Elements of Itô calculus

1 Introduction

Some basic facts about probability theory and stochastic processes are recalled in chapter III of [4]. The purely analytic introduction to Itô calculus given here is based on [1] (in French, english translation available as an appendix in [5]). Another good reference is [3] in particular chapter I.

2 Finite dimensional distributions of a stochastic process

Definition 2.1 (*Stochastic process*)**.** *A collection of random variables* $\{\boldsymbol{\xi}_t | t \geq 0\}$

$$\boldsymbol{\xi} : \Omega \times \mathbb{R}_+ \to \mathbb{R}^d$$

is called a stochastic process.

Realizations of stochastic process are now paths rather than numbers:

Definition 2.2 (*Sample path*)**.** *For each* $\omega \in \Omega$ *the mapping*

$$t \to \boldsymbol{\xi}_t(\omega)$$

is called the sample path *of the stochastic process.*

In most applications stochastic processes are characterized by means of the family of all *finite dimensional joint distributions associated* to them. This means that for a stochastic process valued on $\mathbb{R}$, for any discrete sequence $\{t_i\}_{i=1}^n$ we consider the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ and $B_1, \dots, B_n\, \mathcal{B}$ and consider

$$P_{\xi_t}(B_1, t_1, \dots, B_n, t_n) \equiv P(\xi_{t_1} \in B_1, \dots, \xi_{t_n} \in B_n)$$

The so defined families of joint probability yield a consistent description of a stochastic process

$$\boldsymbol{\xi} : \Omega \times \mathbb{R}_+ \to \mathbb{R}^d$$

if the following *Kolmogorov consistency conditions* are satisfied

- i $P(\mathbb{R}^d, t) = 1$ for any t
- ii $P_{\xi_t}(B_1, t_1, \dots, B_n, t_n) \geq 0$
- iii $P_{\xi_t}(B_1, t_1, \dots, B_n, t_n) = P_{\xi_t}(B_1, t_1, \dots, B_n, t_n, \mathbb{R}^d, t_{m+1})$
- iv $P_{\xi_t}(B_{\pi(1)}, t_{\pi(1)}, \dots, B_{\pi(n)}, t_{\pi(n)}) = P_{\xi_t}(B_1, t_1, \dots, B_n, t_n, \mathbb{R}^d, t)$

3 Wiener process

The above definitions allow us to characterize the Wiener process as a stochastic process:

Definition 3.1 (*Wiener Process aka Brownian motion*)**.** *A real valued stochastic process*

$$w_t \,:\, \Omega \times \mathbb{R}_+ \,\to\, \mathbb{R}$$

is called a **Wiener process** *or* **Brownian motion** *if*

i $w_0 = 0$

ii any increment $w_t - w_s$ *has Gaussian probability density*

$$w_t - w_s \overset{d}{=} \frac{e^{-\frac{x^2}{2\,(t-s)}}}{\sqrt{2\,\pi\,(t-s)}} \tag{3.1}$$

for all $t \,\geq\, s \,\geq\, 0.$

iii For all times

$$t_1 \,<\, t_2 \,<\, \ldots \leq t_n$$

the random variables

$$w_{t_1}\,, w_{t_2} - w_{t_1}\,, \ldots\,, w_{t_n} - w_{t_{n-1}}$$

are independent *(the process has independent increments).*

3.1 Consequences of the definition

Some observations are in order

- It is not restrictive to consider the one dimensional case. A d-dimensional Wiener process a vector valued stochastic process whose components are each independent one-dimensional Wiener processes. More explicitly, the probability density of Brownian motion on $\mathbb{R}^d$ is given by

$$p_{\boldsymbol{w}_t}(\boldsymbol{x}) = \prod_{i=1}^{d} p_{w_t^i}(x_i)$$

- By i and ii we have that

$$p_{w_t}(x) = \frac{e^{-\frac{x^2}{2\,\sigma^2\,t}}}{(2\,\pi\,\sigma^2\,t)^{\frac{1}{2}}} \qquad \& \qquad p_{w_{t_2}-w_{t_1}}(x) = \frac{e^{-\frac{x^2}{2\,\sigma^2\,(t_2-t_1)}}}{[2\,\pi\,\sigma^2\,(t_2-t_1)]^{\frac{1}{2}}} \quad t_2 \,>\, t_1 \tag{3.2}$$

By iii The joint probability of w_{t_1} and $w_{t_2} - w_{t_1}$ is

$$p_{w_{t_1},w_{t_2}-w_{t_1}}(x_1, y) = \frac{e^{-\frac{x_1^2}{2\,\sigma^2\,t_1}}}{(2\,\pi\,\sigma^2\,t_1)^{\frac{1}{2}}} \frac{e^{-\frac{y^2}{2\,\sigma^2\,(t_2-t_1)}}}{[2\,\pi\,\sigma^2\,(t_2-t_1)]^{\frac{1}{2}}}$$

By definition of probability density we can also write

$$p_{w_{t_1},w_{t_2}-w_{t_1}}(x_1,y) = p_{w_{t_1},w_{t_2}-x_1}(x_1,y) = p_{w_{t_1},w_{t_2}}(x_1,y+x_1)$$

since

$$w_{t_2} = (w_{t_2} - w_{t_1}) + w_{t_1}$$

Recalling the definition of *conditional probability* we must also have

$$p_{w_{t_1},w_{t_2}}(x_1,y+x_1) = p_{w_{t_2}|w_{t_1}}(x_1+y,t_2\,|\,x_1,t_1)\,p_{w_{t_1}}(x_1) \quad \forall\, x_1\,,x_2\,,t_2\,>\,t_1$$

whence

$$p_{w_{t_2}|w_{t_1}}(x_1+y,t_2\,|\,x_1,t_1) = \frac{p_{w_{t_1},w_{t_2}}(x_1,y+x_1)}{p_{w_{t_1}}(x_1)} = \frac{p_{w_{t_1},w_{t_2}-w_{t_1}}(x_1,y)}{p_{w_{t_1}}(x_1)} = \frac{e^{-\frac{y^2}{2\,\sigma^2\,(t_2-t_1)}}}{[2\,\pi\,\sigma^2\,(t_2-t_1)]^{\frac{1}{2}}}$$

Finally, upon setting $x_2 = y + x_1$ we get into:

$$p_{w_{t_2}|w_{t_1}}(x_2,t_2\,|\,x_1,t_1) = \frac{e^{-\frac{(x_2-x_1)^2}{2\,\sigma^2\,(t_2-t_1)}}}{[2\,\pi\,\sigma^2\,(t_2-t_1)]^{\frac{1}{2}}}$$

3.2 Continuity and non-differentiability of the Wiener process

We can ask what in the probability that realizations of the Wiener process are continuous. Since increments $w_t - w_s$ of the Wiener process are *stationary* as their distribution depends only upon $t - s$, To answer the question it is sufficient to estimate

$$P\left(|w_t| \geq \varepsilon\right) = 2\int_{\varepsilon}^{\infty} dx\,\frac{e^{-\frac{x^2}{2\,t}}}{\sqrt{2\,\pi\,t}} \leq 2\int_{0}^{\infty} dx\,\frac{x}{\varepsilon}\,\frac{e^{-\frac{x^2}{2\,t}}}{\sqrt{2\,\pi\,t}} = \frac{1}{\varepsilon}\sqrt{\frac{2\,t}{\pi}} \tag{3.3}$$

We see that for any *fixed* ε the probability of having $|w_t| \geq \varepsilon$ vanishes. A more refined analysis based on this observation proves that the Wiener process $w_t \equiv w(t,\omega)$ is:

1. continuous $\forall\, t \in \mathbb{R}_+$, **almost everywhere** with respect to $\omega \in \Omega$ (i.e. a part from a measure zero set);
2. non-differentiable $\forall\, t \in \mathbb{R}_+$, **almost everywhere** with respect to $\omega \in \Omega$.

An indication of non-differentiability comes from the evaluation of the expected value of the absolute value of the incremental ratio of the Wiener process:

$$\mathrm{E}\frac{|w_t|}{t} = 2\int_{0}^{\infty} dx\,\frac{x}{t}\,\frac{e^{-\frac{x^2}{2\,t}}}{\sqrt{2\,\pi\,t}} = \sqrt{\frac{2}{\pi\,t}} \tag{3.4}$$

which diverges as t tends to zero.

4 Terminology: filtration and process adapted to a filtration

Let I an ordered set with respect to the $\leq$ binary operation. This means that all pair elements of in the set are satisfy

1. *reflexivity*: $a \leq a \, \forall \, a \in I$
2. *antisymmetry*: $a \leq b \;\&\; b \leq a \;\Rightarrow\; a = b \;\forall\; a, b \in I$
3. *transitivity*: $a \leq b \;\&\; b \leq c \;\Rightarrow\; a \leq b \;\forall\; a, b, c \in I$

In practice this is technical jargon used in [4] to say "take I either $I \subset \mathbb{N}$ or $I \subset \mathbb{R}_+$". We have then the following definition

Definition 4.1 (*Filtration*)**.** *A family $\boldsymbol{\mathcal{F}}_I = \{\mathcal{F}_t \,:\, \forall t \in I\}$ of σ-algebras is called a filtration if*

$$\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \quad if \quad t_1 \leq t_2 \;\forall t_1, t_2 \in I \tag{4.1}$$

Note that $\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$ means that any event which is in $\mathcal{F}_{t_1}$ is also in $\mathcal{F}_{t_2}$. If we think of t as a "time" variable this means that as t increases more and more detailed information becomes available about events occurring in the sample space Ω. The description of this events calls for finer and finer partitions of Ω which is why the condition (4.1) must hold. Let

$$\xi_t \colon \Omega \times I \mapsto \mathbb{R} \tag{4.2}$$

be a stochastic process, we say that

Definition 4.2. *$\xi_t, t \in I$ is adapted to a filtration $\boldsymbol{\mathcal{F}}_I$ if for all t in I, ξ_t is $\mathcal{F}_t$-measurable, that is, if for any t, $\mathcal{F}_t$ contains all the information about ξ_t (and may contain extra information)*

In other words, ξ_t is $\mathcal{F}_t$-measurable if for any event $\{\xi_t \in A\}$ there is a map ϕ such that A is the image of an event $B \in \mathcal{F}_t$. This also mean that $\{\xi_t \in A\}$ does not depend upon events occurring at later "times" i.e. is independent of the future. In this sense the process is said *non-anticipating*.

Example 4.1. Let w_t a Wiener process for all $t \geq 0$, the function

$$f(t) = \begin{cases} 0 \quad \text{if} & \max\limits_{0 \leq s \leq t} w_s \leq 1 \\ 1 \quad \text{if} & \max\limits_{0 \leq s \leq t} \; > 1 \end{cases}$$

is *non-anticipative* as it depends on the Wiener process up to the time t when the function is evaluated. On the other hand for any $T > t$ the function

$$g(t) = \begin{cases} 0 \quad \text{if} & \max\limits_{0 \leq s \leq T} w_s \leq 1 \\ 1 \quad \text{if} & \max\limits_{0 \leq s \leq T} w_s > 1 \end{cases}$$

is *anticipative* as it depends on realizations of the Wiener process for times s posterior to the sampling time t.

5 Markov processes and Chapman-Kolmogorov equation

A special important class of non-anticipating stochastic processes are Markov processes

Definition 5.1. *Let $\mathcal{F}^{\xi}_{[0,t]}$ the filtration generated by a stochastic process ξ_t. Then ξ_t is Markov if for any $t \geq s$ and any event A*

$$P\left(\xi_t \in A|\boldsymbol{\mathcal{F}}^{\xi}_{[0,s]}\right) = P\left(\xi_t \in A|\mathcal{F}^{\xi}_{s}\right) = P\left(\xi_t \in A|\xi_s\right) \tag{5.1}$$

This means that the state of the system at time s fully specify further evolution independently of what happened before for times smaller s. The system has no "memory". In particular if the transition probability p density of the Markov process $\xi_t : \Omega \times I \mapsto \mathbb{R}$ is available

$$P(\xi_t \in A|\xi_s = y) = \int_A dx\, p(x, t\,|\, y, s) \tag{5.2}$$

then the definition of Markov process implies that the Chapman-Kolmogorov equation

$$p(x_2, t_2\,|\,x_0, t_0) = \int_{\mathbb{R}} dx_1\, p(x_2, t_2\,|\,x_1, t_1)\, p(x_1, t_1\,|\,x_0, t_0) \tag{5.3}$$

must hold true for any (x_2, x_0) and for any t_i, $i = 0, 1, 2$.

6 Functions of finite variation

Let $I = [a, b] \in \mathbb{R}$ and

$$f : I \mapsto \mathbb{R} \tag{6.1}$$

a deterministic function.

Definition 6.1. *We call the variation of f the quantity*

$$V_{[f]}(I) = \sup_{\mathsf{p}} \sum_{t_i \in \mathsf{p}} |f(t_{i+1}) - f(t_i)| \tag{6.2}$$

whether the sup is taken over the partitions $\{\mathsf{p}\}$ (see appendix A)) of I

By triangular inequality we also have

$$V_{[f]}(I) = \lim_{|\mathsf{p}|\downarrow 0} \sum_{t_i \in \mathsf{p}} |f(t_{i+1}) - f(t_i)| \tag{6.3}$$

for p the mesh of the partition.

Proposition 6.1. *If $f \in C^1(I)$ and*

$$\int_I dt\, \left|\frac{df}{dt}\right|(t) < \infty \tag{6.4}$$

then

$$V_{[f]}(I) = \int_I dt\, \left|\frac{df}{dt}\right|(t) \tag{6.5}$$

Proof. Let

$$\dot{f} = \frac{df}{dt} \tag{6.6}$$

the chain of equalities holds

$$\begin{aligned} V_{[f]}(I) &= \lim_{|\mathsf{p}|\downarrow 0} \sum_{t_i \in \mathsf{p}} |f(t_{i+1}) - f(t_i)| = \lim_{|\mathsf{p}|\downarrow 0} \sum_{t_i \in \mathsf{p}} |\dot{f}(t_i)|(t_{i+1} - t_i) \\ &= \lim_{|\mathsf{p}|\downarrow 0} \sum_{t_i \in \mathsf{p}} \left| \int_{t_i}^{t_{i+1}} dt\, \dot{f}(t) \right| = \sup_{\mathsf{p}} \sum_{t_i \in \mathsf{p}} \left| \dot{f}(t_i) \right| (t_{i+1} - t_i) = \int_I dt \left| \frac{df}{dt} \right| (t) \end{aligned} \tag{6.7}$$

thus proving the claim. □

Thus, trajectories solutions of ordinary differential equations

$$\dot{x}_t = v(x_t) \tag{6.8}$$

with v sufficiently smooth as customary in applications are functions of finite variation. A larger class of functions includes those with discontinuities.

Definition 6.2. *A function $f: I \mapsto \mathbb{R}$* **continuous from the right**

$$\lim_{t \downarrow t_o} f(t) = f(t_o) \tag{6.9}$$

with **limit from the left**

$$\lim_{t \uparrow t_o} f(t) = f(t_o -) \tag{6.10}$$

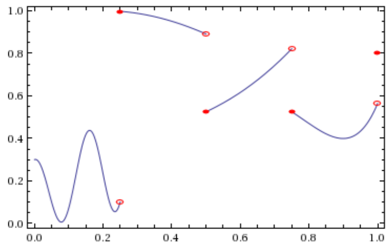
is called a **CADLAG** *function (French: Countinue à Droite Limite à Gauche).*

The difference

$$J_{[f]}(t) = f(t) - f(t-) \tag{6.11}$$

is called a **jump**. A theorem from analysis guarantees that a function defined on an interval $[a, b]$ can have no more than countably many jumps (see e.g. [3] and refs therein).

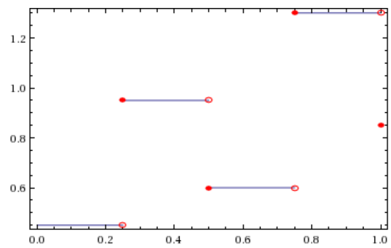
a CADLAG function:
circles denote the limit
dots continuity at jump locations.



If a cadlag function varies only at jump locations $\mathcal{J} \in I$ we have

$$f(t) = \sum_{s \in \mathcal{J}} J_{[f]}(s)$$

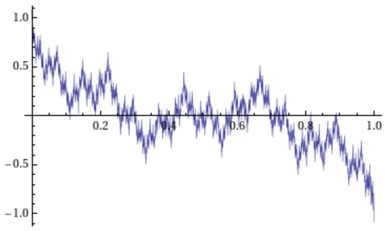
$$V_{[f]}(I) = \sum_{s \mathcal{J}} |J_{[f]}(s)| \tag{6.12}$$



7 Quadratic (co)-variation

In section 3.2 we argued that the Wiener process w_t is everywhere continuous and non-differentiable in t. This phenomenon is not related to randomness but it may occur also for deterministic functions. An example is the *Weierstrass function* $W_t \colon \mathbb{R} \mapsto \mathbb{R}$

$$W_t := \sum_{n=0}^{\infty} a^n \cos(b^n t) \qquad \text{for} \qquad \begin{array}{l} 0 < a < 1 \\ b \in 2\,\mathbb{N}+1 \\ a\,b > 1 \end{array} \tag{7.1}$$



G.H. Hardy [2] proved that the Weierstrass is everywhere continuous and non-differentiable. One can get an intuition of the reason observing that

1. a sequence of continuous functions (i.e. approximations by finite sum) uniformly converging admits as a limit a continuous function;
2. differentiating individual addends in the series one obtains

$$-\sum_{n=0}^{\infty} a^n\, b^n\, \sin(b^n\, t) \tag{7.2}$$

   which is a diverging series.

Hardy also showed that

$$|W_{t+h} - W_t| \leq C\, h^{\alpha} \qquad \alpha = -\frac{\ln a}{\ln b} \tag{7.3}$$

A further consequence is that the Weierstrass function is not of finite variation. It makes sense to consider functions of *finite second variation*:

Definition 7.1 (*Quadratic (co-)variation*)**.** *Let* $\xi_t \colon I \mapsto \mathbb{R}$ *and* $\chi_t \colon I \mapsto \mathbb{R}$ *the limit*

$$V_{[\xi\,,\chi]}(I) = \lim_{|\mathsf{p}_{(n)}|\downarrow 0} \sum_{t_k \in \mathsf{p}_n} (\xi_{t_k} - \xi_{t_{k-1}})(\chi_{t_k} - \chi_{t_{k-1}})$$

is called the **quadratic co-variation** *of the processes. In particular*

$$V_{[\xi\,,\xi]}(I) = \lim_{|\mathsf{p}_{(n)}|\downarrow 0} \sum_{t_k \in \mathsf{p}_n} (\xi_{t_k} - \xi_{t_{k-1}})^2$$

is called the **quadratic variation** *of* ξ_t*.*

Finiteness of the quadratic variation is possible only if the (first) variation of a function diverges. Namely

$$\sum_k [f(t_k) - f(t_{k-1})]^2 \leq \max_k |f(t_k) - f(t_{k-1})| \sum_k |f(t_k) - f(t_{k-1})|$$

so that for a differentiable function

$$\sum_k [f(t_k) - f(t_{k-1})]^2 \leq \max_k |f(t_k) - f(t_{k-1})| \int_0^t dt\, |f'(t)| \overset{\max_k (t_k - t_{k-1}) \to 0}{\longrightarrow} 0$$

In the case of the Wiener process the finiteness of the right hand side implies

$$\sum_k |f(t_k) - f(t_{k-1})| \overset{\max_k (t_k - t_{k-1}) \to 0}{\uparrow} \infty$$

7.1 Quadratic variation of the Wiener process

In the case of the Wiener process we have

Proposition 7.1 (*Quadratic variation of the B.M.*)**.** *The quadratic variation of the Brownian motion in* $[0, t]$ *for any* $t \in \mathbb{R}_+$

$$V_{[w\,,w]}(I) \;= t$$

in the sense of $\mathbb{L}^2(\Omega)$.

Proof. By direct calculation we know that

$$\mathrm{E} w_t^2 = t \tag{7.4}$$

Let p_n a partition paving $[0, t]$ with n sub-intervals:

$$Q_n := \sum_{\mathsf{p}_n} (w_{t_k} - w_{t_{k-1}})^2$$

we have then

$$\mathrm{E}\,(Q_n - t)^2 = \mathrm{E} \sum_{k\,l \in \mathsf{p}_n} \left[(w_{t_k} - w_{t_{k-1}})^2 - (t_k - t_{k-1}) \right] \left[(w_{t_l} - w_{t_{l-1}})^2 - (t_l - t_{l-1}) \right]$$

For non-overlapping intervals, the averaged quantities are independent random variables with zero average. The only contributions to the sum come from overlapping intervals:

$$\mathrm{E}(Q_n - t)^2 = \mathrm{E} \sum_{k \in \mathsf{p}_n} \left[(w_{t_k} - w_{t_{k-1}})^2 - (t_k - t_{k-1}) \right]^2 = 2 \sum_{k \in \mathsf{p}_n} (t_k - t_{k-1})^2$$

whence

$$\mathrm{E}(Q_n - t)^2 \leq 2\, t \max_{k \in \mathsf{p}_n} (t_k - t_{k-1}) \overset{\max_{k \in \mathsf{p}_n}(t_k - t_{k-1}) \downarrow 0}{\rightarrow} 0$$

□

The finite value of the quadratic variation motivates the estimate

$$d w_t \sim O(\sqrt{dt})$$

for typical increments of the Wiener process.

8 Differential calculus for functions of finite quadratic variation

Suppose $x_t \colon\; I \mapsto \mathbb{R}$ is a CADLAG function of finite quadratic variation. Then we have

$$V_{[x\,,x]}(I) = V_{[x\,,x]}^c(I) + \sum_{s \in \mathcal{J}} J_{[x]}^2(s) \tag{8.1}$$

where $V_{[x\,,x]}^c(I)$ is the quadratic variation of the continuous part of x and $\mathcal{J}$ denotes the set of the jump locations over I. The following proposition shows that the jump component of the quadratic variation defines the atomic part of the measure defined by $V_{[x\,,x]}(I)$. In other words, $V^c_{[x\,,x]}(I)$ defines a measure absolutely continuous with respect to the Lebsegue measure

$$dV^c_{[x\,,x]}((0,t]) = g(x_t)dt \tag{8.2}$$

for some positive definite

$$g\colon \mathbb{R} \mapsto \mathbb{R}_+ \tag{8.3}$$

whilst

$$d\sum_{s\in\mathcal{J}} J^2_{[x]}(s) = \sum_{s\in\mathcal{J}} dt\, r_i \delta(t-s) \tag{8.4}$$

for some $r_i \in \mathbb{R}_+$

Proposition 8.1. *Let $\{\mathsf{p}_n\}_{n=0}^{\infty}$ a sequence of partitions of the interval I. For any continuous function*

$$f\colon I \mapsto \mathbb{R} \tag{8.5}$$

the limit

$$\lim_{n\uparrow\infty} \sum_{t\geq t_i\in\mathsf{p}_n} f(x_{t_i})\,(x_{t_{i+1}} - x_{t_i})^2 = \int_{(0,t)} dV_{[x\,,x]}((0,s])\, f(x_{s^-}) \tag{8.6}$$

Proof. Let C be the countable set of points in I where x_t performs jumps of size strictly larger than $O(\varepsilon^2)$ for any arbitrary $\varepsilon > 0$. Let also z_t be the distribution function of finite-size jumps in I

$$z_t = \sum_{s\in C\cap(0,t]} (x_s - x_{s^-}) \tag{8.7}$$

we have

$$\lim_{n\uparrow\infty} \sum_{t\in\mathsf{p}_n} f(x_{t_i})\,(z_{t_{i+1}} - z_{t_i})^2 = \sum_{t\in C\cap(0,s]} f(x_{t^-})\,(x_t - x_{t^-})^2 \tag{8.8}$$

Let now y be the discrete measure such that

$$\sum_{t_i\in\mathsf{p}_n\cap(0,t]} (x_{t_{i+1}} - x_{t_i})^2 = \sum_{t_i\in\mathsf{p}_n\cap(0,t]} (y_{t_{i+1}} + z_{t_{i+1}} - y_{t_i} + z_{t_i})^2 \tag{8.9}$$

i.e. the discrete approximant of the absolute continuous part of the measure defined by the quadratic variation of x_t. Then

$$\sum_{t_i\in\mathsf{p}_n\cap(0,t]} (x_{t_{i+1}} - x_{t_i})^2 = \\ \sum_{t_i\in\mathsf{p}_n\cap(0,t]} (y_{t_{i+1}} - y_{t_i})^2 + \sum_{t_i\in\mathsf{p}_n\cap(0,t]} (z_{t_{i+1}} - z_{t_i})^2 + 2\sum_{t_i\in\mathsf{p}_n\cap(0,t]} (z_{t_{i+1}} - z_{t_i})(y_{t_{i+1}} - y_{t_i}) \tag{8.10}$$

By definition of y the third term of the right hand side converges to zero and the measure associated to y weakly converges to a measure the atomic part thereof only comprises jumps of size less equal $O(\varepsilon^2)$. It follows that

$$\limsup_n \left| \lim_{n\uparrow\infty} \sum_{t\in\mathsf{p}_n} f(x_{t_i})\,(y_{t_{i+1}} - y_{t_i})^2 - \int_{(0,t)} dV^c_{[x\,,x]}((0,s])\, f(x_{s^-}) \right| \leq O(\varepsilon^2) \tag{8.11}$$

which proves the claim □

Föllmer [1] proved the following theorem, which we reproduce here in abridged form.

Theorem 8.1. *Let $x_t \colon I \mapsto \mathbb{R}$ be a CADLAG function of finite quadratic variation and $F \in C^2(\mathbb{R})$. Then the Itô formula*

$$\begin{aligned} F(x_t) - F(x_0) &= \int_0^t dx_s\,(\partial_x F)(x_s-) \\ &+ \frac{1}{2}\int_0^t d\,V^c_{[x\,,x]}([0,s])\,(\partial_x^2 F)(x_s) + \sum_{s\in\mathcal{J}}\big[F(x_s) - F(x_s-) - (\partial_x F)(x_s-)J_{[x]}(s)\big] \end{aligned} \tag{8.12}$$

holds with

$$\int_0^t dx_s\,(\partial_x F)(x_s-) = \lim_{|\mathsf{p}_n|\downarrow 0}\sum_{t_k\in\mathsf{p}_n}(x_{t_{k+1}} - x_{t_k})\,(\partial_x F)(x_{t_k}) \tag{8.13}$$

holds true and the series in (8.13) is absolutely convergent.

Proof. Since x_t is not of finite variation the difficulty of the proof is to prove (8.13). The strategy of the proof is to prove the absolute convergence of the left hand side and of the other terms on the right hand side. As far as the left hand side is concerned, by hypothesis x_t is continuous to the right so guaranteeing the convergence

$$F(x_t) - F(x_0) = \lim_{|\mathsf{p}_n|\downarrow 0}\sum_{t_k\in\mathsf{p}_n}[F(x_{t_{k+1}}) - F(x_{t_k})] \tag{8.14}$$

since the addends give rise to alternating sums. Let us now distinguish two cases

1. Suppose now that x_t is continuous. Then by Taylor's formula

$$F(x_{t_{k+1}}) - F(x_{t_k}) = (x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k}) + \frac{(x_{t_{k+1}} - x_{t_k})^2}{2}(\partial_x F)(x_{\tilde{t}_k}) \tag{8.15}$$

for some $\tilde{t}_k \in (t_k, t_{k+1})$. We can rewrite the equality as

$$\begin{aligned} F(x_{t_{k+1}}) - F(x_{t_k}) &= (x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k}) \\ &+ \frac{(x_{t_{k+1}} - x_{t_k})^2}{2}(\partial_x F)(x_{t_k}) + \frac{(x_{t_{k+1}} - x_{t_k})^2}{2}[(\partial_x F)(x_{\tilde{t}_k}) - (\partial_x F)(x_{t_k})] \end{aligned} \tag{8.16}$$

We then have

$$\begin{aligned} &\lim_{|\mathsf{p}_n|\downarrow 0}\sum_{t_k\in\mathsf{p}_n}(x_{t_{k+1}} - x_{t_k})^2[(\partial_x F)(x_{\tilde{t}_k}) - (\partial_x F)(x_{t_k})] \\ &\qquad \le \lim_{|\mathsf{p}_n|\downarrow 0}\max_{t_k\in\mathsf{p}_n}[(\partial_x F)(x_{\tilde{t}_k}) - (\partial_x F)(x_{t_k})]\sum_{t_k\in\mathsf{p}_n}(x_{t_{k+1}} - x_{t_k})^2 \to 0 \end{aligned} \tag{8.17a}$$

$$\lim_{|\mathsf{p}_n|\downarrow 0}\sum_{t_k\in\mathsf{p}_n}(x_{t_{k+1}} - x_{t_k})^2\,(\partial_x F)(x_{t_k}) \le \lim_{|\mathsf{p}_n|\downarrow 0}\max_{t_k\in\mathsf{p}_n}(\partial_x F)(x_{t_k})\sum_{t_k\in\mathsf{p}_n}(x_{t_{k+1}} - x_{t_k})^2 \,<\, \infty \tag{8.17b}$$

where (8.17a) holds because of the continuity of x_t and F and (8.17b) since F is continuous over a finite closed interval. Gleaning the information provided (8.14) and (8.17a)-(8.17b) we conclude that the claim (8.13) must hold true.

2. Let us now turn to the general case of a CADLAG function. Let $\varepsilon > 0$ We divide the jumps of x_t on $[0,t]$ into two classes:

   (a) $C_1 \equiv C_1(\varepsilon, t)$ with jumps of finite size;

   (b) $C_2 \equiv C_2(\varepsilon, t)$ such that $\sum_{s\in C_2} J^2_{[x]}(s) \leq \varepsilon^2$.

   We then write

$$\sum_{t_k\in \mathsf{p}_n} \left[F(x_{t_{k+1}}) - F(x_{t_k})\right] = \sum_1 \left[F(x_{t_{k+1}}) - F(x_{t_k})\right] + \sum_2 \left[F(x_{t_{k+1}}) - F(x_{t_k})\right] \tag{8.18}$$

   where $\sum_1$ indicates the summation over those $t_k \in \mathsf{p}_n$ for which the interval $]t_k, t_{k+1}]$ contains a jump of class C_1. Clearly we have

$$\lim_{|\mathsf{p}_n|\downarrow 0} \sum_1 \left[F(x_{t_{k+1}}) - F(x_{t_k})\right] = \sum_{s\in\mathcal{J}} [F(x_s) - F(x_{s}-)] \tag{8.19}$$

   On the other hand we can apply Taylor's formula to write

$$\begin{aligned}\sum_2 \left[F(x_{t_{k+1}}) - F(x_{t_k})\right] &= \sum_{t_k\in\mathsf{p}_n} \left[(x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k}) + \frac{(x_{t_{k+1}} - x_{t_k})^2}{2}(\partial_x^2 F)(x_{t_k})\right] \\ &- \sum_1 \left[(x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k}) + \frac{(x_{t_{k+1}} - x_{t_k})^2}{2}(\partial_x^2 F)(x_{t_k})\right] + \frac{1}{2}\sum_2 (x_{t_{k+1}} - x_{t_k})^2 R(\tilde{t}_k, t_k)\end{aligned} \tag{8.20}$$

   with

$$R(\tilde{t}_k, t_k) := \left[(\partial_x F)(x_{\tilde{t}_k}) - (\partial_x F)(x_{t_k})\right] \tag{8.21}$$

   Furthermore

$$\sum_{t_k\in\mathsf{p}_n} (x_{t_{k+1}} - x_{t_k})^2 (\partial_x F)(x_{t_k}) = \left(\sum_1 + \sum_2\right)(x_{t_{k+1}} - x_{t_k})^2 (\partial_x F)(x_{t_k}) \tag{8.22}$$

   we have

$$\begin{aligned}\sum_2 \left[F(x_{t_{k+1}}) - F(x_{t_k})\right] &= \sum_{t_k\in\mathsf{p}_n} (x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k}) \\ &+ \sum_2 (x_{t_{k+1}} - x_{t_k})^2 \left[(\partial_x^2 F)(x_{t_k}) + R(\tilde{t}_k, t_k)\right] - \sum_1 (x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k})\end{aligned} \tag{8.23}$$

   Let us analyze the terms occurring in (8.20) separately.

   (a) Since $\sum_2$ does not contain jumps of finite size

$$\lim_{|\mathsf{p}_n|\downarrow 0} \sum_2 (x_{t_{k+1}} - x_{t_k})^2 R(\tilde{t}_k, t_k) = 0 \tag{8.24}$$

   for the same reasons put forward in the continuous x_t case. Similarly

$$\lim_{|\mathsf{p}_n|\downarrow 0} \sum_2 (x_{t_{k+1}} - x_{t_k})^2 (\partial_x^2 F)(x_{t_k}) = \int_0^t d\, V^c_{[x\,,x]}([0,s])\, (\partial_x^2 F)(x_s) \tag{8.25}$$

(b) The limit of the sum including jumps is dominated by these latter ones

$$\lim_{|\mathsf{p}_n|\downarrow 0} \sum_1 (x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k}) = \sum_{s\in\mathcal{J}} J_{[x]}(s)(\partial_x F)(x_s-) \tag{8.26}$$

We have then

$$\begin{aligned} F(x_t) - F(x_o) &= \lim_{|\mathsf{p}_n|\downarrow 0} \sum_{t_k\in\mathsf{p}_n} [F(x_{t_{k+1}}) - F(x_{t_k})] = \lim_{|\mathsf{p}_n|\downarrow 0} \sum_{t_k\in\mathsf{p}_n} (x_{t_{k+1}} - x_{t_k})(\partial_x F)(x_{t_k}) \\ &+ \frac{1}{2}\int_0^t d\,V^c_{[x\,,x]}([0,s])\,(\partial_x^2 F)(x_s) + \sum_{s\in\mathcal{J}} \left[F(x_s) - F(x_s-) - (\partial_x F)(x_s-)J_{[x]}(s)\right] \end{aligned} \tag{8.27}$$

where Taylor's formula also guarantees that

$$\sum_{s\in\mathcal{J}} \left[F(x_s) - F(x_s-) - (\partial_x F)(x_s-)J_{[x]}(s)\right] \leq C \sum_{s\in\mathcal{J}} J^2_{[x]}(s) \tag{8.28}$$

Thus the last two term on the right hand side of (8.27) denote a finite limit of the approximating sums. This observation yields the claim and concludes the proof.

□

Observation if x_t is CADLAG of finite (first) variation we have

$$F(x_t) - F(x_0) = \int_0^t ds\,(\partial_x F)(x_s-) + \sum_{s\in\mathcal{J}} [F(x_s) - F(x_s-)] \tag{8.29}$$

as an example consider

$$x_t = \begin{cases} t & t\in[0,1/2) \\ t+1 & t\in[1/2,1] \end{cases} \tag{8.30}$$

If we take for F the identity map, the integral on the right hand side of (8.29) over $[0,t]$ with $t\leq 1$ yields

$$\int_0^t ds = t \tag{8.31}$$

thus we see that the jump term accounts for the unit addend guaranteeing the identity between the two sides of (8.29).

Appendix

A Partitions

Definition A.1 (*Partition*)**.** *If $I = [x_-\,,x_+] \subset \mathbb{R}$ is an interval a partition p (subdivision) of I is a finite sequence $\{x_k\}_{k=1}^n$ of points in I such that*

$$x_- = x_1 < \ldots < x_n = x_+$$

Definition A.2 (*Mesh of a partition*)**.** *The mesh size of a partition p of an interval $I = [x_-\,,x_+]$ is*

$$|\mathsf{p}| = \max_{1\leq k\leq n} |x_{k+1} - x_k|$$

Definition A.3 (*Refinement of a partition*)**.** *The refinement of a partition p of the interval I is another partition p' that contains all the points from p and* some additional points*, again sorted by order of magnitude.*

References

[1] H. Föllmer. Calcul ditô sans probabilités. In *Séminaire de probabilités XV (Strasbourg)*, volume 850 of *Spinger Lecture Notes in Mathematics*, pages 143–150. Springer, 1981.

[2] G. H. Hardy. Weierstrass's non-differentiable function. *Transactions of the American Mathematical Society*, 17(3):301–325, July 1916.

[3] F. C. Klebaner. *Introduction to stochastic calculus with applications*. Imperial College Press, 2 edition, 2005.

[4] G. A. Pavliotis and A. M. Stuart. *Multiscale methods: averaging and homogenization*, volume 53 of *Texts in applied mathematics*. Springer, 2008.

[5] D. Sondermann. *Introduction to stochastic calculus for finance: a new didactic approach*, volume 579 of *Lecture Notes in Economics and Mathematical Systems*. Springer, 2006.