Abstract

Title of Document: ON STOCHASTIC DIFFERENTIAL EQUATIONS IN THE ITO AND IN THE STRATONOVICH SENSE

Brett Williams, Master of Arts, 2012

Directed By: Professor Mark Freidlin, Math

In this paper I looked into some modifications of the standard diffusion equation. First I added "look back" in the differential equation and proved that the solution of the new equation converged to the solution of the diffusion equation in the Ito sense. Then I proved that if we use an approximation to the Weiner process as well as "look back" our solution will depend on the order in which we take the limits. Specifically if we first let the look back go to zero then let our approximation to the Weiner process converge to the Weiner process we will converge to the diffusion equation understood in the Stratonovich sense and if we first let our approximation converge to the Weiner process then let our "look back" go to zero we will converge to the Ito integral.

ON STOCHASTIC DIFFERENTIAL EQUATIONS IN THE ITO AND IN THE STRATONOVICH SENSE

by

Brett T. Williams

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Masters of Arts
2012

Advisory Committee:
Professor Mark Freidlin, Chair
Professor Sandra Cerrai
Professor Leonid Koralov

Contents

1 Background

1.1 Defining our process

A Weiner process is a Gaussian random process which has the following properties:

1. $E(W_t) = 0, \;\; t \geq 0$
2. $E(W_t W_s) = min(s, t)$
3. $W_t(\omega)$ is continuous in t a.s.

That such a process exists can be proven (cf, for example, Koralov and Sinai 2007). The Weiner process has several important properties. Of particular interest are that the Weiner process is a.s. nowhere differentiable, has infinite total variation, and the increments of the Weiner process are independent of each other. In addition it can be shown that $E(W_t - W_s)^2 = t - s$ (cf for example Karatzas and Shreve 1991).

So let W_t be a real Weiner process. It can be shown that the Weiner process is a suitable "white noise process" to add to a differential equation (see Oksendal 2007 pp. 21-22). Let us then consider the two differential equations:

$$\dot{x}_t = b(x_t) + \sigma(x_t)\dot{W}_t, \;\; x_0 = x_0 \tag{1}$$

$$\dot{x}_t^S = b(x_t^S) + \sigma(x_t^S) \circ \dot{W}_t, \;\; x_0^S = x_0 \tag{2}$$

Where equation 1 is understood in the Ito sense and equation 2 is understood in the Stratonovich sense. These equations find significance in a variety of applications. However, the solution to these equations, x_t and x_t^S, are not smooth. They are in fact nowhere differentiable.

Because the Weiner process is a.s. nowhere differentiable, $\dot{W}_t$ is non-existent and we are left to consider weak solutions to the above equations. The exact nature of $\int_0^t \sigma(x_s) dW_s$ is unclear, however, due to the Weiner process being of unbounded (infinite) total variation.

When considering how the stochastic integral should be defined we would want to define the integral for simple functions and then extend the definition to some class of measurable functions. This approach, however, leads to problems since it can be shown that when trying to approximate $\int_s^t W_s dW_s$ in this manner we can find two simple approximations $\phi_1(t, \omega, n)$ and $\phi_2(t, \omega, n)$ such that as n goes to infinity both functions converge to the Weiner process but $E[\int_0^t \phi_1(s, \omega) dW_s(\omega)] = 0$ for all n and $E[\int_0^t \phi_2(s, \omega) dW_s(\omega)] = t$ for all n (See Oksendal 2007 p.23). We therefore need to limit the class of functions we can use as approximating functions for the integral.

In 1944 Kiyoshi Ito suggested we use the value at the left hand end point of the interval to make this approximation. This suggestion has led to what is now known as the Ito integral. In 1963 Donald Fisk and in 1966 Ruslan Stratonovich independently suggested a different definition that used the value at the mid-point of the interval as an approximating value. This suggestion has led to the definition of what is now known as the Stratonovich integral.

1.2 The Ito Integral

To solve the problem of different approximations leading to different integrals we need to restrict the class of simple functions in some way. Ito suggested we look at functions $f(t,\omega)$ which are $\mathcal{F}_t$ adapted. This would restrict us to using the value of our function at the left end point as an approximation. This seems like a reasonable criterion since it leads to a property of not "looking ahead." For a full description of Ito's definition see Oksendal 2007 pp.25-30. Essentially once we restrict ourselves to $\mathcal{F}_t$ adapted functions we can define the integral naturally for a simple function $\phi(t,\omega) = \sum_j e_j(\omega)\chi_{[t_j,t_{j+1}]}$ as

$$\int_r^t \phi(s,\omega)dW_s = \sum_j e_j(\omega)[W_{t_{j+1}} - W_{t_j}](\omega) \tag{3}$$

and expand the definition by using simple approximations to more complex functions.

The Ito integral has many useful properties (see Oksendal 2007):

1. $E[(\int_r^t f(s,\omega)dW_s)^2] = E[\int_r^t f^2(s,\omega)ds]$ (The Ito Isometry)
2. $\int_r^t fdW_s = \int_r^u fdW_s + \int_u^t fdW_s$ (Separability)
3. $\int_r^t (cf+g)dW_s = c\int_r^t fdW_s + \int_r^t gdW_s$ (Linearity)
4. $E(\int_r^t fdW_s) = 0$
5. $\int_r^t fdW_s$ is $\mathcal{F}_t$ measurable
6. The Ito integral is a Martingale w.r.t. $\mathcal{F}_t$

However, change of variables is difficult under the Ito integral. For a one dimensional function we get the following theorem:

Theorem 1 (The One Dimensional Ito Formula) *Let x_t be defined as above. Let $g(t,x) \in C_2([0,\infty X\mathbf{R})$. Then $Y_t = g(t,X_t)$ is an Ito process and*

$$dY_t = \frac{\partial g}{\partial t}(t,X_t)dt + \frac{\partial g}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)\cdot(dX_t)^2$$

And for multidimensional functions

Theorem 2 (The Multidimensional Ito Formula) *Let X_t be an n dimensional Ito process. Let $g(t,x) = (g_1(t,x), ..., g_p(t,x))$ be a C_2 map from $([0,\infty)X\mathbf{R}^n)$ to $\mathbf{R}^p$. Then $Y_t = g(t, X_t)$ is an Ito process and Y_k is given by*

$$dY_{t,k} = \frac{\partial g_k}{\partial t}(t, X_t)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t, X_t)dX_{i,t} + \frac{1}{2}\sum_{i,j}\frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t)dX_{i,t}dX_{j,t}$$

For a proof of these theorems see Oksendal 2007 p 46.

1.3 The Stratonvich Integral

Because of its "no look ahead" property, the Ito integral's approximations have to use the left end points of intervals for the approximating functions. In 1966 Stratonovich proposed a different approach. For random processes X_t and Y_t we define

$$\int_s^t Y_s dX_s = l.i.p_{\Delta\to 0}\sum_{j=1}^{N-1}\frac{Y_{t_j} + Y_{t_{j+1}}}{2}, t_j)[X_{t_{j+1}} - X_{t_j}]$$

(See Stratonovich 1966, Ikeda and Watanabe 1989). Effectively this changes from using the left end point of the interval to using the midpoint of the interval in the approximation. Despite this seemingly minor change, the properties of the integrals vary significantly.

Although the two integrals disagree, the Stratonovich integral can be written as a function of the Ito integral. While it can be shown using integration by parts that for smooth functions the Ito and Stratonovich integrals agree, if the function is non-differentiable, like x_t, then the Ito integral and Stratonovich integrals are not the same. Specifically, in one dimension, x_t^S can be related to the Ito integral by the equation:

$$x_t^S - x_0 = \int_0^t b(x_s^S)ds + \int_0^t \sigma(x_s^S)dW_s + \frac{1}{2}\int_0^t \dot{\sigma}(x_s^S)\sigma(x_s^S)ds \quad (4)$$

where the stochastic integral above is interpreted in the Ito sense (see Stratonovich 1966). More generally, we can define a new operation "symetric Ω multiplication:"

$$Y \circ dX = YdX + \frac{1}{2}dXdY \quad (5)$$

And show that this definition leads to an integral defined as above(see Ikeda and Watanabe 1989). This would mean that if x_t^S is defined as above:

$$\int_0^t x_t^S \circ dW_t = \int_0^t x_s^S dW_s + \frac{1}{2} < x_s^S, W_s >_t$$
$$= \int_0^t x_t^S dW_s + \frac{1}{4} \int_0^t \frac{\partial \sigma (x_t^S)^2}{\partial x} \cdot dt$$

See Ikeda and Watanabe, 1989 and Karatzas and Shreve, 1991 for more details.

It can be proven (cf example Ikeda and Watanabe 1989) that the Stratonovich integral is separable and linear. However we lose the property that the integral is a Martingale (Oksendal 2007) and we also lose the fact that the average of the integral is 0 (specifically from the above relation we get $E(\int_0^t f \circ dW_s) = \frac{1}{4}E[\int_0^t \frac{\partial f^2}{\partial x} ds]$ which is in general not zero).

We do, however, get a very nice property when doing change of variables. Unlike the Ito formula, the Stratonovich analogue does not have any second order terms, which makes it so that the chain rule functions similarly to what would be expected for the normal integral. Specifically we get

Theorem 3 (The Multidimensional Ito Formula for Stratonovich Integrals) *Let X_t be an n dimensional Stratonvich process. Let $g(t,x) \in C_2([0,\infty)X\mathbf{R}^n)$. Then $Y_t = g(t, X_t)$ is an Stratonovich process and Y is given by*

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \sum_i \frac{\partial g}{\partial x_i}(t, X_t) \circ dX_{i,t}$$

(see Ikeda and Watanabe 1989). Because of these different properties, both definitions of the integral are desirable in different situations.

1.4 Choosing which interpretation to use

Which interpretation is appropriate depends on how we choose our approximating process. In "On the Convergence of Ordinary Integrals to Stocastic Integrals," Eugene Wong and Moshe Zakai proved that if we approximate the Weiner process with a process, y_n, which has the following proerties:

1. y_n is continuous and of bounded variation a.s.
2. y_n converges a.s. to W_t as $n \to \infty$
3. For almost all ω there exists $n_0(\omega)$ and $k(\omega)$, both finite such that for all $n > n_0$ and all t in [a,b], $y_n(t, \omega) \leq k(\omega)$
4. y_n has a piecewise continuous derivative

And if b and σ are both Lipshitz continuous, then the approximation

$$\dot{x}_t^n = b(x_t^n) + \sigma(x_t^n)\dot{y}_{n,t} \tag{6}$$

converges as $n \to \infty$ to x_t^S a.s. This would lead us to believe that the Stratonovich integral is the more natural choice in certain situations. However the fact that the Ito integral is a Martingale w.r.t. $\mathcal{F}_t$ leads this choice to be more desirable in many situations.

When we use an approximation which uses multiple parameters, we sometimes get different behaviours depending on the order in which we take the limits. For instance, if we consider the system:

$$\mu\dot{p}_t^{\mu,\delta} = b(q_t^{\mu,\delta}) - p_t^{\mu,\delta} + \sigma(q_t^{\mu,\delta})\frac{dV_t^\delta}{dt}, \dot{q}_t^{\mu,\delta} = p_t^{\mu,\delta}, q_0^{\mu,\delta} = q, p_0^{\mu,\delta} = p \tag{7}$$

where V_t^δ is an approximation to W_t and look at the behaviour as μ and δ approach zero, then we will see different behaviours depending on how μ and δ are related. M. Freidlin proved that if $\mu, \delta \downarrow 0$ and $lim(\mu e^{1/\delta}) = 0$, then $q_t^{\mu,\delta}$ converges to x_t and, otherwise, the solution converges to x_t^S. More specifically, if we let δ go to zero and then let μ go to zero, we will understand the integral in the Ito sense, but if we let μ go to zero first and then let δ go to zero, we will understand the integral in the Stratonovich sense. (Freidlin 2004)

In this paper I will consider a different method of smoothing out the process, namely using a mollifier in the differential equation to smooth the solution out. I will first show that if we only use the mollifier, then the solution should be understood in the Ito sense. I will then show that if we use both an approximation to the Weiner process as Wong and Zakai did and a mollifier, then the solution will converge to the Stratonvich integral provided we take the limit of the mollifier first or take both limits at the same time. However, if we first let the process converge to the W_t then let the mollifier converge to zero, the random process will converge as though we were looking at the stochastic integral as the Ito integral.

2 Differential Equations with Mollifiers

First I will consider just the case where we apply a mollifier to our process to smooth out the process. So let x_t^δ be the solution to the equation

$$\dot{x_t^\delta} = b(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha) + \sigma(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha)\dot{W}_s \tag{8}$$

and x_t be the solution to the equation

$$\dot{x_t} = b(x_t) + \sigma(x_t)\dot{W}_s \tag{9}$$

where $\sigma\dot{W}_s$ is understood in the Ito sense. If both b and σ are Lipsitz and h is a mollifier with support between 0 and δ then we have the following theorem:

Theorem 4 *For each T¿0, under the above assumptions x_t^δ converges uniformly on [0,T] in probability as $\delta \downarrow 0$ to x_t*

Proof: To prove this theorem we will need the following lemma:

Lemma 1 *If b and σ are both Lipshitz then $E|x_t^\delta - x_s^\delta|^2 \le K|t-s|$*

Proof of Lemma: Using the boundedness of b and σ and the fact that $s < t < T$ so $(t-s) < T$ we get that:

$$\begin{aligned}
E|x_t^\delta - x_s^\delta|^2 \le 2E|\int_s^t b(\int_0^t h(r-\alpha)x_\alpha^\delta d\alpha)dr)|^2 \\
+ 2E|\int_s^t \sigma(\int_0^t h(r-\alpha)x_\alpha^\delta d\alpha)dWr)|^2 \\
&\le 2(K_1^2(t-s)^2) + 2(\int_s^t \sigma(\int_0^t h(r-\alpha)x_\alpha^\delta d\alpha)^2 dr) \\
&\le 2(K_1^2(t-s)^2) + 2(K_2|t-s|) \\
&\le K|t-s|
\end{aligned}$$

Which proves our lemma.

To show convergence it will suffice to show $\lim_{\delta\downarrow 0} E(X_t^\delta - X_t)^2 = 0$. But using the properties of the Ito integral, the fact that b and σ are Lipshitz, and lemma 1 we get:

$$
\begin{aligned}
E(X_t^\delta - X_t)^2 &\leq 2E(\int_0^t b(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha) - b(x_s)ds)^2 \\
&\quad + 2E(\int_0^t \sigma(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha) - \sigma(x_s)dWs)^2 \\
&\leq 2E(\int_0^t K_3^2 |\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha - x_s|^2 ds) \\
&\quad + 2E\int_0^t K_4^2 |\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha - x_s|^2 ds \\
&\leq 2E(\int_0^t (K_3^2 + K_4^2)|\int_0^t h(s-\alpha)(x_s^\delta - x_s + x_\alpha^\delta - x_s^\delta)d\alpha|^2 ds) \\
&\leq 2K_5 E[\int_0^t 2(\int_0^t h(s-\alpha)(x_s^\delta - x_s)d\alpha)^2 \\
&\quad + 2(\int_0^t h(s-\alpha)(x_\alpha^\delta - x_s^\delta)d\alpha)^2 ds] \\
&\leq 2K_5[\int_0^t E2(\int_0^t h(s-\alpha)(x_s^\delta - x_s)d\alpha)^2 \\
&\quad + 2E(\int_0^t h(s-\alpha)(\max_{s-\delta \leq r \leq s}(x_r^\delta) - x_s^\delta)d\alpha)^2 ds] \\
&\leq 2K_5[\int_0^t 2E(x_s^\delta - x_s)^2 + 2(\int_0^t (K^2\delta)ds] \\
&\leq 4K_5 \int_0^t E(x_s^\delta - x_s)^2 + 4K_5 T(K^2\delta)
\end{aligned}
$$

At this point we use Gronwall's inequality to get $E(X_t^\delta - X_t)^2 \leq 4K_5 T(K^2\delta)e^{4TK_5}$ which clearly goes to zero as $\delta \downarrow 0$.

3 Approximations with mollifiers and a smooth process

If we both use a smooth approximation to the Weiner process and add a mollifier to our process as we did before, then we will get behaviour similar to what M. Freidlin found for $q_t^{\mu,\delta}$. Specifically, we will find that what $x_t^{\mu,\delta}$ converges to will depend on the order in which we take the limits.

Let y_μ be a continuous random process which has the following properties:

1. y_μ is continuous and of bounded variation a.s.
2. y_μ converges a.s. to W_t as $\mu \to 0$
3. For almost all ω there exists $\mu_0(\omega)$ and $k(\omega)$, both finite such that for all $\mu < \mu_0$ and all t in [a,b], $y_\mu(t,\omega) \leq k(\omega)$
4. y_μ has a piecewise continuous derivative

then the following equation will asymptotically approximate either x_t or x_t^S

$$\dot{x}_t^{\delta,\mu} = b(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha) + \sigma(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha)\dot{y}_\mu(s), x_0^{\delta,\mu} = x_0 \tag{10}$$

More specifically if we first take δ to be any function of μ which goes to zero as μ goes to zero, then we get that $x_t^{\delta,\mu} \to x_t^S$ from the following theorem:

Theorem 5 *Let $x_t^{\delta,\mu}$ be defined as above. Let x_t^S be the solution of the Stratonovich equation:*

$$\dot{x_t^S} = b(x_t^S) + \sigma(x_t^S)\dot{W}_s + \frac{1}{2}\sigma(x_t^S)\frac{\partial\sigma(x_t^S)}{\partial x} \tag{11}$$

Then if δ is a function of μ such that when μ is zero then so is δ
and b(x), σ(x) and $\partial\sigma^2(x)/\partial x$ are Lipshitz continuous
and $\partial\sigma(x)/\partial x$ is continuous
and further $\sigma(x) \geq \beta > 0$ (or -$\sigma(x) \geq \beta > 0$)
then $x_t^{\delta,\mu}$ converges to x_t^S almost surely.

Proof: I will prove this similarly to Wong and Zakai (1966) Theorem 2 with some minor modifications. I will need the following lemma which Wong and Zakai proved:

Lemma 2 *Let f(t) be real, non-negative and continuous in $-\infty < a \leq t \leq b < \infty$. Let $0 < \mu < \infty$, $\rho > 0$ and let $\epsilon(t) \geq 0$ and $\int_a^b \epsilon(t)ds <$*

$(\rho\mu e^{\mu\rho(b-a)})^{-1}$ *Suppose that*

$$log(1+f(t)/\mu) \leq log(1+\epsilon(t)) + \rho \int_a^b f(s)ds$$

Then

$$f(t) \leq \mu[\epsilon(t) + \rho\mu e^{\rho\mu(a-b)} \int_a^b \epsilon(t)dt]/[1 - \rho\mu e^{\rho\mu(a-b)} \int_a^b \epsilon(t)dt]$$

Let $\Phi(x) = \int_0^x \sigma(u)du$, then using lemma 1:

$$\begin{aligned}
\Phi(x_t^{\delta,\mu}) &= \int_a^t \frac{\partial\Phi(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha)}{\partial s}ds + \int_a^t \frac{d\Phi(x_s^{\delta,\mu})}{\partial x_s^{\delta,\mu}}dx_s^{\delta,\mu} + \Phi(x_a^{\delta,\mu}) \\
&= \int_a^t \frac{\partial\Phi(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha)}{\partial s}ds + \int_a^t \frac{b(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha)}{\sigma(x_s^{\delta,\mu})}ds + \\
&\quad + \int_a^t \frac{\sigma(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha)}{\sigma(x_s^{\delta,\mu})}dy_\mu + \Phi(x_a^{\delta,\mu}) \\
&\leq \int_a^t \frac{\partial\Phi(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha)}{\partial s}ds + \int_a^t \frac{b(x_s^{\delta,\mu}) + K_1\delta}{\sigma(x_s^{\delta,\mu})}ds + \\
&\quad + (1 + sign(y_\mu(t) - y_\mu(a))K_2\delta)(y_\mu(t) - y_\mu(a)) + \Phi(x_a^{\delta,\mu})
\end{aligned}$$

But this also means that:

$$\begin{aligned}
\Phi(x_t^{\delta,\mu}) \geq \int_a^t \frac{\partial\Phi(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu}d\alpha)}{\partial s}ds + \int_a^t \frac{b(x_s^{\delta,\mu}) - K_1\delta}{\sigma(x_s^{\delta,\mu})}ds + \\
+ (1 - sign(y_\mu(t) - y_\mu(a))K_6\delta)(y_\mu(t) - y_\mu(a)) + \Phi(x_a^{\delta\mu})
\end{aligned}$$

But thanks to a result from Wong and Zakai:

$$\Phi(x_t^S) = \int_a^t \frac{\partial\Phi(x_s^S)}{\partial s}ds + W_t - W_a + \int_a^t \frac{b(x_s^S)}{\sigma(x_s^S)}ds + \Phi(x_a^S) \tag{12}$$

as a special case of equation 11 from their paper.
Since b(x) is Lipshitz, it follows that $|b(x)| \leq K(1+|x|)$ for some K, therefore:

$$\begin{aligned}
|\frac{b(x)}{\sigma(x)} - \frac{b(y)}{\sigma(y)}| &\leq |\frac{b(x)}{\sigma(x)} - \frac{b(y)}{\sigma(x)}| + |\frac{b(y)}{\sigma(x)} - \frac{b(y)}{\sigma(y)}| &(13) \\
&\leq (K_2/\beta)(1+|y|)|x-y| &(14)
\end{aligned}$$

Also, since $\sigma^{-2}(x)$ and $\partial\sigma(x)/\partial t$ are uniformly bounded we have:

$$|\frac{\partial\Phi(x)}{\partial t} - \frac{\partial\Phi(y)}{\partial t}| \leq K_3|x-y| \tag{15}$$

By a special case of equation 14 from Wong and Zakai we know that:

$$|\Phi(x) - \Phi(y)| \geq K_4 log(\frac{(1 + |x - y|)}{1 + |y|}) \tag{16}$$

Let $u = 1 + max_{a \leq t \leq b} x_t^S$ which is finite a.s. Then if we subtract $|\Phi(x_t^{\delta,\mu}) - \Phi(x_t^S)|$ we get the following bound:

$$log(1 + \frac{|x_t^{\delta,\mu} - x_t^S|}{u}) \geq K_5||y_\mu(t) - W_t| + K_6\delta| + K_5||y_\mu(a) - W_a| + K_6\delta| + \\ + K_5 u \int_a^t ||x_s^{\delta,\mu} - x_s^S| + K_7\delta| ds$$

At this point we apply lemma 2 with $\epsilon_\mu(t)) = \epsilon(t) = exp(K_5||y_\mu(t) - W_t| + K_6\delta| + K_5||y_\mu(a) - W_a| + K_6\delta| + K_8(t-a)u\delta) - 1$

Since δ goes to zero as μ goes to zero, $\epsilon_\mu(t)$ goes to zero as μ goes to zero, and therefore by dominated convergence so does $\int_a^b \epsilon_\mu(t)dt$. Therefore $x_t^{\delta,\mu} \to x_t^S$ a.s.

However, if we take $\mu \downarrow 0$ first and assume that $\delta >> \mu$ so that the limits can be taken separately, then we get a different result. Namely we find that $x_t^{\delta,\mu} \to x_t$ from the following theorem:

Theorem 6 *If first $\mu \downarrow 0$ and then $\delta \downarrow 0$, and under the same conditions as above, then $x_t^{\delta,\mu}$ converges to x_t in probability.*

Proof: Since both $x_t^{\delta,\mu}$ and x_t^δ are smooth, we can use integration by parts to prove this. So applying integration by parts:

$$\begin{aligned} x_t^{\delta,\mu} - x_0 &= \int_0^t b(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu} d\alpha)ds + \int_0^t \sigma(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu} d\alpha)dy_\mu(s) \\ &= \int_0^t b(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu} d\alpha) - \frac{\partial}{\partial t}(\sigma(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu} d\alpha))y_\mu(s)ds \\ &\quad - \sigma(o)y_\mu(0) + \sigma(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu} d\alpha)y_\mu(t) \end{aligned}$$

And

$$\begin{aligned} x_t^\delta - x_0 &= \int_0^t b(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha)ds + \int_0^t \sigma(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha)dW_s \\ &= \int_0^t b(\int_0^t h(s-\alpha)x_\alpha^{\delta,} d\alpha) - \frac{\partial}{\partial t}(\sigma(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha))W_s ds - \sigma(o)W_0 \\ &\quad + \sigma(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha)W_t \end{aligned}$$

So subtracting these two we get

$$
\begin{aligned}
|x_t^{\delta,\mu} - x_t^\delta| \leq & \int_0^t |b(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu} d\alpha) - b(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha)|ds + \\
& + \int_0^t |\frac{\partial}{\partial t}(\sigma(\int_0^t h(s-\alpha)x_\alpha^{\delta,\mu} d\alpha))y_\mu(s) - \\
& - \frac{\partial}{\partial t}(\sigma(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha))W_s|ds + \sigma(0)|y_\mu(0) - W_0| + \\
& + |\sigma(\int_0^t h(t-\alpha)x_\alpha^{\delta,\mu} d\alpha)y_\mu(t) - \sigma(\int_0^t h(t-\alpha)x_\alpha^\delta d\alpha)W_t| \\
\leq & K_1 \int_0^t |x_s^{\delta,\mu} - x_s^\delta|ds + \int_0^t |\frac{\partial}{\partial t}\sigma(\int_0^t h(s-\alpha)x_\alpha^\delta d\alpha)||y_\mu(s) - W(s)| + \\
& + \int_0^t K_2|x_\alpha^{\delta,\mu} - x_\alpha^\delta| \int_0^t h(s-\alpha)dy_\mu(s)d\alpha + K_3|y_\mu(0) - W(0)| + \\
& + \sigma(\int_0^t h(t-\alpha)x_\alpha^\delta d\alpha)|y_\mu(t) - W_t| \\
\leq & K_1 \int_0^t |x_s^{\delta,\mu} - x_s^\delta|ds + K_4|y_\mu(s) - W(s)| + K_5 \int_0^t |x_\alpha^{\delta,\mu} - x_\alpha^\delta|d\alpha + \\
& + K_3|y_\mu(0) - W(0)| + K_6|y_\mu(t) - W_t| \\
\leq & K_7 \int_0^t |x_s^{\delta,\mu} - x_s^\delta|ds + K_4|y_\mu(s) - W(s)| + K_3|y_\mu(0) - W(0)| + \\
& + K_6|y_\mu(t) - W_t|
\end{aligned}
$$

Then applying Gronwall's inequality and the rules of expectation we get $E|x_t^{\delta,\mu} - x_t^\delta| \leq E[K_4|y_\mu(s) - W(s)|d\alpha + K_3|y_\mu(0) - W(0)| + K_6|y_\mu(t) - W_t|]e^{K_7 T}$ which clearly goes to zero as $\mu \downarrow 0$. We then apply Theorem 4 to obtain our result.

So we see that if we take the limit as first $\mu \downarrow 0$ then $\delta \downarrow 0$ we need to understand the stochastic integral in the Ito sense and if we take $\delta \downarrow 0$ then $\mu \downarrow 0$ we should understand the integral in the Stratonovich sense.

Bibliography

1. D.L. Fisk. Quasi-martingales and Stochastic Integrals. *Technical Report* **1** Dept. of Math. Michigan State University (1963).
2. M. Freidlin. Some Remarks on the Smoluchowski-Kramers Approximation. *Journal of Statistical Physics* **117** 617-634 (2004).
3. N. Ikeda and S Watanabe. *Stochastic Differential Equations and Diffusion Processes* Second Edition (1989).
4. K. Ito. Ito Stochastic Integral. *Proc. Imp. Acad. Tokyo* **20** 519-524 (1944).
5. I. Karatzas and S. Shreve. *Brownian Motion and Stocastic Calculus* Second Edition (1991)
6. L. Koralov and Y. Sinai. *Theory of Probability and Random Processes* 2nd Edition (2007).
7. B. Oksendal. *Stochastic Differential Equations: an Introduction with Applications* 6th Edition (2007).
8. R.L. Stratonovich. A New representation for stochastic integrals and equations. *J. SIAM Control* **4:** 362-371 (1966).
9. E. Wong and M. Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Stat* **36:** 1560-1564 (1965).