CHAPTER 6

Stochastic Representations for Nonlinear Parabolic PDEs

H. Mete Soner*

Koç University, Istanbul, Turkey
E-mail: msoner@ku.edu.tr

Contents
1. Introduction ........................................ 3
Notation ........................................ 4
2. Linear case: Feynman–Kac representation .................. 4
2.1. Linear monotone semigroups .......................... 5
2.2. Zeroth-order term and forcing ......................... 8
2.3. Diffusions and parabolic PDEs ....................... 9
2.4. Initial value problems ............................ 11
2.5. Discrete Markov processes and simply coupled equations .... 12
2.6. Jump Markov processes and integro-differential equations .... 14
2.7. Dirichlet boundary conditions .................... 16
2.8. Neumann condition and the Skorokhod problem .... 17
2.9. Stationary problems ............................. 18
3. Representation via controlled processes ................ 19
3.1. Viscosity solutions .......................... 22
3.2. Optimal control of diffusion processes ............... 26
3.3. Smooth value function and verification .............. 31
3.4. Optimal control of jump Markov processes .......... 33
3.5. Other type of control problems .................. 34
3.6. Stochastic target and geometric problems .......... 35
4. Backward representations ........................ 38
4.1. Backward stochastic differential equations .......... 38
4.2. Second-order backward stochastic equations ........ 39
5. Monte Carlo methods ................................ 45
5.1. The linear case .................................. 45
5.2. The semilinear case .......................... 46
5.3. The fully nonlinear case ....................... 47
Acknowledgement ................................ 48
References .................................. 48

*Member of the Turkish Academy of Sciences.

HANDBOOK OF DIFFERENTIAL EQUATIONS
Evolutionary Equations, volume 3
Edited by C.M. Dafermos and E. Feireisl
© 2007 Elsevier B.V. All rights reserved
Abstract
We discuss several different representations of nonlinear parabolic partial differential equations in terms of Markov processes. After a brief introduction of the linear case, different representations for nonlinear equations are discussed. One class of representations is in terms of stochastic control and differential games. An extension to geometric equations is also discussed. All of these representations are through the appropriate expected values of the data. Different type of representations are also available through backward stochastic differential equations. A recent extension to second-order backward stochastic differential equations allow us to represent all fully nonlinear scalar parabolic equations.

Keywords: Second-order backward stochastic differential equations, Fully nonlinear parabolic partial differential equations, Viscosity solutions, Superdiffusions, Feynman–Kac formula, BSDE, 2BSDE

MSC: 60H10, 35K55, 60H30, 60H35
1. Introduction

In this chapter we outline several connections between partial differential equations (PDE) and stochastic processes. These are extensions of Feynman–Kac-type representation of solutions to PDEs as the expected value of certain stochastic processes. Possible numerical implications of these connections are discussed as well. Although we restrict the scope of this paper to representation formulae, stochastic analysis provides much more analytical tools for PDEs. In particular, superdiffusions as developed by Dynkin [28–31] are related to nonlinear PDEs with a power-type nonlinearity, for reaction-diffusion equations interesting connections were used by Freidlin [38,39] to prove deep analytical results for these equations, also Barlow and Bass [7] study equations on fractals using random processes. Other important issues such as Martin boundaries, hypoellipticity and Malliavin calculus is not covered in these notes. Moreover, the theory of partial differential equation with stochastic forcing terms is not included. Interested readers may consult the papers [50,51] by Lions and Souganidis and [17] by Buckhadam and Ma and the references therein.

The starting point of most of our analysis is the celebrated Feynman–Kac formula [35, 43] which states that any solution of the linear heat equation

\[
\frac{\partial u}{\partial t}(t, x) = \Delta u(t, x), \quad t > 0, x \in \mathbb{R}^d,
\]

with initial condition \( u(0, x) = f(x) \) with certain growth conditions (see Section 2) is given by

\[
u(t, x) := E\left[f(x + \sqrt{2}W(t))\right],
\]

where \( W(\cdot) \) is the standard \( d \)-dimensional Brownian motion. This well-understood connection can be explained in several different ways. We will employ the semigroups to motivate this connection and to generalize it to more general stochastic processes. In that section we will briefly state Feynman–Kac-type formulae for several class of linear equations. The most general class of equations we will consider are second-order parabolic type integro-differential equations. Boundary value problems of Dirichlet and Neumann type are also discussed.

In Section 3 we extend these results to nonlinear equations of same type by using controlled stochastic processes. Since these equations do not always admit classical (or smooth) solutions, we will employ the theory of viscosity solutions to prove the representation formulae rigorously. The chief tool in this analysis is the dynamic programming principle which was first observed by Bellman [8]. The infinitesimal version of the dynamic programming principle is in fact gives the related partial differential equation. In this context, semigroup motivation plays an important role, as the theory of viscosity solutions is best explained through semigroups and the dynamic programming principle is in fact the semigroup property. We refer to the books by Bensoussan and Lions [9,10], Krylov [46] and Fleming and Soner [36] for more references and the historical development of the theory.
In that section we also provide a more recent representation formulae for geometric type equations. This is achieved by using a nonclassical control problem called target problems \[67,68\].

In this chapter we restrict our attention to only optimal control. However, these methods extend naturally to stochastic differential games. For this extension, we refer to the paper by Fleming and Souganidis \[37\] and Chapter 11 in the second edition of \[36\].

Another type of connection between PDEs and stochastic processes is given by backward stochastic differential equations (BSDEs in short). These formulae is analogous to method of characteristics for first-order equations. Indeed, initially BSDEs were studied by Bismut \[11,12\] then by Peng \[59\] as an extension of Pontryagin maximum principle which itself is an extension of characteristics. We provide a brief introduction to BSDEs and then outline a recent result of Cheredito, Soner, Touzi and Victoir \[21\]. This result extends the representation to formulae to all fully nonlinear, parabolic, second-order partial differential equations.

Last section is devoted to possible numerical implications of these formulae.

### Notation

Let \( d \geq 1 \) be a natural number. We denote by \( \mathcal{M}_{d,k} \) the set of all \( d \times k \) matrices with real components, \( \mathcal{M}^d = \mathcal{M}_{d,d} \). \( B' \) is the transpose of a matrix \( B \in \mathcal{M}^d \) and \( \text{Tr}[B] \) its trace. By \( \mathcal{M}_{d}^{\text{inv}} \) we denote the set of all invertible matrices in \( \mathcal{M}^d \), by \( S^d \) all symmetric matrices in \( \mathcal{M}^d \), and by \( S^d_+ \) all positive semidefinite matrices in \( \mathcal{M}^d \). For \( B, C \in \mathcal{M}^d \), we write \( B \succeq C \) if \( B - C \in S^d_+ \). For \( x \in \mathbb{R}^d \), we set

\[
|x| := \sqrt{x_1^2 + \cdots + x_d^2}
\]

and for \( B \in \mathcal{M}^d \),

\[
|B| := \sup_{x \in \mathbb{R}^d, |x| \leq 1} Bx.
\]

Equalities and inequalities between random variables are always understood in the almost sure sense. \( W(\cdot) \) is a multidimensional Brownian motion on a complete probability space \((\Omega, \mathcal{F}, P)\). For \( t \geq 0 \), we denote by \((\mathcal{F}_t)_{t \geq 0}\) a filtration satisfying the usual conditions and containing the filtration generated by \( \{W(s)\}_{s \in [0,T]} \).

### 2. Linear case: Feynman–Kac representation

Let \( W(\cdot) \) be the standard \( d \)-dimensional Brownian motion, \( a \) be positive constant, and \( f \) be a scalar-valued, continuous function \( f \) on \( \mathbb{R}^d \) satisfying the growth condition

\[
|f(x)| \leq C[1 + |x|^\alpha] \quad \forall x \in \mathbb{R}^d,
\]
Stochastic representations for nonlinear parabolic PDEs

for some constants $C, \alpha \geq 0$. Then, the Feynman–Kac formula [35,43] states that

$$u(t, x) := E\left[f\left(x + a\sqrt{2}W(t)\right)\right],$$

(2.1)

is the unique solution of the heat equation

$$\frac{\partial u}{\partial t}(t, x) := u_t(t, x) = a\Delta u(t, x), \quad t > 0, \quad x \in \mathbb{R}^d,$$

together with the initial condition

$$u(0, x) = f(x), \quad x \in \mathbb{R}^d.$$

Indeed, once we know either if the function defined by the expected value is $C^{1,2}$, or if the heat equation has a smooth solution, then the above representation is an direct application of the Itô formula, Theorem 3.3 in [44]. The above growth condition is sufficient for either one of these conditions; see for instance Section 4.4, Remark 4.4 in [44]. Moreover, in this special case, polynomial growth can be weakened.

The above formula generalizes to a large class of Markov processes, linear equations and boundary problems. In this section we briefly and formally describe all these generalizations.

2.1. Linear monotone semigroups

The connection between the Markov processes and certain linear equations is now well understood and can be explained in many ways. In this chapter we will utilize semigroups to motivate this connection. The semigroup approach has the advantage that it generalizes to the nonlinear setting and it is well adapted to the theory of viscosity solutions. However, we use this approach only to motivate the connection and therefore our discussion of semigroups is only formal. In particular, we will not be precise about the domains of the operators.

In the initial discussion, we assume that the equations are defined on a metric space $\mathbb{D}$ which is equal to either $\mathbb{R}^d$ or to $\mathbb{R}^d \times \{1, 2, \ldots, N\}$. Problems on bounded subsets of $\mathbb{R}^d$ are, of course, common and similar representation results are available for these equations as well. But, in this subsection, we restrict our analysis to problems defined on all of $\mathbb{R}^d$ or $\mathbb{R}^d \times \{1, 2, \ldots, N\}$. Boundary problems will be discussed in the Sections 2.7 and 2.8.

Also, to simplify the presentation, we will consider PDEs that are backward in time. For these equations a terminal data at a given time $T$, instead of an initial data, is given. Then a solution is constructed for all times prior to $T$. Of course, there is a direct connection between terminal value problems and initial value problems through a simple time reversal. We perform this change for diffusion processes in the Section 2.4.

For all $t \geq 0$, let $\hat{L}_t$ be a linear operator on a subset $\hat{D}$ of $C_b(\mathbb{D})$ – bounded, scalar-valued, continuous functions on $\mathbb{R}^d$. Let $\varphi \in C_b(\mathbb{D})$ be a given function. For $T \geq 0$, consider the linear equation

$$-u_t(t, z) = (\hat{L}_t u(t, \cdot))(z), \quad \forall (t, z) \in (-\infty, T) \times \mathbb{D},$$

(2.2)
together with final data
\[
    u(T, z) = \varphi(z) \quad \forall z \in \mathbb{D}. \tag{2.3}
\]

Assume that for every \( T \) and \( \varphi \in C_b(\mathbb{D}) \) this equation has a unique smooth solution and let \( u \) be this unique solution. Clearly this solution depends on \( T \) and \( \varphi \), but this dependence is always suppressed. Now, define a two-parameter family of operators
\[
    (\widehat{T}_{t,T} \varphi)(z) := u(t, z), \quad \forall t \leq T, z \in \mathbb{D}.
\]

By uniqueness, this family is a linear semigroup, i.e.,
\[
    \widehat{T}_{t,T} \varphi = \widehat{T}_{t,r} \left( \widehat{T}_{r,T} \varphi \right) \quad \forall t \leq r \leq T. \tag{2.4}
\]

Moreover, it is clear that the infinitesimal generator of this semigroup is the operator \( \widehat{L}_t \),
\[
    \lim_{h \downarrow 0} \frac{\widehat{T}_{t,t+h} \varphi - \varphi}{h} = \widehat{L}_t \varphi
\]
for every \( \varphi \in \mathbb{D} \).

We continue by constructing a similar semigroup using Markov processes. For this purpose, let \((\Omega, P, \mathcal{F})\) be probability space and \(\{\mathcal{F}_r\}_{r \geq 0}\) be a filtration. Let \(\{Z^{t,z}(r)\}_{r \geq t}\) be a \(\mathbb{D}\)-valued Markov process on this probabilistic structure starting from \(Z^{t,z}(t) = z\). For a continuous bounded function \(\varphi\), define a two parameter semigroup by
\[
    T_{t,T} \varphi(z) := E\left[\varphi(Z^{t,z}(T))\right] \quad \forall t \leq T.
\]

Formally, the Markov property of \( Z \) implies that \( T_{t,T} \) satisfies (2.4). Indeed, for \( t \leq r \leq T \),
\[
    T_{t,r} \left( T_{r,T} \varphi \right)(z) = E\left[\left( T_{r,T} \varphi \right)(Z^{t,z}(r))\right] = E\left[ E\left[ \varphi(Z^{r,z}(T)) \right] \right].
\]

Since, by the Markov property of the process,
\[
    Z^{r,z}(r)(T) = Z^{t,z}(T), \tag{2.5}
\]
we have
\[
    T_{t,r} \left( T_{r,T} \varphi \right)(z) = E\left[ E\left[ \varphi(Z^{t,z}(T)) \right] \right] = T_{t,T} \varphi(z).
\]

Hence, \( T_{t,T} \) is a two-parameter semigroup. Following the theory of semigroups, the infinitesimal generator
\[
    L_t \varphi := \lim_{r' \to r, h \downarrow 0} \frac{T_{t,r+h} \varphi - \varphi}{h} \tag{2.6}
\]
exists for \( \varphi \in D \subset C_b(\mathbb{D}) \), for some subset \( D \) (see Section 5.1 in [44]).

Now, it is clear that to have a Feynman–Kac representation for the equations (2.2) and (2.3), we need construct a Markov process whose infinitesimal generator \( \mathcal{L}_t \) agrees with the linear operator \( \hat{\mathcal{L}}_t \) that appears in (2.2). To see this connection formally, suppose \( \mathcal{L}_t = \hat{\mathcal{L}}_t \). Fix \( T \) and \( \varphi \), and set

\[
\begin{align*}
\text{u}(t, z) := \mathcal{T}_t, T \varphi(z) = E\left[ \varphi\left(Z_t, Z(T)\right) \right].
\end{align*}
\]

We formally claim that \( \text{u} \) solves (2.2). Indeed, assume that \( \text{u} \) is smooth in the sense that it has a continuous time derivative and \( \text{u}(t, \cdot) \in D \) for all \( t \). Then, by the semigroup property,

\[
\begin{align*}
-\text{u}(t, z) &= \lim_{h \downarrow 0} \frac{\text{u}(t-h, z) - \text{u}(t, z)}{h} \\
&= \lim_{h \downarrow 0} \frac{\mathcal{T}_{t-h, T} \varphi(x) - \text{u}(t, z)}{h} \\
&= \lim_{h \downarrow 0} \frac{\mathcal{T}_{t-h, T}(\text{u}(t, \cdot))(z) - \text{u}(t, z)}{h} \\
&= \mathcal{L}_t \left( \text{u}(t, \cdot) \right)(z).
\end{align*}
\]

Hence, \( \text{u} \) defined by (2.7) solves (2.2). The terminal value (2.3) follows from the definition of \( \text{u} \). Hence, if equation (2.2) together with (2.3) has a unique “smooth” solution, then it must be given by (2.7). However, in most cases we can prove this representation directly and obtain uniqueness as a byproduct of the representation.

In the example of a Brownian motion, it is well known that the infinitesimal generator is the Laplacian and therefore we have the representation (2.1) for the heat equation. In the same spirit, we can prove representation results for a large class of linear equations including equations with nonlocal terms. However, the semigroup generated by Markov processes are monotone and this puts a certain restriction on the operators \( \mathcal{L}_t \) that are the infinitesimal generators of the semigroups constructed by Markov processes. Indeed, the monotonicity of the stochastic semigroup (2.7) is a direct consequence of the definition and stated as

\[
\varphi \leq \psi \implies \mathcal{T}_{t, r} \varphi \leq \mathcal{T}_{t, r} \psi.
\]

Suppose that \( \varphi, \psi \in D \) and there exists \( z_0 \in \mathbb{D} \) such that

\[
0 = (\varphi - \psi)(z_0) = \max_{\mathbb{D}} (\varphi - \psi).
\]

Then \( \varphi(z_0) = \psi(z_0) \), \( \varphi \leq \psi \) and by monotonicity, \( \mathcal{T}_{t, r} \varphi \leq \mathcal{T}_{t, r} \psi \) for every \( t \leq r \). By (2.6),

\[
\mathcal{L}_t \varphi(z_0) = \lim_{h \downarrow 0} \frac{\mathcal{T}_{t, t+h} \varphi(z_0) - \varphi(z_0)}{h} \leq \lim_{h \downarrow 0} \frac{\mathcal{T}_{t, t+h} \psi(z_0) - \psi(z_0)}{h} = \mathcal{L}_t \psi(z_0).
\]
Hence we proved that for every $t$, any infinitesimal generator $L_t$ of a Markov process satisfies the maximum principle: For any functions $\varphi$, $\psi$ in the domain of $L_t$, and $z_0 \in D$ satisfying (2.9), we have

$$L_t \varphi(z_0) \leq L_t \psi(z_0).$$

(2.10)

This is essentially the only important restriction for equation (2.2) to have a Feynman-Kac-type representation. Also, it is important to note that the maximum principle is the crucial property for the development of viscosity solutions as well. Moreover, the maximum principle can be directly extended to nonlinear operators and this extension will be discussed in the preceding section.

To see the importance of the maximum principle, let us consider the example of a partial differential operator. So suppose that $D = \mathbb{R}^d$ and $L_t$ be given by

$$L_t \varphi(x) = H(t, x, \varphi(x), D\varphi(x), \ldots, D^k \varphi(x))$$

for some given function $H$. Then, by calculus, we see that $L_t$ has maximum principle if and only if $k = 2$ and

$$H(t, x, u, p, B + B') \leq H(t, x, u, p, B) \quad \forall B' \geq 0.$$

(2.11)

(Here and in the rest of the chapter, for symmetric matrices inequalities are understood in the sense of quadratic forms.) This property means that the corresponding equation is a second-order (possibly degenerate) parabolic equation. These equations are related to diffusion processes that will be discussed in Section 2.3. There are nonlocal operators that have maximum principle and some examples will be discussed in Sections 2.5 and 2.6.

Also, the infinitesimal generators of Markov processes, again by definition, are translation invariant. Indeed, for any $\varphi$ in the domain of $L_t$ and a constant $\beta$, $T_{t,T}(\varphi + \beta) = (T_{t,T} \varphi) + \beta$. Hence

$$L_t(\varphi + \beta) = \lim_{h \downarrow 0} \frac{T_{t,t+h}(\varphi + \beta) - (\varphi + \beta)}{h} = \lim_{h \downarrow 0} \frac{T_{t,t+h} \varphi - \varphi}{h} = L_t \varphi,$$

and therefore, the infinitesimal operators of Markov processes do not contain any zeroth-order terms. However, with a minor modification in the definition of the semigroup, a zeroth-order term and a forcing function can be included in the theory. This is the subject of the next subsection.

### 2.2. Zeroth-order term and forcing

Let $\{Z^{t,z}(s)\}_{s \geq t}$ be as in the previous subsection. To include a term $r(t, z)u(t, z) + h(t, z)$ to equation (2.2), we modify the Markov semigroup as follows. Define random variables,

$$B(t, T; z) := \exp \left( - \int_t^T r(s, Z^{t,z}(s)) \, ds \right)$$
and 

\[ H(t, T; z) := \int_t^T B(t, s; z) h(s, Z^{t,z}(s)) \, ds. \]

For a continuous bounded function \( \varphi \), define a two parameter semigroup by

\[ \mathcal{T}_{t,r} \varphi(z) := E[B(t, r; z)] \varphi(Z^{t,r}(z)) + H(t, r; z). \]

To prove the semigroup property, for \( t \leq r \leq T \), observe that

\[ \mathcal{T}_{t,r} (\mathcal{T}_{r,T} \varphi)(z) = E[B(t, r; z)] \mathcal{T}_{r,T} \varphi(Z^{t,z}(r)) + H(t, r; z). \]

By the Markov property of \( Z(\cdot) \) (or equivalently (2.5)),

\[ B(t, T; z) = B(t, r; z) B(r, T; Z^{r,z}(T)), \]

\[ H(t, T; z) = B(t, r; z) H(r, T; Z^{r,z}(T)) + H(t, r; z). \]

Hence, we have

\[ \mathcal{T}_{t,r} (\mathcal{T}_{r,T} \varphi)(z) = E[E[B(t, T; z)] \varphi(Z^{t,z}(T)) + H(t, r; z)] = \mathcal{T}_{t,r} \varphi(z). \]

Now the infinitesimal generator of this semigroup is given by

\[ \tilde{L}_t \varphi(z) = \lim_{h \downarrow 0} \frac{\mathcal{T}_{t+h,t} \varphi(z) - \varphi(z)}{h} \]

\[ = \lim_{h \downarrow 0} \frac{1}{h} E[B(t, t+h; z)] \varphi(Z^{t,z}(t+h)) + \varphi(Z^{t,z}(t)) - \varphi(z)] \]

\[ + \lim_{h \downarrow 0} \frac{1}{h} E[H(t, t+h; z)] + \lim_{h \downarrow 0} \frac{1}{h} E[\varphi(Z^{t,z}(t+h)) - \varphi(z)] \]

\[ = r(t, z) \varphi(z) + h(t, z) + \mathcal{L}_t \varphi(z). \]

### 2.3. Diffusions and parabolic PDEs

In this section, \( \mathbb{D} = \mathbb{R}^d \) and \( Z = X \) is diffusion process satisfying the stochastic differential equation (SDE),

\[ dX(s) = \mu(s, X(s)) \, ds + \sigma(s, X(s)) \, dW(s), \quad \forall s > t, \quad (2.12) \]
with initial condition

\[ X(t) = x. \]  

(2.13)

We assume that

\[ \mu : \mathbb{R}^+ \times \mathbb{R}^d \mapsto \mathbb{R}^d, \quad \sigma : \mathbb{R}^+ \times \mathbb{R}^d \mapsto \mathcal{M}^{d,k} \]

are given functions satisfying usual conditions (cf. [44]) and \( W(\cdot) \) is a \( \mathbb{R}^k \)-valued Brownian motion. Then the infinitesimal generator of this process is

\[ \mathcal{L}_t = \mu(t,x) \cdot \nabla + \frac{1}{2} a(t,x) : D^2, \]  

(2.14)

where for two \( d \times d \) symmetric matrices \( A \) and \( B \),

\[ A : B := \text{trace}[AB] = \sum_{i,j=1}^{d} A_{i,j} B_{i,j}, \]

\[ a_{i,j}(t,x) = \sum_{l=1}^{k} \sigma_{i,l}(t,x) \sigma_{j,l}(t,x), \]

and \( \nabla, D^2 \) are respectively the gradient and the Hessian with respect to the spatial variable \( x \). Hence, the related partial differential equation is

\[ -u_t(t,x) = \mathcal{L}_t u(t,x) \]

\[ = \mu(t,x) \cdot \nabla u(t,x) + \frac{1}{2} a(t,x) : D^2 u(t,x) \]

\[ = \sum_{i=1}^{d} \mu_i(t,x) u_{x_i}(t,x) \]

\[ + \frac{1}{2} \sum_{i,j=1}^{d} a_{i,j}(t,x) u_{x_i x_j}(t,x) \text{ on } (\mathbb{R}^d) \times (-\infty, T). \]  

(2.15)

The connection between the diffusion processes and the above equation can be proved directly by using the Itô calculus as well. Indeed, suppose that (2.15) together with the final data (2.3) has a smooth solution \( u \). Fix \( t < T \) and \( x \in \mathbb{R}^d \) and let \( \{ X(s) = X_{t,s}(s) \}_{s \geq t} \) be the solution of the stochastic differential equation (2.12), (2.13). By Itô formula (Theorem 3.3.6 in [44]),

\[ d(u(s, X(s))) = [u_s(s, X(s)) + (\mathcal{L}_s u(s, \cdot))(X(s))] ds \]

\[ + \nabla u(s, X(s)) \cdot \sigma(s, X(s)) dW(s). \]
Stochastic representations for nonlinear parabolic PDEs

By (2.15), the $ds$ term in the above equation is zero. Hence, the process $Y(s) := u(s, X(s))$ is a local martingale. Under suitable growth conditions on $u$ or on $\nabla u$, we can show that

$$Y(t) = E\left[Y(T)\right] \implies u(t, x) = E\left[u(T, X^{t,x}(T))\right] = E\left[\varphi(X^{t,x}(T))\right].$$

Note that the above proof of representation using the Itô calculus has the advantage that it also proves uniqueness under some growth conditions.

A linear term and a forcing function to equation (2.15) can be added by the technique developed in Section 2.2.

### 2.4. Initial value problems

In this subsection, we will briefly discuss how we may translate the above results to initial value problems. Consider an initial value problem

$$v_t(t, x) = \left(\tilde{L}_t(v(t, \cdot))\right)(x) \text{ on } (0, \infty) \times \mathbb{R}^d,$$

together with

$$v(0, x) = \varphi(x),$$

where

$$\tilde{L}_t = \tilde{\mu}(t, x) \cdot \nabla + \frac{1}{2} \tilde{a}(t, x) : D^2.$$

Fix $(t, x) \in (0, \infty) \times \mathbb{R}^d$ and set $\tilde{X}(s)$ be the solution of

$$d\tilde{X}(s) = \tilde{\mu}(t-s, \tilde{X}(s)) \, ds + \tilde{\sigma}(t-s, \tilde{X}(s)) \, dW(s),$$

with initial data $\tilde{X}(0) = x$. Apply the Itô rule to the process $Y(s) := v(t-s, \tilde{X}(s))$. The result is

$$dY(s) = \left[-v_t(t-s, \tilde{X}(s)) + \left(\tilde{L}_s(v(t-s, \cdot))\right)(\tilde{X}(s))\right] \, ds + (\cdots) \, dW(s).$$

Again $ds$ term is zero by the equation and the stochastic term is a local martingale. Under suitable growth assumptions,

$$Y(0) = E\left[Y(t)\right] \implies v(t, x) = E\left[v(0, \tilde{X}(t))\right] = E\left[\varphi(\tilde{X}(t))\right].$$

This result can also be directly derived from the results of Section 2.3 by a time reversal. Indeed, for a given $T$ and a solution $v$ of the above initial value problem, set

$$u(t, x) := v(T-t, x) \quad \forall t \leq T, x \in \mathbb{R}^d.$$
so that \( u \) solves (2.15) with
\[
\mu(t, x) = \tilde{\mu}(T-t, x), \quad a(t, x) = \tilde{a}(T-t, x).
\]
Let \( X(s) = X^{T-t, x}(s) \) be the solution of (2.12) with initial data \( X^{T-t, x}(T-t) = x \). Then, \( X(s) = \tilde{X}(s-T+t) \). In particular, \( X(T) = \tilde{X}(t) \), and in view of the representation proved in Section 2.3,
\[
v(t, x) = u(T-t, x) = E[\varphi(X(T))] = E[\varphi(\tilde{X}(t))].
\]

2.5. Discrete Markov processes and simply coupled equations

In this subsection, we will first consider Markov processes on a discrete set \( \{1, 2, \ldots, N\} \) and then couple these processes with diffusion processes of Section 2.3.

Let \( \nu(\cdot) \) be a Markov process on a discrete set \( \Sigma := \{1, 2, \ldots, N\} \). For \( (i, j) \in \Sigma \) and \( r \geq t \geq 0 \), let
\[
P_{i,j}(t, r) := P(\nu(r) = j | \nu(t) = i),
\]
be the transition probabilities. Assume that the rate functions
\[
p_{i,j}(t) := \frac{\partial}{\partial r} P_{i,j}(t, t)
\]
exist. Since
\[
\sum_j P_{i,j}(t, r) = 1, \quad P_{i,j}(t, r) \geq 0 = P_{i,j}(t, t) \quad \forall i \neq j, r \geq t,
\]
we conclude that
\[
p_{i,j}(t) \geq 0 \quad \text{for } i \neq j \quad \text{and} \quad p_{i,i}(t) = -\sum_{j \neq i} p_{i,j}(t).
\]
Moreover, the infinitesimal generator is given by
\[
\mathcal{L}_i \varphi(i) = \sum_{j=1}^{N} p_{i,j}(t) \varphi(j) = \sum_{j \neq i} p_{i,j}(t) \left[ \varphi(j) - \varphi(i) \right].
\]
Therefore, the simple difference equation
\[
-u_t(t, i) = \sum_{j \neq i} p_{i,j}(t) \left[ u(t, j) - u(t, i) \right] \quad \forall t \leq T, i = 1, \ldots, N, \quad (2.16)
\]
with terminal data
\[ u(T, \cdot) = \varphi(\cdot) \quad \text{on} \quad \Sigma, \]
has the representation
\[ u(t, i) = E[\varphi(\nu(t)) \mid \nu(t) = i]. \]
Also a linear term and a forcing term can be added to this equation by the exponential
discounting technique developed in Section 2.2. Hence, the representation result covers all
equations of the form
\[ -u_t(t, i) = \sum_{j=1}^{N} p_{i,j}(t)[u(t, j) - u(t, i)] \]
\[ + r_i(t)u(t, i) + h_i(t) \quad \forall t \leq T, i = 1, \ldots, N, \]  \hfill (2.17)
for given functions \( p_{i,j}(t) \geq 0 \) for \( i \neq j \) and general functions \( r_i(t), h_i(t) \) without sign
restrictions. Indeed, any solution of (2.17) with terminal data (2.16), has the stochastic
representation,
\[ u(t, i) = E[H(t, T; i) + B(t, T; i)\varphi(\nu(T)) \mid \nu(t) = i], \]
where as in Section 2.2,
\[ B(t, T; i) := \exp\left(\int_{t}^{T} r_{\nu(s)}(s) \, ds\right), \]
\[ H(t, T; i) := \int_{t}^{T} B(t, s; i)h_{\nu(s)}(s) \, ds. \]
We will now combine the above representation with the results of Section 2.3 to obtain
a representation for a simply couple system of parabolic equations as well. Indeed, let
\( Z := (X, \nu) \in \mathbb{R}^d \times \{1, 2, \ldots, N\} \) be a Markov process constructed as follows. Let \( \nu \in \{1, 2, \ldots, N\} \) be a discrete Markov process as above, and for each \( i \in \{1, 2, \ldots, N\} \), \( X_{i,x}^t \)
be an independent diffusion processes solving the SDE
\[ dX_{j,x}^t(s) = \mu_i(s, X_{j,x}^t(s)) \, ds + \sigma_i(s, X_{j,x}^t(s)) \, dW(s), \]
with initial condition (2.13). Given an initial condition \( t, z := (x, i) \), we start the process
\[ Z^{t,z}(t) = (X(t), \nu(t)) = z. \]
Then, there are strictly increasing stopping times \( t < \tau_1 < \tau_2 < \cdots \), so that
\[ \nu(s) = i, \quad s \in [t, \tau_1), \]
\[ \nu(s) = i_1, \quad s \in [\tau_1, \tau_2), \ldots, \nu(s) = i_N, \quad s \in [\tau_N, \tau_{N+1}). \]
Given these stopping times, we define a continuous $X$ process recursively by

$$X(s) = X^{i,x}_{T_1}(s), \quad s \in [t, \tau_1],$$

$$X(s) = X^{N,X_{\tau_N}}_{\tau_N}(s), \quad s \in (\tau_N, \tau_{N+1}], N = 1, 2, \ldots.$$ 

It is clear that $Z = Z^{t,x,i}(\cdot)$ is a Markov process with an infinitesimal generator

$$L_t \psi(x, i) = \mu_i(t, x) \cdot \nabla \psi(x, i)$$

$$+ \frac{1}{2} a_{ij}(t, x) : D^2 \psi(x, i) + \sum_{j \neq i} p_{ij} [\psi(x, j) - \psi(x, i)].$$

Therefore

$$u(t, x, i) = E[\psi(X(T), \nu(T)) | Z(t) = (X(t), \nu(t)) = (x, i)],$$

is a solution of the coupled parabolic equation

$$-u_t(t, x, i) = (Lu(t, \cdot, \cdot))(x, i) \quad \forall t < T, (x, i) \in \mathbb{R}^d \times \{1, 2, \ldots, N\},$$

with final data

$$u(T, x, i) = \psi(x, i).$$

Notice that one may see the above equation as a system of coupled parabolic equations with a solution

$$v(t, \cdot) := (u(t, \cdot, 1), \ldots, u(t, \cdot, N)) : \mathbb{R}^d \mapsto \mathbb{R}^N$$

for all $t$. However, this coupling is only through the zeroth-order terms and the coupling constants $p_{ij}$'s are all nonnegative. For that reason, we would like to view the above equation as a scalar valued function

$$u(t, \cdot, \cdot) : \mathbb{R}^d \times \{1, 2, \ldots, N\} \mapsto \mathbb{R}^1.$$

These two different point of views have been effectively used by Freidlin in his pioneering work [38] on the analysis of some reaction–diffusion equations by stochastic methods.

### 2.6. Jump Markov processes and integro-differential equations

In this subsection, we will consider Markov processes which solve a stochastic differential equation which is more general than the one considered in the Section 2.3. This is done by adding a stochastic integral to the standard diffusion equation (2.12). This stochastic
integral is generally independent of the Brownian motion and it is driven by a random martingale measure. In particular, this measure contains jump terms and as such they generalize all the processes considered in the previous subsections.

Precisely, let $\pi$ a positive Borel measure on $\mathbb{R}^d$, called the \textit{compensator}, satisfying

$$
\int_{\mathbb{R}^d} \left[ 1 \wedge |\xi|^2 \right] \pi(d\xi) < \infty.
$$

Given this compensator measure $\pi$, there exists a random counting measure $p$ on the Borel subsets of $\mathbb{R}^+ \times \mathbb{R}^d$ so that, for any Borel set $A \subset \mathbb{R}^+ \times \mathbb{R}^d$, $p(A)$ has a Poisson distribution with mean

$$
\lambda(A) := \int_A \pi(d\xi) \, dt.
$$

Moreover, $\tilde{p} := p - \lambda$ is a martingale measure. We refer to a manuscript of Skorokhod [62] for a construction of such measures, or to a recent book by Oksendall and Sulem [55].

In the manuscript of Skorokhod [62] and in the paper of Fujiwara and Kunita [40] existence and uniqueness of stochastic differential equations are also proved. Indeed, let $\mu, \sigma$ be as in Section 2.3 and let

$$
f : \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d
$$

be a function satisfying

$$
\int_{\mathbb{R}^d} \left[ 1 \wedge |f(t,x,\xi)|^2 \right] \pi(d\xi) < \infty.
$$

We assume the standard Lipschitz condition

$$
|\mu(t,x) - \mu(t,y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 + \int_{\mathbb{R}^d} \left| f(t,x,\xi) - f(t,y,\xi) \right|^2 \pi(d\xi) \leq C|x - y|^2
$$

for all $t \in \mathbb{R}^1, x, y \in \mathbb{R}^d$ for some constant $C$. Then there exists a unique solution to

$$
X(r) = x + \int_t^r \mu(s, X(s)) \, ds + \int_t^r \sigma(s, X(s)) \, dW(s)
$$

$$
+ \int_t^r \int_{\mathbb{R}^d} f(s, X(s), \xi) \tilde{p}(ds \times d\xi) \quad \forall r \geq t,
$$

(2.18)

for any initial condition $(t, x)$, a random measure $\tilde{p}$ constructed as above and an independent standard Brownian motion $W(\cdot)$. 

Moreover, the process $X$ is a Markov process with an infinitesimal generator,

$$L_t \phi(x) = \mu(t, x) \cdot \nabla \phi(x) + \frac{1}{2} a(t, x) : D^2 \phi(x) + \tilde{L}_t \phi(x),$$

where the part corresponding to the random measure is given by

$$\tilde{L}_t \phi(x) = \int_{\mathbb{R}^d} \left[ \phi(x + f(t, x, \xi)) - \phi(x) - f(t, x, \xi) \cdot \nabla \phi(x) \right] \pi(d\xi).$$

(2.19)

See for instance, p. 94 of [62] or [55] for a proof. Hence, we have the stochastic representation discussed earlier for the integro-differential equation,

$$-u_t(t, x) - \mu(t, x) \cdot \nabla u(t, x) - \frac{1}{2} a(t, x) : D^2 u(t, x) = \left( \tilde{L}_t u(t, \cdot) \right)(x) = 0.$$  

A linear term and forcing can be added as before and also with further coupling with a discrete Markov process would yield a system of integro-differential equations.

### 2.7. Dirichlet boundary conditions

In the previous subsections, for the ease of exposition, we restricted our discussion to problems on all of $\mathbb{R}^d$. However, with a simple absorption rule at the boundary of a given region, we can include all Dirichlet problems into this theory. In this subsection, we outline the main tools that can be used for almost all processes. Indeed, let $O$ be an open set with smooth boundary and $L_t$ be as in (2.14) and consider the boundary value problem

$$-u_t(t, x) - \mu(t, x) \cdot \nabla u(t, x) - \frac{1}{2} a(t, x) : D^2 u(t, x) = 0 \quad \forall t < T, x \in O,$$

(2.20)

together with the terminal data

$$u(T, x) = \varphi(x) \quad \forall x \in O,$$

and the boundary condition,

$$u(t, x) = g(t, x) \quad \forall t < T, x \in \partial O,$$

(2.21)

for some given function $g$. Usually, we require a compatibility condition, $g(T, x) = \varphi(x)$ for all $x$ on the boundary of $O$. We can now view the solution $u(t, \cdot)$ as the value of the semigroup $T_{t, T}$ applied to the terminal data $\varphi$. We include the boundary conditions in the definition of the domain of this semigroup, which can be taken as

$$C_g(t, \cdot) := \{ v : C(\overline{O}) | v(x) = g(t, x) \forall x \in \partial O \}.$$
The stochastic semigroup is defined as follows: given $\varphi$, $g$, $T$ as above, and an initial condition $x \in \mathbb{R}^d$, $t < T$, let $X^{t,x}(\cdot)$ be the solution of the SDE (2.12), (2.13). Let $\theta$ be the exit time from the domain $O \times [t, T]$,

$$\theta := \inf \{ s \geq t : X^{t,x}(s) \notin O \} \wedge T.$$ 

Set

$$T_{t,T} \varphi(x) := E\left[ \varphi(X^{t,x}(T)) \chi_{\{\theta = T\}} + g(\theta, X^{t,x}(\theta)) \chi_{\{\theta < T\}} \right].$$

Under suitable growth and regularity conditions, we can show that the two semigroups are equal to each other. Thus, we have a stochastic representation for the boundary value problem. Integro-differential equations can be dealt with similarly. However, boundary data on all of $\mathbb{R}^d \setminus O$ is needed instead of data only on $\partial O$.

**2.8. Neumann condition and the Skorokhod problem**

Neumann-type boundary conditions are included into the theory with some care. Consider the same parabolic equation with boundary condition

$$-u_\nu(t,x) := -\nu(t,x) \cdot \nabla u(t,x) = g(t,x) \quad \forall t < T, x \in \partial O,$$

for some given function $g$ and a given unit vector field $\nu(t,x)$. We require that

$$\nu(t,x) \cdot n(x) > 0,$$

where $n(x)$ is unit inward normal to the boundary $\partial O$ at $x \in \partial O$. To obtain a representation we use the local time on $\partial O$. Indeed, we modify the SDE (2.12) in the following way.

Given an initial condition $x \in O$ and $t < T$, we look for continuous processes $X^{t,x}(\cdot)$ and $l(\cdot)$ satisfying (2.13) and for $s \in [t, T]$,

$$X^{t,x}(s) = x + \int_t^s \mu(r, X^{t,x}(r)) \, dr + \int_t^s \sigma(r, X^{t,x}(r)) \, dW(r)$$

$$+ \int_t^r \nu(r, X^{t,x}(r)) \, dl(r) \in \overline{O},$$

$$l(s) = \int_t^s \chi_{\{X^{t,x}(r) \in \partial O\}} \, dl(r),$$

$$l(0) = 0 \quad \text{and} \quad l \text{ is nondecreasing and continuous.}$$

In the literature the solution $X^{t,x}(\cdot)$ called the reflected diffusion process, $l$ is the local time and the above set of equations are called the Skorokhod problem. Under the usual Lipschitz conditions on $\mu$, $\sigma$, $\nu$, and smoothness assumption on the boundary $\partial O$, the Skorokhod problem has a unique solution. This and more was proved by Lions and Sznitman [52].
We also refer to Lions [48] for the connection to partial differential equations and viscosity solutions.

The only difference between (2.24) and (2.12) is the last $dl$ integral and the important requirement that $X^{t,x}(s) \in \overline{O}$ for all $s$. Notice that (2.25) ensures that $dl$ increases only when the diffusion processes $X^{t,x}$ is on the boundary $\partial O$. Hence, formally $X^{t,x}$ processes is “reflected” on the boundary $\partial O$ in the direction $v(t,x)$. In view of the condition (2.23) and the fact that $n$ is the inward normal, the reflection direction $v(t,x)$ points inward from $x \in \partial O$. These guarantee that the process $X^{t,x}$ takes values in $\overline{O}$.

We now define Markov stochastic semigroup,

$$u(t,x) = E\left[\varphi(X^{t,x}(T)) + \int_t^T g(s, X^{t,x}(s)) \, dl(s)\right].$$

We claim that any smooth solution $v \in C^{1,2}((0, T) \times \overline{O})$ of (2.15), (2.3) and (2.22) is equal to $u$. Indeed, let $X^{t,x}, l$ be a solution of the Skorokhod problem with initial data $X^{t,x}(t) = x$. Since $l$ is a monotone function, $X$ is a semimartingale and with the use of Itô’s rule we obtain

$$\varphi(X^{t,x}(T)) = v(T, X^{t,x}(T))$$

$$= v(t,x) + M(T) + \int_t^T [v_t + L v](s, X^{t,x}(s)) \, ds$$

$$+ \int_t^T \nabla v(s, X^{t,x}(s)) \cdot v(s, X^{t,x}(s)) \, dl(s),$$

where $M$ is a local martingale with $M(t) = 0$. By (2.15), the first integrand is zero, and by (2.22) the second integrand is equal to $-g(s, X^{t,x}(s))$. Also under some suitable growth conditions $E[M(T)] = 0$. We use these observations and then take the expected value. The result is $v = u$.

Once again, a linear term and a forcing function can be added into the theory as in the Section 2.2.

2.9. Stationary problems

Time homogeneous linear problems also have similar stochastic representations. Indeed, let $r(x) \geq \beta > 0$ be a given function. Let $Z^x(\cdot)$ be a time homogeneous Markov process with infinitesimal generator $L$ and initial condition $Z^x(0) = z$. Given an open set $O$ with smooth boundary, consider the boundary value problem

$$r(z)u(z) - Lu(z) = h(z) \quad \forall z \in O,$$

together with the boundary condition,

$$u(z) = g(z) \quad \forall z \in \mathbb{R}^d \setminus O,$$

(2.26)
for some given function \( g \).

To obtain a stochastic representation, let \( \theta \) be the exit time from the domain \( O \),

\[
\theta^z := \inf \{ s \geq 0 : Z^z(s) \notin O \}.
\]

Set

\[
u(z) := E \left[ \int_0^{\theta^z} B(s) h \left( Z^z(s) \right) ds + B(\theta^z) g \left( Z^z(\theta^z) \right) \chi_{\{\theta^z < \infty\}} \right].
\]

where

\[
B(t) := \exp \left( -\int_0^t r(s, Z^z(s)) ds \right).
\]

Notice that due our strict positivity assumption on \( r \), \( B(s) \geq e^{-\beta s} \) for all \( s \geq 0 \). Therefore, the integral term in the above expression is integrable under reasonable growth assumptions on \( h \).

For \( T > 0 \), define a stochastic semigroup by

\[
T_T \varphi(z) := E \left[ \int_0^{T \wedge \theta^z} B(s) L \left( Z^z(s) \right) ds \right.
\]

\[
+ B(\theta^z) \varphi \left( Z^z(\theta^z) \right) \chi_{\{\theta^z < T\}} + B(T) g \left( Z^z(T) \right) \chi_{\{\theta^z \geq T\}} \right].
\]

Time homogeneity of the Markov process implies the semigroup property,

\[
T_{T+S} \varphi = T_T (T_S \varphi) \quad \forall T, S \geq 0.
\]

Also, \( u \) is a fixed point of this semigroup for every \( T \); that is, \( u = T_T u \) for all \( T \geq 0 \). Then, under suitable growth and regularity conditions, we can show that \( u \) is the unique solution of the linear equation. Thus, we have a stochastic representation for the stationary boundary value problem as well. For partial differential equation, the boundary condition is needed only on \( \partial O \). Neumann boundary conditions are handled as in the previous subsection.

### 3. Representation via controlled processes

In this section we will consider nonlinear equations of the form

\[
-u_t(t, z) + \mathcal{H}(t, z, u(t, \cdot)) = 0 \quad \forall (t, x) \in (-\infty, T) \times \mathbb{D}, \tag{3.1}
\]

where \( \mathbb{D} \) is as before. The most general form of the nonlinearity \( \mathcal{H} \) is of the form

\[
\mathcal{H}(t, z, \varphi(\cdot)) := \inf_{\beta \in B} \sup_{\alpha \in A} \left\{ -L_t^\alpha \varphi(z) - r(t, z, \alpha, \beta) \varphi(z) - L(t, z, \alpha, \beta) \right\}, \tag{3.2}
\]
where for a set of parameters $\alpha, \beta$ in some control sets $A$ and $B$, $\mathcal{L}_t^{\alpha,\beta}$ is the infinitesimal generator of Markov process on $\mathbb{D}$, as in Section 2.1, and $r$ and $L$ are given functions. These equations are related to stochastic differential games. An interesting and an important class of equations related to stochastic optimal control is obtained by taking $B$ to be a singleton. In this case, the nonlinearity reduces to

$$\mathcal{H}(t, z, \varphi(\cdot)) := \sup_{\alpha \in A} \left\{ -\mathcal{L}_t^{\alpha} \varphi(z) - r(t, z, \alpha)\varphi(z) - L(t, z, \alpha) \right\}. \quad (3.3)$$

In this chapter we only discuss operators of the above form. For differential games, we refer the interested reader to Chapter 11 in the second edition of [36].

Finally, note that the nonlinearity in (3.2) has the maximum principle as defined in Section 2.1 (cf. (2.10)). Recall that as a consequence of maximum principle all local operators $\mathcal{H}$ on $\mathbb{D} = \mathbb{R}^d$ of the above form must be given by

$$\mathcal{H}(t, x, \varphi(\cdot)) = H(t, x, \varphi(x), \nabla \varphi(x), D^2 \varphi(x)).$$

for some given function $H$ satisfying (2.11).

Consider the nonlinear equation (3.1). If this equation together with the terminal data (2.3) has a unique solution, then we define a nonlinear semigroup acting on the terminal data $\varphi$ by

$$\hat{T}_{T, t}(\varphi)(z) := u(t, z).$$

By uniqueness, this is a semigroup. To obtain the related stochastic semigroup, we consider $\mathcal{H}$ as in (3.3).

Let $A$ be the set of all bounded, progressively measurable processes $\alpha(t) \in A$. Again as in the linear case, given an initial condition $t, z$ and a processes $\alpha(\cdot) \in A$, consider a class of processes $Z^{t, z, \alpha(\cdot)}$. We assume that, for every fixed $\alpha \in A$, the infinitesimal generator of the processes $Z^{t, z, \alpha}$ is equal to $\mathcal{L}_t^{\alpha}$, where $\bar{\alpha}$ is process which is equal to the constant $\alpha$ everywhere. Define the value function $v$ by

$$v(t, z) := \inf_{\alpha \in A} \left( J_{T, t}^{\alpha}(\varphi)(z), \right) \quad (3.4)$$

where with $Z = Z^{t, z, \alpha(\cdot)}$,

$$J_{T, t}^{\alpha}(\varphi)(z) := E \left[ \int_t^T B(s) L(s, z(s), \alpha(s)) \, ds + B(T) \varphi(Z(T)) \right],$$

$$B(T) = B(t, T; z, \alpha(\cdot)) = \exp \left( - \int_t^T r(s, z(s), \alpha(s)) \, ds \right).$$

Bellman’s dynamic programming in this context states that, for any stopping time
Stochastic representations for nonlinear parabolic PDEs

\[ \theta \in [t, T], \]

\[ v(t, z) = \inf_{\alpha(\cdot)} \left\{ E \left[ \int_t^\theta B(s)L(s, Z(s), \alpha(s)) \, ds + B(\theta)v(\theta, Z(\theta)) \right] \right\}, \]

\[ = \inf_{\alpha(\cdot)} \left\{ (\mathcal{J}^a_{t, \theta}(v(\theta, \cdot)))(z) \right\}. \]

We refer to [66] for a general abstract proof of dynamic programming under some structural assumptions. The crucial structure needed to prove the above result is the additive structure given by the Markov assumption on the process \( Z \) and the fact that

\[ B(t, T; z, \alpha(\cdot)) = B(t, \theta; z, \alpha(\cdot))B(\theta, T; Z_t, z, \alpha(\cdot), \alpha(\cdot)). \]

Now we define the stochastic semigroup by

\[ (T_{t,T} \varphi) = v(t, \cdot) = \inf_{\alpha(\cdot)} (\mathcal{J}^a_{t, T}(\varphi)), \]

so that the dynamic programming principle implies that, for any stopping time \( \theta \in [t, T], \)

\[ (T_{t,T} \varphi) = (T_{t,\theta} u(\theta, \cdot)) = (T_{t,\theta} [T_{\theta,T} \varphi]). \]

Hence, dynamic programming implies that \( T_{t,T} \) is a semigroup. Indeed, the dynamic programming principle and the semigroup properties are essentially equivalent.

It now remains to show that the infinitesimal generator of this stochastic semigroup agrees with \( \mathcal{H} \) given in (3.3). We show this connection only formally here. We then introduce the theory of viscosity solutions of Crandall and Lions to prove it rigorously. Indeed, we need to compute the following limit,

\[ \lim_{\theta \downarrow t} \frac{(T_{t,\theta} \varphi)(z) - \varphi(z)}{\theta - t} \]

\[ = \lim_{\theta \downarrow t} \inf_{\alpha(\cdot) \in \mathcal{A}} \left[ \frac{E \int_t^\theta B(s)L(s, Z(s), \alpha(s)) \, ds}{\theta - t} + \frac{E(B(\theta)\varphi(Z(\theta)) - \varphi(z))}{\theta - t} \right]. \]

Now, formally, assume that we may interchange the order of limit and infimum. Also, again formally, assume that the infimum can be taken only over all \( \tilde{\alpha} \), where \( \tilde{\alpha}(s) = \alpha \) for all \( s \).

Then, formally,

\[ \lim_{\theta \downarrow t} \frac{(T_{t,\theta} \varphi)(z) - \varphi(z)}{\theta - t} = \inf_{\alpha(\cdot) \in \mathcal{A}} \lim_{\theta \downarrow t} \left[ J_1(t, \theta, \alpha)(z) + (J_2(t, \theta, \alpha)\varphi)(z) \right], \]

where

\[ J_1(t, \theta, \alpha)(z) = \frac{E \int_t^\theta B(s)L(s, Z(s), \alpha) \, ds}{\theta - t}. \]
and
\[(J_2(t, \theta, \alpha)\varphi)(z) = \frac{E[B(\theta)\varphi(Z(\theta)) - \varphi(z)]}{\theta - t},\]

Since the infinitesimal generator of the controlled process \(Z\) with control process \(\alpha\) is assumed to be \(\mathcal{L}_t^\alpha\), the limit of \(J_2\) is equal to \(\mathcal{L}_t^\alpha \varphi(z) + r(t, z, \alpha)\varphi(z)\). Also, by the continuity of the processes, the limit of \(J_1\) is equal to \(L(t, z, \alpha)\). Hence, formally, we compute the infinitesimal generator is
\[
\lim_{\theta \downarrow t} \frac{(T_{t,\theta} \varphi)(z) - \varphi(z)}{\theta - t} = \inf_{\alpha \in \mathcal{A}} \{ L_t^\alpha \varphi(z) + r(t, z, \alpha)\varphi(z) + L(t, z, \alpha) \} = -\mathcal{H}(t, z, \varphi(\cdot)).
\]

Hence, we have shown that the generators of the semigroup \(\hat{T}_{t,T}\) related to the PDE and the stochastic semigroup \(T_{t,T}\) have the same infinitesimal generator. Therefore, as in the linear case, if the PDE has unique solution in a certain class containing \(v\) then, the value function \(v\) is the unique solution of the PDE.

The main focus of the preceding subsections, is to make the above calculations rigorous and to extend these results to general nonlinearities.

### 3.1. Viscosity solutions

This subsection follows very closely [36].

Let \(\mathcal{D}\) be closed subset of a Banach space and \(\mathcal{C}\) be a collection of functions on \(\mathcal{D}\) which is closed under addition, i.e.,
\[
\phi, \psi \in \mathcal{C} \Rightarrow \phi + \psi \in \mathcal{C}.
\]

As in the previous sections, the main object of our analysis is a two parameter family of operators \(\{T_{t,r}: t \leq r \leq T\}\) with the common domain \(\mathcal{C}\). In the applications the exact choice of \(\mathcal{C}\) is not important. However, when \(\mathcal{D}\) is compact, we will require that \(\mathcal{C}\) contains \(C(\mathcal{D})\). For noncompact \(\Sigma\), additional conditions are often imposed. Indeed, in most of our examples, we will require that \(\mathcal{C}\) contains \(\mathcal{M}(\mathcal{D}) \cap C_p(\mathcal{D})\) (\(\mathcal{M}(\mathcal{D})\) is set of all real-valued functions which are bounded from below, \(C_p(\mathcal{D})\) is set of all continuous, real-valued functions which are polynomially growing). We assume that \(T_{tt}\) is the identity.

Next we want to state the semigroup property. When, \(T_{rT} \varphi\) belongs to \(\mathcal{C}\), the semigroup property is (2.4). However, \(T_{rT} \varphi\) may not be in the domain. So, in general we assume that, for all \(\phi, \psi \in \mathcal{C}\) and \(t \leq r \leq t \leq s \leq T\),
\[
T_{tr} \phi \leq T_{ts} \psi \quad \text{if} \quad \phi \leq T_{rs} \psi, \quad (3.5)
\]
\[
T_{tr} \phi \geq T_{ts} \psi \quad \text{if} \quad \phi \geq T_{rs} \psi. \quad (3.6)
\]
By taking \( r = s \) in (3.2) we conclude that the above conditions imply \( T_{tr} \) is monotone, (2.8). Moreover, if \( T_{rs} \psi \in \mathcal{C} \), by taking \( \phi = T_{rs} \psi \), we obtain (2.4). So, in general, (3.5), (3.6) is a convenient way of stating monotonicity and the semigroup properties.

All the linear semigroups introduced in the previous section satisfy the above conditions. We will now give the example of a semigroup generated by stochastic optimal. This example will then be studied in detail in the next subsection.

EXAMPLE 3.1 (Controlled diffusion processes). We follow a construction as in (3.4). Let \( \mathcal{A} \) be a control set, \( O \) be an open subset of \( \mathbb{R}^d \). Set \( \mathcal{D} = \overline{O} \) and \( \mathcal{C} = \mathcal{M}(\Sigma) \), set of all measurable functions bounded from below. Let \( \mu, \sigma, L, g \) be functions satisfying the standard Lipschitz conditions (see [36]), i.e., for any function

\[
\phi : (-\infty, T] \times \mathbb{D} \times \mathcal{A} \mapsto \mathcal{M},
\]

where \( \mathcal{M} \) is any normed space (in our applications \( \mathcal{M} \) is either \( \mathbb{R}^d \) or the set of real matrices with usual norm), we say that \( \phi \) satisfies the standard Lipschitz condition if \( \phi \) is continuous in the \((t, x)\) variables and

\[
\| \phi(t, x, \alpha) - \phi(t, x', \alpha) \|_{\mathcal{M}} \leq C |x - x'| \quad \forall t \in (-\infty, T], x, x' \in \overline{O}, \alpha \in A,
\]

(3.7)

with a constant independent of all variables. Let \((\Omega, P, \mathcal{F})\) be a probability space, \( W(\cdot) \) be a standard \( \mathbb{R}^k \) Brownian motion and \( \{\mathcal{F}_t\} \) be the filtration satisfying the usual conditions as in [44].

Let \( \mathcal{A} \) be all bounded, progressively measurable, \( \mathcal{A} \)-valued random processes. We all \( \mathcal{A} \) the set of admissible controls. In some applications further restrictions on the controls are needed. These can be modeled easily by introducing \((t, x)\) depended subsets of \( \mathcal{A} \). However, in that case certain conditions must be satisfied as discussed in [36,66].

Given a process \( \alpha(\cdot) \in \mathcal{A} \) and an initial condition (2.13), we consider the controlled stochastic differential equation

\[
dX(s) = \mu(s, X(s), \alpha(s)) \, ds + \sigma(s, X(s), \alpha(s)) \, dW(s).
\]

(3.8)

For a given boundary function \( g \), a running cost function \( L \) and a function \( \psi \), set

\[
(\mathcal{J}_{t, T}^{\alpha(\cdot)} \psi)(x) = E \left[ \int_t^{\theta \wedge T} B(s)L(s, X(s), \alpha(s)) \, ds + B(\theta)g(\theta, X(\theta)) \chi_{\theta < r} + B(r)\psi(X(r)) \chi_{\theta \geq r} \right],
\]

(3.9)

where

\[
B(T) = B(t, T; z, \alpha(\cdot)) = \exp \left( -\int_t^T r(s, Z(s), \alpha(s)) \, ds \right),
\]

(3.10)
θ is the exit time of \((s, X(s))\) from \(\overline{Q} = [t, T] \times \overline{O}\). The nonlinear semigroup is given by

\[
(\mathcal{T}_{t,r} \psi)(x) := \inf_{\alpha(\cdot) \in A} J^{\alpha(\cdot)}_{t,r}(x).
\]

Assume that \(L, g\) and \(\psi\) are all bounded from below. Then, \(\mathcal{T}_{t,r} \psi\) is also bounded from below and therefore, for every \(\psi \in \mathcal{C}\), \(\mathcal{T}_{t,r} \psi\) is well defined and belongs to \(\mathcal{C}\). Clearly \(\mathcal{T}_{t,r}\) is monotone (2.8). Also, dynamic programming for optimal control (cf. [36,66]) implies the semigroup property (2.4).

Notice that the infinitesimal generator of the controlled process is

\[
L^a_t = \mu(t, x, \alpha) \cdot \nabla + \frac{1}{2} a(t, x, \alpha) : D^2,
\]

where “\(\cdot\)” is as before and

\[
a_{i,j}(t, x, \alpha) = \sum_{l=1}^{k} \sigma_{i,l}(t, x, \alpha) \sigma_{j,l}(t, x, \alpha).
\]

Hence, in view of the formal argument given in the introduction of this section, the related partial differential equation is (3.1) with \(\mathcal{H}\) as in (3.3) with the above \(L^a_t\).

In the next section, we will rigorously prove the connection between the dynamic programming equation and the above semigroup.

For \(\psi \in \mathcal{C}\), \(t \leq T\), \(x \in \mathbb{D}\), set

\[
v(t, x) = (\mathcal{T}_{t,T} \psi)(x).
\]

In analogy with control problems, we call \(v(t, x)\) the value function. Using the semigroup property, we conclude that the value function satisfies

\[
v(t, x) = (\mathcal{T}_{t,r} v(r, \cdot))(x) \quad \forall x \in \mathbb{D}, t \leq r \leq T,
\]

provided that \(v(r, \cdot) \in \mathcal{C}\). This identity is just a restatement of the dynamic programming principle when the semigroup is related to an optimal control problem. Hence, we refer to (3.12) as the (abstract) dynamic programming principle.

Having formulated the dynamic programming principle abstractly, we proceed to derive the corresponding dynamic programming equation. Let \(r = t + h\) in (3.12) for some \(h > 0\) and small. Assume that \(v(t + h, \cdot) \in \mathcal{C}\). Then

\[
-\frac{1}{h} \left[ (\mathcal{T}_{t+h} v(t + h, \cdot))(x) - v(t, x) \right] = 0,
\]

for all \(x \in \mathbb{D}\) and \(t < t + h \leq T\). To continue even formally, we need to assume that the above quantity has a limit as \(h \downarrow 0\), when \(v\) is “smooth”. So we assume that there exist an open set \(\mathbb{D}' \subset \mathbb{D}\), a set of smooth functions \(\mathcal{D} \subset C((-\infty, T) \times \mathbb{D}')\) and a one-parameter
Stochastic representations for nonlinear parabolic PDEs

family of nonlinear operators $\{G_t\}_{t \leq T}$ of functions of $\mathbb{D}$, satisfying the following conditions with

$$Q = (-\infty, T) \times \mathbb{D},$$

$$\varphi_t(t, x), \ (G_t \varphi(t, \cdot))(x) \in C(Q) \quad \text{and} \quad \varphi(t, \cdot) \in C \quad \forall t \leq T,$$

(3.14)

$$\varphi, \tilde{\varphi} \in \mathcal{D}, \lambda \geq 0 \implies \varphi + \ldots \in \mathcal{D}, \lambda \varphi \in \mathcal{D},$$

(3.15)

$$\lim_{h \downarrow 0} \frac{1}{h} \left[ \left( T_{t+h} \varphi(t+h, \cdot) \right)(x) - \varphi(t, x) \right] = \varphi_t(t, x) - \left( G_t \varphi(t, \cdot) \right)(x)$$

(3.16)

for all $\varphi \in \mathcal{D}, (t, x) \in Q$. We refer to the elements of $\mathcal{D}$ as test functions and $\mathcal{G}_t$ as the infinitesimal generator of the semigroup $\{T_{t,r}\}$. Note that, if $\varphi$ is any test function, then $\varphi(t, x)$ is defined for all $(t, x) \in (-\infty, T) \times \mathbb{D}$ even though (3.16) is required to hold only for $(t, x) \in Q$.

Like the choice of $\mathcal{C}$, the exact choice of $\mathcal{D}$ is not important. One should think of $\mathcal{D}$ as the set of “smooth” functions. For example, if $\mathbb{D}' = O$ is a bounded subset of $\mathbb{R}^d$ and $\mathbb{D} = \overline{O}$, then we require that $\mathcal{D}$ contains $C^\infty(\overline{O})$. Indeed, this requirement will be typical when $\mathcal{G}_t$ is a partial differential operator.

In most applications, $\mathbb{D}'$ is simply the interior of $\mathbb{D}$. However, in the case of a controlled jump Markov process which is stopped after the exit from an open set $O \subset \mathbb{R}^d$, we have $\mathbb{D}' = O$, while $\mathbb{D}$ is the closure of the set that can be reached from $O$.

Now suppose that $v \in \mathcal{D}$ and let $h$ go to zero in (3.13). The result is

$$-v_t(t, x) + \left( G_t v(t, \cdot) \right)(x) = 0, \quad (t, x) \in Q.$$  

(3.17)

In analogy with optimal control, the above equation is called the (abstract) dynamic programming equation.

In general, the value function is not in $\mathcal{D}$ and therefore it is not a classical solution of (3.17). In that case the equation (3.17) has to be interpreted in a weaker sense. This will be the subject of viscosity solutions.

We are now in a position to give the definition of viscosity solutions in the abstract setting. This is a straightforward generalization of the original definition given by Crandall and Lions [25]. Also see Crandall, Evans and Lions [23]. Let $Q = (-\infty, T) \times \mathbb{D}'$, $\mathcal{D}$ and $\mathcal{C}$ as before.

In the below definition, we assume continuity to simplify the presentation. However, for the definition we only need the solution to be locally bounded, see [4,36].

**Definition 3.2 (Viscosity solutions).** Let $w \in C((-\infty, T] \times \mathbb{D})$. Then

(i) $w$ is a viscosity subsolution of (3.17) in $Q$ if for each $\varphi \in \mathcal{D}$,

$$-\varphi_t(\bar{t}, \bar{x}) + \left( G_{\bar{t}} \varphi(\bar{t}, \cdot) \right)(\bar{x}) \leq 0,$$

(3.18)

at every $(\bar{t}, \bar{x}) \in Q$ which is a maximizer of $w - \varphi$ on $(-\infty, T] \times \mathbb{D}$ with $w(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x})$.

(ii) $w$ is a viscosity supersolution of (3.17) in $Q$ if for each $\varphi \in \mathcal{D}$,

$$-\varphi_t(\bar{t}, \bar{x}) + \left( G_{\bar{t}} \varphi(\bar{t}, \cdot) \right)(\bar{x}) \geq 0,$$

(3.19)
H.M. Soner

at every \((\bar{t}, \bar{x}) \in Q\) which is a minimizer of \(w - \varphi\) on \((-\infty, T] \times \mathbb{D}\) with \(w(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x})\). 

(iii) \(w\) is a viscosity solution of (3.17) in \(Q\) if it is both a viscosity subsolution and a viscosity supersolution of (3.17) in \(Q\).

It follows from the monotonicity and the semigroup properties and the definitions that any classical solution of (3.17) is also a viscosity solution, see for instance [36]. Another immediate consequence is the following.

**Theorem 3.3.** Assume (3.5), (3.6), (3.14)–(3.16). Suppose that the value function \(v\) defined by (3.11) is continuous. Then, \(v\) is a viscosity solution of (3.17) in \(Q\).

**Proof.** Let \(\varphi \in \mathcal{D}\) and \((\bar{t}, \bar{x}) \in Q\) be a maximizer of the difference \(v - \varphi\) on \(\overline{Q}\) satisfying \(v(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x})\). Then, \(\varphi \geq v\). Using (3.6) with \(\phi = \varphi(r, \cdot)\) and \(s = T\), we obtain for every \(r \in [\bar{t}, T]\),

\[
(T_{\bar{t}, r} \varphi(r, \cdot))(\bar{x}) \geq (T_{\bar{t}, T} \psi)(\bar{x}) = v(\bar{t}, \bar{x}) = \varphi(\bar{t}, \bar{x}).
\]

Recall that by (3.14), \(\varphi(r, \cdot)\) is in the domain of \(T_{\bar{t}, r}\). Take \(r = \bar{t} + h\) and use (3.16) to arrive at

\[
-\varphi_t(\bar{t}, \bar{x}) + (\mathcal{G}_{\bar{t}} \varphi(\bar{t}, \cdot))(\bar{x}) = -\lim_{h \downarrow 0} h^{-1} \left[(T_{\bar{t}, \bar{t} + h} \varphi(\bar{t} + h, \cdot))(\bar{x}) - \varphi(\bar{t}, \bar{x})\right] \leq 0.
\]

Hence (3.18) is satisfied and consequently \(v\) is a viscosity subsolution of (3.17) in \(Q\). The supersolution property of \(v\) is proved exactly the same way as the subsolution property. \(\square\)

### 3.2. Optimal control of diffusion processes

In this subsection, we will prove that the value function of an optimal control problem is the unique viscosity solution of the dynamic programming equation (3.1) with \(\mathcal{H}\) as in (3.3) with \(\mathcal{L}^\alpha\) given by (3.10). In order to achieve this, we will define a stochastic nonlinear semigroup as in the introduction of this section and in Example 3.1. Then, we will verify the assumptions of Theorem 3.3 to show that the value function of the stochastic optimal control problem is the viscosity solution.

If the controlled Markov processes are uniformly parabolic, then there are classical solutions to the dynamic programming equation (3.17) and uniqueness is standard under natural conditions well known in the PDE literature. Combined with these PDE results, the viscosity property of the value function provides a stochastic representation. However, we do not, in general, assume the uniform parabolicity and therefore we only expect the value function to be a viscosity solution. Still in this case, there are uniqueness results for viscosity solutions (see [24,36]) and a representation result follows. The main difference between the smooth (or equivalently the uniformly elliptic) case and the nonsmooth case is that, as for the linear problems, smooth solutions with certain growth conditions can be directly shown to be the value function. This point is further developed in the next subsection. In
Stochastic representations for nonlinear parabolic PDEs

1 this section, we consider the boundary value problems with no exponential discounting
2 (i.e., equation with no linear term: \( r \equiv 0 \)). However, our results easily extend to the other
3 cases as described in Section 2.
4 Now let \( T_t \) be the semigroup defined in Example 3.1. To simplify the presentation, we
take \( r \equiv 0 \). So that the nonlinear operators on \( C \) are given by
5 \[
6 (T_t \varphi)(x) = \inf_{\alpha(\cdot) \in A} E \left\{ \int_t^{\theta \wedge r} L(s, X(s), \alpha(s)) \, ds 
7 + g(\theta, X(\theta)) \chi_{\theta < r} + \varphi(X(r)) \chi_{\theta \geq r} \right\},
8\]
9 where \( t \leq r \leq T, \varphi \in C, \theta \) is the exit time of \( (s, X(s)) \) from \( Q = (-\infty, T] \times \overline{O} \) and
10 \( g \in C(\overline{Q}) \) is a given function, which we call the lateral boundary data. Clearly \( T_t \)
satisfies (3.5) and (3.6). The semigroup property however, is equivalent to the dynamic program-
11 ming principle. We refer to Chapter 5 in [36] or [66] for the general structure of dynamic
12 programming in a certain context.
13 To apply the results of Theorem 3.3, we also have to verify (3.16). Indeed we shall prove
14 that (3.16) holds with \( D = \overline{O}, D' = O \) (hence \( Q = (-\infty, T] \times O \)), \( D = C^{1,2}(\overline{Q}) \) and the
15 infinitesimal generator,
16 \[
17 (G_t \varphi)(x) = \mathcal{H}(t, x, \varphi(\cdot)), \quad (t, x) \in Q,
18\]
19 where \( \mathcal{H} \) is as in (3.3) and \( \mathcal{L}_t^g \) as in (3.10). In view of Theorem 3.3, this result will im-
20 ply that the value function is a viscosity solution of the dynamic programming equation
21 provided that it is continuous.
22 Recall \( O \) is assumed to be bounded. For the unbounded case, we refer to [36].
23
24 THEOREM 3.4. Suppose that \( f, \sigma \) satisfy (3.7), \( A \) is compact and \( g, L \) are continuous.
25 Then, for every \( w \in D \) and \( (t, x) \in Q \), we have
26 \[
27 \lim_{h \downarrow 0} \frac{1}{h} [ (T_{t+h} w(t+h, \cdot))(x) - w(t, x) ] = w_t(t, x) - (G_t w(t, \cdot))(x).
28\]
29
30 PROOF. We start with a probabilistic estimate. Let \( X(\cdot) \) be the solution of (3.8) with con-
31 trol \( \alpha(\cdot) \) and initial condition \( X(t) = x \in O \). Since \( Q \) is bounded, \( f \) and \( \sigma \) are bounded,
32 for any positive integer \( m \) and \( h \in (0, 1] \), we have
33 \[
34 E \sup_{t \leq \rho \leq t+h} |X(\rho) - x|^{2m} 
35 = E \sup_{t \leq \rho \leq t+h} \left| \int_t^{\rho} f(s, X(s), \alpha(s)) \, ds 
36 + \int_t^{\rho} \sigma(s, X(s), \alpha(s)) \, dW(s) \right|^{2m}
37\]
38
\[ H.M. \text{ Soner} \]

\[
\leq \hat{C}_m E \left( \int_t^{t+h} \left| f(s, X(s), \alpha(s)) \right|^2 \, ds \right)^{2m}
+ \hat{C}_m E \sup_{t \leq \rho \leq t+h} \left( \int_t^{\rho} \left| \sigma(s, X(s), \alpha(s)) \right| \, dW(s) \right)^{2m}
\leq \hat{C}_m \| f \|^2 2m h^{2m} + \hat{C}_m \| \sigma \|^2 2m h^m \leq C_m h^m, \tag{3.20}
\]

where \( \| \cdot \| \) denotes the sup-norm on \( Q \) and \( C_m, \hat{C}_m, \tilde{C}_m \) are suitable constants. Set \( d(x) = \text{dist}(x, \partial O) \) and recall that \( \theta \) is the exit time from \( Q \). Then for \( t + h \leq T \),

\[
P(\tau \leq t + h) \leq P \left( \sup_{t \leq \rho \leq t+h} |X(\rho) - x| \geq d(x) \right)
\leq \left( E \sup_{t \leq \rho \leq t+h} |X(\rho) - x|^{2m} \right) (d(x))^{-2m}
\leq \frac{C_m h^m}{(d(x))^{2m}}. \tag{3.21}
\]

Fix \( \alpha \in A \) and let \( \bar{\alpha} \equiv \alpha \). Then the definition of \( T_{t,t+h} \) yields

\[
I(h) = \frac{1}{h} \left[ (T_{t,t+h} w(t + h, \cdot))(x) - w(t, x) \right]
\leq \frac{1}{h} E \int_t^{(t+h)\land \theta} L(s, X(s), \alpha) \, ds
+ \frac{1}{h} E \left[ w(t + h, X(t + h)) - w(t, x) \right] \chi_{\theta \geq t+h}
+ \frac{1}{h} E \left[ g(\theta, X(\theta)) - w(t, x) \right] \chi_{\theta < t+h}. \tag{3.22}
\]

The estimate (3.21) with \( m = 2 \) yields

\[
\lim_{h \downarrow 0} \frac{1}{h} P_t \left( \tau \leq t + h \right) = 0
\]

for every \( (t, x) \in Q \). Hence

\[
\lim_{h \downarrow 0} \frac{1}{h} E \int_t^{(t+h)\land \theta} L(s, X(s), \alpha) \, ds = L(t, x, \alpha)
\]

and

\[
\lim_{h \downarrow 0} \frac{1}{h} E \left[ g(\theta, X(\theta)) - w(t, x) \right] \chi_{\theta < t+h} = 0.
\]
Also, by Itô’s formula

\[ \lim_{h \downarrow 0} \frac{1}{h} E \left[ w(t + h, x(t + h)) - w(t, x) \right] \mathbf{1}_{\theta \geq t + h} \]

\[ = \lim_{h \downarrow 0} \frac{1}{h} E \left[ w((t + h) \wedge \theta, X((t + h) \wedge \theta)) - w(t, x) \right] \]

\[ = \lim_{h \downarrow 0} \frac{1}{h} \int_t^{(t+h) \wedge \theta} \left[ w_t(s, X(s)) + (L^\alpha t w(s, \cdot))(X(s)) \right] ds \]

\[ = \left( L^\alpha t w(t, \cdot) \right)(x). \]

Substitute the above into (3.21) to obtain

\[ \limsup_{h \downarrow 0} I(h) \leq L(t, x, \alpha) + w_t(t, x) + \left( L^\alpha t w(t, \cdot) \right)(x) \]

for all \( \alpha \in A \). We take the infimum over \( \alpha \). The result is

\[ \limsup_{h \downarrow 0} I(h) \leq w_t(t, x) - \left( G_t w(t, \cdot) \right)(x). \]

For any sequence \( h_n \downarrow 0 \), there exists \( \alpha_n(\cdot) \) satisfying

\[ (T_{t, t_n} w(t_n, \cdot))(x) \]

\[ \geq E \left[ \int_t^{\theta_n} L(s, x_n(s), \alpha_n(s)) ds + g(\theta_n, X_n(\theta_n)) \mathbf{1}_{\theta_n < t_n} \right. \]

\[ \left. \mathbf{1}_{\theta_n = t_n} \right] - (h_n)^2, \]

where \( t_n = t + h_n, \theta_n = \hat{\theta}_n \wedge t_n, X_n(\cdot) \) is the solution of (3.8), (2.13) with control \( \alpha_n \), and \( \hat{\theta}_n \) is the exit time of \((s, X_n(s))\) from \( Q \). Therefore

\[ I(h_n) \geq \frac{1}{h_n} E \int_t^{\theta_n} L(s, x_n(s), \alpha_n(s)) ds \]

\[ + \frac{1}{h_n} E \left[ w(t_n, X(t_n)) - w(t, x) \right] \mathbf{1}_{\theta_n = t_n} \]

\[ + \frac{1}{h_n} E \left[ g(\theta_n, X_n(\theta_n)) - w(t, x) \right] \mathbf{1}_{\theta_n < t_n} - h_n. \]  

(3.23)

The probabilistic estimate (3.21) with \( m = 2 \) implies that the limit of the third term is zero
and

\[
\lim_{n \to \infty} \frac{1}{h_n} E \left( \int_t^{t_n} L(t, x, \alpha_n(s)) \, ds - \int_t^{\theta_n} L(s, X_n(s), \alpha_n(s)) \, ds \right) \leq \lim_{n \to \infty} \frac{1}{h_n} \left[ \|L\|_{\infty} E(t_n - \theta_n) \right. \\
\left. + E \int_t^{t_n} \left| L(t, x, \alpha_n(s)) - L(s, X_n(s), \alpha_n(s)) \right| \, ds \right].
\] (3.24)

Since $Q \times A$ is compact, $L$ is uniformly continuous. Also (3.20) implies that, for every $\delta > 0$,

\[
\lim_{n \to \infty} P \left( \sup_{t \leq \rho \leq t + h_n} |X_n(\rho) - x| \geq \delta \right) = 0.
\]

Therefore the uniform continuity of $L$ and (3.20) imply that the limits in (3.24) are zero. We now use (3.21) and Itô formula to obtain

\[
\lim_{n \to \infty} \frac{1}{h_n} E \left( \int_t^{t_n} w(t, X_n(t)) - w(t, x) \chi_{\theta_n = t_n} \right. \\
\left. - w(t, x) + \left( \mathcal{L}_{\alpha_n(s)}^\theta w(t, \cdot) \right)(x) \right) \, ds \leq \lim_{n \to \infty} \frac{1}{h_n} \sup_{\alpha} \left| w(t, x) + \left( \mathcal{L}_{\alpha_n(s)}^\theta w(t, \cdot) \right)(x) \right| E(t_n - \theta_n) \\
+ \lim_{n \to \infty} \frac{1}{h_n} E \int_t^{\theta_n} \left| \left( \mathcal{L}_{\alpha_n(s)}^\theta w(s, \cdot) \right)(X_n(s)) - \left( \mathcal{L}_{\alpha_n(s)}^\theta w(s, \cdot) \right)(x) \right| \, ds \\
+ \lim_{n \to \infty} \frac{1}{h_n} E \int_t^{\theta_n} \left| w(t, X_n(s)) - w(t, x) \right| \, ds.
\]

Since $w \in C^{1,2}(Q), \mathcal{L}_{\alpha_n}^\theta w(s, y)$ is a uniformly continuous function of $Q$. As in (3.24), the dominated convergence theorem and (3.20) imply that the above limit is zero. Combine this with (3.23) and (3.24) to obtain

\[
I(h_n) \geq L^n + G^n - e(n),
\]

where

\[
(L^n, G^n) := \frac{1}{h_n} \left( E \int_t^{t+h_n} L(t, x, \alpha_n(s)) \, ds, \\
E \int_t^{t+h_n} w(t, x) + \left( \mathcal{L}_{\alpha_n(s)}^\theta w(t, \cdot) \right)(x) \, ds \right).
\]
and the error term $e(n)$ converges to zero as $n \to \infty$. Define a set

$$\hat{A} = \{ (L, G) \in \mathbb{R}^2 : L = L(t, x, \alpha),$$

$$G = w(t, x) + (L_\alpha^\alpha w(t, \cdot))(x) \text{ for some } \alpha \in A \}. $$

Then $(L^n, G^n) \in \overline{co}(\hat{A})$, where $\overline{co}$ denotes the convex, closed hull of $\hat{A}$. Also,

$$L^n + G^n \geq \inf \{ L + G : (L, G) \in \overline{co}(\hat{U}) \}$$

$$= \inf \{ L + G : (L, G) \in \hat{U} \}$$

$$= w(t, x) - (G_t w(t, \cdot))(x).$$

As in Section 3.1, let $v$ be the value function. Then, in view of Theorem 3.3, we have the following representation result for (3.1) with nonlinearity $H$ given in (3.3) with $L^\alpha$ given in (3.10) and $r \equiv 0$. However, this restriction that $r \equiv 0$ can easily be removed by the techniques developed in Section 2. For that reason we state the result including the linear term $r(t, z, \alpha)v(t, x)$. For uniqueness we need the boundary conditions. It is clear that $v$ satisfies the terminal condition (2.20). Also, under some conditions, the value function satisfies the Dirichlet boundary condition (2.21) (see Chapter 5 in [36]). Also, in degenerate cases, (2.21) may hold only in the viscosity sense. We refer the interested reader to the book of Barles [4], or Section 7.6 in [36].

**Corollary 3.5 (Control representation for (3.1)).** Suppose that $v \in C(\overline{Q})$. Then $v$, a viscosity solution of the dynamic programming equation (3.1), with nonlinearity $H$ given in (3.3) with the infinitesimal generator as in (3.10), i.e.,

$$-v_t(t, x) + H(t, x, v(t, x), D(t, x), D^2 v(t, x)) = 0 \quad \text{on } (-\infty, T) \times O,$$

(3.25)

where

$$H(t, x, v, p, B)$$

$$= \sup_{\alpha \in A} \left\{ -r(t, x, \alpha)v - \mu(t, x, \alpha) \cdot p - \frac{1}{2} a(t, x, \alpha) : B - L(t, x, \alpha) \right\}. $$

In particular, if $v$ satisfies (2.21) and if there is only one continuous viscosity solution of (3.25) together with (2.21), (2.20), then this solution is given as the value function of the stochastic optimal control problem.

**3.3. Smooth value function and verification**

In this subsection, we assume that there exits a $u \in C^{1,2}((-\infty, T] \times \overline{O})$ that solves the dynamic programming equation (3.25) together with boundary conditions (2.21), (2.20).
Then, we will show by a direct application of Itô calculus that this solution must be equal to the value of the stochastic semigroup defined in Example 3.1. This, in particular, proves uniqueness of $C^{1,2}$ solutions to equations (3.25), (2.20), (2.21).

**Theorem 3.6 (Verification).** Let $u \in C^{1,2}((-\infty,T] \times \mathcal{O})$ be a solution of (3.25), (2.20), (2.21). Then, for every $\alpha(\cdot) \in \mathcal{A}$,

$$u(t,x) \leq (\mathcal{J}^{\alpha(\cdot)}_{t,T} \varphi)(x).$$

In addition, suppose that there exists an optimal control $\alpha^*(\cdot) \in \mathcal{A}$ so that, for Lebesque almost all $s \in [t,T]$,

$$\alpha^*(s) \in \arg \min_{\alpha \in \mathcal{A}} \left\{ -r(s,X^*(s),\alpha^*(s))v(s,X^*(s)) \\
- \mu(s,X^*(s),\alpha^*(s)) \cdot Du(s,X^*(s)) \\
- \frac{1}{2} a(s,X^*(s),\alpha^*(s)) : D^2 u(s,X^*(s)) \\
- L(s,X^*(s),\alpha^*(s)) \right\},$$

where $X^*(s)$ is the solution of (3.8), with initial data (2.13) and control $\alpha^*(\cdot)$. Then

$$u(t,x) = (\mathcal{J}^{\alpha^*(\cdot)}_{t,T} \varphi)(x).$$

**Proof.** Fix $(t,x)$ and $\alpha(\cdot)$. Let $X(\cdot)$ be the corresponding state process. Apply the Itô rule to $Y(\cdot) := B(\cdot)u(s,X(s))$. The result is

$$dY(s) = B(s) \left[ r(s,X(s),\alpha(s))u(s,X(s)) + \mu(s,X(s),\alpha(s)) \cdot Du(s,X(s)) \\
+ \frac{1}{2} a(s,X(s),\alpha(s)) : D^2 u(s,X(s)) + u_t(s,X(s)) \right] ds \\
+ B(s)Du(s,X(s)) dW(s).$$

We integrate the above on $[t,\theta]$, take the expected value and then use the equations (3.25), (2.21), (2.20). The result is

$$Y(t) = u(t,x) = E \left[ B(\theta)u(\theta,X(\theta)) \right]$$

$$- E \left( \int_t^\theta B(s) \left[ u_t(s,X(s)) + r(s,X(s),\alpha(s))u(s,X(s)) \\
+ \mu(s,X(s),\alpha(s)) \cdot Du(s,X(s)) \right] ds \right).$$
Stochastic representations for nonlinear parabolic PDEs

\[ + \frac{1}{2} a(s, X(s), \alpha(s)) : D^2 u(s, X(s)) \right) ds \]
\[ \geq E \left[ B(\theta) u(\theta, X(\theta)) \right] + E \left( \int_{t}^{\theta} B(s) L(s, X(s), \alpha(s)) ds \right) \]
\[ = \left( \mathcal{J}_{t, T} \varphi \right)(x). \]

This proves the first part of the statement. To prove the second part, we repeat the above calculations with the control \( \alpha^*(\cdot) \) and \( X^*(\cdot) \). The inequality in the above calculation is now an equality and the optimality of \( \alpha^*(\cdot) \) follows. \( \square \)

3.4. Optimal control of jump Markov processes

In this subsection, we briefly discuss the extension of the representation results for the integro-differential equations discussed in Section 2.6. We refer to the books [36,55] and the paper [2] for more information.

As in the diffusion case, we will introduce controlled diffusion equations driven by a Markov process and a random measure. Indeed, let \( A \) be a control set, and \( \mu, \sigma, L \) be as in Section 3.2. Further, let \( \pi \) be a compensator measure on \( \mathbb{R}^d \) and \( f \) be function satisfying

\[ \sup_{\alpha \in A} \int_{\mathbb{R}^d} \left[ 1 \wedge \left| f(t, x, \xi, \alpha) \right|^2 \right] \pi(d\xi) < \infty. \]

We also assume the standard Lipschitz condition

\[ \int_{\mathbb{R}^d} \left| f(t, x, \xi, \alpha) - f(t, y, \xi, \alpha) \right|^2 \pi(d\xi) \leq C |x - y|^2 \]

for all \( t \in \mathbb{R}^1, x, y \in \mathbb{R}^d \) for some constant \( C \). As before, let \( A \) be all bounded, progressively measurable, \( A \)-valued random processes. Then, given a control process \( \alpha(\cdot) \in A \) and initial point \( (t, x) \), there exists a unique solution to

\[ X(T) = x + \int_{t}^{T} \mu(s, X(s), \alpha(s)) ds + \int_{t}^{T} \sigma(s, X(s), \alpha(s)) dW(s) \]
\[ + \int_{t}^{T} \int_{\mathbb{R}^d} f(s, X(s), \xi, \alpha(s)) \tilde{p}(ds \times d\xi), \] (3.26)

where \( \tilde{p} \) is a martingale random measure with compensator \( \pi \) and \( W(\cdot) \) is an independent standard Brownian motion.

For a constant control \( \alpha(\cdot) \equiv \alpha \), the infinitesimal generator is given by

\[ \mathcal{L}_t^\alpha \varphi(x) = \mu(t, x, \alpha) \cdot \nabla \varphi(x) + \frac{1}{2} a(t, x, \alpha) : D^2 \varphi(x) + \tilde{L}_t^\alpha \varphi(x), \]
where the part corresponding to the random measure is as in (2.19),
\[
\widetilde{L}_t \alpha \varphi(x) = \int_{\mathbb{R}^d} \left[ \varphi(x + f(t, x, \xi, \alpha)) - \varphi(x) - f(t, x, \xi, \alpha) \cdot \nabla \varphi(x) \right] \pi(d\xi).
\]

Now we define the pay-off functional and the value function as in the diffusion case (see (3.9)),
\[
\left( J_{t,T}^{\alpha(\cdot)} \psi \right)(x) = E \left[ \int_t^T B(s) L(s, X(s), \alpha(s)) \, ds + B(\theta) g(\theta, X(\theta)) \chi_{\theta < T} + B(T) \psi(X(\theta)) \chi_{\theta \geq T} \right],
\]
where
\[
B(T) = B(t, T; z, \alpha(z)) = \exp \left( - \int_t^T r(s, Z(s), \alpha(s)) \, ds \right).
\]
and \(\theta\) is the exit time of \((s, X(s))\) from \(\overline{Q} = [t, T] \times \overline{O}\). The nonlinear semigroup is given by
\[
(T_{t,r}^{\alpha(\cdot)} \psi)(x) := \inf_{\alpha(\cdot) \in A} J_{t,r}^{\alpha(\cdot)} \psi(x).
\]
Again it follows that \(T_{t,r}\) is monotone, semigroup by dynamic programming for optimal control.

We can then show that the value function is a viscosity solution of the corresponding dynamic programming equation. Such a result for jump Markov processes was first proved in \([63]\), and then by Sayah \([61]\). We refer to the recent book of Oksendall and Sulem \([55]\) for more information.

Given the form of the generator \(L_{t}^{\alpha}\) and the formal discussion given in the beginning of this section, the related dynamic programming equation is (3.1), with \(H\) in (3.3) and \(L_{t}^{\alpha}\) as above.

The boundary conditions are (2.20) and (2.21). But the important point to emphasize is that (2.21) holds for all \(x\) not only on \(\partial O\) but in all of \(\mathbb{R}^d \setminus O\) as \(X(\theta) \in \mathbb{R}^d \setminus O\).

### 3.5. Other type of control problems

Several other types of control problems have been studied in the literature. These problems are related to so-called quasivariational inequalities. Indeed, stopping time problems are related to obstacle problems \([9]\). Impulse or switching controls yield quasivariational inequalities \([10]\). Singular control problems allow the state processes to be discontinuities \([36]\). Dynamic programming equations for singular control problem are again quasivariational inequalities but with constraints on the first derivative of the solution. Equations with constraints on the second derivatives are much rare and obtained only in \([20]\).
3.6. Stochastic target and geometric problems

In this section, we consider a special class of nonlinear parabolic equations. These equations are related to geometric flows of manifolds embedded in \( \mathbb{R}^d \). The nonlinearities \( H \) that appear in these equations are, in addition to being parabolic (2.11), also geometric, i.e.,

\[
\mathcal{H}(t, x, \lambda p, B + \mu p \otimes p) = \lambda \mathcal{H}(t, x, p, B) \quad \forall \lambda \geq 0, \mu \in \mathbb{R}^1.
\] (3.27)

It was shown in [67,68] that a large subclass of these nonlinearities have a stochastic representation similar to that discussed in Section 3.2. In this representation, however, a new class of control problems called stochastic target problems are used [65,66]. A stochastic target problem is a nonclassical control problem in which the controller tries to steer a controlled stochastic process into a given target set \( G \) by judicial choices of controls. The chief object of study is the set of all initial positions from which the controlled process can be steered into \( G \) with probability one in an allowed time interval. Clearly these reachability sets depend on the allowed time. Thus, they can be characterized by an evolution equation which is the analogue of the dynamic programming equation of stochastic optimal control.

Geometric equations express the velocity of the boundary as a possibly nonlinear function of the normal and the curvature vectors. In [66,67] it was shown that smooth solutions of these geometric equations, when exist, are equal to the reachability sets. However, as a Cauchy problem, these equations in general do not admit classical smooth solutions and a weak formulation is needed. Several such formulations were given starting with the pioneering work of Brakke [15]. Here we consider the viscosity formulation given independently by Chen, Giga and Goto [19] and by Evans and Spruck [34]. The main idea of this approach is to characterize the geometric solution as the zero level set of a continuous function. Then, this function solves a partial differential equation (3.25) with a geometric \( H \) satisfying (3.27).

The chief goal of this subsection is to give a stochastic characterization of the unique level set solutions of [19,34] in terms of the target problem. The stochastic semigroup is given by

\[
v(t, x) := \inf_{\alpha(\cdot) \in \mathcal{A}} \text{ess sup}_{\omega \in \Omega} \varphi(X_{t,x}^{\alpha(\cdot)}(T, \omega)),
\] (3.28)

where for initial data \((t, x) \in (-\infty, T) \times \mathbb{R}^d\), control process \(\alpha(\cdot) \in \mathcal{A}\), the controlled process \(\{X(s) := X_{t,x}^{\alpha(\cdot)}(s)\}_{s \geq t}\) is the solution of (3.8) and (2.13).

The following representation result is proved in [66].

**Theorem 3.7.** Suppose that the standard Lipschitz assumption (3.7) holds and that \( H \) is locally Lipschitz on \( \{p \neq 0\} \). Then, \( v \) defined in (3.28) satisfies (2.20) pointwise and it is a discontinuous viscosity solution of (3.25) with

\[
H(t, x, p, B) := \sup_{\nu \in \mathcal{N}(t, x, p)} \left\{ -\mu(t, x, \alpha) \cdot p - \frac{1}{2} a(t, x, \alpha) : B \right\},
\] (3.29)
where
\[
\mathcal{N}(t, x, p) := \{ \alpha \in A : \sigma(t, x, \alpha)p = 0 \} \quad \text{for } p \neq 0 \quad \text{and} \quad \mathcal{N}(t, x, 0) := A.
\]

(3.30)

Observe that \( H(t, x, p, B) \) defined above is geometric and also it is singular at \( p = 0 \) because \( \mathcal{N}(t, x, 0) = A \).

The above theorem, in fact, follows from a more geometric result that connects the evolution equations more manifolds and stochastic target problems. In this context, the semigroup \( \mathcal{T}_{t,T} \) acts on subsets of \( \mathbb{R}^d \). Indeed, for a given Borel subset \( G \) of \( \mathbb{R}^d \), the target reachability set is given by
\[
\mathcal{T}_{t,T} G := v^G(t) := \{ x \in \mathbb{R}^d : X^\alpha(t, x)(T) \in G \text{ a.s. for some } \alpha(\cdot) \in A \}.
\]

Dynamic programming principle for these problems is proved in [66]: for all \( t \leq r \leq T \),
\[
\mathcal{T}_{t,T} G = \{ x \in \mathbb{R}^d : X^\alpha(r)(r) \in \mathcal{T}_{r,T} G \text{ a.s. for some } \alpha(\cdot) \in A \}.
\]

This is exactly the semigroup property
\[
\mathcal{T}_{t,T} G = \mathcal{T}_{t,r}(\mathcal{T}_{r,T} G).
\]

The infinitesimal generator of this semigroup can be stated purely in terms of geometric quantities such as the normal vector and second quadratic form of the set. Indeed, in [66] the characteristic functions of the reachability sets are shown to be viscosity solutions of the geometric dynamic programming equations in the sense defined in [64]. In particular, this result implies that the reachability set is included in the zero sublevel set of the solutions constructed in [19,34]. In view of the techniques developed by Barles, Soner and Souganidis [6], and [64], these purely geometric results are equivalent to Theorem 3.7. To state the main result in this direction we need the following definition:
\[
K(t, z) := \{ (\mu(t, x, \alpha), \sigma(t, x, \alpha)) : \alpha \in A \}.
\]

THEOREM 3.8. Let the conditions of Theorem 3.7 hold. Suppose that \( \varphi \) is bounded and uniformly continuous, and (3.25) with \( \mathcal{H} \) as in (3.29) has comparison. Let \( v \) is the unique bounded continuous viscosity solution of (3.25), (2.20). Assume further that the set \( K(t, x) \) is closed and convex for all \( (t, x) \in (-\infty, T) \times \mathbb{R}^d \). Then
\[
v^G(t) = \{ x \in \mathbb{R}^d : v(t, x) \leq 0 \}
\]
with the target set
\[
G := \{ x \in \mathbb{R}^d : \varphi(x) \leq 0 \}.
\]
The proof of this theorem is a straightforward application of Theorem 3.7 and the results of [6]. Observe that the boundedness of \( \varphi \) is not a restriction, as one can replace \( \varphi \) by \( \varphi(1 + |\varphi|)^{-1} \).

The stochastic target problems with jump-diffusion processes are discussed by Bouchard [13]. Also, target problems are related to forward–backward stochastic differential equations (FBSDEs) discussed in Section 4. A similar representation theorem for the special case of the codimension-one mean curvature flow was also obtained by Buckdahn, Cardaliaguet and Quincampoix [16].

We close this subsection by the important example of mean curvature flow.

**EXAMPLE 3.9.** Consider the example with \( A = \mathcal{P}^{k,d} \) be the set of all projection matrices on \( \mathbb{R}^d \) onto a hyperplane of dimension \( k \), \( \mu \equiv 0 \) and \( \sigma(t,x,\alpha) = \sqrt{2}\alpha \). Then, the state equation (3.8) reduces to

\[
\frac{dX(s)}{ds} = \sqrt{2}\alpha(s) \, dW(s),
\]

where \( W(\cdot) \) is the standard \( d \)-dimensional Brownian motion. Hence, at each time \( s \), the controller decides on which \( k \)-dimensional space \( X(\cdot) \) should diffuse. Then the related PDE has the form

\[
\mathcal{H}^k(p,B) = \sup\{\alpha : B | \alpha \in \mathcal{P}^{k,d} \text{ and } \alpha p = 0\}.
\]

This is exactly the same nonlinear function used by Ambrosio and Soner [1] to describe the weak flow of codimension \( d - k \) mean curvature flow. In the special case of \( k = d - 1 \), any \( \alpha \in \mathcal{P}^{d-1,d} \) is given by \( \alpha = I - v \otimes v \) for some unit vector \( v \in \mathbb{R}^d \). Also for such a matrix \( \alpha \) and \( p \neq 0 \),

\[
\alpha p = 0 \implies [I - v \otimes v]p = 0 \implies v = \pm \frac{p}{|p|} := \bar{p}.
\]

Hence,

\[
\mathcal{H}^{d-1}(p,B) = \sup\{[I - v \otimes v] : B | v = \pm \bar{p}\} = \text{trace}(B) - B \bar{p} \cdot \bar{p},
\]

and equation (3.25) has the form

\[
-v(t,x) - \Delta v(t,x) + \frac{D^2v(t,x)\nabla v(t,x) \cdot \nabla v(t,x)}{|
abla v(t,x)|^2} = 0.
\]

This is the level set equation (in reversed time) for the mean curvature flow [19,34]. Then, in the special case of this example, results of this subsection can be stated as follows. The unique viscosity solution of the above level set equation of the mean curvature flow has the stochastic representation

\[
v(t,x) := \inf_{\alpha(t) \in \mathcal{P}^{k,d}} \sup_{\omega \in \Omega} \varphi \left( x + \int_t^T \alpha(s) \, dW(s) \right).
\]
4. Backward representations

In this section we outline a different connection between PDEs and stochastic processes. Vaguely, this connection is analogous to the connection between ordinary differential equations and first-order PDEs through the method of characteristics. Indeed, it is first observed by Bismut [11] in his seminal work on the extension of Pontryagin maximum principle to stochastic optimal control. Pontryagin’s maximum principle itself is the extension of the Hamilton–Jacobi theory of classical mechanics to deterministic optimal control and provides conditions for maximality through a set of ordinary differential equations. For stochastic optimal control, Bismut achieved this using stochastic processes. As well known, the method of characteristics and its mentioned generalizations have both initial and terminal boundary data to be satisfied. In the stochastic context, due to the adaptability conditions, this makes the problem harder. However, a deep theory is now developed through the recent works of Peng, Pardoux and others [32,54,56–58,60]. This theory known as Backward stochastic differential equations (BSDEs) will be outlined in the next subsection. BSDEs have a natural connection with PDEs and several numerical methods have been developed [27,53,69]. However, the PDEs connected to BSDEs are always quasilinear. Recently, Cheredito, Soner, Touzi and Victoir [21] extended this theory to cover all fully nonlinear, parabolic, second-order PDEs. This extension and the possible numerical implications are outlined below. For a more complete introduction to BSDEs we refer to the survey paper of El-Karoui, Peng and Quenez [32].

4.1. Backward stochastic differential equations

Let $X(\cdot) := X^{t,x}(\cdot)$ be the solution of (2.12), (2.13). Given real-valued, nonlinear function $f$ and terminal data $\varphi$ consider the equation

$$dY(s) = f(s, X(s), Y(s))\, ds + Z(s) \cdot \sigma(s, X(s))\, dW,$$

(4.1)

with terminal data

$$Y(T) = \varphi(X(T)).$$

(4.2)

The problem is to find processes $Y(\cdot)$ and $Z(\cdot)$ that are integrable and adapted to the filtration $\mathcal{F}_t$. Adaptedness condition is a serious technical condition as the given data for $Y(\cdot)$ is specified at the terminal time $T$. In the probabilistic literature the BSDE is defined more generally. Random $f$ and more general $X(\cdot)$ process with $Y(\cdot)$ dependence are also considered. Here we restrict ourselves to the above framework to simplify the presentation.

Let us assume that the solution $Y(s)$ is given as a deterministic function of time and $X(s)$, i.e., assume that there is a deterministic function $v$ so that

$$Y(s) = v(s, X(s)) \quad \forall s \in [t, T].$$
If we also assume that $v$ is smooth, then by the Itô formula, we have

$$d[v(s, X(s))] = [v_t(s, X(s)) + (\mathcal{L}_t v)(s, \cdot)(X(s))] \, ds$$

$$+ \nabla v(s, X(s)) \cdot \sigma(s, X(s)) \, dW(s),$$

where $\mathcal{L}_t$ is as in (2.14). If $\sigma$ has full rank, then equating the above equation to (4.1) yields

$$Z(s) = \nabla v(s, X(s)),$$

and $v$ must solve

$$-v_t - \mathcal{L}_t v + f(t, x, v, \nabla v) = 0 \text{ on } (-\infty, T] \times \mathbb{R}^d.$$ 

Hence, smooth solutions of the above semilinear PDE has the representation in terms of the BSDE (4.1). Numerical implication of this connection is discussed in Section 5.2. Also the rigorous connection between the PDE and the BSDE is given in the references cited before.

### 4.2. Second-order backward stochastic equations

In the BSDE literature it has not been possible to consider PDEs with a nonlinear second-order term. Only quasilinear PDEs were shown to have connection with the BSDEs. This is achieved by introducing a $Y, Z$ dependence in the dynamics of $X$.

In recent work [21], BSDEs were generalized by restricting the $Z$ process to be a semi-martingale. Precisely a second-order backward stochastic differential equation (2BSDE in short) has the $X$ and $Y$ equations, (2.12), (2.13) and (4.1), (4.2), and an additional equation

$$dZ(s) = a(s) \, ds + \Gamma(s) \sigma(s, X(s)) \, dW(s)$$

for some processes $a(\cdot)$ and $\Gamma(\cdot)$.

For simplicity, let us assume that $\sigma \equiv I_d$. Then, we rewrite the 2BSDE as

$$dX(s) = \mu(s, X(s)) \, ds + dW(s),$$

$$dY(s) = H(s, X(s), Y(s), Z(s), \Gamma(s)) \, ds + Z(s) \circ dX(s),$$

$$dZ(s) = A(s) \, ds + \Gamma(s) \circ dX(s),$$

$$Y(T) = \varphi(X(T)), \quad X(t) = x,$$

where $H$ is a given function and the Fisk–Stratonovich integral $\circ$ is given by

$$Z(s) \circ dW(s) = Z(s) \, dW(s) + \frac{1}{2} \text{trace} [\Gamma(s)] \, ds.$$
Below, we will give the precise function spaces in which we look for the solutions. However, to establish the connection between the PDEs let us formally assume that there is a solution and is given by $Y(s) = v(s, X(s))$. Then, by the Itô formula (using the definition of the Fisk–Stratonovich integral),

$$d[v(s, X(s))] = v_t(s, X(s)) ds + \nabla v(s, X(s)) \circ dX(s).$$

Comparing this to the $dY$ equation in (4.4), again we conclude that $Z(s) = \nabla v(s, X(s))$ and

$$-v_t(s, X(s)) + H(s, X(s), Y(s), Z(s), \Gamma(s)) = 0 \quad \forall s \in [t, T]. \quad (4.5)$$

Now we apply the Itô rule to $\nabla v(s, X(s))$. The result is

$$d[\nabla v(s, X(s))] = [\nabla v_t(s, X(s)) + \mathcal{L} \nabla v(s, X(s))] ds + D^2 v(s, X(s)) dW(s).$$

We compare this to the $dZ$ equation in (4.4) to conclude that $\Gamma(s) = D^2 v(s, X(s))$. We substitute these into (4.5) to obtain

$$-v_t(s, X(s)) + H(s, X(s), v(s, X(s)), \nabla v(s, X(s)), D^2 v(s, X(s))) = 0 \quad \forall s \in [t, T].$$

If the $X$ process has full support, then we conclude that

$$-v_t(t, x) + H(t, x, \nabla v(t, x), D^2 v(t, x)) = 0 \quad \forall (t, x) \in (-\infty, T] \times \mathbb{R}^d. \quad (4.6)$$

Notice that we did not make any assumptions on $H$. In particular, no parabolicity is assumed. Although, the above formal calculations do not require parabolicity, the existence of a solution to 2BSDE fails without parabolicity. In the remainder of this subsection, we will give the precise definitions of the functions spaces and the assumptions needed. Then, we will state the main representation result of [21] without proof.

In addition to usual local Lipschitz conditions, we assume there are constants $C \geq 0$ and $p_1 \in [0, 1]$ such that

$$|\mu(t, x)| \leq C(1 + |x|^{p_1}), \quad (t, x) \in (-\infty, T] \times \mathbb{R}^d.$$

**Definition 4.1.** Let $(t, x) \in (-\infty, T) \times \mathbb{R}^d$ and $(Y(\cdot), Z(\cdot), \Gamma(\cdot), A(\cdot))$ be a quadruple of $\mathcal{F}$-progressively measurable processes on $[t, T]$ with values in $\mathbb{R}$, $\mathbb{R}^d$, $\mathcal{S}^d$ and $\mathbb{R}^d$, respectively. Then we call $(Y, Z, \Gamma, A)$ a solution to the second-order backward stochastic differential equation (2BSDE) corresponding to $(X^{t, x}, H, \varphi)$ if they solve (4.4).
Equations (4.4) can be viewed as a whole family of 2BSDEs indexed by \((t,x) \in [0,T) \times \mathbb{R}^d\). We have formally argued that the solution of these equations are related to the fully nonlinear partial differential equation (4.6).

Since \(Z\) is a semimartingale, the use of the Fisk–Stratonovich integral in (4.4) means no loss of generality, but it simplifies the notation in the PDE (4.6). Alternatively, (4.4) could be written in terms of the Itô integral as

\[
dY(s) = \tilde{H}(s, X^{t,x}(s), Y(s), Z(s), \Gamma(s)) \, ds + Z(s) \cdot dX^{t,x}(s),
\]

where (recalling that \(\sigma\) in the \(X\) equation is taken to be the identity)

\[
\tilde{H}(t, x, y, z, \gamma) = H(t, x, y, z, \gamma) + \frac{1}{2} \text{trace}[\gamma].
\]

In terms of \(\tilde{H}\), the PDE (4.6) can be rewritten as

\[
-v_t(t, x) + \tilde{H}(t, x, v(t, x), Dv(t, x), D^2v(t, x)) - \frac{1}{2} \Delta v(t, x) = 0.
\]

Finally, notice that the form of the PDE (4.6) does not depend on the functions the dynamics of the \(X\) process. So, we could restrict our attention to the case where \(\mu \equiv 0\) and \(\sigma \equiv I_d\), the \(d \times d\) identity matrix. But the freedom to choose the dynamics of \(X\) from a more general class of diffusions provides additional flexibility in the design of the Monte Carlo schemes discussed in Section 5.

**From a solution of the PDE to a solution of the 2BSDE.** Assume \(v : [0,T] \times \mathbb{R}^d \to \mathbb{R}\) is a continuous function such that

\[v_t, Dv, D^2v, \mathcal{L}Dv\] exist and are continuous on \([0,T) \times \mathbb{R}^d\),

and \(v\) solves the PDE (4.6) with terminal condition (2.20). Then it follows directly from Itô’s formula that for each pair \((t,x) \in (-\infty,T) \times \mathbb{R}^d\), the processes

\[
Y(s) = v(s, X^{t,x}(s)), \quad s \in [t,T],
\]

\[
Z(s) = Dv(s, X^{t,x}(s)), \quad s \in [t,T],
\]

\[
\Gamma(s) = D^2v(s, X^{t,x}(s)), \quad s \in [t,T],
\]

\[
A(s) = \mathcal{L}Dv(s, X^{t,x}(s)), \quad s \in [t,T],
\]

solve the 2BSDE corresponding to \((X^{t,x}, H, \phi)\).

**From a solution of the 2BSDE to a solution of the PDE.** In all of this subsection, we assume that

\[H : (-\infty,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{S}^d \to \mathbb{R}\] and \(\phi : \mathbb{R}^d \to \mathbb{R}\).
are continuous functions that satisfy the following Lipschitz and growth assumptions:

(A1) For every $N \geq 1$ there exists a constant $F_N$ such that

$$|H(t, x, y, z, \gamma) - H(t, x, \tilde{y}, z, \gamma)| \leq F_N |y - \tilde{y}|$$

for all $t \in (-\infty, T]$, $x, z \in \mathbb{R}^d$, $y, \tilde{y} \in \mathbb{R}^1$, $\gamma \in S^d$ with $\max\{|x|, |y|, |\tilde{y}|, |z|, |\gamma|\} \leq N$.

(A2) There exist constants $F$ and $p_2 \geq 0$ such that

$$|H(t, x, y, z, \gamma)| \leq F(1 + |x|^{p_2} + |y| + |z|^{p_2} + |\gamma|^{p_2})$$

for all $(t, x, y, z, \gamma) \in (-\infty, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times S^d$.

(A3) There exist constants $G$ and $p_3 \geq 0$ such that

$$|\varphi(x)| \leq G(1 + |x|^{p_3})$$

for all $x \in \mathbb{R}^d$.

**Admissible strategies.** We fix constants $p_4, p_5 \geq 0$ and denote for all $(t, x) \in (-\infty, T] \times \mathbb{R}^d$ and $m \geq 0$ by $\mathcal{A}^{t,x}_m$ the class of all processes of the form

$$Z(s) = z + \int_t^s A(r) \, dr + \int_t^s \Gamma( r) \, dX_t,x (r), \quad s \in [t, T],$$

where $z \in \mathbb{R}^d$, $(A(\cdot), \Gamma(\cdot)) \in \mathbb{R}^d \times S^d$ progressively measurable processes satisfying

$$\max\{|Z(s)|, |A(s)|, |\Gamma(s)|\} \leq m(1 + |X_t,x (s)|^{p_4}) \quad \forall s \in [t, T], \quad (4.7)$$

and

$$|\Gamma(r) - \Gamma(s)| \leq m(1 + |X_t,x (r)|^{p_5} + |X_t,x (s)|^{p_5}) \times (|r - s| + |X_t,x (r) - X_t,x (s)|) \quad \forall r, s \in [t, T]. \quad (4.8)$$

Set $\mathcal{A}^{t,x} := \bigcup_{m \geq 0} \mathcal{A}^{t,x}_m$. It follows from the assumptions (A1) and (A2) on $H$ and the condition (4.7) on $Z$ that for all $y \in \mathbb{R}$ and $Z \in \mathcal{A}^{t,x}$, the forward SDE

$$dY(s) = f(s, X^{t,x}(s), Y(s), Z(s), \Gamma(s)) \, ds + Z(s) \circ dX^{t,x}(s), \quad s \in [t, T],$$

with $Y(t) = y$, has a unique strong solution $Y^{t,x,y,Z}(\cdot)$ (this can, for instance, be shown with the arguments in the proofs of Theorems 2.3, 2.4 and 3.1 in Chapter IV of Ikeda and Watanabe [41]).

Notice that $Z \in \mathcal{A}^{t,x}$ is a solution of the 2BSDE if $Y^{t,x,y,Z}(T) = \varphi(X^{t,x}(T))$.

We will show that solutions to 2BSDE in the class $\mathcal{A}^{t,x}$ has at most one solution.
Stochastic representations for nonlinear parabolic PDEs

Uniqueness of 2BSDE in $A^{t,x}$. For our last assumption and the statement of Theorem 4.3, we need the following definition.

**Definition 4.2.** Let $q \geq 0$.

1. We call a function $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ a viscosity solution with growth $q$ of the PDE (4.6) with terminal condition (2.20) if $v$ is a viscosity solution of (4.6) on $(-\infty, T) \times \mathbb{R}^d$ such that $|v(t, x)| \leq C(1 + |x|^q)$ for all $(t, x) \in (-\infty, T] \times \mathbb{R}^d$.

2. We say that the PDE (4.6) with terminal condition (2.20) has comparison with growth $q$ if the following holds:

   If $w : (-\infty, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is lower semicontinuous and a viscosity supersolution of (4.6) on $(-\infty, T) \times \mathbb{R}^d$ and $u : (-\infty, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ upper semicontinuous and a viscosity subsolution of (4.6) on $(-\infty, T) \times \mathbb{R}^d$ such that $w(T, x) \geq g(x) \geq u(T, x)$ for all $x \in \mathbb{R}^d$ and there exists a constant $C \geq 0$ with $w(t, x) \geq -C(1 + |x|^p)$ and $u(t, x) \leq C(1 + |x|^p)$ for all $(t, x) \in (-\infty, T) \times \mathbb{R}^d$.

   then $w \geq u$ on $(-\infty, T] \times \mathbb{R}^d$.

With this definition our last assumption on $H$ and $\phi$ is

(A4) The PDE (4.6) with terminal condition (2.20) has comparison with growth $p = \max\{p_2, p_3, p_2 p_4, p_4 + 2 p_1\}$.

The following result is proved in [20].

**Theorem 4.3 (Uniqueness of 2BSDE).** Assume (A1)–(A4) and that $H$ is parabolic (2.11). For $x_0 \in \mathbb{R}^d$ suppose that the 2BSDE corresponding to $(X^{0,x_0}, H, \phi)$ has a solution with $Z^{0,x_0} \in A^{0,x_0}$. Then

(i) The associated PDE (4.6) with terminal condition (2.20) has a unique viscosity solution $v$ with growth $p = \max\{p_2, p_3, p_2 p_4, p_4 + 2 p_1\}$, and $v$ is continuous on $[0, T] \times \mathbb{R}^d$.

(ii) For all $(t, x) \in [0, T] \times \mathbb{R}^d$, there exists exactly one solution $(Y^{t,x}, Z^{t,x}, \Gamma^{t,x}, A^{t,x})$ to the 2BSDE corresponding to $(X^{t,x}, H, \phi)$ such that $Z^{t,x} \in A^{t,x}$ and

$$Y^{t,x}(s) = v(s, X^{t,x}(s)), \quad s \in [t, T],$$

where $v$ is the unique continuous viscosity solution with growth $p$ of (4.6) and (2.20).

**Remark 4.4.** 1. Under the hypothesis of the above theorem, the solution of the 2BSDE satisfies $Y^{t,x}(t) = v(t, x)$. Hence, $v(t, x)$ can be approximated by backward simulation.
of the process \((Y^{t,x}(s))_{s\in[0,T]}\). If \(v\) is \(C^2\), it follows from Itô’s lemma that 
\[ Dv(s, X^{t,x}(s)), s \in [t,T]. \]
Then \(Dv(t,x)\) can also be approximated by backward simulation. Moreover, for \(v\) is \(C^3\), 
\(\Gamma^{t,x}(s) = D^2v(s, X^{t,x}(s))\) can be simulated in this way.

A formal discussion of a potential numerical scheme for the backward simulation of the 
processes \((Y^{t,x}, Z^{t,x})\) and \((\Gamma^{t,x})\) is provided in Section 5.3.

2. We have already shown that a classical solution \(v\) of (4.6) and (2.20) and its deriv-
atives provide a solution of the 2BSDE.

3. The parabolicity assumption (2.11) is natural from the PDE viewpoint. If \(H\) is 
uniformly elliptic: there exists a constant \(C > 0\) such that 
\[ H(t,x,y,\gamma - B) \geq H(t,x,y,\gamma + C \text{Tr}[B]) \quad \forall B \geq 0. \]
Then the PDE (4.6) is uniformly parabolic, and there exist general results on existence, 
quickness and smoothness of solutions, see for instance, [45]. When \(H\) is linear in the 
\(\gamma\) variable (in particular, for the semi- and quasilinear equations discussed in Section 5.2), 
uniform ellipticity essentially guarantees existence, uniqueness and smoothness of solu-
tions to the PDE (4.6) and (2.20); see for instance, Section 5.4 in [47].

4. Condition (A4) is an implicit assumption on the functions \(H\) and \(\phi\) as we find it 
more convenient to assume comparison directly in the form (A4) instead of placing tech-
nical assumptions on \(H\) and \(\phi\) which guarantee that the PDE (4.6) with terminal condi-
tion (2.20) has comparison. However, several comparison results for nonlinear PDEs are 
available in the literature; see for example, Crandall, Ishii and Lions [24], Fleming and 
Soner [36], Cabre and Caffarelli [18]. However, most results are stated for equations in 
bounded domains. For equations in the whole space, the critical issue is the interplay be-
tween the growth of solutions at infinity and the growth of the nonlinearity. We list some 
typical situations where comparison holds:

(a) Comparison with growth 1. Assume (A1)–(A3) and there exists a function 
\(h: [0, \infty] \to [0, \infty]\) with \(\lim_{x \to 0} h(x) = 0\) such that 
\[ |H(t,x,y,\alpha(x - \tilde{x}),A) - H(t,\tilde{x},y,\alpha(x - \tilde{x}),B)| \leq h(\alpha|x - \tilde{x}|^2 + |x - \tilde{x}|) \]
for all \((t,x,\tilde{x},y), \alpha > 0\) and \(A, B\) satisfying 
\[ -\alpha \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \leq \begin{bmatrix} A & 0 \\ 0 & -B \end{bmatrix} \leq \alpha \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}. \]
Then it follows from Theorem 8.2 in [24] that equations of the form (4.6), (2.20) have 
comparison with growth 0 if the domain is bounded. If the domain is unbounded, it follows 
from the modifications outlined in Section 5.D of Crandall et al. [24] that (4.6) and (2.20) 
have comparison with growth 1.

(b) For the dynamic programming equation (3.25) related to a stochastic optimal con-
trol problem, a comparison theorem for bounded solutions is given in [36], Section 5.9, 
Theorem V.9.1.

(c) Many techniques in dealing with unbounded solutions were developed by Ishii [42] 
for first-order equations (that is, when \(f\) is independent of \(\gamma\)). These techniques can be
extended to second-order equations. Some related results can be found in [4,5]. In [5], in addition to comparison results for PDEs, one can also find BSDEs based on jump Markov processes.

5. Monte Carlo methods

In this section we provide a formal discussion of the numerical implications of our representation results. We start by recalling some well-known facts in the linear case. We then review some results for the semilinear and the quasilinear cases. Then, we conclude with the fully nonlinear case related to Theorem 4.3.

5.1. The linear case

In this subsection we assume that the function $H$ is of the form

$$H(t, x, y, z, \gamma) = -\alpha(t, x) - \beta(t, x)y - \mu(x) \cdot z - \frac{1}{2} a(x) : \gamma.$$  

Then (4.6) is a linear parabolic equation and we discussed already that the Feynman–Kac representation has the form

$$v(t, x) = E\left[\int_t^T B_{t,s} \alpha(s, X^{t,x}(s)) \, ds + B_{t,T} g(X^{t,x}(T))\right],$$

where

$$B_{t,s} := \exp\left(\int_t^s \beta(r, X^{t,x}(r)) \, dr\right)$$

(see, for instance, Theorem 5.7.6 in [44]). This representation suggests a numerical approximation of the function $v$ by means of the so-called Monte Carlo method.

(i) Given $J$ independent copies \(\{X^j, 1 \leq j \leq J\}\) of the process $X^{t,x}$, set

$$\tilde{v}^{(J)}(t, x) := \frac{1}{J} \sum_{j=1}^J \int_t^T B_{t,s}^j \alpha(s, X^j(s)) \, ds + B_{t,T}^j g(X^j(T)),$$

where $B_{t,s}^j := \exp(\int_t^s \beta(r, X^j(r)) \, dr)$. Then, it follows from the law of large numbers and the central limit theorem that

$$\tilde{v}^{(J)}(t, x) \to v(t, x) \quad \text{a.s. and}$$

$$\sqrt{J} \left(\tilde{v}^{(J)}(t, x) - v(t, x)\right) \to N(0, \rho) \quad \text{in distribution},$$
where \( \rho \) is the variance of the random variable \( \int_t^T B_{t,s} \alpha(s, X_t^{t,x}(s)) \, ds + B_{t,T} g(X_t^{t,x}(T)) \). Hence, \( \hat{v}^{(J)}(t,x) \) is a consistent approximation of \( v(t,x) \). Moreover, in contrast to finite differences or finite elements methods, the error estimate is of order \( J^{-1/2} \), independent of the dimension \( d \).

(ii) In practice, it is not possible to produce independent copies \( \{X_j, 1 \leq j \leq J\} \) of the process \( X_t^{t,x} \), except in trivial cases. In most cases, the above Monte Carlo approximation is performed by replacing the process \( X_t^{t,x} \) by a suitable discrete-time approximation \( X^N \) with time step of order \( N^{-1} \) for which independent copies \( \{X^N_j, 1 \leq j \leq J\} \) can be produced. The simplest discrete-time approximation is the following discrete Euler scheme:

Set \( X^N_t = x \) and for \( 1 \leq n \leq N \),

\[
X^N_{tn} = X^N_{tn-1} + \mu(X^N_{tn-1}) (t_n - t_{n-1}) + \sigma(X^N_{tn-1}) (W_{tn} - W_{tn-1}),
\]

where \( t_n := t + n(T-t)/N \). We refer to [69] for a survey of the main results in this area.

5.2. The semilinear case

We next consider the case where \( H \) is given by

\[
H(t,x,y,z,\gamma) = \varphi(t,x,y,z) - \mu(x) \cdot z - \frac{1}{2} a(x) : \gamma.
\]

Then the PDE (4.6) is semilinear. We assume that the assumptions of Theorem 4.3 are satisfied. In view of the connection between Fisk–Stratonovich and Itô integration, the 2BSDE (4.4) reduces to an uncoupled forward–backward SDE (FBSDE) of the form

\[
dY(s) = \varphi(s, X_t^{t,x}(s), Y(s), Z(s)) \, ds + Z(s) \cdot \sigma(X_t^{t,x}(s)) \, dW(s),
\]

with terminal data \( Y(T) = g(X_t^{t,x}(T)) \) (compare to Peng [60], Pardoux and Peng [57]). For \( N \geq 1 \), we denote \( t_n := t + n(T-t)/N, n = 0, \ldots, N \), and we define the discrete-time approximation \( Y^N \) of \( Y \) by the backward scheme

\[
Y^N_T := g(X^T_{t,x})
\]

and, for \( n = 1, \ldots, N \),

\[
Y^N_{tn-1} := E[Y^N_{tn} | X^N_{tn-1} = \varphi(t_{n-1}, X^N_{tn-1}, Y^N_{tn-1}, Z^N_{tn-1}) (t_n - t_{n-1})], \quad (5.9)
\]

\[
Z^N_{tn-1} := \frac{1}{t_n - t_{n-1}} \left( \sigma(X^N_{tn-1}) \right)^{-1} E[(W_{tn} - W_{tn-1}) Y^N_{tn} | X^N_{tn-1}]. \quad (5.10)
\]

Then, we have

\[
\limsup_{N \to \infty} \sqrt{N} | Y^N_t - v(t,x) | < \infty,
\]
Stochastic representations for nonlinear parabolic PDEs

1 and in case that \( v \) is \( C^2 \) also,

\[
\limsup_{N \to \infty} \sqrt{N} |Z_t^N - Dv(t, x)| < \infty,
\]

see for instance, Bally and Pagès [3], Bouchard and Touzi [14]. The practical implementation of this backward scheme requires the computation of the conditional expectations appearing in (5.9) and (5.10). This suggests the use of a Monte Carlo approximation, as in the linear case. But at every time step, we need to compute conditional expectations based on \( J \) independent copies \( \{X_j, 1 \leq j \leq J\} \) of the process \( X_{t,x} \). Recently, several approaches to this problem have been developed. We refer to Bally and Pagès [3], Bouchard and Touzi [14], Lions and Regnier [49] and the references therein for the methodology and the analysis of such nonlinear Monte Carlo methods.

We refer to Chevance [22] and to a recent article by Delarue and Menozzi [26] for Monte Carlo simulations for the quasilinear case using forward–backward stochastic equations.

5.3. The fully nonlinear case

We now discuss the case of a general \( H \) as in the previous section. Let \( \mu, \sigma \) be as in the dynamics (2.12)

\[
\tilde{H}(t, x, y, z, \gamma) = H(t, x, y, z, \gamma) + \mu(t, x) \cdot z + \frac{1}{2} \alpha(t, x) : \gamma.
\]

Then, for all \((t, x)\) the 2BSDE corresponding to \((X_{t,x}, H, \varphi)\) can be written as

\[
dY(s) = \tilde{H}(s, X_{t,x}(s), Y(s), Z(s), \Gamma(s)) \, ds \\
+ Z(s) \cdot \sigma(s, X_{t,x}(s)) \, dW(s), \quad s \in [t, T),
\]

\[
dZ(s) = \mathcal{A}(s) \, ds + \Gamma(s) \, dX_{t,x}(s), \quad s \in [t, T),
\]

\[
Y(T) = g(X_{t,x}(T)).
\] (5.11)

We assume that the conditions of Theorem 4.3 hold true, so that the PDE (4.6) has a unique viscosity solution \( v \) with growth \( p = \max\{p_2, p_3, p_2p_4, p_4 + 2p_1\} \), and there exists a unique solution \((Y_{t,x}^{L}, Z_{t,x}^{L}, \Gamma_{t,x}^{L})\) to the 2BSDE (5.11) with \( Z_{t,x}^{L} \in A_{t,x}^{L} \).

Comparing with the backward scheme (5.9), (5.10) in the semilinear case, we suggest the following discrete-time approximation of the processes \( Y_{t,x}^{L}, Z_{t,x}^{L} \) and \( \Gamma_{t,x}^{L} \):

\[
Y_{tn}^{N} := g(X_{tn}^{L}), \quad Z_{tn}^{N} := Dg(X_{tn}^{L}),
\]

and, for \( n = 1, \ldots, N, \)

\[
Y_{tn-1}^{N} := E[Y_{tn}^{N} | X_{tn-1}^{L}] - \varphi(t_{n-1}, X_{tn-1}^{L}, Y_{tn-1}^{N}, Z_{tn-1}^{N}, \Gamma_{tn-1}^{N})(t_n - t_{n-1})
\]

\[
Z_{tn-1}^{N} := \frac{1}{t_n - t_{n-1}} (\sigma(X_{tn-1}^{L})^{-1} E[(W_{tn} - W_{tn-1})Y_{tn}^{N} | X_{tn-1}^{L}]).
\]
and

\[
\Gamma_{t_{n-1}}^N := \frac{1}{t_n - t_{n-1}} E \left[ Z_{t_n}^N (W_{t_n} - W_{t_{n-1}}) \right] \left| X_{t_{n-1}}^{I,x} \right| \sigma (X_{t_{n-1}}^{I,x})^{-1}.
\]

We expect that

\[
(Y_t^N, Z_t^N, \Gamma_t^N) \rightarrow (v(t,x), Dv(t,x), D^2v(t,x)) \quad \text{as } N \rightarrow \infty.
\]

However, proof of the above assertion is not yet available.

Acknowledgement

This work was partly supported by the Turkish Academy of Sciences.

References

Stochastic representations for nonlinear parabolic PDEs

H.M. Soner