NOTE PER IL CORSO DI DIDATTICA DELLA MATEMATICA II

ENRICO GREGORIO

Note scritte per la SSIS Veneto, anno accademico 2004-2005.

1. Lezione 1

Si analizzeranno alcuni brani tratti da libri di testo per le scuole superiori. Ciascuno li commenti.

1.1. **Geometria.**

- La retta è il secondo ente fondamentale, possiamo immaginarla come un *insieme consecutivo e infinito di punti aventi sempre la stessa direzione.*
- Disegnate una retta qualsiasi sul piano e dite qual è l'insieme di tutti i punti del piano *equidistanti* da questa retta.
- Consideriamo tre punti qualsiasi A, B e C e disegniamo la retta che passa per questi tre punti. Ci accorgiamo subito che se i tre punti non sono *allineati*, non esiste alcuna retta che passi contemporaneamente per tutti e tre i punti. Diremo quindi che *per tre punti passa una e una sola retta se sono allineati, nessuna in caso contrario.*
- Un poligono equilatero e equiangolo si dice regolare. Un poligono si dice irregolare quando ha *tutti i lati e gli angoli disuguali.*
- Sono figure dotate di centro di simmetria tutte quelle in cui *le diagonali si bisecano* e, ovviamente, il cerchio.
- La retta a è tale che tutti i suoi punti appartengono al piano α; si dice che la retta giace nel piano. *Usando la simbologia insiemistica* scriveremo

$$a \in \{\alpha\}, \quad \{a\} \subset \{\alpha\}, \quad \{a\} \cap \{\alpha\} = \{a\}.$$

1.2. **Numeri.**

- La successione dei numeri naturali è in corrispondenza biunivoca con i *punti equidistanti* di una retta.
- Consideriamo due numeri naturali *divisibili tra loro.*
- Se due frazioni sono equivalenti, sappiamo che rappresentano la stessa grandezza, *quindi sono anche uguali.*
- Nell'uso corrente, comunque, chiameremo numero razionale, o frazione, *una frazione qualsiasi irriducibile.*
- Prima di effettuare qualsiasi operazione con le frazioni, esse *vanno ridotte ai minimi termini.*
- Se vogliamo rappresentare quindi un determinato numero irrazionale *basterà scriverlo come somma di due quadrati perfetti, di cui uno uguale a* 1, e quindi costruire il triangolo rettangolo che ha come cateti la radice quadrata di questi numeri.
- Come potrai leggere nella scheda storica, π è un numero trascendente, *ovvero un numero la cui parte decimale è infinita e non periodica.*
- In seguito è stato dimostrato che π è un numero irrazionale, ma che non può essere rappresentato sulla retta usando riga e compasso, *per cui viene definito numero trascendente.*
- Per riconoscere se un numero è primo lo dividiamo per i successivi numeri primi 2, 3, 5, 7, 11, ... *senza tralasciarne alcuno.*
- Per stabilire se un numero è primo *basta consultare le tavole* che sono in fondo al testo e che riportano i numeri primi minori di 5000.

1.3. **Geometria analitica.**

- Le due rette orientate perpendicolari prendono il nome di assi cartesiani e, precisamente, *quella orizzontale* asse delle ascisse o asse x, *quella verticale* asse delle ordinate o asse y.
- L'equazione della retta passante per i due punti $A(x_1, y_1)$ e $B(x_2, y_2)$ è data da
$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}.$$

1.4. **Ricorsione.** Un calcolatore è in grado di eseguire qualsiasi compito *purché adeguatamente programmato.*

1.5. **Analisi matematica.**

- Si chiama successione *un insieme numerabile di numeri reali.*
- La funzione $y = x - \sqrt{1 - x^2}$ è un ramo della *funzione più generale* $y = x \pm \sqrt{1 - x^2}$.
- La funzione $x^2y^5 - 3x^3y^4 + 5xy^6 - 2y^7$ non è esplicitabile in quanto *nessuno sa risolvere le equazioni dal quinto grado in su.*
- Teorema di Heine-Borel: ogni funzione continua è uniformemente continua.

2. Lezione 2

2.1. **Che cos'è "uguale"?** Uno dei primi simboli matematici che si incontrano è quello di "=". Tutti (o quasi) in prima elementare imparano che $1 + 1 = 2$ e a scrivere operazioni in colonna come

$$\begin{array}{r} 134+ \\ 278= \\ \hline 412 \end{array}$$

Più avanti imparano le cosiddette *equivalenze*:

$$13\,\mathrm{m}^2 = 130\,000\,\mathrm{cm}^2$$

e quindi a usare il simbolo = in modo diverso dal precedente.

Con le frazioni le cose si complicano ancora: $1/2$ e $2/4$ sono uguali o no?

Alla scuola media inferiore, il simbolo = comincia a diventare un *segnale di attivazione*: si scrive

$$\frac{4}{3} \cdot \left\{ 3 + \left[\frac{1}{2} \left(\frac{1}{3} + 6 \right) - 1 \right] \cdot \frac{1}{5} \right\} =$$

e da lì si parte per sviluppare e semplificare l'espressione.

Ancora più avanti, con l'introduzione del calcolo letterale, l'idea che si fanno di questo simbolo diventa probabilmente confusa: che significa infatti

$$3ab^4 + 5a^2b - 4ab^4 = -ab^4 + 5ab^2$$

ai loro occhi?

Ancor più quando si trovano davanti definizioni come:

> *Dicesi equazione un'uguaglianza fra espressioni algebriche soddisfatta per uno o più valori delle lettere.*

- Quella di prima è un'equazione oppure no?
- Quanto fatto fino a quel momento ha a che fare con le equazioni?
- Perché le equazioni si trattano in modo diverso dalle espressioni?

- Quando si trova
$$0 = 0$$
abbiamo un'equazione? Dove sono le lettere cui sostituire i valori?
- Qual è l'incognita in un'equazione letterale?

Si corre il serio rischio di far trattare il tutto a livello *sintattico* e non più *semantico*: il simbolo = non dice più nulla, ma è solo un segnaposto per attivare l'applicazione di certe regole.

Tanto più se ci si trova di fronte a "soluzioni" del seguente tenore:

$$\frac{1}{2}x + \frac{1}{3} = \frac{3}{4}x - \frac{1}{6}$$

$$\frac{6x + 4 = 9x - 2}{12}$$

$$6x + 4 = 9x - 2$$
$$6x - 9x = -2 - 4$$
$$-3x = -6$$

$$x = \frac{-6}{-3}$$

$$x = 2$$

o, peggio,

$$4 = x - 1$$
$$-x = -1 - 4$$
$$-x = -5$$
$$x = 5$$

È chiaro che la seconda equazione andrebbe prima di tutto *osservata*: se da x tolgo 1 trovo 4, quindi $x = 5$. Questo modo di risolvere i problemi è difficile da insegnare, naturalmente. Ma uno studente sveglio viene mortificato se non gli si dà la possibilità di trattare i problemi con il metodo più efficiente. Allo studente meno portato può essere comunque insegnato ad attivare le *procedure* opportune per la soluzione: si arriva allo stesso risultato, forse solo più lentamente.

Si può anche mostrare con esempi che le "soluzioni a occhio" non sempre sono corrette e che, viceversa, a volte occorre stare attenti e applicare le regole:

$$s = 1 + a + \cdots + a^n + \cdots = 1 + a(1 + a + \cdots + a^n + \cdots) = 1 + as$$

da cui $s = 1/(1 - a)$, che invece vale solo per $|a| < 1$ (in questo errore è incorso perfino Eulero).

2.2. **Che cos'è un'equazione?** Premesso che la definizione data poc'anzi è scorretta, quale dovremmo dare?

La mia opinione è che *non* dovremmo darne affatto una definizione.

Le equazioni non esistono!

Si tratta ovviamente di un'affermazione paradossale. Tutti noi consciamente o inconsciamente usiamo equazioni e sappiamo benissimo che cosa sono.

Il problema, come spesso accade in matematica, è che al momento di dare una definizione precisa non sappiamo più che cosa siano.

Per usare una bella citazione latina:

> *Quid ergo est tempus? Si nemo ex me quaerit, scio: si quaerenti explicare velim, nescio.*
>
> (Agostino, Confessioni, Libro XI, Capitolo XIV)

In matematica allora si ricorre spesso a definizioni operative: si pensi alla definizione di "avere lo stesso numero di elementi" per gli insiemi infiniti.

È facile dire quando due insiemi hanno lo stesso numero di elementi: quando esiste una biiezione del primo sul secondo. Ma possiamo definire la "cardinalità"?

Il primo tentativo, quello di Frege, fondato sul ripartire la classe universale in classi di equivalenza rispetto alla relazione di *equipotenza* è fallito a causa del paradosso di Russell.

Una possibile definizione è tramite la teoria degli ordinali, come spiegata nel libro di Halmos; ma è complicata e, pur dando ottimi risultati, non è comprensibile nemmeno a uno studente universitario alle prime armi. La definizione operativa, invece, è intuitiva e utile: le proprietà fondamentali dei numeri cardinali possono essere ottenute più o meno facilmente.

Per tornare all'argomento. Che cos'è un'equazione?

È un modo per risolvere un problema; dobbiamo trovare una certa quantità sotto certe condizioni (per esempio il lato di un triangolo). In termini di questa quantità possiamo calcolare *in due modi diversi* una seconda quantità, che è nota dai dati del problema.

Bene: questi due modi, $f(x)$ e $g(x)$, devono essere uguali:

$$f(x) = g(x).$$

Il buon insegnante sa a questo punto proporre vari esempi di ciò che vuole spiegare.

Ora è facile, usando le proprietà dell'addizione e della moltiplicazione (se è necessario), arrivare a scrivere

$$f(x) - g(x) = 0.$$

Dunque il problema si riduce a calcolare quei valori per i quali l'espressione $f(x) - g(x)$ è zero.

Dobbiamo far capire che tutte le equazioni nascono così. Magari poi si fa esercizio senza scrivere esplicitamente il problema che porta a quella equazione. Ma se si definiscono le equazioni senza dire *prima* a che servono, non si dà modo di comprendere che cosa siano.

Non parlerei mai di "principi di equivalenza" delle equazioni. Piuttosto di operazioni reversibili.

È sensato dire che le equazioni $x^2 = -1$ e $x^6 + 4 = 0$ sono equivalenti perché hanno lo stesso insieme di soluzioni? Direi di no. Due equazioni sono equivalenti se si può passare dall'una all'altra con operazioni di addizione o di moltiplicazione *reversibili*.

Quindi sommare ad ambo i membri la stessa quantità $h(x)$ è ammesso, perché posso tornare indietro sommando $-h(x)$ (esempi, non scritture formali!). È ammesso moltiplicare per 3 perché posso moltiplicare per $1/3$ e tornare alla forma precedente. Non è ammesso moltiplicare per x, a meno che non sappia che $x \neq 0$. E così via.

2.3. **Altri "uguale".** Supponiamo di avere l'equazione

$$\left(x + \frac{1}{x}\right)^2 - 3\left(x + \frac{1}{x}\right) - 4 = 0.$$

L'ovvia tecnica per risolverla è porre

$$t = x + \frac{1}{x}$$

e di calcolare le soluzioni di $t^2 - 3t - 4 = 0$. Qui abbiamo un diverso uso del simbolo, apparentemente. È chiaro che non è così, ma allo studente può apparire sconcertante cambiare l'incognita.

Forse lo sconcerto può essere alleviato se fin da subito si abituano i ragazzi a scomporre le espressioni in parti.

Prendiamo una delle famose "espressioni a quattro piani":

$$\frac{\frac{1}{3} + \frac{2}{5}}{\frac{5}{4} - \frac{2}{3}}.$$

Naturalmente sui testi se ne possono trovare di molto più complesse, ma il discorso è identico. Chiamiamo E il valore dell'espressione, A il numeratore e B il denominatore:

$$A = \frac{1}{3} + \frac{2}{5} = \frac{5+6}{15} = \frac{11}{15},$$

$$B = \frac{5}{4} - \frac{2}{3} = \frac{15-8}{12} = \frac{7}{12}.$$

Dunque

$$E = \frac{A}{B} = \frac{11}{15}\frac{12}{7} = \frac{44}{35}.$$

Otteniamo, a mio parere, almeno tre risultati:

- gli studenti imparano a lavorare su blocchi più piccoli e a scomporre i calcoli complicati in pezzi maneggevoli;
- ogni pezzo è un certo numero e l'espressione stessa *è* un certo numero che possiamo determinare;
- diamo un senso preciso a quel misterioso segno di "=".

Scomporre in parti è utile anche perché minimizza la riscrittura di parti dell'espressione e la possibilità (molto concreta) di errori di copiatura.

Ho sempre considerato un assurdo supplizio quello di dover riscrivere per cinque o magari più righe una parte di espressione che non andava modificata fino al termine. Perché non fare una bella riga sotto un pezzo, dargli un nome e sviluppare quello?

L'algebra simbolica ha avuto inizio quando è diventato necessario considerare più di un'incognita alla volta. Invece di parlare della "prima cosa" e

della "seconda cosa", si è imparato a denotarle ciascuna con un simbolo. Da lì il cammino è stato rapido e ha portato a considerare perfino equazioni in cui gli stessi coefficienti sono simbolici: molti problemi possono essere risolti simultaneamente.

Un altro uso delle sostituzioni è molto interessante: valutare l'espressione

$$\frac{1}{1+10^{200}}+\frac{1}{1+10^{-200}}.$$

Se proviamo a sviluppare l'espressione ci troviamo in imbarazzo per via dei calcoli con numeri poco maneggevoli. Invece, ponendo

$$a=10^{200},$$

l'espressione diventa

$$\frac{1}{1+a}+\frac{1}{1+\dfrac{1}{a}}=1.$$

Un numero troppo complicato (o grande come 10^{200}) viene tolto di mezzo e non ci spaventa più.

2.4. **Soluzione di equazioni.** Tutti conoscono la formula

$$x_{1,2}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}.$$

Ovviamente una tale formula non significa nulla: chi è questo misterioso $x_{1,2}$? È quello con il "più" o quello con il "meno"?

Nessuno dei due, è chiaro. Un numero non può avere due valori, nemmeno la x può avere due valori. Più corretto è dire: le soluzioni dell'equazione $ax^2+bx+c=0$ sono i numeri

$$\frac{-b-\sqrt{b^2-4ac}}{2a} \qquad \text{e} \qquad \frac{-b+\sqrt{b^2-4ac}}{2a}.$$

Quindi, se $ax^2+bx+c=0$, allora x è il primo *oppure* il secondo numero. Quante volte si sente dire: "$x=2$ *e* $x=3$"? Come può un numero essere sia 2 che 3?

No: *le soluzioni sono i numeri* 2 *e* 3.

La formula, imparata acriticamente, spinge lo studente a usarla sempre:

$$x^2-5x+6\leq 0$$

$$x_{1,2}=\frac{5\pm\sqrt{25-24}}{2}$$

$$x_1=2, \qquad x_2=3$$

$$x\leq 2, \qquad x\leq 3$$

oppure, più comune,

$$x^2\leq 4$$

$$x\leq\pm 2.$$

Ma non è solo questo aspetto che voglio sottolineare: più rilevante è quello dell'uso di un simbolo (o di parole a esso equivalenti) in modi diversi. Non è così! Il simbolo di uguale indica identità: gli oggetti da una parte e dall'altra sono *nomi della stessa cosa*.

Questa cosa può non essere nota: un'equazione è essenzialmente una condizione *necessaria* affinché accada una certa situazione. Talvolta è anche una condizione sufficiente, ma non sempre.

> Se indichiamo con x la misura del raggio della circonferenza, allora dobbiamo avere... che si traduce nell'equazione
>
> $$x^2 = 2x + 3$$
>
> e quindi abbiamo $x = 3$ oppure $x = -1$. Siccome il raggio è positivo, l'unica soluzione è che la misura del raggio è 3.

Non è che la seconda soluzione sia "inaccettabile": non è una soluzione del problema. Qui l'equazione porge una condizione necessaria ma non sufficiente.

Ovviamente anche l'altra soluzione dell'equazione va sottoposta all'analisi per verificarne la corrispondenza con i dati e le limitazioni del problema.

2.5. **Un altro "uguale"?** Il principio di identità dei polinomi introduce un nuovo concetto di uguaglianza: abbiamo una nuova struttura e siamo *liberi* di definire che cosa intendiamo per uguaglianza.

2.6. **Congruente o uguale?** L'uguaglianza di figure geometriche è uno scoglio per molti. Infatti, con ottimi motivi, si usa parlare di *congruenza* di figure piuttosto che di uguaglianza.

Su questo la scelta va ponderata: Euclide parlava di uguaglianza e non aveva soverchi problemi.

La dimostrazione del *pons asinorum*, nella quale un triangolo isoscele è considerato uguale a sé stesso da due punti di vista diversi è notevole per illustrare il concetto.

Dobbiamo considerare il triangolo isoscele ABC di vertice A sia come triangolo ABC che come triangolo ACB: per il primo criterio di congruenza, i due triangoli sono "uguali" e perciò hanno gli elementi corrispondenti "uguali". Quindi l'angolo in B è "uguale" all'angolo in C.

Pochi studenti del biennio sono in grado di comprendere questa dimostrazione, probabilmente.

I due triangoli sono solo uno! Ma visto, come abbiamo detto, in due modi distinti. Qui abbiamo due triangoli "uguali" ma *diversi*!

La soluzione naturalmente sta nel considerare una simmetria assiale e infatti, con il metodo delle simmetrie, la dimostrazione è più evidente. Però c'è un inghippo: dobbiamo definire isoscele un triangolo che ha un asse di simmetria.

3. Lezione 3

3.1. **Saper scegliere.** Molti argomenti (quasi tutti) possono essere trattati in più modi. Alcuni sono semplici variazioni, in altri la differenza può essere molto rilevante.

Facciamo qualche esempio.

3.1.1. *Trigonometria.* Cerchio goniometrico oppure triangoli rettangoli?

3.1.2. *Numeri reali.* Sezioni di Dedekind, classi contigue, successioni di Cauchy oppure scatole cinesi? O addirittura impostazione assiomatica?

3.1.3. *Geometria piana.* Assiomi di Euclide, di Hilbert o impostazione kleiniana?

3.1.4. *Calcolo integrale.* Prima gli integrali definiti o quelli indefiniti?

3.2. **Trigonometria.** La scelta dell'impostazione dipende molto anche da chi ascolta e dall'impiego successivo dei concetti.

Per insegnare la trigonometria ai futuri geometri può essere sensato parlare direttamente di triangoli rettangoli, puntando sui problemi pratici di risoluzione dei triangoli.

Viceversa, dove l'uso delle funzioni trigonometriche è più astratto (si pensi a un istituto tecnico a indirizzo elettronico), si può preferire l'approccio con il cerchio goniometrico, che mostra quasi direttamente la dipendenza funzionale.

Occorre poi tenere presente un fattore fondamentale: come si dimostra la "formula di sottrazione del coseno"?

Le dimostrazioni usuali impiegano il cerchio goniometrico, ma sono tutte lacunose e inutilmente complicate. Più semplice, forse, calcolare la misura della corda sottesa da un angolo al centro α; facili considerazioni sul piano cartesiano mostrano che questa corda è $\sqrt{2-2\cos\alpha}$. Basta riferire il piano a un opportuno sistema di coordinate.

A questo punto basta considerare un angolo β e calcolare. Si può allora osservare che la dimostrazione non usa affatto che $\beta \geq \alpha$ né che α o β siano positivi: si tratta semplicemente di costruire l'angolo $\beta - \alpha$ e ruotarlo di α.

Potrebbe essere istruttivo anche dimostrare la formula per angoli acuti usando i triangoli rettangoli, per abituare gli studenti alle risoluzioni di triangoli.

Occorre comunque precisare che l'uso della geometria analitica nello studio della trigonometria va ampliato in modo che non sia un mero strumento occasionale. Sono anche importantissime le considerazioni trigonometriche per ottenere modi diversi di calcolare, per esempio, equazioni di rette o la distanza punto-retta.

Può essere anche molto utile parlare di equazioni parametriche delle rette, che permettono di trattare senza eccezioni problemi sui fasci: l'equazione parametrica della retta generica passante per il punto (x_0, y_0) è

$$\begin{cases} x = x_0 + t\cos\alpha \\ y = y_0 + t\operatorname{sen}\alpha \end{cases}$$

dove α è un angolo qualunque. Molto spesso è più conveniente questa rappresentazione, piuttosto che quella con l'equazione $ax + by + c = 0$. Il significato geometrico di α è evidente.

3.3. **Numeri reali.** Ciascun metodo di costruzione dei numeri reali ha i suoi pregi e i suoi difetti. Quello assiomatico ha il pregio della chiarezza: si mettono in evidenza le operazioni ammesse e le regole di calcolo. Qual è il prezzo che si paga?

Sezioni di Dedekind: sono veramente una risposta al problema?

Classi contigue: dov'è il punto fondamentale che le distingue dalle sezioni di Dedekind?

Le scatole cinesi sono solo un modo meno noioso di parlare di classi contigue.

Successioni di Cauchy: si definiscono bene le operazioni; ma qual è il vero dominio?

A conti fatti sembrerebbe che il metodo assiomatico sia il migliore. Del resto nessuno si fa problemi ad accettare i numeri razionali, sui quali operare in base alle regole di calcolo (che derivano dagli assiomi di campo); perché mai porsi troppi problemi su un insieme numerico del quale abbiamo un'intuizione data dalla retta?

Quello che poi conta è di saper *approssimare* un dato numero reale con la precisione voluta. Qui si apre un campo vastissimo, quello del calcolo numerico: calcolo iterativo, frazioni continue, metodo di Newton, tangenti e secanti. Oppure la classica esaustione di Eudosso e Archimede.

Gli assiomi dei numeri reali garantiscono la possibilità dell'approssimazione, perché i razionali (o le frazioni decimali o quelle binarie) sono densi nei reali e ogni insieme limitato ha estremo superiore.

Insistere sul concetto di estremo superiore può anzi essere più produttivo di lunghi calcoli e spiegazioni sulle classi contigue.

Si può allora anche passare a strutture ordinate più generali, come reticoli e algebre di Boole, utili in molti ambiti.

Ho sempre trovato molto attraente il fatto che la definizione della divisibilità nei naturali sia l'analogo moltiplicativo dell'ordine usuale:

$$a \leq b \iff \text{esiste } c \text{ con } b = a + c,$$
$$a \mid b \iff \text{esiste } c \text{ con } b = ac.$$

Come l'ordine usuale è importante, altrettanto lo è quello per divisibilità, con i concetti collegati di massimo comun divisore e minimo comune multiplo.

La definizione formale è analoga, le proprietà delle due relazioni diversissime. Eppure devono avere qualcosa in comune: proprio l'essere relazioni d'ordine.

3.4. **Geometria piana.** Gli assiomi di Euclide sono lacunosi. Questo dato di fatto è emerso molto tempo fa; non tanto riguardo al problema delle parallele, del quale tanti hanno cercato di trovare soluzione, quanto per via di molti assiomi usati in modo implicito.

A tale riguardo si può menzionare l'assioma di Pasch: supponiamo di avere quattro punti su una retta, A, B, C e D. Supponiamo anche di sapere che B sta fra A e C e che C sta fra B e D. Possiamo concludere qualcosa su A, B e D? L'intuizione spaziale ci dice che B sta fra A e D, ma gli assiomi euclidei non permettono di dimostrarlo!

L'assioma di Pasch, riconosciuto indipendente e consistente rispetto agli altri assiomi di Hilbert, stabilisce: *se una retta incontra un vertice di un triangolo e non contiene altri vertici, allora incontra il lato opposto del triangolo.*

Può essere utile riflettere sulla necessità di questo assioma nella dimostrazione del fatto che B sta fra A e D.

Da un altro lato, infliggere agli studenti l'assiomatizzazione di Hilbert della geometria è forse troppo: gli assiomi sono certamente rigorosi, ma troppo

tecnici in alcuni casi e rischiano di oscurare la semplicità dell'intuizione geometrica.

D'altra parte anche certe dimostrazioni euclidee sono oscure e complicate. È proprio necessario imporle a studenti del primo anno di scuola superiore?

La mia idea è sì. Ma non con lo scopo di imbottirli di dimostrazioni da imparare a memoria, quanto di

- abituarli a ragionare con un certo insieme di dati;
- cercare analogie con casi già noti;
- farsi una chiara idea di ciò che occorre per poter provare un certo fatto;
- saper riconoscere le forme e applicare le giuste tecniche;
- eccetera.

Problemi ipercomplicati non servono a nulla se non a spaventare lo studente che non sa da dove cominciare. Viceversa, problemi semplici permettono poi di saperne scomporre uno più complesso in sottoproblemi. Non vogliamo fare dei nostri studenti i geni matematici che non sono: saper risolvere un problema complicato non è nemmeno indice assoluto di abilità matematica.

Sono i problemi *difficili* a dire se uno è bravo oppure no; ma difficile può essere un problema che si enuncia in due righe.

> Sia ABC un triangolo nel quale le bisettrici relative agli angoli B e C sono congruenti. È vero che il triangolo ABC è isoscele sulla base BC?

Un problema del genere, sul quale invito a riflettere, non è per niente un problema facile nonostante la formulazione sia semplicissima. Il risultato si chiama tradizionalmente "Teorema di Steiner-Lehmus"; la prima dimostrazione risale al 1844, dovuta a Steiner, proprio su sollecitazione di Lehmus che ne trovò un'altra nel 1850. La prima dimostrazione diretta è del 1970.

3.5. **Calcolo integrale.** Non dovrebbe stupire la mia personale preferenza per la trattazione dell'integrale definito *prima* di quella del cosiddetto integrale indefinito (che chiamerei più semplicemente *una primitiva*).

Parlare di primitive ancor prima di trattare l'integrale sarebbe come insegnare la somma in colonna senza sapere che cosa sia l'addizione.

Un metodo di calcolo è ben diverso dal concetto! L'integrale, introdotto dapprima come area, può benissimo essere applicato a molte situazioni diverse, come per esempio lunghezza di una curva o lavoro di una forza.

Fra l'altro la ricerca di una primitiva non è l'unico metodo per calcolare un integrale: si pensi a una funzione importante come $x \mapsto e^{-x^2}$ che non ammette primitive esprimibili tramite "funzioni elementari".

La funzione $x \mapsto e^{-x^2}$ non ammette primitive: questa è una delle cose che talvolta si sentono dire. Ovviamente, trattandosi di una funzione continua su tutto l'asse reale ammette primitiva eccome:

$$x \mapsto \int_0^x e^{-t^2} dt$$

è una primitiva, per il teorema fondamentale del calcolo. Funzione che possiamo studiare nel modo consueto, calcolandone dominio, derivata, eccetera.

Non sappiamo niente di più sulla funzione

$$x \mapsto \int_1^x 4t \log t \, dt$$

quando la scriviamo come

$$x \mapsto 2x^2 \log x - x^2 + 1.$$

Ne abbiamo solo scritto un'altra forma. Se volessimo, per esempio, studiare gli zeri di questa funzione, non avremmo altro modo che calcolarne la derivata per stabilirne crescenza e decrescenza: cioè proprio la funzione che abbiamo integrato.

Ovvio che, in molti casi, la conoscenza esplicita della primitiva può dare informazioni essenziali: il valore esatto dell'integrale è una.

Un altro esempio. Supponiamo di voler calcolare la lunghezza di un tratto della curva $f(x) = \cosh x = (e^x + e^{-x})/2$. La risposta è facile conoscendo

$$\int_0^x \sqrt{1 + (f'(t))^2} \, dt = \int_0^x \sqrt{1 + \operatorname{senh}^2 t} \, dt = \int_0^x \cosh t \, dt = \operatorname{senh} x.$$

Il problema è rilevante, sapendo che la curva data descrive la forma di una fune sospesa alle estremità e soggetta alla sola forza di gravità (catenaria).

Lo stesso problema per la curva $g(x) = \operatorname{senh} x$ si traduce in un integrale ellittico non esprimibile tramite funzioni elementari che si può ridurre al calcolo della lunghezza di un'opportuna ellisse. Un modo di procedere come questo dà spunti di riflessione: a volte occorre, come nel caso del logaritmo o delle funzioni trigonometriche, definire nuove funzioni che risolvano certi problemi.

Li risolvono davvero? Sì e no. Possiamo studiare le proprietà qualitative di queste funzioni e da queste trarre deduzioni sui nostri problemi. E possiamo, avendo una rappresentazione di queste funzioni come integrali, calcolare i valori di queste funzioni con la precisione desiderata. Che si vuole di più?

Certo, non conosciamo *veramente* queste funzioni. Ma conosciamo il logaritmo? O il seno?

Un altro dettaglio sull'integrazione; ho già detto che abolirei il termine "integrale indefinito", parlando invece di primitiva. La solita definizione:

$$\int f(x) \, dx$$

è l'insieme di tutte le primitive di f non ha alcun senso né alcuna utilità. Dove mai si usa questo insieme di funzioni? Che proprietà ha? *Serve davvero, dunque?*

No, naturalmente: meglio tagliar corto e dire che $\int f(x) \, dx$ denota *una* primitiva di f. Tanto, per calcolare l'integrale esteso a un intervallo, una vale l'altra: su quell'intervallo due primitive differiscono per una costante.

Non vorrei tacere il fatto che spesso, dopo la definizione dell'integrale indefinito come insieme delle primitive si trovano le seguenti amenità:

$$\int f(x)g(x) \, dx = F(x)g(x) - \int F(x)g'(x) \, dx$$

dove F è una primitiva di f. Possiamo dunque operare algebricamente sull'insieme delle primitive di una funzione? Parrebbe di sì. E allora come la mettiamo con il seguente ragionamento?

Sappiamo che

$$\int kf(x)\,dx = k\int f(x)\,dx,$$

dove k è un numero reale. Ma per $k = 0$ abbiamo allora dimostrato che la funzione nulla ha una sola primitiva.

Facciamo un esempio: le funzioni $f(x) = \log x$ e $g(x) = \log(2x)$ sono entrambe primitive di $x \mapsto 1/x$ (definita per $x > 0$). Che cosa ci fa preferire l'una all'altra? Nulla: entrambe danno lo stesso risultato quando si applichi il teorema fondamentale per il calcolo dell'integrale definito di $1/x$.

Essere pedanti e dire che *la più generale primitiva di* $x \mapsto 1/x$ (*definita su* $\mathbf{R} \setminus \{0\}$) *è*

$$F(x) = \begin{cases} \log x + c_1 & \text{per } x > 0 \\ \log(-x) + c_2 & \text{per } x < 0 \end{cases}$$

non aggiunge *nulla* alla conoscenza degli allievi. Insisto: il calcolo della primitiva serve solo a trovare il valore di un integrale; una vale l'altra.

4. Lezione 4: il concetto di funzione

Pochi concetti matematici sono così fondamentali come quello di funzione e, più in generale, di relazione. Non a caso si è tentato di porli come concetti indefiniti sui quali costruire la fondazione della matematica.

Purtroppo una formulazione di questo tipo è poco adatta allo studio elementare, anche se ha un grande fascino.

Il tentativo però mette in luce un aspetto molto importante: ancor più del concetto di funzione è rilevante quello di *composizione di funzioni*.

Esistono molti approcci alle funzioni, cui darò nomi forse non usuali.

Concreto: che parte dalla "dipendenza di una variabile da un'altra";
Analitico: dove le funzioni sono espresse tramite formule;
Astratto: una funzione è un insieme di coppie ordinate.

Nessuno di questi da solo riesce a cogliere tutti gli aspetti.

Il primo traduce l'idea *dinamica* di funzione: facciamo percorrere alla variabile indipendente tutto il dominio e otteniamo i corrispondenti valori della variabile dipendente.

Il secondo andrebbe bandito da qualunque trattazione seria.

Il terzo approccio considera una funzione come un'ente matematico *statico* e tutti sappiamo che è così.

Il problema è di riuscire a comunicare quest'idea agli studenti.

È nota la cosiddetta "definizione di Dirichlet":

> Si chiama funzione qualunque legge che a ogni valore della variabile indipendente associa uno e un solo valore della variabile dipendente.

Ci accorgiamo subito che questo non definisce nulla: che cos'è una "legge"? Stiamo definendo un concetto con uno che è del tutto equivalente. E chi sono queste fantomatiche variabili? Perché una è indipendente? Da che cosa è indipendente?

Cito un vecchio slogan dei sindacati negli anni '70-'80: "Il salario è una variabile indipendente".

Troppo spesso concetti matematici vengono abusati nel linguaggio comune; questo è un caso: l'equivoco nasce ovviamente dal non avere la minima idea di che cosa significhi "variabile indipendente".

Potrebbe essere un'idea per un lavoro interdisciplinare quella di analizzare dal punto di vista economico quello slogan.

Ma non è questo che mi interessa, quanto il capire da dove proviene il fraintendimento.

L'idea di funzione nasce piuttosto tardi nello sviluppo della matematica. Basti pensare che la notazione $f(x)$ è stata proposta da Lagrange.

Esisteva già ovviamente qualche prodromo, soprattutto in Newton che parlava di "flussioni" e "fluenti": in sostanza altri nomi per variabile dipendente e indipendente. Da Newton provengono le notazioni $\dot{x}$, $\ddot{x}$ per indicare le derivate prime e seconde, ancor oggi usate in fisica (con un preciso significato tecnico).

Le notazioni di Leibniz erano le classiche

$$\frac{dy}{dx}$$

dove y denota la quantità dipendente dalla grandezza x.

La notazione funzionale di Lagrange ha aperto la strada alla definizione di Dirichlet, il quale intendeva ampliare la classe delle funzioni ammissibili fino a comprendere quella che porta il suo nome:

$$f(x) = \begin{cases} 0 & \text{se } x \text{ è razionale;} \\ 1 & \text{se } x \text{ è irrazionale.} \end{cases}$$

Autorevoli matematici la rifiutavano come funzione.

Il lavoro di Cantor sugli insiemi infiniti portò a un altro ampliamento del concetto e alla definizione come insieme di coppie ordinate.

Definire una funzione come "legge" porta a un circolo vizioso. Come in geometria non si definisce il "punto" o la "retta", anche nel caso delle funzioni si preferisce evitare di addentrarsi in problemi più filosofici che matematici e ci si limita a considerare una funzione come un insieme f dotato di certe proprietà:

- gli elementi di f sono coppie ordinate;
- due coppie ordinate distinte in f hanno primi elementi distinti.

La condizione di univocità (cioè la seconda proprietà), è spesso scritta

$$\text{se } (a,b),(a,b') \in f, \text{ allora } b = b'.$$

Non è facile, per un principiante, capire il senso di una proprietà così formulata.

Si può tentare di chiarirla in questo modo: supponiamo di avere una funzione f (data quindi come insieme di coppie ordinate) e supponiamo di conoscere un elemento a del dominio.

Possiamo allora cercare fra gli elementi di f quella coppia che ha a come primo elemento (sappiamo che ce n'è una) e troviamo (a,b).

Supponiamo ora che con un procedimento diverso (magari perché possediamo un algoritmo capace di descrivere tutti gli elementi di f) scopriamo che la coppia $(a,b') \in f$.

Siccome sappiamo che f è una funzione, possiamo affermare che b e b' sono nomi dello stesso oggetto. È esperienza nota che uno stesso oggetto matematico può avere "nomi" diversi: $\log 4$ e $2\log 2$, per esempio.

Tuttavia ritengo che, almeno al principio, la definizione più comprensibile di univocità sia quella data sopra.

Analogamente per l'iniettività: la funzione f è iniettiva se elementi distinti di f hanno anche secondi elementi distinti.

Due coppie ordinate (a, b) e (c, d) sono distinte quando

$$a \neq b \text{ oppure } c \neq d.$$

L'univocità afferma che due coppie in f sono distinte quando hanno *primi elementi distinti*.

L'iniettività afferma che due coppie in f sono distinte quando hanno *primi elementi distinti e secondi elementi distinti*.

Quando vogliamo verificare se un insieme di coppie ordinate f è una funzione, dobbiamo proprio andare a vedere se esistono in f coppie distinte con primi elementi uguali: in tal caso f non è una funzione.

Diverso è il caso se f è dato mediante una procedura "algoritmica":

$$f = \{\, (n+1, n) : n \in \mathbf{N} \,\}$$

è una funzione? In questo caso, proprio perché abbiamo un algoritmo, il secondo metodo è preferibile: supponiamo che $(a, b), (a, b') \in f$.

Allora $(a, b) = (x + 1, x)$ e $(a, b') = (y + 1, y)$ per opportuni $x, y \in \mathbf{N}$. Dunque

$$a = x + 1 = y + 1$$

da cui segue che $x = y$ e quindi $b = b'$.

Non sappiamo che cosa sia un algoritmo (qui la citazione da S. Agostino va ancora bene), ma quando ne abbiamo uno davanti lo sappiamo riconoscere. È una procedura che ci permette di eseguire "calcoli", non solo con numeri ovviamente.

Vale la stessa cosa per la verifica dell'iniettività. Se f è data come insieme di coppie ordinate, si cerca se ci siano in f coppie con uguale secondo elemento. Se invece f è quella di prima, si può provare così: supponiamo che $(a, b), (a', b) \in f$. Questo non è vietato dalla condizione di univocità.

Avremo $(a, b) = (x + 1, x)$ e $(a', b) = (y + 1, y)$. Ma allora $b = x = y$ e quindi $a = x + 1 = b + 1 = y + 1 = a'$ e quindi la funzione è iniettiva.

Più complessa è la questione su dominio e codominio, per la quale suggerisco di usare una terminologia precisa.

Un'*applicazione* $f\colon A \to B$ è data dai seguenti ingredienti:

- un insieme A detto *dominio*;
- un insieme B detto *codominio*;
- una funzione f.

Si richiede anche che:

- per ogni $a \in A$ esiste $b \in B$ tale che $(a, b) \in f$.

Sappiamo già dalla definizione di funzione che, dato $a \in A$, l'elemento $b \in B$ con la proprietà che $(a, b) \in f$ è unico.

Distinguere fra funzione e applicazione può, a mio parere, evitare fraintendimenti.

Ogni funzione può essere vista come applicazione in molti modi; ma, in osservanza della definizione, il dominio è univocamente determinato. Non lo è il codominio e questo è, in un certo senso, nella natura delle cose.

Il moto di un corpo su una retta in funzione del tempo può benissimo essere considerato come applicazione da un certo intervallo dei numeri reali (il dominio) verso l'insieme dei punti della retta o anche di un qualunque piano che la contenga; dopo tutto ci immaginiamo che questo moto avvenga nello spazio.

Trovo quindi scorretto parlare della "funzione $f(x)$". Quello che conta è f e non il nome che diamo all'inesistente variabile.

Spesso si trova la scrittura $y = f(x)$; questa ha senso solo se la intendiamo come abbreviazione di

$$f = \{\, (x, f(x)) : x \in A \,\},$$

a sua volta abbreviazione di

$$f = \{\, (x, y) : \text{esiste } x \in A \text{ tale che } y = f(x) \,\},$$

dove A è un opportuno insieme di numeri reali *preassegnato* o calcolabile secondo precise definizioni. Infatti abbiamo identificato ogni funzione con il suo grafico; più precisamente questo è vero per le applicazioni, per le quali possiamo dire che la funzione (il terzo ingrediente) è un sottoinsieme di $A \times B$.

Dare un nome alla "variabile indipendente" porta a problemi nella costruzione delle composizioni di funzioni. Che, in realtà, sono facili.

Possiamo certamente dare un senso a scritture del tipo $x \mapsto f(x)$. Spesso $f(x)$ è rappresentato con una formula del tipo $x^3 - 2x$, ma non necessariamente. Se ora abbiamo $f\colon A \to B$ e $g\colon B \to C$, chi è la loro composizione?

$$g \circ f = \{\, (a, c) : \text{esiste } b \in B \text{ tale che } (a, b) \in f \text{ e } (b, c) \in g \,\}.$$

Abbiamo esplicitamente dato $g \circ f$ come insieme di coppie ordinate e possiamo completare assegnando dominio A e codominio C.

La verifica che A può essere preso come dominio dice anche come si calcola $g \circ f$: sia $a \in A$. Per ipotesi esiste $b \in B$ tale che $(a, b) \in f$ e quindi anche $c \in C$ tale che $(b, c) \in g$.

In tal caso abbiamo $b = f(a)$ e quindi $c = g(b) = g(f(a))$ e la dimostrazione è completa. Di più: sappiamo che, per ogni $a \in A$,

$$g \circ f(a) = g(f(a)).$$

Perciò, se $f\colon \mathbf{R} \to \mathbf{R}$ è l'applicazione $x \mapsto x^2$ e $g\colon \mathbf{R} \to \mathbf{R}$ è l'applicazione $x \mapsto x + 1$, avremo, per ogni $a \in \mathbf{R}$:

$$g \circ f(a) = g(f(a)) = g(a^2) = a^2 + 1$$

e quindi $g \circ f\colon \mathbf{R} \to \mathbf{R}$ è l'applicazione $x \mapsto x^2 + 1$.

È utile imparare a leggere le formule indipendentemente dal nome delle lettere che compaiono. L'applicazione g di prima è "sommare 1", la f è "elevare al quadrato".

Una notazione (non molto popolare) è addirittura quella di non assegnare alcuna lettera:

$$\begin{aligned} f\colon \mathbf{R} &\to \mathbf{R} \\ * &\mapsto (*)^2 \end{aligned}$$

dove l'asterisco svolge lo stesso ruolo di x, ma è meno carico di significati esterni. Al posto di $*$ possiamo mettere qualunque numero reale, come indicato dal fatto che la funzione ha dominio $\mathbf{R}$. Così

$$\begin{aligned} g\colon \mathbf{R} &\to \mathbf{R} \\ * &\mapsto (*) + 1 \end{aligned}$$

e quindi $g(f(a))$ si calcola proprio "sommando 1 a $f(a) = a^2$".

Parlare subito di composizione permette di dimostrare fatti che sono importanti riguardo all'iniettività e alla suriettività: un'applicazione è iniettiva se e solo se è "cancellabile a sinistra" ed è suriettiva se e solo se è "cancellabile a destra".

La composizione di due applicazioni iniettive (suriettive) è iniettiva (suriettiva); se la composizione $g \circ f$ è iniettiva (suriettiva), allora f è iniettiva (g è suriettiva).

Questi fatti possono rendere più comprensibile l'importanza dei concetti di iniettività e suriettività: sapere di un'applicazione che è iniettiva o suriettiva ci può semplificare alcuni calcoli.

5. Lezione 5: la tesi di Church

Che cos'è un algoritmo?

Abbiamo già avuto occasione di porre domande del genere e la risposta è tutt'altro che buona: *non lo sappiamo.*

È una limitazione? Forse sì o forse no.

Come in tanti campi della matematica, non è tanto importante saper precisare nei dettagli un concetto, quanto saperlo usare.

Gli algoritmi hanno origine nella più remota antichità: ogni metodo di calcolo è un algoritmo e i noti papiri egizi o le tavolette sumere e babilonesi ne contengono moltissimi. Del resto anche noi cominciamo a usare algoritmi fin da bambini: l'addizione e la moltiplicazione in colonna lo sono.

La parola *algoritmo* deriva probabilmente dal nome di Al Khuwarizmi, l'arabo ritenuto ideatore del metodo delle equazioni. Può sembrare curioso, ma *algorista* nella Spagna uscita dalla dominazione araba indicava ciò che chiameremmo ora *giustaossi* (in veneto): forse *colui che rimette le ossa a posto.*

Sappiamo riconoscere un algoritmo quando ce lo troviamo davanti; più difficile è dare una definizione che possa comprendere tutte le possibilità che sono state aperte dall'avvento del calcolo automatico.

Caratteristica comune sembrerebbe essere quella di "saper operare in modo non ambiguo, prendendo sempre un unico percorso". Ciò non è vero: si pensi agli innumerevoli casi in cui un programma, di fronte a due possibilità, ne sceglie una "a caso".

Non ci deve sorprendere: ci sono molte costruzioni geometriche con riga e compasso che si eseguono scegliendo un punto e tracciando una circonferenza con quel centro e raggio arbitrario.

Alcune delle costruzioni note dai corsi di disegno tecnico non rispettano rigorosamente le prescrizioni della geometria greca classica, nella quale il compasso non poteva essere usato per riportare la lunghezza di un segmento: si pensava al compasso come uno strumento che si chiude non appena una punta viene sollevata dall'appoggio. Si può però dimostrare che l'uso moderno del compasso non cambia l'insieme delle costruzioni possibili: ogni costruzione geometrica con riga e compasso (usato in modo classico o no) si può eseguire con la riga e *un unico cerchio dato*.

Abbiamo qui un aspetto importante, sul quale torneremo: possiamo confrontare fra loro metodi diversi, giungendo alla conclusione che danno gli stessi risultati. A volte, invece, ciò non accade: con la sola riga non è possibile eseguire tutte le costruzioni possibili con riga e compasso. Con riga, compasso e squadra si possono eseguire costruzioni che risolvono problemi di terzo grado (di ciò erano consapevoli anche i Greci).

Si possono caratterizzare mediante l'algebra, tutte le costruzioni possibili con riga e compasso. Limitandoci al piano, un punto di coordinate (a, b) è costruibile se e solo se a e b appartengono a un sottocampo F dei reali per il quale esista una successione

$$Q = F_0 \subset F_1 \subset F_2 \subset \cdots \subset F_{n-1} \subset F_n = F$$

dove gli F_i sono sottocampi dei reali e $F_i = F_{i-1}[c_i]$ con $c_i^2 \in F_{i-1}$, per $i = 1, 2, \ldots, n$; Q denota i numeri razionali.

Un risultato come questo prova che certe costruzioni sono impossibili: un numero costruibile è algebrico e ha polinomio minimo sui razionali di grado potenza di due. Tutti i numeri trascendenti non sono costruibili né lo sono quelli di grado non potenza di due. Ovviamente esistono anche numeri di grado potenza di due che non sono costruibili; infatti esistono equazioni di grado 4 le cui radici dipendono essenzialmente da radici cubiche e quindi non sono costruibili.

Ma questa caratterizzazione apre nuove prospettive. Abbiamo un procedimento effettivo per sapere se un numero algebrico è costruibile (purché ne conosciamo il polinomio minimo). Nello sviluppo della logica a partire da Boole si pone in modo naturale il problema: possiamo trovare un procedimento effettivo per sapere se un certo enunciato ammette una dimostrazione?

La conoscenza di questo procedimento permetterebbe di eliminare molti problemi della matematica: basterebbe applicare il procedimento, attenderne l'esito, e sapere se possediamo una dimostrazione o no.

Un primo successo fu la dimostrazione della decidibilità (cioè l'esistenza di una procedura effettiva) del calcolo proposizionale. Il metodo delle tavole di verità permette di sapere se una certa formula è una tautologia e anzi di conoscerne il valore di verità in funzione dei valori di verità delle formule elementari che la compongono.

Più precisamente, si dimostra che, da un certo insieme di formule prese come assiomi, si possono dedurre con le regole di deduzione ammesse tutte e sole le tautologie. Inoltre esiste una procedura effettiva per:

(1) decidere se una formula è un assioma;
(2) decidere se una formula è una tautologia;
(3) trovare una deduzione di una tautologia a partire dagli assiomi.

Il problema è quindi completamente risolto. Che dire del calcolo dei predicati? Il calcolo proposizionale è tutto sommato molto limitato: non si va molto oltre i classici sillogismi. Il calcolo dei predicati, invece, comprende tutte le teorie matematiche note: ogni teoria può essere formalizzata in un linguaggio che comprenda predicati (cioè simboli di relazione) e nel quale si possa

decidere se una formula è un assioma.

È ovvio che, se prendiamo come assioma ogni enunciato vero, ogni enunciato vero è dimostrabile. Ma avremmo solo spostato il problema: come facciamo a sapere se un enunciato è vero?

Più ancora: che cosa è "vero"? L'analisi di questo ci porterebbe troppo in là. Limitiamoci a problemi più maneggevoli.

La scuola di Hilbert riconobbe la necessità di provare la coerenza dell'aritmetica; una volta dimostrata questa, saremmo stati in possesso di dimostrazioni di coerenza di ogni teoria nota. Questo almeno era il programma.

Scosso alle fondamenta dal teorema di incompletezza dimostrato da Gödel nel 1939.

5.1. **Formalizzazione della consistenza dell'aritmetica.** Gödel si chiese se fosse possibile dimostrare la coerenza dell'aritmetica rimanendo al suo interno. Una teoria è coerente se e solo se esistono in essa enunciati indimostrabili. L'idea fu allora di trasformare l'asserzione metamatematica: "non esiste una dimostrazione dell'enunciato φ" in un enunciato dell'aritmetica stessa del quale poter dare una dimostrazione. Basta questo per dire che l'aritmetica è decidibile?

No, naturalmente: dato un enunciato di cui si possa dimostrare che è non dimostrabile, occorre anche dare una dimostrazione della sua negazione.

Ed ecco che arriva la sorpresa: esiste almeno un enunciato φ dell'aritmetica tale che né φ né la sua negazione siano dimostrabili.

Usando questo fatto Gödel riuscì a compiere il salto decisivo: *non è possibile dimostrare la coerenza dell'aritmetica all'interno dell'aritmetica stessa.* In effetti l'attribuzione precisa di questo risultato deve comprendere anche Rosser; Gödel riuscì solo a dimostrare una forma debole del teorema e qualche anno dopo Rosser ne diede la dimostrazione completa.

Non ci interessa analizzare questo tipo di risultati, vogliamo piuttosto considerare alcuni dei metodi usati da Gödel.

Il primo passo fu quello di assegnare a ogni formula e a ogni dimostrazione formale un numero che le caratterizzassero. Il metodo è ingegnoso.

Un linguaggio formale consiste di

(1) connettivi e quantificatori: "$\vee$", "$\neg$", "$\forall$";
(2) predicati "P_n", per ogni numero naturale n;
(3) variabili "v_n, per ogni numero naturale n.

Per semplicità supporremo che i predicati siano binari. In molte trattazioni si introducono anche simboli funzionali e simboli di costante, che però possono essere eliminati a prezzo di aumentare la complessità delle formule.

Assegniamo numeri ai simboli in questo modo:

(1) 7 a "$\vee$", 13 a "$\neg$", 19 a "$\forall$";
(2) $3 + 6n$ a "P_n";
(3) $5 + 6n$ a "v_n".

Dato un numero naturale possiamo decidere se è il numero di un simbolo e di quale.

Una formula è una successione finita di simboli del linguaggio; se indichiamo con g_0 la funzione che a ogni simbolo associa il suo numero, possiamo, data la formula $\varphi = \alpha_1\alpha_2 \dots \alpha_n$ considerare il numero

$$g_1(\varphi) = p_1^{g_0(\alpha_1)} p_2^{g_0(\alpha_2)} \dots p_n^{g_0(\alpha_n)}$$

dove $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, ... è la successione dei numeri primi. Per via dell'unicità della decomposizione in fattori primi, dato un numero possiamo stabilire se è il numero di una certa successione di simboli e quale. Certe successioni non sono formule, altre sì, ma sappiamo quali in base alle regole.

Una dimostrazione Δ è una successione finita di formule:

$$\varphi_1 \ \varphi_2 \ \dots \ \varphi_m.$$

A ogni successione finita Δ di successioni finite di simboli possiamo associare il numero

$$g_2(\Delta) = p_1^{g_1(\varphi_1)} p_2^{g_1(\varphi_2)} \dots p_n^{g_1(\varphi_n)}$$

che è un numero certamente diverso dal numero di una successione finita di simboli, perché l'esponente del 2 è dispari in $g_1(\varphi)$, mentre è pari in $g_2(\Delta)$.

Esempio. Il numero di Gödel della formula

$$\forall v_0 \forall v_1 \vee \neg P_0 v_0 v_1 P_0 v_1 v_0$$

è

$$2^{19}3^5 5^{19} 7^{11} 13^7 17^{13} 19^3 23^5 29^{11} 31^5 37^{11} 41^5.$$

Ammettendo che il predicato "P_0" stia per l'uguaglianza, una formula di questo tipo si troverebbe più spesso scritta come

$$\forall v_0 \forall v_1 (v_0 = v_1 \to v_1 = v_0)$$

ma non è questo il punto.

Il fatto è che esiste una procedura effettiva per decidere se una successione finita di simboli è una formula secondo le regole. Che cosa diventa guardando i numeri di Gödel? Per essere più precisi, cerchiamo una funzione $\mathbf{f}_1 : \mathbf{N} \to \mathbf{N}$ che valga 0 sui numeri che sono numeri di Gödel di una formula e 1 sui numeri che non sono numeri di Gödel di una formula.

5.2. **Funzioni ricorsive.** Ci rendiamo facilmente conto che la funzione $\mathbf{f}_1$ si può calcolare con una procedura effettiva:

(1) si scompone un numero n in fattori primi;
(2) si listano i fattori primi così ottenuti, q_1, q_2, ..., q_k;
(3) se $q_k > p_k$, poniamo $\mathbf{f}_1 = 1$; altrimenti si va al passo successivo;
(4) si listano gli esponenti dei numeri primi, e_1, e_2, ..., e_k;
(5) si pone $t = 0$;
(6) si pone $t := t + 1$; se e_t mod 6 è pari, si pone $\mathbf{f}_1 = 1$;
(7) se e_t mod $6 = 1$ e $e_t = 1$ oppure $e_t > 19$, si pone $\mathbf{f}_1 = 1$;

(8) si pone a_t uguale al simbolo che corrisponde a e_t e si torna al passo 6.

Non insistiamo troppo su questa enumerazione di regole, ma alla fine troviamo una lista a_1, a_2, ..., a_k di simboli oppure abbiamo già calcolato $\mathbf{f}_1 = 1$. Nel primo caso possiamo usare le regole per stabilire se la successione di simboli è una formula ed è chiaro che queste regole possono essere espresse in termini puramente numerici. Per esempio, il simbolo "$\forall$" deve essere seguito da un simbolo di variabile, mentre il connettivo "$\neg$" da una formula, e così via.

Possiamo realizzare in modo analogo una funzione $\mathbf{f}_2$ che restituisca il valore 0 se e solo se l'argomento è il numero di Gödel $g_2(\Delta)$ di una deduzione corretta nella teoria?

La risposta è sì e si basa sul fatto che le regole di deduzione sono esprimibili algoritmicamente.

La dimostrazione di Gödel, con le integrazioni di Rosser, mostra, con un procedimento che ricorda quello diagonale di Cantor, che esistono funzioni che non rientrano nella classe delle funzioni esprimibili con i metodi algoritmici derivanti da queste regole e un enunciato tale che la dimostrazione sua o della sua negazione, produrrebbe una funzione del tipo non esprimibile che sarebbe esprimibile: contraddizione.

Quali sono le funzioni ottenibili con i metodi suddetti? La risposta di Gödel è: le funzioni ricorsive.

Considereremo solo funzioni definite su $\mathbf{N}^n$, dove n è un numero naturale: $\mathbf{N}^n$ è l'insieme delle n-uple di numeri naturali. Con $\mathbf{N}^0$ intendiamo un insieme formato da un solo elemento, che è 0.

I valori delle nostre funzioni saranno sempre in $\mathbf{N} \cup \{\infty\} = \mathbf{N} \cup \{\infty\}$, dove ∞ è un elemento che non è una m-upla di numeri naturali, per alcun m.

Prima di procedere all'introduzione delle funzioni ricorsive, esaminiamo una convenzione: considereremo in realtà funzioni parziali, cioè non definite su tutto $\mathbf{N}^n$; dire che $f(\mathbf{x}) = \infty$ o $f(\mathbf{x}) \neq \infty$ è un modo comodo di dire che f non è definita o è definita su $\mathbf{x}$.

Se $g_1, g_2, \ldots, g_k$ sono funzioni $\mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$ e f è una funzione $\mathbf{N}^k \to \mathbf{N} \cup \{\infty\}$, definiamo

$$f(g_1, g_2, \ldots, g_k)(\mathbf{x}) = f(g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_k(\mathbf{x}))$$

purché $g_i(\mathbf{x}) \neq \infty$ $(i = 1, 2, \ldots, k)$; altrimenti poniamo $f(g_1, g_2, \ldots, g_k)(\mathbf{x}) = \infty$.

5.3. **Funzioni ricorsive iniziali.** Fissato i con $1 \leq i \leq n$ consideriamo

$$I_{n,i} \colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$$

definita, per $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{N}^n$, da $I_{n,i}(\mathbf{x}) = x_i$.

Consideriamo anche la funzione Z di "zero variabili", cioè definita su $\mathbf{N}^0$, $Z() = 0$.

Consideriamo per finire la funzione $S \colon \mathbf{N} \to \mathbf{N} \cup \{\infty\}$: $S(n) = n + 1$.

5.4. **Funzioni ricorsive.** Ogni funzione ricorsiva iniziale è una funzione ricorsiva.

Se $f \colon \mathbf{N}^k \to \mathbf{N} \cup \{\infty\}$ e $g_1, g_2, \ldots, g_k \colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$ sono funzioni ricorsive, allora $f(g_1, g_2, \ldots, g_k) \colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$ è una funzione ricorsiva.

Supponiamo di avere una funzione $g\colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$ e una funzione $h\colon \mathbf{N}^{n+2} \to \mathbf{N} \cup \{\infty\}$. Allora esiste un'unica funzione $f\colon \mathbf{N}^{n+1} \to \mathbf{N} \cup \{\infty\}$ caratterizzata dalle identità:

$$f(\mathbf{x}, 0) = g(\mathbf{x})$$
$$f(\mathbf{x}, S(n)) = h(\mathbf{x}, n, f(\mathbf{x}, n))$$

Notiamo che, se $f(\mathbf{a}, b) = \infty$, allora anche $f(\mathbf{a}, c) = \infty$, per ogni $c > b$. In effetti stiamo usando la convenzione della composizione di funzioni vista sopra: quando uno degli argomenti è ∞, la funzione vale ∞ per convenzione.

Se g e h sono ricorsive, allora f così ottenuta è ricorsiva.

5.5. **Minimizzazione.** Sia data una funzione $g\colon \mathbf{N}^{n+1} \to \mathbf{N} \cup \{\infty\}$. Definiamo allora una funzione $f\colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$ con la seguente procedura:

- fissiamo $\mathbf{a} \in \mathbf{N}^n$: esiste al più un $b \in \mathbf{N}$ tale che, per ogni $y < b$, $g(\mathbf{a}, y) > 0$ e $g(\mathbf{a}, y) \neq \infty$;
- se tale b esiste, poniamo $f(\mathbf{a}) = b$;
- se tale b non esiste, poniamo $f(\mathbf{a}) = \infty$.

La funzione $f\colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$ così ottenuta si denota con μg.

Ogni funzione che si ottenga da una funzione ricorsiva per *minimizzazione* è ricorsiva. In altre parole, se g è ricorsiva, allora μg è ricorsiva.

Per esempio, se

$$g(x_1, x_2, x_3) = \begin{cases} x_2 - x_1 x_3 & \text{se } x_2 - x_1 x_3 \geq 0 \\ 0 & \text{se } x_2 - x_1 x_3 < 0 \end{cases}$$

allora $\mu g(a_1, a_2)$ è il minimo numero b tale che, per ogni $y < b$, $g(a_1, a_2, y) = a_2 - a_1 y > 0$, se tale b esiste. Esaminiamo i vari casi.

Se $a_1 = a_2 = 0$, abbiamo $b = 0$; infatti l'asserzione

$$\text{“per ogni } y, \text{ se } y < 0 \text{ allora } g(0, 0, y) > 0 \text{ e } g(0, 0, y \neq \infty\text{”}$$

è vera.

Se invece $a_2 > 0$ e $a_1 = 0$, dal momento che $g(a_1, 0, y) = a_1 > 0$, per ogni y, dice che questo minimo b non esiste.

Supponiamo allora $a_1 > 0$ e $a_2 > 0$. Esiste un minimo n tale che $a_1 n \geq a_2$ e $n > 0$. È evidente allora che, per ogni $y < n$, $a_1 y < a_2$, cioè $g(a_1, a_2, y) = a_2 - a_1 y > 0$. Se invece $m > n$, abbiamo $g(a_1, a_2, n) = 0$. Dunque $\mu g(a_1, a_2) = n$.

In ogni caso, $\mu g(a_1, a_2)$ è il minimo intero n tale che $a_1 n \geq a_2$, tranne quando un tale n non esiste, cioè per $a_1 = 0$ e $a_2 > 0$.

L'introduzione di questo “operatore” μ è giustificata dal principio del minimo: ogni insieme non vuoto di numeri naturali ha minimo. Perciò, se

$$\{\, n \in \mathbf{N} : \text{esiste } y < n \text{ con } g(\mathbf{a}, y) = 0 \text{ oppure } g(\mathbf{a}, y) = \infty \,\}$$

non è vuoto, esso ha minimo. Questo minimo non può essere 0 e quindi sarà scrivibile come $b + 1$. Ma allora, per definizione,

$$\text{per ogni } y, \text{ se } y < b, \text{ allora } g(\mathbf{a}, y) > 0 \text{ e } g(\mathbf{a}, y) \neq \infty.$$

5.6. **Esempi.** (1) *Permutazioni e identificazioni di variabili.* Sia $g\colon \mathbf{N}^k \to \mathbf{N} \cup \{\infty\}$ una funzione ricorsiva e, per $1 \leq i \leq n$, sia j_i sia un naturale $1 \leq j_i \leq k$. Allora

$$f = g(I_{k,j_1}, I_{k,j_2}, \ldots, I_{k,j_n})$$

è una funzione ricorsiva. In pratica stiamo calcolando

$$f(x_1, x_2, \ldots, x_n) = g(x_{i_1}, x_{i_2}, \ldots, x_{j_k})$$

Quindi ci sentiremo liberi di eseguire permutazioni e identificazioni di variabili senza usare la notazione più complicata.

(2) *La somma.* Si prenda $g = I_{1,1}$ e $h(x_1, x_2, x_3) = S(x_3)$ (si scriva h come composizione); la funzione σ definita per ricorsione ha le seguenti proprietà:

$$\sigma(x_1, 0) = I_{1,1}(x_1) = x_1$$
$$\sigma(x_1, S(x_2)) = h(x_1, x_2, \sigma(x_1, x_2)) = S(\sigma(x_1, x_2))$$

(3) *La funzione $Z_n\colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$, $Z_n(\mathbf{x}) = 0$ è ricorsiva.* Infatti è ottenibile per composizione con la funzione zero.

(4) *Il prodotto.* Si prenda $g = Z_1$ e $h(x_1, x_2, x_3) = \sigma(x_1, x_3)$. Allora

$$\pi(x_1, 0) = g(x_1) = 0$$
$$\pi(x_1, S(x_2)) = h(x_1, x_2, \pi(x_1, x_2)) = \sigma(x_1, \pi(x_1, x_2))$$

(5) *Elevamento a potenza e fattoriale.* Esercizi. È a volte comodo evitare il riferimento alle g e h; la definizione usuale per la potenza è

$$x^0 = 1, \quad x^{S(y)} = \pi(x^y, x).$$

Scrivere le funzioni g e h necessarie.

(6) $\delta(x) = x - 1$ se $x > 0$, altrimenti $\delta(x) = 0$. Si ottiene per ricorsione con $\delta(0) = 0$ e $\delta(S(x)) = x$.

(7) $\mu(x_1, x_2) = x_1 - x_2$ se $x_1 - x_2 \geq 0$, altrimenti $\mu(x_1, x_2) = 0$. Si pone

$$\mu(x_1, 0) = x_1, \quad \mu(x_1, x_2 + 1) = \delta(\mu(x_1, x_2)).$$

(8) $\alpha(x_1, x_2) = x_1 - x_2$ se $x_1 - x_2 \geq 0$, altrimenti $\alpha(x_1, x_2) = x_2 - x_1$. Basta scrivere

$$\alpha(x_1, x_2) = \sigma(\mu(x_1, x_2), \mu(x_2, x_1)).$$

(9) $\zeta(x) = 0$ se $x = 0$, altrimenti $\zeta(x) = 1$. Per ricorsione $\zeta(0) = 0$, $\zeta(S(x)) = 1$. Possiamo poi porre $\bar{\zeta}(x) = \mu(1, \zeta x)$.

(10) *La funzione minimo.* Definiamola per due variabili:

$$\min(x_1, x_2) = \mu(x_1, \mu(x_1, x_2)).$$

Se poi l'abbiamo definita per n variabili,

$$\min(x_1, \ldots, x_n, x_{n+1}) = \min(\min(x_1, \ldots, x_n), x_{n+1}).$$

(11) *La funzione massimo.* Come prima:

$$\max(x_1, x_2) = \sigma(\mu(x_1, x_2), x_2),$$
$$\max(x_1, \ldots, x_n, x_{n+1}) = \max(\max(x_1, \ldots, x_n), x_{n+1}).$$

(12) *La funzione resto.* Vogliamo scrivere il resto della divisione di x_2 per x_1 e chiamarlo $\rho(x_1, x_2)$. Possiamo porre

$$\rho(x_1, 0) = 0, \quad \rho(x_1, S(x_2)) = \pi(S(\rho(x_1, x_2)), \zeta(\alpha(x_1, S(\rho(x_1, x_2))))).$$

(13) *Numero di divisori.* Consideriamo la funzione D così definita:

$$D(x) = \sum_{y \leq x} \bar{\zeta}(\rho(y, x)),$$

per $x > 0$ e $D(0) = 1$. Si verifichi che è ricorsiva. Abbiamo:

$$D(1) = \sum_{y \leq 1} \bar{\zeta}(\rho(y, 1)) = 1,$$
$$D(2) = \sum_{y \leq 2} \bar{\zeta}(\rho(y, 2)) = 2,$$
$$D(3) = \sum_{y \leq 3} \bar{\zeta}(\rho(y, 3)) = 2,$$
$$D(4) = \sum_{y \leq 4} \bar{\zeta}(\rho(y, 4)) = 3.$$

Ne segue che l'essere un numero primo è una proprietà esprimibile ricorsivamente.

5.7. **Relazioni ricorsive.** Quello di prima è un esempio di proprietà ricorsiva. Più in generale possiamo definire ricorsiva una relazione n-aria R quando la sua funzione caratteristica è ricorsiva.

5.8. **Sommatorie.** Abbiamo usato un simbolo che sembrerebbe vietato, quello di sommatoria. In realtà lo possiamo definire ricorsivamente:

$$\sum_{y<0} f(x_1, \ldots, x_n, y) = 0$$
$$\sum_{y<z+1} f(x_1, \ldots, x_n, y) = \sigma\Big(\sum_{y<z} f(x_1, \ldots, x_n, y), f(x_1, \ldots, x_n, z)\Big)$$
$$\sum_{y \leq z} f(x_1, \ldots, x_n, y) = \sum_{y<z+1} f(x_1, \ldots, x_n, y)$$

5.9. **Esistono funzioni non ricorsive?** La risposta è sì. Infatti

> *l'insieme delle funzioni ricorsive con dominio* $\mathbf{N}^n$ *è numerabile.*

Il modo di mostrare questo fatto è di "esprimere ogni funzione ricorsiva" nel linguaggio formale dell'aritmetica, cioè associare a ogni funzione ricorsiva una formula del linguaggio in modo iniettivo. Siccome le formule sono numerabili, anche le funzioni ricorsive lo sono.

I dettagli sono troppo complicati per essere trattati qui, ma l'idea di fondo non dovrebbe essere difficile: le funzioni iniziali sono esprimibili certamente con una formula; le regole di costruzione delle funzioni ricorsive (che sono a loro volta ricorsive), possono essere formalizzate nell'aritmetica.

Calcolabilità

Descriviamo una macchina ideale che ci permetta di calcolare le funzioni ricorsive. La chiameremo *MI*.

Questa macchina possiede un *registro* R_i per ogni numero naturale, nel quale può essere memorizzato un numero naturale. Non è rilevante sapere come. Diremo che i è l'*indirizzo* del registro R_i.

C'è anche un *contatore* K, che in ogni momento contiene un numero naturale.

Un registro o il contatore è *vuoto* se contiene il numero 0. *Cancellare* un registro o il contatore significa scriverci 0.

L'ipotesi di infinitezza del numero di registri non è irrealistica: in ogni calcolo solo un numero finito di registri sarà non vuoto. Possiamo quindi pensare che in caso di bisogno si possano aggiungere le celle di memoria necessarie. La macchina è comunque ideale, esistono limitazioni fisiche alla sua grandezza.

Un *programma* è una successione finita di *comandi*, scelti fra la lista che segue. Se

$$P = \langle \mathsf{C}_0, \mathsf{C}_1, \ldots, \mathsf{C}_{h-1} \rangle$$

è un programma, la *posizione* del comando C_i è i. Daremo poi alcune condizioni che questa successione deve soddisfare.

Comandi Z: il comando Z_i (i è un numero naturale) cancella il registro R_i e aggiunge 1 al contatore K; notazione intuitiva: $R_i := 0$;

Comandi A: il comando $A_{i,j}$ (i e j sono numeri naturali) assegna il contenuto del registro R_j al registro R_i e aggiunge 1 a K; notazione intuitiva: $R_i := R_j$;

Comandi S: il comando S_i (i è un numero naturale) aggiunge 1 a R_i e a K; notazione intuitiva: $R_i := R_i + 1$;

Comandi W: il comando $\mathsf{W}_{i,j,k}$ (i, j e k sono numeri naturali) viene eseguito così: indicando con r_i e r_j i contenuti di R_i e R_j rispettivamente,
- se $r_i = r_j$ si cerca nella successione dei comandi da eseguire il comando E_k e si pone nel contatore la sua posizione aumentata di 1;
- se $r_i \neq r_j$ si aggiunge 1 a K;

notazione intuitiva: **while** $R_i \neq R_j$ **do**;

Comandi E: il comando E_k (k è un numero naturale) cerca la posizione del comando $\mathsf{W}_{i,j,k}$ e pone nel contatore la sua posizione; notazione intuitiva **end while**.

Le condizioni da soddisfare perché P sia un programma sono:

(1) per ogni k c'è al più un comando $\mathsf{W}_{i,j,k}$ e, in tal caso, c'è esattamente un comando E_k e lo segue;
(2) il comando $\mathsf{W}_{i,j,k}$ precede il comando $\mathsf{W}_{i,j,l}$ se $k < l$;
(3) se tra il comando $\mathsf{W}_{i,j,k}$ e il comando E_k c'è un comando $\mathsf{W}_{i,j,l}$ (e quindi $k < l$), allora il comando E_l precede E_k;
(4) se tra il comando $\mathsf{W}_{i,j,k}$ e il comando E_k c'è un comando E_l, allora il comando $\mathsf{W}_{i,j,l}$ corrispondente precede E_k (e quindi $k < l$);

Come si vede queste condizioni riguardano solo i cicli "while", che devono essere "annidati" oppure "disgiunti". Possiamo anche dire che la successione dei comandi che compaiono fra un $\mathsf{W}_{i,j,k}$ e il corrispondente E_k è essa stessa un programma.

Come si comporta *MI*? Quando viene inserito un programma, essa esegue i comandi che vi compaiono, leggendo il contatore. Quando nel contatore c'è il numero n, *MI* esegue il comando di posto n; ma se n è maggiore della lunghezza del programma, allora *MI* si spegne.

Possiamo certamente supporre che, al momento della partenza, il contatore sia vuoto; se il programma inserito è vuoto, *MI* si spegne, altrimenti esegue il comando C_0. Gli unici registri che possono essere modificati nel processo sono quelli il cui indirizzo compare come indice di comandi Z, S, A o anche fra i primi due indici dei comandi W.

Non facciamo ipotesi sul contenuto iniziale dei registri; ogni programma può infatti essere usato come sottoprogramma di un altro, per esempio come ciclo "while". Non è difficile convincersi che due programmi che differiscono solo per gli indici dei comandi W ed E hanno lo stesso effetto finale, partendo dalla stessa configurazione iniziale dei registri. Quindi non è complicato capire come si possano *concatenare* due programmi: basta modificare opportunamente gli indici dei comandi del secondo, per esempio aggiungendo a ciascuno l'indirizzo massimo dei registri che compaiono nel primo e modificando in modo ovvio gli indici dei cicli "while".

Esempio. Il programma il cui effetto è di copiare il numero memorizzato in R_j nel registro R_i ($i \neq j$) è $\langle \mathsf{A}_{i,j} \rangle$. Lo stesso effetto si ottiene con il programma

$$\langle \mathsf{Z}_i, \mathsf{W}_{i,j,0}, \mathsf{S}_i, \mathsf{E}_0 \rangle.$$

Come possiamo calcolare una funzione? Supponiamo che P sia un programma e consideriamo una n-upla $(a_1, \ldots, a_n) \in \mathbf{N}^n$. Possiamo scrivere i

numeri a_i nei registri R_i ($1 \leq i \leq n$) e avviare l'esecuzione. Se, quando la macchina si ferma, nel registro R_0 c'è il numero b, poniamo

$$f_P(a_1, \ldots, a_n) = b.$$

Se l'esecuzione non si arresta, poniamo $f_P(a_1, \ldots, a_n) = \infty$ (trascuriamo il problema della determinazione dell'arresto). Diremo allora che P *calcola* la funzione f_P. Una funzione $f\colon \mathbf{N}^n \to \mathbf{N} \cup \{\infty\}$ è *calcolabile* se esiste un programma che la calcola.

Esempio. La funzione $\sigma\colon (x_1, x_2) \mapsto x_1 + x_2$ è calcolabile dal programma

$$\langle \mathsf{A}_{0,1}, \mathsf{Z}_3, \mathsf{W}_{2,3,0}, \mathsf{S}_0, \mathsf{S}_3, \mathsf{E}_0 \rangle$$

che azzera i registri R_0 e R_3 e, fino a che il contenuto di R_3 è uguale al contenuto di R_2, somma 1 sia a R_0 che a R_3.

Le funzioni ricorsive iniziali sono tutte calcolabili:

- la funzione Z dal programma $\langle \mathsf{Z}_0 \rangle$;
- la funzione S dal programma $\langle \mathsf{A}_{0,1}, \mathsf{S}_0 \rangle$;
- la funzione $I_{n,i}$ dal programma $\langle \mathsf{A}_{0,i} \rangle$.

Supponiamo che g, h_1, …, h_k siano calcolabili e sia $f = g(h_1, \ldots, h_k)$ (con gli opportuni domini). Siano P, P_1, …, P_k programmi che le calcolano. L'unico vero problema che abbiamo è di conservare il risultato che producono e fare in modo che i registri da 1 a n rimangano intoccati.

Si tratta di specificare le operazioni necessarie: come sempre, è più facile avere l'idea intuitiva di come fare che scrivere i dettagli. Per farla breve, possiamo immagazzinare i valori $h_i(a_1, \ldots, a_n)$ invece che nel registro R_0 in k registri opportuni e usare quelli per il calcolo di

$$f(h_1(a_1, \ldots, a_n), \ldots, h_k(a_1, \ldots, a_n)).$$

Passiamo invece allo schema di ricorsione

$$f(\mathbf{x}, 0) = g(\mathbf{x})$$
$$f(\mathbf{x}, S(y)) = h(\mathbf{x}, y, f(\mathbf{x}, y))$$

supponendo di avere programmi G e H che calcolano g e h. Bisogna allora applicare la funzione h a valori crescenti di y e dell'ultima variabile che all'inizio sono dati da g e successivamente dai valori della stessa h. Quindi riserveremo un registro per contare le iterazioni e un altro in cui porre i successivi valori da sostituire nell'ultima variabile. Il fatto che esistano i comandi W ed E ci permette di eseguire le iterazioni.

Con la minimizzazione il discorso è analogo.

Viceversa

La macchina *MI* permette (in teoria) di calcolare tutte le funzioni ricorsive. Ogni comando di *MI* è realizzabile in un linguaggio di programmazione sufficientemente ricco: assegnazioni di valori dati a registri, azzeramento di registri, sommare uno al contenuto di un registro, cicli "while" sono realizzabili facilmente o già forniti.

Di fatto è facile convincersi che ogni altra funzione predefinita in un linguaggio di programmazione è costruibile a partire da questi. Ovviamente scrivere un programma nel linguaggio spartano che abbiamo usato fin qui è

un esercizio puramente teorico: i linguaggi di alto livello sono nati proprio per togliere al programmatore la fatica di dover inventare l'acqua calda (o quella fredda, talvolta) per ogni compito debba svolgere.

Alcuni linguaggi non possiedono il ciclo "while", ma usano invece i condizionali. Può essere utile pensare come realizzare un ciclo "while" avendo solo a disposizione "if". (Suggerimento: si usa una ricorsione!)

Del resto i linguaggi di alto livello ammettono una traduzione in "linguaggio macchina" e questo ha di solito comandi del tipo di quelli usati prima. La situazione è apparentemente complicata dal fatto che in un calcolatore il programma stesso risiede in alcuni registri della macchina, ma questo è del tutto secondario anche se pone qualche domanda interessante.

Non è inconcepibile un programma che possa modificare sé stesso, dal momento che il programma risiede sui registri della macchina. Tuttavia ciò che accade è analogo alla nostra *MI*: le *istruzioni* seguite sono astratte e non risiedono in alcun posto; il calcolatore esegue in successione le istruzioni che comunque sono del tipo detto. Non bisogna confondere il programma astratto con la lista dei contenuti dei registri.

Più interessante è la dimostrazione che la funzione associata a ogni programma della macchina *MI* è ricorsiva.

Non entreremo in dettagli, ma l'idea è la stessa di Gödel: associare un numero a ogni programma in modo univoco; poiché un programma è una successione finita di comandi, basta assegnare a ogni comando un numero distinto e usare la tecnica di Gödel. La funzione

$$\mathrm{Prog}(x) = \begin{cases} x & \text{se } x \text{ è il numero di un programma,} \\ 1 & \text{se } x \text{ non è il numero di un programma,} \end{cases}$$

è chiaramente ricorsiva. Se x non è il numero di un programma, possiamo convenire che $\{x\}$ sia il programma vuoto, altrimenti $\{x\}$ indica il programma di cui x è il numero.

Anche a ogni stato della macchina (cioè a ogni configurazione dei registri) possiamo associare un numero:

$$2^k 3^{r_0} 5^{r_1} 7^{r_2} \dots p_h^{r_{h-1}} \cdots$$

sfruttando il fatto che i registri sono tutti vuoti tranne un numero finito; k è il contenuto di K, r_i quello del registro R_i.

Possiamo allora definire una funzione Next: $\mathrm{Next}(x, y)$ è il numero corrispondente allo stato successivo della macchina quando il programma in esecuzione è $\{x\}$ e lo stato attuale è y. Questa funzione è ricorsiva.

A partire da essa non è troppo complicato dimostrare che le funzioni calcolabili da *MI* sono ricorsive.

La tesi di Church, finalmente

Supponiamo di saper arricchire la dotazione di comandi che la macchina *MI* può eseguire. L'insieme delle funzioni calcolabili aumenterà?

La tesi di Church è che questo non è possibile, se questi comandi sono formulabili in termini di "algoritmi", cioè secondo la nostra intuizione di ciò che è un algoritmo.

Se il linguaggio è più ricco, possiamo assegnare numeri ai nuovi comandi, la funzione Prog cambierà, ma rimarrà ricorsiva. Possiamo pensare che la nuova funzione Next non sia ricorsiva?

Bisognerebbe entrare nel dettaglio della definizione di Next e ci si accorgerebbe che per renderla non ricorsiva ci vorrebbero comandi di tipo del tutto inconcepibile al momento attuale.

Di fatto tutte le possibili definizioni proposte di calcolabilità risultano equivalenti (mediante una *dimostrazione*).

La tesi di Church non è dimostrabile, perché richiederebbe di precisare il concetto di algoritmo e a quel punto sarebbe nient'altro che una nuova definizione di calcolabilità, equivalente alle precedenti, visto che non abbiamo ancora trovato definizioni di algoritmo che conducano fuori dalle funzioni ricorsive.

La dimostrazione della ricorsività delle funzioni calcolabili da *MI* tramite la tecnica di associare numeri di Gödel è un punto a favore della tesi.