Arbitrary topics

Asaf Ferber [*]

June 8, 2020

Contents

[*] Department of Mathematics, UCI. Email: asaff@uci.edu.

1 Useful inequalities and estimates

In this note we will survey few useful inequalities and estimates which we like. Many of the inequalities are summarized from the lovely book of Steele entitled "Cauchy-Schwarz master class", which is highly recommended to read.

1.1 Rearrangement inequality

In this section we prove the following useful inequality

Lemma 1.1 (Rearrangement inequality)**.** *Suppose $-\infty < a_1 \leq a_2 \leq \ldots \leq a_n < \infty$ and $-\infty < b_1 \leq b_2 \leq \ldots \leq b_n < \infty$. Then, for every permutation $\sigma \in S_n$ we have*

$$a_1 b_n + a_2 b_{n-1} + \ldots + a_n b_1 \leq a_1 b_{\sigma(1)} + a_2 b_{\sigma(2)} + \ldots + a_n b_{\sigma(n)} \leq a_1 b_1 + a_2 b_2 + \ldots + a_n b_n.$$

Proof. We will only prove the upper bound and leave the lower bound (which can be obtained as an immediate corollary) as an exercise for the reader.

Let $\sigma \in S_n$ be such that the sum

$$\sum_{i=1}^{n} a_i b_{\sigma(i)}$$

is maximal, and among all permutations which give a maximal sum, σ has the maximum number of fixed points.

We will show that if σ is not the identity permutation, then one can find a permutation τ that attains the maximum and has more fixed points than σ. This will be a contradiction.

Let $1 \leq j \leq n$ be the first index for which $\sigma(j) \neq j$. Since σ is not the identity permutation, j is well defined. Now, let k be such that $\sigma(k) = j$, and observe that, since j is the first point which is not fixed under σ, we have $\sigma(j) > j$, and $k > j$. Therefore, by the assumption on the labeling of the sequences, we have that

1. $a_j \leq a_k$, and
2. $b_j \leq b_{\sigma(j)}$.

In particular, this implies that

$$(a_k - a_j)(b_{\sigma(j)} - b_j) \geq 0,$$

which translates to

$$a_k b_{\sigma j} + a_j b_j \geq a_j b_{\sigma(j)} + a_k b_j.$$

Therefore, the permutation τ which is obtained from σ by setting $\tau(j) = j$ and $\tau(k) = \sigma(j)$, and $\tau(i) = \sigma(i)$ for all $i \neq j, k$ attains the maximum as well, and has more fixed points. Contradiction. □

Exercise 1.2. *Let $\mathcal{S}^n$ denote the n-dimensional unit sphere. Show that for all $x \in \mathcal{S}^n$ we have that*

$$\sum_i x_i^2 = \max_{y \in \mathcal{S}^n} \langle x, y \rangle.$$

1.2 Reverse Cauchy-Schwarz

Here the objective is to determine, under certain circumstances, when can we write an inequality of the following form:

$$\left(\sum_{k=1}^{n} a_k^2\right)^{1/2} \left(\sum_{k=1}^{n} b_k^2\right)^{1/2} \leq \rho \sum_{k=1}^{n} a_k b_k. \tag{1}$$

To get some feeling it makes sense to study the first non-trivial case where $n = 2$ (and for convenience we will normalize both $a_1 = b_1 = 1$). In this case (1) translates to

$$(1 + a^2)(1 + b^2) \leq \rho \cdot ab.$$

Observe that if, for example, we choose $b = \frac{1}{a}$, and we allow $a \to \infty$, then the left hand side is unbounded while the right hand sided is always ρ. In particular, it means that obtaining (1) is hopeless unless we add some constraints on the ratios $\frac{a_k}{b_k}$.

Suppose that $m \leq \frac{a_k}{b_k} \leq M$ holds for all k. In order to obtain a bound that involves both quadratic and linear elements, it might seem natural to consider the following:

$$\left(M - \frac{a_k}{b_k}\right)\left(\frac{a_k}{b_k} - m\right) \geq 0.$$

This translates to

$$a_k^2 + (mM) \cdot b_k^2 \leq (m + M) a_k b_k.$$

Now, summing over all k we obtain that

$$\sum_k a_k^2 + (mM) \sum_k b_k^2 \leq (m + M) \sum_k a_k b_k. \tag{2}$$

Recall the AM-GM inequality

$$\left(\prod_{k=1}^{n} x_k\right)^{1/n} \leq \frac{1}{n}\left(\sum_{k=1}^{n} x_k\right),$$

which by combining with (2) gives

$$\left(\sum_{k=1}^{n} a_k^2\right)^{1/2} \left((mM)\sum_{k=1}^{n} b_k^2\right)^{1/2} \leq \frac{m + M}{2}\left(\sum_k a_k b_k\right),$$

as desired. We summarize the above in the following theorem

Theorem 1.3. *Let $(a_k)_{k=1}^n$ and $(b_k)_{k=1}^n$ be two sequences of non-negative real numbers. Suppose that for some $m \leq M$ we have $m \leq \frac{a_k}{b_k} \leq M$ for all k. Then the following inequality holds:*

$$\left(\sum_{k=1}^{n} a_k^2\right)^{1/2} \left(\sum_{k=1}^{n} b_k^2\right)^{1/2} \leq \frac{m + M}{2\sqrt{mM}}\left(\sum_k a_k b_k\right).$$

1.3 Basic convexity inequalities

Recall that a function $f : [a, b] \to \mathbb{R}$ is *convex* if for all $x, y \in [a, b]$ and for all $p \in [0, 1]$ we have

$$f(px + (1-p)y) \leq pf(x) + (1-p)f(y). \tag{3}$$

This basic property serves as the key property for establishing many relatively simple, but yet powerful, inequalities. Let us start with the most fundamental inequality due to Jensen:

Lemma 1.4 (Jensen's inequality)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is a convex function and suppose that the nonnegative numbers $p_1, \dots, p_n$ satisfy*

$$p_1 + p_2 + \dots + p_n = 1.$$

Then, for all $x_j \in [a, b]$, $j = 1, \dots, n$ we have

$$f\left(\sum_{j=1}^{n} x_j p_j\right) \leq \sum_{j=1}^{n} p_j f(x_j).$$

Proof. Observe that for $n = 2$ Jensen's inequality is just (3) which is just the definition of convexity, and therefore it is natural to proceed by induction. For the induction step, note that:

$$\sum_{j=1}^{n} p_j x_j = p_1 x_1 + (1-p_1) \sum_{j=2}^{n} \frac{p_j}{1-p_1} x_j.$$

Therefore, by the induction hypothesis we have that

$$f\left(\sum_{j=1}^{n} x_j p_j\right) \leq p_1 f(x_1) + (1-p_1) f\left(\sum_{j=2}^{n} \frac{p_j}{1-p_1} x_j\right),$$

which again, by induction hypothesis is at most

$$\sum_{j=1}^{n} p_j f(x_j),$$

as desired. This completes the proof. □

Among other examples, the functions $e^x, x^2, -\log x$ are all convex in their entire domain (can you prove it? if not, wait for the next section).

As a quick application, let us prove the AM-GM (Arithmetic Mean- Geometric Mean) inequality:

Lemma 1.5 (The AM-GM inequality)**.** *Let $x_1, \dots, x_n \geq 0$. Then*

$$(x_1 \cdots x_n)^{1/n} \leq \frac{\sum_{j=1}^{n} x_j}{n}.$$

Proof. Write

$$(x_1 \cdots x_n)^{1/n} = \exp\left(\frac{1}{n} \sum_{j=1}^{n} \log x_j\right),$$

and observe that the function e^x is convex, and therefore, by Jensen's inequality we have that

$$\exp\left(\frac{1}{n}\sum_{j=1}^{n}\log x_j\right) \leq \frac{1}{n}\cdot\sum_{j=1}^{n} e^{\log x_j} = \frac{\sum_{j=1}^{n} x_j}{n},$$

as desired. This completes the proof. □

Exercise 1.6. *Prove the following general version of the AM-GM inequality: Let $x_1, \ldots, x_n \geq 0$, and let $p_1, \ldots, p_n \geq 0$ satisfy $p_1 + \ldots p_n = 1$. Then,*

$$\prod_{j=1}^{n} x_j^{p_j} \leq \sum_{j=1}^{n} p_j x_j.$$

Let us now give a quick application of the AM-GM inequality to achieve another useful inequality.

Lemma 1.7. *Let $a, b, c \geq 0$. Then*

$$ab + ac + bc \leq a^2 + b^2 + c^2.$$

Proof. By the AM-GM inequality we have that

$$ab \leq \frac{1}{2}(a^2 + b^2), \text{ and } ac \leq \frac{1}{2}(a^2 + c^2), \text{ and } bc \leq \frac{1}{2}(b^2 + c^2).$$

Summing up all the above bounds give the desired. □

A more general setting: suppose that $a_1, \ldots, a_n \geq 0$, and let G be any graph on n vertices. Define the operation

$$\Sigma(G; a_1, \ldots, a_n) := \sum_{ij \in E(G)} a_i a_j.$$

For example, observe that the LHS in 1.7 is just $\Sigma(G; a, b, c)$, where $G = K_3$ is the complete graph on 3 vertices.

Note that the above operation can be naturally generalized to hypergraphs as well: given any subset $S \subseteq [n]$, we define $a_S := \prod_{i \in S} a_i$. Therefore, given any *set system* $\mathcal{H} \subseteq 2^{[n]}$ (also referred to as a *hypergraph*), we can define

$$\Sigma(\mathcal{H}; a_1, \ldots, a_n) = \sum_{S \in \mathcal{H}} a_S.$$

Lemma 1.8. *Let H be a k-uniform, d-regular hypergraph on n vertices. Then, for all $a_1, \ldots, a_n \geq 0$ we have*

$$\Sigma(H; a_1, \ldots, a_n) \leq \frac{d}{k}\sum_{j=1}^{n} a_j^k.$$

1.3.1 Differential criterion for convexity

Note that in order to use Jensen's inequality on some function f, we first need to establish the convexity of f. It turns out that even though the definition for convexity is very simple, in most cases it is not that simple to prove that a function f is convex by applying the definition to f directly. The most common way to prove convexity is by applying the following differential criterion:

Lemma 1.9 (Differential criterion for convexity)**.** *Suppose that $f : (a, b) \to \mathbb{R}$ is twice differentiable and that $f''(x) \geq 0$ for all $x \in (a, b)$. Then, f is convex on (a, b).*

Proof. Recall the fundamental theorem of calculus that gives us, for a differentiable function and some x_0 in its domain, that

$$f(x) = f(x_0) + \int_{x_0} f'(t)dt, \tag{4}$$

holds for all x in its domain.

Moreover, since $f''(x) \geq 0$ for all x, we have that $f'(t)$ is a non-decreasing function.

Now, let $x < y \in (a, b)$ and $p \in [0, 1]$. We wish to show that (3) holds. In particular, setting

$$\Delta := pf(x) + (1 - p)f(y) - f(px + (1 - p)y),$$

we wish to show that $\Delta \geq 0$.

To this end, observe that by applying the fundamental theorem of calculus (4) to x, y and $x_0 := px + (1 - p)y$, we obtain that

$$\Delta = p\left(f(x_0) + \int_{x_0}^{x} f'(t)dt\right) + (1 - p)\left(f(x_0) + \int_{x_0}^{y} f'(t)dt\right) - f(x_0),$$

which translates to

$$\Delta = p\int_{x_0}^{x} f'(t)dt + (1 - p)\int_{x_0}^{y} f'(t)dt = (1 - p)\int_{x_0}^{y} f'(t)dt - p\int_{x}^{x_0} f'(t)dt. \tag{5}$$

By monotonicity of f' we obtain the bounds

$$(1 - p)\int_{x_0}^{y} f'(t)dt \geq (1 - p)f'(x_0)(y - x_0) = (1 - p)pf'(x_0)(y - x),$$

and

$$p\int_{x}^{x_0} f'(t)dt \leq pf'(x_0)(x_0 - x) = p(1 - p)(y - x),$$

which are matching bounds. Plugging these estimates into (5) completes the proof. □

Exercise 1.10. *Show that $f(x) = -\log(\cos x)$ is convex in $-\pi/2 < x < \pi/2$.*

Exercise 1.11. *Show that for $-\pi/2 < \theta_i < \pi/2$, $i = 1, \ldots, n$, we have*

$$\prod_{j=1}^{n} \cos(\theta_j) \leq \cos\left(\frac{\theta_1 + \ldots + \theta_n}{n}\right)^n.$$

1.3.2 An application: on the maximum of the product of two edges

As a neat application to Jensen's inequality we solve the following problem:

Problem 1.12. *In an equilateral triangle with area A, the product of any two sides is $(4/\sqrt{3})A$. Show that every triangle has two sides with product which is at least $(4/\sqrt{3})A$.*

Recall that for a triangle with sides of lengths a, b, c and with angles α, β, γ opposite to the sides, respectively, the area A satisfies:

$$A = \frac{1}{2}ab\sin\gamma = \frac{1}{2}ac\sin\beta = \frac{1}{2}bc\sin\alpha.$$

By averaging, we obtain that

$$\frac{1}{3}(ab + ac + bc) = \frac{2A}{3}\left(\frac{1}{\sin\alpha} + \frac{1}{\sin\beta} + \frac{1}{\sin\gamma}\right).$$

Now, observe that the function $f(x) = 1/\sin x$ satisfies $f''(x) = 1/\sin x + 2\frac{\cos^2 x}{\sin^3 x}$ which is positive for all $x \in (0, \pi)$. Therefore, $f(x)$ is convex in $(0, \pi)$. Moreover, since $\alpha + \beta + \gamma = \pi$, by Jensen's inequality we obtain

$$\frac{2}{\sqrt{3}} = \frac{1}{\sin(\pi/3)} \leq \frac{1}{3}\left(\frac{1}{\sin\alpha} + \frac{1}{\sin\beta} + \frac{1}{\sin\gamma}\right).$$

Plugging it into the first equality, we obtain that

$$\frac{1}{3}(ab + ac + bc) \geq \frac{4A}{\sqrt{3}},$$

which gives us the desired.

1.3.3 Estimating the error in Jensen's inequality

Lemma 1.13 (Hölder's Defect Formula). *Let $f : [a, b] \to \mathbb{R}$ be twice differentiable, and assume that $0 \leq m \leq f''(x) \leq M$ holds for all $x \in [a, b]$. Then, for any choice of $a \leq x_i \leq b$, where $i = 1, \ldots, n$, and nonnegative $p_1, \ldots, p_n$ with $p_1 + \ldots + p_n = 1$, there exists $\mu \in [m, M]$ for which*

$$\sum_k p_k f(x_k) - f\left(\sum_k p_k x_k\right) = \frac{1}{4}\mu \sum_{j=1}^{n}\sum_{k=1}^{n} p_j p_k (x_j - x_k)^2.$$

Proof. Probably the main difficulty in approaching this problem is to understand how to use the condition $0 \leq m \leq f''(x) \leq M$.

Define two auxiliary functions $g(x) = \frac{1}{2}Mx^2 - f(x)$ and $h(x) = f(x) - \frac{1}{2}mx^2$. Observe that by the assumption on f'' we have that both g and h are convex (check their second derivatives!). Now, let $\bar{x} = \sum_j p_j x_j$, and observe that, by Jensen's, we have

$$g(\bar{x}) \leq \sum_{j=1}^{n} p_j \left(\frac{1}{2}Mx_j^2 - f(x_j)\right).$$

By rearranging, this bound translates to

$$\left(\sum_j p_j f(x_j)\right) - f(\bar{x}) \leq \frac{1}{2}M\left\{\left(\sum p_j x_j^2\right) - \bar{x}^2\right\} = \frac{1}{2}M\sum p_j (x_j - \bar{x})^2.$$

Doing the same with h, we obtain

$$\left(\sum_j p_j f(x_j)\right) - f(\bar{x}) \geq \frac{1}{2}m\left\{\left(\sum p_j x_j^2\right) - \bar{x}^2\right\} = \frac{1}{2}m\sum p_j(x_j - \bar{x})^2.$$

In order to complete the proof one need to observe that

$$\sum p_j(x_j - \bar{x})^2 = \frac{1}{2}\sum_j\sum_k p_j p_k(x_j - x_k)^2,$$

which is left as an easy exercise. □

2 Combinatorial linear algebra

In this note we introduce few known and original results in what I consider as “combinatorial linear algebra”.

2.1 Background in linear algebra

In this section we provide a (very) brief background in linear algebra.

2.1.1 Some basic notions and results

Suppose that A is an $n \times n$ real-valued matrix and that $x \in \mathbb{R}^n$. Then,

- Its *spectral norm* is defined as $\|A\| := \max_{\|x\|_2=1}\|Ax\|_2$. Moreover, it is well known that $\|A\| = \max_{\|x\|_2=\|y\|_2=1}|x^TAy|$, and that

$$\|A\| = \max\left\{\sqrt{\lambda} \mid \ \lambda \text{ is an eigenvalue of } A^TA\right\}.$$

- $W(A) = \left(\sum_{i,j} a_{ij}^2\right)^{1/2}$ is the *Euclidean norm* of A.
- $tr(A) = \sum_{i=1}^n a_{ii}$ is its *trace*. Moreover, $tr(A) = \sum_i \lambda_i$, where λ_i are all the eigenvalues of A (with multiplicities).

2.1.2 Spectral decomposition

Theorem 2.1 (Spectral decomposition)**.** *Let A be a real-valued, symmetric matrix. Then, A can be decomposed as*

$$A = \sum_{i=1}^n \lambda_i v_i v_i^T,$$

where λ_i's are its eigenvalues and the v_i's are corresponding eigenvectors which form an orthonormal basis.

Proof. Let $v_1, \ldots, v_n$ be any orthonormal basis of $\mathbb{R}^n$. Then, every vector $x \in \mathbb{R}^n$ can be written as

$$x = \sum_{i=1}^n \langle v_i, x\rangle v_i,$$

where $\langle u, v\rangle = u^T v$ is the standard inner product.

Now, observe that for every x we have

$$x = \sum_{i=1}^{n} \langle v_i, x\rangle v_i = \sum_{i=1}^{n} v_i \langle v_i, x\rangle = \left(\sum_{i=1}^{n} v_i v_i^T\right) x,$$

and therefore, we obtain that

$$I_n = \sum_{i=1}^{n} v_i v_i^T.$$

Next, since A is a real-valued, symmetric matrix, one can find an orthonormal basis $v_1, \dots, v_n$ which consists of eigenvectors of A (this fact is called "the spectral theorem"). For such a basis, if we multiply the above identity by A we obtain

$$A = AI_n = \sum_{i=1}^{n} Av_i v_i^T = \sum_{i=1}^{n} \lambda_i v_i v_i^T.$$

This completes the proof. □

2.2 Gram matrices and applications

2.3 Intersection of vector spaces with the boolean hypercube

In this section we investigate problems related to the size of the intersection between a vector subspace of $\mathbb{R}^n$ and the *boolean hypercube* $Q_n := \{0, 1\}^n$.

As a first natural question one should probably ask

Question 2.2. *Suppose $V \subseteq \mathbb{R}^n$ is a subspace of dimension r. How large can $|V \cap Q_n|$ be?*

A simple observation due to Odlyzko answers the above question:

Observation 2.3 (Odlyzko). *Suppose $V \subseteq \mathbb{R}^n$ is a subspace of dimension r. Then,*

$$|V \cap Q_n| \leq 2^r.$$

Proof sketch. Let $v_1, \dots, v_r$ be a basis of V. Since $\dim(V) = r$, all the linear combination of $v_1, \dots, v_r$ depend on r "free" coordinates (and the rest are forced). Therefore, there are at most 2^r many combinations for 0/1 vectors on these coordinates. □

Next, let us try to investigate the structure of such subspaces V with $\dim(V) = r$ and with a maximal intersection $|V \cap Q_n| = 2^r$. As it turns out, these subspaces have a very simple structure.

Lemma 2.4. *Let $V \subseteq \mathbb{R}^n$ be a subspace of dimension r and with $|V \cap Q_n| = 2^r$. Then, there exists a partition $[n] = I_1 \cup \ldots \cup I_{r+1}$ such that for every $v \in V$ and for every $k, \ell \in I_j$ (where $1 \leq j \leq r+1$) we have $v_k = v_\ell$.*

In order to prove Lemma 2.4 we will prove something slightly stronger:

Lemma 2.5. *Let $V \subseteq \mathbb{R}^n$ be a subspace of dimension r, and let $v_1, \dots, v_r$ be any basis of V. Let M be the $r \times n$ matrix with the v_i's as its rows. Then, there exists a partition $[n] = I_1 \cup \ldots \cup I_{r+1}$ such that for all $1 \leq j \leq r+1$, all the columns of M with indices in I_j are the same.*

Exercise 2.6. *Show that Lemma 2.5 indeed implies Lemma 2.4.*

Proof. Given a matrix A, we let $\mathrm{span}(A)$ be the subspace spanned by its rows. In particular, we have that $\mathrm{span}(M) = V$. Now, observe that any matrix M' obtained by applying gaussian elimination to M still has $\mathrm{span}(M') = V$. Moreover, observe that by exchanging columns of M, even though the span of its rows changes, we still have the same size of intersection with Q_n. Therefore, we can assume without loss of generality that the first r columns of M form a non-singular matrix, and then, by applying a guassian elimination to M, we can obtain a matrix M' for which its first r columns form the identity matrix I_r, and for which $|\mathrm{span}(M') \cap Q_n| = |V \cap Q_n| = 2^r$. The main advantage is that, for M', every $v \in Q_r$ is such that $v^T M' \in Q_n$. Finally, observe that every columns has support of size at most 1. Indeed, let $u \in \mathbb{R}^r$, if its support is larger than 1, then it cannot be that $v^T u \in \{0, 1\}$ for all $v \in Q_r$. Moreover, if the support of u is of size 1, then the only non-zero element must be 1. To summarize, in M' every column belongs to $\{e_1, \ldots, e_r, \bar{0}\}$, and therefore, in M, let $u_1, \ldots, u_r$ be the first r columns, then we have that every columns is in $\{u_1, \ldots, u_r, \bar{0}\}$. This completes the proof. □

2.4 Siegel's lemma

Suppose A is an $M \times N$ matrix with $M < N$. Clearly, as $M < N$, there must be non-trivial solution to $Ax = 0$. The problem that we are interested at is in finding a non-trivial solution $x \in \mathbb{R}^N$ for which $\|x\|_{|infty} = \max_{1 \le j \le N}\{|x_j|\}$ is as small as possible. Such a bound is given by the following lemma due to Siegel.

Lemma 2.7 (Siegel's Lemma). *Let*

$$L_m(x) = \sum_{n=1}^{N} a_{mn} x_n, \;\; m = 1, \ldots, M,$$

be M non-trivial linear forms in N variables x_n, and assume that $a_{mn} \in \mathbb{Z}$ for all m, n.

For $A_m := \sum_{n=1}^{N} |a_{mn}|$, $m = 1, \ldots, M$, we have that there exists a non-trivial solution $z \in \mathbb{Z}^N$ to the system of linear equations with

$$1 \le \max_{1 \le n \le N} |z_n| \le \lfloor (A_1 \cdots A_M)^{\frac{1}{N-M}} \rfloor.$$

Proof. Let $A := \lfloor (A_1 \cdots A_M)^{\frac{1}{N-M}} \rfloor$, and observe that

$$A_1 \cdots A_M < (A + 1)^{N-M}.$$

In particular, we have that

$$\prod_{j=1}^{M} (A_j A + 1) \le \prod_{j=1}^{M} (A_j (A + 1)) < (A + 1)^N.$$

Now, consider the N-dimensional box $B_1 := [0, A]^N$ in $\mathbb{Z}^N$, and observe that it contains $(A+1)^N$ many distinct vectors. Moreover, observe that for each $m = 1, \ldots, M$ and for each $x \in B_1$ we have

$$\sum_{a_{mn} < 0} a_{mn} x_n \le L_m(x) \le \sum_{a_{mn} > 0} a_{mn} x_n.$$

Let $N_m := -\sum_{a_{mn}<0} a_{mn}$ and $P_m := \sum_{a_{mn}>0} a_{mn}$ be the absolute values of the sums of the negative and positive coefficients of L_m, respectively. Observe that $N_m + P_m = A_m$ and that

$$-N_m A \leq L_m(x) \leq P_m A,$$

for all $x \in B_1$.

Now, define the box $B_2 := \prod_{m=1}^{M} [-N_m A, P_m A]$ in $\mathbb{Z}^M$, and observe that it contains

$$\prod_{m=1}^{M} ((N_m + P_m)A + 1) = \prod_{m=1}^{M} (A_m A + 1) < (A+1)^N$$

many vectors.

In particular, there are two distinct vectors $x, y \in B_1$ which give the same solution in B_2. This implies that $z := x - y \neq 0$ is also a solution and it clearly satisfies $\|z\|_\infty \leq A$. □

2.5 Quantitative Halász-type inequality

In connection with their study of random polynomials, Littlewood and Offord introduced the following problem. Let $\boldsymbol{a} := (a_1, \ldots, a_n) \in (\mathbb{Z} \backslash \{0\})^n$ and let $\epsilon_1, \ldots, \epsilon_n$ be independent and identically distributed (i.i.d.) Rademacher random variables, i.e., each ϵ_i independently takes values ±1 with probability 1/2 each. Estimate the largest atom probability $\rho(\boldsymbol{a})$, which is defined by

$$\rho(\boldsymbol{a}) := \sup_{x \in \mathbb{Z}} \Pr\left[\epsilon_1 a_1 + \cdots + \epsilon_n a_n = x\right].$$

They showed that $\rho(\boldsymbol{a}) = O(n^{-1/2} \log n)$ for any such $\boldsymbol{a}$. Soon after, Erdős used Sperner's theorem (we will give a simple proof using Fourier's in a later section) to give a simple combinatorial proof of the refinement $\rho(\boldsymbol{a}) \leq \binom{n}{\lfloor n/2 \rfloor}/2^n = O(n^{-1/2})$, which is tight, as is readily seen by taking $\boldsymbol{a}$ to be the all ones vector.

The results of Littlewood–Offord and Erdős generated considerable interest and inspired further research on this problem. One such direction of research was concerned with improving the bound of Erdős under additional assumptions on $\boldsymbol{a}$. The first such improvement was due to Erdős and Moser, who showed that if all coordinates of $\boldsymbol{a}$ are distinct, then $\rho(\boldsymbol{a}) = O(n^{-3/2} \log n)$. Subsequently, Sárkőzy and Szemerédi improved this estimate to $O(n^{-3/2})$, which is asymptotically optimal. Soon afterwards, Halász proved the following very general theorem relating the "additive structure" of the coordinates of $\boldsymbol{a}$ to $\rho(\boldsymbol{a})$.

Theorem 2.8 (Halász's inequality)**.** *Let $\boldsymbol{a} := (a_1, \ldots, a_n) \in (\mathbb{Z} \setminus \{0\})^n$. For an integer $k \geq 1$, let $R_k(\boldsymbol{a})$ denote the number of solutions to $\pm a_{i_1} \pm a_{i_2} \cdots \pm a_{i_{2k}} = 0$, where repetitions are allowed in the choice of $i_1, \ldots, i_{2k} \in [n]$. There exists an absolute constant $C > 0$ such that*

$$\rho(\boldsymbol{a}) \leq \frac{C\sqrt{k} R_k(\boldsymbol{a})}{2^{2k} n^{2k+1/2}} + e^{-n/\max\{k,C\}}.$$

It is easy to see that Halász's inequality, applied with $k = 1$, yields the estimate $\rho(\boldsymbol{a}) = O(n^{-1/2})$ for every $\boldsymbol{a} \in (\mathbb{Z} \backslash \{0\})^n$; if one further assumes that the coordinates of $\boldsymbol{a}$ are distinct, then $R_1(\boldsymbol{a}) \leq 2n$ and one obtains the stronger bound $\rho(\boldsymbol{a}) = O(n^{-3/2})$, recovering the result of Sárkőzy and Szemerédi. We emphasize that Theorem 2.8 is valid even when k grows with n (the constant C does not depend on either k, n, or $\boldsymbol{a}$). This fact will prove to be crucial for the results in this section.

Here, we wish to investigate the structure of vector with large atom probability over finite fields. The starting point for our approach is the anti-concentration inequality of Halász 2.8. For a vector $\boldsymbol{a} \in \mathbb{F}_p^n$, we define $\rho_{\mathbb{F}_p}(\boldsymbol{a})$ and $R_k(\boldsymbol{a})$ as above, except that all arithmetic is done over the p-element field $\mathbb{F}_p$, and we let $\operatorname{supp}(\boldsymbol{a}) = \{i \in [n] : a_i \neq 0 \mod p\}$.

Theorem 2.9 (Halász's inequality over $\mathbb{F}_p$)**.** *There exists an absolute constant C such that the following holds for every odd prime p, integer n, and vector $\boldsymbol{a} := (a_1, \ldots, a_n) \in \mathbb{F}_p^n \setminus \{\mathbf{0}\}$. Suppose that an integer $k \geq 0$ and positive real M satisfy $30M \leq |supp(\boldsymbol{a})|$ and $80kM \leq n$. Then,*

$$\rho_{\mathbb{F}_p}(\boldsymbol{a}) \leq \frac{1}{p} + \frac{CR_k(\boldsymbol{a})}{2^{2k}n^{2k} \cdot M^{1/2}} + e^{-M}.$$

The proof of this theorem is a straightforward adaptation of Halász's original argument. For the reader's convenience, we provide complete details in Section 2.6.

Note that Halász's inequality may be viewed as a *partial inverse Littlewood–Offord theorem.* Indeed, if $\rho_{\mathbb{F}_p}(\boldsymbol{a})$ is "large", then it must be the case that $R_k(\boldsymbol{a})$ is also "large". Hence, an upper bound on the number of vectors $\boldsymbol{a}$ for which $R_k(\boldsymbol{a})$ is "large" is also an upper bound on the number of vectors with "large" $\rho_{\mathbb{F}_p}(\boldsymbol{a})$. Moreover, since $\rho_{\mathbb{F}_p}(\boldsymbol{a}) \leq \rho_{\mathbb{F}_p}(\boldsymbol{b})$ for every subvector $\boldsymbol{b} \subseteq \boldsymbol{a}$, when $\rho_{\mathbb{F}_p}(\boldsymbol{a})$ is "large", so is $R_k(\boldsymbol{b})$ for *every* $\boldsymbol{b} \subseteq \boldsymbol{a}$. As we shall show, the number of vectors $\boldsymbol{a}$ with such "hereditary" property can be bounded from above quite efficiently using direct combinatorial arguments. Consequently, our approach yields strong bounds on the number of vectors $\boldsymbol{a}$ with $\rho_{\mathbb{F}_p}(\boldsymbol{a}) \geq \rho$ for a significantly wider range of ρ than the range amenable to the "structural" approach of Tao-Vu described above (NOT YET, ADD TO NOTES).

Instead of working directly with $R_k(\boldsymbol{a})$, however, we will find it more convenient to work with the following closely related quantity.

Definition 2.10. *Suppose that $\boldsymbol{a} \in \mathbb{F}_p^n$ for an integer n and a prime p and let $k \in \mathbb{N}$. For every $\alpha \in [0, 1]$, we define $R_k^\alpha(\boldsymbol{a})$ to be the number of solutions to*

$$\pm a_{i_1} \pm a_{i_2} \cdots \pm a_{i_{2k}} = 0 \mod p$$

that satisfy $|\{i_1, \ldots, i_{2k}\}| \geq (1 + \alpha)k$.

It is easily seen that $R_k(\boldsymbol{a})$ cannot be much larger than $R_k^\alpha(\boldsymbol{a})$. This is formalized in the following simple lemma.

Lemma 2.11. *For all integers k, n with $k \leq n/2$, any prime p, vector $\boldsymbol{a} \in \mathbb{F}_p^n$, and $\alpha \in [0, 1]$,*

$$R_k(\boldsymbol{a}) \leq R_k^\alpha(\boldsymbol{a}) + \left(40k^{1-\alpha}n^{1+\alpha}\right)^k.$$

Proof. By definition, $R_k(\boldsymbol{a})$ is equal to $R_k^\alpha(\boldsymbol{a})$ plus the number of solutions to $\pm a_{i_1} \pm a_{i_2} \cdots \pm a_{i_{2k}} = 0$ that satisfy $|\{i_1, \ldots, i_{2k}\}| < (1 + \alpha)k$. The latter quantity is bounded from above by the number of sequences $(i_1, \ldots, i_{2k}) \in [n]^{2k}$ with at most $(1 + \alpha)k$ distinct entries times 2^{2k}, which is the number of choices for the ± signs. Thus

$$R_k(\boldsymbol{a}) \leq R_k^\alpha(\boldsymbol{a}) + \binom{n}{(1+\alpha)k}\left((1+\alpha)k\right)^{2k}2^{2k} \leq R_k^\alpha(\boldsymbol{a}) + \left(4e^{1+\alpha}k^{1-\alpha}n^{1+\alpha}\right)^k,$$

where the final inequality follows from the well-known bound $\binom{a}{b} \leq (ea/b)^b$. Finally, noting that $4e^{1+\alpha} \leq 4e^2 \leq 40$ gives us the desired bound. □

The following counting theorem provides an upper bound on the number of sequences $\boldsymbol{a}$ for which every "relatively large" subsequence $\boldsymbol{b}$ has "large" $R_k^\alpha(\boldsymbol{b})$. In particular, the sequences $\boldsymbol{a}$ that are not counted have a "relatively large" subsequence $\boldsymbol{b}$ with "small" $R_k^\alpha(\boldsymbol{b})$ and thus also "small" $R_k(\boldsymbol{b})$ (by Lemma 2.11), and hence small $\rho_{\mathbb{F}_p}(\boldsymbol{b})$ (by Theorem 2.9). Since $\rho_{\mathbb{F}_p}(\boldsymbol{a}) \le \rho_{\mathbb{F}_p}(\boldsymbol{b})$ whenever $\boldsymbol{b} \subseteq \boldsymbol{a}$, each sequence $\boldsymbol{a}$ that is not counted has "small" $\rho_{\mathbb{F}_p}(\boldsymbol{a})$.

Theorem 2.12. *Let p be a prime, let $k, n \in \mathbb{N}$, $s \in [n]$, $t \in [p]$, and let $\alpha \in (0,1)$. Denoting*

$$\boldsymbol{B}_{k,s,\ge t}^{\alpha}(n) := \left\{ \boldsymbol{a} \in \mathbb{F}_p^n : R_k^\alpha(\boldsymbol{b}) \ge t \cdot \frac{2^{2k} \cdot |\boldsymbol{b}|^{2k}}{p} \textit{ for every } \boldsymbol{b} \subseteq \boldsymbol{a} \textit{ with } |\boldsymbol{b}| \ge s \right\},$$

we have

$$|\boldsymbol{B}_{k,s,\ge t}^{\alpha}(n)| \le \left(\frac{s}{n}\right)^{2k-1} (\alpha t)^{s-n} p^n.$$

2.6 Halász's inequality

2.7 VC-dimension and the Sauer-Shelah lemma

Let $\mathcal{F}$ be a collection of subsets of $[n]$. For a subset $S \subseteq [n]$, we define the *projection of $\mathcal{F}$ onto S* as

$$\Pi_{\mathcal{F}}(S) = \{F \cap S \mid F \in \mathcal{F}\}.$$

We say that a set S is *shattered* by $\mathcal{F}$ if $|\Pi_{\mathcal{F}}(S)| = 2^{|S|}$. That is, a set is shattered by $\mathcal{F}$ is and only if every subset of S can be obtained as $F \cap S$ for some $F \in \mathcal{F}$.

With this notation in hands we can define the VC-dimension of $\mathcal{F}$ as follows:

Definition 2.13. *Let $\mathcal{F}$ be a collection of subsets of $[n]$. The* VC-dimension *of $\mathcal{F}$ is defined to be*

$$VC(\mathcal{F}) = \max\{|S| \mid S \textit{ is shattered by } \mathcal{F}\}.$$

The following lemma was first proved by Vapnik-Chervonenkis and rediscovered many times. It is nowadays known as the Sauer-Shelah lemma.

Lemma 2.14 (Sauer-Shelah lemma)**.** *Let $\mathcal{F}$ be a collection of subsets of n with $VC(\mathcal{F}) = d < \infty$, and define*

$$\Pi_{\mathcal{F}}(m) = \max\{|\Pi_{\mathcal{F}}(S)| \mid S \subseteq [n],\ |S| = m\}.$$

Then,

$$\Pi_{\mathcal{F}}(m) \le \sum_{k=0}^{d} \binom{m}{k} = O(m^d).$$

Note that it is often convenient to use the following equivalent form of this lemma

Lemma 2.15 (Sauer-Shelah lemma)**.** *Let $\mathcal{F}$ be a collection of subsets of n with $VC(\mathcal{F}) = d < \infty$. Suppose that*

$$\Pi_{\mathcal{F}}(m) > \sum_{k=0}^{d} \binom{m}{k} = O(m^d),$$

then there exists a subset $S \subseteq [n]$ of size $|S| = d+1$ which is shattered by $\mathcal{F}$.

There are many proofs for this lemma and here we will give lovely proof by Peter Frankl and Janos Pach which is a bit less known.

Proof. Fix a set S of size m, and consider the family of subsets $\mathcal{S} := \Pi_{\mathcal{F}}(S)$. One can naturally view $\mathcal{S}$ as a family of subsets of $[m]$ (simply identify $S = [m]$ arbitrarily). Moreover, as the lemma is trivial for $m = d$, we assume that $m > d$. Let $\binom{[m]}{\leq d}$ denote the collection of all subsets of $[m]$ of size at most d. Clearly, we have

$$\Phi_d(m) := \left|\binom{[m]}{\leq d}\right| = \sum_{k=0}^{d} \binom{m}{k}.$$

For each $A \in \mathcal{S}$ define a function $f_A : \binom{[m]}{\leq d} \to \{0, 1\}$ as follows: $f_A(X) = 1$ if and only if $X \subseteq A$. Note that since the functions f_A can be viewed as vectors in $\mathbb{R}^{\Phi_d(m)}$, it is enough to show that they are linearly independent.

Assume towards a contradiction, that there are coefficients α_A, $A \in \mathcal{S}$, not all of them are 0, for which

$$\sum_{A\in\mathcal{S}} \alpha_A f_A = 0.$$

Our goal is to show that there exists a subset $T \subseteq [m]$ of size at least $d+1$ which is shattered by $\mathcal{F}$. This will clearly be a contradiction.

In order to prove that, we first define, for every $X \subseteq [m]$, the parameter

$$\sigma(X) = \sum_{A\in\mathcal{S}, X\subseteq A} \alpha_A.$$

Next, we crucially observe that for every $X \in \binom{[m]}{\leq d}$ we have

$$0 = \sum_{A\in\mathcal{S}} \alpha_A f_A(X) = \sigma(X).$$

Now, let $T \subseteq [m]$ be the smallest set for which $\sigma(T) \neq 0$ (show that it is well defined!). Clearly, we have that $|T| \geq d + 1$. It thus remain to show that T is shattered by $\mathcal{S}$ (and therefore, it is also shattered by $\mathcal{F}$ which is a contradiction).

To this end, consider any $X \subseteq T$. We wish to show that there exists some $A \in \mathcal{S}$ for which $A \cap T = X$. We show that by proving

$$\sum_{A\in\mathcal{S}, X=A\cap T} \alpha_A \neq 0.$$

Recall the following generalized version of the inclusion-exclusion formula:

$$\sum_{A\in\mathcal{S}, X=A\cap T} \alpha_A = \sum_{X\subseteq Y\subseteq T} (-1)^{|Y-X|}\sigma(Y).$$

Now, since $\sigma(Y) = 0$ for all $X \subseteq Y \subset T$ (recall that T is minimal with respect to $\sigma(T) \neq 0$), we conclude that

$$\sum_{A\in\mathcal{S}, X=A\cap T} \alpha_A = (-1)^{|T-X|}\sigma(T) \neq 0.$$

This completes the proof.

□

3 Cancellation

Given any sum of real or complex numbers, we can always upper bound it by applying the triangle inequality and taking the absolute values of the summands. As this approach is tight whenever all the numbers are positive, it is intuitively clear that in many cases we should be able to do much better! This is the aim of this chapter.

3.1 Abel's inequality

As a first example illustrating the power of cancellations we give the following powerful (and yet simple) inequality due to Abel.

Lemma 3.1 (Abel's inequality)**.** *Let $z_1, \ldots, z_n \in \mathbb{C}$, and let $S_k = z_1 + \ldots + z_k$, for all k. Then, for each sequence of real numbers $a_1 \geq a_2 \geq \ldots \geq a_n \geq 0$ we have*

$$|\sum_i a_i z_i| \leq a_1 \max_{1 \leq k \leq n} |S_k|.$$

Proof. Since $S_k - S_{k-1} = z_k$ for all $k \geq 2$ and $S_1 = z_1$, we have that

$$a_1 z_1 + a_2 z_2 + \ldots + a_n z_n = a_1 S_1 + a_2 (S_2 - S_1) + \ldots + a_n (S_n - S_{n-1}).$$

By rearranging, we obtain that

$$\sum_i a_i z_i = (a_1 - a_2) S_1 + (a_2 - a_3) S_2 + \ldots + (a_{n-1} - a_n) S_{n-1} + a_n S_n.$$

Using the triangle inequality and the fact that $a_j - a_{j+1} \geq 0$ for all j, we obtain

$$|\sum_i a_i z_i| \leq (\max |S_k|) \left((a_1 - a_2) + (a_2 - a_3) + \ldots + (a_{n-1} - a_n) + a_n\right) = a_1 \max |S_k|.$$

This completes the proof. □

3.2 Exponential sums

First, we need some notation: we define the *exponential function* $\mathbf{e}(t) := \exp(2\pi i t)$, where $\exp(2\pi i t) = \cos 2\pi t + i \sin 2\pi t$. Moreover, for every $t \in \mathbb{R}$ we define its *distance from the nearest integer* $\|t\| = \min\{|t - k| \; : \; k \in \mathbb{Z}\}$.

The following observations are going to be our bread and butter:

Observation 3.2. $\int_0^1 \mathbf{e}(t) dt = 0$.

Or, alternatively, in the discrete case, we have

Observation 3.3. $\sum_{k=0}^{n-1} \mathbf{e}(k/n) = 0$.

Observation 3.4. $\frac{\mathbf{e}(t)+\mathbf{e}(-t)}{2} = \cos 2\pi t$*, and* $\frac{\mathbf{e}(t)-\mathbf{e}(-t)}{2i} = \sin 2\pi t$.

The above observation we mostly be used to bring sin or cos into play as follows:

Observation 3.5. *For all $t, k, n \in \mathbb{R}$ we have*

$$\frac{\mathbf{e}(kt)-1}{\mathbf{e}(nt)-1} = \frac{\mathbf{e}(kt/2)}{\mathbf{e}(nt/2)} \cdot \frac{\mathbf{e}(kt/2)-\mathbf{e}(-kt/2)}{\mathbf{e}(nt/2)-\mathbf{e}(-nt/2)} = \frac{\mathbf{e}(kt/2)}{\mathbf{e}(nt/2)} \cdot \frac{\sin \pi kt}{\sin \pi nt}.$$

We will also make use of the following simple observation

Observation 3.6. *Let $z_1, \dots, z_n \in \mathbb{C}$. Then,*

$$|\sum_{k=1}^{n} z_n|^2 = \sum_{k=1}^{n} |z_k|^2 + 2Re \sum_{h=1}^{n-1} \sum_{m=1}^{n-h} z_{m+h} \bar{z}_m.$$

We encourage the reader, as a first warm-up, to prove these observations.

As a first non-trivial example, we prove the following simple bound for linear exponential sums:

Lemma 3.7 (Linear exponential sums). *For all $t \in \mathbb{R}$ and all $M, N \in \mathbb{Z}$ we have*

$$\left| \sum_{k=M+1}^{M+N} \mathbf{e}(kt) \right| \leq \min \left\{ N, \frac{1}{|\sin \pi t|} \right\} \leq \min \left\{ N, \frac{1}{2\|t\|} \right\}.$$

Proof. First, since $|\mathbf{e}(t)| = 1$ for all $t \in \mathbb{R}$, we trivially have that

$$\left| \sum_{k=M+1}^{M+N} \mathbf{e}(kt) \right| \leq N.$$

Second, observe that the above sum is a geometric summation, and therefore we have

$$\sum_{k=M+1}^{M+N} \mathbf{e}(kt) = \mathbf{e}\left((M+1)t\right) \cdot \left(\frac{\mathbf{e}(Nt)-1}{\mathbf{e}(t)-1} \right).$$

Now, by Observation 3.5, the RHS equals

$$\mathbf{e}\left((M+1)t\right) \frac{\mathbf{e}(Nt/2)}{\mathbf{e}(t/2)} \frac{\sin \pi Nt}{\sin \pi t},$$

and therefore, by putting absolute values, we obtain that

$$\left| \sum_{k=M+1}^{M+N} \mathbf{e}(kt) \right| \leq \left| \frac{\sin \pi Nt}{\sin \pi t} \right| \leq \frac{1}{|\sin \pi t|}.$$

Finally, observe that $2\|t\| \leq |\sin \pi t|$ holds for all t (prove it!). □

Then, we prove a similar type of bound for quadratic exponential sums:

Lemma 3.8 (Quadratic exponential sums). *For $b, c \in \mathbb{R}$ and all integers $0 \leq M < N$ we have*

$$\left| \sum_{k=1}^{M} \mathbf{e}\left((k^2+bk+c)/N\right) \right| \leq \sqrt{2N(1+\log N)}.$$

Proof. Given a polynomial $P(k) = \alpha k^2 + \beta k + \gamma$, we want to estimate the sum

$$S_M(P) = \sum_{k=1}^{M} \mathbf{e}(P(k)).$$

Setting $z_k = \mathbf{e}(P(k))$ for all k and $n = M$ in Observation 3.6, we obtain that

$$|S_M(P)|^2 = M + 2\mathrm{Re} \sum_{h=1}^{M-1} \sum_{m=1}^{M-h} \mathbf{e}\left(P(m+h) - P(m)\right).$$

Now, since

$$P(m+h) - P(m) = 2\alpha mh + \alpha h^2 + \beta h,$$

we obtain that

$$|S_M(P)|^2 = M + 2\mathrm{Re} \sum_{h=1}^{M-1} \sum_{m=1}^{M-h} \mathbf{e}\left(\alpha h^2 + \beta h\right) \mathbf{e}\left(2\alpha mh\right)$$

which equals

$$M + 2\mathrm{Re} \sum_{h=1}^{M-1} \mathbf{e}\left(\alpha h^2 + \beta h\right) \left(\sum_{m=1}^{M-h} \mathbf{e}\left(2\alpha mh\right) \right).$$

Putting absolute values and using Lemma 3.7 on the inner sum, we obtain

$$|S_M(P)|^2 \leq M + 2 \sum_{h=1}^{M-1} \frac{1}{|\sin 2\pi h\alpha|} \leq N + \sum_{h=1}^{N-1} \frac{1}{\|2h\alpha\|}.$$

From here it is a simple exercise to complete the proof. □

The above proof relied on the following simple corollary of Observation 3.6

$$\left| \sum_{n=1}^{N} c_n \right|^2 \leq \sum_{n=1}^{N} |c_n|^2 + 2 \sum_{h=1}^{N-1} \left| \sum_{m=1}^{N-h} c_{m+h} \bar{c}_m \right|. \tag{6}$$

It is thus natural to examine the inner sum

$$\rho_N(h) = \sum_{m=1}^{N-h} c_{m+h} \bar{c}_m \text{ for all } 1 \leq h < N.$$

Note that, at least intuitively, if the sums $\rho_N(h)$ are small in average, then we also expect (6) to be small. For example, in the proof of Lemma 3.8 we used the sharp estimates from Lemma 3.7 to upper bound $|\rho_N(h)|$, which in turns gave us the more general upper bound in 3.8. Unfortunately, for general sums, it is harder to bound $\rho_N(h)$ that tightly. The general problem we want to discuss now is the following:

Problem 3.9. *Suppose that $c_1, \ldots, c_N \in \mathbb{C}$ and satisfy:*

- $|c_n| \leq 1$ *for all n, and*
- $\lim_{N\to\infty} \frac{\rho_N(h)}{N} = 0$ *for all $h = 1, 2, \ldots$.*

Does it follow that

$$\lim_{N\to\infty} \frac{|c_1 + \ldots + c_N|}{N} = 0?$$

Observe that (6) is not very helpful, since, for example, if we take a sequence $(c_n)_{n=1}^{\infty}$ with $|\rho_N(h)| = \Theta(hN^{1/2})$, then the conditions in Problem 3.9 are still satisfied but the RHS in (6) is completely useless as it is larger than N^2 (CHECK IT!).

Therefore, in order to solve Problem 3.9 (spoiler, the answer is YES!), we need to do something smarter than the trivial (6). This leads us to prove the following lemma which is due to van der Corput:

Lemma 3.10 (A qualitative van der Corput inequality)**.** *Let $c_1, \ldots, c_N \in \mathbb{C}$, and let $1 \leq H < N$. Then,*

$$\left|\sum_{n=1}^{N} c_n\right|^2 \leq \frac{4N}{H+1}\left(\sum_{n=1}^{N} |c_n|^2 + \sum_{h=1}^{H} |\rho_N(h)|\right).$$

Before proceeding to the proof, try to show that this indeed solves Problem 3.9.

Proof. For simplicity, let us extend the sequence c_n for all $n \in \mathbb{Z}$ by setting $c_n = 0$ for all $n \notin [1, N]$. Then, as a simple exercise we obtain that

$$(H+1)\sum_{n=1}^{N} c_n = \sum_{n=1}^{N+H}\sum_{h=0}^{H} c_{n-h}.$$

Now, let us square this identity and use the triangle inequality to obtain

$$(H+1)^2\left|\sum_{n=1}^{N} c_n\right|^2 = \left|\sum_{n=1}^{N+H}\sum_{h=0}^{H} c_{n-h}\right|^2 \leq \left(\sum_{n=1}^{N+H}\left|\sum_{h=0}^{H} c_{n-h}\right|\right)^2.$$

To continue, observe that by Cauchy-Schwarz on the RHS we obtain that

$$(H+1)^2\left|\sum_{n=1}^{N} c_n\right|^2 \leq (N+H)\sum_{n=1}^{N+H}\left|\sum_{h=0}^{H} c_{n-h}\right|^2.$$

Next, we need to expand the square in the RHS to bring in the $\rho_N(h)$ parameter:

$$\sum_{n=1}^{N+H}\left|\sum_{h=0}^{H} c_{n-h}\right|^2 = \sum_{n=1}^{N+H}\left(\sum_{j=0}^{H} c_{n-j}\sum_{k=0}^{H} \bar{c}_{n-k}\right),$$

which with a bit of algebra (that we omit) is at most

$$\leq (H+1)\sum_{n=1}^{N} |c_n|^2 + 2\sum_{h=1}^{H}(H+1-h)\left|\sum_{n=1}^{N} c_n\bar{c}_{n+h}\right|.$$

Finally, to complete the proof, all we need to do is to bring in the parameters $\rho_N(h)$ and bound the coefficients in the obvious way. This is left as an exercise. □