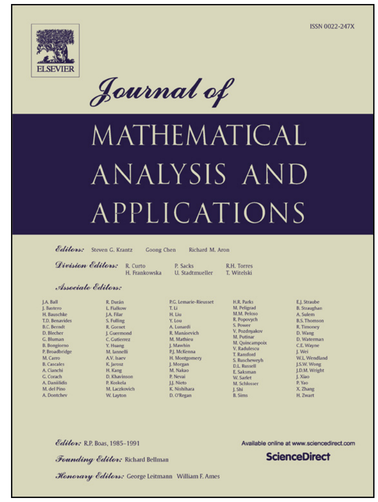


# Differentiation of measures on an arbitrary measurable space

Oleksii Mostovyi, Pietro Siorpaes

PII: S0022-247X(23)00441-9
DOI: https://doi.org/10.1016/j.jmaa.2023.127438
Reference: YJMAA 127438

To appear in: *Journal of Mathematical Analysis and Applications*

Received date: 27 November 2022

# DIFFERENTIATION OF MEASURES ON AN ARBITRARY MEASURABLE SPACE

OLEKSII MOSTOVYI AND PIETRO SIORPAES

ABSTRACT. Let $\mu, \nu$ be positive finite measures on an *arbitrary* measurable space $(\Omega, \mathcal{F})$, $\nu = \nu^a + \nu^s$ be the Lebesgue decomposition of $\nu$ with respect to $\mu$, $\mathcal{P}$ be the family $\mathcal{P}$ of all finite partitions $\pi \subseteq \mathcal{F}$ of $\Omega$, and

$$f_\pi(\mu) := \sum_{A\in\pi:\mu(A)>0} 1_A \frac{\nu(A)}{\mu(A)}, \quad \pi \in \mathcal{P}.$$

We recall that $(f_\pi(\mu))_{\pi\in\mathcal{P}} \to \frac{d\nu^a}{d\mu}$ in $L^1(\mu)$, as is (essentially) known. Here we *identify* $(\pi_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}$ such that

$$f_{\pi_n}(\mu) \to \frac{d\nu^a}{d\mu} \quad \mu \text{ a.s. as } n \to \infty;$$

in the setting of separable $\mathcal{F}$, a (rather trivial) way to do this was already known. To do all this, we characterise the case of equality in Jensen's *conditional* inequality (generalising the known case of the standard Jensen inequality), and use this to determine how, given a $p$-uniformly integrable martingale $(f_i)_{i\in I}$, one can *identify a sequence* $(i_n)_{n\in\mathbb{N}} \subseteq I$ such that $(f_{i_n})_n$ converges in $L^p$ to some $f$ which closes *the whole net* $(f_i)_i$. We also give a new proof of the (already-known) characterisation of $p$-uniformly integrable martingales, without relying on the martingale a.s. convergence theorem.

**Keywords:** Radon-Nikodym derivative, measure, closed martingale, net, convergence, convex, conditional Jensen inequality, Hilbert space.
**MSC 2020 Classification:** 60G46, 28A50, 39B62.

## 1. Introduction

The following theorem, due to Lebesgue[1], Radon and Nikodym, has been called 'probably the most important theorem in measure theory' in the classic book [12].

**Theorem 1.** *Given finite positive measures $\nu, \mu$ on a measurable space $(\Omega, \mathcal{F})$, there exists unique positive measures $\nu^a, \nu^s$ s.t. $\nu = \nu^a + \nu^s$, $\nu^a \ll \mu$ and $\nu^s \perp \mu$, and there exists unique $f \in L^1(\mu) := L^1(\Omega, \mathcal{F}, \mu)$ s.t. $\nu^a = f \cdot \mu$.*

Of course Theorem 1 admits variants for the cases of real, complex, and sigma-finite measures, which readily follow from the statement above. A way to *construct* the function $f = \frac{d\nu^a}{d\mu}$ if $\Omega = \mathbb{R}^N$ is using the following classical theorem of differentiation of measures (see [6, Chapter 1, Section 6], which calls it 'the fundamental theorem of calculus for Radon measures in $\mathbb{R}^n$').

[1]Although this is commonly referred to as the Radon-Nikodym theorem, the first version of the existence of the density of a measure on $\mathbb{R}^n$ absolutely continuous with respect to the Lebesgue measure, is due to Lebesgue; Radon extended this result to Radon measures, and Nikodym to general measures (see [3, footnote 18, p. 155]). Moreover, the existence of the decomposition $\nu = \nu^a + \nu^s$ is also due to Lebesgue.

**Theorem 2.** *Given $\mu, \nu$ positive Borel measures on $\mathbb{R}^n$, finite on compacts, let*[2]

$$h_\epsilon(x) := \frac{\nu(B_\epsilon(x))}{\mu(B_\epsilon(x))}, \quad \text{for } B_\epsilon(x) := \{y \in \mathbb{R}^n : ||y - x|| \leq \epsilon\}, \quad \epsilon > 0.$$

*Then $h_\epsilon \to \frac{d\nu^a}{d\mu}$ $\mu$ a.e. as $\epsilon \downarrow 0$, and $\mu$ a.e. $x$ is a Lebesgue point of $f := \frac{d\nu^a}{d\mu}$, i.e.*

$$\lim_{\epsilon \downarrow 0} \frac{1}{\mu(B_\epsilon(x)))} \int_{B_\epsilon(x)} |f(y) - f(x)|\mu(dy) = 0 \qquad \text{for } \mu \text{ a.e. } x,$$

*and more generally if $f := \frac{d\nu^a}{d\mu} \in L^p(\mu)$ for $p \in [1, \infty)$ then*[3]

$$\lim_{\epsilon \downarrow 0} \frac{1}{\mu(B_\epsilon(x)))} \int_{B_\epsilon(x)} |f(y) - f(x)|^p \mu(dy) = 0 \qquad \text{for } \mu \text{ a.e. } x \tag{1}$$

Theorem 2 confirms that the function $f := \frac{d\nu^a}{d\mu}$, defined via the identity $\nu^a = f \cdot \mu$, does indeed correspond to the more intuitive notion of derivative of measures $f(x)$ as a limit of the form $\lim_{i \in I} \frac{\nu(B_i(x))}{\mu(B_i(x))}$, for some family of sets $B_i(x) \ni x, i \in I$ which shrinks to $x$ in the limit.

Even if on a general measurable space one cannot consider the balls $B_\epsilon(x)$, Theorem 2 admits an analogue which holds without the metric structure, and which is reasonably well known to probabilists (though not to analysts). Such analogue is obtained by calculating $\lim_{\pi \in \mathcal{P}} \frac{\nu(B_\pi(x))}{\mu(B_\pi(x))}$, where $\mathcal{P} := \mathcal{P}(\mathcal{F})$ is the family of all finite partitions $\pi \subseteq \mathcal{F}$ of $\Omega$, $B_\pi(x)$ is the element of $\pi \in \mathcal{P}$ which contains $x$, and $\frac{\nu(B_\pi(x))}{\mu(B_\pi(x))}$ is defined to be 0 on any $B_\pi(x)$ such that $\mu(B_\pi(x)) = 0$. Indeed, when $\mathcal{F}$ is *separable*, i.e. there exists a sequence of sets $(B_j)_{j \in \mathbb{N}} \subseteq \Omega$ s.t. $\mathcal{F} = \sigma((B_j)_{j \in \mathbb{N}})$, and $\pi_n$ is the[4] partition of $\Omega$ s.t. $\sigma(\pi_n) = \sigma((B_j)_{j=0}^n)$, then it is known that

$$f_{\pi_n} := \frac{\nu(B_{\pi_n}(x))}{\mu(B_{\pi_n}(x))} = \sum_{A \in \pi_n : \mu(A) > 0} 1_A(x) \frac{\nu(A)}{\mu(A)} \to \frac{d\nu^a}{d\mu}(x) \quad \mu \text{ a.e. } x \text{ as } n \to \infty, \tag{2}$$

and if $\nu \ll \mu$ the convergence is also in $L^1(\mu)$; if instead $\nu \not\ll \mu$, $\nu_{|\sigma(\pi_n)} \ll \mu_{|\sigma(\pi_n)}$ then $(f_{\pi_n})_n$ is a martingale but it is *not* uniformly integrable, and so it does *not* converge in $L^1(\mu)$, see [15, Theorem 5.2.6].

For general (not necessarily separable) $\mathcal{F}$, it is known that if $\nu \ll \mu$ holds then the net $(\frac{\nu(B_\pi(x))}{\mu(B_\pi(x))})_{\pi \in \mathcal{P}}$ is a uniformly integrable martingale, and in this case such net converges to $\frac{d\nu}{d\mu}$ in $L^1(\mu)$ (all these results can be found in [2, Chapter 5, Items 56-57]; see also [16, Chapter 14, Section 13],[5, Theorem 1.3.2],[13, proof of Theorem 19.2]). While we were not able to find in the literature any explicit statement about the convergence of $\left(\frac{\nu(B_\pi(x))}{\mu(B_\pi(x))}\right)_{\pi \in \mathcal{P}}$ (or sequences thereof) when $\mathcal{F}$ is not separable

[2] Here we use the convention that $h_\epsilon(x) := \infty$ for all $x$ for which $\mu(B_\epsilon(x)) = 0$.

[3] Here we also use the convention that the term on the left of the = symbol in eq. (1) equals $\infty$ at all $x$ for which $\mu(B_\epsilon(x)) = 0$ for some $\epsilon > 0$.

[4] The elements of $\pi_n$ are the atoms of $\sigma((B_j)_{j=0}^n)$, and are the sets of the form $\cap_{j=0}^n C_j$ where $C_j \in \{B_j, \Omega \setminus B_j\}$.

and $\nu \not\ll \mu$, it is easy[5] to see that also in this case

$$\left(\frac{\nu(B_\pi(x))}{\mu(B_\pi(x))}\right)_{\pi\in\mathcal{P}} \to \frac{d\nu}{d\mu}(x) \quad \text{in } L^1(\mu(dx)).$$

This may seem surprising, since in general $\left(\frac{\nu(B_\pi)}{\mu(B_\pi)}\right)_{\pi\in\mathcal{P}}$ is not[6] uniformly integrable; yet there is no contradiction, because if a net $(g_i)_{i\in I} \subseteq L^1$ is converging in probability, then the fact that it is converging in $L^1 \iff$ it is uniformly integrable holds if $I = \mathbb{N}$ or if $(g_i)_{i\in I}$ is a martingale [13, Theorems 16.6 and 19.4], but in general only the implication $\Longleftarrow$ holds (taking $(g_i)_{i\in\mathbb{Z}}$ such that $(g_n)_{1\leq n\in\mathbb{N}}$ is converging in $L^1$ but $(g_{-n})_{n\in\mathbb{N}}$ is not uniformly integrable shows that $\Longrightarrow$ fails).

In theorem 4, one of our two main contributions, we generalise the above theorem of differentiation of measures from the setting of separable $\mathcal{F}$ to the case of an arbitrary $\sigma$-algebra $\mathcal{F}$, by identifying a sequence $(\pi_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}(\mathcal{F})$ such that

$$f_{\pi_n}(\mu) := \sum_{A\in\pi_n:\mu(A)>0} 1_A \frac{\nu(A)}{\mu(A)} \to \frac{d\nu^a}{d\mu} \quad \mu \text{ a.s..}$$

Essentially the same proof also shows that $(f_\pi(\mu))_{\pi\in\mathcal{P}} \to \frac{d\nu^a}{d\mu}$ holds in probability $\mu$, even if $\nu \not\ll \mu$; this argument, unlike that in footnote 5, does not make use of the existence of the Lebesgue decomposition (which is instead obtained as a bi-product of our proof); however, it only yields convergence in probability, not in $L^1$.

To prove our result on the differentiation of measures we use theorem 12, our second main contribution, which shows how to identify, given a $p$-uniformly integrable martingale $(f_i)_{i\in I}$ with an arbitrary (upward-directed) time-index $I$, an appropriate sequence $(i_n)_{n\in\mathbb{N}} \subseteq I$ such that $(f_{i_n})_n$ converges in $L^p$ to some $f$ which closes *the whole net* $(f_i)_i$ (not just the sequence $(f_{i_n})_{n\in\mathbb{N}}$). Importantly, this fact easily implies (see theorem 19) the (well-known) fact that $(f_i)_i \to f$ in $L^p$, and since to follow our whole proof of theorem 12 one does not need to know anything about nets (one just needs to check that the definition of convergence of a net is satisfied), this approach provides an elementary proof of this fact. This highlights how theorem 12 allows to study martingales with an arbitrary time-index $I$, by reducing to martingales indexed by $\mathbb{N}$ (i.e. sequences). In our opinion, this is a much more transparent approach than the only alternative known to us, which is presented in [11, Lemma V-1-1, Proposition V-1-2] and which is, in the words of [16, Section 14.13], ‘a piece of abstract nonsense’.

To prove theorem 12 we need to consider the case of equality in Jensen’s *conditional* inequality. While the case of equality in Jensen’s inequality (the standard one) was already known [9, Remark following Exc.3.34], though it does not appear

[5] Indeed, since this fact holds for $\nu^a$, for any $\epsilon > 0$ there exists $\pi_\epsilon \in \mathcal{P}$ such that

$$\left\|\frac{\nu^a(B_\pi)}{\mu(B_\pi)} - \frac{d\nu}{d\mu}\right\|_{L^1(\mu)} < \epsilon, \quad \text{for all } \pi \in \mathcal{P} \text{ such that } \pi \supseteq \pi_\epsilon;$$

since $\mu \perp \nu^s$, there exists $S \in \mathcal{F}$ such that $\mu(S) = 0 = \nu^s(\Omega \setminus S)$; then

$$\pi_\epsilon^S := \{P\cap S, P\setminus S : P \in \pi\} \setminus \{\emptyset\} \quad \text{satisfies} \quad \mathcal{P} \ni \pi_\epsilon^S \supseteq \pi_\epsilon$$

and $\frac{\nu^s(B_\pi)}{\mu(B_\pi)} = 0$ for all $\pi \in \mathcal{P}$ such that $\pi \supseteq \pi_\epsilon^S$, proving the thesis.

[6] This happens for example if $\exists S \in \mathcal{F}$ and $\pi_n \in \mathcal{P}$ such that $\nu(S) \neq 0 = \mu(S)$, $\nu_{|\sigma(\pi_n)} \ll \mu_{|\sigma(\pi_n)}$ and $S \in \mathcal{H} := \sigma(\cup_n \sigma(\pi_n))$, in which case $\left(\frac{\nu(B_{\pi_n})}{\mu(B_{\pi_n})}\right)_n$ is not uniformly integrable, as it follows applying [15, Theorem 5.2.6] with $\mathcal{F}$ replaced by $\mathcal{H}$, and so such is also $\left(\frac{\nu(B_\pi)}{\mu(B_\pi)}\right)_{\pi\in\mathcal{P}}$.

in most books on measure theory, we were not able to locate such result for the *conditional* Jensen's inequality anywhere in the literature, and so we investigate it here. Since this is a result of independent interest, we strived to give a statement at the appropriate level of generality, without introducing the unnecessary simplifying assumptions (on finiteness and integrability) which plague the existing literature on (the two versions of) Jensen's inequality. We then show with counter-examples that our assumptions cannot simply be dispensed with.

In the proof of theorem 12 we also use the (well-known) characterisation of $p$-uniformly integrable martingales (as those which are $L^p$-convergent, or equivalent which are closed by a function in $L^p$), applied to martingale sequences. We give a new proof of such characterisation, which is shorter and more elementary than the only alternative one of which we are aware of, since we completely avoid the martingale convergence theorem (which states a.s. convergence). Instead we rely on Hilbert spaces arguments for $L^2$-bounded martingales, and to reduce the general case to the $L^2$ case by a truncation argument. We were pleasantly surprised by the fact that such a proof proved possible. As mentioned above, we then use theorem 12 to easily extend such characterisation of $p$-uniformly integrable martingales to the case of an arbitrary time-index $I$, see theorem 19.

## 2. Main theorems

In this section we state in detail and discuss all our main theorems. First, we need to recall a few definitions and notations. We consider throughout the paper (real-valued) martingales indexed by a (general) *upward-directed* set $I$. In other words, $(I, \leq)$ is a partially ordered set such that for any $i, j \in I$ there exists $k \in I$ such that $i \leq k, j \leq k$. Concretely, we will be interested in the case where $I$ is $\mathcal{P}(\mathcal{F})$ (the family of all finite partitions $\pi \subseteq \mathcal{F}$ of $\Omega$), endowed with the order of the inclusion of sets (i.e. $\pi \leq \pi'$ if $\pi \subseteq \pi'$). Given an increasing[7] family of sub $\sigma$-algebras $(\mathcal{F}_i)_{i\in I}$ on the probability space $(\Omega, \mathcal{F}, \mu)$, and a family $(f_i)_{i\in I} \subseteq L^1(\mu)$, we say that $(f_i, \mathcal{F}_i)_{i\in I}$ is a *martingale* (with index set $I$) if $f_i$ is $\mathcal{F}_i$-measurable and satisfies $\mathbb{E}[f_j|\mathcal{F}_i] = f_i$ for all $j \geq i$. A martingale $(f_i, \mathcal{F}_i)_{i\in I}$ is said to be *closed* by $g \in L^1$ if $\mathbb{E}[g|\mathcal{F}_i] = f_i$ for all $i \in I$; equivalently, if $(f_i, \mathcal{F}_i)_{i\in I\cup\{\infty\}}$ is a martingale, where by definition

$$f_\infty := g, \quad \mathcal{F}_\infty := \mathcal{F}, \quad \infty \geq i \quad \text{for all} \quad i \in I.$$

A family $(g_i)_{i\in I}$ of random variables is said to be *$p$-uniformly integrable* if $(|g_i|^p)_{i\in I}$ is uniformly integrable. A family $(f_i)_{i\in I} \subseteq \mathcal{T}$ of elements of some topological space $\mathcal{T}$ is called a *net*; such net is said to converge to $f \in \mathcal{T}$ if for every neighbourhood $U(f)$ of $f$ there exist $j \in I$ such that for every $i \in I, i \geq j$ one has $f_i \in U(f)$. The space of (equivalence classes of) real-valued random variables $L^0(\mu)$ is endowed with the convergence in probability $\mu$, which is metrisable.

**Definition 3.** Given a function $a : I \to [-\infty, \infty]$, we will say that a sequence $(i_n)_{n\in\mathbb{N}} \subseteq I$ *asymptotically maximises* $a$ if

$$\sup_{i\in I} a_i = \sup_{n\in\mathbb{N}} a_{i_n}.$$

[7] I.e. $\mathcal{F}_i \subseteq \mathcal{F}_j \subseteq \mathcal{F}$ for all $i \leq j$.

2.1. **On the differentiation of measures.** We need to recall a few well known facts. If $\pi = \{A_k\}_{k=1,\dots,m}$ is a *finite partition of* $(\Omega, \mathcal{F})$ (i.e. $\pi \subseteq \mathcal{F}$ and it is a finite partition of $\Omega$), define

$$f_\pi(\mu) := \sum_{k:\mu(A_k)>0} 1_{A_k} \frac{\nu(A_k)}{\mu(A_k)} = \sum_{A\in\pi:\mu(A)>0} 1_A \frac{\nu(A)}{\mu(A)}. \tag{3}$$

Notice that, if we restrict the measures $\mu, \nu$ to the finite $\sigma$-algebra $\sigma(\pi)$, which is the family of unions of elements of $\pi$, then trivially the union $B$ of all the sets $A \in \pi$ such that $\mu(A) > 0$ belongs to $\sigma(\pi)$, and it's complement $B^c := \Omega \setminus B$ is the biggest $\mu_{\sigma(\pi)}$-null set (meaning it is a null set and $\mu(C) = 0, C \in \sigma(\pi)$ imply $C \subseteq B^c$), and thus $\nu(B \cap C) = 0$. Thus, the Lebesgue decomposition of $\nu_{|\sigma(\pi)}$ into $(\nu_{|\sigma(\pi)})^a + (\nu_{|\sigma(\pi)})^s$ exists and is given by

$$(\nu_{|\sigma(\pi)})^a = \nu_{|\sigma(\pi)}(B \cap \cdot) = f_\pi(\mu) \cdot \mu_{|\sigma(\pi)}, \quad (\nu_{|\sigma(\pi)})^s = \nu_{|\sigma(\pi)}(B^c \cap \cdot), \tag{4}$$

so in particular

$$f_\pi(\mu) = \frac{d(\nu_{|\sigma(\pi)})^a}{d\mu_{|\sigma(\pi)}}. \tag{5}$$

We now state our theorem on the differentiation of measures.

**Theorem 4.** *If $\nu, \mu$ are finite positive measures on $(\Omega, \mathcal{F})$ then $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ converges to $\frac{d\nu^a}{d\mu}$ in probability $\mu$. Moreover, a* sequence *$\pi_n \in \mathcal{P}, n \in \mathbb{N}$ such that $f_{\pi_n}(\mu) \to \frac{d\nu^a}{d\mu}$ in probability $\mu$ (and thus*[8] *also $\mu$ a.e.) can be identified as follows. Let $f_\pi(\gamma)$ be defined via (2) using $\gamma := \mu + \nu$, i.e.*

$$f_\pi(\gamma) := \sum_{A\in\pi:\mu(A)>0} 1_A \frac{\nu(A)}{\gamma(A)}, \tag{6}$$

*and choose*[9] *$\pi_n \in \mathcal{P}, n \in \mathbb{N}$ to be increasing and such that $(f_{\pi_n}(\gamma))_n$ asymptotically maximises the function $\mathcal{P} \ni \pi \mapsto \int_\Omega f_\pi(\gamma)^2 d\gamma$.*

*Remark* 5. Since $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ is converging in $L^0$, it follows from [11, Lemma V-1-1] that $(f_{\pi_n}(\mu))_n$ converges in $L^0$ for *any* increasing sequence $(\pi_n)_n \subseteq \mathcal{P}$; however, in general the limit is not $\frac{d\nu^a}{d\mu}$. As we will prove in remark 44, any increasing $(\pi_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}$ such that $\frac{d\nu^a}{d\mu}$ is measurable with respect to $\bigvee_n \sigma(\pi_n)$ is such that $f_{\pi_n}(\mu) \to \frac{d\nu^a}{d\mu}$ in a.s.. An important added value of theorem 4 is that we identify $(\pi_n)_n$ (such that the limit of $(f_{\pi_n}(\mu))_n$ is $\frac{d\nu^a}{d\mu}$) in a more constructive way, by asking that $(\pi_n)_n$ asymptotically maximises a specific function of $\pi$ whose values can be calculated from the inputs $\mu, \nu$, rather than from $\frac{d\nu^a}{d\mu}$, which is not known a priori.

*Remark* 6. It is easy to write variants of theorem 4, by considering $\pi_n$ determined by different choices of the function $x \mapsto x^2$ and the measure $\gamma$. For one example, see our alternative proof of item 3 in remark 9; for another, notice that theorem 4 holds if $x \mapsto x^2$ is replaced by any[10] strictly convex function $\phi : \mathbb{R} \to \mathbb{R}$, as it follows using theorem 12 instead of corollary 29 in its proof.

[8]The a.e. convergence follows from the martingale convergence theorem; to prove convergence in probability we can instead rely on our much simpler corollary 29.

[9]As we will see in lemma 25, such $\pi_n$ can easily be shown to exist.

[10]The family $(\phi(f_\pi))_{\pi\in\mathcal{P}}$ is always uniformly integrable, since $(f_\pi(\gamma))_{\pi\in\mathcal{P}}$ is a *bounded* martingale.

*Remark* 7. Our proof of theorem 4 automatically shows the existence of the Lebesgue decomposition $\nu = \nu^a + \nu^s$ of $\nu$ with respect to $\mu$, and the existence of $\frac{d\nu^a}{d\mu}$; in particular it proves the Radon-Nikodym theorem for positive finite $\nu \ll \mu$.

*Remark* 8. A disadvantage of considering the whole net $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ instead of the sequence $(f_{\pi_n}(\mu))_n$ is that one cannot expect[11] essential convergence of $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ to hold, since uniformly integrable martingales with arbitrary time-index, unlike those indexed by $\mathbb{N}$, in general do not converge essentially (even when their time-index is countable [5, Example 4.2.1]), unless some covering conditions are satisfied (for a discussion of covering conditions see [5, Chapter 4], or also [11, Chapter V.1]). We remind the reader that essential convergence is the proper generalisation of a.e. convergence when dealing with nets (instead of sequences) of random variables, and is defined as follows. A net $(f_i)_{i\in I} \subseteq L^0(\mu)$ is said to *converge essentially* to $f_\infty \in L^0(\mu)$ if

$$f_\infty = \operatorname*{ess\,inf}_{s\in I} \operatorname*{ess\,sup}_{t\in I, t\geq s} f_t = \operatorname*{ess\,sup}_{s\in I} \operatorname*{ess\,inf}_{t\in I, t\geq s} f_t \quad \text{a.s.},$$

where ess inf denotes the essential infimum, i.e. the infimum in $L^0(\mu)$, and analogously for ess sup.

*Remark* 9. To further the analogy with theorem 2, we mention here that under the assumptions of theorem 4 the following results also hold. We do not state them as part of theorem 4 to clarify that they are not new. From now on we assume that $\mu$ is a probability, which is without loss of generality (since we can reduce to it by dividing $\mu$ times $\mu(\Omega)$, as the case $\mu = 0$ has no interest); we do it since we want to talk of $\mu$-martingales, which are normally defined when $\mu$ is a *probability* (a text which considers more general $\mu$ is [15, Chapter 6.1]).

(1) $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ is a positive $\mu$-supermartingale, and it is a $\mu$-martingale if and only if $\nu_{|\sigma(\pi)} \ll \mu_{|\sigma(\pi)}$ for all $\pi \in \mathcal{P}$.
(2) If $(\pi_n)_n \subseteq \mathcal{P}$ is any increasing sequence then $(f_{\pi_n}(\mu))_{n\in\mathbb{N}}$ converges $\mu$ a.s..
(3) $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ is a uniformly integrable martingale if $\nu \ll \mu$, and then $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ converges to $\frac{d\nu}{d\mu}$ in $L^1$. More generally, for $p \in [1,\infty)$, $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ is $p$-uniformly integrable martingale if and only if $\nu \ll \mu$ and $\frac{d\nu}{d\mu} \in L^p$, and in this case $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ converge in $L^p$ to $\frac{d\nu}{d\mu}$ (and thus also does $(f_{\pi_n}(\mu))_{n\in\mathbb{N}}$, for $\pi_n$ as in theorem 4).

Item 1 is trivial to prove, and is well known (see e.g. [11, Proposition 3.1.5]).

Item 2 follows from the supermartingale convergence theorem [16, Theorem 11.5], which states that $(f_{\pi_n}(\mu))_n$, being a supermartingale bounded[12] in $L^1$, converges a.e.. Notice that obtaining a.s. convergence instead of just convergence in probability comes at the cost of having to rely on the supermartingale convergence theorem, instead of on our (much simpler) corollary 29, on which the proof of theorem 4 is based. Moreover, as mentioned in remark 5, the main point of theorem 4 is not that $(f_{\pi_n}(\mu))_n$ converges, but rather the identification of its limit.

Item 3 for $p = 1$ is well known [16, Chapter 14, Section 13]. Notice that one can also prove this via a simplified variant of our proof of theorem 4, as follows. Once

[11] Unfortunately, we are not aware of an example in which essential convergence does not hold for a martingale *of the form* $(f_\pi(\mu))_{\pi\in\mathcal{P}}$, so the cited example only suggests, and does not prove, that $(f_\pi(\mu))_{\pi\in\mathcal{P}}$ does not always converge essentially.

[12] A positive supermartingale $(f_n)_{n\in\mathbb{N}}$ is always bounded in $L^1$ since $\|f_n\|_{L^1} = \mathbb{E}[f_n] \leq \mathbb{E}[f_0] < \infty$.

proved as usual that $(f_\pi(\mu), \mathcal{F}_\pi)_{\pi\in\mathcal{P}}$ is a uniformly integrable $\mu$-martingale and

$$\nu(A) = \int_A f_\pi(\mu) d\mu \quad \text{for all } A \in \mathcal{F}_\pi, \tag{7}$$

to obtain that $(f_\pi(\mu))_\pi$ converges in $L^1$ to some $f_\infty$ (so that eq. (7) shows that $\frac{d\nu}{d\mu}$ exists and equals $f_\infty$) we can simply apply theorem 12 with $\phi = \int \tan^{-1}$ (see remark 13), whereas the standard probabilistic proof of the Radon-Nikodym theorem has to rely on theorem 19.

The case of general $p \in [1, \infty)$ then easily follows using theorem 19: if $(f_\pi(\mu))_{\pi\in\mathcal{P}} \to \frac{d\nu}{d\mu}$ in $L^1(\mu)$ then the convergence holds also in $L^p$, since $\frac{d\nu}{d\mu} \in L^p$ closes $(f_\pi(\mu))_{\pi\in\mathcal{P}}$.

*Remark* 10. For an even closer analogy with theorem 2 one should consider the limit not of $f_\pi(\mu)$ but of

$$h_\pi(\mu) := \sum_{A\in\pi} 1_A \frac{\nu(A)}{\mu(A)},$$

where $\frac{\nu(A)}{\mu(A)}$ is defined to be $\infty$ whenever $\mu(A) = 0$. However, doing this makes no difference, because $f_\pi(\mu)$ differs from $h_\pi(\mu)$ only[13] on a $\mu$-null set.

*Remark* 11. The martingale-based method used to prove (2) can also be used to investigate what families of sets one can use in Theorem 2 instead of $(B_\epsilon(x))_{\epsilon>0,x\in\mathbb{R}^n}$; for an exhaustive study of the topic of derivation and its relation to martingales one can consult [7], and for a shorter and readable account of the most important results see [5, Chapter 7].

2.2. **The identification of a special sequence.** To prove theorem 4 we will make use of the following theorem 12 (or of the less general corollary 29, which is simpler to prove).

**Theorem 12.** *Consider a uniformly integrable martingale $(f_i, \mathcal{F}_i)_{i\in I}$. Let each $f_i$ take values a.s. in a closed interval[14] $J$, and $\phi : J \to \mathbb{R}$ be a strictly convex, continuous[15] function such that $(\phi(f_i))_{i\in I}$ is uniformly integrable. Then:*

*(a) There exists an* increasing *sequence $(i_n)_{n\in\mathbb{N}} \subseteq I$ which asymptotically maximises the function $I \ni i \mapsto \mathbb{E}\phi(f_i)$*
*(b) For any $(i_n)_n \subseteq I$ as in item (a)*

$$(f_{i_n})_{n\in\mathbb{N}} \text{ converges in } L^1 \text{ to a } f_\infty \text{ which closes } (f_i, \mathcal{F}_i)_{i\in I}. \tag{8}$$

*(c) If (8) holds then $(f_i)_{i\in I} \to f_\infty$ in $L^1$. More generally, if $p \in [1, \infty)$ and $(f_{i_n})_{n\in\mathbb{N}}$ is p-uniformly integrable then (8) implies $(f_i)_{i\in I} \to f_\infty$ in $L^p$.*

*Remark* 13. In theorem 12 one can consider the case $J = \mathbb{R}, \phi := \int_0^\cdot \tan^{-1}(x)dx$, which has the advantage that it works for any choice of uniformly integrable $(f_i)_i$. Other interesting examples of $J, \phi$ are:

---

[13]They are both equal to $\frac{\nu(A)}{\mu(A)}$ on every $A \in \pi$ such that $\mu(A) > 0$. On the $A \in \pi$ such that $\mu(A) = 0$ instead $f_\pi(\mu) = 0 \neq \infty = h_\pi(\mu)$, so $\{f_\pi(\mu) \neq h_\pi(\mu)\}$ equals the biggest $\mu_{|\sigma(\pi)}$-null set $B^c$ introduced shortly before eq. (4).

[14]By *interval* we mean a convex set $J \subseteq \mathbb{R}$; we do not assume that $J$ is bounded.

[15]Of course $\phi$ is automatically continuous in the interior of $J$, and upper-semicontinuous on $J$, since it is convex.

(1) $J = \mathbb{R}, \phi(x) := |x|^p$ for $p \in (1, \infty)$, if $(f_i)_i$ is $p$-uniformly integrable (which is equivalent to assuming that $(f_i)_i$ is bounded in $L^p$, by Doob's $L^p$ inequality). The case $p = 2$ is of particular importance, because the corresponding proof simplifies considerably. Thus, to ease the reader's task, we will first consider such case separately (see corollary 29).

(2) $J = [0, \infty), \phi(t) = e^{-t}$, if $(f_i)_i$ is positive: in this case $(\phi(f_i))_{i \in I}$ is (defined and) uniformly integrable (because $e^{-t}$ is bounded on $t \in (-1, \infty)$).

(3) $J = \mathbb{R}$ and a Lipschitz $\phi$: in this case $(\phi(f_i))_i$ is uniformly integrable since $|\phi(x)| \leq a + b|x|$ for all $x \in \mathbb{R}$ (for some $a, b \geq 0$), and $(f_i)_i$ is uniformly integrable. This case of course subsumes the case $\phi := \int_0^{\cdot} \tan^{-1}(x)dx$.

*Remark* 14. An important feature of theorem 12 is that our proof of its last item requires absolutely no knowledge of nets (we just directly verify that the definition of convergence of a net is satisfied), which enables us to give an elementary proof of the well-know characterisation of $L^p$-converging martingales stated in theorem 19.

*Remark* 15. The novelty of theorem 12 is *not* the fact that it implies that a uniformly integrable martingale $(f_i)_i$ converges (a well-known fact, see theorem 19); instead, it is the identification of an increasing sequence $(i_n)_n$ such that $(f_{i_n})_n$ converges to some $f$ which closes $(f_i)_i$; this in turn easily allows us to give an elementary proof of the fact that $(f_i)_i$ converges (to the same limit $f$).

2.3. **The case of equality in the conditional Jensen inequality.** To prove theorem 12 we need to characterise the case of equality in the conditional Jensen inequality, generalising the analogous result for (the standard, unconditional) Jensen's inequality found in [9, Remark following Exc.3.34]

We now first recall such inequality, which is essentially well-known [16, Section 9.7(h)], [13, Corollary 23.13]; however, all the literature we consulted only considers the case of a *finite-valued* convex function $\phi$, and makes unnecessary integrability assumptions on $f$ and $\phi(f)$. In contrast, we consider (as is standard in convex analysis) $\phi$ to belong to the set

$$\mathcal{C} := \{\phi : \mathbb{R} \to \mathbb{R} \cup \{\infty\} \text{ is proper, convex, lower-semicontinuous}\}, \tag{9}$$

and (as only occasionally done in probability) we use the notion of *extended* conditional expectation $\mathbb{E}(f|\mathcal{G})$ defined for all $f$ such that $\mathbb{E}(f^-|\mathcal{G}) < \infty$ or $\mathbb{E}(f^+|\mathcal{G}) < \infty$ as $\mathbb{E}(f|\mathcal{G}) := \mathbb{E}(f^+|\mathcal{G}) - \mathbb{E}(f^-|\mathcal{G})$.

**Theorem 16** (Jensen inequality)**.** *Consider, $\phi \in \mathcal{C}$, a $\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$, and a random variable $f$ such that $\mathbb{E}(|f||\mathcal{G}) < \infty$, so that $g := \mathbb{E}(f|\mathcal{G})$ is defined, and is a real-valued random variable. Then $\mathbb{E}[\phi(f)|\mathcal{G}], \phi(g)$ are defined and satisfy*

$$\mathbb{E}[\phi(f)|\mathcal{G}] \geq \phi(g). \tag{10}$$

*If $f \in L^1$ (resp. if $\phi(f)^+ \in L^1$) then $\phi(f)^-, \phi(g)^- \in L^1$ (resp. $\phi(g)^+ \in L^1$), so $\mathbb{E}\phi(f), \mathbb{E}\phi(g)$ are defined and satisfy*

$$\mathbb{E}\phi(f) \geq \mathbb{E}\phi(g). \tag{11}$$

Then we state our contribution, the (apparently not yet known) case of equality, at the natural level of generality.

**Theorem 17** (Equality in Jensen inequality). *As in theorem 17 assume $\phi \in \mathcal{C}, \mathbb{E}(|f||\mathcal{G}) < \infty$ and define $g := \mathbb{E}(f|\mathcal{G})$. Consider the conditions*[16]*:*

*(1)* $f = g$.
*(2)* $\mathbb{E}[\phi(f)|\mathcal{G}] = \phi(g)$.
*(3)* $\mathbb{E}\phi(f) = \mathbb{E}\phi(g)$.

*Trivially item 1 $\implies$ item 2, and if $f \in L^1$ or $\phi(f)^+ \in L^1$ then $\mathbb{E}\phi(f), \mathbb{E}\phi(g)$ are defined and item 2 $\implies$ item 3.*

*For the converse implications, let $J$ be*[17] *the smallest interval such that $f \in J$ a.s., and assume that $\phi$ restricted to $J$ is real-valued and* strictly *convex. Then*

*(a) If $\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty$ then $\mathbb{E}[|\phi(f)||\mathcal{G}] < \infty$ and item 2 $\implies$ item 1.*
*(b) If $\mathbb{E}[\phi(f)^+] < \infty$* and *$\mathbb{E}[|f|] < \infty$ then $\phi(f), \phi(g) \in L^1$, and item 3 $\implies$ item 2 $\implies$ item 1.*

*Remark* 18. We will show with counter-examples that the integrability assumptions

$$\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty, \quad \mathbb{E}[\phi(f)^+] < \infty, \quad \mathbb{E}[|f|] < \infty,$$

in theorem 17 are really needed to get the implications item 2 $\implies$ item 1 and item 3 $\implies$ item 2; this is essentially due to the fact that if $a \geq b \geq 0$ satisfy $\mathbb{E}a = \mathbb{E}b$ (resp. $\mathbb{E}[a|\mathcal{G}] = \mathbb{E}[b|\mathcal{G}]$), to conclude that $a = b$ we need to assume $a \in L^1$ (resp. $\mathbb{E}[a|\mathcal{G}] < \infty$).

2.4. **On the characterisation of $L^p$-converging martingales.** We will use theorem 12 to give a transparent proof the difficult implication (which is (3) $\implies$ (1)) of the following important characterisation of $L^p$-converging martingales [11, Proposition V-1-2], [13, Theorem 23.15]. To be clear, the whole point of using theorem 12 to prove theorem 19 is that it allows to reduce to the case of sequences (instead of nets). In fact, to prove theorem 12 we will actually use theorem 19 applied to the case of sequences (i.e. for $I = \mathbb{N}$), for which we will also provide a new proof, which does not rely on the martingale convergence theorem.

**Theorem 19.** *Given a martingale $(f_i, \mathcal{F}_i)_{i\in I}$ with $f_i \in L^p$ for all $i \in I$, where $p \in [1, \infty)$, the following are equivalent, and they imply $\mathbb{E}[f|\mathcal{G}] = g$ for $\mathcal{G} := \sigma(\cup_i \mathcal{F}_i)$:*

*(1)* $(f_i)_{i\in I}$ *is convergent in $L^p$ to some $g$.*
*(2)* $(f_i, \mathcal{F}_i)_{i\in I}$ *is closed by a $f \in L^p$.*
*(3)* $(f_i)_{i\in I}$ *is p-uniformly integrable.*

Having thus obtained an accessible (elementary and relative short) proof of Theorem 19 has the following pleasant consequence. Recall that, given $f \in L^1(\mu)$, one can easily define the conditional expectation $\mathbb{E}[f|\mathcal{H}]$ for a *finite* $\sigma$-algebra $\mathcal{H}$, replacing $f$ with its local average on every atom of $\mathcal{H}$ (i.e. on every set of the $\pi \in \mathcal{P}$ such that $\sigma(\pi) = \mathcal{H}$). One could then *define* $\mathbb{E}[f|\mathcal{G}]$ for an arbitrary $\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$ as the limit in $L^1$ of $\mathbb{E}[X|\mathcal{H}]$ as the finite sigma algebras $\mathcal{H} \subseteq \mathcal{G}$ become finer and finer. This approach has the advantage of being an intuitive way of defining $\mathbb{E}[f|\mathcal{G}]$ as a local average, certainly better from a pedagogical point of view than

[16] Of course the equalities in items 1 and 2, being equalities between random variables on $(\Omega, \mathcal{F}, \mu)$, are meant to be satisfied only $\mu$ almost surely.
[17] Such an interval exists, as it equals the closed convex hull of the support of the law of $f$ (which is a closed interval), minus its endpoints on which the law of $f$ has no atoms.

the standard way (due to Kolmogorov) to define the conditional expectation $\mathbb{E}[f|\mathcal{G}]$ as the $\mathcal{G}$-measurable $g \in L^1(\mu)$ such that

$$\int_A f d\mu = \int_A g d\mu \quad \text{for all } A \in \mathcal{G}. \tag{12}$$

Somewhat analogously, for $\nu \ll \mu$ one could define $\frac{d\nu}{d\mu}$ as the $L^1$-limit of the martingale $(f_\pi(\mu))_{\pi\in\mathcal{P}}$, instead of as the $g \in L^1(\mu)$ which satisfies $\nu = g \cdot \mu$; however, in this case the change is less compelling, as the classic definition of $\frac{d\nu}{d\mu}$ is certainly more intuitive than the classic definition of $\mathbb{E}[f|\mathcal{G}]$. The disadvantage of this alternate approach to defining $\mathbb{E}[f|\mathcal{G}]$ is that it is not obvious that such limit exists. This however trivially follows from theorem 19, as follows.

**Corollary 20.** *Given a $\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$, let $\mathbb{H}(\mathcal{G})$ be the family of all finite $\sigma$-algebras $\mathcal{H} \subseteq \mathcal{G}$, ordered by inclusion. If $f \in L^1(\mu)$ then the net $(\mathbb{E}[f|\mathcal{H}])_{\mathcal{H}\in\mathbb{H}(\mathcal{G})}$ converges in $L^1$ to $\mathbb{E}[f|\mathcal{G}]$.*

*Proof.* Trivially $(\mathbb{E}[f|\mathcal{H}])_{\mathcal{H}\in\mathbb{H}(\mathcal{G})}$ is a martingale closed by $f \in L^1$, and so the thesis follows from applying Theorem 19. □

Thus, the fact that we gave an accessible proof of corollary 20 thereof enables lecturers to effectively introduce the notion of conditional expectation as a limit. Of course Kolmogorov's definition should still be introduced, being an equivalent characterisation which is useful in some proofs[18], and whose equivalence is trivial to prove[19].

2.5. **Outline of the paper.** The first part of this paper is dedicated to the proof of theorem 12, and the second to derive its consequences theorem 4 and theorem 19.

In section 3 we prove theorem 22, which is a result about Hilbert spaces analogous to theorem 12. We then apply theorem 22 to the Hilbert space $L^2$ in section 4 to obtain corollary 29, which is the special case of theorem 12 in which one considers $L^2$-bounded martingales and $\phi(x) = x^2$. We take this approach, instead of proving directly theorem 12, because to prove theorem 4 it is enough to use the simpler corollary 29 instead of theorem 12, and for pedagogical reasons: the proof of theorem 12 is similar to that of corollary 29, but needs the following two additional results of independent interest. One, any uniformly integrable martingale sequence is converging in $L^1$ (we prove this in section 5); two, the case of equality in the conditional Jensen inequality (considered in section 6). In section 7 we finally prove theorem 12. We then apply theorem 12 to derive theorem 4 in section 8, and to derive theorem 19 in section 9.

## 3. The case of Hilbert spaces

In this section we prove theorem 22, which is a result about Hilbert spaces which is analogous to theorem 12. We denote with $(H, \langle\cdot,\cdot\rangle)$ a (real) Hilbert space; whenever we talk of convergence of elements of $H$, we mean in the norm topology.

[18]While most properties of $\mathbb{E}[X|\mathcal{G}]$ are trivial consequences of the analogous properties of $\mathbb{E}[X|\mathcal{H}]$ (which are very easy to prove), some, like the conditional dominated convergence theorem, and the conditional Fatou lemma, are better proved using Kolmogorov's characterisation.
[19]Clearly the $L^1$-limit of $(\mathbb{E}[f|\mathcal{H}])_{\mathcal{H}}$ satisfies such characterisation, and the proof of uniqueness is trivial.

Consider $f = (f_i)_{i\in I} \subseteq H$ and a family $(V_i)_{i\in I}$ of closed vector subspaces of $H$ such that

$$f_i \in V_i, \qquad V_i \subseteq V_j, \qquad \text{for all} \quad i, j \in I, i \leq j. \tag{13}$$

We will say that $(f_i, V_i)_{i\in I}$ has *increments orthogonal to the past* if

$$\langle f_k - f_j, h_j\rangle = 0 \qquad \text{for } h_j \in V_j, \quad j \leq k, \quad j, k \in I. \tag{14}$$

Sometimes, given an element $f_\infty \in H$, we will add an element, denoted by $\infty$, to the index $I$, and consider $(f_i, V_i)_{i\in I\cup\{\infty\}}$; in this case we take $V_\infty := H$, and extend the order $\leq$ from $I$ to $I \cup \{\infty\}$ by setting

$$i \leq \infty \quad \text{ for all } i \in I.$$

In this case, if $(f_i, V_i)_{i\in I}$ has increments orthogonal to the past, then so does $(f_i, V_i)_{i\in I\cup\{\infty\}}$ if and only if

$$\langle f_\infty - f_j, h_j\rangle = 0 \qquad \text{for } h_j \in V_j, \quad j \in I. \tag{15}$$

*Remark* 21. While the above definition of increments orthogonal to the past looks perhaps unintuitive, it is what we will need to treat martingales, and it is closely connected to the following more intuitive condition (of which we will never make use). Since (13) implies that $V_j$ always contains the closed vector space $V_j^f$ generated by $(f_i)_{i\leq j}$, (14) implies

$$\langle f_k - f_j, f_i\rangle = 0 \qquad \text{for } i \leq j \leq k, \quad i, j, k \in I. \tag{16}$$

Moreover, if $V_j = V_j^f$ for every $j \in I$, then conditions (14) and (16) are equivalent. However, $V_j = V_j^f$ almost never holds, and so the condition (16) is rarely enough even to treat the case of a martingale $(f_i)_i$ with its natural filtration $\mathcal{F}_j^f := \sigma((f_i)_{i\leq j}), j \in I$. Indeed, in this case we will need to consider the space $V_j := L^2(\Omega, \mathcal{F}_j^f, \mu)$, which is normally (much) bigger than $V_j^f$: for example, if $I$ is finite then each $V_j^f$ is finite dimensional, whereas $L^2(\mathcal{F}_j^f, \mu)$ is infinite dimensional unless each $f_i, i \leq j$ takes only finitely many values a.s..

We will later use the obvious fact that (16) implies that $(f_i)_i$ has *orthogonal increments*, i.e.

$$\langle f_k - f_j, f_j - f_i\rangle = 0 \qquad \text{for } i \leq j \leq k, \quad i, j, k \in I;$$

notice that the opposite implication holds if $I$ has a minimum $m$ and $f_m = 0$.

Here comes our main theorem in the setting of Hilbert spaces.

**Theorem 22.** *Assume that $(f_i, V_i)_{i\in I}$ has increments orthogonal to the past. If $(f_i)_{i\in I}$ is bounded then it converges in $H$ to some $f_\infty$ such that $(f_i, V_i)_{i\in I\cup\{\infty\}}$ has increments orthogonal to the past. Moreover $f_{i_n} \to f_\infty$ for any* increasing *$(i_n)_{n\in\mathbb{N}} \subseteq I$ such that*

$$\sup_{n\in\mathbb{N}} \|f_{i_n}\|^2 = \sup_{i\in I} \|f_i\|^2,$$

*and there always exists such a $(i_n)_{n\in\mathbb{N}}$.*

To prove the above theorem we need a couple of lemmas, the first one of which also provides a simple and elementary proof of lemma 28.

**Lemma 23.** *Let $(f_n)_{n\in\mathbb{N}} \subseteq H$ be a bounded sequence in a Hilbert space $H$. If $(f_n)_n$ has orthogonal increments then it is convergent.*

*Proof.* For $m < n$, we can write $f_n - f_m$ using the telescopic sum of the increments

$$f_n - f_m = \sum_{k=m+1}^{n} (f_k - f_{k-1}).$$

Since these are orthogonal, it follows that

$$\|f_n - f_m\|^2 = \sum_{k=m+1}^{n} \|f_k - f_{k-1}\|^2. \tag{17}$$

Since $(f_n)_n$ is bounded, it follows from (17) that $\sum_{k=1}^{\infty} \|f_k - f_{k-1}\|^2 < \infty$. Thus (17) implies that $(f_n)_n$ is Cauchy, and thus it is convergent. □

**Lemma 24.** *Assume that $(f_i, V_i)_{i \in I}$ has increments orthogonal to the past. If $(f_i)_{i\in I}$ is bounded, $(i_n)_{n\in\mathbb{N}}, (j_n)_{n\in\mathbb{N}} \subseteq I$ are increasing and $i_n \leq j_n$ for all $n$ then $(f_{i_n})_n$ and $(f_{j_n})_n$ converge in $H$ to some $f_\infty$ and $g$ such that $f_\infty$ equals the projection $P_V g$ of $g$ on the smallest closed vector subspace $V \subseteq H$ which contains every $V_{i_n}, n \in \mathbb{N}$.*

*Proof.* Since $(f_{i_n})_n$ has increments orthogonal to the past, it has orthogonal increments, so by Lemma 23 it is convergent; analogously for $(f_{j_n})_n$. Let us denote with $f_\infty$ and $g$ their limits. By assumption for any $h_k \in V_k$

$$0 = \langle f_{j_m} - f_{j_n}, h_k \rangle = \langle f_{j_n} - f_{i_n}, h_k \rangle = \langle f_{i_n} - f_{i_m}, h_k \rangle, \text{ for } \quad k \leq i_n, n \leq m,$$

so taking $m \to \infty$ we conclude

$$0 = \langle g - f_{j_n}, h_k \rangle = \langle f_{j_n} - f_{i_n}, h_k \rangle = \langle f_{i_n} - f_\infty, h_k \rangle, \text{ for } \quad k \leq i_n. \tag{18}$$

Adding all the terms in eq. (18) we get

$$\langle g - f_\infty, h \rangle = 0, \qquad \text{for all } h \in \cup_{n\in\mathbb{N}} V_{i_n}, \tag{19}$$

which trivially implies $\langle g - f_\infty, h \rangle = 0$ for all $h \in V$, i.e. $f_\infty = P_V g$. □

In the following lemma we do not ask that $(i_n)_n$ be *strictly* increasing, since in principle it can happen that $I$ admits a maximum.

**Lemma 25.** *Let $a : I \to [-\infty, \infty]$ be an* increasing *function. Then, there exists an* increasing sequence *$(i_n)_{n\in\mathbb{N}} \subseteq I$ which asymptotically maximises $a$.*

*Proof.* Since $a$ is increasing, we can choose a sequence $(k_n)_{n\in\mathbb{N}} \subseteq I$ such that $a_{k_n} \uparrow a := \sup_{i\in I} a_i$, and then use the fact that $I$ is upper-directed to define $i_n$ by induction:

$$i_0 := k_0, \quad i_{n+1} \in I \text{ chosen such that } i_{n+1} \geq i_n, k_{n+1}.$$

Such $(i_n)_n$ is increasing and $a_{k_n} \leq a_{i_n} \leq a$ for all $n$ we conclude that $a_{i_n} \uparrow a$. □

*Proof of theorem 22.* By assumption if $i, j \in I, i \leq j$ then $\langle f_j - f_i, f_i \rangle = 0$, and so

$$\|f_j\|^2 = \|f_j - f_i\|^2 + \|f_i\|^2 \geq \|f_i\|^2,$$

so the map $I \ni i \mapsto a_i := \|f_i\|^2$ is increasing. By lemma 25 we can choose an *increasing* $(i_n)_{n\in\mathbb{N}}$ which asymptotically maximises $(a_i)_{i\in I}$. By Lemma 23 $(f_{i_n})_n \to f_\infty$. Below we will show that $(f_i)_{i\in I\cup\{\infty\}}$ has increments orthogonal to the past, i.e. eq. (15) holds; from it, it follows that $\langle f_\infty - f_i, f_i - f_{i_n} \rangle = 0$ for any $i_n \leq i$, and so

$$\|f_{i_n} - f_\infty\|^2 = \|f_{i_n} - f_i\|^2 + \|f_i - f_\infty\|^2 \geq \|f_i - f_\infty\|^2,$$

and since $f_{i_n} \to f_\infty$, simply by definition of convergence we get $(f_i)_{i\in I} \to f_\infty$.

To conclude, let us prove (15). Fix an arbitrary $j \in I$. Construct by induction an *increasing* $(j_n)_n \subseteq I$ such that $j_n \geq j, i_n$ for each $n \in \mathbb{N}$ (to do so, choose $j_0 \geq j, i_0$, and then for $n \in \mathbb{N}$ choose $j_{n+1} \geq j_n, i_{n+1}$). By Lemma 23 $(f_{j_n})_n$ converges to some $g \in H$. Since $a_{i_n} \leq a_{j_n} \leq a$ and $a_{i_n} \to a$ we get $a_{j_n} \to a$ and so

$$\|f_\infty\|^2 = \lim_n \|f_{i_n}\|^2 = a = \lim_n \|f_{j_n}\|^2 = \|g\|^2. \tag{20}$$

Lemma 24 gives that $f_\infty$ is the projection of $g$ on a subspace, and so $\langle f_\infty, g - f_\infty\rangle = 0$, which implies

$$\|f_\infty\|^2 + \|g - f_\infty\|^2 = \|g\|^2,$$

so eq. (20) implies $f_\infty = g$. Since $j \leq j_n \leq j_m$ for all $n \leq m$, (16) gives that

$$\langle f_{j_m}, h_j\rangle = \langle f_{j_n}, h_j\rangle, \quad \langle f_{j_n}, h_j\rangle = \langle f_j, h_j\rangle$$

and taking $m \to \infty$ in the former we get $\langle g, h_j\rangle = \langle f_{j_n}, h_j\rangle$. From the last two equalities and $f_\infty = g$ we conclude that

$$\langle f_\infty, h_j\rangle = \langle g, h_j\rangle = \langle f_{j_n}, h_j\rangle = \langle f_j, h_j\rangle$$

and since $j \in J$ was arbitrary we conclude that (15) holds. □

## 4. The case of martingales in $L^2$

In this section we prove corollary 29, which is the special case of theorem 12 with $p = 2$ and $\phi(x) = x^2$, as a corollary of theorem 22. The connection between martingales bounded in $L^2$ and Hilbert spaces is given by the following lemma 26. In this section we consider the Hilbert space $H = L^2(\mu)$ and the $\sigma$-algebra $\mathcal{F}_\infty := \mathcal{F}$, and we will use the abbreviated notations $L^2(\mu) := L^2(\Omega, \mathcal{F}, \mu)$ and $L^2(\mathcal{F}_i, \mu) := L^2(\Omega, \mathcal{F}_i, \mu)$.

**Lemma 26.** *Consider $(f_i)_{i\in I} \subseteq L^2(\mu)$, and $\sigma$-algebras $\mathcal{F}_i \subseteq \mathcal{F}, i \in I$ such that $f_i \in L^2(\mathcal{F}_i, \mu)$ and $\mathcal{F}_i \subseteq \mathcal{F}_j$ for all $i, j \in I, i \leq j$. Then $(f_i, \mathcal{F}_i)_{i\in I}$ is a martingale if and only if $(f_i, L^2(\mathcal{F}_i, \mu))_{i\in I}$ has increments orthogonal to the past. In particular, $(f_i, \mathcal{F}_i)_{i\in I}$ is a martingale and $f_\infty \in L^2(\mu)$ closes $(f_i, \mathcal{F}_i)_{i\in I}$ if and only if $(f_i, L^2(\mathcal{F}_i, \mu))_{i\in I\cup\{\infty\}}$ has increments orthogonal to the past.*

*Proof.* Consider $i, j, k \in I$ such that $i \leq j \leq k$. Since $\mathbb{E}[\mathbb{E}[\cdot|\mathcal{F}_i]] = \mathbb{E}$, we get

$$\mathbb{E}[(f_k - f_j)g_i] = \mathbb{E}[\mathbb{E}[(f_k - f_j)g_i|\mathcal{F}_i]] = \mathbb{E}[g_i\mathbb{E}[f_k - f_j|\mathcal{F}_i]], \quad \text{for all} \quad g_i \in L^2(\mathcal{F}_i, \mu),$$

and so $(f_i, L^2(\mathcal{F}_i, \mu))_{i\in I}$ has increments orthogonal to the past iff $\mathbb{E}[f_k - f_j|\mathcal{F}_i] = 0$ for all $i \leq j \leq k$, i.e. iff $(f_i, \mathcal{F}_i)_{i\in I}$ is a martingale. This equivalence, applied to $I \cup \{\infty\}$ instead of $I$, concludes the proof. □

*Remark* 27. We could of course have defined $(f_i, V_i)_{i\in I}$ to have increments independent from the past for $f_i \in A$, $V_i$ closed subspace of $B$, where $A, B$ are two vector spaces in separating duality. In this case taking $A = L^1, V_i = L^\infty(\mathcal{F}_i, \mu)$, $B = L^\infty(\mu)$ would have given the analogue of lemma 26 for all martingales (not just martingales in $L^2$). We choose not to take this point of view since most people are more comfortable talking of $L^1$ and $L^\infty$ than of spaces in separating duality. On the other hand, it is simpler to talk of a general Hilbert space than of $L^2(\mu)$, since the latter has additional structure.

The convergence of $L^2$-bounded martingales, which is normally proved via the martingale convergence theorem [16, Chapter 14.1], and the $L^2$ version of theorem 12 for martingale sequences, now follow trivially from our results for Hilbert spaces.

**Lemma 28.** *Any martingale $(f_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ bounded in $L^2$ is convergent in $L^2$.*

*Proof.* By lemma 26 $(f_i, L^2(\mathcal{F}_i, \mu))_i$ has increments orthogonal to the past, and thus the thesis follows from lemma 23. □

**Corollary 29.** *If $(f_i)_{i\in I}$ is a $L^2$-bounded martingale then it converges in $L^2$ to some $f_\infty$ which closes $(f_i)_{i\in I}$. Moreover $f_{i_n} \to f_\infty$ in $L^2$ for any* increasing $(i_n)_{n\in\mathbb{N}} \subseteq I$ *such that*

$$\sup_{n\in\mathbb{N}} \|f_{i_n}\|_{L^2}^2 = \sup_{i\in I} \|f_i\|_{L^2}^2$$

*and there always exists such a $(i_n)_{n\in\mathbb{N}}$.*

*Proof.* The thesis follows trivially from lemma 26 and theorem 22. □

*Remark* 30. To obtain from theorem 19 with $p = 2$ the same conclusion as in corollary 29, one needs to additionally apply Doob's $L^2$ inequality to obtain that $(f_i)_{i\in I}$ is 2-uniformly integrable.

## 5. Uniformly integrable martingale sequences converge in $L^1$

The key step in proving theorem 12 for general $p \geq 1$ it to prove it for $p = 1$; to do so, we will need the following lemma 31 (which is a sub-statement of theorem 19), whose role in the proof of corollary 29 was taken up by lemma 28. This section if dedicated to its proof.

**Lemma 31.** *Any uniformly integrable martingale $(f_n)_{n\in\mathbb{N}}$ converges in $L^1$.*

Normally lemma 31 is proved using the corollary of the martingale convergence theorem which states that $(f_m)_{m\in\mathbb{N}}$ converges a.s.; we present below a novel proof, which (like our proof of lemma 28) avoids the martingale convergence theorem, and is in particular much more elementary.

While the proof of the $L^2$ analogue of lemma 31, i.e. of lemma 28, was simply obtained using Hilbert spaces, to prove lemma 31 we will first apply a weaker statement, namely that a uniformly integrable family $(f_m)_{m\in\mathbb{N}}$ admits a forward convex combination which is convergent. We then show that, if $(f_m)_m$ is a martingale, this easily implies that it is closed by some $f \in L^1$. If $f \in L^2$, since $\mathbb{E}[\cdot|\mathcal{F}_m]$ is a contraction in $L^2$ we get that $(f_m)_m = (\mathbb{E}[f|\mathcal{F}_m])_{m\in\mathbb{N}}$ is $L^2$-bounded, and so $L^2$ convergent by lemma 28. Thus, to prove convergence of $(f_m)_m$ for general $f$ we just need to apply a truncation argument to reduce to the case of $f \in L^2$. We now give the details.

The following result is given an elementary and short proof in [1, Lemma 2.1]; their proof considers Hilbert spaces and a truncation argument, just like our proof of lemmas 31 and 34. Alternatively, notice that this is a special case of Mazur's lemma, but the proof of the latter relies on the use of the weak topology, and is thus less elementary.

**Lemma 32.** *If $(f_n)_{n\in\mathbb{N}}$ is uniformly integrable, then for every $m \in \mathbb{N}$ there exists a convex combination $g_m$ of $(f_k)_{k\geq m}$ such that $g_m$ converges in $L^1$ as $m \to \infty$.*

The above result easily allows to prove the following one.

**Lemma 33.** *If* $(f_m)_{m\in\mathbb{N}}$ *is a uniformly integrable martingale then it is closed.*

*Proof.* By lemma 32 there exists a (finite) convex combination $g_m$ of $(f_k)_{k\geq m}$ such that $g_m$ converges in $L^1$ to some $f$. Since $\mathbb{E}[g_m|\mathcal{F}_k] = f_k$, taking $m \to \infty$ we get $\mathbb{E}[f|\mathcal{F}_k] = f_k$, so $f$ closes $(f_m)_m$. □

We now truncate $f$ to show that a martingale closed by $f$ is convergent.

**Lemma 34.** *If* $(f_m)_{m\in\mathbb{N}}$ *is a closed martingale then it is convergent in* $L^1$.

*Proof.* Let $f \in L^1$ close $(f_m)_{m\in\mathbb{N}}$, and consider its truncation $f^n := (f \wedge n) \vee (-n)$, for $n \in \mathbb{N}$. Writing $f = (f - f^n) + f^n$, bound $||\mathbb{E}[f|\mathcal{F}_m] - \mathbb{E}[f|\mathcal{F}_k]||_{L^1}$ above by

$$||\mathbb{E}[f - f^n|\mathcal{F}_m]||_{L^1} + ||\mathbb{E}[f - f^n|\mathcal{F}_k]||_{L^1} + ||\mathbb{E}[f^n|\mathcal{F}_m] - \mathbb{E}[f^n|\mathcal{F}_k]||_{L^1}.$$

Since $\mathbb{E}[\cdot|\mathcal{F}_m]$ and $\mathbb{E}[\cdot|\mathcal{F}_k]$ are contractions in $L^1$, we get

$$||\mathbb{E}[f|\mathcal{F}_m] - \mathbb{E}[f|\mathcal{F}_k]||_{L^1} \leq 2||f - f^n||_{L^1} + ||\mathbb{E}[f^n|\mathcal{F}_m] - \mathbb{E}[f^n|\mathcal{F}_k]||_{L^1} \tag{21}$$

Since $f^n \in L^2$, $(\mathbb{E}[f^n|\mathcal{F}_m])_{m\in\mathbb{N}}$ converges in $L^2$ by lemma 28; in particular it is Cauchy in $L^2$, thus in $L^1$ since $\|\cdot\|_{L^1(\mu)} \leq \|\cdot\|_{L^2(\mu)}$. Thus eq. (21) and $f^n \to f \in L^1$ imply that $(\mathbb{E}[f|\mathcal{F}_m])_{m\in\mathbb{N}}$ is Cauchy in $L^1$, and thus $L^1$ convergent.

□

*Proof of lemma 31.* It follows from lemmas 33 and 34. □

## 6. The conditional Jensen inequality

Notice that in the proof of theorem 22 we concluded $f_\infty = g$ from $\|f_\infty\|^2 = \|g\|^2$ and $\langle f_\infty, g - f_\infty\rangle = 0$. If we want to prove theorem 12 in full generality, this step needs to change, since we are no longer working in a Hilbert space, and so we cannot use identities involving $\|\cdot\|^2$. This step will thus be replaced by theorem 17, which characterises the case equality in the conditional Jensen inequality, to which this section is dedicated to. In the rest of the section we first provide some preliminaries about the extended conditional expectation and properties of convex functions; then we prove theorems 16 and 17; and finally we show with three counter-examples that the integrability assumptions

$$\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty, \quad \mathbb{E}[\phi(f)^+] < \infty, \quad \mathbb{E}[|f|] < \infty,$$

in theorem 17 are really needed.

*Remark* 35. Notice that to prove theorem 19 it is enough[20] to consider $f \in L^1$ and $\phi$ (real-valued and) continuously differentiable, Lipschitz (equivalently $\phi'$ is bounded), with strictly increasing derivative $\phi'$. In this case the proofs of theorems 16 and 17 simplify considerably (mostly since all quantities considered are finite), and most conveniently they only make use of the standard notion of conditional expectation (defined on $L^1$) and its properties, and of simple facts from calculus; moreover, the conditional statements (inequality, case of equality) could be easily derived from the corresponding unconditional statements by disintegration (though this is a rather overly-sophisticated approach), i.e. by considering the regular conditional probability given $\mathcal{G}$. Thus, the reader interested only to this setting can safely skip the upcoming preliminaries about the *extended* conditional expectation and convex

[20]Indeed, it is enough to consider $\phi := \int_0^{\cdot} \tan^{-1}(x)dx$.

functions, and a good part of the proof of theorem 19. Nonetheless, we prefer to state and prove theorem 17 in the appropriate generality, since it is of independent interest.

We now define the extended conditional expectation $\mathbb{E}(f|\mathcal{G})$ (similarly[21] to what done in [10, Chapter 1, eq. 1.1] and [13, Definition 23.5]), and state some of its properties (whose simple proofs appear in [13, Theorem 23.8]), which we will use without further mention. To extend $\mathbb{E}(f|\mathcal{G})$ (assuming to be initially defined only for $f \in L^1$), we first define it for all $[0,\infty]$-valued measurable $f$ as the $[0,\infty]$-valued random variable

$$\mathbb{E}(f|\mathcal{G}) := \sup_{n\in\mathbb{N}} \mathbb{E}(f \wedge n|\mathcal{G}) = \lim_n \mathbb{E}(f \wedge n|\mathcal{G}), \quad \text{if } f \geq 0;$$

we can then extend it further, and define it as the $[-\infty,\infty]$-valued random variable

$$\mathbb{E}(f|\mathcal{G}) := \mathbb{E}(f^+|\mathcal{G}) - \mathbb{E}(f^-|\mathcal{G})$$

if $f$ is $[-\infty,\infty]$-valued and either $\mathbb{E}(f^-|\mathcal{G}) < \infty$ or $\mathbb{E}(f^+|\mathcal{G}) < \infty$. The linearity property

$$\mathbb{E}[af + b|\mathcal{G}] = a\mathbb{E}[f|\mathcal{G}] + \mathbb{E}[b|\mathcal{G}] \tag{22}$$

holds whenever $\mathbb{E}(|f||\mathcal{G}) < \infty$ if the random variables $a, b$ are such that $a$ is $\mathcal{G}$-measurable, and $\mathbb{E}(|b||\mathcal{G}) < \infty$ (which in particular holds if $b$ is $\mathcal{G}$-measurable). If $f \geq g \geq 0$ then $\mathbb{E}[\mathbb{E}(f|\mathcal{G})] = \mathbb{E}[f] \in [0,\infty]$ and $\mathbb{E}(f|\mathcal{G}) \geq \mathbb{E}(g|\mathcal{G})$. Thus $\mathbb{E}[\mathbb{E}(f|\mathcal{G})] = \mathbb{E}[f]$ holds if either $f^+$ or $f^-$ is in $L^1$; and if $f \geq g$ and either $\mathbb{E}(f^+|\mathcal{G}) < \infty$ or $\mathbb{E}(g^-|\mathcal{G}) < \infty$ holds, then $\mathbb{E}(f|\mathcal{G}), \mathbb{E}(g|\mathcal{G})$ are defined and satisfy $\mathbb{E}(f|\mathcal{G}) \geq \mathbb{E}(g|\mathcal{G})$ (the case $\mathcal{G} = \{\emptyset, \Omega\}$ gives the same statement for $\mathbb{E} = \mathbb{E}[\cdot|\{\emptyset,\Omega\}]$).

We now recall some facts about convex functions of a real variable, found in most convex analysis books (see e.g. [8]). Given an *open* interval $J$ and a (strictly) convex function $\phi : J \to \mathbb{R}$, the right-derivative $\phi' := \phi'_r$ always exists (finite), it is right-continuous and (strictly) increasing, and satisfies

$$\phi(t) - \phi(s) = \int_s^t \phi'(u)du, \quad \text{for all } s, t \in J. \tag{23}$$

In general, we will consider a $\phi$ which is not necessarily real-valued. To avoid the trivial case, we only consider $\phi$ *proper*, i.e. we assume that $\phi$ is not the constant $\infty$; to avoid pathologies, we only consider lower-semicontinuous functions, leading us to consider the family $\mathcal{C}$. Recall that any $\phi \in \mathcal{C}$ can be written as $\sup_{n\in\mathbb{N}} \phi_n$ for a sequence of affine functions $\phi_n$ (i.e. functions of the form $x \mapsto ax + b$, where $a, b \in \mathbb{R}$). If $\phi \in \mathcal{C}$ then $\{\phi < \infty\}$ is (non-empty and) convex, so it is one of the sets $[a,b], (a,b], [a,b), (a,b)$, where

$$-\infty \leq a := \inf\{\phi < \infty\} \leq b := \sup\{\phi < \infty\} \leq \infty.$$

Recall that $\phi$ is continuous (also on the boundary) when restricted to $\{\phi < \infty\}$, and the right-derivative $\phi'_r$ and left-derivative $\phi'_l$ of $\phi$ always exists and are finite in the interior of $\{\phi < \infty\}$; if $\phi(a) < \infty$ then $\phi'_r(a)$ is defined and can take any value

[21]The only slight difference is that we define $\mathbb{E}(f|\mathcal{G})$ as possibly taking the values $\pm\infty$, whereas [10] only allows $\infty$ and [13] only allows real-values.

in $[-\infty, \infty)$, and analogously if $\phi(b) < \infty$ then $\phi'_l(b)$ is defined and can take any value in $(-\infty, \infty]$. We define

$$\phi'(x) := \begin{cases} -\infty & \text{for all } x \le a \\ \phi'_r(x) & \text{if } x \in \{\phi < \infty\} \cap [a, b) \\ \phi'_l(b) & \text{if } x = b \in \{\phi < \infty\} \\ \infty & \text{for all } x \ge b \end{cases} \tag{24}$$

Such $\phi'$ is increasing, and it is continuous from the right on $\{\phi < \infty\} \cap [a, b)$, and is continuous from the left at $b$ if $\phi(b) < \infty$. With such definition of $\phi'$, we will consider the following quantity:

$$\Delta\phi(s, t) := \phi(t) - \phi(s) - \phi'(s)(t - s), \tag{25}$$

which is defined if $s, t \in \{\phi < \infty\}$, making use of the conventions (standard in measure theory)

$$\pm\infty \cdot 0 = 0, \quad \pm\infty \cdot a = \pm\infty, \quad \pm\infty \cdot (-a) = \mp\infty, \text{ for all } a \in (0, \infty), \tag{26}$$

to define the term $-\phi'(s)(t - s) \in (-\infty, \infty]$.
Finally, recall that any if $\phi \in \mathcal{C}$ is bounded below by an affine function, and so if $\phi \in \mathcal{C}$ and $\mathbb{E}(|h||\mathcal{G}) < \infty$ then $\mathbb{E}(\phi(h)^-|\mathcal{G}) < \infty$, and in particular $\mathbb{E}(\phi(h)|\mathcal{G})$ is defined.

We now prove theorem 16, and then theorem 17. Though theorem 16 is a more general statement than normally found in the literature, the key step of the standard proof of eq. (10) as found in [16, Section 9.7(h)] applies unchanged; we now give the simple details, and show that eq. (11) follows easily from eq. (10) under the appropriate integrability assumption.

*Proof of theorem 16.* If $\phi$ is affine, then by linearity of the conditional expectation $\mathbb{E}[\phi(f)|\mathcal{G}] = \phi(g)$. Since we can write $\phi$ as $\sup_{n\in\mathbb{N}} \phi_n$ for a sequence of affine functions $\phi_n$, we get

$$\mathbb{E}[\phi(f)|\mathcal{G}] \ge \mathbb{E}[\phi_n(f)|\mathcal{G}] = \phi_n(g)$$

and taking $\sup_n$ we conclude that eq. (10) holds.

If we assume $\phi(f)^+ \in L^1$, then applying eq. (10) to $\phi^+$ gives $\mathbb{E}[\phi(f)^+|\mathcal{G}] \ge \phi(g)^+$, and taking expectations now shows that $\phi(g)^+ \in L^1$; in particular considering the convex function $x \mapsto |x|$ shows that $f \in L^1$ implies $g \in L^1$. Thus, if $f \in L^1$ then $\phi(f)^-, \phi(g)^- \in L^1$, since $g \in L^1$ and $\phi$ is bounded below by an affine function.

Thus, whether $\phi(f)^+ \in L^1$ or $f \in L^1$, either $\mathbb{E}\phi(f)^+$ and $\mathbb{E}\phi(g)^+$ are finite, or $\mathbb{E}\phi(f)^-$ and $\mathbb{E}\phi(g)^-$ are finite; either way, $\mathbb{E}\phi(f), \mathbb{E}\phi(g)$ are defined, so eq. (10) implies eq. (11). □

To prove theorem 17, we now provide below a different proof of theorem 16, which only works under the assumption $\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty$ a.s.; following this proof closely will allows us to characterise the case of equality in eq. (10) (and in eq. (11)). We will first need the following technical lemma, due to the fact that $\phi'$ can take the value $-\infty$ (resp. $\infty$) on the left (resp. right) endpoint of $\{\phi < \infty\}$.

**Lemma 36.** *As in theorem 17 assume $\mathbb{E}(|f||\mathcal{G}) < \infty, \phi \in \mathcal{C}$ and define $g := \mathbb{E}(f|\mathcal{G})$. If $\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty$ then*

$$\mathbb{P}(\{\phi'(g) \in \{\pm\infty\}\} \cap \{\phi(g) < \infty\} \cap \{f \ne g\}) = 0.$$

*Proof.* To fix ideas assume $\{\phi < \infty\} = [a, b)$ (the other cases are analogous). Then $\{\phi' \in \{\pm\infty\}\} \cap \{\phi < \infty\} \subseteq \{a\}$, so it is enough to show that $\{g = a\} \cap \{f \neq g\}$ is a null set. Recall the conventions in eq. (26). Notice that $\infty \cdot 1_{\{f<a\}} \leq \phi(f)^+$, and so $\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty$ implies that

$$\mathbb{E}[\infty \cdot 1_{\{f<a\}}|\mathcal{G}] = \sup_{n\in\mathbb{N}} \mathbb{E}[n \cdot 1_{\{f<a\}}|\mathcal{G}] = \infty \cdot \mathbb{E}[1_{\{f<a\}}|\mathcal{G}]$$

is finite, and thus $\mathbb{E}[1_{\{f<a\}}|\mathcal{G}] = 0$ and so $\mathbb{P}(f < a) = 0$. Thus the identity

$$\mathbb{E}[f 1_{\{g=a\}}] = \mathbb{E}[g 1_{\{g=a\}}] = \mathbb{E}[a 1_{\{g=a\}}]$$

implies that $f = a$ a.s. on $\{g = a\}$ (since the random variable $(f - a)1_{\{g=a\}}$ is positive and has 0 expectation, and so it equals 0), i.e. $\mathbb{P}(\{g = a\} \cap \{f \neq g\}) = 0$. □

*Proof of theorem 17.* Notice that $f \in J$ a.s. implies $g \in J$ a.s.. We first prove eq. (10). Since $\phi'$ is increasing, the inequality

$$\phi(t) - \phi(s) = \int_s^t \phi'(u)du \geq \int_s^t \phi'(s)du = \phi'(s)(t - s) \tag{27}$$

holds for all $s, t$ in the interior of the interval $\{\phi < \infty\}$, and so on the whole $\{\phi < \infty\}$ by taking limits; moreover if $s, t \in J$ then eq. (27) holds with equality iff $t = s$ (since $\phi'$ is strictly increasing on the interior of $J$). Using the notation $\Delta\phi$ introduced in eq. (25), this means that $\Delta\phi(s, t) \in [0, \infty]$ for all $s, t \in \{\phi < \infty\}$, and if $s, t \in J$ then $\Delta\phi(s, t) = 0$ holds iff $t = s$. Since $\phi$ is continuous and $\phi'$ is increasing they are Borel, so $\phi \circ f$ is $\mathcal{F}$-measurable, and $\phi \circ g$, $\phi' \circ g$ are $\mathcal{G}$-measurable.

Assume from now on that $\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty$, so that $\mathbb{E}[|\phi(f)||\mathcal{G}] < \infty$ (in particular $\phi(f) < \infty$). Then theorem 16 implies that $\phi(g) < \infty$, and by lemma 36

$$\phi'(g)(f - g) = \phi'(g)1_{\{\phi'(g)\in\mathbb{R}\}}(f - g),$$

so the quantity $\Delta\phi(g, f)$ is well defined and finite, and from eq. (22) it follows that

$$\mathbb{E}[\Delta\phi(g, f)|\mathcal{G}] = \mathbb{E}[\phi(f)|\mathcal{G}] - \phi(g), \tag{28}$$

and so $\Delta\phi \geq 0$ implies eq. (10).

Let us now consider the case of equality. Since a positive random variable has 0 expectation iff it equals 0, it has 0 *conditional* expectation iff it equals 0. Thus eq. (28) shows that eq. (10) holds with equality iff $\Delta\phi(g, f) = 0$, i.e. iff $f = g$ (since $f, g \in J$ a.s.).

Assume now that $\mathbb{E}[\phi(f)^+] < \infty$ and $f \in L^1$; then theorem 16 gives $\phi(g)^+, g \in L^1$. Since $\phi$ is bounded below by an affine function and $f, g \in L^1$ we get $\phi(f)^-, \phi(g)^- \in L^1$. Thus $\phi(f), \phi(g) \in L^1$, so the positive random variable $h := \mathbb{E}[\phi(f)|\mathcal{G}] - \phi(g)$ has the same expectation as $\phi(f) - \phi(g)$; so, if we assume $\mathbb{E}[\phi(f)] = \mathbb{E}[\phi(g)]$ we get that $\mathbb{E}h = 0$, and since $h \geq 0$ we conclude $h = 0$. □

We now show with counter-examples that the integrability assumptions in theorem 17 are really needed.

*Example* 37. Let's see that the implication item 3 $\Longrightarrow$ item 2 fails without the assumption $\mathbb{E}[\phi(f)^+] < \infty$ (even if $\mathbb{E}[|f|] < \infty$ holds). Consider two independent random variables $u, v$ such that

$$u, v \in L^1 \setminus L^2, \quad u \geq 0 \neq v, \quad \mathbb{E}[v] = 0,$$

and define

$$f := u + v, \quad \mathcal{G} := \sigma(u),$$

so that $\mathbb{E}[|f|] < \infty$ (and in particular $\mathbb{E}[|f||\mathcal{G}] < \infty$), $g = \mathbb{E}[f|\mathcal{G}] = u + \mathbb{E}[v] = u$. The function $\phi(x) := x^2$ is in $\mathcal{C}$ and

$$\mathbb{E}[\phi(f)|\mathcal{G}] = \mathbb{E}[u^2+v^2+2uv|\mathcal{G}] = u^2+\mathbb{E}[v^2]+2u\mathbb{E}[v] = u^2+\mathbb{E}[v^2] = \infty > u^2 = \phi(g),$$

yet taking expectations we get

$$\mathbb{E}[\phi(f)] = \mathbb{E}[u^2+v^2] = \infty = \mathbb{E}[u^2] = \mathbb{E}[\phi(g)].$$

*Example* 38. Let's see that the implication item 3 $\implies$ item 2 fails without the assumption $\mathbb{E}[|f|] < \infty$ (even if $\mathbb{E}[\phi(f)^+] < \infty$ holds). Consider two independent random variables $a, b$ such that

$$a, b \geq 0, \quad b \notin L^1 \ni a, \quad a \text{ is not a.s. constant,}$$

and define

$$f := a - b, \quad \mathcal{G} := \sigma(b).$$

Then $f^+ \in L^1$, $f^- \notin L^1$, and $\mathbb{E}[|f||\mathcal{G}] \leq \mathbb{E}[a|\mathcal{G}] + b < \infty$ since $a \in L^1$. Moreover $f$ is not (a.s. equal to a random variable which is) $\mathcal{G}$-measurable, since otherwise $a$ would be $\mathcal{G}$-measurable and independent from $\mathcal{G}$ and thus constant. Now let $g : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function such that $\frac{1}{2} \leq g \leq 2$, and define

$$c(x) := \begin{cases} 2x & \text{if } x \leq 0 \\ \frac{x}{2} & \text{if } x \geq 0 \end{cases}, \quad d(x) := \begin{cases} \frac{x}{2} & \text{if } x \leq 0 \\ 2x & \text{if } x \geq 0 \end{cases}.$$

Then the function $\phi(x) := \int_0^x g(t)dt, x \in \mathbb{R}$, is real-valued, strictly convex and Lipschitz, and satisfies $c \leq \phi \leq d$, and so

$$\phi(f)^+ \leq 2f1_{\{f \geq 0\}} = 2f^+, \quad \phi(f)^- \geq \frac{1}{2}f1_{\{f \leq 0\}} = \frac{1}{2}f^- \tag{29}$$

Since $f^+ \in L^1$ and $f^- \notin L^1$, eq. (29) implies $\phi(f)^+ \in L^1, \phi(f)^- \notin L^1$. Thus theorem 16 implies $-\infty = \mathbb{E}[\phi(f)] \geq \mathbb{E}[\phi(g)]$, and so $\mathbb{E}[\phi(f)] = \mathbb{E}[\phi(g)]$. Yet eq. (10) does not hold with equality, since otherwise theorem 17 would imply that $f = g := \mathbb{E}[f|\mathcal{G}]$, which is not possible since $f$ is not $\mathcal{G}$-measurable.

*Example* 39. Now let's see that the implication item 2 $\implies$ item 1 fails without the assumption $\mathbb{E}[\phi(f)^+|\mathcal{G}] < \infty$. Let $u, v, \mathcal{G}, f$ (and thus $g$) be as in example 37, but now consider

$$\phi(x) := \begin{cases} 0 & \text{for } x \leq -1 \\ \infty & \text{for } x > -1 \end{cases}.$$

Then $\phi \in \mathcal{C}$ and since $g = u \geq 0$ we have $\phi(g) = \infty$, and so eq. (10) implies $\mathbb{E}[\phi(f)|\mathcal{G}] = \infty = \phi(g)$, yet $g = u \neq u + v = f$.

## 7. Proof of theorem 12

This section is devoted to the proof of theorem 12, which is conveniently analogous to that of theorem 22. First, the role of lemma 23 will be taken up by lemma 31. Next, the following result is the analogue of lemma 24, since $\mathbb{E}[\cdot|\mathcal{G}]$ is the orthogonal projection of $L^2(\mathcal{F}, \mu)$ onto $L^2(\mathcal{G}, \mu)$.

**Lemma 40.** *If $(f_i, \mathcal{F}_i)_{i \in I}$ is a uniformly integrable martingale, $(i_n)_{n \in \mathbb{N}}, (j_n)_{n \in \mathbb{N}} \subseteq I$ are increasing and $i_n \leq j_n$ for all $n$, then $(f_{i_n})_n$ and $(f_{j_n})_n$ converge in $L^1$ to some $f_\infty$ and $g$ such that $\mathbb{E}[g|\mathcal{G}] = f_\infty$ for $\mathcal{G} := \sigma(f_{i_n}, n \in \mathbb{N})$.*

*Proof.* By lemma 31 $(f_{i_n})_n$ is convergent in $L^1$, and analogously for $(f_{j_n})_n$; let us denote with $f_\infty$ and $g$ their limits. The proof of lemma 24 then applies unchanged, provided that $\langle f, h\rangle$ is interpreted as $\mathbb{E}[fh]$ for $f \in \{f_\infty, g, f_i : i \in I\}$ and $h \in V_k := L^\infty(\mathcal{F}_k, \mu)$, and $P_V g$ is interpreted as $\mathbb{E}[g|\mathcal{G}]$. □

*Proof of theorem 12.* So as to be in the setting of theorems 16 and 17, we can extend $\phi$ to a function in $\mathcal{C}$, which we still improperly denote by $\phi$: for example we can set $\phi := \infty$ outside the closure of $J$, and define $\phi(l) := \phi(l+)$ (resp. $\phi(r) := \phi(r-)$) on the left-endpoint (resp. right-endpoint) of $J$ (such limits exist [8, Proposition 3.1.2]).

(Item (a)) By theorem 16 the map $I \ni i \mapsto a_i := \mathbb{E}\phi(f_i)$ is increasing, so by lemma 25 we can choose an *increasing* $(i_n)_{n\in\mathbb{N}}$ which asymptotically maximises $(a_i)_{i\in I}$, where

$$a_i := \mathbb{E}\phi(f_i),$$

so that $a_{i_n} \to a := \sup_{i\in I} a_i$.

(Item (b)) Let $(i_n)_n$ be as in the statement. Lemma 40 gives that $(f_{i_n})_n \to f$ in $L^1$. Since $(f_{i_n})_n \to f$ in probability, $f_{i_n} \in J$ a.s., $J$ is closed and $\phi$ is continuous on $J$, we conclude $(\phi(f_{i_n}))_n \to \phi(f_\infty)$ in probability, and so also in $L^1$ (by the uniform integrability assumption). To conclude, let us show that $f_\infty$ closes $(f_i, \mathcal{F}_i)_{i\in I}$. Fix an arbitrary $j \in I$. Construct by induction an *increasing* $(j_n)_n \subseteq I$ such that $j_n \geq j, i_n$ for each $n \in \mathbb{N}$ (to do so, choose $j_0 \geq j, i_0$, and then for $n \in \mathbb{N}$ choose $j_{n+1} \geq j_n, i_{n+1}$). Then lemma 40 shows that $(f_{j_n})_n$ converges in $L^1$ to some $g$ which satisfies $\mathbb{E}[g|\mathcal{G}] = f_\infty$; reasoning as for $(f_{i_n})_n$ shows that $(\phi(f_{j_n}))_n \to \phi(g)$ in $L^1$. Since $a_{i_n} \leq a_{j_n} \leq a := \sup_{i\in I} a_i$ and $a_{i_n} \to a$ we get $a_{j_n} \to a$ and so

$$\mathbb{E}\phi(f_\infty) = \lim_n \mathbb{E}\phi(f_{i_n}) = \lim_n a_{i_n} = a = \lim_n a_{j_n} = \lim_n \mathbb{E}\phi(f_{j_n}) = \mathbb{E}\phi(g),$$

and so theorem 17 yields $f_\infty = g$ (since $\mathbb{E}[g|\mathcal{G}] = f_\infty$ and $\phi(g) \in L^1$). Fix an arbitrary $h_j \in L^\infty(\mathcal{F}_j)$. Since $j \leq j_n \leq j_m$ for all $n \leq m$, the tower property of the conditional expectation gives that

$$\mathbb{E}[f_{j_m} h_j] = \mathbb{E}[f_{j_n} h_j], \quad \mathbb{E}[f_{j_n} h_j] = \mathbb{E}[f_j h_j],$$

and taking $m \to \infty$ in the former we get $\mathbb{E}[g h_j] = \mathbb{E}[f_{j_n} h_j]$. From the last two equalities and $f_\infty = g$ we conclude that

$$\mathbb{E}[f_\infty h_j] = \mathbb{E}[g h_j] = \mathbb{E}[f_{j_n} h_j] = \mathbb{E}[f_j h_j]$$

and since $j \in J$, $h_j \in L^\infty(\mathcal{F}_j)$ were arbitrary we conclude that $f_\infty$ closes $(f_i, \mathcal{F}_i)_{i\in I}$.

(Item (c)) Since $f_\infty$ closes $(f_i, \mathcal{F}_i)_{i\in I}$ we have $f_i - f_{i_n} = \mathbb{E}[f_\infty - f_{i_n}|\mathcal{F}_i]$ for $i \geq i_n$ and so, since $\mathbb{E}[\cdot|\mathcal{F}_i]$ is a contraction on $L^p$ (as it follows for example from theorem 16), it follows that

$$\|f_i - f_{i_n}\|_{L^p} = \|\mathbb{E}[f_\infty - f_{i_n}|\mathcal{F}_i]\|_{L^p} \leq \|f_\infty - f_{i_n}\|_{L^p}, \quad i \geq i_n.$$

Thus

$$\|f_\infty - f_i\|_{L^p} \leq \|f_\infty - f_{i_n}\|_{L^p} + \|f_{i_n} - f_i\|_{L^p} \leq 2\|f_\infty - f_{i_n}\|_{L^p}, \qquad i \geq i_n. \tag{30}$$

If $(f_{i_n})_{n\in\mathbb{N}}$ is p-uniformly integrable then (8) shows that it converges in $L^p$ to $f_\infty$, and so eq. (30) implies $(f_i)_i \to f$ in $L^p$ by definition of convergence of $(f_i)_i$. □

## 8. Proof of theorem 4

In this section we prove Theorem 4 using corollary 29. A useful idea, which not only allows to reduce to the case of $\nu \ll \mu$ but also allows to consider martingales which are bounded (instead of just uniformly integrable), is to replace $\mu$ with $\gamma := \mu + \nu \geq \nu$. If $\nu \ll \mu$, we can then use that the Radon-Nikodym derivative of $\nu$ and of $\gamma$ with respect to $\mu$ are linked by the following formula

$$\frac{d\nu}{d\gamma} = \frac{\frac{d\nu}{d\mu}}{1 + \frac{d\nu}{d\mu}},$$

which is obtained by formally dividing times $d\mu$ both numerator and denominator on the left-hand side. It turns out that such formula essentially works also when $\nu \not\ll \mu$, as specified by lemma 41. Notice that lemma 41 in particular shows that $\frac{d\nu^a}{d\mu}$ exists if $\frac{d\nu}{d\gamma}$ exists. All this is used in [14, Theorem 7.2.12] as a step in a (completely different from ours) proof of the Radon-Nikodym theorem (and of the existence of the Lebesgue decomposition $\nu^a + \nu^s$ of $\nu$ with respect to $\mu$), to reduce to the case of the existence of $\frac{d\nu}{d\gamma}$. We extract from there the proof of lemma 41, which we include below for convenience of the reader.

**Lemma 41.** *Given positive measures $\mu, \nu$ on $(\Omega, \mathcal{F})$. Let $\gamma := \mu + \nu$ and assume*[22] *that $f(\gamma) := \frac{d\nu}{d\gamma}$ exists. Define the continuous bijection $\psi : [0,1] \to [0,\infty]$ as*

$$\psi(x) := \begin{cases} \frac{x}{1-x} & \text{for } x \in [0,1), \\ \infty & \text{for } x = 1, \end{cases} \quad \text{whose inverse is } \psi^{-1}(y) := \begin{cases} \frac{y}{1+y} & \text{for } y \in [0,\infty), \\ 1 & \text{for } y = \infty. \end{cases}$$

*Then $g := g(\gamma) := \psi \circ f(\gamma)$ satisfies*

$$\nu^a := \nu(\{g < \infty\} \cap \cdot) = g 1_{\{g<\infty\}} \cdot \mu \ll \mu, \quad \nu^s := \nu(\{g = \infty\} \cap \cdot) \perp \mu, \tag{31}$$

*so in particular $\frac{d\nu^a}{d\mu} = g 1_{\{g<\infty\}} = g$ $\mu$ a.s., $\mu(g = \infty) = 0$.*

*Proof.* Since $\gamma \geq \nu$, $h := f(\gamma)$ is $[0,1]$-valued and satisfies

$$\int_\Omega \phi(1-h) d\nu = \int_\Omega \phi h d\mu \quad \text{for all } \phi \geq 0. \tag{32}$$

Fix an arbitrary $\Gamma \in \mathcal{F}$. Notice that on $\{g < \infty\} = \{h < 1\}$ we have $1 - h > 0$, so $\phi_\Gamma := \frac{1}{1-h} 1_{\{g<\infty\} \cap \Gamma}$ is well defined and finite. Apply (32) with $\phi = \phi_\Gamma$ to get

$$\nu(\{h < 1\} \cap \Gamma) = \int_\Omega \phi_\Gamma (1-h) d\nu = \int_\Gamma 1_{\{g<\infty\}} g d\mu$$

i.e. $\nu(\{g < \infty\} \cap \cdot) = g 1_{\{g<\infty\}} \cdot \mu$. The thesis then follows the fact that (32) with $\phi := 1_{\{g=\infty\}}$ gives

$$\mu(\{g = \infty\}) = \int_{\{h=1\}} h d\mu = \int_{\{h=1\}} (1-h) d\nu = 0.$$

□

[22]Of course we will show (in theorem 4) that $\frac{d\nu}{d\gamma}$ always exist.

**Lemma 42.** *Under the assumption of lemma 41, take $\pi \in \mathcal{P}$, and let $f_\pi(\gamma)$ be defined as in eq.* (6)*. Then $f_\pi(\gamma) = \frac{d\nu_{|\sigma(\pi)}}{d\gamma_{|\sigma(\pi)}}$, and $g_\pi(\gamma) := \psi \circ f_\pi(\gamma)$ satisfies*

$$g_\pi(\gamma) 1_{\{g_\pi(\gamma)<\infty\}} = \frac{d(\nu_{|\sigma(\pi)})^a}{d\mu_{|\sigma(\pi)}} = f_\pi(\mu).$$

*Proof.* The thesis follows from eq. (4), and lemma 41 with $\mathcal{F} = \sigma(\pi)$. □

**Lemma 43.** *Let $(f_i)_{i\in I}, f$ be $[0,1]$-valued random variables. If $(f_i)_i \to f$ in probability and $f < 1$ a.s. then $f_i 1_{\{f_i<1\}} \to f$ in probability.*

*Proof.* We have to show that for small enough $\delta, \epsilon > 0$ there exists $j \in I$ such that

$$A_\delta := \{|f_i 1_{\{f_i<1\}} - f| \geq \delta\}$$

has probability $\mu$ at most $\epsilon$. Since $\mu(f < 1 - \frac{1}{n}) \uparrow \mu(f < 1) = 1$ as $n \to \infty$, there exists $n$ such that $\mu(f \geq 1 - \frac{1}{n}) < \epsilon/2$. Since $(f_i)_i \to f$ in $L^0(\mu)$ there exists $j \in I$ such that

$$\mu(\{|f_i - f| \geq \delta\}) < \epsilon/2 \quad \text{for all } i \geq j.$$

The thesis thus follows from the fact that, for any $\delta < \frac{1}{n}$, $|f_i(\omega) - f(\omega)| < \delta$ and $f(\omega) < 1 - \frac{1}{n}$ imply $f_i(\omega) = f_i(\omega)1_{\{f_i(\omega)<1\}}$, and so

$$A_\delta \cap \left\{ f < 1 - \frac{1}{n} \right\} \subseteq \{|f_i - f| \geq \delta\}.$$

□

*Proof of Theorem 4.* We use the notations of lemmas 41 and 42. For all $\pi \in \mathcal{P}$ let $\mathcal{F}_\pi := \sigma(\pi)$. Notice that $f_\pi(\gamma)$ has values in $[0,1]$, is $\mathcal{F}_\pi$-measurable and satisfies

$$\nu(A) = \textstyle\int_A f_\pi(\gamma) d\gamma \tag{33}$$

for all $A \in \pi$, and thus for all $A \in \mathcal{F}_\pi$ (since any element of $\mathcal{F}_\pi$ is a finite union of elements of $\pi$), so $(f_\pi(\gamma), \mathcal{F}_\pi)_{\pi\in\mathcal{P}}$ is a $[0,1]$-valued $\gamma$-martingale.

By corollary 29 $(f_\pi(\gamma))_\pi$ converges in $L^2(\gamma)$ (and so in $L^0(\mu)$) to some $f_\infty$ which closes $(f_\pi(\gamma), \mathcal{F}_\pi)_\pi$, and so does $(f_{\pi_n}(\gamma))_n$, and so eq. (33) shows that $f_\infty = f(\gamma) := \frac{d\nu}{d\gamma}$; in particular $\frac{d\nu}{d\gamma}$ exists, and so by lemma 41 also $\frac{d\nu^a}{d\mu}$ and the decomposition $\nu = \nu^a + \nu^s$ exist.

Since lemma 41 gives $g(\gamma) < \infty$ $\mu$ a.e., we have $f(\gamma) < 1$ $\mu$ a.e., so lemma 43 applied to $f_i = f_\pi(\gamma)$ gives that $(f_\pi(\gamma)1_{\{f_\pi(\gamma)<1\}})_{\pi\in\mathcal{P}} \to f(\gamma)$ in $L^0(\mu)$. From this and the fact that $\psi$ is an isometry[23] we get

$$f_\pi(\mu) = g_\pi(\gamma)1_{\{g_\pi(\gamma)<\infty\}} = \psi(f_\pi(\gamma)1_{\{f_\pi(\gamma)<1\}}) \to \psi(f(\gamma)) = g(\gamma) = \frac{d\nu^a}{d\mu} \text{ in } L^0(\mu)$$

where the first (resp. third) equality comes from lemma 42 (resp. lemma 41), the second follows from $\{\psi = \infty\} = \{1\}$ and the definition $g_\pi(\gamma) := \psi(f_\pi(\gamma))$. Replacing $(\pi)_{\pi\in\mathcal{P}}$ with $(\pi_n)_{n\in\mathbb{N}}$ throughout the proof shows that $f_{\pi_n}(\mu) \to \frac{d\nu^a}{d\mu}$ in $L^0(\mu)$. □

[23] We are here assuming that $[0,\infty]$ is endowed with the metric $d_\psi$ which makes $\psi$ an isometry (i.e. $d_\psi(x,y) := |\psi^{-1}(x) - \psi^{-1}(y)|$), which induces on $[0,\infty]$ the usual topology. Of course one could also use any distance $d^*$ such that $\forall\epsilon > 0 \exists\delta > 0$ such that $d(x,y) < \delta$ implies $d^*(x,y) < \epsilon$, since all that matters is that $\psi$ is uniformly continuous (it does not need to be an isometry).

*Remark* 44. In [4, Chapter 11, Section 17] it is shown that if $\pi_n \in \mathcal{P}, n \in \mathbb{N}$ is increasing then

$$f_{\pi_n} \to \mathbb{E}\left[\frac{d\nu^a}{d\mu}\middle|\tilde{\mathcal{F}}\right] \mu \text{ a.e. as } n \to \infty, \tag{34}$$

where $\tilde{\mathcal{F}} := \bigvee_n \sigma(\pi_n) = \sigma(\cup_n \mathcal{F}_{\pi_n}) \subseteq \mathcal{F}$ and $\mathbb{E}[\cdot|\tilde{\mathcal{F}}]$ denotes the conditional $\mu$-expectation w.r.t. $\tilde{\mathcal{F}}$. In general $\tilde{\mathcal{F}}$ will in general be strictly[24] smaller than $\mathcal{F}$ for the order $\preceq_\mu$ of $\mu$ a.s. inclusion[25], and so in general $\mathbb{E}\left[g\middle|\tilde{\mathcal{F}}\right] \neq g$ for some $g \in L^1(\mu)$.

Nonetheless, eq. (34) shows that, with the $\pi_n$ as in Theorem 4, one gets that $\mathbb{E}\left[\frac{d\nu^a}{d\mu}\middle|\tilde{\mathcal{F}}\right] = \frac{d\nu^a}{d\mu}$, i.e. $\frac{d\nu^a}{d\mu}$ is $\tilde{\mathcal{F}}$-measurable. So, in hindsight, this is the reason why it is enough to take the limits along the sequence we chose, instead of using the whole net. Notice that it is obvious[26] that $\frac{d\nu^a}{d\mu}$ is measurable with respect to $\sigma(\cup_n \mathcal{F}_{\pi_n})$ for some choice of $\pi_n$, and so $f_{\pi_n}(\mu) \to \frac{d\nu^a}{d\mu}$ $\mu$ a.e. for any such $(\pi_n)_n$.

## 9. The proof of theorem 19

We now show how theorem 19 can be easily proved using theorem 12 and the definition of convergence of a net. Notice that the *definition* of convergence of a net is, formally speaking, analogous to the definition of limit as $t \to +\infty$ of a function of a real variable $t \in \mathbb{R}$ (familiar to all students), and as such presents no difficulties. However, proving results about nets instead is not always achievable by a straightforward generalisation of the proof for sequences. Instead, we show how theorem 19 provides an alternative way, in the specific setting at hand.

*Proof of theorem 19.* (3) $\implies$ (1): Consider $J = \mathbb{R}, \phi := \int_0^\cdot \tan^{-1}(x)dx$. Since the derivative $\phi'$ of $\phi$ is (strictly increasing and) bounded, $\phi$ is (strictly convex and) Lipschitz; thus $(\phi(f_i))_{i\in I}$ is uniformly integrable, since such is $(f_i)_{i\in I}$. Thus, we can apply the first two items of theorem 12 to obtain that there exists some $(i_n)_n$ for which $(f_{i_n})_{n\in\mathbb{N}}$ converges in $L^1$ to a $f$ which closes $(f_i, \mathcal{F}_i)_{i\in I}$. Since by assumption $(f_{i_n})_{n\in\mathbb{N}}$ is p-uniformly integrable, by the last item of theorem 12 $(f_i)_{i\in I}$ converges to $f$ in $L^p$.

(1) $\implies$ ((2) and $\mathbb{E}[f|\mathcal{G}] = g$): We present all the details of this rather trivial implication, to clarify that they trivially follow from the *definition* of convergence of a net. For every $i, j \in I, i \leq j$ and $A \in \mathcal{F}_i$ we have $\mathbb{E}[1_A f_i] = \mathbb{E}[1_A f_j]$, and taking the limit over $j \in I$ gives $\mathbb{E}[1_A f_i] = \mathbb{E}[1_A g]$, since otherwise there exists $\varepsilon >$ such that $\varepsilon < |\mathbb{E}[1_A f_i] - \mathbb{E}[1_A g]|$, which contradicts the existence of $j_\varepsilon$ such that $\|g - f_j\|_{L^1} < \varepsilon$ for all $j \geq j_\varepsilon$ since

$$\varepsilon < |\mathbb{E}[1_A f_i] - \mathbb{E}[1_A g]| = |\mathbb{E}[1_A f_j] - \mathbb{E}[1_A g]| \leq \|g - f_j\|_{L^1} < \varepsilon.$$

It follows that $g$ closes $(f_i)_{i\in I}$. If $f$ is another function which closes $(f_i)_{i\in I}$ then $\mathbb{E}[1_A f] = \mathbb{E}[1_A g]$ holds for all $A \in \mathcal{F}_i$ for every $i \in I$, and thus it holds for every

[24] In particular this happens whenever $\mathcal{F}$ is not a.s. separable (meaning it equals a separable $\sigma$-algebra up to null sets), or equivalently [13, Theorem 23.21] when $L^1(\Omega, \mathcal{F}, \mu)$ is not separable.

[25] By definition $\mathcal{G} \preceq_\mu \mathcal{H}$ if every set in $\mathcal{G}$ differs from some set in $\mathcal{H}$ only by a $\mu$-null set.

[26] Any $\mathcal{F}$-measurable $g : \Omega \to \mathbb{R}$ is also $\tilde{\mathcal{F}}$-measurable for some separable sigma algebra $\tilde{\mathcal{F}} \subseteq \mathcal{F}$, because the Borel sets of $\mathbb{R}$ form a separable sigma algebra, and so also $\sigma(g)$ is separable.

$A \in \mathcal{G} := \sigma(\cup_i \mathcal{F}_i)$. This implies $\mathbb{E}[f|\mathcal{G}] = g$, since $g$ is $\mathcal{G}$-measurable (because $(f_i)_{i\in I} \subseteq L^1(\mathcal{G})$ and $L^1(\mathcal{G})$ is closed[27] in $L^1(\mathcal{F})$).

(2) $\implies$ (3): The classic proof is found in [11, Lemma IV-2-4] for $p = 1$; the case $p > 1$ follows easily from Doob's $L^p$ martingale inequality and the fact that $\mathbb{E}[\cdot|\mathcal{F}_i]$ is a contraction[28] in $L^p$ and so

$$\| \text{ess-sup}_i |f_i| \|_{L^p} \leq c_p \sup_i \|f_i\|_{L^p} \leq \|f\|_{L^p},$$

which shows that $(|f_i|^p)_i$ is dominated by the *integrable* random variable ess-sup$_i |f_i|^p$, and so in particular it is uniformly integrable. We now present a novel proof, which has the advantage that it does not rely on Doob's inequality, and it unifies the cases $p = 1$ and $p > 1$.

By one implication of de la Vallée Poussin criterion there exists a positive even convex function $\phi : \mathbb{R} \to \mathbb{R}$ such that $\lim_{t\to\infty} \phi(t)/t^p = \infty$ and $\phi(f) \in L^1$. Since $f_i = \mathbb{E}[f|\mathcal{F}_i]$ and $\phi$ is continuous, the conditional Jensen inequality (see [16, Section 9.7(h)] or theorem 16) gives that $\phi(f_i) \leq \mathbb{E}[\phi(f)|\mathcal{F}_i]$, and thus $\mathbb{E}\phi(f_i) \leq \mathbb{E}\phi(f) < \infty$; so, the opposite implication of de la Vallée Poussin criterion shows that $(f_i)_{i\in I}$ is $p$-uniformly integrable. □

**Acknowledgment.** We thank an anonymous referee for providing helpful comments. Oleksii Mostovyi is supported by the National Science Foundation under grant No. DMS-1848339 (2019-2024). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

[1] M. Beiglboeck, W. Schachermayer, and B. Veliyev. A short proof of the Doob–Meyer theorem. *Stochastic Processes and their Applications*, 122(4):1204–1209, 2012.

[2] C. Dellacherie and P.-A. Meyer. *Probabilities and potential. B*, volume 72 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1982. Theory of martingales, Translated from the French by J. P. Wilson.

[3] E. DiBenedetto and E. Debenedetto. *Real analysis*. Springer, 2002.

[4] J. L. Doob. *Measure theory*, volume 143. Springer Science & Business Media, 2012.

[5] G. A. Edgar and L. Sucheston. *Stopping times and directed processes*, volume 47. Cambridge University Press, 1992.

[6] L. C. Evans and L. Gariepy. *Measure theory and fine properties of functions*. Routledge, 2018.

[7] C. A. Hayes and C. Y. Pauc. *Derivation and martingales*, volume 49. Springer Science & Business Media, 2012.

[8] J. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms, Vol 1*, volume 1. Springer, 1993.

[9] J. Hoffman-Jorgensen. *Probability with a view toward statistics*, volume 1. Routledge, 2017.

[10] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[27] If $C \subseteq L^1$ is closed under convergence of sequences then it is closed under convergence of nets, since by definition if $(f_i)_{i\in I} \to f$ then $\forall n \in \mathbb{N}\, \exists i_n \in I$ such that $\|f - f_{i_n}\|_{L^1} < 1/n$.

[28] The standard proof of this is a trivial application of the conditional Jensen inequality (theorem 16). Here a more elementary proof, which has the advantage of yielding also the case $p = \infty$: if $p \in [1, \infty]$ has conjugate exponent $q$ then Hölder's inequality and its converse give that

$$||\mathbb{E}[f|\mathcal{G}]||_{L^p} = \sup_{||g||_{L^q(\mathcal{G})}\leq 1} \mathbb{E}[\mathbb{E}[f|\mathcal{G}]g] = \sup_{||g||_{L^q(\mathcal{G})}\leq 1} \mathbb{E}[fg] \leq \sup_{||g||_{L^q(\mathcal{F})}\leq 1} \mathbb{E}[fg] = ||f||_{L^p}.$$

[11] J. Neveu. *Discrete-parameter martingales*. North-Holland Publishing Company, 1975.
[12] W. Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 2006.
[13] R. L. Schilling. *Measures, integrals and martingales*. Cambridge University Press, 2017.
[14] D. W. Stroock. *A concise introduction to the theory of integration*. Springer Science & Business Media, 1998.
[15] D. W. Stroock. *Probability theory: an analytic view*. Cambridge university press, 2010.
[16] D. Williams. *Probability with martingales*. Cambridge university press, 1991.

Oleskii Mostovyi, University of Connecticut, Department of Mathematics, Storrs, CT 06269, United States
*Email address*: oleksii.mostovyi@uconn.edu

Pietro Siorpaes: Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom
*Email address*: p.siorpaes@imperial.ac.uk