UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Métodos entrópicos en Combinatoria Aditiva y la Conjetura de Marton.

Mateo Mauri

Director/a: Román Sasyk

22 de Agosto de 2025

Indice

Prólogo

El objetivo de esta tesis es presentar una exposición de la noción de **grupos aproximados**, uno de los temas centrales de la **Combinatoria Aditiva**. El motor de la discusión será exponer los recientes resultados de Gowers, Green, Manners y Tao, quienes probaron un ansiado teorema que caracteriza la estructura de estos grupos aproximados, hasta ahora conocido como la **Conjetura de Marton** o **Conjetura polinomial de Freiman-Ruzsa** (acortado, por sus siglas en inglés, como **PFR**) para grupos abelianos de torsión finita.

En 2009 Terence Tao publica en arXiv un artículo titulado "Sumset and inverse sumset theory for Shannon entropy" [31], que, en sus propias palabras (ver [30]), es la evolución de una "escena borrada" de su libro (en conjunto con Vu) "Additive combinatorics" [35]. En aquel paper se desarrollan definiciones y resultados típicos de combinatoria aditiva, pero reformulados a través de la entropía de Shannon. Casi 15 años más tarde, en 2023, Tao publica en conjunto con Ben Green y Freddie Manners un artículo llamado "Sumset and entropy revisited" [16] en el cual hacen uso intensivo de esta reformulación entrópica para demostrar varias cosas, pero principalmente para dar un paso fundamental en la Conjetura de Marton, ya que allí logran demostrar una versión inicial de esta conjetura (conocida como el caso del 99%, cosa que explicaremos en la Introducción), haciendo uso de distintas herramientas técnicas que les provee el contexto entrópico, pero fundamentalmente la buena relación que tiene la entropía con los morfismos de grupos. Este artículo fue claramente un paso acertado (y casi premonitorio de la demostración definitiva de la conjetura), ya que casi 6 meses luego de su publicación fue acompañado por otro llamado "On a conjecture of Marton" [7], donde Tim Gowers, Ben Green, Freddie Manners y Terence Tao dieron definitivamente una solución por la afirmativa a la Conjetura de Marton en $\mathbf{F}_2$, a través de un elegante argumento inductivo (cuyo desarrollo haremos en el Capítulo 4) que parte del caso que habían demostrado en el artículo anterior. A lo largo de esta tesis desarrollaremos las ideas necesarias para entender el camino que estos matemáticos diagramaron para demostrar la Conjetura de Marton, por lo que el autor de la misma quiere dejar en claro que la intención no es dar resultados originales ni dejar de celebrar el mérito y autoría de las distintos personas que trabajaron para la solución de la misma. Por esta última razón es que se intentará dejar en claro en cada sección las fuentes principales que fueron utilizadas para la escritura de la tesis, en muchas ocasiones haciendo fundamentalmente una traducción de los artículos originales,

y en otras haciendo leves modificaciones a las demostraciones, siempre buscando facilitar el trabajo del lector.

Sumado a esto, en el Capítulo 5 incluímos la demostración de la Conjetura de Marton en grupos de torsión arbitraria, que fue dada también por Gowers, Green, Manners y Tao en "Marton's Conjecture in abelian groups with bounded torsion" [7], aunque había sido anunciada en el artículo en el cual resolvieron el caso en $\mathbf{F}_2$ de la misma conjetura. También daremos, en el último capítulo, la principal aplicación, hasta la actualidad, de la Conjetura de Marton, que es a un problema inverso de las **normas de Gowers**, cosa que esencialmente ya había sido demostrada por Green y Tao, en "An inverse theorem for the Gowers $U^3(G)$ norm" [14].

Para lograr su objetivo, la presente tesis está dividida en distintos capítulos.

El primer capítulo es introductorio, en él se busca enunciar y contextualizar brevemente los principales resultados a probar, incluyendo un recorrido por las ideas que se entrelazan en la demostración de la Conjetura de Marton. Es aquí donde aparece por primera vez la noción de **entropía**, concepto ya estándar en la Teoría de la Información, que apareció en los últimos años en la Combinatoria Aditiva, y que es la principal novedad en el área que permitió demostrar la conjetura. La Introducción finaliza con una breve sección para fundamentar la denominación de grupos aproximados, y entender qué se busca saber de estos.

En el segundo capítulo se desarrollarán las ideas de entropía necesarias, yendo desde su definición hasta las distintas reformulaciones de los conceptos de distancia y tamaño de conjuntos que se desarrollaron para entender su relación con la Combinatoria Aditiva. En este capítulo incluimos también una serie de desigualdades típicas de Teoría de la Información y muchas aplicaciones de estas a Combinatoria Aditiva, en general intentando recuperar varios resultados clásicos en un contexto entrópico. En este sentido, se observará en repetidas ocasiones que lo que se intenta hacer es trasladar la mayor parte de la teoría clásica a un lenguaje entrópico, y ver si es que en este contexto se pueden conseguir mejores cotas. El capítulo contiene también una versión de la conjetura en lo que se conoce como el **mundo del 99%** y una versión entrópica del lema de **Balog-Szeméredi-Gowers**.

El tercer capítulo se encarga de recorrer algunas versiones equivalentes de la Conjetura de Marton, por un lado para ofrecer al lector distintas formas de pensar el resultado, todas estándar en Combinatoria Aditiva, pero fundamentalmente para reformular esta conjetura de una manera entrópica, ya que está es la versión que demostraremos en la tesis.

Los Capítulos 4 y 5 tienen el contenido principal de la demostración de la Conjetura de Marton, donde se hace uso de la mayoría de las ideas desarrolladas en el Capítulo 2. Cada uno de los capítulos presenta una versión distinta de la conjetura: el primero se refiere al caso de grupos de torsión par, mientras que el segundo se encarga de grupos de torsión arbitraria. Esta costumbre, usual en el área, se debe más que nada a que las ideas esenciales se pueden ver en el caso de 2 torsión, que es más sencillo, y no preocuparse por complicaciones técnicas que surgen al considerar torsíon impar.

La versión de la Conjetura de Marton en $\mathbf{Z}$ es un problema abierto, y se discute

brevemente en la Introducción. El sexto capítulo está dedicado a una aplicación de la conjetura resuelta en el caso de torsión par a los enteros, obteniéndose allí una versión débil de la Conjetura de Marton, cosa que también fue probada por Gowers, Green, Manners y Tao en "Sumsets and entropy revisited" [16].

Los últimos dos capítulos se dedican a relacionar la Conjetura de Marton con las **normas de Gowers**, siendo el primer capítulo una introducción general a estas, motivadas a través de la demostración usando análisis de Fourier del teorema de Szeméredi ofrecida por Gowers, y el segundo una explicación de como podemos usar el resultado central de esta tesis para deducir un teorema inverso de estas normas.

Capítulo 1

Introducción

1.1 Resultados principales

¿Es posible encontrar progresiones aritméticas de longitud arbitraria en los números primos? Es decir, si fijamos una cantidad $k \in \mathbf{N}$, ¿existen un intervalo t y una lista $\{n, n+t, n+2t, ..., n+kt\}$ de números primos?

La respuesta por la afirmativa a esta pregunta es uno de los resultados más célebres de la Teoría Aditiva de Números de los últimos años, dada por Ben Green y Terence Tao. Uno de los artículos fundamentales para que este resultado haya existido es en cual Gowers dió una nueva demostración del Teorema de Szemerédi [6], donde se da origen al estudio de las **normas de Gowers**, y así también al análisis de Fourier en grados altos, temas que se volvieron centrales en la **Combinatoria Aditiva**.

La Combinatoria Aditiva es el área interesada por estudiar distintas estructuras aproximadas que aparecen en la matemática al trabajar en conjuntos con una operación, estudiando nociones como grupos aproximados, anillos aproximados, cuerpos aproximados, polinomios aproximados, etc. Es este el espacio dentro de la matemática en el cual vamos a estar trabajando a lo largo de esta tesis, que, como se verá, es un lugar sumamente interesante en el cual distintas ideas y técnicas se ponen en práctica, pasando, por ejemplo, por Combinatoria Elemental, Análisis Armónico, Geometría Convexa, Teoría de Grafos, Probabilidad, Teoría Ergódica, Geometría de Incidencia, etc.

Como se podrá suponer, es un área con demasiadas aristas, por lo que en este texto se dejarán varias puertas sin abrir, pero buscaremos centrarnos en ciertas ideas relacionadas con **grupos aproximados**, demostrando resultados recientes de Gowers, Green, Manners y Tao (sin lugar a duda, nombres centrales en este campo), y conectando esto con las ya antes mencionadas normas de Gowers, cuyo interés, al menos para el autor de esta tesis, viene dado por las aplicaciones de estas a la teoría de números. El estudio de las normas de Gowers se relaciona con el desarrollo de una teoría de **polinomios aproximados**, como se verá más adelante.

Una primera lectura de los problemas clásicos de la Combinatoria Aditiva puede ser dada en los números enteros ($\mathbf{Z}$), pero al tratarse de propiedades que suelen estar definidas usando las características aditivas de los conjuntos, pueden reformularse en cualquier grupo. Siguiendo este espíritu, es habitual resolver distintos problemas primero en el contexto de los cuerpos finitos o, más en general, de los espacios vectoriales sobre cuerpos finitos, lugares en los cuales podemos, además de usar herramientas de teoría de grupos de torsión finita, valernos de todo el arsenal del álgebra lineal. Es por esta razón que a lo largo de esta tesis nos concentraremos en este modelo, sin perder de vista las generalizaciones de los distintos resultados hacia los enteros. Para más información sobre las interacciones entre los enteros y los modelos en espacios vectoriales sobre cuerpos finitos dentro del área, recomendamos fuertemente los surveys "Finite field models in additive combinatorics"[11] y "Finite field models in arithmetic combinatorics – twenty years on"[23] (el primero es, a esta altura, un clásico moderno en este área, mientras que el segundo es una suerte de actualización del primero). Con esto no queremos decir que el estudio de la combinatoria aditiva se restringe solamente a los enteros y a los cuerpos finitos. Por ejemplo, hay varios resultados en conjuntos matriciales, como $SL_2(\mathbf{F}_p)$ o $SL_2(\mathbf{C})$ (ver [2], [17]), pero en esta tesis nos centraremos en los espacios vectoriales de característica finita, o más en general, sobre grupos de torsión finita.

Entonces, como ya mencionamos, una de las ideas principales a trabajar en esta tesis es la de grupos aproximados. En particular, nos interesan resultados con cierto sabor "inverso" [1], es decir, responder a la pregunta: Si tengo un conjunto que se comporta como un grupo en el sentido aditivo, ¿Cuanta estructura algebraica tiene este conjunto?

Claramente, la formulación es totalmente vaga y pueden encontrarse distintas formas de responder a esta inquietud. Para comenzar, la forma que nos interesa en esta tesis de medir si un conjunto es "casi" un grupo viene asociada a la idea de la **constante de duplicación** σ, definida como

$$\sigma(A) := \frac{|A+A|}{|A|},$$

donde A es un subconjunto cualquiera de un grupo, y entendemos que

$$A+B := \{x+y : x \in A, y \in B\}$$

cuando B es otro subconjunto del grupo. Con este lenguaje, diremos que un conjunto es "casi" un grupo si tenemos que esta constante σ es pequeña. Es sencillo ver las cotas triviales

$$1 \leq \sigma(A) \leq |A|,$$

y se puede probar que la cota inferior se realiza si y solo si el conjunto A es la coclase de algún subgrupo, lo que le da sentido a pensar que un conjunto con

[1] Se dice inverso ya que la pregunta "directa" sería: Si tengo un grupo, ¿Qué propiedades tiene este?

constante de duplicación pequeña debería tener estructura similar a la de un grupo.

Un resultado en esta dirección es el obtenido por Green y Ruzsa (ver [10]):

Teorema 1.1 ([10, Teorema 2.1])**.** *Sea $A \subseteq \mathbf{F}_2^\infty$ un conjunto finito tal que $|A+A| \leq K|A|$. Luego A está contenido en alguna coclase de un subgrupo $H \leq \mathbf{F}_2^\infty$ con $|H| \leq K^2 2^{2K^2-2}|A|$.*

En cierto sentido débil, este teorema es una descripción completa de los conjuntos con una constante de duplicación pequeña. Este resultado afirma que si tenemos un conjunto con constante de duplicación menor a K, lo podemos contener en un subespacio afín de tamaño menor o igual a $K^2 2^{2K^2-2}|A|$, y de manera inversa, si tenemos un conjunto contenido en un subespacio de estas características, es claro que $|A+A| \leq K^2 2^{2K^2-2}|A|$. El problema con este resultado es que sería provechoso poder "ir y volver" de la propiedad aditiva a la propiedad algebraica sin tener esta pérdida exponencial en K. Por ejemplo, uno podría apuntar a conseguir un teorema del cual se pueda obtener algo de la forma

$$\begin{aligned} \text{constante de duplicación } K &\implies \text{estructura algebraica} \\ &\implies \text{constante de duplicación } K' \end{aligned}$$

donde K' se relaciona de forma *polinomial* con respecto a K.

Sin embargo, es claro que un resultado que nos garantice esto debería ser realmente distinto al enunciado más arriba, y podemos ver esto tomando un conjunto $A \subseteq \mathbf{F}_2^\infty$ compuesto por un subespacio H junto con K puntos $\{x_1, ..., x_k\}$ linealmente independientes tales que $\langle x_1, ..., x_k\rangle \cap H = \{0\}$. Reflexionando un poco, se llega a que $|A+A| \leq K|A|$, pero el menor subespacio afín que contiene a A tiene tamaño, más o menos, $2^K|A|$, lo que nos obliga a considerar esta pérdida exponencial.

El principal teorema a demostrar en esta tesis es la solución a una conjetura de Katalin Marton (ver "An analog of Freiman's theorem in groups" [27], donde Ruzsa le atribuye esta conjetura a Marton) conocida en la literatura como la **Conjetura polinomial de Freiman-Ruzsa** (o **Conjetura de Marton**, pero extrañamente, esta última forma de decirle no es muy tradicional en la literatura), que viene a ser la reformulación del teorema de arriba, pero donde se busca cubrir al conjunto por varias coclases, en lugar de solo una, y recupera la posibilidad de mantener el crecimiento polinomial:

Teorema 1.2 ([7, Teorema 1.3])**.** *Sea G un grupo abeliano con torsión m. Supongamos que $A \subseteq G$ es un conjunto finito no vacio tal que $|A+A| \leq K|A|$. Luego A puede cubrirse por como mucho $(2K)^{O(m^3)}$ coclases de algún subgrupo $H \leq G$ de tamaño menor o igual $|A|$. Más aún, H está contenido en $\ell A - \ell A$ para algún $\ell \ll (2 + m\log K)^{O(m^3 \log m)}$.*

Este teorema fue demostrado recientemente por Gowers, Green, Manners y Tao, en una serie de cuatro papers: "On a conjecture of Marton" [8], "Marton's

Conjecture in abelian groups with bounded torsion" [7], "Sumset and inverse sumset theory for Shannon entropy" [31] y "Sumsets and entropy revisited" [16]. Para mantener la notación sencilla y las ideas claras, demostraremos primero el caso de $\mathbf{F}_2$, por lo que enunciaremos los resultados en este contexto, pero a lo largo del texto se podrán apreciar las distintas generalizaciones a $\mathbf{F}_p$ con p impar, y en realidad también a cualquier grupo finito de m torsión. Siguiendo esta línea, el enunciado del teorema central en el paper más importante de Green, Gowers, Manners y Tao en estos temas, "On a conjecture of Marton" [8], es:

Teorema 1.3 ([8, Teorema 1.2])**.** *Supongamos que A es un subconjunto finito no vacío de $\mathbf{F}_2^n$ tal que $|A+A| \leq K|A|$. Luego A puede cubrirse por como mucho $2K^{12}$ coclases de algún subgrupo $H \leq \mathbf{F}_2^n$ de tamaño menor o igual $|A|$.*

Notar que si A esta cubierto por r coclases de un subespacio H de tamaño menor o igual a $|A|$, se tiene que $|A+A| \leq (\binom{r}{2}+1)\,|A|$, y entonces el teorema de arriba nos dice que podemos pasar de la propiedad de estar cubierto por coclases de un subgrupo a tener una constante de duplicación pequeña y de vuelta, solo con una pérdida polinomial. En este sentido, es un fuerte puente entre propiedades aditivas y algebraicas de un conjunto.

Exponer este teorema fue el motor principal de la misma tesis, y comprender bien su demostración el objetivo principal del tesista.

Un recurso muy útil para entender este resultado es la siguiente afirmación atribuida a Ruzsa, por Green, en el survey "Notes on the polynomial Freiman–Ruzsa conjecture" [10], centrada en distintas versiones equivalentes del teorema:

Proposición 1.4 ([10, Proposición 2.2])**.** *Las siguientes afirmaciones son equivalentes.*

1. *Si $A \subseteq \mathbf{F}_2^\infty$ cumple que $|A+A| \leq K\,|A|$, entonces existe $A' \subseteq A$, $|A'| \geq |A|\,/C_1(K)$, contenido en una coclase de algun subespacio de tamaño como mucho $C_2(K)\,|A|$*

2. *Si $A \subseteq \mathbf{F}_2^\infty$ cumple que $|A+A| \leq K\,|A|$, entonces A puede cubrirse por como mucho $C_3(K)$ coclases de algun subespacio de tamaño como mucho $C_4(K)\,|A|$.*

3. *Si $A \subseteq \mathbf{F}_2^\infty$ cumple que $|A+A| \leq K\,|A|$, y adicionalmente existe un conjunto B, $|B| \leq K$, tal que $A+B = A+A$, entonces A puede cubrirse por, como mucho, $C_5(K)$ coclases de algun subespacio de tamaño como mucho $C_6(K)\,|A|$.*

4. *Supongamos que $f : \mathbf{F}_2^m \to \mathbf{F}_2^\infty$ es una función con la propiedad de que*

$$|\{f(x)+f(y)-f(x+y) : x,y \in \mathbf{F}_2^m\}| \leq K.$$

   Luego f puede escribirse como $g+h$, donde g es lineal, y $|Im(h)| \leq C_7(K)$.

5. *Supongamos que $f : \mathbf{F}_2^m \to \mathbf{F}_2^\infty$ es una función con la propiedad de que para al menos $2^{3m}/K$ de las tuplas $(x_1,x_2,x_3,x_4) \in \mathbf{F}_2^m$ con $x_1+x_2 = x_3+x_4$*

vale que $f(x_1) + f(x_2) = f(x_3) + f(x_4)$*. Luego, existe una función lineal afín* $g : \mathbf{F}_2^m \to \mathbf{F}_2^\infty$ *tal que* $f(x) = g(x)$ *para al menos* $2^m/C_8(K)$ *valores de* x*.*

Mas aún, si $C_i(K)$ *esta acotada por un polinomio en* K *para todo* $i \in I$*, donde* I *es alguno de los conjuntos* $\{1,2\},\{3,4\},\{5,6\},\{7\},\{8\}$*, en realidad* $C_i(K)$ *es acotada por un polinomio de* K *para todo* i*.*

Para ver todo un poco más en contexto, hagamos un recorrido por algunos de los distintos resultados previos hasta la resolución de la Conjetura de Marton. El primero en conseguir un teorema de este estilo fue Imre Ruzsa [27], que además de ser quien atribuyó esta conjetura a Marton, logró probar que el resultado era válido con una cota superior de $2K^2 2^{K^4}$, en lugar del $2K^C$ deseado. Luego de eso, la mejor cota fue la obtenida por Sanders [29], que consiguió controles del estilo de $\ll \exp(\log^{C_1} K)$ (en particular, lo demostró para $C_1 = 4+\varepsilon$), utilizando técnicas de la transformada de Fourier, muy usadas en el área.

Es importante destacar también que muy poco tiempo después de la aparición del preprint [8] en arXiv, Jyun-Jie Liao refinó los argumentos del trabajo hasta lograr bajar la constante a $2K^9$ (ver [18]).

La Conjetura de Marton es, en si misma, un muy buen resultado de caracterización en la teoría de grupos aproximados, pero también tiene varias aplicaciones interesantes. La primera de estas responde a una de las preguntas más naturales que uno puede hacerse cuando sabe que vale este resultado: ¿Qué pasa con los números enteros?

Teorema 1.5 ([8, Teorema 1.3])**.** *Sea* A *un subconjunto finito de* $\mathbf{Z}^D$ *para algún* $d \in \mathbf{N}$*, y supongamos que* $|A+A| \leq K|A|$ *para algún* K*. Entonces existe un subconjunto* $A' \subseteq A$*,* $|A'| \geq K^{-C_1/2}|A|$*, con* $dim(A') \leq C_2 \log(K)$*, para ciertas constantes absolutas* $C_1, C_2 \geq 0$*.*

Esto es conocido como la conjetura polinomial de Freiman-Ruzsa "débil" sobre $\mathbf{Z}$, y la deducción de este resultado desde la conjetura en $\mathbf{F}_2$ fue probada con anterioridad en "Sumsets and entropy revisited" [16, Teorema 1.11]. La noción de dimensión que usamos aquí es la **dimensión afín**, que define la dimensión de A como la dimensión del espacio vectorial real generado por $A - A$. Existe también una conjetura "fuerte" sobre $\mathbf{Z}$, pero esta se mantiene abierta (no es trivial siquiera enunciar el resultado que se desea demostrar, esto se trabaja en [21], [20]). Hasta ahora la mejor formulación que tenemos es:

Conjetura 1.6 (Conjetura polinomial fuerte de Freiman-Ruzsa en $\mathbf{Z}^D$, [21])**.** *Si* $G = \mathbf{R}^m$ *o* $\mathbf{Z}^m$ *y* $A \subseteq G$ *es un conjunto finito con la propiedad de que* $|A+A| \leq K|A|$*. Existe entonces una progresión convexa*

$$P = \{a_0 + n_1 a_1 + ... + n_d a_d : (n_1, ..., n_d) \in \mathbf{Z}^d \cap B\},$$

donde $B \subseteq \mathbf{R}^d$ *es un conjunto convexo centralmente simétrico (es decir, el conjunto es convexo y cumple que* $B = -B$*) y* $a_0, ..., a_d$ *son números dados, con rango* $d = O(\log(2K))$ *y tamaño* $\left|B \cap \mathbf{Z}^d\right| = O(K^{O(1)})|A|$*; y un conjunto* $X \subseteq G, |X| = O(K^{O(1)})$*; tales que* $A \subseteq P + X$*.*

Otra aplicación de la validez de la Conjetura de Marton particularmente agradable es la siguiente:

Corolario 1.7 ([8, Corolario 1.5])**.** *Existe una constante C con la siguiente propiedad. Supongamos que tenemos una función $f : \mathbf{F}_2^m \to \mathbf{F}_2^n$ tal que, si escogemos una pareja x, y de valores de forma aleatoria (con una distribución uniforme) en $\mathbf{F}_2^m$, $\mathbf{P}(f(x+y) = f(x) + f(y)) \geq 1/K$. Entonces existe un morfismo $g : \mathbf{F}_2^m \to \mathbf{F}_2^n$ tal que $\mathbf{P}(f(x) = g(x)) \gg K^{-C}$*

Otra consecuencia de la Conjetura de Marton, que generaliza un resultado de Bourgain y Chang en los números enteros (ver [1]), fue demostrado por Mudgal (ver [22]):

Corolario 1.8. *Sea $A \subset \mathbf{R}$ un conjunto finito con al menos dos elementos. Entonces debe ocurrir que alguno de los conjuntos $mA, A^{(m)}$ tiene cardinal al menos $|A|^{f(m)}$, donde $f(m) \to \infty$ cuando $m \to \infty$.*

La última aplicación que mencionaremos en este contexto es el siguiente teorema inverso de las normas de Gowers, que serán trabajadas en los últimos dos capítulos del texto:

Corolario 1.9 (Corolario 1.6, [8])**.** *Existe una constante C con la siguiente propiedad. Sea $f : \mathbf{F}_2^n \to \mathbf{C}$ una función acotada por 1, tal que $\|f\|_{U^3(\mathbf{F}_2^n)} \geq 1/K$. Luego existe un polinomio cuadrático $P : \mathbf{F}_2^n \to \mathbf{F}_2$ tal que*

$$|\frac{1}{2^n} \sum_{x \in \mathbf{F}_2^n} f(x)(-1)^{P(X)}| \gg K^{-C}.$$

Como referencia, en [19] se muestran distintas consecuencias de la PFR a la Teoría de la Computación.

Lo último que aclararemos en esta introducción es sobre la línea final del Teorema 1.2. Una conjetura muy conocida también en combinatoria aditiva es la conjetura polinomial de Bogolyubov:

Conjetura 1.10. *(conjetura polinomial de Bogolyubov) Sea G un grupo de torsión m, y A un subconjunto finito tal que $|A + A| \leq K|A|$. Existe entonces un subgrupo $H \subseteq 2A - 2A$ tal que $|H| \leq K^{O_m(1)}|A|$.*

Esto sigue siendo un problema abierto, incluso en el caso $m = 2$, pero las técnicas utilizadas en "Marton's Conjecture in abelian groups with bounded torsion" [7] permiten obtener el resultado más débil de esta conjetura, es decir, en lugar de meter un subespacio grande en $2A - 2A$, lograron hacerlo en $\ell A - \ell A$, con ℓ teniendo dependencia *polilogarítmica* en K. Hasta la fecha, Sanders (ver [29, Teorema 11.1]) había demostrado que $2A - 2A$ contiene un subespacio H' de tamaño al menos $exp(-C_m \log^4(2K))|A|$. Se sabía que la Conjetura de Bogolyubov implica la Conjetura de Marton, por lo que uno estaría tentado a intentar resolver 1.10.

1.2 Métodos entrópicos y técnicas de demostración

El principal ingrediente que traen a la mesa Gowers, Green, Manners y Tao es el de **entropía de Shannon**, idea que claramente los antecede, pero que no había sido utilizada todavía de forma fuerte en Combinatoria Aditiva. Las principales fuentes de referencia para estudiar la aplicación de estas técnicas en Combinatoria Aditiva son [31] y [16]. Procedamos entonces en esta sección a hacer un sumario de las ideas fundamentales para la demostración de la conjetura, buscando poner en contexto al lector (la idea en esto no es ser completamente riguroso, sino dar una estructura general de lo que haremos en los capítulos a venir). La presente sección combina ideas de dos entradas del blog de Tao, ver [33] y [34]. A lo largo de la discusión, si bien enunciaremos cirtos resultados para grupos abelianos en general, vamos a asumir que estamos trabajando en $\mathbf{F}_2^n$.

Veamos entonces como podríamos demostrar el Teorema 1.3, introduciendo en el camino las ideas de entropía necesarias. La idea principal es hacer inducción sobre la constante de duplicación $\sigma(A) = |A + A| / |A|$.

En particular, supongamos que cualquier conjunto A con constante de duplicación K sea "conmensurable" (parecido, en un sentido combinatorio aditivo) con respecto a otro conjunto A' pero con menor constante de duplicación, por ejemplo $K^{0.99}$. Una forma de la cual podriamos entender esta idea de conmensurabilidad es la distancia de Ruzsa $d(A; B)$ de los conjuntos A y B, que definiremos como

$$d(A; B) := \log \frac{|A - B|}{|A|^{1/2} |B|^{1/2}}, \tag{1.1}$$

cosa que uno querría controlar por $O(\log K)$. La terminología de distancia se justifica gracias al siguiente resultado:

Lema 1.11 (Desigualdad triangular de la distancia de Ruzsa)**.** *La distancia de Ruzsa $d(A, B)$ es no negativa, simétrica, y obedece la desigualdad triangular*

$$d(A, C) \leq d(A, B) + d(B, C)$$

para cualesquiera tres subconjuntos A, B, C en un grupo ambiente G.

(Para una demostración, ver [35, Lema 2.6])

La distancia de Ruzsa satisface todos los axiomas de una métrica, excepto por el hecho de que $d(A, B) = 0$ no implica que $A = B$ (tampoco vale que $d(A, A) = 0$ para todos los subconjuntos A). El siguiente resultado, cuya demostración puede buscarse en [35, Proposición 2.7], da algunas ideas de como se comporta esta distancia:

Proposición 1.12. *Supongamos que $A \subseteq G$ es un subconjunto de un grupo aditivo G. Entonces las siguientes afirmaciones son equivalentes:*

1. *$\sigma(A) = 1$ (i.e. $|A + A| = |A|$);*
2. *$\delta(A) = 1$ (i.e. $|A - A| = |A|$, or $d(A, A) = 0$) [2];*

[2] Esta notación es estándar, tanto como antes definimos la constante de duplicación, se puede hacer una definición análoga para controlar el tamaño del conjunto resta, $\delta(A) := \frac{|A-A|}{|A|}$

3. $d(A, B) = 0$ *para al menos un subconjunto* B*;*

4. $|nA - mA| = |A|$ *para al menos un par de números enteros no negativos* n, m *con* $n + m \geq 2$*;*

5. $|nA - mA| = |A|$ *para todos los pares de números enteros no negativos* n, m*;*

6. A *es una coclase de un subgrupo finito* H *de* G*.*

Con todo esto, aunque no sea de verdad una métrica, es muy útil seguir la heurística de pensar a la distancia de Ruzsa como una verdadera distancia, que mide cuan lejos están dos conjuntos, pero en un sentido aditivo.

Volviendo al argumento, si uno puede fabricarse un A' a una distancia chica de A pero con menor constante de duplicación, podriamos repetir este procedimiento hasta lograr que la constante de duplicación decaiga tanto como uno quiera. Luego, tendríamos el resultado, ya que se puede demostrar de forma más sencilla que, si la constante de duplicación de A es menor a $3/2$, entonces $A - A$ es un subespacio de $\mathbf{F}_2^n$ (esto lo probaremos más adelante, en la Sección 2.3, cuando trabajemos el caso del 99%, pero también puede verse que el resultado de arriba ya menciona algo similar a esto, en el caso de que la distancia sea efectivamente 0).

En efecto, podemos ahora usar la desigualdad triangular de la distancia de Ruzsa para concluir lo deseado: como las constantes de duplicación decaen como una sucesión geométrica de parámetro 0.99, aplicando reiteradas veces esta desigualdad nos fabricamos un conjunto A'' conmensurable con respecto a A, de forma que $A'' - A''$ es un subespacio. Habiendo llegado hasta esta situación, es posible utilizar lemas de cubrimientos (como el que daremos en el Capítulo 3, Lema 3.1) para concluir que A se puede cubrir con pocas coclases de un subespacio.

Con esta idea en la cabeza, sería entonces lógico preocuparnos por encontrar un método para fabricar, dado un conjunto A con pequeña constante de duplicación, un A' con las propiedades descriptas arriba. Para pensar en esto es útil jugar con algunos conjuntos y ver que hacemos en cada caso.

El primer tipo de conjunto A que uno puede plantear para relacionar con un cubrimiento por coclases es tomar algún subgrupo cualquiera H y considerar a A como un conjunto aleatorio $1/K$ denso allí. Es claro que podemos cubrir a A con el subgrupo H, y este subgrupo cumplirá las relaciones de tamaño que buscamos mantener con respecto a A: sería conmensurable con respecto a este y tendría constante de duplicación 1, por lo que en este caso tomaríamos $A' = H$. Para obtener H a partir de A de forma "intrínseca" (es decir, sin saber que construimos a A a partir de H) podríamos considerar el conjunto $A + A$, que, como construímos a A de manera aleatoria, debería llenar todo el conjunto H[3].

[3]La idea fundamental para comprender estos dos ejemplos es que, dado un conjunto A "sin estructura aditiva" debería ocurrir que $|A + A| \approx \binom{|A|}{2} \approx |A|^2$, por lo que bastaría pedir una noción de densidad relativamente débil para creerse que, si el conjunto es aleatorio, $A + A$ llene todo el subespacio.

El problema con esta estrategia, la de tomar $A' = A + A$ como "mejor" conjunto, es que podemos fácilmente fabricarnos un ejemplo en el cual esto no funcione: si consideramos como A a un conjunto formado por K coclases de algún subgrupo H, es fácil ver que $A + A$ será esencialmente K^2 coclases del mismo subgrupo H, por lo que su constante de duplicación claramente empeorará. En este caso, la manera de fabricarnos un "mejor" conjunto a partir de A es quedarse solo con una coclase, cosa que podemos reescribir como tomar el conjunto $A' = A \cap (A + h)$, para algún $h \in A + A$ típico.

Luego hay, al menos, dos formas bastante razonables para intentar modificar los conjuntos, que llamaremos "jugadas":

1. Reemplazar A con $A + A$.

2. Reemplazar A con $A \cap (A + h)$ con un "típico" $h \in A + A$.

Lamentablemente, existen conjuntos para los cuales ninguna de estas dos jugadas sirven (esencialmente se puede verificar que si uno fabrica a A tomando conjuntos $1/\sqrt{K}$ densos en $\sqrt{K}$ coclases de algún subgrupo H, entonces la constante de duplicación se mantiene más o menos en K aplicando cualquiera de las dos estrategias), pero quizás sí una combinación de ambas jugadas (en el caso del ejemplo, hacer una de las jugadas y después la otra, en cualquier orden, sirve).

Esto sugiere una posible estrategia: demostrar que alguna de las dos jugadas mejoran la constante de duplicación y, si esto no ocurre, hacer alguna operación más complicada para conseguir la disminución deseada.

Una idea interesante que también le da un poco más de fuerza a esta estrategia es que si tomamos un conjunto A con constante de duplicación K, entonces $A \times A$ tiene constante de duplicación K^2, y proyectando este producto a una recta mediante la proyección $\pi : \mathbf{F}_2^n \times \mathbf{F}_2^n \to \mathbf{F}_2^n$ que cumple $\pi(x, y) = x + y$, obtenemos que $\pi(A) = A + A$. Pero $\pi^{-1}(\{h\})$ es esencialmente $A \cap (A + h)$, de lo que nos gustaría decir que

$$\sigma(A \times A) \text{ se comporta como } \sigma(A + A).\sigma(A \cap (A + h_0)), \tag{1.2}$$

donde h_0 es algún elemento con la propiedad de que su fibra se comporta, en algún sentido, como el promedio de las demás fibras. Esto nos garantizaría que al hacer las dos jugadas que propusimos antes para mejorar nuestro conjunto siempre tendríamos un caso donde, en el peor caso, la constante de duplicación se mantiene estable, y en el mejor, disminuye. El problema de este argumento es que no se puede escribir de forma rigurosa con la constante de duplicación clásica.

Por suerte, en el contexto de entropía todo funciona mejor. Ya definiremos todas las nociones entrópicas equivalentes necesarias, pero lo importante de este nuevo contexto es que podemos dar un resultado que, de forma cuantitativa, nos diga cuan expresable es la entropía de $A \times A$ con respecto a las de $A + A$ y $A \cap (A + h)$, dándonos incluso una expresión del error, de lo que nos tomaremos para trabajar el caso donde ninguna de las dos jugadas sea muy efectiva. Llamaremos a este resultado la **identidad del fibrado**.

Continuemos entonces definiendo la entropía de Shannon. Sea A un conjunto finito. Trabajaremos con el conjunto de todas las medidas de probabilidad en A con soporte compacto (en este caso finito), es decir, funciones $p : A \to [0, 1]$ distintas de 0 solo para un número finito de valores y cuya suma total es 1. Definimos una **variable aleatoria** en A como una variable aleatoria X que toma valores en un subconjunto finito $\text{rango}(X) := \{x \in A : \mathbf{P}(x \in X) \neq 0\}$, de modo que la función de distribución $p_X(x) := \mathbf{P}(x \in X)$ de X es una medida de probabilidad. Escribimos $X \equiv Y$ si $p_X = p_Y$, es decir, si X e Y tienen la misma distribución. Nos referimos a las variables aleatorias que toman valores en un conjunto finito como **variables aleatorias discretas**.

La **entropía de Shannon** $\mathbf{H}(p)$ de una distribución de probabilidad p está dada por la fórmula

$$\mathbf{H}(p) := \sum_{x \in A} F(p(x))$$

donde $F : \mathbf{R}^+ \to \mathbf{R}^+$ es la función

$$F(x) := x \log \frac{1}{x}$$

bajo la convención de que $F(0) = 0$. Dada una variable aleatoria A-valuada X, definimos entonces $\mathbf{H}(X) := \mathbf{H}(p_X)$.

Un ejemplo central de variable aleatoria que utilizaremos varias veces es la de una **variable uniforme** en un conjunto A dado, cosa que denotaremos por $\mathbf{U_A}$. En este caso particular vale que

$$\mathbf{H}(U_A) = \log(|A|),$$

lo que nos da una relación muy directa entre el cardinal del conjunto A, y la entropía de la variable aleatoria U_A. Es esta idea la que justifica una heurística que utilizaremos a lo largo de todo el texto, y es que **la entropía de una variable aleatoria será nuestro análago al tamaño de un conjunto**. Es claro entonces que las variables aleatorias discretas son generalizaciones de un conjunto, donde ahora nos permitimos ponderar los elementos del conjunto como nos sea útil, y que la entropía generaliza nuestra noción de tamaño.

Un pequeño ejemplo que ya muestra el beneficio de pensar en variables aleatorias en lugar de conjuntos es la diferencia entre el conjunto $A + B$ y la variable aleatoria dada por sumar una uniforme en A y una uniforme en B, es decir $U_A + U_B$, donde usamos la típica notación de variable uniforme en un conjunto. Mientras que $A + B$ es simplemente el conjunto de sumas, la variable aleatoria $U_A + U_B$ tiene soporte en $A + B$, pero también codífica la información de la cantidad de maneras de sumar cada elemento de $A + B$.

Vamos a entonces a reiniciar esta digresión, ahora reemplazando las nociones clásicas de combinatoria aditiva con estas ideas de entropía:

El concepto clave con el cual comenzaremos es el de la **distancia entrópica de Ruzsa** $\mathbf{d[X; Y]}$ [4] entre dos variables aleatorias X, Y que toman valores en

[4] La principal diferencia notacional que utilizaremos entre la distancia de Ruzsa y la distancia

$\mathbf{F}_2^n$ (a partir de ahora trabajaremos en este grupo, aunque ciertas definiciones que daremos podrían entenderse en contextos más amplios), definida por

$$d[X;Y] := \mathbf{H}(X' - Y') - \frac{1}{2}\mathbf{H}(X') - \frac{1}{2}\mathbf{H}(Y'), \tag{1.3}$$

donde X', Y' son copias independientes de X, Y.

Una vez más: La idea a tener en mente es que nuestro conjunto A puede interpretarse como una variable aleatoria con distribución uniforme en A, que denotaremos U_A, y lo que antes entendiamos como el cardinal de A va a ser reemplazado por la entropía de esta variable, $\mathbf{H}(U_A)$. Siguiendo con esta idea, es más claro como la distancia de Ruzsa entrópica es una versión análoga de la clásica, definida en (1.1). Se puede demostrar también en este contexto la desigualdad triangular

$$d[X;Y] \leq d[X;Z] + d[Z;Y]$$

para cualesquiera tres variables aleatorias X, Y, Z, lo que justifica el nombre de distancia (ver el Lema 2.4 de esta tesis), pero tampoco será una verdadera métrica, ya que tendrá problemas similares a la distancia de Ruzsa clásica.

Podemos definir también una constante de duplicación entrópica de una variable aleatoria X como

$$\sigma[X] := \mathbf{H}(X + X') - \mathbf{H}(X),$$

donde X' es una copia independiente de X. Una cosa a notar importante entre las versiones combinatorias y entrópicas de la constante de duplicación es que la constante de duplicación de una variable aleatoria puede ser significativamente más chica que la del conjunto imagen de esa variable. Por ejemplo, sea A el intervalo $[0, N)$ junto con otros $\sqrt{N}$ enteros en posición "aleatoria", donde N lo pensamos como un número grande. Entonces puede verse que la constante de duplicación de A es más o menos $\sqrt{N}$, pero la distribución uniforme en A tiene constante de duplicación $O(1)$. Vemos entonces que un poco de "ruido" (como los $\sqrt{N}$ enteros en posición aleatoria) puede afectar mucho la constante de duplicación de un conjunto, pero puede tener impacto casi nulo en la constante de duplicación de la variable uniforme en ese conjunto. Esto nos dice que podemos heurísticamente entender a la teoría entrópica de sumas de conjuntos como una versión "tolerante al ruido" de la combinatoria clásica.

En este lenguaje, puede reescribirse la conjetura polinomial de Freiman-Ruzsa como:

Teorema 1.13 (Conjetura de Marton entrópica)**.** *Sea X una variable aleatoria con valores en $\mathbf{F}_2^n$ tal que $d[X;X] \leq \log K$. Existe entonces una variable aleatoria uniforme U_H sobre un subgrupo $H \leq \mathbf{F}_2^n$ tal que $d[X;U_H] \leq C \log K$ para alguna constante absoluta C.*

entrópica de Ruzsa será hablar de una con paréntesis y de la otra con corchetes. Esta diferencia notacional no será solo para la distancia, sino que en general hablaremos de las nociones combinatorio aditivas con paréntesis y de las entrópicas con corchetes.

Notar aquí que la hipótesis sobre la constante de duplicación fue reemplazada por "$d[X;X] \leq \log K$", ya que, en $\mathbf{F}_2^n$, vale que

$$d[X;X] = \mathbf{H}(X - X') - \mathbf{H}(X) = \mathbf{H}(X + X') - \mathbf{H}(X) = \sigma[X].$$

Esto último es nuestra noción análoga de constante de duplicación, y lo que estamos usando fuertemente es que en un grupo de 2-torsión sumar es lo mismo que restar (sin embargo, este uso de la torsión no es para nada fundamental, puede ser arreglado tranquilamente para otros primos distintos a 2, la necesidad de trabajar en $\mathbf{F}_2$ para simplificar las cuentas se verá más adelante). Por último, con todo lo desarrollado previamente, puede entenderse la desigualdad

$$d[X;U_H] \leq C \log K$$

como que X se puede cubrir por una cantidad de coclases controlada por K del subgrupo H (esto lo formalizaremos con ciertos lemas de cubrimientos en el Capítulo 3, pero ahora no es importante).

El siguiente paso es formalizar esta idea de achicar la constante de duplicación con algo conmensurable, establecido en la siguiente proposición:

Proposición 1.14 (Decrecimiento de la distancia)**.** *Si X, Y son dos variables aleatorias $\mathbf{F}_2^n$-valuadas, entonces podemos encontrar dos variables aleatorias $\mathbf{F}_2^n$ valuadas X', Y' tales que*

$$d[X';Y'] \leq (1-\eta)d[X;Y]$$

y

$$d[X;X'], d[Y;Y'] \leq Cd[X;Y]$$

para algunas constantes absolutas $C, \eta \geq 0$.

Notar, nuevamente, que en esta proposición la primera desigualdad debería tomarse como una disminución de la constante de duplicación, mientras que la segunda como una noción de conmensurabilidad entre las variables en cuestión, pero por lo antes notado todo puede sintetizarse en la noción de distancia entrópica de Ruzsa.

Observemos que sucede al asumir como verdadera a esta proposición. Arrancando con X, Y ambas igual a X e iterando, uno puede encontrar sucesiones de variables aleatorias X_n, Y_n, con $X_0 = Y_0 = X$, de forma tal que

$$d[X_n;Y_n] \leq (1-\eta)^n d[X;X],$$

y

$$d[X_{n+1};X_n], d[Y_{n+1};Y_n] \leq C(1-\eta)^n d[X;X].$$

En particular, por la desigualdad triangular y el comportamiento de la serie geométrica,

$$d[X_n;X], d[Y_n;X] \leq \frac{C}{\eta} d[X;X].$$

Con un argumento de compacidad puede demostrarse que alguna subsucesión de (X_n, Y_n) debe converger a una tupla compuesta por dos variables aleatorias X_∞, Y_∞, que, por propiedades de continuidad de la distancia de Ruzsa, cumplen que

$$d[X_\infty; Y_\infty] = 0$$

y

$$d[X_\infty; X], d[Y_\infty; X] \leq \frac{C}{\eta} d[X; X].$$

Si obtuviésemos esto, por todo lo discutido anteriormente en el caso clásico, obtendríamos la Conjetura de Marton, ya que estamos considerando que la variable aleatoria X es análoga al conjunto A, y X' a A', por lo que nos fabricamos una "variable aleatoria X_∞ conmensurable con respecto a X y con constante de duplicación nula". Para hacer esto riguroso, también tendremos que demostrar algún resultado que nos indique que, cuando la constante de duplicación es suficientemente pequeña, podemos demostrar una versión de la Conjetura de Marton en el contexto entrópico, y de eso se encargará la Sección del Capítulo 2 dedicado al caso del 99%. Concentrémonos entonces en esquematizar la prueba de la Proposición 1.14, que reformularemos como:

Proposición 1.15 (Formulación contrarecíproca)**.** *Si X e Y son variables aleatorias $\mathbf{F}_2^n$-valuadas, con la propiedad de que*

$$d[X'; Y'] > d[X; Y] - \eta(d[X; Y] + d[X'; X] + d[Y'; Y]) \tag{1.4}$$

para todas las variables aleatorias $\mathbf{F}_2^n$-valuadas X' e Y' y una constante absoluta lo suficientemente pequeña $\eta > 0$, entonces se obtiene un absurdo.

Efectivamente, asumiendo por verdadero el contrarrecíproco de la proposición de arriba, es decir, que existen X', Y' tales que

$$d[X'; Y'] \leq d[X; Y] - \eta(d[X; Y] + d[X'; X] + d[Y'; Y]),$$

lo que implica la proposición 1.14 con $C = 1/\eta$.

Ahora todo el juego se reduce a usar desigualdades de la entropía de Shannon y técnicas del **cálculo entrópico de Ruzsa** para deducir una contradicción como la buscada en la Proposición 1.15, para algún η suficientemente pequeño, intentando entrelazar las ideas intuitivas que planteamos para fabricarnos jugadas que nos ayuden a acercarnos a ese absurdo.

Vamos a hacer eso, pero primero hagamos unos cambios a (1.4) que facilitarán las ideas. Lo primero es optar por la versión condicional de esta desigualdad, es decir,

$$d[X'|Z; Y'|W] \leq d[X; Y] - \eta(d[X; Y] + d[X'|Z; X] + d[Y'|W; Y])$$

para cualesquiera dos variables aleatorias Z, W, posiblemente acopladas a X', Y' respectivamente. No hace falta entender a esta altura que significa condicionar a estas variables dentro del símbolo de distancia, pero se puede pensar como que se especifica cierta información de los conjuntos.

En particular, vamos a decir que una variable aleatoria X' es **relevante** (condicionada con respecto a otra variable aleatoria Z) si

$$d[X'|Z;X] = O(d[X;Y]), \tag{1.5}$$

o equivalentemente (por la desigualdad triangular)

$$d[X'|Z;Y] = O(d[X;Y]),$$

por lo que tenemos la cota inferior (asumiendo la hipótesis de la Proposición 1.15)

$$d[X'|Z;Y'|W] \geq (1 - O(\eta))d[X;Y] \tag{1.6}$$

siempre que X' e Y'sean relevantes condicionadas a Z, W respectivamente. Esta noción de relevancia es análoga a la idea de armar un conjunto conmensurable con respecto al original.

Esta desigualdad es muy útil, ya que hay que tener en mente que las reglas del cálculo entrópico de Ruzsa nos van a decir que, a groso modo, cualquier variable aleatoria que construyamos a partir (y esto es muy importante) de sumar copias de X, Y y condicionando con respecto a otras sumas, serán relevantes. De manera informal, se puede entender a esta desigualdad como que el espacio de variables aleatorias relevantes esta $(1 - O(\eta))d[X;Y]$-separado con respecto a la distancia entrópica de Ruzsa, y lo que haremos para llegar al absurdo será fabricar dos variables relevantes que estén cerca en esta distancia.

Enunciamos ahora la identidad del fibrado antes mencionada:

Proposición 1.16 (Identidad del fibrado)**.** *Sea $\pi : G \to H$ un morfismo de grupos. Para cualesquiera dos variables aleatorias G-valuadas X, Y, uno tiene que*

$$d[X;Y] = d[\pi(X);\pi(Y)] + d[X|\pi(X);Y|\pi(Y)] \\ + \mathbf{I}\left[X - Y \,:\, \pi(X), \pi(Y) | \pi(X) - \pi(Y)\right].$$

No es importante todavía entender qué significa la información mutua de X e Y ($\mathbf{I}\left[X : Y\right]$), pero puede entenderse como una especie de covarianza entre ambas variables.

Recordar (otra vez) que en $\mathbf{F}_2^n$ podemos reemplazar sumas por restas y viceversa. Si especializamos esta proposición al caso donde $G = \mathbf{F}_2^n \times \mathbf{F}_2^n$, $H = \mathbf{F}_2^n$, $\pi : G \to H$ es la función sumar $\pi(x + y) = x + y$, y $X = (X_1, X_2)$, $Y = (Y_1, Y_2)$ son pares de variables aleatorias independientes en $\mathbf{F}_2^n$, obtenemos el siguiente corolario:

Corolario 1.17. *Sean X_1, X_2, Y_1, Y_2 variables aleatorias $\mathbf{F}_2^n$-valuadas independientes. Tenemos entonces la identidad*

$$d[X_1, Y_1] + d[X_2; Y_2] = d[X_1 + X_2; Y_1 + Y_2] + d[X_1|X_1 + X_2; Y_1|Y_1 + Y_2] \\ + \mathbf{I}\left[(X_1 + Y_1; X_2 + Y_2) \,:\, (X_1 + X_2, Y_1 + Y_2) | X_1 + X_2 + Y_1 + Y_2\right]$$

Esta es justamente la identidad fundamental que nos da el caso entrópico, donde se ve claramente en el lado derecho de la ecuación como hay un término para la duplicación de $A+A$ y otro para la de $A\cap(A+h)$ (pensar a $X|X+Y=h$ como "los elementos de X que están en $Y+h$"), junto con un tercer "término de error".

Una forma de ver la utilidad de esta identidad puede ser descartando el término (siempre positivo) de la información mutua, obteniendo

$$d[X_1,Y_1]+d[X_2;Y_2]\geq d[X_1+X_2;Y_1+Y_2]+d[X_1|X_1+X_2;Y_1|Y_1+Y_2]. \quad (1.7)$$

Si tomamos X_1,X_2,Y_1,Y_2 como copias independientes de X,Y,Y,X (notar el intercambio de las últimas dos variables) obtenemos

$$2d[X;Y]\geq d[X+Y;X+Y]+d[X_1|X_1+X_2;Y_1|Y_1+Y_2]. \quad (1.8)$$

Mediante el cálculo entrópico de Ruzsa, se puede chequear que $X+Y$, $X_1|X_1+X_2$, y $Y_1|Y_1+Y_2$ son todas variables aleatorias relevantes (de la manera definida en (1.5)), por lo que, de (1.6), deducimos las cotas inferiores y superiores que nos dan

$$d[X+Y;X+Y]=(1+O(\eta))d[X;Y]. \quad (1.9)$$

Un buen aporte de esta estimación es que ahora podemos trabajar en el caso simétrico $X=Y$, sin perdida de generalidad. Efectivamente, si tomamos $X^*=X+Y$, obtenemos de (1.6) y (1.9) que

$$d[X'|Z;Y'|W]\geq(1-O(\eta))d[X^*;X^*] \quad (1.10)$$

siempre que $X'|Z$, $Y'|W$ sean relevantes, que por el cálculo entrópico de Ruzsa es equivalente a pedir que

$$d[X'|Z;X^*],d[Y'|W;X^*]=O(d[X^*;X^*]).$$

Usando otra vez la identidad del fibrado dada en el Corolario 1.17, renombrando Y_1,Y_2 como X_3,X_4 y tomando X_1,X_2,X_3,X_4 como copias independientes de X^*, concluimos que

$$\begin{aligned}2d[X^*;X^*]&=d[X_1+X_2;X_3+X_4]+d[X_1|X_1+X_2;X_3|X_3+X_4]\\&\quad+\mathbf{I}\left[(X_1+X_3;X_2+X_4)\,:\,(X_1+X_2,X_3+X_4)|X_1+X_2+X_3+X_4\right].\end{aligned}$$

Como antes, las variables X_1+X_2, X_3+X_4, $X_1|X_1+X_2$ y $X_3|X_3+X_4$ son todas relevantes, y de (1.10) obtenemos

$$d[X_1+X_2;X_3+X_4],d[X_1|X_1+X_2;X_3|X_3+X_4]\geq(1-O(\eta))d[X^*;X^*].$$

Notar que de todo esto conseguimos acotar fuertemente a la información mutua:

$$\mathbf{I}\left[(X_1+X_3;X_2+X_4)\,:\,(X_1+X_2,X_3+X_4)|X_1+X_2+X_3+X_4\right]=O(\eta)d[X^*;X^*].$$

Una desigualdad conocida dentro del estudio de la entropía de Shannon, la **desigualdad del procesamiento de datos** (que demostraremos en el Lema 2.3), nos ayuda a limpiar un poco esta expresión hasta obtener

$$\mathbf{I}\left[X_1 + X_3 \,:\, X_1 + X_2 | X_1 + X_2 + X_3 + X_4\right] = O(\eta) d[X^*; X^*].$$

Para proceder con la construcción de un absurdo, definamos las variables aleatorias

$$S \coloneqq X_1 + X_2 + X_3 + X_4, U \coloneqq X_1 + X_2, \text{ y } V \coloneqq X_1 + X_3,$$

las cuales dan que

$$\mathbf{I}\left[U \,:\, V | S\right] = O(\eta) d[X^*; X^*].$$

Intuitivamente, esto nos dice que U y V son casi independientes dado S, si pudiéramos lograr que $d[X^*; X^*]$ sea chica, que es uno de los objetivos. Por simplicidad del argumento, vamos a asumir que efectivamente son independientes, cosa que más adelante veremos, puede corregirse mediante una aplicación de una versión entrópica del Teorema de Balog-Szemerédi-Gowers, demostrada por Tao en [31]. Asumimos entonces que

$$\mathbf{I}\left[U \,:\, V | S\right] = 0.$$

Aquí es donde usaremos de forma fundamental el hecho de estar en un grupo con 2-torsión, ya que, en este contexto, $U + V$ tiene la misma forma que U y V:

$$U + V = X_2 + X_3,$$

es decir, la suma de dos variables con la misma distribución. Como ya dijimos, el uso de la 2-torsión, fundamental para establecer esta suma, puede arreglarse, pero va a requerir de hacer varios cambios de definiciones y notación, lo que empasta los argumentos.

En particular, por la simetría de permutar, se obtiene

$$\mathbf{H}\left(U + V | S\right) = \mathbf{H}\left(U | S\right) = \mathbf{H}\left(V | S\right),$$

y por la definición de distancia condicional de Ruzsa (que definiremos en el Capítulo 2), tenemos un decrecimiento muy significativo de la distancia

$$\mathbf{E}_s d[U | S = s; V | S = s] = 0,$$

contradiciendo (1.4) como era deseado.

Esto fue solo un recorrido por arriba de algunas ideas que se trabajan en la demostración, le dedicaremos dos capítulos enteros a rehacer recorrido detalladamente, primero para el caso de 2 torsión, y luego para el general. En ambos casos, tomaremos varios desvíos con la finalidad de hacer las estimaciones más finas, cosa que a veces puede distraer a uno de las ideas, pero recomendamos al lector tener en cuenta este esqueleto de demostración que se desarrollo recién: formulación del problema con una desigualdad entrópica, estimaciones a través

del calculo de Ruzsa y desarrollo de una contradicción a través de una información mutua muy pequeña.

Por último, aclaramos que las ideas hasta aquí relatadas "cubren" lo trabajado en [8], pero el "caso base" de esta idea inductiva fue hecha en [16], y también nos tomará la mayor parte del Capítulo 2, donde daremos una versión análoga del resultado de Ruzsa que decía que los conjuntos con constante de duplicación menor o igual a $\frac{3}{2}$ son espacios afines, de lo cual uno partía a intentar demostrar que puede hacer un algoritmo para ir achicando la constante de duplicación de su conjunto hasta situarla debajo de esta constante.

1.3 Grupos aproximados

Vamos a desarrollar, antes de volver a trabajar de lleno con entropía, la idea clásica de grupo aproximado, siguiendo la presentación de Green en su survey "Approximate Groups and Their Applications: Work of Bourgain, Gamburd, Helfgott and Sarnak" [12], con la intención de familiarizar un poco más al lector con el tipo de problemas a resolver en el área. Vamos a trabajar sobre todo con conjuntos finitos, siempre adentro de un grupo abeliano G, y vamos a tomar a este grupo como un grupo aditivo, en general (en la exposición de Green que seguimos, el autor de la misma trabaja en grupos arbitrarios, sin asumir conmutatividad, por lo que la notación y ciertos resultados son levemente distintos en esta presentación). Si $A \subseteq G$ es un conjunto finito, vamos a considerar los conjuntos suma $A + A$ y resta $A - A$ como los ya mencionados en la introducción. La siguiente proposición, cuya demostración dejamos al lector, plantea distintas maneras de pensar a un grupo como un conjunto con ciertas propiedades aditivas.

Proposición 1.18. *Sea A un subconjunto finito en un grupo ambiente G. Tenemos entonces las siguientes afirmaciones:*

1. *$|A - A| \geq |A|$, con igualdad si y sólo si $A = H + x$ para algún subgrupo $H \leq G$ y algún elemento $x \in G$;*

2. *$|A + A| \geq |A|$, con igualdad si y sólo si $A = H + x$ para algún subgrupo $H \leq G$ y algún elemento $x \in G$;*

3. *La cantidad de tuplas $(a_1, a_2, a_3, a_4) \in A^4$ con $a_1 - a_2 = a_3 - a_4$ es como mucho $|A|^3$, con igualdad si y sólo si $A = H + x$ para algún subgrupo $H \leq G$ y algún elemento $x \in G$;*

4. *La cantidad de tuplas $(a_1, a_2, a_3, a_4) \in A^4$ with $a_1 + a_2 = a_3 + a_4$ es como mucho $|A|^3$, con igualdad si y sólo si $A = Hx$ para algún subgrupo $H \leq G$ y algún elemento $x \in G$;*

5. *$\mathbf{P}(a_1 + a_2 \in A | a_1, a_2 \in A) \leq 1$, con igualdad si y sólo si $A = H$ para algún subgrupo $H \leq G$;*

6. $\mathbf{P}(a_1 - a_2 \in A | a_1, a_2 \in A) \leq 1$, *con igualdad si y sólo si* $A = H$ *para algún subgrupo* $H \leq G$.

Esta presentación puede parecer un tanto trivial y al mismo tiempo extraña para alguien con conocimientos en teoría de grupos, pero está formulada de manera que uno pueda preguntarse que pasa con los conjuntos donde las desigualdades están cerca de ser igualdades. Vamos entonces a introducir un parámetro $K \geq 1$ que indique, de cierta forma, cuan cerca están nuestros conjuntos de la igualdad; los valores grandes de K van a indicar que nuestros conjuntos aproximan más vagamente a un grupo, mientras que los valores cercanos a 1, mucha similaridad con uno.

En general, tenemos tres tipos de escenarios: K vale exactamente 1, K está muy cerca de ser 1, o K es bastante más grande que 1. Los dos primeros casos serán llamados, respectivamente, el caso del 100% y el caso del 99%, mientras que el último será el caso general.

Nos centramos entonces en la siguiente lista de propiedades que un conjunto $A \subseteq G$ podría tener:

(1) $|A - A| \leq K|A|$;

(2) $|A + A| \leq K|A|$;

(3) La cantidad de tuplas $(a_1, a_2, a_3, a_4) \in A^4$ con $a_1 - a_2 = a_3 - a_4$ es al menos $|A|^3/K$;

(4) La cantidad de tuplas $(a_1, a_2, a_3, a_4) \in A^4$ con $a_1 + a_2 = a_3 + a_4$ es al menos $|A|^3/K$;

(5) $\mathbf{P}(a_1 + a_2 \in A | a_1, a_2 \in A) \geq 1/K$;

(6) $\mathbf{P}(a_1 - a_2 \in A | a_1, a_2 \in A) \geq 1/K$.

La primera diferencia que encontramos con la Proposición 1.18 es que ahora estás propiedades ya no son equivalentes. Por ejemplo, consideremos al grupo ambiente $\mathbf{Z}$, y tomemos el conjunto

$$A = \{1, \dots, n\} \cup \{2^{n+1}, 2^{n+2}, \dots, 2^{2n}\}.$$

Es sencillo verificar que (3) y (4) se satisfacen para cualquier $K > 12$ si $n \to \infty$, porque hay $\frac{2}{3}n^3(1 + o(1))$ soluciones de $a_1 + a_2 = a_3 + a_4$ con $a_1, a_2, a_3, a_4 \in \{1, \dots, n\}$. Por otro lado, el conjunto $A + A$ contiene los números $2^{n+i} + j$ para cada par i, j con $0 < i, j \leq n$. Como estos números son todos distintos, tenemos que

$$|A + A| \geq n^2 = |A|^2/4,$$

lo que significa que si n es suficientemente grande con respecto a K entonces no se satisface (2).

Lo notable es que existe una forma de entender a todas estas propiedades como equivalentes, y para esto vamos a definir una noción de equivalencia débil:

Definición 1.19 (Equivalencia débil)**.** *Supongamos que A y B son dos conjuntos finitos en un grupo abeliano ambiente y que $K \geq 1$ es algún parámetro. Escribiremos entonces $A \sim_K B$ para denotar que existe algún x en el grupo ambiente tal que $|A \cap B + x| \geq \max(|A|, |B|)/K$. Decimos entonces que A y B son débilmente equivalentes (con parámetro K).*

Volviendo a lo anterior, fijemos cualquier elección de $j, j' \in \{1, \ldots, 6\}$, y supongamos que algún conjunto A cumple la condición (j) en la lista de arriba, con parámetro K. Existe entonces un conjunto B que satisface la condición (j') con parámetro $K' = P(K)$ (algún polinomio en K) tal que $A \sim_{K'} B$. La prueba de esto no es trivial, y referimos al lector interesado al survey [12] (si bien ahí no está probado, se dan suficientes referencias y comentarios adicionales).

Podemos dar entonces una definición formal de grupo aproximado, aunque hay que tener en cuenta las variantes que nos propone la discusión de arriba.

Definición 1.20. *Supongamos que A es un conjunto finito en algún grupo ambiente y que $K \geq 1$ es algún parámetro. Decimos entonces que A es un grupo K-aproximado si es simétrico (si $a \in A$ entonces $a^{-1} \in A$, y A incluye a la identidad) y existe un conjunto X en el grupo ambiente con $|X| \leq K$ tal que $A + A \subseteq X + A$.*

Resulta que esta noción es más o menos equivalente a las propiedades (1) - (6) de arriba, pero con algunas ventajas, como relacionarse bien con morfismos, o un buen control sobre sus conjuntos de sumas iteradas: vale que $A + A + A \subseteq X + X + A$, lo que implica que $|3A| = |A + A + A| \leq K^2|A|$, y similarmente $|nA| \leq K^{n-1}|A|$ donde nA denota el conjunto de las sumas $a_1 + \cdots + a_n$ con $a_1, \ldots, a_n \in A$.

Con todo esto en mente, podemos plantear el siguiente problema de clasificación de grupos aproximados:

Problema 1.1. *Consideremos la colección $\mathcal{C}$ de todos los grupos K-aproximados A en un grupo ambiente G. ¿Existe alguna subcolección "altamente estructurada" $\mathcal{C}'$ de forma que todo $A \in \mathcal{C}$ es débilmente equivalente a algún conjunto $B \in \mathcal{C}'$ con párametro K', donde K' depende solamente de K?*

Claramente esta pregunta puede hacerse en una amplia variedad de contextos y con mucha generalidad, pero en nuestro caso la gracia es trabajar con la noción de grupo aproximado dada por la propiedad (1), mientras que también buscamos una dependencia polinomial de K' para con K.

Para cerrar este capítulo, notemos que si queremos formalizar la idea de otras estructuras aproximadas podriamos hacerlo inspirados en el caso de grupos aproximados. Por ejemplo, si quisiéramos definir la noción de **anillo aproximado**, podríamos pedir que nuestro conjunto A cumpla

$$|A + A| \leq K\,|A| \;\text{ y }\; |A \cdot A| \leq K\,|A|\,,$$

donde definimos al conjunto $A \cdot A$ de la manera esperada. Naturalmente, esto es similar a pedir que

$$\max\big(|A + A|\,, |A \cdot A|\big) \leq K\,|A|\,,$$

y dejamos al lector pensar como esto se relaciona con el Corolario 1.8.

Capítulo 2

La Entropía en combinatoria aditiva

El objetivo de este capítulo es definir la entropía de Shannon de una variable aleatoria que toma valores en un grupo, y ver como las propiedades que esta cumple pueden servirnos para decir cosas sobre grupos aproximados. Para todos los resultados técnicos de entropía referimos al lector a un libro de Teoría de la Información, como por ejemplo [9].

Dedicaremos la primer sección del capítulo a dar la definiciones básicas para trabajar ideas entrópicas de Combinatoria Aditiva, basando la discusión principalmente en el apéndice que Tao le dedica a la entropía en su artículo "Sumset and inverse sumset theory for Shannon entropy" [31, Apéndice A], junto con un pequeño apartado donde dejaremos sentadas varias desigualdades y resultados útiles. En la segunda sección comenzará el camino hacia probar el caso del 99% de la Conjetura de Marton (caso en el cual asumiermos que la constante de duplicación está acotada por un ε tan pequeño como necesitemos), por lo que se trabajará un resultado que se apoya sobre la divergencia de Kullback-Leibler, un concepto extraído de la Teoría de la Información. En la sección posterior se llevará a cabo efectivamente la demostración de este caso inicial de la Conjetura de Marton. Como fue anticipado en el Prólogo, este fue el primer resultado fundamental que demostraron Green, Manners y Tao en "Sumsets and entropy revisited" [16] para demostrar la Conjetura de Marton, y es este artículo la fuente principal que el autor de la tesis utilizó para escribir las Secciones 2.2 y 2.3.

Por otro lado, ya habiendo cerrado la demostración del caso del 99%, la cuarta sección del capítulo cambia de rumbo, y se dedica a exponer una discusión sobre el Teorema de Balog-Szemerédi-Gowers, tanto en su versión clásica como en la entrópica. La discusión es original, pero busca explicitar la relación entre la versión clásica y entrópica del Teorema de Balog-Szemerédi-Gowers, y para esto utiliza principalmente lo desarrollado por Tao en "Sumset and inverse sumset theory for Shannon entropy" [31] (para el caso entrópico), y por Tao y Vu en el

libro "Additive Combinatorics" [35] (para el caso clásico).

Por último, la sección final cubre algunos lemas técnicos sobre entropía que serán necesarios para el caso de Marton en torsión impar, y puede ser omitida hasta el Capítulo 5.

2.1 Nociones básicas de entropía

Para comenzar, repasaremos la entropía de Shannon, siguiendo lo explicado en "Sumset and inverse sumset theory for Shannon entropy" [31, Apéndice A] por Tao. Vamos a trabajar con variables aleatorias G-valuadas, donde G es un grupo cualquiera, es decir, variables aleatorias que toman valores en G. En general, vamos a ir de estas variables aleatorias a distribuciones, comunmente notadas por p, indistintamente. En este espíritu, vamos a definir la entropía de Shannon $\mathbf{H}(p)$ de una distribución de una variable aleatoria G-valuada (con G finito) como

$$\mathbf{H}(p) := \sum_{x \in G} F(p(x)),$$

donde $F : \mathbf{R}^+ \to \mathbf{R}^+$ es la función

$$F(x) := x \log \frac{1}{x},$$

con la convención de que $F(0) = 0$. Dada una variable aleatoria X, vamos a definir $\mathbf{H}(X) := \mathbf{H}(p_X)$.

Una forma de entender este objeto de manera intuitiva es la siguiente: Supongamos que tenemos una variable aleatoria X con función de densidad $p(x) = \mathbf{P}(X = x)$, y queremos definir de alguna forma la noción de cuanto nos "sorprende" el resultado de esta variable. Para modelar esta noción, querríamos que ciertas condiciones se cumplan:

- La sorpresa de una variable aleatoria determinística debería ser 0.
- Eventos con probabilidad estrictamente menor deberían ser estrictamente más sorprendentes.
- Dos eventos independientes tienen que ser exactamente tan sorprendentes como la suma de sus sorpresas por separado.

Partiendo de esto, uno llega a convencerse de que la sorpresa de que una variable aleatoria tome un valor particular puede darse por

$$I_p(x) := -\ln(p_x),$$

y así podríamos entender a la entropía de forma intuitiva como **la esperanza de la sorpresa de la variable X**, ya que

$$\mathbf{H}(X) = \mathbf{E}_p\big[I_p(X)\big].$$

Comencemos entonces viendo, y no siempre demostrando, algunas propiedades que cumple la función F que define la entropía. Varias de las cosas que diremos sobre esta función serán estimaciones básicas, generalmente utilizando solamente propiedades de funciones cóncavas, por lo que recomendamos al lector revisar el texto de Gray [9] en caso de que algo no parezca claro, la idea es no deternos demasiado en las pruebas clásicas de la construcción de la entropía, sino ver como se usa en combinatoria aditiva.

La función F tiene como primera derivada a

$$F'(x) = log\frac{1}{x} - 1$$

y como segunda a

$$F''(x) = -\frac{1}{x},$$

para $x > 0$. De esto podemos deducir que F es cóncava en $\mathbf{R}^+$, y creciente para $x < 1/e$ (para convencerse de estas cosas, también es útil observar la Figura 2.1). En particular, vale que

$$F(x) \leq F(1/e) = 1/e$$

y

$$F(y) \leq F(x) + F'(x)(y - x) \tag{2.1}$$

para todo $y \geq 0, x > 0$. También vale la propiedad de subaditividad

$$F(x + y) \leq F(x) + F(y) \tag{2.2}$$

para todo $x, y \geq 0$, y en particular la desigualdad triangular

$$|F(a) - F(b)| \leq F(|a - b|)$$

para $0 \leq a, b \leq 1/e$. De la identidad

$$F(x) + F'(x)(y - x) - F(y) = y(\frac{x}{y} - 1 - \log\frac{x}{y})$$

y (2.1) se obtiene la cota

$$F(x) + F'(x)(y - x) - F(y) = y \log_+ \frac{y}{x} + O(x) + O(y) \tag{2.3}$$

donde $\log_+ x := \max(\log x, 0)$. Por último, de la identidad

$$F(ax) = F(a)F(x)(\frac{1}{\log\frac{1}{a}} + \frac{1}{\log\frac{1}{x}})$$

podemos deducir

$$F(ax) \leq 2F(a)F(x)$$

siempre que $0 \leq a, x \leq 1/e$.

Tenemos entonces la primer estimación:

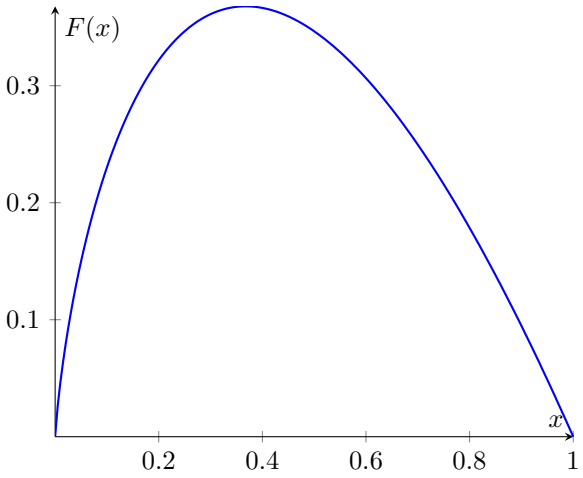


Figure 2.1: Gráfico de $F(x) = x \log \frac{1}{x}$.

Lema 2.1. *Sea A un conjunto finito, y sea X una variable aleatoria $A-$valuada. Luego $\mathbf{H}(X) \leq \log |A|$. Más aún, si $\mathbf{H}(X) \geq \log |A| - \log K$ para algún $K \geq 1$, entonces*

$$\sum_{k=1}^{\infty} 2^k \mathbf{P}(X \in A_k) \ll 1 + \log(K),$$

donde

$$A_k := \{x \in A : 2^{2^{k-1}} \leq p_X(x)\,|A| \leq 2^{2^k}\}.$$

Demostración. Para la primer cota, observamos que

$$\begin{aligned}\mathbf{H}(X) &= \sum_{x \in A} F(p_X(x)) \\ &\leq \sum_{x \in A} \Big(F\Big(\frac{1}{|A|}\Big) + F'\Big(\frac{1}{|A|}\Big)\Big(p_X(x) - \frac{1}{|A|}\Big)\Big) \\ &= \log |A|,\end{aligned}$$

tal como se requiere, donde la desigualdad del medio se justifica con (2.1). De manera similar, si $\mathbf{H}(X) \geq \log |A| - \log K$, entonces el argumento anterior muestra que

$$\sum_{x \in A} \Big(F\Big(\frac{1}{|A|}\Big) + F'\Big(\frac{1}{|A|}\Big)\Big(p_X(x) - \frac{1}{|A|}\Big) - F(p_X(x))\Big) \leq \log K.$$

Por (2.1), el sumando es no negativo; y por (2.3), el sumando es

$$p_X(x) \log\Big(|A|\, p_X(x)\Big) + O\big(p_X(x)\big)$$

para $p_X(x) \geq 1/|A|$. La segunda afirmación se sigue al descomponer la variable x en los conjuntos A_k. □

También vamos a trabajar con versiones condicionadas de la entropía, siempre en un contexto finito. Definimos la **entropía condicional** $\mathbf{H}(X|Y)$ con la fórmula

$$\mathbf{H}(X|Y) := \sum_{y \in Im(Y)} p_Y(y)\mathbf{H}(X|Y=y), \tag{2.4}$$

donde a la derecha aparece una variable aleatoria condicionada.

Simplemente haciendo la cuenta se puede ver que

$$\mathbf{H}(X|Y) = \mathbf{H}(X,Y) - \mathbf{H}(Y) \tag{2.5}$$

y que

$$\mathbf{H}(X|Y) = \mathbf{H}(X,Y|Y).$$

Usando la fórmula de probabilidad total, la definición de entropía condicional, y la desigualdad de Jensen con la concavidad de F, se puede concluir que

$$\mathbf{H}(X|Y) \leq \mathbf{H}(X) \tag{2.6}$$

donde la igualdad vale si y solo si $(X|Y=y) \equiv X$ para todo $y \in Im(Y)$, o, en otras palabras, si X e Y son independientes, ya que la función es estrictamente cóncava. Con esto y (2.5), se concluye que

$$\mathbf{H}(X,Y) \leq \mathbf{H}(X) + \mathbf{H}(Y), \tag{2.7}$$

con igualdad si y solo si son independientes.

Diremos que una variable aleatoria Y esta **determinada** por otra variable X si vale que $Y = f(X)$, donde $f : Im(X) \to Im(Y)$ es una función. Se ve directo de la propiedad de subaditividad (2.2) que

$$\mathbf{H}(Y) \leq \mathbf{H}(X) \tag{2.8}$$

siempre que X determine a Y. Por ejemplo, como (X,Y) determina a X y a Y, vale que

$$\mathbf{H}(X), \mathbf{H}(Y) \leq \mathbf{H}(X,Y),$$

y usando las relaciones (2.5), (2.6) y (2.8) puede deducirse que

$$\mathbf{H}(X) - \mathbf{H}(Y) \leq \mathbf{H}(X|Y) \leq \mathbf{H}(X). \tag{2.9}$$

Si X determina a Y, entonces X y (X,Y) se determinan entre ellas, de lo que se deduce que $\mathbf{H}(X,Y) = \mathbf{H}(X)$. En particular,

$$\mathbf{H}(X|Y) = \mathbf{H}(X) - \mathbf{H}(Y),$$

cuando X determina a Y, y $\mathbf{H}(Y|X) = 0$.

La siguiente desigualdad, clásica en Teoría de la Información, será fundamental a lo largo de todo el texto:

Lema 2.2 (Desigualdad de **submodularidad**). *Si X_0, X_1, X_2, X_{12} son variables aleatorias tales que X_1 y X_2 cada una de ellas determina a X_0, y (X_1, X_2) determina X_{12}, entonces*

$$\mathbf{H}(X_{12}) + \mathbf{H}(X_0) \leq \mathbf{H}(X_1) + \mathbf{H}(X_2).$$

Demostración. Por (2.5) y (2.9) basta demostrar que

$$\mathbf{H}(X_{12}|X_0) \leq \mathbf{H}(X_1|X_0) + \mathbf{H}(X_2|X_0).$$

Por (2.4) basta demostrar que

$$\mathbf{H}(X_{12}|X_0 = x_0) \leq \mathbf{H}(X_1|X_0 = x_0) + \mathbf{H}(X_2|X_0 = x_0)$$

para todo $x_0 \in \text{rango}(X_0)$. Pero, por hipótesis, $(X_1|X_0 = x_0)$ y $(X_2|X_0 = x_0)$ determinan $(X_{12}|X_0 = x_0)$, y la afirmación se sigue de (2.7) y (2.8). □

Un caso especial de este lema es la observación de que

$$\mathbf{H}(Y|Z) \leq \mathbf{H}(X|Z) \tag{2.10}$$

siempre que (X, Z) determine a Y. De forma similar, vale que

$$\mathbf{H}(X, Y|Z) \leq \mathbf{H}(X|Z) + \mathbf{H}(Y|Z) \tag{2.11}$$

para cualquier terna X, Y, Z, con igualdad si y solo si X e Y son condicionalmente independientes con respecto a Z.

Vamos a definir la **información mutua $\mathbf{I}[\mathbf{X} : \mathbf{Y}]$** mediante la fórmula

$$\begin{aligned}\mathbf{I}[X : Y] &:= \mathbf{H}(X) + \mathbf{H}(Y) - \mathbf{H}(X, Y) \\ &= \mathbf{H}(X) - \mathbf{H}(X|Y) \\ &= \mathbf{H}(Y) - \mathbf{H}(Y|X).\end{aligned}$$

Es claro por lo visto antes que esta cantidad es no negativa, y que se anula exactamente cuando X e Y son independientes.

Se puede definir también la **información mutua condicional $\mathbf{I}[\mathbf{X} : \mathbf{Y}|\mathbf{Z}]$** como

$$\mathbf{I}[X : Y|Z] := \sum_z p_Z(z)\mathbf{I}[(X|Z = z) : (Y|Z = z)]. \tag{2.12}$$

Observemos que la **submodularidad** es equivalente a la afirmación de que

$$\mathbf{I}[X : Y|Z] \geq 0, \tag{2.13}$$

ya que

$$\mathbf{I}[X : Y|Z] = \mathbf{H}(X, Z) + \mathbf{H}(Y, Z) - \mathbf{H}(X, Y, Z) - \mathbf{H}(Z).$$

A lo largo del texto, vamos a referirnos como desigualdad de submodularidad indistintamente a esta versión, o al Lema 2.2.

Otra desigualdad clásica de teoría de la información que vamos a usar es la **desigualdad de procesamiento de datos**, que se enuncia a través de la información mutua.

Lema 2.3. *Sean X, Y, Z variables aleatorias. Para cualesquiera dos funciones f, g definidas en los rangos de X, Y respectivamente, vale que*

$$\mathbf{I}[f(X) : g(Y)|Z] \leq \mathbf{I}[X : Y|Z].$$

Demostración. Basta con probar la versión no condicionada $\mathbf{I}[f(X) : g(Y)] \leq \mathbf{I}[X : Y]$ de esta desigualdad, ya que la versión condicionada resulta de condicionar con los eventos $Z = z$, multiplicar por $p_Z(z)$, y sumar sobre todo z.

Por simetría e iterando, basta con probar la desigualdad que involucra a uno solo de los lados

$$\mathbf{I}[f(X) : Y] \leq \mathbf{I}[X : Y],$$

o equivalentemente que $\mathbf{H}[Y|X] \leq \mathbf{H}[Y|f(X)]$.

Pero como X determina $f(X)$, se tiene que $\mathbf{H}[Y|X] = \mathbf{H}[Y|f(X), X]$, y el resultado se sigue aplicando submodularidad (2.13). □

2.1.1 Entropía en grupos

Hasta ahora, no usamos en estos resultados ni definiciones el hecho de que las variables aleatorias tomen valores sobre sobre un grupo. En lo que sigue, vamos a pensar que G es un grupo abeliano. Siempre que X, Y son dos variables aleatorias G-valuadas, vale que

$$\mathbf{H}(X \pm Y) \geq \mathbf{H}(X \pm Y|Y) = \mathbf{H}(X|Y) = \mathbf{H}(X) - \mathbf{I}[X : Y], \tag{2.14}$$

y de forma análoga para los roles de X e Y invertidos, por lo que obtuvimos

$$\max(\mathbf{H}(X), \mathbf{H}(Y)) - \mathbf{I}[X : Y] \leq \mathbf{H}(X \pm Y).$$

De forma análoga se puede obtener también que

$$\max(\mathbf{H}(X|Z), \mathbf{H}(Y|Z)) - \mathbf{I}[X : Y|Z] \leq \mathbf{H}(X \pm Y|Z),$$

y, si X e Y son independientes, que

$$\max(\mathbf{H}(X), \mathbf{H}(Y)) \leq \mathbf{H}(X \pm Y). \tag{2.15}$$

Si continuamos en la suposición de independencia y tomando $d[X; Y]$ como la distancia entrópica de Ruzsa, definida en (1.3), vale que

$$d[X; Y] = \mathbf{H}(X - Y) - \frac{1}{2}\mathbf{H}(X) - \frac{1}{2}\mathbf{H}(Y),$$

que combinado con (2.15) nos devuelve que

$$|\mathbf{H}(X) - \mathbf{H}(Y)| \leq 2d[X; Y] \tag{2.16}$$

y que

$$\mathbf{H}(X - Y) - \mathbf{H}(X), \mathbf{H}(X - Y) - \mathbf{H}(Y) \leq 2d[X; Y].$$

Una de las propiedades más importantes de la distancia entrópica es la **desigualdad triangular de Ruzsa**:

Lema 2.4. *Sean X, Y, Z tres variables aleatorias G-valuadas. Vale entonces que*

$$d[X;Y] \leq d[X;Z] + d[Z;Y].$$

Demostración. Sin pérdida de generalidad podemos asumir que las variables son independientes, cosa que asumiremos al inicio de la gran mayoría de las demostraciones, debido a que en la definición de distancia de Ruzsa se toman copias independientes de las variables. Se tiene entonce que nuestro interés está en demostrar

$$\mathbf{H}(X-Y) \leq \mathbf{H}(X-Z) + \mathbf{H}(Y-Z) - \mathbf{H}(Z). \tag{2.17}$$

Por submodularidad, vale que

$$\mathbf{H}(X-Y) + \mathbf{H}(X-Z, Y, X-Y) \leq \mathbf{H}(X-Z, X-Y) + \mathbf{H}(Y, X-Y)$$

y observando que

$$\begin{aligned} &\mathbf{H}(X-Z, X-Y) = \mathbf{H}(X-Z, Y-Z) \leq \mathbf{H}(X-Z) + \mathbf{H}(Y-Z), \\ &\mathbf{H}(Y, X-Y) = \mathbf{H}(X,Y), \\ \text{y } &\mathbf{H}(X-Z, Y, X-Y) = \mathbf{H}(X,Y,Z) = \mathbf{H}(X,Y) + \mathbf{H}(Z), \end{aligned}$$

se concluye el resultado. □

Definiremos también variantes condicionales de la distancia. Si (X, Z) e (Y, W) son variables aleatorias (donde X e Y son G-valuadas), definiremos

$$d[X|Z;Y|W] := \sum_{z,w} p_Z(z) p_W(w) d[(X|Z=z);(Y|W=w)] \tag{2.18}$$

como la **distancia condicional de Ruzsa**. De forma alternativa, se pueden tomar $(X', Z'), (Y', W')$ copias independientes de las variables anteriores, y observar que

$$d[X|Z;Y|W] = \mathbf{H}(X'-Y'|Z', W') - \frac{1}{2}\mathbf{H}(X'|Z') - \frac{1}{2}\mathbf{H}(Y'|W').$$

2.1.2 Algunos resultados útiles.

Habiendo introducido las definiciones básicas de los objetos a trabajar, pasamos a hacer un rejunte de distintos lemas y proposiciones que necesitaremos usar en el texto, todos involucrando la entropía. La idea de esta sección es que sirva de referencia en los otros capítulos, y puede resultar algo disconexa entre resultado y resultado, por lo que se recomienda leerla a sabiendas de esto.

El primer resultado es una desigualdad muy parecida a la de submodularidad:

Lema 2.5. *Sean X, Y, Z variables aleatorias independientes tomando valores en un grupo abeliano. Entonces*

$$\mathbf{H}(X+Y+Z) - \mathbf{H}(X+Y) \leq \mathbf{H}(Y+Z) - \mathbf{H}(Y).$$

Demostración. Por la definición de la información mutua condicional (2.12) tenemos

$$\begin{aligned}\mathbf{I}\left[X\,:\,Z\mid X+Y+Z\right] &= \mathbf{H}[X, X+Y+Z] + \mathbf{H}[Z, X+Y+Z] \\ &\quad - \mathbf{H}[X, Z, X+Y+Z] - \mathbf{H}[X+Y+Z].\end{aligned}$$

Sin embargo, usando el hecho de que podemos separar la entropía de un vector aleatorio como la suma de las entropías de sus coordenadas, si son independientes, obtenemos que

$$\mathbf{H}[X, X+Y+Z] = \mathbf{H}[X, Y+Z] = \mathbf{H}[X] + \mathbf{H}[Y+Z],$$

$$\mathbf{H}[Z, X+Y+Z] = \mathbf{H}[Z, X+Y] = \mathbf{H}[Z] + \mathbf{H}[X+Y],$$

y

$$\mathbf{H}[X, Z, X+Y+Z] = \mathbf{H}[X, Y, Z] = \mathbf{H}[X] + \mathbf{H}[Y] + \mathbf{H}[Z].$$

Tras un breve cálculo, vemos que la desigualdad enunciada es equivalente a afirmar que

$$\mathbf{I}\left[X\,:\,Z\mid X+Y+Z\right] \geq 0,$$

lo cual, por supuesto, es una instancia de la no negatividad de la información mutua (2.13). □

Podemos generalizar esto al siguiente resultado, donde también lo relacionamos a una desigualdad con respecto a las distancias de varias variables:

Proposición 2.6. *Sea G un grupo abeliano, y sean X, Y, Z variables aleatorias G-valuadas. Si $X, Y_1, \ldots, Y_n$ son variables aleatorias independientes, entonces*

$$\mathbf{H}\Big[X + \textstyle\sum_{i=1}^n Y_i\Big] - \mathbf{H}[X] \leq \displaystyle\sum_{i=1}^n \big(\mathbf{H}[X+Y_i] - \mathbf{H}[X]\big) \tag{2.19}$$

y

$$\mathrm{d}\Big[X; \textstyle\sum_{i=1}^n Y_i\Big] \leq 2 \displaystyle\sum_{i=1}^n d[X; Y_i]. \tag{2.20}$$

Demostración. La desigualdad (2.19) puede demostrarse haciendo inducción a partir del caso $n = 2$, que es el Lema 2.5. Si nos dirigimos ahora a (2.20), podemos usar la desigualdad (2.19) en la definición de distancia de Ruzsa, reescribiendo Y_i por $-Y_i$, para obtener

$$\begin{aligned}d\big[X; \textstyle\sum_{i=1}^n Y_i\big] + \tfrac{1}{2}\Big(\mathbf{H}\Big[\textstyle\sum_{i=1}^n Y_i\Big] - \mathbf{H}[X]\Big) \\ \leq \sum_{i=1}^n \Big(d[X; Y_i] + \tfrac{1}{2}\big(\mathbf{H}[X] - \mathbf{H}[Y_i]\big)\Big).\end{aligned}$$

Por el hecho de que $\mathbf{H}[X+Y] \geq \max(\mathbf{H}[X], \mathbf{H}[Y])$ en el caso de independencia tenemos

$$\mathbf{H}\big[\sum_{i=1}^n Y_i\big] \geq \tfrac{1}{n} \sum_{i=1}^n \mathbf{H}[Y_i],$$

entonces podemos reordenar las cosas para llegar a

$$d\big[X;\textstyle\sum_{i=1}^n Y_i\big] \le \displaystyle\sum_{i=1}^n d[X;Y_i] + \frac{n-1}{2n}\sum_{i=1}^n \big(\mathbf{H}[Y_i] - \mathbf{H}[X]\big).$$

La desigualdad (2.20) ahora se deduce usando que $\mathbf{H}[Y_i] - \mathbf{H}[X] \le 2d[X;Y_i]$ que probamos en (2.16); de hecho podría reemplazarse el 2 de (2.20) por la constante $(2n-1)/n$, que es apenas mejor. □

El siguiente lema es un resultado del estilo del teorema de Balog-Szemerédi-Gowers, pero con la distancia entrópica. Este resultado es crucial para la demostración de la Conjetura de Marton, en particular para los argumentos utilizados en [8], y es discutido en profundidad en la Sección 2.4, pero lo enunciamos y probamos ahora porque puede verse como una desigualdad que involucra a la distancia entrópica de Ruzsa.

Lema 2.7. *Sea (A,B) una variable aleatoria G^2-valuada, y sea $Z := A+B$. Entonces*

$$\sum_z p_Z(z) d[(A|Z=z);(B|Z=z)] \le 3\mathbf{I}[A:B] + 2\mathbf{H}(Z) - \mathbf{H}(A) - \mathbf{H}(B). \tag{2.21}$$

En esta demostración necesitaremos la noción de **ensayos condicionalmente independientes** de un par de variables aleatorias (X,Y) (no necesariamente independientes). Decimos que X_1, X_2 son ensayos condicionalmente independientes de X respecto a Y al declarar que $(X_1 \mid Y=y)$ y $(X_2 \mid Y=y)$ son copias independientes de $(X \mid Y=y)$ para todo y en el rango de Y. De este modo, tenemos que

$$\mathbf{H}[(X_1 \mid Y=y),(X_2 \mid Y=y)] = 2\mathbf{H}[X \mid Y=y]$$

para todo y, lo cual, al sumar sobre y (ponderado por $p_Y(y)$), da

$$\mathbf{H}[X_1, X_2 \mid Y] = 2\mathbf{H}[X \mid Y],$$

y por lo tanto, usando (5.8), se ve que

$$\begin{aligned}\mathbf{H}[X_1,X_2,Y] = \mathbf{H}[X_1,X_2 \mid Y] + \mathbf{H}[Y] &= 2\mathbf{H}[X \mid Y] + \mathbf{H}[Y] \\ &= 2\mathbf{H}[X,Y] - \mathbf{H}[Y]. \end{aligned} \tag{2.22}$$

Notar también que las distribuciones marginales de (X_1,Y) y (X_2,Y) coinciden con la distribución original (X,Y).

Demostración del Lema 2.7. Sean (A_1,B_1) y (A_2,B_2) ensayos condicionalmente independientes de (A,B) respecto a Z, de modo que (A_1,B_1) y (A_2,B_2) están acoplados mediante la variable aleatoria $A_1+B_1 = A_2+B_2$, a la que, por abuso de notación, llamaremos también Z.

Observamos que el lado izquierdo de (2.21) es

$$\mathbf{H}[A_1 - B_2 \mid Z] - \tfrac{1}{2}\mathbf{H}[A_1 \mid Z] - \tfrac{1}{2}\mathbf{H}[B_2 \mid Z], \tag{2.23}$$

puesto que, crucialmente, $(A_1 \mid Z = z)$ y $(B_2 \mid Z = z)$ son independientes para todo z.

Aplicando la submodularidad (Lema 2.2) se obtiene

$$\begin{aligned} &\mathbf{H}[A_1 - B_2] + \mathbf{H}[A_1 - B_2, A_1, B_1] \\ &\quad \leq \mathbf{H}[A_1 - B_2, A_1] + \mathbf{H}[A_1 - B_2, B_1]. \end{aligned} \tag{2.24}$$

Procedemos a estimar el segundo, tercer y cuarto término que aparecen en esta desigualdad. Primero notar, teniendo en cuenta que la tupla $(A_1 - B_2, A_1, B_1)$ determina la tupla (A_1, A_2, B_1, B_2) ya que $A_1 + B_1 = A_2 + B_2$, que

$$\begin{aligned} \mathbf{H}[A_1 - B_2, A_1, B_1] &= \mathbf{H}[A_1, B_1, A_2, B_2] \\ &= \mathbf{H}[A_1, B_1, A_2, B_2 | Z] + \mathbf{H}[Z] = 2\mathbf{H}[A, B] - \mathbf{H}[Z]. \end{aligned} \tag{2.25}$$

Luego, observemos que

$$\mathbf{H}[A_1 - B_2, A_1] = \mathbf{H}[A_1, B_2] \leq \mathbf{H}[A] + \mathbf{H}[B], \tag{2.26}$$

y finalmente, tenemos que

$$\mathbf{H}[A_1 - B_2, B_1] = \mathbf{H}[A_2 - B_1, B_1] = \mathbf{H}[A_2, B_1] \leq \mathbf{H}[A] + \mathbf{H}[B]. \tag{2.27}$$

Sustituyendo (2.25), (2.26) y (2.27) en (2.24) se deduce

$$\mathbf{H}[A_1 - B_2] \leq 2\mathbf{I}\,[A \,:\, B] + \mathbf{H}[Z],$$

y así, como condicionar disminuye la entropía (2.6),

$$\mathbf{H}[A_1 - B_2 \mid Z] \leq 2\mathbf{I}\,[A \,:\, B] + \mathbf{H}[Z].$$

Dado que

$$\begin{aligned} \mathbf{H}[A_1 \mid Z] &= \mathbf{H}[A_1, A_1 + B_1] - \mathbf{H}[Z] \\ &= \mathbf{H}[A, B] - \mathbf{H}[Z] \\ &= \mathbf{H}[A] + \mathbf{H}[B] - \mathbf{I}\,[A \,:\, B] - \mathbf{H}[Z], \end{aligned}$$

y de forma similar para $\mathbf{H}[B_2 \mid Z]$, se deduce que la expresión en (2.23) está acotada por

$$3\mathbf{I}\,[A \,:\, B] + 2\mathbf{H}[Z] - \mathbf{H}[A] - \mathbf{H}[B],$$

tal como se afirma. □

Lo siguiente es una relación básica entre la distancia de dos variables y esa misma distancia, pero cuando una de ellas resta, cosa que es algo que ignoramos en el caso de 2-torsión, pero será útil para los otros casos. Por lo general, el conjunto suma $A + A$ no se comporta exactamente igual a $A - A$, pero tenemos estimaciones como esta que son útiles:

Lema 2.8. $d[X;-Y] \leq 3d[X;Y]$

Demostración. Sean X, Y variables aleatorias G-valuadas independientes. Sean $(X_1, Y_1), (X_2, Y_2)$ ensayos condicionalmente independientes de (X, Y) relativos a $X - Y$; como (X, Y) determina $X - Y$, concluimos que $X_1 - Y_1 = X_2 - Y_2$. Sea (X_3, Y_3) otro ensayo de (X, Y), independiente de X_1, X_2, Y_1, Y_2, entonces tenemos la identidad

$$X_3 + Y_3 = (X_3 - Y_2) - (X_1 - Y_3) + X_2 + Y_1.$$

Luego $(X_3 - Y_2, X_1 - Y_3, X_2, Y_1)$ y (X_3, Y_3) cada uno determinan a $X_3 + Y_3$, mientras que $(X_3 - Y_2, X_1 - Y_3, X_2, Y_1)$ y (X_3, Y_3) juntos determinan a $(X_1, X_2, X_3, Y_1, Y_2, Y_3)$; por lo que aplicamos submodularidad y concluimos que

$$\begin{aligned}\mathbf{H}(X_1, X_2, X_3, Y_1, Y_2, Y_3) + \mathbf{H}(X_3 + Y_3) \leq \\ \mathbf{H}(X_3 - Y_2, X_1 - Y_3, X_2, Y_1) + \mathbf{H}(X_3, Y_3).\end{aligned}$$

Pero de (2.7) y la hipótesis de independencia podemos ver que

$$\begin{aligned}\mathbf{H}(X_1, X_2, X_3, Y_1, Y_2, Y_3) &= 2\mathbf{H}(X, Y) - \mathbf{H}(X - Y) + \mathbf{H}(X) + \mathbf{H}(Y) \\ \mathbf{H}(X_3 + Y_3) &= \mathbf{H}(X + Y) \\ \mathbf{H}(X_3 - Y_2, X_1 - Y_3, X_2, Y_1) &\leq 2\mathbf{H}(X - Y) + \mathbf{H}(X) + \mathbf{H}(Y) \\ \mathbf{H}(X_3, Y_3) &= \mathbf{H}(X) + \mathbf{H}(Y)\end{aligned}$$

y por ende

$$\mathbf{H}(X + Y) \leq 3\mathbf{H}(X - Y) - \mathbf{H}(X) - \mathbf{H}(Y) \tag{2.28}$$

que se puede reorganizar para concluir el resultado. □

Una variante de la distancia entrópica de Ruzsa que nos será útil es la **distancia entrópica maximal de Ruzsa**, definida como:

$$d^*[X;Y] \coloneqq \sup_{X',Y'} \big(\mathbf{H}(X' - Y') - \frac{1}{2}\mathbf{H}(X') - \frac{1}{2}\mathbf{H}(Y')\big),$$

donde las (X', Y') recorren todas las parejas de copias de (X, Y), sin que les pidamos independencia.

El siguiente resultado es una desigualdad triangular entrópica más fuerte que el Lema 2.4:

Lema 2.9. *Sean X, Y, Z variables aleatorias con valores en G. Entonces:*

(i) *Se tiene que $d^*[X;Z] \leq d[X;Y] + d[Y;Z]$.*

(ii) *Se tiene que $d[X;Y] \leq d^*[X;Y] \leq 3d[X;Y]$.*

Demostración del Lema 2.9. Comenzamos con la parte (i). Basta con demostrar que

$$\mathbf{H}(X - Z) - \frac{1}{2}\big(\mathbf{H}(X) + \mathbf{H}(Z)\big) \leq d[X;Y] + d[Y;Z].$$

Esto es equivalente a establecer que

$$\mathbf{H}(X-Z) \leq \mathbf{H}(X-Y) + \mathbf{H}(Y-Z) - \mathbf{H}(Y)$$

siempre que Y sea independiente de (X, Z) (aunque no se requiere que X y Z sean independientes entre sí).

Aplicamos la desigualdad de submodularidad y obtenemos

$$\mathbf{H}(X-Y, Z, X-Z) + \mathbf{H}(X-Z) \leq \mathbf{H}(X-Y, X-Z) + \mathbf{H}(Z, X-Z).$$

Con estas elecciones tenemos

$$\mathbf{H}(X-Y, Z, X-Z) = \mathbf{H}(X, Y, Z) = \mathbf{H}(X, Z) + \mathbf{H}(Y),$$

$$\mathbf{H}(X-Y, X-Z) = \mathbf{H}(X-Y, Y-Z) \leq \mathbf{H}(X-Y) + \mathbf{H}(Y-Z)$$

y

$$\mathbf{H}(Z, X-Z) = \mathbf{H}(X, Z).$$

Reordenando todo esto, obtenemos la primer desigualdad del enunciado.

Para la parte (ii), la primera desigualdad $d[X;Y] \leq d^*[X;Y]$ es directa de las definiciones. Para la segunda desigualdad, aplicamos (i) y la desigualdad triangular para concluir que

$$\begin{aligned} d^*[X;Y] &\leq d[X;Y] + d[Y;Y] \\ &\leq d[X;Y] + d[Y;X] + d[X;Y], \end{aligned}$$

lo que prueba la afirmación. □

Un resultado que se puede probar usando la distancia maximal es el siguiente:

Lema 2.10. *Sea G un grupo sin torsión, y sean X e Y dos variables aleatorias con valores en G. Vale entonces que $d[X;2Y] \leq 5d[X;Y]$.*

Demostración. Podemos asumir que X e Y son independientes. Luego

$$\begin{aligned} \mathbf{H}(X-2Y) &= \mathbf{H}((X-Y)-Y) \\ &\leq d^*[X-Y;Y] + \frac{1}{2}\mathbf{H}(X-Y) + \frac{1}{2}\mathbf{H}(Y) \end{aligned} \tag{2.29}$$

por la definición de la distancia maximal. Por el Lema 2.9,

$$\begin{aligned} d^*[X-Y;Y] &\leq d[Y;Y] + d[X-Y;Y] \\ &\leq 2d[X;Y] + d[X-Y;Y]. \end{aligned} \tag{2.30}$$

Tomando Y_1 e Y_2 como dos copias independientes de Y (también independientes de X) tenemos

$$d[X-Y;Y] = \mathbf{H}(X-Y_1-Y_2) - \frac{1}{2}\mathbf{H}(X-Y) - \frac{1}{2}\mathbf{H}(Y). \tag{2.31}$$

Escribiendo $A \coloneqq Y_1$, $B \coloneqq Y_2$ y $C \coloneqq X - Y_1 - Y_2$, resulta

$$\mathbf{H}(A, B, C) = \mathbf{H}(X, Y_1, Y_2) = \mathbf{H}(X) + 2\mathbf{H}(Y),$$

y

$$\mathbf{H}(A, C) = \mathbf{H}(A, C + A) = \mathbf{H}(Y_1, X - Y_2) = \mathbf{H}(Y) + \mathbf{H}(X - Y_2),$$
$$\mathbf{H}(B, C) = \mathbf{H}(B, C + B) = \mathbf{H}(Y_2, X - Y_1) = \mathbf{H}(Y) + \mathbf{H}(X - Y_1)$$

y al aplicar la desigualdad de submodularidad obtenemos

$$\mathbf{H}(X - Y_1 - Y_2) \leq \mathbf{H}(X - Y_1) + \mathbf{H}(X - Y_2) - \mathbf{H}(X).$$

Combinando esto con (2.31) nos da

$$d[X - Y; Y] \leq \frac{3}{2}\mathbf{H}(X - Y) - \mathbf{H}(X) - \frac{1}{2}\mathbf{H}(Y)$$

que, junto con (2.29) y (2.30), nos da

$$\mathbf{H}(X - 2Y) \leq 2d[X; Y] + 2\mathbf{H}(X - Y) - \mathbf{H}(X) = 4d[X; Y] + \mathbf{H}(Y)$$

y por ende

$$d[X; 2Y] \leq 4d[X, Y] + \frac{1}{2}(\mathbf{H}(Y) - \mathbf{H}(X)) \leq 5d[X; Y]$$

donde usamos (2.16) para el último paso.

□

Casi para finalizar la sección, dejamos el siguiente resultado relacionando las constantes de duplicación combinatoria y entrópica, que fueron discutidas en la Introducción. Para enunciarlo utilizaremos la noción de **energía** de un conjunto A, que notaremos como

$$\mathrm{E}(A) \coloneqq |\{(a, b, c, d) \in A^4 : a + b = c + d\}|,$$

y recordamos que las constantes de duplicación clásica y entrópica eran

$$\sigma(A) = e^{d(A;A)} = |A + A| / |A|$$

y

$$\sigma[X] = d[X; X] = \mathbf{H}[X + X'] - \mathbf{H}[X],$$

donde X' y X son variables aleatorias independientes con la misma distribución.

Proposición 2.11. $\frac{|A|^3}{\mathrm{E}(A)} \leq \sigma[A] \leq \sigma(A)$.

Demostración. Denotamos $X \coloneqq U_A + U'_A$ como la suma de dos variables aleatorias uniformes independientes en A. La desigualdad del lado derecho es inmediata

a partir de la desigualdad $\mathbf{H}(X) \leq \log|A+A|$. En cuanto a la desigualdad del lado izquierdo, se observa que

$$p_X(x) = \frac{|A \cap (x-A)|}{|A|^2}.$$

La desigualdad aritmética-geométrica ponderada dice que dados $a_1, \ldots, a_n > 0$ números reales positivos, y $w_1, w_2, \ldots, w_n \geq 0$ pesos reales no negativos tales que

$$w_1 + w_2 + \cdots + w_n = 1,$$

entonces se cumple:

$$a_1^{w_1} a_2^{w_2} \cdots a_n^{w_n} \leq w_1 a_1 + w_2 a_2 + \cdots + w_n a_n,$$

con igualdad si y sólo si $a_1 = a_2 = \cdots = a_n$ (cuando todos los $w_i > 0$).

De esto y la definición de entropía puede deducirse que

$$e^{-\mathbf{H}(X)} = \prod_x p_X(x)^{p_X(x)} \leq \sum_x p_X(x)^2 = \frac{\mathrm{E}(A)}{|A|^4}.$$

El resultado sigue inmediatamente. □

Es conveniente dejar por sentado también que, como $\mathbf{H}(X) \geq -\ln p_{max}$, se tiene

$$\max_x p_X(x) \geq e^{-\mathbf{H}(X)}. \tag{2.32}$$

Fue mencionada en la Introducción, en la Proposición 1.16, la relación que hay entre la entropía y los morfismos. Pasamos a dar un ejemplo de este fenómeno, aplicado a cocientes de grupos. Sea X una variable aleatoria con valores en G, y sea H un subgrupo finito de G. Denotemos por $\pi : G \to G/H$ la aplicación cociente. Sea U_H una variable aleatoria uniforme en H, independiente de X. Entonces tenemos

$$\begin{aligned}
\mathbf{H}(X + U_H) &= \mathbf{H}(X + U_H, \pi(X)) \\
&= \mathbf{H}(\pi(X)) + \mathbf{H}(X + U_H | \pi(X)) \\
&= \mathbf{H}(\pi(X)) + \mathbf{H}(U_H) \\
&= \mathbf{H}(\pi(X)) + \log|H|.
\end{aligned}$$

Para justificar la tercer igualdad primero notamos que

$$p_{X+U_H|\pi(X)=c}(y) = \sum_{h \in H} p_{U_H}(h) p_{X|\pi(X)=c}(y-h) = \frac{1}{|H|} \sum_{h \in H} p_{X|\pi(X)=c}(y-h),$$

y que

$$\sum_{h \in H} p_{X|\pi(X)=c}(y-h) = \sum_{x \in c+H} p_{X|\pi(X)=c}(x) = 1,$$

por probabilidad total, o 0 si $\pi(y) \neq c$. Luego, usamos las definiciones de entropía para verificar que

$$\begin{aligned}
\mathbf{H}\left(X + U_H | \pi(X)\right) &= \sum_{c \in G/H} p_{\pi(X)}(c) \mathbf{H}\left(X + U_H | \pi(X) = c\right) \\
&= \sum_{c \in G/H} p_{\pi(X)}(c) \big(- \sum_{y \in G} p_{X+U_H|\pi(X)=c}(y) \ln(p_{X+U_H|\pi(X)=c}(y)) \big) \\
&= \sum_{c \in G/H} p_{\pi(X)}(c) \big(- \sum_{y \in c+H} p_{X+U_H|\pi(X)=c}(y) \ln(p_{X+U_H|\pi(X)=c}(y)) \big) \\
&= \sum_{c \in G/H} p_{\pi(X)}(c) \mathbf{H}\left(U_H\right) \\
&= \mathbf{H}\left(U_H\right).
\end{aligned}$$

Se deduce entonces que

$$d[X; U_H] \;=\; \mathbf{H}\left(\pi(X)\right) \;+\; \frac{1}{2}\big(\log|H| \;-\; \mathbf{H}\left(X\right)\big). \tag{2.33}$$

A partir de esto y de la relación entre la diferencia de las entropías y la distancia dada en (2.16), obtenemos

$$\mathbf{H}\left(\pi(X)\right) \;\leq\; 2\, d[X; U_H]. \tag{2.34}$$

Además, por el Lema 2.9 y ya que, como ahora veremos, $d[U_H; U_H] = 0$, se observa que

$$d^*[X; U_H] \;=\; d[X; U_H].$$

En efecto, la distancia entre una variable aleatoria uniforme en un subgrupo consigo misma es 0, cosa que es una instancia del fenómeno más general, muchas veces útil en el contexto de grupos finitos, de que la entropía de la suma de una variable aleatoria soportada en un grupo H y una uniforme independiente en ese mismo grupo siempre es $\log(|H|)$. Esto puede verse ya que

$$\begin{aligned}
\mathbf{H}[U_H + X] &= -\sum_h p_{U_H+X}(h) \log(p_{U_H+X}(h)) \\
&= -\sum_h p_{U_H}(h) \log(p_{U_H}(h)) = \mathbf{H}[U_H] = \log(|H|),
\end{aligned}$$

donde usamos que $p_{U_H+X}(h) = p_{U_H}(h)$ siempre que X sea una variable aleatoria soportada en H.

2.2 Divergencia de Kullback-Leibler y variables entre variables.

El objetivo central de esta sección es demostrar la Proposición 2.12, esencialmente un resultado que nos permite, dadas dos variables aleatorias X e Y, crear una

variable aleatoria Z cercana en distancia de Ruzsa a ambas, pero con una constante de duplicación bien acotada. Esto lo vamos a aprovechar en la siguiente sección para probar la Conjetura de Marton en el caso en el cual la constante de duplicación es bastante pequeña (en particular, la formulación entrópica de este caso de la Conjetura de Marton es 2.12). La fuente trabajada para desarrollar estos temas es "Sumsets and entropy revisited" [16], un artículo escrito por Green, Manners y Tao unos meses antes de publicar la solución de la Conjetura de Marton, en el cual continúan el trabajo de Tao en "Sumset and inverse sumset theory for Shannon entropy"[31] desarrollando una teoría entrópica de la combinatoria aditiva, y prueban la versión de la conjetura recién mencionada. Si bien omitiremos las referencias explícitas en cada uno de los lemas, proposiciones y teoremas que probemos, todo el contenido de estas dos secciones puede encontrarse en ese artículo.

Es importante dejar en claro que lo que esta sección busca es demostrar la Proposición 2.12, y el contenido fundamental de esta demostración está encapsulado en el Lema 2.13, donde haremos un argumento probabilístico para fabricarnos un conjunto S con las características necesarias para demostrar la Proposición 2.12.

Para demostrar la Proposición 2.12, es fundamental la **divergencia de Kullback-Leibler**, que nos dará una nueva noción de distancia entre dos variables aleatorias. Supongamos que X, Y son variables aleatorias con funciones de distribución p_X, p_Y, respectivamente. Entonces definimos

$$D_{\mathrm{KL}}(X\|Y) \coloneqq \sum_t p_X(t) \log\left(\frac{p_X(t)}{p_Y(t)}\right).$$

Es convencional definir que el sumando aquí sea 0 si $p_X(t) = 0$ e ∞ si $p_Y(t) = 0$ pero $p_X(t) \neq 0$; en la práctica, evitaremos esta última situación. Una referencia para estudiar la divergencia de Kullback-Leibler es [9, Capítulo 3.2].

Pensando otra vez a la noción de entropía como promedio de la sorpresa asociada a la variable aleatoria X, tenemos que

$$D_{\mathrm{KL}}(X\|Y) = \mathbf{E}_p\big(I_q(X) - I_p(X)\big),$$

si q es la función de distribución de Y, lo que nos dice que, si pensamos que Y es una distribución que conocemos, con la cual estamos intentando estimar la distribución de X, la divergencia de Kullback-Leibler nos está indicando la diferencia entre la esperanza de la sorpresa de nuestro modelo Y contra la verdadera esperanza de la sorpresa del modelo original X. Tenemos entonces una nueva noción de "distancia" entre dos variables aleatorias.

Es conveniente relacionar a la divergencia de Kullback-Leibler con la **entropía cruzada**

$$\mathbf{H}(X : Y) \coloneqq \sum_t p_X(t) \log \frac{1}{p_Y(t)} \tag{2.35}$$

(donde se aplican las mismas convenciones para no dividir por 0). Así,

$$D_{\mathrm{KL}}(X\|Y) = \mathbf{H}(X : Y) - \mathbf{H}(X). \tag{2.36}$$

En particular, si X toma valores en un conjunto finito S, entonces $\mathbf{H}(X : U_S) = \log|S|$ y, por lo tanto,

$$D_{\mathrm{KL}}(X\|U_S) = \log|S| - \mathbf{H}(X). \tag{2.37}$$

De una aplicación estándar de la desigualdad de Jensen se obtiene la **desigualdad de Gibbs**

$$D_{\mathrm{KL}}(X\|Y) \geq 0; \tag{2.38}$$

también tenemos la **desigualdad de Pinsker** (cuya demostración se puede encontrar, por ejemplo, en [9, Lema 6.2])

$$\sum_t |p_X(t) - p_Y(t)| \leq \sqrt{2D_{\mathrm{KL}}(X\|Y)}. \tag{2.39}$$

Si X, Y, Z son variables aleatorias con valores en un grupo G (no necesariamente independientes), observamos a partir de la desigualdad de Gibbs la siguiente cota útil

$$\begin{aligned}\mathbf{H}(Z-Y) - \mathbf{H}(Y) &\leq \mathbf{H}(Z-Y:X) - \mathbf{H}(Y)\\ &= \sum_z p_Z(z)\big(\mathbf{H}(z-Y:X) - \mathbf{H}(z-Y)\big)\\ &= \sum_z p_Z(z) D_{\mathrm{KL}}(z-Y\|X)\\ &= \mathbf{E}[D_{\mathrm{KL}}(\tilde{X}-Y\|X)]\end{aligned} \tag{2.40}$$

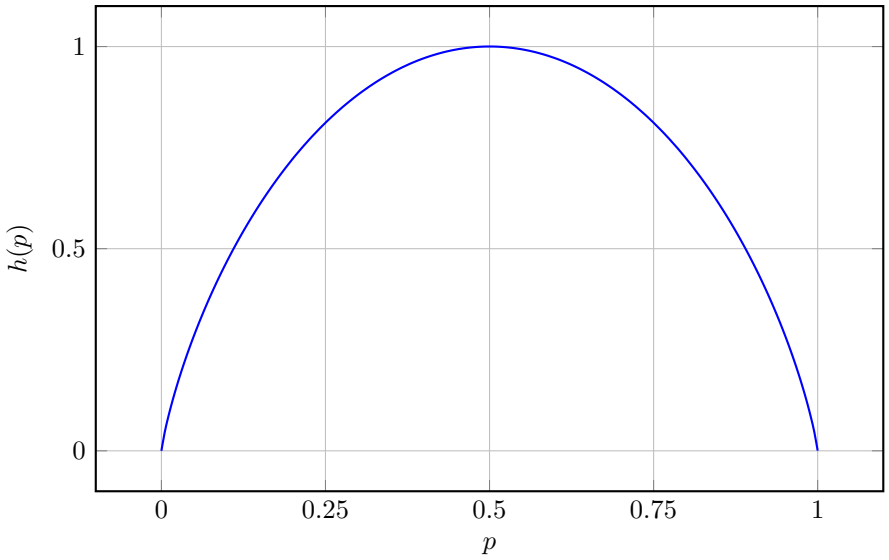
donde hemos usado la invarianza por traslación por constantes de la entropía de Shannon para observar que $\mathbf{H}(z-Y) = \mathbf{H}(Y)$, así como el hecho de que $p_{Z-Y}(t) = \sum_z p_Z(z)p_{z-Y}(t)$ (notar también que en la última expresión estamos tomando esperanza sobre la variable aleatoria $\tilde{X}$, que es el único objeto "aleatorio" en esa expresión, ya que los otros se entienden fijos al tomar esa esperanza). Nótese que tenemos igualdad en (2.40) cuando $X = Z - Y$.

Debido a que condicionar una variable aleatoria a eventos que ocurren o no ocurren es estándar, suele ser útil trabajar con variables aleatorias con distribución Bernoulli, y es por esto que de aquí en adelante, dado un parámetro real $p \in (0,1)$, vamos a escribir

$$\mathrm{h}(p) \coloneqq p\log(\frac{1}{p}) + (1-p)\log(\frac{1}{1-p})$$

como la entropía de una variable aleatoria Bernoulli de parámetro p.

Entropía de una variable Bernoulli: $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$



Ahora estamos en condiciones de enunciar el principal resultado de esta sección:

Proposición 2.12. *Sea $C \geq 4$ un parámetro real. Para cualquier par de variables aleatorias G-valuadas X, Y existe un subconjunto finito no vacio S de G tal que, si U_S es una variable aleatoria uniforme en S, entonces*

$$d^*[U_S; Y] \leq (C+2)d[X;Y] + \mathrm{h}\left(1 - \frac{2}{C}\right) \tag{2.41}$$

y

$$\log \frac{|S-S|}{|S|} \leq (2C+4)d[X;Y] + 2\mathrm{h}\left(1 - \frac{2}{C}\right). \tag{2.42}$$

Aislamos un caso especial de esta proposición para poder usar en la siguiente sección: Si $d[X;Y] = \varepsilon$ para algún $0 < \varepsilon \leq \frac{1}{16}$, entonces al tomar $C := \varepsilon^{-1/2}$ obtenemos las cotas $d[U_S;Y] \ll \varepsilon^{1/2}\log\frac{1}{\varepsilon}$ y $|S-S| \leq \big(1 + O(\varepsilon^{1/2}\log\frac{1}{\varepsilon})\big)|S|$. En efecto, para observar esto hay que notar que el término predominante en

$$(\varepsilon^{-1/2}+2)\varepsilon + (1-2\varepsilon^{1/2})\log(\frac{1}{1-2\varepsilon^{1/2}}) + (2\varepsilon^{1/2})\log(\frac{1}{2\varepsilon^{1/2}}),$$

cuando ε es muy pequeño, es

$$(2\varepsilon^{1/2})\log(\frac{1}{2\varepsilon^{1/2}}) = \varepsilon^{1/2}\log(\frac{1}{2\varepsilon}) = \varepsilon^{1/2}\log(\frac{1}{\varepsilon}) + \varepsilon^{1/2}\log(\frac{1}{2}),$$

y lo que domina a este último término es $\varepsilon^{1/2}\log(\frac{1}{\varepsilon})$, que era lo que queríamos ver.

Para probar la Proposición 2.12 primero necesitamos el siguiente resultado, en cuya demostración se encapsula la idea para fabricar el conjunto que nos interesa, dejando para después la tarea de refinarlo hasta que satisfaga todo lo que exige el enunciado de la Proposición 2.12. Para este lema asumiremos que C, X, Y son como en el enunciado de la Proposición 2.12.

Lema 2.13. *Existe un subconjunto finito no vacío S de G tal que*

$$\log|S| \geq \mathbf{H}(Y) - 2\mathrm{h}\left(1 - \frac{2}{C}\right) - 4d[X;Y] \tag{2.43}$$

y tal que

$$d^*[Z;Y] \leq Cd[X;Y] + \frac{1}{2}\big(\mathbf{H}(Y) - \mathbf{H}(Z)\big) \tag{2.44}$$

para toda variable aleatoria Z con valores en S.

Demostración. Como X e Y son independientes, tenemos

$$\mathbf{H}(X-Y) = \frac{1}{2}\mathbf{H}(X) + \frac{1}{2}\mathbf{H}(Y) + d[X;Y], \tag{2.45}$$

y por lo tanto, usando la cota sobre la diferencia de entropías y la distancia (2.16), se sigue que

$$\mathbf{H}(X-Y) - \mathbf{H}(Y) \leq 2d[X;Y].$$

Aplicando el caso de igualdad en (2.40), con $X = X - Y$ y $Z = X$, obtenemos también que

$$\sum_x p_X(x) D_{\mathrm{KL}}(x - Y \| X - Y) \leq 2d[X;Y]. \tag{2.46}$$

Notar que lo que aquí hicimos fue desglosar la distancia en un promedio compuesto por varios pedazos, deduciendo que si la distancia de Ruzsa de X e Y es pequeña, entonces la suma ponderada de las divergencias entre algunas fibras y $X - Y$ también debe ser pequeña. Inspirados por esto, definimos S mediante la fórmula

$$S := \big\{x : p_X(x) > 0, D_{\mathrm{KL}}(x - Y \| X - Y) \leq Cd[X;Y]\big\}. \tag{2.47}$$

Denotamos por A la variable aleatoria $A := 1_{X\in S}$, y escribimos $p := \mathbf{P}(A = 1) = \mathbf{P}(X \in S)$. Por la desigualdad de Markov, que dice que

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}(X)}{a},$$

aplicada a la variable aleatoria $D_{\mathrm{KL}}(x - Y \| X - Y)$ con $a = Cd[X;Y]$ y (2.46), se sigue que

$$p = \mathbf{P}(X \in S) \geq 1 - \frac{2}{C} \geq \frac{1}{2}, \tag{2.48}$$

ya que $C > 4$.

Ahora hacemos algunas observaciones. Primero,

$$\begin{aligned}\mathbf{H}(X) &= \mathbf{H}(X, A)\\ &= \mathbf{H}(X|A) + \mathbf{H}(A)\\ &= p\mathbf{H}(X|A=1) + (1-p)\mathbf{H}(X|A=0) + \mathrm{h}(p). \end{aligned} \tag{2.49}$$

Segundo, dado que Y es independiente de X y de A, se sigue, usando que la entropía de la suma de variables independientes acota a las entropías de las variables por separado (2.15), que

$$\mathbf{H}(X - Y|A = i) \geq \mathbf{H}(Y), \mathbf{H}(X|A = i)$$

para $i = 0, 1$; por lo tanto,

$$\begin{aligned}\mathbf{H}(X - Y) &\geq \mathbf{H}(X - Y|A)\\ &= p\mathbf{H}(X - Y|A = 1) + (1 - p)\mathbf{H}(X - Y|A = 0)\\ &\geq p\mathbf{H}(Y) + \frac{1-p}{2}\big(\mathbf{H}(Y) + \mathbf{H}(X|A = 0)\big). \end{aligned} \tag{2.50}$$

Combinando (2.49) y (2.50) con (2.45) concluimos, tras un breve cálculo, que

$$k \geq \frac{p}{2}\mathbf{H}(Y) - \frac{p}{2}\mathbf{H}(X|A = 1) - \frac{1}{2}\mathrm{h}(p). \tag{2.51}$$

Como X condicionada a $A = 1$ está soportada en S, tenemos que $\mathbf{H}(X|A = 1) \leq \log|S|$. Sustituyendo en (2.51) y reorganizando se obtiene

$$\log|S| \geq \mathbf{H}(Y) - \frac{\mathrm{h}(p)}{p} - \frac{2d[X;Y]}{p}.$$

Usando (2.48) (y el hecho de que $\mathrm{h}(p)$ es decreciente para $p \geq 1/2$), se obtiene (2.43).

Ahora demostraremos (2.44). A partir de (2.40) (reemplazando X por $X - Y$) tenemos

$$\mathbf{H}(Z - Y) - \mathbf{H}(Y) \leq \sum_z p_Z(z) D_{\mathrm{KL}}(z - Y\|X - Y).$$

Primero, observemos que la divergencia de Kullback–Leibler está bien definida y es finita. Efectivamente, Z toma valores $z \in S$, y por la definición de S tenemos que $p_X(z) > 0$ para tales z. Así, si $p_{z-Y}(t) > 0$, entonces

$$p_{X-Y}(t) = \sum_x p_X(x) p_{x-Y}(t) \geq p_X(z) p_{z-Y}(t) > 0.$$

Por la definición de S, $D_{\mathrm{KL}}(z - Y\|X - Y) \leq Cd[X;Y]$ para z en el rango de Z, y se sigue la afirmación (2.44), ya que

$$\begin{aligned} d^*[Z;Y] &= \mathbf{H}[\tilde{Z} - Y] - \frac{1}{2}\mathbf{H}[Y] - \frac{1}{2}\mathbf{H}[Z]\\ &\leq Cd[X;Y] + \frac{1}{2}(\mathbf{H}[Y] - \mathbf{H}[Z]), \end{aligned}$$

donde la primera igualdad es válida para alguna copia $\tilde{Z}$ de Z, pero las dos expresiones de los extremos la ignoran, ya que $\mathbf{H}[Z] = \mathbf{H}[\tilde{Z}]$. □

Ahora estamos listos para la demostración de la Proposición 2.12.

Demostración de la Proposición 2.12. Comenzamos estableciendo la primera desigualdad (2.41). Sea S como en el Lema 2.13. Al tomar $Z = U_S$ en (2.44), tenemos

$$d^*[U_S; Y] \leq Cd[X; Y] + \frac{1}{2}\big(\mathbf{H}(Y) - \log|S|\big).$$

Entonces, la desigualdad pedida en (2.41) se deduce de (2.43).

Ahora probamos (2.42). Sea Z, Z' un par arbitrario de variables aleatorias con valores en S. A partir de la desigualdad triangular para la distancia maximal y (2.44) tenemos

$$\begin{aligned} d^*[Z; Z'] &\leq d[Z; Y] + d[Z'; Y] \\ &\leq 2Cd[X; Y] + \mathbf{H}(Y) - \frac{1}{2}\mathbf{H}(Z) - \frac{1}{2}\mathbf{H}(Z') \end{aligned}$$

o, equivalentemente,

$$\mathbf{H}(Z - Z') \leq 2Cd[X; Y] + \mathbf{H}(Y). \tag{2.52}$$

Ahora observamos que es posible elegir Z, Z' con soporte en S de manera que $Z - Z'$ tenga la distribución uniforme en $S - S$. Para ello, basta con tomar (Z, Z') con función de distribución

$$p_{(Z,Z')}(s_1, s_2) \coloneqq \frac{1}{|S - S|\,|\{(t_1, t_2) \in S : t_1 - t_2 = s_1 - s_2\}|}$$

para $s_1, s_2 \in S$. En este caso, $\mathbf{H}(Z - Z') = \log|S - S|$. Utilizando (2.52) y (2.43), se deduce (2.42). □

2.3 El caso del 99%

Esta sección tiene por objetivo, como ya anticipamos, probar el **caso del 99%** de la Conjetura de Marton, es decir, representa el caso en el cual queremos cubrir con pocas coclases de un subgrupo a un conjunto con una constante de duplicación muy pequeña. La versión entrópica principal que daremos de este resultado es:

Proposición 2.14. *Existe una constante absoluta $\varepsilon_0 > 0$ tal que lo siguiente es verdadero. Sean X, Y variables aleatorias G-valuadas, y supongamos que $d[X; Y] \leq \varepsilon_0$. Existe entonces un subgrupo finito H de G tal que $d[X; U_H]$, $d[Y; U_H] \leq 12d[X; Y]$.*

La versión clásica de este resultado (es decir, la versión dada en un contexto aditivo, sin entropía) es un resultado de Freiman (ver [3]), será el Teorema 2.15. Es fundamental destacar que este es el resultado que la Proposición 2.14 busca reformular en un contexto entrópico, y es para esto que desarrollamos tanto la presente sección, como la anterior. El objetivo de probar este tipo de resultados

en la presente tesis es para usarlos como caso base de un argumento inductivo que nos permitirá probar la Conjetura de Marton en los capítulos 4 y 5, en los cuales utilizaremos la versión concreta del caso del 100%, que formulamos como Corolario 2.20. Esta sección, como la anterior, recorre resultados de "Sumsets and entropy revisited" [16].

Primero tratamos el caso $X = Y$ (el cual se establecerá en la Proposición 2.14 con la constante mejorada de 6 en lugar de 12). Asumamos de aquí en adelante que X es una variable aleatoria con valores en G y que $d[X;X] = \varepsilon \leq \varepsilon_0$.

Observamos primero que una versión débil de la Proposición 2.14 se deduce rápidamente de la Proposición 2.12: efectivamente, como se observó después del enunciado de la Proposición 2.12, se tiene que existe un conjunto S tal que

$$d[U_S;Y] \ll \varepsilon^{1/2} \log \frac{1}{\varepsilon} \tag{2.53}$$

y

$$|S-S| \leq \left(1 + O\left(\varepsilon^{1/2} \log \frac{1}{\varepsilon}\right)\right)|S| < \frac{3}{2}|S|, \tag{2.54}$$

donde la última desigualdad se cumple si ε_0 es lo suficientemente pequeño.

Notar que aquí no logramos obtener todavía un subgrupo, pero sí un conjunto con constante de duplicación muy pequeña, que vendría a ser un buen paso en esa dirección. Tiene sentido entonces que para este paso nos ayude el caso del 99% del problema clásico, y justamente es lo que usaremos ahora. Esto es en realidad una observación bien conocida de Freiman (ver [3]) que implica que $H \coloneqq S - S$ es un grupo. Recordamos aquí la breve demostración.

Teorema 2.15. *Sea S un conjunto en un grupo abeliano G tal que la constante de duplicación de S es menor o igual a $\frac{3}{2}$. Se tiene entonces que $H \coloneqq S - S$ es un subgrupo de G.*

Demostración. Para cualesquiera $x, y \in S$, los conjuntos $x - S$ e $y - S$ están contenidos en $S - S$, por lo que

$$|(x-S) \cap (y-S)| > \frac{1}{2}|S|.$$

Es decir, existen más de $\frac{1}{2}|S|$ pares $(u,v) \in S \times S$ tales que $x - u = y - v$. Para cada uno de estos pares se tiene $x - y = u - v$. Ahora, sean $x', y' \in S$ otros elementos cualesquiera. De manera similar, existen más de $\frac{1}{2}|S|$ pares $(u', v') \in S \times S$ tales que $x' - y' = u' - v'$. Existen más de $\frac{1}{2}|S|$ valores de v y más de $\frac{1}{2}|S|$ valores de u', y todos estos valores pertenecen a S; por lo tanto, debe ocurrir que $v = u'$ para algún par de estos elementos. Se sigue que

$$(x-y) + (x'-y') = (u-v) + (u'-v') = u - v' \in S - S.$$

Puesto que x, y, x', y' fueron elegidos arbitrariamente, se deduce que $S - S$ es cerrado bajo la suma. Dado que contiene al 0 y es cerrado bajo la toma de inversos, debe ser un grupo. □

A partir del estudio que hicimos sobre la relación de la entropía con con las proyecciones al cociente (2.33) (notando que S está contenido en una única coclase de H) y (2.54), tenemos que

$$d[U_H, U_S] = \frac{1}{2}\log\frac{|H|}{|S|} \ll \varepsilon^{1/2}\log\frac{1}{\varepsilon}.$$

Por lo tanto, a partir de (2.53) y de la desigualdad triangular se tiene

$$d[U_H, Y] \ll \varepsilon^{1/2}\log\frac{1}{\varepsilon}. \tag{2.55}$$

Esto es más débil que la Proposición 2.14 únicamente en la dependencia no lineal respecto a ε.

Deduciremos la afirmación de la Proposición 2.14 partiendo desde esta cota. La estrategia que vamos a usar para mejorar estas cotas es, a grandes rasgos, descomponer la distancia de X a X en dos objetos, uno relacionado a una variable aleatoria concentrada en un punto, y otro relacionado a una variable casi uniforme, y para controlar estos dos casos vamos a enunciar los Lemas 2.16 y 2.18. Trabajemos primero el caso en que X está altamente concentrada cerca de un punto:

Lema 2.16. *Existe $\delta_0 > 0$ tal que se cumple lo siguiente. Supongamos que X es una variable aleatoria con valores en G y que $x_0 \in G$ es un valor tal que $\mathbf{P}(X = x_0) \geq 1 - \delta_0$. Entonces $\mathbf{H}(X) \leq 2d[X; X]$.*

Demostración. Reemplazando X por $X - x_0$ si es necesario, podemos asumir sin pérdida de generalidad que $x_0 = 0$. Sean X_1, X_2 copias independientes de X. Entonces, por la definición de la distancia entrópica, nuestra tarea es demostrar que

$$\mathbf{H}(X_1 - X_2) \geq \frac{3}{2}\mathbf{H}(X). \tag{2.56}$$

Definamos $p := \mathbf{P}(X \neq 0)$ (por lo tanto, $p \leq \delta_0$), y sea A la función indicadora del suceso de que $X_1, X_2 \neq 0$; entonces $\mathbf{P}(A = 0) = 1 - p^2$ y $\mathbf{P}(A = 1) = p^2$. Como consecuencia, tenemos

$$\begin{aligned}\mathbf{H}(X_1 - X_2) &\geq \mathbf{H}(X_1 - X_2|A)\\ &= (1-p^2)\mathbf{H}(X_1 - X_2|A = 0) + p^2\mathbf{H}(X_1 - X_2|X_1, X_2 \neq 0)\\ &\geq (1-p^2)\mathbf{H}(X_1 - X_2|A = 0) + p^2\mathbf{H}(X|X \neq 0).\end{aligned} \tag{2.57}$$

Observamos ahora que para cualquier z, si $A = 0$ y $X_1 - X_2 = z$, entonces (X_1, X_2) solo puede tomar dos valores, $(z, 0)$ y $(0, -z)$ si $z \neq 0$, y un único valor $(0, 0)$ si $z = 0$. Por lo tanto,

$$\begin{aligned}\mathbf{H}(X_1, X_2|A = 0) - \mathbf{H}(X_1 - X_2|A = 0)&\\ = \mathbf{H}(X_1, X_2|X_1 - X_2, A = 0)&\\ \leq \mathbf{P}(X_1 - X_2 \neq 0|A = 0)\log 2 = \frac{2p(1-p)}{1-p^2}\log 2.&\end{aligned}$$

Combinando con (2.57), se obtiene

$$\mathbf{H}(X_1 - X_2) \geq (1-p^2)\mathbf{H}(X_1, X_2|A=0) + p^2\mathbf{H}(X|X\neq 0) - 2p(1-p)\log 2. \tag{2.58}$$

Observamos también que

$$\begin{aligned} 2\,\mathbf{H}(X) &= \mathbf{H}(X_1, X_2) = \mathbf{H}(X_1, X_2, A) = \mathbf{H}(X_1, X_2|A) + \mathbf{H}(A) \\ &= (1-p^2)\mathbf{H}(X_1, X_2|A=0) + p^2\mathbf{H}(X_1, X_2|A=1) + \mathrm{h}(p^2) \\ &= (1-p^2)\mathbf{H}(X_1, X_2|A=0) + 2p^2\mathbf{H}(X|X\neq 0) + \mathrm{h}(p^2). \end{aligned} \tag{2.59}$$

Además, notamos que, definiendo I como la función indicadora de $X \neq 0$, como X deteremina a I, de 2.15 se obtiene que

$$\mathbf{H}(X) = \mathbf{H}(X|I) + \mathbf{H}(I) = p\,\mathbf{H}(X|X\neq 0) + \mathrm{h}(p). \tag{2.60}$$

Restando p veces (2.60) de (2.59) se obtiene

$$(2-p)\mathbf{H}(X) = (1-p^2)\mathbf{H}(X_1, X_2|A=0) + p^2\mathbf{H}(X|X\neq 0) + \mathrm{h}(p^2) - p\,\mathrm{h}(p).$$

Combinando esto con (2.58), se obtiene

$$\mathbf{H}(X_1 - X_2) \geq (2-p)\mathbf{H}(X) + p\,\mathrm{h}(p) - 2p(1-p)\log 2 - \mathrm{h}(p^2). \tag{2.61}$$

Recordemos que nuestro objetivo es demostrar (2.56). Para obtenerlo a partir de (2.61), notamos primero que,

$$2p(1-p)\log 2 + \mathrm{h}(p^2) \leq \left(\frac{1}{2} - p\right)\mathrm{h}(p) \tag{2.62}$$

siempre que δ_0 sea lo suficientemente pequeño: el lado izquierdo es $\sim 2p\log 2$, mientras que el lado derecho es $\sim \frac{1}{2}p\log\frac{1}{p}$ (Recordar que $p \leq \delta_0$). (Un análisis más cuidadoso muestra que $\delta_0 = \frac{1}{20}$ es suficiente.) También tenemos

$$\mathrm{h}(p) = \mathbf{H}(I) \leq \mathbf{H}(X). \tag{2.63}$$

La cota deseada (2.56) se sigue inmediatamente de (2.61), (2.62) y (2.63). □

Observación 2.17. *La constante* 2 *en la afirmación del Lema 2.16 puede ser reemplazada por cualquier número mayor que* 1*, a costa de hacer* δ_0 *más pequeño. Esto se puede demostrar con modificaciones muy leves del argumento anterior.*

A continuación, consideramos el caso de una variable aleatoria con valores en H, que ahora suponemos suficientemente uniforme.

Lema 2.18. *Supongamos que* X *es una variable aleatoria con valores en un grupo* H *y que* $\mathbf{H}(X) \geq \log|H| - \frac{1}{8}$*. Entonces*

$$\log|H| - \mathbf{H}(X) \leq 2d[X, X].$$

Para demostrar esto utilizaremos el siguiente lema acerca de acoplamientos de variables aleatorias casi uniformes, que resulta de interés independiente. Aquí, para una distribución de probabilidad p en un grupo H, se escribe

$$\|p - u_H\|_1 := \sum_{x \in H} \big|p(x) - \frac{1}{|H|}\big|$$

para la distancia ℓ^1 de p a la distribución uniforme.

Lema 2.19. *Supongamos que $p_1, p_2, p_3 \colon H \to \mathbf{R}_{\geq 0}$ son tres distribuciones de probabilidad en H tales que*

$$\|p_1 - u_H\|_1 + \|p_2 - u_H\|_1 + \|p_3 - u_H\|_1 \leq 1. \tag{2.64}$$

Entonces existe un par de variables aleatorias (X, Y) sobre H (no necesariamente independientes) que tienen como distribuciones marginales $p_X = p_1$, $p_Y = p_2$ y $p_{X-Y} = p_3$.

Demostración. Queremos demostrar que el trío de distribuciones $(p_1, p_2, p_3) \in \mathbf{R}^H \times \mathbf{R}^H \times \mathbf{R}^H$ se encuentra en la cápsula convexa del conjunto

$$\Sigma := \{(\delta_x, \delta_y, \delta_{x-y}) : x, y \in H\} \subseteq \mathbf{R}^H \times \mathbf{R}^H \times \mathbf{R}^H.$$

Aquí usamos la notación habitual $\delta_t(u) = 1$ si $u = t$, y $\delta_t(u) = 0$ en caso contrario (dejamos al lector observar que efectivamente probar esto implica el resultado deseado). Por el Teorema de Hahn–Banach (del cual usamos una versión sencilla, la de un espacio vectorial de dimensión finita), esto es equivalente a demostrar que no existe ningún hiperplano que separe (p_1, p_2, p_3) de Σ, o lo que es lo mismo, que siempre que $f_1, f_2, f_3 \colon H \to \mathbf{R}$ sean funciones tales que

$$f_1(x) + f_2(y) + f_3(x - y) \geq 0 \tag{2.65}$$

para todo $x, y \in H$, se tiene también

$$\sum_{x \in H} f_1(x)p_1(x) + \sum_{y \in H} f_2(y)p_2(y) + \sum_{z \in H} f_3(z)p_3(z) \geq 0. \tag{2.66}$$

A partir de ahora, asumamos (2.65). Notemos que tanto (2.65) como (2.66) no se ven afectados si se desplazan f_1, f_2, f_3 por constantes c_1, c_2, c_3 cuya suma sea cero. Por lo tanto, podemos normalizar de modo que

$$\min f_1 = \min f_2 = \min f_3,$$

simplemente resolviendo un sistema lineal de ecuaciones. Si este valor es no negativo, entonces (2.66) es inmediata, por lo que podemos asumir que es negativo. Reescalando, podemos normalizar de modo que

$$\min f_1 = \min f_2 = \min f_3 = -1. \tag{2.67}$$

En particular, existe $x_0 \in H$ tal que $f_1(x_0) = -1$. De (2.65) y (2.67) se concluye que para todo $y \in H$ se tiene

$$f_2(y) + f_3(x_0 - y) \geq 1 \quad \text{y} \quad \min(f_2(y), f_3(x_0 - y)) \geq -1.$$

Esto implica que

$$\begin{aligned}
&f_2(y)p_2(y) + f_3(x_0 - y)p_3(x_0 - y) \\
&\quad \geq (f_2(y) + f_3(x_0 - y)) \min(p_2(y), p_3(x_0 - y)) \\
&\qquad + \min(f_2(y), f_3(x_0 - y))\big|p_2(y) - p_3(x_0 - y)\big| \\
&\quad \geq \min(p_2(y), p_3(x_0 - y)) - \big|p_2(y) - p_3(x_0 - y)\big| \\
&\quad = \frac{p_2(y) + p_3(x_0 - y)}{2} - \frac{3}{2}\big|p_2(y) - p_3(x_0 - y)\big| \\
&\quad \geq \frac{p_2(y) + p_3(x_0 - y)}{2} - \frac{3}{2}\left|p_2(y) - \frac{1}{|H|}\right| - \frac{3}{2}\left|p_3(x_0 - y) - \frac{1}{|H|}\right|.
\end{aligned}$$

Sumando sobre y, se concluye que

$$\sum_{y \in H} f_2(y)p_2(y) + \sum_{z \in H} f_3(z)p_3(z) \geq 1 - \frac{3}{2}\|p_2 - u_H\|_1 - \frac{3}{2}\|p_3 - u_H\|_1.$$

Permutando cíclicamente los roles de f_1, f_2, f_3 y de p_1, p_2, p_3 y promediando, la cota deseada (2.66) se sigue de (2.64). □

Demostración del Lema 2.18. Por (2.37) y la desigualdad de Pinsker (2.39) se tiene que

$$\|p_X - u_H\|_1 \leq \sqrt{2(\log|H| - \mathbf{H}(X))} \leq \frac{1}{2}. \tag{2.68}$$

Aplicando el Lema 2.19 (con $p_1 = p_2 = p_X$ y $p_3 = \frac{1}{|H|}$), se deduce que existe un par de variables aleatorias (X_1, X_2) tales que X_1 y X_2 tienen cada una la misma distribución marginal que X, y $X_1 - X_2$ es uniforme en H.

Finalmente, el Lema 2.9 da

$$\begin{aligned}
\log|H| = \mathbf{H}(X_1 - X_2) &\leq \mathbf{H}(X) + d^*[X; X] \\
&\leq \mathbf{H}(X) + d[X; X] + d[X; X],
\end{aligned}$$

lo cual implica inmediatamente el resultado. □

Ya habiendo desarrollado las dos situaciones extremas, X uniforme o X muy concentrada, demostramos la primera versión de la Conjetura de Marton entrópica del texto:

Demostración de la Proposición 2.14. Recordar que primero vamos a establecer el caso $X = Y$ (con la constante 12 reemplazada por 6). Tenemos, por hipótesis, que $d[X; X] = \varepsilon \leq \varepsilon_0$. Sea $\pi : G \to G/H$ la proyección al cociente, tomando como H al grupo definido utlizando el Teorema de Freiman, Teorema 2.15, que

venía dado como $H = S - S$ para un S particular, que se elegía en función de ε. Recordemos por (2.33) que

$$d[X, U_H] = \mathbf{H}(\pi(X)) + \frac{1}{2}\big(\log|H| - \mathbf{H}(X)\big). \tag{2.69}$$

Lo que sigue de la demostración consistirá en acotar los términos de la derecha de la expresión de arriba, primero $\mathbf{H}(\pi(X))$ con el Lema 2.16, y luego $\log|H| - \mathbf{H}(X)$ mediante el Lema 2.18. Notamos que, como estamos suponiendo que "X se parece mucho a la distribución uniforme en un subgrupo H", esperaríamos que ambos términos sean muy chicos, y lo fundamentaremos con los últimos lemas.

Tenemos la cota débil (2.55)

$$d[X, U_H] \ll \varepsilon^{1/2} \log \frac{1}{\varepsilon}.$$

Por lo tanto, de (2.69) se ve que

$$\mathbf{H}(\pi(X)) \ll \varepsilon^{1/2} \log \frac{1}{\varepsilon} \tag{2.70}$$

y que

$$\mathbf{H}(X) \geq \log|H| - O\bigg(\varepsilon^{1/2} \log \frac{1}{\varepsilon}\bigg). \tag{2.71}$$

Notar que (2.70) nos está diciendo que, como la entropía de esta variable aleatoria es muy pequeña, la probabilidad debe concentrarse mucho en un solo valor, lo que nos indica que el Lema 2.16 puede ayudarnos a acotar.

Ahora, aplicando la Proposición 4.3[1] (reemplazando H allí por G/H, y recordando que $d[X; X] = \varepsilon$) obtenemos

$$d\big[\pi(X), \pi(X)\big] \leq \varepsilon, \tag{2.72}$$

y

$$\sum_{y_1, y_2 \in G/H} p_{\pi(X)}(y_1) p_{\pi(X)}(y_2) d[X|\pi(X) = y_1, X|\pi(X) = y_2] \leq \varepsilon. \tag{2.73}$$

Por (2.70) , (2.32) y la serie de Taylor de la exponencial, vemos que existe algún $y_0 \in G/H$ tal que

$$\mathbf{P}(\pi(X) = y_0) \geq 1 - O\big(\varepsilon^{1/2} \log \frac{1}{\varepsilon}\big).$$

Luego de trasladar X de ser necesario, podemos suponer sin pérdida de generalidad que $y_0 = 0$, es decir,

$$p_{\pi(X)}(0) = \mathbf{P}(X \in H) \geq 1 - O\bigg(\varepsilon^{1/2} \log \frac{1}{\varepsilon}\bigg) \geq \max\bigg(\frac{10}{11}, 1 - \delta_0\bigg), \tag{2.74}$$

[1] La Proposición 4.3 estudia en general la relación de la distancia entrópica de Ruzsa con los morfismos de grupos, y será una clave fundamental también de la demostración del caso general. La demostración de este resultado, si bien aparece más adelante en el texto, es completamente independiente a las cosas aquí desarrolladas y podría ir a leerse directamente ahora.

donde δ_0 es la constante del Lema 2.16, asumiendo que ε_0 (del enunciado de la Proposición 2.14) es lo suficientemente pequeño.

Aplicando el Lema 2.16 a $\pi(X)$ utilizando (2.72) y (2.74), concluimos que

$$\mathbf{H}(\pi(X)) \leq 2\varepsilon. \tag{2.75}$$

Por otro lado, descartando todos los términos en la suma sobre y_1 en (2.73) excepto el término con $y_1 = 0$, y usando (2.74), se deduce que

$$\sum_{y \in G/H} p_{\pi(X)}(y) d[X|\pi(X) = 0, X|\pi(X) = y] \leq 1.1\varepsilon.$$

Por (2.16), esto implica que

$$\sum_{y \in G/H} p_{\pi(X)}(y) \left|\mathbf{H}(X|\pi(X) = y) - \mathbf{H}(X|\pi(X) = 0)\right| \leq 2.2\varepsilon,$$

y, por la desigualdad triangular,

$$\left|\mathbf{H}(X \mid \pi(X)) - \mathbf{H}(X|\pi(X) = 0)\right| \leq 2.2\varepsilon.$$

Utilizando que

$$\mathbf{H}(X) = \mathbf{H}(X \mid \pi(X)) + \mathbf{H}(\pi(X))$$

y (2.75), concluimos que

$$\left|\mathbf{H}(X) - \mathbf{H}(X|\pi(X) = 0)\right| \leq 4.2\varepsilon. \tag{2.76}$$

En particular, a partir de (2.71) deducimos

$$\mathbf{H}(X|\pi(X) = 0) \geq \log|H| - O\left(\varepsilon^{1/2} \log \frac{1}{\varepsilon}\right). \tag{2.77}$$

Ahora, descartando todos los términos en (2.73) excepto aquel con $y_1 = y_2 = 0$, y usando (2.74), tenemos

$$d[X|\pi(X) = 0, X|\pi(X) = 0] \leq 1.21\varepsilon.$$

Se deduce del Lema 2.18 que $\mathbf{H}(X|\pi(X) = 0) \geq \log|H| - 2.42\varepsilon$, y por tanto, a partir de (2.76) obtenemos

$$\mathbf{H}(X) \geq \log|H| - 6.62\varepsilon.$$

Combinando esto con (2.69) y (2.75) se concluye que

$$d[X; U_H] \leq 5.31\varepsilon \leq 6\varepsilon,$$

lo cual es la afirmación de la Proposición 2.14 (con una constante mejorada) en el caso simétrico $X = Y$.

Finalmente, deducimos el caso general en el que X e Y pueden ser diferentes. Supongamos ahora que

$$d[X;Y] = \varepsilon \leq \varepsilon_0',$$

donde $\varepsilon_0' := \varepsilon_0/2$, siendo ε_0 la constante anterior. Por la desigualdad triangular,

$$d[X;X] \leq 2\varepsilon \leq \varepsilon_0,$$

y, por lo tanto, en el caso simétrico de la Proposición 2.14 establecido anteriormente, tenemos que

$$d[X;U_H] \leq 12\varepsilon$$

para algún subgrupo $H \leq G$. De manera similar, se tiene que

$$d[Y;U_{H'}] \leq 12\varepsilon$$

para algún subgrupo $H' \leq G$.

Resta demostrar que $H = H'$. Para ello, observamos que por la desigualdad triangular se tiene

$$d[U_H;U_{H'}] \leq 25\varepsilon. \tag{2.78}$$

Si $H \neq H'$, entonces $H + H'$ es un subgrupo de G que contiene propiamente a H y H' y, por lo tanto, tiene tamaño al menos $2\max(|H|,|H'|)$. Dado que $U_H - U_{H'}$ es uniforme en $H + H'$ cuando estas variables uniformes son independientes, se tiene que

$$d[U_H;U_{H'}] \geq \log 2,$$

lo cual contradice (2.78) si ε_0 es lo suficientemente pequeño. Por lo tanto, efectivamente, $H = H'$, y esto concluye la demostración. □

Obtuvimos en particular el caso del 100% de la Conjetura de Marton en su versión entrópica, que es la que utilizaremos en el Capítulo 4 para demostrar la Conjetura de Marton en $\mathbf{F}_2$:

Corolario 2.20. *Supongamos que X_1, X_2 son dos variables aleatorias G-valuadas tales que $d[X_1;X_2] = 0$. Existe entonces un subgrupo $H \leq G$ tal que $d[X_1;U_H] = d[X_2;U_H] = 0$.*

2.4 Teorema de Balog-Szemerédi-Gowers

Ya habiendo probado el caso inicial de la Conjetura de Marton, en esta sección vamos a cambiar de rumbo y clarificar un poco qué es el teorema de Balog-Szemerédi-Gowers original, del cual ya dijimos que el Lema 2.7 es una versión entrópica. Este teorema es sumamente relevante en Combinatoria Aditiva (es, por ejemplo, utilizado por Tim Gowers en su célebre artículo "A new proof of Szemerédi's theorem" [6]), pero toma particular importancia en nuestra tesis ya que es un paso clave para demostrar la versión general de la Conjetura de Marton, cosa que veremos en el Capítulo 4.

Para enunciar este teorema, primero definimos los **conjuntos de suma parcial**

$$A \overset{E}{+} B := \{a + b : (a, b) \in E\}$$

con E un subconjunto cualquiera de $A \times B$.

Teorema 2.21 (Teorema de Balog-Szemerédi-Gowers)**.** *Supongamos que A, B son subconjuntos finitos y no vacíos de un grupo aditivo G, y sea $E \subset A \times B$ tal que $|E| \geq |A||B|/K$ y $|A \overset{E}{+} B| \leq K|A|^{1/2}|B|^{1/2}$ para algún $K \geq 1$. Entonces existen subconjuntos $A' \subset A, B' \subset B$ con $|A'| \gg |A|/K$, $|B'| \gg |B|/K$ tales que $|A' + B'| \ll K^7|A'|^{1/2}|B|^{1/2}$.*

En pocas palabras, podriamos entender que este resultado nos dice que, si tenemos dos conjuntos A y B tales que su suma parcial con respecto a otro conjunto E es pequeña, entonces podemos refinar estos conjuntos a unos subconjuntos A' y B' de tamaño considerable que cumplan que su suma total es también pequeña.

Para una demostración y un análisis más profundo de este resultado, referimos al lector a [35, Capítulo 6.4].

La idea análoga de una suma parcial $A \overset{E}{+} B$ en el contexto entrópico es una suma $X + Y$ de variables aleatorias **no independientes** X, Y. Por ejemplo, si $E \subset A \times B$ es un conjunto no vacío, y (X, Y) es una variable aleatoria uniforme en E, entonces $X + Y$ es una variable aleatoria que toma valores en $A \overset{E}{+} B$. Por otro lado, la noción de un refinamiento A' de un conjunto A con tamaño comparable al de A es un condicionamiento de la variable aleatoria X, con respecto a otra variable aleatoria Z, de forma que $\mathbf{H}(X|Z)$ sea todavía comparable con $\mathbf{H}(X)$. El Teorema entrópico de Balog-Szemerédi-Gowers nos dice entonces que si dos variables aleatorias **débilmente dependientes** X, Y tienen una suma de poca entropía, podemos fabricarnos condicionamientos de X, Y (que capturen casi toda la entropía) cuya suma **independiente** tenga una entropía pequeña.

En concreto, la primera versión de este resultado entrópico es la siguiente:

Teorema 2.22 (Teorema original de Balog–Szemerédi–Gowers entrópico,[31, Teorema 3.1])**.** *Sea G un grupo aditivo y sean X, Y variables aleatorias con valores en G que son **débilmente dependientes** en el sentido de que*

$$\log K \ \geq \ \mathbf{I}[X : Y]$$

para algún $K \geq 1$. Supongamos además que

$$\mathbf{H}(X + Y) \ \leq \ \frac{1}{2}\mathbf{H}(X) + \frac{1}{2}\mathbf{H}(Y) + \log K.$$

Sea $(X_1, Y), (X_2, Y)$ un par de ensayos condicionalmente independientes de (X, Y) condicionados a Y. A su vez, sea (X_1, X_2, Y) y (X_1, Y') un par de ensayos condicionalmente independientes de (X_1, X_2, Y) y (X_1, Y) condicionados

a X_1. Entonces X_2 y Y' son condicionalmente independientes respecto de X_1, Y, y se cumple

$$\begin{aligned}\mathbf{H}\left(\left(X_2 \,\middle|\, X_1, Y\right)\right) &\geq \mathbf{H}(X) - \log K,\\ \mathbf{H}\left(\left(Y' \,\middle|\, X_1, Y\right)\right) &\geq \mathbf{H}(Y) - \log K,\\ \mathbf{H}\left(\left(X_2 + Y' \,\middle|\, X_1, Y\right)\right) &\leq \frac{1}{2}\mathbf{H}(X) + \frac{1}{2}\mathbf{H}(Y) + 7\log K.\end{aligned}$$

Para una demostración de este Teorema, referimos al lector a [31, Teorema 3.1]. Si uno sigue esta demostración puede notar que el paso fundamental es el siguiente lema:

Lema 2.23 (Lema débil de Balog-Szemerédi-Gowers)**.** *Vale que, en las condiciones de arriba,*

$$\mathbf{H}(X_1 - X_2|Y) \leq \mathbf{H}(X) + 4\log K.$$

Y justamente esta es la versión que tiene el mismo espiritu del Lema 2.7, solo que este último tiene mejores cotas.

Para ser más precisos, si uno observa la demostración del Lema 2.7, podemos notar que vale que

$$\sum_z p_Z(z) d[(A'|Z=z);(B'|Z=z)] \leq 3\mathbf{I}[A:B] + 2\mathbf{H}(Z) - \mathbf{H}(A) - \mathbf{H}(B)$$

donde A', B' son ensayos condicionalmente independientes de A, B con respecto a $Z = A + B$. En particular, debe existir $z \in Im(Z)$ tal que

$$d[(A'|Z=z);(B'|Z=z)] \leq 3\mathbf{I}[A:B] + 2\mathbf{H}(Z) - \mathbf{H}(A) - \mathbf{H}(B),$$

cosa de que si uno venía asumiendo que

$$\mathbf{I}[A:B] \leq \log(K) \text{ y } \mathbf{H}(Z) - \frac{1}{2}\mathbf{H}(A) - \frac{1}{2}\mathbf{H}(B) \leq \log(K)$$

implica que

$$d[(A'|Z=z);(B'|Z=z)] \leq 5\log(K),$$

pero ahora con el beneficio de que $(A'|Z=z)$ y $(B'|Z=z)$ son independientes (notar que la distancia entrópica entre estas dos variables está midiendo el tamaño de la suma de las variables aleatorias).

2.5 Resultados para el caso de m-torsión.

Esta última sección está enteramente dedicada a algunos resultados que se necesitarán para la Conjetura de Marton en el caso de m-torsión, por lo que recomendamos al lector venir simplemente a chequear los resultados de ser necesarios mientras uno lea la demostración de este caso. Por esto mismo, vamos a asumir sabidas cosas que se dicen en el capítulo del caso general, como por ejemplo la definición de multidistancia.

Proposición 2.24. *Sea G un grupo abeliano finito. Supongamos que I es un conjunto de índices de tamaño $m \geq 2$. Supongamos que X_I es un conjunto de variables aleatorias con valores en G tal que $\mathrm{D}[X_I] < c_0$ para una constante absoluta $c_0 > 0$ suficientemente pequeña. Entonces, existe un subgrupo $H \leq G$ tal que*

$$\sum_{i \in I} d[X_i; U_H] \ll m \, \mathrm{D}[X_I].$$

Además, si todas las X_i toman valores en un conjunto simétrico $S \subseteq G$ que contiene el origen, entonces podemos tomar $H \subseteq 6S$.

Demostración. Definimos $\varepsilon := \mathrm{D}[X_I]$. Por el Lema 5.9(i) y pensando en el promedio, podemos encontrar un $i \in I$ tal que

$$\sum_{k \neq i} d[X_i; -X_k] \leq (m-1)\varepsilon. \tag{2.79}$$

Aplicando la idea del promedio nuevamente, podemos hallar $j \neq i$ tal que

$$d[X_i; -X_j] \leq \varepsilon.$$

Si ε es suficientemente chico, podemos aplicar la Proposición 2.14 para concluir que existe un subgrupo finito H de G tal que

$$d[X_i; U_H] \leq 12\varepsilon.$$

Mirando bien la demostración de este resultado podemos observar que H es de la forma $H = S' - S'$, donde todos los elementos y de S' tienen divergencias de Kullback–Leibler

$$D_{\mathrm{KL}}(y - X_j || X_i - X_j)$$

finita; esto implica que $S' \subseteq 3S$ y, por lo tanto, que $H \subseteq 6S$.

Dado que $d[-Y; U_H] = d[Y; -U_H] = d[Y; U_H]$ para cualquier variable aleatoria Y, la ecuación (2.79) y la desigualdad triangular nos dan

$$\sum_{k \in I} d[X_k; U_H] \leq (13m - 1)\varepsilon,$$

como queríamos. □

Lema 2.25. *Sea G un grupo abeliano, supongamos que I es un conjunto de índices finito, $|I| \geq 2$, y que $(X_i)_{i \in I}$ son variables aleatorias G-valuadas conjuntamente independientes.*

(i) *Para cualquier variable aleatoria Y, y cualquier $i_0 \in I$, vale que*

$$\mathrm{d}\Big[Y; \textstyle\sum_{i \in I} X_i\Big] \leq d[Y; X_{i_0}] + \frac{1}{2}\Big(\mathbf{H}\Big[\textstyle\sum_{i \in I} X_i\Big] - \mathbf{H}[X_{i_0}]\Big).$$

(ii) *Para cualquier conjunto de índices finito J, cualquier conjunto de variables aleatorias conjuntamente independientes $(Y_j)_{j\in J}$ y conjuntamente independientes de $(X_i)_{i\in I}$, y cualquier función $f\colon J\to I$, tenemos que*

$$\mathbf{H}\Big[\sum_{j\in J} Y_j\Big] \le \mathbf{H}\Big[\sum_{i\in I} X_i\Big] + \sum_{j\in J}\big(\mathbf{H}[Y_j - X_{f(j)}] - \mathbf{H}[X_{f(j)}]\big).$$

(iii) *Si escribimos $W = \sum_{i\in I} X_i$ entonces $d[W;-W] \le 2\,\mathrm{D}[X_I]$.*

Demostración. Podemos asumir sin pérdida de generalidad que Y, $(Y_j)_{j\in J}$ y X_I son todas independientes entre ellas. Aplicando la Proposición 2.6 con $n=2$, $X = X_{i_0}$, $Y_1 = -Y$ y $Y_2 = \sum_{i\neq i_0} X_i$ obtenemos

$$\mathbf{H}\Big[-Y + \sum_{i\in I} X_i\Big] \le \mathbf{H}[X_{i_0} - Y] + \mathbf{H}\Big[\sum_{i\in I} X_i\Big] - \mathbf{H}[X_{i_0}]. \tag{2.80}$$

Usando la definición de la distancia de Ruzsa concluimos (i).

Para probar (ii), escribimos $W := \sum_{i\in I} X_i$. Luego

$$\begin{aligned}\mathbf{H}\Big[\sum_{j\in J} Y_j\Big] &\le \mathbf{H}\Big[-W + \sum_{j\in J} Y_j\Big]\\ &\le \mathbf{H}[W] + \sum_{j\in J}\big(\mathbf{H}[Y_j - W] - \mathbf{H}[W]\big)\\ &\le \mathbf{H}[W] + \sum_{j\in J}\big(\mathbf{H}[Y_j - X_{f(j)}] - \mathbf{H}[X_{f(j)}]\big)\end{aligned}$$

donde usamos (2.19) en el anteúltimo paso y (2.80) en el último.

Para (iii), tomemos $(X_i')_{i\in I}$ como copias independientes de $(X_i)_{i\in I}$ y escribamos $W' = \sum_{i\in I} X_i'$. Fijemos tambíen $a, b \in I$.

Aplicando la Proposición 2.6 con $n=2$, $X = X_a$, $Y_1 = \sum_{i\neq a} X_i$ y $Y_2 = W'$, obtenemos

$$\mathbf{H}[W + W'] \le \mathbf{H}[W] + \mathbf{H}[X_a + W'] - \mathbf{H}[X_a]. \tag{2.81}$$

Aplicando la Proposición 2.6 con $n=2$, $X = X_b'$, $Y_1 = X_a$ y $Y_2 = \sum_{i\neq b} X_i'$ conseguimos

$$\mathbf{H}[X_a + W'] \le \mathbf{H}[X_a + X_b] + \mathbf{H}[W'] - \mathbf{H}[X_b'].$$

Combinando ahora esto con (2.81) y luego aplicando la Proposición 2.6, (2.14) nos da que

$$\begin{aligned}\mathbf{H}[W + W'] &\le 2\mathbf{H}[W] + \mathbf{H}[X_a + X_b] - \mathbf{H}[X_a] - \mathbf{H}[X_b]\\ &\le 3\mathbf{H}[W] - \mathbf{H}[X_a] - \mathbf{H}[X_b].\end{aligned}$$

Promediando esto a través de todas las combinaciones de (a,b) da que $\mathbf{H}[W] + 2\,\mathrm{D}[X_I]$, y reordenando se obtiene (iii). □

Lema 2.26. *Sea G un grupo abeliano, sean X, Y variables aleatorias independientes G-valuadas, y sea $a \in \mathbf{Z}$. Tenemos entonces las siguientes dos desigualdades:*

(i) $\mathbf{H}[X - aY] - \mathbf{H}[X] \leq 4|a|d[X;Y]$.

(ii) $\mathbf{H}[X - aY] - \mathbf{H}[X] \leq (4 + 10\lfloor \log_2 |a| \rfloor) d[X;Y]$.

Demostración. Sea a un entero cualquiera, y sea X' otra copia independiente de X. Vamos entonces a probar las dos desigualdades:

$$\mathbf{H}[X - (a \pm 1)Y] \leq \mathbf{H}[X - aY] + \big(\mathbf{H}[X - Y - X'] - \mathbf{H}[X]\big) \tag{2.82}$$

y

$$\mathbf{H}[X - 2aY] \leq \mathbf{H}[X - aY] + \big(\mathbf{H}[X - 2X'] - \mathbf{H}[X]\big). \tag{2.83}$$

Primero veamos que nos basta con eso. Por (2.19) con $n = 2$ tenemos

$$\mathbf{H}[Y - X + X'] \leq \mathbf{H}[Y - X] + \mathbf{H}[Y + X'] - \mathbf{H}[Y],$$

que puede reordenarse para darnos

$$\mathbf{H}[X - Y - X'] - \mathbf{H}[X] \leq d[X;Y] + d[X;-Y].$$

Por el Lema 2.8, esto es como mucho $4d[X;Y]$, y luego (2.82) nos da

$$\mathbf{H}[X - (a \pm 1)Y] \leq \mathbf{H}[X - aY] + 4d[X;Y]. \tag{2.84}$$

Haciendo una inducción directa en a obtenemos (i).

Ahora sea X'' otra copia independiente de X. Por el caso $a = 1$ de (2.82) aplicado a X, X' obtenemos

$$\mathbf{H}[X - 2X'] \leq \mathbf{H}[X - X'] + \mathbf{H}[X - X' - X''] - \mathbf{H}[X],$$

y otra aplicación de (2.19) con $n = 2$ nos da

$$\mathbf{H}[X - X' - X''] \leq \mathbf{H}[X - X'] + \mathbf{H}[X - X''] - \mathbf{H}[X] = \mathbf{H}[X] + 2d[X;X].$$

Luego,

$$\mathbf{H}[X - 2X'] \leq \mathbf{H}[X] + 3d[X;X] \leq \mathbf{H}[X] + 6d[X;Y]$$

donde estamos usando la desigualdad triangular en la segunda cota. Así, (2.83) aplicado a X y Y implica

$$\mathbf{H}[X - 2aY] \leq \mathbf{H}[X - aY] + 6d[X;Y]. \tag{2.85}$$

Las desigualdades (2.84) y (2.85) implican que podemos acotar de forma recursiva

$$\mathbf{H}[X - aY] - \mathbf{H}[X] \leq f(a)d[X;Y],$$

donde $f\colon \mathbf{Z} \to \mathbf{Z}_{\geq 0}$ es la función más pequeña que obedece que $f(0) = 0$, $f(a \pm 1) \leq f(a) + 4$ y $f(2a) \leq f(a) + 6$. En efecto, podemos insertar un nuevo dígito binario al inicio de a a costa de aumentar $f(a)$ por como mucho 10, lo que nos da la cota $f(a) \leq 4 + 10\lfloor \log_2 |a| \rfloor$, como era requerido para (ii).

Nos queda ahora el trabajo de establecer (2.82) y (2.83). Para la primera, escribimos $X-(a+1)Y$ como $(X-Y)-aY$ y aplicamos la desigualdad (2.17), de la Proposición 2.6, con las variables (X, Y, Z) reemplazadas por $(X-Y, X', aY)$ para obtener

$$\mathbf{H}[X-Y-aY] \leq \mathbf{H}[X-Y-X'] + \mathbf{H}[X'-aY] - \mathbf{H}[X'].$$

Renombrando las cosas, esto es exactamente (2.82) en el caso $a+1$. Para el caso $a-1$, simplemente reemplazamos todos los términos $X-Y$ con $X+Y$, notando que $\mathbf{H}[X-X'+Y] = \mathbf{H}[X-X'-Y]$.

Para (2.83) aplicamos la desigualdad triangular para obtener

$$\mathbf{H}[X-2aY] \leq \mathbf{H}[X-2X'] + \mathbf{H}[2X'-2aY] - \mathbf{H}[2X'].$$

Nos basta entonces con probar que

$$\mathbf{H}[2X'-2aY] - \mathbf{H}[2X'] \leq \mathbf{H}[X'-aY] - \mathbf{H}[X']. \tag{2.86}$$

Para ver esto, observamos que la desigualdad de submodularidad nos da que

$$\begin{aligned}\mathbf{H}[X'-aY \,|\, 2X'-2aY] &\geq \mathbf{H}[X'-aY \,|\, 2X'-2aY, aY] \\ &= \mathbf{H}[X' \,|\, 2X'-2aY, aY] \\ &= \mathbf{H}[X' \,|\, 2X', aY] = \mathbf{H}[X' \,|\, 2X'],\end{aligned}$$

y esto se reordena para obtener la desigualdad deseada (2.86). □

Capítulo 3

Algunas equivalencias

El objetivo de esta sección es mostrar las distintas formas en las cuales podemos enunciar la Conjetura de Marton. Lo primero que haremos será probar la equivalencia entre la conjetura en su forma clásica y su versión entrópica, lo que es fundamental porque esta segunda es la que demostraremos en esta tesis. Luego, en la segunda sección, probaremos la Proposición 1.4 de la Introducción, mediante ideas de combinatoria aditiva, sin hacer mención a la entropía. La primer sección está basada en el survey de Green "Finite field models in additive combinatorics" [11], mientras que la segunda es esencialmente la equivalencia probada en "Sumsets and entropy revisited" [16].

Dejaremos aquí sentado un lema de cubrimientos, atribuído a Ruzsa, que usaremos en este capítulo:

Lema 3.1. *Sean S y T dos subconjuntos de un grupo abeliano tales que $|S+T| \leq K'|S|$. Existe entonces un conjunto $X \subseteq T$, $|X| \leq K'$, tal que $T \subseteq S - S + X$.*

Demostración. Tomemos un conjunto maximal $X \subseteq T$ tal que los conjuntos $S+x$, $x \in T$, son disjuntos dos a dos. Como $\bigcup_{x\in X} S+x \subseteq S+T$, debemos tener que $|S||X| \leq K'|S|$, lo que nos dice que $|X| \leq K'$. Tomemos ahora $t \in T$. Por maximalidad, debe haber algún $x \in X$ tal que $(S+t)\cap(S+x) \neq \emptyset$, lo que significa que $t \in S - S + X$. □

También enunciamos una desigualdad del estilo de las **desigualdades de Plünnecke**:

Proposición 3.2. *Supongamos que A y B son subconjuntos de un grupo abeliano G, tales que $|A+B| \geq K|A|$. Luego, para dos enteros no negatios cualesquiera k, l, vale que $|kA - lB| \leq K^{k+l}|A|$.*

No daremos una prueba de esto, pero referimos al lector interesado a [35, Corolario 6.29].

Por último, enunciamos una versión del Teorema de Balog-Szemerédi-Gowers similar al Teorema 2.21:

Lema 3.3. *Sea A un subconjunto de un grupo abeliano. Supongamos que $|A| = n$ y que hay al menos cn^3 tuplas $(a_1, a_2, a_3, a_4) \in A^4$ tales que $a_1 + a_2 = a_3 + a_4$. Existe entonces un conjunto $A' \subseteq A$ tal que $|A'| \geq 2^{-19}c^{12}n$ y $|A' - A'| \leq 2^{57}c^{-36}|A'|$.*

Una referencia de esto es [6, Capítulo 6, Proposición 12].

3.1 PFR aditiva vs. PFR entrópica.

Vamos entonces a probar la equivalencia de la PFR entre su versión entrópica y su versión aditivo combinatoria. Lo haremos para el caso de $\mathbf{F}_2^n$, ya que, entre otras cosas, para el caso de torsión impar de la Conjetura de Marton necesitaremos un resultado levemente más fuerte (en el caso de torsión impar nos interesará controlar la ubicación de ciertos conjuntos, para dar un resultado del estilo de la Conjetura de Bogolyuvob), y esta demostración no nos alcanza. Sin embargo, daremos la versión pertinente en el Capítulo 5.

Proposición 3.4. *Las siguientes dos afirmaciones son equivalentes:*

(1) Si $A \subseteq \mathbf{F}_2^n$ y si $\sigma(A) \leq K$, entonces A está cubierto por $O(K^{O(1)})$ coclases de algún subespacio $H \leq \mathbf{F}_2^n$ de tamaño a lo sumo $|A|$.

(2) Si X, Y son dos variables aleatorias con valores en $\mathbf{F}_2^n$, entonces existe algún subgrupo $H \leq \mathbf{F}_2^n$ tal que $d[X, U_H], d[Y, U_H] \ll d[X, Y]$.

Demostración. Primero derivamos la afirmación entrópica a partir de la afirmación combinatoria.

Definamos $k := d[X, Y]$ y fijemos $K := e^k$. Podemos suponer que $k \geq \epsilon_0$, donde ϵ_0 es la constante en la Proposición 2.14, ya que en caso contrario la afirmación se sigue inmediatamente de dicha proposición.

Aplicando la Proposición 2.12 con $C = 4$, obtenemos un conjunto $S \subseteq \mathbf{F}_2^n$ con

$$d[X, U_S] \ll k \tag{3.1}$$

y, utilizando que podemos reemplazar sumas con restas, obtenemos que

$$|S + S| \ll K^{O(1)}|S|. \tag{3.2}$$

Por (1), existe un subgrupo $H \leq \mathbf{F}_2^n$, con $|H| \leq |S|$, tal que S está cubierto por $O(K^{O(1)})$ coclases de H. En particular, $S + H$ está contenido en la unión de estas coclases, por lo que $|S + H| \ll K^{O(1)} \min(|S|, |H|)$.

Ahora, para cualquier par de conjuntos A, B, tenemos

$$\begin{aligned} d[U_A, U_B] &= \mathbf{H}(U_A - U_B) - \frac{1}{2}(\mathbf{H}(U_A) + \mathbf{H}(U_B)) \\ &\leq \log|A - B| - \frac{1}{2}\big(\log|A| + \log|B|\big) \\ &= \log\left(\frac{|A - B|}{|A|^{1/2}|B|^{1/2}}\right). \end{aligned}$$

(Esta es la versión bipartita de la Proposición 2.11.)

Aplicando esto con $A = S$ y $B = H$ (y observando que $H = -H$), obtenemos $d[U_S, U_H] \ll k$. Por la desigualdad triangular y (3.1), se sigue que $d[X, U_H] \ll k$, lo que concluye la demostración (2).

Pasamos ahora a la implicación inversa, derivando la afirmación combinatoria a partir de la afirmación entrópica.

Supongamos que $A \subseteq \mathbf{F}_2^n$ es un conjunto y definamos $K := \sigma(A)$ y $k := \log K$. Entonces, por (2.11), tenemos que $d[U_A, U_{-A}] \leq \log \sigma(A) = k$.

Asumiendo la segunda afirmación, existe algún subgrupo finito $H \leq \mathbf{F}_2^n$ tal que $d[U_A, U_H] \ll k$. Aplicando (2.16) y usando el hecho de que $\mathbf{H}(U_A) = \log |A|$ y $\mathbf{H}(U_H) = \log |H|$, obtenemos

$$K^{-O(1)}|A| \ll |H| \ll K^{O(1)}|A|. \tag{3.3}$$

Denotemos por $p(x)$ la función de densidad de $U_A - U_H$, de modo que $p(x) = \frac{|A \cap (H+x)|}{|A||H|}$. A partir de (2.32), se sigue que existe algún x_0 tal que

$$\begin{aligned} p(x_0) &\geq e^{-\mathbf{H}(U_A - U_H)} \\ &= e^{-d[U_A, U_H]}|A|^{-1/2}|H|^{-1/2} \\ &= K^{-O(1)}|A|^{-1}, K^{-O(1)}|H|^{-1}, \end{aligned}$$

es decir, que $|A \cap (H + x_0)| = K^{-O(1)}|H|, K^{-O(1)}|A|$.

El lema de cubrimiento de Ruzsa, Lema 3.1, aplicado con $S = A \cap (H + x_0)$ y $T = A$, y usando el hecho de que $S + T \subseteq A + A$ y $S - S \subseteq H$, obtenemos que A está cubierto por $O(K^{O(1)})$ traslaciones de H. En efecto, esto puede confirmarse observando que

$$|S + T| \leq |A + A| \leq K|A| \leq K^{O(1)}|S|.$$

Si $|H| \leq |A|$, hemos terminado. Si $|H| > |A|$, pasamos a un subgrupo $H' \leq H$ de tamaño en el rango $\left(\frac{1}{2}|A|, |A|\right]$; entonces A está cubierto por $O(K^{O(1)})$ traslaciones de H', y la demostración se completa en este caso también. En efecto, notar que, como $|H| \leq CK^{O(1)}|A|$ para alguna constante C, podemos ir achicando a este subespacio H sacando de a un generador a la vez, en cada paso dividiendo el tamaño de H en dos, necesitando como mucho $\log_2(CK^{O(1)})$ pasos, por lo que nuestras estimaciones polinomiales no se ven afectadas, ya que podremos cubrir a A con como mucho $\log_2(CK^{O(1)})O(K^{O(1)}) = O(K^{O(1)})$ coclases de nuestro subespacio final. □

Observación 3.5. *Si uno persigue las constantes, puede probarse que si la segunda afirmación es correcta con la constante C, es decir $d[X, U_H], d[Y, U_H] \leq Cd[X, Y]$, entonces en la primera afirmación vale la constante $C + 1$, es decir, que A está cubierto por $O(K^{C+1})$ coclases. Esto es necesario para obtener exactamente la versión de la Conjetura de Marton de la introducción, ya que hablaba de cubrir al conjunto con $2K^{12}$ coclases de un subgrupo, y lo que vamos a probar es la segunda afirmación pero con la constante $C = 11$. Referimos al lector interesado a los apéndices de [8].*

3.2 Unas cuantas versiones más.

Probaremos ahora la Proposición 1.4, enunciada en la Introducción, que nos daba múltiples caracterizaciones equivalentes de la PFR en $\mathbf{F}_2$. La demostración que proveemos está sacada de el survey de Ben Green "Finite field models in additive combinatorics" [11], y es un lindo ejemplo de como interactúan distintos resultados e ideas típicas de Combinatoria Aditiva. Es común en el área encontrar resultados de este estilo, mostrando que esencialmente tenemos una "equivalencia polinomial" entre distintas nociones de grupo aproximado, cosa que ya habíamos anticipado en la Introducción, y que habíamos hecho explícito con la Proposición 1.18. Por comodidad del lector, enunciamos la proposición una vez más:

Proposición 3.6 ([10, Proposición 2.2])**.** *Las siguientes afirmaciones son equivalentes.*

1. *Si $A \subseteq \mathbf{F}_2^\infty$ cumple que $|A+A| \leq K|A|$, entonces existe $A' \subseteq A$, $|A'| \geq |A|/C_1(K)$, contenido en una coclase de algun subespacio de tamaño como mucho $C_2(K)|A|$.*

2. *Si $A \subseteq \mathbf{F}_2^\infty$ cumple que $|A+A| \leq K|A|$, entonces A puede cubrirse por como mucho $C_3(K)$ coclases de algun subespacio de tamaño como mucho $C_4(K)|A|$.*

3. *Si $A \subseteq \mathbf{F}_2^\infty$ cumple que $|A+A| \leq K|A|$, y adicionalmente existe un conjunto B, $|B| \leq K$, tal que $A+B = A+A$, entonces A puede cubrirse por, como mucho, $C_5(K)$ coclases de algun subespacio de tamaño como mucho $C_6(K)|A|$.*

4. *Supongamos que $f : \mathbf{F}_2^m \to \mathbf{F}_2^\infty$ es una función con la propiedad de que*

$$|\{f(x)+f(y)-f(x+y) : x,y \in \mathbf{F}_2^m\}| \leq K.$$

   Luego f puede escribirse como $g+h$, donde g es lineal, e $|Im(h)| \leq C_7(K)$.

5. *Supongamos que $f : \mathbf{F}_2^m \to \mathbf{F}_2^\infty$ es una función con la propiedad de que al menos $2^{3m}/K$ de las tuplas $(x_1,x_2,x_3,x_4) \in \mathbf{F}_2^m$ con $x_1+x_2 = x_3+x_4$ vale que $f(x_1)+f(x_2) = f(x_3)+f(x_4)$. Luego, existe una función lineal afín $g : \mathbf{F}_2^m \to \mathbf{F}_2^\infty$ tal que $f(x) = g(x)$ para al menos $2^m/C_8(K)$ valores de x.*

Mas aún, si $C_i(K)$ esta acotada por un polinomio en K para todo $i \in I$, donde I es alguno de los conjuntos $\{1,2\},\{3,4\},\{5,6\},\{7\},\{8\}$, en realidad $C_i(K)$ es acotada por un polinomio de K para todo i.

Demostración. (1) $\iff$ (2).

Es fácil ver que (2) $\implies$ (1), por lo que demostraremos la otra dirección. Supongamos que $A \subseteq \mathbf{F}_2^\infty$ cumple que $|A+A| \leq K|A|$. Usando (1), tomamos $A' \subseteq A$ un subconjunto con $|A'| \geq |A|/C_1(K)$ y tal que $|A'|$ está incluido en la coclase de un subespacio de tamaño como mucho $C_2(K)|A|$. Usando el lema

de cubrimiento de Rusza, Lema 3.1, y tomando $S = A'$, $T = A$, obtenemos un $X \subseteq T$ tal que $|X| \geq KC_1(K)$ y $A \subseteq A' - A' + X$. De esto se deduce inmediatamente (2), con $C_3(K) \geq KC_1(K)$ y $C_4(K) = C_2(K)$.

(2) $\iff$ (3).

Es evidente que (2) $\implies$ (3). Para demostrar la otra dirección, vamos a aplicar (3) al conjunto $D = A - A$. Por la desigualdad de Plünnecke, Proposición 3.2, tenemos que

$$|D + D| = |2A - 2A| \leq K^4 |A| \leq K^4 |D| .$$

Afirmamos que existe un conjunto B, $|B| \leq K^8$, tal que $D + B = D + D$. Para ver esto, aplicamos otra vez el lema de cubrimiento con $S = A$, $T = 2A - A$ y $K' = K^4$ (se puede chequear que esto cumple las hipótesis usando la desigualdad de Plünnecke una vez más). De esta forma, obtenemos un conjunto X, $|X| \leq K^4$, tal que $2A - A \subseteq X + (A - A)$, lo que implica que

$$2A - 2A \subseteq X + (A - 2A) \subseteq X - X + (A - A).$$

Esto muestra que la afirmación es correcta, tomando $B = X - X$. Ahora podemos aplicar (3) para obtener que D, y por ende A, puede ser cubierto por como mucho $C_5(K^8)$ coclases de algun subespacio de tamaño como mucho $C_6(K^8) |D| \leq K^2 C_6(K^8) |A|$.

(4) $\implies$ (3).

Supongamos que tenemos un conjunto $A \subseteq \mathbf{F}_2^\infty$ con $|A + A| \leq K |A|$, junto a un conjunto B, $|B| \leq K$, tal que $A + A = A + B$, y asumimos también que $0 \in A$, cosa que podemos hacer sin perdida de generalidad. Sea ahora H_0 un subespacio minimal con la propiedad de que existe una función $\pi : A \to H_0$ lineal e inyectiva. Luego,

$$\pi(A + A) = \pi(A - A) = H_0.$$

En efecto, notemos que en $\mathbf{F}_2^n$ la suma es lo mismo que la resta, y asumamos que existe $v \in H_0$ tal que $v \notin \pi(A + A) = \pi(A - A)$. Podemos considerar entonces el espacio $H_0/\langle v \rangle$ y la función proyección ϕ natural a este cociente, que componiendo con π nos da la proyección $\psi \coloneqq \phi \circ \pi$. Resulta que ψ también es una proyección inyectiva en A, ya que $\psi(a) - \psi(a') = 0$ si y sólo si $\psi(a - a') = 0$, cosa que ocurre sólo si $\pi(a - a') \in \langle v \rangle$, ya que π era inyectiva. Esto nos dice que $v \in \pi(A - A) = \pi(A + A)$, lo que nos dice que ψ es inyectiva. El problema con esto es que la dimensión de $H_0/\langle v \rangle$ es estrictamente menor que la de H_0, lo que contradice la minimalidad de H_0.

Vamos a definir ahora una función $f : H_0 \to \mathbf{F}_2^\infty$ de la siguiente forma: fijamos un orden total en B, y para cada $x \in H_0$, de entre todos los pares $(a, b) \in A \times B$ tales que $x = \pi(a + b)$ elegimos el a asociado al b más pequeño. Tomamos $f(x) = a$.

Afirmamos que $|\{f(x) + f(y) - f(x + y) : x, y, \in H_0\}| \leq K^7$. Para ver esto, escribimos $x = \pi(a_1 + b_1), y = \pi(a_2 + b_2)$ y $x + y = \pi(a_3 + b_3)$. Así,

$$f(x) + f(y) - f(x + y) = a_1 + a_2 - a_3. \tag{3.4}$$

Ahora escogemos $a_4 \in A, b_4 \in B$ tales que $a_1 + a_2 = a_4 + b_4$ y también $a_5 \in A, b_5 \in B$ tales que $a_3 + a_4 = a_5 + b_5$. Sumar todo esto resulta en

$$a_1 + a_2 - a_3 = a_5 + b_4 + b_5. \tag{3.5}$$

De aquí vemos, como π es lineal y la característica es 2, que

$$\pi(a_5) = \pi(b_1 + b_2 + b_3 + b_4 + b_5)$$

(para ver esto, recordar que $\pi(a_1 + b_1) + \pi(a_2 + b_2) = \pi(a_3 + b_3)$). Como π era inyectiva en A, de esta igualdad se puede deducir que la cantidad de posibles valores a_5 debe estar acotada por K^5 (al ir variando los b_i). De (3.2) podemos ver que hay como mucho K^7 posibles valores de $a_1 + a_2 - a_3$, lo que, en vista de (3.4), implica la afirmación.

Ahora usamos que habiamos asumido (4) para obtener que $f = g + h$, donde $g : H_0 \to \mathbf{F}_2^\infty$ es lineal y $|Im(h)| \leq C_7(K^7)$. Con esto, si tomamos $H = \pi^{-1}\big(g(H_0)\big)$, se deduce (3) para $C_5 \leq C_7(K^7)$ y $C_6(K) \leq K$.

(1) $\implies$ (5).

Supongamos que $f : \mathbf{F}_2^m \to \mathbf{F}_2^\infty$ es una función con la propiedad del enunciado, es decir, que al menos $2^{3m}/K$ de las tuplas $(x_1, x_2, x_3, x_4) \in \mathbf{F}_2^m$ con $x_1 + x_2 = x_3 + x_4$ también cumplen $f(x_1) + f(x_2) = f(x_3) + f(x_4)$. Consideremos el gráfico $\Gamma = \{(x, f(x)) : x \in \mathbf{F}_2^m\}$ de f, cuyo cardinal es $N = 2^m$, y la cantidad de soluciones a la ecuación $t_1 + t_2 = t_3 + t_4$ con $t_i \in \Gamma$ es al menos N^3/K.

Se deduce del resultado de Balog-Szemerédi-Gowers (3.3) que debe existir un conjunto $\Gamma' \subseteq \Gamma, |\Gamma'| \geq 2^{-19}K^{-12}N$, tal que $|\Gamma' - \Gamma'| \leq 2^{57}K^{36}|\Gamma'|$.

Aplicando (2), obtenemos que Γ' puede cubrirse por $l = C_3(2^{57}K^{36})$ coclases $H + x_1, ..., H + x_l$ de algún subespacio $H \subseteq \mathbf{F}_2^m \times \mathbf{F}_2^\infty$, $|H| \leq C_4(2^{57}K^{36})|\Gamma'| \leq C_4(2^{57}K^{36})N$. Haciendo crecer l a $C_9(K) := C_3(2^{57}K^{36})^2 C_4(2^{57}K^{36})$ de ser necesario, podemos asumir que la proyección π de H en el primer factor $\mathbf{F}_2^m$ es un monomorfismo (lo que estariamos haciendo es achicar H, a costa de aumentar l). Por palomar, debe existir algún i tal que

$$|\Gamma' \cap (H + x_i)| \geq |\Gamma'|/C_9(K) \geq 2^{-19}K^{-12}N/C_9(K).$$

Escribimos ahora $\Gamma'' = \Gamma' \cap (H + x_i)$, y definimos $E = \pi(\Gamma'')$.

Es claro que $f|_E$ es una función lineal afín, y esto termina de confirmar (5), con $C_8(K) = 2^{19}K^{12}C_9(K)$.

(5) $\implies$ (4).

Fijemos $N = 2^n$, y supongamos que $f : \mathbf{F}_2^n \to \mathbf{F}_2^\infty$ es una función con la propiedad de que $|B| \leq K$, donde $B := \{f(x) + f(y) - f(x + y) : x, y \in \mathbf{F}_2^n\}$. La cantidad de tuplas (x_1, x_2, x_3, x_4) tales que $x_1 + x_2 = x_3 + x_4$ y $f(x_1) + f(x_2) = f(x_3) + f(x_4)$ es entonces al menos N^3/K, cosa que puede verificarse con una cuenta corta, utilizando la desigualdad de Cauchy-Schwarz. En efecto, llamemos $t(x, y) := x + y$, $b(x, y) := f(x) + f(y)$, y $g(t, b) := |\{(x, y) : t(x, y) = t \text{ y } b(x, y) = b\}|$, cosa de que

$$\begin{aligned} Q &:= |\{(x_1, x_2, x_3, x_4) : x_1 + x_2 = x_3 + x_4 \text{ y } f(x_1) + f(x_2) = f(x_3) + f(x_4)\}| \\ &= \sum_{t,b} g(t, b)^2. \end{aligned}$$

Como $f(x)+f(y)-f(x+y)=b(x,y)-f(t)$ toma a lo sumo $|B|\leq K$ valores cuando (x,y) recorren el conjunto $\{(x,y):x+y=t\}$, se tiene que

$$|(t,b):g(t,b)\neq 0|\leq KN.$$

La desigualdad de Cauchy-Schwarz nos dice entonces que

$$\begin{aligned}\sum_{t,b} g(t,b)^2 \sum_{t,b} \mathbf{1}_{g(t,b)\neq 0} &\geq \big(\sum_{t,b} g(t,b)\mathbf{1}_{g(t,b)\neq 0}\big)^2 \\ &= \big(\sum_{t,b} g(t,b)\big)^2,\end{aligned}$$

y al pasar dividiendo obtenemos

$$\sum_{t,b} g(t,b)^2 \geq \frac{N^4}{KN}=\frac{N^3}{K},$$

que era justamente lo que afirmamos.

Tenemos luego que existe un conjunto $E\subseteq \mathbf{F}_2^n, |E|\geq N/C_8(K)$, tal que $f|_E$ es una función lineal afín. Escribimos ahora $g:\mathbf{F}_2^n\to\mathbf{F}_2^\infty$ como la extensión de esta función a todo $\mathbf{F}_2^n$.

Ahora el lema de cubrimiento, Lema 3.1, nos dice que existe un conjunto $T, |T|\leq C_8(K)$, tal que $T+E-E=\mathbf{F}_2^n$. Es sencillo confirmar también que

$$\begin{aligned}f(t+e_1-e_2) &= f(t)+f(e_1)-f(e_2)+b_1-b_2 \\ &= g(t+e_1-e_2)-g(t)+g(0)+f(t)+b_1-b_2\end{aligned}$$

para algunos $b_1,b_2\in B$. Luego, $|Im(f-g)|=|T|^2|B|^2\leq C_8(K)^2K^2$, y con esto se termina la demostración. □

Capítulo 4

La Conjetura de Marton en $\mathbf{F}_2$.

> Terry,
>
> I know that your blog has a wide readership and so it would be irresponsible of me to endorse gambling in any way. However I should perhaps mention the informal bet I had with Jean Bourgain, whereby he wins 100 dollars from me if a counterexample to PFR is found and I win the same if a proof of the conjecture is found within 5 years.
>
> —Ben Green, *comentario en el blog de Tao, 2008.*

El objetivo de este capítulo es demostrar el siguiente resultado:

Teorema 4.1. *Sea $G = \mathbf{F}_2^n$, y sean X_1^0, X_2^0 dos variables aleatorias $G-$valuadas. Existe entonces un subgrupo $H \leq G$ tal que*

$$d[X_1^0, U_H] + d[U_H, X_2^0] \leq 11 d[X_1^0, X_2^0]$$

donde U_H denota la distribución uniforme en el conjunto H. Más aún, tanto $d[X_1^0, U_H]$ como $d[U_H, X_2^0]$ son menores que $6d[X_1^0, X_2^0]$.

Por la Proposición 3.4 esto prueba la Conjetura de Marton en $\mathbf{F}_2$.

Este capítulo cubre los resultados esenciales de "On a conjecture of Marton" [8], que es el artículo principal a estudiar en esta tesis. La primera sección recuerda los caminos de demostración discutidos en la Introducción, pero con un poco más de formalidad, estableciendo exactamente como es que vamos a frasear todo para concluir el resultado. Le siguen luego tres secciones, donde se trabajan las tres ideas principales que tiene el paper [8]: establecer la relación entre la distancia entrópica y los morfismos, apoyarse de eso para ver como aplicar las distintas jugadas acota ciertas informaciones mutuas, y utilizar esto junto al teorema de Balog-Szemerédi-Gowers para concluir.

4.1 Plan de la demostración

Antes que nada, el lector debería leer esta sección intentando ver como la ruta que plantearemos en lo siguiente se relaciona con lo discutido en la Introducción, en particular en la Sección 1.2, donde presentamos una versión simplificada de la demostración.

Vamos a fijar por este capítulo a $G = \mathbf{F}_2^n$ y a X_0^1, X_0^2 como en el Teorema 4.1. Como fue comentado antes, la estrategia principal de la demostración va a ser una especie de inducción en la distancia $d[X;Y]$. Para esto, lo primero que haremos será la reducción del Teorema 4.1 a la Proposición 4.2, la cual nos permitirá hacer esta inducción.

Fijemos antes un poco de notación: en el resto de este capítulo, η será la constante $\frac{1}{9}$. Dadas dos variables aleatorias G-valuadas X_1, X_2, introducimos el funcional

$$\tau[X_1;X_2] \coloneqq d[X_1;X_2] + \eta d[X_1;X_0^1] + \eta d[X_2;X_0^2]. \tag{4.1}$$

Cabe destacar que este funcional solo depende de las distribuciones p_{X_1}, p_{X_2}, y no de si las variables son o no independientes entre ellas, o con respecto a las variables X_0^1, X_0^2.

Enunciamos la proposición principal de la cual deduciremos el Teorema 4.1:

Proposición 4.2. *Sean X_1, X_2 dos variables aleatorias $G-$valuadas tales que $d[X_1;X_2] > 0$. Deben existir entonces dos variables aleatorias X_1', X_2' $G-$valuadas también tales que*

$$\tau[X_1';X_2'] < \tau[X_1;X_2] \tag{4.2}$$

Para deducir el Teorema 4.1 de la Proposición 4.2, vamos a usar también el **caso "100%"** de nuestro teorema principal, el Corolario 2.20, de la siguiente forma:

Demostración del Teorema 4.1 asumiendo la Proposición4.2. Escogemos un par de variables X_1, X_2 (en realidad, sus distribuciones en G) que minimicen al operador τ. Notar que este problema de optimización es sobre el producto cartesiano del espacio de distribuciones de probabilidad sobre G. Este espacio con la topología usual es compacto, y claramente $d[-;-]$ es continua, por lo que el mínimo deseado existe.

Como (X_1, X_2) minimizan τ, por el contrarecíproco de la Proposición 4.2, $d[X_1;X_2] = 0$, de lo que el Corolario 4.1 nos garantiza que debe existir un grupo $H \leq G$ tal que $d[X_1;U_H] = d[X_2;U_H] = 0$. Finalmente, obtenemos

$$\begin{aligned}\eta(d[X_1^0;U_H] + d[X_2^0;U_H]) &\leq \eta(d[X_1^0;X_1] + d[X_2^0;X_2]) \\ &= \tau[X_1;X_2] \\ &\leq \tau[X_2^0;X_1^0] = (1+2\eta)d[X_1^0;X_2^0]\end{aligned}$$

Recordando que $\eta = \frac{1}{9}$, se obtiene el primer resultado del teorema. Para deducir lo segundo, observar que $\left|d[X_1^0;U_H] - d[X_2^0;U_H]\right| \leq d[X_1^0;X_2^0]$. □

A lo largo de este capítulo denotaremos con k a la cantidad $k := d[X_1; X_2]$.

Al igual que como hacimos en la Introducción (ver la Sección 1.2) al construir variables relevantes, la idea principal para probar la Proposición 4.2 es construir distintas variables X_1', X_2' que se relacionen a X_1, X_2, y probar que alguna elección funciona, es decir, que alguna de estas parejas fabricadas achican el operador τ, con respecto a lo que vale en X_1 y X_2. Estas distintas variables construidas serían los análogos de los "mejores" conjuntos que armabamos en la Introducción, que estaban a una distancia controlada del conjunto original, pero con mejor constante de duplicación. Recordemos que en el contexto de conjuntos, habíamos sugerido que nuestras jugadas eran esencialmente dos: sumar el conjunto consigo mismo, o quedarse con alguna fibra particular de la suma.

Consideremos entonces la tupla de 4 variables aleatorias $X_1, X_2, \tilde{X}_1, \tilde{X}_2$ donde $X_i, \tilde{X}_i$ son copias independientes de X_i. En realidad, podemos pensar que $X_1, X_2, \tilde{X}_1, \tilde{X}_2, X_1^0, X_2^0$ son todas independientes. Nuestras opciones principales para (X_1', X_2') son las sumas

$$X_1' = X_1 + \tilde{X}_2, X_2' = X_2 + \tilde{X}_1 \tag{4.3}$$

y

$$X_1' = X_1 + \tilde{X}_1, X_2' = X_2 + \tilde{X}_2 \tag{4.4}$$

o alternativamente las "fibras"

$$X_1' = (X_1|X_1 + \tilde{X}_2 = g), X_2' = (X_2|X_2 + \tilde{X}_1 = g') \tag{4.5}$$

y

$$X_1' = (X_1|X_1 + \tilde{X}_1 = g), X_2' = (X_2|X_2 + \tilde{X}_2 = g') \tag{4.6}$$

con $g, g' \in G$. Notar que esencialmente respetamos el mismo tipo de jugadas que teníamos en el mundo de los conjuntos, solo que ahora tenemos algunas variantes técnicas; nos permitimos distintas formas de representar la variable aleatoria asociada a $A + A$, ya que esto nos permitirá más tarde conseguir mejores cotas para la información, y porque ahora estamos trabajando en una versión del problema donde no tenemos un sólo conjunto A, sino que en realidad estaríamos trabajando con dos conjuntos A, B a una distancia acotada, y queremos hablar sobre las distintas formas de representar $A + B$, $A + A$, o $B + B$ dependiendo de lo que sea más útil en este contexto.

Más adelante, con el **lema del fibrado**, la Proposición 4.3, vamos a poder controlar las distancias de estas X_i'.

La conclusión es que o bien alguna de estas variables debe servir (es decir, cumplir la desigualdad (4.2)), o bien todas deben fallar *por poco* (por $O(\eta k)$), y esto último ocurre solo si todas las desigualdades usadas están cerca de ser igualdades. Motivado por esto, uno puede ir hacia atrás y ver qué pasa si la desigualdad del lema del fibrado, Proposición 4.3, es muy fina. Lo que ocurrirá en este caso es que si ni (4.3) ni (4.5) sirven para probar (4.2), obtendremos una cota superior

$$I_1 := \mathbf{I}\left[X_1 + X_2 \,:\, \tilde{X}_1 + X_2 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2\right] \leq 2\eta k \tag{4.7}$$

De manera informal, esto nos está diciendo que $X_1 + X_2$ y $\tilde{X}_1 + X_2$ son casi independientes condicionados por $X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2$.

Acotando de forma similar, si ni (4.4) ni (4.6) sirven para demostrar (4.2), obtenemos que

$$I_2 := \mathbf{I}\left[X_1 + X_2 \,:\, X_1 + \tilde{X}_1 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2\right] \leq 2\eta k + \frac{2\eta(2\eta k - I_1)}{1-\eta}, \quad (4.8)$$

y que

$$I_3 := \mathbf{I}\left[\tilde{X}_1 + X_2 \,:\, X_1 + \tilde{X}_1 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2\right] \leq 2\eta k + \frac{2\eta(2\eta k - I_1)}{1-\eta}, \quad (4.9)$$

debido a que $I_2 = I_3$, cosa que se vuelve evidente al cambiar los nombres entre $\tilde{X}_1$ y X_1.

Si ninguna de las opciones (4.3), (4.4), (4.5), (4.6) cumplen (4.2) (es decir, si estamos trabajando con una pareja de variables para las cuales ninguna de las jugadas "típicas" funcionan), entonces las tres estimaciones (4.7), (4.8) y (4.9) deben ser ciertas. En este punto, llegamos a una parte del argumento que los autores del paper [8] denominaron "final del juego". Esta misma parte vendría a ser la parte del final de la Introducción, donde nos encargamos de ver lo que ocurría cuando teníamos la información muy acotada.

Supongamos primero que las informaciones mutuas en (4.7), (4.8) y (4.9) son 0 en lugar de ser meramente pequeñas. Se sigue luego que para cualquier s en el soporte de $X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2$, las tres variables aleatorias

$$T_1 = (X_1 + X_2 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2), \quad (4.10)$$

$$T_2 = (X_1 + \tilde{X}_1 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2), \quad (4.11)$$

$$T_3 = (\tilde{X}_1 + X_2 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2) \quad (4.12)$$

son independientes de a pares. Notemos también que $T_1 + T_2 + T_3$ es constante (en particular es identicamente cero, ya que estamos en $\mathbf{F}_2^n$), y se puede ver haciendo la cuenta de la distancia que para cualquier terna (T_1, T_2, T_3) de variables aleatorias independientes dos a dos tales que $T_1 + T_2 + T_3$ es constante, se tiene que

$$d[T_1; T_2] = \mathbf{H}(T_3) - \frac{1}{2}\mathbf{H}(T_1) - \frac{1}{2}\mathbf{H}(T_2)$$

y de forma similar para permutaciones cíclicas del orden de las variables, por lo que

$$d[T_1; T_2] + d[T_1; T_3] + d[T_2; T_3] = 0 \quad (4.13)$$

y, al tomar X_1' y X_2' como algún par de T_1, T_2, T_3, obtenemos $d[X_1'; X_2'] = 0$. Esta conclusión es muy fuerte, y haciendo algunas cuentas más podemos probar (4.2) con estas dos variables.

Si las tres variables T_1, T_2 y T_3 son meramente "casi independientes dos a dos", como en (4.7), (4.8) y (4.9), entonces las cosas son menos directas. Recordamos que filosóficamente, sumas de variables aleatorias no independientes

son a variables aleatorias independientes lo que sumas de conjuntos *parciales* $A +_E B$ a través de un grafo bipartito E son a sumas totales de conjuntos $A + B$. Luego, es natural aplicar la variante del Teorema entrópico de Balog-Szemerédi-Gowers, Lema 2.7, para pasar de las variables aleatorias casi independientes (T_1, T_2) de arriba, para las cuales $\mathbf{H}(T_1 + T_2) - \frac{1}{2}\mathbf{H}(T_1) - \frac{1}{2}\mathbf{H}(T_2)$ es pequeño (o incluso negativo), a otras variables X_1', X_2' con $d[X_1'; X_2'] = O(\eta k)$. Una vez que obtengamos estas variables, ellas nos darán (4.2).

Si bien todo lo que recién describimos es la forma de demostrar la Proposición 4.2, y la manera en la cual los autores del artículo [8] dicen haberlo pensado, para estructurar bien la escritura de la demostración vamos a tomar el camino del contrarecíproco, al igual que en la Introducción, donde lo formulamos en la Proposición 1.15. Encarando esta dirección, supongamos que tenemos un par (X_1, X_2) de variables aleatorias G-valuadas con $d[X_1; X_2] = k$ para algún k, y supongamos que

$$\tau[X_1', X_2'] \geq \tau[X_1, X_2] \tag{4.14}$$

para cualquier par de variables aleatorias G-valuadas X_1', X_2'. A partir de la defínción de τ, podemos escribir la desigualdad de arriba como

$$d[X_1'; X_2'] \geq k - \tau(d[X_1^0; X_1'] - d[X_1^0; X_1]) - \tau(d[X_2^0; X_2'] - d[X_2^0; X_2]). \tag{4.15}$$

La idea es entonces es testear la desigualdad (4.15) con varias elecciones de X_1', X_2', generadas a partir de X_1, X_2, con el objetivo de demostrar que k debe valer 0. El sabor general de lo que vamos a hacer será: probar con las distintas jugadas, y al asumir que cada jugada no sirve, ir deduciendo distintas desigualdades que debería cumplir nuestra pareja de variables aleatorias. Testear (4.15) con (4.3) y (4.5), nos conducirá demostrar (4.7). Testear (4.15) con (4.4) y (4.6), nos conducirá demostrar (4.8), y claramente la misma cota puede deducirse para I_3, ya que $I_2 = I_3$. Los detalles de estas deducciones están en la Sección 4.4.

Finalmente, armados con la información otorgada por (4.7) y (4.8), procedemos al *final del juego*, usando las elecciones de X_1', X_2' dadas por (4.10) y el teorema entrópico de Balog-Szemerédi-Gowers, para demostrar que $k = 0$. Estos detalles están en la Sección (4.5).

La razón principal para argumentar por contradicción es que simplifica enormemente nuestras discusiones sobre variables aleatorias condicionadas. Nótese que (4.15) implica una versión condicionada de sí misma: para cualquier par de variables aleatorias (X_1', Y_1) y (X_2', Y_2), podemos aplicar (4.15) con X_1', X_2' reemplazados por cada una de las variables aleatorias condicionadas $(X_1'|Y_1 = y_1), (X_2'|Y_2 = y_2)$, obteniendo

$$\begin{aligned} d[(X_1'|Y_1 = y_1); (X_2'|Y_2 = y_2)] \\ \geq k - \eta(X_1'|Y_1 = y_1) - \eta(X_2'|Y_2 = y_2). \end{aligned}$$

Multiplicando por $p_{Y_1}(y_1)p_{Y_2}(y_2)$ y sumando, tenemos

$$\begin{aligned} d[X_1'|Y_1; X_2'|Y_2] \geq k - \eta\big(d[X_1^0; X_1'|Y_1] - d[X_1^0; X_1]\big) \\ - \eta\big(d[X_2^0; X_2'|Y_2] - d[X_2^0; X_2]\big). \end{aligned} \tag{4.16}$$

4.2 Algunos ejemplos

Para justificar más la estrategia que se les ocurrió a Gowers, Green, Manners y Tao, es útil considerar algunos casos de la forma $X_1 = U_{A_1}$, $X_2 = U_{A_2}$ para varios conjuntos $A_1, A_2 \subseteq \mathbf{F}_2^n$, y discutir qué elecciones de X_1', X_2' dan la estimación deseada (4.2). Esta discusión será la versión entrópica de la construcción de conjuntos con mejor constante de duplicación, que hizo en la Introducción, Sección 1.2. En lo siguiente será conveniente escribir $K := e^k$, donde k era $d[X_1; x_2]$. Una vez más, instamos al lector a comparar, durante esta discusión, con lo hecho la Introducción, y en como estos ejemplos y elecciones se relacionan con el caso de Combinatoria Aditiva clásica.

Ejemplo 1: Consideremos primero el caso en el que $A_1 = A_2 = A$, y A es un subconjunto aleatorio de algún subgrupo $H \leq \mathbf{F}_2^n$ con densidad $\frac{1}{K}$ en H. La intuición fundamental a tener en este caso es que, si uno trabaja con un conjunto "sin estructura aditiva" A, uno esperaría que

$$|A + A| \approx \binom{|A|}{2} \approx |A|^2,$$

de lo que uno intuye que, si A es suficientemente denso (es decir, $|A| \geq \sqrt{|H|}$), la suma $A + A$ llena todo el subgrupo H. Además, uno esperaría que $A + A$ no solo llene el subgrupo H, sino que lo haga de forma relativamente uniforme (de no ser así A estaría concentrado en alguna coclase de otro subgrupo), por lo que entenderemos que $X_1 + X_2$ tiene una distribución similar a la de U_H. Notar entonces que se pueden hacer las cuentas y llegar a que $d[X_1; X_2] \approx k$.

En este caso, la elección (4.4), es decir $X_1' = X_1 + \tilde{X}_1$ y $X_2' = X_2 + \tilde{X}_2$, establece inmediatamente (4.2). Efectivamente, tanto X_1' como X_2' están cerca de la distribución uniforme U_H sobre H, así que $d[X_1'; X_2'] \approx 0$ y $d[X_i'; X_i] \approx k/2$ para $i = 1, 2$ (cosa que se deduce al escribir la definición de la distanca de Ruzsa entrópica), y (4.2) se sigue, con margen de sobra, por la desigualdad triangular:

$$\begin{aligned}
\tau[X_1', X_2'] &= d[X_1', X_2'] + \eta d[X_1', X_0^1] + \eta d[X_2', X_0^2] \\
&\approx \eta d[X_1', X_0^1] + \eta d[X_2', X_0^2] \\
&\leq \eta(d[X_1', X_1] + d[X_1, X_0^1]) + \eta(d[X_2', X_2] + d[X_2, X_0^2]) \\
&\approx \eta(k/2 + d[X_1, X_0^1]) + \eta(k/2 + d[X_2, X_0^2]) \\
&\leq \eta k + \eta d[X_1, X_0^1] + \eta d[X_2, X_0^2] \\
&\leq k + \eta d[X_1, X_0^1] + \eta d[X_2, X_0^2] \\
&\approx \tau[X_1, X_2].
\end{aligned}$$

Ejemplo 2: Consideremos ahora el caso en el que $A_1 = \bigcup_{i=1}^m (x_i + H)$ y $A_2 = \bigcup_{i=1}^m (y_i + H)$, donde $H < \mathbf{F}_2^n$ es un subgrupo y los $x_i + H$ e $y_i + H$ son linealmente independientes en G/H. Tomando $m := K$, se obtiene $d[X_1; X_2] \approx k$ siguiendo las mismas ideas que recién, ya que uno esperaría que ahora la suma $A_1 + A_2$ llene m^2 coclases de H, debido a la independencia lineal de los $x_i + H$ e $y_i + H$.

En este caso se puede obtener (4.2) con la elección (4.5), es decir, $X_1' = (X_1|X_1 + \tilde{X}_2 = g)$ y $X_2' = (X_2|X_2 + \tilde{X}_1 = g')$ para algunos $g, g' \in A_1 + A_2$. Efectivamente, X_1' es la distribución uniforme sobre $A_1 \cap (A_2 + g)$, que si $g \in x_i + y_j + H$ es exactamente la coclase $x_i + H$ de H (por independencia lineal). De forma similar, X_2' es uniforme sobre una coclase $y_j' + H$, por lo que $d[X_1'; X_2'] \approx 0$ y nuevamente se sigue (4.2), siguiendo cuentas casi idénticas a las de arriba.

Ejemplo 3: En ambos ejemplos anteriores, alguna de las elecciones (4.3), (4.4), (4.5) o (4.6) ya da la estimación (4.2) y por lo tanto no es necesario pasar al *final del juego*.

Aquí hay un tercer ejemplo, una especie de combinación de los dos anteriores, para el cual las elecciones (4.3), (4.4), (4.5) y (4.6) todas fallan. Sean A_1', A_2' subconjuntos aleatorios (independientes) de densidad $1/m$ de los conjuntos A_1, A_2 del ejemplo anterior, donde ahora $m = \sqrt{K}$.

Se puede verificar que $X_1 + \tilde{X}_1, X_2 + \tilde{X}_2$ se parecen a la distribución uniforme sobre la unión de aproximadamente $K/2$ coclases de H, ya que ahora tenemos subconjuntos densos de $\sqrt{K}$ coclases de H, por lo que si asumimos que no hay muchas sumas repetidas tendremos que $A_1' + A_1'$ llena varias coclases de H, y cuando decimos varias nos referimos a las $\binom{m}{2}$ formas de combinar los x_i, que en este caso es aproximadamente $K/2$. Debido a esto, $d[X_1'; X_2'] \approx k - O(1)$ y no se va a obtener (4.2) con la elección (4.4), cosa que se puede ver al intentar rehacer las cuentas de más arriba (en los dos ejemplos anteriores obteniamos $d[X_1'; X_2'] \approx 0$, cosa que era fundamental para hacer las cuentas). El caso de (4.3) es muy similar, ya que las estimaciones que haremos serán las mismas, todo apoyado sobre la idea de que la suma de dos conjuntos sin mucha estructura se distribuye de forma homogénea en el subgrupo.

Por otro lado, las variables $(X_1|X_1 + \tilde{X}_2 = g)$, $(X_2|X_2 + \tilde{X}_1 = g')$ serán uniformes en $A_1 \cap (A_2 + g)$ y $A_2 \cap (A_1 + g')$ respectivamente, y dichos conjuntos típicamente (cuando no son vacíos) se asemejan a subconjuntos aleatorios de una coclase de H, de densidad $1/m^2 \approx 1/K$. Para estas variables también se tiene $d[X_1'; X_2'] \approx k$, por lo que no se obtiene (4.2) con la elección (4.5). El caso de (4.6) es similar.

Por lo tanto, procedemos al final del juego y consideramos las tres variables en (4.10). Ya que $X_1 + X_2$ es aproximadamente uniforme sobre $B := \bigcup_{i,j}(x_i + y_j + H)$, un análisis similar al del segundo ejemplo muestra que

$$(X_1 + X_2|X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2 = s)$$

es aproximadamente uniforme sobre $B \cap (B + s)$, que típicamente es una unión de cuatro coclases de H (pensar que $g = (x_1 + y_1) + (x_2 + y_2) = (x_1 + y_2) + (x_2 + y_1)$), y de forma similar para $(\tilde{X}_1 + X_2|X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2 = s)$.

Además, se puede verificar que estas dos variables son "50% independientes", en el sentido de que conocer una de ellas reduce la elección de la otra a dos coclases de H. Similarmente, también podríamos haber hecho algo válido arrancando con $X_1 + \tilde{X}_1$ y $X_2 + \tilde{X}_2$.

Si aplicamos entonces el Lema 2.7, por ejemplo con

$$A = (X_1 + \tilde{X}_1 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2 = s)$$

y

$$B = (X_1 + X_2 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2 = s),$$

y luego con las otras permutaciones de las variables, encontramos variables X_1', X_2' de la forma

$$\begin{aligned} X_1' =& (X_1 + X_2 | \tilde{X}_1 + X_2 + X_1 + \tilde{X}_2 = s, \tilde{X}_1 + X_2 = t) \\ =& (X_1 + X_2 | \tilde{X}_1 + X_2 = t, X_1 + \tilde{X}_2 = s + t), \\ X_2' =& (X_1 + \tilde{X}_1 | \tilde{X}_1 + X_2 + X_1 + \tilde{X}_2 = s, \tilde{X}_1 + X_2 = t) \\ =& (X_1 + \tilde{X}_1 | \tilde{X}_1 + X_2 = t, X_1 + \tilde{X}_2 = s + t). \end{aligned}$$

En otras palabras, X_1' es una suma de dos variables de la forma $(X_1 | X_1 + \tilde{X}_2 = a)$ y $(X_2 | X_2 + \tilde{X}_1 = b)$. Cada una será uniforme sobre un subconjunto aleatorio de una coclase de H, y sumarlas dará algo cercano a uniforme sobre una coclase de H. El caso de X_2' es similar, así que obtenemos $d[X_1'; X_2'] \approx 0$ y se sigue (4.2), escribiendo las mismas desigualdades que desarrollamos arriba. Con esto terminamos concluyendo que, al menos en este caso, cuando las dos jugadas típicas fallan, la variable fabricada con el lema de Balog-Szemerédi-Gowers nos da una pareja de variables útil.

4.3 Lema del fibrado

Comenzando con la demostración, vamos a demostrar el tan anticipado lema del fibrado, en su versión general:

Proposición 4.3 (Lema del fibrado)**.** *Sea $\pi\colon H \to H'$ un morfismo entre grupos abelianos y sean Z_1, Z_2 variables aleatorias con valores en H. Entonces tenemos*

$$d[Z_1; Z_2] \geq d[\pi(Z_1); \pi(Z_2)] + d[Z_1 | \pi(Z_1); Z_2 | \pi(Z_2)]. \tag{4.17}$$

Además, si Z_1, Z_2 son independientes, la diferencia entre los dos lados es

$$\mathbf{I}\big[Z_1 - Z_2 : (\pi(Z_1), \pi(Z_2)) \mid \pi(Z_1 - Z_2)\big]. \tag{4.18}$$

Demostración. Para probar (4.17), por la definición de distancia de Ruzsa entrópica, podemos tomar Z_1, Z_2 independientes. Se tiene que

$$\begin{aligned} & d\big[Z_1 | \pi(Z_1); Z_2 | \pi(Z_2)\big] \\ &= \mathbf{H}\big[Z_1 - Z_2 | \pi(Z_1), \pi(Z_2)\big] - \tfrac{1}{2}\mathbf{H}[Z_1 | \pi(Z_1)] - \tfrac{1}{2}\mathbf{H}[Z_2 | \pi(Z_2)] \\ &\leq \mathbf{H}\big[Z_1 - Z_2 | \pi(Z_1 - Z_2)\big] - \tfrac{1}{2}\mathbf{H}[Z_1 | \pi(Z_1)] - \tfrac{1}{2}\mathbf{H}[Z_2 | \pi(Z_2)] \\ &= d[Z_1; Z_2] - d[\pi(Z_1); \pi(Z_2)]. \end{aligned}$$

Donde el último paso se deduce de que

$$\mathbf{H}(A|B) = \mathbf{H}(A) - \mathbf{H}(B),$$

siempre que A determine a B, y de la definición de distancia de Ruzsa entrópica, mientras que, para justificar el paso del medio, hace falta usar submodularidad, Lema 2.2, recordando que esta propiedad puede pensarse también como que

$$\mathbf{H}(A,B,C) + \mathbf{H}(C) \leq \mathbf{H}(A,C) + \mathbf{H}(B,C)$$

o, lo que nos es más útil acá,

$$\mathbf{H}(A|B,C) \leq \mathbf{H}(A|C).$$

Con esto ya probamos la desigualdad (4.17). Notar ahora de la identidad

$$\mathbf{H}(A|B) - \mathbf{H}(A|B,C) = \mathbf{I}[A : C|B],$$

que podríamos tomar $A := Z_1 - Z_2, B = \pi(Z_1 - Z_2)$, $C := (\pi(Z_1), \pi(Z_2))$, y como en este caso C determina a B, valde que $\mathbf{H}(A|B,C) = \mathbf{H}(A|C)$. Todo esto nos da que

$$\begin{aligned}\mathbf{H}(Z_1 - Z_2|\pi(Z_1 - Z_2)) - \mathbf{H}(Z_1 - Z_2|\pi(Z_1), \pi(Z_2)) \\ = \mathbf{I}[Z_1 - Z_2 : (\pi(Z_1), \pi(Z_2))|\pi(Z_1 - Z_2)],\end{aligned}$$

que era lo último que nos faltaba probar. □

Recordamos que la desigualdad (4.17) es una formulación precisa de la idea intuitiva de que la constante de duplicación de un subconjunto de G debería, al aplicarsele un morfismo $\pi : G \to H$, ser al menos la constante de duplicación del conjunto "base" multiplicada por alguna combinación o promedio de las constantes de duplicación de las fibras. Como ya dijimos en la Introducción, cuando mencionamos que (1.2) no es verdadera, esta formulación no es posible de hacerse en el paradigma clásico de constante de duplicación combinatoria, sino que hay que pasar necesariamente a la formulación entrópica, y es este uso de la entropía un paso crucial en la prueba.

Enunciamos ahora una versión específica de la Proposición 4.3 que vamos a utilizar varias veces:

Corolario 4.4. *Sean Y_1, Y_2, Y_3 e Y_4 variables aleatorias independientes que toman valores en algún grupo abeliano G. Entonces,*

$$\begin{aligned}d[Y_1 - Y_3; Y_2 - Y_4] + d\big[Y_1|Y_1 - Y_3; Y_2|Y_2 - Y_4\big] \\ + \mathbf{I}\big[Y_1 - Y_2 : Y_2 - Y_4|Y_1 - Y_2 - Y_3 + Y_4\big] = d[Y_1; Y_2] + d[Y_3; Y_4].\end{aligned}$$

Demostración. Aplicamos el lema del fibrado, Proposición 4.3, con $H := G \times G$, $H' := G$, y π el morfismo de resta $\pi(x,y) := x - y$, con las variables aleatorias $Z_1 := (Y_1, Y_3)$ y $Z_2 := (Y_2, Y_4)$. Por independencia tenemos entonces

$$d[Z_1; Z_2] = d[Y_1; Y_2] + d[Y_3; Y_4],$$

y al aplicar la definición del morfismo se tiene que

$$d[\pi(Z_1);\pi(Z_2)] = d[Y_1 - Y_3; Y_2 - Y_4].$$

Además,

$$d[Z_1|\pi(Z_1); Z_2|\pi(Z_2)] = d[Y_1|Y_1 - Y_3; Y_2|Y_2 - Y_4],$$

ya que $Z_1 = (Y_1, Y_3)$ e Y_1 están vinculados por una transformación afín invertible una vez que se fija $\pi(Z_1) = Y_1 - Y_3$, cosa que se ve de manera similar para Z_2 e Y_2.

Finalmente, tenemos

$$\begin{aligned}
&\mathbf{I}\big[Z_1 - Z_2 : (\pi(Z_1), \pi(Z_2)) \,|\, \pi(Z_1) + \pi(Z_2)\big] \\
&\quad = \mathbf{I}\big[(Y_1 - Y_2, Y_3 - Y_4) : (Y_1 - Y_3, Y_2 - Y_4) \,|\, Y_1 - Y_2 - Y_3 + Y_4\big] \\
&\quad = \mathbf{I}\big[Y_1 - Y_2 : Y_2 - Y_4 \,|\, Y_1 - Y_2 - Y_3 + Y_4\big]
\end{aligned}$$

donde en la última línea utilizamos el hecho de que $(Y_1 - Y_2, Y_1 - Y_2 - Y_3 + Y_4)$ determinan de manera única a $Y_3 - Y_4$ y, de manera similar, $(Y_2 - Y_4, Y_1 - Y_2 - Y_3 + Y_4)$ determinan de manera única a $Y_1 - Y_3$. □

Por supuesto, en nuestro contexto de característica 2, los signos negativos en todos estos resultados pueden ser reemplazados por signos positivos, cosa que haremos sin cuidado.

4.4 Estimaciones sobre la información mutua

El objetivo de esta sección es acotar efectivamente las informaciones mutuas de (4.7), (4.8) y (4.9), a partir de "testear" con distintas jugadas la desigualdad (4.15). Para esto, recordemos que $X_1, X_2, \tilde{X}_1, \tilde{X}_2$ son variables aleatorias independientes, donde $X_1, \tilde{X}_1$ son copias de X_1 y $X_2, \tilde{X}_2$ son copias de X_2.

Caso 1: Primero vamos a establecer la cota superior (4.7), que, asumiendo (4.15), afirma que

$$I_1 \coloneqq \mathbf{I}\big[X_1 + X_2 : \tilde{X}_1 + X_2 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2\big] \leq 2\eta k.$$

Para ver esto, aplicamos el Corolario 4.4 con la elección

$$(Y_1, Y_2, Y_3, Y_4) \coloneqq (X_1, X_2, \tilde{X}_2, \tilde{X}_1),$$

que nos da

$$\begin{aligned}
&d[X_1 + \tilde{X}_2; X_2 + \tilde{X}_1] + d\big[X_1 | X_1 + \tilde{X}_2; X_2 | X_2 + \tilde{X}_1\big] \\
&\quad + \mathbf{I}\big[X_1 + X_2 : \tilde{X}_1 + X_2 \,|\, X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2\big] = 2k,
\end{aligned} \tag{4.19}$$

como $d[X_1; X_2] = k$.

Aplicando (4.15) y (4.16), obtenemos

$$d[X_1+\tilde{X}_2;X_2+\tilde{X}_1] \geq k-\eta\big(d[X_1^0;X_1+\tilde{X}_2]-d[X_1^0;X_1]\big) \\ -\eta\big(d[X_2^0;X_2+\tilde{X}_1]-d[X_2^0;X_2]\big)$$

y

$$d[X_1|X_1+\tilde{X}_2;X_2|X_2+\tilde{X}_1] \\ \geq k-\eta\big(d[X_1^0;X_1|X_1+\tilde{X}_2]-d[X_1^0;X_1]\big) \tag{4.20} \\ -\eta\big(d[X_2^0;X_2|X_2+\tilde{X}_1]-d[X_2^0;X_2]\big). \tag{4.21}$$

Por lo tanto, para demostrar (4.7), basta demostrar que

$$(d[X_1^0;X_1+\tilde{X}_2]-d[X_1^0;X_1])+(d[X_2^0;X_2+\tilde{X}_1]-d[X_2^0;X_2]) \\ +(d[X_1^0;X_1|X_1+\tilde{X}_2]-d[X_1^0;X_1]) \\ +(d[X_2^0;X_2|X_2+\tilde{X}_1]-d[X_2^0;X_2])\leq 2k. \tag{4.22}$$

Para probar esto, primero nos detenemos a enunciar algunos lemas que nos serán de utilidad en todo lo que resta de la demostración. El espíritu de estos lemas en general es controlar como se modifica la distancia entre dos variables aleatorias al hacerle ciertas modificaciones, por lo general de un sabor aditivo o de condicionamiento.

El primero es una cota que relaciona las variables condicionadas y no condicionadas con la distancia de Ruzsa.

Lema 4.5. *Supongamos que (X,Z) y (Y,W) son variables aleatorias, donde X,Y toman valores en algún grupo abeliano. Entonces,*

$$d[X|Z;Y|W]\leq d[X;Y]+\tfrac{1}{2}\mathbf{I}\left[X\ :\ Z\right]+\tfrac{1}{2}\mathbf{I}\left[Y\ :\ W\right].$$

Demostración. Utilizando la expresión alternativa de la distancia de Ruzsa condicional (2.18), si $(X',Z'),(Y',W')$ son copias independientes de las variables (X,Z), (Y,W), tenemos

$$\begin{aligned} d[X|Z;Y|W] &= \mathbf{H}[X'-Y'|Z',W']-\tfrac{1}{2}\mathbf{H}[X'|Z']-\tfrac{1}{2}\mathbf{H}[Y'|W'] \\ &\leq \mathbf{H}[X'-Y']-\tfrac{1}{2}\mathbf{H}[X'|Z']-\tfrac{1}{2}\mathbf{H}[Y'|W'] \\ &= d[X';Y']+\tfrac{1}{2}\mathbf{I}\left[X'\ :\ Z'\right]+\tfrac{1}{2}\mathbf{I}\left[Y'\ :\ W'\right]. \end{aligned}$$

La última igualdad se obtiene usando las definiciones de $d[-;-]$ e $\mathbf{I}\left[-\ :\ -\right]$. □

Tenemos también la siguiente estimación:

Lema 4.6. *Sean X,Y,Z variables aleatorias que toman valores en algún grupo abeliano, con Y,Z independeentes. Tenemos entonces que*

$$\begin{aligned} d[X;Y-Z]-d[X;Y] &\leq \tfrac{1}{2}(\mathbf{H}[Y-Z]-\mathbf{H}[Y]) \\ &= \tfrac{1}{2}d[Y;Z]+\tfrac{1}{4}\mathbf{H}[Z]-\tfrac{1}{4}\mathbf{H}[Y] \end{aligned} \tag{4.23}$$

y

$$d[X;Y|Y-Z]-d[X;Y] \le \tfrac{1}{2}\big(\mathbf{H}[Y-Z]-\mathbf{H}[Z]\big) \tag{4.24}$$
$$= \tfrac{1}{2}d[Y;Z]+\tfrac{1}{4}\mathbf{H}[Y]-\tfrac{1}{4}\mathbf{H}[Z]. \tag{4.25}$$

Demostración. Demostramos primero (4.23). Podemos asumir (tomando una copia independiente) que X es independiente de Y, Z. Tenemos entonces que

$$\begin{aligned} &d[X;Y-Z]-d[X;Y] \\ &\qquad = \mathbf{H}[X-Y+Z]-\mathbf{H}[X-Y]-\tfrac{1}{2}\mathbf{H}[Y-Z]+\tfrac{1}{2}\mathbf{H}[Y]. \end{aligned}$$

Combinando esto con el Lema 2.5 (aplicado con Y reemplazado por $-Y$) nos da la cota deseada. Notar que la igualdad en (4.23) es directo de la definición de $d[Y;Z]$, ya que Y, Z son independientes.

Para probar la desigualdad (4.25), tenemos

$$\begin{aligned} \mathbf{I}[Y\ :\ Y-Z] &= \mathbf{H}[Y]+\mathbf{H}[Y-Z]-\mathbf{H}[Y,Y-Z] \\ &= \mathbf{H}[Y]+\mathbf{H}[Y-Z]-\mathbf{H}[Y,Z] = \mathbf{H}[Y-Z]-\mathbf{H}[Z], \end{aligned}$$

y luego (4.25) es una consecuencia del Lema 4.5. Una vez más, la segunda forma es un resultado de la definición de $d[Y;Z]$. □

Volvemos ahora a nuestra labor de probar el caso 1, que es establecer (4.22), y por ende (4.7). Por el Lema 4.6 (y recordando que k está definida como $d[X_1;X_2]$) tenemos que

$$d[X_1^0;X_1+\tilde{X}_2]-d[X_1^0;X_1] \le \tfrac{1}{2}k+\tfrac{1}{4}\mathbf{H}[X_2]-\tfrac{1}{4}\mathbf{H}[X_1],$$

$$d[X_2^0;X_2+\tilde{X}_1]-d[X_2^0;X_2] \le \tfrac{1}{2}k+\tfrac{1}{4}\mathbf{H}[X_1]-\tfrac{1}{4}\mathbf{H}[X_2],$$

$$d[X_1^0;X_1|X_1+\tilde{X}_2]-d[X_1^0;X_1] \le \tfrac{1}{2}k+\tfrac{1}{4}\mathbf{H}[X_1]-\tfrac{1}{4}\mathbf{H}[X_2] \tag{4.26}$$

y que

$$d[X_2^0;X_2|X_2+\tilde{X}_1]-d[X_2^0;X_2] \le \tfrac{1}{2}k+\tfrac{1}{4}\mathbf{H}[X_2]-\tfrac{1}{4}\mathbf{H}[X_1]. \tag{4.27}$$

Sumando todas estas desigualdades, obtenemos (4.22).

Para usar en las próximas dos secciones, remarcamos que restar (4.20) de (4.19), y combinando la desigualdad resultante con (4.26) y (4.27) nos da la cota

$$d[X_1+\tilde{X}_2;X_2+\tilde{X}_1] \le (1+\eta)k-I_1,$$

que es equivalente a

$$\mathbf{H}\big[X_1+X_2+\tilde{X}_1+\tilde{X}_2\big] \le \tfrac{1}{2}\mathbf{H}[X_1]+\tfrac{1}{2}\mathbf{H}[X_2]+(2+\eta)k-I_1. \tag{4.28}$$

Caso 2: Seguimos ahora con la prueba de la estimación (4.8), que era

$$I_2 \coloneqq \mathbf{I}\big[X_1+X_2 : X_1+\tilde{X}_1|X_1+X_2+\tilde{X}_1+\tilde{X}_2\big] \le 2\eta k+\frac{2\eta(2\eta k-I_1)}{1-\eta}.$$

Con la elección

$$(Y_1, Y_2, Y_3, Y_4) \coloneqq (X_2, X_1, \tilde{X}_2, \tilde{X}_1),$$

el Corolario 4.4 ahora se reescribe como

$$\begin{aligned} &d[X_1 + \tilde{X}_1; X_2 + \tilde{X}_2] + d\big[X_1 | X_1 + \tilde{X}_1; X_2 | X_2 + \tilde{X}_2\big] \\ &\quad + \mathbf{I}\big[X_1 + X_2 : X_1 + \tilde{X}_1 \,|\, X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2\big] = 2k, \end{aligned}$$

recordando otra vez que $k \coloneqq d[X_1; X_2]$. Aplicando (4.15) y (4.16) como en el caso 1, obtenemos

$$\begin{aligned} d[X_1 + \tilde{X}_1; X_2 + \tilde{X}_2] \geq k &- \eta\big(d[X_1^0; X_1 + \tilde{X}_1] - d[X_1^0; X_1]\big) \\ &- \eta\big(d[X_2^0; X_2 + \tilde{X}_2] - d[X_2^0; X_2]\big) \end{aligned} \tag{4.29}$$

y

$$\begin{aligned} d[X_1 | X_1 + \tilde{X}_1; X_2 | X_2 + \tilde{X}_2] \geq k &- \eta\big(d[X_1^0; X_1 | X_1 + \tilde{X}_1] - d[X_1^0; X_1]\big) \\ &- \eta\big(d[X_2^0; X_2 | X_2 + \tilde{X}_2] - d[X_2^0; X_2]\big). \end{aligned}$$

Ahora el Lema 4.6 nos da

$$d[X_1^0; X_1 + \tilde{X}_1] - d[X_1^0; X_1] \leq \tfrac{1}{2} d[X_1; X_1], \tag{4.30}$$

$$d[X_2^0; X_2 + \tilde{X}_2] - d[X_2^0; X_2] \leq \tfrac{1}{2} d[X_2; X_2], \tag{4.31}$$

$$d[X_1^0; X_1 | X_1 + \tilde{X}_1] - d[X_1^0; X_1] \leq \tfrac{1}{2} d[X_1; X_1],$$

y

$$d[X_2^0; X_2 | X_2 + \tilde{X}_2] - d[X_2^0; X_2] \leq \tfrac{1}{2} d[X_2; X_2].$$

Combinando todas esta desigualdades y cancelando los términos se obtiene

$$\mathbf{I}\left[X_1 + X_2 \,:\, X_1 + \tilde{X}_1 | X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2\right] \leq \eta(d[X_1; X_1] + d[X_2; X_2]). \tag{4.32}$$

Uno podría acotar el lado derecho con $4\eta k$ usando la desigualdad triangular de Ruzsa, pero el siguiente razonamiento da mejores constantes, con el costo de volverse un poco más rebuscado. Primero, combinando (4.29), (4.30) y (4.31), obtenemos

$$d[X_1 + \tilde{X}_1; X_2 + \tilde{X}_2] \geq k - \frac{\eta}{2}(d[X_1; X_1] + d[X_2; X_2]). \tag{4.33}$$

también podemos expandir

$$\begin{aligned} &d[X_1 + \tilde{X}_1; X_2 + \tilde{X}_2] \\ &= \mathbf{H}[X_1 + \tilde{X}_1 + X_2 + \tilde{X}_2] - \tfrac{1}{2}\mathbf{H}[X_1 + \tilde{X}_1] - \tfrac{1}{2}\mathbf{H}[X_2 + \tilde{X}_2] \\ &= \mathbf{H}[X_1 + \tilde{X}_1 + X_2 + \tilde{X}_2] - \tfrac{1}{2}\mathbf{H}[X_1] - \tfrac{1}{2}\mathbf{H}[X_2] \\ &\qquad\qquad - \tfrac{1}{2}\left(d[X_1; X_1] + d[X_2; X_2]\right), \end{aligned}$$

y luego por (4.28)

$$d[X_1 + \tilde{X}_1; X_2 + \tilde{X}_2] \leq (2+\eta)k - \tfrac{1}{2}\left(d[X_1; X_1] + d[X_2; X_2]\right) - I_1.$$

Combinando esto con (4.33) obtenemos

$$d[X_1; X_1] + d[X_2; X_2] \leq 2k + \frac{2(2\eta k - I_1)}{1-\eta}, \tag{4.34}$$

y reemplazando en (4.32) obtenemos la cota deseada (4.8).

4.5 Final del juego

Llegamos ahora al paso final de la demostración de la Proposición 4.2. El principal resultado desde el cual se podrá concluir el resultado es el Lema 4.8, pero para poder aplicarlo, primero haremos ciertas estimaciones sobre algunas variables aleatorias, utilizando lo que hicimos en las últimas dos secciones.

Comencemos probando una desigualdad que será utilizada varias veces en las cuentas de abajo, que es simplemente un reordenamiento de los lemas previos.

Lema 4.7. *Sean X, Y, Z, Z' variables aleatorias que toman valores en un grupo, con Y, Z, Z' independientes. Tenemos entonces que*

$$\begin{aligned} &d[X; Y - Z | Y - Z - Z'] - d[X; Y] \\ &\quad \leq \tfrac{1}{2}(\mathbf{H}[Y - Z - Z'] + \mathbf{H}[Y - Z] - \mathbf{H}[Y] - \mathbf{H}[Z']). \end{aligned} \tag{4.35}$$

Demostración. Por (4.25) (con un cambio de variables) obtenemos

$$d[X; Y - Z | Y - Z - Z'] - d[X; Y - Z] \leq \tfrac{1}{2}(\mathbf{H}[Y - Z - Z'] - \mathbf{H}[Z']).$$

Agregando esto a (4.23) nos da el resultado. □

Sean $X_1, X_2, \tilde{X}_1, \tilde{X}_2$ como antes, e introducimos las variables aleatorias

$$U \coloneqq X_1 + X_2, \qquad V \coloneqq \tilde{X}_1 + X_2, \qquad W \coloneqq X_1 + \tilde{X}_1$$

y

$$S \coloneqq X_1 + X_2 + \tilde{X}_1 + \tilde{X}_2.$$

De las definiciones (4.7), (4.8), (4.9) de I_1, I_2, I_3, y con la notación de arriba, vemos que

$$I_1 = \mathbf{I}[U : V \,|\, S], \qquad I_2 = \mathbf{I}[W : U \,|\, S], \qquad I_3 = \mathbf{I}[V : W \,|\, S].$$

A partir de (4.7), (4.8) y (4.9) tenemos las desigualdades

$$\mathbf{I}[V : W \,|\, S], \mathbf{I}[W : U \,|\, S] \leq 2\eta k + \frac{2\eta(2\eta k - I_1)}{1-\eta}.$$

Sumando estas dos desigualdades y la igualdad $\mathbf{I}[U : V \mid S] = I_1$ da

$$\begin{aligned}\mathbf{I}[U : V \mid S] + \mathbf{I}[V : W \mid S] + \mathbf{I}[W : U \mid S] & \\ &\leq I_1 + 4\eta k + \frac{4\eta(2\eta k - I_1)}{1-\eta} \\ &= 6\eta k - \frac{1-5\eta}{1-\eta}(2\eta k - I_1). \end{aligned} \tag{4.36}$$

Recomendamos al lector leer el argumento asumiendo primero que $I_1 = 2\eta k$, caso en el cual las cuentas son mucho más limpias. Primero hacemos un rejunte de estimaciones preliminares sobre las distancias:

Por el Lema 4.7 (otra vez cambiando todos los signos de resta con signos de suma, y tomando $X = X_1^0$, $Y = X_1$, $Z = X_2$ y $Z' = \tilde{X}_1 + \tilde{X}_2$, de forma que $Y + Z = U$ y $Y + Z + Z' = S$) tenemos, notando que $\mathbf{H}[Y + Z] = \mathbf{H}[Z']$,

$$d[X_1^0; U|S] - d[X_1^0; X_1] \leq \tfrac{1}{2}(\mathbf{H}[S] - \mathbf{H}[X_1]).$$

Más aplicaciones del Lema 4.7 dan

$$\begin{aligned} d[X_2^0; U|S] - d[X_2^0; X_2] &\leq \tfrac{1}{2}(\mathbf{H}[S] - \mathbf{H}[X_2]) \\ d[X_1^0; V|S] - d[X_1^0; X_1] &\leq \tfrac{1}{2}(\mathbf{H}[S] - \mathbf{H}[X_1]) \\ d[X_2^0; V|S] - d[X_2^0; X_2] &\leq \tfrac{1}{2}(\mathbf{H}[S] - \mathbf{H}[X_2]) \end{aligned}$$

y

$$d[X_1^0; W|S] - d[X_1^0; X_1] \leq \tfrac{1}{2}(\mathbf{H}[S] + \mathbf{H}[W] - \mathbf{H}[X_1] - \mathbf{H}[W']),$$

donde $W' := X_2 + \tilde{X}_2$. Para tratar con $d[X_2^0; W|S]$, primero notar que es igual a $d[X_2^0; W'|S]$, ya que para una elección fija s de S tenemos $W' = W + s$. Ahora podemos aplicar una vez más el Lema 4.7 para obtener

$$d[X_2^0; W'|S] - d[X_2^0; X_2] \leq \tfrac{1}{2}(\mathbf{H}[S] + \mathbf{H}[W'] - \mathbf{H}[X_2] - \mathbf{H}[W]).$$

Sumando estas seis estimaciones y usando (4.28), podemos concluir que

$$\begin{aligned}\sum_{i=1}^{2} \sum_{A \in \{U,V,W\}} \big(d[X_i^0; A|S] - d[X_i^0; X_i]\big) & \\ &\leq 3\mathbf{H}[S] - \tfrac{3}{2}\mathbf{H}[X_1] - \tfrac{3}{2}\mathbf{H}[X_2] \\ &\leq (6 - 3\eta)k + 3(2\eta k - I_1). \end{aligned} \tag{4.37}$$

Llegamos ahora a la observación clave que explotaremos, que es que

$$U + V + W = 0. \tag{4.38}$$

Acá estamos usando fuertemente el hecho de estar en característica 2. Esta es la única vez que se usa de forma crítica esto en el argumento, en el sentido de que no puede esquivarse con arbitrarias inserciones de signos negativos o aceptando constantes levemente peores invocando la desigualdad $d[X; -Y] \leq 3d[X; Y]$ y estimaciones relacionadas.

Para ver la fuerza de (4.36) y (4.38), registramos el siguiente resultado.

Lema 4.8. *Sea $G = \mathbf{F}_2^n$ y sea (T_1, T_2, T_3) una variable aleatoria G^3-valuada tal que $T_1 + T_2 + T_3 = 0$. Sea*

$$\delta := \sum_{1 \le i < j \le 3} \mathbf{I}\left[T_i \,:\, T_j\right]. \tag{4.39}$$

entonces existen variables aleatorias T_1', T_2' tales que

$$\begin{aligned} d[T_1'; T_2'] + \eta(d[X_1^0; T_1'] - d[X_1^0; X_1]) + \eta(d[X_2^0; T_2'] - d[X_2^0; X_2]) \\ \le \delta + \frac{\eta}{3}\Big(\delta + \sum_{i=1}^{2}\sum_{j=1}^{3}(d[X_i^0; T_j] - d[X_i^0; X_i])\Big). \end{aligned}$$

Demostración. Aplicamos la variante del teorema Balog–Szemerédi–Gowers entrópico probado en el Lema 2.7, tomando $(A, B) = (T_1, T_2)$. Como $T_1 + T_2 = T_3$, la conclusión es que

$$\begin{aligned} \sum_{t_3} p_{T_3}(t_3) d[(T_1 | T_3 = t_3); (T_2 | T_3 = t_3)] \\ \le 3\mathbf{I}\left[T_1 \,:\, T_2\right] + 2\mathbf{H}[T_3] - \mathbf{H}[T_1] - \mathbf{H}[T_2]. \end{aligned} \tag{4.40}$$

El lado derecho en (4.40) se puede arreglar como

$$\begin{aligned} &2(\mathbf{H}[T_1] + \mathbf{H}[T_2] + \mathbf{H}[T_3]) - 3\mathbf{H}[T_1, T_2] \\ &= 2(\mathbf{H}[T_1] + \mathbf{H}[T_2] + \mathbf{H}[T_3]) - \mathbf{H}[T_1, T_2] - \mathbf{H}[T_2, T_3] - \mathbf{H}[T_1, T_3] = \delta, \end{aligned}$$

usando el hecho de que los tres términos $\mathbf{H}[T_i, T_j]$ son iguales a $\mathbf{H}[T_1, T_2, T_3]$, y por ende iguales entre ellos. También tenemos

$$\begin{aligned} &\sum_{t_3} p_{T_3}(t_3)\big(d[X_1^0; (T_1 | T_3 = t_3)] - d[X_1^0; X_1]\big) \\ &\quad = d[X_1^0; T_1 | T_3] - d[X_1^0; X_1] \le d[X_1^0; T_1] - d[X_1^0; X_1] + \tfrac{1}{2}\mathbf{I}\left[T_1 \,:\, T_3\right] \end{aligned}$$

por el Lema 4.5, y similarmente

$$\begin{aligned} \sum_{t_3} p_{T_3}(t_3)(d[X_2^0; (T_2 | T_3 = t_3)] - d[X_2^0; X_2]) \\ \le d[X_2^0; T_2] - d[X_2^0; X_2] + \tfrac{1}{2}\mathbf{I}\left[T_2 \,:\, T_3\right]. \end{aligned}$$

Definamos por ahora

$$\psi[Y_1; Y_2] := d[Y_1; Y_2] + \eta(d[X_1^0; Y_1] - d[X_1^0; X_1]) + \eta(d[X_2^0; Y_2] - d[X_2^0; X_2]).$$

Juntando las observaciones de arriba, tenemos

$$\begin{aligned} \sum_{t_3} p_{T_3}(t_3)\psi[(T_1 | T_3 = t_3); (T_2 | T_3 = t_3)] \le \delta + \eta(d[X_1^0; T_1] - d[X_1^0; X_1]) \\ + \eta(d[X_2^0; T_2] - d[X_2^0; X_2]) + \tfrac{1}{2}\eta\mathbf{I}\left[T_1 \,:\, T_3\right] + \tfrac{1}{2}\eta\mathbf{I}\left[T_2 \,:\, T_3\right]. \end{aligned}$$

Escogiendo algún t_3 en el soporte de T_3 que minimiza el valor $\psi[-;-]$, y tomando $T'_{1,3} \coloneqq (T_1|T_3 = t_3)$, $T'_{2,3} \coloneqq (T_2|T_3 = t_3)$, tenemos

$$\begin{aligned}\psi[T'_{1,3};T'_{2,3}] \le \delta + \eta(d[X_1^0;T_1] - d[X_1^0;X_1]) + \eta(d[X_2^0;T_2] - d[X_2^0;X_2]) \\ + \tfrac{1}{2}\eta\mathbf{I}\left[T_1 \,:\, T_3\right] + \tfrac{1}{2}\eta\mathbf{I}\left[T_2 \,:\, T_3\right].\end{aligned} \tag{4.41}$$

Repetimos ahora este mismo análisis para todas las permutaciones de $\{T_1, T_2, T_3\}$ para obtener variables $T'_{\alpha,\gamma}, T'_{\beta,\gamma}$ con $\{\alpha, \beta, \gamma\}$ moviendose entre todas las seis permutaciones de $\{1, 2, 3\}$. Promediando las desigualdades (4.41), y recordando la definición (4.39) de δ, obtenemos

$$\tfrac{1}{6}\sum_{\alpha,\beta,\gamma}\psi[T'_{\alpha,\gamma};T'_{\beta,\gamma}] \le \delta + \frac{\eta}{3}\Big(\delta + \sum_{i=1}^{2}\sum_{j=1}^{3}(d[X_i^0;T_j] - d[X_i^0;X_i])\Big),$$

de lo que se deduce el resultado (tomando T'_1, T'_2 como $T'_{\alpha,\gamma}, T'_{\beta,\gamma}$ para algunos (α, β, γ) que dan como mucho el valor promedio). □

Aplicando el Lema 4.8 con variables aleatorias cualesquiera (T_1, T_2, T_3) tales que $T_1 + T_2 + T_3 = 0$ vale identicamente, y aplicando (4.15) con $X'_1 = T'_1$, $X'_2 = T'_2$, deducimos que

$$k \le \delta + \frac{\eta}{3}\Big(\delta + \sum_{i=1}^{2}\sum_{j=1}^{3}(d[X_i^0;T_j] - d[X_i^0;X_i])\Big).$$

En particular podemos aplicar esto con

$$T_1 = (U|S = s), \qquad T_2 = (V|S = s), \qquad T_3 = (W|S = s)$$

para s en el rango de S (que es una elección válida por (4.38)) y luego promediamos sobre s con pesos $p_S(s)$, para obtener

$$k \le \tilde{\delta} + \frac{\eta}{3}\Big(\tilde{\delta} + \sum_{i=1}^{2}\sum_{A\in\{U,V,W\}}\big(d[X_i^0;A|S] - d[X_i^0;X_i]\big)\Big), \tag{4.42}$$

donde

$$\tilde{\delta} \coloneqq \mathbf{I}\left[U \,:\, V|S\right] + \mathbf{I}\left[V \,:\, W|S\right] + \mathbf{I}\left[W \,:\, U|S\right].$$

Juntando esto con (4.36) y (4.37), concluimos que

$$\begin{aligned}k &\le \Big(1 + \frac{\eta}{3}\Big)\Big(6\eta k - \frac{1-5\eta}{1-\eta}(2\eta k - I_1)\Big) + \frac{\eta}{3}\Big((6-3\eta)k + 3(2\eta k - I_1)\Big) \\ &= (8\eta + \eta^2)k - \left(\frac{1-5\eta}{1-\eta}\Big(1 + \frac{\eta}{3}\Big) - \eta\right)(2\eta k - I_1) \\ &\le (8\eta + \eta^2)k\end{aligned}$$

ya que la cantidad $2\eta k - I_1$ es no negativa (por (4.7)), y su coeficiente en la expresión de arriba es no positivo dado que $\eta(2\eta + 17) \le 3$, que es ciertamente

el caso por nuestra elección $\eta = \frac{1}{9}$ (y en realidad para cualquier $\eta \leq \frac{1}{6}$). Más aún, para $\eta = \frac{1}{9}$ tenemos $8\eta + \eta^2 < 1$. Se sigue que $k = 0$, que era justo lo que buscabamos demostrar en nuestro intento de demostrar la Proposición 4.2 de forma contrarecíproca, cosa que hicimos al asumir la desigualdad (4.15). La demostración de la Proposición 4.2, y así de la Conjetura de Marton en el caso $\mathbf{F}_2$, queda completa.

Capítulo 5

El caso general

Vamos en este capítulo a demostrar la Conjetura de Marton en el caso de grupos de torsión m, que es en esencia muy similar al caso de $\mathbf{F}_2$, pero con la fundamental diferencia de que, para en el *final del juego* poder concluir, debemos reordenar el argumento de manera tal de armarnos ciertas variables aleatorias más complejas que sumen 0. En el caso de 2-torsión hacíamos esto trabajando con ciertas permutaciones de las sumas de 3 variables independientes, y para poder generar algo similar, vamos ahora a tener que redefinir muchas nociones con las que ya habiamos trabajado, para trabajar con más variables a la vez. La principal fuente de referencia para este capítulo es el artículo "Marton's Conjecture in abelian groups with bounded torsion" [7] de Gowers, Green, Manners y Tao, si bien no aclararemos esta referencia en cada resultado enunciado.

En concreto, lo primero que definiremos es una noción de **multidistancia**, pero sobre eso hablaremos bien en la próxima sección. Por ahora, nos contentamos con enunciar y discutir el resultado a probar:

Teorema 5.1. *Sea G un grupo abeliano con torsión m para algún $m \geq 2$. Supongamos que $A \subseteq G$ es un conjunto finito no vacío tal que $|A+A| \leq K|A|$. Luego, A puede cubrirse por como mucho $(2K)^{O(m^3)}$ coclases de algún subgrupo $H \leq G$ de tamaño al menos $|A|$. Más aún, H está contenido en $\ell A - \ell A$ para algún $\ell \ll (2 + m \log K)^{O(m^3 \log m)}$.*

Otra cosa distinta que incluimos en esta versión, aunque también se podía incluir en la del caso de $\mathbf{F}_2$, es la contención del subgrupo H dentro de $\ell A - \ell A$ para algún ℓ, cosa relacionada a la Conjetura de Bogolyubov (que enunciamos como Conjetura 1.10), o más en general a distintas conjeturas que generalizan la Conjetura de Marton. Entre ellas podemos mencionar:

Conjetura 5.2. *Supongamos que $A \subset \mathbf{F}_2^n$ es un conjunto con un complemento aditivo de tamaño K (es decir, existe un conjunto S de tamaño $|S| = K$ tal que $A + S = \mathbf{F}_2^n$). Entonces $2A$ contiene alguna coclase de algún subespacio de codimensión $O_K(1)$.*

Esta es una versión cualitativa, ya que no se especifica la dependencia en K, pero si la codimensión fuese controlada por $O(\log K)$, esto implicaría la Conjetura de Bogolyubov.

Otra conjetura todavía más fuerte es:

Conjetura 5.3. *Supongamos que $\mathbf{F}_2^n$ está particionada en $A_1, \dots, A_K$. Entonces $2A_i$ contiene una coclase de un subespacio de codimensión $O_K(1)$ para algún i.*

Una de las diferencias más palpables entre la Conjetura de Marton y estas conjeturas es que la de Marton se encarga de cubrir con subespacios, mientras que estas otras de incluir subespacios.

A grandes rasgos, la demostración del caso de torsión m arbitraria no tiene ideas muy importantes propias, sino que es una adaptación técnica de la del caso de torsión 2, pero recomendamos al lector prestarle atención a estos intentos por incluir subespacios adentro de los conjuntos, ya que es lo único verdaderamente nuevo.

Como antes, obtendremos el Teorema 5.1 a través de una versión entrópica de este:

Teorema 5.4. *Supongamos que G es un grupo abeliano de torsión m. Supongamos que X, Y son variables aleatorias G-valuadas. Luego existe un subgrupo $H \leq G$ tal que*

$$d[X; U_H], d[Y; U_H] \ll m^3 d[X; Y].$$

Más aún, si X, Y toma valores en un conjunto simétrico $S \subseteq G$ que contiene al origen, entonces H puede tomarse contenido en ℓS para algún $\ell \ll (2 + md[X; Y])^{O(m^3 \log m)}$.

La deducción de que un resultado implica al otro será discutido en la Sección 5.5.2.

5.1 Plan de la demostración

Para describir el argumento, vamos a introducir la noción de **multidistancia**. Esta es una idea útil introducida en el artículo [7], la principal referencia de este capítulo, que no hacia falta en $\mathbf{F}_2$. Para esto, es conveniente antes introducir la siguiente notación: si I es un conjunto finito de indices, entonces X_I denota una tupla $(X_i)_{i \in I}$ de variables aleatorias. Usualmente todas las variables en tal tupla van a ser G-valuadas para algún grupo abeliano G.

Definición 5.5. *Sea G un grupo abeliano y sea X_I una tupla finita no vacia de variables aleatorias G-valuadas. Definimos entonces*

$$\mathrm{D}[X_I] := \mathbf{H}\big[\textstyle\sum_{i \in I} \tilde{X}_i\big] - \dfrac{1}{|I|} \displaystyle\sum_{i \in I} \mathbf{H}[\tilde{X}_i],$$

donde las $\tilde{X}_i$ son copias independientes de las X_i.

De (2.15) vemos que $\mathbf{H}\big[\sum_{i\in I}\tilde{X}_i\big] \geq \mathbf{H}[X_i]$ para todo $i \in I$, y luego promediando podemos concluir que la multidistancia es siempre no negativa. Es también claramente invariante respecto a las permutaciones de las X_i. Remarcamos que, en el caso $I = \{1,2\}$, $\mathrm{D}[X_{\{1,2\}}]$ es igual a $d[X_1; -X_2]$ (y luego, si G es un espacio vectorial de característica 2, es lo mismo que $d[X_1; X_2]$). Esta observación explica por que se usa el término multidistancia; de todas formas, uno no debería tomar esta terminología con demasiada seriedad.

Vamos a deducir el Teorema 5.4 a partir del siguiente enunciado sobre la multidistancia:

Proposición 5.6. *Sea G un grupo abeliano de torsión m, y sea $I = \{1, 2, \dots, m\}$. Si X_I es una tupla de variables aleatorias G-valuadas entonces existe un subespacio $H \leq G$ tal que*

$$\sum_{i\in I} d[X_i; U_H] \ll m^3\, \mathrm{D}[X_I].$$

Más aún, si todas las X_i toman valores en algún conjunto simétrico $S \subseteq G$ que contiene el origen, entonces H puede tomarse para ser contenido en ℓS para algún $\ell \ll (2 + \mathrm{D}[X_I])^{O(m^3 \log m)}$.

La deducción del Teorema 5.4 a partir de esto no es muy complicada, la haremos en la próxima sección.

Demostraremos la Proposición 5.6 mediante una especie de inducción sobre la multidistancia, al igual que antes. El resultado técnico que impulsa este argumento es la siguiente proposición:

Proposición 5.7. *Sea G un grupo abeliano con torsión m. Definimos $\eta := c/m^3$ para una constante absoluta $c > 0$ suficientemente pequeña, y establecemos $I := \{1, 2, \dots, m\}$. Si X_I es un conjunto de variables aleatorias con valores en G tal que $\mathrm{D}[X_I] > 0$, entonces existe otro conjunto de variables aleatorias X'_I con valores en G que satisface la disminución de multidistancia*

$$\mathrm{D}[X'_I] \leq (1-\eta)\,\mathrm{D}[X_I] - \eta \sum_{i\in I} d[X_i; X'_i]. \tag{5.1}$$

Además, si todas las X_i toman valores en un conjunto simétrico $S \subseteq G$ que contiene el origen, entonces los X'_i pueden elegirse de modo que tomen valores en $m^3 S$.

La demostración de esta proposición constituye la mayor parte del capítulo, al igual que en el caso de 2-torsión, ya que es el paso inductivo de nuestro argumento. También necesitaremos el siguiente resultado sobre valores muy pequeños de la multidistancia, que forma el caso base de la inducción:

Proposición 5.8. *Sea G un grupo abeliano finito. Supongamos que I es un conjunto de índices de tamaño $m \geq 2$. Supongamos que X_I es un conjunto de variables aleatorias con valores en G tal que $\mathrm{D}[X_I] < c_0$ para una constante*

absoluta $c_0 > 0$ suficientemente pequeña. Entonces, existe un subgrupo $H \leq G$ tal que

$$\sum_{i \in I} d[X_i; U_H] \ll m \, \mathrm{D}[X_I].$$

Además, si todas las X_i toman valores en un conjunto simétrico $S \subseteq G$ que contiene el origen, entonces podemos tomar $H \subseteq 6S$.

Esto es una consecuencia directa de la Proposición 2.14, y dejamos la demostración para la Sección 5.5.1, solo por completitud. Para concluir esta linea de ideas, mostramos cómo el resultado principal se deduce de las Proposiciones 5.7 y 5.8.

Demostración de la Proposición 5.6 asumiendo la Proposición 5.7 y la Proposición 5.8. Aplicamos la Proposición 5.7 de manera iterativa, obteniendo para cada $t \geq 0$ un conjunto de variables aleatorias $X_I^{(t)}$ con soporte en $m^{3t}S$, con $X_I^{(0)} = X_I$ (es decir, $X_i^{(0)} = X_i$ para $i \in I$), y

$$\mathrm{D}\big[X_I^{(t+1)}\big] \leq (1-\eta)\, \mathrm{D}\big[X_I^{(t)}\big] - \eta \sum_{i \in I} d\big[X_i^{(t)}; X_i^{(t+1)}\big], \tag{5.2}$$

lo que en particular implica que $\mathrm{D}\big[X_I^{(t+1)}\big] \leq (1-\eta)\, \mathrm{D}\big[X_I^{(t)}\big]$. Definimos $k := \mathrm{D}[X_I]$. Mediante una inducción sencilla, se sigue que

$$\mathrm{D}\big[X_I^{(t)}\big] \leq (1-\eta)^t k. \tag{5.3}$$

Aplicamos esta iteración hasta que t alcanza el valor

$$s := \lfloor C m^3 \log(2+k) \rfloor, \tag{5.4}$$

donde C es una constante absoluta suficientemente grande.

Por (5.3), si C es lo suficientemente grande, tendremos $\mathrm{D}\big[X_I^{(s)}\big] < c_0$, donde c_0 es la constante en la Proposición 5.8. A partir de dicha proposición, se deduce que existe un subgrupo $H \leq G$ tal que

$$\sum_{i \in I} d\big[X_i^{(s)}; U_H\big] \ll m \, \mathrm{D}\big[X_I^{(s)}\big].$$

Además, podemos tomar $H \subseteq \ell S$ para algún $\ell \leq 6m^{3s} \ll (2+k)^{O(Cm^3 \log m)}$.

A partir de varias aplicaciones de la desigualdad triangular y (5.2), obtenemos

$$\begin{aligned}
\sum_{i \in I} d[X_i; U_H] &\leq \sum_{i \in I} d\big[X_i^{(s)}; U_H\big] + \sum_{t=0}^{s-1} \sum_{i \in I} d\big[X_i^{(t)}; X_i^{(t+1)}\big] \\
&\ll m \, \mathrm{D}[X_I^{(s)}] + \sum_{t=0}^{s-1} \frac{1}{\eta} \Big(\mathrm{D}\big[X_I^{(t+1)}\big] - (1-\eta)\, \mathrm{D}[X_I^{(t)}] \Big) \\
&\ll (m + \eta^{-1})\, \mathrm{D}\big[X_I^{(s)}\big] + \sum_{t=1}^{s-1} \mathrm{D}\big[X_I^{(t)}\big].
\end{aligned}$$

Por (5.3), deducimos

$$\sum_{i\in I} d[X_i; U_H] \ll \big((m+\eta^{-1})(1-\eta)^s + \eta^{-1}\big)k \ll m^3 k.$$

Esto concluye la demostración. □

Un detalle importante de esta demostración, en relación con las conjeturas antes mencionadas, es la relación que se preserva entre ℓ y el k, cosa de la cual nos gustaría deshacernos para probar los problemas abiertos. Esta demostración se encarga de hacer la sucesión de variables de la inducción de una forma más artesanal que en la versión de $\mathbf{F}_2$, sin usar argumentos de compacidad, principalmente para poder tener algún control sobre este s que se escoge.

5.2 Relacionando la distancia de Ruzsa y la multidistancia

Desarrollamos en lo siguiente algunas propiedades básicas de la multidistancia, comenzando por relacionar esta noción con la noción más estándar de distancia de Ruzsa entrópica. Esta sección se encarga más que nada de sentar resultados elementales que vamos a necesitar más adelante.

Lema 5.9. *Sea G un grupo abeliano, sea I un conjunto de índices de tamaño $m \geq 2$, y sea X_I un conjunto de variables aleatorias con valores en G. Entonces, se cumple lo siguiente:*

(i) $\sum_{\substack{j,k\in I\\ j\neq k}} d[X_j; -X_k] \leq m(m-1)\,\mathrm{D}[X_I]$;

(ii) $\sum_{j\in I} d[X_j; X_j] \leq 2m\,\mathrm{D}[X_I]$;

(iii) *Si todas las $(X_i)_{i\in I}$ tienen la misma distribución, entonces $\mathrm{D}[X_I] \leq md[X_j; X_j]$ para cualquier $j \in I$.*

Demostración. Sin pérdida de generalidad, podemos suponer que los X_i son independientes entre sí.

Siguiendo desigualdades ya trabajadas de entropía (por ejemplo (2.14)), se tiene que para cualesquiera $j, k \in I$ distintos,

$$\mathbf{H}[X_j + X_k] \leq \mathbf{H}\big[\textstyle\sum_{i\in I} X_i\big],$$

y, por lo tanto, usando la definición de la distancia de Ruzsa,

$$d[X_j; -X_k] \leq \mathbf{H}\big[\textstyle\sum_{i\in I} X_i\big] - \frac{1}{2}\mathbf{H}[X_j] - \frac{1}{2}\mathbf{H}[X_k].$$

Sumando esta desigualdad sobre todos los pares (j, k), con $j \neq k$, obtenemos (i).

Usando la desigualdad triangular,

$$d[X_j; X_j] \leq 2d[X_j; -X_k],$$

y aplicando esto a cada sumando en (i), se obtiene (ii) (tras dividir por $m-1$).

Para demostrar (iii), aplicamos la Proposición 2.6 con X siendo una copia independiente de $-X_j$, e $Y_1, \dots, Y_n$ un reordenamiento de $(X_i)_{i\in I}$, obteniendo

$$\mathbf{H}\big[-X + \textstyle\sum_{i\in I} X_i\big] - \mathbf{H}[X] \le md[X;X].$$

Como

$$\mathbf{H}\big[-X + \textstyle\sum_{i\in I} X_i\big] \ge \mathbf{H}\big[\sum_{i\in I} X_i\big]$$

por (2.14), esto implica (iii). □

Ya estamos en condiciones de reducir el Teorema 5.4 a la Proposición 5.6.

Demostración del Teorema 5.4 asumiendo la Proposición 5.6. Primero, afirmamos que basta demostrar el Teorema 5.4 en el caso $X = Y$ (a costa de empeorar las constantes implícitas por un factor un poco mayor que 2). En efecto, para X, Y generales, por la desigualdad triangular tenemos $d[X;X] \le 2d[X;Y]$, por lo que (suponiendo el caso $X = Y$ del Teorema 5.4) existe un subgrupo H con

$$d[X;U_H] \ll m^3 d[X;Y].$$

Entonces también se tiene

$$d[Y;U_H] \le d[X;U_H] + d[X;Y] \ll m^3 d[X;Y].$$

Esto prueba la afirmación.

Supongamos ahora que $X = Y$. Tomemos $I = \{1, 2, \dots, m\}$ y definamos $X_i = X$ para todo $i \in I$. Por el Lema 5.9(iii), se tiene

$$\mathrm{D}[X_I] \le md[X;X].$$

Por lo tanto, aplicando la Proposición 5.6, podemos encontrar un subgrupo H de G tal que $H \subseteq \ell S$ para algún

$$\ell \ll (2 + \mathrm{D}[X_I])^{O(m^3 \log m)} \le (2 + md[X;X])^{O(m^3 \log m)}$$

y además

$$\sum_{i\in I} d[X;U_H] \ll m^3 \,\mathrm{D}[X_I] \le m^4 d[X;X].$$

Dado que el lado izquierdo es simplemente $md[X;U_H]$, se sigue el resultado. □

5.3 La regla de la cadena de la multidistancia

En esta sección establecemos una desigualdad clave para el comportamiento de la multidistancia bajo morfismos, junto con algunas de sus consecuencias. El lema fundamentalque daremos, Lema 5.10, es una "regla de la cadena" para la multidistancia, análoga a la regla de la cadena (2.5) para la entropía de

Shannon, así como al lema del fibrado para la distancia de Ruzsa entrópica, la Proposición 4.3.

Ya hemos mencionado nuestra convención de escribir $X_I = (X_i)_{i\in I}$ para un conjunto de variables aleatorias indexado por un conjunto finito I. Si estas variables aleatorias toman valores en G, es conveniente introducir dos notaciones adicionales.

Primero, si $\pi\colon G \to H$ es un morfismo, escribimos

$$\pi(X_I) \coloneqq (\pi(X_i))_{i\in I}.$$

Segundo, si $Y_I = (Y_i)_{i\in I}$ es otro conjunto de variables aleatorias con valores en G, escribimos

$$X_I + Y_I \coloneqq (X_i + Y_i)_{i\in I}.$$

También necesitaremos introducir la noción de **multidistancia condicional**. Si X_I y Y_I son conjuntos de variables aleatorias, con los X_i tomando valores en G, definimos

$$\mathrm{D}[X_I|Y_I] \coloneqq \mathbf{H}\Big[\textstyle\sum_{i\in I}\tilde{X}_i\big|(\tilde{Y}_j)_{j\in I}\Big] - \dfrac{1}{|I|}\displaystyle\sum_{i\in I}\mathbf{H}[\tilde{X}_i|\tilde{Y}_i], \tag{5.5}$$

donde $(\tilde{X}_i, \tilde{Y}_i)$ para $i \in I$ son copias independientes de (X_i, Y_i) para $i \in I$ (pero aquí **no** asumimos que X_i es independiente de Y_i, ni que $\tilde{X}_i$ es independiente de $\tilde{Y}_i$).

De manera equivalente, se tiene

$$\mathrm{D}[X_I|Y_I] = \sum_{(y_i)_{i\in I}} \Big(\prod_{i\in I} p_{Y_i}(y_i)\Big)\, \mathrm{D}\big[(X_i \,|\, Y_i = y_i)_{i\in I}\big], \tag{5.6}$$

donde cada y_i varía en el soporte de p_{Y_i} para $i \in I$.

Presentamos entonces la regla de la cadena:

Lema 5.10. *Sea $\pi\colon G \to H$ un morfismo de grupos abelianos y sea X_I un conjunto de variables aleatorias con valores en G que son independientes entre sí. Entonces, $\mathrm{D}[X_I]$ se expresa como*

$$\mathrm{D}\big[X_I|\pi(X_I)\big] + \mathrm{D}\big[\pi(X_I)\big] + \mathbf{I}\Big[\textstyle\sum_{i\in I}X_i : \pi(X_I) \;\big|\; \pi\big(\sum_{i\in I}X_i\big)\Big]. \tag{5.7}$$

Demostración. Para abreviar la notación en esta demostración, escribimos

$$S := \sum_{i\in I} X_i.$$

Expandiendo la definición (2.12) de la información mutua $\mathbf{I}[S : \pi(X_I)|\pi(S)]$ y usando el hecho de que $\pi(S)$ está determinado tanto por S como por $\pi(X_I)$, obtenemos

$$\mathbf{I}[S : \pi(X_I)|\pi(S)] = \mathbf{H}[S] + \mathbf{H}[\pi(X_I)] - \mathbf{H}[S, \pi(X_I)] - \mathbf{H}[\pi(S)].$$

Aplicando la regla de la cadena (2.5), el lado derecho es igual a

$$\mathbf{H}[S] - \mathbf{H}[S|\pi(X_I)] - \mathbf{H}[\pi(S)].$$

Por lo tanto,

$$\mathbf{H}[S] = \mathbf{H}[S|\pi(X_I)] + \mathbf{H}[\pi(S)] + \mathbf{I}[S : \pi(X_I)|\pi(S)]. \tag{5.8}$$

A partir de otra aplicación de la regla de la cadena (2.5), se obtiene

$$\mathbf{H}[X_i] = \mathbf{H}[X_i \,|\, \pi(X_i)] + \mathbf{H}[\pi(X_i)] \tag{5.9}$$

para todo $i \in I$. Promediando (5.9) sobre i y restando esto de (5.8), obtenemos el resultado como consecuencia de la definición de multidistancia. □

Necesitaremos concatenar varias aplicaciones de la regla de la cadena para la multidistancia, por lo que será conveniente establecer una versión condicional de esta regla, como sigue.

Lema 5.11. *Sea $\pi\colon G \to H$ un morfismo de grupos abelianos. Sea I un conjunto finito de índices y sea X_I un conjunto de variables aleatorias con valores en G. Sea Y_I otro conjunto de variables aleatorias (no necesariamente con valores en G). Supongamos que los pares (X_i, Y_i) son independientes entre sí (pero X_i no necesita ser independiente de Y_i). Entonces, se cumple*

$$\begin{aligned}\mathrm{D}[X_I|Y_I] = \mathrm{D}\big[X_I \,|\, \pi(X_I), Y_I\big] + \mathrm{D}\big[\pi(X_I) \,|\, Y_I\big] \\ + \mathbf{I}\Big[\textstyle\sum_{i\in I} X_i : \pi(X_I) \;\Big|\; \pi\big(\sum_{i\in I} X_i\big), Y_I\Big].\end{aligned} \tag{5.10}$$

En efecto, para cada y_i en el soporte de p_{Y_i}, podemos aplicar la regla de la cadena recién establecida con X_i reemplazado por la variable aleatoria condicionada $(X_i|Y_i = y_i)$, y la ecuación (5.10) se sigue promediando (5.7) sobre y_i usando los pesos p_{Y_i}.

Podemos iterar el lema anterior de la siguiente manera.

Lema 5.12. *Sea m un entero positivo. Supongamos que tenemos una sucesión*

$$G_m \to G_{m-1} \to \cdots \to G_1 \to G_0 = \{0\} \tag{5.11}$$

de morfismos entre grupos abelianos $G_0, \ldots, G_m$, y para cada $d = 0, \ldots, m$, sea $\pi_d\colon G_m \to G_d$ el morfismo de G_m a G_d inducido por composición en esta sucesión (por ejemplo, π_m es la identidad y π_0 es el morfismo trivial). Sea I un conjunto finito de índices y sea $X_I = (X_i)_{i\in I}$ un conjunto de variables aleatorias con valores en G_m que son independientes entre sí. Entonces, se cumple

$$\begin{aligned}\mathrm{D}[X_I] = \sum_{d=1}^{m} \mathrm{D}\big[\pi_d(X_I) \,|\, \pi_{d-1}(X_I)\big] \\ + \sum_{d=1}^{m-1} \mathbf{I}\Big[\textstyle\sum_i X_i : \pi_d(X_I) \;\Big|\; \pi_d\big(\sum_i X_i\big), \pi_{d-1}(X_I)\Big].\end{aligned} \tag{5.12}$$

En particular, dado que todos los términos $\mathbf{I}[-]$ *son no negativos, tenemos*

$$\begin{aligned}\mathrm{D}[X_I] \geq \sum_{d=1}^{m} \mathrm{D}\big[\pi_d(X_I) | \pi_{d-1}(X_I)\big] \\ + \mathbf{I}\Big[\textstyle\sum_i X_i : \pi_1(X_I) \;\Big|\; \pi_1\big(\sum_i X_i\big)\Big].\end{aligned} \tag{5.13}$$

Demostración. Por el Lema 5.11 (tomando $Y_I = \pi_{d-1}(X_I)$ y $\pi = \pi_d$ en ese lema, y notando que $\pi_d(X_I)$ determina Y_I), se tiene

$$\begin{aligned}\mathrm{D}\big[X_I \,|\, \pi_{d-1}(X_I)\big] = \mathrm{D}\big[X_I \,|\, \pi_d(X_I)\big] + \mathrm{D}\big[\pi_d(X_I) \,|\, \pi_{d-1}(X_I)\big] \\ + \mathbf{I}\Big[\textstyle\sum_{i\in I} X_i : \pi_d(X_I) \;\Big|\; \pi_d\big(\sum_{i\in I} X_i\big), \pi_{d-1}(X_I)\Big]\end{aligned}$$

para $d = 1, \ldots, m-1$. El resultado se sigue al aplicar una suma telescópica, notando que $\mathrm{D}[X_I|\pi_0(X_I)] = \mathrm{D}[X_I]$ y que $\pi_m(X_I) = X_I$. □

En nuestra aplicación necesitaremos el siguiente caso particular del lema anterior.

Corolario 5.13. *Sea* G *un grupo abeliano y sea* $m \geq 2$. *Supongamos que* $X_{i,j}$, *para* $1 \leq i, j \leq m$, *son variables aleatorias independientes con valores en* G. *Entonces, se cumple*

$$\begin{aligned}&\mathbf{I}\Big[\big(\textstyle\sum_{i=1}^m X_{i,j}\big)_{j=1}^m : \big(\sum_{j=1}^m X_{i,j}\big)_{i=1}^m \;\Big|\; \sum_{i=1}^m \sum_{j=1}^m X_{i,j}\Big] \\ &\quad\leq \sum_{j=1}^{m-1} \Big(\mathrm{D}\big[(X_{i,j})_{i=1}^m\big] - \mathrm{D}\Big[(X_{i,j})_{i=1}^m \;\Big|\; (X_{i,j} + \cdots + X_{i,m})_{i=1}^m\Big]\Big) \\ &\qquad\qquad + \mathrm{D}\big[(X_{i,m})_{i=1}^m\big] - \mathrm{D}\big[\big(\textstyle\sum_{j=1}^m X_{i,j}\big)_{i=1}^m\big],\end{aligned}$$

donde todas las multidistancias aquí involucran el conjunto de índices $\{1, \ldots, m\}$.

Demostración. En el Lema 5.12 tomamos $G_d := G^d$ con las aplicaciones $\pi_d \colon G^m \to G^d$ para $d = 1, \ldots, m$ definidas por

$$\pi_d(x_1, \ldots, x_m) := (x_1, \ldots, x_{d-1}, x_d + \cdots + x_m)$$

con $\pi_0 = 0$. Dado que $\pi_{d-1}(x)$ se puede obtener de $\pi_d(x)$ aplicando un morfismo, obtenemos una sucesión de la forma (5.11).

Ahora aplicamos el Lema 5.12 con $I = \{1, \ldots, m\}$ y $X_i := (X_{i,j})_{j=1}^m$. Usando la independencia conjunta, obtenemos

$$\mathrm{D}[X_I] = \sum_{j=1}^{m} \mathrm{D}\big[(X_{i,j})_{i\in I}\big].$$

Por otro lado, para $1 \leq j \leq m-1$, notamos que una vez que $\pi_j(X_i)$ está fijado, $\pi_{j+1}(X_i)$ está completamente determinado por $X_{i,j}$ y viceversa, por lo que

$$\mathrm{D}\big[\pi_{j+1}(X_I) \mid \pi_j(X_I)\big] = \mathrm{D}\big[(X_{i,j})_{i\in I} \mid \pi_j(X_I)\big].$$

Dado que los $X_{i,j}$ son independientes entre sí, podemos simplificar aún más:

$$\mathrm{D}\big[(X_{i,j})_{i\in I} \mid \pi_j(X_I)\big] = \mathrm{D}\big[(X_{i,j})_{i\in I} \mid (X_{i,j} + \cdots + X_{i,m})_{i\in I}\big].$$

Sustituyendo todo esto en la conclusión de el Lema 5.12, obtenemos

$$\begin{aligned}\sum_{j=1}^{m} \mathrm{D}[(X_{i,j})_{i\in I}] \geq \sum_{j=1}^{m-1} \mathrm{D}\big[(X_{i,j})_{i\in I} \mid (X_{i,j} + \cdots + X_{i,m})_{i\in I}\big] \\ + \mathrm{D}\Big[\big(\textstyle\sum_{j=1}^{m} X_{i,j}\big)_{i\in I}\Big] \\ + \mathbf{I}\Big[\big(\textstyle\sum_{i=1}^{m} X_{i,j}\big)_{j=1}^{m} : \big(\textstyle\sum_{j=1}^{m} X_{i,j}\big)_{i=1}^{m} \mid \textstyle\sum_{i=1}^{m}\sum_{j=1}^{m} X_{i,j}\Big]\end{aligned}$$

y la afirmación se sigue reorganizando los términos. □

5.4 El argumento principal

Comenzamos ahora una discusión preliminar a la demostración de la Proposición 5.7. Al igual que en el caso de $\mathbf{F}_2$, es conveniente trabajar con la formulación contrarrecíproca, lo que nos permite aprovechar la notación de la entropía condicional.

Supongamos, en lo que sigue, que tenemos variables aleatorias G-valuadas X_i, para $i \in I = \{1, \ldots, m\}$, que toman valores en un conjunto $S \subseteq G$, con

$$k := \mathrm{D}[X_I], \tag{5.14}$$

y supongamos que no podemos achicarlas como en la Proposición 5.7, es decir, que

$$\mathrm{D}[X'_I] \geq (1-\eta)k - \eta \sum_{i\in I} d[X_i; X'_i] \tag{5.15}$$

para cualquier conjunto X'_I de variables aleatorias G-valuadas que toman valores en $m^3 S$. El objetivo es demostrar que $k = 0$, lo cual es equivalente a probar la Proposición 5.7.

Ahora observamos que (5.15) implica una versión condicionada de sí misma,

$$\mathrm{D}[X'_I|Y_I] \geq (1-\eta)k - \eta \sum_{i\in I} d[X_i; X'_i|Y_i] \tag{5.16}$$

para cualquier conjunto X'_I de variables aleatorias G-valuadas que toman valores en $m^3 S$ y para cualquier conjunto Y_I de variables aleatorias. Para obtener (5.16) a partir de (5.15), simplemente reemplazamos X'_i por $(X'_i|Y_i = y_i)$ y luego sumamos ponderando con $\prod_{i\in I} p_{Y_i}(y_i)$.

La desigualdad (5.16) puede reescribirse de la siguiente manera conveniente:

$$k - \mathrm{D}[X'_I|Y_I] \leq \eta\Big(k + \sum_{i\in I} d[X_i; X'_i|Y_i]\Big).$$

También resulta útil notar que

$$k - \mathrm{D}[X_I'|Y_I] \leq \eta\left(k + \sum_{i\in I} d[X_{\sigma(i)}; X_i'|Y_i]\right) \tag{5.17}$$

para cualquier permutación $\sigma : I \to I$, ya que la multidistancia es invariante bajo permutaciones.

5.4.1 Acotando la información mutua

El primer paso clave del argumento es observar que (5.17) se combina con el Corolario 5.13 para obtener la siguiente desigualdad:

Proposición 5.14. *Sea G un grupo abeliano. Sea $m \geq 2$, y supongamos que $X_{i,j}$, para $1 \leq i, j \leq m$, son variables aleatorias G-valuadas independientes entre sí, tales que para cada $j = 1, \ldots, m$, las variables aleatorias $(X_{i,j})_{i=1}^m$ tienen la misma distribución que alguna permutación de las variables aleatorias $X_I = (X_i)_{i=1}^m$. Definimos*

$$\mathcal{I} := \mathbf{I}\Big[\big(\textstyle\sum_{i=1}^m X_{i,j}\big)_{j=1}^m : \big(\sum_{j=1}^m X_{i,j}\big)_{i=1}^m \mid \sum_{i=1}^m \sum_{j=1}^m X_{i,j}\Big].$$

Entonces, suponiendo que (5.14) y (5.15) se cumplen, tenemos

$$\mathcal{I} \leq 2\eta m\left(k + \sum_{i=1}^m d[X_i; X_i]\right) \leq 2m(2m+1)\eta k. \tag{5.18}$$

Para cada $j \in \{1, \ldots, m\}$, llamamos al conjunto $(X_{i,j})_{i=1}^m$ una **columna**, y para cada $i \in \{1, \ldots, m\}$, llamamos al conjunto $(X_{i,j})_{j=1}^m$ una **fila**. Por hipótesis, cada columna es una permutación de $X_I = (X_i)_{i=1}^m$.

Demostración. El Corolario 5.13 establece que

$$\mathcal{I} \leq \sum_{j=1}^{m-1} A_j + B, \tag{5.19}$$

donde

$$A_j := \mathrm{D}[(X_{i,j})_{i=1}^m] - \mathrm{D}\big[(X_{i,j})_{i=1}^m \mid (X_{i,j} + \cdots + X_{i,m})_{i=1}^m\big]$$

y

$$B := \mathrm{D}\big[(X_{i,m})_{i=1}^m\big] - \mathrm{D}\Big[\big(\textstyle\sum_{j=1}^m X_{i,j}\big)_{i=1}^m\Big].$$

Consideremos primero los términos A_j, para un j fijo en $\{1, \ldots, m-1\}$. Por simetría bajo permutaciones de la multidistancia y nuestra hipótesis sobre las columnas, se tiene que

$$\mathrm{D}[(X_{i,j})_{i=1}^m] = \mathrm{D}[(X_i)_{i=1}^m] = \mathrm{D}[X_I] = k.$$

Sea $\sigma = \sigma_j \colon I \to I$ una permutación tal que $X_{i,j} = X_{\sigma(i)}$, y definimos $X'_i \coloneqq X_{i,j}$ y $Y_i \coloneqq X_{i,j} + \cdots + X_{i,m}$. Notemos que todas estas variables aleatorias toman valores en mS. Aplicando (5.17), concluimos que

$$A_j \leq \eta\Big(k + \sum_{i=1}^{m} d\big[X_{i,j}; X_{i,j} | X_{i,j} + \cdots + X_{i,m}\big]\Big). \tag{5.20}$$

Ahora consideramos B. Por simetría de permutación en la última columna,

$$\mathrm{D}\big[(X_{i,m})_{i=1}^{m}\big] = \mathrm{D}[X_I] = k.$$

Para cada $i \in I$, definimos la suma de la fila i como

$$V_i \coloneqq \sum_{j=1}^{m} X_{i,j}.$$

Si aplicamos (5.17) nuevamente, ahora con $X_{\sigma(i)} = X_{i,m}$, $X'_i \coloneqq V_i$, y tomando Y_i trivial (es decir, usando la versión no condicionada (5.15)), obtenemos

$$B \leq \eta\Big(k + \sum_{i=1}^{m} d[X_{i,m}; V_i]\Big). \tag{5.21}$$

Ahora queda acotar las distancias que aparecen en (5.20) y (5.21) utilizando técnicas del cálculo de Ruzsa.

Para $1 \leq j \leq m-1$ y $1 \leq i \leq m$, por (4.24), que se obtiene al aplicar el Lema 4.6, tenemos

$$\begin{aligned} d\big[X_{i,j}; X_{i,j} | X_{i,j} + \cdots + X_{i,m}\big] &\leq d[X_{i,j}; X_{i,j}] \\ &\quad + \tfrac{1}{2}\big(\mathbf{H}[X_{i,j} + \cdots + X_{i,m}] - \mathbf{H}[X_{i,j+1} + \cdots + X_{i,m}]\big). \end{aligned}$$

Para cada i, sumando sobre $j = 1, \ldots, m-1$ se obtiene

$$\begin{aligned} &\sum_{j=1}^{m-1} d\big[X_{i,j}; X_{i,j} | X_{i,j} + \cdots + X_{i,m}\big] \\ &\qquad \leq \sum_{j=1}^{m-1} d[X_{i,j}; X_{i,j}] + \frac{1}{2}\big(\mathbf{H}[V_i] - \mathbf{H}[X_{i,m}]\big). \end{aligned} \tag{5.22}$$

Por otro lado, por el Lema 2.25(i) (ya que $X_{i,m}$ aparece en la suma V_i), tenemos

$$d[X_{i,m}; V_i] \leq d[X_{i,m}; X_{i,m}] + \frac{1}{2}\big(\mathbf{H}[V_i] - \mathbf{H}[X_{i,m}]\big). \tag{5.23}$$

Combinando (5.19), (5.20) y (5.21) con (5.22) y (5.23) (sumando las últimas dos sobre i), obtenemos

$$\begin{aligned} \frac{1}{\eta}\mathcal{I} &\leq mk + \sum_{i,j=1}^{m} d[X_{i,j}; X_{i,j}] + \sum_{i=1}^{m} (\mathbf{H}[V_i] - \mathbf{H}[X_{i,m}]) \\ &= mk + m\sum_{i=1}^{m} d[X_i; X_i] + \sum_{i=1}^{m} \mathbf{H}[V_i] - \sum_{i=1}^{m} \mathbf{H}[X_i]. \end{aligned} \tag{5.24}$$

Por el Lema 2.25(ii) (con f asignando a cada j el índice j' tal que $X_{i,j}$ es una copia de $X_{j'}$), obtenemos la cota

$$\mathbf{H}[V_i] \leq \mathbf{H}\Big[\sum_{j=1}^m X_j\Big] + \sum_{j=1}^m d[X_{i,j}; X_{i,j}].$$

Finalmente, sumando sobre i y usando que $\mathrm{D}[X_I] = k$, obtenemos

$$\begin{aligned}\sum_{i=1}^m \mathbf{H}[V_i] - \sum_{i=1}^m \mathbf{H}[X_i] &\leq \sum_{i,j=1}^m d[X_{i,j}; X_{i,j}] + mk \\ &= m\sum_{i=1}^m d[X_i; X_i] + mk,\end{aligned}$$

donde en el segundo paso usamos la hipótesis de permutación. Combinando esto con (5.24), obtenemos la primera desigualdad en (5.18). La segunda se sigue inmediatamente del Lema 5.9(ii). □

5.4.2 Final del juego

Ahora definimos un conjunto de variables aleatorias independientes $(Y_{i,j})_{i,j\in\mathbb{Z}/m\mathbb{Z}}$ de la siguiente manera: por un leve abuso de notación, identificamos $\mathbb{Z}/m\mathbb{Z}$ con $\{1,\dots,m\}$ de la manera obvia, y definimos $Y_{i,j}$ como una copia independiente de X_i.

Nos interesaremos en las siguientes variables aleatorias derivadas de $(Y_{i,j})_{i,j\in\mathbb{Z}/m\mathbb{Z}}$:

$$W \coloneqq \sum_{i,j\in\mathbb{Z}/m\mathbb{Z}} Y_{i,j}$$

y

$$Z_1 \coloneqq \sum_{i,j\in\mathbb{Z}/m\mathbb{Z}} iY_{i,j}, \quad Z_2 \coloneqq \sum_{i,j\in\mathbb{Z}/m\mathbb{Z}} jY_{i,j}, \quad Z_3 \coloneqq \sum_{i,j\in\mathbb{Z}/m\mathbb{Z}} (-i-j)Y_{i,j}.$$

La suma $(-i-j)$ se realiza en $\mathbb{Z}/m\mathbb{Z}$. Notemos que, dado que asumimos que G tiene torsión m, está bien definida la multiplicación de elementos de G por elementos de $\mathbb{Z}/m\mathbb{Z}$. Además, observamos que los Z_i están todos soportados en m^3S.

Dado que estas cantidades aparecerán frecuentemente, también definimos para $i,j,r \in \mathbb{Z}/m\mathbb{Z}$ las variables

$$P_i \coloneqq \sum_{j\in\mathbb{Z}/m\mathbb{Z}} Y_{i,j}, \quad Q_j \coloneqq \sum_{i\in\mathbb{Z}/m\mathbb{Z}} Y_{i,j}, \quad R_r \coloneqq \sum_{\substack{i,j\in\mathbb{Z}/m\mathbb{Z}\\ i+j=-r}} Y_{i,j}. \tag{5.25}$$

Notamos las siguientes identidades:

$$Z_1 = \sum_{i\in\mathbb{Z}/m\mathbb{Z}} iP_i, \qquad Z_2 = \sum_{j\in\mathbb{Z}/m\mathbb{Z}} jQ_j, \qquad Z_3 = \sum_{r\in\mathbb{Z}/m\mathbb{Z}} rR_r. \tag{5.26}$$

Hay varios hechos clave a destacar en esta situación. Uno de ellos es la afirmación fácilmente verificable de que

$$Z_1 + Z_2 + Z_3 = 0 \tag{5.27}$$

se cumple identicamente. Otro hecho importante es la siguiente proposición, que en términos generales establece que Z_1, Z_2, Z_3 son "casi" independientes de a pares, condicionados a W.

Proposición 5.15. *Suponiendo que* (5.14) *y* (5.15) *se cumplen, se tiene que*

$$\mathbf{I}[Z_1 : Z_2 \,|\, W],\ \mathbf{I}[Z_2 : Z_3 \,|\, W],\ \mathbf{I}[Z_1 : Z_3 \,|\, W] \leq t$$

donde

$$t \coloneqq 2m(2m+1)\eta k. \tag{5.28}$$

Demostración. Analizamos estas variables aplicando la Proposición 5.14 de distintas formas.

En la primera aplicación, tomamos $X_{i,j} = Y_{i,j}$. Notemos que cada columna $(X_{i,j})_{i=1}^m$ es efectivamente una permutación de $X_1, \ldots, X_m$; de hecho, la permutación trivial. Además, para cada $i \in \mathbb{Z}/m\mathbb{Z}$, la suma de la fila es

$$\sum_{j=1}^{m} X_{i,j} = \sum_{j \in \mathbb{Z}/m\mathbb{Z}} Y_{i,j} = P_i$$

y para cada $j \in \mathbb{Z}/m\mathbb{Z}$, la suma de la columna es

$$\sum_{i=1}^{m} X_{i,j} = \sum_{i \in \mathbb{Z}/m\mathbb{Z}} Y_{i,j} = Q_j.$$

Finalmente, notemos que

$$\sum_{i,j=1}^{m} X_{i,j} = W.$$

La conclusión de la Proposición 5.14 establece entonces que

$$\mathbf{I}\Big[(P_i)_{i \in \mathbb{Z}/m\mathbb{Z}} : (Q_j)_{j \in \mathbb{Z}/m\mathbb{Z}} \,\Big|\, W\Big] \leq t,$$

con t dado en (5.28). Dado que Z_1 es una función de $(P_i)_{i \in \mathbb{Z}/m\mathbb{Z}}$ por (5.26), y de manera similar Z_2 es una función de $(Q_j)_{j \in \mathbb{Z}/m\mathbb{Z}}$, se sigue inmediatamente de la desigualdad de procesamiento de datos, Lema 2.3, que

$$\mathbf{I}[Z_1 : Z_2 \,|\, W] \leq t.$$

En la segunda aplicación de la Proposición 5.14, tomamos en cambio $X'_{i,j} = Y_{i-j,j}$. De nuevo, para cada j fijo, el conjunto $(X'_{i,j})_{i=1}^m$ es una permutación de $X_1, \ldots, X_m$. En este caso, las sumas de las filas para $i \in \{1, \ldots, m\}$ son

$$\sum_{j=1}^{m} X'_{i,j} = \sum_{j \in \mathbb{Z}/m\mathbb{Z}} Y_{i-j,j} = R_{-i}.$$

De manera similar, las sumas de las columnas para $j \in \{1, \dots, m\}$ son

$$\sum_{i=1}^{m} X'_{i,j} = \sum_{i \in \mathbb{Z}/m\mathbb{Z}} Y_{i-j,j} = Q_j.$$

Como antes,

$$\sum_{i,j=1}^{m} X'_{i,j} = W.$$

Por lo tanto, usando (5.26) y la desigualdad de procesamiento de datos nuevamente, la conclusión de la Proposición 5.14 nos dice que

$$\mathbf{I}[Z_3 : Z_2 \,|\, W] \le \mathbf{I}\Big[(R_i)_{i \in \mathbb{Z}/m\mathbb{Z}} : (Q_j)_{j \in \mathbb{Z}/m\mathbb{Z}} \,\Big|\, W\Big] \le t.$$

En la tercera aplicación, tomamos $X''_{i,j} = Y_{i,j-i}$. Las sumas de las columnas y filas son, respectivamente,

$$\sum_{j=1}^{m} X''_{i,j} = \sum_{j \in \mathbb{Z}/m\mathbb{Z}} Y_{i,j-i} = P_i$$

y

$$\sum_{i=1}^{m} X''_{i,j} = \sum_{i \in \mathbb{Z}/m\mathbb{Z}} Y_{i,j-i} = R_{-j}.$$

Por lo tanto, la Proposición 5.14 y la desigualdad de procesamiento de datos nos dan

$$\mathbf{I}[Z_1 : Z_3 \,|\, W] \le \mathbf{I}\Big[(P_i)_{i \in \mathbb{Z}/m\mathbb{Z}} : (R_j)_{j \in \mathbb{Z}/m\mathbb{Z}} \,\Big|\, W\Big] \le t,$$

lo que completa la demostración. □

En este punto, estamos en una situación muy similar a la de $\mathbf{F}_2$: al condicionar sobre un valor típico $W = w$, las variables aleatorias Z_1, Z_2, Z_3 son "casi independientes dos a dos", en el sentido de que la información mutua entre cualquiera de ellas es pequeña. Usando el Teorema entrópico de Balog–Szemerédi–Gowers, podemos encontrar variables relacionadas con una duplicación muy pequeña y luego usarlas como candidatos en (5.16) para obtener una contradicción (a menos que $k = 0$).

Para poner esto en práctica, primero recopilamos algunas estimaciones sobre las variables W y Z_2.

Lema 5.16. *Para W y Z_2 definidos anteriormente, se cumplen las siguientes desigualdades:*

(i) $\mathbf{H}[W] \le (2m-1)k + \frac{1}{m}\sum_{i=1}^{m} \mathbf{H}[X_i]$;

(ii) $\mathbf{H}[Z_2] \le \big(28(m-1)\log_2 m\big)k + \frac{1}{m}\sum_{i=1}^{m} \mathbf{H}[X_i]$;

(iii) $\mathbf{I}[W : Z_2] \le 2(m-1)k$;

(iv) $\sum_{i=1}^m d[X_i; Z_2|W] \leq 15(m^2 \log_2 m)k$.

Demostración. Sin pérdida de generalidad, podemos suponer que $X_1, \ldots, X_m$ son independientes. Escribimos $S = \sum_{i=1}^m X_i$. Notemos que, para cada $j \in \mathbb{Z}/m\mathbb{Z}$, la suma Q_j de (5.25) tiene la misma distribución que S. Por la Proposición 2.6, tenemos

$$\begin{aligned}\mathbf{H}[W] = \mathbf{H}\Big[\textstyle\sum_{j\in\mathbb{Z}/m\mathbb{Z}} Q_j\Big] &\leq \mathbf{H}[S] + \sum_{j=2}^{m} (\mathbf{H}[Q_1 + Q_j] - \mathbf{H}[S]) \\ &= \mathbf{H}[S] + (m-1)d[S; -S].\end{aligned}$$

Por el Lema 2.25(iii), se tiene

$$d[S; -S] \leq 2k \tag{5.29}$$

y, por lo tanto,

$$\mathbf{H}[W] \leq 2k(m-1) + \mathbf{H}[S].$$

Dado que, por la definición de multidistancia,

$$\mathbf{H}[S] = k + \frac{1}{m}\sum_{i=1}^{m} \mathbf{H}[X_i], \tag{5.30}$$

esto implica (i).

Pasamos a (ii). Observamos que

$$\mathbf{H}[Z_2] = \mathbf{H}\Big[\textstyle\sum_{j\in\mathbb{Z}/m\mathbb{Z}} jQ_j\Big].$$

Aplicando nuevamente la Proposición 2.6, obtenemos

$$\mathbf{H}[Z_2] \leq \sum_{i=2}^{m-1} \mathbf{H}[Q_1 + iQ_i] - (m-2)\mathbf{H}[S].$$

Usando el Lema 2.26(ii) y (5.29), deducimos

$$\begin{aligned}\mathbf{H}[Z_2] &\leq \mathbf{H}[S] + (10\lfloor \log_2 m \rfloor + 4)(m-2)d[S; -S] \\ &\leq \mathbf{H}[S] + (20\lfloor \log_2 m \rfloor + 8)(m-2)k.\end{aligned}$$

Aplicando (5.30) (y usando estimaciones brutas para simplificar términos en m), se obtiene (ii). Cabe mencionar que, si en lugar del Lema 2.26(ii) utilizáramos la versión más débil pero más sencilla del Lema 2.26(i), los resultados principales seguirían obteniendo cotas similares.

Para (iii), tenemos naturalmente que

$$\mathbf{I}[W : Z_2] = \mathbf{H}[W] - \mathbf{H}[W|Z_2].$$

Dado que $Z_2 = \sum_{j=1}^{m-1} jQ_j$ y $W = \sum_{j=1}^{m} Q_j$, se sigue que

$$\mathbf{H}[W|Z_2] \geq \mathbf{H}[W \,|\, Q_1, \ldots, Q_{m-1}] = \mathbf{H}[Q_m] = \mathbf{H}[S].$$

Por lo tanto, usando (i) y (5.30), obtenemos

$$\mathbf{I}[W : Z_2] \leq \mathbf{H}[W] - \mathbf{H}[S] \leq 2(m-1)k,$$

lo que prueba (iii).

Por último, consideramos (iv). Para cada $i \in \{1, \dots, m\}$, usando el Lema 2.25(i) (observando que la suma Z_2 contiene X_i como sumando), tenemos

$$d[X_i; Z_2] \leq d[X_i; X_i] + \tfrac{1}{2}(\mathbf{H}[Z_2] - \mathbf{H}[X_i]). \tag{5.31}$$

Además, aplicando la desigualdad (4.5), obtenemos

$$d[X_i; Z_2|W] \leq d[X_i; Z_2] + \tfrac{1}{2}\mathbf{I}[W : Z_2].$$

Combinando con (5.31) y (iii), se sigue que

$$d[X_i; Z_2|W] \leq d[X_i; X_i] + \tfrac{1}{2}(\mathbf{H}[Z_2] - \mathbf{H}[X_i]) + (m-1)k.$$

Sumando sobre i y aplicando (ii), obtenemos

$$\sum_{i=1}^{m} d[X_i; Z_2|W] \leq \sum_{i=1}^{m} d[X_i; X_i] + \frac{m}{2}(28(m-1)\log_2 m)k + m(m-1)k.$$

Finalmente, aplicando el Lema 5.9(ii) (y usando estimaciones brutas para los términos en m), se obtiene (iv). □

A continuación, demostramos el siguiente resultado utilizando la versión entrópica de Balog–Szemerédi–Gowers, que es prácticamente igual al Lema 4.8, pero con algunos cambios para la multidistancia:

Lema 5.17. *Sea G un grupo abeliano y sea (T_1, T_2, T_3) una variable aleatoria con valores en G^3 tal que $T_1 + T_2 + T_3 = 0$ se cumple identicamente. Definimos*

$$\delta := \mathbf{I}[T_1 : T_2] + \mathbf{I}[T_1 : T_3] + \mathbf{I}[T_2 : T_3].$$

Sea $Y_1, \dots, Y_n$ otro conjunto de variables aleatorias con valores en G y sea $\alpha > 0$ una constante. Entonces, existe una variable aleatoria U, con soporte contenido en el de T_2, tal que

$$d[U; U] + \alpha \sum_{i=1}^{n} d[Y_i; U] \leq \Big(2 + \frac{\alpha n}{2}\Big)\delta + \alpha \sum_{i=1}^{n} d[Y_i; T_2]. \tag{5.32}$$

Demostración. Aplicamos el Teorema entrópico de Balog–Szemerédi–Gowers (2.7) con $X = T_1$ y $Y = T_2$. Dado que $T_1 + T_2 = -T_3$, obtenemos

$$\begin{aligned} &\sum_z p_{T_3}(z) d[T_1 \,|\, T_3 = z; T_2 \,|\, T_3 = z] \\ &\qquad \leq 3\mathbf{I}[T_1 : T_2] + 2\mathbf{H}[T_3] - \mathbf{H}[T_1] - \mathbf{H}[T_2] \\ &\qquad = \mathbf{I}[T_1 : T_2] + \mathbf{I}[T_1 : T_3] + \mathbf{I}[T_2 : T_3] = \delta, \end{aligned} \tag{5.33}$$

donde la última línea se obtiene observando que

$$\mathbf{H}[T_1, T_2] = \mathbf{H}[T_1, T_3] = \mathbf{H}[T_2, T_3] = \mathbf{H}[T_1, T_2, T_3]$$

ya que cualesquiera dos de T_1, T_2, T_3 determinan al tercero, y al descomponer la definición de información mutua.

Por (5.33) y la desigualdad triangular,

$$\sum_z p_{T_3}(z) d[T_2 \,|\, T_3 = z; T_2 \,|\, T_3 = z] \leq 2\delta.$$

Además, por el Lema 4.5, para cada Y_i, se tiene que

$$\begin{aligned}&\sum_z p_{T_3}(z) d[Y_i; T_2 \,|\, T_3 = z] \\ &\qquad = d[Y_i; T_2 \,|\, T_3] \leq d[Y_i; T_2] + \frac{1}{2}\mathbf{I}[T_2 : T_3] \leq d[Y_i; T_2] + \frac{\delta}{2}.\end{aligned}$$

Por lo tanto,

$$\begin{aligned}&\sum_z p_{T_3}(z) \left(d[T_2 \,|\, T_3 = z; T_2 \,|\, T_3 = z] + \alpha \sum_{i=1}^{n} d[Y_i; T_2 \,|\, T_3 = z] \right) \\ &\qquad \leq \left(2 + \frac{\alpha n}{2}\right)\delta + \alpha \sum_{i=1}^{n} d[Y_i; T_2].\end{aligned}$$

El resultado se sigue al tomar $U = (T_2 \,|\, T_3 = z)$ para algún z tal que la cantidad dentro del paréntesis en el lado izquierdo sea como máximo el valor medio ponderado. □

Finalmente, podemos reunir todos estos elementos. Para cada valor $W = w$, aplicamos el Lema 5.17 con

$$T_1 = (Z_1 \,|\, W = w), \quad T_2 = (Z_2 \,|\, W = w), \quad T_3 = (Z_3 \,|\, W = w)$$

tomando $Y_i = X_i$ y $\alpha = \eta/m$, donde η es la constante en el enunciado de la Proposición 5.7. Definimos

$$\delta_w \coloneqq \mathbf{I}[T_1 : T_2] + \mathbf{I}[T_1 : T_3] + \mathbf{I}[T_2 : T_3]$$

para esta elección, y notamos que

$$\begin{aligned}\delta_* \coloneqq \sum_w p_W(w)\delta_w &= \mathbf{I}[Z_1 : Z_2 \,|\, W] + \mathbf{I}[Z_1 : Z_3 \,|\, W] + \mathbf{I}[Z_2 : Z_3 \,|\, W] \\ &\leq 6m(2m+1)\eta k \ll \eta m^2 k \qquad (5.34)\end{aligned}$$

por la Proposición 5.15.

Sea U_w la variable aleatoria cuya existencia está garantizada por el Lema 5.17, de modo que (5.32) nos da

$$d[U_w; U_w] \leq \Big(2 + \frac{\alpha m}{2}\Big)\delta_w + \alpha \sum_{i=1}^{m} \big(d[X_i; T_2] - d[X_i; U_w]\big). \tag{5.35}$$

Notemos que, por el Lema 5.17, el soporte de U_w está contenido en el de Z_2. Recordando que $Z_2 = \sum_{i,j \in \mathbb{Z}/m\mathbb{Z}} jY_{i,j}$, donde los $Y_{i,j}$ son copias de X_i y, por lo tanto, están soportados en S, concluimos que el soporte de U_w está contenido en $m^3 S$.

Sea $(U_w)_I$ la tupla que consiste en la misma variable U_w repetida m veces. Por el Lema 5.9(iii),

$$\mathrm{D}[(U_w)_I] \leq m d[U_w; U_w]. \tag{5.36}$$

Por otro lado, aplicando nuestra hipótesis (5.15), obtenemos

$$\mathrm{D}[(U_w)_I] \geq (1 - \eta)k - \eta \sum_{i=1}^{m} d[X_i; U_w]. \tag{5.37}$$

Combinando (5.35), (5.36) y (5.37) y promediando sobre w (con peso $p_W(w)$), y recordando el valor de $\alpha = \eta/m$, obtenemos

$$m\Big(2 + \frac{\eta}{2}\Big)\delta_* + \eta \sum_{i=1}^{m} d[X_i; Z_2 | W] \geq (1 - \eta)k$$

ya que los términos $d[X_i; U_w]$ se cancelan por nuestra elección de α. Aplicando la desigualdad del Lema 5.16(iv) y (5.34), y usando el hecho de que $2 + \frac{\eta}{2} < 3$, obtenemos

$$m^3 \eta k + \eta(m^2 \log_2 m)k \gg k.$$

Recordemos que, en el enunciado de la Proposición 5.7, η se tomó como c/m^3. Si la constante c es lo suficientemente pequeña, esto lleva a una contradicción a menos que $k = 0$. Esto es lo que necesitábamos demostrar, y la prueba de la Proposición 5.7 está completa.

5.5 Últimos detalles

5.5.1 Caso base

En esta sección simplemente vamos a demostrar el caso del 99% de la conjetura de Marton cuando la torsión no es par. En particular, tenemos lo siguiente:

Demostración de la Proposición 5.8. Definamos $\varepsilon := \mathrm{D}[X_I]$. Luego por el Lema 5.9(i) y tomando promedios, podemos encontrar $i \in I$ tal que

$$\sum_{k \neq i} d[X_i; -X_k] \leq (m - 1)\varepsilon, \tag{5.38}$$

y al tomar promedio otra vez encontramos $j \neq i$ que cumple que

$$d[X_i; -X_j] \leq \varepsilon.$$

Para un ε suficientemente pequeño, podemos aplicar la Proposición 2.14 para concluir que existe un grupo finito H de G tal que $d[X_i; U_H] \leq 12\varepsilon$. Siguiendo la demostración de ese resultado con atención, podemos ver que H es de la forma $H = S' - S'$ (cosa que, a su vez, provenía del Teorema 2.15 de Freiman), donde todos los elementos y de S' tienen divergencia de Kullback–Leibler $D_{\mathrm{KL}}(y - X_j || X_i - X_j)$ finita; esto implica que $S' \subseteq 3S$ y luego que $H \subseteq 6S$. Como $d[-Y; U_H] = d[Y; -U_H] = d[Y; U_H]$ para cualquier variable aleatoria Y, (5.38) y la desigualdad triangular nos permiten ver que

$$\sum_{k \in I} d[X_k; U_H] \leq (13m - 1)\varepsilon,$$

como queríamos. □

5.5.2 De la versión entrópica a la combinatoria

Esta sección final del capítulo tiene como único objetivo demostrar que el Teorema 5.4 implica al Teorema 5.1, cosa que se hace prácticamente de la misma forma que en el caso de $\mathbf{F}_2$, pero teniendo un especial cuidado con observar donde es que está contenido el subgrupo H.

Demostración de que el Teorema 5.4 implica al Teorema 5.1. Sean m, A, K como en el Teorema 5.1. Podemos asumir, trasladando de ser necesario, que A contiene 0. La cota sobre la constante de duplicación $|A + A| \leq K|A|$ y una cota simple nos dan que

$$d[U_A; -U_A] \leq \log K.$$

Podemos entonces, por el Teorema 5.4 (con S tomado como $A \cup -A$, que claramente es simétrico y contiene al 0), hallar un subgrupo H de G, tal que $H \subseteq \ell S \subseteq \ell A - \ell A$ para algún $\ell \ll (2 + m \log K)^{O(m^3 \log m)}$, tal que

$$d[U_A; U_H] \ll m^3 \log K.$$

Ahora, como $\frac{1}{2}|\mathbf{H}[X] - \mathbf{H}[Y]| \leq d[X; Y]$ concluimos que

$$\log |H| = \log |A| + O(m^3 \log K)$$

y que

$$\mathbf{H}[U_A - U_H] = \log |H| + O(m^3 \log K).$$

Debe existir un punto $x_0 \in \mathbf{F}_p^n$ tal que

$$p_{U_A - U_H}(x_0) \geq e^{-\mathbf{H}[U_A - U_H]} \geq K^{-O(m^3)}/|H|,$$

o en otras palabras

$$|A \cap (H + x_0)| \geq K^{-O(m^3)}|H|,$$

debido a lo observado en (2.32). Aplicando el lema de cubrimiento de Ruzsa, Lema 3.1, podemos cubrir a A por como mucho $K^{O(m^3)}$ trasladados de

$$(A \cap (H + x_0) - A \cap (H + x_0)) \subseteq H.$$

Subdividiendo H en coclases H' de un subgrupo con cardinal entre $|A|/m$ y $|A|$ de ser necesario, se concluye el resultado. □

Capítulo 6

PFR débil en los enteros

El objetivo de esta sección será demostrar el Teorema 1.5, que daba un resultado similar a la conjetura polinomial de Freiman-Ruzsa, pero en los números enteros. La fuente principal de este capítulo es "Sumsets and entropy revisited" [16]. Por comodidad del lector, enunciaremos otra vez el resultado:

Teorema 6.1. *Sea A un subconjunto finito de $\mathbf{Z}^D$ para algún $d \in \mathbf{N}$, y supongamos que $|A+A| \leq K|A|$ para algún K. Entonces existe un subconjunto $A' \subseteq A$, $|A'| \geq K^{-C_1/2}|A|$, con $dim(A') \leq C_2 log(K)$, para ciertas constantes absolutas $C_1, C_2 \geq 0$.*

Recordamos que la noción de dimensión que aquí usamos es la de dimensión afín, donde entendemos a

$$\dim(A)$$

como a la dimensión del espacio vectorial real generado por $A - A$.

En [16, Sección 7] se explica que este resultado es una versión mejorada de un resultado previo de Manners, donde él ya había desarrollado ciertas ideas para pasar de $\mathbf{F}_2^D$ a $\mathbf{Z}^D$, pero que se mejoran con las nociones de entropía.

A lo largo de este capítulo probaremos varios lemas, concluyendo con la demostración del Teorema 6.6, del cual deduciremos inmediatamente el Teorema 6.1.

Motivemos primero un poco el primer lema, donde se ven la principal idea que vincula $\mathbf{Z}$ con $\mathbf{F}_2$. Tomando $A \subset \mathbf{Z}^D$ como un conjunto con una constante de duplicación (combinatoria) K pequeña, se sigue que el dilatado $2 \cdot A$, que está contenido en $A+A$, es conmensurable (quizás con una relación polinomial con K) con respecto a A. Al proyectar módulo 2, uno esperaría que la proyección $\pi(A)$ sea conmensurable con la proyección $\pi(2 \cdot A) = \{0\}$. En el contexto entrópico, el Lema 2.10, que decía que

$$d[X, 2Y] \leq 5d[X, Y],$$

actuará en lugar de la observación de que $2 \cdot A$ está contenido en $A + A$. Formalizamos esta relación entre los conjuntos con pequeña constante de duplicación y la proyección módulo 2 en el siguiente lema:

Lema 6.2. *Sean X, Y variables aleatorias con valores en $\mathbf{Z}^D$ para algún $D \geq 0$. Denotemos por $\phi : \mathbf{Z}^D \to \mathbf{F}_2^D$ el morfismo natural. Entonces, se cumple que*

$$\mathbf{H}(\phi(X)), \mathbf{H}(\phi(Y)) \leq 10d[X, Y].$$

Demostración. Por la Proposición 4.3 y el Lema 2.10, se tiene que

$$d[\phi(X), \phi(2Y)] \leq d[X, 2Y] \leq 5d[X, Y]. \tag{6.1}$$

Sin embargo, $\phi(2Y)$ es idénticamente cero, por lo que

$$d[\phi(X), \phi(2Y)] = d[\phi(X), 0] = \frac{1}{2}\mathbf{H}(\phi(X)).$$

Combinando esto con (6.1), obtenemos la cota deseada para $\mathbf{H}(\phi(X))$. La misma cota para $\mathbf{H}(\phi(Y))$ se obtiene de manera análoga. □

El argumento principal para justificar este teorema es una inducción, en la cual necesitaremos fabricarnos varios conjuntos, tanto para el caso base como para la inducción en sí. Los Lemas 6.3 y 6.4 serán los principales resultados técnicos que vamos a necesitar para construir estos conjuntos, los enunciamos y demostramos a continuación.

El primer lema nos da una forma cuantitativa de escoger ciertas fibras densas en un conjunto:

Lema 6.3. *Sea $\phi\colon G \to H$ un morfismo y sean $A, B \subseteq G$ subconjuntos finitos. Para $x, y \in H$, definimos*

$$A_x = A \cap \phi^{-1}(x), \quad B_y = B \cap \phi^{-1}(y)$$

como las fibras de A y B, y escribimos $\alpha_x := \frac{|A_x|}{|A|}$ y $\beta_y := \frac{|B_y|}{|B|}$. Definimos $k = d[U_A, U_B]$, $\overline{k} = d[\phi(U_A), \phi(U_B)]$ y $M = \mathbf{H}(\phi(U_A)) + \mathbf{H}(\phi(U_B))$. Entonces, existen $x, y \in H$ tales que A_x, B_y son no vacíos y cumplen

$$\overline{k} \log \frac{1}{\alpha_x \beta_y} \leq M\big(k - d[U_{A_x}, U_{B_y}]\big). \tag{6.2}$$

Demostración. Observemos primero que las variables aleatorias $(U_A \mid \phi(U_A) = x)$ y $(U_B \mid \phi(U_B) = y)$ tienen la misma distribución que U_{A_x} y U_{B_y}, respectivamente, es decir, corresponden a distribuciones uniformes en las fibras.

Aplicando el lema del fibrado, Lema 4.3, se sigue que

$$\sum_{x,y \in H} \alpha_x \beta_y d[U_{A_x}, U_{B_y}] \leq k - \overline{k}, \tag{6.3}$$

ya que el término de la derecha es lo mismo que $d[U_A, U_B | \phi(U_A), \phi(U_B)]$ Por

otra parte, la definición de M puede reescribirse como definición

$$\begin{aligned} M &= \sum_{x\in H} \alpha_x \log\frac{1}{\alpha_x} + \sum_{y\in H} \beta_y \log\frac{1}{\beta_y} \\ &= \sum_{x\in H} \alpha_x (\sum_{y\in H} \beta_y) \log\frac{1}{\alpha_x} + \sum_{y\in H} \beta_y (\sum_{x\in H} \alpha_x) \log\frac{1}{\beta_y} \qquad (6.4) \\ &= \sum_{x,y\in H} \alpha_x\beta_y (\log\frac{1}{\alpha_x} + \log\frac{1}{\beta_y}) \qquad (6.5) \\ &= \sum_{x,y\in H} \alpha_x\beta_y \log\frac{1}{\alpha_x\beta_y}, \qquad (6.6) \end{aligned}$$

y por lo tanto, combinando (6.3) con (6.4), obtenemos

$$\sum_{x,y\in H} \alpha_x\beta_y \Big(Md[U_{A_x}, U_{B_y}] + \overline{k} \log\frac{1}{\alpha_x\beta_y} \Big) \leq Mk.$$

Por el principio del palomar concluimos entonces que existe al menos una elección de x, y con $\alpha_x, \beta_y > 0$ tal que

$$Md[U_{A_x}, U_{B_y}] + \overline{k} \log\frac{1}{\alpha_x\beta_y} \leq Mk.$$

Reordenando términos, obtenemos la desigualdad (6.2). □

En los siguientes lemas la constante C con la que trabajamos es la constante que aparece del lado derecho de la desigualdad en la versión entrópica de la Conjetura de Marton 4.1. En este texto se vió que se puede tomar $C = 11$ en la versión entrópica de la Conjetura de Marton, pero en realidad, si se consiguieran mejores cotas estas se arrastrarían directamente a los siguientes resultados, por lo que preferimos dejar la constante como C.

Lema 6.4. *Supongamos que X e Y son variables aleatorias con valores en $\mathbf{F}_2^D$. Entonces, existe un subgrupo $H \leq \mathbf{F}_2^D$ tal que, denotando por $\pi\colon \mathbf{F}_2^D \to \mathbf{F}_2^D/H$ la proyección natural, se cumple que*

$$\log|H| \leq 2(\mathbf{H}(X) + \mathbf{H}(Y)) \qquad (6.7)$$

y

$$\mathbf{H}(\pi(X)) + \mathbf{H}(\pi(Y)) \leq 8Cd[\pi(X); \pi(Y)]. \qquad (6.8)$$

Este resultado (otra vez) lo demostraremos mediante la construcción inductiva de una sucesión de subespacios. Para proceder con la construcción, será útil tener el siguiente lema demostrado por separado:

Lema 6.5. *Sea $n \in \mathbf{N}$. Sean X, Y variables aleatorias con valores en $\mathbf{F}_2^n$. Suponemos que*

$$\mathbf{H}(X) + \mathbf{H}(Y) > 8Cd[X; Y]. \qquad (6.9)$$

Entonces, existe un subgrupo no trivial $H \leq \mathbf{F}_2^n$ *tal que*

$$\log |H| \leq \mathbf{H}(X) + \mathbf{H}(Y) \tag{6.10}$$

y (denotando por $\pi\colon \mathbf{F}_2^n \to \mathbf{F}_2^n/H$ *la proyección natural)*

$$\mathbf{H}(\pi(X)) + \mathbf{H}(\pi(Y)) \leq \frac{1}{2}\big(\mathbf{H}(X) + \mathbf{H}(Y)\big). \tag{6.11}$$

Demostración. Aplicando la versión entrópica de la Conjetura de Marton, Teorema 4.1, obtenemos un subgrupo H tal que

$$d[X; U_H], d[Y; U_H] \leq C d[X; Y].$$

Usando (2.34) y (6.9), se deduce que

$$\mathbf{H}(\pi(X)) + \mathbf{H}(\pi(Y)) \leq 4C d[X; Y] < \frac{1}{2}(\mathbf{H}(X) + \mathbf{H}(Y)),$$

lo que prueba (6.11).

Para demostrar (6.10), aplicamos (2.16), lo que nos da

$$\log |H| - \mathbf{H}(X) \leq 2d[X; U_H] \leq 2C d[X; Y],$$

y de manera similar para Y. Luego, usando (6.9), obtenemos

$$\log |H| \leq \frac{1}{2}(\mathbf{H}(X) + \mathbf{H}(Y)) + 2C d[X; Y] < \mathbf{H}(X) + \mathbf{H}(Y),$$

lo que da la cota deseada en (6.10).

Si H fuera trivial, tendríamos $\pi(X) = X$, $\pi(Y) = Y$, y por (6.11) se seguiría que $\mathbf{H}(X) + \mathbf{H}(Y) = 0$, lo que contradice la hipótesis (6.9). □

Demostración del Lema 6.4. Definimos iterativamente una sucesión de subgrupos $\{0\} = H_0 < H_1 < \cdots$ de $\mathbf{F}_2^D$. Denotamos por $\pi_i\colon \mathbf{F}_2^D \to \mathbf{F}_2^D/H_i$ el operador de proyección en el i-ésimo paso, y definimos $k_i := d[\phi_i(X), \phi_i(Y)]$. Vamos a parar la iteración en el paso i si se cumple

$$\mathbf{H}(\pi_i(X)) + \mathbf{H}(\pi_i(Y)) \leq 8C k_i. \tag{6.12}$$

Si esta condición no se cumple, aplicamos el Lema 6.5 a $\pi_i(X), \pi_i(Y)$, obteniendo un subgrupo no trivial $H_{i+1}/H_i \leq \mathbf{F}_2^D/H_i$ tal que

$$\log \frac{|H_{i+1}|}{|H_i|} \leq \mathbf{H}(\pi_i(X)) + \mathbf{H}(\pi_i(Y)) \tag{6.13}$$

y

$$\mathbf{H}(\pi_{i+1}(X)) + \mathbf{H}(\pi_{i+1}(Y)) \leq \frac{1}{2}\big(\mathbf{H}(\pi_i(X)) + \mathbf{H}(\pi_i(Y))\big). \tag{6.14}$$

Claramente, de la aplicación iterada de (6.14) obtenemos

$$\mathbf{H}(\pi_i(X)) + \mathbf{H}(\pi_i(Y)) \leq 2^{-i}(\mathbf{H}(X) + \mathbf{H}(Y)).$$

Luego, aplicando un argumento telescópico a (6.13), obtenemos

$$\log |H_i| \leq 2(\mathbf{H}(X) + \mathbf{H}(Y)). \tag{6.15}$$

Dado que los grupos H_i forman una sucesión estrictamente creciente, la iteración se detiene en algún paso i. En ese momento, se cumplen tanto (6.12) como (6.15), y por lo tanto, tomando $\pi = \pi_i$, se concluye la demostración del Lema 6.4. □

Con todos estos resultados en la mano ya estamos en condiciones de probar el principal teorema de esta sección, que es una versión bipartita de 1.5, y la implica directamente:

Teorema 6.6. *Sea $D \in \mathbf{N}$, y supongamos que $A, B \subseteq \mathbf{Z}^D$ son conjuntos finitos y no vacíos, y definimos $k := d[U_A, U_B]$. Entonces, existen subconjuntos no vacíos $A' \subseteq A$, $B' \subseteq B$ tales que*

$$\log \frac{|A|}{|A'|} + \log \frac{|B|}{|B'|} \leq C_1 k$$

y además cumplen que $\dim A', \dim B' \leq C_2 k$, con C_1, C_2 constante absolutas.

Demostración. Procederemos por inducción sobre $|A| + |B|$.

También podemos suponer que A, B no están contenidos en coclases de algún subgrupo propio de $\mathbf{Z}^D$, pues en caso contrario podemos reemplazar $\mathbf{Z}^D$ por dicho subgrupo.

Sea $\phi : \mathbf{Z}^D \to \mathbf{F}_2^D$ el morfismo natural. Por el Lema 6.2, tenemos

$$\mathbf{H}(\phi(U_A)), \mathbf{H}(\phi(U_B)) \leq 10k. \tag{6.16}$$

Aplicando el lema 6.4 a $\phi(U_A), \phi(U_B)$, encontramos un subgrupo $H \leq \mathbf{F}_2^D$ y la proyección asociada $\psi : \mathbf{F}_2^D \to \mathbf{F}_2^D/H$ tal que, denotando por

$$\tilde{\phi} = \psi \circ \phi : \mathbf{Z}^D \to \mathbf{F}_2^D/H,$$

la proyección compuesta natural, se cumple que

$$\log |H| \leq 2(\mathbf{H}(\phi(U_A)) + \mathbf{H}(\phi(U_B))) \leq 40k \tag{6.17}$$

y

$$\mathbf{H}(\tilde{\phi}(U_A)) + \mathbf{H}(\tilde{\phi}(U_B)) \leq 8Cd\big[\tilde{\phi}(U_A), \tilde{\phi}(U_B)\big]. \tag{6.18}$$

Ahora, combinando (6.16) y (2.8), también obtenemos

$$\mathbf{H}(\tilde{\phi}(U_A)) + \mathbf{H}(\tilde{\phi}(U_B)) \leq 20k. \tag{6.19}$$

Si H es todo $\mathbf{F}_2^D$, entonces de (6.17) (tomando $C_2 = 40/\log 2$) se deduce que $D \leq C_2 k$, y por lo tanto el resultado se cumple simplemente tomando $A' = A$, $B' = B$. Esto vendría a probar el caso base de nuestra inducción.

Supongamos, entonces, que H no es todo $\mathbf{F}_2^D$. Para $x, y \in \mathbf{F}_2^D/H$, definimos los conjuntos

$$A_x \coloneqq A \cap \tilde{\phi}^{-1}(x), \quad B_y \coloneqq B \cap \tilde{\phi}^{-1}(y),$$

que corresponden a las fibras de A y B sobre x, y respectivamente. Dado que asumimos que A, B no están contenidos en cosets de un subgrupo propio de $\mathbf{Z}^D$, podemos suponer que al menos uno de $\tilde{\phi}(A)$, $\tilde{\phi}(B)$ no es un conjunto unitario. Por lo tanto, se tiene

$$|A_x| + |B_y| < |A| + |B|$$

y $\mathbf{H}(\tilde{\phi}(U_A)) + \mathbf{H}(\tilde{\phi}(U_B)) > 0$, lo que implica que $d\big[\tilde{\phi}(U_A), \tilde{\phi}(U_B)\big] > 0$ por (6.18).

Aplicando el Lema 6.3, junto con (6.18), encontramos $x, y \in \mathbf{F}_2^D/H$ tales que

$$\log \frac{|A|}{|A_x|} + \log \frac{|B|}{|B_y|} \leq 8C\big(k - d[U_{A_x}, U_{B_y}]\big). \tag{6.20}$$

Definimos $k' = d[U_{A_x}, U_{B_y}]$. Por hipótesis inductiva sobre $|A_x|, |B_y|$, podemos encontrar subconjuntos $A' \subseteq A_x$ y $B' \subseteq B_y$ tales que $\dim A', \dim B' \leq C_2 k' \leq C_2 k$ y

$$\log \frac{|A_x|}{|A'|} + \log \frac{|B_y|}{|B'|} \leq C_1 k'.$$

Sumando esto a (6.20), obtenemos

$$\log \frac{|A|}{|A'|} + \log \frac{|B|}{|B'|} \leq C_1 k' + 8C(k - k'). \tag{6.21}$$

Si $C_1 \geq 8C$, el lado derecho de (6.21) es a lo sumo $C_1 k$, cerrando así la inducción. La demostración queda completa. □

Capítulo 7

Normas de Gowers

> "In a sense, the normed spaces just defined encapsulate all the information we need about the arithmetical properties of the functions we consider. In their definitions they bear some resemblance to Sobolev spaces. Although I cannot think of any potential applications, I still feel that it would be interesting to investigate them further."
>
> *Timothy Gowers, "A new proof of Szemerédi's theorem"* [6].

El objetivo de esta sección es dar una introducción a las normas de Gowers, objeto central en la Combinatoria Aditiva, para luego en Capítulo 8 demostrar las aplicaciones de la resolución de la Conjetura de Marton a estas. La principal fuente trabajada para este capítulo es el artículo "An inverse theorem for the Gowers $U^3(G)$ norm"[14], de Green y Tao, donde, además de probar un teorema inverso sobre las normas de Gowers, dan una introducción muy buena de ellas, que aquí presentamos con leves modificaciones.

Un teorema famoso y profundo de Szemerédi afirma que cualquier conjunto de enteros con densidad superior positiva contiene progresiones aritméticas de longitud arbitrariamente grande. Más precisamente:

Teorema 7.1 (Teorema de Szemerédi [6])**.** *Sea A un subconjunto* $\mathbf{Z}$ *cuya* ***densidad superior***

$$\limsup_{N\to\infty}(2N+1)^{-1}|A\cap[-N,N]|$$

es estrictamente positiva. Entonces, para todo $k\geq 1$, el conjunto A contiene infinitas progresiones aritméticas $\{a,a+r,\ldots,a+(k-1)r\}$, con $r\neq 0$, de longitud k.

El primer caso no trivial de este teorema es cuando $k=3$ fue tratado por Roth mediante un argumento basado en análisis de Fourier (ver [25]). El caso de valores más altos de k resultó más resistente a los métodos Fourier-analíticos, y la primera demostración completa de este teorema fue lograda por Szemerédi

usando métodos combinatorios. Posteriormente, Furstenberg introdujo una demostración basada en teoría ergódica (ver [4]). Por último, hace 25 años, Gowers [6] presentó una demostración de naturaleza tanto combinatoria como Fourier-analítica, sustancialmente más cercana en espíritu al argumento original de Roth que las demás pruebas.

La demostración completa del teorema de Szemerédi por Gowers en [6] es bastante extensa e involucra muchas ideas nuevas y profundas. Sin embargo, es posible descomponerla en una serie de pasos más simples y, todos salvo uno, bastante directos. En primer lugar, es fácil mostrar que para cualquier k fijo, el teorema 7.1 es equivalente a la siguiente versión:

Teorema 7.2 (Teorema de Szemerédi, segunda versión [6]). *Sean $\delta > 0$ y $k \geq 1$. Entonces existe un entero $N_0 = N_0(\delta, k)$ tal que, siempre que $N \geq N_0$ y $A \subseteq [1, N]$ sea tal que $|A|/|[1, N]| \geq \delta$, entonces A contiene al menos una progresión aritmética propia de longitud k.*

La siguiente observación, debida a Roth, es que se puede intentar demostrar este teorema realizando una inducción descendente sobre el parámetro de densidad δ (siendo trivial o vacuo el caso $\delta \geq 1$). En particular, para cualquier k fijo, el Teorema 7.2 es equivalente a la siguiente afirmación.

Teorema 7.3 (Ausencia de progresiones implica incremento de densidad). *Sean $\delta > 0$ y $k \geq 1$. Sea $N \geq 1$, y sea $A \subseteq [1, N]$ sin progresiones aritméticas propias de longitud k, tal que $|A|/|[1, N]| \geq \delta$. Entonces, si N es suficientemente grande dependiendo de k y δ, existe una progresión aritmética $P \subseteq [1, N]$ con $|P| \geq \omega(N, \delta)$ para alguna función $\omega(N, \delta)$ de N que tiende a infinito cuando $N \to \infty$ para cada δ fijo, tal que se tiene el incremento de densidad $|A \cap P|/|P| \geq \delta + c(\delta)$, donde $c(\delta) > 0$ es una función de δ que está acotada inferiormente por una constante positiva siempre que δ esté acotada inferiormente por una constante positiva.*

La deducción del Teorema 7.2 a partir del teorema 7.3 es un argumento inductivo directo.

El siguiente paso es un poco técnico, y consiste en pasar del intervalo $[1, N]$ a un grupo cíclico $\mathbf{Z}/N\mathbf{Z}$ para algún primo N. De hecho, utilizando el postulado de Bertrand (es decir, que siempre hay un primo entre X y $2X$) y un argumento simple de recubrimientos para dividir progresiones en $\mathbf{Z}/N\mathbf{Z}$ en progresiones dentro de $[1, N]$, se puede demostrar que el Teorema 7.3 es, para cada k fijo, equivalente (salvo por cambios menores en las cotas $\omega(N, \delta)$ y $c(\delta)$) a la siguiente afirmación:

Teorema 7.4 (Ausencia de progresiones implica incremento de densidad). *Sean $\delta > 0$ y $k \geq 1$. Sea $N \geq 1$ un número primo, y sea Q una progresión propia en $\mathbf{Z}/N\mathbf{Z}$ tal que $|Q| \geq c_0 N$ para algún $0 < c_0 \leq 1$. Sea $A \subseteq Q$ tal que $|A| \geq \delta N$, y tal que A no contiene progresiones aritméticas propias de longitud k. Entonces, si N es suficientemente grande dependiendo de k y δ, existe una progresión aritmética propia $P \subseteq \mathbf{Z}/N\mathbf{Z}$ con $|P| \geq \omega(N, \delta, c_0, k)$ para alguna*

función $\omega(N, \delta, c_0, k)$ de N que tiende a infinito cuando $N \to \infty$ para cada δ, c_0, k fijos, tal que se tiene el incremento de densidad

$$\frac{|A \cap P|}{|P|} \geq \frac{|A \cap Q|}{|Q|} + c(\delta, c_0, k),$$

donde $c(\delta, c_0, k) > 0$ está acotado inferiormente por una constante positiva siempre que δ, c_0 estén acotados inferiormente por constantes positivas y k esté fijado.

La deducción del Teorema 7.3 a partir del Teorema 7.4 no es difícil. Por supuesto, queda demostrar el Teorema 7.4, pero al menos ahora podemos concentrarnos en mirar solo lo que sucede en los grupos $\mathbf{Z}/N\mathbf{Z}$ con N primo. Esto fue logrado en el caso $k = 3$ por Roth utilizando métodos Fourier-analíticos. Para extender estos argumentos al caso de valores mayores de k, Gowers introdujo una colección de herramientas que forman parte de una teoría que podría denominarse "**análisis de Fourier de grado superior**", por razones que quedarán claras más adelante. En particular, para tratar el caso $k = 4$ se requirió "**análisis de Fourier cuadrático**".

Aunque el argumento original de Gowers tiene lugar en un grupo cíclico $\mathbf{Z}/N\mathbf{Z}$ de orden primo, nosotros trabajaremos en el entorno apenas más general de espacios vectoriales sobre cuerpos finitos.

Tomemonos un momento para definir algunos de los objetos con los que vamos a trabajar:

Definición 7.5. *Si $f : G \to H$ es una función de un grupo aditivo a otro, y $h \in H$, definimos el operador de traslación T^h aplicado a f mediante la fórmula $T^h f(x) := f(x + h)$, y el operador de diferencia $h \cdot \nabla := T^h - 1$ aplicado a f mediante la fórmula $(h \cdot \nabla) f(x) := f(x+h) - f(x)$. Extendemos estas definiciones a funciones de varias variables poniendo como subíndice la variable a la que se aplica el operador. Por ejemplo, si $f(x, y)$ es una función de dos variables, definimos $T^h_x f(x, y) := f(x + h, y)$ y $h \cdot \nabla_x f(x, y) := f(x + h, y) - f(x, y)$, si x, h varían dentro de un grupo aditivo G, y de manera similar para la variable y.*

Es útil pensar a este operador de diferencia como una suerte de derivada, pero con diferencias discretas. A continuación introducimos una forma multilineal $\Lambda_k(f_1, \ldots, f_k)$ que es útil para contar progresiones aritméticas. Aquí, resulta conveniente adoptar la notación de esperanza (típica en los artículos de Tao y sus colaboradores), lo cual permite ocultar algunos factores de normalización molestos como $1/N$ en nuestros argumentos. Así, si $f : G \to \mathbf{C}$ es una función con valores complejos definida en un conjunto finito G, y $B \subseteq G$ es un subconjunto no vacío de G, usaremos

$$\mathbf{E}_{x \in B} f(x) := \frac{1}{|B|} \sum_{x \in B} f(x)$$

para denotar el promedio de f sobre B. Abreviaremos $\mathbf{E}_{x \in G} f(x)$ como $\mathbf{E}(f)$ cuando el dominio G de f sea claro por el contexto.

Ahora bien, si G es un grupo aditivo finito y $f_0, \ldots, f_{k-1} : G \to \mathbf{C}$ son funciones complejas, definimos la forma k-lineal $\Lambda_k(f_0, \ldots, f_{k-1}) \in \mathbf{C}$ mediante

$$\Lambda_k(f_0, \ldots, f_{k-1}) := \mathbf{E}_{x,r \in G} f_0(x) T^r f_1(x) \ldots T^{(k-1)r} f_{k-1}(x).$$

Observemos que si $A \subseteq G$ y $f_0 = \ldots = f_{k-1} = 1_A$, donde $1_A : G \to \{0,1\}$ denota la función indicadora de A, entonces $\Lambda_k(1_A, \ldots, 1_A)$ es simplemente el número de progresiones de longitud k (incluyendo aquellas con diferencia común 0), dividido por el factor de normalización N^2. En particular, si $(N, (k-1)!) = 1$ y A no contiene progresiones propias de longitud k, entonces se tiene que $\Lambda_k(1_A, \ldots, 1_A) = |A|/N^2$, lo cual será bastante pequeño cuando N sea grande.

Por lo tanto, es de interés determinar bajo qué condiciones $\Lambda_k(1_A, \ldots, 1_A)$ es pequeña o grande. Con este fin, Gowers introdujo (lo que ahora se conoce como) las **normas de uniformidad de Gowers** $\|f\|_{U^d(G)}$ para cualquier función compleja $f : G \to \mathbf{C}$:

Definición 7.6 (Normas de uniformidad de Gowers)**.** *Sea $d \geq 0$, y sea $f : G \to \mathbf{C}$ una función. Definimos la norma de uniformidad de Gowers $\|f\|_{U^d(G)} \geq 0$ de f como la cantidad*

$$\|f\|_{U^d(G)} := \big(\mathbf{E}_{x \in G, h \in G^d} \prod_{\omega \in \{0,1\}^d} \mathcal{C}^{|\omega|} T^{\omega \cdot h} f(x)\big)^{1/2^d},$$

donde $\omega = (\omega_1, \ldots, \omega_d)$, $h = (h_1, \ldots, h_d)$, $\omega \cdot h := \omega_1 h_1 + \ldots + \omega_d h_d$, $|\omega| := \omega_1 + \ldots + \omega_d$, y $\mathcal{C}$ es el operador de conjugación definido por $\mathcal{C} f(x) := \overline{f(x)}$.

Una definición equivalente muy útil de las normas $U^d(G)$ está dada por las siguientes fórmulas recursivas:

$$\|f\|_{U^0(G)} = \mathbf{E}(f); \quad \|f\|_{U^1(G)} = |\mathbf{E}(f)|; \quad \|f\|_{U^d} := \Big(\mathbf{E}_{h \in G} \|T^h f \overline{f}\|_{U^{d-1}(G)}^{2^{d-1}}\Big)^{1/2^d} \tag{7.1}$$

para todo $d \geq 1$.

Una configuración de la forma $(x + \omega \cdot h)_{\omega \in \{0,1\}^d}$ se denomina un **cubo de dimensión** d. Por lo tanto, $\|f\|_{U^d(G)}^{2^d}$ es un promedio ponderado de f sobre cubos; por ejemplo, $\|1_A\|_{U^d(G)}^{2^d}$ es igual al número de cubos contenidos en A, dividido por el factor de normalización N^{d+1}.

Los casos $d = 0, 1$ en la definición son algo degenerados, y de hecho $U^d(G)$ no es una norma en esos casos. Sin embargo, para $d > 1$, se puede demostrar que $\|\cdot\|_{U^d(G)}$ es efectivamente una norma, es decir, es homogénea, no negativa, no degenerada y satisface la desigualdad triangular.

Para comenzar el camino hacia la prueba de que estos objetos son efectivamente normas y afianzar algunas ideas, puede ser útil hacer algunos cálculos

sobre $\|.\|_{U^1}$ y $\|.\|_{U^2}$, que son los casos más sencillos. Primero, observar que:

$$\begin{aligned}\|f\|_{U^1} &= \big(\mathbf{E}_{x\in G,h\in G}\prod_{\omega\in\{0,1\}}\mathcal{C}^{|\omega|}T^{\omega\cdot h}f(x)\big)^{1/2}\\ &= \big(\mathbf{E}_{x\in G,h\in G}f(x)\overline{f(x+h)}\big)^{1/2}\\ &= \big(\mathbf{E}_{x\in G}f(x)\big(\mathbf{E}_{h\in G}\overline{f(x+h)}\big)\big)^{1/2}\\ &= \big(\mathbf{E}_{x\in G}f(x)\big(\mathbf{E}(\overline{f})\big)\big)^{1/2}\\ &= \big(\mathbf{E}(f)\big)\big(\mathbf{E}(\overline{f})\big)\big)^{1/2}\\ &= |\mathbf{E}(f)|,\end{aligned}$$

lo que justifica la definición alternativa de las normas.

Para la siguiente cuenta, que puede ser salteada si el lector no está cómodo con las ideas, será necesario trabajar con nociones de transformada de Fourier en grupos. Recomendamos para una introducción a estos tema el libro "Fourier Analysis on Number Fields" [24]. De todas formas, lo único que utilizaremos es la fórmula de inversión de Fourier y la identidad de Plancherel, que ahora repasaremos.

Dado un grupo abeliano finito G, definimos el **espacio dual de Pontryagin** $\hat{G}$ como el espacio de todos los morfismos $\xi : G \to \mathbf{R}/\mathbf{Z}$. Para cada función $f : G \to \mathbf{C}$, definimos la **transformada de Fourier** $\hat{f} : \hat{G} \to \mathbf{C}$ mediante la fórmula

$$\hat{f}(\xi) \coloneqq \mathbf{E}_{x\in G}f(x)e(-\xi(x)). \tag{7.2}$$

Tenemos también la **fórmula de inversión de Fourier**:

$$f(x) \coloneqq \sum_{\xi\in\hat{G}}\hat{f}(\xi)e(\xi(x)), \tag{7.3}$$

y, ya que estamos, dejamos sentada para uso futuro la **identidad de Plancherel**:

$$\mathbf{E}_x(|f(x)|^2) = \sum_{\xi\in\widehat{G}}|\widehat{f}(\xi)|^2. \tag{7.4}$$

Con estas herramientas podemos observar, usando la fórmula de inversión en la

segunda igualdad, que

$$\begin{aligned}
\|f\|_{U^2}^4 &= \mathbf{E}_{x,h_1,h_2\in G} f(x)\overline{f(x+h_1)}f(x+h_1+h_2)\overline{f(x+h_2)} \\
&= \mathbf{E}_{x,h_1,h_2\in G} \sum_{\xi_1,\xi_2,\xi_3,\xi_4\in\hat{G}} \big[\hat{f}(\xi_1)\hat{f}(\xi_2)\hat{f}(\xi_3)\hat{f}(\xi_4) \\
&\qquad\qquad \times e(\xi_1(x)+\xi_2(x+h_1+h_2)+\xi_3(x+h_1)+\xi_4(x+h_2))\big] \\
&= \sum_{\xi_1,\xi_2,\xi_3,\xi_4\in\hat{G}} \big[\hat{f}(\xi_1)\hat{f}(\xi_2)\hat{f}(\xi_3)\hat{f}(\xi_4) \\
&\qquad \times \mathbf{E}_{x,h_1,h_2\in G} e(\xi_1(x)+\xi_2(x+h_1+h_2)-\xi_3(x+h_1)-\xi_4(x+h_2))\big] \\
&= \sum_{\xi_1,\xi_2,\xi_3,\xi_4} f(\hat{\xi}_1)f(\hat{\xi}_2)f(\hat{\xi}_3)f(\hat{\xi}_4) \\
&\qquad \times \big[\mathbf{E}_x e(\xi_1(x)+\xi_2(x)-\xi_3(x)-\xi_4(x)) + \mathbf{E}_{h_1} e(\xi_2(h_1)-\xi_3(h_1)) \\
&\qquad\quad + \mathbf{E}_{h_1} e(\xi_2(h_1)-\xi_4(h_2))\big] \\
&= \sum_{\xi} \hat{f}(\xi)^4 = \|\hat{f}\|_{l^4(\hat{G})}^4,
\end{aligned}$$

donde el último paso se justifica notando que

$$\sum_{x\in G} e(\xi(x))) = 0,$$

siempre que el caracter sea no trivial, lo que en la expresión anterior indica que solo sobreviven los términos asociados a $\xi_1 = \xi_2 = \xi_3 = \xi_4$. En particular, esto nos está diciendo que en el caso $d = 2$ tenemos una norma.

Es importante notar que si se hace una cuenta similar pero para el caso de $\|.\|_{U^3}$, las mismas ideas fallan, ya que

$$\begin{aligned}
\|f\|_{U^3(G)} \quad := \quad |G|^{-4} \sum_{x,a,b,c\in G} (f(x)\overline{f(x+a)f(x+b)f(x+c)}f(x+a+b)\times \\
\times f(x+b+c)f(x+c+a)\overline{f(x+a+b+c)})^{1/8}.
\end{aligned}$$

y cuando aquí hacemos la misma cuenta mediante inversión de Fourier, las ecuaciones que nos quedan sobre los caracteres dejan de tener una solución sencilla.

Por este ejemplo es que utilizaremos otras técnicas para demostrar que esta y las demás $\|.\|_{U^d}$ son normas.

Para lograr esto, comenzaremos definiendo el **producto interno de Gowers**, como

$$\big\langle (f_\omega)_{\omega\in\{0,1\}^d} \big\rangle_{U^d(G)} := \mathbf{E}_{x,h_1,\ldots,h_d\in G} \prod_{\omega_1,\ldots,\omega_d\in\{0,1\}} \mathcal{C}^{\omega_1+\ldots+\omega_d} f_\omega(x+h_1\omega_1+\ldots+h_d\omega_d)$$

para cualquier 2^d-upla de funciones $f_\omega : G \to \mathbf{C}$, de forma que en particular vale

$$\big\langle (f)_{\omega\in\{0,1\}^d} \big\rangle_{U^d(G)} = \|f\|_{U^d}^{2^d}.$$

Una simple aplicación de la desigualdad clásica de Cauchy-Schwarz nos permite ver que

$$\big\langle (f_\omega)_{\omega\in\{0,1\}^d}\big\rangle_{U^d(G)} \leq \prod_{j=0,1} \big|\big\langle (f_\omega)_{\pi_{i,j}(\omega)\in\{0,1\}^d}\big\rangle_{U^d(G)}\big|^{1/2}$$

para todo $1 \leq i \leq d$, donde $\pi_{i,j}(\omega) \in \{0,1\}^d$ está formado por ω pero con la i-ésima coordenada reemplazada por j. Iterando esto puede observarse que

$$\big\langle (f_\omega)_{\omega\in\{0,1\}^d}\big\rangle_{U^d(G)} \leq \prod_{\omega\in\{0,1\}^d} \|f_\omega\|_{U^d(G)},$$

desigualdad conocida como la **desigualdad de Gowers-Cauchy-Schwarz**.

Podemos ver entonces que

$$\begin{aligned}\|f+g\|_{U^d}^{2^d} &= \mathbf{E}_{x\in G, h\in G^d} \prod_{\omega\in\{0,1\}^d} \mathcal{C}^{|\omega|} T^{\omega\cdot h}(f+g)(x)\\ &= \mathbf{E}_{x\in G, h\in G^d} \sum_{i=1}^{2^{2^d}} \prod_{\omega\in\{0,1\}^d} \mathcal{C}^{|\omega|} T^{\omega\cdot h} F_{i,\omega}(x)\end{aligned}$$

donde cada $F_{i,\omega}$ es una copia de f o de g, cosa que resultó de distribuir el producto del binomio. Continuando,

$$\begin{aligned}\sum_{i=1}^{2^{2^d}} \mathbf{E}_{x\in G, h\in G^d} \prod_{\omega\in\{0,1\}^d} \mathcal{C}^{|\omega|} T^{\omega\cdot h} F_{i,\omega}(x) &= \sum_{i=1}^{2^{2^d}} \big\langle (F_{i,\omega})_{\omega\in\{0,1\}^d}\big\rangle_{U^d(G)}\\ \leq \sum_{i=1}^{2^{2^d}} \prod_{\omega\in\{0,1\}^d} \|F_{i,\omega}\|_{U^d(G)} &= \sum_{i=0}^{2^d} \binom{2^d}{i} \|f\|_{U^d(G)}^{2^d-i} \|g\|_{U^d(G)}^{i}\\ &= \big(\|f\|_{U^d(G)} + \|g\|_{U^d(G)}\big)^{2^d},\end{aligned}$$

y de aquí se deduce la desigualdad triangular al quitar las potencias.

Daremos ahora un ejemplo para iluminar qué es lo que las normas de Gowers buscan encapsular. Supongamos que f tiene la forma $f(x) := e\big(\phi(x)\big)$ para alguna función de fase $\phi : G \to \mathbf{R}/\mathbf{Z}$, donde $e : \mathbf{R}/\mathbf{Z} \to \mathbf{C}$ es el mapa exponencial $e(x) := e^{2\pi i x}$. Entonces, un cálculo sencillo muestra que

$$\|f\|_{U^d}^{2^d} = \mathbf{E}_{x,h_1,\ldots,h_d\in G} e((h_1\cdot\nabla_x)\ldots(h_d\cdot\nabla_x)\phi(x)).$$

Así, la norma U^d mide, en cierto sentido, la oscilación presente en la "d-ésima derivada" de la fase. Por lo general, uno esperaría que sumar muchas exponenciales elevados a números complejos al azar de por resultado 0, ya que se presupone que habrá mucha cancelación (pensar en lo que sucede al sumar números de norma uno escogidos al azar). Por el otro lado, uno espera que, si la suma da grande, es porque los exponentes presentan bastante estructura, y se

están "complotando" para alinearse y no cancelarse. En particular, esperamos que la norma U^d sea grande si la fase se comporta como un "polinomio" de grado a lo sumo $d-1$, es decir que tomar d de estas derivadas aditivas anula la fase, y pequeña si la fase se comporta como una función "aleatoria". Esta idea de "polinomio" es muy importante para el estudio de las normas de Gowers, y será bien definido más adelante en el texto. Esta es una de las primeras instancias en las cuales las normas de Gowers nos permiten distinguir entre funciones "aleatorias" y "estructuradas".

Si entendemos que tomar transformada de Fourier de una función es calcular el producto interno de la función con respecto a una fase lineal exponencial, podemos entender entonces que lo que hace el análisis de Fourier clásico es comparar funciones con fases lineales. De esta forma, la idea trabajada en el párrafo anterior comienza a reflejar por qué es que el estudio de las normas de Gowers suele llamarse "análisis de Fourier de grado superior", ya que lo que hace es comparar a las funciones con fases polinomiales. En general, esperamos que estas fases polinomiales sean las únicas funciones con una norma de Gowers grande, y nos gustaría poder afirmar que una función cualquiera con una norma de Gowers grande debe estar muy correlacionada (en el sentido de un producto interno grande) con una de estas fases.

Observamos que, como consecuencia inmediata de la fórmula recursiva de las normas de Gowers (7.1) y un argumento inductivo, se tiene la propiedad de monotonicidad:

$$\|f\|_{U^d(G)} \leq \|f\|_{U^{d+1}(G)} \text{ para } d = 0, 1, 2, \ldots \tag{7.5}$$

En efecto, observar que

$$\begin{aligned}
\|f\|_{U^{d+1}(G)} &= \left(\mathbf{E}_{h\in G}\|T^h f\overline{f}\|_{U^d(G)}^{2^d}\right)^{1/2^{d+1}} \geq \left(\mathbf{E}_{h\in G}\|T^h f\overline{f}\|_{U^{d-1}(G)}^{2^d}\right)^{1/2^{d+1}} \\
&= \left(\mathbf{E}_{h\in G}(\|T^h f\overline{f}\|_{U^{d-1}(G)}^{2^{d-1}})^2\right)^{1/2^{d+1}} \geq \left(\mathbf{E}_{h\in G}\|T^h f\overline{f}\|_{U^{d-1}(G)}^{2^{d-1}}\right)^{1/2^d} \\
&= \|f\|_{U^d(G)},
\end{aligned}$$

donde la última desigualdad se justifica mediante una generalización de que $\sqrt{\frac{a^2+b^2}{2}} \geq \frac{a+b}{2}$.

Por último, es interesante notar que las normas de Gowers son invariantes ante ciertas transformaciones:

$$\|f\|_{U^d(G)} = \|T^h f\|_{U^d(G)} = \|\overline{f}\|_{U^d(G)} = \|fe(\phi(x))\|_{U^d(G)}, \tag{7.6}$$

si ϕ es una fase polinomial de grado como mucho $d-1$. La última de estas igualdades es la única que requiere un comentario adicional, y dejamos al lector justificarlo notando que la derivada multiplicativa de una fase polinomial de grado $d-1$ es una fase polinomial de grado $d-2$, en conjunto con la definición recursiva de las normas de Gowers (7.1).

La relevancia de las normas de uniformidad de Gowers para las progresiones aritméticas radica en el siguiente resultado. Denotemos $\mathcal{D} := \{z \in \mathbb{C} : |z| \leq 1\}$ al disco unitario.

Proposición 7.7 (Teorema de von Neumann generalizado)**.** *Sea G un grupo abeliano finito tal que $(N,(k-1)!)=1$. Sean $f_0,\dots,f_{k-1}:G\to\mathcal{D}$ funciones. Entonces se cumple que*

$$|\Lambda_k(f_0,\dots,f_{k-1})|\leq \min_{1\leq j\leq k}\|f_j\|_{U^{k-1}(G)}.$$

Demostración. Basta con probar la afirmación más general

$$|\mathbf{E}_{x,h}\prod_{j\in J}T^{a_jh}f_j(x)|\leq \|f_{j_0}\|_{U^{|J|-1}(G)} \tag{7.7}$$

para todo conjunto finito J con $|J|\geq 1$, todo $j_0\in J$, todas las funciones acotadas $(f_j)_{j\in J}$, y todos los enteros distintos $(a_j)_{j\in J}$ tales que $a_j-a_{j'}$ es coprimo con $|G|$ para todo par distinto $j,j'\in J$.

Procedemos por inducción sobre $|J|$. Cuando $|J|=1$ la afirmación es trivial a partir de (7.1), así que supongamos $|J|\geq 2$ y que la afirmación ya fue demostrada para valores menores de J. Sea j_1 un elemento de $J\backslash\{j_0\}$. Haciendo el cambio de variable $x\to x+a_{j_1}h$ si es necesario, podemos suponer que $a_{j_1}=0$. Dado que f_{j_1} es acotada, podemos entonces expresar el lado izquierdo de (7.7) como

$$\big|\mathbf{E}_{x,h\in G}\big(\mathbf{b}(x)\prod_{j\in J\backslash\{j_1\}}T^{a_jh}f_j(x)\big)\big|$$

para alguna función $\mathbf{b}(x)$ de norma infinito menor o igual a 1, lo cual, por (8.2)[1], se puede acotar por

$$\mathbf{E}_{k\in G}\big(\mathbf{E}_{x,h\in G}\big(\prod_{j\in J\backslash\{j_1\}}T^{a_jh}(T^{a_jk}f_j\overline{f_j})(x)\big)\big)^{1/2}.$$

Aplicando la hipótesis inductiva (7.7) a la esperanza interior, esto se puede acotar a su vez por

$$\mathbf{E}_{k\in G}(\|T^{a_jk}f_j\overline{f_j}\|_{U^{|J|-2}(G)})^{1/2},$$

lo cual, por la desigualdad de Hölder y haciendo el cambio de variable $h:=a_jk$ (notando que $(N,a_j-a_{j_1})=(N,a_j)=1$), se acota por

$$\mathbf{E}_{h\in G}\big(\|T^hf_j\overline{f_j}\|^{2^{|J|-2}}_{U^{|J|-2}(G)}\big)^{1/2^{|J|-1}}.$$

La afirmación se deduce entonces de la fórmula recursiva de las normas de Gowers (7.1). □

Es instructivo continuar con el ejemplo de la fase dado anteriormente. Si $f_j=e(\phi_j)$, entonces

$$\Lambda_k(f_0,\dots,f_{k-1})=\mathbf{E}_{x,r\in G}e(\phi_0(x)+\phi_1(x+r)+\ldots+\phi_{k-1}(x+(k-1)r)).$$

[1]Esta es una versión de la desigualdad de Van Der Corput, que probaremos más adelante como una aplicación de la desigualdad de Cauchy-Schwarz, de forma completamente independiente a lo que estamos haciendo ahora.

Así, $\Lambda_k(f_0, \ldots, f_{k-1})$ está midiendo la oscilación presente en la expresión $\phi_0(x) + \phi_1(x+r) + \ldots + \phi_{k-1}(x+(k-1)r)$. La Proposición 7.7 puede entonces interpretarse como una afirmación que dice que si esta expresión no oscila, entonces tampoco lo hacen las expresiones $(h_1 \cdot \nabla_x) \ldots (h_{k-1} \cdot \nabla_x)\phi_j(x)$ para todo $1 \leq j \leq k$. Obsérvese que este hecho se sigue moralmente de "diferenciar" la expresión $\phi_0(x) + \phi_1(x+r) + \ldots + \phi_k(x+(k-1)r)$ en $k-1$ direcciones diferentes para eliminar todos los términos salvo uno.

Con esto en la mano podemos retomar nuestro recorrido por la demostración del Teorema de Szemerédi:

Corolario 7.8 (Ausencia de progresiones implica norma de uniformidad grande, [6]). *Sea $k \geq 3$, sea G un grupo aditivo finito con $(N, (k-1)!) = 1$, y sea $A \subseteq G$, $|A| = \alpha N$, un conjunto no vacío tal que A no contiene progresiones aritméticas propias de longitud k. Si $N \geq 2/\alpha^{k-1}$, entonces se cumple que $\|1_A - \alpha\|_{U^{k-1}(G)} \geq 2^{-k-1}\alpha^{k-1}$.*

Más generalmente, sea P una progresión aritmética propia en G tal que $|P| \geq c_0 N$. Sea $A \subseteq P$, $|A| = \alpha|P|$, un conjunto no vacío que no contiene progresiones aritméticas propias. Si $N > N_0(c_0, k, \alpha)$, entonces se cumple $\|1_A - \alpha 1_P\|_{U^{k-1}(G)} \geq c(c_0, \alpha, k) > 0$, donde la cantidad $c(c_0, \alpha, k)$ se mantiene acotada inferiormente por una constante positiva cuando c_0, α están acotados inferiormente por constantes positivas y k está fijado.

Demostración. Comenzamos con la primera afirmación. Como $1_A = \alpha + (1_A - \alpha)$, podemos descomponer la expresión $\Lambda_k(1_A, \ldots, 1_A)$ como la suma de 2^k términos, uno de los cuales es $\Lambda_k(\alpha, \ldots, \alpha)$, y los otros $2^k - 1$ pueden acotarse en módulo por $\|1_A - \alpha\|_{U^{k-1}(G)}$ gracias a la Proposición 7.7. En particular, concluimos que

$$|\Lambda_k(\alpha, \ldots, \alpha) - \Lambda_k(1_A, \ldots, 1_A)| \leq 2^k \|1_A - \alpha\|_{U^{k-1}(G)}.$$

Pero claramente $\Lambda_k(\alpha, \ldots, \alpha) = \alpha^k$, mientras que como A no tiene progresiones aritméticas propias, se cumple $\Lambda_k(1_A, \ldots, 1_A) = |A|/N^2 = \alpha/N \leq \alpha^k/2$. De aquí se deduce la primera afirmación.

La segunda afirmación procede de manera similar, pero se basa en la descomposición $1_A = \alpha 1_P + (1_A - \alpha 1_P)$, y en la observación de que $\Lambda_k(1_P, \ldots, 1_P) \geq c(c_0, k) > 0$ para cierta cantidad positiva $c(c_0, k)$ que depende de c_0 y k; dejamos los detalles al lector. □

Comparando esta proposición con el Teorema 7.4, vemos entonces que para demostrar el teorema de Szemerédi para un k fijo, basta probar lo siguiente:

Teorema 7.9 (Norma de uniformidad grande implica incremento de densidad [6]). *Sean $\eta > 0$ y $k \geq 3$. Sea $G = \mathbf{Z}/N\mathbf{Z}$ un grupo cíclico de orden primo, y sea $f : G \to \mathcal{D}$ una función acotada de valores reales tal que $\mathbf{E}(f) = 0$ y $\|f\|_{U^{k-1}(G)} \geq \eta$. Entonces, si $N > N_0(k, \eta)$, existe una progresión aritmética propia $P \subseteq G$ con $|P| \geq \omega(N, \eta, k)$ tal que $\mathbf{E}_{x \in P} f(x) \geq c(\eta, k)$, donde:*

- *$\omega(N, \eta) \to \infty$ cuando $N \to \infty$ para η fija;*

- $c(\eta, k) > 0$ *está acotado inferiormente por una constante positiva cuando* η *está acotada inferiormente y* k *está fijo.*

En efecto, el Teorema 7.4 se deduce aplicando el Corolario 7.8 y luego invocando el Teorema 7.9 con $f := 1_A - (\mathbf{E}_{x \in Q} 1_A(x)) 1_Q$. El Teorema 7.9 se deduce en realidad en [6] a partir del siguiente teorema más fuerte:

Teorema 7.10 (Teorema inverso débil para $U^{k-1}(\mathbf{Z}/N\mathbf{Z})$, [6])**.** *Sean* $\eta > 0$ *y* $k \geq 3$. *Sea* $G = \mathbf{Z}/N\mathbf{Z}$ *un grupo cíclico de orden primo, y sea* $f : G \to \mathcal{D}$ *una función acotada tal que* $\|f\|_{U^{k-1}(G)} \geq \eta$. *Entonces, si* $N \geq \exp(C_k \eta^{-C_k})$ *para alguna constante suficientemente grande* $C_k > 0$, *se puede partir* $\mathbf{Z}/N\mathbf{Z}$ *en progresiones aritméticas* $(P_j)_{j \in J}$, *cada una de tamaño* $|P_j| \geq c_k \eta^{-C_k} N^{c_k \eta^{-C_k}}$ *para ciertos* $c_k, C_k > 0$, *tales que*

$$\sum_{j \in J} |\mathbf{E}(f\, 1_{P_j})| \geq c_k \eta^{C_k}$$

para ciertos $c_k, C_k > 0$.

El Teorema 7.9 (y por lo tanto el Teorema 7.1) se deduce rápidamente de este resultado y de la hipótesis de media cero $\sum_{j \in J} \mathbf{E}(f\, 1_{P_j}) = \mathbf{E}(f) = 0$. De hecho, se obtiene un resultado cuantitativo bastante bueno para el Teorema 7.2, con $N_0 = \exp(\exp(C_k \delta^{-C_k}))$ para algún $C_k > 0$ explícito.

Nos referimos al Teorema 7.10 como un teorema inverso *débil* porque proporciona un criterio simplemente necesario para que una función acotada f tenga norma $U^{k-1}(G)$ grande, y por lo tanto una condición suficiente para que la norma $U^{k-1}(G)$ sea pequeña. Además, notar que efectivamente este es un resultado inverso, en el sentido de que es posible demostrar que si la función f del enunciado es efectivamente una fase polinomial, debería cumplir la tesis del teorema, cosa que se detalla en [32, Ejercicio 1.5.6]. Como se discutió anteriormente, este teorema es lo suficientemente fuerte como para implicar el teorema de Szemerédi.

Es por ello de interés obtener un mejor teorema inverso para la norma $U^{k-1}(\mathbf{Z}/N\mathbf{Z})$, que proporcione una condición más fácil de verificar para saber cuándo esta norma es pequeña. Idealmente, nos gustaría que esta condición fuera tanto necesaria como suficiente, al menos hasta pérdidas constantes.

Presentamos ahora algunos ejemplos y resultados que ilustrarán cómo debería ser el teorema inverso. Recordemos que la norma U^d de una función $e(\phi)$ mide la oscilación en la derivada d-ésima de la fase ϕ. También recordemos que un polinomio de grado a lo sumo $d-1$ es una función cuya derivada d-ésima se anula. Generalizamos este concepto, como ya anticipamos, del siguiente modo:

Definición 7.11. ***(Funciones de fase polinomial locales)*** *Si* B *es un subconjunto no vacío de un grupo aditivo finito* G *y* $d \geq 1$, *diremos que una función* $\phi : B \to \mathbf{R}/\mathbf{Z}$ *es una* ***función de fase polinomial de orden a lo sumo d-1 localmente en*** B *si se cumple que*

$$(h_1 \cdot \nabla_x) \dots (h_{d+1} \cdot \nabla_x) \phi(x) = 0$$

siempre que el cubo $(x + \omega_1 h_1 + \ldots + \omega_d h_d)_{\omega_1,\ldots,\omega_d \in \{0,1\}}$ *esté contenido en* B. *Si* $f : B \to \mathbf{C}$ *es una función, definimos la* ***tendencia polinómica local de orden d en B****, denotada* $\|f\|_{u^d(B)}$, *como la cantidad*

$$\|f\|_{u^d(B)} := \sup |\mathbf{E}_{x \in B}(f(x)e(-\phi(x)))|$$

donde el supremo se toma sobre todas las funciones fase polinómicas locales de orden a lo sumo $d-1$ *en* B.

Comenzaremos trabajando en el caso global $B = G$, aunque más adelante será necesario considerar también el caso local. Nos referiremos a las funciones de fase polinomial de grado a lo sumo 1 como **funciones de fase lineal**, y a las de grado a lo sumo 2 como **funciones de fase cuadrática**, utilizando los modificadores "local" o "global" según corresponda.

La cantidad $\|.\|_{u^d(B)}$ es claramente una seminorma, y comparte varias propiedades con la norma $U^d(G)$. En primer lugar, al igual que la norma $U^d(G)$, se cumple la monotonía $\|f\|_{u^d(B)} \leq \|f\|_{u^{d+1}(B)}$, y cuando $B = G$ también tenemos la invarianza por traslación $\|T^h f\|_{u^d(G)} = \|f\|_{u^d(G)}$. Además, se cumple la simetría por conjugación $\|\overline{f}\|_{u^d(B)} = \|f\|_{u^d(B)}$, y la invarianza de fase $\|fe(\phi)\|_{u^d(B)} = \|f\|_{u^d(B)}$ siempre que ϕ sea una función de fase polinomial local de grado a lo sumo $d-1$ en B.

A partir de esta invarianza y de (7.1), (7.5), concluimos que

$$\|f\|_{U^d(G)} = \|fe(-\phi)\|_{U^d(G)} \geq \|fe(-\phi)\|_{U^1(G)} = |\mathbf{E}_{x \in G}(f(x)e(-\phi(x)))|$$

siempre que ϕ sea una función de fase polinomial global de grado a lo sumo $d-1$. Tomando el supremo sobre todas las ϕ, obtenemos la desigualdad

$$\|f\|_{U^d(G)} \geq \|f\|_{u^d(G)} \tag{7.8}$$

para todo $d \geq 1$, todo grupo aditivo G, y toda función $f : G \to \mathbf{C}$.

Ahora es natural preguntarse si la desigualdad (7.8) puede invertirse. Cuando $d = 1$, es fácil verificar (usando (7.1) y el hecho de que los polinomios de grado a lo sumo 0 son constantes) que en efecto hay igualdad:

$$\|f\|_{U^1(G)} = \|f\|_{u^1(G)}.$$

Consideremos ahora el caso $d = 2$. Recordemos que:

$$\|f\|_{U^2(G)}^4 = \sum_{\xi \in \widehat{G}} |\widehat{f}(\xi)|^4. \tag{7.9}$$

A continuación, observemos que si $\phi : G \to \mathbf{R}/\mathbf{Z}$ es una función de fase polinomial global de grado a lo sumo 1, entonces la función $x \mapsto \phi(x) - \phi(0)$ es un morfismo de G en $\mathbf{R}/\mathbf{Z}$, y por lo tanto existe $\xi \in \widehat{G}$ tal que $\phi(x) = \xi \cdot x + \phi(0)$. A partir de esto, se deduce fácilmente que

$$\|f\|_{u^2(G)} = \sup_{\xi \in \widehat{G}} |\widehat{f}(\xi)|. \tag{7.10}$$

Notemos que

$$\|f\|_{U^2(G)}^4 = \sum_{\xi\in\widehat{G}}(|\hat{f}|^4) \leq (\sum_{\xi\in\widehat{G}}|\hat{f}|^2)^2 \leq (\mathbf{E}_{x\in G}|f|^2)^2$$
$$= (\mathbf{E}_{x\in G} f\overline{f})^2 \leq (\mathbf{E}_{x\in G} f e(\phi(x)))^2$$

Concluimos, tomando supremo, lo siguiente:

Proposición 7.12 (Teorema inverso para la norma $U^2(G)$)**.** *Sea $f : G \to \mathrm{D}$ una función acotada. Entonces*

$$\|f\|_{u^2(G)} \leq \|f\|_{U^2(G)} \leq \|f\|_{u^2(G)}^{1/2}.$$

A partir de evidencias como la Proposición 7.12, uno se ve tentado a conjeturar que las normas $U^d(G)$ y $u^d(G)$ también están relacionadas para valores mayores de d, en el sentido de que si f es acotada y una de las dos normas $\|f\|_{U^d(G)}$, $\|f\|_{u^d(G)}$ es pequeña, entonces la otra también lo es.

Por (7.8) ya sabemos que una de las direcciones es cierta: la pequeñez de la norma $U^d(G)$ implica la pequeñez de la norma $u^d(G)$. Lo difícil es desarrollar teoría para demostrar la otra dirección.

En el artículo en el cual nos basamos para armar está sección [14], Green y Tao demostraron una relación entre ambas normas con una pérdida cuasi polinomial, cosa que se pudo mejorar a través de la resolución de la Conjetura de Marton a una relación polinomial:

Corolario 7.13 ([7, Corolario 1.2])**.** *Sea p un primo impar. Sea $f\colon \mathbf{F}_p^n \to \mathbf{C}$ una función 1-acotada, y supongamos que $\|f\|_{U^3(\mathbf{F}_p^n)} \geq \eta$ para algún η, $0 < \eta \leq \frac{1}{2}$. Existe entonces un polinomio cuadrático $\phi\colon \mathbf{F}_p^n \to \mathbf{F}_p$ tal que*

$$|\mathbf{E}_{x\in\mathbf{F}_p^n} f(x) e_p(-\phi(x))| \gg \eta^{O(p^3)},$$

donde $e_p(x) \coloneqq e^{2\pi i x/p}$.

La versión de este mismo resultado en $\mathbf{F}_2$ es:

Corolario 7.14 ([8, Corolario 1.6])**.** *Existe una constante C con la siguiente propiedad. Sea $f : \mathbf{F}_2^n \to \mathbf{C}$ una función acotada por 1, tal que $\|f\|_{U^3(\mathbf{F}_2^n)} \geq 1/K$. Luego existe un polinomio cuadrático $P : \mathbf{F}_2^n \to \mathbf{F}_2$ tal que*

$$|\frac{1}{2^n}\sum_{x\in\mathbf{F}_2^n} f(x)(-1)^{P(X)}| \gg K^{-C}.$$

Dedicaremos la siguiente sección a la deducción del primero de estos resultados desde la Conjetura de Marton.

Para concluir, nos parece pertinente volver a la discusión dada en la Introducción, en la cual motivamos el estudio de los grupos aproximados, y comentamos que en la Combinatoria Aditiva se estudian distintas estructuras

aproximadas. En este espíritu, la forma de estudiar a estas fases polinomiales y a las normas de Gowers nos da una forma de entender polinomios aproximados, a través de ciertas propiedades aditivas que los polinomios tradicionales deben cumplir. Además, es importante aclarar que el estudio de estas normas no se restringe al estudio de progresiones aritméticas, sino que son fundamentales para estudiar amplias familias de progresiones lineales, en distintos tipos de conjuntos, a veces no densos en el sentido clásico, como lo son los **primos**. En particular, Green y Tao probaron que:

Teorema 7.15 (Teorema de Szemerédi en los números primos, [13])**.** *Sea A un subconjunto de los primos con* densidad superior relativa *positiva; es decir,*

$$\limsup_{N\to\infty} \frac{|A \cap [1,N]|}{\pi(N)} > 0,$$

donde $\pi(N)$ denota la cantidad de números primos acotados por N. Luego, para cualquier entero $k \geq 1$, el conjunto A contiene infinitas progresiones aritméticas de longitud k.

La demostración de este resultado involucra dos ingredientes muy importantes, siendo el primero un desarrollo extenso de las normas de Gowers, para poder controlar las progresiones aritméticas en ciertos conjuntos a través de una generalización del teorema de Szemerédi, mientras que la segunda parte es un argumento de transferencia de este tipo de resultados a los números primos. La primera de estas partes está bastante más relacionada con las cosas recién discutidas, mientras que la transferencia a los números primos se hace mediante argumentos de Teoría Analítica de Números.

Capítulo 8

Una aplicación a la norma U_3

El objetivo de este capítulo es demostrar que la Conjetura de Marton implica un teorema inverso de las normas de Gowers $\|.\|_{U^3\mathbf{F}_p^n}$, cosa que fue enunciada al final de la sección anterior. Además, es posible dar un resultado en la otra dirección, indicando una equivalencia entre la Conjetura de Marton y el teorema de las normas de Gowers.

Vamos a contentarnos en hacer el caso de $\mathbf{F}_p^n$ con p impar, que es levemente distinto al caso de torsión par, principalmente debido a que en estos espacios se puede dividir por 2, y esto suele resultar bastante útil al trabajar con formas cuadráticas. La dificultad es técnica, no sustancial, y el lector interesado puede seguir el trabajo de Samorodnitsky en "Low-degree tests at large distances" [28], haciendo ciertas modificaciones detalladas en un apéndice de "On a conjecture of Marton" [8], para obtener el resultado. Las principales fuentes de esta sección son "An inverse theorem for the Gowers $U^3(G)$ norm" [14] y "An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm" [15], ambos de Green y Tao, de donde también sugerimos al lector interesado buscar la demostración de que el teorema inverso de las normas de gowers implica la Conjetura de Marton, cosa que mustra que ambos resultados son equivalentes. Para ser exactos, las demostraciones que daremos en este capítulo no se encuentran en la bibliografía, pero se obtienen siguiendo las indicaciones dadas en "Marton's Conjecture in abelian groups with bounded torsion" [7, Apéndice C] aplicadas a los resultados de [14].

8.1 Preliminares

El objetivo de este capítulo es demostrar el siguiente resultado, que enunciamos una vez más por comodidad del lector:

Corolario 8.1. *Sea p un primo impar. Sea $f\colon \mathbf{F}_p^n \to \mathbf{C}$ una función 1-acotada,*

y supongamos que $\|f\|_{U^3(\mathbf{F}_p^n)} \geq \eta$ *para algún* η, $0 < \eta \leq \frac{1}{2}$. *Existe entonces un polinomio cuadrático* $\phi\colon \mathbf{F}_p^n \to \mathbf{F}_p$ *tal que*

$$|\mathbf{E}_{x\in\mathbf{F}_p^n} f(x)e_p(-\phi(x))| \gg \eta^{O(p^3)},$$

donde $e_p(x) \coloneqq e^{2\pi ix/p}$.

El corazón de la demostración de este resultado es la Sección 8.2, pero para poder llegar a ahí primero vamos a establecer algunos resultados y notaciones que serán de utilidad.

Cuando $g(x_1, \ldots, x_n)$ es una función a valores complejos con $\|g\|_\infty \leq 1$, nos referiremos a $g(x_1, \ldots, x_n)$ simplemente como $\mathbf{b}(x_1, \ldots, x_n)$. La notación $\mathbf{b}$ denota así una función con $\|\mathbf{b}\|_\infty \leq 1$, pero esta notación puede hacer referencia a distintas funciones de una línea a otra, o incluso dentro de una misma línea (de manera similar a la notación O, o al uso de constantes no especificadas C, el objetivo de esta notación es simplificar las expresiones, haciendo énfasis en distintos momentos en las cuentas que importan).

Lo siguiente que enunciamos es una versión de la desigualdad de Cauchy-Schwarz, acompañada de un resultado del estilo de los lemas de **Van der Corput**. Estos lemas, en general, vienen asociados al control de sumas exponenciales, y suelen decir cosas del estilo "si considero la suma exponencial $\sum_x e^{\phi(x)}$, esta va a ser pequeña si las derivadas de ϕ son grandes", idea que vale intuitivamente ya que uno esperaría mucha cancelación.

Lema 8.2. ***(Cauchy-Schwarz)*** *Sean* X, Y *conjuntos finitos, y sea* $f : X \times Y \to \mathbf{C}$ *una función. Entonces, para cualquier función acotada* $\mathbf{b}(x)$ *definida en* X, *se cumple que*

$$\begin{aligned}|\mathbf{E}_{x\in X, y\in Y} f(x,y)\mathbf{b}(x)| \leq \mathbf{E}_{x\in X}|\mathbf{E}_{y\in Y} f(x,y)| \leq (\mathbf{E}_{x\in X}|\mathbf{E}_{y\in Y} f(x,y)|^2)^{1/2} \\ = (\mathbf{E}_{x\in X, y,y'\in Y} f(x,y')\overline{f(x,y)})^{1/2}. \quad (8.1)\end{aligned}$$

En el caso particular en que $Y = G$ es un grupo, se deduce en particular la **desigualdad de Van der Corput**:

$$|\mathbf{E}_{x\in X, y\in G} f(x,y)\mathbf{b}(x)| \leq |\mathbf{E}_{x\in X, y,h\in G} T_y^h f(x,y)\overline{f(x,y)}|^{1/2} \quad (8.2)$$

El objetivo de ambos resultados es simplificar expresiones que involucren a funciones $\mathbf{b}$, y como la demostración es una aplicación directa del Cauchy-Schwarz clásico, la omitiremos.

Otro lema que usaremos es:

Lema 8.3 (Forma trilineal grande implica correlación con fase lineal)**.** *Sean* B, B' *dos subconjuntos no vacíos de un grupo aditivo* G. *Entonces se tiene que*

$$\|f\|_{u^2(B')} \geq \frac{\mathbf{E}(1_B)}{\mathbf{E}(1_{B+B'})}|\mathbf{E}_{z\in B', x\in B}(f(z)\mathbf{b}_1(x)\mathbf{b}_2(z+x))|$$

para toda función $f : G \to \mathbf{C}$ *y cualesquiera dos funciones acotadas* $\mathbf{b}_1$, $\mathbf{b}_2$.

Demostración. Sin pérdida de generalidad, podemos suponer que f, $\mathbf{b}_1$ y $\mathbf{b}_2$ se anulan fuera de B', B y $B+B'$, respectivamente. A partir del desarrollo en serie de Fourier tenemos

$$\begin{aligned}\mathbf{E}_{z\in B', x\in B}(f(z)\mathbf{b}_1(x)\mathbf{b}_2(z+x)) &= \frac{1}{\mathbf{E}(1_{B'})\mathbf{E}(1_B)}\mathbf{E}_{y\in G}(f * \mathbf{b}_1(y)\mathbf{b}_2(y)) \\ &= \frac{1}{\mathbf{E}(1_{B'})\mathbf{E}(1_B)}\sum_{\xi\in\widehat{G}} \widehat{f}(\xi)\widehat{\mathbf{b}}_1(\xi)\widehat{\mathbf{b}}_2(-\xi).\end{aligned}$$

Por otro lado, por Plancherel tenemos que

$$\sum_{\xi\in\widehat{G}} |\widehat{\mathbf{b}}_1(\xi)|^2 \leq \mathbf{E}(1_B)$$

y

$$\sum_{\xi\in\widehat{G}} |\widehat{\mathbf{b}}_2(-\xi)|^2 \leq \mathbf{E}(1_{B+B'}).$$

Aplicando la desigualdad de Hölder, concluimos que

$$|\mathbf{E}_{z\in B', x\in B}(f(z)\mathbf{b}_1(x)\mathbf{b}_2(z+x))| \leq \frac{\mathbf{E}(1_{B+B'})}{\mathbf{E}(1_{B'})\mathbf{E}(1_B)} \sup_{\xi\in\widehat{G}} |\widehat{f}(\xi)|,$$

y por lo tanto existe $\xi \in \widehat{G}$ tal que

$$|\mathbf{E}_{z\in B'}(f(z)e(-\xi\cdot z))| \geq |\mathbf{E}_{z\in B', x\in B}(f(z)\mathbf{b}_1(x)\mathbf{b}_2(z+x))| \frac{\mathbf{E}(1_B)}{\mathbf{E}(1_{B+B'})}.$$

La afirmación se deduce. □

Por último, enunciamos y demostramos un sencillo lema de promedios sobre grupos:

Lema 8.4 (Promediado en un subgrupo)**.** *Sea G un grupo aditivo, sea H un subgrupo finito de G, y sea $A \subseteq H$ un subconjunto no vacío. Sea $f : H \to \mathbf{C}$ una función. Entonces*

$$\mathbf{E}_{x\in H}\mathbf{E}_{y\in x+A} f(y) = \mathbf{E}_{y\in H} f(y).$$

En particular, por el principio del palomar, existe $x \in H$ tal que

$$|\mathbf{E}_{y\in x+A} f(y)| \geq |\mathbf{E}_{y\in H} f(y)|.$$

Demostración. Como $H + h = H$ para todo $h \in A$, se tiene que

$$\mathbf{E}_{x\in H} f(x+h) = \mathbf{E}_{y\in H} f(y) \text{ para todo } h \in A.$$

Promediando esto sobre todos los $h \in A$ se obtiene la primera afirmación, y la segunda se deduce del principio del palomar. □

8.2 Una idea de Gowers

La primer parte del argumento es establecer una "derivada de fase" $h \to \xi_h$ para la función f y establecer ciertas propiedades aditivas en esta derivada de fase. Estos argumentos se aplican a grupos finitos aditivos arbitrarios G; más tarde nos adentraremos en el caso particular de los cuerpos finitos.

Proposición 8.5 (Grandes normas U^3 dan varias tuplas aditivas, [5])**.** *Sea G un grupo finito aditivo cualquiera, y sea $f : G \to \mathcal{D}$ una función tal que $\|f\|_{U^3(G)} \geq \eta$ para algún $\eta > 0$. Entonces existe un subconjunto $H \subseteq G$, y una función $\xi : H \to \widehat{G}$ cuyo gráfico $\Gamma := \{(h, \xi_h) : h \in H\} \subseteq G \times \widehat{G}$ cumple con la estimación*

$$|\{(z_1, z_2, z_3, z_4) \in \Gamma^4 : z_1 + z_2 = z_3 + z_4\}| \geq 2^{-8}\eta^{64}N^3. \tag{8.3}$$

Más aún, para cada $(h, \xi_h) \in \Gamma$ tenemos

$$|\mathbf{E}_x T^h f(x)\overline{f}(x)e(-\xi_h \cdot x)| \geq \eta^4/2^{1/2}.$$

Demostración. De (7.1) tenemos

$$\mathbf{E}_{h\in G}\|T^h f\overline{f}\|^4_{U^2(G)} \geq \eta^8.$$

Aplicando la Proposición 7.12 concluimos que

$$\mathbf{E}_{h\in G}\|T^h f\overline{f}\|^2_{u^2(G)} \geq \eta^8.$$

Entonces si llamamos

$$H := \{h \in G : \|T^h f\overline{f}\|^2_{u^2(G)} \geq \eta^8/2\},$$

obtenemos

$$\mathbf{E}_{h\in G}\|T^h f\overline{f}\|^2_{u^2(G)} 1_{G\setminus H}(h) \leq \eta^8/2,$$

y luego

$$\mathbf{E}_{h\in G}\|T^h f\overline{f}\|^2_{u^2(G)} 1_H(h) \geq \eta^8/2.$$

En particular tenemos

$$\mathbf{E}(1_H) \geq \eta^8/2, \tag{8.4}$$

debido a que f esta acotada en norma por 1. Por (7.10) y la definición de H, podemos encontrar un mapeo $h \mapsto \xi_h$ de H a $\widehat{G}$ tal que

$$|\mathbf{E}_x T^h f(x)\overline{f}(x)e(-\xi_h \cdot x)| \geq \eta^4/2^{1/2} \text{ para todo } h \in H. \tag{8.5}$$

Fijemos esta aplicación $h \mapsto \xi_h$. Sumamos los cuadrados de la expresión de arriba indexando en h y usamos (8.4) para concluir

$$\mathbf{E}_h|\mathbf{E}_x T^h f(x)\overline{f}(x)e(-\xi_h \cdot x)|^2 1_H(h) \geq \eta^{16}/4.$$

Pero de la identidad

$$|\mathbf{E}_x T^h f(x)\overline{f}(x)e(-\xi_h\cdot x)|^2 = \mathbf{E}_{x,k} T^k(T^h f)(x)\overline{T^h f}(x)\overline{T^k f(x)}f(x)e(\xi_h\cdot k)$$

concluimos

$$|\mathbf{E}_{x,h,k} T^k(T^h f)(x)\overline{T^h f}(x)e(\xi_h\cdot k)1_H(h)\overline{T^k f(x)}f(x)| \geq \eta^{16}/4. \tag{8.6}$$

En este punto suprimimos las menciones explicitas de f y escribimos esto simplemente como

$$|\mathbf{E}_{x,h,k\in G}(\mathbf{b}(x+h,k)1_H(h)e(\xi_h\cdot k)\mathbf{b}(x,k))| \geq \eta^{16}/4.$$

Aplicando (8.2) para eleminiar el factor de $\mathbf{b}(x,k)$, concluimos

$$\mathbf{E}_{h,h_1,x,k}\mathbf{b}(x+h+h_1,k)\mathbf{b}(x+h,k)e((h_1\cdot\nabla_h)\xi_h\cdot k)1_H(h+h_1)1_H(h) \geq \eta^{32}/16. \tag{8.7}$$

A través de la sustitución $y = x+h$, obtenemos

$$\mathbf{E}_{h,y,h_1,k}e((h_1\cdot\nabla_h)\xi_h\cdot k)1_H(h+h_1)1_H(h)\mathbf{b}(y,h_1,k) \geq \eta^{32}/16. \tag{8.8}$$

Aplicando (8.2) otra vez, concluimos

$$\mathbf{E}_{h,h_1,h_2,y,k}e((h_2\cdot\nabla_h)(h_1\cdot\nabla_h)\xi_h\cdot k)1_H(h+h_1+h_2)1_H(h+h_2)1_H(h+h_1)1_H(h) \geq 2^{-8}\eta^{64}. \tag{8.9}$$

Sumando esto en k usando la formula de inversión de Fourier, y descartando el promediado irrelevante en y, inferimos

$$\mathbf{E}_{h,h_1,h_2}1_{(h_2\cdot\nabla_h)(h_1\cdot\nabla_h)\xi_h=0}1_H(h+h_1+h_2)1_H(h+h_2)1_H(h+h_1)1_H(h) \geq \eta^{64}/256. \tag{8.10}$$

Se deduce entonces el resultado sustituyendo $z_1 := (h,\xi_h)$, $z_2 := (h+h_1,\xi_{h+h_1})$, $z_3 := (h+h_2,\xi_{h+h_2})$, y $z_4 := (h+h_1+h_2,\xi_{h+h_1+h_2})$. □

Para explotar la conclusión de (8.3), vamos a requerir dos resultados muy útiles de Combinatoria Aditiva, ambos ya enunciados en distintas versiones en la presente tesis. El primer resultado, una variante de Balog-Szemerédi-Gowers, afirma que un conjunto Γ' con cierta estructura aditiva parcial (en el sentido de que contiene varias cuadruplas aditivas) se puede refinar a tener una estructura aditiva más completa (en el sentido de que su conjunto de suma es chico). Resultados de este estilo se encuentran en distintos lugares dentro de la bibliografía, pero recomendamos al lector interesado buscar referncias en [14, Sección 5].

Escribimos $\Gamma'-\Gamma' := \{z-z' : z,z'\in\Gamma'\}$ para el conjunto resta de Γ'.

Teorema 8.6. *(Teorema Balog-Szemerédi-Gowers) Sea G un conjunto aditivo, y sea Γ un subconjunto finito no vacio de G tal que*

$$|\{(z_1,z_2,z_3,z_4)\in\Gamma : z_1+z_2 = z_3+z_4\}| \geq |\Gamma|^3/K$$

para algún $K\geq 1$. Entonces existe un subconjunto $\Gamma'\subseteq\Gamma$ tal que

$$|\Gamma'| \geq 2^{-6}K^{-1}|\Gamma| \text{ and } |\Gamma'-\Gamma'| \leq 2^{42}K^6|\Gamma'|.$$

La otra herramienta que necesitamos es una desigualdad del tipo de Plünnecke, de la cual se puede encontrar una demostración en [26].

Teorema 8.7. *(Desigualdad de Plünnecke [26])*[1] *Sea G un grupo aditivo cualquiera, y sea Γ' un subconjunto finito no vacio de G tal que $|\Gamma' - \Gamma'| \leq K|\Gamma'|$ para algún $K \geq 1$. Tenemos entonces que $|k\Gamma' - l\Gamma'| \leq K^{k+l}|\Gamma'|$ para todo $k, l \geq 1$, donde $k\Gamma'$ denota el conjunto k-veces de suma de Γ'.*

También referimos al lector a [35], donde distintas versiones de estos últimos dos teoremas son tratadas, poniendolos más en contexto y dando intuiciones sobre por qué son resultados naturales en estas áreas.

Combinando la Proposición 8.5 con el Teorema 8.6 y el Teorema 8.7, concluimos:

Proposición 8.8 ($h \mapsto \xi_h$ es casi lineal afín)**.** *Sea G un grupo finito aditivo cualquiera, y sea $f : G \to \mathrm{D}$ una función acotada tal que $\|f\|_{U^3(G)} \geq \eta$ para algún $\eta > 0$. Existe entonces un subconjunto $H' \subseteq G$ y una función $\xi : H' \to \widehat{G}$ cuyo gráfico $\Gamma' := \{(h, \xi_h) : h \in H'\} \subseteq G \times \widehat{G}$ obedece las estimaciones*

$$|\Gamma'| \geq 2^{-14}\eta^{64}N \tag{8.11}$$

y

$$|k\Gamma' - l\Gamma'| \leq (2^{90}\eta^{-384})^{k+l}|\Gamma'| \text{ para todo } k, l \geq 1.$$

Más aún para cada $(h, \xi_h) \in \Gamma'$ tenemos que

$$|\mathbf{E}_{x \in G}(T^h f(x)\overline{f(x)}e(-\xi_h \cdot x))| \geq \eta^4/2. \tag{8.12}$$

Vamos ahora a aprovechar que no estamos trabajando en cualquier grupo G, sino en espacios vectoriales de dimensión finita.

8.3 El caso de los cuerpos finitos

Gracias a la Proposición 8.8, ya conseguimos una derivada de fase $h \mapsto \xi_h$ que exhibe cierto comportamiento lineal. El primer paso (siguiendo a Gowers [5]) es mostrar que ξ de hecho coincide con una fase lineal en un subespacio grande; luego vamos a mostrar que incluso ξ coincide con una fase lineal *autoadjunta* en un subespacio grande (esta es la parte substancialmente nueva en el trabajo de Green y Tao [14]). Finalmente, vamos a seguir ideas de Gowers [5] y conjugar f con una fase cuadrática para eliminar esta derivada de fase lineal y concluir el argumento.

Paso 1: Linealización de la derivada de fase.

[1] Esta versión de la desigualdad de Plünnecke es casi la misma que dimos en la Proposición 3.2, pero aquella era una versión bipartita del resultado. Son esencialmente dos reformulaciones de lo mismo, y para lo que haremos ahora nos bastaría con la versión ya dada en el Capítulo 3, pero decidimos enunciar esta otra versión por comodidad del lector.

Aplicando la Proposición 8.8 (e identificando $\mathbf{F}_p^n$ con su dual de Pontryagin $\widehat{\mathbf{F}_p^n}$ de la forma usual), uno puede encontrar un subconjunto H' de $\mathbf{F}_p^n$ y una función $\xi : H' \to \mathbf{F}_p^n$ cuyo gráfico $\Gamma' := \{(h, \xi_h) : h \in H'\} \subseteq \mathbf{F}_p^n \times \mathbf{F}_p^n$ obedece las estimaciones

$$|\Gamma'| \gg \eta^{O(1)} p^n$$

y

$$|2\Gamma'| \ll \eta^{-O(1)} p^n,$$

además de que

$$|\mathbf{E}_{x \in \mathbf{F}_p^n} f(x+h)\overline{f(x)} e(-\xi_h \cdot x)| \gg \eta^{O(1)}$$

para todo $(h, \xi_h) \in \Gamma'$, donde $\cdot : \mathbf{F}_p^n \times \mathbf{F}_p^n \to \mathbf{F}_p$ es el producto interno usual.

Aplicando la versión de la Conjetura de Marton para grupos de torsión impar, Teorema 5.1, al conjunto H', podemos cubrir Γ' con $O(\eta^{-O(p^3)})$ trasladados de un subgrupo H de $\mathbf{F}_p^n \times \mathbf{F}_p^n$ de cardinal como mucho $|H'| \leq p^n$. Si introducimos los grupos

$$H_1 := \{y \in \mathbf{F}_p^n : (0, y) \in H\}$$

y

$$H_0 := \{x \in \mathbf{F}_p^n : (x, y) \in H \text{ para algún } y \in \mathbf{F}_p^n\}$$

entonces $|H| = |H_0||H_1|$, y H' puede cubrirse con $O(\eta^{-O(p^3)})$ trasladados de H_0. Luego

$$|H_0| \gg \eta^{O(p^3)} p^n$$

y así

$$|H_1| = |H|/|H_0| \ll \eta^{-O(p^3)}.$$

Considerando un subespacio complementario de $\{0\}^n \times H_1$ en H, por ejemplo $\mathbf{F}_p^n \times \{0\}^n \cap H$, podemos escribir

$$H = \{(x, y) : x \in H_0, y - M_0 x \in H_1\}$$

para alguna transformación lineal (la proyección a este espacio complementario que dimos de ejemplo) $M_0 \colon H_0 \to \mathbf{F}_p^n$, que podemos luego extender (de forma más o menos arbitraria) a un morfismo de $\mathbf{F}_p^n$ a $\mathbf{F}_p^n$. Como estamos en característica impar, podemos escribir $M_0 = 2M$[2] para alguna otra transformación lineal $M \colon \mathbf{F}_p^n \to \mathbf{F}_p^n$. Uno puede entonces cubrir H por $O(|H_1|) = O(\eta^{-O(p^3)})$ trasladados del gráfico $\{(x, 2Mx) : x \in \mathbf{F}_p^n\}$, y luego Γ' puede también ser cubierto por $O(\eta^{-O(p^3)})$ trasladados de este gráfico. Concluimos que

$$\mathbf{E}_{h \in \mathbf{F}_p^n} 1_{H'}(h) 1_{\xi_h = 2Mh + \xi_0}(h) \gg \eta^{O(p^3)}$$

para algún $\xi_0 \in \mathbf{F}_p^n$.

Con todo esto, obtenemos que

$$\mathbf{E}_{h \in \mathbf{F}_p^n} |\mathbf{E}_{x \in \mathbf{F}_p^n} f(x+h)\overline{f(x)} e(-(2Mh + \xi_0) \cdot x)| \gg \eta^{O(p^3)} \tag{8.13}$$

[2]Esta es la situación en la cual utilizamos que estamos trabajando en torsión impar, y es desde aquí de donde se parten los caminos con respecto al caso de la torsión par.

Paso 2: El argumento de simetría.

Establecemos ahora algunas propiedades de simetría en M. Como vamos a estar concentrados más en M que en f en este paso, vamos a reprimir los términos involucrando a f (y a ξ_0 también) usando la notación de $\mathbf{b}$. Por (8.13) tenemos que

$$|\mathbf{E}_{x,h\in\mathbf{F}_p^n}\mathbf{b}(x+h)\mathbf{b}(x)\mathbf{b}(h)e(-2Mh\cdot x)| \gg \eta^{O(p^3)} \tag{8.14}$$

En efecto, para llegar de (8.13) a (8.14) hay que agregar el factor de $\mathbf{b}(h)$ por la manera en la cual modificamos los módulos, utilizando una especie de función de signo dependiendo de h. Aplicando Cauchy-Schwarz, Lema 8.2, para eliminar $\mathbf{b}(h)$, vemos que

$$|\mathbf{E}_{x,y,h\in\mathbf{F}_p^n}\mathbf{b}(y+h)\mathbf{b}(y)\mathbf{b}(x+h)\mathbf{b}(x)e(-2Mh\cdot(y-x))| \gg \eta^{O(p^3)}; \tag{8.15}$$

Haciendo el cambio de variables $z = x+y+h$, esto se convierte en

$$|\mathbf{E}_{x,y,z\in\mathbf{F}_p^n}\mathbf{b}(z,y)\mathbf{b}(z,x)e(-2M(z-x-y)\cdot(y-x))| \gg \eta^{O(p^3)}; \tag{8.16}$$

Absorbiendo todos los términos de las fases que podemos en las funciones $\mathbf{b}$, inferimos

$$|\mathbf{E}_{x,y,z\in\mathbf{F}_p^n}\mathbf{b}(z,y)\mathbf{b}(z,x)e(2\{x,y\})| \gg \eta^{O(p^3)}, \tag{8.17}$$

donde $\{x,y\}$ es la forma antisimétrica definida por $\{x,y\} = Mx\cdot y - My\cdot x$. Por el principio del palomar en z se concluye que

$$|\mathbf{E}_{x,y\in\mathbf{F}_p^n}\mathbf{b}(y)\mathbf{b}(x)e(2\{x,y\})| \gg \eta^{O(p^3)} \tag{8.18}$$

para algunas funciones acotadas $\mathbf{b}(y)$, $\mathbf{b}(x)$. Aplicando otra vez Cauchy-Schwarz, Lema 8.2, para eliminar el factor de $\mathbf{b}(x)$, deducimos que

$$|\mathbf{E}_{x,y,y'\in\mathbf{F}_p^n}\mathbf{b}(y)\mathbf{b}(y')e(2\{x,y'-y\})| \gg \eta^{O(p^3)}. \tag{8.19}$$

Por la desigualdad triangular luego obtenemos

$$\mathbf{E}_{y,y'\in\mathbf{F}_p^n}|\mathbf{E}_{x\in\mathbf{F}_p^n}e(2\{x,y'-y\})| \gg \eta^{O(p^3)}; \tag{8.20}$$

Haciendo la sustitución $h = y'-y$ concluimos

$$\mathbf{E}_{h\in\mathbf{F}_p^n}|\mathbf{E}_{x\in\mathbf{F}_p^n}e(2\{x,h\})| \gg \eta^{O(p^3)}. \tag{8.21}$$

Podemos pensar al mapeo $x\mapsto 2\{x,h\}$ como un morfismo de $\mathbf{F}_p^n$ a $\mathbf{R}/\mathbf{Z}$. Entonces si escribimos

$$W := \{h\in\mathbf{F}_p^n : \{x,h\} = 0 \text{ para todo } x\in\mathbf{F}_p^n\}$$

se tiene que W es un subespacio lineal de $\mathbf{F}_p^n$ y

$$\mathbf{E}_{x\in\mathbf{F}_p^n}e(2\{x,h\}) = 1_W(h).$$

Luego, en vista de (8.21), observamos que W es muy grande con respecto a $\mathbf{F}_p^n$:

$$|W|/p^n = |W|/|\mathbf{F}_p^n| = \mathbf{E}_{y,h\in\mathbf{F}_p^n} 1_W(h) \gg \eta^{O(p^3)}. \tag{8.22}$$

Por construcción de W vemos que M es autoadjunta en W, lo que es lo mismo que decir que

$$Mw \cdot w' = Mw' \cdot w \text{ para todo } w, w' \in W. \tag{8.23}$$

Paso 3: Eliminar la componente de la fase cuadrática.

Vamos ahora a la parte final. Volvemos a (8.13), que reescribimos como

$$|\mathbf{E}_{x,h\in\mathbf{F}_p^n} \mathbf{b}(h)\mathbf{b}(x+h)\overline{f(x)}e(-2Mh\cdot x)| \gg \eta^{O(p^3)},$$

donde distribuimos el factor de la fase $e(-\xi_0 \cdot x) = e(-\xi_0 \cdot (x+h))e(\xi_0 \cdot h)$ en las funciones $\mathbf{b}$. Acá el enfoque va a estar puesto en el factor $\overline{f(x)}$, el objetivo siendo demostrar que esta función exhibe un "sesgo" cuadrático. Primero observamos que el simple argumento de promedios del Lema 8.4 nos permite encontrar $h' \in \mathbf{F}_p^n$ tal que

$$|\mathbf{E}_{x\in G,h\in W} \mathbf{b}(h+h')\mathbf{b}(x+h+h')\overline{f(x)}e(-2M(h+h')\cdot x)| \gg \eta^{O(p^3)}.$$

Una vez más podemos absorver el $e(-2Mh'\cdot x)$ en las funciones $\mathbf{b}$, y concluir que

$$|\mathbf{E}_{x\in G,h\in W} \mathbf{b}(h)\mathbf{b}(x+h)\overline{f(x)}e(-2Mh\cdot x)| \gg \eta^{O(p^3)},$$

lo que implica que

$$|\mathbf{E}_{y\in G,x,h\in W} \mathbf{b}(h)\mathbf{b}(x+h+y)\overline{f(x+y)}e(-2Mh\cdot (x+y))| \gg \eta^{O(p^3)}.$$

Usando la desigualdad triangular podemos deducir

$$\mathbf{E}_{y\in G}|\mathbf{E}_{x,h\in W} \mathbf{b}(h,y)\mathbf{b}(x+h,y)\overline{f(x+y)}e(-2Mh\cdot x)| \gg \eta^{O(p^3)}. \tag{8.24}$$

Ahora observamos de (8.23) que tenemos la identidad

$$2Mh\cdot x = M(x+h)\cdot(x+h) - Mx\cdot x - Mh\cdot h,$$

y por ende

$$e(-2Mh\cdot x) = \mathbf{b}(x+h)\mathbf{b}(h)e(Mx\cdot x).$$

Luego (8.24) implica que

$$\mathbf{E}_{y\in G}|\mathbf{E}_{x,h\in W} \mathbf{b}(h,y)\mathbf{b}(x+h,y)\overline{f(x+y)}e(-Mx\cdot x)| \gg \eta^{O(p^3)}.$$

Aplicando el Lema 8.3 para cada $y \in G$ por separado, y con $B = B' = W$, concluimos que

$$\mathbf{E}_{y\in G}\|\overline{f(x+y)}e(Mx\cdot x)\|_{u^2(W)} \gg \eta^{O(p^3)}, \tag{8.25}$$

y por ende de la monotonía de las normas u^k de Gowers obtenemos

$$\mathbf{E}_{y\in G}\|\overline{f(x+y)}e(Mx\cdot x)\|_{u^3(W)} \gg \eta^{O(p^3)}.$$

Pero la norma $u^3(W)$ es invariante bajo modulaciones de fase cuadráticas, conjugación, y traslación, y así concluimos que

$$\mathbf{E}_{y\in G}\|f\|_{u^3(y+W)} \gg \eta^{O(p^3)}.$$

En efecto, notemos que la función $P(x) \coloneqq Mx\cdot x$ es una fase cuadrática, debido a que

$$\begin{aligned}(z\cdot\Delta)(y\cdot\Delta)P(x) &= P(x+y+z)-P(x+y)-P(x+z)+P(x)\\ &= M(x+y+z)\cdot(x+y+z)-M(x+y)\cdot(x+y)\\ &\quad -M(x+z)\cdot(x+z)+M(x)\cdot(x)\\ &= 0,\end{aligned}$$

donde la última igualdad se obtiene distribuyendo los productos internos en x, y, z, y utilizando el hecho de que la matriz M es autoadjunta en W.

Por el Teorema [14, Teorema 2.3(ii)][3] obtenemos que

$$\|f\|_{u^3(\mathbf{F}_p^n)} \geq p^{-n}\,|W|\,\mathbf{E}_{y\in G}\|f\|_{u^3(y+W)}$$

para todo $y \in \mathbf{F}_p^n$, y luego por (8.22) se tiene que

$$\|f\|_{u^3(\mathbf{F}_p^n)} \gg \eta^{O(p^3)}.$$

Por la definición de la norma u^3, esto nos da el Corolario 8.1.

[3] No enunciamos ni probamos esto aquí para no complejizar más las ideas, pero es un resultado que permite relacionar las normas locales de Gowers con las globales. También aclaramos que, si bien el resultado de Green y Tao se da en el caso de $p = 5$, no se utiliza nada especial de ese primo en particular además del hecho de ser impar, y sus argumentos pueden modificarse para cubrir el caso general de primos impares.

Bibliografía

[1] J. Bourgain and M.-C. Chang. On the size of k-fold sum and product sets of integers. *J. Amer. Math. Soc.*, 17(2):473–497, 2004. ISSN 0894-0347. doi: 10.1090/S0894-0347-03-00446-6. URL https://doi.org/10.1090/S0894-0347-03-00446-6.

[2] M.-C. Chang. Product theorems in SL_2 and SL_3. *J. Inst. Math. Jussieu*, 7 (1):1–25, 2008. ISSN 1474-7480. doi: 10.1017/S1474748007000126. URL https://doi.org/10.1017/S1474748007000126.

[3] G. Freiman. Number-theoretic studies in the markov spectrum and in the structural theory of set addition. In *Groups and the Inverse Problems of Additive Number Theory*, pages 175–183. Kalinin State University, Moscow, 1973.

[4] H. Furstenberg, Y. Katznelson, and D. Ornstein. The ergodic theoretical proof of Szemerédi's theorem. *Bull. Amer. Math. Soc. (N.S.)*, 7(3):527–552, 1982. ISSN 0273-0979. doi: 10.1090/S0273-0979-1982-15052-2. URL https://doi.org/10.1090/S0273-0979-1982-15052-2.

[5] W. T. Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998. ISSN 1016-443X. doi: 10.1007/s000390050065. URL https://doi.org/10.1007/s000390050065.

[6] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11 (3):465–588, 2001. ISSN 1016-443X. doi: 10.1007/s00039-001-0332-9. URL https://doi.org/10.1007/s00039-001-0332-9.

[7] W. T. Gowers, B. Green, F. Manners, and T. Tao. Marton's conjecture in abelian groups with bounded torsion. 2024.

[8] W. T. Gowers, B. Green, F. Manners, and T. Tao. On a conjecture of Marton. *Ann. of Math. (2)*, 201(2):515–549, 2025. ISSN 0003-486X. doi: 10.4007/annals.2025.201.2.5. URL https://doi.org/10.4007/annals.2025.201.2.5.

[9] R. M. Gray. *Entropy and information theory*. Springer, New York, second edition, 2011. ISBN 978-1-4419-7969-8; 978-1-4419-7970-4. doi: 10.1007/978-1-4419-7970-4. URL https://doi.org/10.1007/978-1-4419-7970-4.

[10] B. Green. Notes on the polynomial freiman–ruzsa conjecture. URL https://people.maths.ox.ac.uk/greenbj/papers/PFR.pdf. Nota no publicada.

[11] B. Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005. doi: 10.1017/CBO9780511734885.002. URL https://doi.org/10.1017/CBO9780511734885.002.

[12] B. Green. Approximate groups and their applications: Work of bourgain, gamburd, helfgott and sarnak. 2009. URL https://arxiv.org/abs/0911.3354.

[13] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008. ISSN 0003-486X. doi: 10.4007/annals.2008.167.481. URL https://doi.org/10.4007/annals.2008.167.481.

[14] B. Green and T. Tao. An inverse theorem for the Gowers $U^3(G)$ norm. *Proc. Edinb. Math. Soc. (2)*, 51(1):73–153, 2008. ISSN 0013-0915. doi: 10.1017/S0013091505000325. URL https://doi.org/10.1017/S0013091505000325.

[15] B. Green and T. Tao. An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm. *Math. Proc. Cambridge Philos. Soc.*, 149(1):1–19, 2010. ISSN 0305-0041. doi: 10.1017/S0305004110000186. URL https://doi.org/10.1017/S0305004110000186.

[16] B. Green, F. Manners, and T. Tao. Sumsets and entropy revisited. *Random Structures Algorithms*, 66(1):Paper No. e21252, 33, 2025. ISSN 1042-9832. doi: 10.1002/rsa.21252. URL https://doi.org/10.1002/rsa.21252.

[17] H. A. Helfgott. Growth and generation in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$. *Ann. of Math. (2)*, 167(2):601–623, 2008. ISSN 0003-486X. doi: 10.4007/annals.2008.167.601. URL https://doi.org/10.4007/annals.2008.167.601.

[18] J.-J. Liao. Improved exponent for marton's conjecture in $\mathbb{F}_2^n$. 2024.

[19] S. Lovett. An exposition of sanders' quasi-polynomial freiman–ruzsa theorem. *Theory of Computing Library Graduate Surveys*, 6, 2015.

[20] S. Lovett and O. Regev. A counterexample to a strong variant of the polynomial Freiman-Ruzsa conjecture in Euclidean space. *Discrete Anal.*, pages Paper No. 8, 6, 2017. doi: 10.19086/da.1640. URL https://doi.org/10.19086/da.1640.

[21] F. Manners. Finding a low-dimensional piece of a set of integers. *Int. Math. Res. Not. IMRN*, (15):4673–4703, 2017. ISSN 1073-7928. doi: 10.1093/imrn/rnw153. URL https://doi.org/10.1093/imrn/rnw153.

[22] A. Mudgal. An Elekes-Rónyai theorem for sets with few products. *Int. Math. Res. Not. IMRN*, (13):10410–10424, 2024. ISSN 1073-7928. doi: 10.1093/imrn/rnae087. URL https://doi.org/10.1093/imrn/rnae087.

[23] S. Peluse. Finite field models in arithmetic combinatorics—twenty years on. In *Surveys in combinatorics 2024*, volume 493 of *London Math. Soc. Lecture Note Ser.*, pages 159–199. Cambridge Univ. Press, Cambridge, 2024.

[24] D. Ramakrishnan and R. J. Valenza. *Fourier analysis on number fields*, volume 186 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999. ISBN 0-387-98436-4. doi: 10.1007/978-1-4757-3085-2. URL https://doi.org/10.1007/978-1-4757-3085-2.

[25] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953. ISSN 0024-6107. doi: 10.1112/jlms/s1-28.1.104. URL https://doi.org/10.1112/jlms/s1-28.1.104.

[26] I. Ruzsa. An application of graph theory to additive number theory. *Sci. Ser. A Math. Sci. (N.S.)*, 3:97–109, 1989. ISSN 0716-8446.

[27] I. Ruzsa. An analog of Freiman's theorem in groups. Number 258, pages xv, 323–326. 1999. Structure theory of set addition.

[28] A. Samorodnitsky. Low-degree tests at large distances. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 506–515. ACM, New York, 2007. doi: 10.1145/1250790.1250864. URL https://doi.org/10.1145/1250790.1250864.

[29] T. Sanders. On the Bogolyubov-Ruzsa lemma. *Anal. PDE*, 5(3):627–655, 2012. ISSN 2157-5045. doi: 10.2140/apde.2012.5.627. URL https://doi.org/10.2140/apde.2012.5.627.

[30] T. Tao. Sumset and inverse sumset theorems for shannon entropy, 2009. URL https://terrytao.wordpress.com/2009/06/25/sumset-and-inverse-sumset-theorems-for-shannon-entropy/. Blog post on *What's new*.

[31] T. Tao. Sumset and inverse sumset theory for Shannon entropy. *Combin. Probab. Comput.*, 19(4):603–639, 2010. ISSN 0963-5483. doi: 10.1017/S0963548309990642. URL https://doi.org/10.1017/S0963548309990642.

[32] T. Tao. *Higher order Fourier analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. ISBN 978-0-8218-8986-2. doi: 10.1090/gsm/142. URL https://doi.org/10.1090/gsm/142.

[33] T. Tao. On a conjecture of marton, 2023. URL https://terrytao.wordpress.com/2023/11/13/on-a-conjecture-of-marton/. Publicado el 13 de noviembre de 2023.

[34] T. Tao. An abridged proof of marton's conjecture, 2024. URL https://terrytao.wordpress.com/2024/06/22/an-abridged-proof-of-martons-conjecture/. Publicado el 22 de junio de 2024.

[35] T. Tao and V.H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-13656-3. Paperback edition [of MR2289012].