The Annals of Probability
1996, Vol. 24, No. 1, 196–205

ON THE EXISTENCE OF UNIVERSAL FUNCTIONAL SOLUTIONS TO CLASSICAL SDE'S[1]

BY OLAV KALLENBERG

Auburn University

Assume that weak existence and pathwise uniqueness hold for solutions to the equation $dX_t = \sigma(t, X)\,dB_t + b(t, X)\,dt$ starting at fixed points. Then there exists a Borel measurable function F, such that any solution (X, B) satisfies $X = F(X_0, B)$ a.s. This strengthens a fundamental result of Yamada and Watanabe, where F may depend on the initial distribution μ.

1. Introduction and main results. Throughout the paper we shall consider a general stochastic differential equation (SDE) of the form

$$dX_t^i = \sigma_j^i(t, X)\,dB_t^j + b^i(t, X)\,dt, \tag{1}$$

where $\sigma = (\sigma_j^i)$ and $b = (b^i)$ are predictable processes of suitable dimension, defined on the canonical space $C(\mathbb{R}_+, \mathbb{R}^d)$ equipped with the raw induced filtration. By a *solution* to (1) is meant a pair (X, B) of processes on some probability space with filtration $\mathscr{F}$, such that B is an $\mathscr{F}$-Brownian motion in $\mathbb{R}^r$ while X is a continuous $\mathscr{F}$-adapted process satisfying (1). In order for the integrals in (1) to make sense, the process X must fulfill the integrability condition

$$\int_0^t (\|a(s, X)\| + |b(s, X)|)\,ds < \infty \quad \text{a.s., } t \geq 0, \tag{2}$$

where $a^{ij} = \sigma_k^i \sigma_k^j$ or $a = \sigma\sigma^T$, while $\|\cdot\|$ is any norm in the space of $d \times d$ matrices.

By *weak existence* with initial distribution μ we shall mean the existence (on a suitable filtered probability space) of a solution (X, B), such that X_0 has distribution μ. In contrast, *strong existence* means that (1) can be solved on any given probability space with a complete filtration $\mathscr{F}$, an associated Brownian motion B and an $\mathscr{F}_0$-measurable random variable ξ with distribution μ, in such a way that $X_0 = \xi$ a.s. The corresponding notions of uniqueness are those of *uniqueness in law*, meaning that the distribution of X depends only on μ, and of *pathwise uniqueness*, which means that if X and Y are two solutions with $X_0 = Y_0$ a.s. and initial distribution μ, defined on the

Received December 1994; revised May 1995.

[1]Research supported by NSF Grant DMS-93-03050.

AMS 1991 *subject classifications*. Primary 60H10; secondary 60G44.

Key words and phrases. Weak and strong solutions, pathwise uniqueness, local martingale problem.

same filtered probability space with a common Brownian motion B, then X and Y are indistinguishable.

In a remarkable classical paper, Yamada and Watanabe (1971) proved the following fundamental result. *Assume that weak existence and pathwise uniqueness hold for solutions with a fixed initial distribution* μ. *Then strong existence and uniqueness in law hold for the same initial distribution. Furthermore, there exists a Borel measurable function* F_μ: $\mathbb{R}^d \times C(\mathbb{R}_+, \mathbb{R}^r) \to C(\mathbb{R}_+, \mathbb{R}^d)$, *such that any solution* (X, B) *with initial distribution* μ *satisfies* $X = F_\mu(X_0, B)$ *a.s.*

Now assume in addition that weak existence and pathwise uniqueness hold for solutions starting at arbitrary fixed points x, that is, for all initial distributions δ_x, and write $F_x(w) = F_{\delta_x}(x, w)$. A simple conditioning argument then yields

$$\mu\left\{x \in \mathbb{R}^d : F_\mu(x, B) = F_x(B) \text{ a.s.}\right\} = 1, \tag{3}$$

as noted already by Yamada and Watanabe. Unfortunately, one cannot conclude without product measurability that $X = F_\mu(X_0, B) = F_{X_0}(B)$ a.s. In order to obtain the desired universal representation $X = F(X_0, B)$ a.s., it is necessary first to construct a measurable modification of the process $F_x(w)$ on the canonical path space. Our primary aim is to strengthen the Yamada–Watanabe theorem by showing that such a universal representation function F does exist.

THEOREM 1. *Assume that weak existence and pathwise uniqueness hold for solutions starting at arbitrary fixed points. Then strong existence and uniqueness in law hold for every initial distribution. Furthermore, there exists a Borel measurable and universally predictable function* F: $\mathbb{R}^d \times C(\mathbb{R}_+, \mathbb{R}^r) \to C(\mathbb{R}_+, \mathbb{R}^d)$ *such that any solution* (X, B) *satisfies* $X = F(X_0, B)$ *a.s.*

To explain the predictability assertion, let $\mathscr{G}$ denote the induced filtration on the canonical space $\mathbb{R}^d \times C(\mathbb{R}_+, \mathbb{R}^r)$, write $\mathscr{G}^\mu$ for the completion of $\mathscr{G}$ with respect to $\mu \times W^r$, where W^r is Wiener measure on $C(\mathbb{R}_+, \mathbb{R}^r)$, and form $\overline{\mathscr{G}}_t = \bigcap_\mu \mathscr{G}_t^\mu$ for $t \geq 0$. Then F is $\overline{\mathscr{G}}$-predictable.

Our proof extends immediately to the case when weak existence and pathwise uniqueness are only assumed for solutions starting at fixed points in some Borel set $D \subset \mathbb{R}^d$. Still we get a functional representation $X = F(X_0, B)$ a.s., but only for solutions X with $X_0 \in D$ a.s. A similar remark applies to the subsequent propositions.

CAUTION. In the original statement of Yamada and Watanabe [(1971), Corollary 3], the authors actually claim the existence of a universally measurable function F, such that any solution (X, B) to (1) satisfies $X = F(X_0, B)$ a.s. However, the argument provided is insufficient to prove the claim, and a corrected version appears in Ikeda and Watanabe [(1989), Theorem IV.1.1].

Even there the functional solution is written in universal form as $X = F(X_0, B)$ a.s., but then the authors define the "function" F to be equal to F_μ whenever X_0 has distribution μ.

For the proof of Theorem 1, we shall need the measurability part of the following proposition, which is only a modest extension of a result by Stroock and Varadhan [(1979), Exercise 6.7.4]. A weaker version with a different proof appears in Yamada and Watanabe [(1971), Corollary 2]. The last assertion, also essentially due to Stroock and Varadhan, follows easily by the same proof and is added here because of its independent interest.

PROPOSITION 1. *Assume that weak existence and uniqueness in law hold for solutions starting at arbitrary fixed points. Then the two properties remain valid for arbitrary initial distributions. Moreover, the corresponding distributions P_x form a kernel from $\mathbb{R}^d$ to $C(\mathbb{R}_+, \mathbb{R}^d)$, and in the diffusion case they further satisfy the strong Markov property.*

In particular, it is interesting to note from the last two results that weak existence, in the stated combination with pathwise uniqueness or uniqueness in law, extends from degenerate to arbitrary initial distributions. For the two uniqueness properties, the extension can be established directly by a simple conditioning argument, as pointed out in Ikeda and Watanabe [(1989), Remarks IV.1.2 and IV.1.4]. For the weak existence property alone, the situation may be less obvious and worth recording.

PROPOSITION 2. *Assume that weak existence holds for solutions starting at arbitrary fixed points. Then weak existence holds for any initial distribution.*

The last two results may be rephrased in terms of the corresponding local martingale problem, which will play an important role in our proofs. Here we define

$$M_t^f = f(X_t) - f(X_0) - \int_0^t A_s f(X)\, ds, \qquad t \geq 0,\ f \in C_K^\infty,$$

where C_K^∞ denotes the class of infinitely differentiable functions $f\colon \mathbb{R}^d \to \mathbb{R}$ with bounded support, while the operators A_s are defined by

$$A_s f(x) = \tfrac{1}{2} a^{ij}(s, x) f''_{ij}(x_s) + b^i(s, x) f'_i(x_s), \qquad s \geq 0,\ f \in C_K^\infty.$$

Recall that a process X or its distribution P is said to solve the *local martingale problem* for (a, b), if M^f is a local martingale for every $f \in C_K^\infty$. By a fundamental result of Stroock and Varadhan, a measure P has the stated property iff there exists a solution (X, B) to (1) such that X has distribution P.

The area of this paper is of course very classical, and the basic definitions and results may be found in many textbooks on stochastic calculus. Inexperienced readers may find the detailed discussion in Rogers and Williams

[(1987), Sections V.8–10, 16–17 and 19–21] particularly helpful. Note incidentally that, in their terminology, Theorem 1 states that equation (1) is exact iff it has a unique (up to measurable modifications) strong solution (which may eliminate the need for a dual terminology).

A referee kindly called attention to a paper by Kaneko and Nakao (1988), where a Borel measurable strong solution is obtained for pathwise unique SDE's of Markovian type with continuous coefficients of linear growth. This result was obtained as a by-product of their main result that the associated Euler–Maruyama scheme converges a.s. for some subsequence independent of the initial values.

2. Some technical lemmas. We begin with some elementary properties of kernels. By a (*probability*) *kernel* between two measurable spaces $(S, \mathscr{S})$ and $(T, \mathscr{T})$ is meant a mapping μ: $S \times \mathscr{T} \to [0, \infty]$, such that $\mu(s, B)$ is $\mathscr{S}$-measurable in $s \in S$ for fixed $B \in \mathscr{T}$ and a (probability) measure in $B \in \mathscr{T}$ for fixed $s \in S$. Alternatively, μ may be regarded as a measurable map from S to $\mathscr{M}(T)$ [or $\mathscr{P}(T)$], the space of (probability) measures m on T, endowed with the σ-field induced by the functions $m \mapsto mB$, $B \in \mathscr{T}$.

LEMMA 1. *Fix three measurable spaces $(S, \mathscr{S})$, $(T, \mathscr{T})$ and $(U, \mathscr{U})$, and let μ be a probability kernel from S to T. Then*:

(i) *For any $\nu \in \mathscr{P}(U)$, the mapping $s \mapsto \mu_s \otimes \nu$ is a probability kernel from S to $T \times U$.*

(ii) *For any measurable function f: $S \times T \to U$, the mapping $s \mapsto \mu_s \circ (f(s, \cdot))^{-1}$ is a probability kernel from S to U.*

PROOF. (i) Note that $(\mu_s \otimes \nu)A$ is $\mathscr{S}$-measurable when $A = B \times C$ with $B \in \mathscr{T}$ and $C \in \mathscr{U}$, and extend by a monotone class argument.

(ii) For any $D \in \mathscr{U}$ we need to show that $\mu_s\{t \in T;\ f(s, t) \in D\}$ is $\mathscr{S}$-measurable. Equivalently we may prove for any set $A \in \mathscr{S} \otimes \mathscr{T}$ with sections $A_s = \{t \in T;\ (s, t) \in A\}$ that the function $s \mapsto \mu_s A_s$ is $\mathscr{S}$-measurable. Again this is obvious for product sets $A = B \times C$ with $B \in \mathscr{S}$ and $C \in \mathscr{T}$, and it follows in general by a monotone class argument. $\square$

For the next result, say that a measure on a metric space is *degenerate* if its support contains at most one point.

LEMMA 2. *Let μ be a kernel from a measurable space $(S, \mathscr{S})$ to a separable metric space T, and let D denote the set of all $s \in S$, such that the measure μ_s is degenerate. Then $D \in \mathscr{S}$.*

PROOF. Fix a countable topological base $B_1, B_2, \dots$ in T, define $J = \{(i, j);\ B_i \cap B_j = \varnothing\}$ and note that

$$D = \left\{ s \in S:\ \sum_{(i, j) \in J} (\mu_s B_i)(\mu_s B_j) = 0 \right\}. \qquad \square$$

We proceed with a kernel version of the existence theorem for regular conditional distributions.

LEMMA 3. *Fix a measurable space $(S, \mathscr{S})$ and two Polish spaces T and U with Borel σ-fields $\mathscr{T}$ and $\mathscr{U}$, and let μ be a probability kernel from S to $T \times U$. Then the Radon–Nikodym densities*

$$\nu(s, t, B) = \frac{\mu(s, dt \times B)}{\mu(s, dt \times U)}, \qquad s \in S, t \in T, B \in \mathscr{U},$$

have versions which form a probability kernel from $S \times T$ to U.

PROOF. For each $B \in \mathscr{U}$ we may use Doob's martingale approach to Radon–Nikodym densities, as described in Dellacherie and Meyer [(1980), Theorem V.58], to produce a product-measurable version of the function $(s, t) \mapsto \nu(s, t, B)$. Then proceed as in the usual construction of regular conditional distributions, noting that in each step the exceptional (s, t)-set A lies in $\mathscr{S} \otimes \mathscr{T}$ and satisfies $\mu_s A_s = 0$ for all $s \in S$, where $A_s = \{t \in T; (s, t) \in A\}$. $\square$

To motivate the next result, consider some random elements $\xi_1, \xi_2, \dots$ and ξ in a complete metric space S, such that $\xi_n \to_P \xi$. Then ξ is the a.s. limit of a subsequence and hence a.s. of the form $\xi = F(\xi_1, \xi_2, \dots)$ for some measurable function F: $S^\infty \to S$. However, the subsequence and then also F will depend on the distribution μ of (ξ_n), and for varying μ the representation becomes instead $\xi = F(\mu, \xi_1, \xi_2, \dots)$ a.s. This shows that, whenever a random object is constructed by successive approximations in probability, we are forced to consider approximating sequences of the form $\xi_n = F_n(\mu, X)$. In this setting, the measurability of the representing functions extends to the limit.

LEMMA 4. *Fix a measurable space S, a complete metric space (T, ρ) and a sequence of measurable functions $F_1, F_2, \dots : \mathscr{P}(S) \times S \to T$. Then there exists a measurable function F: $\mathscr{P}(S) \times S \to T$, such that whenever X is a random element in S with distribution μ for which $F_n(\mu, X)$ converges in probability, the limit equals $F(\mu, X)$ a.s.*

PROOF. For any $k \in \mathbb{N}$ and $\mu \in \mathscr{P}(T^\infty)$ we define

$$n_k(\mu) = \inf\left\{n \in \mathbb{N}; \sup_{m \geq n} \int [\rho(t_m, t_n) \wedge 1] \, \mu(dt) \leq 2^{-k}\right\}$$

and we note that the functions n_k are measurable. If (ξ_n) is a random sequence in T with distribution μ satisfying $\xi_n \to_P \xi$, then $n_k(\mu) < \infty$ for all k and we have

$$E \sum_k \left\{\rho\left(\xi_{n_k(\mu)}, \xi_{n_{k+1}(\mu)}\right) \wedge 1\right\} \leq \sum_k 2^{-k} < \infty.$$

Hence the sum on the left is a.s. finite, so the sequence $(\xi_{n_k(\mu)})$ is a.s. Cauchy, and therefore $\xi_{n_k(\mu)} \to \xi$ a.s. Thus $\xi = G(\mu, \xi_1, \xi_2, \dots)$ a.s., where

$G(\mu, t_1, t_2, \dots) = \lim_k t_{n_k(\mu)}$ whenever the $n_k(\mu)$ are finite and such that the limit exists, while otherwise $G(\mu, t_1, t_2, \dots) = t_0$ for some fixed $t_0 \in T$.

To see that G is measurable, let L denote the set of convergent sequences $t = (t_n) \in T^\infty$, define $l(t) = \lim_n t_n$ on L and put $l(t) = t_0$ on L^c. Since $G(\mu, t) = l \circ (t_{n_k(\mu)})$ on the set where the $n_k(\mu)$ are finite and the sequence $(t_{n_k(\mu)})$ is measurable on the same set, it remains to show that l is measurable. Then note that L is a measurable subset of T^∞, since $t \in L$ is equivalent to the Cauchy convergence of (t_n). Furthermore $f(t_n) \to f \circ l(t)$ on L for any continuous function f: $T \to \mathbb{R}$, so $f \circ l$ is measurable, and by a simple approximation it follows that $1_G \circ l$ is measurable for any open set $G \subset T$, which implies the desired measurability of l.

Now consider a random element X in S with distribution μ and let $\xi_n = F_n(\mu, X)$, $n \in \mathbb{N}$, for some measurable functions F_n: $\mathscr{P}(S) \times S \to T$. If $\xi_n \to_P \xi$, the previous argument yields $\xi = G(\nu, \xi_1, \xi_2, \dots)$ a.s., where ν denotes the distribution of the sequence (ξ_n). Thus it remains to show that ν is a measurable function of μ. However, this is clear from Lemma 1(ii), applied to the kernel $K(\mu, B) = \mu B$ from $\mathscr{P}(S)$ to S and the function $f = (F_n)$: $\mathscr{P}(S) \times S \to T^\infty$. $\square$

The last result will now be applied to stochastic integrals with respect to solution processes X. Write $L(X^i)$ for the class of predictable processes Y that are locally integrable with respect to X^i, and denote the corresponding stochastic integral processes by $Y \cdot X^i$. Integrals with respect to Lebesgue measure λ will be written as $Y \cdot \lambda$.

LEMMA 5. *Fix any predictable function f: $C(\mathbb{R}_+, \mathbb{R}^d) \to C(\mathbb{R}_+, \mathbb{R}^d)$. Then there exists some measurable function*

$$F: \mathscr{P}\big(C(\mathbb{R}_+, \mathbb{R}^d)\big) \times C\big(\mathbb{R}_+, \mathbb{R}^d\big) \to C(\mathbb{R}_+, \mathbb{R}), \tag{4}$$

such that whenever X is a process with distribution μ that solves the local martingale problem for (a, b) and satisfies $f^i(X) \in L(X^i)$ for all i, we have $f^i(X) \cdot X^i = F(\mu, X)$ a.s.

PROOF. Recall that X is a semimartingale with drift component $b(X) \cdot \lambda$ and covariation processes $[X^i, X^j] = a^{ij}(X) \cdot \lambda$ [cf. Rogers and Williams (1987), Section V.20]. Hence $f^i(X) \in L(X^i)$ for all i, iff the processes $(f^i)^2 a^{ii}(X)$ and $f^i b^i(X)$ are a.s. Lebesgue integrable, and in that case $f_n^i(X) \cdot X^i \to_P f^i(X) \cdot X^i$ in the sense of uniform convergence on bounded intervals, whenever a.s.

$$\big(f_n^i - f^i\big)^2 a^{ii}(X) \cdot \lambda \to 0, \qquad \big(f_n^i - f^i\big) b^i(X) \cdot \lambda \to 0. \tag{5}$$

Assuming that $f_n^i(X) \cdot X^i = F_n(\mu, X)$ a.s. for some measurable functions F_n as in (4), we may then conclude by Lemma 4 that the limit $f^i(X) \cdot X^i$ has a representation of the same kind. This argument will now be applied in three steps.

First we define $f_n(x) = f(x)1\{|f(x)| \le n\}$. Then (5) holds by dominated convergence, so $f_n^i(X) \cdot X^i \to_P f^i(X) \cdot X^i$. Thus if $f_n^i(X) \cdot X^i = F_n(\mu, X)$ a.s. for some measurable functions F_n, we get a similar representation for the process $f^i(X) \cdot X^i$, which reduces the discussion to the case of bounded f.

Next we may introduce the moving averages

$$f_n(t,x) = n\int_{(t-n^{-1})_+}^{t} f(s,x)\,ds, \qquad t \ge 0,\, n \in \mathbb{N},\, x \in C\big(\mathbb{R}_+, \mathbb{R}^d\big),$$

which are clearly continuous and adapted, hence predictable. By Lebesgue's differentiation theorem we have

$$\lambda\big\{s \ge 0;\, f_n(s,x) \nrightarrow f(s,x)\big\} = 0, \qquad x \in C\big(\mathbb{R}_+, \mathbb{R}^d\big),$$

so (5) holds by dominated convergence, and again $f_n^i(X) \cdot X^i \to_P f^i(X) \cdot X^i$. By the previous argument, this reduces the discussion to the case of bounded and continuous f.

We may finally reduce to the case of simple predictable integrands, by taking $f_n(s,x) = f(2^{-n}[2^n s], x)$ and using dominated convergence and Lemma 4 as before. However, for simple predictable f, the values of the process $f^i(X) \cdot X^i$ are finite sums of products $f^i(s,X)(X_t^i - X_s^i)$, so in this case we have trivially a representation $f^i(X) \cdot X^i = F(X)$ for some measurable function $F\colon C(\mathbb{R}_+, \mathbb{R}^d) \to C(\mathbb{R}_+, \mathbb{R})$. $\square$

The previous result will now be applied to the construction of a weak solution to (1) from a solution to the associated local martingale problem.

LEMMA 6. *There exists a measurable function*

$$F\colon \mathscr{P}\big(C\big(\mathbb{R}_+, \mathbb{R}^d\big)\big) \times C\big(\mathbb{R}_+, \mathbb{R}^{d+r}\big) \to C\big(\mathbb{R}_+, \mathbb{R}^r\big),$$

such that whenever X is a process with distribution μ that solves the local martingale problem for (a,b) while Y is an independent Brownian motion in $\mathbb{R}^r$, the process $B = F(\mu, X, Y)$ is another Brownian motion in $\mathbb{R}^r$, such that the pair (X, B) with induced filtration solves (1).

PROOF. We may follow the usual construction of B, as described in, for example, Rogers and Williams [(1987), Theorem V.20]. This involves, as the only nonelementary step, a stochastic integration with respect to the continuous semimartingale (X, Y), where the integrand is a predictable function of X obtained by some elementary matrix algebra. Now even (X, Y) is a solution to a local martingale problem, so Lemma 5 yields the desired functional representation. $\square$

3. Proofs of main results.

PROOF OF PROPOSITION 1. Write $\mathscr{P} = \mathscr{P}(C(\mathbb{R}_+, \mathbb{R}^d))$, let $\mathscr{P}_M$ denote the class of measures $P \in \mathscr{P}$ that solve the local martingale problem for (a,b) with a degenerate initial distribution and note that $\mathscr{P}_M = \{P_x;\ x \in \mathbb{R}^d\}$.

Following the argument outlined in Stroock and Varadhan [(1979), Exercise 6.7.4], we need to show that $\mathscr{P}_M$ is a Borel subset of $\mathscr{P}$, since the asserted measurability will then follow by Kuratowski's theorem, the fact that any measurable bijection between Borel subsets of Polish spaces has a measurable inverse [cf. Parthasarathy (1967), Section I.3].

To this aim, we note that the class C_K^∞ in the formulation of the local martingale problem may be replaced by a countable subclass $\mathscr{C}$, consisting of suitably truncated versions of the coordinate functions x^i and their products $x^i x^j$. In fact, in the proof of the fundamental equivalence between weak solutions to (1) and solutions to the associated local martingale problem, one needs to consider the processes M^f only for such functions f [cf. Rogers and Williams (1987), Theorem V.20].

Now introduce the canonical process X in $C(\mathbb{R}_+, \mathbb{R}^d)$ and let D denote the class of measures $P \in \mathscr{P}$ such that $P \circ X_0^{-1}$ is degenerate. Further let I consist of all measures $P \in \mathscr{P}$, such that X satisfies the integrability condition (2). Finally put $\tau_n^f = \inf\{t;\ |M_t^f| \geq n\}$ and let L denote the class of measures $P \in \mathscr{P}$, such that the processes $M^{f,n} = M^f(\cdot \wedge \tau_n^f -)$ are martingales under P for arbitrary $f \in \mathscr{C}$ and $n \in \mathbb{N}$. Then $\mathscr{P}_M = D \cap I \cap L$ and it suffices to show that D, I and L are measurable.

For D the measurability is clear by Lemma 2. Even I is measurable, since the integrals on the left of (2) are measurable by Fubini's theorem. Finally L is measurable, since the relation $P \in L$ is equivalent to countably many conditions of the form $E[M_t^{f,n} - M_s^{f,n}; A] = 0$, with $f \in \mathscr{C}$, $n \in \mathbb{N}$, $s < t$ in $\mathbb{Q}_+$ and $A \in \mathscr{F}_s$.

To prove the strong Markov property in the diffusion case, it is enough to show for each x and any bounded stopping time τ that $P_x[\theta_\tau \in \cdot | \mathscr{F}_\tau] \in L$ a.s. However, this is equivalent to countably many relations of the form

$$E_x\big[\big(M_t^{f,n} - M_s^{f,n}\big)1_A \circ \theta_\tau | \mathscr{F}_\tau\big] = 0 \quad \text{a.s.},$$

each of which follows by Doob's optional sampling theorem from the local martingale property of M^f under P_x. $\square$

PROOF OF THEOREM 1. By the Yamada–Watanabe theorem there is uniqueness in law for solutions starting at fixed points, and by Proposition 1 the corresponding distributions P_x form a kernel from $\mathbb{R}^d$ to $C(\mathbb{R}_+, \mathbb{R}^d)$. Let G denote the function in Lemma 6, and note that if X is a process with distribution P_x while Y is an independent Brownian motion in $\mathbb{R}^r$, then $B = G(P_x, X, Y)$ is another Brownian motion in $\mathbb{R}^r$, such that the pair (X, B) solves (1). Writing Q_x for the distribution of (X, B), we may conclude from Lemma 1, (i) and (ii), that the mapping $x \mapsto Q_x$ is a kernel from $\mathbb{R}^d$ to $C(\mathbb{R}_+, \mathbb{R}^{d+r})$.

Now change the notation and write (X, B) for the canonical process on $C(\mathbb{R}_+, \mathbb{R}^{d+r})$. By the Yamada–Watanabe theorem we have $X = F_x(B)$ a.s. Q_x, so

$$Q_x[X \in \cdot | B] = \delta_{F_x(B)} \quad \text{a.s., } x \in \mathbb{R}^d. \tag{6}$$

By Lemma 3 the conditional distributions $Q_x[X \in \cdot | B \in dw]$ have versions $\nu_{x,w}$ which form a probability kernel from $\mathbb{R}^d \times C(\mathbb{R}_+, \mathbb{R}^r)$ to $C(\mathbb{R}_+, \mathbb{R}^d)$, and (6) shows that $\nu_{x,w}$ is a.s. degenerate for each x. Since the set of degenerate measures is measurable by Lemma 2, we may modify ν such that $\nu_{x,w}$ becomes degenerate for all x and w, and hence of the form

$$\nu_{x,w} = \delta_{F(x,w)}, \qquad x \in \mathbb{R}^d, w \in C(\mathbb{R}_+, \mathbb{R}^r), \tag{7}$$

for some function $F: \mathbb{R}^d \times C(\mathbb{R}_+, \mathbb{R}^r) \to C(\mathbb{R}_+, \mathbb{R}^d)$. The kernel property of ν implies that F is product measurable, while a comparison of (6) and (7) yields $F(x, B) = F_x(B)$ a.s. for each x.

Now fix an arbitrary probability measure μ on $\mathbb{R}^d$, and conclude from Proposition 1 and its proof that $P_\mu = \int P_x\, \mu(dx)$ solves the local martingale problem for (a, b) with initial distribution μ. Hence there exists a solution (X, B) to (1), such that X_0 has distribution μ. Since (1) remains valid under conditioning on X_0, the pathwise uniqueness in the degenerate case implies $P[X = F(X_0, B)|X_0] = 1$ a.s., and we get $X = F(X_0, B)$ a.s. In particular, pathwise uniqueness holds for arbitrary initial distributions μ.

We return to the canonical setting and write (ξ, B) for the identity mapping on the space $\mathbb{R}^d \times C(\mathbb{R}_+, \mathbb{R}^r)$ with probability measure $\mu \otimes W^r$ and induced completed filtration $\mathscr{G}^\mu$. By the Yamada–Watanabe theorem there exists a $\mathscr{G}^\mu$-adapted solution $X = F_\mu(\xi, B)$ to (1) with $X_0 = \xi$ a.s., and by the previous discussion we have even $X = F(\xi, B)$ a.s. Hence F is adapted to $\mathscr{G}^\mu$, and since μ is arbitrary, the adaptedness extends to the universal completion $\overline{\mathscr{G}}_t = \bigcap_\mu \mathscr{G}_t^\mu$, $t \geq 0$. The asserted $\overline{\mathscr{G}}$-predictability now follows since F has continuous paths. □

PROOF OF PROPOSITION 2. Define $\mathscr{P}$ and $\mathscr{P}_M$ as before, and recall from the proof of Proposition 1 that $\mathscr{P}_M$ is a Borel subset of the Polish space $\mathscr{P}$, hence Borel isomorphic to a Borel set in $\mathbb{R}_+$. Letting $\varphi(P)$ denote the starting point associated with a measure $P \in \mathscr{P}_M$, it is further seen that φ is continuous and hence a measurable map from $\mathscr{P}_M$ to $\mathbb{R}^d$. Since $\mathbb{R}^d$ is separable, it follows by a simple approximation that the graph $G = \{(P, \varphi(P)); P \in \mathscr{P}_M\}$ is a Borel subset of $\mathscr{P} \times \mathbb{R}^d$.

Now fix a probability measure μ on $\mathbb{R}^d$. By the section theorem in Dellacherie and Meyer [(1975), III.44(b)], there exists a measurable mapping $x \mapsto P_x^\mu$ from $\mathbb{R}^d$ to $\mathscr{P}_M$, such that $(P_x^\mu, x) \in G$ for $x \in \mathbb{R}^d$ a.e. μ. The measures P_x^μ clearly solve the local martingale problem for (a, b), while the corresponding initial distributions equal δ_x a.e. μ. Thus $P_\mu = \int P_x^\mu \mu(dx)$ has initial distribution μ, and from the proof of Proposition 1 we note that even P_μ solves the local martingale problem for (a, b). Thus (1) has a weak solution with initial distribution μ. □

REFERENCES

DELLACHERIE, C. and MEYER, P.-A. (1975, 1980). *Probabilités et Potentiel* Chapters I–VIII. Hermann, Paris.

IKEDA, N. and WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. North-Holland, Amsterdam.

KANEKO, H. and NAKAO, S. (1988). A note on approximation for stochastic differential equations. *Séminaire de Probabilités XXII*. *Lecture Notes in Math*. **1321** 155–162. Springer, Berlin.

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

ROGERS, L. C. G. and WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales* **2**. Wiley, New York.

STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer, New York.

YAMADA, T. and WATANABE, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ*. **11** 155–167.

DEPARTMENT OF MATHEMATICS
228 PARKER HALL
AUBURN UNIVERSITY
AUBURN, ALABAMA 36849-5310
E-mail: clark@mail.auburn.edu