# Quantitative Lebesgue Differentiation Theorem and Regularity Lemma

Li-Fu Chen

June 12, 2020

## 1 Introduction

In the field of analysis, it is common to distinguish hard (quantitative) analysis from soft (qualitative) analysis. Hard analysis deals with finite quantities and their quantitative properties. Soft analysis, on the other hand, tends to deal with more infinitary objects and their qualitative properties. It is well known that the results obtained by hard and soft analysis respectively can be connected to each other by various correspondence principles. In this report, we will develop some quantitative notions in hard analysis by refining the corresponding ones in soft analysis and then apply them to deduce the quantitative Lebesgue differentiation theorem and regularity lemma.

## 2 Preliminary

Before our discussion of the main topics, we should introduce two important theorems that will be used frequently in our analysis. The first one is **finite convergence principle** and the other one is **Hardy-Littlewood maximal inequality**.

**Theorem 2.1** (**Finite convergence principle**)**.** If $\varepsilon > 0$, $F : \mathbb{Z}^+ \to \mathbb{Z}^+$ is a function, and $0 \leq x_1 \leq x_2 \leq \ldots \leq x_M \leq 1$ is such that $M$ is sufficiently large depending on $F$ and $\varepsilon$, then there exists $1 \leq N < N + F(N) \leq M$ such that $|x_n - x_m| \leq \varepsilon$ for all $N \leq n, m \leq N + F(N)$.

**Proof.** First, we define the indexes recursively by $i_1 = 1$ and $i_{j+1} = i_j + F(i_j)$. That $M$ is sufficiently large depending on $F$ and $\varepsilon$ means $M \geq i_{\frac{1}{\varepsilon}+1}$.

Suppose it were wrong. Then for any $N$ with $1 \le N < N + F(N) \le M$, there exists $n, m$ with $N \le n, m \le N + F(N)$ such that $|x_n - x_m| > \varepsilon$. Considering $x_{i_1}, x_{i_2}, ..., x_{i_{\frac{1}{\varepsilon}+1}}$, we have $x_{i_{k+1}} - x_{i_k} > \varepsilon$ for $k = 1, 2, ..., \frac{1}{\varepsilon}$. Hence, $x_M - x_1 \ge \sum_{n=1}^{\frac{1}{\varepsilon}} (x_{i_{n+1}} - x_{i_n}) > \frac{1}{\varepsilon}\varepsilon = 1$, which is a contradiction. Thus, the finite convergence principle holds. $\square$

**Definition 2.1** (**Hardy-Littlewood maximal function**)**.** Given a locally integrable function $f : \mathbb{R}^d \to \mathbb{R}$. The function $M(f) : \mathbb{R}^d \to \mathbb{R}$ is defined by $M(f)(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\, dy$.

**Theorem 2.2** (**Hardy-Littlewood maximal inequality**)**.** For $d \ge 1$ and $f \in L^1(R^d)$, there exists a constant $C_d > 0$ such that for all $\lambda > 0$, we have $|\{x : M(f)(x) > \lambda\}| < \frac{C_d}{\lambda} \|f\|_{L^1}$.

**Proof.** We will apply the Vitali covering lemma to prove the inequality. Since the proof of the lemma can be easily deduced by modifying the proof learned in class, we omit the detailed proof and just state the result here.

**Vitali Covering Lemma**: Given $\mathcal{F}$ a family of open balls with bounded diameter, $\mathcal{F}$ contains a countable subfamily $\mathcal{F}'$ consisting of disjoint balls such that $\bigcup_{B \in \mathcal{F}} B \subset \bigcup_{B \in \mathcal{F}'} 5B$, where $5B$ is $B$ with 5 times radius.

If $M(f)(x) > \lambda$, then we can find a ball $B_x$ centered at $x$ such that $\int_{B_x} |f| dy > \lambda |B_x|$. By the Vitali covering lemma, we can find a sequence of disjoint balls $B_j$ from $\{B_x : M(f)(x) > \lambda\}$ such that the union of $5B_j$ covers $\{x : M(f)(x) > \lambda\}$. It follows: $|\{x : M(f)(x) > \lambda\}| \le 5^d \sum_j |B_j| \le \frac{5^d}{\lambda} \int |f| dy = \frac{C_d}{\lambda} \|f\|_{L^1}$. $\square$

# 3 Qualitative Lebesgue differentiation theorem

First, let's recall the Lebesgue differentiation theorem we have learned in the class:

**Theorem 3.1** (**Qualitative Lebesgue differentiation theorem**)**.** If $f : [0,1] \to [0,1]$ is Lebesgue measurable, then for almost every $x \in [0,1]$ we have $f(x) = \lim_{r \to 0} \frac{1}{r} \int_x^{x+r} f(y) dy$.

Since we have proved the qualitative Lebesgue differentiation theorem in class, we omit its proof here. However, it is worth a mention that the version of Lebesgue differentiation theorem is qualitative in nature. It only asserts that $\frac{1}{r} \int_x^{x+r} f(y)\, dy$ eventually gets close to $f(x)$ for almost every $x$ by taking $r$ small enough, but does not give a quantitative bound for how small $r$ has to be. The following simple example can show why there is a problem. Let $A_n$ be $[0, 1/2^n] \cup [2/2^n, 3/2^n] \cup \ldots \cup [1 - 2/2^n, 1 - 1/2^n]$. Define $f_n$ as $\mathbb{1}_{A_n}$. From the qualitative Lebesgue differentiation theorem, $\frac{1}{r} \int_x^{x+r} f_n(y)\, dy$ eventually goes to $f_n(x)$, which is either 0 or 1, for almost every $x$. However, it is not hard to see if $|r| > 1/2^n$, $\frac{1}{r} \int_x^{x+r} f_n(y)\, dy$ will be neither 0 nor 1. Actually, when $|r| \gg 1/2^n$, $\frac{1}{r}\int_x^{x+r} f_n(y)\ dy$ is roughly 1/2. Intuitively, while each $A_n$ is certainly measurable, they are getting **less measurable** as $n$ increases.

# 4 Quantitative Lebesgue differentiation theorem

The simple example in the qualitative Lebesgue differentiation theorem motivates us to think about what **less measurable** means. The rate of convergence in the qualitative Lebesgue differentiation theorem heavily depends on how measurable the sets $A_n$ are. Hence, if we want to quantify the Lebesgue differentiation theorem, one of the natural ways is to develop the quantification of the notion of measurability and apply it to refine the Lebesgue differentiation theorem. There are so many ways to make such a quantification. Here we introduce some typical proposals:

**Definition 4.1** (**$(\varepsilon, n)$-measurability for a set**)**.** A set $A \subset [0, 1]$ is $(\varepsilon, n)$-*measurable* if there exists a set $B$ which is the union of dyadic intervals $[\frac{j}{2^n}, \frac{j+1}{2^n}]$ at scale $2^{-n}$, such that $A$ and $B$ only differ on a set of Lebesgue measure $\varepsilon$.

**Definition 4.2** (**$(\varepsilon, n)$-measurability for a function**)**.** A function $f : [0, 1] \to [0, 1]$ is $(\varepsilon, n)$-*measurable* if there exists a function $g$ which is constant on the dyadic intervals $[\frac{j}{2^n}, \frac{j+1}{2^n}]$ such that $\int_0^1 |f(x) - g(x)|dx \le \varepsilon$.

The next two theorems will connect the quantitative notion of $(\epsilon, n)$-measurability with the traditional qualitative notion of Lebesgue measurability.

**Theorem 4.1** (**Lebesgue approximation theorem for a set**)**.** Let $A \subset [0, 1]$ be measurable. Then for every $\varepsilon > 0$ there exists $n$ such that $A$ is $(\varepsilon, n)$-measurable.

**Proof.** In the following, any set considered is contained in $[0, 1]$.

(a) The claim is closed under countable unions:

Proof: Given a sequence of sets $\{A_i\}_{i=1}^{\infty}$ such that for every $i \in \mathbb{Z}^+$ and $\varepsilon > 0$, there exists $n$ such that $A_i$ is $(\varepsilon, n)$-measurable. Then we claim that for every $\varepsilon > 0$, there exists $n$ such that $A = \cup_{i=1}^{\infty} A_i$ is $(\varepsilon, n)$-measurable.

Given $\varepsilon > 0$. For each $i$, there exists $n_i$ such that $A_i$ is $(\varepsilon/2^{i+1}, n_i)$-measurable. Hence, for each $A_i$, there exists $B_i$ which is the union of dyadic intervals at scale $2^{-n_i}$ such that $A_i$ and $B_i$ only differ on a set of Lebesgue measure $\varepsilon/2^{i+1}$. Take $B = \cup_{i=1}^{\infty} B_i$. Then $A$ and $B$ differ on a set of Lebesgue measure $\sum_{i=1}^{\infty} \varepsilon/2^{i+1} = \varepsilon/2$. Since $| \cup_{i=1}^{n} B_i| \to |B| < \infty$, there exists a large $N$ such that $A$ and $\cup_{i=1}^{N} B_i$ differ on a set of Lebesgue measure $\varepsilon$. Take $n = max\{n_1, n_2, ..., n_N\}$. Then

$A$ is $(\varepsilon, n)$-measurable.

(b) The claim is closed under complement:

Clearly, if $A$ is $(\varepsilon, n)$-measurable, then $A^c$ is $(\varepsilon, n)$-measurable since $B^c$, where $B$ only differs from $A$ on a set of Lebesgue measure $\varepsilon$, only differs from $A^c$ on a set of Lebesgue measure $\varepsilon$.

(c) The claim is closed under countable intersections:

By writing countable intersections as the complement of countable unions and (a) and (b), one can quickly deduce (c).

Next, we prove that the theorem holds for some basic sets and use (a),(b),(c) to extend them to all Lebesgue measurable sets.

(1) $A$ is the finite union of dyadic intervals:

Choose sufficiently large $n$ such that all of the dyadic intervals that make up $A$ has the larger scale than $2^{-n}$. Then we can just take $B$ in $(\epsilon, n)$-measurability as $A$ and prove A is $(\varepsilon, n)$-measurable.

(2) $A$ is compact:

**Claim**: $A = \bigcap_{n=1}^{\infty} A^{(n)}$, where $A^{(n)}$ is the union of all the closed dyadic intervals at scale $2^{-n}$ that intersect $A$.

Proof. Clearly, $A \subseteq \bigcap_{n=1}^{\infty} A^{(n)}$. It remains to show that $\bigcap_{n=1}^{\infty} A^{(n)} \subseteq A$. Given $x \in \bigcap_{n=1}^{\infty} A^{(n)}$. Suppose $x \in A^c$. Since $A$ is closed, $A^c$ is open and thus there exists a small enough dyadic interval $D^{(k)}$ at scale $2^{-k}$ such that $x \in D^{(k)} \subseteq A^c$. Hence, $x \notin A^{(k)}$, which contradicts that $x \in \bigcap_{n=1}^{\infty} A^{(n)}$. Thus, $x \in A$ and $A = \bigcap_{n=1}^{\infty} A^{(n)}$. □

Then by (1) and (c), we prove $A$ is $(\varepsilon, n)$-measurable.

(3) The theorem holds for all Borel-measurable sets:

Since Borel-measurable sets are generated by open sets under countable unions, countable intersections, and complement, one can use (a),(b),(c), and (2) to show that the theorem holds for all Borel-measurable sets.

(4) The theorem holds for all Lebesgue-measurable sets:

Clearly, the theorem holds for zero sets. Hence, with (3) it holds for all Lebesgue-measurable sets. □

**Theorem 4.2** (**Lebesgue approximation theorem for a function**). Let $f : [0,1] \to [0,1]$ be measurable. Then for every $\varepsilon > 0$ there exists $n$ such that $f$ is $(\varepsilon, n)$-measurable.

**Proof.** Given $\varepsilon > 0$. Define $B_k$ by $B_k = f^{pre}([k\varepsilon/2, (k+1)\varepsilon/2))$ [1]. Since $f$ is measurable, each $B_k$ is measurable. Hence, by Theorem 4.1, for each $B_k$, there exists a finite union of

[1] $\mathrm{B}_{2/\varepsilon-1} = f^{pre}([1-\varepsilon/2, 1])$

dyadic intervals $D_k$ at scale $2^{-n_k}$ such that $B_k$ and $D_k$ only differ on a set of Lebesgue measure $\varepsilon^2/4$. Write $f(x)$ as $f(x) = \sum_{k=0}^{2/\varepsilon-1} f(x)\mathbb{1}_{B_k}$ and define $g(x)$ as $g(x) = \sum_{k=0}^{2/\varepsilon-1} k\frac{\varepsilon}{2}\mathbb{1}_{D_k}$. Clearly, $g$ is a function which is constant on dyadic intervals at scale $2^{-max\{n_0,n_1,...n_{2/\varepsilon-1}\}}$. Take $n = max\{n_0, n_1, ...n_{2/\varepsilon-1}\}$. Next, we show that $f$ is $(\varepsilon, n)$-measurable.

$$
\begin{aligned}
|f(x) - g(x)| &\leq \sum_{k=0}^{2/\varepsilon-1} |f(x)\mathbb{1}_{B_k} - k\frac{\varepsilon}{2}\mathbb{1}_{D_k}| \\
&= \sum_{k=0}^{2/\varepsilon-1} |f(x)\mathbb{1}_{B_k\cap D_k} - k\frac{\varepsilon}{2}\mathbb{1}_{B_k\cap D_k} + f(x)\mathbb{1}_{B_k-D_k} - k\frac{\varepsilon}{2}\mathbb{1}_{D_k-B_k}| \\
&\leq \sum_{k=0}^{2/\varepsilon-1} (\frac{\varepsilon}{2}\mathbb{1}_{B_k\cap D_k} + \mathbb{1}_{B_k-D_k} + \mathbb{1}_{D_k-B_k}).
\end{aligned}
$$

Since $B_k$ are disjoint, $B_k \cap D_k$ are disjoint in $[0, 1]$ . Then we have $\int_0^1 |f(x) - g(x)|dx \leq \int_0^1 \sum_{k=0}^{2/\varepsilon-1}(\frac{\varepsilon}{2}\mathbb{1}_{B_k\cap D_k} + \mathbb{1}_{B_k-D_k} + \mathbb{1}_{D_k-B_k})dx \leq \frac{\varepsilon}{2} + \frac{2}{\varepsilon}\frac{\varepsilon^2}{4} = \varepsilon$. □

**Remark.** Actually, we can first show **Theorem 4.2** and then apply it to prove **Theorem 4.1**. That is, these two theorems are equivalent.

Using the concept of $(\varepsilon, n)$-measurability, we have quantified the notion of measurability. We are now ready to see how these concepts we have developed work in helping quantify the Lebesgue differentiation theorem.

**Theorem 4.3** (**Quantitative Lebesgue differentiation theorem**)**.** Let $f : [0, 1] \to [0, 1]$ be $(\varepsilon, n)$-measurable. Then for all $x$ in $[0, 1]$ outside of a set of measure $O(\sqrt{\epsilon})$, we have $\frac{1}{r}\int_x^{x+r} f(y)\ dy = f(x) + O(\sqrt{\epsilon})$ for all $0 < r < \sqrt{\epsilon}2^{-n}$.

**Proof.** Let $g$ be the function such that $\int_0^1 |f(x) - g(x)|dx \leq \varepsilon$ in $(\varepsilon, n)$-measurability. Define $f_r(x)$ as $f_r(x) = \frac{1}{r}\int_x^{x+r} f(t)dt$ and $\Omega(f)(x)$ as $\Omega(f)(x) = \sup_{0<r<r_0} f_r(x) - \inf_{0<r<r_0} f_r(x)$, where $r_0$ will be determined later. Let $A$ be the set of points in $[0, 1]$ with the property: the distance between the point and the endpoint of the dyadic interval at scale $2^n$ in which it lies is larger than $r_0$. Then $|A| = 1-2^n r_0$ and one can easily see that for $x \in A$, $\sup_{0<r<r_0} g_r(x) = \inf_{0<r<r_0} g_r(x) = g(x)$ since $g$ is constant on $[x, x + r_0]$. Hence, $\Omega(g)(x) = 0$ for $x \in A$. Set $h = f - h$ and we have $\|h\|_{L^1} \leq \varepsilon$. For $x \in A$, $\Omega(f)(x) = \Omega(g + h)(x) \leq \Omega(g)(x) + \Omega(h)(x) = \Omega(h)(x)$. Clearly, $\Omega(h)(x) = \sup_{0<r<r_0} h_r(x) - \inf_{0<r<r_0} h_r(x) \leq 2M(h)(x)$, where $M$ is the Hardy-Littlewood maximal operator. Hence, $\{x \in [0, 1] : |\Omega(h)(x)| > \lambda\} \subseteq |\{x \in [0, 1] : |M(h)(x)| > \lambda/2\}$. Then by the Hardy-Littlewood maximal inequality, we have $|\{x \in [0, 1] : |\Omega(h)(x)| > \lambda\}| \leq |\{x \in [0, 1] : |M(h)(x)| > \lambda/2\}| \leq \frac{C}{\lambda}\|h\|_1 < \frac{C}{\lambda}\varepsilon$. Take $\lambda = \sqrt{\varepsilon}$. It shows that $\Omega(h)(x) \leq \sqrt{\varepsilon}$

outside a set $B$ of measure less than $C\sqrt{\varepsilon}$. Now, it is clear that we should choose $r_0 = 2^{-n}\sqrt{\varepsilon}$ so that $|A| = 1-\sqrt{\varepsilon}$. Hence, $|A-B| \geq 1-(C+1)\sqrt{\varepsilon}$ and $\Omega(f)(x) \leq \sqrt{\varepsilon}$ for $x \in A-B$. By the qualitative Lebesgue differentiation theorem, we know the limit of $f_r(x)$ is $f(x)$ for almost every $x$. Then, $\inf_{0<r'<2^{-n}\sqrt{\varepsilon}} f_{r'}(x) \leq f(x), f_r(x) \leq \sup_{0<r'<2^{-n}\sqrt{\varepsilon}} f_{r'}(x)$ for almost every $x$ and for all $r < 2^{-n}\sqrt{\varepsilon}$. Hence, for almost every $x \in A-B$, $|f_r(x) - f(x)| \leq \Omega(f)(x) \leq \sqrt{\varepsilon}$ for all $r < 2^{-n}\sqrt{\varepsilon}$. $\square$

# 5 Lebesgue regularity lemma

In $(\varepsilon, n)$-measurability, we try to approximate a set or a function by a finite union of intervals, or a piecewise constant function. However, one can adopt a different philosophy. One clue is to look at the simple examples $A_n$ and $f_n$ discussed in the section of Qualitative Lebesgue differentiation theorem. Observe that if one averages $f_n$ on any reasonable sized interval $J$, one gets something very close to the global average of $f_n = 1/2$. In other words, the integral of $f_n$ on an interval $J$ is close to the global average of $f_n$ times |J|. This motivates the following definition.

**Definition 5.1** (**$\varepsilon$-regularity**)**.** A function $f : [0,1] \to [0,1]$ is said to be $\varepsilon$-regular on a dyadic interval $I$ if we have $|\int_J f(x)dx - |J| ⨍_I f(x)\,dx| \leq \varepsilon|I|$ for all dyadic subintervals $J \subseteq I$.

Next, we will connect the quantitative notion of $\varepsilon$-regularity with the traditional qualitative notion of Lebesgue measurability.

**Lemma 5.1** (**Lebesgue regularity lemma**)**.** If $\varepsilon > 0$ and $f : [0,1] \to [0,1]$ is measurable, then there exists an positive integer $n < \frac{1}{\varepsilon^3}\log_2\frac{1}{\varepsilon} + 1$, such that $f$ is $\varepsilon$-regular on all but at most $\varepsilon 2^n$ of the $2^n$ dyadic intervals of length $2^{-n}$.

**Proof.** $\forall n \in \mathbb{N}$, let $f^{(n)} : [0,1] \to [0,1]$ be the conditional expectation of $f$ to the dyadic intervals of length $2^{-n}$, that is, $f^{(n)} = \sum_{i=1}^{2^n} c_{i,n}\mathbb{1}_{I_i^n}$, where $I_i^n = [(i-1)2^{-n}, i2^{-n})$ and $c_{i,n} = ⨍_{I_i^n} f(x)\,dx$.

**Claim A**: $E_n := \int_0^1 |f^{(n)}(x)|^2 dx$ is an increasing sequence in $n$ and bounded between 0 and 1. Proof. It is clear that $E_n$ is bounded by 0 and 1 since $f^{(n)}$ is between 0 and 1. We check that

$E_{n+1} - E_n \geq 0$ and thus $E_n$ are increasing.

$$E_{n+1} - E_n = \int_0^1 [(\sum_{i=1}^{2^{n+1}} c_{i,n+1} \mathbb{1}_{I_i^{n+1}}(x))^2 - (\sum_{i=1}^{2^n} c_{i,n} \mathbb{1}_{I_i^n}(x))^2] dx$$

$$= \int_0^1 [\sum_{i=1}^{2^{n+1}} c_{i,n+1}^2 \mathbb{1}_{I_i^{n+1}}(x) - \sum_{i=1}^{2^n} c_{i,n}^2 \mathbb{1}_{I_i^n}(x)] dx$$

$$= 2^{-n-1} \sum_{i=1}^{2^n} [c_{2i-1,n+1}^2 + c_{2i,n+1}^2 - 2c_{i,n}^2].$$

Note that $c_{2i-1,n+1} + c_{2i,n+1} = 2c_{i,n}$. Hence, $c_{2i-1,n+1}^2 + c_{2i,n+1}^2 - 2c_{i,n}^2 \geq 0$ and thus $E_{n+1} - E_n \geq 0$.$\square$

$E_n$ is an increasing sequence between 0 and 1. Then we can apply the finite convergence principle with $F(x) = \log_2 \frac{1}{\varepsilon}$ and $M = \frac{1}{\varepsilon^3} \log_2 \frac{1}{\varepsilon} + 1$ to $E_n$. Thus, we can find $n < M$ such that $E_{n+\log_2 \frac{1}{\varepsilon}} - E_n \leq \varepsilon^3$.

After some tedious calculation, we have

$$\int_0^1 |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|^2 dx = 2^{-n} \varepsilon \sum_{i=1}^{2^n} \sum_{j=1}^{\frac{1}{\varepsilon}} |c_{\frac{i-1}{\varepsilon}+j, n+\log_2 \frac{1}{\varepsilon}} - c_{i,n}|^2$$

Note that

$$\sum_{i=1}^{2^n} \sum_{j=1}^{\frac{1}{\varepsilon}} c_{\frac{i-1}{\varepsilon}+j, n+\log_2 \frac{1}{\varepsilon}} c_{i,n} = \sum_{i=1}^{2^n} \frac{1}{\varepsilon} c_{i,n}^2 = \sum_{i=1}^{2^n} \sum_{j=1}^{\frac{1}{\varepsilon}} c_{i,n}^2$$

Thus, we have

$$\int_0^1 |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|^2 dx = 2^{-n} \varepsilon \sum_{i=1}^{2^n} \sum_{j=1}^{\frac{1}{\varepsilon}} (c_{\frac{i-1}{\varepsilon}+j, n+\log_2 \frac{1}{\varepsilon}}^2 - c_{i,n}^2) = E_{n+\log_2 \frac{1}{\varepsilon}} - E_n \leq \varepsilon^3$$

Next, let $X : \{I_i^n\}_{i=1}^{2^n} \to \mathbb{R}$ be the discrete random variable defined by

$$X(I_i^n) = \int_{I_i^n} |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|^2 dx$$

and set $P\{X(I_i^n)\} = \frac{1}{2^n}$.

By Markov inequality, $P(X \geq \varepsilon^2 2^{-n}) \leq \frac{E[X]}{\varepsilon^2 2^{-n}}$.

$P(X \geq \varepsilon^2 2^{-n}) = \frac{1}{2^n} \times$ the number of $I_i^n$ on which $X \geq \varepsilon^2 2^{-n}$ and $E[X] = \frac{\int_0^1 |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|^2 dx}{2^n}$.

Hence, $P(X \geq \varepsilon^2 2^{-n}) \leq \varepsilon$, which means that the number of $I_i^n$ on which $X \geq \varepsilon^2 2^{-n}$ is less than or equal to $\varepsilon 2^n$. It implies that $\int_{I_i^n} |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|^2 dx \leq \varepsilon^2 2^{-n}$ for all but at most $\varepsilon 2^n$ of the dyadic intervals $I_i^n$. On these dyadic intervals, by Cauchy-Schwarz inequality,

$$(\int_{I_i^n} |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)| \times 1 dx)^2 \leq \int_{I_i^n} |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|^2 dx \times 2^{-n} \leq (\varepsilon 2^{-n})^2$$

Thus, we have $\int_{I_i^n} |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)| dx \leq \varepsilon 2^{-n}$ on all but at most $\varepsilon 2^n$ of the dyadic intervals $I_i^n$.

**Claim B**: $f$ is $\varepsilon$-regular on a dyadic interval $I_i^n$ with $\int_{I_i^n} |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|dx \leq \varepsilon 2^{-n}$

Proof. Given a dyadic sub-interval $J \subseteq I_i^n$.

(1) $|J| \leq \frac{1}{2}\varepsilon 2^{-n}$:

$|\int_J f(x)dx - |J| \fint_{I_i^n} f(x)\, dx| \leq 2|J| \leq \varepsilon 2^{-n} = \varepsilon |I_i^n|$

(2) $\varepsilon 2^{-n} \leq |J| \leq 2^{-n}$:

$|J| = 2^l \varepsilon |I_i^n|$ for some $l$ such that $1 \leq 2^l \leq \frac{1}{\varepsilon}$. That implies $J$ is the union of $2^l$ consecutive dyadic sub-intervals of length $\varepsilon 2^{-n}$, say $J = \cup_{k=1}^{2^l} J_k$.

$$\int_J |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|dx = \sum_{k=1}^{2^l} \varepsilon |I_i^n| |\fint_{J_k} f(x)\, dx - \fint_{I_i^n} f(x)\, dx| \leq \varepsilon 2^{-n}$$

$$\varepsilon |I_i^n| |\sum_{k=1}^{2^l} (\fint_{J_k} f(x)\, dx - \fint_{I_i^n} f(x)\, dx)| = |\sum_{k=1}^{2^l} \varepsilon |I_i^n| \fint_{J_k} f(x)\, dx - 2^l \varepsilon |I_i^n| \fint_{I_i^n} f(x)\, dx)| \leq \varepsilon 2^{-n}$$

$$\begin{aligned} |\sum_{k=1}^{2^l} \varepsilon |I_i^n| \fint_{J_k} f(x)\, dx - 2^l \varepsilon |I_i^n| \fint_{I_i^n} f(x)\, dx| &= |\sum_{k=1}^{2^l} \int_{J_k} f(x)dx - |J| \fint_{I_i^n} f(x)dx| \\ &= |\int_J f(x)dx - |J| \fint_{I_i^n} f(x)dx| \\ &\leq \varepsilon 2^{-n} = \varepsilon |I_i^n|. \end{aligned}$$

Hence, from (1) and (2), we prove the claim. □

Since $\int_{I_i^n} |f^{(n+\log_2 \frac{1}{\varepsilon})}(x) - f^{(n)}(x)|dx \leq \varepsilon 2^{-n}$ on all but at most $\varepsilon 2^n$ of the dyadic intervals $I_i^n$ and by Claim B, we prove that $f$ is $\varepsilon$-regular on all but at most $\varepsilon 2^n$ of the $2^n$ dyadic intervals of length $2^{-n}$. □

Actually, one can get a stronger result at the cost of worsening the bound on $n$, It comes as no surprise since we haven't fully exploited the finite convergence principle when proving the regularity lemma. To motivate the stronger version, first observe that if a function $f$ is $\varepsilon$-regular on an interval $I$, then on that interval we have a decomposition $f = c + h$ on $I$ where $c = \frac{1}{|I|}\int_I f(y)dy$ is the mean of $f$ on $I$, and $h$ has small averages in the sense that $|\int_J h(y)\ dy| \leq \epsilon |I|$ for all dyadic sub-intervals $J$ of $I$. We can even do better than this, in which the notion of strong $(\varepsilon, m)$-regularity will be introduced.

**Definition 5.2** (**Strong** $(\varepsilon, m)$**-regularity**)**.** A function $f : [0,1] \to [0,1]$ is said to be *strongly* $(\varepsilon, m) - regular$ on a dyadic interval $I$ if there exists a decomposition $f = c+e+h$ on $I$, where $c = \frac{1}{|I|}\int_I f(y)dy$ is the mean of $f$ on $I$, $e$ is small in the sense that $\frac{1}{|I|}\int_I |e(y|dy \leq \varepsilon$, and $h$

has vanishing averages in the sense that $\int_J h(y)dy = 0$ for all dyadic sub-intervals $J \subseteq I$ with $|J| \geq 2^{-m}|I|$.

**Corollary 5.1.** If $2^{-m} \leq \varepsilon$, then strong $(\varepsilon, m)$-regularity implies $\varepsilon$-regularity.

**Proof.** Given a dyadic sub-interval $J \subseteq I$ and $f$ is $(\varepsilon, m)$-regular with $2^{-m} \leq \varepsilon$.
(1) $|J| \geq 2^{-m}|I|$:
$|\int_J (f(x) - c)dx| = |\int_J (e(x) + h(x))dx| = |\int_J e(x)dx| \leq \int_I |e(x)|dx \leq \varepsilon|I|$
(2) $|J| < 2^{-m}|I|$:
First notice that $|J| \leq 2^{-m+1}|I| \leq \frac{1}{2}\varepsilon|I|$.
$|\int_J (f(x) - c)dx| \leq \int_J |f(x)|dx + \int_J |c|dx \leq 2|J| \leq \varepsilon|I|$ □

The parameter $m$ here suggests that when the scale is not too fine($\geq 2^{-m}|I|$), strong $(\epsilon, m)$-regularity offers much better control on the fluctuation of $f$ at finer scales.
With the notion of strong $(\epsilon, m)$-regularity, we then can get a stronger version of Lebesgue regularity lemma.

**Lemma 5.2 (Strong Lebesgue regularity lemma).** If $\varepsilon > 0$, $F : \mathbb{N} \to \mathbb{N}$, and $f : [0, 1] \to [0, 1]$ is measurable, then there exists a positive integer $n = O_{\varepsilon,F}(1)$ such that $f$ is $(\varepsilon, F(n))$-regular on all but at most $\varepsilon 2^n$ of the $2^n$ dyadic intervals of length $2^{-n}$.

**Proof.** $\forall n \in \mathbb{N}$, let $f^{(n)} : [0, 1] \to [0, 1]$ be the conditional expectation of $f$ to the dyadic intervals of length $2^{-n}$, that is, $f^{(n)} = \sum_{i=1}^{2^n} c_{i,n} \mathbb{1}_{I_i^n}$, where $I_i^n = [(i-1)2^{-n}, i2^{-n})$ and $c_{i,n} = ⨍_{I_i^n} f(x)\, dx$.
**Claim A**: $E_n := \int_0^1 |f^{(n)}(x)|^2 dx$ is an increasing sequence in $n$ and bounded between 0 and 1.
Proof. It is clear that $E_n$ is bounded by 0 and 1 since $f^{(n)}$ is between 0 and 1. We check that $E_{n+1} - E_n \geq 0$ and thus $E_n$ are increasing.

$$\begin{aligned}
E_{n+1} - E_n &= \int_0^1 [(\sum_{i=1}^{2^{n+1}} c_{i,n+1} \mathbb{1}_{I_i^{n+1}}(x))^2 - (\sum_{i=1}^{2^n} c_{i,n} \mathbb{1}_{I_i^n}(x))^2] dx \\
&= \int_0^1 [\sum_{i=1}^{2^{n+1}} c_{i,n+1}^2 \mathbb{1}_{I_i^{n+1}}(x) - \sum_{i=1}^{2^n} c_{i,n}^2 \mathbb{1}_{I_i^n}(x)] dx \\
&= 2^{-n-1} \sum_{i=1}^{2^n} [c_{2i-1,n+1}^2 + c_{2i,n+1}^2 - 2c_{i,n}^2].
\end{aligned}$$

Note that $c_{2i-1,n+1} + c_{2i,n+1} = 2c_{i,n}$. Hence, $c_{2i-1,n+1}^2 + c_{2i,n+1}^2 - 2c_{i,n}^2 \geq 0$ and thus $E_{n+1} - E_n \geq 0$.□
$E_n$ is an increasing sequence between 0 and 1. Then we can apply the finite convergence

principle to $E_n$. Thus, we can find $n = O_{\varepsilon,F}(1)$ such that $E_{n+F(n)} - E_n \leq \varepsilon^3$.

After some tedious calculation, we have

$$\int_0^1 |f^{(n+F(n))}(x) - f^{(n)}(x)|^2 dx = 2^{-n-F(n)} \sum_{i=1}^{2^n} \sum_{j=1}^{2^{F(n)}} |c_{(i-1)2^{F(n)}+j,n+F(n)} - c_{i,n}|^2$$

Note that

$$\sum_{i=1}^{2^n} \sum_{j=1}^{2^{F(n)}} c_{(i-1)2^{F(n)}+j,n+F(n)} c_{i,n} = \sum_{i=1}^{2^n} 2^{F(n)} c_{i,n}^2 = \sum_{i=1}^{2^n} \sum_{j=1}^{2^{F(n)}} c_{i,n}^2$$

Thus, we have

$$\int_0^1 |f^{(n+F(n))}(x) - f^{(n)}(x)|^2 dx = 2^{-n-F(n)} \sum_{i=1}^{2^n} \sum_{j=1}^{2^{F(n)}} (c_{(i-1)2^{F(n)}+j,n+F(n)}^2 - c_{i,n}^2) = E_{n+F(n)} - E_n \leq \varepsilon^3$$

Next, let $X : \{I_i^n\}_{i=1}^{2^n} \to \mathbb{R}$ be the discrete random variable defined by

$$X(I_i^n) = \int_{I_i^n} |f^{(n+F(n))}(x) - f^{(n)}(x)|^2 dx$$

and set $P\{X(I_i^n)\} = \frac{1}{2^n}$.

By Markov inequality, $P(X \geq \varepsilon^2 2^{-n}) \leq \frac{E[X]}{\varepsilon^2 2^{-n}}$.

$P(X \geq \varepsilon^2 2^{-n}) = \frac{1}{2^n}$ times the number of $I_i^n$ on which $X \geq \varepsilon^2 2^{-n}$ and $E[X] = \frac{\int_0^1 |f^{(n+F(n))}(x) - f^{(n)}(x)|^2 dx}{2^n}$.

Hence, $P(X \geq \varepsilon^2 2^{-n}) \leq \varepsilon$, which means that the number of $I_i^n$ on which $X \geq \varepsilon^2 2^{-n}$ is less than or equal to $\varepsilon 2^n$. It implies that $\int_{I_i^n} |f^{(n+F(n))}(x) - f^{(n)}(x)|^2 dx \leq \varepsilon^2 2^{-n}$ for all but at most $\varepsilon 2^n$ of the dyadic intervals $I_i^n$. On these dyadic intervals, by Cauchy-Schwarz inequality,

$$(\int_{I_i^n} |f^{(n+F(n))}(x) - f^{(n)}(x)| \times 1 dx)^2 \leq \int_{I_i^n} |f^{(n+F(n))}(x) - f^{(n)}(x)|^2 dx^{-n} \leq (\varepsilon 2^{-n})^2$$

Thus, we have $\int_{I_i^n} |f^{(n+F(n))}(x) - f^{(n)}(x)| dx \leq \varepsilon |I_i^n|$ on all but at most $\varepsilon 2^n$ of the dyadic intervals $I_i^n$. □

**Claim B**: $f$ is $(\varepsilon, F(n))$-regular on a dyadic interval $I_i^n$ with $\int_{I_i^n} |f^{(n+F(n))}(x) - f^{(n)}(x)| dx \leq \varepsilon |I_i^n|$

Proof. Since $f(x) = f^{(n)}(x) + [f^{(n+F(n))}(x) - f^{(n)}(x)] + [f(x) - f^{(n+F(n))}(x)]$ on $I_i^n$, we take $c = \frac{1}{|I|} \int_{I_i^n} f(y) dy, e(x) = f^{(n+F(n))}(x) - f^{(n)}(x)$, and $h(x) = f(x) - f^{(n+F(n))}(x)$.

By assumption, $\int_{I_i^n} |e(x)| dx \leq \varepsilon |I_i^n|$. It remains to show that $\int_J h(y) dy = 0$ for all dyadic sub-intervals $J \subseteq I_i^n$ with $|J| \geq 2^{-F(n)} |I_i^n|$.

Given a dyadic sub-interval $J \subseteq I_i^n$ with $|J| \geq 2^{-F(n)} |I_i^n|$.

Since $|J| \geq 2^{-F(n)} |I_i^n|, \int_J f^{(n+F(n))}(y) dy = \int_J f(y) dy$.

Hence, $\int_J h(y) dy = \int_J f(y) - f^{(n+F(n))}(y) dy = 0$. Thus, we prove that $f$ is $(\varepsilon, F(n))$-regular on $I_i^n$. □

Since $\int_{I_i^n} |f^{(n+F(n))}(x) - f^{(n)}(x)| dx \leq \varepsilon 2^{-n}$ on all but at most $\varepsilon 2^n$ of the dyadic intervals $I_i^n$ and by Claim B, we prove that $f$ is $(\varepsilon, F(n))$-regular on all but at most $\varepsilon 2^n$ of the $2^n$ dyadic intervals of length $2^{-n}$. □

Finally, we want to mention two applications of strong Lebesgue regularity lemma.
The first one is the version of quantitative Lebesgue differentiation theorem in terms of strong Lebesgue regularity lemma.

**Theorem 5.1** (**Quantitative Lebesgue differentiation theorem**). If $\varepsilon > 0$, $F : \mathbb{N} \to \mathbb{N}$, and $f : [0,1] \to [0,1]$ is measurable, then there exists a positive integer $n = O_{\varepsilon,F}(1)$ such that $\forall x \in [0,1]$ outside a set of measure $O(\varepsilon)$ we have the Cauchy sequence property: $\mid \frac{1}{r}\int_x^{x+r} f(y)dy - \frac{1}{s}\int_x^{x+s} f(y)dy \mid \leq \varepsilon$ for all $2^{-n-F(n)} < r, s < 2^{-n}$.

**Remark.** In Tao's blog[2], he gives the hint that this theorem can be deduced by combining the strong regularity lemma with the Hardy-Littlewood maximal inequality.

The other one is the one-dimensional case of the Rademacher differentiation theorem.

**Theorem 5.2** (**The Rademacher differentiation theorem for 1-dimensional case**). A Lipschitz continuous function from [0,1] to $\mathbb{R}$ is almost everywhere differentiable.

**Remark.** We have proved this theorem in class. Anyone who is interested in the proof in terms of strong Lebesgue regularity lemma can find it in Tao's blog[2] and read his creative proof.

# References


[1] Terence Tao: Soft analysis, hard analysis, and the finite convergence principle, `https://terrytao.wordpress.com/2007/05/23/soft-analysis-hard-analysis-and-the-finite-convergence-principle/`

[2] Terence Tao: The Lebesgue differentiation theorem and the Szemeredi regularity lemma, `https:// terrytao.wordpress.com/ 2007/06/18/ the- lebesgue- differentiation-theorem-and-the-szemeredi-regularity-lemma/`

[3] Quantitative Lebesgue Differentiation Theorem proof, `https:// math.stackexchange.com/ questions/ 1926553/ quantitative- lebesgue-differentiation-theorem-proof`

[4] Wikipedia: Hardy—Littlewood maximal function,
https://en.wikipedia.org/wiki/Hardy-Littlewood_maximal_function