Probabilistic Methods in Analysis
Kent State University

Swapnamoy Kader
Summer 2026

This entire document is based on the coursework during the semester. This document is strictly meant to help students learn, recap and prepare for Probabilistic Methods in Analysis.

Contents

1 Discrete Probability Spaces, Random Variables, and Expectation

1.1 Why Probability?

Probability gives a mathematical language for situations where there are several possible outcomes and we do not know in advance which one occurs. In this course, probability is not only an object of study; it becomes a tool for proving existence theorems in analysis and combinatorics.

Typical examples:

- Tossing a coin.
- Rolling a die.
- Choosing a random subset of $\{1, \dots, N\}$.
- Choosing random signs ± 1 in a polynomial.
- Randomly coloring columns of a matrix.

1.2 Sample Space

Definition 1.1. A *sample space* is the collection of all possible outcomes. It is denoted by Ω.
In this course, the sample spaces are usually finite or countable:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}.$$

Example 1.2 (Coin toss)**.** For a coin toss,

$$\Omega = \{H, T\}.$$

Example 1.3 (Die roll)**.** For a die,

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

1.3 Probability Measure

Definition 1.4. A probability measure assigns a number $p_i \geq 0$ to each outcome ω_i, with

$$\sum_i p_i = 1.$$

We write

$$\mathbb{P}(\omega_i) = p_i.$$

The condition $\sum_i p_i = 1$ means that one of the outcomes must occur.

1.4 Events

Definition 1.5. An *event* is a subset of the sample space:

$$E \subset \Omega.$$

If $E \subset \Omega$, then

$$\mathbb{P}(E) = \sum_{\omega_i \in E} p_i.$$

Example 1.6. For a fair die, let

$$E = \{2, 4, 6\}.$$

Then E is the event that the die shows an even number, and

$$\mathbb{P}(E) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

1.5 Probability Space

Definition 1.7. The pair

$$(\Omega, \mathbb{P})$$

is called a probability space.

1.6 Product Probability Spaces and Independence

Suppose

$$(\Omega_1, \mathbb{P}_1), \qquad (\Omega_2, \mathbb{P}_2)$$

are probability spaces. The product sample space is

$$\Omega = \Omega_1 \times \Omega_2.$$

Its elements are pairs (ω_1, ω_2). Define

$$\mathbb{P}((\omega_1, \omega_2)) = \mathbb{P}_1(\omega_1)\mathbb{P}_2(\omega_2).$$

This models independent experiments.

Example 1.8 (Two fair coins)**.**

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}.$$

Each outcome has probability $1/4$.

If $E_1 \subset \Omega_1$ and $E_2 \subset \Omega_2$, then

$$\mathbb{P}(E_1 \times E_2) = \mathbb{P}_1(E_1)\mathbb{P}_2(E_2).$$

This is the basic formula for independence.

1.7 Random Variables

Definition 1.9. A random variable is a function

$$X : \Omega \to \mathbb{R}.$$

For each outcome ω, the value $X(\omega)$ is a real number.

Example 1.10. For a coin toss, define

$$X(H) = 1, \qquad X(T) = 0.$$

Then X records whether the toss was heads.

1.8 Indicator Functions

Definition 1.11. For an event $E \subset \Omega$, the indicator random variable $\mathbf{1}_E$ is

$$\mathbf{1}_E(\omega) = \begin{cases} 1, & \omega \in E, \\ 0, & \omega \notin E. \end{cases}$$

Indicator functions convert events into random variables.

1.9 Expectation

Definition 1.12. The expectation of a random variable $X : \Omega \to \mathbb{R}$ is

$$\mathbb{E}(X) = \sum_i X(\omega_i)p_i.$$

Example 1.13 (Fair die)**.** Let $X(i) = i$. Then

$$\mathbb{E}(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5.$$

1.10 Expectation of an Indicator

If $X = \mathbb{1}_E$, then

$$\mathbb{E}(\mathbb{1}_E) = \sum_{\omega \in E} \mathbb{P}(\omega) = \mathbb{P}(E).$$

Therefore

$$\boxed{\mathbb{E}(\mathbb{1}_E) = \mathbb{P}(E).}$$

This identity is used constantly.

1.11 Linearity of Expectation

For random variables X, Y and scalars a, b,

$$\boxed{\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).}$$

Proof. By definition,

$$\mathbb{E}(aX + bY) = \sum_i (aX(\omega_i) + bY(\omega_i))p_i.$$

Distributing the sum gives

$$= a \sum_i X(\omega_i)p_i + b \sum_i Y(\omega_i)p_i = a\ \mathbb{E}(X) + b\ \mathbb{E}(Y).$$

□

Key point. Linearity of expectation does not require independence. Independence becomes important when multiplying random variables or computing variance of sums.

2 Variance, Independence, Markov's Inequality, and the Weak Law

2.1 Motivation

Expectation gives the average value, but it does not measure how much a random variable fluctuates.

For example, suppose

$$\mathbb{P}(X = 99) = \mathbb{P}(X = 101) = \frac{1}{2},$$

and

$$\mathbb{P}(Y = 0) = \mathbb{P}(Y = 200) = \frac{1}{2}.$$

Both have expectation 100, but Y fluctuates much more.

2.2 Variance

Definition 2.1. The variance of X is

$$\boxed{\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}(X))^2\big].}$$

Variance measures the typical squared distance from the mean.

Expanding,

$$(X - \mathbb{E}(X))^2 = X^2 - 2X\mathbb{E}(X) + (\mathbb{E}(X))^2.$$

Taking expectations gives

$$\boxed{\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.}$$

Since variance is an expectation of a square,

$$\mathrm{Var}(X) \geq 0.$$

Also,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \leq \mathbb{E}(X^2).$$

2.3 Independence of Random Variables

Definition 2.2. Random variables X and Y are independent if

$$\mathbb{P}(X = a, Y = b) = \mathbb{P}(X = a)\mathbb{P}(Y = b)$$

for all possible values a, b.

Proposition 2.3. *If X and Y are independent, then*

$$\boxed{\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y).}$$

Proof. Using independence,

$$\mathbb{E}(XY) = \sum_{a,b} ab\mathbb{P}(X = a, Y = b) = \sum_{a,b} ab\mathbb{P}(X = a)\mathbb{P}(Y = b).$$

Separating the sums,

$$\mathbb{E}(XY) = \left(\sum_{a} a\mathbb{P}(X = a)\right)\left(\sum_{b} b\mathbb{P}(Y = b)\right) = \mathbb{E}(X)\,\mathbb{E}(Y).$$

□

2.4 Variance of a Sum

If X and Y are independent, then

$$\boxed{\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).}$$

More generally, if $X_1, \ldots, X_n$ are independent, then

$$\boxed{\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k).}$$

2.5 Markov's Inequality

Theorem 2.4 (Markov)**.** *If $X \geq 0$ and $t > 0$, then*

$$\boxed{\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.}$$

Proof. Let $A = \{X \geq t\}$. Since $X \geq t$ on A,

$$\mathbb{E}(X) \geq \sum_{\omega \in A} X(\omega)\mathbb{P}(\omega) \geq \sum_{\omega \in A} t\mathbb{P}(\omega) = t\mathbb{P}(A).$$

Therefore $\mathbb{P}(A) \leq \mathbb{E}(X)/t$. □

2.6 Chebyshev's Inequality

Apply Markov to the nonnegative variable $(X - \mathbb{E}(X))^2$. Then

$$\mathbb{P}(|X - \mathbb{E}(X)| > a) = \mathbb{P}((X - \mathbb{E}(X))^2 > a^2) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

Thus

$$\boxed{\mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{\mathrm{Var}(X)}{a^2}.}$$

2.7 Weak Law of Large Numbers

Suppose $X_1, X_2, \ldots$ are independent, $\mathbb{E}(X_k) = \mu$, and $\mathrm{Var}(X_k) \leq C$. Define

$$A_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Then

$$\boxed{\mathbb{P}(|A_n - \mu| > \varepsilon) \to 0 \quad \text{as } n \to \infty.}$$

Proof. First,

$$\mathbb{E}(A_n) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(X_k) = \mu.$$

Second, by independence,

$$\mathrm{Var}(A_n) = \frac{1}{n^2} \sum_{k=1}^{n} \mathrm{Var}(X_k) \leq \frac{nC}{n^2} = \frac{C}{n}.$$

By Chebyshev,

$$\mathbb{P}(|A_n - \mu| > \varepsilon) \leq \frac{\mathrm{Var}(A_n)}{\varepsilon^2} \leq \frac{C}{n\varepsilon^2} \to 0.$$

□

Key point. The Weak Law says that averages of independent variables concentrate near their expected value. This is the first appearance of concentration in the course.

3 Exponential Moments and Concentration Inequalities

3.1 Why Chebyshev Is Not Enough

Chebyshev gives polynomial decay. For independent sums, the true tail is often exponentially small. This stronger decay is crucial for random subsets, random sign polynomials, and discrepancy.

Let $X_1, \ldots, X_n$ be independent random signs:

$$\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = \frac{1}{2}.$$

Put

$$S_n = X_1 + \cdots + X_n.$$

Then $\mathbb{E}(S_n) = 0$ and $\mathrm{Var}(S_n) = n$. Chebyshev gives

$$\mathbb{P}(|S_n| > t) \leq \frac{n}{t^2}.$$

For $t = 10\sqrt{n}$, this gives only $1/100$, while the true probability is exponentially small.

3.2 The Exponential Trick

For any random variable X, any $a \in \mathbb{R}$, and any $t > 0$,

$$\{X > a\} \subset \{e^{tX} > e^{ta}\}.$$

By Markov,

$$\mathbb{P}(X > a) = \mathbb{P}(e^{tX} > e^{ta}) \leq e^{-ta}\mathbb{E}(e^{tX}).$$

Thus

$$\boxed{\mathbb{P}(X > a) \leq e^{-ta}\mathbb{E}(e^{tX}) \quad (t > 0).}$$

This is the Chernoff or exponential moment method.

3.3 Moment Generating Function

The quantity

$$M_X(t) = \mathbb{E}(e^{tX})$$

is called the moment generating function.

3.4 Exponential Moment Lemma

Lemma 3.1. *Suppose $X_1, \ldots, X_N$ are independent random variables such that*

$$|X_i| \leq 1, \qquad \mathbb{E}(X_i) = 0.$$

Then, for every $|t| \leq 1$,

$$\mathbb{E}\left(\exp\left(t\sum_{i=1}^{N} X_i\right)\right) \leq \exp\left(t^2 \sum_{i=1}^{N} \mathrm{Var}(X_i)\right).$$

Proof. Since $\mathbb{E}(X_i) = 0$, we also have

$$\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2 = \mathbb{E}(X_i^2).$$

We want to estimate

$$\mathbb{E}(e^{t\sum_{i=1}^{N} X_i}).$$

Since the random variables $X_1, \ldots, X_N$ are independent,

$$e^{t\sum_{i=1}^{N} X_i} = \prod_{i=1}^{N} e^{tX_i}.$$

Therefore

$$\mathbb{E}(e^{t\sum_{i=1}^{N} X_i}) = \mathbb{E}\left(\prod_{i=1}^{N} e^{tX_i}\right) = \prod_{i=1}^{N} \mathbb{E}(e^{tX_i}).$$

So it is enough to estimate each factor

$$\mathbb{E}(e^{tX_i}).$$

Let

$$y = tX_i.$$

Since $|X_i| \leq 1$ and $|t| \leq 1$, we have

$$|y| \leq 1.$$

For $|y| \leq 1$, we use the elementary inequality

$$e^y \leq 1 + y + y^2.$$

Hence

$$e^{tX_i} \leq 1 + tX_i + t^2X_i^2.$$

Taking expectation,

$$\mathbb{E}(e^{tX_i}) \leq 1 + t\mathbb{E}(X_i) + t^2\mathbb{E}(X_i^2).$$

Since

$$\mathbb{E}(X_i) = 0,$$

we get

$$\mathbb{E}(e^{tX_i}) \leq 1 + t^2\mathbb{E}(X_i^2).$$

Using

$$1 + u \leq e^u,$$

we obtain

$$\mathbb{E}(e^{tX_i}) \leq e^{t^2\mathbb{E}(X_i^2)}.$$

But because $\mathbb{E}(X_i) = 0$,

$$\mathbb{E}(X_i^2) = \mathrm{Var}(X_i).$$

Therefore

$$\mathbb{E}(e^{tX_i}) \leq e^{t^2\,\mathrm{Var}(X_i)}.$$

Now multiply over i:

$$\mathbb{E}(e^{t\sum_{i=1}^N X_i}) = \prod_{i=1}^N \mathbb{E}(e^{tX_i}) \leq \prod_{i=1}^N e^{t^2\,\mathrm{Var}(X_i)}.$$

Thus

$$\mathbb{E}(e^{t\sum_{i=1}^N X_i}) \leq e^{t^2\sum_{i=1}^N \mathrm{Var}(X_i)}.$$

This proves the lemma. □

3.5 Tail Estimate

Let

$$S_N = \sum_{i=1}^N X_i.$$

We want to estimate

$$\mathbb{P}(S_N > \varepsilon N).$$

For $t > 0$, Markov's inequality gives

$$\mathbb{P}(S_N > \varepsilon N) = \mathbb{P}(e^{tS_N} > e^{t\varepsilon N}) \leq e^{-t\varepsilon N}\mathbb{E}(e^{tS_N}).$$

Using the exponential moment lemma,

$$\mathbb{P}(S_N > \varepsilon N) \leq e^{-t\varepsilon N}e^{t^2\sum_{i=1}^N \mathrm{Var}(X_i)}.$$

Since $|X_i| \leq 1$ and $\mathbb{E}X_i = 0$,

$$\mathrm{Var}(X_i) = \mathbb{E}X_i^2 \leq 1.$$

Therefore

$$\sum_{i=1}^{N} \mathrm{Var}(X_i) \leq N.$$

Hence

$$\mathbb{P}(S_N > \varepsilon N) \leq e^{-t\varepsilon N + t^2 N}.$$

Now choose

$$t = \frac{\varepsilon}{2}.$$

Then

$$-t\varepsilon N + t^2 N = -\frac{\varepsilon^2 N}{2} + \frac{\varepsilon^2 N}{4} = -\frac{\varepsilon^2 N}{4}.$$

Therefore

$$\mathbb{P}(S_N > \varepsilon N) \leq e^{-\varepsilon^2 N/4}.$$

Similarly, applying the same argument to $-S_N$, we get

$$\mathbb{P}(S_N < -\varepsilon N) \leq e^{-\varepsilon^2 N/4}.$$

Thus

$$\mathbb{P}(|S_N| > \varepsilon N) \leq \mathbb{P}(S_N > \varepsilon N) + \mathbb{P}(S_N < -\varepsilon N).$$

Hence

$$\boxed{\mathbb{P}(|S_N| > \varepsilon N) \leq 2e^{-\varepsilon^2 N/4}.}$$

3.6 One Random Sign

Let $X = \pm 1$ with probability $1/2$. Then

$$\mathbb{E}(e^{tX}) = \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \cosh t.$$

Using power series,

$$\cosh t \leq e^{t^2/2}.$$

Therefore

$$\boxed{\mathbb{E}(e^{tX}) \leq e^{t^2/2}.}$$

3.7 Sum of Independent Random Signs

For $S_n = X_1 + \cdots + X_n$, independence gives

$$\mathbb{E}(e^{tS_n}) = \prod_{k=1}^{n} \mathbb{E}(e^{tX_k}) \leq e^{nt^2/2}.$$

Thus

$$\mathbb{P}(S_n > a) \leq e^{-ta+nt^2/2}.$$

Minimize the exponent by choosing $t = a/n$. Then

$$\boxed{\mathbb{P}(S_n > a) \leq e^{-a^2/(2n)}.}$$

Applying the same estimate to $-S_n$,

$$\boxed{\mathbb{P}(|S_n| > a) \leq 2e^{-a^2/(2n)}.}$$

3.8 Weighted Random Signs

Let

$$X = \sum_{k=1}^{n} a_k \varepsilon_k,$$

where $\varepsilon_k = \pm 1$ independently with probability $1/2$. Then

$$\mathbb{E}(e^{tX}) = \prod_{k=1}^{n} \mathbb{E}(e^{ta_k\varepsilon_k}) = \prod_{k=1}^{n} \cosh(ta_k) \leq \prod_{k=1}^{n} e^{t^2 a_k^2/2}.$$

Hence

$$\boxed{\mathbb{E}(e^{tX}) \leq e^{t^2 \sum a_k^2/2}.}$$

Applying the exponential trick and optimizing,

$$\boxed{\mathbb{P}(|X| > \lambda) \leq 2\exp\left(-\frac{\lambda^2}{2\sum a_k^2}\right).}$$

Key point. This estimate is the main probabilistic tool in the May part of the course. It is used for random subsets, random polynomials, and discrepancy estimates.

4 Random Subsets and the Probabilistic Method

4.1 The Probabilistic Method

The basic philosophy is:

> If a randomly chosen object has positive probability of being good, then at least one good object exists.

We often do not explicitly construct the object. Instead, we prove that the probability of success is positive.

4.2 Random Subsets of $\{1, \dots, N\}$

For each $k = 1, \dots, N$, choose independently

$$X_k = \begin{cases} 1, & \text{if } k \text{ is selected,} \\ 0, & \text{otherwise,} \end{cases}$$

with

$$\mathbb{P}(X_k = 1) = p, \qquad \mathbb{P}(X_k = 0) = 1 - p.$$

Then the random subset is

$$S = \{k : X_k = 1\}.$$

Its size is

$$|S| = \sum_{k=1}^{N} X_k.$$

4.3 Expectation and Variance

Since $\mathbb{E}(X_k) = p$,

$$\boxed{\mathbb{E}(|S|) = Np.}$$

Since $X_k^2 = X_k$,

$$\mathrm{Var}(X_k) = \mathbb{E}(X_k^2) - (\mathbb{E}(X_k))^2 = p - p^2 = p(1-p).$$

Independence gives

$$\boxed{\mathrm{Var}(|S|) = Np(1-p).}$$

The standard deviation is

$$\sqrt{Np(1-p)}.$$

So $|S|$ is typically within order $\sqrt{Np}$ of Np.

4.4 Centering the Variables

We want to estimate

$$\mathbb{P}\{||S| - Np| \geq m\}.$$

Since

$$|S| - Np = \sum_{k=1}^{N} (X_k - p),$$

define

$$Y_k = X_k - p.$$

Then

$$Y_k = \begin{cases} 1-p, & \text{with probability } p, \\ -p, & \text{with probability } 1-p. \end{cases}$$

Notice that

$$\mathbb{E}(Y_k) = p(1-p) + (1-p)(-p) = p(1-p) - p(1-p) = 0.$$

Also,

$$|Y_k| \leq 1.$$

Now compute the variance. Since $\mathbb{E}(Y_k) = 0$,

$$\mathrm{Var}(Y_k) = \mathbb{E}(Y_k^2).$$

Hence

$$\mathrm{Var}(Y_k) = p(1-p)^2 + (1-p)p^2.$$

Factor:

$$\mathrm{Var}(Y_k) = p(1-p)\big((1-p) + p\big).$$

Therefore

$$\boxed{\mathrm{Var}(Y_k) = p(1-p).}$$

4.5 Exponential Moment Estimate

By the exponential moment lemma, for $|t| \leq 1$,

$$\mathbb{E}(e^{t\sum_{k=1}^{N} Y_k}) \leq \exp\left(t^2 \sum_{k=1}^{N} \mathrm{Var}(Y_k)\right).$$

Since

$$\mathrm{Var}(Y_k) = p(1-p),$$

we have

$$\sum_{k=1}^{N} \mathrm{Var}(Y_k) = Np(1-p).$$

Therefore

$$\boxed{\mathbb{E}(e^{t\sum_{k=1}^{N} Y_k}) \leq e^{t^2 Np(1-p)}.}$$

4.6 Upper Tail Estimate

We estimate

$$\mathbb{P}\left\{\sum_{k=1}^{N} Y_k \geq m\right\}.$$

For $t > 0$, Markov's inequality gives

$$\mathbb{P}\left\{\sum_{k=1}^{N} Y_k \geq m\right\} = \mathbb{P}\left\{e^{t\sum_{k=1}^{N} Y_k} \geq e^{tm}\right\}.$$

Therefore

$$\mathbb{P}\left\{\sum_{k=1}^{N} Y_k \geq m\right\} \leq e^{-tm}\mathbb{E}(e^{t\sum_{k=1}^{N} Y_k}).$$

Using the exponential moment estimate,

$$\mathbb{P}\left\{\sum_{k=1}^{N} Y_k \geq m\right\} \leq e^{-tm}e^{t^2 Np(1-p)}.$$

Hence

$$\boxed{\mathbb{P}\left\{\sum_{k=1}^{N} Y_k \geq m\right\} \leq \exp\left(t^2 Np(1-p) - tm\right).}$$

4.7 Choice of t

We want to minimize

$$t^2 Np(1-p) - tm.$$

The best choice would be

$$t = \frac{m}{2Np(1-p)}.$$

However, our exponential moment lemma was proved only for

$$|t| \leq 1.$$

Therefore we choose

$$\boxed{t = \min\left\{1, \frac{m}{2Np(1-p)}\right\}.}$$

4.8 First Case

Suppose

$$m \leq 2Np(1-p).$$

Then

$$t = \frac{m}{2Np(1-p)}.$$

Substituting this into the exponent gives

$$t^2Np(1-p) - tm = \frac{m^2}{4Np(1-p)} - \frac{m^2}{2Np(1-p)}.$$

Thus

$$t^2Np(1-p) - tm = -\frac{m^2}{4Np(1-p)}.$$

Therefore

$$\boxed{\mathbb{P}\left\{\sum_{k=1}^{N} Y_k \geq m\right\} \leq \exp\left(-\frac{m^2}{4Np(1-p)}\right).}$$

Since

$$\sum_{k=1}^{N} Y_k = |S| - Np,$$

we get

$$\boxed{\mathbb{P}\{|S| - Np \geq m\} \leq \exp\left(-\frac{m^2}{4Np(1-p)}\right)}$$

whenever

$$m \leq 2Np(1-p).$$

4.9 Second Case

Suppose

$$m > 2Np(1-p).$$

Then

$$t = 1.$$

Therefore

$$t^2Np(1-p) - tm = Np(1-p) - m.$$

Since

$$m > 2Np(1-p),$$

we have

$$Np(1-p) < \frac{m}{2}.$$

Hence

$$Np(1-p) - m < -\frac{m}{2}.$$

Therefore

$$\boxed{\mathbb{P}\{|S| - Np \geq m\} \leq e^{-m/2}}$$

whenever

$$m > 2Np(1-p).$$

4.10 Conclusion

Combining both cases, we obtain

$$\boxed{\mathbb{P}\{|S| - Np \geq m\} \leq \begin{cases} \exp\left(-\dfrac{m^2}{4Np(1-p)}\right), & \text{if } m \leq 2Np(1-p), \\ e^{-m/2}, & \text{if } m > 2Np(1-p). \end{cases}}$$

The same argument applied to $-Y_k$ gives the corresponding lower-tail estimate. Therefore,

$$\mathbb{P}\{||S| - Np| \geq m\} \leq 2 \begin{cases} \exp\left(-\dfrac{m^2}{4Np(1-p)}\right), & \text{if } m \leq 2Np(1-p), \\ e^{-m/2}, & \text{if } m > 2Np(1-p). \end{cases}$$

4.11 Remark on the Small p Case

The first bound cannot hold for all m, especially when p is small.

Indeed, take

$$m = N(1-p).$$

Then the event

$$|S| - Np \geq N(1-p)$$

is the same as

$$|S| = N.$$

This means every integer is chosen. Hence

$$\mathbb{P}(|S| = N) = p^N.$$

Now

$$p^N = e^{-N\log(1/p)}.$$

But the Gaussian-type estimate would give roughly

$$\exp\left(-\frac{N(1-p)}{4p}\right),$$

which is much smaller than p^N when p is very small.

Therefore the Gaussian bound is not sharp in the large deviation regime for small p. A sharper bound in that case requires a more precise Chernoff estimate.

4.12 Chebyshev Estimate

Chebyshev gives

$$\mathbb{P}\left(\big||S| - Np\big| > t\right) \leq \frac{Np(1-p)}{t^2}.$$

This proves concentration, but only polynomially.

4.13 Exponential Estimate

The exponential moment method gives a much stronger bound. In one common form,

$$\boxed{\mathbb{P}\left(\big||S| - Np\big| > m\right) \leq 2\exp\left(-c\frac{m^2}{Np}\right)}$$

for deviations in the standard range, with a universal constant $c > 0$.

More refined Chernoff bounds are possible, especially when p is small. The class notes explicitly ask for the correct sharp bound in this regime.

4.14 Random Signs Again

Instead of 0 or 1, one often chooses

$$\varepsilon_k = \pm 1$$

with probability 1/2. These are called Rademacher variables.

Basic facts:

$$\mathbb{E}\varepsilon_k = 0, \qquad \mathrm{Var}(\varepsilon_k) = 1, \qquad \varepsilon_k^2 = 1.$$

For a weighted sign sum

$$X = \sum_{k=1}^{n} a_k \varepsilon_k,$$

we have

$$\mathbb{E}X = 0, \qquad \mathrm{Var}(X) = \sum_{k=1}^{n} a_k^2.$$

The exponential tail is

$$\boxed{\mathbb{P}(|X| > \lambda) \leq 2\exp\left(-\frac{\lambda^2}{2\sum a_k^2}\right).}$$

4.15 A Sharper Chernoff Bound for Small p

Let

$$S \subset \{1, 2, \ldots, N\}$$

be a random subset obtained by choosing each integer independently with probability p.

Define

$$Y = |S|.$$

Then

$$Y \sim \mathrm{Bin}(N, p),$$

and

$$\mathbb{E}(Y) = Np.$$

We want a good estimate for

$$\mathbb{P}\{Y - Np \geq m\}.$$

Put

$$\mu = Np.$$

Then the event becomes

$$\{Y - \mu \geq m\} = \{Y \geq \mu + m\}.$$

Claim

For every $m > 0$,

$$\boxed{\mathbb{P}\{Y - \mu \geq m\} \leq \exp\left(-\left[(\mu + m)\log\left(1 + \frac{m}{\mu}\right) - m\right]\right).}$$

Equivalently,

$$\boxed{\mathbb{P}\{|S| - Np \geq m\} \leq \exp\left(-\left[(Np + m)\log\left(1 + \frac{m}{Np}\right) - m\right]\right).}$$

Thus we may take

$$\boxed{g(m, N, p) = (Np + m)\log\left(1 + \frac{m}{Np}\right) - m.}$$

Proof

Since

$$Y = \sum_{k=1}^{N} X_k,$$

where

$$X_k = \begin{cases} 1, & \text{with probability } p, \\ 0, & \text{with probability } 1 - p, \end{cases}$$

we have

$$\mathbb{E}(e^{tY}) = \prod_{k=1}^{N} \mathbb{E}(e^{tX_k}).$$

For one variable,

$$\mathbb{E}(e^{tX_k}) = (1 - p)e^0 + pe^t = 1 - p + pe^t.$$

Therefore

$$\mathbb{E}(e^{tY}) = (1 - p + pe^t)^N.$$

Using

$$1 + u \leq e^u,$$

we get

$$1 - p + pe^t = 1 + p(e^t - 1) \leq e^{p(e^t - 1)}.$$

Hence

$$\mathbb{E}(e^{tY}) \leq e^{Np(e^t-1)} = e^{\mu(e^t-1)}.$$

Now apply Markov's inequality. For $t > 0$,

$$\mathbb{P}\{Y \geq \mu + m\} = \mathbb{P}\{e^{tY} \geq e^{t(\mu+m)}\}.$$

Thus

$$\mathbb{P}\{Y \geq \mu + m\} \leq e^{-t(\mu+m)}\mathbb{E}(e^{tY}).$$

Using the estimate above,

$$\mathbb{P}\{Y \geq \mu + m\} \leq \exp\left(-t(\mu+m) + \mu(e^t - 1)\right).$$

We now choose t optimally. Minimize

$$-t(\mu+m) + \mu(e^t - 1).$$

Differentiate:

$$-(\mu+m) + \mu e^t = 0.$$

Hence

$$e^t = \frac{\mu+m}{\mu} = 1 + \frac{m}{\mu}.$$

Therefore

$$t = \log\left(1 + \frac{m}{\mu}\right).$$

Substituting this value of t, we get

$$-t(\mu+m) + \mu(e^t - 1) = -(\mu+m)\log\left(1 + \frac{m}{\mu}\right) + \mu\left(\frac{\mu+m}{\mu} - 1\right).$$

But

$$\mu\left(\frac{\mu+m}{\mu} - 1\right) = m.$$

Therefore

$$\mathbb{P}\{Y - \mu \geq m\} \leq \exp\left(-(\mu+m)\log\left(1 + \frac{m}{\mu}\right) + m\right).$$

Hence

$$\boxed{\mathbb{P}\{Y - \mu \geq m\} \leq \exp\left(-\left[(\mu+m)\log\left(1 + \frac{m}{\mu}\right) - m\right]\right).}$$

Interpretation for Small p

Recall that

$$\mu = Np.$$

When p is small and $m \gg Np$, we have

$$1 + \frac{m}{Np} \approx \frac{m}{Np}.$$

Therefore

$$g(m, N, p) = (Np + m) \log \left(1 + \frac{m}{Np}\right) - m$$

behaves like

$$m \log \left(\frac{m}{Np}\right).$$

Thus, in the large-deviation regime,

$$\boxed{\mathbb{P}\{|S| - Np \geq m\} \lesssim \exp\left(-cm \log\left(\frac{m}{Np}\right)\right).}$$

This is much better than the crude bound

$$e^{-m/2}$$

when p is small.

Sharpness Example

Take

$$m = N(1 - p).$$

Then

$$|S| - Np \geq N(1 - p)$$

is the same as

$$|S| \geq N.$$

But since $|S| \leq N$, this event is exactly

$$|S| = N.$$

That means every integer was chosen. Therefore

$$\mathbb{P}\{|S| = N\} = p^N.$$

Hence

$$\boxed{\mathbb{P}\{|S| - Np \geq N(1-p)\} = p^N = e^{-N\log(1/p)}.}$$

Now compute our exponent:

$$g(N(1-p), N, p) = (Np + N(1-p))\log\left(1 + \frac{N(1-p)}{Np}\right) - N(1-p).$$

Since

$$Np + N(1-p) = N,$$

and

$$1 + \frac{N(1-p)}{Np} = 1 + \frac{1-p}{p} = \frac{1}{p},$$

we get

$$g(N(1-p), N, p) = N\log\left(\frac{1}{p}\right) - N(1-p).$$

For small p,

$$g(N(1-p), N, p) \sim N\log\left(\frac{1}{p}\right).$$

Therefore

$$e^{-g(N(1-p),N,p)}$$

has essentially the same exponential size as

$$p^N = e^{-N\log(1/p)}.$$

This shows that the bound is sharp up to constants in the exponent.

Sharper Entropy Form

A sharper form of the Chernoff bound is

$$\boxed{\mathbb{P}\{|S| - Np \geq m\} \leq e^{-ND(q\|p)},}$$

where

$$q = p + \frac{m}{N},$$

and

$$D(q\|p) = q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}.$$

Here $D(q\|p)$ is the relative entropy between Bernoulli random variables with parameters q and p.

In the extreme case

$$m = N(1-p),$$

we have

$$q = 1.$$

Then

$$D(1\|p) = \log \frac{1}{p}.$$

So

$$e^{-ND(1\|p)} = e^{-N\log(1/p)} = p^N.$$

Thus the entropy form gives the correct exponential behavior exactly.

5 Random Sign Polynomials and Rudin-Shapiro Polynomials

5.1 The Problem

Consider polynomials

$$p(z) = \sum_{k=0}^{n-1} c_k z^k, \qquad c_k = \pm 1.$$

We want to make

$$\sup_{|z|=1} |p(z)|$$

as small as possible.

The trivial bound is

$$|p(z)| \leq \sum_{k=0}^{n-1} |c_k| = n.$$

But this is not sharp.

5.2 Why $\sqrt{n}$ Is the Natural Scale

On the unit circle, put $z = e^{it}$. Then observe that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |p(e^{i\theta})|^2 \, d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{k=0}^{n-1} c_k e^{ik\theta} \right) \left(\sum_{\ell=0}^{n-1} \overline{c_\ell} e^{-i\ell\theta} \right) d\theta.$$

Expanding the product,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |p(e^{i\theta})|^2 \, d\theta = \sum_{k,\ell=0}^{n-1} c_k \overline{c_\ell} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(k-\ell)\theta} \, d\theta.$$

But

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(k-\ell)\theta} \, d\theta = \begin{cases} 1, & k = \ell, \\ 0, & k \neq \ell. \end{cases}$$

Therefore

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |p(e^{i\theta})|^2 \, d\theta = \sum_{k=0}^{n-1} |c_k|^2.$$

Since $c_k = \pm 1$, we have $|c_k|^2 = 1$. Hence

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |p(e^{i\theta})|^2 \, d\theta = n.$$

Thus

$$\|p\|_{L^2(\mathbb{T})} = \sqrt{n}.$$

Since

$$\|p\|_{L^2(\mathbb{T})} \leq \|p\|_{L^\infty(\mathbb{T})},$$

we must have

$$\max_{|z|=1} |p(z)| \geq \sqrt{n}.$$

So the best possible order one can hope for is $\sqrt{n}$.

5.3 Random Signs

Choose

$$c_k = \varepsilon_k,$$

where the ε_k are independent random signs. Fix $z = e^{i\theta}$. Then

$$p(z) = \sum_{k=0}^{n-1} \varepsilon_k e^{ik\theta}.$$

The real and imaginary parts are

$$\operatorname{Re} p(z) = \sum_{k=0}^{n-1} \varepsilon_k \cos(k\theta),$$

$$\operatorname{Im} p(z) = \sum_{k=0}^{n-1} \varepsilon_k \sin(k\theta).$$

Each is a weighted sum of independent random signs.
Since $\sum \cos^2(k\theta) \leq n$,

$$\mathbb{P}(|\operatorname{Re} p(z)| > t) \leq 2e^{-t^2/(2n)}.$$

Similarly,

$$\mathbb{P}(|\operatorname{Im} p(z)| > t) \leq 2e^{-t^2/(2n)}.$$

Now suppose

$$|p(z)| > t.$$

Then either

$$|\operatorname{Re} p(z)| > \frac{t}{\sqrt{2}}$$

or

$$|\operatorname{Im} p(z)| > \frac{t}{\sqrt{2}}.$$

Therefore, by the union bound,

$$\mathbb{P}\{|p(z)| > t\} \leq \mathbb{P}\left\{|\operatorname{Re} p(z)| > \frac{t}{\sqrt{2}}\right\} + \mathbb{P}\left\{|\operatorname{Im} p(z)| > \frac{t}{\sqrt{2}}\right\}.$$

Hence

$$\mathbb{P}\{|p(z)| > t\} \leq 4e^{-t^2/(8n)}.$$

Thus, for every fixed z with $|z| = 1$,

$$\boxed{\mathbb{P}\{|p(z)| > t\} \leq 4e^{-t^2/(8n)}.}$$

5.4 Controlling Finitely Many Points

Now choose M points

$$z_1, z_2, \ldots, z_M$$

on the unit circle.

Then

$$\mathbb{P}\left\{\max_{1\leq j\leq M} |p(z_j)| > t\right\} = \mathbb{P}\left(\bigcup_{j=1}^{M}\{|p(z_j)| > t\}\right).$$

By the union bound,

$$\mathbb{P}\left\{\max_{1\leq j\leq M} |p(z_j)| > t\right\} \leq \sum_{j=1}^{M} \mathbb{P}\{|p(z_j)| > t\}.$$

Using the previous estimate,

$$\mathbb{P}\left\{\max_{1\leq j\leq M} |p(z_j)| > t\right\} \leq 4Me^{-t^2/(8n)}.$$

Therefore, if

$$4Me^{-t^2/(8n)} < 1,$$

then there exists a choice of signs $c_k = \pm 1$ such that

$$|p(z_j)| \leq t$$

for every $j = 1, \ldots, M$.

The condition

$$4Me^{-t^2/(8n)} < 1$$

is equivalent to

$$e^{-t^2/(8n)} < \frac{1}{4M}.$$

Taking logarithms,

$$-\frac{t^2}{8n} < -\log(4M).$$

Thus it is enough to choose

$$t > \sqrt{8n \log(4M)}.$$

5.5 Passing from Finitely Many Points to the Whole Circle

The previous argument only controls p at the points

$$z_1, \ldots, z_M.$$

But we want to control

$$\max_{|z|=1} |p(z)|.$$

Choose the points $z_1, \ldots, z_M$ equally spaced on the unit circle. Then for every z with $|z| = 1$, there exists some z_j such that the arc distance between z and z_j is at most

$$\frac{\pi}{M}.$$

By the triangle inequality,

$$|p(z)| \leq |p(z_j)| + |p(z) - p(z_j)|.$$

We estimate the second term using the Fundamental Theorem of Calculus along the arc from z_j to z:

$$|p(z) - p(z_j)| \leq \int_{z_j}^{z} |p'(w)| \, |dw|.$$

Since the arc length is at most π/M, we get

$$|p(z) - p(z_j)| \leq \frac{\pi}{M} \max_{|w|=1} |p'(w)|.$$

Now

$$p(z) = \sum_{k=0}^{n-1} c_k z^k,$$

so

$$p'(z) = \sum_{k=1}^{n-1} k c_k z^{k-1}.$$

If $|z| = 1$, then $|z^{k-1}| = 1$. Also $|c_k| = 1$. Hence

$$|p'(z)| \leq \sum_{k=1}^{n-1} k|c_k||z|^{k-1} = \sum_{k=1}^{n-1} k.$$

Therefore

$$|p'(z)| \leq \frac{n(n-1)}{2}.$$

Thus

$$\max_{|z|=1} |p'(z)| \leq \frac{n(n-1)}{2}.$$

Consequently,

$$|p(z) - p(z_j)| \leq \frac{\pi}{M} \frac{n(n-1)}{2}.$$

Now choose

$$M = n^2.$$

Then

$$\frac{\pi}{M} \frac{n(n-1)}{2} = \frac{\pi}{n^2} \frac{n(n-1)}{2} = \frac{\pi(n-1)}{2n} \leq \frac{\pi}{2}.$$

Therefore, for every z on the unit circle,

$$|p(z)| \leq |p(z_j)| + \frac{\pi}{2}.$$

Since we have chosen the signs so that

$$|p(z_j)| \leq t$$

for all j, we obtain

$$|p(z)| \leq t + \frac{\pi}{2}.$$

Hence

$$\max_{|z|=1} |p(z)| \leq t + \frac{\pi}{2}.$$

5.6 Final Estimate

Take

$$M = n^2.$$

Then

$$t > \sqrt{8n \log(4M)} = \sqrt{8n \log(4n^2)}.$$

Since

$$\log(4n^2) = \log 4 + 2 \log n \lesssim \log n,$$

we get

$$t \lesssim \sqrt{n \log n}.$$

Therefore there exists a choice of signs $c_k = \pm 1$ such that

$$\max_{|z|=1} |p(z)| \leq C\sqrt{n \log n}$$

for some absolute constant $C > 0$.

Thus we have proved:

$$\boxed{\text{There exists } p(z) = \sum_{k=0}^{n-1} c_k z^k, \quad c_k = \pm 1, \quad \text{such that} \quad \max_{|z|=1} |p(z)| \leq C\sqrt{n \log n}.}$$

5.7 Can the Logarithm Be Removed?

The random method gives $\sqrt{n \log n}$. The natural question is whether one can achieve $C\sqrt{n}$. Rudin-Shapiro polynomials give a deterministic construction with the optimal order.

5.8 Rudin-Shapiro Polynomials

Define

$$P_0(z) = Q_0(z) = 1.$$

Then recursively set

$$P_{k+1}(z) = P_k(z) + z^{2^k} Q_k(z),$$
$$Q_{k+1}(z) = P_k(z) - z^{2^k} Q_k(z).$$

The coefficients remain ± 1.

5.9 The Main Identity

On $|z| = 1$,

$$\begin{aligned} |P_{k+1}|^2 + |Q_{k+1}|^2 &= |P_k + z^{2^k} Q_k|^2 + |P_k - z^{2^k} Q_k|^2 \\ &= 2|P_k|^2 + 2|Q_k|^2. \end{aligned}$$

Thus

$$\boxed{|P_{k+1}|^2 + |Q_{k+1}|^2 = 2(|P_k|^2 + |Q_k|^2).}$$

By induction,

$$|P_k|^2 + |Q_k|^2 = 2^{k+1}.$$

If $N = 2^k$, then

$$\boxed{|P_k(z)| \leq \sqrt{2N} \qquad (|z| = 1).}$$

This is essentially optimal.

5.10 Open Question

Can one get, for every $\varepsilon > 0$, signs such that

$$\left| \sum_{k=0}^{N-1} (\pm 1) z^k \right| \leq (1 + \varepsilon)\sqrt{N}$$

for all $|z| = 1$ and sufficiently large N? This is a natural sharpening beyond Rudin-Shapiro.

6 Pigeonhole Principle and Dirichlet Approximation

6.1 The Pigeonhole Principle

Theorem 6.1 (Pigeonhole principle)**.** *If $N+1$ objects are placed into N boxes, then some box contains at least two objects.*

More generally, if n objects are placed into k boxes, then some box contains at least n/k objects, up to rounding.

6.2 Fractional Part

For $x \in \mathbb{R}$, define the fractional part

$$\{x\} = x - \lfloor x \rfloor.$$

Then

$$0 \leq \{x\} < 1.$$

Also define distance to the nearest integer:

$$\|x\| = \min_{m \in \mathbb{Z}} |x - m|.$$

6.3 Dirichlet Approximation Theorem

Theorem 6.2 (Dirichlet)**.** *Let $\alpha \in \mathbb{R}$. For every positive integer Q, there exist integers p, q, with*

$$1 \leq q \leq Q,$$

such that

$$\boxed{\left|\alpha - \frac{p}{q}\right| < \frac{1}{qQ}.}$$

Proof. Consider the $Q+1$ numbers

$$0, \{\alpha\}, \{2\alpha\}, \ldots, \{Q\alpha\}$$

in $[0,1)$. Divide $[0,1)$ into Q intervals of length $1/Q$. By the pigeonhole principle, two of the fractional parts lie in the same interval. Hence for some $0 \leq m < n \leq Q$,

$$|\{n\alpha\} - \{m\alpha\}| < \frac{1}{Q}.$$

Let $q = n - m$. Then $1 \leq q \leq Q$. Also

$$q\alpha = (n-m)\alpha = p + \delta$$

for some integer p and some $|\delta| < 1/Q$. Thus

$$|q\alpha - p| < \frac{1}{Q}.$$

Dividing by q,

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{qQ}.$$

□

6.4 Simultaneous Approximation

Let

$$\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}.$$

We want to approximate each α_j by a rational number with the same denominator:

$$\alpha_j \approx \frac{p_j}{q}.$$

The error is measured by

$$\varepsilon = \max_{1 \leq j \leq m} \left|\alpha_j - \frac{p_j}{q}\right|.$$

Trivially, if q is fixed, we can choose p_j to be the nearest integer to $q\alpha_j$. Then

$$\left|\alpha_j - \frac{p_j}{q}\right| \leq \frac{1}{q}.$$

The goal is to do better by choosing q carefully.

Theorem 6.3 (Simultaneous Dirichlet Approximation)**.** *Let*

$$\alpha_1, \ldots, \alpha_m \in \mathbb{R}.$$

For every positive integer Q, there exist integers

$$1 \leq q \leq Q$$

and

$$p_1, \ldots, p_m \in \mathbb{Z}$$

such that

$$\max_{1\le j\le m}\left|\alpha_j-\frac{p_j}{q}\right|\le\frac{C_m}{qQ^{1/m}},$$

where C_m is a constant depending only on m.

If Q is an exact m-th power, one may take $C_m=1$, so that

$$\max_{1\le j\le m}\left|\alpha_j-\frac{p_j}{q}\right|<\frac{1}{qQ^{1/m}}.$$

Proof. We first prove the theorem in the special case when Q is an exact m-th power.

Suppose

$$Q=L^m$$

for some positive integer L.

For each integer q, define the vector

$$v_q=(\{q\alpha_1\},\{q\alpha_2\},\dots,\{q\alpha_m\})\in[0,1)^m,$$

where

$$\{x\}=x-\lfloor x\rfloor$$

denotes the fractional part of x.

Now consider the $Q+1$ points

$$v_0,v_1,\dots,v_Q$$

inside the unit cube

$$[0,1)^m.$$

Since

$$Q=L^m,$$

we divide the cube $[0,1)^m$ into $L^m=Q$ smaller cubes, each of side length

$$\frac{1}{L}=\frac{1}{Q^{1/m}}.$$

There are $Q+1$ points and only Q boxes. Therefore, by the pigeonhole principle, two of the points lie in the same small cube.

Thus there exist integers

$$0\le q'<q\le Q$$

such that v_q and $v_{q'}$ lie in the same small cube.

Therefore, for each $j=1,\dots,m$,

$$|\{q\alpha_j\}-\{q'\alpha_j\}|<\frac{1}{L}.$$

Now write

$$q\alpha_j = \lfloor q\alpha_j \rfloor + \{q\alpha_j\},$$

and similarly

$$q'\alpha_j = \lfloor q'\alpha_j \rfloor + \{q'\alpha_j\}.$$

Subtracting,

$$(q - q')\alpha_j = (\lfloor q\alpha_j \rfloor - \lfloor q'\alpha_j \rfloor) + (\{q\alpha_j\} - \{q'\alpha_j\}).$$

Let

$$Q_1 = q - q'.$$

Then

$$1 \leq Q_1 \leq Q.$$

Define

$$p_j = \lfloor q\alpha_j \rfloor - \lfloor q'\alpha_j \rfloor.$$

Then $p_j \in \mathbb{Z}$, and

$$|Q_1\alpha_j - p_j| < \frac{1}{L}.$$

Dividing by Q_1, we get

$$\left|\alpha_j - \frac{p_j}{Q_1}\right| < \frac{1}{Q_1 L}.$$

Since

$$L = Q^{1/m},$$

we obtain

$$\left|\alpha_j - \frac{p_j}{Q_1}\right| < \frac{1}{Q_1 Q^{1/m}}.$$

Thus, if we rename Q_1 as q, we have found integers

$$1 \leq q \leq Q$$

and

$$p_1, \ldots, p_m \in \mathbb{Z}$$

such that

$$\max_{1 \leq j \leq m} \left|\alpha_j - \frac{p_j}{q}\right| < \frac{1}{qQ^{1/m}}.$$

This proves the theorem when Q is an exact m-th power.

Now suppose Q is not an exact m-th power.

Let

$$L = \lfloor Q^{1/m} \rfloor.$$

Then

$$L^m \leq Q.$$

We divide $[0,1)^m$ into L^m smaller cubes, each of side length

$$\frac{1}{L}.$$

Now consider the $L^m + 1$ points

$$v_0, v_1, \ldots, v_{L^m}.$$

There are $L^m + 1$ points and only L^m boxes. Hence, by the pigeonhole principle, two of these points lie in the same box.

Thus there exist

$$0 \leq q' < q \leq L^m$$

such that for each $j = 1, \ldots, m$,

$$|\{q\alpha_j\} - \{q'\alpha_j\}| < \frac{1}{L}.$$

Set

$$Q_1 = q - q'.$$

Then

$$1 \leq Q_1 \leq L^m \leq Q.$$

As before, define

$$p_j = \lfloor q\alpha_j \rfloor - \lfloor q'\alpha_j \rfloor.$$

Then

$$|Q_1\alpha_j - p_j| < \frac{1}{L}.$$

Dividing by Q_1,

$$\left|\alpha_j - \frac{p_j}{Q_1}\right| < \frac{1}{Q_1 L}.$$

Since

$$L = \lfloor Q^{1/m} \rfloor,$$

we have, for large Q,

$$L \geq \frac{1}{2} Q^{1/m}.$$

Therefore

$$\frac{1}{L} \leq \frac{2}{Q^{1/m}}.$$

Hence

$$\left|\alpha_j - \frac{p_j}{Q_1}\right| < \frac{2}{Q_1 Q^{1/m}}.$$

Renaming Q_1 as q, we obtain

$$1 \le q \le Q$$

and

$$p_1, \ldots, p_m \in \mathbb{Z}$$

such that

$$\max_{1\le j\le m}\left|\alpha_j - \frac{p_j}{q}\right| < \frac{2}{qQ^{1/m}}.$$

Thus the theorem holds for arbitrary Q, with a harmless constant. □

6.5 Why This Appears Here

Dirichlet approximation and Spencer's theorem share the same skeleton:

$$\text{many objects} + \text{few boxes} \Longrightarrow \text{two objects collide.}$$

In Dirichlet, the objects are fractional-part vectors. In Spencer's theorem, the objects are sign vectors and the boxes are geometric covering sets.

7 Discrepancy Theory and Spencer's Theorem

7.1 The Discrepancy Problem

We are given an $m \times n$ matrix

$$A = (a_{ij}), \qquad 1 \le i \le m, \quad 1 \le j \le n,$$

with

$$|a_{ij}| \le 1.$$

We want to choose signs

$$\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$$

so that all the row sums

$$y_i = \sum_{j=1}^{n} a_{ij}\varepsilon_j$$

are small simultaneously.

Equivalently, we want to make

$$\max_{1 \le i \le m} |y_i|$$

as small as possible.

This is a discrepancy problem.

7.2 Discrepancy

Definition 7.1. Define the discrepancy of the matrix A by

$$\operatorname{disc}(A) = \min_{\varepsilon_j = \pm 1} \max_{1 \le i \le m} \left| \sum_{j=1}^{n} a_{ij}\varepsilon_j \right|.$$

So the problem is:

Can we choose signs so that every row is balanced?

The trivial estimate is easy. Since

$$|a_{ij}| \le 1,$$

we have

$$|y_i| = \left| \sum_{j=1}^{n} a_{ij}\varepsilon_j \right| \le \sum_{j=1}^{n} |a_{ij}| \le n.$$

Therefore,

$$\max_i |y_i| \leq n.$$

Thus

$$\boxed{\text{disc}(A) \leq n.}$$

But this is a poor estimate. We want something closer to $\sqrt{n}$.

7.3 Random Signs Give $\sqrt{n \log m}$

First choose the signs randomly:

$$\varepsilon_j = \pm 1$$

independently with probability $1/2$.

For a fixed row i, define

$$Y_i = \sum_{j=1}^{n} a_{ij}\varepsilon_j.$$

Since

$$\mathbb{E}(\varepsilon_j) = 0,$$

we get

$$\mathbb{E}(Y_i) = 0.$$

Also,

$$\text{Var}(Y_i) = \sum_{j=1}^{n} a_{ij}^2.$$

Since

$$|a_{ij}| \leq 1,$$

we have

$$a_{ij}^2 \leq 1.$$

Hence

$$\text{Var}(Y_i) = \sum_{j=1}^{n} a_{ij}^2 \leq n.$$

By the exponential moment estimate, for a fixed i,

$$\mathbb{P}(|Y_i| > t) \leq 2e^{-t^2/(4n)}.$$

Now we want this for every row $i = 1, \dots, m$. Use the union bound:

$$\mathbb{P}\left(\max_{1\le i\le m} |Y_i| > t\right) \le \sum_{i=1}^{m} \mathbb{P}(|Y_i| > t).$$

Therefore

$$\mathbb{P}\left(\max_{i} |Y_i| > t\right) \le 2me^{-t^2/(4n)}.$$

To make this probability less than 1, it is enough to choose t such that

$$2me^{-t^2/(4n)} < 1.$$

Taking logarithms, this means

$$\frac{t^2}{4n} > \log(2m).$$

So we may choose

$$t \sim \sqrt{n \log m}.$$

Thus random signs prove

$$\boxed{\operatorname{disc}(A) \le C\sqrt{n \log m}.}$$

This is good, but not optimal when m is comparable to n.

7.4 Why $\sqrt{n}$

A Hadamard matrix H is an $n \times n$ matrix with entries ± 1 such that the rows are orthogonal. Equivalently,

$$H^T H = nI.$$

Take

$$A = H.$$

For any sign vector

$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n),$$

define

$$y = H\varepsilon.$$

Then

$$\sum_{i=1}^{n} y_i^2 = \|H\varepsilon\|_2^2.$$

But

$$\|H\varepsilon\|_2^2 = \varepsilon^T H^T H\varepsilon.$$

Since

$$H^T H = nI,$$

we get

$$\varepsilon^T H^T H\varepsilon = n\varepsilon^T\varepsilon.$$

Also,

$$\varepsilon^T\varepsilon = \sum_{j=1}^{n} \varepsilon_j^2 = n.$$

Therefore

$$\sum_{i=1}^{n} y_i^2 = n^2.$$

Hence at least one coordinate satisfies

$$y_i^2 \geq n.$$

So

$$\max_i |y_i| \geq \sqrt{n}.$$

Thus, in general, we cannot hope for better than order

$$\sqrt{n}.$$

So the ideal theorem would say

$$\text{disc}(A) \leq C\sqrt{n}.$$

Spencer's theorem says this is true when

$$m \leq n.$$

7.5 Spencer's Theorem

Theorem 7.2 (Spencer's Theorem)**.** *Suppose*

$$m \leq n$$

and

$$|a_{ij}| \leq 1.$$

Then there exist signs

$$\varepsilon_j = \pm 1$$

such that

$$\boxed{\max_{1\le i\le m}\left|\sum_{j=1}^{n} a_{ij}\varepsilon_j\right| \le C\sqrt{n}.}$$

This theorem is often called

$$\boxed{\text{Six standard deviations suffice.}}$$

Random signs give $\sqrt{n\log m}$. Spencer removes the logarithmic loss when $m \le n$.

7.6 Random Signs and the Cross-Polytope

Let

$$X_j = \pm 1$$

independently with probability $1/2$.

Define

$$Y_i = \sum_{j=1}^{n} a_{ij}X_j.$$

We know

$$\mathbb{E}(Y_i^2) = \sum_{j=1}^{n} a_{ij}^2 \le n.$$

Therefore

$$\mathbb{E}\left(\sum_{i=1}^{m} Y_i^2\right) = \sum_{i=1}^{m} \mathbb{E}(Y_i^2) \le mn.$$

By Markov's inequality,

$$\mathbb{P}\left(\sum_{i=1}^{m} Y_i^2 > 4mn\right) \le \frac{\mathbb{E}\left(\sum_{i=1}^{m} Y_i^2\right)}{4mn}.$$

Since

$$\mathbb{E}\left(\sum_{i=1}^{m} Y_i^2\right) \le mn,$$

we get

$$\mathbb{P}\left(\sum_{i=1}^{m} Y_i^2 > 4mn\right) \leq \frac{1}{4}.$$

So with probability at least 3/4,

$$\sum_{i=1}^{m} Y_i^2 \leq 4mn.$$

Now use Cauchy–Schwarz:

$$\left(\sum_{i=1}^{m} |Y_i|\right)^2 \leq m \sum_{i=1}^{m} Y_i^2.$$

Thus, with positive probability,

$$\sum_{i=1}^{m} |Y_i| \leq 2m\sqrt{n}.$$

Normalize by defining

$$Z_i = \frac{Y_i}{\sqrt{n}}.$$

Then

$$\sum_{i=1}^{m} |Z_i| \leq 2m.$$

So many sign choices produce points

$$Z = (Z_1, \ldots, Z_m)$$

inside the body

$$K_m = \left\{ z \in \mathbb{R}^m : \sum_{i=1}^{m} |z_i| \leq 2m \right\}.$$

This body is called the cross-polytope. It is the ℓ^1-ball of radius $2m$.

The important conclusion is:

$$\boxed{\text{At least } 2^{n-1} \text{ sign choices land inside } K_m.}$$

We use 2^{n-1} instead of $3 \cdot 2^n/4$ because the weaker estimate is enough.

7.7 Covering the Cross-Polytope by Boxes

We now want to cover

$$K_m = \left\{ z \in \mathbb{R}^m : \sum_i |z_i| \leq 2m \right\}$$

by cubes of side length B.

Let F_m be the least number of such cubes needed.

The goal is to prove

$$F_m < 2^{n-1}.$$

Then since at least 2^{n-1} sign choices land inside K_m, two of them must land in the same cube.

That is where the pigeonhole principle enters.

7.8 The Recursive Covering Estimate

Take

$$z \in K_m.$$

Define the set of large coordinates:

$$I = \left\{ i : |z_i| \geq \frac{B}{2} \right\}.$$

Since

$$\sum_i |z_i| \leq 2m,$$

and each coordinate in I contributes at least $B/2$, we get

$$|I| \frac{B}{2} \leq 2m.$$

Hence

$$|I| \leq \frac{4m}{B}.$$

Put

$$m' = \frac{4m}{B}.$$

Thus every point in K_m has at most m' large coordinates.

This is the key geometric observation.

To cover K_m, we first choose the possible set I of large coordinates. There are at most

$$\binom{m}{m'}$$

choices.

For fixed I, the coordinates outside I are small and can be covered by one interval of length B. The coordinates inside I form a smaller-dimensional problem of dimension m'.

This gives a recurrence of the form

$$F_m \leq \binom{m}{m'} B^{m'} F_{m'}.$$

Using the standard binomial estimate

$$\binom{m}{m'} \leq \left(\frac{em}{m'}\right)^{m'},$$

and

$$m' = \frac{4m}{B},$$

we have

$$\frac{m}{m'} = \frac{B}{4}.$$

Hence

$$\binom{m}{m'} \leq \left(\frac{eB}{4}\right)^{m'}.$$

For $B \geq 16$, this is bounded by something like

$$B^{m'}.$$

Therefore

$$F_m \leq B^{m'} B^{m'} F_{m'}.$$

Thus

$$F_m \leq B^{2m'} F_{m'}.$$

Since

$$m' = \frac{4m}{B},$$

we get

$$\boxed{F_m \leq B^{8m/B} F_{4m/B}.}$$

Now iterate this recurrence.

The exponents form a geometric series:

$$\frac{8m}{B} + \frac{8m}{B}\frac{4}{B} + \frac{8m}{B}\left(\frac{4}{B}\right)^2 + \cdots .$$

If $B \geq 16$, then

$$\frac{4}{B} \leq \frac{1}{4}.$$

So the sum is bounded by

$$\frac{16m}{B}.$$

Therefore

$$\boxed{F_m \leq B^{16m/B}.}$$

This is the covering estimate.

7.9 Choosing B

We want

$$F_m \leq 2^{n-1}.$$

It is enough to require

$$B^{16m/B} \leq 2^{n/2}.$$

Taking logarithms,

$$\frac{16m}{B} \log B \leq \frac{n}{2} \log 2.$$

Roughly, this means B should satisfy

$$B \gtrsim \frac{m}{n} \log B.$$

If

$$m \leq n,$$

then $m/n \leq 1$, so we may take B to be a large absolute constant.

Thus when $m \leq n$, the cross-polytope can be covered by fewer than 2^{n-1} boxes of constant side length.

Therefore, by the pigeonhole principle, at least two sign choices land in the same box.

When $m > n$, one needs B growing roughly like

$$B \sim C \log\left(\frac{m}{n}\right)$$

or a related logarithmic quantity. This gives a weaker but still useful discrepancy bound.

7.10 The Pigeonhole Step

We have at least

$$2^{n-1}$$

sign choices landing inside K_m.

If K_m is covered by fewer than

$$2^{n-1}$$

boxes, then some box contains at least two sign choices.

In the refined version of the notes, one actually gets a box containing at least

$$2^{n/2-1}$$

sign choices.

Let that box be

$$\mathcal{B}.$$

Then many sign vectors

$$X = (X_1, \ldots, X_n) \in \{-1, 1\}^n$$

have their normalized image

$$z_i(X) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{ij} X_j$$

inside the same box $\mathcal{B}$.

If two sign choices X' and X'' lie in the same box of side length B, then for every row i,

$$|z_i(X') - z_i(X'')| \leq B.$$

This closeness is what will give a small discrepancy.

7.11 We Need Two Sign Choices Far Apart

If we simply choose two sign vectors in the same box, they might differ in only one coordinate.

That would not be useful.

We need two sign choices

$$X', X''$$

which are:

(1) in the same box, so their images are close;

(2) far apart in Hamming distance, so they differ in many coordinates.

The Hamming distance between X' and X'' is the number of indices j such that

$$X'_j \neq X''_j.$$

7.12 Counting Nearby Sign Choices

Fix one sign vector X'.

How many sign choices can be obtained from X' by flipping at most n_0 signs?

The answer is

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n_0}.$$

This is the size of a Hamming ball of radius n_0.

We estimate it using the binomial theorem.

For $0 < t < 1$,

$$(1+t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k.$$

For $k \leq n_0$, we have

$$t^k \geq t^{n_0}.$$

Therefore

$$(1+t)^n \geq \sum_{k=0}^{n_0} \binom{n}{k} t^k \geq t^{n_0} \sum_{k=0}^{n_0} \binom{n}{k}.$$

Hence

$$\sum_{k=0}^{n_0} \binom{n}{k} \leq t^{-n_0}(1+t)^n.$$

Optimizing gives

$$t = \frac{n_0}{n - n_0}.$$

Write

$$n_0 = q_0 n.$$

Then the estimate becomes roughly

$$\sum_{k=0}^{q_0 n} \binom{n}{k} \leq \left[\left(\frac{1}{q_0} \right)^{q_0} \left(\frac{1}{1 - q_0} \right)^{1-q_0} \right]^n.$$

For $q_0 > 0$ small enough, this is less than

$$2^{n/2-1}.$$

Thus, inside a box containing at least $2^{n/2-1}$ sign choices, not all of them can lie within distance $q_0 n$ of X'.

Therefore there exists another sign vector X'' in the same box such that

$$X'_j \neq X''_j$$

for at least

$$q_0 n$$

indices j.

7.13 Constructing the Partial Coloring

Now define

$$\varepsilon_j = \frac{X'_j - X''_j}{2}.$$

Since

$$X'_j, X''_j \in \{-1, 1\},$$

we have

$$\varepsilon_j \in \{-1, 0, 1\}.$$

Indeed:

$$X'_j = X''_j \quad \Longrightarrow \quad \varepsilon_j = 0,$$

while

$$X'_j \neq X''_j \quad \Longrightarrow \quad \varepsilon_j = \pm 1.$$

Since X' and X'' differ in at least $q_0 n$ coordinates, at least $q_0 n$ of the ε_j's are nonzero.

Thus ε is a partial coloring: it assigns signs to a positive proportion of the coordinates and leaves the rest uncolored.

7.14 Bounding the Error of the Partial Coloring

For each row i,

$$\sum_{j=1}^{n} a_{ij}\varepsilon_j = \frac{1}{2}\sum_{j=1}^{n} a_{ij}(X'_j - X''_j).$$

Using the definition of z_i,

$$z_i(X') = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{ij}X'_j,$$

and

$$z_i(X'') = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} a_{ij}X''_j.$$

Therefore

$$\sum_{j=1}^{n} a_{ij}(X'_j - X''_j) = \sqrt{n}\left(z_i(X') - z_i(X'')\right).$$

Hence

$$\left|\sum_{j=1}^{n} a_{ij}\varepsilon_j\right| = \frac{\sqrt{n}}{2}|z_i(X') - z_i(X'')|.$$

But X' and X'' lie in the same box of side length B, so

$$|z_i(X') - z_i(X'')| \leq B.$$

Therefore

$$\boxed{\left|\sum_{j=1}^{n} a_{ij}\varepsilon_j\right| \leq \frac{B\sqrt{n}}{2}}$$

for every row i.

This is the partial coloring lemma.

It says:

We can color at least $q_0 n$ coordinates while keeping every row sum bounded by $CB\sqrt{n}$.

7.15 Iterating the Partial Coloring

After the first step, at least $q_0 n$ coordinates have been colored.

So the number of uncolored coordinates is at most

$$n(1-q_0).$$

Now repeat the argument on the remaining uncolored columns.

At step 0, there are n columns.

At step 1, there are at most

$$n(1-q_0)$$

columns.

At step 2, there are at most

$$n(1-q_0)^2$$

columns.

At step k, there are at most

$$n(1-q_0)^k$$

columns.

Each step colors a fixed positive proportion of the remaining columns.

Eventually the number of uncolored columns becomes 0.

7.16 Error Accumulation

At the k-th step, the number of active columns is

$$n_k = n(1-q_0)^k.$$

The error contributed at that step is roughly

$$C\sqrt{n_k}\left[\log\left(\frac{m}{n_k}\right)\right].$$

Since

$$n_k = n(1-q_0)^k,$$

we have

$$\log\left(\frac{m}{n_k}\right) = \log\left(\frac{m}{n}\right) + k\log\left(\frac{1}{1-q_0}\right).$$

So the total error is bounded by

$$C\sqrt{n}\sum_{k\geq 0}(1-q_0)^{k/2}\left[\log\left(\frac{m}{n}\right) + k\log\left(\frac{1}{1-q_0}\right)\right].$$

This sum converges because

$$(1-q_0)^{k/2}$$

decays geometrically.

Therefore

$$\max_i \left|\sum_{j=1}^{n} a_{ij}\varepsilon_j\right| \leq C\sqrt{n}\left(1+\log\left(\frac{m}{n}\right)\right).$$

So the cross-polytope argument gives a bound of the form

$$\boxed{\operatorname{disc}(A) \leq C\sqrt{n}\left(1+\log\left(\frac{m}{n}\right)\right).}$$

When $m \leq n$, the logarithmic part is harmless, and we get

$$\boxed{\operatorname{disc}(A) \leq C\sqrt{n}.}$$

This is the Spencer-type conclusion.

7.17 Why the Euclidean Ball Appears

At the end of the notes, we ask:

$$\text{What if we cover } \left\{z : \sum_{i=1}^{m} z_i^2 \leq 2m\right\}?$$

This is the Euclidean ball, or ℓ^2-ball.

Previously we covered the cross-polytope:

$$\sum_i |z_i| \leq 2m.$$

That is an ℓ^1-body.

The Euclidean ball is smaller and rounder:

$$\sum_i z_i^2 \leq 2m.$$

Covering the Euclidean ball can give a better covering estimate than covering the cross-polytope.

That improvement is related to the sharper Spencer bound

$$\boxed{\operatorname{disc}(A) \leq C\sqrt{n \log\left(\frac{2m}{n}\right)}}$$

for $m \geq n$, and in particular

$$\operatorname{disc}(A) \leq C\sqrt{n}$$

when $m \leq n$.

So the Euclidean-ball question is not random. It is the next natural improvement after the cross-polytope argument.

7.18 Motivation

In the proof of Spencer's theorem, we previously considered the cross-polytope

$$\left\{ z \in \mathbb{R}^m : \sum_{i=1}^{m} |z_i| \leq 2m \right\}.$$

We ask:

$$\text{What do we get if we instead cover } \left\{ z \in \mathbb{R}^m : \sum_{i=1}^{m} z_i^2 \leq 2m \right\}?$$

This new set is the Euclidean ball of radius $\sqrt{2m}$:

$$B_2^m(\sqrt{2m}) = \left\{ z \in \mathbb{R}^m : \sum_{i=1}^{m} z_i^2 \leq 2m \right\}.$$

The point is that covering the Euclidean ball gives a better covering estimate than covering the cross-polytope.

The reason is simple:

$$\ell^1 \text{ condition: } \sum_i |z_i| \leq 2m$$

only charges a large coordinate linearly, while

$$\ell^2 \text{ condition: } \sum_i z_i^2 \leq 2m$$

charges a large coordinate quadratically.

Thus large coordinates are more expensive in the Euclidean ball.

7.19 Recall the Setting

Let

$$A = (a_{ij})$$

be an $m \times n$ matrix with

$$|a_{ij}| \leq 1.$$

Choose random signs

$$X_j = \pm 1$$

independently with probability $1/2$, and define

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j.$$

Normalize by

$$Z_i = \frac{Y_i}{\sqrt{n}}.$$

Thus

$$Z_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{ij} X_j.$$

Earlier we used the body

$$\sum_{i=1}^{m} |Z_i| \leq 2m.$$

Now we instead use the stronger condition

$$\sum_{i=1}^{m} Z_i^2 \leq 2m.$$

So define

$$E_m = \left\{ z \in \mathbb{R}^m : \sum_{i=1}^{m} z_i^2 \leq 2m \right\}.$$

This is the Euclidean ball of radius $\sqrt{2m}$.

7.20 Many Sign Choices Land in the Euclidean Ball

We first show that many sign choices produce points inside E_m.

For each i,

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j.$$

Since the X_j's are independent random signs,

$$\mathbb{E}(X_j) = 0, \qquad \mathbb{E}(X_j^2) = 1.$$

Therefore

$$\mathbb{E}(Y_i^2) = \mathbb{E}\left(\sum_{j=1}^{n} a_{ij} X_j\right)^2.$$

Expanding,

$$\mathbb{E}(Y_i^2) = \sum_{j=1}^{n} a_{ij}^2 \mathbb{E}(X_j^2) + \sum_{j \neq k} a_{ij} a_{ik} \mathbb{E}(X_j X_k).$$

Since X_j and X_k are independent and mean zero when $j \neq k$,

$$\mathbb{E}(X_j X_k) = \mathbb{E}(X_j)\mathbb{E}(X_k) = 0.$$

Also,

$$\mathbb{E}(X_j^2) = 1.$$

Hence

$$\mathbb{E}(Y_i^2) = \sum_{j=1}^{n} a_{ij}^2.$$

Since

$$|a_{ij}| \leq 1,$$

we get

$$\mathbb{E}(Y_i^2) \leq n.$$

Now

$$Z_i = \frac{Y_i}{\sqrt{n}}.$$

Therefore

$$\mathbb{E}(Z_i^2) = \mathbb{E}\left(\frac{Y_i^2}{n}\right) = \frac{1}{n}\mathbb{E}(Y_i^2) \leq 1.$$

Summing over $i = 1, \ldots, m$,

$$\mathbb{E}\left(\sum_{i=1}^{m} Z_i^2\right) = \sum_{i=1}^{m} \mathbb{E}(Z_i^2) \leq m.$$

By Markov's inequality,

$$\mathbb{P}\left\{\sum_{i=1}^{m} Z_i^2 > 2m\right\} \leq \frac{\mathbb{E}\left(\sum_{i=1}^{m} Z_i^2\right)}{2m}.$$

Since

$$\mathbb{E}\left(\sum_{i=1}^{m} Z_i^2\right) \leq m,$$

we obtain

$$\mathbb{P}\left\{\sum_{i=1}^{m} Z_i^2 > 2m\right\} \leq \frac{1}{2}.$$

Therefore

$$\mathbb{P}\left\{\sum_{i=1}^{m} Z_i^2 \leq 2m\right\} \geq \frac{1}{2}.$$

There are 2^n total sign choices. Hence at least

$$2^{n-1}$$

sign choices land inside

$$E_m = \left\{z \in \mathbb{R}^m : \sum_{i=1}^{m} z_i^2 \leq 2m\right\}.$$

7.21 The Covering Problem

We now want to cover

$$E_m = \left\{z \in \mathbb{R}^m : \sum_{i=1}^{m} z_i^2 \leq 2m\right\}$$

by axis-parallel cubes of side length B.

Let

$$N_m(B)$$

denote the number of cubes of side length B needed to cover E_m.

If

$$N_m(B) < 2^{n-1},$$

then since at least 2^{n-1} sign choices land in E_m, two sign choices must land in the same cube.

This is the pigeonhole step.

Thus our goal is to estimate $N_m(B)$.

7.22 Why the Euclidean Ball Improves the Estimate

For the cross-polytope, we used

$$\sum_{i=1}^{m} |z_i| \leq 2m.$$

If

$$|z_i| \geq \frac{B}{2},$$

then that coordinate contributes at least $B/2$ to the ℓ^1-sum.

Therefore the number of large coordinates is at most

$$\frac{2m}{B/2} = \frac{4m}{B}.$$

So for the cross-polytope, the number of large coordinates is bounded by

$$\boxed{\frac{4m}{B}.}$$

For the Euclidean ball, we instead have

$$\sum_{i=1}^{m} z_i^2 \leq 2m.$$

If

$$|z_i| \geq \frac{B}{2},$$

then

$$z_i^2 \geq \frac{B^2}{4}.$$

Therefore the number of large coordinates is at most

$$\frac{2m}{B^2/4} = \frac{8m}{B^2}.$$

Thus for the Euclidean ball, the number of large coordinates is bounded by

$$\boxed{\frac{8m}{B^2}.}$$

This is the key improvement.

The cross-polytope gives a dimension drop of order

$$\frac{m}{B},$$

whereas the Euclidean ball gives a dimension drop of order

$$\frac{m}{B^2}.$$

Thus the dimension drops much faster when using the Euclidean ball.

7.23 Covering Estimate for the Euclidean Ball

We claim that, for $B \geq 2$,

$$\boxed{N_m(B) \leq B^{Cm/B^2}}$$

for some absolute constant $C > 0$.

Equivalently,

$$\boxed{\log N_m(B) \leq C\frac{m}{B^2}\log B.}$$

This should be compared with the cross-polytope estimate

$$\boxed{\log F_m(B) \leq C\frac{m}{B}\log B.}$$

So the Euclidean ball improves

$$\frac{m}{B}$$

to

$$\frac{m}{B^2}.$$

7.24 Proof Idea of the Euclidean Covering Estimate

Cover $\mathbb{R}^m$ by cubes of side length B, centered at lattice points

$$Bk, \qquad k = (k_1, \ldots, k_m) \in \mathbb{Z}^m.$$

So each cube is

$$Q_k = \prod_{i=1}^{m} \left[Bk_i - \frac{B}{2}, Bk_i + \frac{B}{2} \right].$$

We only need to count how many such cubes intersect

$$E_m = \left\{ z \in \mathbb{R}^m : \sum_{i=1}^{m} z_i^2 \leq 2m \right\}.$$

Suppose Q_k intersects E_m. Then there exists

$$z \in E_m \cap Q_k.$$

Thus, for each coordinate,

$$|z_i - Bk_i| \leq \frac{B}{2}.$$

If $k_i \neq 0$, then $|k_i| \geq 1$. Hence

$$|z_i| \geq B|k_i| - \frac{B}{2}.$$

Since $|k_i| \geq 1$,

$$|k_i| - \frac{1}{2} \geq \frac{|k_i|}{2}.$$

Therefore

$$|z_i| \geq \frac{B}{2}|k_i|.$$

Squaring,

$$z_i^2 \geq \frac{B^2}{4} k_i^2.$$

Summing over all coordinates with $k_i \neq 0$, we get

$$\sum_{i=1}^{m} z_i^2 \geq \frac{B^2}{4} \sum_{i=1}^{m} k_i^2.$$

But $z \in E_m$, so

$$\sum_{i=1}^{m} z_i^2 \leq 2m.$$

Therefore

$$\frac{B^2}{4} \sum_{i=1}^{m} k_i^2 \leq 2m.$$

Hence

$$\sum_{i=1}^{m} k_i^2 \leq \frac{8m}{B^2}.$$

So every cube that intersects E_m corresponds to an integer vector

$$k \in \mathbb{Z}^m$$

satisfying

$$\boxed{\sum_{i=1}^{m} k_i^2 \leq \frac{8m}{B^2}.}$$

Therefore $N_m(B)$ is bounded by the number of integer lattice points in the Euclidean ball

$$\left\{ k \in \mathbb{Z}^m : \sum_{i=1}^{m} k_i^2 \leq \frac{8m}{B^2} \right\}.$$

A standard lattice-counting estimate gives

$$\# \left\{ k \in \mathbb{Z}^m : \sum_{i=1}^{m} k_i^2 \leq \frac{8m}{B^2} \right\} \leq B^{Cm/B^2}.$$

Thus

$$\boxed{N_m(B) \leq B^{Cm/B^2}.}$$

This is the desired Euclidean covering estimate.

7.25 Comparison with the Cross-Polytope Estimate

For the cross-polytope, the covering estimate was roughly

$$F_m(B) \leq B^{Cm/B}.$$

Equivalently,

$$\log F_m(B) \leq C\frac{m}{B}\log B.$$

For the Euclidean ball, we get

$$N_m(B) \leq B^{Cm/B^2}.$$

Equivalently,

$$\log N_m(B) \leq C\frac{m}{B^2}\log B.$$

Thus the denominator improves from B to B^2.

This is a real gain.

7.26 Applying the Pigeonhole Principle

We have at least

$$2^{n-1}$$

sign choices inside the Euclidean ball.

If

$$N_m(B) < 2^{n-1},$$

then two sign choices land in the same cube.

It is enough to require

$$B^{Cm/B^2} \leq 2^{n/2}.$$

Taking logarithms,

$$C\frac{m}{B^2}\log B \leq \frac{n}{2}\log 2.$$

Thus we need

$$\frac{m}{B^2}\log B \lesssim n.$$

Equivalently,

$$B^2 \gtrsim \frac{m}{n}\log B.$$

So when $m \geq n$, we may choose

$$\boxed{B \sim \sqrt{\frac{m}{n}\log\left(\frac{2m}{n}\right)}.}$$

In the important case

$$m \leq n,$$

we can take B to be an absolute constant.

7.27 What Happens If Two Sign Choices Land in the Same Cube?

Let

$$X' = (X'_1, \dots, X'_n)$$

and

$$X'' = (X''_1, \dots, X''_n)$$

be two sign choices whose normalized images land in the same cube of side length B. That means for every i,

$$|Z_i(X') - Z_i(X'')| \leq B.$$

Recall

$$Z_i(X) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{ij} X_j.$$

Therefore

$$\left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} a_{ij}(X'_j - X''_j) \right| \leq B.$$

Multiplying by $\sqrt{n}$,

$$\left| \sum_{j=1}^{n} a_{ij}(X'_j - X''_j) \right| \leq B\sqrt{n}.$$

Now define

$$\varepsilon_j = \frac{X'_j - X''_j}{2}.$$

Since

$$X'_j, X''_j \in \{-1, 1\},$$

we have

$$\varepsilon_j \in \{-1, 0, 1\}.$$

Therefore

$$\left| \sum_{j=1}^{n} a_{ij}\varepsilon_j \right| \leq \frac{B\sqrt{n}}{2}.$$

So the partial coloring has error at most

$$\boxed{CB\sqrt{n}.}$$

Using the choice of B, this becomes

$$CB\sqrt{n} \sim C\sqrt{m \log\left(\frac{2m}{n}\right)}.$$

If $m \leq n$, then B is constant, and this becomes

$$\boxed{C\sqrt{n}.}$$

7.28 Comparison of the Two Methods

The cross-polytope method gave a covering number like

$$F_m(B) \leq B^{Cm/B}.$$

To make this less than 2^n, one needs approximately

$$B \gtrsim \frac{m}{n} \log\left(\frac{m}{n}\right).$$

Then the partial coloring error is roughly

$$B\sqrt{n} \sim \frac{m}{n} \log\left(\frac{m}{n}\right) \sqrt{n}.$$

The Euclidean ball gives

$$N_m(B) \leq B^{Cm/B^2}.$$

So it is enough to take

$$B \sim \sqrt{\frac{m}{n} \log\left(\frac{2m}{n}\right)}.$$

Then the error is roughly

$$B\sqrt{n} \sim \sqrt{m \log\left(\frac{2m}{n}\right)}.$$

Thus the Euclidean ball is better because it replaces a linear dependence on m/n by a square-root dependence.

7.29 Main Takeaway

The cross-polytope uses

$$\sum_i |z_i| \leq 2m.$$

Large coordinates cost size B, so there can be about

$$\frac{m}{B}$$

large coordinates.

The Euclidean ball uses

$$\sum_i z_i^2 \leq 2m.$$

Large coordinates cost size B^2, so there can be only about

$$\frac{m}{B^2}$$

large coordinates.

That is why the covering improves from

$$\boxed{B^{Cm/B}}$$

to

$$\boxed{B^{Cm/B^2}}.$$

8 Conditional Probability and Conditional Expectation

8.1 Motivation

Suppose a fair die is rolled. Before knowing anything, the probability of rolling a six is

$$\mathbb{P}(6) = \frac{1}{6}.$$

If someone tells us that the roll was even, then the possible outcomes become

$$\{2, 4, 6\}.$$

Now

$$\mathbb{P}(6 \mid \text{even}) = \frac{1}{3}.$$

The probability changed because our information changed. This is the main idea behind conditional probability.

8.2 Conditional Probability

Let $A, B \subset \Omega$ be events and assume $\mathbb{P}(B) > 0$.

Definition 8.1. The conditional probability of A given B is

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.}$$

Interpretation: once B is known to occur, we treat B as the new universe. We ask what proportion of B also lies inside A.

Example 8.2. Let $A = \{6\}$ and $B = \{2, 4, 6\}$. Then

$$\mathbb{P}(A \cap B) = \frac{1}{6}, \qquad \mathbb{P}(B) = \frac{1}{2}.$$

Therefore

$$\mathbb{P}(A \mid B) = \frac{1/6}{1/2} = \frac{1}{3}.$$

8.3 Conditional Probability Space

Given an event B with $\mathbb{P}(B) > 0$, define

$$\mathbb{P}_B(A) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Then $\mathbb{P}_B$ is a probability measure on the new universe B.
Indeed:

- $\mathbb{P}_B(A) \geq 0$.
- $\mathbb{P}_B(B) = 1$.
- If $A_1, A_2, \ldots$ are disjoint, then

$$\mathbb{P}_B\left(\bigcup_j A_j\right) = \sum_j \mathbb{P}_B(A_j).$$

8.4 Law of Total Probability

Suppose $B_1, \ldots, B_n$ form a partition of Ω. This means

$$\Omega = B_1 \cup \cdots \cup B_n, \qquad B_i \cap B_j = \varnothing \quad (i \neq j).$$

Then for every event A,

$$\boxed{\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j).}$$

Proof. Since the B_j form a partition,

$$A = (A \cap B_1) \cup \cdots \cup (A \cap B_n),$$

and these pieces are disjoint. Hence

$$\mathbb{P}(A) = \sum_j \mathbb{P}(A \cap B_j).$$

But

$$\mathbb{P}(A \cap B_j) = \mathbb{P}(A \mid B_j)\mathbb{P}(B_j).$$

Therefore

$$\mathbb{P}(A) = \sum_j \mathbb{P}(A \mid B_j)\mathbb{P}(B_j).$$

□

8.5 Bayes' Formula

Using

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

and

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A)\mathbb{P}(A),$$

we get

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.}$$

If $A_1, \ldots, A_n$ partition Ω, then

$$\boxed{\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}.}$$

8.6 Conditional Expectation on an Event

Let X be a random variable. We define the average of X on B by

$$\boxed{\mathbb{E}(X \mid B) = \frac{1}{\mathbb{P}(B)} \sum_{\omega \in B} X(\omega)\mathbb{P}(\omega).}$$

Equivalently,

$$\boxed{\mathbb{E}(X \mid B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)}.}$$

Example 8.3. Let $X(i) = i$ on a fair die, and let $B = \{2, 4, 6\}$. Then

$$\mathbb{E}(X\mathbb{1}_B) = 2 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 2.$$

Since $\mathbb{P}(B) = 1/2$,

$$\mathbb{E}(X \mid B) = \frac{2}{1/2} = 4.$$

This is exactly the average of $2, 4, 6$.

8.7 Partition Version

If

$$\Omega = B_1 \cup \cdots \cup B_n$$

is a partition, then define a new random variable by

$$\mathbb{E}(X \mid \mathcal{F}) = \mathbb{E}(X \mid B_j) \quad \text{on } B_j.$$

Thus conditional expectation is constant on each information block.

Key point. Conditional expectation is averaging with respect to the information available. If we only know which block B_j contains the outcome, then the best prediction of X is the average of X over that block.

9 Sigma-Algebras, Filtrations, and Conditional Expectation

9.1 Information as a Partition

Suppose we toss a coin twice:

$$\Omega = \{HH, HT, TH, TT\}.$$

If we only know the first toss, then we cannot distinguish HH from HT, and we cannot distinguish TH from TT. Thus the information gives the partition

$$\{HH, HT\}, \qquad \{TH, TT\}.$$

Each block consists of outcomes that look identical under the information available.

9.2 Sigma-Algebras: Informal View

In finite probability spaces, a sigma-algebra can be viewed as a collection of events whose occurrence can be determined from the information available.

For example, if we know only the first toss, then the event "first toss is heads" is observable, but the event "second toss is heads" is not observable.

9.3 Filtrations

Information usually grows over time.

Definition 9.1. A filtration is an increasing sequence of sigma-algebras

$$\boxed{\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots .}$$

Here $\mathcal{F}_k$ represents the information available at time k.

Example 9.2 (Coin tosses)**.** Before tossing any coin, we know nothing:

$$\mathcal{F}_0 = \{\varnothing, \Omega\}.$$

After the first toss, $\mathcal{F}_1$ contains events determined by the first toss. After the second toss, $\mathcal{F}_2$ contains events determined by the first two tosses. Thus

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2.$$

9.4 Conditional Expectation with Respect to Information

Suppose $\mathcal{F}$ is generated by a partition

$$\Omega = B_1 \cup \cdots \cup B_m.$$

Then

$$\boxed{\mathbb{E}(X \mid \mathcal{F}) = \mathbb{E}(X \mid B_j) \quad \text{on } B_j.}$$

So $\mathbb{E}(X \mid \mathcal{F})$ is a new random variable, constant on each atom B_j.

Example 9.3. Let $\Omega = \{1, 2, 3, 4\}$ with equal probabilities. Define

$$X(1) = 1, \quad X(2) = 3, \quad X(3) = 2, \quad X(4) = 6.$$

Let $B_1 = \{1, 2\}$ and $B_2 = \{3, 4\}$. Then

$$\mathbb{E}(X \mid B_1) = \frac{1+3}{2} = 2,$$

and

$$\mathbb{E}(X \mid B_2) = \frac{2+6}{2} = 4.$$

Therefore

$$\mathbb{E}(X \mid \mathcal{F}) = \begin{cases} 2, & \omega \in B_1, \\ 4, & \omega \in B_2. \end{cases}$$

9.5 Average Preservation

Theorem 9.4. *For any integrable random variable X,*

$$\boxed{\mathbb{E}(\mathbb{E}(X \mid \mathcal{F})) = \mathbb{E}(X).}$$

Proof. Let $\mathcal{F}$ be generated by $B_1, \ldots, B_m$. Then

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{F})) = \sum_j \mathbb{E}(X \mid B_j)\mathbb{P}(B_j).$$

Using the definition,

$$= \sum_j \frac{\mathbb{E}(X\mathbb{1}_{B_j})}{\mathbb{P}(B_j)}\mathbb{P}(B_j) = \sum_j \mathbb{E}(X\mathbb{1}_{B_j}) = \mathbb{E}(X).$$

□

9.6 Tower Property

If $\mathcal{G} \subset \mathcal{F}$, then

$$\boxed{\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}) \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G}).}$$

This says: averaging to a fine level and then averaging again to a coarser level is the same as averaging directly to the coarser level.

9.7 Dyadic Intervals and Dyadic Filtration

On $[0, 1)$, split into dyadic intervals.

Level 0:

$$[0, 1).$$

Level 1:

$$[0, 1/2), \quad [1/2, 1).$$

Level 2:

$$[0, 1/4),\ [1/4, 1/2),\ [1/2, 3/4),\ [3/4, 1).$$

Level k consists of intervals of length 2^{-k}.

Let $\mathcal{F}_k$ be the sigma-algebra generated by dyadic intervals of level k. Then

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots .$$

9.8 Dyadic Conditional Expectation

For $f \in L^1([0, 1])$, define

$$\boxed{f_k = \mathbb{E}(f \mid \mathcal{F}_k).}$$

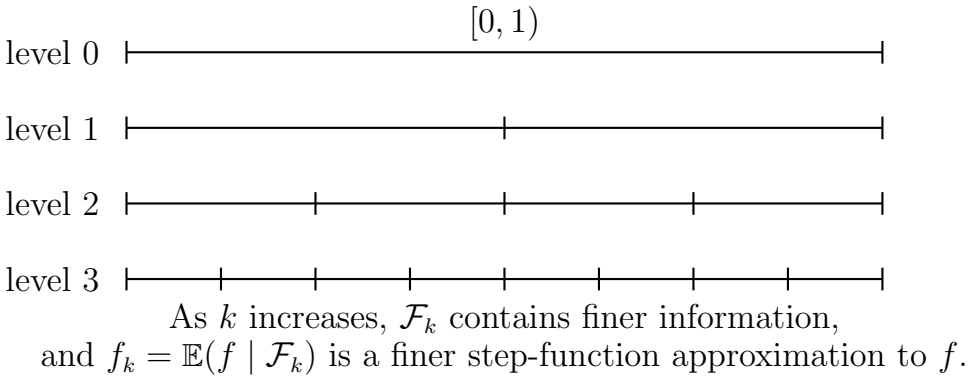


Figure 1: The dyadic filtration on $[0, 1)$ refines the interval into smaller dyadic pieces.

On each dyadic interval I of level k,

$$\boxed{f_k(x) = \frac{1}{|I|}\int_I f(t)\,dt \quad (x \in I).}$$

Thus f_k is the piecewise constant approximation obtained by replacing f on each dyadic interval by its average.

Remember

The sequence $f_k = \mathbb{E}(f \mid \mathcal{F}_k)$ is the central example for the rest of the course. It will become a martingale.

10 Martingales and Dyadic Martingales

10.1 What Is a Martingale?

A martingale is a mathematical model of a fair game. Given all current information, the best prediction of the next value is the present value.

Definition 10.1. Let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ be a filtration. A sequence $X_0, X_1, X_2, \ldots$ is a martingale if X_k is $\mathcal{F}_k$-measurable and

$$\boxed{\mathbb{E}(X_{k+1} \mid \mathcal{F}_k) = X_k}$$

for every k.

Equivalent formulation:

$$\boxed{\mathbb{E}(X_n \mid \mathcal{F}_k) = X_k \qquad (k \leq n).}$$

10.2 Example: Simple Random Walk

Let $\varepsilon_k = \pm 1$ with probability $1/2$, independently. Define

$$S_n = \varepsilon_1 + \cdots + \varepsilon_n.$$

Let $\mathcal{F}_n$ be the information generated by $\varepsilon_1, \ldots, \varepsilon_n$. Then

$$S_{n+1} = S_n + \varepsilon_{n+1}.$$

Taking conditional expectation,

$$\mathbb{E}(S_{n+1} \mid \mathcal{F}_n) = S_n + \mathbb{E}(\varepsilon_{n+1} \mid \mathcal{F}_n) = S_n.$$

Hence S_n is a martingale.

10.3 Martingale Differences

Define

$$\boxed{d_k = X_k - X_{k-1}.}$$

Then

$$X_n = X_0 + \sum_{k=1}^{n} d_k.$$

The martingale property implies

$$\boxed{\mathbb{E}(d_k \mid \mathcal{F}_{k-1}) = 0.}$$

Indeed,

$$\mathbb{E}(d_k \mid \mathcal{F}_{k-1}) = \mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) = X_{k-1} - X_{k-1} = 0.$$

10.4 Constant Expectation

If X_k is a martingale, then

$$\boxed{\mathbb{E}X_n = \mathbb{E}X_0.}$$

This follows by the tower property:

$$\mathbb{E}X_n = \mathbb{E}(\mathbb{E}(X_n \mid \mathcal{F}_{n-1})) = \mathbb{E}X_{n-1} = \cdots = \mathbb{E}X_0.$$

10.5 The Dyadic Martingale

Let $f \in L^1([0,1])$ and define

$$\boxed{f_k = \mathbb{E}(f \mid \mathcal{F}_k).}$$

Then f_k is a martingale:

$$\boxed{\mathbb{E}(f_{k+1} \mid \mathcal{F}_k) = f_k.}$$

Proof. Since $f_{k+1} = \mathbb{E}(f \mid \mathcal{F}_{k+1})$,

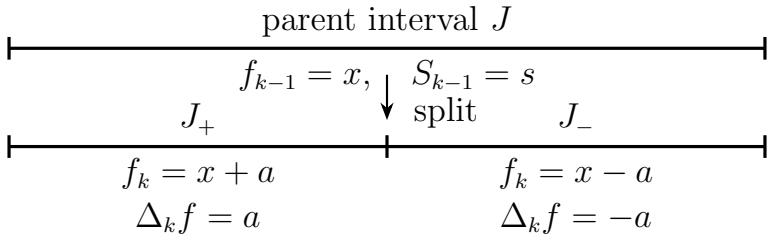
$$\mathbb{E}(f_{k+1} \mid \mathcal{F}_k) = \mathbb{E}(\mathbb{E}(f \mid \mathcal{F}_{k+1}) \mid \mathcal{F}_k).$$

By the tower property,

$$\mathbb{E}(\mathbb{E}(f \mid \mathcal{F}_{k+1}) \mid \mathcal{F}_k) = \mathbb{E}(f \mid \mathcal{F}_k) = f_k.$$

□

10.6 Local Dyadic Picture



On both children, $(\Delta_k f)^2 = a^2$.
Therefore $S_k^2 = s^2 + a^2$ is constant on the whole parent interval J.

Figure 2: The local dyadic martingale picture on a parent interval J.

Let J be a dyadic interval at level $k-1$. On J, both f_{k-1} and S_{k-1} are constant. Write

$$f_{k-1} = x.$$

The interval J splits into two children J_+ and J_-. Since f_k must average to f_{k-1} on J, we can write

$$f_k = x + a \quad \text{on one child,}$$

and

$$f_k = x - a \quad \text{on the other child.}$$

Therefore

$$\Delta_k f = f_k - f_{k-1} = \pm a.$$

This local picture is used repeatedly in the square-function proof.

11 Martingale Differences, Orthogonality, and Square Functions

11.1 Martingale Differences

For the dyadic martingale $f_k = \mathbb{E}(f \mid \mathcal{F}_k)$, define

$$\boxed{\Delta_k f = f_k - f_{k-1}.}$$

Then

$$f_n = f_0 + \sum_{k=1}^{n} \Delta_k f.$$

Moreover,

$$\boxed{\mathbb{E}(\Delta_k f \mid \mathcal{F}_{k-1}) = 0.}$$

11.2 Orthogonality

Theorem 11.1. *If $j \neq k$, then*

$$\boxed{\int \Delta_j f \, \Delta_k f = 0.}$$

Proof. Assume $j < k$. Then $\Delta_j f$ is $\mathcal{F}_j$-measurable, hence also $\mathcal{F}_{k-1}$-measurable. Therefore

$$\int \Delta_j f \, \Delta_k f = \int \Delta_j f \, \mathbb{E}(\Delta_k f \mid \mathcal{F}_{k-1}).$$

But

$$\mathbb{E}(\Delta_k f \mid \mathcal{F}_{k-1}) = 0.$$

Hence the integral is zero. □

11.3 Pythagoras' Formula

Since

$$f_n = f_0 + \sum_{k=1}^{n} \Delta_k f$$

and the terms are orthogonal in L^2,

$$\boxed{\|f_n\|_2^2 = \|f_0\|_2^2 + \sum_{k=1}^{n} \|\Delta_k f\|_2^2.}$$

11.4 Square Function

Definition 11.2. The square function is

$$\boxed{S_k^2 = \sum_{j=1}^{k} (\Delta_j f)^2,}$$

so that

$$S_k = \left(\sum_{j=1}^{k} (\Delta_j f)^2 \right)^{1/2}.$$

The limiting square function is

$$\boxed{S = \left(\sum_{j=1}^{\infty} (\Delta_j f)^2 \right)^{1/2}.}$$

11.5 Local Formula

In the local dyadic model,

$$f_k = x \pm a.$$

Thus

$$\Delta_k f = \pm a,$$

so

$$(\Delta_k f)^2 = a^2.$$

Therefore

$$\boxed{S_k^2 = S_{k-1}^2 + a^2.}$$

This is one of the most important computational identities in the course.

11.6 The Fundamental L^2 Identity

By definition,

$$\int S_k^2 = \sum_{j=1}^{k} \int (\Delta_j f)^2.$$

Using orthogonality,

$$\sum_{j=1}^{k} \int (\Delta_j f)^2 = \int f_k^2 - \int f_0^2.$$

Thus

$$\boxed{\int f_k^2 = \int f_0^2 + \int S_k^2.}$$

Letting $k \to \infty$,

$$\boxed{\int f^2 = \int f_0^2 + \int S^2.}$$

If $f_0 = 0$, then

$$\boxed{\int f^2 = \int S^2.}$$

11.7 The Question for $p \neq 2$

For $p = 2$, orthogonality gives equality. For $p \neq 2$, we want constants $c_p, C_p > 0$ such that

$$\boxed{c_p \|f\|_p \leq \|S(f)\|_p \leq C_p \|f\|_p}$$

for $1 < p < \infty$.

This is the dyadic Littlewood–Paley theorem.

12 Supermartingales and the Burkholder Method for $p > 2$

12.1 The Goal

Let f be a periodic function and let

$$f_k = \mathbb{E}(f \mid \mathcal{F}_k)$$

be the dyadic martingale generated by f. Define the martingale differences by

$$\Delta_k f = f_k - f_{k-1}.$$

The dyadic square function is defined by

$$S_k^2 = S_{k-1}^2 + (\Delta_k f)^2,$$

with the convention

$$S_0 = |f_0|.$$

Thus

$$S_k^2 = f_0^2 + \sum_{j=1}^{k} (\Delta_j f)^2.$$

Let

$$S = \lim_{k \to \infty} S_k.$$

The claim is that for $1 < p < \infty$,

$$\int_I |f|^p$$

and

$$\int_I S^p$$

are comparable. In this note we prove the upper square-function estimate for $p > 2$:

$$\boxed{\int_I S^p \leq C_p \int_I |f|^p.}$$

Equivalently,

$$\boxed{\|S\|_p \leq C_p \|f\|_p.}$$

12.2 Supermartingales

Definition 12.1. A sequence X_k is a supermartingale with respect to $\mathcal{F}_k$ if

$$\boxed{\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) \leq X_{k-1}.}$$

Then

$$\mathbb{E}X_k \leq \mathbb{E}X_{k-1} \leq \cdots \leq \mathbb{E}X_0.$$

12.3 The Case $p > 2$

For $p > 2$, we no longer have such an exact identity. Instead, we prove an inequality by constructing a supermartingale.

Fix $p > 2$. We define

$$X_k = S_k^p - AS_k^{p-2} f_k^2,$$

where $A > 0$ is a large constant to be chosen later.

The goal is to prove

$$\boxed{\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) \leq X_{k-1}.}$$

This means that (X_k) is a supermartingale.

Once we know this, taking expectations will give

$$\mathbb{E}X_k \leq \mathbb{E}X_0.$$

This will eventually imply the desired estimate

$$\int S^p \leq C_p \int |f|^p.$$

12.4 The Local Dyadic Picture

Fix one dyadic interval J at level $k-1$.

On J, both f_{k-1} and S_{k-1} are constant. Write

$$f_{k-1} = x, \qquad S_{k-1} = s.$$

The interval J splits into two children:

$$J_+ \quad \text{and} \quad J_-.$$

Since f_k has conditional expectation f_{k-1} on J, the two child values average to x. Therefore we may write

$$f_k = x + a \quad \text{on } J_+,$$

and

$$f_k = x - a \quad \text{on } J_-.$$

Hence

$$\Delta_k f = a \quad \text{on } J_+,$$

and

$$\Delta_k f = -a \quad \text{on } J_-.$$

Therefore

$$(\Delta_k f)^2 = a^2$$

on both children.

Thus

$$S_k^2 = S_{k-1}^2 + (\Delta_k f)^2 = s^2 + a^2$$

on both children.

So S_k is constant on the whole parent interval J. Let

$$r = S_k = \sqrt{s^2 + a^2}.$$

This is the important dyadic "miracle":

$$\boxed{S_k \text{ is constant on the parent interval } J.}$$

12.5 Estimate the S_k^p Part

We compare

$$S_k^p \quad \text{and} \quad S_{k-1}^p.$$

In the local notation,

$$S_k = r, \qquad S_{k-1} = s,$$

where

$$r^2 = s^2 + a^2.$$

Therefore

$$S_k^p - S_{k-1}^p = r^p - s^p = (s^2 + a^2)^{p/2} - s^p.$$

Let

$$\phi(u) = u^{p/2}.$$

Since $p > 2$, we have $p/2 > 1$, and ϕ is increasing and convex.
By the mean value theorem,

$$(s^2 + a^2)^{p/2} - s^p = \phi(s^2 + a^2) - \phi(s^2)$$

is bounded by

$$\phi'(s^2 + a^2)a^2.$$

Since

$$\phi'(u) = \frac{p}{2}u^{p/2-1},$$

we get

$$(s^2 + a^2)^{p/2} - s^p \leq \frac{p}{2}(s^2 + a^2)^{p/2-1}a^2.$$

But

$$(s^2 + a^2)^{1/2} = r.$$

Thus

$$(s^2 + a^2)^{p/2-1} = r^{p-2}.$$

Therefore

$$r^p - s^p \leq \frac{p}{2}r^{p-2}a^2.$$

In martingale notation, this becomes

$$\boxed{S_k^p - S_{k-1}^p \leq \frac{p}{2}S_k^{p-2}(\Delta_k f)^2.}$$

12.6 Estimate the $S_k^{p-2}f_k^2$ Part

Now we compare

$$S_k^{p-2}f_k^2$$

with

$$S_{k-1}^{p-2}f_{k-1}^2.$$

Recall that on the parent interval J,

$$S_k = r, \qquad S_{k-1} = s, \qquad f_{k-1} = x.$$

Also,

$$f_k = x + a \quad \text{on } J_+,$$

and

$$f_k = x - a \quad \text{on } J_-.$$

Therefore

$$\mathbb{E}(f_k^2 \mid \mathcal{F}_{k-1}) = \frac{1}{2}(x+a)^2 + \frac{1}{2}(x-a)^2.$$

Expanding,

$$\frac{1}{2}(x+a)^2 + \frac{1}{2}(x-a)^2 = \frac{1}{2}(x^2+2ax+a^2) + \frac{1}{2}(x^2-2ax+a^2).$$

The cross terms cancel, so

$$\mathbb{E}(f_k^2 \mid \mathcal{F}_{k-1}) = x^2 + a^2.$$

Because $S_k = r$ is constant on the parent interval,

$$\mathbb{E}(S_k^{p-2} f_k^2 \mid \mathcal{F}_{k-1}) = r^{p-2}(x^2+a^2).$$

Therefore

$$\mathbb{E}(S_k^{p-2} f_k^2 \mid \mathcal{F}_{k-1}) - S_{k-1}^{p-2} f_{k-1}^2$$

equals

$$r^{p-2}(x^2+a^2) - s^{p-2}x^2.$$

Rewrite this as

$$r^{p-2}a^2 + (r^{p-2} - s^{p-2})x^2.$$

Since $r \geq s$ and $p > 2$, we have

$$r^{p-2} - s^{p-2} \geq 0.$$

Therefore

$$r^{p-2}a^2 + (r^{p-2} - s^{p-2})x^2 \geq r^{p-2}a^2.$$

Thus

$$\boxed{\mathbb{E}(S_k^{p-2} f_k^2 \mid \mathcal{F}_{k-1}) - S_{k-1}^{p-2} f_{k-1}^2 \geq S_k^{p-2}(\Delta_k f)^2.}$$

12.7 Prove the Supermartingale Property

Recall

$$X_k = S_k^p - A S_k^{p-2} f_k^2.$$

Then

$$X_k - X_{k-1} = (S_k^p - S_{k-1}^p) - A\left(S_k^{p-2} f_k^2 - S_{k-1}^{p-2} f_{k-1}^2\right).$$

Take conditional expectation with respect to $\mathcal{F}_{k-1}$.

Using the first estimate,

$$\mathbb{E}(S_k^p - S_{k-1}^p \mid \mathcal{F}_{k-1}) \leq \frac{p}{2} S_k^{p-2} (\Delta_k f)^2.$$

Using the second estimate,

$$\mathbb{E}\left(S_k^{p-2} f_k^2 - S_{k-1}^{p-2} f_{k-1}^2 \mid \mathcal{F}_{k-1}\right) \geq S_k^{p-2} (\Delta_k f)^2.$$

Therefore

$$\mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \leq \frac{p}{2} S_k^{p-2} (\Delta_k f)^2 - A S_k^{p-2} (\Delta_k f)^2.$$

Hence

$$\mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \leq \left(\frac{p}{2} - A\right) S_k^{p-2} (\Delta_k f)^2.$$

If

$$A \geq \frac{p}{2},$$

then

$$\mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \leq 0.$$

Therefore

$$\boxed{\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) \leq X_{k-1}.}$$

Thus (X_k) is a supermartingale.

12.8 Take Expectations

Since X_k is a supermartingale,

$$\mathbb{E}X_k \leq \mathbb{E}X_0.$$

Now

$$X_0 = S_0^p - A S_0^{p-2} f_0^2.$$

Since

$$S_0 = |f_0|,$$

we have

$$X_0 = |f_0|^p - A|f_0|^{p-2} f_0^2.$$

But

$$|f_0|^{p-2} f_0^2 = |f_0|^p.$$

Therefore

$$X_0 = (1 - A)|f_0|^p.$$

If $A \geq 1$, then

$$X_0 \leq 0.$$

Thus

$$\mathbb{E}X_k \leq 0.$$

Hence

$$\mathbb{E}\left(S_k^p - AS_k^{p-2}f_k^2\right) \leq 0.$$

Therefore

$$\mathbb{E}S_k^p \leq A\mathbb{E}(S_k^{p-2}f_k^2).$$

In integral form,

$$\boxed{\int S_k^p \leq A\int S_k^{p-2}f_k^2.}$$

12.9 Apply Hölder's Inequality

We estimate

$$\int S_k^{p-2}f_k^2.$$

Use Hölder's inequality with conjugate exponents

$$\frac{p}{p-2} \quad \text{and} \quad \frac{p}{2}.$$

Indeed,

$$\frac{p-2}{p} + \frac{2}{p} = 1.$$

We write

$$S_k^{p-2} = (S_k^p)^{(p-2)/p}$$

and

$$f_k^2 = (|f_k|^p)^{2/p}.$$

Therefore Hölder gives

$$\int S_k^{p-2}f_k^2 \leq \left(\int S_k^p\right)^{(p-2)/p}\left(\int |f_k|^p\right)^{2/p}.$$

Using the previous inequality,

$$\int S_k^p \leq A\left(\int S_k^p\right)^{(p-2)/p}\left(\int |f_k|^p\right)^{2/p}.$$

Assume first that

$$\int S_k^p \neq 0.$$

Divide both sides by

$$\left(\int S_k^p\right)^{(p-2)/p}.$$

Then

$$\left(\int S_k^p\right)^{2/p} \leq A\left(\int |f_k|^p\right)^{2/p}.$$

Raise both sides to the power $p/2$:

$$\int S_k^p \leq A^{p/2}\int |f_k|^p.$$

If $\int S_k^p = 0$, the inequality is trivial.

Therefore

$$\boxed{\int S_k^p \leq A^{p/2}\int |f_k|^p.}$$

12.10 Use Jensen's Inequality

Recall

$$f_k = \mathbb{E}(f \mid \mathcal{F}_k).$$

Since the function

$$x \mapsto |x|^p$$

is convex for $p > 1$, Jensen's inequality gives

$$|f_k|^p = |\mathbb{E}(f \mid \mathcal{F}_k)|^p \leq \mathbb{E}(|f|^p \mid \mathcal{F}_k).$$

Integrating,

$$\int |f_k|^p \leq \int |f|^p.$$

Hence

$$\int S_k^p \leq A^{p/2}\int |f|^p.$$

Thus

$$\boxed{\int S_k^p \leq A^{p/2}\int |f|^p.}$$

12.11 Let $k \to \infty$

Since

$$S_k^2 = f_0^2 + \sum_{j=1}^{k} (\Delta_j f)^2,$$

the sequence S_k is increasing pointwise:

$$S_k \leq S_{k+1}.$$

Also,

$$S_k \to S.$$

Therefore

$$S_k^p \to S^p$$

and S_k^p increases pointwise to S^p.

By the monotone convergence theorem,

$$\int S^p = \lim_{k\to\infty} \int S_k^p.$$

Since

$$\int S_k^p \leq A^{p/2} \int |f|^p$$

for every k, we get

$$\boxed{\int S^p \leq A^{p/2} \int |f|^p.}$$

Choosing

$$A = \frac{p}{2}$$

gives

$$\boxed{\int S^p \leq \left(\frac{p}{2}\right)^{p/2} \int |f|^p.}$$

Thus, for $p > 2$,

$$\boxed{\|S\|_p \leq C_p \|f\|_p.}$$

12.12 What the Miracle Means

The key special feature of the dyadic martingale is this:

On a parent interval J,

$$\Delta_k f = +a$$

on one child and

$$\Delta_k f = -a$$

on the other child.

Therefore

$$(\Delta_k f)^2 = a^2$$

on both children.

Hence

$$S_k^2 = S_{k-1}^2 + (\Delta_k f)^2 = s^2 + a^2$$

is constant on the whole parent interval J.

This is why, in conditional expectations,

$$S_k^{p-2}$$

can be treated like a constant on the parent interval.

That is the "miracle" in the proof.

Without this dyadic symmetry, the computation is much less direct.

12.13 Conclusion

For $p > 2$, by constructing the supermartingale

$$X_k = S_k^p - AS_k^{p-2} f_k^2,$$

with

$$A \geq \frac{p}{2},$$

we obtain

$$\int S^p \leq C_p \int |f|^p.$$

Equivalently,

$$\boxed{\|S\|_p \leq C_p \|f\|_p.}$$

This proves the upper square-function estimate for $p > 2$.

13 Burkholder Method for $1 < p < 2$

13.1 Statement of the Result

Let f be a periodic function and let

$$f_k = \mathbb{E}(f \mid \mathcal{F}_k)$$

be the dyadic martingale generated by f. Define the martingale differences by

$$\Delta_k f = f_k - f_{k-1}.$$

The dyadic square function is defined by

$$S_k^2 = S_{k-1}^2 + (\Delta_k f)^2,$$

with

$$S_0 = |f_0|.$$

Equivalently,

$$S_k^2 = f_0^2 + \sum_{j=1}^{k} (\Delta_j f)^2.$$

Let

$$S = \lim_{k \to \infty} S_k.$$

The goal is to prove, for $1 < p < 2$,

$$\boxed{\int S^p \leq C_p \int |f|^p.}$$

Equivalently,

$$\boxed{\|S\|_p \leq C_p \|f\|_p.}$$

In the argument below, we first assume $f \geq 0$, which is the setting used in the class notes.

13.2 Why the $p > 2$ Argument Changes

For $p > 2$, one uses the supermartingale

$$S_k^p - A S_k^{p-2} f_k^2.$$

But when

$$1 < p < 2,$$

the exponent $p - 2$ is negative. Hence

$$S_k^{p-2}$$

can become singular near $S_k = 0$.

So for $1 < p < 2$, we instead use the modified expression

$$\boxed{X_k = (S_k^2 + f_k^2)^{p/2} - A f_k^p.}$$

The goal is to prove that (X_k) is a supermartingale for A sufficiently large.

That is, we want

$$\boxed{\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) \leq X_{k-1}.}$$

13.3 The Local Dyadic Picture

Fix a dyadic interval J at level $k - 1$.

On J, the functions f_{k-1} and S_{k-1} are constant. Write

$$f_{k-1} = x, \qquad S_{k-1} = s.$$

The interval J splits into two children, J_+ and J_-. Since f_k averages to f_{k-1} on J, the two child values may be written as

$$f_k = x + a \quad \text{on } J_+,$$

and

$$f_k = x - a \quad \text{on } J_-.$$

Thus

$$\Delta_k f = a \quad \text{on } J_+,$$

and

$$\Delta_k f = -a \quad \text{on } J_-.$$

Therefore

$$(\Delta_k f)^2 = a^2$$

on both children.

Hence

$$S_k^2 = S_{k-1}^2 + (\Delta_k f)^2 = s^2 + a^2$$

on both children.

So S_k is constant on the parent interval J.

Since we are assuming $f \geq 0$, the child values satisfy

$$x + a \geq 0, \qquad x - a \geq 0.$$

Therefore

$$|a| \leq x.$$

This fact will be used when applying Taylor estimates to f_k^p.

13.4 The Supermartingale

Define

$$X_k = (S_k^2 + f_k^2)^{p/2} - A f_k^p,$$

where $A > 0$ will be chosen later.

We want to prove

$$\mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \leq 0.$$

In local notation,

$$X_{k-1} = (s^2 + x^2)^{p/2} - A x^p.$$

On the two children,

$$X_k = \left(s^2 + a^2 + (x+a)^2\right)^{p/2} - A(x+a)^p,$$

and

$$X_k = \left(s^2 + a^2 + (x-a)^2\right)^{p/2} - A(x-a)^p.$$

We estimate the two parts separately.

13.5 Estimate the $(S_k^2 + f_k^2)^{p/2}$ Part

We first estimate

$$(S_k^2 + f_k^2)^{p/2} - (S_{k-1}^2 + f_{k-1}^2)^{p/2}.$$

Let

$$B = s^2 + x^2.$$

On the two children,

$$S_k^2 + f_k^2 = s^2 + a^2 + (x+a)^2$$

and

$$S_k^2 + f_k^2 = s^2 + a^2 + (x-a)^2.$$

Compute:

$$s^2 + a^2 + (x+a)^2 = s^2 + x^2 + 2a^2 + 2xa = B + 2a^2 + 2xa,$$

and

$$s^2 + a^2 + (x-a)^2 = s^2 + x^2 + 2a^2 - 2xa = B + 2a^2 - 2xa.$$

Let

$$\phi(u) = u^{p/2}.$$

Since

$$1 < p < 2,$$

we have

$$0 < \frac{p}{2} < 1.$$

Therefore ϕ is concave.

For a concave function,

$$\phi(B+h) - \phi(B) \leq \phi'(B)h.$$

Also,

$$\phi'(u) = \frac{p}{2}u^{p/2-1}.$$

Therefore,

$$\phi(B + 2a^2 + 2xa) - \phi(B) \leq \frac{p}{2}B^{p/2-1}(2a^2 + 2xa),$$

and

$$\phi(B + 2a^2 - 2xa) - \phi(B) \leq \frac{p}{2}B^{p/2-1}(2a^2 - 2xa).$$

Averaging over the two children gives

$$\mathbb{E}\left[(S_k^2 + f_k^2)^{p/2} - (S_{k-1}^2 + f_{k-1}^2)^{p/2} \mid \mathcal{F}_{k-1}\right]$$

$$\leq \frac{1}{2} \cdot \frac{p}{2}B^{p/2-1}(2a^2 + 2xa) + \frac{1}{2} \cdot \frac{p}{2}B^{p/2-1}(2a^2 - 2xa).$$

The $2xa$ terms cancel, so

$$\mathbb{E}\left[(S_k^2 + f_k^2)^{p/2} - (S_{k-1}^2 + f_{k-1}^2)^{p/2} \mid \mathcal{F}_{k-1}\right] \leq pB^{p/2-1}a^2.$$

Since

$$B = s^2 + x^2 \geq x^2$$

and

$$\frac{p}{2} - 1 < 0,$$

we have

$$B^{p/2-1} \leq (x^2)^{p/2-1} = x^{p-2}.$$

Thus

$$\boxed{\mathbb{E}\left[(S_k^2 + f_k^2)^{p/2} - (S_{k-1}^2 + f_{k-1}^2)^{p/2} \mid \mathcal{F}_{k-1}\right] \leq px^{p-2}a^2.}$$

In martingale notation, this is

$$\boxed{\mathbb{E}\left[(S_k^2 + f_k^2)^{p/2} - (S_{k-1}^2 + f_{k-1}^2)^{p/2} \mid \mathcal{F}_{k-1}\right] \leq pf_{k-1}^{p-2}(\Delta_k f)^2.}$$

13.6 A Taylor Estimate for f_k^p

We now estimate

$$\mathbb{E}(f_k^p - f_{k-1}^p \mid \mathcal{F}_{k-1}).$$

Locally this is

$$\frac{1}{2}(x+a)^p + \frac{1}{2}(x-a)^p - x^p.$$

Because $f \geq 0$, we have

$$|a| \leq x.$$

Set

$$u = \frac{a}{x}.$$

Then

$$|u| \leq 1.$$

We need the following elementary inequality.

Lemma 13.1. *Let $1 < p < 2$. There exists a constant $c_p > 0$ such that for every $-1 \leq z \leq 1$,*

$$\boxed{(1+z)^p - 1 \geq pz + c_p z^2.}$$

One may take, for example,

$$c_p = \frac{p-1}{4}.$$

Proof. Let

$$g(z) = (1+z)^p.$$

Then

$$g'(z) = p(1+z)^{p-1},$$

and

$$g''(z) = p(p-1)(1+z)^{p-2}.$$

Since $1 < p < 2$, the exponent $p-2$ is negative. On the interval $[-1, 1]$, the quantity $(1+z)^{p-2}$ is bounded below by a positive constant away from the endpoint issue in the Taylor estimate, and one obtains a positive quadratic lower bound.

Thus there exists $c_p > 0$ such that

$$(1+z)^p \geq 1 + pz + c_p z^2$$

for $-1 \leq z \leq 1$.

A rough usable choice is

$$c_p = \frac{p-1}{4}.$$

□

Applying the lemma to $z = u$, we get

$$(1+u)^p - 1 \geq pu + c_p u^2.$$

Applying it to $z = -u$, we get

$$(1-u)^p - 1 \geq -pu + c_p u^2.$$

Adding the two inequalities and dividing by 2,

$$\frac{1}{2}(1+u)^p + \frac{1}{2}(1-u)^p - 1 \geq c_p u^2.$$

Multiplying by x^p, we obtain

$$\frac{1}{2}(x+a)^p + \frac{1}{2}(x-a)^p - x^p \geq c_p x^p \left(\frac{a}{x}\right)^2.$$

Therefore

$$\frac{1}{2}(x+a)^p + \frac{1}{2}(x-a)^p - x^p \geq c_p x^{p-2} a^2.$$

Hence

$$\boxed{\mathbb{E}(f_k^p - f_{k-1}^p \mid \mathcal{F}_{k-1}) \geq c_p f_{k-1}^{p-2} (\Delta_k f)^2.}$$

Using the rough constant from the notes, we may take

$$c_p = \frac{p-1}{4}.$$

13.7 Combine the Two Estimates

Recall

$$X_k = (S_k^2 + f_k^2)^{p/2} - Af_k^p.$$

Then

$$X_k - X_{k-1} = \left[(S_k^2 + f_k^2)^{p/2} - (S_{k-1}^2 + f_{k-1}^2)^{p/2}\right] - A(f_k^p - f_{k-1}^p).$$

Taking conditional expectation and using the estimates above,

$$\mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1})$$

$$\leq pf_{k-1}^{p-2}(\Delta_k f)^2 - Ac_p f_{k-1}^{p-2}(\Delta_k f)^2.$$

Therefore

$$\mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \leq (p - Ac_p)f_{k-1}^{p-2}(\Delta_k f)^2.$$

If

$$A \geq \frac{p}{c_p},$$

then

$$\mathbb{E}(X_k - X_{k-1} \mid \mathcal{F}_{k-1}) \leq 0.$$

Thus

$$\boxed{\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) \leq X_{k-1}.}$$

So X_k is a supermartingale.

Using the rough value

$$c_p = \frac{p-1}{4},$$

it is enough to take

$$\boxed{A \geq \frac{4p}{p-1}.}$$

13.8 Initial Value

Now compute X_0.

Since

$$S_0 = |f_0|,$$

and in this part $f \geq 0$, we have

$$S_0 = f_0.$$

Thus

$$X_0 = (S_0^2 + f_0^2)^{p/2} - Af_0^p.$$

Since

$$S_0^2 + f_0^2 = 2f_0^2,$$

we get

$$X_0 = (2f_0^2)^{p/2} - Af_0^p.$$

Hence

$$X_0 = 2^{p/2} f_0^p - Af_0^p.$$

Therefore

$$X_0 = (2^{p/2} - A)f_0^p.$$

If

$$A \geq 2^{p/2},$$

then

$$X_0 \leq 0.$$

So choose

$$\boxed{A \geq \max\left\{2^{p/2}, \frac{4p}{p-1}\right\}.}$$

Then X_k is a supermartingale and $X_0 \leq 0$.

13.9 Take Expectations

Since X_k is a supermartingale,

$$\mathbb{E}X_k \leq \mathbb{E}X_0.$$

But

$$X_0 \leq 0.$$

Therefore

$$\mathbb{E}X_k \leq 0.$$

That is,

$$\mathbb{E}\left[(S_k^2 + f_k^2)^{p/2} - Af_k^p\right] \leq 0.$$

Hence

$$\mathbb{E}(S_k^2 + f_k^2)^{p/2} \leq A\mathbb{E}f_k^p.$$

In integral notation,

$$\boxed{\int (S_k^2 + f_k^2)^{p/2} \leq A \int f_k^p.}$$

13.10 Deduce the Square-Function Estimate

Since

$$S_k^2 \leq S_k^2 + f_k^2,$$

we have

$$S_k^p \leq (S_k^2 + f_k^2)^{p/2}.$$

Therefore

$$\int S_k^p \leq \int (S_k^2 + f_k^2)^{p/2}.$$

Using the previous estimate,

$$\int S_k^p \leq A \int f_k^p.$$

Now

$$f_k = \mathbb{E}(f \mid \mathcal{F}_k).$$

Since $x \mapsto |x|^p$ is convex for $p > 1$, Jensen's inequality gives

$$|f_k|^p = |\mathbb{E}(f \mid \mathcal{F}_k)|^p \leq \mathbb{E}(|f|^p \mid \mathcal{F}_k).$$

Integrating,

$$\int |f_k|^p \leq \int |f|^p.$$

Therefore

$$\boxed{\int S_k^p \leq A \int |f|^p.}$$

13.11 Let $k \to \infty$

Since

$$S_k^2 = f_0^2 + \sum_{j=1}^{k} (\Delta_j f)^2,$$

we have

$$S_k \leq S_{k+1}.$$

Also,

$$S_k \to S.$$

Thus

$$S_k^p \to S^p$$

monotonically.

By the monotone convergence theorem,

$$\int S^p = \lim_{k\to\infty} \int S_k^p.$$

Since

$$\int S_k^p \le A \int |f|^p$$

for every k, we obtain

$$\boxed{\int S^p \le A \int |f|^p.}$$

Thus, for $1 < p < 2$,

$$\boxed{\|S\|_p \le C_p \|f\|_p.}$$

13.12 Main Idea of the Proof

For $p > 2$, the useful supermartingale was

$$S_k^p - A S_k^{p-2} f_k^2.$$

For $1 < p < 2$, this is not good because S_k^{p-2} has a negative exponent.

So instead we use

$$\boxed{X_k = (S_k^2 + f_k^2)^{p/2} - A f_k^p.}$$

The proof works because the term

$$(S_k^2 + f_k^2)^{p/2}$$

increases slowly, while the term

$$f_k^p$$

increases enough on average to dominate it.

The key Taylor estimate is

$$\boxed{(1+z)^p - 1 \ge pz + c_p z^2, \qquad |z| \le 1.}$$

This is the main inequality in the $1 < p < 2$ case.

13.13 Conclusion

For $1 < p < 2$, choosing

$$A \geq \max\left\{2^{p/2}, \frac{4p}{p-1}\right\},$$

the process

$$X_k = (S_k^2 + f_k^2)^{p/2} - Af_k^p$$

is a supermartingale with $X_0 \leq 0$.

Therefore

$$\int S^p \leq C_p \int |f|^p.$$

Equivalently,

$$\boxed{\|S\|_p \leq C_p \|f\|_p.}$$

13.14 Summary of the Two Upper-Bound Cases

For $p > 2$ use

$$X_k = S_k^p - AS_k^{p-2} f_k^2.$$

For $1 < p < 2$ use

$$X_k = (S_k^2 + f_k^2)^{p/2} - A|f_k|^p.$$

Both are designed so that

$$\mathbb{E}(X_k \mid \mathcal{F}_{k-1}) \leq X_{k-1}.$$

Both imply

$$\boxed{\|S(f)\|_p \leq C_p \|f\|_p.}$$

14 Reverse Inequality and Completion of Littlewood–Paley

14.1 The Theorem

Let $1 < p < \infty$, and let

$$q = \frac{p}{p-1}$$

be the conjugate exponent, so that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Let $f_k = \mathbb{E}(f \mid \mathcal{F}_k)$ be the dyadic martingale generated by f, and define the martingale differences by

$$\Delta_k f = f_k - f_{k-1}.$$

The square function associated to f is

$$S_f^2 = f_0^2 + \sum_{k \geq 1} (\Delta_k f)^2.$$

Previously, we proved the upper bound

$$\boxed{\|S_f\|_p \leq C_p \|f\|_p.}$$

Now we prove the lower bound

$$\boxed{\|f\|_p \leq C_p \|S_f\|_p.}$$

Equivalently,

$$\boxed{\int |f|^p \leq C_p \int S_f^p.}$$

Together,

$$\boxed{\|f\|_p \asymp_p \|S(f)\|_p}$$

for $1 < p < \infty$.

This proof uses L^p-duality.

14.2 L^p-L^q Duality

We use the following standard fact.

Theorem 14.1 (L^p-duality)**.** *Let $1 < p < \infty$, and let q be the conjugate exponent:*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then

$$\boxed{\|f\|_p = \sup_{\|g\|_q \leq 1} \left| \int fg \right|.}$$

Proof. First, by Hölder's inequality,

$$\left| \int fg \right| \leq \left(\int |f|^p \right)^{1/p} \left(\int |g|^q \right)^{1/q}.$$

Thus

$$\left| \int fg \right| \leq \|f\|_p \|g\|_q.$$

If $\|g\|_q \leq 1$, then

$$\left| \int fg \right| \leq \|f\|_p.$$

Therefore

$$\sup_{\|g\|_q \leq 1} \left| \int fg \right| \leq \|f\|_p.$$

For the reverse inequality, assume $f \neq 0$. Define

$$g = \frac{|f|^{p-1} \operatorname{sgn}(f)}{\left(\int |f|^p \right)^{1/q}}.$$

Then

$$|g|^q = \frac{|f|^{(p-1)q}}{\int |f|^p}.$$

Since

$$(p-1)q = p,$$

we have

$$|g|^q = \frac{|f|^p}{\int |f|^p}.$$

Hence

$$\int |g|^q = 1.$$

So

$$\|g\|_q = 1.$$

Also,

$$\int fg = \frac{\int |f|^p}{\left(\int |f|^p\right)^{1/q}}.$$

Therefore

$$\int fg = \left(\int |f|^p\right)^{1-1/q}.$$

Since

$$1 - \frac{1}{q} = \frac{1}{p},$$

we get

$$\int fg = \left(\int |f|^p\right)^{1/p} = \|f\|_p.$$

Thus

$$\|f\|_p \leq \sup_{\|g\|_q \leq 1} \left|\int fg\right|.$$

Combining both inequalities,

$$\boxed{\|f\|_p = \sup_{\|g\|_q \leq 1} \left|\int fg\right|.}$$

This proves the duality formula. □

14.3 Martingale Expansions

Let

$$f_k = \mathbb{E}(f \mid \mathcal{F}_k), \qquad g_k = \mathbb{E}(g \mid \mathcal{F}_k).$$

Define

$$\Delta_k f = f_k - f_{k-1}, \qquad \Delta_k g = g_k - g_{k-1}.$$

Then

$$f_k = f_0 + \sum_{i=1}^{k} \Delta_i f,$$

and

$$g_k = g_0 + \sum_{i=1}^{k} \Delta_i g.$$

Passing to the limit, formally,

$$f = f_0 + \sum_{i\geq 1} \Delta_i f,$$

and

$$g = g_0 + \sum_{i\geq 1} \Delta_i g.$$

The martingale differences are orthogonal in L^2. Therefore,

$$\int fg = \int f_0 g_0 + \sum_{i\geq 1} \int \Delta_i f \, \Delta_i g.$$

So we have the identity

$$\boxed{\int fg = \int f_0 g_0 + \sum_{i\geq 1} \int \Delta_i f \, \Delta_i g.}$$

This is the martingale analogue of Parseval's identity.

14.4 Pointwise Cauchy–Schwarz

At each point x, consider the two sequences

$$\big(f_0(x), \Delta_1 f(x), \Delta_2 f(x), \dots\big)$$

and

$$\big(g_0(x), \Delta_1 g(x), \Delta_2 g(x), \dots\big).$$

By the Cauchy–Schwarz inequality,

$$\left| f_0 g_0 + \sum_{i\geq 1} \Delta_i f \, \Delta_i g \right| \leq \left(f_0^2 + \sum_{i\geq 1} (\Delta_i f)^2 \right)^{1/2} \left(g_0^2 + \sum_{i\geq 1} (\Delta_i g)^2 \right)^{1/2}.$$

But

$$S_f^2 = f_0^2 + \sum_{i\geq 1} (\Delta_i f)^2,$$

and

$$S_g^2 = g_0^2 + \sum_{i\geq 1}(\Delta_i g)^2.$$

Hence

$$\left| f_0 g_0 + \sum_{i\geq 1} \Delta_i f \, \Delta_i g \right| \leq S_f S_g.$$

Integrating, we get

$$\boxed{\left| \int fg \right| \leq \int S_f S_g.}$$

This is the key estimate.

14.5 Apply Hölder's Inequality

Now apply Hölder's inequality to

$$\int S_f S_g.$$

Since $1/p + 1/q = 1$,

$$\int S_f S_g \leq \left(\int S_f^p \right)^{1/p} \left(\int S_g^q \right)^{1/q}.$$

Thus

$$\left| \int fg \right| \leq \|S_f\|_p \|S_g\|_q.$$

Now we use the upper square-function inequality, already proved for every exponent $1 < r < \infty$:

$$\|S_h\|_r \leq C_r \|h\|_r.$$

Apply this with $h = g$ and $r = q$. Then

$$\|S_g\|_q \leq C_q \|g\|_q.$$

Therefore

$$\left| \int fg \right| \leq C_q \|S_f\|_p \|g\|_q.$$

If

$$\|g\|_q \leq 1,$$

then

$$\left| \int fg \right| \leq C_q \|S_f\|_p.$$

14.6 Take the Supremum

Now take the supremum over all g such that

$$\|g\|_q \leq 1.$$

Using L^p-duality,

$$\|f\|_p = \sup_{\|g\|_q \leq 1} \left| \int fg \right|.$$

But we proved that for every such g,

$$\left| \int fg \right| \leq C_q \|S_f\|_p.$$

Therefore

$$\boxed{\|f\|_p \leq C_q \|S_f\|_p.}$$

Raising both sides to the p-th power,

$$\boxed{\int |f|^p \leq C_q^p \int S_f^p.}$$

This is the desired lower bound.

14.7 Final Square-Function Theorem

Combining the upper and lower bounds, we obtain

$$\boxed{c_p \int |f|^p \leq \int S_f^p \leq C_p \int |f|^p.}$$

Equivalently,

$$\boxed{\|S_f\|_p \sim_p \|f\|_p.}$$

This holds for every

$$1 < p < \infty.$$

14.8 Summary of the Lower Bound Proof

The lower bound follows from the chain of inequalities

$$\int fg = \int f_0 g_0 + \sum_i \int \Delta_i f \, \Delta_i g,$$

then

$$\left|\int fg\right| \leq \int S_f S_g,$$

then

$$\int S_f S_g \leq \|S_f\|_p \|S_g\|_q,$$

and finally

$$\|S_g\|_q \leq C_q \|g\|_q.$$

Thus

$$\left|\int fg\right| \leq C_q \|S_f\|_p \|g\|_q.$$

Taking the supremum over all g with

$$\|g\|_q \leq 1$$

gives

$$\|f\|_p \leq C_q \|S_f\|_p.$$

Therefore

$$\boxed{\int |f|^p \leq C_p \int S_f^p.}$$

15 The Hilbert Transforms and Martingale Transforms

15.1 The Hilbert Transform

The Hilbert transform is formally defined by

$$Hf(x) = \int_{\mathbb{R}} \frac{f(y)}{x-y}\, dy.$$

The problem is that the kernel

$$\frac{1}{x-y}$$

has a singularity when

$$x = y.$$

Thus the integral is not an ordinary Lebesgue integral.

Instead, the Hilbert transform is defined as a principal value:

$$Hf(x) = \text{p. v.} \int_{\mathbb{R}} \frac{f(y)}{x-y}\, dy.$$

That means

$$Hf(x) = \lim_{\varepsilon \to 0} \int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\, dy.$$

Equivalently, define the truncated kernel

$$K_\varepsilon(t) = \begin{cases} \dfrac{1}{t}, & |t| > \varepsilon, \\ 0, & |t| \le \varepsilon. \end{cases}$$

Then

$$H_\varepsilon f(x) = \int_{\mathbb{R}} K_\varepsilon(x-y) f(y)\, dy,$$

and

$$Hf(x) = \lim_{\varepsilon \to 0} H_\varepsilon f(x).$$

The main theorem one wants to prove is the L^p-boundedness of the Hilbert transform.

Theorem 15.1 (Boundedness of the Hilbert Transform)**.** *For every*

$$1 < p < \infty,$$

there exists a constant $C_p > 0$ such that

$$\boxed{\|Hf\|_{L^p(\mathbb{R})} \le C_p \|f\|_{L^p(\mathbb{R})}.}$$

The Hilbert transform is difficult to study directly because its kernel is singular and non-local.

So we first study a dyadic model operator built from Haar functions.

15.2 Dyadic Intervals and Haar Functions

A dyadic interval is an interval of the form

$$I = \left[k2^{-n}, (k+1)2^{-n}\right),$$

where

$$k, n \in \mathbb{Z}.$$

Each dyadic interval I splits into a left half and a right half:

$$I = I_l \cup I_r.$$

Define the unnormalized Haar function by

$$h_I = \mathbb{1}_{I_l} - \mathbb{1}_{I_r}.$$

Thus

$$h_I(x) = \begin{cases} 1, & x \in I_l, \\ -1, & x \in I_r, \\ 0, & x \notin I. \end{cases}$$

The Haar functions have mean zero:

$$\int h_I = 0.$$

Also,

$$h_I^2 = \mathbb{1}_I,$$

so

$$\int h_I^2 = |I|.$$

The Haar functions are orthogonal:

$$\int h_I h_J = 0$$

whenever

$$I \neq J.$$

More precisely,

$$\int h_I h_J = \begin{cases} 0, & I \neq J, \\ |I|, & I = J. \end{cases}$$

Thus the Haar functions form an orthogonal system, although not an orthonormal one.

15.3 Haar Expansion of a Function

Assume for simplicity that f has mean zero:

$$\int f = 0.$$

Then f has a Haar expansion

$$f = \sum_I c_I h_I.$$

The coefficient c_I is obtained by orthogonal projection:

$$c_I = \frac{\langle f, h_I \rangle}{\langle h_I, h_I \rangle}.$$

Since

$$\langle h_I, h_I \rangle = |I|,$$

we get

$$\boxed{c_I = \frac{1}{|I|} \int f h_I.}$$

Thus

$$f = \sum_I c_I h_I.$$

This is the Haar analogue of a Fourier expansion.

In martingale language, the martingale difference associated to I is

$$\Delta_I f = c_I h_I.$$

Therefore

$$f = \sum_I \Delta_I f = \sum_I c_I h_I.$$

15.4 The Dyadic Square Function

The dyadic square function of f is

$$S_f(x) = \left(\sum_I |\Delta_I f(x)|^2 \right)^{1/2}.$$

Since

$$\Delta_I f = c_I h_I$$

and

$$h_I^2 = \mathbb{1}_I,$$

we get

$$|\Delta_I f(x)|^2 = |c_I|^2 \mathbb{1}_I(x).$$

Hence

$$\boxed{S_f(x)^2 = \sum_I |c_I|^2 \mathbb{1}_I(x).}$$

The martingale square-function theorem says that for

$$1 < p < \infty,$$

we have

$$\boxed{\|f\|_{L^p} \simeq_p \|S_f\|_{L^p}.}$$

That is, there exist constants $c_p, C_p > 0$ such that

$$c_p \|f\|_{L^p} \leq \|S_f\|_{L^p} \leq C_p \|f\|_{L^p}.$$

15.5 The Dyadic Model Operator

Instead of studying the Hilbert transform directly, we define a dyadic model operator T.

We define T by specifying its action on Haar functions:

$$\boxed{T(h_I) = h_{I_l} - h_{I_r}.}$$

So if

$$f = \sum_I c_I h_I,$$

then

$$Tf = \sum_I c_I T(h_I).$$

Therefore

$$\boxed{Tf = \sum_I c_I(h_{I_l} - h_{I_r}).}$$

This operator is called a Haar shift.

It is a dyadic analogue of the Hilbert transform.

The Hilbert transform shifts oscillations in a continuous way, while T shifts Haar oscillations from a dyadic interval to its children.

15.6 What Happens to the Square Function?

The key question is:

What happens to S_f when we apply T?

That is, how does

$$S_{Tf}$$

compare to

$$S_f?$$

The answer is:

$$\boxed{S_{Tf} = S_f.}$$

This is the main miracle of the dyadic model.

15.7 Proof That $S_{Tf} = S_f$

Suppose

$$f = \sum_I c_I h_I.$$

Then

$$S_f(x)^2 = \sum_I |c_I|^2 \mathbb{1}_I(x).$$

Now apply T:

$$Tf = \sum_I c_I(h_{I_l} - h_{I_r}).$$

For each dyadic interval I, the coefficient c_I is moved to the two children I_l and I_r, with opposite signs.

However, the square function does not see signs.

It only sees squares of coefficients.

For a fixed $x \in I$, exactly one of the two children I_l or I_r contains x. Therefore the contribution of the coefficient c_I to the square function of Tf is still

$$|c_I|^2$$

on the interval I.

Hence

$$S_{Tf}(x)^2 = \sum_I |c_I|^2 \mathbb{1}_I(x).$$

But this is exactly

$$S_f(x)^2.$$

Therefore

$$\boxed{S_{Tf} = S_f.}$$

15.8 L^p-Boundedness of the Haar Shift

Using the square-function theorem, we prove that T is bounded on L^p.

For

$$1 < p < \infty,$$

we know

$$\|h\|_{L^p} \simeq_p \|S_h\|_{L^p}.$$

Apply this to Tf:

$$\|Tf\|_{L^p} \lesssim_p \|S_{Tf}\|_{L^p}.$$

But

$$S_{Tf} = S_f.$$

Therefore

$$\|Tf\|_{L^p} \lesssim_p \|S_f\|_{L^p}.$$

Apply the square-function theorem again to f:

$$\|S_f\|_{L^p} \lesssim_p \|f\|_{L^p}.$$

Thus

$$\boxed{\|Tf\|_{L^p} \leq C_p \|f\|_{L^p}.}$$

So the dyadic Haar shift is bounded on L^p for every

$$1 < p < \infty.$$

The argument can be summarized by the chain

$$\boxed{\|Tf\|_{L^p} \simeq_p \|S_{Tf}\|_{L^p} = \|S_f\|_{L^p} \simeq_p \|f\|_{L^p}.}$$

This is exactly the line from the notes:

$$\|f\|_{L^p} \approx \|S^f\|_{L^p} = \|S^{Tf}\|_{L^p} \approx \|Tf\|_{L^p}.$$

15.9 Writing T as an Integral Operator

Now we rewrite T in kernel form.

We have

$$Tf = \sum_I c_I (h_{I_l} - h_{I_r}),$$

where

$$c_I = \frac{1}{|I|} \int f(y) h_I(y)\, dy.$$

Substitute the formula for c_I:

$$Tf(x) = \sum_I \left(\frac{1}{|I|} \int f(y) h_I(y)\, dy \right) \left[h_{I_l}(x) - h_{I_r}(x) \right].$$

Move the sum inside the integral:

$$Tf(x) = \int \left(\sum_I \frac{h_I(y) \left[h_{I_l}(x) - h_{I_r}(x) \right]}{|I|} \right) f(y)\, dy.$$

Therefore

$$\boxed{Tf(x) = \int K(x,y) f(y)\, dy,}$$

where the kernel is

$$\boxed{K(x,y) = \sum_I \frac{h_I(y) \left[h_{I_l}(x) - h_{I_r}(x) \right]}{|I|}.}$$

Thus T is an integral operator, just like the Hilbert transform.

15.10 Comparison with the Hilbert Transform

The Hilbert transform has kernel

$$K_H(x,y) = \frac{1}{x-y}.$$

So

$$Hf(x) = \text{p. v.} \int_{\mathbb{R}} K_H(x,y) f(y)\, dy.$$

The dyadic Haar shift has kernel

$$K(x,y) = \sum_I \frac{h_I(y)\left[h_{I_l}(x) - h_{I_r}(x)\right]}{|I|}.$$

Both kernels have similar behavior.

1. Singularity near the diagonal

The Hilbert kernel

$$\frac{1}{x-y}$$

blows up when

$$x = y.$$

The Haar-shift kernel is also concentrated around dyadic intervals containing both x and y. Its singular behavior occurs near the diagonal

$$x = y.$$

2. Size

For the Hilbert transform,

$$|K_H(x,y)| = \frac{1}{|x-y|}.$$

For the Haar-shift kernel, each term has size roughly

$$\frac{1}{|I|}.$$

When I is the smallest dyadic interval containing both x and y, the length $|I|$ is comparable to $|x-y|$. Therefore the dyadic kernel behaves like

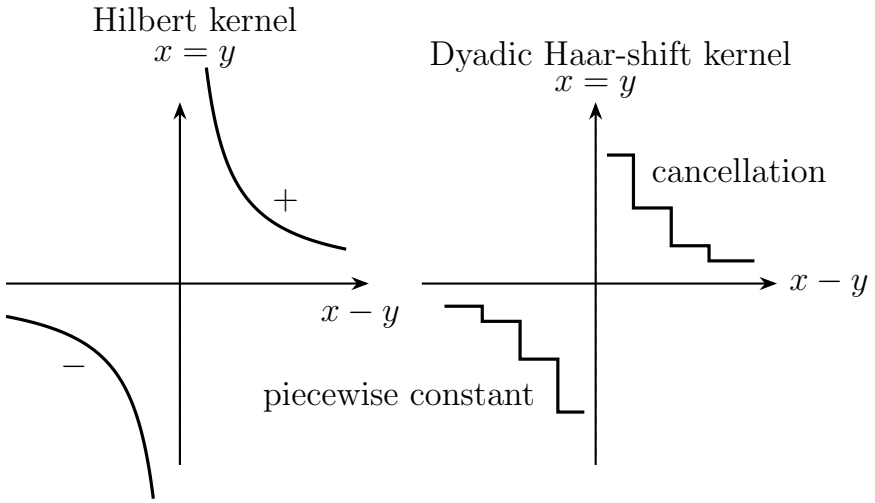
$$\frac{1}{|x-y|}.$$

3. Cancellation

The Hilbert kernel changes sign across the diagonal.

The Haar-shift kernel also has cancellation because the Haar functions take values $+1$ and -1.

Thus the Haar shift is a dyadic model of the Hilbert transform.



Both kernels are singular near $x = y$, have size comparable to $1/|x - y|$, and have cancellation across the diagonal.

Figure 3: The dyadic Haar-shift kernel is a piecewise constant dyadic model of the Hilbert kernel.

15.11 What Has Been Proved?

The notes shown prove the L^p-boundedness of the dyadic Haar shift:

$$\boxed{\|Tf\|_{L^p} \leq C_p \|f\|_{L^p}}$$

for

$$1 < p < \infty.$$

This is not yet the full Hilbert transform theorem, but it is the model argument.

The point is: Hilbert transform $\approx$ dyadic Haar shifts.

So if one can bound Haar shifts uniformly on L^p, one is on the path toward proving the boundedness of the Hilbert transform.

Key point. The Haar system is unconditional in L^p, $1 < p < \infty$. Changing signs of Haar coefficients does not destroy the L^p norm.

16 Averaging Haar Shifts and Recovering the Hilbert Transform

16.1 The Main Idea

In the previous notes, we introduced a dyadic Haar-shift operator T, which has the form

$$Tf(x) = \int_{\mathbb{R}} K(x,y) f(y)\, dy.$$

The Hilbert transform is

$$Hf(x) = \text{p. v.} \int_{\mathbb{R}} \frac{f(y)}{x-y}\, dy.$$

The goal of this section is to explain how the Hilbert transform kernel

$$\frac{1}{x-y}$$

arises by averaging dyadic Haar-shift kernels over translations and dilations.

The guiding principle is:

The Hilbert transform kernel is an average of dyadic Haar-shift kernels.

16.2 Recall the Haar-Shift Kernel

Let I be a dyadic interval and write

$$I = I_l \cup I_r$$

for its left and right halves.

The unnormalized Haar function is

$$h_I = \mathbf{1}_{I_l} - \mathbf{1}_{I_r}.$$

The dyadic Haar shift is defined by

$$T(h_I) = h_{I_l} - h_{I_r}.$$

If

$$f = \sum_I c_I h_I,$$

then

$$Tf = \sum_I c_I(h_{I_l} - h_{I_r}).$$

Since

$$c_I = \frac{1}{|I|}\int f(y)h_I(y)\,dy,$$

we may write

$$Tf(x) = \sum_I \left(\frac{1}{|I|}\int f(y)h_I(y)\,dy\right)\left[h_{I_l}(x) - h_{I_r}(x)\right].$$

Moving the sum inside the integral gives

$$Tf(x) = \int_{\mathbb{R}} \left(\sum_I \frac{h_I(y)\left[h_{I_l}(x) - h_{I_r}(x)\right]}{|I|}\right) f(y)\,dy.$$

Thus

$$\boxed{Tf(x) = \int_{\mathbb{R}} K(x,y)f(y)\,dy,}$$

where

$$\boxed{K(x,y) = \sum_I \frac{h_I(y)\left[h_{I_l}(x) - h_{I_r}(x)\right]}{|I|}.}$$

This is the dyadic Haar-shift kernel.

16.3 Why Averaging Is Needed

The kernel $K(x,y)$ depends on the dyadic grid.

However, the Hilbert transform kernel

$$\frac{1}{x-y}$$

is translation-invariant and scale-invariant.

Therefore, to recover the Hilbert transform from dyadic shifts, we average the dyadic kernel over:

(1) translations;

(2) dilations.

The idea is:

$$\text{dyadic kernel} \quad \longrightarrow \quad \text{average over translations and dilations} \quad \longrightarrow \quad \frac{c}{x-y}.$$

16.4 Averaging Over Translations

First average over translations.

For a translation parameter $t \in \mathbb{R}$, consider

$$K(x+t, y+t).$$

Since the dyadic grid is not translation-invariant, changing t changes the kernel.

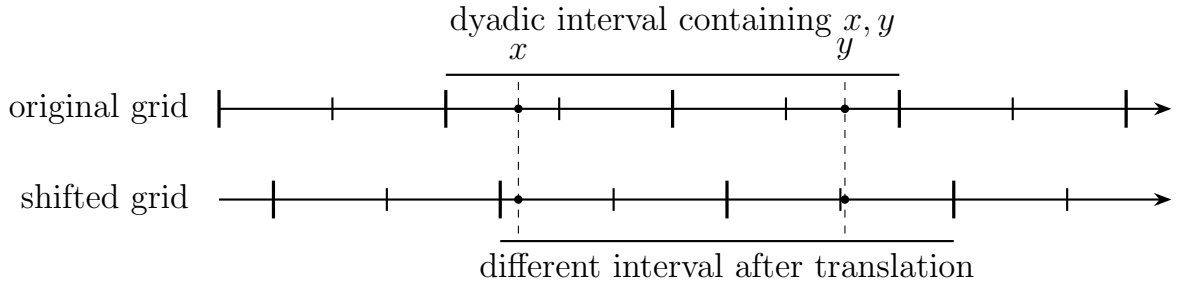


Figure 4: Translating the dyadic grid changes the dyadic interval geometry seen by the pair (x, y).

However, after averaging in t, the result depends only on the difference

$$z = y - x.$$

That is,

$$\mathrm{Av}_t\, K(x+t, y+t)$$

is a function of $z = y - x$.

In the notes, this averaged kernel has the form

$$\boxed{\mathrm{Av}_t\, K(x+t, y+t) = \sum_{k \in \mathbb{Z}} 2^k \varphi_0(2^k z), \qquad z = y - x.}$$

Here φ_0 is the basic function drawn in the notes.

The k-th term

$$2^k \varphi_0(2^k z)$$

represents the contribution from dyadic intervals of length approximately 2^{-k}.

16.5 Why the Scaling Looks Like This

At scale 2^{-k}, dyadic intervals have length

$$|I| = 2^{-k}.$$

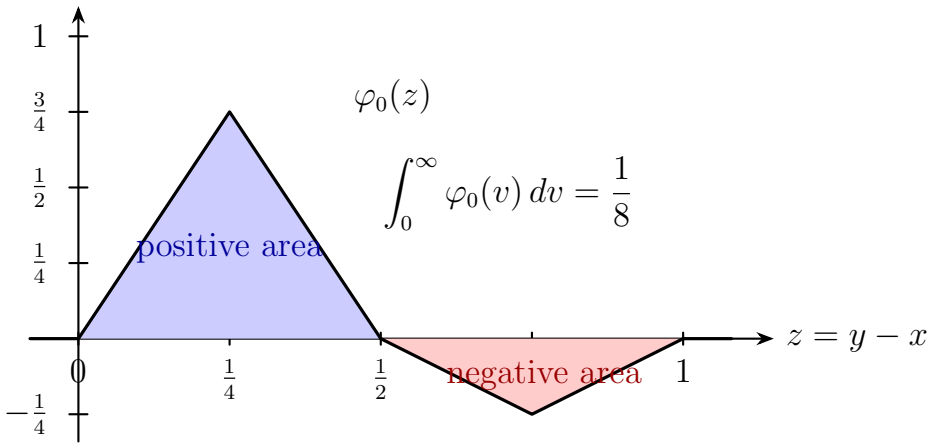


Figure 5: The basic function φ_0. The signed area gives the constant in the averaged kernel.

The Haar-shift kernel has size roughly

$$\frac{1}{|I|} = 2^k.$$

Also, the geometry at that scale depends on the normalized distance

$$\frac{z}{|I|} = 2^k z.$$

Therefore the contribution from scale k has the form

$$\boxed{\varphi_k(z) = 2^k \varphi_0(2^k z).}$$

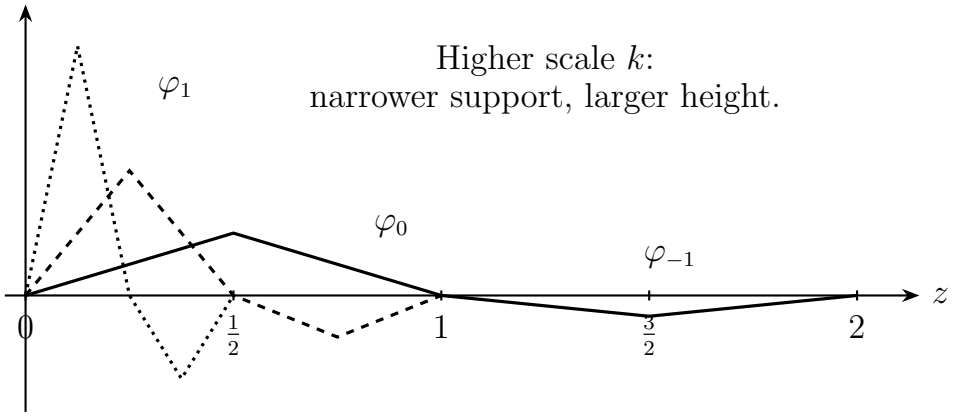


Figure 6: Scaled copies $\varphi_k(z) = 2^k \varphi_0(2^k z)$.

This is the usual scaling rule:

$$\text{height } 2^k, \qquad \text{width } 2^{-k}.$$

Thus the translation-averaged kernel is

$$\sum_{k\in\mathbb{Z}} \varphi_k(z) = \sum_{k\in\mathbb{Z}} 2^k \varphi_0(2^k z).$$

16.6 Averaging Over Dilations

The translation average is still dyadic because it only sees scales

$$2^k.$$

To remove the dyadic scale dependence, we also average over dilations

$$\lambda \in [1,2]$$

with respect to the measure

$$\frac{d\lambda}{\lambda}.$$

This is the natural measure because it is invariant under multiplicative scaling.
We consider

$$\mathrm{Av}_\lambda \left(\sum_k 2^k \lambda \varphi_0(2^k \lambda z) \right),$$

where the factor λ comes from the scaling of the kernel.
The normalized average over $[1,2]$ is

$$\frac{1}{\log 2} \int_1^2 \sum_k 2^k \lambda \varphi_0(2^k \lambda z) \frac{d\lambda}{\lambda}.$$

Therefore

$$\mathrm{Av}_\lambda \left(\sum_k 2^k \lambda \varphi_0(2^k \lambda z) \right) = \frac{1}{\log 2} \int_1^2 \sum_k 2^k \lambda \varphi_0(2^k \lambda z) \frac{d\lambda}{\lambda}.$$

Canceling the factor λ, we get

$$= \frac{1}{\log 2} \sum_k \int_1^2 2^k \varphi_0(2^k \lambda z)\, d\lambda.$$

16.7 The Change of Variables

For each fixed k, make the change of variables

$$u = 2^k \lambda.$$

Then

$$du = 2^k \, d\lambda.$$

When

$$\lambda = 1,$$

we have

$$u = 2^k.$$

When

$$\lambda = 2,$$

we have

$$u = 2^{k+1}.$$

Hence

$$\int_1^2 2^k \varphi_0(2^k \lambda z) \, d\lambda = \int_{2^k}^{2^{k+1}} \varphi_0(uz) \, du.$$

Therefore

$$\frac{1}{\log 2} \sum_k \int_1^2 2^k \varphi_0(2^k \lambda z) \, d\lambda = \frac{1}{\log 2} \sum_k \int_{2^k}^{2^{k+1}} \varphi_0(uz) \, du.$$

But the intervals

$$[2^k, 2^{k+1}]$$

partition $(0, \infty)$. Thus

$$\sum_k \int_{2^k}^{2^{k+1}} \varphi_0(uz) \, du = \int_0^\infty \varphi_0(uz) \, du.$$

Therefore

$$\mathrm{Av}_\lambda \left(\sum_k 2^k \lambda \varphi_0(2^k \lambda z) \right) = \frac{1}{\log 2} \int_0^\infty \varphi_0(uz) \, du.$$

16.8 Recovering the Kernel $1/z$

Now assume first that

$$z > 0.$$

Make the change of variables

$$v = uz.$$

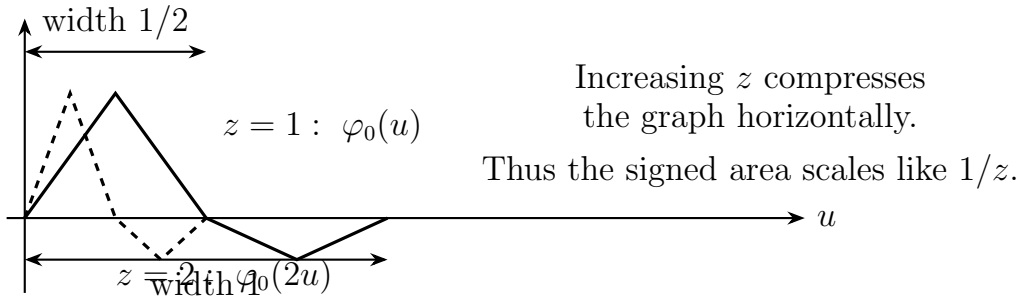


Figure 7: The substitution $v = uz$ shows that $\int_0^\infty \varphi_0(uz)\,du$ is proportional to $1/z$.

Then

$$du = \frac{dv}{z}.$$

Thus

$$\frac{1}{\log 2}\int_0^\infty \varphi_0(uz)\,du = \frac{1}{z\log 2}\int_0^\infty \varphi_0(v)\,dv.$$

The integral

$$\int_0^\infty \varphi_0(v)\,dv$$

is just a fixed constant.

In the notes, this constant is written as

$$\int_0^\infty \varphi_0(v)\,dv = \frac{1}{8}.$$

Hence

$$\frac{1}{\log 2}\int_0^\infty \varphi_0(uz)\,du = \frac{1}{8\log 2}\frac{1}{z}.$$

Thus

$$\boxed{\mathrm{Av}_{\lambda,t}\, K(\lambda(x+t),\lambda(y+t)) = \frac{c}{z}, \qquad z = y - x,}$$

where

$$c = \frac{1}{8\log 2}.$$

Since

$$z = y - x,$$

this becomes

$$\frac{c}{y-x} = -\frac{c}{x-y}.$$

This is the Hilbert-transform kernel up to a harmless constant and sign.

Therefore,

The averaged Haar-shift kernel is a constant multiple of the Hilbert kernel.

16.9 Truncated Kernels

In practice, one often truncates the sum over scales.

Define

$$K^{(N)}(x,y)$$

by keeping only the scales

$$-N \le k \le N.$$

Then define

$$T_N f(x) = \int_{\mathbb{R}} K^{(N)}(x,y) f(y)\, dy.$$

The notes indicate that, except for jump discontinuities coming from the Haar functions,

$$T_N f(x) \longrightarrow cHf(x)$$

as

$$N \to \infty.$$

That is,

$$\boxed{T_N f(x) = \int_{\mathbb{R}} K^{(N)}(x,y) f(y)\, dy \longrightarrow cHf(x).}$$

The jumps occur because Haar functions are piecewise constant, so their kernels are not smooth.

However, away from the diagonal

$$x = y,$$

the averaged truncated kernels converge to

$$\frac{c}{x-y}.$$

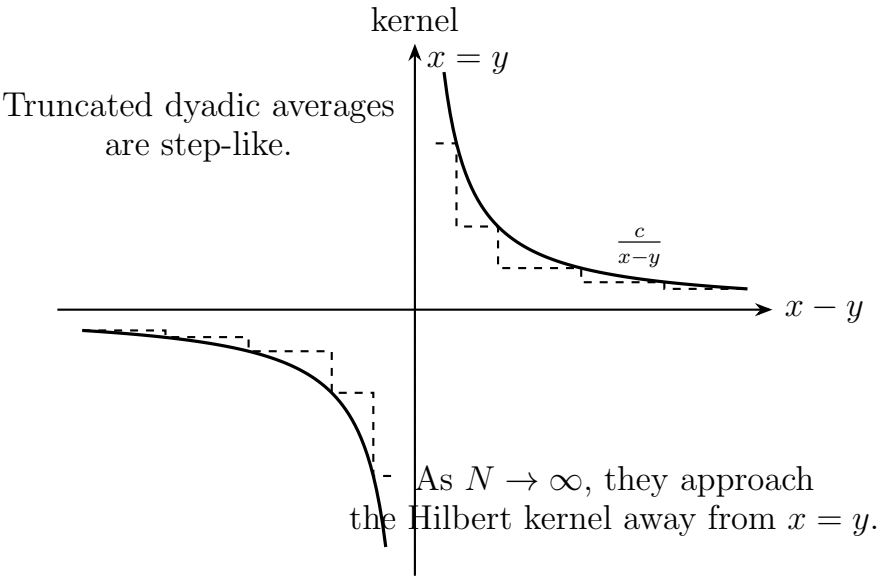


Figure 8: Averaged truncated kernels approximate the singular Hilbert kernel away from the diagonal.

16.10 Why This Proves Boundedness of the Hilbert Transform

We already proved that the dyadic Haar shift T satisfies

$$\boxed{\|Tf\|_{L^p} \leq C_p \|f\|_{L^p}}$$

for every

$$1 < p < \infty.$$

Translations and dilations of T have the same L^p boundedness, up to harmless normalization.

Therefore each shifted and dilated operator $T_{\lambda,t}$ satisfies

$$\|T_{\lambda,t} f\|_{L^p} \leq C_p \|f\|_{L^p}.$$

An average of bounded operators is also bounded with the same type of estimate. Hence

$$\|\mathrm{Av}_{\lambda,t}\, T_{\lambda,t} f\|_{L^p} \leq C_p \|f\|_{L^p}.$$

But the averaged operator is a constant multiple of the Hilbert transform:

$$\mathrm{Av}_{\lambda,t}\, T_{\lambda,t} = cH.$$

Therefore

$$\|cHf\|_{L^p} \leq C_p \|f\|_{L^p}.$$

Since $c \neq 0$, we obtain

$$\boxed{\|Hf\|_{L^p(\mathbb{R})} \leq C_p \|f\|_{L^p(\mathbb{R})}.}$$

This proves the L^p-boundedness of the Hilbert transform for

$$1 < p < \infty.$$

16.11 The Case of Piecewise Constant Compactly Supported Functions

We now explain how the previous averaging argument answers the following question.

Let f be a piecewise constant function with compact support. Show that

$$\|Hf\|_{L^p(\mathbb{R})} \leq C_p \|f\|_{L^p(\mathbb{R})}, \qquad 1 < p < \infty.$$

Recall that the Hilbert transform is defined by the principal value formula

$$Hf(x) = \text{p. v.} \int_{\mathbb{R}} \frac{f(y)}{x-y}\, dy.$$

Equivalently,

$$Hf(x) = \lim_{\varepsilon \to 0} \int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\, dy.$$

Since f is piecewise constant and compactly supported, this principal value is well-defined for almost every x. There may be issues at jump points of f, but these points form a set of measure zero and therefore do not affect the L^p-norm.

Theorem 16.1. *Let $1 < p < \infty$. If f is piecewise constant and compactly supported on $\mathbb{R}$, then*

$$\boxed{\|Hf\|_{L^p(\mathbb{R})} \leq C_p \|f\|_{L^p(\mathbb{R})}.}$$

Proof. From the dyadic Haar-shift argument, we already know that the Haar shift T satisfies

$$\boxed{\|Tf\|_{L^p} \leq C_p \|f\|_{L^p}}$$

for every $1 < p < \infty$.

Indeed, this followed from the square-function identity

$$S_{Tf} = S_f$$

and the martingale square-function theorem

$$\|f\|_{L^p} \simeq_p \|S_f\|_{L^p}.$$

Hence

$$\|Tf\|_{L^p} \simeq_p \|S_{Tf}\|_{L^p} = \|S_f\|_{L^p} \simeq_p \|f\|_{L^p}.$$

Now consider translated and dilated copies of the Haar shift. We denote these operators by

$$T_{\lambda,t},$$

where $t \in \mathbb{R}$ is a translation parameter and $\lambda \in [1,2]$ is a dilation parameter.

Translations and dilations preserve the same type of L^p-estimate, so each $T_{\lambda,t}$ satisfies

$$\|T_{\lambda,t}f\|_{L^p} \leq C_p\|f\|_{L^p},$$

with a constant independent of λ and t.

Now truncate the Haar-shift kernel to finitely many dyadic scales. Let

$$T_{\lambda,t}^{(N)}$$

denote the shifted and dilated Haar-shift operator using only the scales

$$-N \leq k \leq N.$$

Define the averaged operator

$$A_N f = \mathrm{Av}_{\lambda,t}\, T_{\lambda,t}^{(N)} f.$$

Since each $T_{\lambda,t}^{(N)}$ is bounded on L^p, the average is also bounded. Therefore

$$\|A_N f\|_{L^p} \leq C_p\|f\|_{L^p}.$$

The kernel computation from the previous section shows that the averaged kernels converge to a constant multiple of the Hilbert transform kernel. More precisely,

$$A_N f(x) \longrightarrow cHf(x)$$

as $N \to \infty$, for almost every x, where $c \neq 0$ is an absolute constant.

This is exactly where the averaging over translations and dilations is used. The averaged dyadic kernel becomes

$$\frac{c}{x-y},$$

which is the Hilbert-transform kernel up to a constant.

Now apply Fatou's lemma:

$$\|cHf\|_{L^p}^p = \int_{\mathbb{R}} |cHf(x)|^p\, dx.$$

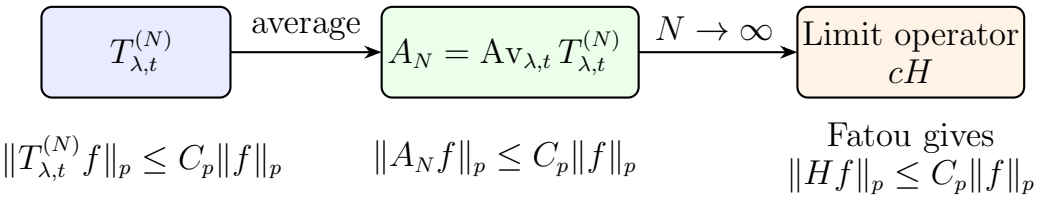


Figure 9: Structure of the averaging argument proving the L^p boundedness of the Hilbert transform.

Since

$$A_N f(x) \to cHf(x)$$

for almost every x, we have

$$\int_{\mathbb{R}} |cHf(x)|^p \, dx = \int_{\mathbb{R}} \lim_{N\to\infty} |A_N f(x)|^p \, dx.$$

By Fatou's lemma,

$$\int_{\mathbb{R}} \lim_{N\to\infty} |A_N f(x)|^p \, dx \le \liminf_{N\to\infty} \int_{\mathbb{R}} |A_N f(x)|^p \, dx.$$

Hence

$$\|cHf\|_{L^p}^p \le \liminf_{N\to\infty} \|A_N f\|_{L^p}^p.$$

Using the uniform bound

$$\|A_N f\|_{L^p} \le C_p \|f\|_{L^p},$$

we obtain

$$\|cHf\|_{L^p}^p \le C_p^p \|f\|_{L^p}^p.$$

Since $c \neq 0$, we can divide by $|c|^p$. Thus

$$\|Hf\|_{L^p}^p \le C_p \|f\|_{L^p}^p.$$

Taking p-th roots gives

$$\boxed{\|Hf\|_{L^p} \le C_p \|f\|_{L^p}.}$$

This proves the desired estimate for piecewise constant compactly supported functions. □

Key point. The assumption that f is piecewise constant and compactly supported is a convenient starting class. Once the estimate is proved on this dense class, the Hilbert transform can be extended by density to all of $L^p(\mathbb{R})$, $1 < p < \infty$.

17 Covering Density on $\mathbb{Z}$

17.1 The main question

Let

$$E \subset \mathbb{Z}$$

be a finite set. We want to cover long intervals of integers using shifts, or translates, of E.

A shift of E by an integer z is

$$z + E = \{z + e : e \in E\}.$$

The guiding question is:

> How many shifts of E are needed to cover a long interval of integers?

This is a discrete covering problem. The word *density* will measure the asymptotic number of translates needed per integer.

17.2 Definition of m_n

Let

$$[1, n]_{\mathbb{Z}} = \{1, 2, \ldots, n\}.$$

Define m_n to be the minimal number of shifts of E needed to cover $[1, n]_{\mathbb{Z}}$. Thus m_n is the smallest integer for which there exist shifts $z_1, \ldots, z_{m_n} \in \mathbb{Z}$ satisfying

$$[1, n]_{\mathbb{Z}} \subset \bigcup_{j=1}^{m_n} (z_j + E).$$

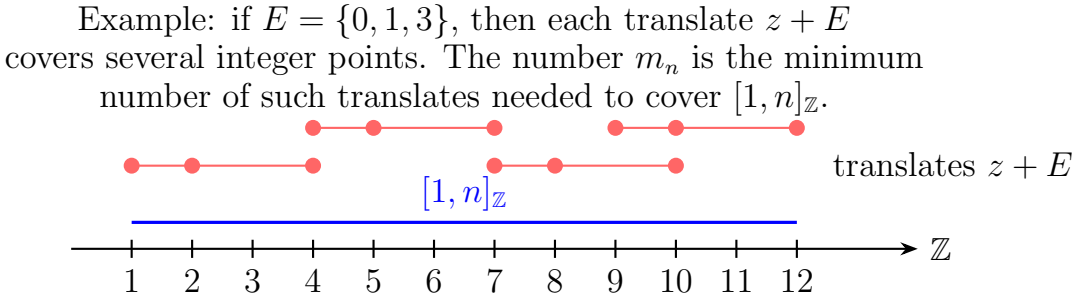


Figure 10: Covering a long interval of integers by translates of a finite set E.

Equivalently, because translating the interval does not change the problem, m_n is also the minimal number of shifts needed to cover any interval of n consecutive integers.

Example 17.1. If $E = \{0\}$, then each translate covers exactly one integer. Hence

$$m_n = n, \qquad \frac{m_n}{n} = 1.$$

So the covering density is 1.

Example 17.2. If $E = \{0, 1\}$, one translate covers two consecutive integers. To cover $[1, n]$, we need roughly $n/2$ translates. More precisely,

$$m_n = \left\lceil \frac{n}{2} \right\rceil, \qquad \lim_{n\to\infty} \frac{m_n}{n} = \frac{1}{2}.$$

17.3 Subadditivity

The first important observation is that m_n is subadditive.

Proposition 17.3 (Subadditivity)**.** *For all positive integers n, n',*

$$\boxed{m_{n+n'} \leq m_n + m_{n'}.}$$

Proof. Cover $[1, n]$ using m_n shifts of E. Then cover $[n+1, n+n']$, which is an interval of length n', using $m_{n'}$ shifts of E. Combining these two coverings gives a covering of $[1, n + n']$ using $m_n + m_{n'}$ shifts. Since $m_{n+n'}$ is minimal, we get

$$m_{n+n'} \leq m_n + m_{n'}.$$

□

17.4 Existence of the limiting density

Because m_n is subadditive, the average m_n/n has a limit.

Theorem 17.4 (Existence of covering density)**.** *The limit*

$$\boxed{d(E) = \lim_{n\to\infty} \frac{m_n}{n}}$$

exists. Moreover,

$$\boxed{d(E) = \inf_{n\geq 1} \frac{m_n}{n}.}$$

Proof. Let

$$L = \inf_{n\geq 1} \frac{m_n}{n}.$$

Clearly

$$L \leq \frac{m_n}{n}$$

for every n, so

$$L \leq \liminf_{n\to\infty} \frac{m_n}{n}.$$

It remains to show

$$\limsup_{n\to\infty} \frac{m_n}{n} \leq L.$$

Fix $\varepsilon > 0$. By definition of infimum, there exists N_ε such that

$$\frac{m_{N_\varepsilon}}{N_\varepsilon} \leq L + \varepsilon.$$

Now cover a long interval $[1, n]$ by consecutive blocks of length N_ε, plus possibly one leftover block. The number of such blocks is at most

$$k \leq \frac{n}{N_\varepsilon} + 1.$$

Each full block can be covered using m_{N_ε} shifts. Hence

$$m_n \leq k m_{N_\varepsilon}.$$

Therefore

$$\frac{m_n}{n} \leq \frac{k m_{N_\varepsilon}}{n} \leq \frac{m_{N_\varepsilon}}{N_\varepsilon} + \frac{m_{N_\varepsilon}}{n}.$$

Letting $n \to \infty$, the second term tends to 0, and so

$$\limsup_{n\to\infty} \frac{m_n}{n} \leq L + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary,

$$\limsup_{n\to\infty} \frac{m_n}{n} \leq L.$$

Combining this with the earlier lower bound gives

$$\lim_{n\to\infty} \frac{m_n}{n} = L.$$

□

Definition 17.5 (Covering density)**.** The number

$$d(E) = \lim_{n\to\infty} \frac{m_n}{n}$$

is called the *covering density* of E.

17.5 Local density principle

The number $d(E)$ is not only a limit for finite intervals. It also describes the local density that any infinite covering must have on long intervals.

Suppose $Z \subset \mathbb{Z}$ is a set of shifts such that

$$\mathbb{Z} = \bigcup_{z \in Z} (z + E).$$

Let $I \subset \mathbb{Z}$ be a finite interval. We count the number of shifts whose translate meets I:

$$N_Z(I) = \#\{z \in Z : (z + E) \cap I \neq \varnothing\}.$$

17.6 A local estimate for economical covers

Let

$$D = \operatorname{diam}(E) = \max E - \min E.$$

Since $E \subset \mathbb{Z}$ is finite, $D < \infty$.

For each n, let $Z_n \subset \mathbb{Z}$ be a set of shifts giving a most economical cover of $[1, n]_{\mathbb{Z}}$. Thus

$$[1, n]_{\mathbb{Z}} \subset \bigcup_{z \in Z_n} (z + E)$$

and

$$|Z_n| = m_n.$$

The following lemma says that if we take a subinterval I of $[1, n]_{\mathbb{Z}}$, then not too many shifts from the optimal cover of $[1, n]_{\mathbb{Z}}$ can meet I.

Proposition 17.6 (Local counting lemma)**.** *Let $I \subset [1, n]_{\mathbb{Z}}$ be an interval of length $n' \leq n$. Then*

$$\boxed{\#\{z \in Z_n : (z + E) \cap I \neq \varnothing\} \leq m_{n'} + 2 \operatorname{diam}(E).}$$

Equivalently,

$$\boxed{N_{Z_n}(I) \leq m_{n'} + 2D.}$$

Proof. Write

$$I = [a, b]_{\mathbb{Z}}, \qquad |I| = n',$$

so

$$b - a + 1 = n'.$$

We split the shifts $z \in Z_n$ whose translates meet I into two types.

First, consider the shifts whose translates are completely contained in I:

$$\mathcal{A}_{\text{in}} = \{z \in Z_n : z + E \subset I\}.$$

We claim that

$$|\mathcal{A}_{\text{in}}| \leq m_{n'}.$$

Indeed, suppose for contradiction that

$$|\mathcal{A}_{\text{in}}| > m_{n'}.$$

Remove all translates $z + E$ with $z \in \mathcal{A}_{\text{in}}$ from the optimal cover of $[1, n]_{\mathbb{Z}}$. These removed translates are contained in I, so outside I the covering is unchanged.

Now replace them by a most economical cover of I. Since I has length n', it can be covered by $m_{n'}$ shifts of E. Therefore we have replaced more than $m_{n'}$ translates by only $m_{n'}$ translates. This gives a cover of $[1, n]_{\mathbb{Z}}$ using fewer than m_n shifts, contradicting the minimality of Z_n.

Hence

$$|\mathcal{A}_{\text{in}}| \leq m_{n'}.$$

Now consider the boundary shifts:

$$\mathcal{A}_{\text{bd}} = \{z \in Z_n : (z + E) \cap I \neq \varnothing \text{ and } z + E \not\subset I\}.$$

These are translates that meet I but are not completely contained in I. Such a translate must stick out through the left endpoint or the right endpoint of I.

Let

$$e_{\min} = \min E, \qquad e_{\max} = \max E.$$

Then

$$D = e_{\max} - e_{\min}.$$

If $z + E$ meets $I = [a, b]_{\mathbb{Z}}$, then for some $e \in E$,

$$z + e \in [a, b]_{\mathbb{Z}}.$$

Thus

$$z \in [a - e_{\max},\, b - e_{\min}]_{\mathbb{Z}}.$$

On the other hand, $z + E \subset I$ exactly when

$$z + e_{\min} \geq a$$

and

$$z + e_{\max} \leq b.$$

Equivalently,

$$z \in [a - e_{\min},\, b - e_{\max}]_{\mathbb{Z}}.$$

Therefore the shifts that meet I but are not contained in I lie in the two boundary regions

$$[a - e_{\max},\, a - e_{\min} - 1]_{\mathbb{Z}}$$

and

$$[b - e_{\max} + 1,\, b - e_{\min}]_{\mathbb{Z}}.$$

Each of these intervals has length at most

$$e_{\max} - e_{\min} = D.$$

Hence

$$|\mathcal{A}_{\mathrm{bd}}| \le 2D.$$

Combining the inside and boundary parts, we get

$$\#\{z \in Z_n : (z + E) \cap I \neq \varnothing\} \le |\mathcal{A}_{\mathrm{in}}| + |\mathcal{A}_{\mathrm{bd}}|.$$

Using the estimates above,

$$\#\{z \in Z_n : (z + E) \cap I \neq \varnothing\} \le m_{n'} + 2D.$$

Since $D = \mathrm{diam}(E)$, this proves

$$\boxed{N_{Z_n}(I) \le m_{n'} + 2\,\mathrm{diam}(E).}$$

□

Key point. The meaning of the lemma is simple. Inside the interval I, an optimal cover of the large interval cannot use more than $m_{n'}$ fully internal translates. Otherwise, we could replace them by a better cover of I and improve the global cover. The only extra translates are the ones crossing the two endpoints of I, and there are at most $2\,\mathrm{diam}(E)$ of them.

17.7 Counting translates contained in an interval

Sometimes it is useful to count only those translates that are fully contained in I. Define

$$R_Z(I) = \#\{z \in Z : z + E \subset I\}.$$

We claim that for every $\varepsilon > 0$, there exists $K \ge 1$ such that if $I \subset \mathbb{Z}$ is an interval with

$$|I| \ge K,$$

then

$$\boxed{\frac{R_Z(I)}{|I|} \geq d(E) - \varepsilon.}$$

Indeed, the translates that meet I but are not contained in I must cross one of the two endpoints of I. Since E has diameter $D = \operatorname{diam}(E)$, there are at most $2D$ such boundary translates. Hence

$$N_Z(I) \leq R_Z(I) + 2D.$$

Equivalently,

$$R_Z(I) \geq N_Z(I) - 2D.$$

From the previous proposition,

$$N_Z(I) \geq m_{|I|}.$$

Therefore

$$R_Z(I) \geq m_{|I|} - 2D.$$

Dividing by $|I|$, we get

$$\frac{R_Z(I)}{|I|} \geq \frac{m_{|I|}}{|I|} - \frac{2D}{|I|}.$$

Since

$$\frac{m_{|I|}}{|I|} \to d(E)$$

as $|I| \to \infty$, and

$$\frac{2D}{|I|} \to 0,$$

we can choose K large enough so that for every $|I| \geq K$,

$$\frac{m_{|I|}}{|I|} \geq d(E) - \frac{\varepsilon}{2}$$

and

$$\frac{2D}{|I|} \leq \frac{\varepsilon}{2}.$$

Thus

$$\frac{R_Z(I)}{|I|} \geq d(E) - \varepsilon.$$

This proves the contained-translate version.

17.8 Existence of an infinite covering with optimal local density

We now prove the converse direction: there exists an infinite set of shifts $Z \subset \mathbb{Z}$ whose local density is at most $d(E)$.

More precisely, we prove that there exists $Z \subset \mathbb{Z}$ such that

$$\mathbb{Z} = \bigcup_{z \in Z} (z + E)$$

and for every $\varepsilon > 0$, there exists $K \geq 1$ such that whenever $I \subset \mathbb{Z}$ is an interval with

$$|I| \geq K,$$

we have

$$\boxed{\frac{N_Z(I)}{|I|} \leq d(E) + \varepsilon.}$$

Together with the local lower-density principle, this shows that the best possible local density is exactly $d(E)$.

Theorem 17.7 (Existence of an optimal-density covering)**.** *There exists a set of shifts $Z \subset \mathbb{Z}$ such that*

$$\mathbb{Z} = \bigcup_{z \in Z} (z + E)$$

and for every $\varepsilon > 0$, there exists $K \geq 1$ such that for every interval $I \subset \mathbb{Z}$ with $|I| \geq K$,

$$\boxed{N_Z(I) \leq (d(E) + \varepsilon)|I|.}$$

Equivalently,

$$\boxed{\frac{N_Z(I)}{|I|} \leq d(E) + \varepsilon.}$$

Proof. Let

$$D = \operatorname{diam}(E).$$

We split $\mathbb{Z}$ into consecutive finite intervals

$$\mathbb{Z} = \bigcup_{\ell \in \mathbb{Z}} J_\ell$$

such that

$$|J_\ell| \to \infty$$

as $|\ell| \to \infty$.

For each interval J_ℓ, choose a most economical cover of J_ℓ by shifts of E. Since J_ℓ has length $|J_\ell|$, this cover uses exactly

$$m_{|J_\ell|}$$

shifts.

Let $Z_\ell \subset \mathbb{Z}$ be the set of shifts used to cover J_ℓ. Thus

$$J_\ell \subset \bigcup_{z \in Z_\ell} (z + E)$$

and

$$|Z_\ell| = m_{|J_\ell|}.$$

Now define

$$Z = \bigcup_{\ell \in \mathbb{Z}} Z_\ell,$$

removing repetitions if necessary.

Then

$$\mathbb{Z} = \bigcup_{\ell \in \mathbb{Z}} J_\ell \subset \bigcup_{\ell \in \mathbb{Z}} \bigcup_{z \in Z_\ell} (z + E) = \bigcup_{z \in Z} (z + E).$$

Hence Z gives a covering of $\mathbb{Z}$.

Now let $I \subset \mathbb{Z}$ be an interval. We want to estimate

$$N_Z(I) = \#\{z \in Z : (z + E) \cap I \neq \varnothing\}.$$

The interval I intersects some of the partition intervals J_ℓ. Let

$$\mathcal{L}(I) = \{\ell : J_\ell \cap I \neq \varnothing\}.$$

Then

$$I \subset \bigcup_{\ell \in \mathcal{L}(I)} J_\ell.$$

For most of these J_ℓ's, the whole J_ℓ lies inside I. At most two of them are boundary intervals, namely the intervals containing the left and right endpoints of I.

For each $\ell \in \mathcal{L}(I)$, the shifts coming from the cover of J_ℓ that intersect I are at most

$$m_{|J_\ell|} + 2D.$$

The $2D$ term accounts for translates that cross the boundary of I. Since there are at most two boundary intervals, the total boundary error is at most

$$6D$$

after allowing a harmless extra constant.

Thus

$$N_Z(I) \leq \sum_{\ell \in \mathcal{L}(I)} m_{|J_\ell|} + 6D.$$

Now fix $\varepsilon > 0$. Since

$$\frac{m_n}{n} \to d(E),$$

there exists $K \geq 1$ such that for all $n \geq K$,

$$\frac{m_n}{n} \leq d(E) + \frac{\varepsilon}{2}.$$

Therefore, if

$$|J_\ell| \geq K,$$

then

$$m_{|J_\ell|} \leq \left(d(E) + \frac{\varepsilon}{2}\right) |J_\ell|.$$

Separate the sum into short and long J_ℓ's:

$$\sum_{\ell \in \mathcal{L}(I)} m_{|J_\ell|} = \sum_{\substack{\ell \in \mathcal{L}(I) \\ |J_\ell| < K}} m_{|J_\ell|} + \sum_{\substack{\ell \in \mathcal{L}(I) \\ |J_\ell| \geq K}} m_{|J_\ell|}.$$

For the long intervals,

$$\sum_{\substack{\ell \in \mathcal{L}(I) \\ |J_\ell| \geq K}} m_{|J_\ell|} \leq \left(d(E) + \frac{\varepsilon}{2}\right) \sum_{\substack{\ell \in \mathcal{L}(I) \\ |J_\ell| \geq K}} |J_\ell|.$$

Since the J_ℓ's intersecting I cover I, and since only two boundary intervals may extend beyond I, we get

$$\sum_{\ell \in \mathcal{L}(I)} |J_\ell| \leq |I| + O(K).$$

In the rough estimate from the notes, the short intervals contribute at most

$$K(K+2).$$

Thus

$$\sum_{\ell \in \mathcal{L}(I)} m_{|J_\ell|} \leq K(K+2) + \left(d(E) + \frac{\varepsilon}{2}\right) |I|.$$

Hence

$$N_Z(I) \leq K(K+2) + 6D + \left(d(E) + \frac{\varepsilon}{2}\right) |I|.$$

Now choose $|I|$ sufficiently large so that

$$K(K+2)+6D \leq \frac{\varepsilon}{2}|I|.$$

Then

$$N_Z(I) \leq \frac{\varepsilon}{2}|I| + \left(d(E)+\frac{\varepsilon}{2}\right)|I|.$$

Therefore

$$N_Z(I) \leq (d(E)+\varepsilon)|I|.$$

Dividing by $|I|$, we obtain

$$\boxed{\frac{N_Z(I)}{|I|} \leq d(E)+\varepsilon.}$$

This proves the theorem. □

Key point. The construction is simple: cover larger and larger blocks optimally, then paste those optimal covers together. The only possible loss occurs near the boundaries of the blocks, and this boundary loss becomes negligible on long intervals.

17.9 Summary

For a finite set $E \subset \mathbb{Z}$, we defined m_n as the minimum number of shifts of E needed to cover $[1,n]$. The key property is subadditivity:

$$m_{n+n'} \leq m_n + m_{n'}.$$

Therefore the limit

$$d(E) = \lim_{n\to\infty} \frac{m_n}{n}$$

exists and equals

$$\inf_{n\geq 1} \frac{m_n}{n}.$$

This number is the asymptotic covering density of E.

18 Upper Bound for Covering Density

18.1 Statement of the main estimate

Let

$$E \subset \mathbb{Z}, \qquad |E| = \ell.$$

The goal of this lecture is to prove the two-sided estimate

$$\boxed{\frac{1}{\ell} \leq d(E) \leq C\frac{1+log\ell}{\ell}.}$$

The lower bound is elementary. The upper bound is proved by the probabilistic method.

18.2 The easy lower bound

Each translate $z + E$ contains exactly ℓ points. To cover n points, at least n/ℓ translates are needed. Hence

$$m_n \geq \frac{n}{\ell}.$$

Dividing by n and letting $n \to \infty$, we obtain

$$\boxed{d(E) \geq \frac{1}{\ell}.}$$

18.3 Probabilistic proof of the upper bound

We now prove the upper bound

$$\boxed{d(E) \leq C\frac{1+\log|E|}{|E|}.}$$

Let

$$|E| = \ell$$

and write

$$D = \operatorname{diam}(E).$$

Recall that

$$d(E) = \inf_{n \geq 1} \frac{m_n}{n}.$$

It is enough to prove that for large n,

$$\frac{m_n}{n} \lesssim \frac{1+\log \ell}{\ell}.$$

We cover the interval

$$[0, n-1]_{\mathbb{Z}} = \{0, 1, \ldots, n-1\}.$$

This is equivalent to covering $[1, n]_{\mathbb{Z}}$, since translating the interval does not change m_n.

Let

$$\widetilde{n} = n + D.$$

Choose m shifts independently and uniformly at random from the interval

$$[-D, n-1]_{\mathbb{Z}}.$$

This interval has length $\widetilde{n}$.

Fix a point

$$x \in [0, n-1]_{\mathbb{Z}}.$$

We ask: what is the probability that one random shift $z + E$ covers x?

The point x belongs to $z + E$ exactly when

$$x = z + e$$

for some $e \in E$. Equivalently,

$$z = x - e.$$

Since E has ℓ elements, there are exactly ℓ shifts z for which

$$x \in z + E.$$

Also these shifts all lie in $[-D, n-1]_{\mathbb{Z}}$. Hence

$$\mathbb{P}\{x \text{ is covered by one random shift}\} = \frac{\ell}{\widetilde{n}}.$$

Therefore

$$\mathbb{P}\{x \text{ is not covered by one random shift}\} = 1 - \frac{\ell}{\widetilde{n}}.$$

Since the m shifts are chosen independently,

$$\mathbb{P}\{x \text{ is not covered by any of the } m \text{ shifts}\} = \left(1 - \frac{\ell}{\widetilde{n}}\right)^m.$$

Let

$$U = \#\{x \in [0, n-1]_{\mathbb{Z}} : x \text{ is not covered after the } m \text{ random shifts}\}.$$

Then, by linearity of expectation,

$$\mathbb{E}[U] = \sum_{x=0}^{n-1} \mathbb{P}\{x \text{ is not covered}\}.$$

Thus

$$\mathbb{E}[U] = n\left(1 - \frac{\ell}{\widetilde{n}}\right)^m.$$

Now, after the m random shifts have been chosen, suppose U points remain uncovered. We can cover each remaining uncovered point by adding one extra shift of E. Therefore the total number of shifts needed is at most

$$Y = m + U.$$

Taking expectations,

$$\mathbb{E}[Y] = m + \mathbb{E}[U].$$

Hence

$$\mathbb{E}[Y] = m + n\left(1 - \frac{\ell}{\widetilde{n}}\right)^m.$$

By the averaging principle, there exists at least one choice of the m random shifts for which

$$Y \leq m + n\left(1 - \frac{\ell}{\widetilde{n}}\right)^m.$$

Indeed, if every choice gave a larger value of Y, then the expectation would also be larger.

Thus there exists a deterministic cover of $[0, n-1]_{\mathbb{Z}}$ using at most

$$m + n\left(1 - \frac{\ell}{\widetilde{n}}\right)^m$$

shifts. Since m_n is minimal, we get

$$m_n \leq m + n\left(1 - \frac{\ell}{\widetilde{n}}\right)^m.$$

Dividing by n, we obtain

$$\frac{m_n}{n} \leq \frac{m}{n} + \left(1 - \frac{\ell}{\widetilde{n}}\right)^m.$$

Now choose

$$m = \left\lceil \frac{\widetilde{n}}{\ell} \log \ell \right\rceil .$$

Using the standard inequality

$$1 - u \leq e^{-u}, \qquad 0 < u < 1,$$

we get

$$\left(1 - \frac{\ell}{\widetilde{n}}\right)^m \leq \exp\left(-\frac{m\ell}{\widetilde{n}}\right).$$

By the choice of m,

$$\frac{m\ell}{\widetilde{n}} \geq \log \ell.$$

Therefore

$$\left(1 - \frac{\ell}{\widetilde{n}}\right)^m \leq e^{-\log \ell} = \frac{1}{\ell}.$$

Hence

$$\frac{m_n}{n} \leq \frac{m}{n} + \frac{1}{\ell}.$$

Since

$$m = \left\lceil \frac{\widetilde{n}}{\ell} \log \ell \right\rceil ,$$

we have

$$m \leq \frac{\widetilde{n}}{\ell} \log \ell + 1.$$

Therefore

$$\frac{m}{n} \leq \frac{\widetilde{n}}{n} \frac{\log \ell}{\ell} + \frac{1}{n}.$$

Recall that

$$\widetilde{n} = n + D.$$

Thus

$$\frac{\widetilde{n}}{n} = 1 + \frac{D}{n}.$$

So

$$\frac{m_n}{n} \leq \left(1 + \frac{D}{n}\right) \frac{\log \ell}{\ell} + \frac{1}{n} + \frac{1}{\ell}.$$

Now let $n \to \infty$. Since D and ℓ are fixed,

$$\frac{D}{n} \to 0 \qquad \text{and} \qquad \frac{1}{n} \to 0.$$

Therefore

$$d(E) = \lim_{n\to\infty} \frac{m_n}{n} \leq \frac{\log \ell}{\ell} + \frac{1}{\ell}.$$

Hence

$$\boxed{d(E) \leq \frac{1 + \log \ell}{\ell}.}$$

Since $\ell = |E|$, this becomes

$$\boxed{d(E) \leq \frac{1 + \log |E|}{|E|}.}$$

Combining this with the lower bound

$$d(E) \geq \frac{1}{|E|},$$

we obtain

$$\boxed{\frac{1}{|E|} \leq d(E) \leq C\frac{1 + \log |E|}{|E|}.}$$

In fact, with the normalization above, one may take $C = 1$ up to harmless endpoint conventions.

18.4 Final two-sided estimate

Combining the lower and upper estimates gives

$$\boxed{\frac{1}{\ell} \leq d(E) \leq C\frac{1 + \log \ell}{\ell}.}$$

18.5 Interpretation

The lower bound says that one cannot beat $1/\ell$, because one translate covers at most ℓ points. The upper bound says that this lower bound is always achievable up to a logarithmic factor.

Thus every finite set E of size ℓ can cover $\mathbb{Z}$ with density no worse than about

$$\frac{\log \ell}{\ell}.$$

19 Sharpness of the Logarithmic Bound

19.1 The natural question

We have proved

$$d(E) \leq C\frac{1+\log|E|}{|E|}.$$

A natural question is whether the logarithm is an artifact of the proof. Could it be true that

$$d(E) \leq \frac{C}{|E|}$$

for every finite set $E \subset \mathbb{Z}$?

The answer is no. The logarithmic factor is necessary in general.

19.2 Main sharpness statement

There exist finite sets $E \subset \mathbb{Z}$ such that

$$\boxed{d(E) \geq c\frac{\log|E|}{|E|}.}$$

Therefore, in the worst case,

$$\boxed{d(E) \asymp \frac{\log|E|}{|E|}.}$$

19.3 Why regular sets do not work

If

$$E = \{0, 1, 2, \ldots, \ell - 1\},$$

then one translate covers a whole block of ℓ consecutive integers. The covering density is essentially

$$\frac{1}{\ell}.$$

So this example does not force a logarithm. To make covering difficult, E must be more irregular or spread out.

19.4 A probabilistic construction showing sharpness

Proposition 19.1 (Sharpness of the logarithmic upper bound)**.** *There exists a universal constant $c > 0$ such that for every integer $\ell \geq 1$, one can choose a finite set*

$$E \subset \mathbb{Z}$$

with

$$|E| = \ell$$

such that

$$\boxed{d(E) \geq c\frac{1+\log \ell}{\ell}.}$$

Proof. It is enough to prove the estimate for large ℓ, since small values of ℓ can be absorbed into the constant c.

We shall construct E probabilistically.

Consider the interval

$$[0, 2\ell - 1]_{\mathbb{Z}}.$$

Choose a random subset

$$E \subset [0, 2\ell - 1]_{\mathbb{Z}}$$

by selecting each integer independently with probability $1/2$.

Then

$$\mathbb{E}|E| = \ell.$$

Moreover, by symmetry of the binomial distribution,

$$\mathbb{P}(|E| \geq \ell) \geq \frac{1}{2}.$$

We shall show that, with positive probability, this random set E cannot cover the interval

$$[0, 2\ell - 1]_{\mathbb{Z}}$$

using too few shifts.

Fix a set of shifts

$$Z \subset \mathbb{Z}$$

with

$$|Z| = m.$$

Only shifts z in the range

$$[-2\ell + 1, 2\ell - 1]_{\mathbb{Z}}$$

can matter for covering $[0, 2\ell - 1]_{\mathbb{Z}}$. Indeed, since

$$E \subset [0, 2\ell - 1]_{\mathbb{Z}},$$

if $z + E$ intersects $[0, 2\ell - 1]_{\mathbb{Z}}$, then necessarily

$$-2\ell + 1 \leq z \leq 2\ell - 1.$$

Thus the number of possible choices of Z with $|Z| = m$ is at most

$$\binom{4\ell}{m} \leq (4\ell)^m.$$

Now fix such a set Z. For a point

$$x \in [0, 2\ell - 1]_{\mathbb{Z}},$$

we have

$$x \in Z + E$$

if and only if

$$x - z \in E$$

for some $z \in Z$. Equivalently,

$$(x - Z) \cap E \neq \varnothing.$$

Therefore the event $x \in Z + E$ is completely determined by the membership of the points in

$$x - Z = \{x - z : z \in Z\}$$

inside the random set E.

Since $|Z| = m$, there are at most m possible points $x - z$. Hence

$$\mathbb{P}(x \notin Z + E) \geq 2^{-m}.$$

Consequently,

$$\mathbb{P}(x \in Z + E) \leq 1 - 2^{-m}.$$

We now choose many points $x_1, \ldots, x_k \in [0, 2\ell - 1]_{\mathbb{Z}}$ so that the sets

$$x_j - Z$$

are pairwise disjoint.

This can be done greedily. Indeed, once x_j is chosen, it rules out only those x for which

$$(x - Z) \cap (x_j - Z) \neq \varnothing.$$

This means that for some $z, z' \in Z$,

$$x - z = x_j - z',$$

or equivalently,

$$x = x_j + z - z'.$$

There are at most m^2 such forbidden values of x. Therefore we may choose at least

$$k \geq \frac{2\ell}{m^2}$$

points $x_1, \ldots, x_k$, up to harmless integer-part errors, such that the sets $x_j - Z$ are pairwise disjoint.

For these chosen points, the events

$$x_j \in Z + E$$

are independent, because they depend on disjoint sets of random choices defining E.

Therefore

$$\mathbb{P}\left(Z + E \supset [0, 2\ell - 1]_{\mathbb{Z}}\right) \leq \mathbb{P}(x_j \in Z + E \text{ for all } j = 1, \ldots, k).$$

By independence,

$$\mathbb{P}(x_j \in Z + E \text{ for all } j = 1, \ldots, k) \leq (1 - 2^{-m})^k.$$

Using

$$k \geq \frac{2\ell}{m^2},$$

we get

$$\mathbb{P}\left(Z + E \supset [0, 2\ell - 1]_{\mathbb{Z}}\right) \leq (1 - 2^{-m})^{2\ell/m^2}.$$

Since

$$1 - u \leq e^{-u},$$

we have

$$(1 - 2^{-m})^{2\ell/m^2} \leq \exp\left(-2^{-m}\frac{2\ell}{m^2}\right).$$

Now take the union bound over all possible sets Z with $|Z| = m$. The probability that there exists some Z with $|Z| = m$ such that

$$Z + E \supset [0, 2\ell - 1]_{\mathbb{Z}}$$

is at most

$$(4\ell)^m \exp\left(-2^{-m}\frac{2\ell}{m^2}\right).$$

Equivalently, this is

$$\exp\left(m\log(4\ell) - 2^{-m}\frac{2\ell}{m^2}\right).$$

We now choose

$$m = \lfloor \delta \log_2 \ell \rfloor,$$

where $0 < \delta < 1$ is a small absolute constant.

Then

$$2^{-m} \approx \ell^{-\delta}.$$

Hence

$$2^{-m}\frac{2\ell}{m^2} \approx \frac{2\ell^{1-\delta}}{(\log_2 \ell)^2}.$$

On the other hand,

$$m\log(4\ell) \approx \delta(\log_2 \ell)\log(4\ell),$$

which grows only like a constant times $(\log \ell)^2$.

Since

$$\frac{\ell^{1-\delta}}{(\log \ell)^2} \to \infty$$

as $\ell \to \infty$, the negative term dominates. Therefore, for all sufficiently large ℓ,

$$(4\ell)^m \exp\left(-2^{-m}\frac{2\ell}{m^2}\right) < \frac{1}{2}.$$

Thus, with probability greater than $1/2$, no collection of m shifts covers $[0, 2\ell - 1]_{\mathbb{Z}}$.

Also,

$$\mathbb{P}(|E| \geq \ell) \geq \frac{1}{2}.$$

Hence with positive probability, both of the following occur:

$$|E| \geq \ell,$$

and no m shifts of E cover $[0, 2\ell - 1]_{\mathbb{Z}}$.

Choose one such set E. If $|E| > \ell$, choose any subset

$$E' \subset E$$

with

$$|E'| = \ell.$$

Since

$$E' \subset E,$$

any cover using shifts of E' would also be a cover using shifts of E. Therefore E' also cannot cover $[0, 2\ell - 1]_{\mathbb{Z}}$ using m shifts.

Replacing E by E', we have constructed a set $E \subset \mathbb{Z}$ with

$$|E| = \ell$$

such that

$$m_{2\ell} > m.$$

Thus

$$m_{2\ell} \geq m \geq c_1 \log_2 \ell$$

for some absolute constant $c_1 > 0$.

We now pass from this finite interval estimate to a lower bound for the density $d(E)$.

Since

$$E \subset [0, 2\ell - 1]_{\mathbb{Z}},$$

we have

$$\operatorname{diam}(E) \leq 2\ell - 1.$$

For $n \in \mathbb{N}$, consider the interval

$$[0, 4\ell n - 1]_{\mathbb{Z}}.$$

Inside this interval, choose n disjoint subintervals of length 2ℓ, separated from each other by gaps of length 2ℓ. For example, take

$$I_j = [4\ell j, 4\ell j + 2\ell - 1]_{\mathbb{Z}}, \qquad j = 0, 1, \ldots, n - 1.$$

Because the gaps have length 2ℓ, and because every translate of E has diameter at most $2\ell - 1$, no single translate of E can meet two different intervals I_j.

Therefore, to cover all of

$$[0, 4\ell n - 1]_{\mathbb{Z}},$$

one must cover each I_j separately. Each I_j has length 2ℓ, so covering I_j requires at least

$$m_{2\ell}$$

shifts.

Hence

$$m_{4\ell n} \geq n\, m_{2\ell}.$$

Dividing by $4\ell n$, we obtain

$$\frac{m_{4\ell n}}{4\ell n} \geq \frac{n\, m_{2\ell}}{4\ell n} = \frac{m_{2\ell}}{4\ell}.$$

Using

$$m_{2\ell} \geq c_1 \log_2 \ell,$$

we get

$$\frac{m_{4\ell n}}{4\ell n} \geq c_2 \frac{\log_2 \ell}{\ell}$$

for another absolute constant $c_2 > 0$.

Now let

$$n \to \infty.$$

Since

$$d(E) = \lim_{N\to\infty} \frac{m_N}{N},$$

we obtain

$$d(E) \geq c_2 \frac{\log_2 \ell}{\ell}.$$

Changing from $\log_2$ to the natural logarithm only changes the constant. Hence

$$d(E) \geq c \frac{\log \ell}{\ell}.$$

Finally, since for $\ell \geq 2$,

$$1 + \log \ell \leq 2 \log \ell,$$

after adjusting the constant $c > 0$, we get

$$\boxed{d(E) \geq c \frac{1 + \log \ell}{\ell}.}$$

This proves the proposition. □

19.5 Conclusion

The logarithm in the upper bound is not merely a weakness of the proof. In general it is necessary:

$$\boxed{\exists E \subset \mathbb{Z} \text{ finite such that } d(E) \geq c \frac{\log |E|}{|E|}.}$$

Together with the previous upper bound, this gives the worst-case order

$$\boxed{d(E) \asymp \frac{\log |E|}{|E|}.}$$

20 Rogers' Covering Theorem

20.1 From $\mathbb{Z}$ to $\mathbb{R}^n$

The final topic moves from discrete covering to continuous covering.

Previously, we covered intervals of integers using translates of a finite set

$$E \subset \mathbb{Z}.$$

Now we cover Euclidean space using translates of a convex body.

Let

$$K \subset \mathbb{R}^n$$

be a convex body, meaning that K is compact, convex, and has nonempty interior.

A translate of K is

$$x + K = \{x + y : y \in K\}.$$

We want to cover $\mathbb{R}^n$ by translates of K:

$$\mathbb{R}^n = \bigcup_i (x_i + K).$$

In the notes, the proof is written for the Euclidean unit ball

$$B = B_1(0) \subset \mathbb{R}^n.$$

So we first prove the result for B. The same idea is the basis of Rogers' theorem for general convex bodies.

20.2 Covering density in $\mathbb{R}^n$

Let

$$B_r = B_r(0)$$

denote the Euclidean ball of radius r centered at the origin.

Suppose

$$\mathbb{R}^n = \bigcup_{z \in Z} (z + B).$$

The covering density of this covering is measured by

$$\limsup_{r \to \infty} \frac{\#\{z \in Z : (z + B) \cap B_r \neq \varnothing\} \operatorname{vol}(B)}{\operatorname{vol}(B_r)}.$$

For the unit ball B, define

$$\theta(B) = \inf_{Z} \limsup_{r\to\infty} \frac{\#\{z \in Z : (z+B) \cap B_r \neq \varnothing\} \operatorname{vol}(B)}{\operatorname{vol}(B_r)},$$

where the infimum is over all coverings

$$\mathbb{R}^n = \bigcup_{z\in Z} (z+B).$$

The trivial lower bound is

$$\theta(B) \geq 1.$$

Indeed, to cover B_r, the total volume of the relevant unit balls must be at least $\operatorname{vol}(B_r)$. Therefore

$$\#\{z \in Z : (z+B) \cap B_r \neq \varnothing\} \operatorname{vol}(B) \geq \operatorname{vol}(B_r).$$

20.3 Rogers' theorem

Theorem 20.1 (Rogers' covering theorem, ball version)**.** *There exists a universal constant $C > 0$ such that*

$$\boxed{\theta(B) \leq Cn \log n.}$$

More generally, Rogers' theorem says that for every convex body $K \subset \mathbb{R}^n$,

$$\boxed{\theta(K) \leq Cn \log n.}$$

The proof below explains the probabilistic method behind the theorem.

20.4 Choosing random centers

Fix

$$r > 0$$

large, and let

$$0 < \delta < 1.$$

Choose

$$m$$

points

$$z_1, \ldots, z_m$$

independently and uniformly from the enlarged ball

$$B_{r+1}.$$

Let

$$Z' = \{z_1, \dots, z_m\}.$$

Instead of first covering by balls of radius 1, we first look at the smaller balls

$$z + (1-\delta)B.$$

Define the random uncovered set

$$R = B_{r+\delta} \setminus \bigcup_{z \in Z'} \big(z + (1-\delta)B\big).$$

So R is the part of $B_{r+\delta}$ not covered by the smaller balls.

20.5 Probability that a point is uncovered

Fix

$$x \in B_{r+\delta}.$$

If a center z lies in

$$x - (1-\delta)B,$$

then

$$x \in z + (1-\delta)B.$$

Also,

$$x - (1-\delta)B \subset B_{r+1},$$

because

$$|x| \le r + \delta$$

and every point of $(1-\delta)B$ has norm at most $1-\delta$. Hence

$$r + \delta + (1-\delta) = r + 1.$$

Therefore

$$\mathbb{P}\{x \text{ is covered by one random small ball}\} = \frac{\operatorname{vol}((1-\delta)B)}{\operatorname{vol}(B_{r+1})}.$$

Since

$$\operatorname{vol}((1-\delta)B) = (1-\delta)^n \operatorname{vol}(B)$$

and

$$\operatorname{vol}(B_{r+1}) = (r+1)^n \operatorname{vol}(B),$$

we get

$$\mathbb{P}\{x \text{ is covered by one random small ball}\} = \frac{(1-\delta)^n}{(r+1)^n}.$$

Thus

$$\mathbb{P}\{x \text{ is not covered by one random small ball}\} = 1 - \frac{(1-\delta)^n}{(r+1)^n}.$$

Since the m centers are chosen independently,

$$\mathbb{P}\{x \in R\} = \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m.$$

20.6 Expected uncovered volume

By Fubini's theorem,

$$\mathbb{E}[\text{vol}(R)] = \int_{B_{r+\delta}} \mathbb{P}\{x \in R\}\, dx.$$

Hence

$$\mathbb{E}[\text{vol}(R)] = \text{vol}(B_{r+\delta}) \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m.$$

Since

$$\text{vol}(B_{r+\delta}) = (r+\delta)^n \,\text{vol}(B),$$

we have

$$\mathbb{E}[\text{vol}(R)] = (r+\delta)^n \,\text{vol}(B) \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m.$$

Therefore there exists one deterministic choice of centers Z' such that

$$\text{vol}(R) \leq (r+\delta)^n \,\text{vol}(B) \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m.$$

Fix such a choice of Z'.

20.7 The interior of the uncovered region

Define

$$R' = \{x \in R : B_\delta(x) \subset R\}.$$

So R' is the part of the uncovered region whose entire δ-neighborhood is still uncovered.

The important claim is that

$$B_r \setminus R' \subset \bigcup_{z \in Z'} (z + B).$$

Indeed, let

$$x \in B_r \setminus R'.$$

There are two cases.

Case 1: $x \notin R$**.**

Then x is covered by one of the smaller balls:

$$x \in z + (1 - \delta)B$$

for some $z \in Z'$. Since

$$z + (1 - \delta)B \subset z + B,$$

we get

$$x \in z + B.$$

Case 2: $x \in R \setminus R'$**.**

Since $x \notin R'$, the ball $B_\delta(x)$ is not contained in R. Hence there exists

$$y \in B_\delta(x)$$

such that

$$y \notin R.$$

Because $x \in B_r$ and $|x - y| < \delta$, we have

$$y \in B_{r+\delta}.$$

Since $y \notin R$, the point y is covered by one of the smaller balls:

$$y \in z + (1 - \delta)B$$

for some $z \in Z'$.

Thus

$$|y - z| < 1 - \delta.$$

Also

$$|x - y| < \delta.$$

Therefore

$$|x - z| \leq |x - y| + |y - z| < \delta + (1 - \delta) = 1.$$

Hence

$$x \in z + B.$$

Thus, in both cases,

$$x \in \bigcup_{z \in Z'} (z + B).$$

Therefore

$$\boxed{B_r \setminus R' \subset \bigcup_{z \in Z'} (z + B).}$$

20.8 Covering the remaining set by a maximal separated set

Now choose a maximal δ-separated subset

$$Z'' \subset R'.$$

This means that for distinct $z, z' \in Z''$,

$$|z - z'| \geq \delta,$$

and Z'' is maximal with respect to this property.

By maximality,

$$R' \subset \bigcup_{z \in Z''} B_\delta(z).$$

Since $\delta < 1$, we have

$$B_\delta(z) \subset z + B.$$

Therefore

$$R' \subset \bigcup_{z \in Z''} (z + B).$$

Combining this with the previous claim, we get

$$B_r \subset \bigcup_{z \in Z'} (z + B) \cup \bigcup_{z \in Z''} (z + B).$$

Thus

$$Z = Z' \cup Z''$$

gives a covering of B_r by unit balls.

20.9 Packing estimate for the extra centers

Since Z'' is δ-separated, the balls

$$B_{\delta/2}(z), \qquad z \in Z'',$$

are pairwise disjoint.

Moreover, since $z \in R'$, we have

$$B_\delta(z) \subset R.$$

In particular,

$$B_{\delta/2}(z) \subset R.$$

Therefore

$$\#Z'' \operatorname{vol}(B_{\delta/2}) \leq \operatorname{vol}(R).$$

Since

$$\operatorname{vol}(B_{\delta/2}) = \left(\frac{\delta}{2}\right)^n \operatorname{vol}(B),$$

we get

$$\#Z'' \leq \frac{\operatorname{vol}(R)}{\left(\frac{\delta}{2}\right)^n \operatorname{vol}(B)}.$$

Using the bound on $\operatorname{vol}(R)$, we obtain

$$\#Z'' \leq \frac{(r+\delta)^n \operatorname{vol}(B) \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m}{\left(\frac{\delta}{2}\right)^n \operatorname{vol}(B)}.$$

Hence

$$\#Z'' \leq \left(\frac{2}{\delta}\right)^n (r+\delta)^n \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m.$$

20.10 Density estimate for the finite covering

We have

$$\#Z \leq \#Z' + \#Z'' = m + \#Z''.$$

Therefore

$$\frac{\#Z \operatorname{vol}(B)}{\operatorname{vol}(B_r)} \leq \frac{m \operatorname{vol}(B)}{\operatorname{vol}(B_r)} + \frac{\#Z'' \operatorname{vol}(B)}{\operatorname{vol}(B_r)}.$$

Since

$$\operatorname{vol}(B_r) = r^n \operatorname{vol}(B),$$

the first term is

$$\frac{m}{r^n}.$$

For the second term, using the estimate for $\#Z''$, we get

$$\frac{\#Z'' \operatorname{vol}(B)}{\operatorname{vol}(B_r)} \leq \left(\frac{2}{\delta}\right)^n \left(\frac{r+\delta}{r}\right)^n \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m.$$

Therefore

$$\frac{\#Z \operatorname{vol}(B)}{\operatorname{vol}(B_r)} \leq \frac{m}{r^n} + \left(\frac{2}{\delta}\right)^n \left(\frac{r+\delta}{r}\right)^n \left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m.$$

20.11 Choosing the number of random centers

Now choose

$$m = \lfloor \alpha r^n \rfloor,$$

where $\alpha > 0$ is a parameter to be chosen later.

Letting

$$r \to \infty,$$

we get

$$\frac{m}{r^n} \to \alpha,$$

and

$$\left(\frac{r+\delta}{r}\right)^n \to 1.$$

Also,

$$\left(1 - \frac{(1-\delta)^n}{(r+1)^n}\right)^m \to e^{-\alpha(1-\delta)^n}.$$

Hence the limiting density is bounded by

$$\boxed{\alpha + \left(\frac{2}{\delta}\right)^n e^{-\alpha(1-\delta)^n}.}$$

Thus

$$\theta(B) \leq \alpha + \left(\frac{2}{\delta}\right)^n e^{-\alpha(1-\delta)^n}.$$

20.12 Optimizing the parameters

We now choose δ and α.

Take

$$\delta = \frac{1}{n}.$$

Then

$$\left(\frac{2}{\delta}\right)^n = (2n)^n.$$

Also, for $n \geq 2$,

$$\left(1 - \frac{1}{n}\right)^n \geq \frac{1}{4}.$$

Hence

$$\theta(B) \leq \alpha + (2n)^n e^{-\alpha/4}.$$

Choose

$$\alpha = 8n \log(2n).$$

Then

$$(2n)^n e^{-\alpha/4} = (2n)^n e^{-2n\log(2n)} = (2n)^n (2n)^{-2n} = (2n)^{-n} \leq 1.$$

Therefore

$$\theta(B) \leq 8n \log(2n) + 1.$$

Thus, for some universal constant $C > 0$,

$$\boxed{\theta(B) \leq Cn \log n.}$$

This proves Rogers' covering estimate for the Euclidean unit ball.

20.13 Passing from finite balls to all of $\mathbb{R}^n$

The argument above gives, for every large r, a covering of B_r with density at most

$$Cn \log n$$

up to an error tending to 0 as $r \to \infty$.

To obtain a covering of all of $\mathbb{R}^n$, one uses a standard limiting or periodic continuation argument. For example, one can carry out the same proof inside a large cube, periodically repeat the resulting pattern, and then let the cube size tend to infinity.

Therefore there exists a covering

$$\mathbb{R}^n = \bigcup_{z \in Z} (z + B)$$

such that

$$\boxed{\theta(B) \leq Cn \log n.}$$

20.14 Why the logarithm appears

The estimate obtained from the proof is

$$\alpha + \left(\frac{2}{\delta}\right)^n e^{-\alpha(1-\delta)^n}.$$

The first term,

$$\alpha,$$

is the density contribution of the random centers.

The second term,

$$\left(\frac{2}{\delta}\right)^n e^{-\alpha(1-\delta)^n},$$

is the density contribution of the extra centers needed to fill the holes.

If δ is too large, the factor $(1-\delta)^n$ becomes very small, so the random balls do not cover enough.

If δ is too small, the packing factor

$$\left(\frac{2}{\delta}\right)^n$$

becomes too large.

The balance occurs around

$$\delta \sim \frac{1}{n}.$$

Then

$$\left(\frac{2}{\delta}\right)^n \sim (2n)^n,$$

so to kill this factor we need

$$e^{-\alpha/e} \approx (2n)^{-n}.$$

Thus

$$\alpha \sim en \log(2n).$$

This explains why the final bound has the form

$$Cn \log n.$$

20.15 Lower bounds and near-optimality

The trivial lower bound is

$$\theta(B) \geq 1.$$

For high-dimensional balls, one can prove much stronger lower bounds. In particular, there are lower bounds of order

$$cn.$$

Thus Rogers' theorem is close to optimal:

$$cn \leq \theta(B) \leq Cn \log n.$$

The remaining gap is logarithmic.

Exercises

1. A fair coin is tossed three times. Write the sample space and compute the probability of exactly two heads.
2. For a fair die, compute $\mathbb{P}(\{2,4,6\})$.
3. Prove $\mathbb{E}(\mathbb{1}_E) = \mathbb{P}(E)$.
4. Let $X = \mathbf{1}_E$ and $Y = \mathbf{1}_F$. Prove $\mathbb{E}(X+Y) = \mathbb{P}(E) + \mathbb{P}(F)$.
5. Compute the variance of a fair die.
6. Prove $\mathrm{Var}(cX) = c^2 \mathrm{Var}(X)$.
7. Prove Chebyshev's inequality using Markov's inequality.
8. Let X be Bernoulli with $\mathbb{P}(X=1) = p$. Compute $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.
9. Compute $\mathbb{E}(e^{tX})$ for a random sign X.
10. Prove $\cosh t \le e^{t^2/2}$.
11. Derive $\mathbb{P}(|S_n| > a) \le 2e^{-a^2/(2n)}$.
12. Derive the weighted-sign estimate.
13. If $p = 1/4$ and $N = 1000$, compute $\mathbb{E}|S|$ and $\mathrm{Var}(|S|)$.
14. Let $X = \varepsilon_1 + \cdots + \varepsilon_{20}$. Compute $\mathbb{E}X$ and $\mathrm{Var}(X)$.
15. Prove $\mathrm{Var}(\sum a_k \varepsilon_k) = \sum a_k^2$.
16. Use the exponential method to prove the weighted sign tail estimate.
17. Prove $\mathrm{Var}(\sum \varepsilon_k \cos(k\theta)) = \sum \cos^2(k\theta)$.
18. Prove $\mathbb{P}(|p(z)| > u) \le 4e^{-u^2/(8n)}$.
19. Compute P_1, Q_1, P_2, Q_2 for the Rudin-Shapiro recursion.
20. Prove $|A+B|^2 + |A-B|^2 = 2|A|^2 + 2|B|^2$.
21. Deduce $|P_k(z)| \le \sqrt{2N}$.
22. Prove that among any $n+1$ integers, two have the same remainder modulo n.
23. Prove Dirichlet's approximation theorem carefully.
24. Prove the simultaneous approximation theorem.

25. Explain precisely where the pigeonhole principle is used.
26. Show $\mathrm{Var}(Y_i) = \sum_j a_{ij}^2$.
27. Derive $\mathbb{P}(|Y_i| > \lambda) \le 2e^{-\lambda^2/(2n)}$.
28. Use the union bound to prove $D(\varepsilon) \lesssim \sqrt{n \log m}$.
29. Explain where the pigeonhole principle enters Spencer's proof.
30. Why is the condition $m \le n$ important in the covering argument?
31. Prove that $\mathbb{P}_B(A) = \mathbb{P}(A \mid B)$ is a probability measure.
32. Prove the law of total probability.
33. Derive Bayes' theorem.
34. For a fair die, let $X(i) = i$ and $B = \{1, 3, 5\}$. Compute $\mathbb{E}(X \mid B)$.
35. Show that $\mathbb{E}(\mathbb{1}_A \mid B) = \mathbb{P}(A \mid B)$.
36. Verify $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F})) = \mathbb{E}(X)$ directly in a four-point probability space.
37. Prove the tower property for finite partitions.
38. Draw the first four levels of the dyadic filtration.
39. Compute f_0, f_1, f_2 for $f(x) = x$.
40. Explain why f_k is constant on each dyadic interval of length 2^{-k}.
41. Show that simple random walk is a martingale.
42. Prove that martingale differences have conditional mean zero.
43. Prove $\mathbb{E}X_n = \mathbb{E}X_0$ for any martingale.
44. Compute f_0, f_1, f_2 for $f(x) = x$.
45. Verify directly that $\mathbb{E}(f_{k+1} \mid \mathcal{F}_k) = f_k$ for the dyadic martingale.
46. Prove $\mathbb{E}(\Delta_k f \mid \mathcal{F}_{k-1}) = 0$.
47. Prove the orthogonality relation.
48. Derive the L^2 identity.
49. Verify $S_k^2 = S_{k-1}^2 + a^2$ in the local model.
50. Explain why orthogonality is no longer enough for $p \neq 2$.

51. Verify the formula for $\mathbb{E}(X_k \mid \mathcal{F}_{k-1})$.
52. Show that $\dfrac{(x+a)^2+(x-a)^2}{2} = x^2 + a^2$.
53. Carry out the Taylor expansion of $(1+u)^{p/2}$.
54. Complete the Holder step carefully.
55. Explain why choosing A large makes X_k a supermartingale.
56. Compute $\mathbb{E}(X_k \mid \mathcal{F}_{k-1})$ for the $1 < p < 2$ Burkholder function.
57. Explain why the linear terms cancel.
58. Prove that $|x|^p$ is convex for $p > 1$.
59. Explain why the proof gives only $\|S(f)\|_p \le C_p \|f\|_p$.
60. Prove the identity $\int fg = \sum_k \int \Delta_k f \Delta_k g$.
61. Apply pointwise Cauchy–Schwarz to the sum over k.
62. Complete the duality argument.
63. Explain why the reverse inequality uses the upper square-function estimate for exponent q.
64. Prove $S(Tf) = S(f)$ when $\varepsilon_k = \pm 1$.
65. Use Littlewood–Paley to prove $\|Tf\|_p \le C_p \|f\|_p$.
66. Write a martingale transform in Haar notation.
67. Explain why this is a model for singular integral boundedness.
68. State the definition of $T(h_I)$.
69. Explain why T resembles the Hilbert transform.
70. Prove the L^p boundedness of T using Littlewood–Paley.
71. Define the principal value integral for Hf.
72. Explain why the kernel $1/(x-y)$ is singular.
73. State how translations and dilations of dyadic grids enter the proof.
74. Explain why uniform L^p bounds for dyadic operators imply the L^p bound for their average.
75. Compute m_n and $d(E)$ for $E = \{0\}$.

76. Compute m_n and $d(E)$ for $E = \{0, 1\}$.

77. Prove directly that $m_{n+n'} \leq m_n + m_{n'}$.

78. Prove that $d(E) = \inf_{n \geq 1} m_n/n$.

79. Explain why boundary errors in local coverings are controlled by $\operatorname{diam}(E)$.

80. Prove the lower bound $d(E) \geq 1/\ell$.

81. For a fixed point x, prove that exactly ℓ shifts z satisfy $x \in z + E$.

82. Prove $(1-t)^m \leq e^{-mt}$ for $0 \leq t \leq 1$.

83. Complete the proof of the upper bound using $m = (\widetilde{n}/\ell) \log \ell$.

84. Explain why the term involving $\operatorname{diam}(E)$ disappears in the definition of $d(E)$.

85. Explain why the interval set $E = \{0, 1, \ldots, \ell - 1\}$ has covering density about $1/\ell$.

86. Derive the estimate
$$\mathbb{E}U \approx N \left(1 - \frac{\ell}{N}\right)^m.$$

87. Use $(1-t)^m \approx e^{-mt}$ to explain why m must be at least of order $N \log \ell/\ell$.

88. Explain the analogy with the coupon collector problem.

89. State clearly why this proves the logarithmic factor cannot always be removed.

90. Define the covering density $\theta(K)$ in $\mathbb{R}^n$.

91. Derive the estimate
$$\mathbb{E}U \leq \operatorname{vol}(Q_R) \exp\left(-\frac{m \operatorname{vol}(K)}{\operatorname{vol}(Q_R)}\right).$$

92. Explain why choosing
$$m \approx \frac{\operatorname{vol}(Q_R)}{\operatorname{vol}(K)} n \log n$$
makes the uncovered volume small.

93. Explain the role of maximal separated sets in filling the uncovered region.

94. Compare the proof of Rogers' theorem with the proof of the discrete upper bound for $d(E)$.