Besicovitch's covering theorem and differentiation

Notation and Facts

- For $x = (x_1, \ldots, x_d)$ in $\mathbb{R}^d$, define $|x| := \sqrt{\sum_i x_i^2}$, the usual ℓ^2 norm on $\mathbb{R}^d$, and use $\langle \cdot, \cdot \rangle$ for the corresponding inner product.
- Write $\mathfrak{m}_d$ for Lebesgue measure on $\mathbb{R}^d$.
- Let $B[a, r] := \{x \in \mathbb{R}^d : |x - a| \leq r\}$ denote the closed ball in $\mathbb{R}^d$ with center a and radius r. If $r = 0$ then the ball is degenerate; it reduces to the singleton set $\{a\}$. The corresponding open ball is denoted by $B(a, r)$. That is $B(a, r) := \{x \in \mathbb{R}^d : |x - a| < r\}$.
- The symbol Δ will denote $\mathfrak{m}_d B[0, 1]$. Fortunately, it is not necessary to know that $\Delta = \pi^{d/2}/\Gamma(1 + d/2)$. Invariance properties of Lebesgue measure give $\mathfrak{m}_d B[a, r] = \Delta r^d$ for each $a \in \mathbb{R}^d$ and $r \geq 0$.
- A measure λ on $\mathcal{B}(\mathcal{R}^d)$ is a called a Radon measure if $\lambda K < \infty$ for each compact K and $\lambda D = \sup\{\lambda K : K \subset D \text{ and } K \text{ is compact }\}$ for each Borel set D. Such a measure is necessarily locally finite: for each x in $\mathbb{R}^d$ there is an open neighborhood U of x with $\lambda U < \infty$, because $\mathbb{R}^d$ is locally compact. (In fact every locally finite λ on $\mathcal{B}(\mathbb{R}^d)$ must be a Radon measure.) Radon measures are also outer regular: $\lambda D = \inf\{\lambda G : G \supset D \text{ and } G \text{ open }\}$

- The support of a Radon measure λ, henceforth denoted by S_λ or $\text{SUPP}(\lambda)$, is the smallest closed set F for which $\lambda F^c = 0$. Each nondegenerate open ball $B(x,r)$ for x in $\text{SUPP}(\lambda)$ has nonzero λ-measure, for otherwise we would have $\lambda\left(F^c \cup B(x,r)\right) = 0$. The ratio $\mu B[x,r]/\lambda[x,r]$ is well defined for each $r > 0$ and $x \in \text{SUPP}(\lambda)$

This note collects together some ideas that I have learned by reading parts of:

- S := Simon (1983)
- M := Mattila (1999)
- EG := Evans and Gariepy (2015)

I also found that many of the methods from P :=Pollard (2001, Chap 3), for $\mathfrak{m}_d$ and differentiation of measures dominated by $\mathfrak{m}_d$ carry over to more general measures on $\mathcal{B}(\mathbb{R}^d)$ once the analog (Theorem <7>) of the Vitali Covering Lemma (P, page 68) is established.

1 Motivation

S:motivation

Several years ago I was asked to referee a paper on the interpretation of conditional probability distributions. The authors made heavy use of facts about measures on Euclidean spaces. In particular they relied on theory described by the excellent book of Evans and Gariepy (2015), particularly those parts involving the "area formula". Eventually, because my knowledge in this general area was mostly rusted away, I bought that book then read it up to end of their chapter 3.

Along the way to the 'area formula' I learned a lot about differentiation theorems. In particular, I learned that such theorems were not just restricted to measures on $\mathbb{R}^d$ that had densities with respect to Lebesgue measure but could also be extended to general Radon measures on Euclidean spaces. The main tool was a result due to Besicovitch (1945), which took over the role of the Vitale covering theorem that I learned many years ago.

Bes <1> **Theorem.** *(Besicovitch) Let A be a bounded subset of $\mathbb{R}^d$ and $\mathcal{B}_A$ be a set of closed balls $\{B[a,r(a)] : a \in A\}$, where $r(a) > 0$ for each a in A. Then there is a finite or countably infinite sequence $T = (a_1, a_2, \dots)$ in A, with $r(a_i)$ a decreasing function of i, for which*

(i) $\mathbb{1}\{x \in A\} \le \sum_{a\in T} \mathbb{1}\{x \in B[a,r(a)]\} \le 15^d$. for each x in $\mathbb{R}^d$. That is, $\{B[a,r(a)] : a \in T\}$ covers A but no point of $\mathbb{R}^d$ is contained in more than 15^d members of the covering.

(ii) T can be partitioned into subsequences $T_1, \dots, T_n$ with $n \le N_d := 1 + 60^d$ such that, for each γ, all the balls $\{B[a,r(a)] : a \in T_\gamma\}$ are disjoint.

□

To simplify notation I also abbreviate $B[a, r(a)]$ to B_a.

After working my way through the proof given by Evans and Gariepy (2015, 39–46) I stumbled on another excellent account of the Besicovish theorem given by Mattila (1999, pages 28–34). I even tried to read the original 8 page paper by Besicovitch (1945), without much gain in my understanding; and I got totally lost trying to read Federer (1969), which seems to be the source for real experts in this area. One of those experts is Leon Simon, whose oft revised lecture notes (Simon, 1983) were very helpful to me, particularly during my initial struggles.

The proof of Theorem <1> (given in the next Section) consists mainly of repeated applications of two greedy algorithms, which I very imaginatively call GREEDY$_1$ and GREEDY$_2$.

The set T will be constructed as a finite or countably infinite sequence made up of finite blocks, $T = (S_1, S_2, \dots)$, with each S_γ being obtained by an appeal to GREEDY$_1$. This T will satisfy (i). It will also have the useful property that $r(a)$ is decreasing (maybe better: nonincreasing, as one is forced to say to when there is any danger that decreasing might be misinterpreted to imply strict inequality) as we move along the sequence.

Then disjoint subsequences $T_1, T_2, \dots$ of T will be defined using GREEDY$_2$. For each i, the balls $\{B_a : a \in T_i\}$ will be disjoint. Finally T will be replaced by $\cup_{i \le N} T_i$ for a cunningly chosen N, which will be shown to preserve property (i). By construction, the new T will satisfy (ii).

For me, it was at first puzzling that Lebesgue measure $\mathfrak{m}_d$ should figure so prominently in an argument that is mostly intended for application to other measures on $\mathbb{R}^d$, measures that need not have a density with respect to $\mathfrak{m}_d$. Eventually it dawned on me that it is the invariance and scaling properties of $\mathfrak{m}_d$ that provide the vital information about geometrical properties of sets of closed balls. Those properties translate into pointwise inequalities, which can then be integrated with respect to any finite measure.

2 Description of the greedies, for use while walking

S:walk

Remember that we have a bounded set A and for each point of A we are given a nondegenerate ball centered at that point. (Nondegenerate means that the radius is not zero). We are trying to find a sequence (T) of centers, sorted in order of decreasing radius of the corresponding balls, such the union covers A. We allow some overlap in the balls in the sequence but want no point in the underlying euclidean space $\mathbb{R}^d$ to be covered more than some fixed number (depending on the dimension of the space) of times.

If we want the final sequence to be in order of decreasing radius, it might seems that we should just start with the biggest ball and work our way down. Of course that can't work because an uncountable set of radii need not have a largest member. Instead we can find the supremum—call it R—of the radii in the given collection of balls then initially consider only those balls with radius lying between R and $R/2$.

Start by picking any center in A whose ball has radius in the $R/2$ to R range. Carve that ball away from A then choose another center (again in the $R/2$ to R range) from the bit of A left after the first carving. And so on.

By construction, the center of each ball in the sequence lies outside all the previous balls in the sequence, which implies that all the selected centers are at least $R/2$ apart. There can be only finitely many such centers in a bounded region, so the selections from the $R/2$ to R range run out of centers to choose from after a finite number of steps. Take that finite sequence of balls, sort them into the order of decreasing radius, then use them to start the T sequence.

We have now reduced the set A to a subset. Call it A_1. For all the centers in A_1 the radius of the corresponding ball is at most $R/2$. (Otherwise, such a ball would have been chopped out by the initial greedy procedure.)

Now start all over again from the centers in A_1. And so on.

The bit about the number of balls in the sequence that contain a given point in $\mathbb{R}^d$ needs a picture (see Lemma <2>). It reduces to the task of bounding the number of points that can be placed on the unit sphere with mutual separations of at least $1/2$. (The idea is: the vectors joining the common point to each of the centers of the balls cut the unit sphere at points that inherit a separation from the assumed separation of the centers.)

The partitioning of T into subsequences, with the balls in each subsequence being disjoint, is also just an exercise in greed. Think of T as a sequence of balls, not just centers. For T_1 start with the first ball in T. Call it B_1. Then hunt along T for the first ball disjoint from B_1. Call it B_2. Then continue the hunt for the first ball that is disjoint from both B_1 and B_2. And so on. It might take quite a while to process the whole of T in this way. You might want to hire an inductive assistant who can process an infinite sequence in a finite time.

Remove T_1 from the sequence T, then repeat the procedure to construct T_2, starting from the first ball of T that is not in T_1. You will probably need many inductive assistants.

Actually, you won't need an infinite number of assistants: a little piece of geometry will show that $T_1, T_2, \dots, T_n$, for some n depending on the dimension, will be enough to cover the original A.

3 Proof of the Besicovich Covering Theorem

`S:covering`

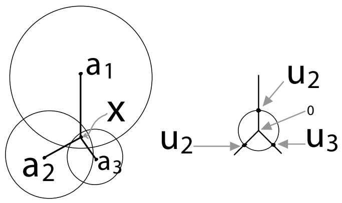
This Section is based on M(28-34). My contributions consist mainly of minor modifications, such as drawing pictures and emphasizing that the construction is based on greedy arguments. The proof makes use of two simple geometric properties of Euclidean space.

`separation` <2> **Lemma.** *Let $B[a_i, r_i]$ for $1 \le i \le k$ be closed balls in $\mathbb{R}^d$ with the properties that $\bigcap_{i=1}^k B[a_i, r_i] \ne \emptyset$ and*

`\E@ centers` <3> $$a_i \notin B[a_j, r_j] \qquad \text{for each pair } (i, j) \text{ with } i \ne j.$$

Then $k \le 3^d$.

> **Remark.** Condition <3> could also be written as: $|a_i - a_j| > \max(r_i, r_j)$ for all $i \neq j$.



Proof. Let x be a point in $\bigcap_{i=1}^k B[a_i, r_i]$. If r_i were 0 then x would equal a_i, which would violate <3>. Thus $r_i > 0$ for each i.

For each i we have $a_i = x + s_i u_i$ with u_i a unit vector and $0 < s_i \le r_i$. Consider a pair with $i \neq j$. Without loss of generality suppose $s_i \ge s_j$. Then, by <3>,

$$s_i^2 \le \max(r_i^2, r_j^2) < |a_i - a_j|^2 = |s_i u_i - s_j u_j|^2 = s_i^2 + s_j^2 - 2 s_i s_j \langle u_i, u_j \rangle,$$

which implies $2\langle u_i, u_j\rangle < s_j^2/(s_i s_j) \le 1$. It follows that

$$|u_i - u_j|^2 = 2 - 2\langle u_i, u_j \rangle > 1 \qquad \text{for } i \neq j.$$

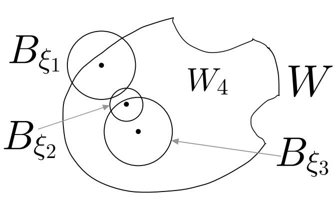
The balls $B[u_i, 1/2]$ are disjoint and each has Lebesgue measure $\Delta(1/2)^d$. The union $\cup_{i \le k} B[u_i, 1/2]$ is a subset of $B[0, 3/2]$, which has Lebesgue measure $\Delta(3/2)^d$. It follows that $k\Delta(1/2)^d \le \Delta(3/2)^d$. □

> **Remark.** The final paragraph of the proof used the standard trick for bounding packing numbers in Euclidean spaces, exploiting invariance properties of Lebesgue measure. The inequality $\langle u_i, u_j \rangle < 1/2$ implies that the angle between the two unit vectors is greater than $60°$, as asserted by M(Lemma 2.5).

`shrink` *<4>* **Lemma.** *Suppose $b \in B[a, r]$ with $r > 0$. Then for each R in $[0, r]$ there is an x for which $b \in B[x, R] \subset B[a, r]$.*

Proof. Without loss of generality $a = 0$, so that $b = su$ for some unit vector u and some s in $[0, r]$. If $s \le R$ choose $x = 0$. If $s = R + \delta$ with $\delta > 0$ choose $x = \delta u$. Then $|b - x| = s - \delta = R$, implying $b \in B[x, R]$. Also, if $y \in B[x, R]$ then $|y| \le |x| + R = s \le r$, implying $B[x, R] \subset B[0, r]$. □

As explained in Section 1, most of the proof for part (i) of the Theorem consists of repeated applications of the following greedy procedure. For the first application, W will equal A.



procedure GREEDY$_1(W)$
  ▷ *W can be any subset of A.* ◁
  Initialize: $R \leftarrow \sup\{r(a) : a \in W\}$ and $j \leftarrow 1$ and $W_1 \leftarrow W$.
  ▷ *On the first pass through the loop j is 1 and WD_j equals W.* ◁
  while $\{a \in W_j : r(a) > R/2\} \neq \emptyset$ **do**
    Arbitrarily choose some member ξ_j of W_j with $r(\xi_j) > R/2$.
    ▷ *Any such ξ_j suffices; no cleverness here.* ◁
    $W_{j+1} \leftarrow W_j \backslash B[\xi_j, r(\xi_j)]$
    Increment j by 1.

▷ *The boundedness of A forces the loop to exit after some finite number of iterations. After the loop completes k iterations we have $\xi_j \notin B[\xi_i, r(\xi_i)]$ for $1 \le i < j \le k$. Thus $|\xi_j - \xi_i| > r(\xi_i) > R/2$ for $i < j$. The closed balls $B[\xi_j, R/4]$ for $1 \le j \le k$ are disjoint with union contained in $B[0, \mathrm{diam}(A) + R/4]$. It follows that $k\Delta(R/4)^d \le \Delta\left(\mathrm{diam}(A) + R/4\right)^d$, which provides an upper bound for k.* ◁

Suppose the loop completes J iterations before exiting.

Define S as the finite sequence $\{a_1, \dots, a_J\}$ obtained by sorting $(\xi_1, \dots, \xi_J)$ to ensure that $i \mapsto r(a_i)$ is decreasing.

return (W_{J+1}, S, R).

The proof for part (ii) of the Theorem consists of repeated applications of GREEDY$_2$. Initially, D will be the sequence T constructed in the first part of the proof. It will generate a subsequence E of D for which the balls B_a with $a \in E$ are disjoint.

procedure GREEDY$_2(D)$

  ▷ *$D = (a_i : i \in I)$ is a sequence of points in A, with $I = \mathbb{N}$ or $I = \{1, 2, \dots, n\}$ for some finite n, with $r(a_i)$ decreasing.* ◁

  Initialize: Attach label 'extract' to a_1 and set j equal to 2.

  while $j \in I$ **do**

    Attach label 'extract' to a_j if $B[a_j, r(a_j)]$ has an empty intersection with $B[a_i, r(a_i)]$ for each i that is $< j$ and is labelled 'extract'. Otherwise attach label 'keep'.

    Increment j by 1.

  Let E be the subsequence of D consisting of the elements labelled 'extract' and K be the subsequence of D labelled 'keep'.

  return (E, K).

  ▷ *The balls B_a for $a \in E$ are disjoint. The procedure* GREEDY$_2(K)$ *will then extract another disjoint sequence. And so on.* ◁

Proof (of Theorem <1>). Without loss of generality, suppose $0 \in A$, so that the set A is contained within the ball $B[0, \mathrm{diam}(A)]$. Also, the whole Theorem is trivial if there exists an $a \in A$ for which $r(a)$ is larger than $\mathrm{diam}(A)$. So assume that $\sup\{r(a) : a \in A\} \le \mathrm{diam}(A)$. Define $A_0 := A$.

The construction starts off with $(A_1, S_1, R_1) \leftarrow$ GREEDY$_1(A_0)$. The set S_1 is finite with $\mathrm{diam}(A) \ge R_1 \ge r(a) > R_1/2$ for each a in S_1 and $r(a') \le R_1/2$ for each a' in the set $A_1 := A \backslash \cup_{a \in S_1} B_a$. For each pair a, a' of distinct members of S_1 we have $|a - a'| > R_1/2$ because $\min\left(r(a), r(a')\right) > R_1/2$ and either $a' \notin B[a, r(a)]$ or $a \notin B[a', r(a')]$. Take S_1 to be the first $|S_1|$ members of the sequence T.

> **Remark.** The next three paragraphs could be compressed into a more cryptic inductive assertion. I prefer a more verbose description, both to provide a check on notation and to make sure there are no special complications with the initial step. Many authors would prefer to

> write: inductively, assuming the ith step is completed, leaving a nonempty A_{i+1}, for the $(i+1)$st step we ... .

If $A_1 \neq \emptyset$, run $(A_2, S_2, R_2) \leftarrow \text{GREEDY}_1(A_1)$ to define the next $|S_2|$ members of T. We then have:

- $A_2 = A_1 \setminus \cup_{a \in S_2} B[a, r(a)] = A \setminus \cup_{a \in S_1 \cup S_2} B[a, r(a)]$;
- $R_2 := \sup\{r(a) : a \in A_1\} \leq R_1/2$;
- $R_2 \geq r(a) > R_2/2$ for each a in S_2 ;
- $R_2/2 \geq r(a')$ for each a' in A_2 ;
- For each pair a, a' of distinct members of S_2 we have $|a - a'| > R_2/2$;
- For each a in S_1 and each a' in A_2 we have $|a - a'| > r(a) > R_1/2 \geq r(a')$ so that $|a - a'| > \max(r(a), r(a'))$, an instance of the property needed by Lemma <2>.

If $A_2 \neq \emptyset$ run $(A_3, S_3, R_3) \leftarrow \text{GREEDY}_1(A_2)$, and so on.
In summary, if $A_{\gamma-1} \neq \emptyset$ then $(A_\gamma, S_\gamma, R_\gamma) \leftarrow \text{GREEDY}_1(A_{\gamma-1})$ produces:

(a) $A_\gamma = A \setminus \cup \{B[a, r(a)] : a \in \cup_{\alpha \leq \gamma} S_\alpha\}$;

(b) $R_\gamma := \sup\{r(a) : a \in A_{\gamma-1}\} \leq R_{\gamma-1}/2$;

(c) $R_\gamma \geq r(a) > R_\gamma/2$ for each a in S_γ ;

(d) $R_\gamma/2 \geq r(a')$ for each a' in A_γ ;

(e) $|a - a'| > R_\gamma/2$ for each pair a, a' of distinct members of S_γ ;

(f) $|a - a'| > r(a) = \max(r(a), r(a'))$ for each a in S_γ and each a' in A_γ. Consequently, if $a \in S_\gamma$ and $a' \in S_\beta$, with, $\gamma \neq \beta$, then $|a - a'| > \max(r(a), r(a'))$.

The lower bound in assertion (i) of Theorem <1> is an easy consequence of (a)—(f): If $A_\gamma = \emptyset$ for some γ then

$$A \subset \cup\{B[a, r(a)] : a \in \cup_{\alpha \leq \gamma} S_\alpha\} \subseteq \cup_{a \in T} B[a, r(a)];$$

and if $A_\gamma \neq \emptyset$ for all γ then $r(a) \leq R_\gamma/2 \leq \text{diam}(A)/2^\gamma$ for all a in A_γ. If $x \in A$ and $r(x) > \text{diam}(A)/2^\gamma$ then $x \notin A_\gamma$, so that $x \in \cup_{\alpha \leq \gamma} \cup_{a \in S_\alpha} B[a, r(a)]$.

The upper bound in assertion (i) takes a little more work and help from Lemma <2>. For each $x \in \mathbb{R}^d$ the sum $\sum_{a \in T} \mathbb{1}\{x \in B[a, r(a)]\}$ counts the number of times that x is included in a covering ball.

First consider how many balls containing x could come from a single S_γ. Suppose there are ℓ of them. The centers of those balls would lie in $B[x, R_\gamma]$ and each pair of them would be separated by at least $R_\gamma/2$. The balls of radius $R_\gamma/4$ around those centers would be disjoint and would all be contained within $B[x, 5R_\gamma/4]$. It follows that $\ell\Delta(R_\gamma/4)^d \le \Delta(5R_\gamma/4)^2$, or $\ell \le 5^d$.

Next consider

$$\Gamma := \{\gamma : x \in B[t, r(t)] \text{ for at least one } t \text{ in } S_\gamma \ \}.$$

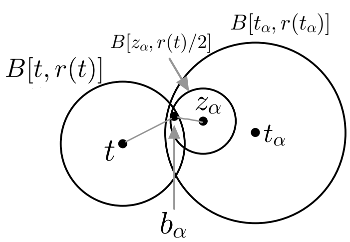
By (f), we have $|t - t'| > \max(r(t), r(t'))$ if $t \in S_\gamma$ and $t' \in S_\beta$, for distinct γ, β in Γ. Lemma <2> then ensures that $|\Gamma| \le 3^d$.

Thus there are at most $3^d \times 5^d$ balls centered in T that contain x.

For the proof of assertion (ii) of the Theorem, first decompose T into the union of disjoint subsequences $T_1, T_2, \dots$ with the balls $\{B[a, r(a)] : a \in T_\gamma\}$ disjoint for each γ. Start with $(T_1, K_1) = \text{GREEDY}_2(T)$. If $K_1 \ne \emptyset$, apply $(T_2, K_2) = \text{GREEDY}_2(K_1)$. And so on. The sequence T is the union of all the T_γ's.

The Theorem effectively claims that we only need a given finite number of the T_γ's to get balls that cover A. If we replace T by $\widetilde{T} = \cup_{\gamma \le N} T_\gamma$ then the upper bound in (i) is still holds, because $\widetilde{T} \subset T$. It remains only to show that the balls $\{B_a : a \in \widetilde{T}\}$ still cover A. The proof again involves invariance of Lebesgue measure.

Consider any $a \in A$. By assertion (i), there exists a t in T for which $a \in B[t, r(t)]$. There exists an n for which $t \in T_n$. We just need to show that $n \le 1 + 60^d$.



Think about why t ended up in T_n rather than in some T_α for an α with $1 \le \alpha < n$. For each such α there must be some $t_\alpha \in T_\alpha$ appearing earlier than t in the T sequence for which the corresponding ball has a nonempty intersection with $B[t, r(t)]$. That is, there exists at least one point b_α with $b_\alpha \in B[t_\alpha, r(t_\alpha)] \cap B[t, r(t)]$. Moreover, the decreasing property of the radius function on T ensures that $r(t_\alpha) \ge r(t)$.

By Lemma <4> there exists a z_α for which $b_\alpha \in B[z_\alpha, r(t)/2] \subset B[t_\alpha, r(t_\alpha)]$. If $x \in B[z_\alpha, r(t)/2]$ then

$$|t - x| \le |t - b_\alpha| + |b_\alpha - z_\alpha| + |z_\alpha - x| \le r(t) + r(t)/2 + r(t)/2 = 2r(t).$$

Thus $B[z_\alpha, r(t)/2] \subset B[t, 2r(t)]$.

Define $f(x) := \sum_{\alpha < n} \mathbb{1}\{x \in B[z_\alpha, r(t)/2]\}$ for $x \in \mathbb{R}^d$. The previous inclusion shows that

\E@ f=0 <5>

$$f(x) = 0 \qquad \text{for } x \notin B[t, 2r(t)].$$

The upper bound in (i) shows that no point in $\mathbb{R}^d$ is covered by more than 15^d balls of the form $B[t, r(t)]$ for $t \in T$. Consequently,

\E@ f.le15d <6>

$$f(x) \le \sum\nolimits_{\alpha < n} \mathbb{1}\{x \in B[t_\alpha, r(t_\alpha)]\} \le 15^d \qquad \text{for each } x \text{ in } \mathbb{R}^d.$$

Taken together, <5> and <6> give the neater bound:

$$f(x) \le 15^d \mathbb{1}\{x \in B[t, 2r(t)]\} \qquad \text{for each } x \text{ in } \mathbb{R}^d.$$

Integrate with respect to $\mathfrak{m}_d$:

$$\mathfrak{m}_d f = (n-1)\Delta\left(r(t)/2\right)^d \leq 15^d \Delta\left(2r(t)\right)^d,$$

which implies $n \leq 1 = 15^d \times 4^d$. To cover each point of a we only need balls $B[t, r(t)]$ with t in $\widetilde{T} := \cup_{\gamma \leq N} T_\gamma$. We can replace T by $\widetilde{T}$ and still have (i) holding. And we also get (ii). □

4 A Vitali-like result for Radon measures

S:Vitali

A collection $\mathcal{F}$ of closed subsets of $\mathbb{R}^d$ is said to ***cover a set*** A ***in the Vitali sense*** if to each a in A and each $\epsilon > 0$ there is an F in $\mathcal{F}$ with $\operatorname{diam}(F) < \epsilon$ for which $a \in F$. The collection is also called a ***Vitali covering*** for A.

Under some regularity assumptions that prevent the sets in such an $\mathcal{F}$ from being too 'thin', a classical result of Vitali asserts: for each Borel set A with $\mathfrak{m}_d A$ finite there is a finite or countably infinite subset $\mathcal{F}_A$ of $\mathcal{F}$ for which $\mathfrak{m}_d\left(A \backslash \cup \mathcal{F}_A\right) = 0$. (See Pollard, 2001, Section 3.5 for a proof.) With the help of Theorem <1>, the Theorem can be extended to arbitrary finite measures on $\mathcal{B}(\mathbb{R}^d)$, for $\mathcal{F}$ a Vitali covering consisting of closed balls. The next Section shows how this extension leads to differentiation theorems for pairs of finite measures on $\mathcal{B}(\mathbb{R}^d)$.

Vitali.Radon <7> **Theorem.** *Suppose A is a bounded, Borel-measurable subset of $\mathbb{R}^d$ and λ is a Radon measure. Let $\mathcal{F}$ be a collection of nondegenerate closed balls (that is, with radius > 0), with centers in A, that covers A in the Vitali sense. Then for each open set G containing A there exists a countable (or finite) subcollection $\mathcal{F}_A$ of $\mathcal{F}$ consisting of pairwise disjoint balls for which $\cup\mathcal{F}_A \subset G$ and $\lambda\left(A \backslash \cup \mathcal{F}_A\right) = 0$.*

Remark. The important role of G (and other G_i's in the proof) is not obvious until one sees the step in the argument where $\lambda(G_0 \backslash F_0)$ is used to bound $\lambda(A \backslash F_0)$. Surreptitiously, the Theorem is trying to break A into a disjoint union of closed balls plus a λ-negligible set. Clearly that is not literally possible. For example, think what would happen if a λ-negligible, countable, dense subset were removed from A, which would prevent it from containing any nondegerate closed ball. The next best thing is to break the enclosing G into a disjoint union of closed balls plus a a set of very small λ measure.

Proof. The following argument is based on M(page 34) and EG(page 45).

The Theorem is trivial unless $\lambda A > 0$. With N_d as in Theorem <1>, define $\epsilon = (4N_d)^{-1}$ and $\rho := 1 - \epsilon < 1$. Choose an open set G_0 with $G \supset G_0 \supset A$ and $\lambda G_0 < (1+\epsilon)\lambda A$. For each point a in A, choose an $\mathcal{F}$-ball $B_a := B[a, r(a)]$ with $r(a)$ small enough that $B_a \subset G_0$.

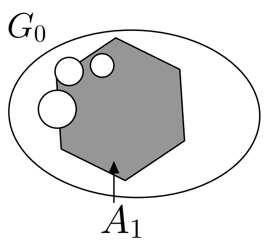
Remark. The strategy is to carve out from G_0 a closed set F_0 (a finite union of $\mathcal{F}$-balls) with λ measure at least $2\epsilon\lambda A$, thereby ensuring that $\lambda\left(A \backslash F_0\right)) \leq \rho\lambda A$. With a sequence of such steps we can whittle A down to a set with zero λ measure.

Invoke Theorem <1> to find T and $T_1, \dots, T_n$ with $n \le N_d$ for which $A \subset \cup_{\gamma \le n} D_\gamma$ where $D_\gamma := \cup_{a \in T_\gamma} B_a$. For each γ the balls $\{B_a : a \in T_\gamma\}$ are disjoint. Then we have

$$\lambda A \le \sum_{\gamma \le n} \lambda D_\gamma \le N_d \max_\gamma \lambda D_\gamma .$$

For the γ that maximizes λD_γ find k for which the closed set

$$F_1 := \cup\{B_a : a \text{ is one of the first } k \text{ members of } T_\gamma\}$$



satisfies the inequalities $\lambda F_1 \ge \lambda(D_\gamma)/2 \ge \lambda A/(2N_d)$. Remember that the balls contributing to F_1 are disjoint and their centers lie in A.

The set $A_1 := A \backslash F_1$ is contained in the open set $G_0 \backslash F_1$ and

$$\lambda A_1 \le \lambda G_0 \backslash F_1 = \lambda G_0 - \lambda F_1 < (1+\epsilon)\lambda A - (2N_d)^{-1}\lambda A = \rho\lambda A .$$

Choose another open set G_1 for which $A_1 \subset G_1 \subset G_0 \backslash F_1$ and $\lambda G_1 < (1+\epsilon)\lambda A_1$ then repeat the argument to find a closed subset F_2 that is a finite union of $\mathcal{F}$-balls contained in G_1 such that $\lambda F_2 \ge \lambda A_1/(2N_d)$. The set $A_2 := A_1 \backslash F_2$ then has $\lambda A_2 \le \rho^2 \lambda A$. Moreover $F_1 \cap F_2 = \emptyset$. And so on.

In this way we generate disjoint closed sets F_i, each a union of finitely many closed balls in G, for which $\lambda\left(A \backslash \cup_i F_i\right)) = 0$. The closed balls that make up all the F_i's are disjoint. □

5 Differentiation of measures

S:density

As shown by P(Sections 3.1 and 3.2), general results about densities (such as the Lebesgue decomposition) can be deduced from the special case where λ and μ are finite measures on $\mathcal{B}(\mathbb{R}^d)$ with $\lambda \ge \mu$. For that case, the Radon-Nikodym theorem shows that μ has a density $d\mu/d\lambda = h$ with respect to λ for which $0 \le h \le 1$. That is, $\mu A = \lambda h(x)\{x \in A\}$ for each Borel set A.

density.balls <8> **Theorem.** *Suppose λ and μ are finite measures on $\mathcal{B}(\mathbb{R}^d)$ with $\lambda \ge \mu$ and $h = d\mu/d\lambda$. Then*

$$\lim_{r \searrow 0} \frac{\mu B[x,r]}{\lambda[x,r]} = h(x) \qquad \text{for } \lambda \text{ almost all } x \text{ in } S_\lambda, \text{ the support of } \lambda.$$ □

It will be easy to extend this Theorem to a full Lebesgue decomposition for pairs of Radon measures. See Corollary <13>.

To avoid a lot fiddling with special cases, let me first note that, for fixed x in S_λ, the function $\Lambda_x(r) := \lambda B[x,r]$ is continuous from the right and nondecreasing. The limit from the left is given by $\Lambda_x(r-) = \lambda B(x,r)$. If there is a discontinuity in Λ_x at r then the size of the jump, $\Lambda_x(r) - \Lambda_x(r-)$, equals $\lambda\{y : |x-y| = r\}$. Disjointness of the spheres $\{y : |x-y| = r\}$ for different r values ensures that Λ_x can have at most countably many points of discontinuity. Consequently,

for each $r_0, \delta, \epsilon > 0$ there exists a continuity point r of Λ_x such that

\E@ cty.points <9> $r_0 < r < r_0 + \delta$ abd $|\Lambda_x(r) - \Lambda_x(r_0)| < \epsilon$.

Start of proof of Theorem <8>.
Write $\psi(x,r)$ for the ratio $\mu B[x,r]/\lambda B[x,r]$ for x in S_λ.

Let $\{\epsilon_n\}$ be any sequence that decreases to zero, such as $\epsilon_n = n^{-1}$. For each n and $x \in S_\lambda$ define

$$f_n(x) := \sup\{\psi(x,r) : 0 < r < \epsilon_n\},$$
$$g_n(x) := \inf\{\psi(x,r) : 0 < r < \epsilon_n\}.$$

The values $f_n(x)$ decrease to a limit $f(x)$; the values $g_n(x)$ increase to a limit $g(x)$. At each x in S_λ we have

$$1 \geq f_n(x) \geq f(x) \geq g(x) \geq g_n(x) \geq 0.$$

Theorem <8> will be an easy consequence of f and g being Borel-measurable functions for which

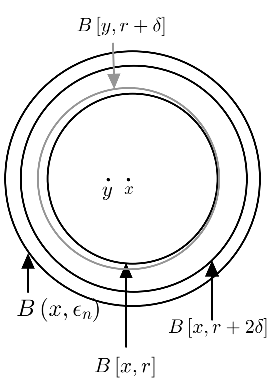
\E@ density.balls2 <10> $$f(x) = h(x) = g(x) \text{ ae}[\lambda] \text{ on } S_\lambda.$$

To establish <10>, it helps to first show (Lemma <11>) that each f_n and g_n (and hence f and g) is Borel-measurable. The main idea in the proof is that μ should be bigger than $s\lambda$ for sets where $f > s$ and smaller than $t\lambda$ for sets where $g < t$, whose proof (Lemma <12>) uses the Besicovitch results in the form of Theorem <7>. After that step the rest of the argument is a standard measure-theoretic exercise.

semi.cty *<11>* **Lemma.** *For each $t > 0$, both $\{f_n > t\} \cap S_\lambda$ and $\{g_n < t\} \cap S_\lambda$ belong to $\mathcal{B}(\mathbb{R}^d)$.*

Proof. Actually, as this Proof will show, both $G_t := \{x \in S_\lambda : f_n(x) > t\}$ and $H_t := \{x \in S_\lambda : g_n(x) < t\}$ are open as subsets of S_λ. That is, they are of the form $U \cap S_\lambda$, where U is an open subset of $\mathbb{R}^d$. To prove this assertion for G_t, we have to show that if $x \in G_t$ then $B(x,\delta) \cap S_\lambda \subset G$ for some $\delta > 0$.

Consider an x in S_λ at which $f_n(x) > t$. By definition of 'supremum', there must exist a ball $B = B[x, r_x]$ with $0 < r_x < \epsilon_n$ and $\psi(x, r_x) > t$. We may assume that r_x is a continuity point of Λ_x. (Actually, for this part of the argument it suffices to invoke continuity from the right. Continuity from the left is useful for the g_n result.) Choose $\delta > 0$ for which $B[x, r_x + 2\delta] \subset B(x, \epsilon_n)$. If $|y - x| < \delta$ then



$$B[x, r_x] \subset B[y, r_x + \delta] \subset B[x, r_x + 2\delta] \subset B(x, \epsilon_n).$$

Hence

$$f_n(y) \geq \psi(y, r_x + \delta) = \frac{\mu B[y, r_x + \delta]}{\lambda B[y, r_x + \delta]} \geq \frac{\mu B[x, r_x]}{\lambda B[x, r_x + 2\delta]}.$$

If δ is small enough then $\lambda B[x, r_x + 2\delta]$ is close enough to $\lambda B[x, r_x]$ to ensure that the final ratio is $> t$. That is, $B(x,\delta) \cap S_\lambda \subset G_t$.

The argument for g_n is similar, except that we need $0 < 2\delta < t$ and $\lambda B[x, r_x - 2\delta]$ close to $\lambda B[x, r_x]$, for which left-continuity of Λ_x at r_x is needed. □

inequalities <12> **Lemma.** *Suppose A is a bounded Borel-measurable subset of S_λ.*

(i) If $f(x) > s$ for all x in A then $\mu A \geq s\lambda A$.

(ii) If $t > g(x)$ for all x in A then $\mu A \leq t\lambda A$.

Proof. Let G be an open set with $A \subset G$. For (i), let $\mathcal{F}$ denote the set of all nondegenerate balls $B = B[a,r]$ with $a \in A$ and $B \subset G$ and $\mu B \geq s\lambda B$. I claim that $\mathcal{F}$ is a Vitali covering for A.

For each a in A there is a strictly positive δ_a for which $B[a,r] \subset G$ when $r < \delta_a$. In particular, if $\epsilon_n < \delta_a$ the inequality

$$s < f(a) \leq f_n(a) = \sup\{\psi(a,r) : 0 < r < \epsilon_n\}$$

implies the existence of a ball $B = B[a,r]$ with $0 < r < \epsilon_n$ and $B \subset G$ and $\psi(a,r) = \mu B/\lambda B > s$. That is, $B \in \mathcal{F}$.

By Theorem <7>, there is a subset $\mathcal{F}_A$ of $\mathcal{F}$ consisting of disjoint balls for which $\lambda(A \backslash D) = 0$, where $D := \cup \mathcal{F}_A$. The set D is Borel-measurable because $\mathcal{F}_A$ is at worst countably infinite. Then we have

$$\mu G \geq \mu D = \sum\nolimits_{B \in \mathcal{F}_A} \mu B \geq \sum\nolimits_{B \in \mathcal{F}_A} s\lambda B = s\lambda D.$$

Via the inequality $\lambda A \leq \lambda D + \lambda(A \backslash D) = \lambda D$ it then follows that $\mu G \geq s\lambda A$. Take the infimum over all open G with $A \subset G$ to deduce that $\mu A \geq s\lambda A$.

For (ii) the argument is similar, except that $\mathcal{F}$ should now be the set of all nondegenerate balls $B = B[a,r]$ with $a \in A$ and $B \subset G$ and $\mu B \leq t\lambda B$. □

To get <10>, first consider all the sets of the form

$$A = A_{s,t,R} := \{x \in S_\lambda : \ |x| \leq R \text{ and } f(x) > s > t > h(x)\}$$

By Lemma <12> part (i) and definition of the density h,

$$t\lambda A \geq \lambda h(x)\{x \in A\} = \mu A \geq s\lambda A.$$

The requirement $s > t$ forces a contradiction unless $\lambda A = 0$. Take a union over R in $\mathbb{N}$ and s,t ranging over pairs of positive rationals with $s > t$ to conclude that $f(x) \leq h(x)$ ae$[\lambda]$ on S_λ.

Argue similarly for the sets $\{x \in S_\lambda : \ |x| \leq R \text{ and } h(x) > s > t > g(x)\}$ to deduce that $g(x) \geq h(x)$ ae$[\lambda]$ on S_λ.

The fact that $f \geq g$ on S_λ then leads to the desired equality <10>, whence the desired the limit assertion <8>.

□ **End of proof of Theorem <8>.**

Radon <13> **Corollary.** *Suppose μ and ν are Radon measures on $\mathbb{R}^d$. Define $\lambda = \mu + \nu$. Then*

$$\frac{\mu B[x,r]}{\nu[x,r]} \to \frac{d\mu}{d\nu}(x) \qquad \text{as } r \downarrow 0, \text{ for } \nu \text{ almost all } x \text{ in } \textsc{supp}(\nu).$$

Proof. It is enough to prove the assertion for x in $B(0,R)$ for each finite R. Thus we lose no generality in working with the restrictions of μ and ν to that open set. Equivalently, we can just assume that $\lambda = \mu + \nu$ is finite.

As explained by Pollard (2001, Sections 3.1 and 3.2), there is a measurable function h with $0 \le h \le 1$ such that

$$\mu f = \lambda h f \text{ and } \nu f = \lambda(1-h)f \qquad \text{for each } f \text{ in } \mathcal{M}^+.$$

Here $\mathcal{M}^+$ denotes the set of all $[0,\infty]$-valued, Borel measurable functions on $\mathbb{R}^d$.

Necessarily $\mu\{h=0\} = 0 = \nu\{h=1\}$. The restrictions of the measures μ and ν to the set $\{0 < h < 1\}$ are mutually absolutely continuous. On the set $\{h < 1\}$ the measure μ has density $g(x) = h(x)/(1-h(x))$ with respect to ν. That is, $\mu f\{h<1\} = \nu f g\{h<1\}$ for each f in $\mathcal{M}^+$. By Theorem <8>,

$$\frac{\nu B[x,r]}{\nu B[x,r] + \mu B[x,r]} \to 1 - h(x) \qquad \text{as } r \downarrow 0, \text{ for } \lambda \text{ almost all } x \text{ in } \textsc{supp}(\lambda).$$

Excluding the ν-negligible set $\{h=1\}$, we can transform this assertion to

$$1 + \frac{\mu B[x,r]}{\nu B[x,r]} \to \frac{1}{1-h(x)} \qquad \text{for } \nu \text{ almost all } x \text{ in } \textsc{supp}(\nu),$$

which is equivalent to assertion of the Corollary. □

References

Besicovitch1945PCPS Besicovitch, A. S. (1945). A general form of the covering principle and relative differentiation of additive functions. *Mathematical Proceedings of the Cambridge Philosophical Society 41*, 103 – 110.

EvansGariepy2015 Evans, L. C. and R. F. Gariepy (2015). *Measure theory and fine properties of functions* (Revised ed.). CRC press.

Federer69book Federer, H. (1969). *Geometric Measure Theory*. Springer-Verlag.

Mattila1999geometry Mattila, P. (1999). *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*. Cambridge University Press.

PollardUGMTP Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

Simon1983gmt Simon, L. (1983). *Simon Lectures on geometric measure theory*. Proceedings of the Centre for Mathematics and its Applications. ANU Mathematical Sciences Institute (MSI) Canberra, Australia. Available online from https://openresearch-repository.anu.edu.au/handle/1885/13501; updated version online at https://web.stanford.edu/class/math285/ts-gmt.pdf.