1995

Smooth Densities for Degenerate Stochastic Delay Equations with Hereditary Drift

Denis R. Bell
University of North Florida

Salah-Eldin A. Mohammed
Southern Illinois University Carbondale, salah@sfde.math.siu.edu

Published in Annals of Probability, 23(4), 1875-1894.

The Annals of Probability
1995, Vol. 23, No. 4, 1875–1894

SMOOTH DENSITIES FOR DEGENERATE STOCHASTIC DELAY EQUATIONS WITH HEREDITARY DRIFT

BY DENIS R. BELL[1] AND SALAH-ELDIN A. MOHAMMED[2]

University of North Florida and Southern Illinois University at Carbondale

We establish the existence of smooth densities for solutions of $\mathbf{R}^d$-valued stochastic hereditary differential systems of the form

$$dx(t) = H(t, x)\,dt + g(t, x(t-r))\,dW(t).$$

In the above equation, W is an n-dimensional Wiener process, r is a positive time delay, H is a nonanticipating functional defined on the space of paths in $\mathbf{R}^d$ and g is an $n \times d$ matrix-valued function defined on $[0,\infty) \times \mathbf{R}^d$, such that gg^* has degeneracies of polynomial order on a hypersurface in $\mathbf{R}^d$. In the course of proving this result, we establish a very general criterion for the hypoellipticity of a class of degenerate parabolic second-order time-dependent differential operators with space-independent principal part.

1. Introduction. In an earlier paper [2], the authors used the Malliavin calculus to establish the existence of smooth densities for $\mathbf{R}^d$-valued stochastic differential delay equations of the form

$$dx(t) = g(x(t-r))\,dW(t). \tag{1.1}$$

Here r denotes a strictly positive *time delay*, W is an n-dimensional Wiener process and g is an $n \times d$ matrix-valued function defined on $\mathbf{R}^d$, such that gg^* has a single point of degeneracy of linear order. The novelty of this result lies in the fact that, in contrast to a classical (i.e., nondelayed) diffusion, the solution process x in (1.1) is *non-Markov*. In fact, if we define x_t, $t \geq 0$, by $x_t(s) := x(t+s)$, $-r \leq s \leq 0$, then $\{x_t: t \geq 0\}$ is a Feller process with values in the space $C([-r,0],\mathbf{R}^d)$ of initial paths ([9], Chapters 3 and 4). The infinitesimal generator of the trajectory process $\{x_t: t \geq 0\}$ is a second-order differential operator on infinite-dimensional space whose principal part degenerates on a surface of *finite* codimension. Thus (1.1) does not correspond to a differential operator on $\mathbf{R}^d$ (cf. [9]), and the result in [2] cannot be obtained via existing techniques from partial differential equations.

Received July 1994.
[1]Research supported in part by NSF Grant DMS-91-21406.
[2]Research supported in part by NSF Grants DMS-89-07857 and DMS-92-06785.
AMS 1991 *subject classifications.* Primary 60H07, 60H30, 34K50; secondary 60H10, 60H20, 34F05.
Key words and phrases. Hereditary delay systems, Malliavin calculus, smooth densities.

The Malliavin calculus is a powerful tool for studying the regularity of measures induced by stochastic differential equations. When applied to classical diffusions, it has been made to yield very sharp criteria for the existence of smooth densities (see, e.g., Kusuoka and Stroock [8] and Bell and Mohammed [3, 4]). The crucial part of this work consists of the derivation of probabilistic lower bounds on the *Malliavin covariance matrix* of the underlying diffusion process. However, the delicate analysis developed to obtain these lower bounds for degenerate Markovian systems is heavily dependent on the existence of an invertible stochastic flow. This does not hold for stochastic delay equations such as (1.1), where the coefficients in the equation depend on the past history of the solution process x. In fact (1.1) does not admit a stochastic flow on the state space $C([-r, 0], \mathbf{R}^d)$ [9, 10]. As a result of this pathological behavior, little is known about the existence of densities for degenerate non-Markov stochastic differential equations. Indeed, as far as we are aware, [2] is the first result of this nature (although Kusuoka and Stroock [7] have solved the problem under *strong ellipticity* hypotheses on the hereditary process).

In view of this dearth of information concerning regularity of non-Markovian systems, it is interesting to develop new methods to extend the highly specific result in [2] to the more general setting

$$dx(t) = H(t, x)\, dt + g(t, x(t-r))\, dW(t). \tag{A}$$

This is the object of the present article.

The main ideas in our analysis of (A) are as follows:

1. We reduce the stochastic hereditary equation (A) to an equation of a more elementary form (2.5), by *conditioning on the past history* of the solution process.
2. We use the Malliavin calculus to derive a new result (Theorem 2.2) concerning the existence of smooth densities for (2.5). As a by-product of this analysis we obtain a new and very general *mean-ellipticity* criterion for hypoellipticity of second-order time-dependent parabolic operators with space-independent principal terms (Theorem 2.3). It is not clear to us how this result could be obtained using existing techniques from the field of partial differential equations.
3. Introducing a process of *partial integration by parts* over the Wiener space, we show (Lemma 4.1) that the existence of smooth densities for (A) follows from an invertibility condition (4.3) on the Malliavin covariance matrix corresponding to the conditioned system (2.5).
4. Our main result is Theorem 2.1. This theorem asserts that the solution process in (A) admits smooth densities at all positive times, under hypotheses that allow the noise covariance to degenerate on a moving hypersurface in $\mathbf{R}^d$. The initial datum for (A) is a deterministic path defined on the time interval $[-r, 0]$. We assume that the initial path has only limited contact with the degeneracy surface, in a sense made precise

in condition (2.4). We then use the delay structure of the equation to *propagate forward* a probabilistic analogue of condition (2.4) (Lemma 4.2). This procedure yields probabilistic lower bounds on segments of the solution process x that allow us to establish condition (4.3). Theorem 2.1 then follows as indicated above in statement 3.

2. Statement of results. We will extend the result discussed above in the following directions:

1. A general class of equations of type (A) is considered, incorporating a drift term of hereditary type, which may depend arbitrarily on the history of the solution path x.
2. The function g is allowed to be time-dependent. Furthermore, the "single point degeneracy" hypothesis on g in [2] is relaxed to allow gg^* to degenerate anywhere on a *large class of moving hypersurfaces* in $\mathbf{R}^d$. In particular, degeneracy is allowed to occur on obstacles, such as the boundaries of moving spheres in $\mathbf{R}^d$.
3. The function $\mathbf{R}^d \ni x \mapsto \det gg^*(x)$ is allowed to vanish at any *polynomial* rate as x approaches the degeneracy hypersurface.

The following set of hypotheses and notation will be used throughout.

HYPOTHESES H. (i) W: $[0,\infty) \times \Omega \to \mathbf{R}^n$ *is a standard n-dimensional Wiener process, defined on a complete filtered probability space* $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, P)$.

(ii) g: $[0,\infty) \times \mathbf{R}^d \to \mathbf{R}^{d\times n}$ *is a continuous map into the space of $d \times n$ matrices, with bounded Fréchet derivatives in the space variables of all orders. The space $\mathbf{R}^{d\times n}$ is furnished with the Hilbert–Schmidt norm.*

(iii) *r is a positive real number and η: $[-r,0] \to \mathbf{R}^d$ is a continuous initial path.*

(iv) *C is the space of all continuous paths ξ: $[-r,\infty) \to \mathbf{R}^d$ given the compact-open topology. For every $t > 0$ we will denote by $C([-r,t],\mathbf{R}^d)$ the Banach space of all continuous paths ξ: $[-r,t] \to \mathbf{R}^d$ furnished with the supremum norm*

$$\|\xi\|_\infty = \sup_{-r\leq u\leq t} \|\xi(u)\|.$$

H: $[0,\infty) \times C \to \mathbf{R}^d$ is a globally bounded continuous map such that, for every $t > 0$ and $\xi \in C$, $H(t,\xi)$ depends only on $\{\xi(s)$: $-r \leq s \leq t\}$ and has partial Fréchet derivatives of all orders with respect to $\xi \in C([-r,t],\mathbf{R}^d)$, which are globally bounded in $(t,\xi) \in (0,\infty) \times C$. The symbol $H_\xi(t,\xi)$ will denote the partial Fréchet derivative of H with respect to ξ. Set

$$\alpha_t := \sup\{\|H_\xi(u,\xi)\|: u \in [0,t], \xi \in C([-r,u],\mathbf{R}^d)\}, \qquad t > 0,$$

and

$$\alpha_\infty := \sup\{\|H_\xi(u,\xi)\|: u \in [0,\infty), \xi \in C([-r,u],\mathbf{R}^d)\},$$

where $\|H_\xi(u, \xi)\|$ *stands for the operator norm of* $H_\xi(u, \xi): C([-r, u], \mathbf{R}^d) \to \mathbf{R}^d$.

Under the above hypotheses, it is known that the stochastic differential equation

$$(2.1) \qquad \begin{aligned} dx(t) &= H(t, x)\,dt + g(t, x(t-r))\,dW(t), \qquad t > 0, \\ x(t) &= \eta(t), \qquad -r \le t \le 0, \end{aligned}$$

has a unique continuous $(\mathscr{F}_t)_{t \ge 0}$-adapted solution x: $[-r, \infty) \times \Omega \to \mathbf{R}^d$ (cf. [7, 9]). Our main result is the following theorem.

THEOREM 2.1. *Assume Hypotheses* H. *Suppose there exist positive constants* ρ, δ, *an integer* $p \ge 2$ *and a function* ϕ: $[0, \infty) \times \mathbf{R}^d \to \mathbf{R}$ *satisfying the following conditions*:

(i) *For* $(t, x) \in [0, \infty) \times \mathbf{R}^d$,

$$(2.2) \qquad g(t, x)g(t, x)^* \ge \begin{cases} |\phi(t, x)|^p \mathbf{I}, & |\phi(t, x)| < \rho, \\ \delta \mathbf{I}, & |\phi(t, x)| \ge \rho. \end{cases}$$

(ii) $\phi(t, x)$ *is* C^1 *in* t *and* C^2 *in* x, *with bounded first derivatives in* (t, x) *and bounded second derivatives in* $x \in \mathbf{R}^d$.

(iii) *There is a positive constant* c *such that*

$$(2.3) \qquad \|\nabla \phi(t, x)\| \ge c > 0,$$

for all $(t, x) \in [0, \infty) \times \mathbf{R}^d$, *with* $|\phi(t, x)| \le \rho$. *In* (2.3), ∇ *denotes the gradient operator with respect to the space variable* $x \in \mathbf{R}^d$.

(iv) *There is a positive number* δ_0 *such that* $\delta_0 < (3\alpha_\infty)^{-1} \wedge r$ *and for every Borel set* $\mathbf{J} \subseteq [-r, 0]$ *of Lebesgue measure* δ_0 *the following holds*:

$$(2.4) \qquad \int_{\mathbf{J}} \phi(t + r, \eta(t))^2\,dt > 0.$$

Define $s_0 \in [-r, 0]$ *by*

$$s_0 := \sup\left\{ s \in [-r, 0]: \int_{-r}^{s} \phi(u + r, \eta(u))^2\,du = 0 \right\}.$$

Then for all $t > s_0 + r$ *the solution* $x(t)$ *of* (2.1) *is absolutely continuous with respect to the d-dimensional Lebesgue measure and has a* C^∞ *density.*

REMARK. If we take $\alpha_\infty = 0$, we see that Theorem 1 of [2] is a special case of the above theorem.

The remainder of the paper is laid out as follows. In Section 3 we set up the machinery needed to prove Theorem 2.1. This requires the computation of the Malliavin covariance matrix corresponding to the Itô map $W \mapsto x(t)$. This computation is effected by exploiting the delay structure of (2.1) and by *conditioning* on the past history of the solution process x. The problem is thus reduced to the computation of a covariance matrix $C(t)$ arising from a simpler equation of the form

$$(2.5)\qquad \begin{aligned} dy(t) &= H(t,y)\,dt + F(t)\,dW(t), \qquad t > a,\\ y(t) &= x(t), \qquad 0 \le t \le a, \end{aligned}$$

where a is a fixed conditioning time, $F\colon [a,\infty) \to \mathbf{R}^{d\times n}$ is a *deterministic* map and x is a *deterministic* initial path. In the sequel, (2.5) will be referred to as the *conditioned equation*. This equation will be analyzed under the following hypotheses which are analogous to H.

HYPOTHESES H′. (i) *W satisfies $H(i)$.*

(ii) *The maps $F\colon [a,\infty) \to \mathbf{R}^{d\times n}$ and $x\colon [0,a] \to \mathbb{R}^d$ are continuous.*

(iii) *C' is the space of all continuous paths $\xi\colon [0,\infty) \to \mathbf{R}^d$ given the compact-open topology. $H\colon [a,\infty)\times C' \to \mathbf{R}^d$ is a globally bounded continuous map such that for every $t > a$ and $\xi \in C'$, $H(t,\xi)$ depends only on $\{\xi(s)\colon 0 \le s \le t\}$ and has partial Fréchet derivatives of all orders with respect to $\xi \in C([0,t],\mathbf{R}^d)$, which are globally bounded in $(t,\xi) \in (a,\infty)\times C'$. Denote the partial Fréchet derivative of H with respect to ξ by $H_\xi(t,\xi)$. Set*

$$\alpha_t' := \sup\left\{\|H_\xi(u,\xi)\|\colon u \in [a,t],\ \xi \in C\big([0,u],\mathbf{R}^d\big)\right\}, \qquad t > a,$$

where $\|H_\xi(u,\xi)\|$ stands for the operator norm of $H_\xi(u,\xi)\colon C([0,u],\mathbf{R}^d) \to \mathbf{R}^d$.

Our analysis of (2.5) yields the following result.

THEOREM 2.2. *Assume that* (2.5) *satisfies Hypotheses* H′. *Let $t > a$ and let α_t' be defined as in Hypothesis* H′(iii). *Suppose there exists $\delta^* < 1/(3\alpha_t')$ such that*

$$(2.6)\qquad \int_{t-\delta^*}^{t} \mu_1(s)\,ds > 0,$$

where $\mu_1(s)$, $s \ge a$, denotes the smallest eigenvalue of the nonnegative definite matrix $F(s)F(s)^*$. Then the solution $y(t)$ of (2.5) *has an absolutely continuous distribution with respect to d-dimensional Lebesgue measure and has a C^∞ density.*

In the special case when the drift $H(t,y)$ in (2.5) has the form $B(t,y(t))$, for some Lipschitz function $B\colon \mathbf{R}^+\times\mathbf{R}^d \to \mathbf{R}^d$, then y is a (time-inhomogeneous) diffusion process. In this case the methods used to prove Theorem 2.2 yield the following (apparently new) result in partial differential equations:

THEOREM 2.3. *For each $t > 0$, let $A(t) = [a_{ij}(t)]_{i,j=1}^d$ denote a symmetric nonnegative definite $d\times d$ matrix. Let $\mu_2(t)$ be the smallest eigenvalue of $A(t)$. Assume the following*:

(i) *The map $t \mapsto A(t)$ is continuous.*

(ii) *There exists $T > 0$ such that*

$$(2.7)\qquad \int_0^T \mu_2(s)\,ds > 0.$$

(iii) *The functions* b_i, $i = 1, \dots, d$, c: $\mathbf{R}^+ \times \mathbf{R}^d \to \mathbf{R}$ *are bounded, jointly continuous in* (t, x) *and have partial derivatives of all orders in* x, *all of which are bounded in* (t, x).

Let $T_0 := \sup\{T > 0: \int_0^T \mu_2(s)\, ds = 0\}$ *and let* $L_{t,x}$ *denote the differential operator*

$$(2.8) \qquad L_{t,x} := \frac{1}{2} \sum_{j=1}^{d} a_{ij}(t) \frac{\partial^2}{\partial x_i \, \partial x_j} + \sum_{i=1}^{d} b_i(t, x) \frac{\partial}{\partial x_i} + c(t, x).$$

Then the parabolic equation $\partial u / \partial t = L_{t,x} u$ *has a fundamental solution* $\Gamma(t, x, y)$ *defined on* $(T_0, \infty) \times \mathbf{R}^{2d}$, *which is* C^1 *in* t *and* C^∞ *in* (x, y). *Furthermore, if the coefficients* $a_{ij}(t)$, $b_i(t, x)$ *and* $c(t, x)$, $i, j = 1, \dots, d$, *are* C^∞ *in* (t, x) *and*

$$(2.9) \qquad \lim_{t \to T_0^+} (t - T_0) \log\left\{ \int_{T_0}^{t} \mu_2(s)\, ds \right\} = 0,$$

then $\partial / \partial t - L_{t,x}$ *is a hypoelliptic operator on* $(T_0, \infty) \times \mathbf{R}^d$.

We note that the *mean-ellipticity* hypothesis (2.7) is much weaker than the *pointwise-ellipticity* condition which often appears in the classical literature on partial differential equations.

The proof of Theorem 2.1 is completed in Section 4. The crucial part of the argument consists of obtaining a probabilistic lower bound for the Malliavin covariance matrix arising from the conditioned form of (2.1). We obtain this lower bound from the following asymptotic estimate, which we establish for the solution process x in (2.1):

$$(2.10) \qquad P\left(\int_{(t - \delta_0) \vee 0}^{t} |\phi(u, x(u - r))|^2 \, du < \varepsilon \right) = o(\varepsilon^k),$$

as $\varepsilon \to 0+$ for all $t > s_0 + r$ and all $k \geq 1$.

Lemma 4.2 in Section 4 plays a key role in the derivation of (2.10). This lemma enables us to exploit the time delay in (2.1) to propagate condition (2.4) from the initial path η to the segment $\{x(u): t - \delta_0 \leq u \leq t\}$.

The existence of the asymptotic estimate (2.10) is interesting because we have no reason to expect an analogous result to hold for classical diffusions.

3. Analysis of the conditioned equation. Proofs of Theorems 2.2 and 2.3. We begin by introducing the basic terminology and machinery of the Malliavin calculus. Our presentation of this subject is succinct and assumes the results in [2], Section 2. The reader is referred to Bell [1] (also Stroock [11]) for a comprehensive introduction to the Malliavin calculus.

Let L: $D \to L^2(\Omega, \mathbf{R})$ denote the number operator with domain $D \subset L^2(\Omega, \mathbf{R})$, and let $\langle \cdot, \cdot \rangle$: $D \times D \to L^2(\Omega, \mathbf{R})$ be an associated bilinear form, defined as the unique closed bilinear extension of the following form (which is denoted by the same symbol):

$$\langle f_1, f_2\rangle := L(f_1 f_2) - f_1 L(f_2) - f_2 L(f_1), \qquad f_1, f_2 \in D\colon f_1 f_2 \in D.$$

See [11] and [1].

Assume Hypotheses H′ throughout this section. Consider the stochastic differential equation (2.5), where $F\colon [a,\infty) \to \mathbf{R}^{d\times n}$ is a fixed continuous map. This equation has a unique $(\mathscr{F}_t)_{t\geq 0}$-adapted pathwise continuous solution $y \in L^2(\Omega, C([-r,T],\mathbf{R}^d))$ for every $0 < T < \infty$. This solution can be constructed by the standard Picard iteration scheme (Itô and Nisio [6], Theorem 11; cf. Mohammed [9], pages 36–39). Furthermore, if we write $y = (y_1,\dots,y_d)$, then it follows from the work of Kusuoka and Stroock ([7], Lemma 2.9) that, for each $1 \le i \le d$ and $T > a$, the process $(y_i(t), a \le t \le T)$ lies in the class of Malliavin smooth processes R_T introduced in [2], page 79. Let $t > a$ denote a fixed time satisfying the hypothesis of Theorem 2.2. The Malliavin covariance matrix associated with the random variable $y(t)$ is defined to be the random $d \times d$-matrix.

$$C(t) := \left[\langle y_i(t), y_j(t)\rangle\right]_{i,j=1}^{d}. \tag{3.1}$$

Our object is to prove that $y(t)$ admits a smooth density with respect to Lebesgue measure on $\mathbf{R}^d$. As is well known, it suffices to prove that $C(t)$ is a.s. invertible and

$$\det C(t)^{-1} \in \bigcap_{q=1}^{\infty} L^q(\Omega,\mathbf{R}) \tag{3.2}$$

(cf. Stroock [11] and Bell [1], Section 2.3, Theorem 2.5).

We begin by deriving a suitably tractable integral representation for $C(t)$. This will require some further notation. Denote by $K^{(n)}(t) \subset C([0,t],\mathbf{R}^n)$ the n-dimensional Cameron–Martin space. This is the Hilbert space consisting of all absolutely continuous paths $k\colon [0,t] \to \mathbf{R}^n$ such that $k(0) = 0$ and $\int_0^t \|k'(u)\|^2\,du < \infty$. The inner product on $K^{(n)}(t)$ is defined by

$$\langle k_1, k_2\rangle_{K^{(n)}(t)} := \int_0^t k_1'(u)\cdot k_2'(u)\,du, \qquad k_1, k_2 \in K^{(n)}(t),$$

where the dot $(\cdot)$ denotes the Euclidean inner product on $\mathbf{R}^n$.

In the sequel we will identify $K^{(n)}(t)$ with the following subspace of $C([-r,\, t],\mathbf{R}^n)$:

$$\Big\{\xi\colon \xi \in C([-r,t],\mathbf{R}^n),\ \xi(s) = 0 \text{ for all } s \in [-r,0],$$
$$\xi \text{ is absolutely continuous on } [0,t] \text{ and } \int_0^t \|\xi'(s)\|^2\,ds < \infty\Big\}.$$

This identification gives a continuous linear embedding $K^{(n)}(t) \hookrightarrow C([-r,t],\mathbf{R}^n)$. The Cameron–Martin space $K^{(d)}(t) \hookrightarrow C([-r,t],\mathbf{R}^d)$ is defined similarly. In view of this embedding, we will consider $H_\xi(u,z)$ as a continuous linear map $H_\xi(u,z)\colon K^{(d)}(t) \to \mathbf{R}^d$, for each $(u,z) \in [0,t] \times C([-r,t],\mathbf{R}^d)$. Using the identifications $K^{(d)}(t) \equiv K^{(d)}(t)^*$ and $\mathbf{R}^d \equiv \mathbf{R}^{d*}$, we let $H_\xi(u, z)^*$: $\mathbf{R}^d \to K^{(d)}(t)$ represent the corresponding Hilbert space adjoint of $H_\xi(u, z)$.

The precise form of the covariance matrix $C(t)$ in (3.1) is determined by the action of the map Y: $W \mapsto y$ on the subspace $K^{(n)}(t)$, where y is the pathwise solution of (2.5). More specifically, consider the Fréchet smooth map X: $K^{(n)}(t) \mapsto K^{(d)}(t)$, $v \mapsto z$, where z solves the equation

$$(3.3)\qquad \begin{aligned} dz(s) &= \{H(s,z) + F(s)v'(s)\}\,ds, \qquad a < s \le t,\\ z(s) &= x(s), \qquad 0 \le s \le a. \end{aligned}$$

Define the map X_t: $K^{(n)}(t) \to \mathbf{R}^d$ by $X_t := e_t \circ X$, where e_t denotes evaluation at time t. Let X_t^i, $1 \le i \le d$, be the ith component of X_t. If D denotes Fréchet differentiation of C^1 functions on $K^{(n)}(t)$, then for each $\sigma \in K^{(n)}(t)$, we will identify $DX_t^i(\sigma) \in K^{(n)}(t)^*$, with its image in $K^{(n)}(t)$ under the canonical isomorphism $K^{(n)}(t)^* \equiv K^{(n)}(t)$. Following the method developed in [1], Chapter 4, the Malliavin covariance $C(t)$ corresponding to $y(t)$ will be obtained as a stochastic extension of the map $B(t)$: $K^{(n)}(t) \to \mathbf{R}^{d\times d}$ defined by

$$(3.4)\qquad [B(t)(v)]_{ij} := \langle DX_t^i(v), DX_t^j(v)\rangle_{K^{(n)}(t)}, \qquad v \in K^{(n)}(t).$$

Our computation of (3.4) will require the following lemma. In the statement of the lemma and its proof, the dot $(\cdot)$ indicates the unspecified argument of a function.

LEMMA 3.1. *For all $v, k \in K^{(n)}(t)$, the path $DX(v)^*k$ is given by*

$$(3.5)\qquad DX(v)^* k = \int_a^{a\vee\cdot} F(s)^* \sigma(s)\,ds,$$

where $F(s)^*$ denotes the transpose of the matrix $F(s)$ and σ satisfies the Volterra integral equation

$$(3.6)\quad \sigma(s) = k'(s) + \int_s^t \big[H_\xi(u, X(v))^*(\sigma(u))\big]'(s)\,du, \qquad a \le s \le t.$$

PROOF. Let $h, v \in K^{(n)}(t)$ and set $z := X(v)$. Differentiating with respect to v in equation (3.3), we observe that $\eta := DX(v)h$ satisfies the integral equation

$$\eta(s) = \int_a^{a\vee s}\{H_\xi(u,z)\eta + F(u)h'(u)\}\,du, \qquad 0 \le s \le t.$$

Thus for $\rho \in K^{(d)}(t)$,

$$\begin{aligned}\langle \eta, \rho\rangle_{K^{(d)}(t)} &= \int_a^t \{H_\xi(u,z)\eta + F(u)h'(u)\}\cdot\rho'(u)\,du\\ &= \left\langle \eta, \int_a^t H_\xi(u,z)^*\rho'(u)\,du\right\rangle_{K^{(d)}(t)} + \int_a^t h'(u)\cdot\{F(u)^*\rho'(u)\}\,du.\end{aligned}$$

This gives

$$\left\langle \eta, \rho - \int_a^t H_\xi(u,z)^* \rho'(u)\,du \right\rangle_{K^{(d)}(t)} = \left\langle h, \int_a^{a\vee\cdot} F(u)^* \rho'(u)\,du \right\rangle_{K^{(n)}(t)}.$$

Defining $k := \rho - \int_a^t H_\xi(u,z)^* \rho'(u)\,du \in K^{(d)}(t)$ and $\sigma := \rho'$, we have

$$\langle DX(v)h, k\rangle_{K^{(d)}(t)} = \left\langle h, \int_a^{a\vee\cdot} F(u)^* \sigma(u)\,du \right\rangle_{K^{(n)}(t)}, \tag{3.7}$$

where

$$\sigma(s) = k'(s) + \int_a^t \left[H_\xi(u,z)^*(\sigma(u))\right]'(s)\,du, \qquad a \le s \le t. \tag{3.8}$$

The nonanticipating property of H implies that the integrand in (3.8) is zero if $u \le s$. The lemma now follows from (3.7) and (3.8). □

Let $\{e_1,\dots,e_d\}$ denote the standard orthonormal basis of $\mathbf{R}^d$. Lemma 3.1 implies that $[DX_t^i(v)]^*$ is the path

$$\int_a^{a\vee\cdot} F(u)^* \sigma_i(u)\,du, \tag{3.9}$$

where

$$\sigma_i(s) = e_i + \int_s^t \left[H_\xi(u,z)^* \sigma_i(u)\right]'(s)\,du, \qquad a \le s \le t. \tag{3.10}$$

From (3.9) and (3.10), we deduce the following representation for the matrix $B(t)$ in (3.4):

$$B(t) = \int_a^t M(s)F(s)F(s)^* M(s)^*\,ds, \tag{3.11}$$

where M is the $d \times d$ matrix-valued function satisfying the equation

$$M(s) = I + \int_s^t M(u)\left[\{H_\xi(u,z)^*(\cdot)\}'(s)\right]^*\,du, \qquad a \le s \le t. \tag{3.12}$$

In the above equation, the term $[\{H_\xi(u,z)^*(\cdot)\}'(s)]^*$ is the transpose of the linear map

$$\mathbf{R}^d \ni w \mapsto \{H_\xi(u,z)^*(w)\}'(s) \in \mathbf{R}^d$$

regarded as a $d \times d$ matrix with respect to the canonical basis in $\mathbf{R}^d$.

Equations (3.11) and (3.12) yield the following integral representation for the Malliavin covariance $C(t)$ defined in (3.1).

LEMMA 3.2. *Let $C(t)$ be the Malliavin covariance matrix of the solution $y(t)$ of* (2.5). *Then*

$$C(t) = \int_a^t Z(s)F(s)F(s)^* Z(s)^*\,ds, \tag{3.13}$$

where

$$Z(s) = I + \int_s^t Z(u)\left[\{H_\xi(u,y)^*(\cdot)\}'(s)\right]^*\,du, \qquad a \le s \le t. \tag{3.14}$$

PROOF. For each $\omega \in \Omega$, let $\{\omega_m\}_{m=1}^{\infty} \subset K^{(n)}(t)$ denote the sequence of piecewise-linear approximations to ω defined by

$$\omega_m(u) := [(j+1) - mu/t]\,\omega(jt/m) + [mu/t - j]\,\omega((j+1)t/m),$$

for $u \in (jt/m, (j+1)t/m]$, $j = 0, 1, 2, \dots, m-1$, $m \geq 1$.

Then as $m \to \infty$, $W_m \to W$ a.s. in $C([0,t], \mathbf{R}^n)$ and $z_m := X(W_m) \to Y(W) = y$ in the space $C([0,t], \mathbf{R}^d)$ in L^p for every $p \geq 1$. Furthermore, it follows from work of Kusuoka and Stroock [7] (see [2], page 81) that

$$C(t) = \lim_{m \to \infty} B(t)(W_m) \tag{3.15}$$

in L^1. Now replace z in (3.12) by z_m and denote the corresponding solution process by M_m. A standard convergence argument shows that $M_m \to Z$ in the space $C([0,t], \mathbf{R}^{d \times d})$ in L^p for every $p \geq 1$, where Z is as defined in (3.14). The result now follows from (3.11) and (3.15). □

We note that in, general, the matrix-valued process $Z(s)$ in (3.14) *need not be invertible for all values of* $s \in [a,t]$. This reflects, at an analytic level, the non-Markovian character of the solution process y in (2.5). For this reason, it is possible to obtain lower bounds on the integrand in (3.13) only for values of s close to t [where $Z(s)$ *will* be invertible]. The next two lemmas will be used to produce these lower bounds.

LEMMA 3.3. *Fix $u \in [0,t]$ and $\xi \in C([-r,t], \mathbf{R}^d)$. Recall that $\|H_\xi(u,\xi)\|$ denotes the uniform operator norm of the Fréchet derivative $H_\xi(u,\xi) \in L(C([-r,u], \mathbf{R}^d), \mathbf{R}^d)$. Then, for all $v \in \mathbf{R}^d$ and a.e. $s \in [0,t]$, we have*

$$\big\|[H_\xi(u,\xi)^*(v)]'(s)\big\| \leq \big(\|H_\xi(u,\xi)\|\big)\|v\|.$$

PROOF. Let $\theta \in L^2([-r,t], \mathbf{R}^d)$. Define $\phi \in K^{(d)}(t)$ by

$$\phi(s) := \int_0^s \theta(u)\,du, \qquad 0 \leq s \leq t.$$

Then

$$\begin{aligned} \left|\int_0^t [H_\xi(u,\xi)^*(v)]'(s) \cdot \theta(s)\,ds\right| &= |\langle H_\xi(u,\xi)^*(v), \phi\rangle_{K^{(d)}(t)}| \\ &= |v \cdot H_\xi(u,\xi)(\phi)| \\ &\leq \|v\|\big(\|H_\xi(u,\xi)\|\big)\|\theta\|_1, \end{aligned} \tag{3.16}$$

where $\|\theta\|_1$ denotes the L^1-norm

$$\|\theta\|_1 := \int_0^t \|\theta(u)\|\,du$$

on $L^1([0,t], \mathbf{R}^d)$. Define a continuous linear functional μ: $L^2([0,t], \mathbf{R}^d) \to \mathbf{R}$ by

$$\mu(\theta) := \int_0^t [H_\xi(u,\xi)^*(v)]'(s) \cdot \theta(s)\,ds,$$

for all $\theta \in L^2([0,t],\mathbf{R}^d)$. Since $L^2([0,t],\mathbf{R}^d)$ is dense in $L^1([0,t],\mathbf{R}^d)$, it follows from (3.16) that μ has a unique continuous linear extension $\lambda \in [L^1([0,t],\mathbf{R}^d)]^*$ with

$$\|\lambda\| = \|\mu\| \le \left\|\left[H_\xi(u,\xi)^*(v)\right]'(s)\right\| \le \left(\|H_\xi(u,\xi)\|\right)\|v\|.$$

It follows by the duality $[L^1([0,t],\mathbf{R}^d)]^* \equiv L^\infty([0,t],\mathbf{R}^d)$ that $[H_\xi(u,\xi)^*(v)]'(s)$ is essentially bounded in $s \in [0,t]$ and

$$\operatorname{ess\,sup}\left\{\left\|\left[H_\xi(u,\xi)^*(v)\right]'(s)\right\| : s \in [0,t]\right\} = \|\mu\| \le \left(\|H_\xi(u,\xi)\|\right)\|v\|.$$

Hence the lemma is proved. □

LEMMA 3.4. *Let K: $[0,t] \times [0,t] \to \mathbf{R}^{d\times d}$ be a Borel measurable essentially bounded $d \times d$ matrix-valued function with*

$$\|K\|_\infty := \operatorname{ess\,sup}\{\|K(u,s)\| : (u,s) \in [0,t] \times [0,t]\}.$$

Then the matrix-valued integral equation

$$M(s) = I + \int_s^t K(u,s)M(u)\,du, \qquad s \in [0,t], \tag{3.17}$$

has a unique solution $M \in L^\infty([0,t],\mathbf{R}^{d\times d})$. If $0 < t - t_0 < 1/(3\|K\|_\infty)$, then for all $s \in (t_0, t)$, $M(s)$ is invertible and $\|M(s)^{-1}\| \le 2$.

The proof of the above lemma is elementary.

PROOF OF THEOREM 2.2. Fix $t > a$ and let $\delta^* > 0$ be as in the statement of the theorem. Lemmas 3.3 and 3.4 imply that $Z(u)^*$ is invertible for all $u \in (t - \delta^*, t)$ and $\|Z(u)^{*-1}\| \le 2$, where Z is the $d \times d$ matrix-valued process in (3.14). Let λ denote the smallest eigenvalue of the matrix $C(t)$ in (3.13). Using (2.6), we obtain

$$\lambda \ge \tfrac{1}{4}\int_{t-\delta^*}^t \mu_1(u)\,du > 0.$$

Thus $C(t)$ satisfies (3.2) and we conclude that $y(t)$ has a smooth density with respect to d-dimensional Lebesgue measure. □

PROOF OF THEOREM 2.3. Since, for each $t > 0$, the matrix $A(t)$ is symmetric and nonnegative definite, there exists a unique $d \times d$ nonnegative definite symmetric matrix $F(t)$ such that $A(t) = F(t)^2$. Let $T > 0$ be given. Define functions

$$\begin{aligned}\Theta(t) &:= F(T_0 + T - t),\\ \psi(t,x) &:= c(T_0 + T - t, x),\\ h(t,x) &:= \left(b_1(T_0 + T - t, x), \dots, b_d(T_0 + T - t, x)\right)^*,\end{aligned}$$

for all $0 < t < T$ and $x \in R^d$.

Consider the stochastic differential equation

$$(3.18)\quad \begin{aligned} dx^{s,z}(t) &= h(t, x^{s,z}(t))\,dt + \Theta(t)\,dW(t), \qquad s < t < T,\\ x^{s,z}(s) &= z \in \mathbf{R}^d, \end{aligned}$$

where W: $[0,T] \times \Omega \to \mathbf{R}^d$ is the d-dimensional standard Wiener process. Denote by $C_0^\infty(\mathbf{R}^d, \mathbf{R})$ the space of all C^∞-functions ϕ: $\mathbf{R}^d \to \mathbf{R}$, which vanish at infinity and have all derivatives globally bounded. For each $\phi \in C_0^\infty(\mathbf{R}^d, \mathbf{R})$ define V: $[0,T] \times \mathbf{R}^d \to \mathbf{R}$ by

$$V(s,z) := E\left\{\phi(x^{s,z}(T))\exp\left\{\int_s^T \psi(u, x^{s,z}(u))\,du\right\}\right\},$$

for $s \in [0,T]$ and $z \in \mathbf{R}^d$. Define $L'_{t,z} := L_{T_0+T-t,z}$. Then the smoothness of $h(t,\cdot)$, $\psi(t,\cdot)$ and the continuity of ψ imply that V is C^1 in the first variable, C^∞ in the second variable and satisfies the forward equation

$$(3.19)\quad \begin{aligned} \frac{\partial V(s,z)}{\partial s} + L'_{s,z}V(s,z) &= 0, \qquad 0 < s < T,\\ V(T,z) &= \phi(z), \qquad z \in \mathbf{R}^d. \end{aligned}$$

Equation (3.19) is an easy consequence of Itô's formula and the Markov property of the process x in (3.18) (cf. Gihman and Skorohod [5], Theorem 5, page 297). Let $\nu(s,z,t,\cdot)$ denote the family of finite Borel measures

$$(3.20)\quad \nu(s,z,t,B) := \int_{x^{s,z}(t)^{-1}(B)} \exp\left\{\int_s^t \psi(u, x^{s,z}(u))\,du\right\} dP,$$

for $0 < s \le t$, $z \in \mathbf{R}^d$ and any Borel subset B of $\mathbf{R}^d$. Note that the function V can be expressed in terms of these measures by the formula

$$(3.21)\quad V(s,z) = \int_{\mathbf{R}^d} \phi(y)\nu(s,z,T,dy), \qquad 0 < s < T,\ z \in \mathbf{R}^d.$$

Now let $C(s,z,T)$ denote the Malliavin covariance matrix of $x^{s,z}(T)$, for each $0 < s < T$. Applying Lemma 3.2 in the context of (3.18) yields the following integral representation for $C(s,z,T)$:

$$(3.22)\quad C(s,z,T) = \int_s^T Z^{s,z}(u)\Theta(u)\Theta(u)^* Z^{s,z}(u)^*\,du,$$

where

$$Z^{s,z}(u) = I + \int_u^T Z^{s,z}(v) h_x(v, x^{s,z}(v))\,dv, \qquad s \le u \le T.$$

Here h_x denotes the Fréchet derivative of h in the space variable. It is easy to see that for each $s \le u \le T$ the matrix $Z^{s,z}(u)$ is invertible with inverse $Y^{s,z}(u)$ given by the equation

$$Y^{s,z}(u) = I - \int_u^T h_x(v, x^{s,z}(v)) Y^{s,z}(v)\,dv, \qquad s \le u \le T.$$

Since h_x is bounded on $\mathbf{R}^+ \times \mathbf{R}^d$, it follows from this equation that $Y^{s,z}(u)$ is uniformly bounded in $0 \le s \le u \le T$, $z \in \mathbf{R}^d$, $\omega \in \Omega$. Let N denote a deterministic upper bound for $Y^{s,z}(u)$, and let $\lambda(s,z,T)$ denote the smallest eigenvalue of $C(s,z,T)$. From (3.22) we obtain

$$(3.23)\quad \lambda(s,z,T) \ge \frac{1}{N^2}\int_s^T \mu_2(T_0 + T - u)\,du = \frac{1}{N^2}\int_{T_0}^{T_0+T-s} \mu_2(u)\,du,$$

where $\mu_2(u)$ is the smallest eigenvalue of $A(u)$. It follows that for every $s < T$ there exists an interval (s_1, s_2) containing s and a deterministic positive constant δ such that

$$(3.24)\qquad \inf\{\lambda(u,z,T): u \in (s_1,s_2),\, z \in \mathbf{R}^d\} \ge \delta.$$

Using (3.24) and a modification of the proof of Theorem 3.17 in Kusuoka and Stroock [8], we deduce that the measure $\nu(s,z,T,\cdot)$ defined in (3.20) has a density $p(s,z,T,y)$ for each $0 < s < T$, $z \in \mathbf{R}^d$, with the property that $p(s,z,T,y)$ is C^1 in $s \in (0,T)$ and C^∞ in $(z,y) \in \mathbf{R}^d \times \mathbf{R}^d$. We may now write (3.21) in the form

$$(3.25)\qquad V(s,z) = \int_{\mathbf{R}^d} \phi(y)p(s,z,T,y)\,dy.$$

Now define $\Gamma: (T_0, T_0 + T) \times \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ by

$$(3.26)\qquad \Gamma(t,z,y) := p(T_0 + T - t, z, T, y).$$

Note that Γ is C^1 in t and C^∞ in (z,y). We will now show that Γ is a fundamental solution for the parabolic equation $\partial u/\partial t = L_{t,z}u$. First, by an argument of Kusuoka and Stroock ([8], proof of Theorem 3.17), the inequality (3.24) implies that

$$(3.27)\qquad \sup\left\{\left|\frac{\partial P(s,z,T,y)}{\partial s}\right| : s \in (s_1,s_2),\, z,y \in \mathbf{R}^d\right\} < \infty$$

and

$$(3.28)\qquad \sup\{|L'_{s,z}p(s,z,T,y)|: s \in (s_1,s_2),\, z,y \in \mathbf{R}^d\} < \infty$$

for every interval $(s_1,s_2) \subset (0,T)$. Then (3.27) enables us to compute $\partial V(s,z)/\partial s$ by differentiation with respect to s under the integral sign in (3.25). Similarly, by (3.28), we may compute $L'_{s,z}V(s,z)$ by applying $L'_{s,z}$ to the integrand in (3.25). Using (3.19), we then obtain

$$\int_{\mathbf{R}^d} \phi(y)\left\{\frac{\partial p(s,z,T,y)}{\partial s} + L'_{s,z}p(s,z,T,y)\right\}dy = 0,$$

for all $\phi \in C_0^\infty(\mathbf{R}^d, \mathbf{R})$, $z \in \mathbf{R}^d$ and $s \in (0,T)$. This implies

$$\frac{\partial p(s,z,T,y)}{\partial s} = -L'_{s,z}p(s,z,T,y),$$

for all $s \in (0,T)$, $z, y \in \mathbf{R}^d$. In view of the definition of Γ [i.e., (3.26)], the above relation immediately gives

$$\frac{\partial \Gamma(t, z, y)}{\partial t} = L_{t,z}\Gamma(t, z, y), \tag{3.29}$$

for all $t \in (T_0, T_0 + T]$, $y, z \in \mathbf{R}^d$. Thus Γ is a fundamental solution for $\partial u/\partial t = L_{t,z}u$, and the first part of Theorem 2.3 is proved.

To prove the hypoellipticity assertion, we note that (2.9) and (3.23) imply that, for every positive integer q,

$$\lim_{t \to T_0^+} (t - T_0)\|\det C(T_0, z, t)^{-1}\|_q = 0,$$

where $\|\cdot\|_q$ denotes the L_q-norm. It now follows from a straightforward extension of Theorem (8.13) of [8] that $\partial/\partial t - L_{t,z}$ is hypoelliptic on $\mathbf{R}^+ \times \mathbf{R}^d$. □

4. Proof of Theorem 2.1. Throughout this section, we assume the hypotheses and notation of Theorem 2.1. Let $t > 0$ be arbitrary, choose n to be the integer such that $t \in (nr, (n+1)r]$ and denote nr by R. Conditioning (2.1) on $\mathscr{F}_R$, we denote by $C^R(t)$ the covariance matrix generated by the map $\mathbf{W}|[R, t] \mapsto y(t)$, where y is defined by (2.5) with $a = R$ and $F(s) = g(s, x(s - r))$, $s \in [R, t]$. According to the result of Section 3,

$$C^R(t) = \int_R^t Z(s) g(s, x(s-r)) g(s, x(s-r))^* Z(s)^* \, ds, \tag{4.1}$$

where

$$Z(s) = I + \int_s^t Z(u)\left[\{H_\xi(u, x)^*(\cdot)\}'(s)\right]^* du, \qquad R \le s \le t. \tag{4.2}$$

The next result enables us to prove Theorem 2.1 by what may be considered a process of *partial integration by parts* over the Wiener space.

LEMMA 4.1. *Suppose that* $C^R(t) \in \mathrm{GL}(d)$ *and*

$$(\det C^R(t))^{-1} \in \bigcap_{p=1}^{\infty} L^p. \tag{4.3}$$

Then $x(t)$ *in* (2.1) *has a* C^∞ *density with respect to Lebesgue measure on* $\mathbf{R}^d$.

PROOF. Let $n \ge 1$ and let $h_1, \dots, h_n$ denote arbitrary unit vectors in $\mathbf{R}^d$. For any C^∞ test function $\phi\colon \mathbf{R}^d \to \mathbf{R}$,

$$\begin{aligned} |E(D^n\phi(x(t))(h_1, \dots, h_n))| &= \left|E\left[E(D^n\phi(x(t))(h_1, \dots, h_n)|\mathscr{F}_R)\right]\right| \\ &= \left|E\left[E(\phi(x(t))X_{h_1,\dots,h_n}|\mathscr{F}_R)\right]\right| &(4.4)\\ &\le \|\phi\|_\infty E\left[|X_{h_1,\dots,h_n}|\right], &(4.5)\end{aligned}$$

where (4.4) is obtained by integrating by parts with respect to the conditioned Wiener space. Note that $X_{h_1,\dots,h_n}$ is a multilinear form in $C^R(t)^{-1}$ and random variables obtained by a finite number of applications of the operations L and $\langle \cdot, \cdot \rangle$ to the map $W|[R,t] \mapsto y(t)$. The hypotheses of the lemma, together with the fact that H and g have bounded space derivatives of all orders, imply that $E|X_{h_1,\dots,h_n}| < \infty$. Thus (4.5) implies that $x(t)$ has a C^∞ density (cf., e.g., [1], Lemma 1.14). □

For the remainder of this section, we let ϕ: $[0,\infty) \times \mathbf{R}^d \to \mathbf{R}$ denote the function introduced in the statement of Theorem 2.1. Let $\xi(t)$ denote the process $\phi(t + r, x(t))$, $t \geq -r$.

In the sequel, we will adopt the following notation regarding the probability $P(E)$ of an event $E \in \mathscr{F}$. We shall write

$$P(E) = o(\varepsilon^k)$$

if

$$P(E) = o(\varepsilon^k)$$

as $\varepsilon \to 0+$ for every integer $k \geq 1$.

The next result plays a key role in our verification of (4.3).

LEMMA 4.2. *Suppose that, for some* $-r < a < b$, ξ *satisfies*

$$P\left(\int_a^b \xi(t)^2\,dt < \varepsilon\right) = o(\varepsilon^k). \tag{4.6}$$

Then

$$P\left(\int_{a+r}^{b+r} \xi(t)^2\,dt < \varepsilon\right) = o(\varepsilon^k). \tag{4.7}$$

The proof of Lemma 4.2 will require the following preliminary result.

LEMMA 4.3. *Fix* $b > a > 0$. *Let* y: $[0,\infty) \times \Omega \to \mathbf{R}$ *be a measurable stochastic process such that* $E\sup_{a \leq t \leq b}|y(t)|^p$ *is finite for every positive integer* p. *Suppose that* y *satisfies*

$$P\left(\int_a^b y(t)^2\,dt < \varepsilon\right) = o(\varepsilon^k).$$

Then for every positive constant α,

$$P\left(\int_a^b \left\{y(t)^2 \wedge \alpha\right\}dt < \varepsilon\right) = o(\varepsilon^k).$$

PROOF. This result was proved by the authors as Lemma 3 in [2], pages 91–94. We give here another, more elementary proof.

Let A and B denote, respectively, the sets $\{s \in [a,b]: y(s)^2 \le \alpha\}$ and $\{s \in [a,b]: y(s)^2 > \alpha\}$. Then

$$\begin{aligned} P\left(\int_a^b \{y(t)^2 \wedge \alpha\}\,dt < \varepsilon\right) &= P\left(\int_A y(t)^2\,dt + \alpha\lambda(B) < \varepsilon\right) \\ &\le P\left(\int_A y(t)^2\,dt < \varepsilon,\ \alpha\lambda(B) < \varepsilon\right) \\ &= P\left(\int_a^b y(t)^2\,dt < \varepsilon + \int_B y(t)^2\,dt,\ \lambda(B) < \varepsilon/\alpha\right) \\ &\le P_1 + P_2. \end{aligned}$$

Here

$$P_1 := P\left(\int_a^b y(t)^2\,dt < \varepsilon + \int_B y(t)^2\,dt,\ \lambda(B) < \varepsilon/\alpha,\ \sup_{a \le t \le b} y(t)^2 \le 1/\sqrt{\varepsilon}\right)$$

and

$$P_2 := P\left(\sup_{a \le t \le b} y(t)^2 > 1/\sqrt{\varepsilon}\right).$$

Note that

$$P_1 \le P\left(\int_a^b y(t)^2\,dt < \varepsilon + \sqrt{\varepsilon}/\alpha\right) = o(\varepsilon^k)$$

by hypothesis.

Using the finite-moment hypothesis on y and applying the Markov–Chebyshev inequality to the probability P_2, we obtain $P_2 = o(\varepsilon^k)$. This completes the proof of the lemma. □

PROOF OF LEMMA 4.2. Write $g = (g_1 \cdots g_n)$, where g_i, $1 \le i \le n$, are column d-vectors. Computing $\xi(s) = \phi(s + r, x(s))$, $s > 0$, by Itô's formula gives

$$(4.8) \qquad \begin{aligned} d\xi(s) = &\sum_{i=1}^n \nabla\phi(s + r, x(s)) \cdot g_i(s, x(s - r))\,dW_i(s) \\ &+ G(s)\,ds, \qquad s > 0, \end{aligned}$$

where G is a bounded $(\mathscr{F}_s)_{s \ge 0}$-adapted real-valued process. We may write

$$P\left(\int_{a+r}^{b+r} \xi(s)^2\,ds < \varepsilon\right) = P_1 + P_2,$$

where

$$\begin{aligned} P_1 := P\Bigg(&\int_{a+r}^{b+r} \xi(s)^2\,ds < \varepsilon, \\ &\sum_{i=1}^n \int_{a+r}^{b+r} [\nabla\phi(s + r, x(s)) \cdot g_i(s, x(s - r))]^2\,ds \ge \varepsilon^{1/18}\Bigg) \end{aligned}$$

and

$$P_2 := P\left(\int_{a+r}^{b+r} \xi(s)^2\, ds < \varepsilon,\right.$$
$$\left.\sum_{i=1}^{n} \int_{a+r}^{b+r} [\nabla\phi(s+r, x(s)) \cdot g_i(s, x(s-r))]^2\, ds < \varepsilon^{1/18}\right).$$

In view of (4.8), an inequality of Kusuoka and Stroock (cf. [1], Lemma 6.5) implies that $P_1 = o(\varepsilon^k)$. Thus it is sufficient to show that P_2 also has this property.

Write

$$a(s) = \sum_{i=1}^{n} [\nabla\phi(s+r, x(s)) \cdot g_i(s, x(s-r))]^2.$$

Then, by (2.2) and (2.3), it follows that $a(s) \ge c^2(|\xi(s-r)|^p \wedge \delta)$ if $|\xi(s)| \le \rho$.

Define

$$A := \{s \in [a+r, b+r] : |\xi(s)| \le \rho\}$$

and

$$B := \{s \in [a+r, b+r] : |\xi(s)| > \rho\}.$$

Then

$$\begin{aligned}
P_2 &= P\left(\int_{a+r}^{b+r} \xi^2(s)\, ds < \varepsilon, \int_{a+r}^{b+r} a(s)\, ds < \varepsilon^{1/18}\right)\\
&\le P\left(\int_{a+r}^{b+r} \xi^2(s)\, ds < \varepsilon, \int_A c^2(|\xi(s-r)|^p \wedge \delta)\, ds < \varepsilon^{1/18}\right)\\
&= P\left(\int_{a+r}^{b+r} \xi^2(s)\, ds < \varepsilon, \int_{a+r}^{b+r} c^2(|\xi(s-r)|^p \wedge \delta)\, ds\right.\\
&\qquad\qquad \left. < \varepsilon^{1/18} + \int_B c^2(|\xi(s-r)|^p \wedge \delta)\, ds\right)\\
&\le P\left(\int_{a+r}^{b+r} \xi^2(s)\, ds < \varepsilon, \int_{a+r}^{b+r} c^2(|\xi(s-r)|^p \wedge \delta)\, ds < \varepsilon^{1/18} + c^2\delta\lambda(B)\right).
\end{aligned}$$

However, $\int_{a+r}^{b+r} \xi^2(s)\, ds < \varepsilon$ implies $\lambda(B) < \varepsilon/\rho^2$. Thus the preceding probability is

$$\begin{aligned}
&\le P\left(\int_{a+r}^{b+r} c^2(|\xi(s-r)|^p \wedge \delta)\, ds < \varepsilon^{1/18} + c^2\delta\varepsilon/\rho^2\right)\\
&\le P\left(\int_a^b (|\xi(s)|^p \wedge \delta)\, ds \le c'\varepsilon^{1/18}\right)
\end{aligned}$$

for some positive constant c' and for small enough ε. Assumption (4.6), Lemma 4.3 and Jensen's inequality allow us to conclude that the probability on the right-hand side of the above inequality is $o(\varepsilon^k)$. This implies that $P_2 = o(\varepsilon^k)$, and the proof of Lemma 4.2 is complete. □

COMPLETION OF THE PROOF OF THEOREM 2.1. Let s_0 be as defined in Theorem 2.1(iv). Fix $t > s_0 + r$. Then $t \in (nr, (n+1)r]$ for some integer $n \geq 0$. Choose δ_0 so that condition (iv) of the theorem is satisfied. Let λ denote the smallest eigenvalue of $C^R(t)$, where $R = nr$. By Lemmas 3.3 and 3.4 it follows that $Z(u)$ in (4.2) is invertible a.s. for all $u \in [(t-\delta_0) \vee 0, t]$ and $\|Z(u)^{-1}\| \leq 2$. It follows from (4.1) and (2.2) that

$$\lambda \geq \tfrac{1}{4}\int_{(t-\delta_0)\vee 0}^{t} \{|\xi(u-r)|^p \wedge \delta\}\, du. \tag{4.9}$$

We will show that

$$P(\lambda < \varepsilon) = o(\varepsilon^k). \tag{$\star$}$$

Since it is easy to check that ($\star$) implies (4.3), the theorem will then follow from Lemma 4.1.

We break the verification of ($\star$) into the following exhaustive cases:

Case A: $n = 0$. (i) $t > \delta_0$. By hypothesis (iv) of the theorem, we have

$$\int_{t-\delta_0-r}^{t-r} \phi(u+r, \eta(u))^2\, du > 0. \tag{4.10}$$

It follows from the above inequality and the continuity of ϕ and η that

$$\int_{t-\delta_0}^{t} \{|\xi(u-r)|^p \wedge \delta\}\, du > 0. \tag{4.11}$$

From (4.11) and (4.9) we have

$$P(\lambda < \varepsilon) = 0, \tag{4.12}$$

for sufficiently small $\varepsilon > 0$. Thus ($\star$) trivially holds for this case.

(ii) $t < \delta_0$. Then from (4.9)

$$\lambda \geq \tfrac{1}{4}\int_{0}^{t}\{|\phi(u, \eta(u-r))|^p \wedge \delta\}\, du \tag{4.13}$$

a.s. Now, by hypothesis, $t - r > s_0$. Hence $\int_{-r}^{t-r}\phi(u+r, \eta(u))^2\, du > 0$. Together with (4.13), this implies (4.12), and ($\star$) follows as before.

Case B: $n \geq 1$. (i) $t - \delta_0 \geq nr$. Then by hypothesis (iv) of the theorem,

$$\int_{t-\delta_0-(n+1)r}^{t-(n+1)r} \phi(u+r, \eta(u))^2\, du > 0. \tag{4.14}$$

Using Lemma 4.2 we now propagate this condition forward through $(n+1)$ delay periods so as to obtain

$$P\left(\int_{t-\delta_0}^{t} |\xi(u-r)|^2\, du < \varepsilon\right) = o(\varepsilon^k). \tag{4.15}$$

Applying Lemma 4.3, Jensen's inequality and (4.9) to the above inequality, we obtain (★).

(ii) $(n-1)r \le t - \delta_0 < nr$. By hypothesis (iv), we either have

$$\int_{t-\delta_0-nr}^{0} \phi(u+r, \eta(u))^2 \, du > 0 \tag{4.16}$$

or

$$\int_{-r}^{t-(n+1)r} \phi(u+r, \eta(u))^2 \, du > 0. \tag{4.17}$$

In case (4.16) holds, we can propagate it forward through n delay periods using Lemma 4.2. Together with Lemma 4.3, this gives

$$P\left(\int_{t-\delta_0}^{nr} \{|\xi(u-r)|^2 \wedge \delta\} \, du < \varepsilon\right) = o(\varepsilon^k).$$

Thus

$$P\left(\int_{t-\delta_0}^{t} \{|\xi(u-r)|^2 \wedge \delta\} \, du < \varepsilon\right) = o(\varepsilon^k). \tag{4.18}$$

Hence by Jensen's inequality and (4.9), (★) holds. On the other hand, if (4.17) holds, then a similar propagation argument through $(n+1)$ delay periods gives

$$P\left(\int_{nr}^{t} \{|\xi(u-r)|^2 \wedge \delta\} \, du < \varepsilon\right) = o(\varepsilon^k). \tag{4.19}$$

This implies (4.18) and (★) holds in this case too.

The proof of Theorem 2.1 is now complete. □

In conclusion, we note that the conditioning argument used above leads to a very simple proof of the existence of densities for the solution of (2.1) in the special case where gg^* is *uniformly positive definite*. Consider first the zero-drift version of the equation

$$dz(t) = g(t, z(t-r)) \, dW(t), \qquad t > 0,$$

and assume that $gg^* \ge \delta I$, for some $\delta > 0$. Then, conditioned on $\mathscr{F}_{mr}$, the segment $\{z(t),\ mr < t \le (m+1)r\}$ has a Gaussian distribution, with mean $z(mr)$ and a nondegenerate covariance matrix

$$\int_{mr}^{t} g(s, z(s-r)) g(s, z(s-r))^* \, ds.$$

Thus $z(t)$ can be shown to have a smooth density by integrating the conditional Gaussian density. Furthermore, by the Girsanov theorem, the random variable $x(t)$ in (2.1) has a distribution equivalent to that of $z(t)$. It follows that $x(t)$ has a positive density with respect to Lebesgue measure on $\mathbf{R}^d$ (although it is not clear that the density is smooth).

Since Theorem 2.1 allows the matrix function gg^* to have points of degeneracy, the drift term $H(t, x)$ in (2.1) might assume values which lie outside the range of the matrix $g(t, x(t - r))$. Thus the above application of the Girsanov theorem is no longer valid. In this situation the Malliavin calculus appears to be the only technique currently available for establishing the existence of smooth densities.

REFERENCES

[1] BELL, D. R. (1987). *The Malliavin Calculus.* Longman, London.

[2] BELL, D. R. and MOHAMMED, S.-E. A. (1991). The Malliavin calculus and stochastic delay equations. *J. Funct. Anal.* **99** 75–99.

[3] BELL, D. R. and MOHAMMED, S.-E. A. (1993). Opérateurs paraboliques hypoelliptiques avec dégénérescences exponentielles. *C. R. Acad. Sci. Paris Sér. I Math.* **317** 1059–1064.

[4] BELL, D. R. and MOHAMMED, S.-E. A. (1995). An extension of Hörmander's theorem for infinitely degenerate second-order operators. *Duke Math J.* **78** 453–475.

[5] GIHMAN, I. I. and SKOROHOD, A. V. (1981). *Stochastic Differential Equations.* Springer, Berlin.

[6] ITÔ, K. and NISIO, M. (1964). On stationary solutions of a stochastic differential equation. *J. Math. Kyoto Univ.* **4** 1–75.

[7] KUSUOKA, S. and STROOCK, D. W. (1982). Applications of the Malliavin calculus, I. *Taniguchi Sympos. SA Katata* 271–306.

[8] KUSUOKA, S. and STROOCK, D. W. (1985). Applications of the Malliavin calculus, II. *J. Fac. Sci. Univ. Tokyo Sect. 1A Math.* **32** 1–76.

[9] MOHAMMED, S.-E. A. (1984). *Stochastic Functional Differential Equations.* Pitman, London.

[10] MOHAMMED, S.-E. A. (1986). Nonlinear flows of stochastic linear delay equations. *Stochastics* **17** 207–212.

[11] STROOCK, D. W. (1981). The Malliavin calculus, a functional analytic approach. *J. Funct. Anal.* **44** 212–257.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NORTH FLORIDA
JACKSONVILLE, FLORIDA 32224

DEPARTMENT OF MATHEMATICS
SOUTHERN ILLINOIS UNIVERSITY AT CARBONDALE
CARBONDALE, ILLINOIS 62901