Available online at www.sciencedirect.com
SciVerse ScienceDirect
Stochastic Processes and their Applications 123 (2013) 446–474
stochastic processes and their applications
www.elsevier.com/locate/spa

On truncated variation, upward truncated variation and downward truncated variation for diffusions

Rafał M. Łochowski [a,b], Piotr Miłoś [c,*]

[a] *Department of Mathematics and Mathematical Economics, Warsaw School of Economics, Madalińskiego 6/8, 02-513 Warszawa, Poland*
[b] *African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg 7945, South Africa*
[c] *Faculty of Mathematics, Informatics, and Mechanics, Banacha 2, 02-097 Warszawa, Poland*

Received 31 August 2011; received in revised form 17 April 2012; accepted 22 August 2012
Available online 27 August 2012

Abstract

The truncated variation, TV^c, is a fairly new concept introduced in Łochowski (2008) [5]. Roughly speaking, given a càdlàg function f, its truncated variation is "the total variation which does not pay attention to small changes of f, below some threshold $c > 0$". The very basic consequence of such approach is that contrary to the total variation, TV^c is always finite. This is appealing to the stochastic analysis where so-far large classes of processes, like semimartingales or diffusions, could not be studied with the total variation. Recently in Łochowski (2011) [6], another characterization of TV^c has been found. Namely TV^c is the smallest possible total variation of a function which approximates f uniformly with accuracy $c/2$. Due to these properties we envisage that TV^c might be a useful concept both in the theory and applications of stochastic processes.

For this reason we decided to determine some properties of TV^c for some well-known processes. In course of our research we discover intimate connections with already known concepts of the stochastic processes theory.

First, for semimartingales we proved that TV^c is of order c^{-1} and the normalized truncated variation converges almost surely to the quadratic variation of the semimartingale as $c \searrow 0$. Second, we studied the rate of this convergence. As this task was much more demanding we narrowed to the class of diffusions (with some mild additional assumptions). We obtained the weak convergence to a so-called Ocone martingale. These results can be viewed as some kind of law of large numbers and the corresponding central limit theorem.

Finally, for a Brownian motion with a drift we proved the behavior of TV^c on intervals going to infinity. Again, we obtained a LLN and CLT, though in this case they have a different interpretation and were easier to prove.

* Corresponding author. Tel.: +48 600973193, +48 22 55 44 122; fax: +48 22 55 44 300.
E-mail addresses: rlocho@sgh.waw.pl (R.M. Łochowski), pmilos@mimuw.edu.pl (P. Miłoś).

doi:10.1016/j.spa.2012.08.007

All the results above were obtained in a functional setting, viz. we worked with processes describing the growth of the truncated variation in time. Moreover, in the same respect we also treated two closely related quantities—the so-called upward truncated variation and downward truncated variation.

Keywords: Stochastic processes; Semimartingales; Diffusions; Truncated variation; Total variation

1. Introduction and results

Recently, the following notion of *truncated variation* has been introduced in [5]:

$$\mathrm{TV}^c(f, [a; b]) := \sup_n \sup_{a \le t_1 < t_2 < \cdots < t_n \le b} \sum_{i=1}^{n-1} \phi_c\left(|f(t_{i+1}) - f(t_i)|\right), \tag{1.1}$$

where $\phi_c(x) = \max\{x - c, 0\}$, $c \ge 0$ and $f : [a; b] \mapsto \mathbb{R}$ is a càdlàg function. The trivial observation is that TV^0 is nothing else than the total variation (which will also be denoted by TV). The introduction of the truncation parameter c makes it possible to circumvent a classical problem of stochastic analysis; namely, that the total variation of the Brownian motion as well as of a 'non-trivial' diffusion process is almost surely infinite. This alone makes TV^c an interesting research object. Other properties of TV^c were found, amongst which the variational characterization of the truncated variation, which is given by

$$\mathrm{TV}^c(f, [a; b]) = \inf\left\{\mathrm{TV}(g, [a; b]) : g \text{ such that } \|g - f\|_\infty \le \frac{1}{2}c\right\}, \tag{1.2}$$

where $\|g\|_\infty := \sup\{|g(x)| : x \in [a; b]\}$. In other words, truncated variation is the lower bound for the total variation of functions approximating f with accuracy c. It appears that the inf in the above expression is attained at some function g^c. The properties just listed give hope that TV^c could be used in the stochastic analysis. This question is an active field of research, some promising results are contained in [7], like definition of a stochastic integral with respect to a semimartingale as a limit of the pathwise Riemann–Stieltjes stochastic integrals, and others are being investigated. A detailed description would be too vast for our introduction; therefore we refer the reader to [7,6], and its debriefing in Section 2.

Having agreed that TV^c might be a useful tool, an important task is to describe the behavior of TV^c for a vast class of stochastic processes. This is the main aim of this paper. We will derive first order properties for *continuous semimartingales* and second order properties for *continuous diffusions* (under some mild technical assumptions) when $c \searrow 0$. Intuitively, these answer the question of how fast TV^c converges to the total variation, that is how fast it diverges to infinity. In the case of the Brownian motion with drift we will also study the behavior of TV^c on large time intervals.

Before presenting our results we define two concepts closely related to TV^c. The upward truncated variation given by

$$\mathrm{UTV}^c(f, [a; b]) := \sup_n \sup_{a \le t_1 < s_1 < t_2 < s_2 < \cdots < t_n < s_n \le b} \sum_{i=1}^{n} \phi_c\left(f(s_i) - f(t_i)\right), \tag{1.3}$$

and the downward truncated variation given by

$$\mathrm{DTV}^c(f,[a;b]) := \sup_n \sup_{a\le t_1<s_1<t_2<s_2<\cdots<t_n<s_n\le b} \sum_{i=1}^n \phi_c\left(f(t_i)-f(s_i)\right).$$

The relation between TV^c, UTV^c, DTV^c will become clear in Section 2.1. Given a càdlàg process $\{X_t\}_{t\ge 0}$ we define the following families of processes $\{\mathrm{TV}^c(X,t)\}_{t\ge 0}$, $\{\mathrm{UTV}^c(X,t)\}_{t\ge 0}$ and $\{\mathrm{DTV}^c(X,t)\}_{t\ge 0}$ by

$$\mathrm{TV}^c(X,t) := \mathrm{TV}^c(X,[0;t]), \qquad \mathrm{UTV}^c(X,t) := \mathrm{UTV}^c(X,[0;t]),$$
$$\mathrm{DTV}^c(X,t) := \mathrm{DTV}^c(X,[0;t]),$$

where all the above definitions are understood in a pathwise fashion. Obviously, all three processes are increasing. Moreover, for semimartingales and $c \searrow 0$, under weak non-degeneracy conditions, their values diverge up to infinity. Thus a natural question arises what the growth rate of the (upward, downward) truncated variation is? Under a proper normalization we expect also some convergence to a non-trivial object. These questions are answered in the following section.

1.1. Behavior as $c \searrow 0$. First order properties for continuous semimartingales

For a continuous semimartingale $\{X\}_{t\in[0;T]}$ we will denote its decomposition by

$$X_t := X_0 + M_t + A_t, \quad t \in [0;T],$$

where M is a continuous local martingale such that $M_0 = 0$ and A is a continuous finite variation process such that $A_0 = 0$. Given $T > 0$, by $\mathcal{C}([0;T],\mathbb{R})$ we denote the usual space of continuous functions on $[0;T]$ endowed with the topology given by norm $\|\cdot\|_\infty$.

Theorem 1. *Let $T > 0$ and let $\{X\}_{t\in[0;T]}$ be a continuous semimartingale as above. We have*

$$\lim_{c\searrow 0} cTV^c(X,t) \to \langle X\rangle_t, \quad a.s.$$
$$\lim_{c\searrow 0} cUTV^c(X,t) \to \langle X\rangle_t/2, \quad a.s.$$

and

$$\lim_{c\searrow 0} cDTV^c(X,t) \to \langle X\rangle_t/2, \quad a.s.$$

In all cases the convergence is understood in the $\mathcal{C}([0;T],\mathbb{R})$ topology.

Remark 2. One can see that TV^c is of order c^{-1}. Hence by the discussion above this is also the lower bound of the total variation of the approximation of X in $\|\cdot\|_\infty$-ball of radius $c/2$. For diffusions we will find finer estimates in the next section.

Assumptions of Theorem 1 could be weakened slightly. Without additional effort we can prove the theorem for A not being necessary continuous. This is however cumbersome from notational point of view, as we cannot work in $\mathcal{C}([0;T],\mathbb{R})$ space. The problem of non-continuous semimartingales will be treated in full extent in future papers.

Remark 3. Theorem 1 could be considered as some kind of a law of large numbers. We will now provide a rough justification using the Wiener process W as an example. One can imagine splitting an interval $[0;1]$ into c^{-2} parts. On each part W performs a motion of order c. The contribution of the part to the total truncated variations is not negligible and is of order c. The contributions are random and "almost" independent for non-neighboring parts. Therefore there is no randomness in the limit.

Remark 4. The heuristics presented in the previous remark is nice at the intuitive level; however a more precise description is required to perform the proof. In the case of a Wiener process with drift this will be a precise characterization of $t_1, t_2, \ldots$ for which the sup in definition (1.1) is attained, which will lead to a natural renewal structure. In the case of a general semimartingale following the same path seems to be hopeless. To circumvent the problem we employed an abstract approach based on time change techniques in spirit of the Dambis, Dubins–Schwarz theorem [11, Chapt. V, Theorem 1.6].

Having explained "the law of large numbers nature" of the above result a natural question arises about the corresponding central limit theorem. This will be addressed in the next section for $\{X_t\}_{t\geq 0}$ being a diffusion satisfying some mild conditions.

1.2. Behavior as $c \searrow 0$. Second order properties for diffusions

Let us now consider a general diffusion defined with equation

$$\mathrm{d}X_t = \sigma(X_t)\mathrm{d}W_t + \mu(X_t)\mathrm{d}t, \qquad X_0 = 0. \tag{1.4}$$

We will always assume that σ, μ are Lipschitz functions and $\sigma > 0$. It is well known, [11, Sect. IX.2], that under these conditions the equation admits a unique strong solution. The main result of this section is given below.

Theorem 5. *Let $T > 0$ then*

$$\left(X, UTV^c(X,t) - \frac{1}{2}\left(\frac{\langle X\rangle_t}{c} + X_t\right), DTV^c(X,t) - \frac{1}{2}\left(\frac{\langle X\rangle_t}{c} - X_t\right),\right.$$
$$\left. TV^c(X,t) - \frac{\langle X\rangle_t}{c}\right) \to^d (X, \tilde{M}_t, \tilde{M}_t, 2\tilde{M}_t), \quad \text{as } c \searrow 0, \tag{1.5}$$

where $\tilde{M}$ is given by the change time formula:

$$\tilde{M}_t := 12^{-1/2} B_{\langle X\rangle_t}, \tag{1.6}$$

where B is a standard Brownian motion such that B and X are independent. The convergence is understood as the weak convergence in $\mathcal{C}([0;T],\mathbb{R})^4$ topology.

Remark 6. Let us notice that by [3, Proposition 5.33] from the joint convergence of X and three other processes related to UTV, DTV and TV one obtains their stable convergence as described in [3, Sect. VIII.5].

Remark 7. Let us now present an intuitive explanation of the result on the example of a Wiener process with drift, W and the truncated variation. Theorem 1 reads as

$$c\mathrm{TV}^c(W,t) \to t, \quad \text{a.s.}$$

and by Theorem 5 and the fact that $\langle W\rangle_t = t$ we obtain

$$\mathrm{TV}^c(X,t) - \frac{t}{c} \to^d 3^{-1/2} B_t.$$

In this case the theorems are indeed an "almost classical" law of large numbers and central limit theorem. This stems from the fact that TV^c in this case has a particularly nice, renewal structure.

On the intuitive level, by (1.2) one may say that for any path of W on interval $[a; b]$, minimal "vertical" length of graph of any random function $f : [a; b] \to \mathbb{R}$, uniformly close to this path must be at least equal to

$$\frac{b-a}{c} + \sqrt{\frac{b-a}{3}} R_c,$$

where $c = 2\sup_{t\in[a;b]} |f(t) - W_t|$, and R_c is a random variable such that it tends in distribution to a standard normal distribution $\mathcal{N}(0, 1)$ as $c \searrow 0$. Note that for small c's this lower bound is almost deterministic.

Remark 8. It is easy to check that $\langle \tilde{M} \rangle = \langle X \rangle$. Let M be the local martingale in the semimartingale decomposition of X. It is natural to ask how the laws of M and $\tilde{M}$ are related. The martingales of the form given by (1.6) were introduced in [9] and are called Ocone martingales. By results of [15] it follows that M is an Ocone martingale only if $\sigma = const$ (i.e. X is a Brownian motion with some stochastic drift).

Let us also notice that $\sigma = const$ is also the only case when $\langle X \rangle_t$ is a deterministic process.

Ocone martingales have particularly simple structure which sometimes makes it easy to draw conclusion about them. As an example we consider a situation when $\sigma \le C$. Then

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0;T]} \tilde{M}_t \ge a\right) &\le \mathbb{P}\left(\sup_{t\in[0;CT]} 12^{-1/2} B_t > a\right)\\ &= \mathbb{P}\left(\sup_{t\in[0;T]} B_t > (12/C)^{1/2} a\right);\end{aligned}$$

hence $\tilde{M}$ has a Gaussian concentration. Further properties and references can be found in [15].

Remark 9. The assumption $\sigma > 0$ is equivalent to $\sigma \neq 0$. This follows by the fact that σ is continuous so, under the assumption that $\sigma \neq 0$, either $\sigma > 0$ for any x or $\sigma < 0$. In the latter case one can simply take $-\sigma$ instead of σ and obtain a diffusion with the same law.

The case when σ may attain value 0 requires further studies. To see this let us consider "a very degenerate case" when $\sigma = 0$ on an interval $[x_0; x_1]$ for $x_0 < x_1$. For any $x \in (x_0; x_1)$ the diffusion degenerates locally to a deterministic process, a solution of an ordinary differential equation, with a bounded total variation. Hence the above formulation of the CLT does not make sense. While this case was relatively easier, the situation becomes more involved for border points x_0, x_1 or "isolated" 0's. We suspect that in such cases a non-trivial correction term containing the local time may be required.

Remark 10. Similarly as in the case of the law of large numbers (see Remark 4) the proof splits into technically different parts.

The first one deals with the Wiener process with drift $X_t = W_t + \mu t$. We use here the fact that $\mathrm{TV}^c(X, t)$ has a fairly simple renewal-like structure. Moreover, it is possible to derive explicit formulas for the Laplace transform of the increments of the truncated variation. Then a very simple argument allows to treat random drift, i.e. the case where μ is a random variable independent of W.

The second step deals with diffusions with $\sigma = const$. Namely, on a small interval we have $X_{\Delta t+t} - X_t \approx \sigma(W_{\Delta t+t} - W_t) + \mu(X_t)\Delta t := Y_{\Delta t}$ which is essentially a Wiener process with a random drift as above. It turns out that we may control the quality of the approximation to conclude the proof using some metric-theoretic tricks and the Prohorov metric in this case.

As explained in Remark 26, this approach fails in the case of non-constant σ. Here we appeal to a time change technique and a Rényi mixing-like argument (see e.g. [13, p. 309]). A reader familiar with this kind of reasoning may recognize that this is why we get the independence in (1.6).

1.3. Large time results

For the Wiener process with drift it is possible to derive results for large time. In this section, we put

$$X := W_t + \mu t,$$

First, we present the following fact.

Fact 11. *Let $T > 0$ and $c > 0$. We have*

$$\lim_{n\to+\infty} TV^c(X, nt)/n \to m^c_\mu t, \quad a.s.,$$

where the convergence is understood in $\mathcal{C}([0;T],\mathbb{R})$ topology and

$$m^c_\mu = \begin{cases} \mu\coth(c\mu) & \text{if } \mu \neq 0, \\ c^{-1} & \text{if } \mu = 0. \end{cases} \tag{1.7}$$

Analogously we have

$$\lim_{n\to+\infty} UTV^c(X, nt)/n \to \frac{1}{2} n^c_\mu t, \quad a.s.,$$

and

$$\lim_{n\to+\infty} DTV^c(X, nt)/n \to \frac{1}{2} n^c_{-\mu} t, \quad a.s.$$

where again the convergence is understood in $\mathcal{C}([0;T],\mathbb{R})$ topology and

$$n^c_\mu = \begin{cases} \mu\coth(c\mu) + \mu & \text{if } \mu \neq 0, \\ c^{-1} & \text{if } \mu = 0. \end{cases} \tag{1.8}$$

The quality of the above approximation is studied in the following.

Theorem 12. *Let $T > 0$ and $c > 0$. We have*

$$\frac{TV^c(X, nt) - m^c_\mu nt}{\sigma^c_\mu \sqrt{n}} \to^d B_t, \quad \text{as } n \to +\infty,$$

where $\to^d$ is understood as weak convergence in $\mathcal{C}([0;T],\mathbb{R})$ topology; m^c_μ is given by (1.7) and

$$\left(\sigma^c_\mu\right)^2 = \begin{cases} \dfrac{2 - 2c\mu\coth(c\mu)}{\sinh^2(c\mu)} + 1 & \text{if } \mu \neq 0, \\ 1/3 & \text{if } \mu = 0. \end{cases}$$

Theorem 13. *Let $T > 0$ and $c > 0$. We have*

$$\frac{UTV^c(X, nt) - \frac{1}{2}n^c_\mu nt}{\rho^c_\mu \sqrt{n}} \to^d B_t, \quad \text{as } n \to +\infty,$$

and

$$\frac{DTV^c(X, nt) - \frac{1}{2}n^c_{-\mu} nt}{\rho^c_\mu \sqrt{n}} \to^d B_t, \quad \text{as } n \to +\infty,$$

where $\to^d$ is understood as weak convergence in $\mathcal{C}([0; T], \mathbb{R})$ topology; n^c_μ is given by (1.8) and

$$\left(\rho^c_\mu\right)^2 = \begin{cases} \dfrac{2\exp(4c\mu)\,(\sinh(2c\mu) - 2c\mu)}{(\exp(2c\mu) - 1)^3} & \text{if } \mu \neq 0, \\ 1/3 & \text{if } \mu = 0. \end{cases}$$

Remark 14. Fact 11 could be considered as a kind of law of large numbers. Indeed, TV^c builds up over time (cf. Section 2.1) and because of the homogeneity of X its truncated variation can be decomposed into a number of independent increments. These increments are also square integrable; therefore Theorems 12 and 13 hold.

The task of proving analogous facts for more general classes of processes seems to be elusive at the moment. First, our methods failed in this case, but the reason seems to lie deeper than that. It is connected with the fact that the truncated variation depends on the paths in a rather complicated way, simplifying only when $c \searrow 0$. We suspect that it is possible to prove similar results for ergodic Markov processes. This however seem a little unsatisfactory as in this case the convergence stems merely from the fact that on distant intervals the process itself is nearly independent.

Remark 15. It is possible for the finite dimensional distributions of the normalized truncated variation processes appearing in Theorems 12 and 13 to obtain even stronger results, namely the Berry–Esséen-type estimates of the rate of convergence to the normal distribution. The straightforward way to obtain such estimates is to use the already mentioned cumulative structure of the truncated variation processes of a Brownian motion with drift and [12, Theorem 8.2]. One can check that the appropriate moments exist (see formula (3.15)) and observe that inter-renewal times in this case have the same distribution as the exit time of Brownian motion with drift from a strip. Thus we obtain that the difference between the cdf of the multidimensional projection of the limit distribution and the cdf of the finite dimensional distributions of the normalized truncated variation processes in Theorems 12 and 13 is of order $\log(n)/\sqrt{n}$. We suspect that the results of Theorem 5 can be strengthened in a similar way. This will be a subject of further studies.

Let us now comment on the structure of the paper. In the next section we gather facts about the truncated variation and discuss potential application to the theory of stochastic processes. Section 3 is devoted to the proof of Theorem 5. In Section 4 we present the proof of Theorem 1. Finally in Section 5 we sketch the proof of the large time results presented just above.

2. Properties of the truncated variation

This section is based on results of [6]. For reader's convenience we keep much of the notation introduced there.

Arguably the most interesting property of the TV^c was listed in (1.2). Another closely related property is given by

$$TV^c(f, [a; b]) = \inf\left\{TV(g, [a; b]) : g \text{ such that } \|g - f\|_{osc} \le c, g(a) = f(a)\right\}, \quad (2.1)$$

where $\|h\|_{osc} := \sup\{|h(x) - h(y)| : x, y \in [a; b]\}$. The infimum in (2.1) is attained for some $g^{0,c} : [a; b] \mapsto \mathbb{R}$, which is unique. Moreover, we also have the following explicit representation:

$$g^{0,c}(s) = f(a) + UTV^c(f, [a; s]) - DTV^c(f, [a; s]) \quad (2.2)$$

and

$$\|g^{0,c} - f\|_\infty \le c. \quad (2.3)$$

$g^{0,c}$ is also closely related to the solution of the problem stated in (1.2). Let us put $\alpha_0 := -\inf\left\{g^{0,c}(s) - f(s) : s \in [a; b]\right\} - \frac{1}{2}\|g^{0,c} - f\|_{osc}$. The function g^c for which inf in (1.2) is attained is given by

$$g^c(s) := \alpha_0 + g^{0,c}(s). \quad (2.4)$$

The problem posed by (2.1) seems a little artificial at first. Its formulation has however a substantial advantage over the problem of (1.2) when considered in stochastic setting. Namely, when working with stochastic processes the solution given by (2.2) is adaptable to the same filtration as the process itself while the solution obtained in (2.4) requires some "knowledge of future". We would like also to mention that condition $\|g - f\|_{osc} \le c$ in (2.1) implies that the increments of f are uniformly approximated by the increments of $g^{0,c}$ with accuracy c. This property might be useful for applications to numerical stochastic integration.

To give the reader some intuition about the functions introduced above we rephrase [6, Remark 2.4]: "g^c is the most lazy function possible, which changes its value only if it is necessary to stay in the tube defined by $\|g^c - f\|_\infty \le c/2$". This can be seen on the following picture (Fig. 2.1)

We hope that we convince the reader that the truncated variation is an interesting research object. Moreover, we hope that it will be useful both in the theory of stochastic processes and in their applications. The first step towards this goal was undertaken in [8,7] e.g. in [8] was calculated the Laplace transform of UTV^c and DTV^c for a Brownian motion with drift and in [7] are presented possible applications to the approximation of stochastic processes and stochastic integration.

We plan to report shortly on further findings.

2.1. *Joint structure of TV^c, UTV^c and DTV^c*

We will now describe the structure of TV^c, UTV^c and DTV^c. The construction is described in more details in [6, Section 2]. Let $-\infty < a < b < +\infty$ and let $f : [a; b] \to \mathbb{R}$ be a càdlàg function. For $c > 0$ let us assume that

$$T_U^c f := \inf\left\{s \ge a : \sup_{t\in[a;s]} f(t) - f(s) \ge c\right\}$$
$$\le T_D^c f := \inf\left\{s \ge a : f(s) - \inf_{t\in[a;s]} f(t) \ge c\right\} \quad (2.5)$$

i.e. the first upward jump of function f of size c appears before the first downward jump of the same size c or both times are infinite, i.e. there is no upward or downward jump of size c. Note

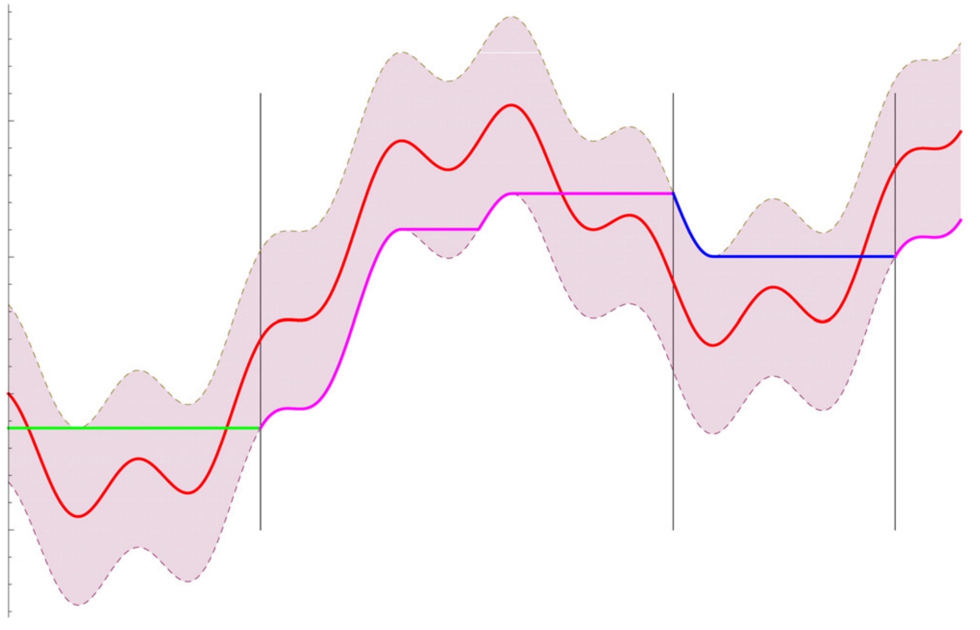

Fig. 2.1. An example of function (in red) and g^c (in various colors). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that when this condition fails one may simply consider function $-f$. Now we define sequences $\left(T^c_{U,k}\right)_{k=0}^{\infty}, \left(T^c_{D,k}\right)_{k=-1}^{\infty}$, in the following way: $T^c_{D,-1} = a$, $T^c_{U,0} = T^c_U f$ and for $k \geq 0$:

$$T^c_{D,k} := \inf\left\{ s \geq T^c_{U,k} : \sup_{t\in\left[T^c_{U,k};s\right]} f(t) - f(s) \geq c \right\},$$

$$T^c_{U,k+1} = \inf\left\{ s \geq T^c_{D,k} : f(s) - \inf_{t\in\left[T^c_{D,k};s\right]} f(t) \geq c \right\}.$$

Next let us define two sequences of non-decreasing functions $m^c_k : \left[T^c_{D,k-1}; T^c_{U,k}\right) \to \mathbb{R}$ and $M^c_k : \left[T^c_{U,k}; T^c_{D,k}\right) \to \mathbb{R}$ for $k \geq 0$ such that $T^c_{D,k-1} < \infty$ and $T^c_{U,k} < \infty$ respectively, with the formulas

$$m^c_k(s) := \inf_{t\in\left[T^c_{D,k-1};s\right]} f(t), \qquad M^c_k(s) = \sup_{t\in\left[T^c_{U,k};s\right]} f(t).$$

Similarly, let us define two finite sequences of real numbers $\left\{m^c_k\right\}$ and $\left\{M^c_k\right\}$, for such k's that $T^c_{D,k-1} < \infty$ and $T^c_{U,k} < \infty$ by

$$m^c_k := m^c_k\left(T^c_{U,k}-\right) = \inf_{t\in\left[T^c_{D,k-1};T^c_{U,k}\right)} f(t), \tag{2.6}$$

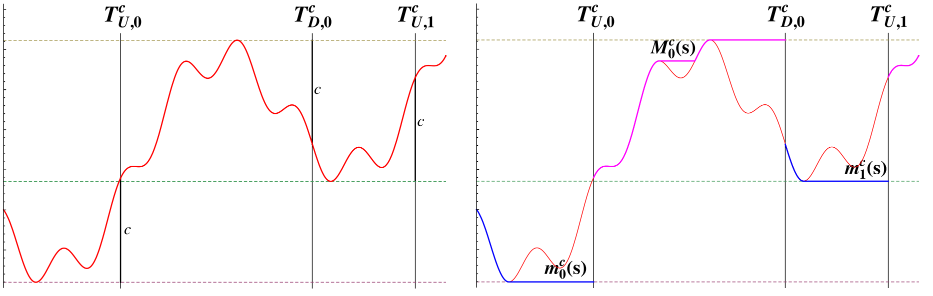

Fig. 2.2. Example of definition of $T^c_{U,k}$, $T^c_{D,k}$ and M^c_k, m^c_k.

$$M^c_k := M^c_k\left(T^c_{D,k}-\right) = \sup_{t\in\left[T^c_{U,k};T^c_{D,k}\right)} f(t). \tag{2.7}$$

The above definitions are simple however may be hard to read without pictures. We hope the following will be helpful. Note that we present the same function as in the previous example (Fig. 2.2).

The main result of this section is (cf. [6, Theorem 2.3]) the following.

Theorem 16. *For any càdlàg function $f : [a; b] \mapsto \mathbb{R}$ such that $T^c_U f \le T^c_D f$ we have*

$$UTV^c(f, [a; s]) = DTV^c(f, [a; s]) = 0,$$

when $s \in \left[a; T^c_{U,0}\right)$ and

$$UTV^c(f,[a;s]) := \begin{cases} \sum_{i=0}^{k-1}\left\{M^c_i - m^c_i - c\right\} + M^c_k(s) - m^c_k - c & \text{if } s\in\left[T^c_{U,k};T^c_{D,k}\right), \\ \sum_{i=0}^{k}\left\{M^c_i - m^c_i - c\right\} & \text{if } s\in\left[T^c_{D,k};T^c_{U,k+1}\right), \end{cases}$$

$$DTV^c(f,[a;s]) := \begin{cases} \sum_{i=0}^{k-1}\left\{M^c_i - m^c_{i+1} - c\right\} & \text{if } s\in\left[T^c_{U,k};T^c_{D,k}\right), \\ \sum_{i=0}^{k-1}\left\{M^c_i - m^c_{i+1} - c\right\} \\ \quad + M^c_k - m^c_{k+1}(s) - c & \text{if } s\in\left[T^c_{D,k};T^c_{U,k+1}\right). \end{cases}$$

Moreover, for any càdlàg function $f : [a; b] \mapsto \mathbb{R}$ and any $s \in [a; b]$ we have

$$TV^c(f, [a; s]) = UTV^c(f, [a; s]) + DTV^c(f, [a; s]). \tag{2.8}$$

2.2. Basic properties of $TV^c(f, [a; b])$, $UTV^c(f, [a; b])$ and $DTV^c(f, [a; b])$

We will now list some properties, most of which is used in the paper. These are taken from [6, Section 2.4, Section 2.5]. Unless stated otherwise the functions considered below are càdlàg.

- For any strictly increasing and continuous function $s : \mathbb{R} \to \mathbb{R}$

$$\mathrm{TV}^c(f, [a; b]) = \mathrm{TV}^c\left(f \circ s^{-1}, [s(a); s(b)]\right), \tag{2.9}$$

the analogous equalities hold for UTV^c and DTV^c.

- For any $f : [a; b] \mapsto \mathbb{R}$ and any $c > 0$ we have
$$\mathrm{DTV}^c(f, [a; b]) = \mathrm{UTV}^c(-f, [a; b]) . \tag{2.10}$$
- For any $s \in (a; b)$ we have
$$\mathrm{TV}^c(f, [a; b]) \geq \mathrm{TV}^c(f, [a; s]) + \mathrm{TV}^c(f, [s; t]) , \tag{2.11}$$
and the analogous inequalities hold for UTV^c and DTV^c.
- On the other hand, for any $s \in (a; b)$ we have
$$\mathrm{TV}^c(f, [a; b]) \leq \mathrm{TV}^c(f, [a; s]) + \mathrm{TV}^c(f, [s; t]) + c, \tag{2.12}$$
and the analogous inequalities hold for UTV^c and DTV^c.
- For any $f, g : [a; b] \to \mathbb{R}$ and $c_1, c_2 \geq 0$ we have
$$\mathrm{TV}^{c_1+c_2}(f + g, [a; b]) \leq \mathrm{TV}^{c_1}(f, [a; b]) + \mathrm{TV}^{c_2}(g, [a; b]) , \tag{2.13}$$
and the analogous inequalities hold for UTV^c and DTV^c. Note that above we admit some quantities to be infinite in case $c_1 = 0$ or $c_2 = 0$. In particular
$$\left|\mathrm{TV}^c(f + g, [a; b]) - \mathrm{TV}^c(f, [a; b])\right| \leq \mathrm{TV}(g, [a; b]) . \tag{2.14}$$
These facts were not proved in [6]. We offer a proof in Fact 17 below.
- For any $f : [a; b] \mapsto \mathbb{R}$ mapping
$$(0, +\infty) \ni c \mapsto \mathrm{TV}^c(f, [a; b]) ,$$
is convex and decreasing hence continuous. The same holds true for UTV^c and DTV^c. Moreover, though not mentioned in [6], it can be easily upgraded to functional setting. E.g. we define functional $T : (0; +\infty) \mapsto \mathcal{D}$ (Skorohod space of càdlàg functions) given by $T(c)(t) := \mathrm{TV}^c(f, [a; t])$ is convex and decreasing in a point-wise sense.
- For any $f : [a; b] \mapsto \mathbb{R}$ we have
$$\lim_{c \searrow 0} \mathrm{TV}^c(f, [a; b]) = \mathrm{TV}(f, [a; b]) , \tag{2.15}$$
we recall that the right-hand side might be infinite.

Fact 17. *For any $f, g : [a; b] \to \mathbb{R}$ and $c_1, c_2 \geq 0$ we have*
$$TV^{c_1+c_2}(f + g, [a; b]) \leq TV^{c_1}(f, [a; b]) + TV^{c_2}(g, [a; b]) , \tag{2.16}$$
and the analogous inequalities hold for UTV^c and DTV^c.

Proof. The inequality for UTV^c holds by definition (1.3) and the inequality
$$\begin{aligned} &\max\{f(s) + g(s) - f(t) - g(t) - c_1 - c_2, 0\} \\ &\quad = \max\{f(s) - f(t) - c_1 + g(s) - g(t) - c_2, 0\} \\ &\quad \leq \max\{f(s) - f(t) - c_1, 0\} + \max\{g(s) - g(t) - c_2, 0\} . \end{aligned}$$
By (2.10) we have similar property for DTV^c. Finally, to obtain (2.16) it is enough to utilize (2.8). □

3. Proof of Theorem 5

The proof structure reflects the outline contained in Remark 10. We start with the following.

3.1. Proof for Wiener process with drift

In our proof we will use an Anscombe-like result. It is not much more than a reformulation of [13, Theorem 4.5.5] to our specific needs. From now on we will use "$\lesssim$" to denote the situation when an equality or inequality holds with some constant which is irrelevant for calculations. Our setting is as follows. Let us fix some $T > 0$ and

$$(D_i(c), Z_i(c)), \quad i \geq 1,$$

be sequences of i.i.d. random vectors indexed by certain parameter $c \in (0, 1]$. We define

$$M_c(t) := \min\left\{ i \geq 0 : \sum_{i=1}^{i+1} D_i(c) > t \right\}, \tag{3.1}$$

$$P_c(t) := \left(\sum_{i=1}^{M_c(t)} Z_i(c) \right) - \frac{\mathbb{E}Z_1(c)}{\mathbb{E}D_1(c)} t, \quad t \in [0; T]. \tag{3.2}$$

Let us observe that such defined M_c, P_c are càdlàg processes. We will use the following assumptions.

(A1) For any $c > 0$ we have $D_1(c) > 0$ a.s. and $\mathbb{E}D_1(c) \to 0$ as $c \searrow 0$.

(A2) We denote $X_i(c) := Z_i(c) - (\mathbb{E}Z_1(c)/\mathbb{E}D_1(c))D_i(c)$. We have $\mathbb{E}X_i(c) = 0$. We assume that there exists $\sigma > 0$ such that

$$\frac{\mathbb{E}X_1(c)^2}{\mathbb{E}D_1(c)} \to \sigma^2, \quad \text{as } c \searrow 0.$$

(A3) There exists $\delta \in (0, 2]$ such that

$$\frac{\mathbb{E}|X_1(c)|^{2+\delta}}{\mathbb{E}D_1(c)} \to 0, \quad \text{as } c \searrow 0.$$

(A4) There exists $\delta > 0, C > 0$ such that for any $c \in (0; 1]$ we have

$$\mathbb{E}|D_1(c)|^{1+\delta} \leq C(\mathbb{E}D_1(c))^{1+\delta}.$$

Before the formulation of the fact we define

$$\mathcal{D} := \mathcal{D}([0; T], \mathbb{R}) := \{f : [0; T] \mapsto \mathbb{R} : f \text{ is càdlàg}\}, \tag{3.3}$$

we equip this space with $\|\cdot\|_\infty$-norm. This may seem unusual, as the Skorohod metric (see [1, Chapter 3]) is a more natural choice for space $\mathcal{D}$. Let us note however that in all cases we will obtain the convergence to a continuous limit. In such case both notions are equivalent (see [1, Section 18].

Fact 18. *Let $T > 0$ and assume that (A1)–(A4) hold. Then*

$$P_c \to^d \sigma B, \quad \text{as } c \searrow 0,$$
$$(\mathbb{E}D_1(c))M_c \to^d id, \quad \text{as } c \searrow 0, \tag{3.4}$$

where σ^2 is the same as in (A2), $id(x) = x$, and the convergence is understood as weak convergence in $\mathcal{D}([0; T], \mathbb{R})$.

Proof. We define

$$S_c(n) := \sum_{i=1}^{n} Z_i(c), \qquad V_c(n) := \sum_{i=1}^{n} D_i(c), \quad n \in \mathbb{N}. \tag{3.5}$$

Moreover, let us denote $f(c) := \frac{\mathbb{E}Z_1(c)}{\mathbb{E}D_1(c)}$ and we recall that $X_i(c) := Z_i(c) - f(c)D_i(c)$. We define a family of auxiliary processes

$$P_c^1(t) := H_c(\lfloor g(c)t \rfloor), \quad t \geq 0, \tag{3.6}$$

where $H_c(n) := S_c(n) - f(c)V_c(n)$ and $g(c) := (\mathbb{E}D_1(c))^{-1}$. By (A1) $g(c) \to +\infty$ as $c \searrow 0$.

Now the proof follows by [13, Theorem 4.5.5, p. 290]. The assumptions of [13, Theorem 4.5.5] consist of seven conditions denoted by $\mathcal{T}_4, \mathcal{S}_4, \mathcal{S}_5, \mathcal{S}_7, \mathcal{S}_8, \mathcal{S}_9$ and $\mathcal{J}_{20}$. These conditions read as:

- $(\mathcal{T}_4)$: $\left(\kappa_{\varepsilon,k}, \xi_{\varepsilon,k}\right)$, $k = 1, 2, \ldots$, is a sequence of i.i.d. random vectors that take values in $[0; +\infty) \times \mathbb{R}$;
- $(\mathcal{S}_4)$: $n_\varepsilon \mathbb{P}\left(\kappa_{\varepsilon,k} > u\right) \to \pi_1(u)$ as $\varepsilon \to 0$ for all $u > 0$, which are points of continuity of the limit function $\pi_1(u)$;
- $(\mathcal{S}_5)$: $n_\varepsilon \mathbb{E}\kappa_{\varepsilon,k} 1_{\{\kappa_{\varepsilon,k} \leq u\}} \to c(u)$ as $\varepsilon \to 0$ for some $u > 0$, which is a point of continuity of $\pi_1(u)$;
- $(\mathcal{S}_7)$: $n_\varepsilon \mathbb{P}\left(\left|\xi_{\varepsilon,k}\right| > u\right) \to 0$ as $\varepsilon \to 0$ for every $u > 0$;
- $(\mathcal{S}_8)$: $n_\varepsilon \mathbb{E}\left|\xi_{\varepsilon,k}\right| 1_{\{|\xi_{\varepsilon,k}| \leq u\}} \to a$ as $\varepsilon \to 0$ for some $u > 0$;
- $(\mathcal{S}_9)$: $n_\varepsilon \mathbb{D}^2\left|\xi_{\varepsilon,k}\right| 1_{\{|\xi_{\varepsilon,k}| \leq u\}} \to b^2$ as $\varepsilon \to 0$ for some $u > 0$;
- $(\mathcal{J}_{20})$: $c = c(u) - \int_0^u s d\pi_1(s) > 0$, where $\pi_1(s)$ and $c(u)$ are obtained in $(\mathcal{S}_4)$ and $(\mathcal{S}_5)$ respectively.

Before verifying assumptions we list how our notation translates to the one of [13], c is ϵ, $\lceil g(c) \rceil$ is n_ϵ, $D_i(c)$ is $\kappa_{\epsilon,i}$ and $X_i(c)$ is $\xi_{\epsilon,i}$. Condition $\mathcal{T}_4$ (p. 287) is obviously fulfilled. Conditions $\mathcal{S}_4$ and $\mathcal{S}_5$ (p. 283) hold with $\pi(u) = 0$ and $c(u) = 1$ respectively. Indeed, let us fix $u > 0$. $\mathcal{S}_4$ writes as

$$\lceil g(c) \rceil \mathbb{P}(D_1(c) > u) \leq \lceil g(c) \rceil u^{-1-\delta} \mathbb{E}|D_1(c)|^{1+\delta} \lesssim \lceil g(c) \rceil g(c)^{-(1+\delta)} \to 0, \quad \text{as } c \searrow 0,$$

where we used assumption (A1), (A4) and the Chebyshev inequality. We will use a few times an obvious inequality

$$|x|^{\delta_1+\delta_2} \geq |u|^{\delta_1}|x|^{\delta_2}, \tag{3.7}$$

valid for any $\delta_1, \delta_2 > 0$ and $|x| \geq |u|$. We check that

$$\lceil g(c) \rceil \mathbb{E}D_1(c) 1_{\{D_1(c) > u\}} \leq \lceil g(c) \rceil u^{-\delta} \mathbb{E}|D_1(c)|^{1+\delta} \to 0, \quad \text{as } c \searrow 0, \tag{3.8}$$

again by (A1), (A4) and (3.7). The expression in condition $\mathcal{S}_5$ writes in our notation as $\lceil g(c) \rceil \mathbb{E}D_1(c) 1_{\{D_1(c) \leq u\}}$. By (3.8) its limit does not depend on u and is the same as the one of

$$\lceil g(c) \rceil \mathbb{E}D_1(c) \to 1, \quad \text{as } c \searrow 0,$$

which follows by (A1) and the definition of $g(c)$.

We will now verify conditions $\mathcal{S}_7, \mathcal{S}_8, \mathcal{S}_9$ (p. 287–288) with $a = 0$ and $b^2 = \sigma^2$. Let $u > 0$, the condition $\mathcal{S}_7$ writes as

$$\lceil g(c) \rceil \mathbb{P}(|X_1(c)| \geq u) \leq \lceil g(c) \rceil u^{-(2+\delta)} \mathbb{E}|X_1(c)|^{2+\delta} \to 0, \quad \text{as } c \searrow 0,$$

where we used assumption (A3) and the Chebyshev inequality. Further we have

$$\lceil g(c)\rceil \mathbb{E}\left(|X_1(c)|1_{\{|X_1(c)|>u\}}\right) \le \lceil g(c)\rceil u^{-(1+\delta)}\mathbb{E}|X_1(c)|^{2+\delta} \to 0, \quad \text{as } c \searrow 0,$$

where we used assumption (A3) and (3.7). Now $\mathcal{S}_8$ follows directly from above and the equality

$$\mathbb{E}\left(X_1(c)1_{\{|X_1(c)|>u\}}\right) = -\mathbb{E}\left(X_1(c)1_{\{|X_1(c)|\le u\}}\right)$$

which is a consequence of the fact that $\mathbb{E}X_i(c) = 0$. Let us now observe that

$$\lceil g(c)\rceil \mathbb{E}\left(X_1(c)^2 1_{\{|X_1(c)|>u\}}\right) \le \lceil g(c)\rceil u^{-\delta}\mathbb{E}|X_1(c)|^{2+\delta} \to 0, \quad \text{as } c \searrow 0,$$

where we again used assumption (A3) and (3.7). By the above considerations we have that $\lim_{c\searrow 0}\lceil g(c)\rceil \mathrm{Var}(X_1(c)1_{\{|X_1(c)|\le u\}})$ is the same as $\lim_{c\searrow 0}\lceil g(c)\rceil \mathrm{Var}(X_1(c))$. Now $\mathcal{S}_9$ follows directly from (A2). Finally, $\mathcal{J}_{20}$ (p.285) holds with $c = 1$; see also [13, (4.5.2)].

Now, it is straightforward to identify the limit using the description in [13, p. 284 and p. 288]. Indeed, the process κ_0 (p. 284) is simply given by $\kappa_0(t) = t$ (notice that on the right hand side of formula (4.5.1) in [13, p. 284] one should replace z by y) so its inverse ν_0 is also $\nu_0(t) = t$ (which proves (3.4)). The process ξ_0 is the same as in $\mathcal{A}_{65}$ (p. 288). Let us note that Silvestrov's $\to^U$ is the same convergence we need; see [13, Definition 2.4.2]. □

Let W be a standard Wiener process and $\mu \in \mathbb{R}$. We denote a Wiener process with drift μ by

$$X_t := W_t + \mu t, \quad t \ge 0. \tag{3.9}$$

Our first result is the following.

Lemma 19. *Let $T > 0$ and X be a Wiener process with drift given by (3.9). We have*

$$\left(X_t - \mu t, TV^c(X,t) - \frac{t}{c}\right) \to^d \left(W_t, 3^{-1/2}B_t\right), \quad \text{as } c \searrow 0, \tag{3.10}$$

where (W, B) are independent standard Wiener processes. The convergence is understood as weak convergence in $\mathcal{C}([0;T],\mathbb{R})^2$ topology.

Proof. We fix $a, b \in \mathbb{R}$ and define $A_t^c := a\mathrm{TV}^c(X,t) + bX_t - \left(\frac{a}{c} + b\mu\right)t$. Assume that we proved that

$$A^c \to^d \left(a^2/3 + b^2\right)\tilde{B}, \quad \text{as } c \searrow 0, \tag{3.11}$$

weakly in topology of $\mathcal{C}([0;T],\mathbb{R})$, where $\tilde{B}$ is some standard Brownian motion. The convergence for $(a,b) = (1,0)$ yields that $\left\{\mathrm{TV}^c(X,t) - \frac{t}{c}\right\}_{c>0}$ is tight; hence also is the sequence of vectors on the left hand side of (3.10). Now, applying the Cramér–Wold device [1, Theorem 7.7] we easily justify the convergence of finite-dimensional distributions; hence (3.10) indeed holds.

Now we are to prove (3.11). We transparently transfer all quantities of Section 2 to the stochastic setting by applying them in a pathwise fashion, i.e. $f(t) = X_t$. We denote

$$G_i(c) := \left(M_i^c - m_i^c - c\right) + \left(M_i^c - m_{i+1}^c - c\right),$$
$$H_i(c) := \left(M_i^c - m_i^c - c\right) - \left(M_i^c - m_{i+1}^c - c\right).$$

By Theorem 16 and continuity of X we have $\mathrm{TV}^c\left(X, T_{U,k}^c\right) = \sum_{i=0}^{k-1} Y_i(c)$ (in fact this holds under additional assumption (2.5) but this is irrelevant in the limit). By (2.3) and again by

Theorem 16 we have

$$\|X_{T^c_{U,k}} - \sum_{i=0}^{k-1} H_i(c)\|_\infty \le c, \quad \text{a.s.} \tag{3.12}$$

(Note that $X_0 = 0$.) We fix some $a, b \in \mathbb{R}$ and for any $i \ge 0$ write

$$Z_i(c) := aG_i(c) + bH_i(c). \tag{3.13}$$

We denote also

$$D_i(c) := T^c_{U,i} - T^c_{U,i-1}, \quad i \ge 1, \text{ and } D_0(c) := T^c_{U,0}. \tag{3.14}$$

The following simple observation will be crucial for the further proof. Let us notice that by the strong Markov property of X and its space homogeneity we have that $\{Z_i(c)\}_{i\ge 1}$ and $\{D_i(c)\}_{i\ge 1}$ are i.i.d. sequences. For $i = 0$ the distributions are different because of "starting conditions". The first part, i.e. the values for $i = 0$ disappear in the limit. For notational simplicity from now on, we will implicitly assume that $i \ge 1$.

We will proceed now in the direction of utilizing Fact 18. To do this, we need to calculate moments; fortunately enough [14] provides us with sufficient tools. Using the notation from [14] we may write

$$(T^c_{D,i} - T^c_{U,i}, M^c_i - m^c_i - c) =^d (T_c, X(T_c) + c),$$

where T_c, X are defined in [14, Introduction]. Hence the formula [14, (1.1)] reads as

$$\mathbb{E}\exp(\alpha\left(M^c_i - m^c_i - c\right) - \beta\left(T^c_{D,i} - T^c_{U,i}\right)) = \frac{\delta\exp(-(\alpha+\mu)c)\exp(\alpha c)}{\delta\cosh(\delta c) - (\alpha+\mu)\sinh(\delta c)}, \tag{3.15}$$

where $\delta = \sqrt{\mu^2 + 2\beta}$. This formula is valid if $\alpha < \delta\coth(\delta c) - \mu$ and $\beta > 0$. If $\mu \ne 0$ we may also put $\beta = 0$. One may check that the pair

$$\left(T^c_{U,i+1} - T^c_{D,i}, M^c_i - m^c_{i+1} - c\right)$$

is independent of $(T^c_{D,i} - T^c_{U,i}, M^c_i - m^c_i - c)$. It becomes obvious when one recalls definitions of Section 2 ((2.6) and (2.7) in particular) and apply the strong Markov property of X. Moreover, we notice that the law of $\left(T^c_{U,i+1} - T^c_{D,i}, M^c_i - m^c_{i+1} - c\right)$ is the same as the one of $(T^c_{D,i} - T^c_{U,i}, M^c_i - m^c_i - c)$ if we change the drift coefficient to $-\mu$. Therefore, by [14, (1.1)] we get

$$\mathbb{E}\exp(\alpha\left(M^c_i - m^c_{i+1} - c\right) - \beta\left(T^c_{U,i+1} - T^c_{D,i}\right)) = \frac{\delta\exp(-(\alpha-\mu)c)\exp(\alpha c)}{\delta\cosh(\delta c) - (\alpha-\mu)\sinh(\delta c)}, \tag{3.16}$$

where $\delta = \sqrt{\mu^2 + 2\beta}$ (with the same restrictions as before). These are enough information to check the moment conditions required in Fact 18. Calculations are easy and straightforward, however lengthy. We decided not to include all of them in the paper. Instead, we list crucial steps and provide the reader with the Mathematica notebook with all details.[1] Combining the above equations and putting $\alpha = 0$ (note that this is always possible for c's small enough) we get

$$\mathbb{E}\exp(-\beta D_i(c)) = \frac{2\beta + \mu^2}{\beta + \mu^2 + \beta\cosh\left(2c\sqrt{2\beta + \mu^2}\right)}.$$

[1] http://www.mimuw.edu.pl/~pmilos/moments.nb. The file can be viewed with a free application available on http://www.wolfram.com/products/player/.

Differentiation yields

$$\mathbb{E}D_i(c) = \frac{2\sinh(c\mu)^2}{\mu^2} = 2c^2 + O(c^4). \tag{3.17}$$

One can check that the formula given above is valid for $\mu = 0$ when we take the limit. This applies also to the subsequent moments formulas. Moreover

$$\mathbb{E}D_i(c)^2 = \frac{16}{3}c^4 + O(c^6), \qquad \mathbb{E}D_i(c)^4 = \frac{7936}{105}c^8 + O(c^{10}).$$

This is enough to check conditions (A1) of Fact 18 as well as (A4) with $\delta = 3$. Analogously, by putting $\beta = 0$ we calculate that

$$\mathbb{E}\exp(\alpha Z_i(c)) = \frac{4\mu^2}{\left((a-b)\left(1-e^{-2c\mu}\right)\alpha - 2\mu\right)\left((a+b)\left(1-e^{2c\mu}\right)\alpha + 2\mu\right)}. \tag{3.18}$$

Again, by differentiation one gets

$$\mathbb{E}Z_i(c) = \frac{2\sinh(c\mu)(a\cosh(c\mu) + b\sinh(c\mu))}{\mu}. \tag{3.19}$$

And therefore

$$\frac{\mathbb{E}Z_i(c)}{\mathbb{E}D_i(c)} = \mu(b + a\coth(c\mu)) = \frac{a}{c} + b\mu + O(c). \tag{3.20}$$

Now we have

$$\mathbb{E}\exp\left(\alpha Z_i(c) - \beta D_i(c)\right)$$
$$= \frac{2\left(2\beta+\mu^2\right)}{-a^2\alpha^2 + b^2\alpha^2 + 2b\alpha\mu + 2\left(\beta+\mu^2\right) + \left(a^2\alpha^2 + 2\beta - b\alpha(b\alpha+2\mu)\right)\cosh(2c\sqrt{2\beta+\mu^2}) - 2a\alpha\sqrt{2\beta+\mu^2}\sinh(2c\sqrt{2\beta+\mu^2})}.$$

Following axiom (A2) we denote $X_i(c) := Z_i(c) - (\mathbb{E}Z_1(c)/\mathbb{E}D_1(c))D_i(c)$.

Using this one may check that

$$\mathbb{E}X_i(c)^2 = \frac{3a^2 - b^2 - 4abc\mu + \left(a^2+b^2\right)\cosh(2c\mu) - 4a^2c\mu\coth(c\mu) + 2ab\sinh(2c\mu)}{\mu^2}.$$

Now it is straightforward to check (A2) of Fact 18, viz.

$$\frac{\mathbb{E}X_i(c)^2}{\mathbb{E}D_1} = \frac{1}{2}\operatorname{cshs}(c\mu)^2(3a^2 - b^2 - 4abc\mu + \left(a^2+b^2\right)\cosh(2c\mu)$$
$$- 4a^2c\mu\coth(c\mu) + 2ab\sinh(2c\mu) = \left(\frac{a^2}{3} + b^2\right) + \frac{4}{3}abc\mu + O(c^2).$$

Finally, one can check that $\mathbb{E}X_i(c)^4 \lesssim c^4$ and hence (A3) is verified with $\delta = 2$. Having checked all conditions we conclude that for $P_c(t)$ defined by (3.2) and (3.13), (3.14) we have

$$P_c(t) - \left(\frac{a}{c} + b\mu\right)t \to^d \left(\frac{a^2}{3} + b^2\right)^{1/2}\tilde{B}, \quad \text{as } c \searrow 0.$$

Therefore in order to prove (3.11) it is enough to show that $P_c(t) - a\mathrm{TV}^c(X,t) - bX_t \to^d 0$. By the property (3.12) and the continuity of X it follows easily that it suffices to concentrate on the case $(a, b) = (1, 0)$, that is $A_t = \mathrm{TV}^c(X, t)$. Since $D_0(c)$ has a different distribution than $D_i(c)$ for $i \geq 1$ we introduce two auxiliary objects

$$\tilde{M}_c(t) := \min\left\{n \geq 0 : \sum_{i=0}^{n} D_i(c) > t\right\}, \qquad \tilde{S}_c(n) = \sum_{i=0}^{n} Z_i(c),$$

and

$$\tilde{P}_c(t) := \tilde{S}_c(\tilde{M}_c(t)).$$

This differs slightly from P_c; however, one easily checks that $\tilde{P}_c - P_c \to^d 0$. By Theorem 16 we see that the processes $\mathrm{TV}^c(X, t)$ and $\tilde{S}_c\left(\tilde{M}_c(t)\right)$ coincide at random times $T^c_{U,i}$, $i \geq 1$; moreover, both are increasing, hence, for any $T \geq 0$ and $\varepsilon > 0$

$$\mathbb{P}\left(\sup_{t\in[0,T]}\left|\mathrm{TV}^c(X, t) - \tilde{S}_c(\tilde{M}_c(t))\right| > \varepsilon\right) \leq \mathbb{P}\left(\sup_{t\in[0;T]} Z_{\tilde{M}_c(t)}(c) > \varepsilon\right).$$

Using this we estimate

$$\mathbb{P}\left(\sup_{t\in[0,T]}\left|TV^{\mu}_c(t) - \tilde{S}_c\left(\tilde{M}_c(t)\right)\right| > \varepsilon\right)$$
$$\leq \mathbb{P}\left(\max_{k\leq 2T/\mathbb{E}D_1(c)+1} Z_k(c) \geq \varepsilon\right) + \mathbb{P}\left(\tilde{M}_c(T) \geq \frac{2T}{\mathbb{E}D_1(c)} + 1\right).$$

The first term could be estimated by the Chebyshev inequality and the estimates of $\mathbb{E}Z_1(c)^4$ and $\mathbb{E}D_1(c)$

$$\mathbb{P}\left(\max_{k\leq 2T/\mathbb{E}D_1(c)+1} |Z_i(c)| > \varepsilon\right) \leq \left(\frac{2T}{\mathbb{E}D_1(c)} + 1\right)\frac{\mathbb{E}Z_1(c)^4}{\varepsilon^4} \to 0, \quad \text{as } c \to 0.$$

The convergence of the second term to 0 could be established by Fact 18. □

3.2. *Proof for diffusions with $\sigma = const$*

We start with a yet simpler case. Namely, let W be a standard Brownian motion and X be a random variable. Let us define process Z by

$$Z_t := W_t + Xt, \quad t \geq 0.$$

Lemma 20. *Let $T > 0$. Let us assume that W and X are independent; then*

$$\left(X, W, TV^c(Z, t) - \frac{t}{c}\right) \to^d \left(X, W, 3^{-1/2}B\right), \quad \text{as } c \searrow 0$$

where B is a standard Brownian motion and X, W, B are independent. The convergence is understood in weak sense in the product topology of $\mathbb{R} \times \mathcal{C}([0; T], \mathbb{R})^2$.

Proof. We will proceed by the very definition of the weak convergence. Let $f : \mathbb{R} \times \mathcal{C}([0; T], \mathbb{R})^2 \mapsto \mathbb{R}$ be a bounded continuous function. We have

$$\lim_{c\searrow 0} \mathbb{E}f\left(X, W, \mathrm{TV}^c(Z, t) - \frac{t}{c}\right)$$

$$= \lim_{c\searrow 0} \mathbb{E}\mathbb{E}\left(f\left(x, W, \mathrm{TV}^c(Z,t) - \frac{t}{c}\right)\middle| X = x\right)$$
$$= \mathbb{E}\lim_{c\searrow 0} \mathbb{E}\left(f\left(x, W, \mathrm{TV}^c(Z,t) - \frac{t}{c}\right)\middle| X = x\right)$$
$$= \mathbb{E}\mathbb{E}\left(f\left(x, W, 3^{-1/2}B\right)\middle| X = x\right) = \mathbb{E} f\left(X, W, 3^{-1/2}B\right)$$

where we used Lemma 19 and the Lebesgue dominated convergence theorem. □

We will deal now with diffusion given by an equation

$$dX_t = \mathrm{d}W_t + \mu(X_t)\mathrm{d}t, \quad X_0 = 0, \tag{3.21}$$

i.e. we set $\sigma \equiv 1$ in (1.4). We assume also that μ is bounded and Lipschitz. This process is essentially a Brownian motion with "a variable drift". We denote

$$\mu^* = \sup_{x\in\mathbb{R}} |\mu(x)| < +\infty. \tag{3.22}$$

We will use the discretion technique. To this end we need to be able to control the increments of X. The following simple lemma is the first, most crude step of our analysis

Lemma 21. *Let $t \geq 0$ and $\delta > 0$; then for any $b > 0$ we have*

$$\mathbb{P}\left(\sup_{s\in[t;t+\delta]} |X_s - X_t| \geq (\mu^* + b)\delta\right) \leq 2\exp\left(-b^2\delta/2\right).$$

Proof. We know that

$$X_t = X_0 + W_t + \int_0^t \mu(X_s)\mathrm{d}s.$$

Hence, we have $X_s - X_t \in (W_s - W_t - \mu^*(s-t), W_s - W_t + \mu^*(s-t))$. Now the lemma follows by [11, Proposition II.1.8]. □

Let us fix $T > 0, n = 1, 2, \ldots$ and denote $t_i^n := i\frac{T}{n}, i \in \{0, 1, \ldots, n\}$. We define the "approximated" truncated variation process by

$$ATV^{n,c}(t) := \sum_{i=0}^{\lfloor nt\rfloor -1} \mathrm{TV}^c\big(X, [t_i^n; t_{i+1}^n]\big) + \mathrm{TV}^c\left(X, [t_{\lfloor nt\rfloor}^n; t]\right),. \tag{3.23}$$

Its name is justified by the following lemma.

Lemma 22. *We have*

$$ATV^{n,c}(t) - TV^c(X,t) \to 0, \ \textit{a.s.} \quad \textit{when } c \searrow 0,$$

and the convergence is understood in $C([0;T],\mathbb{R})$ topology.

Proof. By (2.11) one easily verifies that $ATV^{n,c}(t) \leq \mathrm{TV}^c(X,t)$. On the other hand, by (2.12), $\mathrm{TV}^c(X,t) - ATV^{n,c}(t) \leq nc$, for any $t \in [0;T]$. □

We will take now a detour of the main flow of the proof in order to collect weak convergence facts used below. First we recall the Prokhorov metric. Let (S, d) be a metric space and $\mathcal{P}(S)$ be the space of Borel probability measures on S. We topologise $\mathcal{P}(S)$ with the Prokhorov metric

$$d_P(P, Q) := \inf\left\{\epsilon > 0 : P(F) \le Q(F^\epsilon) + \epsilon, \text{ for all closed } F \subset S\right\}, \tag{3.24}$$

in the above expression $F^\epsilon := \left\{x \in S : \inf_{y\in F} d(x, y) < \epsilon\right\}$. It is well-known that when (S, d) is separable then convergence with respect to $d_P(\cdot, \cdot)$ is equivalent to weak convergence. We refer the reader to [2, Chapter 3] and [2, Theorem 3.3.1] in particular. Given two random variables X, Y with values in the same space we will write

$$d_P(X, Y) := d_P(\mathcal{L}(X), \mathcal{L}(Y)),$$

where $\mathcal{L}(X)$ denotes the law of X.

In some parts of our analysis we will need the space of càdlàg functions $\mathcal{D}([0; T], \mathbb{R})$ introduced by (3.3). We will also use the following product space

$$\mathcal{C} \times \mathcal{D} := \mathcal{C}([0; T], \mathbb{R}) \times \mathcal{D}([0; T], \mathbb{R}), \tag{3.25}$$

always with the norm given by $\|(f, g)\| := \|f\|_\infty + \|g\|_\infty$.

Lemma 23. *Let (X, Y) be random variables with values in $\mathcal{C} \times \mathcal{D}$; moreover let A be an event. Then*

$$d_P((X, Y), (X1_A, Y)) \le 2(1 - \mathbb{P}(A)).$$

Proof. It is enough to apply [2, Theorem 3.1.2] with $\mu = \mathcal{L}((X, Y), (X1_A, Y))$. □

Lemma 24. *Let $X := (X_1, X_2)$ and $Y := (Y_1, Y_2)$ be a random variable with values in $\mathcal{C} \times \mathcal{D}$ such that*

$$\mathbb{P}\left(\|X_1 - Y_1\|_\infty \ge \epsilon/2\right) \le \epsilon/2 \quad \textit{and} \quad \mathbb{P}\left(\|X_2 - Y_2\|_\infty \ge \epsilon/2\right) \le \epsilon/2,$$

then

$$d_P(X, Y) \le \epsilon.$$

Proof. We calculate

$$\mathbb{P}\left(\|(X_1 - Y_1, X_2 - Y_2)\| \ge \epsilon\right) \le \mathbb{P}\left(\|X_1 - Y_1\| \ge \epsilon/2\right) + \mathbb{P}\left(\|X_2 - Y_2\| \ge \epsilon\right) \le \epsilon,$$

now the proof follows directly by the application of [2, Theorem 3.1.2]. □

We are ready to prove the main result of this part of the proof which is an upgrade of Lemma 19 to "simplified diffusions" given by (3.21).

Fact 25. *Let $T > 0$. We have*

$$\left(X, TV^c(X, t) - \frac{t}{c}\right) \to^d (X, 3^{-1/2}B), \quad \textit{as } c \searrow 0, \tag{3.26}$$

where the convergence is understood as weak convergence in $\mathcal{C}([0; T], \mathbb{R}^d)^2$ topology and B is a Brownian motion independent of X.

Proof. We recall that $t_i^n := \frac{i}{n}T$, fix some $A \geq \mu^* + 1$ and define random sets

$$A_i^n := [X_{t_i^n} - A/n^{1/4}; X_{t_i^n} + A/n^{1/4}].$$

We also define random variables

$$\mu_i^n := \mu(X_{t_i^n})$$

and events

$$E_i^n := \left\{X_s \in A_i^n, \text{ for } s \in [t_i^n; t_{i+1}^n]\right\}, \quad E^n := \bigcap_{i\in\{0,1,\ldots,n-1\}} E_i^n.$$

Using Lemma 21 we check that for n large enough we have $\mathbb{P}\left(E_i^n\right) \geq 1 - 2n\exp(-n^{1/2}/2)$. Consequently, $\mathbb{P}(E^n) \to 1$ as $n \to +\infty$. For $n \in \mathbb{N}$ we define càdlàg processes $\left\{X_t^n\right\}_{t\in[0;T]}$ which approximate our diffusion:

$$X_t^n := X_{t_i^n} + \mu_i^n(t - t_i^n) + W_t - W_{t_i^n}, \quad \text{whenever } t \in [t_i^n; t_{i+1}^n).$$

One easily checks that $X^n \to X$ a.s. with respect to $\|\cdot\|_\infty$. Let us recall (3.23), we define its counterpart for X^n, viz.,

$$H^{n,c}(t) := \sum_{i=0}^{\lfloor nt\rfloor-1} \mathrm{TV}^c\left(X^n, [t_i^n; t_{i+1}^n]\right) + \mathrm{TV}^c\left(X^n, [t_{\lfloor nt\rfloor}^n; t]\right).$$

One checks (using the same method as in the proof of Lemma 22) that

$$H^{n,c}(t) - \mathrm{TV}^c\left(X^n, t\right) \to 0, \text{ a.s.} \quad \text{when } c \searrow 0,$$

norm $\|\cdot\|_\infty$. On each interval $t \in [t_i^n; t_{i+1}^n)$ we have

$$X_t^n - X_t = \int_{t_i^n}^t (\mu_i^n - \mu(X_s))\mathrm{d}s. \tag{3.27}$$

We observe that conditionally on E_i^n this expression defines a function of t which is Lipschitz with constant $w_n \leq Ln^{-1/4}$ for some $L > 0$. This follows by the fact that μ is a Lipschitz function itself. By (2.13) applied with $c_1 = c$ and $c_2 = 0$, conditionally on E_i^n, we have that

$$\mathrm{TV}^c\left(X^n, [t_i^n, t]\right) - w_n(t - t_i^n) \leq \mathrm{TV}^c\left(X, [t_i^n, t]\right) \leq \mathrm{TV}^c\left(X^n, [t_i^n, t]\right) + w_n(t - t_i^n),$$

for any $t \in [t_i^n; t_{i+1}^n]$. Further

$$1_{E^n} H^{n,c}(t) - w_nT \leq 1_{E^n} ATV^{n,c}(t) \leq 1_{E^n} H^{n,c}(t) + w_nT, \tag{3.28}$$

for any $t \in [0; T]$. In other words: $\|1_{E^n} ATV^{n,c}(t) - 1_{E^n} H^{n,c}(t)\|_\infty \leq 2w_nT$. Lemma 24 implies that

$$d_P\left((X^n, 1_{E^n} ATV^{n,c}), (X^n, 1_{E^n} H^{n,c})\right) \leq 4w_nT. \tag{3.29}$$

It will be crucial that this estimate is uniform in c. Let us denote $L^{n,c} := (H^{n,c}(t) - c/t)$. Lemma 20 applied term by term to $H^{n,c}$ yields the functional convergence

$$(L^{n,c}, X^n) \to^d (3^{-1/2}B, X^n), \quad \text{as } c \searrow 0, \tag{3.30}$$

where B and X^n are independent. In the above, we understand the convergence as the functional one in $\mathcal{C} \times \mathcal{D}$ (see also (3.25)).

The rest of the proof will follow by metric-theoretic considerations. Let us denote

$$X_1(c) := \left(\mathrm{TV}^c(X,t) - t/c, X\right), \qquad X_2(c,n) := \left(\mathrm{TV}^c(X,t) - t/c, X^n\right),$$
$$X_3(c,n) := \left(ATV^{n,c}(t) - t/c, X^n\right), \qquad X_4(c,n) := \left(1_{E^n}(ATV^{n,c}(t) - t/c), X^n\right),$$
$$X_5(c,n) := \left(1_{E^n}(H^{n,c}(t) - t/c), X^n\right), \qquad X_6(c,n) := \left(H^{n,c}(t) - t/c, X^n\right),$$
$$X_7(n) := \left(3^{-1/2}B, X^n\right), \qquad X_8 := \left(3^{-1/2}B, X\right).$$

Let us fix some $\epsilon > 0$. We find $n_{1,2}$ such that for any $n \geq n_{1,2}$ we have $d_P\left(X_1(c), X_2(c,n)\right) \leq \epsilon$ which is possible by Lemma 24 and convergence $X^n \to X$. We find $n_{3,4}$ such that for any $n \geq n_{3,4}$ we have $d_P\left(X_3(c,n), X_4(c,n)\right) \leq \epsilon$ which is possible by Lemma 23 and estimation of the probability of E^n. Further we find $n_{4,5}$ such that for any $n \geq n_{4,5}$ we have $d_P\left(X_4(c,n), X_5(c,n)\right) \leq \epsilon$ which is given by (3.29). Next, we check that for any $n \geq n_{3,4}$ we have $d_P\left(X_5(c,n), X_6(c,n)\right) \leq \epsilon$ as well. Finally, we choose $n_{7,8}$ such that for any $n \geq n_{7,8}$ we have $d_P\left(X_7(n), X_8\right) \leq \epsilon$ which holds by Lemma 24. We denote $N = \max(n_{1,2}, n_{3,4}, n_{4,5}, n_{7,8})$; obviously for this N all the above inequalities hold simultaneously for any $c > 0$.

Now we choose c_0 such that for any $c \leq c_0$ we have $d_P\left(X_2(c,N), X_3(c,N)\right) \leq \epsilon$ and $d_P\left(X_6(c,N), X_7(c,N)\right) \leq \epsilon$. The first one is possible by Lemmas 22 and 24 and the second one by (3.30) and again Lemma 24. Using the triangle inequality multiple times one obtains

$$d_P\left(X_1(c), X_8\right) \leq 8\epsilon, \quad \text{for any } c \leq c_0.$$

This yields convergence (3.26) since ϵ was arbitrary. □

Remark 26. We strongly believe that it is not possible to improve the above proof to general diffusions. The main reason is that without $\sigma = const$ assumption equation (3.27) is no longer true. Consequently, the estimate in (3.29) depends not only on w_n but also on c. Even worse, one can check that the estimate diverges to infinity as $c \searrow 0$. We could change n and c simultaneously in a smart way so that the estimate is still useful. However a new problem emerges then, namely estimate in Lemma 22 also depend on n and c. It appears that it is not possible to change n and c in such a way that both estimates converge to 0 when $c \searrow 0$.

3.3. *Proof for general diffusion*

Now we proceed to the general case. Before proving Theorem 5 we present some measure-theoretic considerations. In the reasoning below by $\mathbb{W}$ we denote the Wiener measure on $\mathcal{C}([0;T],\mathbb{R})$, see e.g. [11, Proposition I.3.3], and by H we denote the Cameron–Martin space, see [11, Definition VIII.2.1]. Moreover by $\mathcal{H}$ we denote algebra (i.e. class closed under finite sums and finite intersections) generated by open balls with centers in H. We have the following lemma.

Lemma 27. *Let $h : \mathcal{C}([0;T],\mathbb{R}) \mapsto \mathbb{R}_+$ be a measurable mapping such that $\int h(f)\mathbb{W}(\mathrm{d}f) = 1$. Then for any $\epsilon > 0$ there exists $m \in \mathbb{N}$, sets $A_1, A_2, \ldots, A_m \in \mathcal{H}$ and $h_1, h_2, \ldots, h_m \in \mathbb{R}_+$ such that*

$$\int_{\mathcal{C}} |h_\epsilon(f) - h(f)|\mathbb{W}(\mathrm{d}f) \leq \epsilon, \tag{3.31}$$

where

$$h_\epsilon(f) := \sum_{i=1}^{m} h_i 1_{A_i}(f).$$

Moreover, one may choose such $A_1, A_2, \ldots, A_m$ that for all $i \le m$, $\mathbb{W}(\partial A_i) = 0$.

Proof. In the proof we will write $\mathcal{C}$ instead of $\mathcal{C}([0;T],\mathbb{R})$ and $B(f,r)$ will denote an open ball with convention $B(f,0)=\emptyset$. Let us notice that without loss of generality we can assume that h is bounded by some $l>0$ and has compact support, say contained in ball $B(0,R)$. Indeed for any function h and any $\epsilon>0$ we can choose l, R such that $\int_{\mathcal{C}} |h(f)1_{\{h\le l\}}1_{\{f\in B(0,R)\}} - f(f)|\mathbb{W}(f) \le \epsilon/2$. Now it is enough to approximate $h(f)1_{\{h\le l\}}1_{\{f\in B(0,R)\}}$ with accuracy $\epsilon/2$. Therefore from now on we will work implicitly with the assumptions listed above.

Let us denote

$$S_k := \{f \in \mathcal{C} : h(f) \in (k\epsilon/2, (k+1)\epsilon/2]\},$$

for $k \in \{0,1,\ldots,2l/\epsilon\}$. We note that by our assumption sets S_k are bounded. We put $\delta := \epsilon^2/(4l^2)$. By the regularity of $\mathbb{W}$ (see [1, Theorem 1.1.1]) we can find open sets O_k such that

$$S_k \subset O_k \quad \text{and} \quad \mathbb{W}(O_k \setminus S_k) \le \delta/2. \tag{3.32}$$

It is well known that $\mathcal{C}$ is a separable space and H is its dense subspace so one can easily find a countable subset $\{f_1, f_2, \ldots\} \subset H$ which is dense in $\mathcal{C}$. For each f_i we define $r_i^k := \sup\{r : B(f_i, r) \subset O_k\}/2$ (by convention we put $r_i^k := 0$ if the set is empty). One promptly proves that $O_k = \bigcup_i B(f_i, r_i^k)$. By the continuity of measure there exists $i_k \in \mathbb{N}$ such that

$$\mathbb{W}(O_k) - \mathbb{W}\left(\bigcup_{i\le i_k} B(f_i, r_i^k)\right) \le \delta/2.$$

Let us denote $A_k := \bigcup_{i\le i_k} B(f_i, r_i^k)$. We now define

$$h_\epsilon(f) := \sum_k \left(\frac{k}{2}\epsilon\right) 1_{A_k}(f).$$

We will now show that h_ϵ is a good approximating function. We recall that by the construction $A_k \subset O_k$ and $W(O_k \setminus A_k) \le \delta/2$. This together with (3.32) yields that $\mathbb{W}(S_k \Delta A_k) \le \delta$, where Δ denotes the symmetric difference. We have

$$\begin{aligned}
\int_{\mathcal{C}} |h_\epsilon(f) - h(f)|\mathbb{W}(\mathrm{d}f) &= \int_{\mathcal{C}} \left|\sum_k (k\epsilon/2)\, 1_{A_k}(f) - \sum_k h(f) 1_{S_k}(f)\right| \mathbb{W}(\mathrm{d}f) \\
&\le \sum_k \int_{\mathcal{C}} \left|(k\epsilon/2)1_{A_k}(f) - h(f)1_{S_k}(f)\right| \mathbb{W}(\mathrm{d}f) \\
&\le \frac{\epsilon}{2}\sum_k \mathbb{W}(A_k \cap S_k) + l\sum_k \mathbb{W}(S_k \Delta A_k) \le \frac{\epsilon}{2} + l\frac{2l}{\epsilon}\delta = \epsilon.
\end{aligned}$$

To check $\mathbb{W}(\partial A_k) = 0$ it is enough to prove that for any $f \in H$ and any $r>0$ we have $\mathbb{W}(\partial B(f,r)) = 0$. By [11, Theorem VIII.2.2] it is enough to show that $\mathbb{W}(\partial B(0,r)) = 0$. This holds by the fact that sup of the Wiener process has a continuous density (see [11, Section III.3]). □

Finally we present the following.

Proof. (of Theorem 5). We will first show that in order to prove (1.5) it is enough to prove

$$\left(X, \mathrm{TV}^c(X,t) - \frac{\langle X_t\rangle}{c}\right) \to^d (X, 2M), \quad \text{as } c \searrow 0, \tag{3.33}$$

where M is the same as in Theorem 5. Since $X_0 = 0$ by (2.3) there exists process $\{R_c(t)\}_{t\in[0;T]}$ such that $\|R_c\|_\infty \le c$ almost surely and

$$X_t = \mathrm{UTV}^c(X,t) - \mathrm{DTV}^c(X,t) + R_c(t). \tag{3.34}$$

This together with (2.8) yields that

$$\mathrm{UTV}^c(X,t) = \frac{1}{2}\left(\mathrm{TV}^c(X,t) + X_t - R_c(t)\right). \tag{3.35}$$

Therefore

$$\mathrm{UTV}^c(X,t) - \frac{1}{2}\left(\frac{\langle X_t\rangle}{c} + X_t\right) = \frac{1}{2}\left(\mathrm{TV}^c(X,t) - \frac{\langle X_t\rangle}{c}\right) - \frac{1}{2}R_c(t).$$

Now the convergence follows simply by the fact that $\mathrm{TV}^c(X,t) - \langle X_t\rangle/c$ is a continuous transformation of (3.33) and by [1, Corollary 2, p.31], [1, Theorem 4.1]. A completely analogous argument proves the convergence of $\mathrm{DTV}^c(X,t) - \frac{1}{2}\left(\langle X_t\rangle/c - X_t\right)$. The joint convergence in (1.5) can be established in the same way.

It will be more convenient to work with the additional assumption that

$$C_1 \ge \sigma \ge C_2 > 0, \tag{3.36}$$

for some constants $C_1, C_2 > 0$. At the end of the proof we will remove this assumption. Diffusion (1.4) writes in the integral form as

$$X_t = \int_0^t \sigma(X_s)\mathrm{d}W_s + \int_0^t \mu(X_s)\mathrm{d}s.$$

Let us define $\beta_t := \int_0^t \sigma(X_s)^2\mathrm{d}s = \langle X_t\rangle$, its inverse $\alpha_t := \inf\{s \ge 0 : \beta_s > t\}$ and

$$\tilde{X}_t := X_{\alpha_t}, \quad t \in [0; T_0], \text{ where } T_0 := C_2^2 T.$$

By the time-change formula [10, Theorem 8.5.7] we obtain that $\tilde{X}$ is also a diffusion fulfilling equation

$$\tilde{X}_t = \tilde{W}_t + \int_0^t \frac{\mu(\tilde{X}_s)}{\sigma^2(\tilde{X}_s)}\mathrm{d}s,$$

for some Brownian motion $\tilde{W}$. We chose such T_0 that the definition is valid (i.e. $\alpha_{T_0} \le T$). We note also that $x \mapsto \frac{\mu(x)}{\sigma^2(x)}$ is a Lipschitz function. Let us now denote the natural filtration of $\tilde{X}$ (and $\tilde{W}$) by $\mathcal{F}$. Making the reverse change of time we get $X_t = \tilde{X}_{\beta_t}$. We denote also $\mathcal{G}_t := \mathcal{F}_{\beta_t}$. Now we can apply Fact 25. We know that

$$\left(\mathrm{CTV}^c(\tilde{X},t), \tilde{X}\right) \to^d (B, \tilde{X}), \tag{3.37}$$

where $CTV^c(X,t) := TV^c(X,t) - \frac{c}{t}$ and B and $\tilde{X}$ are independent. Let us also note that CTV^c can be regarded as a measurable mapping $\mathrm{CTV}^c : \mathcal{C}([0;T],\mathbb{R}) \mapsto \mathcal{C}([0;T],\mathbb{R})$.

Now, let $K \in \mathcal{H}$ be a non-empty set. We check that the measure $\mathbb{P}\left(\tilde{X} \in \cdot | X \in K\right)$ is absolutely continuous with respect to $\mathbb{P}\left(\tilde{X} \in \cdot\right)$. Indeed one needs only to check that $\mathbb{P}(X \in K) > 0$.

By the Radon–Nikodym theorem [4, Theorem A.1.3] there exists a measurable function h such that

$$\mathbb{P}\left(\tilde{X} \in \mathrm{d}f | X \in K\right) = h(f)\mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right). \tag{3.38}$$

Using this fact we can leverage (3.37). Let us first note that by the portmanteau theorem [1, Theorem I.2.1] and [1, Theorem I.2.2] and standard topological considerations we know that (3.37) is equivalent to

$$\mathbb{P}\left(\left\{CTV^c(\tilde{X}) \in K_1\right\} \cap \left\{\tilde{X} \in K_2\right\}\right) \to \mathbb{P}(B \in K_1)\,\mathbb{P}\left(\tilde{X} \in K_2\right), \quad \forall_{K_1, K_2 \in \mathcal{H}}. \tag{3.39}$$

Further, by (3.38), we have

$$\begin{aligned}
a_c &:= \mathbb{P}\left(\left\{CTV^c(\tilde{X}) \in K_1\right\} \cap \left\{\tilde{X} \in K_2\right\} \cap \{X \in K\}\right) = \mathbb{P}(X \in K) \\
&\mathbb{P}\left(\left\{CTV^c(\tilde{X}) \in K_1\right\} \cap \left\{\tilde{X} \in K_2\right\} | X \in K\right) \\
&\quad = \mathbb{P}(X \in K) \int_{\mathcal{C}} h(f) 1_{\{CTV^c(f) \in K_1\}} 1_{\{f \in K_2\}} \mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right),
\end{aligned}$$

where $\mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right)$ is the same as the Wiener measure. We now approximate h with accuracy $\epsilon = 1/n$ with simple function h^n satisfying conditions of Lemma 27 (we use additional superscript ^{n} to denote the case we are referring to). Hence we have

$$\int_{\mathcal{C}} |h^n(f) - h(f)| \mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right) \le \frac{1}{n}. \tag{3.40}$$

We define

$$a_c^n := \mathbb{P}(X \in K) \int_{\mathcal{C}} h^n(f) 1_{\{CTV^c(f) \in K_1\}} 1_{\{f \in K_2\}} \mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right).$$

One easily checks that for any $c > 0$ there is $|a_c - a_c^n| \le 1/n$. Applying (3.39) we obtain

$$\begin{aligned}
a_c^n &= \mathbb{P}(X \in K) \sum_{i=1}^{m^n} h_i^n \mathbb{P}\left(\left\{CTV^c(\tilde{X}) \in K_1\right\} \cap \left\{\tilde{X} \in A_i^n \cap K_2\right\}\right) \\
&\quad \to_{c \searrow 0} \mathbb{P}(B \in K_1)\,\mathbb{P}(X \in K) \sum_{i=1}^{m^n} h_i^n \mathbb{P}\left(\tilde{X} \in A_i^n \cap K_2\right) \\
&= \mathbb{P}(B \in K_1)\,\mathbb{P}(X \in K) \int_{K_2} h^n(f) \mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right) =: a^n.
\end{aligned}$$

It is easy to check that $|a^n - a| \le 1/n$, where

$$\begin{aligned}
a &:= \mathbb{P}(B \in K_1)\,\mathbb{P}(X \in K) \int_{K_2} h(f) \mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right) \\
&= \mathbb{P}(B \in K_1)\,\mathbb{P}\left(\left\{\tilde{X} \in K_2\right\} \cap \{X \in K\}\right).
\end{aligned}$$

Using the standards arguments we obtain that for any $K_1, K_2, K \in \mathcal{H}$

$$\mathbb{P}\left(\left\{CTV^c(\tilde{X}) \in K_1\right\} \cap \left\{\tilde{X} \in K_2\right\} \cap \{X \in K\}\right)$$

$$\to_{c\searrow 0} \mathbb{P}(B \in K_1)\,\mathbb{P}(X \in K) \int_{K_2} h(f)\mathbb{P}\left(\tilde{X} \in \mathrm{d}f\right)$$
$$= \mathbb{P}(B \in K_1)\,\mathbb{P}\left(\left\{\tilde{X} \in K_2\right\} \cap \{X \in K\}\right).$$

Using [1, Theorem I.2.2] in the same spirit as in the case of (3.39) we get

$$(CTV^c(\tilde{X}), \tilde{X}, X) \to^d (B, \tilde{X}, X), \quad \text{as } c \searrow 0$$

where B is independent of $(\tilde{X}, X)$ hence also of β. Changing the time according to this process we obtain

$$\mathrm{TV}^c\left(\tilde{X}, \beta_t\right) - \frac{\beta_t}{c} \to^d B_{\beta_t}.$$

This equation is well-defined as long as $t \le T_{00} = T_0/C_1^2 = TC_2^2/C_1^2$. Our final step is to use (2.9) in order to get

$$\mathrm{TV}^c(X, t) - \frac{\langle X\rangle_t}{c} \to^d B_t.$$

So far we have obtained convergence in the space $\mathcal{C}([0; T_{00}], \mathbb{R})$. Taking the initial value of T larger (which is possible as our diffusion is well defined on the whole line) we can obtain the convergence in $\mathcal{C}([0; T], \mathbb{R})$.

We are yet to remove assumption (3.36). For any $N > 0$ we put

$$\sigma^N(x) := \begin{cases} \sigma(x), & \text{if } |x| \le N, \\ \sigma(N), & \text{if } x > N, \\ \sigma(-N), & \text{if } x < -N, \end{cases} \qquad \mu^N(x) := \begin{cases} \mu(x), & \text{if } |x| \le N, \\ \mu(N), & \text{if } x > N, \\ \mu(-N), & \text{if } x < -N. \end{cases}$$

We define a family of diffusions by

$$\mathrm{d}X_t^N := \sigma^N(X_t^N)\mathrm{d}W_t + \mu(X_t^N)\mathrm{d}t, \qquad X_0^N = 0.$$

We assume that this diffusion is driven by the same W as in (1.4) and that X, X^N are coupled in such a way that $X_t^N = X_t$ and $\langle X_t^N\rangle = \langle X_t\rangle$ whenever $t \le \tau^N := \inf\left\{t \ge 0 : |X_t^N| > N\right\} = \inf\{t \ge 0 : |X_t| > N\}$. The solution of (1.4) is a continuous process and exists on the whole line; therefore for any $T > 0$ we have

$$1_{\{\tau^N \le T\}} \to_{N\to+\infty} 0, \quad \text{a.s.}$$

We notice now that X^N fulfills (3.36); hence the thesis of Theorem 5 is already proved for it. The quantities studied in the proof are equal for X^N and X on the set $\left\{\tau^N \le T\right\}$. Using the metric-theoretic arguments as in the proof of Fact 25 one easily concludes the proof. □

4. Proof of Theorem 1

As indicated in the Introduction the proof splits into two parts. In the first one we will prove Theorem 1 in the case when X is a Wiener process with a drift. This will serve as a key step for the second part of the proof in which, using time change techniques we will elevate the result to a general class of semimartingales.

4.1. Proof for the Wiener process with drift

This is much simpler compared to the proof of Lemma 19; therefore we provide only a sketch leaving details to the reader. Let X be a Wiener process with drift, i.e.

$$X_t := W_t + \mu t,$$

for a standard Wiener process W and $\mu \in \mathbb{R}$. We have the following lemma.

Lemma 28. *Let $T > 0$ and let $\{X\}_{t\in[0;T]}$ be a Wiener process with drift. Then*

$$\lim_{c\searrow 0} cTV^c(X,t) \to \langle X\rangle_t, \quad a.s.$$

The converge is understood in the $\mathcal{C}([0;T],\mathbb{R})$ topology.

Proof. First, we recall $S_c(n)$ defined in (3.5) and $Z_i(c)$ given by (3.13). We want to show that process $X_t(c) := cS_c(\lceil g(c)t\rceil)$ converges to a linear function. Let us consider

$$M_n(c) := c\sum_{i=1}^{n}(Z_i(c) - \mathbb{E}Z_i(c)).$$

It is a centered martingale. Differentiation of (3.18) yields that (we have $a = 1, b = 0$ in this case)

$$\mathbb{E}(Z_i(c) - \mathbb{E}Z_i(c))^2 = \frac{2\cosh(2c\mu)\sinh(c\mu)^2}{\mu^2} = 2c^2 + O(c^3).$$

Therefore, by the Doob inequality and (3.17) we have

$$\mathbb{E}\left(\sup_{t\in[0;T]}[cS_c(\lceil g(c)t\rceil) - c\lceil g(c)t\rceil\mathbb{E}Z_i(c))]^2\right) \le LTc^2,$$

for some constant L. Using (3.19) and (3.20) one obtains

$$X_t(c) \to id, \quad \text{a.s.},$$

where $id(t) = t$ and the convergence holds in $\mathcal{C}([0;T],\mathbb{R})$ topology. Now one proves an analogous convergence for process $V_c(\lceil g(c)t\rceil)$, which is, roughly speaking, the inverse of M_c. To finish the proof one needs to argue similarly as in the second part of the proof of Lemma 19. □

4.2. Proof for semimartingales

Now we assume that $X_t = X_0 + M_t + A_t$, where M is a continuous local martingale and A is a process with bounded variation.

To avoid notational inconveniences we assume that M, A are defined on $[0;+\infty)$ (one can simply put a constant process after T). Let us now introduce an additional standard Brownian motion β independent of X and denote

$$X^\epsilon := X + \epsilon\beta, \quad \epsilon > 0.$$

This is a simple trick to avoid the case when $\langle X\rangle$ is not strictly increasing. Indeed we have $\langle X_t^\epsilon\rangle = \langle X_t\rangle + \epsilon t$. Obviously this is a strictly increasing function. Moreover its inverse, denoted by α, is almost surely Lipschitz with constant smaller than ϵ^{-1}. Let us denote $M^{\epsilon} := M + \epsilon\beta$. The DDS theorem [11, Theorem V.1.6] ensures that there exists a Brownian motion B such that

$$M_t^{\epsilon} := B_{\langle X_t^{\epsilon}\rangle}, \quad t \in [0; T].$$

Using (2.9) we have

$$c\mathrm{TV}^c\left(X^{\epsilon}, t\right) = c\mathrm{TV}^c\left(B_t + A_{\alpha_t}, \langle X^{\epsilon}\rangle_t\right). \tag{4.1}$$

Let us fix $N > 0$. Applying (2.14) to the paths of $B_t + A_{\alpha_t}$ we get

$$|\mathrm{TV}^c\left(B_t + A_{\alpha_t}, \langle X_t^{\epsilon}\rangle \wedge N\right) - \mathrm{TV}^c\left(B_t, \langle X_t^{\epsilon}\rangle \wedge N\right)| \le \mathrm{TV}\left(A_{\alpha_t}, \langle X^{\epsilon}\rangle_t \wedge N\right).$$

Using Lemma 28 and the above estimate one gets that

$$c\mathrm{TV}^c\left(B_t + A_{\alpha_t}, \langle X_t^{\epsilon}\rangle \wedge N\right) \to_{c\searrow 0} \langle X_t^{\epsilon}\rangle \wedge N \quad \text{a.s.}$$

where convergence is understood in $\mathcal{C}([0; T], \mathbb{R})$ topology. Moreover, the limit agrees with the limit of (4.1) on the set $\{\langle X^{\epsilon}\rangle \le N\}$. Hence we obtain

$$c\mathrm{TV}^c\left(X^{\epsilon}, t\right) \to \langle X^{\epsilon}\rangle, \quad \text{a.s.}$$

Our aim now is to get rid of ϵ. We fix some $\alpha \in (0, 1)$ and notice that by (2.13) we have

$$c\mathrm{TV}^c(X, t) \le c\mathrm{TV}^{\alpha c}\left(X^{\epsilon}, t\right) + c\mathrm{TV}^{(1-\alpha)c}(\epsilon B, t).$$

Therefore we have

$$\begin{aligned}\limsup_{c\searrow 0} c\mathrm{TV}^c(X, t) &\le \alpha^{-1}\langle X_t^{\epsilon}\rangle + (1-\alpha)^{-1}\epsilon^2 t \\ &= \alpha^{-1}\langle X_t\rangle + \left((1-\alpha)^{-1} + \alpha^{-1}\right)\epsilon^2 t.\end{aligned}$$

By converging $\epsilon \to 0$ and $\alpha \to 1$ one can obtain

$$\limsup_{c\searrow 0} c\mathrm{TV}^c(X, t) \le \langle X_t\rangle, \quad \text{a.s.}$$

Analogously one obtains a lower-bound for lim inf. Therefore we proved that for any $t > 0$ we have

$$\lim_{c\searrow 0} c\mathrm{TV}^c(X, t) = \langle X_t\rangle, \quad \text{a.s.}$$

This is a one dimensional convergence but one easily extends it to the finite dimensional one. Moreover, since the trajectories are almost surely increasing the finite dimensional convergence can be upgraded to the functional one. This follows by the simple fact that if $f_n \in \mathcal{C}([0; T], \mathbb{R})$ is a sequence of continuous increasing functions converging point-wise to a continuous function then the convergence is in fact uniform.

In order to prove the convergence for UTV^c it suffices to use (3.35). DTV^c follows similarly.

5. Proof of large times results

In this section we will only prove Theorem 12. It follows by a similar argument as in the proof of Lemma 19. This time c is fixed and n will go to infinity. The analogs of (3.13) and (3.14) are given by

$$Z_i(n) := n^{-1/2} Y_i(c)$$

and

$$D_i(n) := n^{-1}(T^c_{U,i} - T^c_{U,i-1}), \quad i \geq 1, \quad \text{and} \quad D_0(n) := n^{-1}T^c_{U,0}.$$

By (3.20) (with $a = n^{1/2}$, $b = 0$) we have

$$\frac{\mathbb{E}Z_i(n)}{\mathbb{E}D_i(n)} = n^{-1/2}\mu\coth(c\mu).$$

We define $X_i(n) := Z_i(n) - (\mathbb{E}Z_i(n)/\mathbb{E}D_i(n))D_i(n)$. Repeating calculations as in the proof of Lemma 19 one obtains

$$\mathrm{Var}(X_i(n)) = \frac{3 + \cosh(2c\mu) - 4c\mu\coth(c\mu)}{n\mu^2}.$$

One checks that $\mathrm{Var}(X_i(n))/\mathbb{E}D_i(n) = \left(\sigma^c_\mu\right)^2$ as in Theorem 12. Now in order to obtain this theorem it is enough to apply Fact 18. This is an easy task. Above we have already checked that (A1), (A2) and (A3) are trivial. (A4) holds with any $\delta > 0$.

We skip the proof of 11 which is a simpler version of the proof in Section 4.1. Proofs of Theorem 13 follows similarly to the one above with an exception that $Z_i(n) = M^c_i - m^c_i - c$ in the case of UTV^c and $Z_i(n) = M^c_i - m^c_{i+1} - c$ in the case of DTV^c.

Acknowledgments

The authors would like to thank Radosław Adamczak for help in the proof of Lemma 27. The authors also thank the anonymous referee for useful comments. The second author was partially supported by the National Science Centre under decision no. DEC-2011/01/B/ST1/05089 and by the Ministry of Science grant N N201 397537.

This research was supported in part by the National Science Centre under decision no. DEC-2011/01/B/ST1/05089.

References

[1] Patric Billingsley, Convergence of Probability Measures, John Wiley, New York, 1968.

[2] Stewart N. Ethier, Thomas G. Kurtz, Markov processes: Characterization and convergence, in: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986.

[3] Jean Jacod, Albert N. Shiryaev, Limit Theorems for Stochastic Processes, in: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer, Berlin, Heidelberg, New York, 2003.

[4] Olav Kallenberg, Foundations of Modern Probability, in: Probability and its Applications, Springer, 1997.

[5] R. Łochowski, On truncated variation of Brownian motion with drift, Bull. Pol. Acad. Sci. Math. 56 (3–4) (2008) 267–281.

[6] R. Łochowski, On pathwise uniform approximation of processes with càdlàg trajectories by processes with minimal total variation, eprint arXiv:1106.3199, 2011.

[7] R. Łochowski, On pathwise uniform approximation of processes with càdlàg trajectories by processes with finite total variation, eprint arXiv:1106.2630v4, 2012.

[8] R. Łochowski, Truncated variation, upward truncated variation and downward truncated variation of Brownian motion with drift—Their characteristics and applications, Stochastic Process. Appl. 121 (2011) 378–393.

[9] D.L. Ocone, A symmetry characterization of conditionally independent increment martingales, in: Barcelona Seminar on Stochastic Analysis (St. Feliu de Guíxols, in: Progr. Probab., vol. 32, Birkhäuser, Basel, 1993, pp. 147–167.

[10] Bernt Øksendal, Stochastic differential equations, in: An Introduction with Applications, sixth edn., Universitext. Springer-Verlag, Berlin, 2003.
[11] Daniel Revuz, Marc Yor, Continuous Martingales and Brownian Motion, in: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1991.
[12] A. Roginsky, A central limit theorem for cumulative processes, Adv. in Appl. Probab. 26 (1994) 104–121.
[13] Dmitrii S. Silvestrov, Limit theorems for randomly stopped stochastic processes, in: Probability and its Applications (New York), Springer-Verlag, London Ltd., London, 2004.
[14] H.M. Taylor, A stopped Brownian motion formula, Ann. Probab. 3 (1975) 234–246.
[15] L. Vostrikova, M. Yor, Some invariance properties (of the laws) of Ocone's martingales, in: Séminaire de Probabilités, XXXIV, in: Lecture Notes in Math., vol. 1729, Springer, Berlin, 2000, pp. 417–431.