

Il Paradosso di Banach–Tarski*

Emanuele Paolini

15 giugno 2001

Informazioni più dettagliate sul paradosso di Banach–Tarski si possono trovare nel libro *S. Wa-
gon: The Banach–Tarski Paradox*.

Indice

1	Introduzione	1
2	Misure	2
3	Equiscindibilità	3
4	Decomposizione paradossale di un gruppo libero	6
5	Decomposizione paradossale di quasi tutta S^2	7
6	Il teorema di Bernstein	8
7	Assorbimento	9
8	Forma forte	10
9	Alcune considerazioni finali	10

1 Introduzione

Il paradosso di Banach–Tarski può essere enunciato così: “È possibile suddividere una palla in 10 parti e poi ricomporre le parti per formare due palle identiche alla prima”.

Prima di discuterlo, dobbiamo però dare una definizione precisa di cosa si intende per “palla”, per “parti” e per “ricomporre”. Quelle che seguono sono forse le definizioni più naturali che darebbe chiunque abbia una minima conoscenza di insiemistica.

Indichiamo con $B^3 \subset \mathbb{R}^3$ la palla aperta unitaria ovvero l’insieme di tutti i punti di \mathbb{R}^3 che distano meno di 1 dall’origine. Con $S^2 \subset \mathbb{R}^3$ indichiamo invece la superficie sferica, ovvero l’insieme dei punti con distanza 1 dall’origine. Introduciamo inoltre il simbolo di “unione disgiunta” $A \sqcup B$ che indica l’usuale unione $A \cup B$ di due insiemi ma, allo stesso tempo, afferma che $A \cap B = \emptyset$.

Definizione 1.1 (equidecomponibilità). *Diciamo che due sottoinsiemi $A, B \subset \mathbb{R}^n$ sono equidecomponibili, e scriviamo $A \sim B$, se è possibile trovare degli insiemi A_1, \dots, A_N e delle isometrie dirette (rototraslazioni) $\theta_1, \dots, \theta_N : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ tali che*

$$A = A_1 \sqcup \dots \sqcup A_N \quad B = \theta_1(A_1) \sqcup \dots \sqcup \theta_N(A_N).$$

Possiamo quindi enunciare il paradosso di Banach–Tarski.

*Questo documento è disponibile all’indirizzo <http://www.math.unifi.it/paolini/diletto/banach-tarski/>

Teorema 1.2 (Banach–Tarski). *La palla B^3 è equidecomponibile a due copie di se stessa: $B^3 \sim B^3 \sqcup B^3$.*

Nota: Scrivendo $B^3 \sim B^3 \sqcup B^3$ abbiamo abusato della notazione appena introdotta per il simbolo “ \sqcup ” in quanto chiaramente $B^3 \cap B^3 \neq \emptyset$. In questo caso (e in altri casi simili nel seguito) si intende che una delle due B^3 va traslata in modo che risulti disgiunta dall'altra.

2 Misure

Come mai il Teorema di Banach–Tarski ci sembra paradossale? Perché abbiamo l'idea che ci sia una funzione che ad ogni sottoinsieme di \mathbb{R}^3 associa il suo volume, e che questa funzione si debba conservare per equidecomposizione.

Quello che sarebbe comodo poter avere è una misura invariante definita sulle parti di \mathbb{R}^n , cioè una funzione $\mu : \mathcal{P}(\mathbb{R}^n) \rightarrow [0, +\infty]$ con le seguenti proprietà:

- (i) $\mu(\bigsqcup_{k \in \mathbb{N}} A_k) = \sum_{k \in \mathbb{N}} \mu(A_k)$ (numerabile additività);
- (ii) $\mu(\theta(A)) = \mu(A)$ se $\theta : \mathbb{R}^n \rightarrow \mathbb{R}^n$ è un'isometria (invarianza);
- (iii) $0 < \mu(B^n) < \infty$ (non banalità).

Una conseguenza del paradosso di Banach–Tarski è che questa misura su \mathbb{R}^3 non esiste. In realtà, in queste ipotesi, c'è una dimostrazione più semplice della non esistenza di una misura siffatta, che vale in generale per ogni dimensione $n \in \mathbb{N}$.

Teorema 2.1 (Vitali). *È possibile trovare degli insiemi disgiunti $C_k \subset [0, 1)$ tutti equidecomponibili allo stesso insieme T e tali che*

$$[0, 1) = \bigsqcup_{k \in \mathbb{N}} C_k.$$

Se esistesse una misura invariante sulle parti di \mathbb{R} seguirebbe $\mu([0, 1)) = \sum_{k \in \mathbb{N}} \mu(C_k) = \sum_{k \in \mathbb{N}} \mu(T)$, se fosse $\mu(T) = 0$ allora avremmo $\mu([0, 1)) = 0$ se invece $\mu(T) > 0$ si avrebbe $\mu([0, 1)) = \infty$.

In ogni caso μ sarebbe una misura banale.

Vediamo la dimostrazione di questo teorema, (meglio conosciuto come *esempio di Vitali*) che ha alcune idee in comune con la dimostrazione del paradosso di Banach–Tarski.

Dimostrazione del Teorema di Vitali. Diciamo che due numeri in $[0, 1)$ sono equivalenti se la loro differenza è razionale. L'assioma della scelta ci assicura che è possibile trovare un insieme $T \subset [0, 1)$ che contiene esattamente un elemento per ogni classe di equivalenza. Esiste cioè T con la proprietà che se $x, y \in T$ allora $x - y$ non è razionale, inoltre per ogni numero $t \in [0, 1)$ esiste $x \in T$ tale che $x - t$ è razionale.

Consideriamo ora una numerazione q_k dei razionali in $[0, 1)$ cioè $\mathbb{Q} \cap [0, 1) = \{q_k : k \in \mathbb{N}\}$. Definiamo poi

$$\begin{aligned} A_k &= (T + q_k) \cap [0, 1) \\ B_k &= (T + q_k - 1) \cap [0, 1) \\ C_k &= A_k \sqcup B_k. \end{aligned}$$

Chiaramente $T \sim C_k = A_k \sqcup B_k$. Inoltre, per costruzione, si verifica facilmente che $[0, 1) = \bigsqcup_{k \in \mathbb{N}} C_k$. □

Questo esempio si può poi estendere facilmente ad ogni dimensione in quanto

$$[0, 1)^n = [0, 1) \times [0, 1)^{n-1} = \bigsqcup_{k \in \mathbb{N}} C_k \times [0, 1)^{n-1}.$$

Il paradosso di Banach–Tarski, invece, può essere dimostrato in \mathbb{R}^n solo se $n \geq 3$. Questo paradosso ha delle conseguenze più forti rispetto all'esempio di Vitali, in quanto implica la non esistenza di una misura invariante μ sulle parti di \mathbb{R}^n nemmeno se al posto della *numerabile additività* si richiede solo l'*additività finita*:

$$\mu(A \sqcup B) = \mu(A) + \mu(B).$$

È stato invece dimostrato che su \mathbb{R} ed \mathbb{R}^2 esistono effettivamente delle misure invarianti finitamente additive.

Il fatto di non poter avere delle misure invarianti definite su tutti i sottoinsiemi di \mathbb{R}^n , oltre ad essere controintuitiva, è una cosa piuttosto spiacevole in teoria della misura. La soluzione consiste nel definire la misura μ soltanto su di una classe ristretta di insiemi, che verranno chiamati *insiemi misurabili*, per i quali valgano le proprietà di cui sopra.

Ritornando al paradosso di Banach–Tarski, segue che le parti in cui viene divisa la sfera non possono essere tutte misurabili.

3 Equiscindibilità

La nozione di “equidecomponibilità” data sopra è molto semplice e naturale per chi è abituato a pensare che una figura geometrica è l'insieme dei propri punti. Questo concetto però, è relativamente recente (1800 circa). Basti pensare a come Euclide aveva invece posto gli assiomi della geometria euclidea: rette e punti sono entità separate, l'appartenenza di un punto ad una retta è una relazione tra oggetti distinti. D'altra parte per poter fare geometria con gli insiemi di punti occorre conoscere l'insieme \mathbb{R} dei numeri reali, che già contiene in sé una struttura talmente ricca da essere, in un certo senso, paradossale.

Pensiamo ad esempio al rettangolo $[0, 2] \times [0, 1] \subset \mathbb{R}^2$. La nostra idea intuitiva di rettangolo 2×1 ci dice che è possibile tagliarlo a metà per ottenere due quadrati 1×1 . Ma dal punto di vista insiemistico come posso fare questa “decomposizione”? Gli insiemi $[0, 1] \times [0, 1]$ e $[1, 2] \times [0, 1]$ non sono una decomposizione perché hanno intersezione non vuota data da $\{1\} \times [0, 1]$. In questo senso non è chiaro se $[0, 2] \times [0, 1]$ sia equidecomponibile a $[0, 1] \times [0, 1] \sqcup [2, 3] \times [0, 1]$.

Vediamo ora come questa decomposizione può essere effettivamente fatta, sebbene i pezzi usati siano piuttosto “brutti”.

L'idea fondamentale è innanzitutto quella di sfruttare il “paradosso” intrinseco degli insiemi infiniti: “Un insieme infinito può essere messo in corrispondenza biunivoca con una sua parte propria”. L'esempio più semplice è dato dal più piccolo insieme infinito, \mathbb{N} i numeri naturali, che può essere messo in corrispondenza con $\mathbb{N} + 1 = \mathbb{N} \setminus \{0\}$ mediante una traslazione. Questa corrispondenza è addirittura isometrica, quindi \mathbb{N} e $\mathbb{N} \setminus \{0\}$ risultano essere equidecomponibili (tramite un unico pezzo).

Come si può sfruttare questa idea nel caso del nostro rettangolo? Notiamo che per provare che $[0, 2] \times [0, 1] \sim [0, 1] \times [0, 1] \sqcup [2, 3] \times [0, 1]$, sarà sufficiente provare che $[0, 1] \times [0, 1] \sim [0, 1] \times [0, 1]$. Quello che dobbiamo riuscire a fare è “assorbire” il segmento $\{1\} \times [0, 1]$ nella decomposizione, così come abbiamo fatto con il numero 0 per i naturali.

Per trovare una struttura simile alle traslazioni di \mathbb{N} su un insieme limitato come $[0, 1] \times [0, 1]$ ricorriamo alle rotazioni. Consideriamo la trasformazione $\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ che rappresenta una rotazione di centro $(1/2, 1/2)$ di un angolo irrazionale rispetto a π . L'idea è che il semigrupp (non ci sono gli elementi inversi, come in \mathbb{N}) delle rotazioni $\{\theta^n : n \in \mathbb{N}\}$ è isomorfo a \mathbb{N} stesso. Infatti $\theta^n = \theta^m$ implica $\theta^{n-m} = Id$ che significa che l'angolo di rotazione di θ^{n-m} è multiplo di 2π il che, per la nostra scelta dell'angolo, può accadere solo quando $n = m$. Consideriamo dunque il segmento $T = \{1/2\} \times]0, 1/3[$ e l'insieme $D = \bigcup_{n \in \mathbb{N}} \theta^n(T)$. Siccome su T non ci sono punti fissi di θ si vede facilmente che $\theta(D) = D \setminus T$, iterando si ottiene

$$\theta^3(D) = D \setminus (T \cup \theta(T) \cup \theta^2(T)).$$

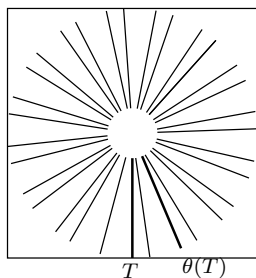


Figura 1: L'assorbimento di un segmento.

Dunque potendo dividere il segmento $\{1\} \times [0, 1]$ in tre segmenti isometrici a T , si ottiene che

$$\begin{aligned}
[0, 1] \times [0, 1] &= [0, 1] \times [0, 1] \setminus D \sqcup D \\
&= [0, 1] \times [0, 1] \setminus D \sqcup \{1\} \times [0, 1] \sqcup (1, 1) \sqcup T \sqcup \theta^3(D) \\
&\sim [0, 1] \times [0, 1] \setminus D \sqcup \{(1, 1)\} \sqcup T \sqcup \theta(T) \sqcup \theta^2(T) \sqcup \theta^3(D) \\
&= [0, 1] \times [0, 1] \setminus D \sqcup \{(1, 1)\} \sqcup D \\
&\sim [0, 1] \times [0, 1] \sqcup \{(1, 1)\}
\end{aligned}$$

(si veda la Figura 1). Ci avanza dunque il solo punto $(1, 1)$ che può essere “riassorbito” in maniera analoga al segmento, considerando, ad esempio, l’orbita tramite θ del punto $(1/2, 3/8)$, i dettagli sono lasciati al lettore.

L’equidecomposizione appena fatta si basa su tecniche “moderne”. Un concetto più “classico” che rappresenta meglio l’idea del “tagliare con le forbici” è invece il seguente.

Definizione 3.1. *Diciamo che due n -poliedri (segmenti se $n = 1$, poligoni se $n = 2$) $A, B \subset \mathbb{R}^n$ sono equiscindibili se è possibile trovare degli n -poliedri A_1, \dots, A_N e delle isometrie dirette $\theta_1, \dots, \theta_N$ tali che*

$$A = A_1 \cup \dots \cup A_N \quad B = \theta_1(A_1) \cup \dots \cup \theta_N(A_N)$$

e tali che le intersezioni $A_i \cap A_j$, $\theta_i(A_i) \cap \theta_j(A_j)$ siano contenute in piani $(n - 1)$ -dimensionali (ovvero abbiano parte interna vuota).

L’equiscindibilità viene comunemente usata per “dimostrare” elementarmente le formule usate per calcolare le aree dei poligoni nel piano. Si assume a priori che l’area di due figure equiscindibili sia la stessa e si riconduce l’area di un poligono qualunque a quella di un rettangolo.

Ad esempio, un poligono regolare di perimetro p e apotema a è equiscindibile ad un rettangolo di base $p/2$ e altezza a , da cui la formula per l’area di un poligono regolare (Figura 2).

Si può dimostrare senza troppe difficoltà che due poligoni con la stessa area sono equiscindibili (in realtà è addirittura possibile equiscindere i due poligoni usando solo traslazioni). Innanzitutto un poligono può essere scisso in triangoli (è in un certo senso la definizione di poligono). Poi ogni triangolo è equiscindibile ad un rettangolo. Una volta che ho ottenuto dal poligono una unione disgiunta di rettangolini, posso trasformare tutti i rettangolini in rettangoli con la stessa altezza (due rettangoli con la stessa area sono equiscindibili) e formare quindi un unico rettangolo che è equiscindibile ad un quadrato con la stessa area (Figura 3).

Un fatto sorprendente è l’esistenza del seguente risultato, in un certo senso opposto a quello del paradosso di Banach–Tarski.

Teorema 3.2 (Dehn). *Un tetraedro e un cubo dello stesso volume non sono equiscindibili.*

La dimostrazione di questo teorema si basa sull’esistenza di funzioni additive non lineari. Come per l’esempio di Vitali e per il paradosso di Banach–Tarski, anche qui serve l’assioma della scelta (il lettore attento scoprirà dove, sebbene non venga esplicitamente evidenziato).

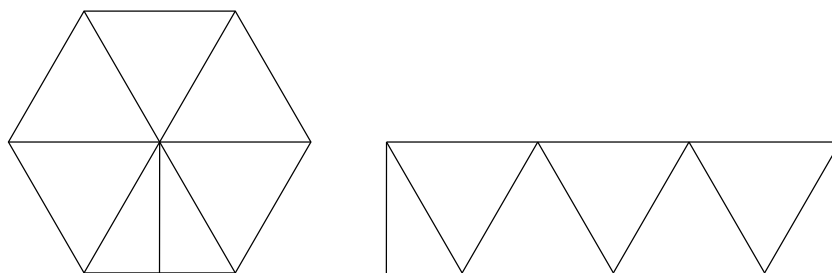


Figura 2: La trasformazione di un poligono regolare in un rettangolo.

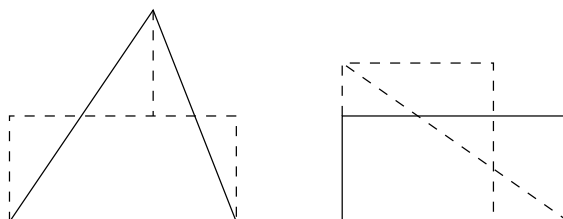


Figura 3: Un triangolo è equiscindibile ad un rettangolo, e un rettangolo ad un quadrato (se il rettangolo è troppo stretto prima lo devo dividere a metà e “ingrossarlo”).

Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione tale che $f(x + y) = f(x) + f(y)$ (funzione additiva). Ci chiediamo se è vero che $f(x) = xf(1)$, cioè f è una funzione lineare.

Notiamo che $f(2) = f(1) + f(1) = 2f(1)$ e per induzione $f(n) = f(1) + f(n - 1) = nf(1)$. Più in generale $f(n(x/n)) = nf(x/n)$ e quindi $f(x/n) = f(x)/n$. Questo ci permette di concludere che $f((p/q)x) = (p/q)f(x)$ ovvero la funzione f è \mathbb{Q} -lineare. Viceversa, ogni funzione \mathbb{Q} -lineare è ovviamente additiva.

Se allora consideriamo lo spazio vettoriale $\mathbb{R}_{\mathbb{Q}}$ (cioè \mathbb{R} visto come spazio vettoriale sul campo \mathbb{Q}) possiamo definire una funzione lineare su $\mathbb{R}_{\mathbb{Q}}$ fissandone a piacere i valori su una base. Poiché esiste almeno una base che contiene i vettori indipendenti 1 e (ad esempio) π , imponendo $f(1) = 1$, $f(\pi) \neq \pi$ e a piacere sui restanti elementi della base, si ottiene una funzione additiva su \mathbb{R} ma non lineare.

Tornando al teorema di Dehn, l'idea è quella di costruire una funzione F che associa ad ogni poliedro un numero reale, e che sia invariante per equiscissione, cioè se A e B equiscindono C si ha $F(C) = F(A) + F(B)$. Se troviamo una tale funzione che assuma valori diversi sul cubo e sul tetraedro il teorema è dimostrato.

Dimostrazione del Teorema di Dehn. Consideriamo la seguente funzione F

$$F(P) = \sum_s f(\alpha_s)l(s)$$

dove P è un generico poliedro, la somma è fatta su tutti gli spigoli s di P , α_s è l'angolo formato dalle due facce di P che si appoggiano su s , $l(s)$ è la lunghezza dello spigolo s ed f è una qualunque funzione additiva tale che $f(\pi) = 0$ (da quanto detto prima si ha allora $f(\pi/2) = f(2\pi) = 0$).

Verifichiamo ora che F è invariante per equiscissione. Sia P equiscisso in P_1, \dots, P_N , quando andiamo a considerare uno spigolo s di P , questo spigolo, sarà diviso in un numero finito di spigoli dei pezzi P_i . La lunghezza di s però sarà esattamente la somma della lunghezza dei vari pezzi, e l'angolo α_s sarà la somma degli angoli dei pezzi che si appoggiano su s . D'altra parte i nuovi spigoli che si sono venuti a creare nella equiscissione, non contribuiscono a F in quanto la somma degli angoli su tali spigoli sarà 2π se lo spigolo è interno a P e π se gli spigoli si trovano sulle facce

di P . In ogni caso dopo aver sommato otteniamo $f(\pi) = f(2\pi) = 0$. Segue che F è invariante per equiscissione.

Ora F calcolata sul cubo dà 0 in quanto gli angoli sono tutti $\pi/2$ ma $f(\pi/2) = 0$. Gli angoli tre le facce del tetraedro misurano invece $\arccos(1/3)$ ed essendo $\arccos(1/3)$ irrazionale con π è possibile scegliere f in modo che $f(\arccos(1/3)) \neq 0$, da cui l'assurdo se fossero equiscindibili. \square

Il teorema di Dehn è dunque dimostrato. Si noti che ciò implica che, a differenza di quanto succede nel piano, dove l'area di un qualunque poligono può essere ricondotta all'area del rettangolo, nello spazio non c'è un modo analogo per "dimostrare" la formula del volume di una piramide: "Area di base per altezza diviso tre".

Di fatto la dimostrazione, o meglio, la giustificazione di tale formula richiede l'uso degli integrali.

4 Decomposizione paradossale di un gruppo libero

La possibilità di dimostrare il paradosso di Banach–Tarski si basa sull'esistenza di un sottogruppo libero generato da due elementi nel gruppo $SO(3)$ delle rotazioni nello spazio \mathbb{R}^3 .

Consideriamo due rotazioni $\theta, \varphi \in SO(3)$ (cioè due rotazioni dello spazio con centro l'origine) e consideriamo poi il gruppo G generato da queste due rotazioni, ovvero l'insieme di tutte le rotazioni che possono essere ottenute mediante la composizione di $\theta, \varphi, \theta^{-1}$ e φ^{-1} . Ogni elemento di G può dunque essere rappresentato da una "parola" scritta con le "lettere" $\theta, \theta^{-1}, \varphi$ e φ^{-1} .

Ci chiediamo se è possibile, a forza di comporre tra loro $\theta, \varphi, \theta^{-1}, \varphi^{-1}$ ottenere l'identità e . Chiaramente la risposta, in questo contesto, è sì, ad esempio $\theta\theta^{-1} = e$ ma anche $\theta\varphi\varphi^{-1}\theta^{-1} = \theta\theta^{-1} = e$. Dunque per rendere la cosa più consistente, richiediamo che se durante la composizione si usa una rotazione (ad esempio θ) allora al passo immediatamente successivo non si può usare la sua inversa (θ^{-1}). Ciò richiediamo che le parole non contengano mai consecutivamente una lettera e la sua inversa. Con questa restrizione, la possibilità di ottenere l'identità dipende dalla scelta delle rotazioni θ e φ . Nel caso in cui non sia possibile ottenere l'identità diciamo che G è *liberamente generato* da θ e φ o il *gruppo libero* su θ e φ (per essere precisi l'identità di fatto è la composizione di 0 rotazioni, quello che si richiede è che non sia possibile ottenerla con un numero positivo di rotazioni).

Notiamo poi che se il gruppo G è liberamente generato, allora ogni elemento di G si scrive in modo *unico* come composizione di $\theta, \theta^{-1}, \varphi$ e φ^{-1} ferma restando l'imposizione di non usare mai di seguito una rotazione e la sua inversa. Ad esempio se fosse $\theta\theta\varphi\theta = \theta\varphi\theta\varphi$ otterrei innanzitutto $\theta\varphi\theta = \varphi\theta\varphi$ (ho moltiplicato a sinistra per θ^{-1}) e quindi $\varphi^{-1}\theta^{-1}\varphi^{-1}\theta\varphi\theta = \varphi^{-1}\theta^{-1}\varphi^{-1}\varphi\theta\varphi = e$.

Si noti che se ci restringiamo al gruppo delle rotazioni del piano $SO(2)$ non è possibile trovare un sottogruppo libero. Infatti date due rotazioni $\theta, \varphi \in SO(2)$ si ha sempre $\theta\varphi = \varphi\theta$ e quindi $\theta^{-1}\varphi^{-1}\theta\varphi = e$.

Per quanto riguarda $SO(3)$ (e in generale $SO(n)$ con $n \geq 3$) è possibile trovare due rotazioni θ e φ che generano un gruppo libero. Una possibile scelta è data dalle rotazioni attorno agli assi z e x di un angolo pari a $\arccos \frac{1}{3}$ cioè (in notazione matriciale)

$$\theta = \begin{pmatrix} \frac{1}{3} & -\frac{2\sqrt{2}}{3} & 0 \\ \frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \varphi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & -\frac{2\sqrt{2}}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{3} \end{pmatrix}.$$

La verifica del fatto che queste due rotazioni sono "indipendenti" (cioè generano un gruppo libero) è elementare e lasciata al lettore interessato.

Sia allora G il gruppo libero generato da questi due elementi θ e φ . Una possibile rappresentazione grafica di G è quella data in Figura 4. Il punto al centro è l'elemento neutro e , ogni moltiplicazione a destra per θ mi fa procedere verso destra, per θ^{-1} verso sinistra, una moltiplicazione (a destra) per φ mi fa procedere verso l'alto e per φ^{-1} verso il basso. La moltiplicazione a sinistra è leggermente più complicata. Se moltiplico un elemento g a sinistra per θ , (cioè voglio calcolare

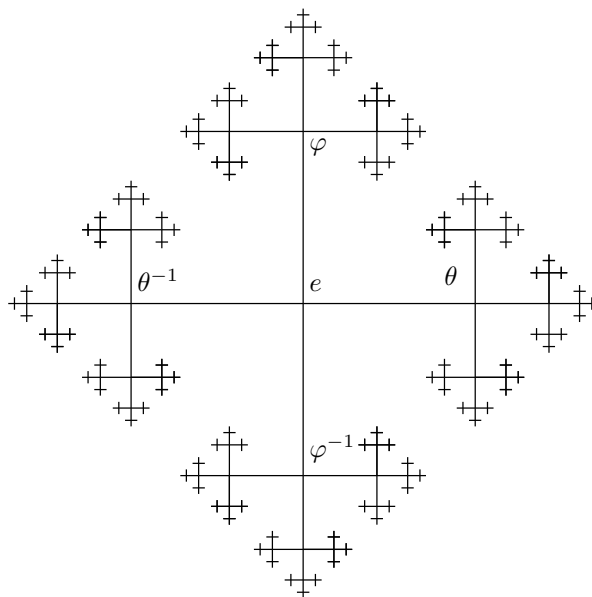


Figura 4: Una rappresentazione grafica del gruppo libero generato da due elementi θ, φ .

θg), devo considerare l'intero ramo che congiunge e a g e spostarlo tutto a destra riscalandolo opportunamente.

Questa rappresentazione di G è chiaramente frattale ed autosimile, dunque si presta ad essere utilizzata per decomporre G in copie di sè stesso. Si consideri ad esempio l'insieme G_θ dato dall'insieme degli elementi di G che cominciano con la lettera θ . Questo insieme, nella rappresentazione grafica, è il "ramo di destra" uscente da e . Allo stesso modo consideriamo gli insiemi $G_{\theta^{-1}}, G_\varphi$ e $G_{\varphi^{-1}}$ che sono gli altri tre rami del grafico. Si ottiene dunque la seguente decomposizione

$$G = \{e\} \sqcup G_\theta \sqcup G_{\theta^{-1}} \sqcup G_\varphi \sqcup G_{\varphi^{-1}}.$$

Ora cerchiamo invece di capire come è fatto l'insieme $\theta G_{\theta^{-1}}$ (cioè l'insieme delle parole di $G_{\theta^{-1}}$ moltiplicate a sinistra per θ). Basta prendere l'insieme $G_{\theta^{-1}}$ e spostare il "gambo" che lo congiunge ad e sull'elemento θ . Quello che si ottiene è l'insieme di tutte le parole che *non* iniziano per θ . Abbiamo allora ottenuto la seguente decomposizione

$$G = G_\theta \sqcup \theta G_{\theta^{-1}}$$

e analogamente

$$G = G_\varphi \sqcup \varphi G_{\varphi^{-1}}.$$

Questa decomposizione può essere facilmente verificata anche senza il supporto del grafico, basta notare che una qualunque parola g di G o inizia con la lettera θ (e quindi sta in G_θ) oppure può essere scritta come $g = \theta \theta^{-1} g$ dove $\theta^{-1} g$ è effettivamente una parola che inizia per θ^{-1} (non ci può essere cancellazione perché g non inizia per θ).

In conclusione abbiamo che G contiene quattro "pezzi" $G_\theta, G_{\theta^{-1}}, G_\varphi$ e $G_{\varphi^{-1}}$ che possono essere ricomposti (usando le traslazioni a sinistra del gruppo) in due copie di G stesso.

5 Decomposizione paradossale di quasi tutta \mathbb{S}^2 .

Vogliamo sfruttare questa decomposizione di G per ottenere una decomposizione di \mathbb{S}^2 ovvero la superficie sferica di raggio unitario e centro l'origine in \mathbb{R}^3 .

Innanzitutto consideriamo l'insieme $D' \subset \mathbb{S}^2$ formato dall'intersezione con \mathbb{S}^2 di tutti gli assi di rotazione degli elementi di G . Siccome G è un insieme numerabile (posso mettere tutte le parole che rappresentano elementi di G in ordine alfabetico e numerarle una alla volta) anche D' è numerabile. Definiamo poi D come l'insieme di tutte le immagini dei punti di D' tramite una qualunque rotazione di G (cioè $D = G(D')$). Anche D sarà un insieme numerabile. Inoltre il gruppo G agisce su $\mathbb{S}^2 \setminus D$, nel senso che preso un punto $x \in \mathbb{S}^2 \setminus D$ e una rotazione $g \in G$ allora anche il punto $g(x)$ appartiene a $\mathbb{S}^2 \setminus D$. Il fatto che g sia una rotazione ci assicura che $g(x) \in \mathbb{S}^2$; inoltre se fosse $g(x) \in D$ allora sarebbe $g(x) = h(z)$ per qualche $z \in D'$ e $h \in G$ e quindi anche $x = g^{-1}h(z)$ sarebbe elemento di D .

Dato un punto $x \in \mathbb{S}^2 \setminus D$ definiamo l'orbita $Orb(x)$ come l'insieme dei punti $g(x)$ al variare di $g \in G$. Chiaramente due punti x, y appartengono alla stessa orbita, $Orb(x) = Orb(y)$, se e solo se esiste un elemento $g \in G$ tale che $g(x) = y$.

L'insieme delle orbite risulta essere una partizione di $\mathbb{S}^2 \setminus D$. In base all'assioma della scelta possiamo trovare una "sezione" della partizione, ovvero un insieme $M \subset \mathbb{S}^2 \setminus D$ che contiene uno ed un solo punto per ogni orbita. Dunque ogni elemento $x \in M$ ha la proprietà che $g(x) \notin M$ qualunque sia $g \in G$. Inoltre ogni orbita interseca M e quindi dato un qualunque punto $y \in \mathbb{S}^2 \setminus D$ esistono $x \in M$ e $g \in G$ tali che $g(x) = y$.

Se g è una rotazione in G , l'insieme $g(M)$ è l'insieme di tutti i punti di M ruotati secondo g (ovvero l'immagine di M tramite la rotazione g). Vogliamo dimostrare ora che se g, h sono due diverse rotazioni di G allora $g(M)$ e $h(M)$ sono disgiunti. Supponiamo per assurdo che esistano $x, y \in M$ tali che $g(x) = h(y)$. Se $x = y$ allora avrei $h^{-1}g(x) = x$ ma questo è impossibile perché significherebbe che x sta nell'asse di rotazione di $h^{-1}g$ e quindi $x \in D$. Se invece $x \neq y$ avrei $h^{-1}g(x) = y$ e quindi x e y dovrebbero stare nella stessa orbita cosa che ho escluso essendo $x, y \in M$.

Se H è un sottoinsieme di G posso definire inoltre $H(M)$ come l'unione di $h(M)$ al variare di h in H . Per quanto detto prima si ha $G(M) = \mathbb{S}^2 \setminus D$.

Riprendendo ora la decomposizione di G trovata nella sezione precedente, ottengo che gli insiemi

$$A_1 = G_\theta(M), A_2 = G_{\theta^{-1}}(M), A_3 = G_\varphi(M), A_4 = G_{\varphi^{-1}}(M)$$

sono a due a due disgiunti, e inoltre si ha

$$A_1 \sqcup \theta A_2 = \mathbb{S}^2 \setminus D, \quad A_3 \sqcup \varphi A_4 = \mathbb{S}^2 \setminus D.$$

Abbiamo così ottenuto una decomposizione di un sottoinsieme di $\mathbb{S}^2 \setminus D$ (ricordiamo che $\mathbb{S}^2 \setminus D = A_1 \sqcup A_2 \sqcup A_3 \sqcup A_4 \sqcup M$) in due copie identiche di $\mathbb{S}^2 \setminus D$.

La definizione seguente ci permetterà di scrivere il risultato appena ottenuto così: $\mathbb{S}^2 \setminus D \sqcup \tau(\mathbb{S}^2 \setminus D) \preceq \mathbb{S}^2 \setminus D$ (dove τ è una traslazione).

Definizione 5.1. Diciamo che $A \preceq B$ se A è equidecomponibile ad un sottoinsieme di B ovvero esiste un insieme B' tale che $A \sim B' \subset B$.

Ci sarà dunque utile avere a disposizione il seguente teorema, che dimostreremo nella prossima sezione.

Teorema 5.2. Se $A \preceq B$ e $B \preceq A$ allora $A \sim B$.

6 Il teorema di Bernstein

Per avere una visione più intuitiva del teorema precedente, il seguente esempio dovrebbe essere piuttosto chiarificatore.

Allarghiamo per un momento la classe di trasformazioni ammissibili, aggiungendo alle isometrie anche le omotetie. La questione che ci poniamo è: "È possibile suddividere un cerchio in un numero finito di parti e ricomporle usando isometrie e omotetie in modo da ottenere un quadrato?"

A prima vista può sembrare molto difficile risolvere questa questione, ma in realtà la soluzione (affermativa) è piuttosto semplice. L'idea fondamentale è che è possibile rimpicciolire il quadrato e

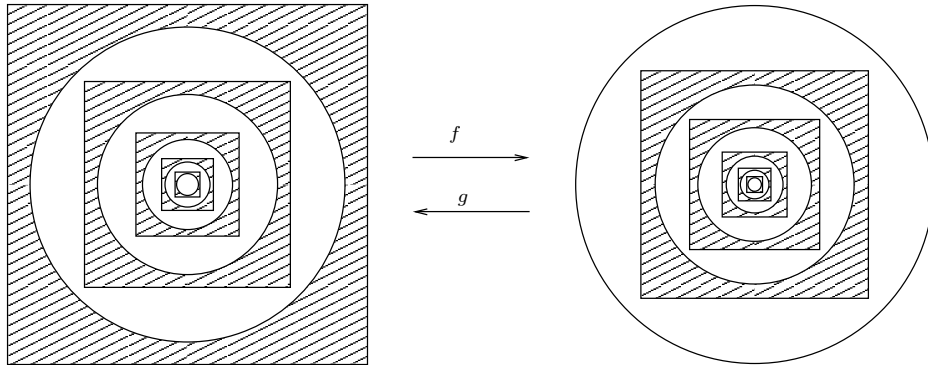


Figura 5: Come decomporre un cerchio in due parti da ricomporre (mediante omotetie) in un quadrato.

mandarlo in un sottoinsieme del cerchio (tramite un'omotetia f) e viceversa è possibile rimpicciolire il cerchio e mandarlo in un sottoinsieme del quadrato (tramite g). Iterando più volte (infinite volte!) f e g otteniamo gli insiemi inscatolati come in Figura 5. È chiaro dunque che la parte ombreggiata del quadrato è omotetica (tramite f) alla parte ombreggiata nel cerchio, e viceversa la parte bianca del quadrato è omotetica (tramite g) alla parte bianca del cerchio.

Il fatto di avere usato omotetie come funzioni f e g non è rilevante in questo esempio. Se come f e g prendiamo due funzioni iniettive qualunque abbiamo dimostrato il seguente teorema, fondamentale nello studio della cardinalità. Se A e B sono insiemi si dice che A e B hanno la stessa cardinalità, $\#A = \#B$, se esiste una funzione bigettiva $f : A \rightarrow B$. Si dice che A ha cardinalità più piccola di B , $\#A \leq \#B$, se esiste una funzione iniettiva $f : A \rightarrow B$.

Teorema 6.1 (Bernstein). *Siano $f : A \rightarrow B$ e $g : B \rightarrow A$ funzioni iniettive. Allora esiste $h : A \rightarrow B$ bigettiva. Ovvero*

$$\#A \leq \#B, \#B \leq \#A \Rightarrow \#A = \#B.$$

In particolare il precedente teorema permette di dimostrare facilmente che due insiemi di \mathbb{R}^n con parte interna non vuota possono essere messi in corrispondenza biunivoca.

Ora la dimostrazione del Teorema 5.2 si svolge nello stesso modo. Se $A \preceq B$ allora esiste una funzione iniettiva $f : A \rightarrow B$, (f sarà definita sui pezzi A_i della decomposizione come l'isometria θ_i) e viceversa se $B \preceq A$ sarà definita una funzione iniettiva $g : B \rightarrow A$. Si possono dividere quindi A e B in due parti, come nel caso del quadrato e del cerchio. Si prendono poi i pezzi A_i di A e si intersecano con la parte "ombreggiata" di A . Questi nuovi pezzi vengono mandati (tramite f che è una isometria su ogni singolo pezzo) esattamente nella parte ombreggiata di B . Viceversa la parte bianca di B si decompone coi pezzi B_i di B e viene mandata tramite g nella parte bianca di A . Questa idea può essere effettivamente formalizzata in una dimostrazione del teorema (esercizio per il lettore).

7 Assorbimento

Ricordando quanto ottenuto nelle sezioni precedenti, siamo arrivati alla seguente decomposizione paradossale

$$\mathbb{S}^2 \setminus D \sim \mathbb{S}^2 \setminus D \sqcup \tau(\mathbb{S}^2 \setminus D).$$

Il problema di eliminare l'insieme D è un problema di "assorbimento" come quello che abbiamo avuto per dimostrare che il rettangolo può essere diviso in due parti uguali.

Siccome D è numerabile, esisterà sicuramente una rotazione $\sigma \in SO(3)$ tale che $\sigma^n(D) \cap \sigma^m(D) = \emptyset$. Infatti è possibile fissare a priori un asse di rotazione, e considerare l'insieme X degli

angoli assunti (rispetto all'asse di rotazione fissato) da due qualunque punti di D . Siccome D è numerabile anche X è numerabile, e quindi esisterà un angolo che non sta in X . Preso ora $T = \bigsqcup_{k \in \mathbb{N}} \sigma^k(D)$ si trova che

$$T = D \sqcup \bigsqcup_{k \in \mathbb{N}} \sigma^{k+1}(D) = D \sqcup \sigma(T)$$

cioè $T \setminus D \sim T$. Dunque

$$\mathbb{S}^2 \setminus D = \mathbb{S}^2 \setminus T \sqcup T \setminus D \sim \mathbb{S}^2 \setminus T \sqcup T = \mathbb{S}^2$$

cioè $\mathbb{S}^2 \setminus D \sim \mathbb{S}^2$.

In definitiva, abbiamo trovato una decomposizione paradossale di \mathbb{S}^2

$$\mathbb{S}^2 \sim \mathbb{S}^2 \sqcup \tau(\mathbb{S}^2)$$

(per una certa traslazione τ).

La conclusione è ormai facile. Prendendo i coni formati dai pezzi della decomposizione di \mathbb{S}^2 con vertice nell'origine, si ottiene subito una decomposizione paradossale di $B^3 \setminus \{0\}$. Non ci resta che “assorbire” il punto 0.

Per far questo prendiamo come prima una qualunque rotazione ξ tale che $\xi^n(0) \neq \xi^m(0)$ e consideriamo l'orbita $T = \bigsqcup_{n \in \mathbb{N}} \xi^n(0)$. Come prima si ottiene $T \sim T \setminus \{0\}$ e quindi si ottiene facilmente la decomposizione cercata $B^3 \setminus \{0\} \sim B^3$ e quindi la tesi.

8 Forma forte

È molto semplice ora generalizzare il paradosso di Banach–Tarski nella seguente forma forte.

Teorema 8.1. *Se A e B sono due insiemi limitati di \mathbb{R}^3 con parte interna non vuota allora $A \sim B$.*

Essendo riusciti a duplicare la sfera, possiamo iterare il procedimento, ottenendo $B^3 \sim B^3 \sqcup B^3 \sim \dots \sim B^3 \sqcup \dots \sqcup B^3$.

Il fatto che A ha parte interna non vuota significa che A contiene una palletta B_ε , d'altra parte, essendo B limitato, B può essere ricoperto da un numero finito di pallette B_ε , e quindi $B \preceq B_\varepsilon \sqcup \dots \sqcup B_\varepsilon \sim B_\varepsilon$. D'altra parte lo stesso ragionamento vale scambiando A e B da cui $A \sim B_\varepsilon$ (supponiamo che anche B contenga una palletta di raggio ε in caso contrario bisognava scegliere all'inizio un ε più piccolo che andasse bene per entrambi) e quindi $A \sim B$.

9 Alcune considerazioni finali

L'assioma della scelta AC , su cui si basano pesantemente i risultati precedenti, è indipendente dagli altri assiomi della teoria degli insiemi di Zermelo–Fraenkel ZF . Questo da un lato significa che AC non può essere dimostrato in ZF dall'altro significa che se ZF è coerente anche $ZF + AC$ è coerente. In pratica un matematico può decidere a suo piacere se assumere che valga AC oppure no.

Abbiamo visto che AC ha molte “spiacevoli” e strane conseguenze: il paradosso di Banach–Tarski, la non esistenza di misure invarianti sulle parti di \mathbb{R}^n , la non equiscindibilità di cubo e tetraedro. Ciò porterebbe a pensare che sarebbe molto ragionevole non accettare AC come assioma.

Questo è vero solo in parte. In primo luogo AC ha conseguenze fondamentali in vari campi della matematica (ad esempio in teoria dei gruppi o in analisi funzionale), ma a prescindere da questo, il fatto di non prendere AC come assioma non aiuta molto, infatti senza AC non potremmo comunque dimostrare che esiste una misura invariante sulle parti di \mathbb{R}^n o che non è possibile

duplicare la sfera... semplicemente queste affermazioni sarebbero “non decidibili” in ZF . D'altra parte sarebbe anche inutile prendere la negazione di AC come assioma (cosa che sembra, tra l'altro, molto poco naturale), perché anche questo non sarebbe sufficiente a rendere decidibili tali affermazioni.

Una possibilità sarebbe invece prendere invece come assioma l'esistenza di una misura invariante sulle parti di \mathbb{R}^n . Questa possibilità è contemplata nel libro citato all'inizio, e porta a svariate conseguenze, in particolare in teoria degli insiemi, che dato il loro contenuto tecnico non è possibile discutere in questa sede. Di certo in tal caso il Teorema di Banach–Tarski sarebbe falso.

In conclusione vogliamo soltanto sottolineare che, indipendentemente, dall'accettazione o meno dell'assioma della scelta, la natura paradossale del Teorema di Banach–Tarski (e dell'esempio di Vitali e del Teorema di Dehn) è una conseguenza della teoria moderna degli insiemi. Il fatto di supporre che “esistono” numeri irrazionali risulta essere molto utile ma non ha un riscontro concreto ed è una delle fonti di questi “paradossi”.

Ancor più “paradossale” è (nello stesso senso) l'esistenza di insiemi infiniti. Gran parte dei problemi matematici possono essere enunciati usando solo insiemi finiti, così come non è del tutto assurda l'ipotesi che l'universo stesso sia finito (nel senso di “finitamente descrivibile”). L'utilizzo di questi concetti astratti, è giustificato semplicemente dal fatto che risultano, a posteriori, utili nella risoluzione dei problemi.

L'avversione spontanea di ognuno di noi (e di molti matematici professionisti) rispetto al paradosso di Banach–Tarski probabilmente non è molto diversa dall'avversione che doveva provare Pitagora rispetto all'esistenza di $\sqrt{2}$.

Aggiornamenti apportati al documento

15 giugno 2001: prima versione definitiva.

13 novembre 2003: inserimento degli *hyperlinks* e dell'indice.

1 dicembre 2003: ho rifatto tutte le figure.