Semimartingale decomposition of convex functions of continuous semimartingales by Brownian perturbation

Nastasiya F. Grinberg
Department of Statistics, University of Warwick
N.Grinberg@warwick.ac.uk

May 3, 2010

In this note we prove that the local martingale part of a convex function $f$ of a $d$-dimensional semimartingale $X = M + A$ can be written in terms of an Itô stochastic integral $\int H(X) \, dM$, where $H(x)$ is some particular measurable choice of subgradient $\nabla f(x)$ of $f$ at $x$, and $M$ is the martingale part of $X$. This result was first proved by Bouleau in [2]. Here we present a new treatment of the problem. We first prove the result for $\tilde{X} = X + \epsilon B$, $\epsilon > 0$, where $B$ is a standard Brownian motion, and then pass to the limit as $\epsilon \to 0$, using results in [1] and [4].

1 Introduction

Consider a general convex function $f : \mathbb{R}^d \to \mathbb{R}$, not necessarily everywhere differentiable. Every differentiable point $x \in \mathbb{R}^d$ has a unique tangential hyperplane, while at non-differentiable points there will be a whole variety of different supporting hyperplanes. For a continuous semimartingale $X$ with decomposition $X = M + A$ we prove that the (local) martingale part of $f(X)$ can be expressed in terms of a stochastic integral of a measurable selection of subgradient $\nabla f(x)$ against $M$. For piecewise linear 1-dimensional convex functions this follows from the Meyer-Tanaka formula. In particular for $f(x) = |x|$ we have $\nabla f(x) = \text{sgn}(x)$, where $\text{sgn}(x) = -1$ if $x \leq 0$ and 1 otherwise, which is just the left hand side derivative. So at the origin, which is the only point where derivative is not defined, we can take the supporting line to be $y = -x$.

The main result of this note is the following

**Theorem 1.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex and let $X$ be a continuous $\mathbb{R}^d$-valued semimartingale with Meyer decomposition $X_t = X_0 + M_t + A_t$ which is defined on filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$. Then $f(X_t)$ is again a continuous semimartingale; in particular its local martingale part is given by

$$
\int_0^t \nabla f(X_s) \, dM_s
$$

where $\nabla f(x)$ is some choice of subgradient of $f$ at $x$, such that $f(X_t) \in m\mathcal{F}_t$ for all $t \geq 0$.

The first part of the theorem stating that $f(X_t)$ is a semimartingale was proved by Meyer [11] and later by Carlen and Protter [4]. Meyer just proves that $f(X_t)$ is a semimartingale, while Carlen and Protter express the martingale and finite variation process.
parts of the decomposition in terms of certain limits. Neither of the papers however give an explicit semimartingale decomposition of \( f(X_t) \). In [2] Bouleau took a step further and proved that at each \( x \in \text{dom}(f) \) there exists a choice \( H(x) \) of subgradient \( \nabla f(x) \) of \( f \) such that the martingale part of the decomposition of \( f(X_t) \) can be expressed as an Itô stochastic integral \( \int H(X) \, dM \). In the follow-up paper [3] he proves the conjecture stated in [2] that in fact any measurable choice of \( H(x) \) can be used. In this note we are proving the first of the two results using an approach completely different to that in [2].

There are many other papers on extending the Itô’s formula by considering different classes of functions \( f \) or stochastic processes or both. In [13], for example, Russo and Vallois derive Itô’s formula for \( C^1(\mathbb{R}^d) \)-functions of continuous semimartingales whose time-reversals are also continuous semimartingales. They also extend the formula to the case of \( C^1(\mathbb{R}^d) \)-functions with first order derivatives being Hölder-continuous with any parameter and the process given by a stochastic flow generated by a so-called \( C^0(\mathbb{R}^d, \mathbb{R}^d) \)-semimartingale. In both cases the quadratic variation process is expressed in terms of the generalised quadratic covariation process \( \langle f'(X), X \rangle_t \) introduced by the authors in an earlier paper [12] (see also paper by Fuhrman and Tessitore [9], where authors extend the notion of the generalised quadratic covariation further to the infinite-dimensional case and non-differentiable functions). In [8] Föllmer, Protter and Shiryaev consider the case of an absolutely continuous function \( f \) with a locally square integrable derivative and \( X \) a 1-dimensional Brownian motion, for which a version of Itô’s formula is derived with the finite variation part expressed again in terms of the quadratic covariation \( \langle f'(B), B \rangle_t \).

The multidimensional case (where \( f \)) is treated in \([7]\). In [10] Kendall discusses semimartingale decomposition of \( r(B) \), where \( r \) is the distance function of Brownian motion on a manifold. The problem tackled in [10] is similar to ours as \( r \) fails to be differentiable on a set of measure zero, called the cut-locus. It is proved in [10] that \( r(B) \) is a semimartingale and its canonical decomposition is found explicitly in the sequel [5].

The layout of the paper is as follows. In section 2 and 3 we introduce some notation and preliminary results concerning convex functions, including some important theorems on differentiability: in particular in section 3 we explain that a proper convex function is everywhere differentiable (i.e. has a unique supporting hyperplane) except on a set of measure zero. Hence, by virtue of observing that a Brownian perturbation of our semimartingale \( \tilde{X}^{(\epsilon)}_t = X_t + \epsilon B_t \) has a probability density at every time \( t \), we show that for convex \( f \) the gradient \( \nabla f(\tilde{X}^{(\epsilon)}_t) \) is defined for all \( t \) almost everywhere. To show that the martingale part of \( f(\tilde{X}^{(\epsilon)}_t) \) is \( \int \nabla f(\tilde{X}^{(\epsilon)}_t) \, d\tilde{M}^{(\epsilon)} \), where \( \tilde{M}^{(\epsilon)} = M + \epsilon B \) and \( \nabla f \) is some choice of subgradient, we approximate \( f \) by a sequence of \( C^2 \) convex functions \( f_n : \mathbb{R}^d \to \mathbb{R} \), \( n \geq 1 \). The martingale part of each \( f_n(\tilde{X}^{(\epsilon)}_t) \) is known explicitly from Itô’s formula and its convergence to \( \nabla f \) almost everywhere is ensured by results of section 3. Convergence of the stochastic integral \( \int \nabla f_n(\tilde{X}^{(\epsilon)}_t) \, d\tilde{M}^{(\epsilon)} \) to \( \int \nabla f(\tilde{X}^{(\epsilon)}_t) \, d\tilde{M}^{(\epsilon)} \) is ensured by a result of Carlen and Protter [4]. We conclude by proving the convergence \( \lim_{\epsilon \to 0} \int \nabla f(\tilde{X}^{(\epsilon)}_t) \, d\tilde{M}^{(\epsilon)} = \int \nabla f(X) \, dM \).

Section 4 deals with a special case when \( f \) is piecewise linear. By proving a generalised version of Meyer-Tanaka formula we find the local martingale part of \( f(X_t) \) and thus prove Theorem 1 for such \( f \). We conclude by giving a particular example of a subgradient that satisfies Theorem 1.

2 Convex functions: some notation and results

In order to prove the main result of this report we need to introduce some notation
and results from the differential theory of convex functions. Proofs of the results stated in this section and more details on convex functions are given in [13]. Let \( f \) be any function living on \( \mathbb{R}^d \) and taking values in \([-\infty, +\infty]\). At any point \( x \in \mathbb{R}^d \) we define the one-directional derivative of \( f \) with respect to a vector \( y \in \mathbb{R}^d \), if it exists, as follows

\[
Df(x)[y] := \lim_{\lambda \to 0} \frac{f(x + \lambda y) - f(x)}{\lambda}
\]

The two sided derivative at \( x \) in direction \( y \) exists if and only if \( Df(x)[y] = -Df(x)[-y] \) (1)

Now if the function \( f \) is convex, then the one-directional derivative always exists and moreover we may write

\[
Df(x)[y] = \inf_{\lambda > 0} \frac{f(x + \lambda y) - f(x)}{\lambda}
\] (2)

Furthermore \( Df(x)[y] \) is positively homogeneous (i.e. \( Df(x)[\lambda y] = \lambda Df(x)[y] \) for \( \lambda \in (0, \infty) \)) and convex in \( y \) with \( Df(x)[0] = 0 \) [13, Thm. 23.1] and

\[
Df(x)[y] \geq -Df(x)[-y]
\] (3)

We also mention the upper semicontinuity of the one-sided derivative of a convex \( f \) with respect to \( y \):

\[
Df(x)[y] = \lim_{z \to y} Df(x)[z] \geq \lim_{z \to y} Df(x)[z] \quad \text{for any} \quad x \in \mathbb{R}^d
\]

If for a general \( f \) all directional derivatives exist and are two-sided and finite then we define the gradient of \( f \) at \( x = (x_1, \ldots, x_d) \) by

\[
\nabla f(x) := \left( \frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_d}(x) \right)
\]

and for any non-zero vector \( y = (y_1, \ldots, y_d) \) we define its directional version by

\[
\langle \nabla f(x), y \rangle := \frac{\partial f}{\partial x_1}(x)y_1 + \ldots + \frac{\partial f}{\partial x_d}(x)y_d
\]

Also note \( Df(x)[y] = \langle \nabla f(x), y \rangle \) for all \( y \).

Of course a general convex function \( f \) is not necessarily everywhere differentiable, a simple example being \( f(x) = |x| \) which is not differentiable at \( x = 0 \). We can however define a set of subgradients at such a “troublesome” point.

**Definition 2.** Let \( f : \mathbb{R}^d \to \mathbb{R} \) be a convex function. A subgradient \( \nabla f(x) \) of \( f \) at \( x \in \mathbb{R}^d \) is a gradient of an affine hyperplane \( h(x) = \alpha + \beta^T x, \alpha, \beta \in \mathbb{R}^d \), passing through the point \( (x, f(x)) \) and satisfying

\[
h(x) \leq f(x)
\]

for all other \( x \). We denote any subgradient at \( x \) with respect to \( y \in \mathbb{R}^n \) by \( \langle \nabla f(x), y \rangle \).
We say \( h(x) \) is a **supporting hyperplane** of \( f \) at point \((x, f(x))\). Clearly at differentiable points \( h(x) \) is unique and is just the tangent of \( f \). Conversely, at points where \( f \) is not differentiable we can construct infinitely many tangential hyperplanes \( h(x) \). The set of all subgradients at \( x \) is called the **subdifferential** of \( f \) at \( x \), denoted \( \partial f(x) \). Hence a convex function with finite values is subdifferentiable everywhere. In subsequent sections we will need the following result

**Theorem 3.** ([13, Thm. 23.2]) Let \( f \) be a convex function and \( x \) a point at which \( f \) is finite. Then \( \nabla f(x) \) is a subgradient of \( f \) at \( x \) if and only if

\[
Df(x)[y] \geq \langle \nabla f(x), y \rangle \quad \forall y \in \mathbb{R}^d \setminus \{0\}
\]  

Relation (4) is called the subgradient inequality and can be used as an alternative definition of a subgradient.

### 3 Differential theory of convex functions

This short section is devoted to studying the set \( \mathcal{D} \) of points in the domain of \( f \) at which the supporting hyperplane is unique. It is known [13, Thm. 25.2] that in order to have a unique supporting hyperplane it suffices for the partial derivatives with respect to the basis vectors of \( \mathbb{R}^d \) to exist. Furthermore it turns out [13, Thm. 25.4] that the set of points at which \( f \) fails to have a two-sided directional derivative has measure zero and \( Df(x)[y] \) is a continuous function of \( x \) on the set \( \mathcal{D}_y, y \neq 0 \), of points at which \( Df(x)[y] = -Df(x)[-y] \).

Now suppose \( \{e_1, ..., e_d\} \) is a basis of \( \mathbb{R}^d \). Let \( \mathcal{D}_i \) be a subset of \( \mathbb{R}^d \) which consists of points where the two-sided derivative in the direction of \( e_i \) \( \partial f/\partial x_i(e_i) \) exists and let \( \mathcal{D}_i^c \) be its compliment in \( \text{int}(\text{dom}f) \). Here \( \text{dom}f \) is the effective domain of \( f \), which consists of values of \( x \) at which \( f(x) \) is finite, i.e. \( \text{dom}f = \{x \in \mathbb{R}^d; -\infty < f(x) < +\infty\} \). Then \( \mathcal{D} = \mathcal{D}_1 \cap ... \cap \mathcal{D}_d \). Now by [13, Thm. 25.4] \( \mathcal{D}_i^c \) has measure zero for all \( i \). But \( \mathcal{D}^c = \mathcal{D}_1^c \cup ... \cup \mathcal{D}_d^c \) is then the union of null sets is also a null set. Hence \( \mathcal{D}^c \) has measure zero. Finally, since each \( \partial f/\partial x_i \) is continuous on its corresponding \( \mathcal{D}_i \) (again by [13, Thm. 25.4]) \( \nabla f(x) = (\partial f/\partial x_1, ..., \partial f/\partial x_d) \) is continuous on \( \mathcal{D} \).

Thus we see that set \( \mathcal{D}^c \) of points at which a convex function \( f \) does not have a unique supporting hyperplane has measure zero. Subsequently any process which has a probability density at each time \( t \) spends time of measure zero in \( \mathcal{D}^c \), an important fact we will use in the sequel.

To prove Theorem 1 for a general convex \( f \) we will approximate it by a sequence of twice continuously differentiable convex functions \( f_n : \mathbb{R}^d \to \mathbb{R} \), to which we know Itô’s formula can be applied. So to conclude this section we state a result saying that for such a sequence \( \{f_n\}_{n \geq 1} \) with \( f_n \to f \), \( \nabla f_n(x) \) converges uniformly to \( \nabla f(x) \) for all \( x \in \mathcal{D} \).

**Theorem 4.** ([13, Thm. 25.7]) Let \( f \) be a convex function defined on \( \mathbb{R}^d \) and \( \{f_n\} \) a sequence of smooth convex functions on \( \mathbb{R}^d \) such that \( \lim_{n \to \infty} f_n(x) = f(x) \forall x \in \mathbb{R}^d \). Let \( \mathcal{D} \subseteq \mathbb{R}^d \) be the set of points where \( f \) is differentiable.

\[
\lim_{n \to \infty} \nabla f_n(x) = \nabla f(x) \quad \forall x \in \mathcal{D}
\]  

(5)
This result will be used several times in sections 5 and 6.

4 Piecewise linear convex functions and Meyer-Tanaka formula

In this section we start our analysis of the martingale part of $f(X)$. However instead of treating the case of a general convex $f$ we first prove Theorem 1 in a special case when $f$ is piecewise linear. Recall the simple one-dimensional example of $f(x) = |x|$. Suppose $X$ is a continuous martingale with canonical decomposition $X_t = X_0 + M_t + A_t$. Then we cannot apply the usual Itô’s formula to $f(X)$ since $f$ is not differentiable at $x = 0$ and so $df/dx$ is not well defined at the origin. The way around this problem is to choose the gradient at the origin from a set of possible subgradients ranging from -1 to +1. This is exactly what the Meyer-Tanaka formula does (Tanaka formula, if $X = B$ is a standard Brownian motion.)

**Theorem 5.** (Meyer-Tanaka formula for continuous semimartingales) Let $X$ be a continuous semimartingale. Define the function $\text{sgn}(x)$ to be $-1$ if $x \leq 0$ and $1$ otherwise. Then $f(X)$, where $f(x) = |x|$, is again a semimartingale and in particular

$$|X_t| = |X_0| + \int_0^t \text{sgn}(X_s)dX_s + L_t^0$$

where $L_t^0$ is the local time of $X$ at 0.

So in this case $\nabla f(x)$ is the left-hand side derivative. Moreover, because as we know Brownian motion spends zero time in Lebesgue-null sets, we can in fact choose $\nabla f(0)$ to be any number in the interval $[-1, +1]$. Using the Meyer-Tanaka formula we can prove a more general result. Namely we will prove that any piecewise linear convex function of a continuous semimartingale is itself a continuous semimartingale and find the martingale part of the decomposition explicitly.

**Proposition 6.** Let $X = (X^1, ..., X^d)$ be a continuous semimartingale living on $\mathbb{R}^d$, with $i$th component having decomposition $X^i_t = X^i_0 + M^i_t + A^i_t$, $i \in \{1, ..., d\}$. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function defined by $f(x) = l_1(x) \lor ... \lor l_k(x)$, $x \in \mathbb{R}^d$, where $l_i(x) = \alpha_i + \sum_{j=1}^d \beta_{ij}x_j = \alpha_i + \beta^T_i x$, $\alpha_i, \beta_i \in \mathbb{R}^d$, $i \in \{1, ..., k\}$ and $x \lor y := \sup\{x, y\}$. Then $f(X)$ is a semimartingale with decomposition

$$f(X_t) = f(X_0) + \sum_{i=1}^k \int_0^t 1_{B_i}(X_s)\beta^T_i dX_s + \frac{1}{2}L_t$$

where $B_i = \{x : \min\{k : \sup_j l_j(x) = l_k(x)\} = i\}$ and $L_t$ is an increasing process, constant on the complement of $\{t : l_i(X_t) = l_j(X_t) \text{ for any } i \neq j\}$. In particular the local martingale part of $f(X)$ is given by

$$\sum_{i=1}^k \int_0^t 1_{B_i}(X_s)\beta^T_i dM_s$$

5
Proof. We prove the proposition for the case when $k = 2$ and any $d \geq 1$ and general case follows by induction. Consider $f(x) = l_1(x) \lor l_2(x)$. Denote $l_1(X_t) = Y_t$ and $l_2(X_t) = Z_t$. Since $X_t$ is a continuous semimartingale so are affine functionals, $Y_t$ and $Z_t$, of $X_t$. Let the corresponding decompositions be $Y = M + A$ and $Z = N + S$. Consider $f(x) = l_1(x) \lor l_2(x) = y \lor z$. We can rewrite $y \lor z$ as follows

$$y \lor z = \frac{1}{2} (|y-z| + y + z)$$

Hence, using the differential notation for simplicity, we obtain

$$d(Y_t \lor Z_t) = \frac{1}{2} d(|Y_t - Z_t| + Y_t + Z_t) = \frac{1}{2} (d(|W_t|) + dY_t + dZ_t)$$

where $W := Y - Z$, and so $W = (M - N) + (A - S)$. Using Meyer-Tanaka formula the above becomes

$$\frac{1}{2} (\text{sgn}(W_t) dW_t + dL^0_t + dY_t + dZ_t)$$

where $L^0_t$ is the local time of $W$ at 0. Next

$$\frac{1}{2} (\text{sgn}(W_t) d(M_t - N_t) + \text{sgn}(W_t) d(A_t - S_t) + d(M_t + A_t) + d(N_t + S_t) + dL^0_t) =$$

$$= \frac{1}{2} (\text{sgn}(Y_t - Z_t) + 1) dM_t - (\text{sgn}(Y_t - Z_t) - 1) dN_t +$$

$$+ (\text{sgn}(Y_t - Z_t) + 1) dA_t - (\text{sgn}(Y_t - Z_t) - 1) dS_t + dL^0_t$$

Now $\text{sgn}(W_t) = \text{sgn}(Y_t - Z_t) = 1_{[Y_t > Z_t]} - 1_{[Y_t \leq Z_t]}$ and so $\text{sgn}(W_t) + 1 = 21_{[Y_t > Z_t]}$ and $\text{sgn}(W_t) - 1 = -21_{[Y_t \leq Z_t]}$. Hence we obtain

$$d(Y_t \lor Z_t) = 1_{[Y_t > Z_t]} dM_t + 1_{[Y_t \leq Z_t]} dN_t + 1_{[Y_t > Z_t]} dA_t + 1_{[Y_t \leq Z_t]} dS_t + \frac{1}{2} dL_t =$$

$$= 1_{[Y_t > Z_t]} dY_t + 1_{[Y_t \leq Z_t]} dZ_t + \frac{1}{2} dL_t$$

or

$$Y_t \lor Z_t = Y_0 \lor Z_0 + \int_0^t 1_{[Y_s > Z_s]} dY_s + \int_0^t 1_{[Y_s \leq Z_s]} dZ_s + \frac{1}{2} dL_t$$

where $L_t$ is a continuous increasing process, constant on the complement of $\{t : l_1(X_t) = l_2(X_t)\}$. The above expression is exactly (6) for $n = 2$. Noticing that $x \lor y \lor z = (x \lor y) \lor z$, the general case follows by induction.

\[\square\]

Clearly the integrand in (7) is a measurable selection of the multivalued map $\partial f(x)$ and so Theorem 1 holds in the special case of convex piecewise linear functions. To illustrate this result we consider our simple example again: for $f(x) = |x| d = 1$, $k = 2$ and $l_1(x) = -x$ and $l_2(x) = x$ and so $B_1 = \{x : x < 0\}$, $B_2 = \{x : x \geq 0\}$ and $L_t$ is an increasing process constant on the complement of $\{t : X_t = 0\}$. 

6
This result, although not essential, is a nice warm-up before we start dealing with a more general situation in the next sections. We refer reader to [12, Ch. VI.1] for a detailed discussion of classical Tanaka and Itô-Tanaka formulas (for $d = 1$). One might also find a discussion of convex functions in section 3 of the Appendix of [12] useful.

5 Semimartingale decomposition of $f(\tilde{X}_t)$

We are now ready to start the analysis of the general case of a convex $f$ defined over the whole of Euclidian space $\mathbb{R}^d$. Let $X$ be a continuous semimartingale with decomposition $X = M + A$ defined on some filtered probability space $(\Omega, \mathcal{F}_t, P)$. Let $(\tilde{\Omega}, \tilde{\mathcal{F}}_t, \tilde{P})$ be some enlargement of this space such that $B$ is an $(\tilde{\mathcal{F}}_t)$-standard Brownian motion independent of $X$.

Define the perturbed process $\tilde{X}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}}_t, \tilde{P})$ by

$$\tilde{X}_t(\epsilon) := \tilde{X}_t := X_t + \epsilon B_t; \quad \epsilon > 0, \quad t \geq 0$$

For simplicity of notation we shall suppress the superscript $(\epsilon)$ wherever possible. For simplicity also but without loss of generality we can assume that $X_0 = \tilde{X}_0 = 0$.

In this section we find the martingale part of $f(\tilde{X}(\epsilon))$ explicitly in order to take the limit as $\epsilon \to 0$ in the next section and hence prove Theorem 1. The reasoning behind adding a small amount of Brownian motion to $X_t$ is as follows: we know very little about the behaviour of $X_t$ as it is a general semimartingale. For instance, it can at some times be trivial, i.e. constant. Hence it might spend positive amount of time in those points where $f$ is not differentiable, that is where it has more than one supporting hyperplane, with positive probability. To avoid this happening we perturb $X_t$ by adding $\epsilon B_t$. Then

Lemma 7. $\tilde{X}_t$ has a probability density at each $t > 0$ and in particular spends zero time in any null set.

Proof. It suffices to prove that $\tilde{P}(\tilde{X}_t \in N) = 0$ for any $t > 0$ and $N \subset \mathbb{R}$ with Leb$(N) = 0$. Then $\tilde{P}$ is absolutely continuous with respect to Lebesgue measure and the corresponding Radon-Nikodym derivative is the probability density of $\tilde{X}_t$. So, for any Lebesgue-null set $N$ we have

$$\tilde{P}(\tilde{X}_t \in N) = \mathbb{E}\left[\tilde{P}(\tilde{X}_t \in N | \mathcal{F}_t)\right]$$

where $\mathcal{F}_t \equiv \sigma(\{X_s; 0 \leq s \leq t\})$ and we use the tower property of conditional expectation. Next we express $\tilde{X}_t$ in terms of $X_t$ and $B_t$ and use the fact that $B_t$ is independent of $X_t$, and hence of $\mathcal{F}_t$, to obtain

$$\mathbb{E}\left[\tilde{P}(X_t + \epsilon B_t \in N | \mathcal{F}_t)\right] = \int \tilde{P}(x + \epsilon B_t \in N) d\mu_t(x)$$

where $\mu_t$ is the law of $X_t$. Observe that $\tilde{B}_t := x + \epsilon B_t$ is a Brownian motion starting at $x$ with $\langle \tilde{B}_t, \tilde{B}_t \rangle = \epsilon^2 t^2$. But we know that Brownian motion hits null-sets with probability zero. Hence the above integral is equal to zero and the lemma is proved.

In section 3 we have seen that, $D^c$, the set of points at which $f$ is not differentiable is Lebesgue-null. Consequently $\tilde{X}$ spends zero time at those “ambiguous” points in $D^c$ (or at least spends zero time traveling in “ambiguous” directions in which a gradient is not uniquely specified). Hence $\nabla f(\tilde{X})$ is almost surely everywhere defined. Because
of Lemma 7 a particular measurable choice of \( \nabla f(x) \in \partial f(x) \) at \( x \in D^c \) is therefore unimportant as it does not change the value of the stochastic integral \( \int_0^t \nabla f(X_s) dM_s \), which we will show is the martingale part of \( f(\tilde{X}) \). To do that we approximate \( f \) by a sequence of convex twice continuously differentiable functions.

Let \( \{f_n\}_{n \leq 1} \) be a sequence of \( C^2 \) convex functions defined on \( \mathbb{R}^d \). We equip the set of convex functions on \( \mathbb{R}^d \) with uniform convergence on compact sets with the corresponding metric \( \rho \), defined by \( \rho(f, g) = \sum_{k=1}^{\infty} 2^{-k} \rho_k(f, g) \) where

\[
\rho_k(f, g) = \sup_{|x| \leq k} |f(x) - g(x)| \left/ \left( 1 + \sup_{|x| \leq k} |f(x) - g(x)| \right) \right.
\]

Let \( f_n \) increase to \( f \) and \( \lim_{n \to \infty} \rho(f_n, f) = 0 \). From Theorem 4 we know that \( \nabla_n f(x) \) converges to \( \nabla f(x) \) for all \( x \in D \) and from Lemma 7 that particular choices of \( \nabla f(x) \) for \( x \in D^c \) are unimportant when we deal with Brownian perturbation \( \tilde{X} \). We now have to prove that stochastic integral \( \int_0^t \nabla f_n(X_s) dM_s \), the martingale part of \( f_n(\tilde{X}) \), converges in some sense to \( \int_0^t \nabla f(X_s) dM_s \) and that it is indeed the martingale part of \( f(\tilde{X}) \). It turns out that the convergence is in \( \mathcal{H}^1 \) norm: for a continuous semimartingale \( X \) with decomposition \( X = M + A \) we define

\[
\|X\|_{\mathcal{H}^1} = \| (M, M) \|_{\infty}^{1/2} + \int_0^\infty |dA_s| \|_{L^p}
\]

The \( \mathcal{H}^p \)-space consists of all semimartingales \( X \) such that \( \|X\|_{\mathcal{H}^p} < \infty \). Once the convergence is established, the fact that \( \int \nabla f(\tilde{X}) d\tilde{X} \) is a local martingale part of \( f(\tilde{X}) \) will follow from [4, Thm. 2] of Carlen and Protter.

Suppose \( (X^n)_{n \geq 1} \) is a sequence of continuous semimartingales with the decomposition \( X^n = X^n_0 + M^n + A^n \), such that \( \lim_{n \to \infty} \mathbb{E}[(X^n - X)^+] = 0 \). Here \( X^* = \sup_{t \in [0, T]} |X_t| \). Barlow and Protter prove ([1, Thm. 1]) that under some regulating conditions imposed on \( M^n \) and \( A^n \) not only that the limiting process \( X \) is again a continuous semimartingale but that there is also convergence of the corresponding martingale and finite variation process parts of the decompositions. More specifically the topological space \( \mathcal{H}^1 \) of local martingales is complete and so is \( A^1 \), the space of processes of finite variation with norm \( \|A\|_{A^1} = \|\int_0^\infty |dA_s|\|_{L^1} \) (see Émery [6]).

In [4] Carlen and Protter prove that the assumptions of [1, Thm. 1] are satisfied in case when the sequence of \( C^2 \) convex functions \( \{f_n\}_{n \geq 1} \) of (a not necessarily continuous) semimartingale \( X = M + A \) converges (increases) to a convex \( f \), thus making the result applicable in our situation.

We need to note the following two inequalities

\[
\sup \sup_{|x| \leq r} |\nabla f_n(x)| \leq C_r < \infty; \quad \forall r > 0 \quad (8)
\]

and

\[
\sup_{|x| \leq r} |\nabla f(x)| \leq C_r < \infty; \quad \forall r > 0 \quad (9)
\]

where \( C_r \) is some constant only depending on \( r \). To see why inequality (8) is true first notice that since \( \lim_{n \to \infty} \rho(f_n, f) = 0 \) the variation of \( f_n \) is uniformly bounded in \( n \) on \( \{|x| \leq r + 1\} \) for any \( r > 0 \). Denote this bound by \( C_r \). Let \( x_n \) be such that
∇f_n(x_n) = \sup_{|x| \leq r} |\nabla f_n(x)|
and let \( u_n := \nabla f_n(x_n)/|\nabla f_n(x_n)| \). Then

\[
|\nabla f_n(x_n)| = \langle \nabla f_n(x_n), \nabla f_n(x_n) \rangle = \langle \nabla f_n(x), u_n \rangle
\]

\[
\leq \inf_{\lambda > 0} \frac{f_n(x_n + \lambda u_n) - f_n(x_n)}{\lambda} \leq f_n(x_n + u_n) - f_n(x_n)
\]

(10)

But since \( |x_n + u_n| \leq r + 1 \) the above is less than \( C_r \) for all \( n \) and (8) follows.

Now since \( f_n \) converges to \( f \) uniformly on compact sets we also have \( f_n \to f \) pointwise. Therefore for any \( x, y \) with \( |x|, |y| < r + 1 \) the inequality \( f_n(x) - f_n(y) \leq C_r \), \( n \geq 1 \), implies \( f(x) - f(y) \leq C_r \) by virtue of taking the limit \( n \to \infty \). Expression (9) then follows by the same argument we used to prove (8).

We are now ready to prove the following

Lemma 8. The local martingale part of \( f(\tilde{X}_t) \) is given by the limit

\[
\int_0^t \nabla f(\tilde{X}_s) d\tilde{M}_s = \lim_{n \to \infty} \int_0^t \nabla f_n(\tilde{X}_s) d\tilde{M}_s
\]

(11)

locally in \( \mathcal{H}_1 \), where \( \nabla f(x) \in \partial f(x) \) is some measurable choice of subgradient of \( f \) at \( x \).

Proof. Since for each \( n \geq 1 \) \( f_n \) is in \( C^2 \), the martingale part of \( f_n(\tilde{X}) \) is given by \( \int \nabla f_n(\tilde{X}) d\tilde{M} \), where \( \tilde{M} = M + \epsilon B \). The result of Carlen and Protter [4, Thm. 2] ensures that the martingale part of the limiting process \( f(\tilde{X}_t) \) is given by the limit of \( \int \nabla f_n(\tilde{X}) d\tilde{M} \) as \( n \) tends to infinity, locally in \( \mathcal{H}_1 \). Our aim is to prove that this limit is given by \( \int \nabla f(\tilde{X}) d\tilde{M} \) for some measurable choice of subgradient \( \nabla f \in \partial f \).

Since \( \tilde{X} \) is a continuous semimartingale, by means of localisation we can assume that \( \tilde{X} \) is contained in an open ball of radius \( r > 0 \) centered at the origin, denoted by \( B(r) \). Localisation can also be done in such a way that the localised process is simultaneously in \( \mathcal{H}_1 \); we know that continuous semimartingales are at least locally in \( \mathcal{H}_1 \). In particular we assume that \( \langle \tilde{M}, \tilde{M} \rangle_t \) is bounded in \( B(r) \).

Define \( \tilde{T}_r = \inf\{t : \tilde{X}_t \notin B(r)\} \) and consider the stopped process \( \tilde{X}_{\tilde{T}_r \wedge \tilde{T}_r} \). By Lemma [7] within ball \( B(r) \) the localised process has the density. It doesn’t matter if \( \tilde{X}_{\tilde{T}_r} \) is not in \( \mathcal{D} \) since the value of \( \nabla f(\tilde{X}) \) at time \( \tilde{T}_r \) does not change the value of the integral \( \int_0^{\tilde{T}_r} \nabla f_n(\tilde{X}_t) d\tilde{M}_t \).

Notice that convergence of a continuous (local) martingale \( \tilde{M} \) in \( \mathcal{H}^p \) is equivalent to convergence of \( \langle \tilde{M}, \tilde{M} \rangle^{1/2} \) in \( \mathcal{D}^p \), and so convergence in \( \mathcal{H}^p \) implies convergence in \( \mathcal{H}_1 \) for \( 1 \leq l < p \) in this case. In our case it is easier to prove convergence (11) in \( \mathcal{H}_1^2 \) and then deduce convergence in \( \mathcal{H}_1 \). Now, for any measurable selection \( \nabla f \in \partial f \) we have

\[
\lim_{n \to \infty} \left\| \int_0^{\tilde{T}_r} \left( \nabla f_n(\tilde{X}_t) - \nabla f(\tilde{X}_t) \right) d\tilde{M}_t \right\|_{\mathcal{H}_1^2}
= \lim_{n \to \infty} \mathbb{E} \left[ \int_0^{\tilde{T}_r} \left( \nabla f_n(\tilde{X}_t) - \nabla f(\tilde{X}_t) \right)^2 d\langle \tilde{M}, \tilde{M} \rangle_t \right]^{1/2}
\]

(12)
Using inequalities (8) and (9) we see that the integrand in (12) is bounded above by \(4C_2^2\). Recall that \(\langle \tilde{M}, \tilde{M} \rangle\) must also be bounded in \(\mathcal{B}(r)\). Hence both the integrand and the expression inside the expectation sign must be bounded and we can use the bounded convergence theorem to pull the limit inside the expectation and the integral sign, so that (12) is equal to
\[
\mathbb{E} \left[ \int_0^{\tilde{T}_r} \lim_{n \to \infty} \left( \nabla f_n(\tilde{X}_t) - \nabla f(\tilde{X}_t) \right)^2 d\langle \tilde{M}, \tilde{M} \rangle_t \right]^{1/2}
\]

We can then use almost sure convergence of \(\nabla f_n(\tilde{X}_t)\) to \(\nabla f(\tilde{X}_t)\) for all \(\tilde{X}_t \in \mathcal{D}\) and the fact that particular choices \(\nabla f(\tilde{X}_t) \in \partial f(\tilde{X}_t)\) for \(\tilde{X}_t \in \mathcal{D}\) are not charged by the integral to conclude that \(\int_0^{\tilde{T}_r} \nabla f_n(\tilde{X}_s)d\tilde{M}_s\) converges to \(\int_0^{\tilde{T}_r} \nabla f(\tilde{X}_s)d\tilde{M}_s\) in \(\mathcal{H}^2\) and hence in \(\mathcal{H}^1\). This is true for any radius \(r > 0\) of the region of localisation \(\mathcal{B}(r)\) and so (11) follows.

\[\square\]

6 Proof of Theorem 1

Finally we need to derive the analogous result for our original object of interest, continuous semimartingale \(X\).

**Proof of Theorem 1.** We have \(\lim_{\epsilon \to 0} \tilde{X}^{(\epsilon)} = X\) almost surely and thus \(\lim_{\epsilon \to 0} f(\tilde{X}^{(\epsilon)}) = f(X)\) almost surely. Note that the limiting process \(\tilde{X}^{(\epsilon)}\) as \(\epsilon\) tends to zero lives in the enlarged probability space \((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\), even though the original process \(X\) is defined on \((\Omega, \mathcal{F}, \mathbb{P})\). Crucially by It\(\hat{o}\)’s lemma \(f(\tilde{X}^{(\epsilon)}_t)\) is a continuous semimartingale for every \(\epsilon > 0\). Hence we can apply result of Barlow and Protter [11 Thm. 1] if we can show that the conditions of the theorem are satisfied in our case. Luckily most of the work has been done for us in [4] and we just have to slightly amend the arguments therein.

We first need to suitably localise our process. Let \(\mathcal{B}(r)\) be an open ball of radius \(r\) and \(\mathcal{B}(r')\) an open ball of radius \(r' > r > 0\). For all \(r, r' > 0\) define stopping times \(T_r := \inf\{t: X_t \notin \mathcal{B}(r)\}\) and \(\tilde{T}_{r'} := \inf\{t: \tilde{X}_t \notin \mathcal{B}(r')\}\) and take \(T = T_r \wedge \tilde{T}_{r'}\). Assume also that \(\tilde{X}_{t \wedge T}, X_{t \wedge T} \in \mathcal{H}^1\) for all \(t \geq 0\) and in particular that \(\langle M, M \rangle_t\) is bounded in \(\mathcal{B}(r)\). We consider the stopped process \(\tilde{X}_{t \wedge T}\). Note that \(X_{t \wedge T} \in \mathcal{B}(r) \subset \mathcal{B}(r')\) and \(\tilde{X}_{t \wedge T} \in \mathcal{B}(r')\) for all \(t \geq 0\). The localised process is absolutely continuous until it is stopped at time \(T\). Again the value of \(\tilde{X}\) at \(T\) does not affect the stochastic integral \(\int_0^T \nabla f(\tilde{X}_{s \wedge T})d\tilde{M}_{s \wedge T}\).

In what follows we will need the fact that \(f\) is Lipschitz in the ball \(\mathcal{B}(r)\) for any \(r > 0\) (see for example [13 Thm. 24.7]). Hence by calculation similar to [10] we have
\[
|\nabla f(\tilde{X}_{t \wedge T})| \leq C_{r'} < \infty; \quad t \geq 0
\]
where \(C_{r'}\) is some finite constant only depending on \(r'\).

Now to apply results of Barlow and Protter we need to prove
\[
\lim_{\epsilon \to 0} \mathbb{E} \left[ \sup_{t \leq T} |f(\tilde{X}_t^{(\epsilon)}) - f(X_t)| \right] = 0 \tag{14}
\]

\[
\sup_{\epsilon} \mathbb{E} \left[ \sup_{t \leq T} |\tilde{N}_t^{(\epsilon)}| \right] \leq K_{r,r'} \tag{15}
\]

\[
\sup_{\epsilon} \mathbb{E} \left[ \int_0^T |d\tilde{S}_s^{(\epsilon)}| \right] \leq K_{r,r'} \tag{16}
\]

where \( \tilde{N}^{(\epsilon)} := \tilde{N} \) and \( \tilde{S}^{(\epsilon)} := \tilde{S} \) are the martingale and finite variation parts of the semimartingale decomposition of \( f(\tilde{X}) \) respectively and \( K_{r,r'} \) is some finite constant which only depends on \( r \) and \( r' \).

We first look at expression (14). Since \( f \) is Lipschitz in the ball \( B(r') \) we have

\[
\sup_{t \leq T} |f(\tilde{X}_t) - f(X_t)| \leq \epsilon K_{r'} \sup_{t \leq T} |B_t|
\]

where \( K_{r'} < \infty \) is a Lipschitz constant depending on \( r' \). Taking the limit \( \epsilon \to 0 \) gives (14).

To prove (15) first note that by Lemma 8 for each \( \epsilon > 0 \) the martingale part of \( f(\tilde{X}_t) \) is \( \tilde{N}^{(\epsilon)} = \int \nabla f(\tilde{X}_t) d\tilde{M}_t \) where \( \nabla f \) is some measurable choice of subgradient. Then by the Burkholder-Davis-Gundy inequality we have for some constant \( p < \infty \)

\[
\mathbb{E} \left[ \sup_{t \leq T} |\tilde{N}_t| \right] \leq p \mathbb{E} \left[ (\tilde{N}, \tilde{N})^{1/2} T \right]
\]

\[
= p \mathbb{E} \left[ \left( \int_0^T |\nabla f(\tilde{X}_t)|^2 d(\tilde{M}, \tilde{M})_t \right)^{1/2} \right]
\]

\[
\leq pC_{r'} \mathbb{E} \left[ (\tilde{M}, \tilde{M})^{1/2} T \right]
\]

where the second inequality follows by inequality (13). To finish we need to bound \( \langle \tilde{M}, \tilde{M} \rangle_T \) by some constant independent of \( \epsilon \). Now \( \langle \tilde{M}, \tilde{M} \rangle_T = \langle M, M \rangle_T + \epsilon T \) which for sufficiently small \( \epsilon \), and hence eventually for all \( \epsilon \), is less than \( \langle M, M \rangle_T + T \) which is in turn bounded above by \( \langle M, M \rangle_T + T_r \), since \( T_r \geq T = T_r \wedge \tilde{T}_r \). Moreover recall that \( \langle M, M \rangle_{T_\wedge r} \), and hence \( \langle M, M \rangle_T \), is bounded in \( \mathcal{B}(r) \) by some constant only depending on \( r \). Thus for all sufficiently small \( \epsilon \) \( \langle \tilde{M}, \tilde{M} \rangle_T \) is bounded above by some constant that only depends on the radius \( r \) (and not on \( \epsilon \)) and (15) follows.

Proof of (16) mimics the argument in Carlen and Protter [4, pp. 4-5], modulo obvious simplifications to allow for the fact that our case is continuous and using the fact that

\[
|f(\tilde{X}_{T_\wedge r}) - f(\tilde{X}_{0})| \leq K_{r'} |\tilde{X}_{T_\wedge r} - \tilde{X}_{0}| \leq r' K_{r'}
\]

for some constant \( K_{r'} < \infty \) by Lipschitz-continuity of \( f \) in \( \mathcal{B}(r') \).
All is left to prove now is that this limit is given by

\[ \lim_{\epsilon \to 0} \int_0^t \nabla f(\tilde{X}^{(s)}) d\tilde{M}^{(s)} = \int_0^t \nabla f(X_{s \land T}) dM_{s \land T}, \]

in \( \mathcal{H}^1 \).

Proving the above convergence will require us to consider the limit of \( \nabla f(\tilde{X}^{(s)}) \) as \( \epsilon \) tends to 0. Of course in general when \( \lim_{n \to \infty} x_n = x \) for \( x \in \text{dom} f \), \( x_n \in \text{dom} f \) \( \forall n \geq 1 \), \( \lim_{n \to \infty} \nabla f(x_n) \) need not exist. However the situation when \( x_n = x + \epsilon_n y \) for some \( y \in \mathbb{R}^d \) and \( \epsilon_n \to 0 \) as \( n \to \infty \), i.e. when \( x_n \) approaches \( x \) from a single direction \( y \), is special. In this case it is known that \( \nabla f(x_n) \) converges to the part of the boundary of \( \partial f(x) \) consisting of points at which \( y \) is normal to \( \partial f(x) \). Moreover

**Lemma 9.** Let \( f : \mathbb{R}^d \to \mathbb{R} \) be a convex function. For any \( x \in \mathbb{R}^d \), for almost all \( y \in S^{d-1} \), where \( S^{d-1} \) is the unit sphere in \( \mathbb{R}^d \),

\[ \lim_{\epsilon \downarrow 0} \nabla f(x + \epsilon y) \]

exists, belongs to \( \partial f(x) \) and is unique for any selection \( \nabla f(x + \epsilon y) \in \partial f(x + \epsilon y) \), we may make from the subdifferential of \( f \) at \( x + \epsilon y \) for any \( \epsilon > 0 \).

**Proof.** See appendix. \( \square \)

Using the above result we see that for all \( t \geq 0 \) for almost all values of \( B_t \) the limit \( \lim_{\epsilon \downarrow 0} \nabla f(X_t + \epsilon B_t) \) exists and belongs to \( \partial f(X_t) \). Denote this limit by \( \nabla f(X_t) \). Also for any path of \( X \) and \( B \) for small enough \( \epsilon \), i.e. eventually, we have \( T_r < T_{r+} \). So \( T \to T_r \) as \( \epsilon \to 0 \) a.s. and

\[ \lim_{\epsilon \downarrow 0} \nabla f(X_{t \land T} + \epsilon B_{t \land T}) = \nabla f(X_{t \land T}) \quad \text{a.s.} \]  \hspace{1cm} (18)

Again we consider convergence in \( \mathcal{H}^2 \) first and convergence in \( \mathcal{H}^1 \) follows. We have, using the fact that \( \lim_{\epsilon \to 0} \tilde{M}_{t \land T} = \lim_{\epsilon \to 0} M_{t \land T} \)

\[
\lim_{\epsilon \downarrow 0} \left| \int_0^t \nabla f(\tilde{X}_{s \land T}) d\tilde{M}_{s \land T} - \int_0^t \nabla f(X_{s \land T}) dM_{s \land T} \right|_{\mathcal{H}^2} &= \lim_{\epsilon \downarrow 0} \mathbb{E} \left[ \int_0^T \nabla f(\tilde{X}_s)^2 d\langle M_s, M_s \rangle + \int_0^{T_r} \nabla f(X_s)^2 d\langle M_s, M_s \rangle \
- 2 \int_0^\infty \nabla f(\tilde{X}_{s \land T}) \nabla f(X_{s \land T}) d\langle M_{s \land T}, M_{s \land T} \rangle \right]^{1/2} \]
\]  \hspace{1cm} (19)
Once again we can use inequality \((13)\) to see that the first and third integrands in \((19)\) are bounded above by \(C_r^2 < \infty\) and \(C_r C_{r'} < \infty\) respectively. The integrals themselves are bounded since \((M, M)\) is bounded in \(\mathcal{B}(r)\). Thus we can interchange the limit with the expectation and the integration signs. Convergence \((18)\) and the fact that \(T \rightarrow T_r\) a.s. then finally yield \((17)\). Noticing that the above is true for all \(r' > r > 0\) concludes the proof.

As was mentioned before, in \([3]\) Bouleau has proved that any measurable choice of subgradient \(\nabla f(X_t)\) works for the stochastic integral of Theorem 1. A function \(\nabla^e f(x) = \lim_{\theta \downarrow 0} \mathbb{E}[\nabla f(x + \theta N)]\) (20) where \(N\) is a standard \(d\)-dimensional Gaussian random variable, is a particular example. \(\nabla^e f(x)\) can be regarded as a sort of an average of (sub)gradients within the vicinity of \(x\). To verify that it does indeed define a subgradient of \(f\) at each \(x \in \mathbb{R}^d\) we check the subgradient inequality (4) of Theorem 3. For any \(y \in \mathbb{R}^d \setminus \{0\}\) we have

\[
\langle \nabla^e f(x), y \rangle = \langle \lim_{\theta \downarrow 0} \mathbb{E}[\nabla f(x + \theta N)], y \rangle = \lim_{\theta \downarrow 0} \mathbb{E}[\langle \nabla f(x + \theta N), y \rangle] \tag{21}
\]

Now \(N\) is almost surely finite and also \(x + \theta N \in \mathcal{B}(|x + \theta N|) \subset \mathcal{B}(|x| + |N|)\) for small enough \(\theta\) and so eventually for all \(\theta\). Hence by the Lipschitz property of \(f\) and by the subgradient inequality (4) we have

\[
\langle \nabla^e f(x + \theta N), y \rangle \leq D(x + \theta N)[y] = \inf_{\lambda > 0} \frac{f(x + \theta N + \lambda y) - f(x + \theta N)}{\lambda} \\
\leq f(x + \theta N + y) - f(x + \theta N) \leq K|y|
\]

for Lipshitz constant \(K < \infty\) depending on \(x\) and \(N\). Appealing to the bounded convergence theorem now allows us to take the limit inside the expectation in equation (21) above

\[
\langle \nabla^e f(x), y \rangle = \mathbb{E}[\langle \lim_{\theta \downarrow 0} \nabla f(x + \theta N), y \rangle] \tag{22}
\]

According to Lemma 9 \(\lim_{\theta \downarrow 0} \nabla f(x + \theta N)\) exists, is unique and belongs to \(\partial f(x)\) for almost all \(N\). Denote this limit by \(z_N\). Then (22) equals

\[
\mathbb{E}[\langle z_N, y \rangle] \leq \mathbb{E}[Df(x)[y]] = Df(x)[y]
\]

Hence we have \(\langle \nabla^e f(x), y \rangle \leq Df(x)[y]\) for any \(y \in \mathbb{R}^d \setminus \{0\}\) for all \(x\) and so \(\nabla^e f(x)\) is a well defined subgradient of \(f\).

**Appendix: Proof of Lemma 9**
Proof. First of all recall that $Df(x)[y] = \lim_{\epsilon \to 0} (f(x + \epsilon y) - f(x))/\epsilon$ is a positively homogenous function, convex in $y$ with $Df(x)[0] = 0$. Let $g(y) := Df(x)[y]$. Hence $\nabla g(\lambda y)$ exists and is unique for all $\lambda > 0$ for almost all $y \in \mathbb{R}^d$. Fix $x, y \in \mathbb{R}^d$ and without loss of generality, by adding a suitable affine function to $f$, assume that

$$f(x) = g(y) = \nabla g(y) = 0$$

The fact that the limit $\lim_{\epsilon \to 0} \nabla f(x + \epsilon y)$, if it exists, belongs to $\partial f(x)$ follows from [13, Thm. 24.5.1]. To prove existence we argue by contradiction. If theorem fails then we can find a subsequence $\epsilon_n \to 0$ and a selection $\nabla f(x + \epsilon_n y) \in \partial f(x + \epsilon_n y)$ such that

$$\lim_{n \to \infty} \nabla f(x + \epsilon_n y) = h \neq 0$$  \hspace{1cm} (23)

and also a vector $u \in \mathbb{R}^d$ with $\langle h, u \rangle > 0$. For such $u$ consider

$$\frac{f(x + \epsilon_n y + \epsilon_n u) - f(x + \epsilon_n y)}{\epsilon_n} = \lambda \frac{f(x + \epsilon_n y + \epsilon_n \lambda u) - f(x + \epsilon_n y)}{\epsilon_n \lambda}$$

Using (2) and homogeneity of $g(y)$ the above is greater or equal to

$$\frac{\lambda}{\epsilon_n} Df(x + \epsilon_n y)[\epsilon_n u] = \lambda Df(x + \epsilon_n y)[u] \geq \lambda \langle \nabla f(x + \epsilon_n y), u \rangle \geq \lambda \langle h, u \rangle + O(1)$$

where the last two inequality signs come from expressions (4) and (23) respectively. Thus we obtain

$$\frac{f(x + \epsilon_n y + \epsilon_n \lambda u) - f(x + \epsilon_n y)}{\epsilon_n} \geq \lambda \langle h, u \rangle + O(1)$$  \hspace{1cm} (24)

where $O(1) \to 0$ as $n \to \infty$. On the other hand, since $f(x) = g(y) = 0$, we have

$$\frac{f(x + \epsilon_n y) - f(x)}{\epsilon_n} = \frac{f(x + \epsilon_n y) - f(x)}{\epsilon_n} = O(1)$$  \hspace{1cm} (25)

Hence combining (24) and (25) obtain

$$\frac{f(x + \epsilon_n y + \epsilon_n \lambda u) - f(x)}{\epsilon_n} = \frac{f(x + \epsilon_n y + \epsilon_n \lambda u) - f(x + \epsilon_n y)}{\epsilon_n} + \frac{f(x + \epsilon_n y) - f(x)}{\epsilon_n} \geq \lambda \langle h, u \rangle + O(1)$$

Letting $n \to \infty$, i.e. $\epsilon_n \to 0$, the above inequality becomes

$$Df(x)[y + \lambda u] = g(y + \lambda u) \geq \lambda \langle h, u \rangle > 0$$

$$\Rightarrow \frac{g(y + \lambda u)}{\lambda} = \frac{g(y + \lambda u) - g(y)}{\lambda} \geq \langle h, u \rangle > 0$$

14
And so letting $\lambda \to 0$ one obtains
\[
\langle \nabla g(y), u \rangle \geq \langle h, u \rangle > 0
\]
But this contradicts the assumption that $\nabla g(y) = 0$.

Acknowledgements. This note is a part of my PhD thesis and therefore I would like to thank my supervisor professor Wilfrid Kendall for all the help he gave me and acknowledge funding from the Statistics department of Warwick University which supports my studies. I would also like to thank Michel Émery for helpful discussions during Probability Summer School at Saint-Flour in July 2008 and Larbi Alili and the anonymous referee for helpful comments on the earlier versions of this note.

References

[1] Martin T. Barlow and Philip Protter. On convergence of semimartingales. In Séminaire de Probabilités, XXIV, 1988/89, volume 1426 of Lecture Notes in Math., pages 188–193. Springer, Berlin, 1990.

[2] Nicolas Bouleau. Semi-martingales à valeurs $\mathbb{R}^d$ et fonctions convexes. C. R. Acad. Sci. Paris Sér. I Math., 292(1):87–90, 1981. ISSN 0151-0509.

[3] Nicolas Bouleau. Formules de changement de variables. Ann. Inst. H. Poincaré Probab. Statist., 20(2):133–145, 1984. ISSN 0246-0203.

[4] Eric Carlen and Philip Protter. On semimartingale decompositions of convex functions of semimartingales. Illinois J. Math., 36(3):420–427, 1992. ISSN 0019-2082. URL http://projecteuclid.org/getRecord?id=euclid.ijm/1255987418.

[5] Michael Cranston, Wilfrid S. Kendall, and Peter March. The radial part of Brownian motion. II. Its life and times on the cut locus. Probab. Theory Related Fields, 96(3):353–368, 1993. ISSN 0178-8051. doi: 10.1007/BF01292677. URL http://dx.doi.org/10.1007/BF01292677.

[6] Michel Émery. Une topologie sur l’espace des semimartingales. In Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78), volume 721 of Lecture Notes in Math., pages 260–280. Springer, Berlin, 1979.

[7] Hans Föllmer and Philip Protter. On Itô’s formula for multidimensional Brownian motion. Probab. Theory Related Fields, 116(1):1–20, 2000. ISSN 0178-8051.

[8] Hans Föllmer, Philip Protter, and Albert N. Shiryaev. Quadratic covariation and an extension of Itô’s formula. Bernoulli, 1(1-2):149–169, 1995. ISSN 1350-7265.
[9] Marco Fuhrman and Gianmario Tessitore. Generalized directional gradients, backward stochastic differential equations and mild solutions of semilinear parabolic equations. *Appl. Math. Optim.*, 51(3):279–332, 2005. ISSN 0095-4616. doi: 10.1007/s00245-004-0814-x. URL http://dx.doi.org/10.1007/s00245-004-0814-x.

[10] Wilfrid S. Kendall. The radial part of Brownian motion on a manifold: a semimartingale property. *Ann. Probab.*, 15(4):1491–1500, 1987. ISSN 0091-1798.

[11] Paul-André Meyer. Un cours sur les intégrales stochastiques. In *Séminaire de Probabilités, X (Seconde partie: Théorie des intégrales stochastiques, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, pages 245–400. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[12] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. ISBN 3-540-64325-7.

[13] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[14] Francesco Russo and Pierre Vallois. The generalized covariation process and Itô formula. *Stochastic Process. Appl.*, 59(1):81–104, 1995. ISSN 0304-4149. doi: 10.1016/0304-4149(95)93237-A. URL http://dx.doi.org/10.1016/0304-4149(95)93237-A.

[15] Francesco Russo and Pierre Vallois. Itô formula for $C^1$-functions of semimartingales. *Probab. Theory Related Fields*, 104(1):27–41, 1996. ISSN 0178-8051. doi: 10.1007/BF01303801. URL http://dx.doi.org/10.1007/BF01303801.