Stochastic flows and stochastic differential equations

HIROSHI KUNITA

Professor, Department of Applied Science, Kyushu University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1990
First paperback edition 1997

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Kunita, H.
Stochastic flows and stochastic differential equations / Hiroshi Kunita.
p. cm. – (Cambridge studies in advanced mathematics: 24)
Includes bibliographical references.
ISBN 0 521 35050 6
1. Stochastic analysis. 2. Flows (Differentiable dynamical systems)
3. Stochastic differential equations. I. Title. II. Series.
QA274.2.K86 1990
519.2–dc20 89-70813 CIP

ISBN 0 521 35050 6 hardback
ISBN 0 521 59925 3 paperback

Transferred to digital printing 2002

Contents

1

Stochastic processes and random fields

1.1 Preliminaries

Probability spaces and random variables

Let Ω be a set. A collection $\mathscr{F}$ of subsets of Ω is called a *σ-field* if it contains an empty set and is closed under the operations of countable unions and complements. The pair $(\Omega, \mathscr{F})$ is called a *measurable space*. Elements of Ω are called *samples* and those of $\mathscr{F}$ are called *events*. Let P be a σ-additive measure on $(\Omega, \mathscr{F})$. It is called a *probability* if $P(\Omega) = 1$. The triple $(\Omega, \mathscr{F}, P)$ is called a *probability space*.

A finite collection of events $\{A_1, \ldots, A_n\}$ is called *independent* if $P(\bigcap_{l=1}^k A_{i_l}) = \prod_{l=1}^k P(A_{i_l})$ holds for any subset $\{A_{i_1}, \ldots, A_{i_k}\}$ of $\{A_1, \ldots, A_n\}$. A collection of infinite events $\{A_\lambda : \lambda \in \Lambda\}$ is called independent if any finite subcollection of $\{A_\lambda : \lambda \in \Lambda\}$ is independent. Let $\mathscr{G}$ be a subset of $\mathscr{F}$. If $\mathscr{G}$ is a σ-field, it is called a *sub σ-field* of $\mathscr{F}$. Suppose now we are given a finite collection $\{\mathscr{F}_1, \ldots, \mathscr{F}_n\}$ of sub σ-fields of $\mathscr{F}$. It is called independent if for every choice $A_i \in \mathscr{F}_i$. $i = 1, \ldots, n$, the collection $\{A_1, \ldots, A_n\}$ is independent. An infinite collection $\{\mathscr{F}_\lambda : \lambda \in \Lambda\}$ of sub σ-fields of $\mathscr{F}$ is called independent if its arbitrary finite subcollection is independent.

A real valued measurable function $X(\omega)$ defined on $(\Omega, \mathscr{F})$ is called a *real random variable* or simply a *random variable*. The random variable $X(\omega)$ may take values $\pm\infty$, but we assume that $X(\omega)$ takes finite values for almost all ω unless otherwise mentioned. We often suppress the sample ω and write it as X. If we can define the integral of X by the measure P, we denote it by $E[X]$ and call it the *expectation* of X, i.e.,

$$E[X] = \int_\Omega X(\omega)\,\mathrm{d}P(\omega). \tag{1}$$

In later discussions we will often use the following notation

$$E[X : A] = \int_A X(\omega)\,\mathrm{d}P(\omega). \tag{2}$$

Let S be a complete separable metric space and $\mathscr{B}(S)$ be its topological Borel field. A measurable map X from $(\Omega, \mathscr{F})$ into $(S, \mathscr{B}(S))$ is called a *random variable with values in S* or *S-valued random variable*. If X is a random variable with values in S and B is an element of $\mathscr{B}(S)$, then the set $\{\omega : X(\omega) \in B\}$ belongs to $\mathscr{F}$. For simplicity this set is denoted by $\{X \in B\}$ or $X \in B$. Now the collection of the sets $\{X \in B : B \in \mathscr{B}(S)\}$ is a sub σ-field of $\mathscr{F}$. It is called the *σ-field generated by the random variable* X and is denoted by $\sigma(X)$. Let $\{X_\lambda : \lambda \in \Lambda\}$ be a collection of random variables. The smallest sub σ-field of $\mathscr{F}$ containing $\bigcup_{\lambda \in \Lambda} \sigma(X_\lambda)$ is denoted by $\sigma(X_\lambda : \lambda \in \Lambda)$ and is called the σ-field generated by $\{X_\lambda : \lambda \in \Lambda\}$.

Two families of random variables $\{X_\lambda : \lambda \in \Lambda\}$ and $\{X_\gamma : \gamma \in \Gamma\}$ are called independent if $\sigma(X_\lambda : \lambda \in \Lambda)$ and $\sigma(X_\gamma : \gamma \in \Gamma)$ are independent. The independence of infinite families of random variables is defined similarly.

An event A is called a null event or a null set if $P(A) = 0$ holds. If a proposition holds except for ω belonging to a certain null set, it is said to hold *almost everywhere* (abbreviated as a.e.) or *almost surely* (abbreviated as a.s.). As an example for two random variables X and Y, '$X = Y$ a.s.' means that $\{\omega : X(\omega) \neq Y(\omega)\}$ is a null set. We often do not distinguish these X and Y and write simply $X = Y$.

For a sequence $X_1, X_2, \ldots, X$ of real random variables, we introduce three types of convergence.

(a) $\{X_n\}$ is said to *converge to X almost everywhere* or *almost surely* if for almost all ω, $\{X_n(\omega)\}$ converges to $X(\omega)$.
(b) Let $p \geq 1$. Denote by L^p the totality of random variables Y such that $E[|Y|^p] < \infty$ and define the L^p-norm by $\|Y\|_p = E[|Y|^p]^{1/p}$. If $X_1, X_2, \ldots, X$ are in L^p and $\|X_n - X\|_p \to 0$ is satisfied, $\{X_n\}$ is said to *converge to X in L^p*.
(c) $\{X_n\}$ is said to *converge to X in probability* if for any $\varepsilon > 0$ $P(|X_n - X| > \varepsilon)$ converges to 0.

We give the well known relations on these three convergences without proofs.

Theorem 1.1.1 *The almost everywhere convergence implies the convergence in probability. The L^p-convergence implies the convergence in probability.* $\square$

A collection of real random variables $\{X_\lambda\}$ is called *uniformly integrable* if

$$\sup_\lambda \int_{|X_\lambda| > c} |X_\lambda| \, \mathrm{d}P \xrightarrow[c \to \infty]{} 0 \tag{3}$$

is satisfied. For a sequence of uniformly integrable random variables, the convergence in probability implies the convergence in L^1.

Theorem 1.1.2 *Let $\{X_n\}$ be a sequence of uniformly integrable random variables. If $\{X_n\}$ converges to X in probability, then it converges in L^1.* □

Let $P_n : n = 1, 2, \ldots, P$ be a sequence of probabilities on $(S, \mathscr{B}(S))$. The sequence $\{P_n\}$ is said to *converge weakly* to P if for any bounded continuous function f on S, $\{\int f \, dP_n\}$ converges to $\int f \, dP$. Now for an S-valued random variable X, we define its *law* by the probability P_X on $(S, \mathscr{B}(S))$ such that

$$P_X(B) = P(X \in B), \qquad \text{for all } B \in \mathscr{B}(S). \tag{4}$$

A sequence $\{X_n\}$ of S-valued random variables is then said to *converge weakly* if the corresponding sequence of the laws converges weakly.

We shall quote some basic properties of the weak convergence. Proofs of the following theorems (1.1.3–1.1.5) can be found in Billingsley [8] and Ikeda–Watanabe [49].

Theorem 1.1.3 *Let $\{P_n : n = 1, 2, \ldots, P\}$ be a sequence of probabilities on $(S, \mathscr{B}(S))$. The following statements are equivalent.*

(a) *$\{P_n\}$ converges to P weakly.*
(b) *$\overline{\lim}_{n\to\infty} P_n(F) \leq P(F)$ holds for any closed subset F of S.*
(c) *$\underline{\lim}_{n\to\infty} P_n(G) \geq P(G)$ holds for any open subset G* of S. □

A family $\{P_\lambda : \lambda \in \Lambda\}$ of probabilities over $(S, \mathscr{B}(S))$ is called *relatively compact* if any subset of $\{P_\lambda : \lambda \in \Lambda\}$ contains a subsequence converging weakly. A useful criterion for the relative compactness of the measures is the tightness: a family $\{P_\lambda : \lambda \in \Lambda\}$ of probabilities is called *tight* (or *uniformly tight*) if for any $\varepsilon > 0$ there exists a compact subset K_ε of S such that $P_\lambda(K_\varepsilon) > 1 - \varepsilon$ holds for all $\lambda \in \Lambda$.

Theorem 1.1.4 *The family of probabilities $\{P_\lambda : \lambda \in \Lambda\}$ on $(S, \mathscr{B}(S))$ is relatively compact if and only if it is tight.* □

The following theorem, due to Skorohod, shows that the weak convergence and the strong convergence are equivalent if the corresponding random variables are defined on a suitable probability space.

Theorem 1.1.5 *Let $\{P_n\}$ be a sequence of probabilities on $(S, \mathscr{B}(S))$ converging weakly to P. Then on a suitable probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$ we can construct S-valued random variables $\tilde{X}_n$, $n = 1, 2, \ldots$ and $\tilde{X}$ satisfying the following properties.*

(a) *The laws of* $\tilde{X}_n$, $n = 1, 2, \ldots,$ *and* $\tilde{X}$ coincide with P_n, $n = 1, 2, \ldots$ *and* P, *respectively.*
(b) $\{\tilde{X}_n\}$ *converges to* $\tilde{X}$ *almost everywhere.* □

Conditional expectations

Let $\mathscr{G}$ be a sub σ-field of $\mathscr{F}$ and let X be an integrable random variable. An integrable $\mathscr{G}$-measurable random variable $\hat{X}$ is called the *conditional expectation of X with respect to* $\mathscr{G}$ if $\hat{X}$ satisfies

$$\int_A X \, \mathrm{d}P = \int_A \hat{X} \, \mathrm{d}P, \qquad \text{for all } A \in \mathscr{G}. \tag{5}$$

The conditional expectation exists uniquely. We denote it by $E[X|\mathscr{G}]$.

Theorem 1.1.6 *Let* X, Y, X_n, $n = 1, 2, \ldots$ *be integrable random variables and* $\mathscr{G}, \mathscr{H}$ *be sub* σ*-fields of* $\mathscr{F}$.

(c.1) *If* a, b *are constants, then* $E[aX + bY|\mathscr{G}] = aE[X|\mathscr{G}] + bE[Y|\mathscr{G}]$ *a.s.*
(c.2) *If* X *is* $\mathscr{G}$*-measurable and* XY *is integrable, then* $E[XY|\mathscr{G}] = XE[Y|\mathscr{G}]$ *a.s.*
(c.3) *If* $\mathscr{H} \subset \mathscr{G}$, *then* $E[E[X|\mathscr{G}]|\mathscr{H}] = E[E[X|\mathscr{H}]|\mathscr{G}] = E[X|\mathscr{H}]$ *a.s.*
(c.4) (Jensen's inequality) *Let* $f(x)$ *be a convex function. If* $f(X)$ *is integrable, then* $f(E[X|\mathscr{G}]) \le E[f(X)|\mathscr{G}]$ *a.s. In particular, if* $|X|^p$ *is integrable for* $p \ge 1$, *then* $|E[X|\mathscr{G}]|^p \le E[|X|^p|\mathscr{G}]$ *a.s.*
(c.5) *Let* $p \ge 1$. *If* $\{X_n\}$ *converges to* X *in* L^p, *then* $\{E[X_n|\mathscr{G}]\}$ *converges to* $E[X|\mathscr{G}]$ *in* L^p.
(c.6) *If* $\{X_n\}$ *converges to* X *with respect to the weak topology of* L^p, *then* $\{E[X_n|\mathscr{G}]\}$ *converges to* $E[X|\mathscr{G}]$ *with respect to the weak topology of* L^p. □

For the proof, see Neveu [103].

The *conditional probability of the event A given the* σ*-field* $\mathscr{G}$ is defined by

$$P(A|\mathscr{G}) = E[\chi(A)|\mathscr{G}], \tag{6}$$

where $\chi(A)$ is the indicator function of the set A. Then it has these three properties:

(a) $0 \le P(A|\mathscr{G}) \le 1$ a.s.
(b) $P(\Omega|\mathscr{G}) = 1$, $P(\phi|\mathscr{G}) = 0$ a.s.
(c) for any pairwise disjoint sets $A_1, A_2, \ldots,$ $P(\bigcup_{n=1}^{\infty} A_n|\mathscr{G}) = \sum_{n=1}^{\infty} P(A_n|\mathscr{G})$ a.s.

These are easily verified from the definition of the conditional probability.

Exercise 1.1.7 Let X and Y be independent random variables with values in complete separable metric spaces S and S', respectively. Let $g(x, y)$ be a real measurable function on $S \times S'$ such that $g(X, Y)$ is integrable. Show that

$$E[g(X, Y)|\sigma(Y)] = \int g(x, Y)P_X(\mathrm{d}x) \qquad \text{a.s.}$$

where P_X is the law of X.

1.2 Stochastic processes

Brownian motions

A collection of random variables X_t, $t \in \mathbb{T}$ with values in a complete separable metric space S where $\mathbb{T}$ is a time set is called a *stochastic process with state space S*. If $\mathbb{T}$ is an interval, it is called a *stochastic process with continuous parameter*. If $\mathbb{T}$ is a discrete subset of $\mathbb{R}$, it is called a *stochastic process with discrete parameter*. When a sample ω is fixed, $X_t(\omega)$, $t \in \mathbb{T}$ can be regarded as a function of t. It is called a *sample path* (or *sample function*) of the stochastic process. In this book we shall mainly consider stochastic processes with continuous parameter. In most cases the time set $\mathbb{T}$ will be the finite interval $[0, T]$, but the infinite interval $\mathbb{T} = [0, \infty)$ or $\mathbb{T} = (-\infty, 0]$ will be dealt with in some cases.

In this section we define three basic stochastic processes called Brownian motions, martingales and Markov processes, which are the central topics in this book. We first introduce some general notions on stochastic processes.

Let X_t, $t \in \mathbb{T}$ be a stochastic process with continuous parameter. It is called *measurable* if $X : \mathbb{T} \times \Omega \to S$ is measurable with respect to the product σ-field $\mathscr{B}(\mathbb{T}) \otimes \mathscr{F}$. The continuity of a stochastic process is defined similarly as the convergence of random variables. Let X_t, $t \in \mathbb{T}$ be a real valued stochastic process. It is called *continuous in probability* if for any $t \in \mathbb{T}$, X_{t+h} converges to X_t in probability as h tends to 0. If X_t is in L^p and $\lim_{h\to 0} \|X_{t+h} - X_t\|_p = 0$ holds for any t, it is called *continuous in L^p*. Obviously a stochastic process continuous in L^p is continuous in probability.

If the sample function $X_t(\omega)$, $t \in \mathbb{T}$ is a continuous function of t for almost all ω, X_t is called a *continuous stochastic process*. If the sample function $X_t(\omega)$, $t \in \mathbb{T}$ is a right continuous function of t for almost all ω, X_t is called a *right continuous stochastic process*.

A stochastic process $\tilde{X}_t$, $t \in \mathbb{T}$ is called a *modification* of X_t, $t \in \mathbb{T}$ if $P(X_t = \tilde{X}_t) = 1$ holds for all t of $\mathbb{T}$. In most cases we do not distinguish between a stochastic process and its modification. However the properties of sample functions can depend on the choice of modification, so it is sometimes necessary to take a good modification of a given stochastic process. In Section 1.4 we shall give a criterion for a given stochastic process to have a modification of a continuous stochastic process.

Now let $X_t = (X_t^1, \ldots, X_t^d)$ be a continuous process with values in $\mathbb{R}^d$ having the mean vector $m(t) = E[X_t]$ and covariance matrix $V(s, t) = E[(X_s - m(s))(X_t - m(t))^t]$ where $(\ )^t$ stands for the transpose of the vector $(\ )$. It is called a *Brownian motion* if it has independent increments, i.e. for any $0 \le t_0 < t_1 < \cdots < t_n$ of $\mathbb{T}$, X_{t_0}, $X_{t_{i+1}} - X_{t_i}$: $i = 0, \ldots, n-1$ are independent random variables. Now if X_t is a Brownian motion, $X_t' = X_t - m(t)$ is also a Brownian motion. Further increments X_{t_0}', $X_{t_{i+1}}' - X_{t_i}'$: $i = 0, \ldots, n-1$ are orthogonal to each other, i.e.

$$E[X_{t_0}'(X_{t_{j+1}}' - X_{t_j}')^t] = 0, \qquad E[(X_{t_{i+1}}' - X_{t_i}')(X_{t_{j+1}}' - X_{t_j}')^t] = 0$$

holds for any $i \neq j$. (See Exercise 1.1.7.) Then the covariance $V(s, t)$ satisfies the following:

(i) $V(s, t) = V(r, r)$ where $r \equiv \min\{s, t\}$,
(ii) $V(t) \equiv V(t, t)$ increases with t.

A Brownian motion is called *standard* if $m(t) = 0$ and $V(t, t) = tE$ where E is the identity matrix.

Martingales

Let $\{\mathscr{F}_t : t \in \mathbb{T}\}$ be a family of sub σ-fields of $\mathscr{F}$. It is called a *filtration* of sub σ-fields of $\mathscr{F}$ if it satisfies the following three properties:

(i) $\mathscr{F}_s \subset \mathscr{F}_t$ if $s < t$,
(ii) $\bigcap_{\varepsilon>0} \mathscr{F}_{t+\varepsilon} = \mathscr{F}_t$,
(iii) each $\mathscr{F}_t$ contains all null sets of $\mathscr{F}$.

A stochastic process X_t, $t \in \mathbb{T}$ is called *$(\mathscr{F}_t)$-adapted* if for each t, X_t is $\mathscr{F}_t$-measurable.

Suppose that we are given a stochastic process X_t, $t \in \mathbb{T}$, first. Let $\mathscr{F}_t$ be the smallest σ-field including $\bigcap_{\varepsilon>0} \sigma(X_s : s \le t + \varepsilon)$ and all null sets of $\mathscr{F}$. Then $\{\mathscr{F}_t\}$ is a filtration and X_t is an $(\mathscr{F}_t)$-adapted process. It is called the *filtration generated by the process X_t*.

In the sequel we assume that a filtration $\{\mathscr{F}_t\}$ is given and is fixed unless otherwise mentioned. Let X_t be a real valued $(\mathscr{F}_t)$-adapted process such that for each t, X_t is integrable. It is called a *martingale* if it satisfies

$$E[X_t | \mathscr{F}_s] = X_s \qquad \text{a.s. for any } t > s. \tag{1}$$

It is called a *submartingale* if it satisfies

$$E[X_t|\mathscr{F}_s] \geq X_s \qquad \text{a.s. for any } t > s. \tag{2}$$

Further if the converse inequalities $E[X_t|\mathscr{F}_s] \leq X_s$ hold a.s., it is called a *supermartingale.*

Let X_t be a (sub)martingale and let $f: \mathbb{R} \to \mathbb{R}$ be an (increasing) convex function. If $f(X_t)$ is integrable for any t, it is a submartingale because of Jensen's inequality in Theorem 1.1.6. In particular $X_t^+ \equiv \max\{X_t, 0\}$ is a submartingale if X_t is a submartingale. Next let X_t be a martingale and let $p \geq 1$. If $E[|X_t|^p] < \infty$ holds for any t, it is called an L^p*-martingale.* In this case $|X_t|^p$ is a submartingale.

One of the most important examples of martingales is a Brownian motion. We will give Lévy's characterization of a Brownian motion through certain martingale properties.

Theorem 1.2.1 *Let $X_t = (X_t^1, \ldots, X_t^d)$, $t \in [0, T]$ be a continuous stochastic process with $X_0 = 0$ having the mean vector 0 and the covariance matrix $V(s, t)$. The following statements are equivalent.*

(i) *X_t is a Brownian motion.*
(ii) *Both X_t^i and $X_t^i X_t^j - V^{ij}(t, t)$ $i, j = 1, \ldots, d$ are martingales with respect to the filtration generated by X_t.*
(iii) *X_t is a Gaussian process, i.e. $(X_{t_0}, \ldots, X_{t_n})$ is subject to a Gaussian distribution for any $0 \leq t_0 < \cdots < t_n \leq T$. Further its covariance $V(s, t)$ coincides with $V(r, r)$, where $r \equiv \min\{s, t\}$.* □

The proof will be given in Section 2.3, see Theorem 2.3.13.

We next quote some theorems on martingales due to Doob, without giving proofs (Theorems 1.2.2, 1.2.3, 1.2.5–1.2.7 below). The details are found in Doob [26], Meyer [99] and other books dealing with the martingale theory.

Theorem 1.2.2 *Let X_t be a submartingale such that $E[X_t]$ is right continuous with respect to t. Then it has a modification $\tilde{X}_t$ such that its sample paths are right continuous with left hand limits a.s.* □

Theorem 1.2.3

(i) *Let X_t, $t \in [0, \infty)$ be a right continuous submartingale. Suppose $\sup_t E[X_t^+] < \infty$. Then $X_\infty = \lim_{t\uparrow\infty} X_t$ exists a.s. and X_∞ is integrable. Furthermore if $\{X_t^+ : t \in [0, \infty)\}$ is uniformly integrable, then X_t,*

$t \in [0, \infty]$ is a submartingale. In particular if X_t, $t \in [0, \infty)$ is a uniformly integrable martingale, then X_t, $t \in [0, \infty]$ is a martingale.

(ii) *Let X_t, $t \in (-\infty, 0]$ be a right continuous submartingale. Then $X_{-\infty} = \lim_{t \to -\infty} X_t$ exists a.s.* □

Stopping times play important roles in the theory of submartingales. Let $\mathbb{T}$ be $[0, \infty)$ or $[0, T]$. A random variable τ with values in $\overline{\mathbb{T}}$ (closure of $\mathbb{T}$ in $[0, \infty]$) is called a *stopping time* if $\{\tau \le t\} \in \mathscr{F}_t$ holds for any t. For a given stopping time τ, we set

$$\mathscr{F}_\tau = \{A \in \mathscr{F} : A \cap \{\tau \le t\} \in \mathscr{F}_t \text{ holds for all } t \in \mathbb{T}\}. \tag{3}$$

It is easily verified that $\mathscr{F}_\tau$ is a sub σ-field of $\mathscr{F}$.

Now let X_t be an $(\mathscr{F}_t)$-adapted process, right continuous with left hand limits. Then X_τ is $\mathscr{F}_\tau$-measurable. Indeed, if X_t is an $(\mathscr{F}_t)$-adapted *simple process*, i.e. there exists a finite partition $\{0 = t_0 < t_1 < \cdots < t_l\}$ of $\mathbb{T}$ such that $X_t = X_{t_k}$ holds for any $t \in [t_k, t_{k+1})$, then X_τ is written as $\sum_k X_{t_k} \chi(t_k \le \tau < t_{k+1})$ where $\chi(A)$ is the indicator function of the set A. It is clearly $\mathscr{F}_\tau$-measurable. Since any process which is right continuous with left hand limits is approximated uniformly by a sequence of the above simple processes a.s., X_τ is $\mathscr{F}_\tau$-measurable.

Let τ and σ be stopping times. Then the following properties hold.

(i) The sets $\{\tau < \sigma\}$, $\{\tau = \sigma\}$ and $\{\tau \le \sigma\}$ belong to both $\mathscr{F}_\tau$ and $\mathscr{F}_\sigma$.
(ii) If $\tau \le \sigma$, then $\mathscr{F}_\tau \subset \mathscr{F}_\sigma$.

Indeed, we have $\{\tau < \sigma\} \cap \{\tau \le t\} = \bigcup \{\tau < s, s < \sigma\}$, where the union is taken for all rationals s less than or equal to t. Then the set belongs to $\mathscr{F}_t$. This proves that the set $\{\tau < \sigma\}$ belongs to $\mathscr{F}_\tau$. The other assertions in (i) can be shown similarly. Next suppose $\tau \le \sigma$. If $B \in \mathscr{F}_\tau$, then $B \cap \{\sigma \le t\} = B \cap \{\tau \le t\} \cap \{\sigma \le t\} \in \mathscr{F}_t$. Therefore we have $B \in \mathscr{F}_\sigma$. This proves the second assertion.

We give an example of a stopping time, which will be used later in order to localize martingales.

Example 1.2.4 Let X_t be a right continuous $(\mathscr{F}_t)$-adapted process. Let G be an open subset of $\mathbb{R}$. The *hitting time* of X_t to the set G or the *first time such that $X_t \in G$* is defined by

$$\tau_G = \inf\{t \in \mathbb{T} : X_t \in G\} \qquad (= \sup\{t \in \mathbb{T}\} \text{ if } \{\cdots\} = \emptyset). \tag{4}$$

Then τ_G is a stopping time. In fact, we have $\{\tau_G \ge t\} = \bigcap (X_r \in G^c\}$, where the intersection is taken for all rationals r less than t. Therefore $\{\tau_G \geq t\} \in \mathscr{F}_t$ holds for any t.

Any stopping time can be approximated from the above by a decreasing sequence of stopping times with discrete values. Indeed, given a stopping time τ, we define the sequence τ_n, $n = 1, 2, \ldots$ by

$$\tau_n = \begin{cases} 0, & \text{if } \tau = 0, \\ \min\left\{\frac{k+1}{2^n}, \sup\{t \in \mathbb{T}\}\right\}, & \text{if } \frac{k}{2^n} < \tau \leq \frac{k+1}{2^n}, \quad k = 0, 1, 2, \ldots \end{cases}$$

We can easily verify that each τ_n is a stopping time and the sequence $\{\tau_n\}$ decreases to τ.

We will now quote Doob's *optional stopping time theorem* and its two consequences called *Doob's inequalities.* In the first theorem, we consider submartingales with finite time interval.

Theorem 1.2.5 *Let X_t, $t \in [0, T]$ be a continuous submartingale and let τ, σ be any two stopping times. Then X_τ is integrable and satisfies*

$$E[X_\tau | \mathscr{F}_\sigma] \geq X_{\min\{\tau,\sigma\}}. \tag{5}$$

In particular if X_t is a martingale, the equality holds in (5). □

Theorem 1.2.6 *Let X_t be a submartingale. Then*

$$cP\left(\sup_{s\leq t} X_s > c\right) \leq \int_{\sup_{s\leq t} X_s > c} X_t \, \mathrm{d}P \tag{6}$$

holds for any $c > 0$ and $t \in \mathbb{T}$. □

Theorem 1.2.7 *Let X_t be a positive submartingale. Then for any $p > 1$ we have*

$$E\left[\sup_{s\leq t} X_s^p\right] \leq q^p E[X_t^p] \qquad \textit{for all } t \in \mathbb{T}, \tag{7}$$

where q is a positive number such that $p^{-1} + q^{-1} = 1$. □

Markov processes

Let S be a locally compact, complete separable metric space and let $\mathscr{B}(S)$ be the set of all Borel subsets of S. By a *Borel measure* on S we mean a regular measure μ on $\mathscr{B}(S)$ such that $\mu(K) < \infty$ holds for any compact subset K of S. In particular if $\mu(S) = 1$ it is called a probability. A family of Borel measures $\{K(x, \cdot): x \in S\}$ on S is called a *kernel* if $K(x, E)$ is $\mathscr{B}(S)$-measurable with respect to x for each E of $\mathscr{B}(S)$. In the following a $\mathscr{B}(S)$-measurable function is called simply measurable. Let f be a real valued measurable function on S. We use the notation

$$Kf(x) = \int K(x, \mathrm{d}y)f(y) \tag{8}$$

if the integral is well defined for any x. The function $Kf(x)$ is measurable obviously.

Let $\{P_{s,t}(x, \cdot)\}$ be a family of kernels consisting of probability distributions on S, where $s < t$ are elements of $\mathbb{T}$. It is called a *transition probability* if it satisfies the Chapmann–Kolmogorov equation:

$$P_{s,u}(x, E) = \int_S P_{t,u}(y, E)P_{s,t}(x, \mathrm{d}y), \tag{9}$$

for every $s < t < u$, $x \in S$ and $E \in \mathscr{B}(S)$.

Suppose that a filtration $\{\mathscr{F}_t : t \in \mathbb{T}\}$ of sub σ-fields of $\mathscr{F}$ is given. Let X_t, $t \in \mathbb{T}$ be a stochastic process with state space S adapted to $(\mathscr{F}_t)$. The process X_t is called a *Markov process with transition probability* $\{P_{s,t}(x, \cdot)\}$ if it has the *Markov property* with respect to $\{\mathscr{F}_t\}$:

$$P(X_t \in E|\mathscr{F}_s) = P_{s,t}(X_s, E) \qquad \text{for every } s < t \text{ and } E \in \mathscr{B}(S). \tag{10}$$

A Markov process X_t is called *temporally homogeneous* if the transition probability $\{P_{s,t}(x, A)\}$ depends only on $t - s$.

In the following we will consider a temporally homogeneous Markov process with continuous time parameter $\mathbb{T} = [0, \infty)$. The transition probability $P_{0,t}(x, \cdot)$ is often denoted by $P_t(x, \cdot)$. For a bounded measurable function f on S, we denote $P_t f$ by $T_t f$, namely

$$T_t f(x) = \int_S P_t(x, \mathrm{d}y)f(y). \tag{11}$$

It is again a bounded measurable function of x. The family of operators $\{T_t : t \in \mathbb{T}\}$ satisfies the semigroup property $T_{t+s}f = T_t T_s f$ by the Chapmann–Kolmogorov equation.

Let $C(S)$ be the set of all real valued continuous functions on S. If S is a compact space, it is a separable Banach space with the supremum norm. If S is a noncompact space, denote by $C_\infty(S)$ the subset of $C(S)$ such that $\lim_{x\to\infty} f(x)$ exists and equals 0, where ∞ is the infinity adjoined to S as a one point compactification. Then $C_\infty(S)$ is also a Banach space with the supremum norm. Now suppose that S is compact (or noncompact). If T_t defined by (11) maps $C(S)$ (or $C_\infty(S)$) into itself and is strongly continuous i.e. for every t, $T_{t+h}f$ converges to $T_t f$ strongly as $h \to 0$ for any f, the semigroup $\{T_t\}$ of linear operators on $C(S)$ (or $C_\infty(S)$) is called a *Feller semigroup*.

Theorem 1.2.8 *Let X_t be a Markov process with a Feller semigroup $\{T_t\}$. Then X_t has a modification such that its sample functions are right continuous with left hand limits.*

Proof We prove the theorem in the case where the state space S is compact. The case for noncompact S is left to the reader (see Exercises 1.2.11 and 1.2.12). By the Markov property (10), we have for $s < t < u$

$$E[T_{u-t}f(X_t)|\mathscr{F}_s] = T_{t-s}T_{u-t}f(X_s) = T_{u-s}f(X_s) \quad \text{a.s.}$$

Then if $\alpha > 0$ and f is non-negative we obtain

$$E\left[\int_t^\infty e^{-\alpha u}T_{u-t}f(X_t)\,du \,\middle|\, \mathscr{F}_s\right] \le \int_s^\infty e^{-\alpha u}T_{u-s}f(X_s)\,du.$$

Setting

$$U_\alpha f(x) = \int_0^\infty e^{-\alpha u}T_u f(x)\,du, \tag{12}$$

the above inequality is written as $e^{-\alpha t}E[U_\alpha f(X_t)|\mathscr{F}_s] \le e^{-\alpha s}U_\alpha f(X_s)$. Therefore $e^{-\alpha t}U_\alpha f(X_t)$ is a bounded supermartingale. Then by Theorem 1.2.2, $e^{-\alpha t}U_\alpha f(X_t)$ has a modification such that its sample functions are right continuous with left hand limits. The latter property is valid for any f of $C(S)$.

Now let $\{f_n\}$ be a countable dense subset of $C(S)$. Then the set of functions $g_n \equiv U_\alpha f_n$, $n = 1, 2, \ldots$ where $\alpha > 0$ is fixed, is again a dense subset of $C(S)$. Indeed, in view of the resolvent equation $U_\alpha f - U_\beta f + (\alpha - \beta)U_\alpha U_\beta f = 0$ (*cf.* Lemma 1.3.1) the range of $C(S)$ by the map U_α is independent of α, and further it is a dense subset of $C(S)$ since $\alpha U_\alpha f$ converges to f strongly as α tends to infinity. Now let $g_n(X_t)^\sim$ be a modification of $g_n(X_t)$ such that its sample functions are right continuous and have left hand limits. Let $\tilde{\Omega}$ be the set of all samples ω such that $g_n(X_t)^\sim(\omega)$, $n = 1, 2, \ldots$ are all right continuous with left hand limits. Take any ω from $\tilde{\Omega}$. Then for every t there exists a unique point $\tilde{X}_t(\omega)$ in S such that $g_n(X_t)^\sim(\omega) = g_n(\tilde{X}_t(\omega))$ holds for any n. Thus $g_n(\tilde{X}_t(\omega))$ is right continuous with left hand limits with respect to t for all n. This implies that $\tilde{X}_t(\omega)$ itself is right continuous with left hand limits in the space S since $\{g_n\}$ is dense in $C(S)$. Since $P(\tilde{\Omega}) = 1$, we have $g_n(X_t) = g_n(\tilde{X}_t)$ for all n a.s. for each t. Therefore $\tilde{X}_t$ is a modification of X_t. $\square$

Suppose we are given a temporally homogeneous transition probability $\{P_t(x, \cdot)\}$ such that it defines a Feller semigroup. We shall construct a Markov process X_t associated with $\{P_t(x, \cdot)\}$ such that its sample paths are right continuous with left hand limits. Let $\overline{W}$ be the set of all maps $\bar{w}$ from $\mathbb{T}$ into S. We denote the elements of $\overline{W}$ by $\bar{w}$ and their values at $t \in \mathbb{T}$ by $\bar{w}(t)$ or $\bar{w}_t$. A subset A of $\overline{W}$ represented by

$$A = \{\bar{w} : (\bar{w}(t_1), \ldots, \bar{w}(t_n)) \in E_n\}, \tag{13}$$

where $0 \le t_1 < \cdots < t_n$ and E_n is a Borel set in S^n, is called a *cylinder set of* $\overline{W}$. Let $\mathscr{B}_{t_1,\ldots,t_n}$ be the collection of all cylinder sets represented above where $t_1, \ldots, t_n$ are fixed and E_n are running over all Borel sets in S^n. Then it is a σ-field of $\overline{W}$. For each x of S we define a probability measure $P_x^{(t_1,\ldots,t_n)}$ on $\mathscr{B}_{t_1,\ldots,t_n}$ by

$$P_x^{(t_1,\ldots,t_n)}(A)$$
$$= \int \cdots \int_{E_n} P_{t_1}(x, \mathrm{d}x_1) P_{t_2-t_1}(x_1, \mathrm{d}x_2) \ldots P_{t_n-t_{n-1}}(x_{n-1}, \mathrm{d}x_n), \tag{14}$$

where A is an element of $\mathscr{B}_{t_1,\ldots,t_n}$ defined by (13). Then the family of measures $\{P_x^{(t_1,\ldots,t_n)}\}$ is consistent for each x, i.e. if $(t'_1, \ldots, t'_m)$ is a subset of $(t_1, \ldots, t_n)$ and A is an element of $\mathscr{B}_{t_1,\ldots,t_n}$ belonging to $\mathscr{B}_{t'_1,\ldots,t'_m}$, then $P_x^{(t_1,\ldots,t_n)}(A) = P_x^{(t'_1,\ldots,t'_m)}(A)$ holds. Now let $\mathscr{A}(\overline{W})$ be the algebra $\bigcup \mathscr{B}_{t_1,\ldots,t_n}$ where the union runs over all $t_1, \ldots, t_n$ of $\mathbb{T}$ and $n = 1, 2, \ldots$ Then there exists a unique measure $\hat{P}_x$ on the algebra $\mathscr{A}(\overline{W})$ such that its restriction to $\mathscr{B}_{t_1,\ldots,t_n}$ coincides with $P_x^{(t_1,\ldots,t_n)}$. Let $\mathscr{B}(\overline{W})$ be the smallest σ-field containing $\mathscr{A}(\overline{W})$. Then the measure $\hat{P}_x$ can be extended uniquely to a measure $\bar{P}_x$ on $\mathscr{B}(\overline{W})$ by Kolmogorov–Hopf's theorem. Denote by $\mathscr{F}(\overline{W})$ the completion of $\mathscr{B}(\overline{W})$ with respect to $\bar{P}_x$ (x being fixed).

Let $\{\mathscr{F}_t\}$ be the filtration generated by the stochastic process $\bar{w}(t)$. We show that $\bar{w}(t)$ has the Markov property with respect to $\{\mathscr{F}_t\}$ for each measure $\bar{P}_x$. Let $s < t$ and A be an element of $\mathscr{A}(\overline{W})$ represented as in (13) where $t_n \le s$. Then we have by (14)

$$\bar{P}_x(A \cap \{\bar{w}(t) \in E\}) = \int \cdots \int_{E_n \times S} P_{t_1}(x, \mathrm{d}x_1) \ldots P_{s-t_n}(x_n, \mathrm{d}y) P_{t-s}(y, E).$$

Denote by $\bar{E}_x$ the expectation with respect to the measure $\bar{P}_x$. Then $\bar{E}_x[P_{t-s}(\bar{w}(s), E) : A]$ is also written as the right hand side of the above. This implies

$$\bar{E}_x[f(\bar{w}(t)) : A] = \bar{E}_x[T_{t-s} f(\bar{w}(s)) : A] \tag{15}$$

for any f of $C_\infty(S)$ (or of $C(S)$ if S is compact) and A of $\sigma(\bar{w}(r) : r \le s)$. Further $f(\bar{w}(t))$ is right continuous in L^2. In fact we have by (15)

$$\begin{aligned}&\bar{E}_x[(f(\bar{w}(s+\varepsilon)) - f(\bar{w}(s)))^2]\\ &\quad = \bar{E}_x[T_\varepsilon(f^2)(\bar{w}(s)) - 2T_\varepsilon f(\bar{w}(s))f(\bar{w}(s)) + f(\bar{w}(s))^2]\end{aligned}$$

which converges to 0 as $\varepsilon \to 0$. Then the above equality (15) is valid for any A of $\mathscr{F}_s$. Therefore $\bar{w}(t)$ is a Markov process with the Feller semigroup $\{T_t\}$. Then $\bar{w}(t)$ has a modification $X_t(\bar{w})$ which is right continuous with left hand limits by the previous theorem.

Now let W be the set of all $w \in \overline{W}$ such that $w(t)$ is right continuous with left hand limits. Then for almost all $\bar{w}$, the right continuous modification $X(\bar{w}) \equiv \{X_t(\bar{w}) : t \in \mathbb{T}\}$ can be regarded as an element of W. Let $\mathscr{B}(W)$ be the smallest σ-field of W containing all cylinder sets of W. Then the set $\{\bar{w} : X(\bar{w}) \in B\}$ belongs to $\mathscr{F}(\overline{W})$ for any B of $\mathscr{B}(W)$. We define the law of $(X_t, \bar{P}_x)$ on the space $(W, \mathscr{B}(W))$ by

$$P_x(B) = \bar{P}_x(\{\bar{w} : X(\bar{w}) \in B\}). \tag{16}$$

The expectation by the measure P_x is denoted by E_x. Note that $P_x(B)$ is measurable with respect to x for any B of $\mathscr{B}(W)$. The triple $(W, \mathscr{B}(W), P_x : x \in S)$ is called a *right continuous Markov process with the Feller semigroup* $\{T_t\}$ or simply a *Feller process.*

Now for $s \in [0, \infty)$, let θ_s be a map from W into itself such that $(\theta_s w)_t = w(s+t)$ holds for all t. It is a measurable map from $(W, \mathscr{B}(W))$ into itself. The family $\{\theta_s\}$ satisfies the semigroup property $\theta_s\theta_t = \theta_{s+t}$.

We shall extend the Markov property of the Feller process.

Theorem 1.2.9 *Let $(W, \mathscr{B}(W), P_x : x \in S)$ be a Feller process. Then each P_x satisfies*

$$P_x(\theta_s^{-1}B|\mathscr{F}_s) = P_{w(s)}(B) \qquad \textit{for every } B \in \mathscr{B}(W), \tag{17}$$

where $\{\mathscr{F}_t\}$ is the filtration generated by $w(t)$.

Proof Let A and B be the cylinder sets $\{w : (w(t_1), \ldots, w(t_n)) \in E_n\}$ and $\{w : (w(u_1), \ldots, w(u_m)) \in E_m\}$ respectively, where $t_n \le s$ and E_n and E_m are Borel sets of S^n and S^m respectively. Using the Markov property, we can show that both of $P_x(\theta_s^{-1}B \cap A)$ and $E_x[P_{w(s)}(B) : A]$ are represented by the same quantity below:

$$\int\cdots\int_{E_n \times E_m} P_{t_1}(x, dx_1)\cdots P_{u_1+s-t_n}(x_n, dy_1)\cdots P_{u_m-u_{m-1}}(y_{m-1}, dy_m).$$

Therefore we have $P_x(\theta_s^{-1}B \cap A) = E_x[P_{w(s)}(B) : A]$ for the above cylinder sets A and B. The equality can be extended to any B of $\mathscr{B}(W)$ and A of $\mathscr{F}_s$. Thus the theorem is established. □

Exercise 1.2.10 Suppose that a filtration $\{\mathscr{F}_t : t \in [0, \infty)\}$ (or $\{\mathscr{F}_t : t \in (-\infty, 0]\}$) of sub σ-fields of $\mathscr{F}$ is given. Let X be a real random variable such that $E[|X|^p] < \infty$ for some $p > 1$. Show that $X_t = E[X|\mathscr{F}_t]$ is a martingale. Let $\tilde{X}_t$ be a right continuous modification of X_t. Show that $X_\infty = \lim_{t\to\infty} \tilde{X}_t$ (or $X_{-\infty} = \lim_{t\to-\infty} \tilde{X}_t$ respectively) exists and equals $E[X|\mathscr{F}_\infty]$ (or $E[X|\mathscr{F}_{-\infty}]$ respectively) where $\mathscr{F}_\infty$ is the least σ-field including $\bigcup_t \mathscr{F}_t$ (or $\mathscr{F}_{-\infty} = \bigcap_t \mathscr{F}_t$, respectively). (*Hint*: Show that $\{\tilde{X}_t : t \in [0, \infty)\}$ is uniformly integrable and then show $E[X_\infty : A] = E[X : A]$ holds for any A of $\mathscr{F}_\infty$.)

Exercise 1.2.11 (*Strong Markov property*) Let X_t be a right continuous Markov process associated with the Feller semigroup $\{T_t\}$. Let $\{\mathscr{F}_t\}$ be the filtration generated by X_t. Let τ be a stopping time and $\mathscr{F}_\tau$ be the σ-field defined by (3). Show that

$$E[f(X_{t+\tau}) : A] = E[T_t f(X_\tau) : A] \qquad \text{for every } A \in \mathscr{F}_\tau$$

holds for any $f \in C(S)$ (or $C_\infty(S)$ if S is noncompact). (*Hint*: show the above first in the case where τ is a stopping time with discrete values. Then for general τ, approximate it from the above by a decreasing sequence of stopping times with discrete values.)

Exercise 1.2.12 (*Proof of Theorem* 1.2.8 *for noncompact case*) Suppose that S is noncompact in Theorem 1.2.8. Let $\hat{S} = S \cup \{\infty\}$ be the one point compactification of S.

(i) Show that X_t has a modification $\tilde{X}_t$ with values in $\hat{S}$ such that its sample paths are right continuous with left hand limits with respect to the topology of $\hat{S}$.

(ii) Let $\{G_n\}$ be a decreasing sequence of open neighborhoods of ∞ such that $\bigcap_n G_n = \{\infty\}$. Let σ_n be the first time that $\tilde{X}_t$ hits the set G_n and let $\sigma_\infty = \lim_n \sigma_n$. Let $\alpha > 0$. Show

$$E\left[\int_0^\infty e^{-\alpha t} f(\tilde{X}_t)\, dt\right] = E\left[\int_0^{\sigma_\infty} e^{-\alpha u} f(\tilde{X}_u)\, du\right] + E[e^{-\alpha\sigma_\infty} U_\alpha f(\tilde{X}_{\sigma_\infty -})],$$

where $\tilde{X}_{\sigma_\infty -} = \lim_n \tilde{X}_{\sigma_n}$ and $U_\alpha f$ is defined by (12). Deduce from this that $\sigma_\infty = \infty$ a.s. (*Hint*: use the strong Markov property of $\tilde{X}_t$).

1.3 Ergodic properties of Markov processes

Recurrent and transient processes

Let $(W, \mathscr{B}(W), P_x : x \in S)$ be a Feller process constructed in the previous section associated with the Feller semigroup $\{T_t\}$, where S is a locally compact complete separable metric space. We assume that the transition probability has a positive, continuous density function.

Condition (A) *There exists a Borel measure μ on S supported by S, and a strictly positive function $p_t(x, y)$ continuous in $(t, x, y) \in (0, \infty) \times S^2$ such that the transition probability $P_t(x, \mathrm{d}y)$ equals $p_t(x, y)\mu(\mathrm{d}y)$.* □

Let $B(S)$ be the set of all bounded measurable functions on S and let $BC(S)$ be the set of all bounded continuous functions on S. Each T_t maps $B(S)$ into $BC(S)$. Indeed if f of $B(S)$ satisfies $0 \le f \le 1$, both of $T_t f$ and $T_t(1 - f)$ are lower semicontinuous and their sum is constant 1. Then $T_t f$ has to be continuous. Obviously the last property is valid for any f of $B(S)$. The semigroup with this property is called a *strong Feller semigroup.*

The Feller process is called *recurrent in the sense of Harris* if

$$\int_0^\infty \chi(A)(w(t))\,\mathrm{d}t = \infty \qquad \text{a.s. } P_x \tag{1}$$

is satisfied for every $x \in S$ whenever $\mu(A) > 0$. Further, the Feller process is called *transient* if

$$\sup_{x \in S} E_x\left[\int_0^\infty \chi(K)(w(t))\,\mathrm{d}t\right] < \infty \tag{2}$$

holds for any compact subset K of S.

We will show that any Feller process satisfying Condition (A) is either transient or recurrent in the sense of Harris. Our discussion of this problem is similar to Revuz [112]. As an intermediate step we introduce a suitable Markov process with discrete time parameter and show first a similar transient–recurrent dichotomy for this process (Lemma 1.3.3 below). Then it will be applied to our Feller process.

Let h be a bounded non-negative measurable function on S. Set

$$U_h f(x) = E_x\left[\int_0^\infty \exp\left\{-\int_0^t h(w(s))\,\mathrm{d}s\right\} f(w(t))\,\mathrm{d}t\right]. \tag{3}$$

It is well defined; at least f is a non-negative measurable function. Setting $U_h(x, E) \equiv U_h\chi(E)(x)$, it defines a kernel if $U_h(x, K) < \infty$ holds for any compact K.

Lemma 1.3.1

(i) *The following relation holds:*

$$U_h h(x) = 1 - E_x\left[\exp\left\{-\int_0^\infty h(w(s))\,\mathrm{d}s\right\}\right]. \tag{4}$$

(ii) *Let $h \geq k \geq 0$ and $f \geq 0$. Then U_h and U_k satisfy the resolvent equation:*

$$U_k f - U_h f = U_h(h-k)U_k f = U_k(h-k)U_h f. \tag{5}$$

(iii) *If $U_h f$ is a bounded function, it is continuous.*

Proof Note the equality

$$\int_0^\infty \exp\left\{-\int_0^t h(w(s))\,\mathrm{d}s\right\} h(w(t))\,\mathrm{d}t = 1 - \exp\left\{-\int_0^\infty h(w(s))\,\mathrm{d}s\right\}.$$

Then equality (4) is immediate. We have by Theorem 1.2.9 (Markov property),

$$U_k f(w_t) = E_x\left[\int_t^\infty \exp\left\{-\int_t^v k(w_u)\,\mathrm{d}u\right\} f(w_v)\,\mathrm{d}v \,\middle|\, \mathscr{F}_t\right].$$

Therefore,

$$\begin{aligned}
&U_h(h-k)U_k f(x)\\
&= E_x\left[\int_0^\infty \exp\left\{-\int_0^t h(w_s)\,\mathrm{d}s\right\}(h(w_t)-k(w_t))\right.\\
&\qquad \left.\times\left(\int_t^\infty \exp\left\{-\int_t^v k(w_u)\,\mathrm{d}u\right\} f(w_v)\,\mathrm{d}v\right)\mathrm{d}t\right]\\
&= E_x\left[\int_0^\infty f(w_v)\left\{\int_0^v \exp\left\{-\int_0^t (h(w_s)-k(w_s))\,\mathrm{d}s\right\}\right.\right.\\
&\qquad \left.\left.\times\,(h(w_t)-k(w_t))\,\mathrm{d}t\right\}\exp\left\{-\int_0^v k(w_u)\,\mathrm{d}u\right\}\mathrm{d}v\right]\\
&= E_x\left[\int_0^\infty f(w_v)\left[1-\exp\left\{-\int_0^v (h(w_s)-k(w_s))\,\mathrm{d}s\right\}\right]\right.\\
&\qquad \left.\times\exp\left\{-\int_0^v k(w_u)\,\mathrm{d}u\right\}\mathrm{d}v\right]\\
&= U_k f(x) - U_h f(x).
\end{aligned} \tag{6}$$

This proves the resolvent equation (5).