Normas, desigualdades y dualidad[1]

H. A. Helfgott

[1]Versión preliminar del 3 de junio de 2005

Índice general

Prefacio

NOTACIÓN. Sean f, g funciones definidas en un subconjunto de los reales. Al escribir $g \ll f$, queremos decir que $|g(x)| < c_1|f(x)|$ para todo $x > c_2$, donde c_1 y c_2 son constantes positivas. Por $O(f(x))$ entendemos una función g no especificada tal que $g \ll f$. Por lo tanto, $h(x) = O(f(x))$ es lo mismo que $h \ll f$.

Por $o(f(x))$ denotamos cualquier función g tal que, para todo $\epsilon > 0$, hay un X tal que $|g(x)| < \epsilon|f(x)|$ para todo $x > X$. Finalmente, escribimos $f \sim g$ si, para todo $\epsilon > 0$, hay un X tal que $|f(x) - g(x)| < \epsilon|g(x)|$ para todo $x > X$. (Decimos también que f y g son *asintóticas*.)

Si las constantes c_1, c_2, ϵ ("constantes implícitas") no son en verdad completamente constantes ("constantes absolutas"), sino que dependen de, digamos, A y B, entonces escribimos A y B bajo la relación: $\ll_{A,B}$, $O_{A,B}(\cdots)$; una constante no absoluta se escribe $C_{A,B}$. Lo mismo vale si la relación entre ϵ y X depende de A y B: $o_{A,B}(\cdots)$, $\sim_{A,B}$.

CAPÍTULO 1

Principios básicos

1. La norma ℓ_r. La desigualdad del triángulo.

Consideremos elementos $\vec{x}$ de $\mathbb{R}^n$, esto es, vectores $\vec{x} = (x_1, \cdots, x_n)$ con x_i real, $1 \leq i \leq n$. Recuerden que la longitud de $\vec{x}$ es simplemente $\sqrt{x_1^2 + \cdots + x_n^2}$, por el teorema de Pitágoras. Trabajaremos con un concepto mas general de longitud, llamado *norma*. Para cada $\vec{x} \in \mathbb{R}^n$ y cada número real r que no sea igual a cero, podemos definir la norma[1] ℓ_2 de la manera siguiente:

$$|\vec{x}|_r = \left(\frac{1}{n} \sum_{x=1}^{n} |x_i|^r \right)^{1/r}. \tag{1.1.1}$$

En particular, si $r = 2$, la norma $|\vec{x}|_2$ es igual a la longitud de $\vec{x}$ dividida por $\sqrt{n}$.

En el espacio ordinario, la linea es el camino más corto entre dos puntos; en particular, dados vectores $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_m$, la longitud del vector $\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_m$ es menor o igual a la suma de las longitudes de $\vec{x}_1, \cdots, \vec{x}_m$. La misma aseveracion es cierta de la norma ℓ_r, si es que $r \geq 1$:

$$|\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_m|_r \leq |\vec{x}_1|_r + |\vec{x}_2|_r + \cdots + |\vec{x}_m|_r. \tag{1.1.2}$$

Esta aseveracion lleva el nombre de *desigualdad de Minkowski* o simplemente *desigualdad del triángulo para la norma ℓ_r*.

Si $r > 1$, la desigualdad del triangulo es una igualdad sí y sólo sí $\vec{x}_1, \ldots, \vec{x}_m$ son multiplos no negativos uno del otro, i.e., sí y solo sí todos apuntan en exactamente la misma dirección.

Es fácil probar (1.1.2) para $m > 2$ asumiendo que (1.1.2) es verdad para $m = 2$. Por lo tanto, necesitamos probar (1.1.2) sólo para $m = 2$. Por conveniencia, escribiremos en lo inmediato $\vec{x}$ en vez de $\vec{x}_1$ y $\vec{y}$ en vez de $\vec{x}_2$.

Problema 1 (La prueba original de Minkowski). Muchos problemas clásicos de máximos y mínimos – como, por ejemplo, el problema de encontrar la figura de perímetro dado que rodee un máximo de área – ceden al enfoque siguiente. Digamos que queremos mostrar que el máximo se obtiene sólo en un caso en particular. (La estrategia para encontrar un mínimo es la misma.) Supongamos que se llega al máximo en algún otro caso; el plan es mostrar que una pequeña variación de este supuesto máximo, escogida según nuestra conveniencia, lleva a un incremento en el valor de la función a ser maximizada. Así llegamos a una contradicción.

Entonces no hemos terminado todavía, ya que es posible que una función no tenga un máximo en ningún lugar. Esta posibilidad puede ser eliminada – si es que en verdad no es el caso – mediante el uso de algún afín del hecho siguiente: una función continua restringida a una caja cerrada $\{\vec{y} \in \mathbb{R}^n : |y_i| \leq N\}$ debe adoptar un mínimo y un máximo[2].

Enfoquemos la desigualdad del triángulo como un problema de máximos y mínimos: para $\vec{x}$ fijo, mostrar que $f_x(\vec{y}) = |\vec{x} + \vec{y}|_r - |\vec{y}|_r$ toma su máximo sí y sólo sí $\vec{x} = t\vec{y}$ para algún número $t \geq 0$. Sea $r > 1$.

(a) Reduzca el problema al caso de $\vec{x}, \vec{y} \in (\mathbb{R}_0^+)^n$, i.e., $\vec{x}$, $\vec{y}$ con coordenadas no negativas solamente. Muestre que el caso $r = 1$ es trivial.

[1] Los analistas funcionales generalmente utilizan p en vez de r en este contexto; por lo tanto, hablan de la *norma* ℓ_p, lo cual es la misma cosa que la norma ℓ_r. Seguimos **[6]** en su utilizacion de ℓ_r, en parte para evitar cualquier confusion con los números primos p.

[2] En general, la imágen de un conjunto compacto bajo una función continua es compacto, y un subconjunto de $\mathbb{R}^n$ es compacto sí y sólo sí es cerrado y acotado (teorema de Heine-Borel).

(b) Sea $r > 1$. Fije $\vec{x} \in (\mathbb{R}_0^+)^n$. Tome $\vec{y} \in (\mathbb{R}_0^+)^n$ tal que $\vec{y} \neq t\vec{x}$ para $t \geq 0$. Muestre, entonces, que existe un índice i tal que $\frac{\partial f_{\vec{x}}(\vec{y})}{\partial y_i} < 0$.

(c) Sea $N > 0$ arbitrario. Considere la caja cerrada $\{\vec{y} \in \mathbb{R}^n : |y_i| \leq N\}$. Muestre que el máximo de $f_{\vec{x}}(\vec{y})$ en la caja puede ser alcanzado sólo cuando $\vec{y} = t\vec{x}$ para algún $t \geq 0$. (Se debe considerar el caso de $\vec{y}$ en la superficie de la caja. Use (b) en toda su fuerza.)

(d) Muestre que el máximo de $f_{\vec{x}}(\vec{y})$ dentro de una caja cerrada es alcanzado sí y sólo sí $\vec{y} = t\vec{x}$ para algún $t \geq 0$.

(e) Deduzca de (d) que $f_{\vec{x}}(\vec{y})$ alcanza su máximo dentro de $\mathbb{R}^n$ sí y sólo sí $\vec{y} = t\vec{x}$ para algún $t \geq 0$.

(f) Concluya que la desigualdad del triángulo es cierta, y que es una igualdad sólo cuando $\vec{y} = t\vec{x}$ para algún $t \geq 0$.

La desigualdad del triángulo no se cumple si $r < 1$. En verdad, la desigualdad opuesta es cierta: para $r < 1$ y $\vec{x}_1, \cdots, \vec{x}_m \in (\mathbb{R}_0^+)^n$,

$$|\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_m|_r \geq |\vec{x}_1|_r + |\vec{x}_2|_r + \cdots + |\vec{x}_m|_r, \tag{1.1.3}$$

con igualdad sólo si $\vec{x}_1, \ldots, \vec{x}_m$ son multiplos no negativos uno del otro.

Problema 2. Adapte su solución al problema 1 para probar (1.1.3).

* * *

Sea V un espacio vectorial sobre el cuerpo $k \subset \mathbb{C}$. (Piénsese de $k = \mathbb{Q}$, $k = \mathbb{R}$, $k = \mathbb{C}$.) Una *norma* sobre V es un símbolo $|\cdot|$ que toma valores reales y satisface las propiedades siguientes:

(a) Para cada $\vec{v} \in V$, tenemos $|\vec{v}| \geq 0$. Mas aún, $|\vec{v}| = 0$ sí y sólo sí $\vec{v} = 0$.

(b) $|cv| = |c||v|$ para cada $c \in k$.

(c) Para $\vec{v}, \vec{w} \in V$ cualesquiera, la desigualdad del triángulo se cumple: $|\vec{v} + \vec{w}| \leq |\vec{v}| + |\vec{w}|$.

A un espacio vectorial V con una norma $|\cdot|$ se le llama, naturalmente, *espacio con norma*.

En la terminología dada, ℓ_r es una norma $r \geq 1$ (gracias a (1.1.2)) pero no una norma para $r < 1$ (gracias a (1.1.3)).

En la prueba de Minkowski de la desigualdad del triángulo para ℓ_r, $r \geq 1$, el hecho que V sea una espacio vectorial sobre $\mathbb{R}$ de una manera crucial: tomamos derivadas. Por supuesto, se pueden sacar derivadas sobre $\mathbb{C}$, y, de manera formal, también sobre cuerpos mucho mas generales. Aún así, será bueno tener una prueba de la desigualdad del triángulo que no use derivadas en absoluto. Ver la §4.

2. Comparaciones entre normas

Para cada $\vec{v} \in \mathbb{R}^n$ y todo r, s con $-\infty < r < s < \infty$,

$$|\vec{v}|_r \leq |\vec{v}|_s \quad \text{(desigualdad de Jensen)}, \tag{1.2.1}$$

con igualdad sí y sólo sí $v_1 = v_2 = \cdots = v_n$.

Problema 3 (Prueba de la desigualdad de Jensen). (a) Asumiendo que (1.2.1) es cierta para $\vec{v} \in \mathbb{R}^{n-1}$, reduzca (1.2.1) para $\vec{v} \in \mathbb{R}^n$ a la desigualdad $\left|\frac{n-1}{n}x^r + \frac{y^r}{n}\right|^{1/r} < \left|\frac{n-1}{n}x^s + \frac{y^s}{n}\right|^{1/s}$ para $y > x \geq 0$.

(b) Reduzca $\left|\frac{n-1}{n}x^r + \frac{1}{n}y^r\right|^{1/r} < \left|\frac{n-1}{n}x^s + \frac{1}{n}y^s\right|^{1/s}$ para $y > x \geq 0$ a $\left|\frac{n-1}{n} + \frac{1}{n}u^r\right|^{1/r} < \left|\frac{n-1}{n} + \frac{1}{n}u^s\right|^{1/s}$ para $u > 1$.

(c) Pruebe que, para $u \geq 0$ dado, la función $r \mapsto \left|\frac{n-1}{n} + \frac{1}{n}u^r\right|^{1/r}$ es estrictamente creciente.

Definimos

$$|\vec{v}|_\infty = \lim_{r\to\infty} |\vec{v}|_r, \quad |\vec{v}|_{-\infty} = \lim_{r\to-\infty} |\vec{v}|_r.$$

Problema 4. (a) Pruebe que $|\vec{v}|_\infty = \max_i |v_i|$, $|\vec{v}|_{-\infty} = \min_i |v_i|$.

(b) Muestre que la desigualdad de Jensen vale aún si r, s o ambos son iguales a $-\infty$ o ∞ (y $r < s$).

PROBLEMA 5. Muestre que

$$\lim_{r\to 0^+} |\vec{v}|_r = \lim_{r\to 0^-} |\vec{v}|_r = (|v_1|\cdots|v_n|)^{1/n} \quad \text{(la media geométrica, MG).}$$

Como

$$|\vec{v}|_{-1} = \frac{n}{\frac{1}{|v_1|}+\cdots+\frac{1}{|v_n|}} \quad \text{(la media harmónica, MH)}$$

y

$$|\vec{v}|_1 = \frac{|v_1|+\cdots+|v_n|}{n} \quad \text{(la media aritmética, MA)}$$

La desigualdad de Jensen implica directamente que

$$\text{mín}\,|v_i| \le MH \le MG \le MA \le \text{máx}\,|v_i|,$$

con igualdad sí y solo sí $\vec{v}=0$.

* * *

Dada una function integrable (de Riemann, de Lebesgue...) $f : X \to \mathbb{R}$ en un espacio X de medida 1 (digamos $X=[0,1]$), podemos definir

$$|f|_r = \left(\int_X |f(x)|^r dx\right)^{1/r}. \tag{1.2.2}$$

PROBLEMA 6. Muestre utilizando su proceso de límite favorito que las desigualdades de Minkowski y de Jensen valen para (1.2.2). Mas adelante, muestre que la desigualdad de Hölder (y por lo tanto la de Cauchy) tambien vale para (1.2.2). (Por razones históricas, la desigualdad de Cauchy para (1.2.2) se llama desigualdad de Cauchy-Bunyakovsky-Schwarz, o a veces simplemente Cauchy-Schwarz, a pesar de la prioridad de Bunyakovsky sobre Schwarz en la generalización a la integral (1.2.2). La desigualdad de Cauchy es en verdad más fácil de transferir a (1.2.2) que la de Minkowski o Jensen. Ver el problema 22.

PROBLEMA 7. Estrictamente hablando, el espacio vectorial de todas las funciones integrables f de Riemann o Lebesgue sobre X no es un espacio con norma, ya que $|f|_r$ puede ser 0 aún si f no es idénticamente cero. Cómo puede esto ser remediado?

PROBLEMA 8 (Normalización). Sea la medida de X diferente de 1. Mantenga la definición de $|f|_r$ como en (1.2.2).

(a) Muestre que la desigualdad de Minkowski todavia vale, pero la de Jensen (1.2.1) en general no. (Más tarde, muestre que las desigualdades de Hölder y Cauchy todavía valen cuando la medida de X no es 1.)
(b) Muestre como (1.2.1) puede ser modificada para $|X|\neq 1$, $|X|\neq\infty$ de tal manera que la desigualdad de Jensen valga.
(c) Si $|X|=\infty$, reemplaze (1.2.1) por una desigualdad entre $|f|_1$, $|f|_r$ y $|f|_s$ (para $1<r<s$).
(d) Reinterprete la ℓ_r norma en $\mathbb{R}^n$ (ya sea con su definición (1.1.1) o con la más simple, y a veces usada, $|\vec{x}|_r = |\sum_{i=1}^n x_i^r|^{1/r}$, como un caso especial de la norma (1.2.2) para funciones sobre $X=\{1,2,\ldots,n\}$ con una medida apropiada. Concluir que la desigualdad de Minkowski (pero no la de Jensen) todavía vale para $|\vec{x}|_r = |\sum_{i=1}^n x_i^r|^{1/r}$.

3. Normas y dualidad

En general, una *función linear* f de un espacio vectorial V_1 sobre un cuerpo k a un espacio vectorial V_2 también sobre k es simplemente una function $f : V_1 \to V_2$ tal que[3]

(a) $f(v+w)=f(v)+f(w)$ para todo $v,w\in V_1$,
(b) $f(cv)=cf(v)$ para todo $v\in V_1$ y para todo $c\in k$.

[3]En adelante, omitiremos, como es costumbre, las flechas sobre los vectores a menos que haya riesgo de confusión.

En particular, una función linear f de $\mathbb{R}^n$ a $\mathbb{R}$ es una function $f : \mathbb{R}^n \to \mathbb{R}$ que satisface (a) $f(v+w) = f(v) + f(w)$ para todo $v, w \in \mathbb{R}^n$, y (b) $f(cv) = cf(v)$ para todo $v \in \mathbb{R}^n$ y todo $c \in \mathbb{R}$.

Dados $v, w \in \mathbb{R}^n$, definimos el *producto escalar*

$$\langle v, w\rangle = \frac{1}{n}\sum_{i=1}^{n} v_i w_i. \tag{1.3.1}$$

PROBLEMA 9. (a) Sea $w \in \mathbb{R}^n$. Muestre que la función $f_w : v \mapsto \langle v, w\rangle$ es linear.
(b) Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función linear. Muestre que existe un $w \in \mathbb{R}^n$ tal que la función $v \mapsto \langle v, w\rangle$ de $\mathbb{R}^n$ a $\mathbb{R}$ es idéntica a f.

Acabamos de ver que $\mathbb{R}^n$ puede ser identificado *como conjunto* con el conjunto de funciones lineares de $\mathbb{R}^n$ a $\mathbb{R}$ (también llamado el *dual* de $\mathbb{R}^n$). Es ahora fácil probar que $\mathbb{R}^n$ y el dual de $\mathbb{R}^n$ pueden ser identificados como *espacios vectoriales*.

PROBLEMA 10. (a) Muestre que $f_{cw} = cf_w$ para todo $w \in \mathbb{R}^n$ y todo $c \in \mathbb{R}$.
(b) Muestre que $f_{w_1+w_2} = f_{w_1} + f_{w_2}$ para $w_1, w_2 \in \mathbb{R}^n$ cualesquiera.

Queda por probar que $\mathbb{R}^n$ y el dual de $\mathbb{R}^n$ se pueden identificar como *espacios de normas*. Para ello, debemos escoger una norma ℓ_r para $\mathbb{R}^n$, y también debemos explicar que quiere decir una norma en un espacio vectorial de funciones lineares, tal como el dual de $\mathbb{R}^n$.

Sea $f : V_1 \to V_2$ una función linear de un espacio con norma a otro. Definimos la *norma de operador* de f como

$$|f| = \sup_{v \in V_1, v \neq 0} \frac{|f(v)|}{|v|}, \tag{1.3.2}$$

donde la norma en el numerador es la norma de V_2 y la norma en el denominador es la norma de V_1.

Ahora sean $V_1 = \mathbb{R}^n$ y $V_2 = \mathbb{R}$, donde V_1 tiene la norma de ℓ_r, $r \geq 1$, y a V_2 se le da simplemente el valor absoluto en $\mathbb{R}$ como norma. Tenemos entonces una norma de operador (dada por (1.3.2)) sobre el dual de $\mathbb{R}^n$. Qué es una norma, concretamente?

Veremos que, bajo la identificacion del dual de $\mathbb{R}^n$ con $\mathbb{R}^n$, la norma de operador del dual de $\mathbb{R}^n$ (donde a $\mathbb{R}^n$ se le da la norma ℓ_r) es simplemente la norma ℓ_s sobre $\mathbb{R}^n$, donde $s = \frac{1}{1-\frac{1}{r}}$, o, para decirlo mas simétricamente, s es el número tal que $1 = \frac{1}{r} + \frac{1}{s}$. Decimos que s es el *exponente dual* a r.

PROBLEMA 11. Muestre que lo que debemos probar se puede replantear como sigue: para $v, w \in \mathbb{R}^n$ y todo $r, s \geq 1$ con $\frac{1}{r} + \frac{1}{s} = 1$,

$$|\langle v, w\rangle| \leq |v|_r |w|_s \quad \text{(desigualdad de Hölder)}, \tag{1.3.3}$$

donde la igualdad se logra para por lo menos un v por cada w.

Por supuesto, aún tenemos que probar la desigualdad de Hölder.

PROBLEMA 12. Muestre que la desigualdad de Hölder no depende de la normalización – como la desigualdad de Minkowski, y al contrario de la de Jensen: en otras palabras, la desigualdad de Hölder es equivalente a

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^r\right)^{1/r} \left(\sum_{i=1}^{n} y_i^s\right)^{1/s}.$$

PROBLEMA 13. (a) Muestre que es suficiente probar la desigualdad de Hölder para $n = 2$.
(b) Pruebe que la desigualdad de Hölder para $n = 2$ puede reducirse a la aseveración siguiente: $(1+x)^{1/r} \cdot (1+y)^{1-1/r} \geq 1 + x^{1/r} y^{1-1/r}$ para $x, y \geq 0$, $r \geq 1$.
(c) Muestre que, para $y \geq 0$, $r \geq 1$ fijos, $x \mapsto (1+x)^{1/r} \cdot (1+y)^{1-1/r} - (1 + x^{1/r} y^{1-1/r})$ toma su mínimo (para $x \in [0, 1)$ cuando $x = y$, y que este mínimo es 0.
(d) Concluya que la desigualdad de Hölder es verdadera. Muestre también que es una igualdad solo cuando $|v_i|^r = c|w_i|^s$ para algún $c \in \mathbb{R}$ y todo $i = 1, 2, \cdots, n$.

4. Auto-dualidad, productos escalares y desigualdad de Cauchy

Un caso importante de la desigualdad de Hölder (1.3.3) ocurre cuando $r = s = 2$:

$$|\langle v, w\rangle| \leq |v|_2 |w|_2.$$

En efecto, el dual de $\mathbb{R}^n$ como un espacio con norma ℓ_2 se identifica con $\mathbb{R}^n$ como un espacio con norma ℓ_2. En otras partes, $\mathbb{R}^n$ con norma ℓ_2 es su propio dual.

La desigualdad de Cauchy tiene muchas pruebas independientes de las de Hölder.

PROBLEMA 14. (a) Como antes, reduzca el caso general al caso $n = 2$. Muestre también que el factor de $\frac{1}{n}$ aparece en ambos lados y por lo tanto puede ser eliminado.

(b) Como antes, muestre que es suficiente probar que

$$1 + xy \leq (1 + x^2)^{1/2}(1 + y^2)^{1/2}$$

for $x, y \geq 0$.

(c) Pruebe lo mismo sin utilizar límites. Puede usar el obvio pero extremadamente útil hecho que el cuadrado de un número real es siempre no-negativo[4].

Como $\langle \cdot, \cdot \rangle$ y $|\cdot|$ son el producto escalar usual y la norma Euclideana $|\cdot|_2$, debemos ser capaces de ver la desigualdad de Cauchy en términos geométricos.

PROBLEMA 15. (a) Reduzca la desigualdad de Cauchy al caso donde $v_2 = \cdots = v_n = 0$ y $w_3 = \cdots = w_n = 0$ por medio de rotaciones.

(b) Escriba $v = (a, 0, \ldots, 0)$, $w = (b\cos\alpha, b\sin\alpha, 0, \ldots, 0)$. Muestre que $\langle v, w\rangle = ab\cos\alpha$, $|v|_2 = a$, $|w|_2 = b$. Deduzca la desigualdad de Cauchy.

En esta última prueba, utilizamos el hecho que el producto escalar y la norma Euclideana $|\cdot|_2$ son invariantes bajo rotaciones. En otras palabras, hay un gran *grupo* de transformaciones lineares (es decir, el grupo de rotaciones) que preservan el producto escalar y la norma ℓ_2.

PROBLEMA 16. Muestre que ninguna norma ℓ_r para $r \neq 2$ es invariante bajo un grupo tan rico de transformaciones lineares como el grupo de rotaciones.

* * *

Podemos definir la norma ℓ_2 en terminos del producto escalar en $\mathbb{R}^n$: $|v|_2 = \sqrt{\langle v, v\rangle}$. En general, así como podemos definir un espacio con norma, podemos definir un *espacio con producto escalar*.

Sea V un espacio vectorial sobre un cuerpo $k \subset \mathbb{R}$. Un *producto escalar* sobre V es un símbolo $\langle \cdot, \cdot \rangle$ que toma valores reales y satisface los axiomas siguientes:

(a) $\langle cv, w\rangle = c\langle v, w\rangle = \langle v, cw\rangle$,
(b) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$,
(c) $\langle v, w\rangle = \langle w, v\rangle$,
(d) $\langle v, v\rangle \geq 0$, con igualdad sí y sólo sí $v = 0$.

De (b) o (c) se deduce que $\langle v, 0\rangle = \langle 0, v\rangle = 0$ para todo v. De la condición para la igualdad en (d) se deduce que, para todo $v \neq 0$, hay un w tal que $\langle v, w\rangle \neq 0$; en otras palabras, bajo nuestra definición, el producto escalar debe ser *no degenerado*.

Para todo producto escalar $\langle \cdot, \cdot \rangle$, el símbolo $|v| = \sqrt{\langle v, v\rangle}$ satisface las condiciones para ser una norma. Ahora veremos que esta norma siempre satisface la desigualdad de Cauchy; tendremos una prueba general de la desigualdad de Cauchy, en la cual, necesariamente, usaremos sólo el axioma (d) como única desigualdad de la cual partir (ya que es la única que tenemos).

PROBLEMA 17.
Muestre que $\langle v, w\rangle \leq \frac{1}{2}\langle v, v\rangle + \frac{1}{2}\langle w, w\rangle = \frac{1}{2}|v|^2 + \frac{1}{2}|w|^2$ para $v, w \in V$ cualesquiera. Qué desigualdad podríamos usar para concluir, si solo la desigualdad usada fuera en la dirección opuesta, o una igualdad? Cuando es esa desigualdad una igualdad?

[4]En cierto sentido, una desigualdad es siempre un enunciado *analítico*, en el sentido que utiliza, si no límites, por lo menos el ordenamiento de los reales, o como mínimo el ordenamiento de los enteros. Los enunciados algebraicos tienden a ser igualdades o equivalencias. Para probar un enunciado analítico, se necesita generalmente un germen analítico, por así decirlo; el hecho que el cuadrado de un real sea no-negativo juega el rol de un germen prácticamente mínimo. Como pronto veremos, otros enunciados muy similares pueden jugar el mismo rol.

Por el mismo argumento que en la parte 17, tenemos $\langle v_1, tv_2\rangle \leq \frac{1}{2}|v_1|^2 + \frac{1}{2}|tv_2|^2$. Para que valor de t es la desigualdad aludida anteriormente una igualdad? Escoja t apropiadamente y pruebe Cauchy.

* * *

Surge la siguiente pregunta: cuando es una norma inducida por un producto escalar? En otras palabras, dada una norma $|\cdot|$ en un espacio vectorial V, cuando hay un producto escalar $\langle\cdot,\cdot\rangle$ en V tal que $|v| = \sqrt{\langle v, v\rangle}$ para todo $v \in V$?

La siguiente condición es necesaria y suficiente: para $v, w \in V$ cualesquiera,

$$|v + w| + |v - w| = 2(|v| + |w|). \tag{1.4.1}$$

PROBLEMA 18. (a) Muestre que (1.4.1) es una condición necesaria.
(b) Suponga que (1.4.1) vale. Defina un producto escalar $\langle\cdot,\cdot\rangle$ en términos de $|\cdot|$. Muestre que el producto escalar satisface axiomas (a)–(d), y que, mas aún, es no degenerado.

PROBLEMA 19. Muestre que, si $r \neq 2$, la norma ℓ_r no es inducida por un producto escalar.

* * *

Es posible probar la desigualdad de Hölder utilizando la desigualdad de Cauchy, y la desigualdad de Minkowski a través de la de Hölder, de manera bastante abstracta. El motivo es el siguiente: hay, como he visto, pruebas casi completamente algebraicas de la desigualdad de Cauchy; de esta manera conseguimos pruebas de las desigualdades de Hölder y Minkowski de naturaleza más bien formal. En particular, estas pruebas valdrán sin la menor alteración para las normas dadas por integrales – sin necesidad de engorrosos procesos de límite[5].

Podemos ver tanto las normas (1.1.1) y el producto escalar (1.3.1), así como las normas y productos escalares no normalizados

$$|x|_r = \left(\sum_i |x_i|^r\right)^{1/r}, \qquad \langle v, w\rangle = \sum_{i=1}^n v_i w_i \tag{1.4.2}$$

pueden verse como casos especiales de las normas y el producto escalar para integrales de Lebesgue. Para obtener (1.1.1) y (1.3.1), definimos $X = \{1, 2, \ldots, n\}$ y escogemos para X la medida μ tal que $\mu(x) = \frac{1}{n}$; para obtener (1.4.2), definimos $X = \{1, 2, \ldots, n\}$ y escogimos $\mu(x) = 1$.

De una vez extenderemos nuestra definición del producto escalar a espacios vectoriales complejos. Un producto escalar en un espacio vectorial complejo V es una función $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$ que satisface los mismos axiomas que un producto escalar en un espacio vectorial real, con las siguientes diferencias:

(a) $\langle cv, w\rangle = \overline{c}\langle v, w\rangle$, en vez de $\langle v, cw\rangle = c\langle v, w\rangle$,
(b) $\langle v, w\rangle = \overline{\langle w, v\rangle}$, en vez de $\langle v, w\rangle = \langle w, v\rangle$.

Denotamos por $\overline{z}$ el *conjugado* de un número complejo $z = a + bi$, esto es, $\overline{z} = a - bi$. Como antes, $\langle v, v\rangle$ es un real no negativo, y es igual a cero sólo sí $v = 0$.

PROBLEMA 20. La prueba de la desigualdad de Cauchy esbozada en el problema 17 se transfiere con facilidad a los espacios vectoriales complejos. Verifique esta aseveración.

PROBLEMA 21. Para $v, w \in \mathbb{C}^n$, defina el producto escalar como sigue:

$$\langle v, w\rangle = \frac{1}{n}\sum_{j=1}^n \overline{v_j} w_j. \tag{1.4.3}$$

Muestre que el producto así definido satisface todos los axiomas del producto escalar, con las modificaciones listadas arriba para el caso complejo. Porqué es que (1.3.1) es un caso particular de (1.4.3)?

[5]Si r es irracional, siempre sera necesario algún proceso de límite en la prueba de por lo menos algunas aseveraciones sobre la norma $|\cdot|_r$. Empero, este sera un proceso de límite sobre los reales, y no sobre las funciones integrables.

Sea X un espacio de medida finita. Podemos definir, como antes, la norma ℓ_r sobre las funciones $f : X \to \mathbb{C}$:

$$|f|_r = \left(\int_X |f(x)|^r \right)^{1/r}. \tag{1.4.4}$$

Tambien definimos el producto escalar de $f, g : X \to \mathbb{C}$:

$$\langle f, g \rangle = \int_X \overline{f(x)} g(x) dx. \tag{1.4.5}$$

Ya que las normas y el producto escalar en $\mathbb{R}^n$ son casos especiales de (1.4.4) y (1.4.5), trabajaremos con integrales en el resto de la sección.

PROBLEMA 22. Muestre que (1.4.4) y (1.4.5) satisfacen los axiomas de normas y productos escalares. La desigualdad de Cauchy para las integrales (1.4.4) y (1.4.5) se deduce directamente de la prueba abstracta en el problema 17, la cual utiliza solamente los axiomas mencionados.

PROBLEMA 23 (Hölder a través de Cauchy). (a) Utilizando la desigualdad de Cauchy repetidamente, muestre que, para $m \geq 1$, $u = 2^m$,

$$\left(\int_X |f_1(x)| \cdots |f_u(x)| \right)^u \leq \left(\int_X |f_1(x)|^u \right) \cdot \left(\int_X |f_2(x)|^u \right) \cdots \left(\int_X |f_u(x)|^u \right),$$

donde $f_1, \ldots, f_u : X \to \mathbb{C}$ son funciones integrables cualesquiera.

(b) Muestre que la desigualdad de Hölder vale para r de la forma $r = \frac{2^m}{k}$, donde $k \in \{1, 2, \cdots, 2^m\}$.

(c) Pruebe la desigualdad de Hölder en general, por continuidad.

(d) Muestre que la desigualdad de Hölder es una igualdad sólo en el caso especificado en el Problema 13, parte d.

PROBLEMA 24 (Minkowski a través de Hölder; prueba de F. Riesz). Tratemos de probar la desigualdad de Minkowski

$$\left(\int_X |f(x) + g(x)|^r dx \right)^{1/r} \leq \left(\int_X |f(x)|^r dx \right)^{1/r} + \left(\int_X |g(x)|^r dx \right)^{1/r}, \tag{1.4.6}$$

donde $f, g : X \to \mathbb{R}$, a través de la desigualdad de Hölder (1.3.3).

(a) Para comenzar, reduzca (1.4.6) al caso de $f, g : X \to \mathbb{R}$ con valores no negativos solamente. (Entonces los valores absolutos ya no son necesarios.)

(b) Necesitamos expresar $\int_X (f(x) + g(x))^r dx$ como un producto escalar $\langle v, w \rangle$. Explore las posibilidades.

(c) Si $v = v_1 + v_2$, tenemos $\langle v, w \rangle = \langle v_1, w \rangle + \langle v_2, u \rangle$. Que se obtiene si se aplica la desigualdad de Hölder a ambos terminos (por separado) del lado derecho de esta igualdad?

En estas notas trabajamos siempre con X de medida finita. Gran parte de lo que estamos desarrollando vale también para X de medida infinita, el cual es un caso con numerosas aplicaciones; empero, también es un caso con muchas dificultades adicionales, debidas en parte al hecho que, para una norma dada, una función continua $f : X \to \mathbb{C}$ de un espacio de medida infinita a $\mathbb{C}$ no tiene integral $\int_X f(x)$ finita las más de las veces. (Hay complicaciones severas que aparecen incluso cuando el espacio X es de medida finita pero no compacto. Ver el problema 30.) El caso más sencillo con X infinito es el de $X = \mathbb{N}$ con $\mu(\{x\}) = 1$. Entonces una función $f : X \to \mathbb{C}$ de norma ℓ_1 es lo mismo que una serie absolutamente convergente, el espacio de funciones de $f : X \to \mathbb{C}$ de norma ℓ_2 es el espacio de Hilbert de dimensión enumerable, cuya teoriá dista mucho de ser trivial.

* * *

En la teoría de números, el siguiente caso especial de la desigualdad de Cauchy es muy utilizado. En su primera forma, también es un caso especial de la desigualdad de Jensen.

PROBLEMA 25. (a) Sean $a_1, \ldots, a_n \in \mathbb{R}$. Muestre que

$$\left(\sum_{i=1}^n a_i \right)^2 \leq n \sum_{i=1}^n a_i^2. \tag{1.4.7}$$

(b) Supongamos que a lo mas m de los a_i son no nulos. Muestre que, entonces,

$$\left(\sum_{i=1}^{n} a_i\right)^2 \leq m \sum_{i=1}^{n} a_i^2. \tag{1.4.8}$$

La situacion siguiente es muy común en la teoría de numeros analítica.

PROBLEMA 26. Sean $a_{i,j} \in \mathbb{C}$ para $1 \leq i \leq n$, $1 \leq j \leq m$, donde no todos los $a_{i,j}$ son reales positivos. Queremos acotar $|\sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}|$. Muestre que el cuadrado de $|\sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}|$ es a lo más

$$\sqrt{n \sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \left(\sum_{i=1}^{n} \overline{a_{i,j_1}} a_{i,j_2}\right)} \tag{1.4.9}$$

Típicamente, separariamos entonces la suma doble exterior de (1.4.9) en una parte "diagonal"

$$\sum_{j=1}^{m} \left(\sum_{i=1}^{n} \overline{a_{i,j_1}} a_{i,j_2}\right)^2,$$

que consiste de todos los términos con $j_1 = j_2$, y una parte "no diagonal"

$$\sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{m} \sum_{j_2=1}^{m} \overline{a_{i,j_1}} a_{i,j_2},$$

que consiste de todos los otros términos. Entonces acotaríamos la parte diagonal por fuerza bruta (gracias a su pequeño número de términos) y trataríamos de obtener cancelación en las sumas de la forma

$$\sum_{i=1}^{n} \overline{a_{i,j_1}} a_{i,j_2},$$

para la mayoría de los pares (j_1, j_2) con $j_1 \neq j_2$. ("Obtener cancelación" consiste en mostrar que los argumentos (ángulos) de los números complejos $\overline{a_{i,j_1}} a_{i,j_2}$ son lo suficientemente distintos como para que las contribuciones de los termines $\overline{a_{i,j_1}} a_{i,j_2}$ a la suma se eliminen en gran parte las unas a las otras. Claro está, $\overline{a_{i,j_1}} a_{i,j_2}$ podrían ser todos reales; en ese caso, el signo juega el rol del argumento ($\arg(r) = 0$ para $r > 0$, $\arg(r) = \pi$ para $r < 0$).

5. Operadores duales. Principio de la gran criba

Sean V, W espacios vectoriales sobre $\mathbb{R}$ o $\mathbb{C}$ dotados de productos escalares. Sea $A : V \to W$ un operador linear (es decir, una función linear) de V a W. Decimos que un operador linear $A^* : W \to V$ es el *operador dual* a A, si para $v \in V$, $w \in W$ cualesquiera,

$$\langle w, Av \rangle = \langle A^* w, v \rangle.$$

Si $V = \mathbb{R}^m$, $W = \mathbb{R}^n$, o $V = \mathbb{C}^m$, $W = \mathbb{C}^n$, y $\langle \cdot, \cdot \rangle$ es el producto escalar usual[6], todo operador linear $A : V \to W$ tiene un único dual. Esto se puede ver de la manera siguiente. El producto escalar se puede expresar como sigue:

$$\langle \vec{x}, \vec{y} \rangle = t_{\overline{\vec{x}}} \vec{y} = (x_1, x_2, \cdots, x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

para $\vec{x}, \vec{y} \in \mathbb{C}^n$ o $\vec{x}, \vec{y} \in \mathbb{R}^n$, y lo mismo con m en vez de m para $\vec{x}, \vec{y} \in \mathbb{C}^m$, o $\vec{x}, \vec{y} \in \mathbb{R}^m$. (El producto de $t_{\overline{\vec{x}}}$ y $\vec{y}$ es simplemente un producto matricial. Denotamos por t_v la transposición de

[6]Por cierto, cualquier producto escalar en $\mathbb{C}^n$ es equivalente al producto escalar usual en $\mathbb{C}^n$ bajo alguna transformación linear invertible de $\mathbb{C}^n$ a $\mathbb{C}^n$. Esto es equivalente al hecho que toda matriz Hermitiana ($M = t_{\overline{M}}$) es diagonalizable (en C).

un vector vertical en uno horizontal, y, más generalmente, denotamos por t_A la transposición de una matriz A.) El operador A se puede expresar como una matriz con m columnas y n filas:

$$A(\vec{v}) = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$$

para $v \in V$. Por lo tanto, para $v \in V$, $w \in W$,

$$\langle w, Av\rangle = t_{\overline{w}} \cdot Av = t_{\overline{w}} A \cdot v = t_{\overline{t_{\overline{A}} w}} v = \langle t_{\overline{A}} w, v\rangle,$$

por lo cual el dual A^* de A existe y es simplemente la transposición $t_{\overline{A}}$ del conjugado $\overline{A}$ de la expresión como matriz del operador A.

Problema 27. Muestre que, para dos espacios lineares cualesquiera V, W sobre $\mathbb{R}$ o $\mathbb{C}$, el dual A^* de un operador linear A de V a W es único, si es que existe.

Sean $V = \mathbb{R}^m$, $W = \mathbb{R}^n$ o $V = \mathbb{C}^m$, $W = \mathbb{C}^n$. Asignemos la norma ℓ_{r_1} a V y la norma ℓ_{r_2} a W, para $1 \le r_1, r_2 \le \infty$ arbitrarios. Un operador linear A de V a W tiene una norma $|A|_{r_1,r_2}$ dada por (1.3.2). El operador dual A^* de A va de W a V. Véase A^* como un operador linear de W con la norma ℓ_{s_2} a V con la norma ℓ_{s_1}, donde s_1 y s_2 son los exponentes duales a r_1 y r_2, respectivamente.

La norma de A^* como operador de un espacio de norma ℓ_{s_2} a un espacio de norma ℓ_{s_1} no es otra sino la norma de A:

$$|A^*|_{s_2,s_1} = |A|_{r_1,r_2} \tag{1.5.1}$$

Esta igualdad es la esencia de la *gran criba*. En su corazón yace una desigualdad.

Problema 28 (Prueba de (1.5.1)). (a) Por definición, $|A|_{r_1,r_2} = \sum_{v\in V\, v\neq 0} \frac{|Av|_{r_2}}{|v|_{r_1}}$. Demuestre que

$$\sup_{\substack{v\in V\\ v\neq 0}} \frac{|Av|_{r_2}}{|v|_{r_1}} = \sup_{\substack{w\in W\\ w\neq 0}} \frac{\langle w, Av\rangle}{|w|_{s_2}|v|_{r_1}}.$$

(b) Pruebe que, en general, se puede invertir el orden de los sup:

$$\sup_{x\in X}\sup_{y\in Y} f(x,y) = \sup_{y\in Y}\sup_{x\in X} f(x,y).$$

(c) Muestre que

$$\sup_{\substack{v\in V\\ v\neq 0}}\sup_{\substack{w\in W\\ w\neq 0}} \frac{\langle w, Av\rangle}{|w|_{s_2}|v|_{r_1}} = \sup_{\substack{v\in V\\ v\neq 0}}\sup_{\substack{w\in W\\ w\neq 0}} \frac{\langle A^*w, v\rangle}{|w|_{s_2}|v|_{r_1}} = \sup_{\substack{w\in W\\ w\neq 0}} \frac{|A^*w|_{s_1}}{|w|_{s_2}}.$$

(d) Concluya que $|A|_{r_1,r_2} = |A^*|_{s_2,s_1}$.

Las aplicaciones de (1.5.1) (ver §2) se hacen a menudo a partir de la forma concreta siguiente.

Problema 29. Sean dados $m, n \in \mathbb{Z}$, $\{a_{i,j}\}_{1\le i\le m, 1\le j\le n}$ con $a_{i,j} \in \mathbb{C}$, $r_1, r_2 \in [1,\infty]$. Supongamos que se nos pide probar que

$$\left(\sum_{j=1}^{n}\left|\sum_{i=1}^{m} a_{i,j}x_j\right|^{r_2}\right)^{1/r_2} \le C\left(\sum_{i=1}^{m}|x_i|^{r_1}\right)^{1/r_1}$$

para todo $\vec{x} \in \mathbb{R}^m$. Muestre que basta probar que

$$\left(\sum_{i=1}^{m}\left|\sum_{j=1}^{n} \overline{a_{i,j}}y_j\right|^{s_1}\right)^{1/s_1} \le C\left(\sum_{j=1}^{n}|y_j|^{s_2}\right)^{1/s_2}$$

para todo $\vec{y} \in \mathbb{R}^n$.

* * *

PROBLEMA 30. Un operador linear $A : V \to W$ de un espacio vectorial con norma a otro se llama *acotado* si la norma $|A|$ (en el sentido de (1.3.2) es finita. En general, todo operador acotado tiene un dual, auń cuando V o W tienen dimension infinita. Veamos un caso particular pero sumamente importante.

Sean tanto V como W iguales al espacio de funciones integrables $f : X \to \mathbb{C}$ con X compacto (y por ende de medida finita). Demos a V la norma ℓ_{r_1} y a W la norma ℓ_{r_2}. Supongamos que el operador A es

$$A(f) = \int_X \phi(x, y) f(x) dx,$$

donde $\phi : X \times X \to \mathbb{C}$ es una función integrable y acotada (llamada *núcleo del operador* A; de A mismo se dice que es un *operador integral*). Muestre que A tiene un operador dual.

El núcleo ϕ juega el rol de la matriz de A cuando V, W son de dimensión finita. Nótese que la prueba de (1.5.1) es válida para V y W de dimensión general.

6. Analísis de Fourier en $\mathbb{Z}/p$. Transformada de Fourier como isometría.

El cuerpo $\mathbb{Z}/p$ se puede definir simplemente[7] como el conjunto $\{0, 1, \ldots, p-1\}$ dotado de la adición y la multiplicación módulo p:

$$3 + 5 = 1, \quad 3 - 5 = 5, \quad 3 \cdot 5 = 1, \quad 3^{-1} = 5 \quad \text{en } \mathbb{Z}/p, \quad p = 7.$$

PROBLEMA 31. Pruebe que $\mathbb{Z}/p$ satisface todos los axiomas de un cuerpo. El paso menos sencillo es mostrar que todo $x \in \mathbb{Z}/p$ no nulo tiene un inverso $x^{-1} \in \mathbb{Z}/p$. Muestre que esto se deduce del siguiente enunciado: dados $a, b \in \mathbb{Z}$ primos entre sí, existen $m, n \in \mathbb{Z}$ tales que $am + bn = 1$. Pruebe esto a su vez modificando el algoritmo de Euclides, el cual encuentra el máximo común denominador de a y b (igual a 1, en nuestro caso).

El análisis de Fourier sobre $\mathbb{R}/\mathbb{Z}$, el cual no es sino el análisis de Fourier de las funciones de período 1 de $\mathbb{R}$ a $\mathbb{C}$, se basa en el hecho que las funciones $\psi_n : x \mapsto e^{2\pi i n x}$, $n \in \mathbb{Z}$,

(a) son ortogonales entre sí (respecto al producto escalar para funciones de $\mathbb{R}/\mathbb{Z}$ a $\mathbb{C}$);
(b) son de norma ℓ_2 igual a 1;
(c) son caracteres aditivos, i.e., $\psi_n(x + y) = \psi_n(x)\psi(y)$;
(d) generan un subespacio denso del espacio de funciones integrables de $\mathbb{R}/\mathbb{Z}$ a $\mathbb{R}$.

Análogamente, consideremos las funciones $\psi_y : x \mapsto e^{2\pi i x y/p}$ para $y \in \mathbb{Z}/p$.

PROBLEMA 32. Demuestre que las funciones ψ_y tienen período p, y por lo tanto pueden ser vistas como funciones de $\mathbb{Z}/p$ a $\mathbb{C}$. Dote a $\mathbb{Z}/p$ de la medida μ tal que $\mu(\{x\}) = 1/p$. Muestre que las funciones ψ_y son ortogonales entre sí, tienen norma ℓ_2 igual a 1, son caracteres aditivos, y generan el espacio de funciones integrables de $\mathbb{R}/\mathbb{Z}$.

PROBLEMA 33. Utilizando el problema 32, pruebe que toda función $f : \mathbb{Z}/p \to \mathbb{C}$ puede ser expresada como una *serie de Fourier*

$$f(x) = \sum_{y \in \mathbb{Z}/p} a_y \psi_y(x) = \sum_{y \in \mathbb{Z}/p} a_y e^{2\pi i x y/p},$$

dónde $a_y \in \mathbb{Z}/p$. Muestre que

$$a_y = \langle \psi_y, f \rangle = \frac{1}{p} \sum_{y \in \mathbb{Z}/p} e^{2\pi i x y}.$$

Los coeficientes a_y son denominados *coeficientes de Fourier*. Definimos la *transformada de Fourier* $\hat{f} : \mathbb{Z}/p \to \mathbb{C}$ de la manera siguiente:

$$\hat{f}(y) = a_y.$$

Veremos ahora que dos propiedades de la transformación de Fourier $f \to \hat{f}$ para $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ también funcionan para $f : \mathbb{Z}/p \to \mathbb{C}$.

[7]Los puristas prefieren definir $\mathbb{Z}/p$ como el conjunto de clases de equivalencia de $\mathbb{Z}$ módulo $p\mathbb{Z}$, dotado de las operaciones inducidas por la adición y la multiplicación en $\mathbb{Z}$.

PROBLEMA 34. Pruebe que $|\hat{f}|_2 = \frac{1}{p}|f|_2$. En otras palabras, la transformación $\hat{h}$ de Fourier es una isometría multiplicada por $\frac{1}{p}$. (Una *isometría* es una transformación que preserva la norma.)

PROBLEMA 35. La *convolución* $f * g$ de dos funciones $f, g : \mathbb{Z}/p \to \mathbb{C}$ es la función $x \mapsto \frac{1}{p}\sum_{y\in\mathbb{Z}/p} f(y)g(x-y)$. Muestre que $\widehat{f * g} = \hat{f} \cdot \hat{g}$.

CAPÍTULO 2

Aplicaciones en la teoría de números

1. La desigualdad de Cauchy y el análisis de Fourier en la combinatoria aditiva

La combinatoria aditiva se ocupa de las maneras de representar elementos de un grupo (generalmente abeliano) como suma de elementos dados.

PROBLEMA 36. Sea A un conjunto finito de enteros. Sea k un entero positivo fijo. Definimos $r(n)$ como el número de maneras de expresar n de la manera siguiente:

$$n = a_1 + \cdots + a_k - a_{k+1} - \cdots - a_{2k},$$

donde $a_1, a_2, \ldots, a_{2k}$ son elementos cualesquiera de A. Queremos probar que $r(n) \geq r(0)$ para todo n.

(a) Sea $t(m)$ el número de soluciones de

$$a_1 + \cdots + a_k = m$$

con $a_1, \ldots, a_k \in A$. Muestre que $r(n) = \sum_{m\in\mathbb{Z}} t(m)t(m+n)$.

(b) Pruebe que

$$\sum_{m\in\mathbb{Z}} t(m)t(m+n) \leq \sum_{m} (t(m))^2 = r(0).$$

* * *

Dado un subconjunto $A \subset \mathbb{Z}/p$, denotamos por A su número de elementos, y por $\chi_A : \mathbb{Z}/p \to \mathbb{C}$ su *función característica*

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Por lo tanto, $|A| = |\chi_A|_1 = |\chi_A|_2^2$.

Definimos

$$A + A = \{x + y : x, y \in A\}, \qquad \xi A = \{\xi x : x \in A\} \qquad \text{para } \xi \in \mathbb{Z}/p.$$

Es de esperarse que, para un A "tıpico" o "dentro de lo normal", el número de elementos de $A + A$ no sea menor que una constante por $|A|^2$. Por otra parte, si $A = \{1, 2, \ldots, k\}$, entonces $|A + A|$ es sólamente $2|A| - 1$. Lo mismo es cierto si A es una progresión aritmética cualquiera, módulo p.

Mostraremos que, si bien $|A + A|$ puede ser pequeño, $|A + \xi A|$ no lo será para algún $\xi \in S$, donde $S \subset \mathbb{Z}/p$ es un subconjunto cualquiera. (Para que el resultado no sea trivial o vacuo, es necesario que S tenga un cierto número mínimo de elementos.)

PROBLEMA 37 ([**10**], Lemma 2). (a) Pruebe que, para cualquier $\xi \in \mathbb{Z}/p$,

$$p \cdot |\chi_A * \chi_{\xi A}|_2^2 = |A|^2 + |\{(a_1, b_1, a_2, b_2) : a_1, b_1, a_2, b_2 \in A, a_1 \neq a_2, a_1 + b_1\xi = a_2 + b_2\xi\}|.$$

(b) Muestre que, en consecuencia,

$$\sum_{\xi\in S} p \cdot |\chi_A * \chi_{\xi A}|_2^2 \leq |S||A|^2 + |A|^4.$$

Que cota inferior se le puede dar entonces a $\max_{\xi\in S} p \cdot |\chi_A * \chi_{\xi A}|_2^2$?

(c) Acote $|A + \xi A|$ en términos de $|\chi_A * \chi_{\xi A}|_2^2$ y $|\chi_A * \chi_{\xi A}|_1$. Deduzca que existe un $\xi \in S$ tal que

$$|A + \xi A| \geq |A|^2|S|/(|A|^2 + |S|).$$

PROBLEMA 38 ([**7**], Lemma 2.5). Veremos que se puede obtener una cota para $|A + \xi A|$ complementaria a aquella obtenida en el problema 37: a veces es más fuerte que esta, y a veces más debil. Usaremos tanto el análisis de Fourier como la desigualdad de Cauchy.

(a) Muestre que

$$\sum_{\xi\in S} |\chi_A * \chi_{\xi A}|_2^2 = p \sum_{\xi\in S} |\widehat{\chi_A * \chi_{\xi A}}|_2^2 \leq |S||\hat{A}(0)|^4 + p^2(|\widehat{\chi_A}|_2^2)^2 = \frac{1}{p^4}|S||A|^4 + \frac{1}{p^2}|A|^2.$$

(b) Partiendo de (a), muestre que existe un $\xi \in S$ tal que

$$|\chi_A * \chi_{\xi A}|_2^2 \leq \frac{|A|^4}{p^4} + \frac{1}{p^2|S|}|A|^2.$$

(c) Demuestre que existe un $\xi \in S$ tal que

$$|A + A| \leq p \cdot \frac{|\chi_A * \chi_{\xi A}|_1^2}{|\chi_A * \chi_{\xi A}|_2^2} = \left(\frac{1}{p} + \frac{1}{|S||A|^2/p}\right)^{-1}.$$

(d) Concluya que, si $|S||A|^2 > p^2$, existe un $\xi \in S$ tal que $|A + \xi A|$ contiene más de la mitad de los elementos de $\mathbb{Z}/p$. Muestre también que, si $|S||A| > p^{1+\epsilon}$, entonces $|A + \xi A| > \frac{1}{2}\min(p, |A|^{1+\epsilon})$.

Qué cota inferior se le puede dar a $|A + \xi A|$ para por lo menos la mitad de los elementos de $\xi \in S$? Qué cota se puede dar para por lo menos 9 de cada 10 elementos de S? (Retorne a la parte (b).)

2. La gran criba: desigualdades

> **criba.** (De *cribo*). **1. f.** Cuero ordenadamente agujereado y fijo en un aro de madera, que sirve para cribar. También se fabrica de plancha metálica con agujeros, o con red de malla de alambre.
> **2. f.** Cada uno de los aparatos mecánicos que se emplean en agricultura para cribar semillas, o en minería para lavar y limpiar los minerales.
>
> [**13**], "Criba"

Una *criba* es un método que nos permite contar cuantos elementos de un conjunto dado no son eliminados por una sucesión de criterios. Generalmente, se trata de conjuntos de enteros, y de criterios de divisibilidad.

Podemos utilizar una criba para encontrar primos: comenzamos con el conjunto $\mathscr{A}$ de todos los enteros de 1 a n; para cada primo $p \leq n^{1/2}$, eliminamos todos los elementos de $\mathscr{A}$ divisibles por p; si llevamos cuenta de cuantos elementos son eliminados en cada paso, cuidandonos de no contar ningún elemento por partida doble, terminaremos con el número de primos de 1 a n.

PROBLEMA 39 (Criba de Erastóstenes–Legendre). Hay n enteros de 1 a n. De éstos, $\lfloor n/2 \rfloor$ son divisibles por 2, y $\lfloor n/3 \rfloor$ son divisibles por 3. El número de enteros de 1 a n que no son divisibles ni por dos ni por tres no es $n - \lfloor \frac{n}{2} \rfloor - \lfloor \frac{n}{3} \rfloor$, ya que los elementos divisibles por 6 son contados por partida doble. Compensando este hecho, obtenemos que el número de enteros de 1 a n es $n - \lfloor \frac{n}{2} \rfloor - \lfloor \frac{n}{3} \rfloor + \lfloor \frac{n}{6} \rfloor$.

(a) Muestre que, en general, el número de enteros de 1 a n sin factores primos con D es

$$\sum_{d|D} \mu(d) \left\lfloor \frac{n}{d} \right\rfloor = n \sum_{d|D} \mu(d) \frac{1}{d} + O(\tau(D)) = n \prod_{p|D} \left(1 - \frac{1}{p}\right) + O(\tau(D)), \tag{2.2.1}$$

donde $\tau(d)$ es el número de divisores (enteros positivos) de d, $O(x)$ denota una cantidad acotada en valor absoluto por Cx, donde C es una constante (en este caso 1) y $\mu(d)$ es

la *función de Möbius*

$$\mu(d) = \begin{cases} 1 & \text{si } d \text{ tiene un número par de divisores primos, ninguno repetido} \\ -1 & \text{si } d \text{ tiene un número impar de divisores primos, ninguno repetido} \\ 0 & \text{si } p^2|d \text{ para algún primo } p. \end{cases}$$

Puede usar (y demostrar) la *formula de inversión de Möbius*:

$$\sum_{d|D} \mu(d) = \begin{cases} 1 & \text{si } D = 1, \\ 0 & \text{si } D > 1. \end{cases} \tag{2.2.2}$$

(b) Denotemos por $\pi(x)$ el número de primos de 1 a x. Pruebe que

$$\pi(n) - \pi(n^{1/2}) \ll \frac{n}{\prod_{p \le \frac{1}{2}\log n}\left(1 + \frac{1}{p}\right)}. \tag{2.2.3}$$

Gracias al teorema de Mertens (2.3.7), podemos ver que la cota dada por (2.2.3) es simplemente

$$\pi(n) - \pi(n^{1/2}) \ll \frac{n}{\log\log n},$$

lo cual es bastante pobre comparado con la realidad (ver el apéndice 3). El método del problema 39 tiene la virtud de ser adaptable al estudio de muchos conjuntos aparte del de los números primos – por ejemplo, una modificación nos da la asintótica correcta para el número de enteros de 1 a n libres de cuadrados, i.e., no divisibles por ningún cuadrado d^2, $d > 1$. Aún así, podemos ver que se trata de un método mas bien débil.

La idea de las *pequeñas cribas* (criba pura de Brun, criba de Brun, criba de Selberg, criba de Rosser-Iwaniec) consiste en aproximar y truncar $\mu(d)$ en (2.2.1) de tal manera que el error así incurrido sea mucho menor que la mejora en el término de error (inicialmente $O(\tau(d))$) de (2.2.1).

La *gran criba* sigue un enfoque distinto, aunque, como veremos más tarde, con ciertos reencuentros con el de las pequeñas cribas (en particular, la criba de Selberg). La primera gran criba fue concebida y formulada por Linnik **[11]** para situaciones en las cuales hay un gran número de residuos por primo a ser excluidos. Rényi encontró un resultado mas fuerte que el de Linnik mediante un enfoque probabilístico. Desde ese entonces, la gran criba ha sido mejorada, pero por lo menos su enunciado ha conservado un fuerte resabio probabilístico, por lo menos en el sentido formal.

* * *

Sea $\mathscr{A} \subset \mathbb{Z}$ un conjunto finito. Definimos

$$Z = |\mathscr{A}|, \qquad Z(d) = |\{a \in \mathscr{A} : d|a\}|.$$

Lo razonable es suponer que $Z(p)$ sea aproximadamente Z/p para un prímo "típico" p. Esto no puede cierto para todo $\mathscr{A}$ y todo p; tomemos, por ejemplo, el caso de $\mathscr{A}$ igual al conjunto de todos los primos de 1 a n. Empero, el conjunto de primos es en cierto sentido penalizado por su excepcionalidad; queremos mostrar que un conjunto puede ser excepcional sólo en cuanto sea pequeño.

Problema 40. Estimaremos la norma ℓ_2 de la desviación de $Z(p)$ con respecto a Z/p.

(a) Sea $\mathscr{A} \subset \{1, 2, \dots, n\}$. Se nos da el problema de mostrar que

$$\sum_{p \le X} p\left(Z(p) - \frac{Z}{p}\right)^2 \le C_{X,n}|\mathscr{A}| \tag{2.2.4}$$

para alguna constante $C_{X,n}$ dependiente sólo de X y n. Reformule (2.2.4) como una aseveración sobre la norma[1] $|A|_{2,2}$ de un operador $A : \mathbb{R}^n \to \mathbb{R}^m$, donde m es el número de primos $\le X$. Note que $|\mathscr{A}| = |\vec{x}|_2^2$, donde $x_j = 1$ para $j \in \mathscr{A}$ y $x_j = 0$ para $j \notin \mathscr{A}$.

(b) Encuentre el dual A^* de A. Escriba su norma $|A^*|_{2,2}$.

[1]En toda esta sección, puede utilizar la norma ℓ_2 no normalizada $\sum_{j=1}^n |x_j|^2$, si esto le resulta más conveniente.

(c) Proceda como en (1.4.9) y lo que le sigue, inviertiendo el orden de la suma y separando los términos diagonales de los no diagonales. En el problema presente, la parte no diagonal dará el término de error, y la parte diagonal dará el término principal para valores típicos de X y n.

(d) Estime las contribuciones de la partes diagonal y no diagonal. Concluya que

$$\sum_{p\le X} p\left(Z(p)-\frac{Z}{p}\right)^2 \le (n+X^2)|\mathscr{A}|. \tag{2.2.5}$$

(e) Qué cota superior para el número de primos de 1 a n nos da (2.2.5)?

Como puede verse, el enunciado (2.2.5) es tán débil como la criba de Erastóstenes-Legendre cuando se trata de contar primos. Hay dos direcciones posibles en las que podemos proceder:

(a) podemos tratar de mejorar la cota (2.2.5);

(b) podemos tratar de generalizarla para que detecte no sólo las desviaciones en el número de elementos de $\mathscr{A}$ divisibles por p, sino las deviaciones en el número de elementos de $\mathscr{A}$ congruentes $\mod p$ a distintos elementos de $\mathbb{Z}/p$.

Comenzaremos por (b), y en el proceso encontraremos una manera de realizar (a). Primero veremos que la desviaciones del promedio en $\mathbb{Z}/p$ son visibles en la transformada de Fourier.

Problema 41. (a) Dada una función $f:\mathbb{Z}/p\to\mathbb{C}$ una función, sea $E_f=\frac{1}{p}\sum_{j=0}^{p-1} f(j)$. Muestre que

$$\sum_{x=0}^{p-1}|f(x)-E_f|^2 = p\sum_{x=1}^{p-1}|\hat{f}(x)|^2.$$

(b) Defina

$$Z(d,a)=|\{x\in\mathscr{A}: x\equiv a \bmod d\}|, \quad S(r)=\sum_{x\in\mathscr{A}} e^{2\pi i r x}$$

para a, d enteros, r real. Muestre que

$$p\sum_{a=0}^{p-1}\left(Z(p,a)-\frac{Z}{p}\right)^2 = \sum_{a=1}^{p-1}|S(a/p)|^2. \tag{2.2.6}$$

(c) Dado $\vec{x}\in\mathbb{C}^n$, podemos plantear en general

$$Z(d,a)=\sum_{\substack{1\le j\le n\\ j\equiv a \bmod d}} x_j, \quad S(r)=\sum_{1\le j\le n} x_j e^{2\pi i r j}.$$

Muestre que (2.2.6) aún vale.

Problema 42. Pruebe que, para n y α cualesquiera,

$$\sum_{j=1}^{n} e^{2\pi i j\alpha} = e(\pi i(n+1)\alpha)\frac{\sin \pi n\alpha}{\sin \pi\alpha}.$$

(Utilize la identidad $\sum_{j=1}^{n} g^j=\frac{g^{n+1}-g}{g-1}$.)

Problema 43. Nuestra meta ahora es acotar

$$\sum_{p\le X} p\sum_{a=0}^{p-1}\left(Z(p,a)-\frac{Z}{p}\right)^2.$$

Gracias a (2.2.6), nos bastará con acotar

$$\sum_{p\le X}\sum_{a=1}^{p-1}|S(a/p)|^2. \tag{2.2.7}$$

Exprese (2.2.7) de la forma $|A\vec{x}|_2^2$ para un operador linear $A:\mathbb{C}^n\to\mathbb{C}^m$, donde m es el número de pares (p,a) con $p\le X$ y $1\le a<p$. Proceda como antes: exprese $|A^*\vec{x}|_2^2$ para $\vec{x}\in\mathbb{C}^m$ como una

suma de términos diagonales y no diagonales. Simplifique las sumas $\sum_{j=1}^{n} e^{2\pi ixj}$ provenientes de los términos no diagonales mediante la identidad obtenida en el problema 42; no las acote todavía. (O acótelas, después de todo; probablemente obtendrá una cota de $\sqrt{n + O(X^2 \log X)}$ para lo norma, lo cual fue el mejor resultado en algun momento, pero ya ha sido superado.)

La contribución total de los términos diagonales es $n \sum_{j=1}^{m} |x_j|^2$. La suma de los términos no diagonales es

$$\sum_{\substack{p_1 \le X \\ (p_1,a_1) \neq (p_2,a_2)}} \sum_{p_2 \le X} \sum_{a_1=1}^{p_1-1} \sum_{a_2=1}^{p_2-1} \overline{x_{(p_1,a_1)} e^{-\pi i (n+1) a_1/p_1}} x_{(p_2,a_2)} e^{-\pi i (n+1) a_2/p_2} \frac{\sin \pi n \alpha}{\sin \pi \alpha}. \tag{2.2.8}$$

Problema 44 (Un lema técnico). Sean $\alpha_r \in \mathbb{R}$ dados tales que $|\alpha_{r_1} - \alpha_{r_2}| \ge \delta$ para r_1, r_2 distintos cualesquiera, donde $\delta > 0$ es un positivo arbitrario. Queremos obtener la cota

$$\left| \sum_{\substack{r_1 \\ r_1 \neq r_2}} \sum_{r_2} \frac{\overline{z_{r_1}} z_{r_2}}{\alpha_{r_1} - \alpha_{r_2}} \right| \le \frac{\pi}{\delta} \sum_r |z_r|^2 \tag{2.2.9}$$

para $z_1, z_2, \dots, z_n \in \mathbb{C}$ cualesquiera.

(a) Probar (2.2.9) es lo mismo que mostrar que, para todo $\vec{z}$, $|\langle \vec{z}, A\vec{z} \rangle| \le \frac{\pi}{\delta} |\vec{z}|_2^2$, donde $A_{i,j} = \frac{1}{\alpha_i - \alpha_j}$ si $i \neq j$, y $A_{i,i} = 0$. Dado que A es una matriz antisimétrica, que estamos tratando de probar acerca del mayor autovalor (en valor absoluto) λ_0 de A? (Por lo mismo que A es antisimétrica, todos sus autovalores son puramente imaginarios.)

(b) Cual es el mayor autovalor de $t_A A$, en términos de λ_0? Sea $\vec{v}_0$ un autovector de A correspondiente a λ_0. Cuál es el valor de $\overline{t_{\vec{v}_0}} t_A A \vec{v}_0$, en términos de λ_0?

(c) Desarrolle $\overline{t_{\vec{v}}} t_A A \vec{v}$, utilizando la identidad

$$\frac{1}{\alpha_i - \alpha_k} \frac{1}{\alpha_j - \alpha_k} = \frac{1}{\alpha_j - \alpha_i} \left(\frac{1}{\alpha_i - \alpha_k} - \frac{1}{\alpha_j - \alpha_k} \right).$$

Asumiendo $\vec{v}$ es un autovector de A (i.e., $\sum_{j \neq i} \frac{v_j}{\alpha_i - \alpha_j} = \lambda v_i$ para todo i y algún λ puramente imaginario), simplifique y obtenga tanta cancelación en $\overline{t_{\vec{v}}} t_A A \vec{v}$ como sea posible. Despues del uso de la más simple de las desigualdades que conocemos en parte de lo que queda, debe obtener la cota superior

$$3 \sum_i |v_i|^2 \sum_k \frac{1}{(\alpha_i - \alpha_k)^2}. \tag{2.2.10}$$

(d) Utilize el hecho que $|\alpha_i - \alpha_j| > \delta$, asi como la conocida igualdad $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ (Euler), para acotar (2.2.10). Obtenga, finalmente, (2.2.9).

Problema 45 (Corolarios de (2.2.9)). (a) Sean $\alpha_r \in \mathbb{R}$ dados tales que $|\alpha_{r_1} - \alpha_{r_2}| \ge \delta$ para r_1, r_2 distintos cualesquiera, donde $\delta > 0$ es un positivo arbitrario. Procediendo a partir de (2.2.9), queremos obtener la cota

$$\left| \sum_{\substack{r_1 \\ r_1 \neq r_2}} \sum_{r_2} \frac{\overline{z_{r_1}} z_{r_2}}{\sin \pi(\alpha_{r_1} - \alpha_{r_2})} \right| \le \frac{1}{\delta} \sum_r |z_r|^2 \tag{2.2.11}$$

para $z_1, z_2, \dots, z_n \in \mathbb{C}$ cualesquiera. Aplique (2.2.9) a los números con doble índice $\alpha_{m,r} = m + \alpha_r$, $z_{m,r} = (-1)^m z_r$, donde $1 \le m \le K$ y K es un parámetro a fijar más tarde. Reemplaze la condición $(m_1, r_1) \neq (m_2, r_2)$ por $r \neq s$ (porqué es que esto no cambia el resultado?), defina $k = m - n$ y reemplaze la suma con respecto a m_1 y m_2 por una suma con respecto a k, con pesos $K - |k|$. Divida por K, deje que $K \to \infty$ y utilize la igualdad $\lim_{K \to \infty} \sum_{k=-K}^{K} \frac{(-1)^k}{k+\alpha} = \frac{\pi}{\sin \pi \alpha}$ (la cual, como la suma de Euler antes utilizada, puede probarse por integración de contornos en el plano complejo).

(b) Ahora queremos obtener

$$\left|\sum_{\substack{r_1\\ r_1\neq r_2}}\sum_{r_2}\overline{z_{r_1}}z_{r_2}\frac{\sin 2\pi x(\alpha_{r_1}-\alpha_{r_2})}{\sin\pi(\alpha_{r_1}-\alpha_{r_2})}\right|\leq\frac{1}{\delta}\sum_r|z_r|^2 \tag{2.2.12}$$

para $x\in\mathbb{R}$ y $z_1,z_2,\dots,z_n\in\mathbb{C}$ cualesquiera (y los $\alpha_r\in\mathbb{R}$ satisfacen $|\alpha_{r_1}-\alpha_{r_2}|\geq\delta$. Basta con aplicar (2.2.11 dos veces por separado, con $\vec{z}$ ligeramente diferente del $\vec{z}$ dado en cada caso (de manera que los dos nuevos $\vec{z}$ incorporen x), y sumar los dos resultados.

Ahora podemos finalmente probar la formas modernas[2] (2.2.14) y (2.2.15) de la gran criba.

PROBLEMA 46. (a) Aplique (2.2.12) al término no-diagonal (2.2.8) para obtener la *criba de Rényi*[3]:

$$\sum_{p\leq X}p\sum_{a=0}^{p-1}\left(Z(p,a)-\frac{Z}{p}\right)^2=\sum_{p\leq X}\sum_{a=1}^{p-1}|S(a/p)|^2\leq(n+X^2)\sum_{j=1}^n|x_j|^2. \tag{2.2.13}$$

(b) A decir verdad, para obtener (2.2.13), utilizamos el dato que p es primo sólo mediante el hecho que, cuando p_1, p_2 son primos $\leq X$ y a_1, a_2 son enteros tales que $(p_1,a_1)\neq(p_2,a_2)$, $a_1\nmid p_1$, $a_2\nmid p_2$, entonces

$$\left|\frac{a_1}{p_1}-\frac{a_2}{p_2}\right|\geq\frac{1}{p_1p_2}\geq\frac{1}{X^2}.$$

Ahora bien, esto es cierto para enteros generales $q_1,q_2\leq X$ en vez de primos $p_1,p_2\leq X$, con tal que $\mathrm{mcd}(a_1,q_1)=1$, $\mathrm{mcd}(a_2,q_2)=1$. Deduzca que

$$\sum_{1\leq q\leq X}\sum_{\substack{a=0\\ \mathrm{mcd}(a,q)=1}}^{q-1}|S(a/q)|^2\leq(n+X^2)\sum_{j=1}^n|x_j|^2 \tag{2.2.14}$$

(Roth **[15]**, Bombieri **[3]**)

(c) Queremos expresar $\sum_{0\leq a<q:\mathrm{mcd}(a,q)=1}|S(a/q)|^2$ en términos de $Z(d,a)$, para así poder replantear (2.2.14) de una manera parecida a (2.2.5). Muestre primero que

$$q\sum_{d|q}\frac{\mu(d)}{d}Z\left(\frac{q}{d},h\right)=\sum_{\substack{1\leq a\leq q\\ \mathrm{mcd}(a,q)=1}}S(q,a)e^{-2\pi i\frac{ah}{q}}.$$

(Utilize la siguiente consecuencia inmediata (y ubicua) de la fórmula de inversión de Möbius (2.2.2): $\sum_{1\leq a\leq q:\mathrm{mcd}(a,q)=1}c_a=\sum_d\mu(d)\sum_{1\leq a\leq q:d|q}c_a$ para c_a cualesquiera.) Luego muestre que

$$\sum_{\substack{a=1\\ \mathrm{mcd}(a,q)=1}}^{q}|S(a/q)|^2=q\sum_{h=1}^q\left(\sum_{d|q}\frac{\mu(d)}{d}Z\left(\frac{q}{d},h\right)\right)^2.$$

Concluya que (2.2.14) tiene la forma equivalente

$$\sum_{1\leq q\leq X}q\sum_{a=1}^q\left(\sum_{d|q}\frac{\mu(d)}{d}Z\left(\frac{q}{d},a\right)\right)^2\leq(n+X^2)\sum_{j=1}^n|x_j|^2. \tag{2.2.15}$$

Las desigualdades (2.2.14) y (2.2.15) son esencialmente óptimas. Lo que queda es derivar sus consecuencias.

---

[2]"Aditivas", ya que los términos $j\mapsto e^{2\pi ija/q}$ implícitos en $S(a/q)$ son "caracteres aditivos", esto es, homomorfismos de $\mathbb{Z}$ (como grupo aditivo) al círculo de radio unidad (como grupo multiplicativo). Existen también formulaciones "multiplicativas" de la gran criba; ver, e.g., **[2]**, §4.

[3]Con un factor $(n+X^2)$ mucho más pequeño, y por lo tanto mucho mejor, que el original de Rényi – por lo menos para $X>n^{1/3}$.

3. La gran criba como tal

Comenzemos por definir un problema de criba de manera un tanto general. Sean dados

(a) un conjunto $\mathscr{S} \subset \{1, 2, \dots, n\}$,
(b) un conjunto de primos $\mathscr{P}$, y
(c) para cada $p \in \mathscr{P}$, un subconjunto $\Omega_p \subset \mathbb{Z}/p$ de cardinalidad $\omega_p = |\Omega_p|$,

tales que, para $p \in \mathscr{P}$, $a \in \Omega_p$ y $s \in \mathscr{S}$ cualesquiera, tenemos $s \not\equiv a \bmod p$. En otras palabras, las clases en Ω_p les son prohibidas a los elementos de $\mathscr{S}$. Nuestra tarea es entonces acotar superiormente el número de elementos $|\mathscr{S}|$ de $\mathscr{S}$.

Si ω_p esta acotado, decimos que tenemos una situación de *pequeña criba*; si, por el contrario, ω_p crece – y a menudo lo hace de manera proporcional a p – decimos que se requiere una *gran criba*.

Por razones históricas, a (2.2.14) y (2.2.15) se les llama *desigualdades de gran criba*; empero, mostraremos como deducir de ellas tanto la primera gran criba (Linnik, 1941) como una criba moderna válida para ω_p cualesquiera. En el caso de ω_p acotado, este criba general es equivalente a una versión un tanto restringida de una de las pequeñas cribas más conocidas: la criba cuadrática de Selberg.

Problema 47 (Gran criba de Linnik). Sea $\mathscr{P} = \{p \text{ primo} : p \leq \sqrt{n}\}$. Utilizando (2.2.13), pruebe que, para todo α, $0 < \alpha < 1$,

$$|\mathscr{S}| \ll \frac{n}{\alpha^2 \cdot |\{p \in \mathscr{P} : \omega_p > \alpha p\}|}, \tag{2.3.1}$$

donde la constante implícita es absoluta.

La primera aplicación dada a la gran criba (por Linnik mismo) fue una respuesta parcial a la siguiente interrogante. Decimos que a es un *residuo cuadrático* modulo p si existe un entero b tal que $a \equiv b^2 \bmod p$. Ahora bien, uno espera que los residuos y los no residuos cuadráticos estén entremezclados, en parte porque no hay razón por la cual no lo estén. He aquí la pregunta: que cota inferior podemos dar para el no residuo más pequeño? En otras palabras, podemos encontrar k dependiente de p tal que al menos un elemento de $\{1, 2, \dots, k\}$ deba ser un no residuo cuadrático?

Linnik logró probar que uno puede tomar k igual a una potencia arbitrariamente pequeña de p, salvo para un número muy limitado de excepciones. Antes de seguir sus pasos, necesitamos un lema por lo demás muy útil.

Problema 48 (Números lisos o "friables"). Probaremos lo siguiente: para todo $\epsilon > 0$, existe un $\delta > 0$ tal que, para todo N, el número de enteros $n \in \{1, 2, \dots, N\}$ todos cuyos divisores primos son $\leq N^\epsilon$ es por lo menos δN. En otras maneras, una proporción positiva de todos los enteros son bastante "friables". En verdad, probaremos algo mas preciso: sea $l(N, z)$ el número de enteros $n \in \{1, 2, \dots, N\}$ todos cuyos divisores primos son $\leq z$; entonces, para todo $u > 0$,

$$l(N, N^u) \sim \delta(u)N, \tag{2.3.2}$$

donde $\delta : (0, \infty) \to \mathbb{R}$ es una función diferenciable definida por iteración:

$$\begin{aligned} &\delta(u) = 1 \quad \text{si } u \geq 1, \\ &\delta(u) = 1 - \int_u^1 \delta\left(\frac{v}{1-v}\right)\frac{dv}{v} \quad \text{si } u < 1. \end{aligned} \tag{2.3.3}$$

Nótese que, para definir el valor $\delta(u)$ para $\frac{1}{k+1} \leq u < \frac{1}{k}$, se utilizan sólo los valores de $\delta(w)$ para $w \geq \frac{u}{1-u} \geq \frac{1}{k}$. Se sugiere una estrategia de inducción: probar (2.3.2) para $1/2 \leq u \leq 1$, y, luego, para $\frac{1}{k+1} \leq u < \frac{1}{k}$, asumiendo (2.3.2) para $u \geq \frac{1}{k}$.

(a) Cuántos factores primos $p > N^{1/2}$ puede tener un número $n \leq N$? Pruebe (2.3.2) para $1/2 \leq u \leq 1$, excluyendo, para cada $p > N^{1/2}$, los $1 \leq n \leq N$ divisibles por p.
(b) El problema de evitar excluir un elemento $n \in \{1, \dots, N\}$ más de una vez torna el enfoque directo de (a) improcedente[4]. Sigamos una de las estrategias más comunes. Pruebe el

[4]Esta es, en general, la razón por la cual cribar no es trivial.

lemma de Buchstab:

$$l(N,p') = l(N,p) - \sum_{k=1}^{\infty} l\left(\frac{N}{p^k}, p'\right),$$

donde p' y p son dos números primos consecutivos cualesquiera ($p' < p$). Deduzca que, para y, z cualesquiera con $y < z$,

$$l(N,y) = l(N,z) - \sum_{y<p\le z} l\left(\frac{N}{p}, p'\right) - \sum_{k=2}^{\infty} l\left(\frac{N}{p^k}, p'\right), \tag{2.3.4}$$

donde p' denota el primo inmediatamente precedente a p.

(c) Sea $u_0 \in \left[\frac{1}{k+1}, \frac{1}{k}\right)$. Defina $y = N^{u_0}$, $z = N^k$. Acote el último término del lado derecho de (2.3.4) por $o(N)$ y estime el segundo término asumiendo (2.3.2) para $u \ge \frac{1}{k}$. Concluya que (2.3.3) vale para u_0.

Problema 49 (Teorema de Linnik). Sea $\epsilon > 0$. Para N arbitrario, sea $\mathscr{P}$ el conjunto de primos $p \le N$ tales que $1, 2, 3, \dots, \lfloor N^\epsilon \rfloor$ son todos residuos cuadráticos $\mod p$. Probaremos que $|\mathscr{P}| < C_\epsilon$, donde C_ϵ es una constante que depende sólo de ϵ.

(a) Sea $N = n^2$, $\mathscr{P}$ como dicho, y $\mathscr{S} = \{1 \le s \le N : q \text{ primo}, q|s \Rightarrow q < N^\epsilon\}$. Dados $s \in \mathscr{S}$ y $p \in \mathscr{P}$, que se puede decir acerca de los factores primos de s, considerados módulo p? Qué se puede decir acerca de s en sí, módulo p? En que clases Ω_p es que s no puede estar?

(b) Cuánto es ω_p? Aplique la gran criba de Linnik (2.3.1) para concluir.

(c) Pruebe el siguiente corolario: para todo $\epsilon > 0$, el número de primos[5] $p \le N$ tales que $1, 2, 3, \dots, \lfloor p^\epsilon \rfloor$ son todos residuos cuadráticos $\mod p$ esta acotado por $\ll_\epsilon \log\log N$.

* * *

Derivaremos ahora de (2.2.14) la criba general prometida.

Problema 50. Sea dado un problema de criba por $\mathscr{S} \subset [1,n]$, $\mathscr{P}$, y $\{\Omega_p\}_{p\in\mathscr{P}}$. Sea $\mathscr{Q}$ el conjunto de todos los $q \le X$ iguales a productos de primos distintos en $\mathscr{P}$. Probaremos que

$$|\mathscr{S}| \le \frac{N + X^2}{\sum_{q\in\mathscr{Q}} \prod_{p|q} \frac{\omega_p}{p-\omega_p}} \qquad \text{(Montgomery [12])} \tag{2.3.5}$$

(a) Defina $J(q) = \prod_{p|q} \frac{\omega_p}{p-\omega_p}$. Verifique que basta probar

$$\sum_{\substack{1\le a\le q\\ \gcd(a,q)=1}} |S(a/q)|^2 \ge |S(0)|^2 J(q), \tag{2.3.6}$$

donde $S(r) = \sum_{1\le j\le n} e^{2\pi i j r} x_j$ (como de costumbre), $q \in \mathscr{Q}$ y $\vec{x} \in \mathbb{C}^n$ es tal que $x_j = 0$ para $j \notin \mathscr{S}$.

(b) Muestre que, si sabemos (2.3.6) para q primo y *todo* $\vec{x} \in \mathbb{C}^n$ tal que $x_j = 0$ para $j \notin \mathscr{S}$, podemos probar (2.3.6) para q no primo y cualquier $\vec{x} \in \mathbb{C}^n$ dado con $x_j = 0$ para $j \notin \mathscr{S}$.

(c) Sea $Z(p,a) = \sum_{1\le j\le n : j\equiv a \bmod p} x_j$. Demuestre que

$$\sum_{a=1}^{p-1} |S(a/p)|^2 = p \sum_{a=1}^{p} |Z(p,a)|^2 - |S(0)|^2.$$

Utilizando el hecho que $Z(p,a) = 0$ para $a \notin \Omega_p$, pruebe también que

$$|S(0)|^2 \le (p - \omega_p) \sum_{a=1}^{p} |S(p,a)|^2,$$

donde se aplica una de las desigualdades que conocemos. Concluya que (2.3.6) vale para $q = p$ primo.

[5]Es una vieja conjetura de Vinogradov que el número de tales primos está en verdad acotado superiormente por una constante. Esta conjetura esta aún abierta. Si se asume la hipótesis generalizada de Riemann, se puede probar la aseveración de la conjetura, aún con p^ϵ reemplazado por $2(\log p)^2$ (Ankeny, [1]).

* * *

Veamos ahora algunas aplicaciones de (2.3.5) como pequeña criba, es decir, para ω_p pequeño.

Problema 51 (Brun-Titchmarsh).

Problema 52 (Densidad de primos gemelos).

Apéndice: lemas sobre los primos.

Sea $\pi(n)$ el número de primos de 1 a n. El teorema de los números primos (Hadamard – de la Vallée-Poussin) afirma que $\pi(n) \sim \frac{n}{\log n}$. El primer resultado parcial hacia el teorema de los números primos (conjeturado por Gauss y Legendre) fue el teorema de Chebyshev (1848):

$$\log 2 \cdot (1 + o(1))\frac{n}{\log n} < \pi(n) < 2 \log 2 \cdot (1 + o(1))\frac{n}{\log n}.$$

Se puede probar por integración

$$\sum_{n=1}^{x} \frac{1}{n} = \log x + \gamma + o(1),$$

donde γ es una constante absoluta (llamada *constante de Euler*). El teorema de Chebyshev basta para probar el teorema de Mertens:

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log x}. \tag{2.3.7}$$

Para las pruebas, ver, por ejemplo, [**9**], cáp. 2.

Bibliografía

[1] Ankeny, N. C., The least quadratic non-residue, *Ann. of Math.* **55** (1952), no. 1, 65–72.
[2] Bombieri, E., Le grand crible dans la théorie analytique des nombres, *Astérisque* **18** (1987).
[3] Bombieri, E., On the large sieve, *Mathematika* **12** (1965), 201–225.
[4] Charles, D. X., *Sieve methods*, manuscrito, `http://www.cs.wisc.edu/~cdx/Sieve.pdf`
[5] Davenport, H., *Multiplicative number theory*, Markham, Chicago, 1967.
[6] Hardy, G. H., Littlewood, J. E., y G. Pólya, *Inequalities*, 2da ed., Cambridge University Press, 1952.
[7] Helfgott, H. A., Growth and generation in $SL_2(\mathbb{Z}/p)$, prepublicación.
[8] Iwaniec, H., *Sieve methods*, manuscrito.
[9] Iwaniec, H., y E. Kowalski, *Analytic number theory*, AMS, Providence, RI, 2004.
[10] Konyagin, S. V., A sum-product estimate in fields of prime order, prepublicación.
[11] Linnik, Yu. V., La gran criba, *Dokl. Akad. Nauk SSSR* **30** (1941), 292–294 (en ruso).
[12] Montgomery, H. L., A note on the large sieve, *J. London Math. Soc.* **43** (1968), 93–98.
[13] Real Academia Española, *Diccionario de la lengua española*, 22da ed., Espasa-Calpe, Madrid, 2001.
[14] Rényi, A., On the large sieve of Ju. V. Linnik, *Compositio Math.* **8** (1950), 68–75.
[15] Roth, K., On the large sieves of Linnik and Renyi, *Mathematika* **12** (1965), 1–9.