**3. Exercise.** Let $b_t$ be a $d$-dimensional process, $b_t \in \mathcal{S}$. Prove that

$$\big(\exp(\int_0^t b_t\,dw_t - (1/2)\int_0^t |b_t|^2\,dt), \mathcal{F}_t\big)$$

is a supermartingale.

## 5. Itô's formula

In the usual calculus, after the notion of integral is introduced one discusses the rules of integration and compiles the table of "elementary" integrals. The most important tools of integration are change of variable and integration by parts, which are proved on the basis of the formula for differentiating superpositions. The formula for the *stochastic* differential of a superposition is called *Itô's formula*. This formula was discovered in [**It**] as a curious fact and then became the main tool of modern stochastic calculus.

**1. Definition.** Let $(\Omega, \mathcal{F}, P)$ be a complete probability space carrying a $d_1$-dimensional Wiener process $(w_t, \mathcal{F}_t)$ and a continuous $d$-dimensional $\mathcal{F}_t$-adapted process $\xi_t$. Assume that we are also given a $d \times d_1$ matrix valued process $\sigma_t$ and a $d$-dimensional process $b_t$ such that $\sigma \in \mathcal{S}$ and $b$ is jointly measurable in $(\omega, t)$, $\mathcal{F}_t$-adapted, and $\int_0^T |b_s|\,ds < \infty$ (a.s.) for any $T < \infty$. Then we write

$$d\xi_t = \sigma_t\,dw_t + b_t\,dt$$

if and only if (a.s.) for all $t$

$$\xi_t = \xi_0 + \int_0^t \sigma_s\,dw_s + \int_0^t b_s\,ds. \tag{1}$$

In that case one says that $\xi_t$ has *stochastic differential* equal to $\sigma_t\,dw_t + b_t\,dt$.

From calculus we know that if $f(x)$ and $g(t)$ are differentiable, then

$$df(g(t)) = f'(g(t))\,dg(t).$$

It turns out that stochastic differentials possess absolutely different properties. For instance, consider $d(w_t^2)$ for one-dimensional $w_t$. If the usual rules were true, we would have $dw_t^2 = 2w_t\,dw_t$, that is,

$$w_t^2 = 2\int_0^t w_s\,dw_s.$$

However, this is impossible since

$$Ew_t^2 = t, \quad E\int_0^t w_s^2\, ds < \infty, \quad E\int_0^t w_s\, dw_s = 0.$$

Still, there is a case in which the usual formula holds. This case was found by Hitsuda. Let $(w_t', w_t'')$ be a two-dimensional Wiener process and define the *complex* Wiener process by

$$z_t = w_t' + iw_t''.$$

It turns out (see Exercise 5) that, for any analytic function $f(z)$, we have $df(z_t) = f'(z_t)\, dz_t$, that is,

$$f(z_t) = f(0) + \int_0^t f(z_s)\, dz_s. \tag{2}$$

We have what would be "the usual formula" if $z_t$ were piecewise differentiable.

We have introduced formal $d_1$-dimensional expressions $\sigma_t\, dw_t + b_t\, dt$. Now we define rules of operating with them. We assume that while multiplying them by constants, adding up, and evaluating their scalar products the usual algebraic rules of factoring out and combining similar terms are enforced *along with* the following multiplication table (which, by the way, keeps the products of stochastic differentials in the set of stochastic differentials):

$$dw_t^i dw_t^j = \delta^{ij}\, dt, \quad dw_t^i dt = (dt)^2 = 0. \tag{3}$$

A crucial role in the proof of Itô's formula is played by the following.

**2. Lemma.** *Let $\xi_t, \eta_t$ be real-valued processes having stochastic differentials. Then $\xi_t\eta_t$ also has a stochastic differential, and*

$$d(\xi_t\eta_t) = \eta_t\, d\xi_t + \xi_t\, d\eta_t + (d\xi_t)d\eta_t.$$

Proof. Let

$$\xi_t = \xi_0 + \int_0^t \sigma_s\, dw_s + \int_0^t b_s\, ds, \quad \eta_t = \eta_0 + \int_0^t \tilde{\sigma}_s\, dw_s + \int_0^t \tilde{b}_s\, ds,$$

where $\sigma_s$ and $\tilde{\sigma}_s$ are vector-valued processes and $b_s$ and $\tilde{b}_s$ are real-valued ones. By the above rules, assuming the summation convention, we can write

$$\eta_t d\xi_t = \eta_t(\sigma^k_t\,dw^k_t + b_t\,dt) = \eta_t\sigma^k_t\,dw^k_t + \eta_t b_t\,dt = \eta_t\sigma_t\,dw_t + \eta_t b_t\,dt,$$

$$\xi_t d\eta_t = \xi_t\tilde\sigma_t\,dw_t + \xi_t\tilde b_t\,dt, \quad (d\xi_t)d\eta_t = \sigma^j_t\,dw^j_t\tilde\sigma^k_t\,dw^k_t = \sigma^j_t\tilde\sigma^j_t\,dt = \sigma_t\cdot\tilde\sigma_t\,dt.$$

Therefore our assertion means that, for all $t \in [0,\infty)$ at once, with probability one,

$$\xi_t\eta_t = \xi_0\eta_0 + \int_0^t (\eta_s\sigma_s + \xi_s\tilde\sigma_s)\,dw_s + \int_0^t (\eta_s b_s + \xi_s\tilde b_s + \sigma_s\cdot\tilde\sigma_s)\,ds. \qquad (4)$$

First, notice that the right-hand side of (4) makes sense because (a.s.)

$$\int_0^t |\eta_s b_s|\,ds \le \max_{s\le t}|\eta_s| \int_0^t |b_s|\,ds < \infty,$$

$$\int_0^t |\eta_s\sigma^j_s|^2\,ds \le \max_{s\le t}|\eta_s| \int_0^t |\sigma^j_s|^2\,ds < \infty,$$

$$\int_0^t |\sigma_s\cdot\tilde\sigma_s|\,ds \le \int_0^t |\sigma_s|^2\,ds + \int_0^t |\tilde\sigma_s|^2\,ds < \infty.$$

Next, notice that if $d\xi'_t = \sigma'_t\,dw_t + b'_t\,dt$ and $d\xi''_t = \sigma''_t\,dw_t + b''_t\,dt$ and (4) holds with $\xi',\sigma',b'$ and $\xi'',\sigma'',b''$ in place of $\xi,\sigma,b$, then it also holds for $\xi'+\xi'',\sigma'+\sigma'',b'+b''$. It follows that we may concentrate only on two possibilities for $d\xi_t$: $d\xi_t = \sigma_t\,dw_t$ and $d\xi_t = b_t\,dt$. We have the absolutely similar situation with $\eta$. Therefore, we have to deal only with four pairs of $d\xi_t$ and $d\eta_t$. To finish our preparation, we also notice that both sides of (4) are continuous in $t$, so that to prove that they coincide with probability one for all $t$ at once, it suffices to prove that they are equal almost surely for each particular $t$.

Thus, fix $t$, and first let $d\xi_t = b_t\,dt$ and $d\eta_t = \tilde b_t\,dt$. Then (4) follows from the usual calculus (or is proved as in the following case).

The two cases, (i) $d\xi_t = \sigma_t\,dw_t$ and $d\eta_t = \tilde b_t\,dt$ and (ii) $d\xi_t = b_t\,dt$ and $d\eta_t = \tilde\sigma_t\,dw_t$, are similar, and we concentrate on (i).

Let $0 = t_{m0} \le t_{m1} \le ... \le t_{mk_m} = t$ be a sequence of partitions of $[0,t]$ such that $\max_i(t_{m,i+1} - t_{mi}) \to 0$ as $m \to \infty$. Define

$$\kappa_m(s) = t_{mi}, \quad \tilde\kappa_m(s) = t_{m,i+1} \quad \text{if} \quad s \in [t_{mi}, t_{m,i+1}).$$

Obviously $\kappa_m(s), \tilde\kappa_m(s) \to s$ uniformly on $[0,t]$. In addition, the formula

$$ab - cd = (a-c)d + (b-d)a$$

and Theorem 3.14 show that (a.s.)

$$\xi_t\eta_t - \xi_0\eta_0 = \sum_{i=0}^{k_m-1}(\xi_{t_{m,i+1}}\eta_{t_{m,i+1}} - \xi_{t_{mi}}\eta_{t_{mi}})$$

$$= \sum_{i=0}^{k_m-1}\eta_{t_{mi}}\int_{t_{mi}}^{t_{m,i+1}}\sigma_s\,dw_s + \sum_{i=0}^{k_m-1}\xi_{t_{m,i+1}}\int_{t_{mi}}^{t_{m,i+1}}\tilde{b}_s\,ds$$

$$= \int_0^t \eta_{\kappa_m(s)}\sigma_s\,dw_s + \int_0^t \xi_{\tilde{\kappa}_m(s)}\tilde{b}_s\,ds. \tag{5}$$

Furthermore, as $m\to\infty$, we have (a.s.)

$$\Big|\int_0^t \xi_{\tilde{\kappa}_m(s)}\tilde{b}_s\,ds - \int_0^t \xi_s\tilde{b}_s\,ds\Big| \le \sup_{s\le t}|\xi_{\tilde{\kappa}_m(s)} - \xi_s|\int_0^t|\tilde{b}_s|\,ds \to 0,$$

$$\int_0^t |\eta_{\kappa_m(s)} - \eta_s|^2(\sigma_s^j)^2\,ds \le \sup_{s\le t}|\eta_{\kappa_m(s)} - \eta_s|^2\int_0^t(\sigma_s^j)^2\,ds \to 0,$$

and the last relation by Theorem 3.5 (iii) implies that

$$\int_0^t \eta_{\kappa_m(s)}\sigma_s\,dw_s \xrightarrow{P} \int_0^t \eta_s\sigma_s\,dw_s. \tag{6}$$

Now by letting $m\to\infty$ in (5) we get (4) (a.s.) in our particular case.

Thus it only remains to consider the case $d\xi_t = \sigma_t\,dw_t$, $d\eta_t = \tilde{\sigma}_t\,dw_t$, and prove that

$$\xi_t\eta_t = \xi_0\eta_0 + \int_0^t(\eta_s\sigma_s + \xi_s\tilde{\sigma}_s)\,dw_s + \int_0^t \sigma_s\cdot\tilde{\sigma}_s\,ds. \tag{7}$$

Notice that we may assume that $\xi_0 = \eta_0 = 0$, since in the initial reduction to four cases we could absorb the initial values in the terms with $dt$.

Now we again use bilinearity and conclude that, since $\sigma$ and $\tilde{\sigma}$ can be represented as sums of vector-valued processes each of which has only one nonidentically zero element, we only have to prove (7) for such simple vector-valued processes. Furthermore, keeping in mind that each $f\in\mathcal{S}$ can be approximated by $f^n\in H_0$ (see, for instance, the proof of Theorem 3.14), we see that we may assume that $\sigma^j, \tilde{\sigma}^j \in H_0$.

In this way we conclude that to prove (7) in the general case, it suffices to prove that, if $f,g\in H_0$, $\xi_r = \int_0^r f_s\,dw_s^i$, and $\eta_r = \int_0^r g_s\,dw_s^j$, then (a.s.)

$$\xi_t\eta_t = \int_0^t f_s\eta_s\,dw^i_s + \int_0^t g_s\xi_s\,dw^j_s + \int_0^t f_s g_s\delta^{ij}\,ds. \tag{8}$$

Remember that $t$ is fixed, and without losing generality assume that the partitions corresponding to $f$ and $g$ coincide and $t$ is one of the partition points. Let $\{t_0, t_1, ...\}$ be the common partition with $t = t_k$. Next, as above we take the sequence of partitions defined by $t_{mi}$ of $[0,t]$ and again without loss of generality assume that each $t_i$ lying in $[0,t]$ belongs to $\{t_{mi} : i = 0, 1, ...\}$. We use the formula

$$ab - cd = (a-c)d + (b-d)c + (a-c)(b-d) \tag{9}$$

and Theorem 3.14. Fix a $q = 0, ..., k-1$ and, by default summing up with respect to those $r$ for which $t_q \le t_{mr} < t_{q+1}$, write (a.s.)

$$\xi_{t_{q+1}}\eta_{t_{q+1}} - \xi_{t_q}\eta_{t_q} = \sum(\xi_{t_{m,r+1}}\eta_{t_{m,r+1}} - \xi_{t_{mr}}\eta_{t_{mr}})$$

$$= \sum \eta_{t_{mr}}\int_{t_{mr}}^{t_{m,r+1}} f_s\,dw^i_s + \sum \xi_{t_{mr}}\int_{t_{mr}}^{t_{m,r+1}} g_s\,dw^j_s$$

$$+\sum\int_{t_{mr}}^{t_{m,r+1}} f_s\,dw^i_s\int_{t_{mr}}^{t_{m,r+1}} g_s\,dw^j_s = \int_{t_q}^{t_{q+1}}\eta_{\kappa_m(s)}f_s\,dw^i_s +$$

$$+\int_{t_q}^{t_{q+1}}\xi_{\kappa_m(s)}g_s\,dw^j_s + f_{t_q}g_{t_q}\sum(w^i_{t_{m,r+1}} - w^i_{t_{mr}})(w^j_{t_{m,r+1}} - w^j_{t_{mr}}). \tag{10}$$

In the expression after the last equality sign the first two terms converge in probability to

$$\int_{t_q}^{t_{q+1}}\eta_s f_s\,dw^i_s, \quad \int_{t_q}^{t_{q+1}}\xi_s g_s\,dw^j_s$$

respectively, which is proved in the same way as (6). If $i = j$, the last term converges in probability to

$$f_{t_q}g_{t_q}(t_{q+1} - t_q) = \int_{t_q}^{t_{q+1}} f_s g_s\,ds$$

by Theorem 2.2.6. Consequently, by letting $m \to \infty$ in (10) and then adding up the results for $q = 0, ..., k-1$, we come to (7) if $i = j$. For $i \ne j$ one uses the same argument complemented by the observation that the last sum in (10) tends to zero in probability, since its mean is zero due to the independence of $w^i$ and $w^j$, and

$$E\big[\sum(w^i_{t_{m,r+1}} - w^i_{t_{mr}})(w^j_{t_{m,r+1}} - w^j_{t_{mr}})\big]^2 = \text{Var}\,[...]$$

$$= \sum E(w^i_{t_{m,r+1}} - w^i_{t_{mr}})^2(w^j_{t_{m,r+1}} - w^j_{t_{mr}})^2$$

$$= \sum(t_{m,r+1} - t_{mr})^2 \le \max_i(t_{m,i+1} - t_{mi})t \to 0.$$

The lemma is proved.

**3. Exercise.** Explain why in the treatment of the fourth case one cannot use a formula similar to (5) in place of (10).

**4. Theorem** (Itô's formula)**.** *Let a $d_1$-dimensional process $\xi_t$ have stochastic differential, and let $u(x) = u(x^1, ..., x^{d_1})$ be a real-valued twice continuously differentiable function of $x \in \mathbb{R}^{d_1}$. Then $u(\xi_t)$ has a stochastic differential, and*

$$du(\xi_t) = u_{x^i}(\xi_t)\,d\xi^i_t + (1/2)u_{x^ix^j}(\xi_t)\,d\xi^i_t d\xi^j_t. \tag{11}$$

Proof. Let $C^2$ be the set of all real-valued twice continuously differentiable function on $\mathbb{R}^{d_1}$. We are going to use the fact that for every $u \in C^2$ there is a sequence of polynomials $u^m$ such that $u^m, u^m_{x^i}, u^m_{x^ix^j}$ converge to $u, u_{x^i}, u_{x^ix^j}$ uniformly on each ball. For such a sequence and any $\omega, t, i, j$

$$\sup_{s\le t}|u^m_{x^i}(\xi_t) - u_{x^i}(\xi_t)| + \sup_{s\le t}|u^m_{x^ix^j}(\xi_t) - u_{x^ix^j}(\xi_t)| \to 0,$$

since each trajectory of $\xi_s$, $s \le t$, lies in a ball. It follows easily that, if (11) is true for $u^m$, then it is also true for $u$.

Thus, we only need to prove (11) for polynomials, and to do this it obviously suffices to show that (11) holds for linear function and also for the product of any two functions $u$ and $v$ for each of which (11) holds.

For linear $u$ formula (11) is obvious. If (11) holds for $u$ and $v$, then by Lemma 2

$$d(u(\xi_t)v(\xi_t)) = u(\xi_t)\,dv(\xi_t) + v(\xi_t)\,du(\xi_t) + (du(\xi_t))dv(\xi_t)$$

$$= [uv_{x^i} + vu_{x^i}](\xi_t)\,d\xi^i_t + (1/2)[uv_{x^ix^j} + vu_{x^ix^j}](\xi_t)\,d\xi^i_t d\xi^i_t + u_{x^i}v_{x^j}(\xi_t)\,d\xi^i_t d\xi^i_t$$

$$= (uv)_{x^i}(\xi_t)\,d\xi^i_t + (1/2)(uv)_{x^ix^j}(\xi_t)\,d\xi^i_t d\xi^i_t.$$

The theorem is proved.

Itô's formula (11) looks very much like Taylor's formula with two terms. Usually one rewrites it in a different way. Namely, let $d\xi_t = \sigma_t \, dw_t + b_t \, dt$, $a = (1/2)\sigma_t \sigma_t^*$. Simple manipulations show that $(d\xi_t^i) d\xi_t^j = 2a_t^{ij} \, dt$ and hence

$$du(\xi_t) = L_t u(\xi_t) \, dt + \sigma_t^* u_x(\xi_t) \, dw_t,$$

where $u_x = \operatorname{grad} u$ is a column vector and $L_t$ is the second-order differential operator given by

$$L_t v(x) = a_t^{ij} v_{x^i x^j}(x) + b_t^i v_{x^i}(x).$$

In this notation (11) means that for all $t$ (a.s.)

$$u(\xi_t) = u(\xi_0) + \int_0^t L_s u(\xi_s) \, ds + \int_0^t \sigma_s^* u_x(\xi_s) \, dw_s. \tag{12}$$

**5. Exercise.** Prove that (2) holds for analytic functions $f$.

Itô's formula leads to extremely important formulas relating the theory of stochastic integration with the theory of partial differential equations. One of them is the following theorem.

**6. Theorem.** *Let $\xi_0$ be nonrandom, let $Q$ be a domain in $\mathbb{R}^{d_1}$, let $\xi_0 \in Q$, let $\tau$ be the first exit time of $\xi_t$ from $Q$, and let $u$ be a function which is continuous in $\bar{Q}$ and has continuous first and second derivatives in $Q$. Assume that*

$$P(\tau < \infty) = 1, \quad E \int_0^\tau |L_s u(\xi_s)| \, ds < \infty.$$

*Then*

$$u(\xi_0) = E u(\xi_\tau) - E \int_0^\tau L_s u(\xi_s) \, ds.$$

We give no proof to this theorem because it is just a particular result, and usually when one needs such results it is easier and shorter to prove what is needed directly instead of trying to find the corresponding result in the literature. We will see examples of this in Sec. 7.

Roughly speaking, to prove Theorem 6 one plugs $\tau$ in place of $t$ in (12) and takes expectations. The main difficulties on the way are caused by the fact that $u$ is not even given in the whole $\mathbb{R}^{d_1}$ and the expectation of a stochastic integral does not necessarily exist, let alone equal zero. One overcomes these difficulties by taking smaller domains $Q_m \uparrow Q$, extending $u$ outside $Q_m$, taking $\tau$ even smaller that the first exit time from $Q_m$, and then passing to the limit.

## 6. An alternative proof of Itô's formula

The approach we have in mind is based on using stopping times and stochastic intervals. It turns out that these tools could be used right from the beginning, even for defining Itô integral. First we briefly outline how to do this, to give the reader one more chance to go through the basics of the theory and also to show a way which is valid for integrals against more general martingales.

**1. Definition.** Let $\tau = \tau(\omega)$ be a $[0, \infty)$-valued function on $\Omega$ taking only finitely many values, say $t_1, ..., t_n \geq 0$. We say that $\tau$ is a *simple stopping time* (relative to $\mathcal{F}_t$) if $\{\omega : \tau(\omega) = t_k\} \in \mathcal{F}_{t_k}$ for any $k = 1, ..., n$. The set of all simple stopping times is denoted by $\mathcal{M}$.

Below in this section we only use simple stopping times.

**2. Exercise*.** (i) Prove that simple stopping times are stopping times, and that $\{\omega : \tau(\omega) \geq t\} \in \mathcal{F}_t$ for any $t$.

(ii) Derive from (i) that if $\tau_1$ and $\tau_2$ are simple stopping times, then $\tau_1 \wedge \tau_2$ and $\tau_1 \vee \tau_2$ are simple stopping times as well.

**3. Lemma.** *For a real-valued function $\gamma(\omega)$, define the* stochastic interval $(\!(0, \gamma]\!]$ *as the set $\{(\omega, t) : \omega \in \Omega, 0 < t \leq \gamma(\omega)\}$ and let $\Pi$ be the collection of all stochastic intervals $(\!(0, \tau]\!]$ with $\tau$ running through the set of all simple stopping times. Finally, for $\Delta = (\!(0, \tau]\!] \in \Pi$, define $\zeta(\Delta) = w_\tau$. Then $\zeta$ is a random orthogonal measure on $\Pi$ with reference measure $\mu = P \times \ell$ and $E\zeta(\Delta) = 0$ for any $\Delta \in \Pi$.*

Proof. Let $\tau$ be a simple stopping time and $\{t_1, ..., t_n\}$ the set of its values. Then

$$w_\tau = w_{t_1} I_{\tau = t_1} + ... + w_{t_n} I_{\tau = t_n}$$

and, since $Ew_s^2 < \infty$, $E\zeta^2((\!(0, \tau]\!]) = Ew_\tau^2 < \infty$.

Next we will be using the simple fact that, if $\tau$ is a simple stopping time and the set $\{0 = t_0 < t_1 < ... < t_n\}$ contains all possible values of $\tau$, then

$$w_\tau = \sum_{i=0}^{n-1} f_{t_i}(w_{t_{i+1}} - w_{t_i}), \quad \tau = \sum_{i=0}^{n-1} f_{t_i}(t_{i+1} - t_i), \tag{1}$$

where $f_t := I_{\tau > t}$ is $\mathcal{F}_t$-measurable (Exercise 2). Since $\{\omega : \tau(\omega) > t_i\} \in \mathcal{F}_{t_i}$ and $w_{t_{i+1}} - w_{t_i}$ is independent of $\mathcal{F}_{t_i}$, we have

$$Ef_{t_i}(w_{t_{i+1}} - w_{t_i}) = Ef_{t_i}E(w_{t_{i+1}} - w_{t_i}) = 0, \quad E\zeta((\!(0, \tau]\!]) = 0.$$

Now, let $\tau$ and $\sigma$ be simple stopping times, $\{t_1, ..., t_n\}$ the ordered set of their values, and $\Delta_1 = (0, \tau]\!]$ and $\Delta_2 = (0, \sigma]\!]$. By using (1) we have

$$E\zeta(\Delta_1)\zeta(\Delta_2) = Ew_\tau w_\sigma = \sum_{i,j=0}^{n-1} Ef_{t_i}g_{t_j}(w_{t_{i+1}} - w_{t_i})(w_{t_{j+1}} - w_{t_j}),$$

which, in the same way as in the proofs of Theorem 2.7.3 or Lemma 1.3 used in other approaches, is shown to be equal to

$$\sum_{i=0}^{n-1} Ef_{t_i}g_{t_i}(t_{i+1} - t_i) = E\sum_{i=0}^{n-1} I_{\tau\wedge\sigma > t_i}(t_{i+1} - t_i) = E\tau \wedge \sigma.$$

Since

$$E\tau \wedge \sigma = \int_\Omega \int_0^\infty I_{(0,\tau]\!]\cap(0,\sigma]\!]}(\omega, t)\, P(d\omega)dt = \mu(\Delta_1 \cap \Delta_2),$$

the lemma is proved.

From this lemma we derive the following version of Wald's identities.

**4. Corollary.** *Let $\tau_1$ and $\tau_2$ be simple stopping times. Then $Ew^2_{\tau_1} = E\tau_1$ and $E(w_{\tau_1} - w_{\tau_2})^2 = E|\tau_1 - \tau_2|$.*

Indeed, we get the first equality from the proof of Lemma 3 by taking $\sigma = \tau$. To prove the second one, define $\tau = \tau_1 \vee \tau_2$, $\sigma = \tau_1 \wedge \tau_2$ and notice that

$$E(w_{\tau_1} - w_{\tau_2})^2 = E(w_\tau - w_\sigma)^2 = Ew^2_\tau - 2Ew_\tau w_\sigma + Ew^2_\sigma$$
$$= E\tau - E\sigma = E(\tau - \sigma) = E|\tau_1 - \tau_2|.$$

**5. Exercise.** Carry over the result of Corollary 4 to all bounded stopping times.

**6. Remark.** Lemma 3 and the general Theorem 2.3.13 imply that there is a stochastic integral operator, say $I$, defined on $L_2(\Pi, \mu)$ with values in $L_2(\mathcal{F}, P)$. Since $\Pi$ is a $\pi$-system of subsets of $\Omega \times (0, \infty)$, we have $L_2(\Pi, \mu) = L_2(\sigma(\Pi), \mu)$ due to Theorem 2.3.19.

**7. Remark.** It turns out that $\sigma(\Pi) = \mathcal{P}$. Indeed, on the one hand the indicators of the sets $(0, \tau]\!]$ generating $\sigma(\Pi)$ are left-continuous and $\mathcal{F}_t$-adapted, hence predictable (Exercise 2.8.3). In other words, $(0, \tau]\!] \in \mathcal{P}$ and $\sigma(\Pi) \subset \mathcal{P}$. On the other hand, if $A \in \mathcal{F}_s$, $s \geq 0$, and for $n > s$ we define $\tau_n = s$ on $A$ and $\tau_n = n$ on $\Omega \setminus A$, then $\tau_n$ are simple stopping times and

$$(0, \tau_n]\!] = \{(\omega, t) : 0 < t \leq \tau_n(\omega)\}$$

$$= \{(\omega, t) : 0 < t \le s, \omega \in A\} \bigcup \{(\omega, t) : 0 < t \le n, \omega \in A^c\},$$

$$\bigcup_n (\!(0, \tau_n]\!] = (A \times (0, s]) \cup (A^c \times (0, \infty)) \in \sigma(\Pi),$$

so that $(\bigcup_n (\!(0, \tau_n]\!])^c = A \times (s, \infty) \in \sigma(\Pi)$. It follows that the set generating $\mathcal{P}$ is a subset of $\sigma(\Pi)$ and $\mathcal{P} \subset \sigma(\Pi)$.

**8. Remark.** Remark 7 and the definition of $L_2(\Pi, \mu)$ imply the somewhat unexpected result that for every $f \in L_2(\mathcal{P}, \mu)$, in particular, $f \in H$, there are simple stopping times $\tau_i^m$ and *constants* $c_i^m$ defined for $m = 1, 2, ...$ and $i = 1, ..., k(m) < \infty$ such that

$$E \int_0^\infty |f_t - \sum_{i=1}^{k(m)} c_i^m I_{(\!(0, \tau_i^m]\!]}(t)|^2 \, dt \to 0$$

as $m \to \infty$.

**9. Exercise.** Find simple stopping times $\tau_i^m$ and constants $c_i^m$ such that, for the one-dimensional Wiener process $w_t$,

$$E \int_0^\infty |I_{t \le 1} w_t - \sum_{i=1}^{k(m)} c_i^m I_{(\!(0, \tau_i^m]\!]}(t)|^2 \, dt \to 0$$

as $m \to \infty$.

**10. Remark.** The operator $I$ from Remark 6 coincides on $L_2(\Pi, \mu)$ with the operator of stochastic integration introduced before Remark 1.6. This follows easily from the uniqueness of continuation and Theorem 2.7, showing that the old stochastic integral coincides with the new one on the indicators of $(\!(0, \tau]\!]$ and both are equal to $w_\tau$.

After making sure that we deal with the same objects as in Sec. 5, we start proving Itô's formula, allowing ourselves to use everything proved before Sec. 5. As in Sec. 5, we need only prove Lemma 5.2. Define $\kappa_n(t) = 2^{-n}[2^n t]$.

Due to (5.9) we have

$$w_t^i w_t^j = \int_0^t w^i_{\kappa_n(s)} \, dw_s^j + \int_0^t w^j_{\kappa_n(s)} \, dw_s^i$$

$$+ \sum_{k=0}^\infty (w^i_{\frac{k+1}{2^n} \wedge t} - w^i_{\frac{k}{2^n} \wedge t})(w^j_{\frac{k+1}{2^n} \wedge t} - w^j_{\frac{k}{2^n} \wedge t}), \quad i, j = 1, ..., d \quad \text{(a.s.)}. \qquad (2)$$

By sending $n$ to infinity, from the theorem on quadratic variation of the Wiener process we get that (a.s.)

$$w_t^i w_t^j = \int_0^t w_s^i \, dw_s^j + \int_0^t w_s^j \, dw_s^i + \delta^{ij} t, \quad i, j = 1, ..., d. \tag{3}$$

Furthermore, for $\gamma, \tau \in \mathcal{M}$, $\gamma \leq \tau$, by using the fact that the sets of all values of $\gamma, \tau$ are finite, we obtain that

$$\int_0^\infty w_\gamma^j I_{\gamma < s \leq \tau} \, dw_s^i = w_\gamma^j (w_\tau^i - w_\gamma^i) \quad \text{(a.s.)}.$$

Hence and from (3) for $i, j = 1, ..., d$, $\tau, \sigma \in \mathcal{M}$, $\gamma = \tau \wedge \sigma$ we have (a.s.)

$$w_\tau^i w_\sigma^j = (w_\tau^i - w_\gamma^i) w_\gamma^j + (w_\sigma^j - w_\gamma^j) w_\gamma^i + w_\gamma^i w_\gamma^j$$

$$= \int_0^\infty w_\gamma^j I_{\gamma < s \leq \tau} \, dw_s^i + \int_0^\infty w_\gamma^i I_{\gamma < s \leq \sigma} \, dw_s^j$$

$$+ \int_0^\infty w_s^j I_{s \leq \gamma} \, dw_s^i + \int_0^\infty w_s^i I_{s \leq \gamma} \, dw_s^j + \delta^{ij} \gamma$$

$$= \int_0^\infty w_{s \wedge \sigma}^j I_{s \leq \tau} \, dw_s^i + \int_0^\infty w_{s \wedge \tau}^i I_{s \leq \sigma} \, dw_s^j + \int_0^\infty I_{s \leq \tau} I_{s \leq \sigma} \, ds.$$

By replacing here $\tau, \sigma$ by $\tau \wedge t, \sigma \wedge t$, we conclude that (a.s.)

$$w_{t \wedge \tau}^i = \int_0^t I_{s \leq \tau} \, dw_s^i, \quad w_{t \wedge \sigma}^j = \int_0^t I_{s \leq \sigma} \, dw_s^j,$$

$$w_{t \wedge \tau}^i w_{t \wedge \sigma}^j = \int_0^t w_{s \wedge \sigma}^j I_{s \leq \tau} \, dw_s^i + \int_0^t w_{s \wedge \tau}^i I_{s \leq \sigma} \, dw_s^j + \int_0^t I_{s \leq \tau} I_{s \leq \sigma} \, ds. \tag{4}$$

Next, similarly to our argument about (2) and (3), by replacing $w_t^j$ with $t$ and then $w_t^i$ with $t$ as well, instead of (4) we get

$$t \wedge \sigma = \int_0^t I_{s \leq \sigma} \, ds, \quad t \wedge \tau = \int_0^t I_{s \leq \tau} \, ds,$$

$$
\begin{aligned}
(t \wedge \sigma) w^i_{t\wedge\tau} &= \int_0^t (s \wedge \sigma) I_{s\le\tau}\, dw^i_s + \int_0^t w^i_{s\wedge\tau} I_{s\le\sigma}\, ds, \\
(t \wedge \tau)(t \wedge \sigma) &= \int_0^t (s \wedge \sigma) I_{s\le\tau}\, ds + \int_0^t (s \wedge \tau) I_{s\le\sigma}\, ds.
\end{aligned}
\tag{5}
$$

To finish the preliminaries, we observe that for each $\mathcal{F}_0$-measurable random variable $\xi_0$, obviously

$$
\xi_0 w^i_{t\wedge\tau} = \int_0^t \xi_0 I_{s\le\tau}\, dw^i_s, \quad (t\wedge\tau)\xi_0 = \int_0^t \xi_0 I_{s\le\tau}\, ds. \tag{6}
$$

Now we recall the notion of stochastic differential from before Lemma 5.2, and the multiplication table (5.3). Then we automatically have the following.

**11. Lemma.** *All the formulas* (4), (5), *and* (6) *can be written in one and the same way: If* $\xi_t, \eta_t$ *are real-valued processes and*

$$
d\xi_t = \sigma_t\, dw_t + b_t\, dt, \quad d\xi_t = \sigma'_t\, dw_t + b'_t\, dt, \tag{7}
$$

*where all entries of* $\sigma_t, \sigma'_t$ *and of* $b_t$, $b'_t$ *are indicators of elements of* $\Pi$, *then*

$$
d(\xi_t\eta_t) = \xi_t\, d\eta_t + \eta_t\, d\xi_t + (d\xi_t)(d\eta_t). \tag{8}
$$

Also notice that since both sides of equality (8) are linear in $\xi$ and in $\eta$, equality (8) immediately extends to all processes $\xi_t, \eta_t$ satisfying (7) with functions $\sigma, \sigma', b, b'$ of class $S(\Pi)$.

Now we are ready to prove Lemma 5.2, saying that (8) holds true for all scalar processes $\xi_t, \eta_t$ possessing stochastic differentials. To this end, assume first that $\sigma', b' \in S(\Pi)$ and take a sequence of processes $\sigma_n, b_n$ of class $S(\Pi)$ such that (a.s.)

$$
\int_0^T (|\sigma_t - \sigma_{nt}|^2 + |b_t - b_{nt}|)\, dt \to 0 \quad \forall T \in [0, \infty).
$$

Define also processes $\xi^n_t$, replacing $\sigma, b$ in (6) by $\sigma_n, b_n$. As is well known, in probability

$$\sup_{t\leq T}[|\int_0^t(\sigma_s-\sigma_{ns})\,dw_s|+|\int_0^t(b_s-b_{ns})\,ds|]\to 0,$$

$$\sup_{s\leq T}|\xi_t-\xi_t^n|\to 0\quad\forall T\in[0,\infty).\tag{9}$$

If necessary, we take a subsequence and we assume that the convergences in (9) hold almost surely. Then by the dominated convergence theorem we have (a.s.)

$$\int_0^T|\xi_t-\xi_t^n|(|\sigma_t'|^2+|b_t'|)\,dt\to 0,$$

$$\int_0^T|\eta_t|(|\sigma_t-\sigma_{nt}|^2+|b_t-b_{nt}|)\,dt\to 0,$$

$$\int_0^T|\sigma_t\cdot\sigma_t'-\sigma_{nt}\cdot\sigma_t'|\,dt$$

$$\leq(\int_0^T|\sigma_t-\sigma_{nt}|^2\,dt)^{1/2}(\int_0^T|\sigma_t'|^2\,dt)^{1/2}\to 0\quad\forall T\in[0,\infty).$$

This and an argument similar to the one which led us to (9) show that in the integral form of (8), with $\xi_t^n$ instead of $\xi_t$, we can pass to the limit in probability and get (8) for the limit process $\xi_t$. Of course, after this we fix the process $\xi_t$ and we carry out a similar limit passage in (8) affecting the second factor. In this way we get Lemma 5.2 in a straightforward way from the quite elementary Lemma 11.

## 7. Examples of applying Itô's formula

In this section $w_t$ is a $d$-dimensional Wiener process.

**1. Example.** Let $\tau$ be the first exit time of $w_t$ from $B_R=\{x:|x|<R\}$, where $R>0$ is a number. As we know, $\tau$ is a stopping time. Take

$$u(x)=(1/d)(R^2-|x|^2)$$

and apply Itô's formula to $u(w_t)$. Here $\xi_t=w_t$, $\sigma$ is the identity matrix, $b=0$, and the corresponding differential operator $L_t=(1/2)\Delta$. We have (a.s.)

$$u(w_t)=-t-\int_0^t(2/d)w_s\,dw_s+(1/d)R^2\quad\forall t.$$

Substitute $t\wedge\tau$ in place of $t$, take expectations, and notice that, since $|w_t|\leq R$ before $\tau$, we have $0\leq u(w_{t\wedge\tau})\leq(1/d)R^2$ and