{\mathrm{b}}\big(\mathbb{R}^N;\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)\big)$, then so is $a^{\frac12}$.

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

Set $d=\mathbf{I}-\frac{a}{\Lambda}$. Obviously d is symmetric, $0\mathbf{I}\le d\le\big(1-\frac{\epsilon}{\Lambda}\big)\mathbf{I}$, and $a=\Lambda(\mathbf{I}-d)$ Thus, if $\binom{\frac12}{0}=1$ and

$$\binom{\frac12}{m}=\frac{\prod_{\ell=1}^{m-1}\left(\frac12-m\right)}{m!}$$

are the coefficients in the Taylor expansion of $x\rightsquigarrow(1+x)^{\frac12}$ around 0, then

$$\sum_{m=0}^\infty(-1)^m\binom{\frac12}{m}d^m$$

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

$$\left(\Lambda^{\frac12}\sum_{m=0}^\infty(-1)^m\binom{\frac12}{m}d^m(\mathbf{y})\right)\boldsymbol{\xi}=\lambda^{\frac12}\boldsymbol{\xi}.$$

Hence,

$$(*)\qquad a^{\frac{1}{2}} = \Lambda^{\frac{1}{2}} \sum_{m=0}^{\infty} (-1)^m \binom{\frac{1}{2}}{m} d^m.$$

Now assume that a is continuously differentiable in a neighborhood of $\mathbf{0}$. Then it is easy to see from (*) that

$$\partial_{x_i} a^{\frac{1}{2}}(\mathbf{0}) = -\Lambda^{-\frac{1}{2}} \left(\sum_{m=1}^{\infty} m(-1)^m \binom{\frac{1}{2}}{d}^{m-1} (\mathbf{0}) \right) \partial_{x_i} a(\mathbf{0}),$$

where again the series converges in the operator norm. Furthermore, because $(-1)^m \binom{\frac{1}{2}}{m} \geq 0$ for all $m \geq 0$ and $d(\mathbf{0}) \leq 1 - \frac{\epsilon}{\Lambda}$,

$$\left\| \sum_{m=1}^{\infty} m(-1)^m \binom{\frac{1}{2}}{m} d^{m-1}(\mathbf{0}) \right\|_{\text{op}}$$
$$le \sum_{m=1}^{\infty} m(-1)^m \binom{\frac{1}{2}}{m} \|d(\mathbf{0})\|_{\text{op}}^{m-1} = \tfrac{1}{2}\big(1 - \|d(\mathbf{0})\|_{\text{op}}\big)^{-\frac{1}{2}} = \frac{\Lambda^{\frac{1}{2}}}{2\epsilon^{\frac{1}{2}}},$$

and so the first assertion is now proved.

Turning to the second assertion, one again uses (*) to see that if $\|\boldsymbol{\alpha}\| = n$ then

$$\partial^{\boldsymbol{\alpha}} a(\mathbf{0}) = \sum_{k=1}^{n} (-1)^k \Lambda^{\frac{1}{2}-k} \left(\sum_{m=k}^{\infty} \frac{m!}{(m-k)!} (-1)^m \binom{\frac{1}{2}}{m} d^{m-k}(\mathbf{0}) \right)$$
$$\times \left(\sum_{\boldsymbol{\alpha}_1 + \cdots + \alpha_k = \boldsymbol{\alpha}} \partial^{\boldsymbol{\alpha}_1} a(\mathbf{0}) \cdots \partial^{\boldsymbol{\alpha}_k} a(\mathbf{0}) \right).$$

Proceeding as above, one see that

$$\left\| \sum_{m=k}^{\infty} \frac{m!}{(m-k)!} (-1)^m \binom{\frac{1}{2}}{m} d^{m-k}(\mathbf{0}) \right\|_{\text{op}} \leq \left| \binom{\frac{1}{2}}{k} \right| \left(\frac{\Lambda}{\epsilon} \right)^{k-\frac{1}{2}},$$

and from here it is easy to complete the proof. □

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 σ. 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

$$a(\mathbf{x}) = \begin{pmatrix} 1 & 0 \\ 0 & |\mathbf{x}|^2 \end{pmatrix} \quad \text{for } \mathbf{x} \in \mathbb{R}^2.$$

Again $a^{\frac12}$ is Lipschitz continuous but not differentiable. On the other hand, if

$$\sigma(\mathbf{x}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & x_1 & x_2 \end{pmatrix},$$

then $a = \sigma\sigma^\top$ and σ is smooth. However, it can be shown that in general there is no smooth choice of σ 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 smaooth σ. 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*

$$\|a^{\frac12}(\mathbf{y}) - a^{\frac12}(\mathbf{x})\|_{\mathrm{H.S.}} \le N\sqrt{2^{-1}K}|\mathbf{y}-\mathbf{x}|.$$

PROOF: The proof turns on a simple fact about functions $f : \mathbb{R} \longrightarrow [0,\infty)$ that have two continuous derivatives and whose second derivative is bounded. Namely,

$$(*) \qquad |f'(0)| \le \sqrt{2\|f''\|_{\mathrm{u}} f(0)}.$$

To prove this simply use Taylor's theorem to write

$$0 \le f(h) \le f(0) + hf'(0) + \frac{h^2}{2}\|f''\|_{\mathrm{u}}$$

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 $a+\epsilon\mathbf{I}$ and then let $\epsilon \searrow 0$. Assuming this uniform positivity, we know that $a^{\frac12}$ has two continuous derivatives, and we need to show is that $|\partial_{x_k} a^{\frac12}_{ij}| \le \sqrt{2^{-1}K}$. For this purpose, let $\mathbf{x}$ be given, and, without loss in generality, assume that $a(\mathbf{x})$ is diagonal. Then, because $a = a^{\frac12}a^{\frac12}$,

$$\partial_{x_k} a_{ij}(\mathbf{x}) = \partial_{x_k} a^{\frac12}_{ij}(\mathbf{x})\big(\sqrt{a_{ii}(\mathbf{x})} + \sqrt{a_{jj}(\mathbf{x})}\big) \ge \partial_{x_k} a^{\frac12}_{ij}(\mathbf{x})\sqrt{a_{ii}(\mathbf{x}) + a_{jj}(\mathbf{x})},$$

and so

$$\big|\partial_{x_k} a^{\frac12}_{ij}(\mathbf{x})\big| \le \frac{\big|\partial_{x_k} a_{ij}(\mathbf{x})\big|}{\sqrt{a_{ii}(\mathbf{x}) + a_{jj}(\mathbf{x})}}.$$

When $i=j$, apply (*) to $f(h)=a(\mathbf{x}+h\mathbf{e}_k)$, and conclude that

$$\left|\partial_{x_k} a^{\frac12}(\mathbf{x})\right| \le \sqrt{2Ka_{ii}(\mathbf{x})},$$

which means that

$$\left|\partial_{x_k} a_{ii}^{\frac12}(\mathbf{x})\right| \le \sqrt{2^{-1}Ka_{ii}(\mathbf{x})}.$$

When $i\ne j$, set

$$f_\pm(h) = a_{ii}(\mathbf{x}+h\mathbf{e}_k) \pm 2a_{ij}(\mathbf{x}+h\mathbf{e}_k) + a_{jj}(\mathbf{x}+h\mathbf{e}_k).$$

Then, by (*),

$$\left|\partial_{x_k} a_{ij}(\mathbf{x})\right| \le \frac{|f'_+(0)|+|f'_-(0)|}{4} \le \sqrt{2^{-1}K\big(a_{ii}(\mathbf{x})+a_{jj}(\mathbf{x})\big)},$$

and so

$$\left|\partial_{x_k} a_{ij}^{\frac12}(\mathbf{x})\right| \le \sqrt{2^{-1}K}. \quad \square$$

Euler's Approach to Itô Integral Equations

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

Given $\sigma:\mathbb{R}^N\longrightarrow \mathrm{Hom}(\mathbb{R}^M;\mathbb{R}^N)$ and $b:\mathbb{R}^N\longrightarrow\mathbb{R}^N$ satisfying (27), define $X_n(t,\mathbf{x})$ for $n\ge 0$ and $(t,\mathbf{x})\in[0,\infty)\times\mathbb{R}^N$ so that $X_n(0,\mathbf{x})=\mathbf{x}$ and

$$X_n(t,\mathbf{x}) = X_n([t]_n,\mathbf{x}) + \sigma\big(X_n([t]_n,\mathbf{x})\big)\big(w(t)-w([t]_n)\big) + b\big(X_n([t]_n,\mathbf{x})\big)(t-[t]_n).$$

Then, for each $(t,\mathbf{x})\in[0,\infty)\times\mathbb{R}^N$ and $n\ge0$, the distribution of $X_n(t,\mathbf{x})$ under $\mathcal{W}$ is the measure $P_n(t,\mathbf{x})$ described in the discussion preceding Theorem 26. In particular, for each $T>0$,

$$\sup_{n\ge 0}\ \sup_{(t,\mathbf{x})\in[0,T]\times\mathbb{R}^N} \frac{\mathbb{E}^{\mathcal{W}}\big[|X_n(t,\mathbf{x})|^2\big]}{1+|\mathbf{x}|^2} < \infty$$

for all $T>0$. In addition, it is clear that $X_n(t,\mathbf{x})$ is W_t-measurable.

Set

$$I_n(t,\mathbf{x}) = \sum_{m<2^n t} \sigma\big(X_n(m2^{-n},\mathbf{x})\big)\big(w((m+1)2^{-n}\wedge t) - w(m2^{-n}\big)$$

and

$$B_n(t,\mathbf{x}) = \sum_{m<2^n t} b\big(X_n(m2^{-n},\mathbf{x})\big)\big(((m+1)2^{-n}\wedge t) - (m2^{-n}\big).$$

Note that if $m2^{-n} \le s < t \le (m+1)2^{-n}$, then, because $w(t) - w(s)$ is independent of W_s,

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

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

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[\big(\boldsymbol{\xi}, I_n(t,\mathbf{x})\big)^2_{\mathbb{R}^N} - \big(\boldsymbol{\xi}, I_n(s,\mathbf{x})\big)^2_{\mathbb{R}^N}\,\big|\,W_s\big]\\
&= \mathbb{E}^{\mathcal{W}}\Big[\Big(\big(\boldsymbol{\xi}, I_n(t,\mathbf{x})\big)_{\mathbb{R}^N} - \big(\boldsymbol{\xi}, I_n(t,\mathbf{x})\big)_{\mathbb{R}^N}\Big)^2\,\Big|\,W_s\Big]\\
&= \sum_{i,j=1}^{N}\big(\sigma(X_n(m2^{-n},\mathbf{x})^\top\boldsymbol{\xi}\big)_i\big(\sigma(X_n(m2^{-n},\mathbf{x})^\top\boldsymbol{\xi}\big)_j\mathbb{E}^{\mathcal{W}}\big[\big(w(t)_i - w(s)_i\big)\big(w(t)_j - w(s)_j\big)\,\big|\,W_s\big]\\
&= \big(\boldsymbol{\xi}, a(X(m2^{-n},\mathbf{x}))\boldsymbol{\xi}\big)_{\mathbb{R}^N}(t-s).
\end{aligned}$$

Thus,

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

is a martingale, and so

$$\begin{aligned}
\mathbb{E}^{\mathcal{W}}\big[|I_n(t,\mathbf{x}) - I_n(s,\mathbf{x})|^2\big] &= \mathbb{E}^{\mathcal{W}}\big[|I_n(t,\mathbf{x})|^2 - |I_n(s,\mathbf{x})|^2\big]\\
&= \mathbb{E}^{\mathcal{W}}\left[\int_s^t \operatorname{Trace}\big(a(X_n([\tau]_n,\mathbf{x}))\big)\,d\tau\right].
\end{aligned}$$

At the same time, $B_n(t,\mathbf{x}) = \int_0^t b([\tau]_n,\mathbf{x})\,d\tau$, and so

$$\mathbb{E}^{\mathcal{W}}\big[|B_n(t,\mathbf{x}) - B_n(s,\mathbf{x})|^2\big] \le (t-s)\int_s^t \mathbb{E}^{\mathcal{W}}\big[|b(X_n([\tau]_n,\mathbf{x}))|^2\big]\,d\tau.$$

Hence, for each $T > 0$,

$$(*)\qquad K(T) \equiv \sup_{n\ge 0}\sup_{0\le s<t\le T}\frac{\mathbb{E}^{\mathcal{W}}\big[|X_n(t,\mathbf{x}) - X_n(s,\mathbf{x})|^2\big]}{(1+|\mathbf{x}|^2)(t-s)} < \infty.$$

Obviously,

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

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

$$\left(|I_n(t,\mathbf{x}) - I_{n-1}(t,\mathbf{x})|^2 - \int_0^t \big\|\sigma\big(X_n([\tau]_n)\big) - \sigma\big(X_{n-1}([\tau]_{n-1})\big\|^2_{\mathrm{H.S.}}\,d\tau, W_t, \mathcal{W}\right)$$

is a continuous martingate. Hence, since $\big(I_n(t,\mathbf{x})-I_{n-1}(t,\mathbf{x}),W_t,\mathcal{W}\big)$ is a continuous martingale, Doob's inequality says that[4]

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}\big[\|I_n(\,\cdot\,,\mathbf{x})-I_{n-1}(\,\cdot\,,\mathbf{x})\|_{[0,T]}^2\big] &\le 4\mathbb{E}^{\mathcal{W}}\big[|I_n(T,\mathbf{x})-I_{n-1}(T,\mathbf{x})|^2\big]\\ &= 4\int_0^T \mathbb{E}^{\mathcal{W}}\big[\|\sigma(X_n([\tau]_n,\mathbf{x}))-\sigma(X_{n-1}([\tau]_{n-1},\mathbf{x}))\|_{\mathrm{H.S.}}^2\big]\,d\tau\\ &\le 8\int_0^T \mathbb{E}^{\mathcal{W}}\big[\|\sigma(X_n([\tau]_n,\mathbf{x}))-\sigma(X_n([\tau]_{n-1},\mathbf{x}))\|_{\mathrm{H.S.}}^2\big]\,d\tau\\ &\quad+8\int_0^T \mathbb{E}^{\mathcal{W}}\big[\|\sigma(X_n([\tau]_{n-1},\mathbf{x}))-\sigma(X_{n-1}([\tau]_{n-1},\mathbf{x}))\|_{\mathrm{H.S.}}^2\big]\,d\tau.\end{aligned}$$

At the same time,

$$\begin{aligned}|B_n(t,\mathbf{x})-B_{n-1}(t,\mathbf{x})|^2 &\le t\int_0^t \big|b\big(X_n([\tau]_n,\mathbf{x})\big)-b\big(X_{n-1}([\tau]_{n-1},\mathbf{x})\big)\big|^2\,d\tau\\ \le 2t\bigg(\int_0^t \big|b\big(X_n([\tau]_n,\mathbf{x})\big)-b\big(X_n([\tau]_{n-1},\mathbf{x})\big)\big|^2\,d\tau\\ &+\int_0^t \big|b\big(X_n([\tau]_{n-1},\mathbf{x})\big)-b\big(X_{n-1}([\tau]_{n-1},\mathbf{x})\big)\big|^2\,d\tau\bigg).\end{aligned}$$

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

$$\begin{aligned}&\mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x})-X_{n-1}(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big]\\ &\quad\le C(T)(1+|\mathbf{x}^)2^{-n}+\lambda\int_0^t \mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x})-X_{n-1}(\,\cdot\,.\mathbf{x})\|_{[0,\tau]}^2\big]\,d\tau\end{aligned}$$

when $t\in[0,T]$. Hence, by Gronwall's lemma,

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

Starting from the preceding, one sees that, for any $m\ge 0$,

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

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

$$w\in C\big([0,\infty);\mathbb{R}^M\big)\longmapsto X(\,\cdot\,,\mathbf{x})\in C\big([0,\infty);\mathbb{R}^N\big)$$

[4] If $\psi:[0,\infty)\longrightarrow\mathbb{R}^N$, then $\|\psi\|_{[s,t]}=\sup\{|\psi(\tau)|:\,\tau\in[s,t]\}$.

and another constant $C(T)<\infty$ such that

(31) $$\mathbb{E}^{\mathcal{W}}\big[\|X(\,\cdot\,,\mathbf{x})-X_n(\,\cdot\,,\mathbf{x})\|_{[0,T]}^2\big]\le 2^{-n}C(T)(1+|\mathbf{x}|^2).$$

Maringales Everywhere: Since, $X_n(t,\mathbf{x})$ and $I_n(t,\mathbf{x})$ are

$$W_t=\sigma\big(\{w(\tau):\,\tau\in[0,t]\}\big)\text{-measurable},$$

$X(t,\mathbf{x})$,

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

are $\overline{W_t}$-measurable.[5] In addition, because $\big(I_n(t,\mathbf{x}),W_t,\mathcal{W}\big)$ and, for each $\boldsymbol{\xi}\in\mathbb{R}^N$,

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

are martingales for all $n\ge0$, it follows from (31) that $\big(I(t,\mathbf{x}),\overline{W)t},\mathcal{W}\big)$ is also a continuous, square integrable martingale and, for each $\boldsymbol{\xi}\in\mathbb{R}^N$, that

$$\left(\big(\boldsymbol{\xi},I(t,\mathbf{x})\big)_{\mathbb{R}^N}^2-\int_0^t\big(\boldsymbol{\xi},a(X(\tau,\mathbf{x}))\boldsymbol{\xi}\big)_{\mathbb{R}^N}\,d\tau,\overline{W_t},\mathcal{W}\right)$$

is a martingale. Using the polarization identity

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

one sees that

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

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

(32) $$\begin{gathered}\big(I(t,\mathbf{x}),\overline{W_t},\mathcal{W}\big)\ \text{and}\\ \left(I(t,\mathbf{x})\otimes I(t,\mathbf{x})-\int_0^t a\big(X(\tau,\mathbf{x})\big)\,d\tau,\overline{W_t},\mathcal{W}\right)\\ \text{are martingales.}\end{gathered}$$

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

$$|I_n(t,\mathbf{x})|\le\|a\|_{\mathrm{u}}^{\frac12}\sum_{m<2^nt}|w(t\wedge(m+1)2^{-n})-w(m2^{-n})|,$$

[5] I use $\overline{W_t}$ here to denote the completion of W_t with respect to $\mathcal{W}$.

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

$$A_n(t,\mathbf{x}) = \int_0^t a\big(X([\tau]_n,\mathbf{x})\big)\,d\tau$$
$$\text{and } E_{\boldsymbol{\zeta},n}(t,\mathbf{x}) = \exp\Big(\big(\boldsymbol{\zeta}, I_n(t,\mathbf{x})\big)_{\mathbb{R}^N} - \tfrac{1}{2}\big(\boldsymbol{\zeta}, A_n(t,\mathbf{x})\boldsymbol{\zeta}\big)_{\mathbb{R}^N}\Big).$$

Our first goal is to show that $\big(E_{\boldsymbol{\zeta},n}(t,\mathbf{x}), W_t, \mathcal{W}\big)$ is a martingale when $|||a|||_{\mathrm{u}} < \infty$, and the arguement is very similar to the one given to prove the martingale property for $I_n(t,\mathbf{x})$. 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

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}\big[E_{\boldsymbol{\zeta},n}(t,\mathbf{x})\,\big|\,W_s\big] = E_{\boldsymbol{\zeta},n}(s,\mathbf{x})\mathbb{E}^{\mathcal{W}}\Big[&\exp\Big(\big(\boldsymbol{\zeta}, \sigma(X_n(m2^{-n},\mathbf{x}))(w(t)-w(s)\big)_{\mathbb{R}^N} \\ &- \tfrac{t-s}{2}\big(\boldsymbol{\zeta}, a(X(m2^{-n},\mathbf{x}))\boldsymbol{\zeta}\big)_{\mathbb{R}^N}\Big)\,\Big|\,W_s\Big] \\ = E_{\boldsymbol{\zeta},n}(s,\mathbf{x}),&\end{aligned}$$

from which the asserted martingale property follows immediately.

As a consequence of the preceding, we know that

$$\mathbb{E}^{\mathcal{W}}\Big[e^{(\boldsymbol{\xi},I_n(t,\mathbf{x}))_{\mathbb{R}^N} - \frac{t|||a|||_{\mathrm{u}}}{2}|\boldsymbol{\xi}|^2}\Big] \leq 1 \quad \text{for all } \boldsymbol{\xi} \in \mathbb{R}^N,$$

and therefore

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

Hence

$$\mathbb{E}^{\mathcal{W}}\Big[e^{\frac{|I_n(t,\mathbf{x})|^2}{2(1+t|||a|||_{\mathrm{u}})}}\Big] \leq (1+t|||a|||_{\mathrm{u}})^{\frac{N}{2}}.$$

Knowing this, and defining

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

it follows from Fatou's Lemma that

$$\mathbb{E}^{\mathcal{W}}\Big[e^{\frac{|I(t,\mathbf{x})|^2}{2(1+t|||a|||_{\mathrm{u}})}}\Big] \leq (1+t|||a||_{\mathrm{u}})^{\frac{N}{2}}. \tag{33}$$

In addition, if

$$A(t,\mathbf{x}) = \int_0^t a\big(X(\tau,\mathbf{x})\big)\,d\tau$$

and

$$E_{\boldsymbol{\zeta}}(t,\mathbf{x}) = \exp\Big(\big(\boldsymbol{\zeta}, I(t,\mathbf{x})\big)_{\mathbb{R}^N} - \tfrac{1}{2}\big(\boldsymbol{\zeta}, A(t,\mathbf{x})\boldsymbol{\zeta}\big)_{\mathbb{R}^N}\Big),$$

then $\{E_{\boldsymbol{\zeta},n}(t,\mathbf{x}) : n \geq 0\}$ converges to $E_{\boldsymbol{\zeta}}(t,\mathbf{x})$ in $L^1(\mathcal{W};\mathbb{C})$, and so

$$\big(E_{\boldsymbol{\zeta}}(t,\mathbf{x}), \overline{\mathcal{W}_t}, \mathcal{W}\big) \quad \text{is a martingale for each } \boldsymbol{\zeta} \in \mathbb{C}^N. \tag{34}$$

Continuing under the assumption that $\|a\|_{\mathrm{u}} < \infty$, we will now show that

$$\begin{gathered} \left(\varphi\big(M(t,\mathbf{x})\big) - \tfrac{1}{2}\int_0^t \mathrm{Trace}\big(a(X(\tau,\mathbf{x}))\nabla^2\varphi(M(\tau,\mathbf{x})\big)\,d\tau, \overline{\mathcal{W}_t}, \mathcal{W}\right) \\ \text{is a martingale for all } \varphi \in C^2(\mathbb{R}^N;\mathbb{C}) \text{ for which there is a } \lambda \geq 0 \\ \text{such that } \sup_{\mathbf{y}\in\mathbb{R}^N} e^{-\lambda|\mathbf{y}|}\|\nabla^2\varphi(\mathbf{y})\|_{\mathrm{H.S.}} < \infty, \end{gathered} \tag{35}$$

and, under the additional that assumption that $\|b\|_{\mathrm{u}} < \infty$, that

$$\begin{gathered} \left(\varphi\big(X(t,\mathbf{x})\big) - \int_0^t L\varphi\big(X(\tau,\mathbf{x})\big)\,d\tau, \overline{\mathcal{W}_t}, \mathcal{W}\right) \\ \text{is a martingale for all } \varphi \in C^2(\mathbb{R}^N;\mathbb{C}) \text{ for which there is a } \lambda \geq 0 \\ \text{such that } \sup_{\mathbf{y}\in\mathbb{R}^N} e^{-\lambda|\mathbf{y}|}\|\nabla^2\varphi(\mathbf{y})\|_{\mathrm{H.S.}} < \infty, \end{gathered} \tag{36}$$

which is the pathspace statement that the distribution of $X(\,\cdot\,,\mathbf{x})$ satisfies Kolmogorov's forward equation.

To this end, begin by observing that, given (33), standard approximation arguements show that it suffices to handle $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{C})$. Furthermore, if $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{C})$, then

$$\begin{aligned} &(2\pi)^N\left(\varphi\big(X(t,\mathbf{x})\big) - \int_0^t L\varphi\big(X(\tau,\mathbf{x})\big)\,d\tau\right) \\ &= \int_{\mathbb{R}^N}\Bigg(e^{-i(\boldsymbol{\xi},X(t,\mathbf{x})\mathbf{x})_{\mathbb{R}^N}} \\ &\qquad + \int_0^t e^{-i(\boldsymbol{\xi},X(\tau,\mathbf{x}))_{\mathbb{R}^N}}\Big(i(\boldsymbol{\xi}, b(X(\tau,\mathbf{x})))_{\mathbb{R}^N} \\ &\qquad\qquad + \tfrac{1}{2}\big(\boldsymbol{\xi}, a(X(\tau,\mathbf{x}))\boldsymbol{\xi}\big)_{\mathbb{R}^N}\Big)\,d\tau\Bigg)\hat{\varphi}(\boldsymbol{\xi})\,d\boldsymbol{\xi}, \end{aligned}$$

and so it suffices to prove (36) when $\varphi(\mathbf{y}) = e^{(\boldsymbol{\zeta},\mathbf{y})_{\mathbb{R}^N}}$. The same line of reasoning shows that it suffices to prove (35) for $e^{(\boldsymbol{\zeta},\mathbf{y})_{\mathbb{R}^N}}$, 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-σ-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 of $\mathbb{C}$-valued random variables such that $V(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$ and $t \rightsquigarrow V(t)(\omega)$ is a continuous function that has bounded variation $\big(|V|(T)\big)(\omega)$ on $[0,T]$ for each $T > 0$. If*

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

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

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\big[M(t)V(t) - M(s)V(s),\, \Gamma\big] \\
&\quad = \sum_{m=0}^{n-1} \mathbb{E}^{\mathbb{P}}\big[M(\tau_{m+1,n})V(\tau_{m+1,n} - M(\tau_{m,n})V(\tau_{m,n}),\, \Gamma\big] \\
&\quad = \mathbb{E}^{\mathbb{P}}\left[\sum_{m=0}^{n-1} M(\tau_{m+1,n})\big(V(\tau_{m+1,n}) - V(\tau_{m,n})\big),\, \Gamma\right],
\end{aligned}$$

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

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

and therefore that

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

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

$$E_{\boldsymbol{\zeta}}(t,\mathbf{x}) \exp\left(\int_0^t \tfrac{1}{2}\big(\boldsymbol{\zeta}, a(X(\tau,\mathbf{x})\boldsymbol{\zeta}\big)_{\mathbb{R}^N}\, d\tau\right)$$

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

$$e^{(\boldsymbol{\zeta},\mathbf{x})_{\mathbb{R}^N}} E_{\boldsymbol{\zeta}}(t,\mathbf{x}) \exp\left(\int_0^t \Big(\big(\boldsymbol{\zeta}, b(X(t,\mathbf{x})\big)_{\mathbb{R}^N} + \tfrac{1}{2}\big(\boldsymbol{\zeta}, a(X(\tau,\mathbf{x}))\boldsymbol{\zeta}\big)_{\mathbb{R}^N}\Big)\, d\tau\right).$$

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* $\tilde\sigma : \mathbb{R}^N \longrightarrow \mathrm{Hom}(R^M;\mathbb{R}^N)$ *and* $\tilde b : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ *are another pair of Lipschitz continuous function, and define* $\tilde X(t,\mathbf{x})$ *accordingly. If* $\tilde\sigma = \sigma$ *and* $\tilde b = b$ *on an open set* $G \ni \mathbf{x}$ *and* $\zeta = \inf\{t \ge 0 :\ X(t,\mathbf{x}) \notin G\}$, *then*

$$\mathcal{W}\big(X(t\wedge\zeta,\mathbf{x}) = \tilde X(t\wedge\zeta,\mathbf{x}) \text{ for } t \ge 0\big) = 1.$$

PROOF: Set

$$\tilde I(t,\mathbf{x}) = \tilde X(t,\mathbf{x}) - \mathbf{x} - \int_0^t \tilde b\big(\tilde X(\tau,\mathbf{x})\big)\,d\tau.$$

By applying the second part of (32) when σ and b are replaced by $\begin{pmatrix}\sigma\\ \tilde\sigma\end{pmatrix}$ and $\begin{pmatrix}b\\ \tilde b\end{pmatrix}$, one sees that

$$\left(|I(t,\mathbf{x}) - \tilde I(t,\mathbf{x})|^2 - \int_0^t \big\|\sigma(X(\tau,\mathbf{x}) - \tilde\sigma(\tilde X(\tau,\mathbf{x}))\big\|_{\mathrm{H.S.}}^2\,d\tau, \overline{\mathcal{W}_t}, \mathcal{W}\right)$$

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

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[|I(t\wedge\zeta) - \tilde I(t\wedge\zeta,\mathbf{x})|^2\big]\\
&= \mathbb{E}^{\mathcal{W}}\left[\int_0^{t\wedge\zeta} \big\|\tilde\sigma\big(X(\tau,\mathbf{x})\big) - \tilde\sigma\big(\tilde X(\tau,\mathbf{x})\big)\big\|_{\mathrm{H.S.}}^2\,d\tau\right]\\
&\le \int_0^t \mathbb{E}^{\mathcal{W}}\big[\big\|\tilde\sigma\big(X(\tau\wedge\zeta,\mathbf{x})\big) - \tilde\sigma\big(\tilde X(\tau\wedge\zeta,\mathbf{x})\big)\big\|_{\mathrm{H.S.}}^2\,d\tau\\
&\le C\int_0^t \mathbb{E}^{\mathcal{W}}\big[|X(\tau\wedge\zeta,\mathbf{x}) - \tilde X(\tau\wedge\zeta,\mathbf{x})|^2\big]\,d\tau,
\end{aligned}$$

where C is the square of the Lipschitz constant for $\tilde\sigma$. At the same time,

$$\begin{aligned}
|X(t\wedge\zeta,\mathbf{x}) - \tilde X(t\wedge\zeta,\mathbf{x})|^2 \le{}& 2|I(t\wedge\zeta,\mathbf{x}) - \tilde I(t\wedge\zeta,\mathbf{x})|^2\\
&+ 2\int_0^{t\wedge\zeta} \big|\tilde b\big(X(\tau,\mathbf{x})\big) - \tilde b\big(\tilde X(\tau,\mathbf{x})\big)\big|\,d\tau,
\end{aligned}$$

and

$$\int_0^{t\wedge\zeta} \big|b\big(X(\tau,\mathbf{x})\big) - b\big(\tilde X(\tau,\mathbf{x})\big)\big|\,d\tau \le Ct\int_0^t |X(\tau\wedge\zeta) - \tilde X(\tau\wedge\zeta)|^2\,d\tau,$$

where this time C is the square of the Lipschitz constant for $\tilde b$. Hence, after combining these, we see that there is a $C < \infty$ such that

$$\mathbb{E}^{\mathcal{W}}\big[|X(t\wedge\zeta,\mathbf{x}) - \tilde X(t\wedge\zeta,\mathbf{x})|^2\big] \le C(1+t)\int_0^t \mathbb{E}^{\mathcal{W}}\big[|X(t\wedge\zeta,\mathbf{x}) - \tilde X(t\wedge\zeta,\mathbf{x})|^2\big]\,d\tau.$$

Since, by Gronwall's lemma, this means that

$$\mathbb{E}^{\mathcal{W}}\big[|X(t\wedge\zeta,\mathbf{x})-\tilde{X}(t\wedge\zeta,\mathbf{x})|^2\big]=0,$$

it follows that $X(t\wedge\zeta,\mathbf{x})=\tilde{X}(t\wedge\zeta,\mathbf{x})$ (a.s.,$\mathcal{W}$), first for each and then, by continuity, simultaneously for all $t\geq 0$. $\square$

Lemma 38 will allow us to reduce proofs of many results to the case when σ and b are bounded.

LEMMA 39. *Given $p\in[2,\infty)$, set*

$$\lambda_p=\sup_{\mathbf{y}\in\mathbb{R}^N}\frac{\frac{p(p-1)}{2}\mathrm{Trace}\big(a(\mathbf{y})\big)+\big(\mathbf{y},b(\mathbf{y})\big)^+_{\mathbb{R}^N}}{(1+|\mathbf{y}|^p)^{\frac{2}{p}}},$$

$$\lambda_p^0=\sup_{\mathbf{y}\in\mathbb{R}^N}\frac{\frac{p(p-1)}{2}\mathrm{Trace}\big(a(\mathbf{y})\big)\big)^+_{\mathbb{R}^N}}{(1+|\mathbf{y}|^p)^{\frac{2}{p}}},$$

and

$$\beta_p=\sup_{\mathbf{y}\in\mathbb{R}^N}\frac{|b(\mathbf{y})|^p}{(1+|\mathbf{y}|^p)}.$$

Then there exists a constant $C_p<\infty$, depending only on p, λ_p^0, and β_p, such that

$$\mathbb{E}^{\mathcal{W}}\big[\|I(\,\cdot\,,\mathbf{x})-I(s,\mathbf{x})\|^p_{[s,t]}\big]^{\frac{1}{p}}\leq C_pe^{\frac{\lambda_pt}{p}}(1+|\mathbf{x}|)(t-s)^{\frac{1}{2}}$$

and

$$\mathbb{E}^{\mathcal{W}}\big[\|X(\,\cdot\,,\mathbf{x})-X(s,\mathbf{x})\|^p_{[s,t]}\big]^{\frac{1}{p}}\leq C_pe^{\frac{\lambda_pt}{p}}(1+|\mathbf{x}|)(t-s)^{\frac{1}{2}}\vee(t-s)$$

for $0\leq s<t$ and $\mathbf{x}\in\mathbb{R}^N$.

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

Begin with the observation that, if $\varphi_p(\mathbf{y})=|\mathbf{y}|^p$

$$\nabla\varphi_p(\mathbf{y})=p|\mathbf{y}|^{p-2}\mathbf{y}\text{ and }H_p(\mathbf{y})\equiv\nabla^2\varphi_p(\mathbf{y})=|\mathbf{y}|^{p-2}\left(p(p-2)\frac{\mathbf{y}\otimes\mathbf{y}}{|\mathbf{y}|^2}+p\mathbf{I}\right).$$

Hence

$$L\varphi_p=|\mathbf{y}|^{p-2}\left(\frac{p(p-1)\big(\mathbf{y},a(\mathbf{y})\mathbf{y}\big)_{\mathbb{R}^N}}{2|\mathbf{y}|^2}+\frac{p\mathrm{Trace}\big(a(\mathbf{y})\big)}{2}+\big(b(\mathbf{y}),\mathbf{y}\big)_{\mathbb{R}^N}\right)\leq\lambda_p(1+|\mathbf{y}|^p),$$

and therefore, by (36),

$$\mathbb{E}^{\mathcal{W}}\big[1+|X(t,\mathbf{x})|^p\big] \le 1+|\mathbf{x}|^p+\lambda_p\int_0^t \mathbb{E}^{\mathcal{W}}\big[1+|X(\tau,\mathbf{x})|^p\big]\,d\tau.$$

Therefore, by Gronwall's lemma,

$$(*)\qquad \mathbb{E}^{\mathcal{W}}\big[1+|X(t,\mathbf{x})|^p\big] \le (1+|\mathbf{x}|^p)e^{\lambda_p t}.$$

Since

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

it follows that

$$(**)\qquad \mathbb{E}^{\mathcal{W}}\big[\|B(\,\cdot\,,\mathbf{x})-B(s,\mathbf{x})\|_{[s,t]}^p\big] \le \beta_p(1+|\mathbf{x}|^p)(t-s)^p e^{\lambda_p t}.$$

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

$$\left(|I(t,x)-\boldsymbol{\xi}|^p-\tfrac{1}{2}\int_0^t \operatorname{Trace}\big(a(X(\tau,\mathbf{x}))H_p(I(\tau,\mathbf{x})-\boldsymbol{\xi})\big)\,d\tau,\overline{W_t},\mathcal{W}\right)$$

is a martingale, and, as a consequence, for any $s\ge 0$,

$$\left(|I(t\vee s,x)-I(s,\mathbf{x})|^p-\tfrac{1}{2}\int_s^{t\vee s} \operatorname{Trace}\big(a(X(\tau,\mathbf{x}))H_p(I(\tau,\mathbf{x})-I(s,\mathbf{x}))\big)\,d\tau,\overline{W_t},\mathcal{W}\right)$$

is a martingale. In particular, for $0\le s<t$,

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[|I(t,\mathbf{x})-I(s,\mathbf{x})|^p\big] \le \lambda_p^0\int_s^t \mathbb{E}^{\mathcal{W}}\big[(1+|X(\tau,\mathbf{x})|^p)^{\frac{2}{p}}|I(\tau,\mathbf{x})-I(s,\mathbf{x})|^{p-2}\big]\,d\tau\\
&\le \lambda_p^0\left(\int_s^t \mathbb{E}^{\mathcal{W}}\big[(1+|X(\tau,\mathbf{x})|^p)\big]\,d\tau\right)^{\frac{2}{p}}\left(\int_s^t \mathbb{E}^{\mathcal{W}}\big[|I(\tau,\mathbf{x})-I(s,\mathbf{x})|^p\big]\,d\tau\right)^{1-\frac{2}{p}}\\
&\le \lambda_p^0(t-s)e^{\frac{2\lambda_p}{p}t}\mathbb{E}^{\mathcal{W}}\big[|I(t,\mathbf{x})-I(s,x)|^p\big]^{1-\frac{2}{p}},
\end{aligned}$$

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 $I(t\vee s,\mathbf{x})-I(s,\mathbf{x})$ is a martingle, $\mathbb{E}^{\mathcal{W}}\big[|I(t\vee s,\mathbf{x})-I(s,\mathbf{x})|^p\big]$ is a non-decreasing function of t. By combining the preceding with

$$\mathbb{E}^{\mathcal{W}}\big[\|I(\,\cdot\,,\mathbf{x})-I(s,\mathbf{x})\|_{[s,t]}^p\big]^{\frac{1}{p}} \le \frac{p}{p-1}\mathbb{E}^{\mathcal{W}}\big[|I(t,\mathbf{x})-I(s,\mathbf{x})|^p\big]^{\frac{1}{p}},$$

the first estimate follows, and, since $X(t,\mathbf{x})-\mathbf{x}=I(t,\mathbf{x})+B(t,\mathbf{x})$, 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

$$\mathbf{y}^{(R)} = \big((y_1 \wedge R) \vee (-R), \dots, (y_N \wedge R) \vee (-R)\big),$$

and define $\sigma_R(\mathbf{y}) = \sigma(\mathbf{y}^{(R)})$ and $b_R(\mathbf{y}) = b(\mathbf{y}^{(R)})$. Obviously, the quanties λ_p, λ_p^0, and β_p corresponding to $a_R \equiv \sigma_R \sigma_R^\top$ and b_R are dominatied by the original ones, and, by Lemma 38, if $I_R(\,\cdot\,,\mathbf{x})$ and $X_R(\,\cdot\,,\mathbf{x})$ are the stochastic processes determined by σ_R and b_R, then

$$\|I_R(\,\cdot\,,\mathbf{x}) - I_R(s,\mathbf{x})\|_{[s,t]} = \|I(\,\cdot\,,\mathbf{x}) - I(x,\mathbf{x})\|_{[s,t]}$$

and

$$\|X_R(\,\cdot\,,\mathbf{x}) - X_R(s,\mathbf{x})\|_{[s,t]} = \|X(\,\cdot\,,\mathbf{x}) - X(s,\mathbf{x})\|_{[s,t]}$$

if $t \le \zeta_R \equiv \inf\{\tau :\ X(\tau,\mathbf{x}) \in [-R,R]^N\}$. Hence

$$\mathbb{E}^{\mathcal{W}}\big[\|I(\,\cdot\,,\mathbf{x}) - I(s,\mathbf{x})\|_{[s,t]}^p,\, \zeta_R \ge t\big] \le \mathbb{E}^{\mathcal{W}}\big[\|I_R(\,\cdot\,,\mathbf{x}) - I_R(s,\mathbf{x})\|_{[s,t]}^p\big]$$

and

$$\mathbb{E}^{\mathcal{W}}\big[\|X(\,\cdot\,,\mathbf{x}) - X(s,\mathbf{x})\|_{[s,t]}^p,\, \zeta_R \ge t\big] \le \mathbb{E}^{\mathcal{W}}\big[\|X_R(\,\cdot\,,\mathbf{x}) - X_R(s,\mathbf{x})\|_{[s,t]}^p\big].$$

Because $\zeta_R \nearrow \infty$ as $R \to \infty$, this completes the proof. □

Theorem 40. *For any $\varphi \in C^2(\mathbb{R}^N;\mathbb{C})$ whose second derivatives have at most polynomial growth,*

$$\left(\varphi\big(I(t,\mathbf{x})\big) - \tfrac12 \int_0^t \mathrm{Trace}\big(a(X(\tau,\mathbf{x}))\nabla^2\varphi(I(\tau,\mathbf{x})\big)\, d\tau, \overline{\mathcal{W}_t}, \mathcal{W}\right)$$

and

$$\left(\varphi\big(X(t,\mathbf{x})\big) - \int_0^t L\varphi\big(X(\tau,\mathbf{x})\big)\, d\tau, \overline{\mathcal{W}_t}, \mathcal{W}\right)$$

are martingales.

Proof: Determine $X_R(t,\mathbf{x})$, $I_R(t,\mathbf{x})$, and ζ_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

$$\left(\varphi\big(I(t\wedge\zeta_R,\mathbf{x})\big) - \tfrac12 \int_0^{t\wedge\zeta_R} \mathrm{Trace}\big(a(X(\tau,\mathbf{x}))\nabla^2\varphi(I(\tau,\mathbf{x})\big)\, d\tau, \overline{\mathcal{W}_t}, \mathcal{W}\right)$$

and

$$\left(\varphi\big(X(t\wedge\zeta_R,\mathbf{x})\big) - \int_0^{t\wedge\zeta_R} L\varphi\big(X(\tau,\mathbf{x})\big)\, d\tau, \overline{\mathcal{W}_t}, \mathcal{W}\right)$$

are martingales for all $R > 0$. Furthmore, by the estimates in Lemma 39,

$$
\begin{aligned}
&I(t\wedge\zeta_R,\mathbf{x})\big) - \tfrac{1}{2}\int_0^{t\wedge\zeta_R} \operatorname{Trace}\big(a(X(\tau,\mathbf{x}))\nabla^2\varphi(I(\tau,\mathbf{x})\big)\,d\tau \\
&\qquad \longrightarrow I(t,\mathbf{x})\big) - \tfrac{1}{2}\int_0^{t} \operatorname{Trace}\big(a(X(\tau,\mathbf{x}))\nabla^2\varphi(I(\tau,\mathbf{x})\big)\,d\tau
\end{aligned}
$$

and

$$
\begin{aligned}
&\varphi\big(X(t\wedge\zeta_R,\mathbf{x})\big) - \int_0^{t\wedge\zeta_R} L\varphi\big(X(\tau,\mathbf{x})\big)\,d\tau \\
&\qquad \longrightarrow \varphi\big(X(t,\mathbf{x})\big) - \int_0^{t} L\varphi\big(X(\tau,\mathbf{x})\big)\,d\tau
\end{aligned}
$$

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

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 σ and b, such that*

$$
\mathbb{E}^{\mathcal{W}}\big[\|X(\,\cdot\,,\mathbf{x}) - X(\,\cdot\,,\tilde{\mathbf{x}})\|_{[0,T]}^p\big]^{\frac{1}{p}} \le C_p(T)|\mathbf{x}-\tilde{\mathbf{x}}|.
$$

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

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

Then

$$
\nabla\Phi_p(\mathbf{y},\tilde{\mathbf{y}}) = |\mathbf{y}-\tilde{\mathbf{y}}|^{p-1}\begin{pmatrix}\mathbf{y}-\tilde{\mathbf{y}}\\ -(\mathbf{y}-\tilde{\mathbf{y}})\end{pmatrix} \text{ and } \nabla^2\Phi_p(\mathbf{y},\tilde{\mathbf{y}}) = \begin{pmatrix} H_p(\mathbf{y},\tilde{\mathbf{y}}) & -H_p(\mathbf{y},\tilde{\mathbf{y}})\\ -H_p(\mathbf{y},\tilde{\mathbf{y}}) & H_p(\mathbf{y},\tilde{\mathbf{y}})\end{pmatrix}.
$$

Next set

$$
\Sigma(\mathbf{y},\tilde{\mathbf{y}}) = \begin{pmatrix}\sigma(\mathbf{y})\\ \sigma(\tilde{\mathbf{y}})\end{pmatrix} \text{ and } B(\mathbf{y},\tilde{\mathbf{y}}) = \begin{pmatrix} b(\mathbf{y})\\ b(\tilde{\mathbf{y}})\end{pmatrix}.
$$

Then

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

and

$$
\begin{aligned}
&\operatorname{Trace}\big(\Sigma\Sigma^\top(\mathbf{y},\tilde{\mathbf{y}})^\top\nabla^2\Phi_p(\mathbf{z},\tilde{\mathbf{z}})\big) \\
&= p(p-2)|\mathbf{z}-\tilde{\mathbf{z}}|^{p-2}\left(\frac{\big|\big(\sigma(\mathbf{y})-\sigma(\tilde{\mathbf{y}})\big)^\top(\mathbf{z}-\tilde{\mathbf{z}})\big|^2}{|\mathbf{z}-\tilde{\mathbf{z}}|^2} + p\|\sigma(\mathbf{y})-\sigma(\tilde{\mathbf{y}})\|_{\mathrm{H.S.}}^2\right).
\end{aligned}
$$

Hence, since, by Theorem 40,

$$\bigg(\Phi_p\big(I(t,\mathbf{x}),I(t,\tilde{\mathbf{x}})\big) - \tfrac{1}{2}\int_0^t \mathrm{Trace}\big(\Sigma\Sigma^\top(X(\tau,\mathbf{x}),X(\tau,\tilde{\mathbf{x}}))\nabla^2\Phi_p(I(\tau,\mathbf{x}),I(\tau,\tilde{\mathbf{x}})\big)\,d\tau, \overline{\mathcal{W}_t},\mathcal{W}\bigg)$$

and

$$\bigg(\Phi_p\big(X(t,\mathbf{x}),X(t,\tilde{\mathbf{x}})\big) - \int_0^t \Big(\big(\nabla\Phi_p(X(\tau,\mathbf{x}),X(\tau,\tilde{\mathbf{x}})),B(X(\tau,\mathbf{x}),X(\tau,\tilde{\mathbf{x}}))\big)_{\mathbb{R}^{2N}} + \tfrac{1}{2}\mathrm{Trace}\big(\Sigma\Sigma^\top(X(\tau,\mathbf{x}),X(\tau,\tilde{\mathbf{x}}))\nabla^2\Phi_p(X(\tau,\mathbf{x}),X(\tau,\tilde{\mathbf{x}}))\Big)\,d\tau, \overline{\mathcal{W}_t},\mathcal{W}\bigg)$$

are martingales, we see that there is a $K_p<\infty$, depending only on p and the Lipschitz norms of σ and b, such that

$$\mathbb{E}^{\mathcal{W}}\big[|I(t,\mathbf{x})-I(t,\tilde{\mathbf{x}})|^p\big] \le K_p\int_0^t \mathbb{E}^{\mathcal{W}}\big[|I(\tau,\mathbf{x})-I(\tau,\tilde{\mathbf{x}})|^{p-2}|X(\tau,\mathbf{x})-X(\tau,\tilde{\mathbf{x}})|^2\big]\,d\tau$$

and

$$\mathbb{E}^{\mathcal{W}}\big[|X(t,\mathbf{x})-X(t,\tilde{\mathbf{x}})|^p\big] \le |\mathbf{x}-\tilde{\mathbf{x}}|^p + K_p\int_0^t \mathbb{E}^{\mathcal{W}}\big[|X(\tau,\mathbf{x})-X(\tau,\tilde{\mathbf{x}})|^p\big]\,d\tau.$$

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*

$$w\in C\big([0,\infty);\mathbb{R}^M\big)\longmapsto Y(\,\cdot\,,\,\cdot\cdot\,)(w)\in C\big([0,\infty)\times\mathbb{R}:\mathbb{R}^N\big)$$

such that $X(\,\cdot\,,\mathbf{x})=Y(\,\cdot\,,\mathbf{x})$ *(a.s.,* $\mathcal{W}$*) for each* $\mathbf{x}\in\mathbb{R}^N$*. Moreover, for each* $\alpha\in[0,1)$*,* $p\in[1,\infty)$*, and* $R>0$

$$\mathbb{E}^{\mathcal{W}}\left[\left(\sup_{\substack{\mathbf{x},\mathbf{y}\in[-R,R]^N\\ \mathbf{x}\ne\mathbf{y}}}\frac{\|Y(\,\cdot\,,\mathbf{y})-Y(\,\cdot\,,\mathbf{x})\|_{[0,T]}}{|\mathbf{y}-\mathbf{x}|^\alpha}\right)^p\right]<\infty$$

and

$$\mathbb{E}^{\mathcal{W}}\left[\left(\sup_{\substack{(s,\mathbf{x}),(t,\mathbf{y})\in[-R,R]^{N+1}\\(s,\mathbf{x})\neq(t,\mathbf{y})}}\frac{|Y(t,\mathbf{y})-Y(s,\mathbf{x})|}{|(t,\mathbf{y})-(s,\mathbf{x})|^{\frac{\alpha}{2}}}\right)^p\right]<\infty.$$

PROOF: Given the estimates in Corollary 41, the existence of the map $w \rightsquigarrow Y(\,\cdot\,,\,\cdot\cdot\,)(w)$ as well as the estimate on its Hölder continuity as a function of $\mathbf{x}$ are immediate consequences of Kolmogorov's continuity criterion. As for the estimate on its Hölder continuity as a function of $(t,\mathbf{x})$, 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

$$\mathbb{E}^{\mathcal{W}}\big[|X(t,\mathbf{y})-X(s,\mathbf{x})|^p\big]\le C(R)|(t,\mathbf{y})-(s,\mathbf{x})|^{\frac{p}{2}}\text{ for }(s,\mathbf{x}),(t,\mathbf{y})\in[0R]^{N+1},$$

and so another application of Kolmogorov's criterion completes the proof. $\square$

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

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

Before explaining what this means, observe that, as an application of Corollary 42, we know that $\mathbf{x}\rightsquigarrow\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}[\Phi]$ is continuous for any bounded, continuous Φ on $C\big([0,\infty);\mathbb{R}^N\big)$ and therefore that it is Borel measurable for any Borel measurable Φ on $C\big([0,\infty);\mathbb{R}^N\big)$ that is either non-negative or in $\bigcap_{\mathbf{x}\in\mathbb{R}^N}L^1(\mathbb{P}_{\mathbf{x}};\mathbb{C})$. Next let ψ denote a generic element of $C\big([0,\infty);\mathbb{R}^N\big)$ and set $\mathcal{B}_t=\sigma\big(\{\psi(\tau):\tau\in[0,t]\}\big)$. Then a **stopping time** ζ relative to $\{\mathcal{B}_t: t\ge 0\}$ is a measurable map $\zeta: C\big([0,\infty);\mathbb{R}^N\big)\longrightarrow[0,\infty]$ with the property that $\{\zeta\le t\}\equiv\{\psi:\zeta(\psi)\le t\}\in\mathcal{B}_t$ for all $t\ge0$. Associated with a stopping time ζ is the σ-algebra $\mathcal{B}_\zeta$ of sets $\Gamma\subseteq C\big([0,\infty);\mathbb{R}^N\big)$ with the property that $\Gamma\cap\{\zeta\le t\}\in\mathcal{B}_t$ for all $t\ge0$. One can show that $\mathcal{B}_\zeta=\sigma\big(\{\psi(t\wedge\zeta):t\ge0\}\big)$.[6] Next define the time-shift maps $\Sigma_s: C\big([0,\infty);\mathbb{R}^N\big)\longrightarrow C\big([0,\infty);\mathbb{R}^N\big)$ so that $\Sigma_s\psi(t)=\psi(s+t)$. Then the Markov property that we will prove says that for any stopping time ζ and any non-negative, Borel measurable function Φ on $C\big([0,\infty);\mathbb{R}^N\big)^2$,

$$\begin{aligned}&\int_{\{\zeta<\infty\}}\Phi(\psi^\zeta,\Sigma_\zeta\psi)\,\mathbb{P}_{\mathbf{x}}(d\psi)\\&\qquad=\int\left(\int\Phi(\psi_1,\psi_2)\,\mathbb{P}_{\psi_1(\zeta)}(d\psi_2)\right)\mathbb{P}_{\mathbf{x}}(d\psi_1),\end{aligned}\tag{43}$$

[6] See Exercise 7.1.21 in the 2nd edition of my *Probability Theory, An Analytic View.*

where $\psi(\zeta) \equiv \psi\big(\zeta(\psi)\big)$, $\psi^\zeta(t) \equiv \psi(t\wedge\zeta)$, and $\Sigma_\zeta\psi \equiv \Sigma_{\psi(\zeta)}\psi$. In particular,

$$\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\Phi\circ\Sigma_\zeta\,\big|\,\mathcal{B}_\zeta\big](\psi) = \mathbb{E}^{\mathbb{P}_{\psi(\zeta)}}[\Gamma] \text{ for } \mathbb{P}_{\mathbf{x}}\text{-a.e. } \psi\in\{\zeta<\infty\} \tag{44}$$

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

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

$$\begin{aligned}&\int \Phi\big(X_n^\zeta(\,\cdot\,,\mathbf{x})(w), X_n(\,\cdot\,+\zeta,\mathbf{x})(w)\,\mathcal{W}(dw)\\ &= \sum_{m=0}^\infty \int_{\{\zeta=m2^{-\ell}\}} \Phi\big(X_n(\,\cdot\,,\mathbf{x})(w^{m2^{-\ell}}), X_n(\,\cdot\,,X_n(m2^{-\ell},\mathbf{x})(\delta_{m2^{-\ell}}w)\big)\,\mathcal{W}(dw).\end{aligned}$$

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

$$\begin{aligned}&\int_{\{\zeta=m2^{-\ell}\}} \Phi\big(X_n(\,\cdot\,,\mathbf{x})(w^{m2^{-\ell}}), X_n(\,\cdot\,,X_n(m2^{-\ell},\mathbf{x})(\delta_{m2^{-\ell}}w)\big)\,\mathcal{W}(dw)\\ &= \int_{\{\zeta=m2^{-\ell}\}}\bigg(\int \Phi\big(X_n^{m2^{-\ell}}(\,\cdot\,,\mathbf{x})(w_1), X_n(\,\cdot\,,X_n(m2^{-\ell},\mathbf{x})(w_2))\,\mathcal{W}(dw_2)\bigg)\,\mathcal{W}(dw_1),\end{aligned}$$

and so

$$\begin{aligned}&\int \Phi\big(X_n^\zeta(\,\cdot\,,\mathbf{x})(w), X_n(\,\cdot\,+\zeta,\mathbf{x})(w^\zeta)\,\mathcal{W}(dw)\\ &= \int\bigg(\int \Phi\big(X_n^\zeta(\,\cdot\,,\mathbf{x})(w_1), X_n(\,\cdot\,,X_n(\zeta(w_1),\mathbf{x})(w_2))\,\mathcal{W}(dw_2)\bigg)\,\mathcal{W}(dw_1).\end{aligned}$$

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

$$\begin{aligned}&\int \Phi\big(X^\zeta(\,\cdot\,,\mathbf{x})(w), X(\,\cdot\,+\zeta,\mathbf{x})(w^\zeta)\,\mathcal{W}(dw)\\ &= \int\bigg(\int \Phi\big(X^\zeta(\,\cdot\,,\mathbf{x})(w_1), X(\,\cdot\,,X(\zeta(w_1),\mathbf{x})(w_2))\,\mathcal{W}(dw_2)\bigg)\,\mathcal{W}(dw_1)\end{aligned}$$

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

$$\begin{aligned}\int_{\{\zeta<T\}}\Phi(\psi^\zeta,\Sigma_\zeta\psi)\,\mathbb{P}_{\mathbf{x}}(d\psi) &= \int \mathbf{1}_{[0,T)}\big(\zeta(\psi)\wedge T\big)\Phi(\psi^{\zeta\wedge T},\Sigma_{\zeta\wedge T}\psi)\,\mathbb{P}_{\mathbf{x}}(d\psi)\\ &= \int_{\{\zeta<T\}}\left(\int\Phi(\psi_1,\psi_2)\,\mathbb{P}_{\psi_1(\zeta)}(d\psi_2)\right)\mathbb{P}_{\mathbf{x}}(d\psi_1).\end{aligned}$$

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

Characterizing the Distribtuion: We have constructed the measure $\mathbb{P}_{\mathbf{x}}$ and learned something about its properties, but, as yet, we have not characterized it. That is, starting with coefficinets a and b, we have given a construction which led to an associated $\mathbb{P}_{\mathbf{x}}$, 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 $\mathbb{P}_{\mathbf{x}}$ to which our consstruction led has the property that

$$\left(\varphi\big(\psi(t)\big)-\varphi(\mathbf{x})-\int_0^t L\varphi\big(\psi(\tau)\big)\,d\tau,\mathcal{B}_t,\mathbb{P}_{\mathbf{x}}\right) \tag{45}$$
$$\text{is a martingale with mean value 0 for all } \varphi\in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R}),$$

and we will say that such a measure solves the **martingale problem for** L **starting at** $\mathbf{x}$. Notice that if $P(t,\mathbf{x})$ is any transition function that satisfies Kolmogorov's forward equation

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

and if $\mathbb{P}_{\mathbf{x}}$ is a measure on pathspace determined by

$$\mathbb{P}_{\mathbf{x}}\big(\psi(0)=\mathbf{x}\big)=1 \text{ and } \mathbb{P}_{\mathbf{x}}\big(\psi(t)\in\Gamma\,\big|\,\mathcal{B}_s\big)=P\big(t-s,\psi(s),\Gamma\big)\ (\text{a.s.},\mathbb{P}_{\mathbf{x}}),$$

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

$$\begin{aligned}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(t)\big)-\varphi\big(\psi(s)\big)\,\big|\,\mathcal{B}_x\big] &= \big\langle\varphi,P\big(t-s,\psi(s)\big)\big\rangle-\psi\big(\psi(s)\big)\\ &= \int_0^{t-s}\big\langle L\varphi,P\big(\tau,\psi(s)\big)\big\rangle\,d\tau=\int_0^{t-s}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[L\varphi\big(\psi(s+\tau)\,\big|\,\mathcal{B}_s\big]\\ &= \mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\left[\int_s^t L\varphi\big(\psi(\tau)\big)\,d\tau\,\middle|\,\mathcal{B}_s\right],\end{aligned}$$

and so $\mathbb{P}_{\mathbf{x}}$ 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 $X(\,\cdot\,,\mathbf{x})$ is a solution to the O.D.E.

$$(*) \qquad \dot{X}(t,\mathbf{x}) = b\big(X(t,\mathbf{x})\big) \text{ with } X(0,\mathbf{x}) = \mathbf{x},$$

then the measure $\mathbb{P}_{\mathbf{x}} = \delta_{X(\,\cdot\,,\mathbf{x})}$ will be a solution. Conversely, as the following lemma shows, when $a = 0$, any solution $\mathbb{P}_{\mathbf{x}}$ 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 $\mathbb{P}_{\mathbf{x}}$ satisfies (45). Then, for any $\varphi \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$,*

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

is $\mathbb{P}_{\mathbf{x}}$-martingle relative to $\{\mathcal{B}_t : t \geq 0\}$ with mean value 0. In particular, if $a = 0$, then

$$\mathbb{P}_{\mathbf{x}}\left(\psi(t) = \mathbf{x} + \int_0^t b\big(\psi(\tau)\big)\,d\tau \text{ for } t \geq 0\right) = 1.$$

Proof: Without loss in generality, we will assume that $\varphi(\mathbf{x}) = 0$.

Set

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

Then, by (45) and Lemma 37,

$$\begin{aligned}
M(t)^2 &= \varphi^2\big(\psi(t)\big) - 2M(t)V(t) - V(t)^2 \\
&\simeq \int_0^t L\varphi^2\big(\psi(\tau)\big)\,d\tau - 2\int_0^t M(\tau)\,dV(\tau) - V(t)^2 \\
&= \int_0^t \big(L\varphi^2 - 2\varphi L\varphi\big)\big(\psi(\tau)\big)\,d\tau + 2\int_0^t V(\tau)\,dV(\tau) - V(t)^2 \\
&= \int_0^t \big(\nabla\varphi, a\nabla\varphi\big)_{\mathbb{R}^N}\big(\psi(\tau)\big)\,d\tau,
\end{aligned}$$

where "$\simeq$" is used here when the difference between the quantities it separaties is a $\mathbb{P}_{\mathbf{x}}$-martingale relative to $\{\mathcal{B}_t : t \geq 0\}$, and, in the second line, I have used the identities $V(t)^2 = 2\int_0^t V(\tau)\,dV(\tau)$ and $L\varphi^2 - 2\varphi L\varphi = \big(\nabla\varphi, a\nabla\varphi\big)_{\mathbb{R}^N}$.

Now assume that $a = 0$. Then the preceding says that

$$\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\left[\left(\varphi\big(\psi(t)\big) - \varphi(\mathbf{x}) - \int_0^t \big(b, \nabla\varphi\big)_{\mathbb{R}^N}\big(\psi(\tau)\big)\,d\tau\right)^2\right] = 0$$

for all $\varphi \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$ 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 unequeness 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 $\mathbb{P}_{\mathbf{x}}$ is a solution to (45). If $v \in C_{\mathrm{b}}\big([0,T] \times \mathbb{R}^N;\mathbb{R}\big) \cap C^{1,2}\big((0,T) \times \mathbb{R}^N;\mathbb{R}\big)$ and $f \equiv (\partial_\tau + L)v$ is bounded on $(0,T) \times \mathbb{R}^N$, then*

$$\left(v\big(t\wedge T, \psi(t \wedge T)\big) - \int_9^{t\wedge T} f\big(\tau, \psi(\tau)\big)\,d\tau, \mathcal{B}_t, \mathbb{P}_{\mathbf{x}}\right)$$

is a martingale. In particular, if $f = 0$, then

$$\Big(v\big(t\wedge T, \psi(t\wedge T)\big), \mathcal{B}_t, \mathbb{P}_{\mathbf{x}}\Big)$$

is a martingale.

PROOF: Assume for the moment that $v \in C^{1,2}_{\mathrm{b}}\big((0,T)\times\mathbb{R}^N;\mathbb{R}\big)$. Given $0 \leq s < t < T$, set $\tau_{m,n} = s + \frac{m}{n}(t-s)$. Then

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[v\big(t,\psi(t)\big) - v\big(s,\psi(s)\big)\,\big|\,\mathcal{B}_s\big]\\
&\quad = \sum_{m=0}^{n-1}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[v\big(\tau_{m+1,n},\psi(\tau_{m+1,n})\big) - v\big(\tau_{m,n},\psi(\tau_{m,n})\big)\,\big|\,\mathcal{B}_s\big]\\
&\quad = \sum_{m=0}^{n-1}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\left[\int_{\tau_{m,n}}^{\tau_{m+1.n}}\Big(\partial_\tau v\big(\tau,\psi(\tau_{m+1,n})\big) + Lv\big(\tau_{m,n},\psi(\tau)\big)\,d\tau\,\right|\mathcal{B}_s\Bigg]\\
&\quad \longrightarrow \mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\left[\int_s^t(\partial_\tau - L)v\big(\tau,\psi(\tau)\big)\,d\tau\,\right|\mathcal{B}_s\Bigg]
\end{aligned}$$

as $n \to \infty$, and therefore the asserted martingale property holds in this case. To treat the general case, set $\zeta_R = \inf\{t \geq 0 : \psi(t) \notin B(\mathbf{x},R)\} \wedge T$, and choose $v_R \in C^{1,2}_{\mathrm{b}}\big((0,T)\times\mathbb{R}^N;\mathbb{R}\big)$ so that $v = v_R$ on $[0,T] \times \overline{B(\mathbf{0},R)}$. Then, by the preceding applied to v_R and Doob's stopping time theorem,

$$\left(v\big(t\wedge\zeta_R, \psi(t\wedge\zeta_R)\big) - \int_0^{t\wedge\zeta_R} f\big(\tau,\psi(\tau)\big)\,d\tau, \mathcal{B}_t, \mathbb{P}_{\mathbf{x}}\right)$$

is a martingale, and so the desired result follows after one lets $R \to \infty$. $\square$

Assume that $\mathbb{P}_{\mathbf{x}}$ is a solution to (45), and, for a given $\varphi \in C_{\mathrm{b}}(\mathbb{R};\mathbb{R})$ suppose that $u_\varphi \in C_{\mathrm{b}}\big([0,\infty)\times\mathbb{R}^N;\mathbb{R}\big) \cap C^{1,2}\big((0,\infty)\times\mathbb{R}^N;\mathbb{R}\big)$ is a solution to the Kolmogorov backward equation

$$\partial_t u_\varphi = L u_\varphi \text{ and } u_\varphi(0,\,\cdot\,) = \varphi. \tag{48}$$

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

$$\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(T)\big)\,\big|\,\mathcal{B}_s\big] = u_\varphi\big(T-s,\psi(s)\big).$$

Now assume that there is a solution u_φ to (48) for each φ in a determining class[7] Φ of non-negative, bounded, continuous functions. Then, if $\mathbb{P}_{\mathbf{x}}$ and $\tilde{\mathbb{P}}_{\mathbf{x}}$ are solutions to (45),

$$\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(t)\big)\big] = u_\varphi(t,\mathbf{x}) = \mathbb{E}^{\tilde{\mathbb{P}}_{\mathbf{x}}}\big[\varphi\big(\psi(t)\big)\big]$$

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

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(t_{n+1})\big),\ \big(\psi(t_1),\dots,\psi(t_n)\big) \in \Gamma\big] \\
&\quad = \mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[u_\varphi\big(t_{n+1}-t_n,\psi(t_n)\big),\ \big(\psi(t_1),\dots,\psi(t_n)\big) \in \Gamma\big] \\
&\quad = \int_\Gamma u_\varphi\big(t_{n+1}-t_n,\mathbf{y}_n\big)\,\mu_{t_1,\dots,t_n}(d\mathbf{y}_1 \times \cdots \times d\mathbf{y}_n),
\end{aligned}$$

and similarly

$$\begin{aligned}
&\mathbb{E}^{\tilde{\mathbb{P}}_{\mathbf{x}}}\big[\varphi\big(\psi(t_{n+1})\big),\ \big(\psi(t_1),\dots,\psi(t_n)\big) \in \Gamma\big] \\
&\quad = \int_\Gamma u_\varphi\big(t_{n+1}-t_n,\mathbf{y}_n\big)\,\mu_{t_1,\dots,t_n}(d\mathbf{y}_1 \times \cdots \times d\mathbf{y}_n),
\end{aligned}$$

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

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

$$(\lambda - L)u_{\lambda,\varphi} = \varphi \tag{49}$$

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

has a solution in $C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$ for $\lambda > 0$ and φ's in a determining class Φ of non-negative, bounded continuous functions φ. 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 $0 \le s < t$,

$$\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(t)\big)\,\big|\,\mathcal{B}_s\big] = \mathbb{E}^{\tilde{\mathbb{P}}_{\mathbf{x}}}\big[\varphi\big(\psi(t)\big)\,\big|\,\mathcal{B}_s\big]$$

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

$$\int_0^\infty e^{-\lambda\tau}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(s+\tau)\big)\,\big|\,\mathcal{B}_s\big]\,d\tau = \int_0^\infty e^{-\lambda\tau}\mathbb{E}^{\tilde{\mathbb{P}}_{\mathbf{x}}}\big[\varphi\big(\psi(s+\tau)\big)\,\big|\,\mathcal{B}_s\big]\,d\tau,$$

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

$$\left(e^{-\lambda t}u_{\lambda,\varphi}\big(\psi(t)\big) + \int_0^t e^{-\lambda\tau}\varphi\big(\psi(\tau)\big)\,d\tau, \mathcal{B}_t, \mathbb{P}_{\mathbf{x}}\right)$$

is a martingale, and therefore

$$e^{-\lambda t}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[u_{\lambda,\varphi}\big(\psi(t)\big)\,\big|\,\mathcal{B}_s\big] = e^{-\lambda s}u_{\lambda,\varphi}\big(\psi(s)\big) - \int_s^t e^{-\lambda\tau}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(\tau)\big)\,\big|\,\mathcal{B}_s\big]\,d\tau.$$

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

$$\int_0^\infty e^{-\lambda\tau}\mathbb{E}^{\mathbb{P}_{\mathbf{x}}}\big[\varphi\big(\psi(s+\tau)\big)\,\big|\,\mathcal{B}_s\big]\,d\tau = u_{\lambda,\varphi}\big(\psi(s)\big)$$

for any $\mathbb{P}_{\mathbf{x}}$ satisfying (45).

When a and b are uniformly bounded and a is uniformly positive definite, in the sense that $a \ge \epsilon\mathbf{I}$ 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 σ and b have two bounded, continuous derivatives, one can use the construction that we gave of $X(t,\mathbf{x})$ to show that (11) can be solved for $\varphi \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$.

Itô Integration

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

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

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 integratiion and show that all these martingales can be expressed as stochastic integrals.

The Paley–Wiener Integral: Recall that if $f : [0,\infty) \longrightarrow \mathbb{R}$ is continuous and $g : [0,\infty) \longrightarrow \mathbb{R}$ 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 $\big(B(t), \mathcal{F}_t, \mathbb{P}\big)$ 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 $\boldsymbol{\eta} : [0,\infty) \longrightarrow \mathbb{R}^M$ of locally bounded variation, set

$$I_{\boldsymbol{\eta}}(t) = \int_9^t \big(\boldsymbol{\eta}(\tau), dB(\tau)\big)_{\mathbb{R}^M},$$

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

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

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

$$(50) \qquad \big(I_{\boldsymbol{\eta}}(t), \mathcal{F}_t, \mathbb{P}\big), \quad \left(I_{\boldsymbol{\eta}}(t)^2 - \int_0^t |\boldsymbol{\eta}(\tau)|^2\, d\tau, \mathcal{F}_t, \mathbb{P}\right),$$

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

$$(51) \qquad \left(\exp\left(\zeta I_{\boldsymbol{\eta}}(t) - \frac{\zeta^2}{2}\int_0^t |\boldsymbol{\eta}(\tau)|^2\, d\tau\right), \mathcal{F}_t, \mathbb{P}\right)$$

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

$$(52) \qquad \tfrac{1}{4}\mathbb{E}^{\mathbb{P}}\left[\sup_{t\ge 0} I_{\boldsymbol{\eta}}(t)^2\right] \le \sup_{t\ge 0} \mathbb{E}^{\mathbb{P}}\big[I_{\boldsymbol{\eta}}(t)^2\big] = \int_0^\infty |\boldsymbol{\eta}(\tau)|^2\, d\tau.$$

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

$$\int_0^\infty |\boldsymbol{\eta}_k(\tau) - \boldsymbol{\eta}(\tau)|^2\, d\tau \longrightarrow 0.$$

Then, by (52),

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{t\geq 0}|I_{\boldsymbol{\eta}_\ell}(t)-I_{\boldsymbol{\eta}_k}(t)|^2\right]=\mathbb{E}^{\mathbb{P}}\left[\sup_{t\geq 0}|I_{\boldsymbol{\eta}_\ell-\boldsymbol{\eta}_k}(t)|^2\right]\leq 4\int_0^\infty|\boldsymbol{\eta}_\ell(\tau)-\boldsymbol{\eta}_k(\tau)|^2\,d\tau,$$

and so there exists a measurable function $I_{\boldsymbol{\eta}}$, known as the **Paley-Wiener integral** of $\boldsymbol{\eta}$, such that $I_{\boldsymbol{\eta}}(t)$ is $\mathcal{F}_t$ measurable for each $t\geq 0$, $t\rightsquigarrow I_{\boldsymbol{\eta}}(t,\omega)$ is continuous for all $\omega\in\Omega$, and

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{t\geq 0}|I_{\boldsymbol{\eta}}(t)-I_{\boldsymbol{\eta}_k}(t)|^2\right]\leq 4\int_0^\infty|\boldsymbol{\eta}(\tau)-\boldsymbol{\eta}_k(\tau)|^2\,d\tau\longrightarrow 0.$$

Clearly, apart from a $\mathbb{P}$-null set, $I_{\boldsymbol{\eta}}$ does not depend on the choice of the approximating sequence and is a linear function of $\boldsymbol{\eta}$. Moreover, for each $t\geq 0$, $I_{\boldsymbol{\eta}}(t)$ is a centered Gaussian random variable with variance $\int_0^t|\boldsymbol{\eta}(\tau)|^2\,d\tau$, the expressions in (50) and (51) are martingales, and (52) is continues to hold. What is no longer true is that $I_{\boldsymbol{\eta}}(t)(\omega)$ is given by a Riemann-Stieltjes integral or that it even defined ω by ω.

Itô's Integral: Although they resemble one another, the difference between Ito'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 $I(t,\mathbf{x})$. 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 extention to more general integrands. Further remember that he already familiar with the virtues of quantities like $I_n(t,\mathbf{x})$ 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 $\boldsymbol{\eta}$ that are **adapted** to the filtration in the sense that $\boldsymbol{\eta}(t)$ is $\mathcal{F}_t$-measurable for each $t\geq 0$ and for which

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^\infty|\boldsymbol{\eta}(\tau)|^2\,d\tau\right]<\infty. \tag{53}$$

If $\boldsymbol{\eta}$ is such an integrand and if, in addition, $\boldsymbol{\eta}(\,\cdot\,)(\omega)$ 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

$$\tfrac{1}{4}\mathbb{E}^{\mathbb{P}}\left[\sup_{t\geq 0}I_{\boldsymbol{\eta}}(t)^2\right]\leq\sup_{t\geq 0}\mathbb{E}^{\mathbb{P}}\big[I_{\boldsymbol{\eta}}(t)^2\big]=\mathbb{E}^{\mathbb{P}}\left[\int_0^\infty|\boldsymbol{\eta}(\tau)|^2\,d\tau\right] \tag{54}$$

in place of (52).

Ito's next step was to use (54) to extend his definition, and for this purpose he chose a $\rho \in C^\infty\big(\mathbb{R};[0,\infty)\big)$ that vanishes off $(0,1)$ and has total integral 1, and set

$$\boldsymbol{\eta}_k(t) = \int_0^\infty \rho_{\frac{1}{k}}(t-\tau)\boldsymbol{\eta}(\tau)\,d\tau,$$

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

$$\int_0^\infty |\boldsymbol{\eta}_{\frac{1}{k}}(\tau)|^2\,d\tau \le \int_0^\infty |\boldsymbol{\eta}(\tau)|^2\,d\tau \text{ and } \lim_{k\to\infty}\int_0^\infty |\boldsymbol{\eta}_{\frac{1}{k}}(\tau) - \boldsymbol{\eta}(\tau)|^2\,d\tau = 0.$$

Hence,

$$\lim_{k\to\infty}\mathbb{E}^{\mathbb{P}}\left[\int_0^\infty |\boldsymbol{\eta}_{\frac{1}{k}}(\tau) - \boldsymbol{\eta}(\tau)|^2\,d\tau\right] = 0.$$

Furthermore, because the construction of $\boldsymbol{\eta}_{\frac{1}{k}}(t)$ involves only $\boldsymbol{\eta}(\tau)$ for $\tau \le t$, Itô, without further comment, claimed that $\boldsymbol{\eta}_{\frac{1}{k}}$ 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 $t \ge 0$. The great advantage of this notion is that a function is progressively measurable if and only if it is measurable with respect to the σ-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 $\boldsymbol{\eta}_{\frac{1}{k}}$ is progressively measurable if $\boldsymbol{\eta}$ 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 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 $\boldsymbol{\eta}$ which are progressively measurable and satisfy (53), and given $\boldsymbol{\eta} \in \mathcal{P}^2(\mathbb{R}^M)$, one can use (54) to show that $I_{\boldsymbol{\eta}_k}$ converges to an $I_{\boldsymbol{\eta}}$ which, up to a $\mathbb{P}$-null set, is independent of the choice of approximants and for which (50) and (54) hold. In particular, $\boldsymbol{\eta} \rightsquigarrow I_{\boldsymbol{\eta}}$ is a linear map of $\mathcal{P}^2(\mathbb{R}^M)$ into the space $M^2_{\mathrm{c}}(\mathbb{R})$ of continuous martingales $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ for which

$$\|M\|_{M^2_c(R)} \equiv \sup_{t\ge 0}\|M(t)\|_{L^2(\mathbb{P};\mathbb{R})} < \infty.$$

In fact, if

$$\|\boldsymbol{\eta}\|_{\mathcal{P}^2(\mathbb{R}^M)} \equiv \mathbb{E}^{\mathbb{P}}\left[\int_0^\infty |\boldsymbol{\eta}(\tau)|^2\,d\tau\right]^{\frac{1}{2}},$$

then $\|I_{\boldsymbol\eta}\|_{M^2_c(\mathbb{R})} = \|\boldsymbol\eta\|_{\mathcal{P}^2(\mathbb{R}^M)}$, and so $\boldsymbol\eta \rightsquigarrow I_{\boldsymbol\eta}$ is an isometry between the metrics determined by the norms $\|\cdot\|_{\mathcal{P}^2(\mathbb{R}^M)}$ and $\|\cdot\|_{M^2_c(\mathbb{R})}$.

Because it shares many properties with standard integrals, the quantity $I_{\boldsymbol\eta}(t)$ is usually denoted by

$$\int_9^t \big(\boldsymbol\eta(\tau), dB(\tau)\big)_{\mathbb{R}^M}$$

and is called the **Itô stocastic integral**, or just the **stochastic integral**, of $\boldsymbol\eta$ with respect to B.

Some Properties and Extentions: Given $\boldsymbol\eta_1, \boldsymbol\eta_2 \in \mathcal{P}^2(\mathbb{R}^M)$, a simple polarization argument shows that

$$\left(I_{\boldsymbol\eta_1}(t)I_{\boldsymbol\eta_3}(t) - \int_0^t \big(\boldsymbol\eta_1(\tau), \boldsymbol\eta_2(\tau)\big)_{\mathbb{R}^M}\, d\tau, \mathcal{F}_t, \mathbb{P}\right)$$

is a martingale.

Now suppose that $\boldsymbol\eta \in \mathcal{P}^2(\mathbb{R}^M)$ and that ζ is a stopping time relative to $\{\mathcal{F}_t : t \ge 0\}$, Then, by Doob's stopping time theorem,

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[I_{\boldsymbol\eta}(t\wedge\zeta)\int_0^t \mathbf{1}_{[0,\zeta}(\tau)\big(\boldsymbol\eta(\tau), dB(\tau)\big)_{\mathbb{R}^M}\right]\\ &\quad = \mathbb{E}^{\mathbb{P}}\left[I_{\boldsymbol\eta}(t\wedge\zeta)\int_0^{t\wedge\zeta} \mathbf{1}_{[0,\zeta}(\tau)\big(\boldsymbol\eta(\tau), dB(\tau)\big)_{\mathbb{R}^M}\right] = \mathbb{E}^{\mathbb{P}}\left[\int_0^{t\wedge\zeta} |\boldsymbol\eta(\tau)|^2\, d\tau\right],\end{aligned}$$

and so

$$\mathbb{E}^{\mathbb{P}}\left[\left|I_{\boldsymbol\eta}(t\wedge\zeta) - \int_0^t \mathbf{1}_{[0,\zeta)}(\tau)\big(\boldsymbol\eta(\tau), dB(\tau)\big)_{\mathbb{R}^M}\right|^2\right] = 0.$$

Hence

$$I_{\boldsymbol\eta}(t\wedge\zeta) = \int_0^t \mathbf{1}_{[0,\zeta)}(\tau)\big(\boldsymbol\eta(\tau), dB(\tau)\big)_{\mathbb{R}^M}.$$

In particular, if ζ_1 and ζ_2 are a pair of stopping times and $\zeta_1 \le \zeta_2$, then

$$\int_0^{t\wedge\zeta_1} \mathbf{1}_{[0,\zeta_2)}(\tau)\big(\boldsymbol\eta(\tau), dB(\tau)\big)_{\mathbb{R}^M} = \int_0^t \mathbf{1}_{[0,\zeta_1)}(\tau)\big(\boldsymbol\eta(\tau), dB(\tau)\big)_{\mathbb{R}^M}.$$

The preceding considerations afford us the opportunity to integrate $\boldsymbol\eta$'s that are not in $\mathcal{P}^2(\mathbb{R}^M)$. Namely, let $\mathcal{P}^2_{\mathrm{loc}}(\mathbb{R}^M)$ be the set of progressively measurable, $\mathbb{R}^M$-valued functions with the property that $\int_0^t |\boldsymbol\xi(\tau)|^2\, d\tau < \infty$ for all $t \in [0,\infty)$. Then, if $\boldsymbol\xi_k = \mathbf{1}_{[0,\zeta_k)}\boldsymbol\xi$ where

$$\zeta_k = \inf\left\{t \ge); \int_0^t |\boldsymbol\xi(\tau)|^2\, d\tau \ge k\right\},$$

$I_{\boldsymbol{\eta}_{k+1}}(t \wedge \zeta_k) = I_{\boldsymbol{\eta}_k}(t)$, and so not only does

$$I_{\boldsymbol{\eta}}(t) = \lim_{k\to\infty} I_{\boldsymbol{\eta}_k}(t)$$

exist, but also $I(t\wedge\zeta_k) = I_{\boldsymbol{\eta}_k}(t)$ for all $k \geq 1$. Of course, in general $\big(I_{\boldsymbol{\eta}}(t), \mathcal{F}_t, \mathbb{P}\big)$ will not be a martingale since $I_{\boldsymbol{\eta}}(t)$ need not be even $\mathbb{P}$-integrable. On the other hand, for each $k \geq 1$,

$$\big(I_{\boldsymbol{\eta}}(t\wedge\zeta_k), \mathcal{F}_t, \mathbb{P}\big) \text{ and } \left(I_{\boldsymbol{\eta}}(t)^2 - \int_0^{t\wedge\zeta_k} |\boldsymbol{\eta}(\tau)|^2\, d\tau, \mathcal{F}_t, \mathbb{P}\right)$$

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 $\big(M(t\wedge\zeta_k), \mathcal{F}_t, \mathbb{P}\big)$ is a martingale for each $k \geq 1$. Observe that if $\boldsymbol{\eta} \in \mathcal{P}^2_{\text{loc}}(\mathbb{R}^M)$ and ζ is a stopping time, then, since

$$I_{\boldsymbol{\eta}}(t\wedge\zeta) = \lim_{k\to\infty} I_{\boldsymbol{\eta}}(t\wedge\zeta_k\wedge\zeta) = \lim_{k\to\infty} \int_0^{t\wedge\zeta_k} \mathbf{1}_{[0,\zeta)}(\tau)\big(\boldsymbol{\eta}(\tau), dB(\tau)\big)_{\mathbb{R}^M},$$

it is still true that

$$I_{\boldsymbol{\eta}}(t\wedge\zeta) = \int_0^t \mathbf{1}_{[0,\zeta)}(\tau)\big(\boldsymbol{\eta}(\tau), dB(\tau)\big)_{\mathbb{R}^M}.$$

Thus, if $\mathbb{E}^{\mathbb{P}}\left[\int_0^\zeta |\boldsymbol{\eta}(\tau)|^2\, d\tau\right] < \infty$, then

$$\big(I_{\boldsymbol{\eta}}(t\wedge\zeta), \mathcal{F}_t, \mathbb{P}\big) \text{ and } \left(I_{\boldsymbol{\eta}}(t\wedge\zeta)^2 - \int_0^{t\wedge\zeta} |\boldsymbol{\eta}(\tau)|^2\, d\tau, \mathcal{F}_t, \mathbb{P}\right)$$

are martingales. In particular, if $\mathbb{E}^{\mathbb{P}}\left[\int_0^t |\boldsymbol{\eta}(\tau)|^2\, d\tau\right] < \infty$ for all $t \geq 0$, then

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

are martingales. Finally, if $\boldsymbol{\eta}$ is an $\mathbb{R}^M$-valued, adapted function and $\boldsymbol{\eta}(\,\cdot\,)(\omega)$ is continuous for all $\omega \in \Omega$, then $\boldsymbol{\eta}$ is progressively measurable and, by taking

$$\zeta_k = \inf\left\{t \geq 0 : \int_0^t |\boldsymbol{\eta}(\tau)|^2\, d\tau \geq k\right\},$$

one sees that $\boldsymbol{\eta} \in \mathcal{P}^2_{\text{loc}}(\mathbb{R}^M)$.

Identify $\mathbb{R}^N\otimes\mathbb{R}^M$ with $\mathrm{Hom}(\mathbb{R}^M;\mathbb{R}^N)$, 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 σ with the property that

$$\|\sigma\|_{\mathcal{P}^2(\mathbb{R}^N\otimes\mathbb{R}^M)} \equiv \mathbb{E}^{\mathbb{P}}\left[\int_0^\infty \|\sigma(\tau)\|_{\mathrm{H.S.}}^2\,d\tau\right]^{\frac12} < \infty,$$

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

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

so that

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

It is then an easy exercise to check that

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

are martingales and that

$$\begin{aligned}\tfrac12\mathbb{E}^{\mathbb{P}}\left[\sup_{t\ge0}|I_\sigma(t)|^2\right]^{\frac12} &\le \|I_\sigma\|_{M_c^2(\mathbb{R}^N\otimes\mathbb{R}^M)} \equiv \sup_{t\ge0}\mathbb{E}^{\mathbb{P}}\big[|I_\sigma(t)|^2\big]^{\frac12}\\ &= \mathbb{E}^{\mathbb{P}}\left[\int_0^\infty \|\sigma(\tau)\|_{\mathrm{H.S.}}^2\,d\tau\right]^{\frac12}.\end{aligned}$$

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

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

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

Recall that, for all $T>0$,

$$\lim_{n\to\infty}\mathbb{E}^{\mathcal{W}}\big[\|I(\,\cdot\,,\mathbf{x})-I_n(\,\cdot\,,\mathbf{x})\|_{[0,T]}^2 \vee \|X(\,\cdot\,,\mathbf{x})-X_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big]=0$$

and

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \mathbb{E}^{\mathcal{W}}\big[|X(t,\mathbf{x}) - X([t]_n,\mathbf{x})|^2\big] = 0$$

Hence,

$$\begin{aligned}\overline{\lim_{n\to\infty}}\, \mathbb{E}^{\mathcal{W}}\left[\sup_{t\in[0,T]}\left|\int_0^t \sigma\big(X(\tau,\mathbf{x})\big)\,d\tau - I_n(t,\mathbf{x})\right|^2\right] \\ \le 4\,\overline{\lim_{n\to\infty}}\, \mathbb{E}^{\mathcal{W}}\left[\int_0^T \big\|\sigma\big(X(\tau,\mathbf{x})\big) - \sigma\big(X_n([\tau]_n,\mathbf{x})\big)\big\|_{\text{H.S.}}^2\,d\tau\right] = 0,\end{aligned}$$

and so

$$I(t,\mathbf{x}) = \int_0^t \sigma\big(X(\tau,\mathbf{x})\big)\,dw(\tau).$$

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

$$X(t,\mathbf{x}) = \mathbf{x} + \int_0^t \sigma\big(X(\tau,\mathbf{x})\big)\,dw(\tau) + \int_0^t b\big(X(\tau,\mathbf{x})\big)\,d\tau. \tag{56}$$

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

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}\big[|X(t\wedge\zeta_R,\mathbf{x}) - \tilde X(t\wedge\zeta_R,\mathbf{x})|^2\big] &\le 2\mathbb{E}^{\mathcal{W}}\left[\int_0^{t\wedge\zeta_R} \big\|\sigma\big(X(\tau,\mathbf{x})\big) - \sigma\big(\tilde X(\tau,\mathbf{x})\big)\big\|_{\text{H.S.}}^2\,d\tau\right] \\ &\quad + 2t\mathbb{E}^{\mathcal{W}}\left[\int_0^{t\wedge\zeta_R} \big|b\big(X(\tau,\mathbf{x})\big) - b\big(\tilde X(\tau,\mathbf{x})\big)\big|^2\,d\tau\right] \\ &\le C(1+t)\int_0^t \mathbb{E}^{\mathcal{W}}\big[\big|X(\tau\wedge\zeta_R,\mathbf{x}) - \tilde X(\tau\wedge\zeta_R,\mathbf{x})\big|^2\big]\,d\tau,\end{aligned}$$

for some $C < \infty$, and so, by Gromwall's lemma, $X(t,\mathbf{x}) = \tilde X(t,\mathbf{x})$ for $t \in [0,\zeta_R)$. Since $\zeta_R \nearrow \infty$, this proves that $\tilde X(\,\cdot\,,\mathbf{x}) = X(\,\cdot\,,\mathbf{x})$.

Having described $X(\,\cdot\,,\mathbf{x})$ 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 $\tilde X_0(t,\mathbf{x}) = \mathbf{x}$ and

$$\tilde X_{n+1}(t,\mathbf{x}) = \mathbf{x} + \int_0^t \sigma\big(\tilde X_n(\tau,\mathbf{x})\big)\,dw(\tau) + \int_0^t b\big(\tilde X_n(\tau,\mathbf{x})\big)\,d\tau$$

for $n \geq 0$. Set $\Delta_n(t) = \|\tilde{X}_{n+1}(\,\cdot\,,\mathbf{x}) - \tilde{X}_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}$. Then

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}\big[\Delta_n(t)^2\big] \leq& 8\int_0^t \mathbb{E}^{\mathcal{W}}\big[\big\|\sigma\big(\tilde{X}_n(\tau,\mathbf{x})\big) - \sigma\big(\tilde{X}_{n-1}(\tau,\mathbf{x})\big)\big\|_{\mathrm{H.S.}}^2\big]\,d\tau \\ &+ 2t\int_0^t \mathbb{E}^{\mathcal{W}}\big[\big|b\big(\tilde{X}_n(\tau,\mathbf{x})\big) - b\big(\tilde{X}_{n-1}(\tau,\mathbf{x})\big)\big|^2\big]\,d\tau \\ \leq& C(1+t)\int_0^t \mathbb{E}^{\mathcal{W}}\big[\Delta_{n-1}(\tau)^2\big]\,d\tau\end{aligned}$$

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

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

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

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

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

The Crown Jewel: Itô's Formula: Although we interpreted $X(\,\cdot\,,\mathbf{x})$ as a stochastic integral, we have yet to interpret the martingale

$$\big(M_\varphi(t), \overline{\mathcal{B}_t}, \mathcal{W}\big) \text{ where } M_\varphi(t) = \varphi\big(X(t,\mathbf{x})\big) - \varphi(\mathbf{x}) - \int_0^t L\varphi\big(X(\tau.\mathbf{x})\big)\,d\tau$$

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

$$\varphi\big(X(t,\mathbf{x})\big) - \varphi(\mathbf{x}) = \int_0^t \Big(\sigma\big(X(\tau,\mathbf{x})\big)^\top \nabla\varphi\big(X(\tau,\mathbf{x})\big), dw(\tau)\Big)_{\mathbb{R}^N} + \int_0^t b\big(X(\tau,\mathbf{x})\big)\,d\tau,$$

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 indentify. That is, we are guessing that

$$\varphi\big(X(t,\mathbf{x})\big) - \varphi(\mathbf{x}) = \int_0^t \Big(\sigma\big(X(\tau,\mathbf{x})\big)^\top \nabla\varphi\big(X(\tau,\mathbf{x})\big), dw(\tau)\Big)_{\mathbb{R}^N} + \int_0^t L\varphi\big(X(\tau.\mathbf{x})\big)\,d\tau$$

and that the appearance of the term

$$\tfrac{1}{2}\int_0^t \mathrm{Trace}\big(a(X(\tau,\mathbf{x}))\nabla^2\varphi(X(\tau,\mathbf{x}))\big)\,d\tau$$

reflects the fact that $dw(t)$ is on the order of $\sqrt{dt}$, and therefore is not a true infinitesmal. Further evidence for this conjecture is provided by Lemma 46, from which it follows that

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

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 $\sigma \in \mathcal{P}^2_{\mathrm{loc}}(\mathbb{R}^{N_2}\otimes\mathbb{R}^M)$, 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 φ is an element of $C^{1,2}(\mathbb{R}^{N_1}\times\mathbb{R}^{N_2};\mathbb{R})$, then

$$\begin{aligned}
&\varphi\big(V(t), I_\sigma(t)\big) - \varphi(V(0),\mathbf{0})\\
&\qquad = \int_0^t \big(\nabla_{(1)}\varphi(V(\tau), I_\sigma(\tau)), dV(\tau)\big)_{\mathbb{R}^{N_1}}\\
&\qquad\quad + \int_0^t \big(\sigma(\tau)^\top\nabla\varphi(I_\sigma(\tau)), dB(\tau)\big)_{\mathbb{R}^M}\\
&\qquad\quad + \tfrac{1}{2}\int_0^t \operatorname{Trace}\big(\sigma\sigma^\top(\tau)\nabla^2_{(2)}\varphi(I_\sigma)(\tau)\big)\,d\tau,
\end{aligned}\tag{57}$$

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 σ and V are bounded, $\sigma(\,\cdot\,)$ is continuous, and $\varphi\in C^\infty_{\mathrm{c}}(\mathbb{R}^{N_1}\times\mathbb{R}^{N_2};\mathbb{R})$. Thus I will proceed under those assumptions.

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

$$\left(\boldsymbol{\xi}, \int_s^t \sigma(\tau)\,dB(\tau)\right)_{\mathbb{R}^{N_2}} = \int_s^t \big(\sigma(\tau)^\top\boldsymbol{\xi},\, dB(\tau)\big)_{\mathbb{R}^M} \text{ for } t\ge s.$$

To check this, simply observe that

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\left[\left|\left(\boldsymbol{\xi}, \int_s^t \sigma(\tau)\,dB(\tau)\right)_{\mathbb{R}^{N_2}} - \int_s^t \big(\sigma(\tau)^\top\boldsymbol{\xi}, dB(\tau)\big)_{\mathbb{R}^M}\right|^2\right]\\
&= \mathbb{E}^{\mathbb{P}}\left[\int_s^t |\sigma(\tau)^\top\boldsymbol{\xi}|^2\,d\tau\right] - 2\mathbb{E}^{\mathbb{P}}\left[\int_s^t |\sigma(\tau)^\top\boldsymbol{\xi}|^2\,d\tau\right] + \mathbb{E}^{\mathbb{P}}\left[\int_s^t |\sigma(\tau)^\top\boldsymbol{\xi}|^2\,d\tau\right] = 0.
\end{aligned}$$

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

$$\varphi\big(V(t), I_\sigma(t)\big) - \varphi\big(V(0),\mathbf{0}\big) = \lim_{n\to\infty}\Big(\varphi\big(V(t), I_{\sigma_n}(t)\big) - \varphi\big(V(0),\mathbf{0}\big)\Big)$$

and

$$\varphi\big(V(t), I_{\sigma_n}(t)\big) - \varphi\big(V(0(, \mathbf{0}\big) = \sum_{m<2^n t} \Big(\varphi\big(V(t_{m+1,n}), I_{m+1,n})\big) - \varphi\big(V(t_{m,n}, I_{m,n})\big)\Big),$$

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

$$\begin{aligned}\varphi\big(V(t_{m+1,n}), I_{m+1,n}\big) &- \varphi\big(V(t_{m,n}, I_{m+1,n})\big)\\ &= \int_{t_{m,n}}^{t_{m+1,n}} \big(\nabla_{(1)}\varphi(V(\tau), I_{m+1}), dV(\tau)\big)_{\mathbb{R}^{N_1}},\end{aligned}$$

$$\begin{aligned}\varphi\big(V(t), I_{\sigma_n}(t)\big) - \varphi\big(V(0), \mathbf{0}\big) = &\int_0^t \big(\nabla_{(1)}\varphi(V(\tau), I_{\sigma_n}([\tau]_n)), dV(\tau)\big)_{\mathbb{R}^{N_1}}\\ &+ \sum_{m<2^n t} \Big(\varphi\big(V(t_{m,n}), I_{m+1,n}\big) - \varphi\big(V(t_{m,n}, I_{m,n})\big)\Big),\end{aligned}$$

and obviously

$$\int_0^t \big(\nabla_{(1)}\varphi(V(\tau), I_{\sigma_n}([\tau]_n)), dV(\tau)\big)_{\mathbb{R}^{N_1}} \longrightarrow \int_0^t \big(\nabla_{(1)}\varphi(V(\tau), I_{\sigma}(\tau)), dV(\tau)\big)_{\mathbb{R}^{N_1}}.$$

Next, by Taylor's theorem,

$$\begin{aligned}&\varphi\big(V(t_{m,n}), I_{m+1,n}\big) - \varphi\big(V(t_{m,n}), I_{m,n}\big)\\ &= \Big(\nabla_{(2)}\varphi(V(t_{m,n}), I_{m,n}), \Delta_{m,n}\Big)_{\mathbb{R}^{N_2}}\\ &\quad + \tfrac{1}{2}\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}), I_{m,n})\Delta_{m,n} \otimes \Delta_{m,n}\Big) + E_{m,n},\end{aligned}$$

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

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

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

$$\begin{aligned}&\sum_{m<2^n} \Big(\nabla_{(2)}\varphi(V(t_{m,n}), I_{m,n}), \Delta_{m,n}\Big)_{\mathbb{R}^{N_2}}\\ &\quad = \sum_{m<2^n t} \int_{t_{m,n}}^{t_{m+1,n}} \Big(\sigma([\tau]_n)^\top \nabla_{(2)}\varphi\big(V(t_{m,n}, I_{m,n})\big), dB(\tau)\Big)_{\mathbb{R}^M}\\ &\quad = \int_0^t \Big(\sigma(\tau)^\top \nabla_{(2)}\varphi\big(V([\tau]_n), I_{\sigma_n}([\tau]_n))\big), dB(\tau)\Big)_{\mathbb{R}^M}\\ &\quad \longrightarrow \int_0^t \Big(\sigma(\tau)^\top \nabla_{(2)}\varphi\big(V(\tau), I_{\sigma}(\tau))\big), dB(\tau)\Big)_{\mathbb{R}^M}\end{aligned}$$

in $L^2(\mathbb{P};\mathbb{R})$. Finally, write

$$\begin{aligned}&\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})\Delta_{m,n}\otimes\Delta_{m,n}\Big)\\&=\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})A_{m,n}\Big)\\&\qquad+\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})\big(\Delta_{m,n}\otimes\Delta_{m,n}-A_{m,n}\big)\Big),\end{aligned}$$

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

$$\begin{aligned}&\sum_{m<2^nt}\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})A_{m,n}\Big)\\&=\int_0^t\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V([\tau]_n),I_{\sigma_n}([\tau]_n))\sigma\sigma^\top([\tau]_n)\Big)\,d\tau\\&\longrightarrow\int_0^t\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(\tau),I_\sigma(\tau))\sigma\sigma^\top(\tau)\Big)\,d\tau.\end{aligned}$$

At the same time, because

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\big[\Delta_{m,n}\otimes\Delta_{m,n}\,\big|\,\mathcal{F}_{t_{m,n}}\big]\\&\quad=\mathbb{E}^{\mathbb{P}}\big[I_{\sigma_n}(t_{m+1,n})\otimes I_{\sigma_n}(t_{m+1,n})-I_{\sigma_n}(t_{m,n})\otimes I_{\sigma_n}(t_{m,n})\,\big|\,\mathcal{F}_{t_{m,n}}\big]\\&\quad=\mathbb{E}^{\mathbb{P}}\big[A_{m,n}\,\big|\,\mathcal{F}_{t_{m,n}}\big],\end{aligned}$$

the terms

$$\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})\big(\Delta_{m,n}\otimes\Delta_{m,n}-A_{m,n}\big)\Big)$$

are orthogonal in $L^2(\mathbb{P};\mathbb{R})$, and so

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\left(\sum_{m<2^nt}\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})\big(\Delta_{m,n}\otimes\Delta_{m,n}-A_{m,n}\big)\Big)\right)^2\right]\\&\quad=\sum_{m<2^nt}\mathbb{E}^{\mathbb{P}}\Big[\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})\big(\Delta_{m,n}\otimes\Delta_{m,n}-A_{m,n}\big)\Big)^2\Big].\end{aligned}$$

Obviously, $\|A_{m,n}\|_{\mathrm{H.S.}}\le C2^{-n}$ and $\|\Delta_{m,n}\otimes\Delta_{m,n}\|_{\mathrm{H.S.}}\le C|B(t_{m+1,n}-B(t_{m,n})|^2$ for some $C<\infty$. Hence, since

$$\mathbb{E}^{\mathbb{P}}\big[|B(t_{m+1,n}-B(t_{m,n})|^4\big]=2^{-2n}\mathbb{E}^{\mathbb{P}}\big[|B(1)|^4\big],$$

it follows that

$$\sum_{m<2^nt}\mathrm{Trace}\Big(\nabla^2_{(2)}\varphi(V(t_{m,n}),I_{m,n})\Delta_{m,n}\otimes\Delta_{m,n}-A_{m,n}\Big)\longrightarrow 0$$

in $L^2(\mathbb{P};\mathbb{R})$.

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

$$\begin{aligned} &d\varphi\big(V(t),I_\sigma(t)\big)\\ &\quad=\Big(\nabla_{(1)}\varphi\big(V(t),I_\sigma(t)\big),dV(t)\Big)_{\mathbb{R}^{M_1}}+\Big(\sigma(\tau)^\top\nabla_{(2)}\varphi\big(V(t),I_\sigma(\tau)\big),dB(\tau)\Big)_{\mathbb{R}^M}\\ &\qquad+\tfrac12\mathrm{Trace}\Big(\sigma\sigma^\top(t)\nabla^2_{(2)}\varphi\big(V(t),I_\sigma(t)\big)\Big)\,dt. \end{aligned}$$

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

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

$$\begin{aligned}\varphi\big(X(t,\mathbf{x})=&\varphi(\mathbf{x})+\int_0^t\Big(\sigma(X(\tau,\mathbf{x})^\top\nabla\varphi(X(\tau,\mathbf{x})),dB(\tau)\Big)_{\mathbb{R}^M}\\ &+\int_0^t L\varphi\big(X(\tau,\mathbf{x})\big)\,d\tau\end{aligned}$$

becomes a special case of (57) appield to $(\mathbf{x}_1,\mathbf{x}_2)\rightsquigarrow\varphi(\mathbf{x}_1+\mathbf{x}_2)$.

Some Applications: Let $\sigma\in\mathcal{P}^2_{\mathrm{loc}}(\mathbb{R}^N\otimes\mathbb{R}^M)$, and set $A(t)=\int_0^t\sigma\sigma^\top(\tau)\,d\tau$. 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

$$(\mathbf{y},v)\in\mathbb{R}^N\times\mathbb{R}\longmapsto e^{(\boldsymbol{\zeta},\mathbf{y})_{\mathbb{R}^N}-\frac{v}{2}}\in\mathbb{C}$$

to see that

$$\Big(\exp\Big(\big(\boldsymbol{\zeta},I_\sigma(t)\big)_{\mathbb{R}^N}-\tfrac12\big(\boldsymbol{\zeta},A(t)\boldsymbol{\zeta}\big)_{\mathbb{R}^N}\Big),\mathcal{F}_t,\mathbb{P}\Big)$$

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

$$\zeta_R=\inf\big\{t\ge0:\,|I_\sigma(t)|\vee\mathrm{Trace}\big(A(t)\big)\ge R\big\},$$

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

$$\mathbb{E}^{\mathbb{P}}\Big[\exp\Big(\big(\boldsymbol{\xi},I_\sigma(t\wedge\zeta_R)\big)_{\mathbb{R}^N}-\tfrac12\big(\boldsymbol{\xi},A(t\wedge\zeta_R)\boldsymbol{\xi}\big)_{\mathbb{R}^N}\Big)\Big]=1,$$

which, by HW3, means that, for $m\ge1$,

$$\kappa_{2m}^{-1}\mathbb{E}^{\mathbb{P}}\big[\big(\boldsymbol{\xi},A(t\wedge\zeta_R)\boldsymbol{\xi}\big)^m_{\mathbb{R}^N}\big]\le\mathbb{E}^{\mathbb{P}}\big[\big|\big(\boldsymbol{\xi},I_\sigma(t\wedge\zeta_R)\big)_{\mathbb{R}^N}\big|^{2m}\big]\le\kappa_{2m}\mathbb{E}^{\mathbb{P}}\big[\big(\boldsymbol{\xi},A(t\wedge\zeta_R)\boldsymbol{\xi}\big)^m_{\mathbb{R}^N}\big].$$

Since $R \rightsquigarrow \big(\boldsymbol{\xi}, A(t\wedge\zeta_R)\boldsymbol{\xi}\big)_{\mathbb{R}^N}$ is non-decreasing, the monotone convergence theorem together with Fatou's lemma implies that

$$\mathbb{E}^{\mathbb{P}}\big[\big|\big(\boldsymbol{\xi}, I_\sigma(t)\big)_{\mathbb{R}^N}\big|^{2m}\big] \le \kappa_{2m}\mathbb{E}^{\mathbb{P}}\big[\big(\boldsymbol{\xi}, A(t)\boldsymbol{\xi}\big)_{\mathbb{R}^N}^m\big].$$

At the same time, if $\mathbb{E}^{\mathbb{P}}\big[\big|\big(\boldsymbol{\xi}, I_\sigma(t)\big)_{\mathbb{R}^N}\big|^{2m}\big] < \infty$, then $\mathbb{E}^{\mathbb{P}}\big[\big\|\big(\boldsymbol{\xi}, I_\sigma(\,\cdot\,)\big)_{\mathbb{R}^N}\big\|_{[0,t]}^{2m}\big] < \infty$, and so, by Lebesgue's dominated convergence theorem,

$$\kappa_{2m}^{-1}\mathbb{E}^{\mathbb{P}}\big[\big(\boldsymbol{\xi}, A(t)\boldsymbol{\xi}\big)_{\mathbb{R}^N}^m\big] \le \mathbb{E}^{\mathbb{P}}\big[\big|\big(\boldsymbol{\xi}, I_\sigma(t)\big)_{\mathbb{R}^N}\big|^{2m}\big].$$

Hence, we have now shown that

$$\kappa_{2m}^{-1}\mathbb{E}^{\mathbb{P}}\big[\big(\boldsymbol{\xi}, A(t)\boldsymbol{\xi}\big)_{\mathbb{R}^N}^m\big] \le \mathbb{E}^{\mathbb{P}}\big[\big|\big(\boldsymbol{\xi}, I_\sigma(t)\big)_{\mathbb{R}^N}\big|^{2m}\big] \le \kappa_{2m}\mathbb{E}^{\mathbb{P}}\big[\big(\boldsymbol{\xi}, A(t)\boldsymbol{\xi}\big)_{\mathbb{R}^N}^m\big], \tag{58}$$

from which it follows immediately that

$$\begin{aligned}\kappa_{2m}^{-1}N^{1-m}\mathbb{E}^{\mathbb{P}}\big[\mathrm{Trace}\big(A(t)\big)^m\big] &\le \mathbb{E}^{\mathbb{P}}\big[\big|I_\sigma(t)\big|^{2m}\big]\\ &\le \kappa_{2m}N^{m-1}\mathbb{E}^{\mathbb{P}}\big[\mathrm{Trace}\big(A(t)\big)^m\big].\end{aligned}\tag{59}$$

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 there is another, more direct, way 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

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

Again using stopping times to reduce to the case when I_σ and A are bounded, apply (57) and Hölder's inequality to see that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[|I_\sigma(t)|^p\big] &\le p(p-1)\mathbb{E}^{\mathbb{P}}\left[\int_0^t \mathrm{Trace}\big(\sigma\sigma^\top(\tau)\big)|I_\sigma(\tau)|^{p-2}\,d\tau\right]\\ &\le p(p-1)\mathbb{E}^{\mathbb{P}}\left[\left(\int_0^t \mathrm{Trace}\big(\sigma\sigma^\top(\tau)\big)\,d\tau\right)\|I_\sigma(\,\cdot\,)\|_{[0,t]}^{p-2}\right]\\ &\le p(p-1)\mathbb{E}^{\mathbb{P}}\big[A(t)^{\frac{p}{2}}\big]^{\frac{2}{p}}\mathbb{E}^{\mathbb{P}}\big[\|I_\sigma(\,\cdot\,)\|_{[0,t]}^p\big]^{1-\frac{2}{p}},\end{aligned}$$

and then use Doob's inequality to conclude that

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

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

$$\mathbb{E}^{\mathbb{P}}\big[|I_\sigma(t)|^p\big] \le \mathbb{E}^{\mathbb{P}}\big[\|I_\sigma(\,\cdot\,)\|_{[0,t]}\big] \le K_p\mathbb{E}^{\mathbb{P}}\big[A(t)^{\frac{p}{2}}\big] \tag{60}$$

where $K_p = \left(\frac{p^{p+1}}{(p-1)^{p-1}}\right)^{\frac{p}{2}}$.

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 $\big(B(t), \mathcal{F}_t, \mathbb{P}\big)$. To describe local time, define the *occupation time* measures $L(t,\,\cdot\,)$ for $t \geq 0$ by

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

One of Lévy's many remarkable discoveries is that there is a map $\omega \in \Omega \longmapsto \ell(\,\cdot\,\cdot\,\cdot\,) \in C\big([0,\infty) \times \mathbb{R}\big)$ such that $\ell(\,\cdot\,,y)$ a progressively measurable, non-decreasing function for which

$$\mathbb{P}\left(L(t,\Gamma) = \int_\Gamma \ell(t,y)\,dy \text{ for all } t \geq 0 \;\&\; \Gamma \in \mathcal{B}_{\mathbb{R}}\right) = 1. \tag{61}$$

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 $p : [0,\infty) \longrightarrow \mathbb{R}$ is a continuously differentiable path, and let $t \rightsquigarrow \mu(t,\,\cdot\,)$ be its occupation time measures. If $\dot{p} = 0$ on an interval $[a,b]$, then it is clear that, for $t > a$, $\mu(t,\{p(a)\}) \geq (t \wedge b - a)$, 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 $t > a$,

$$\frac{\mu(t,dy) - \mu(a,dy)}{\lambda_{\mathbb{R}}(dy)} = \frac{\mathbf{1}_{[a,t]}(y)}{\dot{p} \circ (p \restriction [p(a),p(t)])^{-1}(y)},$$

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

$$\int_0^t \delta_y\big(B(\tau)\big)\,d\tau,$$

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

$$\frac{1}{2}\ell(t,y) = B(t) \vee y - 0 \vee y - \int_0^t \mathbf{1}_{[y,\infty)}\big(B(\tau)\big)\,d\tau.$$

The key step in the jusification of his idea is the proof that there is a continuous map $(t,y) \rightsquigarrow I(t,y)$ such that

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

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

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

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

$$\mathbb{E}^{\mathbb{P}}\left[\left(\int_0^T \mathbf{1}_{[x,y]}\big(B(\tau)\big)\right)^2 d\tau\right] = 2\mathbb{E}^{\mathbb{P}}\left[\int_{0\leq\tau_1<\tau_2} \mathbf{1}_{[x,y]}\big(B(\tau_1)\big)\mathbf{1}_{[x,y]}\big(B(\tau_2)\big)\,d\tau_1 d\tau_2\right]$$
$$= \frac{1}{2\pi}\int_{0\leq\tau_1<\tau_2} \tau_1^{-\frac{1}{2}}(\tau_2-\tau_1)^{-\frac{1}{2}}\left(\int_{[x,y]^2} e^{-\frac{\xi_1^2}{2\tau_1}}e^{-\frac{(\xi_2-\xi_1)^2}{2(\tau_2-\tau_1)}}\,d\xi_1 d\xi_2\right)d\tau_1 d\tau_2 \leq \frac{2T(y-x)^2}{\pi}.$$

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

We now define

$$\ell(t,y) = 2\big(B(t)\vee y - B(t)\vee 0 - I(t,y)\big).$$

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),

$$f(x) = f(0) + \int_0^t f'\big(B(\tau)\big)\,dB(\tau) + \tfrac{1}{2}\int_0^t \varphi\big(B(\tau)\big)\,d\tau.$$

At the same time,

$$\tfrac{1}{2}\int \varphi(y)\ell(t,y)\,dy = f\big(B(t)\big) - f(0) - \int \varphi(y)I(t,y)\,dy,$$

and because φ and $I(t,\,\cdot\,)$ are continuous and φ has compact support,

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

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

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

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

for each $\varphi \in C_c(\mathbb{R}^N;\mathbb{R})$. 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

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

and we will now show

$$\mathbb{P}\Big(\ell\big(\{t:\,B(t)\neq y\},y\big) = 0\Big) = 1, \tag{62}$$

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 fucntion and $(t,x) \rightsquigarrow f(t,x,\omega)$ is continuous for each $\omega \in \Omega$, then, by using Riemann sum approximations, one can see that

$$\int \left(\int_0^t f(\tau,x,\omega)\,\ell(d\tau,x)(\omega) \right) dx = \int_0^t f\big(\tau,B(\tau)(\omega),\omega\big)\,d\tau \tag{*}$$

for $\mathbb{P}$-almost every ω. Now choose continuous functions $\rho:\mathbb{R}\longrightarrow[0,\infty)$ and $\eta:\mathbb{R}\longrightarrow[0,1]$ 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 $\rho_\epsilon(x) = \epsilon^{-1}\rho\big(\epsilon^{-1}x\big)$ and $\eta_R(x) = \eta(Rx)$, and set

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

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

$$\int \rho_\epsilon(x-y) \left(\int \eta_R\big(B(\tau)-y\big)\,\ell(d\tau,x) \right) dx = 0$$

$\mathbb{P}$-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, $t \rightsquigarrow \ell(t,y)$ 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

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

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

$$B(t)\wedge 0 = B(t) - \int_0^t \mathbf{1}_{[0,\infty)}\big(B(\tau)\big)\,dB(\tau) - \tfrac{1}{2}\ell(t,0) = \int_0^t \operatorname{sgn}\big(B(\tau)\big)\,dB(\tau) - \tfrac{1}{2}\ell(t,0),$$

and

$$|B(t)| = \tilde{B}(t) + \ell(t,0) \text{ where } \tilde{B}(t) \equiv \int_0^t \operatorname{sgn}\big(B(\tau)\big). \tag{63}$$

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

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

is a martingale, and conclude that $\big(\tilde{B}(t), \mathcal{F}_t, \mathbb{P}\big)$ 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 $t \rightsquigarrow |B(t)|$ 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

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

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

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

and therefore both $\ell(t,0)$ and $\tilde{B}(t)$ are $\overline{\sigma\big(\{|B(\tau)| : \tau \in [0,t]\}\big)}$-measurable. In particular, this means that $\operatorname{sgn}\big(B(t)\big)$ 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

$$X(t) = \int_0^t \operatorname{sgn}\big(X(\tau)\big)\,d\tilde{B}(\tau). \tag{*}$$

To see this, all that one has to show is that

$$\int_0^t \xi(\tau)\, d\tilde{B}(\tau) = \int_0^t \xi(\tau)\mathrm{sgn}\big(B(\tau)\big)\, dB(\tau)$$

for bounded $\xi \in \mathcal{P}^2_{\mathrm{loc}}(\mathbb{R})$. 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 stochistic 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 $y \neq 0$,*

$$\mathbb{P}\big(\ell(t,y) > s\big) = \left(\frac{2y^2}{\pi}\right)^{\frac{1}{2}} \int_0^t e^{-\frac{y^2}{2\tau}}\mathbb{P}\big(\ell(t-\tau) > s\big)\, d\tau.$$

In addition,

$$\mathbb{P}\big(\ell(t,0) > 0\big) = 1 \text{ for all } t > 0, \tag{65}$$

and so

$$\mathbb{P}\big(\ell(t,y) > 0\big) = 2P\big(B(t) > y\big).$$

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

Next, set $\zeta_y = \inf\{t \geq 0 :\ B(t) \geq y\}$. Then $\zeta_y < t$ implies that

$$L\big(t,[y,y{+}\delta]\big) = \int_{t\wedge\zeta_y}^t \mathbf{1}_{[y,y+\delta]}\big(B(\tau)\big)\, d\tau = \int_0^{t-t\wedge\zeta_y} \mathbf{1}_{[0,\delta]}\big(B(\tau{+}t{\wedge}\zeta_y){-}B(t{\wedge}\zeta_y)\big)\, d\tau.$$

Since

$$\big(B(\tau + t\wedge\zeta_y) - B(t\wedge\zeta_y), \mathcal{F}_{\tau+\zeta_y}, \mathbb{P}\big)$$

is a Brownian motion that is independent of $\mathcal{F}_{t\wedge\zeta_y}$, it follows that, for any bounded, continuous $f : [0,\infty) \longrightarrow \mathbb{R}$,

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\Big[f\Big(\tfrac{1}{\delta}L\big(t,[y,y+\delta]\big)\Big)\Big]\\ &\quad = \int_{\{\zeta_y<t\}}\left(\int f\Big(\tfrac{1}{\delta}L\big(t-t\wedge\zeta_y(\omega_1),[0,\delta]\big)(\omega_2)\,\mathbb{P}(d\omega_2)\right)\mathbb{P}(d\omega_1).\end{aligned}$$

Hence, after letting $\delta \searrow 0$, we have

$$\mathbb{E}^{\mathbb{P}}\big[f\big(\ell(t,y)\big)\big] = \int_{\{\zeta_y<t\}} \left(\int f\big(\ell\big(t-\zeta_y(\omega_1),0\big)(\omega_2)\big)\,\mathbb{P}(d\omega_2)\right)\mathbb{P}(d\omega_1).$$

Starting from here and using the fact that $\mathbb{P}(\zeta_y < t) = 2\mathbb{P}\big(B(t) > y\big)$, 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

$$\mathbb{P}\big(\exists \tau \in (0,t)\; B(t) > 0\big) = \lim_{y\searrow 0} \mathbb{P}\big(\zeta_y < t\big) = \left(\frac{2}{\pi t}\right)^{\frac{1}{2}} \int_0^\infty e^{-\frac{x^2}{2t}}\,dx = 1. \quad \square$$

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 $N_\epsilon(t)$ of times before t that $B(\,\cdot\,)$ returns to 0 after leaving $(-\epsilon,\epsilon)$. That is, $N_\epsilon(t) \ge n$ if and only if there exist $0 < \tau_1 < \cdots < \tau_{2n} \le t$ such that $|B(\tau_{2m-1})| \ge \epsilon$ and $|B(\tau_{2m}| = 0$ for $1 \le m \le n$. He then showed that

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

One way to prove (66) is to use (63). Indeed, if $\zeta_0 = 0$ and, for $m \ge 1$,

$$\zeta_{2m=1} = \inf\{t \ge \zeta_{2m-2} : |B(t)| \ge \epsilon\} \text{ and } \zeta_{2m} = \inf\{t \ge \zeta_{2m-1} : |B(t)| = 0\},$$

then, because $B(\,\cdot\,)$ is continuous, $\zeta_m \ge t$ for all but a finite number of m's, and

$$\sum_{m\ge 1} \big(|B(\zeta_{2m-1}\wedge t)| - |B(\zeta_{2m}\wedge t)|\big) = \epsilon N_\epsilon(t) + \big(\epsilon - |B(t)|\big) \sum_{m\ge 1} \mathbf{1}_{(\zeta_{2m-1},\zeta_{2m})}(t).$$

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

$$\begin{aligned}\sum_{m\ge 1} \big(|B(\zeta_{2m-1})| - |B(\zeta_{2m})|\big) &= \sum_{m\ge 1} \big(\tilde{B}(\zeta_{2m-1}) - \tilde{B}(\zeta_{2m})\\ &= -\tilde{B}(t) + \sum_{m\ge 0} \big(\tilde{B}(\zeta_{2m+1}) - \tilde{B}(\zeta_{2m})\big) = -|B(t)| + \ell(t,0) + \int_0^t \xi(\tau)\,dB(\tau),\end{aligned}$$

where

$$\xi(\tau) = \sum_{m\ge 0} \operatorname{sgn}\big(B(\tau)\big)\mathbf{1}_{[\zeta_{2m},\zeta_{2m+1}]}(\tau).$$

After combining these, we see that

$$\epsilon N_\epsilon(t)-\ell(t,0) = -|B(t)|\sum_{m\geq 0}\mathbf{1}_{[\zeta_{2m},\zeta_{2m+1}]}(t)-\epsilon\sum_{m\geq 1}\mathbf{1}_{(\zeta_{2m-1},\zeta_{2m})}(t)+\int_0^t \xi(\tau)\,dB(\tau).$$

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 ϵ and $|\xi(\tau)|\leq\mathbf{1}_{[0,\epsilon]}\big(|B(\tau)|\big)$. Hence,

$$\mathbb{E}^{\mathbb{P}}\big[\|\epsilon N_\epsilon(\,\cdot\,)-\ell(\,\cdot\,,0)\|_{[0,t]}^2\big]^{\frac12}\leq\epsilon+2\mathbb{E}^{\mathbb{P}}\left[\int_0^t\mathbf{1}_{[0,\epsilon]}\big(B(\tau)\big)\,d\tau\right]^{\frac12},$$

which means that there is a $C<\infty$ such that

$$\mathbb{E}^{\mathbb{P}}\big[\|\epsilon N_\epsilon(\,\cdot\,)-\ell(\,\cdot\,,0)\|_{[0,t]}^2\big]^{\frac12}\leq C(t\epsilon)^{\frac12}\text{ for }\epsilon\in(0,1].$$

In particular,

$$\sum_{k\geq1}\mathbb{E}^{\mathbb{P}}\big[\|k^{-4}N_{k^{-4}}(\,\cdot\,)-\ell(\,\cdot\,,0)\|_{[0,t]}^2\big]^{\frac12}<\infty,$$

and so $\|k^{-4}N_{k^{-4}}(\,\cdot\,)-\ell(\,\cdot\,,0)\|_{[0,t]}\longrightarrow 0$ (a.s.,$\mathbb{P}$). Finally, note that if $(k+1)^{-4}\leq\epsilon\leq k^{-4}$, then

$$(k+1)^{-4}N_{k^{-4}}(\,\cdot\,)\leq\epsilon N_\epsilon(\,\cdot\,)\leq k^{-4}N_{k+1^{-4}}(\,\cdot\,),$$

and thereby conclude that (66) holds.

In conjuntion with Lemma 64, (66) gives abundant, quantitative evidence that Brownian paths are fuzzy. Indeed, it says that, for any $s\geq0$ and $\delta>0$, with probabiliy 1, $B(\,\cdot\,)$ leaves and revisits $B(s)$ infinitely often before time $s+\delta$.

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

Recall the Euler approximants $X_n(t,\mathbf{x})$, which are the solutions to

$$X_n(t,\mathbf{x})=\mathbf{x}+\int_0^t\sigma\big(X_n([\tau]_n)\big)\,dw(\tau)+\int_0^t b\big(X_n([\tau]_n)\big)\,d\tau,$$

and set $\Delta_n(t,\mathbf{x})=X(t,\mathbf{x})-X_n(t,\mathbf{x})$. 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

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[\|\Delta_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^p\big]^{\frac1p}\leq C_p(t)(1+|\mathbf{x}|)2^{-\frac n2},\\
&\mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x})-X_n(s,\mathbf{x})\|_{[s,t]}^p\big]^{\frac1p}\leq C_p(t)(1+|\mathbf{x}|)(t-s)^{\frac12},\\
&\mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{y})-X_n(\,\cdot\,,\mathbf{x})\|_{[s,t]}^p\big]^{\frac1p}\leq C_p(t)|\mathbf{y}-\mathbf{x}|
\end{aligned}\tag{67}$$

for $n \geq 0$, $0 \leq s < t$, and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$. From these it was clear the $\|\Delta_n(\,\cdot\,,\mathbf{x})\|_{[0,t]} \longrightarrow 0$ both (a.s.,$\mathbb{P}$) and in $L^p(\mathcal{W};\mathbb{R}^N)$, but I now want to use Kolmogorov's continuity criterion to show that this convergence is uniform with respect to $\mathbf{x}$ in compact subsets. This end, note that

$$
\begin{aligned}
&|\Delta_n(t,\mathbf{y}) - \Delta_n(t,\mathbf{x})|^p \\
&\qquad = \big|\big(X(t,\mathbf{y}) - X(t,\mathbf{x})\big) - \big(X_n(t,\mathbf{y}) - X_n(t,\mathbf{x})\big)\big|^{\frac{p}{2}}|\Delta_n(t,\mathbf{y}) - \Delta_n(t,\mathbf{x})|^{\frac{p}{2}},
\end{aligned}
$$

apply Schwarz's inequality to get

$$
\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[\|\Delta_n(\,\cdot\,,\mathbf{y}) - \Delta_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^p\big] \\
&\quad \leq \mathbb{E}^{\mathcal{W}}\big[\|\big(X(\,\cdot\,,\mathbf{y}) - X(\,\cdot\,,\mathbf{x})\big) - \big(X_n(\,\cdot\,,\mathbf{y}) - X_n(\,\cdot\,,\mathbf{x})\big)\|_{[0,t]}^p\big]^{\frac{1}{2}} \\
&\qquad \times \mathbb{E}^{\mathcal{W}}\big[\|\Delta_n(\,\cdot\,,\mathbf{y}) - \Delta_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^p\big]^{\frac{1}{2}},
\end{aligned}
$$

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

$$
\mathbb{E}^{\mathcal{W}}\big[\|\Delta_n(\,\cdot\,,\mathbf{y}) - \Delta_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^p\big]^{\frac{1}{p}} \leq K_p(t,R)2^{-\frac{n}{4}}|\mathbf{y} - \mathbf{x}|^{\frac{1}{2}} \text{ for } \mathbf{x}, \mathbf{y} \in [-R,R]^N.
$$

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

$$
\sup_{n \geq 0} 2^{\frac{n}{4}} \mathbb{E}^{\mathcal{W}}\big[\|\Delta_n\|_{[0,t]\times[-R,R]^N}^p\big] < \infty
$$

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

In order to prove the differentiablity 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*

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

and

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

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

$$
\begin{gathered}
\frac{\|\tilde{\sigma}_n(t,\mathbf{x},\tilde{\mathbf{x}})\|_{\mathrm{H.S.}} \vee |\tilde{b}_n(t,\mathbf{x},\tilde{\mathbf{x}})|}{1 + |\tilde{\mathbf{x}}|} \leq C, \\
\frac{\|\tilde{\sigma}_n(t,\mathbf{x},\tilde{\mathbf{y}}) - \tilde{\sigma}_n(t,\mathbf{x},\tilde{\mathbf{x}})\|_{\mathrm{H.S.}} \vee |\tilde{b}_n(t,\mathbf{x},\tilde{\mathbf{y}}) - \tilde{b}_n(t,\mathbf{x},\tilde{\mathbf{x}})|}{|\tilde{\mathbf{y}} - \tilde{\mathbf{x}}|} \leq C
\end{gathered}
$$

for all $n \geq 0$, $(t, \mathbf{x}, \tilde{\mathbf{x}}) \in [0,\infty) \times \mathbb{R}^N \times \mathbb{R}^{\tilde{N}}$, *and* $\tilde{\mathbf{y}} \neq \tilde{\mathbf{x}}$. *In addition, assume that for each* $p \in [1,\infty)$ *there is a non-decreasing function* $t \rightsquigarrow C_p(t)$ *such that*

$$\frac{\mathbb{E}^{\mathcal{W}}\big[\|\tilde{\sigma}_{n+1}(t,\mathbf{x},\tilde{\mathbf{x}}) - \tilde{\sigma}_n(t,\mathbf{x},\tilde{\mathbf{x}})\|^p_{\mathrm{H.S.}} \vee |\tilde{b}_{n+1}(t,\mathbf{x},\tilde{\mathbf{x}}) - \tilde{b}_n(t,\mathbf{x},\tilde{\mathbf{x}})|^p\big]^{\frac{1}{p}}}{1 + |\mathbf{x}| + |\tilde{\mathbf{x}}|} \leq C_p(t),$$

and

$$\begin{aligned} &\frac{\mathbb{E}^{\mathcal{W}}\big[\|\tilde{\sigma}_n(\,\cdot\,,\mathbf{y},\tilde{\mathbf{x}}) - \tilde{\sigma}_n(\,\cdot\,,\mathbf{x},\tilde{\mathbf{x}})\|^p_{[0,t]} \vee \|\tilde{b}_n(\,\cdot\,,\mathbf{y},\tilde{\mathbf{y}}) - \tilde{b}_n(\,\cdot\,,\mathbf{x},\tilde{\mathbf{x}})\|^p_{[0,t]}\big]^{\frac{1}{p}}}{1 + |\tilde{\mathbf{x}}|} \\ &\qquad \leq C_p(t)|\mathbf{y} - \mathbf{x}| \end{aligned}$$

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

$$\begin{aligned} \frac{\mathbb{E}^{\mathcal{W}}\big[\|\tilde{Y}_n(\,\cdot\,,\mathbf{x})\|^p_{[0,t]}\big]^{\frac{1}{p}}}{1 + |\mathbf{x}|^k} &\leq C_p(t), \\ \frac{\mathbb{E}^{\mathcal{W}}\big[\|\tilde{Y}_n(\,\cdot\,,\mathbf{y}) - \tilde{Y}_n(\,\cdot\,,\mathbf{x})\|^p_{[0,t]}\big]^{\frac{1}{p}}}{1 + |\mathbf{x}|^k + |\mathbf{y}|^k} &\leq C_p(t)|\mathbf{y} - \mathbf{x}|, \\ \frac{\mathbb{E}^{\mathcal{W}}\big[\|\tilde{Y}_n(\,\cdot\,,\mathbf{x}) - \tilde{Y}_n(s,\mathbf{x})\|_{[s,t]}\big]^{\frac{1}{p}}}{1 + |\mathbf{x}|^k} &\leq C_p(t)(t-s)^{\frac{1}{2}} \end{aligned}$$

and

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

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

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

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

$$\begin{aligned} \frac{\mathbb{E}^{\mathcal{W}}\big[\|\tilde{X}_n(\,\cdot\,,\mathbf{x})\|^p_{[0,t]}\big]^{\frac{1}{p}}}{1 + |\mathbf{x}|^k} &\leq \tilde{C}_p(t), \\ \mathbb{E}^{\mathcal{W}}\big[\|\tilde{\Delta}_n(\,\cdot\,,\mathbf{x})\|^p_{[0,t]}\big]^{\frac{1}{p}} &\leq \tilde{C}_p(t)(1 + |\mathbf{x}|^k)2^{-\frac{n}{2}}, \\ \mathbb{E}^{\mathcal{W}}\big[\|\tilde{X}_n(\,\cdot\,,\mathbf{x}) - \tilde{X}_n(s,\mathbf{x})\|^p_{[s,t]}\big]^{\frac{1}{p}} &\leq \tilde{C}_p(t)(1 + |\mathbf{x}|^k)(t-s)^{\frac{1}{2}}, \\ \mathbb{E}^{\mathcal{W}}\big[\|\tilde{X}_n(\,\cdot\,,\mathbf{y}) - \tilde{X}_n(\,\cdot\,,\mathbf{x})\|^p_{[s,t]}\big]^{\frac{1}{p}} &\leq \tilde{C}_p(t)|\mathbf{y} - \mathbf{x}| \end{aligned}$$

for all $n \geq 0$, $0 \leq s < t$, *and* $\mathbf{x},\mathbf{y} \in \mathbb{R}^N$.

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),

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[\|\tilde{X}_n(\,\cdot\,,\mathbf{x})-\tilde{X}_n(s,\mathbf{x})\|_{[s,t]}^p\big]\\
&\qquad\le 3^{p-1}K_p\mathbb{E}^{\mathcal{W}}\left[\left(\int_s^t \|\tilde{\sigma}_n\big(\tau,\mathbf{x},\tilde{X}_n([\tau]_n,\mathbf{x})\big)\|_{\mathrm{H.S.}}^2\,d\tau\right)^{\frac{p}{2}}\right]\\
&\qquad\quad+3^{p-1}\mathbb{E}^{\mathcal{W}}\left[\left(\int_s^t \big|\tilde{b}_n\big(\tau,\mathbf{x},\tilde{X}_n([\tau]_n,\mathbf{x})\big)\big|\,d\tau\right]^p\right]\\
&\qquad\qquad+3^{p-1}\mathbb{E}^{\mathcal{W}}\big[\|\tilde{Y}_n(\,\cdot\,,\mathbf{x})-\tilde{Y}_n(s,\mathbf{x})\|_{[s,t]}^p\big].
\end{aligned}$$

When $s=0$, this says that

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[\|\tilde{X}_n(\,\cdot\,,\mathbf{x})\|_{[s,t]}^p\big]\\
&\qquad\le 3^{p-1}(1+K_p)(t^{\frac{p}{2}}+t^{p-1})C\int_0^t \mathbb{E}^{\mathcal{W}}\big[1+\|\tilde{X}_n(\,\cdot\,,\mathbf{x})\|_{[0,\tau]}^p\big]\,d\tau\\
&\qquad\quad+6^{p-1}C_p(t)^p(1+|\mathbf{x}|^k)^p,
\end{aligned}$$

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

$$\sup_{n\ge 0}\sup_{\mathbf{x}\in\mathbb{R}^N}\frac{\mathbb{E}^{\mathcal{W}}\big[\|\tilde{X}_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^p\big]}{(1+|\mathbf{x}|^k)^p}<\infty.$$

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

The proofs of the other results are left as an exercise. $\square$

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

$$\begin{aligned}
\frac{\partial X_n(t,\mathbf{x})_i}{\partial x_j}=&\frac{\partial X_n(m2^{-n},\mathbf{x})_i}{\partial x_j}\\
&+\sum_{\ell=1}^N\sum_{k=1}^M\frac{\partial\sigma\big(X_n(m2^{-n},\mathbf{x})\big)_{i,k}}{\partial x_\ell}\frac{\partial X_n(m2^{-n},\mathbf{x})_\ell}{\partial x_j}\big(w(t)-w(m2^{-n})\big)_k\\
&\quad+\sum_{\ell=1}^N\frac{\partial b\big(X_n(m2^{-n},\mathbf{x})\big)_i}{\partial x_\ell}\frac{\partial X_n(m2^{-n},\mathbf{x})_\ell}{\partial x_j}(t-m2^{-n})
\end{aligned}$$

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

$$\sigma_n^{(1)}:[0,\infty)\times\mathbb{R}^N\times(\mathbb{R}^N\otimes\mathbb{R}^N)\times C\big([0,\infty);\mathbb{R}^M\big)\longrightarrow\mathrm{Hom}\big(\mathbb{R}^M;\mathbb{R}^N\otimes\mathbb{R}^N\big)$$

and

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

by

$$\sigma_n^{(1)}(t,\mathbf{x},\tilde{\mathbf{x}})_{(ij),k}=\sum_{\ell=1}^N\frac{\partial\sigma_{i,k}}{\partial x_\ell}\big(X_n([t]_n,\mathbf{x})\big)x_{\ell j}^{(1)}$$

and

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

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

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

where $X_n^{(1)}(t,\mathbf{x})_{(ij)}=\frac{\partial X_n(t,\mathbf{x})_i}{\partial x_j}$. As a consequence, the $X_n^{(1)}$'s have the properties proved for the $\tilde{X}_n$'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

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

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

$$\begin{aligned}X^{(1)}(t,\mathbf{x})_{ij}=\mathbf{I}+&\sum_{k=1}^M\sum_{\ell=1}^N\int_0^t\frac{\partial\sigma_{i,k}}{\partial x_\ell}\big(X(\tau,\mathbf{x})\big)X^{(1)}(\tau,\mathbf{x})_{\ell j}\,dw(\tau)_k\\&+\sum_{\ell=1}^N\int_0^t\frac{\partial b_i}{\partial x_\ell}\big(X(\tau,\mathbf{x})\big)X^{(1)}(\tau,\mathbf{x})_{\ell j)}\,d\tau,\end{aligned}\tag{69}$$

and, almost surely, $X_n^{(1)}\longrightarrow X^{(1)}$ 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 $X^{(1)}(t,\mathbf{x})$ is almost surely equal to its Jacobian matrix. In addition, for each $p\in[2,1hi)$, there exists a non-decreasing map $t\in[0,\infty)\longmapsto C^{(1)}(t)\in[0,\infty)$ such that

$$\begin{aligned}&\mathbb{E}^{\mathcal{W}}\big[\|X_{n+1}^{(1)}(\,\cdot\,,\mathbf{x})-X_n^{(1)}(\,\cdot\,,\mathbf{x})\|_{[0,t]}^p\big]^{\frac{1}{p}}\le C_p^{(1)}(t)2^{-\frac{n}{2}},\\&\mathbb{E}^{\mathcal{W}}\big[\|X_n^{(1)}(\,\cdot\,,\mathbf{x})-X_n^{(1)}(s,\mathbf{x})\|_{[s,t]}^p\big]^{\frac{1}{p}}\le C_p^{(1)}(t)(t-s)^{\frac{1}{2}},\\&\mathbb{E}^{\mathcal{W}}\big[\|X_n^{(1)}(\,\cdot\,,\mathbf{y})-X_n^{(1)}(\,\cdot\,,\mathbf{x})\|_{[s,t]}^p\big]^{\frac{1}{p}}\le C_p^{(1)}(t)|\mathbf{y}-\mathbf{x}|.\end{aligned}\tag{70}$$

Next assume that σ and b have three continuous derivatives, all of which are bounded. Then the X_n have three continuous spacial derivatives and

$$X_n^{(2)}(t,\mathbf{x}) = \int_0^t \sigma_n^{(2)}\big(\tau,\mathbf{x},X_n^{(2)}([\tau]_n,\mathbf{x})\big)\,dw(\tau) \\ + \int_0^t b_n^{(2)}\big(\tau,\mathbf{x},X_n^{(2)}([\tau]_n,\mathbf{x})\big)\,d\tau + Y_n^{(2)}(t,\mathbf{x}),$$

where $X_n^{(2)}(t,\mathbf{x})_{ij_1j_2} = \frac{\partial^2 X_n(t,\mathbf{x})_i}{\partial x_{j_1}\partial x_{j_2}}$,

$$\sigma_n^{(2)}(t,\mathbf{x},\mathbf{x}^{(2)})_{(ij_1j_2),k} = \sum_{\ell=1}^N \frac{\partial \sigma_{ik}}{\partial x_\ell}(X_n([\tau]_n,\mathbf{x})\mathbf{x}^{(2)}_{\ell j_1 j_2},$$

$$b_n^{(2)}(t,\mathbf{x},\mathbf{x}^{(2)})_{ij_1j_2} = \sum_{\ell=1}^N \frac{\partial b_i}{\partial x_\ell}(X_n([\tau]_n,\mathbf{x})\mathbf{x}^{(2)}_{\ell j_1 j_2},$$

and $Y_n^{(2)}(t,\mathbf{x})_{ij_1j_2}$ equals

$$\sum_{k=1}^M \sum_{\ell_1,\ell_2=1}^N \int_0^t \frac{\partial^2 \sigma_{ik}}{\partial x_{\ell_1}\partial x_{\ell_2}}\big(X_n([\tau]_n,\mathbf{x})\big)X_n^{(1)}([\tau]_n,\mathbf{x})_{\ell_1 j_1}X_n^{(1)}([\tau]_n,\mathbf{x})_{\ell_2 j_2}\,dw(\tau)_k \\ + \sum_{\ell_1,\ell_2=1}^N \int_0^t \frac{\partial^2 b_i}{\partial x_{\ell_1}\partial x_{\ell_2}}\big(X_n([\tau]_n,\mathbf{x})\big)X_n^{(1)}([\tau]_n,\mathbf{x})_{\ell_1 j_1}X_n^{(1)}([\tau]_n,\mathbf{x})_{\ell_2 j_2}\,d\tau.$$

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 $\frac{\partial^2 X(t,\mathbf{x})_i}{\partial x_{j_1}\partial x_{j_2}} = X^{(2)}(t,\mathbf{x})_{ij_1j_2}$ almost surely and

$$\begin{aligned} X^{(2)}(t,\mathbf{x})_{ij_1j_2} = &\sum_{k=1}^M\sum_{\ell=1}^N \int_0^t \frac{\partial\sigma_{ik}}{\partial x_\ell}\big(X(\tau,\mathbf{x})\big)X^{(2)}(\tau,\mathbf{x})_{\ell j_1 j_2}\,dw(\tau)_k \\ &+ \sum_{\ell=1}^N \int_0^t \frac{\partial b_i}{\partial x_\ell}\big(X(\tau,\mathbf{x})\big)X^{(2)}(\tau,\mathbf{x})_{\ell j_1 j_2}\,d\tau \\ &+ \sum_{k=1}^M \sum_{\ell_1,\ell_2=1}^N \int_0^t \frac{\partial^2\sigma_{ik}}{\partial x_{\ell_1}\partial x_{\ell_2}}\big(X(\tau,\mathbf{x})\big)X^{(1)}(\tau,\mathbf{x})_{\ell_1 j_1}X^{(1)}(\tau,\mathbf{x})_{\ell_2 j_2}\,dw(\tau)_k \\ &+ \sum_{\ell_1,\ell_2=1}^N \int_0^t \frac{\partial^2 b_i}{\partial x_{\ell_1}\partial x_{\ell_2}}\big(X(\tau,\mathbf{x})\big)X^{(1)}(\tau,\mathbf{x})_{\ell_1 j_1}X^{(1)}(\tau,\mathbf{x})_{\ell_2 j_2}\,d\tau. \end{aligned}$$

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 σ 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 nth 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 φ's, and we can now show that it does if σ and b have three bounded derivatives. Indeed, if $\varphi \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$ and $u_\varphi(t,\mathbf{x}) = \mathbb{E}^{\mathcal{W}}\big[\varphi\big(X(t,\mathbf{x})\big]$, then, for any $1 \leq k \leq N$,

$$u_\sigma(t,\mathbf{x})(t,\mathbf{x}+h\mathbf{e}) - u_\sigma(t,\mathbf{x}) = \mathbb{E}^{\mathcal{W}}\big[\varphi\big(X(t,\mathbf{x}+h\mathbf{e}_k)\big) - \varphi\big(X(t,\mathbf{x})\big)\big]$$

where $(\mathbf{e}_k)_j = \delta_{k,j}$. Further,

$$\varphi\big(X(t,\mathbf{x}+h\mathbf{e}_k)\big) - \varphi\big(X(t,\mathbf{x})\big) = \sum_{\ell=1} \int_0^h \partial_{x_\ell}\varphi\big((t,\mathbf{x}+\xi\mathbf{e}_k)\big)X^{(1)}(t,\mathbf{x}+\xi\mathbf{e}_k)_{\ell j}\,d\xi,$$

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

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

Similarly, $\partial_{x_{j_1}}\partial_{x_{j_2}}u_\varphi(t,\mathbf{x})$ exists, is continuous, and is given by

$$\begin{aligned}\sum_{i_1 i_2=1}^N &\mathbb{E}^{\mathcal{W}}\big[\partial_{x_{i_1}}\partial_{x_{i_1}}\varphi\big(X(t,\mathbf{x})X^{(1)}(t,\mathbf{x})_{i_1 j_1}X^{(1)}(t,\mathbf{x})_{i_2 j_2}\big]\\ &+\sum_{i=1}^N \mathbb{E}^{\mathcal{W}}\big[\partial_{x_i}\varphi\big(X(t,\mathbf{x})\big)X^{(2)}(t,\mathbf{x})_{ij_1j_2}\big].\end{aligned}$$

Hence, since, by the Markov property,

$$\begin{aligned}u_\varphi(t+h,\mathbf{x}) - u(t,\mathbf{x}) &= \mathbb{E}^{\mathcal{W}}\big[u_\varphi\big(t,X(h,\mathbf{x})\big) - u_\varphi(t,\mathbf{x})\big]\\ &= \int_0^h \mathbb{E}^{\mathcal{W}}\big[Lu\big(t,X(\xi,\mathbf{x})\big)\big]\,d\xi,\end{aligned}$$

it follows that $\partial_t u_\varphi(t,\mathbf{x})$ exist, is continuous, and is equal to $Lu_\varphi(t,\mathbf{x})$. Since two finite Borel measures on $\mathbb{R}^N$ are equal if their integrals of every $\varphi \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$

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 minumum principle, O. Olenik proved that, when $\varphi \in C^2_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R})$, Kolmogorov's backward equation can 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 σ is a serious drawback. In Corollary 115 below, using a purely probabilistic argument, we will prove that the martingale problem for L has unique solutions if $a^{\frac{1}{2}}$ and b are Lipschitz continuous, which, when combined with Lemma 29, covers the cases to which Olenik's result applies.

Itô's Exponential

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

$$X(t) = 1 + \int_0^t X(\tau)\big(\boldsymbol{\xi}(\tau), dB(\tau)\big)_{\mathbb{R}^N} + \int_0^t X(\tau)\Phi(\tau)\,d\tau, \tag{71}$$

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

$$E_{\boldsymbol{\xi},\Phi}(t) = \exp\left(\int_0^t \big(\boldsymbol{\xi}(\tau), dB(\tau)\big)_{\mathbb{R}^N} + \tfrac{1}{2}\int_0^t \big(\Phi(\tau) - \tfrac{1}{2}|\boldsymbol{\xi}(\tau)|^2\,d\tau\right) \tag{72}$$

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

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

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

$$\begin{aligned}\Bigg(E_{\boldsymbol{\xi},0}(t)\bigg(E_{\mathbf{0},\Phi}(t)\varphi\big(V(t,I_\sigma(t)\big) - \int_0^t E_{\mathbf{0},\Phi}(\tau)\Big(\Phi(\tau)\varphi\big(V(\tau),I_\sigma(\tau)\big)\\ + \Big(\sigma(\tau)\boldsymbol{\xi}(\tau),\nabla_{(2)}\varphi\big(V(\tau),I_\sigma(\tau)\big)\Big)_{\mathbb{R}^{N_2}} + \Big(\nabla_{(1)}\varphi\big(V(\tau),I_\sigma(\tau)\big),dV(\tau)\Big)_{\mathbb{R}^{N1}}\\ + \tfrac{1}{2}\mathrm{Trace}\Big(\sigma\sigma^\top\nabla^2_{(2)}\varphi\big(V(\tau),I_\sigma(\tau)\big)\Big)\,d\tau\Big),\mathcal{F}_t,\mathbb{P}\Bigg)\end{aligned}$$

is a local martingale.

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

By (57),

$$\begin{aligned} M(t) \equiv \varphi\big(V(t, I_\sigma(t)\big) & \\ &- \int_0^t \Big(\big(\nabla_{(1)}\varphi\big(V(\tau), I_\sigma(\tau)\big), \dot V(\tau)\big)_{\mathbb{R}^{N_2}} \\ &\qquad + \tfrac12 \mathrm{Trace}\Big(\sigma\sigma^\top \nabla^2_{(2)}\varphi\big(V(\tau), I_\sigma(\tau)\big)\Big)\, d\tau \\ &= \int_0^t \big(\sigma(\tau)^\top \nabla_{(2)}\varphi\big(V(\tau), I_\sigma(\tau), dB(\tau)\big)_{\mathbb{R}^{N_2}}, \end{aligned}$$

and therefore, again by (57),

$$\begin{aligned} \Bigg(E_{\boldsymbol{\xi},\Phi}(t)M(t) - \int_0^t E_{\boldsymbol{\xi},\Phi}(\tau)\Big(\Phi(\tau)\varphi\big(V(\tau), I_\sigma(\tau)\big) & \\ + \Big(\sigma(\tau)\boldsymbol{\xi}(\tau), \nabla_{(2)}\varphi\big(V(\tau), I_\sigma(\tau)\big)\Big)_{\mathbb{R}^{N_2}}\Big)\, d\tau, \mathcal{F}_t, \mathbb{P}\Bigg) \end{aligned}$$

is a martingale. Equivalently, if

$$\begin{aligned} H(t) = \Phi(t)\varphi\big(V(t), I_\sigma(t)\big) &+ \big(\nabla_{(1)}\varphi\Big(V(t), I_\sigma(t)\big), \dot V(t)\Big)_{\mathbb{R}^{N_1}} \\ &+ \Big(\sigma(t)\boldsymbol{\xi}(t), \nabla_{(2)}\varphi\big(V(t), I_\sigma(t)\Big)_{\mathbb{R}^{N_2}} \\ &\quad + \tfrac12 \mathrm{Trace}\Big(\sigma\sigma^\top(t)\nabla^2_{(2)}\varphi\big(V(t), I_\sigma(t)\big)\Big), \end{aligned}$$

then

$$\Big(E_{\boldsymbol{\xi},\Phi}(t)\varphi\big(V(t), I_\sigma(t)\big) - \int_0^t E_{\boldsymbol{\xi},\Phi}(\tau)H(\tau)\, d\tau, \mathcal{F}_t, \mathbb{P}\Big)$$

is a martingale, and therefore

$$\begin{aligned} &\mathbb{E}^{\mathbb{P}}\big[E_{\boldsymbol{\xi},\Phi}(t)\varphi\big(V(t), I_\sigma(t)\big) - E_{\boldsymbol{\xi},\Phi}(s)\varphi\big(V(s), I_\sigma(s)\big)\,\big|\,\mathcal{F}_s\big] \\ &= \mathbb{E}^{\mathbb{P}}\left[\int_s^t E_{\boldsymbol{\xi},\Phi}(\tau)H(\tau)\, d\tau \,\middle|\, \mathcal{F}_s\right] = \mathbb{E}^{\mathbb{P}}\left[E_{\boldsymbol{\xi},0}(t)\int_s^t E_{\mathbf{0}.\Phi}(\tau)H(\tau)\, d\tau \,\middle|\, \mathcal{F}_s\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[E_{\boldsymbol{\xi},0}(t)\int_0^t E_{\mathbf{0}.\Phi}(\tau)H(\tau)\, d\tau \,\middle|\, \mathcal{F}_s\right] - \mathbb{E}_{\boldsymbol{\xi},0}(s)\int_0^s E_{\mathbf{0},\Phi}(\tau)H(\tau)\, d\tau. \quad \square \end{aligned}$$

Most applications are to cases in which either $\boldsymbol{\xi} = \mathbf{0}$ or $\Phi = 0$. In the following section I will explain some of its applications when $\boldsymbol{\xi} = \mathbf{0}$, and the section after that deals with the case when $\Phi = 0$.

The Feynman-Kac Formula: To see how Lemma 73 gets applied when $\boldsymbol{\xi} = \mathbf{0}$, suppose that $X(\,\cdot\,,\mathbf{x})$ is the solution to (56) and that $F : \mathbb{R}^N \longrightarrow \mathbb{R}$ is a Borel measurable function that is bounded on compact subsets. Then, by taking $\Phi(t) = \int_0^t F\big(X(t,\mathbf{x})\big)\,d\tau$, we see that, for each $T > 0$,

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

is a local martingale if $u \in C^{1,2}\big((0,\infty)\times\mathbb{R}^N;\mathbb{R}\big) \cap C\big([0,\infty)\times\mathbb{R}^N;\mathbb{R}\big)$ 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

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

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

$$\partial_t u = \frac{1}{2}\Big(\partial_x^2 u - x^2 u\Big) \text{ on } (0,\infty)\times\mathbb{R} \text{ with } \lim_{t\searrow 0} u(t,\,\cdot\,) = \varphi. \tag{75}$$

Indeed, if

$$h(t,x,y) = e^{-\frac{t+x^2}{2}} g\big(1-e^{-2t}, y - e^{-t}x\big) e^{\frac{y^2}{2}} \text{ where } g(t,x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2}},$$

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

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

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

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

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

$$\exp\left(-\tfrac{\alpha}{2}\int_0^t \big(x + w(\tau)\big)^2\,d\tau\right) \text{ and } \exp\left(-\tfrac{1}{2}\int_0^{\alpha^{\frac{1}{2}} t} \big(\alpha^{\frac{1}{4}} x + w(\tau)\big)^2\,d\tau\right)$$

have the same distribution. Hence,

$$\mathbb{E}^{\mathcal{W}}\left[\exp\left(-\tfrac{\alpha}{2}\int_0^t \big(x+w(\tau)\big)^2\,d\tau\right)\right] = \big(\cosh(\alpha^{\frac12}t)\big)^{-\frac12}\exp\left(-\frac{\alpha^{\frac12}\tanh(\alpha^{\frac12}t)x^2}{2}\right). \tag{76}$$

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

$$\mathbb{E}^{\mathcal{W}}\left[\exp\left(-\frac{\alpha}{2}\left(\int_0^1 w(\tau)^2\,d\tau-\left(\int_0^1 w(\tau)\,d\tau\right)^2\right)\right)\right]=\sqrt{\frac{\alpha^{\frac12}}{\sinh\alpha^{\frac12}}},$$

and then, again using Brownian scaling,

$$\mathbb{E}^{\mathcal{W}}\Big[\exp\Big(-\frac{\alpha}{2}\mathrm{Var}_{[0,t]}\big(w(\,\cdot\,)\big)\Big)\Big]=\sqrt{\frac{(\alpha t)^{\frac12}}{\sinh(\alpha t)^{\frac12}}}$$
$$\text{where } \mathrm{Var}_{[0,t]}\big(w(\,\cdot\,)\big)=\frac{1}{t}\int_0^t w(\tau)^2\,d\tau-\left(\frac{1}{t}\int_0^t w(\tau)\,d\tau\right)^2 \tag{77}$$

is the variance of the path $w\restriction[0,t]$. From this it follows that

$$\mathcal{W}\Big(\mathrm{Var}_{[0,t]}\big(w(\,\cdot\,)\big)\ge\tfrac{1}{r}\Big)\le e^{-\frac{\alpha}{2r}-\frac{\sqrt{\alpha t}}{2}}\sqrt{\frac{2\sqrt{\alpha t}}{1-e^{-2\sqrt{\alpha t}}}}$$

for all $\alpha>0$, and so, by taking $\alpha=\frac{r^2t}{4}$, we have that

$$\mathcal{W}\Big(\mathrm{Var}_{[0,t]}\big(w(\,\cdot\,)\big)\ge\tfrac{1}{r}\Big)\le e^{-\frac{rt}{4}}\sqrt{\frac{rt}{1-e^{-rt}}}. \tag{78}$$

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 $\big(E_{\boldsymbol{\xi},0}(t),\mathcal{F}_t,\mathbb{P}\big)$ is actually a martingale and not just a local one. For this reason, the following criterion, which was discovered by A. Novikov, can helpful.

LEMMA 79. *For any $\boldsymbol{\xi}\in\mathcal{P}^2_{\mathrm{loc}}(\mathbb{R}^N)$, $\big(E_{\boldsymbol{\xi},0}(t),\mathcal{F}_t,\mathbb{P}\big)$ is a supermartingale, and it is a martingale if and only if $\mathbb{E}^{\mathbb{P}}\big[E_{\boldsymbol{\xi},0}(t)\big]=1$ for all $t\ge0$. Next, set*

$$A_{\boldsymbol{\xi}}(t)\equiv\int_0^t|\boldsymbol{\xi}(\tau)|^2\,d\tau.$$

If

$$\mathbb{E}^{\mathbb{P}}\big[e^{\frac12 A_{\boldsymbol{\xi}}(t)}\big]<\infty \text{ for all } t\ge0,$$

then $\big(E_{\boldsymbol{\xi},0}(t),\mathcal{F}_t,\mathbb{P}\big)$ is a martingale.

PROOF: Choose stopping times $\{\zeta_m : m \geq 1\}$ for $\big(E_{\boldsymbol{\xi},0}(t), \mathcal{F}_t, \mathbb{P}\big)$. Then

$$E_{\boldsymbol{\xi},0}(t\wedge\zeta_m) \longrightarrow E_{\boldsymbol{\xi},0}(t) \text{ (a.s.,}\mathbb{P}) \text{ and } \mathbb{E}^{\mathbb{P}}\big[E_{\boldsymbol{\xi},0}(t\wedge\zeta_m)\big] = 1$$

for all $t \geq 0$. Thus, since $\big(E_{\boldsymbol{\xi},0}(t\wedge\zeta_m), \mathcal{F}_t, \mathbb{P}\big)$ is a martingale for each $m \geq 1$, Fatou's lemma implies that $\big(E_{\boldsymbol{\xi},0}(t), \mathcal{F}_t, \mathbb{P}\big)$ is a supermartingale. Moreover, if $\mathbb{E}^{\mathbb{P}}\big[E_{\boldsymbol{\xi},0}(t)\big] = 1$, then $E_{\boldsymbol{\xi},0}(t\wedge\zeta_m) \longrightarrow E_{\boldsymbol{\xi},0}(t)$ in $L^1(\mathbb{P};\mathbb{R})$, and so $\big(E_{\boldsymbol{\xi},0}(t), \mathcal{F}_t, \mathbb{P}\big)$ 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 $m \geq 1$, $\mathbb{E}^{\mathbb{P}}\big[E_{\pm\boldsymbol{\xi},0}(t\wedge\zeta_m)\big] \leq 1$,

$$\mathbb{E}^{\mathbb{P}}\big[e^{\pm\frac{1}{2}I_{\boldsymbol{\xi}}(t\wedge\zeta_m)}\big] = \mathbb{E}^{\mathbb{P}}\big[E_{\pm\boldsymbol{\xi},0}(t\wedge\zeta_m)^{\frac{1}{2}}e^{\frac{1}{4}A_{\boldsymbol{\xi}}(t)}\big] \leq \Big(\mathbb{E}^{\mathbb{P}}\big[e^{\frac{1}{2}A_{\boldsymbol{\xi}}(t)}\big]\Big)^{\frac{1}{2}},$$

and therefore $\sup_{m\geq 1}\mathbb{E}^{\mathbb{P}}\big[e^{|\frac{1}{2}I_{\boldsymbol{\xi}}(t\wedge\zeta_m)|}\big] < \infty$ is integrable for all $t \geq 0$. In particular, this means that $\big(I_{\boldsymbol{\xi}}(t), \mathcal{F}_t, \mathbb{P}\big)$ is a martingale and therefore that $\big(e^{\alpha I_{\boldsymbol{\xi}}(t)}, \mathcal{F}_t, \mathbb{P}\big)$ 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 $\alpha(\lambda,p) = \frac{p\lambda(p-\lambda)}{1-\lambda^2}$. Then $E_{\lambda\boldsymbol{\xi},0}(t)^{p^2} = \big(e^{\alpha(\lambda,p)I_{\boldsymbol{\xi}}(t)}\big)^{1-\lambda^2}E_{p\boldsymbol{\xi},0}(t)^{\lambda^2}$, and so, by Hölder's inequality and the submartingale property of $e^{\alpha(\lambda,p)I_{\boldsymbol{\xi}}}$, for any stopping time ζ,

$$(*)\qquad \mathbb{E}^{\mathbb{P}}\big[E_{\lambda\boldsymbol{\xi},0}(t\wedge\zeta)^{p^2}\big] \leq \mathbb{E}^{\mathbb{P}}\big[e^{\alpha(\lambda,p)I_{\boldsymbol{\xi}}(t)}\big]^{1-\lambda^2}\mathbb{E}^{\mathbb{P}}\big[E_{p\boldsymbol{\xi},0}(t\wedge\zeta)\big]^{\lambda^2}.$$

Since $\mathbb{E}^{\mathbb{P}}\big[E_{p\boldsymbol{\xi},0}(t\wedge\zeta)\big] \leq 1$, by taking $p = p_\lambda$, we see that

$$\mathbb{E}^{\mathbb{P}}\big[E_{\lambda\boldsymbol{\xi},0}(t\wedge\zeta)^{p_\lambda^2}\big] \leq \mathbb{E}^{\mathbb{P}}\big[e^{\frac{1}{2}I_{\boldsymbol{\xi}}(t)}\big],$$

from which it follows that the local martingale $\big(E_{\lambda\boldsymbol{\xi},0}(t), \mathcal{F}_t, \mathbb{P}\big)$ is a martingale for each $\lambda \in (0,1)$. Finally, take $p = 1$ and $\zeta = t$. Then (*) combined with Jensen's inequality say that

$$\mathbb{E}^{\mathbb{P}}\big[E_{\lambda\boldsymbol{\xi},0}(t)\big] \leq \mathbb{E}^{\mathbb{P}}\big[e^{\frac{1}{2}I_{\boldsymbol{\xi}}(t)}\big]^{2\lambda(1-\lambda)}\mathbb{E}^{\mathbb{P}}\big[E_{\boldsymbol{\xi},0}(t)\big]^{\lambda^2}$$

for all $\lambda \in (0,1)$. Thus, because $\mathbb{E}^{\mathbb{P}}\big[E_{\lambda\boldsymbol{\xi},0}(t)\big] = 1$, after letting $\lambda \nearrow 1$, one sees that $\mathbb{E}^{\mathbb{P}}\big[E_{\boldsymbol{\xi},0}(t)\big] = 1$. $\square$

Novikov's criterion is useful but not definitive. Indeed, consider the case when $N = 1$ and $\xi(t) = B(t)$. Then

$$I_\xi(t) = \int_0^t B(\rho)\,dB(\tau) = \frac{B(t)^2 - t}{2} \text{ and } A_\xi(t) = \int_0^t B(\tau)^2\,d\tau.$$

As we showed $\mathbb{E}^{\mathbb{P}}\big[e^{\frac{1}{2}I_\xi}\big] \le \mathbb{E}^{\mathbb{P}}\big[e^{\frac{1}{2}A_\xi}\big]^{\frac{1}{2}}$, and so, since $\mathbb{E}^{\mathbb{P}}\big[e^{\frac{1}{4}B(t)^2}\big] = \infty$ if $t \ge 2$, $\mathbb{E}^{\mathbb{P}}\big[e^{\frac{1}{2}A_\xi}\big] = \infty$ for $t \ge 2$. Nonetheless, $\big(E_{\xi,0}(t), \mathcal{F}_t, \mathbb{P}\big)$ is a maringale. To see this, choose a sequence $\{\varphi_k : k \ge 1\} \subseteq C_{\mathrm{b}}\big(\mathbb{R}; [0,\infty)\big)$ 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),

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[E_{\xi,0}(t)\big] &= e^{-\frac{t}{2}} \lim_{k\to\infty} \mathbb{E}^{\mathbb{P}}\big[e^{-\frac{1}{2}A_\xi(t)}\varphi_k\big(B(t)\big)\big] \\ &= e^{-\frac{t}{2}} \lim_{k\to\infty} \int h(t,0,y)\varphi_k(y)\,dy = e^{-t}\int g\big(1-e^{-2t},y\big)e^{y^2}\,dy = 1.\end{aligned}$$

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 ξ 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 $\{\mathbb{P}_k : k \ge 1\}$ is a sequence of Borel probability measures on $C\big([0,\infty);\mathbb{R}^N\big)$ with the property that $\mathbb{P}_{k+1} \upharpoonright \mathcal{B}_k = \mathbb{P}_k \upharpoonright \mathcal{B}_k$ for all $k \ge 1$. Then there is a unique Borel probability measure $\mathbb{P}$ on $C\big([0,\infty);\mathbb{R}^N\big)$ such that $\mathbb{P} \upharpoonright \mathcal{B}_k = \mathbb{P}_k \upharpoonright \mathcal{B}_k$ for all $k \ge 1$.*

Proof: This is a relatively straight forward appliction of the fact that a family of Borel probability measures on a Polish space E is realatively 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 $K\complement$ is assigned measure less that ϵ by all members of the family. To apply this result here, one also has to know that $K \subseteq C\big([0,\infty);\mathbb{R}^N\big)$ is compact if and only if $\{\psi \upharpoonright [0,k] : \psi \in K\}$ is compact in $C\big([0,k];\mathbb{R}^N\big)$ for each $k \ge 1$.

The uniquess assertion is trivial. To prove the existence, let $\epsilon > 0$ be given, and for each $k \ge 1$ choose a compact subset K_k of $C\big([0,k];\mathbb{R}^N\big)$ such that

$$\max_{1\le j\le k} \mathbb{P}_j\big(\{\psi : \psi \upharpoonright [0,k] \notin K_k\}\big) le 2^{-k}\epsilon,$$

in which case

$$\sup_{j\ge 1} \mathbb{P}_j\big(\{\psi : \psi \upharpoonright [0,k] \notin K_k\}\big) \le 2^{-k}\epsilon.$$

Finally, set

$$K = \{\psi : \psi \upharpoonright [0,k] \in K_k \text{ for all } k \ge 1\}.$$

Then K is a compact subset of $C\big([0,\infty);\mathbb{R}^N\big)$, and

$$\mathbb{P}_j\big(K\complement\big) \le \sum_{k=1}^{\infty} \mathbb{P}_j\big(\{\psi : \psi \upharpoonright [0,k] \notin K_k\}\big) \le \epsilon.$$

Hence, the sequence $\{\mathbb{P}_k : k \geq 1\}$ is relatively compact. Furthermore, any limit $\mathbb{P}$ will have the property that $\mathbb{P} \restriction \mathcal{B}_k = \mathbb{P}_k \restriction \mathcal{B}_k$ for all $k \geq 1$. Therefore the sequence converges and its limit has the required property. $\square$

Let $X(\,\cdot\,,\mathbf{x})$ be the solution to (56) and $\boldsymbol{\beta} : \mathbb{R}^N \longrightarrow \mathbb{R}^M$ a continuous function. Set

$$E_{\boldsymbol{\beta}}(t,\mathbf{x}) = \exp\left(\int_0^t \Big(\boldsymbol{\beta}\big(X(\tau,\mathbf{x})\big), dw(\tau)\Big)_{\mathbb{R}^M} - \int_0^t \big|\boldsymbol{\beta}\big(X(\tau,\mathbf{x})\big)\big|^2\, d\tau\right),$$

and assume that $\big(E_{\boldsymbol{\beta}}(t,\mathbf{x}), \overline{\mathcal{W}_t}, \mathcal{W}\big)$ is a martingale. If $\mathbb{P}_k$ is the Borel probability measure on $C\big([0,\infty);\mathbb{R}^M\big)$ given by

$$\mathbb{P}_k(\Gamma) = \mathbb{E}^{\mathcal{W}}\big[E_{\boldsymbol{\beta}}(k,\mathbf{x}),\, \Gamma\big],$$

then $\mathbb{P}_{k+1} \restriction \overline{\mathcal{W}_k} = \mathbb{P}_k \restriction \overline{\mathcal{W}_k}$, and so, by Lemma 80, there is a unique $\mathcal{W}_{\boldsymbol{\beta},\mathbf{x}}$ with the property that

$$\text{(81)} \qquad \mathcal{W}_{\boldsymbol{\beta},\mathbf{x}}(\Gamma) = \mathbb{E}^{\mathcal{W}}\big[E_{\boldsymbol{\beta}}(t,\mathbf{x}),\, \Gamma\big] \text{ for all } t \geq 0 \text{ and } \Gamma \in \overline{\mathcal{W}_t}.$$

Furthermore, if

$$L_{\boldsymbol{\beta}}\varphi = L\varphi + \big(\sigma\boldsymbol{\beta}, \nabla\varphi\big)_{\mathbb{R}^N},$$

then, by Lemma 73,

$$\left(\varphi\big(X(t,\mathbf{x})\big) - \int_0^t L_{\boldsymbol{\beta}}\varphi\big(X(\tau,\mathbf{x})\big)\, d\tau, \overline{\mathcal{W}_t}, \mathcal{W}_{\boldsymbol{\beta},\mathbf{x}}\right)$$

is a martingale for all $\varphi \in C^2_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$. Hence the distribution $\mathbb{P}_{\boldsymbol{\beta},\mathbf{x}}$ of $X(\,\cdot\,,\mathbf{x})$ under $\mathcal{W}_{\boldsymbol{\beta},\mathbf{x}}$ is a solution starting at $\mathbf{x}$ to the martingale problem for $L_{\boldsymbol{\beta}}$. Now assume that $\sigma\boldsymbol{\beta}$ is Lipschitz continuous, and let $X_{\boldsymbol{\beta}}(\,\cdot\,,\mathbf{x})$ be the solution to (56) with $b + \sigma\boldsymbol{\beta}$ replacing b. Then the distribution of $X_{\boldsymbol{\beta}}(\,\cdot\,,\mathbf{x})$ is also a solution starting at $\mathbf{x}$ to the martingale problem for $L_{\boldsymbol{\beta}}$, and so, if there is only one solution to that martingale problem, then

$$\text{(82)} \quad \mathcal{W}\big(X_{\boldsymbol{\beta}}(\,\cdot\,,\mathbf{x}) \in \Gamma\big) = \mathbb{E}^{\mathcal{W}}\big[E_{\boldsymbol{\beta}}(t,\mathbf{x}),\, X(\,\cdot\,,\mathbf{x}) \in \Gamma\big] \text{ for all } t \geq 0 \text{ and } \Gamma \in \mathcal{B}_t.$$

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 Leamm 80. To understand how it relates to that discussion, assume that σ is Lipschitz continuous but that $b = \sigma\boldsymbol{\beta}$ is only locally Lipschitz continuous. Choose an $\eta \in C^\infty\big(\mathbb{R}^N, [0,1]\big)$

that is 1 on $\overline{B(\mathbf{0},1)}$ and 0 off of $B(\mathbf{0},2)$, and set $b_k(\mathbf{y}) = \eta(k^{-1}\mathbf{y})b(\mathbf{y})$. For each $k \geq 1$, let $\tilde{X}_k(\,\cdot\,,\mathbf{x})$ be the solution to (56) with b_k in place of b, and define $\tilde{\zeta}_k = \inf\{t \geq 0 : |\tilde{X}_k(t,\mathbf{x})| \geq k\}$. By Lemma 38, $\tilde{X}_{k+1}(t \wedge \tilde{\zeta}_k, \mathbf{x}) = \tilde{X}_k(t \wedge \tilde{\zeta}_k, \mathbf{x})$ (a.s.,$\mathcal{W}$), and so $\tilde{\zeta}_{k+1} \geq \tilde{\zeta}_k$ (a.s.,$\mathcal{W}$). Therefore there exists a measurable $\tilde{\mathfrak{e}} : C\big([0,\infty);\mathbb{R}^M\big) \longrightarrow [0,\infty]$ to which $\{\tilde{\zeta}_k : k \geq 1\}$ converges (a.s.,$\mathcal{W}$), and there exists a progressively measurable function $\tilde{X}(\,\cdot\,,\mathbf{x})$ with values in $\mathbb{R}^N \cup \{\boldsymbol{\infty}\}$ such that $\tilde{X}(\,\cdot\,,\mathbf{x}) \restriction [0,\tilde{\mathfrak{e}})$ is continuous, $\tilde{X}(t,\mathbf{x}) = \boldsymbol{\infty}$ if $t \geq \tilde{\mathfrak{e}}$, and $\tilde{X}(t,\mathbf{x}) = \tilde{X}_k(t,\mathbf{x})$ (a.s.,$\mathcal{W}$) if $t \in [0,\tilde{\zeta}_k)$. Obviously, $\tilde{\mathfrak{e}}$ can be thought of as the *explosion time* for $\tilde{X}(\,\cdot\,,\mathbf{x})$. Now define

$$L_k\varphi = \big(b_k, \nabla\varphi\big)_{\mathbb{R}^N} + \tfrac{1}{2}\mathrm{Trace}\big(\sigma\sigma^\top\nabla^2\varphi\big),$$

and assume that, for each $k \geq 1$, the martingale problem for L_k has only one solution starting at $\mathbf{x}$. Then if $X(\,\cdot\,,\mathbf{x})$ is the solution to (56) when $b = 0$, (82) says that

$$\mathbb{E}^{\mathcal{W}}\big[E_{k,x}(t,\mathbf{x}),\ X(\,\cdot\,,\mathbf{x}) \in \Gamma\big] = \mathcal{W}\big(\tilde{X}_k(\,\cdot\,,\mathbf{x}) \in \Gamma\big) \text{ for } t \geq 0 \text{ and } \Gamma \in \mathcal{B}_t,$$

where

$$E_k(t,\mathbf{x}) = \exp\left(\int_0^t \Big(\boldsymbol{\beta}_k\big(X(\tau,\mathbf{x})\big), dw(\tau)\Big)_{\mathbb{R}^M} - \tfrac{1}{2}\int_0^t |\boldsymbol{\beta}_k(\tau)|^2\,d\tau\right)$$
$$\text{with } \boldsymbol{\beta}_k(\mathbf{y}) \equiv \eta(k^{-1}\mathbf{y})\boldsymbol{\beta}(\mathbf{y}).$$

In particular, if $\zeta_k = \inf\{t \geq 0 : |X(t,\mathbf{x})| \geq k\}$, then

$$E_k(t,\mathbf{x}) = E_{\boldsymbol{\beta}}(t,\mathbf{x}) \text{ (a.s.,}\mathcal{W}\text{) for } t \in [0,\zeta_k),$$

and so

$$\mathbb{E}^{\mathcal{W}}\big[E_{\boldsymbol{\beta}}(t,\mathbf{x}),\ \zeta_k > t\big] = \mathcal{W}\big(\tilde{\zeta}_k > t\big).$$

Finally, since $\zeta_k \nearrow \infty$ (a.s.,$\mathcal{W}$), we have shown that

$$\mathbb{E}^{\mathcal{W}}\big[E_{\boldsymbol{\beta}}(t,\mathbf{x})\big] = \mathcal{W}\big(\tilde{\mathfrak{e}} > t\big).$$

In other words, the expected value of $E_{\boldsymbol{\beta}}(t,\mathbf{x})$ is the probability that $\tilde{X}(\,\cdot\,,\mathbf{x})$ hasn't exploded by time t. Thus, $\tilde{X}(\,\cdot\,,\mathbf{x})$ *never explodes (i.e.,* $\mathcal{W}(\tilde{\mathfrak{e}} = \infty) = 1$*) if and only if* $\big(E_{\boldsymbol{\beta}}(t,\mathbf{x}), \overline{\mathcal{W}_t}, \mathcal{W}\big)$ *is a martingale.* Since $\tilde{X}(\,\cdot\,,\mathbf{x})$ will never explode if $|b(\mathbf{y})| \leq C(1 + |\mathbf{y}|)$, this explains our earlier conclusion that $\big(E_\xi(t), \mathcal{F}_t, \mathbb{P}\big)$ is a martingale when $\xi(t) = B(t)$. Indeed, after translating it into the present setting, one sees that we were dealing with $\sigma = \mathbf{1}$, $b(y) = y$, and $x = 0$. When $\xi(t) = -B(t)$, $\tilde{X}(t,x)$ is the Ornstien-Uhlenbech 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 is an infinite dimensional, separable Banach space over $\mathbb{R}$, and, for $a \in E$, define the translation map $T_a : E \longrightarrow E$ by $T_a = x + a$. Given μ a Borel probability measure on E, there is a dense subset D such that $(T_{ta})_*\mu \perp \mu$ for all $a \in D$.*

To prove this theorem, we will use the folowing lemma.

LEMMA 84. *If K_1 and K_2 are compact subsets of E, then* $\operatorname{int}(K_2 - K_1) = \emptyset$.

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 $\operatorname{int}(K) = \emptyset$ for every compact K.

Given K and $r > 0$, choose $x_1, \dots, x_n \in K$ so that $K \subseteq \bigcup_{m=1}^{n} B_E\big(x_m, \frac{r}{4}\big)$. 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 $1 \le m \le n$. Now choose $a \in E$ so that $\|a\|_E = r$ and $\langle a, a^* \rangle \ge \frac{3r}{4}$. If x were in $K \cap (a + K)$, then there would exist $1 \le m, m' \le n$ such that $\|x - x_m\|_E < \frac{r}{4}$ and $\|x - a - x_{m'}\|_E < \frac{r}{4}$, which would lead to the contradicion that

$$\langle x, a^* \rangle = \langle x - x_m, a^* \rangle < \tfrac{r}{4} \text{ and } \langle x, a^* \rangle = \langle a, a^* \rangle + \langle x - a - x_{m'} \rangle \ge \tfrac{r}{2}.$$

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. □

PROOF OF THEOREM 83: Let μ be a Borel probability measure on E. By Ulam's lemma, for each $m \ge 1$ there exists a compact set K_m such that $\nu(K_m) \ge 1 - \frac{1}{m}$. 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, $\operatorname{int}(K_n - K_m) = \emptyset$ for all $m, n \ge 1$, and so, by the Baire category theorem, $\operatorname{int}(B) = \emptyset$. 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 $a \in D = B\complement$. □

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 infinie dimensional Banach 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 Ω be the space of $w \in C\big([0,\infty);\mathbb{R}^M\big)$ with the properties that $w(0) = \mathbf{0}$ and $\lim_{t\to\infty} \frac{|w(t)|}{t} = 0$, and observe that $\mathcal{W}(\Omega) = 1$. Next define the norm $\|w\|_\Omega = \sup_{t\ge 0} \frac{|w(t)|}{1+t}$. Then it is not hard to show that Ω 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 coninuous $h \in C\big([0,\infty);\mathbb{R}^M\big)$ with the properties that $h(0) = \mathbf{0}$ and $\dot h \in L^2\big([0,\infty);\mathbb{R}{*}M)$. That is, $h(t) = \int_0^t \dot h(\tau)\,d\tau$ where $\dot h \in L^2\big([0,\infty);\mathbb{R}^M\big)$. It is easy to check that $H^1(\mathbb{R}^M)$ becomes a separable Hilbert space if one uses the inner product

$$\big(g,h\big)_{H^1(\mathbb{R}^M)} = \big(\dot g,\dot h\big)_{L^2([0,\infty);\mathbb{R}^M)}.$$

In addition, since $|h(t)| \le t^{\frac12}\|h\|_{H^1(\mathbb{R}^M)}$, $h \in \Omega$ and $\|h\|_\Omega \le \frac12\|h\|_{H^1(\mathbb{R}^M)}$. Finally, it is obvious that $H^1(\mathbb{R}^M)$ is a dense subspace of Ω.

Theorem 85. *Think of $\mathcal{W}$ as a Borel measure on Ω. Given $h \in H^1(\mathbb{R}^M)$, set*

$$I(\dot h) = \int_0^\infty \big(\dot h(\tau), dw(\tau)\big)_{\mathbb{R}^M} \text{ and } R_h = \exp\big(I(\dot h) - \tfrac12\|h\|^2_{H^1(\mathbb{R}^M)}\big).$$

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 $n \ge 1$, $0 \le t_1 < \cdots < t_n$, and $\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_n \in \mathbb{R}^M$, it is obvious that

$$\begin{aligned}&\mathbb{E}^{(T_h)_*\mathcal{W}}\Big[e^{i\sum_{m=1}^n(\boldsymbol{\xi}_m,w(t_m))_{\mathbb{R}^M}}\Big]\\ &\qquad= \exp\left(i\sum_{m=1}^n\big(\boldsymbol{\xi}_m,h(t_m)\big)_{\mathbb{R}^M} - \tfrac12\sum_{m,m'=1}^n(t_m\wedge t_{m'})\big(\boldsymbol{\xi}_m,\boldsymbol{\xi}_{m'}\big)_{R^M}\right).\end{aligned}$$

To compute the same integral with respect to $R_h d\mathcal{W}$, set $h_m(t) = (t\wedge t_m)\boldsymbol{\xi}_m$ and $f(t) = \dot h(t) + i\sum_{m=1}^n \dot h_m(t)$. Then

$$R_h(w)e^{i\sum_{m=1}^n(\boldsymbol{\xi}_m,w(t_m))_{\mathbb{R}^M}} = \exp\left(\int_0^\infty \big(f(\tau),dw(\tau)\big)_{\mathbb{R}^M} - \tfrac12\|h\|^2_{H^1(\mathbb{R}^M)}\right),$$

and so

$$
\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\big[R_h(w)e^{i\sum_{m=1}^n(\boldsymbol{\xi}_m,w(t_m))_{\mathbb{R}^M}}\big]\\
&\qquad=\exp\left(-\tfrac{1}{2}\sum_{j=1}^N\int_0^\infty f(\tau)_j^2\,d\tau-\|h\|_{H^1(\mathbb{R}^M)}^2\right)\\
&\qquad=\exp\left(i\sum_{m=1}^n\big(h_m,h\big)_{H^1(\mathbb{R}^M)}-\tfrac{1}{2}\sum_{m,m'=1}^n\big(h_m,h_{m'}\big)_{H^1(\mathbb{R}^M)}\right)\\
&\qquad=\exp\left(i\sum_{m=1}^n\big(\boldsymbol{\xi}_m,h(t_m)\big)_{\mathbb{R}^M}-\tfrac{1}{2}\sum_{m,m'=1}^n(t_m\wedge t_{m'})\big(\boldsymbol{\xi}_m,\boldsymbol{\xi}_{m'}\big)_{R^M}\right).
\end{aligned}
$$

Hence $d(T_h)_*\mathcal{W}=R_h d\mathcal{W}$.

The proof of the second assertion requires some prepartions. Let L denote the subspace of $H^1(\mathbb{R}^M)$ consisting of twice continuously differentiable functions whose first derivatives have compact support. Clearly L is dense in $H^1(\mathbb{R}^M)$, and so one can find an orthonormal $\{h_m : m\ge 1\}\subseteq L$ for $H^1(\mathbb{R}^M)$. Now define the linear functional Λ on L by

$$
\Lambda(h)=-\int_0^\infty\big(f(\tau),\ddot{h}(\tau)\big)_{\mathbb{R}^M}\,d\tau.
$$

We need to show that $\sum_{m=1}^\infty\Lambda(h_m)^2=\infty$. To this end, suppose that $C\equiv\sqrt{\sum_{m=1}^\infty\Lambda(h_m)^2}<\infty$. Then, if h is in the span of $\{h_m : m\ge 1\}$,

$$
|\Lambda(h)|\le\sum_{m=1}^\infty|\Lambda(h_m)|\big|(h,h_m)_{H^1(\mathbb{R}^M)}\big|\le C\|h\|_{H^1(\mathbb{R}^M)},
$$

and so Λ admits a unique extension as a continuous linear functional on $H^1(\mathbb{R}^M)$. Thus, by the Riesz representation theorem for Hilbert space, there exists an $h_0\in H^1(\mathbb{R}^M)$ such that $\Lambda(h)=\big(h,h_0\big)_{H^1(\mathbb{R}^M)}$, and so

$$
\int_0^\infty\big(f(\tau),\ddot{h}(\tau)\big)_{\mathbb{R}^M}\,d\tau=\int_0^\infty\big(h_0(\tau),\ddot{h}(\tau)\big)_{\mathbb{R}^M}\,d\tau\text{ for all }h\in L.
$$

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

$$
\psi_\epsilon(\tau)=\int_0^\tau\left(\int_0^{\tau_1}\rho_\epsilon(t-\tau_2)\,d\tau_2\right)d\tau_1-\tau.
$$

Then $\dot{\psi}_\epsilon = 0$ off $[0,t)$ and $\ddot{\psi}_\epsilon = \rho_\epsilon$. Hence, $\psi_\epsilon \in L$ and, for any $\boldsymbol{\xi} \in \mathbb{R}^M$,

$$\begin{aligned}&\big(\boldsymbol{\xi}, f(t)\big)_{\mathbb{R}^M} = \lim_{\epsilon \searrow 0} \int_0^\infty \big(f(\tau), \rho_\epsilon(t-\tau)\boldsymbol{\xi}\big)_{\mathbb{R}^M}\, d\tau = -\lim_{\epsilon\searrow 0} \Lambda(\psi_\epsilon \boldsymbol{\xi})\\ &= -\lim_{\epsilon\searrow 0}\big(\psi_\epsilon\boldsymbol{\xi}, h_0\big)_{H^1(\mathbb{R}^M)} = \lim_{\epsilon\searrow 0}\int_0^\infty \big(h_0(\tau), \rho_\epsilon(t-\tau)\boldsymbol{\xi}\big)_{\mathbb{R}^M}\, d\tau = \big(\boldsymbol{\xi}, h_0(t)\big)_{\mathbb{R}^M},\end{aligned}$$

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 $F : \Omega \longrightarrow \mathbb{R}^{\mathbb{Z}^+}$ so that

$$F(w)_m = \int_0^\infty \big(\dot{h}_m(\tau), dw(\tau)\big)_{\mathbb{R}^M} = -\int_0^\infty \big(w(\tau), \ddot{h}_m(\tau)\big)_{\mathbb{R}^M}\, d\tau.$$

Then $F_*\mathcal{W} = \gamma_{0,1}^{\mathbb{Z}^+}$, and, because $F(w+f)_m = F(w)_m + \Lambda(h_m)$,

$$F_*\big((T_f)_*\mathcal{W}\big) = \prod_{m=1}^\infty \gamma_{\Lambda(h_m),1}.$$

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 $F_*\big((T_f)_*\mathcal{W}\big) \perp F_*\mathcal{W}$, 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 Ω. To see this, first observe that, because $H^1(\mathbb{R}^M)$ is dense in Ω, it suffices to show that $\mathcal{W}\big(B_\Omega(h,r)\big) > 0$ for every $h \in H^1(\mathbb{R}^M)$. Second, note that

$$\begin{aligned}\mathcal{W}\big(B_\Omega(0,r)\big) &= \mathbb{E}^{\mathcal{W}}\big[R_{-h}^{-\frac{1}{2}} R_{-h}^{\frac{1}{2}},\, B_\Omega(0,r)\big]\\ &\le \mathbb{E}^{\mathcal{W}}\big[R_{-h}^{-1}\big]^{\frac{1}{2}} (T_{-h})_*\mathcal{W}\big(B_\Omega(0,r)\big)^{\frac{1}{2}} = e^{\frac{\|h\|^2_{H^1(\mathbb{R}^M)}}{2}}\mathcal{W}\big(B_\Omega(h,r)\big),\end{aligned}$$

and therefore it suffices to show that $\mathcal{W}\big(B_\Omega(0,r)\big) > 0$ for all $r > 0$. To this end, let $\big(B(t), \mathcal{F}_t, \mathbb{P}\big)$ be an $\mathbb{R}$-valued Brownian motion, and consider the function

$$u(t,x) = e^{\frac{\pi^2 t}{8r^2}} \sin\frac{\pi}{2}(x+r).$$

Because $\partial_t u + \frac{1}{2}\partial_x^2 u = 0$, $\big(u(t,B(t)), \mathcal{F}_t, \mathbb{P}\big)$ is a martingale. Hence, if $\zeta_r = \inf\{t \ge 0 : |B(t)| \ge r\}$, then, because $u(t, \pm r) = 0$,

$$1 = \mathbb{E}^{\mathbb{P}}\big[u\big(t\wedge\zeta_r, B(t\wedge\zeta_r)\big)\big] = e^{\frac{\pi^2 t}{8r^2}}\mathbb{E}^{\mathbb{P}}\left[\sin\frac{\pi(B(t)+r)}{2}\ \zeta_r > t\right],$$

and so

$$\mathbb{P}\big(\zeta_r > t\big) \ge e^{-\frac{\pi^2 t}{8r^2}}.$$

Because

$$\mathcal{W}\big(\|w(\,\cdot\,)\|_{[0,T]}<r\big)\ge\mathcal{W}\left(\max_{1\le j\le N}\|w(\,\cdot\,)_j\|_{[0,T]}<\frac{r}{N^{\frac12}}\right)=\mathbb{P}\big(\zeta_{N^{-\frac12}r}>T\big)^N,$$

we now know that

$$\mathcal{W}\big(\|w(\,\cdot\,)\|_{[0,T]}<r\big)\ge e^{-\frac{N^2\pi^2T}{8r^2}},$$

and therefore

$$\begin{aligned}\mathcal{W}\big(B_\Omega(0,r)\big)&\ge\mathcal{W}\left(\|w(\,\cdot\,)\|_{[0,T]}<\frac{r}{2}\ \&\ \sup_{t\ge T}\frac{|w(t)-w(T)|}{1+t}\le\frac{r}{2}\right)\\&\ge e^{-\frac{N^2\pi^2T}{2r^2}}\mathcal{W}\left(\sup_{t\ge0}\frac{|w(t)|}{1+t+T}\le\frac{r}{2}\right).\end{aligned}$$

Finally, because $\|w\|_\Omega<\infty$ and $\frac{|w(t)|}{1+t}\longrightarrow0$ (a.s.,$\mathcal{W}$), we can choose $T>0$ so that

$$\mathcal{W}\left(\sup_{t\in[0,\sqrt{T}]}\frac{|w(t)|}{1+t+T}\ge\frac{r}{2}\right)\vee\mathcal{W}\left(\sup_{t\ge\sqrt{T}}\frac{|w(t)|}{1+t}\ge\frac{r}{2}\right)\le\frac14,$$

in which case

$$\begin{aligned}&\mathcal{W}\left(\sup_{t\ge0}\frac{|w(t)|}{1+t+T}\ge\frac{r}{2}\right)\\&\quad\le\mathcal{W}\left(\sup_{t\in[0,\sqrt{T}}\frac{|w(t)|}{1+t+T}\ge\frac{r}{2}\right)+\mathcal{W}\left(\sup_{t\ge\sqrt{T}}\frac{|w(t)|}{1+t}\ge\frac{r}{2}\right)\le\frac12.\end{aligned}$$

Besides what it says about Wiener measure, the preceding result has the following interesting application to partial differential equations. Let $G\ni\mathbf{0}$ be a connected, open subset of $\mathbb{R}^M$ and u a harmonic function on G that achieve its minimum value at $\mathbf{0}$. Then, for bounded, open H with $\bar{H}\subseteq G$, $\mathcal{W}(\zeta^H<\infty)=1$ and $\big(u(w(t\wedge\zeta^H)),W_t,\mathcal{W}\big)$ martingale when $\zeta^H=\inf\{t\ge0:\,w(t)\notin H\}$. Now suppose that there were a point $\mathbf{x}\in G\setminus\{\mathbf{0}\}$ at which $u(\mathbf{x})>u(\mathbf{0})$, and let $p:[0,1]\longrightarrow G$ be a continuous path such that $p(0)=\mathbf{0}$ and $p(1)=\mathbf{x}$. Choose $r>0$ so that $u(\mathbf{y})-u(\mathbf{0})\ge\delta\equiv\frac{u(\mathbf{x})-u(\mathbf{0})}{2}$ for $\mathbf{y}\in\overline{B(\mathbf{x},r)}$ and $|p(t)-\mathbf{y}|>2r$ for all $t\in[0,1]$ and $\mathbf{y}\notin G$. If

$$\zeta(w)=\inf\{t\ge0:\,|w(t)-p(t)|\ge2r\}\text{ and }\zeta'(w)=\inf\{t\ge0:\,w(t)\in\overline{B(\mathbf{x},r)}\},$$

then, $\|w(\,\cdot\,)-p(\,\cdot\,)\|_{[0,1]}<r\implies\zeta'(w)<\zeta(w)$, and so

$$\mathcal{W}(\zeta'<\zeta)\ge\epsilon_r\equiv\mathcal{W}\big(\|w(\,\cdot\,)-p(\,\cdot\,)\|_{[0,1]}<r\big)>0.$$

But this means that

$$u(\mathbf{0}) = \mathbb{E}^{\mathcal{W}}\big[u\big(w(\zeta\wedge\zeta')\big)\big] = \mathbb{E}^{\mathcal{W}}\big[u\big(w(\zeta')\big),\,\zeta'<\zeta\big] + \mathbb{E}^{\mathcal{W}}\big[u\big(w(\zeta)\big),\,\zeta>\zeta'\big]$$
$$\geq u(\mathbf{0}) + \delta\epsilon_r,$$

which is impossible. Therefore no such $\mathbf{x}$ exists, and so we have proved the **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, suppse that $\mathfrak{G}$ is an open subset of $\mathbb{R}\times\mathbb{R}^M$ which is *upward path connected* in the sense that for each $(s,\mathbf{x}),(t,\mathbf{y})\in\mathfrak{G}$ with $t>s$ there is a continuous path $p:[s,t]\longrightarrow\mathbb{R}^M$ such that $p(s)=\mathbf{x}$, $p(t)=\mathbf{y}$, and $\big(\tau,p(\tau)\big)\in\mathfrak{G}$ for all $\tau\in[s,t]$. If $u\in C^{1,2}(\mathfrak{G})$ satisfying $\partial_t u+\frac{1}{2}\Delta u=0$ and $(s,\mathbf{x})\in\mathfrak{G}$, $u(s,\mathbf{x})\leq u(t,\mathbf{y})$ for all $(t,\mathbf{y})\in\mathfrak{G}$ with $t>s$ implies that $u(t,\mathbf{y})=u(s,\mathbf{x})$ for all such $(t,\mathbf{y})$.

A second application is to the development of a calulus for functions on Ω in which Wiener measure plays the role that Lebesgue measure plays in finite dimensions. First observe that

$$\mathbb{E}^{\mathcal{W}}\big[R_h^p\big] = e^{\frac{p(p-1)\|h\|^2_{H^1(\mathbb{R}^M)}}{2}} \quad\text{for all } p\in[1,\infty). \tag{86}$$

Hence, by Hölder's inequality,

$$\|\Phi\circ T_h\|_{L^q(\mathcal{W};\mathbb{R})} \leq e^{\frac{\|h\|^2_{H^1(\mathbb{R}^M)}}{2(p-q)}}\|\Phi\|_{L^p(\mathcal{W};\mathbb{R})} \tag{87}$$
$$\text{for } 1<q<p<\infty \text{ and } \Phi\in L^p(\mathcal{W};\mathbb{R}).$$

Now suppose that $\Phi\in L^p(\mathcal{W};\mathbb{R})$ for some $p\in(1,\infty)$ and that there exists a function $D_h\Phi\in L^p(\mathcal{W};\mathbb{R})$ such that

$$\lim_{\xi\to0}\frac{\Phi\circ T_{\xi h}-\Phi}{\xi}\longrightarrow D_h\Phi \text{ in } L^1(\mathcal{W};\mathbb{R}).$$

Then, because

$$\mathbb{E}^{\mathcal{W}}\big[\Phi\circ T_{\xi h}-\Phi\big] = \mathbb{E}^{\mathcal{W}}\big[(R_{\xi h}-1)\Phi\big]$$

and $\frac{R_{\xi h}-1}{\xi}\longrightarrow I(\dot h)$ in $L^{p'}(\mathcal{W};\mathbb{R})$,[8] it follows that

$$\mathbb{E}^{\mathcal{W}}\big[D_h\Phi\big] = \mathbb{E}^{\mathcal{W}}\big[I(\dot h)\Phi\big]. \tag{88}$$

Next suppose that $\Phi_1,\Phi_2\in L^p(\mathcal{W};\mathbb{R})$ for some $p\in(2,\infty)$ and that

$$\lim_{\xi\to0}\frac{\Phi_i\circ T_{\xi h}-\Phi_i}{\xi}\longrightarrow D_h\Phi_i \text{ in } L^p(\mathcal{W};\mathbb{R}) \text{ for } i\in\{1,2\}.$$

[8] Here and elsewhere, $p'=\frac{p}{p-1}$ denotes the Hölder conjugate of p.

Then another application of Hölder's inequality shows that

$$\frac{(\Phi_1 \circ T_{\xi h})(\Phi_2 \circ T_{\xi h}) - \Phi_1\Phi_2}{\xi} \longrightarrow \Phi_1 D_h \Phi_2 + \Phi_2 D_h \Phi_1 \text{ in } L^1(\mathcal{W};\mathbb{R}),$$

and therefore

$$\mathbb{E}^{\mathcal{W}}\big[\Phi_1 D_h \Phi_2\big] = -\mathbb{E}^{\mathcal{W}}\big[\Phi_2 D_h \Phi_1\big] + \mathbb{E}^{\mathcal{W}}\big[I(\dot{h})\Phi_1\Phi_2\big]. \tag{89}$$

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 Homogenious 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 **homogenious 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,[9] 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 $m \geq 1$ and $t \in [0,\infty]$, $\square^{(m)}(t) \equiv [0,t)^m$ and $\square^{(m)} \equiv \square^{(m)}(\infty)$, we want to assign a meaning to expressions like

$$\tilde{I}_F^{(m)}(t) = \int_{\square^{(m)}(t)} \big(F(\vec{\tau}), d\vec{w}(\vec{\tau})\big)_{(\mathbb{R}^M)^m}$$

when $F \in L^2\big(\square^{(m)};(\mathbb{R}^M)^m\big)$.

With this goal in mind, when $m = 1$ and $F = f \in L^2\big([0,\infty);\mathbb{R}^M\big)$, take $I_F^{(1)}(t) = I_f(t)$, where $I_f(t)$ is the Paley–Wiener integral of f. When $m \geq 2$ and $F = f_1 \otimes \cdots \otimes f_m$ for some $f_1, \dots, f_m \in L^2\big([0,\infty);\mathbb{R}^M\big)$,[10] we use induction to define $I_F^{(m)}(t)$ so that

$$I^{(m)}_{f_1\otimes\cdots\otimes f_m}(t) = \int_0^t I^{(m-1)}_{f_1\otimes\cdots\otimes f_{m-1}}(\tau)\big(f_m(\tau), dw(\tau)\big)_{\mathbb{R}^M}, \tag{90}$$

[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 $f_1\otimes\cdots\otimes f_m$ with the $(\mathbb{R}^M)^m$-valued function F on $[0,\infty)^m$ such that $(\Xi, F(t_1,\dots,t_m))_{(\mathbb{R}^M)^m} = \big(\boldsymbol{\xi}_1, f_1(t_1)\big)^M_{\mathbb{R}} \cdots \big(\boldsymbol{\xi}_m, f_m(t_m)\big)^M_{\mathbb{R}}$ for $\Xi = (\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_m) \in (\mathbb{R}^M)^m$.

where now we need Itô's integral. Of course, in order to do so, we are obliged to check that $\tau \rightsquigarrow f_m(\tau) I^{(m-1)}_{f_1\otimes\cdots\otimes f_{m-1}}(\tau)$ is square integrable. But, assuming that $I^{(m-1)}_{f_1\otimes\cdots\otimes f_{m-1}}$ is well defined, we have that

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}&\left[\int_0^T \left|f_m(\tau) I^{(m-1)}_{f_1\otimes\cdots\otimes f_{m-1}}(\tau)\right|^2 d\tau\right]\\ &= \int_0^T \left|f_m(\tau)\right|^2 \mathbb{E}^{\mathcal{W}}\left[\left|I^{(m-1)}_{f_1\otimes\cdots\otimes f_{m-1}}(\tau)\right|^2\right] d\tau.\end{aligned}$$

Hence, at each step in our induction procedure, we can check that

$$\mathbb{E}^{\mathcal{W}}\left[\left|I^{(m)}_{f_1\otimes\cdots\otimes f_m}(T)\right|^2\right] = \int_{\triangle^{(m)}(T)} \left|f_1(\tau_1)\right|^2 \cdots \left|f_m(\tau_m)\right|^2 d\tau_1\cdots d\tau_m,$$

where $\triangle^{(m)}\ (t) \equiv \left\{(t_1,\dots,t_m) \in \square^{(m)} :\le t_1 < \cdots < t_m < t\right\}$; and so, after polarization, we arrive at

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}&\left[I^{(m)}_{f_1\otimes\cdots\otimes f_m}(T) I^{(m)}_{f'_1\otimes\cdots\otimes f'_m}(T)\right]\\ &= \int_{\triangle^{(m)}(T)} \left(f_1(\tau_1), f'(\tau_1)\right)^M_{\mathbb{R}} \cdots \left(f_m(\tau_m), f'_m(\tau_m)\right)^M_{\mathbb{R}}\, d\tau_1\cdots d\tau_m.\end{aligned}$$

I next introduce

$$\tilde{I}^{(m)}_{f_1\otimes\cdots\otimes f_m}(t) \equiv \sum_{\pi\in\Pi_m} I^{(m)}_{f_{\pi(1)}\otimes\cdots\otimes f_{\pi(m)}}(t), \tag{91}$$

where Π_m is the symmetric group (i.e., the group of permutions) on $\{1,\dots,m\}$. By the preceding, one sees that

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}&\left[\tilde{I}^{(m)}_{f_1\otimes\cdots\otimes f_m}(T) \tilde{I}^{(m)}_{f'_1\otimes\cdots\otimes f'_m}(T)\right]\\ &= \sum_{\pi,\pi'\in\Pi_m} \int_{\triangle^{(m)}(T)} \prod_{\ell=1}^m \left(f_{\pi(\ell)}(\tau_\ell), f'_{\pi'(\ell)}(\tau_\ell)\right)_{\mathbb{R}^M} d\tau_1\cdots d\tau_m\\ &= \sum_{\pi\in\Pi_m} \int_{\square^{(m)}(T)} \prod_{\ell=1}^m \left(f_\ell(\tau_\ell), f'_{\pi(\ell)}(\tau_\ell)\right)_{\mathbb{R}^M} d\tau_1\cdots d\tau_m\\ &= \sum_{\pi\in\Pi_m} \prod_{\ell=1}^m \left(f_\ell, f'_{\pi(\ell)}\right)_{L^2([0,T);\mathbb{R}^M)}.\end{aligned}$$

In preparation for the next step, let $\{g_j : j \ge 1\}$ be an orthonormal basis in $L^2\big([0,\infty);\mathbb{R}^M\big)$, and note that $\left\{g_{j_1}\otimes\cdots\otimes g_{j_m} : (j_1,\dots,j_m) \in (\mathbb{Z}^+)^m\right\}$ is an orthonormal basis in $L^2\big(\square^{(m)};(\mathbb{R}^M)^m\big)$. Next, let $\mathcal{A}$ denote the set of $\alpha\in\mathbb{N}^{\mathbb{Z}^+}$ for which $\|\alpha\|\equiv\sum_1^\infty\alpha_j<\infty$. Finally, given $\alpha\in\mathcal{A}$, set

$$S(\alpha)=\{j\in\mathbb{Z}^+:\,\alpha_j\geq 1\}\text{ and }G_\alpha=\bigotimes_{j\in S(\alpha)}g^{\otimes\alpha_j}.$$

Then, $G_\alpha\in L^2\big(\square^{(\|\alpha\|)};(\mathbb{R}^M)^{\|\alpha\|}\big)$, and, for $\alpha,\beta\in\mathcal{A}$ with $\|\alpha\|=\|\beta\|=m$,

$$\big(\tilde I^{(m)}_{G_\alpha},\tilde I^{(m)}_{G_\beta}\big)_{L^2(\mathcal{W};\mathbb{R})}=\delta_{\alpha,\beta}\alpha!,\tag{92}$$

where $\alpha!=\prod_{j\in S(\alpha)}\alpha_j!$.

Now, given $F\in L^2\big(\square^{(m)};(\mathbb{R}^M)^m\big)$, set

$$\tilde F(t_1,\dots,t_m)\equiv\sum_{\pi\in\Pi_m}F\big(t_{\pi(1)},\dots,t_{\pi(m)}\big),\tag{93}$$

and observe that

$$\begin{aligned}\|\tilde F\|^2_{L^2(\square^{(m)};(\mathbb{R}^M)^m)}&=\sum_{\mathbf{j}\in(\mathbb{Z}^+)^m}\big(\tilde F,g_{j_1}\otimes\cdots\otimes g_{j_m}\big)^2_{L^2\left(\square^{(m)};(\mathbb{R}^M)^m\right)}\\&=\sum_{\|\alpha\|=m}\binom{m}{\alpha}\big(\tilde F,G_\alpha\big)^2_{L^2\left(\square^{(m)};(\mathbb{R}^M)^m\right)},\end{aligned}$$

where $\binom{m}{\alpha}$ is the multinomial coefficient $\frac{m!}{\alpha!}$. Hence, after combining this with calculation in (92), we have that

$$\mathbb{E}^{\mathcal{W}}\left[\left|\sum_{\|\alpha\|=m}\frac{\big(\tilde F,G_\alpha\big)_{L^2(\square^{(m)};(\mathbb{R}^M)^m)}}{\alpha!}\tilde I^{(m)}_{G_\alpha}(\infty)\right|^2\right]=\frac{1}{m!}\|\tilde F\|^2_{L^2(\square^{(m)};(\mathbb{R}^M)^m)}.$$

With these considerations, we have proved the following.

Theorem 94. *There is a unique linear map*

$$F\in L^2\big(\square^{(m)};(\mathbb{R}^M)^m\big)\longmapsto\tilde I^{(m)}_F\in\mathcal{M}^2(\mathcal{W};\mathbb{R})$$

such that $\tilde I^{(m)}_{f_1\otimes\cdots\otimes f_m}$ is given as in (91) and

$$\mathbb{E}^{\mathcal{W}}\Big[\tilde I^{(m)}_F(\infty)\tilde I^{(m)}_{F'}(\infty)\Big]=\frac{1}{m!}\big(\tilde F,\tilde F'\big)_{L^2(\square^{(m)};(\mathbb{R}^M)^m)}.$$

In fact,

$$\tilde I^{(m)}_F=\sum_{\|\alpha\|=m}\frac{\big(\tilde F,G_\alpha\big)_{L^2(\square^{(m)};(\mathbb{R}^M)^m)}}{\alpha!}\tilde I_{G_\alpha},$$

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 $\tilde{I}_F^{(m)}(t)$ as

$$\int_{\square^{(m)}(t)} \big(F(\vec{\tau}), d\vec{w}(\vec{\tau})\big)_{(\mathbb{R}^M)^m}.$$

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 $\tilde{I}_F^{(m)}$ carefully avoids integration across diagonals, whereas the preceding notation gives no hint of that fact.

Take $Z^{(0)}$ to be the subspace of $L^2(\mathcal{W};\mathbb{R})$ consisting of the constant functions, and, for $m \geq 1$, set

$$Z^{(m)} = \big\{\tilde{I}_F^{(m)}(\infty) :\ F \in L^2\big(\square^{(m)}; (\mathbb{R}^M)^m\big)\big\}.$$

Clearly, each $Z^{(m)}$ is a subspace of $L^2(\mathcal{W};\mathbb{R})$. Furthermore, if $\{F_k :\ k \geq 1\} \subseteq L^2\big(\square^{(m)}; (\mathbb{R}^M)^m\big)$ and $\big\{\tilde{I}_{F_k}^{(m)}(\infty) :\ k \geq 1\big\}$ converges in $L^2(\mathcal{W};\mathbb{R})$, then $\{\tilde{F}_k :\ k \geq 1\big\}$ converges in $L^2\big(\square^{(m)}; (\mathbb{R}^M)^m\big)$ to some symmetric function G. Hence, since $\tilde{G} = m!G$, we see that $\tilde{I}_{F_k}^{(m)}(\infty) \longrightarrow \tilde{I}_F^{(m)}(\infty)$ in $L^2(\mathcal{W};\mathbb{R})$ where $F = \frac{1}{m!}G$. That is, each $Z^{(m)}$ is a closed linear subspace of $L^2(\mathcal{W};\mathbb{R})$. 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

$$\begin{aligned}
&\mathbb{E}^{\mathcal{W}}\Big[I^{(m)}_{f_1\otimes\cdots\otimes f_m}(\infty) I^{(m')}_{f'_1\otimes\cdots\otimes f'_{m'}}(\infty)\Big] \\
&\quad = \int_{\Delta^{(m'-m)}} \prod_{\ell=0}^{m-1} \big(f_{m-\ell}(\tau_\ell), f'_{m'-\ell}(\tau_\ell)\big)_{\mathbb{R}^M} \\
&\qquad\qquad\qquad \times \mathbb{E}^{\mathcal{W}}\big[I^{(m'-m)}_{f'_1\otimes\cdots\otimes f'_{m'-m}}(\tau_{m'-m})\big]\, d\tau_1 \cdots d\tau_{m'-m} = 0,
\end{aligned}$$

which completes the proof.

The space $Z^{(m)}$ is the space of *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 $\tilde{I}_F^{(m)}(\infty)$ is of mth order chaos when $m \geq 1$, it is helpful to replace $dw(\tau)$ by the much more ambiguous $\dot{w}(\tau)\, d\tau$ and write

$$\tilde{I}_F^{(m)}(\infty) = \int_{\square^{(m)}} \Big(F(\tau_1,\dots,\tau_m), \big(\dot{w}(\tau_1),\cdots,\dot{w}(\tau_m)\big)\Big)_{(\mathbb{R}^M)^m} d\tau_1 \cdots d\tau_m.$$

In the world of engineering and physics, $\tau \rightsquigarrow \dot{w}(\tau)$ is *white noise.*[11] Thus, $Z^{(m)}$ is the space built out of homogeneous mth order polynomials in white noises

[11] The terminology comes from the observation that, no matter how one interprets $t \rightsquigarrow \dot{w}(t)$, it is a stationary, centered Gaussian process whose covariance is the Dirac delta function times the identity. In particular, $\dot{w}(t_1)$ is independent of $\dot{w}(t_2)$ when $t_1 \neq t_2$.

evaluated at different times.[12] 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

$$L^2(\mathcal{W};\mathbb{R}) = \bigoplus_{m=0}^{\infty} Z^{(m)}. \tag{95}$$

The key to Itô's proof of (95) is found in the following.

LEMMA 96. *If $f \in L^2\big([0,\infty);\mathbb{R}^M\big)$, then and $I_f^{(0)} \equiv 1$, for each $\lambda \in \mathbb{C}$,*

$$e^{\lambda I_f(\infty) - \frac{\lambda^2}{2}\|f\|^2_{L^2([0,\infty);\mathbb{R}^M)}} = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \tilde{I}^{(m)}_{f^{\otimes m}}(\infty) \equiv \lim_{M\to\infty} \sum_{m=0}^{M} \frac{\lambda^m}{m!} \tilde{I}^{(m)}_{f^{\otimes m}}(\infty)$$

$\mathcal{W}$-almost surely and in $L^2(\mathcal{W};\mathbb{R})$. In fact, if

$$R_f^M(\infty,\lambda) \equiv e^{\lambda I_f(\infty) - \frac{\lambda^2}{2}\|f\|^2_{L^2([0,\infty);\mathbb{R}^M)}} - \sum_{m=0}^{M-1} \frac{\lambda^m}{m!} \tilde{I}^{(m)}_{f^{\otimes m}}(\infty),$$

then

$$\mathbb{E}^{\mathcal{W}}\Big[\big|R_f^M(\infty,\lambda)\big|^2\Big] = e^{|\lambda|^2\|f\|^2_{L^2([0,\infty);\mathbb{R}^M)}} - \sum_{m=0}^{M-1} \frac{\big(|\lambda|\|f\|_{L^2([0,\infty);\mathbb{R}^M)}\big)^{2m}}{m!} \le \frac{|\big(\lambda\|f\|_{L^2([0,\infty);\mathbb{R}^M)}\big)^2}{M!}.$$

PROOF: Set $E(t,\lambda) = e^{\lambda I_f(t) - \frac{\lambda^2}{2}\|\mathbf{1}_{[0,t)}f\|^2_{L^2([0,\infty);\mathbb{R}^M)}}$. Then

$$E(t,\lambda) = 1 + \lambda \int_0^t E(\tau,\lambda)\big(f(\tau), dw(\tau)\big)_{\mathbb{R}^M}.$$

Thus, if (cf. (90)) $I^{(m)}(t) \equiv I^{(m)}_{f^{\otimes m}}(t) = \frac{1}{m!}\tilde{I}^{(m)}_{f^{\otimes m}}$ and

$$R^0(t,\lambda) \equiv E(t,\lambda) \quad \text{and} \quad R^{M+1}(t,\lambda) \equiv \lambda \int_0^t R^M(\tau,\lambda)\big(f(\tau), dw(\tau)\big)_{\mathbb{R}^M}$$

for $M \ge 0$, 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 $M \ge 0$. Finally, if $A(t) = \int_0^t |f(\tau)|^2\, d\tau$, then

$$\mathbb{E}^{\mathcal{W}}\big[\big|R^0(t,\lambda)\big|^2\big] \le e^{|\lambda|^2 A(t)}\mathbb{E}^{\mathcal{W}}\big[E(t, 2\mathfrak{Re}\lambda)\big] = e^{|\lambda|^2 A(t)},$$

and

$$\mathbb{E}^{\mathcal{W}}\big[\big|R^{M+1}(t,\lambda)\big|^2\big] = |\lambda|^2 \int_0^t \mathbb{E}^{\mathcal{W}}\big[\big|R^M(\tau,\lambda)\big|^2\big]\dot{A}(\tau)\, d\tau.$$

Hence, the asserted estimate follows by induction on $M \ge 0$. □

[12] Remember that our integrals stay away from the diagonal.

THEOREM 97. *The span of*

$$\mathbb{R} \oplus \left\{\tilde{I}^{(m)}_{f^{\otimes m}}(\infty) : m \geq 1 \;\&\; f \in L^2\big([0,\infty);\mathbb{R}^M\big)\right\}$$

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 $\tilde{I}^{(m)}_{f^{\otimes m}}(\infty)$. By the preceding, we know that $\cos \circ I_f(\infty)$ and $\sin \circ I_f(\infty)$ are in $\mathbb{H}$ for all $f \in L^2\big([0,\infty);\mathbb{R}^M\big)$.

Next, observe that the space of functions $\Phi : C\big([0,\infty);\mathbb{R}^M\big) \longrightarrow \mathbb{R}$ which have the form

$$\Phi = F\big(I_{f_1}(\infty), \dots, I_{f_L}(\infty)\big)$$

for some $L \geq 1$, $F \in \mathscr{S}(\mathbb{R}^L;\mathbb{R})$, and $f_1, \dots, f_L \in L^2\big([0,\infty);\mathbb{R}^M\big)$ is dense in $L^2(\mathcal{W};\mathbb{R}^M)$. Indeed, this follows immediately from the density in $L^2(\mathcal{W};\mathbb{R})$ of the space of functions of the form

$$w \rightsquigarrow F\big(w(t_1), \dots, w(t_L)\big),$$

where $L \geq 1$, $F \in \mathscr{S}(\mathbb{R}^L;\mathbb{R})$, and $0 \leq t_0 < \cdots < t_L$.

Now suppose that $F : \mathscr{S}(\mathbb{R}^L;\mathbb{R})$ is given, and let $\hat{F}$ denote its Fourier transform. Then, by elementary Fourier analysis,

$$F_n(x) \equiv \big(2(4^n+1)\pi\big)^{-L} \sum_{\|\mathbf{m}\|_\infty \leq 4^n} e^{-\sqrt{-1}\,(2^{-n}\mathbf{m},x)_{\mathbb{R}^L}} \hat{F}(\mathbf{m}2^{-n}) \longrightarrow F(x),$$

both uniformly and boundedly, where $\mathbf{m} = (m_1, \dots, m_L) \in \mathbb{Z}^L$ and $\|\mathbf{m}\|_\infty = \max_{1 \leq \ell \leq L} |m_\ell|$. Finally, since $\overline{\hat{F}(\Xi)} = \hat{F}(-\Xi)$ for all $\Xi \in \mathbb{R}L$, we can write

$$\begin{aligned}
&\big(2(4^n+1)\pi\big)^{L} F_n\big(I_{f_1}(\infty), \dots, I_{f_L}(\infty)\big) \\
&\quad = \hat{F}(\mathbf{0}) + 2 \sum_{\substack{\mathbf{m} \in \mathbb{N}^L \\ 1 \leq \|\mathbf{m}\|_\infty \leq 4^n}} \Big(\mathfrak{Re}\big(\hat{F}(\mathbf{m}2^{-n})\big) \cos \circ I_{\mathbf{m},n}(\infty) \\
&\qquad\qquad + \mathfrak{Im}\big(\hat{F}(\mathbf{m}2^{-n})\big) \sin \circ I_{\mathbf{m},n}(\infty)\Big) \in \mathbb{H},
\end{aligned}$$

where $I_{\mathbf{m},n} \equiv 2^{-n} I_{f_{\mathbf{m}}}$ with $f_{\mathbf{m}} = \sum_{\ell=1}^{L} m_\ell f_\ell$. □

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.[13]

[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*

$$(x,\boldsymbol{\xi}) \in \mathbb{R}\times\mathcal{P}^2(\mathbb{R}^M) \longmapsto x + I_{\boldsymbol{\xi}}(\infty) \in L^2(\mathcal{W};\mathbb{R})$$

is a linear, isometric surjection. Hence, for each $\Phi \in L^2(\mathcal{W};\mathbb{R})$ *there is a* $\mathcal{W}$-*almost surely unique* $\boldsymbol{\xi} \in \mathcal{P}^2(\mathbb{R}^M)$ *such that*

$$\Phi = \mathbb{E}^{\mathcal{W}}[\Phi] + \int_0^\infty \big(\boldsymbol{\xi}(\tau), dw(\tau)\big)_{\mathbb{R}^M} \quad \mathcal{W}\text{-almost surely.}$$

In particular,

$$\mathbb{E}^{\mathcal{W}}\big[\Phi \,\big|\, \overline{W_t}\big] = \mathbb{E}^{\mathcal{W}}[\Phi] + \int_0^t \big(\boldsymbol{\xi}(\tau), dw(\tau)\big)_{\mathbb{R}^M} \quad \mathcal{W}\text{-almost surely for each } t \ge 0.$$

Finally, for any $\Phi \in L^1(\mathcal{W};\mathbb{R})$, *there is a version of* $(t,w) \rightsquigarrow \mathbb{E}^{\mathcal{W}}\big[\Phi \,\big|\, \overline{W_t}\big](w)$ *that is is continuous as function of* $t \ge 0$.

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 $f \in L^2\big([0,\infty);\mathbb{R}^M\big)$ and $m \ge 1$,

$$\tilde{I}^{(m)}_{f^{\otimes m}}(\infty) = m\int_0^\infty \big(\tilde{I}^{(m-1)}_{f^{\otimes(m-1)}}(\tau)f(\tau), dw(\tau)\big)_{\mathbb{R}^M},$$

and so, by the first part of Theorem 97, we are done.

Given the first assertion, the second assertion is obvious and shows that $(t,w) \rightsquigarrow E^{\mathcal{W}}[\Phi\,|, \overline{W_t}(w)$ 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 $\{\Phi_k : k \ge 1\} \subseteq L^2(\mathcal{W};\mathbb{R})$ so that $\Phi_n \longrightarrow \Phi$ in $L^1(\mathcal{W};\mathbb{R})$, and let $(t,w) \rightsquigarrow X_k(t,w)$ be a version of $(t,w) \rightsquigarrow E^{\mathcal{W}}[\Phi\,|, \overline{W_t}]$ which is continuous in t. Then, by Doob's inequality, for all $t > 0$,

$$\sup_{\ell > k} \mathcal{W}\big(\|X_\ell(\,\cdot\,) - X_k(\,\cdot\,)\|_{[0,t]} \ge \epsilon\big) \le \frac{\|X_\ell(t) - X_k(t)\|_{L^1(W;\mathbb{R})}}{\epsilon} \longrightarrow 0$$

as $k \to \infty$. Hence there exists a progressively measurable $X : [0,\infty)\times\Omega \longrightarrow \mathbb{R}$ to which $\{X_k(\,\cdot\,) : k \ge 1\}$ $\mathcal{W}$-almost surely converges uniformly on compacts, and, since, for each $t \ge 0$,

$$\lim_{k\to\infty} \big\|X_k(t) - E^{\mathcal{W}}\big[\Phi \,\big|\, \overline{W_t}\big]\big\|_{L^1(\mathcal{W};\mathbb{R})} = 0,$$

it follows that $X(t)$ is a version of $\mathbb{E}^{\mathcal{W}}[\Phi \,|\, \overline{W_t}]$. $\square$

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 $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ for which one knows that there is a progressively measurable function $t \rightsquigarrow A(t)$ which is continuous and non-decreasing in t and for which $\big(M(t)^2 - A(t),\mathcal{F}_t,\mathbb{P}\big)$ is a martingale. At the time, Doob did not know what is now called the Doob-Meyer decompostion 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 Decompostion Theorem: Doob noticed that if $\big(X_n,\mathcal{F}_n,\mathbb{P}\big)$ 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 $\big(X_n - A_n,\mathcal{F}_n,\mathbb{P}\big)$ 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 $\big(X_n - A_n,\mathcal{F}_n,\mathbb{P}\big)$ 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 $\big(X_n - B_n,\mathcal{F}_n,\mathbb{P}\big)$ 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 paratmeter 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 $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ is a continuous, local martingale, then there is a $\mathbb{P}$-almost surely unique continuous, non-decreasing, progressively measurable function $t \rightsquigarrow A(t)$ such that $A(0) = 0$ and $\big(M(t)^2 - A(t),\mathcal{F}_t,\mathbb{P}\big)$ 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 $\big(M(t)^2 - A(t),\mathcal{F}_t,\mathbb{P}\big)$ is a martingale and $\mathbb{E}^{\mathbb{P}}[A(t)] = E^{\mathbb{P}}[M(t)^2]$. In fact, if $\{\zeta_m : m \geq 1\}$ are stopping times for

[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 $\big(M(t)^2 - A(t), \mathcal{F}_t, \mathbb{P}\big)$, then $\mathbb{E}^{\mathbb{P}}[A(t\wedge\zeta_m)] = \mathbb{E}^{\mathbb{P}}[M(t\wedge\zeta_m)^2]$, $A(t\wedge\zeta_m) \nearrow A(t)$, and, as a consequence, $M(t\wedge\zeta_m)^2 - A(t\wedge\zeta_m) \longrightarrow M(t)^2 - A(t)$ in $L^1(\mathbb{P};\mathbb{R})$, and therefore $\big(M(t)^2 - A(t), \mathcal{F}_t, \mathbb{P}\big)$ 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 $\big(M(t\wedge\zeta_m)^2 - A(t\wedge\zeta_m), \mathcal{F}_t, \mathbb{P}\big)$ is a martingale for all $m \geq 1$. In addition, if $A'(\,\cdot\,)$ is a second continuous, non-decreasing, progressively measurable function for which $\big(M(t)^2 - A'(t), \mathcal{F}_t, \mathbb{P}\big)$ is a local martingale, then

$$\big(M(t\wedge\zeta_m)^2 - A'(t\wedge\zeta_m), \mathcal{F}_t, \mathbb{P}\big)$$

is a martingale for all $m \geq 1$, and so, by uniquenss, $A'(t) = A_m(t)$ for $t \in [0, \zeta_m)$. Having made this observtion, 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 $\big(X(t), \mathcal{F}_t, \mathbb{P}\big)$ 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, $\mathbb{E}^{\mathbb{P}}\big[X_R(t)^2\big] = \mathbb{E}^{\mathbb{P}}\Big[\int_0^t X_R(\tau)\,dX_R(\tau)\Big]$, and so $\mathbb{E}^{\mathbb{P}}[X(t\wedge\zeta_R)^2] = 0$. Since $\zeta_R \nearrow \infty$ as $R \to \infty$, it follows that $X(t) = 0$ (a.s.,$\mathbb{P}$). Now suppose that $\big(M(t) - A(t), \mathcal{F}_t, \mathbb{P}\big)$ and $\big(M(t) - B(t), \mathcal{F}_t, \mathbb{P}\big)$ are martingales, and take $\zeta_R = \inf\{t \geq 0 : A(t) \vee B(t) \geq R\}$. Then the preceding result applies to $\big(B(t\wedge\zeta_R) - A(t\wedge\zeta_R), \mathcal{F}_t, \mathbb{P}\big)$ 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 $\zeta_{m,n} \nearrow \infty$ as $m \to \infty$, proceed by induction on $m \geq 1$, and define $\zeta_{m+1,n+1}$ to be

$$\zeta_{\ell,n} \wedge \inf\{t \geq \zeta_{m,n+1} : |M(t) - M(\zeta_{m,n}| \geq 2^{-n}\}$$
$$\text{where } \ell \text{ is the element of } \mathbb{Z}^+ \text{ for which } \zeta_{\ell-1,n} < \zeta_{m,n+1} \leq \zeta_{\ell,n}.$$

Clearly, for each $n \geq 0$, $\{\zeta_{m,n} : n \geq 0\}$ is non-decreasing sequence of bounded stopping times that tend to ∞. 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, $\big(Y_n(t),\mathcal{F}_t,\mathbb{P}\big)$ 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 $\{A_n(\,\cdot\,)\restriction[0,T]:n\geq 0\}$ converges in $L^1\big(\mathbb{P};C([0,T];\mathbb{R})\big)$ or, equivalently, that $\{Y_n(\,\cdot\,)\restriction[0,T]:n\geq 0\}$ 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

$$Y_{n+1}(t)-Y_n(t)=\sum_{m=0}^{\infty}\big(M_{m,n+1}-\tilde{M}_{m,n+1}\big)\Delta_{m,n+1}(t).$$

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,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[\|Y_{n+1}(\,\cdot\,)-Y_n(\,\cdot\,)\|^2_{[0,T]}\big]&\leq 2^{-n}\sum_{m=0}^{\infty}\mathbb{E}^{\mathbb{P}}\big[\Delta_{m,n+1}(t)^2\big]\\&=2^{-n}\sum_{m=0}^{\infty}\big[M(t\wedge\zeta_{m+1,n+1})^2-M(t\wedge\zeta_{m,n+1})^2\big]=2^{-n}\mathbb{E}^{\mathbb{P}}\big[M(t)^2\big],\end{aligned}$$

and the desired convergence follows from this. $\square$

It is important to appreciated 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 $t\rightsquigarrow N(t)$, 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 $\langle\!\langle M\rangle\!\rangle(\,\cdot\,)$ to denote the function $A(\,\cdot\,)$ in Theorem 99. In addition, given a stopping time ζ, M^ζ will be the local martingale $t\rightsquigarrow M(t\wedge\zeta)$. The following lemma contains a few elementary facts that will be used in the next section.

Lemma 100. *Let* $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ *be a continuous local martingale and* ζ *a stopping time. Then* $\langle\langle M^\zeta\rangle\rangle(t)=\langle\langle M\rangle\rangle(t\wedge\zeta)$. *Furthermore, if* $\mathbb{E}^{\mathbb{P}}\big[\langle\langle M\rangle\rangle(\zeta)\big]<\infty$, *then*

$$\big(M^\zeta(t),\mathcal{F}_t,\mathbb{P}\big) \text{ and } \big(M^\zeta(t)^2-\langle\langle M^\zeta\rangle\rangle(t),\mathcal{F}_t,\mathbb{P}\big)$$

are martingales. Finally, if $\alpha:\Omega\longrightarrow\mathbb{R}$ *is a bounded,* $\mathcal{F}_\zeta$*-measurable function and* $\tilde M(t)=\alpha\big(M(t)-M^\zeta(t)\big)$, *then* $\big(\tilde M(t),\mathcal{F}_t,\mathbb{P}\big)$ *is a continuous, local martingale and*

$$\langle\langle\tilde M\rangle\rangle(t)=\alpha^2\big(\langle\langle M\rangle\rangle(t)-\langle\langle M^\zeta\rangle\rangle(t)\big).$$

Proof: Without loos in generality, I will assume that $M(0)=0$.

To prove the first asserion, simply observe that, by Doob's stopping time theorem, $\big(M^\zeta(t)^2-\langle\langle M\rangle\rangle(t\wedge\zeta),\mathcal{F}_t,\big)$ 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 $\mathbb{E}^{\mathbb{P}}\big[\langle\langle M\rangle\rangle(\zeta)\big]<\infty$. Choose stopping times $\{\zeta_m : m\ge 1\}$ for the local martingale $\big(M^\zeta(t),\mathcal{F}_r,\mathbb{P}\big)$. By Doob's inequality,

$$\mathbb{E}^{\mathbb{P}}\big[\|M^{\zeta_m\wedge\zeta}\|^2_{[0,t]}\big]\le 4\mathbb{E}^{\mathbb{P}}\big[\langle\langle M\rangle\rangle(t\wedge\zeta_m\wedge\zeta)\big]\le 4\mathbb{E}^{\mathbb{P}}\big[\langle\langle M\rangle\rangle(\zeta)\big],$$

and therefore $\mathbb{E}^{\mathbb{P}}\big[\|M^\zeta\|^2_{[0,\zeta)}\big]<\infty$. In particular, this means that $M^\zeta(t\wedge\zeta_m)\longrightarrow M^\zeta(t)$ in $L^2(\mathbb{P};\mathbb{R})$, and so, since $\langle\langle M^\zeta\rangle\rangle(t\wedge\zeta_m)\nearrow\langle\langle M^\zeta\rangle\rangle(t)$, there is nothing more to do.

To prove the final assertion, it suffices to prove that

$$\big(\tilde M(t),\mathcal{F}_t,\mathbb{P}\big) \text{ and } \Big(\tilde M(t)^2-\alpha^2\big(\langle\langle M(t)\rangle\rangle(t)-\langle\langle M\rangle\rangle(t\wedge\zeta)\big),\mathcal{F}_t,\mathbb{P}\Big)$$

are maringales if $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ is a bounded martingale. Thus, assume that $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ is a bounded martingale, and let $t>s$ and $\Gamma\in\mathcal{F}_s$ be given. Then

$$\mathbb{E}^{\mathbb{P}}\big[\tilde M(t),\,\Gamma\big]=\mathbb{E}^{\mathbb{P}}\big[\tilde M(t),\,\Gamma\cap\{\zeta\le s\}\big]+\mathbb{E}^{\mathbb{P}}\big[\tilde M(t),\,\Gamma\cap\{s<\zeta\le t\}\big].$$

Because $\Gamma\cap\{\zeta\le s\}\in\mathcal{F}_s$ and $\alpha\mathbf{1}_{\Gamma\cap\{\zeta\le s\}}$ is $\mathcal{F}_s$-measurable,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[\tilde M(t),\,\Gamma\cap\{\zeta\le s\}\big]&=\mathbb{E}^{\mathbb{P}}\big[\alpha\big(M(t)-M(t\wedge\zeta)\big),\,\Gamma\cap\{\zeta\le s\}\big]\\&=\mathbb{E}^{\mathbb{P}}\big[\alpha\big(M(s)-M(s\wedge\zeta)\big),\,\Gamma\cap\{\zeta\le s\}\big]=\mathbb{E}^{\mathbb{P}}\big[\tilde M(t),\,\Gamma\big].\end{aligned}$$

At the same time, $\alpha\mathbf{1}_{\Gamma\cap\{s<\zeta\le t\}}$ is $\mathcal{F}_{t\wedge\zeta}$-measurable, and so, by Hunt's version of the stopping time theorem,

$$\mathbb{E}^{\mathbb{P}}\big[\alpha M(t),\,\Gamma\cap\{s<\zeta\le t\}\big]=\mathbb{E}^{\mathbb{P}}\big[\alpha M(t\wedge\zeta),\,\Gamma\cap\{s<\zeta\le t\}\big],$$

which means that $\mathbb{E}^{\mathbb{P}}\big[\tilde M(t),\,\Gamma\cap\{s<\zeta\le t\}\big]=0$. Thus $\big(\tilde M(t),\mathcal{F}_t,\mathbb{P}\big)$ is a martingale. To show that $\Big(\tilde M(t)^2-\alpha^2\big(\langle\langle M\rangle\rangle(t)-\langle\langle M^\zeta\rangle\rangle(t)\big),\mathcal{F}_t,\mathbb{P}\Big)$ is martingale, use Hunt's theorem to see first that

$$\mathbb{E}^{\mathbb{P}}\big[\alpha^2M(t)M(t\wedge\zeta),\,\Gamma\cap\{s<\zeta\le t\}\big]=\mathbb{E}^{\mathbb{P}}\big[\alpha^2M(t\wedge\zeta)^2,\,\Gamma\cap\{s<\zeta\le t\}\big]$$

and then that

$$\mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(M(t)^2 - M(t\wedge\zeta)^2\big),\, \Gamma\cap\{s<\zeta\le t\}\big]$$
$$= \mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(\langle\!\langle M\rangle\!\rangle(t) - \langle\!\langle M^\zeta\rangle\!\rangle(t)\big),\, \Gamma\cap\{s<\zeta\le t\}\big].$$

Hence,

$$\mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(M(t)-M^\zeta(t)\big)^2,\, \Gamma\cap\{\zeta>s\}\big] = \mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(\langle\!\langle M\rangle\!\rangle(t)-\langle\!\langle M^\zeta\rangle\!\rangle(t)\big),\, \Gamma\cap\{\zeta>s\}\big].$$

Next, since $\alpha\mathbf{1}_{\zeta\le x}$ is $\mathcal{F}_s$-measurable,

$$\mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(M(t)-M^\zeta(t)\big)^2,\, \Gamma\cap\{\zeta\le s\}\big]$$
$$= \mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(M(t)^2 - 2M(t)M^\zeta(s) + M^\zeta(s)^2\big),\, \Gamma\cap\{\zeta\le s\}\big]$$
$$= \mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(M(s)^2 - 2M(s)M^\zeta(s) + M(s)^2 + \langle\!\langle M\rangle\!\rangle(t) - \langle\!\langle M\rangle\!\rangle(s),\, \Gamma\cap\{\zeta\le s\}\big]$$
$$= \mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(M(s)-M^\zeta(s)\big)^2,\, \Gamma\big] + \mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(\langle\!\langle M\rangle\!\rangle(t) - \langle\!\langle M^\zeta\rangle\!\rangle(s),\, \Gamma\cap\{\zeta\le s\}\big].$$

After combining these, one obtains

$$\mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(M(t)-M^\zeta(t)\big)^2 - \alpha^2\big(M(s)-M^\zeta(s)\big)^2,\, \Gamma\big]$$
$$= \mathbb{E}^{\mathbb{P}}\big[\alpha^2\big(\langle\!\langle M\rangle\!\rangle(t) - \langle\!\langle M^\zeta\rangle\!\rangle(s)\big),\, \Gamma\big] \quad \square.$$

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 $(\Omega,\mathcal{F},\mathbb{P})$ be a complete probability space and $\{\mathcal{F}_t : t\ge 0\}$ a non-dereasing filtration of complete sub σ-algebras, use $M^2(\mathbb{P};\mathbb{R})$ and $M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$, respectively, to denote the set of all square integral continuous martingales and local martingales under $\mathbb{P}$ relative to $\{\mathcal{F}_r : t\ge 0\}$. Clearly, both $M^2(\mathbb{P};\mathbb{R})$ and $M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$ are vector spaces over $\mathbb{R}$. Next, given $M_1, M_2 \in M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$, set

$$\langle M_1, M_2\rangle = \frac{\langle\!\langle M_1+M_2\rangle\!\rangle - \langle\!\langle M_1-M_2\rangle\!\rangle}{4},$$

and note that

$$\big(M_1(t)M_2(t) - \langle M_1,M_2\rangle(t), \mathcal{F}_t, \mathbb{P}\big)$$

is a continuous local martingale. Moreover, by exactly the same arguement 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 $A(0)=0$, $A(\,\cdot\,)\restriction[0,t]$ has bounded variation for each $t\ge 0$, and $\big(M_1(t)M_2(t)-A(t),\mathcal{F}_t,\mathbb{P}\big)$ is a continuous local martingale.

LEMMA 101. *The map* $(M_1, M_2) \rightsquigarrow \langle M_1.M_2\rangle$ *is symmetric and bilinear on* $M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})^2$ *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* $0 \leq s < t$,

$$\begin{aligned}&\big|\langle M_1, M_2\rangle(t) - \langle M_1, M_2\rangle(s)\big| \\ &\leq \sqrt{\langle\langle M_1\rangle\rangle(t) - \langle\langle M_1\rangle\rangle(s)}\sqrt{\langle\langle M_2\rangle\rangle(t) - \langle\langle M_2\rangle\rangle(s)}\end{aligned} \tag{102}$$

$\mathbb{P}$*-almost surely. In particular,*

$$\big\|\sqrt{\langle\langle M_2\rangle\rangle} - \sqrt{\langle\langle M_1\rangle\rangle}\big\|_{[0,t]} \leq \sqrt{\langle\langle M_2 - M_1\rangle\rangle(t)} \quad \text{(a.s.,}\mathbb{P}\text{)},$$

and for any stopping time ζ,

$$\langle M_1^\zeta, M_2\rangle(t) = \langle M_1, M_2\rangle(t \wedge \zeta).$$

PROOF: The first assertions are all easy applications of uniqeuess. 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 α and then for all α simultaneously. Thus, $\mathbb{P}$-almost surely,

$$2\big|\langle M_1, M_2\rangle\big| \leq \alpha^2\langle\langle M_1\rangle\rangle + \alpha^{-2}\langle\langle M_2\rangle\rangle,$$

and so, after minimizing with respect to α, one sees that

$$|\langle M_1, M_2\rangle| \leq \sqrt{\langle\langle M_1\rangle\rangle}\sqrt{\langle\langle M_2\rangle\rangle} \quad \text{(a.s., }\mathbb{P}\text{)}.$$

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 inqualities follows in the same way as the triangle inequality follows from Schwarz's inequality. To prove the second, note that, by (102),

$$\big|\langle M_1^\zeta, M_2\rangle(t) - \langle M_1^\zeta, M_2\rangle(t \wedge \zeta)\big| \leq \sqrt{\langle\langle M_1^\zeta\rangle\rangle(t) - \langle\langle M_1^\zeta\rangle\rangle(t \wedge \zeta)}\sqrt{\langle\langle M_2\rangle\rangle(t)} = 0$$

and

$$\big|\langle M_1^\zeta, M_2\rangle(t \wedge \zeta) - \langle M_1, M_2\rangle(t \wedge \zeta)\big| \leq \sqrt{\langle\langle M_1^\zeta - M_1\rangle\rangle(t \wedge \zeta)}\sqrt{\langle\langle M_2\rangle\rangle(t)} = 0,$$

since

$$\langle\!\langle M_1^\zeta - M_1\rangle\!\rangle(t\wedge\zeta) = \langle\!\langle (M_1^\zeta - M_1)^\zeta\rangle\!\rangle(t)$$

and $(M_1^\zeta - M_1)^\zeta = M_1^\zeta - M_1^\zeta = 0$. □

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 $\varphi \in C\big([0,T];\mathbb{R}\big)$,

$$2\left|\int_0^T \varphi(\tau)\,\langle M_1, M_2\rangle(d\tau)\right| \le \alpha^2\int_0^T \varphi(\tau)^2\,\langle\!\langle M_1\rangle\!\rangle(d\tau) + \alpha^{-2}\langle\!\langle M_2\rangle\!\rangle(T),$$

and therefore that

$$\left|\int_0^T \varphi(\tau)\,\langle M_1, M_2\rangle(d\tau)\right| \le \sqrt{\int_0^T \varphi(\tau)^2\langle\!\langle M_1\rangle\!\rangle(d\tau)}\sqrt{\langle\!\langle M_2\rangle\!\rangle(T)}.$$

Because this inequality holds for all continuous φ's on $[0,T]$, it also holds for all Borel measurable ones, and therefore we know that, for all Borel measurable $\varphi : [0,T] \longrightarrow \mathbb{R}$,

$$\left|\int_0^T |\varphi(\tau)|\,|\langle M_1, M_2\rangle|(d\tau)\right| \le \sqrt{\int_0^T \varphi(\tau)^2\langle\!\langle M_1\rangle\!\rangle(d\tau)}\sqrt{\langle\!\langle M_2\rangle\!\rangle(T)}, \tag{103}$$

where $|\langle M_1, M_2\rangle|(dt)$ is the variation measure determined by $\langle M_1, M_2\rangle(dt)$.

With these preparations, we can say how Kunita and Watenable defined stochastic integrals with respect to an element M of $M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$. Use $\mathcal{P}^2_{\mathrm{loc}} * M[\mathbb{R})$ to denote the space of progressively measurable funtions ξ such that $\int_0^t |\xi(\tau)|^2\,\langle\!\langle M\rangle\!\rangle(d\tau) < \infty$ for locally integrable with respect to $|\langle M, M'\rangle|(dt)$ of all $M' \in M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$, and define the stochastic integral I^M_ξ of ξ with respect to M to be the element of $M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$ such that $I^M_\xi(0) = 0$ and $\langle I^M_\xi, M'\rangle(dt) = \xi(t)\langle M, M'\rangle(dt)$ for all $M' \in M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$. It is obvious that this definition uniquely determines I^M_ξ since, if I and J were two such elements of $M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$, 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^M_ξ esists, then $I^{M^\zeta}_\xi$ and $I^M_{\mathbf{1}_{[0,\zeta)}\xi}$ exist and both are equal to $(I^M_\xi)^\zeta$. Next suppose that $\{\xi_n : n \ge 0\} \cup \{\xi\} \subseteq \mathcal{P}^2_{\mathrm{loc}}(M;\mathbb{R})$ and that*

$$\lim_{n\to\infty}\int_0^t \big(\xi_n(\tau) - \xi(\tau)\big)^2\langle\!\langle M\rangle\!\rangle(d\tau) = 0 \quad (\text{a.s.}, \mathbb{P})$$

for all $t \ge 0$. If $I^M_{\xi_n}$ exists for all $n \ge 0$, then I^M_ξ does also.

Proof: The first assertion is an easy application of the last part of Lemma 101. Indeed,

$$\begin{aligned}\langle (I_\xi^M)^\zeta, M'\rangle(t) &= \langle I_\xi^M, M'\rangle(t\wedge\zeta)\\ &= \int_0^{t\wedge\zeta}\xi(\tau)\langle M,M\rangle(d\tau) = \int_0^t \mathbf{1}_{[0,\zeta)}(\tau)\xi(\tau)\langle M,M'\rangle(d\tau),\end{aligned}$$

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

$$\|M\|_{[0,t]} \text{ and } \sup_{n\ge 0}\int_0^t \xi_n(\tau)^2\langle\!\langle M\rangle\!\rangle(d\tau)$$

are uniformly bounded for each $t\ge 0$, and set $I_n = I^M_{\xi_n}$. Then

$$\left(\big(I_n(t)-I_m(t)\big)^2 - \int_0^t \big(\xi_n(\tau)-\xi_m(\tau)\big)^2\langle\!\langle M\rangle\!\rangle(d\tau), \mathcal{F}_t, \mathbb{P}\right)$$

is a martingale, and so

$$\mathbb{E}^{\mathbb{P}}\big[\big(I_n(\,\cdot\,)-I_m(\,\cdot\,)\big)^2\big] = \mathbb{E}^{\mathbb{P}}\left[\int_0^t \big(\xi_n(\tau)-\xi_m(\tau)\big]^2\langle\!\langle M\rangle\!\rangle(d\tau)\right],$$

which, by Doob's inequality, means that

$$\lim_{m\to\infty}\sup_{n>m}\mathbb{E}^{\mathbb{P}}\big[\|I_n(\,\cdot\,)-I_m(\,\cdot\,)\|^2_{[0,t]}\big] = 0.$$

Hence there exists a continuous, square integrable martingle I such that

$$\|I-I_m\|_{[0,t]} \longrightarrow 0 \text{ in } L^2(\mathbb{P};\mathbb{R}) \text{ for all } t\ge 0.$$

In addition, if $M'\in M_{\text{loc}}(\mathbb{P};\mathbb{R})$ and $\langle\!\langle M'\rangle\!\rangle(t)$ is bounded for all $t\ge 0$, then, for $0\le s<t$,

$$\mathbb{E}^{\mathbb{P}}\big[I(t)M'(t)\,\big|\,\mathcal{F}_s\big] = \lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\big[I_n(t)M'(t)\,|\,\mathcal{F}_s\big] = \lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\left[\int_0^t \xi_n(\tau)\langle M,M'\rangle(d\tau)\right]$$

and

$$\int_0^t |\xi(\tau)-\xi_n(\tau)|\,|\langle M,M'\rangle|(d\tau) \le \sqrt{\int_0^t \big(\xi(\tau)-\xi_n(\tau)\big)^2\langle\!\langle M\rangle\!\rangle(d\tau)}\sqrt{\langle\!\langle M'\rangle\!\rangle(t)} \longrightarrow 0$$

$\mathbb{P}$-almost surely. Hence, $\langle I, M'\rangle(dt) = \xi(t)\langle M, M'\rangle(dt)$ (a.s.,$\mathbb{P}$) when $\langle\!\langle M'\rangle\!\rangle(t)$ is bounded for all $t \geq 0$. To prove that the same equality for general M', take $\zeta_k = \inf\{t \geq 0 : \langle\!\langle M\rangle\!\rangle(t) \geq k\}$. 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 $\zeta_k \nearrow \infty$ as $k \to \infty$, it follows that $I = I_\xi^M$.

To remove the boundedness assumptions on M and $\int_0^t \xi_n(\tau)^2 \langle\!\langle M\rangle\!\rangle(d\tau)$, define

$$\zeta_k = \inf\left\{t \geq 0 : |M(t)| vee \sup_{n\geq 0} \int_0^t \xi_n(\tau)^2 \langle\!\langle M\rangle\!\rangle(t) \geq k\right\},$$

and set $M_k = M^{\zeta_k}$. By the preceding, we know that $I_k \equiv I^{M_k}_{\mathbf{1}_{[0,\zeta_k)}\xi}$ exists and is equal to $I_\xi^{M_k}$, 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 $I \in M_{\mathrm{loc}}(\mathbb{P}; \mathbb{R})$ 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 $M' \in M_{\mathrm{loc}}(\mathbb{P}; \mathbb{R})$. $\Box$

In view of the preceding, what remains is to find a sufficiently rich set of ξ's for which we can show that I_ξ^M exists, and as in the case when M was a Brownian motion, a good guess is that ξ'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 $|\psi_n(0)| \vee \mathrm{var}_{[0,t]}(\psi) \leq C < \infty$. If $\psi_n \longrightarrow \psi$ pointwise, then $\mathrm{var}_{[0,t]}(\psi) \leq C$ 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 $\mathrm{var}_{[0,t]}(\psi) \leq C$ is obvious. Next, choose $\{\varphi_k : k \geq 1\} \subseteq C^1(\mathbb{R}; \mathbb{R})$ so that $\|\varphi_k\|_{\mathrm{u}} \leq \|\varphi\|_{[0,t]}$ and $\|\varphi - \varphi_k\|_{[0,t]} \leq \frac{1}{k}$. Then

$$\begin{aligned}&\left|\int_0^t \psi_n(\tau)\, d\varphi(\tau) - \int_0^t \psi_n(\tau)\, d\varphi_k(\tau)\right| \\ &\leq |\varphi(t) - \varphi_k(t)||\psi(t)| + \left|\int_0^t \big(\varphi(\tau) - \varphi_k(\tau)\big)\, d\psi_n(\tau)\right| \leq 2C\|\varphi - \varphi_k\|_{[0,t]} \leq \frac{2C}{k},\end{aligned}$$

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 $k\ge 1$. But, by Lebesgue's dominated convergence theorem,

$$\begin{aligned}\lim_{n\to\infty}\int_0^t \psi_n(\tau)\,d\varphi_k(\tau)&=\lim_{n\to\infty}\int_0^t \psi_n(\tau)\dot\varphi_k(\tau)\,d\tau\\&=\int_0^t \psi(\tau)\dot\varphi_k(\tau)\,d\tau=\int_0^t \psi(\tau)\,d\varphi_k(\tau).\end{aligned}$$

LEMMA 105. *Suppose $M\in M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$ and that ξ is an element of $\mathcal{P}^2_{\mathrm{loc}}(M;\mathbb{R})$ for which $\xi(\,\cdot\,)$ has locally bounded variation. Then I_ξ^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 ξ 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

$$I_n(t)\equiv\int_0^t \xi([\tau]_n)\,dM(\tau)=\sum_{m<2^nt}\xi(m2^{-n})\big(M(t\wedge(m+1)2^{-n})-M(t\wedge m2^{-n})\big),$$

and, because $\langle\!\langle M\rangle\!\rangle(dt)$ is non-atomic and ξ has at most a countable number of discontinuities,

$$\lim_{n\to\infty}\int_0^t\big(\xi(\tau)-\xi_n(\tau)\big)^2\,\langle\!\langle M\rangle\!\rangle(d\tau)=0,$$

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

$$I_n(t)=\sum_{m<2^nt}\Delta_{m2^{-n}}(t\wedge(m+1)2^{-n}).$$

where $\Delta_s(t)\equiv\xi(s)\big(M(t)-M(t\wedge s)\big)$. By Lemma 100, $\Delta_s\in M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$ for each $s\ge 0$, and so $I_n\in M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$. Next, let $M'\in M_{\mathrm{loc}}(\mathbb{P};\mathbb{R})$ be given, and set $X(t)=M(t)M'(t)-\langle M,M'\rangle(t)$. Then, again by Lemma 100,

$$\Big(\xi(s)\big(X(t)-X(t\wedge s)\big),\mathcal{F}_t,\mathbb{P}\Big)\text{ and }\Big(\xi(s)M(s)\big(M'(t)-M'(t\wedge s)\big),\mathcal{F}_t,\mathbb{P}\Big)$$

are local martingales. Hence, since

$$\begin{aligned}&\Delta_s(t)M'(t)-\xi(s)\big(\langle M,M'\rangle(t)-\langle M,M'\rangle(t\wedge s)\big)\\&\qquad=\xi(s)\big(X(t)-X(t\wedge s)\big)-\xi(s)M(s)\big(\langle\!\langle M'\rangle\!\rangle(t)-\langle\!\langle M'\rangle\!\rangle(t\wedge s)\big),\end{aligned}$$

$\langle \Delta_{m2^{-n}}, M' \rangle(t) = \xi(m2^{-n})\big(\langle M, M' \rangle(t) - \langle M, M' \rangle(t \wedge m2^{-n}\big)$, and therefore $\langle X_n, M' \rangle(dt) = \xi_n(t)\langle M, M' \rangle(dt)$.

To remove the boundedness assumption on ξ, for each $n \geq 1$, define $\xi_n(t) = \xi(t)$ if $|\xi(t)| \leq n$ and $\xi_n(t) = \pm n$ if $\pm\xi(t) > n$. Then $\mathrm{var}_{[0,t]}(\xi_n) \leq \mathrm{var}_{[0,t]}(\xi)$, $|\xi_n(t)| \leq n \wedge |\xi(t)|$, and $\int_0^t \big(\xi(\tau) - \xi_n(\tau)\big)^2 \langle\!\langle M \rangle\!\rangle(d\tau) \longrightarrow 0$. Hence, by the preceding combined with the above comment and Lemma 104, I_ξ^M exists,

$$I_{\xi_n}^M(t) \longrightarrow \int_0^t \xi(\tau)\, dM(\tau), \text{ and } I_{\xi_n}^M(t) \longrightarrow I_\xi^M.$$

Finally, to remove the assumption that M and $\langle\!\langle M \rangle\!\rangle$ are bounded, take $\zeta_n = \inf\{t \geq 0 : |M(t)| \vee \langle\!\langle M \rangle\!\rangle(t) \geq n\}$ and $M_n = M^{\zeta_n}$. Then, just as in the proof of Lemma 104, one can define I so that $I(t) = I_\xi^{M_n}(t)$ for $t \in [0, \zeta_n)$, in which case $I = I_\xi^M$ and $I(t) = \int_0^t \xi(\tau)\, dM(\tau)$. $\square$

To complete the program, we have to show that if $\xi \in \mathcal{P}^2_{\mathrm{loc}}(M; \mathbb{R})$ then there exist $\{\xi_n : n \geq 1\} \subseteq \mathcal{P}^2_{\mathrm{loc}}(M; \mathbb{R})$ such that each ξ_n has locally bounded variation and $\int_0^t \big(\xi(\tau) - \xi_n\big)^2 \langle\!\langle M \rangle\!\rangle(d\tau) \longrightarrow 0$. 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 $F : [0,\infty) \longrightarrow \mathbb{R}$ be a continuous, non-decreasing function with $F(0) = 0$, and define*

$$F^{-1}(t) = \inf\{\tau \geq 0 : F(\tau) \geq t\} \text{ for } t \geq 0.$$

Then F^{-1} is a right-continuous and non-decreasing on $\big[0, F(\infty)\big)$, $F \circ F^{-1}(t) = t$, $F^{-1} \circ F(\tau) \leq \tau$, and

$$D \equiv \{\tau : F^{-1} \circ F(\tau) < \tau\}$$

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

$$\int_0^{F(T)} f \circ F^{-1}(t)\, dt = \int_0^T f(d\tau)\, F(d\tau)$$

for non-negative, Borel measurable f on $[0,\infty)$. Finally, set $\rho_\epsilon(t) = \epsilon^{-1}\rho(\epsilon^{-1}t)$ where $\rho \in C^\infty\big(\mathbb{R}; [0,\infty)\big)$ vanishes off of $(0,1)$ and $\int \rho(t)\, dt = 1$, and, given $f \in L^2\big(F; [0,\infty)\big)$, set

$$f_\epsilon(\tau) = \int_0^{F(\infty)} \rho_\epsilon\big(F(\tau) - \sigma\big) f \circ F^{-1}(\sigma)\, d\sigma.$$

Then $\|f_\epsilon\|_{L^2(F;\mathbb{R})} \leq \|f\|_{L^2(F;\mathbb{R})}$ and $\|f - f_\epsilon\|_{L^2(F;\mathbb{R})} \longrightarrow 0$ as $\epsilon \searrow 0$.

PROOF: It is easy to check that F^{-1} is right-continuous, non-decreasing, and satisfies $F^{-1}\circ F(\tau)\le\tau$. 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 $F\big((a,b]\big)=F(a)-F(b)$, it follows that $F(D)=0$.

Next, by the standard change of variables formula for Riemann-Stieltjes integrals,

$$\int_0^T f\circ F(\tau)\,dF(\tau)=\int_0^{F(T)} f(t)\,dt \text{ for } f\in C\big([0,\infty);\mathbb{R}\big),$$

and so

$$\int_0^T f\circ F(\tau)\,F(d\tau)=\int_0^{F(T)} f(t)\,dt$$

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,

$$\begin{aligned}\int_0^\infty f_\epsilon(\tau)^2\,F(d\tau)&=\int_0^{F(\infty)}\rho_\epsilon*(f\circ F^{-1})(t)^2\,dt\\&\le\int_0^{F(\infty)} f\circ F^{-1}(t)^2\,dt=\int_0^\infty f(\tau)^2\,F(d\tau),\end{aligned}$$

and similarly

$$\|f-f_\epsilon\|^2_{L^2(F;\mathbb{R})}=\int_0^{F(\infty)}\Big(f\circ F^{-1}(t)-\rho_\epsilon*(f\circ F^{-1})(t)\Big)^2\,dt\longrightarrow 0\quad\square$$

THEOREM 107. *For each $M\in M_{\text{loc}}(\mathbb{P};\mathbb{R})$ and $\xi\in\mathcal{P}^2_{\text{loc}}(M;\mathbb{R})$ there exists a unique $I^M_\xi\in M_{\text{loc}}(\mathbb{P};\mathbb{R})$ such that $\langle I^M_\xi,M'\rangle(dt)=\xi(t)\langle M,M'\rangle(dt)$ for all $M'\in M_{\text{loc}}(\mathbb{P};\mathbb{R})$. Furthermore, if $\xi(\,\cdot\,)$ has locally bounded variation, then $I^M_\xi(t)$ equals the Riemann-Stieltjes integral on $[0,t]$ of ξ with respect to M.*

PROOF: All that remains is to use the existence assertion when ξ isn't of locally bounded variation. For this purpose, define ξ_ϵ from ξ by the prescription in Lemma 106 with $F=\langle\langle M\rangle\rangle$ and $f=\xi$. Then ξ_ϵ has locally bounded variation and

$$\lim_{\epsilon\searrow 0}\int_0^t\big(\xi(\tau)-\xi_\epsilon(\tau)\big)^2\langle\langle M\rangle\rangle(d\tau)=0.$$

Thus, by Lemma 101, we are done. $\square$

In the future, we will call I_ξ^M the **stochastic integral** of ξ with respect to M and will usually use $\int_0^t \xi(\tau)\,dM(\tau)$ to denote $I_\xi^M(t)$. Notice that if $\xi \in \mathcal{P}^2_{\text{loc}}(M;\mathbb{R})$ and $\eta \in \mathcal{P}^2_{\text{loc}}(I_\xi^M;\mathbb{R})$, then $\xi\eta \in \mathcal{P}^2_{\text{loc}}(M;\mathbb{R})$ and

$$\int_0^t \eta(\tau)\,dI_\xi^M(\tau) = \int_0^t \xi(\tau)\eta(\tau)\,dM(\tau). \tag{108}$$

To check this, set $J(t) = \int_0^t \eta(\tau)\,dI_\xi^M(\tau)$. Then, for any $M' \in M_{\text{loc}}(\mathbb{P};\mathbb{R})$,

$$\langle J, M'\rangle(dt) = \eta(t)\langle I_\xi^M, M'\rangle(dt) = \xi(t)\eta(t)\langle M, M'\rangle(dt).$$

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 $1 \le i \le N_1$, let V_i be a continuous, progressively measurable $\mathbb{R}$-valued function of locally bounded variation, and for each $1 \le j \le N_2$ let $\big(M_j(t).\mathcal{F}_t,\mathbb{P}\big)$ be a continuous local martingale. Set $\vec{V}(t) = \big(V_1,\dots,V_{N_1}\big)$, $\vec{M}(t) = \big(M_1(t),\dots,M_N(t)\big)$, and*

$$A(t) = \big(\big(\langle M_i, M_j\rangle\big)\big)_{1\le i,j\le N}.$$

If $\varphi \in C^{1,2}(\mathbb{R}^{N_1}\times\mathbb{R}^{N_2};\mathbb{C})$, then

$$\begin{aligned}
\varphi\big(\vec{V}(t),\vec{M}(t)\big) = \varphi\big(\vec{V}(0),\vec{M}(0)\big)&\\
+\int_0^t \Big(\nabla_{(1)}\varphi\big(\vec{V}(\tau),\vec{M}(\tau)\big), d\vec{V}(\tau)\Big)_{\mathbb{R}^{N_1}}&\\
+\int_0^t \Big(\nabla_{(2)}\varphi\big(\vec{V}(\tau),\vec{M}(\tau)\big), d\vec{M}(\tau)\Big)_{\mathbb{R}^{N_2}}&\\
+\tfrac12\int_0^t \operatorname{Trace}\Big(\nabla^2_{(2)}\varphi\big(\vec{V}(\tau),\vec{M}(\tau)\big)dA(\tau)\Big)&.
\end{aligned}$$

Proof: Begin by observing that for $0 \le x < t$, $A(t) - A(s)$ is non-negative definite and symmmetric, and therefore $\|A(t)-A(s)\|_{\text{op}} \le \operatorname{Trace}\big(A(t)-A(s)\big)$.

It is easy to reduce to the case when $\varphi \in C_c^\infty(\mathbb{R}^{N_1}\times\mathbb{R}^{N_2};\mathbb{R})$ and $\operatorname{var}_{[0,\infty)}(V)$, $\|M\|_{[0,\infty]}$, and $\sup_{t\ge0}\|A(t)\|_{\text{op}}$ are all uniformly bounded. Thus we will proceed under these assumptions. In particular, this means that $\big(\vec{M}(t),\mathcal{F}_t,\mathbb{P}\big)$ 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 $|\vec{M}(t)-\vec{M}(s)|$.

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 $m \geq 1$,

$$\zeta_{m,n} = \inf\{t \geq 0 : |\vec{M}(t) - \vec{M}(\zeta_{m-1,n}| \vee \operatorname{Trace}\big(A(t) - A(\zeta_{m-1,n}\big) \geq 2^{-n}\big\} \wedge T.$$

By continuity, $\zeta_{m,n} = T$ for all but a finite number of m's, and so

$$\begin{aligned}
&\varphi\big(\vec{V}(T), \vec{M}(T)\big) - \varphi\big(\vec{X}(0)\big) \\
&= \sum_{m=0}^{\infty} \int_{\zeta_{m,n}}^{\zeta_{m+1,n}} \Big(\nabla_{(1)}\varphi\big(\vec{V}(\tau), M(\zeta_{m+1,n})\big), d\vec{V}(\tau)\big) \\
&\quad + \sum_{m=0}^{\infty} \Big(\varphi\big(\vec{V}(\zeta_{m,n}), M(\zeta_{m+1,n})\big) - \varphi\big(\vec{V}(\zeta_{m,n}), M(\zeta_{m,n})\big)\Big).
\end{aligned}$$

Clearly the first sum on the right tends to $\int_0^T \big(\nabla_{(1)}\varphi(\vec{V}(\tau), \vec{M}(\tau), d\vec{V}(\tau)\big)_{R^{N_1}}$ as $n \to \infty$. To handle the second sum, use Taylor's theorem to write

$$\begin{aligned}
&\Big(\varphi\big(\vec{V}(\zeta_{m,n}), M(\zeta_{m+1,n})\big) - \varphi\big(\vec{V}(\zeta_{m,n}), M(\zeta_{m,n})\big)\Big) \\
&\quad = \Big(\nabla_{(2)}\varphi(\vec{X}_{m,n}), \Delta_{m,n}\Big)_{\mathbb{R}^{N_2}} + \tfrac{1}{2}\operatorname{Trace}\Big(\nabla^2_{(2)}\varphi(\vec{X}_{m,n})\Delta_{m,n} \otimes \Delta_{m,n}\Big) + E_{m,n},
\end{aligned}$$

where $\vec{X}_{m,n} = \big(\vec{V}(\zeta_{m,n}), \vec{M}(\zeta_{m,n})\big)$, $\Delta_{m,n} = \vec{M}(\zeta_{m+1,n}) - \vec{M}(\zeta_{m,n})$, and $|E_{m,n}| \leq C|\Delta_{m,n}|^3$ for some $C < \infty$. Using Lemma 100, it is easy to show that

$$\lim_{n\to\infty} \sum_{m=0}^{\infty} \Big(\nabla_{(2)}\varphi(\vec{X}_{m,n}), \Delta_{m,n}\Big)_{\mathbb{R}^{N_2}} = \int_0^T \Big(\nabla_{(2)}\varphi\big(\vec{V}(\tau), \vec{M}(\tau)\big), d\vec{M}(\tau)\Big)_{\mathbb{R}^{N_2}},$$

and, by Hunt's stopping time theorem,

$$\sum_{m=0}^{\infty} \mathbb{E}^{\mathbb{P}}\big[|\Delta_{m,n}|^2\big] = \sum_{m=0}^{\infty} \mathbb{E}^{\mathbb{P}}\big[|M(\zeta_{m+1,n})|^2 - |M(\zeta_{m,n})|^2\big] = \mathbb{E}^{\mathbb{P}}\big[|M(T)|^2\big],$$

and so

$$\sum_{m=0}^{\infty} \mathbb{E}^{\mathbb{P}}\big[|E_{m,n}|\big] \leq 2^{-n} C \mathbb{E}^{\mathbb{P}}\big[|M(T)|^2\big].$$

Finally, set $D_{m,n} = \Delta_{m,n} \otimes \Delta_{m,n} - \big(A(\zeta_{m_1,n}) - A(\zeta_{m,n})\big)$. Then

$$\begin{aligned}
&\sum_{m=0}^{\infty} \operatorname{Trace}\Big(\nabla^2_{(2)}\varphi(\vec{X}_{m,n})\Delta_{m,n} \otimes \Delta_{m,n}\Big) \\
&\quad = \sum_{m=0}^{\infty} \operatorname{Trace}\Big(\nabla^2_{(2)}\varphi(\vec{X}_{m,n})\big(A(\zeta_{m+1,n}) - A(\zeta_{m,n})\big)\Big) \\
&\qquad + \sum_{m=0}^{\infty} \operatorname{Trace}\Big(\nabla^2_{(2)}\varphi(\vec{X}_{m,n}) D_{m,n}\Big),
\end{aligned}$$

and the first sum on the right tends to

$$\int_0^T \mathrm{Trace}\Big(\nabla_{(2)}\varphi\big(\vec{V}(\tau),\vec{M}(\tau)\big)dA(\tau)\Big)$$

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

$$\sum_{m=0}^{\infty}\mathbb{E}^{\mathbb{P}}\big[\|D_{m,n}\|^2_{\mathrm{H.S.}}\big].$$

Since $\|D_{m,n}\|^2_{\mathrm{H.S.}}\le C\big(|\Delta_{m,n}|^4+\|A(\zeta_{m+1,n})-A(\zeta_{m,n})\|^2_{\mathrm{op}}\big)$ for some $C<\infty$, the preceding sum is dominated by a constant times

$$2^{-n}\mathbb{E}^{\mathbb{P}}\big[|\vec{M}(T)|^2+\mathrm{Trace}\big(A(T)\big)\big].\quad\square$$

A particularly striking application of this result is Kunita and Watanabe's derivation of Lévy's characterization of Brownian motion.

Corollary 110. *If $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ is an $\mathbb{R}^N$-valued, continuous local martingale, then it is a Brownian motion if and only if $M(0)=\mathbf{0}$ and $\langle M_i,M_j\rangle=t\delta_{i,j}$ for $1\le i,j\le N$.*

Proof: The necessity is obvious. To prove the sufficiency, first observe that, because $\langle\langle M_i\rangle\rangle(t)$ is bounded for all $1\le i\le N$, $t\ge0$, $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ a martingale. Next, given $\boldsymbol{\xi}\in\mathbb{R}^N$, apply Theorem 109 to show that

$$E_{i\xi}(t)\equiv e^{i(\boldsymbol{\xi},M(t))_{\mathbb{R}^N}+\frac{|\boldsymbol{\xi}|^2t}{2}}=1+\int_0^t e^{i(\boldsymbol{\xi},M(\tau))_{\mathbb{R}^N}+\frac{|\boldsymbol{\xi}|^2\tau}{2}}\big(\boldsymbol{\xi},dM(\tau)\big)_{\mathbb{R}^N}.$$

Finally, if $X(t)$ and $Y(t)$ denote the real and imaginary parts of the preceding stochastic integral, check that $\langle\langle X\rangle\rangle(t)+\langle\langle Y\rangle\rangle(t)\le e^{|\boldsymbol{\xi}|^2t}-1$, and therefore $\big(E_{i\mathbf{x}}(t),\mathcal{F}_t,\mathbb{P}\big)$ is a martingale. Since this means that

$$\mathbb{E}^{\mathbb{P}}\Big[e^{i\big(\boldsymbol{\xi},M(t)-M(s)\big)_{\mathbb{R}^N}}\,\Big|\,\mathcal{F}_s\Big]=e^{-\frac{\|\boldsymbol{\xi}|^2(t-s)}{2}}\ \text{for } 0\le s<t,$$

the proof is complete. $\square$

Another important consequence of Theorem 109 is what it says about the way local martingales transform under smaooth 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 stochasitic processes. Thus given an $\mathbb{R}$-valued progressively measurable function $t \rightsquigarrow X(t)$, one says that $\big(X(t), \mathcal{F}_t, \mathbb{P}\big)$ is a continuous **semi-martingale** if $X = M + V$, where $\big(M(t), \mathcal{F}_t, \mathbb{P}\big)$ is a continuous local martingale with $M(0) = 0$ and V is a progressively measurable fucntion for which $t \rightsquigarrow V(t)$ is a continuous function of locally bounded variation. By the same arguement 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 bounded variation part V of a continuous semi-martingales $X = M + V$, 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 ξ is an $\mathbb{R}$-valued, continuous, progressivley measurable function,

$$\int_0^t \xi(\tau)\, dX(\tau) \equiv \int_0^t \xi(\tau)\, dM(\tau) + \int_0^t \xi(\tau)\, dV(\tau),$$

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 properies of Riemann-Stieltjes integrals, one sees that

$$\int_0^t \eta(\tau)\, d\left(\int_0^\tau \xi(\sigma)\, dX(\sigma)\right) = \int_0^t \xi(\tau)\eta(\tau)\, dX(\tau) \tag{111}$$

if η is a second continuous, progressively measurable function.

The following statement is an immediate consequence of Theorem 109.

Corollary 112. *Suppose that $\vec{X}(t) = \big(X_1(t). \dots, X_N(t)\big)$, where $\big(X_j(t), \mathcal{F}_t, \mathbb{P}\big)$ is a continuous local semi-martingale for each $1 \le j \le N$, and set*

$$A(t) = \Big(\big(\langle X_i(t), X_j(t)\rangle\big)\Big)_{1\le i,j\le N}.$$

If $\varphi \in C^2(\mathbb{R}^N; \mathbb{C})$, then

$$\begin{aligned}\varphi\big(\vec{X}(t)\big) - \varphi\big(\vec{X}(0)\big) = &\int_0^t \Big(\nabla\varphi\big(\vec{X}(\tau)\big), d\vec{X}(\tau)\Big)_{\mathbb{R}^N} \\ &+ \tfrac{1}{2}\int_0^t \operatorname{Trace}\Big(\nabla^2\varphi\big(\vec{X}(\tau)\big) dA(\tau)\Big).\end{aligned}$$

Corollary 113. *If $\big(X_1(t), \mathcal{F}_t, \mathbb{P}\big)$ and $\big(X_2(t), \mathcal{F}_t, \mathbb{P}\big)$ are a pair of continuous local semi-martingales, then, for all $T > 0$,*

$$\begin{aligned}\lim_{n\to\infty} \sup_{t\in[0,T]} \Bigg|\sum_{m=0}^\infty &\big(X_1(t\wedge(m+1)2^{-n}) - X_1(t\wedge m2^{-n}\big) \\ &\times \big(X_2(t\wedge(m+1)2^{-n}) - X_2(t\wedge m2^{-n}\big) - \langle X_1, X_2\rangle(t)\Bigg| = 0\end{aligned}$$

in $\mathbb{P}$-probability.

Proof: First note that, by polarization, it suffices to treat case when $X_1 = X = X_2$ where X and $\langle\!\langle X\rangle\!\rangle$ are uniformly bounded.

Observe that, by Corollary 112,

$$\begin{aligned}\sum_{m=0}^{\infty}&\big(X(t\wedge(m+1)2^{-n}) - X(t\wedge m2^{-n}\big)^2\\ &= \sum_{m=0}^{\infty}\Big(X(t\wedge(m+1)2^{-n})^2 - X(t\wedge m2^{-n})^2\\ &\qquad\qquad - 2X(m2^{-n})\big(X(t\wedge(m+1)2^{-n}) - X(t\wedge m2^{-n})\big)\Big)\\ &= \langle\!\langle X\rangle\!\rangle(t) + 2\int_0^t \big(X(\tau) - X([\tau]_n)\big)\,dX(\tau).\end{aligned}$$

Next let M and V denote the martingale and bounded variation parts of X. Clearly

$$\sup_{t\in[0,T]}\left|\int_0^t \big(X(\tau)-X([\tau]_n)\big)\,dV(\tau)\right| \le \int_0^T \big|X(\tau)-X([\tau]_n)\big|\,d|V|(\tau) \longrightarrow 0$$

as $n\to\infty$. At the same time,

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\sup_{t\in[0,T]}\left(\int_0^t \big(X(\tau)-X([\tau]_n)\big)\,dM(\tau)\right)^2\right]\\ &\qquad\le 4\mathbb{E}^{\mathbb{P}}\left[\int_0^T \big(X(\tau)-X([\tau]_n)\big)^2\,\langle\!\langle M\rangle\!\rangle(d\tau)\right]\longrightarrow 0,\end{aligned}$$

and, when combined with the preceding, that completes the proof. □

Representing Continuous Local Martingales: Suppose that $\big(M(t),\mathcal{F}_t,\mathbb{P}\big)$ is an $\mathbb{R}^N$-valued continuous local martingale with $M(0)=\mathbf{0}$, and set

$$A(t) = \Big(\big(\langle M_i, M_j\rangle(t)\big)\Big)_{1\le i,j\le N}.$$

Further, assume that $t\rightsquigarrow\langle\!\langle M_j\rangle\!\rangle(t)$ is absolutely continuous for each $1\le j\le N$, and therefore, by Lemma 100, for each $1\le i,j\le N$, $t\rightsquigarrow\langle M_i,M_j\rangle(t)$ is also absolutely continuous. Then there exists a progressively measurable $\mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$-valued fucntion 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 $0 \le s < t$, we can and will assume that $a(t)$ is also symmetric and non-negative definite for all $t \ge 0$. By Lemma 29, $\big(\epsilon\mathbf{I} + a\big)^{\frac{1}{2}}$ is progressively measurable for all $\epsilon > 0$, and so $\sigma \equiv a^{\frac{1}{2}} = \lim_{\epsilon\searrow 0}\big(\epsilon\mathbf{I} + a\big)^{\frac{1}{2}}$ is also. Furthermore,

$$\int_0^t \sigma(\tau)^2\, d\tau = A(t),$$

and so one can integrate σ with respect to a Brownian motion.

Now assume that a, and therefore σ, are strictly positive definite. Then σ^{-1} is progressively measurable. In addition,

$$\int_0^t \big(\sigma^{-1}\boldsymbol{\xi}, dA(\tau)\sigma^{-1}\boldsymbol{\eta}\big)_{\mathbb{R}^N} = t(\boldsymbol{\xi},\boldsymbol{\eta})_{\mathbb{R}^N} \text{ for all } \boldsymbol{\xi} \in \mathbb{R}^N,$$

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 $\big(B(t), \mathcal{F}_t, \mathbb{P}\big)$ 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 postitive definite, the preceding arguement 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

$N(t)$ is the null space and $\sigma^{-1}(t) \restriction N(t)^\perp$ is the inverse of $\sigma(t) \restriction N(t)^\perp$. Then $\sigma\sigma^{-1} = \sigma^{-1}\sigma = \Pi^\perp \equiv \mathbf{I} - \Pi$, and, since

$$\Pi(t) = \lim_{\epsilon \searrow 0} a(t)\big(\epsilon\mathbf{I} + a(t)\big)^{-1} \text{ and } \sigma^{-1}(t) = \lim_{\epsilon \searrow 0} \sigma(t)\big(\epsilon\mathbf{I} + a(t)\big)^{-1},$$

both these functions are progressively measurable. In particular, for any $\boldsymbol{\xi} \in \mathbb{R}^N$,

$$\int_0^t \big(\sigma^{-1}(\tau)\boldsymbol{\xi}, dA(\tau)\sigma^{-1}\boldsymbol{\xi}\big)_{\mathbb{R}^N} = \int_0^t |\Pi(\tau)\boldsymbol{\xi}|^2\, d\tau \le t|\boldsymbol{\xi}|^2,$$

and so stochastic integrals of σ^{-1} with respect to M are well defined. Now take $\mathcal{W}$ to be Wiener measure on $C\big([0,\infty);\mathbb{R}^N\big)$, $\tilde{\mathbb{P}} = \mathbb{P}\times\mathcal{W}$, and $\tilde{\mathcal{F}}_t$ to be the completion of $\mathcal{F}_t\times W_t$ with respect to $\tilde{\mathbb{P}}$, and define $\tilde{M}(t)(w,\omega) = M(t)(\omega)$, $\tilde{A}(t)(w,\omega) = A(t)(\omega)$, $\tilde{\Pi}(t)(w,\omega) = \Pi(t)(\omega)$, $\tilde{\sigma}(t)(w,\omega) = \sigma(t)(\omega)$, and $\tilde{\sigma}^{-1}(w,\omega) = \sigma^{-1}(\omega)$. It is then an easy matter to check that $\big(\tilde{M}(t), \tilde{\mathcal{F}}_t, \tilde{\mathbb{P}}\big)$ is a local martingale and that $\big(w(t), \tilde{\mathcal{F}}_t, \tilde{\mathbb{P}}\big)$ is a Brownian motion. In addition, $\langle w_i, \tilde{M}_j\rangle = 0$. Thus, we can take

$$\tilde{B}(t) \equiv \int_0^t \tilde{\sigma}^{-1}(\tau)\, d\tilde{M}(\tau) + \int_0^t \tilde{\Pi}(\tau)^\perp\, dw(\tau),$$

and, using the properties discussed above, one sees that $\langle \tilde{B}_i, \tilde{B}_j\rangle(t) = t\delta_{i,j}$ and therefore that $\big(\tilde{B}(t), \tilde{\mathcal{F}}_t, \tilde{\mathbb{P}}\big)$ is a Brownian motion. Further, if $\tilde{I}(t) = \int_0^t \tilde{\sigma}(\tau)\, d\tilde{B}(\tau)$, 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 $\big(M(t), \mathcal{F}_t, \mathbb{P}\big)$ be an $\mathbb{R}^N$-valued continuous local martingale on $(\Omega, \mathcal{F})$, assume that $\langle\!\langle M_i\rangle\!\rangle$ is absolutely continuous for each $1 \le i \le N$, and define σ as in the preceding. Then there is a probability space $\big(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}\big)$ on which there is a Brownian motion $\big(\tilde{B}(t), \tilde{\mathcal{F}}_t, \tilde{\mathbb{P}}\big)$ and measurable map F to Ω such that $\mathbb{P} = F_*\tilde{\mathbb{P}}$ and*

$$\tilde{M}(t) = \tilde{M}(0) + \int_0^t \tilde{\sigma}(\tau)\, d\tilde{B}(\tau)$$

when $\tilde{M} = M \circ F$ 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 $a : \mathbb{R}^N \longrightarrow \mathrm{Hom}(\mathbb{R}^N;\mathbb{R}^N)$ is a symmetric, non-definite definite matrix valued function and $b : \mathbb{R}^N \longrightarrow \mathbb{R}^N$. If*

$$\sup_{\substack{\mathbf{x},\mathbf{y}\in\mathbb{R}^N \\ \mathbf{y}\ne\mathbf{x}}} \frac{\|a^{\frac{1}{2}}(\mathbf{y}) - a^{\frac{1}{2}}(\mathbf{x})\|_{\mathrm{H.S.}} \vee |b(\mathbf{y}) - b(\mathbf{x})|}{|\mathbf{y}-\mathbf{x}|} < \infty,$$

then, for each $\mathbf{x} \in \mathbb{R}^N$, the martingale problem for the L in (12) has precisely one solution starting at $\mathbf{x}$.

PROOF: Consider the solution $X(\,\cdot\,,\mathbf{x})$ to (56) when $\sigma = a^{\frac{1}{2}}$ and $M = N$. We know that the distribution of $X(\,\cdot\,,\mathbf{x})$ under $\mathcal{W}$ solves the martingale problem for L starting at $\mathbf{x}$. In addition, if $\big(\tilde{B}(t),\tilde{\mathcal{F}}_t,\tilde{\mathbb{P}}\big)$ is any $\mathbb{R}^N$-valued Brownian motion and

$$(*)\qquad \tilde{X}(t,\mathbf{x}) = \mathbf{x} + \int_0^t \sigma\big(\tilde{X}(\tau,\mathbf{x})\big)\,d\tilde{B}(\tau) + \int_0^t b\big(\tilde{X}(\tau,\mathbf{x})\big)\,d\tau,$$

then $\tilde{X}(\,\cdot\,,\mathbf{x})$ has the same distribution under $\tilde{\mathbb{P}}$ as $X(\,\cdot\,,\mathbf{x})$ has under $\mathcal{W}$. Indeed, (*) has only one solution and one can construct that solution from $\tilde{B}(\,\cdot\,)$ by the same Euler approximation procedure as we used to construct $X(\,\cdot\,,\mathbf{x})$ 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 $\tilde{X}(\,\cdot\,,\mathbf{x})$ and $X(\,\cdot\,,\mathbf{x})$. Thus, all that we have to do is show that if $\mathbb{P}$ solves the martingale problem for L starting at $\mathbf{x}$, then there is a Brownian motion $\big(\tilde{B}(t),\tilde{\mathcal{F}}_t,\tilde{\mathbb{P}}\big)$ such that $\mathbb{P}$ is the distribution of the solution to (*). To that end, set $M(t) = \psi(t) - \mathbf{x} - \int_0^t b\big(\psi(\tau)\big)\,d\tau$. Then $\big(M(t),\mathcal{B}_t,\mathbb{P}\big)$ is a continuous martingale, and so Theorem 114 says that there is a $\big(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}}\big)$, a Brownian motion $\big(\tilde{B}(t),\tilde{\mathcal{F}}_t,\tilde{\mathbb{P}}\big)$, and a measure preserving map $F:\tilde{\Omega}\longrightarrow C\big([0,\infty);\mathbb{R}^N\big)$ such that

$$\tilde{X}(t,\mathbf{x}) - \mathbf{x} - \int_0^t b\big(\tilde{X}(\tau,\mathbf{x})\big)\,d\tau = \int_0^t \sigma\big(\tilde{X}(\tau,\mathbf{x})\big)\,d\tilde{B}(\tau)$$

when $\tilde{X}(\,\cdot\,,\mathbf{x}) = X(\,\cdot\,,\mathbf{x})\circ F$. $\square$

Stratonovich Integration: When a probabilist looks at an operator

$$L = \tfrac{1}{2}\sum_{i,j}^{N} a_{i,j}(\mathbf{x})\partial_{x_i}\partial_{x_j} + \sum_{i=1}^{N} b_i(\mathbf{x})\partial_{x_i},$$

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 diffisive 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 $a(\mathbf{x}) = \begin{pmatrix} x_2^2 & -x_1x_2 \\ -x_1x_3 & x_1^2 \end{pmatrix}$ 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 $\sigma(\mathbf{x}) = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$, and so an Itô representation of the associated diffusion is

$$\begin{pmatrix} X_1(t,\mathbf{x}) \\ X_2(t,\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \int_0^t \begin{pmatrix} -X_2(\tau,\mathbf{x}) \\ X_1(\tau,\mathbf{x}) \end{pmatrix} dw(\tau),$$

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 $Z(t,z) = X_1(t,\mathbf{x}) + iX_2(t,\mathbf{x})$. Then

$$Z(t,z) = z + i\int_0^t Z(\tau,z)\,d\tau,$$

and so $Z(t,z) = ze^{iw(t)+\frac{t}{2}}$, or, equivalently,

$$X(t,\mathbf{x}) = e^{\frac{t}{2}}\begin{pmatrix} x_1\cos w(t) - x_2\sin w(t) \\ x_2\cos w(t) + x_1\sin w(t)\end{pmatrix}.$$

In particular, $|X(t,\mathbf{x})| = |\mathbf{x}|e^{\frac{t}{2}}$, which, if $\mathbf{x}\neq\mathbf{0}$, means that the distance of $X(t,\mathbf{x})$ from $\mathbf{0}$ is deterministic and growing exponentially fast.

What the preceding example reflects is that the interpretation of a and b as governing the diffusive and deterministic parts of L is naïve because it is too coordinate dependent. Indeed, if one represents the preceeding L in terms of polar coordinates, one finds that it is equal to

$$\tfrac{1}{2}\big(\partial_\theta^2 + r\partial_r\big),$$

which makes it clear that, although the angular coordinate of $X(t,\mathbf{x})$ 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 $a\geq\epsilon\mathbf{I}$, 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 $a\geq\epsilon\mathbf{I}$, 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 represent L in a form that looks the same in all coordinate systems, and such a representation was introduced by L. Hörmander. To describe this representation, for a vector field $V\in C\big(\mathbb{R}^N;\mathbb{R}^N\big)$, use $\mathcal{L}_V$ to denote the directional derivative $\sum_{j=1}^N V_j\partial_{x_h}$, and $a=\sigma\sigma^\top$, where $\sigma\in\mathrm{Hom}(\mathbb{R}^M;\mathbb{R}^N)$. For each $1\leq k\leq M$, let $V_k\in\mathbb{R}^N$ be the kth column of σ. Assuming that σ is continuously differentiable, one then has

$$L = \tfrac{1}{2}\sum_{k=1}^M \mathcal{L}_{V_k}^2 + \mathcal{L}_{V_0}, \tag{116}$$

when one takes $V_0 = b - \frac{1}{2}\sum_{k=1}^M \mathcal{L}_{V_k}V_k$. 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 diffiomorphism 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 $\mathcal{L}_{F_*V}\varphi = \big(\mathcal{L}_V(\varphi \circ F)\big) \circ F^{-1}$ for $\varphi \in C^1\big(F(G);\mathbb{R}\big)$. More explicitly,

$$F_*V = F^{(1)}V \circ F^{-1} \text{ where } F^{(1)} = \big((\partial_{x_j} F_i)\big)_{1 \le i,j \le N}$$

is the Jacobian matrix of F. Then

$$\mathcal{L}_V^2(\varphi \circ F) = \mathcal{L}_V\big(\mathcal{L}_{F_*V}\varphi\big) \circ F = \big(\mathcal{L}_{F_*V}^2\varphi\big) \circ F,$$

and in the coordinate system on $F(G)$ determined by F,

$$L = \tfrac{1}{2}\sum_{k=1}^M \mathcal{L}_{F_*V_k}^2 + \mathcal{L}_{F_*V_0}.$$

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 associated 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 $X(\,\cdot\,,\mathbf{x})$ is the solution to (56) when $\sigma = \big(V_1,\dots,V_M\big)$ and $b = V_0 + \frac{1}{2}\sum_{k=0}^M \mathcal{L}_{V_k}V_k$, then $F \circ X(\,\cdot\,,\mathbf{x})$ should be solution to (56) with σ and b replaced by

$$\big(F_*V_1,\dots,F_*V_M\big) \text{ and } F_*V_0 + \tfrac{1}{2}\sum_{k=0}^M \mathcal{L}_{F_*V_k}F_*V_k,$$

and, with sufficient patience, one can check that this is true. However, it would be helpful to have a formulism that made such computations easier and brought out the relationship between $X(\,\cdot\,,\mathbf{x})$ and integral curves of the V_k's. For that purpose, reconsider the equation

$$\dot{X}(t,\mathbf{x})(w) = \sum_{k=1}^M V_k\big(X(t,\mathbf{x})(w)\big)\dot{w}(t)_k + V_0\big(X(t,\mathbf{x})(w)\big) \text{ with } X(0,\mathbf{x}) = \mathbf{x},$$

where the interpretation now, unlike Itô's interpretation, is that $X(t,\mathbf{x})(w)$ is constructed by taking the limit of integral curves corresponding to mollifications of the paths w. For instance, in our example,

$$F_1(t,\mathbf{x}) = \begin{pmatrix} x_1\cos t - x_2 \sin t \\ x_2 \cos t + x_1 \sin t \end{pmatrix} \text{ and } F_0(t,\mathbf{x}) = e^{\frac{t}{2}}\mathbf{x}$$

are the integral curves of V_1 and V_0 starting at $\mathbf{x}$, and it is easy to check that, for any continuously differentiable w, $F_0\big(t, F_1(w(t),\mathbf{x})\big)$ is the integral curve of $\dot{w}(t)V_1 + V_0$ starting at $\mathbf{x}$. Thus, one might guess that $F_0\big(t, F_1(w(t),\mathbf{x})\big)$ is the diffusion associated with L, as indeed we saw that it is. Below we will see that 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 skepical about it until Itô rationalized it. Itô understood that there is 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 $\big(X(t),\mathcal{F}_t,\mathbb{P}\big)$ and $\big(Y(t),\mathcal{F}_t,\mathbb{P}\big)$, define 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 $m \ge 0$. Set $\Delta^X_{m,n} = X((m+1)2^{-n}) - X(m2^{-n})$ and $\Delta^Y_{m,n}(t) = Y((m+1)2^{-n}) - Y(m2^{-n})$. Then, by Corollary 113,

$$\begin{aligned}
&\int_0^t X_n(\tau)\,dY_n(\tau)\\
&\quad = \sum_{m<2^n t} X(m2^{-n})\Delta^Y_{m,n} + \tfrac12 \sum_{m<2^n t} 4^n\big(t\wedge(m+1)2^{-n} - m2^{-n}\big)\Delta^X_{m,n}\Delta^Y_{m,n}\\
&\quad \longrightarrow \int_0^t X(\tau)\,dY(\tau) + \frac12\langle X,Y\rangle(t).
\end{aligned}$$

Thus Itô said that the Stratonovich integral of $Y(\,\cdot\,)$ with respect to $X(\,\cdot\,)$ should be

$$\int_0^t Y(\tau)\,\bullet dX(\tau) \equiv \int_0^t Y(\tau)\,dX(\tau) + \tfrac12\langle X,Y\rangle(t).$$[16]

Notice, if one adopts this definition, then the equation in Corollary 112 becomes

$$\varphi\big(\vec{X}(t)\big) = \varphi\big(\vec{X}(0)\big) + \sum_{j=1}^N \int_0^t \partial_{x_j}\varphi\big(\vec{X}(\tau)\big)\,\bullet dX_j(\tau) \tag{117}$$

for $\varphi \in C^3(\mathbb{R}^N;\mathbb{R})$. Indeed, by Itô's formula, the martingale part of $\partial_{x_j}\varphi\big(\vec{X}(t)\big)$ is

$$\sum_{i=1}^N \int_0^t \partial_{x_i}\partial_{x_j}\varphi\big(\vec{X}(\tau)\big)\,dM_i(\tau),$$

[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

$$\sum_{j=1}^N \langle \partial_{x_j}\varphi\circ\vec{X}, X_i\rangle(dt) = \sum_{i,j=1}^N (\partial_{x_i}\partial_{x_j}\varphi)\big(\vec{X}(t)\big)\langle X_i, X_j\rangle(dt).$$

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

$$\begin{aligned} D_n &\equiv \int_0^1 \big(X(\tau) - X_n(\tau)\big)\,dY_n(\tau) \\ &= \sum_{m<2^n} 2^n \left(\int_{I_{m,n}} \Big(\big(X(\tau) - X_{m,n}\big) - 2^n(\tau - m2^{-n})\Delta^X_{m,n}\Big)\,d\tau\right)\Delta^Y_{m,n}, \end{aligned}$$

where $I_{m,n} \equiv [m2^{-n}, (m+1)2^{-n}]$ and, for any $F : [0,\infty) \longrightarrow \mathbb{R}$, $\Delta^F_{m,n} \equiv F_{m+1,n} - F_{m,n}$ with $F_{m,n} \equiv F(m2^{-n})$. Obviously

$$4^n \int_{I_{m,n}} (\tau - m2^{-n})\Delta^X_{m,n}\,d\tau = \tfrac{1}{2}\Delta^X_{m,n},$$

and, using integration by parts,

$$\int_{I_{m,n}} 2^n \int_{I_{m,n}} \big(X(\tau) - X_{m,n}\big)\,d\tau = \Delta^X_{m,n} - 2^n\int_{I_{m,n}} (\tau - m2^{-n})\,dX(\tau).$$

Hence, if

$$Z_n(t) \equiv \int_0^t \big(\tfrac{1}{2} - 2^n(\tau - [\tau]_n)\,dX(\tau),$$

then (cf. the proof of Corollary 113),

$$\begin{aligned} D_n &= \sum_{m<2^n} \Delta^{Z_n}\Delta^Y_{m,n} \\ &= \langle Z_n, Y\rangle(1) + \int_0^1 \big(Z_n(\tau) - Z_n([\tau]_n)\big)\,dY(\tau) + \int_0^1 \big(Y(\tau) - Y([\tau]_n)\big)\,dZ_n(\tau). \end{aligned}$$

It is simple to show that the stochastic integrals on the right tend to 0 as $n \to \infty$. To handle the term $\langle Z_n, Y\rangle(1)$, observe that it equals

$$\frac{1}{2}\int_0^1 c(\tau)\,d\tau - 2^n\int_0^1 (\tau - [\tau]_n)c(\tau)\,d\tau,$$

and, since

$$\begin{aligned} &2^n\int_0^1 (\tau - [\tau]_n)c(\tau)\,d\tau \\ &\quad = \tfrac{1}{2}2^{-n}\sum_{m<2^n} c(m2^{-n}) + 2^n\int_0^1 (\tau - [\tau]_n)\big(c(\tau) - c([\tau]_n)\big)\,d\tau \longrightarrow \tfrac{1}{2}\int_0^1 c(\tau)\,d\tau, \end{aligned}$$

it follows that $D_n \longrightarrow 0$.

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 $\partial_{x_j}\varphi\big(\vec{X}(t)\big)$ would be one. Thus, we have obtained a pleasing looking formula in which only first order derivatives of φ appear explicitly, but we have done so at the price of requiring that φ 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.

Stratonvich 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 $0 \le k \le M$, 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

$$X(t,\mathbf{x}) = \mathbf{x} + \int_0^t V_0\big(X(\tau,\mathbf{x})\big)\,d\tau + \sum_{k=1}^M V_k\big(X(\tau,\mathbf{x})\big) \bullet dw(\tau)_k. \tag{118}$$

Of course, this equation is equivalent to (56) when σ is the $N\times M$-matrix whose kth column, for $1 \le k \le M$ 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 in (116). Notice that using this formalism, it is easy to verify that, if F is a diffeomorphism, then $F\circ X(\,\cdot\,,\mathbf{x})$ is the solution to (118) with the V_k's replaced by F_*V_k's. Indeed, by (117),

$$F\big(X(t,\mathbf{x})\big) = F(\mathbf{x}) + \int_0^t \mathcal{L}_{V_0}F\big(X(\tau,\mathbf{x})\big)\,d\tau + \int_0^t \mathcal{L}_{V_k}F\big(X(\tau,\mathbf{x})\big) \bullet dw(\tau)_k,$$

and $\mathcal{L}_{V_k}F = (F_*V_k)\circ F$. Further evidence of the virtues of the Stratonovich formalism is provided by the computation in the following theorem.

Theorem 119. *Let $V_k^{(1)}$ denote the Jacobian matrix of V_k, and let $X(\,\cdot\,,\mathbf{x})$ be the solution to (118). Then $X(t,\,\cdot\,)$ is continuosly differentiable and it Jabobian matrix satisfies*

$$\begin{aligned} X^{(1)}(t,\mathbf{x}) = \mathbf{I} + &\int_0^t V_0^{(1)}\big(X(\tau,\mathbf{x})\big)X^{(1)}(\tau,\mathbf{x})\,d\tau \\ &+ \sum_{k=1}^M \int_0^t V_k^{(1)}\big(X(\tau,\mathbf{x})\big)X^{(1)}(\tau,\mathbf{x}) \bullet dw(\tau)_k. \end{aligned}$$

In addition

$$\det\bigl(X^{(1)}(t,\mathbf{x})\bigr) = \exp\left(\int_0^t \mathrm{div}V_0\bigl(X(\tau,\mathbf{x})\bigr)\,d\tau + \sum_{k=1}^M \int_0^t \mathrm{div}V_k\bigl(X(\tau,\mathbf{x})\bigr) \bullet dw(\tau)_k\right)$$

$$= \exp\biggl(\int_0^t \mathrm{div}V_0\bigl(X(\tau,\mathbf{x})\bigr)\,d\tau + \sum_{k=1}^M \int_0^t \mathrm{div}V_k\bigl(X(\tau,\mathbf{x})\bigr)\,dw(\tau)_k$$
$$+ \tfrac{1}{2}\sum_{k=1}^M \int_0^t \bigl(\mathcal{L}_{V_k}\mathrm{div}V_k\bigr)\bigl(X(\tau,\mathbf{x})\bigr)\,d\tau\biggr).$$

In particular, if $X(\,\cdot\,,\mathbf{x})$ *is the solution to* (56), V_k *is the* k*th column of* σ, *and* $V_0 = b - \frac{1}{2}\sum_{k=1}^M \mathcal{L}_{V_k}V_k$, *then*

$$\det\bigl(X^{(1)}(t,\mathbf{x})\bigr) = \exp\biggl(\int_0^t \mathrm{div}V_0\bigl(X(\tau,\mathbf{x})\bigr)\,d\tau + \int_0^t \mathrm{div}V_k\bigl(X(\tau,\mathbf{x})\bigr)\,dw(\tau)_k$$
$$- \tfrac{1}{2}\int_0^t \mathrm{Trace}\Bigl(\bigl(V_k^{(1)}(V_k^{(1)})^\top\bigr)\bigl(X(\tau,\mathbf{x})\bigr)\Bigr)\,d\tau\biggr).$$

Proof: The facts that $X(t,\,\cdot\,)$ is continuously differentiable and that $X^{(1)}(\,\cdot\,,\mathbf{x})$ satisfies the asserted equation are easily checked by converting (118) to its equivalent Itô form, applying (69), and reconverting it to Stratonovich form. Furthermore, once one knows the first equality, the other ones follow when one uses Itô's prescription for converting Statonovich integrals to Itô ones.

Given an $N\times N$ matrix $A = \bigl((a_{i,j})\bigr)_{1\le i,j\le N}$, let $C(A)$ be the matrix whose (i,j)th entry is $(-1)^{i+j}$ times the determinant of the $(N-1)\times(N-1)$ matrix obtained by deleting the ith row and jth column of A. Using Cramer's rule, it is easy to check that $\partial_{a_{i,j}}\det(A) = C(A)_{i,j}$ and that $AC(A)^\top = \det(A)\mathbf{I}$. Now set $D(t,\mathbf{x}) = \det\bigl(X^{(1)}(t,\mathbf{x})\bigr)$. Using the preceding and applying (117), one sees that

$$D(t,\mathbf{x}) = 1+\int_0^t \mathrm{div}V_0\bigl(X(\tau,\mathbf{x})\bigr)D(\tau,\mathbf{x})\,d\tau$$
$$+\sum_{m=1}^M \int_0^t \mathrm{div}V_m\bigl(X(\tau,\mathbf{x})\bigr)D(\tau,\mathbf{x}) \bullet dw(\tau)_k.$$

Since

$$\exp\left(\int_0^t \mathrm{div}V_0\bigl(X(\tau,\mathbf{x})\bigr)\,d\tau + \sum_{k=1}^M \int_0^t \mathrm{div}V_k\bigl(X(\tau,\mathbf{x})\bigr) \bullet dw(\tau)_k\right)$$

is the one and only solution to this equation, the proof is complete. $\square$

Of course, one could have carried out the preceding computation without using the Statonovich formalism, but it would have been far more tedious to do so. However, such computations are not the only virtue of the expression in (118). Indeed, although we know how to solve (118) by converting it to an Itô stochastic integral equation, that is not the way the only way think about solving it. Instead, for $\boldsymbol{\xi} = (\xi_0, \dots, \xi_M) \in \mathbb{R} \times \mathbb{R}^M$, set $V_{\boldsymbol{\xi}} = \sum_{k=0}^M \xi_k V_k$, and consider the the ordinary differential equation

$$\dot{F}_{\boldsymbol{\xi}}(t, \mathbf{x}) = V_{\boldsymbol{\xi}}\big(F_{\boldsymbol{\xi}}(t, \mathbf{x})\big) \text{ with } F_{\boldsymbol{\xi}}(\mathbf{0}, \mathbf{x}) = \mathbf{x}.$$

In other words, $t \rightsquigarrow F_{\boldsymbol{\xi}}(t, \mathbf{x})$ is the integral curve of $V_{\boldsymbol{\xi}}$ that passes through $\mathbf{x}$ at time 0. Next define $E : (\mathbb{R} \times \mathbb{R}^M) \times \mathbb{R}^N \longrightarrow \mathbb{R}^N$ by $E(\boldsymbol{\xi}, \mathbf{x}) = F_{\boldsymbol{\xi}}(1, \mathbf{x})$. As is easy to check, $E(t\boldsymbol{\xi}, \mathbf{x}) = F_{\boldsymbol{\xi}}(t, \mathbf{x})$ for all $t \in \mathbb{R}$. Furthermore, standard results from the theory of ordinary differential equations show that $E(\boldsymbol{\xi}, \cdot\,)$ is a diffeomorphism from $\mathbb{R}^N$ onto itself, $E(-\boldsymbol{\xi}, \cdot\,)$ is the inverse of $E(\boldsymbol{\xi}, \cdot\,)$, and there exist constants $C < \infty$ and $\nu \in [0, \infty)$ such that

$$\begin{aligned} &|\partial_{\xi_k} E(\boldsymbol{\xi}, \mathbf{x})| \vee |\partial_{x_j} E(\boldsymbol{\xi}, \mathbf{x})| \vee |\partial_{\xi_k} \partial_{\xi_\ell} E(\boldsymbol{\xi}, \mathbf{x})| \\ &\qquad \vee |\partial_{\xi_k} \partial_{x_j} E(\boldsymbol{\xi}, \mathbf{x})| \vee |\partial_{x_i} \partial_{x_j} E(\boldsymbol{\xi}, \mathbf{x})| \le C(1 + |\mathbf{x}|) e^{\nu |\boldsymbol{\xi}|} \end{aligned} \tag{120}$$

for all $1 \le i, j \le M$ and $0 \le k, \ell \le M$.

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 121. *Assume that $V, W \in C^1(\mathbb{R}^N; \mathbb{R}^N)$ have bounded first order derivatives, and determine $t \rightsquigarrow F(t, \mathbf{x})$ and $t \rightsquigarrow G(t, \mathbf{x})$ by*

$$\dot{F}(t, \mathbf{x}) = V\big(F(t, \mathbf{x})\big) \text{ and } \dot{G}(t, \mathbf{x}) = V\big(G(t, \mathbf{x})\big) \text{ with } F(0, \mathbf{x}) = G(0, \mathbf{x}) = \mathbf{x}.$$

If $[V, W] = 0$, then $F\big(s, G(t, \mathbf{x})\big) = G\big(t, F(x, \mathbf{x})\big)$ for all $(s, t) \in \mathbb{R}^2$.

Proof: Set $\Phi(t) = F\big(s, G(t, \mathbf{x})\big)$. Then, for $\varphi \in C^2(\mathbb{R}^N; \mathbb{R})$,

$$\partial_t \varphi \circ \Phi(t) = \big(\mathcal{L}_{F(s,\,\cdot\,)_* W} \varphi\big)\big(G(t, x)\big).$$

If $\varphi_s = \varphi \circ F(s, \cdot\,)$, then, because $F(s + h, \cdot\,) = F\big(h, F(s, \cdot\,)\big)$,

$$\begin{aligned} \partial_s \Big(\big(\mathcal{L}_{F(s,\,\cdot\,)_* W} \varphi\big)\big(F(-s, \mathbf{y})\big) \Big) &= \partial_h \Big(\mathcal{L}_{F(h,\,\cdot\,)_* W} \varphi_s \big) \big(F(-s - h, \mathbf{y})\big) \Big) \Big|_{h=0} \\ &= \big(\mathcal{L}_W \circ \mathcal{L}_V - \mathcal{L}_V \circ \mathcal{L}_W\big) \varphi_s(\mathbf{y}) = 0. \end{aligned}$$

Hence $\big(\mathcal{L}_{F(s,\,\cdot\,)_* W} \varphi\big)\big(F(-s, \mathbf{y})\big) = \mathcal{L}_W \varphi(\mathbf{y})$, which means that $\big(\mathcal{L}_{F(s,\,\cdot\,)_* W} \varphi\big)(\mathbf{y}) = \mathcal{L}_W \varphi\big(F(s, \mathbf{y})\big)$ and therefore that $\partial_t \varphi \circ \Phi(t) = \mathcal{L}_W \varphi\big(\Phi(t)\big)$. Equivalently, $\dot{\Phi}(t) = W\big(\Phi(t)\big)$, and so, since $\Phi(0) = F(s, \mathbf{x})$,

$$F\big(s, G(t, \mathbf{x})\big) = \Phi(t) = G\big(t, F(s, x)\big). \quad \square$$

THEOREM 122. *Let $X(\,\cdot\,,\mathbf{x})$ be the solution to (118). If the V_k's commute, then $X(t,\mathbf{x})(w) = E\big((t,w(t)),\mathbf{x}\big)$.*

PROOF: By (117),

$$E\big((t,w(t)),\mathbf{x}\big) = x + \int_0^t \partial_{\xi_0} E\big((\tau,w(\tau),\mathbf{x}\big)\,d\tau + \sum_{k=1}^N \partial_{\xi_k} E\big((\tau,w(\tau)),\mathbf{x}\big) \bullet dw(\tau)_k,$$

and so all that remains is to prove that $\partial_{\xi_k} E(\boldsymbol{\xi},\mathbf{y}) = V_k\big(E(\boldsymbol{\xi},\mathbf{y})\big)$. To this end, apply Lemma 121 to see that $E\big(\boldsymbol{\xi},E(\boldsymbol{\eta},\mathbf{y})\big) = E\big(\boldsymbol{\eta},E(\boldsymbol{\xi},\mathbf{y})\big)$, and then apply this to check that if $\Phi(t) = E\big(t\boldsymbol{\xi},E(t\boldsymbol{\eta},\mathbf{y})\big)$, then $\dot\Phi(t) = V_{\boldsymbol{\xi}+\boldsymbol{\eta}}\big(\Phi(t)\big)$ with $\Phi(0) = \mathbf{x}$. Hence $E(\boldsymbol{\xi}+\boldsymbol{\eta},\mathbf{y}) = E\big(\boldsymbol{\eta},E(\boldsymbol{\xi},\mathbf{y})\big)$. Taking $\boldsymbol{\eta} = t\mathbf{e}_k$, where $(\mathbf{e}_k)_\ell = \delta_{k,\ell}$, 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 derivatives as a function of w as the V_k's have as a functions of $\mathbf{x}$. 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 communte and is absent when they don't is $\partial_{\xi_k} E(\boldsymbol{\xi},\mathbf{y}) = V_k\big(E(\boldsymbol{\xi},\mathbf{y})\big)$, which is equivalent to the property $E(\boldsymbol{\xi}+\boldsymbol{\eta},\mathbf{y}) = E\big(\boldsymbol{\eta},E(\boldsymbol{\xi},\mathbf{y})\big)$. To build this property into our construction even when the V_k's don't commute, for $n \ge 0$, determine $X_n(\,\cdot\,,\mathbf{x})$ by $X_n(0,\mathbf{x}) = \mathbf{x}$ and

$$X_n(t,\mathbf{x}) = E\big(\Delta_n(t),X_n([t]_n,\mathbf{x})\big)$$
$$\text{where } \Delta_n(t)(w) = \big(t - m2^{-n}, w(t) - w(m2^{-n})\big) \text{ for } m2^{-n} < t \le (m+1)2^{-n}.$$

Then, by (117),

$$\begin{aligned} X_n(t,\mathbf{x}) = \mathbf{x} + &\int_0^t E_0\big(\Delta_n(\tau),X_n([\tau]_n,\mathbf{x})\big)\,d\tau + \\ &\sum_{k=1}^M \int_0^t E_k\big(\Delta_n(\tau),X_n([\tau]_n,\mathbf{x})\big) \bullet dw(\tau)_k, \end{aligned} \tag{123}$$

where $E_k(\boldsymbol{\xi},\mathbf{y}) \equiv \partial_{\xi_k} E(\boldsymbol{\xi},\mathbf{y})$. After converting the Stratonovich integrals to Itô ones, applying Doob's inequality, and using the estimates in (120), one sees that

$$\mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x})\|^2_{[0,t]}\big] \le 2|\mathbf{x}|^2 + C(1+t)\int_0^t \mathbb{E}^{\mathcal{W}}\Big[e^{2\nu\Delta_n(\tau)}\Big(1+\big|X_n([\tau]_n,\mathbf{x})\big|^2\Big)\Big]\,d\tau$$

for some $C < \infty$. Further, since $\Delta_n(t)$ is independent of $X_n([\tau]_n, \mathbf{x})$ and $\sup_{n\ge 0}\sup_{t\ge 0}\mathbb{E}^{\mathcal{W}}\big[e^{2\nu\Delta_n(t)}\big] < \infty$, there is a $C < \infty$ such that

$$\mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big] \le 2|\mathbf{x}|^2 + C(1+t)\int_0^t \mathbb{E}^{\mathcal{W}}\Big[e^{2\nu\Delta_n(\tau)}\Big(1+\big\|X_n([\,\cdot\,,\mathbf{x})\big\|_{[0,\tau]}^2\Big)\Big]\,d\tau,$$

and so, by Gromwall's lemma,

$$\sup_{n\ge 0}\mathbb{E}^{\mathcal{W}}\big[1+\|X_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big] \le Ke^{Kt}(1+|\mathbf{x}|^2) \tag{124}$$

for some $K < \infty$.

Next let $X(\,\cdot\,,\mathbf{x})$ be the solution to (118), and set $W_k(\boldsymbol{\xi},\mathbf{x}) = E_k(\boldsymbol{\xi},\mathbf{x}) - V_k\big(E(\boldsymbol{\xi},\mathbf{x})\big)$. Then

$$\begin{aligned} X_n(t,\mathbf{x}) - X(t,\mathbf{x}) = &\int_0^t \Big(V_0\big(X_n(\tau,\mathbf{x})\big) - V_0\big(X(\tau,\mathbf{x})\big)\Big)\,d\tau \\ &+ \sum_{k=1}^M \int_0^t \Big(V_k\big(X_n(\tau,\mathbf{x})\big) - V_k\big(X(\tau,\mathbf{x})\big)\Big) \bullet dw(\tau)_k + \mathcal{E}_n(t,\mathbf{x}), \end{aligned}$$

where

$$\begin{aligned} \mathcal{E}_n(t,\mathbf{x}) \equiv &\int_0^t W_0\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big)\,d\tau \\ &+ \sum_{k=1}^N \int_0^t W_k\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big) \bullet dw(\tau)_k. \end{aligned}$$

After converting to Itô form the integrals with integrands involving the V_k's, one sees that

$$\begin{aligned} &\mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x}) - X(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big] \\ &\quad\le C(1+t)\int_0^t \mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x}) - X(\,\cdot\,,\mathbf{x})\|_{[0,\tau]}^2\,d\tau + 2\mathbb{E}^{\mathcal{W}}\big[\|\mathcal{E}_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big] \end{aligned} \tag{*}$$

for some $C < \infty$. In order to estimate $\mathcal{E}_n(t,\mathbf{x})$, we will need the following lemma.

Lemma 125. *There is a $C < \infty$ such that*

$$|W_k(\boldsymbol{\xi},\mathbf{x})| \vee |\partial_{\xi_k} W_k(\boldsymbol{\xi},\mathbf{x})| \le C|\boldsymbol{\xi}|e^{2\nu|\boldsymbol{\xi}|}(1+|\mathbf{x}|)$$

for all $0 \le k \le M$

PROOF: By Taylor's theorem,

$$\begin{aligned}E\big(\boldsymbol{\eta}, E(\boldsymbol{\xi},\mathbf{x})\big) &= E(\boldsymbol{\xi},\mathbf{x}) + V_{\boldsymbol{\eta}}\big(E(\boldsymbol{\xi},\mathbf{x})\big) + \tfrac{1}{2}\mathcal{L}_{V_\eta}V_{\boldsymbol{\eta}}\big(E(\boldsymbol{\xi},\mathbf{x})\big) + R_0(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})\\ &= \mathbf{x} + V_{\boldsymbol{\xi}}(\mathbf{x}) + \tfrac{1}{2}\mathcal{L}_{V_{\boldsymbol{\xi}}}V_{\boldsymbol{\xi}}(\mathbf{x}) + V_{\boldsymbol{\eta}}(\mathbf{x}) + \mathcal{L}_{V_{\boldsymbol{\xi}}}V_{\boldsymbol{\eta}}(\mathbf{x}) + \tfrac{1}{2}\mathcal{L}_{V_\eta}V_{\boldsymbol{\eta}}(\mathbf{x})\\ &\qquad + R_0(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) + R_1(\boldsymbol{\xi},\mathbf{x}) + R_2(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) + R_3(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}),\end{aligned}$$

where

$$\begin{aligned}R_0(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) &= \int_0^1 (1-t)\Big(\mathcal{L}_{V_\eta}V_{\boldsymbol{\eta}}\big(E(t\boldsymbol{\eta}, E(\boldsymbol{\xi},\mathbf{x}))\big) - \mathcal{L}_{V_\eta}V_{\boldsymbol{\eta}}\big(E(\boldsymbol{\xi},\mathbf{x})\big)\Big)\,dt\\ R_1(\boldsymbol{\xi},\mathbf{x}) &= \int_0^1 (1-t)\Big(\mathcal{L}_{\boldsymbol{\xi}}V_{\boldsymbol{\xi}}\big(E(t\boldsymbol{\xi},\mathbf{x})\big) - \mathcal{L}_{\boldsymbol{\xi}}V_{\boldsymbol{\xi}}(\mathbf{x})\Big)\,dt\\ R_2(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) &= \int_0^1 \Big(\mathcal{L}_{V_{\boldsymbol{\xi}}}V_{\boldsymbol{\eta}}\big(E(t\boldsymbol{\xi},\mathbf{x})\big) - \mathcal{L}_{V_{\boldsymbol{\xi}}}V_{\boldsymbol{\eta}}(\mathbf{x})\Big)\,dt\\ R_3(\mathbf{x},\boldsymbol{\eta},\mathbf{x}) &= \frac{\mathcal{L}_{V_\eta}V_{\boldsymbol{\eta}}\big(E(\boldsymbol{\xi},\mathbf{x})\big) - \mathcal{L}_{V_\eta}V_{\boldsymbol{\eta}}(\mathbf{x})}{2}.\end{aligned}$$

At the same time,

$$E(\boldsymbol{\xi}+\boldsymbol{\eta},\mathbf{x}) = \mathbf{x} + V_{\boldsymbol{\xi}+\boldsymbol{\eta}}(\mathbf{x}) + \tfrac{1}{2}\mathcal{L}_{V_{\boldsymbol{\xi}+\eta}}V_{\boldsymbol{\xi}+\boldsymbol{\eta}}(x) + R_1(\boldsymbol{\xi}+\boldsymbol{\eta},\mathbf{x}).$$

Hence, $E\big(\boldsymbol{\eta}, E(\boldsymbol{\xi},\mathbf{x})\big) - E(\boldsymbol{\xi}+\boldsymbol{\eta},\mathbf{x}) = \frac{1}{2}[V_{\boldsymbol{\xi}}, V_{\boldsymbol{\eta}}](\mathbf{x}) + R(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x})$, where

$$R(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) = R_0(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) + R_1(\boldsymbol{\xi},\mathbf{x}) + R_2(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) + R_3(\boldsymbol{\xi},\boldsymbol{\eta},\mathbf{x}) - R_1(\boldsymbol{\xi}+\boldsymbol{\eta},\mathbf{x}).$$

Since $W_k(\boldsymbol{\xi},\mathbf{x}) = \frac{1}{2}[V_{\boldsymbol{\xi}}, V_k] + \partial_{\eta_k}R(\boldsymbol{\xi},\mathbf{0},\mathbf{x})$ and

$$\begin{aligned}\partial_{\eta_k}R(\boldsymbol{\xi},\mathbf{0},\mathbf{x}) &= \int_0^1 \Big(\mathcal{L}_{V_{\boldsymbol{\xi}}}V_k\big(E(\boldsymbol{\xi},\mathbf{x})\big) - \mathcal{L}_{V_{\boldsymbol{\xi}}}V_k(\mathbf{x})\Big)\,d\tau\\ &\quad + \int_0^1 (1-t)\Big(\big(\mathcal{L}_{V_{\boldsymbol{\xi}}}V_k + \mathcal{L}_{V_k}V_{\boldsymbol{\xi}}\big)\big(E(t\boldsymbol{\xi},\mathbf{x})\big)\\ &\qquad\qquad - \big(\mathcal{L}_{V_{\boldsymbol{\xi}}}V_k + \mathcal{L}_{V_k}V_{\boldsymbol{\xi}}\big)(\mathbf{x}) + t\mathcal{L}_{V_{\boldsymbol{\xi}}}V_{\boldsymbol{\xi}}\partial_{\xi_k}E(\boldsymbol{\xi},\mathbf{x})\Big)\,dt,\end{aligned}$$

the estimates in (120) now implies the asserted estimate for $W_k(\boldsymbol{\xi},\mathbf{x})$. To prove the one for $\partial_{\xi_k}W_k(\boldsymbol{\xi},\mathbf{x})$, observe that, since $[V_k, V_k] = 0$,

$$[V_{\boldsymbol{\xi}}, V_k] = \sum_{\ell\neq k}\xi_\ell[V_\ell, V_k],$$

and therefore that $\partial_{\xi_k}W_k(\boldsymbol{\xi},\mathbf{x}) = \partial_{\xi_k}\partial_{\eta_k}R(\boldsymbol{\xi},\mathbf{0},\mathbf{x})$. Hence, another application of (120) give the asserted result. □

Returning to the estimate of $\mathcal{E}_n(t,\mathbf{x})$, set

$$\tilde{W}(\boldsymbol{\xi},\mathbf{x}) = W_0(\boldsymbol{\xi},\mathbf{x}) + \tfrac{1}{2}\sum_{k=1}^{M}\partial_{\xi_k}W_k(\boldsymbol{\xi},\mathbf{x}).$$

Then

$$\mathcal{E}_n(t,\mathbf{x}) = \int_0^t \tilde{W}\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big)\,d\tau + \sum_{k=1}^{M}\int_0^t W_k\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big)\,dw(\tau)_k.$$

Thus

$$\begin{aligned}\mathbb{E}^{\mathcal{W}}\big[\|\mathcal{E}_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big] \leq{}& 2t\int_0^t \mathbb{E}^{\mathcal{W}}\big[\big|\tilde{W}\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big)|^2\big]\,d\tau \\ &+ 8\sum_{m=1}^{M}\int_0^t \mathbb{E}^{\mathcal{W}}\big[\big|W_k\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big)\big|^2\big]\,d\tau.\end{aligned}$$

By Lemma 125 and the independence of $\Delta_n(\tau)$ and $X_n([\tau]_n,\mathbf{x})$,

$$\begin{gathered}\mathbb{E}^{\mathcal{W}}\big[\big|\tilde{W}\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big)|^2\big] \text{ and} \\ \sum_{m=1}^{M}\int_0^t \mathbb{E}^{\mathcal{W}}\big[\big|W_k\big(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\big)\big|^2\big]\end{gathered}$$

are dominated by a constant times

$$\mathbb{E}^{\mathcal{W}}\big[|\Delta_n(\tau)|^2 e^{4\nu|\Delta_n(\tau)|}\big]\mathbb{E}^{\mathcal{W}}\big[1+|X_n([\tau]_n,\mathbf{x})|^2\big].$$

Thus, since $\mathbb{E}^{\mathcal{W}}\big[|\Delta_n(\tau)|^2 e^{4\nu|\Delta_n(\tau)|}\big]$ is dominated by a constant times 2^{-n}, we can now use (124) to see that there is a $C(t)<\infty$ such that

$$\mathbb{E}^{\mathcal{W}}\big[\|\mathcal{E}_n(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big] \leq 2^{-n}K(t)(1+|\mathbf{x}|^2).$$

After combinining this with (*) and applying Gromwall's lemma, we conclude that

$$\mathbb{E}^{\mathcal{W}}\big[\|X_n(\,\cdot\,,\mathbf{x}) - X(\,\cdot\,,\mathbf{x})\|_{[0,t]}^2\big] \leq C(t)(1+|\mathbf{x}|)2^{-n} \tag{126}$$

for some $C(t)<\infty$. In particular, this means that $\|X_n(\,\cdot\,,\mathbf{x}) - X(\,\cdot\,,\mathbf{x})\|_{[0,t]} \longrightarrow 0$ (a.s.,$\mathcal{W}$) and in $L^2(\mathcal{W};\mathbb{R})$.

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. *For each $w \in C\big([0,\infty);\mathbb{R}^M\big)$ and $n \geq 0$, let w_n be the polygonal path that equals w at times $m2^{-n}$ and is linear on $[m2^{-n},(m+1)2^{-n}]$, and let $\tilde{X}_n(\,\cdot\,,\mathbf{x},w)$ be the integral curve of the time dependent vector filed $t \rightsquigarrow V_0 + \sum_{k=1}^M \dot{w}_n(t)V_k$ starting at $\mathbf{x}$. Then, for each $t>0$ and $\epsilon>0$,*

$$\mathcal{W}\big(\|\tilde{X}_n(\,\cdot\,,\mathbf{x}) - X(\,\cdot\,,\mathbf{x})\|_{[0,t]} \geq \epsilon\big) = 0.$$

PROOF: For the present assume that the V_k's are bounded by some constant $C<\infty$.

Observe that

$$\tilde{X}_n\big(t,\mathbf{x}) = E\big(2^n(t-m2^{-n})\Delta_n((m+1)2^{-n}), \tilde{X}_n(m2^{-n},\mathbf{x})\big)$$

for $t \in [m2^{-n},(m+1)2^{-n}]$, and so, by induction on $m \geq 0$, $\tilde{X}_n(m2^{-n},\mathbf{x}) = X_n(m2^{-n},\mathbf{x})$ for all $m \geq 0$. At the same time,

$$\begin{aligned}|X_n(t,\mathbf{x})(w) - X_n(m2^{-n},\mathbf{x})(w)| &\leq C|w(t) - w(m2^{-n})|\\ |\tilde{X}_n(t,\mathbf{x}) - \tilde{X}_n(t,\mathbf{x})(w)| &\leq C|w((m+1)2^{-n}) - w(m2^{-n})|.\end{aligned}$$

Hence $\|X_n(\,\cdot\,,\mathbf{x}) - \tilde{X}_n(\,\cdot\,,\mathbf{x})\|_{[0,t]} \longrightarrow 0$ for each $t \geq 0$, and therefore, by (126), the result is proved when the V_k's are bounded.

To remove the boundedness assumption on the V_k's, for each $R>0$, choose $\eta_R \in C_c^\infty\big(\mathbb{R}^N;[0,1]\big)$ so that $\eta_R = 1$ on $\overline{B(\mathbf{0},R+1)}$, set $V_k^{(R)}(\mathbf{y}) = \eta_R(\mathbf{y})V_k(\mathbf{y})$, and define $S^{(R)}(\mathbf{x})$, $X^{(R)}(\,\cdot\,,\mathbf{x})$, and $X_n^{(R)}(\,\cdot\,,\mathbf{x})$ accordingly. If

$$\zeta_R = \inf\{t \geq 0 : |X(t,\mathbf{x})| \geq R\},$$

then, by Lemma 38, $X(\,\cdot\,,\mathbf{x}) = X^{(R)}(\,\cdot\,,\mathbf{x})$ on $[0,\zeta_R)$. Hence, if $t<\zeta_R$ and $\|X^{(R)}(\,\cdot\,,\mathbf{x}) - X_n^{(R)}(\,\cdot\,,\mathbf{x})\|_{[0,t]} < \epsilon < 1$, then $\|X(\,\cdot\,,\mathbf{x}) - X_n(\,\cdot\,,\mathbf{x})\|_{[0,t]} < \epsilon$, and therefore

$$\begin{aligned}\mathcal{W}\big(\|\tilde{X}_n(\,\cdot\,,\mathbf{x}) - X(\,\cdot\,,\mathbf{x})\|_{[0,t]} < \epsilon\big) &\geq \mathcal{W}\big(\|\tilde{X}_n^{(R)}(\,\cdot\,,\mathbf{x}) - X^{(R)}(\,\cdot\,,\mathbf{x})\|_{[0,t\wedge\zeta_R]} < \epsilon\big)\\ &\geq \mathcal{W}\big(\|\tilde{X}_n^{(R)}(\,\cdot\,,\mathbf{x}) - X^{(R)}(\,\cdot\,,\mathbf{x})\|_{[0,t]} < \epsilon\big) - \mathcal{W}(\zeta_R \leq t).\end{aligned}$$

By first letting $n \to \infty$ and then $R \to \infty$, one arrives at the desired conclusion. □

COROLLARY 128. *Define $S(\mathbf{x})$ to be the closure in $C\big([0,\infty);\mathbb{R}^N\big)$ of the set of continuously differentiable paths $p:[0,\infty) \longrightarrow \mathbb{R}^N$ for which there exists a smooth function $\alpha \in C\big([0,\infty);\mathbb{R}^N\big)$ such that*

$$\dot{p}(t) = V_0\big(p(t)\big) + \sum_{k=1}^M \alpha(t)_k V_k\big(p(t)\big) \text{ with } p(0) = \mathbf{x}. \tag{129}$$

If $\mathbb{P}_{\mathbf{x}}$ is the solution to the martingale problem for the L in (116) starting at $\mathbf{x}$, then $\mathbb{P}_{\mathbf{x}}\big(S(\mathbf{x})\big)=1$. In particular, if $M\subseteq\mathbb{R}^N$ is a closed, differentiable manifold and $V_k(\mathbf{y})$ is tangent to M for all $0\le k\le M$ and $\mathbf{y}\in M$, then

$$\mathbb{P}_{\mathbf{x}}\big(\psi(t)\in M \text{ for all } t\in[0,\infty)\big)=1.$$

PROOF: Since, if $\mathbf{x}\in M$, every $p\in S(\mathbf{x})$ stays on M, the last assertion follows from the first. Next observe that for any piecewise constant $\alpha:[0,\infty)\longrightarrow\mathbb{R}^N$, the solution to (129) is the limit in $C\big([0,\infty);\mathbb{R}^N\big)$ of solutions to (129) for smooth α's. Thus, $S(\mathbf{x})$ contains all limits in $C\big([0,\infty);\mathbb{R}^N\big)$ of solutions to (129) with piecewise constant α's. In particular, if $\tilde{X}_n(\,\cdot\,,\mathbf{x})$ is defined as in the preceding theorem, then $\tilde{X}_n(\,\cdot\,,\mathbf{x})(w)\in S(\mathbf{x})$ for all $n\ge 0$ and w, and so, by that theorem, $\mathbb{P}_{\mathbf{x}}\big(S(\mathbf{x})\big)=\mathcal{W}\big(X(\,\cdot\,,\mathbf{x})\in S(\mathbf{x})\big)=1$. $\square$

The result in Corollary 128 is the easier half of the *support theorem* which says that $S(\mathbf{x})$ is the smallest closed set of $C\big([0,\infty);\mathbb{R}^N\big)$ to which $\mathbb{P}_{\mathbf{x}}$ gives measure 1. That is, not only does $\mathbb{P}_{\mathbf{x}}\big(S(\mathbf{x})\big)=1$ but also, for each $p\in S(\mathbf{x})$, $\epsilon>0$, and $t>0$,

$$\mathbb{P}_{\mathbf{x}}\big(\|\psi(\,\cdot\,)-p(\,\cdot\,)\|_{[0,t]}<\epsilon\big)>0.$$

In a different direction, one can apply these ideas to show that $X(t,\,\cdot\,)$ is a diffeomorphism of $\mathbb{R}^N$ onto itself. We already know that $X(t,\,\cdot\,)$ is continuously differentiable and, by Theorem 119, that its Jacobian matrix is non-degenerate. Thus, what remains is to show that it is has a globally defined inverse, and (126) provides a way to do so. Namely, for each $\boldsymbol{\xi}$, $E(\boldsymbol{\xi},\,\cdot\,)$ is a diffeomorphism whose inverse is $E(-\boldsymbol{\xi},\,\cdot\,)$, which means that, for $t\in[m2^{-n},(m+1)2^{-n}]$, $E\big(-\Delta_n(t)(w),X_n(t,\mathbf{x})(w)\big) = X_n(m2^{-n},\mathbf{x})(w)$. From this, it is clear that $X_n(t,\,\cdot\,)$ is a diffeomorphism of $\mathbb{R}^N$ onto itself for each $t\ge 0$. In fact, if $\check{E}(\boldsymbol{\xi},\mathbf{x})$ is given by the same prescription as $E(\boldsymbol{\xi},\mathbf{x})$ when each V_k is replaced by $-V_k$ and $\check{X}_n(\,\cdot\,,\mathbf{x})$ is determined by $\check{X}_n(0,\mathbf{x})=\mathbf{0}$ and $\check{X}_n(t,\mathbf{x})=\check{E}\big(\Delta_n(t),\check{X}_n([t]_n,\mathbf{x})\big)$ for $t>0$, then

$$\big(X_n(T,\,\cdot\,)(w)\big)^{-1}=\check{X}_n\big(t,\,\cdot\,\big)(\check{w}^t) \text{ where } \check{w}^t(\tau)\equiv w(t\vee\tau)-w\big((t-\tau)^+\big).$$

Now let $\check{X}(\,\cdot\,,\mathbf{x})$ be the solution to (118) with each V_k replaced by $-V_k$. Using the independent increment characterization of Brownian motion, one sees that the distribution of $\check{w}^t$ under $\mathcal{W}$ is the same of that of w, and so, if we define $\check{X}^t(\,\cdot\,,\mathbf{x})(w)=\check{X}(\,\cdot\,,\mathbf{x})(\check{w}^t)$ and $\check{X}^t_n(\,\cdot\,,\mathbf{x})(w)=\check{X}_n(\,\cdot\,,\mathbf{x})(\check{w}^t)$, then

$$\mathbb{E}^{\mathcal{W}}\big[\|\check{X}^t_n(\,\cdot\,,\mathbf{x})-\check{X}^t(\,\cdot\,,\mathbf{x})\|^2_{[0,t]}\big]\le C(t)(1+|\mathbf{x}|)2^{-n}.$$

In particular, this means that $X_n(t,\,\cdot\,)^{-1}(\mathbf{x})\longrightarrow\check{X}^t(t,\mathbf{x})$ (a.s.,$\mathcal{W}$) for each $t>0$ and $\mathbf{x}\in\mathbb{R}^N$. Hence, if we show that, $\mathcal{W}$-almost surely,

$$X_n(t,\,\cdot\,)\longrightarrow X(t,\,\cdot\,) \text{ and } \check{X}_n(t,\,\cdot\,)\longrightarrow\check{X}(t,\,\cdot\,)$$

uniformly on compact subsets, then we will know that $X(t,\,\cdot\,)^{-1}$ exists and is equal to $\check{X}^t(t,\,\cdot\,)$.

The proof of these convergence results is essentially the same as the proof of the corresponding sort of result in the Itô context. That is, one sets $D_n(t,\mathbf{x}) = X(t,\mathbf{x}) - X_n(t,\mathbf{x})$, estimates the moments of $D_n(t,\mathbf{y}) - D_n(t,\mathbf{x})$ in terms of $|\mathbf{y}-\mathbf{x}|$, and then applies Kolmogorov's convergence criterion, and, of course, the same procedure works equally well for $\check{D}_n(t,\mathbf{x}) = \check{X}(t,\mathbf{x}) - \check{X}_n(t,\mathbf{x})$. Thus, for $p \ge 2$, note that

$$\begin{aligned}|D_n(t,\mathbf{y}) - D_n(t,\mathbf{x})|^{p+1} \le \bigl(|X(t,\mathbf{y}) - X(t,\mathbf{x})| + |X_n(t,\mathbf{y}) - X_n(t,\mathbf{x})|\bigr)^p \\ \times\bigl(|D_n(t,\mathbf{x})| + |D_n(t,\mathbf{y})|\bigr),\end{aligned}$$

and therefore that

$$\begin{aligned}&\mathbb{E}^{\mathcal{W}}\bigl[|D_n(t,\mathbf{y}) - D_n(t,\mathbf{x})|^{p+1}\bigr]^{\frac{1}{p}} \\ &\quad\le \Bigl(\mathbb{E}^{\mathcal{W}}\bigl[|X(t,\mathbf{y}) - X(t,\mathbf{x})|^{2p}\bigr]^{\frac{1}{2p}} + \mathbb{E}^{\mathcal{W}}\bigl[|X_n(t,\mathbf{y}) - X_n(t,\mathbf{x})|^{2p}\bigr]^{\frac{1}{2p}}\Bigr) \\ &\qquad\times\Bigl(\mathbb{E}^{\mathcal{W}}\bigl[|D_n(t,\mathbf{y})|^2\bigr]^{\frac{1}{2}} + \mathbb{E}^{\mathcal{W}}\bigl[|D_n(t,\mathbf{x})|^2\bigr]^{\frac{1}{2}}\Bigr)^{\frac{1}{p}}.\end{aligned}$$

By (126), there is a $C < \infty$ such that

$$\Bigl(\mathbb{E}^{\mathcal{W}}\bigl[|D_n(t,\mathbf{y})|^2\bigr]^{\frac{1}{2}} + \mathbb{E}^{\mathcal{W}}\bigl[|D_n(t,\mathbf{x})|^2\bigr]^{\frac{1}{2}}\Bigr)^{\frac{1}{p}} \le C(1+|\mathbf{x}|+|\mathbf{y}|)2^{-\frac{n}{p}}.$$

As for the first factor on the right, observe that

$$\begin{aligned}X(t,\mathbf{y}) - X(t,\mathbf{x}) =&\,\mathbf{y}-\mathbf{x} + \int_0^t \Bigl(V_0\bigl(X(\tau,\mathbf{y})\bigr) - V_0\bigl(X(\tau,\mathbf{x})\bigr)\Bigr)\,d\tau \\ &+ \sum_{k=1}^M \int_0^t \Bigl(V_k\bigl(X(\tau,\mathbf{y})\bigr) - V_0\bigl(X(\tau,\mathbf{x})\bigr)\Bigr) \bullet dw(\tau)_k,\end{aligned}$$

and so, after converting to Itô integrals, applying (60) and Gromwall's lemma, one sees that

$$\mathbb{E}^{\mathcal{W}}\bigl[|X(t,\mathbf{y}) - X(t,\mathbf{x})|^{p+1}\bigr] \le C|\mathbf{y}-\mathbf{x}|^{\frac{p}{p+1}}$$

for some $C < \infty$. To derive the analogous estimate for $X_n(t,\mathbf{y}) - X_n(t,\mathbf{x})$, remember that

$$\begin{aligned}&X_n(t,\mathbf{y}) - X_n(t,\mathbf{x}) \\ &\quad= \mathbf{y}-\mathbf{x} + \int_0^t \Bigl(E_0\bigl(\Delta_n(\tau), X_n([\tau]_n,\mathbf{y})\bigr) - E_0\bigl(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\bigr)\Bigr)\,d\tau \\ &\quad+ \sum_{k=1}^M \int_0^t \Bigl(E_k\bigl(\Delta_n(\tau), X_n([\tau]_n,\mathbf{y})\bigr) - E_k\bigl(\Delta_n(\tau), X_n([\tau]_n,\mathbf{x})\bigr)\Bigr) \bullet dw(\tau)_k.\end{aligned}$$

Thus, by again converting to Itô integrals, applying (60), using the estimates in (120) and remembering that $\Delta_n(\tau)$ is independent of $X_n([\tau]_n, \mathbf{x})$, one arrives that the same sort of estimate in terms of $|\mathbf{y}-\mathbf{x}|$. After combining these, one has that

$$\mathbb{E}^{\mathcal{W}}\big[|D_n(t,\mathbf{y}) - D_n(t,\mathbf{x})|^{p+1}\big]^{\frac{1}{p+1}} \le C(1+|\mathbf{x}|+|\mathbf{y}|)2^{\frac{n}{p+1}}|\mathbf{y}-\mathbf{x}|^{\frac{p}{p+1}}$$

for some $C<\infty$. Hence, by taking $p>N$, Kolmogorov's criterion says that there exists an $\alpha>0$ such that

$$\mathbb{E}^{\mathcal{W}}\big[\|D_n(t,\,\cdot\,)\|_{[-R,R]^N}^{p+1}\big]^{\frac{1}{p+1}} \le C(R)2^{-\alpha n}$$

for some $C(R)<\infty$. Since the same sort of estimate holds for $\check{D}_n(t,\,\cdot\,)$, we have now proved the following result.

Theorem 130. *Let $X(\,\cdot\,,\mathbf{x})$ be the solution to (118). Then, for each $t\ge 0$, $X(t,\,\cdot\,)$ is a diffeomorphism from $\mathbb{R}^N$ onto itself. Furthermore, for each $t>0$, the distrbution of $X(t,\,\cdot\,)^{-1}$ is the same as the distribution of $\check{X}(t,\,\cdot\,)$, where $\check{X}(\,\cdot\,,\mathbf{x})$ is the solution to (118) with each V_k replaced by $-V_k$.*