A reduction technique for limit theorems in analysis and probability theory

J. Bliedtner and P. Loeb[1]

0. Introduction

The gist of many theorems in analysis and probability theory is that for each measure μ in a given class M_0 and for some reference measure σ, the Radon—Nikodym derivative $d\mu/d\sigma$ is equal σ-almost everywhere to a limit along a directed set $\mathscr{I}$ of ratios $R_i(\mu, \sigma)$, $i \in \mathscr{I}$, defined in terms of μ and σ. In this paper, we strengthen the main result from [6] to develop an equivalent formulation of such limit theorems. The essential idea is that the desired result is established for all $\mu \in M_0$ once it is shown that for any measurable set E and any $\nu \in M_0$ with $\nu(E)=0$, $\lim_i R_i(\nu, \sigma)=0$ σ-almost everywhere on E. Indeed, assuming the class M_0 is closed with respect to scaling, it is enough to show that $\limsup_i R_i(\nu, \sigma) \leqq 1$ σ-almost everywhere on E. We describe the ratios $R_i(\mu, \sigma)$ and the limit process with sufficient generality to make our reduction technique applicable in quite diverse settings. The applications in this paper are boundary limit theorems in potential theory, the martingale convergence theorem in probability theory, and differentiation theorems in measure theory. Here are prototypical theorems in these three areas.

0.1 Radial and Fine Limit Theorems in Potential Theory. Let C be the boundary of the unit disk $D=\{z \in \mathbf{C} \,|\, |z|<1\}$. For each finite, positive Borel measure μ on C, let $P\mu$ denote the harmonic function on D obtained by taking the integral with respect to μ of the Poisson kernel. Fix two positive harmonic functions $h=P\sigma$ and $g=P\mu$ on D. By the ratio Fatou theorem (see [10]), for σ-almost every $z \in C$,

[1] The second author's work was supported in part by a grant from the U.S National Science Foundation (DMS 89—02095).

the limit of g/h exists along the radial line $\{r\cdot z | r\in\mathbf{R},\ 0\leqq r<1\}$, and

$$\lim_{r\to 1-}\frac{g(r\cdot z)}{h(r\cdot z)}=\frac{d\mu}{d\sigma}(z).$$

The fine limit theorem of Fatou—Naïm—Doob [6], [10] replaces limits along radii with limits defined in terms of "minimal fine neighborhoods".

0.2 Martingale Convergence Theorem in Probability Theory. Let $(\Omega, \mathscr{A}, P)$ be a probability space and $\langle\mathscr{A}_n\rangle$ an increasing sequence of σ-algebras such that $\mathscr{A}$ is the smallest σ-algebra containing them all. For each integrable function h on Ω and each $n\in\mathbf{N}$, let $E[h|\mathscr{A}_n]$ denote the conditional expectation of h with respect to $\mathscr{A}_n$. The martingale convergence theorem for generated martingales states that for each integrable h,

$$h(\omega)=\lim_n E[h|\mathscr{A}_n](\omega)\quad\text{for}\quad P\text{-a.e.}\quad \omega\in\Omega.$$

It follows from our general result that this statement is equivalent to the seemingly weaker one that for each measurable set E and each integrable function $h\geqq 0$ which vanishes on E,

$$\limsup_n E[h|\mathscr{A}_n](\omega)\leqq 1\quad\text{for}\quad P\text{-a.e.}\quad \omega\in E.$$

0.3. Lebesgue Differentiation and Density Theorems in Measure Theory.
Let $\mathscr{B}$ denote the Borel subsets of $\mathbf{R}^n$, and let σ be a Borel measure, finite on compact sets, with support $Y\subseteqq\mathbf{R}^n$. For any $E\in\mathscr{B}$, let σ_E be the measure defined at each $A\in\mathscr{B}$ by setting $\sigma_E(A)=\sigma(E\cap A)$. Fix a finite Borel measure μ, and let $d\mu/d\sigma$ denote the Radon—Nikodym derivative of the absolutely continuous part of μ with respect to σ. For each $y\in Y$, let $B(y, r)$ denote the closed ball with center y and radius $r>0$. By the Lebesgue differentiation theorem,

$$\lim_{r\to 0+}\frac{\mu(B(y,r))}{\sigma(B(y,r))}=\frac{d\mu}{d\sigma}(y).$$

for σ-almost every $y\in Y$. The Lebesgue density theorem is the same result with μ restricted to the set of Borel measures of the form σ_E. Covering theorems of Besicovitch, Morse, and Vitali are usually used to establish this measure differentiation theorem.

The setting of our general result is a finite measure space $(X, \mathscr{B}, \sigma)$ with $\sigma(X)>0$. Here, we let M denote the set of all nonnegative finite measures on $(X, \mathscr{B})$. The point of the technique discussed here is that one works not with just one other measure, but with all measures in a subset $M(\sigma)\subseteqq M$ having some additional structure. Of particular importance are the cases $M(\sigma)=M$ and $M(\sigma)=\{h\sigma | h\in L^p(\sigma)\}$, $1\leqq p\leqq\infty$. We will treat the limit process in terms of a nonempty family $\mathscr{F}$ of nonnegative real valued functionals on $M(\sigma)$. With σ-almost every point $x \in X$, we will associate a filter $\mathfrak{F}(x)$ on $\mathscr{F}$; the limit will be in terms of these filters.

Such a class $\mathscr{F}$ is obtained as follows in the setting of the Lebesgue differentiation theorem (Example 0.3). Each closed ball B determines a functional F_B by setting $F_B(\mu)=\mu(B)$ for each Borel measure μ. For each fixed x, the balls $B(x, r)$ centered at x with radius $r \leqq 1/n$ form a typical set in a filter base of such functionals. The Lebesgue theorem is a corollary of the Besicovitch—Morse covering theorem (given a simple proof in Section 5) and a new measure differentiation theorem established using our reduction technique in Theorem 4.2. For general metric spaces with an appropriate reference measure, we use a new Vitali covering theorem (6.1).

A class $\mathscr{F}$ of functionals also appears in the setting of the radial and fine limit theorems (Example 0.1). Here, X is the boundary C of the unit disk, $\mathscr{B}$ is the class of Borel sets in C, and $M(\sigma)=M$ is the class of positive finite Borel measures on $(C, \mathscr{B})$. Each point y inside the disk determines a functional F_y on $M(\sigma)$ by taking the harmonic extension $P\mu$ of a measure μ on C and setting $F_y(\mu)=P\mu(y)$.

For the martingale convergence theorem (Example 0.2) and its generalizations, the reference measure σ is the probability measure P. Here, point evaluations of conditional expectations form the appropriate set $\mathscr{F}$ of functionals on $M(P)$. Problems with the ambiguity of such evaluations in terms of sets of zero probability are anticipated in the formulation of our principal theorem.

1. A general limit theorem

In what follows, $(X, \mathscr{B})$ is a measurable space and M is the set of all nonnegative finite measures on $(X, \mathscr{B})$. We use the notation μ_E for the restriction of a measure μ to a measurable set E; i.e., $\mu_E(A)=\mu(A \cap E)$. We let $\mathbf{R}^+$ denote the nonnegative real numbers, $\mathbf{N}$ the natural numbers, and CE the complement of a set E in X. If μ and η are measures, we write $\mu \leqq \eta$ if $\mu(A) \leqq \eta(A)$ for each $A \in \mathscr{B}$.

We work with a nonzero measure $\sigma \in M$, called a reference measure, and a set $M(\sigma) \subseteqq M$ with $\sigma \in M(\sigma)$. Our theorem is given in terms of a class $\mathscr{F}$ of functions mapping $M(\sigma)$ into $\mathbf{R}^+$ with $F(\sigma)>0$ for each $F \in \mathscr{F}$. Associated with σ-almost every point $x \in X$ is a filter $\mathfrak{F}(x)$ on $\mathscr{F}$. We will assume that the following conditions hold for each $\mu \in M(\sigma)$:

i) For every $r \in \mathbf{R}^+$, $r\mu \in M(\sigma)$ and

$$\{F \in \mathscr{F} \mid F(r\mu) = r \cdot F(\mu)\} \in \mathfrak{F}(x) \quad \text{for} \quad \sigma\text{-a.e.} \quad x \in X.$$

ii) For every $E \in \mathscr{B}$, $\mu_E \in M(\sigma)$ and

$$\{F \in \mathscr{F} \mid F(\mu) = F(\mu_E) + F(\mu_{CE})\} \in \mathfrak{F}(x) \quad \text{for} \quad \sigma\text{-a.e.} \quad x \in X.$$

iii) If $s\cdot\sigma_E \leqq \mu \leqq t\cdot\sigma_E$ for some $E\in\mathscr{B}$ and $s, t\in\mathbf{R}^+$, then

$$\{F\in\mathscr{F} \mid s\cdot F(\sigma_E) \leqq F(\mu) \leqq t\cdot F(\sigma_E)\}\in\mathfrak{F}(x) \quad \text{for} \quad \sigma\text{-a.e.} \quad x\in E.$$

Note that the equalities in Properties i and ii are satisfied by all of the functionals in $\mathscr{F}$ if they are affine linear. The second inequality in Property iii is satisfied by all the functionals if they are affine and increasing. In what follows, we write $\lim_{F,\mathfrak{F}(x)} F(\mu)/F(\sigma)=a$ if for each $\varepsilon>0$ there is a set $\mathscr{S}\in\mathfrak{F}(x)$ such that for all $F\in\mathscr{S}$, $|F(\mu)/F(\sigma)-a|<\varepsilon$. Also, we write $\dfrac{d\mu}{d\sigma}$ for the Radon—Nikodym derivative of the absolutely continuous part of μ with respect to σ.

1.1 Theorem. *Given $\sigma\in M(\sigma)\subseteq M$, a family $\mathscr{F}$ of nonnegative functionals with $F(\sigma)>0$ for each $F\in\mathscr{F}$, and filters $\mathfrak{F}(x)$ on $\mathscr{F}$ defined for σ-almost all x, assume that Properties* i—iii *hold for each $\mu\in M(\sigma)$. Then the following are equivalent:*

(1) *For each $\mu\in M(\sigma)$, $\lim_{F,\mathfrak{F}(x)}\dfrac{F(\mu)}{F(\sigma)}=\dfrac{d\mu}{d\sigma}(x)$ for σ-a.e. $x\in X$.*

(2) *For each $E\in\mathscr{B}$ with $\sigma(E)>0$ and each $\nu\in M(\sigma)$ with $\nu(E)=0$,*

$$\lim_{F,\mathfrak{F}(x)}\frac{F(\nu)}{F(\sigma)}=0 \quad \textit{for} \quad \sigma\text{-a.e.} \quad x\in E.$$

(3) *For each $E\in\mathscr{B}$ with $\sigma(E)>0$ and each $\nu\in M(\sigma)$ with $\nu(E)=0$,*

$$\{F\in\mathscr{F} \mid F(\nu) \leqq F(\sigma)\}\in\mathfrak{F}(x) \quad \textit{for} \quad \sigma\text{-a.e.} \quad x\in E.$$

Proof. ($1\Rightarrow 3$) Assume 1 holds and fix $E\in\mathscr{B}$ with $\sigma(E)>0$ and $\nu\in M(\sigma)$ with $\nu(E)=0$. Then $\lim_{F,\mathfrak{F}(x)}\dfrac{F(\nu)}{F(\sigma)}=\dfrac{d\nu}{d\sigma}(x)=0$ for σ-a.e. $x\in E$, so

$$\{F\in\mathscr{F} \mid F(\nu)/F(\sigma) \leqq 1\}\in\mathfrak{F}(x) \quad \text{for} \quad \sigma\text{-a.e.} \quad x\in E.$$

($3\Rightarrow 2$) Given $E\in\mathscr{B}$ with $\sigma(E)>0$ and $\nu\in M(\sigma)$ with $\nu(E)=0$, we have for each $k\in\mathbf{N}$, $k\nu\in M(\sigma)$ and $k\nu(E)=0$. Let

$$\mathscr{S}=\{F\in\mathscr{F} \mid \forall k\in\mathbf{N},\ k\cdot F(\nu)=F(k\nu)\}.$$

By discarding a set of σ-measure 0 from E, we may assume that $\mathscr{S}\in\mathfrak{F}(x)$ for all $x\in E$. It now follows that $\forall k\in\mathbf{N}$,

$$\{F\in\mathscr{S} \mid F(\nu)/F(\sigma) \leqq 1/k\} = \{F\in\mathscr{S} \mid F(k\nu)/F(\sigma) \leqq 1\}\in\mathfrak{F}(x)$$

for σ-a.e. $x\in E$. Therefore, $\lim_{F,\mathfrak{F}(x)}\dfrac{F(\nu)}{F(\sigma)}=0$ for σ-a.e. $x\in E$.

($2\Rightarrow 1$) Fix a finite, nonnegative, integrable function h on X and a measure $\nu\perp\sigma$ so that $\mu=h\sigma+\nu\in M(\sigma)$. There are disjoint measurable sets X_1 and X_2 with

$X=X_1\cup X_2$ and $\sigma(X_2)=\nu(X_1)=0$. By Property ii, therefore, $h\sigma$ and ν are in $M(\sigma)$, and the set

$$\{F\in\mathcal{F}\,|\,F(\mu)=F(h\sigma)+F(\nu)\}\in\mathfrak{F}(x)$$

for σ-a.e. $x\in X$. By assumption, $\lim_{F,\mathfrak{F}(x)}\dfrac{F(\nu)}{F(\sigma)}=0=\dfrac{d\nu}{d\sigma}(x)$ for σ-a.e. $x\in X_1$ and thus for σ-a.e. $x\in X$. To finish the proof, we must show that for some measurable set U with $\sigma(U)=0$ and for all $x\in X-U$, $\lim_{F,\mathfrak{F}(x)}\dfrac{F(h\sigma)}{F(\sigma)}=h(x)$. Choose an $n\in\mathbf{N}$, and partition $\mathbf{R}^+$ into intervals of length $1/4n$. Let E be the inverse image with respect to h of one of the intervals $[r, r+1/4n]$. If $\sigma(E)=0$, adjoin E to U. If $\sigma(E)>0$, let $\mathcal{S}_E$ be the set of all $F\in\mathcal{F}$ such that

$$F(\sigma)=F(\sigma_E)+F(\sigma_{CE}),\quad F(h\sigma)=F(h\sigma_E)+F(h\sigma_{CE}),$$

and

$$r\cdot F(\sigma_E)\leqq F(h\sigma_E)\leqq\left(r+\frac{1}{4n}\right)\cdot F(\sigma_E).$$

By discarding a set of σ-measure 0, we may assume that $\mathcal{S}_E\in\mathfrak{F}(x)$ for each $x\in E$. Now for any $x\in E$ and any $F\in\mathcal{S}_E$, we have

$$\begin{aligned}F(\sigma)\cdot|F(h\sigma)/F(\sigma)-h(x)| &= |F(h\sigma)-h(x)\cdot F(\sigma)|\\ &\leqq |F(h\sigma)-r\cdot F(\sigma)|+|r\cdot F(\sigma)-h(x)\cdot F(\sigma)|\\ &\leqq |F(h\sigma_{CE})-r\cdot F(\sigma_{CE})|+|F(h\sigma_E)-r\cdot F(\sigma_E)|+\big(h(x)-r\big)F(\sigma)\\ &\leqq F(h\sigma_{CE})+r\cdot F(\sigma_{CE})+\frac{1}{4n}\cdot F(\sigma_E)+\frac{1}{4n}\cdot F(\sigma).\end{aligned}$$

Since $F(\sigma)=F(\sigma_E)+F(\sigma_{CE})$, $F(\sigma_E)/F(\sigma)\leqq 1$, so

$$\left|\frac{F(h\sigma)}{F(\sigma)}-h(x)\right|\leqq\frac{F(h\sigma_{CE})}{F(\sigma)}+r\cdot\frac{F(\sigma_{CE})}{F(\sigma)}+\frac{1}{2n}.$$

By assumption, for σ-almost all $x\in E$,

$$\lim_{F,\mathfrak{F}(x)}\frac{F(h\sigma_{CE})}{F(\sigma)}=0,\quad\text{and}\quad\lim_{F,\mathfrak{F}(x)}r\cdot\frac{F(\sigma_{CE})}{F(\sigma)}=0.$$

Therefore, for σ-a.e. $x\in E$, there is a set $\mathcal{S}_n(x)\in\mathfrak{F}(x)$ such that $|F(h\sigma)/F(\sigma)-h(x)|<1/n$ for all $F\in\mathcal{S}_n(x)$. We obtain the desired result by putting all of the sets of measure 0 together for the sets E corresponding to the partition and repeating the operation for each $n\in\mathbf{N}$. $\square$

1.2 Corollary. *Assume that* $F(0)=0$ *for every* $F\in\mathcal{F}$. *Let* $\hat{M}(\sigma)$ *be the family of all signed measures of the form* $\mu_E-\mu_{CE}$, $\mu\in M(\sigma)$ *and* $E\in\mathcal{B}$. *For such a signed* *measure and any* $F \in \mathcal{F}$, *set*

$$F(\mu_E - \mu_{CE}) = F(\mu_E) - F(\mu_{CE}).$$

Then for each $\varphi \in \hat{M}(\sigma)$, $\lim_{F,\mathfrak{F}(x)} \frac{F(\varphi)}{F(\sigma)} = \frac{d\varphi}{d\sigma}(x)$ *for* σ-a.e. $x \in X$ *if and only if for each* $E \in \mathcal{B}$ *with* $\sigma(E) > 0$ *and each* $\nu \in M(\sigma)$ *with* $\nu(E) = 0$, $\{F \in \mathcal{F} | F(\nu) \leqq F(\sigma)\} \in \mathfrak{F}(x)$ *for* σ-a.e. $x \in E$.

Note that if $M(\sigma)$ is the family of all finite measures on $(X, \mathcal{B})$, then by the Hahn decomposition theorem, $\hat{M}(\sigma)$ is the set of all finite signed measures on $(X, \mathcal{B})$.

The remainder of this paper consists of applications illustrating the use of our general reduction technique. The fine limit theorem is our first example. It was the motivation for the weak version of Theorem 1.1 established in [6]. To help clarify the interpretation of Theorem 1.1, we briefly describe this potential theoretic application.

2. Fatou—Naïm—Doob fine limit Theorem

Let D be the unit disk and C its boundary. We let $P\mu$ denote the Poisson integral of a finite measure μ on C. For each $y \in D$ there is a corresponding functional F_y such that $F_y(\mu) = P\mu(y)$. In terms of the notation of Section 1, $X = C$ and $\mathcal{F} = D$. Here, we set $M(\sigma) = M$ regardless of the choice of σ, but for more general domains and potential theories one may need a subset of M to insure finite harmonic extensions. For each $z \in C$, the fine neighborhood $\mathfrak{F}(z)$ is a filter in D refining the ordinary neighborhood system at z. Using a weak form of Theorem 1.1 and a proof that works in very general potential theoretic settings (e.g., [5], [18]), the following result was established in [6].

2.1 Fatou—Naïm—Doob Theorem. *Fix a harmonic function* $h = P\sigma > 0$. *Also fix a superharmonic function* $s \geqq 0$ *on* D; *that is*, $s = p + P\mu$, *where* p *is a potential and* $\mu \in M$. *Then for* σ-*a.e.* $z \in C$,

$$\lim_{\mathfrak{F}(z)} \frac{s}{h} = \frac{d\mu}{d\sigma}(z).$$

The ratio limit theorem of Fatou stated in Example 0.1 is an immediate corollary of Theorem 2.1 and a result of Brelot and Doob (see [10], Section 1. XII. 21). That is, for positive harmonic functions h and g on D, if $\lim_{\mathfrak{F}(z)} \frac{g}{h} = \alpha$ at a boundary point $z \in C$, then $\lim_{r \to 1-} \frac{g(r \cdot z)}{h(r \cdot z)} = \alpha$.

3. The martingale convergence theorem

Let $(\Omega, \mathscr{A}, P)$ be a probability space and $\mathscr{I}$ a countable subset $\mathbf{R}^+$ with $\mathbf{N} \subseteq \mathscr{I}$. The ordering on $\mathscr{I}$ is the ordering inherited from $\mathbf{R}^+$. Let $\{\mathscr{A}_i | i \in \mathscr{I}\}$ be an increasing family of σ-algebras such that $\mathscr{A}$ is the smallest σ-algebra containing them all. Given an integrable function f on Ω, $f(\omega) = \lim_{i \in \mathscr{I}} E[f|\mathscr{A}_i](\omega)$ for P-almost all $\omega \in \Omega$. We will employ Theorem 1.1 in a proof of this equality and its extension to more general martingales and submartingales. Since $f = (f + |f|) - |f|$, we may assume $f \geqq 0$. In applying Theorem 1.1, we take $\mathscr{B} = \mathscr{A}$, and our reference measure σ is P. Instead of working just with measures of the form $f \cdot P$, however, we let $M(P)$ be the set M of all finite measures on the measurable space $(\Omega, \mathscr{A})$. Our technique then yields a short proof of a result due to Andersen and Jessen [2], [9]. First we need some notation.

For each $\mu \in M$ and $i \in \mathscr{I}$, let μ_i denote the restriction $\mu | \mathscr{A}_i$, and fix a finite, nonnegative version h_i^μ of the Radon—Nikodym derivative $d\mu_i/dP$ (measurable with respect to $\mathscr{A}_i$). We set $h_i^P \equiv 1$ and $h_i^0 \equiv 0$. The functions h_i^μ form a martingale if for each $i \in \mathscr{I}$, $\mu_i \ll P$. (Otherwise, they form a supermartingale; see [9], page 632.) In any case, we will use Theorem 1.1 to show that $\dfrac{d\mu}{dP}(\omega) = \lim_i h_i^\mu(\omega)$ for P-almost every $\omega \in \Omega$. The convergence theorem for martingales generated by integrable functions (Corollary 3.3) is the application of this result to measures absolutely continuous with respect to P.

For each $\mu \in M$, $A \in \mathscr{A}$ and $k \in \mathbf{N}$, let

$$A_\mu^k = \{\omega \in A \,|\, \sup_{i \geqq k} h_i^\mu(\omega) > 1\}.$$

We need the following maximal inequality: If $A \in \mathscr{A}_k$, then $\mu(A_\mu^k) \geqq P(A_\mu^k)$. Recall this result follows by taking an arbitrary finite subset

$$\{k = i_1 < i_2 < \ldots < i_m\} \subseteq \mathscr{I},$$

setting $B_0 = \phi$, $B_n = \{\omega \in A | h_{i_n}^\mu > 1\} - \bigcup_{j=0}^{n-1} B_j$ for $1 \leqq n \leqq m$, and noting that

$$\mu(A_\mu^k) \geqq \sum_{n=1}^m \mu(B_n) \geqq \sum_{n=1}^m \int_{B_n} h_{i_n}^\mu \, dP \geqq \sum_{n=1}^m P(B_n) = P\left(\bigcup_{n=1}^m B_n\right).$$

The maximal inequality has the following corollary which is the key to our use of Theorem 1.1.

3.1 Proposition. *Fix $E \in \mathscr{A}$ with $P(E) > 0$ and $\nu \in M$ with $\nu(E) = 0$. For each $i \in \mathscr{I}$, let h_i^ν be a version of $d\nu_i/dP$. Then*

$$\limsup_{i \in \mathscr{I}} h_i^\nu(\omega) \leqq 1$$

for P-almost every $\omega \in E$.

Proof. We will show that $P(\bigcap_{k=1}^{\infty} E_\nu^k)=0$. Fix $\varepsilon>0$. Since $\mathscr{A}$ is generated by the $\mathscr{A}_i$'s, there is a $k\in\mathrm{N}$ and an $A\in\mathscr{A}_k$ such that $(P+\nu)(A\Delta E)<\varepsilon$. By the maximal inequality, $P(A_\nu^k)\leqq\nu(A_\nu^k)$, so,

$$P(E_\nu^k) \leqq P(E-A)+P(A_\nu^k) \leqq P(E-A)+\nu(A-E) < \varepsilon.$$

It follows that $P(\bigcap_{k=1}^{\infty} E_\nu^k)=0$. □

Proposition 3.1 verifies Condition 3 of Theorem 1.1. To see this, let $F_{i,\omega}(\mu)=h_i^\mu(\omega)$ for each $\omega\in\Omega$. Let $\mathscr{F}=\{F_{i,\omega}|i\in\mathscr{I},\ \omega\in\Omega\}$, and for each $\omega\in\Omega$ let $\mathfrak{F}(\omega)$ be the filter on $\mathscr{F}$ generated by the sets $\{F_{i,\omega}|i\geqq k\}$, $k\in\mathrm{N}$. It is easy to see that M, $\mathscr{F}$, and the filters $\mathfrak{F}(\omega)$ satisfy the Properties i—iii needed to apply Theorem 1.1. (Note that if μ satisfies the hypothesis of Property iii, then $\mu\ll P$.) Given Proposition 3.1, we can now apply Corollary 1.2. Here again, $\mu_i=\mu|\mathscr{A}_i$ for each $i\in\mathscr{I}$.

3.2 Theorem (Andersen and Jessen [2]). *For any finite signed measure μ on $(\Omega,\mathscr{A})$, and for P-almost $\omega\in\Omega$,*

$$\frac{d\mu}{dP}(\omega) = \lim_{i\in\mathscr{I}} \frac{d\mu_i}{dP}(\omega).$$

3.3 Corollary. *If h is an integrable function on Ω, then for P-a.e. $\omega\in\Omega$,*

$$h(\omega) = \lim_{i\in\mathscr{I}} E[h|\mathscr{A}_i](\omega).$$

Corollary 3.3 is the convergence theorem for martingales generated by an integrable function, but not all martingales take this form. In [9], J. L. Doob established the almost everywhere convergence of L^1-bounded submartingales $\{h_i|i\in\mathscr{I}\}$ adapted to $\{\mathscr{A}_i|i\in\mathscr{I}\}$. That is, $\sup_{i\in\mathscr{I}} E[|h_i|]<+\infty$, each h_i is measurable with respect to $\mathscr{A}_i$, and for $i<j$ in $\mathscr{I}$, the conditional expectation $E[h_j|\mathscr{A}_i]\geqq h_i$ P-almost everywhere. ($E[h_j|\mathscr{A}_i]=h_i$ P-a.e. if the h_i's form a martingale.) Johansen and Karush [13] showed that the limit of such a submartingale is the Radon—Nikodym derivative of a signed measure φ defined on $(\Omega,\mathscr{A})$. Here, φ is the countably additive part of the finitely additive signed measure φ_0 defined on the algebra $\mathscr{A}_0=\bigcup_{i\in\mathscr{I}}\mathscr{A}_i$ by setting $\varphi_0(A)=\lim_{i\in\mathscr{I}}\int_A h_i dP$ for each $A\in\mathscr{A}_0$. Note, however, that one can not directly apply Theorem 3.2 to φ_0 and φ does not generate the h_i's even when they form a martingale.

Johansen and Karush [13] (and later Chatterji [7]) actually established almost everywhere convergence for the Radon—Nikodym derivatives $d\varphi_i/dP$ of a bounded, increasing family φ_i, $i\in\mathscr{I}$, consisting of finite signed measures. Here for each $i\in\mathscr{I}$, φ_i is defined on $\mathscr{A}_i$, $\sup_{i\in\mathscr{I}}|\varphi_i|(\Omega)<+\infty$, and $\varphi_j(E)\geqq\varphi_i(E)$ for each $E\in\mathscr{A}_i$ and $j>i$. If, for example, one starts with a martingale or submartingale $\{h_i|i\in\mathscr{I}\}$, then each φ_i is defined by the formula

$$\text{1)} \qquad \varphi_i(E) = \int_E h_i\, dP \quad \forall E\in\mathscr{A}_i.$$

We will establish the convergence of the functions $d\varphi_i/dP$ as a corollary of Theorem 3.2. Our method is to show that $d\varphi_i/dP$ converges almost everywhere on Ω by showing that on a nonstandard extension $^*\Omega$ of Ω, the extended-real valued functions $^\circ(^*d\varphi_i/dP)$, $i\in\mathscr{I}$, converge almost everywhere. An alternate approach of Baez—Duarte [3] for the special case of an L^1-bounded martingale $\{x_n\}$, applies a weak form of the Andersen—Jessen theorem to the image of $\{x_n\}$ in an appropriate product space. For our approach, we need some terminology from [1] and [12].

Fix an $\aleph_1$-saturated, nonstandard extension of a standard structure containing $\mathbf{R}$ and Ω. (See [12].) Suppose one is given an internal σ-algebra $\mathscr{C}\subseteq{}^*\mathscr{A}$ and an internal signed measure ν on $(^*\Omega,\mathscr{C})$ with the standard part of the internal total variation $^\circ|\nu|(^*\Omega)<+\infty$. The finitely additive set function $^\circ\nu$ has a unique, real-valued, countably additive extension $\hat{\nu}$ defined on the smallest σ-algebra $\sigma(\mathscr{C})$ containing $\mathscr{C}$. Following the literature (e.g., [1]), we call $(^*\Omega,\sigma(\mathscr{C}),\hat{\nu})$ the Loeb space generated by $(^*\Omega,\mathscr{C},\nu)$. As in [14], however, we do not assume that a Loeb space is complete since we will be working here with more than one measure.

Let $(^*\Omega,\mathscr{B}_\infty,\hat{P})$ be the Loeb space generated by $(^*\Omega,{}^*\mathscr{A},{}^*P)$. For each $i\in{}^*\mathscr{I}$, let $\mathscr{B}_i$ be the smallest σ-algebra containing $^*\mathscr{A}_i$, and let ψ_i denote the extension of $^\circ(^*\varphi_i)$ from $^*\mathscr{A}_i$ to $\mathscr{B}_i$. Recall that $\mathscr{I}$ is the set of standard indexes. Let $\mathscr{B}$ be the smallest σ-algebra containing $\mathscr{B}_i$ for every $i\in\mathscr{I}$, and fix an infinite $\eta\in{}^*\mathbf{N}-\mathbf{N}$. Now for each $i\in\mathscr{I}$, $\mathscr{B}_i\subseteq\mathscr{B}\subseteq\mathscr{B}_\eta\subseteq\mathscr{B}_\infty$. Let ψ denote the restriction of ψ_η to $\mathscr{B}$, and for each $i\in\mathscr{I}$, let ψ^i denote the restriction of ψ to $\mathscr{B}_i$. It follows that for each $i\in\mathscr{I}$, $\psi^i\geqq\psi_i$ and $\psi(^*\Omega)=\sup_{i\in I}\psi_i(^*\Omega)$. If the φ_i's are defined by Equation 1 using either a submartingale or a martingale, then for each $i\in\mathscr{I}$, $\psi^i\ll\hat{P}$, and for the case of a martingale, $\psi^i=\psi_i$. In general, it follows from a maximal inequality that the Radon—Nikodym derivatives $d(\psi^i-\psi_i)/d\hat{P}$ converge $\hat{P}$-almost everywhere to 0. To see this, fix $k\in\mathbf{N}$, $\varepsilon>0$, and a finite set $\{k=i_1<\ldots<i_m\}\subseteq\mathscr{I}$. Set

$$B_0=\emptyset,\quad B_n=\{\omega\in{}^*\Omega\,|\,d(\psi^{i_n}-\psi_{i_n})/d\hat{P}(\omega)\geqq\varepsilon\}-\textstyle\bigcup_{j=0}^{n-1}B_j,\quad 1\leqq n\leqq m.$$

We then have

$$(\psi^k-\psi_k)(^*\Omega)\geqq\varepsilon\cdot\hat{P}(B_1)+(\psi^{i_2}-\psi_{i_2})(^*\Omega-B_1)\geqq\ldots\geqq\varepsilon\cdot\hat{P}\big(\textstyle\bigcup_{n=1}^m B_n\big).$$

Now by Theorem 3.2, $d\psi^i/d\hat{P}$ converges $\hat{P}$-a.e. on $^*\Omega$ to $d\psi/d\hat{P}$. It follows that $d\psi_i/d\hat{P}$ converges $\hat{P}$-a.e. on $^*\Omega$ to $d\psi/d\hat{P}$. On the other hand, for each $i\in\mathscr{I}$, $d\psi_i/d\hat{P}={}^\circ(^*(d\varphi_i/dP))$. Therefore, by the following measure theoretic result (3.4), the Radon—Nikodym derivatives $d\varphi_i/dP$ converge P-almost everywhere on Ω to a function f such that $^\circ(^*f)=d\psi/d\hat{P}$ on $^*\Omega-U$, where U is a set of $\hat{P}$-measure 0 in $\mathscr{B}_\infty$.

3.4 Proposition. *Let $(Y,\mathscr{A},\lambda)$ be a finite measure space and $\{h_i|i\in\mathscr{I}\}$ a family of finite, measurable functions indexed by a countable, ordered set $\mathscr{I}\subseteq\mathbf{R}^+$*

with the ordering inherited from $\mathbf{R}^+$ *and* $\mathbf{N}\subseteq\mathcal{I}$. *Let* $({}^*Y, \mathcal{B}, \hat{\lambda})$ *denote the Loeb space generated by* $({}^*Y, {}^*\mathcal{A}, {}^*\lambda)$ *in an* $\aleph_1$*-saturated nonstandard extension of a standard structure containing* $\mathbf{R}$ *and* Y. *For each* $i\in\mathcal{I}$, *let* $g_i={}^\circ({}^*h_i)$. *Then the functions* h_i, $i\in\mathcal{I}$, *converge* λ*-almost everywhere on* Y *to a function* h *if and only if the functions* g_i, $i\in\mathcal{I}$, *converge* $\hat{\lambda}$*-almost everywhere on* ${}^*Y$ *to a function* g, *in which case* $g={}^\circ({}^*h)$ $\hat{\lambda}$*-a.e. on* ${}^*Y$.

Proof. If the h_i's converge almost everywhere to h, then by Egoroff's theorem, the g_i's converge almost everywhere to ${}^\circ({}^*h)$. For the converse, fix $\varepsilon>0$ in $\mathbf{R}$, $m\in\mathbf{N}$, and any finite set $m=i_1<i_2<\ldots<i_n$ in $\mathcal{I}$. Set

$$A = \{y\in Y \mid \exists j,\ 1\leqq j\leqq n, \text{ with } |h_m(y)-h_{i_j}(y)|\geqq\varepsilon\},$$

$$B = \{x\in {}^*Y \mid \exists j,\ 1\leqq j\leqq n, \text{ with } |g_{i_j}(x)|=+\infty \text{ or } |g_m(x)-g_{i_j}(x)|\geqq\varepsilon\}.$$

Then ${}^*A\subseteq B$, and so $\lambda(A)=\hat{\lambda}({}^*A)\leqq\hat{\lambda}(B)$. By Egoroff's theorem, $h_i(y)$, $i\in\mathcal{I}$, satisfies a Cauchy condition for λ-a.e. $y\in Y$. □

4. Differentiation of measures

Let (X,ϱ) be a metric space or even a quasi-metric space, where the triangle inequality for ϱ is replaced with the existence of a constant $K\geqq 1$ such that for all $x,y,z\in X$, $\varrho(x,z)\leqq K\cdot[\varrho(x,y)+\varrho(y,z)]$. (See Section 6.) For each set $S\subseteq X$, we write $\Delta(S)$ for the diameter of S and χ_S for the characteristic function of S. Let $\mathcal{B}$ denote the collection of Borel subsets of X, $\mathfrak{P}(\mathcal{B})$ denote the set of all subsets of $\mathcal{B}$, and M denote the set of all finite regular Borel measures on X. For each measure $\sigma\in M$ and each $x\in X$, let $\mathcal{B}_{x,\sigma}=\{S\in\mathcal{B} \mid x\in S \text{ and } \sigma(S)>0\}$, and let σ^* be the outer measure generated by σ.

4.1 *Definition.* Fix $\sigma\in M$ and an arbitrary subset A of X. A mapping $\mathcal{D}\colon A\to\mathfrak{P}(\mathcal{B})$ is called a differentiation basis with respect to σ on A (or just a differentiation basis) if the image of each $a\in A$ is a nonempty subset $\mathcal{D}(a)$ of $\mathcal{B}_{a,\sigma}$ with $\inf\{\Delta(D) \mid D\in\mathcal{D}(a)\}=0$. Such a differentiation basis has the Besicovitch—Vitali property with respect to parameters $\varkappa$ and η if for any differentiation basis $\mathcal{S}$ on A with $\mathcal{S}(a)\subseteq\mathcal{D}(a)$ for each $a\in A$, there is a mapping $a\mapsto S(a)\in\mathcal{S}(a)$ defined on a finite subset $A_0\subseteq A$ with

$$\sum\nolimits_{a\in A_0}\chi_{S(a)}\leqq\varkappa \quad\text{and}\quad \sigma^*(A)\leqq\eta\cdot\sum\nolimits_{a\in A_0}\sigma(S(a)).$$

We say a differentiation basis has the Besicovitch—Vitali property on a subset E of A if its restriction to E has this property with respect to some parameters $\varkappa$ and η, which may, of course, depend on E.

The assumption that A_0 is finite is not a restriction. If the Besicovitch—Vitali property holds without the finiteness condition on A_0, then A_0 must at least be countable since the measure σ is finite and $\sigma(S(a))$ is always positive. Replacing the parameter η with 2η allows one to replace A_0 with a finite subset A_1.

There always exists a differentiation basis with the Besicovitch—Vitali property on subsets of a bounded open set X in a finite dimensional normed vector space. Let σ be a Borel measure on X with support $Y \subseteqq X$. Let $\mathcal{D}$ be a differentiation basis such that for each $y \in Y$, $\mathcal{D}(y)$ is a set of closed balls $B(y, r)$ with centers equal to y. Then $\mathcal{D}$ has the Besicovitch—Vitali property not only on Y, but on every subset A of Y. Indeed, an immediate consequence of Theorems 5.3 and 5.4 is that for each set $A \subseteqq Y$ and each differentiation basis $\mathcal{S}$ with $\mathcal{S}(a) \subseteqq \mathcal{D}(a)$ for $a \in A$, one can choose the set A_0 as a subset of a countable set $A_c \subseteqq A$ such that $A \subseteqq \bigcup_{a \in A_c} S(a)$ (whence $\eta=2$ works) and a $\varkappa$ can be chosen that depends only on the dimension of X. Alternatively, Corollary 5.6 shows that A_0 can be chosen so that the sets $S(a)$, $a \in A_0$, are disjoint; i.e., $\varkappa=1$. It follows from Theorems 5.3 and 5.4, that shapes other than balls are also possible. In Section 6, we show that a differentiation basis with $\varkappa=1$ and $\eta \geqq 2$ exists for our more general space X if for an appropriate $m>1$,

$$\sup\left\{\frac{\sigma(B(a, m \cdot r))}{\sigma(B(a, r))} \,\middle|\, a \in A,\ r>0\right\} \leqq \frac{\eta}{2}.$$

Assuming, for now, that one has a differentiation basis with the Besicovitch—Vitali property, Theorem 1.1 yields a measure differentiation theorem in our general setting. To apply Theorem 1.1, we fix $\sigma \in M$ and set $M(\sigma)=M$. Let Y be the support of σ in X. For each $S \in \mathcal{B}$ and $\mu \in M$, let $F_S(\mu)=\mu(S)$. Let $\mathcal{D}$ be a differentiation basis with respect to σ on Y. Set $\mathcal{F}=\bigcup_{y \in Y}\{F_D | D \in \mathcal{D}(y)\}$, and for each $y \in Y$, let $\mathfrak{F}(y)$ be the filter generated by the sets $\{F_D | D \in \mathcal{D}(y), \Delta(D) \leqq 1/n\}$.

4.2 Theorem. *If $\mathcal{D}$ has the Besicovitch—Vitali property on every subset of Y, then for each $\mu \in M$ and σ-almost every $y \in Y$,*

$$\lim_{D \in \mathcal{D}(y),\, \Delta(D) \to 0} \frac{\mu(D)}{\sigma(D)} = \frac{d\mu}{d\sigma}(y).$$

Proof. We show that Condition 3 of Theorem 1.1 holds by fixing a Borel set $E \subseteqq Y$ and a measure $\nu \in M$ with $\sigma(E)>0$, and $\nu(E)=0$. For each point $x \in E$, let $\mathcal{S}(x)=\{S \in \mathcal{D}(x) | \sigma(S)<\nu(S)\}$. Let A be the largest subset of E on which the mapping $a \mapsto \mathcal{S}(a)$ is a differentiation basis. We must show that $\sigma^*(A)=0$. Given $\varepsilon>0$ and the constants $\varkappa$ and η for A, we fix a compact subset $C \subseteqq X-E$ such that $\nu(X-C)<\varepsilon/(\varkappa \cdot \eta)$. For each $a \in A$, we remove sets intersecting C from $\mathcal{S}(a)$.

By assumption, since $\mathcal{S}$ is still a differentiation basis on A, there is a mapping $a \mapsto S(a) \in \mathcal{S}(a)$ defined on a finite subset $A_0 \subseteq A$ such that

$$\sigma^*(A) \leqq \eta \cdot \sum_{a \in A_0} \sigma(S(a)) \leqq \eta \cdot \sum_{a \in A_0} \nu(S(a)) \leqq \eta \cdot \varkappa \cdot \nu(X-C) < \varepsilon.$$

Since ε is arbitrary, $\sigma^*(A)=0$. □

Theorem 4.2 has an obvious generalization to topological spaces, but the covering theorems needed to establish the conditions of Theorem 4.2 seem to require a metric or quasi-metric space. The theorem we discuss in the next section reduces a covering to one where the overlap is controlled. In Section 6, we prove a general Vitali covering theorem establishing the existence of a disjoint family of sets which can be expanded to form a cover while controlling the increase of the reference measure. Before going to these more general settings, however, we consider the case $X=\mathbf{R}$. The authors are indebted to Jesus Aldaz for the following optimal covering lemma for the real line based on a result of T. Radó [17]. It shows that any differentiation basis formed by intervals has the Besicovitch—Vitali property on any subset of its domain.

4.3 Proposition. *Given any finite Borel measure σ on $\mathbf{R}$ and an arbitrary collection of non-degenerate intervals $\mathcal{I}$, there is for each $\varepsilon>0$ a finite disjoint subset $\{I_1, \cdots, I_n\} \subseteq \mathcal{I}$ such that $(2+\varepsilon) \cdot \sum_{k=1}^n \sigma(I_k) \geqq \sigma\left(\bigcup_{I \in \mathcal{I}} I\right)$.*

Proof. Note that $\left(\bigcup_{I \in \mathcal{I}} I\right)-\left(\bigcup_{I \in \mathcal{I}} I^\circ\right)$ is at most a countable set, so by Lindelöf's theorem we may assume that $\mathcal{I}$ is countable. We employ Radó's result after first reducing $\mathcal{I}$ to a finite collection $\mathcal{I}_f$ such that $\sigma\left(\bigcup_{I \in \mathcal{I}_f} I\right) \geqq (1+\varepsilon/2)^{-1} \sigma\left(\bigcup_{I \in \mathcal{I}} I\right)$ and each $I \in \mathcal{I}_f$ contains a point not in any other interval of $\mathcal{I}_f$. We order these points and the corresponding intervals so that for any indices i, j, and k with $i<j<k$ we have $x_i<x_j<x_k$ and thus $I_i \subseteq (-\infty, x_j)$ and $I_k \subseteq (x_j, +\infty)$. Since the intervals with even indices form a disjoint collection, as do the intervals with odd indices, the desired subset of $\mathcal{I}$ is whichever of these two families has the greater total measure. □

5. Besicovitch—Morse covering theorem

In this section we establish a generalization of a covering theorem first proved for disks in the plane by A. S. Besicovitch [4] and extended by A. P. Morse [16] to balls and more general sets in finite dimensional normed vector spaces. These theorems have two parts. The first part uses geometric reasoning to find an upper bound to the number of balls, or more general sets, that can be in what we shall call τ-satellite configuration. For the reader's convenience, we give a short proof of the geometric part of Morse's theorem. The non-geometric parts of the proofs in the literature (e.g., [11]) are here compressed and generalized using a relatively simple proof in Theorem 5.4. That theorem together with Morse's result, gives a covering theorem independent of any measure for spaces locally isometric to finite dimensional normed vector spaces. Combining these results with the differentiation theorem in Section 4 establishes a differentiation theorem for such spaces that works for any reference measure σ.

The setting of this section is again a metric or quasi-metric space (X, ϱ). (See Section 6.) As before, $B(a, r)$ denotes a closed metric ball with center a and radius $r>0$, and $\Delta(S)$ denotes the diameter of a set S. The geometric results mentioned above concern a configuration of sets which can be described even in this general setting.

5.1 *Definition.* Fix $\tau>1$ and $I_n=\{1, 2, ..., n\}\subset \mathbf{N}$. Let $\{a_i|i\in I_n\}$ and $\{S_i|i\in I_n\}$ be, respectively, an ordered set of points and an ordered set of bounded subsets in X. We say that the ordered collection of sets S_i is in τ-*satellite configuration* with respect to the ordered set of points a_i if the following conditions hold for each $i\in I_n$ and some index $i_0\in I_n$ called the central index:

i) $a_i\in S_i$, ii) $S_{i_0}\cap S_i \neq \emptyset$, iii) $\Delta(S_{i_0}) < \tau\cdot\Delta(S_i)$,

iv) If $i<j\leqq n$, then $a_j\notin S_i$ and $\Delta(S_j) < \tau\cdot\Delta(S_i)$.

Note that if $\tau<\tau_1$ and Conditions iii and iv hold for τ, then they hold for τ_1. In understanding Definition 5.1, it helps to take τ close to 1 and let each S_i be a ball with center a_i and radius r_i. In this case, the balls all intersecting $S_{i_0}=B(a_{i_0}, r_{i_0})$, each center is outside or almost outside every other ball, and r_{i_0} is almost the minimum radius. For the case that X is a finite dimensional normed vector space, there is an upper bound to the number of balls in such a configuration (see [15]). There is also, however, an upper bound for shapes more general than balls. We consider next (Theorem 5.2) a sufficient condition for the existence of such a bound.

In Theorems 5.2 and 5.3, we will assume that X is a finite dimensional vector space and that the metric ϱ is given by a norm $\|\cdot\|$. We will write $\mathrm{Card}(A)$ to denote the cardinality of a finite set A. For each $\gamma>0$, we let $N(\gamma)$ be an upper bound to the cardinality of any set A contained in the closed ball $B(0, 1)$ such that the distance between distinct points in A is at least $1/\gamma$. It is well-known that we can choose $N(\gamma)$ so that it depends only on γ and the dimension of X. Given points a, b, and c in X with $a\neq b$ and $a\neq c$, we set

$$U(a;\, b, c) = \left\| \frac{b-a}{\|b-a\|} - \frac{c-a}{\|c-a\|} \right\|.$$

The following result is essentially due to Morse (Theorem 5.9 in [16]).

5.2 Theorem. *Let* $(X, \|\cdot\|)$ *be a finite dimensional normed vector space. Fix* τ *with* $1<\tau\leqq 2$. *Suppose we are given a set* $I_n=\{1, 2, \dots, n\}\subset \mathbf{N}$ *and an ordered set* $\{S_i | i\in I_n\}$ *of bounded subsets of X in* τ*-satellite configuration with respect to an ordered set* $\{a_i | i\in I_n\}$ *of points in X. Also suppose that a ball* $B(a_i, r_i)$ *with center* a_i *is contained in* S_i *for each* $i\in I_n$. *Given the central index* i_0, *set* $a=a_{i_0}$, $r=r_{i_0}$ *and* $S=S_{i_0}$. *Assume there are constants* $C_0>0$ *and* $C_1\geqq 1$ *with the following property: If i and j are indices in* I_n *such that*

$$C_1\cdot r < \|a_i-a\| \leqq \|a_j-a\| \quad \text{and} \quad U(a; a_i, a_j) \leqq 1/C_0,$$

then a_i *must be a point in* S_j. *Then for* $\lambda=\max_{i\in I_n} \Delta(S_i)/2r_i$, *we have*

$$n \leqq N(2\lambda\cdot C_1)+N(8\lambda^2)\cdot N(C_0). \tag{1}$$

Proof. For $i<j$ in I_n,

$$\|a_i-a_j\| > r_i \geqq \Delta(S_i)/2\lambda \geqq \Delta(S)/4\lambda \geqq r/2\lambda.$$

Therefore, there are at most $N(2\lambda\cdot C_1)$ points a_i with $\|a_i-a\|\leqq C_1\cdot r$. If $i<j$ in I_n and $a_i\in S_j$, then

$$a_j\in B(a_i, \Delta(S_j)) \subseteqq B(a_i, 4\lambda\cdot r_i) \quad \text{and} \quad \|a_j-a_i\| \geqq r_i \geqq r_i/2\lambda.$$

If also $j<k$ in I_n, and $a_i\in S_k$, then $a_k\in B(a_i, 4\lambda\cdot r_i)$ and

$$\|a_k-a_j\| \geqq r_j \geqq \Delta(S_j)/2\lambda \geqq \|a_j-a_i\|/2\lambda \geqq r_i/2\lambda.$$

Therefore, $N(8\lambda^2)\geqq \mathrm{Card}(\{j\in I_n | a_i\in S_j\})$. Finally, consider the sets

$$J = \{i\in I_n | C_1\cdot r < \|a_i-a\|\} \quad \text{and} \quad J' = \{i\in J | \forall j\in J-\{i\},\ a_j\notin S_i\}.$$

By assumption, if i and j are in J', then $U(a; a_i, a_j)>1/C_0$. Therefore, $N(C_0)\geqq \mathrm{Card}(J')$ and $N(8\lambda^2)\cdot N(C_0)\geqq \mathrm{Card}(J)$, so Equation 1 holds. □

As an example, assume that $\lambda=1$, that is, $S_i=B(a_i, r_i)$ for each $i\in I_n$. Given τ sufficiently close to 1, one can show that indices i and j in I_n must be equal if $(3/2)\cdot r<\|a_i-a\|\leqq\|a_j-a\|$ and $U(a; a_i, a_j)\leqq 1/4$. (See [15].) It is no harder, however, to work with sets S_i which are starlike with respect to the points in $B(a_i, r_i)$. That is, for each $y\in B(a_i, r_i)$ and each $x\in S_i$, the line segment $\alpha\cdot x+(1-\alpha)\cdot y$, $0\leqq\alpha\leqq 1$, is contained in S_i. Given $\lambda\geqq 1$ and $a\in X$, we let $\mathscr{S}_\lambda(a)$ denote the collection of all sets $S\subseteqq X$ for which there exists an $r>0$ such that $B(a, r)\subseteqq S\subseteqq B(a, \lambda r)$ and S is starlike with respect to every $y\in B(a, r)$.

5.3 Theorem (Morse). *Let* $(X, \|\cdot\|)$ *be a finite dimensional normed vector space. Fix* $\lambda\geqq 1$ *and fix* τ *with* $1<\tau\leqq 2$. *If* $\{S_i | 1\leqq i\leqq n\}$ *is a finite ordered collection of subsets of X in* τ*-satellite configuration with respect to an ordered set* $\{a_i | 1\leqq i\leqq n\}\subset X$, *and for* $1\leqq i\leqq n$, $S_i\in\mathscr{S}_\lambda(a_i)$, *then*

$$n \leqq N(64\lambda^3)+N(8\lambda^2)\cdot N(16\lambda).$$

Proof. For $1 \leqq i \leqq n$, fix $r_i > 0$ so that $B(a_i, r_i) \subseteqq S_i \subseteqq B(a_i, \lambda r_i)$ and S_i is starlike with respect to every $y \in B(a_i, r_i)$. Let i_0 be the central index, and set $a = a_{i_0}$, $r = r_{i_0}$ and $S = S_{i_0}$. Let i and j be indices such that

$$32 \cdot \lambda^2 \cdot r < \|a_i - a\| \leqq \|a_j - a\| \quad \text{and} \quad U(a;\, a_i, a_j) \leqq 1/16\lambda.$$

By Theorem 5.2, we need only show that a_i must be in S_j. To simplify notation, let $b = a_i$ and $c = a_j$. Let $s = \|c-a\|/\|b-a\|$ and $t = 1/s$. Fix $x \in S \cap S_j$, and let $y = (1-s) \cdot x + s \cdot b$. Then $b = (1-t) \cdot x + t \cdot y$, so we need only show that $\|y - c\| \leqq r_j$. Now,

$$\begin{aligned} y - c &= (1-s)(x-a) + s(b-a) - (c-a) \\ &= (1-s)(x-a) + \|c-a\| \left(\frac{b-a}{\|b-a\|} - \frac{c-a}{\|c-a\|} \right). \end{aligned}$$

Since $16\lambda \cdot \Delta(S) \leqq 32 \cdot \lambda^2 \cdot r < \|b-a\|$, and $U(a; b, c) \leqq 1/16\lambda$,

$$\begin{aligned} \|y-c\| &\leqq s \cdot \|x-a\| + \|c-a\|/16\lambda \leqq s \cdot \|b-a\|/16\lambda + \|c-a\|/16\lambda \\ &= \|c-a\|/8\lambda \leqq (\|c-x\| + \|a-x\|)/8\lambda \leqq \Delta(S_j)/2\lambda \leqq r_j. \quad \square \end{aligned}$$

Theorem 5.3 shows that the hypotheses of the following theorem (5.4) are satisfied by rather general coverings in finite dimensional normed vector spaces and spaces locally isometric to finite dimensional normed vector spaces. Theorem 5.4 requires only the setting of a quasi-metric space, and it applies to shapes more general than balls. Our use of the notion of τ-satellite configuration yields a theorem which is more general than corresponding results in the literature. Even for Euclidean spaces, the use of ordering gives additional structure to those collections whose cardinality must be bounded in order to apply the theorem. We will only need the weaker statements of Corollaries 5.5 and 5.6 for the differentiation of measures. Indeed, for that application, it suffices to work on a bounded set, and then, if the bounded set is compact, the natural numbers suffice for the index set in the following proof.

5.4 Theorem. *Let A be an arbitrary subset of X. With each point $a \in A$, associate a set $S(a)$ containing a so that the diameters $\Delta(S(a))$, $a \in A$, have a finite upper bound. Assume that for some $\tau > 1$, there is an upper bound $\varkappa \in \mathrm{N}$ to the cardinality of any ordered set $\{a_i | 1 \leqq i \leqq n\} \subseteqq A$ with respect to which the ordered set $\{S(a_i) | i \leqq i \leqq n\}$ is in τ-satellite configuration. Then for some $m \leqq \varkappa$, there are disjoint subsets $A_1, \ldots, A_m$ of A such that $A \subseteqq \bigcup_{j=1}^m \bigcup_{a \in A_j} S(a)$ and for each j, $1 \leqq j \leqq m$, the elements of the collection $\{S(a) | a \in A_j\}$ are pairwise disjoint.*

Proof. Let T be a choice function on the nonempty subsets B of A such that $T(B)$ is a point $b \in B$ with $\tau \cdot \Delta(S(b)) > \sup_{a \in B} \Delta(S(a))$. Form a one-to-one correspondence between an initial segment of the ordinal numbers and a subcollection of A as follows. Set $B_1 = A$ and $a_1 = T(B_1)$. Having chosen a_α for $\alpha < \beta$, let $B_\beta =$

$A-\bigcup_{\alpha<\beta} S(a_\alpha)$. If $B_\beta \neq \emptyset$, set $a_\beta = T(B_\beta)$. There exists a first ordinal γ for which $B_\gamma = \emptyset$, that is, $A \subseteq \bigcup_{\alpha<\gamma} S(a_\alpha)$. Note that for $\alpha<\beta<\gamma$, we have $a_\beta \notin S(a_\alpha)$ and $\Delta(S(a_\beta)) < \tau \cdot \Delta(S(a_\alpha))$. Let $A_c = \{a_\alpha | \alpha<\gamma\}$, and let $T_f(B)$ denote the first element of each nonempty subset B of A_c.

Given any nonempty subset B of A_c, form a one-to-one correspondence between an initial segment of the ordinal numbers and a subset $V(B)$ of B as follows. Set $B_1 = B$ and $a(1) = T_f(B_1)$. Having chosen $a(\alpha)$ for $\alpha<\beta$, let

$$B_\beta = \{b \in B \,|\, \forall \alpha < \beta,\ S(b) \cap S(a(\alpha)) = \emptyset\}.$$

If $B_\beta \neq \emptyset$, set $a(\beta) = T_f(B_\beta)$. There exists a first ordinal γ for which $B_\gamma = \emptyset$. Let $V(B) = \{a(\alpha) | \alpha < \gamma\}$. For each $b \in B - V(B)$, there is an $a \in V(B)$ with $S(a) \cap S(b) \neq \emptyset$. If a is the first such point with respect to the ordering on A_c, then $\tau \cdot \Delta(S(a)) > \Delta(S(b))$.

Now for $i \geqq 1$, form the sets A_i by induction as follows. Set $A_1 = V(A_c)$. Having chosen A_i for $1 \leqq i \leqq n$, let $B_n = A_c - \bigcup_{i=1}^n A_i$. If $B_n \neq \emptyset$, set $A_{n+1} = V(B_n)$. Note that for each $a \in B_n$, there are points $a_i \in A_i$, $1 \leqq i \leqq n$, such that $S(a_i) \cap S(a) \neq \emptyset$ and $\tau \cdot \Delta(S(a_i)) > \Delta(S(a))$. Thus, the set $\{S(a_1), \ldots, S(a_n), S(a)\}$ is in τ-satellite configuration with respect to the set $\{a_1, a_2, \ldots, a_n, a\}$ when each is given the ordering inherited from A_c. Therefore, $B_n = \emptyset$ for some $n \leqq \varkappa$. □

5.5 Corollary. *For* $A_c = \bigcup_{j=1}^m A_j$, *we have* $\chi_A \leqq \sum_{a \in A_c} \chi_{S(a)} \leqq \varkappa$.

5.6 Corollary. *Assume that* $S(a)$ *is a Borel set for each* $a \in A$. *Then for any finite Borel measure* μ *on* X, *there is a* j *with* $1 \leqq j \leqq m$ *and a finite subset* $A_\mu \subseteq A_j$ *such that*

$$\mu^*(A) \leqq 2\varkappa \cdot \textstyle\sum_{a \in A_\mu} \mu(S(a)).$$

Proof. Choose the first $j \leqq m$ which maximizes the sum $\sum_{a \in A_j} \mu(S(a))$, and choose a finite subset $A_\mu \subseteq A_j$ so that

$$\tfrac{1}{2} \cdot \textstyle\sum_{a \in A_j} \mu(S(a)) \leqq \sum_{a \in A_\mu} \mu(S(a)). \quad \square$$

5.7 Theorem. *Fix a finite, regular Borel measure* σ *on* X, *and let* Y *be the support of* σ. *Let* $\mathscr{D}$ *be a differentiation basis with respect to* σ *on* Y. *Assume that for some* $\tau > 1$, *there is a finite upper bound to the cardinality of any ordered set* $\{a_i | 1 \leqq i \leqq n\} \subseteq Y$ *with respect to which an ordered set* $\{D(a_i) \in \mathscr{D}(a_i) | i \leqq i \leqq n\}$ *can be in* τ*-satellite configuration. Then for each Borel measure* μ *on* X *and for* σ*-almost every* $y \in Y$,

$$\lim_{D \in \mathscr{D}(y),\, \Delta(D) \to 0} \frac{\mu(D)}{\sigma(D)} = \frac{d\mu}{d\sigma}(y).$$

Proof. The result follows from Theorem 4.2 and either of the above corollaries. □

6. A general Vitali covering theorem

Again, as in the monograph of Coifman and Weiss [8], we work with a set X and a quasi-metric ϱ on X. Since in this section the constant associated with ϱ plays a role, we list all of the properties. That is, a quasi-metric ϱ is a nonnegative function on $X\times X$ such that
1) $\varrho(x, y)>0$ iff $x\neq y$,
2) $\varrho(x, y)=\varrho(y, x)$ $\forall x, y\in X$,
3) $\exists K\geqq 1$ such that $\forall x, y, z\in X$, $\varrho(x, z)\leqq K\cdot[\varrho(x, y)+\varrho(y, z)]$.
Of course, if $K=1$ we have a metric space. Again, $B(x, r)$ denotes the closed ball $\{y\in X|\varrho(x, y)\leqq r\}$.

The following theorem (6.1) is a generalization of the Vitali covering theorem. Because no conclusions are made about the countability of the covering, the statement and proof of Theorem 6.1 are somewhat simpler and more general than the corresponding result in [8]. Countability conditions are automatically fulfilled in the applications of Theorem 6.1. The authors are indebted to Professor Hiroshi Matano of Tokyo University for helpful conversations during the development of this section.

6.1 Theorem. *Let A be an arbitrary subset of X. With each $a\in A$ associate a set $S_a\subseteqq X$ containing a. Assume there is a $\lambda\geqq 1$ such that for each $a\in A$, there is a point $c(a)\in S_a$ and a positive number $r(a)$ with*

$$B(c(a), r(a))\subseteqq S_a\subseteqq B(c(a), \lambda\cdot r(a)).$$

Also assume that $\sup_{a\in A} r(a)<+\infty$. Then there is a subset $A_0\subseteqq A$ with the following properties: If $a\neq b$ in A_0, then $S_a\cap S_b=\emptyset$, and

1) $$A\subseteqq \bigcup_{a\in A_0} B(c(a), 5\cdot K^2\cdot\lambda\cdot r(a)).$$

Proof. For each nonempty subsets E of A, let $T(E)$ be an element $b\in E$ such that $2\cdot r(b)>\sup\{r(a)|a\in E\}$. We form a one-to-one correspondence between an initial segment of the ordinals and a subset of A as follows. Let $a_1=T(A)$. Having chosen a_α for all ordinals $\alpha<\beta$, let

$$E_\beta=\{b\in A|\forall\alpha<\beta,\ b\notin B(c(a_\alpha), 5\cdot K^2\cdot\lambda\cdot r(a_\alpha))\}.$$

If $E_\beta\neq\emptyset$, set $a_\beta=T(E_\beta)$. There must exist a first ordinal γ such that $E_\gamma=\emptyset$. That is, for $A_0=\{a_\alpha|\alpha<\gamma\}$, Equation 1 holds.

Now for $\alpha<\beta<\gamma$, $S_{a_\alpha}\cap S_{a_\beta}=\emptyset$. To prove this, we use notation such as S_α for the set S_{a_α}, c_α for $c(a_\alpha)$ and r_α for $r(a_\alpha)$. Assume there is a point $b\in S_\alpha\cap S_\beta$. By our construction, $2\cdot r_\alpha>r_\beta$ and $\varrho(c_\alpha, a_\beta)>5\cdot K^2\cdot\lambda\cdot r_\alpha$. On the other hand, since $b\in S_\alpha\cap S_\beta$,

$$\varrho(c_\alpha, c_\beta)\leqq K\cdot[\varrho(c_\alpha, b)+\varrho(c_\beta, b)]\leqq K\cdot\lambda\cdot[r_\alpha+r_\beta]<3\cdot K\cdot\lambda\cdot r_\alpha.$$

Therefore,

$$\varrho(c_\alpha, a_\beta) \leqq K \cdot [\varrho(c_\alpha, c_\beta) + \varrho(c_\beta, a_\beta)] < 3 \cdot K^2 \cdot \lambda \cdot r_\alpha + K \cdot \lambda \cdot r_\beta$$
$$< 3 \cdot K^2 \cdot \lambda \cdot r_\alpha + 2 \cdot K \cdot \lambda \cdot r_\alpha \leqq 5 \cdot K^2 \cdot \lambda \cdot r_\alpha,$$

which is impossible. □

Given Theorem 4.2. The following corollary of Theorem 6.1 yields a Lebesgue differentiation theorem (6.3) that is valid for rather general spaces when the reference measure behaves like Lebesgue measure with respect to scaling. Of course, Lebesgue measure on Euclidean space is such a measure, as is a Riemannian measure associated with a compact Riemannian variety.

6.2 Corollary. *Let σ be a finite measure on X such that every ball is σ-measurable. Assume there is a constant $C \geqq 1$ such that for $a \in A$,*

$$\sigma\big(B(c(a), 5 \cdot K^2 \cdot \lambda \cdot r(a))\big) \leqq C \cdot \sigma\big(B(c(a), r(a))\big).$$

Then there is a finite subset A_1 of A_0 such that

$$\sigma^*(A) \leqq 2C \cdot \textstyle\sum_{a \in A_1} \sigma^*(S_a).$$

Proof. The sum $\sum_{a \in A_0} \sigma\big(B(c(a), r(a))\big)$ is finite since σ is a finite measure. We may choose a finite subset A_1 of A_0 so that

$$\sigma^*(A) \leqq 2 \cdot \textstyle\sum_{a \in A_1} \sigma\big(B(c(a), 5 \cdot K^2 \cdot \lambda \cdot r(a))\big) \leqq 2C \cdot \sum_{a \in A_1} \sigma\big(B(c(a), r(a))\big)$$
$$\leqq 2C \cdot \textstyle\sum_{a \in A_1} \sigma^*(S_a). \quad \square$$

6.3 Theorem. *Fix a finite regular Borel measure σ on X and let Y be the support of σ. Let $\mathscr{D}$ be a differentiation basis with respect to σ on Y. Assume that there are constants $\lambda \geqq 1$ and $C \geqq 1$ such that for each set $D \in \bigcup_{y \in Y} \mathscr{D}(y)$ there is a ball $B(c, r)$ with*

$$B(c, r) \subseteqq D \subseteqq B(c, \lambda \cdot r), \quad \text{and} \quad \sigma\big(B(c, 5 \cdot K^2 \cdot \lambda \cdot r)\big) \leqq C \cdot \sigma\big(B(c, r)\big).$$

Then for each regular Borel measure μ on X and for σ-almost every $y \in Y$,

$$\lim_{D \in \mathscr{D}(y), \Delta(D) \to 0} \frac{\mu(D)}{\sigma(D)} = \frac{d\mu}{d\sigma}(y).$$

Proof. The result follows from Theorem 4.2 and Corollary 6.2. □

References

1. Albeverio, S., Fenstad, J. E., Høegh-Krohn, R. and Lindstrøm, T., *Nonstandard Methods in Stochastic Analysis and Mathematical Physics,* Academic Press, Orlando, 1986.
2. Andersen, E. S. and Jessen, B., Some limit theorems on set-functions, *Danske Vid. Selsk. Mat.-Fys. Medd.* 25 (1948), 1—8.

3. Baez-Duarte, L., Another look at the martingale theorem, *J. Math. Anal. Appl.* **23** (1968), 551—557.
4. Besicovitch, A. S., A general form of the covering principle and relative differentiation of additive functions (I), (II), *Proc Cambridge Phil. Soc.* Vol. **41** (1945), pp 103—110, Vol. **42** (1946), 1—10.
5. Bliedtner, J. and Hansen, W., *Potential Theory,* Springer—Verlag, Berlin—Heidelberg—New York, 1986.
6. Bliedtner, J. and Loeb, P. A., A measure-theoretic boundary limit theorem, *Archiv der Math.,* Vol. **43** (1984), 373—376.
7. Chatterji, S. D., Differentiation along algebras, *Manuscripta Math.* **4** (1971), 213—224.
8. Coifman, R. R. and Weiss, G., *Analyse Harmonique Non-Commutative sur certains Espaces Homogenes.* Lect. Notes in Math. 242, Springer—Verlag, Berlin—Heidelberg—New York, 1971.
9. Doob, J. L., *Stochastic Processes,* Wiley, New York, 1953.
10. Doob, J. L., *Classical potential theory and its probabilistic counterpart,* Springer—Verlag, Berlin—Heidelberg—New York, 1984.
11. Federer, H., *Geometric Measure Theory,* Springer—Verlag, Berlin—Heidelberg—New York, 1969.
12. Hurd, A. E. and Loeb, P. A., *An Introduction to Nonstandard Real Analysis,* Academic Press, Orlando, 1985.
13. Johansen, S. and Karush, J., On the semimartingale convergence theorem, *Annals Math. Stat.* **37** (1966), 690—694.
14. Loeb, P. A., Conversion from nonstandard to standard measure spaces and applications in probability theory, *Trans. Amer. Math. Soc.* **211** (1975), 113—122.
15. Loeb, P. A., On the Besicovitch Covering Theorem, *Science University of Tokyo Jour. Math.* **25** (1989), 51—55.
16. Morse, A. P., Perfect Blankets, *Trans. Amer. Math. Soc.* **61** (1947), 418—442.
17. Radó, T., Sur on problème relatif à un théorème de Vitali, *Fundamenta Mathematicae* **11** (1928), 228—229.
18. Taylor, J. C., An elementary proof of the theorem of Fatou—Naïm—Doob, *Can. Math. Soc. Conf. Proc.* **1** (1981), 153-163.

Received May 22, 1990

J. Bliedtner
Fachbereich Mathematik
Universität Frankfurt
Robert Mayer Str. 6—8
D—6000 Frankfurt/M
FRG

P. A. Loeb
Department of Mathematics
University of Illinois
1409 West Green Street
Urbana, IL 61801
USA