A probabilistic interpretation of the divergence and BSDE’s

I.L. Stoica

Université de Bucarest, Faculté de Mathématiques, Str. Academiei 14, Bucharest, Ro 70109, Romania

Received 14 November 2000; received in revised form 18 October 2001; accepted 1 June 2002

Abstract

We prove a stochastic representation, similar to the Feynman–Kac formula, for solutions of parabolic equations involving a distribution expressed as divergence of a measurable field. This leads to an extension of the method of backward stochastic differential equations to a class of nonlinearities larger than the usual one.

MSC: 60K35

Keywords: Backward stochastic differential equations; Forward–backward martingale decomposition; Divergence form elliptic operators; Non-linear parabolic equations; Dirichlet spaces

1. Introduction and preliminaries

1.1. Introduction

The backward stochastic differential equations (shortly BSDE’s) were introduced by Pardoux and Peng (1990) as an extension of the classical Feynman–Kac formula so that to interpret probabilistically a class of nonlinear equations (see the introduction to Pardoux (1996)). Subsequently this point of view led to further insight (see Pardoux and Peng, 1992, 1994; El Karoui, 1997) and the BSDE’s became a useful tool in problems related to nonlinear equations.

The aim of this paper is to extend the method of BSDE’s to a larger class of nonlinear equations. To this end we prove a stochastic integral representation for the...
solution of a parabolic equation involving the divergence of a field. More precisely, let \( u : [0, T] \times \mathbb{R}^N \to \mathbb{R} \) be a solution of the following equation:

\[
(\partial_t + L)u + f - \sum_{ij} \partial_i (a^{ij} g_j) = 0,
\]

where \( L \) is an elliptic divergence form operator

\[
L = \sum_{ij} \partial_i (a^{ij} \partial_j) + \sum_i b^i \partial_i,
\]

and \( f : [0, T] \times \mathbb{R}^N \to \mathbb{R} \), \( g : [0, T] \times \mathbb{R}^N \to \mathbb{R}^N \) are given functions. If \((\Omega, \mathcal{F}, (\mathcal{F}_t), X_t, \theta_t, \mathbb{P}^\cdot)\) is the diffusion process generated by \( L \) in \( \mathbb{R}^N \), then one has (see Theorem 3.2)

\[
u(t, X_t) - \nu(s, X_s) = \sum_i \int_s^t \partial_i \nu(r, X_r) \, dM^i_r - \int_s^t f(r, X_r) \, dr \]

\[
- \frac{1}{2} \int_s^t g * \, dX,
\]

where \( M^i \) represents the martingale part of the component \( X^i_t \) of the process, and the integral denoted with * is a stochastic integral expressed in terms of forward and backward martingales. If \( L \) is symmetric, under the measure \( \mathbb{P}^m \), this integral has the expression

\[
\int_s^t g * \, dX = \sum_i \left( \int_s^t g_i(r, X_r) \, dM^i_r + \int_s^t g_i(r, X_r) \, d\tilde{M}^{m,i}_r \right).
\]

The function \( g \) is in general assumed only measurable so that the term \( \sum_{ij} \partial_i (a^{ij} g_j) \) in the above parabolic equation is a distribution, and so this stochastic integral gives a probabilistic representation for a distribution, in the sense that the solution \( u \) is represented as a stochastic process, in relation (**) , in terms of the function \( f \) and the field \( g \). A corollary of relation (**) is that any solution of (*) admits a uniformly in time quasicontinuous representative (Corollary 3.7).

Now let us suppose that \( f : [0, T] \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \), \( g : [0, T] \times \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N \), are measurable functions satisfying conditions (h1–4) of Section 2. Then a solution \( u \) of the nonlinear equation

\[
(\partial_t + L)u + f(\cdot, \cdot, u, du) - \sum_{ij} \partial_i (a^{ij} g_j(\cdot, \cdot, u, du)) = 0,
\]

with \( du = (\partial_1 u, \ldots, \partial_N u) \), yields a process \( Y_t = u(t, X_t) \) that satisfies the following BSDE:

\[
Y_t = \xi - \sum_i \int_t^T Z_{i,r} \, dM^i_r + \int_t^T f(r, X_r, Y_r, Z_r) \, dr
\]

\[
+ \frac{1}{2} \int_t^T g(r, X_r, Y_r, Z_r) * \, dX,
\]
with \( \zeta = u(T, X_T) \) and \( Z_r = du(r, X_r) \). We prove in Lemma 4.3 (and 4.4) an Ito’s type formula for solutions of BSDE. This was the basic ingredient in the applications so far obtained by using BSDE, but we do not go further in the present work.

The plan of the paper is as follows. In the remainder of this section we describe the general framework and recall some facts needed in the main part. In Section 2 we treat analytically the nonlinear equation relevant to our paper. The method of proof is simply based on the contraction principle and, perhaps, it brings not much novelty for a specialist in nonlinear equations. Section 3 contains the main result, which is relation (**) while Section 4 contains the precise formulation of the corresponding BSDE and its basic stochastic calculus.

1.2. Preliminaries

Let \( a_{ij}, i, j = 1, \ldots, N, \ b_k, \ k = 1, \ldots, N, \) be bounded measurable functions defined in \( \mathbb{R}^N \) such that \( a(x) = (a_{ij}(x)) \) represents a symmetric matrix satisfying the uniform ellipticity condition,

\[
\lambda^{-1} |\xi|^2 \leq \sum_{ij} a_{ij}(x)\xi_i \xi_j \leq \lambda |\xi|^2, \quad x, \zeta \in \mathbb{R}^N.
\]

Since in some concrete cases the matrix \( a \) represents the diffusion coefficients of a diffusion process which is obtained as solution of a stochastic differential equation, we shall assume that it has the form \( a_{ij} = \frac{1}{2} \sum_{k=1}^n \sigma_k^i \sigma_k^j \), where \( x \to \sigma(x) = (\sigma_k^i)(x) \) is a bounded measurable map from \( \mathbb{R}^N \) to the set of all \( N \times n \) real matrices.

We denote by \( L_0 \) the symmetric, divergence form, operator associated to \( a \),

\[
L_0 = \sum_{ij} \partial_i a_{ij} \partial_j.
\]

The main operator we are interested in this paper has the form \( L = L_0 + b \), where \( b \) represents the first order operator corresponding to the vector field \( b(x) = (b^1(x), \ldots, b^N(x)) \), or more explicitly,

\[
L = \sum_{ij} \partial_i a_{ij} \partial_j + \sum_k b_k \partial_k.
\]

The operator \( L \) generates a semigroup \( (P_t)_{t \geq 0} \), which possesses continuous densities \( \{ p_t(x, y), \ t > 0, \ x, y \in \mathbb{R}^N \} \), satisfying the following properties:

1. \( p_t(x, y) > 0 \), everywhere,
2. \( \int p_t(x, y) \, dy = 1, \ x \in \mathbb{R}^N, \)
3. \( \int p_t(x, y) p_s(y, z) \, dy = p_{t+s}(x, z), \)
4. \( p_t(\cdot, \cdot), p_t(\cdot, x) \in H^1, \ \forall t > 0, \ x \in \mathbb{R}^N. \)

This follows by classical results of de Giorgi, Nash, Moser and Aronson (see Aronson, 1968). For a short presentation of the facts needed to treat the Markov process associated to \( L \) see Stroock (1988). The semigroup \( (P_t) \) is Fellerian (i.e. maps the space \( C_0(\mathbb{R}^N) \), of continuous functions vanishing to infinity, into itself).
Let \( X = (\Omega, \mathcal{F}, \mathbb{F}_t, X_t, \theta_t, P^x) \) be the canonical diffusion process associated to \((P_t)\). In particular, the process is conservative and \( \Omega \) is the space of all continuous path defined on \([0, \infty)\) with values in \( \mathbb{R}^N \).

The Lebesgue measure in \( \mathbb{R}^N \), denoted by \( m \), is invariant for the adjoint semigroup \((P^*_t)\). The theory of Dirichlet spaces (see e.g. Fukushima et al., 1994) can be used in our framework and the Sobolev space \( H^1(\mathbb{R}^N) \) is the main space of functions which is involved. The energy form is defined by

\[
E(u;v) = \int \sum_{ij} a_{ij} \partial_i u \partial_j v dm, \quad u,v \in H^1.
\]

We also use the space \( H^1_{\text{loc}} \) of those functions \( u \), such that \( u' \in H^1 \), for any \( u' \in C^1_c \).

For \( \mathbb{R}^N \)-valued functions \( f,g \in L^2(\mathbb{R}^N; \mathbb{R}^N) \) we shall use the notation

\[
\mathcal{A}(f,g) = \int_{\mathbb{R}^N} \sum_{ij} a_{ij} f_i g_j dm.
\]

and \( \mathcal{A}(g) = \mathcal{A}(g,g) \). With this notation, if \( u,v \in H^1 \) and \( du = (\partial_1 u, \ldots, \partial_N u) \), \( dv = (\partial_1 v, \ldots, \partial_N v) \) one writes \( E(u,v) = \mathcal{A}(du,dv) \).

Now let us examine the coordinate martingale parts \( M_i = (M_i^t) \) introduced by Fukushima et al. (1994, p. 246). In our framework these martingales can be defined under each measure \( P^x \), \( x \in \mathbb{R}^N \). (Recall that for general Dirichlet spaces one may have an exceptional polar set.) This is possible e.g. by using Theorem 2 of Lyons and Stoica (1996) (or Corollary 5.3 of Lyons and Stoica (1999); see also Rozkosz (1996) for another, independent, work devoted to the same subject). So, for each \( i = 1, \ldots, N \), we have an \((\mathbb{F}_t)\)-adapted, continuous process \( M_i : [0, \infty) \times \Omega \to \mathbb{R} \) such that

1. \((M_i^t, \mathbb{F}_t, P^x)\) is a martingale for each \( x \in \mathbb{R}^N \),
2. \( M_{i+}^t = M_i^t + M_i^s \circ \theta_t \) a.s.,
3. \( X_i^t - X_0^t - M_i^t \) is locally of zero energy,
4. \( \langle M_i^t \rangle_t = \int_0^t a_{ii}(X_s) ds \),
5. If \( M^u \) is the martingale part in Fukushima decomposition (see Chapter 5 in Fukushima et al., 1994) of \( u(X_t), u \in H^1_{\text{loc}} \), then \( \langle M^t, M^u \rangle_t = \int_0^t \sum_j a_{ij} \partial_j u(X_s) ds \).

Now let \( \mu \) be a fixed probability measure in \( \mathbb{R}^N \) and set \( p^\mu_t(x) = \int p_t(y,x) \mu(dy) \). If \( u \in H^1_{\text{loc}} \) we define

\[
\alpha^\mu_u(s,t) = \int_s^t \sum_{ij} a_{ij} \partial_i u \frac{\partial_j p^\mu_r}{p^\mu_r} (X_r) dr,
\]

\[
\beta^\mu_u(s,t) = \int_s^t \sum_i b_i \partial_i u(X_r) dr.
\]

The martingale part of \( u, (M^\mu_i) \) is defined independent of the initial distribution and represents a continuous (local) martingale additive functional. One has
the representation

\[ M_t^u = \sum_i \int_0^t \partial_i u(X_r) \, dM_t^i, \]

which follows, by approximation, from Corollary 5.6.2 of Fukushima et al. (1994).

Then the process

\[ \overline{M}^{\mu, u}(s, t) = 2(u(X_s) - u(X_t)) + M_t^u - M_s^u - 2\mathcal{X}^{\mu, u}(s, t) + 2\beta^u(s, t) \tag{2} \]

represents a backward local martingale under \( P^\mu \). More precisely, for \( s \in [0, \infty) \) one set \( \mathcal{F}'_s = \sigma(X_r \mid r \in [s, \infty)) \), so defining the “backward” filtration \( (\mathcal{F}'_s, s \in [0, \infty)) \). For fixed \( t \) the above processes are adapted to this filtration with respect to parameter \( s \in [0, t] \) and \( \overline{M}^{\mu, u} \) represents a local martingale under \( P^\mu \). This is proved in Lyons and Stoica (1999, Theorem 3.5) for \( u \in H^1 \) and under the measures \( P^x \), but the same proof gives the result for any measure \( \mu \). The extension to \( u \in H^1_\text{loc} \) follows by standard localization procedures (see also Remark 4.1 in Lyons and Stoica, 1999).

For two functions \( u, v \in H^1_\text{loc} \) we have the bracket relations

\[ \langle M^u, M^v \rangle_r = 2 \int_0^r \sum_{ij} a_{ij} \partial_i u \partial_j v(X_r) \, dr, \tag{3} \]

\[ \langle \overline{M}^{\mu, u}(\cdot, t), \overline{M}^{\mu, v}(\cdot, t) \rangle_s = 2 \int_s^t \sum_{ij} a_{ij} \partial_i u \partial_j v(X_r) \, dr. \tag{4} \]

Observe that \( M^u \) and \( \beta^u \) do not depend on the measure \( \mu \), while \( \mathcal{X}^{\mu, u} \) and \( \overline{M}^{\mu, u} \) do depend, but the brackets of the backward martingales are still independent of \( \mu \).

Since the present paper concerns only the space \( \mathbb{R}^N \) we are changing the more geometric, coordinate independent, notation introduced in Lyons and Stoica (1999). For example, we shall systematically use the coordinate martingales \( M^i \), the processes \( \mathcal{X}^{\mu, i} \), and the backward coordinate martingales \( \overline{M}^{\mu, i} \), associated to the coordinate functions \( u(x) = x^i, i = 1, \ldots, N \).

For a function \( g = (g_1, \ldots, g_N) : [0, T] \times \mathbb{R}^N \to \mathbb{R}^N \) and \( 0 < s \leq t \) we are going to use the notation

\[ \int_s^t g \circ dX = \sum_i \left( \frac{1}{2} \int_s^t g_i(r, X_r) \, dM_r^i + \int_s^t g_i(r, X_r) b^i(X_r) \, dr \right. 
\[ \left. - \int_s^t g_i(r, X_r) \sum_j a^{ij} \partial_j p^{\mu}_{pr}(X_r) \, dr - \frac{1}{2} \int_s^t g_i(r, X_r) \, d\overline{M}^{\mu, i} \right). \]

The integrals with respect to the backward martingales \( \overline{M}^{\mu, i}(r, t) \) are to be understood with respect to the parameter \( r \), performed against the backward adapted integrand.
\[ g_i(r, X_r). \] It may be expressed as a limit of Riemannian sums as follows:
\[
\int_s^t g_i(r, X_r) \, d\tilde{M}^\mu_j = \lim_{\delta(A) \to 0} \sum_{j=0}^k g_i(t_{j+1}, X_{t_{j+1}}) \tilde{M}^\mu_j (t_j, t_{j+1}),
\]
the limit being taken in probability, over a sequence \( (A^n) \) of partitions \( A^n = (s = t_0^n < \cdots < t_{k_n+1} = t) \), with \( \delta(A^n) = \sup_i (t_{i+1}^n - t_i^n) \to 0 \), as \( n \to \infty \). Since one has \( \tilde{M}^\mu_j (s, t) + \tilde{M}^\mu_j (t, l) = \tilde{M}^\mu_j (s, l) \), it is clear that the integral over a fixed interval \([s; t]\) may be taken with any of the backward martingales \( \tilde{M}^\mu_j (, l) \), the result being the same. Also note that, because the sum
\[
\sum_{ij} a_{ij} p^\mu_j (X_r) d \psi_r
\]
which again defines a process independent of the initial distribution. One obviously has
\[
\int_s^t g \circ dX = \sum_i \left( \int_s^t g_i(r, X_r) \, dM^\mu_i + \int_s^t g_i(r, X_r) \, d\tilde{M}^\mu_j \right)
\]
This follows from (2). We shall also employ the notation
\[
\int_s^t g \, dX = \sum_i \left( \int_s^t g_i(r, X_r) \, dM^\mu_i + \int_s^t g_i(r, X_r) \, d\tilde{M}^\mu_j \right) + 2 \int_s^t g_j(r, X_r) \sum_j a_{ij} \partial_j p^\mu_j (X_r) \, dr)
\]
In this paper we are also concerned with evolution equations associated to the parabolic operator \( \partial_t + L \) in a strip of the form \([0, T] \times \mathbb{R}^N\), with \( T > 0 \) a fixed constant. Next we introduce the functional spaces corresponding to \( \partial_t + L \).
We denote by \( H_T \) the space of all functions \( u \in L^2([0, T] \times \mathbb{R}^N) \) such that \( t \to u(t, \cdot) \) is continuous in \( L^2(\mathbb{R}^N) \) on \([0, T]\), \( u(t, \cdot) \in H^1(\mathbb{R}^N) \), for almost every \( t \) and \( \int_0^T \mathcal{E}(u(t, \cdot)) \, dt < \infty \). This is a Banach space with the norm
\[
\|u\|_T = \left( \sup_t \|u(t, \cdot)\|_{L^2(\mathbb{R}^N)}^2 + \int_0^T \mathcal{E}(u(t, \cdot)) \, dt \right)^{1/2}.
\]
The space \( \mathcal{C}_T = \mathcal{C}_\infty ([0, T]) \otimes \mathcal{C}_c^{\infty} (\mathbb{R}^N) \) is dense in \( H_T \), as one may directly verify. If \( f \in L^2([0, T] \times \mathbb{R}^N) \) and \( g \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N) \) we shall consider linear evolution equations of the form,
\[
(\partial_t + L)u + f - \sum_{ij} \partial_i a^{ij} g_j = 0
\]
with terminal condition \( u(T, \cdot) = \Phi \), where \( \Phi \) is a given function in \( L^2(\mathbb{R}^N) \). We say that \( u \in H_T \) is a (weak) solution of Eq. (5) if the following relation holds for any test function \( \varphi \in C_T \), and \( t \in [0, T] \),

\[
\int_t^T \left[ (u(s, \cdot), \partial_s \varphi(s, \cdot)) + \mathcal{E}(u(s, \cdot), \varphi(s, \cdot)) - (bu(s, \cdot), \varphi(s, \cdot)) \right] ds = \int_t^T \left[ (f(s, \cdot), \varphi(s, \cdot)) + \mathcal{A}(g(s, \cdot), \varphi(s, \cdot)) \right] ds + (\Phi, \varphi(T, \cdot)) - (u(t, \cdot), \varphi(t, \cdot)).
\]

We are using here the notation \( d\varphi(t,x) = (\partial_1 \varphi(t,x), \ldots, \partial_N \varphi(t,x)) \) and the same notation will also be used in the sequel for functions in \( H_T \).

2. Analytical treatment of a nonlinear equation

In this section we prove existence and uniqueness of solutions for a certain nonlinear equation which will be probabilistically interpreted in the ensuing section. The nonlinear terms of the equation are given by the functions \( f : [0, T] \times \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R} \) and \( g = (g_1, \ldots, g_N) : [0, T] \times \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^N \) which are assumed to be measurable and fulfill the following conditions

\[
\begin{align*}
(f(\cdot, \cdot, 0, 0) &\in L^2([0, T] \times \mathbb{R}^N), \\
|f(t, x, y, z) - f(t, x, y', z')| &\leq C(|y - y'| + |z - z'|), \\
g(\cdot, \cdot, 0, 0) &\in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N), \\
|g(t, x, y, z) - g(t, x, y', z')|_a &\leq C|y - y'| + \alpha|z - z'|,
\end{align*}
\]

with some constant \( C > 0 \) and \( \alpha \in (0, 1) \). In relations \( (h2) \) and \( (h4) \) the variables are arbitrary in the domain of definition of the functions, i.e. \( t \in [0, T], \ x \in \mathbb{R}^N, \ y, y' \in \mathbb{R} \) and \( z, z' \in \mathbb{R}^n \). The modulus \( |.|_a \) is defined for a function \( u = (u_1, \ldots, u_N) : \mathbb{R}^N \rightarrow \mathbb{R}^N \) by

\[
|u|_a(x) = |u(x)|_{a(x)} = \left( \sum_{ij} a^{ij}(x)u_i(x)u_j(x) \right)^{1/2}, \quad x \in \mathbb{R}^N,
\]

so that in relation \( (h4) \) one has a norm that depends on \( x \), which is uniformly dominated by a quantity independent of \( x \).

The nonlinear evolution equation we will consider has the form

\[
(\partial_t + L_0)u(t,x) + f \left( t, x, u(t,x), \frac{1}{\sqrt{2}} d\mu(t,x)\sigma(x) \right) - \sum_{ij} \partial_i a^{ij} g_j \left( t, x, u(t,x), \frac{1}{\sqrt{2}} d\mu(t,x)\sigma(x) \right) = 0.
\]
A weak solution of Eq. (6) with final condition \( u(T,x) = \Phi(x) \), \( \Phi \in L^2(\mathbb{R}^N) \) is a function \( u \in H_T \) verifying the following relation:

\[
\int_t^T \left[ (u(s,\cdot), \partial_s \varphi(s,\cdot)) + \mathcal{E}(u(s,\cdot), \varphi(s,\cdot)) \right] ds
\]

\[
= \int_t^T \left[ \left( f \left( s, \cdot, u(s,\cdot), \frac{1}{\sqrt{2}} du(s,\cdot)\sigma(\cdot) \right), \varphi(s,\cdot) \right) + \mathcal{A} \left( g \left( s, \cdot, u(s,\cdot), \frac{1}{\sqrt{2}} du(s,\cdot)\sigma(\cdot) \right), d\varphi(s,\cdot) \right) \right] ds
\]

\[
+ (\Phi, \varphi(T,\cdot)) - (u(t,\cdot), \varphi(t,\cdot)),
\]

for each \( \varphi \in \mathcal{C}_T \) and \( t \in [0,T] \).

**Remark 2.1.** (1) The way that the derivatives of \( u \) enter in Eq. (6), that is the expression

\[
\frac{1}{\sqrt{2}} du(t,x)\sigma(x) = \left( \frac{1}{\sqrt{2}} \sum_{i=1}^N \sigma_i^j(x) \partial_j u(t,x), \ldots, \frac{1}{\sqrt{2}} \sum_{i=1}^N \sigma_n^j(x) \partial_j u(t,x) \right),
\]

may appear strange at first look. As a consequence of this writing however, it becomes more comfortable to write the Lipschitz conditions (h2) and (h4); they are as follows:

\[
\left| f \left( t,x,u_1, \frac{1}{\sqrt{2}} du_1\sigma \right) - f \left( t,x,u_2, \frac{1}{\sqrt{2}} du_2\sigma \right) \right| \leq C (|u_1(t,x) - u_2(t,x)| + |du_1(t,x) - du_2(t,x)|_{a(x)}),
\]

\[
\left| g \left( t,x,u_1, \frac{1}{\sqrt{2}} du_1\sigma \right) - g \left( t,x,u_2, \frac{1}{\sqrt{2}} du_2\sigma \right) \right|_{a(x)} \leq C |u_1(t,x) - u_2(t,x)| + 2 |du_1(t,x) - du_2(t,x)|_{a(x)}.
\]

(2) The conditions (h1–4), fulfilled by \( f \) and \( g \), imply that the composed functions

\[
f_u(t,x) = f \left( t,x,u(t,x), \frac{1}{\sqrt{2}} du(t,x)\sigma(x) \right),
\]

\[
g_u(t,x) = g \left( t,x,u(t,x), \frac{1}{\sqrt{2}} du(t,x)\sigma(x) \right)
\]

belong to the spaces \( L^2([0,T] \times \mathbb{R}^N) \), respectively, \( L^2([0,T] \times \mathbb{R}^N; \mathbb{R}^N) \). Therefore, once we have a solution of (6), it may be viewed as a solution of the linear equation (5) with \( b \equiv 0 \), and \( f = f_u \), \( g = g_u \) as functions of \( t \) and \( x \) only.
(3) Eq. (5) may also be viewed as a particular case of (6). To see this, just take \( \tilde{b} \) such that \( b = a \tilde{b} \), that is \( b' = \sum_j a^{ij} \tilde{b}^j \) and use the function \( f(t,x) \) appearing in (5) to define another function, \( f(t,x,y,z) := f(t,x) + 1/\sqrt{2} \sum_{k=1}^N \sum_{j=1}^N \sigma_k^j \tilde{b}^j z_k \). So Eq. (5) takes the form (6) with this new function \( f(t,x,y,z) \) and \( g(t,x,y,z) = g(t,x) \).

(4) In our analytical treatment the drift term appears simply as a perturbation of \( L_0 \). For this reason, in the present section, we take Eq. (6) so that it contains the drift in a nonexplicit form, absorbed in \( f \) as explained above.

Next we are going to treat the linear equation obtained as a particular case of (6) when \( f \) and \( g \) do not depend on \( y \) and \( z \).

**Lemma 2.2.** If \( f \) and \( g \) are independent of \( y \) and \( z \) (according to (h1) and (h3) this implies \( f \in L^2([0,T] \times \mathbb{R}^N) \) and \( g \in L^2([0,T] \times \mathbb{R}^N; \mathbb{R}^N) \)) then Eq. (6) has a unique solution and this solution verifies the following relations:

\[
\begin{align*}
(i) \quad & \frac{1}{2} \|u(t,\cdot)\|_2^2 + \int_t^T \mathcal{E}(u(s,\cdot)) \, ds \\
& = \int_t^T \left[ (f(s,\cdot),u(s,\cdot)) + \mathcal{A}(g(s,\cdot),du(s,\cdot)) \right] \, ds + \frac{1}{2} \|\Phi\|_2^2 \\
(ii) \quad & \|u\|_T^2 \leq e^T(\|\Phi\|_2^2 + \|f\|_{L^2([0,T] \times \mathbb{R}^N)}^2 + \|g\|_{L^2([0,T] \times \mathbb{R}^N)}^2)
\end{align*}
\]

**Proof.** Set \( f^i = \sum_j a^{ij}g_j \) and take sequences \((f_k^i)\) in \( C_T \) such that \( f_k^i \to f^i \) in \( L^2([0,T] \times \mathbb{R}^N) \), \( i = 1, \ldots, N \). Then solve for each \( k \) the equation

\[
(\partial_t + L_0)u_k + f - \sum_i \partial_i f_k^i = 0
\]

with final condition \( u_k(T,\cdot) = \Phi \). By relation (8) of Proposition 4 in Bally et al. (in preparation) one also has

\[
\frac{1}{2} \|u_k(t,\cdot) - u_p(t,\cdot)\|_2^2 + \int_t^T \mathcal{E}(u_k(s,\cdot) - u_p(s,\cdot)) \, ds
\]

\[
= \int_t^T \int_{\mathbb{R}^N} \sum_i (f_k^i(s,x) - f_p^i(s,x)) \partial_i (u_k(s,x) - u_p(s,x)) \, dx \, ds.
\]

This relation shows that \( u_k \) is Cauchy in \( H_T \). Its limit is the desired solution, proving the existence. Proposition 4 in Bally et al. (in preparation) ensure uniqueness. Relation (i) in the statement is obtained first for the approximating solutions, \( u_k \), and then one pass to the limit. Relation (ii) is obtained from Gronwall’s lemma. \( \Box \)

Before stating Proposition 2.4, which gives complete information about solutions in \( H_T \) of Eq. (6) with general \( f \) and \( g \), we are going to prove a technical lemma which ensures applicability of the contraction principle.
Lemma 2.3. Let \( u_1, u_2 \in H_T \) and set \( f_k(t, x) = f(t, x, u_k, (1/\sqrt{2}) du_k \sigma), g_k(t, x) = g(t, x, u_k, (1/\sqrt{2}) du_k \sigma), k = 1, 2. \) For a given function \( \Phi \in L^2(\mathbb{R}^N) \) denote by \( v_k \) the solution of the linear equation

\[
(\partial_t + L_0)v_k + f_k - \sum_{ij} \partial_i a_{ij}^k g_{k,j} = 0,
\]

with boundary condition \( v_k(T, \cdot) = \Phi, k = 1, 2. \) There exist two constants \( \varepsilon > 0 \) and \( \theta \in (0, 1), \) which depend only on the constants \( C \) and \( \alpha \) in \((h2)\) and \((h4), \) such that the following inequality holds for each \( t \in [T - \varepsilon, T]:\)

\[
\sup_{s \in [t, T]} \|v_1(s, \cdot) - v_2(s, \cdot)\|^2_2 + \int_t^T \mathcal{E}(v_1(s, \cdot) - v_2(s, \cdot)) \, ds
\leq \theta \left( \sup_{s \in [t, T]} \|u_1(s, \cdot) - u_2(s, \cdot)\|^2_2 + \int_t^T \mathcal{E}(u_1(s, \cdot) - u_2(s, \cdot)) \, ds \right).
\]

Proof. One starts with relation (i) of the preceding lemma applied to \( v_1 - v_2, \) viewed as a solution of the corresponding linear equation,

\[
\frac{1}{2} \|v_1(t, \cdot) - v_2(t, \cdot)\|^2_2 + \int_t^T \mathcal{E}(v_1(s, \cdot) - v_2(s, \cdot)) \, ds
= \int_t^T (f_1(s, \cdot) - f_2(s, \cdot), v_1(s, \cdot) - v_2(s, \cdot)) \, ds
+ \int_t^T A_g(g_1(s, \cdot) - g_2(s, \cdot), dv_1(s, \cdot) - dv_2(s, \cdot)) \, ds.
\]

The first term in the right-hand side is majorized by using the Lipschitz condition \((h2)\) and then some elementary inequalities, obtaining

\[
\leq \frac{C}{2} \int_t^T \|u_1(s, \cdot) - u_2(s, \cdot)\|^2_2 \, ds + \frac{C}{2} \int_t^T \|v_1(s, \cdot) - v_2(s, \cdot)\|^2_2 \, ds
+ \frac{C \delta}{2} \int_t^T \mathcal{E}(u_1(s, \cdot) - u_2(s, \cdot)) \, ds + \frac{C \delta}{2} \int_t^T \|v_1(s, \cdot) - v_2(s, \cdot)\|^2_2 \, ds.
\]

The second term in the right-hand side of the above equality is similarly majorized by using \((h4), \) then getting

\[
\leq \frac{C}{2 \delta^2} \int_t^T \|u_1(s, \cdot) - u_2(s, \cdot)\|^2_2 \, ds + C \delta' \int_t^T \mathcal{E}(v_1(s, \cdot) - v_2(s, \cdot)) \, ds
+ \frac{\alpha}{2} \int_t^T \mathcal{E}(u_1(s, \cdot) - u_2(s, \cdot)) \, ds + \frac{\alpha}{2} \int_t^T \mathcal{E}(v_1(s, \cdot) - v_2(s, \cdot)) \, ds.
\]
Choosing $\delta$ and $\delta'$ such that $2C\delta + \alpha = 1$ and $C\delta' + \alpha = 1$ one gets

$$
\left\|v_1(t, \cdot) - v_2(t, \cdot)\right\|^2_2 + \int_t^T \mathcal{E}(v_1(s, \cdot) - v_2(s, \cdot)) \, ds
\leq C \left(1 + \frac{1}{\delta'}\right) \int_t^T \left\|u_1(s, \cdot) - u_2(s, \cdot)\right\|^2_2 \, ds + \frac{1 + \alpha}{2} \int_t^T \mathcal{E}(u_1(s, \cdot) - u_2(s, \cdot)) \, ds
+ C \left(1 + \frac{1}{\delta}\right) \int_t^T \left\|v_1(s, \cdot) - v_2(s, \cdot)\right\|^2_2 \, ds.
$$

By Gronwall’s lemma and a suitable choice of $\varepsilon$ one gets the inequality in the statement.

**Proposition 2.4.** Under the assumptions (h1–4), to any function $lBS \in L^2(\mathbb{R}^N)$ one associates a uniquely determined solution $u \in HT$ of Eq. (6), satisfying the terminal condition $u(T, \cdot) = lBS$. Moreover the solution satisfies the following estimate:

$$
\left\|u\right\|^2_T \leq K \exp KT \left(\|\Phi\|^2_T + \|f(\cdot, \cdot, 0, 0)\|_{L^2([0,T] \times \mathbb{R}^N)} + \|g(\cdot, \cdot, 0, 0)\|_{L^2([0,T] \times \mathbb{R}^N)}\right),
$$

where $K$ is a constant which depends only on $C$ and $\alpha$.

**Proof.** By the preceding lemma one may use the contraction principle and get existence and uniqueness of solution on the interval $[T - \varepsilon, T]$. Namely, one consider the space $H(T - \varepsilon, T)$ which consists of functions $u \in L^2([T - \varepsilon, T] \times \mathbb{R}^N)$ such that $t \rightarrow u(t, \cdot)$ is continuous in $L^2(\mathbb{R}^N)$, $u(t, \cdot) \in H^1$ for almost every $t$, and

$$
\int_{T - \varepsilon}^T \mathcal{E}(u(t, \cdot)) \, dt < \infty.
$$

An operator $A : H(T - \varepsilon, T) \rightarrow H(T - \varepsilon, T)$ is defined as follows: for $u \in H(T - \varepsilon, T)$ one takes $Au$ to be solution of equation

$$
(\partial_t + L_0)Au + f (\cdot, \cdot, u, \frac{1}{\sqrt{2}} \, du \, \sigma) - \sum_{ij} \partial_i \left( a^{ij} g_j (\cdot, \cdot, u, \frac{1}{\sqrt{2}} \, du \, \sigma)\right) = 0
$$

over the interval $[T - \varepsilon, T]$ with condition $Au(T, \cdot) = \Phi$, the given function $\Phi$. This operator is a contraction with $\|A\| \leq \theta < 1$. Repeating the argument, over the interval $[T - 2\varepsilon, T - \varepsilon]$ and so on, one succeeds to extend the existence and uniqueness over the interval $[0, T]$. The estimate in the statement is obtained writing relation (i) of Lemma 2.2, using then the Lipschitz conditions, and finally applying again Gronwall’s lemma.

3. **Stochastic integral representation of solutions in divergence form**

The aim of this section is to prove Theorem 3.2 from below, which gives the stochastic representation for solutions of Eq. (5).
Lemma 3.1. Let \( g \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N) \), \( f \in L^2([0, T] \times \mathbb{R}^N) \) satisfy the following relation in the weak sense:

\[
\sum_{ij} \partial_i (a_{ij} g_j) = f.
\]

(This means that for each \( \varphi \in \mathcal{C}_T \) one has

\[
- \int_0^T \int_{\mathbb{R}^N} \sum_{ij} a_{ij} (\partial_i \varphi) g_j = \int_0^T \int_{\mathbb{R}^N} \varphi f.
\]

Then, the following equality holds a.s.:

\[
\int_s^t g * dX = -2 \int_s^t f(r, X_r) \, dr.
\]

Proof. First we are going to show that

\[
\sum_i \int_s^T g_i(r, X_r) \, dM_r^i + 2 \int_s^T \sum_{ij} a_{ij} \frac{\partial_i P_r^\mu}{P_r^\mu} g_j(r, X_r) \, dr + 2 \int_s^T f(r, X_r) \, dr
\]

is a \((\mathcal{F}_s')\) (backward) martingale under \(P^\mu\). Since the reversed process is Markovian, in order to check this property it suffices to show that the following conditional expectation vanishes, for each \(0 < s < T\),

\[
E^\mu \left[ \sum_i \int_s^t g_i(r, X_r) \, dM_r^i + 2 \int_s^t \sum_{ij} a_{ij} \frac{\partial_i P_r^\mu}{P_r^\mu} g_j(r, X_r) \, dr + 2 \int_s^t f(r, X_r) \, dr \right] = 0.
\]

Let \( \varphi \in \mathcal{C}_c(\mathbb{R}^N) \) and set \( u(r, x) = P_{t-r} \varphi(x) \). Then one has \( u(t, x) = \varphi(x) \), of course. Writing \( \varphi(X_t) = u(t, X_t) \) as a martingale

\[
u(t, X_t) = u(s, X_s) + \sum_i \int_s^t \partial_i u(r, X_r) \, dM_r^i,
\]

and using Itô’s formula one deduces

\[
E^\mu \left( \sum_i \int_s^t g_i(r, X_r) \, dM_r^i + 2 \int_s^t \sum_{ij} a_{ij} \frac{\partial_i P_r^\mu}{P_r^\mu} g_j(r, X_r) \, dr + 2 \int_s^t f(r, X_r) \, dr \right) \varphi(X_t)
\]

\[
= E^\mu \left( 2 \int_s^t u(r, X_r) \sum_{ij} a_{ij} \frac{\partial_i P_r^\mu}{P_r^\mu} g_j(r, X_r) \, dr \right)
\]
\[
+ 2 \int_s^t u f(r, X_r) \, dr + 2 \int_s^t \sum_{ij} a^{ij}(\partial_i u) g_j(r, X_r) \, dr
\]

\[
= 2 \int_s^t \int_{\mathbb{R}^N} \sum_{ij} a^{ij}(\partial_i (u p^\mu_i)) g_j \, dr + 2 \int_s^t \int_{\mathbb{R}^N} u p^\mu_i f \, dr.
\]

Since \( w(r, x) = u(r, x) p_r(x) \) is in the closure of \( \mathcal{C}_c^\infty([s, t]) \otimes \mathcal{C}_c(\mathbb{R}^N) \) with respect to the norm
\[
\left( \sup_{r \in [s, T]} \| w(r, \cdot) \|_2^2 + \int_s^T \mathcal{E}(w(r, \cdot)) \, dr \right)^{1/2},
\]
the assumption in the statement implies that the preceding expression is null. We have proved that \((\ast)\) is a backward martingale. Adding to this the next backward martingale
\[
\sum_i \int_s^T g_i(r, X_r) \, dM_r,
\]
one gets the following process:
\[
\int_s^T g * \, dX + 2 \int_s^T f(r, X_r) \, dr.
\]
By Lemma 3.3 from below it has zero quadratic variation. Since a martingale with zero quadratic variation is null, it follows that this process vanishes, which leads to the equality asserted in the statement. \( \square \)

**Theorem 3.2.** Let \( u \in H_T \) be a solution of the equation
\[(\partial_t + L)u + f - \sum_{ij} \partial_i (a^{ij} g_j) = 0,\]
where \( g \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N) \), \( f \in L^2([0, T] \times \mathbb{R}^N) \) and the final condition is \( \Phi \in L^2(\mathbb{R}^N) \). Then, for any \( 0 \leq s \leq t \leq T \), one has the following stochastic representation, \( P^m \)-a.s.,
\[
\begin{array}{c}
\sum_i \int_s^t \hat{\partial}_i u(r, X_r) \, dM_r^i - \int_s^t f(r, X_r) \, dr - \frac{1}{2} \int_s^t g * \, dX.
\end{array}
\]

**Proof.** First we note that the case with \( g \equiv 0 \) was proved in Bally et al. (in preparation) (Proposition 8 and Corollary 10) for symmetric processes, but the same proof works for non-symmetric processes like in our framework. So, the problem reduces to the case \( g \neq 0 \), \( f \equiv 0 \) and we shall only treat this case in what follows. The difficulty is that, because of lack of smoothness of \( a^{ij} g_j \), function \( u \) is also not smooth enough, and so we cannot apply Ito’s formula to \( u(t, X_t) \). To overcome this we are going to make some changes of functions and then an approximation with smoother functions.

By Lemma 3.5 from below we have a function \( h \in L^2([0, T] \times \mathbb{R}^N) \) such that \( h(t, \cdot) \in H^1, \, dt \)-almost everywhere, \( dh = (\partial_1 h, \ldots, \partial_N h) \in L^2([0, T] \times \mathbb{R}^N) \) and \( \sum_{ij} \hat{\partial}_i (a^{ij} g_j) = \cdots \)
\[
\sum_{ij} \partial_i (a^{ij} \partial_j) - h, \text{ in the weak sense. Therefore } u \text{ also solves the equation}
\]
\[
(\partial_t + L)u + h - \sum_{ij} \partial_i (a^{ij} \partial_j) = 0.
\]
Assume for the moment that we have proved the representation corresponding to this equation, namely assume that one has
\[
u(t, X_t) - u(s, X_s) = \sum_i \int_s^t \partial_i u(r, X_r) \, dM^i_r - \int_s^t h(r, X_r) \, dr - \frac{1}{2} \int_s^t dh * dX. (\ast)
\]
By Lemma 3.1 we have
\[
\int_s^t g * \, dX = \int_s^t dh * dX + 2 \int_s^t h(r, X_r) \, dr,
\]
which implies the representation from the statement. It remains to prove relation (\ast). To this end we approximate \( h \) with functions from the space \( C^\infty([0, T]) \otimes \mathcal{D}(L_0) \) with respect to the norm \( (\int_0^T (\|h(t, \cdot)\|_2^2 + E(h(t, \cdot))) \, dt)^{1/2} \). By Proposition 2.4 the corresponding solutions converge and if relation (\ast) holds for approximands it clearly passes to the limit, on account of the bracket relations (3), (4). Thus now it remains to prove the representation (\ast) in the case \( h \in C^\infty([0, T]) \otimes \mathcal{D}(L_0) \). In this case we have \( \sum_{ij} \partial_i (a^{ij} \partial_j) = L_0 h \in L^2([0, T] \times \mathbb{R}^N) \), so that \( u \) is a solution of
\[
(\partial_t + L)u + h - L_0 h = 0.
\]
Then we know the representation
\[
u(t, X_t) - u(s, X_s) = \sum_i \int_s^t \partial_i u(r, X_r) \, dM^i_r - \int_s^t (h - L_0 h)(r, X_r) \, dr.
\]
On the other hand, from Lemma 3.1 it follows
\[
\int_s^t L_0 h(r, X_r) \, dr = - \frac{1}{2} \int_s^t dh * dX,
\]
which leads to relation (\ast), completing the proof. \( \square \)

In general the process \( X \) is not a semimartingale because of non-smoothness of the coefficients \( a^{ij} \). It is interesting to note that in fact the \( * \)-integral essentially contains the noise produced by the discontinuities of these coefficients. To be more precise, if this matrix would were differentiable, one could have written
\[
X^i_t - X^i_0 - M^i_t = \int_0^t Lx^i(X_s) \, ds = \delta^i_t + \beta^i_t,
\]
with \( \beta^i_t = \int_0^t b^i(X_s) \, ds, \delta^i_t = \int_0^t \sum_j \partial_j a^{ij}(X_s) \, ds. \) This suggests us to introduce the following notation, in the general case of non-smooth coefficients,
\[
\delta^i_t = X^i_t - X^i_0 - M^i_t - \beta^i_t.
\]
Then clearly one has
\[ \delta_i^t = -\frac{1}{2}(M^t + \tilde{M}(0, t)) - \frac{1}{2} \int_0^t \epsilon^i * dX, \]
where \(\epsilon^i = (0, \ldots, 1, 0, \ldots, 0)\) is a constant vector with 1 only on the \(i\)th component. Now if \(g = (g_i)\) has the components in the class \(\mathcal{C}_c\), we may express the Stratonovich integral
\[ \int_s^t g_i(r, X_r) \circ dM_r^i = \int_s^t g_i(r, X_r) dM_r^i - \int_s^t \sum_j a^{ij} \partial_j g_i(r, X_r) dr. \]
Further we may write in terms of \(\delta\) the integral denoted with circle in the preliminaries
\[ \int_s^t g_i(r, X_r) \circ dX_r^i = \int_s^t g_i(r, X_r) dM_r^i - \frac{1}{2} \int_s^t g_i(r, X_r) \circ d\tilde{M}^i - \int_s^t \sum_i \int_s^t g_i(r, X_r) \circ d\delta_r^i, \]
and from this one deduces
\[ \frac{1}{2} \int_s^t g * dX = -\int_s^t \sum_{i,j} a^{ij} \partial_j g_i(r, X_r) dr - \sum_i \int_s^t g_i(r, X_r) \circ d\delta_r^i. \]
This shows that the *-integral contains the noise introduced by the non-smooth part of the matrix \((a^{ij})\). If \(g\) is only measurable, it is no more possible to obtain such a decomposition. Note that since \(\delta\) has zero quadratic variation, the Stratonovich integral appearing above is in fact a usual stochastic integral.

3.1. Lemmas used in the proof of the theorem

Let us recall the definition of a zero quadratic variation process. If \(A = (A_t)_{t \in [a,b]}\) is a real valued process defined on a probability space \((\Omega, \mathcal{F}, P)\) and \(\Delta = (a = t_0 < \cdots < t_{k+1} = b)\) is a partition of \([a,b]\), then we set
\[ V^2(A, \Delta) = \sum_{i=0}^k (A_{t_{i+1}} - A_{t_i})^2. \]
The process \(A\) is said to have zero quadratic variation if, for any sequence of partitions \((A_k)\) with diameters tending to zero, one has
\[ \lim_{k \to \infty} V^2(A, A_k) = 0, \quad \text{a.s.} \]
Lemma 3.3. If \( g \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N) \) and \( s > 0 \) is fixed, then the process \( \int_s^t g \ast dX \), \( t \in [s, T] \) has zero quadratic variation.

Proof. Observe first that, if \( u \in H^1 \) and \( f \) is such that \(|du|f \in L^2([0, T] \times \mathbb{R}^N)\), and \( \Delta \) is a partition of \([s, T]\), then one has

\[
E^u V^2 \left( \int_s^T f(r, X_r) \, dM^u_r, \Delta \right) = E^u \int_s^T f^2(r, X_r) 2 \sum_{ij} d_{ij} \partial_i u \partial_j u(X_r) \, dr.
\]

This shows that if \( u_n \to u \) in such a way that \( f^2 |du_n - du|^2 \to 0 \) in \( L^1([0, T] \times \mathbb{R}^N) \), then one has, in \( L^1(P^u) \),

\[
\lim_{n \to \infty} V^2 \left( \int_s^T f(r, X_r) \, dM^u_r, \Delta \right) = 0.
\]

A similar relation holds for the backward martingales,

\[
\lim_{n \to \infty} V^2 \left( \int_s^T f(r, X_r) \, d\tilde{M}^u_r, \Delta \right) = 0.
\]

Also, if \( f_n \to f \) in such a way that \((f_n - f)^2 |du|^2 \to 0 \) in \( L^1 \), then one analogously has the relations

\[
\lim_{n \to \infty} V^2 \left( \int_s^T (f_n - f)(r, X_r) \, dM^u_r, \Delta \right) = 0.
\]

\[
\lim_{n \to \infty} V^2 \left( \int_s^T (f_n - f)(r, X_r) \, d\tilde{M}^u_r, \Delta \right) = 0.
\]

On the other hand, if \( u \in \mathcal{D}(L_0) \) and \( f(t, x) = \phi(t) v(x) \), with \( v \in \mathcal{D}(L_0) \) and \( \phi \) differentiable, then one has, by the same calculation as in the proof of Proposition 4.2 in Lyons and Stoica (1999),

\[
\frac{1}{2} \int_s^t f(r, X_r) \, dM^u_r - \frac{1}{2} \int_s^t f(r, X_r) \, dM^v_r + \int_s^t \sum_i b_i(\partial_i u) f(r, X_r) \, dr
\]

\[
- \int_s^t \sum_i (\partial_i p_i^u) p_i^v(\partial_i u) f(r, X_r) \, dr = \int_s^t f(r, X_r) \circ du(X_r),
\]

where the integral in the right-hand side designates the Stratonovich integral of two semimartingales. One is

\[
u(X_t) = u(X_s) + M^u_t - M^v_t + \int_s^t Lu(X_r) \, dr.
\]

Using the similar decomposition for \( v(X_t) \) one may write the other semimartingale as

\[
f(t, X_t) = f(s, X_s) + \int_s^t \phi(r) \, dM^u_r + \int_s^t (\partial_r + L) f(r, X_r) \, dr.
\]
Therefore the above Stratonovich integral may be also written as
\[
\int_s^t f(r, X_r) \circ d\mu(X_r) = \int_s^t f(r, X_r) \, dM_r^\mu + \int_s^t f(r, X_r) \, d\tilde{M}_r^\mu + \int_s^t \sum_{ij} a_{ij}(\partial_i u) \partial_j f(r, X_r) \, dr.
\]

Equating the two expressions of the Stratonovich integral one deduces that
\[
\int_s^t f(r, X_r) \, d\mu_r + \int_s^t f(r, X_r) \, d\tilde{M}_r^\mu
\]
has finite variation. In particular, this process has zero quadratic variation. Making an approximation based on the above relations (a–d) and then using a localization procedure it follows that
\[
\int_s^t f(r, X_r) \, d\mu_r + \int_s^t f(r, X_r) \, d\tilde{M}_r^\mu,
\]
has zero quadratic variation for each \( f \in L^2([0, T] \times \mathbb{R}^N) \).

**Remark 3.4.** The above lemma could be strengthened by showing that one has uniformly zero energy. This follows by the arguments used in the proofs of Propositions 4.1 and 4.2 in Lyons and Stoica (1999). We are avoiding the notion of “zero energy” in this paper simply because it is not related to the main subject.

**Lemma 3.5.** If \( g \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N) \), there exists a (uniquely determined) function \( h \in L^2([0, T] \times \mathbb{R}^N) \) such that \( h(t, \cdot) \in H^1 \) for almost all \( t \), \( \int_0^T \| h(t, \cdot) \| dt < \infty \), and
\[
\int_0^T [(h(t, \cdot), \varphi(t, \cdot)) + \mathcal{E}(h(t, \cdot), \varphi(t, \cdot))] dt = \int_0^T \mathcal{A}(g(t, \cdot) \, d\varphi(t, \cdot)) dt,
\]
\( \varphi \in C^\infty([0, T]) \otimes C^\infty_c(\mathbb{R}^N) \).

**Proof.** The scalar product
\[
\langle \langle u, v \rangle \rangle = \int_0^T [(u(t, \cdot), v(t, \cdot)) + \mathcal{E}(u(t, \cdot), v(t, \cdot))] dt,
\]
defines a Hilbert space structure on the space of all functions \( u \in L^2([0, T] \times \mathbb{R}^N) \) such that
\[
\int_0^T \mathcal{E}(u(t, \cdot)) dt < \infty.
\]
The map
\[
v \mapsto \int_0^T \mathcal{A}(g(t, \cdot), d\nu(t, \cdot)) dt
\]
is a continuous linear functional on this Hilbert space and Riesz’s representation theorem gives the function \( h \) which fulfils the requirements of the statement.
3.2. Some corollaries

Taking the conditional expectation in the relation from Theorem 3.2 one easily deduces the following formula similar to the Feynman–Kac representation.

**Corollary 3.6.** If \( u \in H_T \) is a solution of the equation

\[
(\partial_t + L)u + f - \sum_{ij} \hat{c}_i(a_{ij}g_j) = 0,
\]

with \( g \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N) \), \( f \in L^2([0, T] \times \mathbb{R}^N) \) and final condition \( u(T, \cdot) = \Phi \in L^2(\mathbb{R}^N) \), then one has

\[
u(t, x) = E^x \left[ \Phi(X_T) + \int_t^T f(r, X_r) \, dr + \frac{1}{2} \int_t^T g(r, X_r) \ast dX \right],
\]

\( m(\,d\,x) \) almost everywhere, for each fixed \( t \in [0, T] \).

The following corollary of Theorem 3.2 asserts an analytical fact, namely that the function \( u \) admits a quasicontinuous version. This quasicontinuity holds uniformly in time.

**Corollary 3.7.** Let \( u \in H_T \) be as in the statement of the theorem. Then there exists a function \( \tilde{u} : [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R} \), which coincide with \( u \) as an element of \( H_T \) and, for each \( \varepsilon > 0 \), there exists an open set \( D_\varepsilon \subset [0, T] \times \mathbb{R}^N \) such that the restriction of \( u \) to \( D_\varepsilon \) is continuous and

\[
P^m\{\omega \in \Omega : \exists t \in [0, T] \text{ s.t. } (t, X_t(\omega)) \in D_\varepsilon \} < \varepsilon.
\]

In particular, the maps \( t \rightarrow \tilde{u}(t, X_t(\omega)) \) are continuous on \([0, T]\), except of a set of null \( P^m \)-measure.

**Proof.** If \( \Phi \in C_0(\mathbb{R}^N) \) and \( f, g_i \in C(\mathbb{R}^N) \times C_c(\mathbb{R}^N) \), then it is known that \( u \) has version that is continuous on \([0, T] \times \mathbb{R}^N \) (see Aronson, 1968). On the other hand, the estimate of Proposition 2.4 and the representation proved in the theorem imply the following:

\[
E^m \left( \sup_{0 \leq t \leq T} u(t, X_t)^2 \right) \leq \text{const} \left( \|\Phi\|_2^2 + \|f\|_{L^2([0, T] \times \mathbb{R}^N)}^2 + \|g\|_{L^2([0, T] \times \mathbb{R}^N)}^2 \right).
\]

Now, for general data \((\Phi, f, g)\), one chooses an approximating sequence \((\Phi^n, f^n, g^n)\) of smooth functions as above and such that

\[
\|\Phi^n - \Phi\|_2 \leq \frac{1}{2^n}, \quad \|f^n - f\|_{L^2([0, T] \times \mathbb{R}^N)}^2 \leq \frac{1}{2^n},
\]

\[
\||g^n - g\|_{L^2([0, T] \times \mathbb{R}^N)}^2 \leq \frac{1}{2^n}.
\]

Let \( u^n \) be the sequence of continuous solutions of the equation in the theorem corresponding to the data \((\Phi^n, f^n, g^n)\). Then set \( D_n = \{|u^n - u^{n+1}| > 1/n^2\} \) and \( D_n' = \bigcup_{k \geq n} D_k \).
Then, for each fixed $n_0$, the sequence $(u_n)$ converges uniformly outside the set $D_{n_0}'$. By the preceding estimate we will have

$$
E^m \left( \sup_{0 \leq t \leq T} [u^{n}(t,X_t) - u^{n+1}(t,X_t)]^2 \right) \leq \text{const} \frac{1}{2^n},
$$

and hence

$$
P^m \{ \omega / \exists t \in [0,T] \text{ s.t. } (t,X_t(\omega)) \in D_n \} \leq \text{const} \frac{n^4}{2^{2n}},
$$

$$
P^m \{ \omega / \exists t \in [0,T] \text{ s.t. } (t,X_t(\omega)) \in D_{n}' \} \leq \text{const} \sum_{k \geq n} \frac{k^4}{2^{2k}}.
$$

It follows that the set $\bigcap_n D_{n}'$ is negligible, and so one may define the function $\bar{u} = \lim_{n} u_n$ on the complement of $\bigcap_n D_{n}'$, which turns out to satisfy the requirements of the statement.

3.3. The case of a solution of a SDE

Let us look at the case of a diffusion associated to a stochastic differential equation,

$$
X_t^{x,i} = x^i + \sum_{k=1}^{n} \int_{0}^{t} \sigma^i_k(X_s) \, dB^k_s + \int_{0}^{t} \beta^i(X_s) \, ds, \quad i = 1, \ldots, N,
$$

where $x \in \mathbb{R}^N$, and $B$ is an $n$-dimensional Brownian motion over a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Assume that the coefficients of the diffusion are smooth so that the infinitesimal generator takes the form

$$
L = \frac{1}{2} \sum_{i,j,k} \sigma^i_k \sigma^j_k \partial_i \partial_j + \sum_i \beta^i \partial_i = \sum_i \partial_i (d^i \partial_j) + \sum_i \left( \beta^i - \sum_j \partial_j a^{ij} \right) \partial_i.
$$

It is natural to try express the preceding stochastic calculus in terms of the driving Brownian motion. Then the martingale parts are easily seen to be

$$
M_t^i = \sum_k \int_{0}^{t} \sigma^i_k(X_s) \, dB^k_s, \quad i = 1, \ldots, N.
$$

In order to get a simple expression for the backward martingale parts we are going to suppose that $\beta^i = \sum_j \partial_j a^{ij}$. Then $L = L_0$, the process is symmetric and the Lebesgue measure, $m$, is invariant so that $p^m \equiv 0$. Setting $\Omega' = \mathbb{R}^N \times \Omega$, $P' = m \otimes P$ and $X_t(x,w) = X_t^x(w)$ we obtain a process $X = (X_t)$ which is a Markov realization with initial distribution $m$ over the measure space $(\Omega', P')$. Under this measure the calculations previously done are easily adapted and we will have $\bar{x}^m \equiv 0$, and

$$
\bar{M}^{m,i}(s,t) = \sum_k \int_{s}^{t} \sigma^i_k(X_r) \, dB^k_r - \sum_{j,k} \int_{s}^{t} (\partial_j \sigma^i_k) \, dB^k_r dr,
$$

where $\bar{B}^k_s = B^k_s - B^k_T$. Trying to give a meaning with respect to the initial Brownian motion, the integral with respect to $\bar{B}$ do not make sense directly, because $X$ is not
adapted to the backward filtration yielded by $\overline{B}$. However it may get a sense as an integral with respect to the direct filtration if we consider it as the difference between the Ito integral and two times the Stratonovich one, $d\overline{B}_k = dB_k - 2 \circ dB_k$. Suppose further that $(\sigma_k^i)$ is a square matrix and let $(\tau_k^i)$ be its inverse. Integrating backward we obtain the following backward martingale:

$$\tilde{N}^i(s,t) = \sum_i \int_s^t \tau^i_k(X_r) \, d\tilde{M}^{m,l} = B^i_s - B^i_t - \sum_j \int_s^t \hat{\sigma}^j_k(X_r) \, dr.$$ 

Calculating the brackets we see that $\tilde{M}^i(s,T)$, $i = 1, \ldots, N$ represent a backward Brownian motion over $(\Omega', P')$ endowed with the backward filtration $\mathcal{F}_{t,T} = \sigma(X_s/s \in [t, T])$. It is this that gives the correct interpretation of the backward martingales,

$$\tilde{M}^{m,l}(s,t) = \sum_i \int_s^t \sigma^i_k \, d\tilde{M}^i.$$ 

Further, writing formally the stochastic integral denoted with $\ast$ we will have

$$\int_s^t g \, dX = \sum_{i,k} \left[ \int_s^t g_i(r,X_r)\sigma^i_k(X_r) \, dB^k_r + \int_s^t g_i(r,X_r)\sigma^i_k(X_r) \, d\overline{B}^k_r \right]$$

$$- \sum_{i,j,k} \int_s^t g_i(r,X_r)(\hat{\sigma}^j_k)\sigma^i_k(X_r) \, dr.$$ 

Since $g$ is only measurable, the integral with respect to $d\overline{B}$ is no more interpretable by means of a Stratonovich integral. Therefore this relation has no meaning in this form.

### 3.4. The case of Brownian motion

Suppose now that $X^i_t = B^i_t$, $i = 1, \ldots, N$ is the Brownian motion started at zero. Then one has $M^i_t = B^i_t$, $p^0_t(x) = 1/(2\pi t)^{N/2} \exp - |x|^2/2t$, and $\tilde{M}^i(s,t) = B^i_s - B^i_t + \int_s^t B^l_r \, dr$. Again, by calculating the bracket, one sees that this is a backward Brownian motion with respect to the backward filtration. The integral denoted with $\ast$ can be expressed in terms of the original Brownian motion, by means of a Stratonovich integral as follows:

$$\int_s^t g \, dX = 2 \sum_i \left( \int_s^t g_i(X_r) \, dB^i_r - \int_s^t g_i(X_r) \circ dB^i_r \right) = \sum_i \int_s^t \hat{\sigma}^i g_i(X_r) \, dr,$$ 

provided $g$ is smooth. For measurable $g$ this again has no meaning. For this integral to be well defined, one should consider the backward filtration.

### 3.5. Note added in the proof

The referee of this paper has made several interesting comments. Its main observation is that the uniform ellipticity condition is not essential, and only the existence of a strictly positive density is needed for the validity of the main result of this paper. He notes that if one considers diffusion processes with smooth coefficients and Hörmander
type conditions to ensure the existence of strictly positive density, the result concerning the forward–backward decomposition (Theorem 3.1 of Lyons and Stoica, 1999) remain valid, as well as Theorem 1 of the present paper. This is indeed true. In fact, in the presence of densities, the representation under $P^m$ or under $P^x$, $x \in \mathbb{R}^N$, are equivalent.

However, if one is interested only in the representation under $P^m$, then the most general frame that ensures the validity of a suitable generalization of Theorem 1 is that of a process associated to a Dirichlet space (see e.g. Bally et al., in preparation). In the forward–backward martingale decomposition the Dirichlet space methods are crucial. We content ourselves with the frame of this paper because it is somehow standard and we may easily refer to the literature.

The author thanks the referee for his remarks, in particular for drawing his attention to the connections of this paper with other related subjects.

4. Backward stochastic differential equations

Definition 4.1. (1) Let $\Phi: \mathbb{R}^N \rightarrow \mathbb{R}$, $f, u: [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}$, $g, \phi: [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}^N$ be measurable functions such that $\Phi \in L^2(\mathbb{R}^N)$, $f \in L^2([0, T] \times \mathbb{R}^N)$, $\phi, g \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N)$ and set $\xi = \Phi(X_T)$, $Y_t = u(t, X_t)$, and $Z_t = \phi(t, X_t)$ (where the components of $Z$ are $Z_{i,t} = \phi_i(t, X_t)$, $i = 1, \ldots, N$). We will say that the pair $(Y, Z)$ is a solution of the linear backward stochastic differential equation (BSDE) with final condition $\xi$ and data $f, g$, provided that the following relation holds:

$$Y_t = \xi - \sum_i \int_t^T Z_{i,r} \, dM'_i + \int_t^T f(r, X_r) \, dr + \frac{1}{2} \int_t^T g(r, X_r) \, d\langle X \rangle_r,$$

for any $0 \leq t \leq T$, $P^m$-a.s.

(2) If $f: [0, T] \times \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ and $g: [0, T] \times \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^N$ satisfy the conditions (h1–4) and $\Phi \in L^2(\mathbb{R}^N)$, $u: [0, T] \times \mathbb{R}^N \rightarrow \mathbb{R}$ is measurable and belongs to $L^2([0, T] \times \mathbb{R}^N)$, $\phi \in L^2([0, T] \times \mathbb{R}^N; \mathbb{R}^N)$ and $\xi, Y, Z$ are defined as above, we will say that $(Y, Z)$ is a solution of the nonlinear BSDE with final condition $\xi$ and data $f, g$, provided that the following relation holds:

$$Y_t = \xi - \sum_i \int_t^T Z_{i,r} \, dM'_i + \int_t^T f \left( r, X_r, Y_r, \frac{1}{\sqrt{2}} Z_r \sigma(X_r) \right) \, dr$$

$$+ \frac{1}{2} \int_t^T g \left( r, X_r, Y_r, \frac{1}{\sqrt{2}} Z_r \sigma(X_r) \right) \, d\langle X \rangle_r,$$

for any $0 \leq t \leq T$, $P^m$-a.s.

Clearly, in some sense, it is only a matter of convenience to distinguish between linear and nonlinear BSDE’s. In fact the linear equation is a particular case of the nonlinear equation, namely represents the case when $f$ and $g$ do not depend on $y$ and $z$. Similarly, once one has a solution, the nonlinear equation may be viewed as a linear equation with $f(t, x) = f_u(t, x) = f(t, x, u(t, x), (1/\sqrt{2}) du(t, x) \sigma(x))$ and $g(t, x) = g_u(t, x) = g(t, x, u(t, x), (1/\sqrt{2}) du(t, x) \sigma(x))$. 


Because the density of the semigroup is strictly positive, it follows that for each \( t > 0 \)
all measures \( P^\mu \circ \theta_t^{-1} \) are equivalent to \( P^m \). Moreover, since the stochastic integrals
with respect to \( dM^i \) or \( *dX \) are additive functionals, by the Markov property it follows
that the above relations (7) and (8) hold with \( t > 0 \) under arbitrary \( P^\mu \) (including \( P^m \))
if and only if they hold under one particular such measure.

Observe, however, that for such a solution one has \( Y_0 \in L^2(P^m) \), while \( Y_0 \) may be
nonintegrable for other measures \( P^\mu \).

Note also that the last term, that involving the stochastic integral denoted by \( *dX \),
do not make sense for arbitrary \((F_t)\) adapted processes. It is necessary that the
integrand be Markovian, that is a function of the present state \( X_t \), in order to perform
backward integration.

Proposition 4.2. (1) Let \( \Phi, f, g, u, \phi \) be as in case (1) of the above definition, so
that \( Y_t = u(t, X_t), \ Z_t = \phi(t, X_t) \) is a solution of the linear BSDE with the condition
\( Y_T = \xi = \Phi(X_T) \). Then \( u \) belongs to \( H_T \) and represents a solution of the linear PDE,

\[
(\partial_t + L)u + f - \sum_{ij} \partial_i a_{ij} g_j = 0,
\]

with \( u(T, \cdot) = \Phi \). Moreover, one has \( \phi = du \). Conversely, if \( u \) is a solution of (9), then
\( Y_t = u(t, X_t), \ Z_t = du(t, X_t) \) represents a solution of the linear BSDE.

(2) If \( \Phi, f, g, u, \phi \) are in case (2) of the above definition so that \( Y_t = u(t, X_t), \ Z_t = \phi(t, X_t) \) is a solution of the non-linear BSDE, then \( u \in H_T \) and it is a solution of the
non-linear PDE,

\[
(\partial_t + L)u + f \left( \cdot, \cdot, u, \frac{1}{\sqrt{2}} du \sigma \right) - \sum_{ij} \partial_i a_{ij} g_j \left( \cdot, \cdot, u, \frac{1}{\sqrt{2}} du \sigma \right) = 0,
\]

with the boundary condition \( u(T, \cdot) = \Phi \) and \( \phi = du \). Conversely, if \( u \) is a solution
of Eq. (10), then \( Y_t = u(t, X_t), \ Z_t = du(t, X_t) \) represents a solution of the non-linear BSDE.

Proof. (1) First note that, by Theorem 3.2, any solution \( u \) of the PDE (9) gives rise to a
solution \( Y_t = u(t, X_t), \ Z_t = du(t, X_t) \) of the linear BSDE. Now let \( Y_t' = u'(t, X_t), \ Z_t' = \phi(t, X_t) \)
be another solution of the linear BSDE (7). Then one has

\[
Y_t - Y_t' = - \int_t^T (Z_t s - Z_t' s) dM^i s.
\]

Conditioning with respect to \( \mathcal{F}_t \) one gets \( Y_t = Y_t' \mbox{ -a.s.} \), which implies \( u'(t, \cdot) = u(t, \cdot) \)
almost everywhere. The bracket of the above martingale is

\[
\int_t^T \sum_{ij} a_{ij}(X_s)(\phi_i(X_s) - \hat{\partial}_i u(s, X_s))(\phi_j(X_s) - \hat{\partial}_j u(s, X_s)) ds = 0,
\]

which implies \( \phi(t, x) = du(t, x), \ dt \otimes dx \)—almost everywhere.

(2) The assertions of the second part of the statement follow easily from the first part. \( \square \)
Now we prove an Ito’s type formula which is appropriate to our framework. The proof is as in the classical case; the only thing that should be noted is that the effect of the backward martingale part results in a new quadratic variation term, which is lastly written in the next lemma.

**Lemma 4.3.** Let \((Y,Z)\) be a solution of the linear BSDE, with the notation of the above definition. If \(\varphi \in \mathcal{C}^2(R)\), then the following relation holds, under \(P^m\):

\[
\varphi(Y_t) + \int_t^T \varphi''(Y_s)|Z_s|^2 ds = \varphi(\xi) - \sum_i \int_t^T \varphi'(Y_s)Z_{i,s} dM^i_s + \int_t^T \varphi'(Y_s)f(s,X_s) ds + \frac{1}{2} \int_t^T \varphi'(Y_s)g(s,X_s) * dX_s + \int_t^T \varphi''(Y_s)\sum_{ij} a_{ij}(X_s)g_i(s,X_s)Z_{j,s} ds.
\]

The stochastic integrals with respect to \(dM^i\) and \(\tilde{M}^i\) in the above formula make sense since, by Corollary 3.7, the processes \(\varphi'(Y_s)\) and \(\varphi''(Y_s)\), are pathwise bounded and the resulting integrals are to be seen as local forward resp. backward martingales.

**Proof.** By analytical properties of the linear equation it is known that the solution \(u\) is bounded, provided the data are bounded (see Aronson, 1968). Therefore, by approximation, we are reduced to the case where \(u\) (resp. \(Y\)) is bounded. Next we actually give a proof for \(\varphi(x) = x^2\), the case of arbitrary \(\varphi\) being similar, based on a second order Taylor expansion.

One takes a partition \(\Delta = (t = t_0 < t_1 < \cdots < t_{k+1} = T)\) and writes

\[
Y^2_T - Y^2_t = 2 \sum_{i=0}^{k} Y_{h_i}(Y_{h_i+1} - Y_{h_i}) + \sum_{i=0}^{k} (Y_{h_i+1} - Y_{h_i})^2.
\]

Since the last term in (7) has zero quadratic variation (by Lemma 2.3), it follows that the quadratic variation of \(Y_\cdot\), which is obtained in the limit from the last term in the above relation, equals the quadratic variation of the martingale part

\[
= 2 \int_t^T \sum_{ij} a_{ij}(X_r)Z_{i,r}Z_{j,r} dr.
\]

To examine further relation (\(\ast\)) we introduce the notation

\[
M_r = \sum_i \int_t^r Z_{i,r} dM^i_r,
\]

\[
N_r = \sum_i \int_t^r g_i(r,X_r) dM^i_r,
\]

\[
\tilde{N}^u(s,r) = \sum_i \int_s^r g_i(r,X_r) d\tilde{M}^{u,i}.
\]
Using relation (7) in order to express \( Y_{t_{i+1}} - Y_t \), one easily sees that
\[
2 \sum_i Y_t Y_{t_{i+1}} - Y_t
\]
represents a Riemann sum convergent to an integral and only one term needs some discussion. Namely the term
\[
- \sum_i Y_t \tilde{N}^\mu (t_i, t_{i+1}) + \sum_i (Y_{t_{i+1}} - Y_t) \tilde{N} (t_i, t_{i+1}).
\]
The first sum approximates the integral
\[
- \int_t^T Y_r \, d\tilde{N}^\mu = - \sum_i \int_t^T Y_r g_i (r, X_r) \, d\tilde{M}^{\mu,i},
\]
while the second sum tends to
\[
\langle M, \tilde{N}^\mu \rangle_t^T = - \langle M, N \rangle_t^T = -2 \int_t^T \sum_{ij} a_{ij} (X_s) g_i (r, X_r) Z_{j,s} \, dr.
\]
This is a consequence of the fact that \( N_r + \tilde{N}^\mu (t, r) \) has zero quadratic variation.

The lemma generalizes to several solutions of BSDE’s as follows.

**Lemma 4.4.** Assume that \( \Phi^l \in L^2 ( \mathbb{R}^N ) \), \( f^l \in L^2 ( [0, T] \times \mathbb{R}^N ) \), \( g^l \in L^2 ( [0, T] \times \mathbb{R}^N ; \mathbb{R}^N ) \), \( l = 1, \ldots, k \), are given and \((Y^l, Z^l)\) are the corresponding solutions of the linear BSDE. If \( F \in \mathcal{C}^2 ( \mathbb{R}^k ) \), then one has
\[
F(Y^1_t, \ldots, Y^k_t) + \int_t^T \sum_{l,r=1}^k \partial_l \partial_r F(Y^1_s, \ldots, Y^k_s) \sum_{ij} a_{ij} (X_s) Z^l_{i,s} Z^r_{j,s} \, ds
\]
\[
= F(\xi^1, \ldots, \xi^k) - \sum_i \int_t^T \sum_l \partial_l F(Y^1_s, \ldots, Y^k_s) Z^l_{i,s} \, dM^l_i
\]
\[
+ \int_t^T \sum_l \partial_l F(Y^1_s, \ldots, Y^k_s) f^l(s, X_s) \, ds
\]
\[
+ \frac{1}{2} \int_t^T \sum_l \partial_l F(Y^1_s, \ldots, Y^k_s) g^l(s, X_s) * \, dX
\]
\[
+ \int_t^T \sum_{l,r} \partial_l \partial_r F(Y^1_s, \ldots, Y^k_s) \sum_{ij} a_{ij} (X_s) Z^l_i g^r_j(s, X_s) \, ds.
\]

**Acknowledgements**

This work was done while the author was visiting the Department of Mathematics of Université du Maine, Le Mans, whose support is gratefully acknowledged. It covers the main part of the talk given at the meeting “Equations differentielles stoch astiques retrogrades et applications, Le Mans, 3–4 June 1999”. Thanks are given to Vlad Bally.
and Laurent Denis for stimulating discussions concerning BSDE’s and particularly the subject of this paper. The author is also grateful for the working atmosphere in the Department of Mathematics.

References

Bally, V., Pardoux, E., Stoica, L. Backward stochastic differential equations associated to a symmetric Markov process (in preparation).