2

Weak Convergence and Martingales

2.0 Outline of the Chapter

This chapter contains a brief review of two of the main mathematical methods for dealing with the convergence of a sequence of approximations to a stochastic process or for showing that a sequence of stochastic processes has a limit and for characterizing it. The first method is the theory of weak convergence of a sequence of probability measures. The theory, which is an extension to a sequence of random processes of the theory of convergence in probability of a sequence of random variables, provides powerful tools for approximation and limit theorems. Once one knows that the sequence of processes of concern has a limit, that limit must be characterized. The methods of the so-called martingale problem are a standard and powerful approach to doing such a characterization, when the limit is a diffusion-type process.

The numerical approximations of concern will be representable as controlled Markov chains with multistep memory. The convergence and approximation theorems of the later chapters, which are based on weak convergence theory and the methods of the martingale problem, show that the expectations of a large set of functionals of these chains converge to the values for the original process, as the approximation parameter goes to zero. In particular, the optimal cost values converge to the optimal value for the original controlled process of interest. Also, suitable interpolations of the sequence of approximating chains, under their optimal controls, converges to an optimal limit process. In addition, for numerical purposes one often approximates the original model and uses that for the numerical computations. Then one must show that these approximations do indeed provide results that are close to those for the original model. The same methods are employed for these purposes. Only an outline of the results that are of main use to us will be given. The comprehensive references [8, 23] contain full details and much additional information. The references [55, 58, 61, 56] contain many applications of these methods to control and communications systems or to numerical approximations.

H.J. Kushner, *Numerical Methods for Controlled Stochastic Delay Systems*,
doi: 10.1007/978-0-8176-4621-9_2,

2.1 Weak Convergence

Let $I\!R^k$ denote Euclidean r-space with canonical value $x = (x_1, \ldots, x_k)$, and let $\{X_n\}$ be a sequence of $I\!R^k$-valued random variables on a probability space $(\Omega, \mathcal{F}, P)$. If there is an $I\!R^k$-valued random variable X such that $Ef(X_n) \to Ef(X)$ for each bounded and continuous real-valued function $f(\cdot)$ on $I\!R^k$, then X_n *converges to* X *in distribution*. The sequence $\{X_n\}$ is said to be *tight* or, equivalently said, *bounded in probability* if

$$\lim_{K\to\infty} \sup_n P\{|X_n| \geq K\} = 0. \tag{1.1}$$

An equivalent definition is that for each small $\mu > 0$ there are finite M_μ and K_μ such that $P\{|X_n| \geq K_\mu\} \leq \mu$ for $n \geq M_\mu$. Convergence in distribution is also called *weak convergence*. Tightness is a necessary and sufficient condition that any subsequence of $\{X_n\}$ have a further subsequence that converges in distribution [10, 23].

Let $\{\xi_n\}$ be a sequence of mutually independent and identically distributed real-valued random variables, with mean zero and unit variance and $w(\cdot)$ a real-valued Wiener process with unit variance parameter. For $t > 0$ define

$$w^n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \xi_i, \tag{1.2}$$

where $[nt]$ denotes the integer part of nt. Then the central limit theorem says that $w^n(t)$ converges in distribution to a normally distributed random variable with mean zero and variance t.

For an integer k, and $0 = t_0 < t_1 < \cdots < t_{k+1}$, the multivariate central limit theorem [10] says that $\{w^n(t_{i+1}) - w^n(t_i),\, i \leq k\}$ converges in distribution to $\{w(t_{i+1}) - w(t_i), i \leq k\}$. Now consider $w^n(\cdot)$ to be a random process with paths that are constant on the intervals $[i/n, (i+1)/n)$. It is then natural to ask whether the sequence of processes $w^n(\cdot)$ converges to $w(\cdot)$ in a stronger sense. For example, will the distribution of the maximum $\max\{w^n(t) : t \leq 1\}$ converge in distribution to $\max\{w(t) : t \leq 1\}$? Donsker's theorem states that $F(w^n(\cdot))$ converges in distribution to $F(w(\cdot))$ for a large class of functionals $F(\cdot)$ [7, 23], for example for measurable $F(\cdot)$ that depend on only a finite segment of the path and are continuous almost everywhere with respect to the measure of $w(\cdot)$. This is an example of the theory of weak convergence.

The two main steps in getting the limit theorems for random processes are analogous to what is done for proving the central limit theorem: First show that there are appropriately convergent subsequences and then identify the limits. For vector-valued random variables, the necessary and sufficient condition (1.1) for the first step says that, neglecting an n-dependent set of small (uniformly in n) probability, the values of the random variables X_n are confined to some compact set. When random processes replace random variables, there will be an analogous condition ensuring that the paths are in a compact set with a "high probability."

2.1.1 Basic Theorems of Weak Convergence

Definitions. Let S denote a metric space with metric $\rho(\cdot)$ and $C(S)$ the set of real-valued continuous functions on S, with $C_b(S)$ being the subset of bounded functions. Let $\mathcal{B}(S)$ denote the collection of Borel subsets of S. Let $\mathcal{P}(S)$ denote the space of probability measures on $(S, \mathcal{B}(S))$. Let $X_n, n < \infty$, and X be S-valued random variables, with distributions $P_n, n < \infty$, and P, respectively. The sequence $\{X_n, n < \infty\}$ is said to *converge in distribution* to X if $Ef(X_n) \to Ef(X)$ for all $f \in C_b(S)$ or, equivalently written, if $\int_S f(s)P_n(ds) \to \int_S f(s)P(ds)$. This is called *weak convergence* and written as $P_n \Rightarrow P$. We will often say that the sequence of random variables X_n converges weakly to X, and denote this by $X_n \Rightarrow X$ as well. The X will be said to be the *weak-sense limit.*

For $\lambda \in \Lambda$, an arbitrary index set, let $P_\lambda \in \mathcal{P}(S)$. The set $\{P_\lambda, \lambda \in \Lambda\}$ is called *tight* if for each $\varepsilon > 0$ there is a compact set $K_\varepsilon \subset S$ such that

$$\inf_{\lambda \in \Lambda} P_\lambda(K_\varepsilon) \geq 1 - \varepsilon. \tag{1.3}$$

If P_λ is the measure defined by an S-valued random variable X_λ, then we will also say that $\{X_\lambda, \lambda \in \Lambda\}$ is tight. If all of the X_λ are defined on the same probability space, then (1.3) is equivalent to

$$\inf_{\lambda \in \Lambda} P\{X_\lambda \in K_\varepsilon\} \geq 1 - \varepsilon. \tag{1.4}$$

The Prohorov metric. Let $P_i \in \mathcal{P}(S), i = 1, 2$. For $A \in \mathcal{B}(S)$, define the set $A^\varepsilon = \{s' : \rho(s', s) < \varepsilon \text{ for some } s \in A\}$. Then the *Prohorov metric* $\pi(\cdot)$ on $\mathcal{P}(S)$ is defined by

$$\pi(P_1, P_2) = \inf\{\varepsilon > 0 : P_1(A) \leq P_2(A^\varepsilon) + \varepsilon \text{ for all closed } A \in \mathcal{B}(S)\},$$

and is always used on the space $\mathcal{P}(S)$. The following two theorems are fundamental.

Theorem 1.1. [23, page 101.] *If S is complete and separable, then $\mathcal{P}(S)$ is complete and separable.*

Theorem 1.2. [23, Theorem 3.2.2.] *If S is complete and separable, then a set $\{P_\lambda, \lambda \in \Lambda\} \subset \mathcal{P}(S)$ has compact closure if and only if $\{P_\lambda, \lambda \in \Lambda\}$ is tight.*

Suppose that S is complete and separable and that a given sequence of probability measures has compact closure (Prohorov metric). Theorem 1.2 then implies the existence of a convergent subsequence [19, Theorem 13, page 21]. The theorem gives a practical method for verifying the compact closure property, as tightness is also a property of the random variables (or processes) associated with the P_λ. These random variables typically have explicit representations (for example, they might be solutions to a stochastic differential equation) that can be used to verify the tightness property. A sequence of vector-valued random variables is tight if the sequence of each of its components is tight, as asserted in the next result.

Corollary 1.3. *Let S_1 and S_2 be complete and separable metric spaces, and define $S = S_1 \times S_2$ with the usual product space topology. For $\{P_\lambda, \lambda \in \Lambda\} \subset \mathcal{P}(S)$, let $P_{\lambda,i}$ be the marginal distribution of P_λ on S_i. Then $\{P_\lambda, \lambda \in \Lambda\}$ is tight if and only if $\{P_{\lambda,i}, \lambda \in \Lambda\}$, $i = 1, 2$, are tight.*

The next theorem contains some statements that are equivalent to weak convergence. Let ∂B be the boundary of the set $B \in \mathcal{B}(S)$.

Theorem 1.4. [23, Theorem 3.3.1.] *Let S be a metric space and let P_n, $n < \infty$, and P be elements of $\mathcal{P}(S)$. Then statements (i)–(iv) below are equivalent and are implied by (v). If S is separable, then (i)–(v) are equivalent:*

$$
\begin{aligned}
(i)&\ P_n \Rightarrow P,\\
(ii)&\ \limsup_n P_n(F) \le P(F) \text{ for closed sets } F,\\
(iii)&\ \liminf_n P_n(O) \ge P(O) \text{ for open sets } O,\\
(iv)&\ \lim_n P_n(B) = P(B) \text{ if } P(\partial B) = 0,\\
(v)&\ \pi(P_n, P) \to 0.
\end{aligned}
$$

The theorem implies that, for separable S, convergence in the Prohorov metric is equivalent to weak convergence. Part (iv) implies the following important extension of the class of functionals that converge in distribution.

Theorem 1.5. [7, Theorem 5.1.] *Let S be a metric space, and let $P_n, n < \infty$, and P be probability measures on $\mathcal{P}(S)$ satisfying $P_n \Rightarrow P$. Let $f(\cdot)$ be a real-valued measurable function on S and define D_f to be the measurable set of points at which $f(\cdot)$ is not continuous. Let X_n and X be random variables that induce the measures P_n and P on S, respectively. Then $f(X_n) \Rightarrow f(X)$ whenever $P\{X \in D_f\} = 0$.*

The Skorokhod representation. Suppose that $X_n \Rightarrow X$, where the X_n and X might be defined on different probability spaces. The probability spaces are unimportant, as weak convergence is a statement on the measures of the random variables. But for the purpose of characterizing the weak-sense limit X, it can be very useful to have all processes defined on the same space and weak convergence replaced by probability one convergence. This can be done without changing the distributions of the X_n or X. The result is known as the *Skorokhod representation* [23].

Theorem 1.6. [23, Theorem 3.1.8.] *Let S be a separable metric space, and assume that $P_n \in \mathcal{P}(S)$ converges weakly to $P \in \mathcal{P}(S)$ as $n \to \infty$. Then there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ on which there are defined random variables $\tilde{X}_n, n < \infty$, and $\tilde{X}$ such that for all Borel sets B and all $n < \infty$,*

$$\tilde{P}\left\{\tilde{X}_n \in B\right\} = P_n(B), \qquad \tilde{P}\left\{\tilde{X} \in B\right\} = P(B), \tag{1.5}$$

and such that

$$\tilde{X}_n \to \tilde{X} \text{ with probability one.} \tag{1.6}$$

2.1.2 The Function Spaces $D(S; I)$

For a complete and separable metric space S, let $D(S; I)$ denote the set of S-valued functions on the interval I that are right continuous and have left-hand limits. Let $C(S; I)$ denote the subset of continuous functions. The interval I will be either $[0, \infty), [t_1, \infty)$ or $[t_1, t_2]$ for some $t_1 < t_2$. Even when the weak-sense limit processes have continuous paths, it is usually easier to prove tightness and weak convergence using the path spaces $D(S; [0, \infty))$. If there are Poisson jumps in the system dynamics or if the control is of an impulsive or singular nature, then $D(S; [0, \infty))$ must be used. We next define the metric.

The Skorokhod metric [23, Chapter 3.5], [7, Chapter 3]. For $T > 0$, let Λ_T denote the space of continuous and strictly increasing functions from $[0, T]$ onto $[0, T]$. The functions in this set will be "allowable timescale distortions." For $\lambda(\cdot) \in \Lambda_T$ define

$$|\lambda| = \sup_{s<t} \left| \log \left\{ \frac{\lambda(t) - \lambda(s)}{t - s} \right\} \right| .$$

The Skorokhod metric $d'_T(\cdot)$ on $D(I\!R^k; [0, T])$ is defined by, for $\lambda(\cdot) \in \Lambda_T$,

$$d'_T(f(\cdot), g(\cdot)) = \inf\{\epsilon : |\lambda| \le \epsilon, \ \sup_{0 \le s \le T} |f(s) - g(\lambda(s))| \le \epsilon, \text{ for some } \lambda(\cdot)\}. \tag{1.7}$$

On the space $D(I\!R^k; [0, \infty))$, the metric is defined by

$$d'(f(\cdot), g(\cdot)) = \int_0^\infty e^{-t} \min\left[1, d'_t(f(\cdot), g(\cdot))\right] dt. \tag{1.8}$$

Now let S be a complete and separable metric space with metric $\rho(\cdot)$. Then the Skorokhod metric on the spaces $D(S; [0, T])$ is defined by the $d'_T(\cdot)$ above, but with $\rho(f(s), g(\lambda(s)))$ used in place of $|f(s) - g(\lambda(s))|$, where both $f(\cdot)$ and $g(\cdot)$ are now points in $D(S; [0, T])$. Define the space $D(S; [0, \infty))$ analogously. If S is complete and separable, then so are $D(S; [0, T])$ and $D(S; [0, \infty))$ [23].

If $f_n(\cdot) \to f(\cdot)$ in $d_T(\cdot)$ where $f(\cdot)$ is continuous, then the convergence must be uniform on $[0, T]$. If there are $\eta_n \to 0$ such that the discontinuities of $f_n(\cdot)$ are less than η_n in magnitude and if $f_n(\cdot) \to f(\cdot)$ in $d_T(\cdot)$, then the convergence is uniform on $[0, T]$ and $f(\cdot)$ must be continuous. Because of the "timescale distortion" that is involved in the definition of the metric $d_T(\cdot)$, we can have (loosely speaking) convergence of a sequence of discontinuous functions where there are only a finite number of discontinuities, where both the locations and the values of the discontinuities converge, and a type of "equicontinuity" condition holds between the discontinuities. See [7, 23] for full detail.

A criterion for tightness in $D(S; [0, T])$ and $D(S; [0, \infty))$. The following criterion for tightness will be used. Recall that for a filtration $\{\mathcal{F}_t, t \geq 0\}$, the random time τ is an $\mathcal{F}_t$-stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in [0, \infty)$.

Theorem 1.7. [49, Theorem 2.7b.] *Let $x^n(\cdot)$ be processes with paths in $D(S; [0, \infty))$, where S is a complete and separable metric space with metric $\rho(\cdot)$. For each $\delta > 0$ and rational $t < \infty$, let there be a compact set $S_{\delta,t} \subset S$ such that*

$$\sup_n P(x^n(t) \notin S_{\delta,t}) \leq \delta. \tag{1.9}$$

Let $\mathcal{F}_t^n$ be the σ-algebra determined by $\{x^n(s),\ s \leq t\}$ and let $\mathcal{T}_n(T)$ be the set of $\mathcal{F}_t^n$-stopping times that are no bigger than T. Suppose that

$$\lim_{\delta \to 0} \limsup_n \sup_{\tau \in \mathcal{T}_n(T)} E \min\{1, \rho(x^n(\tau + \delta), x^n(\tau))\} = 0 \tag{1.10}$$

for each $T < \infty$. Then $\{x^n(\cdot),\ n < \infty\}$ is tight in $D(S; [0, \infty))$.

Let $C(G; [a, b])$ denote the space of S-valued continuous functions on the interval $[a, b]$ with the sup norm topology, where G is a compact subset of a Euclidean space. If the interval $[a, b]$ is unbounded, then the local sup norm topology is used. The next theorem gives a necessary and sufficient condition for tightness of sequence of G-value continuous processes.

Theorem 1.8. [8, Theorem 7.3.] *The sequence of G-valued processes $x^n(\cdot)$ is tight in $C(G; [0, 1])$ if and only if: For each $\epsilon > 0$ and $\eta > 0$, there is a $\delta > 0$ and an $n_0 < \infty$ such that, for $n \geq n_0$,*

$$P\left\{\sup_{i\delta < 1} \sup_{s \leq \delta} |x^n(s) - x^n(i\delta)| \geq \epsilon\right\} \leq \eta.$$

2.2 Martingales and the Martingale Method

2.2.1 Martingales

Definitions. Let $(\Omega, \mathcal{F}, P)$ denote a probability space. It will always be assumed that $\mathcal{F}$ is *complete;* i.e., it contains all subsets of P-null sets.

Let $\overline{\mathcal{F} \times \mathcal{B}([0,\infty))}$ denote the completion of the product σ-algebra with respect to the product measure, with Lebesgue measure used on $\mathcal{B}([0,\infty))$. A function $\phi(\cdot)$ on $\Omega \times [0,\infty)$ and with values $\phi(\omega,t)$ in some metric space S is said to be a *measurable process* if it is a measurable mapping from $(\Omega \times [0,\infty), \overline{\mathcal{F} \times \mathcal{B}([0,\infty))})$ to $(S, \mathcal{B}(S))$. All processes are assumed to be measurable and separable. It is always assumed that S and $C(S;[0,\infty))$ and $D(S;[0,\infty))$ are complete and separable metric spaces ([7, 23]).

A family of σ-algebras $\{\mathcal{F}_t, t \geq 0\}$ is called a *filtration* on this probability space if $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for all $0 \leq s \leq t$. We will always assume that the $\mathcal{F}_t$ are complete in that $\mathcal{F}_t$ contains all the subsets of null sets in $\mathcal{F}$. If A is a collection of random variables defined on $(\Omega, \mathcal{F}, P)$, then we use $\mathcal{F}(A)$ to denote the σ-algebra generated by A. Let $E_{\mathcal{F}_t}$ and $P_{\mathcal{F}_t}$ denote the expectation and probability, respectively, conditioned on the σ-algebra $\mathcal{F}_t$.

Let $M(\cdot)$ be a stochastic process defined on $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t, t \geq 0\}$. If $M(t)$ is $\mathcal{F}_t$-measurable for each t, then $M(\cdot)$ is said to be $\mathcal{F}_t$-adapted. Let $M(\cdot)$ be $\mathcal{F}_t$-adapted and take values in the path space $D(I\!R^k;[0,\infty))$. Then $M(\cdot)$ is said to be an $\mathcal{F}_t$*-martingale* if $E|M(t)| < \infty$ for all $t \geq 0$ and

$$E_{\mathcal{F}_t} M(t+s) = M(t) \text{ w.p.1 for all } s,t \geq 0. \tag{2.1}$$

If the filtration is unimportant or obvious, then we will simply say that $M(\cdot)$ is a martingale. If $M(\cdot)$ is an $\mathcal{F}_t$-martingale, then it is also an $\mathcal{F}(M(s), s \leq t)$-martingale. We say that $\{\mathcal{F}(M(s), s \leq t), t \geq 0\}$ is the filtration generated by $M(\cdot)$.

Martingales are a fundamental tool in stochastic analysis. Processes can often be decomposed into a sum of a process of bounded variation and a martingale. This decomposition can be used to facilitate the analysis, as the bounded-variation term is often relatively easy to handle, and there are many useful techniques for the analysis of martingales. The following inequalities will be useful. Let $M(\cdot)$ be a real or vector-valued $\mathcal{F}_t$-martingale with paths in $D(I\!R^k;[0,\infty))$ for some $k \geq 1$. Then [10, 16, 42, 76] for any $c > 0$ and $0 \leq t \leq T$,

$$P_{\mathcal{F}_t} \left\{ \sup_{t \leq s \leq T} |M(s)| \geq c \right\} \leq E_{\mathcal{F}_t} |M(T)|^2 / c^2 \text{ w.p.1}, \tag{2.2}$$

$$E_{\mathcal{F}_t} \sup_{t \leq s \leq T} |M(s)|^2 \leq 4 E_{\mathcal{F}_t} |M(T)|^2 \quad \text{w.p.1}. \tag{2.3}$$

Stopping time. Let $\{\mathcal{F}_t, t \geq 0\}$ be a filtration. If $M(\cdot)$ is an $\mathcal{F}_t$-martingale and τ is an $\mathcal{F}_t$-stopping time, then the "stopped" process defined by $M(t \wedge \tau)$ is also an $\mathcal{F}_t$-martingale [10, 76]. Let $\mathcal{F}_\tau$ denote the "stopped" σ-algebra that is composed of the sets $A \in \mathcal{F}$ such that $A \cap \{\tau \leq t\} \in \mathcal{F}_t$ for all t.

2.2.2 Verifying That a Process Is a Martingale

We now give a method that will be useful in showing that a process is a martingale. It is only a rewording of the definition of a martingale in terms of conditional expectations.

Let Y be a vector-valued random variable with $E|Y| < \infty$, and let $V(\cdot)$ be a process with paths in $D(S;[0,\infty))$, where S is a complete and separable metric space. Suppose that for some given $t > 0$, each integer p and each set of real numbers $0 \le s_i \le t$, $i = 1,\ldots,p$, and each bounded and continuous real-valued function $h(\cdot)$, $Eh(V(s_i), i \le p)Y = 0$. This fact and the arbitrariness of p, s_i, t, and of the function $h(\cdot)$ imply that

$$E[Y|V(s), s \le t] = 0$$

with probability one [10].

Next, let $U(\cdot)$ be a random process with $E|U(t)| < \infty$ for each t, with values in $D(S;[0,\infty))$, and such that for all p, $h(\cdot)$, $s_i \le t, i \le p$, as given above, and a given real $\tau > 0$,

$$Eh(U(s_i), i \le p)\,[U(t+\tau) - U(t)] = 0. \tag{2.4}$$

Then $E[U(t+\tau) - U(t)|U(s), s \le t] = 0$. If this holds for all t and $\tau > 0$, then by the definition (2.1) of a martingale, $U(\cdot)$ is a martingale with respect to the filtration generated by $U(\cdot)$. It is often more convenient to work with the following more general setup.

Theorem 2.1. *Let $U(\cdot)$ be a random process with paths in $D(I\!R^k;[0,\infty))$ and with $E|U(t)| < \infty$ for each t. Let $V(\cdot)$ be a process with paths in $D(S;[0,\infty))$, where S is a complete and separable metric space. Let $U(t)$ be measurable on the σ-algebra $\mathcal{F}_t^V$ determined by $\{V(s), s \le t\}$. Suppose that for each real $t \ge 0$ and $\tau \ge 0$, each integer p, and each set of real numbers $s_i \le t$, $i = 1,\ldots,p$, and each bounded and continuous real-valued function $h(\cdot)$,*

$$Eh(V(s_i), i \le p)\,[U(t+\tau) - U(t)] = 0. \tag{2.5}$$

Then $U(\cdot)$ is an $\mathcal{F}_t^V$-martingale.

An application. A sufficient condition for a Wiener process. The numerical approximations can be represented as processes that have the drift of the original diffusion and are driven by martingales. For the convergence proofs one needs to prove that these martingales converge to a Wiener process. The following result is useful for this purpose.

A process $v(\cdot)$ is said to be nonanticipative with respect to a Wiener process $w(\cdot)$ if $w(\cdot)$ is a martingale with respect to the filtration generated by $(v(\cdot), w(\cdot))$. Equivalently, for all t, $w(t+\cdot) - w(t)$ is independent of $\{v(s), w(s), s \le t\}$. Let $x(\cdot)$ and $z(\cdot)$ be $I\!R^r$-valued continuous processes with $z(\cdot)$ having bounded variation (w.p.1) on any bounded time interval. Let $b(\cdot), \sigma(\cdot), x(\cdot)$ be measurable processes and define $a(t) = \sigma(t)\sigma'(t) = \{a_{i,j}(t); i,j\}$. For $f(\cdot)$ a real-valued function with compact support that is continuous and bounded together with its first and second derivatives, define

$$\mathcal{L}f(x(t)) = f'_x(x(t))b(t) + \frac{1}{2}\sum_{i,j} a_{i,j}(t) f_{x_i x_j}(x(t)).$$

Let $h(\cdot)$ be a bounded and continuous function of its arguments, and for an integer k, and nonnegative t, T, let $0 \leq t_1 \leq \cdots \leq t_k < t < t+T$, and let $f(\cdot)$ be as above. Suppose that for all such $h(\cdot), f(\cdot), k, t, T, t_i$, we have

$$\begin{aligned} &Eh\left(x(t_i), b(t_i), \sigma(t_i), z(t_i), i \leq k\right) \\ &\qquad \times \left[f(x(t+T)) - f(x(t)) - \int_t^T \left(\mathcal{L}f(x(s))ds + f'_x(x(s))dz(s)\right)\right] = 0. \end{aligned} \tag{2.6}$$

Then there is a standard $I\!R^r$-valued Wiener process on the same probability space (perhaps augmenting the space by adding an "independent" Wiener process) such that [42, Chapter 5, Proposition 4.6] $x(\cdot), z(\cdot), b(\cdot), \sigma(\cdot)$ are nonanticipative with respect to $w(\cdot)$, and

$$x(t) = x(0) + \int_0^t b(s)ds + \int_0^t \sigma(s)dw(s) + z(t).$$

http://www.springer.com/978-0-8176-4534-2

Numerical Methods for Controlled Stochastic Delay
Systems
Kushner, H.
2008, XX, 282 p. 37 illus., Hardcover
ISBN: 978-0-8176-4534-2
A product of Birkhäuser Basel